STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING
Contents

Data visualization tip sheet
Module 1. Introduction to data visualization
Module 2. Data visualization principles and concepts
Module 3. Excel data and charts adjustments
Module 4. Visualization tools comparison
Module 5. Excel chart transformations
Module 6. DATAWRAPPER maps
Module 7. DATAWRAPPER range plots
Module 8. Annotations
<table>
<thead>
<tr>
<th>Chart type</th>
<th>When to use</th>
<th>Chart-specific tips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column chart (vertical bar)</td>
<td>Mostly for one variable</td>
<td>Should use the same color for all bars (unless specifically disaggregated for example by sex).</td>
</tr>
<tr>
<td></td>
<td>Show comparison/ranking</td>
<td>Overall, Total, or National data points can be highlighted by a different color or transformed from a bar to a highlight line.</td>
</tr>
<tr>
<td></td>
<td><strong>Bar</strong> used when variable/category labels are too long to show horizontally in a column chart</td>
<td>The width of the bars should be about twice the width of the space between the bars.</td>
</tr>
<tr>
<td>Bar chart (horizontal bar)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clustered/grouped bar or column chart</td>
<td>Mostly for breaking/grouping one variable into different levels of disaggregation (subgroups)</td>
<td>Should use the same color for all bars (unless specifically disaggregated for example by sex).</td>
</tr>
<tr>
<td></td>
<td>Show comparison/ranking</td>
<td>Make sure multiple levels of disaggregation or group are shown in the axis to avoid labeling each category name individually.</td>
</tr>
<tr>
<td></td>
<td><strong>Bar</strong> used when variable/category labels are too long to show horizontally in a column chart</td>
<td>The gap between the clusters/groups of bars should be half the size of the width of the bar itself.</td>
</tr>
<tr>
<td>Stacked bar or column chart</td>
<td>Show the composition usually out of 100% (part-to-whole relationship of a variable’s categories)</td>
<td>Should have no more than 8 categories.</td>
</tr>
<tr>
<td></td>
<td>Can be simple (one variable) or clustered/grouped (different levels of disaggregation or subgroups)</td>
<td>Legends should be stretched across the top of the chart or to the right, and the order should match the order in the chart.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Should use categorical color palette.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The “Other” category should not be the largest section of the stacked chart. Try breaking the “other” section into smaller sections.</td>
</tr>
</tbody>
</table>
### Line Chart
- Show the trend in variables usually over time (but can also be age ranges, or hours in a day).
- Show multiple variables with multiple lines (if they are on the same scale).
- Show the same variable for multiple observations with multiple lines.
- No more than 4 to 6 lines in a chart.
- When too many lines, break into "small multiples" or "grid of charts" highlighting the lines individually or highlighting the main line in one color and all other lines behind in light grey.
- Option to shade area between male and female line to better show gaps.
- When possible, directly label series. If lines are too close together, use a legend.
- Avoid individual data labels. Individual data labels may be acceptable if there are few data points and data labels aren’t overlapping.

### Area Chart
- Show the trend in composition (part-to-whole relationship of categories) over time. It is a combination of the line chart and stacked column chart.
- Should use categorical color palette.
- The “Other” category should not be the largest section of the stacked chart. Try breaking the “other” section into smaller sections.
- X-axis should have an ordinal variable (time, age, hours, temperature, etc.).

### Pie Chart or Donut Chart
- Show the composition out of 100% (part-to-whole relationship of categories) typically for 3 or 4 but preferably no more than 6 categories.
- Avoid pie charts whenever possible and use stacked charts instead.
- If it is necessary to use a pie chart, there should be no more than 6 slices.
- The “Other” slice should not be the largest slice of the pie chart.

### Range Plot
- Show comparison/ranking
- Show gaps and absolute values
- Order the range plot by value for ease of comprehension.
- Gridlines should always be removed to avoid the length of the gap blending into the gridline.
### Heatmap
- Show comparison with multiple levels or disaggregation with many subgroups or categories
- Most commonly shown over time (hours, years) or across age groups
- Pack a lot of information in one chart
- For continuous data, use sequential color palette or if showing the gender gap, use a diverging color palette.
- X-axis should have an ordinal variable (time, age, hours, temperature, etc.)

### Scatterplot
- Show relationship between two variables
- Could show the relationship between male and female values of the same variable or between two years of same variable.
- Show distribution, outliers
- Mostly for microdata but could plot disaggregation category with many subgroups (i.e., geographic regions, occupations)
- Is typically reserved for visualizing data with many observations (i.e., microdata).
- Add a 45-degree line to show equal values between the two variables. For example, gender parity when plotting male and female values on the axes.
- Using a third variable as dot (bubble) size creates a bubble chart.

### Dot plot
- Show distribution of the whole set of observations
- Easily identify outliers
- Mostly for microdata but could plot disaggregation category with many subgroups (i.e., geographic regions, occupations)
- Is typically reserved for visualizing data with many observations (i.e., microdata).
- Can be shown vertically or horizontally.
- Can be colored by one of the disaggregation categories. For example, plotting all households from microdata and coloring by urban/rural.

### Map
- Show geographic data much more visually than through the aforementioned chart types.
- Can be shown at whichever level data are available, national, administrative regions, provinces, districts, cities, etc.
- Should use sequential colors unless showing a gap which requires a diverging color scale.
- Must use the same color scale and range/steps for male and female for proper side by side comparison.
- For a gap, use text in the legend instead of numbers so as not to confuse the reader with negative numbers. Legend should show parity in the middle and female values higher than male values at one end and vice versa for the opposite end of the scale.
DATA VISUALIZATION CHECKLIST

Overall
Graphs will catch a viewer's attention so only visualize the data that needs attention. Too many graphics of unimportant information dilute the power of visualization.

Does the visual highlight a significant gender-relevant finding or conclusion? If not, can the data structure be adjusted to highlight the gender-relevant insight or is the data not gender-relevant?

Total or aggregate statistics can often be removed from tables and graphs to facilitate comparisons between women and men. If the totals are large, choose a chart type that highlights overall and gender-disaggregated statistics.

Is the type of visual appropriate for the data?
For example, change over time is displayed as a line graph, area chart, slope graph, or dot plot.

Does the visual have appropriate level of precision?
Use a level of precision that meets your audiences' needs. Few numeric labels need decimal places, unless you are speaking with academic peers. Charts intended for public consumption rarely need p values listed.

Do the individual visual elements work together to reinforce the overarching takeaway message? Are any of the visual elements duplicating information?
Choices about graph type, text, arrangement, color, and lines should reinforce the same takeaway message without duplication.

Color
Keep culture-laden color connotations in mind.

Use sites like Color Brewer to find color schemes suitable for reprinting in black-and-white and for colorblindness.

Is the color scheme intentional? Is there a consistent color for male and female?
Colors should be derived from an intentional choice, not the default color schemes. Use your organization's colors.
Avoid colors and figures that reinforce gender stereotypes such as pink for women/girls and blue for men/boys.
Always use one color to represent women and another color to represent men. These colors should not be two different shades of the same color hue (i.e., light blue and dark blue). They should be two distinct shades of color (i.e., blue and orange).

Does the color highlight key patterns?
Action colors should guide the viewer to key parts of the display. Less important, supporting, or comparison data should be a muted color, like gray.

Is the color legible when printed in black and white or for people with colorblindness?
When printed or photocopied in black and white, the viewer should still be able to see patterns in the data. Avoid red-green and yellow-blue combinations when those colors touch one another. Avoid using red to mean bad and green to mean good in the same chart.
Data Visualization Checklist

Color (continued)

For accessibility, don’t rely solely on color to convey the idea. As much as possible, the presentation of the data should still be understandable if the color is removed or there should be enough additional text that identifies the key ideas in the table for the visually impaired audience who can read the report with a screen reader.

**Does the text sufficiently contrast the background?**
Black/very dark text against a white/transparent background is easiest to read.

**Lines**

Excessive lines—gridlines, borders, tick marks, and axes—can add clutter or noise to a graph, so eliminate them whenever they aren’t useful for interpreting the data.

**Are there gridlines? If yes, are they muted?**
Color should be faint gray, not black. There should be a preference for no gridlines. Gridlines, even muted, should not be used when the graph includes numeric labels on each data point.

**Does the visual have a border line?**
Graph should bleed into the surrounding page or slide rather than being contained by a border.

**Do the axes have unnecessary tick marks or axis lines?**
Tick marks can be useful in line graphs (to demarcate each point in time along the y-axis) but are unnecessary in most other graph types. Remove axes lines whenever possible.

**Does the graph have one horizontal and one vertical axis?**
Viewers can best interpret one x-axis and one y-axis. Don’t add a second y-axis. Try a connected scatter plot or two graphs, side by side, instead. (A secondary axis used to hack new graph types is ok, so long as viewers aren’t being asked to interpret a second y-axis.)

Arrangement

Improper arrangement of graph elements can confuse readers at best and mislead viewer at worst. Thoughtful arrangement makes a data visualization easier for a viewer to interpret.

**Are the proportions accurate?**
A viewer should be able measure the length or area of the graph with a ruler and find that it matches the relationship in the underlying data. Y-axis scales should be appropriate. Bar charts start axes at 0. Other graphs can have a minimum and maximum scale that reflects what should be an accurate interpretation of the data.

**Are the data intentionally ordered or sorted?**
DATA VISUALIZATION CHECKLIST

Arrangement (continued)

Data should be displayed in an order that makes logical sense to the viewer. Data may be ordered by frequency counts (e.g., from greatest to least for nominal categories), by groupings or bins (e.g., histograms), by time period (e.g., line charts), alphabetically, etc. Use an order that supports interpretation of the data.

As a general rule, the chart should be ordered by value (for sex-disaggregated data typically by female value). Female values should always appear first in a graph compared to male values (above or to the left of the values for men) – this is the case for bars or columns, for the order of tables, and also for legends.

Are axis intervals equidistant? If not, is there a symbol that signifies the jump?

The spaces between axis intervals should be the same unit, even if every axis interval isn’t labeled. Irregular data collection periods can be noted with markers on a line graph, for example. Some charts should not start at 0 or we will not see a significant pattern, but there should at least be a symbol to make it clear that it does not start at 0).

Is the graph two-dimensional?

Avoid three-dimensional displays, bevels, and other distortions.

Text

Graphs don’t contain much text, so existing text must encapsulate your message and pack a punch.

Do subtitles and/or annotations provide additional information?

Subtitles and annotations (call-out text within the graph) can add explanatory and interpretive power to a graph. Use them to answer questions a viewer might have or to highlight specific data points.

Is the text size hierarchical and readable?
DATA VISUALIZATION CHECKLIST

Is the text horizontal?
Titles, subtitles, annotations, and data labels are horizontal (not vertical or diagonal). Line labels and axis labels can deviate from this rule and still receive full points. Consider switching graph orientation (e.g., from column to bar chart) to make text horizontal. For cultures that write vertically, make sure to follow cultural language practices instead.

Are data labeled directly?
Whenever possible, position data or category/series labels near the data rather than in a separate legend (e.g., on top of or next to bars and next to lines). Eliminate/Embed legends when possible because eye movement back and forth between the legend and the data can interrupt the brain’s attempts to interpret the graph.

Are labels used sparingly?
Focus attention by removing the redundancy. For example, in line charts, label every other year on an axis. Do not add numeric labels *and* use a y-axis scale, since this is redundant. Don’t clutter the labels, make sure there is sufficient and consistent space between the labels.

Are the numbers adjusted correctly?
Numbers rounded to no more than one decimal place, unless it is absolutely necessary to have more decimal places. When viewing large numbers, display the data in thousands or millions and specify the unit in the subtitle or axis label. The abbreviation can also be used (in English: K for thousands, M for millions, B for billions). Don’t compare fractions with different denominators, use percentages instead.

Does the direction of the text respect cultural and linguistic rules?
Make sure that graphs and tables are aligned according to the rules for right-to-left or top-down languages. Make sure that the correct grammar, punctuation, and capitalization follow the rules of the corresponding language. When using English, make sure that the graphics and text are consistent with only one type of English (American or British English).

Adapted from "Data Visualization Checklist" by Stephanie Evergreen & Ann K. Emery
Anatomy of a visual

Chart Title

- **Axis Title**: Internet Access
- **Axis Label**: Internet Access
- **Legend**: Internet Access Female, Internet Access Male, Internet Access Total
- **Data Label**: For each region, showing the percentage of Internet Access for each category.
- **Gridlines**: Horizontal and vertical gridlines for better readability.
- **Border**: Surrounding the chart for visual separation.
- **3D**: Depiction of the chart as a 3D object, enhancing visual impact.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 1: INTRODUCTION TO DATA VISUALIZATION
Contents

1. What is data visualization?
2. Why is data visualization important?
3. Advantages/disadvantages of data visualization.
4. How are data visualizations disseminated?
5. For whom are data visualizations produced?
6. Practice exercise
1. Introduction to data visualization
What is data visualization?

- Data visualization is the graphical representation of information and data.

- Visual elements:
  - Include charts, diagrams, data visualization attributes (shape, position, color, pattern, etc.), images/icons, annotations (lines, shaded areas, arrows, etc.).
  - Provide an accessible way to see and understand trends, outliers, and patterns in data.

- Data visualization involves human perception & cognition.
  - The human mind is slow with mental operations like multiplying, subtracting, or simply comparing numbers.
  - Data visualizations present information that we can visually perceive to better understand the insights and trends in the data.

"Data analytics is the data representation and presentation that exploits our visual perceptions capabilities to amplify the cognition."

- Andy Kirk, author of "Data Visualization: a successful design process"
What insights does the table provide?  
Can you easily tell which level of education has the highest enrollment?  
Can you easily compare enrollment for different cities? From year 1 to year 2?

<table>
<thead>
<tr>
<th>City</th>
<th>Pre-primary education Y1</th>
<th>Pre-primary education Y2</th>
<th>Primary education Y1</th>
<th>Primary education Y2</th>
<th>Secondary education Y1</th>
<th>Secondary education Y2</th>
<th>Pre-primary enrollment Y1 to Y2 Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>City A</td>
<td>11.1</td>
<td>18</td>
<td>66.4</td>
<td>77</td>
<td>19.9</td>
<td>32.2</td>
<td>6.9</td>
</tr>
<tr>
<td>City B</td>
<td>7.5</td>
<td>15.2</td>
<td>62.8</td>
<td>73.8</td>
<td>14.4</td>
<td>23.1</td>
<td>7.7</td>
</tr>
<tr>
<td>City C</td>
<td>23.6</td>
<td>26.6</td>
<td>77.7</td>
<td>86.8</td>
<td>35.2</td>
<td>59.2</td>
<td>3</td>
</tr>
<tr>
<td>City D</td>
<td>10.9</td>
<td>17.5</td>
<td>66.2</td>
<td>76.8</td>
<td>19.7</td>
<td>31.4</td>
<td>6.6</td>
</tr>
<tr>
<td>City E</td>
<td>17.7</td>
<td>25.8</td>
<td>81.9</td>
<td>92.7</td>
<td>42.1</td>
<td>64.2</td>
<td>8.1</td>
</tr>
<tr>
<td>City F</td>
<td>28.2</td>
<td>26.4</td>
<td>77.1</td>
<td>84.5</td>
<td>33.2</td>
<td>55.4</td>
<td>-1.8</td>
</tr>
<tr>
<td>City G</td>
<td>7.2</td>
<td>14.7</td>
<td>62.6</td>
<td>73.5</td>
<td>13.9</td>
<td>22.8</td>
<td>7.5</td>
</tr>
<tr>
<td>City H</td>
<td>18.5</td>
<td>36.1</td>
<td>71.8</td>
<td>85</td>
<td>27.5</td>
<td>61</td>
<td>17.6</td>
</tr>
<tr>
<td>City I</td>
<td>16.7</td>
<td>32.5</td>
<td>64.6</td>
<td>76.5</td>
<td>24.8</td>
<td>54.9</td>
<td>15.84</td>
</tr>
<tr>
<td>City J</td>
<td>9.7</td>
<td>17.9</td>
<td>62.5</td>
<td>75.6</td>
<td>18.3</td>
<td>29.8</td>
<td>8.2</td>
</tr>
<tr>
<td>City K</td>
<td>12.8</td>
<td>18</td>
<td>70.6</td>
<td>78.5</td>
<td>21.6</td>
<td>34.6</td>
<td>5.2</td>
</tr>
</tbody>
</table>
What insights does the chart provide?

- Can you easily tell which level of education has the highest enrollment?
- Can you easily compare enrollment for different cities? From year 1 to year 2?
- Is it easier to answer these questions compared to the table?
Simplified presentation

1) Too much information is being shared in one chart

2) Non intuitive for the audience to read through

3) Not immediately visible what the data is saying – where is the main insight?
Main goal: highlight trends in pre-primary enrollment over time.
• The left chart can be simplified to show only the pre-primary enrollment → as in the right chart.
Main goal: highlight trends in pre-primary enrollment over time.

- The left chart can be simplified to show only the pre-primary enrollment as in the right chart.

1) The most relevant information is being shared

2) The chart is more readable but still has too many options...
Key insight: top 3 cities where pre-primary enrollment has improved the most over time.

- Can you easily tell which improvement in enrollment has occurred in each city?
- Which are the top 3 most improved cities?
**Simplified presentation**

➢ **Key insight:** top 3 cities where pre-primary enrollment has improved the most over time.
  - Option 1: Highlight the **improvement** for each city sorted by improvement level.
  - Option 2: Highlight **only the 3 top cities** but with absolute numbers for year 1 and 2.

**Improvement in Pre Primary enrollment from Y1 to Y2**

- City H: 17.6
- City I: 15.84
- City J: 8.2
- City E: 8.1
- City B: 7.7
- City G: 7.5
- City A: 6.9
- City D: 6.6
- City K: 5.2
- City C: 3
- City F: -1.8

**Top Changes in Pre Primary enrollment**

- City H: 38.5
- City I: 36.1
- City J: 31.5
- City E: 16.7
- City B: 15.1
- City G: 14.3
- City A: 13.5
- City D: 12.7
- City K: 9.7
- City C: 17.9

- Provides a better sense of how pre-primary enrollment has changed. Top 3 easily identifiable.
- Conveys the exact information we are trying to communicate with no extra details. No need to search top cities.
Why is data visualization important?

The importance is simple – data visualization helps people:

- Whether simple or complex, the right visualization can bring everyone on the same page, regardless of their level of expertise.

- While traditional education typically draws a distinct line between creative storytelling and technical analysis, the modern professional world also values those who can cross between the two: Data visualization sits right in the middle of analysis and visual storytelling.

Goals of a data visualization: to explain, to monitor, and to relay information
<table>
<thead>
<tr>
<th>ADVANTAGES</th>
<th>DISADVANTAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Easily identify trends that could be missed in tables.</td>
<td>• Cannot always discern exact values from the visual or they are not provided.</td>
</tr>
<tr>
<td>• Appropriate use of visual elements (colors/patterns) makes it easier to find patterns and relationships.</td>
<td>• Core messages can get lost.</td>
</tr>
<tr>
<td>• Possibility to explore data in interactive formats through dashboards.</td>
<td>• Wrong design or visual practices can lead to inaccurate representations or biased information.</td>
</tr>
</tbody>
</table>
Advantages/disadvantages of data visualization

Can you quickly identify which city has the largest population?

One has to look through all the values in the table to decipher which one has the largest population.
Advantages/disadvantages of data visualization

➢ Can you quickly identify which city has the largest population?

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>City A</td>
<td>47.1</td>
<td>28.2</td>
</tr>
<tr>
<td>City B</td>
<td>102.0</td>
<td>61.2</td>
</tr>
<tr>
<td>City C</td>
<td>63.8</td>
<td>38.3</td>
</tr>
<tr>
<td>City D</td>
<td>96.0</td>
<td>57.6</td>
</tr>
<tr>
<td>City E</td>
<td>98.0</td>
<td>58.8</td>
</tr>
<tr>
<td>City F</td>
<td>50.2</td>
<td>30.1</td>
</tr>
<tr>
<td>City G</td>
<td>3.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

The chart makes it easier to identify that City B and City E have the highest populations, but the difference is very minimal in the visual and ambiguous.
Advantages/disadvantages of data visualization

➢ Can you quickly identify which city has the largest population?

<table>
<thead>
<tr>
<th>City</th>
<th>Population</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>City A</td>
<td>47.1</td>
<td>28.2</td>
</tr>
<tr>
<td>City B</td>
<td>102.0</td>
<td>61.2</td>
</tr>
<tr>
<td>City C</td>
<td>63.8</td>
<td>38.3</td>
</tr>
<tr>
<td>City D</td>
<td>96.0</td>
<td>57.6</td>
</tr>
<tr>
<td>City E</td>
<td>98.0</td>
<td>58.8</td>
</tr>
<tr>
<td>City F</td>
<td>50.2</td>
<td>30.1</td>
</tr>
<tr>
<td>City G</td>
<td>3.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Using appropriate visual principles like color differentiation and sorting by value provides the information even more quickly.
Advantages/disadvantages of data visualization

Can you quickly identify which city has the largest population?

The table easily pinpoints the largest population by exact data point, while the chart without data labels does not specify.

<table>
<thead>
<tr>
<th></th>
<th>Population</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>City A</td>
<td>47.1</td>
<td>28.2</td>
</tr>
<tr>
<td><strong>City B</strong></td>
<td><strong>102.0</strong></td>
<td><strong>61.2</strong></td>
</tr>
<tr>
<td>City C</td>
<td>63.8</td>
<td>38.3</td>
</tr>
<tr>
<td>City D</td>
<td>96.0</td>
<td>57.6</td>
</tr>
<tr>
<td>City E</td>
<td>98.0</td>
<td>58.8</td>
</tr>
<tr>
<td>City F</td>
<td>50.2</td>
<td>30.1</td>
</tr>
<tr>
<td>City G</td>
<td>3.0</td>
<td>1.8</td>
</tr>
</tbody>
</table>

The chart above shows the incidence of disease for each city. The bars represent the percentage of population and disease for each city.
Advantages/disadvantages of data visualization

➢ Can you quickly identify which city has the largest population?

In this chart type, you can understand the proportion of population with and without disease, but you cannot tell the population of the cities, which vary as seen in the table. The proportions are also very similar so there isn’t any comparable insight by city.
Advantages/disadvantages of data visualization

➢ Can you quickly identify which city has the largest population?

<table>
<thead>
<tr>
<th>Population</th>
<th>Disease</th>
<th>Population</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>City A</td>
<td>47.1</td>
<td>28.2</td>
<td></td>
</tr>
<tr>
<td>City B</td>
<td>102.0</td>
<td>61.2</td>
<td></td>
</tr>
<tr>
<td>City C</td>
<td>63.8</td>
<td>38.3</td>
<td></td>
</tr>
<tr>
<td>City D</td>
<td>96.0</td>
<td>57.6</td>
<td></td>
</tr>
<tr>
<td>City E</td>
<td>98.0</td>
<td>58.8</td>
<td></td>
</tr>
<tr>
<td>City F</td>
<td>50.2</td>
<td>30.1</td>
<td></td>
</tr>
<tr>
<td>City G</td>
<td>3.0</td>
<td>1.8</td>
<td></td>
</tr>
</tbody>
</table>

This chart is incorrectly stacking incidence of disease on population data, adding the total incidence of disease with total population, which doesn’t provide any useful information.
How are data visualizations disseminated?

Report (i.e. Gender Factbook)
- Official statistical output with mostly text and analysis supported by visuals

Complementary executive summary/visual brief
- 1-2 pages long summarizing key visuals and insights from report

Infographic
- An eye-catching collection of images, icons, charts with minimal text

Written blog or data story
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

Social media card or factoid card
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

Interactive visual for social media
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements
How are data visualizations disseminated?

Report (i.e. Gender Factbook)
- Official statistical output with mostly text and analysis supported by visuals

Complementary executive summary/visual brief
- 1-2 pages long summarizing key visuals and insights from report

Infographic
- An eye-catching collection of images, icons, charts with minimal text

Written blog or data story
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

Social media card or factoid card
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

Interactive visual for social media
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements
How are data visualizations disseminated?

Report (i.e. Gender Factbook)
- Official statistical output with mostly text and analysis supported by visuals

Complementary executive summary/visual brief
- 1-2 pages long summarizing key visuals and insights from report

Infographic
- An eye-catching collection of images, icons, charts with minimal text

Written blog or data story
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

Social media card or factoid card
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

Interactive visual for social media
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements
How are data visualizations disseminated?

**Report (i.e. Gender Factbook)**
- Official statistical output with mostly text and analysis supported by visuals

**Complementary executive summary/visual brief**
- 1-2 pages long summarizing key visuals and insights from report

**Infographic**
- An eye-catching collection of images, icons, charts with minimal text

**Written blog or data story**
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

**Social media card or factoid card**
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

**Interactive visual for social media**
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements.
How are data visualizations disseminated?

**Report (i.e. Gender Factbook)**
- Official statistical output with mostly text and analysis supported by visuals

**Complementary executive summary/visual brief**
- 1-2 pages long summarizing key visuals and insights from report

**Infographic**
- An eye-catching collection of images, icons, charts with minimal text

**Written blog or data story**
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

**Social media card or factoid card**
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

**Interactive visual for social media**
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements
How are data visualizations disseminated?

Report (i.e. Gender Factbook)
- Official statistical output with mostly text and analysis supported by visuals

Complementary executive summary/visual brief
- 1-2 pages long summarizing key visuals and insights from report

Infographic
- An eye-catching collection of images, icons, charts with minimal text

Written blog or data story
- An article that tells a narrative/story around the data using visuals for general audiences to better comprehend.
- Interactive data stories include innovative visualizations with narrative.

Social media card or factoid card
- A quick snapshot (1-2 charts) of a key insight which could be taken from a report or created with data completely separate from any other product.

Interactive visual for social media
- An interactive visual, collection of animated visuals, or data story that allows users to interact with the chart, hover over parts of the visual to get exact data values and information and explore the various elements
The target audience for communications regarding gender statistics is policymakers and funders who are key decision-makers as well as the people who influence them like the general public (voters), media, advocacy organizations etc.
Mapping data outputs to audiences

**General public**
- Key figures and visualizations
- Short factsheet or infographic
- Social media graphic or blog

**Media, policymakers**
- Summary tables, charts, trends, visualizations, short stories
- Press release or short factsheet

**Development practitioners, gender specialists in government, NGOs**
- Visuals of both summary and disaggregated data
- Analysis of results including trends over time in standard reports, metadata

**Data analysts, researchers, academia**
- Complete granular raw data, questionnaires, codebooks, etc.
- Dedicated/comprehensive databases of microdata and processed gender statistics
- Advanced visualizations often using statistical concepts

Lower level of detail

Moderate level of detail

High level of detail
Go to Sheet named "Exercise 1" in the Excel file "Training Dataset Day 1".
Review the dataset and answer the following questions:

1. What are the key or most useful insights from the data?
   • Which disaggregations do you wish to communicate in the visualization?

2. Who is/are the audience(s) you are communicating to?

3. Through which type of data output would you disseminate the data visualization?
   • If you have multiple audiences, will there be more than one data output? Which types?

4. What type of potential actions can be taken based on the insights in the data visualization?

Time for exercise: 15 minutes
Write your answers in the Sheet named "Ex1 Response".
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 2: DATA VISUALIZATION PRINCIPLES & CONCEPTS
Contents

1. Data visualization overall principles
2. Data visualization steps
3. Data visualization tips and concepts by chart type
1. Overall data visualization principles
Data visualization improves communication

**GUIDING PRINCIPLES**

1. Understand the information/data and prioritize the information for sharing.

2. Choose the correct software/platform and chart type for effective visuals designs keeping in mind the audience, data, output type.

3. Use effective data visualization concepts and characteristics (color, shape, size, pattern, etc.) to maximize the impact of the data.

4. Communicate data meaning clearly, quickly and ethically.

5. Always integrate text elements with the graphs and images to tell a story (consider infographics).

**BASIC RECOMMENDATIONS**

1. Highlight 1-2 key messages per visual.

2. Explore graph and chart design choices beyond the default options.

3. Recognize the right use of visual characteristics especially color as it is incredibly powerful to depict meaning.

4. Reduce the clutter and keep only essential elements (no duplicative visual characteristics)

5. Never mislead the audience/manipulate visual.

6. Use annotations, minimize jargon, acronyms, and technical terms, and choose a font that is easy to read.

**BENEFITS**

Decision-makers more quickly absorb gender-related insights from surveys which in turn facilitates evidence-based policy-making
1. Understand and show the data

➢ To create a great visualization, you must understand the key insights and how to let the data shine through the visual.
➢ Auto-generated charts will not automatically or adequately highlight the important aspects of the data.
➢ You must explore the data first to understand which insights to relay to audience.
  • Do not include too many variables or too much information. You do not have to relay all the data. Other data can go in a table in annex.
  • Include helpful annotations.
2. Select the right type of graph

Match the graph type to the audience’s level of data/statistics comprehension.

- **Non-expert audiences**
  - Bar and column chart (grouped or stacked), line chart, area chart, geographic map, range plot, pie chart

- **Medium expertise**
  - Heatmap, tree map, histogram, distribution chart, dot plot, complex tables

- **Expert audiences**
  - Scatter plots, boxplots, Sankey or alluvial diagram, Gantt chart
2. Select the right type of graph

Match the graph to the type of data and the message you want readers to gain.

- Categorical (bar, column, treemap), continuous (scatterplot, heatmap), time series (line, area chart, heatmap).

**Comparison**
- Showing how values are similar or different, highlighting gaps

**Composition**
- Parts of a total that usually add up to 100%

**Relationship**
- How variables relate to one another

**Distribution**
- Where values fall within the dataset, identifying outliers
3. Select the data visualization characteristics

- Our eyes detect changes or highlights through visual perception of the following characteristics.

<table>
<thead>
<tr>
<th>Shape</th>
<th>Enclosure</th>
<th>Line Width</th>
<th>Saturation</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Shape Image" /></td>
<td><img src="image2" alt="Enclosure Image" /></td>
<td><img src="image3" alt="Line Width Image" /></td>
<td><img src="image4" alt="Saturation Image" /></td>
<td><img src="image5" alt="Color Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Size</th>
<th>Markings</th>
<th>Orientation</th>
<th>Position</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image6" alt="Size Image" /></td>
<td><img src="image7" alt="Markings Image" /></td>
<td><img src="image8" alt="Orientation Image" /></td>
<td><img src="image9" alt="Position Image" /></td>
<td><img src="image10" alt="3D Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Length</th>
<th>Curvature</th>
<th>Density</th>
<th>Closure</th>
<th>Sharpness</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image11" alt="Length Image" /></td>
<td><img src="image12" alt="Curvature Image" /></td>
<td><img src="image13" alt="Density Image" /></td>
<td><img src="image14" alt="Closure Image" /></td>
<td><img src="image15" alt="Sharpness Image" /></td>
</tr>
</tbody>
</table>
4. Make graphs clear and clean

Declutter but don’t oversimplify chart!
• When possible, reduce numbers to simplest form.
• Label clearly, specify units, use a legend when necessary.
• Remove duplicated characteristics (gridlines vs. numeric labels; color vs. data labels).
• Avoid 3D graphics and shadows. They distort data and it’s not easy to visually see differences between the chart cylinders, columns etc.
• Break the chart into smaller multiples if the original is too cluttered.
5. Integrate text elements or annotations

➢ Text can highlight a particular data point or guide the audience through the content.
  • This should be in addition to the supplementary text/context in a report that refers to the visual.
6. Choose colors and fonts wisely

➢ Colors should aid understanding, not distract readers.
  • Try a maximum of 3 colors (unless categorically required).
  • Stick with the same fonts and colors consistently.
  • Do not use “familiar colors in surprising ways” i.e. red for good and green for bad.
  • Check whether the colors and contrast work for color blindness and for printing black & white - you may not see the data labels on the bars.

➢ Gender data and color
  • Do not use stereotypical male and female colors (blue/pink).
  • Stick with the same colors for variable groups consistently throughout a product or report.
    • Female and Male or Urban, Rural, Total.
6. Choose colors and fonts wisely

- **Colors should aid understanding, not distract readers.**
  - **Sequential** or gradual color scales are usually for continuous variables. In some cases, they work for stacked bar/column charts.
  - **Diverging** color scales are usually for gender gaps where the values are moving away from 0 in either direction.
  - **Categorical** color scales should be used for multiple categories or variables, but never be used for a different color for each observation within a category.
  - **Missing data** or “other” categories should be a neutral color like grey or grey patterned for maps.
GROUP DISCUSSION

➢ Discussion – which colors are used for male and female in this report?
➢ Are they consistent?
➢ Is there anything you would change about the colors of these visuals?

![Graphs showing gender distribution in education levels and average achievement scores in mathematics, reading, and natural sciences by sex.]

**Note:** The total number of teaching staff does not correspond to the actual number of natural persons, considering that teachers and assistants may teach at two or more education institutions.

In Bosnia and Herzegovina, almost all educators in pre-school institutions are women (99%). Going towards higher levels of education, the gender structure of teaching staff is changing in a way that women are more common in primary and secondary education, while more than a half of teachers and assistants in the institutions of higher education are men.
GROUP DISCUSSION

Employees in all education levels in the school year 2020/2021

Note: The total number of teaching staff does not correspond to the actual number of personnel, considering that teachers and assistants may teach at two or more education institutions.

In Bosnia and Herzegovina, almost all educators in pre-school institutions are women (99%). Going towards higher levels of education, the gender structure of teaching staff is changing in a way that women are more common in primary and secondary education, while more than a half of teachers and assistants in the institutions of higher education are men.

Average achievements in mathematics, reading and natural sciences by sex in 2016, according to PISA survey (scores)

Vocational education is designed for learners to acquire specific knowledge, skills and competences about a particular occupation or trade. Vocational education includes vocational, technical, religious and art secondary schools. 

General education includes grammar schools.

Reading
GROUP DISCUSSION

➢ Discussion – which colors are used for male and female in this report?
➢ Are they consistent?
➢ Is there anything you would change about the colors of these visuals?
Do not reuse the Female & Male colors for other categories! These two colors should remain distinct in order to not confuse the reader throughout the report or infographic.
7. Avoid misleading the audience

➢ Do not allow for jumps in axis labels.
   • Horizontal axis must have consistently or equivalently spaced years. For example, one cannot skip from 2000 to 2010 and then yearly.
7. Avoid misleading the audience

- Do not allow for jumps in axis labels.
  - If vertical axis does not start at 0 – MUST have a symbol to denote the break in axis.

- Do not add values in a stacked bar or column or in a pie chart that do not total 100% For example, stacking female, male, and total together.

- Do not use inconsistent scales.
8. Highlighting the gender-relevant insights

- NSOs' comparative advantage regarding statistics → highlight detailed levels of disaggregation that international organizations don’t calculate or disseminate.
  1. **Use multiple disaggregations in the same charts whenever possible.**
     - Aesthetically more pleasing than just 2 bars for male and female and provide more gender-relevant insights that tell a more nuanced story.
  2. **Use subnational region maps.**
     - International databases cannot showcase this subnational level as it’s not internationally comparable.
8. Highlighting the gender-relevant insights

➢ Aim to highlight the gender gap rather than the distribution of female and male values across a different category (age, sector, type of employment, etc.).

**Showcasing** the distribution of female/male employment across sectors

**Key insight:** Most women and men work in the agricultural sector

**Employment by sex and sector (% of employment) (modeled ILO estimate)**

- Agriculture
- Industry
- Services

**Showcasing** the gender gap

**Key insight:** Gender gap in employment is biggest in the industrial sector

**Employment by sex and sector (% of employment) (modeled ILO estimate)**

- Female
- Male

- Agriculture
- Industry
- Services
2. Data visualization steps
Steps for creating a data visualization

1. **Upload data**
2. **Explore/try visualization options**
3. **Adjust data structure**
4. **Design, format, and finalize visualization**
5. **Publish and/or download visualization**

Optional: add additional annotations as needed
Data visualization breakdown

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

75%
Properly cleaning, transposing & preparing the data

25%
Formatting & annotating the visual
The type of visualizations you can create depends on the structure of the data tables (columns and rows).

See below two different structures of the same data points.

A. Grouped by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>.6</td>
</tr>
</tbody>
</table>

B. Grouped by type of work

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>.6</td>
</tr>
</tbody>
</table>
The type of visualizations you can create depend on the structure of the data tables (columns and rows).

See below two different structures of the same data points with grouped bar chart examples.

A. Grouped by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Difference/gap between the types of unpaid work for a given gender.

B. Grouped by type of work

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Difference/gap between men's and women's time spent on a given type of unpaid work.
3. Data visualization tips and concepts by chart type
COMPARISON – BAR AND COLUMN CHARTS
Bar and column charts

- Simplest form of charts which allow for easy interpretation by non-expert audiences.
- Bar charts allow the labels to be legible on the left side whereas column charts the labels might be at a 45 to 90 degree angle.
- If not sex-disaggregated (i.e. adolescent fertility), keep all bar colors the same and highlight only the bar with the insight you’re imparting on audience in separate color.
- Order bars by value and not alphabetically (for example geographical administrative regions).
Bar and column charts

- If highlighting female vs male values, group the bars or columns by non-gender category and use only two colors – one for each gender.
- Use gridlines or data labels – don’t use both.
  - Data labels are acceptable for up to 10-15 columns/bars
- Consider using icons when there are few bars. For example, only two bars (Female/Male) or four (i.e. Female/Male with Urban/Rural disaggregation).
- Sometimes column charts are used to show survey years (as an alternative to line charts).

**Figure 5-2**: Evolution of the possession of mobile phone by quintiles (%)
Maps

- Maps will provide nuance to the national statistics by easily identifying the discrepancies within the country regionally.
  - These regional data are often not collected by international organizations but are most crucial to policymakers in order to pinpoint policies and interventions that work for each region.
- Use the same color scale and ranges when using side by side maps to compare female and male values.
- To avoid side by side maps, map the gender gap rather than female and male values.
  - Make sure to use a diverging scale with the gender gap data.
  - Rename the legend for the direction of the gap whether the female value is relatively higher or lower than the male value or vice versa.
  - Negative numbers will confuse the reader. Where possible don’t show the negative numbers.
➢ Discussion – Is it easy to see the main insights? Easy to see the gender gap?

➢ Is there anything you would change about the colors of these visuals?
GROUP DISCUSISON

- Discussion – what is the takeaway from the map?

- Is there anything you would change about the colors of these visuals?

- Potential actions:
  • Financial literacy classes in school.
  • Support to women owned business.
  • Simplify the loan approval process.

Women's access to finance: Current Situation

% of women with bank account [75]
- 4.38 - 24.01 [49]
- 24.01 - 43.63 [17]
- 43.63 - 63.25 [7]
- 63.25 - 82.87 [2]

src: https://emap.nrb.org.np/
Map src: https://gadm.org/maps/NPL.html
TRENDS OVER TIME – LINE CHARTS
Line charts

- Used primarily for **time series with year on x-axis**.
  - **X-axis could also be used for ordinal categories** like age ranges, hours in a day, wealth quintile.
- Policymakers and researchers are most interested in seeing progress over the years for narrowing gender gaps, so where possible, provide time-series for indicators.
- Differentiate lines first by color, then by pattern (dashed or dotted line).
- Especially important in time series to **highlight years that may show stark contrast in data before and after that year**.
  - Year in which methodology/calculation of indicators according the ICLS -19 were implemented in data collection as there may be a stark contrast in the data.
  - Year of historical event or relevant legislation enacted.
There should be **no more than 4-6 lines in a line chart.**

**Use small multiple when there are too many lines.** This is also called a grid of charts and you can easily show male and female values for several disaggregations.

---

**Figure 6B**

Revising the Spaghetti Chart

Initial DI Worker Awards by Major Cause of Disability—Calendar Years 1975–2010

(Mean)

- Mental
- Cancer
- Circulatory
- Musculoskeletal

Source: Social Security Advisory Board (2012)
Line charts

- Show female and male rates while also easily visualizing the gap.
  - Shade the gap area between the female and male lines to make the gender gap more visible.
- Annotate and add highlight ranges for context.
COMPOSITION - PARTS OF A WHOLE
Stacked bar and column charts

- Used for depicting parts of a whole/total typically 100%
- Alternative to the grouped bar or column chart to save space.
- Most commonly used to stack:
  - female and male.
  - age ranges.
  - categories of employment (part time vs full time, sectors, etc.).
  - reasons for not being in labor force.
  - categories of unpaid work.
Stacked bar and stacked area charts

- Decision to group by gender or by disaggregation depends on the message being highlighted. Highlighting gender gap vs highlighting distribution or comparison within disaggregation category.

- Stacked area charts convey the same things as stacked bar or column chart but with the time series dimension.

![Stacked bar and area charts example](image-url)
Discussion – Is it easy to see the main insights? Easy to see the gender gap? Is there anything you would change about the colors of these visuals?
COMPARISON OF GAPS – RANGE PLOTS
Great for highlighting gender gaps very clearly, especially when there are a lot of categories or disaggregations (age, location, region, etc.).

- Show female and male rates using a colored dot.
- Show gender gap using the length of the shaded/colored line between the two dots.
Range plots

- Can add the data labels for both the length of the line (gap) and the dots (female and male values).
- **The values will always be female vs. male will NOT show totals.**
- You could potentially show range plot between two years rather than between female and male values.
- The range plot should always be ordered by value (usually female) so it’s easy to follow.
COMPARISON/TIME TRENDS – HEATMAPS
Heatmaps

- Convey a lot of information with multiple disaggregations in one chart.
- Can have one row of female values and one row of male values for comparison.
- Can visualize the gender gap for a disaggregation category with many categories (quintile, types of ownership, age ranges).
Distribution chart/Dot Plot

- Shows distribution within the dataset/category
- Country value compared to regional values.
- Capital city value compared to administrative region values.
- Distribution of all household values (if using microdata).
- Can show vertically or horizontally and color code by region or other types of disaggregation.
PRACTICE EXERCISES
Exercise – suggest changes to the set of visuals

➢ See Sheet named "Exercise 2" in the Excel file "Training Dataset Day 1" for an example of a set of visuals that are part of the same report.

➢ Write a list of suggested changes that you would make to this set of visuals to ensure that they are consistent and following best practices in data visualization principles.

➢ For best practices in data visualization refer to the "Data Visualization Tip Sheet" handout or Word file.

➢ There should be consistency across the whole report in:
  • colors for the same categories;
  • font type and size;
  • title structure and punctuation;
  • phrasing of categories.

➢ Time for exercise: 20 minutes.
➢ Write your answers in the Sheet named "Ex2 Response".
ANNEX
Financial Times Visual Vocabulary
Visual vocabulary

Designing with data

There are so many ways to visualize data – how do we
know which one to pick? Take the categories across the
top to decide which data relationship is most important
in your story, then look at the different types of chart
within the category to form some initial ideas about what
might work best. This list is not meant to be exhaustive,
not a wizard, but a useful starting point for making
informative and meaningful data visualizations.

ft.com/vocabulary
**Deviation**

Emphasize variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral/negative).

**Example FT uses**
Trade surplus/deficit, climate change

**Correlation**

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

**Example FT uses**
Inflation and unemployment, income and life expectancy

**Flow**

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

**Example FT uses**
Movement of funds, trade, migrants, laws, information, relationships, graphs.

**Diverging bar**
A simple standard bar chart that can handle both negative and positive magnitude values.

**Diverging stacked bar**
Perfect for presenting survey results which involve sentiment (e.g. disagree/neural/agree).

**Spine**
Splits a single value into two contrasting components (e.g. male/female).

**Surplus/deficit filled line**
The shaded area of these charts allows a balance to be shown — either against a baseline or between two series.

**Scatterplot**
The standard way to show the relationship between two continuous variables, each of which has its own axis.

**Column + line timeline**
A good way of showing the relationship between an amount (columns) and a rate (line).

**Connected scatter plot**
Usually used to show how the relationship between 2 variables has changed over time.

**Bubble**
Like a scatter plot, but adds additional detail by sizing the circles according to a third variable.

**XY heatmap**
A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

**Sankey**
Shows changes in flows from one condition to at least one other; good for tracing the eventual outcome of a complex process.

**Waterfall**
Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

**Chord**
A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.

**Network**
Used for showing the strength and inter-connectedness of relationships of varying types.
Ranking
Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

**Example FT uses**
- Wealth, deprivation, league tables,
- Concurrency election results

**Ordered bar**
- Standard bar charts display the ranks of values much more easily when sorted into order.

**Ordered column**
- See above.

**Ordered proportional symbol**
- Use when there are big variations between values and/or seeing fine differences between data is not so important.

**Dot strip plot**
- Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

**Skew**
- Perfect for showing how ranks have changed over time or vary between categories.

**Lollipop**
- Lollipops draw more attention to the data value than standard bar/columns and can also show sets and value effectively.

Distribution
Show values in a dataset and how often they occur. The shape (or 'look') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

**Example FT uses**
- Income distributions, population
- Capital city distribution, revealing inequality

**Histogram**
- The standard way to show a statistical distribution - keep the gaps between column levels to highlight the 'shape' of the data.

**Dot plot**
- A simple way of showing the range or frequency of data across multiple categories.

**Dot strip plot**
- Good for showing the range or frequency of data in a distribution, can be a problem when too many dots have the same value.

**Bar code plot**
- Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

**Boxplot**
- Summarise multiple distributions by showing the median, central and range of the data

**Violin plot**
- Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

**Population pyramid**
- A standard way for showing the age and sex breakdown of a population distribution, effectively split to back histograms.

**Cumulative curve**
- A good way of showing how spread a distribution is, a line is always cumulative frequency x and it always a measure.

**Frequency polygons**
- For displaying multiple distributions of data like a regular line chart but limited to a maximum of 3 or 4 datasets

**Beeswarm**
- Use to emphasise individual points in a distribution. Points can be used to an additional variable, best with medium-sized datasets.
Change over Time

Give emphasis to changing trends; these can be short (hourly) movements or extended across decades or centuries. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses:
Share price movements, economic time series, sectoral changes in a market.

**Line**
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

**Column**
Column work well for showing change over time - but usually best with only one series of data at a time.

**Column + line timeline**
A good way of showing the relationship over time between an agent (column) and a rate (line).

**Step**
Good for showing changing data as long as the data can be simplified into 2 x 3 points without missing a key part of story.

**Area chart**
Use with care - these are good at showing changes in total, but seeing change in components can be very difficult.

**Candlestick**
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projections)
Use to show the uncertainty in future projections - usually this grows the further forward to projection.

Connected scatterplot
A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression.

Calendar heatmap
A great way of showing temporal patterns (daily, weekly, monthly) - at the expense of showing precision in quantity.

Priestley timeline
Great when data and duration are key elements of the story in the data.

Circle timeline
Good for showing discrete values of varying size across multiple categories (e.g., earthquakes by continent).

Vertical timeline
Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Seismogram
Another alternative to the circle timeline for showing series where there are big variations in the data.

Streamgraph
A type of area chart used when seeing changes in proportions over time is more important than individual values.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 3: EXCEL CHARTS AND DATA ADJUSTMENTS
Contents

1. Recap of data visualization steps and breakdown
2. Using data visualization principles to adjust data and chart elements in Excel
   - Grouping or ordering data, chart elements, and legend for the right data presentation
   - Decluttering the chart
   - Appropriate sorting and use of colors
Specific actions for adjusting chart elements

1. **Using the right data structure or grouping to highlight the key gender-relevant insight** (slides 9-16).
2. **Reordering chart elements to ensure visually female values are before the male values** (slides 18-25)
3. **Adjusting data structure for grouping multiple disaggregations in one chart** (slides 27-38)
4. **Decluttering the visual** (slides 41-58)
   - Removing 3D elements (slides 41-43)
   - Removing colored background (slides 44-48)
   - Removing gridlines (slides 49-50)
   - Removing axis title and labels (slides 51-52)
   - Formatting data labels (slides 53-56)
   - Sorting values in chart (slides 57-58)
5. **Adjusting the colors of values in a chart** (slides 60-64)
6. **Removing the legend** (slide 65)
1. Recap of data visualization steps and breakdown
Recap: steps for creating a data visualization

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:
- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

1. Upload data
2. Explore/try visualization options
3. Adjust data structure
4. Design, format, and finalize visualization
5. Publish and/or download visualization

Optional: add additional annotations as needed
Recap: data visualization breakdown

75%
Properly cleaning, transposing & preparing the data

25%
Formatting & annotating the visual

**Three-fourths** of creating a visualization is cleaning and formatting the data in the proper way to:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation
Recap: data visualization breakdown

Layers of a visualization
(or series of visualizations)

Considerations about the information to be visualized

Read from the bottom-up

Dynamics
- Animations
- Interactions
- Personalizations

Visual design

Layout

Annotations
- Titles
- Introductions
- Footnotes
- Textual highlights

Arrangements within the visualization
- Positioning
- Ordering
- Grouping

Encoding
- Marks
- Channels

Scaffolding
- Scales
- Legends
- Labels

Static visualizations

Dynamic visualizations
2. Using data visualization principles in Excel to adjust data and chart elements
USING THE RIGHT DATA STRUCTURE OR GROUPING TO HIGHLIGHT THE KEY GENDER-RELEVANT INSIGHT
Adjusting data grouping

- Time use data (% of time spent of a 24-hour day) – values are grouped by type of unpaid work.
- You decide to select a *stacked bar chart* to visualize parts of a whole.
  - What part/percentage of the whole day do these work activities comprise for women and men?

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>.6</td>
</tr>
</tbody>
</table>
Adjusting data grouping

 ➢ Highlight the data range and click the “Insert” tab in the Excel ribbon (toolbar menu at the top).
 ➢ Under the "Charts" section click “Recommended Charts” and click the stacked bar chart and click “OK”.

1. Highlight the data range.
2. Click the “Insert” tab.
3. Under the "Charts" section click “Recommended Charts”.
4. Select the stacked bar chart.
5. Click “OK” to apply the chart.
Incorrect data structure

- This visual is grouping values by type of work (row) and stacking female and male values (columns).
- The data are not structured properly for this stacked bar chart to show what part/percentage of the whole day these unpaid work activities comprise for women and men.

⚠️ Adjustment required!
This visual should instead be stacking types of unpaid work to show the % of time spent in a day by each gender on the combined unpaid domestic and care work activities.

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>.6</td>
</tr>
</tbody>
</table>
Adjustment option 1: transpose data in tool

First, always try transposing the data in the visualization software.

- Click into the chart and then click on the "Chart Design" tab in the Excel ribbon (top toolbar menu). If there is no "Chart Design" tab try clicking into the chart again because this option will only show up when the chart is selected.
- Then click "Switch Row/Column" in the Excel ribbon (top toolbar menu).
Adjustment option 1: transpose data in tool

➢ First, always try transposing the data in the visualization software.
➢ You should see that the visual has changed. The visual is grouped by gender now. This does not change the data table in the Excel spreadsheet. It only changes the chart orientation.

Easier to see that women spent much higher percentages of their day on unpaid work than men.
Adjustment option 2: change visual type

➢ If you are not set on the chart type, switch to a visualization that better fits the original data structure.
  • Select a different chart type icon that you wish to try (i.e. stacked column, grouped bars/columns).

Emphasis on gap or difference between the types of unpaid work for a given gender.

Emphasis on gender gap within a given type of unpaid work.

Emphasis on gap or difference between the types of unpaid work for a given gender.
Adjustment option 2: change visual type

How do you choose which alternative visual to use?

- Is the visual still emphasizing or highlighting the right message/insight?
- Do you have to adjust the chart area or the bar or column height/width?
- Do you have to add data labels, a color key, or gridlines?
- Do you have to sort, regroup, or re-order the bars or columns?

- Correct message
- Data labels and color key needed
- Make chart area taller and bars slightly thinner

- Correct message
- Minimal aesthetic changes or ordering needed

- Must transpose to highlight the correct insight
- Data labels or gridlines needed
- Make chart area taller and bars slightly thinner
Go to Sheet named "Task 1" in the Excel file "Training Dataset Day 1".
Use either of the options demonstrated (transposing data or changing chart type) to adjust the chart in the sheet so that it highlights the gender gap.
REORDERING CHART ELEMENTS TO ENSURE THAT VISUALLY THE FEMALE VALUES ARE BEFORE THE MALE VALUES
Adjusting the order/display of genders

- Female values should **always be first** in the chart either above the male values or to the left of the male values.
- In this chart, the male values show up first, above the female values.
  - **The data table must be reordered to make the change show up in the chart.**

![Chart Title](chart.png)

Visually, "Male" is ordered above or before "Female".
Adding a blank column

➢ Edit the data table by inserting a column in between the "Female" and "Male" column.
   • Click the letter C at the top of the "Male" column to highlight the entire column C for "Male".
   • Right click the highlighted column and select "Insert".
   • There should now be a blank column in between the "Female" and "Male" columns.
Copying and pasting tips

➢ Right click the highlighted cells and select "Copy" then right click the top cell in the empty column and select "Paste" which might look like a clipboard and paper symbol instead of the word "Paste".

➢ Shortcut: CTRL+C for copy and CTRL+V for paste.
  • After highlighting the cells, hold the CTRL and the C key at the same time to copy the highlighted values.
  • Then click on the top/first cell in the empty column and hold the CTRL and the V key at the same time to paste the values.
Copying and pasting data into blank column

➢ Rearrange the columns so that the male values are to the left of the female values in the table.
  • Highlight the cells in the "Male" column by clicking and holding the first cell where it says "Male" and dragging down to the last row. All three rows should be highlighted.
  • Then copy and paste the male data into the empty middle column.
  • Male values should be duplicated now in both the table and the chart.

---

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Chart Title

- Male
- Unpaid Domestic Work
- Unpaid Care Work
Copying and replacing data in column

➢ Rearrange the columns so that the male data are to the left of the female data in the table.
  • Copy the "Female" column and replace the "Male" column furthest to the right.
  • **Female values should be duplicated now in both the table and the chart.**
Deleting the duplicated column

➢ Rearrange the columns so that the male data are to the left of the female data in the table.
   • Delete the first column of data which is column B for "Female" by right clicking the top of the column where it says "B" and selecting "Delete".
Adjusting the order/display of genders

- You may receive this pop-up warning message. Click "OK".
- The data table will now only have two columns and the chart will show "Female" above "Male".
Go to Sheet named "Task 2" in the Excel file "Training Dataset Day 1".
Correct the Female/Male ordering in the chart in the sheet (preview below).
What We Do

ADJUSTING DATA STRUCTURE FOR GROUPING MULTIPLE DISAGGREGATIONS IN ONE CHART
Adjusting grouping of multiple disaggregations

- The table is showing asset ownership (% of females or males who have sole, joint or both sole and joint ownership of land).
- The data are structured as seen below by gender, type of ownership and location.

<table>
<thead>
<tr>
<th>Ownership</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Capital City</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>21.0</td>
<td>79.0</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>33.4</td>
<td>66.6</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>21.4</td>
<td>78.6</td>
</tr>
<tr>
<td><strong>Urban</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>28.2</td>
<td>71.8</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>15.7</td>
<td>84.3</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>24.9</td>
<td>75.1</td>
</tr>
<tr>
<td><strong>Rural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>22.2</td>
<td>77.8</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>9.5</td>
<td>90.5</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>20.2</td>
<td>79.8</td>
</tr>
</tbody>
</table>
Adding a clustered column chart

➢ Highlight the data range and click the “Insert” tab in the Excel ribbon (toolbar menu at the top).
➢ Under the "Charts" section click “Recommended Charts” and click the clustered column chart and click “OK”.

1. Highlight the data range.
2. Click the “Insert” tab.
3. Click “Recommended Charts”.
4. Select the clustered column chart.
5. Click “OK”.

1. Ownership data range.
2. Insert tab.
3. Recommended Charts.
4. Clustered Column.
5. OK button.
The visual does not register the location grouping in this data structure and thinks it is part of the series.
The data must be restructured so that the Excel visualization tool can cluster the column charts by location.
To restructure the data, first insert a new column to the left of the "Ownership" column (currently column A) by right clicking on the top of Column A and selecting "Insert".

There should now be a blank first column.
Cutting and pasting data into blank column

- Cut the location categories and paste them into the first column in the first row of data for that section.
  - You must cut each category and paste it individually. Repeat for each category that must be moved.
    - Shortcut for the cut function is CTRL+X (holding CTRL and X at the same time after the intended cell has been highlighted).
  - In this example, the name of the location should be in the same row as the “Sole ownership” data.
Opening the select data source window

➢ Right click on the chart and click “Select Data”.
➢ The “Select Data Source” box will pop up.
Selecting the new range of data in data source

- The previous data range will be denoted with a dashed green border.
- Select the full range of data including column A and hit the "Enter" button on the keyboard or click the "OK" button in the popup box.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capital</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>City</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sole</td>
<td>21.0</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>33.4</td>
<td>66.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>21.4</td>
<td>78.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sole and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sole</td>
<td>28.2</td>
<td>71.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>15.7</td>
<td>84.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>24.9</td>
<td>75.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sole and j</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sole</td>
<td>22.2</td>
<td>77.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>9.5</td>
<td>90.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownership</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>20.2</td>
<td>79.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sole and j</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Adjusting grouping of multiple disaggregations

➢ Your visualization should now be grouped by the additional “location” disaggregation but there is a little too much spacing in between the groups.
Removing extra space between groupings

- Highlight the empty rows between the groupings by clicking on the grey number associated with the row.
- The rows can be highlighted and deleted separately or at the same time. To highlight and delete at the same time, highlight the first row as indicated above, then hold down the CTRL button and highlight the second empty row you plan to delete (while still holding the CTRL button) before moving to the next step.
Removing extra space between groupings

➢ Once the rows you want are highlighted, right click the row and select “Delete”.
  • If you accidentally highlight the wrong row, before you delete the row, click anywhere in the Excel sheet to unhighlight, then start again with highlighting the right row as per the previous slide’s instructions.
➢ If you are deleting each row separately, highlight the next row and repeat the process.
Adjusting grouping of multiple disaggregations

Your visualization should now be grouped by the additional “location” disaggregation with no additional space in between the groups.
➢ Go to Sheet named "Task 3" in the Excel file "Training Dataset Day 1".
➢ Adjust the chart in the sheet to have the multiple disaggregations neatly grouped within the chart rather than individually listing all of the different combinations.
Below is a visual that you have received in a report and must apply data visualization principles to fix it.

**Discussion** – What are the immediate issues with this visual? What changes would you make?
What We Do

DECLUTTERING:
REMOVING 3-DIMENSIONAL ELEMENTS
Selecting a 2-dimensional chart

➢ To remove the 3-Dimensional effect, click on the chart and under the “Chart Design” tab select “Change Chart Type” and select the 2-Dimensional clustered columns and click “OK”.

1. 2. 3. 4. 5.
Decluttering visuals

➢ The visualization should now be 2-Dimensional.
DECLUTTERING:
REMOVING COLORED BACKGROUND
Selecting a new chart style

➢ You can remove the background either by changing the whole chart style or by changing the color of the background.
➢ To change the chart style, click on the chart and click on the paintbrush icon to the right and select a new chart style or go to the "Chart Design" tab and select a new chart style from the designs.
The visualization should now have a clear background.
Changing the background color of chart

➢ To just change the background color and none of the other chart elements, click the chart and go to the “Format Chart Area” panel that pops up on the right side of the screen. Under the paint icon under the “Fill” section, click on the "Color" dropdown and select white.
Changing the chart style vs. background color

➢ The left is the visual after changing the chart style while the right is the visual after changing the background back to white.
DECLUTTERING:
REMOVING GRIDLINES
Removing gridlines

➢ You can remove gridlines in multiple ways. You can click on the gridlines in the chart (the blue circles confirm that the gridlines are selected and not another chart element) and then hit the “Delete” key.
➢ Or you can click on the chart and then the “Chart Elements” plus symbol and uncheck the box next to "Gridlines".
DECLUTTERING: REMOVING AXIS TITLE AND LABELS
Removing the axis title and axis labels

- There are now duplications of elements. The y-axis is duplicating the data labels while the x-axis is duplicating the chart title. They can be removed in the same way as the gridlines either by clicking on the chart element and hitting the “Delete” key.
- Or you can click on the chart and then the “Chart Elements” plus symbol and uncheck the box next to "Axes".
DECLUTTERING: REFORMATTING DATA LABELS
Moving data labels inside the bars

- The data labels are better placed inside the columns. They can be changed by clicking on the chart and then on the “Chart Elements” plus symbol and clicking on the “Data Labels” option.
- Select “Inside End” or “Inside Base” for the "Data Labels".
Opening the “Format Data Labels” panel

➢ The data labels will appear inside the bars, but they cannot be seen well. The format should be changed.
➢ Data labels can be edited either by clicking a data label and going to “Format Data Labels” panel on the right side or by clicking the “Chart Elements” plus symbol, going to “Data Labels”, and then “More Options..” which should also open up the “Format Data Labels” panel.
Decluttering visuals

➢ In the “Format Data Labels” panel, click the “Text Options” then select a better contrasting color against the colors of the columns from the “Color” dropdown. Click into each data label and repeat color changes.
SORTING VALUES IN A CHART
Sorting values within the chart

➢ To sort the columns, bars, or other elements for other types of visuals, the sorting must happen in the original data table.

➢ Click in the column for the variable which you wish to sort. Then under the “Home” tab in the Excel ribbon (top toolbar menu, click on the “Sort & Filter” button in the “Editing” section. Then click the “Sort Z to A” or “Sort A to Z” option depending on which direction you require the values to flow.
TASK 4

➢ Go to Sheet named "Task 4" in the Excel file "Training Dataset Day 1".
➢ Adjust the chart in the sheet to remove any unnecessary elements (3D, gridlines, axes, etc.).
➢ Adjust the chart in the sheet to sort the bars in order of value.
ADJUSTING COLORS OF VALUES IN A CHART
The data shown are categorical in the column chart, but the color is not encoding any particular insight. There should only be one color for all of the columns.

When customizing colors you can i) select one of the colors already in the pop-up box or ii) enter a 6-letter-number combination called a HEX code.

- For this example, let’s use the HEX code #440E5F which produces [this color](#).
- If you do not have the option to enter a HEX code you may enter the corresponding RGB code with three numbers. For this color the RGB code is (68, 14, 95). Each of the three numbers corresponds to the three letters R (for Red), G (for Green, and B (for Blue).
Selecting hex codes

- Each color has a code made up of 6 units that are a combination of letters and numbers. If you want to consistently use the same color, write down the code for that color to use in multiple visualizations.
  - All HEX codes have corresponding RGB codes in case you do not have the option to enter HEX codes.
- HEX codes can be found at [htmlcolorcodes.com](http://htmlcolorcodes.com) and [https://www.computerhope.com/htmcolor.htm](https://www.computerhope.com/htmcolor.htm)
  - To find the corresponding RGB code of your HEX code, use [https://htmlcolorcodes.com/hex-to-rgb/](https://htmlcolorcodes.com/hex-to-rgb/)
- You can use the HEX codes that correspond to the colors in the National Statistical Office’s branding.

### Major hexadecimal color codes

Below are some common color names and codes. With these colors, you can also use the color name. For example, in HTML tags and CSS that use color codes, you could use “red” instead of “#FF0000”.

<table>
<thead>
<tr>
<th>Color Name</th>
<th>Color Code</th>
<th>Color Name</th>
<th>Color Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>#FF0000</td>
<td>White</td>
<td>#FFFFFF</td>
</tr>
<tr>
<td>Cyan</td>
<td>#00FFFF</td>
<td>Silver</td>
<td>#C0C0C0</td>
</tr>
<tr>
<td>Blue</td>
<td>#0000FF</td>
<td>Gray or Grey</td>
<td>#808080</td>
</tr>
<tr>
<td>DarkBlue</td>
<td>#00000B</td>
<td>Black</td>
<td>#000000</td>
</tr>
<tr>
<td>LightBlue</td>
<td>#ADD8B6</td>
<td>Orange</td>
<td>#FFA500</td>
</tr>
<tr>
<td>Purple</td>
<td>#800080</td>
<td>Brown</td>
<td>#A52A2A</td>
</tr>
<tr>
<td>Yellow</td>
<td>#FFFF00</td>
<td>Maroon</td>
<td>#800000</td>
</tr>
<tr>
<td>Olive</td>
<td>#00FF00</td>
<td>Green</td>
<td>#008000</td>
</tr>
<tr>
<td>Magenta</td>
<td>#FFFF00</td>
<td>Olive</td>
<td>#800000</td>
</tr>
<tr>
<td>Pink</td>
<td>#FFC0CB</td>
<td>Aquamarine</td>
<td>#7FFFD4</td>
</tr>
</tbody>
</table>
Entering a hex code in the “Format Data Series”

- Click on a column and go to the “Format Data Series” panel that pops up in the right side of the screen.
- Click the paint icon and go to the color dropdown.
- Instead of choosing a color from the pop-up, select “More colors” this will prompt the "Colors" window to pop up. Go to the “Custom” tab.
- At the bottom of the pop-up there is a fields titled “Hex”, "Red", "Green", and "Blue". Replace the HEX or RGB code of the current color with the NSO’s branding color or the agreed upon color for the report/visual. To the right it shows a preview of the new color against the old color. Then click “OK”.
### Adjusting colors

- The color for the column selected is now the purple color. Similar to the data labels, this needs to be repeated for each column.
- Microsoft Excel and Powerpoint usually save the color in the “Recent Colors” section which you will see when you open the "Colors" dropdown again. You will not have to enter the code each time.

#### FEMALE TO MALE EMPLOYMENT RATIO

<table>
<thead>
<tr>
<th>Household with children</th>
<th>Extended Family</th>
<th>Overall</th>
<th>Couple with children</th>
<th>Household without children</th>
<th>Couple without children</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>0.62</td>
<td>0.61</td>
<td>0.61</td>
<td>0.60</td>
<td>0.53</td>
</tr>
</tbody>
</table>

#### FEMALE TO MALE EMPLOYMENT RATIO

<table>
<thead>
<tr>
<th>Household with children</th>
<th>Extended Family</th>
<th>Overall</th>
<th>Couple with children</th>
<th>Household without children</th>
<th>Couple without children</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>0.62</td>
<td>0.61</td>
<td>0.61</td>
<td>0.60</td>
<td>0.53</td>
</tr>
</tbody>
</table>
Removing the legend

- Now that all columns are the same color, there is no need for a legend. Instead, the labels should be on the x-axis under each column.
- One way to add labels (without significant work in Excel) is to delete the legend and add the labels as individual text boxes when annotating the visual. This may be preferable if you are already adding the title and source in another tool (i.e. in MS Word or Powerpoint, Canva/Visme, Adobe Illustrator).
- To remove the legend, click on the chart, then on the “Chart Elements” plus symbol. Uncheck “Legend”.
  - Before removing the legend, ensure that you have the data value-category combinations saved somewhere so that you can add the annotations later. You can always refer to the original Excel file.
- This visual can then be copied into Word or Powerpoint or any other tool to then add the labels.
Go to Sheet named "Task 5" in the Excel file "Training Dataset Day 1".

Adjust the chart in the sheet to change the color of the bars to hex code #1D5934 or RGB code (29, 89, 52).

Adjust the chart in the sheet to change the data label colors so that they are visible against this dark bar color.
Female to Male Employment Ratio

**BEFORE**

- Household with children: 0.64
- Extended Family: 0.62
- Couple with children: 0.61
- Household without children: 0.61
- Couple without children: 0.60

**AFTER**

- Household with children: 0.64
- Extended Family: 0.62
- Overall: 0.61
- Couple with children: 0.61
- Household without children: 0.60
- Couple without children: 0.53
PRACTICE EXERCISE:
Recreate a visual from trainer’s screen using data from the Training Dataset file
Exercise – recreate the visual

➢ Using the data in the Sheet named "Exercise 3" in the Excel file "Training Dataset Day 1", recreate the visual below. The male hex code is #003399 or RGB (0, 51, 153) and the female hex code is #008080 or RGB (0, 128, 128).

➢ If your visual doesn’t look the same, think of the following:
  • Are you using the same chart type?
  • Have you adjusted the data structure or created disaggregation categories?
  • Have you adjusted the gridlines, data labels, colors, chart title?

➢ Time for exercise: 20 minutes
Video for widening the plot area

➢ Click the plot area and drag the right side of the area towards the right to make the plot area wider.
  • To drag, you must hover over the circular point for the mouse to turn into an arrow (see video).
  • You will see a light blue outline if done correctly while you’re dragging the plot area.
  • If you happen to move the plot instead of widen it, you can still adjust it – it might take more steps.

WIDEN PLOT (CORRECT)

MOVE PLOT
What We Do

TRANSFORMING LABELS
Moving the data labels to the base of the bars

- After removing the legend, instead of adding the category labels in another tool, another option is to transform the data labels into the category labels.
- Click on the chart and then on the “Chart Elements” plus symbol. Then click “Data Labels” and select “Inside Base”. You should see the data labels move to the bottom. Repeat and select “More Options”.
- In the “Format Data Labels” panel on the right side, click the bar chart symbol and then “Label Options”.
Transforming data labels into category labels

- Check the box next to “Series Name” and uncheck the boxes next to “Value” and “Show Leader Lines”
- You should now see the name rather than the data label. Repeat for each of the data labels.
Transforming data labels into category labels

➢ Check the box next to “Series Name” and uncheck the boxes next to “Value” and “Show Leader Lines”.
➢ Then click the “Text Options” tab and change the text color to black.
Adjusting the size of the plot area

➢ You should now see black category labels instead of white data labels.
➢ Click the inner plot area with the bars and drag upwards to make the area shorter. Now there is space to drag down each of the category labels.
Moving the labels below the plot area

- Click each category name separately and drag it down into the white space below the purple columns.
- Adjust the sizing of the text by moving the 4 corners of the text box.
- Repeat for each category until they are all below the corresponding purple column.
Changing data labels to category labels

- However, readers can no longer tell what values the columns correspond to.
- You could either add the Y-axis and/or gridlines back or you could copy the visual (now with category labels) into Word, Powerpoint, or another tool (Canva, Visme) and add data labels as separate text boxes.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 4: VISUALIZATION TOOL COMPARISON
## Data visualization tools comparison

<table>
<thead>
<tr>
<th>Feature</th>
<th>Excel</th>
<th>Datawrapper</th>
<th>Flourish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suitable to work with static charts</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Suitable to work with interactive charts</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Operates in multiple languages</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Works without Internet connection</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Offers a wide variety of chart options, types and visuals</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Offers the possibility to safely encrypt and store data</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Has a free version</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Datawrapper

➢ Pros
  • Create polished charts and advanced tables in a few clicks compared to Excel which requires many steps.
  • Easy to change chart or transpose data structure in tool without changing original datasheet in Excel.
  • Built-in country maps allow easy visualization at province, administrative region level.
  • Easy to annotate, highlight one or a few lines, bars, data points using specific colors, or to arrange groups of categories vertically.

➢ Cons
  • Chart types are limited compared to Flourish.
  • There isn’t always sample data to understand the data structure for that particular chart type.
  • For interactive charts, you must publish the chart.
  • Internet required.

Training materials available:
https://www.datawrapper.de/training-materials
Datawrapper advanced tables

Training materials available: https://www.datawrapper.de/training-materials
Flourish

➢ Pros:
• Extensive static and interactive visualization template library with sample data and visuals for all charts.
• Easy to create even advanced chart types with many customization options, and grouping or layout options.
• Create data stories with multiple types of visuals, animation, and scrollytelling.

➢ Cons:
• For interactive charts, you must publish the chart.
• Sometimes you must restructure the data to highlight a certain data point, line, bar, series.
• Compared to Datawrapper, sometimes colors and annotations are less intuitive – take longer
• Internet required.

Flourish beginner course available: https://training.flourish.studio/
Flourish grid of charts

Flourish beginner course available: https://training.flourish.studio/
Flourish data stories

Flourish beginner course available: https://training.fLOURISH.studio/
When to use each data tool

Use this tool if/when you...

- **do not have access to the internet**;
- have data that are not yet public and the data storage/encryption requirements;
- **do not suffice** for the private/sensitive data.

- want to quickly create a standard static or interactive chart or advanced tables (with sparklines or data bars);
- prefer to easily **highlight one or a few main data points, lines, or bars** in the chart;
- want to quickly create a **subnational geographic map, range plot, bullet bar**;
- need to **work from a mobile-phone or tablet**.

- need to create a chart type not available in other tools or a more advanced chart type;
- need to see sample data or visuals to understand how to create the visual;
- want to easily create a **data story**;
- prefer to **group data horizontally** instead of vertically;
- want to visualize data in “small multiples” or a “grid of charts”;
- prefer to **add a highlight line or sort the legend options** in a particular order.
Canva, Visme, Infogram, Venngage, Piktochart etc. produce a collection of imagery, text, and data visualizations to give an eye-catching, creative overview of a topic. All have similar functions.
Dashboards: Tableau and PowerBI

- Tableau and PowerBI provide an interface with interactive visualizations and business intelligence “BI” capabilities allowing users to create their own dashboards or reports mostly to track/monitor progress.
- Dashboards use a combination of interactive visualization types and tables that are sortable, filterable.
- Dashboards are exploratory (not explanatory) visualizations – they allow users to explore various insights rather than highlighting one or w few key insights or narratives/stories.
ANNEX
Comparison of SGS indicators in each data visualization tool
Managerial positions

![Bar chart showing the proportion of women in managerial positions in three regions: URT, TZM, and ZNZ. The percentage is broken down by gender, with the majority of managerial positions held by men.](chart.png)
Time use

Proportion of time spent on unpaid work (% of day), by sex

- Male: 4.5% (Unpaid Domestic Work) 0.6% (Unpaid Care Work)
- Female: 15.9% (Unpaid Domestic Work) 3.4% (Unpaid Care Work)

Source: National Statistical Office Labor Force Survey 2022 - Get the data

Proportion of time spent on unpaid domestic and care work

- Male: 4.5% (Unpaid Domestic Work) 0.6% (Unpaid Care Work)
- Female: 15.9% (Unpaid Domestic Work) 3.4% (Unpaid Care Work)
Employment by sector

United Republic of Tanzania

<table>
<thead>
<tr>
<th></th>
<th>Agriculture</th>
<th>Industry</th>
<th>Services</th>
<th>Not Classifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>60.6</td>
<td>5.3</td>
<td>33.4</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>60.6</td>
<td>5.3</td>
<td>33.4</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>59.6</td>
<td>12.7</td>
<td>27.2</td>
<td></td>
</tr>
</tbody>
</table>

Tanzania Mainland

<table>
<thead>
<tr>
<th></th>
<th>Agriculture</th>
<th>Industry</th>
<th>Services</th>
<th>Not Classifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>60.8</td>
<td>12.5</td>
<td>29.6</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>61.3</td>
<td>5</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>60.4</td>
<td>12.5</td>
<td>26.7</td>
<td></td>
</tr>
</tbody>
</table>

Zanzibar

<table>
<thead>
<tr>
<th></th>
<th>Agriculture</th>
<th>Industry</th>
<th>Services</th>
<th>Not Classifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>33.5</td>
<td>16.8</td>
<td>47.2</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>30.7</td>
<td>14.5</td>
<td>51.0</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>35.4</td>
<td>18.5</td>
<td>44.5</td>
<td></td>
</tr>
</tbody>
</table>

Flourish
Asset Ownership

Household members who have their name on the title deed of the household dwelling, by sex (%)

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>22</td>
<td>78</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>9</td>
<td>91</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Urban</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>16</td>
<td>94</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>Capital City</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>National</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>12</td>
<td>88</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>21</td>
<td>79</td>
</tr>
</tbody>
</table>
Ratio of female to male labor force participation
## Asset ownership

### Household members who have their name on the title deed of the household dwelling, by sex (%)

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>22</td>
<td>78</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>9</td>
<td>93</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td><strong>Urban</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>16</td>
<td>84</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td><strong>Capital City</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td><strong>National</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sole ownership</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td>Joint ownership</td>
<td>12</td>
<td>88</td>
</tr>
<tr>
<td>Both sole and joint ownership</td>
<td>21</td>
<td>79</td>
</tr>
</tbody>
</table>

### Household members who have their name on the title deed of the household dwelling, by sex (%)

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sole ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>22</td>
<td>78</td>
</tr>
<tr>
<td>Urban</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>Capital City</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>National</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td><strong>Joint ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>9</td>
<td>91</td>
</tr>
<tr>
<td>Urban</td>
<td>16</td>
<td>84</td>
</tr>
<tr>
<td>Capital City</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>National</td>
<td>23</td>
<td>77</td>
</tr>
<tr>
<td><strong>Both sole and joint ownership</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>Urban</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Capital City</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>National</td>
<td>21</td>
<td>79</td>
</tr>
</tbody>
</table>
Gender Pay Gap

Average hourly earnings in local currency

<table>
<thead>
<tr>
<th>Category</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers</td>
<td>5.0K</td>
<td>6.3K</td>
</tr>
<tr>
<td>Professionals</td>
<td>6.1K</td>
<td>6.2K</td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>3.8K</td>
<td>4.3K</td>
</tr>
<tr>
<td>Armed forces</td>
<td>3.4K</td>
<td>3.7K</td>
</tr>
<tr>
<td>Clerical support workers</td>
<td>2.4K</td>
<td>3.1K</td>
</tr>
<tr>
<td>Total</td>
<td>2.2K</td>
<td>2.9K</td>
</tr>
<tr>
<td>Skilled agricultural, forestry and fishery workers</td>
<td>1.9K</td>
<td>2.1K</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>1.8K</td>
<td>2.1K</td>
</tr>
<tr>
<td>Service and sales workers</td>
<td>1.4K</td>
<td>1.5K</td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>1.3K</td>
<td>1.6K</td>
</tr>
</tbody>
</table>

Source: Country A Labor Force Survey 2022 - Get the data

Source: Paris21 Communicating Gender Statistics on Women’s Economic Empowerment Course
Contents

1. Recap of data visualization steps and breakdown
2. Using data visualization principles to create better charts in Excel:
   • Transforming a pie chart into a stacked chart
   • Transforming a grouped column chart into a bullet bar or a line chart
Specific actions for creating better charts

1. Changing chart type of an existing chart (slides 15-16)
2. Adjusting data selection for a chart (slides 17-18)
3. Widening the plot area within a chart (slide 19)
4. Using the “Switch Row/Column” button to reformat a stacked bar chart (slide 21)
5. Adding a chart title & adjusting the bar thickness (slide 22)
6. Changing the bar colors (slides 24-28)
   • Choosing appropriate colors (slide 25)
   • Selecting HEX/RGB codes or color palettes (slides 26)
7. Overlapping bars (slides 36-38)
8. Removing secondary axis (slide 41)
9. Adjusting frequency of years on x-axis (slide 48)
10. Adjusting line type (slide 49)
11. Adjusting the size of dots (slide 50)
12. Adjusting the colors of lines and dots (slide 51)
13. Moving the legend and widening the chart (slides 52-53)
1. Recap of data visualization steps and breakdown
Recap: steps for creating a data visualization

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- Match the required data structure, inputs, and features of the intended visualization type
- Highlight the right message and insights
- Enable easy formatting and annotation

---

1. Upload data
2. Explore/try visualization options
3. Adjust data structure
4. Design, format, and finalize visualization
5. Publish and/or download visualization
6. Optional: add additional annotations as needed
Recap: data visualization breakdown

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

75% Properly cleaning, transposing & preparing the data

25% Formatting & annotating the visual
Recap: data visualization breakdown

Layers of a visualization
(or series of visualizations)

Considerations about the information to be visualized

Read from the bottom-up

Dynamics
- Animations
- Interactions
- Personalizations

Visual design

Layout

Annotations
- Titles
- Introductions
- Footnotes
- Textual highlights

Arrangements within the visualization
- Positioning
- Ordering
- Grouping

Encoding
- Marks
- Channels

Scaffolding
- Scales
- Legends
- Labels

Static visualizations

Dynamic visualizations
2. Using data visualization principles in Excel to create better charts
Enhancing Excel charts

PART 1: ADJUSTING DATA OR CHART ELEMENTS

Example 1:
Grouping and ordering data, chart elements, and legend for the right data presentation.

Example 2:
Simplifying (removing unnecessary chart elements), intentional and appropriate sorting, and using the right data visualization attributes to highlight insights (color, shape, pattern, etc.)

PART 2: USING BETTER CHART TYPES

Example 1:
Transforming pie chart to a stacked column chart.

Example 2:
Transforming grouped column chart with multiple disaggregations into a bullet bar.

Example 3:
Transforming grouped bar chart to line chart.

PRACTICE EXERCISES

1. Recreate a visual from trainer’s screen using data from Training Dataset file.
2. Create one’s own visual with data of choice from Training Dataset file.

NOTE: It is recommended to complete PART 1 and the accompanying Practice Exercise from Excel Module 3 prior to moving on to PART 2 (this module).
TRANSFORMING A PIE CHART INTO A STACKED CHART
GROUP DISCUSSION

Discussion – Can you understand the main insight? What other ways could this visual be shown that are more efficient? What would you change in this visual?

Reasons for not seeking employment

- Female
  - Full time student: 22%
  - Too young/Too old: 23%
  - Disabled/Ill: 16%
  - Occupied with home duties: 12%
  - My spouse wouldn’t allow that: 9%
  - Other: 18%

- Male
  - Full time student: 24%
  - Too young/Too old: 43%
  - Disabled/Ill: 9%
  - Occupied with home duties: 16%
  - My spouse wouldn’t allow that: 9%
  - Other: 12%
GROUP DISCUSSION

Discussion – Can you understand the main insights better? Which one shows the gender gap better? Can you still understand the individual reasons as a proportion of all reasons (part-to-whole)?
The 100% stacked bar chart compared to the pie chart...

- shows the values in length which are easier to compare than the pie slices;
- shows the percentage/proportion of the whole;
- is easier to identify gender gap comparison with female visually above male comparing lengths;
- is created in one chart not two separate charts.
Stacked bar and column charts save space

- These images are **not** created in Excel, but serve to explain how pie charts take up a lot of space and can be converted into stacked charts.
- Having two separate pie charts for female and male makes it difficult to easily visualize the gender gap.
Changing multiple pie charts into stacked chart

➢ Click one of the original pie charts then click the “Insert” tab in the Excel ribbon (top toolbar menu).
➢ Under the "Charts" section click the bar chart looking icon and select the stacked bar chart.
Changing multiple pie charts into stacked chart

Alternatively, click one of the original pie charts then click the “Chart Design” tab in the toolbar menu at the top. Then click the “Change chart type” icon and select stacked bar chart.
Adjusting data selection for a chart

- The chart is currently only picking up the "Female" column so it just looks like one bar of data.
- To adjust the data range for the chart, right click the chart and select “Select Data”.
- The “Select Data Source” box will pop up highlighting the current data range in dashed green lines.
Readjusting data for stacked bar chart

- Highlight the full range of data to include both the female and male columns. Then click “OK” or hit the "Enter" key.
- Now there should be both female and male data in the chart.
Widening the plot area

- Click the plot area and drag the right side of the area towards the right to make the plot area wider.
  - To drag, you must hover over the circular point for the mouse to turn into an arrow (see video).
  - You will see a light blue outline if done correctly while you are dragging the plot area.
  - If you happen to move the plot instead of widen it, you can still adjust it – it might take more steps.
➢ Go to Sheet named "Task 6" in the Excel file "Training Dataset Day 2".
➢ Adjust the chart in the sheet by changing one of the two pie charts to a stacked bar chart.
➢ Make sure to readjust the data structure to ensure there are two bar colors on the chart, not one.
Reformatting the stacked bar chart

➢ The stacked bar chart rows are by “Reason for not seeking employment” when they should be by “Gender”.
➢ To reformat, click the chart, then click the “Chart Design” tab in the Excel ribbon (top toolbar menu).
➢ Then click “Switch Row/Column" and you will see the chart transform.
Finalizing the chart title and bar thickness

- The chart title needs to be adjusted as well as the color scheme for the stacked bar charts.
- Click the chart title and type in the new title.
- Click the bar segment and in the “Format Data Series” panel to the right, click the bar icon.
- Then under “Series Options” use the slider to adjust the thickness of the bars.
  - Especially when there are only two bars, the gap between the bars should be smaller than the bars themselves. In the image on the left, the gap is 106% and in the image on the right the gap is 57%.

![Image of chart title and bar thickness adjustment]

Reasons for not seeking employment by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Full time student</th>
<th>Too young/Too old</th>
<th>Disabled/Ill</th>
<th>My spouse wouldn't allow that</th>
<th>Occupied with home duties</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Go to the same chart you adjusted in Task 6 in the Sheet named "Task 6" in the Excel file "Training Dataset Day 2".
Adjust the stacked bar chart by using the "Switch Row/Column" step so that there are only two stacked bars, one for female and another for male.
Adjust the bar thickness of the stacked bar chart.
Changing the bar colors

➢ Click the bar segment that corresponds to “Other” which is currently green and change it to grey as “Other” and “No data” should be neutral colors to minimize attention.
   • Click on the paint bucket icon in the “Format Data Series” panel to the right. Under “Fill”, click on the paint bucket dropdown under “Color” and choose grey.
➢ Repeat the steps to adjust the other bar segment colors. Either select from the colors available or click “More colors” and add hex codes under the “Custom” tab.
Choosing appropriate colors

➢ Choose categorical colors (not sequential or diverging).
➢ Choose colors that will work with the rest of the report. Do not reuse colors that have been assigned for other categories “female”, “male”, “total”, “urban”, “rural”, etc.
➢ For the bar segments that you want to draw attention to like “Occupied with home duties” or “My spouse wouldn’t allow that” choose darker or brighter colors. Choose lighter colors for those bar segments that don’t seem to have much of a gender gap like “Too young/too old”, and “Disabled/ill”.

Categorical

Sequential

Diverging
Selecting hex codes and color palettes

- Hex codes can be found at htmlcolorcodes.com and https://www.computerhope.com/htmcolor.htm
- Color palettes can be found at colorhunt.co or colormind.io
- You can also find these by typing searching in an internet search engine like Google Chrome.
- The hex code will appear upon hovering over the color or written right underneath the color.
Choosing appropriate colors

Intentionally choosing colors transforms a visual from looking “outdated” to “polished”.

Reasons for not seeking employment by gender

- Full time student
- Disabled/Ill
- Occupied with home duties
- Other
- Too young/Too old
- My spouse wouldn't allow that

Male

Female

Male

Female
Alternative versions

➢ Choose colors that will work with the rest of the report. Do not reuse colors that have been assigned for other categories “female”, “male”, “total”, “urban”, “rural”, etc.

![Reasons for not seeking employment by gender](chart1)

![Reasons for not seeking employment by gender](chart2)

![Reasons for not seeking employment by gender](chart3)

![Reasons for not seeking employment by gender](chart4)
Go to the same chart you adjusted in Task 6 and 7 in the Sheet named "Task 6" in the Excel file "Training Dataset Day 2".

Choose a categorical palette for the stacked bar chart out of the following options.

Use the eyedropper tool to find the hex codes on the color palettes below, then assign a color to each category.
Reasons for not seeking employment by gender

**BEFORE**

- **Female**
  - Full time student: 22%
  - Too young/Too old: 16%
  - Disabled/Ill: 18%
  - My spouse wouldn’t allow that: 9%
  - Occupied with home duties: 12%
  - Other: 9%

- **Male**
  - Full time student: 24%
  - Too young/Too old: 12%
  - Disabled/Ill: 16%
  - My spouse wouldn’t allow that: 9%
  - Occupied with home duties: 18%
  - Other: 23%

**AFTER**

- **Female**
  - Full time student: 23%
  - Too young/Too old: 16%
  - Disabled/Ill: 9%
  - My spouse wouldn’t allow that: 18%
  - Occupied with home duties: 24%
  - Other: 43%

- **Male**
  - Full time student: 43%
  - Too young/Too old: 9%
  - Disabled/Ill: 16%
  - My spouse wouldn’t allow that: 9%
  - Occupied with home duties: 22%
  - Other: 24%
TRANSFORMING A GROUPED COLUMN CHART WITH MANY CATEGORIES INTO A BULLET BAR
Discussion – Can you understand the main insight? What other ways could this visual be shown that are more efficient? What would you change in this visual based on data visualization principles?
Discussion – Can you understand the main insights better? Which one shows the gender gap better?

Is there anything you would still change in the visual on the right?
The bullet bar compared to the clustered column chart:

- is easier to identify the gender gap at a glance with the skinny bar extending past the thicker;
- is cleaner and more consolidated (allows for many disaggregations to fit in one chart);
- allows the category labels to be horizontal for ease of reading;
- brings variety to the visualization types within a report, while still remaining close to the traditional bar/column chart.
Changing chart type and removing data labels

➢ Click the original chart and then click the “Insert” tab in the Excel ribbon (top toolbar menu).
➢ Under the "Charts" section, click the bar chart looking icon and select the clustered bar chart.
➢ Then click on the chart and the “Chart Elements” plus symbol and uncheck the box next to “Data labels”.
Overlapping bars to create bullet bars

- Overlapping bars will have either the female values on top or the male values or vice versa.
  - Whichever bar is on top is thinner and the bar behind it is thicker in order to properly see the gap.
  - If the bar behind is thinner, the data may end up being completely hidden by the thicker bar on top.
- Click on the bar series that you would like to show on top. The “Format Data Series” panel will show up.
- Click on the bar chart icon in order to see the series options.
Overlapping bars to create bullet bars

- After selecting the bars to go on top, under the “Series Options – Plot Series On” switch primary axis to secondary axis. You will notice the bars are now overlapped and there is another axis on top of the chart.
- Under “Gap Width”, move the slider to the right. You will see the orange bars getting thinner. The percentage in the box next to the slider will increase – you can also enter a percentage in that box (for example, 430%).
Overlapping bars to create bullet bars

- Select the other colored bar series and go directly to the “Gap Width” section within the “Format Data Series” panel. Do not touch anything in the “Plot Series On” section.
- Under “Gap Width”, this time move the slider to the left. You’ll see the blue bars getting thicker. The percentage in the box next to the slider will decrease – you can also enter a percentage in that box (for example, 50%).
➢ Go to Sheet named "Task 9" in the Excel file "Training Dataset Day 2".
➢ Choose which bar series will go on top.
➢ Adjust the bars so that the bar series you want on top is overlapping the other bar series.
➢ Adjust the top bar series so they are skinnier than the other bar series.
➢ Adjust the bottom bar series so they are thicker.
Sorting the bars by value

➢ Click in the column for the variable which you wish to sort. In this case we will sort by female values.
➢ Then under the “Home” tab in the Excel ribbon (top toolbar menu), click on the “Sort & Filter” button in the “Editing” section. Then click “Sort A to Z” (or “Sort Z to A” for the other direction).
Finalizing the chart

- Click the chart title and add an appropriate chart title.
- Remove the secondary axis at the top of the chart by clicking the axis and hitting the delete button.
- Adjust the colors as per preference or consistency with the report.
Proportion in tertiary education by gender and field of study

BEFORE

AFTER

Proportion in tertiary education by gender and field of study (%)
TRANSFORMING A GROUPED COLUMN CHART INTO A LINE CHART
GROUP DISCUSSION

Discussion – What other ways could this visual be shown that are more efficient? What do you like about this visualization? What would you change in this visual based on data visualization principles?

Prevalence of stunting (% of children under 5) by gender and severity

<table>
<thead>
<tr>
<th>Year</th>
<th>Stunting</th>
<th>Severe Stunting</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>2004</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>2006</td>
<td>37</td>
<td>16</td>
</tr>
<tr>
<td>2011</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>2014</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>2018</td>
<td>34</td>
<td>15</td>
</tr>
</tbody>
</table>

Girls

Boys
Group Discussion

➢ Discussion – Can you understand the time series insights better? Is there anything you would still change in the visual on the right?
The line chart compared to the side-by-side clustered column chart:

- is easier to identify the widening and narrowing of the gender gap over time;
- looks less busy/cluttered and takes up less space in a report;
- Allows better comparison between the “Stunting” and “Severe stunting” not just the gender gap.
Changing chart type to line chart

➢ Click the original chart and then click the “Insert” tab in the Excel ribbon (top toolbar menu).
➢ Under the "Charts" section click the line chart looking icon and select the line chart with the dots.
Adjusting the frequency of years on the x-axis

➢ The current chart is misleading with inconsistent frequency of years along the x-axis. To adjust this, click the x-axis of the chart with the dates. You should see the “Format Axis” panel on the right side.
➢ In the “Format Axis” panel, click the bar chart icon and open the “Axis Options” section.
➢ Under “Axis Options”, change the setting from “Automatically select based on data” to “Date axis”. You should now see that the x-axis has equally spaced years from 1991 to 2017.
Adjusting the line type

- Instead of using four colors, two colors can be used for female/male and two line types can be used for the other disaggregation severe stunting/stunting.
- Click on the line you wish to change and then click the paint bucket icon within the “Format Data Series” panel on the right. Under "Line" go to the "Dash type" dropdown menu and select your preferred line type. Here we selected the first dotted line option.
- Repeat this for all other lines in the same disaggregation category.
Adjusting the size of the dots

➢ To avoid misleading readers, it is essential to use dotted line charts to show the gaps in years of data collection. These dots are called “Markers” and the default size is 5.
➢ To make the dots more visible, click on the line with the dots you wish to change and then click the paint bucket icon within the “Format Data Series” panel on the right. Click "Marker" and go to "Marker Options".
➢ Change the setting from “Automatic” to “Built-in”. Then under "Size", type in either a number or use the arrows to increase the size. Here we increased it to 7.
➢ Repeat this until all the dotted lines have the same size “dots”/”markers”.

1. 2. 3. 4.
To ensure that the female and male colors are consistent in the chart, the solid lines (i.e. severe stunting) need to be changed so that "Severe Stunting Female" is blue and the "Severe Stunting Male" is orange.

1. Click on the yellow line (currently Severe Stunting Male) and then click the paint bucket icon within the “Format Data Series” panel on the right. Under “Line” go to the color dropdown menu and select orange.

2. Under “Marker” go to the color dropdown menu under both “Fill” and “Border” and select orange.

3. Repeat this for the "Severe Stunting Female" grey line, changing both the line and the dots to blue.
Moving the legend and adding a title

➢ Click the chart title and add an appropriate chart title.
➢ To move the legend to the right side, click on the chart and then on the “Chart Elements” plus symbol.
➢ To the right of “Legend”, click on the arrow. Then select the position of the legend – in this case “Right”.

---

Prevalence of stunting (% of children under 5) by gender and severity

- Stunting Female
- Stunting Male
- Severe Stunting Female
- Severe Stunting Male
The chart seems a bit squished making the years on the x-axis unreadable at a 90 degree angle.
Pull the left side of the chart at the circle towards the left until you see the years are horizontal again.
Prevalence of stunting (% of children under 5) by gender and severity

- **BEFORE**
- **AFTER**

![Graph showing prevalence of stunting by gender and severity from 1990 to 2018.](image-url)
➢ Go to Sheet named "Task 10" in the Excel file "Training Dataset Day 2".
➢ Change the line type for one of the location disaggregation lines from solid to dashed (either for national or for urban).
➢ Resize the marks (dots) along the line to be larger (size 7 for example).
➢ Change the color of the other disaggregation so that there are only two colors in the chart – one for female and the other for male.
PRACTICE EXERCISE:
Create a visual of your own with data from the Training Dataset file
Exercise – create your own visual

- Using any data from any sheet in the "Training Dataset Day 1" or "Training Dataset Day 2" Excel files, create your own visual with more than four data points.
- The gender factbook’s colors are
  - Female: #006389; RGB (0, 99, 137)
  - Male: #00B0AB; RGB (0, 176, 171)
- Use the "Data Visualization Tip Sheet" handout/Word file to review your visual prior to finalizing.

- A few major questions to keep in mind (more details in the "Data Visualization Tip Sheet" handout):
  - Is the visual showing the gender-relevant insight or does data need to be transposed or restructured?
  - Does it look clean or cluttered? Are there unnecessary gridlines, labels, axes duplicate encodings?
  - Does the visual have proper sorting, grouping, legend order, etc.?
  - Do the colors align? Are they intentional?
  - Can any aspect of the chart be misleading for the audience?
ANNEX
Video for widening the plot area

- Click the plot area and drag the right side of the area towards the right to make the plot area wider.
  - To drag, you must hover over the circular point for the mouse to turn into an arrow (see video).
  - You will see a light blue outline if done correctly while you’re dragging the plot area.
  - If you happen to move the plot instead of widen it, you can still adjust it – it might take more steps.

**WIDEN PLOT (CORRECT)**

**MOVE PLOT**
ADDITIONAL LINE CHART STEPS - LABELS
Adding data labels

➢ To add data labels, click on the chart and then on the “Chart Elements” plus symbol.
➢ Check the box next to “Data Labels”.

Prevalence of stunting (% of children under 5) by sex and severity

- Stunting Female
- Stunting Male
- Severe Stunting Female
- Severe Stunting Male

Chart Elements
- Axes
- Axis Titles
- Chart Title
- Data Labels
- Data Table
- Error Bars
- Gridlines
- Legend
- Trendline
- Up/Down Bars
Changing position of data labels

- The data labels are currently overlapping. The position of the data labels should be changed so that the line above has data labels above the line and the line below has data labels below the line.
- Click on the line that needs the data label change and then on the “Chart Elements” plus symbol. Click the arrow to the right of “Data Labels” and select “Below”. Repeat this for the other blue line.
Removing gridlines and y-axis labels

- Since there are data labels, there is no longer a need for the y-axis labels or gridlines.
- To remove the gridlines, click on the chart and then on the “Chart Elements” plus symbol and uncheck “Gridlines”.
- To remove the y-axis labels, click on the y-axis labels and hit the "Delete" button on the keyboard.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 6: DATAWRAPPER MAPS
Contents

1. Recap of data visualization steps and breakdown
2. Setting up the Datawrapper tool account login
3. Getting started with visualizing maps in Datawrapper
4. Designing, formatting, & annotating the data visualization
5. Downloading the visualization
Specific actions/functions

1. Selecting the chart type (slides 18-21)
2. Uploading data, reuploading data (slides 22-37, 40-42, 82-85)
3. Cleaning and adjusting data structure (slides 28-39)
   - in Datawrapper (slides 28 – 33)
   - in Excel (slides 34 – 39)
4. Choosing a color scale (slides 45 – 50)
5. Changing the color gradient (slides 51 – 56)
6. Changing the color legend (slides 58-63)
7. Adding legend labels (slides 64-66)
8. Adding data labels (slides 68-71, 87-89)
9. Merging values from two columns into a third column in Excel using =CONCATENATE() function (slides 72 – 80)
10. Finalizing annotations: title, subtitle, source (slide 89)
11. Downloading the visual (slides 91-95)
1. Recap of data visualization steps and breakdown
Recap: steps for creating a data visualization

1. Upload data
2. Explore/try visualization options
3. Adjust data structure
4. Design, format, and finalize visualization
5. Publish and/or download visualization
6. Optional: add additional annotations as needed

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- Match the required data structure, inputs, and features of the intended visualization type
- Highlight the right message and insights
- Enable easy formatting and annotation
Recap: data visualization breakdown

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

75% Properly cleaning, transposing & preparing the data

25% Formatting & annotating the visual
The type of visualizations you can create depend on the structure of the data tables (columns and rows).

See below two different structures of the same data points.

### Recap: data structure

#### A. Grouped by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>.6</td>
</tr>
</tbody>
</table>

#### B. Grouped by type of work

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>.6</td>
</tr>
</tbody>
</table>
Recap: data structure

The type of visualizations you can create depend on the structure of the data tables (columns and rows).

See below two different structures of the same data points with grouped bar chart examples.

A. Grouped by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>.6</td>
</tr>
</tbody>
</table>

Difference/gap between the types of unpaid work for a given gender.

B. Grouped by type of work

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>.6</td>
</tr>
</tbody>
</table>

Difference/gap between men’s and women’s time spent on a given type of unpaid work.
2. Setting up the Datawrapper tool account login
Set up Datawrapper account

➢ Type https://app.datawrapper.de/ in your internet browser.
➢ Click "Create a new account" in the bottom of the grey box.
➢ Sign up with your email or with another account like Google, Microsoft, Github, or Twitter.
Set up Datawrapper account

- Enter your email address and create a password at least 8 characters long.
Set up Datawrapper account

➢ Once you have created an account, it will automatically log you in. You should see this screen.
➢ In the right hand corner you will see a message alerting you that your email address needs to be confirmed.
➢ Check your inbox for the Datawrapper email.
Confirm Datawrapper account

➢ Once you find the email from Datatwrapper in your inbox, open it. It should look like this.
➢ Click the "Confirm your account" link in the email.
➢ Then log in to your Datawrapper account again.

Welcome to Datawrapper! 📧 Inbox 🔗

Datawrapper Team &lt;support@datawrapper.de&gt; to me ▼

Hi there!
Thanks for signing up for Datawrapper.
We’re very excited to have you on board.

To get started using Datawrapper, please confirm your account below:

**Confirm your account**

Thanks,
the Datawrapper Team

P.S. Need help getting started? Just reply to this email with any questions or issues you have. The Datawrapper team is always happy to help you.
Datawrapper account successfully confirmed

➢ After logging in you’ll see a green box in the top right that confirms you have successfully activated your email address for Datawrapper.
3. Getting started with visualizing maps in Datawrapper
Log in to Datawrapper

- Type [https://app.datawrapper.de/](https://app.datawrapper.de/) in your internet browser.
- Click "Sign in with Email".
Log in to Datawrapper

➢ Enter your email address and password.
➢ Click the blue "Login" button.
Select a chart category

➢ Click the "Create New" button or "Nouveau" button.
➢ Then select the chart type: Chart, Map, or Table. For this first visual we will select "Map".
Select a map type

➢ There are three map types: Chloropleth map, Symbol map, and Locator map.
➢ Select the "Chloropleth map".
Select a geographic region

- In the search bar, type a geographic region (i.e. world, region name, country name).
- For our visual, we will type in Cameroon to find the Cameroonian maps.
Select the geographic map level for Cameroon

- There are two geographic map levels: Departments and Provinces.
- Choose the level for which you have data. We will select "Provinces". You will see a preview of the map.
- Then click the "Proceed" button.
Upload data from Excel or CSV file

- Upload an Excel or CSV file by clicking the "Upload file" button. The box to select the file will pop up.
- Select the file from the folder on your computer and click "Open".
Upload data from Excel or CSV file

- A successful upload will show a green check mark next to the "Upload file" button.
- If there is more than one sheet in the Excel file, there will be a dropdown to select which sheet to use.
Match the data with the map

- Next to the "Upload" tab, click the "Match" tab.
- Make sure that the matching key is set to the right category (i.e. Name vs. Postal code for the province).
- Select and/or verify the column for the "Name" as well as the column for the "Values".

![Match data with map screenshot](image)

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addis Ababa</td>
<td>58.7</td>
</tr>
<tr>
<td>Central (Sanaa)</td>
<td>40.3</td>
</tr>
<tr>
<td>Sanaa</td>
<td>73.9</td>
</tr>
<tr>
<td>Est</td>
<td>32.3</td>
</tr>
<tr>
<td>Extreme-Nord</td>
<td>39.8</td>
</tr>
<tr>
<td>Littoral (Sanaa)</td>
<td>54.1</td>
</tr>
<tr>
<td>North</td>
<td>55.5</td>
</tr>
<tr>
<td>Nord-Best</td>
<td>59.9</td>
</tr>
<tr>
<td>Oresti</td>
<td>71.2</td>
</tr>
<tr>
<td>South-west</td>
<td>56.7</td>
</tr>
<tr>
<td>South-west</td>
<td>79.8</td>
</tr>
<tr>
<td>Taddi</td>
<td>78.1</td>
</tr>
</tbody>
</table>

Select matching key (i.e. Name vs. Postal code for the province) and verify columns for "Name" and "Values".
Next to the Match tab, click the Check tab. You’ll notice there is a yellow warning symbol and a yellow warning box that shows two geographic areas do not match with data.

In red, 4 region names do not match: Centre (Sands Yaounde), Yaounde, Littoral (Sans Douala), Douala.

The region names need to be adjusted in the data table to match.
The current data structure separates data for cities Douala and Yaounde from their provinces. The map only takes data for provinces so the data structure must be adjusted in Datawrapper or Excel.

### A. Current Data Structure

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38,7</td>
</tr>
<tr>
<td>Centre (Sans Yaoundé)</td>
<td>42,5</td>
</tr>
<tr>
<td>Douala</td>
<td>78,9</td>
</tr>
<tr>
<td>Est</td>
<td>32,3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35,8</td>
</tr>
<tr>
<td>Littoral (Sans Douala)</td>
<td>54,1</td>
</tr>
<tr>
<td>Nord</td>
<td>31,5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53,9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71,2</td>
</tr>
<tr>
<td>Sud</td>
<td>50,7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72,8</td>
</tr>
<tr>
<td>Yaoundé</td>
<td>70,1</td>
</tr>
</tbody>
</table>

### B. Required Data Structure for Map

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38,7</td>
</tr>
<tr>
<td>Centre</td>
<td>42,5</td>
</tr>
<tr>
<td>Est</td>
<td>32,3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35,8</td>
</tr>
<tr>
<td>Littoral</td>
<td>54,1</td>
</tr>
<tr>
<td>Nord</td>
<td>31,5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53,9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71,2</td>
</tr>
<tr>
<td>Sud</td>
<td>50,7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72,8</td>
</tr>
</tbody>
</table>
CLEANING/ADJUSTING THE DATA
OPTION 1: DATAWRAPPER
Filling in blank cells of data in Datawrapper

➢ Copy or type the values for Centre (Sans Yaounde) and Littoral (Sans Douala) into the blank cells in the value column for the regions Centre and Littoral respectively.
Filling in blank cells of data in Datawrapper

- When the values area added in the blank cells, the rows will be numbered.
- The yellow warning box will disappear as well as the phrase “2 unused” at the bottom of the sheet.
Removing rows of data in Datawrapper

- The rows of data that are no longer needed and still appear red should be deleted.
- Highlight the entire row by clicking the row number on the left. It should highlight the row in blue.
- It should also show a red box with a trash can symbol. Click on the symbol to delete the row.
Removing rows of data in Datawrapper

- You should then see that the row has disappeared and a popup signifying that you deleted a row.
  - In case you deleted the wrong row, you can retrieve it by clicking the “Revert” button.
- One of the red boxes (corresponding to the deleted region row) should also have disappeared and the phrase at the bottom of the sheet should have changed from “4 errors” to “3 errors”.
Removing rows of data in Datawrapper

➢ Repeat for all rows of data that must be deleted until there are no red “errors” at the bottom of the datasheet and the red boxes to the left have become a message in green that “All map regions were matched to a row in your dataset”.

➢ Click the "Proceed" button.
CLEANING/ADJUSTING THE DATA
OPTION 2: EXCEL
Recap: data structure required by map

➢ The current data structure separates data for cities Douala and Yaounde from their provinces.
➢ The map only takes data for provinces so the data structure must be adjusted in Datawrapper or Excel.

A. Current Data Structure

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38.7</td>
</tr>
<tr>
<td>Centre (Sans Yaoundé)</td>
<td>42.5</td>
</tr>
<tr>
<td>Douala</td>
<td>78.9</td>
</tr>
<tr>
<td>Est</td>
<td>32.3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35.8</td>
</tr>
<tr>
<td>Littoral (Sans Douala)</td>
<td>54.1</td>
</tr>
<tr>
<td>Nord</td>
<td>31.5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53.9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71.2</td>
</tr>
<tr>
<td>Sud</td>
<td>50.7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72.8</td>
</tr>
<tr>
<td>Yaoundé</td>
<td>70.1</td>
</tr>
</tbody>
</table>

B. Required Data Structure for Map

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38.7</td>
</tr>
<tr>
<td>Centre</td>
<td>42.5</td>
</tr>
<tr>
<td>Est</td>
<td>32.3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35.8</td>
</tr>
<tr>
<td>Littoral</td>
<td>54.1</td>
</tr>
<tr>
<td>Nord</td>
<td>31.5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53.9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71.2</td>
</tr>
<tr>
<td>Sud</td>
<td>50.7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72.8</td>
</tr>
</tbody>
</table>
Adjust the region names (Edit cell text in Excel)

- For the region « Centre (Sans Yaounde) », adjust the name so it says only « Centre ».
- For the region « Littoral (Sans Douala) », adjust the name so it says only « Littoral ».
- This can be done by clicking into the cell and deleting everything in the parentheses.
### Remove unnecessary rows in Excel

- The cities Douala and Yaounde still show up as their own rows of data. They must be deleted.
- To delete a row, highlight the entire row by clicking the row number on the left. It should highlight the entire row in a shaded grey.
- Right click on the highlighted row and select “Delete”.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Region</td>
<td>Birth</td>
<td>Registration</td>
</tr>
<tr>
<td>2</td>
<td>Adamaoua</td>
<td>38.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Centre</td>
<td>42.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Douala</td>
<td>78.9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Est</td>
<td>32.3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Extrême-Nord</td>
<td>35.8</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Littoral</td>
<td>54.1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Nord</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Nord-Ouest</td>
<td>53.9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Ouest</td>
<td>71.2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Sud</td>
<td>50.7</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Sud-Ouest</td>
<td>72.8</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Yaounde</td>
<td>70.1</td>
<td></td>
</tr>
</tbody>
</table>

1. Highlight the row by clicking the row number on the left.
2. Right click on the highlighted row and select “Delete”.

#### Image Description
- Step 1: Highlighting the row.
- Step 2: Right-click menu showing the “Delete” option.
Remove unnecessary rows in Excel

➢ You should then see that the row has disappeared.
  • In case you deleted the wrong row, you can “undo” the action by clicking on the “Undo” button in the toolbar menu at the top.
➢ The final dataset after removing both rows for Douala and Yaounde is on the right side.
Save the Excel file so that the changes are not lost.

- It can be saved with the save symbol or by clicking File in the toolbar menu on the top.
- If you do not want to overwrite the Excel file, click Save As instead of Save and save the file under a different name in your preferred location. For example, *Cameroon Map Data Updated*. Click "Save".
Reupload data to Datawrapper

➢ Option 1: Reupload Excel/CSV file by clicking the "Upload file" button, selecting a file, and clicking "Open".
➢ Option 2: Copy the data from the Excel file and paste it in the blank box and click the arrow on the right.
Successful upload of the correct data structure

- With the Copy/Paste method, there will be a popup notifying that data was pasted with option to revert.
- To verify that all data are correctly uploaded or pasted, review both the "Match" and "Check" tabs again and adjust accordingly if needed.
Successful upload of the correct data structure

➢ The "Check" tab shows a green check mark and a green box that says "All map regions were matched to a row in your dataset"
➢ Click the "Proceed" button.
4. Designing and formatting the visual
In the "Visualize" tab, you will already see a preview of your map with the default elements. On the left are buttons to customize the colors, legend, and layout.
The default color scale

- The default is a linear continuous color scale with a green/blue color palette ranging from the minimum value of the dataset to the maximum value of the dataset.
- Values are automatically assigned a color based on the range. If the range is changed to the minimum and maximum possible for percentages (0 and 100), then variations will not be visible (see right map).
Continuous vs. Step scale

➢ To have more control over the color scale, under “Colors” within the “Type” section, change “continuous” to “steps”. This will produce steps or buckets of colors that can be adjusted by you.

➢ You can choose the number of buckets, the way values are assigned per buckets, or completely customize the bucket ranges.

- **Continuous**: assigns each value a different color on the gradient
- **Steps**: assigns all values within a specified range (bucket/step) the same color
Default step color scale: linear

- Under “Colors” within the “Steps” section, the default number of steps is 5 and can be changed up to 20.
- The default steps are “linear” meaning each range is equidistant between the minimum and maximum.

Ranges | Range size
--- | ---
31.5-39.76 | 8.26
39.76-48.02 | 8.26
48.02-56.28 | 8.26
56.28-64.54 | 8.26
64.54-72.8 | 8.26
Adjusting color steps: quantiles

- The “linear steps” can be changed to “quantiles”, meaning there is an equal number of observations in each step.
- In the quantile range below, for example, there are two regions per step/color.

<table>
<thead>
<tr>
<th>Range size</th>
<th>5 Quantile Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>31.5-35.1</td>
</tr>
<tr>
<td>5.88</td>
<td>35.1-40.98</td>
</tr>
<tr>
<td>11</td>
<td>40.98-51.98</td>
</tr>
<tr>
<td>5.54</td>
<td>51.98-57.52</td>
</tr>
<tr>
<td>15.28</td>
<td>64.54-72.8</td>
</tr>
</tbody>
</table>
Adjusting color steps: custom

➢ You can also customize the ranges and set the beginning and end value of each range.
➢ Under "Colors", within the “Steps” section select “Custom” from the dropdown list. Then enter the custom beginning and end values for each range.
CHANGING THE COLOR GRADIENT
Selecting a different color palette

➢ Under "Colors", within the “Select palette” section, click the color dropdown list.
➢ Select a different color gradient. Remember some are sequential and others are diverging.
Selecting a different color palette

➢ Under "Colors", within the “Select palette” section, click the color dropdown list.
➢ Select a different color gradient. Remember some are sequential and others are diverging.
Customizing the color palette

➢ If you don’t like the color palette options or must match the branding for a report, they can be changed.
➢ Under Color, within the “Select palette” section, click the wrench button.
➢ A color gradient with several color stops will pop up. Click the reverse button to draw attention to low values rather than high values.
Customizing the color palette

➢ Change the color gradient by sliding one or more of the color stops left or right.
➢ Here by moving the color stops towards the green (left) section, the whole color gradient shifts towards the blue hues than the green ones.
Customizing the color palette

- Change the color gradient by clicking one or more color stops and selecting new colors or entering the hex codes of the desired color.
- On the left, 4 color stops from the original green/blue color palette were changed so that they are all shades of blue (no green).
- On the right, all the color stops were changed to different shades of red.

Adjusting the color

HEX code
In the "Refine" tab, choose a color gradient and change the color scale.
The version you submit should have different colors and scale than what is used in the demonstration walk-through.
Moving the color legend (steps scale)

➢ Under “Legend” in the “Position” section select a different position from the dropdown.
➢ Steps example:
Making the color legend vertical (steps scale)

- Under “Legend” in the “Orientation” section switch from horizontal to vertical.
- The steps are now neatly vertically ordered from lowest values to highest values.
Moving the color legend (continuous scale)

- Under “Legend” in the “Position” section select a different position from the dropdown.
- Continuous example:
For continuous scale in this case moving the legend to the top right is too close to the map so it might be best to make it vertical.
Adjusting the size of the color legend

- Under “Legend” within the “size” section, you can change the size by sliding the bar or entering a number.
- The legend can also be moved slightly from its position under “Legend within the “Offset” section by typing in a number up to 50 in the horizontal box (which moves it inward). Numbers in the vertical box would move the legend up or down from its current position.
ADDING LEGEND LABELS
Adding legend labels

- Under “Legend” in the “Labels” switch from “range” to “custom”.
- It will show a box for Min and Max and potentially for Center if you already specified a median number.
- The default text will say “Low”, “Medium”, “High”
Adding legend labels (gender gap)

- Generally, for gender gaps the values range from negative to positive, which may be confusing to readers.
- When charting gender gaps, you always want to change the legend labels/caption to be more user-friendly.
- Gender gap examples below with diverging color scale vertical position so labels don’t overlap:
In the "Refine" tab, adjust the legend so that it is not using the default position or orientation. Move and change the legend, position, orientation, and (if you prefer) the labels as well.
ADDING DATA LABELS
Adding data labels

➢ To add data labels, switch from the “Refine” tab to the “Annotate” tab.
➢ Under the “Map Labels” section, click “Show labels”
➢ The default is to add names of places, like cities.
Adding data labels

➢ Under “Map Labels”, within the “Type” section, switch from “places” to “columns” which will allow you to select from the dropdown which column within the dataset to display values for.
➢ The column named Birth Registration shows the data values. Region would show only region names.
Adding data labels (with regions and values)

➢ To add labels with *both* region names and data values, there needs to be another column in the dataset with this combined information (similar to the previous data label exercise in Excel).
➢ This column can be created in Excel and then the data can be reuploaded into Datawrapper as before.

A. Required Data Structure for Map

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38.7</td>
</tr>
<tr>
<td>Centre</td>
<td>42.5</td>
</tr>
<tr>
<td>Est</td>
<td>32.3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35.8</td>
</tr>
<tr>
<td>Littoral</td>
<td>54.1</td>
</tr>
<tr>
<td>Nord</td>
<td>31.5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53.9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71.2</td>
</tr>
<tr>
<td>Sud</td>
<td>50.7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72.8</td>
</tr>
</tbody>
</table>

B. Required Data Structure for Map with Region and Data Labels

<table>
<thead>
<tr>
<th>Region</th>
<th>Birth Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adamaoua</td>
<td>38.7</td>
</tr>
<tr>
<td>Centre</td>
<td>42.5</td>
</tr>
<tr>
<td>Est</td>
<td>32.3</td>
</tr>
<tr>
<td>Extrême-Nord</td>
<td>35.8</td>
</tr>
<tr>
<td>Littoral</td>
<td>54.1</td>
</tr>
<tr>
<td>Nord</td>
<td>31.5</td>
</tr>
<tr>
<td>Nord-Ouest</td>
<td>53.9</td>
</tr>
<tr>
<td>Ouest</td>
<td>71.2</td>
</tr>
<tr>
<td>Sud</td>
<td>50.7</td>
</tr>
<tr>
<td>Sud-Ouest</td>
<td>72.8</td>
</tr>
</tbody>
</table>
MERGING VALUES FROM TWO COLUMNS INTO A NEW COLUMN IN EXCEL: COMBINED REGION AND DATA LABELS
Adding a third column title

➢ Open the Excel file and go to the sheet with the data for the correct regions for this map.
➢ Click into the top cell of the column (C) to the right of the “Birth registration” column (B) with data values.
➢ Enter the name of the new column. It can be “Data labels” to easily identify the column later.
Merging information with \( =\text{CONCATENATE()} \)

- Instead of using a copy/paste method to copy information from column A and B into column C, there is a function \( =\text{CONCATENATE(cell1,cell2)} \) that can be entered into the cells for column C.
- Double click into the cell in the second row and enter the following: \( =\text{CONCATENATE(} \)
- After typing the parenthesis \( ( \), click directly into the cell in the second row of the first column. It should highlight the cell and add the cell number into the function in the third column. \( =\text{CONCATENATE(A2} \)
  - It should look like the image on the right. Do NOT click anywhere else.
Using the \texttt{=CONCATENATE()} function

➢ After successfully adding the first cell (A2) in the previous step, then type \texttt{=CONCATENATE(A2," ")}:
  
  • The function should look like \texttt{=CONCATENATE(A2," ")}:
    - This will add a space between the information in column A and column B.
    - It should \textbf{not} look like \texttt{=CONCATENATE(A2,"")}, with the quotation marks next to each other or else it will not provide the space between the region name and birth registration value.

➢ It should look like the image on the right. Do NOT click anywhere else.
Using the `=CONCATENATE()` function

- After successfully adding the quotations, (",", ) in the previous step, then click directly into the cell in the second row of the second column.
  - It should highlight the cell and add the cell number into the function in the third column. The function should look like `=CONCATENATE(A2,"",B2`.
- It should look like the image on the right. Do NOT click anywhere else.
After successfully adding the cell of the second row in the second column (B2) in the previous step, then type a closing parenthesis ) and hit enter.

- It should look like this before you hit enter: `=CONCATENATE(A2,"",B2)`

The final after hitting enter should be the value from the second row in the first column Adamaoua and the value from the second row of the second column 38.7 separated by a space: Adamaoua 38.7

It should look like the image on the right.
Full video of `=CONCATENATE()` function
Replicating the function for multiple rows

- After successfully merging or concatenating information into the first row of the third column, repeat this for all rows in the column.
- The easiest way to repeat is to drag the function down for the whole column.
- Click (do not double click) on the cell with the result/function (second row of the third column).
- Then in the bottom right corner of the same cell, click and drag until the last row of data. You should see a green outline when you’re still clicking, but when you release the mouse, you’ll see the results in each row.
Video replicating the function for multiple rows

- After successfully merging or concatenating information into the first row of the third column, repeat this for all rows in the column.
- The easiest way to repeat is to drag the function down for the whole column.
- Click (do not double click) on the cell with the result/function (second row of the third column).
- Then in the bottom right corner of the same cell, click and drag until the last row of data. You should see a green outline when you’re still clicking, but when you release the mouse, you’ll see the results in each row.
Go to Sheet named "Exercise 5 DW Map" from the "Training Dataset Day 2" Excel file.
Merge values from the first two columns into the third column using the =CONCATENATE() function.
What We Do

REUPLOADING DATASET WITH DATA LABELS
Reupload data to Datawrapper

➢ Go to the « Add your data » tab.
➢ Option 1: Reupload Excel/CSV file by clicking the Upload file button, selecting a file, and clicking Open.
➢ Option 2: Copy the data from the Excel file and paste it in the blank box and click the arrow on the right.
Successful upload of the revised dataset

- With the Copy/Paste method, there will be a popup notifying that data was pasted with option to revert.
- To verify that all data are correctly uploaded or pasted, review both the Match and Check tabs again and adjust accordingly if needed.
Successful upload of the revised dataset

➢ The Check tab shows a green check mark and a green box that says « All map regions were matched to a row in your dataset »
➢ Click the proceed button.
FINALIZING ANNOTATIONS
(TITLE, SOURCE, NEW DATA LABELS)
Go to the “Visualize tab” and then switch from the “Refine” tab to the “Annotate” tab.

Under “Map Labels” within the “Select column” section, you will see the “Birth Registration” column is still selected.
Select new column for data labels

- Under “Map Labels” within the “Select column” section, select the name of the column with the combined region and data labels you just created.
- In the example case, this column is named “Data labels”. The new labels should appear on the map.
Adding the title, description, and source

➢ Under the “Title” section, add a title in the blank box provided.
➢ Under the “Description” section, add any additional details or subtitles in the box provided.
➢ Under the “Data source” section, add the name and year of survey in the box provided.
5. Downloading the visual
Download the visual

- Go to the "Publish & Embed" section after finalizing your visual.
- Click the PNG image to view the download options.
Download the visual

➢ Under “Include”, choose whether to download “Just chart” or the chart with “Full header and footer”.
➢ Under “Background”, choose whether you would like the background to be included (white) or transparent (no background).
➢ Then click “Download image”.

![Image of visualization download options](image-url)
“Transparent” background is a good option for adding the visual to a social media card or infographic that may have a colored background.
“Just chart” is the best option if you want to add your own annotations or if the title and source are being added directly in the report and not in each chart.
A note on publishing

➢ The visual does NOT need to be published in order to download a static image file.
➢ If you wish to create an interactive online version with a URL, you will first need to click Publish Now.
➢ When you publish a visual, the data will be visible to those who have the URL. Data cannot be downloaded, but the data values can be seen on the data labels and the tooltip if either are enabled.
PRACTICE EXERCISE 5:
Create a map in Datawrapper with data from the Training Dataset file
Exercise – create a map in Datawrapper

➢ Using the data in Sheet "Exercise 5 DW Map" from "Training Dataset Day 2" Excel file, create a map in Datawrapper.

➢ Tips to remember:
  • The Gender Parity Index indicator has values above and below 1 with 1 indicating equality (gender parity).
  • Diverging colors scales are used for diverging number scales.
  • Captions, annotations, notes, etc. should be used if the data must be interpreted in a certain way by the reader that may not be obvious.

➢ Time for exercise: 20 minutes.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 7: DATAWRAPPER RANGE PLOTS
Contents

1. Recap of data visualization steps and breakdown
2. Setting up the Datawrapper tool account login
3. Getting started with visualizing maps in Datawrapper
4. Designing, formatting, & annotating the data visualization
5. Downloading the visualization
Specific actions/functions

1. Selecting the chart type (slides 18-19)
2. Exploring a sample dataset (slides 20-22)
3. Uploading data (slides 23-25)
4. Changing the chart type (slides 27 – 34)
5. Sorting data values (slides 35-37)
6. Finalizing annotations: title, subtitle, source (slides 38-39)
7. Downloading the visual (slide 40-43)
1. Recap of data visualization steps and breakdown
Recap: steps for creating a data visualization

Upload data

Explore/try visualization options

Adjust data structure

Design, format, and finalize visualization

Publish and/or download visualization

Optional: add additional annotations as needed
Recap: data visualization breakdown

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

25% Formatting & annotating the visual

75% Properly cleaning, transposing & preparing the data
The type of visualizations you can create depend on the structure of the data tables (columns and rows).

See below two different structures of the same data points.

### Recap: data structure

- **A. Grouped by gender**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

- **B. Grouped by type of work**

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Recap: data structure

➢ The type of visualizations you can create depend on the structure of the data tables (columns and rows)
➢ See below two different structures of the same data points with grouped bar chart examples

A. Grouped by gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Unpaid Domestic Work</th>
<th>Unpaid Care Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>15.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Male</td>
<td>4.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

B. Grouped by type of work

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unpaid Domestic Work</td>
<td>15.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Unpaid Care Work</td>
<td>3.4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Difference/gap between the types of unpaid work for a given gender.

Difference/gap between men’s and women’s time spent on a given type of unpaid work.
2. Setting up the Datawrapper tool account login
Set up Datawrapper account

➢ Type [https://app.datawrapper.de/](https://app.datawrapper.de/) in your internet browser.
➢ Click "Create a new account" in the bottom of the grey box.
➢ Sign up with your email or with another account like Google, Microsoft, Github, or Twitter.
Set up Datawrapper account

➢ Enter your email address and create a password at least 8 characters long.
Set up Datawrapper account

➢ Once you have created an account, it will automatically log you in. You should see this screen.
➢ In the right hand corner you will see a message alerting you that your email address needs to be confirmed.
➢ Check your inbox for the Datawrapper email.
Once you find the email from Datatwrapper in your inbox, open it. It should look like this.

Click the "Confirm your account" link in the email.

Then log in to your Datawrapper account again.

Welcome to Datawrapper!

Datawrapper Team <support@datawrapper.de>

to me ▼

Hi there!

Thanks for signing up for Datawrapper.
We’re very excited to have you on board.

To get started using Datawrapper, please confirm your account below:

Confirm your account

Thanks,
the Datawrapper Team

P.S. Need help getting started? Just reply to this email with any questions or issues you have. The Datawrapper team is always happy to help you.
Datawrapper account successfully confirmed

➢ After logging in you’ll see a green box in the top right that confirms you have successfully activated your email address for Datawrapper.
3. Getting started with visualizing maps in Datawrapper
Log in to Datawrapper

- Type [https://app.datawrapper.de/](https://app.datawrapper.de/) in your internet browser.
- Click "Sign in with Email".
Log in to Datawrapper

- Enter your email address and password.
- Click the blue "Login" button.
Select a chart category

➢ Click the "Create New" button.
➢ Then select the chart type "Chart".
See an example of the data structure for a range plot

➢ There is an example dataset in Datawrapper for range plots.
➢ Click "Select a sample dataset" in the dropdown menu and select the dataset "Gender Pay Gap" under Range Plot.
See an example of the data structure for a range plot

➢ You will see that there are salaries for men and women by level of education for the United States in this sample dataset.
See an example of the data structure for a range plot

Go to the “Visualize” section to see what the sample dataset looks like.
See an example of the data structure for a range plot

➢ Return to the “Check & Describe” section to see how the dataset should look when you upload it for the gender pay gap data.
➢ Based on this you can now create your own dataset in this format and go back to the “Upload Data” section and paste the values into the white box or upload an Excel or CSV.
UPLOADING DATA
Option 1: Copy the data from the Excel file and paste it into the empty box. Click the “Proceed” button in the bottom right.
Uploading data

Option 2: Upload your Excel/CSV file by clicking on the "XLS/CSV Upload" button then "Upload a file", selecting a file, and then clicking "Open".
4. Designing and formatting the visual
CHANGING THE CHART TYPE
Changing the graph type

➢ Go to the “Visualize” section. This is the default chart type chosen for the data.
➢ This type of chart is not appropriate for the data. Select a different chart type by clicking on one of the chart images.
Changing the graph type

- Select the “Range Plot" chart.
- Now you'll see that the data are visualized in a range plot, but it's not in order.
CHANGING THE VISUALIZATION ELEMENTS
Changing the visualization elements

➢ Select the "Refine" tab to display options for editing visualization elements.
➢ Make sure that in the “Range" section, the drop-down menu for “Range start" and the drop-down menu for “Range end" are not the same. Depending on your preference, you can choose “Female" for "Start" and “Male" for "End" or vice versa.
Changing the visualization elements

➢ Under “Labels”, you can choose which data labels to display by toggling on the “Show Values” button and then making a choice under “Visibility”.
➢ Usually, you want to select to display “both”, which has the data labels for the start and end values.
➢ Sometimes this makes the graph too cluttered, so you can choose “difference” to show the gender gap.
Changing the visualization elements

- Sometimes selecting “both” makes the graph too cluttered. Here are the options for “difference” and “% change”.

**Both**

**Difference**

**% change**
Changing the visualization elements

- Under "Labels," click "Label First Range."
- You can now see that the category labels for “Female” and “Male” are visible.
SORTING THE VALUES IN THE CHART
Sorting the values in the chart

➢ Under the "Sorting & Grouping" section, toggle on the "Sort Rows" button.
➢ There are four options for sorting values. "Start", "End", "Difference", and "% change". Usually, the chart is sorted by "start" or "end," which means "Female" or "Male."
➢ This helps readers visually see the order.
Sorting the values in the chart

- You can reverse the order of the values by toggling on the "Reverse order" button.
- It's best to show this way - the higher values above and the lower values are lower.
FINALIZING THE CHART
Finalizing the chart

➢ Select the "Annotate" tab to finalize the chart with annotations.
➢ In addition to a title, it's important to add a subheading to explain the values. In this case, it's local currency.
➢ Under "Title," add a title in the empty box provided.
➢ Under "Description," add additional details or subheadings in the box provided.
➢ Under "Data source," add the name and year of the survey in the box provided.
5. Downloading the visual
The download options are the same as for the map exercise.

Go to the "Publish & Embed" tab after finalizing your design.

Click the PNG image to view the download options.

Under "Include," choose to download "just chart" or the chart with "full header and footer."

Under "Background," choose whether you want the background to be included (white) or transparent (no background).

Then click on “Download image”.
“Transparent” background is a good option for adding the visual to a social media card or infographic that may have a colored background.
“Just chart” is the best option if you want to add your own annotations or if the title and source are being added directly in the report and not in each chart.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>3,799</td>
<td>4,285</td>
</tr>
<tr>
<td>Armed forces</td>
<td>2,405</td>
<td>3,056</td>
</tr>
<tr>
<td>Clerical support workers</td>
<td>2,216</td>
<td>2,344</td>
</tr>
<tr>
<td>Overall</td>
<td>2,216</td>
<td>2,344</td>
</tr>
<tr>
<td>Skilled agricultural, forestry and fishery workers</td>
<td>1,781</td>
<td>2,099</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>1,781</td>
<td>2,099</td>
</tr>
<tr>
<td>Service and sales workers</td>
<td>1,373</td>
<td>1,462</td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>1,262</td>
<td>1,568</td>
</tr>
</tbody>
</table>

Created with Datawrapper
A note on publishing

➢ The visual does NOT need to be published in order to download a static image file.
➢ If you wish to create an interactive online version with a URL, you will first need to click Publish Now.
➢ When you publish a visual, the data will be visible to those who have the URL. Data cannot be downloaded, but the data values can be seen on the data labels and the tooltip if either are enabled.
PRACTICE EXERCISE 6:
Create a range plot in Datawrapper with data from the Training Dataset file
Exercise – create a range plot in Datawrapper

- Using the data from the "Exercise 6 DW Range Plot" worksheet in the "Training Dataset Day 2" file, create a range plot in Datawrapper.

- Tips to remember:
  - Select the data labels.
  - Make sure the category labels are visible.
  - Make sure the chart is sorted correctly.
  - Change the colors to the appropriate female and male colors. Don't use stereotypical colors.

- Time for exercise: 20 minutes.
STRENGTHENING GENDER STATISTICS

DATA VISUALIZATION TRAINING MODULE 8: ANNOTATIONS
Contents

1. Recap of data visualization steps and breakdown
2. Using annotations to highlight key information
   a. Adding text boxes
   b. Incorporating icons
   c. Adding lines or shaded shapes to highlight information
3. Exercise to annotate visuals created in Day 1 and Day 2.
1. Recap of data visualization steps and breakdown
Recap: steps for creating a data visualization

- Upload data
- Explore/try visualization options
- Adjust data structure
- Design, format, and finalize visualization
- Publish and/or download visualization

Optional: add additional annotations as needed
Recap: data visualization breakdown

Three-fourths of creating a visualization is cleaning and formatting the data in the proper way to:

- match the required data structure, inputs, and features of the intended visualization type
- highlight the right message and insights
- enable easy formatting and annotation

75% Properly cleaning, transposing & preparing the data

25% Formatting & annotating the visual
Recap: data visualization breakdown
2. Using annotations to highlight key insights
ADDING LABELS OR TEXT
Adding annotations through text boxes

➢ Under the "Insert" tab in Excel or PowerPoint, click on "Text Box" within the "Text" section and position the cursor or text box where you would like your annotation. Type the information (data point value, category label, etc.) Repeat for each bar/label.
➢ Go to Sheet named "Demo 9 Text Boxes" with the bar chart we created from Module 3.
➢ Add data labels and/or category labels to the chart.
  • Add them directly in Excel or copy and paste the chart to PowerPoint and add the annotations in PowerPoint.
INCORPORATING ICONS
Using icons for highlighting statistics

- Icons can be used to create pictograms using the same repeated icon and coloring/shading to show a proportion like 2 in 5 women representing 40% or a ratio "For every one man there are 5 women who...".

20% of the female population in Tanzania (15+)

is widowed, divorced, or separated
Using icons for highlighting statistics

Infographic-like visuals with just images and numbers can be developed for incorporation in a report, an executive summary, a social media card, etc.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any fertilizer</td>
<td>37%</td>
</tr>
<tr>
<td>Inorganic fertilizer</td>
<td>20%</td>
</tr>
<tr>
<td>Improved Seeds</td>
<td>47%</td>
</tr>
<tr>
<td>Organic fertilizer</td>
<td>23%</td>
</tr>
<tr>
<td>Pesticides/Insecticides</td>
<td>20%</td>
</tr>
<tr>
<td>Extension Services</td>
<td>2%</td>
</tr>
</tbody>
</table>

**Reasons for not attending school**

- 39.2% Satisfied
- 23.3% To work/look for work
- 14.5% Marriage
- 9.1% Financial Constraints
- 7.2% Refusal
- 2.7% Pregnancy/Childcare/Distance
- 1.6% Illness/Sickness

- On average, the staff turnover rate is at 20%
- 7 in 100 jobs are vacant in the current labour market
- One third of the current vacant jobs are for expansion purposes

![NPS 2014/15](green) NPS 2020/21
Using icons for highlighting statistics

Icons can enhance simple visuals like bar charts or pie charts that have very few data points (two or three data points).
Icons can enhance simple visuals like bar charts or pie charts that have very few data points (two or three data points).
Using icons for highlighting statistics

- Icons can enhance simple visuals like bar charts or pie charts that have very few data points (two or three data points).
Using icons for legends

Icons can be added to color legends instead of just having colored or patterned boxes for the legend.

---

**Unemployment rates in low-exposure areas in Bangladesh's Cox Bazar district grew double for men and almost eight-times for women during COVID-19 compared to pre-COVID rates.**

![Unemployment rate graph]


---

In Greece, the proportion of female workers among the workers laid off during COVID-19 was more than double the proportion of workers that are female at the surveyed enterprises.

![Proportion of workers and female workers graph]


---

**Stunting and Severe stunting**

![Stunting chart]

---

Girls
Boys
Inserting an icon

➢ To add an icon to a chart, go to the “Insert” tab in Excel or PowerPoint, click on “Icons” within the “Illustration” section. A box will pop up of icons to choose from and you can enter a key word or choose a topic to search for an appropriate icon. Once you have found an icon, click the “Insert” button.
Searching for and choosing an icon

➢ On the left I have selected the “Landscape” topic button. On the right I typed in “Rural” in the search bar
   • Sometimes you have to use synonyms to find more options perhaps “Farm”, “Land”, “Agriculture”, etc. Or instead of typing “Urban” typing “Building”, “City”, “Road”, “Street”, etc.
➢ There are usually two versions of the same icon – one outlined and one filled.
   • The choice should depend on whether the icon is visible enough with just the outline, or it needs to be more prominent.
Searching for and choosing an icon

➢ When creating a pictogram, typically both outlined and filled icons are selected to demonstrate the proportion out of the whole.
  • For example, 2 filled and 3 outlined versions of the same icon should signify 2 out of 5 or 40%.
  • Alternatively, you could use only filled icons and use color to represent the statistic: one color for the proportion and the other for the remaining icons that make up the whole.
  • To change the color of an icon, click on the icon and the “Format Shape” panel should pop up on the right. Click on the paint bucket icon in the "Color" section and choose a color from the dropdown.
Ensuring the icon is visible

➢ Make sure the icon is in front of the bar and not hiding behind the bar. This can be done in several ways:
  • In the "Home" tab, go to the "Drawing" section and click "Arrange" and choose one of the options to move the icon or graphic forward or backward.
  • Click on the icon and go to the "Shape Format" tab and go to "Bring Forward" and click on "Bring Forward". If you don't see the icon on the screen, try clicking on the chart, go to the "Shape Format" tab and go to "Send Backward" and click "Send to Back".
Go to Sheet named "Demo 10 icons" in the Excel file "Training Dataset Day 2" and copy/paste the chart to Powerpoint.

- Create a color legend using the female and male icons. Female is hex code #C481E and RGB (196, 131, 30) and male is hex code #440E5F and RGB(68, 14, 95).
- Add an icon in front of the bars signifying rural as well as a source.

Proportion of women and men with no schooling in rural areas (%)
Create a pictogram in PowerPoint using the female icon to represent the statistic "4 in 5 women have a national identification card".

Add the text 4 in 5 women have a national identification card with the numbers larger than the text.

Copy the finished product into the Sheet named "Demo 10 icons" in the Excel file "Training Dataset Day 2".
ADDING LINES OR SHAPES FOR HIGHLIGHTING INSIGHTS
Adding arrows for easier interpretation

- Arrows when combined with text can provide more context about how the reader should interpret the scale or units of the chart.
- Arrows can also be used with a data callout for percentage point difference to allow readers to easily quantify the gaps that they are seeing visually.
Adding lines for easier interpretation

- Solid or dotted lines can be added as reference or baseline values. They are often added with text that allows the reader to better interpret data values above and below the lines.
- Lines for gender parity are always added to signify equal male and female values in order to help the readers understand how far from gender equality each observation is.
Adding lines and shading to highlight insights

- Dotted lines, arrows, and shaded areas can help draw the readers’ eyes to gaps, outliers, ranges of years, values, etc., and other insights more easily.

Rural men are 5 times less likely than urban men to have any schooling.

Rural men and women have the lowest education rates, but the urban-rural education gap is bigger for men.

Rural men are 5 times less likely than urban men to have any schooling.
In Country A, **1 in 5 women have no education** and the same value increases to **1 in 4 women in rural areas**.

Far fewer women than men are gaining life skills from school that would improve their chances of having decent, safe, paid work. While education rates are lower in rural areas relative to urban areas for both men and women, the urban-rural discrepancy is much bigger for men than for women.

The **gender gap is biggest in urban areas** where women are 2.6 times less likely than men to have no education.

**Rural men are 5 times less likely than urban men to have no education.**

Percentage of the population age 15+ with no education in Country A

Be consistent in the use of terminology when presenting data. In this case, “no education” should not be substituted with “no schooling”, “individuals without education” or any other terms.
Adding lines and shading to highlight insights

➢ Dotted lines, arrows, and shaded areas can help draw the readers’ eyes to gaps, outliers, ranges of years, values, etc., and other insights more easily.
Arrows, dotted lines, and shading

- Combining the arrows with shaded areas can further highlight gaps.
- It is not always necessary to add a data callout in the annotation - the value of the gap can be mentioned in the supplementary text.
Adding shapes as annotations

➢ To add a shape, go to the “Insert” tab and click “Shapes” within the “Illustrations” section. Select the shape you are looking for then drag the mouse on the screen to create the shape. Adjust the shape further by clicking the dots and moving the sides.
➢ For each shape, there are options for the outline to make it thicker, dotted, change the color, or add arrows only for lines.
Adding shapes as annotations

➢ For shapes like a rectangle, you can choose a fill color or leave it unfilled so that there is an outline. You can also make the fill shape semi-transparent in order to highlight a grey area or a specific light color such as yellow or orange or red.
➢ To make it transparent, click on the shape and go to the "Shape Format" panel on the right. Adjust the transparency with the slider in "Transparency". You can also remove the outline from the transparent shape.
Once you’ve made a shape, if you need to use it somewhere else, you can copy and paste it and position it. You can copy multiple shapes at once by holding down the CTRL key and selecting multiple shapes while holding down the CTRL key.

Here we are copying the arrows, lines, and shaded rectangle for the middle visual.
➢ Go to the Sheet named "Demo 11 Lines Shapes" in the Excel file "Training Dataset Day 2 and copy/paste the chart to Powerpoint.
➢ Add a double-ended arrow to highlight the gap with an annotation like the percentage point difference or some other text that gives context to the gap.

Percentage without access to internet (%)
➢ Add a shaded rectangle or outline of a rectangle to the range plot from the demo in Module 7.
➢ Copy the finished product into the Sheet named "Demo 11 Lines Shapes" in the Excel file "Training Dataset Day 2".

**Gender Wage Gap**

Average hourly earnings in local currency

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service and sales workers</td>
<td>1K</td>
<td>3K</td>
</tr>
<tr>
<td>Clerical support workers</td>
<td>982</td>
<td>1K</td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>1K</td>
<td>2K</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>925</td>
<td>2K</td>
</tr>
<tr>
<td>Managers</td>
<td>413</td>
<td>4K</td>
</tr>
<tr>
<td>Craft and related trades workers</td>
<td>252</td>
<td>753</td>
</tr>
<tr>
<td>Skilled agricultural, forestry and fishery workers</td>
<td>218</td>
<td>753</td>
</tr>
<tr>
<td>Plant and machine operators, and assemblers</td>
<td>74</td>
<td>3K</td>
</tr>
</tbody>
</table>

Source: Country A Labor Force Survey 2022 • Get the data

**Gender Wage Gap**

Average hourly earnings in local currency

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service and sales workers</td>
<td>1K</td>
<td>3K</td>
</tr>
<tr>
<td>Clerical support workers</td>
<td>982</td>
<td>1K</td>
</tr>
<tr>
<td>Technicians and associate professionals</td>
<td>1K</td>
<td>2K</td>
</tr>
<tr>
<td>Elementary occupations</td>
<td>925</td>
<td>2K</td>
</tr>
<tr>
<td>Managers</td>
<td>413</td>
<td>4K</td>
</tr>
<tr>
<td>Craft and related trades workers</td>
<td>252</td>
<td>753</td>
</tr>
<tr>
<td>Skilled agricultural, forestry and fishery workers</td>
<td>218</td>
<td>753</td>
</tr>
<tr>
<td>Plant and machine operators, and assemblers</td>
<td>74</td>
<td>3K</td>
</tr>
</tbody>
</table>

Source: Country A Labor Force Survey 2022 • Get the data
➢ Go to Sheet named "Demo 9 Text" using the chart created in Module 5.
➢ Transform the "Overall" data point in the purple column chart to a line. Add a notation that it is the overall value.
PRACTICE EXERCISES:
Exercise – annotate demo data visualizations

➢ Take any of the visuals we have created in the demonstrations or tasks and polish the visualization
   • Add annotations (labels, arrows, shading, etc.), a title (potentially a subtitle), and source.

➢ Tips to remember:
   • Annotations should use an accent color that is complementary to the visual, so it doesn’t introduce too many colors new colors into the visual.
   • Annotations and highlights should contrast well with the original visual.
   • Annotations and highlights should help the reader, not distract and make the visual cluttered.

➢ Time for exercise: 20 minutes.