Graphing Data

Workshop on data analysis and report writing for civil registration based vital statistics

Nadi, Fiji
30 January – 03 February 2023
Outline of presentation

- Why use charts?
- What makes a good graph/chart?
- Types of graphs
  - Bar/column charts, line graph, pie chart, scatter plot
- Advantages and disadvantages
- Choosing a graph
Graphing your data
Why use charts?

- **Comparison**: how much? Which item is bigger or smaller?
- **Changes over time**: how does a variable evolve?
- **Frequency distribution**: how are items distributed? What are the differences?
- **Correlation**
- **Relative** share of a whole
Checklist for a good chart

A good graph:
• grab’s attention
• Presents information clearly, simply, accurately
• Does not mislead
• Facilitates comparisons and highlights trends
• Illustrates messages, themes or stories in text
Bar charts

Female ambassadors in 2006

- Germany
- Finland
- Latvia
- Estonia
- Slovenia
- Lithuania
- Netherlands
- Georgia
- Cyprus
- Croatia
- Israel
- Spain
- Luxembourg
- Italy
- United Kingdom
- Ukraine

Source: UNECE Statistical Database
Figure 1: Deaths among young people aged 15–24, by age and sex, 2019

Note: These data have not been adjusted for Victorian additional death registrations in 2019. See Technical notes for more details. Chart: AIHW.
Stacked column graphs

Gender split of teachers in Ireland, 2005-2006

%

Type of educational institutions

Primary

Secondary

Tertiary

Source: UNECE Statistical Database
Line graphs

Figure 1. Life expectancy at birth, by sex: United States, 2000–2021

NOTES: Estimates are based on provisional data for 2021. Provisional data are subject to change as additional data are received. Estimates for 2000–2020 are based on final data.

Line graphs

Figure 2.2: Male Age-Specific Mortality Rates (deaths per 1000 people), by 3-year period, 2013–2018.

Figure 2.3: Female Age-Specific Mortality Rates (deaths per 1000 people), by 3-year period, 2013–2018.

Figure 1: Leading causes of infant death, 2017

- Fetus & newborn affected by maternal complications of pregnancy: 8%
- Fetus & newborn affected by complications of placenta, cord and membranes: 9%
- Disorders of short gestation and low birthweight: 8%
- Other perinatal conditions: 29%
- Congenital malformations circulatory system: 16%
- Other congenital anomalies: 7%
- Sudden infant death syndrome (SIDS): 3%
- Signs, symptoms and abnormal findings: All except SIDS: 7%
- All other causes: 14%

Note: Due to rounding the proportions do not sum to 100.
Chart: AIHW. Source: Analysis of AIHW National Mortality Database.
Scatter plot

Under-five mortality and underweight prevalence in Sub-Saharan African countries, 2003

<table>
<thead>
<tr>
<th>Chart type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar chart (vertical)</td>
<td>Simple and clear Works for categories and time series</td>
<td>Not good for long time series Small space for long names</td>
</tr>
<tr>
<td>Bar chart (horizontal)</td>
<td>Good for large number of categories Works for long names</td>
<td>Not appropriate for time series</td>
</tr>
<tr>
<td>Line chart</td>
<td>Simple and clear Best for time series</td>
<td>More than three lines gets confusing</td>
</tr>
<tr>
<td>Pie chart</td>
<td>Shows distribution of one variable</td>
<td>Not good for making comparisons Too many ‘slices’ gets confusing</td>
</tr>
<tr>
<td>Scatter plots</td>
<td>Shows relationships between variables</td>
<td>Can be difficult to interpret</td>
</tr>
</tbody>
</table>
Components of a good graph

INFORMATION BOX 4: Parts of Graphs

(a) title

Tuvalu: Number of Households by Island, 1991

(b) y axis title

Number of households

(c) scale starts at zero

(d) source

Source: Tuvalu 1991 Population and Housing Census

(e) x axis title

(f) x axis data is labelled
Which type should I use?
Sort your data

**BAD EXAMPLE**

Adolescent fertility rate, 2006

<table>
<thead>
<tr>
<th>Country</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>10</td>
</tr>
<tr>
<td>Georgia</td>
<td>35</td>
</tr>
<tr>
<td>Greece</td>
<td>15</td>
</tr>
<tr>
<td>Hungary</td>
<td>25</td>
</tr>
<tr>
<td>Romania</td>
<td>30</td>
</tr>
<tr>
<td>Serbia</td>
<td>20</td>
</tr>
</tbody>
</table>

**GOOD EXAMPLE**

Adolescent fertility rate, 2006

<table>
<thead>
<tr>
<th>Country</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greece</td>
<td>15</td>
</tr>
<tr>
<td>Albania</td>
<td>10</td>
</tr>
<tr>
<td>Hungary</td>
<td>25</td>
</tr>
<tr>
<td>Serbia</td>
<td>20</td>
</tr>
<tr>
<td>Romania</td>
<td>30</td>
</tr>
<tr>
<td>Georgia</td>
<td>35</td>
</tr>
</tbody>
</table>

The data are presented by alphabetical order of countries. The values are very difficult to compare. Attention is on the first and last values, which have no specific relevance.

The data are presented in order from smallest to largest values. It is easy to compare them. Attention is focused on the minimum and maximum values of the dataset.
Re-setting axis scales – caution!
All components have maximum impact. The result is a busy chart, difficult to read, even though it shows only three values.

This chart is much easier to read. Minimal use of support components ensures that data take centre stage.
Keep it simple

BAD EXAMPLE

Population aged 18+ by legal marital status in Iceland, 2004

GOOD EXAMPLE

Population aged 18+ by legal marital status in Iceland, 2004
Keep it simple

BAD EXAMPLE

Adolescent fertility rate, 2006

GOOD EXAMPLE

Adolescent fertility rate, 2006
LOAD OF INFORMATION-

Caution!

**Why** are they doing it?!
Any QUESTIONS?