Introduction to Data Analysis

Greg Keeble
UNESCO Institute for Statistics
Overview

- Objective & Audience
- Forms of presentation
- Tables dimensions
- Components of a table
- Layout of a Table

- Presenting Data
- Purpose of graphs & charts
- Components of graphs & charts
- Types of graphs/charts
- Data presentation
Objective and Audience

• The type of statistics depends on the main objective of the presentation and the target audience

• Consider how best to present the data and indicators:
  – What am I trying to communicate?
  – Who are my audience?
  – What kind of presentation will be most effective?
  – What will help my audience to better understand the data?
Forms of presentation!

Communicate to your audience the meaning of the data using the summary statistics in an informative and interesting manner that is easy to understand:

• **Tables** are useful for presenting data and statistics in numeric form

• **Charts and graphs** may be used to highlight key patterns and trends in a graphical form

• **Descriptive text** can describe and summarize findings in verbal form
Dimensions of a summary table

• Summary tables are used to present counts of students, teachers and schools by a categorical variable.

• e.g. number of teachers by qualification
  number of students by grade.

• Summary tables can be:
  – simple one-dimensional table
    with one categorical variable
  – multi-dimensional table
    with two or more categorical variables.
One-dimensional Summary Table
The table shows the number and percentage of student enrollments by grade in high school.

<table>
<thead>
<tr>
<th>Grades</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 9</td>
<td>813</td>
<td>30.88%</td>
</tr>
<tr>
<td>Grade 10</td>
<td>704</td>
<td>26.74%</td>
</tr>
<tr>
<td>Grade 11</td>
<td>575</td>
<td>21.84%</td>
</tr>
<tr>
<td>Grade 12</td>
<td>541</td>
<td>20.55%</td>
</tr>
<tr>
<td>Total</td>
<td>2,633</td>
<td>100.00%</td>
</tr>
</tbody>
</table>
## Multi-dimensional Summary Table

### Student Enrollments by Elementary Grade by Gender

<table>
<thead>
<tr>
<th>Grades</th>
<th>Male No.</th>
<th>Male %</th>
<th>Female No.</th>
<th>Female %</th>
<th>Total No.</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>647</td>
<td>14.0%</td>
<td>555</td>
<td>13.0%</td>
<td>1,202</td>
<td>13.5%</td>
</tr>
<tr>
<td>Grade 2</td>
<td>684</td>
<td>14.8%</td>
<td>557</td>
<td>13.0%</td>
<td>1,241</td>
<td>13.9%</td>
</tr>
<tr>
<td>Grade 3</td>
<td>637</td>
<td>13.7%</td>
<td>609</td>
<td>14.2%</td>
<td>1,246</td>
<td>14.0%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>586</td>
<td>12.6%</td>
<td>565</td>
<td>13.2%</td>
<td>1,151</td>
<td>12.9%</td>
</tr>
<tr>
<td>Grade 5</td>
<td>608</td>
<td>13.1%</td>
<td>524</td>
<td>12.2%</td>
<td>1,132</td>
<td>12.7%</td>
</tr>
<tr>
<td>Grade 6</td>
<td>528</td>
<td>11.4%</td>
<td>550</td>
<td>12.8%</td>
<td>1,078</td>
<td>12.1%</td>
</tr>
<tr>
<td>Grade 7</td>
<td>478</td>
<td>10.3%</td>
<td>481</td>
<td>11.2%</td>
<td>959</td>
<td>10.8%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>469</td>
<td>10.1%</td>
<td>442</td>
<td>10.3%</td>
<td>911</td>
<td>10.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,637</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>4,283</strong></td>
<td><strong>100.0%</strong></td>
<td><strong>8,920</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>
Multi-dimensional tables: 

- can show detailed data patterns and complex relationships 
- can become very complicated if too many data values are presented 
- need to consider the ability of the reader to understand and interpret multi-dimensional tables 
- enable in-depth analysis of the patterns of school participation by presenting data about the distribution of students by grade, age and gender
# Multi-dimensional Summary Table

## Student Enrollments by Elementary Grade by Age and Gender

<table>
<thead>
<tr>
<th>Age</th>
<th>G 1 M</th>
<th>G 1 F</th>
<th>G 2 M</th>
<th>G 2 F</th>
<th>G 3 M</th>
<th>G 3 F</th>
<th>G 4 M</th>
<th>G 4 F</th>
<th>G 5 M</th>
<th>G 5 F</th>
<th>G 6 M</th>
<th>G 6 F</th>
<th>G 7 M</th>
<th>G 7 F</th>
<th>G 8 M</th>
<th>G 8 F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=5</td>
<td>168</td>
<td>147</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>334</td>
</tr>
<tr>
<td>6</td>
<td>367</td>
<td>335</td>
<td>147</td>
<td>108</td>
<td>13</td>
<td>11</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>983</td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>47</td>
<td>374</td>
<td>315</td>
<td>129</td>
<td>118</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,075</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>15</td>
<td>99</td>
<td>92</td>
<td>295</td>
<td>290</td>
<td>92</td>
<td>111</td>
<td>19</td>
<td>11</td>
<td>5</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td>1,052</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>3</td>
<td>31</td>
<td>23</td>
<td>111</td>
<td>115</td>
<td>277</td>
<td>279</td>
<td>93</td>
<td>97</td>
<td>11</td>
<td>20</td>
<td>1</td>
<td></td>
<td>1,071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td>10</td>
<td>48</td>
<td>39</td>
<td>117</td>
<td>91</td>
<td>232</td>
<td>243</td>
<td>88</td>
<td>111</td>
<td>15</td>
<td>16</td>
<td>1</td>
<td>3</td>
<td>1,034</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>20</td>
<td>23</td>
<td>50</td>
<td>46</td>
<td>135</td>
<td>93</td>
<td>194</td>
<td>221</td>
<td>78</td>
<td>92</td>
<td>7</td>
<td>12</td>
<td></td>
<td>981</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>22</td>
<td>21</td>
<td>67</td>
<td>42</td>
<td>110</td>
<td>104</td>
<td>192</td>
<td>190</td>
<td>70</td>
<td>90</td>
<td></td>
<td>927</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>8</td>
<td>37</td>
<td>26</td>
<td>80</td>
<td>60</td>
<td>113</td>
<td>109</td>
<td>178</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>826</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>7</td>
<td>27</td>
<td>13</td>
<td>53</td>
<td>54</td>
<td>110</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td>377</td>
</tr>
<tr>
<td>15+</td>
<td>1</td>
<td>2</td>
<td>10</td>
<td>4</td>
<td>18</td>
<td>14</td>
<td>25</td>
<td>20</td>
<td>103</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>252</td>
</tr>
<tr>
<td>Total</td>
<td>647</td>
<td>555</td>
<td>684</td>
<td>556</td>
<td>632</td>
<td>608</td>
<td>586</td>
<td>564</td>
<td>608</td>
<td>524</td>
<td>528</td>
<td>550</td>
<td>478</td>
<td>481</td>
<td>469</td>
<td>442</td>
<td>8,912</td>
</tr>
</tbody>
</table>
Components of a table

A properly presented table should include:

– Title
– Headings
– Rows and columns
– Units of measurement
– Degree of accuracy
– Footnotes
– Source of data
### Example 4. The use of title, headings and units of measurement

**Pre-primary enrolment and gross enrolment ratios by region, 1999 and 2006**

<table>
<thead>
<tr>
<th></th>
<th>Total enrolment</th>
<th></th>
<th>Gross enrolment ratios</th>
<th></th>
<th>Change between 1999 and 2006</th>
<th>Change between 1999 and 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>School year ending in 1999</td>
<td>(millions)</td>
<td>School year ending in 2006</td>
<td>(millions)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>World</td>
<td>112</td>
<td>139</td>
<td>24</td>
<td>33</td>
<td>41</td>
<td>26</td>
</tr>
<tr>
<td>Developing countries</td>
<td>80</td>
<td>106</td>
<td>32</td>
<td>27</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>Developed countries</td>
<td>25</td>
<td>26</td>
<td>3</td>
<td>73</td>
<td>79</td>
<td>9</td>
</tr>
<tr>
<td>Countries in transition</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>46</td>
<td>62</td>
<td>36</td>
</tr>
</tbody>
</table>

**Note:** Change is computed using non-rounded figures.
**Source:** Annex, Statistical Table 3B.

Layout

A clear, well-structured layout makes it easier for the reader to interpret and understand the information in the table:

– Font style
– Ordering of rows and columns
– Numbers
– Consistent appearance
– Number the table
– Unnecessary distraction

Note: Complex tables that should logically appear together should be placed in the appendices.
Graphs and charts:

• represent and summarize statistical information in a visual manner to show patterns and trends in the data

• are useful for highlighting and presenting important information

• an ideal method for presenting statistical information to non-technical audiences
Purpose of charts and graphs

Charts and graphs are used:
• to visually represent information that cannot be easily read and interpreted from a table
• to show trends and changes in statistical data
• to make comparisons between two different sets of data or when making to make predictions and forecasts.
Advantages of charts and graphs

– easier to understand than a table of numbers
– highlight patterns and trends in the data
– makes comparisons and analysis easy
– representation of data using different types of graphs and charts
– allow for special designs e.g. age-pyramids and thematic maps
Components of charts and graphs

To understand and interpret the data represented in a graph or chart:-

Graphs and charts should have:

• a title
• axis labels
• labels for subgroups
• footnotes
• references to source data
Distribution of global public expenditure by region, 2004

- Lumphini
- Chakkrawat
- Thung Phaya Thai
- Ban Phan Thom
- Wat Sommanat
- Phra Khannong
- Pathum
- Khlong Tan

Note: Distribution is calculated using PPP US$. Source: UIS (2007, Figure 1, p. 11).
Characteristics of charts and graphs

– present a key message
– have a clear objective
– use an appropriate type of presentation
– have a simple and clear design
Types of graphs and charts

There are many different types of graphs and charts including:

- Pie chart
- Line graph
- Bar chart
- Area graph
- Scatter plot
- Maps
Process of creating a graph or chart

Steps to create a graph or chart:
1. Organize and present data in a table
2. Calculate percentages, ratios and indicators
3. Create graph or chart to illustrate the data
Practical Exercise

• Create Pivot Table from your EMIS dataset
• Calculate a percentage table
• Create a chart/graph from percentage table