How to display data badly

Karl W. Broman
Department of Biostatistics

http://www.biostat.jhsph.edu/~kbroman
Using Microsoft Excel to obscure your data and annoy your readers

Karl W Broman
Department of Biostatistics

http://www.biostat.jhsph.edu/~kbroman
Inspiration

This lecture was inspired by


Dr. Wainer was the first to elucidate the principles of the bad display of data.

The now widespread use of Microsoft Excel has resulted in remarkable advances in the field.
General principles

The aim of good data graphics:

- Display data accurately and clearly.

Some rules for displaying data badly:

- Display as little information as possible.
- Obscure what you do show (with chart junk).
- Use pseudo-3d and color gratuitously.
- Make a pie chart (preferably in color and 3d).
- Use a poorly chosen scale.
- Ignore sig figs.
Example 1
Example 1
Example 1
Example 1
Example 1
Example 1
Example 1
Example 1
### Example 2

**Distribution of genotypes**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>21%</td>
</tr>
<tr>
<td>AB</td>
<td>48%</td>
</tr>
<tr>
<td>BB</td>
<td>22%</td>
</tr>
<tr>
<td>missing</td>
<td>9%</td>
</tr>
</tbody>
</table>
Example 2

Distribution of genotypes

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>20</td>
</tr>
<tr>
<td>AB</td>
<td>50</td>
</tr>
<tr>
<td>BB</td>
<td>20</td>
</tr>
<tr>
<td>NA</td>
<td>5</td>
</tr>
</tbody>
</table>
Example 2

**Distribution of genotypes**

- AA
- AB (highest bar)
- BB
- NA (lowest bar)
Example 2

Distribution of genotypes

- 48% AB
- 22% BB
- 9% NA
- 21% other genotypes
Example 2

Distribution of genotypes

- AA: 21%
- AB: 48%
- BB: 22%
- NA: 9%
Example 2

Distribution of genotypes

- 48%
- 22%
- 9%
- 21%

Legend:
- AA
- AB
- BB
- NA
Example 3
Example 3
Example 3
Example 4
Example 4
Example 4

\[ y = 2.6981 + 1.652 \times \]

\[ \rho = 0.8567 \]
Example 5
Example 5
Example 5
Example 5
Example 5
Example 5
Example 6
Example 6
Example 6
### Example 7

<table>
<thead>
<tr>
<th>$N$</th>
<th>$b/c = 10.0$</th>
<th>$b/c = 10.0$</th>
<th>$b/c = 100.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^*$</td>
<td>$G$</td>
<td>$r^*$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.20</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.26</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.32</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.38</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.45</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.51</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.57</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0.63</td>
<td>4</td>
</tr>
</tbody>
</table>
### Example 7

<table>
<thead>
<tr>
<th>$N$</th>
<th>$b/c = 10.0$</th>
<th>$b/c = 10.0$</th>
<th>$b/c = 100.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^*$</td>
<td>$G$</td>
<td>$r^*$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.26333</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.32333</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.38267</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.446</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.50743</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.56743</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0.62948</td>
<td>4</td>
</tr>
</tbody>
</table>
Example 8

![Graph showing the proportion survived against log dose for Drug A, Drug B, and Drug C. The graph indicates that as the log dose increases, the proportion survived decreases for all three drugs, with Drug C showing the steepest decline.]
Example 8

Proportion survived

![Graph showing the proportion survived across different log doses for different drugs.]

- Drug A
- Drug B
- Drug C
Displaying data well

- Be accurate and clear.
- Let the data speak.
  - Show as much information as possible, taking care not to obscure the message.
- Science not sales.
  - Avoid unnecessary frills — esp. gratuitous 3d.
- In tables, every digit should be meaningful. Don’t drop ending 0’s.
Further reading