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.
Show the data

![Diagram showing response data for Treatment and Control groups.](image)
Show the data
Show the data
Show the data
Show the data
Show the data
Avoid pie charts
Avoid pie charts
Avoid pie charts
Avoid pie charts

via @MonaChalabi (bit.ly/pie_vs_barchart)
Avoid pie charts

via @MonaChalabi (bit.ly/pie_vs_barchart)
Avoid pie charts

Via @MonaChalabi (bit.ly/pie_vs_barchart)
Avoid pie charts

Via @MonaChalabi (bit.ly/pie_vs_barchart)
Consider logs
Consider logs
Consider logs
Consider logs
Consider logs
Consider logs
Consider logs
Consider logs
Take differences
Another “take logs” example

Broman et al., Am J Hum Genet 63:861-869, 1998, Fig. 1
Ease comparisons

(things to be compared should be adjacent)
Ease comparisons

(add a bit of color)
Which comparison is easiest?
Don’t distort the quantities
(value \propto \text{radius})

Wheat (17 Gbp)

Human (3.2 Gbp)

Arabidopsis (0.145 Gbp)
Don’t distort the quantities
(value $\propto$ area)

- Wheat (17 Gbp)
- Arabidopsis (0.145 Gbp)
- Human (3.2 Gbp)
Don’t use areas at all
(value $\propto$ length)
### Encoding data

<table>
<thead>
<tr>
<th>Quantities</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Position</td>
<td>• Shape</td>
</tr>
<tr>
<td>• Length</td>
<td>• Hue <em>(which color)</em></td>
</tr>
<tr>
<td>• Angle</td>
<td>• Texture</td>
</tr>
<tr>
<td>• Area</td>
<td>• Width</td>
</tr>
<tr>
<td>• Luminance <em>(light/dark)</em></td>
<td></td>
</tr>
<tr>
<td>• Chroma <em>(amount of color)</em></td>
<td></td>
</tr>
</tbody>
</table>
### Ease comparisons

#### (align things vertically)

<table>
<thead>
<tr>
<th>Height (in)</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Height (in)**

![Height Distribution](image)
Ease comparisons

(use common axes)
Use labels not legends
Don’t sort alphabetically

Health care spending (% GDP)

United States
United Kingdom
Turkey
Switzerland
Sweden
Spain
Russian Federation
Poland
Norway
Netherlands
Mexico
Korea, Rep.
Japan
Italy
Indonesia
India
Indonesia
Italy
Japan
Korea, Rep.
Mexico
Netherlands
Norway
Poland
Russian Federation
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States
Must you include 0?
A bad table

<table>
<thead>
<tr>
<th>$N$</th>
<th>$r^*$</th>
<th>$G$</th>
<th>$r^*$</th>
<th>$G$</th>
<th>$r^*$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>0.2</td>
<td>2</td>
<td>2.225</td>
<td>2</td>
<td>22.47499</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.26333</td>
<td>2</td>
<td>2.88833</td>
<td>2</td>
<td>29.13832</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.32333</td>
<td>3</td>
<td>3.54167</td>
<td>3</td>
<td>35.79166</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.38267</td>
<td>3</td>
<td>4.23767</td>
<td>3</td>
<td>42.78764</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.446</td>
<td>3</td>
<td>4.901</td>
<td>3</td>
<td>49.45097</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.50743</td>
<td>4</td>
<td>5.5765</td>
<td>4</td>
<td>56.33005</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.56743</td>
<td>4</td>
<td>6.26025</td>
<td>4</td>
<td>63.20129</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0.62948</td>
<td>4</td>
<td>6.92358</td>
<td>4</td>
<td>69.86462</td>
</tr>
<tr>
<td>$N$</td>
<td>$b/c = 10.0$</td>
<td>$b/c = 10.0$</td>
<td>$b/c = 100.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$r^*$</td>
<td>$G$</td>
<td>$r^*$</td>
<td>$G$</td>
<td>$r^*$</td>
<td>$G$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.20</td>
<td>2</td>
<td>2.2</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.26</td>
<td>2</td>
<td>2.9</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0.32</td>
<td>3</td>
<td>3.5</td>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.38</td>
<td>3</td>
<td>4.2</td>
<td>3</td>
<td>43</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.45</td>
<td>3</td>
<td>4.9</td>
<td>3</td>
<td>49</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0.51</td>
<td>4</td>
<td>5.6</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.57</td>
<td>4</td>
<td>6.3</td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0.63</td>
<td>4</td>
<td>6.9</td>
<td>4</td>
<td>70</td>
</tr>
</tbody>
</table>

Fewer digits
In the past two decades globally, noticeable increases in the absolute numbers of people with non-significant increase in low-income and non-significant increase in high-income countries and a 12% [95% CI 6%–17%] statistically non-significant increase in low-income and middle-income countries; table 1). Although DALYs lost (46% in high-income countries and 23% in low-income and middle-income countries), and 28% of DALYs lost in low-income and middle-income countries.

Further more, there was a significant 25% reduction in mortality rate (37% [31–41] in high-income countries and 20% [15–30] in low-income and middle-income countries). Stroke prevalence increased significantly by 27% (19–43) in high-income countries and 22% [18–32], respectively), and mortality-to-incidence ratio (23% [14–29] and 27% [14–38]).

Globally, for 1990–2010, we noted a 25% (13–33) significant increase in stroke incidence in people aged 20–64 years, and a 0·35 (0·32 in high-income countries and 0·36 in low-income countries) increase in people aged 75 years and older (table 1).

Table 1: Age-adjusted annual incidence and mortality rates (per 100 000 person-years), disability-adjusted life-years (DALYs) lost, prevalence (per 100 000 people), and mortality-to-incidence ratio (MIR) by age groups in high-income and low-income and middle-income countries, and globally in 1990, 2005, and 2010

<table>
<thead>
<tr>
<th>Year</th>
<th>Incidence</th>
<th>Prevalence</th>
<th>MIR</th>
<th>DALYs lost</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>6353868</td>
<td>13234062</td>
<td>12</td>
<td>63991864</td>
<td>2301435</td>
</tr>
<tr>
<td>2005</td>
<td>9288048</td>
<td>20187246</td>
<td>.</td>
<td>74855520</td>
<td>2734251</td>
</tr>
<tr>
<td>2010</td>
<td>10469624</td>
<td>23052804</td>
<td>.</td>
<td>73293552</td>
<td>2668499</td>
</tr>
</tbody>
</table>

* p value for the difference in age-adjusted rates between 1990 and 2010 only.

Table 1: Age-adjusted annual incidence and mortality rates (per 100 000 person-years), disability-adjusted life-years (DALYs) lost, prevalence (per 100 000 people), and mortality-to-incidence ratio (MIR) by age groups in high-income and low-income and middle-income countries, and globally in 1990, 2005, and 2010.
### 1990

<table>
<thead>
<tr>
<th>n</th>
<th>Rate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued from previous page)

**Globally**

<table>
<thead>
<tr>
<th>Category</th>
<th>n</th>
<th>Rate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;75 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incidence</td>
<td>6,353,868</td>
<td>159.22 (145.32–174.98)</td>
</tr>
<tr>
<td>Prevalence</td>
<td>13,234,062</td>
<td>324.26 (288.74–374.96)</td>
</tr>
<tr>
<td>MIR</td>
<td>..</td>
<td>0.359 (0.318–0.409)</td>
</tr>
<tr>
<td>DALYs lost</td>
<td>63,991,864</td>
<td>1543.96 (1452.03–1728.25)</td>
</tr>
<tr>
<td>Mortality</td>
<td>2,301,435</td>
<td>57.38 (54.12–64.27)</td>
</tr>
</tbody>
</table>

Feigen et al., Lancet 383:245-255, 2014, Table 1
What was wrong with that?

- **Way** too many digits.
- Numbers aren’t aligned.
- Numbers to be compared aren’t anywhere near each other.
- The interesting comparisons are horizontal rather than vertical.
- It would be much better as a multi-panel figure.
One last example
Scatterplots

Non-distracted

Total crashes vs Non-distracted crashes

Speeding

Total crashes vs Speeding crashes

Alcohol

Total crashes vs Alcohol crashes

Ave Ins Premium

Total crashes vs Ave Ins Premium

Ave Ins Loss

Total crashes vs Ave Ins Loss

Premium vs Loss

Ave Ins Loss vs Ave Ins Premium
Summary I

• Show the data

• Avoid chart junk

• Consider taking logs and/or differences

• Put the things to be compared next to each other

• Use color to set things apart, but consider color blind folks

• Use position rather than angle or area to represent quantities
• Align things vertically to ease comparisons

• Use common axis limits to ease comparisons

• Use labels rather than legends

• Sort on meaningful variables (not alphabetically)

• Must 0 be included in the axis limits?

• Use scatterplots to explore relationships
Inspirations

• Hadley Wickham  (slides at http://courses.had.co.nz)
• Naomi Robbins  (Creating more effective graphs)
• Howard Wainer
• Andrew Gelman
• Dan Carr
• Edward Tufte
Further reading


These slides: bit.ly/graphs2018