How to draw a diagram

Andrew Rice
why draw?
what to draw
how to draw it
Lots of this talk is based on content from:

Summarizing your data can be dangerous

<table>
<thead>
<tr>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
<td>10</td>
<td>7.46</td>
<td>8</td>
<td>6.58</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
<td>8</td>
<td>6.77</td>
<td>8</td>
<td>5.76</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
<td>13</td>
<td>12.74</td>
<td>8</td>
<td>7.71</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
<td>9</td>
<td>7.11</td>
<td>8</td>
<td>8.84</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
<td>11</td>
<td>7.81</td>
<td>8</td>
<td>8.47</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
<td>14</td>
<td>8.84</td>
<td>8</td>
<td>7.04</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
<td>6</td>
<td>6.08</td>
<td>8</td>
<td>5.25</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
<td>4</td>
<td>5.39</td>
<td>19</td>
<td>12.5</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.13</td>
<td>12</td>
<td>8.15</td>
<td>8</td>
<td>5.56</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
<td>7</td>
<td>6.42</td>
<td>8</td>
<td>7.91</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
<td>5</td>
<td>5.73</td>
<td>8</td>
<td>6.89</td>
</tr>
</tbody>
</table>

All sets: same mean for $x$ and $y$, same cross-correlation, same linear regression line

Graphics reveal the details of your data
Graphics reveal the details of your data.

The outlier is not visible in either of the single variable distributions.
Summarize many dimensions of data

Napoleon's Army: size, route and outside temperature against time
3D effects distort your data

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Fuel Economy Standards for Autos
Set by Congress and supplemented by the Transportation Department. In miles per gallon.


3D effects distort your data

Required Fuel Economy Standards

Year

Miles per Gallon

Consider the “Lie Factor”

Lie Factor = \frac{\text{perceived size}}{\text{actual size}}

Area of a circle: \text{perceived} = \text{actual}^{0.8\pm0.3}
Consider the “Lie Factor”

HOW ONE LITRE OF PETROL BREAKS DOWN

DUTY 47.1p
PRODUCT 23.2p
VAT 13.4p
RETAILER / DELIVERY 6.3p
(TOTAL: 90P)

SOURCE: AA Sep 2005

BBC News
Consider the “Lie Factor”

Duty is actually more than 50% of the total cost

- DUTY 47.1p
- PRODUCT 23.2p
- VAT 13.4p
- RETAILER/DELIVERY 6.3p
Consider the data density

Some datasets are better displayed in a table

Original graph: Executive Office of the President, Office of Management and Budget, Social Indicators, 1973

Use vector formats where possible

**Vector**
- Graphs and Diagrams
  - SVG, EPS, PDF, WMF

**Raster**
- Lossless
  - Screen Only
    - BMP, PNG
- Lossy
  - Photographs
    - JPG
Inkscape is a powerful tool for vector graphics

- Open-source
- Windows and Linux versions
Simple diagrams with Inkscape

\[ \delta_i = \frac{d_o - d_i}{2r} \]

(a) Tangential Distance

\[ d_i + \frac{d_o - d_i}{2r} \]

\[ \delta_i = (d_i + \frac{d_o - d_i}{2r}) \sin \frac{\pi}{s} \]

(b) Radial Distance

Figure 4.10: Tangential and radial size calculation

(Greek letters in inkscape are available as unicode characters)
Use SVG as an intermediate format

• Inkscape uses SVG an XML-based vector graphics format

```
<?xml version="1.0"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN"
  "http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">

<svg version="1.1" xmlns="http://www.w3.org/2000/svg">
  <circle cx="100" cy="50" r="40" stroke="black" stroke-width="2" fill="red"/>
</svg>
```

Lots of programs now support SVG import and export

You can easily generate your own SVG files and inkscape will render them
Generate all graphics consistently

- Multiple types of diagram
  - line widths and styles
  - font
  - colour scheme
- Draw at document size
- Keep raw data and automate generation
Gnuplot will generate graphs and output to SVG format

set terminal svg size 400 400 fname 'Times New Roman' fsize 12
set output "fuel.svg"
set border 3
set xtics nomirror
set ytics nomirror
set ylabel "Miles per Gallon"
set xlabel "Year"
set title "Required Fuel Economy Standards"
unset key
plot '-' with linespoints pointsize 1 pointtype 7
1978 18
1979 19
1980 20
1981 22
1982 24
1983 26
1984 27
1985 27.5
e
Graphs from Gnuplot can then be rendered in Inkscape.

Figure 4.24: Real-world and simulated location error across processing pipelines.
Complex graphs can be drawn by your own programs

SVG is a simple text format which you can easily generate programmatically
Finish

- Appropriate use of graphics makes your work much more accessible
- Lots of tools exist and can produce high quality output
- Don't just stick with scatter plots if a better presentation is possible
The only good pie chart

Percentage of the graph which resembles Pac-man

- Resembles Pac-man
- Does not resemble Pac-man