

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:

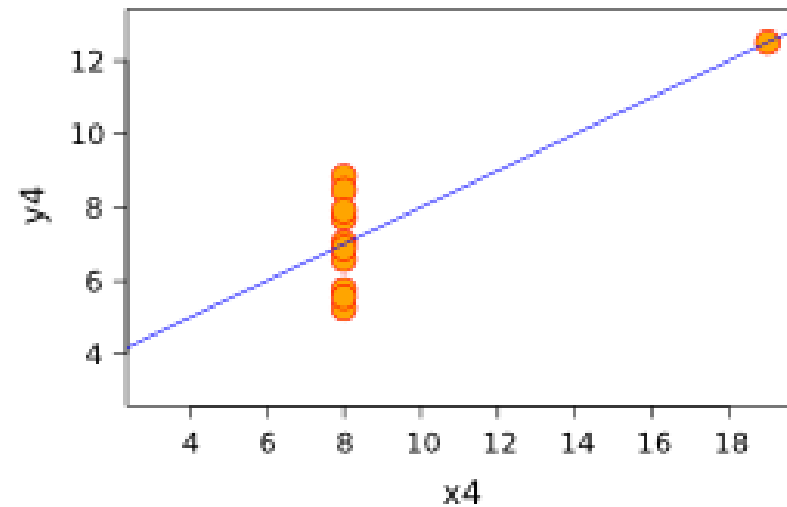
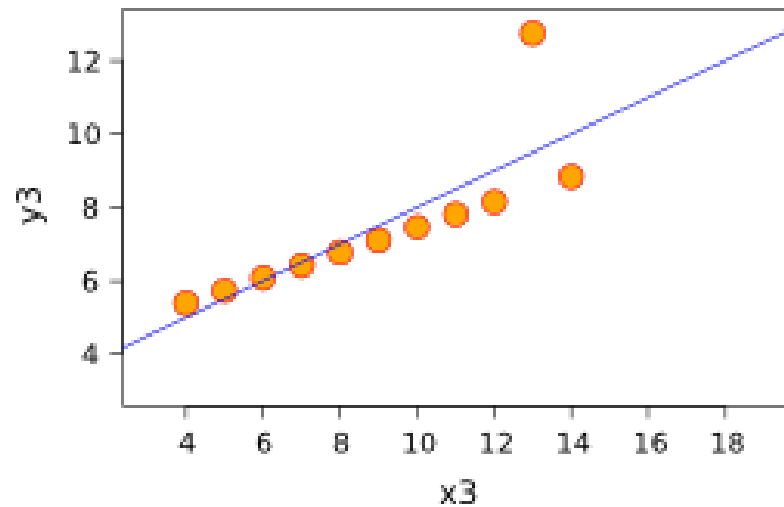
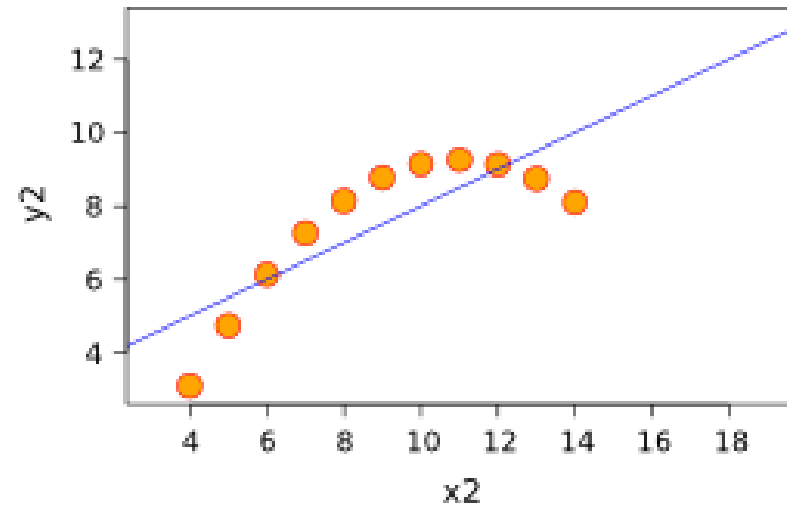
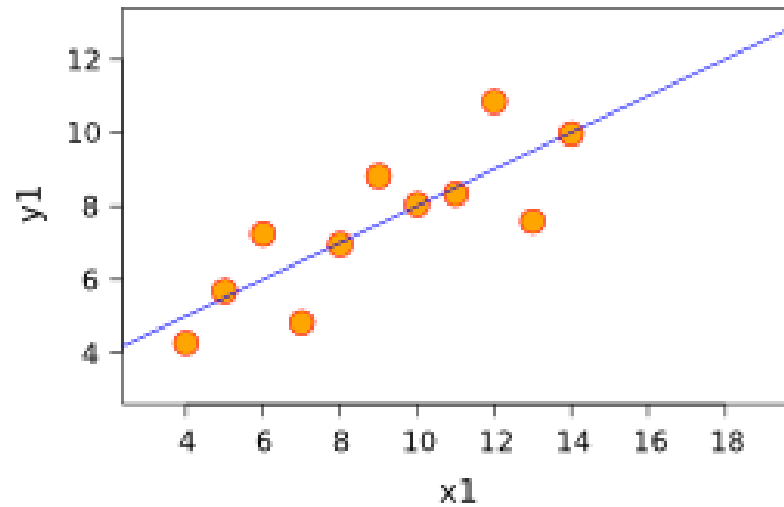
Edward Tufte, “The Visual Display of
Quantitative Information”, Graphics Press
1983

Summarizing your data can be dangerous

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

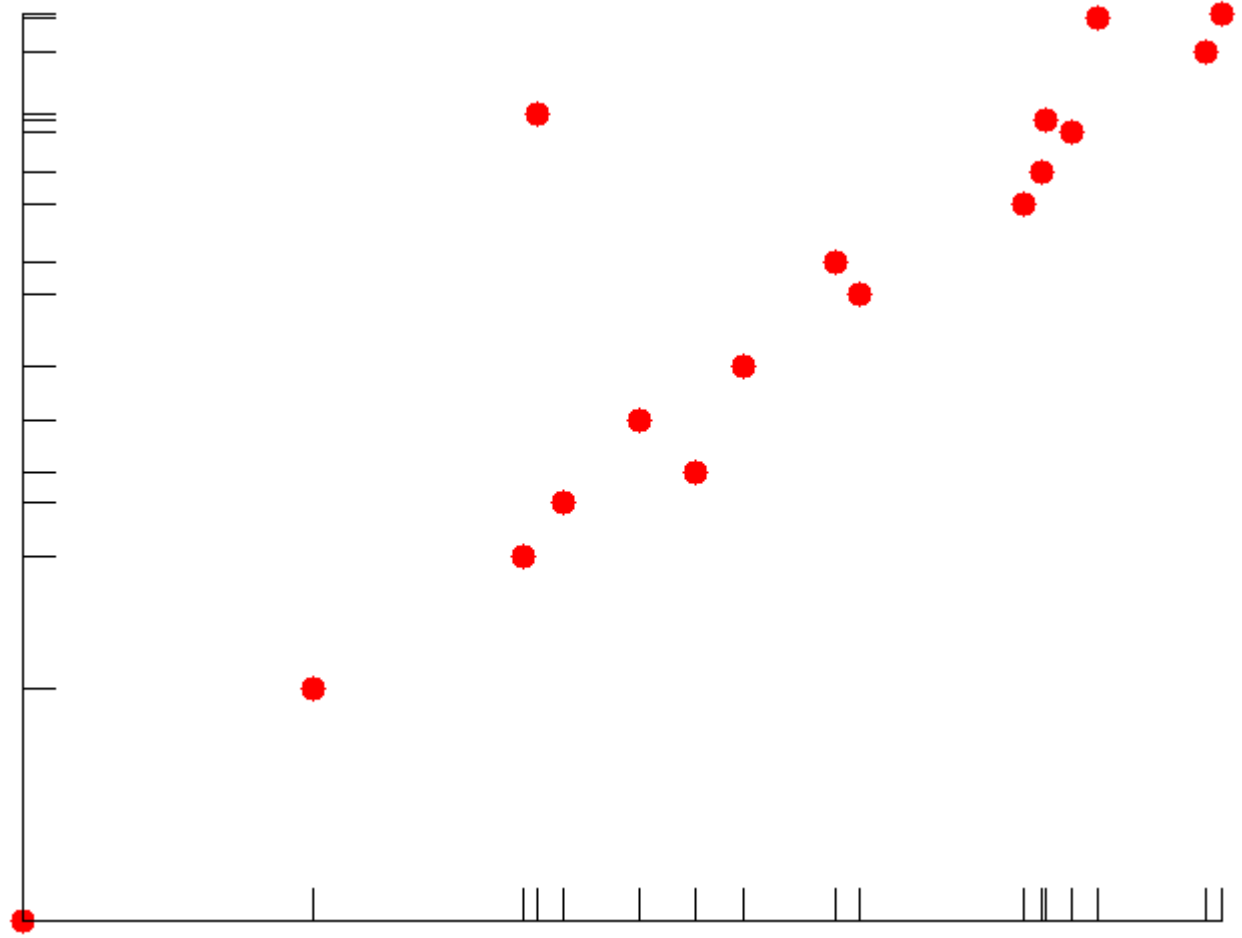
All sets: same mean for x and y, same cross-correlation, same linear regression line

Graphics reveal the details of your data

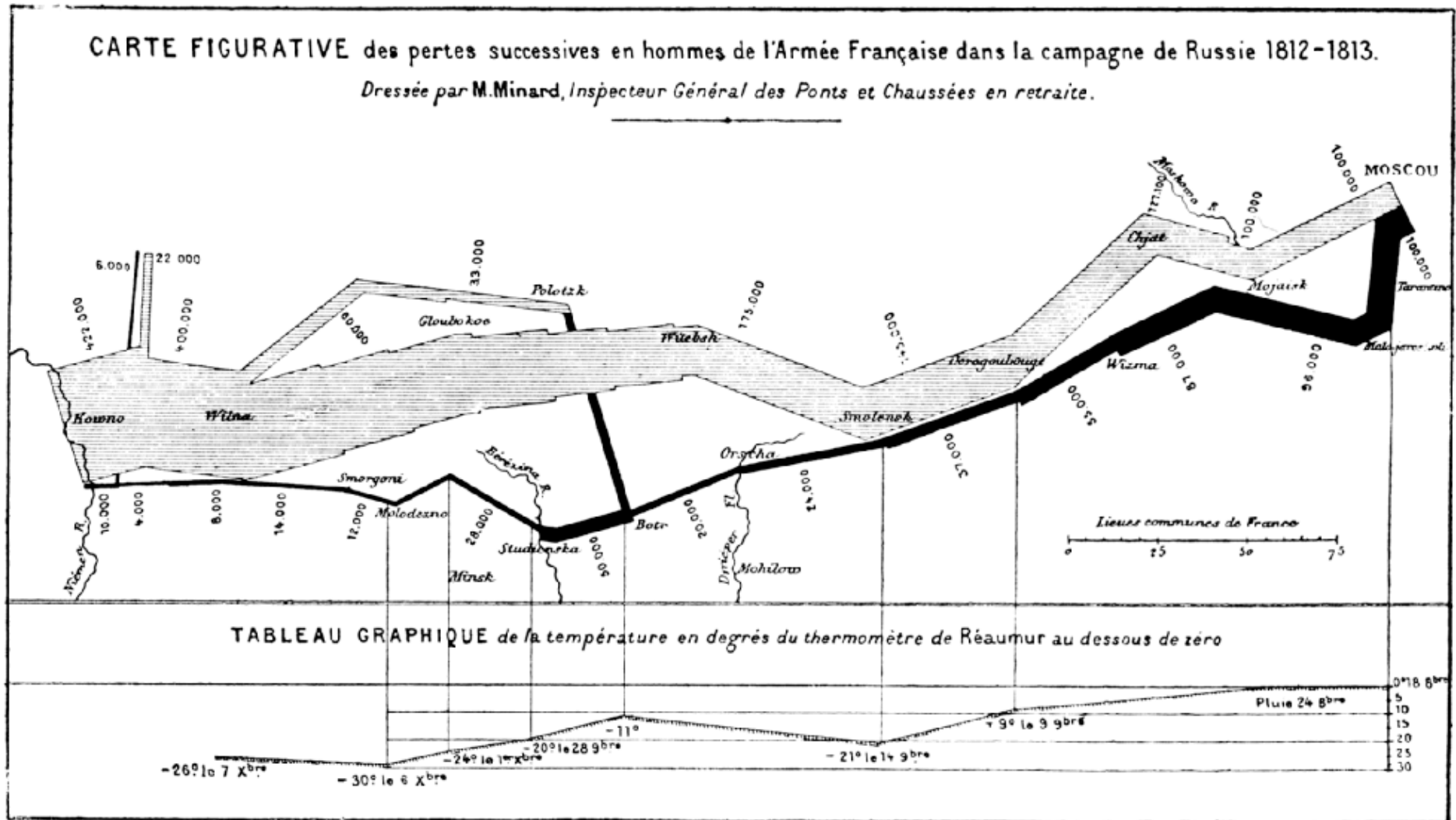


Graphics reveal the details of your data

The outlier is not visible in either of the single variable distributions

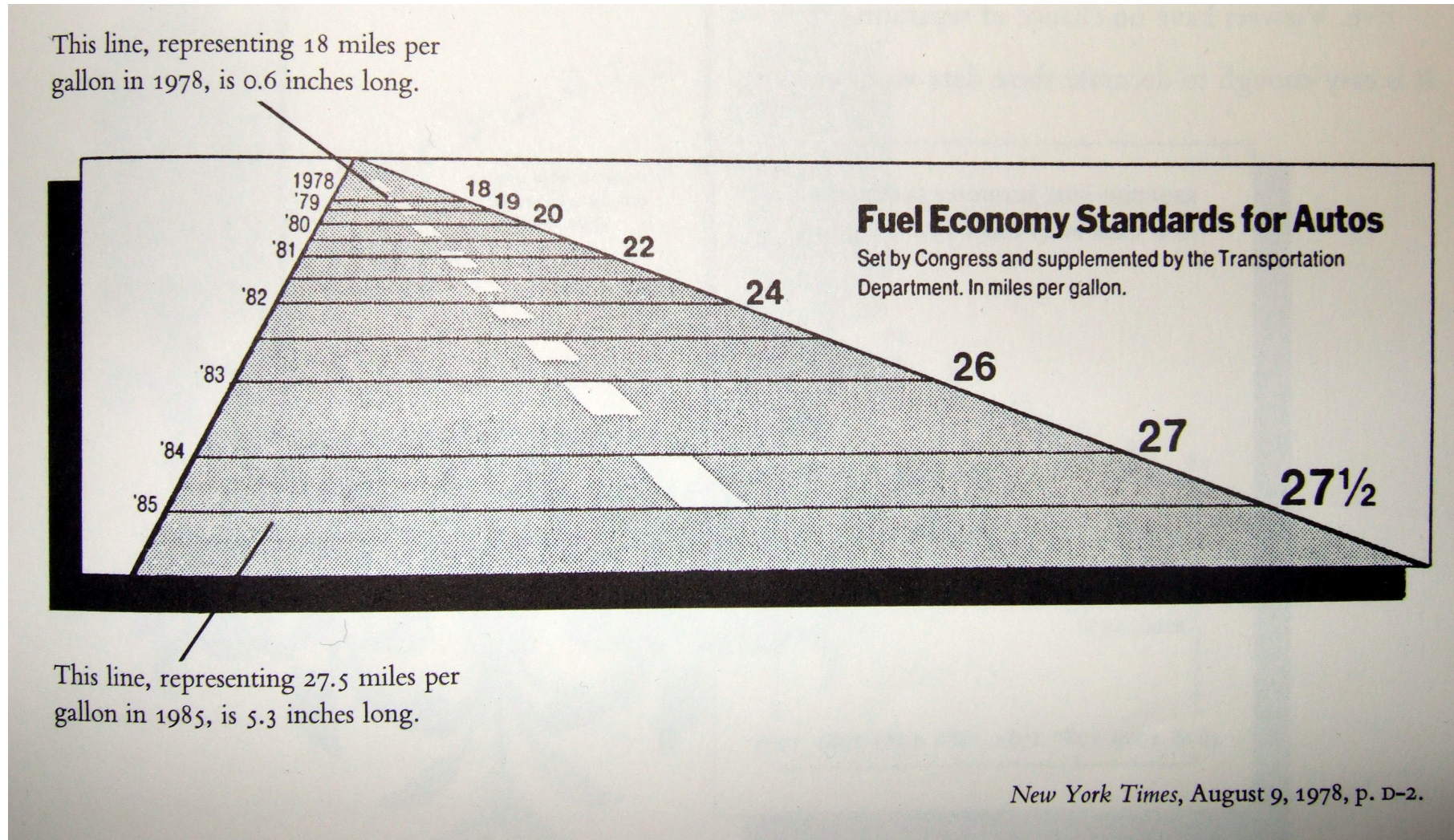


Summarize many dimensions of data



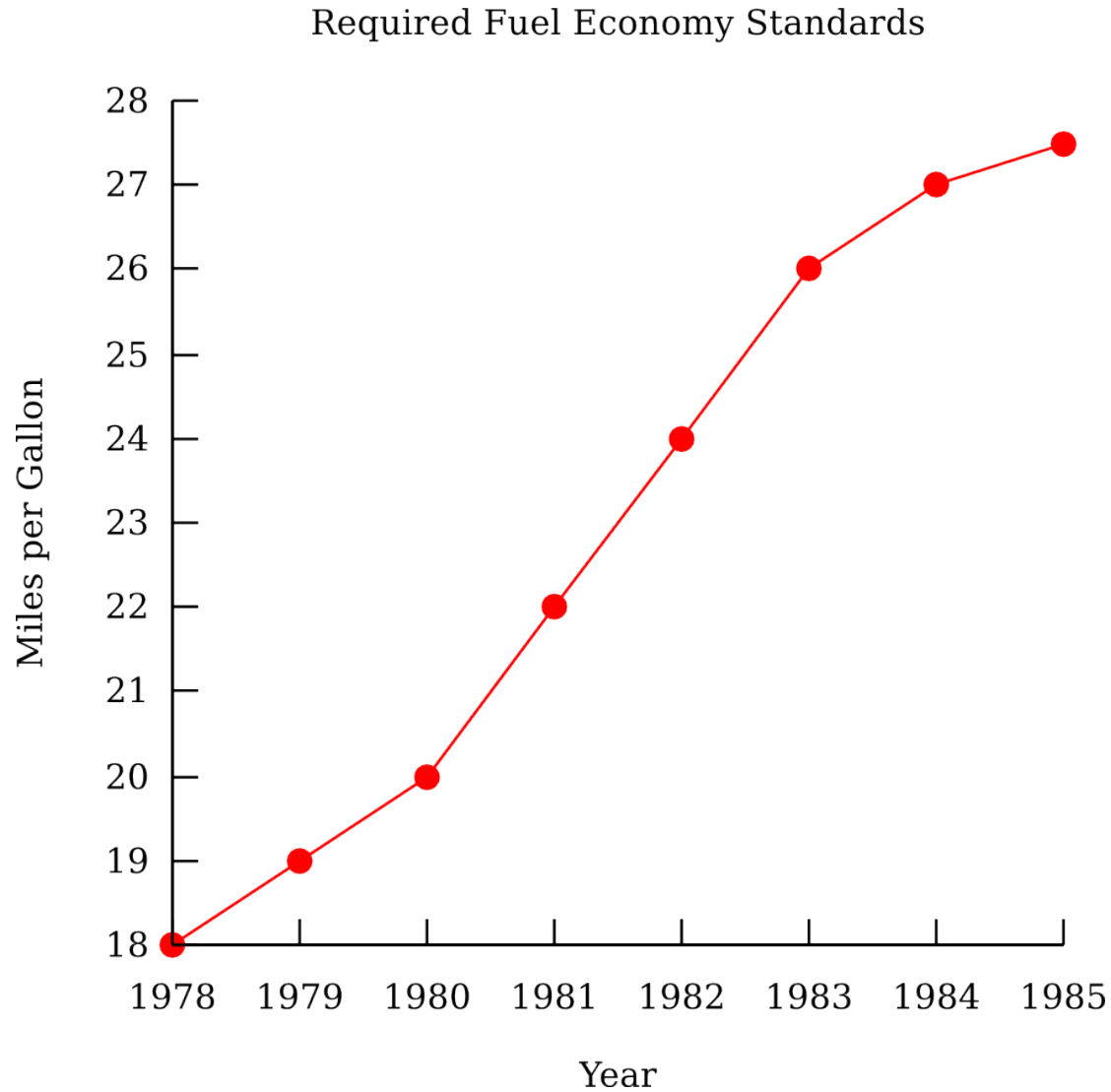
Napoleon's Army: size, route and outside temperature against time

3D effects distort your data



Edward Tufte, "The Visual Display of Quantitative Information",
Graphics Press 1983

3D effects distort your data

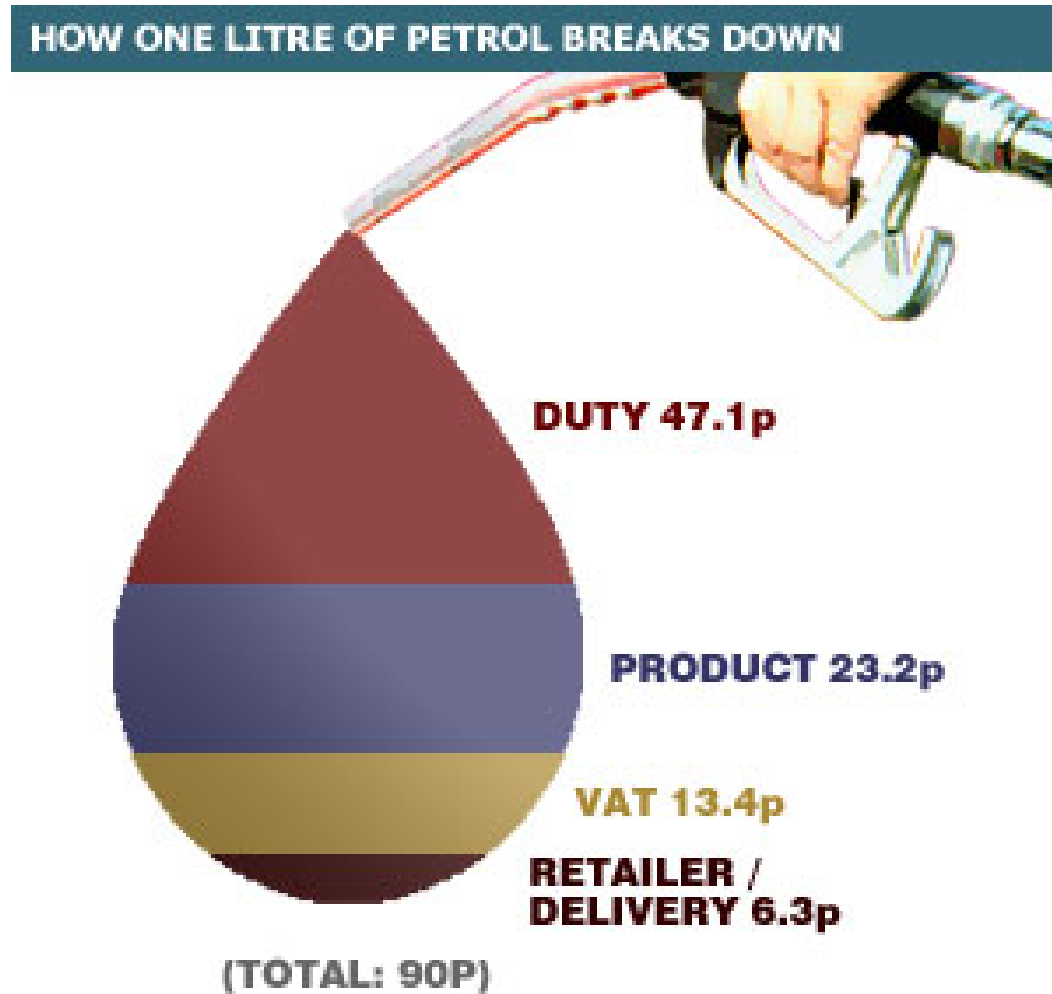


Consider the “Lie Factor”

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

Area of a circle: perceived = actual^(0.8±0.3)

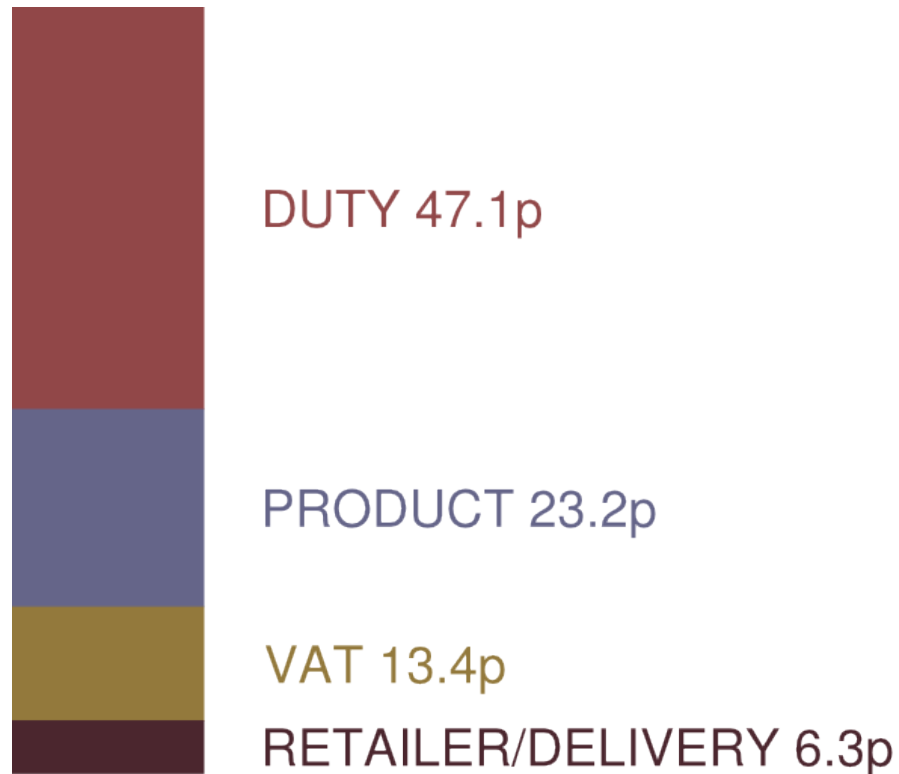
Consider the “Lie Factor”



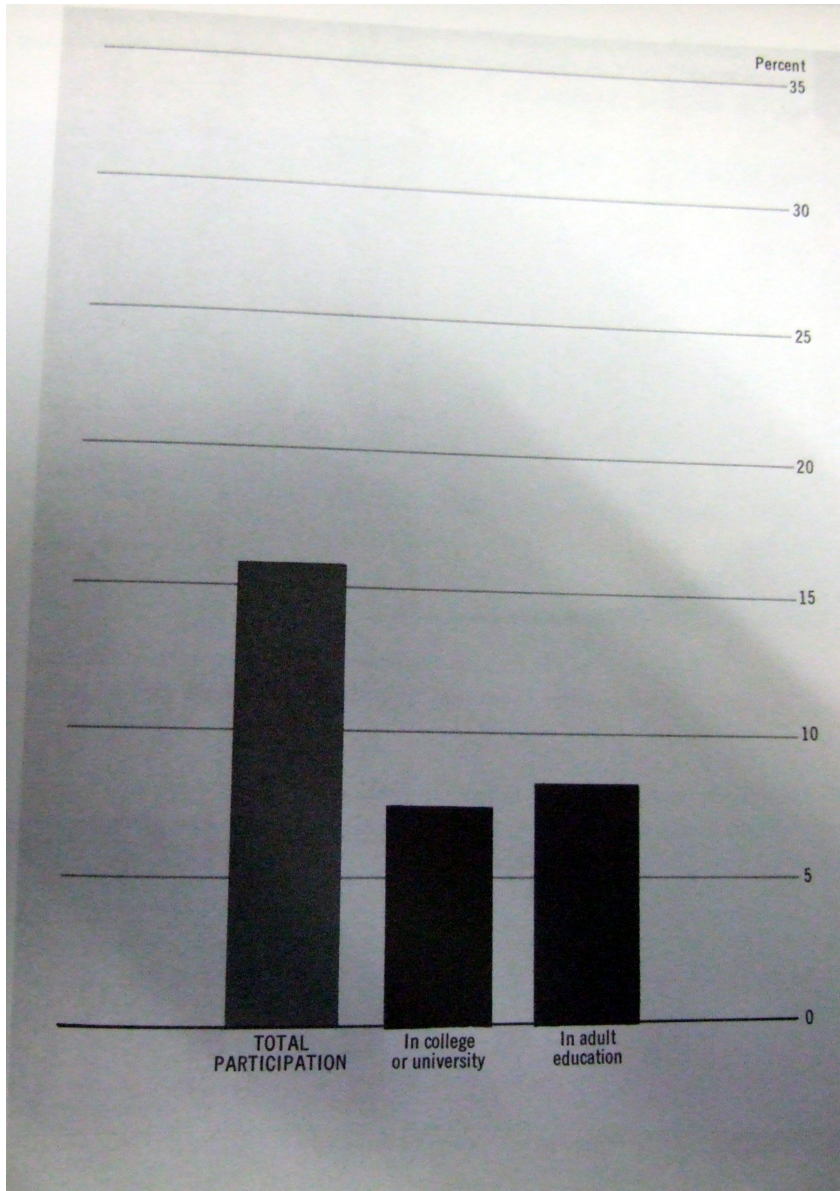
SOURCE: AA Sep 2005

Consider the “Lie Factor”

Duty is actually more than 50% of the total cost



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

Taken from:
Edward Tufte, "The Visual Display of Quantitative Information",
Graphics Press 1983



Earth at Night

Credit: C. Mayhew & R. Simmon (NASA/GSFC),
NOAA/ NGDC, DMSP Digital Archive

Use vector formats where possible

Vector

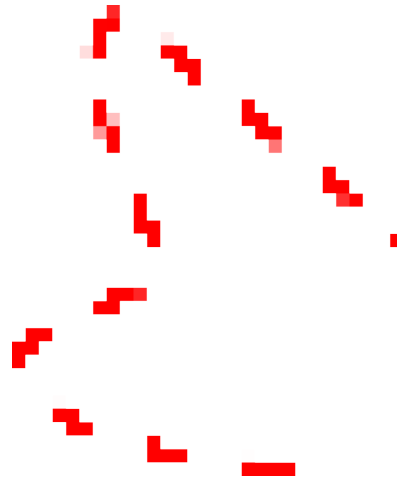


Graphs and Diagrams

SVG, EPS, PDF, WMF

Raster

Lossless



Screen Only

BMP, PNG

Lossy

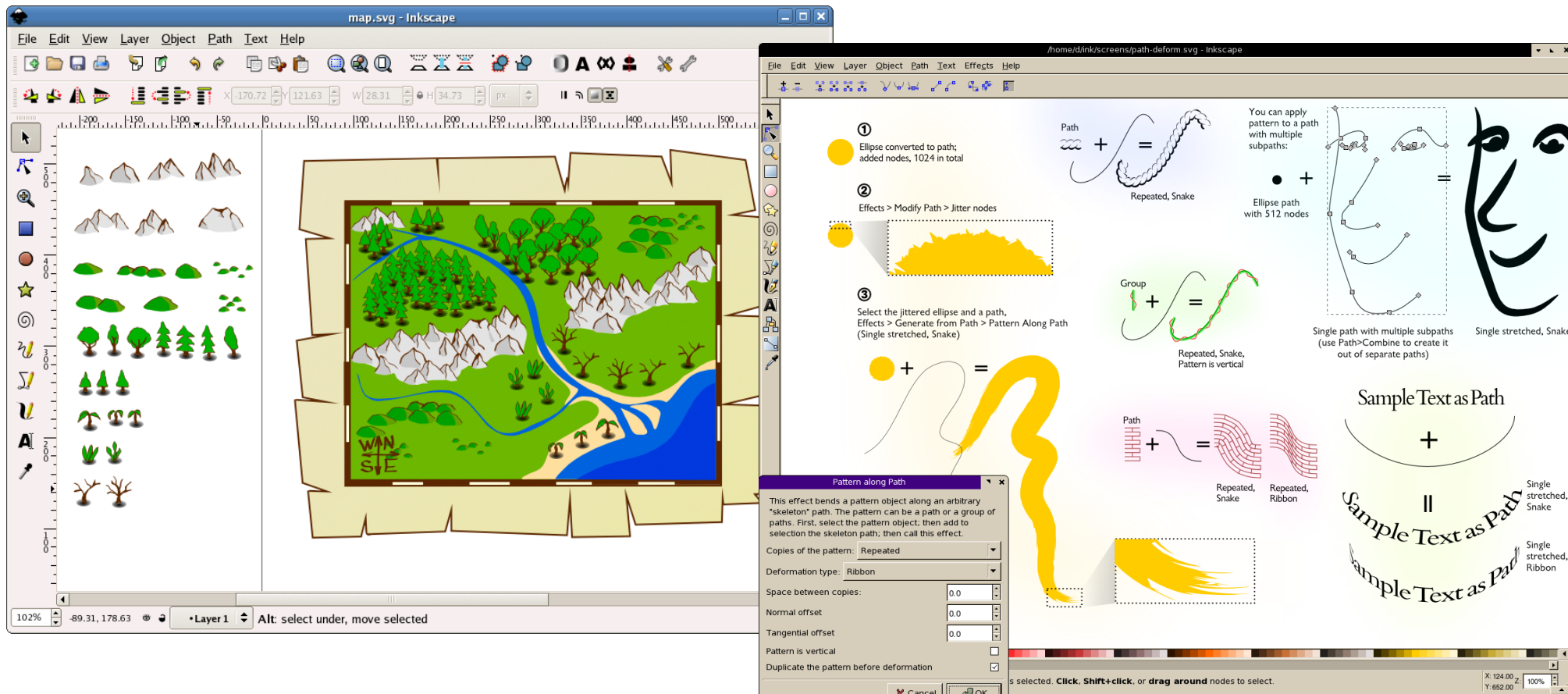


Photographs

JPG

Inkscape is a powerful tool for vector graphics

- Open-source
- Windows and Linux versions



Simple diagrams with Inkscape

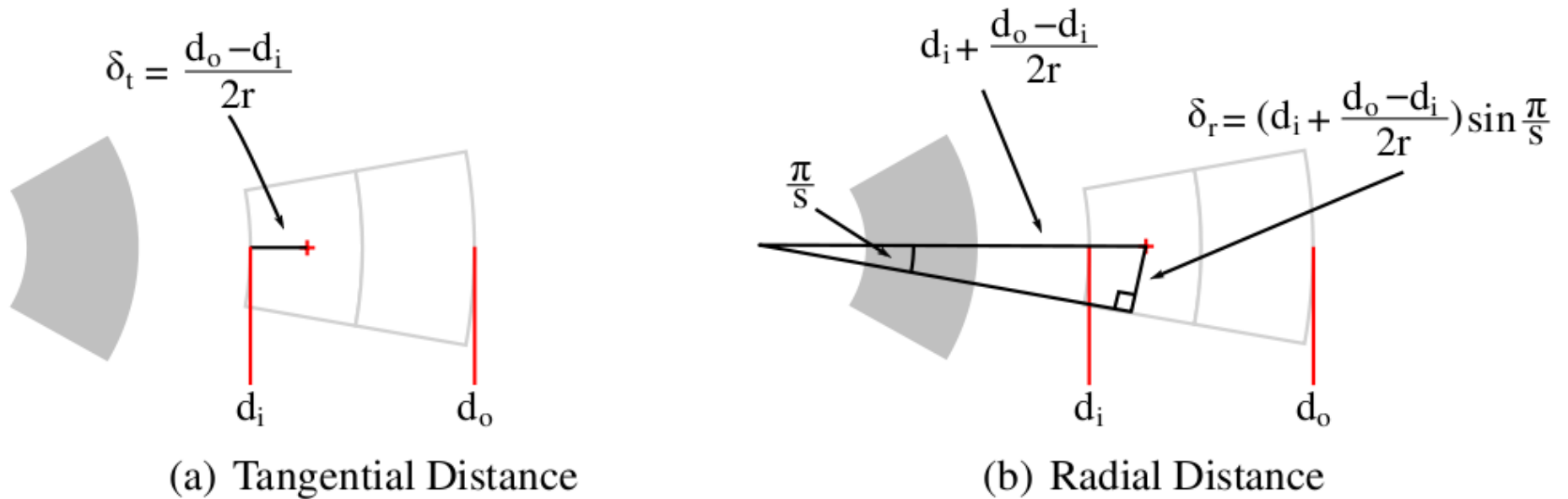


Figure 4.10: Tangential and radial size calculation

(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

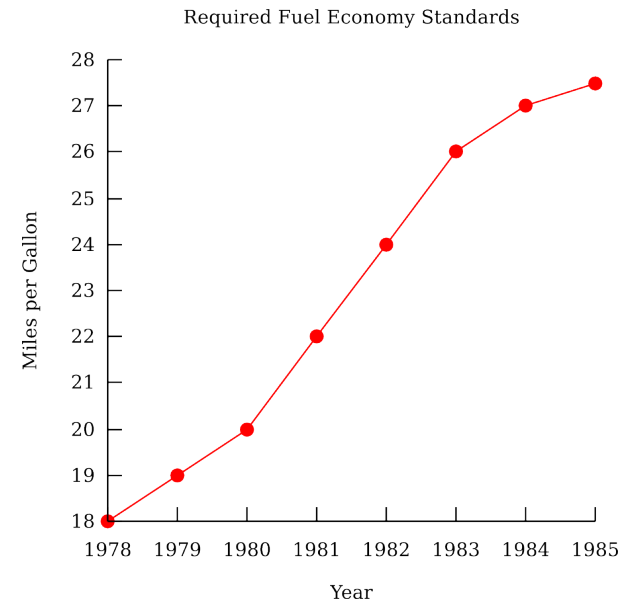
You can easily generate your own SVG files and inkscape will render them

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

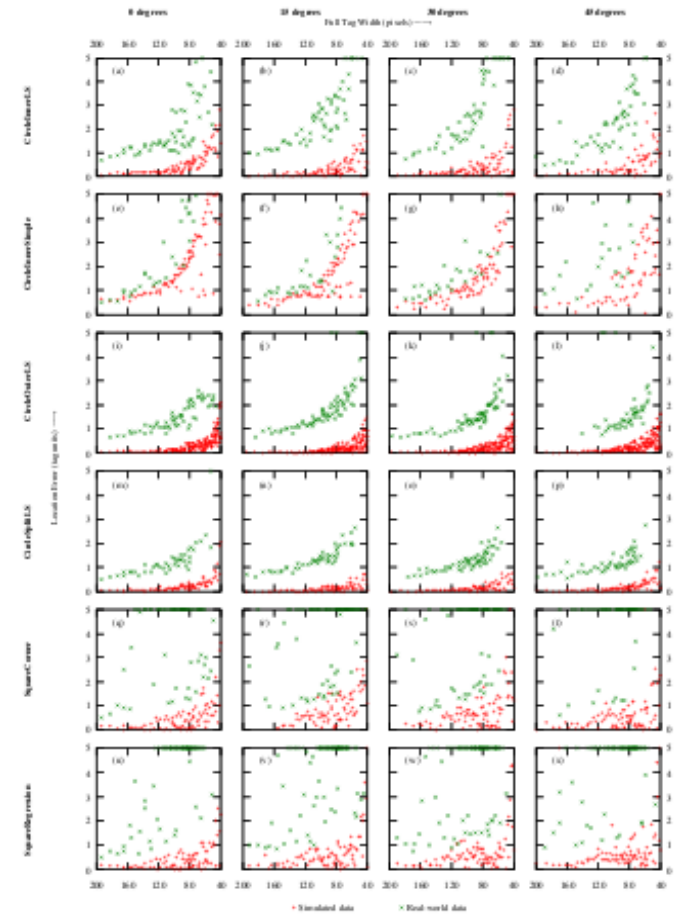
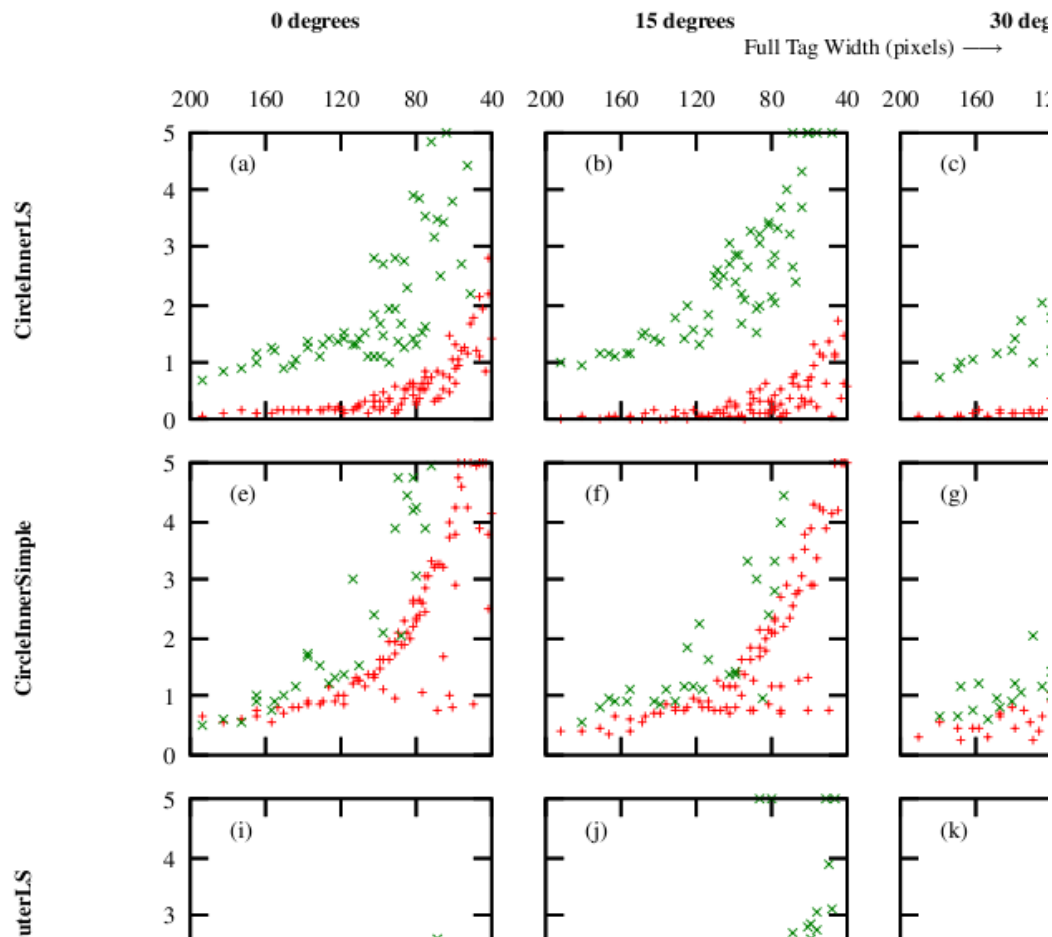
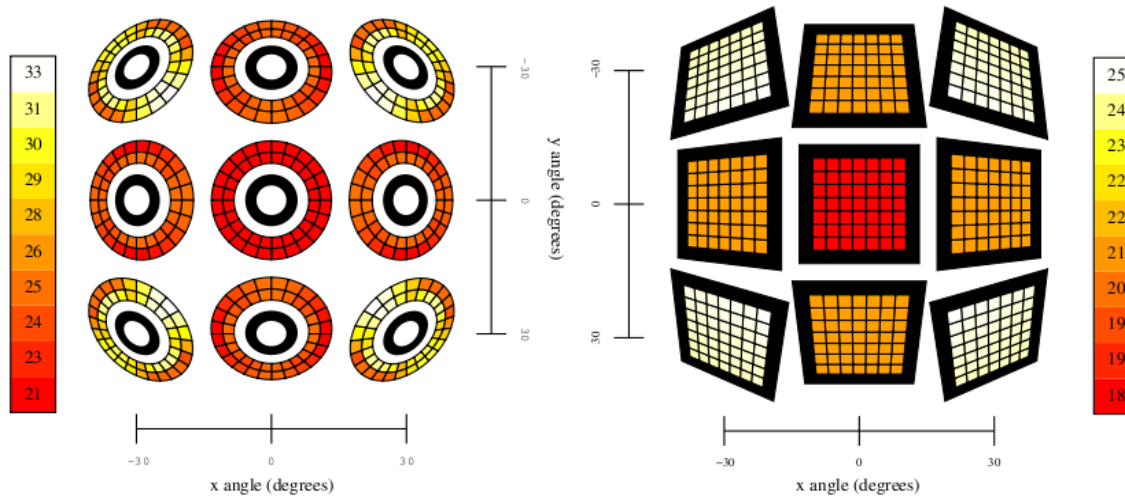
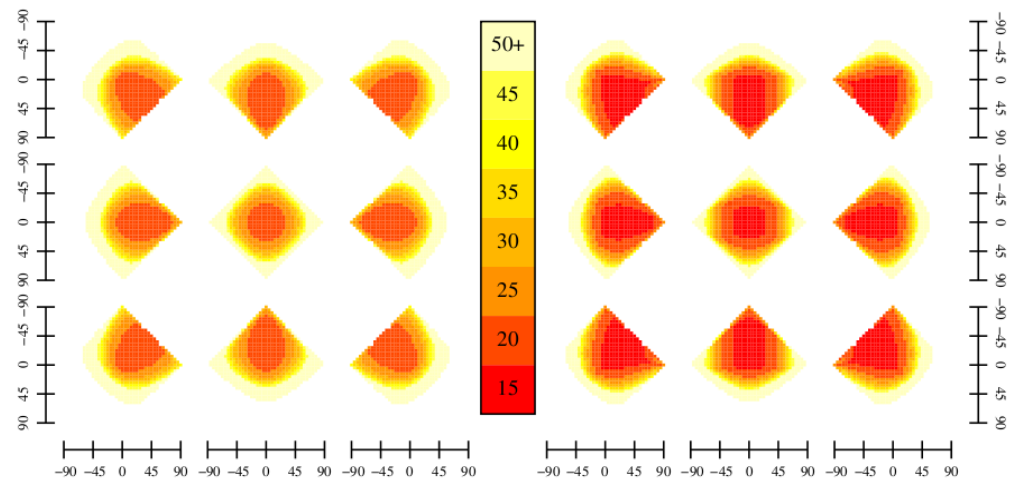


Figure 4.24: Real-world and simulated location error across processing pipelines

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

