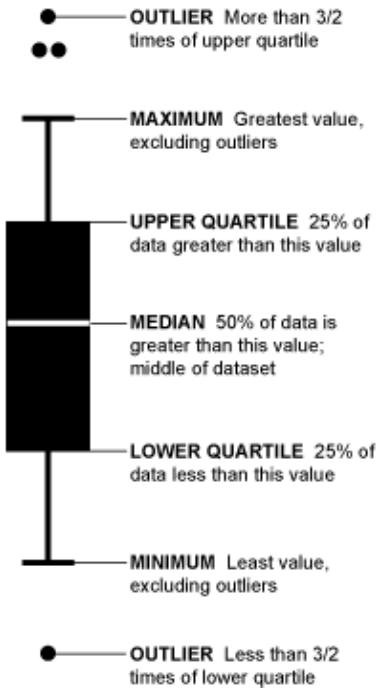
Visualisation

info-20002: foundations of informatics

Univariate - Boxplot



- Invented by J. Tukey,
- Display variability of a data variable
- Five-number summary (Tukey's Hinges)

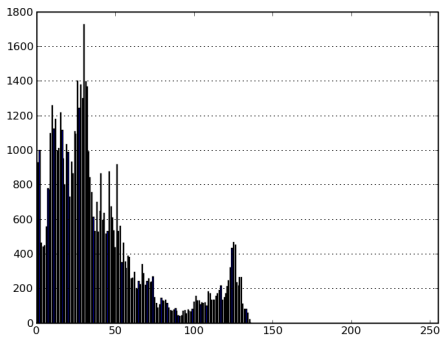
Image source: **How to Read and Use a Box-and-Whisker Plot**
McGill, R., Tukey, J. W., Larsen, W. A. (1978). "Variations of Box Plots". The American Statistician 32(1): 12–16
40 years of boxplots

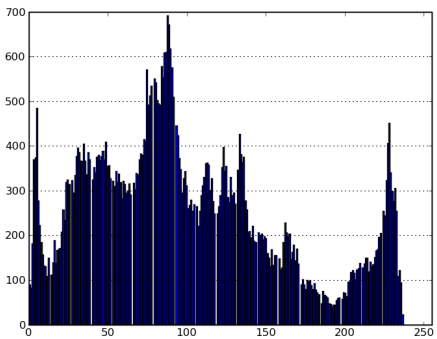
Univariate - World clouds

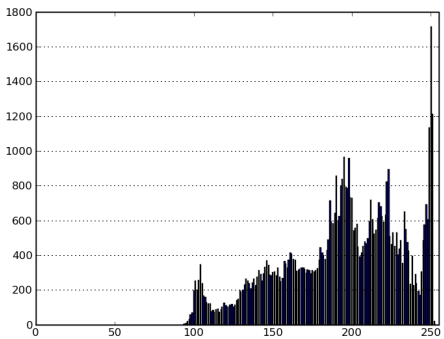
Data source: **Baby names in England and Wales, 2010**
Oliver and Olivia top list of most popular babies' names

Univariate - Histogram

Histograms display the distribution of a data variable. Image histograms show the distribution of pixel values (luminance).

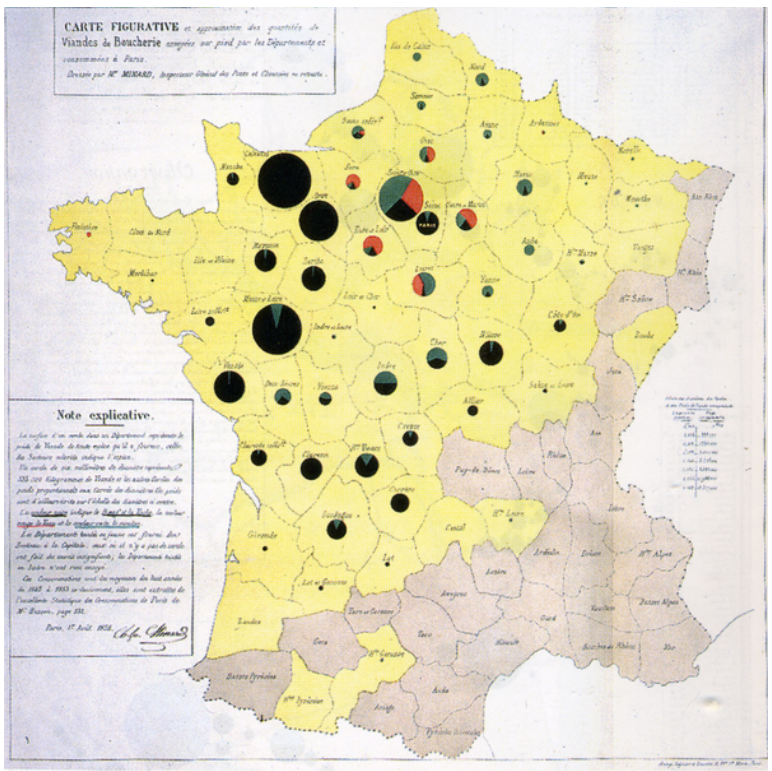






Bivariate - Pie chart

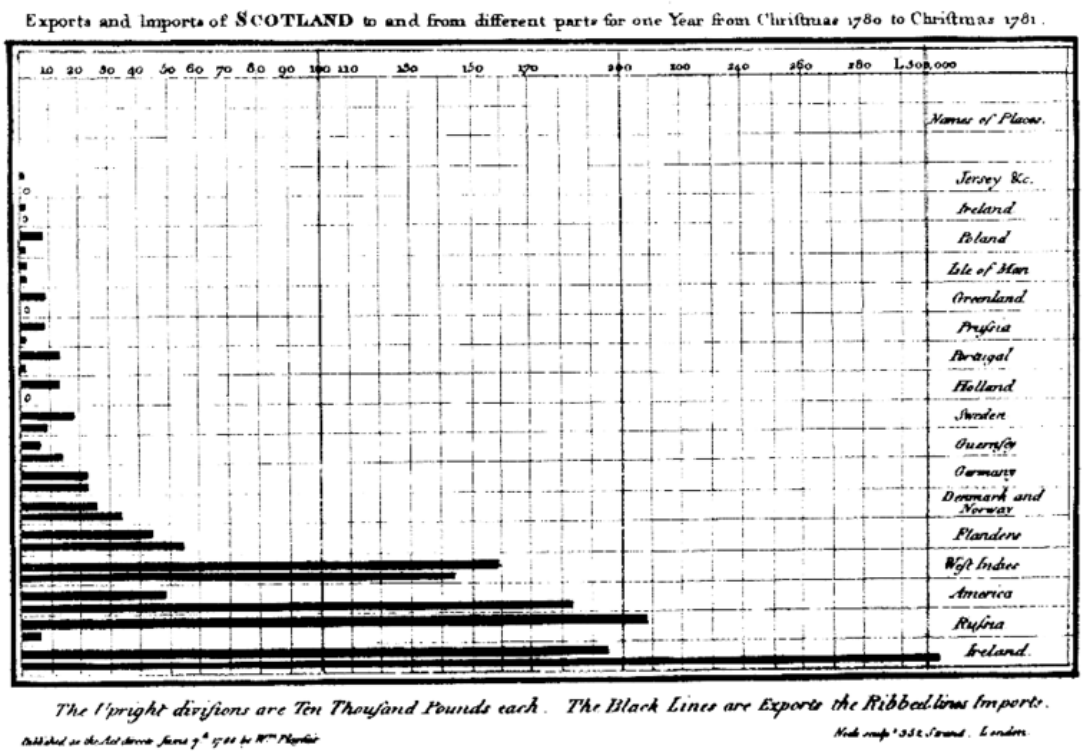
- A small number of nominal values over a single continuous variable
- Communicating the rough proportions



Data source: Stephen Few. **Save the Pies for Dessert.**

Bivariate - Bar chart

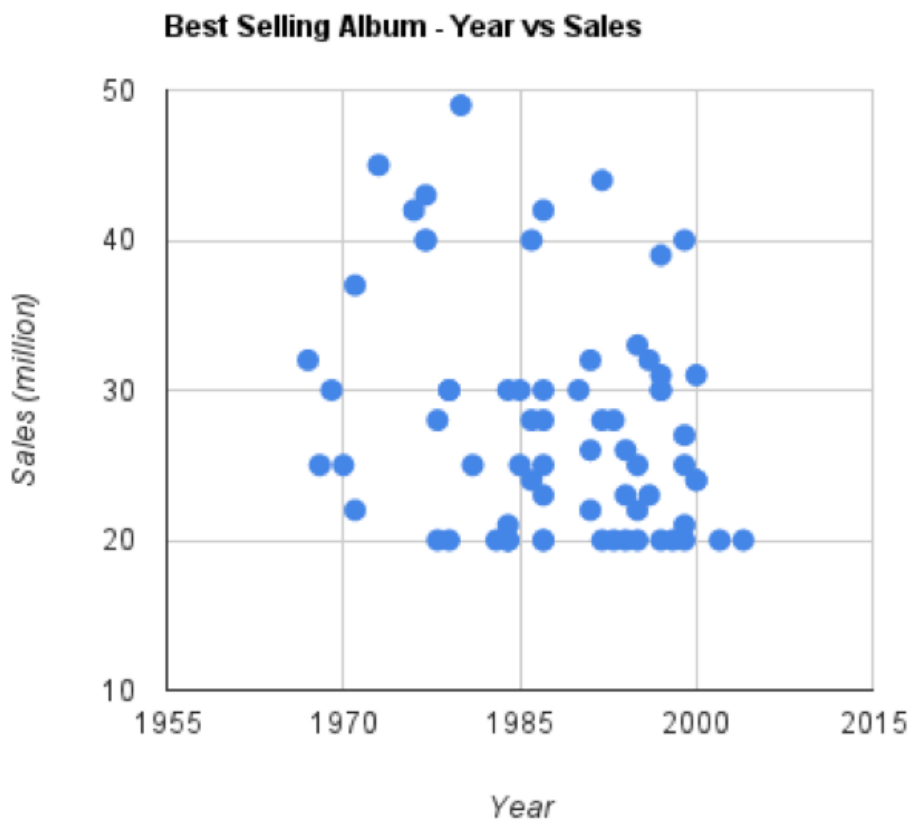
- Continuous values vs nominal values
- Comparing a property of nominal entities



William Playfair (1786). "The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century".

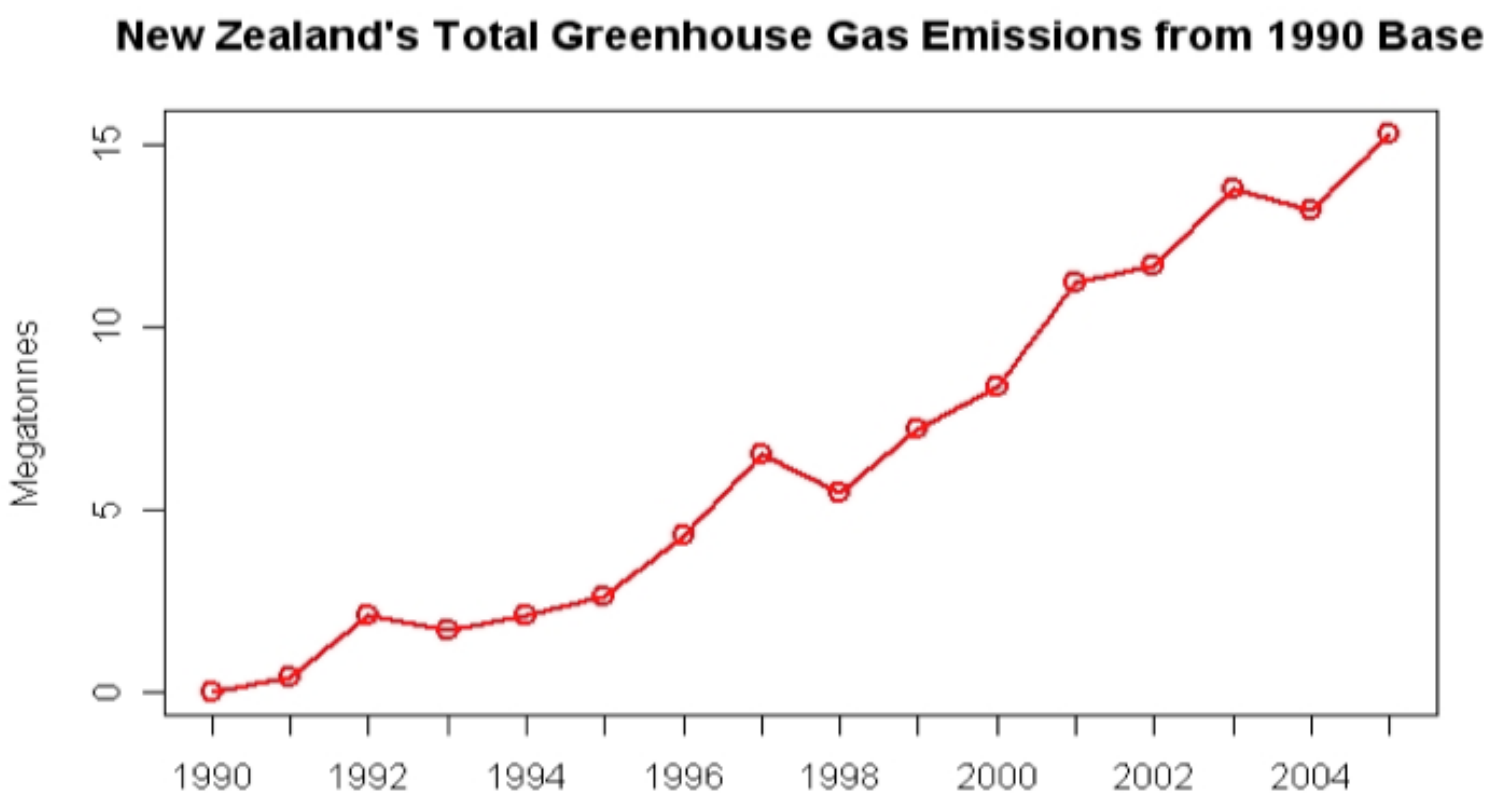
Bivariate - Scatter plot

- Relationship between two continuous variables

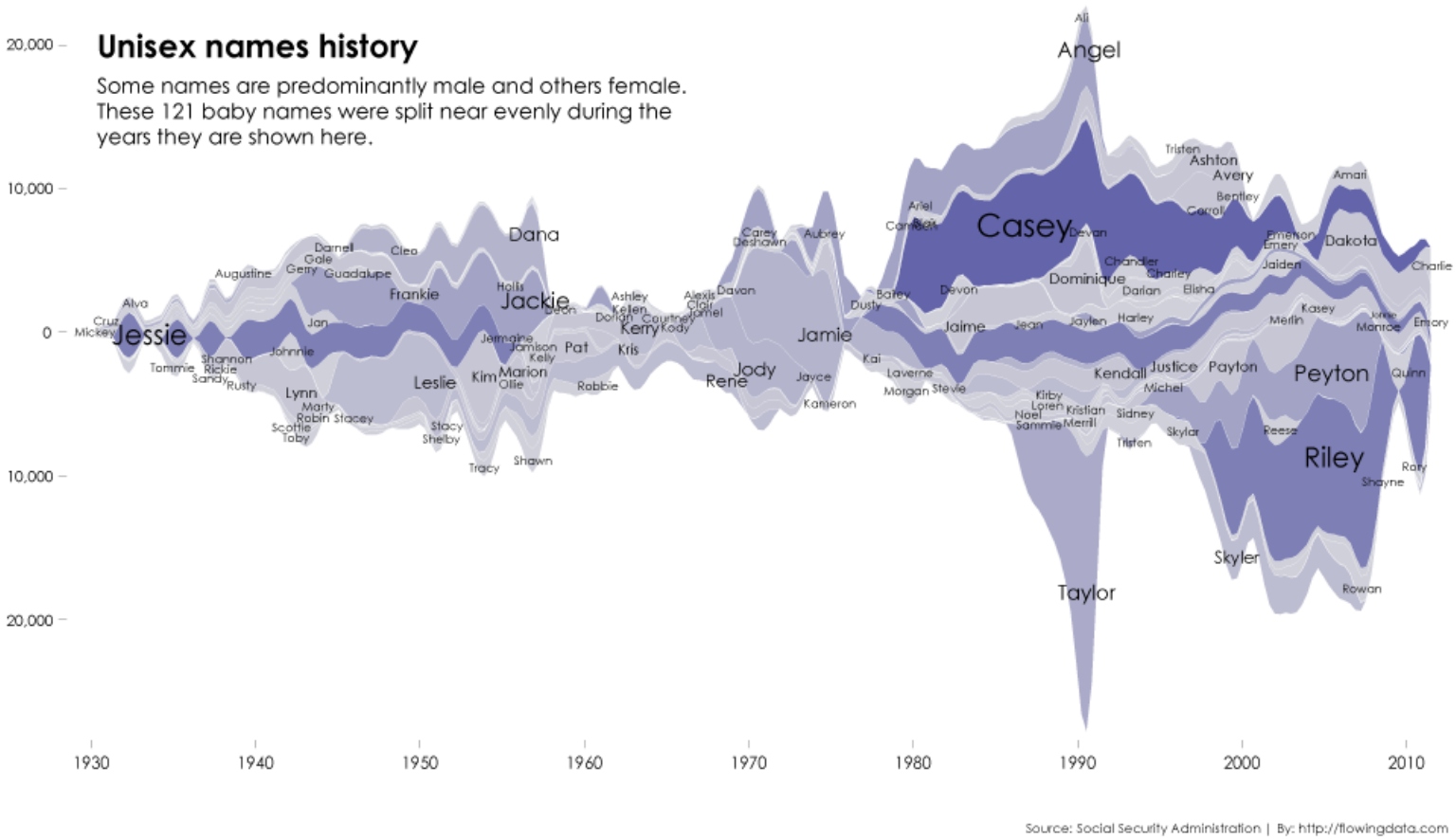


Bivariate - Line chart

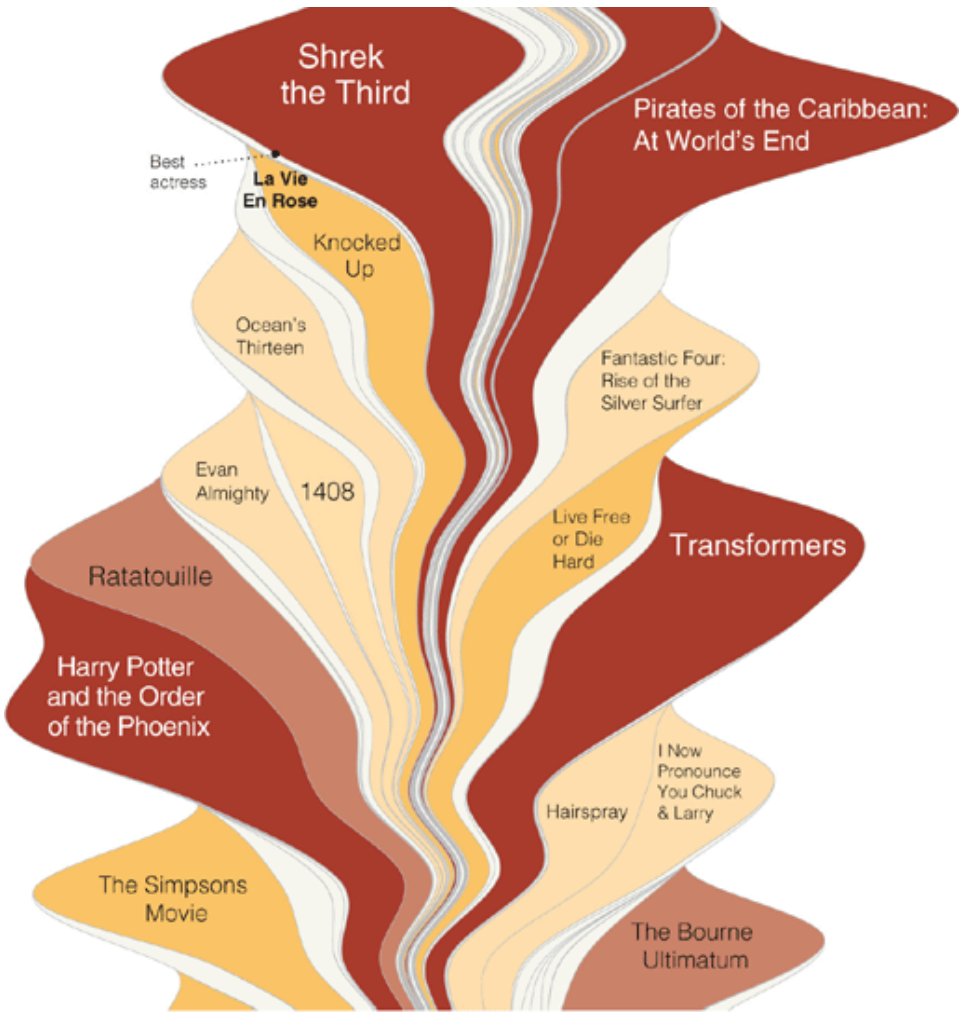
- Display a trend of a continuous variable over time
- Commonly used to compare two or more continuous variables



Multivariate - Stacked area chart/Stream graph



Multivariate - Stacked area chart/Stream graph



Byron, L., Wattenberg, M. (2008). **"Stacked Graphs – Geometry & Aesthetics"**. IEEE Transactions on Visualization and Computer Graphics 14(6): 1245–1252.

Multivariate - Trellis (small multiples)

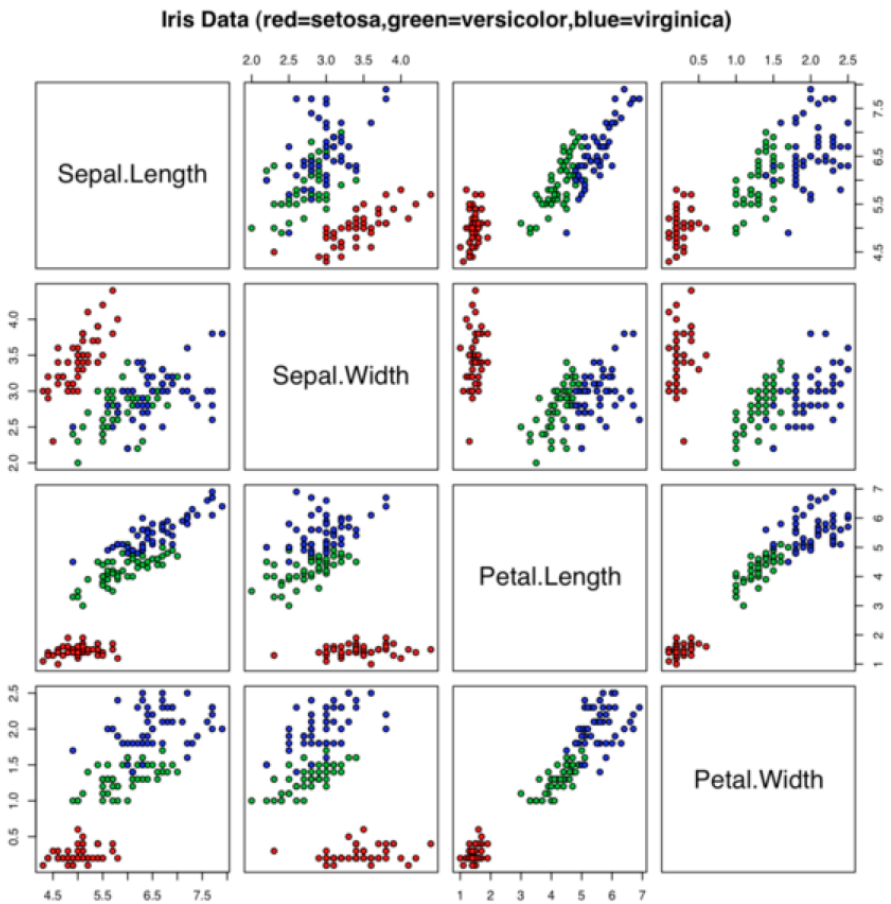
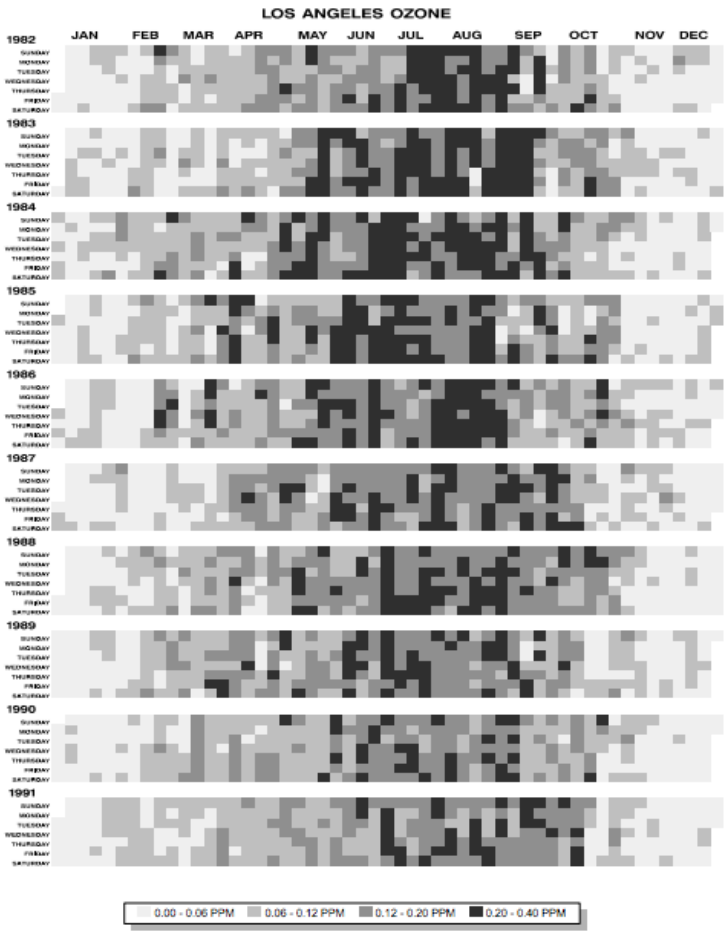


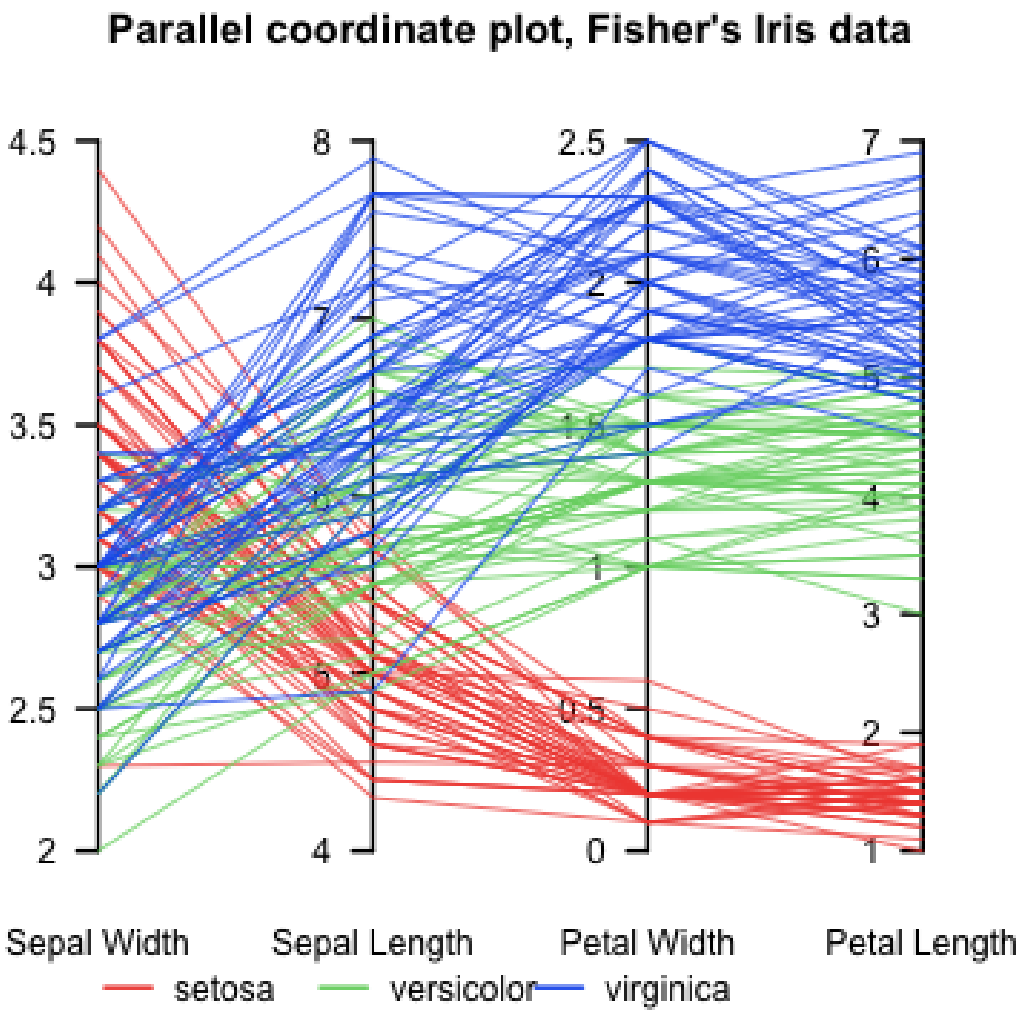
Image source: **Iris versicolor**. CC BY-SA 3.0.

Multivariate - Tile map

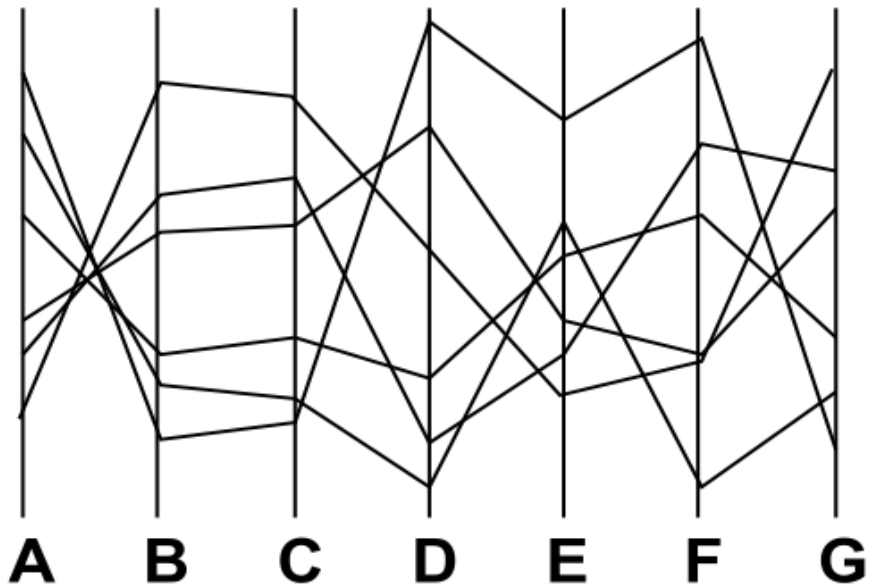


Mintz, D., Fitz-Simons, T. & Wayland, M. "Tracking Air Quality Trends with SAS/GRAPH", SUGI 22 Proceedings, 807-812.

Multivariate - Coordinate Plot (Parallel coordinates)

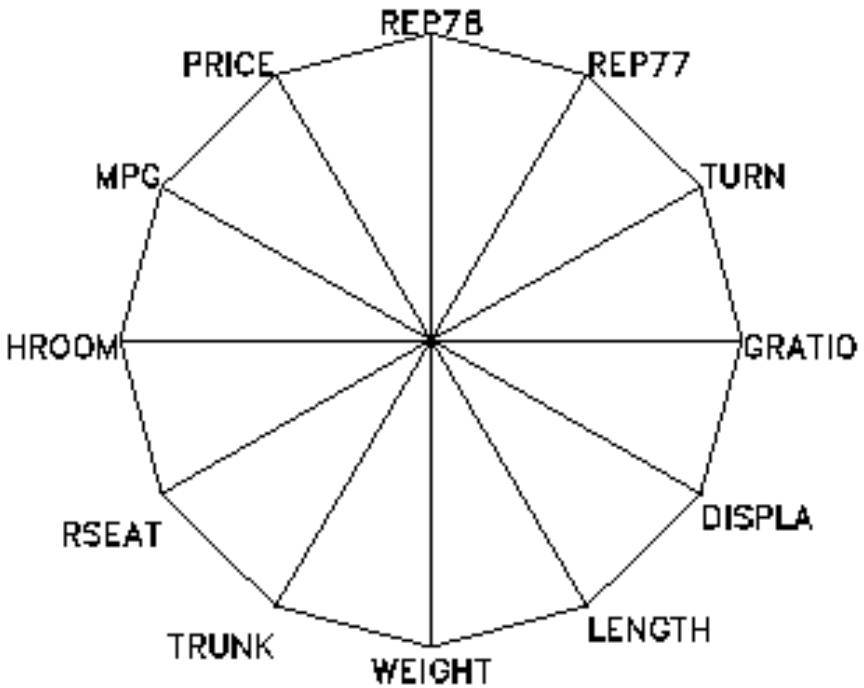


Multivariate - Coordinate Plot (Parallel coordinates)

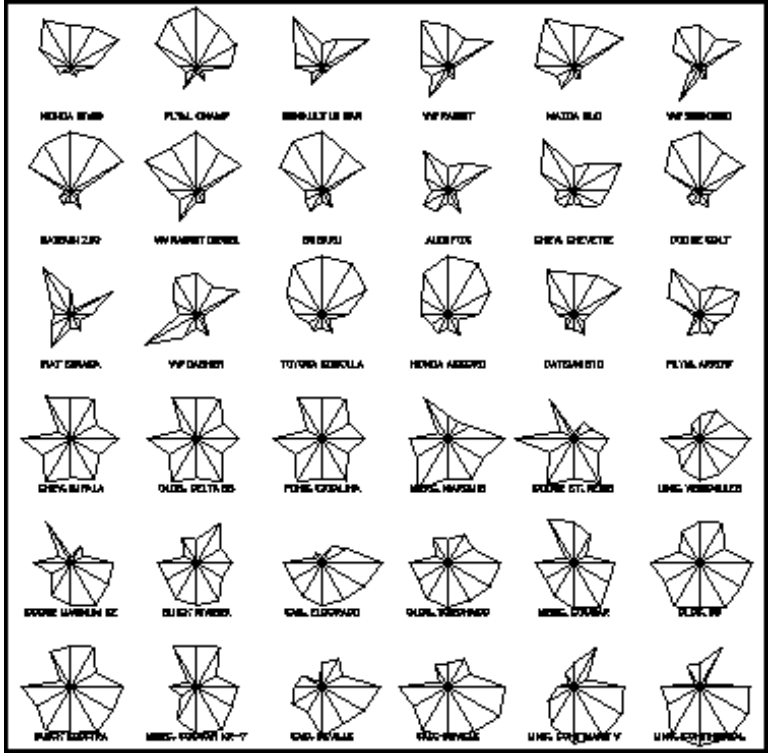


The trade-off between A and B, and the correlation between B and C, are immediately apparent. The trade-off between B and E, and the correlation between C and G, are not.

Multivariate - Star plot/radar chart/spider chart



Variable Assignment Key



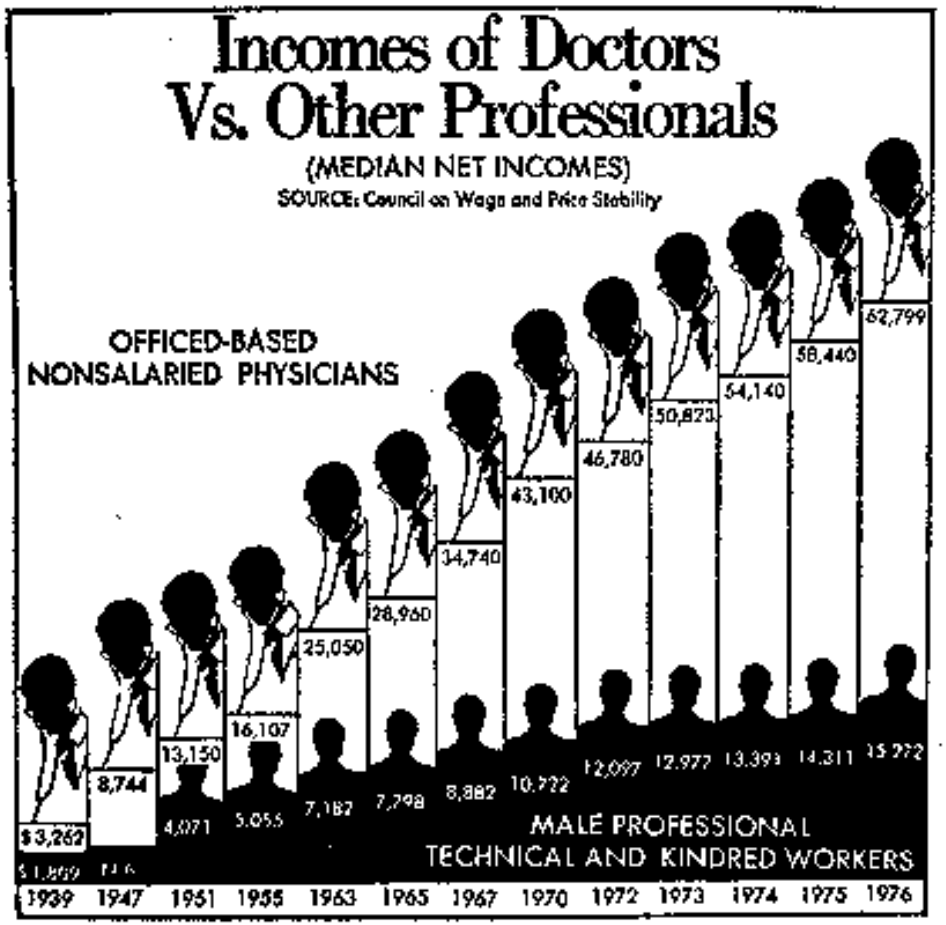
Friendly, M. (1991). **Statistical Graphics for Multivariate Data**. SAS SUGI 16 Conference, Apr, 1991.

Ronald Reagan's Address on Federal Tax Reduction

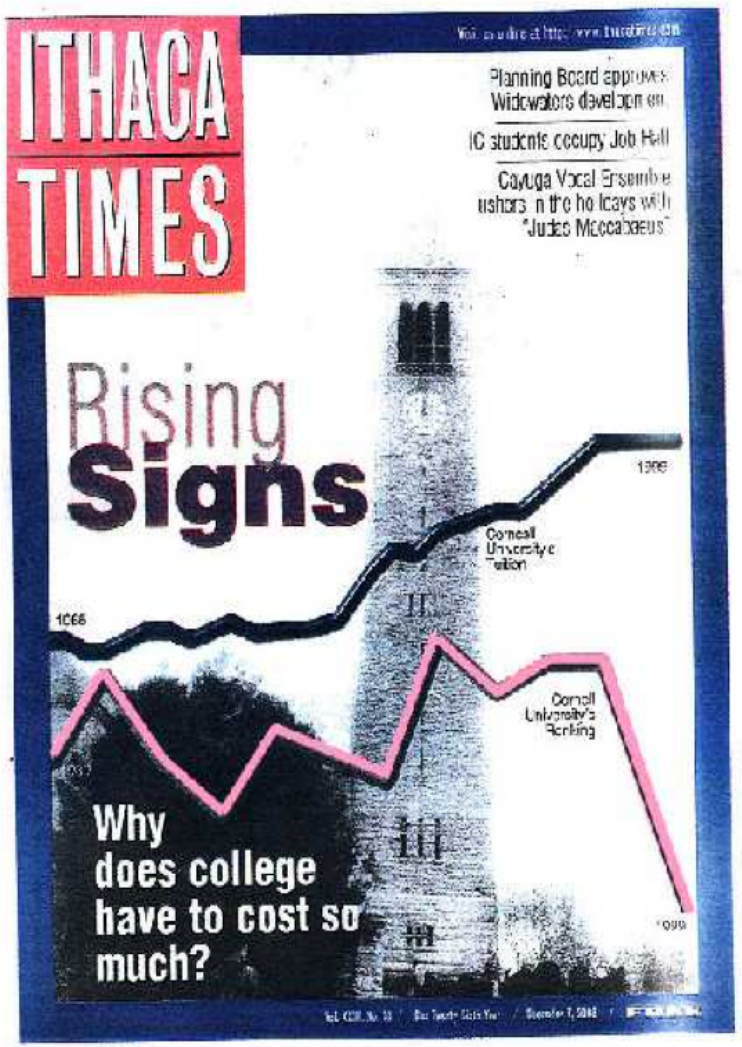
Ross Perot 1992 - Balancing the Budget & Reforming Government

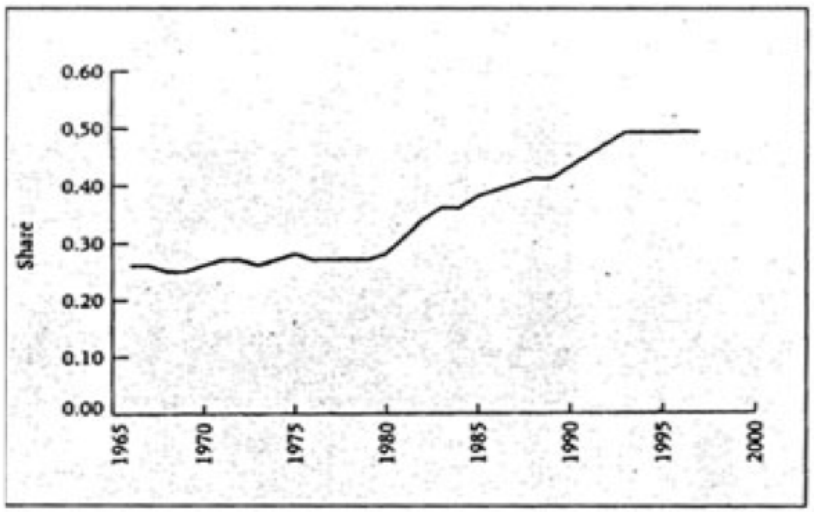
Guidelines

- Don't change the scale across charts which are intended to be contrasted (rubber scaling)
- Don't leave out the origin for ratio variables unless necessary
- Don't use line chart for comparing nominal variables
- Don't change the ordering of ordinal variables

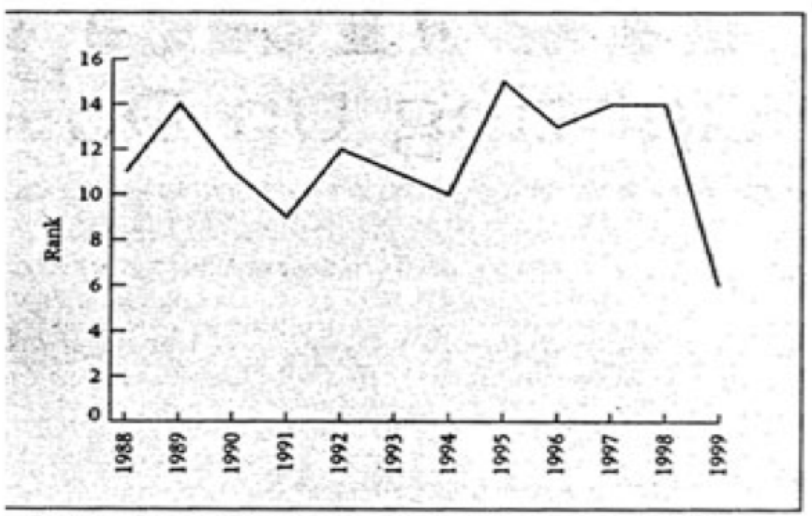


Wainer, H. (1997). Visual Revelations: Graphical Tales of Fate and Deception From Napoleon Bonaparte To Ross Perot





BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.



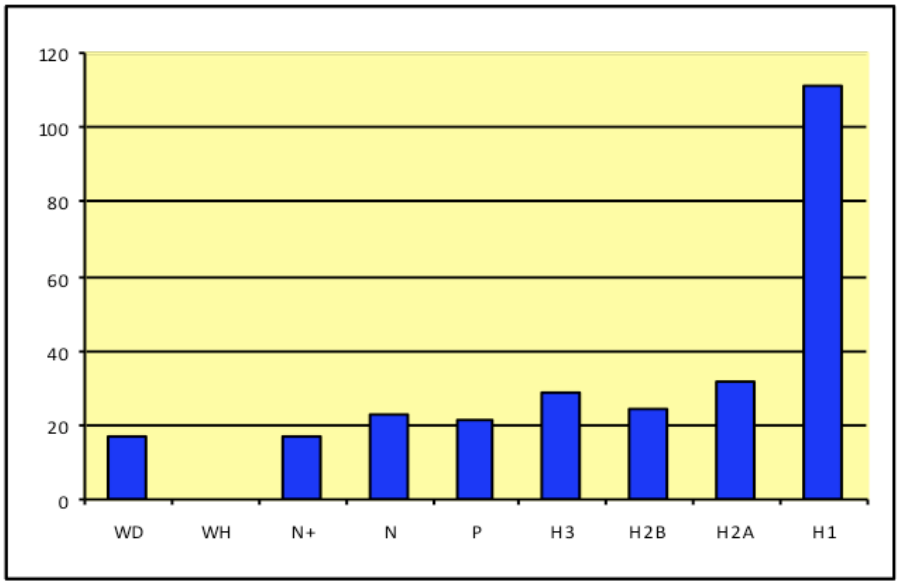
PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.

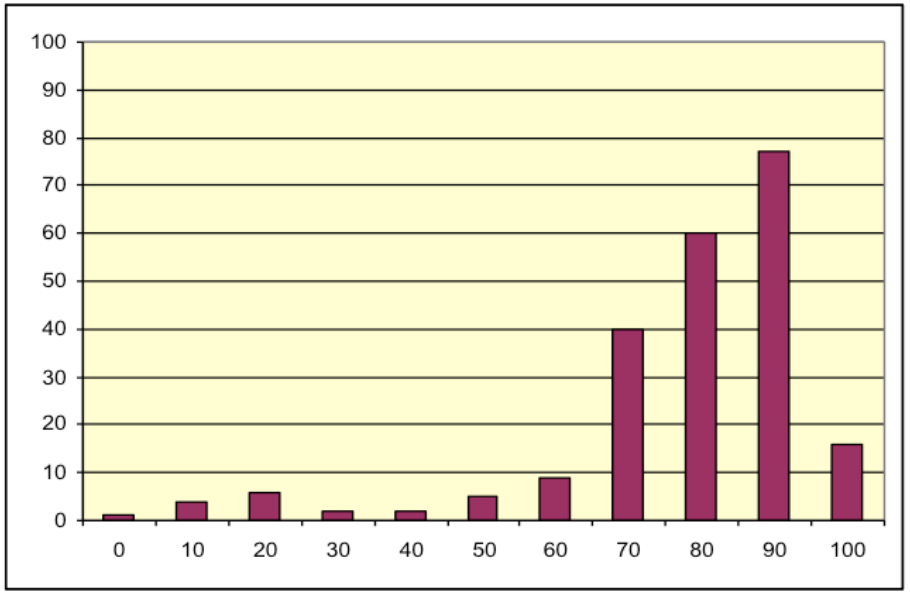
From Ithaca Times (Dec. 7, 2000) via http://www.datavis.ca/

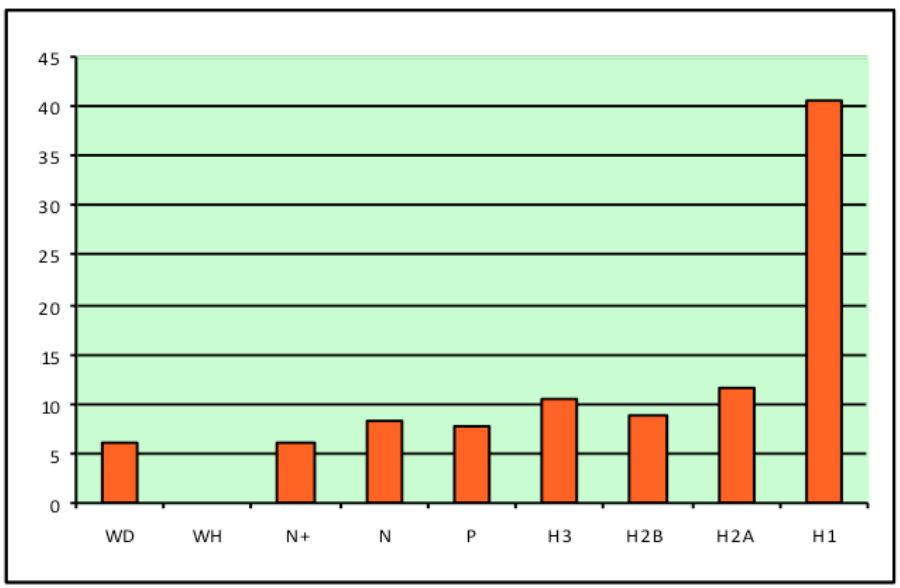
Guidelines

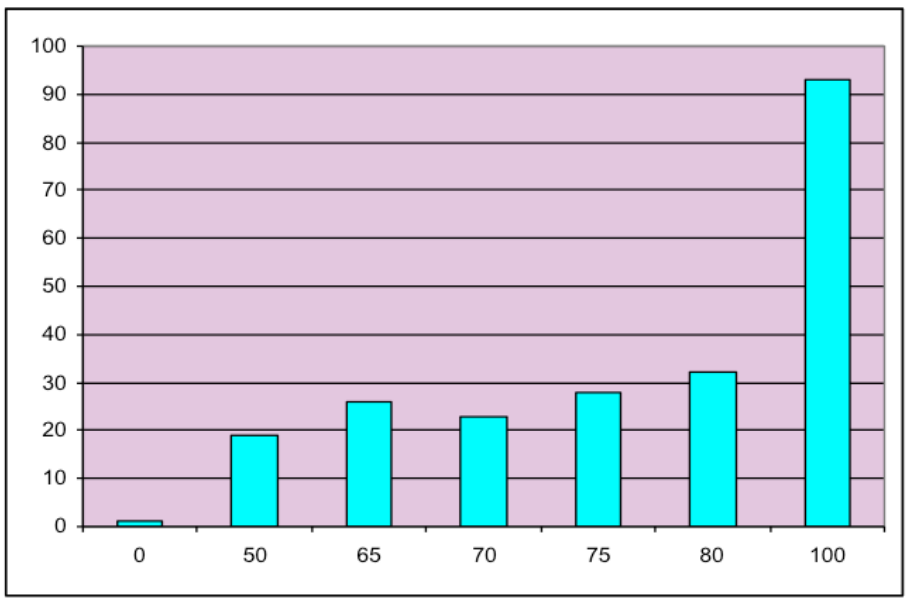
- Red-green color blindness
- Visually impaired people need HIGH contrast
- Set the context for what you are presenting
  - axes labels, legend, units, captions
- Determine what numbers need to be presented to get your message across
  - more numbers vs. less; exact vs. rough; major vs minor axes, scale, absolute vs relative

Tufte E., (1983). The Visual Display of Quantitative Information


120
100
80
60
40
20
0
WD
WH
N+
N
P
H3
H2B
H2A
H1


100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100


45
40
35
30
25
20
15
10
5
0
WD
WH
N+
N
P
H3
H2B
H2A
H1


100
90
80
70
60
50
40
30
20
10
0
0
50
65
70
75
80
100

Measuring visualisation effectiveness

- **data density index (ddi)**
  - the number of numbers plotted per square inch
  - in popular media ranging from .1 to 362
- **data ink ratio**
  - the ink used for data divided by the total ink user for the graphic
  - the proportion of ink used for non-erasable display of information
  - 1.0 - redundant ink

Tufte E., (1983). The Visual Display of Quantitative Information.
Data-ink ratio
Wainer, H. (1984). **How to display data badly**. American Statistician 38(2):137-147

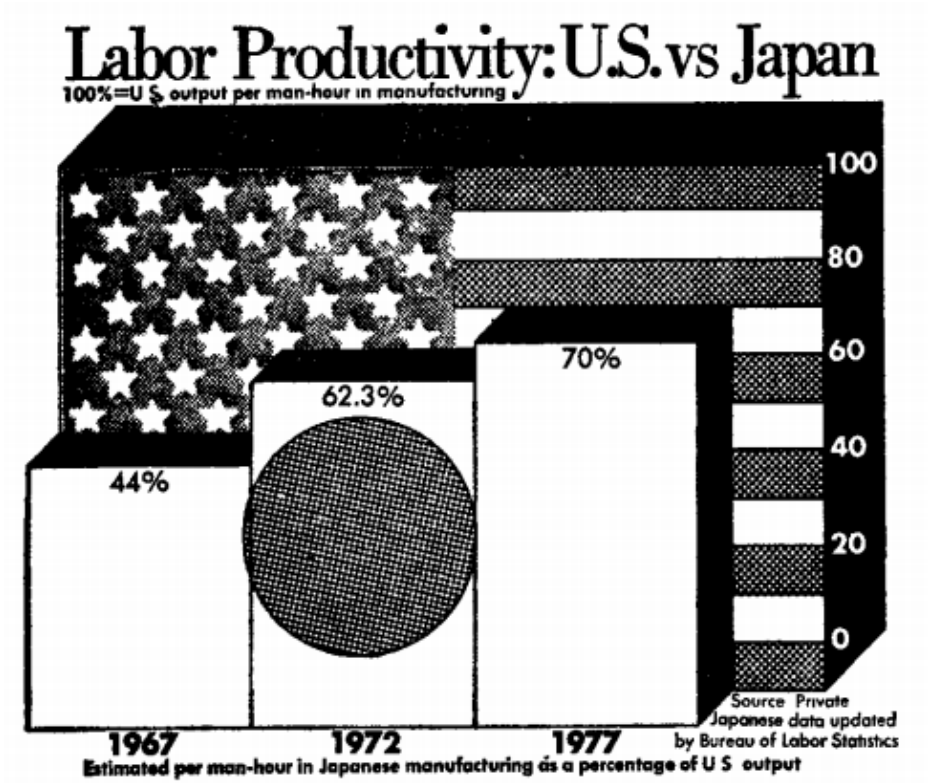
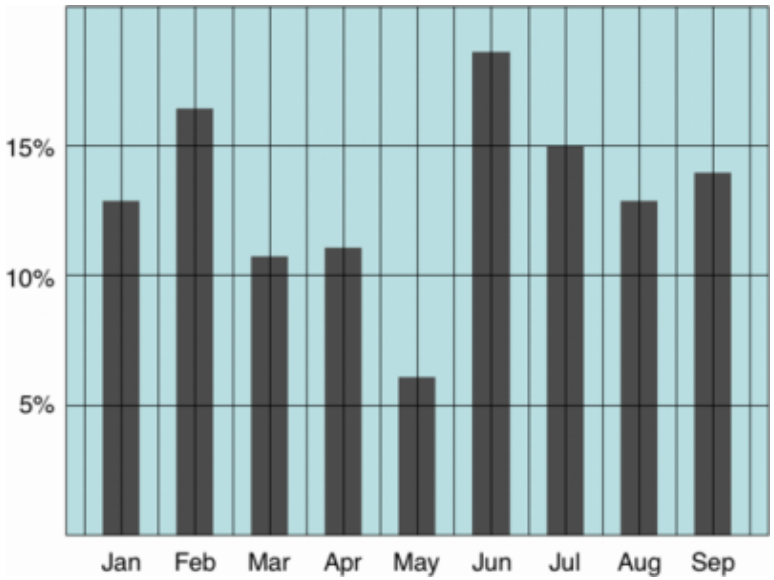
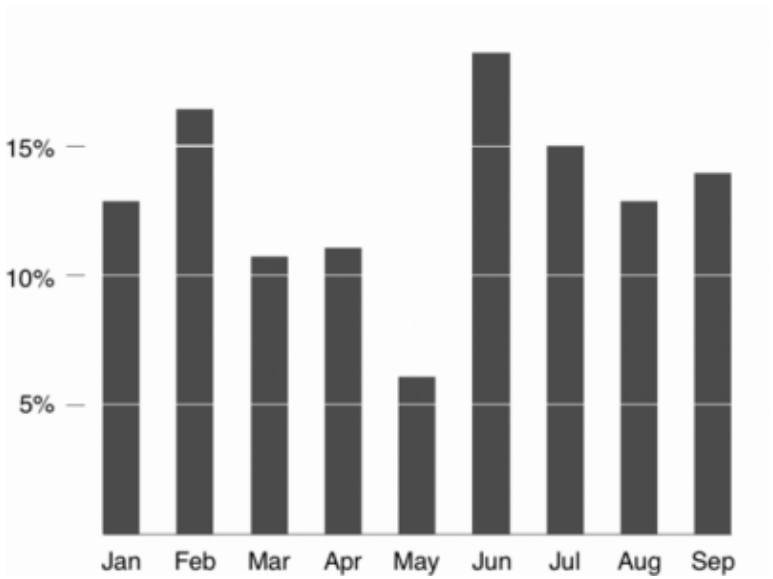


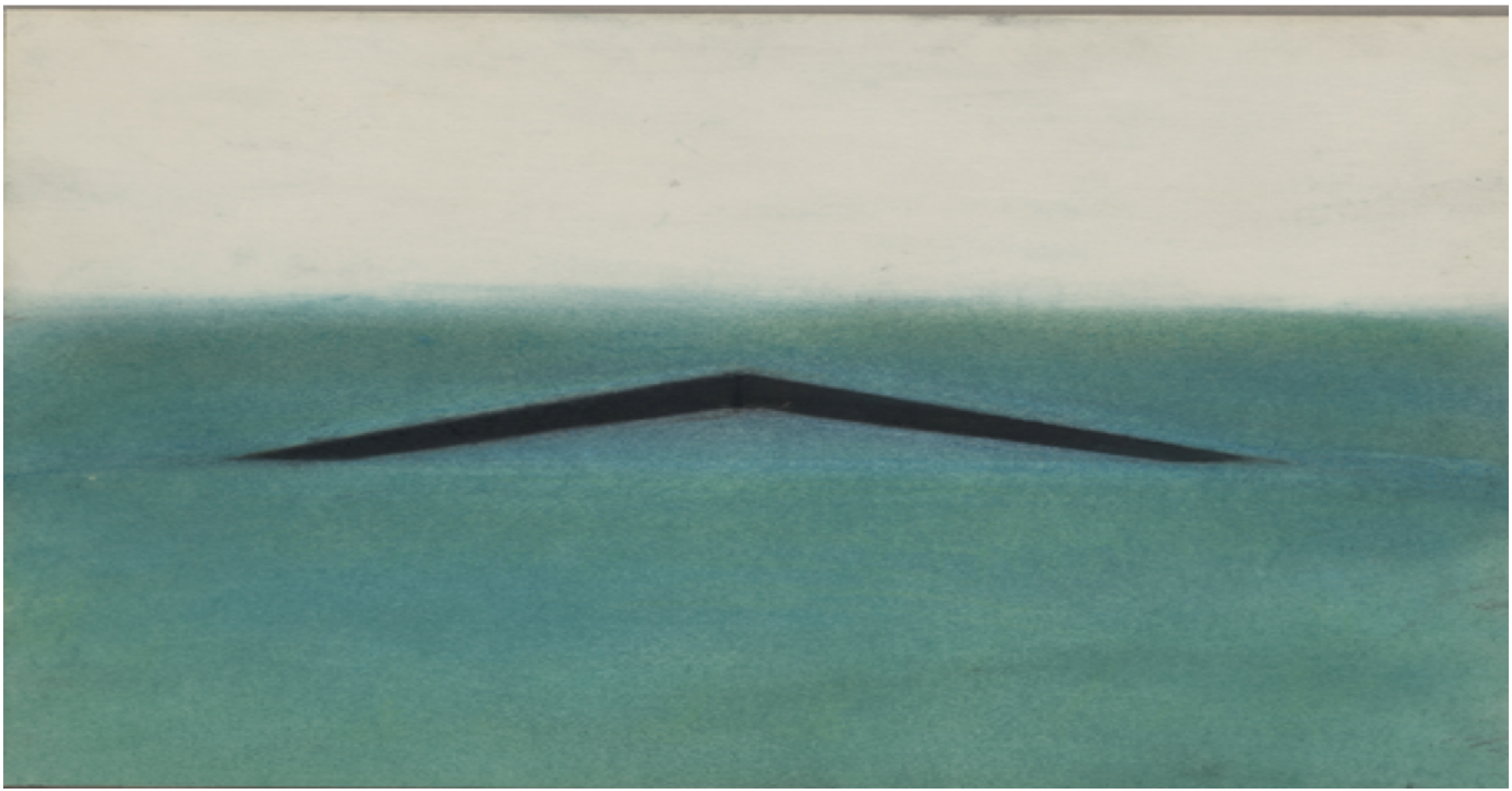
Figure 3 A low density graph (© *1978,* The Washington Post) *with chart-junk to fill in the space (ddi = .2)*


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

Vietnam Veterans Memorial, Washington DC.

Problem: How to display 58,195 names in a memorial space?



Visualisation Steps

- **Data Definition**
  Define the visualisation goal and the supporting data variables
- **Visualisation Selection**
  Select appropriate visual structure
- **Data Pre-Processing**
  Preparing raw data to visualisation-ready data
- **Visual Transformation**
  Mapping data variables to visual elements

matplotlib

- plotting library for python
- produce static/non-interactive visualisation.

http://matplotlib.org/
the tutorial: **http://matplotlib.org/users/pyplot_tutorial.html**
the cookbook: **http://wiki.scipy.org/Cookbook/Matplotlib**
the gallery: **http://matplotlib.org/gallery.html**

matplotlib - structure of usage

- **The `matplotlib` library**

  ```
  >>> import matplotlib
  ```

- **The device dependent backend**
  Specify the drawing engine that renders the visual to a file or a display device.
  Example:

  - 'PS' for creating postscript file
  - 'SVG' for creating scalar vector graphics (SVG file),
  - 'Agg' for creating PNG file:

  ```
  >>> matplotlib.use('Agg')
  ```

- **The `pylab` interface**
  Provide a set of functions on top of the underlying matplotlib library
  Provide functions like `plot`, `boxplot`, and `bar`

  ```
  >>> from pylab import * # or import matplotlib.pyplot as plt
  ```

using matplotlib

To display the plot result in a web page, put this code at the start of your script:

```
import matplotlib
matplotlib.use('Agg')
```

and this code at the end:

```
@app.route('/viz-app', methods=['GET'])
def handler():
  savefig("plot.png", dpi=100)
  body = '<html><body><img src="plot.png" /></body></html>'
  return body, 200, {'Content-Type': 'text/html'}
```

using matplotlib

To send the plot result directly to the browser, put this code at the start of your script:

```
import matplotlib
matplotlib.use('Agg')
```

and this code at the end:

```
@app.route('/viz-app', methods=['GET'])
def handler():
  savefig("plot.png", dpi=100)
  binary = open('plot.png').read()
  return binary, 200, {'Content-Type': 'image/png'}
```