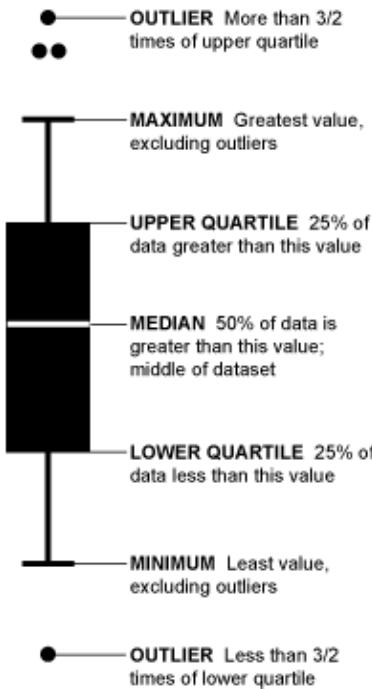


Visualisation

info-20002: foundations of informatics

Univariate - Boxplot



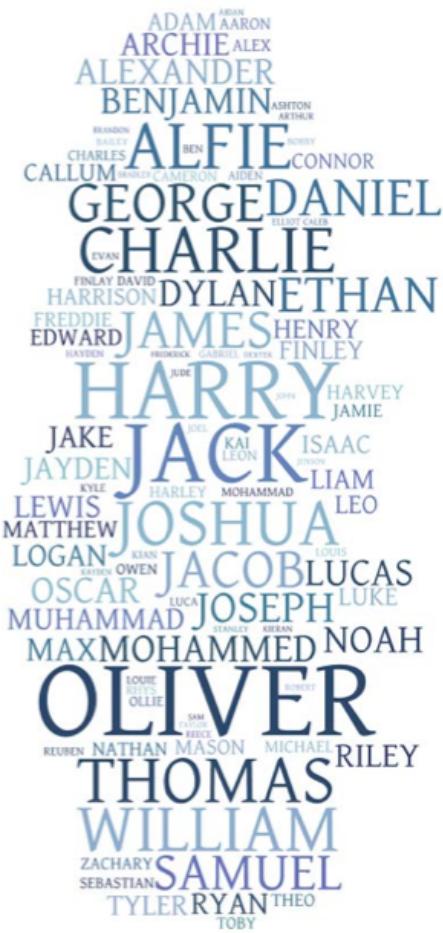
- Invented by J. Tukey,
- Display variability of a data variable
- Five-number summary (Tukey's Hinges)

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**](#)

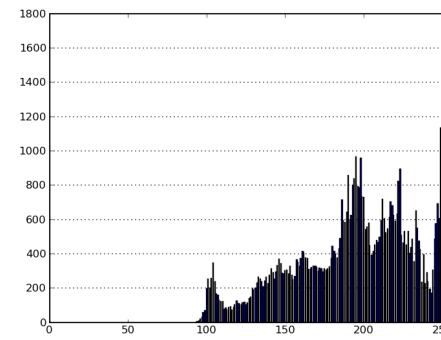
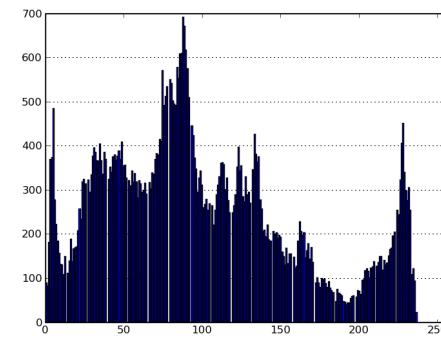
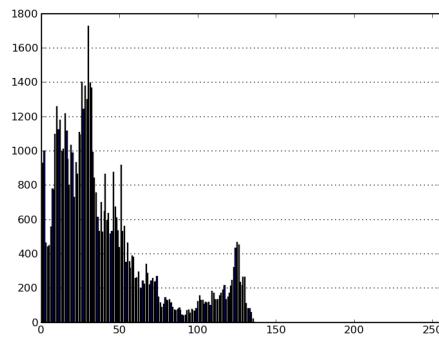
Univariate - World clouds



Data source: [Baby names in England and Wales, 2010](#)
[Oliver and Olivia top list of most popular babies' names](#)

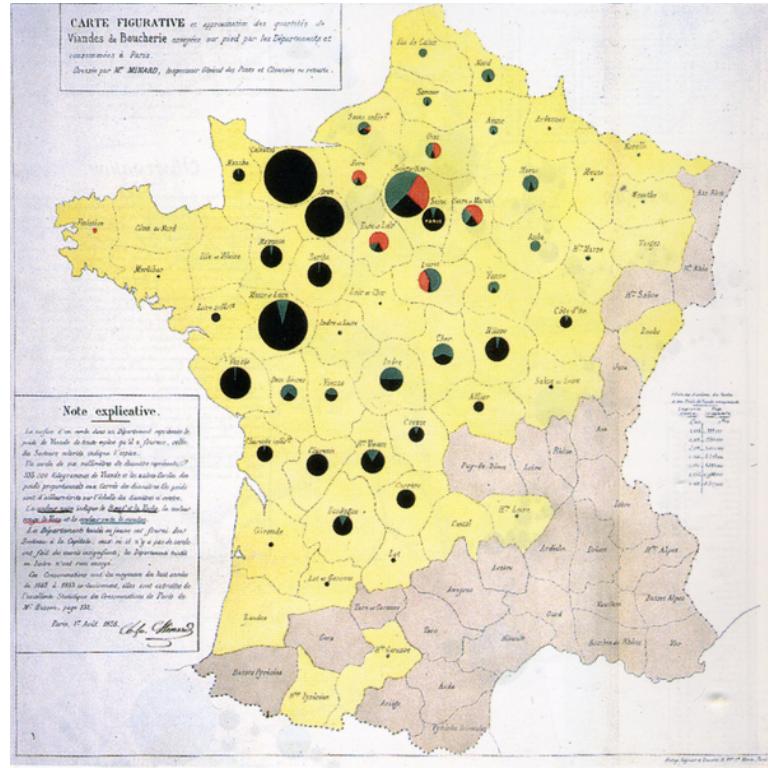
Univariate - Histogram

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

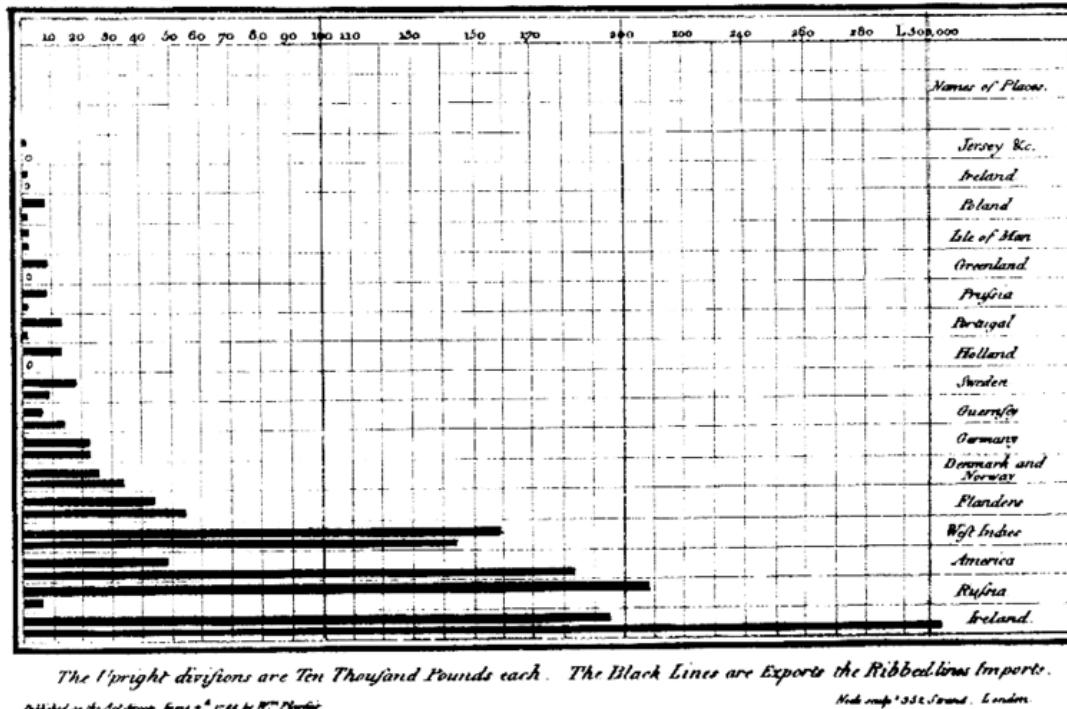


Data source: [Save the Pies for Dessert, Stephen Few](#)

Bivariate - Bar chart

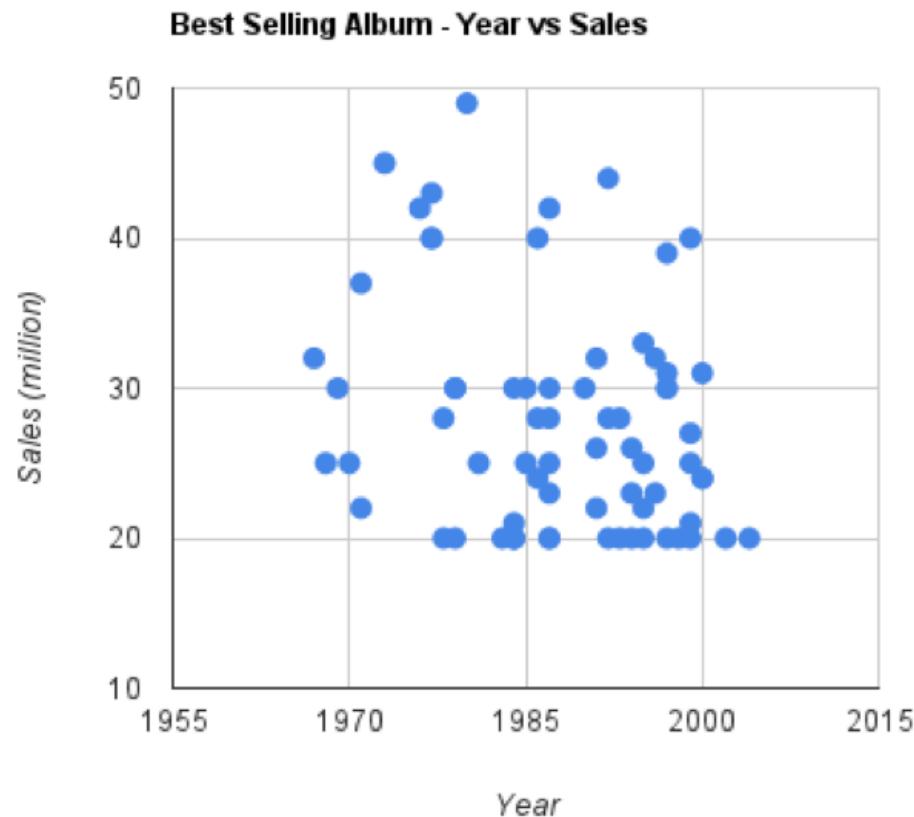
- Continuous values vs nominal values
- Comparing property of nominal entities

Exports and Imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.



Bivariate - Scatter plot

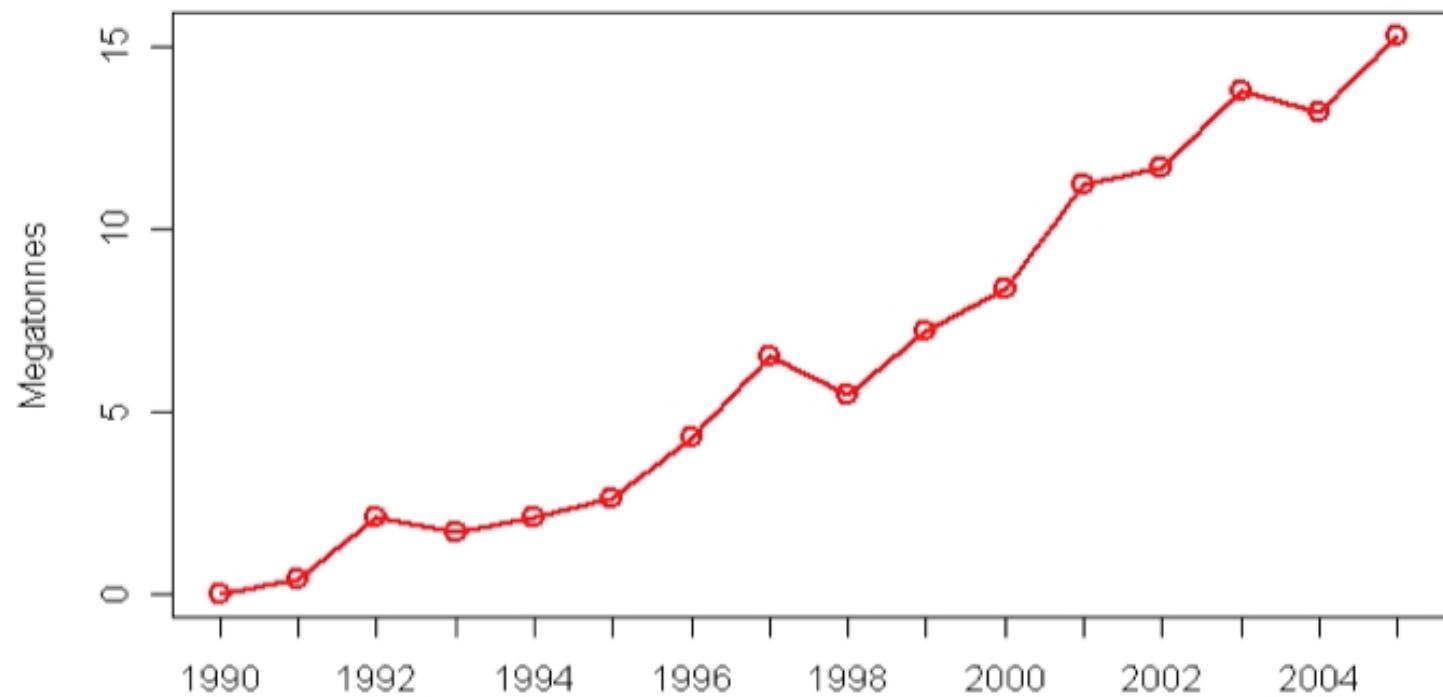
- Relationship between two continuous variables



Bivariate - Line chart

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

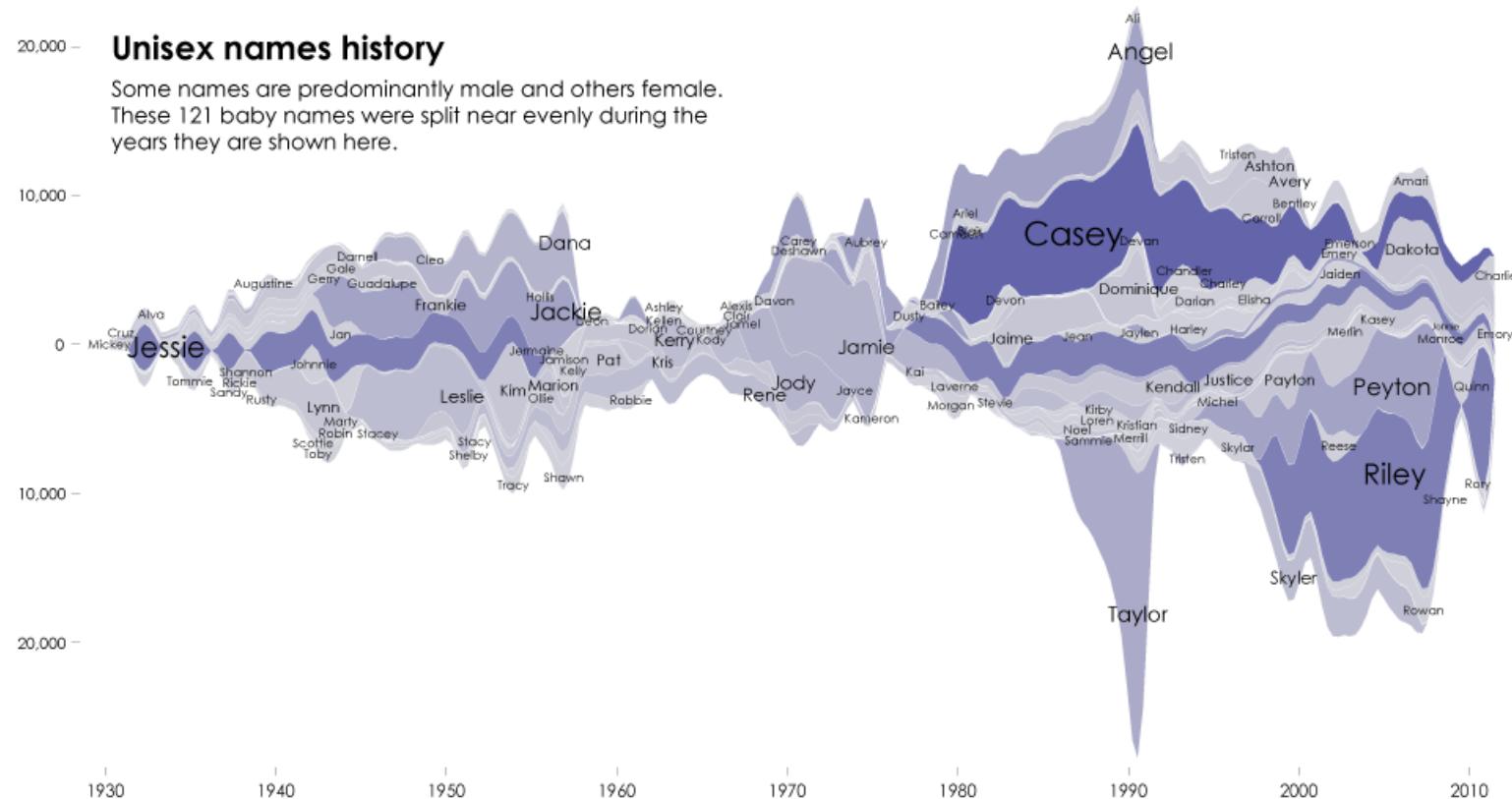
New Zealand's Total Greenhouse Gas Emissions from 1990 Base



Multivariate

20,000 – Unisex names history

Some names are predominantly male and others female. These 121 baby names were split near evenly during the years they are shown here.



Source: Social Security Administration | By: <http://flowingdata.com>

Multivariate

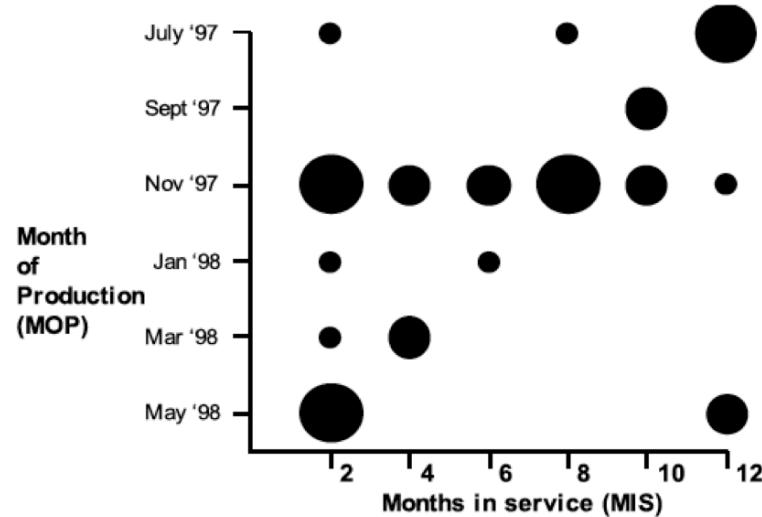
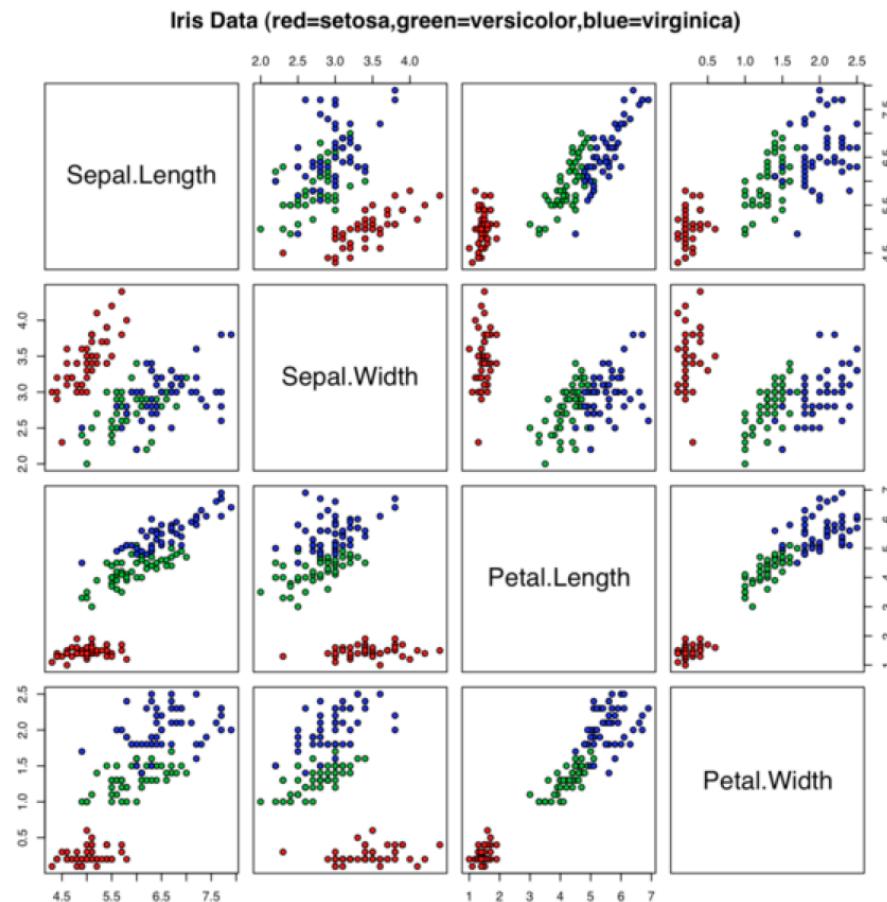
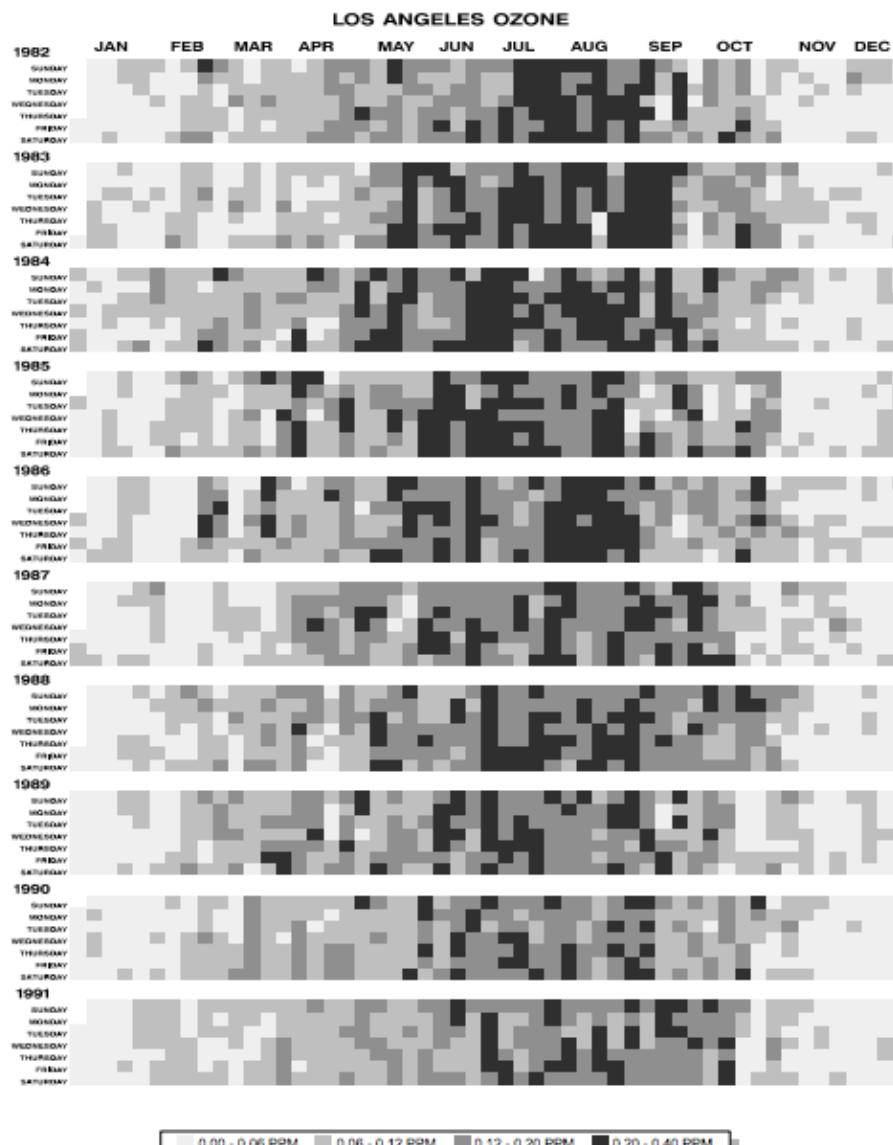


Figure 3.30 A representation of reported product failure, based on month of production (MOP) of the failed product, and total months in service (MIS) before the fault occurred. The radius of each circle indicates the number of faults reported for a given MOP and MIS

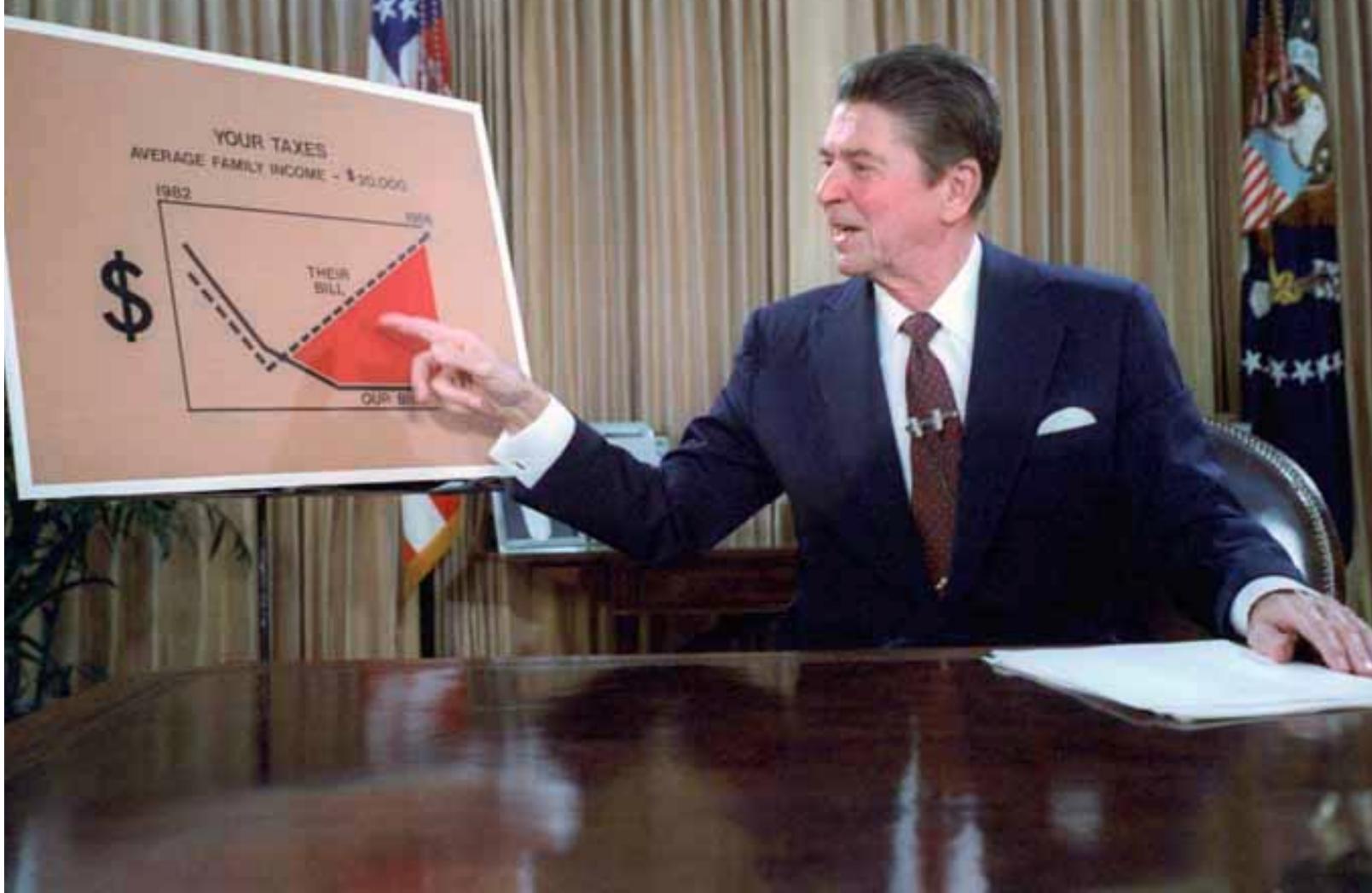
Multivariate



Multivariate









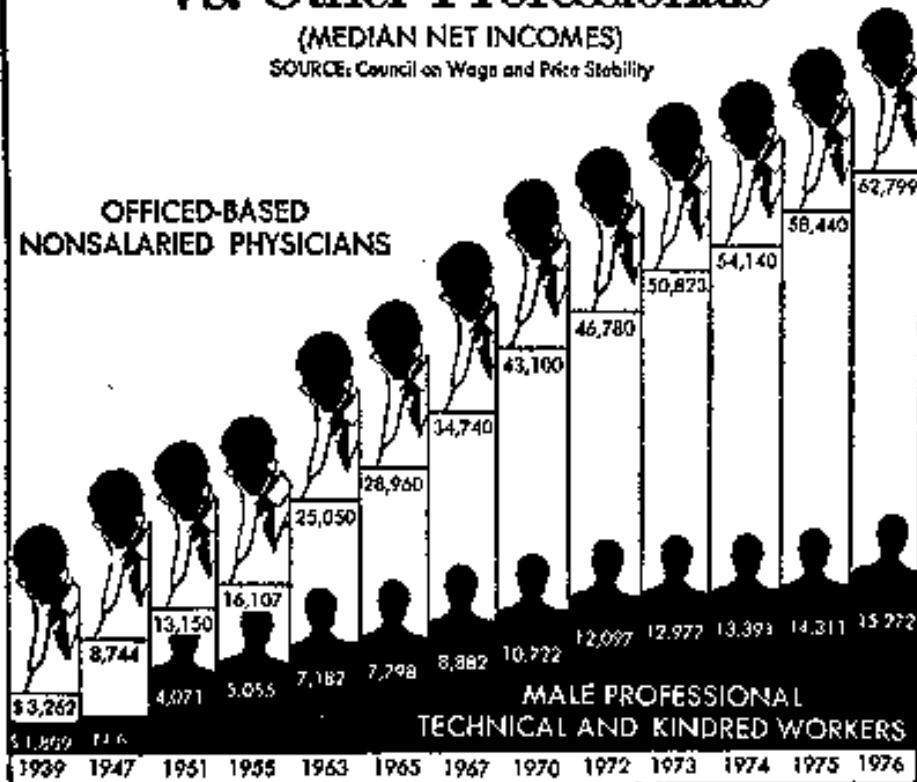
Guidelines

- 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

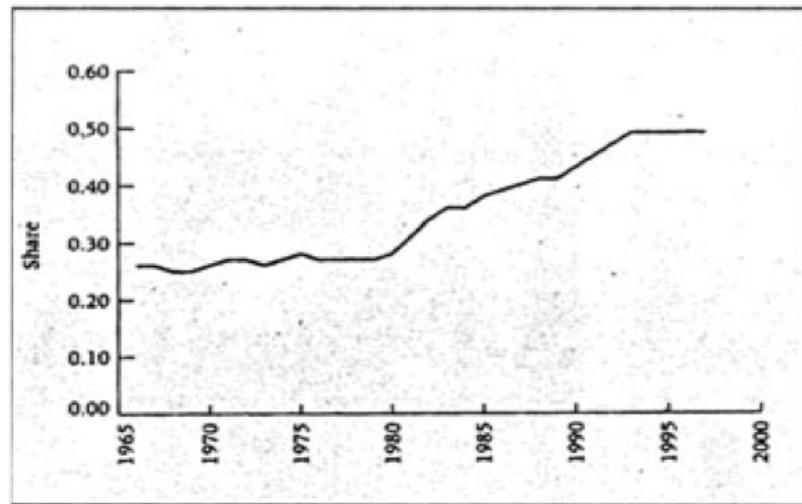
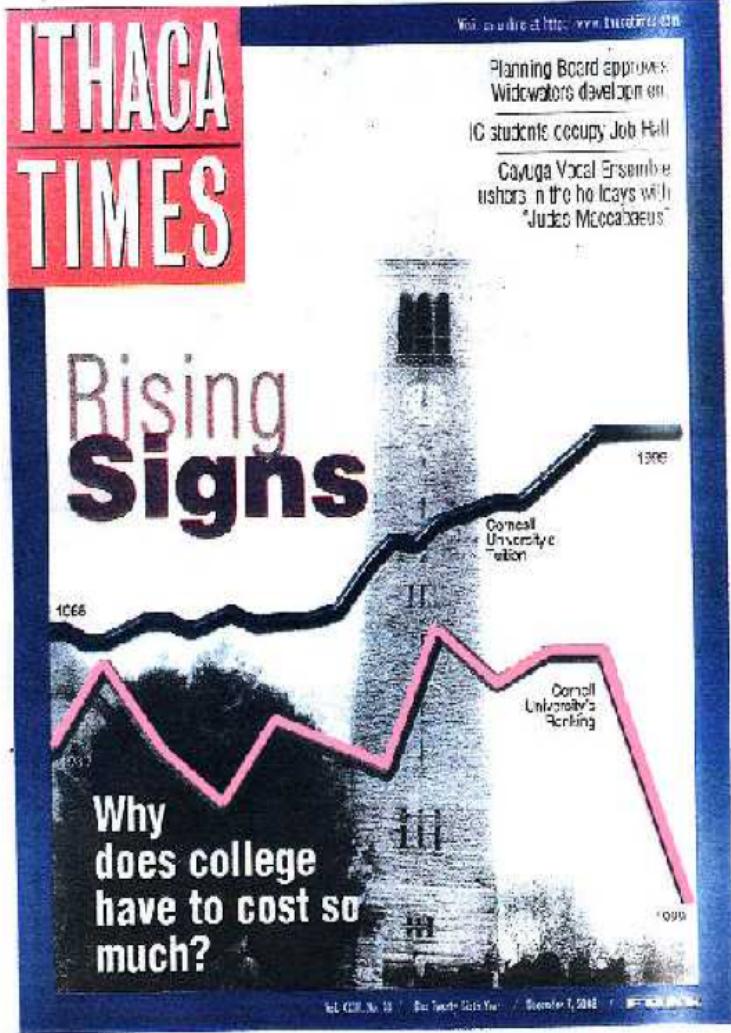
Incomes of Doctors Vs. Other Professionals

(MEDIAN NET INCOMES)

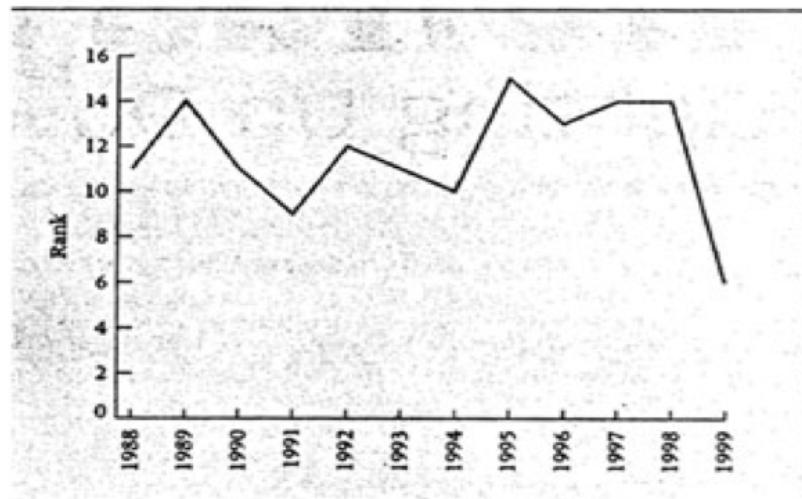
SOURCE: Council on Wage and Price Stability



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.



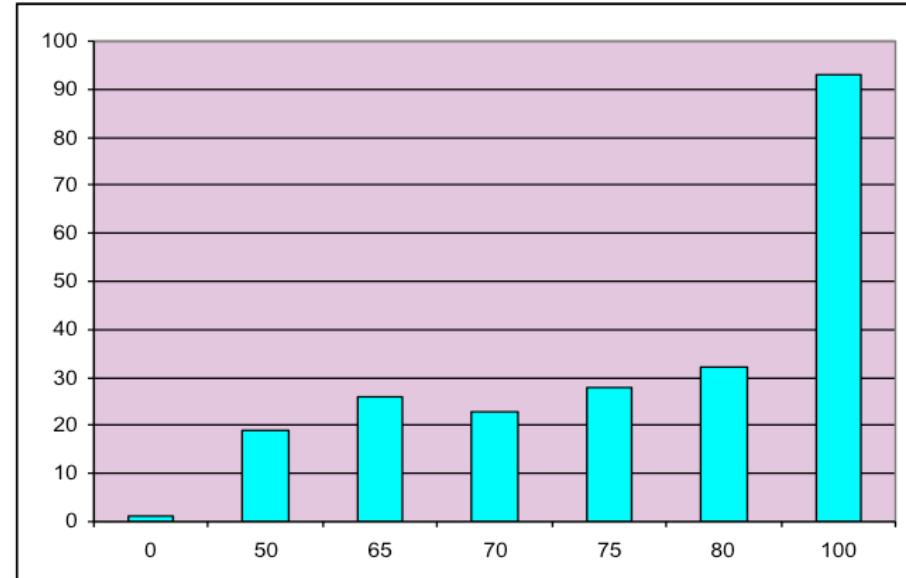
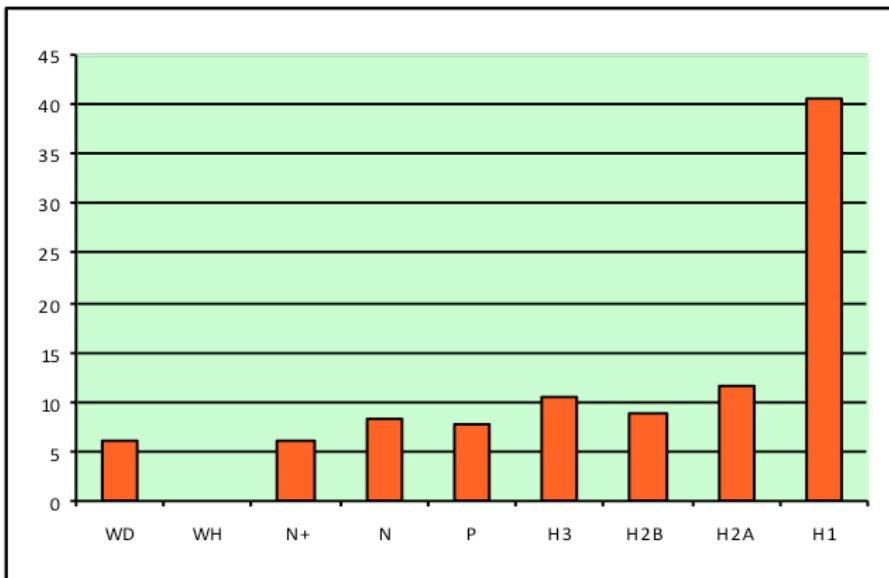
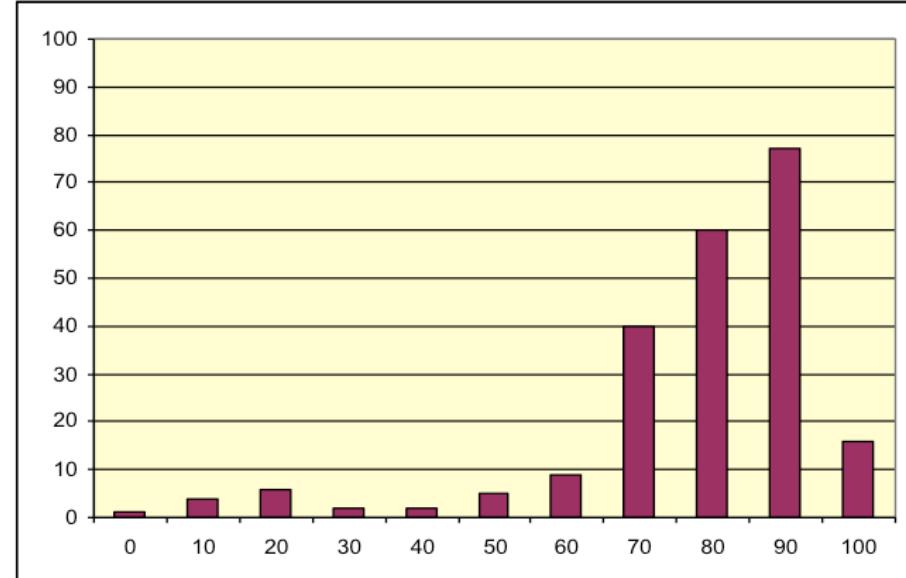
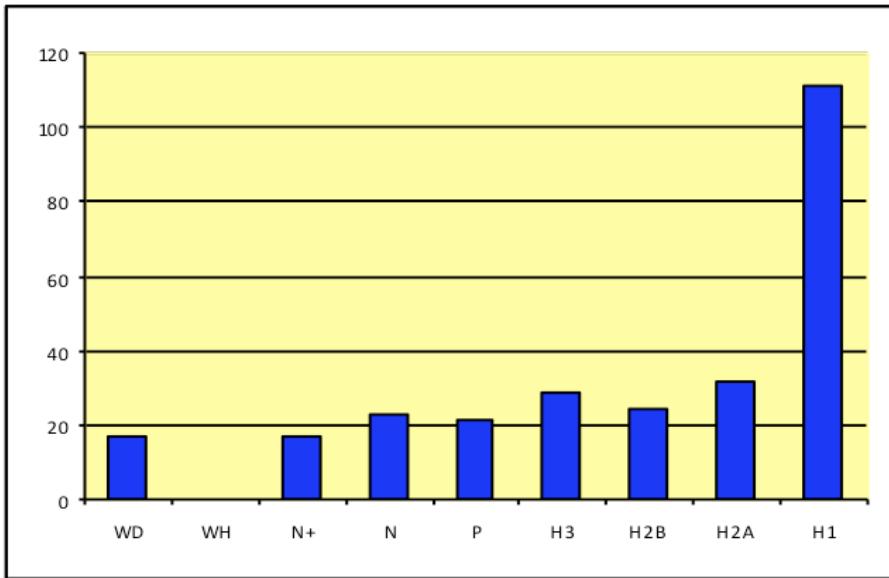
PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN US NEWS & WORLD REPORT HAS RISEN AND FALLEN ERRATICALLY.

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



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 (ddi)**
 - 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

Labor Productivity: U.S. vs Japan

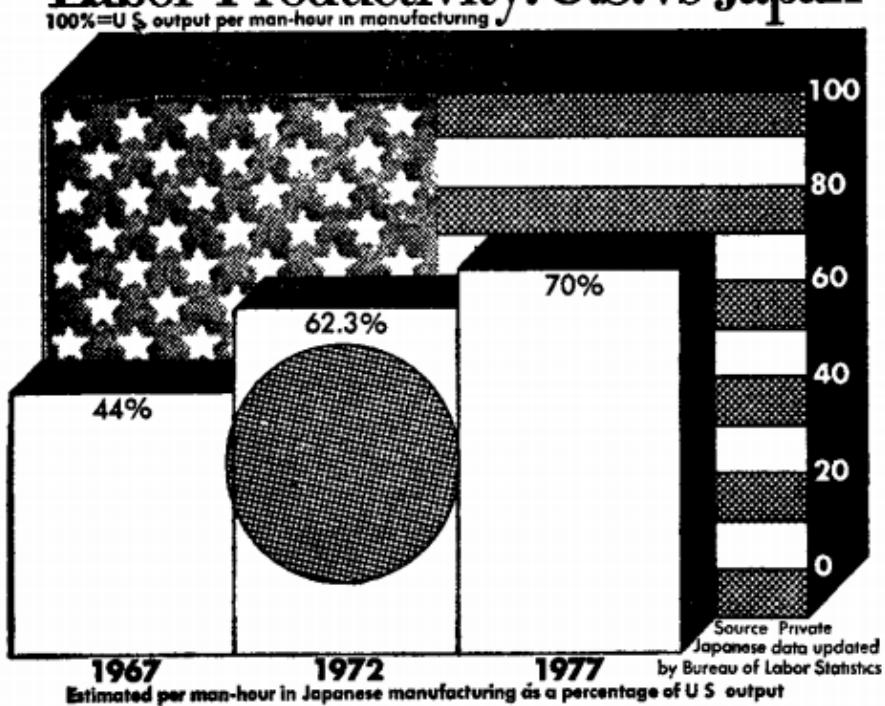
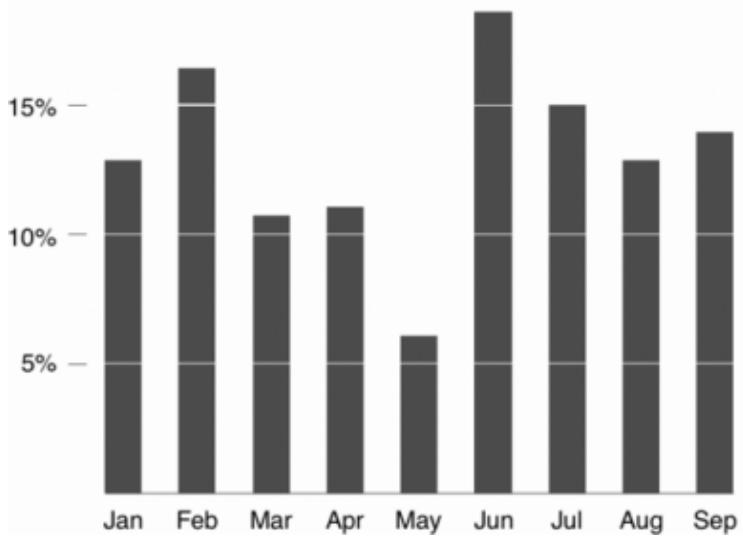
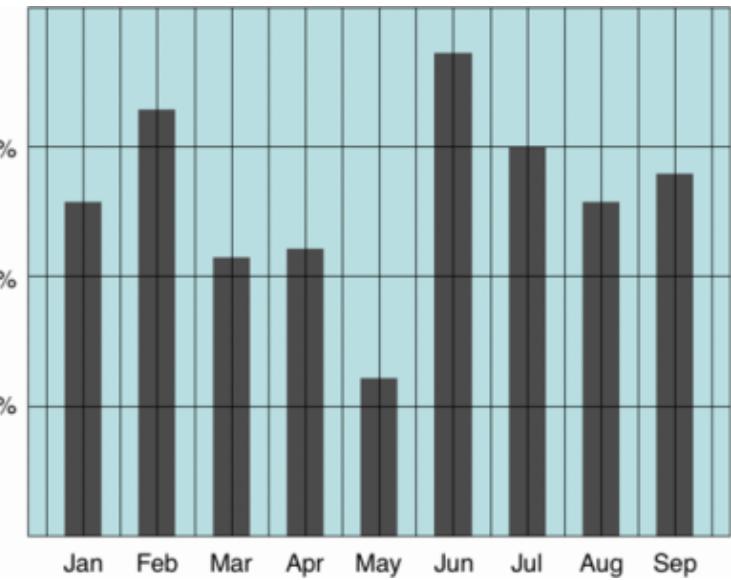
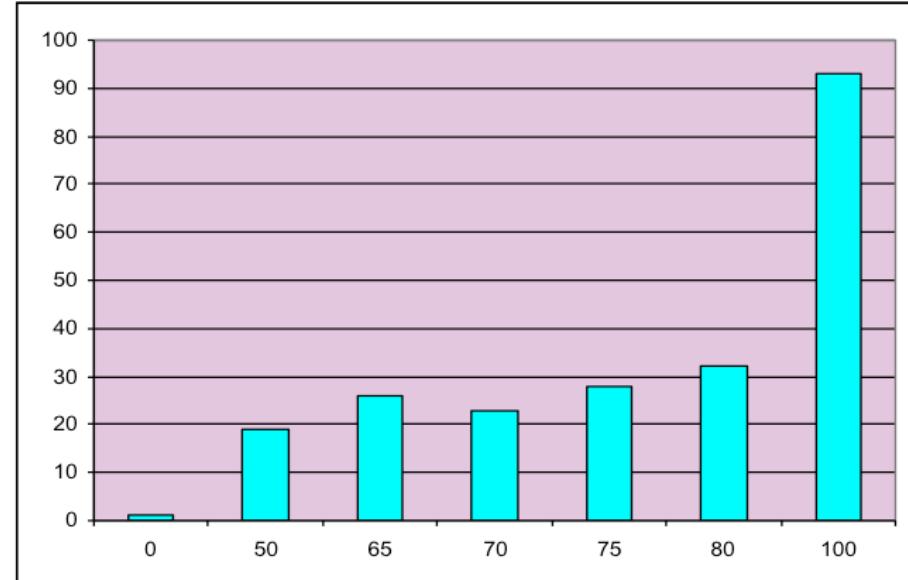
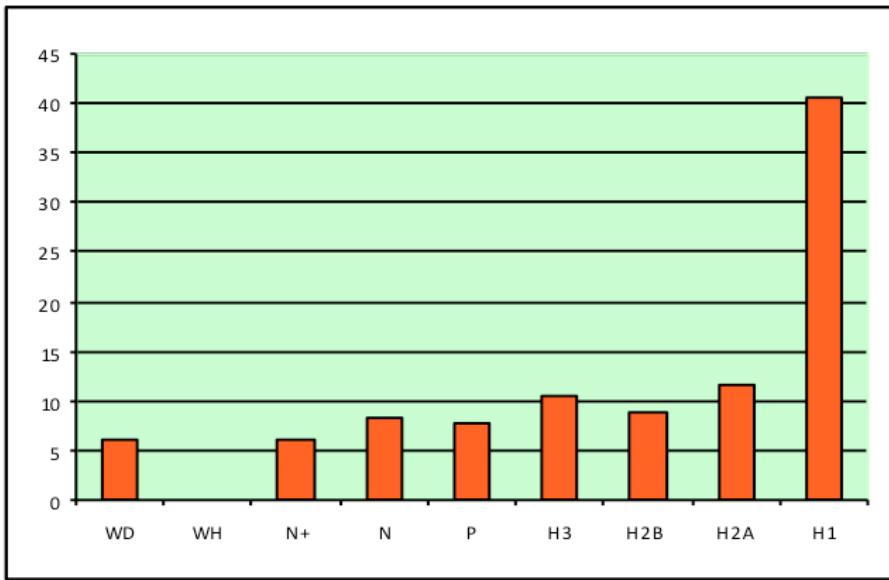
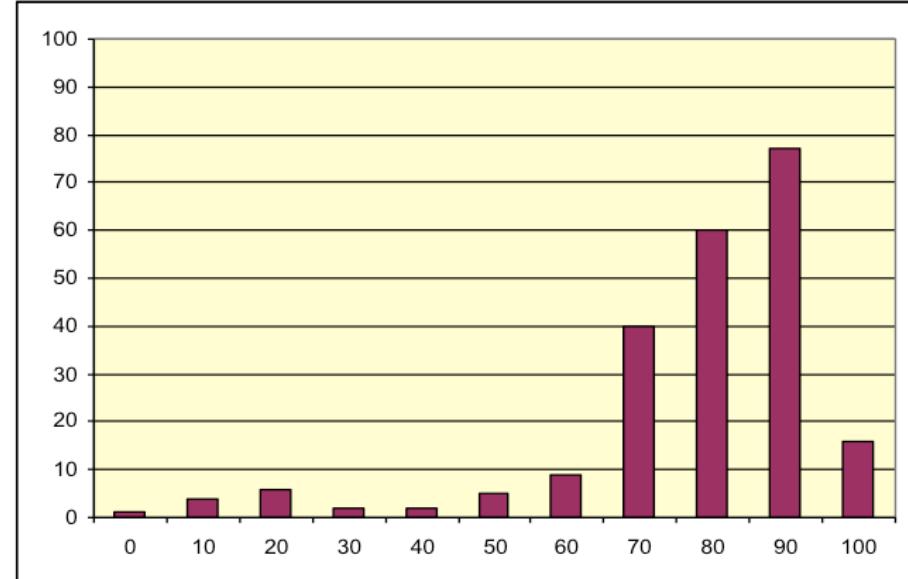
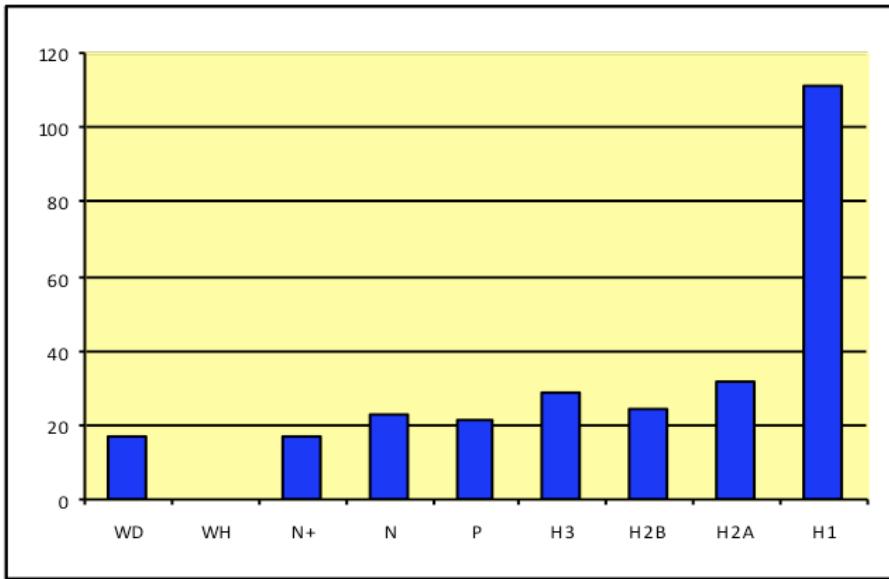


Figure 3 A low density graph (© 1978, The Washington Post) with chart-junk to fill in the space ($ddi = .2$)





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 *
```


using matplotlib - cgi

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:

```
savefig("plot.png", dpi=100)
print 'Content-Type: text/html'
print
print '<html><body>'
print ''
print '</body></html>'
```

using matplotlib - cgi

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:

```
savefig("plot.png", dpi=100)
print 'Content-Type: image/png'
print
print open("plot.png").read()
```