

Napoleon Invasion of Russia, 1812

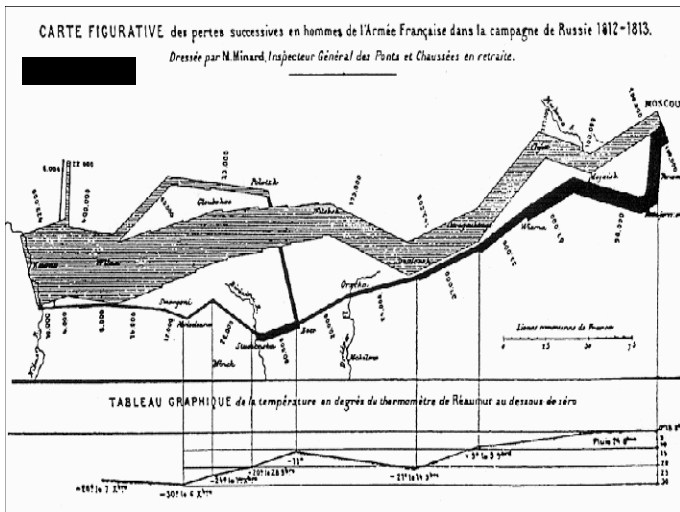
Visualization and Data Mining



Napoleon



2



2/19/2003 9:22:01 7:43 AM Page 1

Map representing the losses over time of French army troops during the Russian campaign, 1812-1813. Constructed by Charles Joseph Minard, Inspector General of Public Works retired. Paris, 20 November 1869

The number of men present at any given time is represented by the width of the grey line; one mm. indicates ten thousand men. Figures are also written besides the lines. Grey designates men moving into Russia; black, for those leaving. Sources for the data are the works of messrs. Thiers, Ségur, Fezensac, Chambray and the unpublished diary of Jacob, who became an Army Pharmacist on 29 October. In order to visualize the army's losses more clearly, I have drawn this as if the units under prince Jerome and Marshall Davoust (temporarily separated from the main body to go to Minsk, and Mikhilov, which then joined up with the main army again) had stayed with the army throughout.

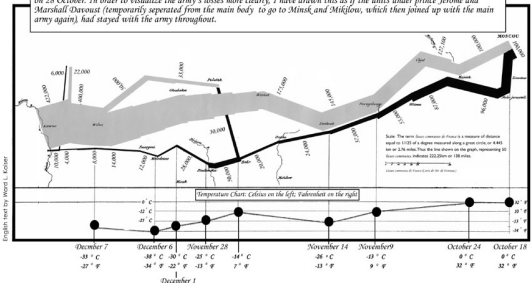
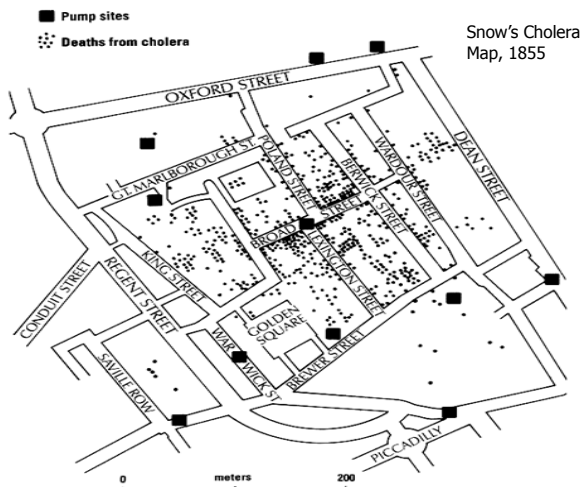
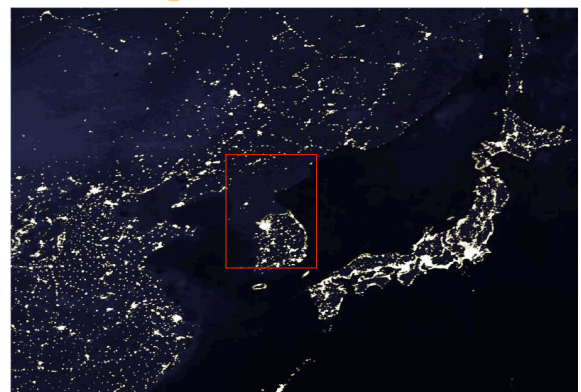


Figure 58. Minard's map of Napoleon's Russian campaign. This graphic has been translated from French to English and modified to most effectively display the temperature data.

4



Asia at night



6

South and North Korea at night

North Korea
Notice how dark
it is

Seoul,
South Korea



7

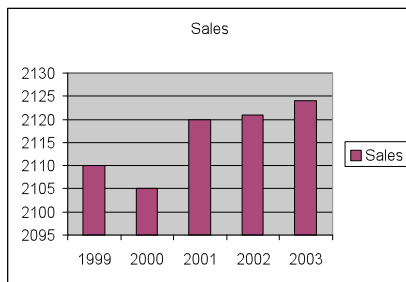
Visualization Role

- Support interactive exploration
- Help in result presentation
- Disadvantage: requires human eyes
- Can be misleading

8

Bad Visualization: Spreadsheet with misleading Y-axis

Year	Sales
1999	2110
2000	2105
2001	2120
2002	2121
2003	2124

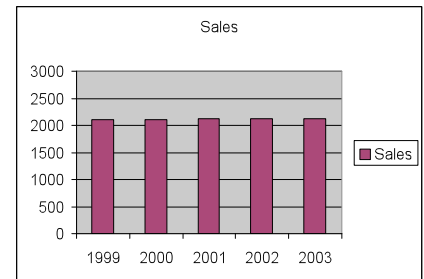


Y-Axis scale gives **WRONG**
impression of big change

9

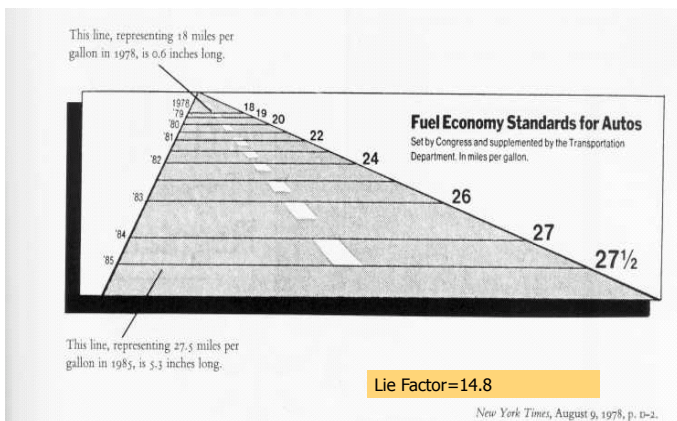
Better Visualization

Year	Sales
1999	2110
2000	2105
2001	2120
2002	2121
2003	2124



Axis from 0 to 2000 scale gives
correct impression of small change

10



(E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition)

11

Lie Factor

$$\text{Lie Factor} = \frac{\text{size of eff esthown in graphic}}{\text{size of eff ean data}} =$$

$$= \frac{(5.3 - 0.6)}{(27.5 - 18.0)} = \frac{7.833}{0.528} = 14.8$$

Tufte requirement: $0.95 < \text{Lie Factor} < 1.05$

(E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition)

12

Tufte's Principles of Graphical Excellence

- Give the viewer
 - the greatest number of ideas
 - in the shortest time
 - with the least ink in the smallest space.

- Tell the truth about the data!

(E.R. Tufte, "The Visual Display of Quantitative Information", 2nd edition)

13

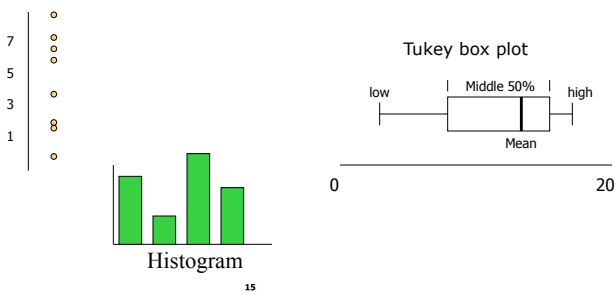
Visualization Methods

- Visualizing in 1-D, 2-D and 3-D
 - well-known visualization methods
- Visualizing more dimensions
 - Parallel Coordinates
 - Other ideas

14

1-D (Univariate) Data

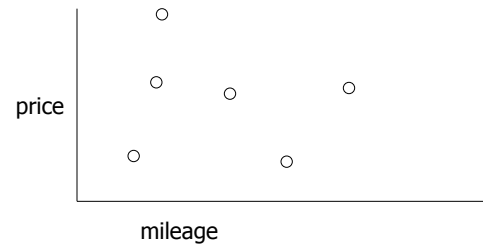
- Representations



15

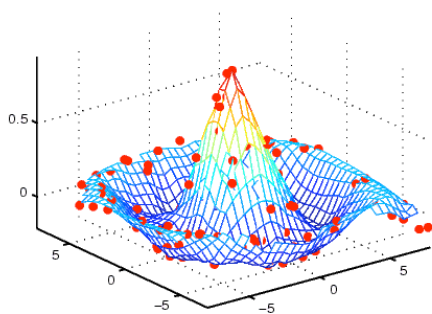
2-D (Bivariate) Data

- Scatter plot, ...



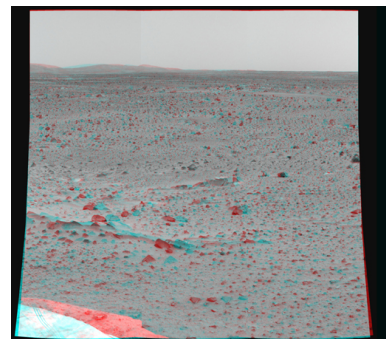
16

3-D Data (projection)



17

3-D image (requires 3-D blue and red glasses)



Taken by Mars Rover Spirit, Jan 2004

18

Visualizing in 4+ Dimensions

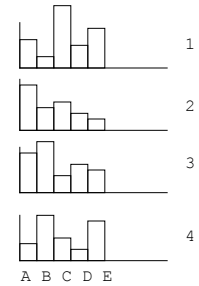
- Scatterplots
- Parallel Coordinates
- Chernoff faces
- ...

19

Multiple Views

Give each variable its own display

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5



Problem: does not show correlations

20

Scatterplot Matrix

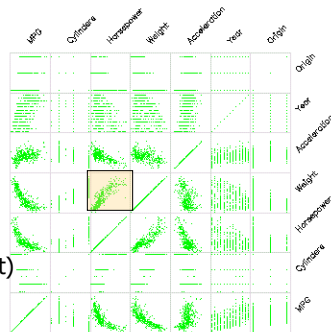
Represent each possible pair of variables in their own 2-D scatterplot (car data)

Q: Useful for what?

A: linear correlations (e.g. horsepower & weight)

Q: Misses what?

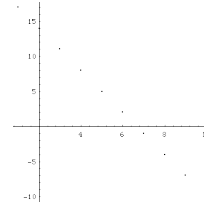
A: multivariate effects



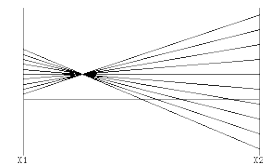
21

Parallel Coordinates

- Encode variables along a horizontal row
- Vertical line specifies values



Dataset in a Cartesian coordinates



Same dataset in parallel coordinates

Invented by Alfred Inselberg while at IBM, 1985



22

Example: Visualizing Iris Data



Iris setosa

sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8



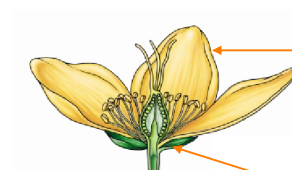
Iris versicolor



Iris virginica

23

Flower Parts

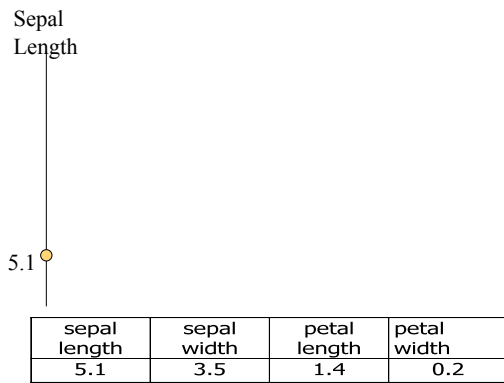


Petal, a non-reproductive part of the flower

Sepal, a non-reproductive part of the flower

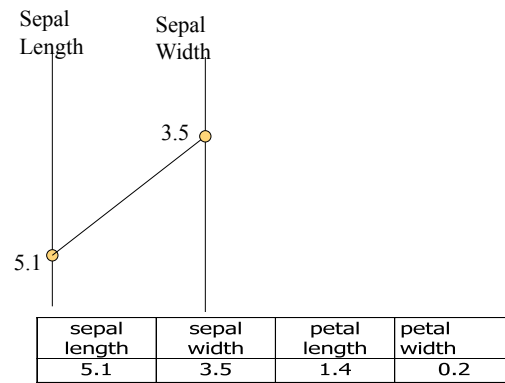
24

Parallel Coordinates



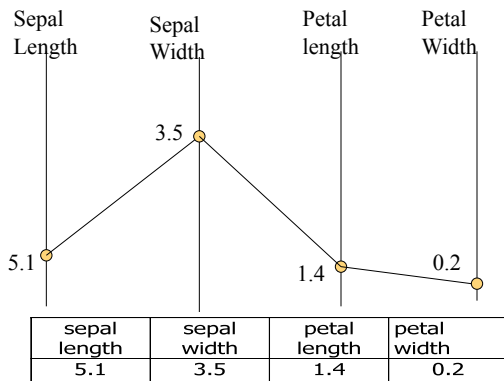
25

Parallel Coordinates: 2 D



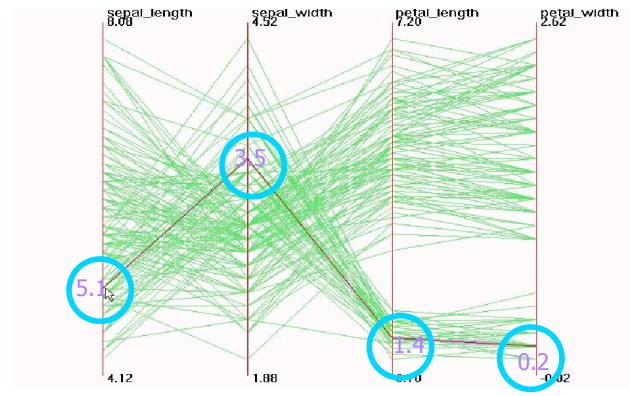
26

Parallel Coordinates: 4 D



27

Parallel Visualization of Iris data



28

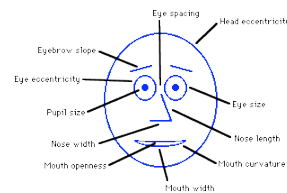
Parallel Visualization Summary

- Each data point is a line
- Similar points correspond to similar lines
- Lines crossing over correspond to negatively correlated attributes
- Interactive exploration and clustering
- Problems: order of axes, limit to ~20 dimensions

29

Chernoff Faces

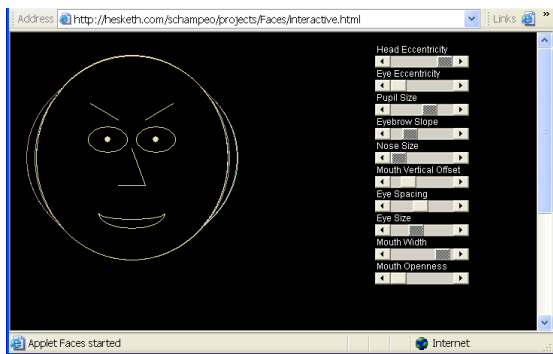
Encode different variables' values in characteristics of human face



Cute applets: <http://www.cs.uchicago.edu/~wiseman/chernoff/>
<http://hesketh.com/schampeon/projects/Faces/chernoff.html>

30

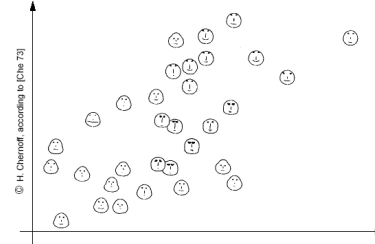
Interactive Face



31

Chernoff faces, example

Chernoff-Faces [Che 73, Tuf 83]



32

Visualization Summary

- Many methods
- Visualization is possible in more than 3-D
- Aim for graphical excellence

33