Visualization

UW CSE 190p

Summer 2012

BARE BONES VISUALIZATION IN PYTHON WITH MATPLOTLIB

matplotlib

- A major design limitation is that it stives to emulate MATLAB
  - More on this in the next lecture
- One important function for HW6:

```
plot(xvalues, yvalues)
```

Plot

```
import matplotlib.pyplot as plt

xs = [1,2,3,4,5]
ys = [x**2 for x in xs]

plt.plot(xs, ys)
```

no return value?

- We are operating on a “hidden” variable representing the figure.
- This is a terrible, terrible trick.
- Its only purpose is to pander to MATLAB users.
- I’ll show you how this works in the next lecture

```
import matplotlib.pyplot as plt

xs = range(-100,100,10)
x2 = [x**2 for x in xs]
negx2 = [-x**2 for x in xs]

plt.plot(xs, x2)
plt.plot(xs, negx2)
plt.xlabel("x")
plt.ylabel("y")
plt.ylim(-2000, 2000)
plt.axhline(0) # horiz line
plt.axvline(0) # vert line
plt.savefig("quad.png")
plt.show()
```

Incrementally modify the figure.

Save your figure to a file

Show it on the screen

```
def myplot(xs, ys, description):
  plt.plot(xs, ys, linewidth=2, color='green', linestyle='-', marker='s', label=description)

def setup_plot():
  plt.xlabel("x")
  plt.ylabel("y")
  plt.axhline(0,linestyle=':',color='red')
  plt.axvline(0,linestyle=':',color='red')

def finish_plot():
  plt.legend()
  plt.show()

setup_plot()
myplot(xs,x2,"x**2")
finish_plot()

setup_plot()
myplot(xs,negx2,"-x**2")
finish_plot()
```

We can group these options into functions as usual, but remember that they are operating on a global, hidden variable

Review

WHY VISUALIZE DATA?

Bill Howe, eScience Institute

Location of deaths in the 1854 London Cholera Epidemic.
X marks the locations of the water pumps

Dr. John Snow

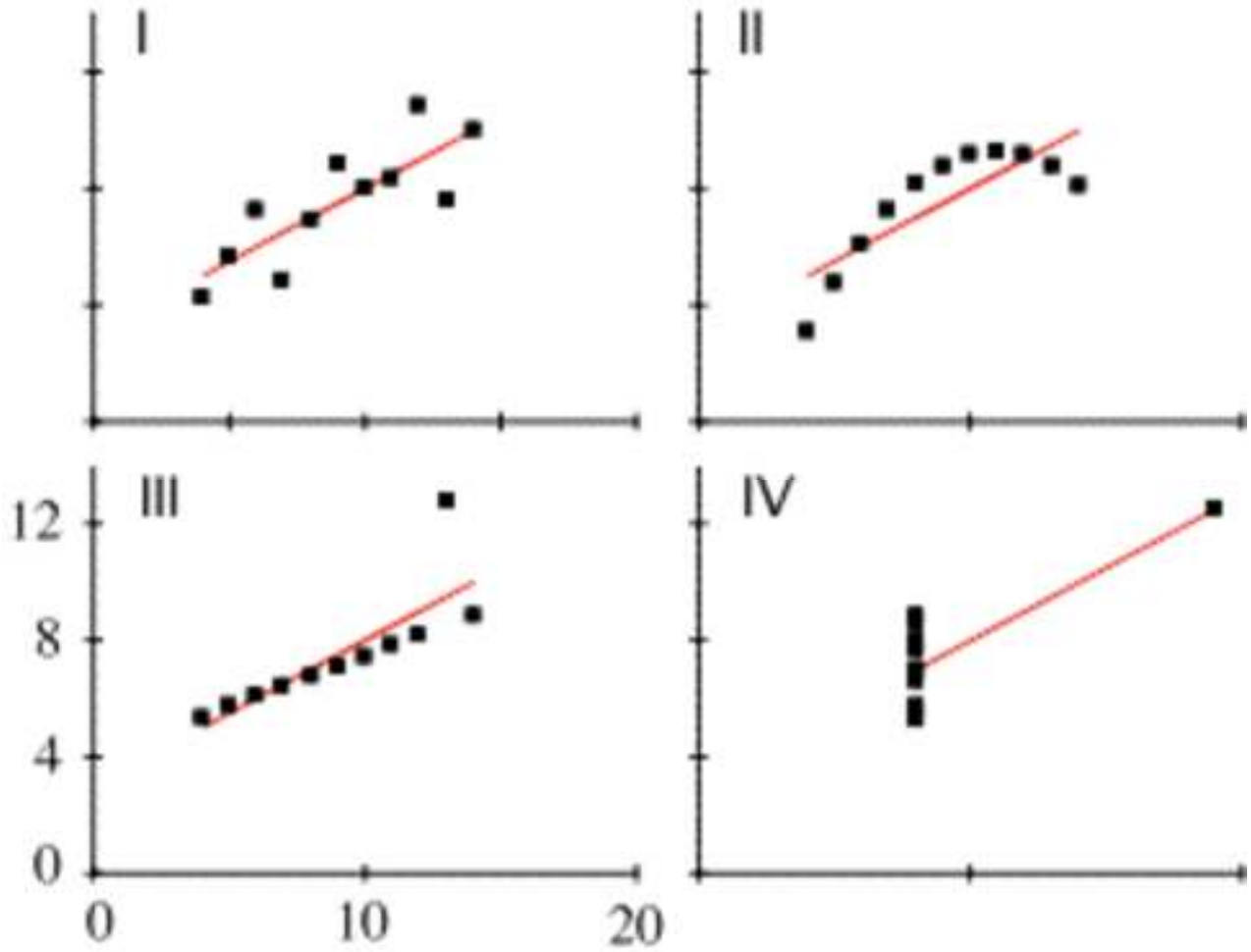
Anscombe's Quartet

I		II		III		IV	
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

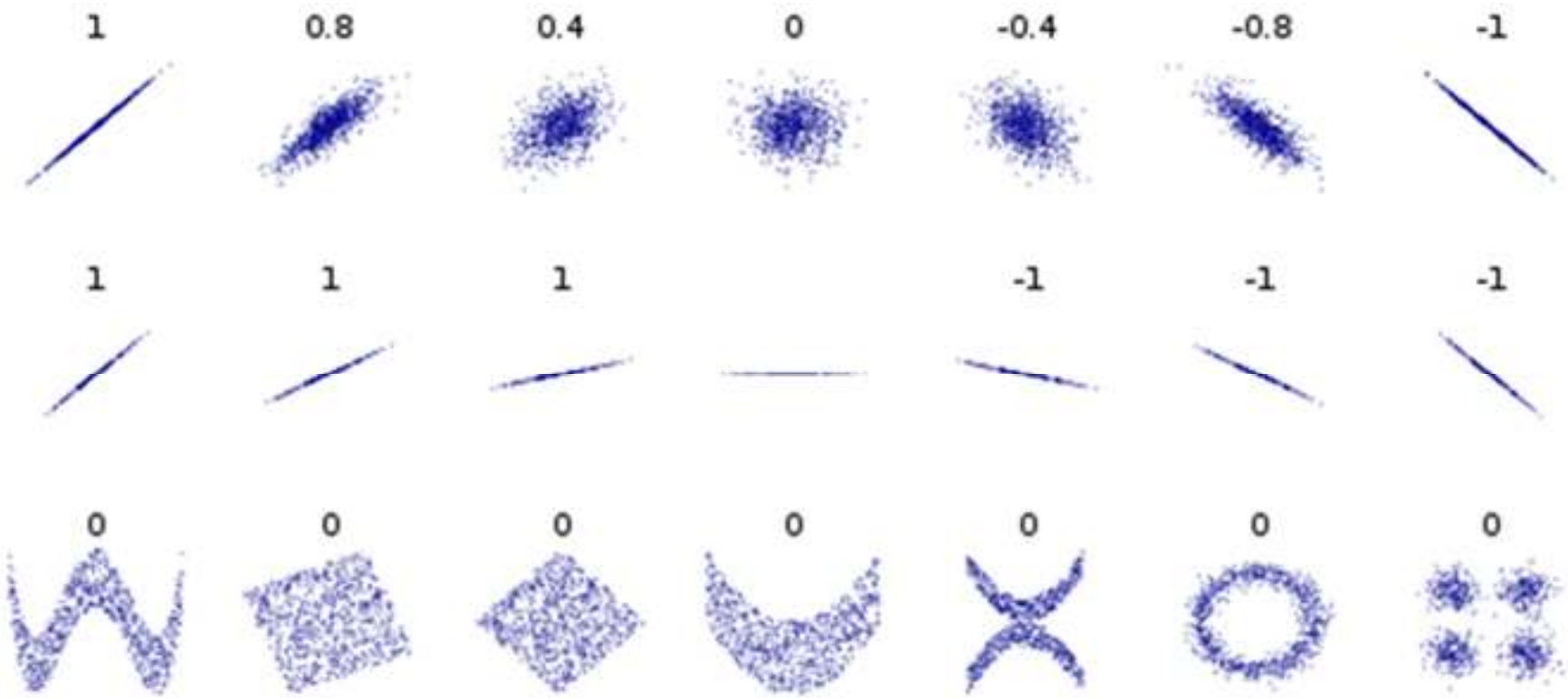
Anscombe's Quartet (2)

- mean of the x values = 9.0
- mean of the y values = 7.5
- equation of the least-squared regression line:
  $y = 3 + 0.5x$
- sums of squared errors (about the mean) = 110.0
- regression sums of squared errors
  (variance accounted for by x) = 27.5
- residual sums of squared errors
  (about the regression line) = 13.75
- correlation coefficient = 0.82
- coefficient of determination = 0.67

Anscombe's Quartet (3)



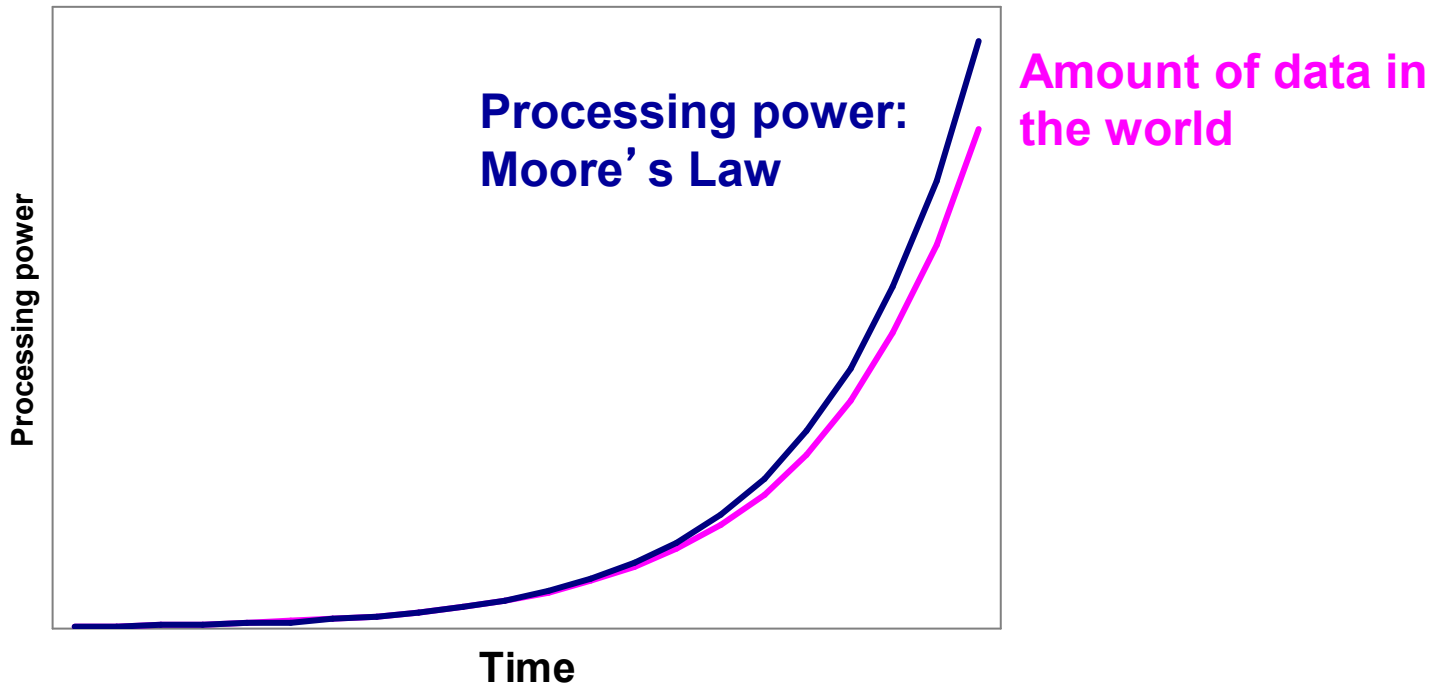
Another example: Pearson Correlation



Other reasons?

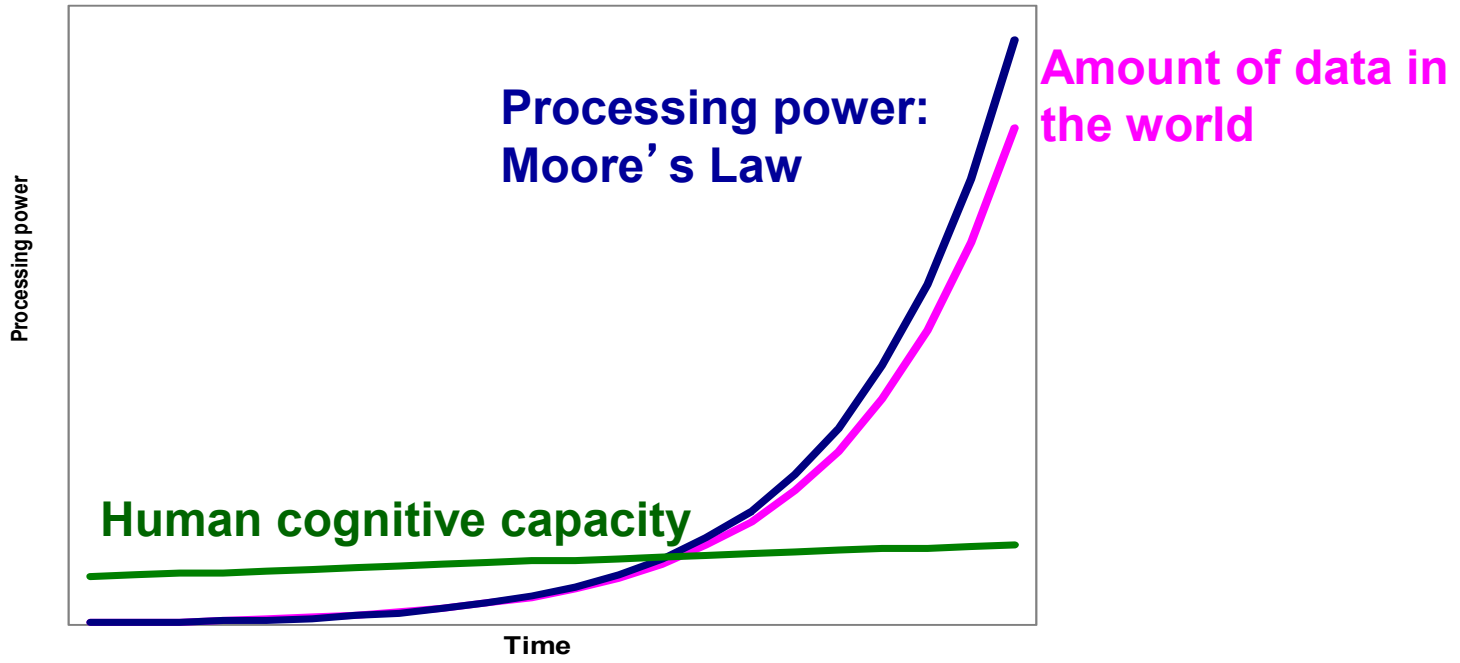
- Visualization is the highest bandwidth channel into the human brain [Palmer 99]
- The visual cortex is the largest system in the human brain; it's wasteful not to make use of it.
- As data volumes grow, visualization becomes a necessity rather than a luxury.
  - "A picture is worth a thousand words"

What is the rate-limiting step in data understanding?



slide src: Cecilia Aragon, UW HCDE

What is the rate-limiting step in data understanding?



Idea adapted from "Less is More" by Bill Buxton (2001)

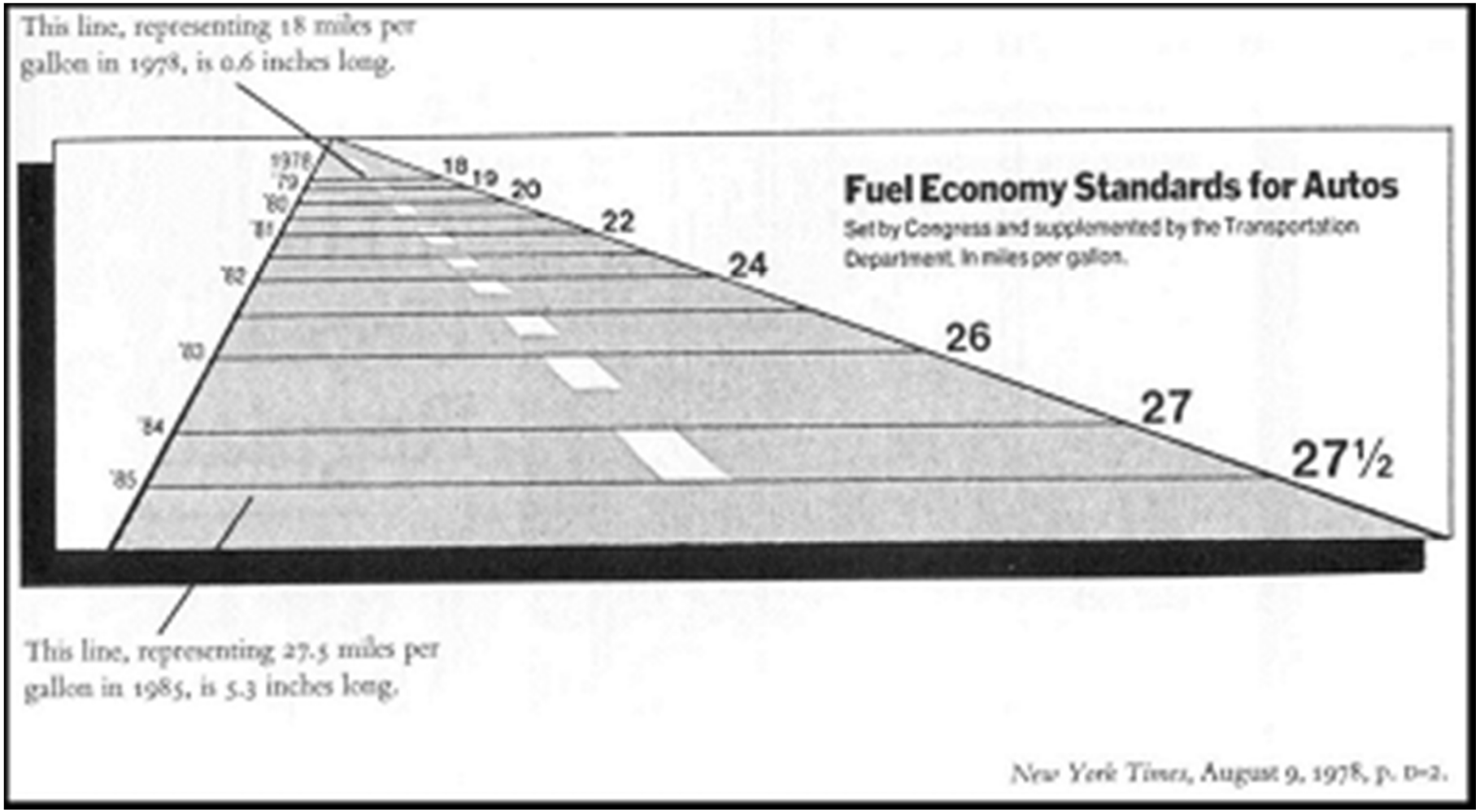
slide src: Cecilia Aragon, UW HCDE

What makes a good visualization?

Edward Tufte: Minimize the Lie Factor

$$\text{Lie Factor} = \frac{\text{Size of effect in the visualization}}{\text{Size of effect in the data}}$$

Example



Tufte 1997

What makes a good visualization?

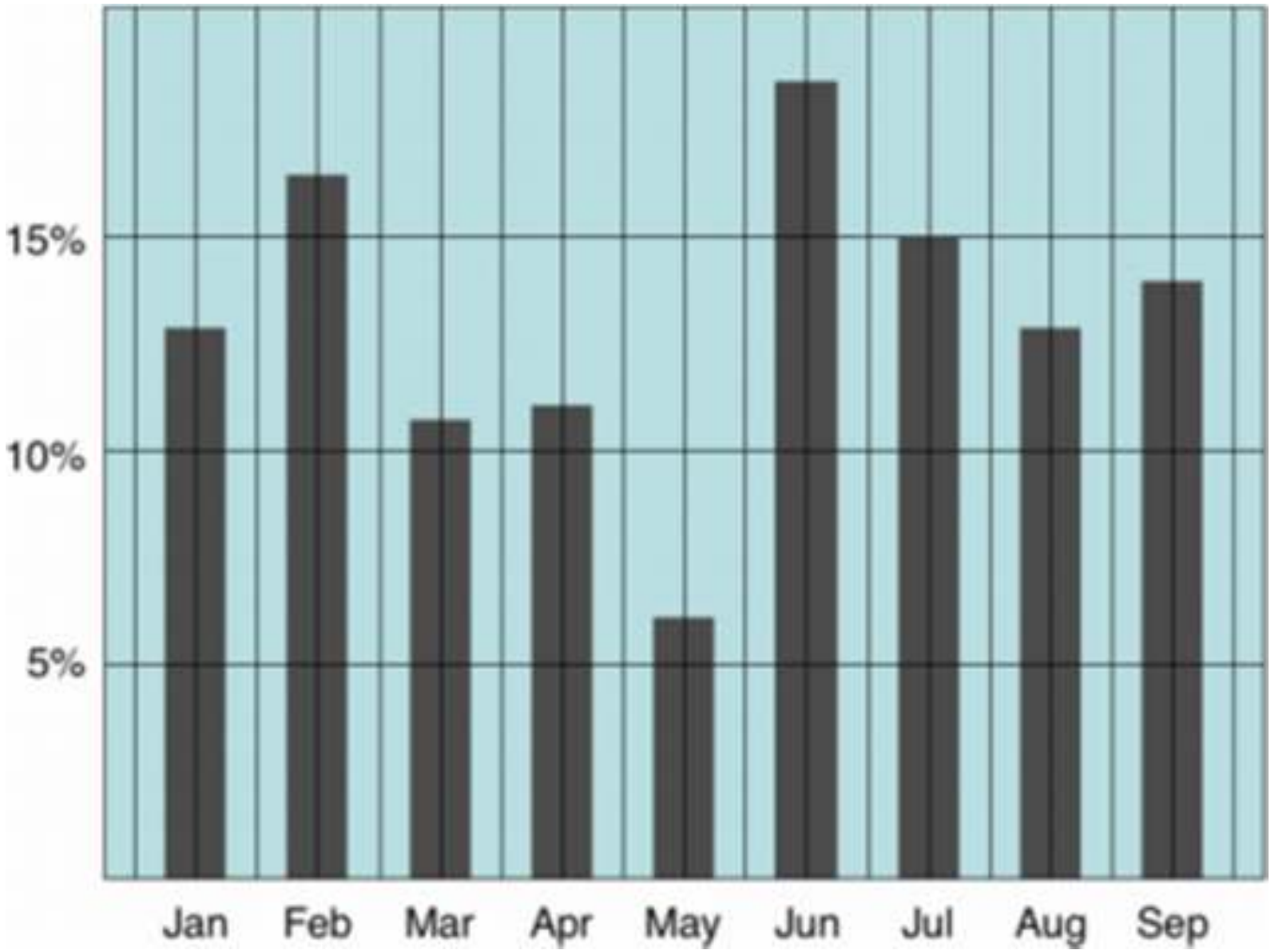
- Edward Tufte: Maximize the data-ink ratio

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

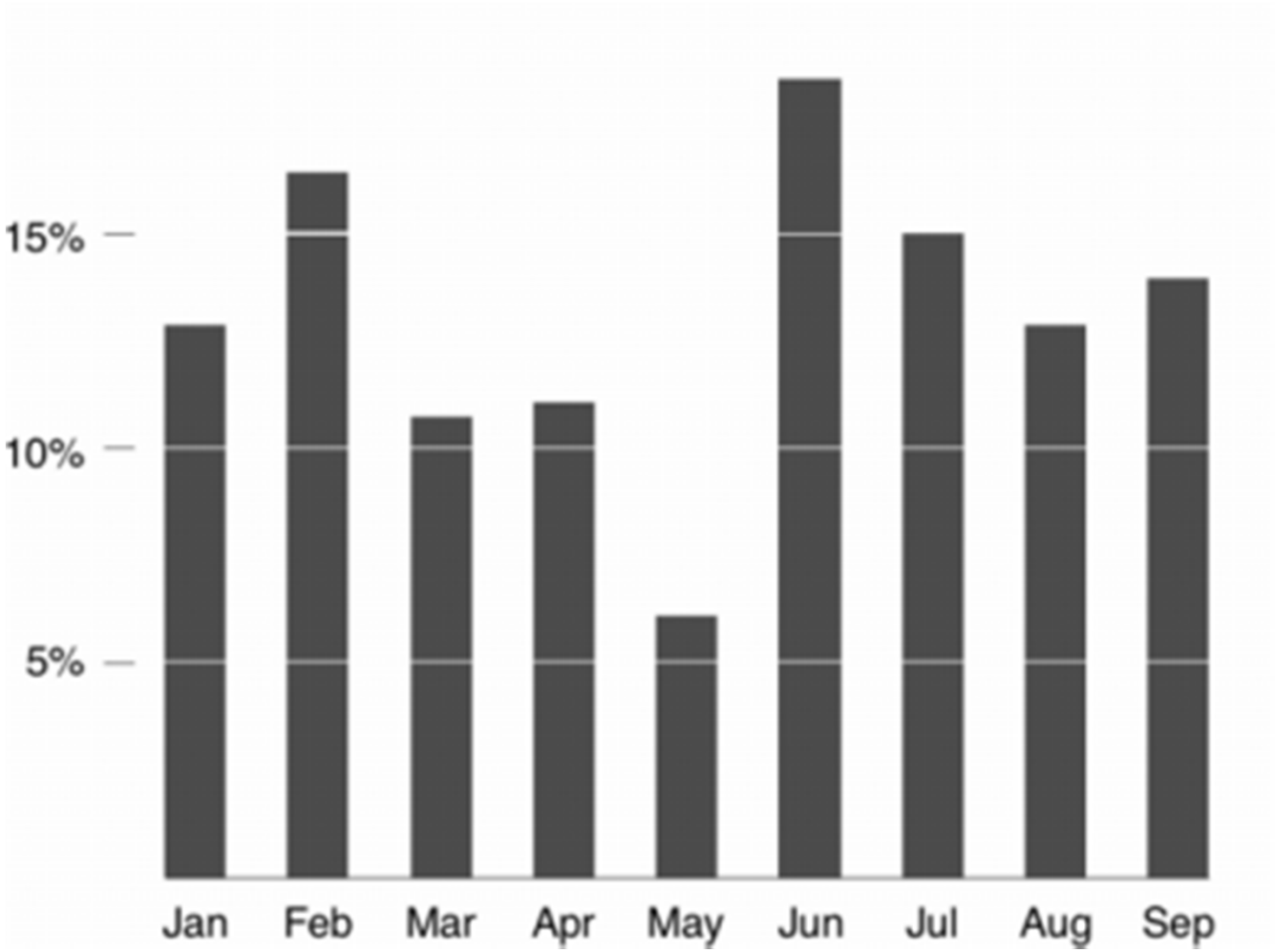
= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased

Example: High or Low Data Ink ratio?



Example: High or Low Data Ink ratio?



MONSTROUS COSTS

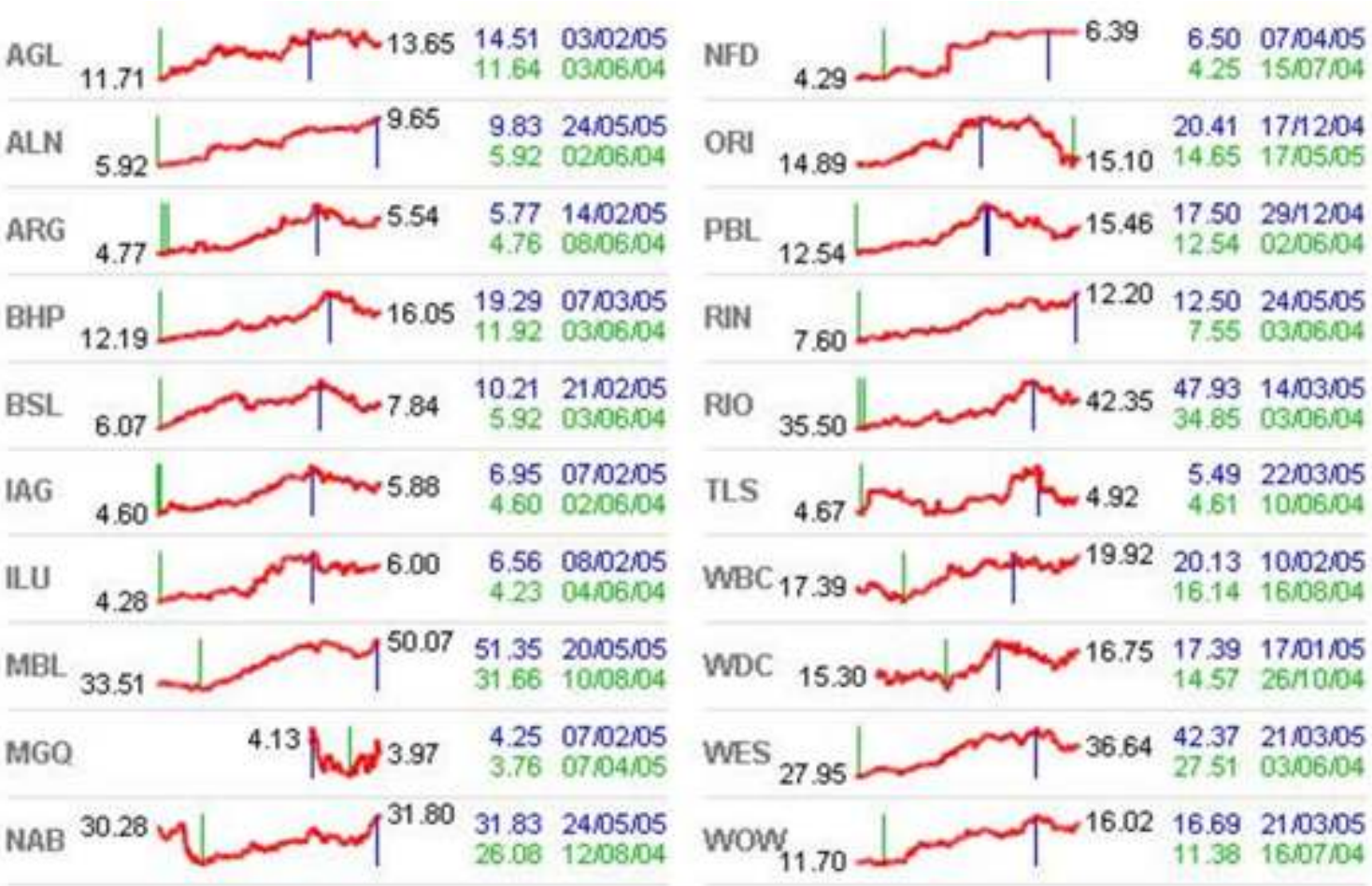
Total House and Senate campaign expenditures, in millions

Bateman et al: The Effects of Visual Embellishment on Comprehension and Memorability of Charts

- There was no significant difference between plain and image charts for interactive interpretation accuracy (i.e., when the charts were visible).
- There was also no significant difference in recall accuracy after a five-minute gap.
- After a long-term gap (2-3 weeks), recall of both the chart topic and the details (categories and trend) was significantly better for Holmes charts.
- Participants saw value messages in the Holmes charts significantly more often than in the plain charts.
- Participants found the Holmes charts more attractive, most enjoyed them, and found that they were easiest and fastest to remember.

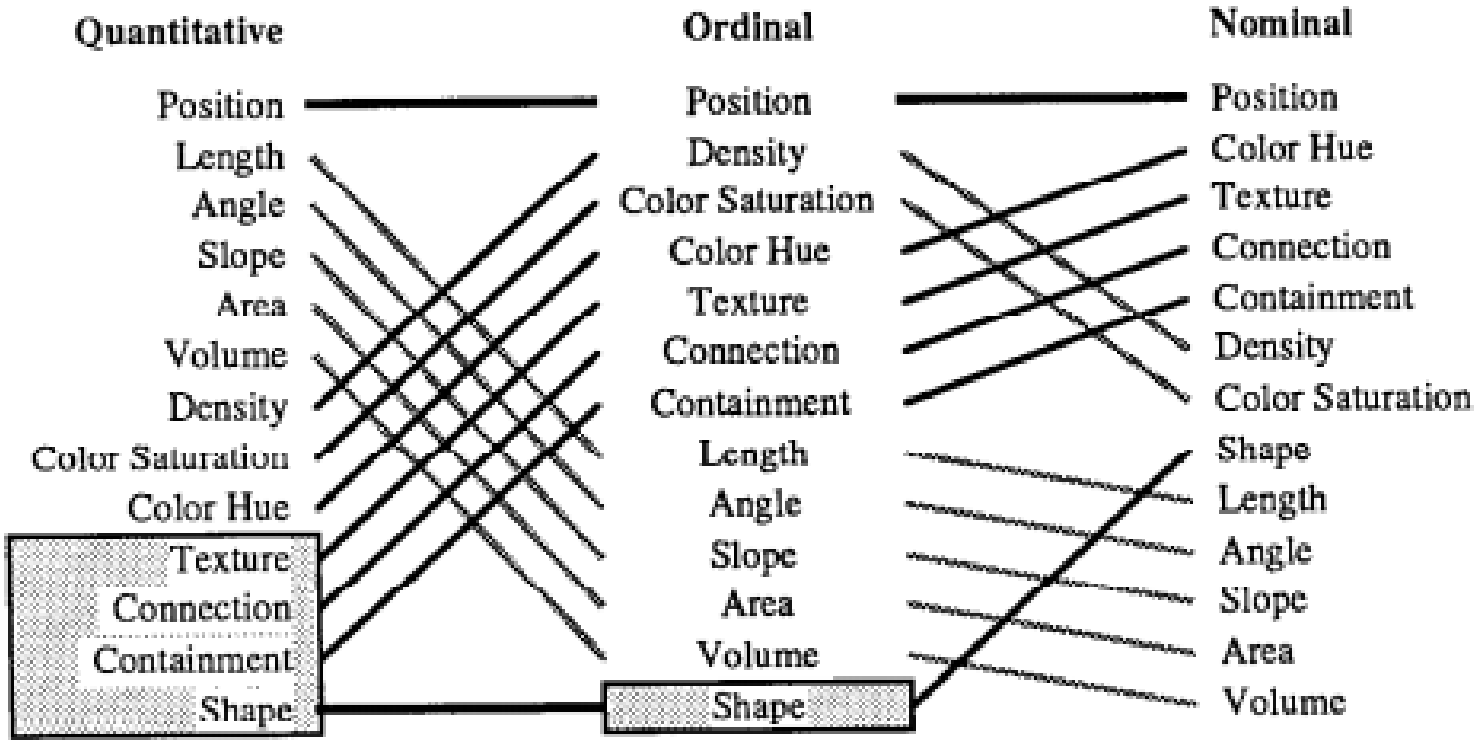
What makes a good visualization?

- Edward Tufte: Small multiples



What makes a good visualization?

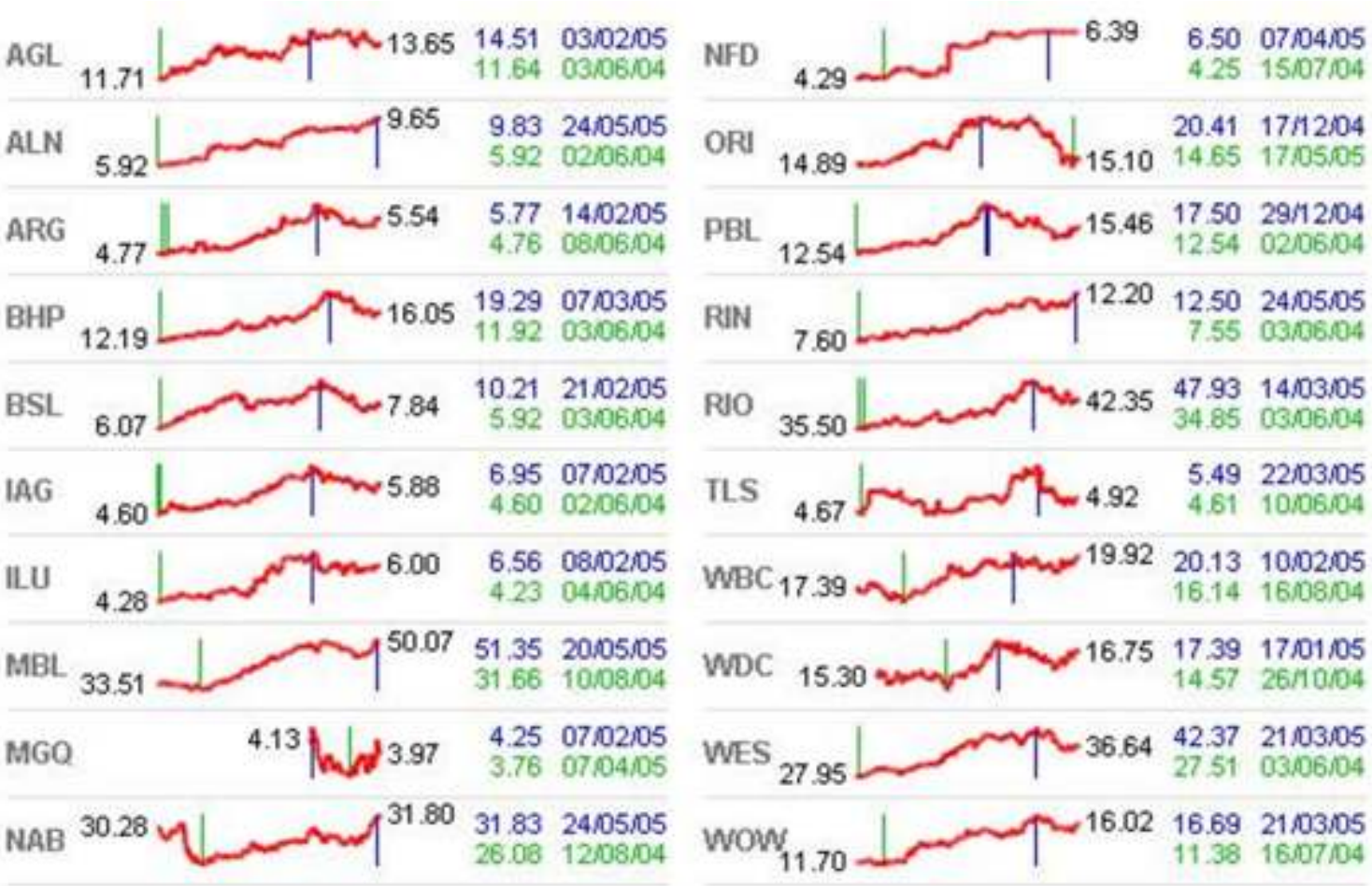
- Jock Mackinlay: Use the appropriate visual element for the relationship and data being analyzed



Conjectured rank effectiveness of each visualization method by data type

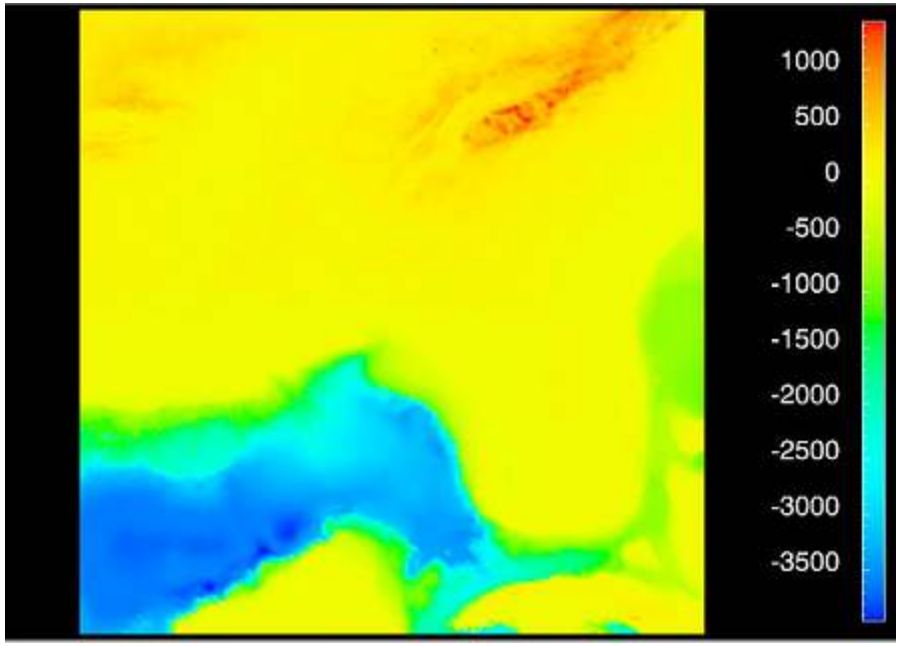
What makes a good visualization?

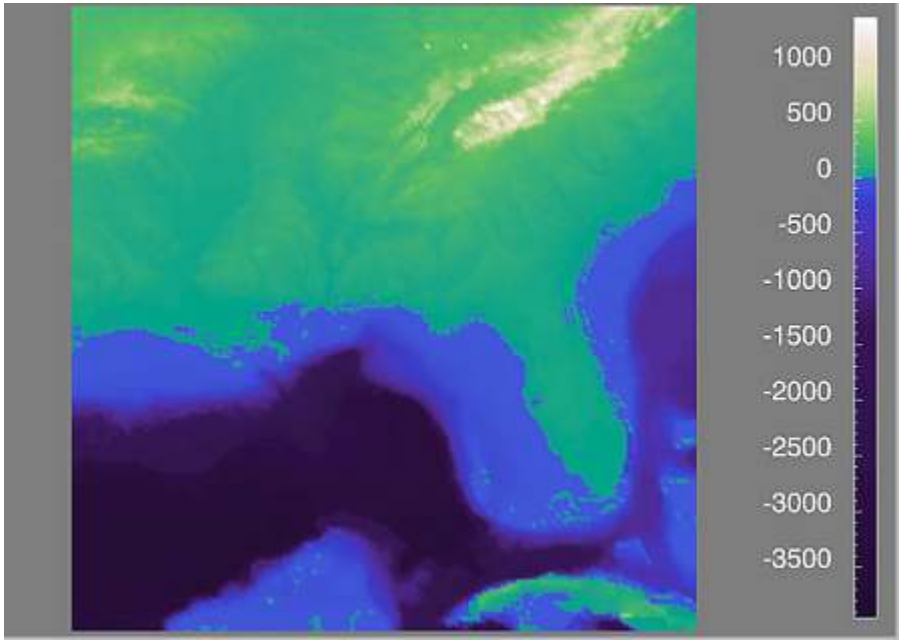
Tufte again: Small multiples



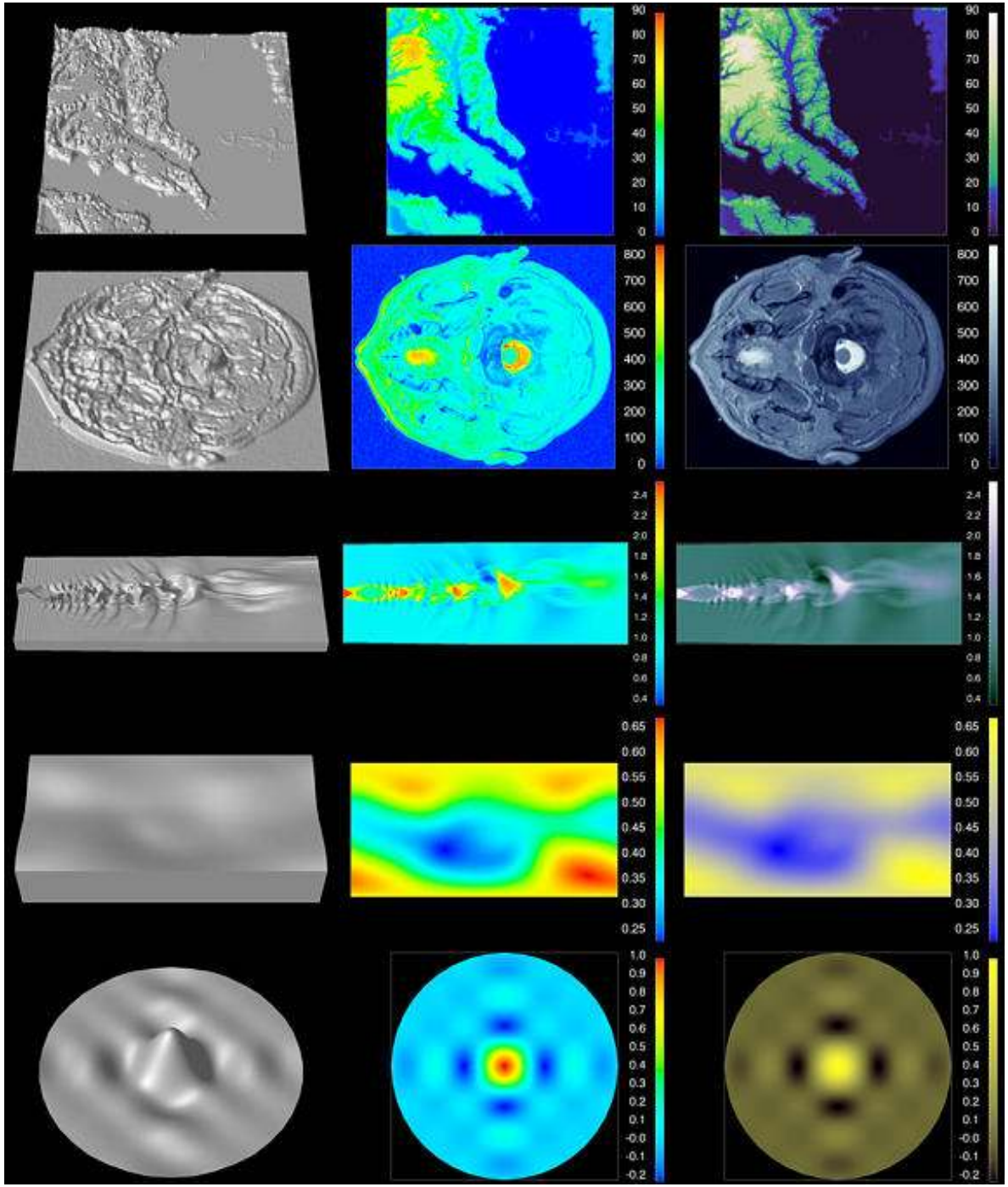
What makes a good visualization?

Lloyd Treinish: Color Matters



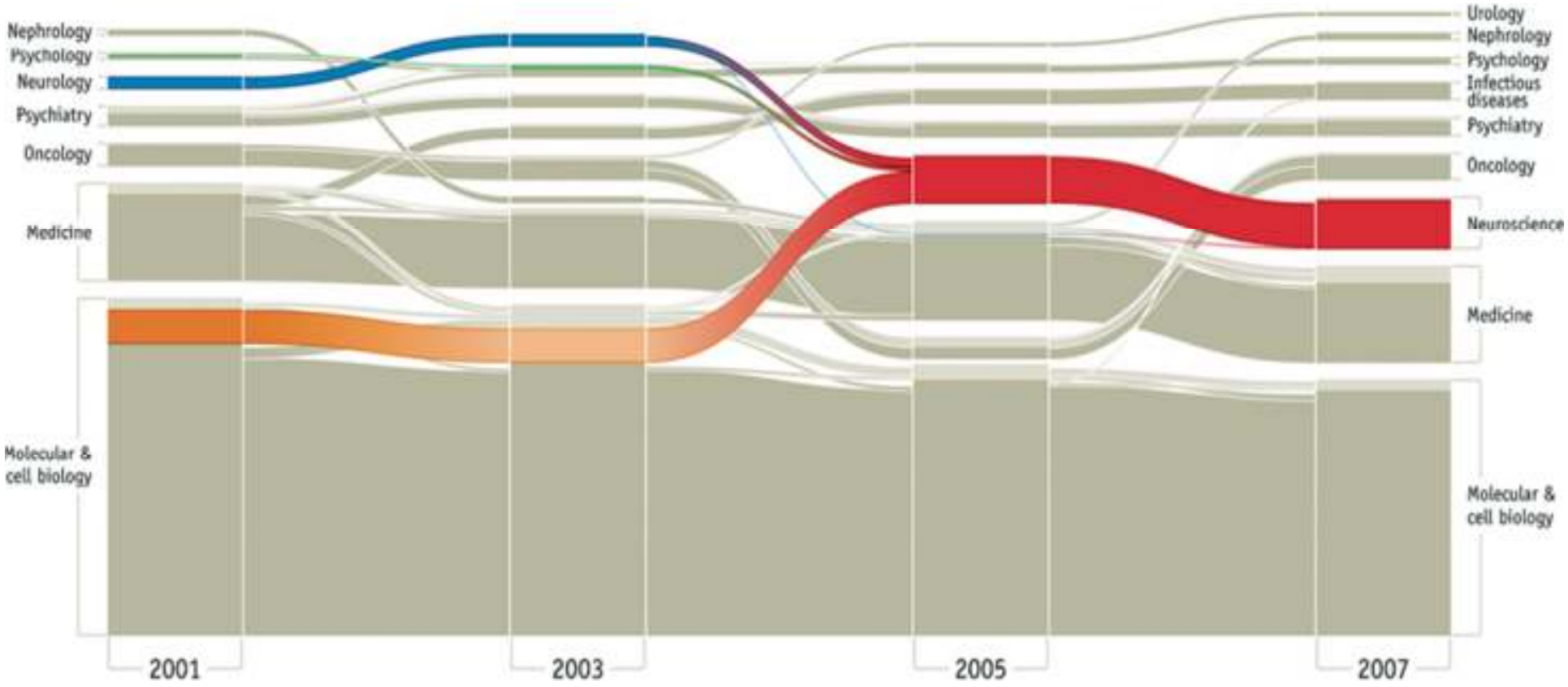


Lloyd Treinish, IBM Research, http://www.research.ibm.com/people/l/lloydt/



Lloyd Treinish, IBM Research, http://www.research.ibm.com/people/l/lloydt/

A Nice Example



Bergstrom, Rosvall, 2011