Knowledge & Visual Analytics

Lecture 1

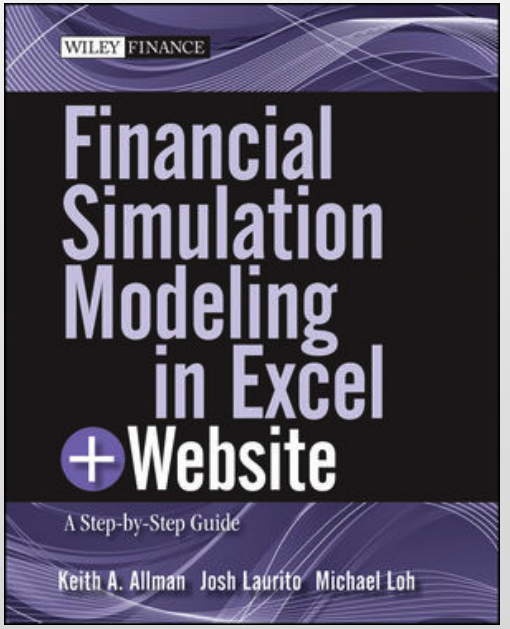
Josh Laurito

Who Am I?

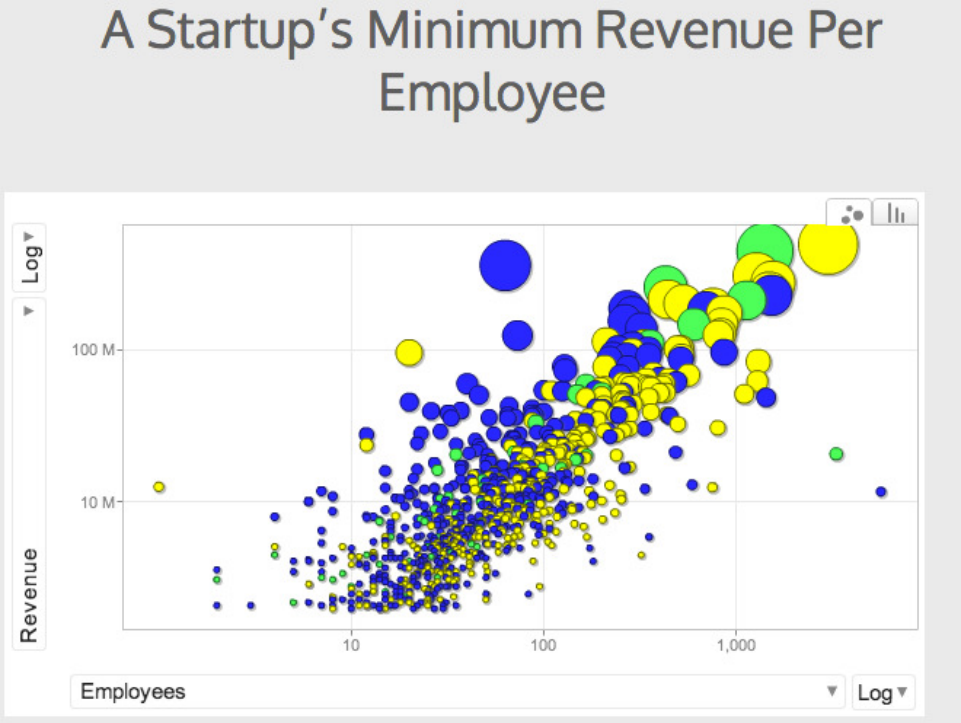
CROWDTWIST
GIVE MORE. GET MORE.

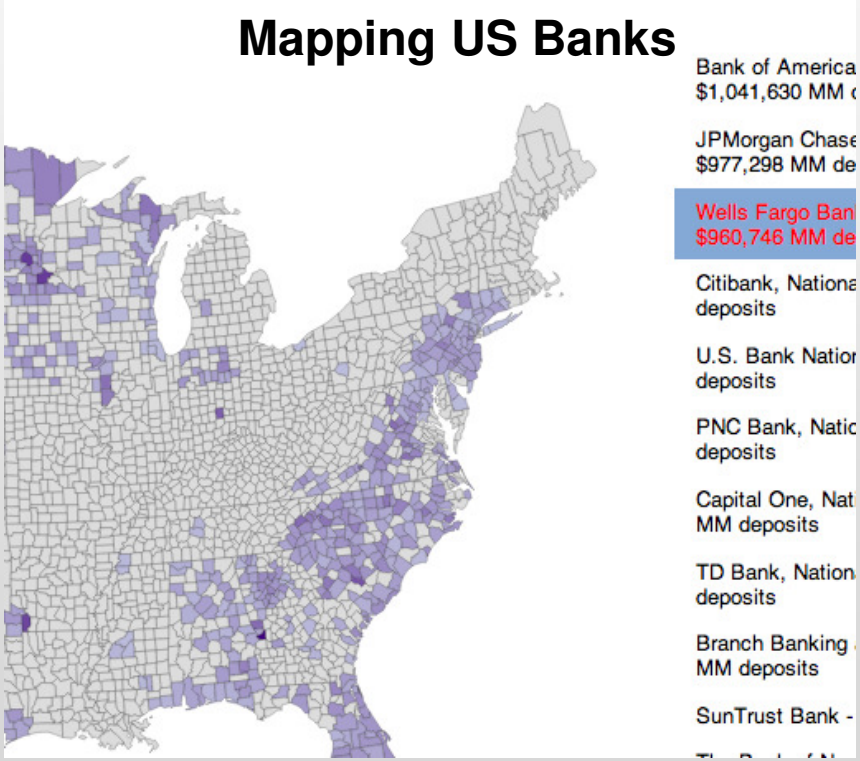
GA GENERAL ASSEMBLY

DIVER
By Lumesis



Recent Work

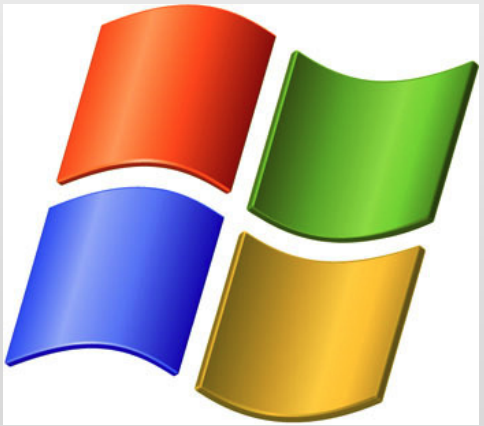

A Startup's Minimum Revenue Per Employee
Log
Revenue
100 M
10 M
10
100
1,000
Employees
Log


Mapping US Banks
Bank of America
$1,041,630 MM
JPMorgan Chase
$977,298 MM de
Wells Fargo Ban
$960,746 MM de
Citibank, Nationa
deposits
U.S. Bank Natior
deposits
PNC Bank, Natic
deposits
Capital One, Nat
MM deposits
TD Bank, Nation
deposits
Branch Banking
MM deposits
SunTrust Bank -

What Do You Do?

Tech? Industries? Goals for this class?

Please fill out initial survey



images via wikipedia

What is this Class About?

How to present information

How to present quantitative information

How to present quantitative information graphically

How to use technology to present quantitative information graphically

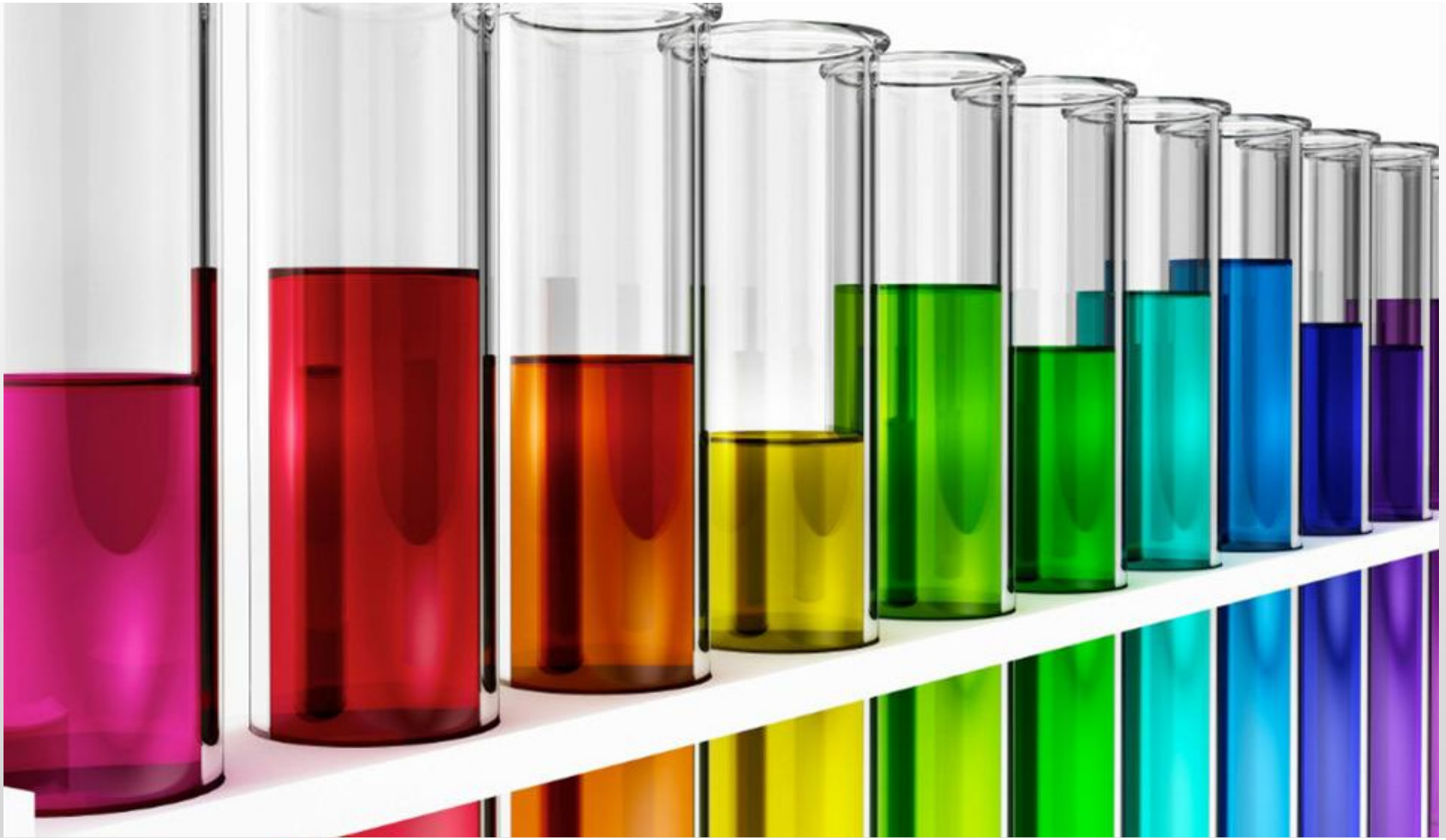
General Class Info

All class dates are in the syllabus

Every class, a short quiz and a homework will be due

At the end of the semester, you will complete a larger, public project

If you have questions, email me at josh.laurito@gmail.com


http://catholiclane.com/wp-content/uploads/rainbo

DataKind

USING DATA IN THE SERVICE OF HUMANITY

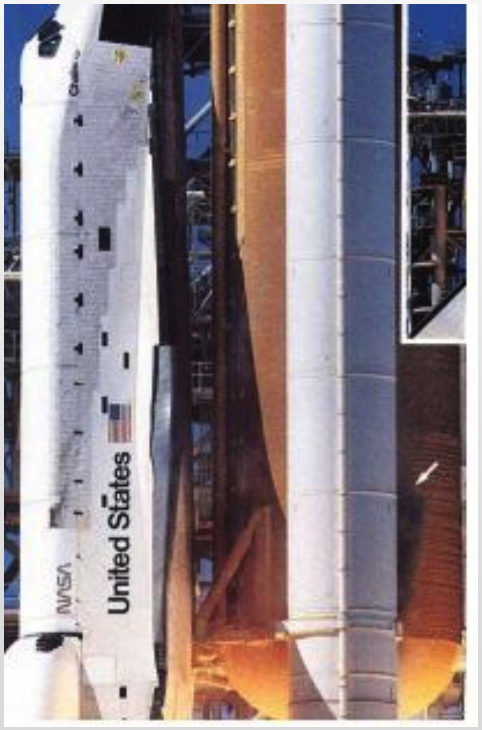

github
SOCIAL CODING

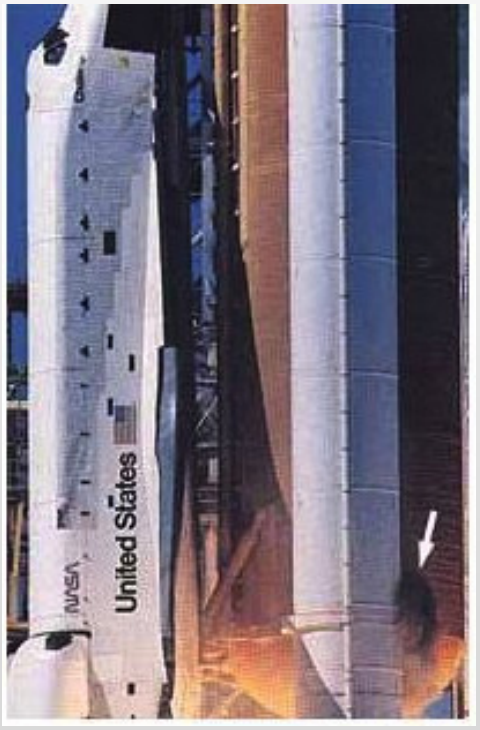
Visualization Technique & Theory

Importance of Data Visualization

http://news.bbc.co.uk/2/shared/spl/hi/pop_ups/06/sci_nat_1986_challenger_disaster/html/1.stm

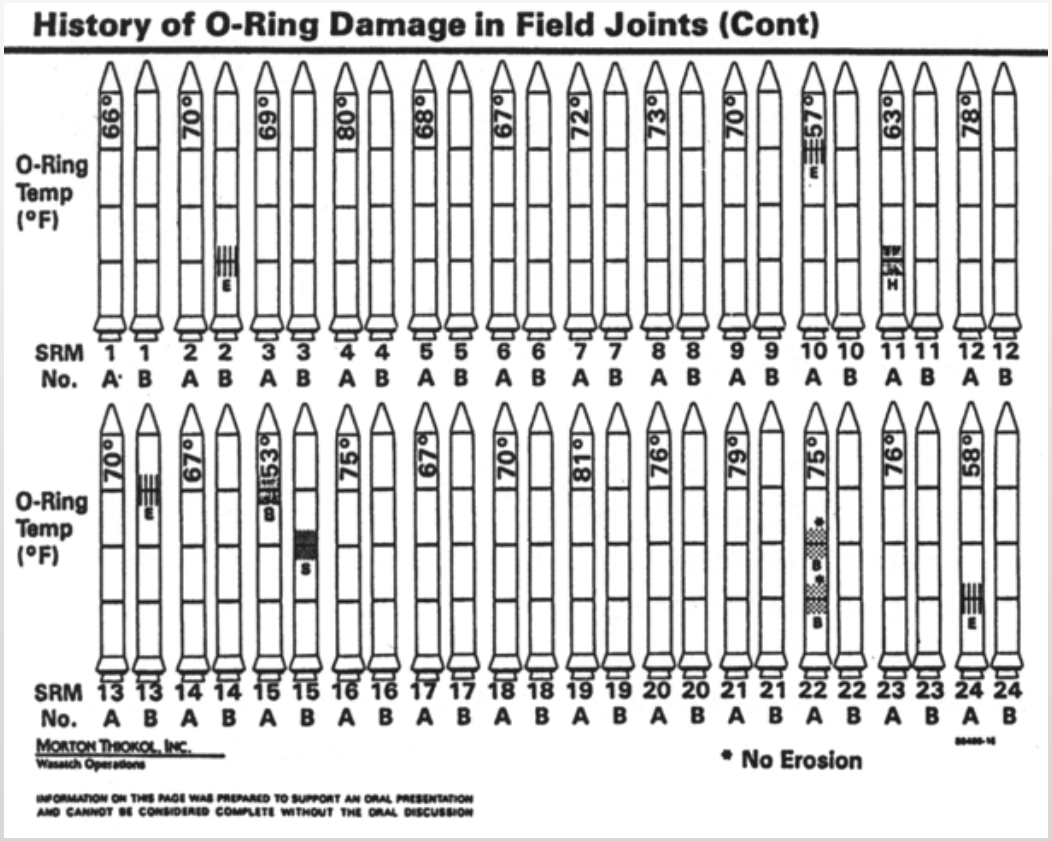
Importance of Data Visualization





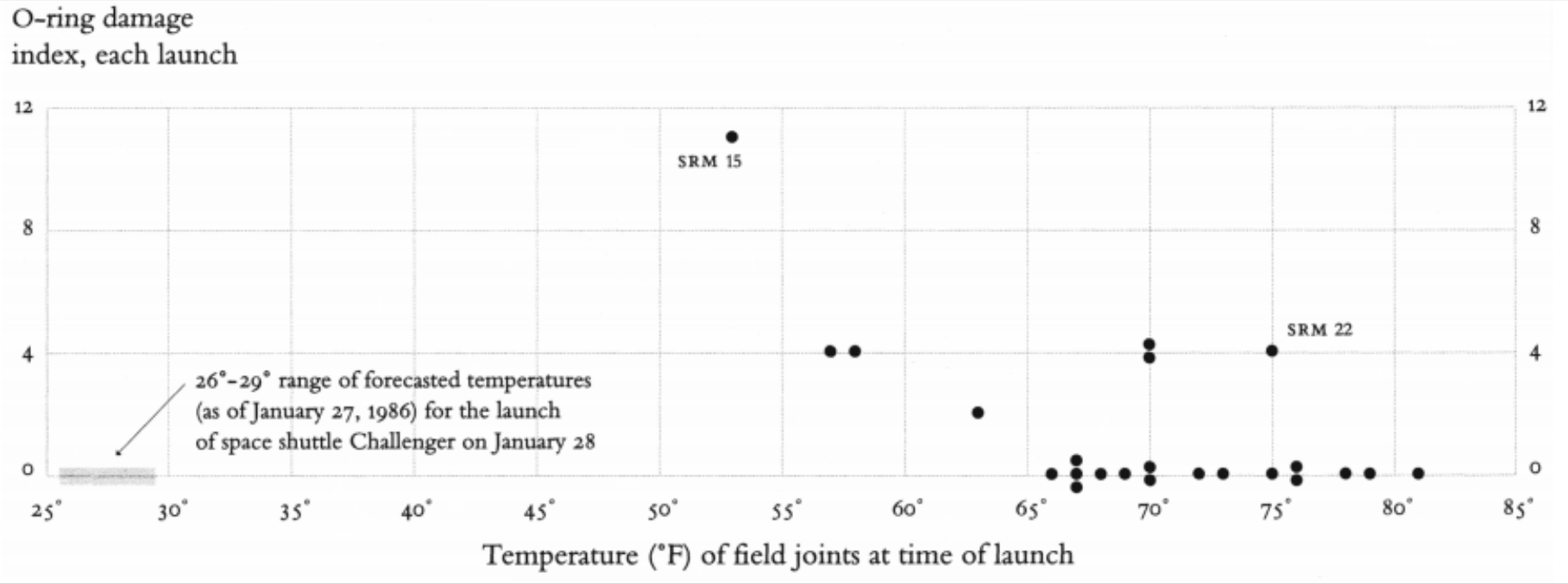
http://www.aerospaceweb.org/question/investigations/q0122.shtml

Importance of Data Visualization



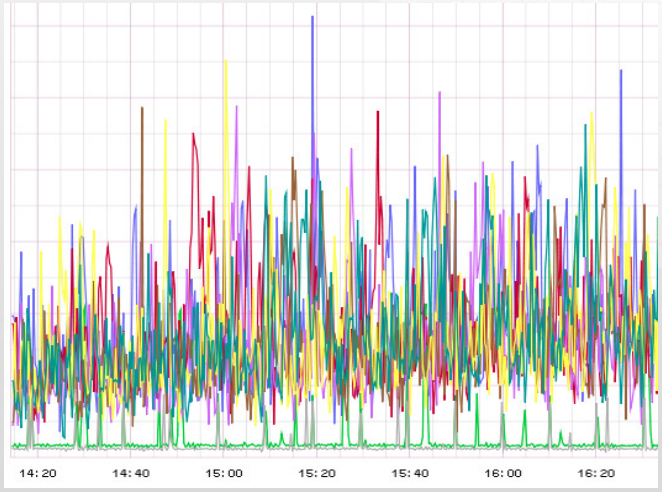
From Tufte's *Visual Explanations*

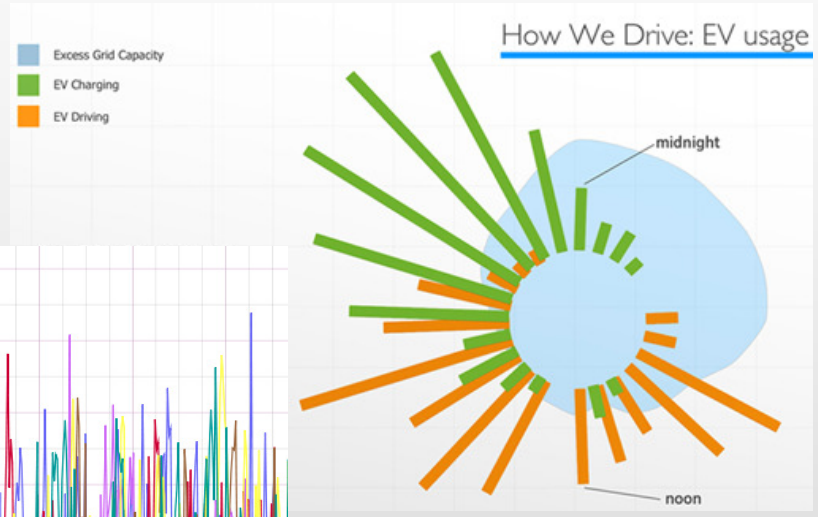
Importance of Data Visualization

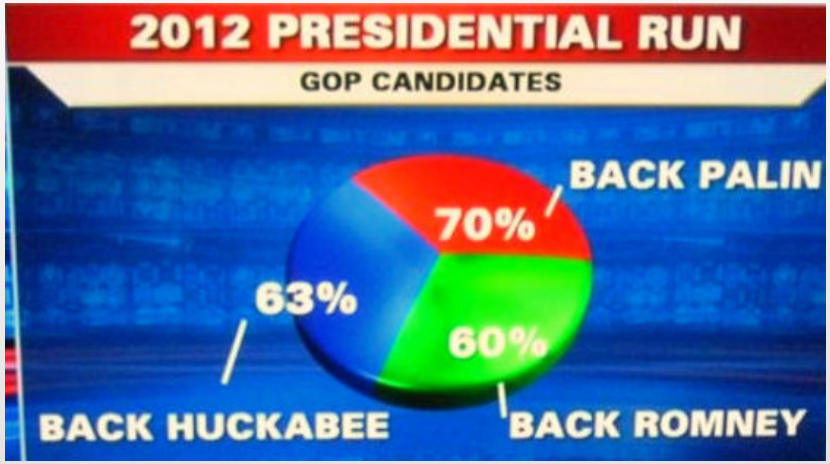


From Tufte's *Visual Explanations*

Other Bad Visualizations







http://blog.visual.ly/poor-visualization-can-do-more-harm-than-good/
http://www.nbcchicago.com/news/local/FOX-News-Chart-Fails-Math-73711092.html
http://themonkeycage.org/2013/01/08/how-2012-stacks-up-the-worst-graph-on-record/

Goals of Data Visualization

Data Visualization has a few primary goals:

- Minimize mental effort (images are more clear than numbers)
- Make information easy to retain (memory is often visual)
- Insight over beauty (‘Aha!’ is more important than ‘wow’)

Visualization Strategies

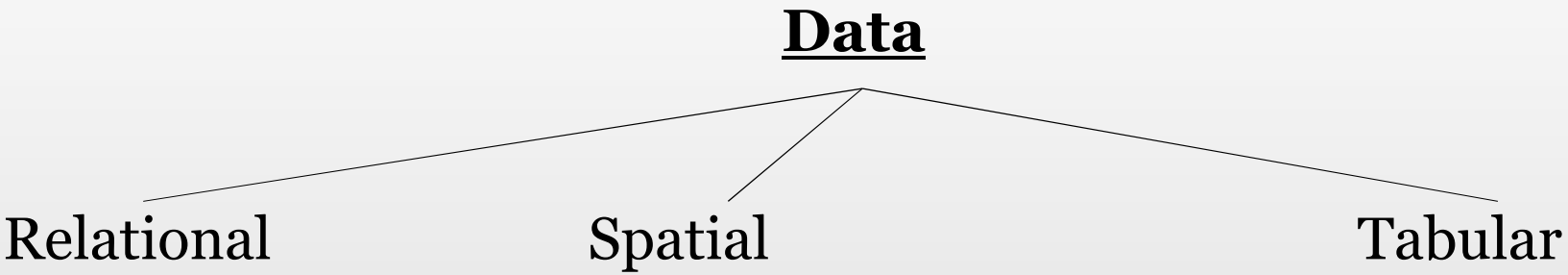
Ways to make visualization **clear**, **memorable** and **insightful**:

Match display to data and purpose

Maximize information-to-ink ratio

Principal of small multiples

Match Display to Data and Purpose



This and the next few slides paraphrase Tamara Munzner http://www.cs.ubc.ca/~tmm/talks/vizbi11/vizbi11.pdf

Match Display to Data and Purpose

Relational

Links between points

Family trees, Int'l trade, Types of data

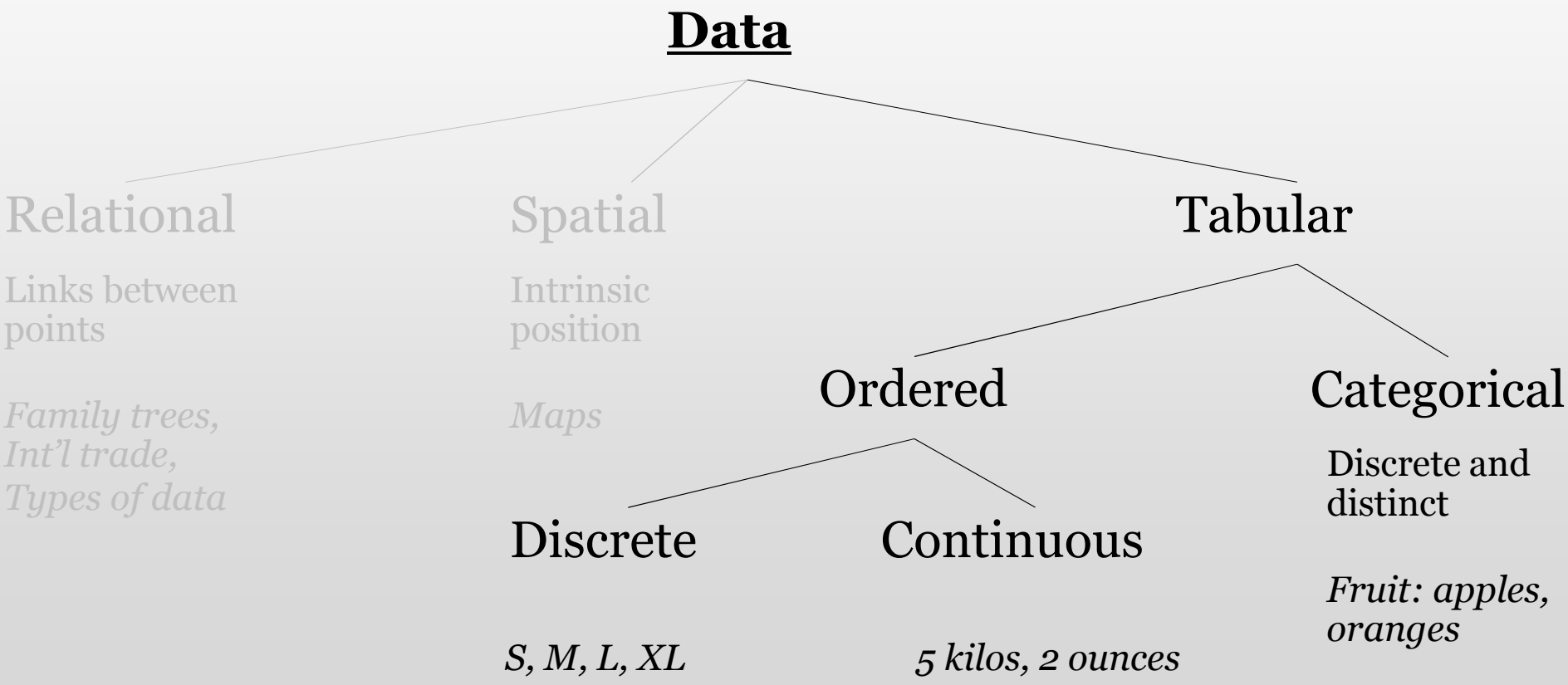
Spatial

Intrinsic position

Maps

Tabular

Match Display to Data and Purpose



Match Display to Data and Purpose

Relational

Spatial

Ordered
discrete/cont.

Categorical

Match Display to Data and Purpose

Relational

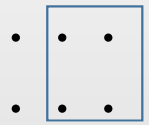
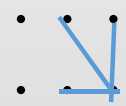
Spatial

Maps

Ordered
discrete/cont.

Categorical

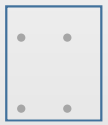
Match Display to Data and Purpose

Relational	Spatial	Ordered discrete/cont.	Categorical
Contain 	Maps 		
Connect 			
Similarity 			
Proximity 			

Match Display to Data and Purpose

Relational

Contain



Connect

Similarity

Proximity

Spatial

Maps

Ordered

discrete/cont.

Position

Length

Area

Shade

Saturation

etc…

Categorical

Match Display to Data and Purpose

Relational

Contain

Connect

Similarity

Proximity

Spatial

Maps

Ordered
discrete/cont.

Position

Length

Area

Shade

Saturation

etc...

Categorical

Position

Color

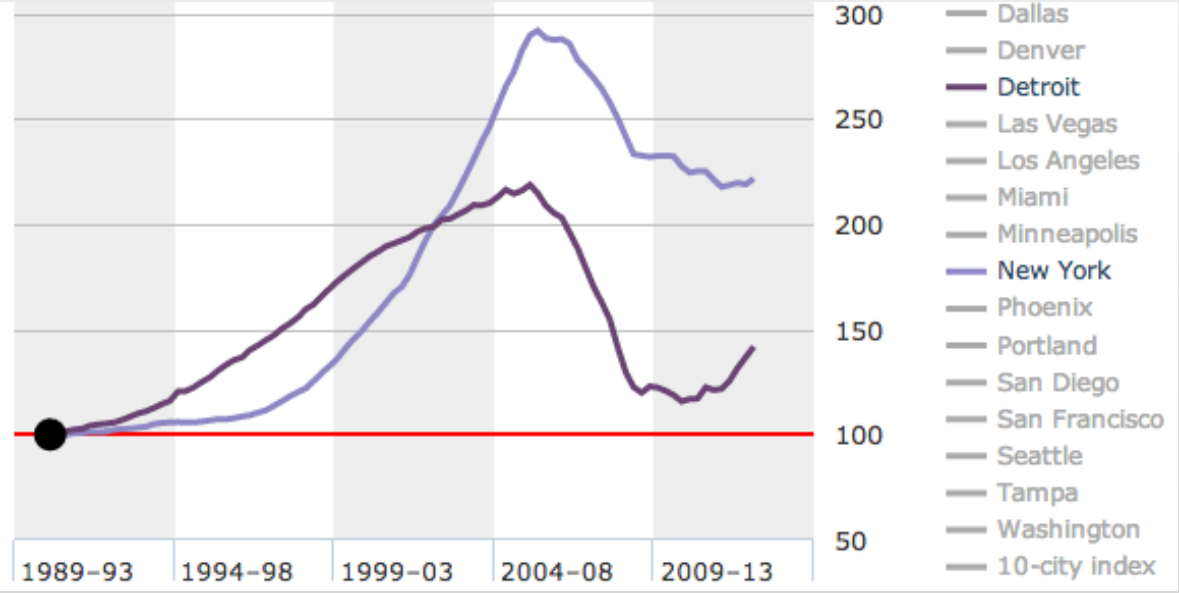
Shape

Patterns

Match Dimensions to Data

Use highly ranked channels for key info to keep memorable

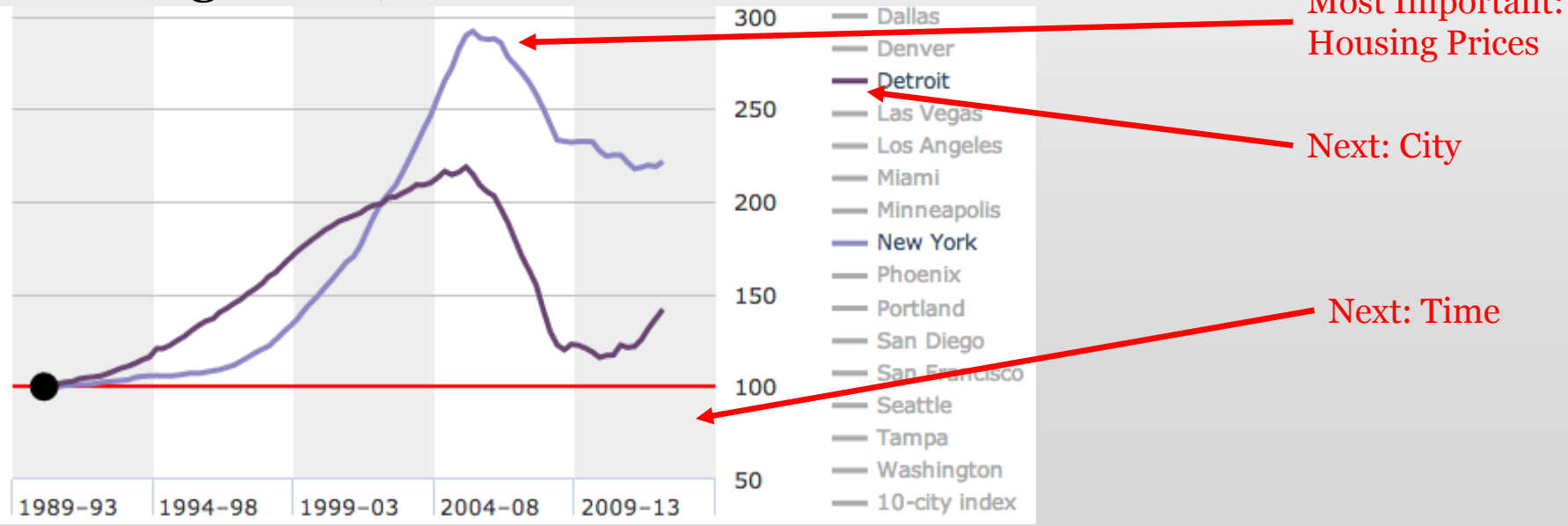
Housing Prices, *The Economist*



Match Dimensions to Data

Use highly ranked channels for key info to keep memorable

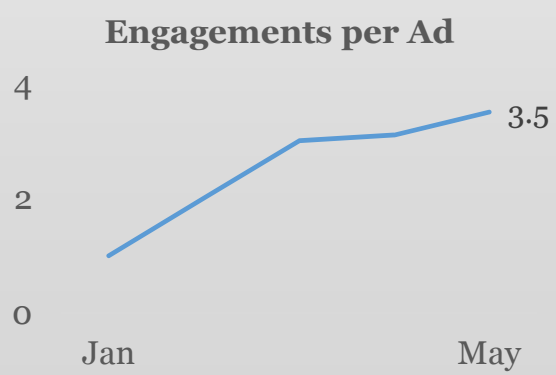
Housing Prices, *The Economist*

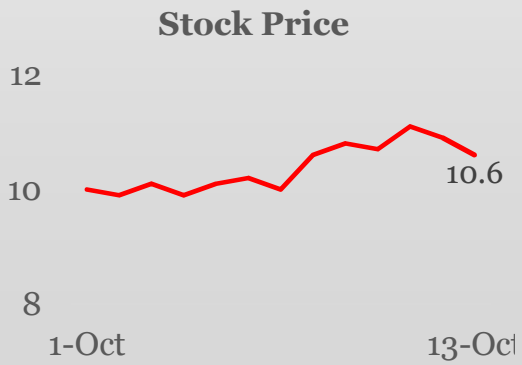


Maximize Information-to-Ink Ratio

Achieve clarity by removing distractions & distracting elements

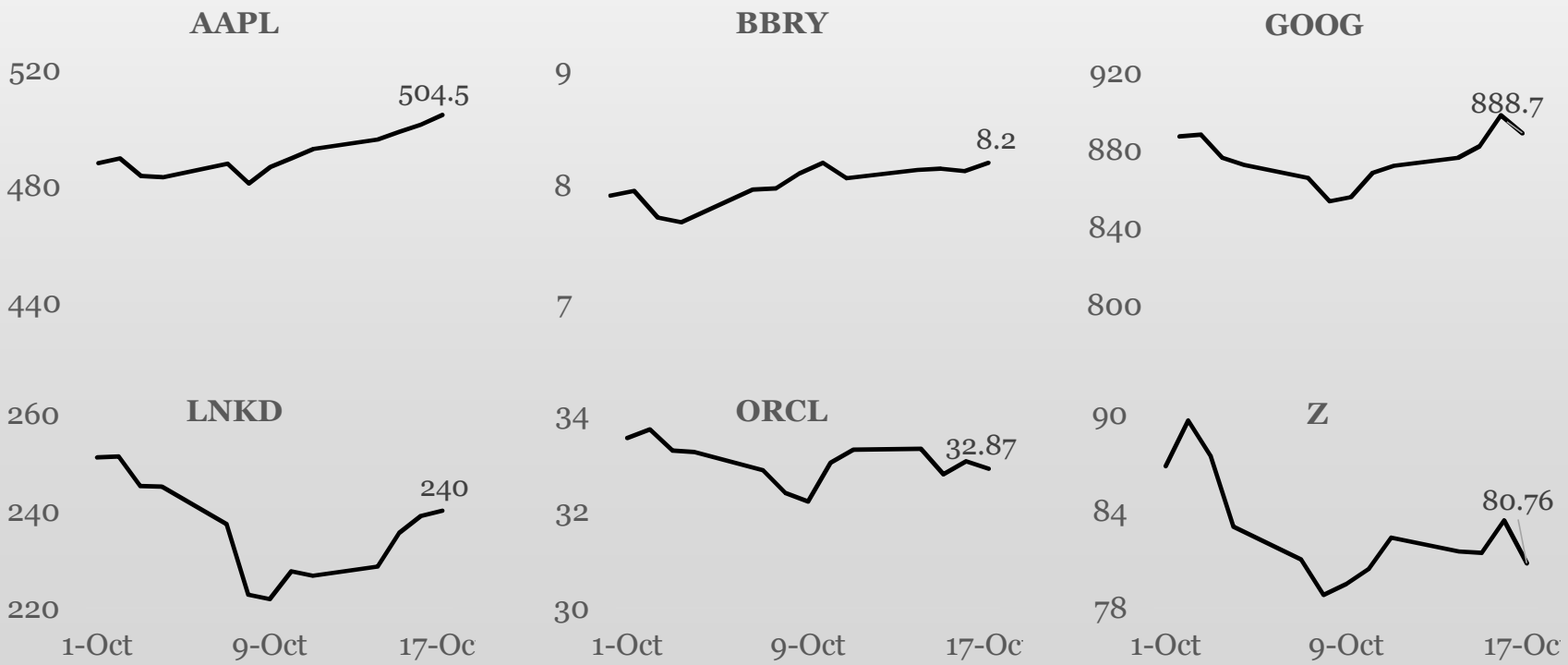
Taken to its logical endpoint: **Sparklines**



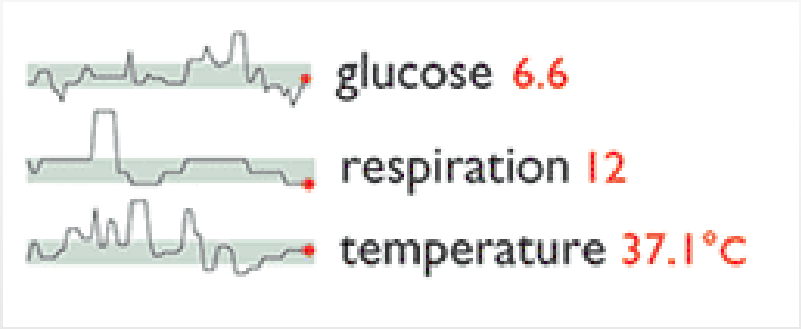


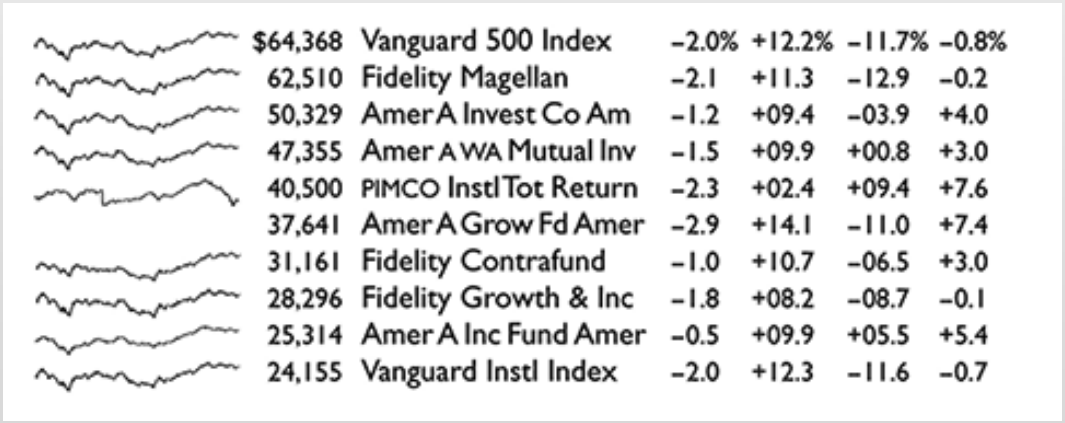
Principal of Small Multiples

It is easier to compare items next to each other than over-plotted



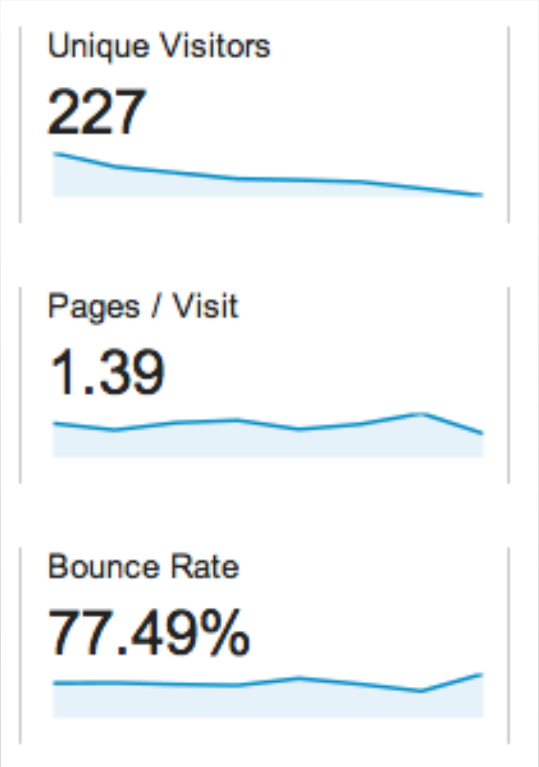
Principal of Small Multiples In Practice





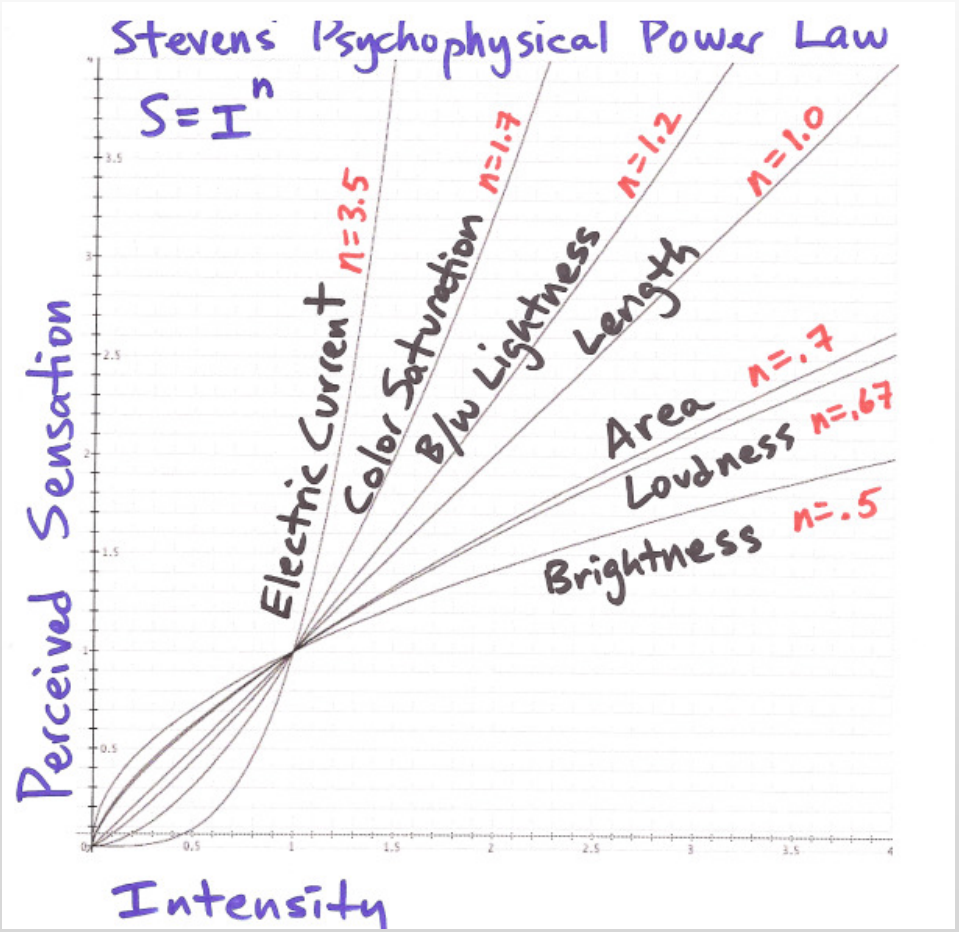
$64,368	Vanguard 500 Index	−2.0%	+12.2%	−11.7%	−0.8%
62,510	Fidelity Magellan	−2.1	+11.3	−12.9	−0.2
50,329	Amer A Invest Co Am	−1.2	+09.4	−03.9	+4.0
47,355	Amer A WA Mutual Inv	−1.5	+09.9	+00.8	+3.0
40,500	PIMCO Instl Tot Return	−2.3	+02.4	+09.4	+7.6
37,641	Amer A Grow Fd Amer	−2.9	+14.1	−11.0	+7.4
31,161	Fidelity Contrafund	−1.0	+10.7	−06.5	+3.0
28,296	Fidelity Growth & Inc	−1.8	+08.2	−08.7	−0.1
25,314	Amer A Inc Fund Amer	−0.5	+09.9	+05.5	+5.4
24,155	Vanguard Instl Index	−2.0	+12.3	−11.6	−0.7

http://www.edwardtufte.com/



https://www.google.com/analytics/

Things To Keep in Mind - Dimensions



Things To Keep in Mind - Discrimination

~3 for Pie Charts 

~7 for colors (not as a scale!) 

~4 overlapping lines

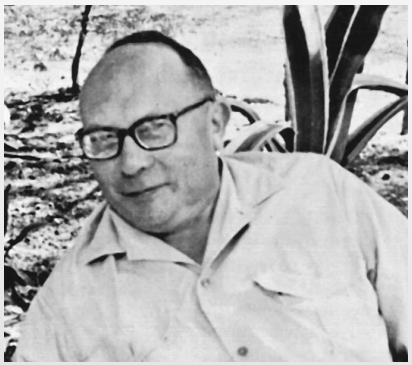
Just the Basics

Data Visualization: seems small & cute, actually a huge topic!

image via http://www.

Further Reading on Visualization

Pioneers: Jacques Bertin



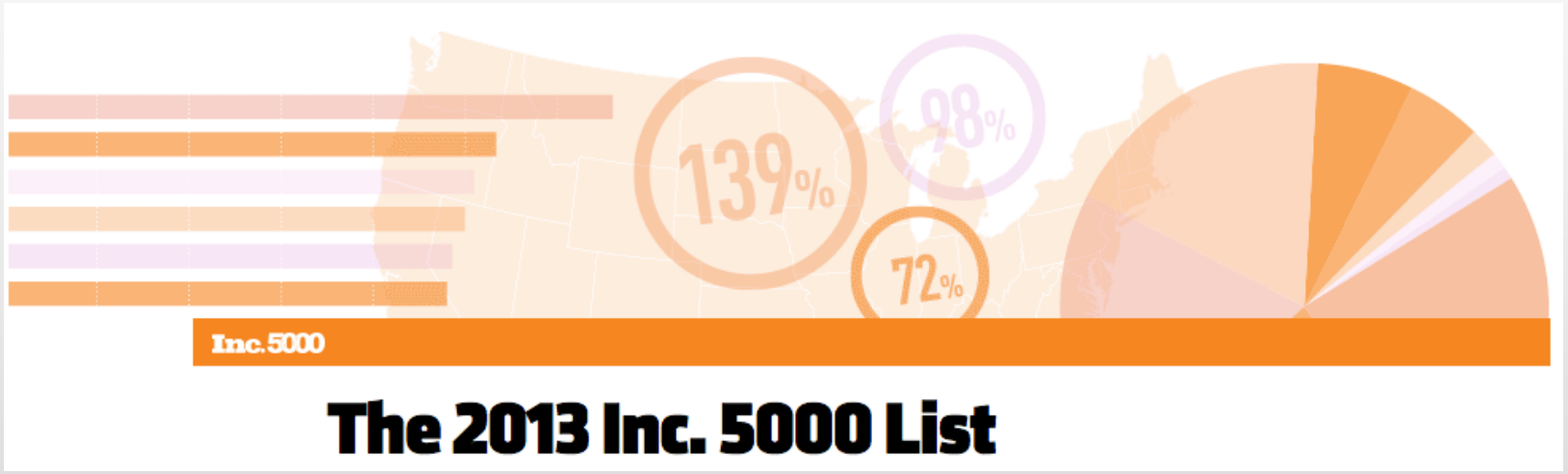
Edward Tufte

Current Leaders: Tamara Munzner

Mike Bostock

Favorites:
Bertin
Tufte
Munzner
Bostock

Assignment 1 – Business Data, ggplot2



Assignment 1 – Business Data, ggplot2

Read ggplot2 presentation

Submit code to me

Post images with some analysis

Contact me with Questions

josh.laurito@gmail.com is best