

a brief history of
Data Visualization



Jeffrey Heer Stanford University

What was the **first**
data visualization?

o BC
|



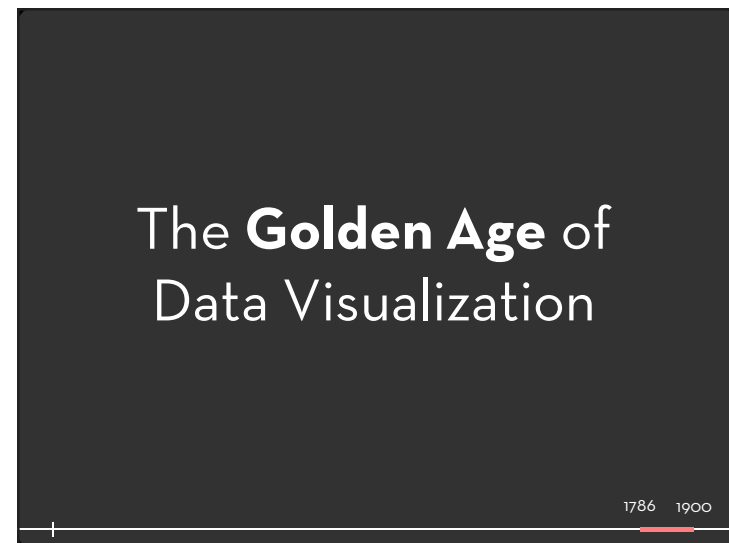
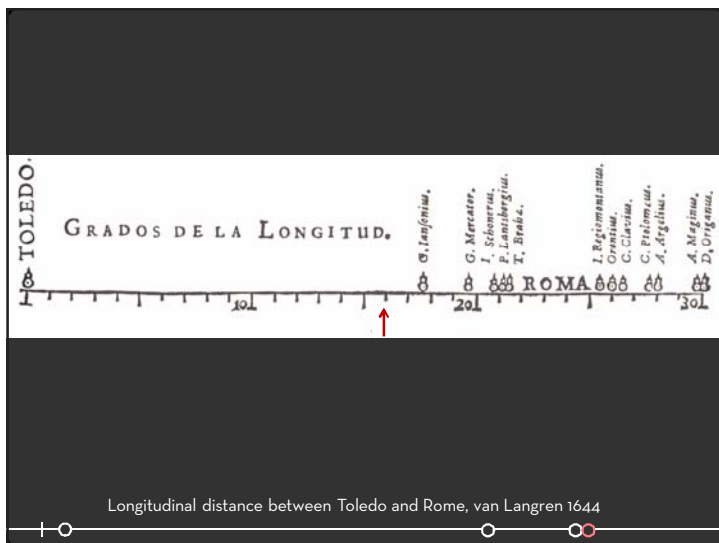
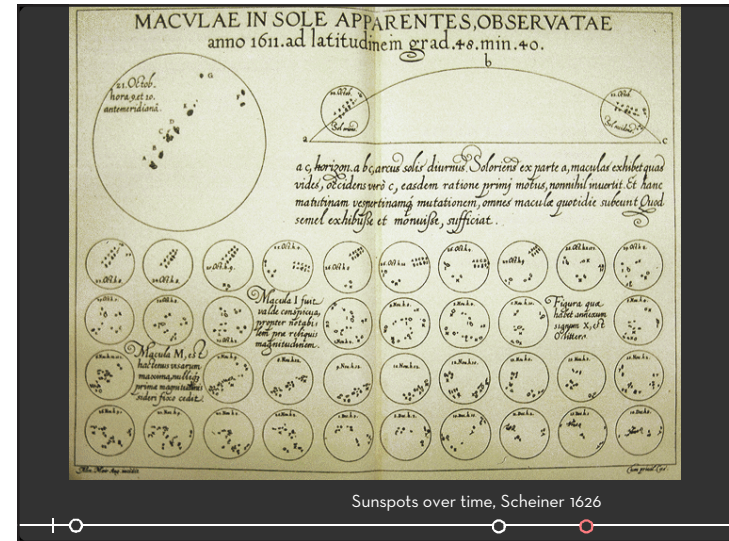
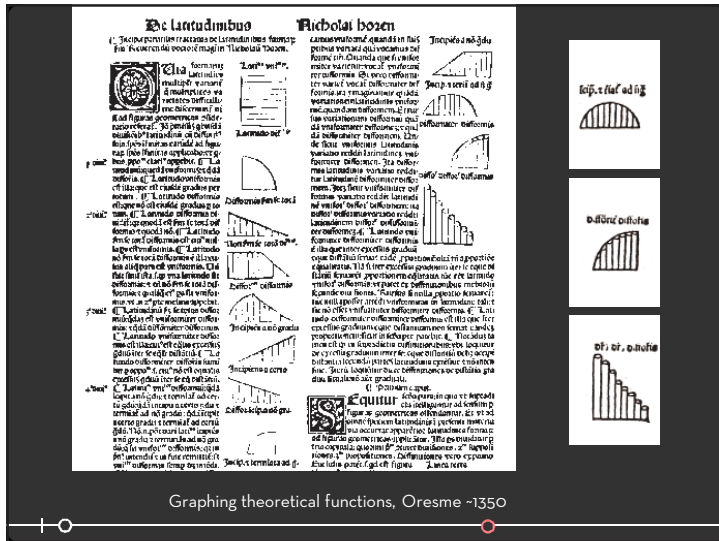
-6200 BC Town Map of Catal Hyük, Konya Plain, Turkey

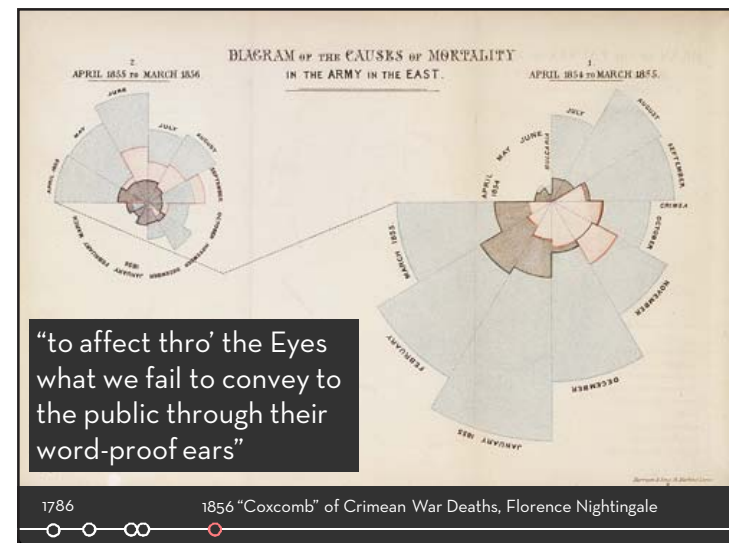
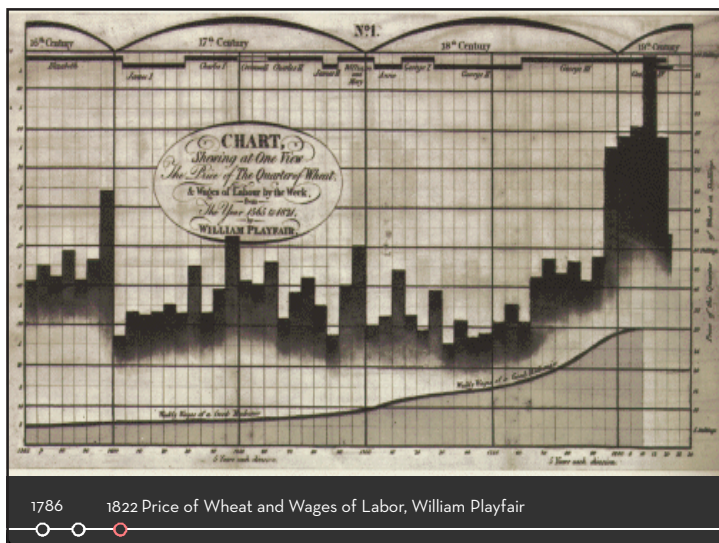
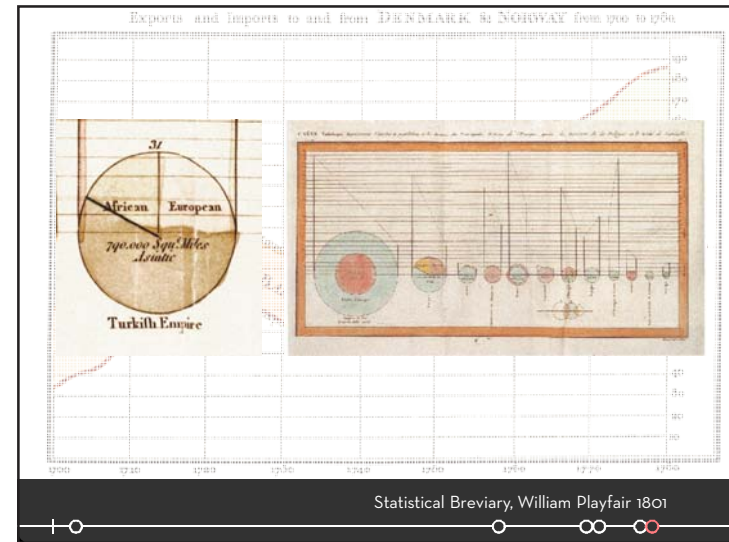
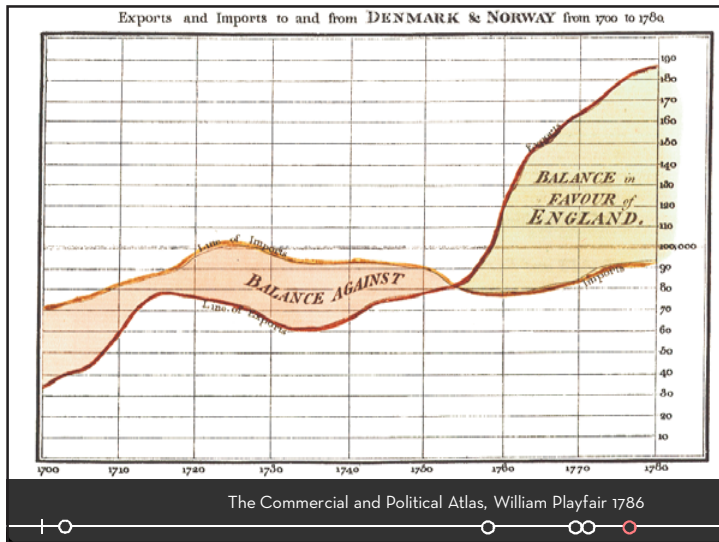
o BC
|

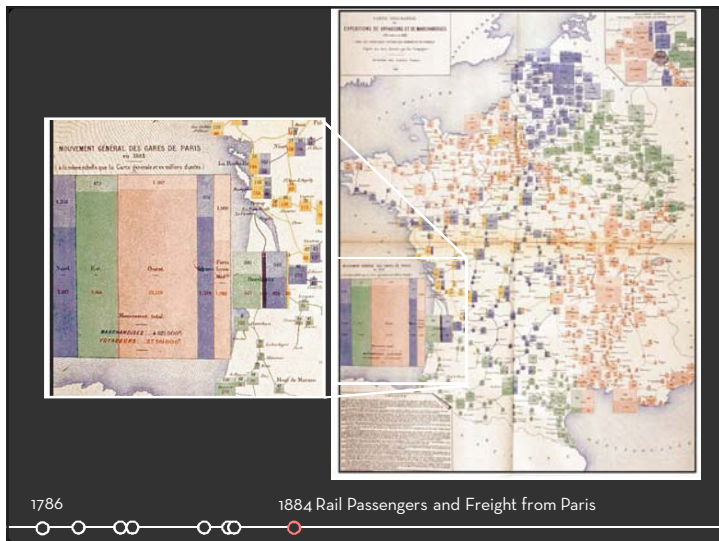
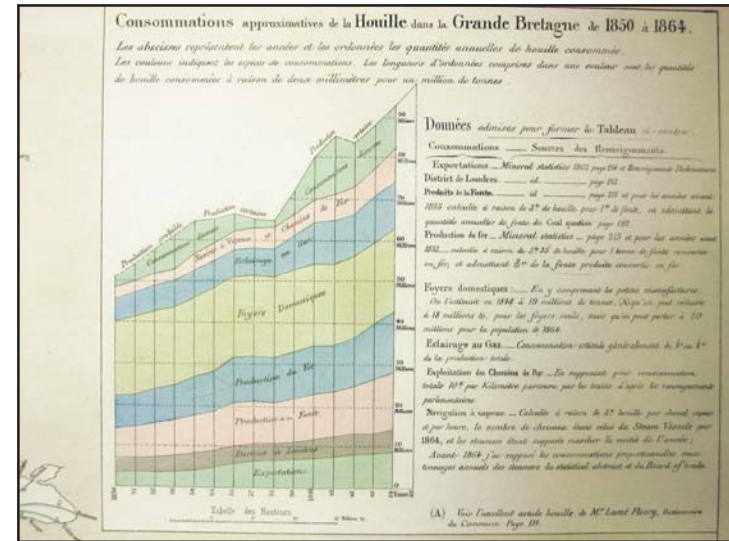


6200 BC

Geographica, Ptolemy 150 AD







Rise of **formal methods** in statistics and social science – Fisher, Pearson, ...

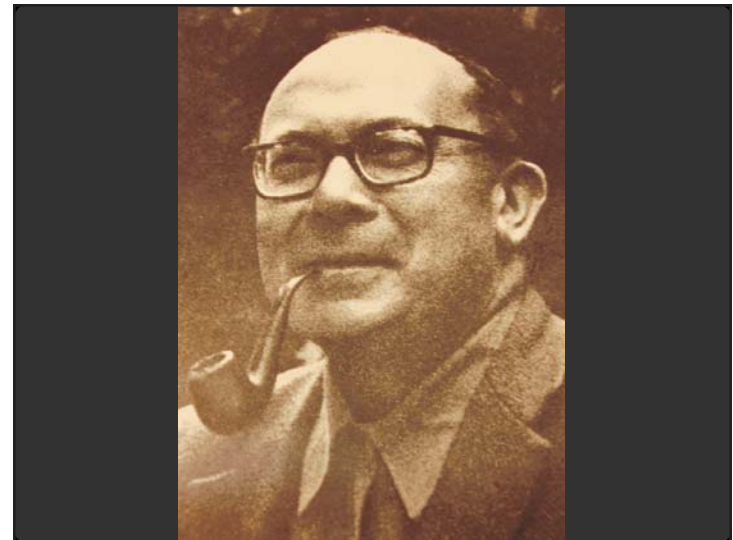
Little innovation in graphical methods

A period of **application and popularization**

Graphical methods enter textbooks, curricula, and **mainstream use**



The Modern Era



LES VARIABLES DE L'IMAGE

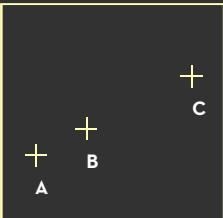
	POINTS			LIGNES		ZONES	
XY 2 DIMENSIONS DU PLAN	x	x	x	/	~	14 15 9 2 18 2 18 21 2 1 21 15 14 15 3 11 2 9	14 15 9 2 18 2 18 21 2 1 21 15 14 15 3 11 2 9
Z TAILLE	■	■	■	/	~	■	■
VALEUR	■	■	■	/	~	■	■

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN	■	■	■	/	~	■	■
COULEUR	■	■	■	/	~	■	■
ORIENTATION	■	■	■	/	~	■	■
FORME	■	■	■	/	~	■	■

1786 Sémiologie Graphique 1967 1983 English Trans.

Bertin's Semiology of Graphics



1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

∴ Encode quantitative variables

"Resemblance, order and proportion are the three signifieds in graphics." - Bertin

1786 Sémiologie Graphique 1967 1983 English Trans.

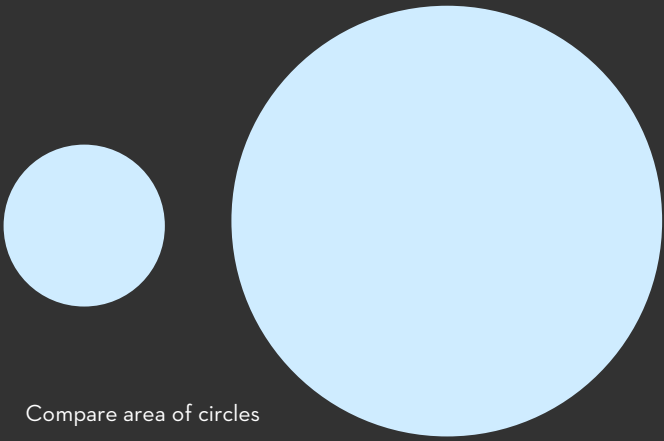
LES VARIABLES DE L'IMAGE

	POINTS			LIGNES		ZONES	
XY 2 DIMENSIONS DU PLAN	x	x	x	/	~	14 15 9 2 18 2 18 21 2 1 21 15 14 15 3 11 2 9	14 15 9 2 18 2 18 21 2 1 21 15 14 15 3 11 2 9
Z TAILLE	■	■	■	/	~	■	■
VALEUR	■	■	■	/	~	■	■

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN	■	■	■	/	~	■	■
COULEUR	■	■	■	/	~	■	■
ORIENTATION	■	■	■	/	~	■	■
FORME	■	■	■	/	~	■	■

1786 Sémiologie Graphique 1967 1983 English Trans.

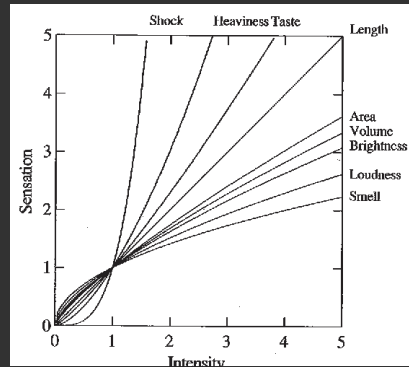


Compare area of circles

Steven's Power Law

$$S = I^p$$

$p < 1$: underestimate
 $p > 1$: overestimate



[graph from Wilkinson 99, based on Stevens 61]

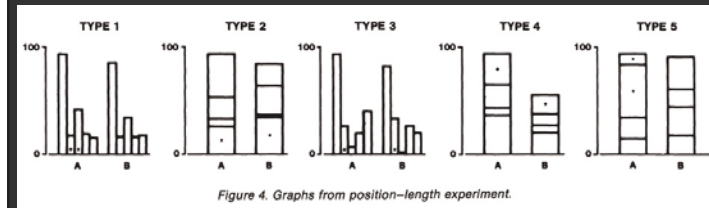


Figure 4. Graphs from position-length experiment.

1786

Cleveland & McGill, Graphical Perception 1984

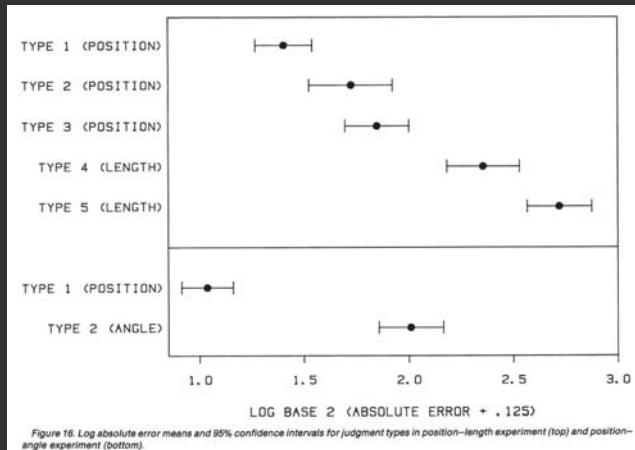
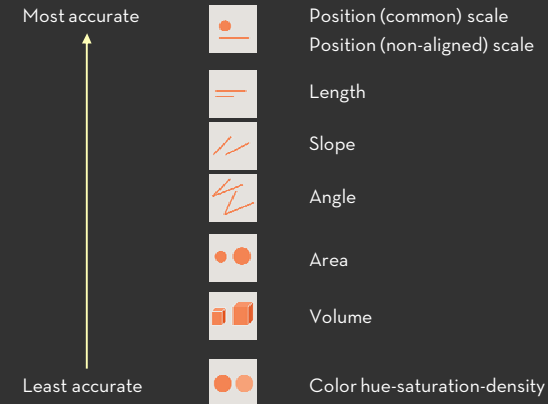


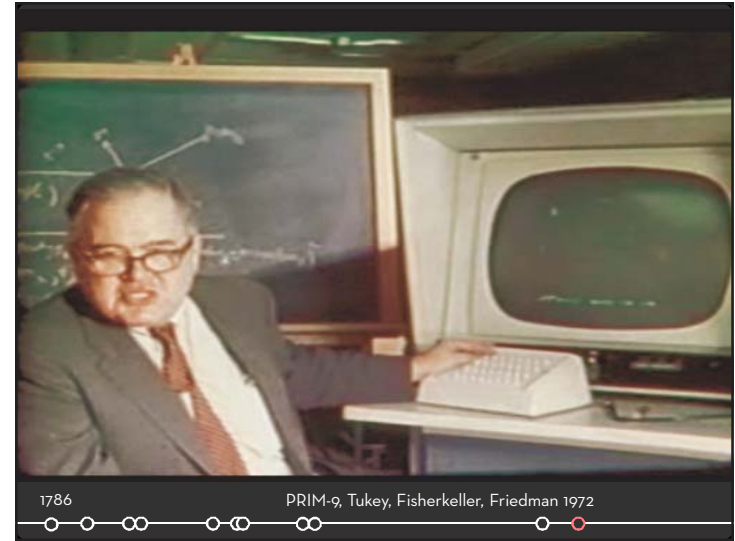
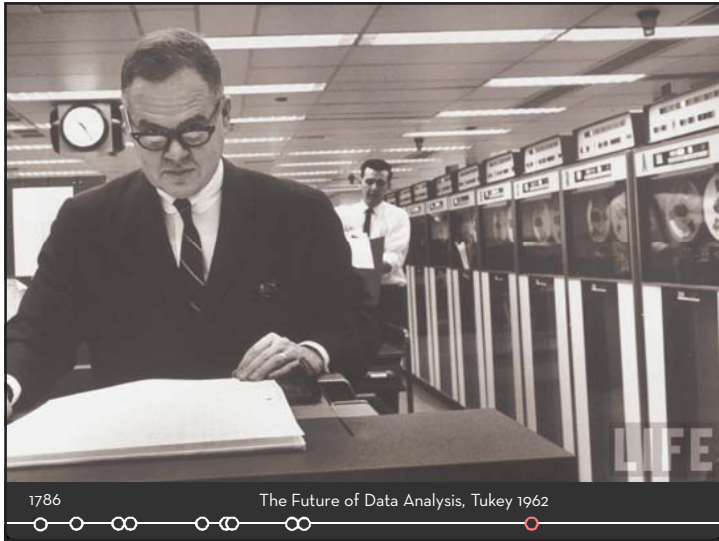
Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position-length experiment (top) and position-angle experiment (bottom).

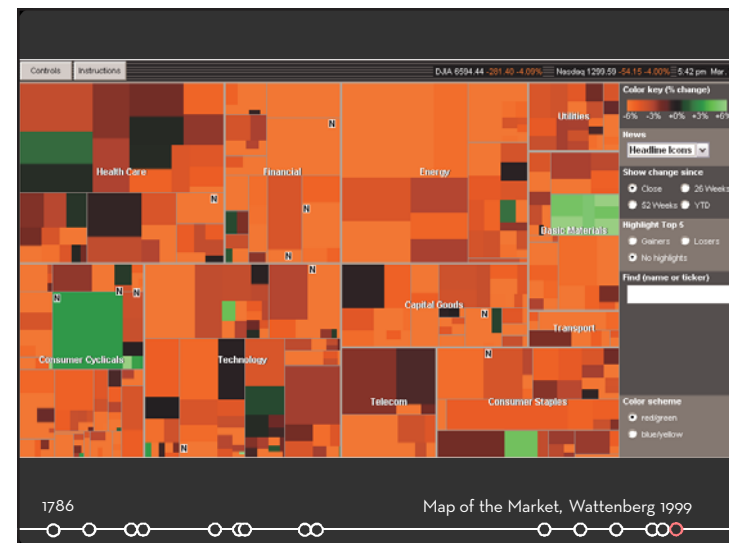
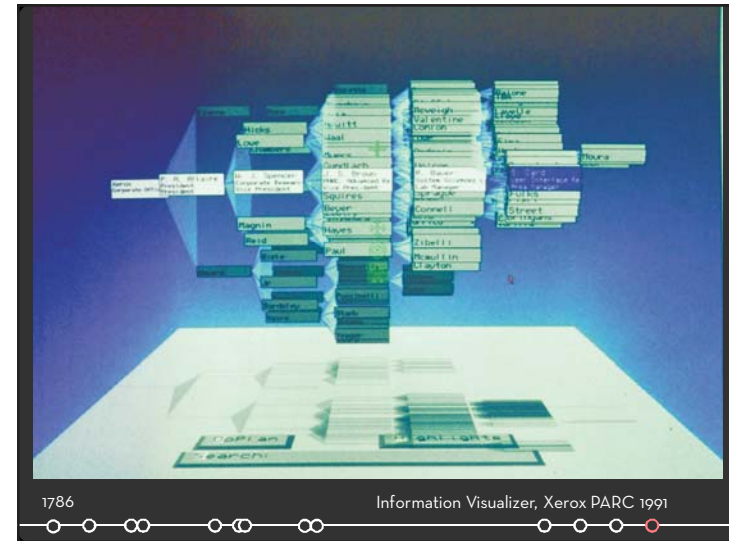
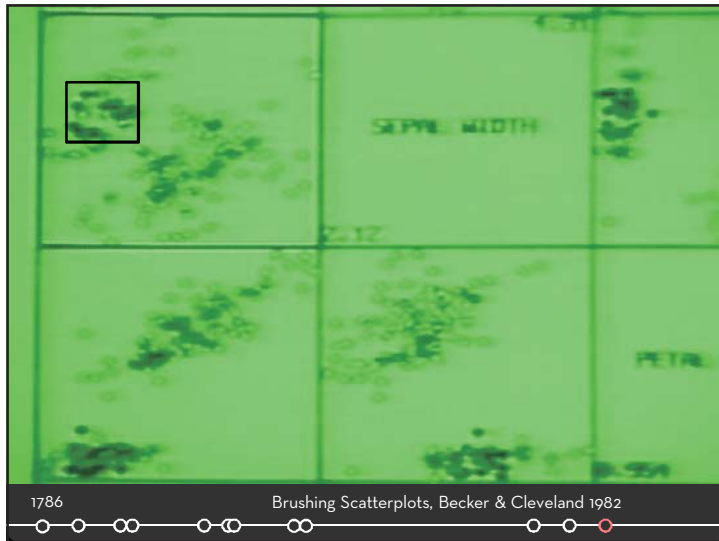
[Cleveland and McGill 84]

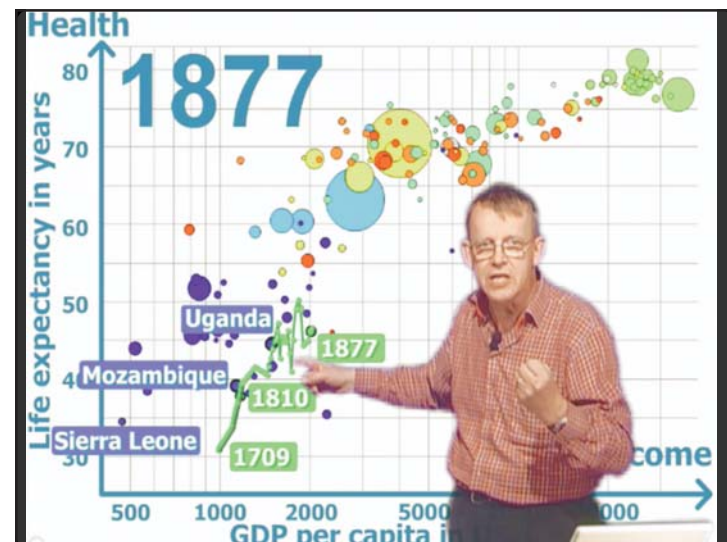
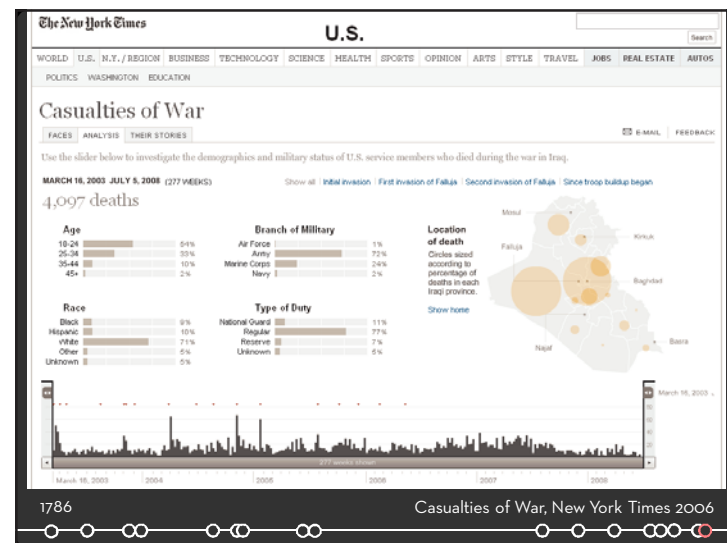


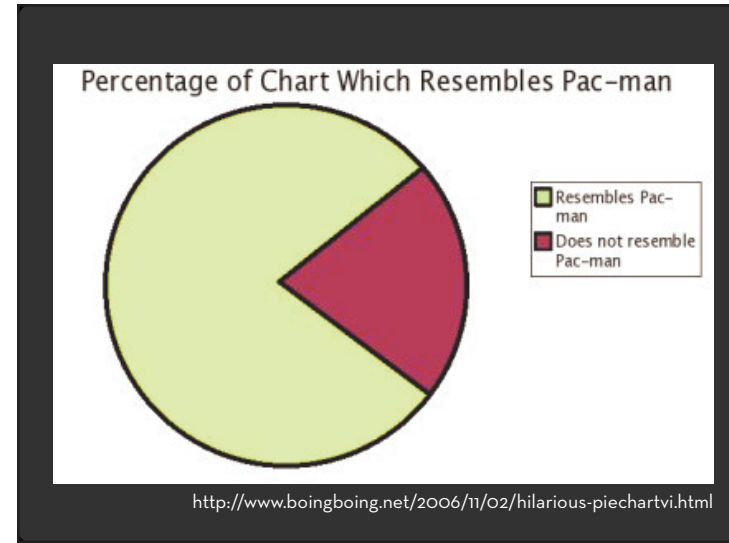
1786

Cleveland & McGill, Graphical Perception 1984

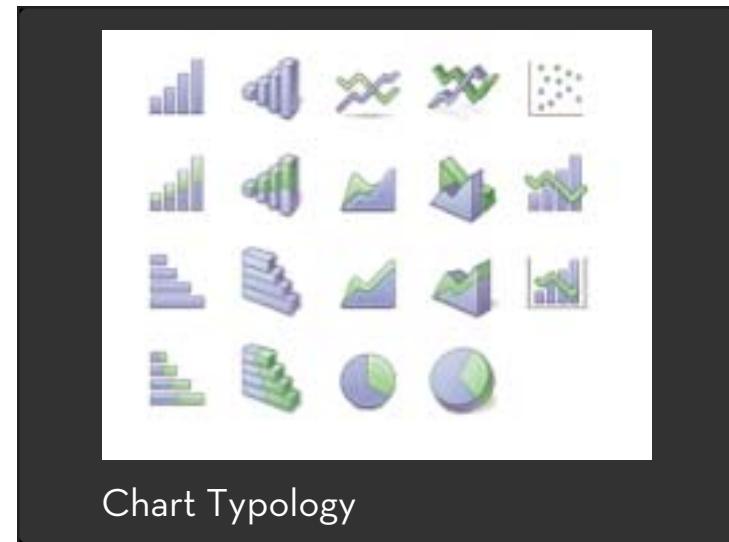








The Rise of Graphical Literacy and Expression



Data Sets : State Quick Facts

Uploaded By: [zuggut](#) Created at: Friday May 18, 3:08 PM
 Data Source: [US Census Bureau](#)
 Description:
 Tags: [people census](#)

[View as text](#) [View as table](#)

	People QuickFacts	Population 2005 estimate	Population percent change April 1 2000 to July 1 2005	Population 2000	Population percent change 1990 to 2000	Persons under 5 years old percent 2004	Persons 18 years old percent 2004	Persons 65 years old and over percent 2004
1	Alabama	4557800	0.03	4447100	0.1	0.07	0.24	0.13
2	Alaska	663661	0.06	626932	0.14	0.08	0.29	0.06
3	Arizona	5939292	0.16	5130632	0.4	0.08	0.27	0.13
4	Arkansas	2779154	0.04	2673400	0.14	0.07	0.25	0.14
5	California	36132147	0.07	33971648	0.14	0.07	0.27	0.11
6	Colorado	4665177	0.08	4301261	0.31	0.07	0.26	0.1
7	Connecticut	3510237	0.03	3405565	0.04	0.06	0.24	0.14
8	Delaware	843524	0.08	783600	0.18	0.07	0.23	0.13
9	Florida	17789064	0.11	15982378	0.24	0.06	0.23	0.17
10	Georgia	9072576	0.11	8186453	0.26	0.08	0.26	0.1
11	Hawaii	1275194	0.05	1211537	0.09	0.07	0.24	0.14
12	Idaho	1429096	0.1	1293963	0.29	0.07	0.27	0.11
13	Illinois	12763371	0.03	12419293	0.09	0.07	0.26	0.12

Choosing a visualization type for State Quick Facts

Analyze a text

Tag Cloud
 How are you using your words? This enhanced tag cloud will show you the words popularity in the given set of text.
[Learn more](#)

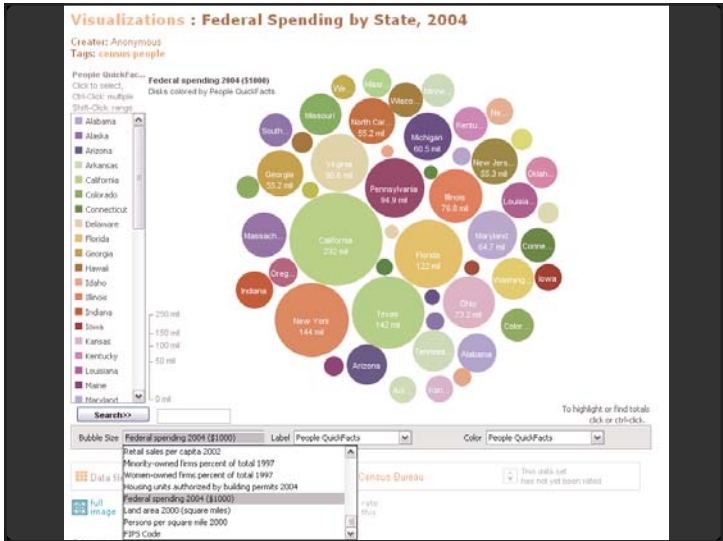
Wordle
 Wordle is a toy for generating "word clouds" from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.
[Learn more](#)

Word Tree
 See a branching view of how a word or phrase is used in a text. Navigate the text by zooming and clicking.
[Learn more](#)

Compare a set of values

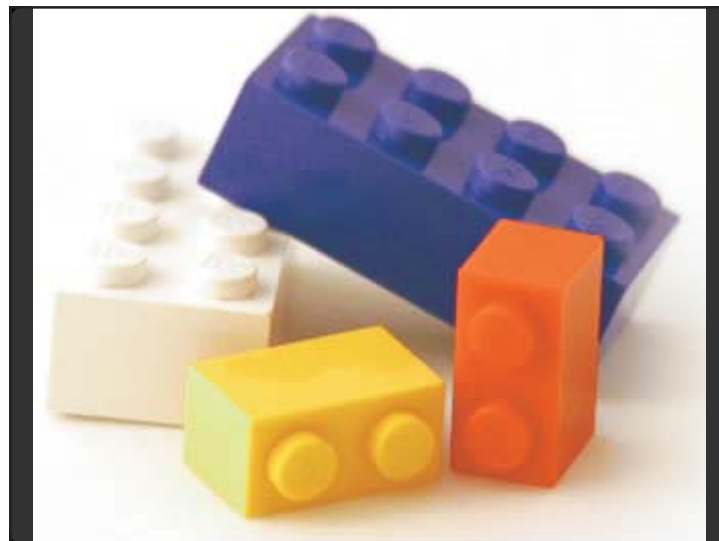
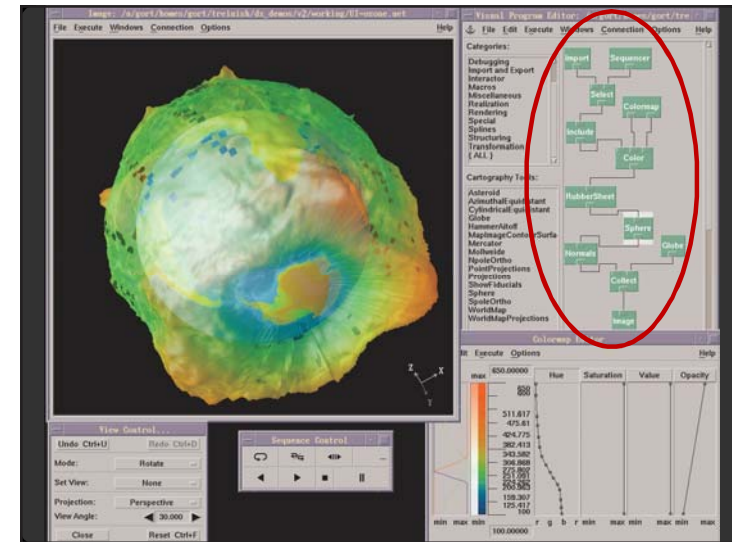
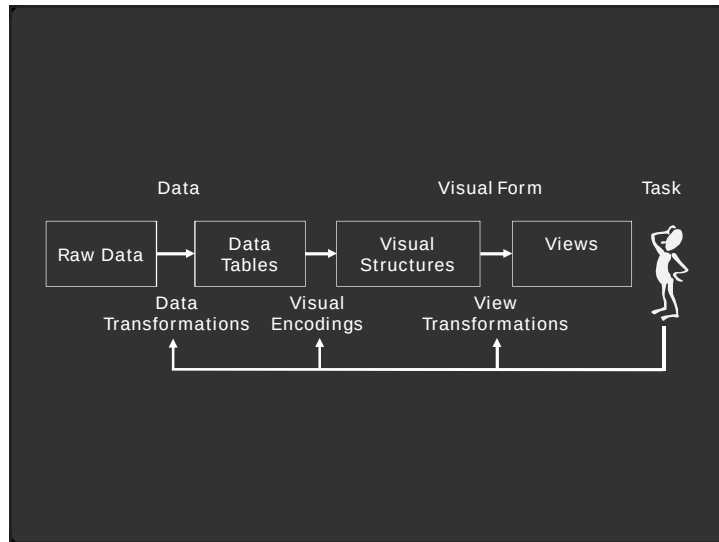
Bar Chart
 How do the items in your data set stack up? A bar chart is a simple and recognizable way to compare values. You can display several sets of bars for multivariate comparisons.
[Learn more](#)

Block Histogram
 This versatile chart lets you get a quick sense of how a single set of data is distributed. Each item in the data is an individually identifiable block.
[Learn more](#)



MAD LIBS
 MY MUSIC LESSON

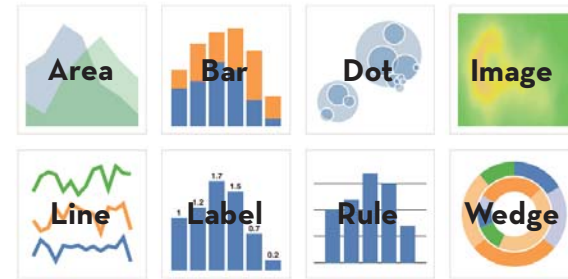
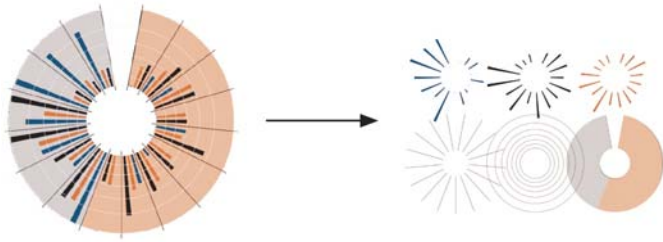
Every Wednesday, when I get home from school, I have a piano lesson. My teacher is a very strict house NOUN. Her name is Hillary Clinton CELEBRITY (FEMALE). Our piano is a Steinway Concert tree NOUN and it has 88 AKK cups RURAL NOUN. It also has a soft pedal and a/an Smily ADJECTIVE pedal. When I have a lesson, I sit down on the piano ALBERTO NOUN and play for 16 minutes PERIOD OF TIME. I do scales to exercise my cats PLURAL NOUN, and then I usually play a minuet by Johann Sebastian washington CELEBRITY (LAST NAME). Teacher says I am a natural Haunted House NOUN and have a good musical leg PART OF THE BODY. Perhaps when I get better I will become a concert Vet PROFESSION and give a recital at Carnegie hospital TYPE OF BUILDING.



Today's first task is not to invent wholly new [graphical] techniques, though these are needed. Rather we need most vitally to recognize and reorganize the essential of old techniques, to make easy their assembly in new ways, and to modify their external appearances to fit the new opportunities.

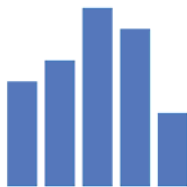
J. W. Tukey, *The Future of Data Analysis*, 1962.

Protovis: A Declarative Language for Visualization <http://protovis.org/>



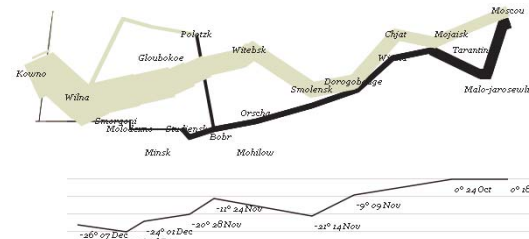
Protovis

Create customized visualizations using a declarative specification language.



```
var vis = new pv.Panel();
vis.add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, .7])
  .bottom(10).width(20)
  .height(function(d) d * 70)
  .left(function() this.index * 25 + 20);
vis.render();
```

Protovis (protovis.org) - JavaScript Visualization Tools



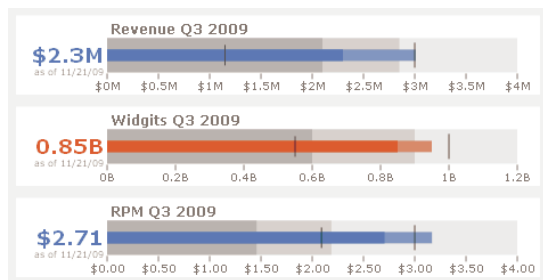
```
var army = pv.nest(napoleon.army, "dir", "group");
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(function() army[this.idx])
  .left(lon).top(lat).size(function(d) d.size/8000)
  .strokeStyle(function() color[army[panelIndex][0].dir]);

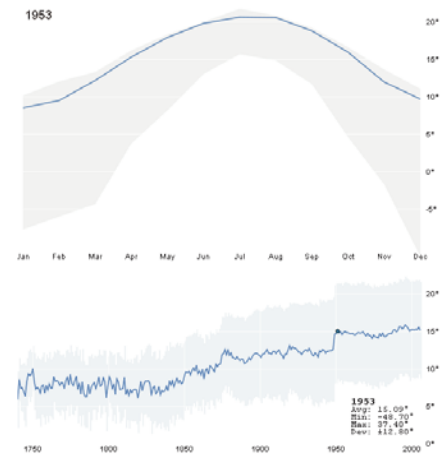
vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text(function(d) d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");

vis.add(pv.Rule).data([0, 10, 20, 30])
  .top(function(d) 300 - 2 * d - 0.5).left(200).right(150)
  .lineWidth(1).strokeStyle("ccc")
  .anchor("right").add(pv.Label)
  .font("italic 10px Georgia")
  .text(function(d) d + "%").textBaseline("center");

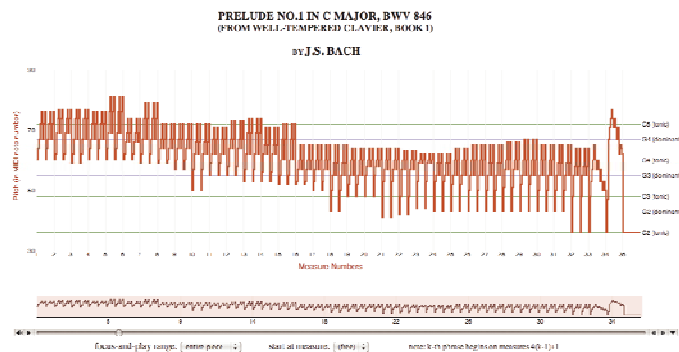
vis.add(pv.Line).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("r")
  .add(pv.Label)
  .top(function(d) 5 + tmp(d))
  .text(function(d) d.temp + "°C").date.substr(0,6)
  .textBaseline("top").font("italic 10px Georgia");
```



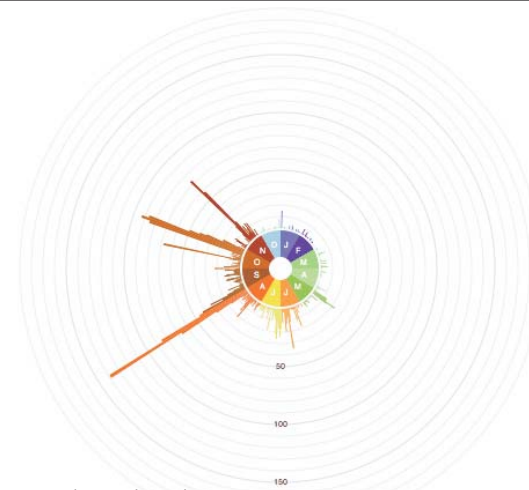
Bullet Charts | Clint Ivy



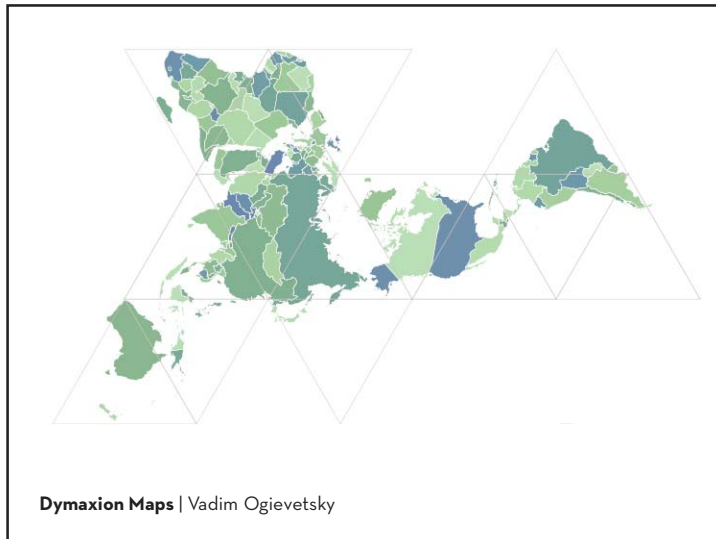
Climate Graph | Robert Kosara



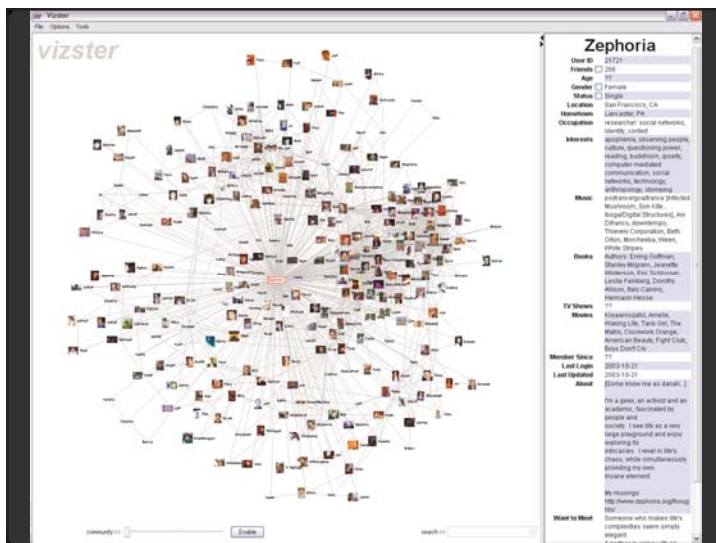
Bach's Prelude #1 in C Major | Jieun Oh



FlickrSeason | Ken-Ichi Ueda



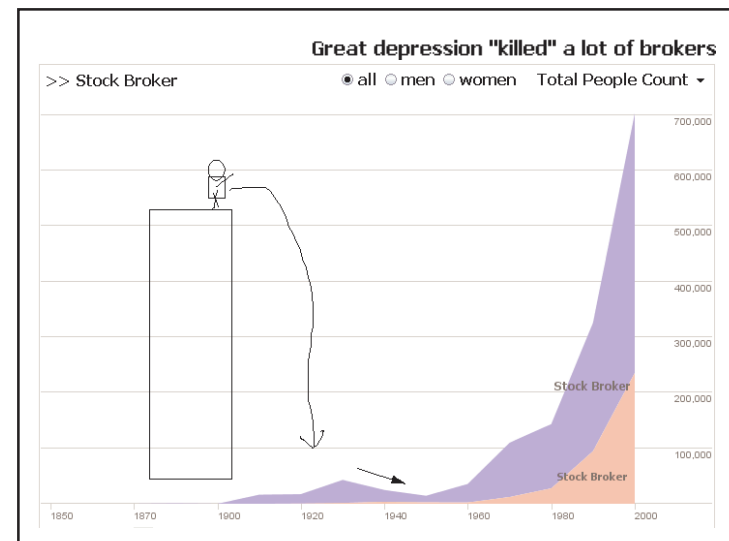
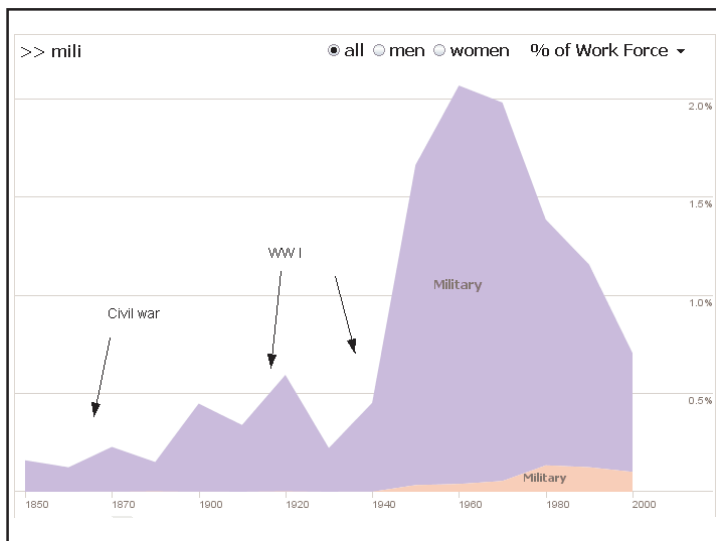
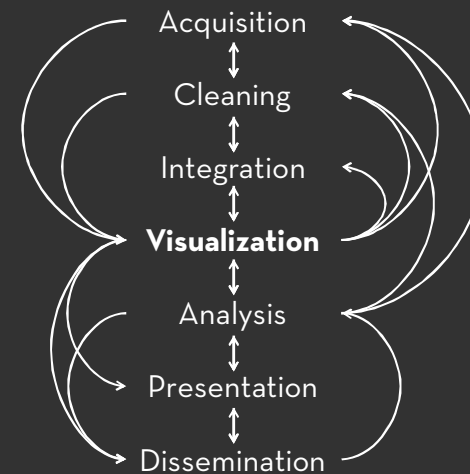
The Reconciliation of Formal and Graphical Methods

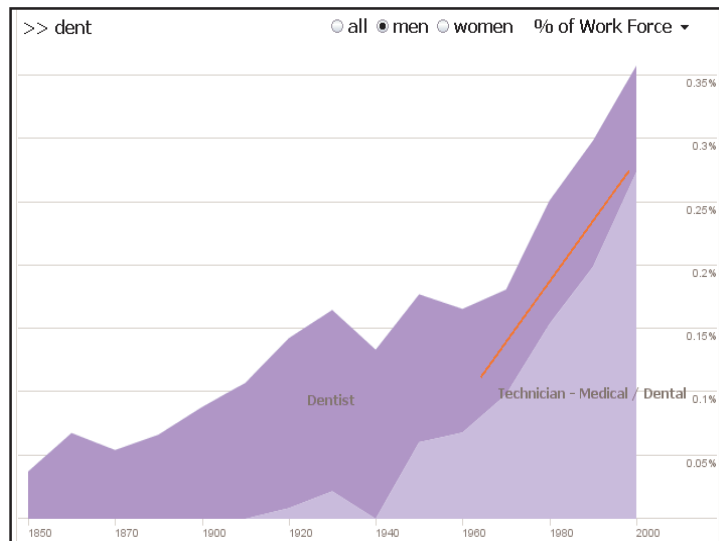


The Data Life-Cycle

While it is often most helpful to “plot the data,” this is rarely enough. We need also to “plot the results of analysis” as a routine matter. There is often more analysis than there was data.

J. W. Tukey, *The Future of Data Analysis*, 1962.

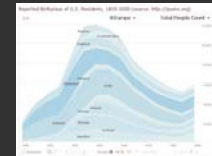




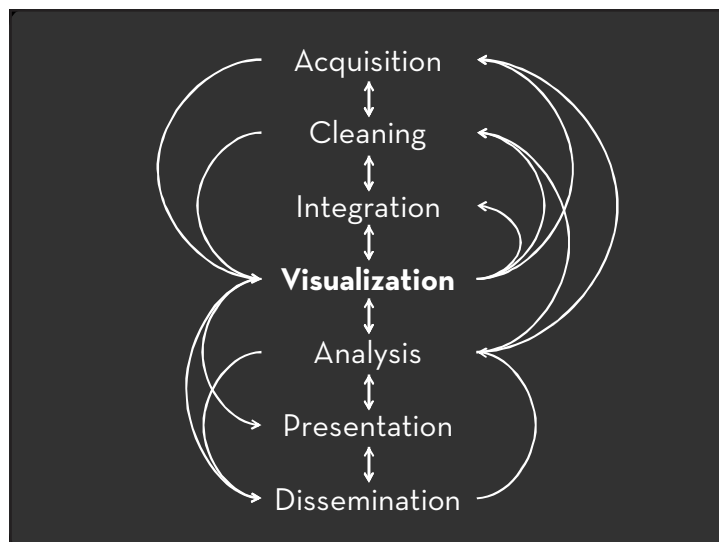
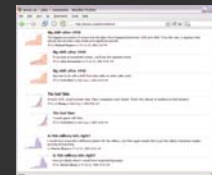
Voyagers and Voyeurs

Complementary faces of analysis

Voyager – focus on visualized data
Active engagement with the data
Serendipitous comment discovery



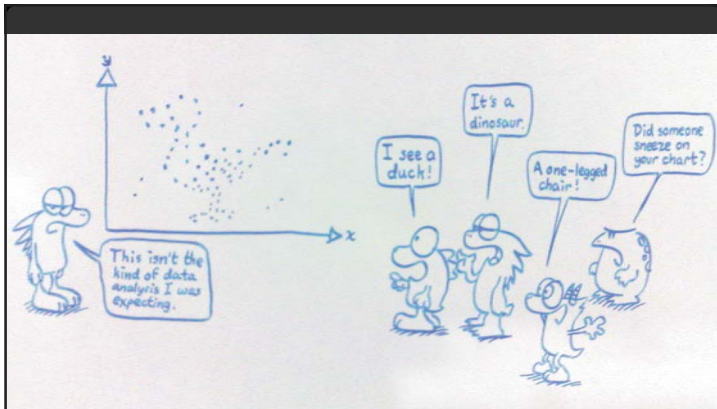
Voyeur – focus on comment listings
Investigate others' explorations
Find people and topics of interest
Catalyze new explorations



The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids.

Because now we really do have **essentially free and ubiquitous data**. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, *The McKinsey Quarterly*, Jan 2009



Jeffrey Heer Stanford University
<http://hci.stanford.edu/jheer>