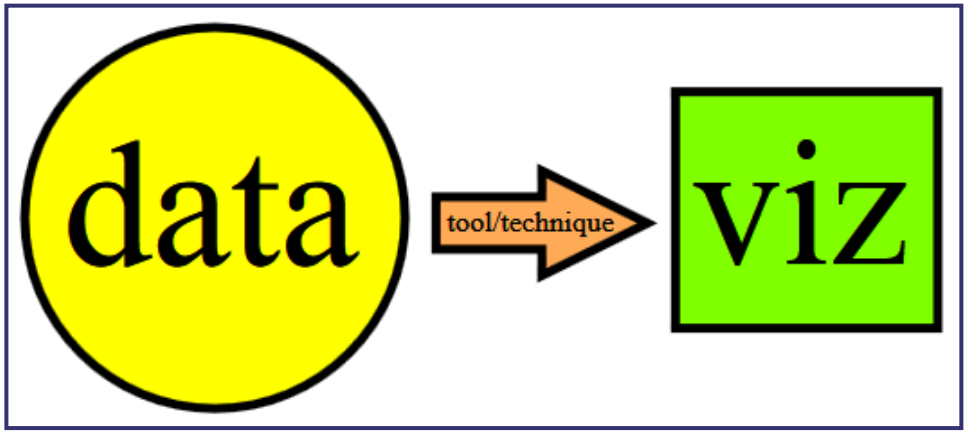
Visualization 

of data

What are we visualizing, and why?

Data!



Data visualization ("data viz") involves (the study of) tools and techniques for turning data into images/graphics - to obtain BETTER INSIGHT into the data.

In other words, this is about graphical depictions of data. Why do it? To understand, communicate, act/decide.

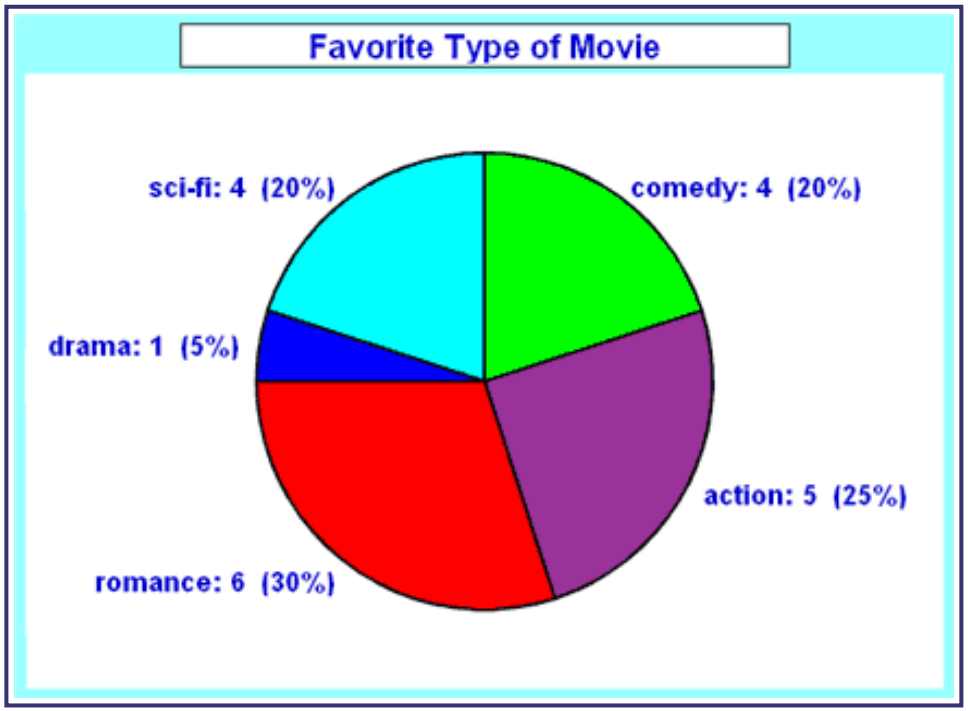
In what follows, we are going to look at what can be visualized, and how. Note: it's not all 'Big Data' viz, it's not all

'mined' results either.

Also, data viz is distinct from ViSC (Visualization in Scientific Computing).

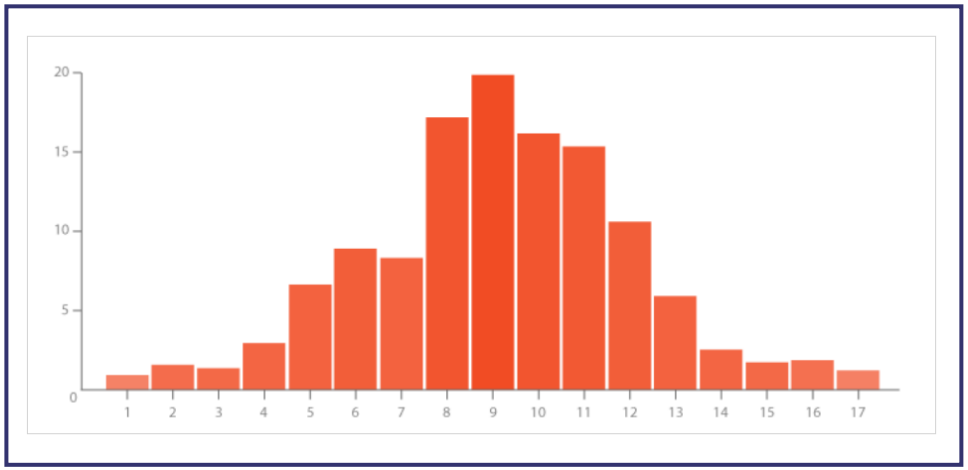
Viz: a single variable

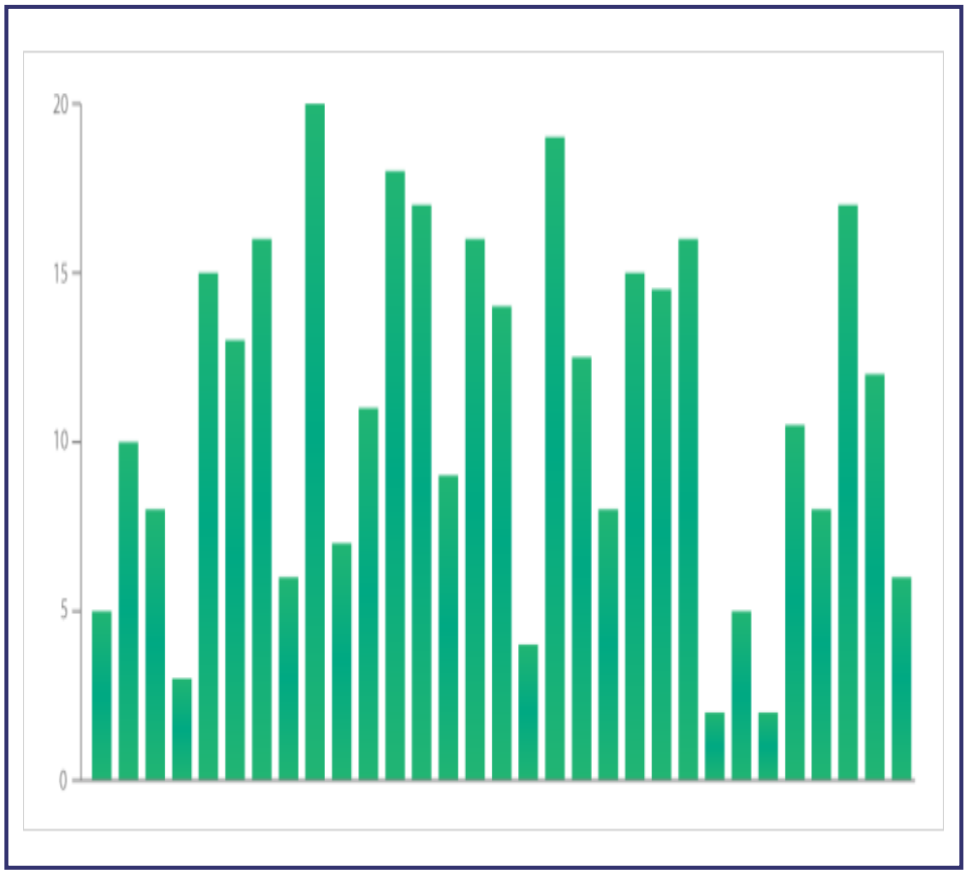
Classically, a pie-chart can be used to express relative fractions of a quantity:



Creating your own pie-chart is simple..

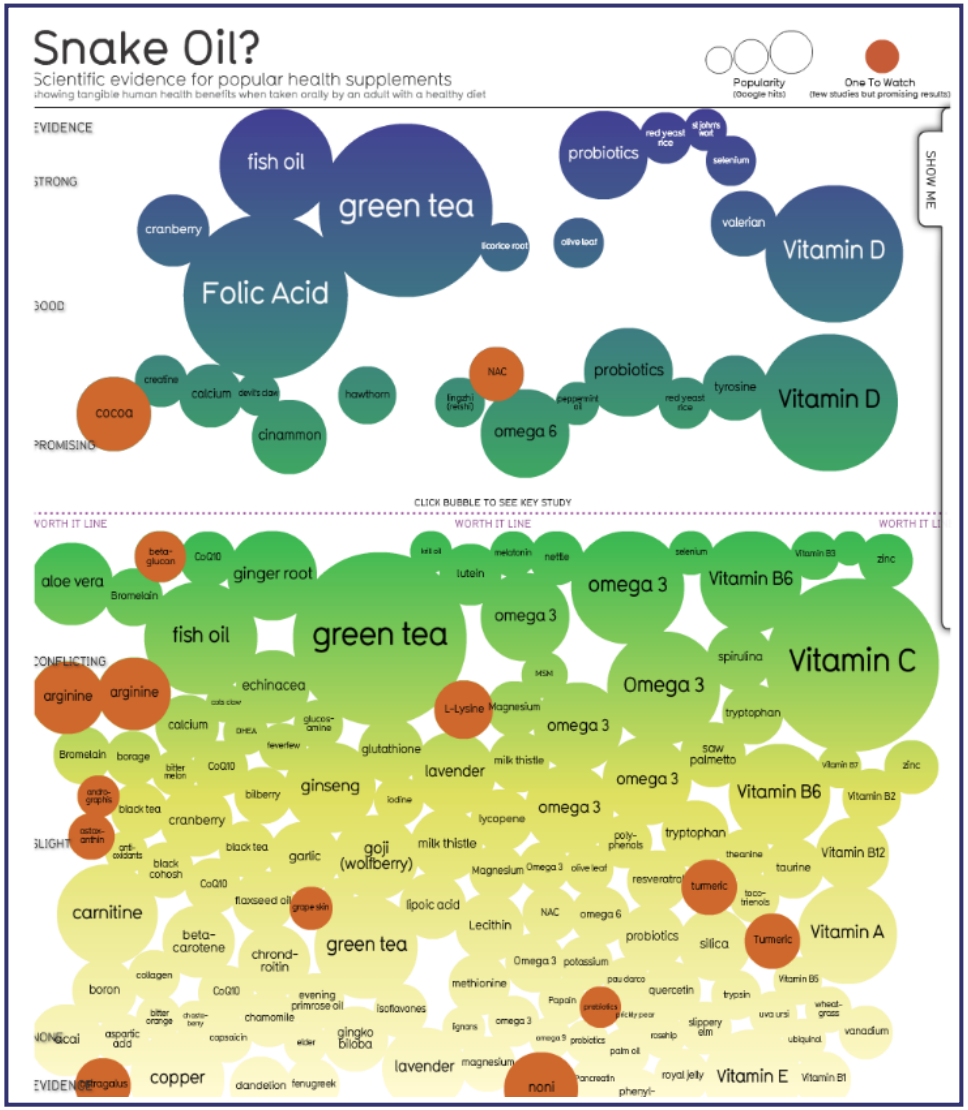
A histogram/bar-chart can be used as well:


20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17


20
15
10
5
0

Also lookup: double histogram, density plot.

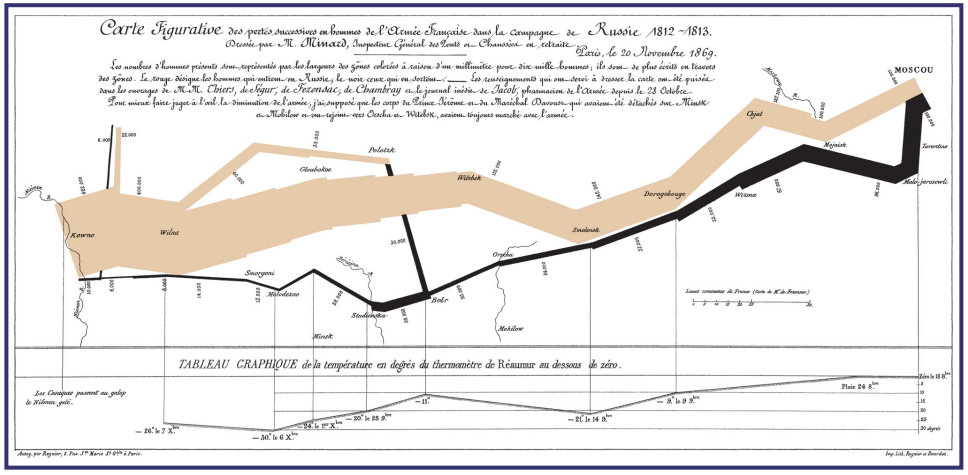
Bubble plots are also useful:



Wordles can be used to indicate relative strengths of keywords/topics. It is easy to create your own.

Multivariate data - possibly the BEST viz ever?!

Edward Tufte calls the graphic below, ""The greatest statistical graphic ever drawn". 'Charles Minard's 1869 chart showing the number of men in Napoleon's 1812 Russian campaign army, their movements, as well as the temperature they encountered on the return path.' (Wikipedia)

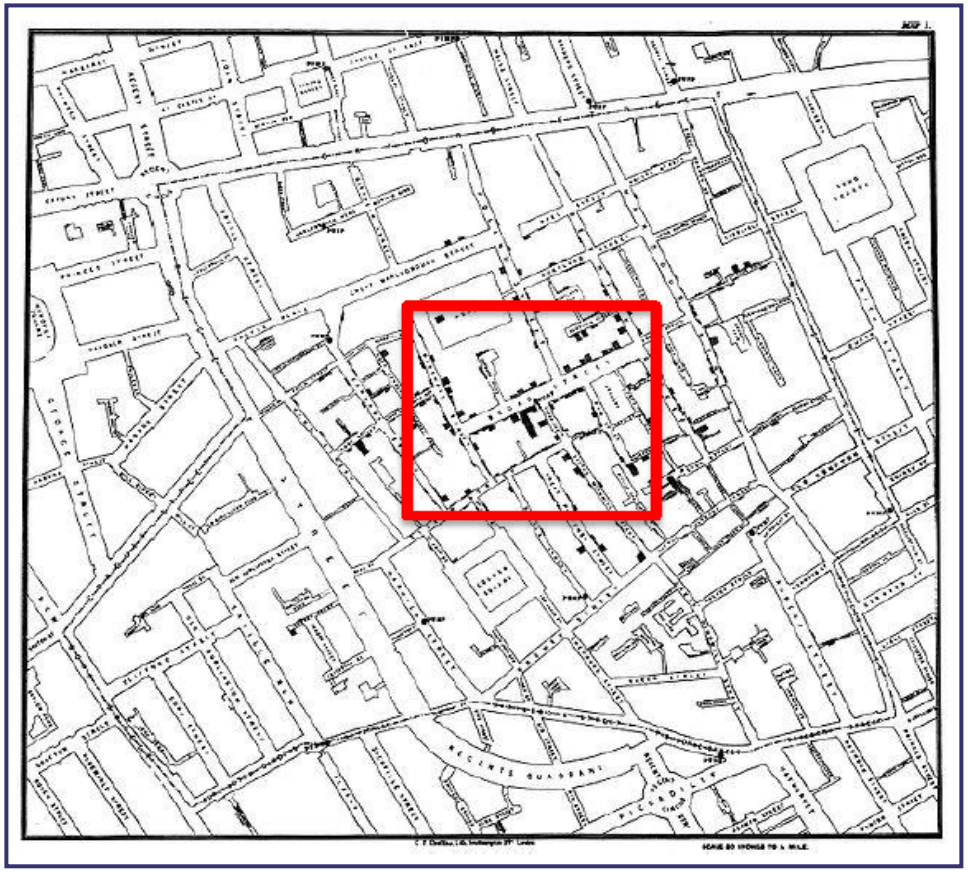


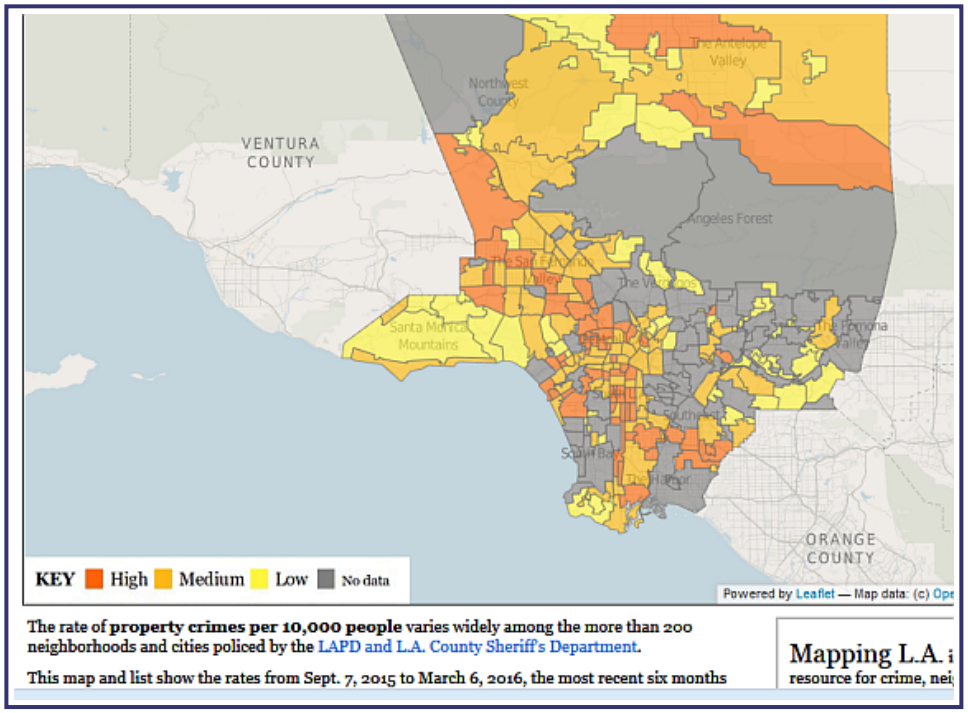
Viz: spatial data

Plotting spatial data (eg. incidence locations) on a map reveals patterns/trends in a direct way.

Cholera outbreak map, 1854, London [plotting of reported cases as black dots reveals the source of the outbreak

[highlighted region]):



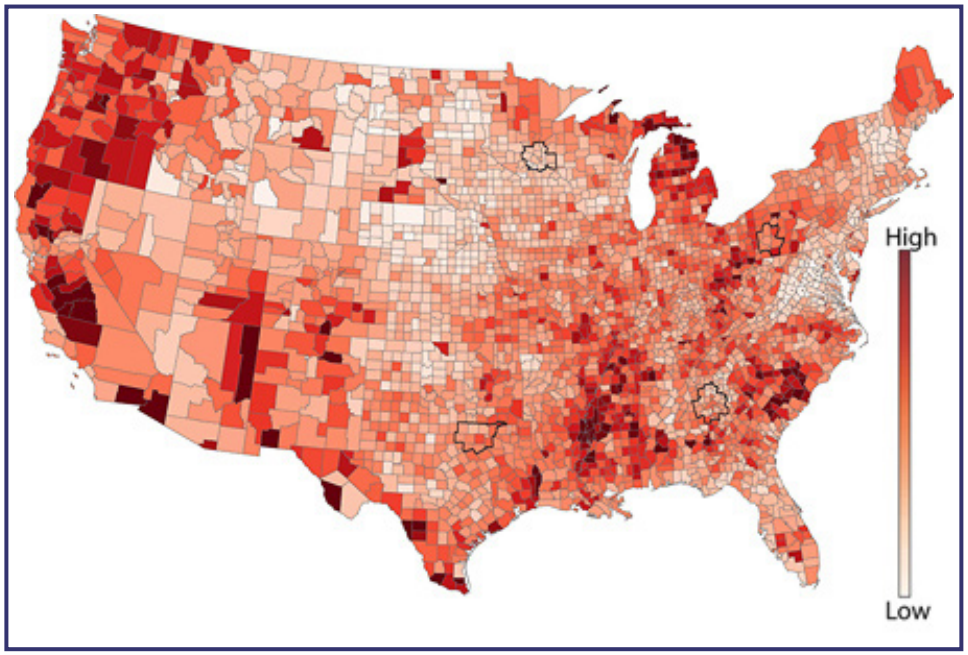


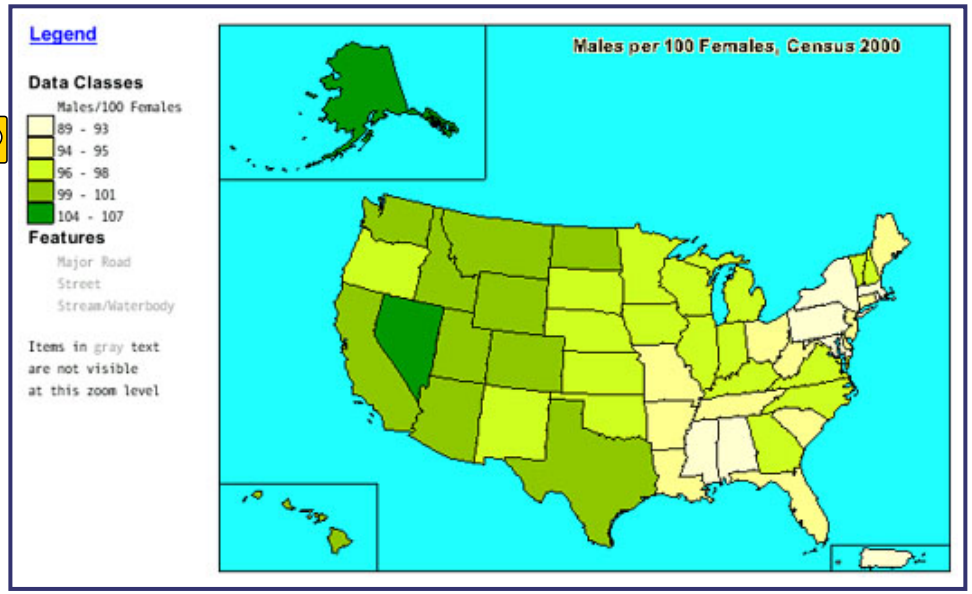
It is quite useful for planning purposes, to visualize data over a map - eg. here are Starbucks locations..

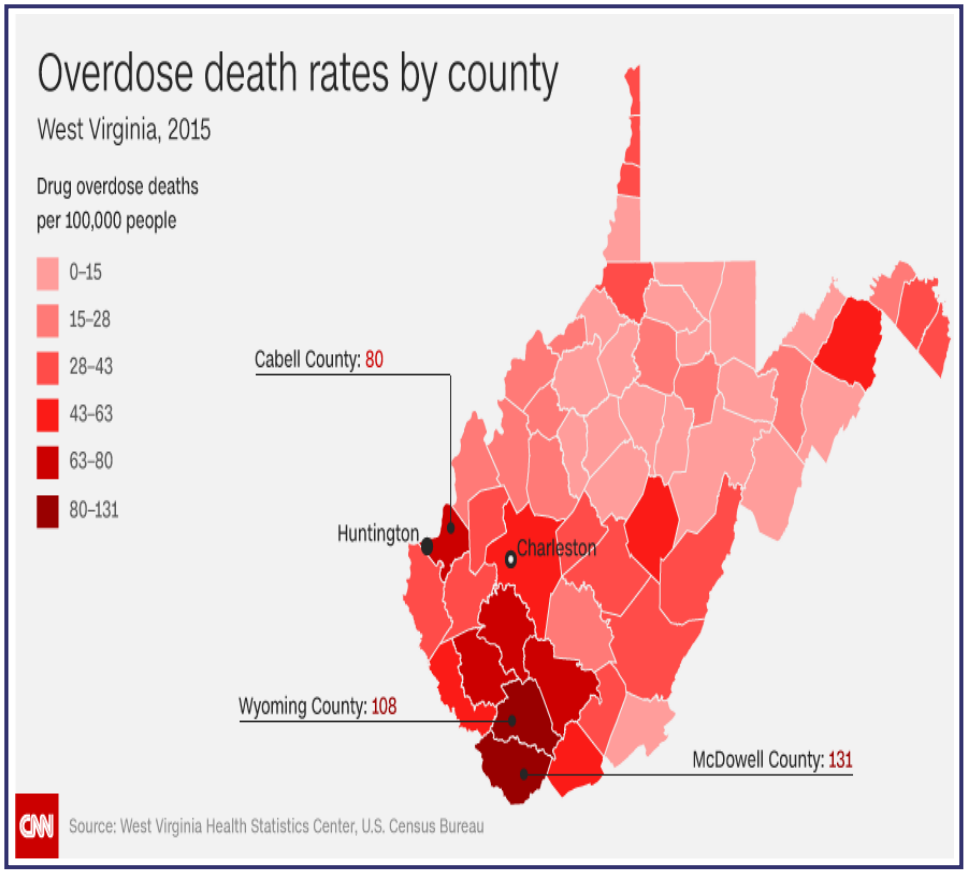
Mined data, eg. associations, can be superposed over a map, eg. in a grocery store. Results can be used to redo the layout. A related topic is product placement.

As we saw earlier, a choropleth map shows spatial, aggregated data (that covers the entire region shown).

These come in two varieties - unclassed (continuous scale), classed (discrete ranges).



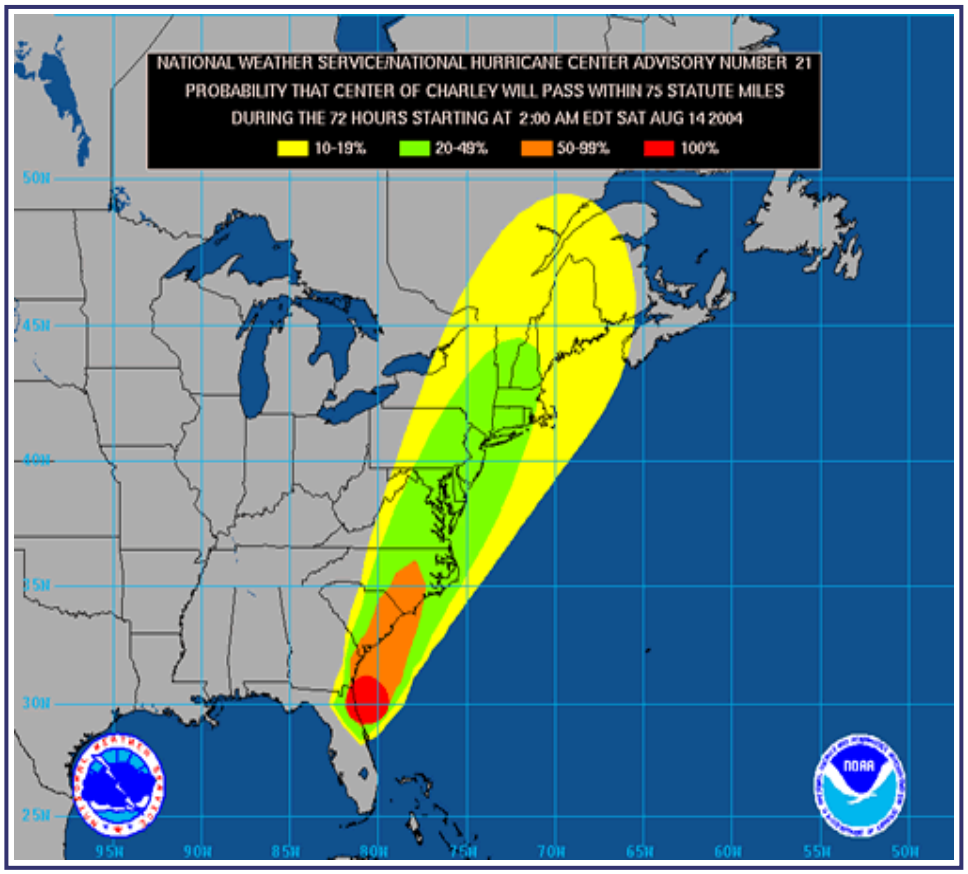

Legend
Data Classes
Males/100 Females
89 - 93
94 - 95
96 - 98
99 - 101
104 - 107
Features
Major Road
Street
Stream/Waterbody
Items in gray text are not visible at this zoom level
Males per 100 Females, Census 2000


Overdose death rates by county
West Virginia, 2015
Drug overdose deaths per 100,000 people
0–15
15–28
28–43
43–63
63–80
80–131
Cabell County: 80
Huntington
Charleston
Wyoming County: 108
McDowell County: 131
CNN
Source: West Virginia Health Statistics Center, U.S. Census Bureau

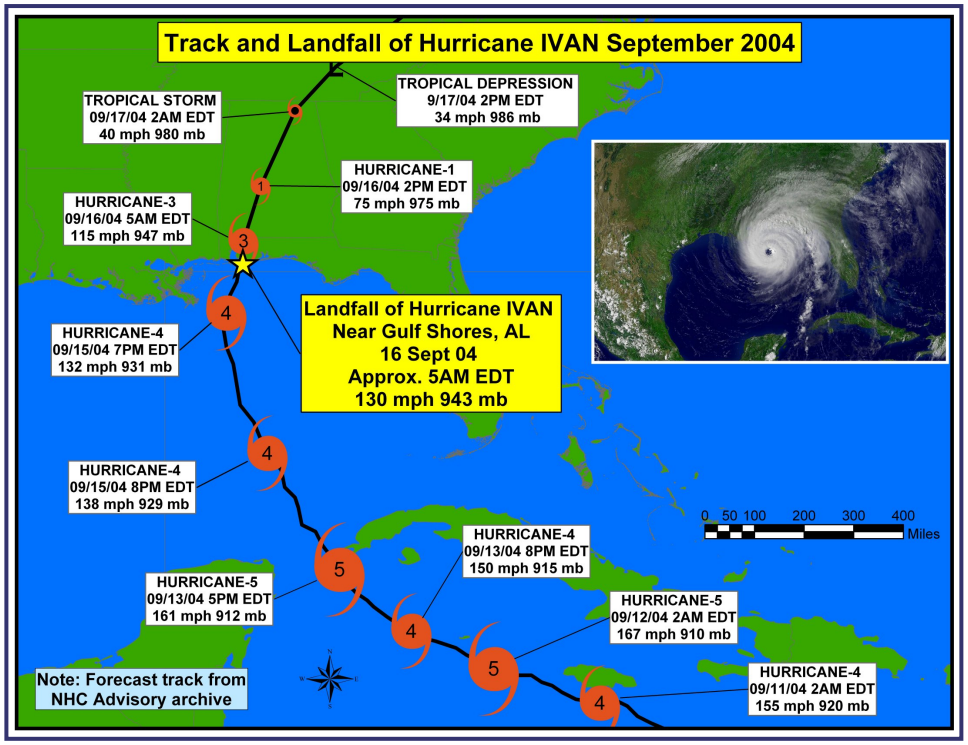
Viz: spatio-temporal data

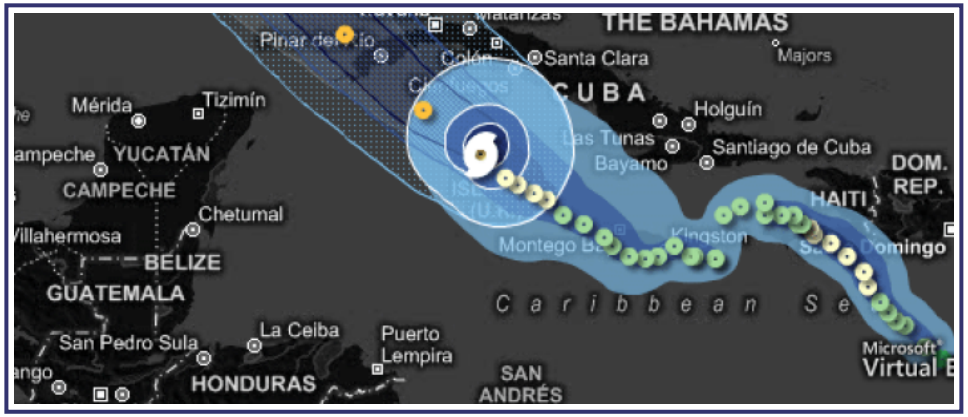
Superposing time-varying data on a map reveals course, trends, etc. Such data could be visualized as animations, too.

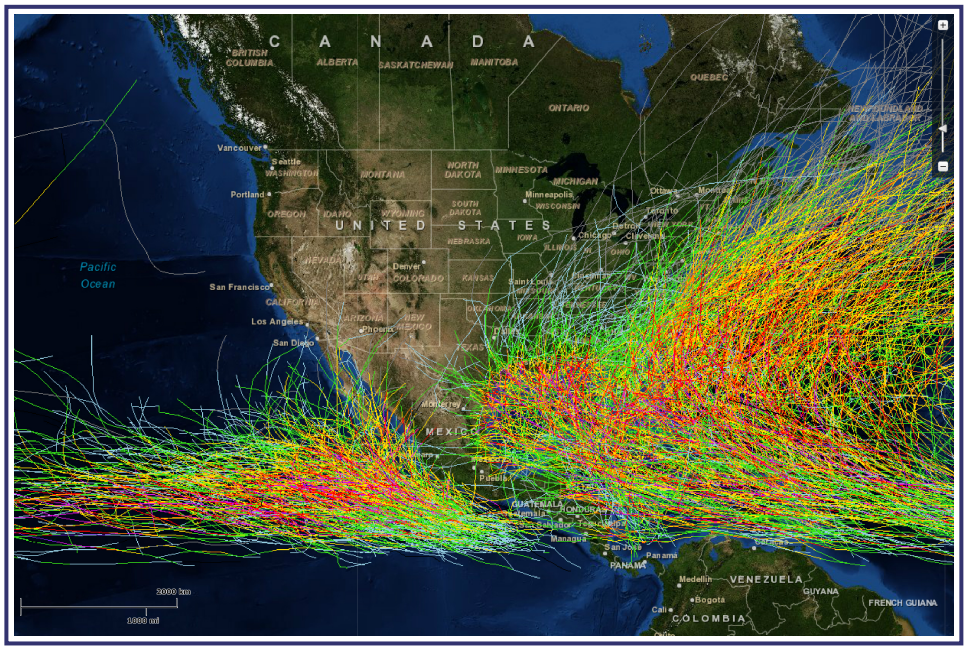
H1-B visas - who gets them? :) [scroll to the bottom of the report]




HURRICANE IKE
PROJECTED PATH
MON AM
SUN AM
SAT AM
FRI AM
THU AM
DEVIATIONS IN TRACK AND/OR INTENSITY FROM CURRENT PROJECTIONS COULD RESULT IN SIGNIFICANT DIFFERENCES FROM THE INFORMATION ON THIS GRAPHIC
The Weather Channel
weather.com
10 Sep 2008 10:41 GMT / 10 Sep 2008 06:41 AM EDT


Track and Landfall of Hurricane IVAN September 2004
TROPICAL STORM
09/17/04 2AM EDT
40 mph 980 mb
TROPICAL DEPRESSION
9/17/04 2PM EDT
34 mph 986 mb
HURRICANE-1
09/16/04 2PM EDT
75 mph 975 mb
HURRICANE-3
09/16/04 5AM EDT
115 mph 947 mb
HURRICANE-4
09/15/04 7PM EDT
132 mph 931 mb
Landfall of Hurricane IVAN
Near Gulf Shores, AL
16 Sept 04
Approx. 5AM EDT
130 mph 943 mb
HURRICANE-4
09/15/04 8PM EDT
138 mph 929 mb
HURRICANE-5
09/13/04 5PM EDT
161 mph 912 mb
HURRICANE-4
09/13/04 8PM EDT
150 mph 915 mb
HURRICANE-5
09/12/04 2AM EDT
167 mph 910 mb
HURRICANE-4
09/11/04 2AM EDT
155 mph 920 mb
0 50 100 200 300 400
Miles
Note: Forecast track from
NHC Advisory archive


THE BAHAMAS
Pinar del Río
Colón
Santa Clara
Majors
CUBA
Mérida
Tizimín
Holguín
Las Tunas
Bayamo
Santiago de Cuba
YUCATÁN
CAMPECHE
DOM. REP.
HAITI
Chetumal
Villahermosa
Kingston
BELIZE
GUATEMALA
Caribbean Sea
La Ceiba
Puerto Lempira
San Pedro Sula
HONDURAS
SAN ANDRÉS
Microsoft Virtual


CANADA
BRITISH COLUMBIA
ALBERTA
SASKATCHEWAN
MANITOBA
ONTARIO
QUEBEC
Vancouver
Seattle
Portland
Minneapolis
Toronto
Ottawa
Montreal
Chicago
Cleveland
Detroit
Pittsburgh
Denver
San Francisco
Los Angeles
San Diego
Phoenix
Pacific Ocean
UNITED STATES
MEXICO
Monterrey
Managua
San José
Panama
PANAMA
Medellín
VENEZUELA
GUYANA
FRENCH GUIANA
Bogotá
Cali
COLOMBIA

Viz: interactivity

Being able to INTERACT with data provides MORE understanding - we can selectively turn items on/off, drill down or roll up, explore the time dimension..

Crime map, USC area :(

Crimean War casualties, by Florence Nightingale (!)

NOAA, Historic Hurricane Tracks

Viz: real-time!

Real-time visualization provides a level of immediacy/freshness/relevance/interest that is simply absent in non-real-time data..

World population growth [even more real-time stats!].

Local traffic.

Earthquakes!

Stocks.

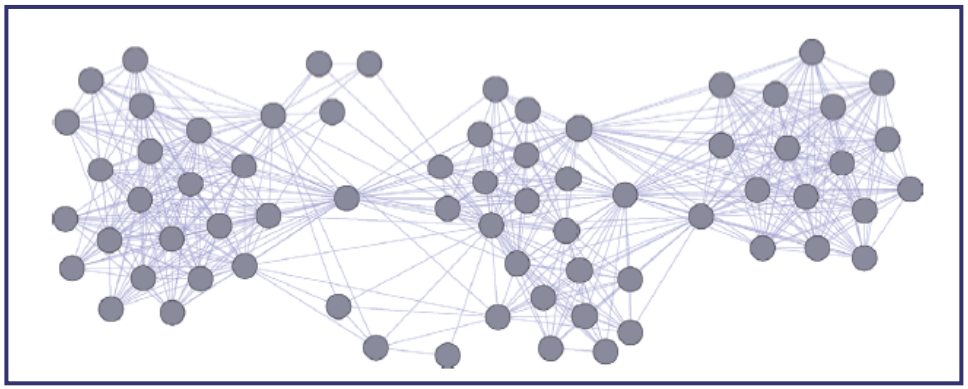
Cybercrimes (!!).

Your to-do: look for real-time weather (eg. Doppler radar) sites. What other real-time updated data can you find or think of?

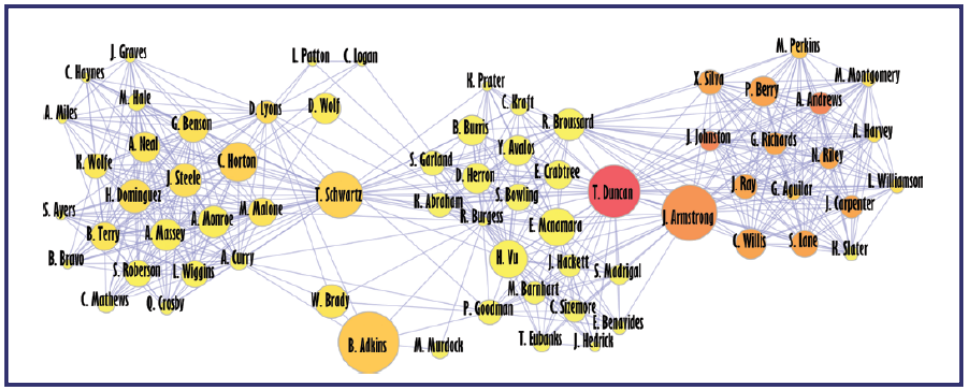
Viz: networks - node attrs

Network visualization is a very popular category - shows **RELATIONSHIPS between entities.**

A diagram that maps email exchanges between family members:

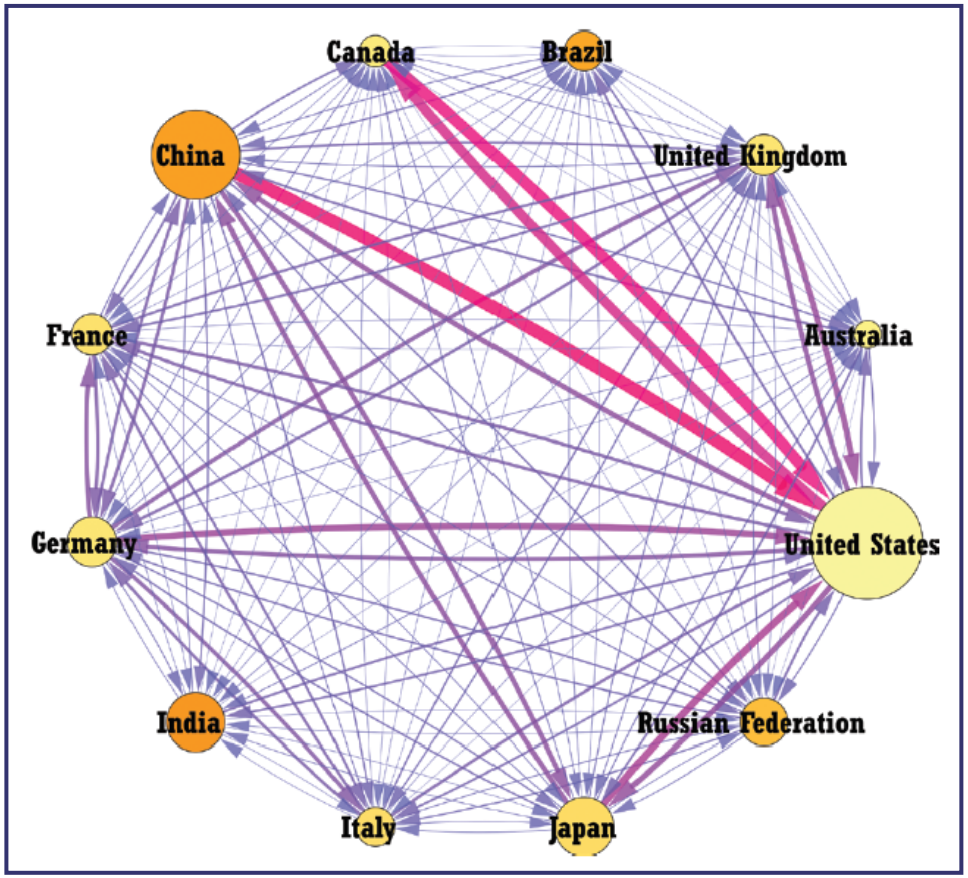


In the above, we are lacking DETAIL that can be **added** using extra **ATTRIBUTES, and LABELS.** Here is an enriched version that uses attrs and labels:


J. Graves
C. Haynes
M. Hale
A. Miles
G. Benson
A. Neal
K. Wolfe
J. Steele
H. Dominguez
S. Ayers
B. Terry
A. Massey
B. Bravo
S. Roberson
L. Wiggins
C. Mathews
Q. Crosby
L. Patton
C. Logan
D. Lyons
D. Wolf
C. Horton
T. Schwartz
A. Monroe
M. Malone
A. Curry
W. Brady
B. Adkins
M. Murdock
K. Prater
C. Kraft
B. Burris
R. Broussard
S. Garland
Y. Avalos
D. Herron
E. Crabtree
S. Bowling
R. Abraham
R. Burgess
T. Duncan
E. Mcnamara
H. Vu
J. Hackett
S. Madrigal
M. Barnhart
P. Goodman
C. Sizemore
E. Benavides
T. Eubanks
J. Hedrick
M. Perkins
X. Silva
P. Berry
M. Montgomery
A. Andrews
J. Johnston
G. Richards
A. Harvey
N. Riley
J. Ray
G. Aguilar
L. Williamson
J. Carpenter
J. Armstrong
C. Wills
S. Lane
H. Slater

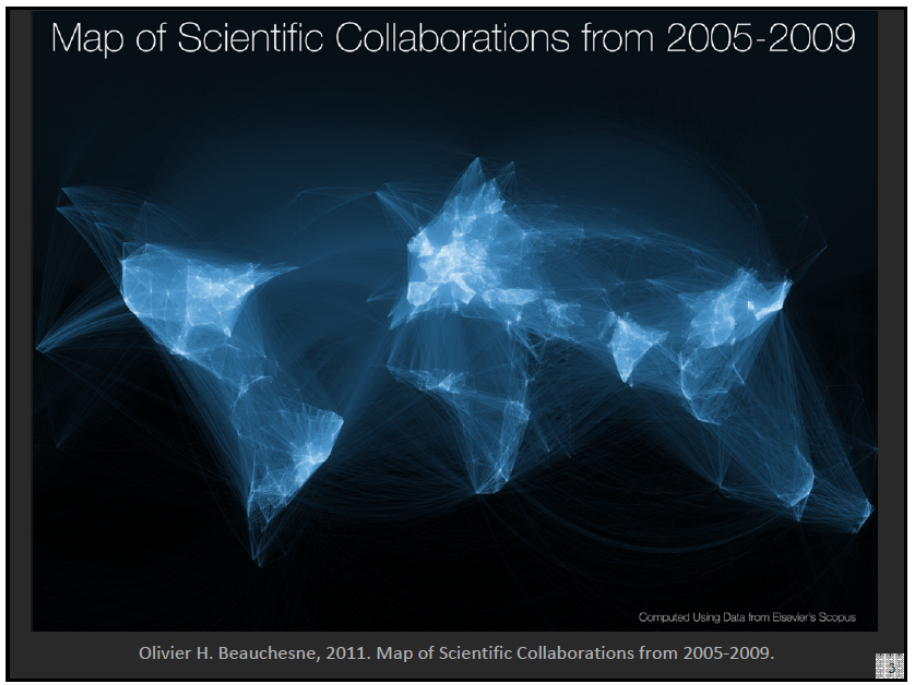
Viz: networks - edge attrs

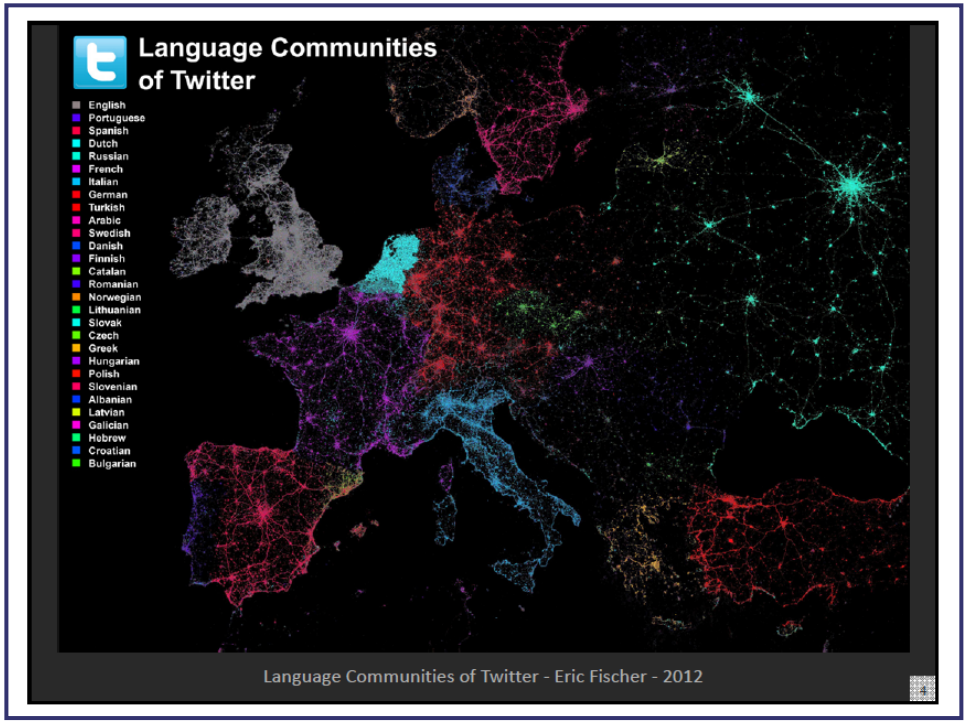
We can use edge attrs (type, eg. arrows, dashes.., color, thickness etc.) to quantify data. The diagram below shows trade quantities between countries (2012, top 12 countries as per GDP):

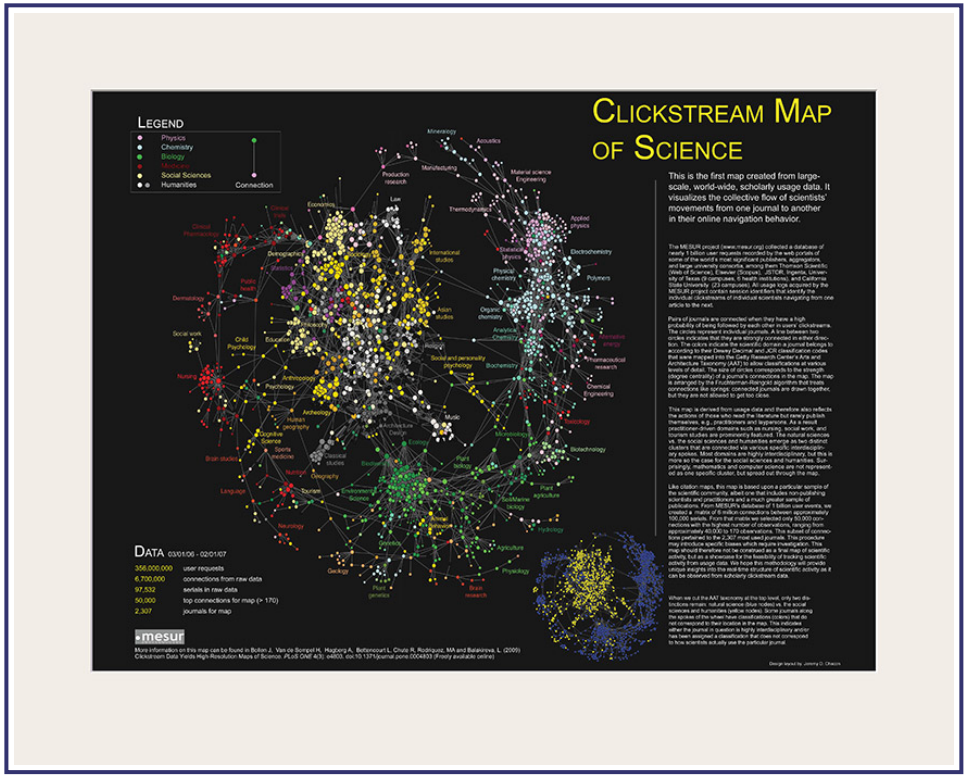


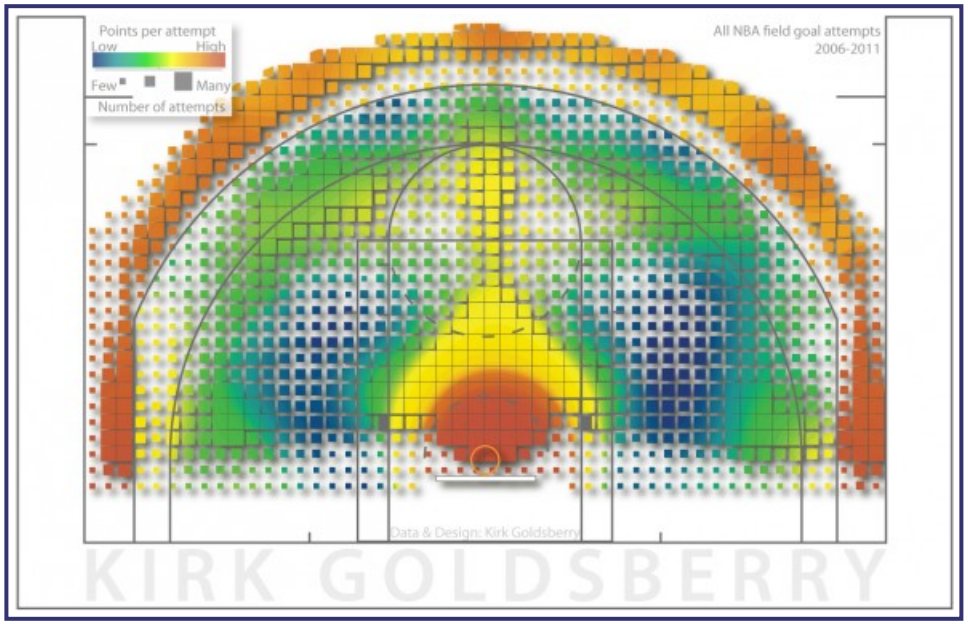
Grab Bag of goodies

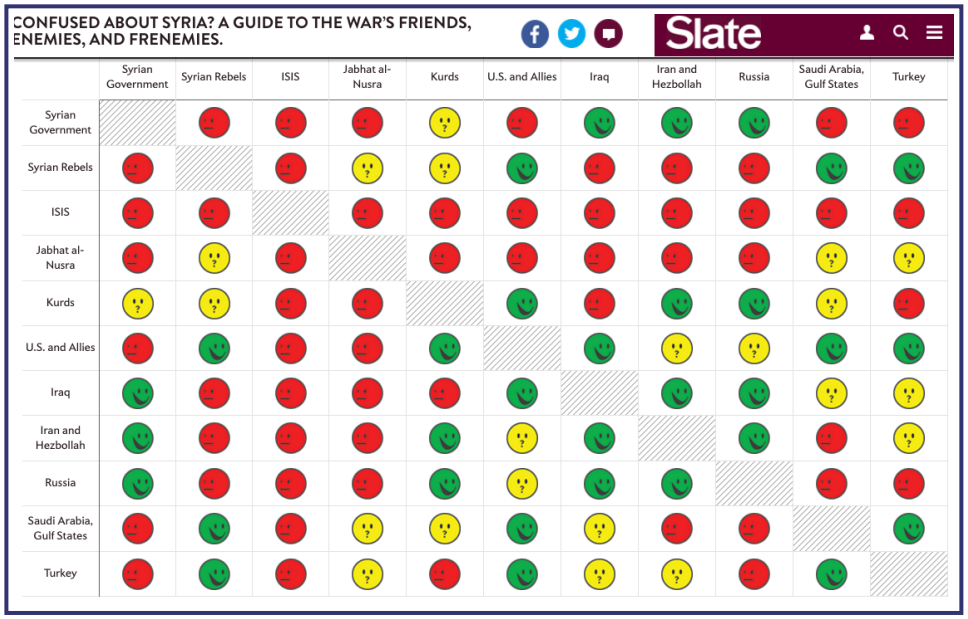
An assortment of 'cool' (visually appealing) and USEFUL data visualizations:

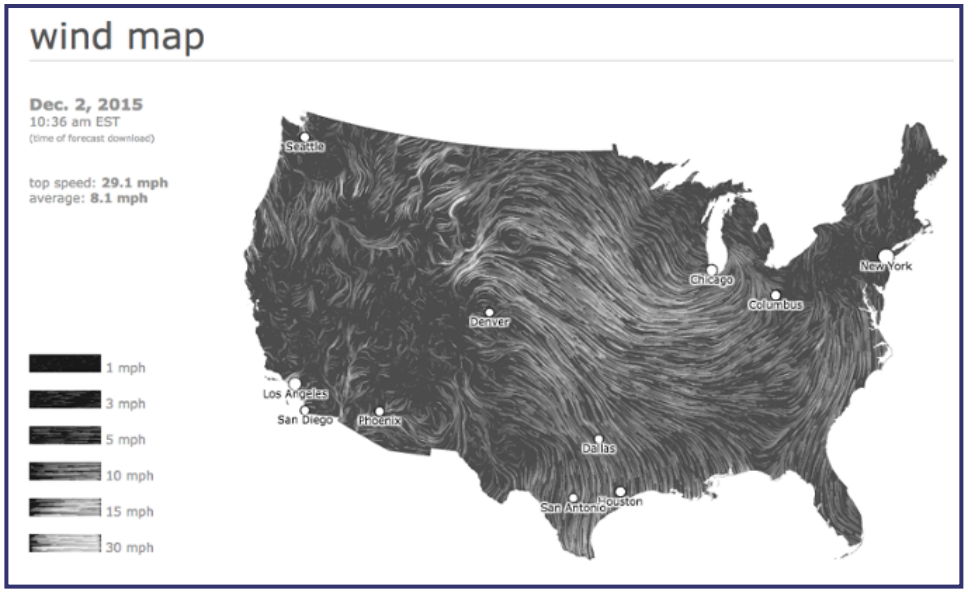



Language Communities of Twitter
English
Portuguese
Spanish
Dutch
Russian
French
Italian
German
Turkish
Arabic
Swedish
Danish
Finnish
Catalan
Romanian
Norwegian
Lithuanian
Slovak
Czech
Greek
Hungarian
Polish
Slovenian
Albanian
Latvian
Galician
Hebrew
Croatian
Bulgarian
Language Communities of Twitter - Eric Fischer - 2012


LEGEND
Physics
Chemistry
Biology
Medicine
Social Sciences
Humanities
Connection
CLICKSTREAM MAP OF SCIENCE
This is the first map created from large-scale, world-wide, scholarly usage data. It visualizes the collective flow of scientists' movements from one journal to another in their online navigation behavior.
DATA 03/01/06 - 02/01/07
358,000,000 user requests
6,700,000 connections from raw data
97,532 serials in raw data
50,000 top connections for map (> 170)
2,307 journals for map
•mesur

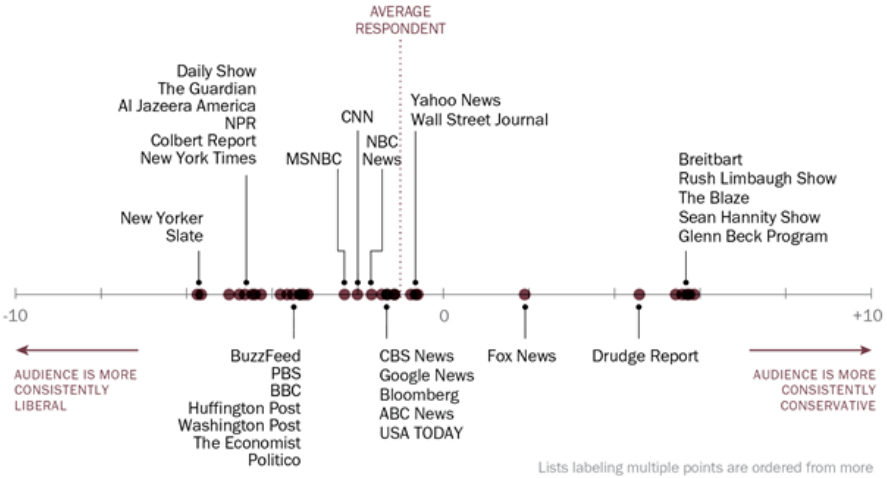

Points per attempt
Low
High
Few
Many
Number of attempts
All NBA field goal attempts
2006-2011
Data & Design: Kirk Goldsberry
KIRK GOLDSBERRY


CONFUSED ABOUT SYRIA? A GUIDE TO THE WAR'S FRIENDS, ENEMIES, AND FRENEMIES.
Slate
Syrian Government
Syrian Rebels
ISIS
Jabhat al-Nusra
Kurds
U.S. and Allies
Iraq
Iran and Hezbollah
Russia
Saudi Arabia, Gulf States
Turkey


wind map
Dec. 2, 2015
10:36 am EST
(time of forecast download)
top speed: 29.1 mph
average: 8.1 mph
1 mph
3 mph
5 mph
10 mph
15 mph
30 mph
Seattle
Chicago
New York
Columbus
Denver
Los Angeles
San Diego
Phoenix
Dallas
San Antonio
Houston

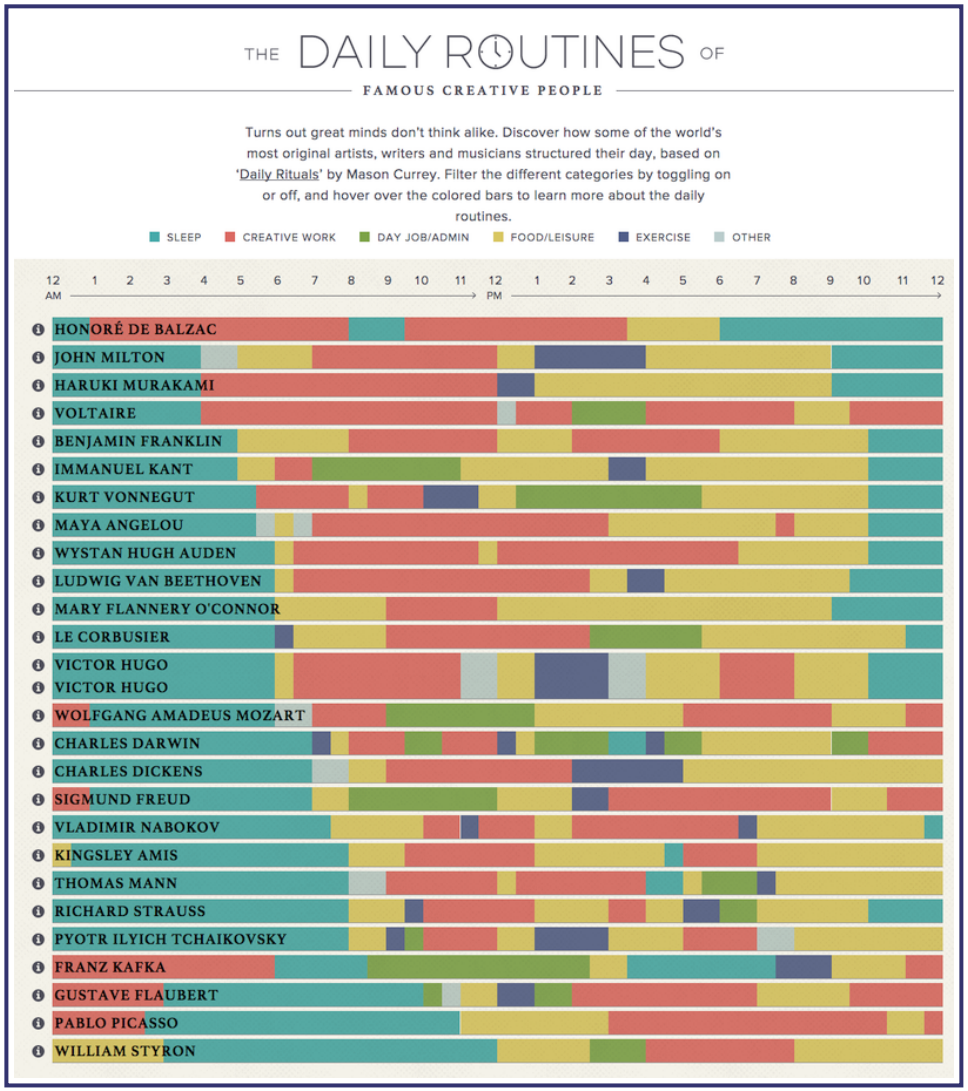
Ideological Placement of Each Source's Audience

Average ideological placement on a 10-point scale of ideological consistency of those who got news from each source in the past week...



American Trends Panel (wave 1). Survey conducted March 19-April 29, 2014. Q22. Based on all web respondents. Ideological consistency based on a scale of 10 political values questions (see About the Survey for more details.) ThinkProgress, DailyKos, Mother Jones, and The Ed Schultz Show are not included in this graphic because audience sample sizes are too small to analyze.

PEW RESEARCH CENTER

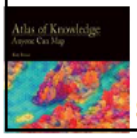

THE DAILY ROUTINES OF
FAMOUS CREATIVE PEOPLE
Turns out great minds don't think alike. Discover how some of the world's most original artists, writers and musicians structured their day, based on 'Daily Rituals' by Mason Currey. Filter the different categories by toggling on or off, and hover over the colored bars to learn more about the daily routines.
SLEEP
CREATIVE WORK
DAY JOB/ADMIN
FOOD/LEISURE
EXERCISE
OTHER
12 AM 1 2 3 4 5 6 7 8 9 10 11 12 PM 1 2 3 4 5 6 7 8 9 10 11 12
HONORÉ DE BALZAC
JOHN MILTON
HARUKI MURAKAMI
VOLTAIRE
BENJAMIN FRANKLIN
IMMANUEL KANT
KURT VONNEGUT
MAYA ANGELOU
WYSTAN HUGH AUDEN
LUDWIG VAN BEETHOVEN
MARY FLANNERY O'CONNOR
LE CORBUSIER
VICTOR HUGO
VICTOR HUGO
WOLFGANG AMADEUS MOZART
CHARLES DARWIN
CHARLES DICKENS
SIGMUND FREUD
VLADIMIR NABOKOV
KINGSLEY AMIS
THOMAS MANN
RICHARD STRAUSS
PYOTR ILYICH TCHAIKOVSKY
FRANZ KAFKA
GUSTAVE FLAUBERT
PABLO PICASSO
WILLIAM STYRON

Where is the 'science'??

Data viz is an art AND a science - there are principles, choices, tradeoffs. As for the principles, these encompass diverse disciplines such as visual perception, color theory, composition (grouping, contrast, harmony, symmetry..), design elements (line, tone, form, texture..), semiotics, etc.

As for what type of graphic to generate for a given type of data analysis, we can follow the guidelines here (from 'Atlas of Knowledge'):

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion



See page 24

Visualization Types (Reference Systems)

1. **Charts:** No reference system—e.g., Wordle.com, pie charts
2. **Tables:** Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.
3. **Graphs:** Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.
4. **Geospatial maps:** Use latitude and longitude reference system. World or city maps.
5. **Network layouts:** Node position might depends on node attributes or node similarity. **Trees:** hierarchies, taxonomies, genealogies. **Networks:** social networks, migration flows.

HOW to GENERATE data viz??

During the past lectures, we've looked at a few data viz examples (eg. GIS data). Here is a systematic breakdown of ways to create all manner of data viz.

Data science software

- Weka
- KNIME
- RapidMiner

Using code

- R, Shiny
- matplotlib
- d3 (JS) and Protovis
- PGFPlots (LaTeX)

Online tools

- datavisual
- infogram
- Online Charts
- Slemma

Math, analysis and plotting packages

- Mathematica
- MATLAB
- OriginLab [and alternatives]
- good old Excel

3rd party data-viz software

- Periscope
- Tableau
- Qlik
- SiSense
- domo
- JMP (SAS)

AR + VR : revolutionary data viz systems?!

The 'coming' revolution in AR and VR is sure to effect sea changes in 'data viz': visualize/interact with => perceive/be immersed in data!

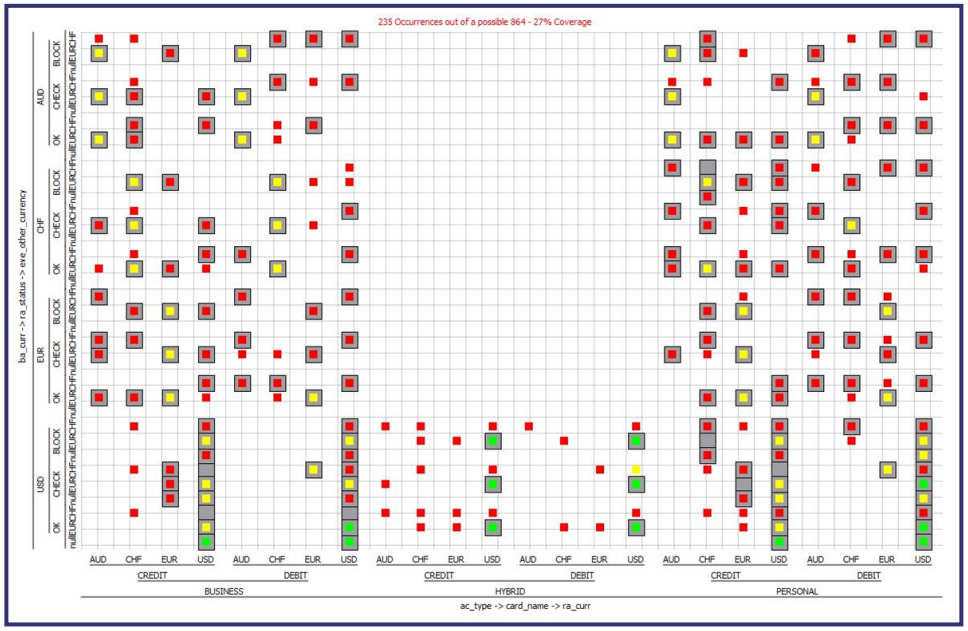
Here is a book on visual data mining, and this is a chapter in it, on using VR to explore data.

In order to make the best use of AR (and VR) for data analysis, we need a thorough understand of how we process visual data.

Great Wave is a company that has created Meta 2, a platform for data-viz using AR. MUCH MORE is possible!

Meta viz

We can even visualize aspects of data that can help us carry out better data analysis - below is a 'coverage map' that shows, well, coverage (missing/inadequate data):



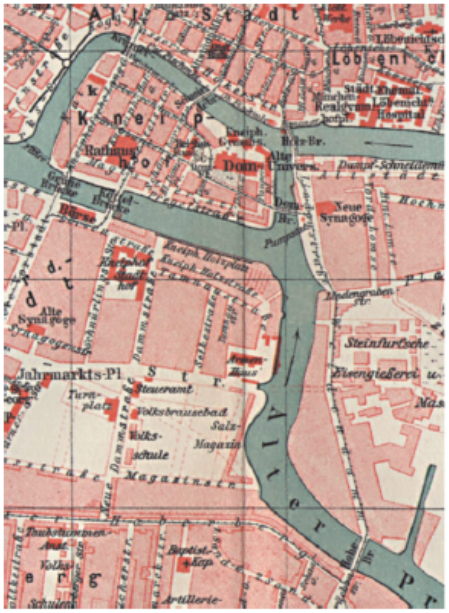
What makes for a good design?

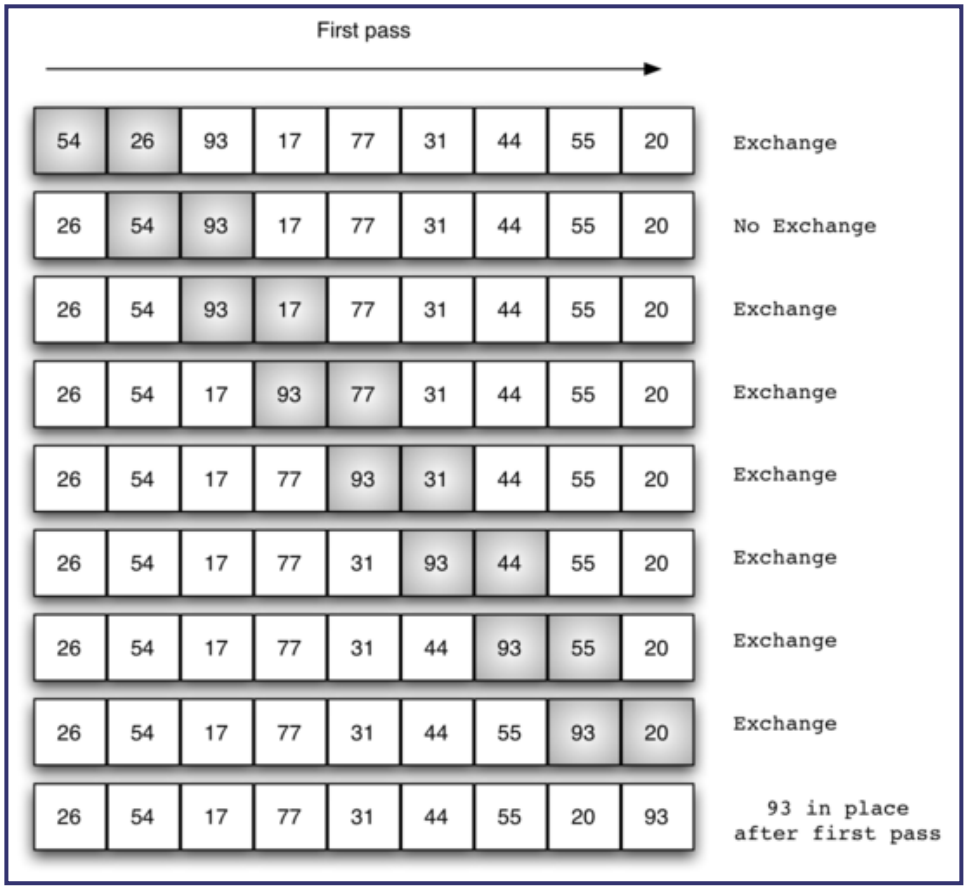
Induce the viewer to think about the substance rather than about methodology, graphic design, the tech of graphic production, or something else. (Edward Tufte)

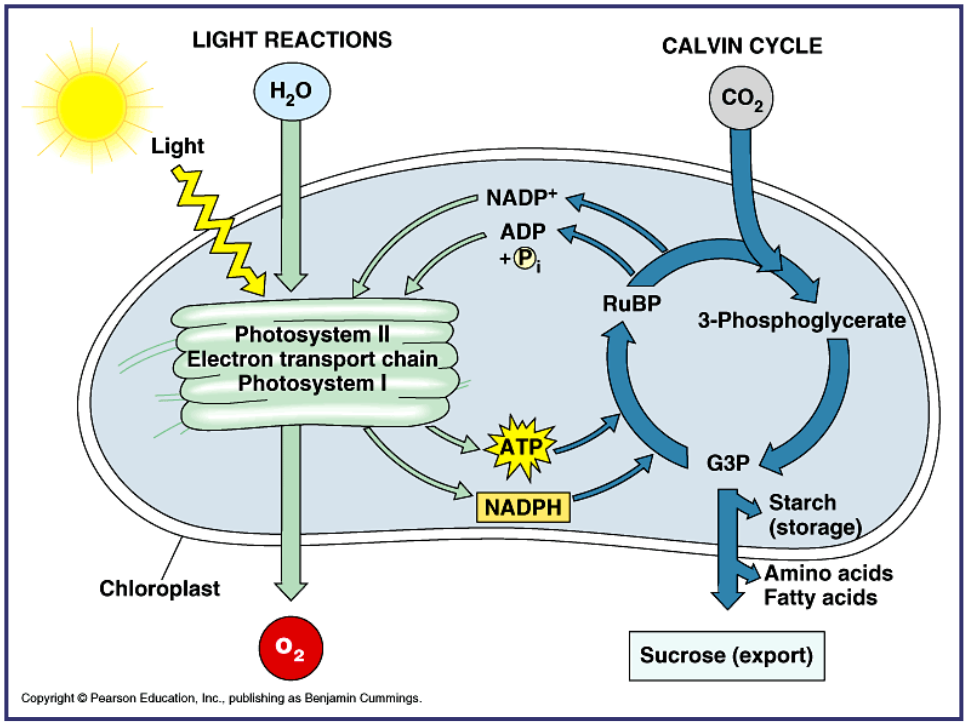
What is NOT data viz?

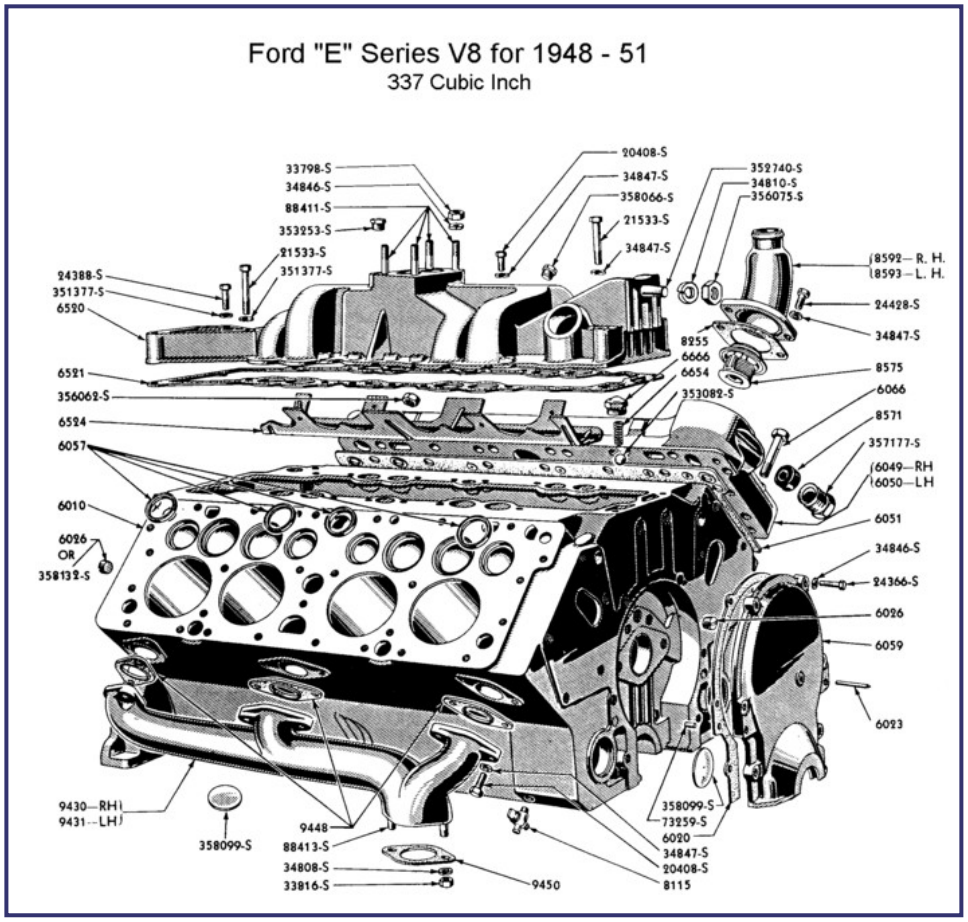
Data viz is specific - involves data=>graphics.

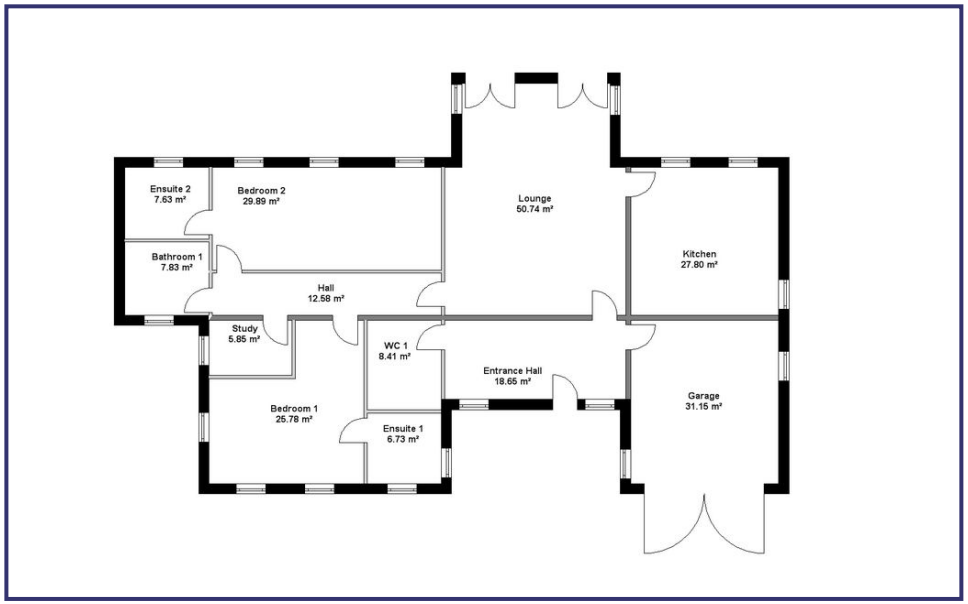
Following are examples of viz, but not necessarily of data.




First pass
54 26 93 17 77 31 44 55 20 Exchange
26 54 93 17 77 31 44 55 20 No Exchange
26 54 93 17 77 31 44 55 20 Exchange
26 54 17 93 77 31 44 55 20 Exchange
26 54 17 77 93 31 44 55 20 Exchange
26 54 17 77 31 93 44 55 20 Exchange
26 54 17 77 31 44 93 55 20 Exchange
26 54 17 77 31 44 55 93 20 Exchange
26 54 17 77 31 44 55 20 93
93 in place after first pass


LIGHT REACTIONS
CALVIN CYCLE
H_2O
CO_2
Light
$NADP^+$
ADP
$+ P_i$
RuBP
3-Phosphoglycerate
Photosystem II
Electron transport chain
Photosystem I
ATP
NADPH
G3P
Starch
(storage)
Amino acids
Fatty acids
Chloroplast
O_2
Sucrose (export)
Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.


Ford "E" Series V8 for 1948 - 51
337 Cubic Inch
33798-S
34846-S
88411-S
353253-S
21533-S
351377-S
24388-S
351377-S
6520
6521
356062-S
6524
6057
6010
6026
OR
358132-S
9430—RH
9431—LH
358099-S
9448
88413-S
34808-S
33816-S
9450
20408-S
34847-S
358066-S
21533-S
34847-S
8255
6666
6654
353089-S
352740-S
34810-S
356075-S
8592—R. H.
8593—L. H.
24428-S
34847-S
8575
6066
8571
357177-S
6049—RH
6050—LH
6051
34846-S
24366-S
6026
6059
6023
358099-S
73259-S
6020
34847-S
20408-S
8115


Ensuite 2
7.63 m²
Bedroom 2
29.89 m²
Lounge
50.74 m²
Kitchen
27.80 m²
Bathroom 1
7.83 m²
Hall
12.58 m²
Study
5.85 m²
WC 1
8.41 m²
Entrance Hall
18.65 m²
Garage
31.15 m²
Bedroom 1
26.78 m²
Ensuite 1
6.73 m²