

1 VISUAL ANALYTICS INTRO

S. Rinzivillo – rinzivillo@isti.cnr.it

WHO I AM?

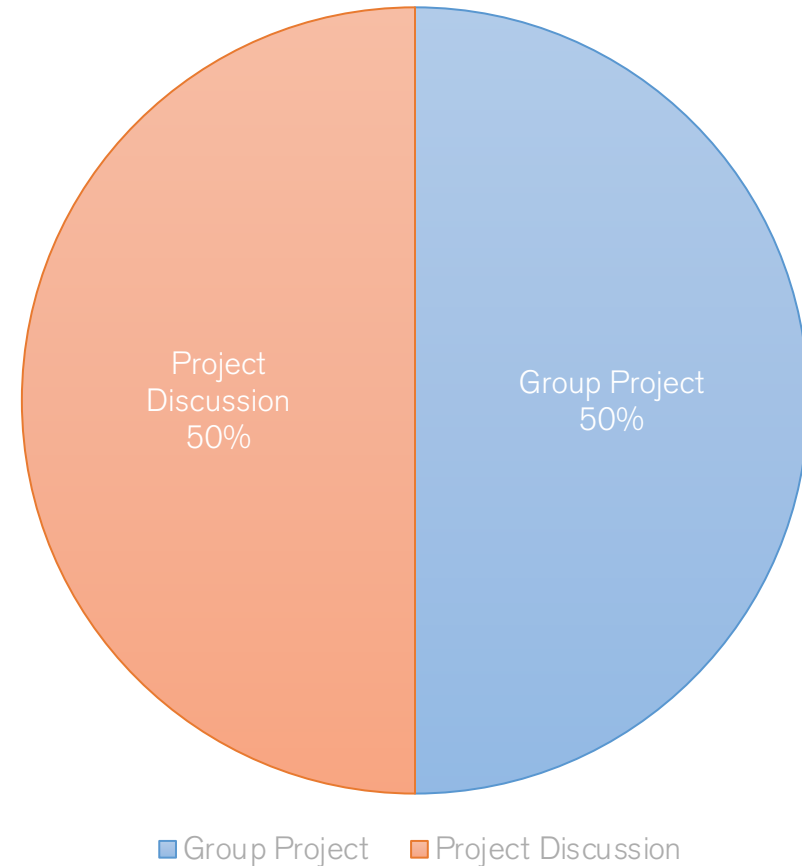
- Salvatore Rinzivillo
 - rinzivillo@isti.cnr.it
- Page course: <http://didawiki.cli.di.unipi.it/>
 - Visual Analytics
- Github page:
 - <https://github.com/va602aa-2021>
- Telegram channel:
 - <https://t.me/va602aa>

SCHEDULE

- On Monday
 - 16:15 to 18:00
 - Online: MS Teams
- On Friday
 - 16:15 to 18:00
 - Online: MS Teams

GRADING

- Project (50%)
 - Up to 2 persons per group (!)
- Project discussion (50%)
- Project topic
 - Multidimensional exploration of a dataset
 - One (or two) dataset(s) assigned for all
 - Specific proposal may be discussed



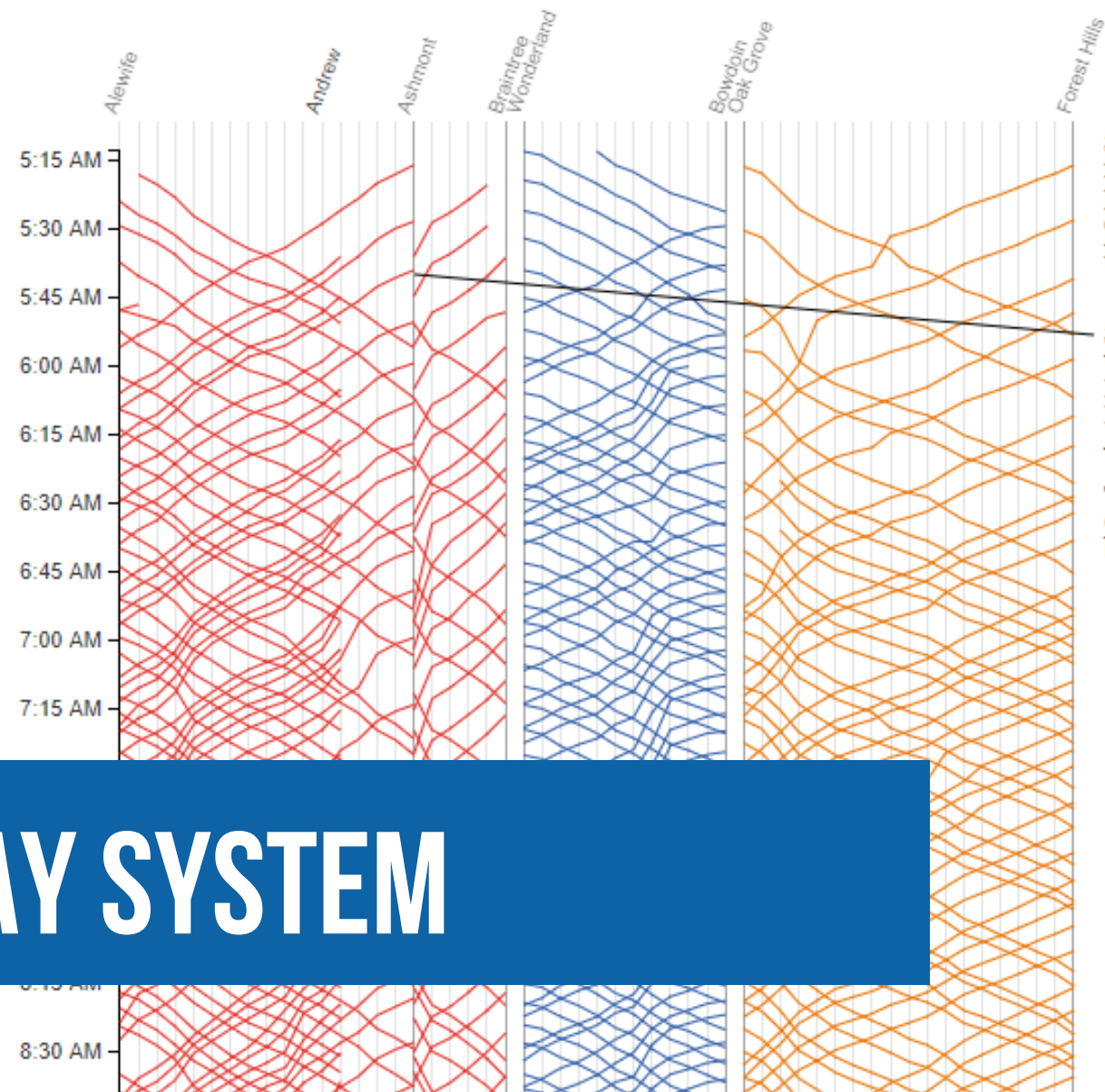
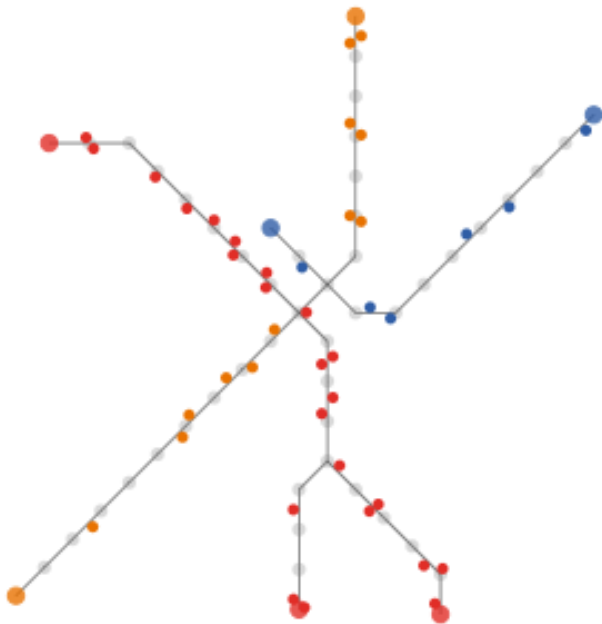
PROJECT FEATURES

- A project should have the following requirements:
 - The application should contain **multiple visual widgets**, each providing insights on a selection of dimensions of the original data
 - It is possible to use state-of-the-art charts (bar charts, line charts, etc.) and libraries (plotly, vega, etc). It is should implement a **novel, original visualization** to present the data in a creative, non-trivial way. (see examples on Vast Challenge 2008 developed in class)
 - **Interactivity** should be implemented, providing toolbars, selections and filters for the data.
 - The visual widget should interact among them, realising a set of **linked display** to browse the data across multiple dimensions

EXAMPLE: SCHOOL DISTRICTS

Where is the best place to
send your child to school in
New York City?





Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

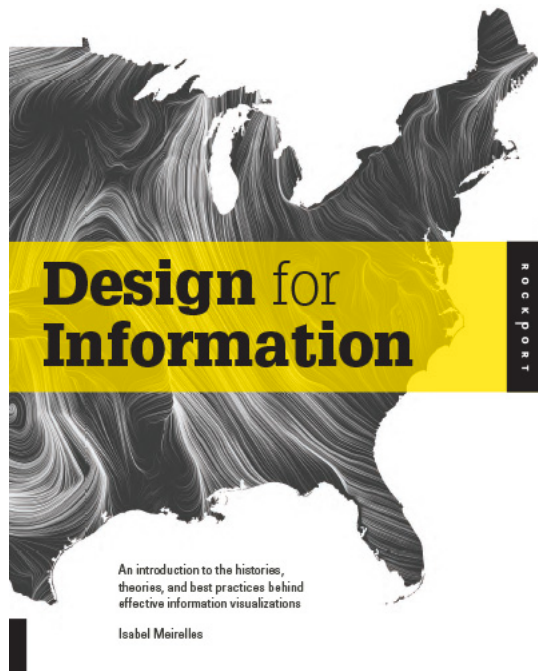
Train frequency increases around 6:30AM as morning rush hour begins.

BOSTON SUBWAY SYSTEM

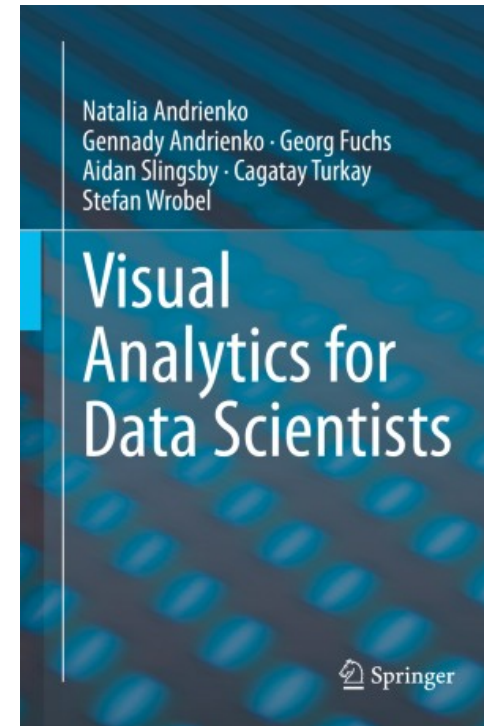
Locati
am. H
differ
Trains
they a
See th
and the

TEXTBOOKS

Design for Information
Isabel Meirelles



Visual Analytics for Data Scientists
Andrienko et al.



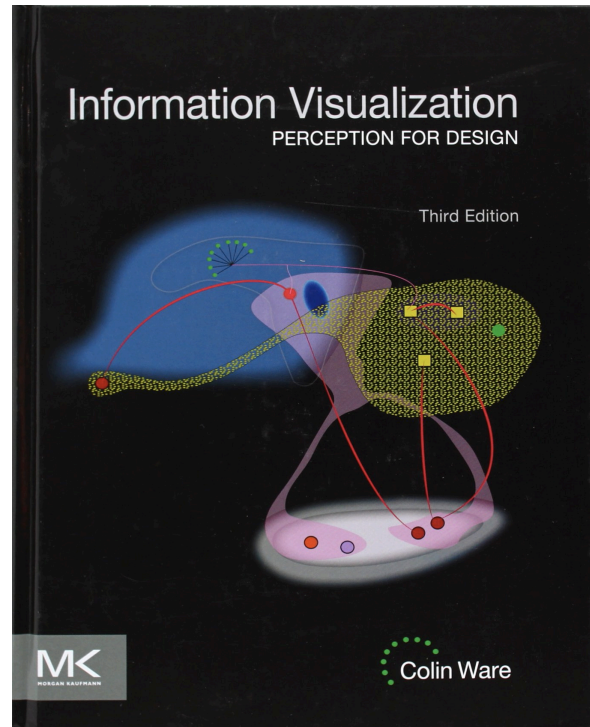
<http://alignedleft.com/tutorials>

Visual Analytics
2020/2021

INTERESTING READINGS

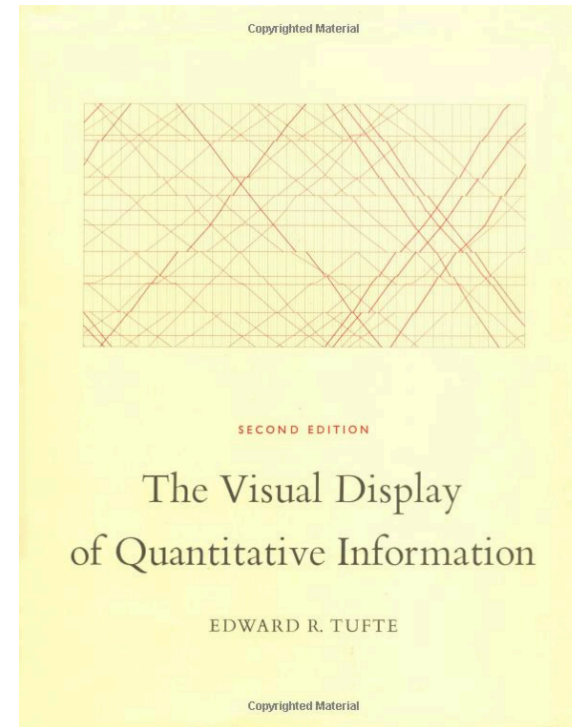
Information Visualization

Colin Ware



The Visual Display of Visual Information

Edward R. Tufte



DATA VISUALIZATION AND VISUAL ANALYTICS

INTRODUCTION

VA - CRASH COURSE

- Effective Visual Representation
 - Vision System
 - Visual Variables
- Toolbox – Bootstrap, Node.js, Vue.js, crossfilter.js
- Toolbox – Base visualizations (Plotly.js, Vega, DC.js)
- Toolbox – D3.js
 - Basics
 - Charts
 - Advanced Visualization
- Scientific Visualization
 - Plotting
 - Geography
- Storytelling



DATA VISUALIZATION

Convey Information through
graphical representation of data

MOTIVATIONS

- Data everywhere
- No value for raw data
 - Need to extract valuable information
- **Information overload:**
 - Irrelevant for current task
 - Processed in an inappropriate way
 - Presented in an inappropriate way

VISUALIZATION GOAL

- Record Information
 - Sketches, photographs, ECG,...
- Analyze data to support decisions (**exploration**)
 - Create and verify hypotheses
 - Identify Patterns
 - Identify Outliers
- Communicate (**explanation**)
 - Share or highlight insights on data
 - Persuade

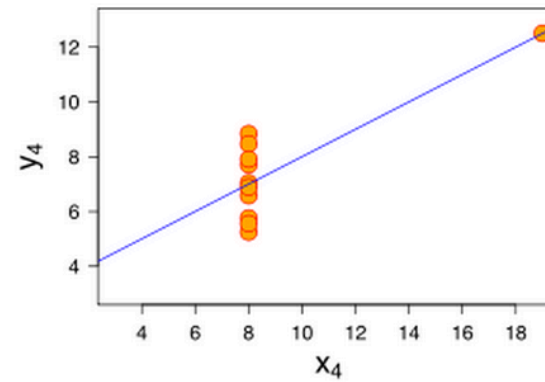
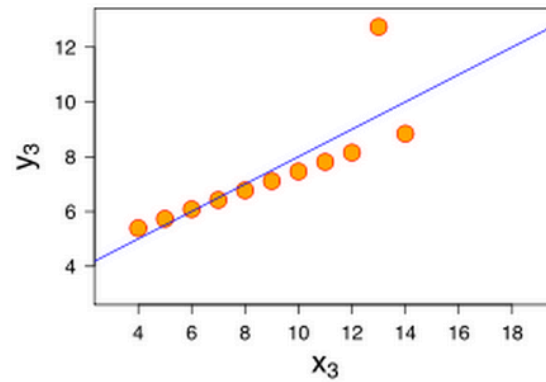
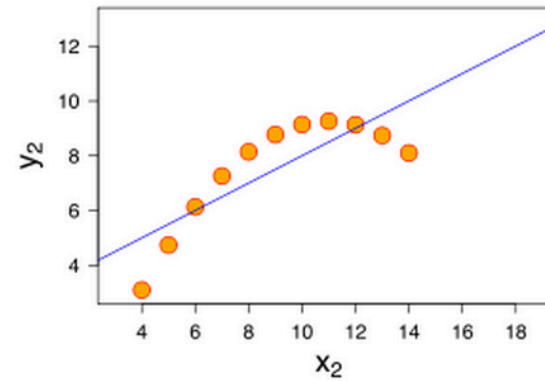
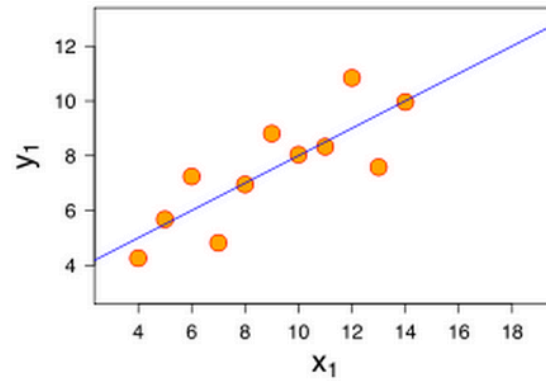
ANALYZE: ANSCOMBE'S QUARTET - DATASETS

<i>Data Set A</i>		<i>Data Set B</i>		<i>Data Set C</i>		<i>Data Set D</i>	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

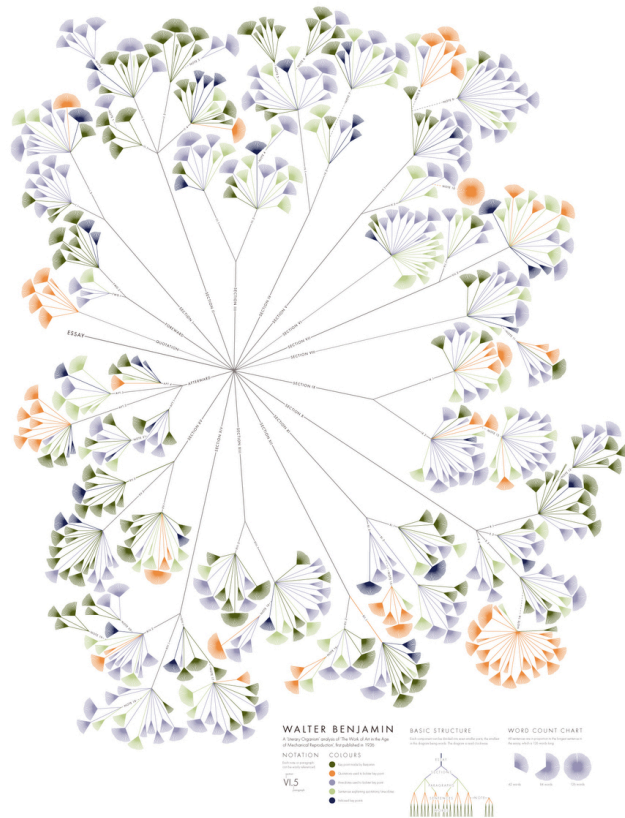
ANALYZE: ANSCOMBE'S QUARTET - PROPERTIES

Property	Value
Mean of x in each case	9 (exact)
Sample variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Sample variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

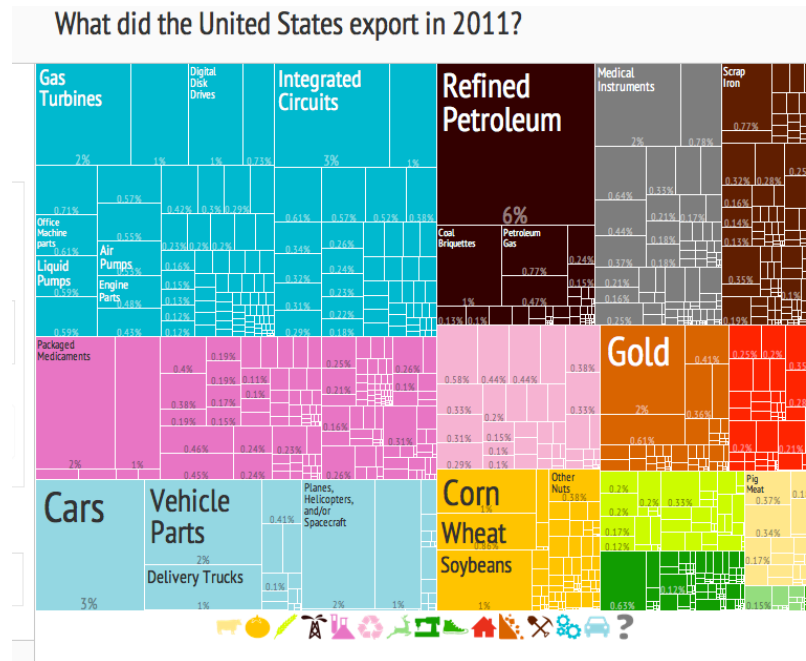
ANALYZE: ANSCOMBE'S QUARTET – GRAPHICS



COMMUNICATE: HIERARCHICAL STRUCTURES

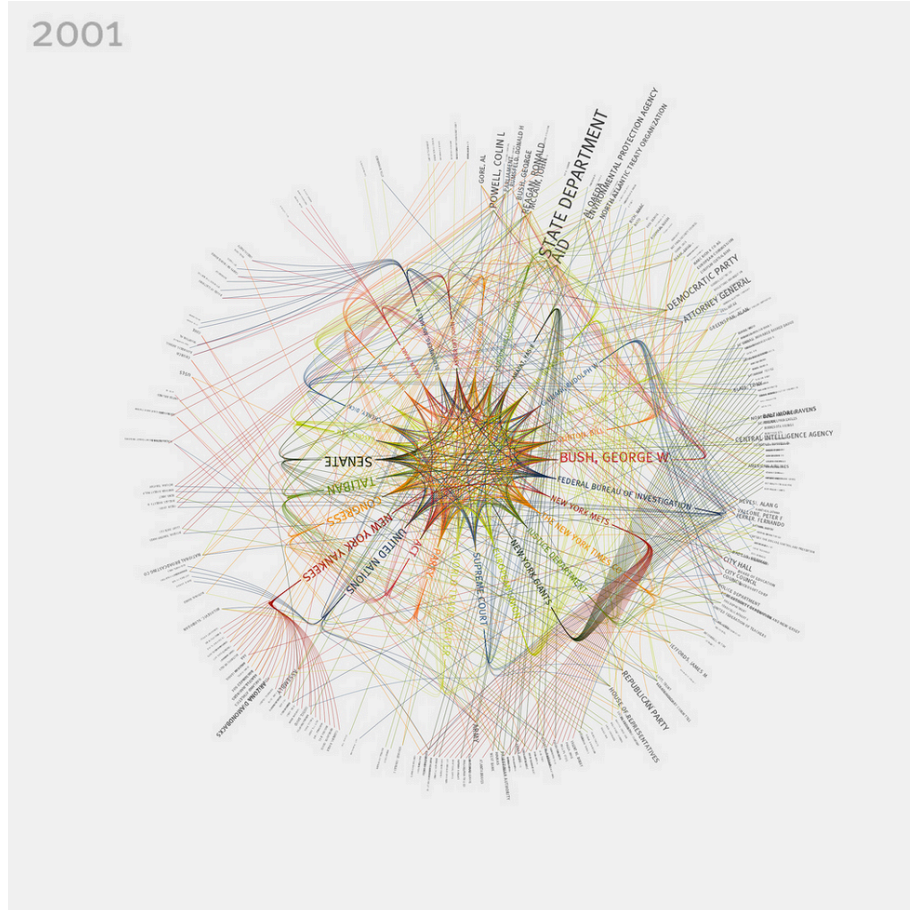


<http://www.stefanieposavec.co.uk/entangled-word-bank/>

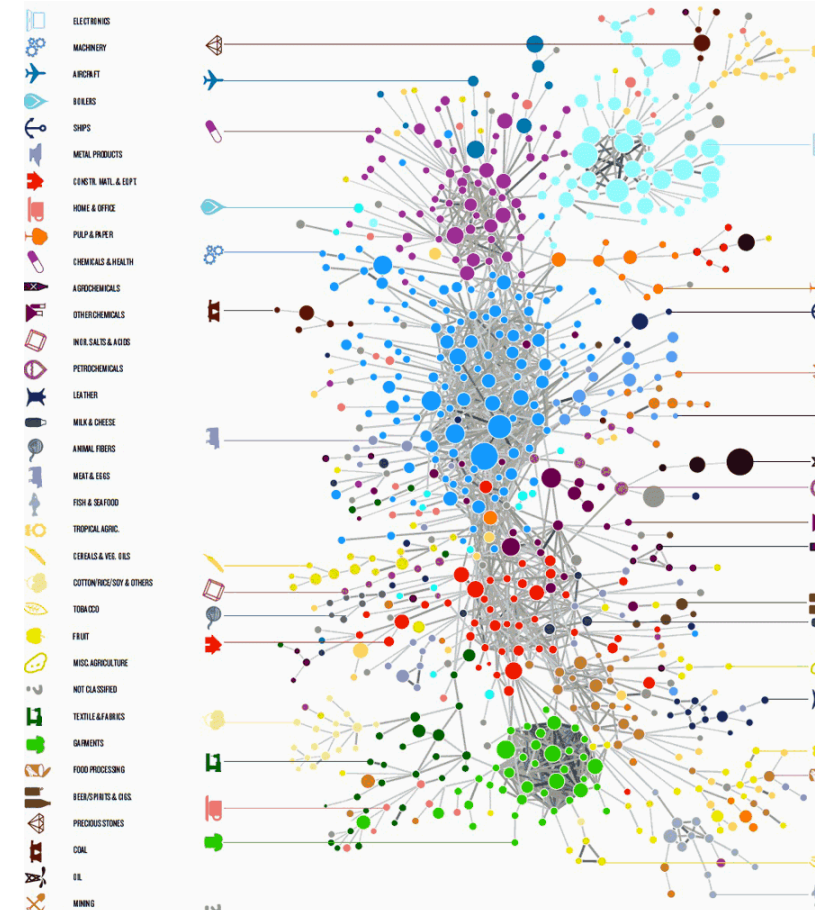


<https://atlas.cid.harvard.edu/>

COMMUNICATE: NETWORKS

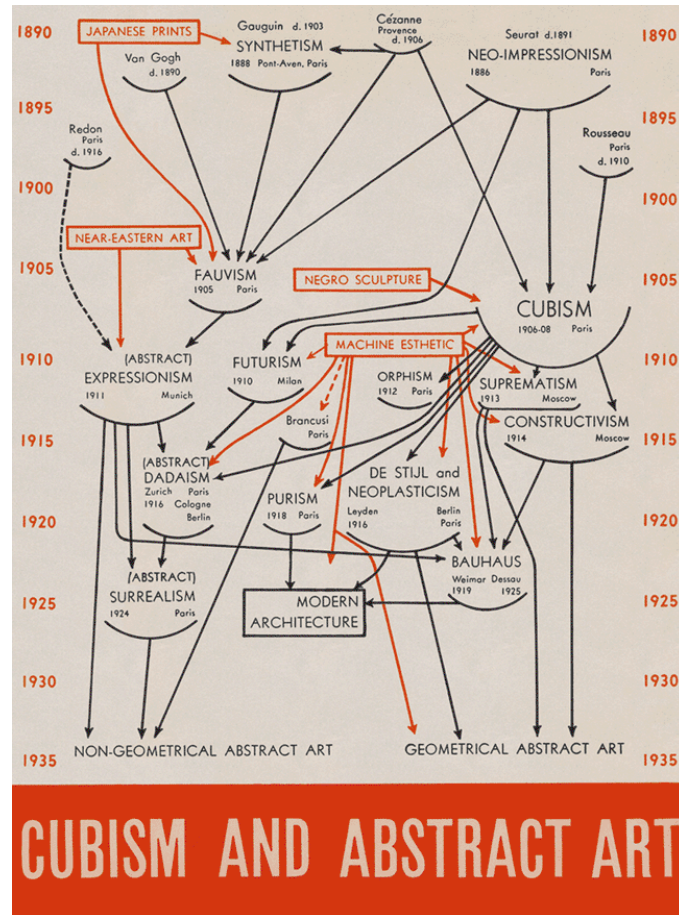


<https://www.flickr.com/photos/blprnt/sets/72157614008027965/>



<https://atlas.cid.harvard.edu/>

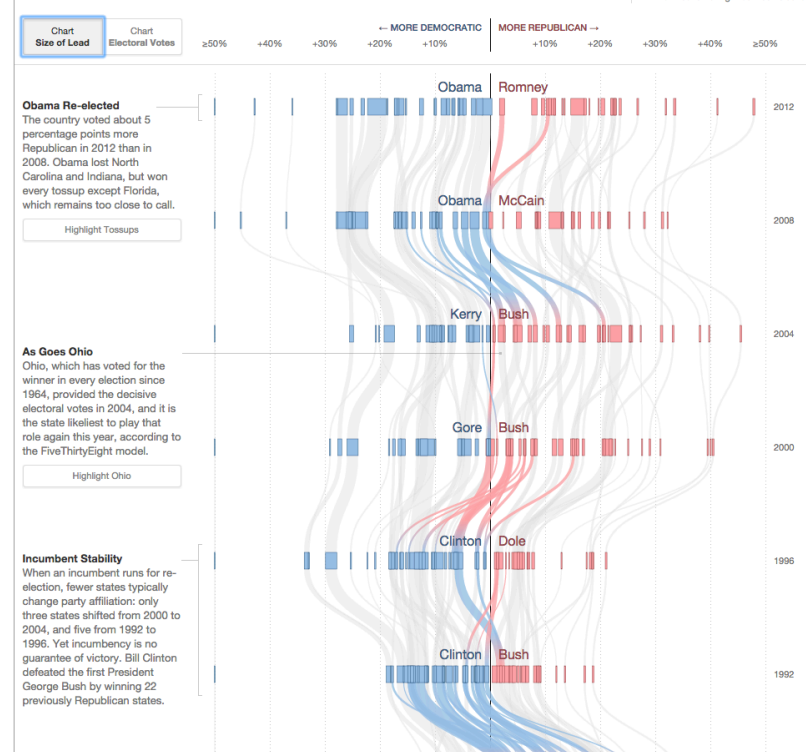
COMMUNICATE: TEMPORAL STRUCTURES



Cubism And Abstract Art (Alfred H. Barr 1936)

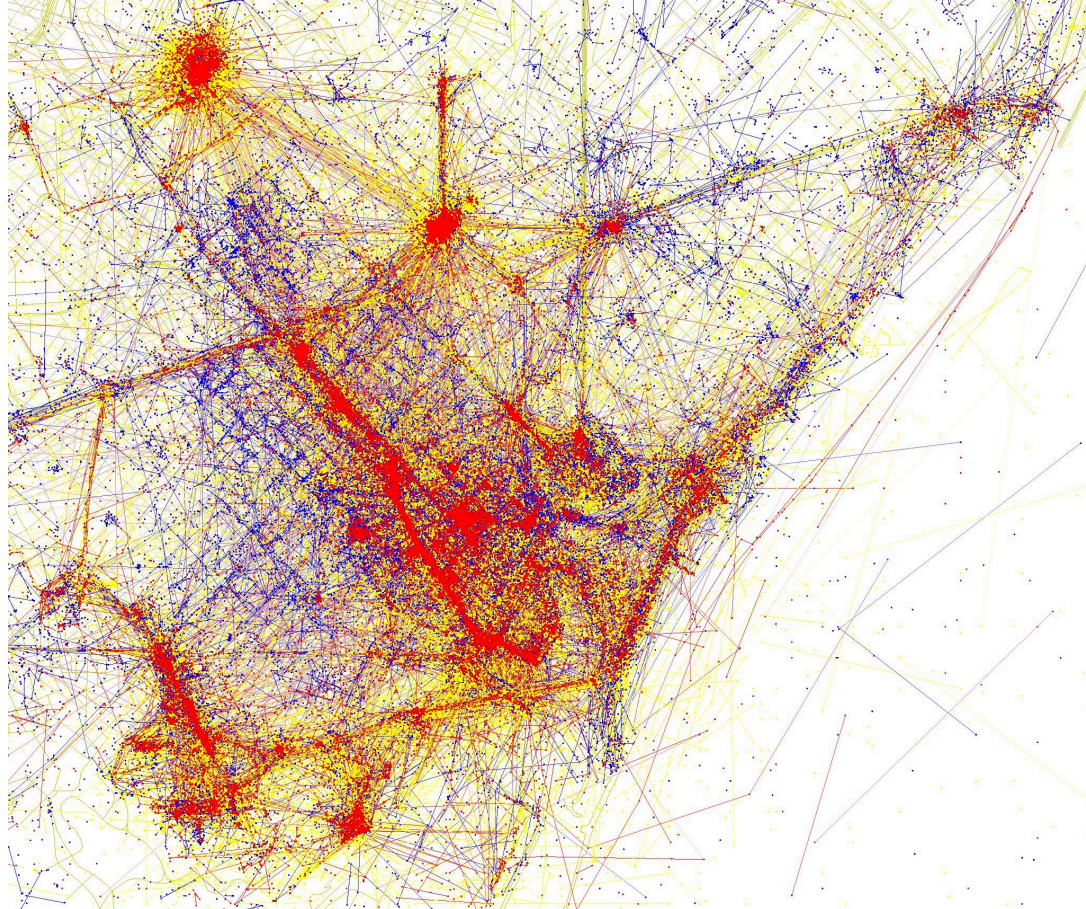
Over the Decades, How States Have Shifted

Recent elections have placed a heavy emphasis on "swing states" — Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states stacked up in the 2012 election and how they have shifted over past elections.

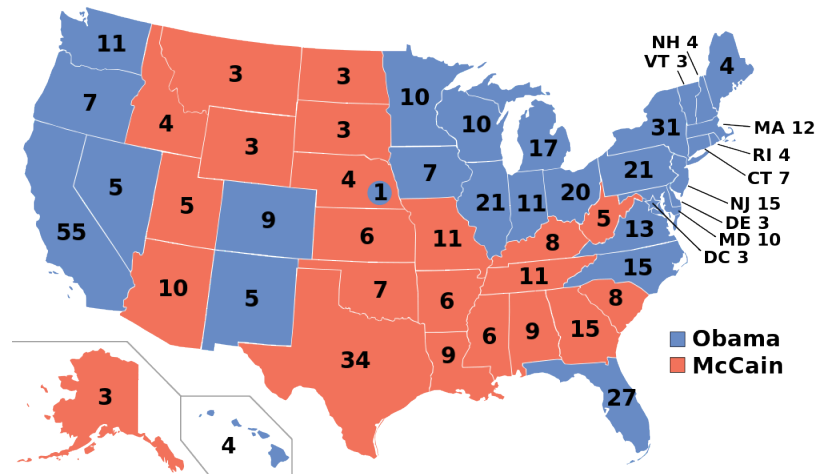


<http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html>

COMMUNICATE: MAPS

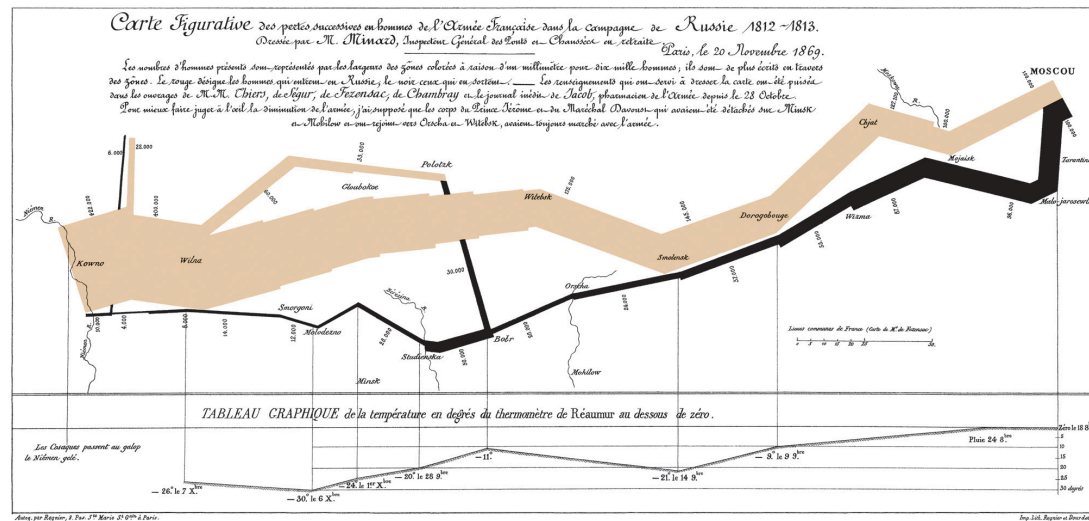


<https://www.flickr.com/photos/walkingsf/sets/72157624209158632/>



"ElectoralCollege2008" by Gage - Own work. Licensed under Public Domain via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:ElectoralCollege2008.svg#mediaviewer/File:ElectoralCollege2008.svg>

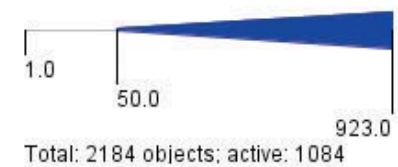
COMMUNICATE: SPATIO-TEMPORAL DATA



"Minard" by Charles Minard (1781-1870) - see upload log. Licensed under Public Domain via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Minard.png#mediaviewer/File:Minard.png>



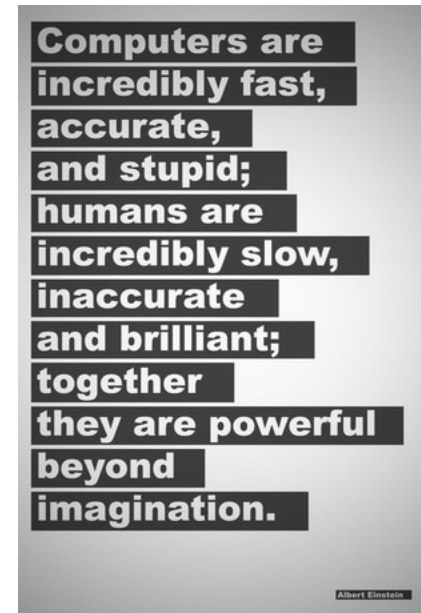
N of moves



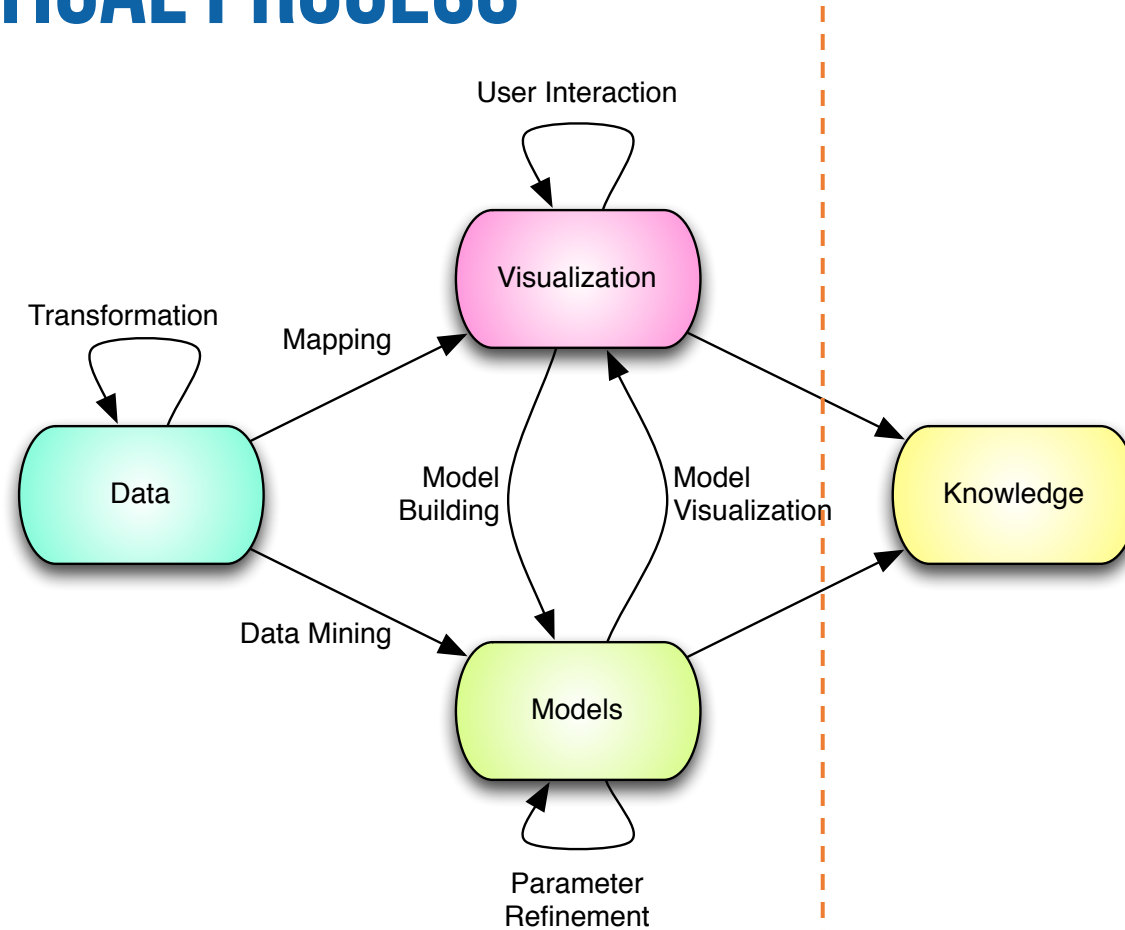
Visual Analytics of Movement.
 G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel
 Springer, 2013

VISUALIZATION AND VISUAL ANALYTICS

- Make data and information processing transparent
- Combine strengths of humans and computers



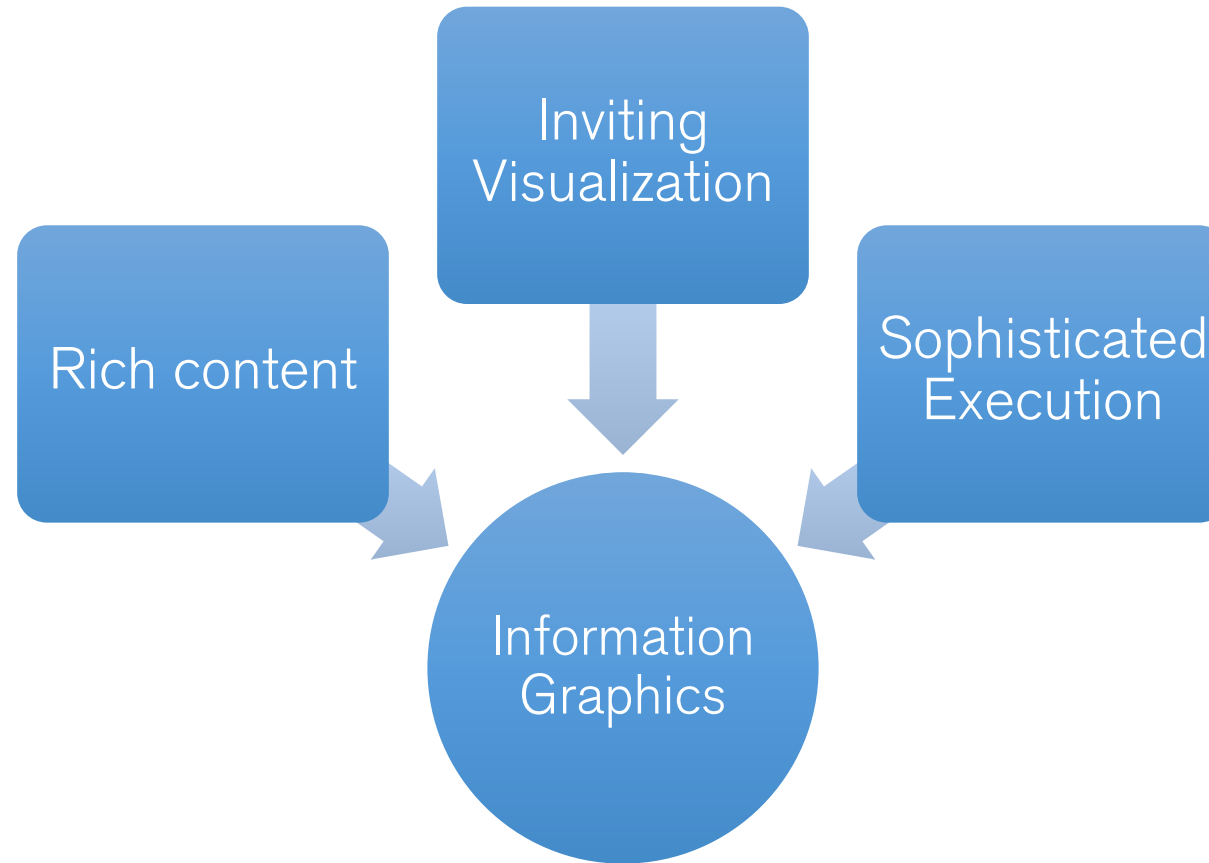
VISUAL ANALYTICAL PROCESS



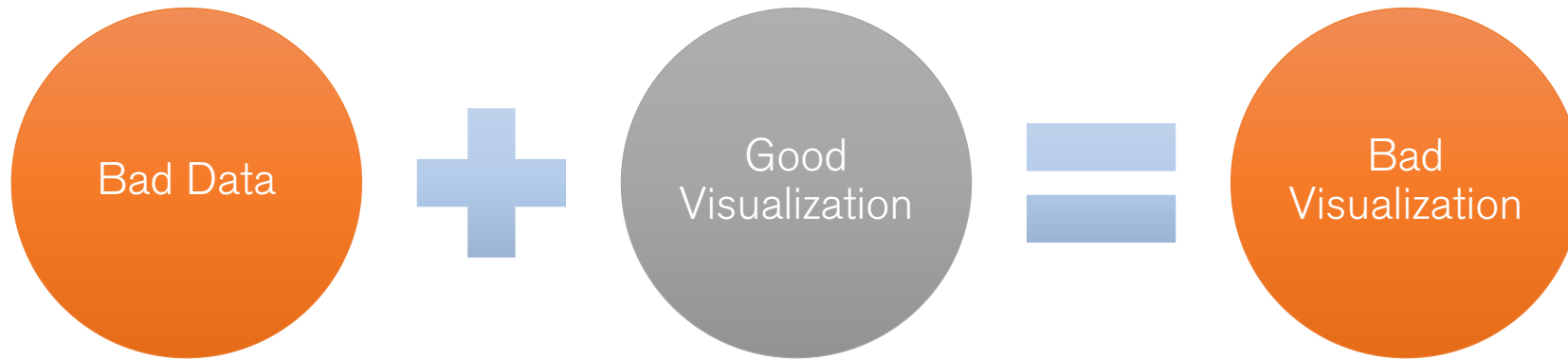
Adapted from:
Mastering the Information Age
Keim, Kohlhammer, Ellis, Mansmann

Exploration | Explanation

ELEMENTS OF GOOD VISUALIZATION



IMPORTANCE OF VALID DATA



OTHER RESOURCES

Observe how others resolved design problems

datavisualization.ch

https://flowingdata.com/

Precinct-level map of 2020 election results

February 2, 2021

Topic
Maps / election, precinct, upshot

NYT's The Upshot published their precinct-level map of 2020 election results. Zoom in to your geographic area and bask in or scratch your head over the detailed variation.

This seems to be a recurring view now, with their "extremely detailed map" making an appearance after the 2016 and 2018 election. They also had their "most detailed maps" in 2014.

However, this year, The Upshot made their precinct-level data available on GitHub, so you can look closer if you like.

infosthetics.com

Age generations in the U.S. Senate, over time

February 1, 2021

Topic
Statistical Visualization / age, generations, government, Senate

