Data Visualization Options

What is data discovery and visualization?

- **Data discovery** is the process of breaking complex **data** collections into information that users can understand and manage. **It** turns incomprehensible mounds of raw **data** into groups, sets, and relationships, making order out of chaos.
- **Data discovery** answers the question, "What does it all mean?"
- Data visualization is its representation.

What are data visualization tools?

- By using visual elements like charts, graphs, and maps, **data visualization tools** provide an accessible way to see and understand trends, outliers, and patterns in **data**.

Is Excel a data visualization tool?

- **Excel** is a spreadsheet **tool**, while Tableau is a **data visualization** one.
- Spreadsheet **tools** are electronic worksheets that display **data** in a tabular format (a table of columns and rows).
- Each **data** point is stored in “cells” and can be manipulated by manually set formulas.

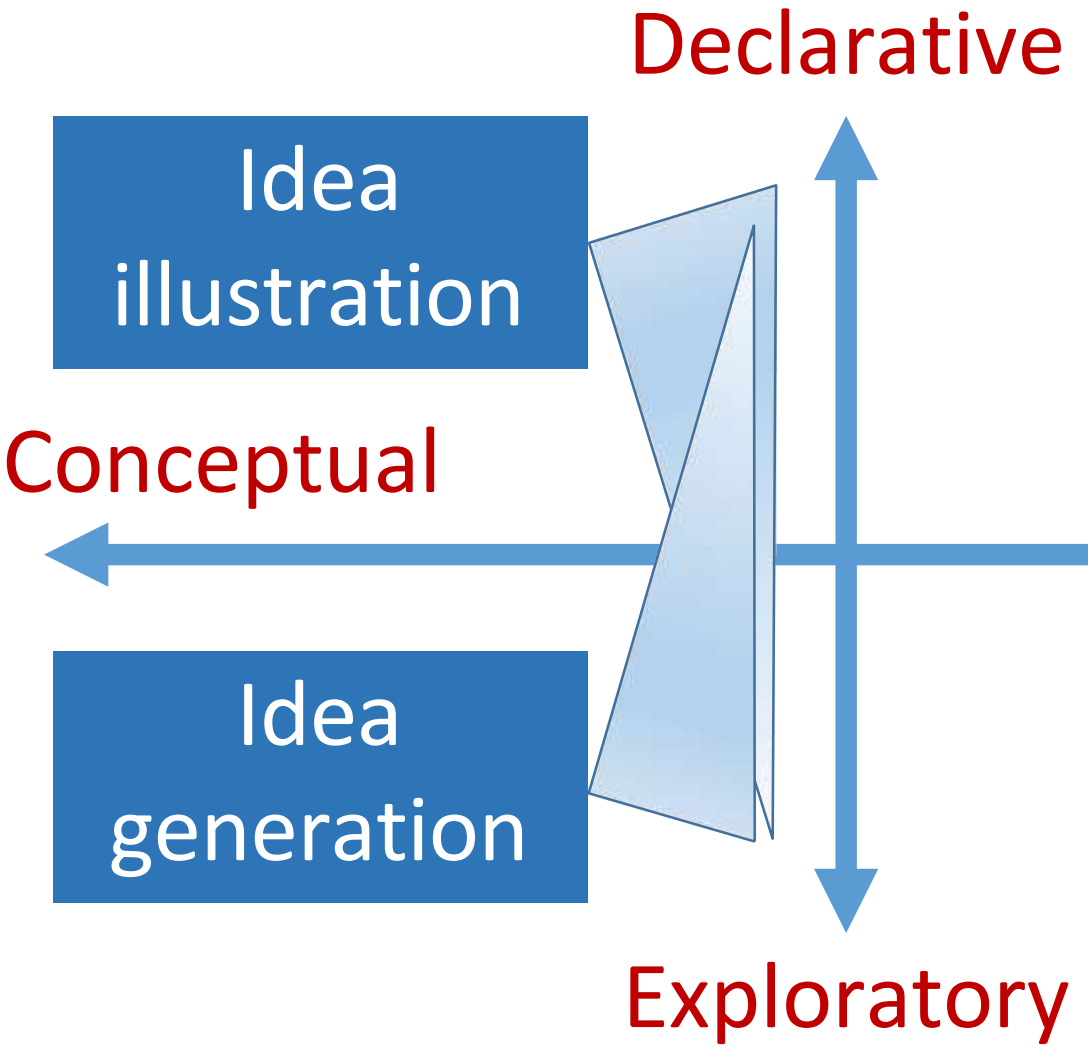
How do you create good data visualization?

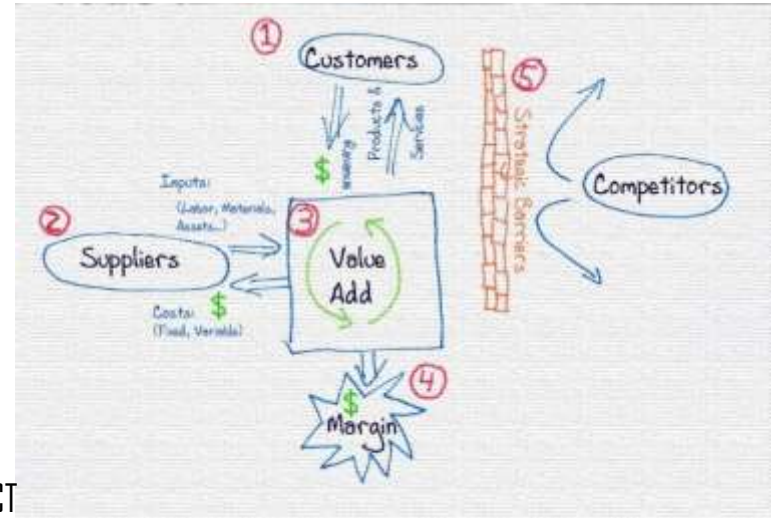
- Use it wisely in your data visualization design.
- Use a single color to represent the same type of data.
- Watch out for positive and negative numbers.
- Make sure there is sufficient contrast between colors.
- Avoid patterns.
- Select colors appropriately.
- Don't use more than 6 colors in a single layout.

What kind of visual communication do you want to create?

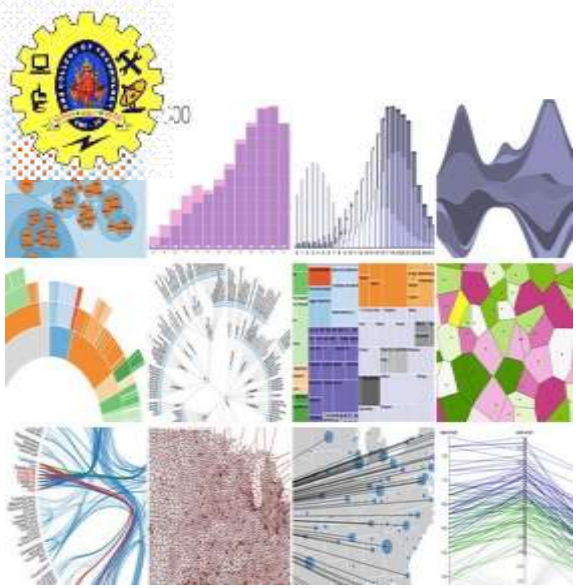
1. Is my information conceptual or data-driven?
   - Conceptual information is qualitative
   - Data-driven information is quantitative
2. Are my visuals meant to be declarative or exploratory?
   - A declarative purpose is to make a statement
   - An exploratory purpose is to look for new ideas

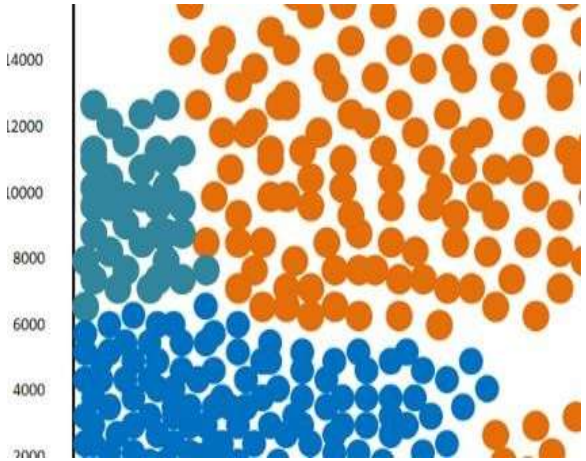
our Types of Data Visualizations





Four Types of Data Visualizations





Declarative

Idea illustration

Everyday dataviz

Conceptual

Data-Driven

Idea generation

Visual discovery

Exploratory

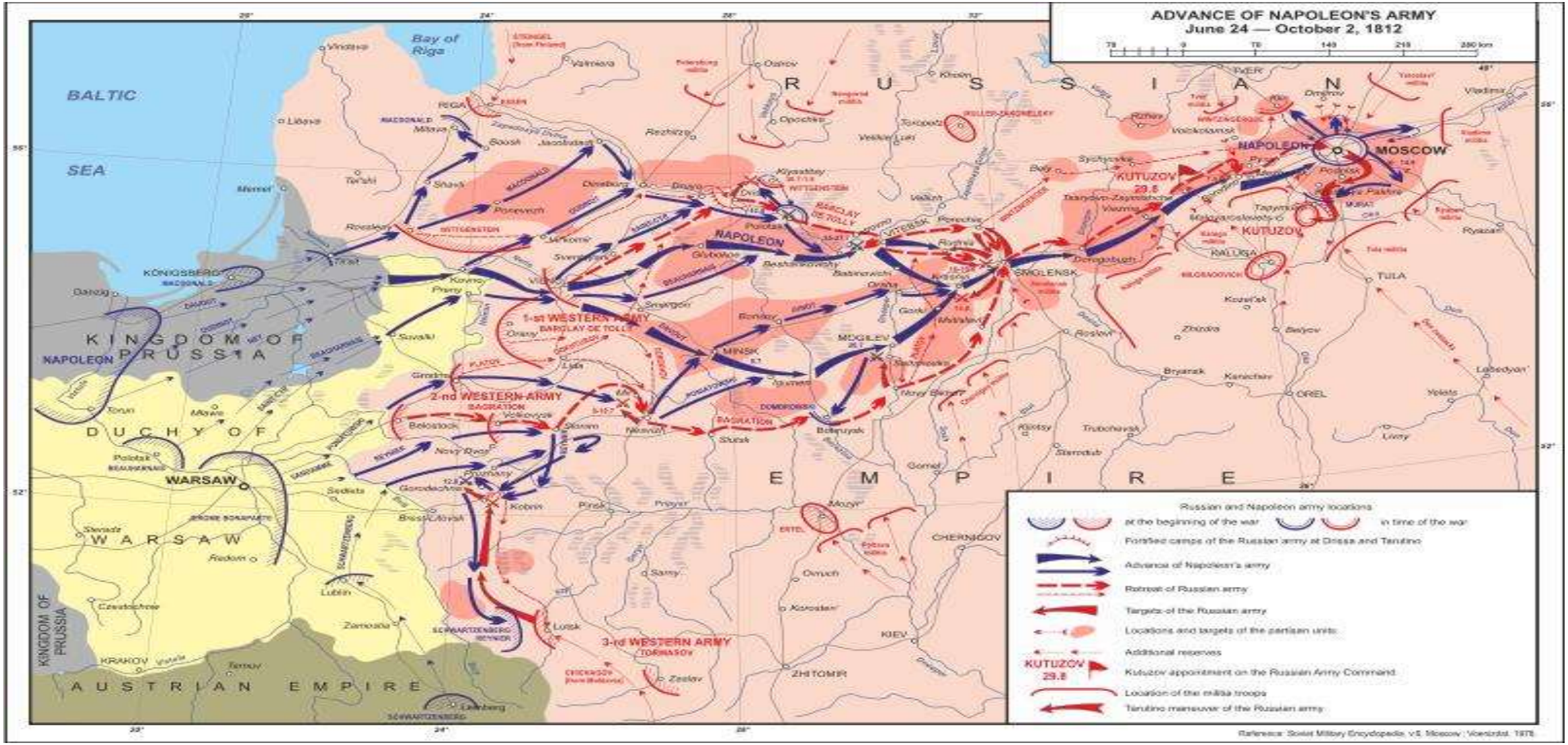
Data Visualization

provide clear understanding of patterns in data

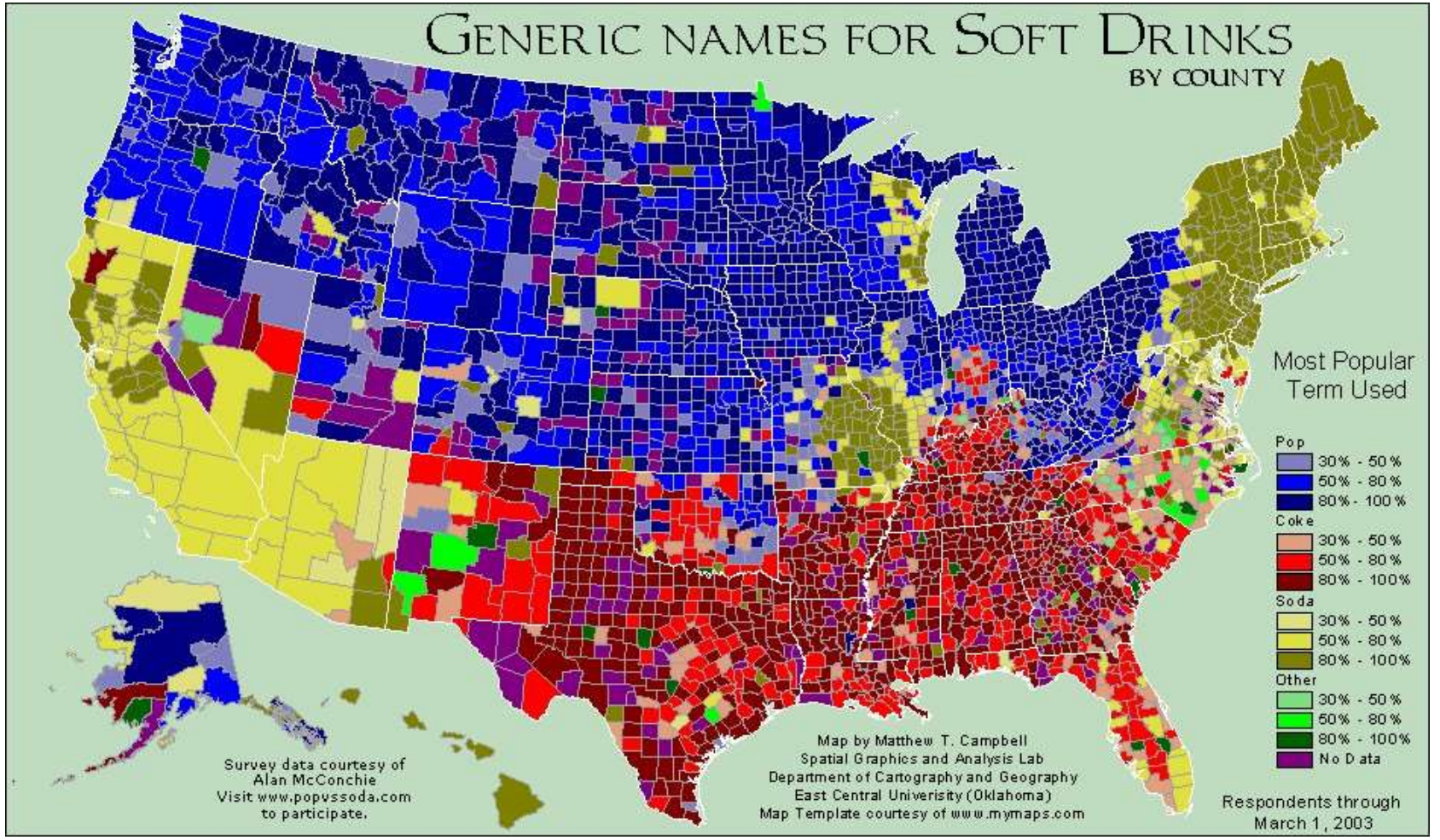
detect hidden structures in data

condense information

What makes a good chart?



What can you learn from this map?



Some basic principles (adapted from Tufte 2009)

1. The chart should tell a story
2. The chart should have graphical integrity
3. The chart should minimize graphical complexity

Tufte's fundamental principle:
Above all else show the data

Principle 1: The chart should tell a story

Graphics should be clear on their own

The depictions should enable meaningful comparison

The chart should yield insight beyond the text

“If the statistics are boring, then you’ve got the wrong numbers.” (Tufte 2009)

Principle 2: The chart should have graphical integrity

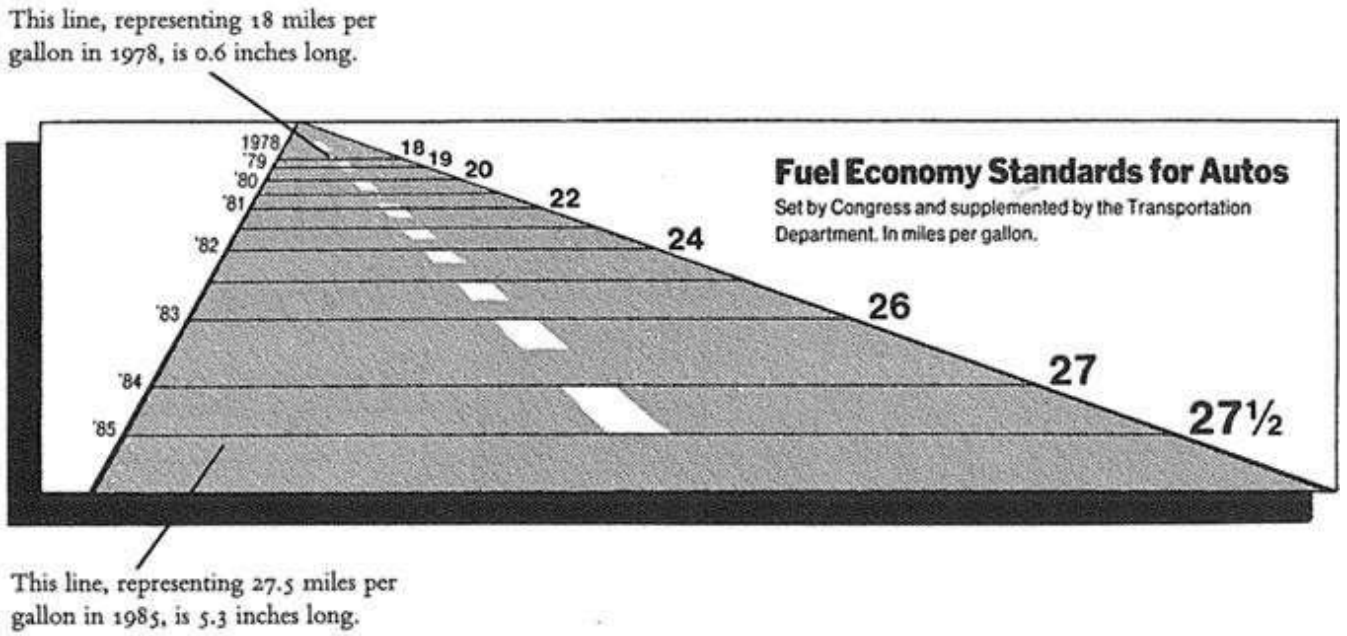
- Basically, it shouldn't "lie" (mislead the reader)
- Tufte's "Lie Factor":
  - $Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data}$

Should be ~ 1

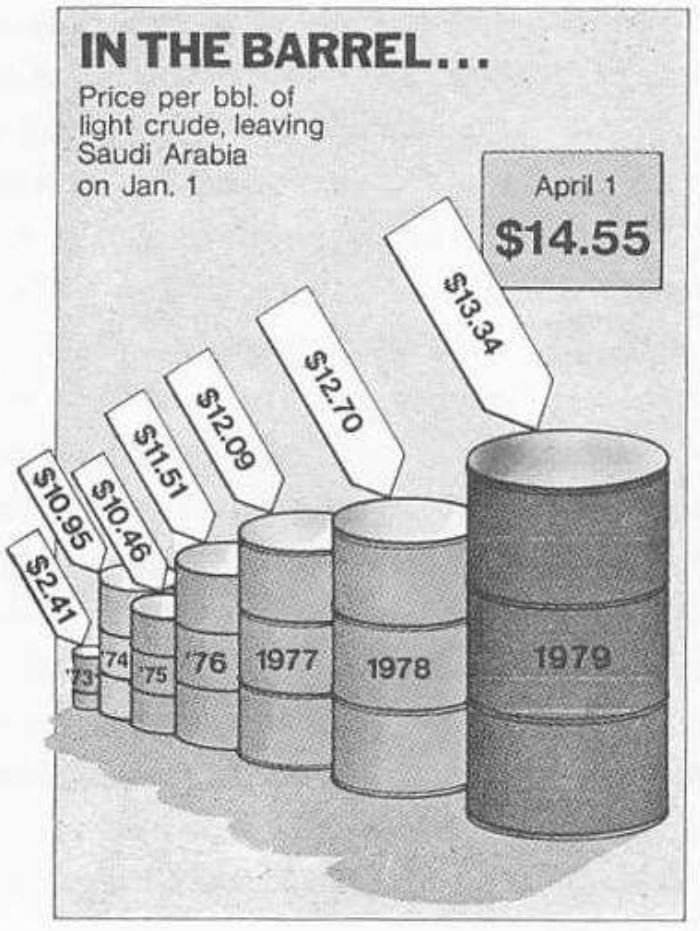
> 1 = exaggerated effect

< 1 = understated effect

Examples of the "lie factor"



$$LF = \frac{5.3/0.6}{27.5/18} = \frac{8.83}{1.53} = 5.77$$



Reprinted from Tufte (2009), p. 57 & p. 62

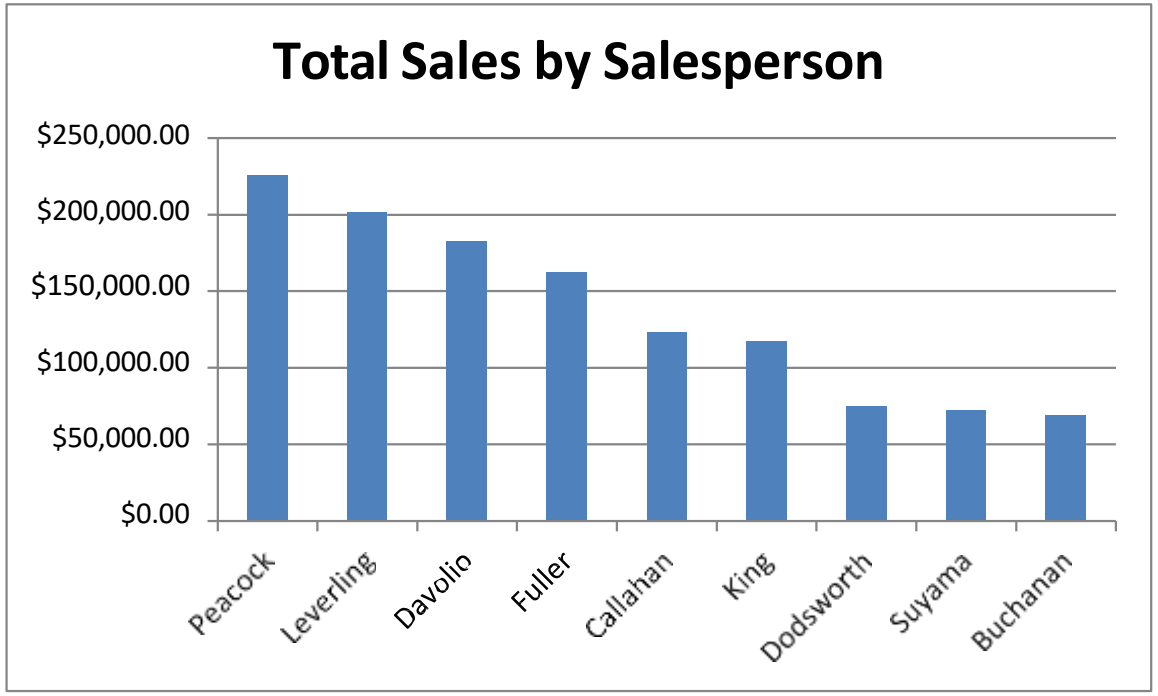
$$LF = \frac{4280\%\ (change\ in\ volume)}{454\%\ (change\ in\ price)} = 9.4$$

Principle 3: The chart should minimize graphical complexity

Generally, the simpler the better…

When a table is better than a chart

For a few data points, a table can do just as well…



Salesperson	Total Sales
Peacock	$225,763.68
Leverling	$201,196.27
Davolio	$182,500.09
Fuller	$162,503.78
Callahan	$123,032.67
King	$116,962.99
Dodsworth	$75,048.04
Suyama	$72,527.63
Buchanan	$68,792.25

The table carries more information in less space and is more precise.

The Ultimate Table: The Box Score

- Large amount of information in a very small space
- So why does this work?
  - Depends on the reader's knowledge of the data

Philadelphia Phillies

Hitters	AB	R	H	RBI	BB	SO	#P	AVG	OBP	SLG
S Victorino CF	3	0	0	0	1	0	16	.000	.250	.000
P Polanco 3B	3	1	0	0	1	0	18	.000	.250	.000
J Rollins SS	4	2	2	0	0	0	14	.500	.500	.500
R Howard 1B	3	1	2	1	0	0	15	.667	.500	.667
R Ibanez LF	4	0	0	1	0	0	14	.000	.000	.000
B Francisco RF	3	1	1	1	1	0	17	.333	.500	.333
C Ruiz C	4	0	1	0	0	0	16	.250	.250	.250
W Valdez 2B	4	0	2	1	0	0	7	.500	.500	.750
R Halladay P	1	0	0	0	0	0	2	.000	.000	.000
a-P Orr PH	1	0	0	0	0	0	3	.000	.000	.000
J Romero P	0	0	0	0	0	0	0	.000	.000	.000
D Herndon P	0	0	0	0	0	0	0	.000	.000	.000
R Madson P	0	0	0	0	0	0	0	.000	.000	.000
b-R Gload PH	1	0	1	0	0	0	3	1.000	1.000	1.000
D Baez P	0	0	0	0	0	0	0	.000	.000	.000
c-J Mayberry Jr. PH	1	0	1	1	0	0	5	1.000	1.000	1.000
Totals	**32**	**5**	**10**	**5**	**3**	**0**	**130**			

a-lined out to first for R Halladay in the 6th
b-singled to left center for R Madson in the 8th
c-singled to deep center for D Baez in the 9th

Data Ink

- The amount of “ink” devoted to data in a chart
- Tufte’s Data-Ink ratio:
  - $Data - ink\ ratio = \frac{data-ink}{total\ ink\ used\ in\ graphic}$

Should be ~ 1

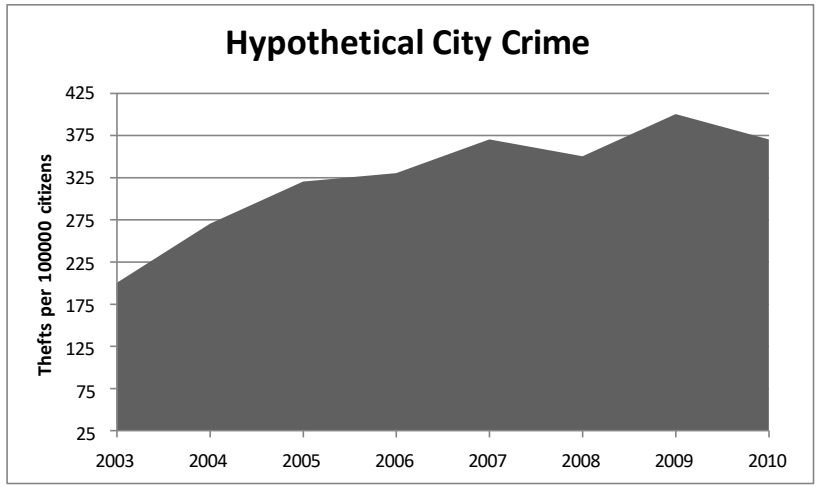
< 1 = more non-data related ink in graphic

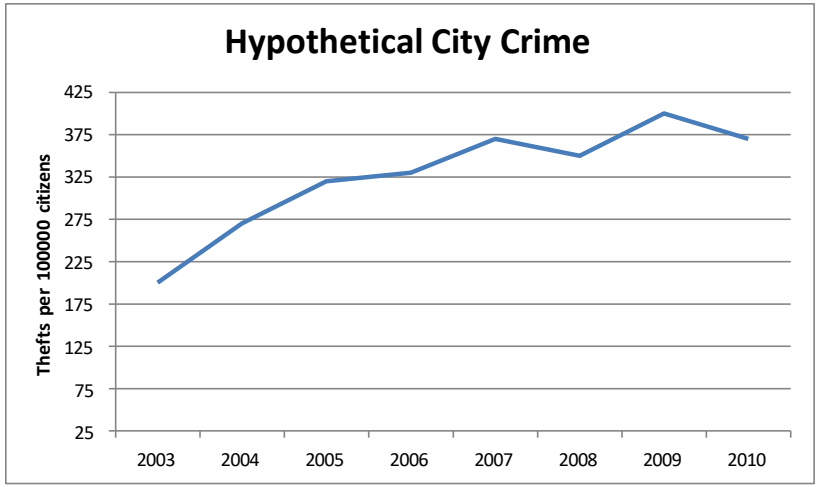
= 1 implies all ink devoted to data

Tufte’s principle:
Erase ink whenever possible

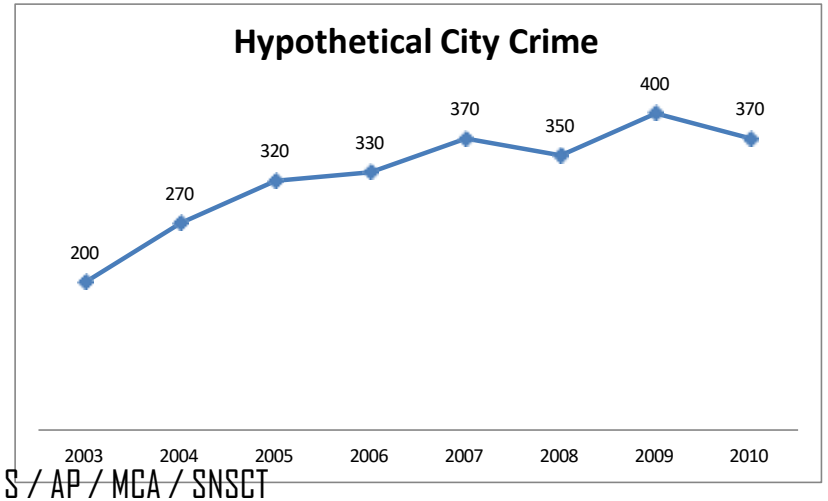
Being conscious of data ink

Lower data-ink ratio
(worse)

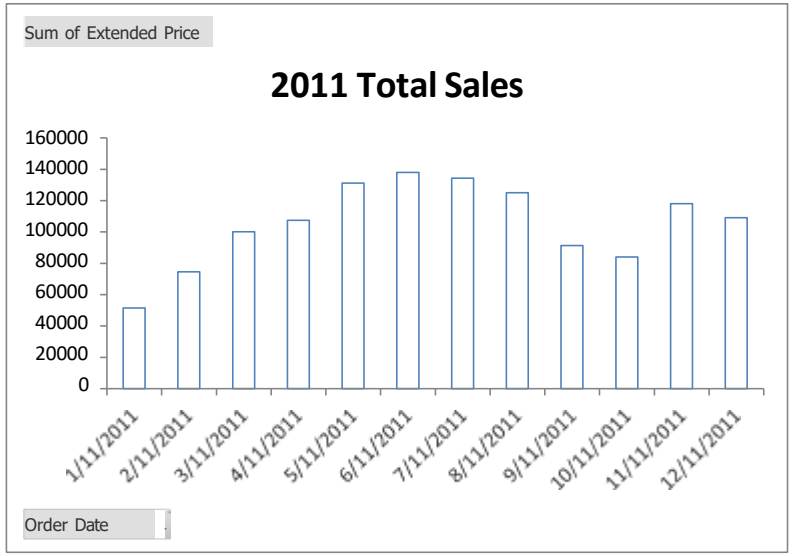


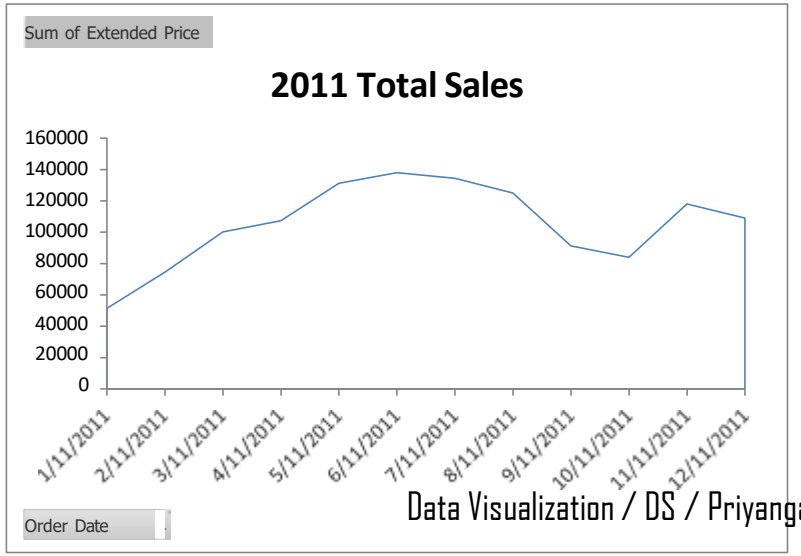


Higher data-ink ratio
(better)



What makes a good chart?





Sometimes it's really a matter of preference.

These both minimize data ink.

Why isn't a table better here?

3-D Charts

Evaluate this from a data-ink perspective. How does it affect the clarity of the chart?

One of the golden rules of data visualization is…..

Never use 3D!

Data Integrity/ Lie Factor	• 3D skews numbers, making them difficult to interpret or compare
Graphical Complexity	• Adding 3D to graphs introduces unnecessary chart elements like side and floor panels

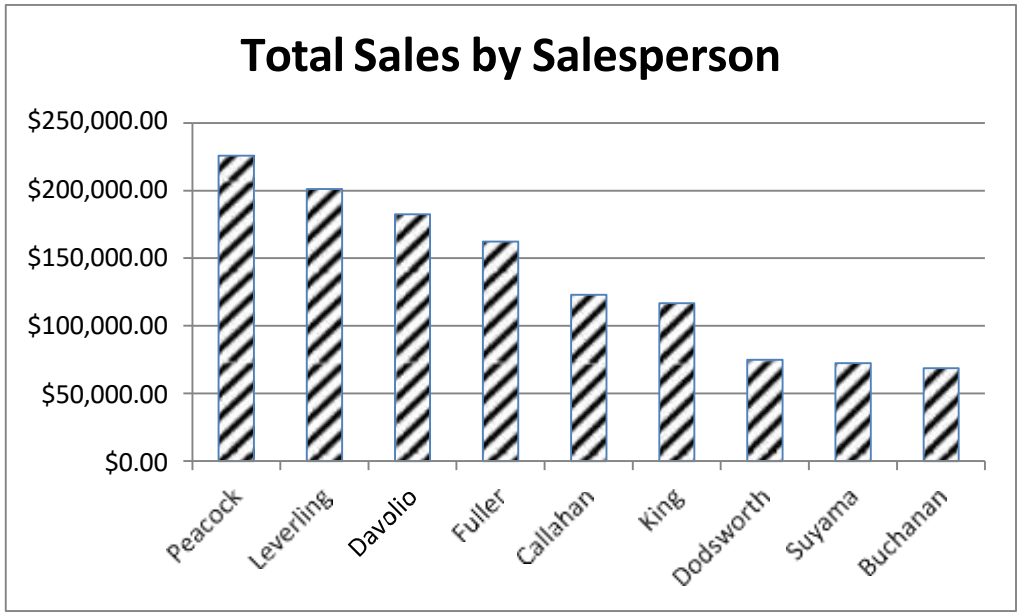
Chartjunk: Data Ink “gone wild”

Unnecessary visual clutter that doesn’t provide additional insight

Distraction from the story the chart is supposed to convey

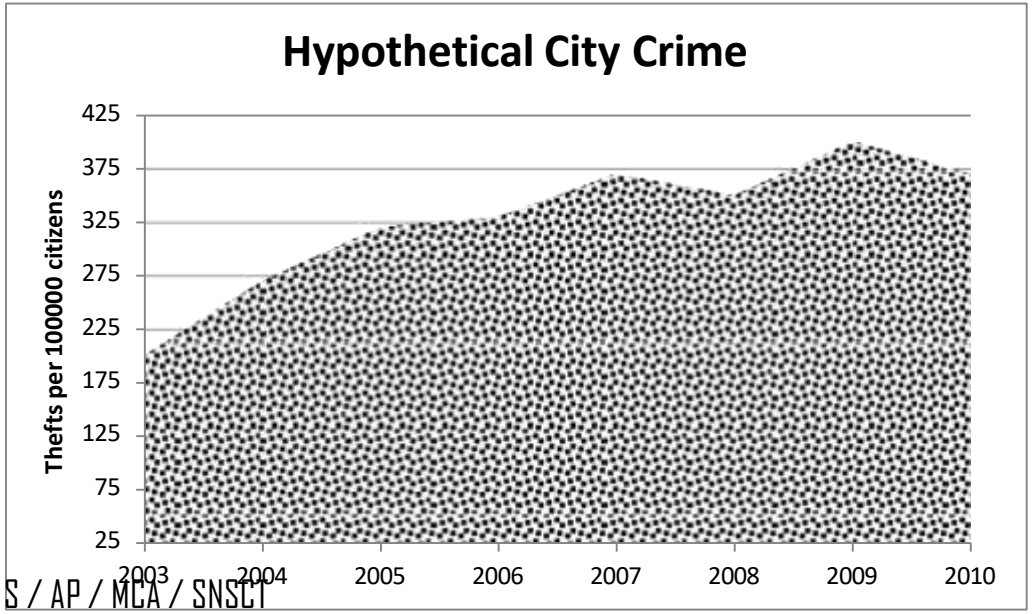
When the data-ink ratio is low, chartjunk is likely to be high

Example: Moiré effects (Tufte 2009)

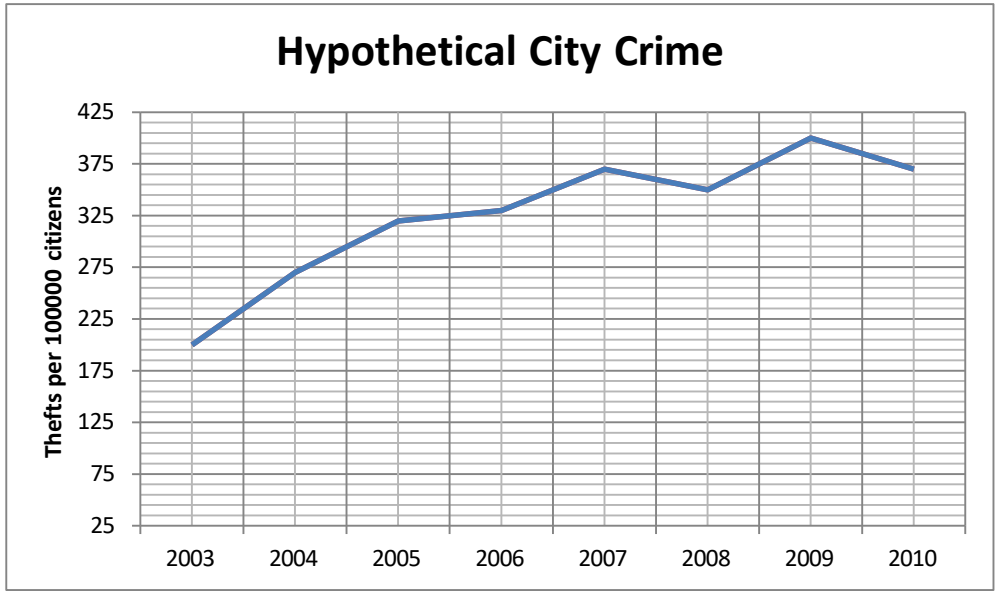


Creates illusion of movement

Stands out, in a bad way

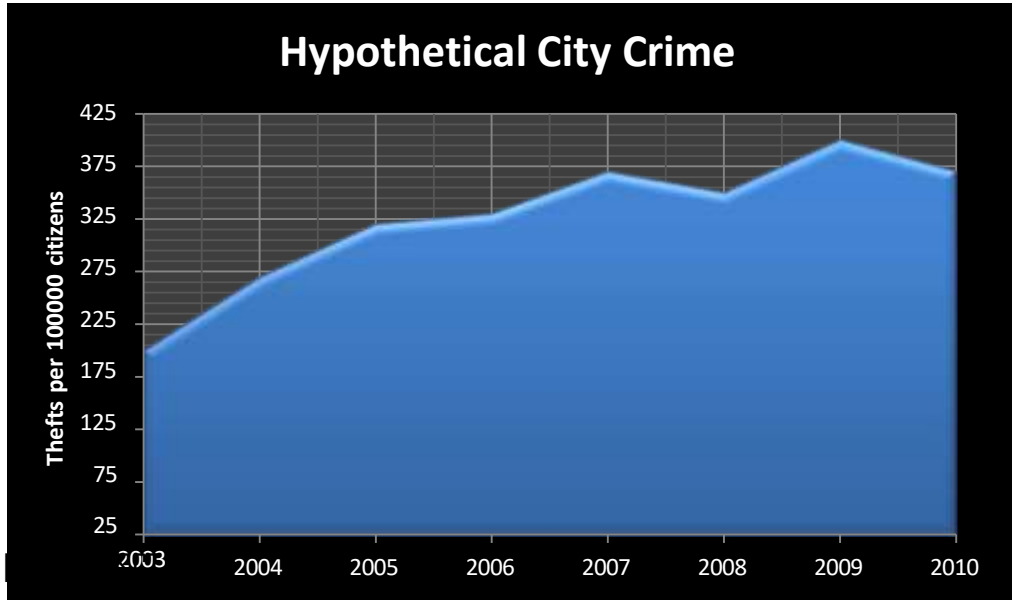


Example: The Grid

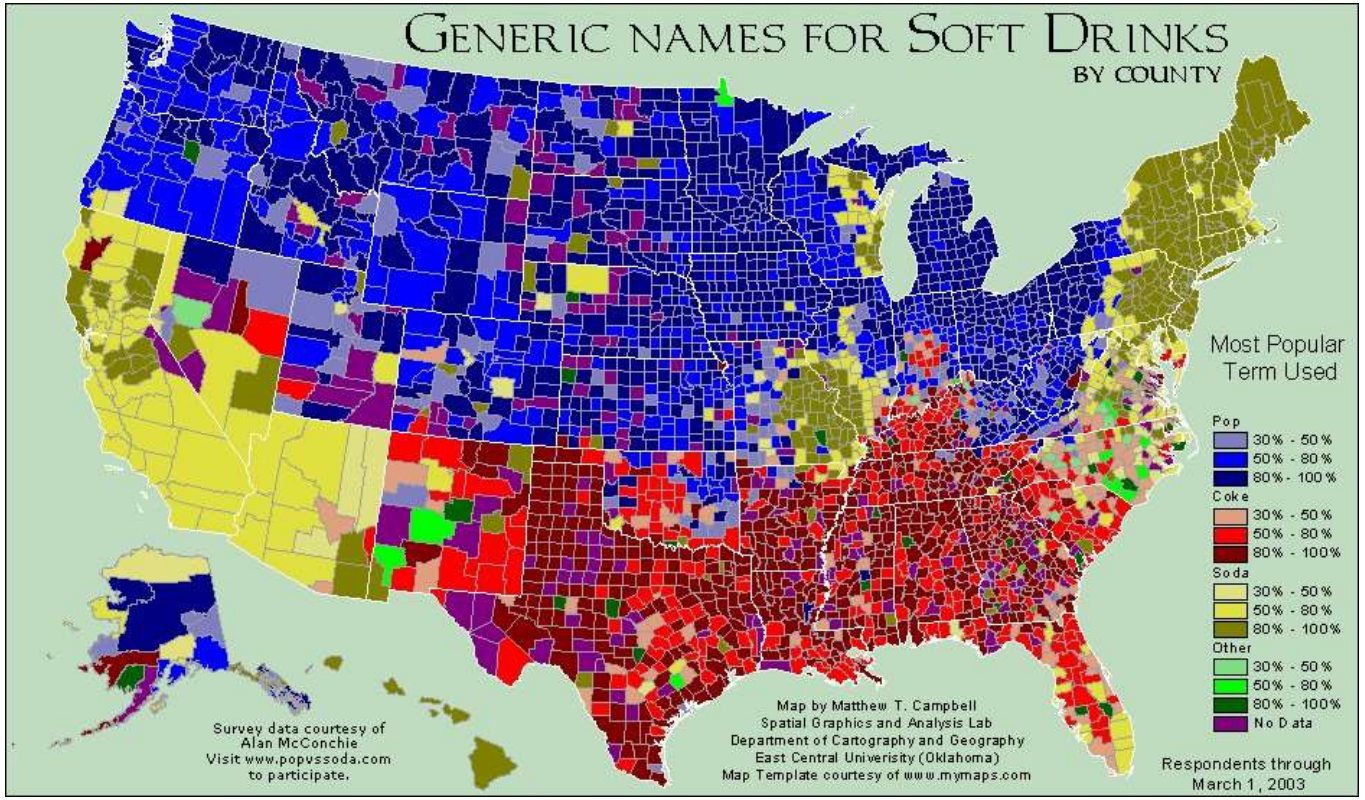


Why are these examples of chartjunk?

What could you do to remedy it?



Data Ink Working For Us



Evaluate this chart in terms of Data Ink.

Imagine this as a bar chart. As a table!!

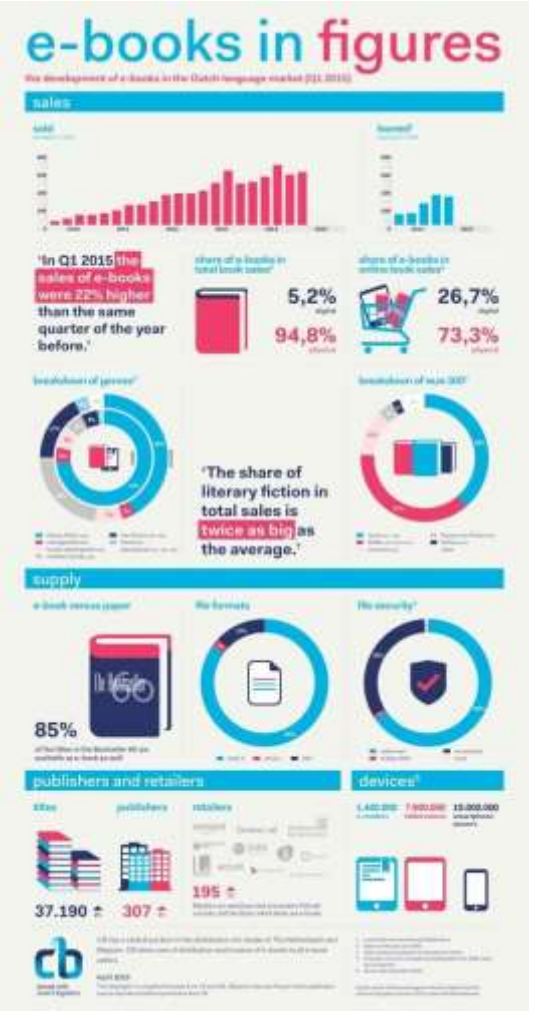
Review: Data principles (adapted from Tufte 2009)

1. The chart should tell a story
2. The chart should have graphical integrity
3. The chart should minimize graphical complexity

Tufte's fundamental principle:
Above all else show the data

Infographics

- Information graphics
- Visualization of information, data or knowledge intended to present information quickly and clearly
- We will have an ICA to create infographics using *Piktochart*.



Summary

- Use data visualization principles to assess a visualization
  - Tell a story
  - Graphical integrity (lie factor)
  - Minimize graphical complexity (data ink, chartjunk)
- Explain how a visualization can be improved based on those principles
- Types of visualization