Big Data and Data Visualization

Dr. Shantanu Ganguly

Fellow, Knowledge Resource Center

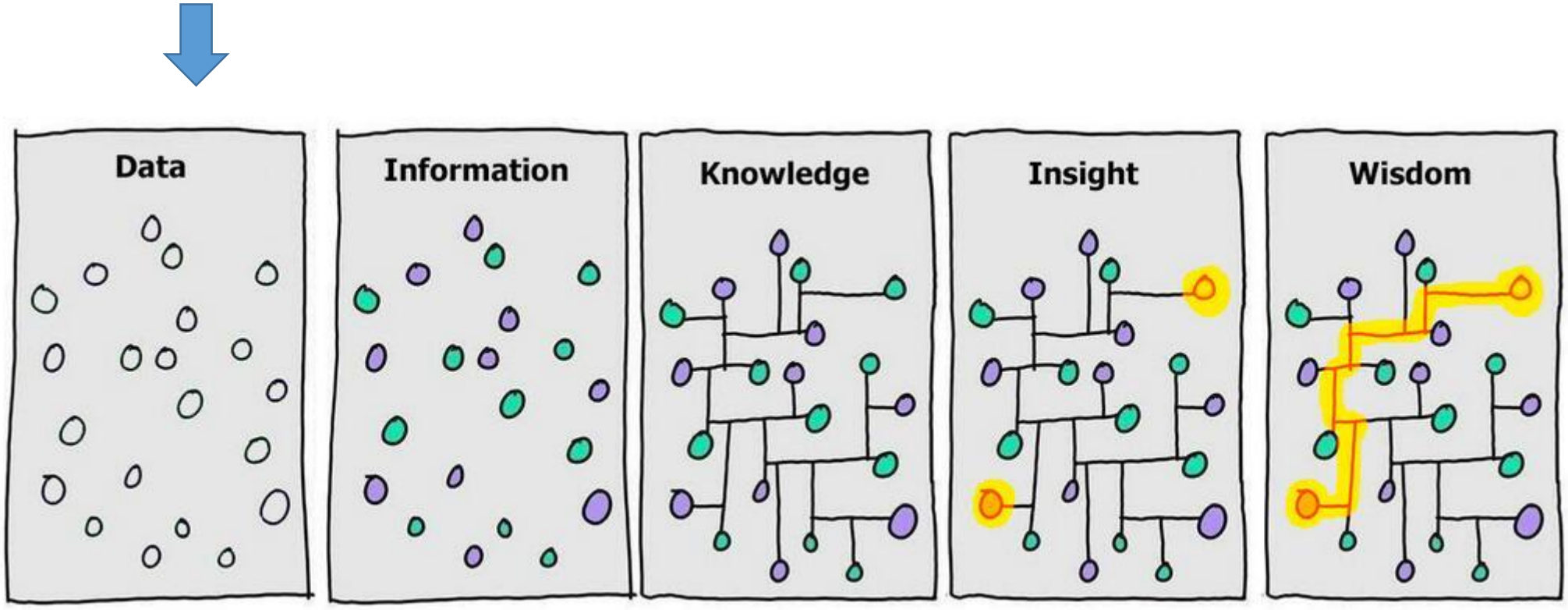
The Energy and Resources Institute (TERI)

Email: shantanu.ganguly@teri.res.in

Lesson Topics

- Big Data
- Big Data and Analytics
- Powerful data visualizations
- Understand visualizations
- Understand design considerations that lead to powerful data visualizations
- Understand effective techniques to create data visualizations
- Understand best practice tips for presenting data visualizations

The Information Continuum



Cartoon by David Somerville, based on a two pane version by Hugh McLeod

Types of Data

Quantitative Data

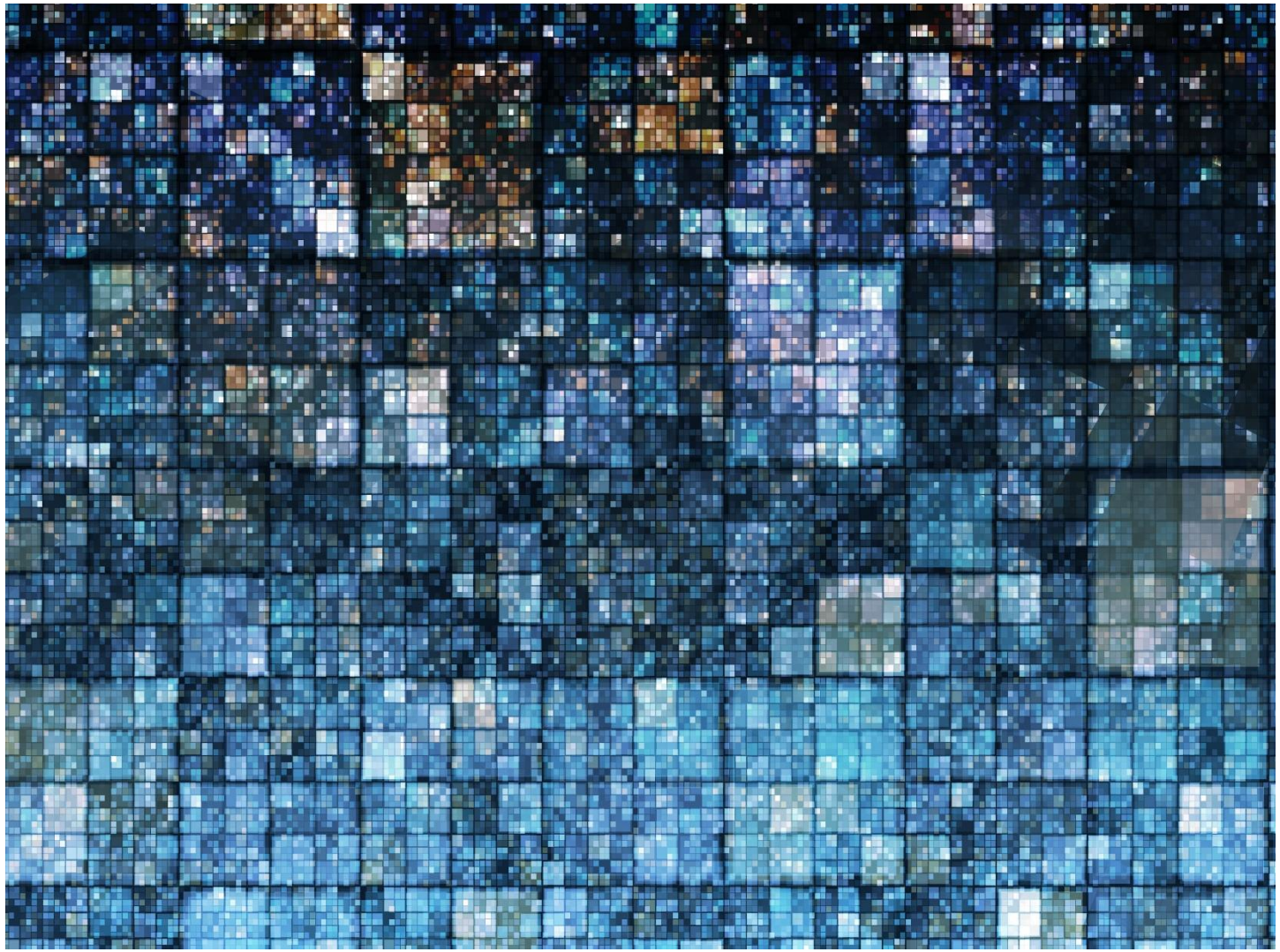
- Measurable
- Collected through measuring things that have a fixed reality
- Close ended

Qualitative Data

- Descriptive
- Collected through observation, field work, focus groups, interviews, recording or filming conversations
- Open ended

Big Data

Data that is too large or too complex to be managed using traditional data processing, analysis, and storage techniques.



What is Big Data?

- Big Data is a collection of large datasets that cannot be adequately processed using traditional processing techniques. Big data is not only data it has become a complete subject, which involves various tools, techniques and frameworks.

- Big data term describes the volume amount of data both structured and unstructured manner that adapted in day-to-day business environment. It’s important that what organizations utilize with these with the data that matters.

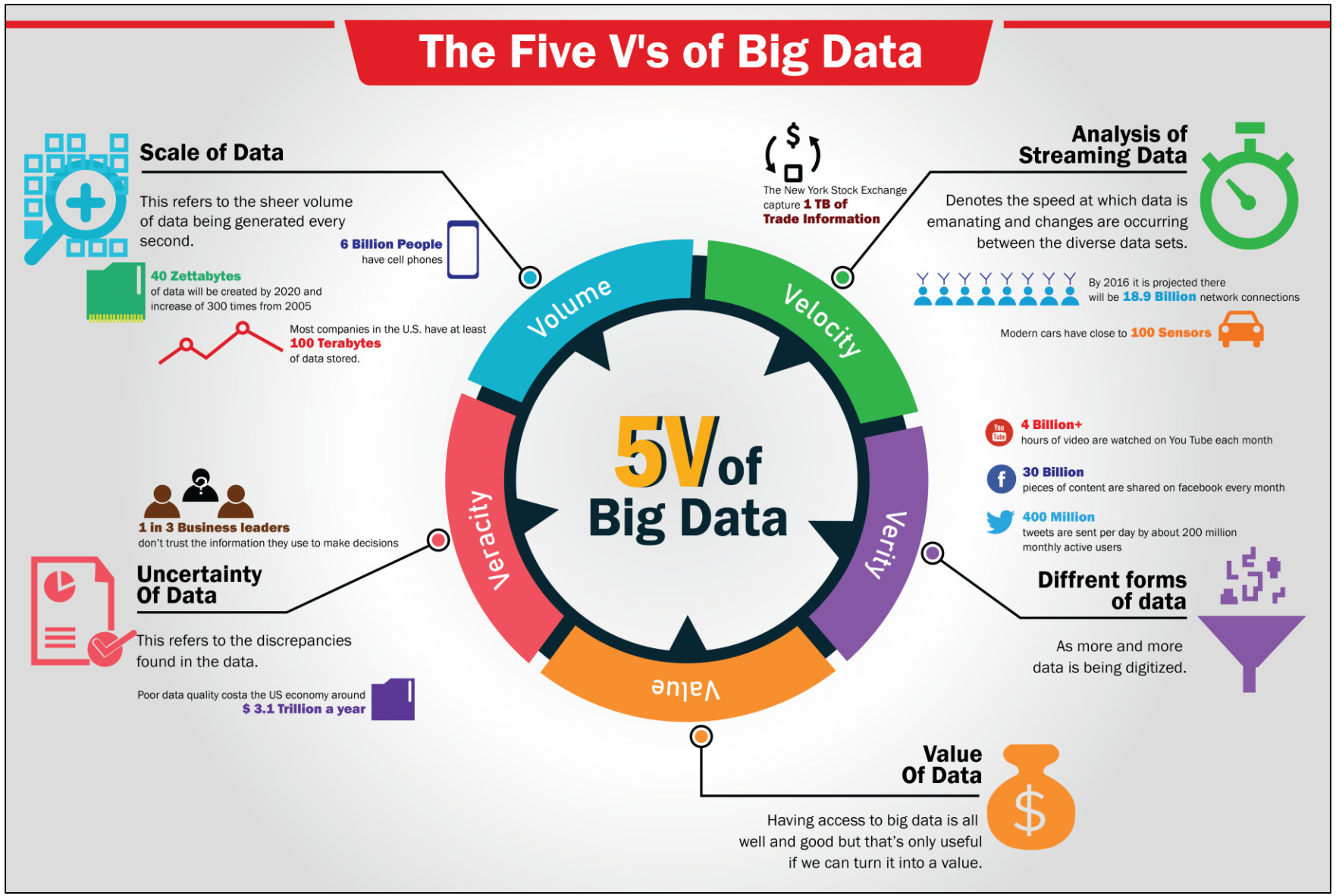
- Big data helps to analyze the in-depth concepts for the better decisions and strategic taken for the development of the organization.

The Evolution of Big Data

The concept of Big Data came into existence in the early 2000s when Industry analyst Doug Laney the definition of big data as the three categories as follows:

- **Volume:** Organizations collects the data from relative sources, which includes business transactions, social media and information from sensor or machine-to-machine data. Before, storage was a big issue but now the advancement of new technologies (such as Hadoop) has reduced the burden.
- **Velocity:** Data streams unparalleled speed of velocity and have improved in timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in real time operations.
- **Variety:** Data comes in all varieties in form of structured, numeric data in traditional databases to unstructured text documents, email, video, audio, stock ticker data and financial transactions.

The Five V's of Big Data



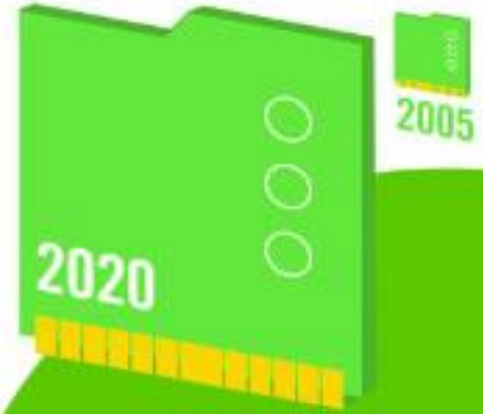
Volume

SCALE OF DATA

40 ZETTABYTES

[43 TRILLION GIGABYTES]

of data will be created by 2020, an increase of 300 times from 2005



It's estimated that

2.5 QUINTILLION BYTES

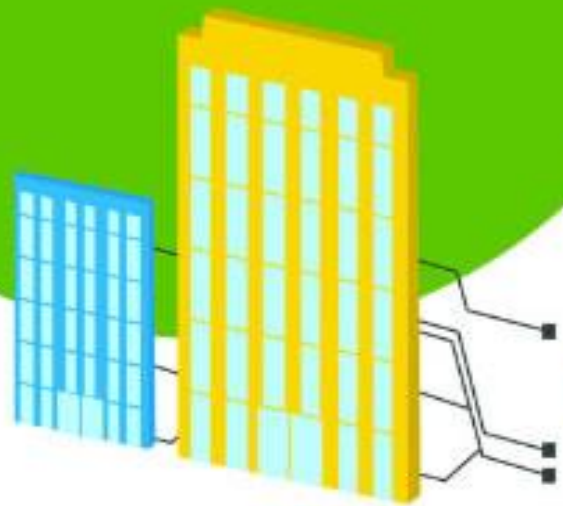
[2.3 TRILLION GIGABYTES]

of data are created each day

6 BILLION PEOPLE

have cell phones

WORLD POPULATION: 7 BILLION



Most companies in the U.S. have at least

100 TERABYTES

[100,000 GIGABYTES]

of data stored

Volume: scale of data

Unit	Value	Size
bit (b)	0 or 1	1/8 of a byte
byte (B)	8 bits	1 byte
kilobyte (KB)	1000^1 bytes	1,000 bytes
megabyte (MB)	1000^2 bytes	1,000,000 bytes
gigabyte (GB)	1000^3 bytes	1,000,000.000 bytes
terabyte (TB)	1000^4 bytes	1,000,000,000,000 bytes
petabyte (PB)	1000^5 bytes	1,000,000,000,000,000 bytes
exabyte (EB)	1000^6 bytes	1,000,000,000,000,000,000 bytes
zettabyte (ZB)	1000^7 bytes	1,000,000,000,000,000,000,000 bytes
yottabyte (YB)	1000^8 bytes	1,000,000,000,000,000,000,000,000 bytes

Volume: scale of data

- 90% of today’s data has been created in just the last 2 years
- Every day we create 2.5 quintillion bytes of data or enough to fill 10 million Blu-ray discs
- 40 zettabytes (4o trillion gigabytes) of data will be created by 2020, an increase of 300 times from 2005, and the equivalent of 5,200 gigabytes of data for every man, woman and child on Earth
- Most companies in the US have over 100 terabytes (100,000 gigabytes) of data stored

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session

Modern cars have close to

100 SENSORS

that monitor items such as fuel level and tire pressure

Velocity

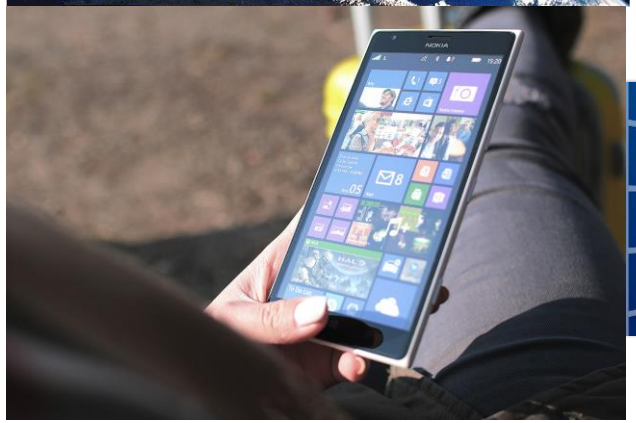
ANALYSIS OF STREAMING DATA

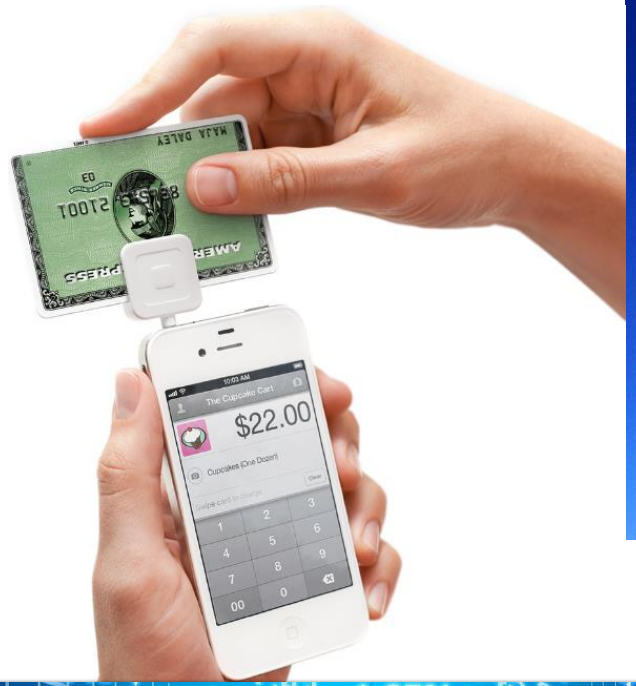
By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth

Velocity: analysis of streaming data





As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]

By 2014, it's anticipated there will be

420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

Variety

DIFFERENT FORMS OF DATA

4 BILLION+ HOURS OF VIDEO

are watched on YouTube each month

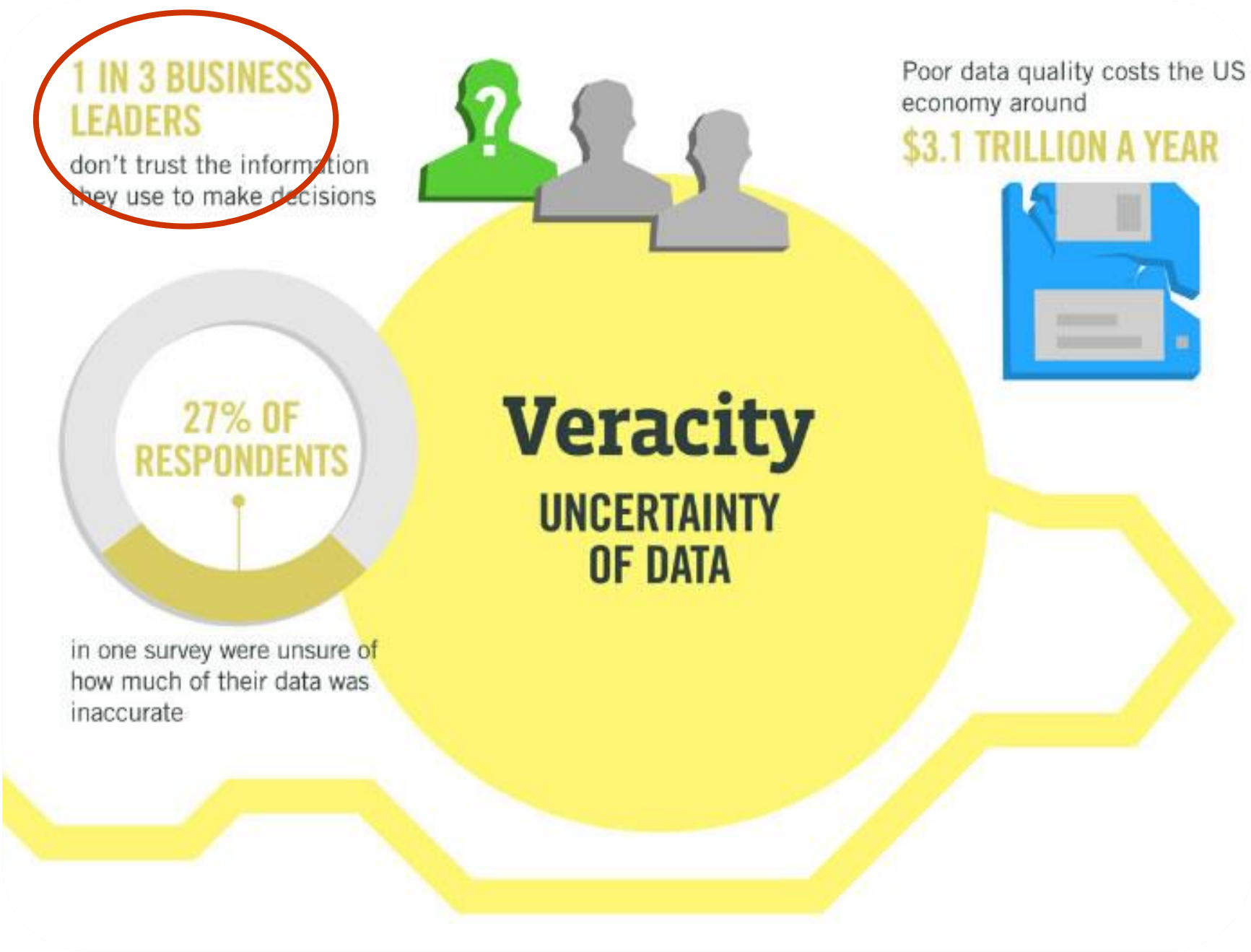
30 BILLION PIECES OF CONTENT

are shared on Facebook every month

400 MILLION TWEETS

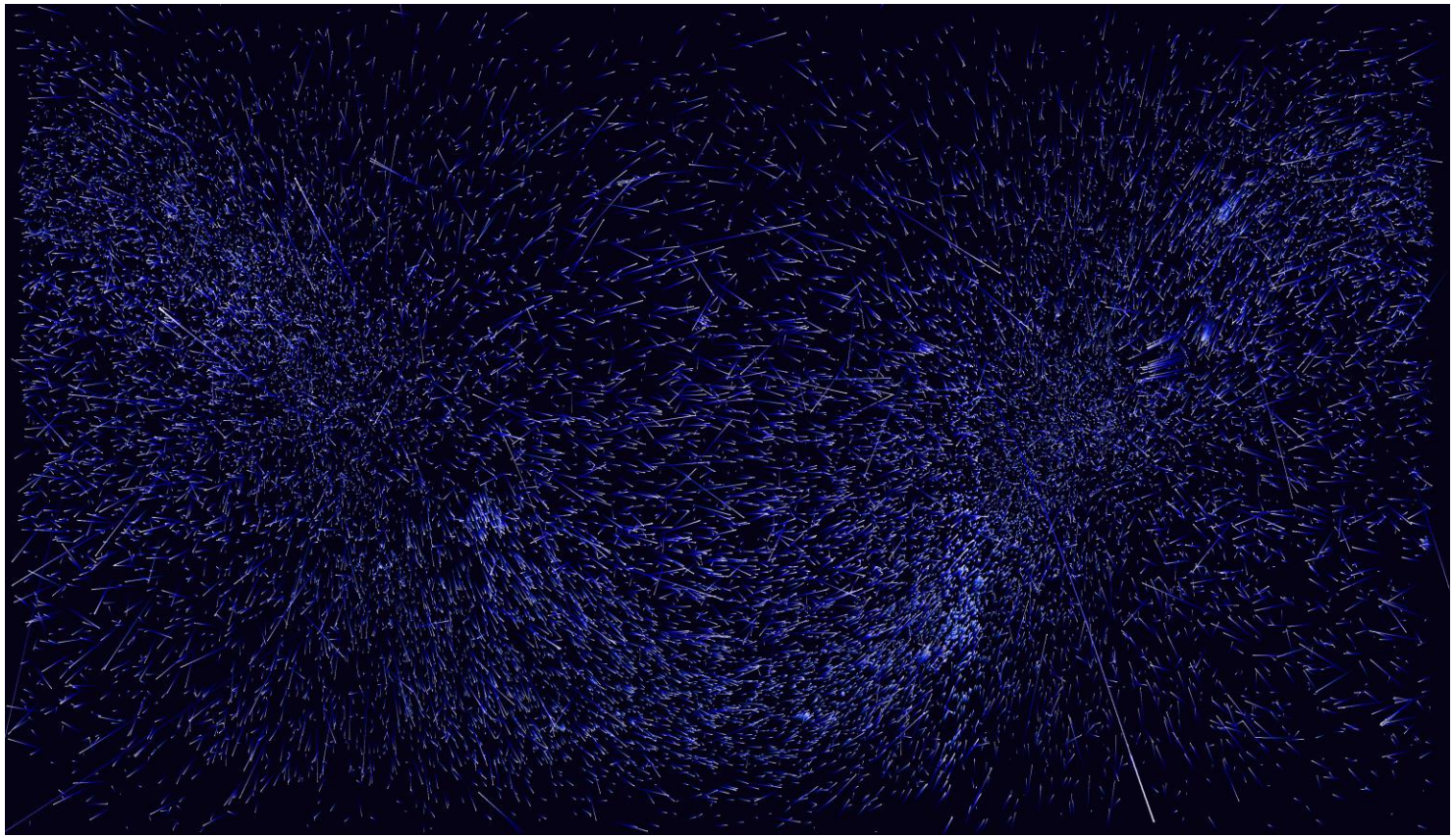
are sent per day by about 200 million monthly active users

Variety: different forms of data


1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions
Poor data quality costs the US economy around
$3.1 TRILLION A YEAR
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate
Veracity
UNCERTAINTY OF DATA

Veracity: trustworthiness of data

- Origin
- Authenticity
- Trustworthiness
- Completeness
- Integrity



Categories of Big Data - I

Big data works on the data produced by various devices and their applications.

- **Black Box Data:** It includes the conversation between crew members and any other communications (alert messages or any order passed) by the technical grounds duty staff.
- **Social Media Data**: Social networking sites such as Facebook and Twitter contains the information and the views posted by millions of people across the globe.
- **Stock Exchange Data**: It holds information (complete details of in and out of business transactions) about the 'buyer' and 'seller' decisions in terms of share between different companies made by the customers.

Categories of Big Data - II

- **Power Grid Data:** The power grid data mainly holds the information consumed by a particular node in terms of base station.
- **Transport Data:** It includes the data's from various transport sectors such as model, capacity, distance and availability of a vehicle.
- **Search Engine Data:** Search engines retrieve a large amount of data from different sources of database.

What is the importance of Big Data?

Combination of big data with high-powered analytics, you can have great impact on your business strategy such as:

- Finding the root cause of failures, issues and defects in real time operations.
- Generating coupons at the point of sale seeing the customer's habit of buying goods.
- Recalculating entire risk portfolios in just minutes.
- Detecting fraudulent behavior before it affects and risks your organization.

Who are the ones who use the Big Data Technology?

1. Banking
2. Government
3. Education
4. Health Care
5. Manufacturing
6. Retail

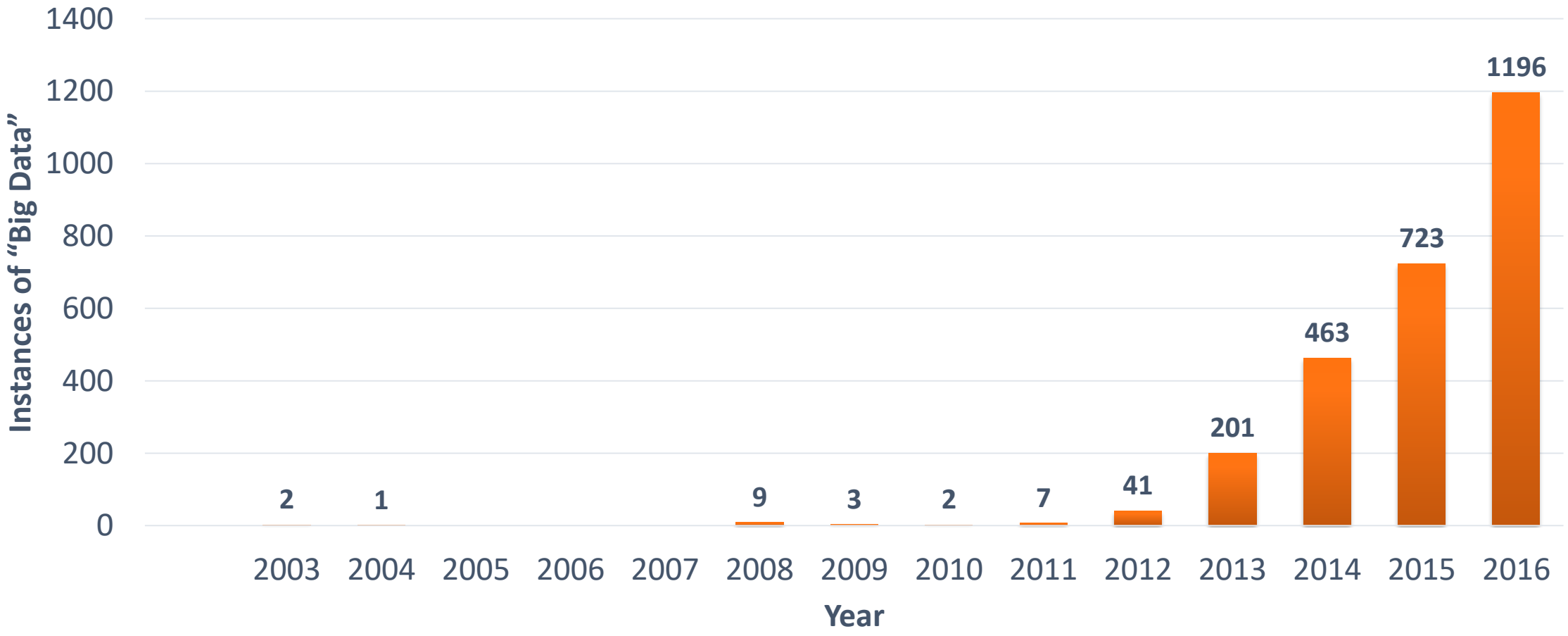
Brief explanation of how exactly businesses are utilizing Big Data

- What customers want ?
- What are the products are rapidly fast moving ?
- What are the expectations of the end users from the customer service?
- How to speed up the time sequence for marketing ?
- Methods on cost reduction.
- Methods to build economies of scale in a highly efficient manner.

Big Data Technologies

- Accurate analysis carried out based on big data which helps to increase and optimizes operational efficiencies, enable cost reductions, and reduce risks for the business operations.
- In order to capitalize on big data one should require infrastructure that manages and processes huge volumes of structured and unstructured data in real-time and can ensure data privacy and security.
- Many technologies are available in the market from different vendors which includes Amazon, IBM, Microsoft, etc., to approach big data.

“Big Data” on PubMed



Big Data and Librarians

- What role will librarians play in the Big Data revolution?
- Do you see yourself playing a part?
- How will you prepare yourself?
- What resources will you use?

Operational Big Data vs. Analytical Big Data

- It includes the applications such as MongoDB which provides operational capabilities for interactive and real time workloads where data is generally captured and stored.
- NoSQL Big Data systems are designed in such a way it capitalizes on new cloud computing architectures, to permit access on massive computations to be run reasonably and efficiently.
- It owns the systems like Massively Parallel Processing database systems and MapReduce which provides the analytical capabilities for re collective and complex analysis.
- MapReduce provides a new method for analyzing the data that flaunts its capabilities provided by SQL, MapReduce that can be scaled up from single servers to thousands of high and low end machines.

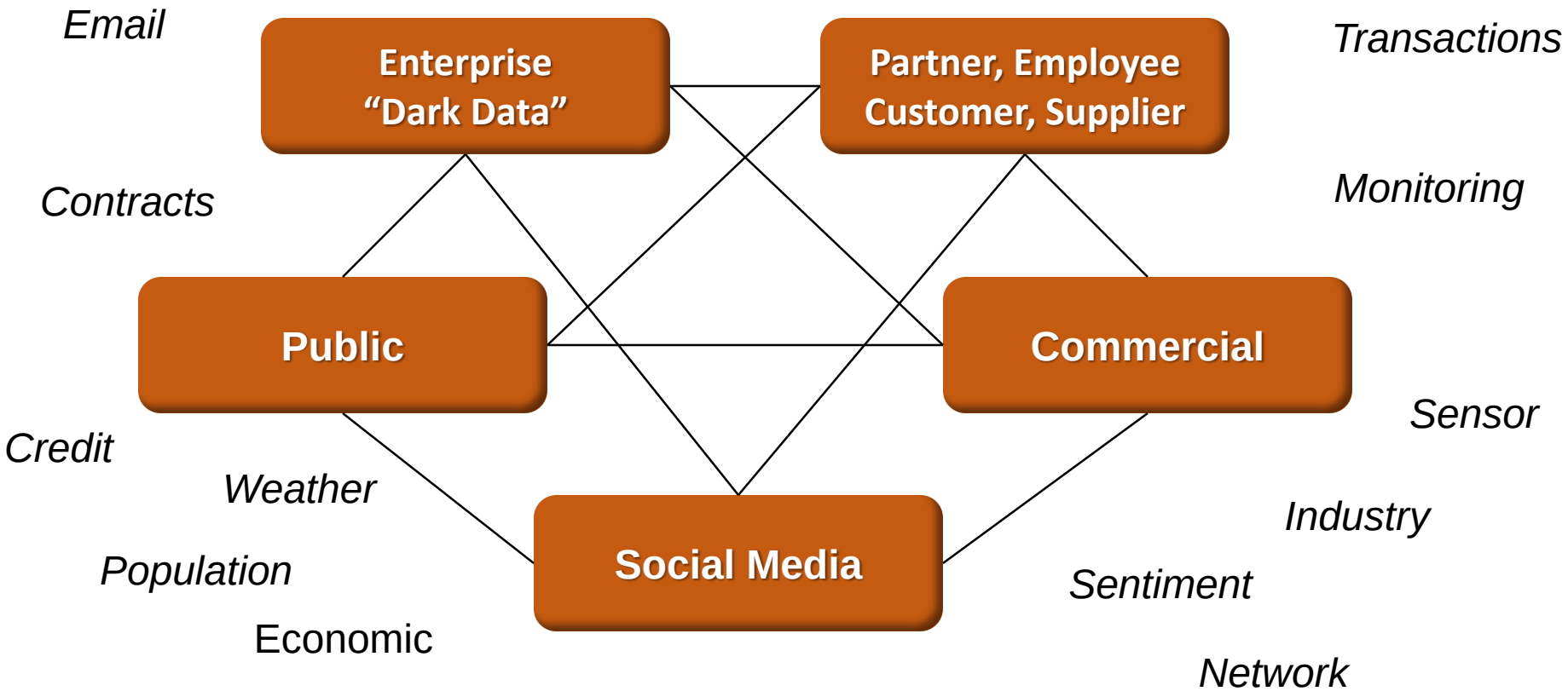
Barriers

Barriers that are imposed on big data are as follows:

- Capture data
- Storage Capacity
- Searching
- Sharing
- Transfer
- Analysis
- Presentation

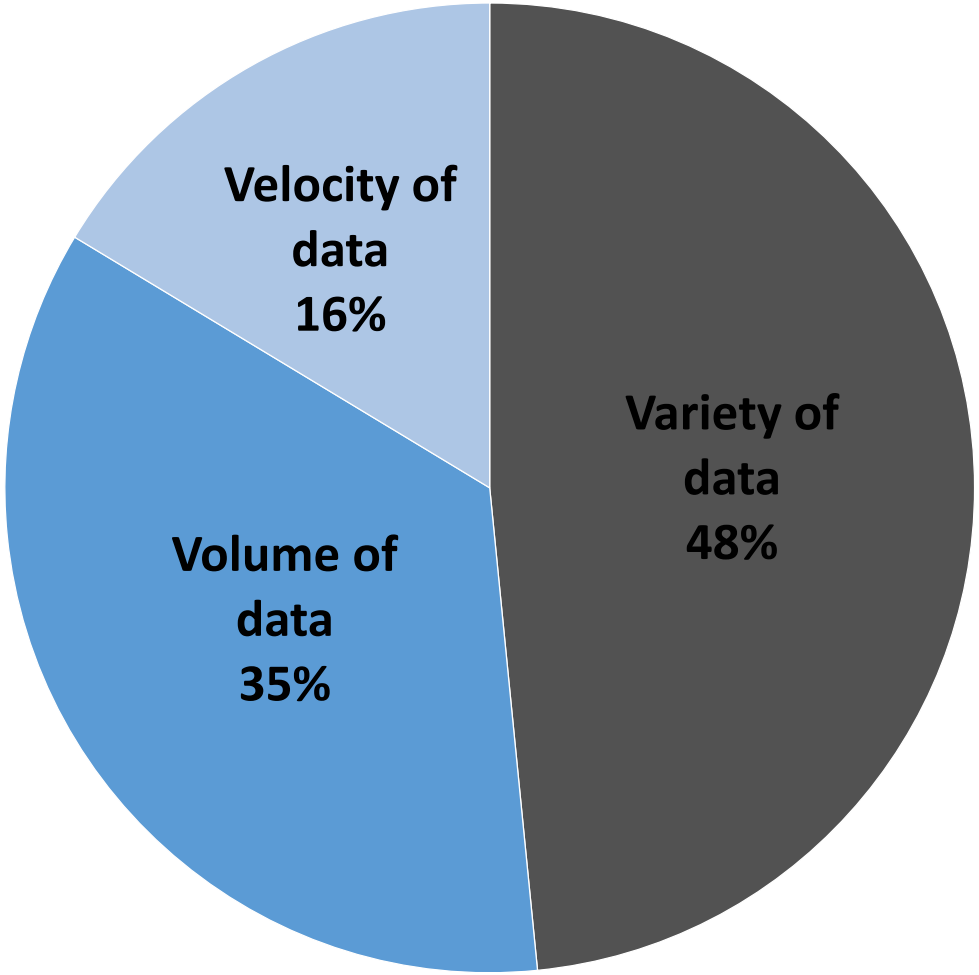
Big Data and Analytics

Where does Big Data come from?



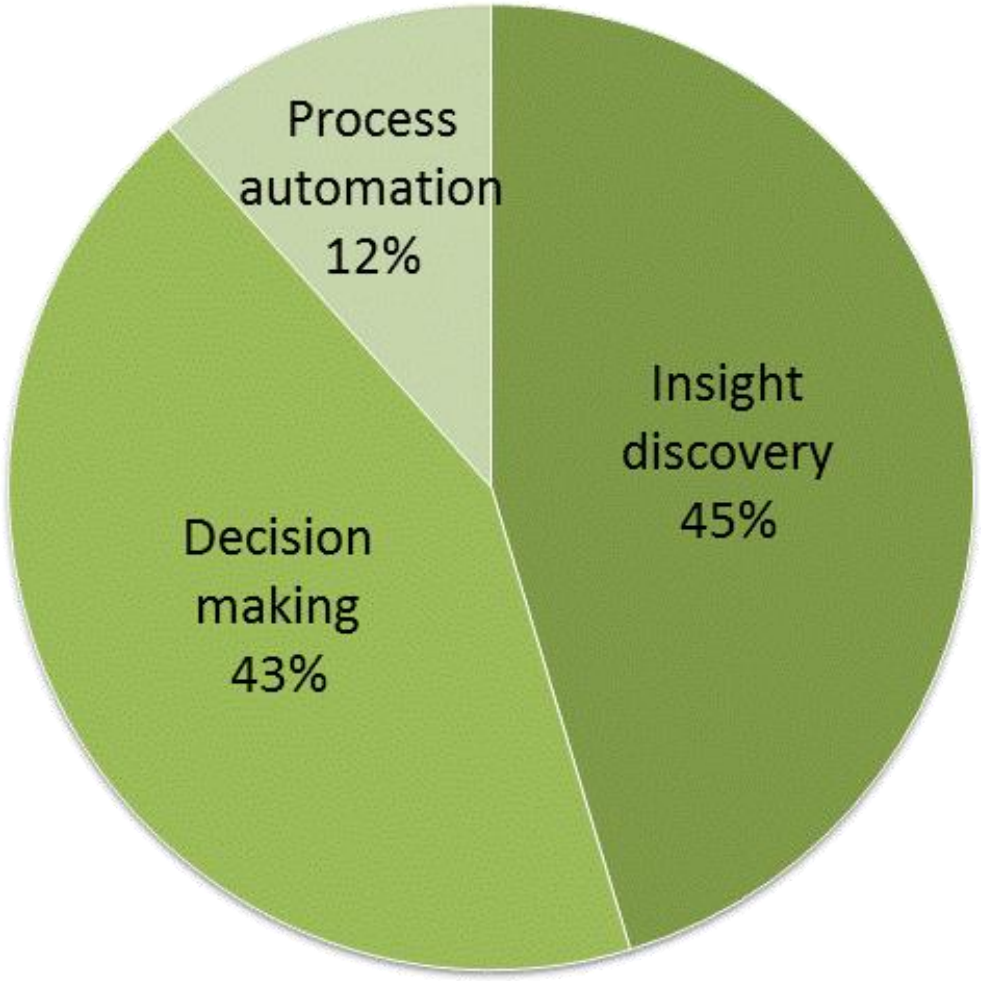
Types of Data

Which Big Data characteristic is the biggest issue for your organization?



Source: Getting Value from Big Data, Gartner Webinar, May 2012

Biggest opportunity for Big Data in your organization?



- 85% of Fortune 500 organizations will be unable to exploit big data for competitive advantage.
- Business analytics needs will drive 70% of investments in the expansion and modernization of information infrastructure.

Source: Getting Value from Big Data, Gartner Webinar, May 2012

Business Analytics

Business Analytics/ Business Intelligence

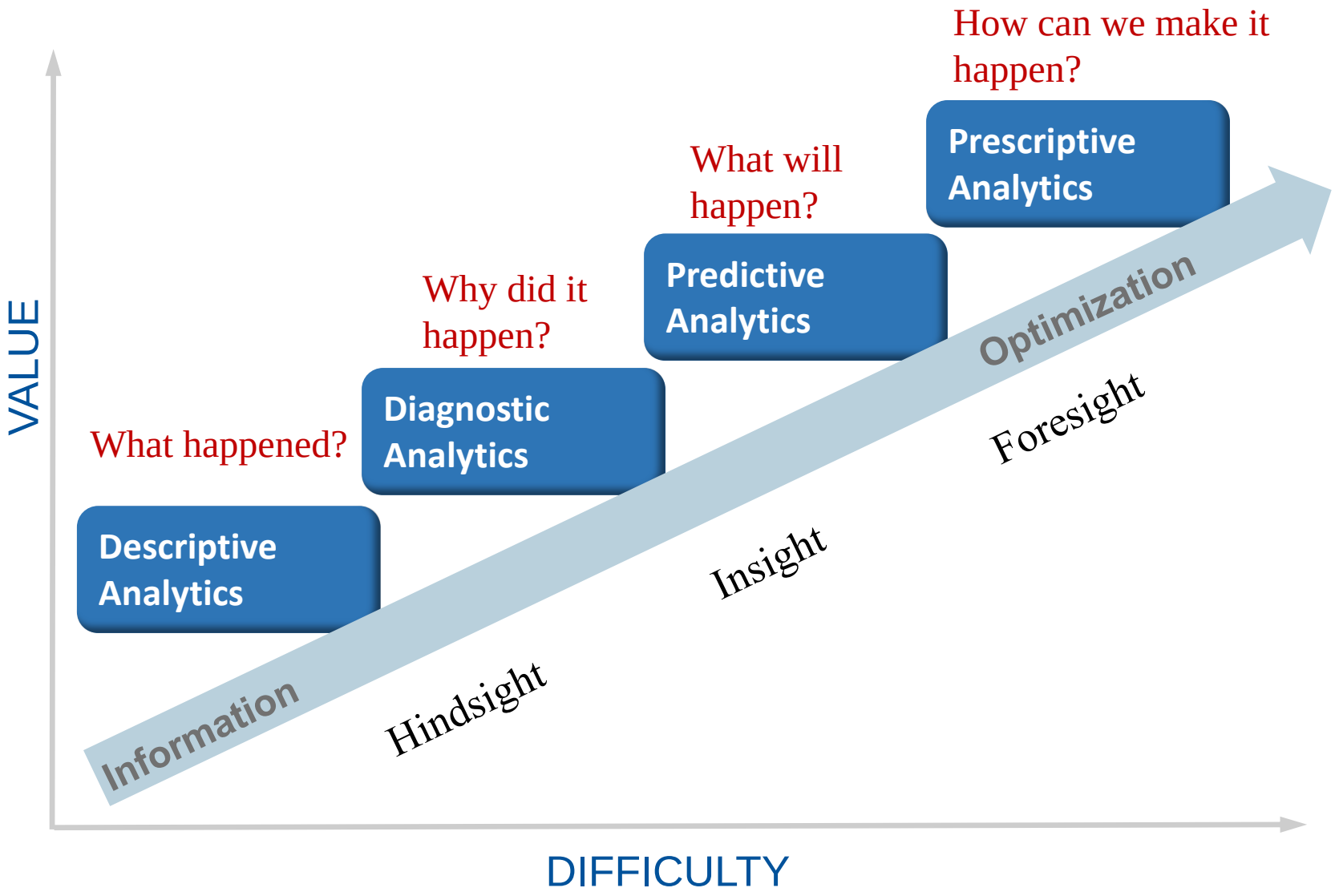
- Business Analytics/Business intelligence (BI) is a broad category of applications, technologies, and processes for:
  - gathering,
  - storing,
  - accessing, and
  - analyzing data
- to help business users make better decisions.

Things Are Getting More Complex

- Many companies are performing new kinds of analytics (**sentiment analysis, etc.), to better and more quickly understand and respond to what customers are saying about them and their products.
- The cloud, and appliances are being used as data stores
- Advanced analytics are growing in popularity and importance

****Sentiment analysis** (also known as **opinion mining**) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.

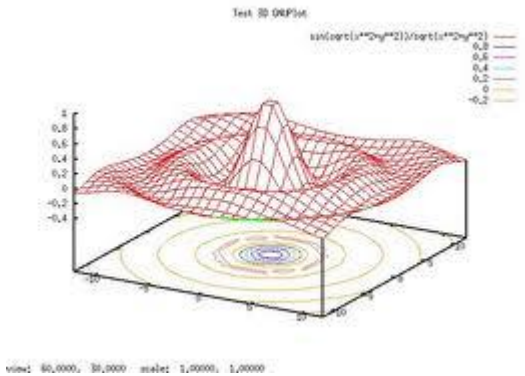
Analytics Models



Descriptive Analytics

- Descriptive analytics, such as reporting/ OLAP, dashboards, and data visualization, have been widely used for some time.
- They are the core of traditional BI.

Year 2000				
	Audio Division		Video Division	
Line Items	Budget	Actual	Budget	Actual
Cost of Goods Sold	$6,851,006.49	$7,132,961.38	$4,322,514.74	$4,526,954.71
Marketing Expense	$750,179.20	$756,596.17	$455,048.05	$462,815.40
Research and Development Expense	$538,243.39	$536,014.73	$329,890.95	$336,808.13
Selling Expense	$1,632,921.64	$1,579,790.18	$986,887.49	$927,970.90
Taxes	$314,659.05	$319,390.19	$202,636.67	$200,205.01
Year 2001				
	Audio Division		Video Division	
Line Items	Budget	Actual	Budget	Actual
Cost of Goods Sold	$2,564,556.31	$2,700,773.16	$1,726,031.16	$1,773,448.08
Marketing Expense	$294,766.22	$290,696.70	$167,757.29	$176,778.55
Research and Development Expense	$200,719.90	$193,236.83	$134,270.95	$125,725.88
Selling Expense	$620,427.30	$611,649.47	$405,092.93	$400,181.91
Taxes	$130,926.70	$122,526.31	$82,450.76	$80,671.87



What has occurred?

Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and predictive analytics.

Predictive Analytics

- Algorithms for predictive analytics, such as regression analysis, machine learning, and neural networks, have also been around for some time.

What will occur?

- Marketing is the target for many predictive analytics applications.
- Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and prescriptive analytics.

Prescriptive Analytics

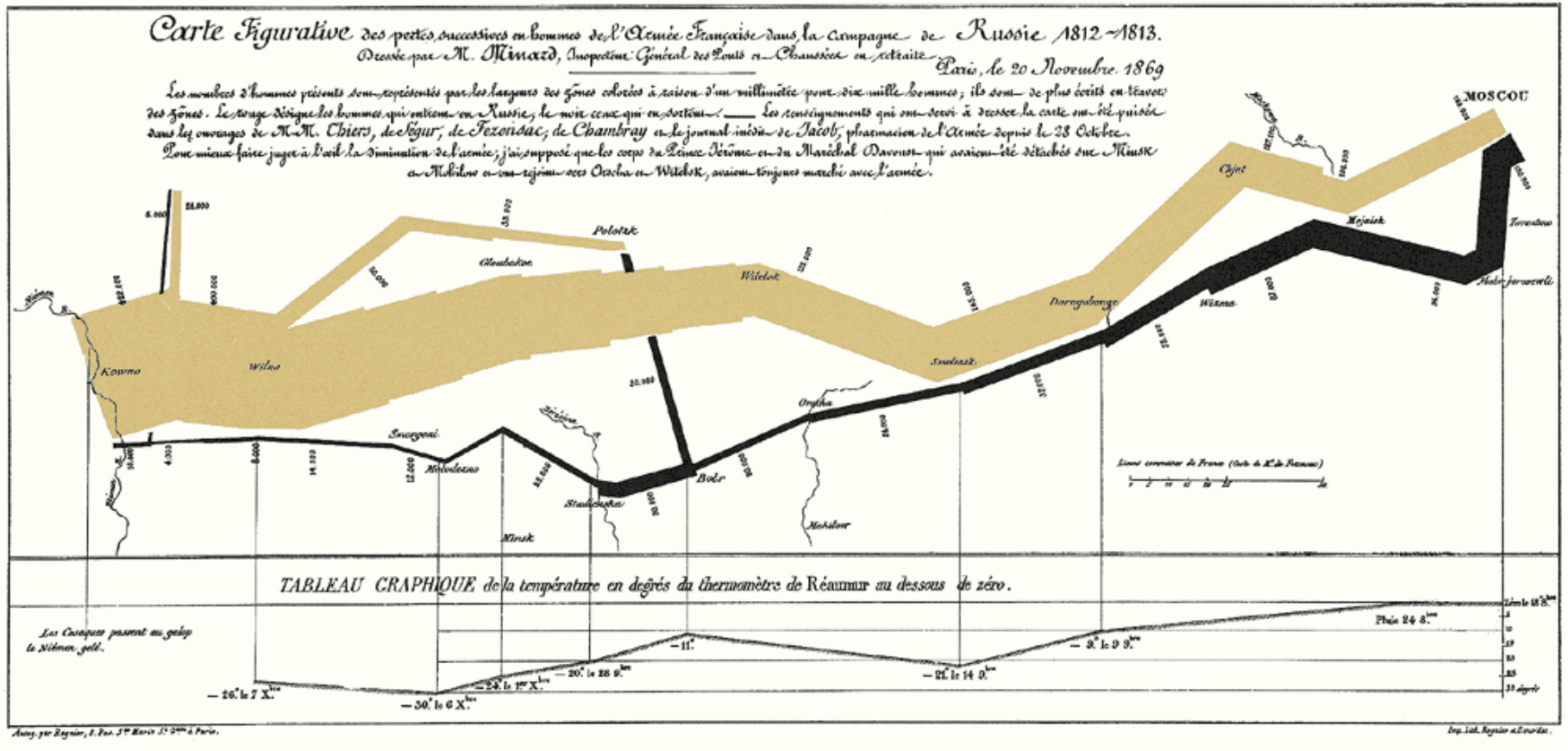
- Prescriptive analytics are often referred to as advanced analytics.
- Often for the allocation of scarce resources
- Optimization

What should occur?

Prescriptive analytics can benefit healthcare strategic planning by using analytics to leverage operational and usage data combined with data of external factors such as economic data, population demographic trends and population health trends, to more accurately plan for future capital investments such as new facilities and equipment utilization as well as understand the trade-offs between adding additional beds and expanding an existing facility versus building a new one.

A Brief History of Data Visualization

What makes a good chart?

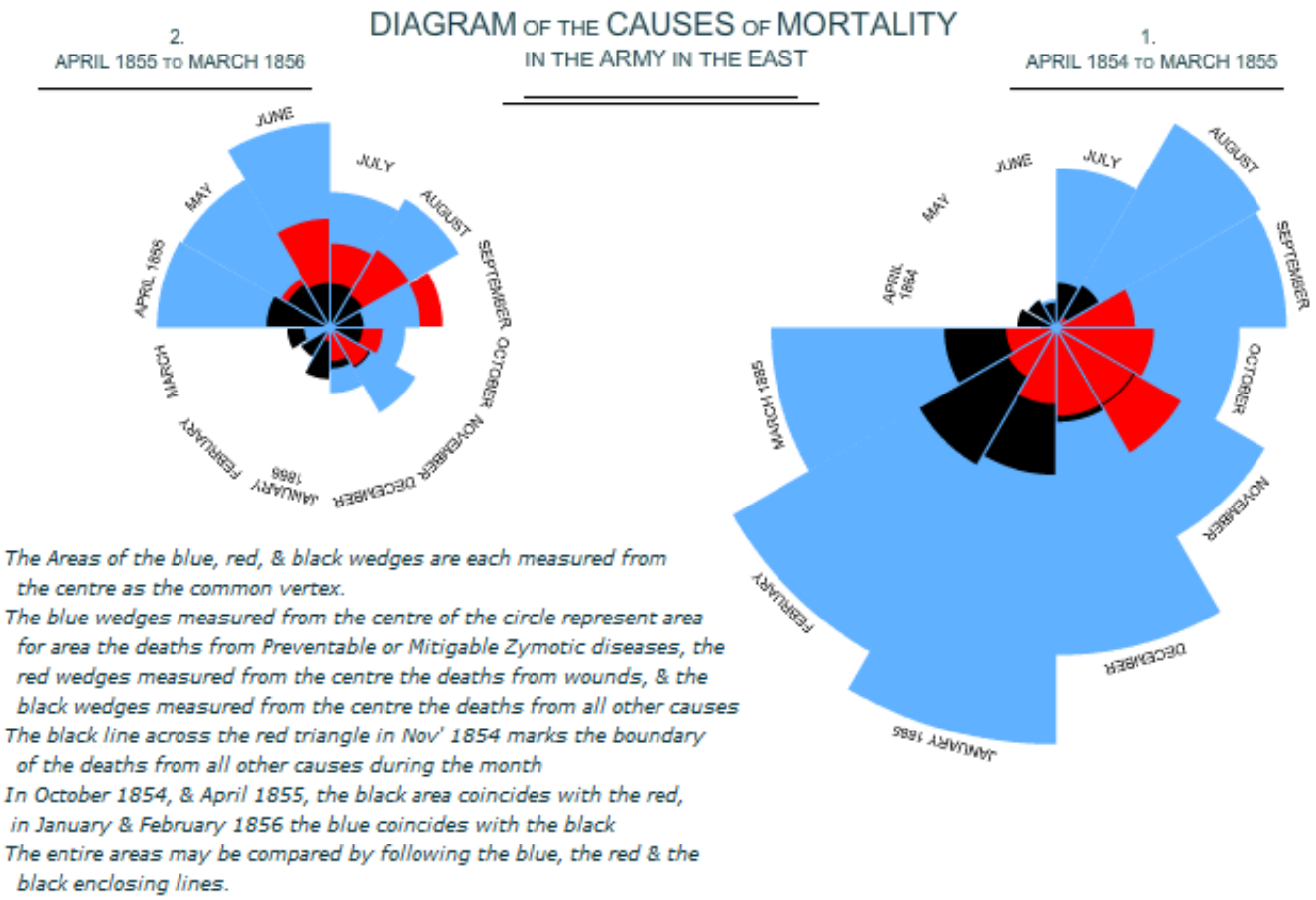


Napoleon's 1812 March by
Charles Joseph Minard

Reprinted in Tufte (2009), p. 41

Perhaps the most famous data presentation…

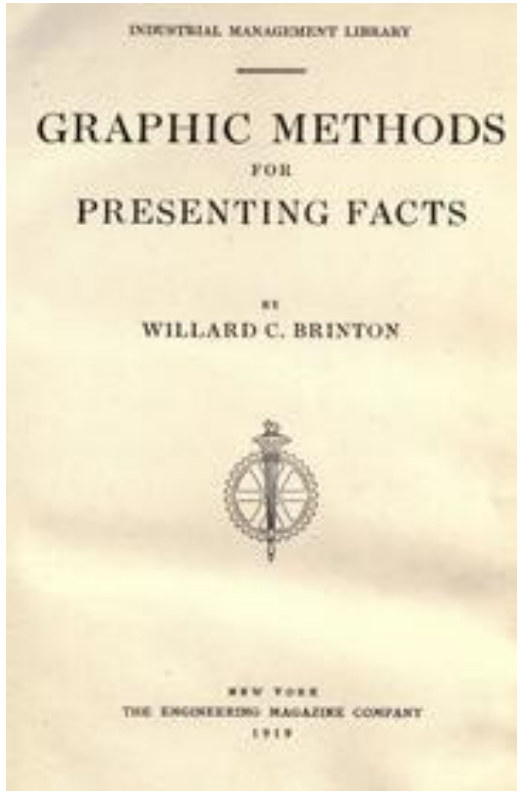
Florence Nightingale's 'Coxcombs‘1858



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month
In October 1854, & April 1855, the black area coincides with the red, in January & February 1856 the blue coincides with the black
The entire areas may be compared by following the blue, the red & the black enclosing lines.

- Pioneer hospital sanitation
- Meticulously gathered data
- Pioneer in applied statistics and visualization
- Nurse

Willard C. Brinton, 1914
First business book about visualization



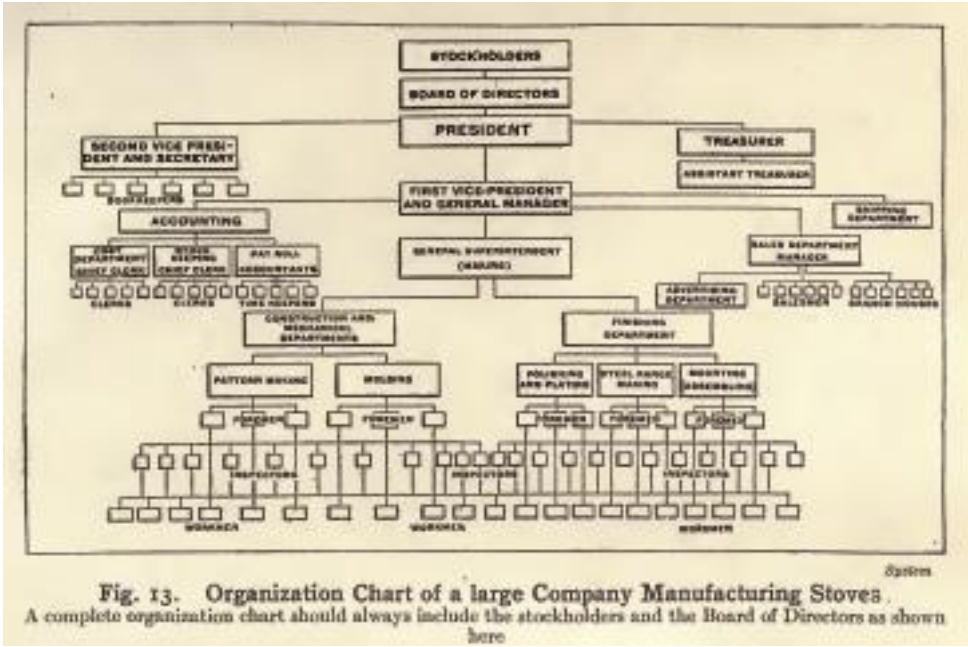
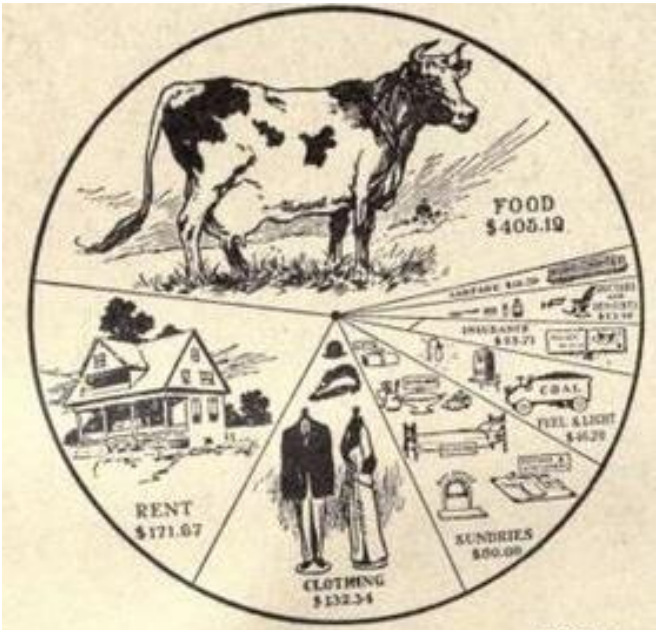


Fig. 13. Organization Chart of a large Company Manufacturing Stoves.
A complete organization chart should always include the stockholders and the Board of Directors as shown here



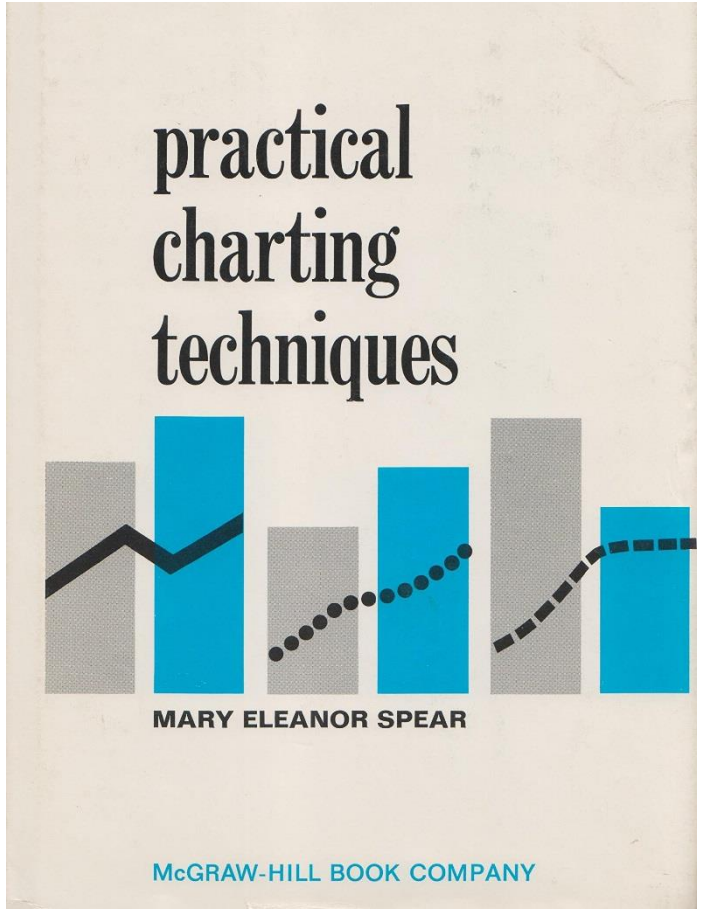
The Survey

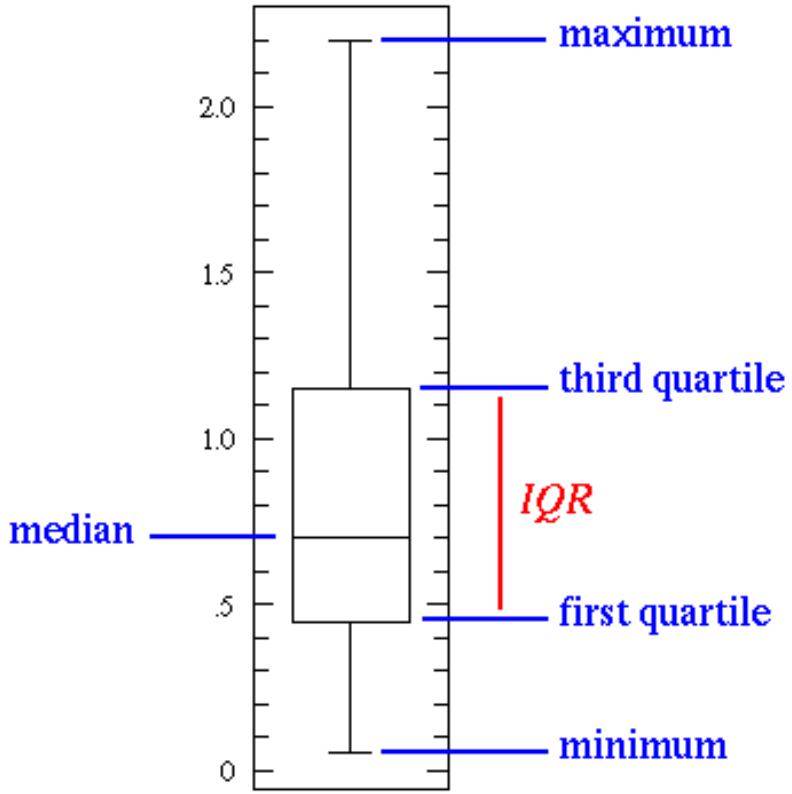
Fig. 3. Disposition of a Family Income of from $900 to $1000
This cut shows an attempt to put figures in popular form. The eye is likely to judge by the size of the pictures rather than by the angles of the sectors

- Rules for presenting data
- American consulting engineer

Mary Eleanor Spear 1952, 1969

- Common-sense advice
- Invented box plot
- Worked for various US government agencies

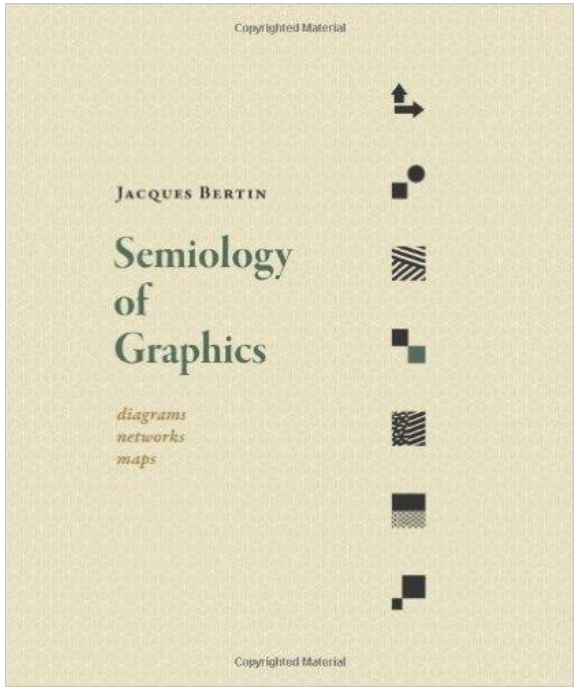


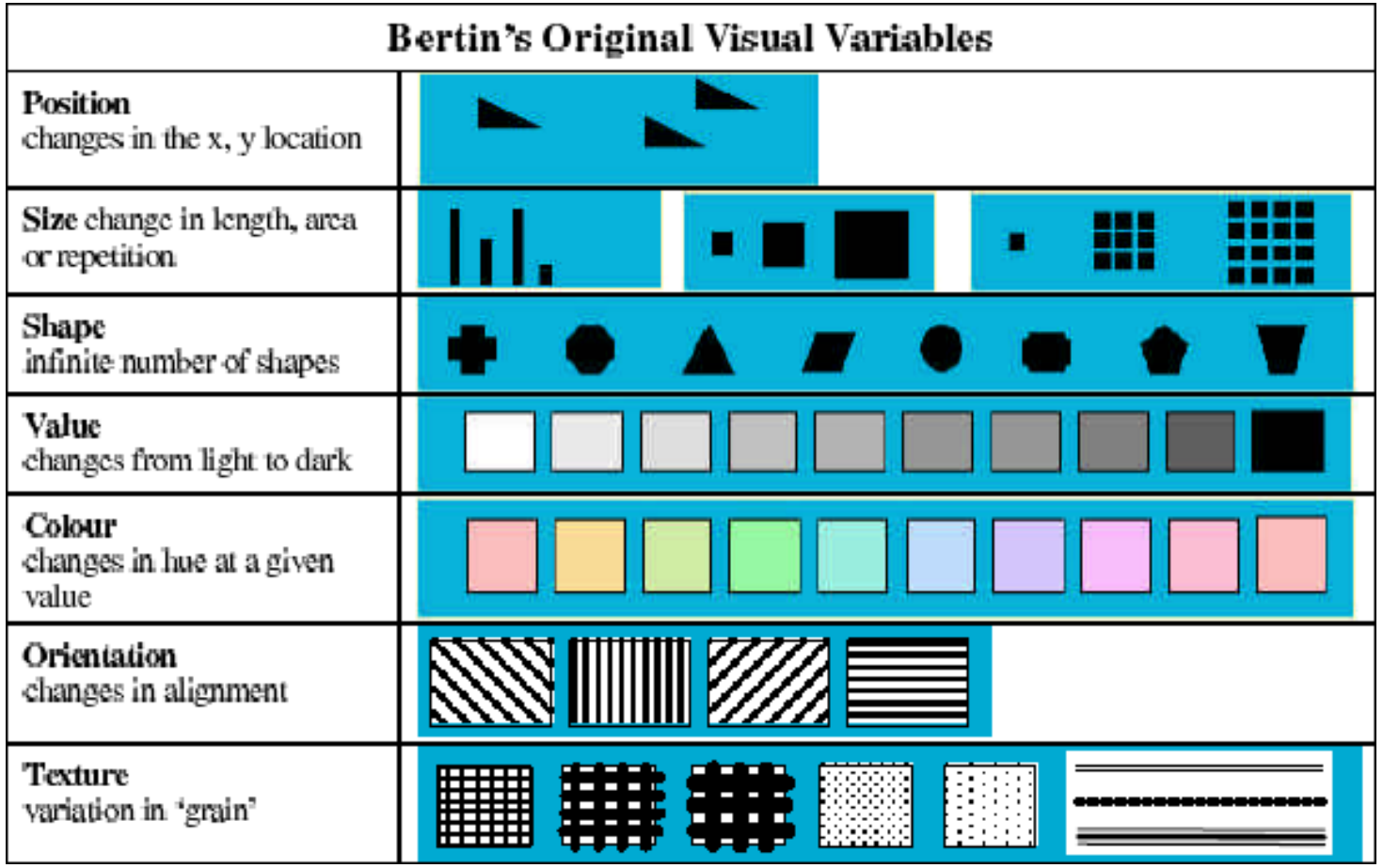


Jacques Bertin1967

- Principle of expressiveness:
  - Say everything you want to say — no more, no less
  - Don’t mislead
- Principle of effectiveness:
  - Use the best method available for showing your data
- Cartographer

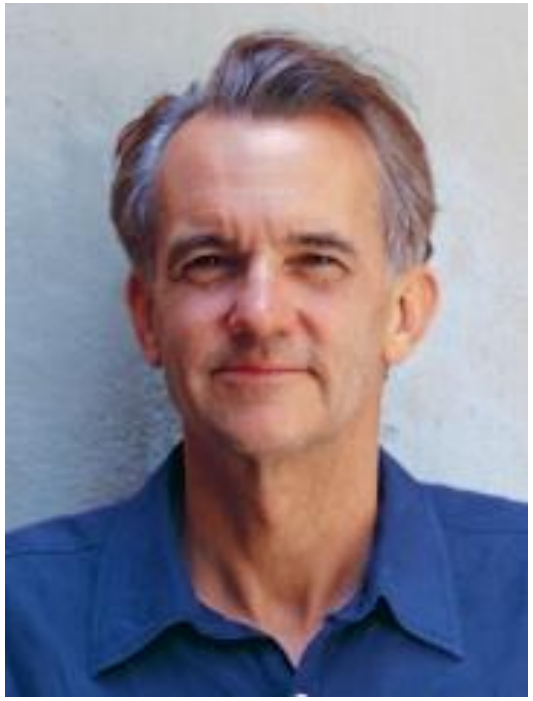
Jacques Bertin
Seven Visual Variables

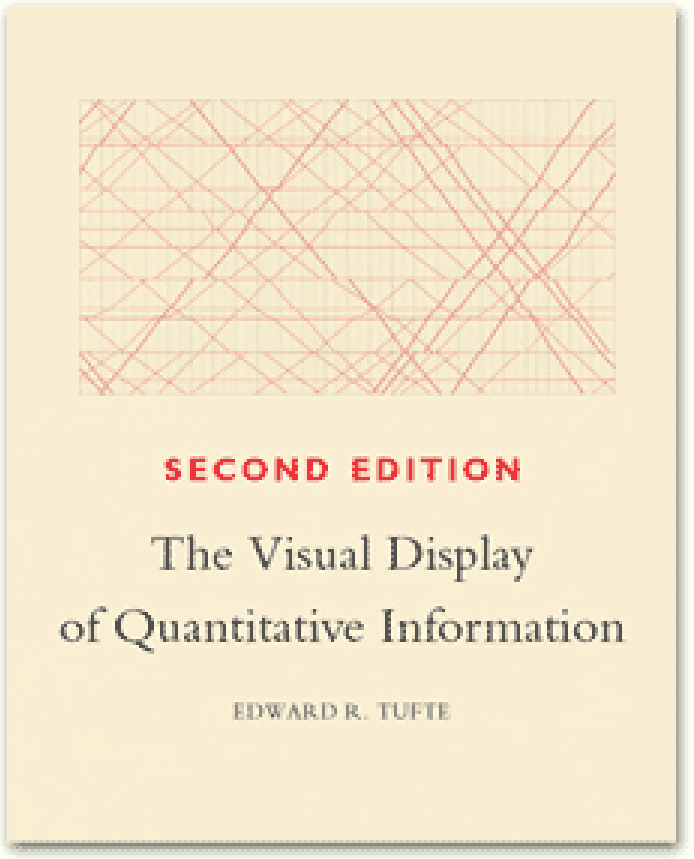




- Position
- Size
- Shape
- Color
- Brightness
- Orientation
- Texture

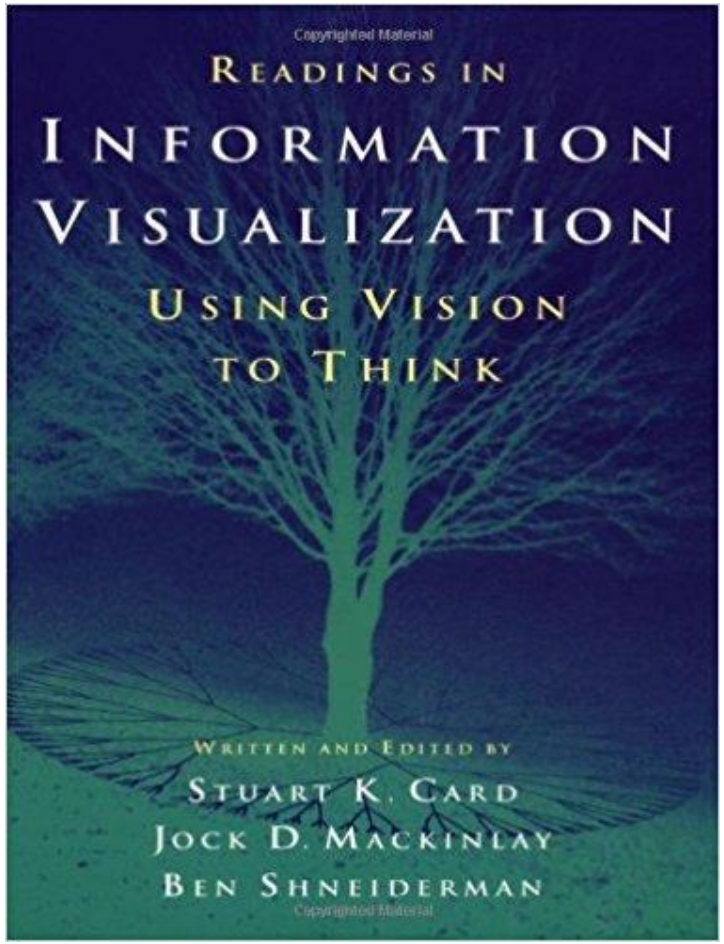
Edward Tufte
1983





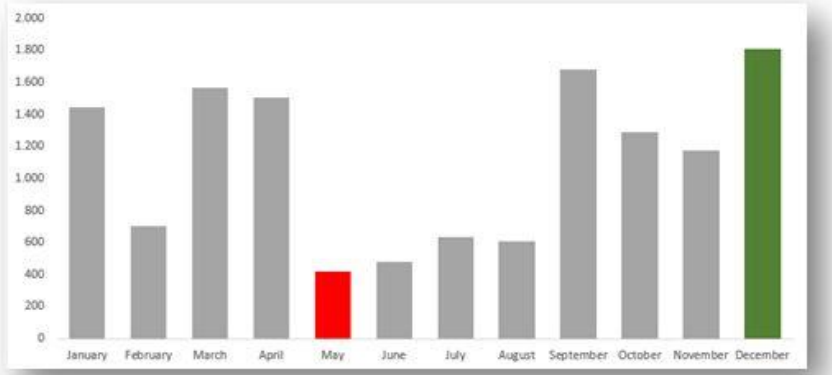
- Disciplined design principles
- Minimalist approach
- Professor emeritus at Yale University

Jock Mackinlay 1986



- Automatically encode data with software
- Enable people to focus on ideas, concepts
- Added eighth variable to Bertin's list: motion
- VP of Research and Design, Tableau Software

When a Chart hits our Eyes



1. Visuals aren't read in a predictable, linear way
   - Create charts spatially, from the visual outward
2. We see first what stands out
   - Whatever stands out should support idea
3. We see only a few visuals at once
   - Plot as few visual elements as possible
4. We seek meaning and make connection
   - Relate visual elements in a meaningful way
5. We rely on conventions and metaphors
   - Embrace deeply ingrained conventions

Example: USA Energy Resources

massima tra gli animali della categoria

Numero di specie scoperte

Percentuale di specie scoperte sul numero di specie stimate

Gli animali formano un **regno immenso**

Mammiferi

5.501

98%

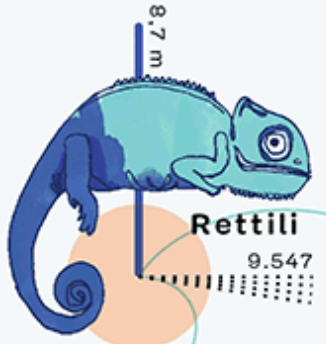
Uccelli

2.7 m

10.064

96%

8.7 m



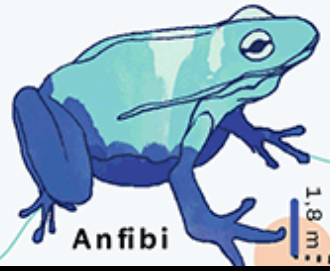
Rettili

9.547

80%

e composto da così tante **specie**,

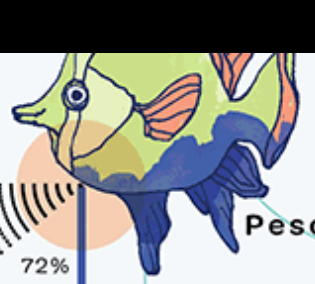
ognuna con le proprie caratteristiche,



Anfibi

1.8 m

6.771



32.400

72%

Pesci

18 m

14 m

43%

85.000

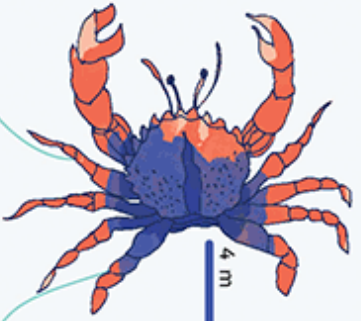
Molluschi

Pare che esistano **20 miliardi** di miliardi di animali nel mondo!

che gli uomini non hanno ancora trovato il tempo

Crostacei

di classificarli **tutti!**



4 m

47.000

31%

30 cm

102.246

17%

Aracnidi

Suddivisi **in ben 8,7 milioni** di specie diverse!

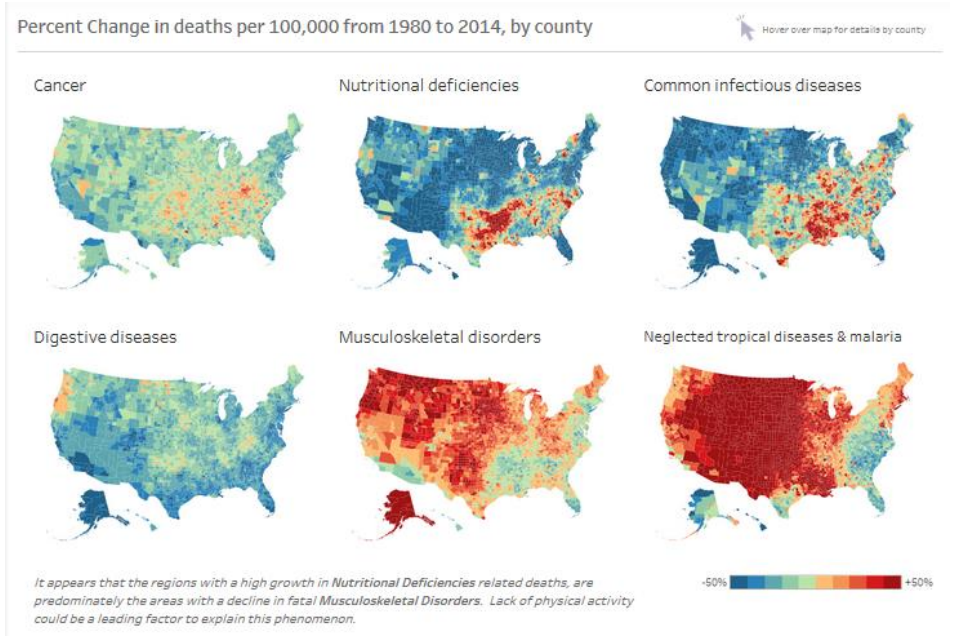
Insetti

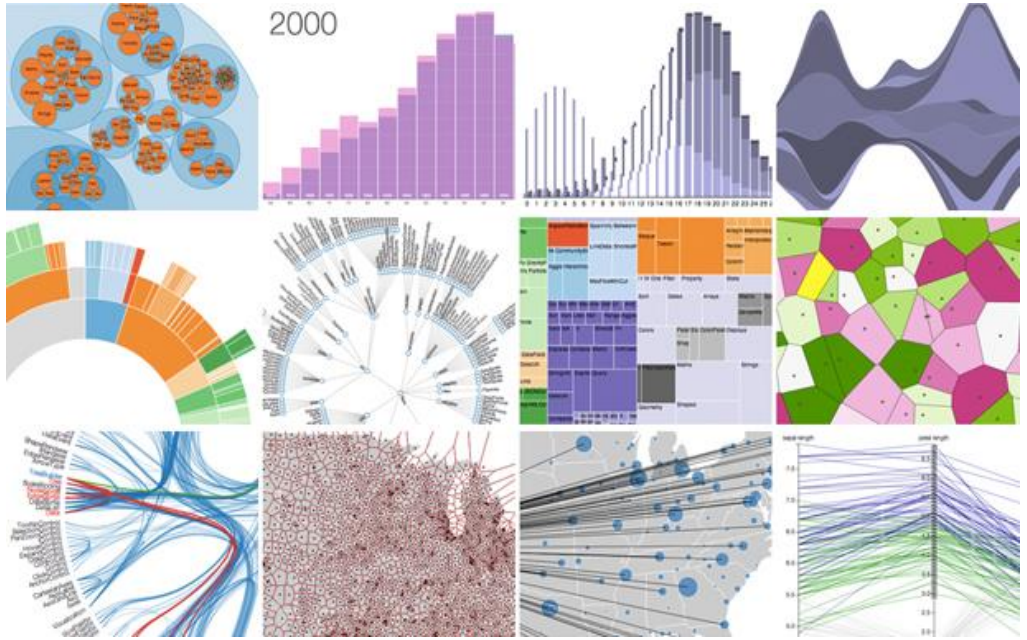
57 cm

20%

1.000.000

Vast Data Visualization Choice





What is data visualization and why is it important?

- **Data visualization** is the act of taking information (**data**) and placing it into a visual context, such as a map or graph. **Data visualizations** make big and small **data** easier for the human brain to understand, and **visualization** also makes it easier to detect patterns, trends, and outliers in groups of **data**.

Is data visualization a part of data science?

- **Data science** and **data visualization** are not two different entities. They are bound to each other. **Data visualization** is a **subset of data science**. **Data science** is not a single process or a method or any workflow.

What are the best data visualization software of 2019?

- WhSisense.
- Looker.
- Periscope Data.
- Zoho Analytics.
- Tableau.
- Domo.
- Microsoft Power BI.
- Qlikview.

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.

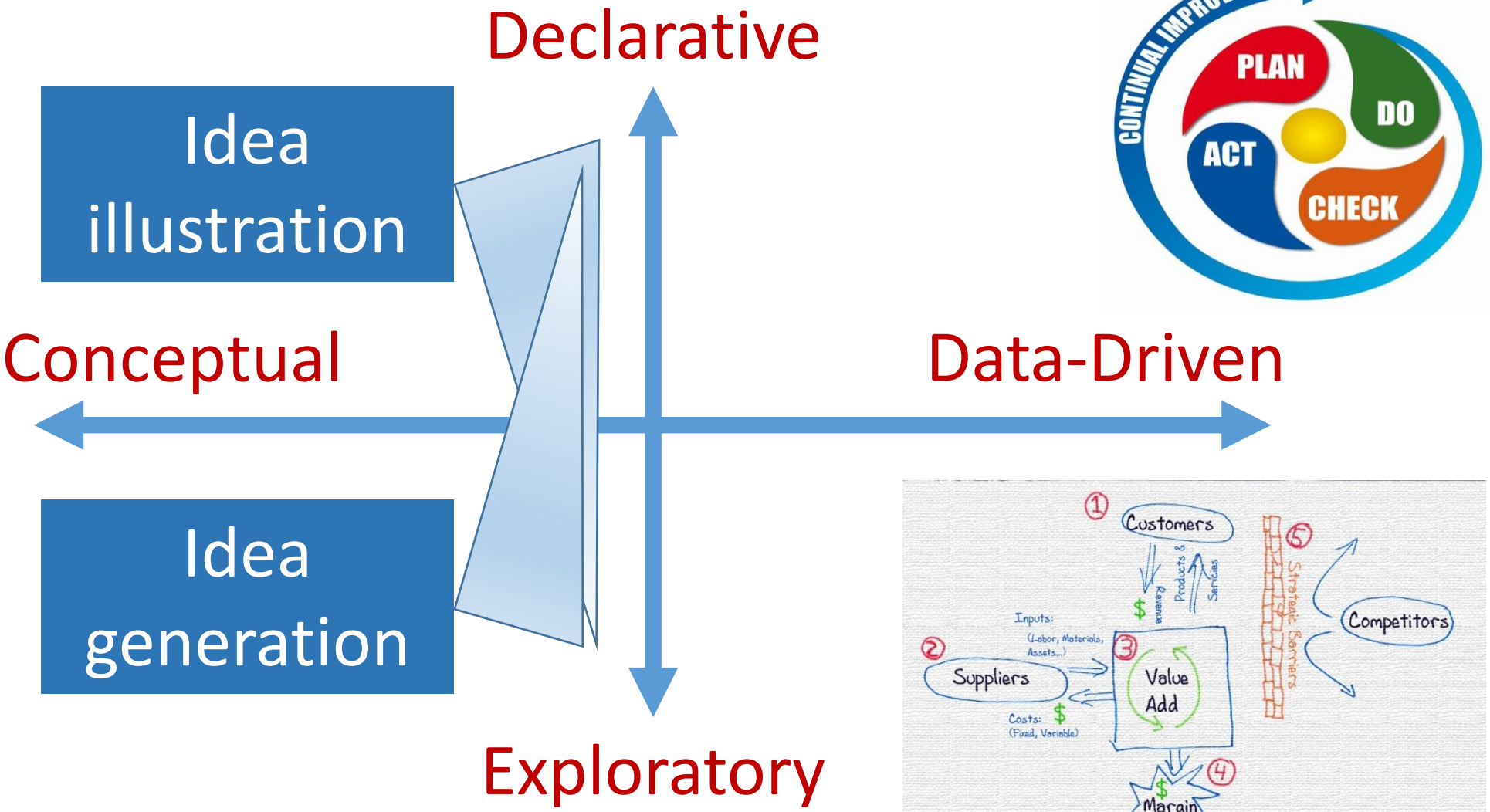
How do you create good data visualization?

- 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.

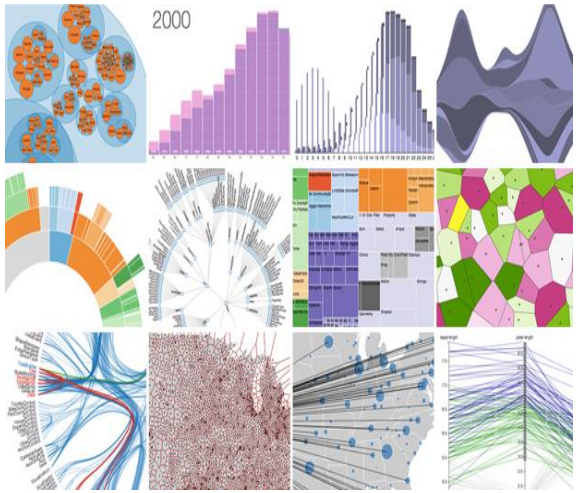
What kind of visual communication do you want to create?

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

Four Types of Data Visualizations



Four Types of Data Visualizations



Declarative

Idea illustration

Everyday dataviz

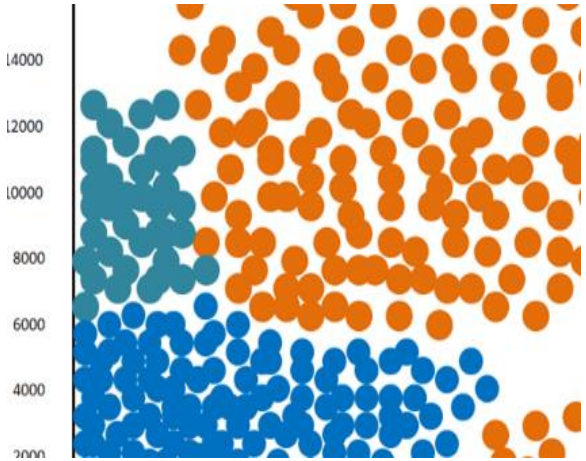
Conceptual

Data-Driven

Idea generation

Visual discovery

Exploratory



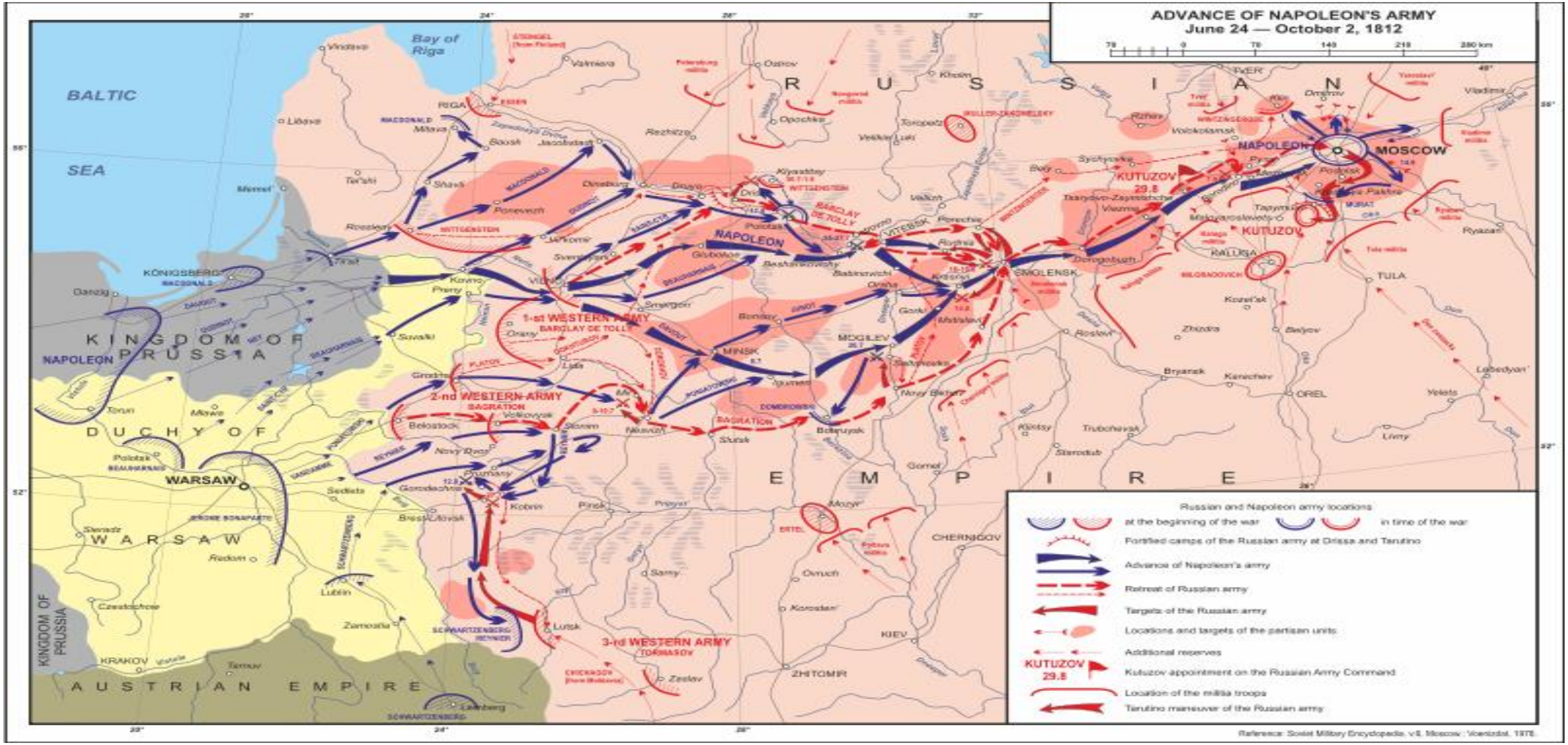
Data Visualization

provide clear understanding of patterns in data

detect hidden structures in data

condense information

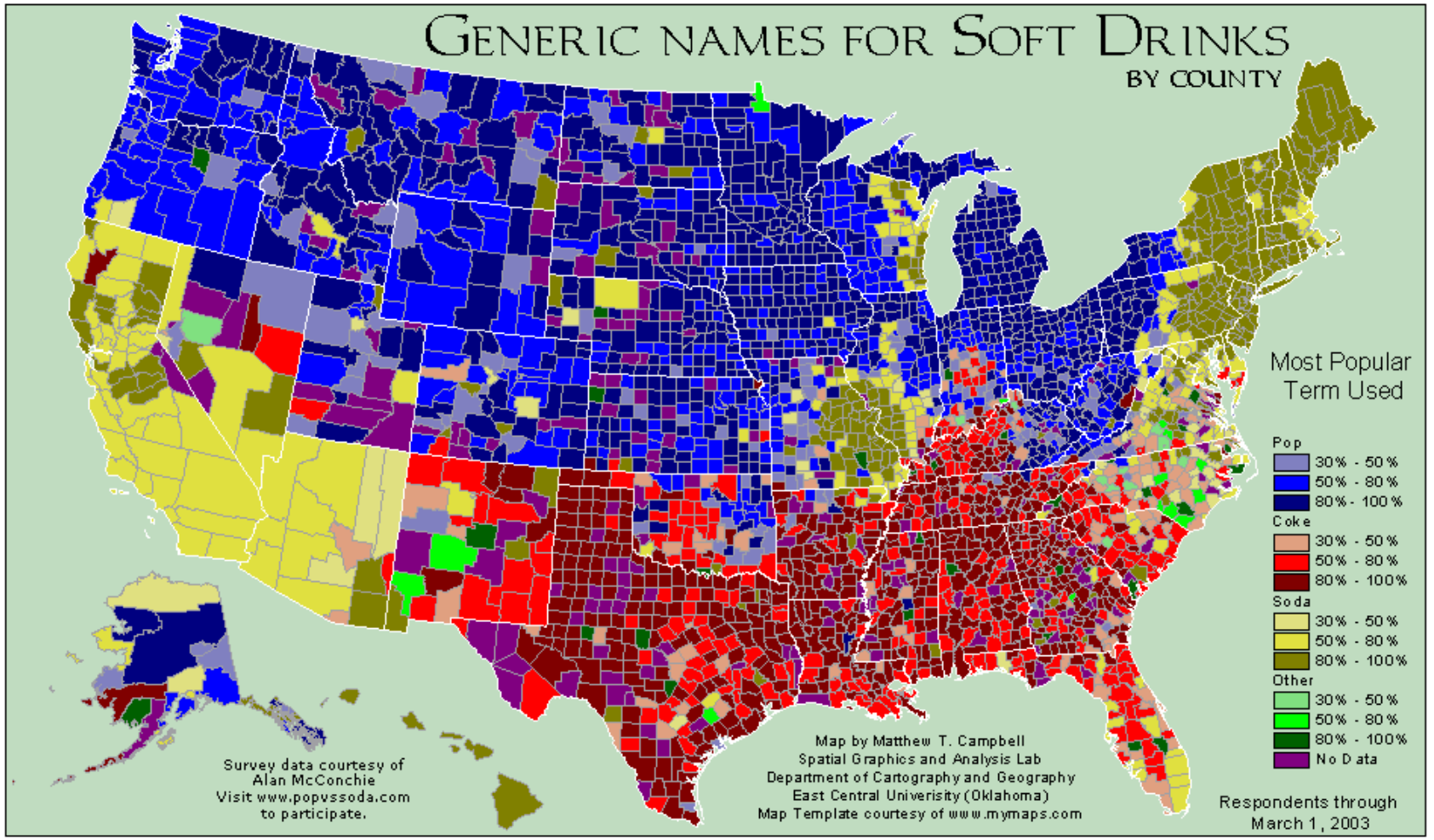
What makes a good chart?



Video: Napoleonic Wars in 8 Minutes Another video

Wikipedia: Patriotic War of 1812
http://en.wikipedia.org/wiki/File:Patriotic_War_of_1812_ENG_map1.svg

What can you learn from this map?



http://www.popvssoda.com/countystats/total-county.html

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

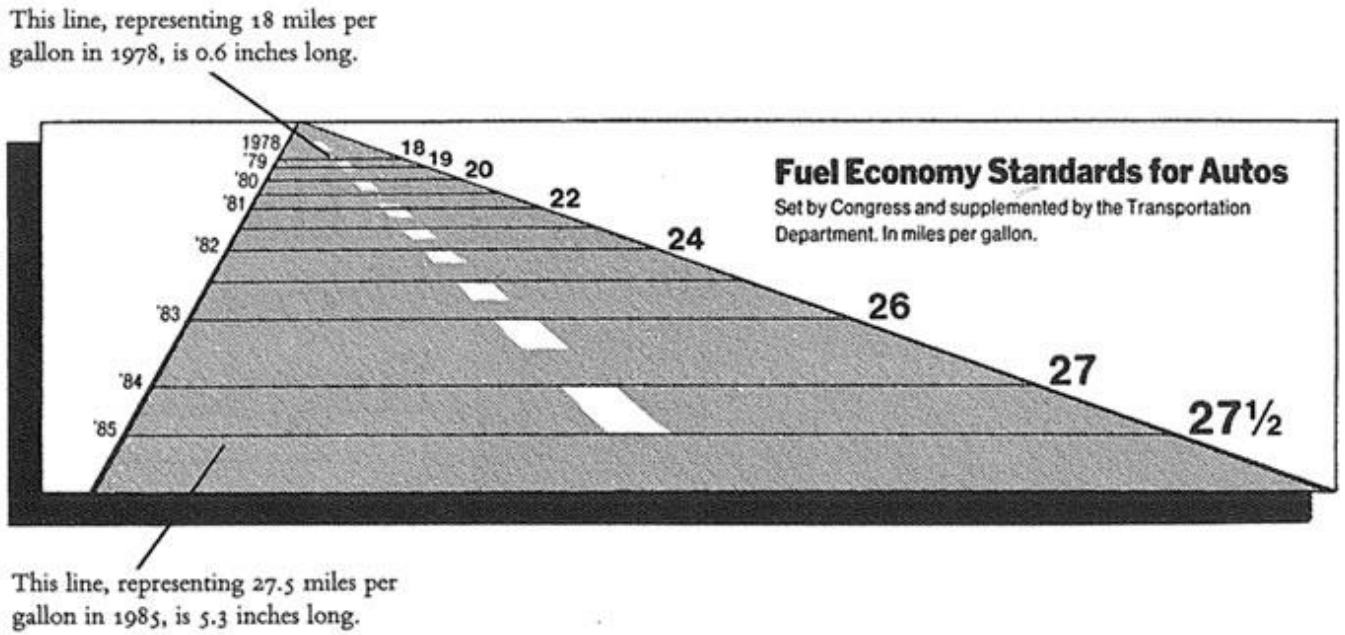
- Basically, it shouldn't "lie" (mislead the reader)
- Tufte's "Lie Factor":
  - $Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data}$

Should be ~ 1

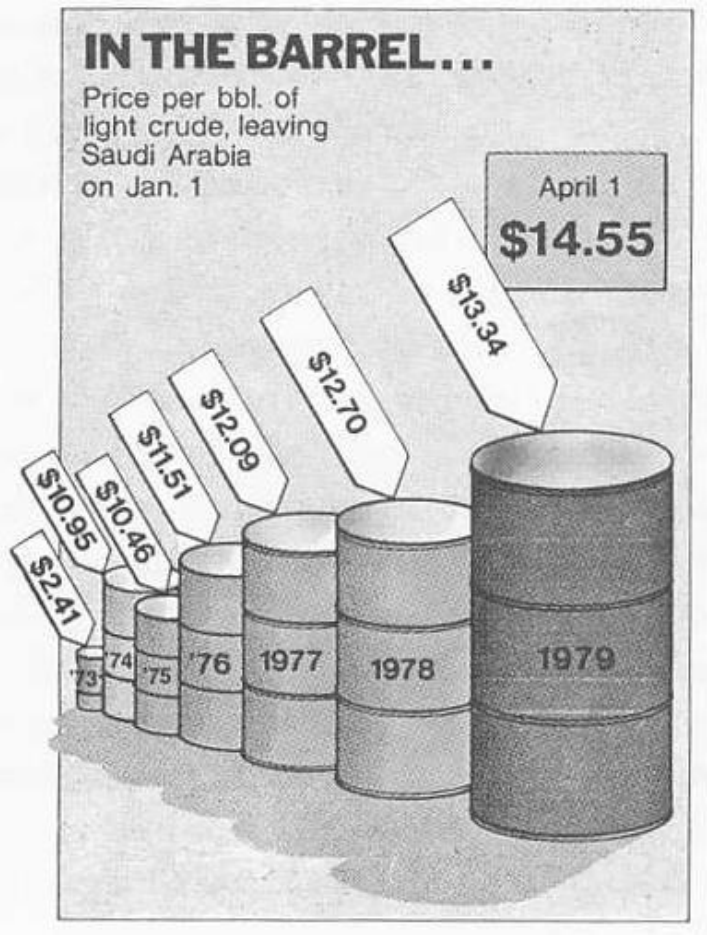
> 1 = exaggerated effect

< 1 = understated effect

Examples of the "lie factor"



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

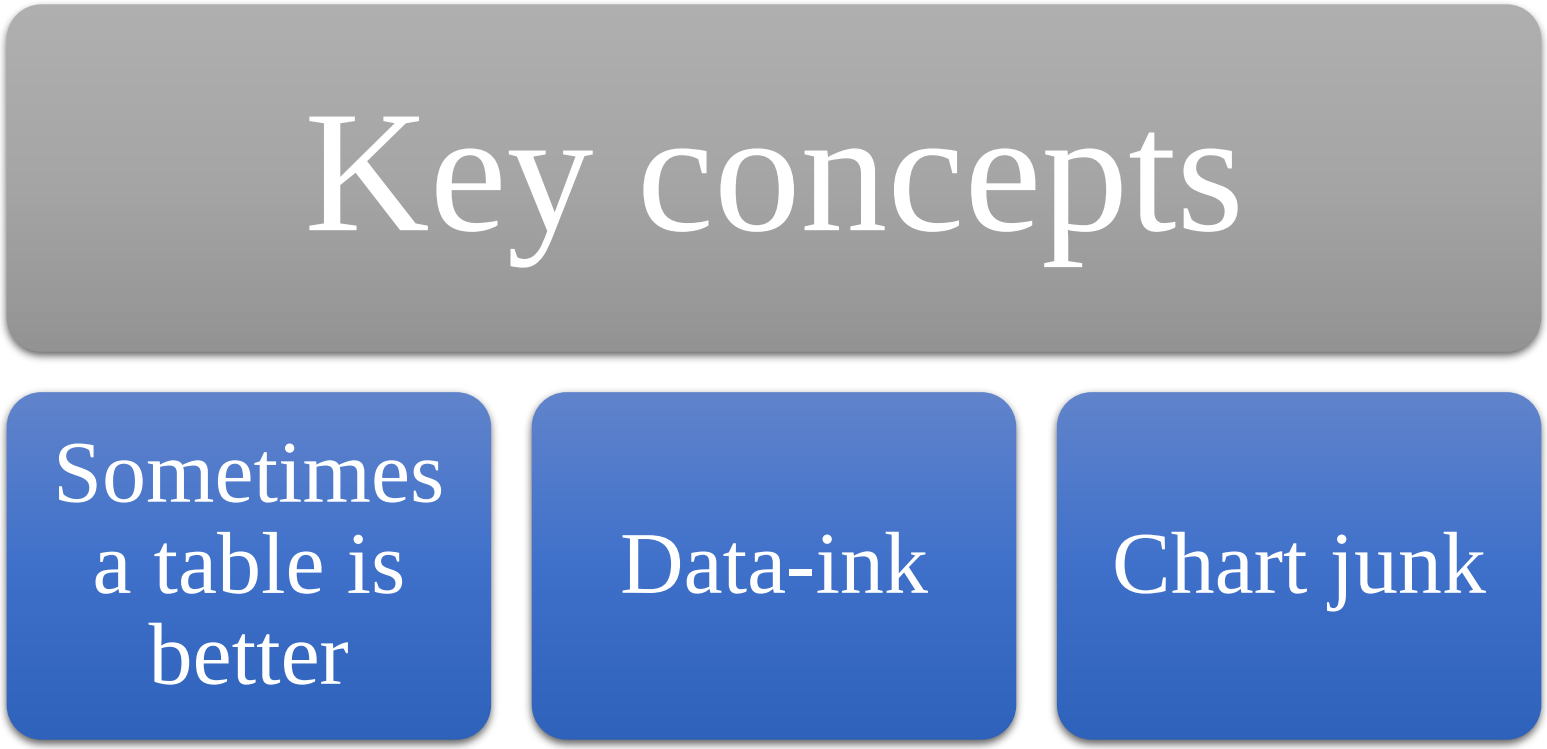


Reprinted from Tufte (2009), p. 57 & p. 62

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

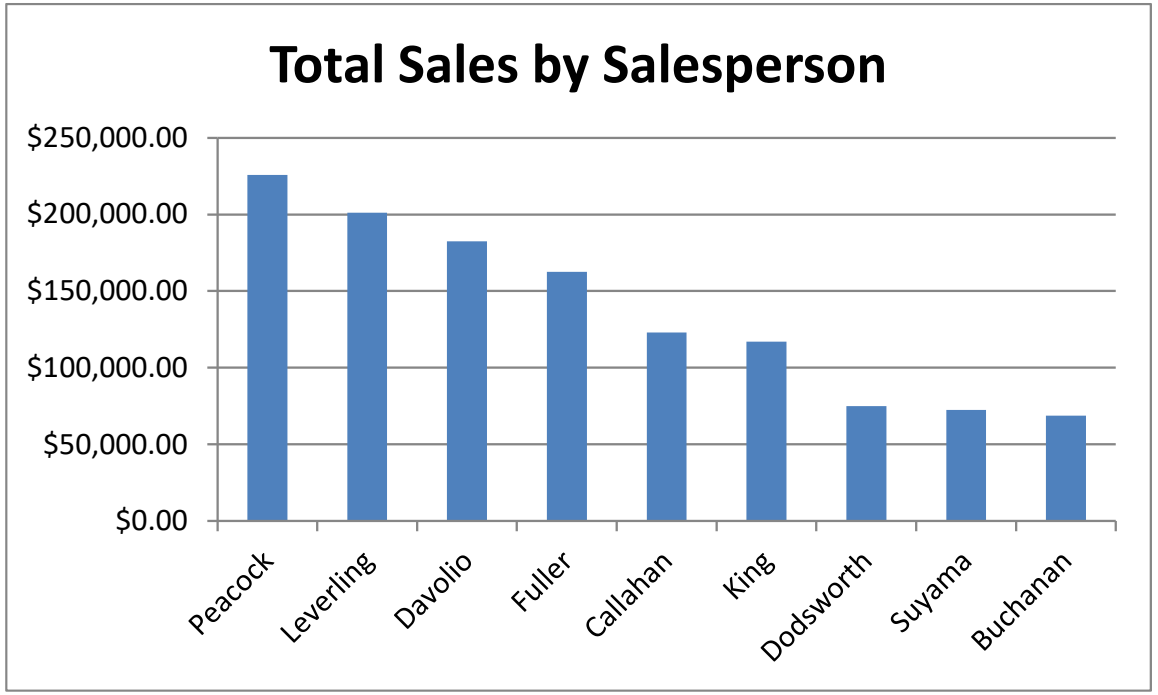
Principle 3: The chart should minimize graphical complexity

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

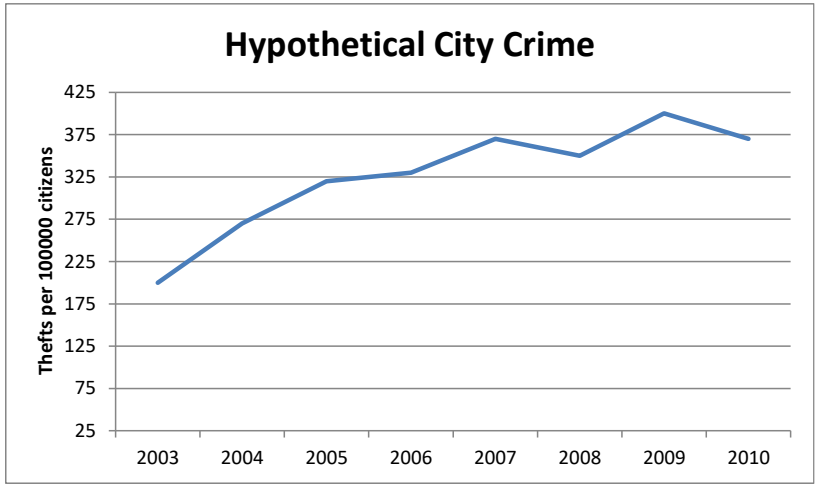
< 1 = more non-data related ink in graphic

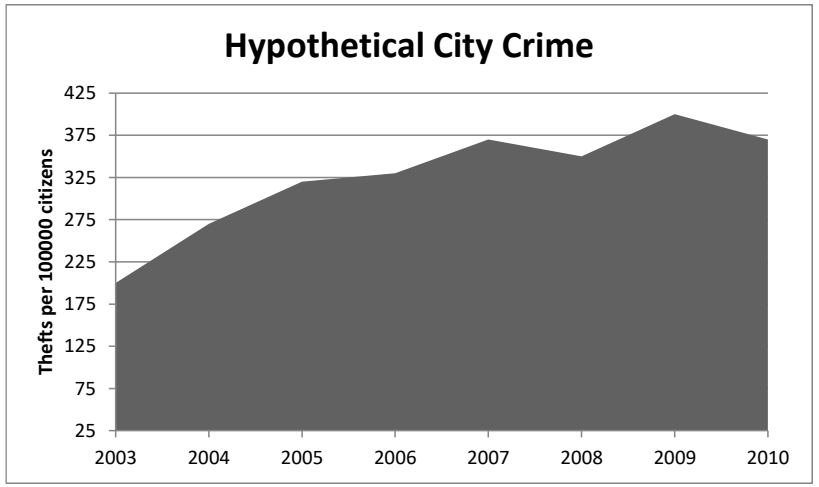
= 1 implies all ink devoted to data

Tufte's principle:
Erase ink whenever possible

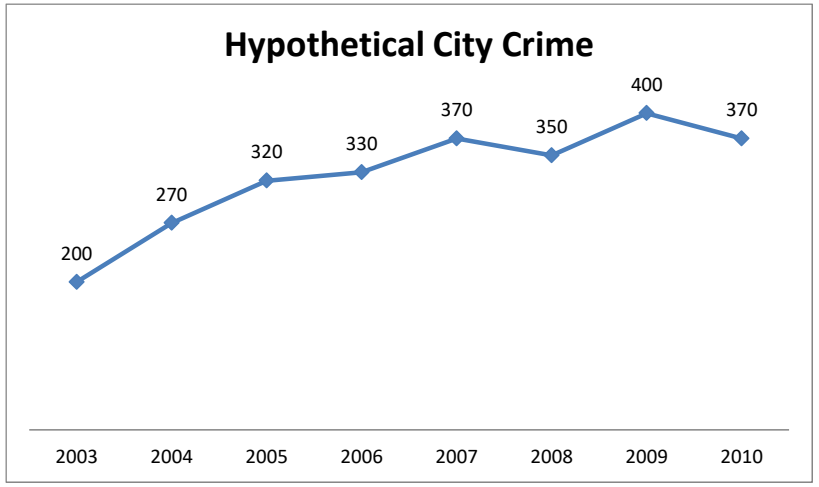
Being conscious of data ink

Lower data-ink ratio
(worse)

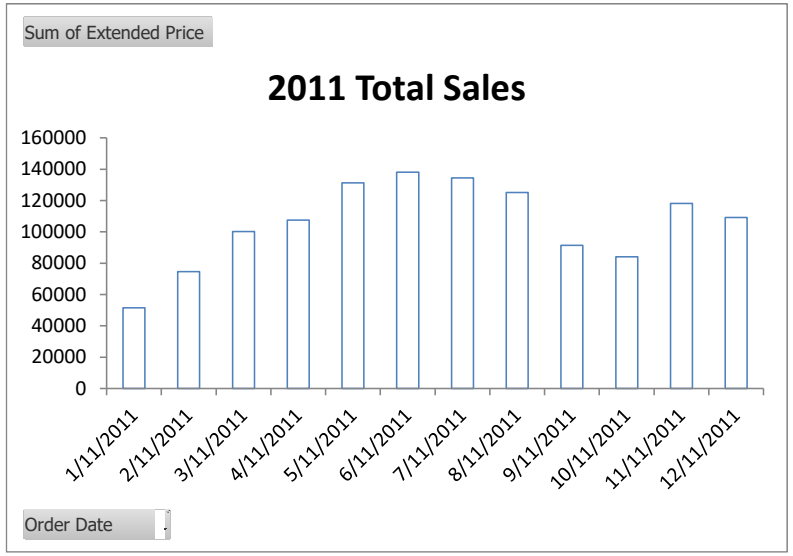


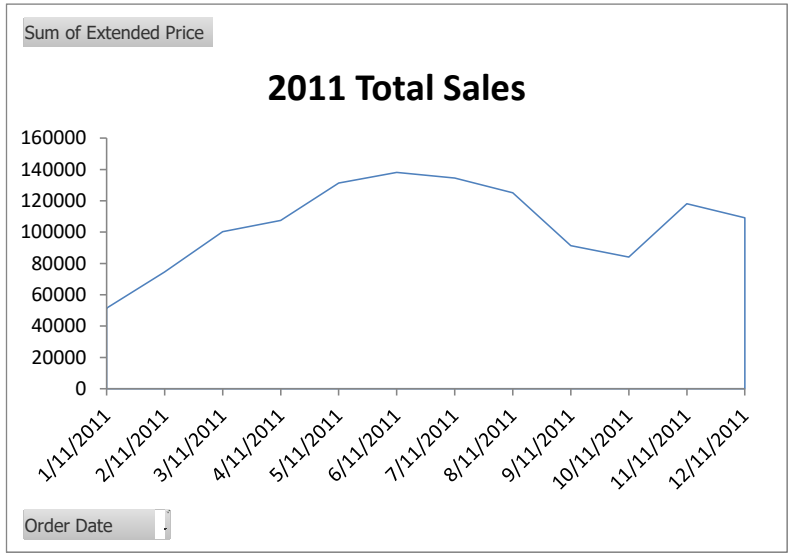


Higher data-ink ratio
(better)



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

Source: Knaflic (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals. Chapter 2.

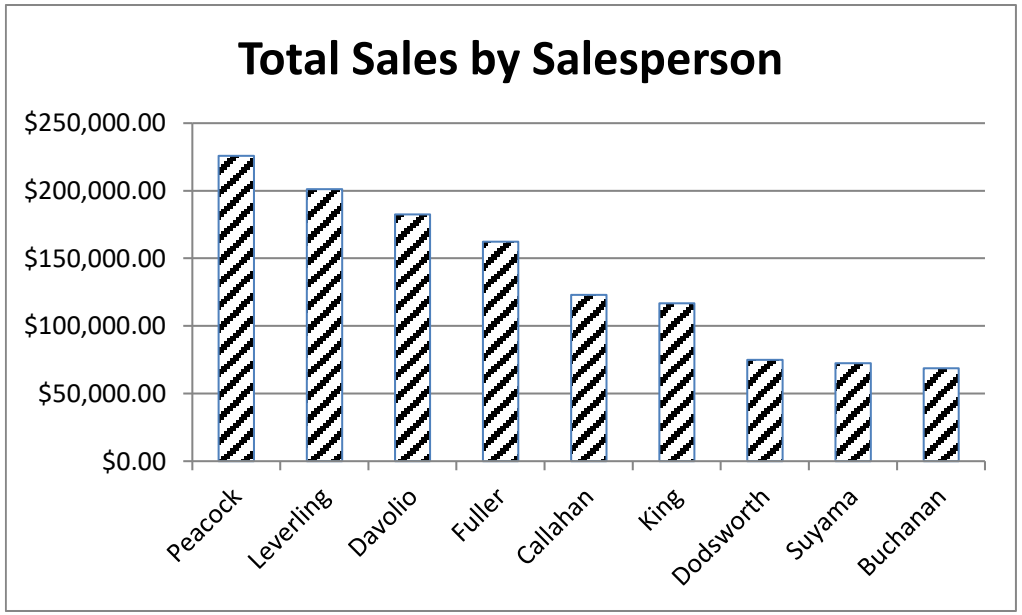
Chartjunk: Data Ink “gone wild”

Unnecessary visual clutter that doesn’t provide additional insight

Distraction from the story the chart is supposed to convey

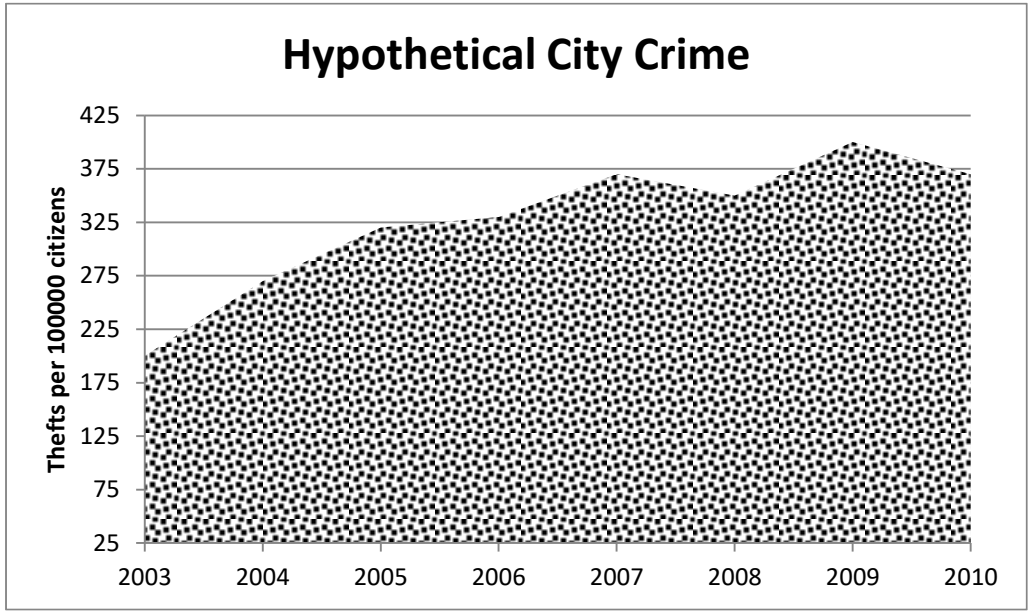
When the data-ink ratio is low, chartjunk is likely to be high

Example: Moiré effects (Tufte 2009)

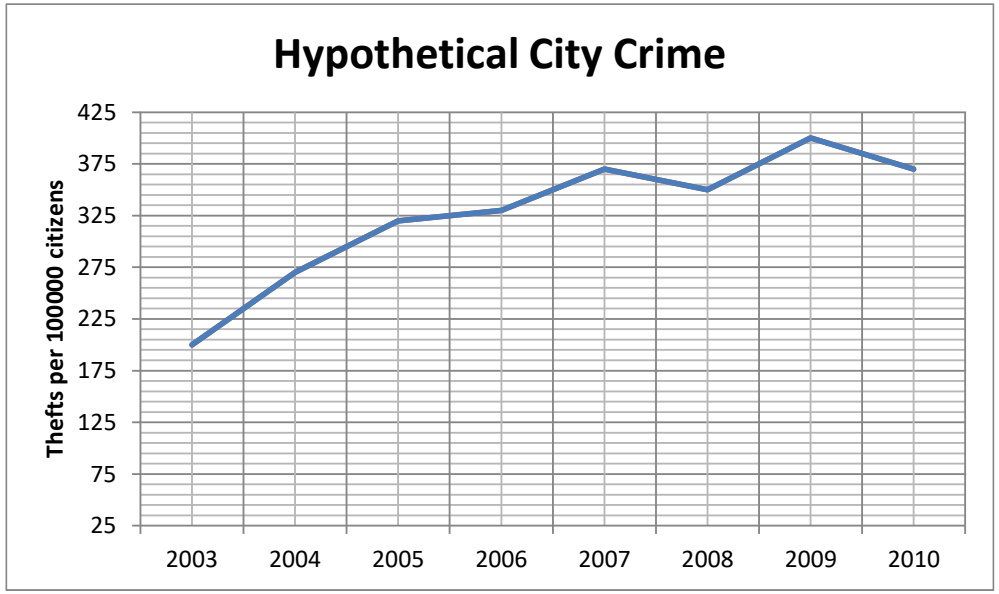


Creates illusion of movement

Stands out, in a bad way

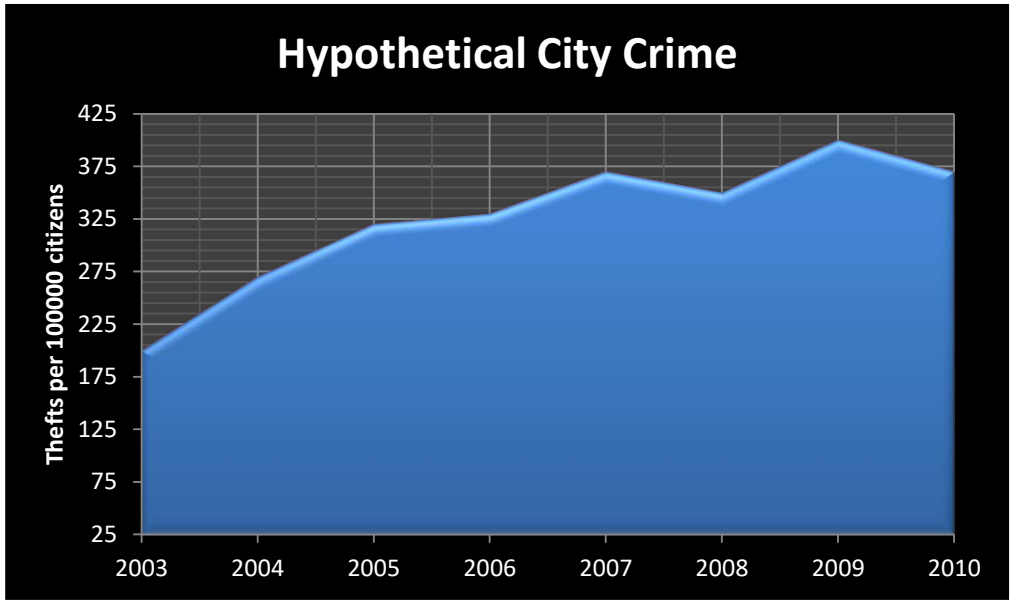


Example: The Grid

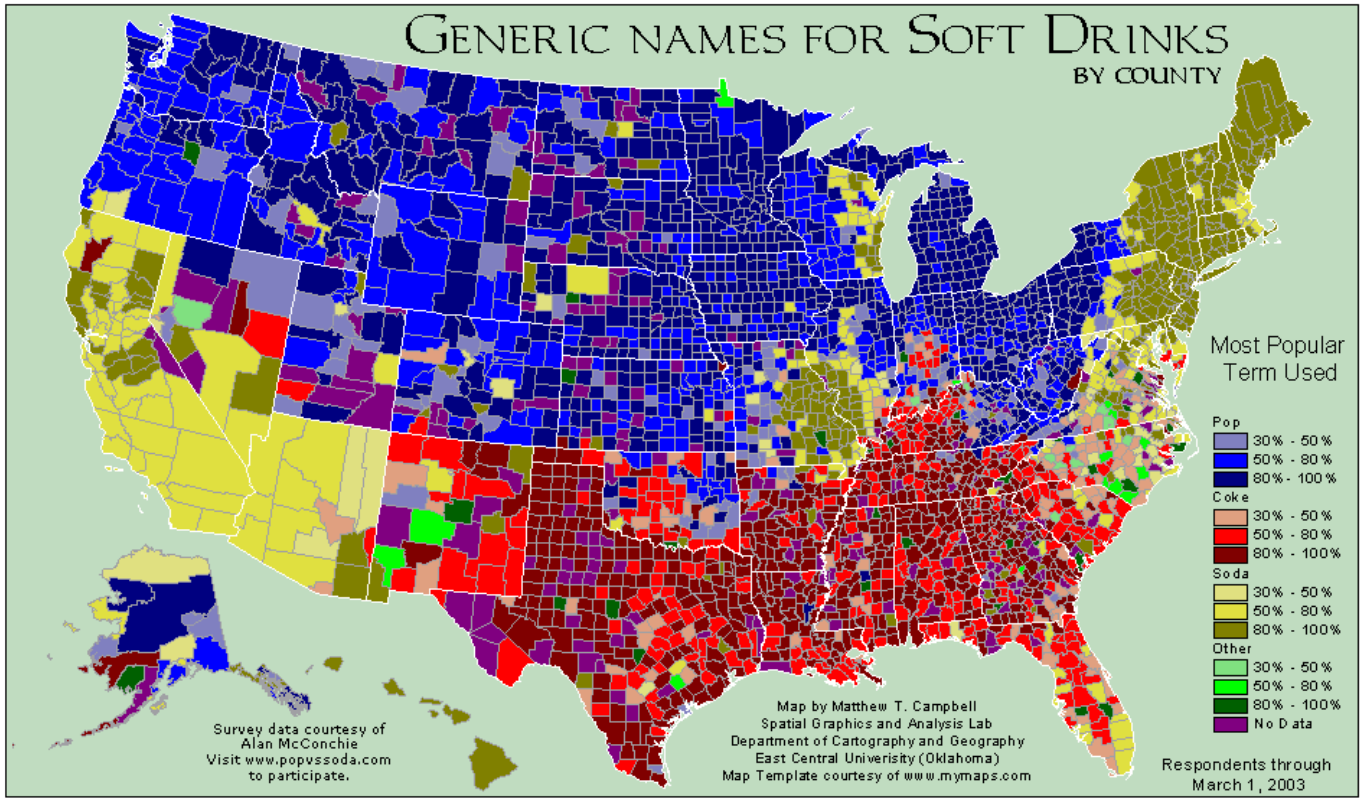


Why are these examples of chartjunk?

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!!

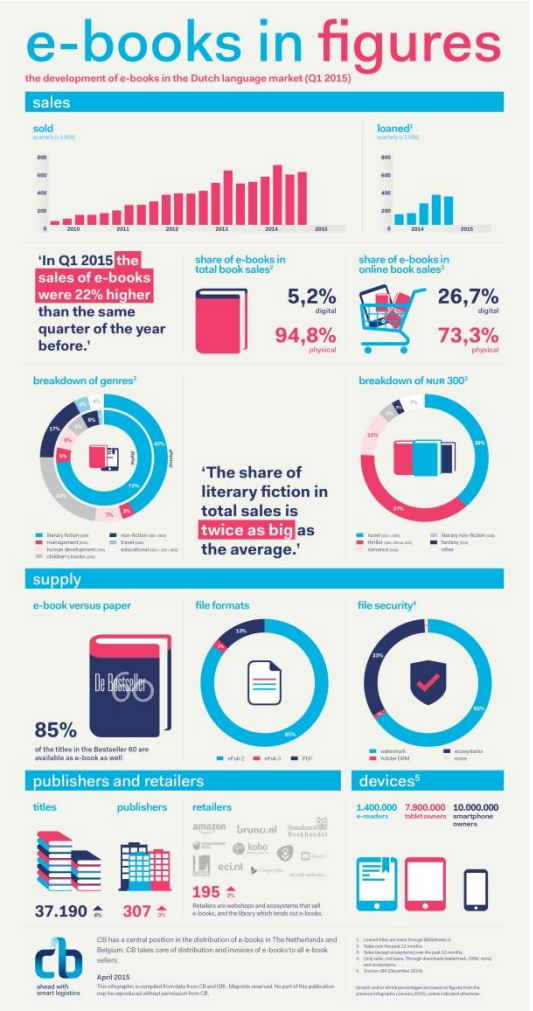
Review: Data principles (adapted from Tufte 2009)

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*.



http://the-digital-reader.com/2015/04/13/infographic-ebooks-on-track-to-double-dutch-ebook-market-in-2014/

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

Resources…

- DataMed https://datamed.org/
- Institute for Health Metrics and Evaluation's Global Health Data Exchange http://ghdx.healthdata.org/
- NNLM RD3: Resources for Data-Driven Discovery https://nnlm.gov/data/
- NNLM's YouTube Channel https://www.youtube.com/channel/UCmZqoegBFKJQF69V8d-05Bw
- OHSU's Big Data to Knowledge https://dmice.ohsu.edu/bd2k/topics.html
- Registry of Research Data Repositories (re3data.org) http://www.re3data.org/

References

- Borgman, Christine L. Big data, little data, no data: Scholarship in the networked world. MIT Press, 2015.
- Federer, Lisa. Beyond the SEA: Data Science 101: An introduction for librarians https://www.youtube.com/watch?v=i78ciP1eGxo&t=3s
- Mayer-Schönberger, Viktor, and Kenneth Cukier. Big data: A revolution that will transform how we live, work and think. Houghton Mifflin Harcourt, 2013.

Bibliography…

- A Good Example of Misleading Visualization
  - http://spatial.ly/2009/09/a-good-example-of-misleading-visualization/
- A quick guide for better data visualizations
  - https://www.tableau.com/good-to-great
- The analysis of visual variables for use in the cartographic design of point symbols for mobile Augmented Reality applications
  - Łukasz Halik, Adam Mickiewicz University Poznan
  - http://www.iag-aig.org/attach/30dee1f85f7bd479367f1f933d48b701/V61N1_2FT.pdf
- The Benefits and Future of Data Visualization
  - StatSilk founder Frank van Cappelle
  - https://www.statsilk.com/blog/benefits-and-future-data-visualization
- Charting Statistics
  - Mary Eleanor Spear
  - https://archive.org/details/ChartingStatistics
- Color Brewer
  - http://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3

Bibliography…

- Data: The language of modern business leaders
  - Steve Proctor, March 17, 2017
  - http://www.itbusiness.ca/sponsored/data-the-language-of-modern-business-leaders
- Data Visualization: The Best Infographic Tools For 2017
  - Bernard Marr, October 10, 2017
  - https://www.huffingtonpost.com/entry/data-visualization-the-best-infographic-tools-for_us_59ca128fe4b0f2df5e83b134
- Data Visualization: The future of data visualization
  - Will Towler, January/February 2015
  - http://analytics-magazine.org/data-visualization-the-future-of-data-visualization
- Data Visualization 101: How to Choose the Right Chart or Graph for Your Data
  - Jami Oetting
  - https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization
- Data-Driven Design: Dare to Wield the Sword of Data – Part I
  - Brent Dykes, December 4, 2012
  - http://www.analyticshero.com/2012/12/04/data-driven-design-dare-to-wield-the-sword-of-data-part-i/
- Datavis.ca
  - http://www.datavis.ca/index.php

Bibliography

- Diverging color schemes: Showing good data isn't enough; you need to show it well
  - Alberto Cairo, June 26, 2016
  - http://www.thefunctionalart.com/2016/06/diverging-color-schemes-showing-good_26.html
- 8 Horrible Data Visualizations That Make No Sense
  - Eric Limer, September 02, 2013
  - http://gizmodo.com/8-horrible-data-visualizations-that-make-no-sense-1228022038
- 55 Striking Data Visualization and Infographic Poster Designs
  - Igor Ovsyannykov, May 16, 2011
  - http://inspirationfeed.com/inspiration/infographics/55-striking-data-visualization-and-infographic-poster-designs/
- 4 Tips for Promoting Predictive Analytics in Your Organization
  - Fern Halper, September 26, 2017
  - https://tdwi.org/articles/2017/09/26/ADV-ALL-4-Tips-for-Promoting-Predictive-Analytics.aspx


DIAMOND JUBILEE
60
YEARS
IIT DELHI


All India Council for Technical Education
AICTE
योगः कर्मसु कौशलम्

THANK YOU


FREE ONLINE EDUCATION
swayam
शिक्षित भारत, उन्नत भारत