

# Information Visualization

## Part 1

Deepak Kumar

## BIG DATA

- Data intensive computing
  - capture
  - curation
  - storage
  - search
  - sharing
  - analysis
  - visualization



# Data Science



**@chl**  
Christian Langreiter

data scientist: statistician who lives in select regions of california or works at bitly

3 hours ago via web ☆ Favorite ↻ Retweet ↩ Reply

11/13/2019

3

# Data Science



11/13/2019

4

# Data Science

## Data science

---

From Wikipedia, the free encyclopedia

**Data science** defines a discipline that incorporates applying varying degrees of [statistics](#), [data visualizations](#), [computer programming](#), [data mining](#), [machine learning](#), and [database engineering](#) to solve complex [data](#) problems. A practitioner of Data Science is called a Data Scientist.

11/13/2019

5

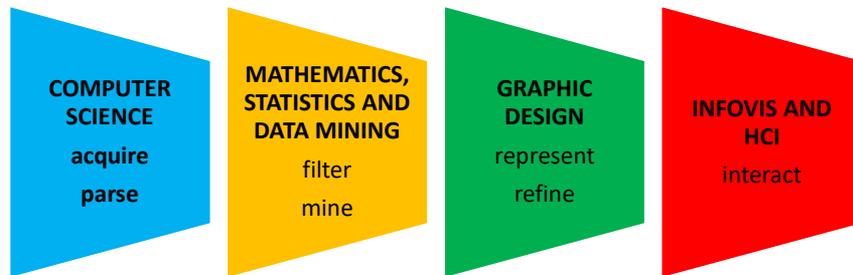
## What is *Data Science*?

- **The process of using data in the *wild*** unstructured, unformatted, multiple sources,...
- **Involves**
  - Acquiring (finding and storing)
  - Analyzing
  - Discovering Patterns/Stories
  - Presenting results

11/13/2019

6

# Data Science



11/13/2019

7

## What is *Data Science*?

- **The process of using data in the *wild*** unstructured, unformatted, multiple sources,...
- **Involves**
  - Acquiring (finding and storing)
  - Analyzing
  - Discovering Patterns/Stories
  - **Presenting results**

11/13/2019

8

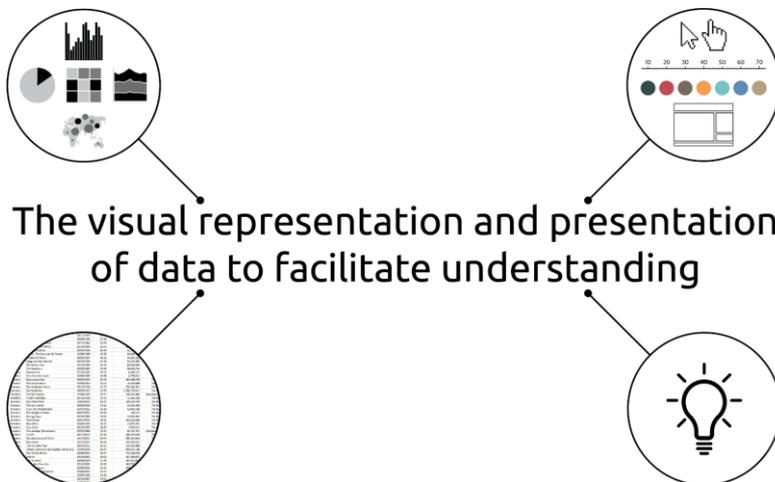
## Visualizing Data



11/13/2019

9

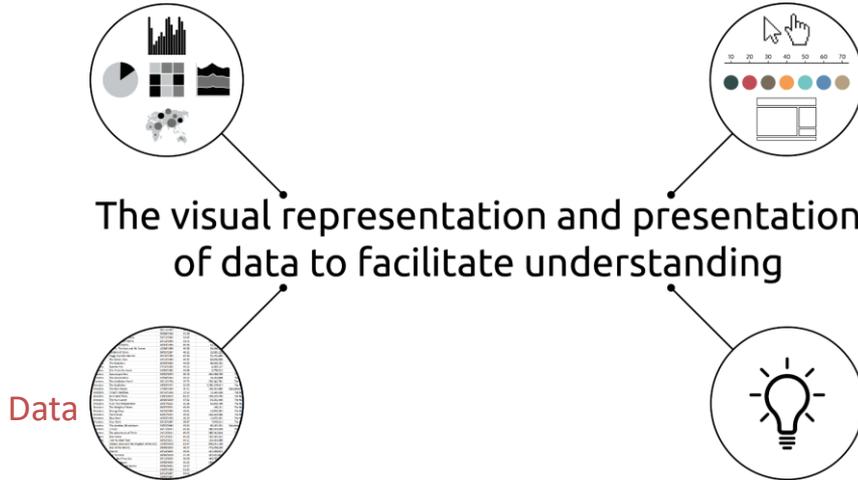
## What is Data Visualization?



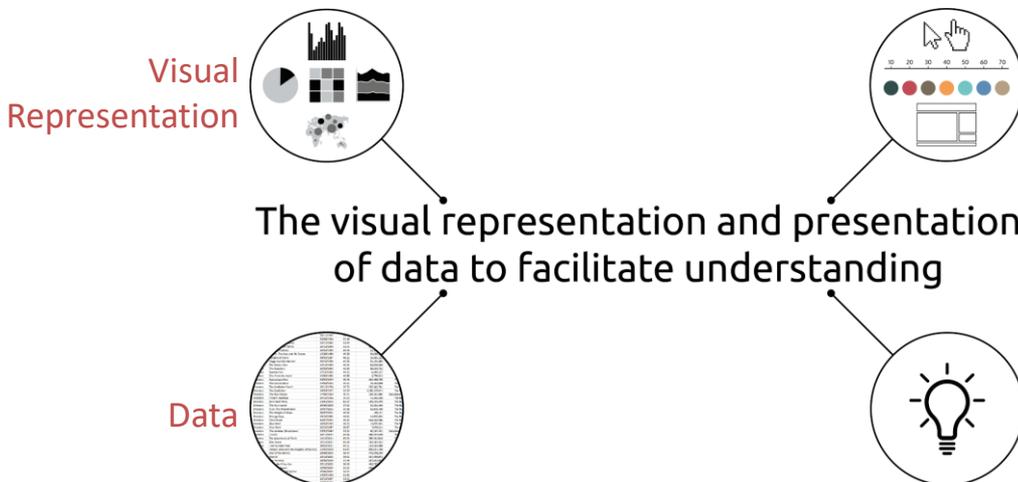
11/13/2019

10

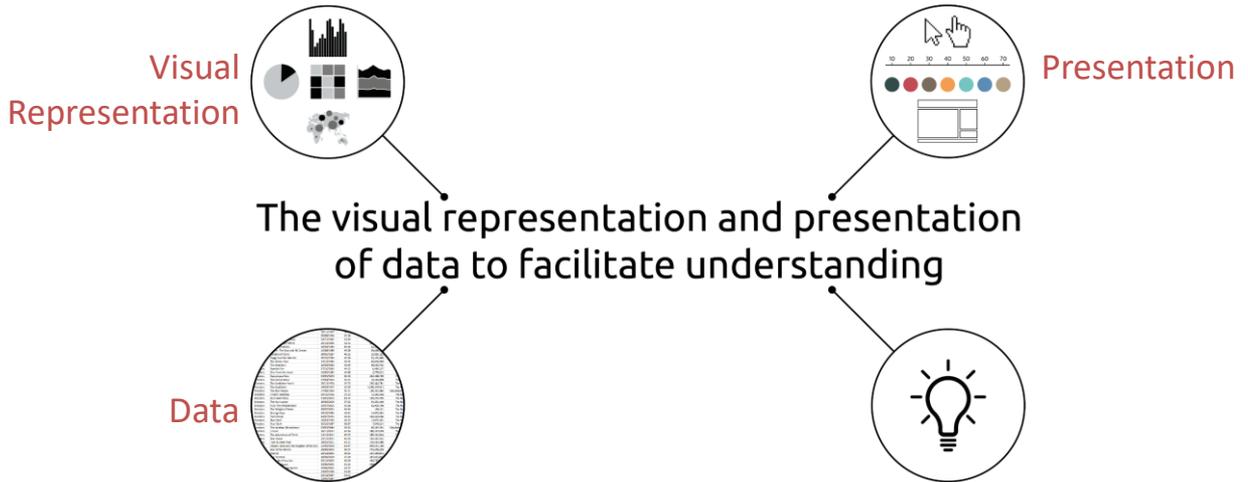
# What is Data Visualization?



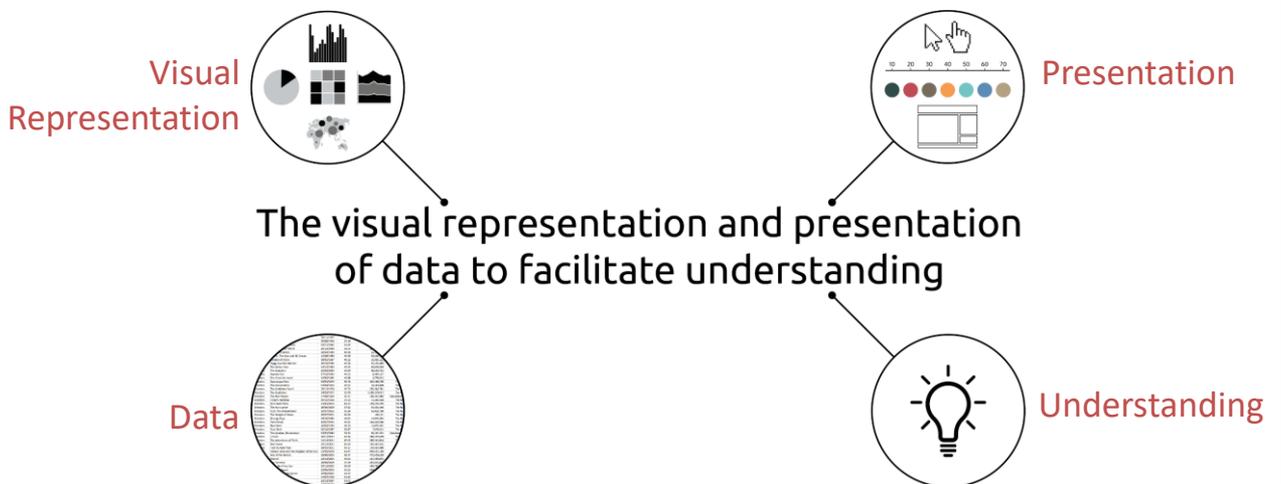
# What is Data Visualization?



# What is Data Visualization?

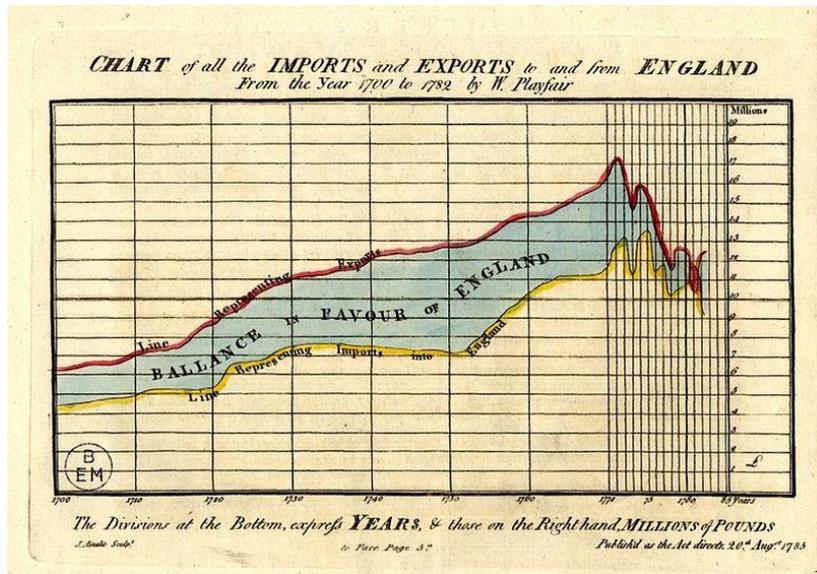


# What is Data Visualization?



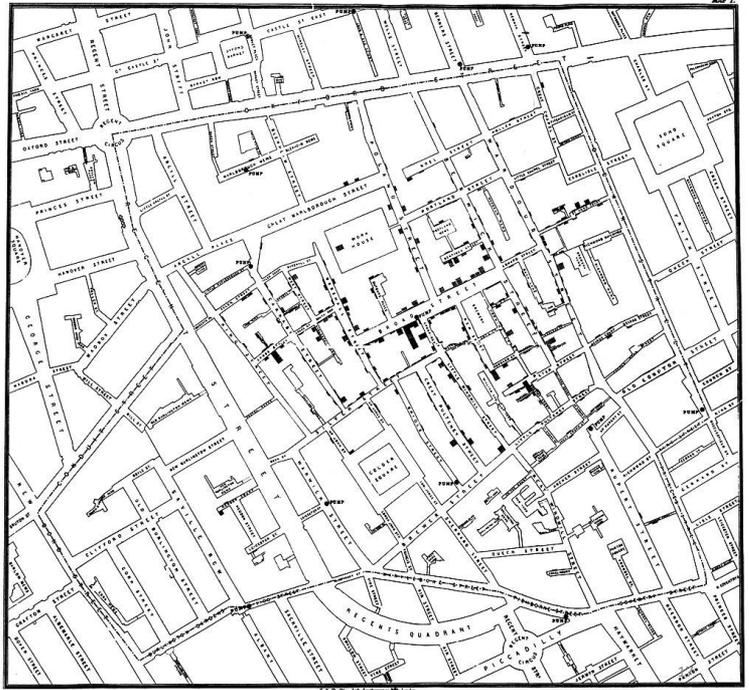
# A Little History...

## William Playfair, 1786



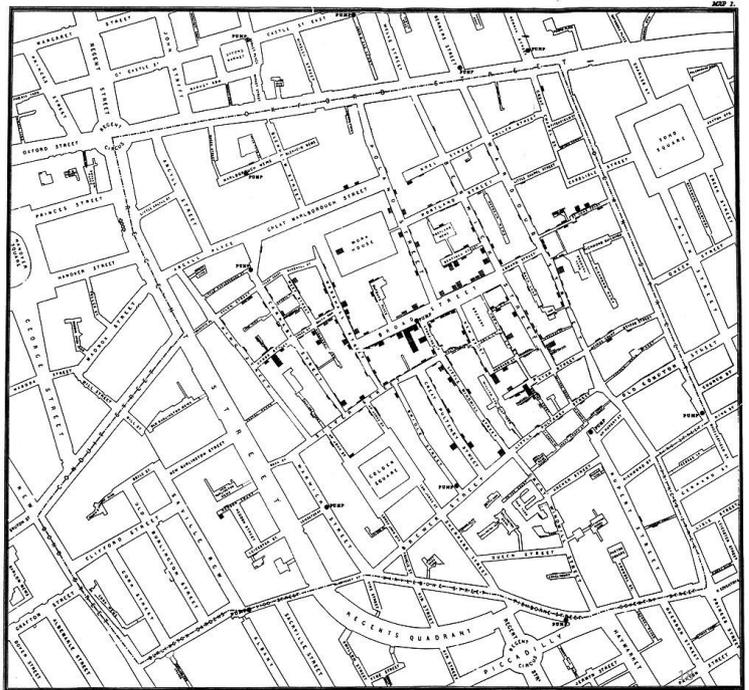
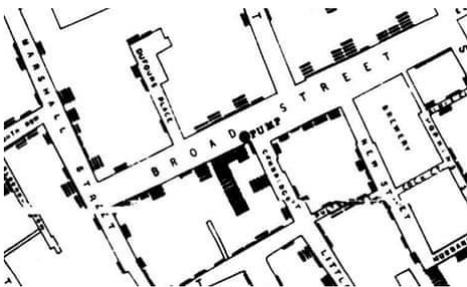
# John Snow, 1854

11/13/2019



# John Snow, 1854

11/13/2019

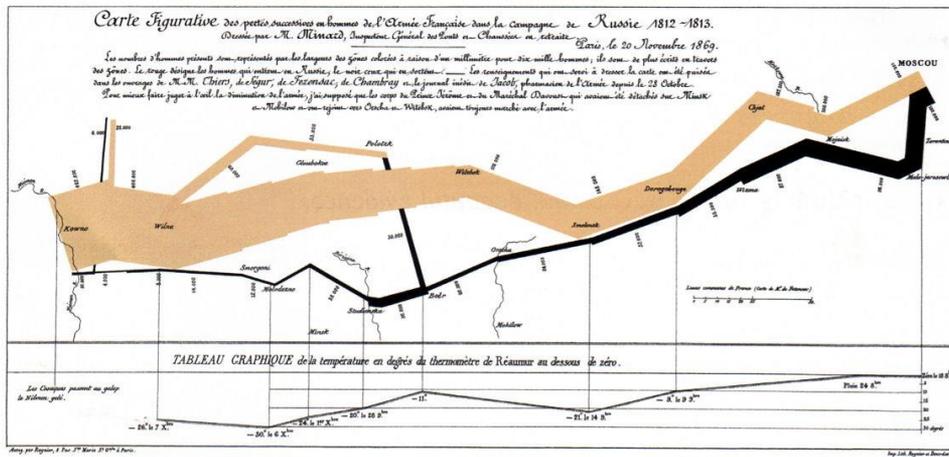


# John Snow Pub, London



11/13/2019

# Charles Joseph Minard, 1869

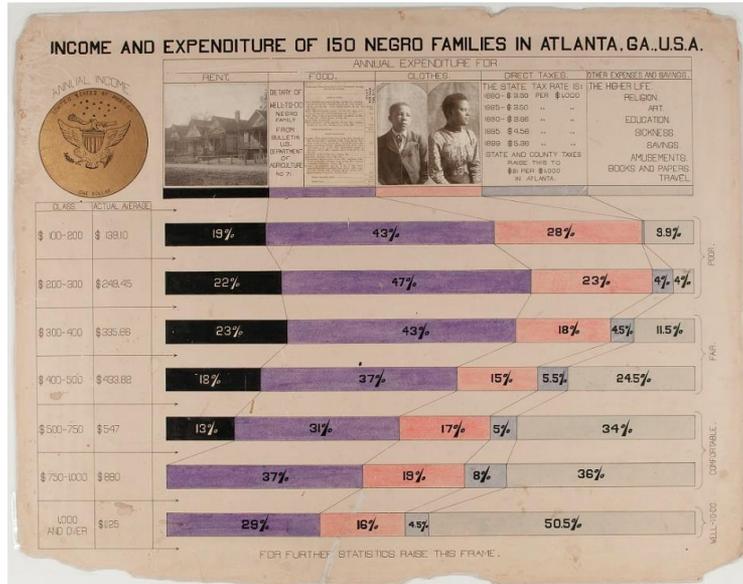


This map drawn by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen, the thick band shows the size of the army (422,000 men) as it invaded Russia. The width of the band indicates the size of the army at each position. In September, the army reached Moscow with 100,000 men. The path of Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. The remains of the Grande Armée struggled out of Russia with 10,000 men. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

11/13/2019

20

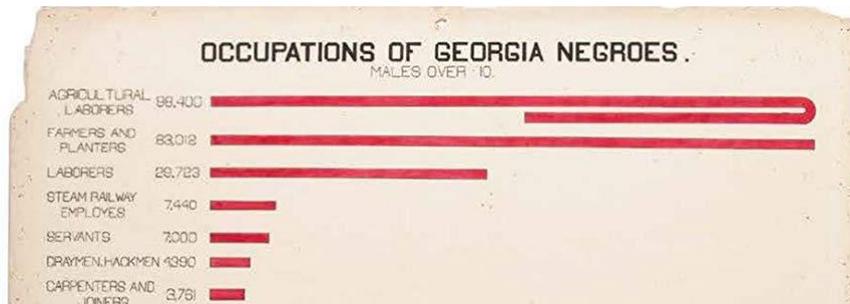
# W.E.B. Du Bois, 1899



11/13/2019

21

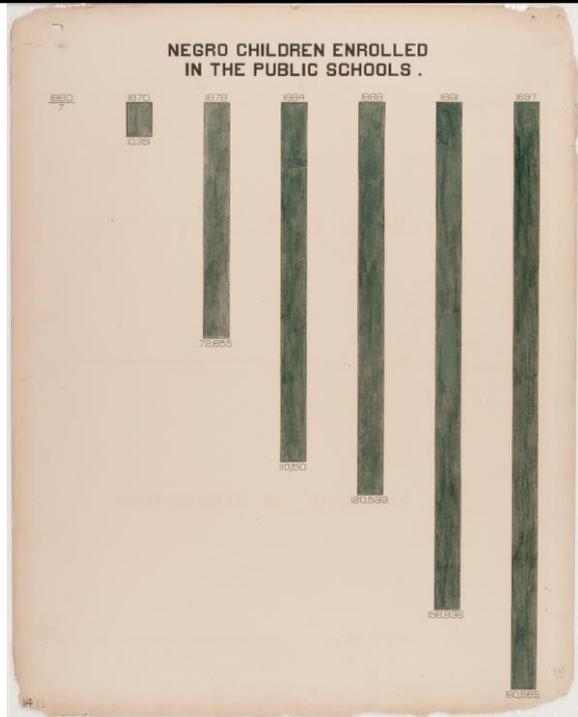
# W.E.B. Du Bois, 1899



11/13/2019

22

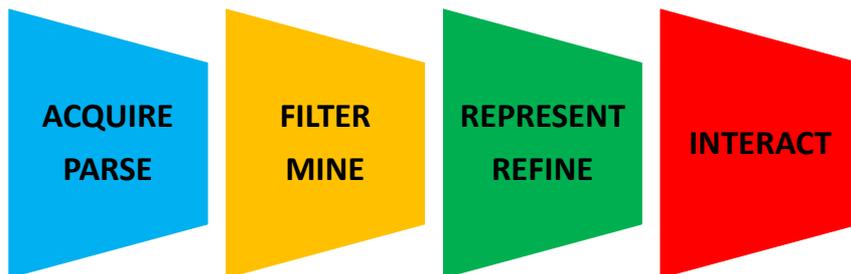
W.E.B. Du Bois, 1899



11/13/2019

23

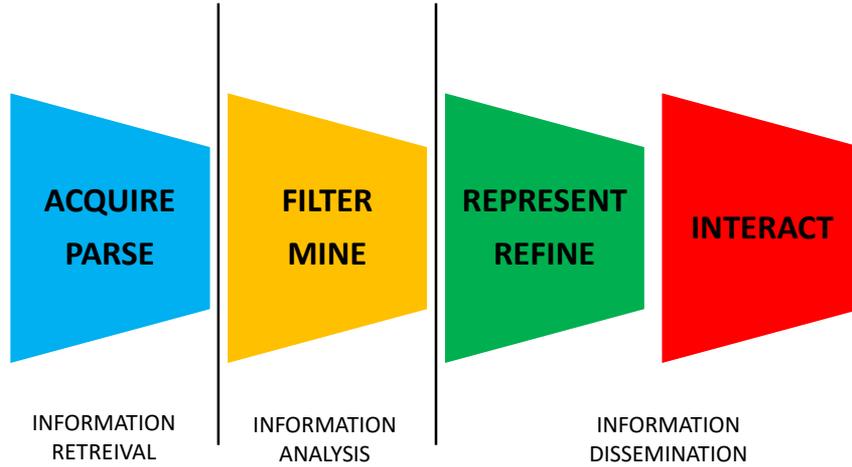
## The Visualization Process



11/13/2019

24

# The Visualization Process



11/13/2019

25

# Example: Data

Proportion of Sales % by Channel Over Time

REPORTING MONTH	STORES	ONLINE	TELEPHONE
May 2011	71	29	0
Jun 2011	72	28	0
Jul 2011	71	28	1
Dec 2011	71	28	1
Jun 2012	73	26	1
Jul 2012	77	22	1
Sep 2012	75	24	1
Nov 2012	75	24	1
Jun 2013	73	26	1
Nov 2013	73	26	1
Jan 2014	73	26	1
Jun 2014	72	27	1
Aug 2014	55	44	1
Sep 2014	60	38	2
Oct 2014	51	48	1
Nov 2014	44	55	1
Jan 2015	52	47	1
Mar 2015	50	48	2
Jun 2015	49	49	2
Jul 2015	37	61	2
Aug 2015	40	58	2
Nov 2015	40	59	1
Dec 2015	22	77	1
Jan 2016	21	77	2
Feb 2016	20	78	2
Apr 2016	14	84	2
Dec 2016	21	77	2
Jun 2018	6	93	1
Jul 2018	6	93	1
Dec 2018	0	100	0

11/13/2019

26

## Example: Data

Proportion of Sales % by Channel Over Time

REPORTING MONTH	STORES	ONLINE	TELEPHONE
May 2011	71	29	0
Jun 2011	72	28	0
Jul 2011	71	28	1
Dec 2011	71	28	1
Jun 2012	73	26	1
Jul 2012	77	22	1
Sep 2012	75	24	1
Nov 2012	75	24	1
Jun 2013	73	26	1
Nov 2013	73	26	1
Jan 2014	73	26	1
Jun 2014	72	27	1
Aug 2014	55	44	1
Sep 2014	60	38	2
Oct 2014	51	48	1
Nov 2014	44	55	1
Jan 2015	52	47	1
Mar 2015	50	48	2
Jun 2015	49	49	2
Jul 2015	37	61	2
Aug 2015	40	58	2
Nov 2015	40	59	1
Dec 2015	22	77	1
Jan 2016	21	77	2
Feb 2016	20	78	2
Apr 2016	14	84	2
Dec 2016	21	77	2
Jun 2018	6	93	1
Jul 2018	6	93	1
Dec 2018	0	100	0

Can be used to answer simple questions:

- What was the % of online sales in April 2016?
- What was the % of store sales in June 2012?
- In January 2015, which channel had the second largest % sales?

What about broader questions?

- Can you identify the trends about channels over time?
- When did % sales from online overtake the store sales?
- During which periods did the channels most accelerated upward or downwards changes?
- What if this table was thousands of rows and dozens of columns?

11/13/2019

27

## Example: Data

Proportion of Sales % by Channel Over Time

REPORTING MONTH	STORES	ONLINE	TELEPHONE
May 2011	71	29	0
Jun 2011	72	28	0
Jul 2011	71	28	1
Dec 2011	71	28	1
Jun 2012	73	26	1
Jul 2012	77	22	1
Sep 2012	75	24	1
Nov 2012	75	24	1
Jun 2013	73	26	1
Nov 2013	73	26	1
Jan 2014	73	26	1
Jun 2014	72	27	1
Aug 2014	55	44	1
Sep 2014	60	38	2
Oct 2014	51	48	1
Nov 2014	44	55	1
Jan 2015	52	47	1
Mar 2015	50	48	2
Jun 2015	49	49	2
Jul 2015	37	61	2
Aug 2015	40	58	2
Nov 2015	40	59	1
Dec 2015	22	77	1
Jan 2016	21	77	2
Feb 2016	20	78	2
Apr 2016	14	84	2
Dec 2016	21	77	2
Jun 2018	6	93	1
Jul 2018	6	93	1
Dec 2018	0	100	0

Can be used to answer simple questions:

- What was the % of online sales in April 2016?
- What was the % of store sales in June 2012?
- In January 2015, which channel had the second largest % sales?

What about broader questions?

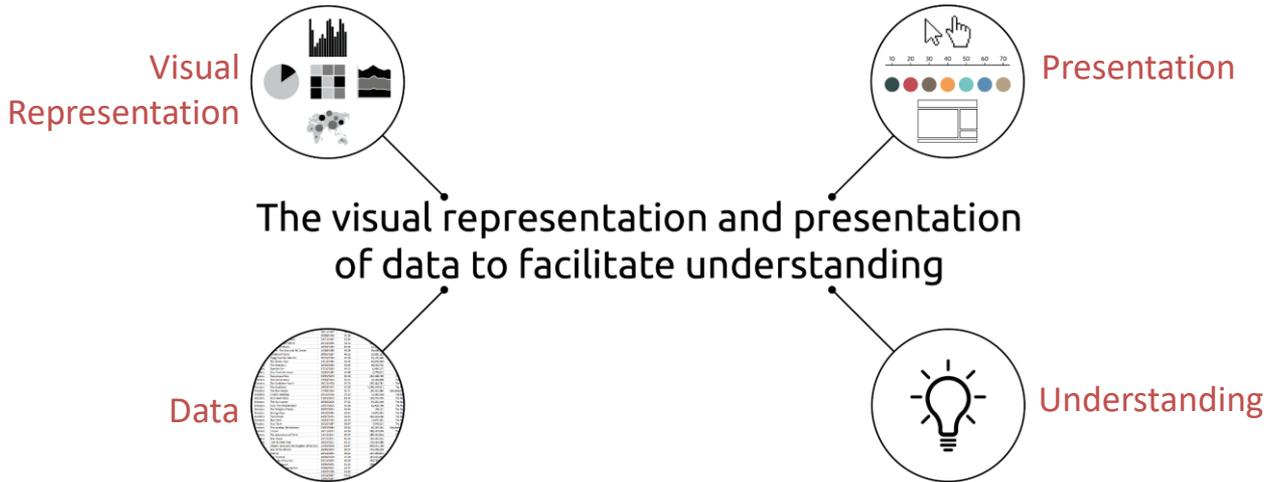
- Can you identify the trends about channels over time?
- When did % sales from online overtake the store sales?
- During which periods did the channels most accelerated upward or downwards changes?
- What if this table was thousands of rows and dozens of columns?

We can *LOOK* at data but we cannot really *SEE* it.

11/13/2019

28

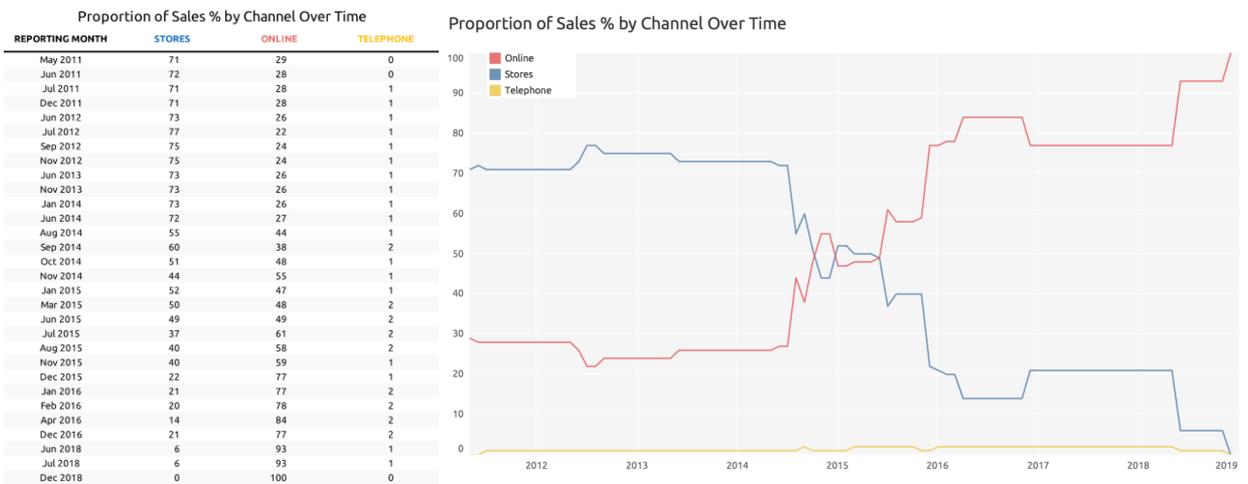
# What is Data Visualization?



11/13/2019

29

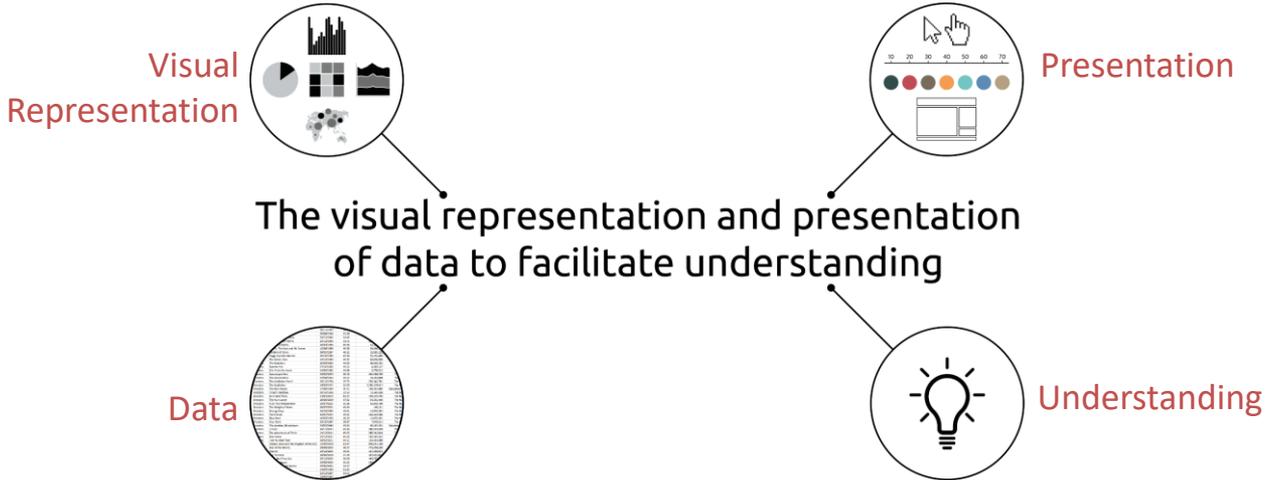
## Example: Data+Visual Representation



11/13/2019

30

# What is Data Visualization?



11/13/2019

31

## Example: Presentation

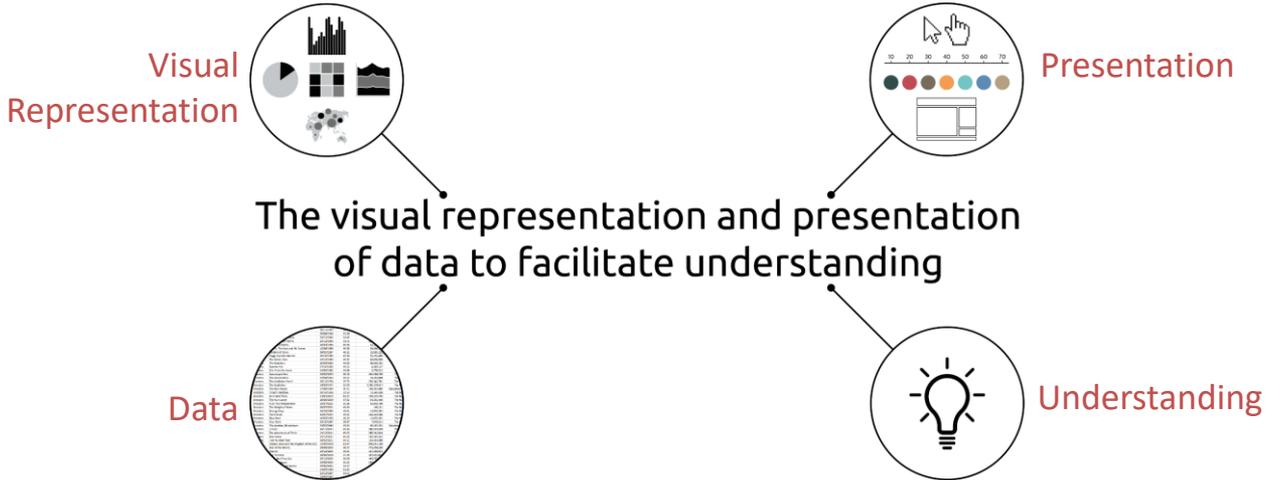
Proportion of Sales % by Channel Over Time



11/13/2019

32

# What is Data Visualization?



11/13/2019

33

# Example: Understanding

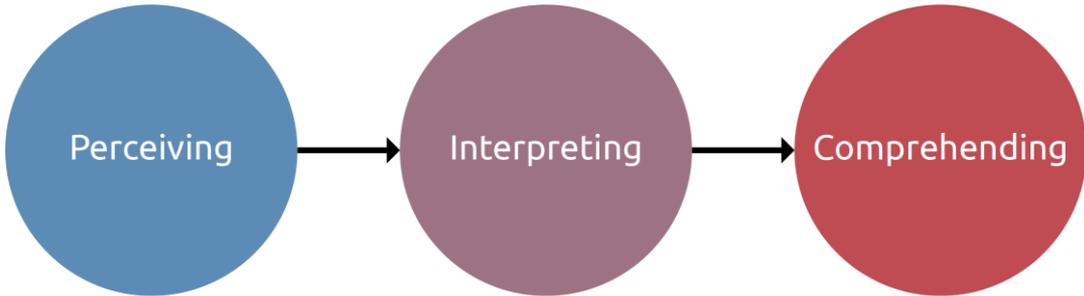
Proportion of Sales % by Channel Over Time



11/13/2019

34

## Understanding: Three Phases



**What do I see?**  
*What data is shown?*  
*How is the data represented?*  
*What features are observable?*

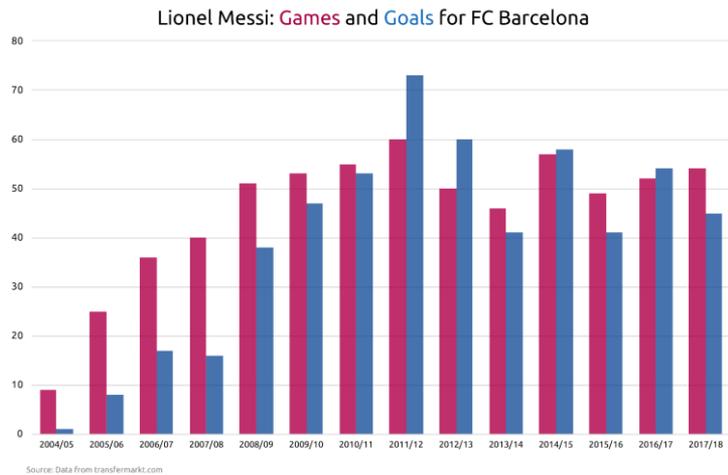
**What does it mean, given the subject?**  
*What features are interesting?*  
*What features are unexpected?*  
*What features are important?*

**What does it mean to me?**  
*What have I learnt?*  
*What do I feel?*  
*What do I now do?*

VISUALISER CONTROL

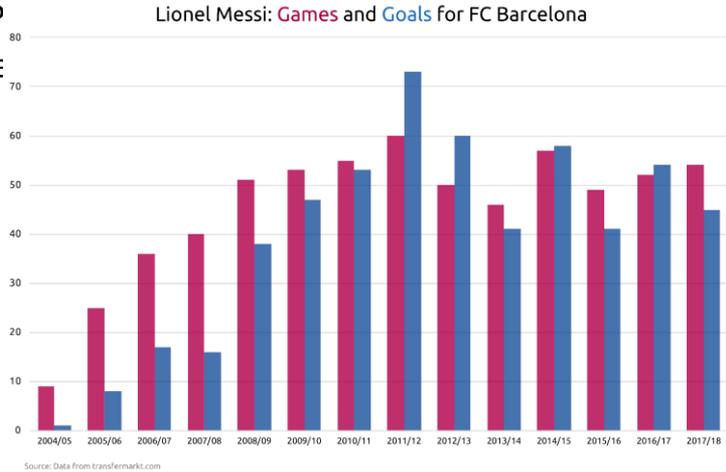
VIEWER CONTROL

## Example: Understanding



## Example: Perceiving

- What chart is being used?
- What items of data do the marks represent?
- What value associations do the attributes represent?
- What range of values are represented?
- Are the data and its representation trustworthy?

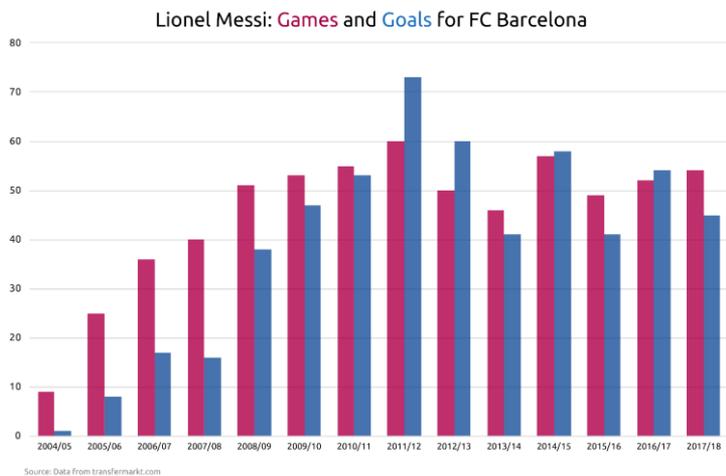


11/13/2019

37

## Example: Perceiving

What do I see?

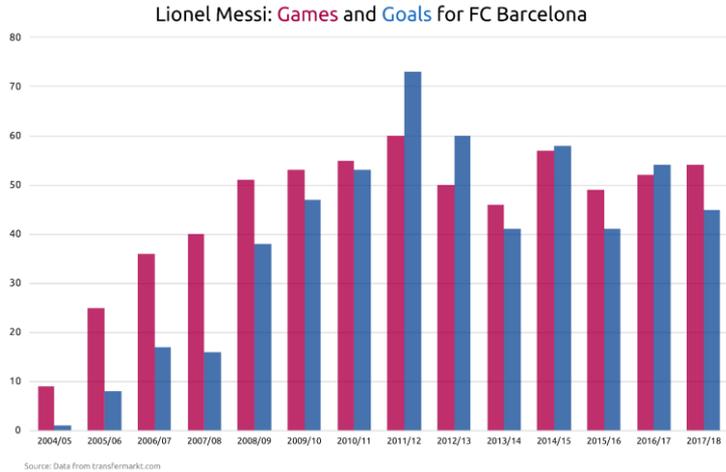


11/13/2019

38

## Example: Interpreting

What does it mean, given the subject?

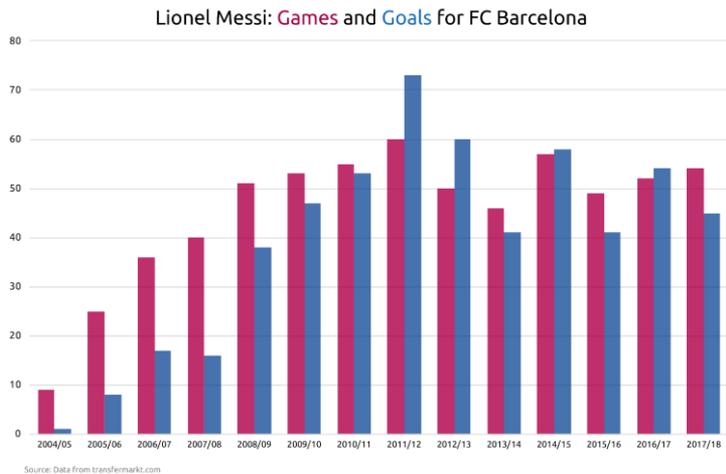


11/13/2019

39

## Example: Comprehending

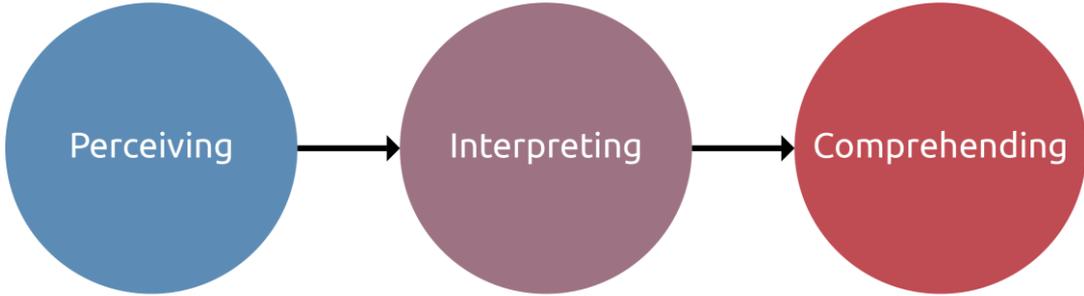
What does it mean to me?



11/13/2019

40

# Understanding: Three Phases



**What do I see?**  
*What data is shown?*  
*How is the data represented?*  
*What features are observable?*

**What does it mean, given the subject?**  
*What features are interesting?*  
*What features are unexpected?*  
*What features are important?*

**What does it mean to me?**  
*What have I learnt?*  
*What do I feel?*  
*What do I now do?*

VISUALISER CONTROL

VIEWER CONTROL

11/13/2019

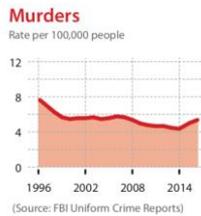
41

# Anatomy of a Chart

Complete chart



Complete chart



11/13/2019

42

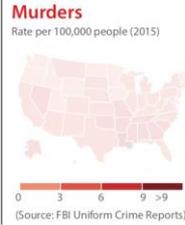
# Anatomy of a Chart

- Scaffolding

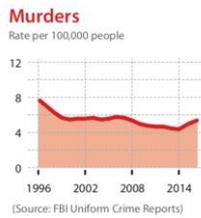
Complete chart



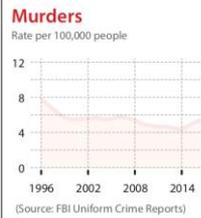
Scaffolding



Complete chart



Scaffolding



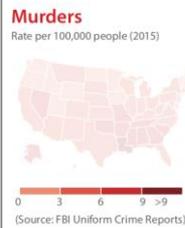
# Anatomy of a Chart

- Scaffolding
- Content

Complete chart



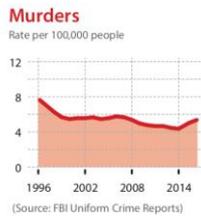
Scaffolding



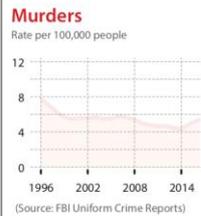
Content



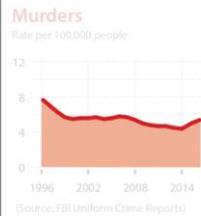
Complete chart



Scaffolding



Content



## Visual Encoding

- **Marks**

Visual placeholders representing data *items* (point, line, shape)

- **Attributes**

Variations in visual appearance of marks to represent values associated with each item (position, size, angle, color, etc.)



11/13/2019

45

## How to read a chart

1. Read the Title, introduction (or caption), and source.
2. Look at the measurements, units, scales, legends.
3. Identify the methods of visual encoding: Color, position, shape, size, etc.
4. Read annotations.
5. Take a bird's-eye-view to spot patterns, trends, and relationships.

11/13/2019

46

# A chart is a *visual argument*.

-: Alberto Cairo

11/13/2019

47

## Chart Types (**CHRTS**)

- **Categorical**  
Comparing categories and distributions of quantitative values
- **Hierarchical**  
Revealing part-whole relationships and hierarchies
- **Relational**  
Exploring correlations and connections
- **Temporal**  
Plotting trends and intervals over time
- **Spatial**  
Mapping spatial patterns through overlays and distortions

11/13/2019

48

# References

- Thorp, Jer: Your Random Numbers – Getting Started with Processing and Data Visualization
- By Jer | April 11, 2010. <http://blog.blprnt.com/blog/blprnt/your-random-numbers-getting-started-with-processing-and-data-visualization>
- The Data Deluge, *The Economist*, Feb 25th 2010.
- Edward Tufte, *Presenting Data and Information: A One-Day Course*. www.edwardtufte.com
- Ira Greenberg, Dianna Xu, Deepak Kumar, *Processing: Creative Coding and Generative Art*, FriendsOfEd, 2012, forthcoming.
- Nathan Yau, *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*, Wiley, 2011.
- Christian Langreiter, Tweet (@chl) at 6:10 AM on September 28, 2011.
- Hal Varian, *On How the Web Challenges Managers*, The McKinsey Quarterly, January 2009.
- National Science Foundation, Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA), Solicitation 12-499, 2012.
- MIT News, *MIT, Intel unveil new initiatives addressing 'big data'*, May 31, 2012.
- Wikipedia, *Data Science*. [http://en.wikipedia.org/wiki/Data\\_science](http://en.wikipedia.org/wiki/Data_science)
- Ben Fry, *Computational Information Design* (PhD Thesis), Massachusetts Institute of Technology, April 2004.
- David Smith, *Revolutions Blog* (<http://blog.revolutionanalytics.com/2011/09/data-science-a-literature-review.html>), September 2011.
- Tweet from CS News Update at 9:54PM on May 21, 2012.
- Matt Stiles, How Common Is Your Birthday? Parts 1 & 2. The Daily Viv Blog, May 12 & 18, 2012. Available at: <http://thedailyviz.com/2012/05/12/how-common-is-your-birthday/>
- Deepak Kumar. Data Science Overtakes Computer Science. ACM Inroads Magazine. Volume 3 Issue 3, September 2012. ACM New York.
- Illuminated Map Displays UK Traffic Casualties Posted by Eugene on December 7, 2011 at 1:00pm (<http://www.mymodernmet.com/profiles/blogs/illuminated-map-displays-uk-traffic-casualties>)
- Min Chen & Luciano Floridi, *An Analysis of Information in Visualization*, Synthese 2012 (to appear), Springer.
- Andy Kirk, *Data Visualisation: A Handbook for Data Driven Design*. 2dn Edition. Sage Publishing, 2019. Website: <http://book.visualisingdata.com>
- Alberto Cairo, *How Charts Lie: Getting Smarter About Visual Information*. Norton, 2019.
- W. Battle-Baptiste and B. Rudert (editors). W. E. B. Data portraits: Visualizing Black America. Princeton Architectural Press, 2018.