

# Data Visualization

FOUNDATIONS (1)

Tea Tušar, Data Science and Scientific Computing

## Outline

What is data visualization?

Why visualize data?

The three principles of good visualization design

- Trustworthiness
- Accessibility
- Elegance

2



# Distinctions in terminology

## Data visualization ≈ information visualization

- Data + meaning = information
- When a distinction is made (we will not make it)
  - Data visualization is concerned with numerical data
  - Information visualization is concerned with abstract data structures

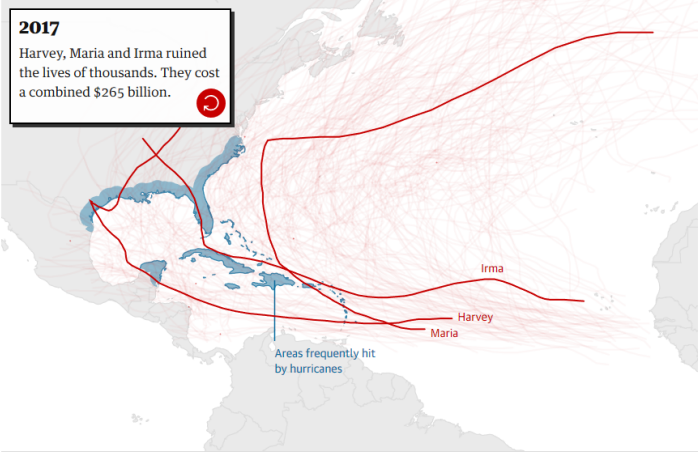
## Scientific Visualization

- Visualization of 3-D phenomena for scientific purposes

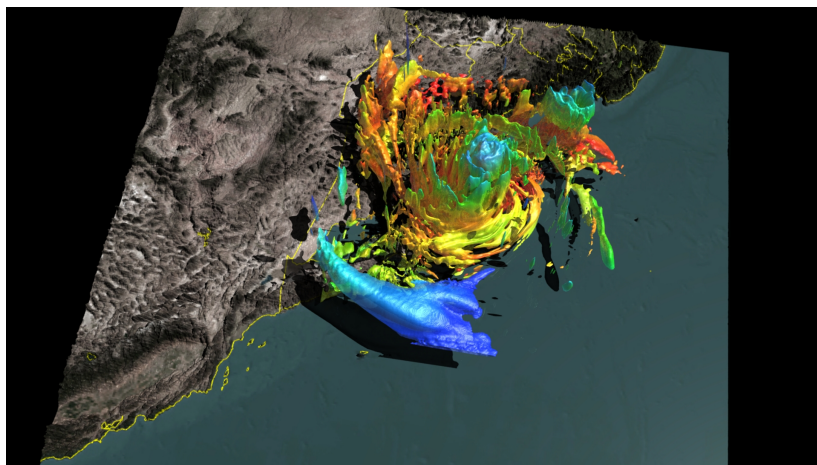
## Infographics

- Use different graphics for explanation (charts, illustrations, photography)
- Traditionally created for print consumption (static)
- Sometimes hard to discern from data visualization

# Data visualization example



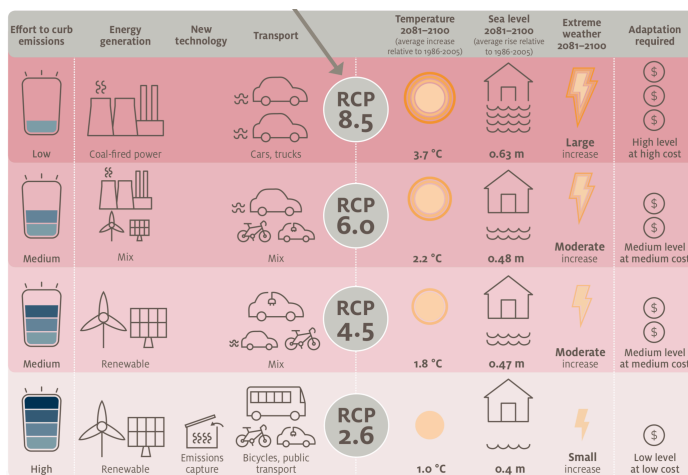
## Scientific visualization example



<https://www2.cisl.ucar.edu/vislab>

7

## Infographic example



<https://coastadapt.com.au/sites/default/files/infographics/15-117-NCCARFINFOGRAPHICS-01-UPLOADED-WEB%2827Feb%29.pdf>

8

## Distinctions in terminology

---

### Interchangeable use

- Chart
- Graph
- Plot
- Diagram
- Map (sometimes!)

9

## Why visualize data?

---

*'A PICTURE IS WORTH A THOUSAND WORDS'*

10

# Anscombe's quartet

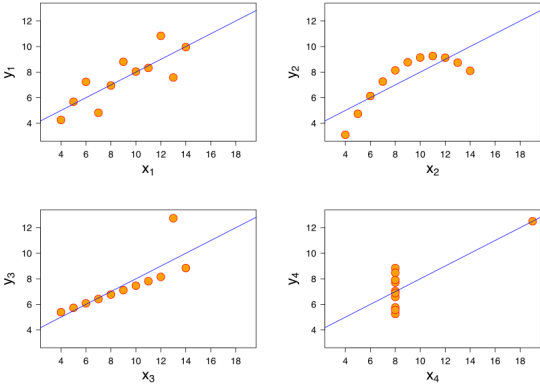
4 datasets with pairs of numbers (x, y) that have nearly identical simple descriptive statistics

Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y	4.125	±0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places

F. J. Anscombe. Graphs in Statistical Analysis. *American Statistician*, 27(1):17-21, 1973  
[https://en.wikipedia.org/wiki/Anscombe's\\_quartet](https://en.wikipedia.org/wiki/Anscombe's_quartet)

# Anscombe's quartet

I		II		III		IV	
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

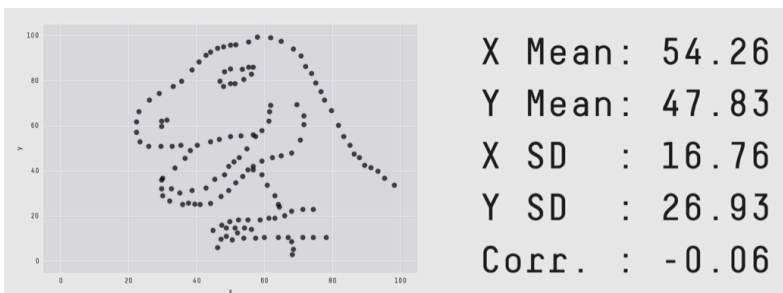


F. J. Anscombe. Graphs in Statistical Analysis. *American Statistician*, 27(1):17-21, 1973  
[https://en.wikipedia.org/wiki/Anscombe's\\_quartet](https://en.wikipedia.org/wiki/Anscombe's_quartet)

# Datasaurus

DrawMyData tool for teaching stats and data science by Robert Grant: <http://robertgrantstats.co.uk/drawmydata.html>

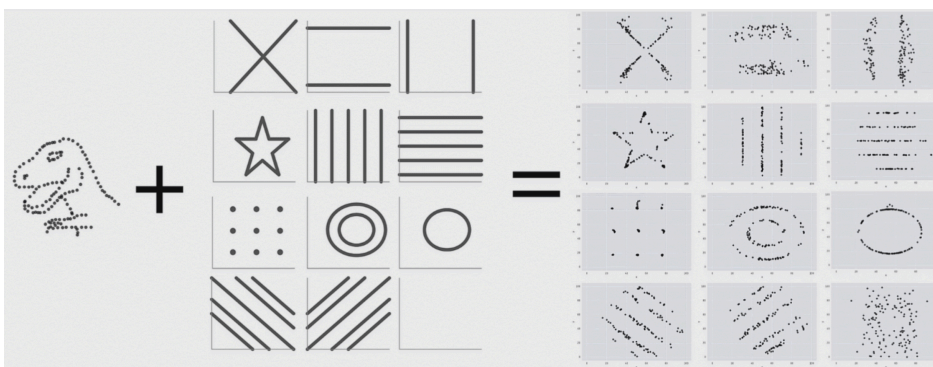
Datasaurus by Alberto Cairo



<http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html>  
<https://www.autodeskresearch.com/publications/samestats>

13

# Datasaurus dozen



Never trust summary statistics alone, always visualize your data

<https://www.autodeskresearch.com/publications/samestats>

14

## Cholera outbreak in London

- In 1854, more than 600 people died of cholera in London's Soho district
- Cause of the disease was unknown at the time
- Two competing theories
  - Cholera is spread by air (predominant)
  - Cholera is spread by water
- Physician John Snow gathered patient data and found the infected water pump
- To convince authorities to close the water pump, he drew a dot distribution map
  - One infected person = one 'dot'
  - Denoted the locations of the water pumps

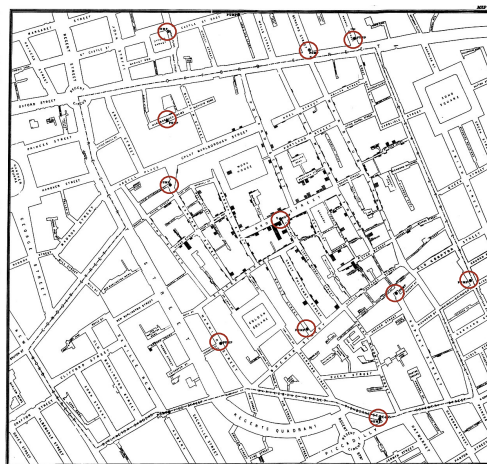
15

## Cholera outbreak in London

Cholera cases clustered around a public water pump on Broad Street



Jo(h)n Snow  
saved the day!



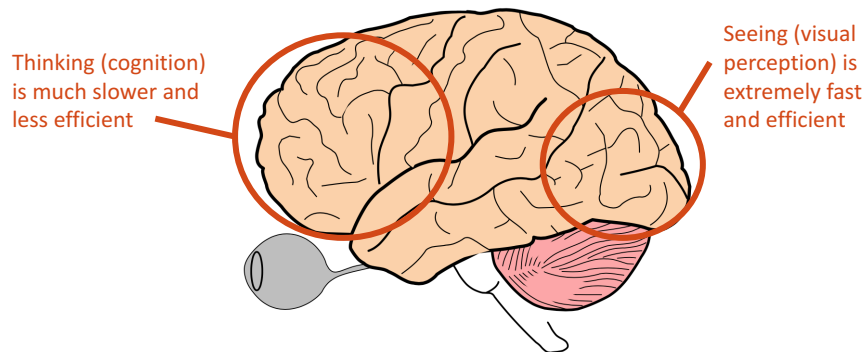
<https://en.wikipedia.org/wiki/File:Snow-cholera-map-1.jpg>

16



## Why does visualization work?

### Perception vs. cognition



Data visualization is effective because it shifts the balance between perception and cognition to take fuller advantage of the brain's abilities

<http://civicmedia.info/resources/stephen-few-data-visualization-for-human-perception>

17

## Purposes of data visualization

### Analyze data to support reasoning

- Develop and assess hypotheses
- Discover errors in data
- Find patterns and correlations

### Communicate information to others

- Present an argument or tell a story
- Inspire

18