Dashing D3.js   **Learn D3 Faster**   Sign In

Why Data Visualizations

A picture is worth 1000 words

A proverb is a simple and concrete saying popularly known and revered, which expresses a truth, based on common sense or the practical experience of humanity. The proverb "A picture is worth 1000 words" is one you have probably heard more than once.

A picture can also be worth 1000 data points. In 1973, the statistician Francis Anscombe demonstrated the importance of graphing data. The Anscombe's Quartet shows how four sets of data with identical simple summary statistics can vary considerably when graphed.

Anscombe's Quartet Data Table

Data Set A		Data Set B		Data Set C		Data Set D	
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

Simple Summary Statistics of Anscombe's Quartet Data Table

Property	Value
Mean of x of each data set	9 (exact)
Variance of x in each data set	11 (exact)
Mean of y in each data set	7.50 (to 2 decimal places)
Variance of y in each data set	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each data set	0.816 (to 3 decimal places)
Linear regression line	y = 3.00 + 0.500x

for each data set	(to 2 and 3 decimal places, respectively)

Graph of Anscombe's Quartet Data Table
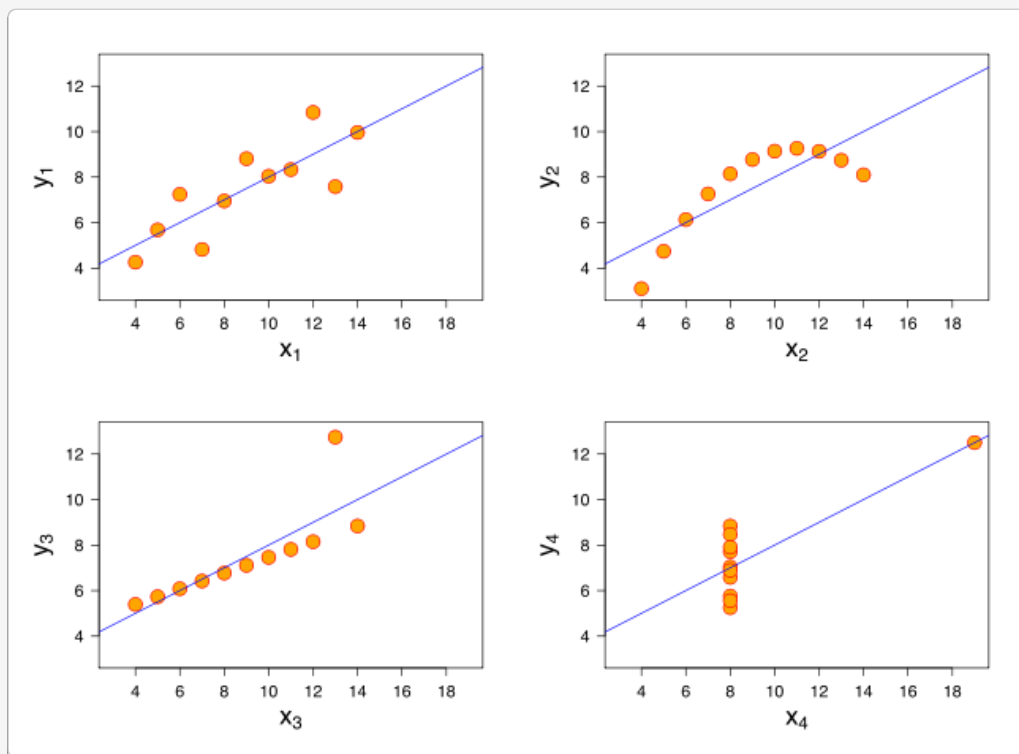
source: Wikimedia Commons

It is hard to tell how the data behaves in the data table. The simple summary statistics table would lead us to believe that all of the data sets are the same. Only when we graph the data do we get a clear picture how the data behaves.
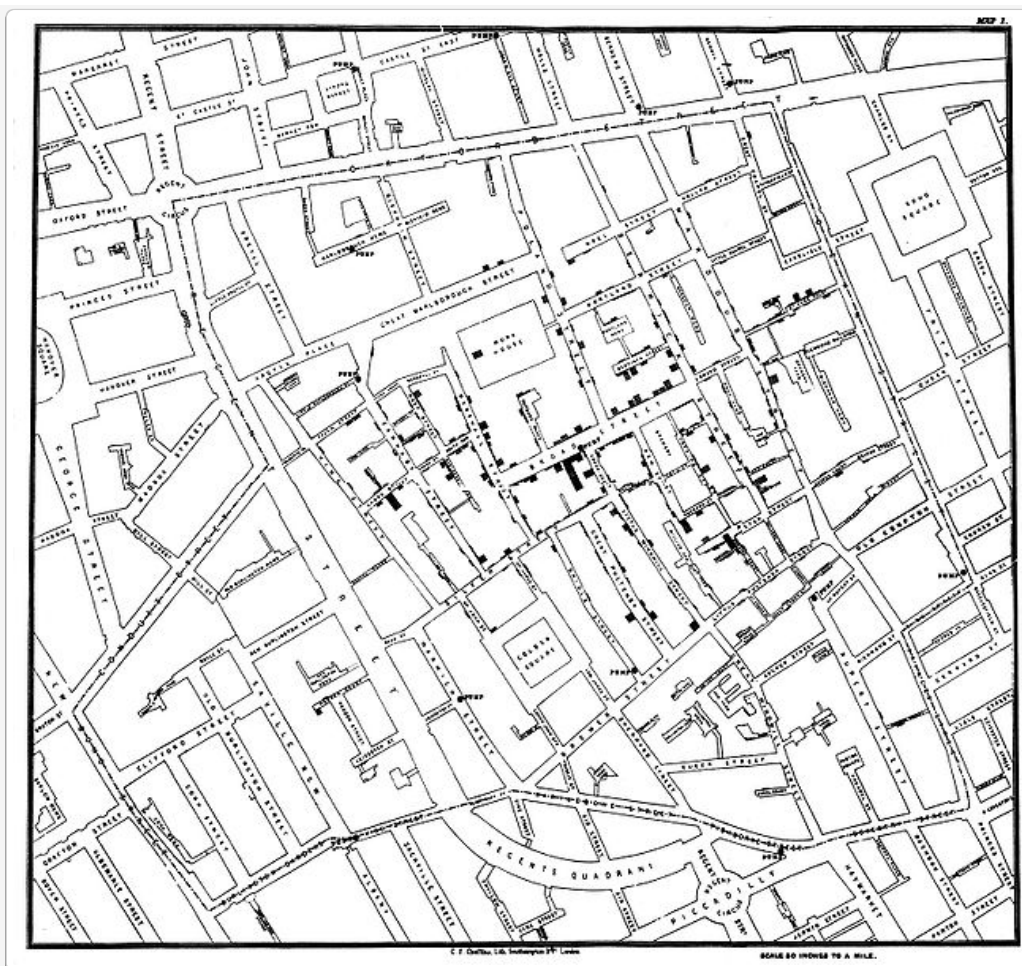
A Famous Data Visualization

In 1854, a cholera outbreak killed 600 people in London. The physician John Snow made this outbreak famous. John Snow used Data Visualization to show that cholera is spread by contaminated water.

At this point in history the germ theory of disease was not known. This theory proposes that micro-organism are the cause of many diseases. Since the theory was not known, the spread of cholera was a mystery to public health officials.

John Snow spoke with locals near the cholera outbreak to discover the source of the germs spreading the disease. He used a "dot distribution map" to show how the cholera cases were clustered. He showed the cases were clustered around a public water pump on Broad Street.

John Snow's Dot Distribution Map of Broad Street Cholera Cases

source: Wikimedia Commons

In addition to this Public Health Data Visualization, John Snow conducted further research. This work is seen as a major event in the history of public health and geography. The work is regarded as the birth of modern Epidemiology.

Data - now and in the future

Gigantic mounts of data are being generated on a daily basis. The amount of data being generated is growing exponentially every year. Below is an info graphic example of the data generated from a single event this past summer - the 2012 Olympic Games.

2012 Olympic Games Data Generated Infographic
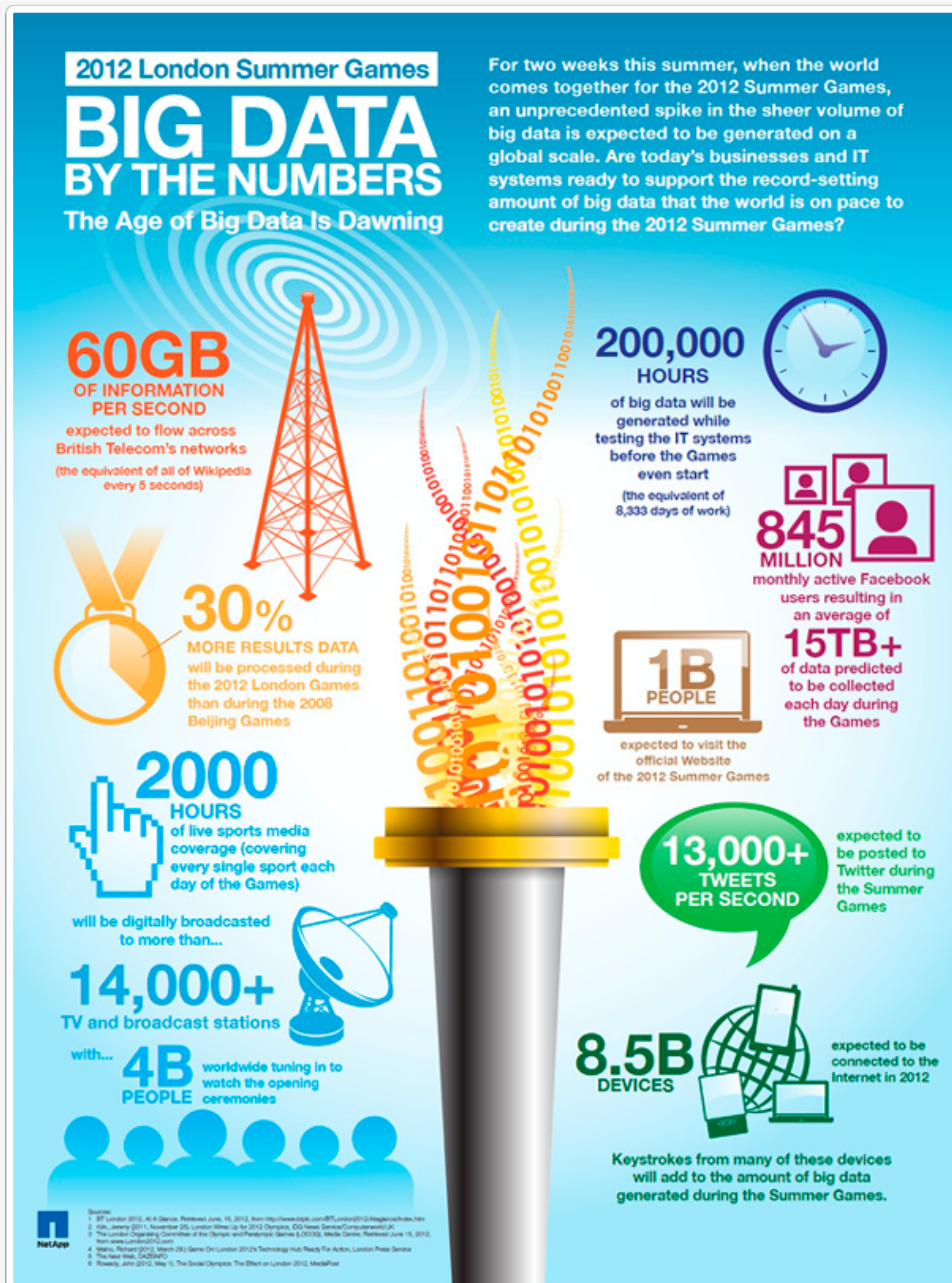
source: NetApp

If one event generates this type of data, just think of how much data is going to be generated and ripe for analysis on a daily basis. This has lead big data analysts to posit that Data is the new Oil.

Data is the new Oil

In 2006, Michael Palmer wrote "Data is the new oil!" declaring "Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."

Since then, several fascinating people have come out saying and supporting the "Data is the new oil" statement. For instance => "Data is the new oil," said Andreas Weigend, Stanford's Head of the Social Data Lab, also the former Chief Scientist at Amazon.

What is interesting for us that that the data has to be "*refined*" to create a valuable entity. Hal Varian, Google's Chief Economist, opined in The McKinsey Quarterly, Jan 2009:

> "The ability to take data—to be able to understand it, to process it, to extract

value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, … because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it."

Which brings us back to the first heading of this section -> "A picture is worth a thousand words." The ability to understand and extract value from data is hugely easier when done through a Data Visualization rather than from looking at the raw data or the simple statistics of the data.

Want to better understand this topic? Check out this step-by-step course => Introductory D3 Training

<div>
←   D3 Tutorial Table of Contents

Why Build Data Visualizations with D3.js   →
</div>

Learn D3.js

D3 Tutorial
D3 Screencasts
D3 Mapping Training
D3 Introductory Training
D3 Intermediate Training
D3 Advanced Training
D3 Corporate Training

DashingD3js.com

Blog
About
Hire Me
D3 Examples
D3 Resources
D3 & Data Viz Newsletter Archive

Data Visualization & D3.js Weekly Newsletter

Get D3.js and Data Visualization news, articles, jobs and more delivered to your inbox every Tuesday:

E-mail

Get the Newsletter

Did you sign up for the newsletter? :)