

# Motivation

## Historical Perspective



# WINTER IS COMING

An Argument from Epidemiology, 1849



# JOHN SNOW'S CHOLERA MAP

An Argument from Epidemiology, 1849



# 1854 Cholera Outbreak

- Tens of thousands people in England are dying of cholera between 1831 and 1854
- Many assumed cholera was airborne (caused by “miasma in the atmosphere”)
- People did not have running water or modern toilets

<http://www.ph.ucla.edu/epi/snow/snowcricketarticle.html> and [http://www.bbc.co.uk/history/historic\\_figures/snow\\_john.shtml](http://www.bbc.co.uk/history/historic_figures/snow_john.shtml)



# 1854 Cholera Outbreak

- Terrible cholera outbreak in 1854 in Soho, near where physician John Snow lived
- Tracked down data from hospitals and public records
- Created simple plot of where victims lived and location of water pumps

<http://www.ph.ucla.edu/epi/snow/snowcricketarticle.html> and [http://www.bbc.co.uk/history/historic\\_figures/snow\\_john.shtml](http://www.bbc.co.uk/history/historic_figures/snow_john.shtml)



Data Visualization  
*Spring 2017*



<http://www.theguardian.com/news/datablog/interactive/2013/mar/15/cholera-map-john-snow-recreated>

# 1854 Cholera Outbreak

- Identified contaminated water pump
- Eventually able to trace many cases to “sherbert” a bubbly drink with a fizzy powder mixed in, served from water coming from the Broad Street area pump
- Pioneered the field of epidemiology *\*maybe*

<http://www.ph.ucla.edu/epi/snow/snowcricketarticle.html> and [http://www.bbc.co.uk/history/historic\\_figures/snow\\_john.shtml](http://www.bbc.co.uk/history/historic_figures/snow_john.shtml)







“How the Ghost Map Helped End a Killer Disease” by Steven Johnson at TedSalon2006, 10:03 min, 2006

[http://www.ted.com/talks/steven\\_johnson\\_tours\\_the\\_ghost\\_map](http://www.ted.com/talks/steven_johnson_tours_the_ghost_map)



# ANSCOMBE'S QUARTET

An Argument from Statistics, 1973



“A computer should make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding.”

– **Francis Anscombe, 1973**

<http://eagereyes.org/criticism/anscombes-quartet>



# Anscombe's Quartet

Group 1		Group 2		Group 3		Group 4	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.56
9.0	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

[http://en.wikipedia.org/wiki/Anscombe's\\_quartet](http://en.wikipedia.org/wiki/Anscombe's_quartet)



# Anscombe's Quartet

Group 1		Group 2		Group 3		Group 4	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.56
9.0	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

[http://en.wikipedia.org/wiki/Anscombe's\\_quartet](http://en.wikipedia.org/wiki/Anscombe's_quartet)



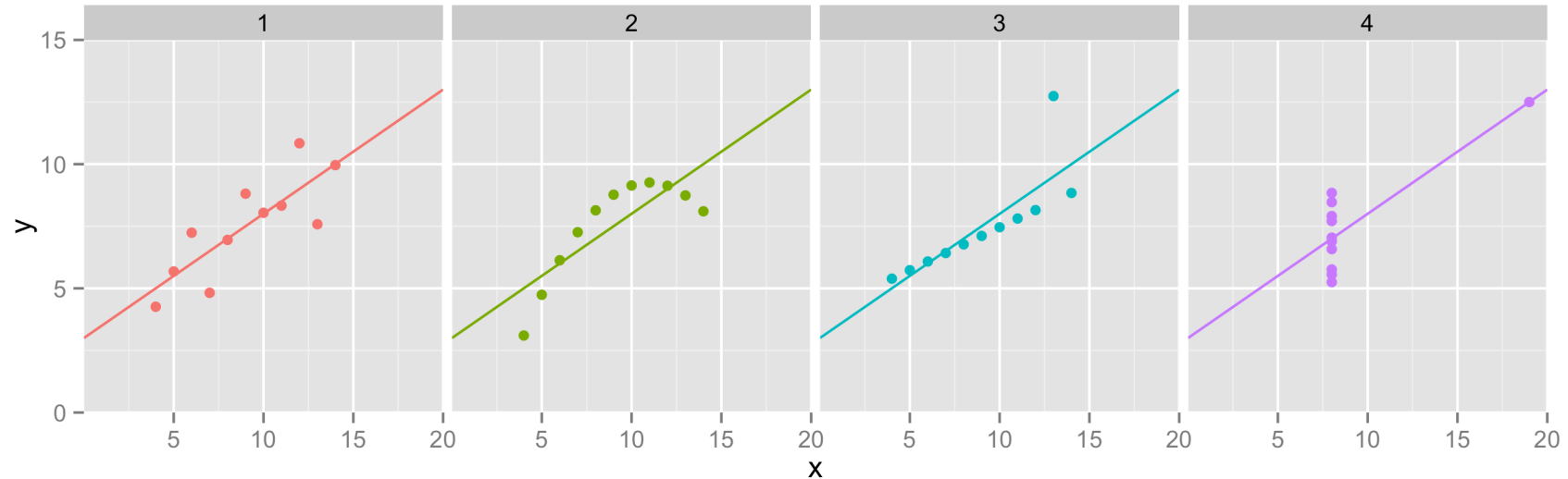
# Simple Statistics

	Group 1	Group 2	Group 3	Group 4
<b>mean(x)</b>	9.00	9.00	9.00	9.00
<b>mean(y)</b>	7.50	7.50	7.50	7.50
<b>var(x)</b>	11.00	11.00	11.00	11.00
<b>var(y)</b>	4.13	4.13	4.12	4.12
<b>correlation</b>	0.82	0.82	0.82	0.82
<b>lm intercept</b>	3.00	3.00	3.00	3.00
<b>lm x effect</b>	0.50	0.50	0.50	0.50

[http://en.wikipedia.org/wiki/Anscombe's\\_quartet](http://en.wikipedia.org/wiki/Anscombe's_quartet)



# Simple Visualization



<http://blog.ouseful.info/2011/08/30/the-visual-difference-%E2%80%93-r-and-anscombe%E2%80%99s-quartet>

# Anscombe's Quartet

- Simple, contrived, but effective example
- Need to analyze **and** visualize your data
- Simple visualizations are often enough to reveal structure

<http://eagereyes.org/criticism/anscombes-quartet>





# WHY VISUALIZE DATA?

Two Perspectives: Statistician vs. Journalist



# Discussion

**Why should we visualize data?**

(i.e. why not let machines take over the world?)





“Let my Dataset Change your Mindset” by Hans Rosling at TED@State, 19:56 min, 2009.

[https://www.ted.com/talks/hans\\_rosling\\_at\\_state](https://www.ted.com/talks/hans_rosling_at_state)





“The Beauty of Data Visualization” by David McCandless at TedGlobal 2010, 17:56 min, 2010  
[https://www.ted.com/talks/david\\_mccandless\\_the\\_beauty\\_of\\_data\\_visualization](https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization)





---

CHANGE THE WORLD FROM HERE