

Descriptive Statistics

Descriptive Statistics

- A descriptive statistic is a description of processed data:
 - A summary statistic
 - A takeaway message
 - Any kind of synthesis
- Fundamental tradeoff:
 - Ease of use vs. Inherent inaccuracy
 - E.g., a batting average reduces a large number of at-bats to one number
 - But does the batter mostly hit singles, or extra-base hits?
 - Are they in a slump? Or in the midst of a hitting streak?
 - Beane discovered on-base-percentage was more informative

Gross Domestic Product (GDP)

- A statistic used by economists to measure economic productivity
- Sum of consumption, investments, government spending, and net exports

2016

| | |
|---------------|--------------|
| United States | \$18,558,130 |
| EU | \$16,477,211 |
| China | \$11,383,030 |
| Poland | \$473,501 |
| South Africa | \$266,213 |
| Iraq | \$148,411 |
| Zimbabwe | \$14,659 |

Gross Domestic Product (GDP)

- A statistic used by economists to measure economic productivity
- Sum of consumption, investments, government spending, and net exports
- Excludes other possibly relevant factors
 - US GDP is \$18.6T and Japanese GDP \$4.4T.
 - Toyota totaled over \$100Bn in US sales in 2016.
 - These \$100Bn in sales are reflected in US GDP, but not necessarily in Japanese GDP.
- If GDP is supposed to measure economic productivity, then we are missing something by excluding these \$100Bn in sales from the Japanese GDP.
- No descriptive statistic is ever perfect. What matters is that it is useful.

Based on this advertisement, would you buy Colgate?



[Image source](#)

But Wait!

- The advertisement makes it seem like 80% of dentists recommend Colgate, while 20% of dentists recommend other toothpastes.
- But what was the alternative?
 - Was it Colgate or bust (i.e., no brushing at all)? Then why only 80%? The other 20% of dentists must really detest Colgate!
 - Or was the alternative other brands of toothpaste? If each dentist could recommend only one brand of toothpaste, and 80% recommend Colgate, then that is meaningful.

But Wait!

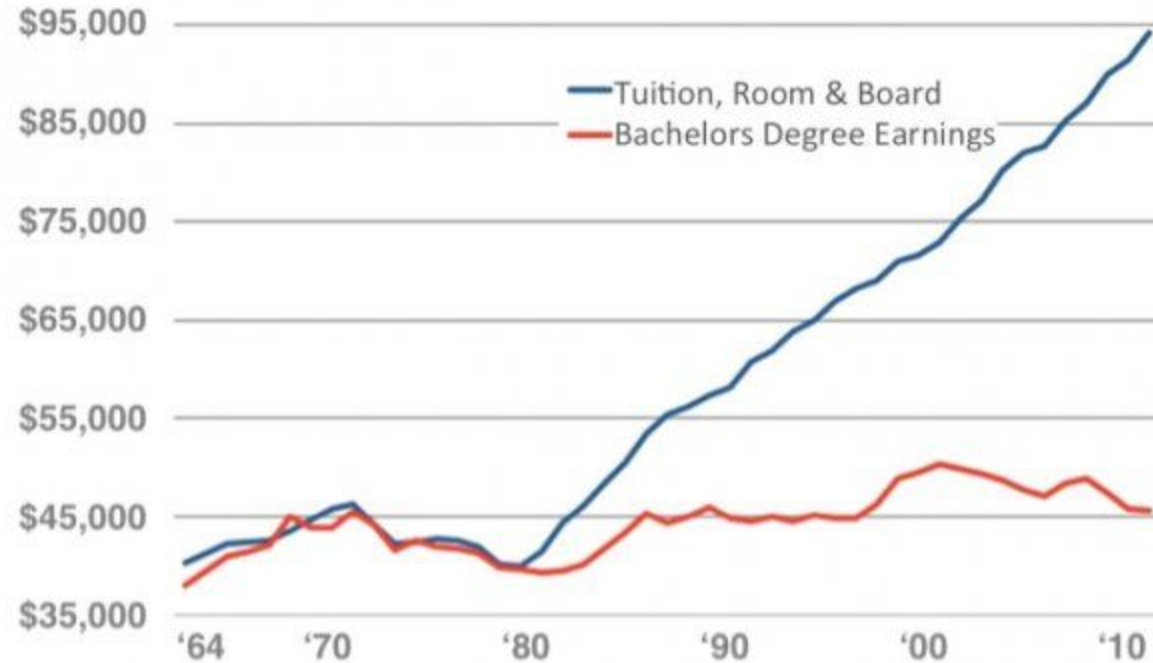
- In fact, this poll used a system of **approval voting**, in which dentists could recommend as many toothpastes as they wanted.
- So every brand of toothpaste could have had an 80% approval rating!
- If Colgate had the highest approval rating, the ad might have read “**More dentists approve of Colgate than any other brand.**” But the fact that it did not suggests that the data do not support this claim.

But Wait!

- The Advertising Standards Association, an organizational advertising watchdog in the UK, forced Colgate to stop showing this ad.
- Interestingly, in a 2013 study that did not allow dentists to recommend multiple toothpastes, Colgate was recommended by 47% of dentists, which was more than any other toothpaste!
- Should Colgate show an ad which touts the number 47%?!
- Maybe they can just say something like: “Colgate comes recommended by a larger percentage of dentists than any other.”

The diminishing financial return of higher education

Costs of 4-yr degree vs. earnings of 4-yr degree



Source: Source: U.S. Census Data & NCES Table 345.

Notes: All figures have been adjusted to 2010 dollars using the Consumer Price Index from the BLS.

[Image source](#)

Cost of College is Rising

- Average cost of a 4 year college degree is approaching \$95,000
- Average annual salary for a college educated worker is \$45,000
- Sounds like a waste of money!

But Wait!

- Average earnings for high school graduates are falling
 - Average high school graduate earns \$12,500 per year
- You earn a salary *annually*
 - Average career lasts 43 years
 - $\$45,000 \times 43 \text{ years} = \$1,935,000$
 - $\$12,500 \times 43 \text{ years} = \$537,500$
 - The difference (well) exceeds \$95,000

Verizon vs. AT&T

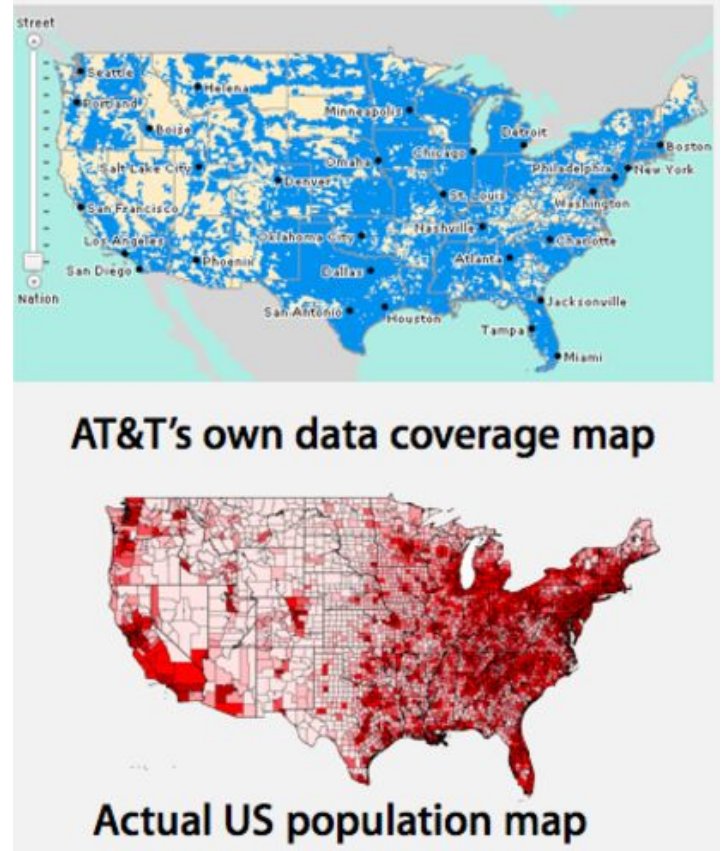
- Verizon is known for comparing its cell phone coverage to its competitors.
- In this ad (from 2009), Verizon claims it has “5 times more 3G coverage than AT&T”.
- AT&T filed a complaint.

The image is a screenshot of the Verizon Wireless website from 2009. At the top, the Verizon logo is on the left, and navigation links for 'Residential', 'Business', and 'Wireless' are in the center. A search bar is on the right. Below this is a red navigation bar with links: 'Phones & Accessories', 'Plans', 'Entertainment & Apps', 'Messaging', 'Business', 'Support', and 'My Verizon'. The main content area features the headline 'America's Largest 3G Network.' followed by the text 'Want 5 times more 3G coverage than AT&T? There's a map for that.*' and 'Browse the web and download music and apps at 3G speed.' Below this are two maps of the United States. The left map, labeled 'Verizon 3G coverage', is predominantly red, indicating extensive coverage. The right map, labeled 'AT&T 3G coverage', is predominantly white with sparse blue patches, indicating much less coverage. At the bottom, there is a small disclaimer: '*COMPARISON BASED ON SQUARE MILES COVERED WITH 3G. VOICE AND DATA SERVICES AVAILABLE OUTSIDE 3G COVERAGE AREA.' and a larger block of fine print: 'THE COMPETITORS' APPROXIMATE 3G COVERAGE IS FROM PUBLICLY AVAILABLE INFORMATION GATHERED PRIOR TO 8/4/2009. THIS DOES NOT SHOW EXACT COVERAGE. WIRELESS SERVICE IS SUBJECT TO NETWORK AND TRANSMISSION LIMITATIONS, PARTICULARLY NEAR BOUNDARIES AND IN REMOTE AREAS. CUSTOMER DEVICE, WEATHER, TOPOGRAPHY AND OTHER ENVIRONMENTAL CONSIDERATIONS AFFECT SERVICE, WHICH MAY VARY SIGNIFICANTLY IN BUILDINGS. MORE RECENT 3G COVERAGE MAY BE AVAILABLE AT VERIZONWIRELESS.COM/3GCOVERAGE AND WIRELESS.ATT.COM/COVERAGEVIEWER. © 2009 VERIZON WIRELESS.'

[Image Source](#)

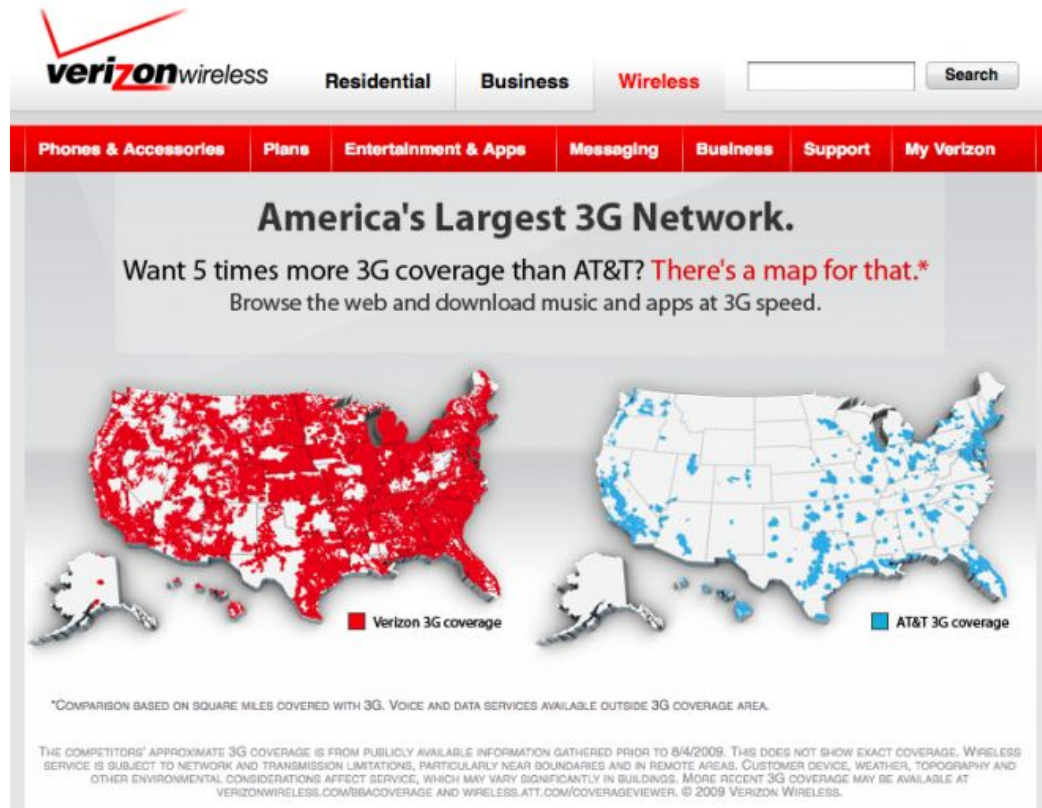
Verizon vs. AT&T

- AT&T argued that from a population standpoint Verizon's claim is false.
- In total, 3G and EDGE reached 303 million people in 2009.
- AT&T's 3G network reached 233 million of them — over 75% of the population.
- Verizon's 3G network covered 284 million people. So, it had only $282/233 = 1.22$ times more coverage.



Verizon vs. AT&T

- Verizon's Coverage Map wasn't based on population 3G coverage. It was based on geographical 3G coverage by square miles.
- AT&T did not argue the fact that by square miles, Verizon did have 5 times more 3G coverage.
- The case was dismissed.



[Image Source](#)

Shootings by Police in the US in 2015

- The Washington Post compiles a [police shootings database](#).
- 990 people were fatally shot in the US by police officers in 2015.
- 49.8% of victims were white.
- 26% of victims were Black.
- 17% of victims were Hispanic.
- What's missing from this picture?

Shootings by Police in the US in 2015

| 2015 | White | Black | Hispanic |
|---------------------|--------------|--------------|-----------------|
| % victims | 49.8 | 26 | 17 |
| % population | 63.7 | 12.2 | 16.3 |

Shootings by Police in the US in 2015

- 22% **fewer** white Americans were shot by police than expected.
 - 493 people versus an expected 630.
 - Less than their representation in the population.
- 112% **more** Black Americans were shot by police than expected.
 - 257 people versus an expected 121.
 - More than twice their representation in the population.
- .7% **more** Hispanic Americans were shot by police than expected.
 - 8 more people than expected.
 - This number falls within a reasonable margin of error.

IL income tax rate increased from 3% to 5%

- Democratic take: “The tax rate increased by 2%.”
 - $5\% - 3\% = 2\%$
- Republican take: “Taxes increased by 67%.”
 - $\$100 \times 5\% = \5 vs. $\$100 \times 3\% = \3
 - $(5 - 3)/3 = \frac{2}{3} = 67\%$
- Both of these claims are factually correct.
- But the messages they send are very different.

Example from *Naked Statistics*, p. 29.

Descriptive Statistics are Summaries

- A summary, by definition, cannot tell the whole story.
- A descriptive statistic is one way of summarizing data, necessarily inaccurate, but sometimes useful nonetheless.
- Only the entire dataset tells the whole story, but is often so impenetrable and unmanageable, that it actually tells you nothing.

Extras

U.S. News College Rankings, 2017

Princeton University

Princeton, NJ

\$47,140

5,400



#1 in National Universities

Harvard University

Cambridge, MA

\$48,949

6,710



#2 in National Universities

University of Chicago

Chicago, IL

\$54,825

5,941



#3 in National Universities (tie)

Yale University

New Haven, CT

\$51,400

5,472



#3 in National Universities (tie)

Columbia University

New York, NY

\$57,208

6,113



#5 in National Universities (tie)

U.S. News College Rankings, 2017

| | |
|--|---|
| Princeton University Princeton, NJ \$47,140 📁 #1 in National Universities | Massachusetts Institute of Technology Cambridge, MA \$49,892 4,524 🔒 □ 📁 #5 in National Universities (tie) |
| Harvard University Cambridge, MA \$48,949 📁 #2 in National Universities | Stanford University Stanford, CA \$49,617 7,034 🔒 □ 📁 #5 in National Universities (tie) |
| University of Chicago Chicago, IL \$54,825 📁 #3 in National Universities (tie) | University of Pennsylvania Philadelphia, PA \$53,534 10,019 🔒 □ 📁 #8 in National Universities |
| Yale University New Haven, CT \$51,400 📁 #3 in National Universities (tie) | Duke University Durham, NC \$53,744 6,609 🔒 □ 📁 #9 in National Universities |
| Columbia University New York, NY \$57,208 📁 #5 in National Universities (tie) | California Institute of Technology Pasadena, CA \$49,908 979 🔒 □ 📁 #10 in National Universities |

U.S. News College Rankings, 2017

| | | |
|--|---|---|
| Princeton University Princeton, NJ \$47,140 🏆 #1 in National Universities | Massachusetts Institute of Technology Cambridge, MA \$49,892 4,524 🏆 #5 in National Universities (tie) | Dartmouth College Hanover, NH \$52,950 4,310 🏆 #11 in National Universities (tie) |
| Harvard University Cambridge, MA \$48,949 🏆 #2 in National Universities | Stanford University Stanford, CA \$49,617 7,034 🏆 #5 in National Universities (tie) | Johns Hopkins University Baltimore, MD \$52,170 6,117 🏆 #11 in National Universities (tie) |
| University of Chicago Chicago, IL \$54,825 🏆 #3 in National Universities (tie) | University of Pennsylvania Philadelphia, PA \$53,534 10,019 🏆 #8 in National Universities | Northwestern University Evanston, IL \$52,678 8,353 🏆 #11 in National Universities (tie) |
| Yale University New Haven, CT \$51,400 🏆 #3 in National Universities (tie) | Duke University Durham, NC \$53,744 6,609 🏆 #9 in National Universities | Brown University Providence, RI \$53,419 6,926 🏆 #14 in National Universities (tie) |
| Columbia University New York, NY \$57,208 🏆 #5 in National Universities (tie) | California Institute of Technology Pasadena, CA \$49,908 979 🏆 #10 in National Universities | Cornell University Ithaca, NY \$52,853 14,566 🏆 #14 in National Universities (tie) |

The Amazon Machine

The VisualCapitalist collected data and created a visualization to compare the size of Amazon to other retailers.

| Company | 2006 | 2016 | % Change |
|-----------|----------|----------|----------|
| Amazon | \$17.5B | \$355.9B | 1934% |
| Walmart | \$214.0B | \$212.4B | 1% |
| Target | \$51.3B | \$40.6B | 21% |
| Best Buy | \$28.4B | \$13.2B | 54% |
| Macy's | \$24.2B | \$11.0B | 55% |
| Kohl's | \$24.2B | \$8.8B | 64% |
| Nordstrom | \$12.4B | \$8.3B | 33% |
| JCPenney | \$18.1B | \$2.6B | 86% |
| Sears | \$27.8B | \$1.1B | 96% |

[Image Source](#)

THE EXTRAORDINARY SIZE OF AMAZON IN ONE CHART

Amazon is bigger than most brick and mortar retailers put together

Market value as of December 30, 2016



Here is how the value of these companies has changed over the last 10 years:



The Amazon Machine

- Amazon increased its value from \$17.5B in 2006 to \$355.9B in 2016.
- The percent increase = $(355.9 - 17.5)/17.5 * 100 = 1,924\%$ increase
- All other retailers decreased in value.

| Company | 2006 | 2016 | % Change |
|-----------|----------|----------|----------|
| Amazon | \$17.5B | \$355.9B | 1934% |
| Walmart | \$214.0B | \$212.4B | 1% |
| Target | \$51.3B | \$40.6B | 21% |
| Best Buy | \$28.4B | \$13.2B | 54% |
| Macy's | \$24.2B | \$11.0B | 55% |
| Kohl's | \$24.2B | \$8.8B | 64% |
| Nordstrom | \$12.4B | \$8.3B | 33% |
| JCPenny | \$18.1B | \$2.6B | 86% |
| Sears | \$27.8B | \$1.1B | 96% |

The Amazon Machine

- Taking a closer look at Amazon vs. the 8 retail stores in the study, Amazon still has a larger market value than all the other stores combined
- Amazon vs. Other = $(355.9 - 297.8) / (297.8) * 100\% = 19.5\%$ larger market value

Market value as of December 30, 2016



21 INDIANS WINS IN A ROW

(AL RECORD)

STATS DURING STREAK

104 RUN
DIFFERENTIAL

| | | |
|--------------|---------------|-------------|
| 41 | 7 | 1.79 |
| HOME RUNS | SHUT- OUTS | TEAM ERA |

NEXT UP:
1916 GIANTS
(26 STRAIGHT WINS)





A handful of statistics that illustrate their dominance during this historic run:

- 105 – Run differential - they've outscored their opponents 142-37.
- 41 – Home runs. Francisco Lindor leads with nine.
- 7 – Shutouts.
- 1.58 – Team ERA.
- 9 – Innings trailed, out of 199 innings.
- 19 – Wins from the starting rotation. Trevor Bauer and Corey Kluber each have four.
- 200-37 – Strikeout-walk ratio.

September 13, 2017