Descriptive Statistics
Raw data are unmanageable

- Impossible to establish trends from looking at it.

name, age, state
Luke, 21, New York
Chuck, 15, Delaware
Denai, 43, Alaska
Sarah, 18, Vermont
Harambe, 76, Illinois
Herold, 34, New York
Chimoiz, 14, North Dakota
Uit, 78, Connecticut
Sarah, 4, Delaware
Martha, 22, South Carolina
Smitty, 25, Maine
Erot, 60, California
Jackie, 12, Virginia
Todo, 89, New York
Jerry, 2, Idaho
Sammy, 17, Florida
Sarah, 88, Georgia
Tammy, 76, Texas
Peter, 54, Kansas
Icto, 35, New York
Louis, 72, Texas
Mayu, 14, Oregon
Jose, 9, Louisiana
Queries and responses

- What state is most common in the dataset?
  - New York
- What name is most common?
  - Sarah
- What letter do most names start with?
  - S
- What is the average age of the people in the dataset?
  - 29.4 years
Raw data are unmanageable

- Impossible to establish trends from looking at it.
- Computational processing is needed for anything beyond a trivial size.
Queries and responses

- What state is most common in the dataset?
  - New York
- What name is most common?
  - Sarah
- What letter do most names start with?
  - S
- What is the average age of the people in the dataset?
  - 31
Descriptive Statistics

- A processed outcome can be:
  - A summary
  - A takeaway
  - Any kind of synthesis

- Fundamental tradeoff:
  - Ease of use vs. Inherent inaccuracy

name, age, state
Luke, 21, New York
Chuck, 15, Delaware
Denai, 43, Alaska
Sarah, 18, Vermont
Harambe, 76, Illinois
Herold, 34, New York
Chimoiz, 14, North Dakota
Uit, 78, Connecticut
Sarah, 4, Delaware
Martha, 22, South Carolina
Smitty, 25, Maine
Erot, 60, California
Jackie, 12, Virginia
Todo, 89, New York
Jerry, 2, Idaho
Sammy, 17, Florida
Sarah, 88, Georgia
Tammy, 76, Texas
Peter, 54, Kansas
Icto, 35, New York
Louis, 72, Texas
Mayu, 14, Oregon
Jose, 9, Louisiana
Gross Domestic Product (GDP)

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

<table>
<thead>
<tr>
<th>Country</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>$18,558,130</td>
</tr>
<tr>
<td>EU</td>
<td>$16,477,211</td>
</tr>
<tr>
<td>China</td>
<td>$11,383,030</td>
</tr>
<tr>
<td>Poland</td>
<td>$473,501</td>
</tr>
<tr>
<td>South Africa</td>
<td>$266,213</td>
</tr>
<tr>
<td>Iraq</td>
<td>$148,411</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>$14,659</td>
</tr>
</tbody>
</table>
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 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.
Rankings

● **The Order of Things**, by Malcolm Gladwell
  ○ Car rankings:
    ■ Porsche Cayman 193
    Chevrolet Corvette 186
    Lotus Evora 182
  ○ Variables: comfort, style, power, “funness”

● **Business School Rankings Calculator**
  ○ You can choose your own weights!

● **2017 Rankings**
  ○ Are the US News and World Report College rankings useful?
Descriptive Statistics are Summaries

- A summary, by definition, cannot tell the whole story.
- A descriptive statistic is one way of summarizing data.
- But even many of them strewn together can still only tell part of the story.
- Only the entire dataset tells the whole story, but is often so impenetrable and unmanageable, that it actually tells you nothing.
Interpreting Statistics
"Weight loss comparison between the active and placebo group was based on the fact that in a 60 day clinical study, the active group lost 2.75 lbs and the placebo group gained 2.18 lbs resulting in a 493% difference between the two groups."

Actually, this is just a difference between the two groups of 4.93 pounds, not a difference of 493%!
iClicker Q: Deciphering Statistics

Based on this advertisement, would you buy Colgate toothpaste?

A: Yes

B: No
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, that is a 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: U.S. Census Data & NCES Table 345.
Notes: All figures have been adjusted to 2010 dollars using the Consumer Price Index from the BLS.
Cost of College is Rising

- Average cost of a 4 year college degree is now almost $100,000
- 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 * 43 years = $1,935,000
  - $12,500 * 43 years = $537,500
  - The difference exceeds $95,000
Summary

- College graduates’ average wages falling doesn't mean college is worse than no college.
- If high school graduates’ wages are falling faster, then college is becoming an increasingly good investment.
Shootings by Police in the US in 2015

- The Washington Post has compiled a police shootings database.
- 990 people were fatally shot in the US by police officers in 2015.
- 49.8% of victims are white.
- 26% of victims are black.
- 17% of victims are hispanic.
- What’s missing from this picture?
# Shootings by Police in the US in 2015

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>% victims</td>
<td>49.8</td>
<td>26</td>
<td>17</td>
</tr>
<tr>
<td>% population</td>
<td>63.7</td>
<td>12.2</td>
<td>16.3</td>
</tr>
</tbody>
</table>
Shootings by Police in the US in 2015

- 14% **fewer** White americans were shot by police than expected.
  - 623 people versus an expected 494.
  - Less than their representation in the population.

- 14% **more** Black americans were shot by police than expected.
  - 258 people versus an expected 119.
  - More than 2 times their representation in the population.

- .7% **more** Hispanic americans were shot by police than expected.
  - 9 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 * 5% = $5 vs. $100 * 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.
Summary

- Descriptive statistics are necessarily inaccurate, but can be useful nonetheless.
- Be careful when interpreting statistics; context matters!
- But even entirely accurate summary statistics, interpreted in context, can favor one interpretation over another.