Misleading With Data

- Lesson 1.1 -
Used by Republican Congressman (Jason Chaffetz) during a congressional hearing.
Thanks, @JoeBiden.

Tweeted by a Democratic House political committee (the DCCC).
Today’s Key Analysis

How can we tell if a data visual is misleading?
Lesson 1.1
Guided Notes

Handout: skewthescript.org/1-1
Topics

1. Quantitative vs. categorical data
2. Misleading graphs
Topics

1. Quantitative vs. categorical data
2. Misleading graphs
Quantitative data: Data that is numerical (think ‘quantities’). The values have an inherent order.

Examples: # of AP classes, SAT score, blood pressure, income, yards per catch, etc.
Quantitative data: Data that is **numerical** (think ‘quantities’). The values have an inherent **order**.

Examples: # of AP classes, SAT score, blood pressure, income, yards per catch, etc.

**Tip:** Finding the average SAT score **makes sense**
Categorical data: Data where values are categories or group labels, which often don’t have an inherent order.

Examples: eye color, ethnicity, favorite sports team, relationship status, etc.
Categorical data: Data where values are categories or group labels, which often don’t have an inherent order.

Examples: eye color, ethnicity, favorite sports team, relationship status, etc.

Tip: Finding the average relationship status doesn’t make sense
## Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>
Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>

Quantitative

- Has inherent order
- Average makes sense
Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>

Categorical

- No inherent order
- Average doesn’t make sense
Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>

Quantitative

- Has inherent order
- Average makes sense
## Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>

### Questions
- Do zip codes have a clear order?
- The ‘average’ zip code?
# Classifying variables

<table>
<thead>
<tr>
<th>Student</th>
<th>Height (in)</th>
<th>Dominant Hand</th>
<th>Final Exam Score</th>
<th>Home Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill</td>
<td>72</td>
<td>Left</td>
<td>77</td>
<td>68494</td>
</tr>
<tr>
<td>Julius</td>
<td>64</td>
<td>Right</td>
<td>83</td>
<td>68492</td>
</tr>
<tr>
<td>Yesenia</td>
<td>67</td>
<td>Right</td>
<td>91</td>
<td>68490</td>
</tr>
</tbody>
</table>

## Categorical

- No inherent order
- Average is meaningless
Topics

1. Quantitative vs. categorical data
2. Misleading graphs
BS
BS
(Bad Statistics)
How to spot a misleading graphic:

1. It may not have axis labels or **scale**.
2. It may **cut off** the x or y axis, or start at a weird place.
3. It may use **pictures** for bar graphs (called a ‘pictograph’).
Was Kobe a ball hog?
Was Kobe a ball hog?

Credit: Stephen Dunn
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

“Ball hog!”
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

(Bad Statistics)
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

The number of “made shots” is double the “assists” per game.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

The two-dimensional ball figure in the pictograph makes the “made shots” appear several times bigger.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

The two-dimensional ball figure in the pictograph makes the “made shots” appear several times bigger.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

The two-dimensional ball figure in the pictograph makes the "made shots" appear several times bigger.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

A regular bar plot shows the difference using **height only**. Now, the “made shots” is double the area of the “assists” bar.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

A regular bar plot shows the difference using height only. Now, the “made shots” is double the area of the “assists” bar.
Was Kobe a ball hog?

In the 2009-2010 season, Kobe made 9.8 shots per game and had 5.0 assists per game, on average.

A regular bar plot shows the difference using height only. Now, the “made shots” is double the area of the “assists” bar.
For the reasons shown, **pictographs** are often misleading.

However, **standard bar graphs** can mislead too!
Advertisement for Chevy truck “dependability”

Example from TED-Ed: “How to spot a misleading graph”
Percent of trucks sold in last 10 years still on road

Example from TED-Ed: “How to spot a misleading graph”
Percent of trucks sold in last 10 years still on road

![Bar chart showing the percentage of trucks sold in the last 10 years still on road for different brands:]

- Chevy: 98.3%
- Ford: 97.5%
- Toyota: 96.4%
- Nissan: 95.4%

Example from TED-Ed: “How to spot a misleading graph”
Percent of trucks sold in last 10 years still on road

Example from TED-Ed: “How to spot a misleading graph”
Percent of trucks sold in last 10 years still on road

Why does 98.3% appear several times greater than 95.4%?

Example from TED-Ed: “How to spot a misleading graph”
Percent of trucks sold in last 10 years still on road

We have a **truncated y-axis** (starts at 95%)

Example from TED-Ed: “*How to spot a misleading graph*”
Percent of trucks sold in last 10 years still on road

Full y-axis (0-100)

Example from TED-Ed: “How to spot a misleading graph”
A place where you’ll find many misleading graphs:

Congress
“That’s the reduction in the breast exams and the red is the increase in the abortions. That’s what is going on in your organization.”

Tweeted by the DCCC, a Democratic House political committee, on Dec. 2nd 2021.
Tweeted by the DCCC, a Democratic House political committee, on Dec. 2\textsuperscript{nd} 2021.
Lesson 1.1
Discussion
Discussion:

- Why might each graphic be misleading? Explain.
PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN

- abortions: 2007, 371 in 2006 vs. 327,000 in 2013
- cancer screening & prevention services: 289,750 in 2006 vs. 935,573 in 2013
Planned Parenthood Services

- Adjusted

Courtesy of Politifact: https://www.politifact.com/factchecks/2015/oct/01/jason-chaffetz/chart-shown-planned-parenthood-hearing-misleading-/
2021 US Weekly Gas Prices, Biden Administration

Source: Energy Information Administration

The Washington Post

Courtesy of Washington Post: https://www.washingtonpost.com/politics/2021/12/02/this-might-be-worst-defense-biden-administration-yet/

Note

Adjusted

skewthescript.org
Lesson 1.1
Practice
Original

Source link: https://obamawhitehouse.archives.gov/blog/2016/10/17/graduation-rate-reaches-new-high-one-student-shares-his-story
Original - scale

U.S. High School Graduation Rates

- 2010-2011: 79%
- 2011-2012: 80%
- 2012-2013: 81.30%
- 2013-2014: 82.10%
- 2014-2015: 83.10%
Adjusted

U.S. High School Graduation Rates

- 2010-2011: 79%
- 2011-2012: 80%
- 2012-2013: 81.30%
- 2013-2014: 82.10%
- 2014-2015: 83.10%
Global Surface Temperature Relative to 1951-1980 Avg. Temp. (Celsius)

Source: NASA/GISS
Adjusted

Global Surface Temperature Relative to 1951-1980 Avg. Temp. (Celsius)

Source: NASA/GISS
Global Surface Temperature Relative to 1951-1980 Avg. Temp. (Celsius)
Average Female Height per country

Graphic found by Sabah Ibrahim on twitter
Average Height Among Women in Different Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Average Height (total inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latvia</td>
<td>65</td>
</tr>
<tr>
<td>Australia</td>
<td>65</td>
</tr>
<tr>
<td>Scotland</td>
<td>65</td>
</tr>
<tr>
<td>Peru</td>
<td>65</td>
</tr>
<tr>
<td>South Africa</td>
<td>65</td>
</tr>
<tr>
<td>India</td>
<td>60</td>
</tr>
</tbody>
</table>