

Describing Data

Part 2: Interpreting Statistics

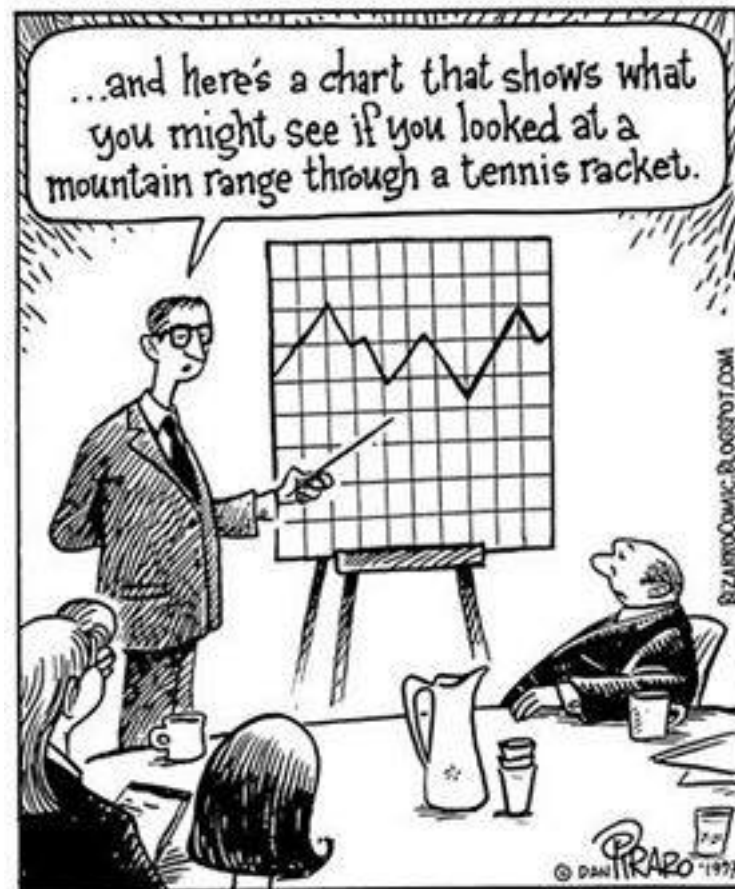
INFO-1301, Quantitative Reasoning 1
University of Colorado Boulder

February 10, 2017

Prof. Michael Paul

Descriptive Statistics

- Purpose: to understand a complex situation through just one or a few numbers
 - Statistics aren't necessarily the complete picture
- What statistics to use?
 - Depends what you value for a problem
- How to interpret statistics?
 - Need to be careful!



Descriptive Statistics

- How to summarize quality of a baseball player?
 - Hitting and running: batting average, home runs, hits, slugging percentage, on base percentage, stolen bases, stolen base percentage, strikeouts, runs batted in, etc.
 - Pitching: wins, winning percentage, saves, earned run average, saves, walks per 9 innings, home runs allowed, complete games, strikeouts, opponents batting average, etc.
 - Fielding: assists, putouts, errors, passed balls, ultimate zone rating, etc.
 - And now many exotic statistics that came out of the Sabermetrics movement

Relative vs Absolute

Illinois state tax rate increased from 3% to 5% by efforts of the Democrats

- In publicity, Democrats focus on the absolute change in the tax rate:
 - **2%** increase
- In publicity, Republicans focus on percentage change in the tax rate:
 - **67%** increase
- Both are correct!



Relative vs Absolute

- Example: Charles Wheelan received a notice that his tax bill to pay for the Tuberculosis Sanitarium District was increasing by 527 percent
 - However, there are not many cases of tuberculosis any more; so the tax bill increase from \$1.15 to about \$6.
- Example: Boss tells you that the company had a good year, so everybody is getting a 10% raise.
 - Your salary is \$35,000 so you are getting \$3500. Your boss's salary is \$200,000 so they are getting \$20,000.



Unit of Analysis

- “Our economy is in the crapper! 30 states had falling incomes last year!”
- “Our economy is showing gains! 70% of Americans had rising incomes last year.”

Both could be correct. How?

- Less populous states (Rhode Island, Delaware, etc.) have falling incomes while more populous states (California, Texas, etc.) have rising incomes

Unit of Analysis

- **Verizon:** we cover a higher percentage of **America** with cell phone service
- **AT&T:** we cover a higher percentage of **Americans** with cell phone service

What's the difference?

- Geographical coverage vs. population coverage

Which is better?

- AT&T better for more people (good in cities!)
- Verizon better if you spend time in less populated places (good for roadtrips!)

Problem with Averages

Bush administration claimed that 92 million Americans would receive an average tax reduction of over \$1000. Fact check:

- Did 92 million Americans get tax cuts?
 - Yes
- Was the mean tax cut over \$1000
 - Yes: \$1083
- Did most families get a cut this large?
 - No: Median tax cut was less than \$100
- Why? Most cuts went to wealthy individuals. Outliers at the top skewed the mean.

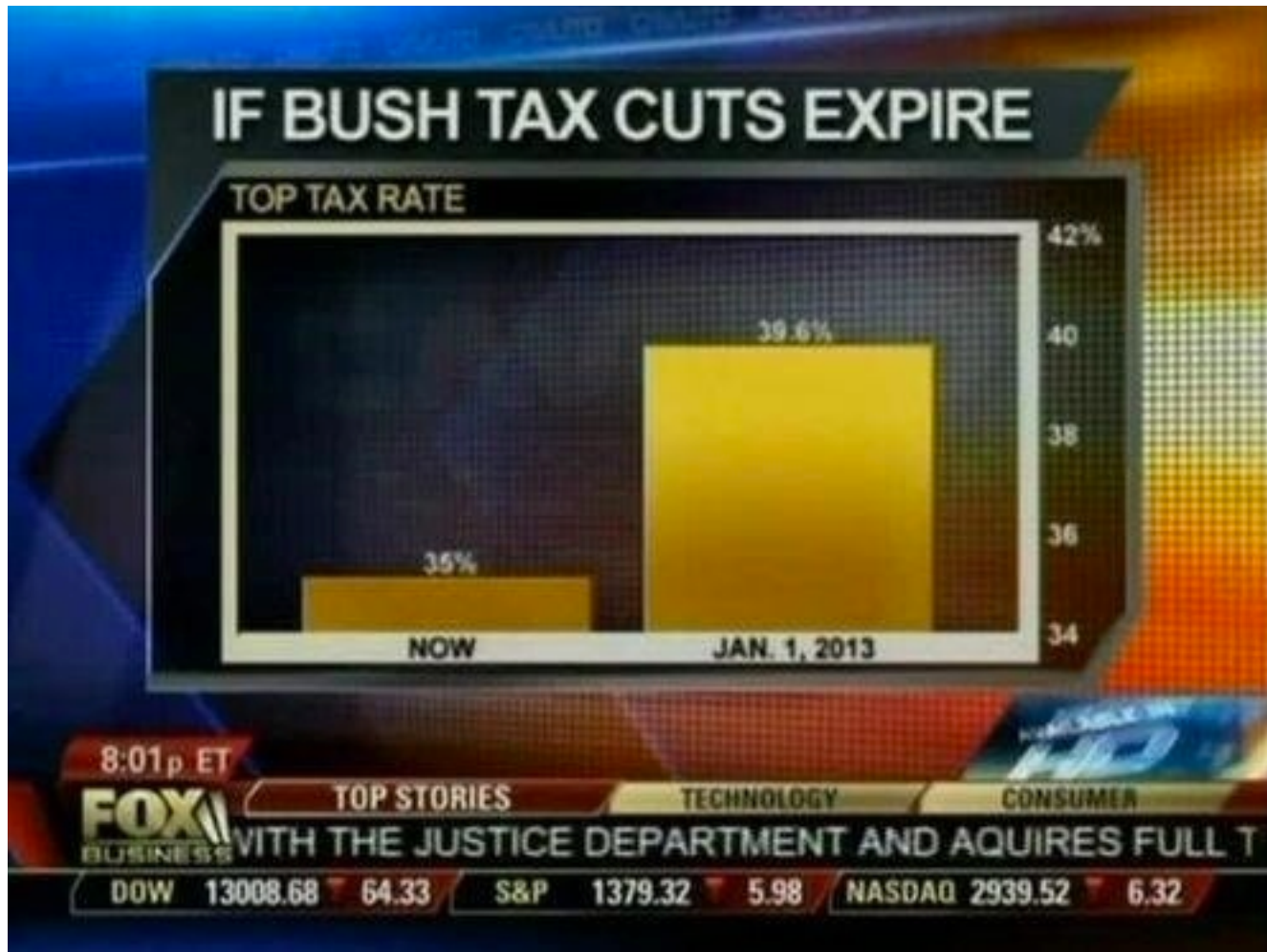
Problem with Medians

- Harvard paleontologist Steven Jay Gould found out that he had a rare form of abdominal cancer (peritoneal mesothelioma)
- Median time from discovery to death: 8 months
- Should he get his life in order because he has less than a year to live?
- Half of the people live longer than the median
- Turns out the mortality distribution is right skewed, so some people live much longer
- Gould lived 20 more years (died from a different cancer)
- He wrote article (playing on Marshall McLuhan) entitled, “The Median Isn’t the Message”

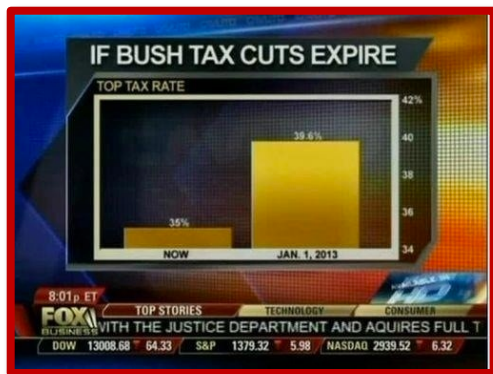
Misleading Data

- Houston public schools reported 1.5% dropout rate: the best rate in the country
- Investigative journalists wanted to find out why:
 - Rod Paige, the Houston school superintendent, gave financial incentives to school principals to have high test scores and low dropout rates; did not monitor how the principals did this.
 - Schools classified almost all dropouts as transferring to another school, returning to their native country, or leaving to pursue a General Equivalency Diploma.
 - Actual annual dropout rate in Houston public schools exceeded 25%.
 - Schools kept standardized test scores high by flunking out poor students before 10th grade (the year in which the standardized test is administered) and in at least one case by making a student take 9th grade 3 times and then promoting him directly to 11th grade.

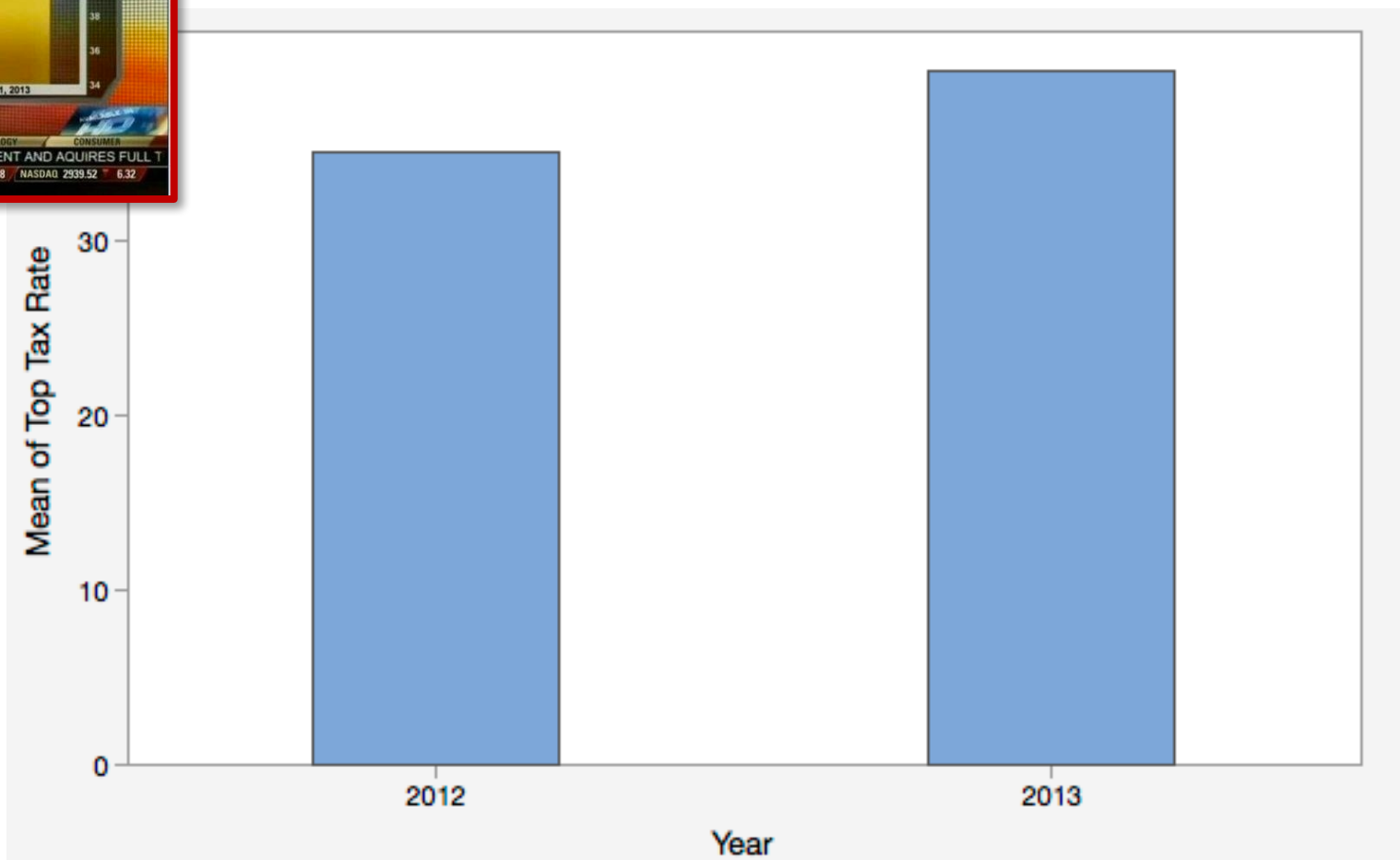
Misleading Visualizations



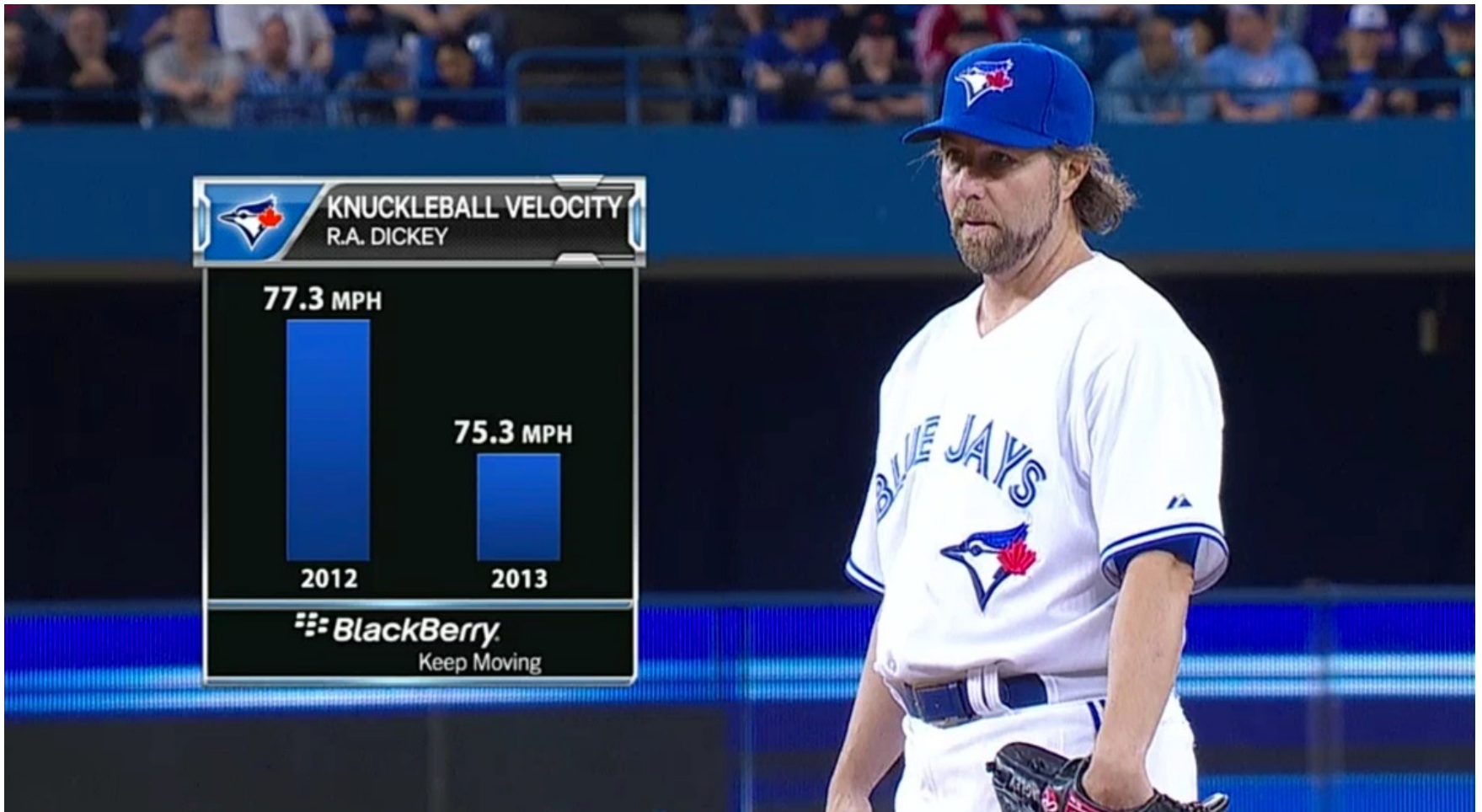
Misleading Visualizations



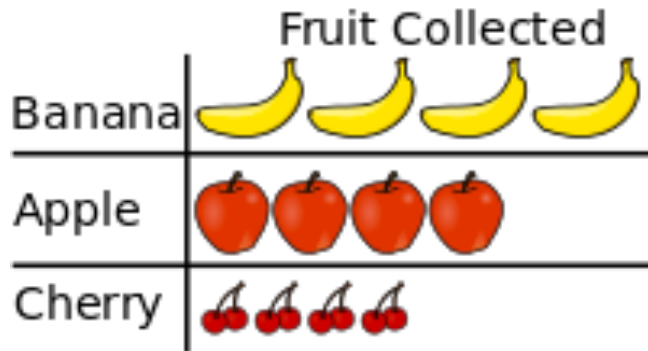
Fixed



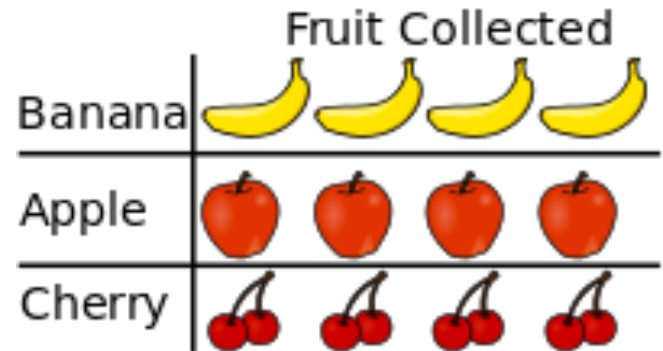
Misleading Visualizations



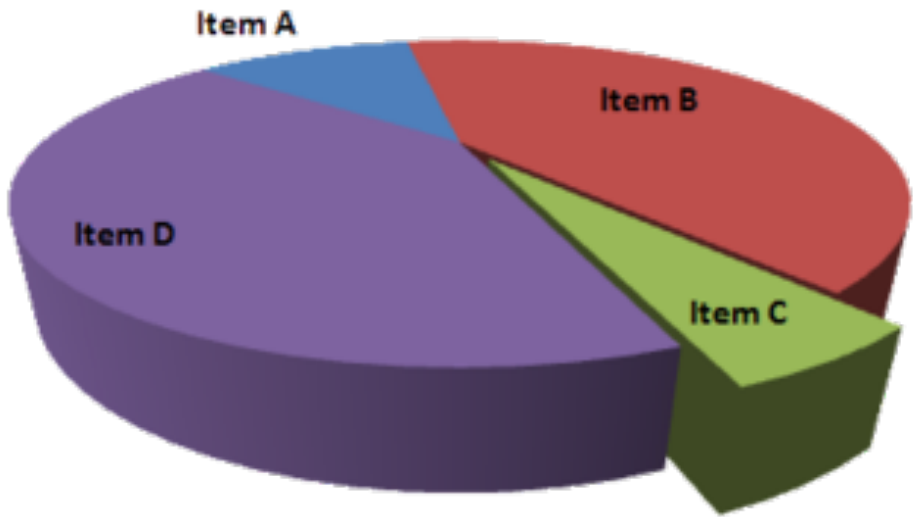
Misleading Visualizations



vs



Misleading Visualizations



VS

