# Section 3: Measures of Central Tendency and Rankings

"The Order of Things" and US News & World Report

## Homework 0 Review

Lies, Damn Lies, and Statistics

### Practice Problem:

Grade Breakdown on a CS Homework Assignment



Score

Total students: 35

Mean: 75 Median: 80 Mode: 85

Mode: 85

If the mean was an A, 21 students would receive an A. If the median was an A, 18 students would receive an A. If the mode was an A, 12 students would receive an A.

Score	# of Students
55	6
60	3
65	1
70	4
75	3
80	6
85	7
90	2
95	2
100	1

- Shape of the probability distribution?
  - Skewed Right
  - (Either diagnosed upon death, or live numerous years after 8 months)
- What did he infer?
  - If diagnosed early, then likely to be in the long, right tail.



My biggest takeaway from Stephen Jay Gould's story is that data represents limitless possibilities in multiple ways. To Gould, the median mortality indicated an indefinite range of possibilities for his longevity post-diagnosis. In other words, the median of the data set only indicated the middle – and as Gould surmised, there's no telling where the end could be and why he shouldn't be a part of it.

More specifically, and to me more excitingly, a single data point can be and often is representative of diverse narratives. Numbers of persons succumbing to Gould's rare abdominal cancer in a certain time frame bears dismally for the curious researcher and inquiring patient. The stories behind those battles, however, can convey so much about access to medical care, wealth, life and family history, personal missions, and more. My primary field is anthropology – I'm intrigued by how the relationships between all those factors impact real people. So while for many, data may be the summation of narratives, for me it is a launching point – as it was for Gould, who read the data present with a sense of opportunity and exceeded all expectations with his will to defy it.

An example data set could be the delay time of airlines. The mean delay time is higher than the median as it is getting pulled to the right by flights that have very long delay times. Most flights leave early, leave on time, or have very short delay times leading to a median delay time that is less than the mean. Therefore if we only focus on the mean of this dataset we might mistakenly believe that a flight we booked could have a very long delay time. If on the other hand we focus on the median and where most of the data lie, we might be able to predict the potential delay time of our flight more accurately.



#### Source

- Senator Ted Cruz said that there had been a "disproportionate increase" since 2009 in funding of earth sciences within NASA "at the expense of exploration and space operations, which I believe should be central to the core mission of NASA."
- There has been an increase, and it is larger than some other NASA areas but spending on earth science is lower now as a percentage of NASA's overall budget than it was in the fiscal year 2009.
- Ted Cruz was looking at the absolute increase in spending, rather than the relative increase.

#### What Sequester?

NASA's 2014 budget would revert to last year's higher level under President Barack Obama's proposal to replace sequestration with a deficit reduction plan that includes tax hikes and entitlement reform. In millions of U.S. dollars.





The unemployment rate is calculated by the Bureau of Labor Statistics based on a survey of 60,000 households that is meant to be representative of the U.S. population. Members of the Census Bureau conduct the surveys, which ask questions like, "Last week, did you do any work for (either) pay (or profit)?" and "Have you been doing anything to find work during the last 4 weeks?"

It determines the unemployment rate by finding the number of people who are employed, unemployed, and the size of the labor force. It bases these calculations on the following definitions, according to the Bureau of Labor Statistics website: "People with jobs are employed. People who are jobless, looking for a job, and available for work are unemployed. The labor force is made up of the employed and the unemployed. People who are neither employed nor unemployed are not in the labor force."

The Bureau of Labor Statistics has outlined some of the efforts made to avoid inaccuracy in reporting the unemployment rate. For instance, it does not use the number of people collecting unemployment insurance benefits to measure unemployment rate because sometimes benefits run out before a person finds a job and some people are ineligible or do not apply for unemployment insurance. However, the reported unemployment rate likely does not tell the whole story. For instance, people who are not employed but not actively looking for a job are considered out of the labor force and not counted in unemployment rate. It also does not count people who may be underemployed, meaning that their job does not allow them to use their skills to their maximal societal benefit. Thus, while the unemployment rate is a useful statistic, it does not answer all pertinent questions about the labor force.

US College rankings have inherent flaws because it **oversimplifies rankings**, **ignoring other important variables** that should have been taken into account.

**Penn State** - public, large and diverse student body, low tuition, rural

vs Yeshiva - private, small and Jewish majority student body, high tuition

Even after we take different variables into account, **weighing** them appropriately is also a difficult problem. In fact, "There's no direct way to measure the quality of an institution—how well a college manages to inform, inspire, and challenge its students..."

US News Report is a self-fulfilling prophecy because the main factor that it judges on is a school's reputation. However, this is a cyclic problem because a school's reputation depends on lists such as US News Report rankings and the rankings depend on that reputation.

"Coffee Actually Makes You Live Longer, New Report Confirms"

 $\rightarrow$  Causation

- Coffee is the direct *cause* of longer lifespan

"Coffee Drinkers Are More Likely To Live Longer. Decaf May Do The Trick, Too."

- $\rightarrow$  Correlation
  - Relationship is found, but coffee may not be direct cause
  - Could be other factors in longer lifespan than just coffee

### US News & World Report

"We use six factors . . . to assess a school's commitment to instruction. Class size has two components, the proportion of classes with fewer than 20 students (30 percent of the faculty resources score) and the proportion with 50 or more students (10 percent of the score). Faculty salary (35 percent) is the average faculty pay, plus benefits,..., adjusted for regional differences in the cost of living. . . . We also weigh the proportion of professors with the highest degree in their fields (15 percent), the student-faculty ratio (5 percent), and the proportion of faculty who are full time (5 percent)."

Potential problems?



### Reputation: 22.5%

"Every year, the magazine sends a survey to the country's university and college presidents, provosts, and admissions deans (along with a sampling of high-school guidance counsellors) asking them to grade all the schools in their category on a scale of one to five. Those at national universities, for example, are asked to rank all two hundred and sixty-one other national universities—and Morse says that the typical respondent grades about half of the schools in his or her category."

Potential problems?

### Biases

- Cyclical relationship
  - High USNWR ranking leads to a good reputation leads to a high USNWR ranking
- Personal biases
  - People are determining these rankings, and people are inherently biased
- Imperfect information
  - Lawyers who were asked to rank law schools ranked Penn State in the middle of the pack.
  - But at the time, Penn State didn't have a law school!
- Unclear objective function
  - What really makes the best school?