

Analysis Basics

But first: The records request project

- Throughout the semester, you are going to have to request and acquire a dataset from the state government that is not online.
- You will have all semester long to do this.
- You must document everything you do, each answer you get from the agency, what you ultimately end up with.
- You will submit your data and give a lightning talk about your experience on Nov. 30.
- DO NOT BLOW THIS OFF. Failure to get something is an F. This may take you a week. This may take you all semester.

A word about math: You are not bad at math.

The secret to this class: Stop. Think. Ask.

The beginner's mind

The very basics: Terminology

1. A data file
2. Records
3. Rows
4. Columns
5. Field/Cells/Objects

Things you can do with data

- View it
- Sort it
- Select things from it
- Exclude things from it
- Group like things together
- Calculate things.

Things you can do with data: Calculate things edition

- Count things (usually have to group them first)
- Measure the middle (averages and medians)
- Measure change (percent change)
- Calculate rates (compare NYC to Lincoln)

Count things

- Raw data versus summary data
- Normalized data
- Demo with Mountain Lions

Measure the middle

- The difference between average and median
- Demo with NU salaries
- When to use medians vs when to use averages

Measure change

- The sad state of newsroom math skills
- Numeric change vs percent change
- Repeat after me: $\text{New} - \text{Old} / \text{Old}$
- Orders of operations: PEMDAS.
- So repeat after me: $(\text{New} - \text{Old}) / \text{Old}$.
- Demo of UCR data

Calculate rates

- Rates: Make big numbers comprehensible and disparate things comparable
- Choose a base.
- $(\text{Number}/\text{Total}) = \text{Proportion}$
- Proportion x Z where Z is the number of people you want to calculate the rate for. 1,000? 100,000?

For next time

- Read Herzog Chapter 8
- Read Numbers in the Newsroom 1 and 2
- Bring your laptop. We get hands on next time.