

What's the FREQ?

Unit 1 - Lab 8

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

USA! USA!

- In the previous lab, we cleaned our American Time Use Survey (ATUS) data to make it usable.
- Now we can actually start to analyze it to learn how people spend their time in the US.
- The process of cleaning and then analyzing data is very common in Data Science.

Summarizing data

- Plots are extremely useful.
 - Help find interesting patterns.
 - Easy to make comparisons.
- Load your cleaned ATUS data set:

```
data(atus_clean)
```

- Create a plot that compares people's genders and the amount of time they spent on homework
 - Using this plot, do you think males or females spend more time doing homework?
 - Find someone in your class who came to the opposite conclusion as you did about which gender, typically, does more homework.
- Instead of relying solely on plots, we want to **make comparisons** using numbers.

How do we summarize categorical variables?

- When we're dealing with categorical variables, we can't just calculate an **average** to describe a *typical* value.
 - (Honestly, what's the average of categories *orange*, *apple* and *banana*, for instance?)
- When trying to describe categorical variables with numbers, we calculate **frequency tables**

Frequency tables?

- When it comes to categories, about all you can do is **count** or **tally** how often each category comes up in the data.
- Let's calculate how many *males* and *females* are in our data set.
- Type the following into the console:

```
tally(~gender, data = atus_clean)
```

- How many more *females* than *males* are there?

2-way Frequency Tables

- Counting the categories of a single variable is nice, but often times we want to make comparisons.
 - For example, what if we wanted to compare the number of people with physical challenges and their genders?
- To make these types of comparisons, we can make a **2-way frequency table**:

```
tally(phys_challenge~gender, data = atus_clean)
```

- Run this command and write down what you notice about numbers of *males* and *females* with physical challenges. Are we still counting?

Interpreting 2-way frequency tables

- Recall that there were 1371 more women than men in our data set.
 - Comparing **counts** then doesn't make sense.
 - If there are more women, then we might expect women to have more physical challenges (compared to men).
- Instead of using **counts**, we use **percentages**
 - So for instance, roughly 89.198% of men do not have a physical difficulty.

Why do we use percentages?

- Percentages let us make comparisons between groups, even if one group is much larger than another.
- Run these two lines of code:

```
tally(phys_challenge~gender, data = atus_clean)
```

```
tally(phys_challenge~gender, data = atus_clean)
```

- ****How many *females* had a __physical challenge__ (using counts)?****
- ****How many *males* had a __physical challenge__ (using counts)?****

Changing our format

- If we did want to make a table with *counts* instead of *percentages*, we can change the **format** of our frequency table.
- Run the following line:

```
tally(phys_challenge~gender, data = atus_clean, format = 'count')
```

- Which gender had more people report NOT having a physical challenge, in terms of counts?
- Which gender had more people report NOT having a physical challenge, in terms of percents?

Adding margins

- Making frequency tables with counts can be misleading.
- To make them less so, we can add in the **margins** or totals for each gender.
- Run this line of code:

```
tally(phys_challenge~gender, data = atus_clean,  
      format = 'count', margins = TRUE)
```

- Explain what happened to the output by including `margins = TRUE` in the function.

On your own

- Which gender has a higher rate of *part time employment*?
- Explore the amount of time each gender socializes
 - Create a subset of data of people who socialized at least 1 minute or more
 - Compare the average amount of socializing done by each gender
 - Does one gender's amount of socializing vary more than the other?