

# Introduction to Data Science

Professional Development

Follow-up #1

Tuesday, October 24, 2017

8:00 am-3:00 pm

# Agenda

- Welcome & Overview
- Effective Student Teams
- Data in the Headlines
- What is an Experiment?-Lesson 2
- Let's Try an Experiment-Lesson 3

## **BREAK**

- Predictions, Predictions-Lesson 4
- Time Perception Experiment-Lesson 5
- RStudio-Lab 3A

## **LUNCH**

- Observational Studies vs. Experiments-Lesson 7
- Monsters that Hide in Observational Studies-Lesson 8
- RStudio-Lab 3B
- Performance Task
- Announcements & PD Evaluation

# What's in Unit 3?

## CCSS-M S&P:

- S-IC 1. Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- S-IC 3. Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- S-IC 6. Evaluate reports based on data.

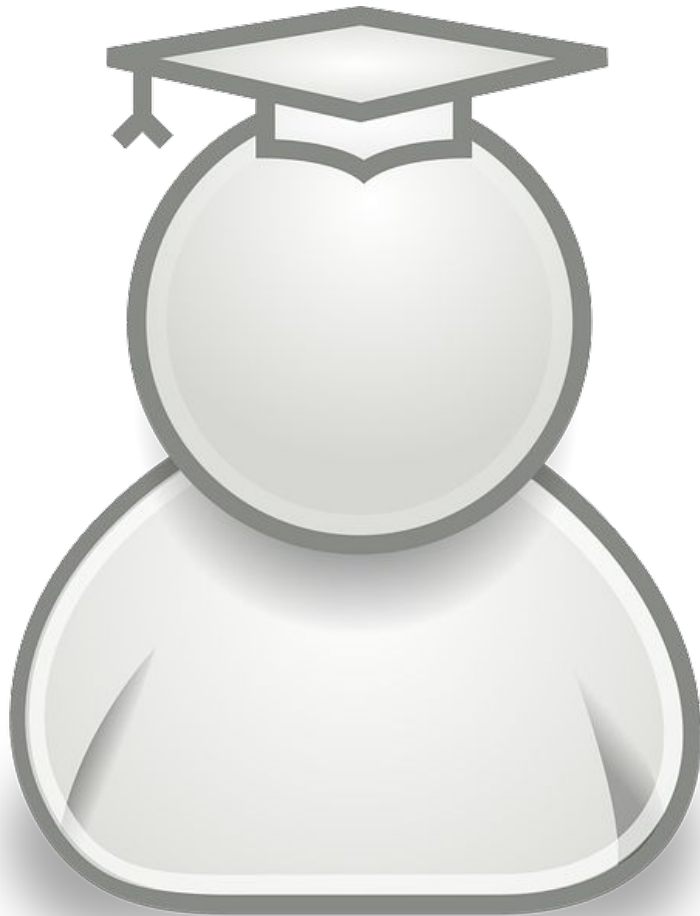
## Data

## Collection/Gathering Methods

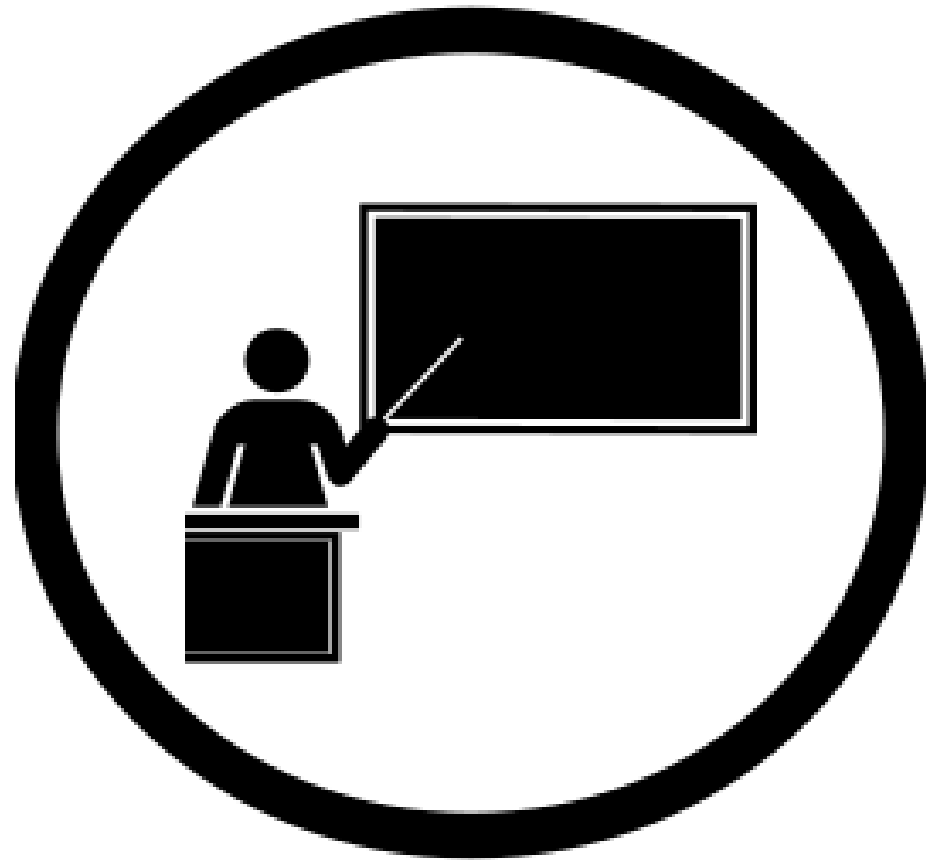
### Unit 3 Themes:

1. Experiments
2. Observational Studies
3. Surveys
4. Sensors
5. Webpages

# Wear the Right Hat



Student Mode



Teacher Mode



# T-Chart



Write one thing  
you feel has  
been a success  
so far this school  
year.

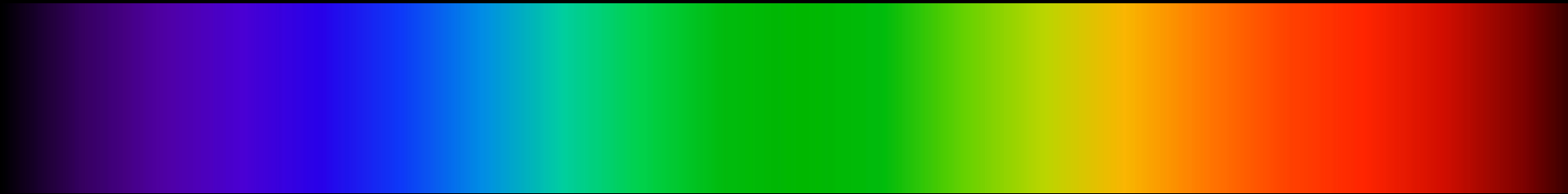
Share one  
challenge you  
face teaching  
IDS.

# Continuum

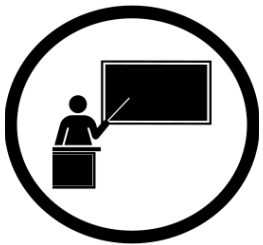
Unit 1  
Lesson 1



Unit ?  
Lesson ?







# Effective Student Teams





“Learning environments for the 21<sup>st</sup> century must be ones in which students are actively engaged with learning tasks and with each other. Today, teachers are in competition with...all sorts of technology, and the expectation that children will learn passively is increasingly unrealistic.”

-Robert E. Slavin

from: *The Nature of Learning: Using Research to Inform Practice*,  
2010

THE QUESTION IS...

WHAT MAKES  
GROUP-WORK  
WORK?



# Making Group-Work Work

<https://www.teachingchannel.org/videos/structured-groups>

Write down anything that you either learned or got validation for as you watch the video.

**Be prepared to share.**

# Elements of Effective Co-operative Learning

```
graph TD; A[Elements of Effective Co-operative Learning] --- B[ ]; B --- C[Group Goals]; B --- D[Individual Accountability];
```

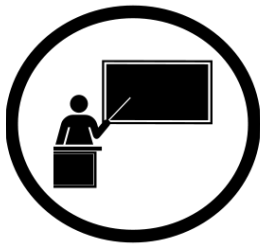
Group Goals

Individual  
Accountability

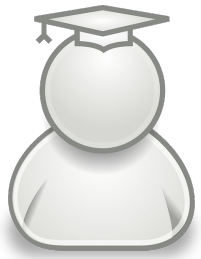


# Group Roles

- #1 Recorder
- #2 Reporter
- #3 Resource Manager
- #4 Faciliatator



# Discussion Norms



During our class discussion, **EVERYONE** is expected to be:

- LISTENING
- CONTRIBUTING
- BUILDING ON EACH OTHER'S IDEAS
- SPEAKING CLEARLY AND LOUDLY ENOUGH FOR ALL TO HEAR



# Data in the Headlines

“People Who Order Coffee Black Are More Likely To Be Psychopaths”

- Huffington Post Article

**To decide whether this is a legitimate claim, what would you ask the researchers?**

**\*you can ask anything**



# Correlation, Association, & Causation

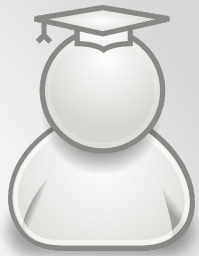
- What's the difference?
- When can we determine causation?



# Lesson 2: What is an Experiment



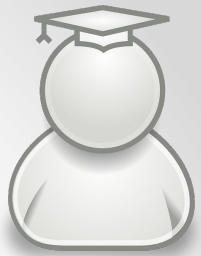
**Objective:** Students will learn about the elements of an experiment and the meaning of “causation”. Student will learn to distinguish claims of causation from claims of association.



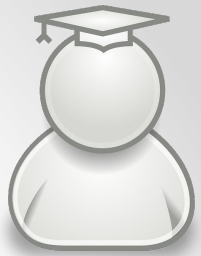
# Headlines

- a. Stop Global Warming: Become a Pirate
  - b. Lack of sleep may shrink your brain
  - c. Early language skills reduce preschool tantrums
  - d. Dogs walked by men are more aggressive
- 
1. What is the headline implying with its wording?
  2. Is it implying causation or association/correlation?
  3. How can you tell the difference between causation and correlation? What words stand out in these headlines?

# “Is Yawning Contagious?”



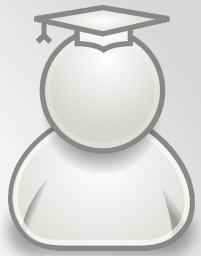
- a. What is the headline implying with its wording?
- b. Is it implying causation or correlation? How do you know?
- c. How can we determine if yawning is contagious?



# Mythbusters

How did the Mythbusters design the investigation?  
What steps did they take? How was their  
approach different than yours?





# Experiment

Mythbusters conducted an **EXPERIMENT** which is one method of data collection.

What were some of the elements of the experiment?



# Discussion

- a. Why did the Mythbusters follow all of these steps to design their experiment?**

We don't know how Mythbusters chose who would be in the treatment group and who would be in the control group. Suppose that the people who showed up first, early in the morning, were assigned to the treatment group, and the last few people, later in the day, ended up in the control group.

- b. Would you believe in the conclusion? Explain.**



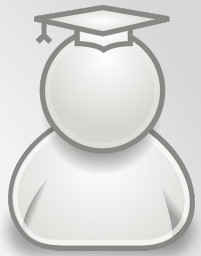


# Discussion

c. In order to make two groups as similar as possible, experimenters usually assign subjects randomly.

**How might we randomly assign about half of the subjects to the treatment group and half to the control?**

d. **Why would random assignment improve the Mythbuster study?**



# Causation

Without **RANDOM ASSIGNMENT** we cannot determine **CAUSATION** because we are not comparing two similar groups.

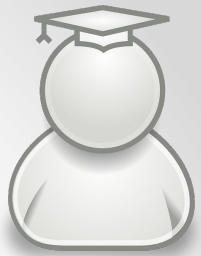
# Lesson 2: What is an Experiment



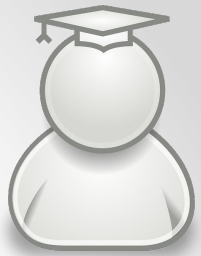
**Objective:** Students will learn about the elements of an experiment and the meaning of “causation”. Student will learn to distinguish claims of causation from claims of association.



# Lesson 3: Let's Try an Experiment!



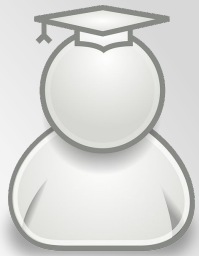
Objective: Students will explore the importance of randomized assignment in experiments. They will understand that without random assignment, there might be confounding variables and will be able to suggest possible confounding variables.



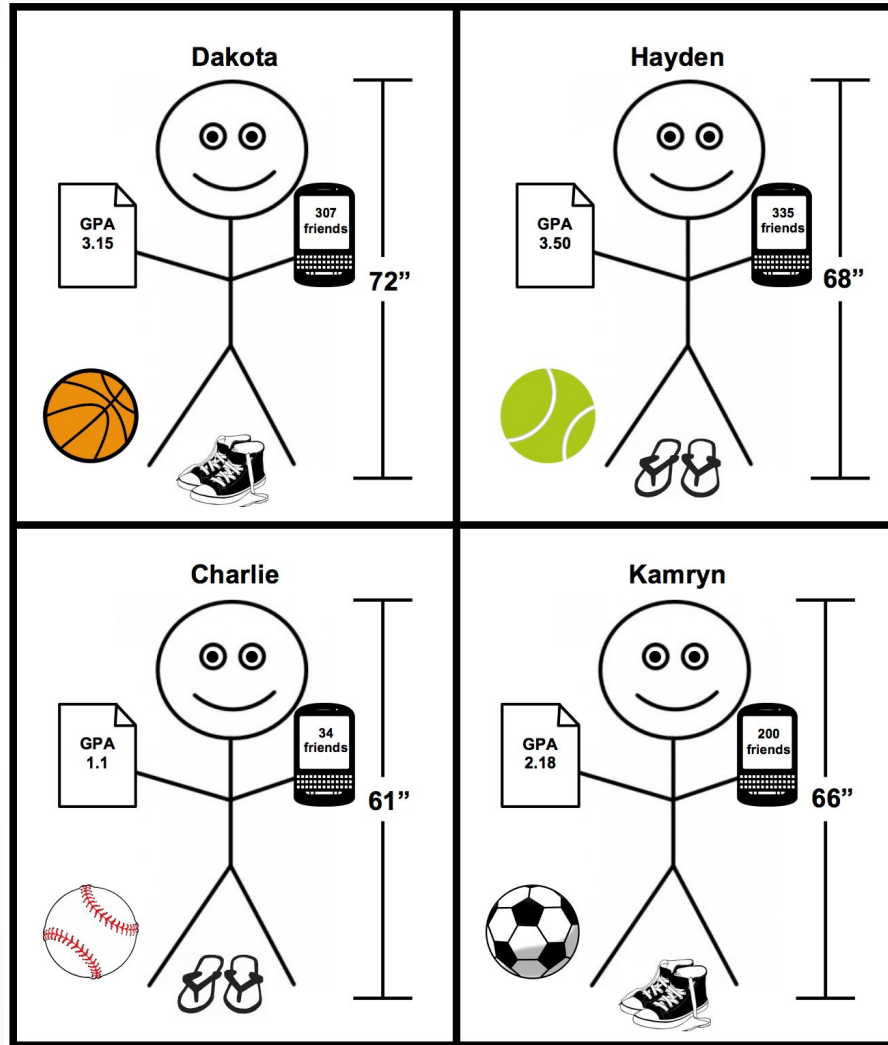
# Stick Figures

Monica's workout vs Heidi's workout

- We want to compare which workout helps the stick figures become more 3-dimensional
- Split the stick figures into two equal numbered groups
- One group will follow Monica's workout and the other group Heidi's workout
- We will measure their "3rd dimension"
- We will take the average of the "3rd dimension"

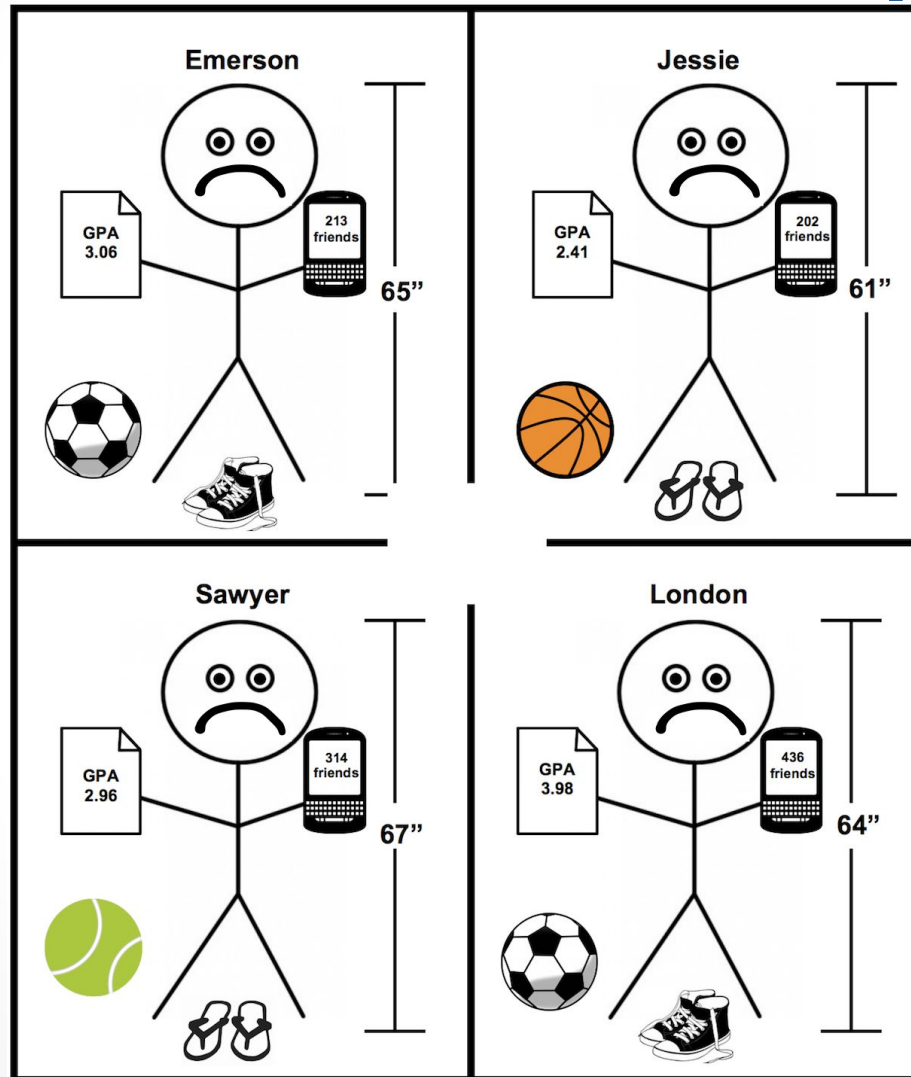


# Monica's Group

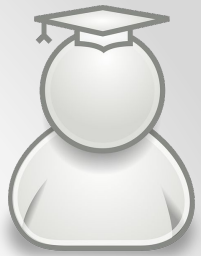




# Heidi's Group







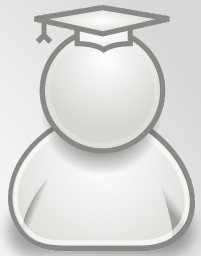
# Results

Monica's group - The average gain in the 3rd dimension was 2.1 cm

Heidi's group - The average gain in the 3rd dimension was 0.1 cm

Is this good evidence that Monica's workout is more effective?

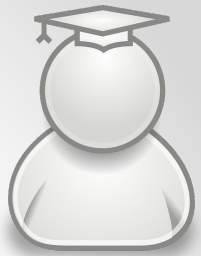
# Experimental Design



Which element was missing from our experiment?

How should the stick figures be put into groups?

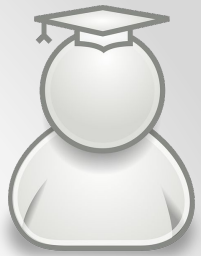
# Let's Try a Well Designed Experiment



**How does our perception of time change when exposed to a stimulus?**

You will be trying to determine the length of one minute without the use of time-aids. They will subject some students to a stimulus and others to no stimulus. They will then analyze the data to determine if subjecting students to a stimulus affects the perception of how long a minute of time lasts.

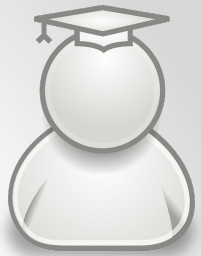
# Elements of the experiment



- a. What is the research question we're interested in addressing?
- b. Who are the subjects that will be participating in the experiment?
- c. How should we randomly assign the subjects into treatment and control groups?
- d. What is the outcome variable that we will be measuring? What unit of measurement should we use?

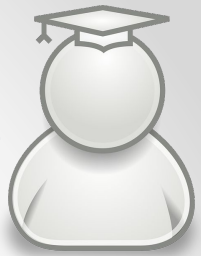


# Lesson 4: Predictions, Predictions



Objective: Students will continue to read articles critically. They will anticipate visualizations about the data that will be collected from the class experiment and make predictions about the outcome.

# Let's Try a Well Designed Experiment



**How does our perception of time change when exposed to a stimulus?**

You will be trying to determine the length of one minute without the use of time-aids. They will subject some students to a stimulus and others to no stimulus. They will then analyze the data to determine if subjecting students to a stimulus affects the perception of how long a minute of time lasts.



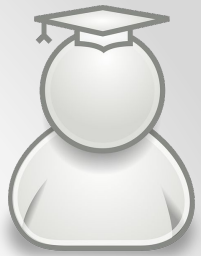
# Treatment

Before break, you designed a class experiment but did not select an actual treatment (stimulus).

To make a decision about what stimulus might have an impact on our perception of time we are going to do some literature review.

[10 Ways Our Minds Warp Time](#)





# Jigsaw

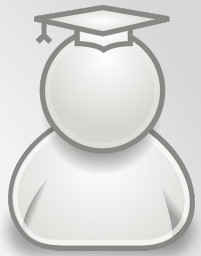
#1 - Read 1. - 2.

#2 - Read 3. - 5.

#3 - Read 6. - 8.

#4 - Read 9. - 10.

Mark the text as you read.



# Decision

You will be trying to determine the length of one minute without the use of time-aids. They will subject some students to a stimulus and others to no stimulus. They will then analyze the data to **determine if subjecting students to a stimulus affects the perception of how long a minute of time lasts.**



# Theory

**How does our perception of time change when exposed to a stimulus?**

Who believes that there will be differences in the estimate of the length of a minute between the treatment and control groups?

How small or large do you think the differences will be?

# Experiment Predictions



Name: \_\_\_\_\_

Date: \_\_\_\_\_

## Experiment Predictions

### Instructions:

Answer the following before conducting your time perception experiment. Remember, the variable that we're measuring is the number of seconds that actually elapse until each person believes one minute has passed.

1. In the boxes below, draw a plot of what you predict the distribution of each group's data will look like. Be sure to add numbers and labels.

Treatment
Control

2. Based on your prediction, write down how the *treatment* group's distribution will compare to the *control* group's in terms of its *center*, *shape* and *spread*.



# Lesson 5: Time Perception

## Experiment

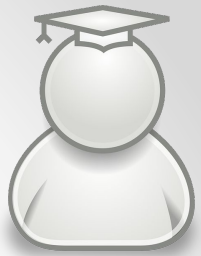


Objective: Students will engage in a collectively designed experiment.

# The stopwatch() function



- Volunteer?
  - Type stopwatch()
  - Click return to start
  - Estimate ONE minute
  - Click return to stop



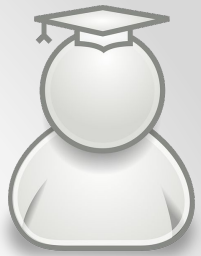
# The Experiment

**How does our perception of time change when exposed to a stimulus?**

- Control Group
  - Sit at desk with eyes closed
  - Run stopwatch()
  - Stop when you believe a minute has passed
- Treatment Group
  - Run stopwatch()
  - Engage in stimulus
  - Stop when you believe a minute has passed



# Time Perception Survey



Record your data from the experiment using the Time Perception Survey

- use the app (UCLA MobilizingCS)

**OR**

- browser-based ([ids.mobilizingcs.org](https://ids.mobilizingcs.org))



# Lab 3A

Create appropriate plots to answer your statistical questions.

Calculate appropriate numerical summaries to answer your statistical questions.

Interpret your plots and summaries.

- Write down a few sentences with your interpretations.

# Shuffling

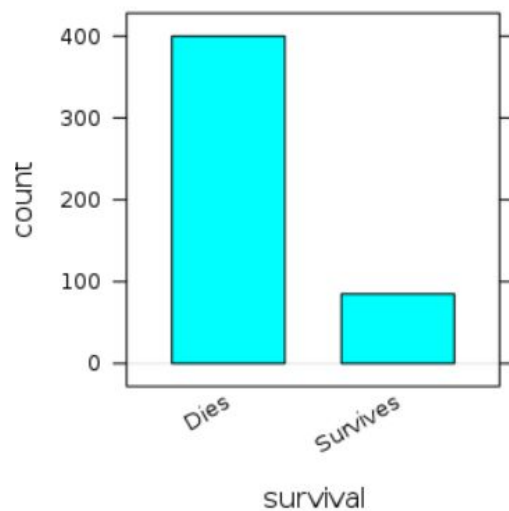
## Wrapping It up

- Is it possible your initial results occurred by chance alone?
  - Use repeated shuffling to determine how likely the typical difference between the two groups occurred by chance alone.
  - Create a plot and use it to justify your answer.

# Slasher Data

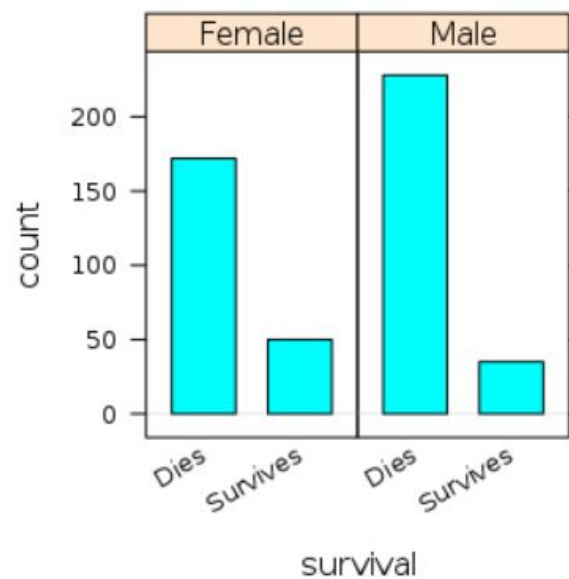
Dies  
0.82

Survives  
0.18



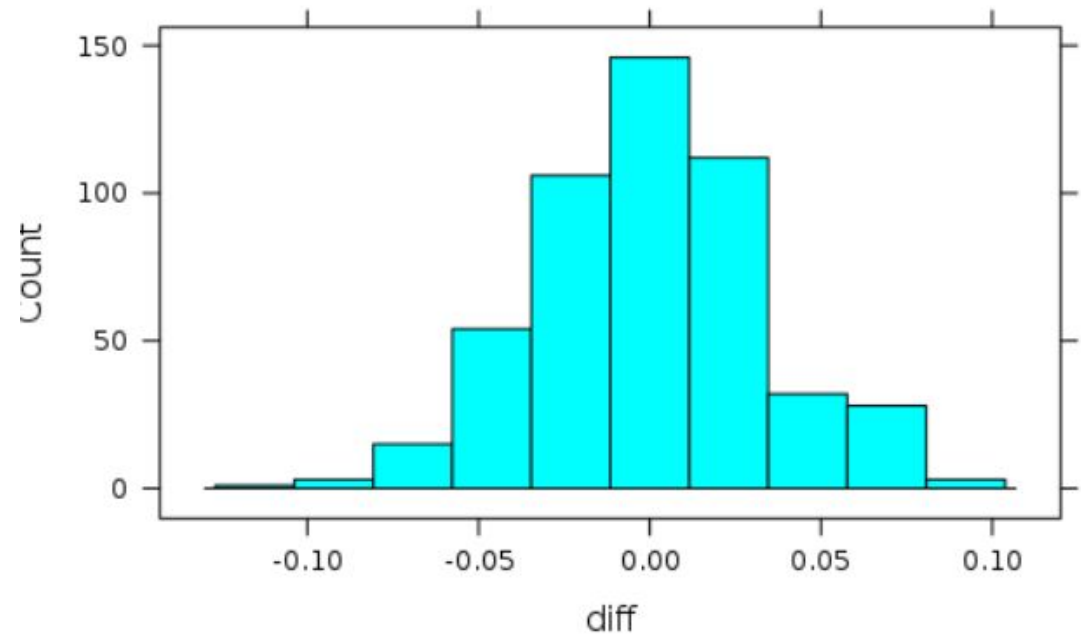
0.77 0.23

0.87 0.13



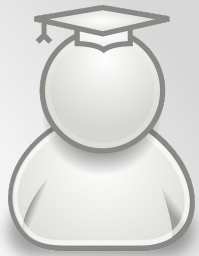
# Conclusion

Survives.Female <sup>^</sup>	Survives.Male <sup>^</sup>
0.1576577	0.1901141
0.1756757	0.1749049
0.1576577	0.1901141
0.1756757	0.1749049
0.1621622	0.1863118
0.1576577	0.1901141
0.1621622	0.1863118
0.1846847	0.1673004
0.1711712	0.1787072
0.1846847	0.1673004





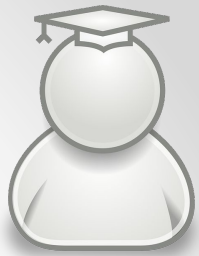
# Lesson 7: Observational Studies vs. Experiments



Objective: Students will learn how observational studies differ from experiments and will classify different research scenarios based on which method would be most appropriate. They will also learn about the roles of ethics, cost limitations, and feasibility when deciding between the two data collection methods.



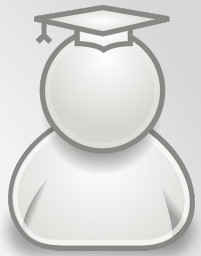
# Observational Studies vs. Experiments



Why can't we always do experiments?

REMEMBER:

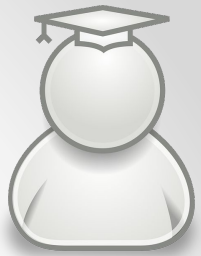
- Observational studies - we cannot randomly assign into control/treatment groups
- Experiments - we **must** randomly assign into these groups



# Experiment... or not

Possible research questions:

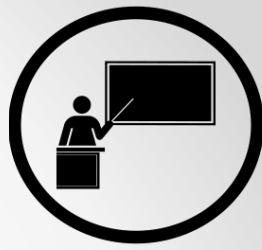
1. Does smoking cause lung cancer?
2. Does drinking water from Mars keep you healthier than drinking water from Earth?
3. Do people with higher IQ scores score better on the SAT than people with lower IQ scores?



# What should we do?

Create a poster:

1. Each team assigned a research question
2. Decide best data collection method
3. Explain why that's the best method



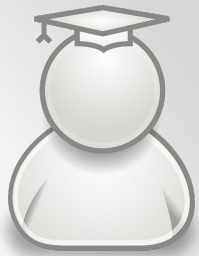
# Exit Ticket

On a post it, create 2 research questions:

- 1 that requires an observational study
- 1 that requires an experiment
- Explain why each data collection method is most appropriate



# Lesson 8: Monsters that Hide in Observational Studies



Objective: Students will learn about confounding factors that may impact the results of an observational study, which is why causation can never be concluded with observational studies, only associations between variables.

# Revisiting Stick Figures

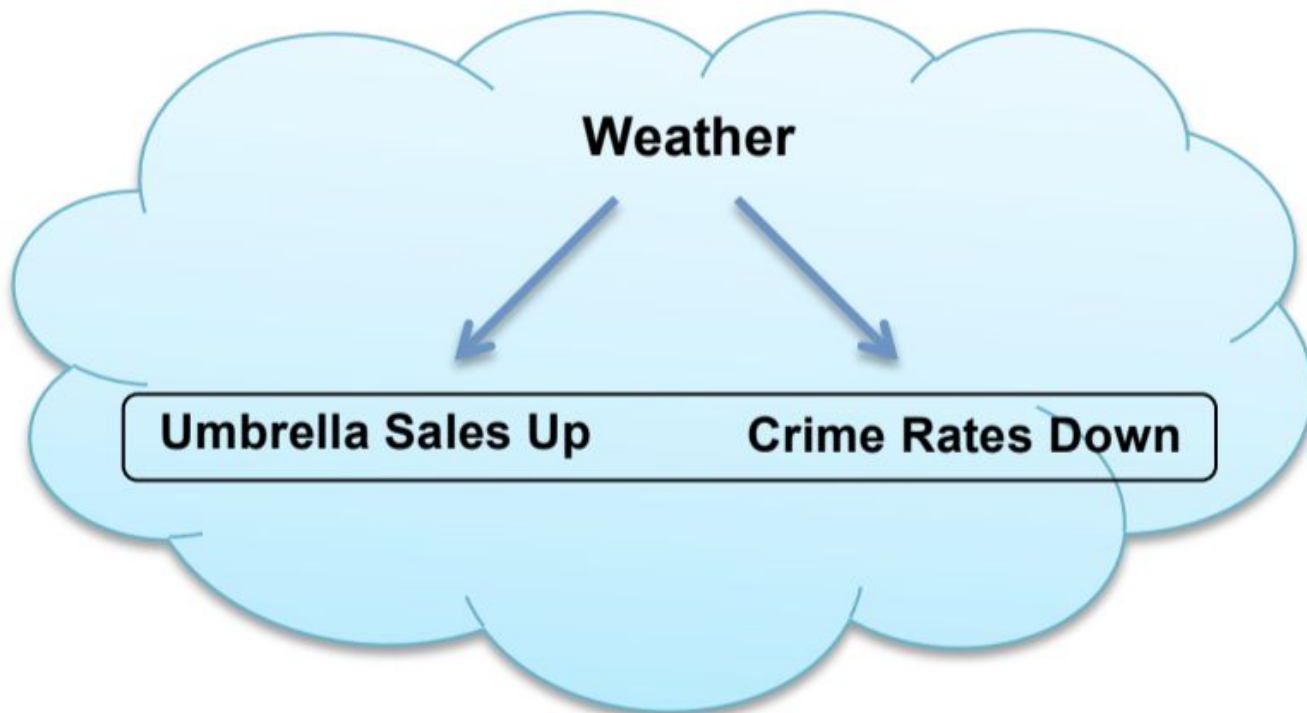


- Does the amount of friends **cause** a person's GPA?
  - What were some of the possible **confounding factors** that contributed to this relationship?
    - remember: confounding factors are variables that are related to both the explanatory variable and the response variable in an observational study.

# The Impact of Confounding Variables



Research suggests that a rise in umbrella sales leads to decreased crime rates

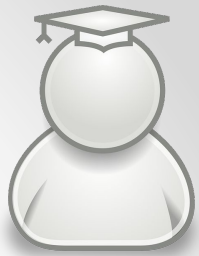


Un

n



# Spurious Correlations



Go to:

<http://www.tylervigen.com/spurious-correlations>

- A. What are the two variables shown in your graph?
- B. Is there a positive association or a negative association between the variables?
- C. Write an interpretation of this plot in the context of the data.
- D. Write the data points in a "spreadsheet format" in a form that RStudio could read. Each row should represent a point on the graph, and each column one of the two variables.
- E. By hand, make a scatterplot of the association. Describe whether the association seems strong or weak or moderate to you.
- F. Do you think that the explanatory variable causes the response variable? Explain.
- G. If you answered 'no' to f, then draw a diagram like in the previous slide (weather/umbrellas/crime) with possible confounding factors.



# Class Scribes

1. Share your graph with your team.
2. Discuss and decide the top 3 topics or understandings from today - be prepared to share



# Lab 3B

- In this lab, we'll read in an *observational study* dataset from a website.
  - We'll use this data to then explore what factors are associated with a person's lung capacity.

*What is the effect of childhood smoking on lung health.*

*Do people who smoke tend to have lower lung capacity than those who do not smoke?*

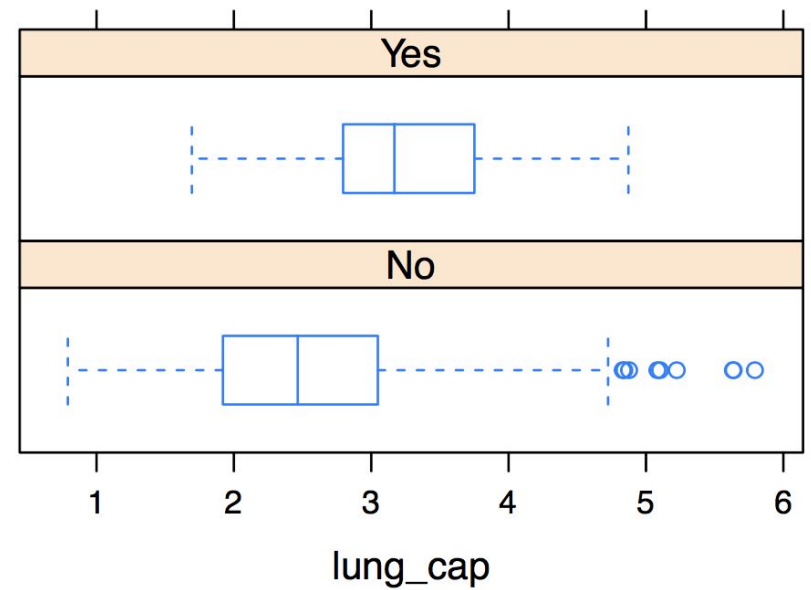
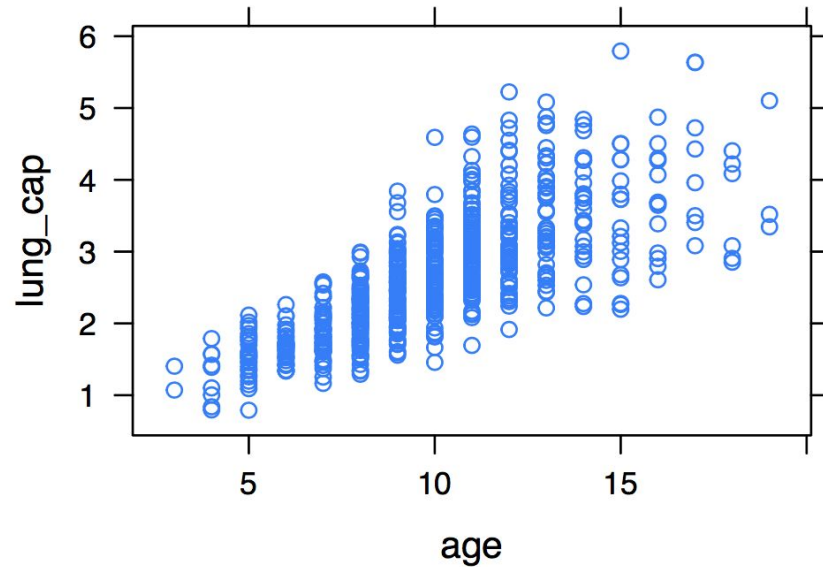
# Clean the Data

- Now that we've got the data loaded, we need to clean it to get it ready for use (*Look at lab 1F for help*). Specifically:
  - We want to name the variables: "age", "lung\_cap", "height", "gender", "smoker", in that order.
  - Change the type of variable for `gender` and `smoker` from *numeric* to *character*.
- After changing the variable types for `gender` and `smoker`:
  - For `gender`, use `revalue` to change "1" to "Male" and "0" to "Female".
  - For `smoker`, use `revalue` to change "1" to "Yes" and "0" to "No".

# Analyze the Data

- Our `lungs` data is from an observational study.
- **Write down a reason the researchers couldn't use an experiment to test the effects of smoking on children's lungs.**
- Observational studies are often helpful for analyzing how variables are related:
  - **Do you think that a person's age affects their lung capacity? Make a sketch of what you think a scatterplot of the two variables would look like and explain.**
- Use the `lungs` data to create an `xyplo`t of `age` and `lung_cap`.
  - **Interpret the plot and describe why the relationship between the two variables makes sense.**

# Visualizations







# Performance Task:

## Observations vs. Experiments

**Objective:** A study of the last 20 years showed that the number of shark attacks on swimmers at the beach was significantly higher in months when sunscreen sales at the beach snack bar were high. A city councilman theorizing that great and identify any confounding variables and how white sharks are attracted to sunscreen, they change the conclusions. They will make advocated a law to ban the use of sunscreen on public beaches. **Scenario:** students are expected to identify the key aspects that make a study an experiment or an observational study. They will analyze the situation and identify any confounding variables and how white sharks are attracted to sunscreen, they change the conclusions. They will make advocated a law to ban the use of sunscreen on public beaches.



# Next Steps

- Office Hours:
  - Thursday 11/2/17 3:00-4:30pm
- Unit 3 (Part 2) & Unit 4 Follow-Up dates:
  - W 11/15/17, Tu 2/20/18, W 3/14/18
- Unit 3 Assessments & Lab Solutions will be posted in Documents Tool at next PD

# PD Evaluation - Gots & Needs

<https://tinyurl.com/IDS-PD-2017-18>

