

# Stats 401 Lab 1

Naomi Giertych

1/2/2018

# Welcome!

- ▶ GSI: Naomi Giertych
- ▶ Email: [ngierty@umich.edu](mailto:ngierty@umich.edu)
- ▶ Office Hours: TBD

## Lab Objectives

- ▶ Develop statistical computation skills (R) and mathematical skills
- ▶ Provide examples and exercises to help prepare you for homework, quiz, and exam questions

# Lab Expectations and Requirements

- ▶ (Printed) Homework is due at the start of class
- ▶ If you need to make up a lab: Please email me and make-up GSI in advance

# What will lab look like?

- ▶ Lab will consist of the following:
  1. Brief review of what was covered in lecture the previous week
  2. A worked example (this may be combined with 1 if reasonable)
  3. An in lab exercise
  4. Lab ticket to complete and hand in (for attendance)
- ▶ If time permits, I'm willing to guide you in homework. However, I will not directly provide answers.

Questions?

## About You!

- ▶ Raise your hand if you've programmed in any language before

## About You!

- ▶ Raise your hand if you've programmed in any language before
- ▶ Raise your hand if you've opened R Studio before this class



## About You!

- ▶ Raise your hand if you've programmed in any language before
- ▶ Raise your hand if you've opened R Studio before this class
- ▶ Raise your hand if you know what a vector is

## Questions?

- ▶ Has everyone been able to install R and R Studio successfully?

## Worked Example Using Swirl

```
# Install Swirl  
install.packages('swirl')  
  
# Call packages  
  
library(swirl)  
  
# Use Swirl  
swirl()  
  
# Navigate to Lesson 2 by selecting 1,  
# then selecting 1 again.
```

## Worked Example Using A New Dataset

```
# Install and call required packages  
# note: I already have faraway  
# installed so I've commented it out  
# install.packages('faraway', dependencies = T)  
library(faraway)  
  
# Load Dataset  
data("uswages")  
  
# What if you want to know what all the variables in  
# this dataset are?  
?uswages
```

```
# Let's find some specific values in the dataset
```

```
# What's the wage of the 1500th observation?
```

```
# Method 1:
```

```
uswages[1500,1]
```

```
## [1] 686.04
```

```
# Method 2
```

```
uswages$wage[1500]
```

```
## [1] 686.04
```

```
# What's the education of the 13th observation?  
# Method 1  
uswages[13,2]
```

```
## [1] 17
```

```
# Method 2  
uswages$educ[13]
```

```
## [1] 17
```

```
# What's the maximum wage found in the dataset?  
max(uswages$wage)
```

```
## [1] 7716.05
```

```
# What's the difference between the wage in the 13th  
# observation and the max?
```

```
# Method 1
```

```
uswages$wage[13]
```

```
## [1] 550.81
```

```
# Difference
```

```
7716.05 - 550.81
```

```
## [1] 7165.24
```

```
# Method 2  
# How do I save a variable?  
wage_13 <- uswages$wage[13]  
max_wage <- max(uswages$wage)
```

```
# Difference  
max_wage - wage_13
```

```
## [1] 7165.24
```



## In Lab Activity

```
# Still using the faraway library
```

```
# Use the SAT dataset
```

```
data("sat")
```

```
# What was the highest total SAT score?
```

```
# What was the lowest total SAT score?
```

```
# Calculate the difference in scores.
```

```
# Figuring out the median ratio of the last 15 states
```

```
# alphabetically
```

```
median(sat$ratio[36:50])
```

```
## [1] 15.7
```

## Key Takeaways

- ▶ Learning a new programming language can be frustrating and fun (sometimes at the same time)
- ▶ Commenting code is **EXTREMELY** important. Future you will thank you!

## Lab Ticket

- ▶ Using the salary dataset from the lab activity, find the following:
  1. What is the median salary?
  2. What is the mean salary?
  3. What do these numbers suggest about the distribution of salaries?