

# **STATS 250 Lab 01**

## **Getting Started with R**

Nick Seewald

[nseewald@umich.edu](mailto:nseewald@umich.edu)

Week of 08/31/2020

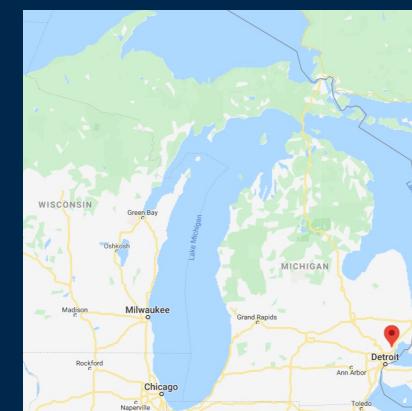
# Who am I? 🧐



Translate this sentence

There is a cow in the  
house!

Il y a une vache dans la maison!



# Who are you?

While you are taking this class, you are a statistician!

 [bit.ly/250labintros](https://bit.ly/250labintros)

See my answers at [bit.ly/nicksurveyanswers](https://bit.ly/nicksurveyanswers)  
(please don't hit submit on my answers!)

# Course Policy

## Grading

Component	Proportion of grade
Exams	30% ( $15\% \times 2$ )
Quizzes	20% ( $10\% \times 2$ )
Homework	20% (12 total, 2 dropped)
MWrite	15% ( $5\% + 10\%$ )
Lab Work	10% (12 total, 1 dropped)
Participation	5% (1% eCoach + 1% Survey + 3% Piazza)

- Letter grades are assigned based on a fixed scale, there is no curve. See syllabus.
- Grades will be traditional A-C then NRC for D-E

# Course Policy

## Lab attendance

- Every week, you have the option of attending either of my synchronous sessions **or** you can complete the lab asynchronously.
  - **Section 108:** Monday 10:00a - 11:30a Eastern
  - **Section 106:** Monday 1:00p - 2:30p Eastern
  - **Section 115:** Asynchronous
- I'd appreciate you telling me your plans each week before lab each week; I'll be sending out a survey on Fridays for the coming week.

# Course Policy

## Office Hours

- You can attend anybody's office hours you want!
  - Full schedule on Canvas
- Completely remote via Zoom (links on Canvas)
- I cannot confirm/deny whether your answers are "right"

## My office hours:

Day	Time (ET)
Tuesday	9:00a - 10:00a
Wednesday	11:00a - 12:00p
Thursday	5:30p - 7:00p
Friday	9:30a - 10:30a

# Course Policy

## Flexibility

- This is going to be an unusual semester.
- Things like the office hours schedule and the schedule of topics **may change**.
- Please try to be flexible with us as we adapt to this new learning environment, as we'll be flexible with you.



# Lab Rules and Responsibilities

## 1. Be Respectful.

- Respect to me includes watching the entirety of these videos, communicating in good faith, and filling out attendance mode forms each week.
- Treat your classmates with dignity and *learn from your differences*
- Respect yourself: *be ready and willing to learn*

# Lab Rules and Responsibilities

## 1. Be Respectful.

- Respect to me includes watching the entirety of these videos, communicating in good faith, and filling out attendance mode forms each week.
- Treat your classmates with dignity and *learn from your differences*
- Respect yourself: *be ready and willing to learn*

## 2. Actively Collaborate.

- Lab activities will involve working with your peers: you should participate and learn from each other.
- In async weeks, **use Piazza**
- Work with friends in other labs (within Dr. Miller's course); your write-ups must be your own.

# Health and Well-Being

College is hard (**especially now**). Your health and well-being are more important than this class.

If you're struggling with mental health, there are resources available to you at [wellbeing.studentlife.umich.edu](http://wellbeing.studentlife.umich.edu).

CAPS: [caps.umich.edu](http://caps.umich.edu) or 734.764.8312

*If you become ill, please report the illness to LSA:*

<https://webapps.lsa.umich.edu/SAA/UGStuAdv/App/Illness/RptIll.aspx>

**Let me know what's going on and I'll do my best to help.**

# Diversity, Equity, and Inclusion

It's important to me to cultivate a learning environment in which everyone's identities (race, ethnicity, gender, sexual orientation, religion, etc.) are respected. To this end:

- If you feel that your performance in the class is being affected by experiences inside **or** outside the classroom, please let me know.
- DEI is a process: I (and many others) am still learning about different perspectives and identities. If anything is said in class (by anyone, including by me) that made you feel uncomfortable, please let me know.

# Is this a math class?

Statistics is a *mathematical science*. This course will involve elements of a typical math course, but also less typical aspects.

# Is this a math class?

Statistics is a *mathematical science*. This course will involve elements of a typical math course, but also less typical aspects.- Often, you will be required to apply statistical knowledge *and* everyday knowledge ("common sense") to solve problems.

- In statistics, there is not always a "correct" answer. You should be prepared to argue (in words) for why your answer is reasonable.
- Written components of assignments are critical.

Adapted from Martha K. Smith, University of Texas at Austin

# Is this a math class?

Statistics is a *mathematical science*. This course will involve elements of a typical math course, but also less typical aspects.- Often, you will be required to apply statistical knowledge *and* everyday knowledge ("common sense") to solve problems.

- In statistics, there is not always a "correct" answer. You should be prepared to argue (in words) for why your answer is reasonable.
- Written components of assignments are critical.

**TL;DR** This is NOT a math class. So if you're "not a math person", don't worry.

Adapted from Martha K. Smith, University of Texas at Austin

# Let's get started!

# Learning Objectives

## Statistical Learning Objectives

1. Visualize categorical data
2. Summarize quantitative and categorical data

## R Learning Objectives

1. Learn the difference between R, R Studio, and R Markdown
2. Become familiar with the R Studio interface
3. Understand key components of an R Markdown document
4. Become familiar with R functions and arguments

# How does lab work? 🧐

Three components each week:

1. Logistics & Demo
2. Try It!
3. Discussion Questions

You'll work through the Try It! and Discussion, and submit an "R Markdown" report.

# Getting Started: What is R?

- Statisticians (that means you!) often use computers to analyze data.
- R is a popular (and powerful) program for manipulating data, doing calculations, and making graphs.
- R works by writing R code.

# Getting Started: What is R?

- Statisticians (that means you!) often use computers to analyze data.
- R is a popular (and powerful) program for manipulating data, doing calculations, and making graphs.
- R works by writing R code.

Wait omg is this a coding class???

# Getting Started: What is R?

- Statisticians (that means you!) often use computers to analyze data.
- R is a popular (and powerful) program for manipulating data, doing calculations, and making graphs.
- R works by writing R code.

Wait omg is this a coding class???



Over the course of the semester, you'll learn how to edit and write some basic R code to help you analyze data to answer research questions. This will always be through the lens of answering statistical questions.

# Time for RStudio Cloud!

**<http://bit.ly/250millerfa20rsc>**

*Log in with Google using your U-M Account (non-umich accounts will be deleted without warning)*



## Spaces

Your Workspace

Stats 250 - Miller - Introduct...

Stats 250 - Romero - Introduct...

New Space

## Learn

Guide

What's New

Primers

Cheat Sheets

## Help

Current System Status

RStudio Community

## Info

## All Projects

New Project ▼Options X

START

lab01-FA20

Nick Seewald

Created Aug 28, 2020 11:15 AM

Search Projects X

List Projects

 All Shared with everyone Yours

Sort Projects

 By name By date createdX

# RStudio Desktop

Follow instructions to get lab materials at

**[github.com/STATS250SBI/lab01-fa20](https://github.com/STATS250SBI/lab01-fa20)**

(Link is also on Canvas in the Lab 1 Assignment)

# Using R as a Calculator

At its most basic, R is a fancy calculator.

Here's a code chunk in which we do some math:

```
```{r calculatorExample}
5 * 7
```
```

Output:

```
5 * 7
```

```
[1] 35
```

# Try it!

In your R Markdown document, edit the code chunk that looks like this (line ~79) to compute 50 divided by 9:

```
```{r tryItCalculator, error = T}
# Write code here! This text is a comment, so it's not run by R. Replace it with
# code to compute 50 divided by 9.
````
```

# Try it!

In your R Markdown document, edit the code chunk that looks like this (line ~79) to compute 50 divided by 9:

```
```{r tryItCalculator, error = T}
# Write code here! This text is a comment, so it's not run by R. Replace it with
# code to compute 50 divided by 9.
````
```

```
50 / 9
```

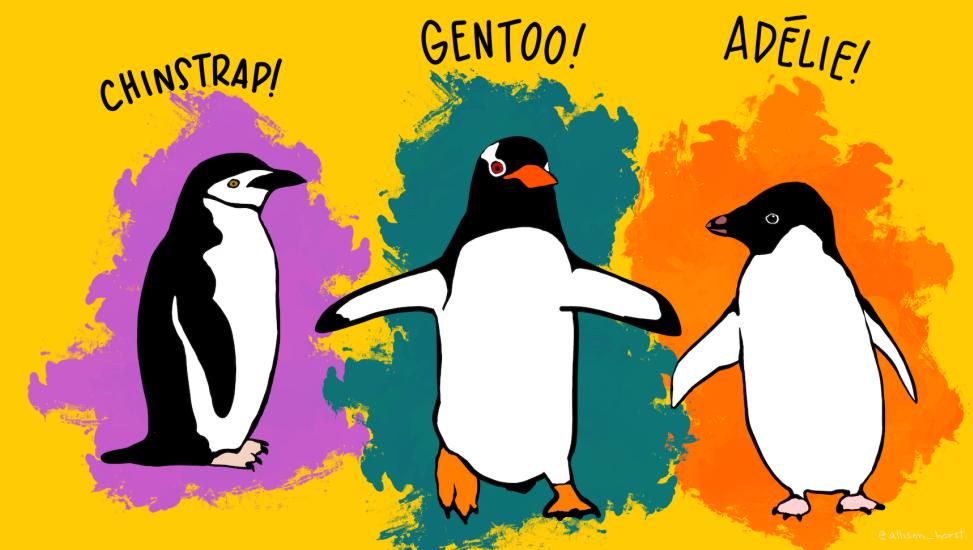
```
[1] 5.555556
```

# Palmer Penguins Data



We're going to learn about some basic R functions using a data set on 333 penguins living on 3 islands in the Palmer Archipelago in Antarctica.

Data were made available by [Dr. Kristen Gorman](#) and the [Palmer Station](#), [Antarctica Long Term Ecological Research area](#), a member of the Long Term Ecological Research Network. The data were prepared by [Dr. Allison Horst](#)



# First Steps

## "Read in" data

```
penguins <- read.csv(url("https://raw.githubusercontent.com/STATS250SBI/palmerpenguins/master/inst/e
```

- You'll learn more about this next week
- For now, the important thing to know is that we've told R we want to use a data set at that URL and we want to call it **penguins**.

# First Steps

## Peek at the first 6 rows of the data

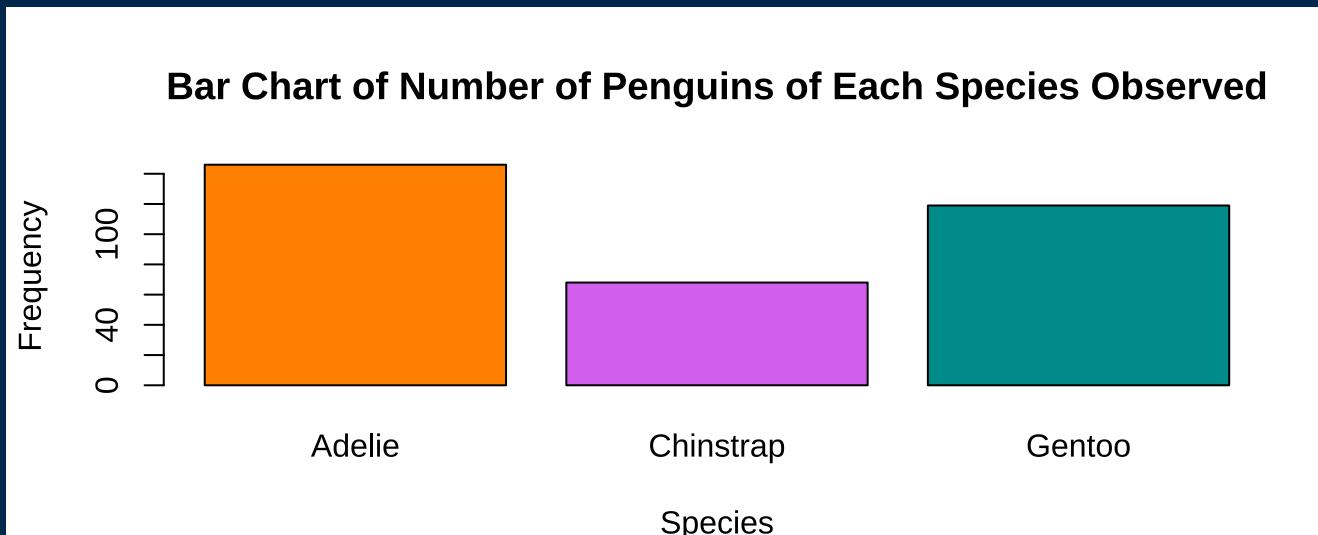
```
head(penguins)
```

```
species      island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
1   Adelie Torgersen      39.1          18.7            181        3750
2   Adelie Torgersen      39.5          17.4            186        3800
3   Adelie Torgersen      40.3          18.0            195        3250
4   Adelie Torgersen      36.7          19.3            193        3450
5   Adelie Torgersen      39.3          20.6            190        3650
6   Adelie Torgersen      38.9          17.8            181        3625
sex year
1   male 2007
2 female 2007
3  female 2007
4  female 2007
5   male 2007
6 female 2007
```

# Bar Charts in R

Let's make a bar chart of the number of penguins observed in each species:

```
barplot(table(penguins$species),  
       xlab = "Species",  
       ylab = "Frequency",  
       main = "Bar Chart of Number of Penguins of Each Species Observed",  
       col = c("darkorange1", "mediumorchid2", "darkcyan"))
```



# Numerical Summaries

We can use R to summarize data numerically. A good way to do this is the **summary()** function.

```
summary(penguins$flipper_length_mm)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|------|---------|------|
| 172  | 190     | 197    | 201  | 213     | 231  |

# Now it's your turn!



# Lab Project



## Your tasks

- Complete the "Try It!" and "Discussion" portions of the lab assignment by copy/pasting and modifying appropriate code from earlier in the document.

## How to get help

- Use the "labs" section of Piazza to ask questions and work with your peers.
- If you use Piazza, please note that in the "Collaborators" list at the top of the discussion section.
- If you're really stuck, email me!  
[nseewald@umich.edu](mailto:nseewald@umich.edu)



# Lab Submission: Finding Your Report

Hit the Knit button one last time, then:

## RStudio Cloud

1. In the Files pane, check the box next to `lab01report.html`
2. Click More > Export...
3. Click Download and save the file on your computer in a folder you'll remember and be able to find later.

## RStudio Desktop (local)

1. Locate the `lab01 report.html` file on your computer. The file will be saved in the location indicated at the top of the files pane.



# Lab Submission: Canvas (Due 9/11 8a ET)

1. Click the "Assignments" panel on the left side of the page. Scroll to find "Lab 1", and open the assignment. Click "Submit Assignment".
2. Towards the bottom of the page, you'll be able to choose `lab01report.html` from the folder you saved it in from RStudio Cloud or noted if you're using RStudio Desktop. **You will only be able to upload a .html file -- do not upload any other file type.**
3. Click "Submit Assignment". You're done!