

# STATS 250 Lab 01

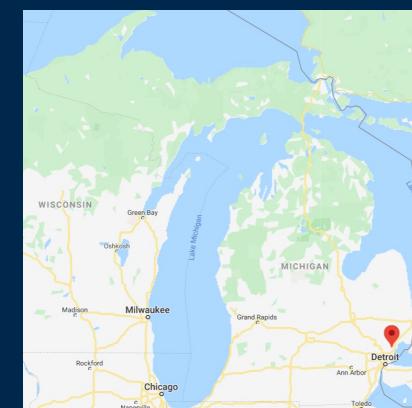
## Getting Started with R

Nick Seewald  
[nseewald+250@umich.edu](mailto:nseewald+250@umich.edu)

Say hi in the chat!

Week of 1/25/2021

# Who am I? 🧐



# Who are you?

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

1. Change your Zoom name to include pronouns

- Right-click (or long press) your name in the Participants list, select "Rename"

2. Add a photo to your Zoom profile

- Go to [umich.zoom.us](https://umich.zoom.us), sign in, and choose "Profile" on the left
- This will be used across all your U-M Zoom meetings

3. In the chat, type your favorite TV show **but don't hit enter it until I say go.**

# Course Policy

## Lab attendance

- It's expected that you attend these synchronous lab sessions and participate in the collaborative breakout sessions.
  - Please do not leave when breakout rooms open!
- If you can't make it to a synchronous session in a particular week, you *must* email me to get asynchronous materials.

# Course Policy

## How to contact me

- Email me! The best address to do this is [nseewald+250@umich.edu](mailto:nseewald+250@umich.edu)
  - Make sure you put "STATS 250" in the subject line
  - If you don't put "STATS 250" in the subject OR email [nseewald+250@umich.edu](mailto:nseewald+250@umich.edu),  
*your email might get lost.*
- I will try to respond to all emails within 24 hours
  - Generally my email hours are 9a - 7p ET on weekdays

# 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

# Lab Rules and Responsibilities

## 1. Be Respectful.

- Come to synchronous sessions! Notify me of conflicts in advance.
- Treat your classmates with dignity and *learn from your differences*
- Respect yourself: *come here to learn*

# Lab Rules and Responsibilities

## 1. Be Respectful.

- Come to synchronous sessions! Notify me of conflicts in advance.
- Treat your classmates with dignity and *learn from your differences*
- Respect yourself: *come here to learn*

## 2. Actively Collaborate.

- Lab activities will involve working with your peers: you should participate and learn from each other.
- The time will pass faster if you're having conversations and working together!

# 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.

- 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.
- Statistics requires **context**: you need to tell me what numbers mean and why we care.  
*No context = no points.*

Adapted from [Martha K. Smith, University of Texas at Austin](#)

# Lab Goals and Structure

- **Overall goal:** Give you hands-on experience working with data in **R**, a powerful statistical computing tool.
  - You'll complete weekly lab assignments using **R Markdown** via **RStudio Cloud**.

## Typical Lab Structure

- **Tutorial** (~30 minutes): We'll walk through how to perform tasks in R together
  - iClicker engagement questions sprinkled throughout (not graded, just for fun)
- **Collaboration** (~40 minutes): You'll work on the lab assignment in breakout rooms
- **Wrap-Up** (~10 minutes): We'll talk about any questions or issues and complete a brief reflection.

# Let's get started!

# Learning Objectives

1. Learn the difference between R, RStudio, and R Markdown
2. Become familiar with the RStudio interface
3. Understand key components of an R Markdown document
4. Learn how to use R to do basic calculations
5. Learn how to import data into R and name objects

# 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/250sbiwn2lrsc>**

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

# What are all these "R" terms?

There are a lot of "R" words floating around. What's going on?

- **R** is a "statistical computing environment" that's designed for manipulating data, generating plots, and performing analyses. It's also a programming language. You'll be *using R* this semester.
- **R Studio** is an "integrated development environment (IDE)" for R (you'll never have to hear the term IDE again in this class). Basically, it's a pretty interface that makes working with R easier. You use R inside of R Studio. If R is ice cream, R Studio is the cone or cup.
- **R Markdown** is a way to write pretty analysis reports that combines R code, R output (plots, analysis results, etc.) and text in one document. This lab document is an R Markdown report!

# R Markdown

- Document creation interface
- Integrate text with R code, R output, and graphs
- Type code in **chunks**:

```
```{r chunkName, <chunk options>}
# Type code here
```

```
```
```

# 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:

```
[1] 35
```

# Try it! (Line ~87)

In your R Markdown document, edit the code chunk that looks like this (line ~87) 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! (Line ~87)

In your R Markdown document, edit the code chunk that looks like this (line ~87) 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
```

# R Functions

- We'll be using R primarily by using *functions*.
  - These work just like math functions: they take inputs and return outputs

```
function.name(argument1, argument2, ...)
```

- Arguments to functions go **inside** parentheses and are separated by commas
- **Finish what you started:** make sure you close parentheses you open!
- Don't put a space between the name of the function and the open parenthesis

# R "Assignment" Syntax (line ~104)

- Often, we want R to remember the results of a calculation so we can use it later.
- We can **assign** these results a name:

```
x <- 36 / 6
```

- Read **<-** as "gets": "x gets 36 divided by 6"

# R "Assignment" Syntax (line ~111)

Now we can ask R what **x** is:

```
x
```

```
[1] 6
```

**NOTE:** R is "case-sensitive", which means that upper-case letters are *different than* lower-case letters. Notice what happens when we ask R for the value of **X**:

```
X
```

```
Error in eval(expr, envir, enclos): object 'X' not found
```

# "Valid" Names in R

- Names can contain only letters, numbers, periods (.) and underscores (\_).
- Must start with a letter or period.

```
tik tok <- 12
```

```
Error: <text>:1:5: unexpected symbol
1: tik tok
   ^
```

```
tiktok <- 12
```



# "Valid" Names in R

- Names can contain only letters, numbers, periods (.) and underscores (\_).
- Must start with a letter or period.

```
tik tok <- 12
```

Error: <text>:1:5: unexpected symbol  
1: tik tok  
  ^

```
tiktok <- 12
```



```
4eva <- 4 * 2
```

Error: <text>:1:1: unexpected input  
1: 4ev  
  ^

```
forever <- 4 * 2
```



# CSV Files

- A common way to store data
- CSV = "Comma Separated Values"

```
"species","island","bill_length_mm","bill_depth_mm","flipper_length_mm","body_mass_g",  
"Adelie","Torgersen",39.1,18.7,181,3750,"male",2007  
"Adelie","Torgersen",39.5,17.4,186,3800,"female",2007  
"Adelie","Torgersen",40.3,18,195,3250,"female",2007  
"Adelie","Torgersen",36.7,19.3,193,3450,"female",2007  
"Adelie","Torgersen",39.3,20.6,190,3650,"male",2007  
"Adelie","Torgersen",38.9,17.8,181,3625,"female",2007
```

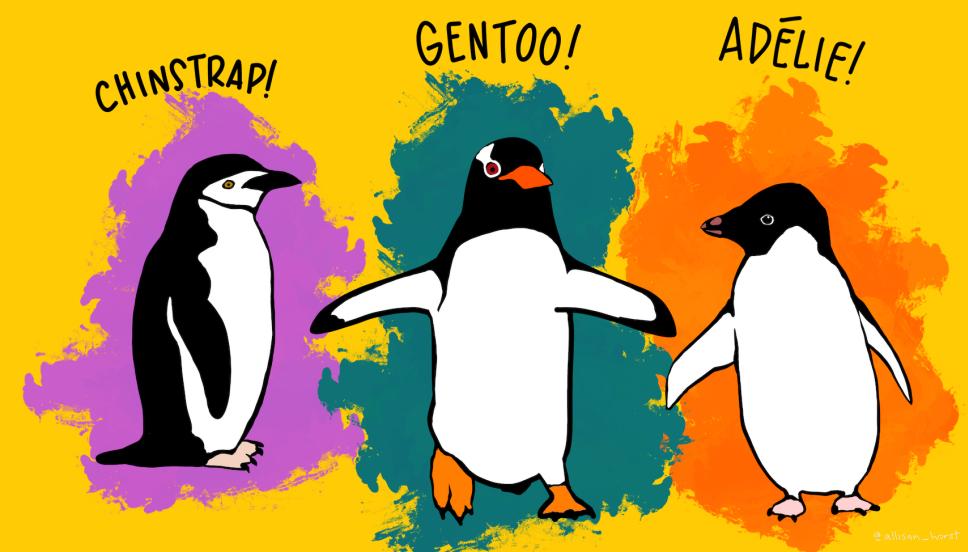
- Each row is a *case* or *observation*
- Each column is a *variable*

# 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](#)



# "Read in" data

```
penguins <- read.csv("penguins.csv", stringsAsFactors = TRUE)
```

- The `read.csv` function takes a CSV file and imports it into R.
- What name did we give the data in R?
- Don't worry about `stringsAsFactors = FALSE` right now.

# 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      |                |               |                   |             |

# View the structure of the data

```
str(penguins)
```

```
'data.frame': 333 obs. of 8 variables:  
 $ species       : Factor w/ 3 levels "Adelie","Chinstrap",...: 1 1 1 1 1 1 1 1 1 ...  
 $ island        : Factor w/ 3 levels "Biscoe","Dream",...: 3 3 3 3 3 3 3 3 3 3 ...  
 $ bill_length_mm: num  39.1 39.5 40.3 36.7 39.3 38.9 39.2 41.1 38.6 34.6 ...  
 $ bill_depth_mm : num  18.7 17.4 18 19.3 20.6 17.8 19.6 17.6 21.2 21.1 ...  
 $ flipper_length_mm: int  181 186 195 193 190 181 195 182 191 198 ...  
 $ body_mass_g   : int  3750 3800 3250 3450 3650 3625 4675 3200 3800 4400 ...  
 $ sex           : Factor w/ 2 levels "female","male": 2 1 1 1 2 1 2 1 2 2 ...  
 $ year          : int  2007 2007 2007 2007 2007 2007 2007 2007 2007 2007 ...
```

- Read this code as "stir of penguins"

# Now it's your turn!



# Lab Project



- You will be moved to a breakout room for the rest of the lab (minus ~10 minutes)
- Introduce yourself to your collaborators!
  - **The person whose birthday is next introduces themselves first**

## Your task

- Open "lab01-report.Rmd" from the Files pane
- Answer each question

```
::::{.answer}  
:::
```

## How to get help

- I'll be floating around between breakout rooms to check on everyone
- Use the "Ask for help" button to flag me down
- Let me know when you're done



# 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 `lab01-report.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.

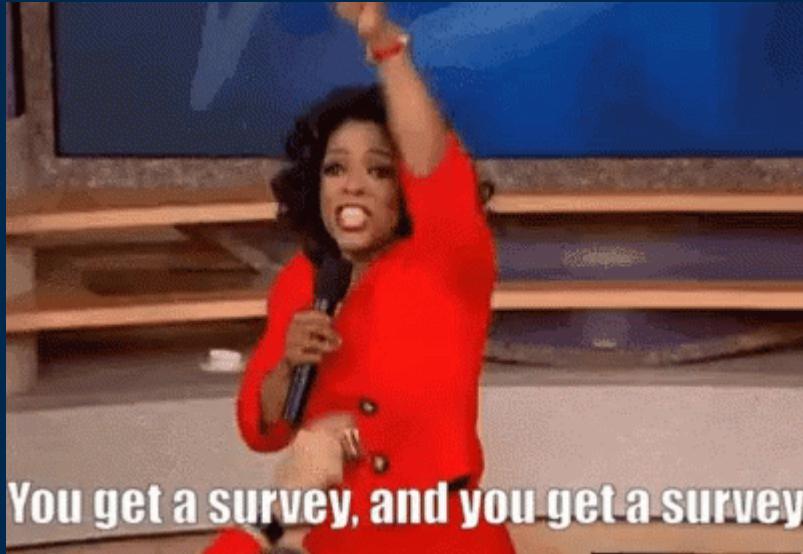


# Lab Submission: Canvas (Due 1/29 8a ET)

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 `lab01-report.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!



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



See my answers at <http://bit.ly/nicksurveyanswers>  
(please don't hit submit)

# R Markdown Cheat Sheet

`*This text will be bold**`

**This text will be bold**

`This text will be italic*`

*This text will be italic*

``This text will look like code``

`This text will look like code`

---

Some tips:

- KNIT AND KNIT OFTEN
- If spacing looks weird, add a blank line

# R Code Cheat Sheet

`<-`

- Assigns a value to a name
- `name <- value`

# R Code Cheat Sheet

`read.csv(file, stringsAsFactors = TRUE)`

- `file` is the name of the file from which data are to be read
- `stringsAsFactors` should generally be set to `TRUE` (all caps): this determines how text-like data in a CSV file is interpreted by R.

# R Code Cheat Sheet

## `head(x, n)`

- `x` is a `data.frame`
- `n` is the number of rows to print

# R Code Cheat Sheet

## `str(object)`

- `object` is the R object about which you want to have some information (i.e., see the structure of).