PSTAT 5LS Lab 1

Professor Miller

June 29 and July 6, 2023

Section 1

Learning Objectives

R 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

Functions covered in this lab

1. `read.csv()`
2. `head()`
3. `str()`

Section 2

Lab Tutorial

Getting Started: What is R?

In Statistics, we often use computers to analyze data. There are a lot of programs that can help you do statistical analyses. One of the most popular (and powerful) is called R. R is a “statistical computing environment” that is designed for manipulating data, doing calculations, and making graphical displays. R works by writing **R code**.

That might sound scary, but *don’t worry*: this is not a programming class. Over the course of the quarter, you’ll learn how to edit and write some basic R code to help you analyze data to answer research questions. Our goal in lab is to help you learn the basics of R and R coding, but through the lens of answering statistical questions.

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 quarter.
- *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!

JupyterHub

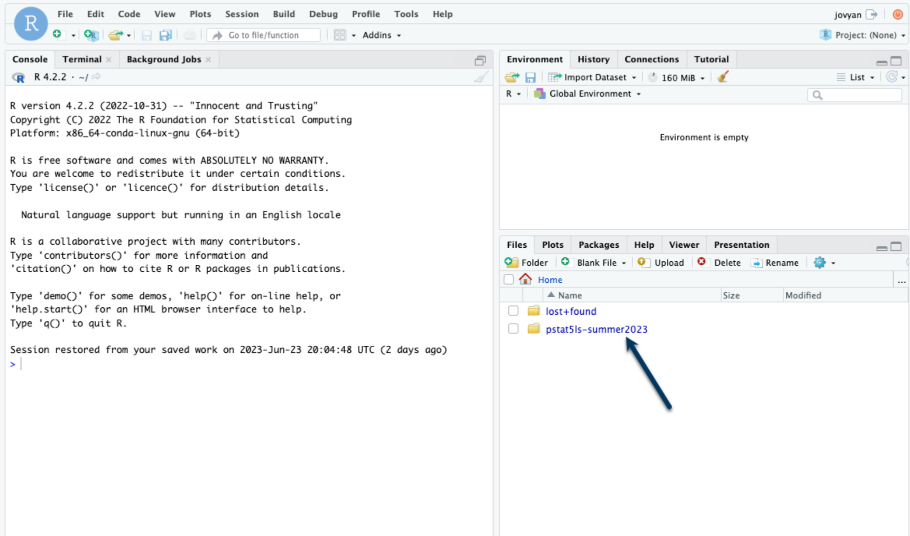
We will be accessing R and RStudio through JupyterHub. JupyterHub is a web-based multi-user environment that will allow us to share with you the files you need for lab. LSIT has provided us with access to JupyterHub.

Each time we begin a lab, navigate to https://tinyurl.com/pstat5ls.

This will open your “instance” of JupyterHub and copy all necessary files for PSTAT 5LS labs.

Initial View of RStudio

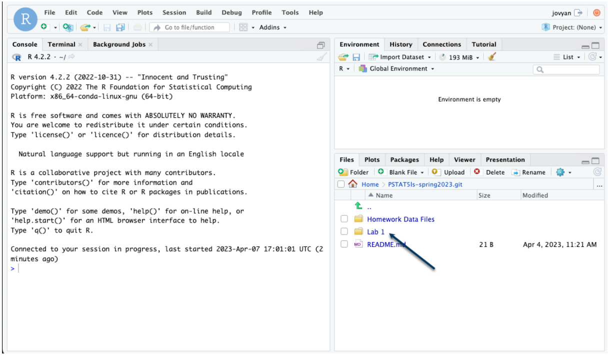
Click on the ‘PSTAT5ls-summer2023’ folder.



```
R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Session restored from your saved work on 2023-Jun-23 20:04:48 UTC (2 days ago)
>
```

Accessing the Lab for the Week

After you have clicked on the 'PSTAT5ls-summer2023' folder, click on the folder for the week's lab.



```
R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

Connected to your session in progress, last started 2023-Apr-07 17:01:01 UTC (2
minutes ago)
>
```

Name	Size	Modified
..		
Homework Data Files		
Lab 1		
README.md	21 B	Apr 4, 2023, 11:21 AM

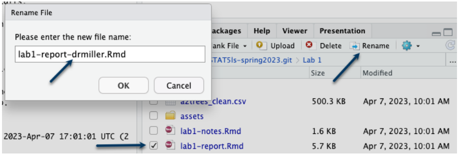
Working with R in JupyterHub

When you navigate to https://tinyurl.com/pstat5ls, you are telling JupyterHub to bring in all of the files that Professor Miller has in our PSTAT 5LS site. This means that files you have worked on will be overwritten.

To avoid your work being overwritten, **rename** both the notes and report files. For Lab 1, you want to rename lab1-notes.Rmd and lab1-report.Rmd. You could add your name to the file name (e.g., lab1-notes-drmiller.Rmd and lab1-report-drmiller.Rmd).

Renaming Files

To rename a file, check the box next to the file you want to rename. Click "Rename" in the files pane, rename the file, then click OK.



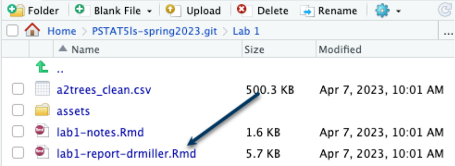
Rename File

Please enter the new file name:

lab1-report-drmiller.Rmd

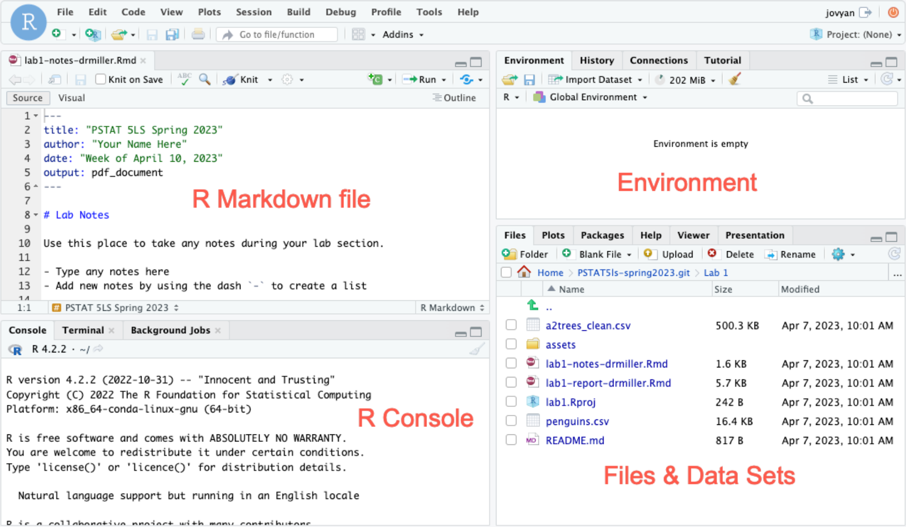
		Size	Modified
□	a2trees_clean.csv	500.3 KB	Apr 7, 2023, 10:01 AM
□	assets		
□	lab1-notes.Rmd	1.6 KB	Apr 7, 2023, 10:01 AM
☑	lab1-report.Rmd	5.7 KB	Apr 7, 2023, 10:01 AM

Your file will then appear with its new name.



	Name	Size	Modified
	..		
□	a2trees_clean.csv	500.3 KB	Apr 7, 2023, 10:01 AM
□	assets		
□	lab1-notes.Rmd	1.6 KB	Apr 7, 2023, 10:01 AM
□	lab1-report-drmiller.Rmd	5.7 KB	Apr 7, 2023, 10:01 AM

RStudio Project Screen



```
---
title: "PSTAT 5LS Spring 2023"
author: "Your Name Here"
date: "Week of April 10, 2023"
output: pdf_document
---

# Lab Notes

Use this place to take any notes during your lab section.

- Type any notes here
- Add new notes by using the dash `-` to create a list
```

Environment is empty

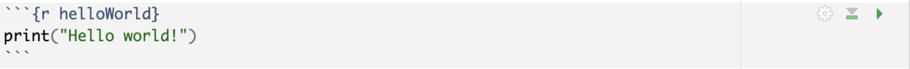
Name	Size	Modified
..		
a2trees_clean.csv	500.3 KB	Apr 7, 2023, 10:01 AM
assets		
lab1-notes-drmiller.Rmd	1.6 KB	Apr 7, 2023, 10:01 AM
lab1-report-drmiller.Rmd	5.7 KB	Apr 7, 2023, 10:01 AM
lab1.Rproj	242 B	Apr 7, 2023, 10:01 AM
penguins.csv	16.4 KB	Apr 7, 2023, 10:01 AM
README.md	817 B	Apr 7, 2023, 10:01 AM

```
R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors
```

R Markdown

R Markdown lets you combine text, R code, and plots in one pretty, reproducible report. If you're curious about this, you can find more details on using R Markdown at http://rmarkdown.rstudio.com.

R Markdown runs code contained in "chunks". A chunk looks like this:

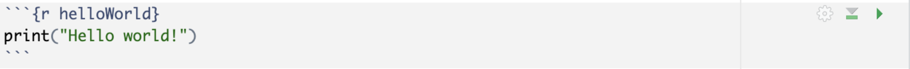


~~~
```{r helloWorld}
print("Hello world!")
```
~~~

```r
print("Hello world!")
```

```
## [1] "Hello world!"
```

R Markdown Chunk

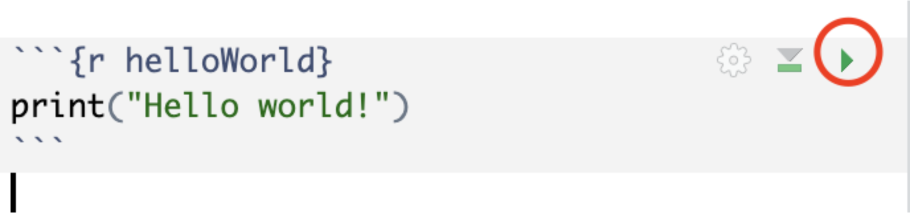


~~~
```{r helloWorld}
print("Hello world!")
```
~~~

Notice that the code, print("Hello world!") is contained between three backticks (```, below the esc key on a US English keyboard; the same key as the tilde) and some stuff in curly braces {}. This is how R Markdown knows where your chunks start and stop. We'll almost always provide pre-made chunks for you to use.

Running a Chunk

You can just run a single chunk by clicking the green “play” button in the upper right corner of the chunk. It’s usually a good idea to click the “Run All Chunks Above” button immediately to the left of the play button first. So you should click Run All Chunks Above, then the play button.



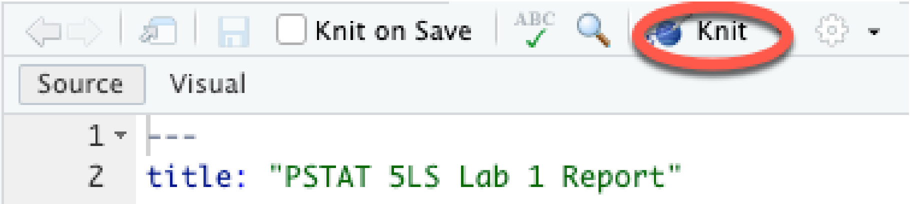
~~~
```{r helloWorld}
print("Hello world!")
```
~~~

If you can’t find the play button, just highlight the code you want to run and click the Run button in the top right corner of the editor pane.

Let’s try this code together in the `tryIt1` chunk in your notes document!

Knitting a Document

When you click the **Knit** button in R Studio, a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.



```
---
title: "PSTAT 5LS Lab 1 Report"
```

Tips for R Markdown

1. **Knit and knit often**: Frequently knitting your document will help you make sure that all your code works and that the document looks the way you want.
2. **Don't be afraid to experiment**: Nobody gets things exactly right the first time, and we all forget how things work sometimes. Keep trying, and you'll eventually get what you want!
3. **Formatting**: You can make text **bold** by surrounding it with two asterisks (**) and *italic* by surrounding it with one asterisk (*). You can make text `look like code` by surrounding it with single backticks.

Using R as a Calculator

At it's most basic, R is a fancy calculator.

Let's try this example together in the `tryIt2` chunk in your lab notes!

Remember, you can run the code in this chunk by clicking the green "play" button in the upper right corner of the chunk.

```
5 * 7
```

```
## [1] 35
```

When you run the chunk, you'll see a `[1]` before the output of 35. *Just ignore this. The result is 35.*

All of the symbols you think you'd use to do math work: + for addition, − for subtraction, * for multiplication, / for division, ^ for exponentiation.

R Function Syntax

We'll be using R primarily through the use of *functions*. A function generally looks something like this:

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

Notice that we have the name of the function, followed immediately by an open parenthesis '(' without a space, then a sequence of *arguments* to the function, separated by commas ',', then a close parenthesis ')'.

R “Assignment” Syntax

Often, we want R to remember the results of a calculation so that we can use it later. We can give the result of some code a name by **assigning** it to something.

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

We read that code as “x *gets* 36 / 6”. The arrow is made using the less-than symbol (<, shift + comma on a US English keyboard) and a hyphen.

Run the `tryIt3` code chunk in your notes to assign 36 divided by 6 to the variable `x`.

R “Assignment” Syntax

Now, we’ve stored the result as x, and R will remember that x is 6. You can see in the environment pane in R Studio (top right) that there’s now a “value” called x and it’s 6. You can also access the value of x by typing x into R. Check it out:

```
x
```

```
## [1] 6
```

Run the tryIt4 code chunk in your notes to see the value that R stored for the variable x.

R is Case-Sensitive

R is “case-sensitive”, which means that upper-case letters are *different from* 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
```

Run the `tryIt5` code chunk in your notes to see what happens.

When giving things names in R, you can only use a combination of letters, numbers, periods, and underscores, and the **names have to start with a letter or a period**. People tend to use underscores or periods instead of spaces.

Common Errors

Watch what happens when you try to assign something to a “bad” name:

```
tik tok <- 12
```

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

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

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

```
_hi_mom <- 5^2
```

```
## Error: <text>:1:2: unexpected symbol
## 1: _hi_mom
##      ^
```

Common Errors

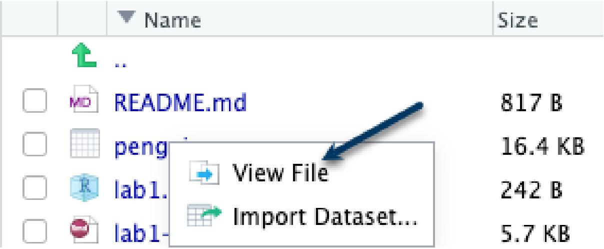
The errors saying “unexpected symbols” or “unexpected input” are R’s way of telling you that these names are not allowed, and that you should use a different name. Here’s how we’d correct these:

```
tiktok <- 12

forever <- 4 * 2

hi_mom <- 5^2
```

CSV files: a common way to store data

CSV stands for “comma separated values” and is a commonly used file type for storing data.

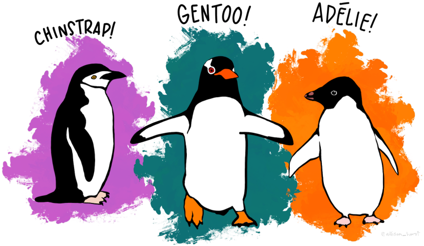
Open the file “penguins.csv” from the files pane (lower right) to see what a .csv file looks like:



Each row of the file is an “observation” or “case”, and consists of one or more variables whose *values* are *separated* by *commas* (hey, look at that). The first row contains the names of the variables contained in the file.

Palmer Penguins Data

We're going to start by working with a data set with data on 333 penguins collected from 3 islands in the Palmer Archipelago in Antarctica. Data were collected and made available by Dr. Kristen Gorman and the Palmer Station, Antarctica LTER, a member of the Long Term Ecological Research Network, and the data were prepared by Dr. Allison Horst.



Reading Data Into R

We can **read** data into R using a function called `read.csv()`. The first argument to `read.csv()` is the name of a .csv file (here, `penguins.csv`), in quotes. We then store the results of `read.csv()` as an object called `penguins`.

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

Run the `tryIt6` chunk in your notes to read in the penguins data file.

Peeking at the Data

Let's see what's in the data. We can peek at the first few (6, specifically) rows of the data using the `head()` function:

```
head(penguins)
```

```
##   species    island bill_length_mm bill_depth_mm flipper_le
## 1  Adelie Torgersen           39.1          18.7
## 2  Adelie Torgersen           39.5          17.4
## 3  Adelie Torgersen           40.3          18.0
## 4  Adelie Torgersen           36.7          19.3
## 5  Adelie Torgersen           39.3          20.6
## 6  Adelie Torgersen           38.9          17.8
##      sex year
## 1   male 2007
## 2 female 2007
## 3 female 2007
## 4 female 2007
```

Data

Variable name	Description
`species`	Penguin species (Adélie, Chinstrap, and Gentoo)
`island`	Island in Palmer Archipeligo, Antarctica, on which the penguin was observed (Biscoe, Dream, or Torgersen)
`bill_length_mm`	A number denoting bill length (in millimeters)
`bill_depth_mm`	A number denoting bill depth (in millimeters)

Data Continued

Variable name	Description
`flipper_length_mm`	A whole number denoting flipper length (in millimeters)
`body_mass_g`	A whole number denoting penguin body mass (in grams)
`sex`	Penguin sex (female, male)
`year`	Study year (2007, 2008, 2009)

Another Way

We can also peek at the data using a function called `str()` (pronounced “stir”, short for “structure”):

```
str(penguins)
```

```
## 'data.frame':    333 obs. of  8 variables:
##  $ species          : Factor w/ 3 levels "Adelie","Chinstra
##  $ island           : Factor w/ 3 levels "Biscoe","Dream",
##  $ bill_length_mm   : num  39.1 39.5 40.3 36.7 39.3 38.9 39
##  $ bill_depth_mm    : num  18.7 17.4 18 19.3 20.6 17.8 19.6
##  $ flipper_length_mm: int  181 186 195 193 190 181 195 182
##  $ body_mass_g      : int  3750 3800 3250 3450 3650 3625 46
##  $ sex              : Factor w/ 2 levels "female","male": 2
##  $ year             : int  2007 2007 2007 2007 2007 2007 20
```

Run the `tryIt8` chunk in your notes examine the structure of the penguins data file.

Knit Your Document One Final Time

Before you turn in your document, you want to be sure to knit one more time. This is just in case you have made any changes.

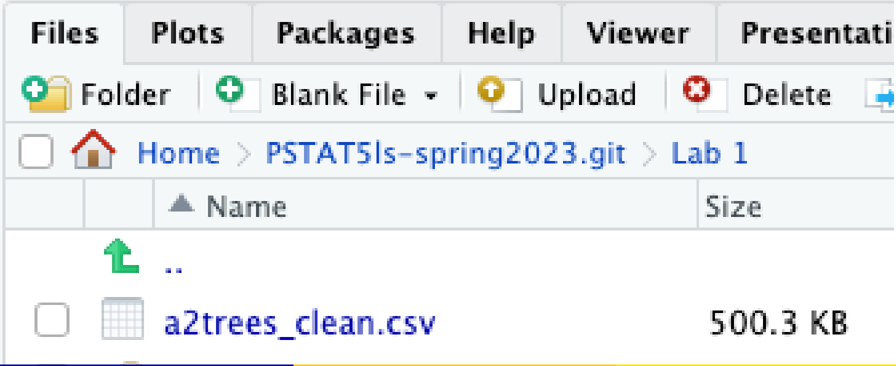
Be sure that a popup window shows you the file! Otherwise there is an error in the creation of the document. You should troubleshoot the markdown file.

Be sure that you read over your file and make sure that it looks right! If pieces are missing, then there is an error in the creation of the document. You should troubleshoot the markdown file.

Export Your .pdf File

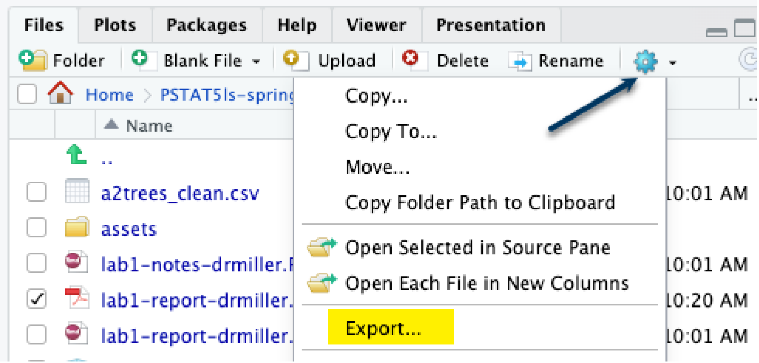
Did you notice that when you knit, the document that is created is a .pdf file? Be sure that you always grab the **.pdf** document to submit your lab report on Gradescope!

- In the Files pane, check the box next to `lab1-report-yourname.pdf`.



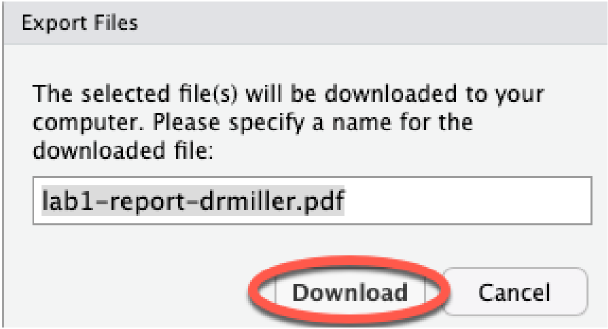
Export Your .pdf File Continued

- Click More > Export. . .



Export Your .pdf File Continued

- Click Download and save the file on your computer in a folder you'll remember and be able to find later. You can just call the file `lab1-report-yourname.pdf` or whatever else you'd like (aslong as you remember what you called it!).



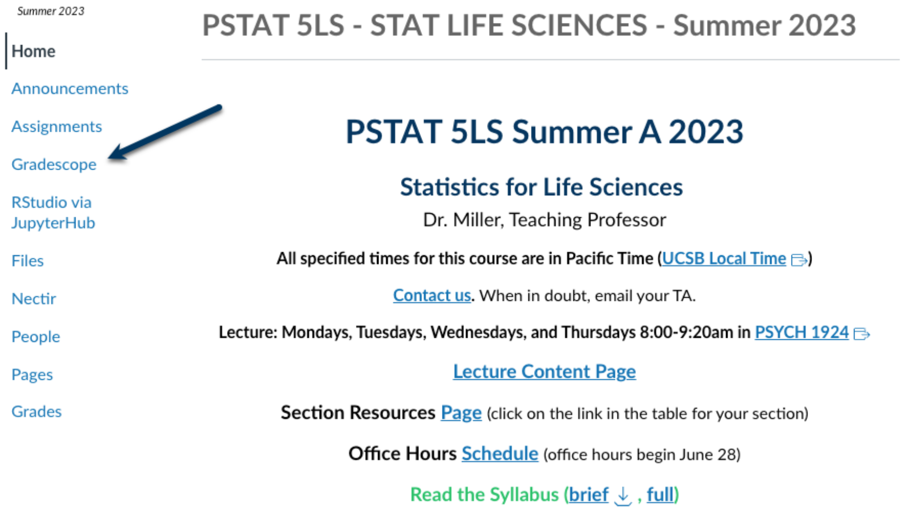
Export Files

The selected file(s) will be downloaded to your computer. Please specify a name for the downloaded file:

lab1-report-drmiller.pdf

Submit to Gradescope

Access Gradescope through Canvas



Summer 2023

- Home
- Announcements
- Assignments
- Gradescope
- RStudio via JupyterHub
- Files
- Nectir
- People
- Pages
- Grades

PSTAT 5LS - STAT LIFE SCIENCES - Summer 2023

PSTAT 5LS Summer A 2023

Statistics for Life Sciences

Dr. Miller, Teaching Professor

All specified times for this course are in Pacific Time (UCSB Local Time)

Contact us. When in doubt, email your TA.

Lecture: Mondays, Tuesdays, Wednesdays, and Thursdays 8:00-9:20am in PSYCH 1924

Lecture Content Page

Section Resources Page (click on the link in the table for your section)

Office Hours Schedule (office hours begin June 28)

Read the Syllabus (brief, full)

Submit to Gradescope Continued

- Click on the assignment name (e.g., Lab 1 Report).
- Click Submit PDF > Select PDF > locate the file > Upload PDF. You can only submit one PDF file per assignment.
- On your screen, you should see a list of the questions/problems in your assignment and thumbnails of your PDF page(s). For each question, click the question on the left and the PDF page(s) that contains the answers on the right.
- Now, click Submit. When your submission is successful, you will be sent to a new page to view your submission, you'll see a success message on your screen, and you'll receive an email. If your submitted file looks good, you see the success message, and you get the confirmation email, you're done!
- If you need to, select the Resubmit button in the bottom right corner of your screen below your submission. Then, repeat the steps above as many times as needed before the assignment due date passes. We will only see your most recent submission when we grade your lab report. All your past submissions are in your Submission History.

Section 3

Questions

What Questions Do You Have?

Section 4

Lab Report

Lab Report

Now it's time to work on the `lab1-report.Rmd` file. It should be open in your Markdown window; if not, go to the Files window in the bottom right corner and click on the file name.

Complete the questions in the lab report document. Be sure to ask questions as you work on this!

Instructions on how to turn the assignment in can be found at the end of the markdown file.