

Name: _____

SID: _____

Week 0 Lab

PHW142

Please remember to put your name and SID on the top of page 1 of all labs and problem sets.

Welcome

Welcome to PHW142: Introduction to Probability and Statistics in Biology and Public Health. For many of you, this will be your first foray into biostatistics. For others, you may have been working in R for several years. No matter your experience level, the teaching team hopes to guide you through this learning experience. Please don't hesitate to reach out to anyone on the teaching team if any questions or issues arise throughout this term.

What is R?

R is an open-source statistical programming language. Unlike many other languages, it is created to perform analyses, as opposed to being engineering-first (like Python). It's open-source nature means there are many many many libraries that have been developed to suit your project needs.

What is R Studio?

R Studio is an "integrated development environment," or IDE, that allows you to run R on your computer (or in the cloud!) and create scripts and see your output all in one program! For the first portion of the course, we'll be using a UC Berkeley-hosted version of RStudio called DataHub.

Side Note: Code Chunks

R Markdown (RMD) files can be interspersed between normal "text" (like this) and code chunks, which contain code. In a typical workflow, you can use RMD files to generate reports or simply have a place to document and write your code. Another benefit of using RMDs is that the output for a particular cell will be displayed immediately below it's code chunk.

Code chunks are delineated by: ````r ...````

- start: three backticks (top left key on your keyboard, to the right of "1") and 'r' - end: three more backticks

*# lines prefaced with a '#' are comments and will not be read
as code in R. use it to annotate/comment on your code!*

if viewing as an RMD (not PDF)

```
# the icons at the top right of the chunk are:  
# - chunk settings  
# - run all previous chunks until this one  
# - run this chunk
```

Also, break your code into lines of approximately 80 characters so it doesn't run off the page when you export. We won't be able to read code that goes off the page!

Adobe Acrobat

We will be using form fillable PDFs for your problem sets. This will help standardize the formatting of your submissions as well as help the teaching team return your assignments faster. It is **required** that you use Adobe Acrobat to complete the problem sets. Not all components of the PDF work in other programs (e.g. inserting graph as an image), so we will require all students to complete their problem sets with Acrobat. Luckily, all students get the entire Adobe Creative Cloud Suite (which includes Adobe Acrobat) for free while they are a student at UC Berkeley. We will be unable to troubleshoot technical issues if you don't use Adobe Acrobat.

Installing Adobe Acrobat

Go to <https://software.berkeley.edu/adobe-creative-cloud> and follow the instructions for your computer's operating system.

Submitting

In this class, our workflow will involve bCourses and Gradescope. Assignments will be submitted via the bCourses assignment dropbox. Grades will be viewable via bCourses, but assignment breakdowns (e.g. rubrics) will be viewable only in Gradescope. If you haven't already, you will likely encounter this website in future quantitative courses (e.g. epidemiology courses). When working off an R Markdown file, it is imperative that you **do not** remove any 'newpage' tags from the file nor chunks of code/original text. This will cause issues for the teaching team and can impact your grade and assignment turn around time.

Our recommended workflow:

1. Download the problem set PDF from bCourses
2. Access R Datahub via the lefthand navigation of bCourses
3. Open the R Markdown file corresponding to this week's problem set
 - In some weeks, we will not provide you with an R Markdown file. If there isn't an Rmd file for that week, create your own blank Rmd.
4. Complete the problem set PDF, pasting code/plots when necessary.

To show that you have successfully completed this lab, please run the following functions and paste your output in the fields below. The code is also available in the `w0-lab-code.Rmd`.

1. Run `Sys.info()` and `R.Version()`

2. Create a graph using the code below. To include the graph in the submission PDF:
 1. Right click and save the image from your Viewer (bottom right pane) or under the code chunk, wherever the plot appears.
 2. In the submission PDF, click on the field below to bring a pop-up box prompting you to browse for the saved file.
 3. Locate the image you saved earlier.
 4. Click Select.

```
library(epitools)
library(ggplot2)

data(wcgs)

ggplot(data = wcgs) +
  geom_point(aes(x = height0,
                 y = weight0))
```

