

# Biostatistics Class Preparation

*Jenine K. Harris*

I am looking forward to our upcoming *Foundations of Public Health: Biostatistics* class next week! The syllabus is now posted in GitHub at the URL: <https://github.com/jenineharris/Foundations-Biostatistics-Fall2018>. We will jump right into learning about data and R on the first day of class, so there are a few things below for you to do before our first class session to ensure a smooth start to the semester. If you run into any problems, please email me: [harrisj@wustl.edu](mailto:harrisj@wustl.edu).

## Installing R & RStudio

The class will be conducted using the R statistical programming environment. It is free and available for most modern operating systems. You will need to have regular access to a desktop computer or laptop where you can install programs. For the class you will want to install R and RStudio, please install these before our first class meeting.

R can be downloaded from <https://cran.r-project.org>. Once R has been installed, then install the RStudio IDE, available at <https://www.rstudio.com>. An IDE is an interactive development environment, which in this case makes R much easier to use.

Open **RStudio** and make sure that everything has installed correctly. There should be a window open on the left hand side of the RStudio screen that says “Console” in small bold print in the top left corner. R is running within this console window.

Check to see if R is working properly by typing in the following commands (shown in shading) at the R > prompt. Press Enter after typing each command to get the results shown:

```
2+2
```

```
## [1] 4
```

```
(4+6)/2
```

```
## [1] 5
```

```
10^2
```

```
## [1] 100
```

```
a <- 3
```

```
a
```

```
## [1] 3
```

## Finding and using the electronic course books and DataCamp

The main book we will use for the course is *Data Analysis with R (second edition)* by *Tony Fischetti*. This book is available electronically for free through the Washington University library website (<https://library.wustl.edu>). The book is also available for purchase through online retailers but will not be available at the campus bookstore.

Please note that, while the Fischetti book is available through the library for free electronically, only 8 users can view the book online at the same time. There may be times when 8 other students have the book open online and you will have to wait or find another source. It has never been a problem in past semesters, but if this is a concern to you, I'd recommend buying a hard copy of the book. There is one hard copy of the first

edition on permanent reserve in the Brown School library, which is also an option since there were not many changes.

We will also read several chapters from *Introductory Statistics with R* by *Peter Dalgaard*. This book is also available electronically from the library website and has no limitations on the number of users accessing the materials.

Finally, we will read a book about the people and events involved in the development of statistics called *The Lady Tasting Tea*. A few copies will be on reserve in the Brown School library and some public libraries also carry the book. You can get an electronic version through Amazon for 10 dollars and there were physical copies for sale for as little as \$2.16 last time I checked.

After installing R and RStudio, I highly recommend accessing the electronic version of the books and browsing through them to get a feel for how long the chapters are and how they are organized. I also highly recommend beginning the DataCamp free Intro to R course here: <https://www.datacamp.com/courses/free-introduction-to-r>.

As part of the course I have worked with DataCamp to provide you free access to all DataCamp classes starting on the first day of class and going for 6 months. I will use your wustl email available to me in the online system to sign you up for the free access, so you may receive an email today or later this week from DataCamp when I sign everyone up. You will need this access for a few weeks during the semester, but can then unsubscribe if you prefer not to use DataCamp resources.

Finally, please plan to bring your laptop with R and RStudio to class every week. If you do not have a laptop, please let me know ASAP so I can reserve one for you and make sure it has the current version of RStudio installed.