Lab 1: Introduction to R

Introduction

R is a free software environment for statistics. R Studio (also free) is an intergrated developement environment for R. You can think of R Studio as a companion program for basic R that helps you by keeping track of your things and organizing your various windows.

In this lab, you will install R and R Studio. You will also learn how to use RMarkdown and how to load basic datasets.

Setup

Installing R

The latest version of R is available for download on Linux, OS X and Windows on the Comprehensive R Archive Network (CRAN): https://cran.r-project.org/.

Installing RStudio

Install R before installing RStudio.

RStudio is available for download on www.rstudio.com. You do **not** need to pay money for RStudio. We will be using RStudio Desktop under the Open Source License.

R Packages

Installing R Packages

In addition to hosting the actual R software, CRAN also hosts a variety of packages of R code. These packages are neatly bundled pieces of R code that add functionality to your basic R environment.

One package that is very handy for data science is Tidyverse (actually a collection of packages). Install Tidyverse by running the following command:

```
install.packages("tidyverse")
```

Loading R Packages

Once installed, packages need to be loaded into your environment before you can use them. You can load the core packages of Tidyverse by running the following command.

```
library("tidyverse")
```

Help! How do I use this?

To quickly open the documentation for package or function using the `help()` function. This can be shortened to just a question mark. In RStudio, this will open a help window on your chosen function.

```
# An example
help(tidyverse)
?tidyverse #this does the same thing
```

You can find more information on getting help with R on the R Project website.

Submitting your assignments with RMarkdown.

RMarkdown is a format that allows you to easily share organized, reproducible code and results with others. RMarkdown support is built into RStudio, so producing RMarkdown documents in RStudio is trivial once you become familiar with the formatting. Make a new RMarkdown file by going to `File > New File > R Markdown ...` in the menu bar of RStudio. Enter an informative title for your new RMarkdown document. You can leave the default output format option set to `HTML` for now. The PDF option is also very useful but may require you to install TeX if you don't have it already—this could take several gigabytes of space so make sure you have room on your computer if you want to go this route.

RMarkdown syntax

Learn more about how to utilize RMarkdown by following along on this quick guide from RStudio.com.

Reading Data into R

R comes with a variety of functions built-in to help you load data. When we load Tidyverse, a package called `readr` is also loaded which contains additional functions to help you load data.

There are two basic ways to load comma-separated value files into R, now that we've loaded the Tidyverse package. We can use the `read.csv` function from base R or the `read_csv` function from the `readr` package which got loaded as part of the Tidyverse core. One notable difference between `read.csv` and `read_csv` is

that `read.csv` will return a standard R dataframe whereas `read_csv` will return a tibble, which like a standard R dataframe but with certain changes to improve their behavior.

Look up the details on how to use one of these two functions using the help documentation.

In addition to `read.csv` or `read_csv`, both base R and the `readr` package contain functions to read many other formats. Explore these on your own.

If they don't contain code that will read in your preferred format, there may be another package in existence that does. For example, if you want to load in data directly from a Microsoft Excel spreadsheet, you could do this by using the readxl package.

Your Turn: Loading a Dataset

You can choose between a publicly available health dataset or you can use your own data if you have any! Here are some suggestions.

- Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011
- U.S. Chronic Disease Indicators (CDI)

Download and load your preferred dataset into R. Print the first 5 rows and 5 columns. Submit your work as a knitted RMarkdown output file (either HTML or PDF format).