Biostat 561: Homework 3

Instructor: Amy Willis, Biostatistics, UW

October 12, 2017

Homework due October 19, 2 p.m.

Link to Homework 3 submission: `https://classroom.github.com/a/yqKcab5N`

You will need to install and load the `tidyverse` for this homework.

Question 1: Getting started with the tidyverse

The dataset `faithful` gives waiting time between eruptions and the duration of the eruption for the Old Faithful geiser in Yellowstone.

a) Use `apply` (without any piping) to calculate the mean waiting time and duration of the eruptions.

b) Use the pipe operator in conjunction with `apply` to calculate the mean waiting time and duration.

c) Using the pipe operator and `summarise`, calculate the mean eruption duration.

d) Calculate the mean waiting time and duration. You need only the pipe operator and one other function from `dplyr`. Find this function at `http://dplyr.tidyverse.org/reference/index.html`.

e) `faithful` is a data frame. How many lines are output by default when you show a data frame in console? Convert `faithful` to a tibble. How many lines are output by default when you show a tibble? What other information is given?

Question 2: Intuitive data exploration

a) `filter()` and `subset()` perform similar functions. List 3 differences between them. Which will you intend to use?

b) Using the data frame `airquality`, calculate the mean and standard deviation of Ozone for Temperatures of 85 degrees or more.

c) Using the data frame `airquality`, calculate the mean and standard deviation of Ozone, stratifying observations with temperatures of 85 degrees or more (versus less than 85 degrees). Verify that this matches your results from last week using `by()`. Which do you find more intuitive?

Question 3: Function composition

Install and load the package `magrittr`. Make extensive use of the documentation `http://magrittr.tidyverse.org/` in answering these questions.

a) What does this do: `f <- . %>% cos %>% sin`?

b) What is the compound assignment pipe operator, `%<>%`?

c) The temperature variable in `airquality` is currently in degrees Fahrenheit. Modify it in place to the same measurement in Celsius, rounded to the nearest degree. Show the output of `head(airquality$Temp)` before and after.