Intro to R Training (class sheet)

1/12/2020

Welcome to R training!

This script is a template for you to complete during the guided in-class R training activity. Follow along with the different subheadings and complete coding activities in each coding “chunk” within the script.

I encourage you to write notes/document all analysis functions and details of Rstudio usage that will be helpful for you in the future. Note areas are indicated by *yellow text between two underscores* and appears as the text in the sentance does.

Your TA will introduce the core concepts of the subheadings below:

## Creating a New R Script

*ADD ANY NOTES HERE*

## The Rstudio Screen

There are four panels to consider within the Rstudio screen: the console, the script tab, the global environment, and the file viewer.

*ADD ANY NOTES HERE*

### The Console

Bottom left. The console is responsible for all of the computational power of the R program. In this tab you can run code, operations, and view recent activity. **Note** anything entered into the console will not be saved when using Rstudio after you close the instance. All saved code must be entered into a script.

*ADD ANY NOTES HERE*

### The Script Tab

Top left. The script tab is the portion of Rstudio where all code, operations and actions can be entered then saved. Think of scripts as documents that contain a combination of code and plain text which Rstudio interprets differently based on the entry. **Note** the script tab will only appear if you have an active script open (it will not open automatically).

*ADD ANY NOTES HERE*

### The Global Environment

Top right. The global environment displays all of the objects/datasets/functions you have created with your R code during your current R session.

*ADD ANY NOTES HERE*

### The File Viewer

Bottom right. The file viewer shows you the active files you are using during your current Rstudio session and should mimic your File Explorer (PC) or Finder (Mac).

*ADD ANY NOTES HERE*

## Markdown Usage

A markdown file is a type of R script that is very useful for learning how to use Rstudio. Think of markdown files as word documents with code “chunks” mixed in with plain readable text.

*ADD ANY NOTES HERE*

### The YAML Header

The colorful lines at the beginning of your markdown file. These provide useful information for the creation and parameters of your script.

*ADD ANY NOTES HERE*

### Code Chunks

This is where you enter code for Rstudio to run. Create new chunks with control+alt+i (PC) or command+option+i (Mac).

*ADD ANY NOTES HERE*

## Libraries

Libraries of the workhorse of R computation. Think of libraries as add-ons that extend the functionality of base R to allow for complex analyses. To use a library you must **install** the library **once** but you must **load** the library for **each instance** of Rstudio where you would like to use it.

library(tidyverse)

## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

iris\_data <- as\_tibble(iris)  
iris\_data

## # A tibble: 150 x 5  
## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## <dbl> <dbl> <dbl> <dbl> <fct>   
## 1 5.1 3.5 1.4 0.2 setosa   
## 2 4.9 3 1.4 0.2 setosa   
## 3 4.7 3.2 1.3 0.2 setosa   
## 4 4.6 3.1 1.5 0.2 setosa   
## 5 5 3.6 1.4 0.2 setosa   
## 6 5.4 3.9 1.7 0.4 setosa   
## 7 4.6 3.4 1.4 0.3 setosa   
## 8 5 3.4 1.5 0.2 setosa   
## 9 4.4 2.9 1.4 0.2 setosa   
## 10 4.9 3.1 1.5 0.1 setosa   
## # ... with 140 more rows

### Installation and Loading

Install syntax: install.packages(“PACKAGE NAME”) Load syntax: library(PACKAGE NAME)

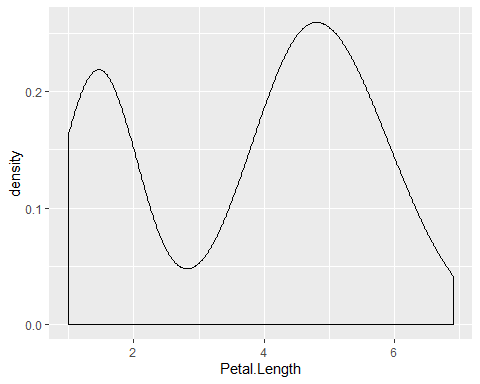
*ADD ANY NOTES HERE*

### The Tidyverse

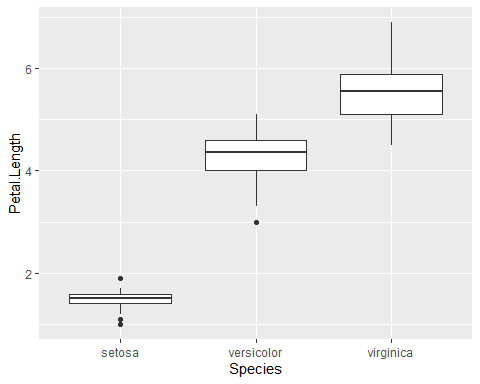
Arguably the most widely used R package. Designed for data organization, processing, tidying, and visualization.

*ADD ANY NOTES HERE*

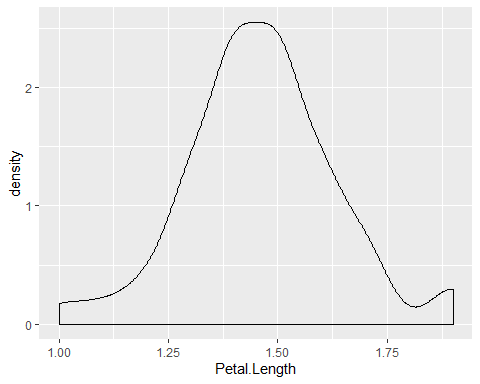
ggplot(iris\_data, aes(Petal.Length)) + geom\_density()



ggplot(iris\_data, aes(Species, Petal.Length)) + geom\_boxplot()



setosa\_only <- filter(iris\_data, Species == "setosa")  
  
ggplot(setosa\_only, aes(Petal.Length)) + geom\_density()



iris\_data <- iris\_data %>% mutate(common\_name = recode(Species,   
 "setosa" = "bristle-pointed iris",  
 "versicolor" = "blue flag",  
 "virginica" = "virginia iris"))

## Coding Crash Course

*ADD ANY NOTES HERE*

### Objects

Things that you create and are saved in R. Most commonly objects are used to store datasets or suubsets. Objects are given names by you when coding. It is best to avoid special characters such as: !@#$ etc. to avoid confusion.

*ADD ANY NOTES HERE*

### Classes

These are types of data that R can use. The common types that will be important for learning R are: Numeric (also called Double or Integer), Character, Factor, and Data Frame(called Tibbles in tidyverse).

*ADD ANY NOTES HERE*

### Data Wrangling

This is the process of organizing data into a format where it is usable in R and/or answers one of your research questions. Data wrangling is highly specific to the dataset you are using and requires a bit of R experience before you can design a plan. We will work ad a group with some aspects of data wrangling with this class.

*ADD ANY NOTES HERE*

### ggplot

ggplot is the core component of the tidyverse that allows researchers to plot high quality data visuals in Rstudio. Basic usage of some of the functionality of ggplot is the overall goal of R usage in this class.

*ADD ANY NOTES HERE*

iris\_mm <- iris\_data %>% mutate(Sepal.Length = Sepal.Length \* 10)