STA130 Homework

Problem Set 0

Josh Speagle & Scott Schwartz

The purpose of this ungraded "Homework 0" is to:

- ☐ Show how to access Jupyterhub and navigate RStudio
- ☐ Introduce .Rmd files, html/R comments, and code chunks [**source pane**]
- ☐ Introduce `parameter=argument` syntax controlling code chunks [**source pane**]
- ☐ Demonstrate R calculator arithmetic (`+`,`*`,`^`, etc.) and logic (`&`,`|`,`!`, etc.) [**console pane**]
- ☐ Discuss loadable packages like `readr`, `tibble`, and `tidyverse` which add functions like `readr::read_csv()` and `tibble::tibble()` to the R calculator
- ☐ Introduce `<-`, `->`, and `=` variable assignments in R [**source/console pane**]
- ☐ Show variable and data set management and the `rm()` function [**environment pane**]
- ☐ Introduce the `<cmd-shift-M>` (Mac) and `<ctrl-shift-M>` (PC) "pipe symbol" `%>%`

- ☐ Show how to run code with `<cmd-shift-return>` (Mac) and `<ctrl-shift-enter>` (PC), and the "play" and "knit" buttons [**source pane**]
- ☐ Show exporting/downloading files from RStudio, organizing and managing files, and course Rmd+pdf submission protocols [**files pane**]
- ☐ Highlight Tutorial, Packages, and Help tabs [**environment/files pane**]

This will both help to serve as a preview of (and complement to) topics that will be covered in later homework assignments.

```
# load tidyverse functionality <-- make code "easy to understand" with comments!
install.packages("tidyverse")
```

```
## Installing package into '/opt/r'
## (as 'lib' is unspecified)
```

```
library(tidyverse) # `install.packages("tidyverse")` isn't needed
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0      v purrr   1.0.0
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
                  #  on jupyterhub for the `tidyverse` package
```

Question 1

Run the R code chunk below that contains `tidyverse` functions to load and show a data set of coffee ratings.

- Use key-binding shortcuts cmd-shift-return (Mac) or ctrl-shift-enter (PC), or the green "play" button.

- The "could not find function" for `read_csv()` or `glimpse()` error is fixed by running `library(tidyverse)`.
- The "cannot open file" error is fixed by putting the file with the right name in the same directory as the .Rmd.

```
library(tidyverse)
coffee_ratings <- read_csv("coffee_ratings.csv") # rather than `read.csv()`
```

```
## Rows: 1338 Columns: 36
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (18): species, owner, country_of_origin, farm_name, mill, company, altit...
## dbl (18): total_cup_points, aroma, flavor, aftertaste, acidity, body, balanc...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(coffee_ratings)        # <- compare what this line of code does
```

```
## # A tibble: 6 x 36
##   total_cup~1 species owner count~2 farm_~3 mill  company altit~4 region produ~5
##         <dbl> <chr>   <chr> <chr>   <chr>   <chr> <chr>   <chr>   <chr>  <chr>
## 1        90.6 Arabica meta~ Ethiop~ "metad~ meta~ metad ~ 1950-2~ guji-~ METAD ~
## 2        89.9 Arabica meta~ Ethiop~ "metad~ meta~ metad ~ 1950-2~ guji-~ METAD ~
## 3        89.8 Arabica grou~ Guatem~ "san m~ <NA>  <NA>    1600 -~ <NA>   <NA>
## 4        89   Arabica yidn~ Ethiop~ "yidne~ wole~ yidnek~ 1800-2~ oromia Yidnek~
## 5        88.8 Arabica meta~ Ethiop~ "metad~ meta~ metad ~ 1950-2~ guji-~ METAD ~
## 6        88.8 Arabica ji-a~ Brazil   <NA>   <NA>  <NA>    <NA>    <NA>   <NA>
## # ... with 26 more variables: in_country_partner <chr>, harvest_year <chr>,
## #   grading_date <chr>, variety <chr>, processing_method <chr>, aroma <dbl>,
## #   flavor <dbl>, aftertaste <dbl>, acidity <dbl>, body <dbl>, balance <dbl>,
## #   uniformity <dbl>, clean_cup <dbl>, sweetness <dbl>, cupper_points <dbl>,
## #   moisture <dbl>, category_one_defects <dbl>, quakers <dbl>, color <chr>,
## #   category_two_defects <dbl>, expiration <chr>, certification_body <chr>,
## #   unit_of_measurement <chr>, altitude_low_meters <dbl>, ...
```

```
#coffee_ratings %>% head() # <- versus what this line of code does
# The `%>%` "pipe symbol" is created with <ctrl-shift-M> or <cmd-shift-M>
```

(a) Use the glimpse() function output above to determine how many rows and columns there are in the coffee ratings dataset.

```
glimpse(coffee_ratings)        # <- compare what this line of code does
```

```
## Rows: 1,338
## Columns: 36
## $ total_cup_points     <dbl> 90.58, 89.92, 89.75, 89.00, 88.83, 88.83, 88.75, ~
## $ species              <chr> "Arabica", "Arabica", "Arabica", "Arabica", "Arab~
## $ owner                <chr> "metad plc", "metad plc", "grounds for health adm~
## $ country_of_origin    <chr> "Ethiopia", "Ethiopia", "Guatemala", "Ethiopia", ~
## $ farm_name            <chr> "metad plc", "metad plc", "san marcos barrancas \~
## $ mill                 <chr> "metad plc", "metad plc", NA, "wolensu", "metad p~
## $ company              <chr> "metad agricultural developmet plc", "metad agric~
## $ altitude             <chr> "1950-2200", "1950-2200", "1600 - 1800 m", "1800-~
## $ region               <chr> "guji-hambela", "guji-hambela", NA, "oromia", "gu~
## $ producer             <chr> "METAD PLC", "METAD PLC", NA, "Yidnekachew Dabess~
## $ in_country_partner   <chr> "METAD Agricultural Development plc", "METAD Agri~
```

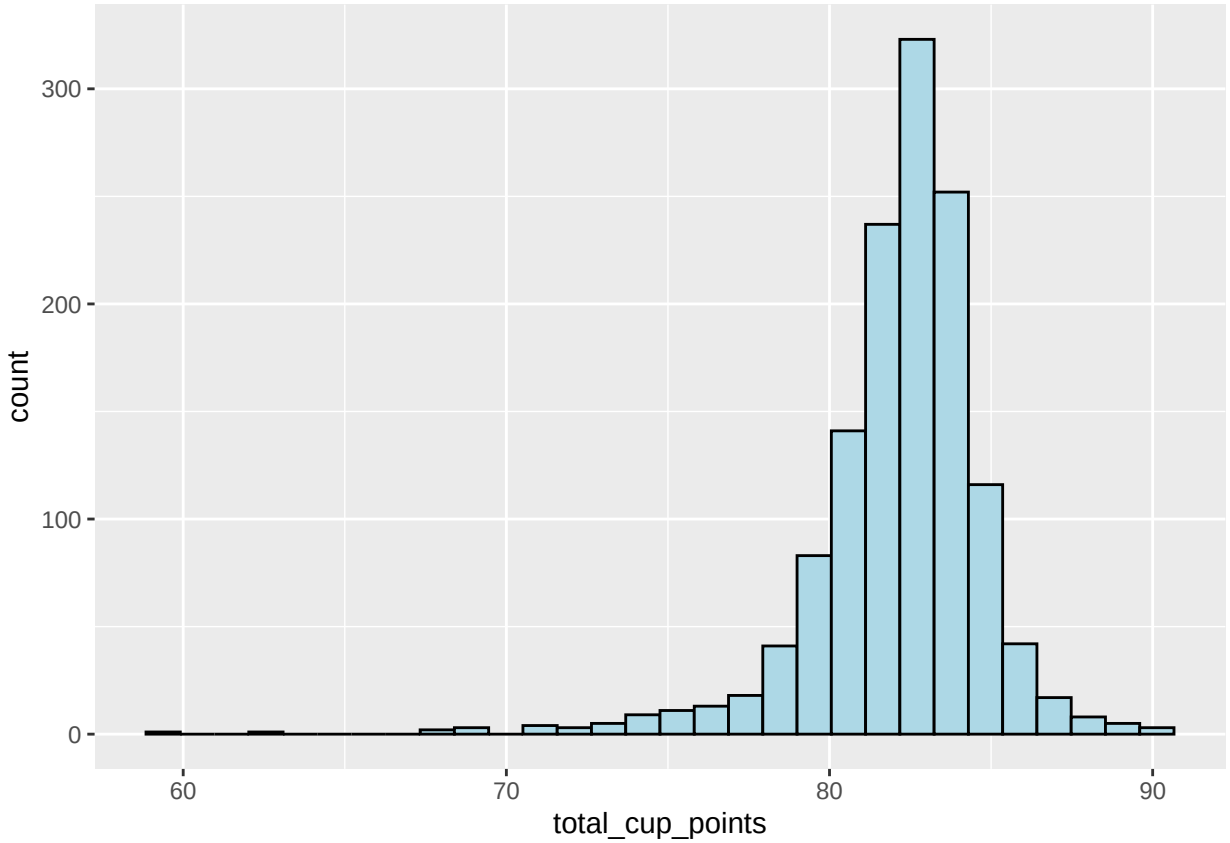
```
## $ harvest_year         <chr> "2014", "2014", NA, "2014", "2014", "2013", "2012~
## $ grading_date         <chr> "April 4th, 2015", "April 4th, 2015", "May 31st, ~
## $ variety              <chr> NA, "Other", "Bourbon", NA, "Other", NA, "Other",~
## $ processing_method    <chr> "Washed / Wet", "Washed / Wet", NA, "Natural / Dr~
## $ aroma                <dbl> 8.67, 8.75, 8.42, 8.17, 8.25, 8.58, 8.42, 8.25, 8~
## $ flavor               <dbl> 8.83, 8.67, 8.50, 8.58, 8.50, 8.42, 8.50, 8.33, 8~
## $ aftertaste           <dbl> 8.67, 8.50, 8.42, 8.42, 8.25, 8.42, 8.33, 8.50, 8~
## $ acidity              <dbl> 8.75, 8.58, 8.42, 8.42, 8.50, 8.50, 8.50, 8.42, 8~
## $ body                 <dbl> 8.50, 8.42, 8.33, 8.50, 8.42, 8.25, 8.25, 8.33, 8~
## $ balance              <dbl> 8.42, 8.42, 8.42, 8.25, 8.33, 8.33, 8.25, 8.50, 8~
## $ uniformity           <dbl> 10.00, 10.00, 10.00, 10.00, 10.00, 10.00, 10.00, ~
## $ clean_cup            <dbl> 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 1~
## $ sweetness            <dbl> 10.00, 10.00, 10.00, 10.00, 10.00, 10.00, 10.00, ~
## $ cupper_points        <dbl> 8.75, 8.58, 9.25, 8.67, 8.58, 8.33, 8.50, 9.00, 8~
## $ moisture             <dbl> 0.12, 0.12, 0.00, 0.11, 0.12, 0.11, 0.11, 0.03, 0~
## $ category_one_defects <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ quakers              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ color                <chr> "Green", "Green", NA, "Green", "Green", "Bluish-G~
## $ category_two_defects <dbl> 0, 1, 0, 2, 2, 1, 0, 0, 0, 4, 1, 0, 0, 2, 2, 0, 0~
## $ expiration           <chr> "April 3rd, 2016", "April 3rd, 2016", "May 31st, ~
## $ certification_body   <chr> "METAD Agricultural Development plc", "METAD Agri~
## $ unit_of_measurement  <chr> "m", "m", "m", "m", "m", "m", "m", "m", "m", "m",~
## $ altitude_low_meters  <dbl> 1950.0, 1950.0, 1600.0, 1800.0, 1950.0, NA, NA, 1~
## $ altitude_high_meters <dbl> 2200.0, 2200.0, 1800.0, 2200.0, 2200.0, NA, NA, 1~
## $ altitude_mean_meters <dbl> 2075.0, 2075.0, 1700.0, 2000.0, 2075.0, NA, NA, 1~
```

```
# coffee_ratings %>% glimpse() # <- versus what this line of code does
# The `%>%` "pipe symbol" is created with <ctrl-shift-M> or <cmd-shift-M>
```

(b) Change the property of the R code chunks above so the knit file shows the code and both the message from `read_csv` and the output from `glimpse`.

(c) Change the property of the R code chunk below so the plot of overall coffee ratings appears in the knitted output.

```
knitr::opts_chunk$set(eval=TRUE, include=TRUE, echo=FALSE, message=FALSE, warning=FALSE)
# Can you read this code? What do you roughly think it's doing?
coffee_ratings %>% # E.g., what does the `%>%` seem to be doing?
  ggplot(aes(x=total_cup_points)) + # E.g., what about `+` !??
  geom_histogram(bins=30, color="black", fill="light blue")
```




```
  # Note here that R code uses the same `parameter=argument` pairing as knitting
```

‘‘‘