Lab 1

Taylor Blair

Math 141, Week 1

Due: Before your Week 2 lab meeting

Goals of this lab

1. Practice running R code from chunks, modifying the text, and knitting an R Markdown document.
2. View and summarize data frames in R.

Problems

For each problem, put your solution between the bars of stars.

Problem 1: Inspecting `pdxTrees`

In this problem, we are going to explore a data frame in R.

a. For data frames, what does a row represent? What does a column represent?

- *A row represents a data point*
- *A column represents a variable or feature*

---

b. Run the following code to load the `pdxTrees` library/package and to grab data on trees in Portland parks.

Pro tip: To run each line of code, put the cursor on the same line as the code and hit Command+Enter (for a Mac) and Control+Enter (for a PC).

```
library(pdxTrees)
pdxTrees_parks <- get_pdxTrees_parks()
```

c. Run the following to begin exploring the data. (Leave the `eval = FALSE`.)

```
View(pdxTrees_parks)
?get_pdxTrees_parks
```

d. What does a row of the data frame represent? How many rows are in the data frame?

- *25,534 rows*
- *A row represents a single tree in the portland area*

e. What does a column of the data frame represent? How many columns are in the data frame?

- *34 columns*
- *A column represents a feature or variable*

f. List three quantitative variables found in the dataset (beyond those discussed in class).

- *Longitude, Latitude, useriD, DBH, Inventory date, tree height, crown width NS and EW, crown base height, structural value, carbon storage pounds, carbon storage value, carbon sequestration pounds, carbon sequest val, stormwater ft, pollution removal val, pollution removal oz, total annual services*

g. List three categorical variables found in the dataset (beyond those discussed in class).

- *Species factoid, origin, nuisance, edible, native, mature size, functional type, scientific name, park, collected by, condition, common name, species, family, genus*

Problem 2: Palmer Penguins

For this problem, let's explore the `palmerpenguins` data frame (which can be found in a package with the same name). The following code loads the data and prints a snapshot.

```
library(palmerpenguins)
penguins
```

```
## # A tibble: 344 x 8
##    species island bill_length_mm bill_depth_mm flipper_length_~ body_mass_g
##    <fct>   <fct>           <dbl>         <dbl>            <int>       <int>
##  1 Adelie  Torge~           39.1          18.7              181        3750
##  2 Adelie  Torge~           39.5          17.4              186        3800
##  3 Adelie  Torge~           40.3          18                195        3250
##  4 Adelie  Torge~           NA            NA                 NA          NA
##  5 Adelie  Torge~           36.7          19.3              193        3450
##  6 Adelie  Torge~           39.3          20.6              190        3650
##  7 Adelie  Torge~           38.9          17.8              181        3625
##  8 Adelie  Torge~           39.2          19.6              195        4675
##  9 Adelie  Torge~           34.1          18.1              193        3475
## 10 Adelie  Torge~           42            20.2              190        4250
## # ... with 334 more rows, and 2 more variables: sex <fct>, year <int>
```

Describe the data. Make sure to identify what the rows and each column represent. For each variable, identify its type. If you want more information, load the help page:

```
# (Leave the eval = FALSE)
?penguins
```

- *344 rows by 8 columns*

  - *Each row represents a penguin in the Palmer Archipelago*
- *Columns/Features*
  - `Species`
  - "*A factor denoting penguin species (Adélie, Chinstrap and Gentoo)*"
  - *Categorical*
  - `Island`
  - "*A factor denoting island in Palmer Archipelago, Antarctica (Biscoe, Dream or Torgersen)*"
  - *Categorical*
  - `bill_length_mm`
  - "*A number denoting bill length (millimeters)*"
  - *Quanitative*
  - `bill_depth_mm`
  - "*A number denoting bill depth (millimeters)*"
  - *Quanitative*
  - `body_mass_g`
  - "*An integer denoting body mass (grams)*""
  - *Quanitative*
  - `Sex`
  - "*A factor denoting penguin sex (female, male)*"
  - *Categorical, sometimes undifined*
  - `Year`
  - "*An integer denoting the study year (2007, 2008, or 2009)*"
  - *Quanitative and Categorical. One of three years so can be expressed either way*
