Homework 5

Jules Merigot (8488256)

November 10, 2022

PSTAT 131/231 Statistical Machine Learning - Fall 2022

Elastic Net Tuning

Before we get started, let's load the Pokemon data in into our workspace.

```
pokemon_data <- read.csv(file = "C:/Users/jules/OneDrive/Desktop/homework-5/data/Pokemon.csv")
head(pokemon_data)
```

```
##   X.                   Name Type.1 Type.2 Total HP Attack Defense Sp..Atk
## 1  1              Bulbasaur  Grass Poison   318 45     49      49      65
## 2  2                Ivysaur  Grass Poison   405 60     62      63      80
## 3  3               Venusaur  Grass Poison   525 80     82      83     100
## 4  3 VenusaurMega Venusaur  Grass Poison   625 80    100     123     122
## 5  4             Charmander   Fire          309 39     52      43      60
## 6  5             Charmeleon   Fire          405 58     64      58      80
##   Sp..Def Speed Generation Legendary
## 1      65    45          1     False
## 2      80    60          1     False
## 3     100    80          1     False
## 4     120    80          1     False
## 5      50    65          1     False
## 6      65    80          1     False
```

Exercise 1

Let's load the janitor package, and use its `clean_names()` function on the Pokémon data. We'll save the results to work with for the rest of the assignment.

```
library(janitor)

Pokemon_data <- clean_names(pokemon_data)
head(Pokemon_data)
```

```
##   x              name type_1 type_2 total hp attack defense sp_atk sp_def
## 1 1         Bulbasaur  Grass Poison   318 45     49      49     65     65
## 2 2           Ivysaur  Grass Poison   405 60     62      63     80     80
## 3 3          Venusaur  Grass Poison   525 80     82      83    100    100
```

```
## 4 3 VenusaurMega Venusaur   Grass Poison    625 80     100      123     122    120
## 5 4               Charmander    Fire          309 39      52       43      60     50
## 6 5               Charmeleon    Fire          405 58      64       58      80     65
##   speed generation legendary
## 1    45          1     False
## 2    60          1     False
## 3    80          1     False
## 4    80          1     False
## 5    65          1     False
## 6    80          1     False
```

As we can see in the data above, the names of each column have been changed to simpler, more efficient, and unique names using strictly the "_" character, numbers, and letters. This shows how useful `clean_names()` is, because it allows for a rapid change in the varaible and predictor names, thus allowing them to be referenced and used more efficiently in the rest of project or assignment being completed.

Exercise 2

Using the entire data set, let's create a bar chart of the outcome variable, `type_1`.

```
Pokemon_data %>%
  ggplot(aes(x=type_1)) +
  geom_bar()
```
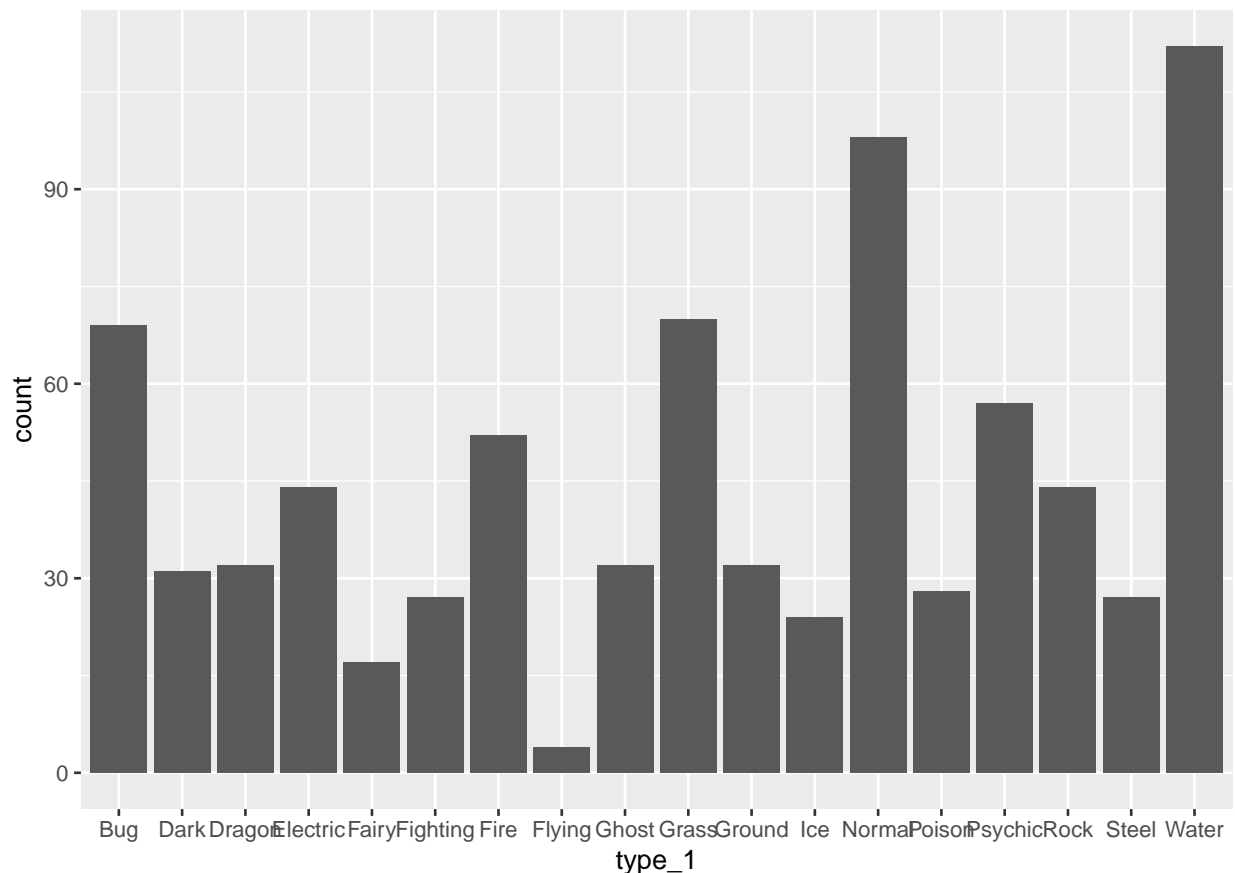

There are 18 classes of the outcome `type_1`, which means there are 18 different types of Pokemon. While

2

there are many Pokemon of the "Water" type, there are very few Pokemon of the "Flying" type.
For this assignment, we'll handle the rarer classes by simply filtering them out. Let's filter the entire data
set to contain only Pokemon whose `type_1` is Bug, Fire, Grass, Normal, Water, or Psychic.

```
filt_types <- c("Bug", "Fire", "Grass", "Normal", "Water", "Pyschic")

Pokemon_data %>%
  filter(type_1 == filt_types) %>%
  head()
```

```
##     x                      name type_1 type_2 total hp attack defense sp_atk
## 1   3                  Venusaur  Grass Poison   525 80     82      83    100
## 2   6 CharizardMega Charizard X   Fire Dragon   634 78    130     111    130
## 3   8                 Wartortle  Water          405 59     63      80     65
## 4  15                  Beedrill    Bug Poison   395 65     90      40     45
## 5  17                 Pidgeotto Normal Flying   349 63     60      55     50
## 6  22                    Fearow Normal Flying   442 65     90      65     61
##   sp_def speed generation legendary
## 1    100    80          1     False
## 2     85   100          1     False
## 3     80    58          1     False
## 4     80    75          1     False
## 5     50    71          1     False
## 6     61   100          1     False
```

Now that we're done filtering, let's convert `type_1` and `legendary` to factors.

```
Pokemon_data$type_1 <- factor(Pokemon_data$type_1)
Pokemon_data$legendary <- factor(Pokemon_data$legendary)
```