**R PRACTICE REPORT**

**PROBABILITY AND STATISTICS ASSIGNMENT**

**WEEK 1**

**MODULE 1**

By : **HARSHIT GAUR**

MASTER OF PROFESSIONAL STUDIES IN ANALYTICS

ALY 6010 : PROBABILITY THEORY AND INTRODUCTORY STATISTICS

NOVEMBER 6, 2021

To : **PROF. AMIN KARIMPOUR**

# **ABSTRACT**

**Data analytics** is a discipline focused on extracting insights from data. It comprises the processes, tools and techniques of data analysis and management, including the collection, organization, and storage of data. The chief aim of data analytics is to apply statistical analysis and technologies on data to find trends and solve problems.

**Sales analytics** is the practice of generating insights from sales data, trends, and metrics to set targets and forecast future sales performance. The best practice for sales analytics is to closely tie all activities to determine revenue outcomes and set objectives for your sales team.

Analysis should focus on improvement and developing a strategy for improving your sales performance in both the short- and long-term.

It provides insights about the top performing and underperforming products/services, the problems in selling and market opportunities, sales forecasting, and sales activities that generate revenue.

# **INTRODUCTION**

It is strongly encouraged to find and choose a data set in an area where one is more interested and is personally motivated to explore about.

Initially, I decided to go with a dataset that will help me hone the skills that I have learned so far in this course and go even beyond it by acquiring more knowledge and broadening my prowess of the analytical skills and R programming language. Then, I came across a dataset which not only interested me very much but also gave me various ideas to implement using that data set.   
  
 I grew up watching an all-time favourite TV Show - Pokémon and the dataset which I chose is about the different types of Pokémon present in the TV series along with their attributes. The dataset contains various numerical and categorical data. It has 1,045 data points with 11 features related to the primary types, attack attributes, defence attributes, hit points, special attributes and more. I decided to put my fondness of the series to better use and up-skill myself in the analytical, visualization, and programmatical aspects of the domain.

The features available in the data set are -

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Feature** | **Dictionary** |
| 1. | Name | Name of the Pokémon |
| 2. | Name2 | Secondary name of the Pokémon |
| 3. | Primary.Type | Primary type |
| 4. | Secondary.Type | Secondary type to which Pokémon belongs to |
| 5. | Attack | Attack attribute |
| 6. | Defense | Defence Quality |
| 7. | HP | Hit Points |
| 8. | Sp.Attack | Special Attack attribute |
| 9. | Sp.Defense | Special Defense attribute |
| 10. | Speed | Speed of the Pokémon |
| 11. | Total | Total Qualities (Summation) |

*Table 1: Features of the data set with their dictionary.*

The data set was obtained from the below URL and will be referred in the bibliography as well :

*The World of Pokemons*. (2021, September 29). Kaggle. https://www.kaggle.com/hamdallak/the-world-of-pokemons

From the structure of the data set, the features, their types, and values can be determined.

|  |
| --- |
| *Figure 1: Structure of the data set.* |

Some of the data points from the summary of the data set are present below.

|  |
| --- |
| *Figure 1: Structure of the data set.* |

# **DATA PRE-PROCESSING AND CLEANING**

The first blush of the data set portrayed the data to be not clean. After an **Initial Data Analysis (IDA)** process assisted with some graph visualizations as well, I found out that the data set requires pre-processing and cleaning in it.

Data cleaning is an important & necessary factor in the data analysis process. As the famous quote says - *"Garbage In, Garbage Out."*

I proceeded with the now needed & necessary step of data cleaning using RStudio application on this data set to wrangle and eliminate all the garbage values which consisted of :

* Missing values
* Duplicate values

Some of the actions performed for data cleaning are :

* 1. The categorical features in the data set were checked for inconsistencies in them. If there were some found, we would need to normalize them using *if..elseif..else* condition with data manipulation.

|  |
| --- |
| *Figure 2: Inconsistencies Check on the feature 'Primary Type' in the data set.* |

* 1. We checked the data set for NA and NULL values in its features using the functions which can be referred from the below snapshot.

|  |
| --- |
| *Figure 3: Checking records with NA, NULL values from the data set.* |

* 1. There were many data points with missing or blank values in the features 'Name2' and 'Secondary.Type'. Removing these data points from the data set made no sense as it would have lost us a majority of the data. Instead, the missing or empty values were replaced by the word "Unknown" using a function ***GSUB()***.

|  |
| --- |
| *Figure 4: Eliminating records with NA, missing values from the data set.* |

* 1. The data set was checked for duplicate values in the combination of 'Name' and 'Name2' because of their uniqueness when combined. We found some duplicate values from the output below where "TRUE" s written.

|  |
| --- |
| Text  Description automatically generated  *Figure 5: Check for Duplication in the combination of features from the data set.* |

* 1. The data set was checked for duplicate values in the combination of 'Name' and 'Name2' because of their uniqueness when combined. We found some duplicate values from the output below where "TRUE" is written.

|  |
| --- |
| *Figure 6: Retrieving the duplicate records from the data set.* |

* 1. The duplicate values found in the above step were eliminated from the data set using **'filter'** function of the ***dplyr*** library.

|  |
| --- |
| *Figure 7: Eliminating the duplicate records from the data set.* |

* 1. After performing the previous steps, the data set was omitted for NA values if present using the below function **'na.omit'**.

|  |
| --- |
| *Figure 8: Removing the NA, missing values from the data set.* |

* 1. The data set do not contain any kind of unbalanced features.

# **EXPLORATORY DATA ANALYSIS**

The features of the data set can be summarised to calculate the statistics -

|  |
| --- |
| *Figure 9: Summary of the data set.* |

The following observations can be made using the statistics found in summary of the data set -

1. The **mean** of the attribute *Attack* is around **80.47** with a **standard deviation** of **32.41** and **quartiles value (Lower Quartile - 55, Higher Quartile - 100).**   
   From the observations, we can anticipate that the data points around **maximum value (190)** can be *outliers* to the feature.
2. The **mean** of the attribute *Defense* is around **74.66** with a **standard deviation** of **31.24** and **quartiles value (Lower Quartile - 50, Higher Quartile - 90).**   
   From the observations, we can anticipate that the data points around **maximum value (250)** can be *outliers* to the feature.
3. The **mean** of the attribute *HP (Hit Point)* is around **70.07** with a **standard deviation** of **26.67** and **quartiles value (Lower Quartile - 50, Higher Quartile - 82).**   
   From the observations, we can anticipate that the data points around **maximum value (255)** can be *outliers* to the feature.
4. The **mean** of the attribute *Sp.Attack (Special Attack)* is around **73.02** with a **standard deviation** of **32.73** and **quartiles value (Lower Quartile - 50, Higher Quartile - 95).**   
   From the observations, we can anticipate that the data points around **maximum value (194)** can be *outliers* to the feature.
5. The **mean** of the attribute *Sp.Defense (Special Defense)* is around **72.29** with a **standard deviation** of **28.07** and **quartiles value (Lower Quartile - 50, Higher Quartile - 90).**   
   From the observations, we can anticipate that the data points around **maximum value (250)** can be *outliers* to the feature.
6. The **mean** of the attribute *Speed* is around **68.80** with a **standard deviation** of **30.21** and **quartiles value (Lower Quartile - 45, Higher Quartile - 90).**   
   From the observations, we can anticipate that the data points around **maximum value (200)** can be *outliers* to the feature.
7. The **mean** of the attribute *Total (Sum of all attributes)* is around **439.32** with a **standard deviation** of **121.97** and **quartiles value (Lower Quartile - 330, Higher Quartile - 515).**   
   From the observations, we can anticipate that the data points around **maximum value (1125)** can be *outliers* to the feature.

The normality of tall he features of the data set was checked in order to understand the type of distribution -

|  |
| --- |
| *Figure 9: Density Plot of features of the data set.* |
| *Figure 10: Shapiro-Wilks Test of the data set.* | |

The density plots of the features gave out the below results about the features of the data set -

1. The density graphs of all the features show that they can be considered as normally distributed.
2. **Shapiro-Wilks Test** was used to verify the normality of a feature of the data set. But, since the p-value was less than 5%, we could verify the normality of the feature.
3. Since, the normality of the data set could not be verified. Therefore, we cannot proceed further with the assumption of normal distribution and the calculation of Skewness and Kurtosis values of these features would not be significant anymore here.

**Primary Type Frequency Distribution**

A Frequency Distribution Table was created for the feature 'Primary Type' as all the Pokémon belong to some of the species available. I used the **factor()** function to check the factor levels of this feature and decided to create a frequency table based on the number of levels and data points available using **table() / count()** function.

|  |
| --- |
| *Figure 10: To check how many factor levels are present for the feature.* |
| *Figure 11: Factor Levels of the 'Primary Type' feature.* |
| A picture containing text, receipt  Description automatically generated  *Figure 12: Frequency Table of the 'Primary Type' feature.* |

1. We can figure out that there are **18 factor levels** of the **Primary Type feature**.
2. The **most occurrences of Pokémon belong to the Primary Type - WATER** with the value of **134.**
3. The **least occurrences of Pokémon belong to the Primary Type - FLYING** with the value of **8.**

|  |
| --- |
| **Chart, bar chart  Description automatically generated**  *Figure 13: Plot of Primary Type Frequency Distribution* |

1. The **highest number of Pokémon** in the data set belong to the **WATER** *species (Primary Type)* with the value of **134.**
2. The **2nd highest number of Pokémon** in the data set belong to the **NORMAL** *species (Primary Type)* with the value of **115.**
3. The **least number of Pokémon** in the data set belong to the **FLYING** *species (Primary Type)* with the value of **8** only**.** *They are abnormally very low in data points present in the data set.*

**Primary Type vs Summation of Attributes of Pokémon**

The below plot can help in visualising the summation of all the attributes of Pokémons grouped according to their Primary Type (Species)

|  |
| --- |
| *Figure 11: Box Plot of Bakery Items vs Promotion Types* |

1. From the grouped line graph, we can figure out that Pokémons of the *primary type (Species)* - **WATER** in the data set have the **cumulative largest abilities** in terms of all the attributes (*Attack, Defense, Hit Point, Special Attack, Special Defense, Speed*) amongst all the other species.
2. Pokémons of the *primary type (Species)* - **FLYING** in the data set have the **cumulative least abilities** in terms of all the attributes (*Attack, Defense, Hit Point, Special Attack, Special Defense, Speed*) amongst all the other species.
3. Pokémons of the *primary type (Species)* - **NORMAL and PSYCHIC** in the data set have the **2nd highest and 3rd cumulative highest abilities respectively** in terms of all the attributes (*Attack, Defense, Hit Point, Special Attack, Special Defense, Speed*) amongst all the other species. *They are very near to each other in terms of abilities.*
4. The depiction also shows that almost all the Pokémon Types (Species) have all of their abilities in accordance with each other amongst themselves.

**Scatterplot Matrix between Abilities of Pokémon Species :**

We can plot, visualize, and figure out the correlations (Positive, Negative, No correlation) between the abilities (features) of Pokémons using the Scatterplot Matrix. This matrix infographic provides the correlations of each abilities in accordance with all the other abilities in order to determine the patterns in them.

|  |
| --- |
| *Figure 12: Scatter Plot Matrix of abilities (attributes)* |

1. We can figure out that Attack is positively correlated to Defense in our data set. The famous statement also holds true in our data set that - *Attack is the best defense.*It seems that those Pokémons who are better at attacking usually develop defensive techniques as well and usually do rely on attacks for defense as well.
2. But, the same cannot be told about the inverse correlation between Defense and Attack. Those Pokémons who are defensive in nature usually do not develop very high attacking prowess.
3. The *Attack, Hit Point* prowess or abilities of Pokémons are also positively correlated to each other for the attacking style Pokémons.
4. We can figure out the types of correlations between all the abilities (Attributes) of the Pokémons from this scatterplot matrix. The types or correlations are :
   1. No Correlation
   2. Positive Correlation, and Negative Correlations

**To find out the Most Powerful Pokémon Species : Type ~ Mean Total Abilities**

We have applied the mean function on the *Total* ability feature which represents the cumulation of all the abilities belonging to a Pokémon. Using mean of this feature for Pokémons belonging to specific *Primary Types (Species*), we can figure out the most powerful and least powerful species in the Pokémon world.

|  |
| --- |
| *Figure 14: Plot of Total Bakery Items Yearly* |

1. It can be observed that the mean of total abilities of the *species* - **DRAGON** exceed all others and can be considered to be the **most powerful species** in the Pokémon world. The total mean value of abilities is around **536**.
2. It can be observed that the mean of total abilities of the *species* - **BUG** is below all others and can be considered to be the **least powerful species** in the Pokémon world. The total mean value of abilities is around **384.**

# **CONCLUSION**

The dataset of 'The World of Pokémons' has provided various insights about the types, abilities of the Pokémons and the patterns between them as well. We performed initial data analysis, exploratory data analysis, calculated various statistics, plotted several visualisation graphs in order to understand the analysis properly. The below points can be inferred from the analysis :

* The **most occurrences of Pokémons belong to the Primary Type - WATER** with the value of **134** and the **least occurrences of Pokémon belong to the Primary Type - FLYING** with the value of **8.**
* Since, the Pokémons with the least occurrences are abnormally low in the data set and we do not have any other information, we can *put up an anticipation that FLYING type Pokémons are very rare in the world of Pokémon*s**. This can only be answered properly if more data or information gets available related to it.**
* The dataset contains the species of WATER type Pokémons whose accumulated abilities are the **Largest/Highest** amongst all the other species of Pokémons present in the data set.   
  All the abilities get accumulated around 10,000 of Water species Pokémons which is very high when compared to other accumulated abilities of Pokémons present in the data set.
* It can be concluded that the **most powerful species** in the Pokémon world is **DRAGON**. The total mean value of abilities is around **536**.*In most cases, dragon species Pokémons would come out as winners in the fights.*
* It can be observed that the mean of total abilities of the *species* - **BUG** is below all others and can be considered to be the **least powerful species** in the Pokémon world. The total mean value of abilities is around **384.**
* It can be concluded that the **least powerful species** in the Pokémon world is **BUG**. The total mean value of abilities is around **384**.  
  *In most cases, bug species Pokémons would come out as losers in the fights.*

# **BIBLIOGRAPHY**

1. *The World of Pokemons*. (2021, September 29). Kaggle. https://www.kaggle.com/hamdallak/the-world-of-pokemons
2. *Guide To Data Cleaning: Definition, Benefits, Components, And How To Clean Your Data*. (2021). Tableau. https://www.tableau.com/learn/articles/what-is-data-cleaning
3. Burns, E. (2021, January 16). *Data Cleaning in R Made Simple - Towards Data Science*. Medium. https://towardsdatascience.com/data-cleaning-in-r-made-simple-1b77303b0b17
4. *Find duplicated rows (based on 2 columns) in Data Frame in R*. (2011, August 8). Stack Overflow. https://stackoverflow.com/questions/6986657/find-duplicated-rows-based-on-2-columns-in-data-frame-in-r
5. *Normality Test in R - Easy Guides - Wiki - STHDA*. (2021). STHDA. http://www.sthda.com/english/wiki/normality-test-in-r
6. Nguyen, C. (2021, September 29). *Guide To Data Visualization With ggplot2 - Towards Data Science*. Medium. https://towardsdatascience.com/guide-to-data-visualization-with-ggplot2-in-a-hour-634c7e3bc9dd
7. *A Grammar of Data Manipulation*. (2021). Dplyr. https://dplyr.tidyverse.org/
8. *ggplot2 - Essentials - Easy Guides - Wiki - STHDA*. (2021). GGPLOT2 | STHDA. http://www.sthda.com/english/wiki/ggplot2-essentials
9. *Home - RDocumentation*. (2021). Functions in R - Documentation. https://www.rdocumentation.org/
10. Holtz, Y. (2021). *The R Graph Gallery – Help and inspiration for R charts*. The R Graph Gallery. https://www.r-graph-gallery.com/index.html

# **APPENDIX**

#---------------------- Week\_1\_Module\_1 R Script ----------------------#

print("Author : Harshit Gaur")

print("Week 1 Assignment - Module 1 R Pratice")

# Importing the packages.

listOfPackages <- c(

"dplyr", "tidyr", "plyr", "tidyverse", "RColorBrewer", "plotrix", "scales", "ggplot2",

"data.table", "reshape", "gridExtra", "vtable", "moments", "ggpubr"

)

for (package in listOfPackages) {

if (package %in% rownames(installed.packages()) == FALSE)

{ install.packages(package) }

# Importing the package.

library(package, character.only = TRUE)

}

# STEP 2: Import data set

# Note: Change the working directory as per the file's location.

setwd("/Users/HarshitGaur/Documents/Northeastern University/MPS Analytics/ALY 6010/Class 1/Assignment/")

pokemon\_dataset <- read.csv("pokemons dataset.csv", header = TRUE)

# Display the data set.

View(pokemon\_dataset)

# Print the structure of 'pokemons dataset.csv' data set

str(pokemon\_dataset)

# Print the summary of 'pokemons dataset.csv' data set

summary(pokemon\_dataset)

st(pokemon\_dataset)

#------------------- Data Cleaning -------------------#

# To check inconsistencies in the 'Primary Type - Character' feature of the data set.

unique(pokemon\_dataset$Primary.Type)

# To check NaN, NULL values in the data set.

sum(is.na(pokemon\_dataset))

sum(is.null(pokemon\_dataset))

# Replacing Empty Values in the features with the word 'Unknown'

pokemon\_dataset$Name2 <- gsub('^$', 'Unknown', pokemon\_dataset$Name2)

pokemon\_dataset$Secondary.type <- gsub('^$', 'Unknown', pokemon\_dataset$Secondary.type)

# Check the duplicate values in a combination of 2 features.

duplicated(pokemon\_dataset[,1:2])

# Retrieving the duplicated records from the data set.

pokemon\_dataset[which(duplicated(pokemon\_dataset[,1:2])),]

# Eliminating the duplicated records using the indexes provided by the above step.

pokemon\_dataset <- pokemon\_dataset %>% filter( !row\_number() %in% 44)

# Removing 'NA, Missing Values' from the data set.

pokemon\_dataset <- na.omit(pokemon\_dataset)

#------------------- Exploratory Data Analysis -------------------#

# Check Normality using Density Graphs of all the univariates.

normality\_attack <- ggdensity(pokemon\_dataset$Attack, main = "Density plot of Attack", xlab = "Attack")

normality\_defense <- ggdensity(pokemon\_dataset$Defense, main = "Density plot of Defense", xlab = "Defense")

normality\_hp <- ggdensity(pokemon\_dataset$HP, main = "Density plot of Hit Points", xlab = "Hit Points")

normality\_spAttack <- ggdensity(pokemon\_dataset$Sp.Attack, main = "Density plot of Special Attack", xlab = "Special Attack")

normality\_spDefense <- ggdensity(pokemon\_dataset$Sp.Defense, main = "Density plot of Special Defense", xlab = "Special Defense")

normality\_speed <- ggdensity(pokemon\_dataset$Speed, main = "Density plot of Speed", xlab = "Speed")

normality\_total <- ggdensity(pokemon\_dataset$Total, main = "Density plot of Total Attributes", xlab = "Total Attributes")

grid.arrange(normality\_attack,normality\_defense,normality\_hp,normality\_spAttack,normality\_spDefense,normality\_speed,normality\_total)

# Check Normality using Shapiro-Wilks Test

shapiro.test(pokemon\_dataset$Attack)

# Check Normality using Q-Q Plot of a single Feature.

qqnorm(pokemon\_dataset$Attack);

qqline(pokemon\_dataset$Attack, col="green", lwd=2)

# Check Skewness of the features

skewness(pokemon\_dataset$Attack)

skewness(pokemon\_dataset$Defense)

skewness(pokemon\_dataset$HP)

skewness(pokemon\_dataset$Sp.Attack)

skewness(pokemon\_dataset$Sp.Defense)

skewness(pokemon\_dataset$Speed)

skewness(pokemon\_dataset$Total)

# Check Skewness of the features

kurtosis(pokemon\_dataset$Attack)

kurtosis(pokemon\_dataset$Defense)

kurtosis(pokemon\_dataset$HP)

kurtosis(pokemon\_dataset$Sp.Attack)

kurtosis(pokemon\_dataset$Sp.Defense)

kurtosis(pokemon\_dataset$Speed)

kurtosis(pokemon\_dataset$Total)

# Check if the factors of 'Primary Type' can be used for Frequency table.

factor(pokemon\_dataset$Primary.Type)

# Create a Frequency Table

freqTable\_primaryType <- count(pokemon\_dataset, 'Primary.Type')

colnames(freqTable\_primaryType) <- c("Primary Type", "Frequency")

# Plot an bar graph for 'Primary Type' using 'ggplot'

# ----------- Plot 1: Primary Type Distribution ----------- #

par(mar = c(2, 6, 3, 2) + 0.1)

ggplot(freqTable\_primaryType)+

geom\_bar(

mapping = aes(

x = `Primary Type`,

y = Frequency

),

stat="identity"

) +

theme(panel.background = element\_rect("#a7f4fc")) +

labs(

title = "Plot 1: Primary Type Distribution",

x = "Primary Type",

y = "Frequency Count"

) +

theme(

plot.title = element\_text(hjust = 0.5, colour = "#7F3D17", face = "bold"),

axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1, size=15)

) +

scale\_y\_continuous(labels = comma) +

geom\_text(aes(x = `Primary Type`, y = Frequency, label = Frequency), vjust=1.8, color = "white", size=4)

# Plot an Grouped Line graph for attack using 'ggplot'

# ----------- Plot 2: Primary Type vs Summation of Attributes ----------- #

# Summation of 'Attributes' according to Primary Type feature

summationTable <- pokemon\_dataset %>% group\_by(pokemon\_dataset$Primary.Type) %>% dplyr::summarise(

sum\_attack = sum(Attack),

sum\_defense = sum(Defense),

sum\_hp = sum(HP),

sum\_spAttack = sum(Sp.Attack),

sum\_spDefense = sum(Sp.Defense),

sum\_speed = sum(Speed),

sum\_total = sum(Total),

)

# Converting the summation Table into a Data Frame.

summationTable <- data.frame((summationTable))

par(mar = c(4, 10, 8, 4) + 0.1)

ggplot(summationTable, aes(x = pokemon\_dataset.Primary.Type, group = 1)) +

geom\_line(aes(y = sum\_attack, color="Attack"), linetype="dotted", size=1) +

geom\_line(aes(y = sum\_defense, color="Defense"), linetype="longdash", size=1) +

geom\_line(aes(y = sum\_hp, color="HP"), linetype="F1", size=1) +

geom\_line(aes(y = sum\_spAttack, color="Special Attack"), linetype="F1", size=1) +

geom\_line(aes(y = sum\_spDefense, color="Special Attack"), linetype="solid", size=1) +

geom\_line(aes(y = sum\_speed, color="Speed"), linetype="twodash", size=1) +

scale\_color\_manual(

name = "Legends",

values = c("Attack" = "red", "Defense" = "blue", "HP" = "yellow", "Special Attack" = "orange", "Special Defense" = "Darkgreen", "Speed" = "magenta")) +

theme(

panel.background = element\_rect("#c9f5e4"),

axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1, size = 12)

) +

labs(

title = "Plot 2: Primary Type vs Summation of Attributes",

x = "Primary Type",

y = "Summation of Attributes"

) + theme(plot.title = element\_text(hjust = 0.5, colour = "#7F3D17", face = "bold"))

# Plot an bar graph for 'Total Attributes' using 'ggplot'

# ----------- Plot 3: Total Attributes Distribution ----------- #

meanTable <- pokemon\_dataset %>% group\_by(pokemon\_dataset$Primary.Type) %>% dplyr::summarise(

mean\_attack = mean(Attack),

mean\_defense = mean(Defense),

mean\_hp = mean(HP),

mean\_spAttack = mean(Sp.Attack),

mean\_spDefense = mean(Sp.Defense),

mean\_speed = mean(Speed),

mean\_total = mean(Total),

)

par(mar = c(2, 6, 3, 2) + 0.1)

ggplot(meanTable)+

geom\_bar(

mapping = aes(

x = `pokemon\_dataset$Primary.Type`,

y = mean\_total

),

stat="identity"

) +

theme(panel.background = element\_rect("#f2c8fa")) +

labs(

title = "Plot 5: Mean Total Attributes Distribution",

x = "Primary Type",

y = "Mean Total Attributes Count"

) + theme(

plot.title = element\_text(hjust = 0.5, colour = "#eb5e7a", face = "bold"),

axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1, size = 18)

) +

scale\_y\_continuous(labels = comma) +

geom\_text(aes(x = `pokemon\_dataset$Primary.Type`, y = mean\_total, label = floor(mean\_total)), vjust=1.8, color = "white", size=4)

# Plot an Scatter Plot Matrix graph for 'Attributes' using 'ggplot'

# ----------- Plot 4: Scatter Plot Matrix between Abilities (Attributes) ----------- #

pairs(~Attack + Defense + HP + Sp.Attack + Sp.Defense + Speed, pokemon\_dataset, col=c("BLUE", "RED"))