Pokemon task

Krzysztof Zabłocki, Sarbel Fadel, Ryan John

2024-05-13

```r
filelocation <- "Data_Pokemon1.csv" #Enter the location of the data of your computer here.
fdata <- read.csv(filelocation)

# Ensure the "Type" column is treated as a factor (needed for next steps)
fdata$Type <- as.factor(fdata$Type)

dim(fdata)
```

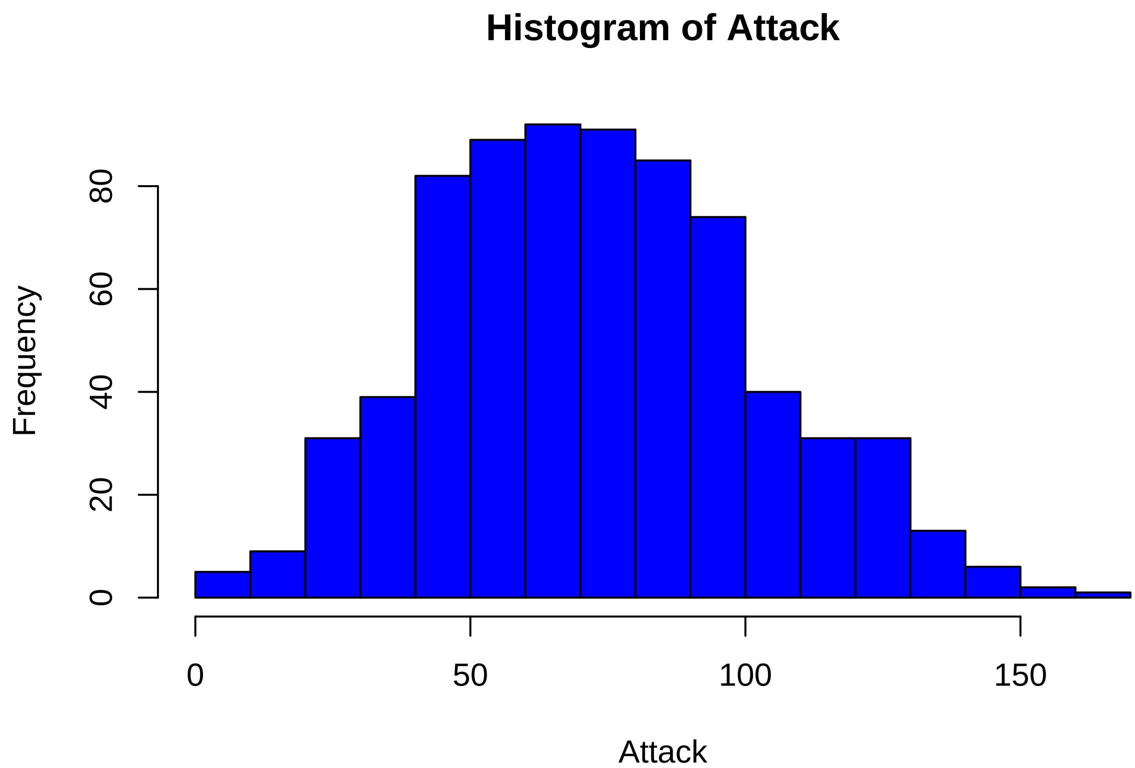
```
## [1] 721   8
```

```r
head(fdata)
```

```
##         Name  Type HP Attack Defense Sp.Attack Sp.Defense Speed
## 1  Bulbasaur Grass 45     49      49        65         65    45
## 2    Ivysaur Grass 60     62      63        80         80    60
## 3   Venusaur Grass 80     82      83       100        100    80
## 4 Charmander  Fire 39     52      43        60         50    65
## 5 Charmeleon  Fire 58     64      58        80         65    80
## 6  Charizard  Fire 78     84      78       109         85   100
```

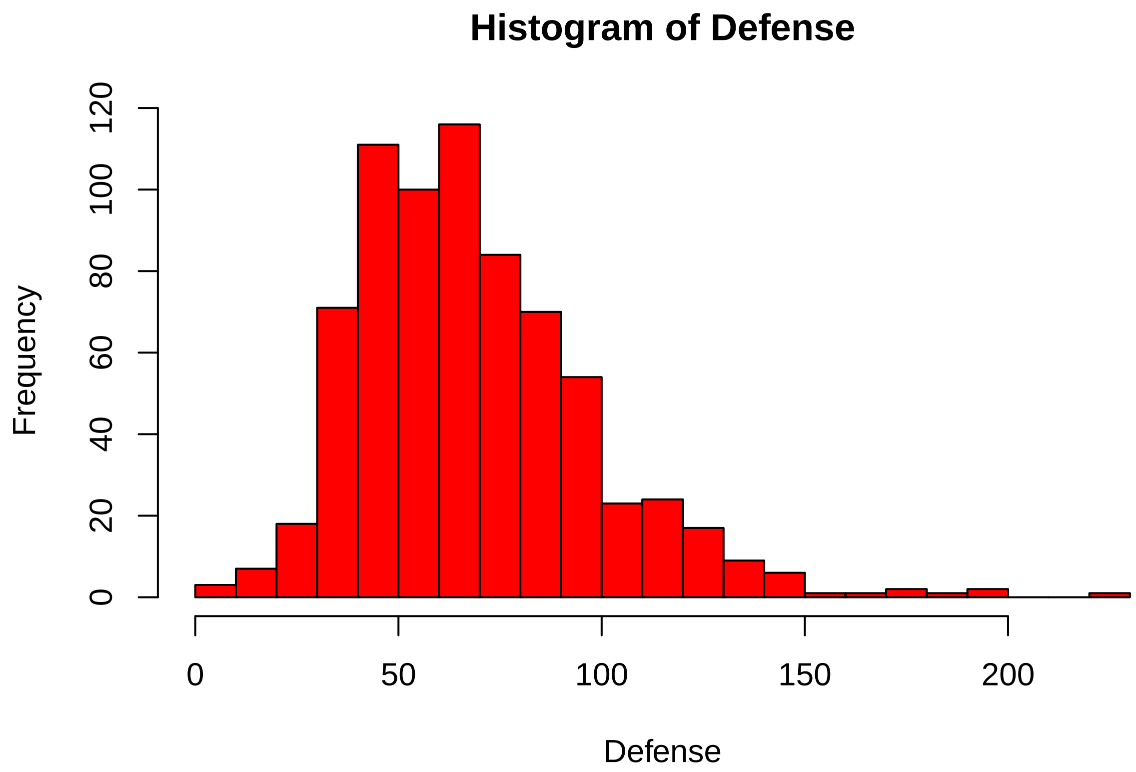
Histograms for Attack and Defense columns

```r
hist(fdata$Attack, main="Histogram of Attack", xlab="Attack", col="blue", breaks=20)
```



This graph shows the distribution of the Attack values of the Pokemon. The x-axis represents the Attack values, while the y-axis represents the frequency of each value. The graph shows that the most common Attack values are between 50 and 100, with a peak around 70. It seems like Gaussian distribution.

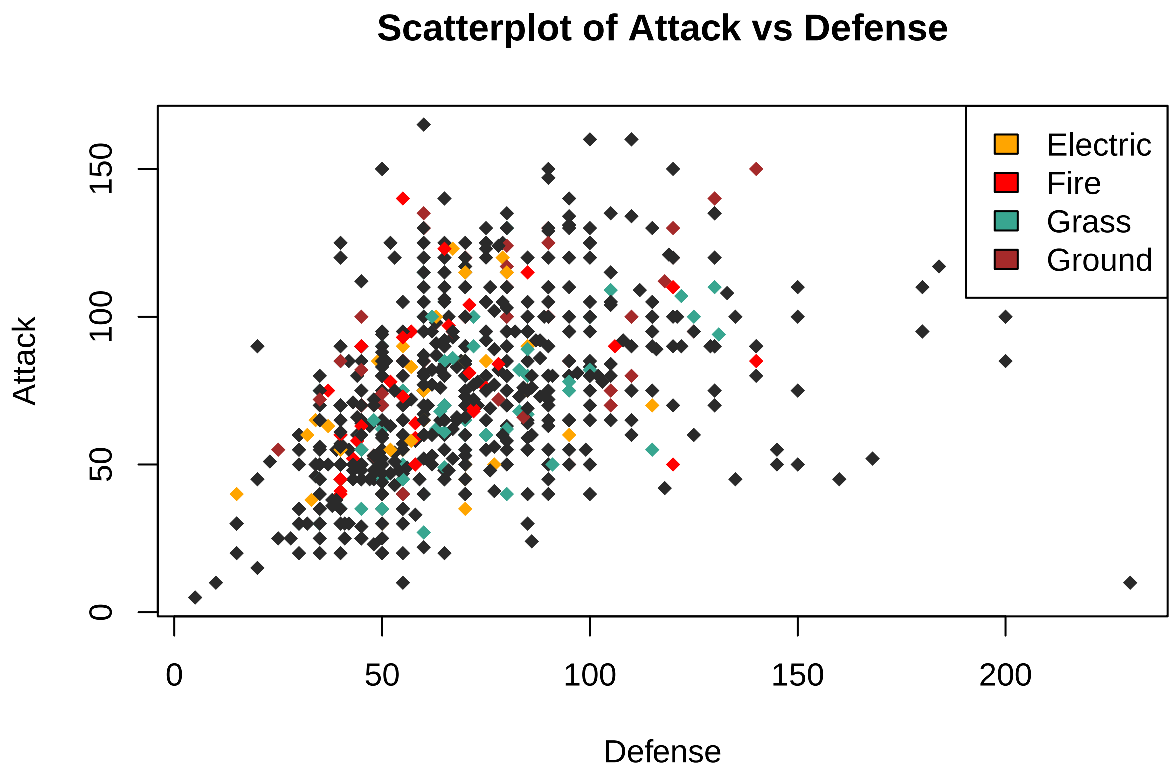
```r
hist(fdata$Defense, main="Histogram of Defense", xlab="Defense", col="red", breaks=20)
```



This graph shows the distribution of the Defense values of the Pokemon. The x-axis represents the Defense values, while the y-axis represents the frequency of each value. The graph is in my opinion still similar to the Gaussian distribution, but it is not as clear as in the case of the Attack values.

Scatter plot of Attack and Defense:

```
colors <- ifelse(fdata$Type == "Electric", "orange",
          ifelse(fdata$Type == "Fire", "red",
          ifelse(fdata$Type == "Grass", "#38a690",
          ifelse(fdata$Type == "Ground","brown","#2a2a2a"))))
plot(fdata$Defense, fdata$Attack,
     main="Scatterplot of Attack vs Defense",
     xlab="Defense",
     ylab="Attack",
     pch=18,          # This is a solid square
     cex=1,          # This makes the points the default size
     col=colors)     # Points are colored
legend("topright", legend = c("Electric","Fire","Grass","Ground"), fill = c("orange", "red","#38a690","
```



This one is a positive corleation and it looks like a y=x graph with some outlying dots (it seems that pokemons attacks and defense stats are somewhat close to each other).

Counting the number of unique types of Pokémon:

```
# Calculate the number of unique types
num_types <- length(levels(fdata$Type))

# Print the number of unique Pokémon types
print(paste("There are", num_types, "unique types of Pokémon."))
```

```
## [1] "There are 18 unique types of Pokémon."
```

Most and least common Pokémon types:

```
# Create a frequency table of Pokémon types
type_table <- table(fdata$Type)

# Find the index of the most common type
most_common_index <- which.max(type_table)
# Find the index of the least common type
least_common_index <- which.min(type_table)

# Extract the most and least common types using the indices
most_common_type <- names(type_table)[most_common_index]
least_common_type <- names(type_table)[least_common_index]
```

```r
# Print the most and least common types and their frequencies
print(paste("The least common type is:", least_common_type, "with", type_table[least_common_index], "occ
```

```
## [1] "The least common type is: Flying with 3 occurrences."
```

```r
print(paste("The most common type is:", most_common_type, "with", type_table[most_common_index], "occur
```

```
## [1] "The most common type is: Water with 105 occurrences."
```

Correlation test between Defense and Attack:

```r
# Perform a correlation test between Defense and Attack
correlation_test <- cor.test(fdata$Defense, fdata$Attack)

# Print the summary of the correlation test
print(correlation_test)
```

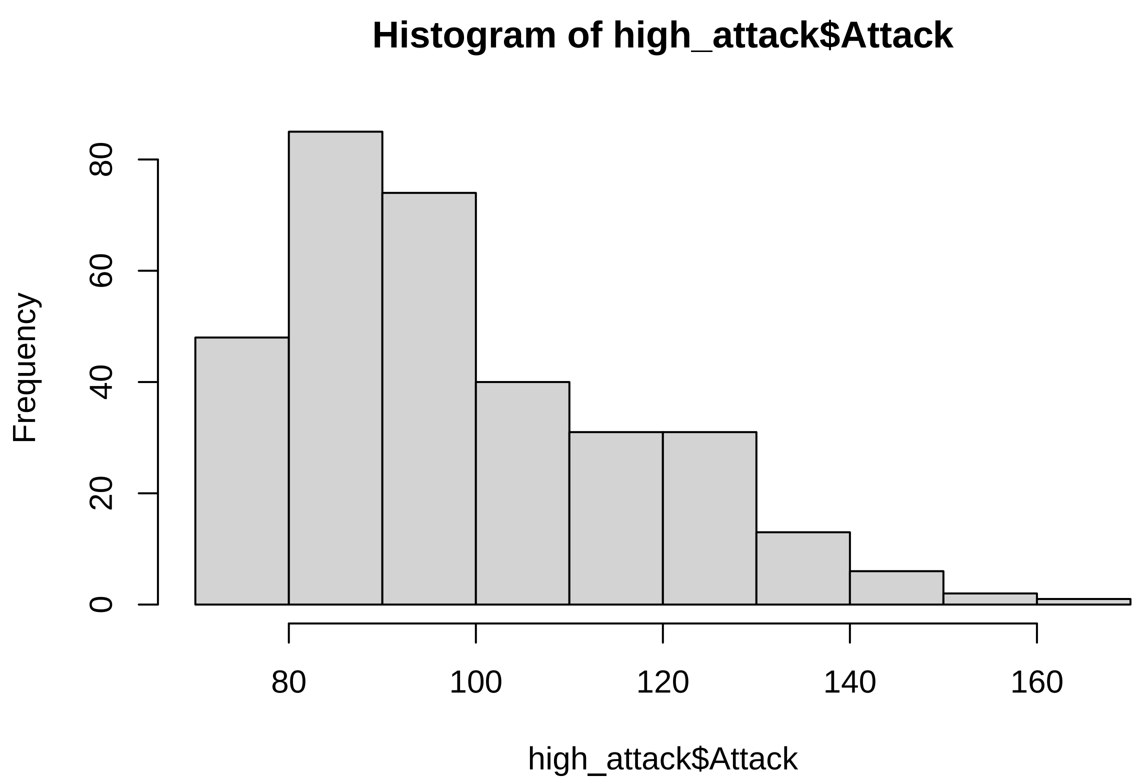
```
##
##  Pearson's product-moment correlation
##
## data:  fdata$Defense and fdata$Attack
## t = 12.918, df = 719, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3728215 0.4914624
## sample estimates:
##       cor
## 0.4340218
```

#####Statistical test & Subset

```r
# Subset the data based on the type of Pokémon
high_attack <- fdata[which(fdata$Attack > 75), ]
dim(high_attack)
```

```
## [1] 331   8
```

```r
hist(high_attack$Attack)
```



Subsets of the data based on the type of Pokémon (Water and Fire):

```r
water <- fdata[which(fdata$Type == "Water"),]
fire <- fdata[which(fdata$Type == "Fire"),]

# Perform the Wilcoxon rank-sum test
wilcox.test(water$Attack, fire$Attack, alternative = "two.sided")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  water$Attack and fire$Attack
## W = 1915.5, p-value = 0.02783
## alternative hypothesis: true location shift is not equal to 0
```

The above test compares the Attack points between Water and Fire Pokémon. The p-value is less than 0.05, which indicates that there is a significant difference between the Attack points of Water and Fire Pokémon.

```
boxplot(fdata$Attack[fdata$Type=="Water"], fdata$Attack[fdata$Type=="Fire"],
        ylab="Attack", names=c("Water", "Fire"), col=c("#3369ff","red"))
```