

615_Assignment3

Xinyi Wang, Hao Qin, Qixuan Zhang, Shiyu Zhang

9/30/2018

Read Data and Data Cleaning

```
#Read data
mydata <- read.csv("pokemon.csv")
#Data Cleaning
names(mydata)

## [1] "X."      "Name"    "Type.1"  "Type.2"  "HP"
## [6] "Attack"  "Defense" "Sp..Atk" "Sp..Def"  "Speed"
## [11] "Generation" "Legendary" "Level"

Attack <- mydata %>% filter(Attack >= 20)
dim(mydata$Speed)

## NULL

head(mydata %>% arrange(desc(Defense)))

##      X.      Name Type.1 Type.2 HP Attack Defense Sp..Atk Sp..Def Speed
## 1 225 Mega Steelix Steel Ground 75 125 230 55 95 30
## 2 231 Shuckle Bug Rock 20 10 230 10 230 5
## 3 334 Mega Aggron Steel 70 140 230 60 80 50
## 4 224 Steelix Steel Ground 75 85 200 55 65 30
## 5 415 Regirock Rock 80 100 200 50 100 50
## 6 790 Avalugg Ice 95 117 184 44 46 28
##      Generation Legendary Level
## 1 2 FALSE 2
## 2 2 FALSE 1
## 3 3 FALSE 2
## 4 2 FALSE 2
## 5 3 TRUE 3
## 6 6 FALSE 3
```

Variable description

The dataset we use is from Kaggle “Pokemon With Stats”. Below is brief description of all variables.

```
knitr::include_graphics('variable.png')
```

Columns

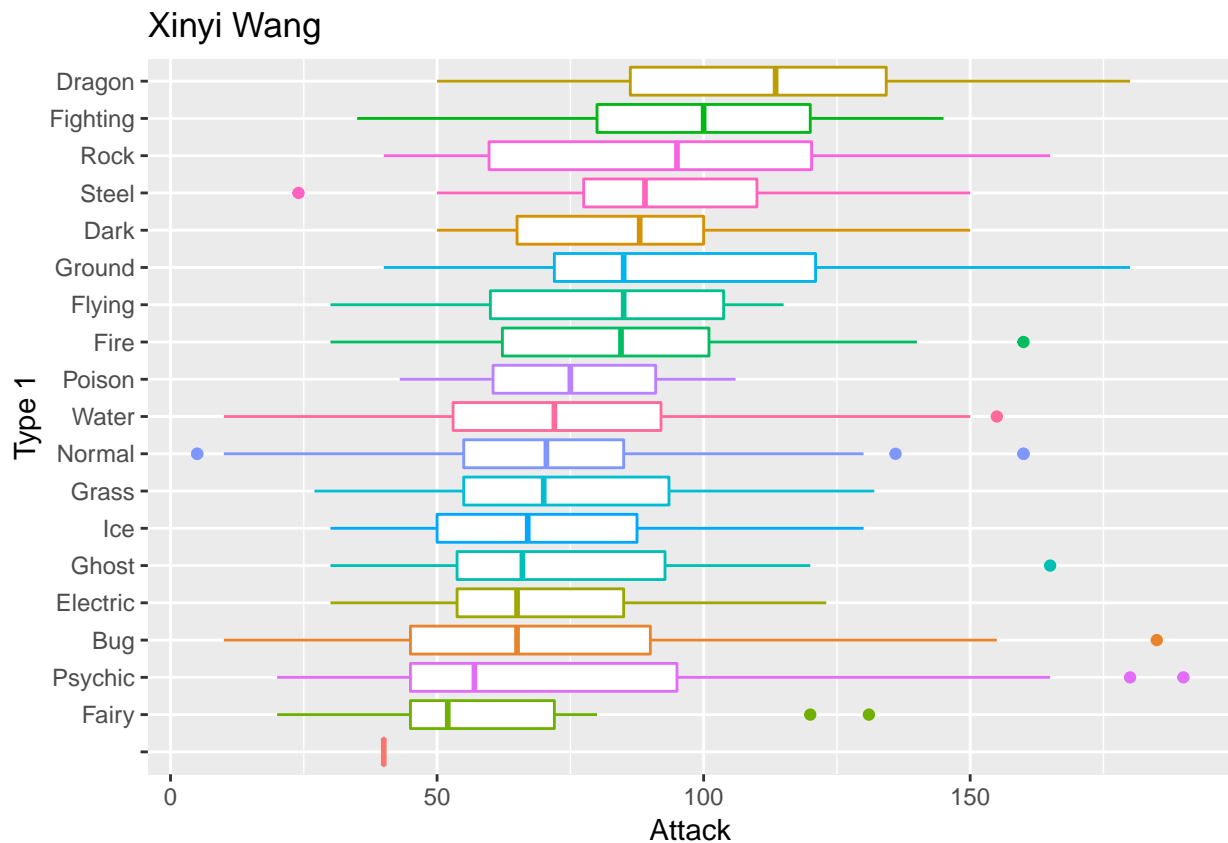
[Edit](#)

- # # PokeDex index number
- A Name Name of the Pokemon
- A Type 1 Type of pokemon
- A Type 2 Other Type of Pokemon
- # Total Sum of Attack, Sp. Atk, Defense, Sp. Def, Speed and HP
- # HP Hit Points
- # Attack Attack Strength
- # Defense Defensive Strength
- # Sp. Atk Special Attack Strength
- # Sp. Def Special Defensive Strength
- # Speed Speed
- # Generation Number of generation
- ✓ Legendary True if Legendary Pokemon False if not (more revision on mythical vs legendary needed)

Plots

1.Box plot

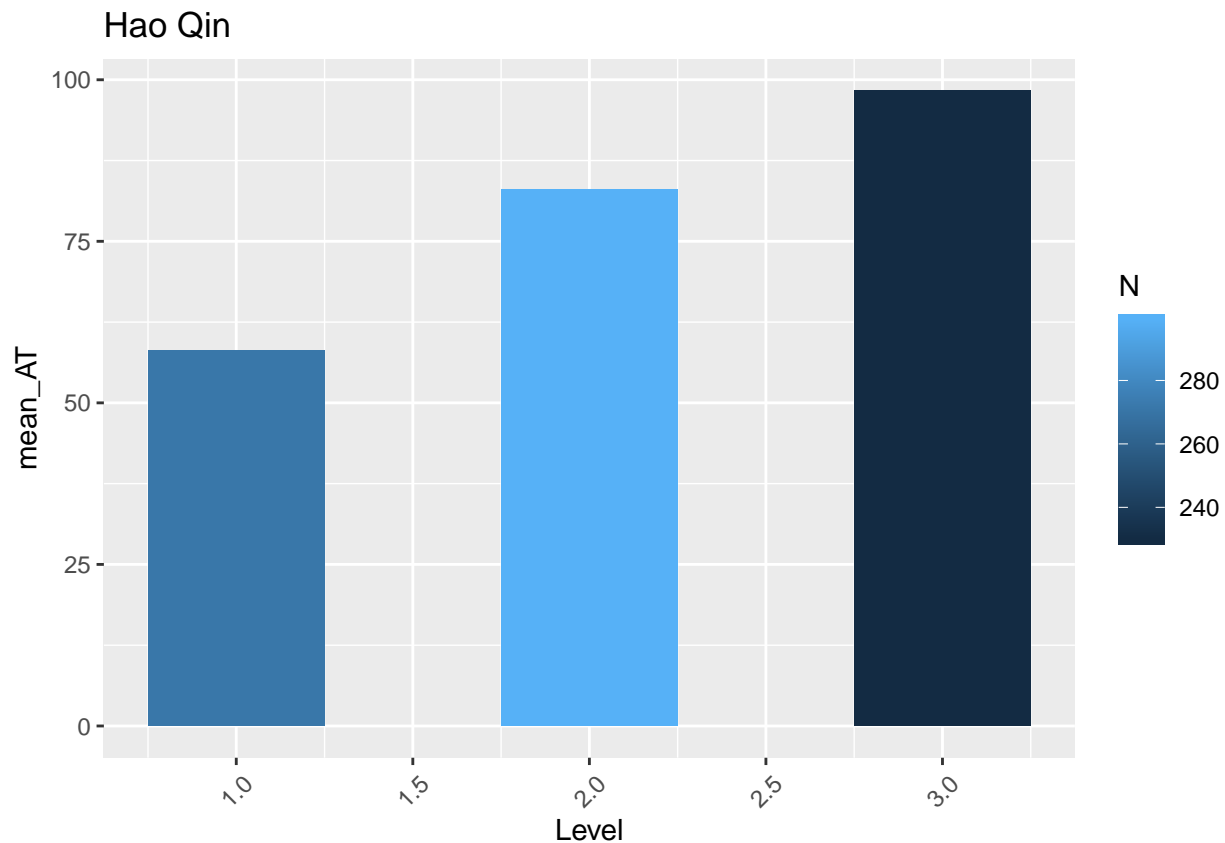
```
#str(mydata)
ggplot(mydata %>%
  group_by(Type.1) %>%
  mutate(med=median(Attack)),
  aes(x=reorder(Type.1, Attack, FUN=median), y=Attack,color = Type.1)) +
coord_flip() +
geom_boxplot(aes(fill=med)) +
scale_fill_gradient(low="white", high="white") +
labs(x="Type 1") +
theme(legend.position="none") +
ggtitle("Xinyi Wang")
```



This box plots shows the spread of all attack grouped by Type 1. We can say from the plot that Dragon type pokemons have an edge over the other types as they have a higher attacks compared to the other types.

2.Bar plot

```
HHHQ=mydata%>%
  group_by(Level)%>%
  summarise(mean_AT=mean(Attack), N=n())
ggplot(HHHQ, aes(x = Level,y =mean_AT, group = factor(1))) + geom_bar(stat = "identity", width = 0.5, a
```



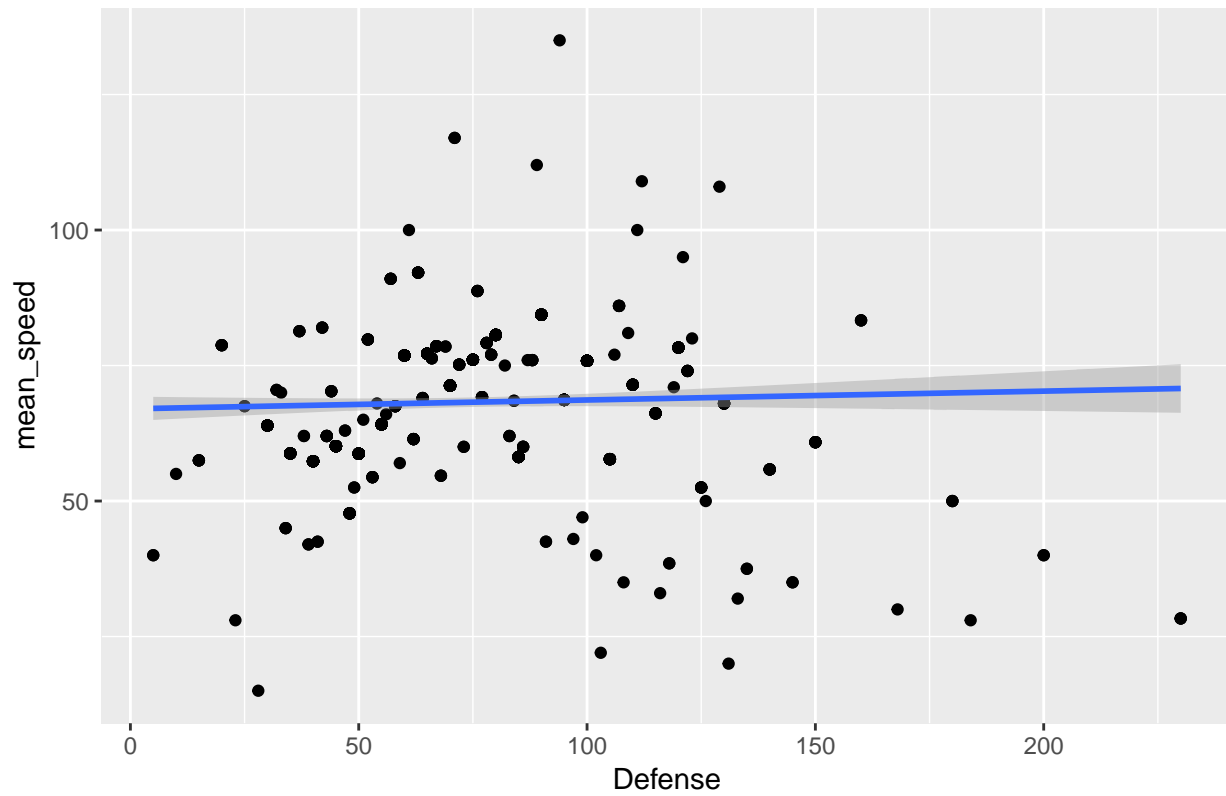
Secondly, we use the bar chart to show the relationship between the level and the average of attack, which grouped by the level. From the outcome, we can get the level3(HP>80) takes the largest proportion, and the level1(HP<60) takes the smallest proportion.

3.Scatter Plot & Bar Plot

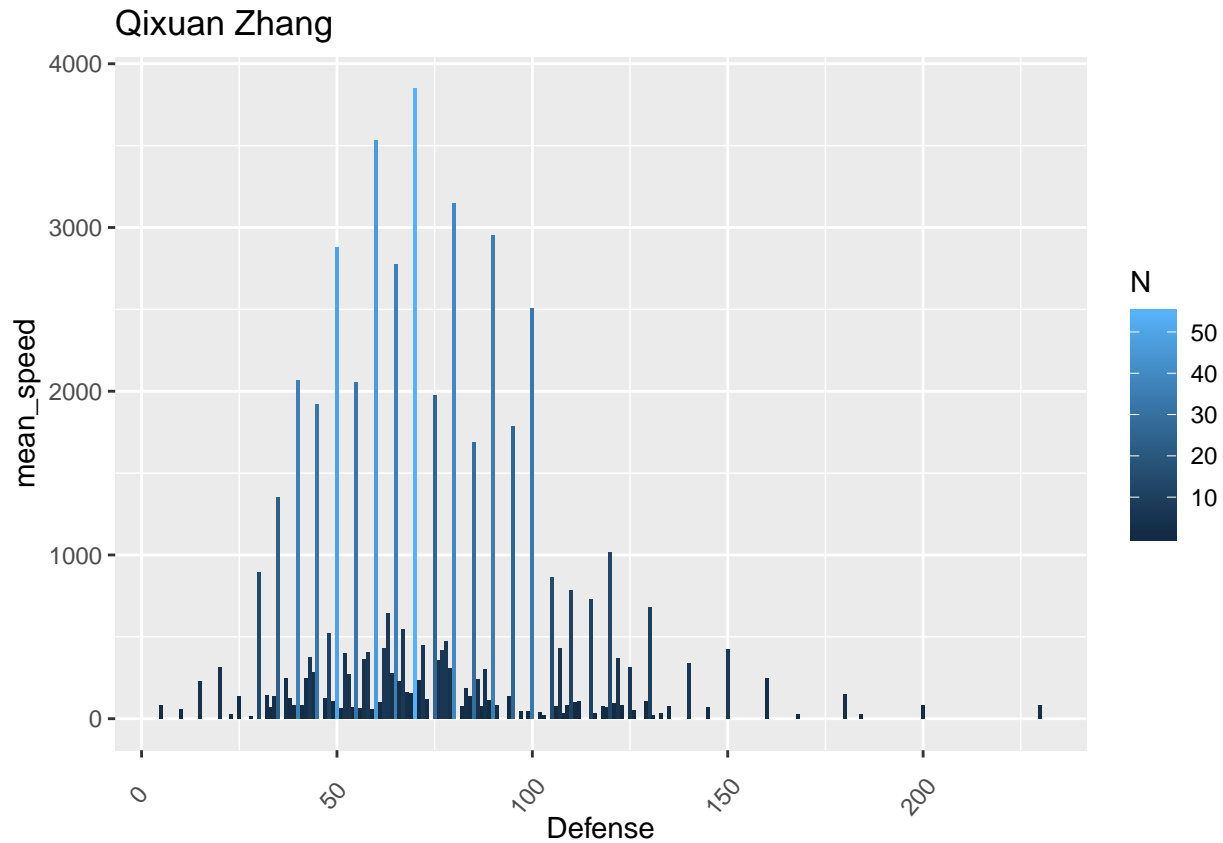
```
mydata1<-mydata %>%
  group_by(Defense) %>%
  mutate(mean_speed=mean(Speed),N=n())

ggplot(mydata1,aes(x=Defense,y=mean_speed))+
  geom_point()+ geom_smooth(method = "lm")+
  ggtitle("Qixuan Zhang")
```

Qixuan Zhang



```
ggplot(mydata1, aes(x = Defense, y = mean_speed, group = factor(2))) +  
  geom_col(width = 0.9, aes(fill=N)) +  
  theme(axis.text.x = element_text(angle = 50, hjust = 0.6, vjust = 0.3)) +  
  ggtitle("Qixuan Zhang")
```



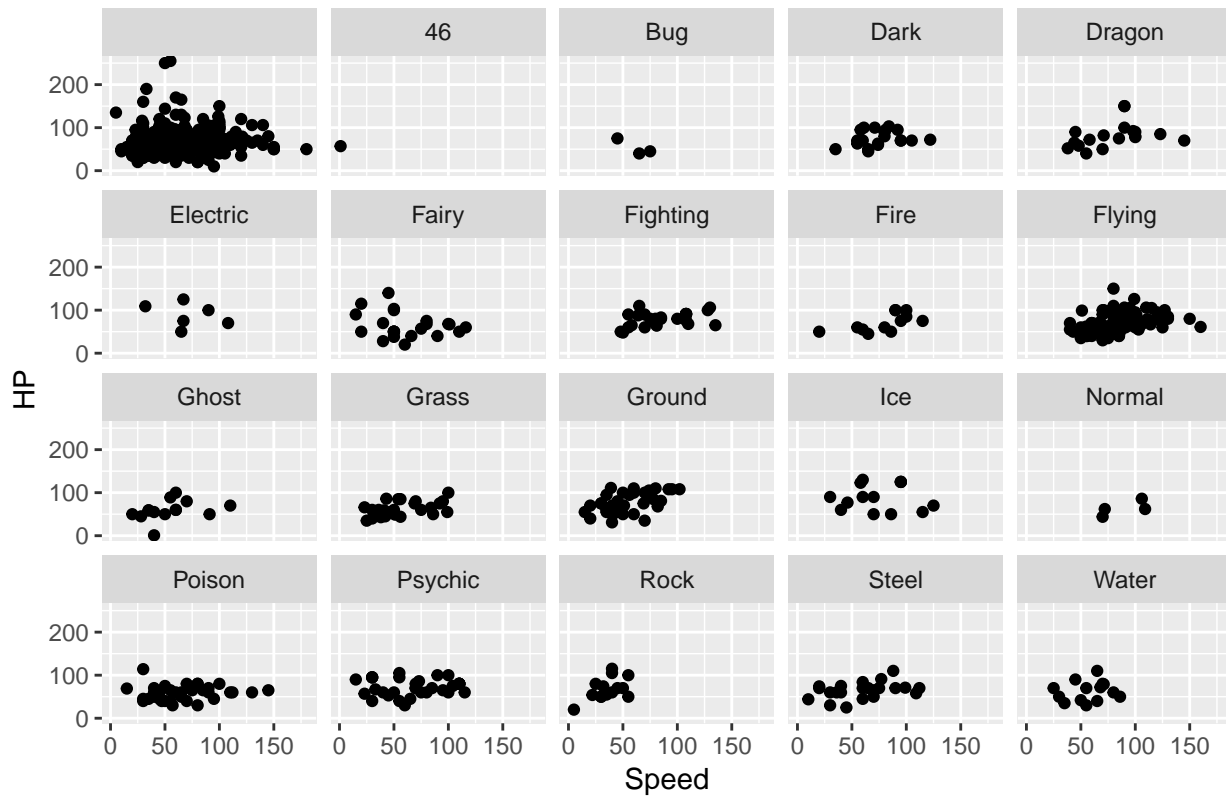
The scatter plot shows the relationship between defense and speed, we can see these two variables seem to show a linear relationship. Besides the scatter plot, we plot a bar plot to see the height of the bars to represent values in the data.

4. Three dimension Scatter Plot

```
mydata4<-mydata %>%
  group_by(Speed) %>%
  mutate(mean_hp=mean(HP), N=n())

ggplot(data = mydata4) +
  geom_point(mapping = aes(x = Speed, y = HP)) +
  facet_wrap(~ Type.2, nrow = 4)+ggtitle("Shiyu Zhang")
```

Shiyu Zhang



The three dimension plot shows that the relationship between speed, HP and type2. from the speed perspective, most of the speed from different type falls between 1 to 100, the 'rock' is a little bit slow (speed falls between 0-50), the 'fighting' is better in general since most of the speed lies between 50 and 150. From the HP perspective, we can see that almost all the data from type 2 lies between 0 and 100, we can say that the HP is not strongly related to the type of pokemon.