**Dataset: Pokémon and their stats**

## Description of Data

This data set includes 1045 Pokémon, including their number, name, and basic stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. It has been of great use when teaching statistics to kids. With certain types you can also give a geeky introduction to machine learning.

Name: Name of each Pokémon.

Total: Sum of all stats that come after this, a general guide to how strong a Pokémon is.

HP: Hitpoints, or health, defines how much damage a Pokémon can withstand before fainting.

Attack: The base modifier for normal attacks (eg. Scratch, Punch).

Defense: The base damage resistance against normal attacks.

SP Attack: Special attack, the base modifier for special attacks (e.g. fire blast, bubble beam).

SP Defence: The base damage resistance against special attacks.

Speed: Determines which pokemon attacks first in each round.

## 1.Read the csv data file.

pokemon\_data=read.csv(file=file.choose(),header = TRUE)

## 2.Display head of data

head(pokemon\_data)

## 3.Display first n rows specified.

head(pokemon\_data,n=10)

## 4.Display tail of data.

tail(pokemon\_data)

## 5.Display n rows specified from bottom

tail(pokemon\_data,n=10)

**6.Determining Type of data**

class(pokemon\_data)

## 7.Table command

## table(pokemon\_data$Attack)

## 

## table(pokemon\_data$HP)

## 8.Determine the structure of data

str(pokemon\_data)

## 9.Summarising the data

summary(pokemon\_data)

## 10.Displaying Dimension of the data

dim(pokemon\_data)

## 11.Displaying length of attack column

length(pokemon\_data$Attack)

## 12.Displaying column names of data

colnames(pokemon\_data)

## 13.Displaying structure of some columns in the data

class(pokemon\_data$Name)typeof(pokemon\_data$Name)

## 14.Displaying type of some data structure in the data

typeof(pokemon\_data$Total)

## 15.List of variables present in pokemon data

ls(pokemon\_data)

## 

## 16.Some pattern matching operations on variable of pokemon\_data

ls(pokemon\_data,pattern="^Sp")

ls(pokemon\_data,pattern="^[AD]")

ls(pokemon\_data,pattern="t.l")

ls(pokemon\_data,pattern="ce$")

## 17.Display 1st row and all columns of data frame

pokemon\_data[1,]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

## 18.Display all rows and 1st column of data frame

pokemon\_data[,1]

## 19.Display data in 2nd row and 3rd column of the data frame

## 

pokemon\_data[2,3]

## 20.Display 1st and 2nd row and all columns

pokemon\_data[1:2,]

## 21.Display all rows and first 3 columns

pokemon\_data[,3]

**22.Selecting data where pokemon name is Venusar with subset operator**

temp\_data=subset(pokemon\_data,Name=="Venusaur")  
temp\_data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

## 23.Renaming a column in data frame

temp\_pokemon=pokemon\_datanames(temp\_pokemon)[names(temp\_pokemon)=="Total"]<-"Total\_Number")  
temp\_pokemon[1,]

**24.Adding a new column to dataframe**

temp\_pokemon[["New\_col"]]<-rep(c(1,2,3,4,5),209)  
temp\_pokemon[1:10,]

## 25.Display Sum of Attack column

sum(pokemon\_data[4])

## 26.Display the maximum value of the Attack column

max(pokemon\_data[4])

## 27.Display the minimum value of the Attack column

min(pokemon\_data[4])

## 28.Attaching Pokemon data

attach(pokemon\_data)

## 29.Now we can use variables inside pokemon data

min(Attack)

tail(Speed)

## 30.Sorting Attack variable in ascending order

sort(Attack)

**31.Sorting Attack variable in descending order**

sort(Attack,decreasing = TRUE)

## 32.Detaching Pokemon data

detach(pokemon\_data)

## 33.Using with operator to use variables inside data

with(pokemon\_data,Attack)

## 34.Finding median of data

median(pokemon\_data$Attack)

## 35.Finding mean of data

mean(pokemon\_data$Attack)

## 36.Finding standard deviation of data

sd(pokemon\_data$Attack)

## 37.Finding variance of data

var(pokemon\_data$Attack)

## 38.Order the Attack column in ascending order

order(pokemon\_data[4])

## 39.Order the attack column in descending order

order(pokemon\_data$Attack,decreasing =TRUE)

## 40.Rank of Speed column

rank(pokemon\_data$Speed)

## 41.Rank of speed column with average as tie breaker

rank(pokemon\_data$Speed,ties.method = "average")

## DPLYR operations

## 42.Usage of mutate function

## library(dplyr)

## Attaching package: 'dplyr'

## head(pokemon\_data %>%mutate(mutated\_Attack=Attack-mean(Attack)))

## 43.Adding extra column with user created vector

## vec = rep(c(1,2,3,4,5),209)

## head(pokemon\_data %>% mutate(newcol = vec))

## 44.Group by function

## new\_pok=pokemon\_data %>% group\_by(Name)

## head(new\_pok)

## 45.Summarise function

## head(pokemon\_data %>% group\_by(Attack,HP) %>% summarise(weight\_Defence=mean(Defence)))

## `summarise()` has grouped output by 'Attack'. You can override using the `.groups` argument.

## 46.Rename operation

## copy\_pokemon=pokemon\_data

## pd\_mod <- copy\_pokemon %>% rename(renamed\_Attack = Attack)

## head(pd\_mod)

## 47.Selecting specific columns

## copy\_pokemon=pokemon\_data

## copy\_pokemon %>% select(Name,HP,Attack,Speed,Total)

## 48.Reordering of columns using select function

## copy\_pokemon=pokemon\_data

## copy\_pokemon %>% select(HP,Attack,Name)

## 49.Filter command

## copy\_pokemon=pokemon\_data

## copy\_pokemon %>% filter(Total >= 500 & Total <=505)

**Histogram**

ggplot(pokemon\_data, aes(x = Attack)) +geom\_histogram()

## 51.Histogram of Attack column and its density

ggplot(pokemon\_data,aes(x=Attack))+  
geom\_histogram(fill="cornsilk",color="blue",  
size=0.2)+geom\_density(color="black")

**52.Line graph of Attack column and its density**

ggplot(pokemon\_data,aes(x=Attack))+  
geom\_density(fill="blue",alpha=.4)

**53.Line graph of Attack column taking two alpha values**

ggplot(pokemon\_data,aes(x=Attack))+  
geom\_line(stat="density")+  
geom\_line(stat="density",adjust=0.25,  
color="red")+geom\_density(fill='blue',alpha=0.2)

**52.Dot Plot**

library(ggplot2)

ggplot(pokemon\_data,aes(x=Attack,y=HP))+geom\_dotplot(binaxis="y",stackdir = "center", binwidth = 4,fill="green")

**53.Box Plot**

ggplot(pokemon\_data, aes(x=Attack,y=HP))+geom\_boxplot(width=0.1,fill='black')+stat\_summary(func='median',fill='white',shape=21)

## 54.Density plot for Attack and HP

ggplot(pokemon\_data,aes(x=Attack,y=HP))+geom\_density2d(aes(colour=..level..))

**55.Violin Plot**

Attack and HP

ggplot(pokemon\_data,aes(x=Attack,y=HP))+geom\_violin()