M. SATHISH KUMAR
E0119052

DATA SCIENCE WITH R PROGRAMMING

CA - 2

1. text = "Hi, I am sathish and 18 years old"

~~print~~

pattern = grepl ('[A-Z]', text) &&
grepl ('[0-9]', text)

pattern

Output:

True

Here we used Regex. As grepl checks all values in the string and gives binary output.

---

2. data = read.csv ("Covid19.csv") # Importing data

Variables

active = data $ Active

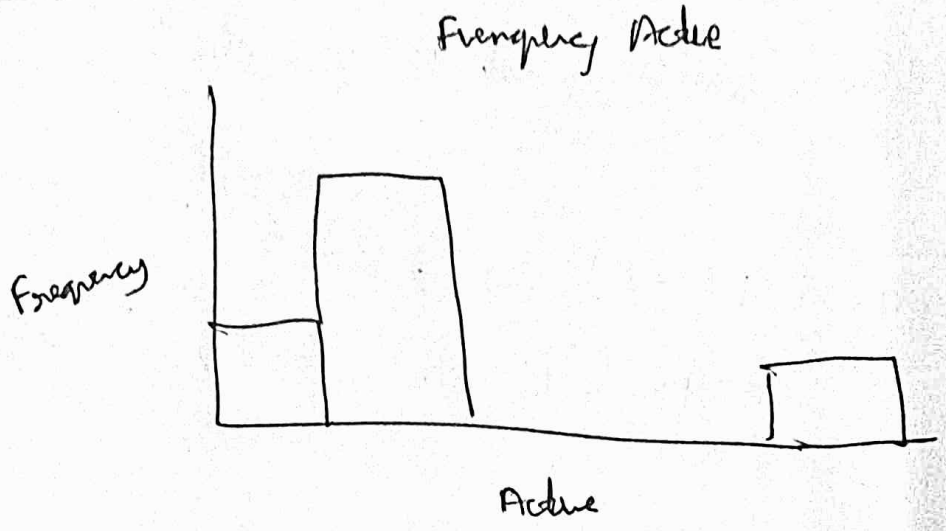
death = data $ Deaths

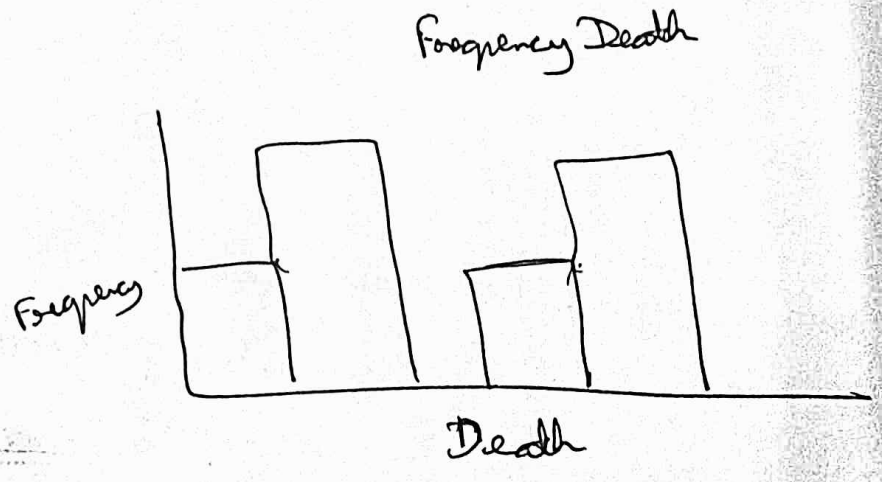
active_hist = hist (active, col = "Yellow", xlab = "Active"
Ylab = "Frequency, Main = "Active Frequency")

death_hist = hist (death, col = "red", xlab = "Death",
Ylab = "Frequency, Main = "Death Frequency")

legend ("top-right", c(df$Active, df$Deaths))

Output :


Frequency Active
Frequency
Active

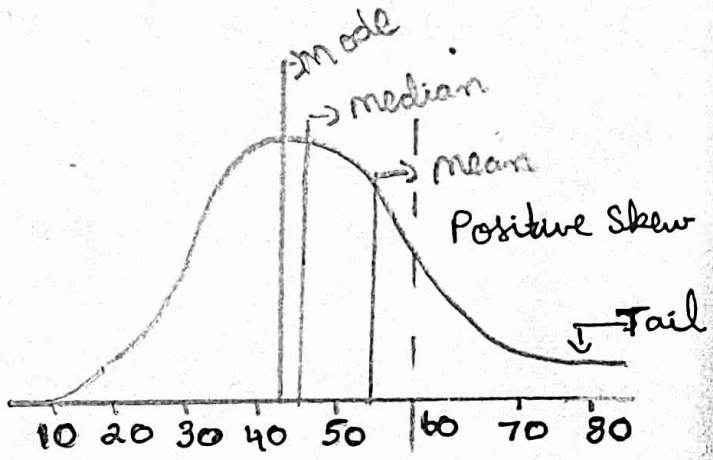

Frequency Death
Frequency
Death

5. **Skewness:**

Measure of the asymmetry of the probability distribution of a real valued random variable about its mean

★ **Types of Skewness:**

- **Positive Skew:**

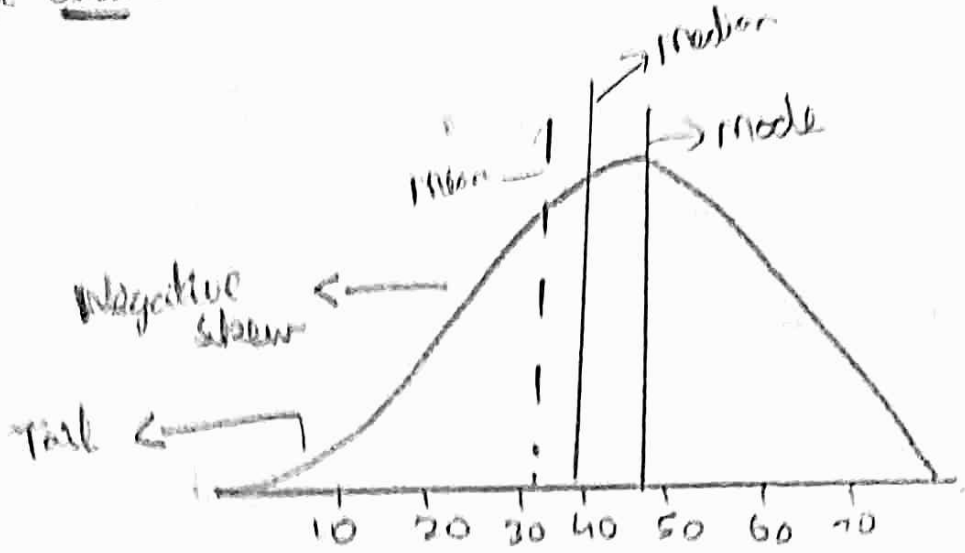


★ A positive skewed distribution with the tail on its right side.

★ Value of skewness for a positively skewed distribution is greater than zero.

Mean >> Median >> Mode

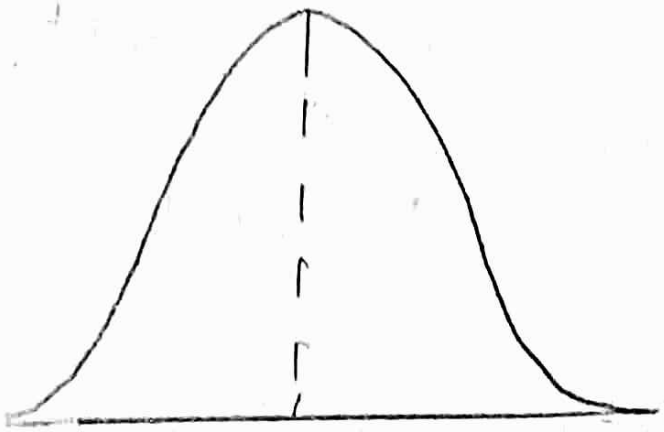
- Negative Skew :



* Negatively Skewed Distribution is the distribution with tail on its left
* Value of Skewness is less than zero

mean < median < mode

- Symmetrical Skew :



* Normal skewness is the ~~poor~~ probability distribution with almost no skewness
* Skewness near to -0

Code:

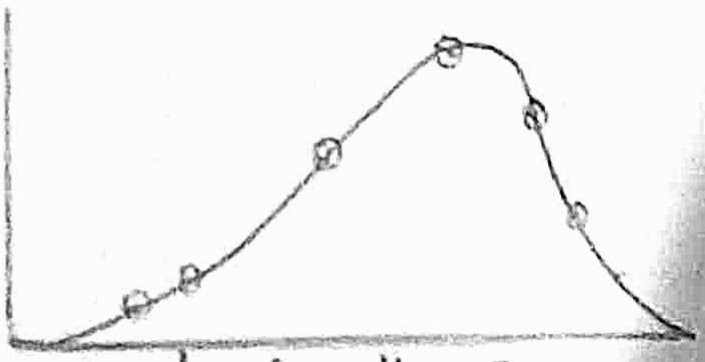
```
# Required Packages:
library (e1071)
df = read.csv ("Python.csv")
skew = skewness (df $ Score)
```

Code:

```
# Required Packages:
library (e1071)
df = read.csv ("Python.csv")
#Applying skewness function
skew = skewness (df $ Score)
print (skew)
summary (skewness (df $ mark))
# Plotting a density curve
polygon (density (df $ Score), col = "red", border = "blue")
```

Output:

-0.1111



Negative skew, so value less than 0

Min	1st Q	Median	Mean	3rd Qu.	Max
10	50	68	60	85	95

4. data.csv

Name	Id
Which	11
Data	21
Big	15
large	11
Storage	13
SETS	18
Search	17
management	15
Storage	12
Science	8

Code:

```
library (wordcloud)
df ← read.csv (words.csv)
print (df)
wordcloud (words = df$Name, freq = df$Id,
           min.freq = 1, max.words = 50, random.order = F)
```

Pattern : Word cloud

Output :

LARGE VOLUME INTENET NO
BIG INFORMATION ANALYSIS NAS
MILLION DATA SEARCH MES
SETS DEVELOPMENT LARGE TIME
NEWS INTER

---

3. Layers in ggplot2

* data
* aesthetic
* geometry
* facets
* statistics
* coordinates

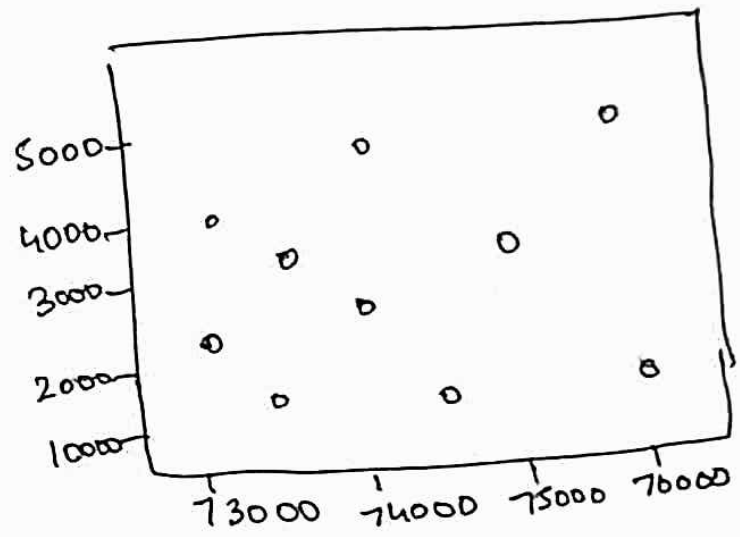
Code :

```
library ("ggplot2")          # Required Packages
library (ggplot2.movies)
~~library ggplot ( data = ~~
df = read.csv ("Covid19.csv")
pl <- ggplot ( df , aes ( df $ Confirmed
```

Code:

```
library (ggplot)
df = read.csv ("Covid 19.csv")
pl ← ggplot (df, aes ( x= df $ Confirmed, y = df $ Active))
pl ← pl + geom_point()
print (pl)
```

Output:



1.