Experiment - 3

Shape of the Data and describing relationships.

Experiment-4

Tables, charts and plots.

Data <— read.csv("filename.csv", header =True).

Frequency and percentage table

Eg:- table (Data)
prop.table (Data).
table (Data $ columnname)
prop.table (table(Data $ column)).

Contingency table :- A special type of frequency distribution table with 2 variables shown simultaneously we use 'x' tabs() to find contingency among 2 variables within an same Data.

Eg:- xtabs(~Variable1 + Variable2, object).
xtabs(~column + Variable3, data).

Charts and plots.

(i) Bar plot

Used to plot a bar graphical representation, by using barplot()

Eg:- x <- c(7,15,23,12,44,56,32)
Barplot(xlab, x = "x-axis", ylab="y-axis", col="Green", col.axis="dark green", col.lab="darkgreen").

(ii) pie chart

It is used to Draw piechart by using pie().

Eg:- x <- c(210,450,250,100,50,90)
names(x) <- c("Algo","DS","Java","C","C++","Python")
pie(x, labels = names(x), col="Yellow", main="SVIT subjects", radius=-1, col.main="darkgreen").

(iii) 3D pie charts :

We use plotrix package, use the library of plotrix. To construct 3D pie charts use the function called pie3D.

Eg:- pie3D(x, labels=names(x), col="red", main="SVIT subjects", label col="darkgreen", col.main="dark green").

(iv) Histogram :

Used to draw graphs with box, by using function called "Hist".

Eg:-

```
Hist(x, main = "histogram", xlab = "values",
    ylab = "freq", col.lab = "darkgreen",
    col.main = "yellow").
```

(v) Scatter plots :

Is used to plot the points to show relationship b/w two Data vectors.

Eg:- plot(x, Y, xlab = "Age", Ylab = "factor",
main = "Age Vs factor", col.lab = "yellow",
col.main = "red, col.axis = "Green).

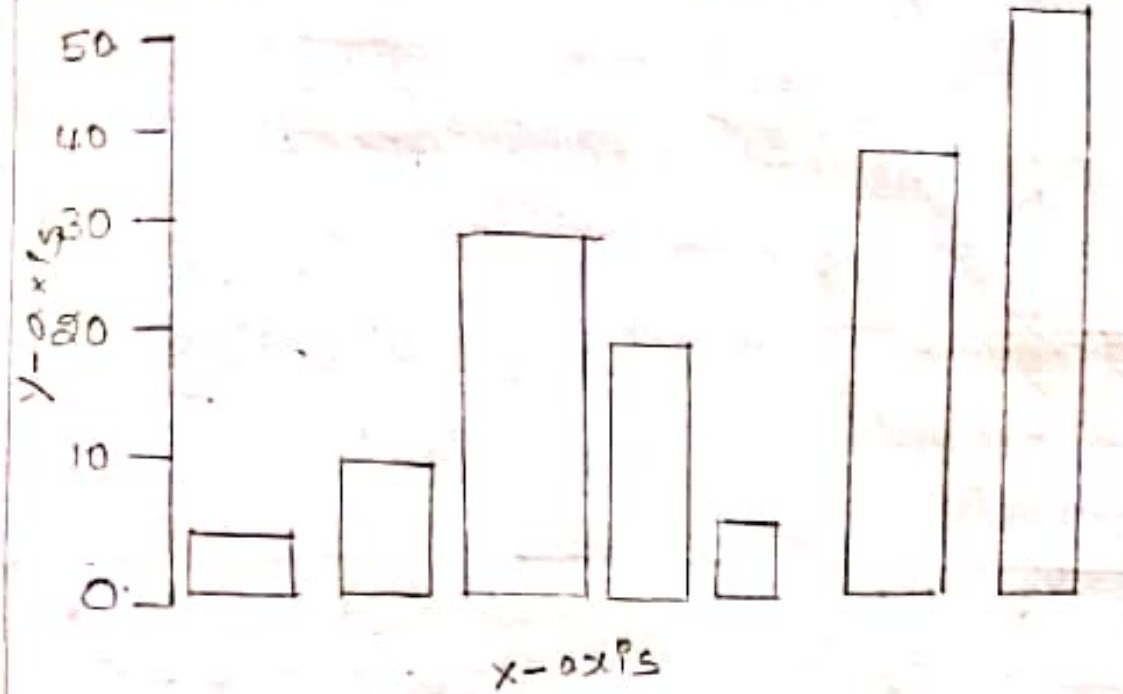
(vi) Box plot :

It is used to show how the data is distributed.

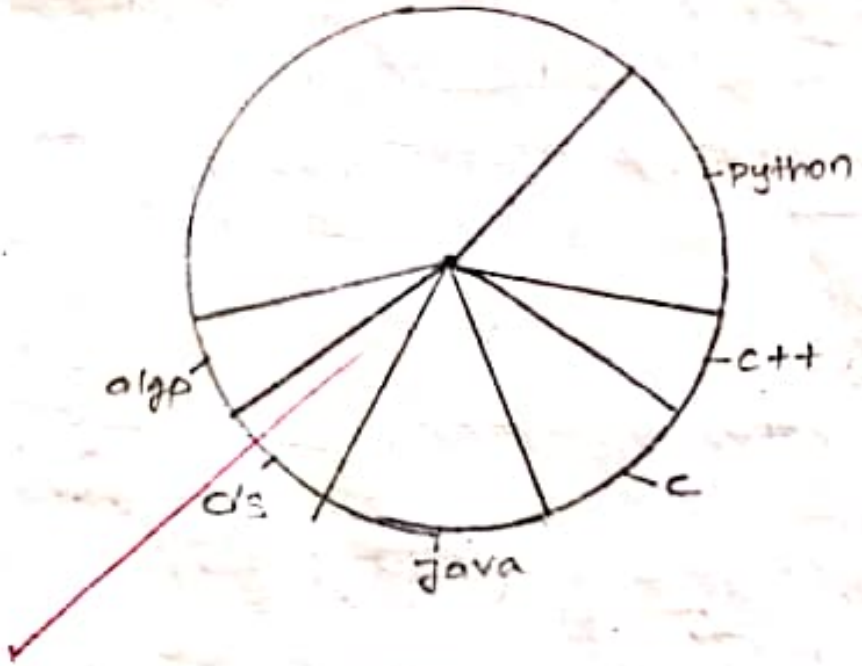
Eg:- boxplot(x, xlab = "prudhvi", ylab =
"kishore", notch).

notch :- If true creates notch on both sides of the box.

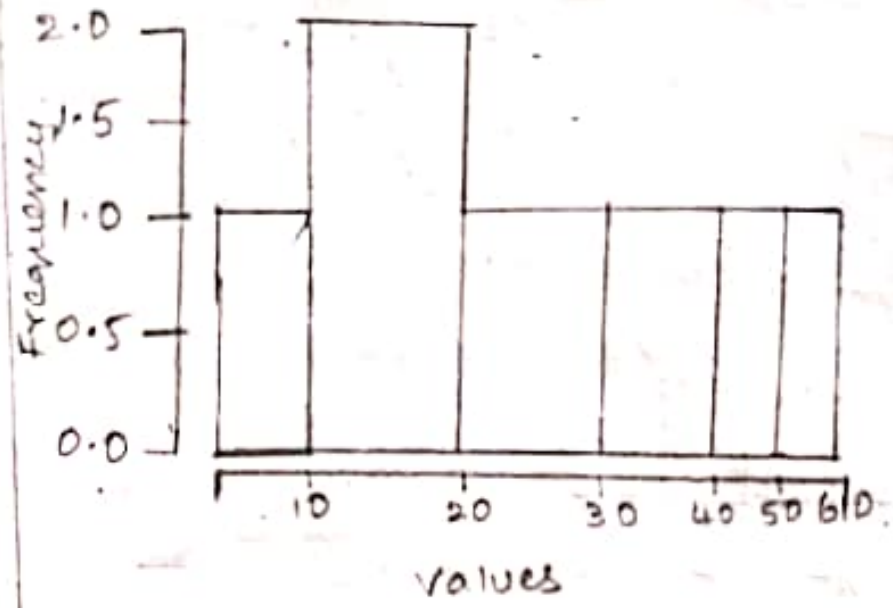
(i) Bar plot



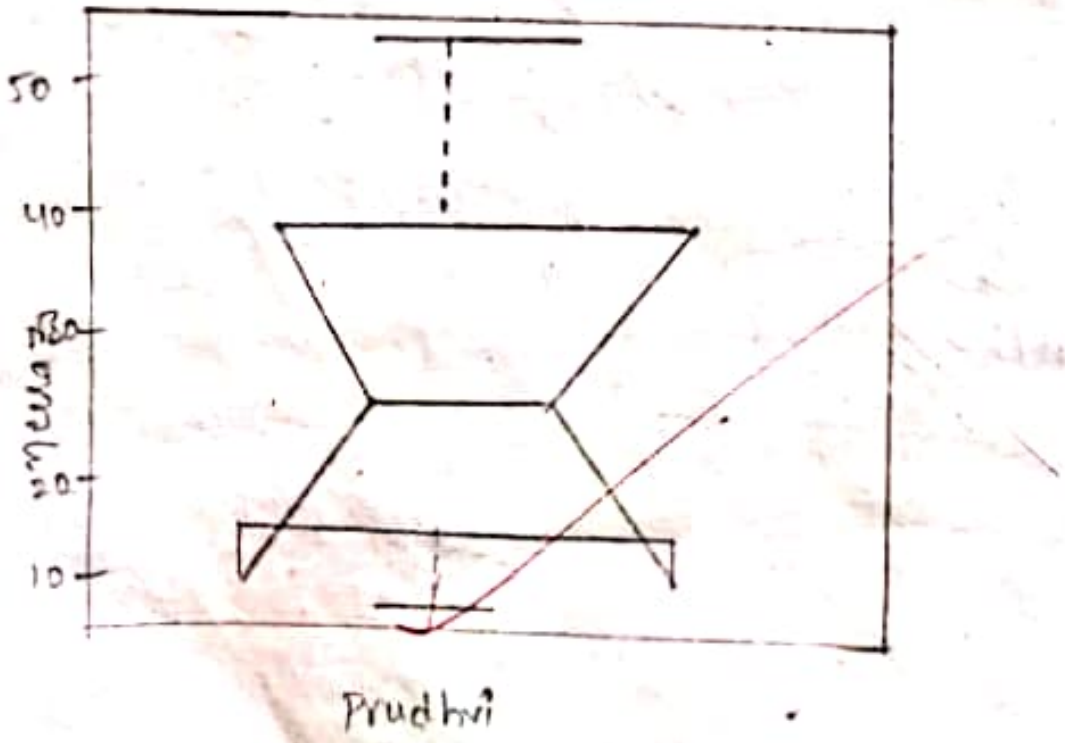
(ii) pie chart



(iv) Histogram



(v) Box plot



Experiment-B

Univariate data, measures of central tendency, frequency distributions, variation, and shape.

Data <- read.csv("filename.csv").

Mean():- Is used to find mean of an variable.

Median():- Is used to find median of an function

IQR():- Is used to find Inter Quartile.

SD():- Is used to find Standard deviation of an variable.

density():- It used to find density of a data, we have to represent in a graph format

```
view(Data)
mean(Data $ columnname).
median(Data $ columnname).
max(Data $ columnname).
min(Data $ columnname).
max(Data $ columname) - min(    ) => Variati
IQR(  "  "  " ).
SD(   "  "  " ).
Boxplot(  "  "  " ).
hist  (  "  "  " ).
```

Plot (density (Data $ columnname)).

Data <— read.csv ("Sample2.csv")

Data

X35
68.02
2.99
3.99
5.94
4.95
7.72
6.22
35.00

Mean(data $ X35).

13.25546

Median (data $ X35)

6.16

max(data $ X35)

164.73

min (data $ X35)

0.49

max(data $ X35) - min (data $ X35)

164.24

IQR (data $ X35)

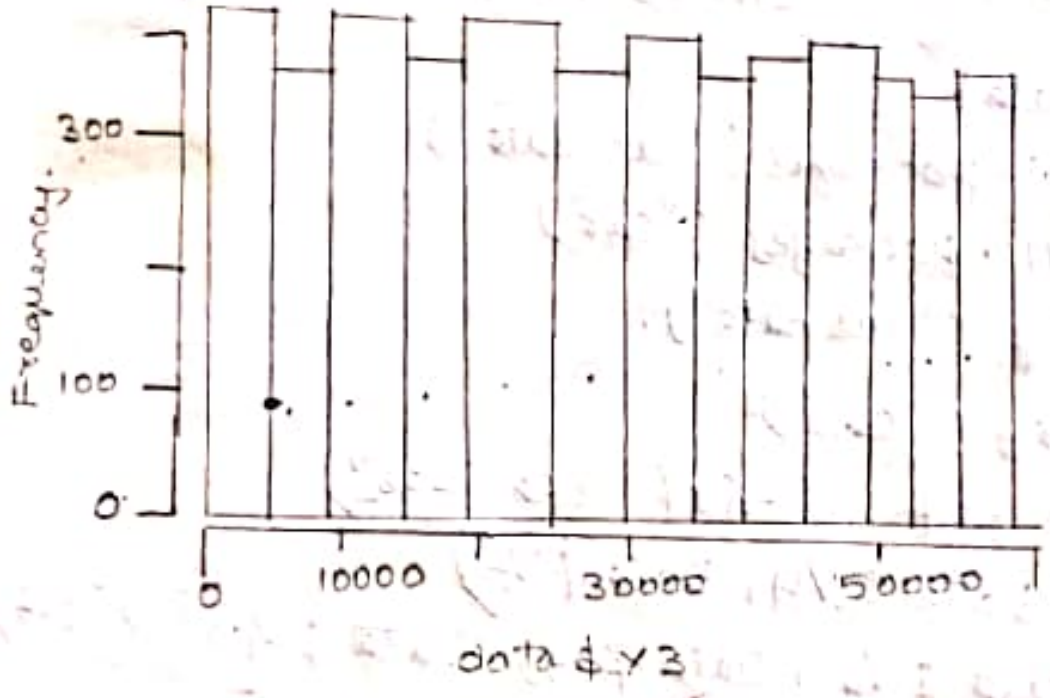
10.63

Sd (data $ X3)

17277.28.

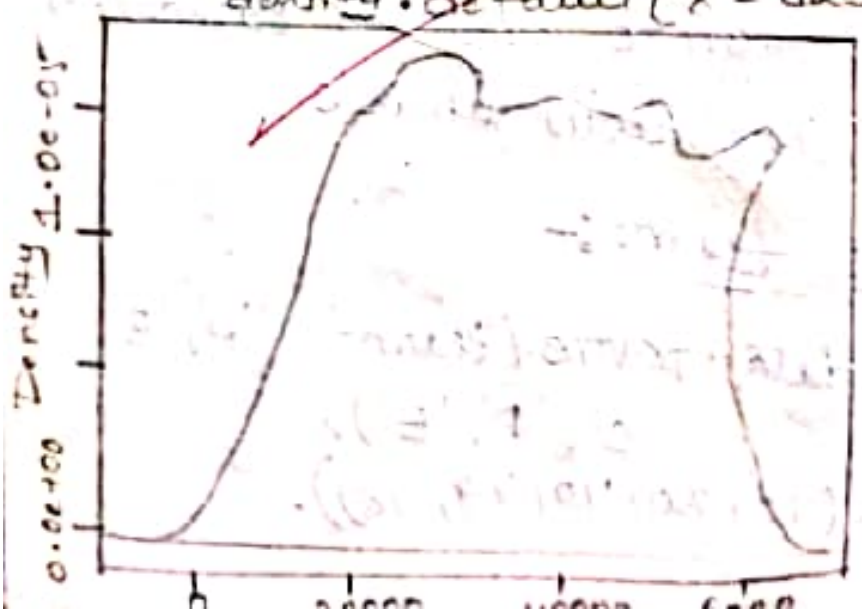
boxplot (data $ x3).

hist (data $ x3).



Plot (density (data $ x3)).

density.default (x = data $ x3)



Experiment-c

Multivariate Data, relationships between a Categorical and a continuous Variable.

Multivariable data :- It is data in which values are several variables i.e more than two variables data are recorded on each unit.

```
Data ← read.csv("filename.csv", header=FALSE,
                strings as factors = True).

View(Data)
Install.packages("car data")
Install.packages("car")
library("car data").
library("car").
Scatter plot matrix( Data [2:6])
plot(Data $ V4, Data $ V5)
text(Data $ V4, Data $ V5), Data $ V1, cex=0.7,
pos=5, col="Red")
```

⇒ Use cut function to create a categorical Variable from continuous Variable

Creation of data frame :-

```
Eg :- df ← data.frame(team=c('A', 'B',
                      'C', 'D', 'E'),
        Points = c(10, 30, 40, 49, 58)).
```

To add new column :- thats cut points into categories

```
df $cat <- cut( df $ points, breaks =
            c(11, 32, 38, 49, 50), labels =
            c('Good', 'Bad', 'OK', 'Great')).
```

```
df
or
view(df).
```

```
Data <- read.csv("sample2.csv", header=
        FALSE, stringsasfactors = True).
```

view(Data).

V1	V3	V4	V5	V6	V7	V8
1	Muhamed	3	-213.5	38.94	35.00	Nunavut
2	Barry	293	45781	46.71	67.00	Nunavut
3	Barry	293	46.71	59.81	78.00	Nunavut
4	Ekaf	483	1198.67	90.90	43.00	Nunavut
5	clay	515	30.94	87.91	22.00	Nunavut
6	Carlos	613	4.3	91.87	16.00	Nunavut
7	carlos	613	12770			

```
df <- data.frame(team = c('a', 'b', 'c', 'd', 'e')
      , points(10, 30, 40, 49, 58))
```

df

	Team	points
1	a	10
2	b	30
3	c	40
4	d	49
5	e	58

df$cat <- cut(df$points, breaks = c(11, 32, 38, 49, 50), labels = c('good', 'bad', 'ok', 'great')).

	Team	points	cat
1	a	10	<NA>
2	b	30	good
3	c	40	ok
4	d	49	ok
5	e	58	<NA>.

Experiment - D.

Relationship between to continuous Variables covaience, co-Relation, co-efficients, Comparing multiple co-relations.

Covariance:- Use cov() for finding co-variance

Eg:- cov(x,y)

x,y are any two continuous Variables

co-Relation:- Use cor() for finding co-relation, co-efficients.

Eg:- cor(x,y)

x,y are any covariance Variables.

$$cov = Sxy = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$\text{corelation co-eff} = \frac{Sxy}{Sx\,Sy}$$

$x = sd(x)$

$y = sd(y)$.

```
x <— c(1,20,56,45,33)
y <— c(45,25,18,22,56):
cov(x,y).
-188.75
cor(x,y)
-0.5340193.
```

Two Categorical Variables :-

Use plot function (or) Hist function

```
df$cat<-cut(df$marks,breaks=c(20,40,60,80,100)
label=c('ok','good','average','great'))
df$bat<-cut(df$marks,breaks=c(20,40,60,80,100)
label=c('ok','good','average','great'))
```

team	marks	cat	bat
kishore	35	ok	ok
Prudhvi	45	good	good
nagasai	55	good	good
babasai	65	Average	Average
akram	75	Average	Average

Two continuous Variables :-

Use plot function (or) Hist function.