ANALYSIS OF DATA FROM CSV FILE

PRANAVCHENDUR T K | 15BCE1097

Aim

To process data from CSV file and represent using different types of graphs and plots including Histogram and Frequency Polygon

Program

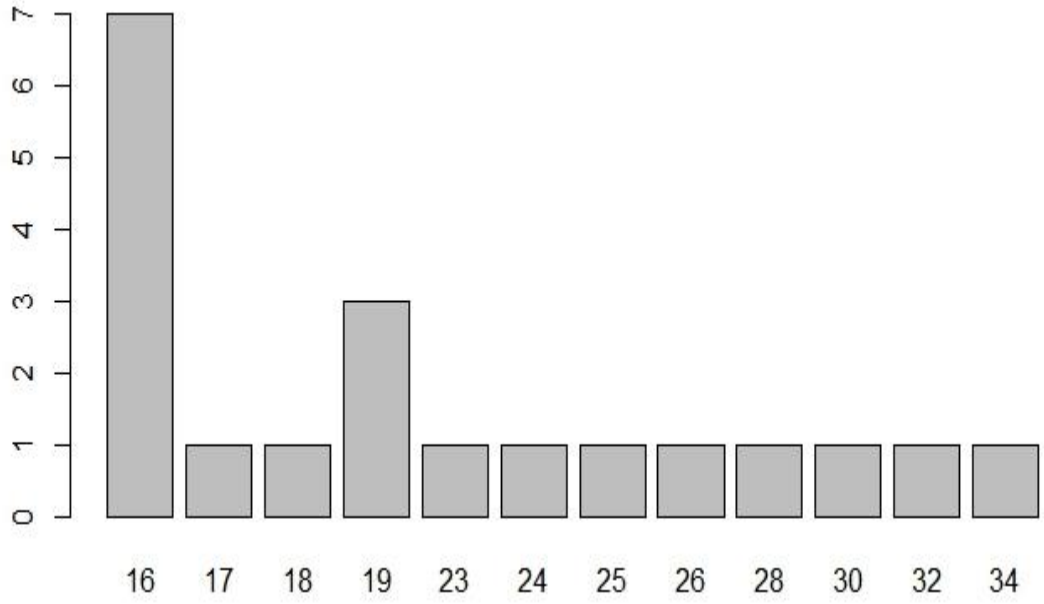
```
dat = read.csv("data.csv", header = TRUE)
dat$Gender=factor(dat$Gender,labels = c("Female","Male"))
dat$Married=factor(dat$Married,labels = c("No","Yes"))

table1=table(dat$Age)
barplot((table1))

table2=table(dat$Gender)
pie(table2)

table3=table(dat$ChildC)
barplot(table3)

table4=table(dat$Married)
pie(table4)

summary(dat)

hist(dat$Age, main = "Histogram of Age", xlab="Age Range")

datainc = dat$Income
datainc
breaks = seq(min(datainc), max(datainc), by=sd(datainc))
datainc.cut = cut(datainc, breaks, right=FALSE)
datainc.freq = table(datainc.cut)
cumfreq = c(0, cumsum(datainc.freq))
plot(breaks, cumfreq, main="Frequency Polygon Of Income", xlab="Income", y
lab="Number of People")
lines(breaks, cumfreq)
```

Table

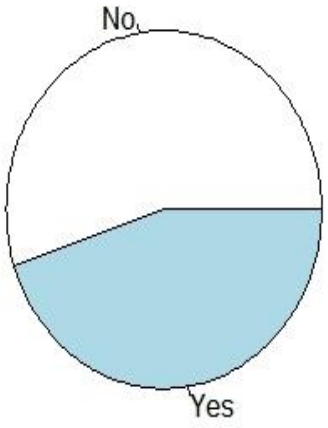
	Subject	Age	Gender	Married	IncomeC	HealthC	ChildC	LifeSatC	SES	Smoke	Spirit	Finish	LifeSat	Income
1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	1	16	Female	No	0	38	0	17	17	1	30	1	22	26
3	2	28	Male	No	0	38	0	16	21	1	39	1	20	15
4	3	16	NA	Yes	16	52	1	39	40	0	30	1	42	88
5	4	23	Male	No	6	51	0	22	31	0	60	1	48	73
6	5	18	Female	Yes	7	52	0	25	38	0	32	0	NA	14
7	6	30	Female	Yes	25	43	2	53	36	1	39	0	33	38
8	7	19	Female	Yes	19	55	0	28	41	0	51	1	33	45
9	8	19	Male	No	0	52	2	17	52	0	35	1	21	16
10	9	34	Female	No	29	60	2	20	56	0	23	1	26	64
11	10	16	Male	No	0	53	0	21	27	0	29	0	37	19
12	11	25	Male	No	3	39	0	18	34	1	61	1	40	56
13	12	16	Male	Yes	1	42	0	31	29	1	58	1	35	70
14	13	16	NA	No	0	43	0	15	28	1	39	1	32	71
15	14	16	Female	Yes	18	54	1	34	38	0	40	0	37	44
16	15	16	Male	No	0	52	0	20	38	0	27	1	35	25
17	16	32	Male	Yes	26	54	1	39	37	0	30	NA	47	38
18	17	19	Female	No	0	46	0	17	25	0	36	1	26	39
19	18	17	Male	Yes	10	55	2	48	53	0	43	0	42	6
20	19	24	Female	No	17	52	0	16	36	0	54	1	38	75
21	20	26	Male	Yes	NA	57	1	39	41	0	32	1	42	67

Graphs

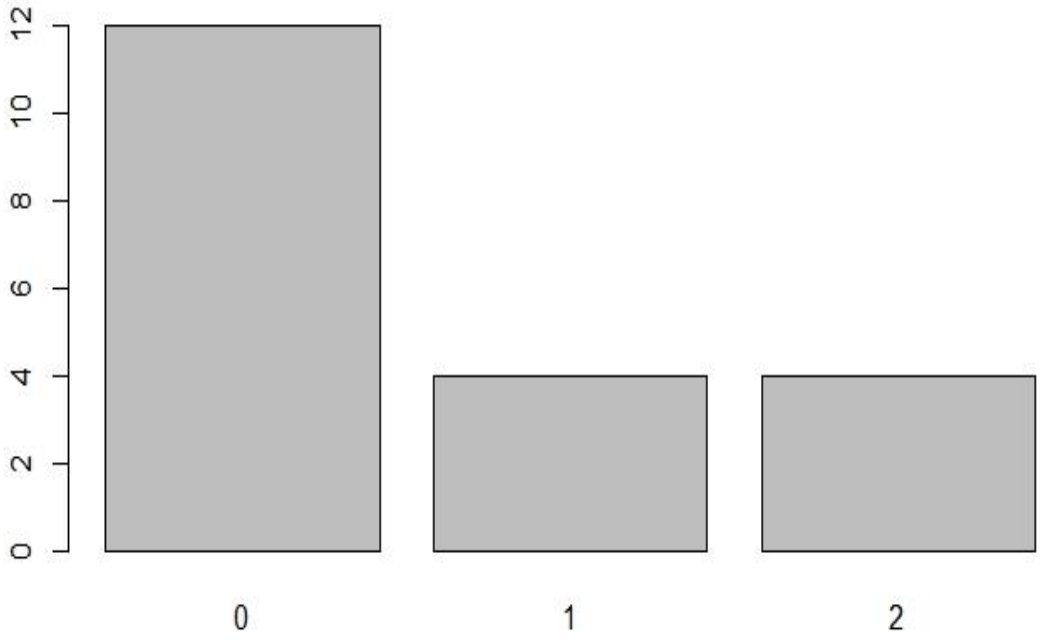
Age Distribution



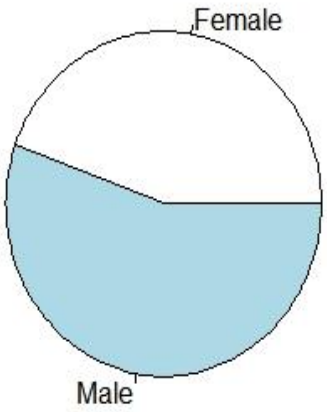
Married



Number of Children



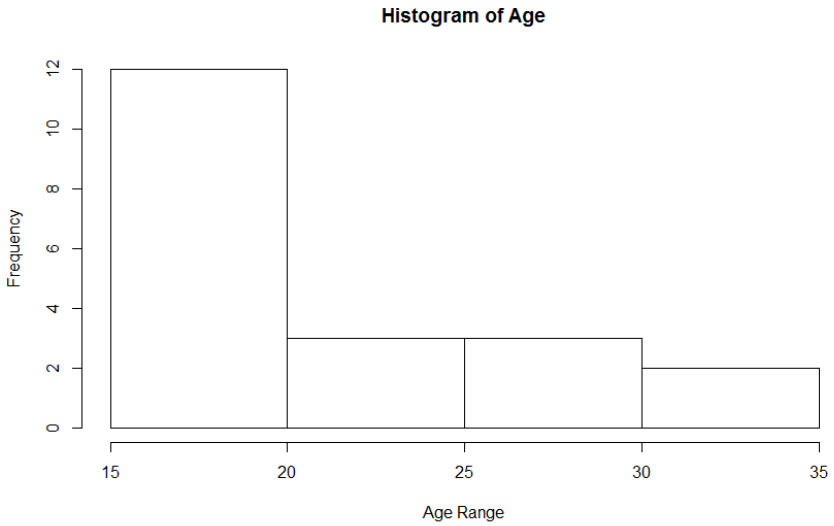
Gender



Summary

```
> summary(dat)
    Subject           Age           Gender   Married     IncomeC          HealthC          ChildC          LifeSatC
 Min.   : 1.00   Min.   :16.00   Female:10   No :11   Min.   : 0.00   Min.   :38.0   Min.   :0.0   Min.   :15.00
 1st Qu.: 5.75   1st Qu.:16.00   Male  :10   Yes: 9   1st Qu.: 0.00   1st Qu.:43.0   1st Qu.:0.0   1st Qu.:17.00
 Median :10.50   Median :19.00                        Median : 4.50   Median :52.0   Median :0.0   Median :21.50
 Mean   :10.50   Mean   :21.30                        Mean   : 8.85   Mean   :49.4   Mean   :0.6   Mean   :26.75
 3rd Qu.:15.25   3rd Qu.:25.25                        3rd Qu.:17.25   3rd Qu.:54.0   3rd Qu.:1.0   3rd Qu.:35.25
 Max.   :20.00   Max.   :34.00                        Max.   :29.00   Max.   :60.0   Max.   :2.0   Max.   :53.00
      SES             Smoke            Spirit          Finish          LifeSat          Income
 Min.   :17.00   Min.   :0.0   Min.   :23.0   Min.   :0.0   Min.   : 0.0   Min.   : 6.00
 1st Qu.:28.75   1st Qu.:0.0   1st Qu.:30.0   1st Qu.:0.0   1st Qu.:26.0   1st Qu.:23.50
 Median :36.50   Median :0.0   Median :37.5   Median :1.0   Median :35.0   Median :41.50
 Mean   :35.90   Mean   :0.3   Mean   :39.4   Mean   :0.7   Mean   :32.8   Mean   :44.45
 3rd Qu.:40.25   3rd Qu.:1.0   3rd Qu.:45.0   3rd Qu.:1.0   3rd Qu.:40.5   3rd Qu.:67.75
 Max.   :56.00   Max.   :1.0   Max.   :61.0   Max.   :1.0   Max.   :48.0   Max.   :88.00
```

Histogram



Frequency Polygon