**TAKE HOME ASSIGNMENT**

**SCS2111 ASSIGNMENT – 02**

**Name – G. S. I. M Perera**

**Index No - 14001111**

Q1.

1. **> library(datasets)**

**> attach(Loblolly)**

**> Loblolly**

1. **> head(Loblolly,15)**

**height age Seed**

1 4.51 3 301

15 10.89 5 301

29 28.72 10 301

43 41.74 15 301

57 52.70 20 301

71 60.92 25 301

2 4.55 3 303

16 10.92 5 303

30 29.07 10 303

44 42.83 15 303

58 53.88 20 303

72 63.39 25 303

3 4.79 3 305

17 11.37 5 305

31 30.21 10 305

c)

**> summary(Loblolly$height)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.46 10.47 34.00 32.36 51.36 64.10

Above is the summary of the statistics of the height of Loblolly data-frame. The height is in a range of 3.46 – 64.10. The average height is 32.36 and the middle value of height is 34.

**> summary(Loblolly$age)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

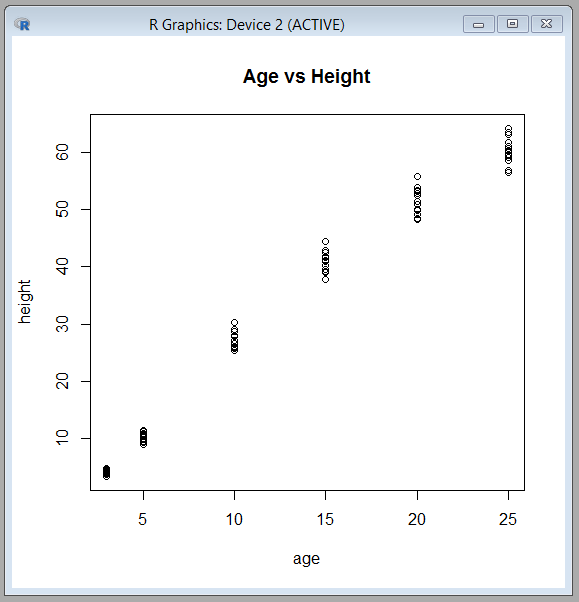
3.0 5.0 12.5 13.0 20.0 25.0

Above is the summary of the statistics of the age of Loblolly data-frame. The age is in a range of 3 – 25. The average age is 13 and the middle value of age is 12.5.

**> summary(Loblolly$Seed)**

329 327 325 307 331 311 315 321 319 301 323 309 303 305

1. 6 6 6 6 6 6 6 6 6 6 6 6 6
2. **> plot(Loblolly$age,Loblolly$height, main="Age vs Height",xlab="age",ylab="height")**



1. **> cor(Loblolly$age,Loblolly$height)**

[1] 0.9899132

F) The correlation coefficient is normalized measurement of how the two variables are linearly related. Here the scatter plot falls almost along straight lines with positive slopes. Also the correlation coefficient obtained in part e for the above variables is so close to 1. Therefore the variable are positively linearly related.

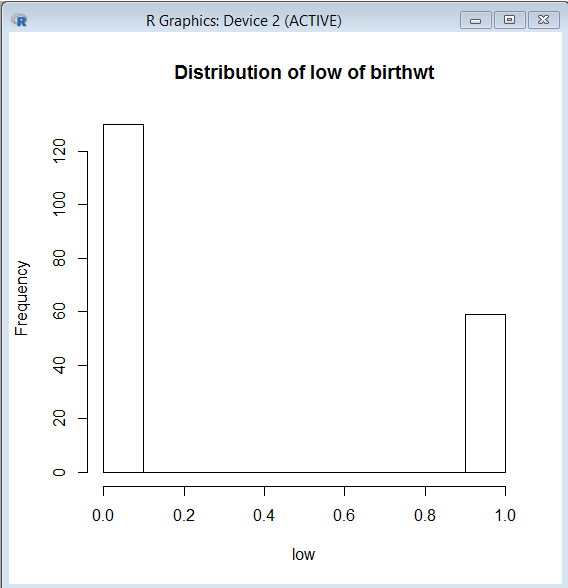
Q2.

1. **> library(MASS)**

**> attach(birthwt)**

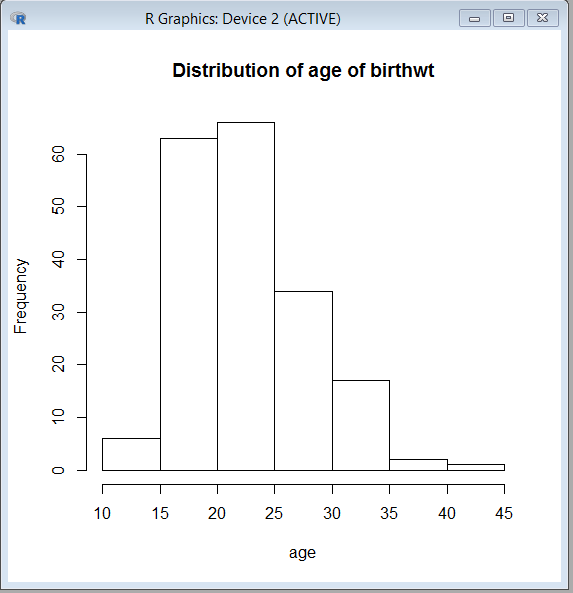
**> birthwt**

1. **> hist(birthwt$low,main="Distribution of low of birthwt",xlab="low")**



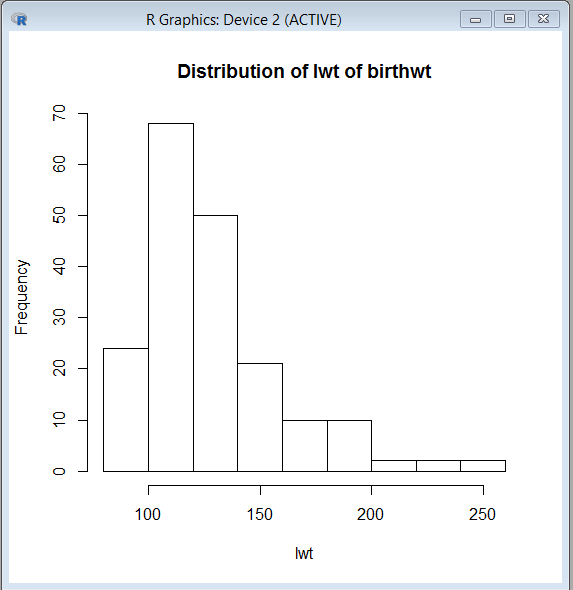
This distribution is a little towards comb distribution. Birth weights are either 0 or 1.

**> hist(birthwt$age,main="Distribution of age of birthwt",xlab="age")**



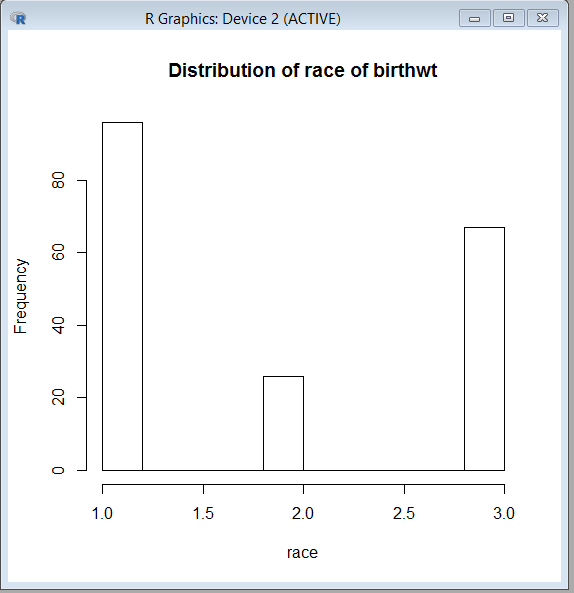
This distribution is a right –skewed distribution. The range of age is 10 – 40.

**> hist(birthwt$lwt,main="Distribution of lwt of birthwt",xlab="lwt")**



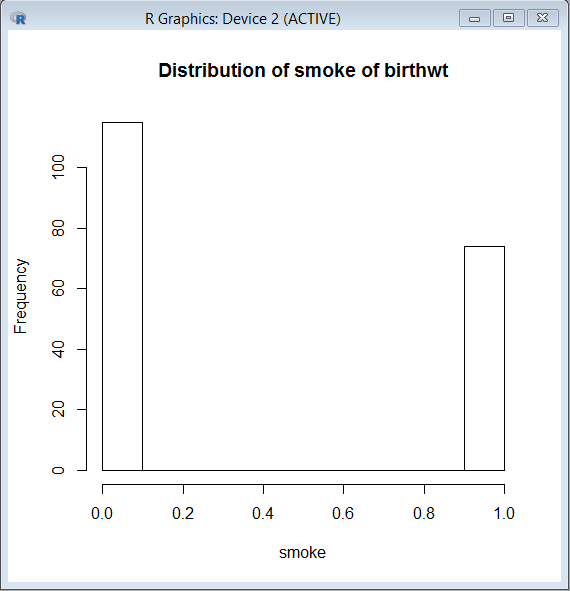
This distribution is a right skewed distribution.

**> hist(birthwt$race,main="Distribution of race of birthwt",xlab="race")**



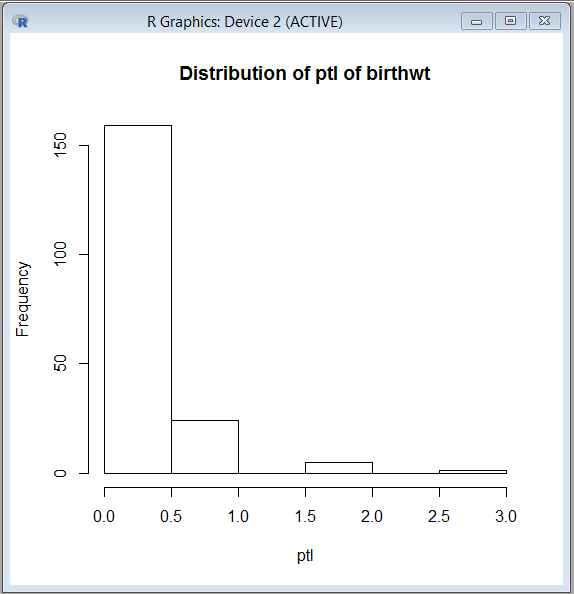
This distribution is a little towards the comb distribution.

**> hist(birthwt$smoke,main="Distribution of smoke of birthwt",xlab="smoke")**



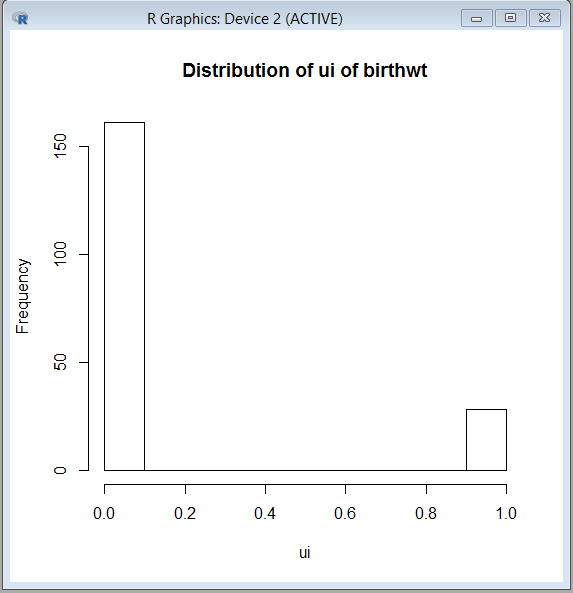
This distribution is little towards comb distribution. The smole values are either 0 or 1.

**> hist(birthwt$ptl,main="Distribution of ptl of birthwt",xlab="ptl")**

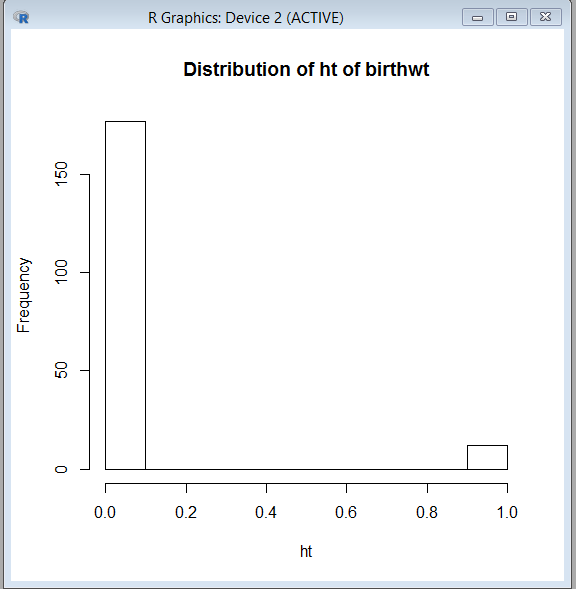


This distribution is little towards the comb distribution.

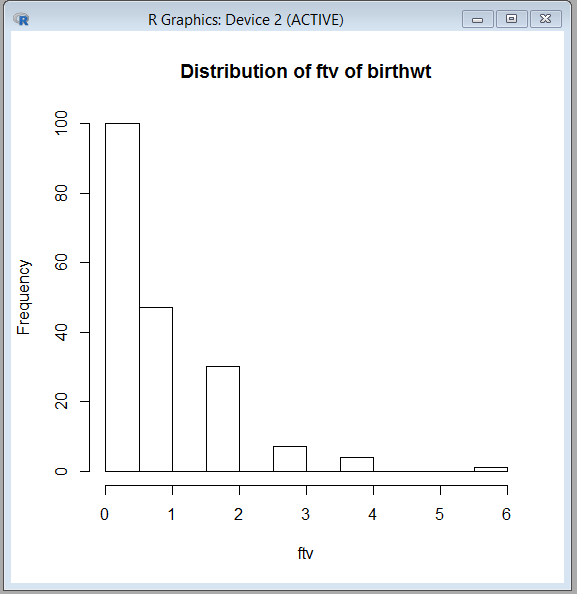
**> hist(birthwt$ui,main="Distribution of ui of birthwt",xlab="ui")**



**> hist(birthwt$ht,main="Distribution of ht of birthwt",xlab="ht")**

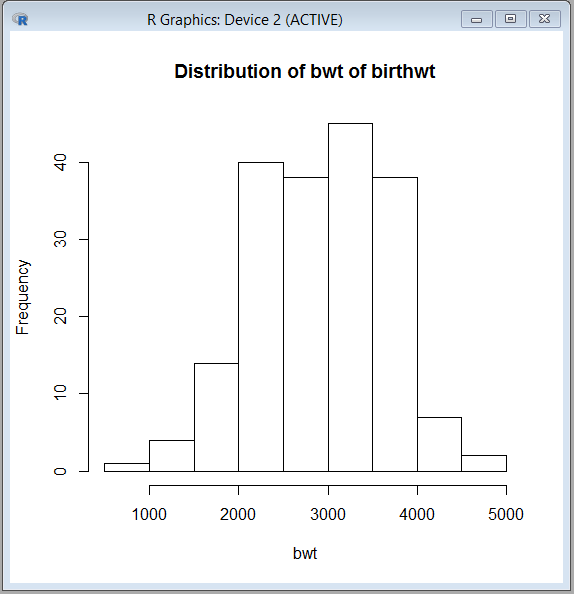


**> hist(birthwt$ftv,main="Distribution of ftv of birthwt",xlab="ftv")**



This distribution is a right skewed distribution with a range of 0 – 6.

**> hist(birthwt$bwt,main="Distribution of bwt of birthwt",xlab="bwt")**



This distribution is normally distributed within the range of values.

1. **> table(birthwt$low)**

0 1

130 59

**> table(birthwt$race)**

1 2 3

96 26 67

**> table(birthwt$smoke)**

0 1

115 74

**> table(birthwt$ht)**

0 1

177 12

**> table(birthwt$ui)**

0 1

161 28

d)

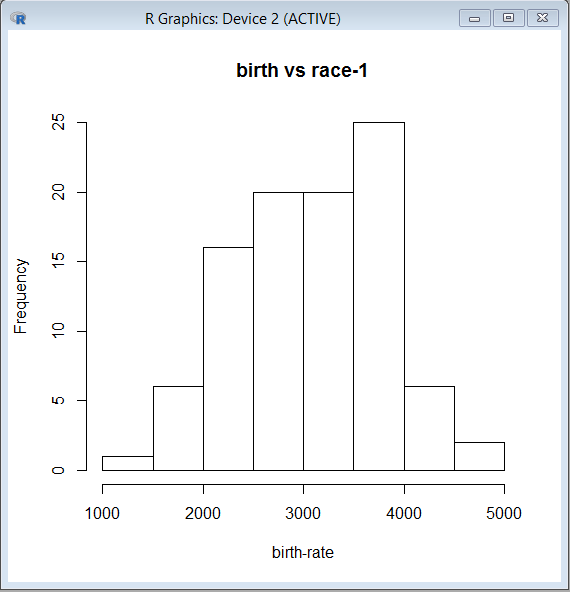
**> Birth\_1 <- birthwt[birthwt$race=="1",]**

**> summary(Birth\_1[,10])**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1021 2585 3062 3103 3651 4990

**> hist(Birth\_1[,10], main="birth vs race-1",xlab="birth-rate")**



This is left-skewed distribution. The range is between 1000 - 5000

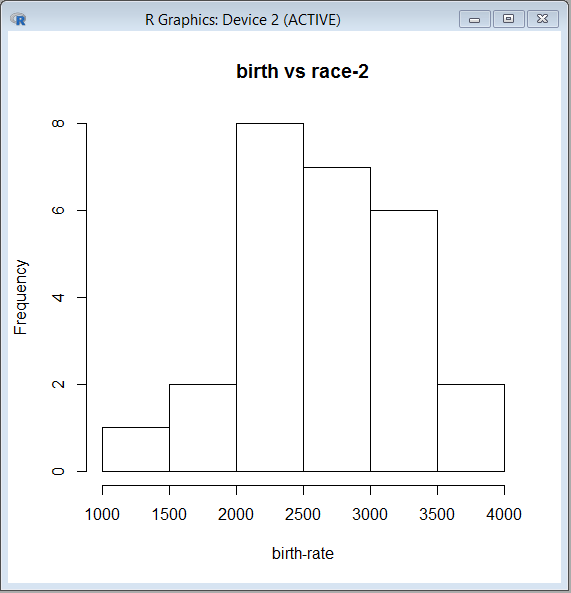
**Birth\_2 <- birthwt[birthwt$race=="2",]**

**> summary(Birth\_2[,10])**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1135 2370 2849 2720 3057 3860

**> hist(Birth\_2[,10], main="birth vs race-2",xlab="birth-rate")**



This is normal distribution. The range is between 1000 - 4000

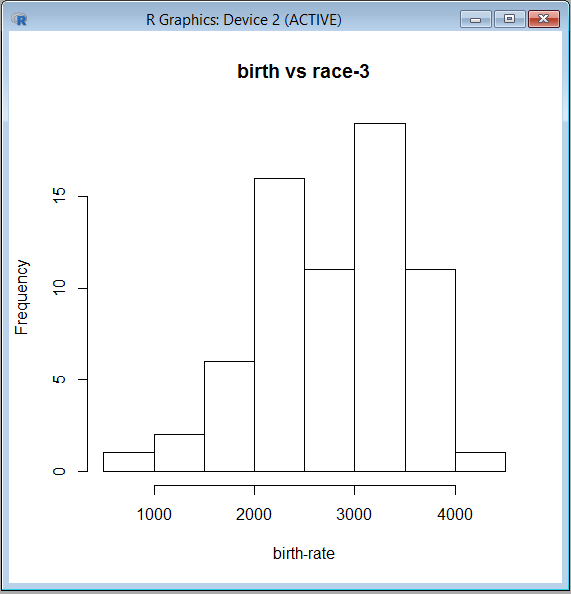
**> Birth\_3 <- birthwt[birthwt$race=="3",]**

**> summary(Birth\_3[,10])**

Min. 1st Qu. Median Mean 3rd Qu. Max.

709 2313 2835 2805 3274 4054

**> hist(Birth\_3[,10], main="birth vs race-3",xlab="birth-rate")**



This is left-skewed distribution.