1-We calculate our statistical findings as shown

> mean(data1$Wage,na.rm = T)

[1] 9.024064

> median(data1$Wage)

[1] 7.78

> var(data1$Wage)

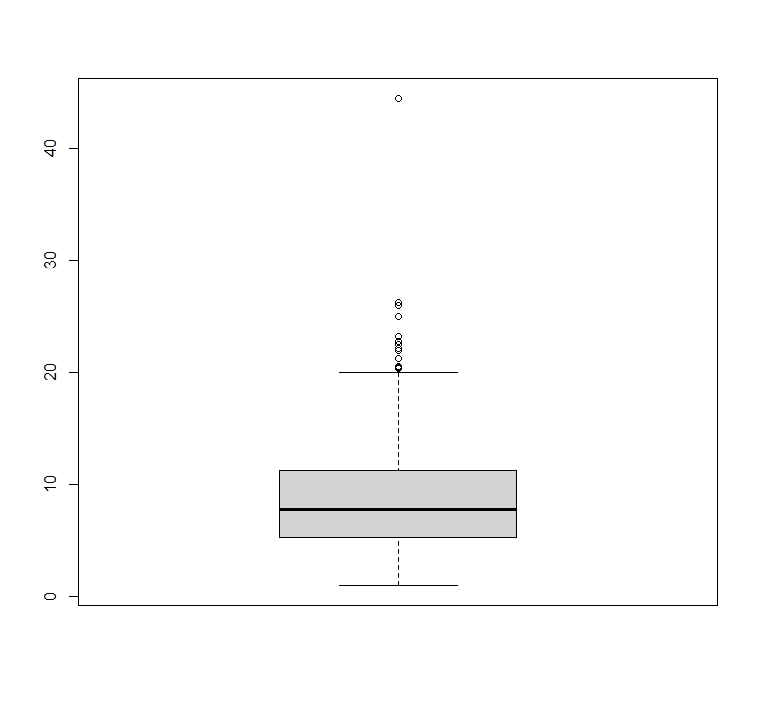
[1] 26.41032

> sd(data1$Wage)

[1] 5.139097

2- boxplot(data1$Wage)

We see that we have some amount of outliers and one huge outlier. We have a “crowd” in the upper side of the candle. The issue can be revisited with log functions.



3- x <- table(data1[10]) y <- table(data1[9])

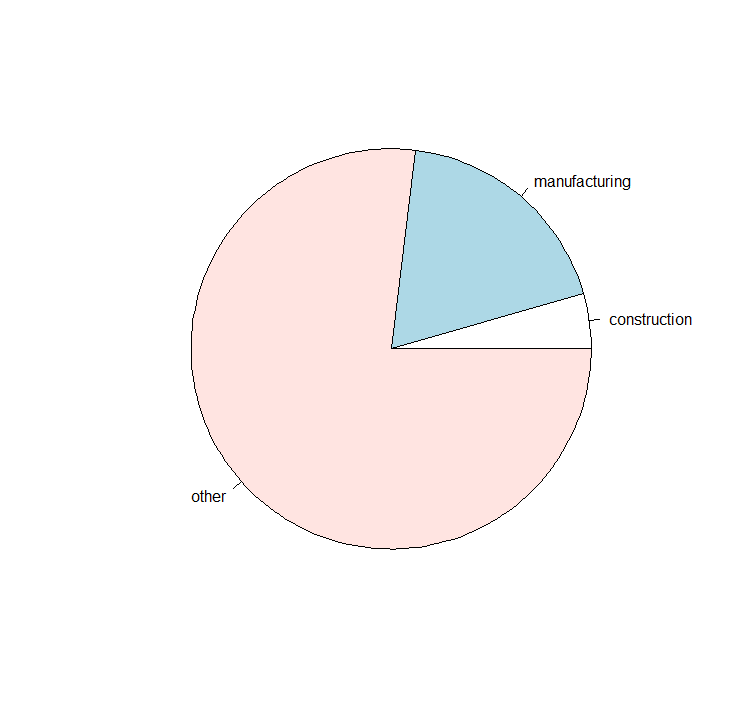
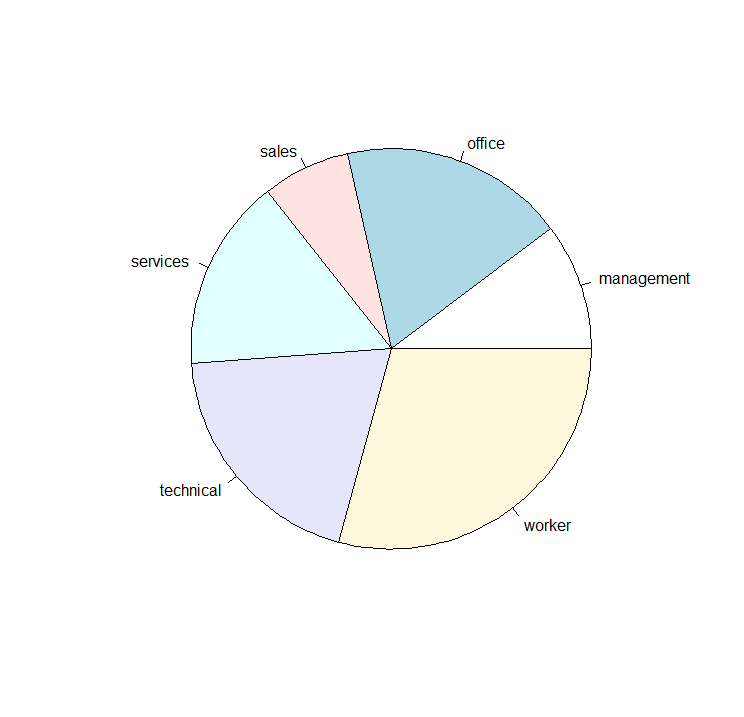
barplot(x) ,barplot(y)

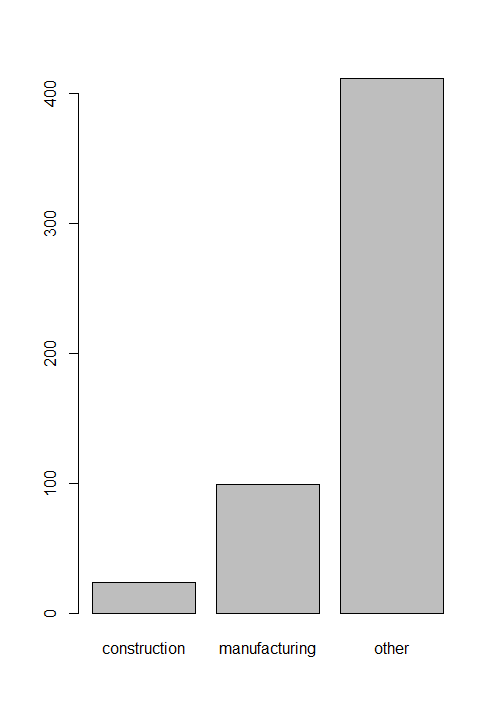
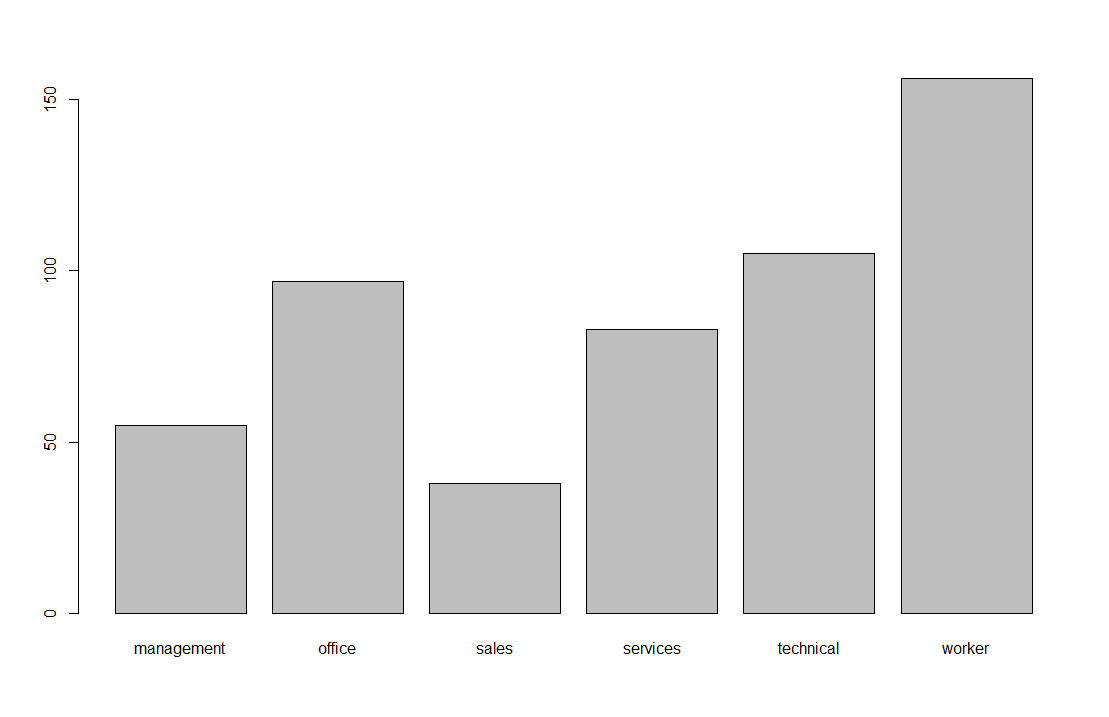
pie(x) ,pie(y)

construction manufacturing other

24 99 411

management office sales services technical worker

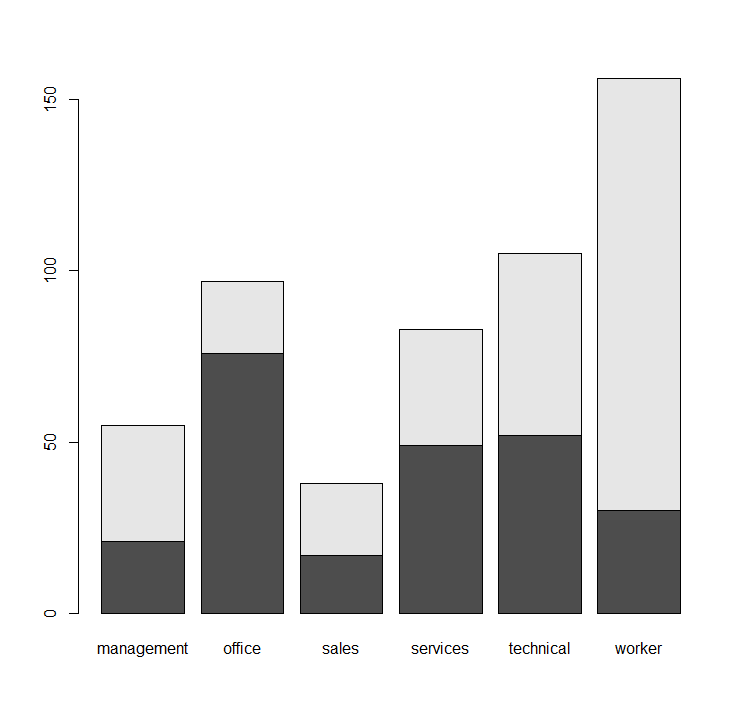
 55 97 38 83 105 156

We can see that construction and manufactoring summed does not exceed half of the population. Varianble other dominates. When we look at below we see that majority of the population is categorized by worker. There is considerably equal distribution between office, services and technical. Sales comes last.

4- Occupation

Gender management office sales services technical worker

female 21 76 17 49 52 30

 male 34 21 21 34 53 126

*White is male*

*Black is Female*

We see that male workers are a lot more than female workes on the other hand we see female majority in the “office”. We observe a considerably equal distribution in sales,services and technical. In managment the population is tended to be male.

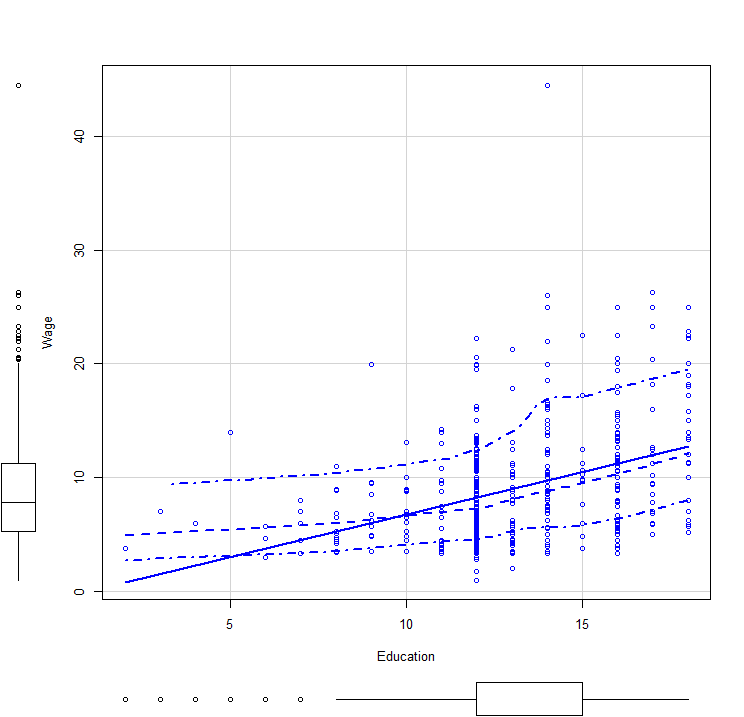
5- cor(wa,edu)

Education ,Wage 0.3819221 Coefficients:

(Intercept) Education

-0.7460 0.7505

We see that without any education wage is negative. But as the party increases its education wage starts to rise. They are positively correlated. By one unit increase in edu wage increases by 0.7505.

6*- Coefficients:*

*(Intercept) Education*

*-0.7460 0.7505*

*(Intercept) Education*

*Estimate -0.746 0.7505*

*Std. Error 1.045 0.0787*

*Adjusted R-squared: 0.1443*

*F-statistic: 90.85 on 1 and 532 DF, p-value: < 2.2e-16*

*Min 1Q Median 3Q Max*

*-7.911 -3.260 -0.760 2.240 34.740*

We experience a low R-squared. Our F-statistics is bigger than the p-value so we can use the metric. İn our Estimaton without any education we have negative wage. But edu and wage is positivly correlated. Ergo, as we exprience a 1 unit increase in edu we observe 0.7505 increase in wage