HW2.R

d0f05lt

2021-10-09

Company : Stevens Project : HW1 Purpose : HW1 First Name : David Last Name : Fu CWID : 10471854 Date : October 1, 2021

Homework 2

```
rm(list=ls())
```

1.

```
cancerData=read.csv("/Users/d0f05lt/School/CS513/HW2/breast-cancer-wisconsin.csv", heade
r=TRUE)
cancerData$F6 <- as.numeric(cancerData$F6)
```

```
## Warning: NAs introduced by coercion
```

I. Summarize

```
summary(cancerData)
```

```
##      Sample               F1               F2               F3
##  Min.   :   61634   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.:  870688   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000
##  Median : 1171710   Median : 4.000   Median : 1.000   Median : 1.000
##  Mean   : 1071704   Mean   : 4.418   Mean   : 3.134   Mean   : 3.207
##  3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000
##  Max.   :13454352   Max.   :10.000   Max.   :10.000   Max.   :10.000
##
##        F4               F5               F6               F7
##  Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.: 1.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 2.000
##  Median : 1.000   Median : 2.000   Median : 1.000   Median : 3.000
##  Mean   : 2.807   Mean   : 3.216   Mean   : 3.545   Mean   : 3.438
##  3rd Qu.: 4.000   3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 5.000
##  Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
##                                    NA's   :16
##        F8               F9            Class
##  Min.   : 1.000   Min.   : 1.000   Min.   :2.00
##  1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:2.00
##  Median : 1.000   Median : 1.000   Median :2.00
##  Mean   : 2.867   Mean   : 1.589   Mean   :2.69
##  3rd Qu.: 4.000   3rd Qu.: 1.000   3rd Qu.:4.00
##  Max.   :10.000   Max.   :10.000   Max.   :4.00
##
```

II. Find Missing Value

```
cancerData[rowSums(is.na(cancerData)) > 0,]
```

```
##      Sample F1 F2 F3 F4 F5 F6 F7 F8 F9 Class
## 24  1057013  8  4  5  1  2 NA  7  3  1     4
## 41  1096800  6  6  6  9  6 NA  7  8  1     2
## 140 1183246  1  1  1  1  1 NA  2  1  1     2
## 146 1184840  1  1  3  1  2 NA  2  1  1     2
## 159 1193683  1  1  2  1  3 NA  1  1  1     2
## 165 1197510  5  1  1  1  2 NA  3  1  1     2
## 236 1241232  3  1  4  1  2 NA  3  1  1     2
## 250  169356  3  1  1  1  2 NA  3  1  1     2
## 276  432809  3  1  3  1  2 NA  2  1  1     2
## 293  563649  8  8  8  1  2 NA  6 10  1     4
## 295  606140  1  1  1  1  2 NA  2  1  1     2
## 298   61634  5  4  3  1  2 NA  2  3  1     2
## 316  704168  4  6  5  6  7 NA  4  9  1     2
## 322  733639  3  1  1  1  2 NA  3  1  1     2
## 412 1238464  1  1  1  1  1 NA  2  1  1     2
## 618 1057067  1  1  1  1  1 NA  1  1  1     2
```

III. Replace Missing Value with mean

```
meanF6 <- mean(cancerData$F6 , na.rm=TRUE)
cancerData$F6[is.na(cancerData$F6)] <- meanF6
```

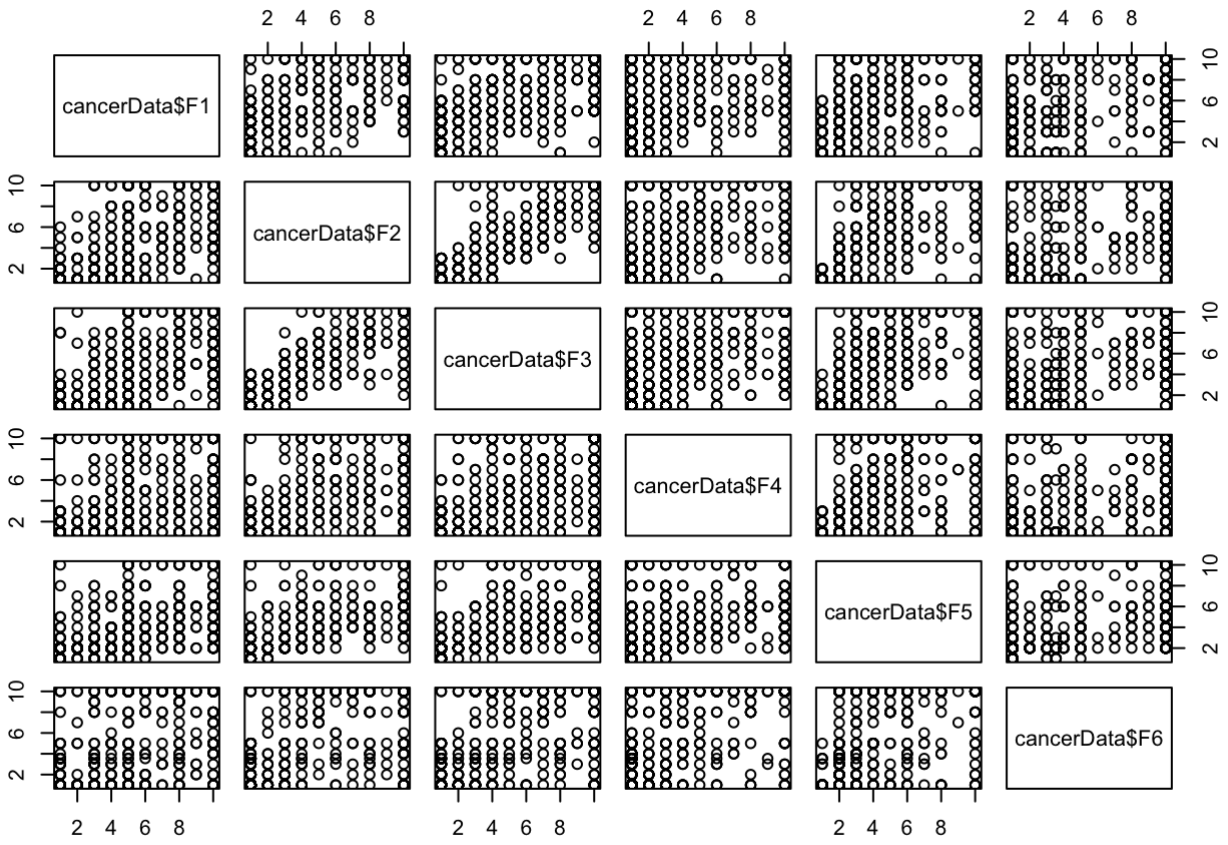
IV. Displaying the frequency table of Class vs F6

```
table(cancerData$Class,cancerData$F6)
```

```
##
##       1   2   3 3.54465592972182   4   5   6   7   8   9  10
##   2 387  21  14                14   6  10   0   1   2   0   3
##   4  15   9  14                 2  13  20   4   7  19   9 129
```

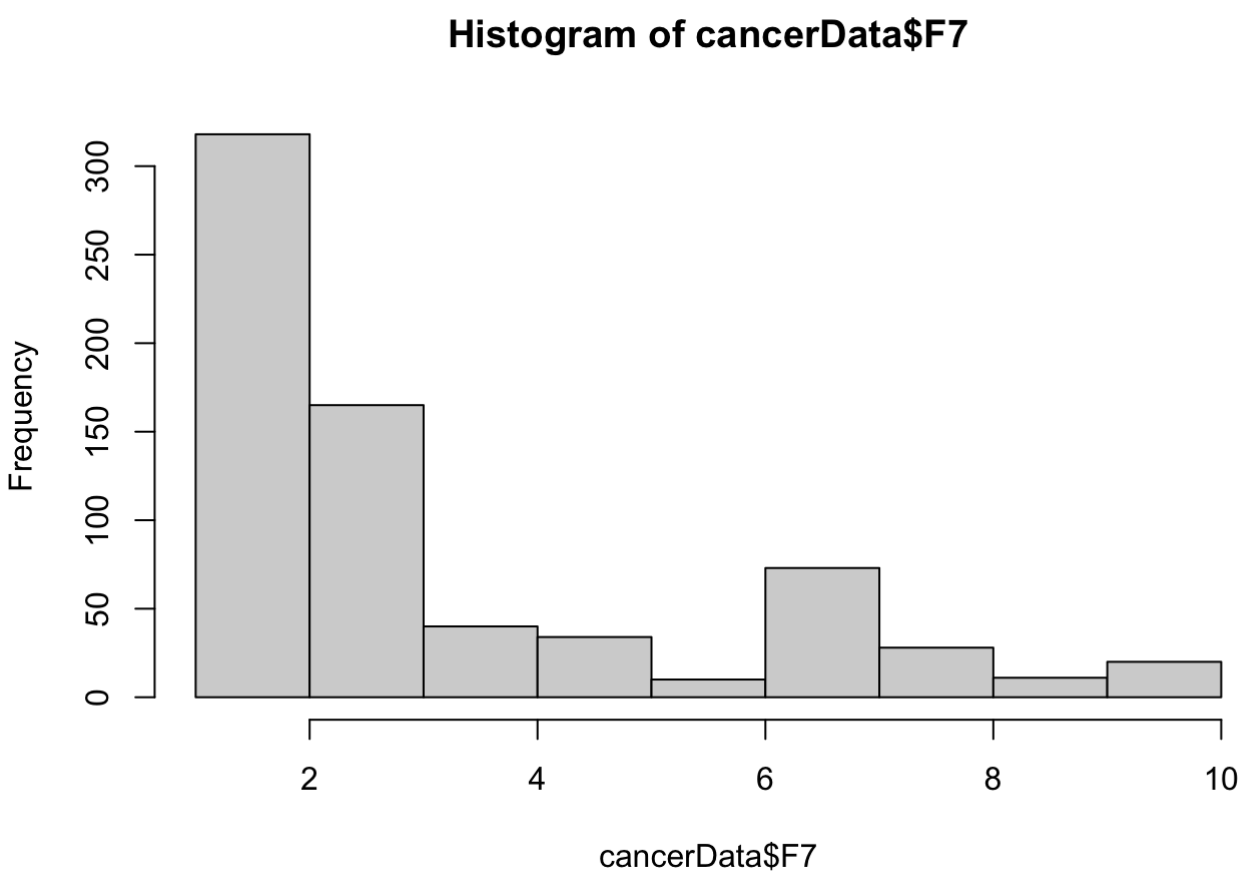
V. Displaying scatter plot of F1 to F6

```
pairs(~cancerData$F1+cancerData$F2+cancerData$F3+cancerData$F4+cancerData$F5+cancerData$F6)
```

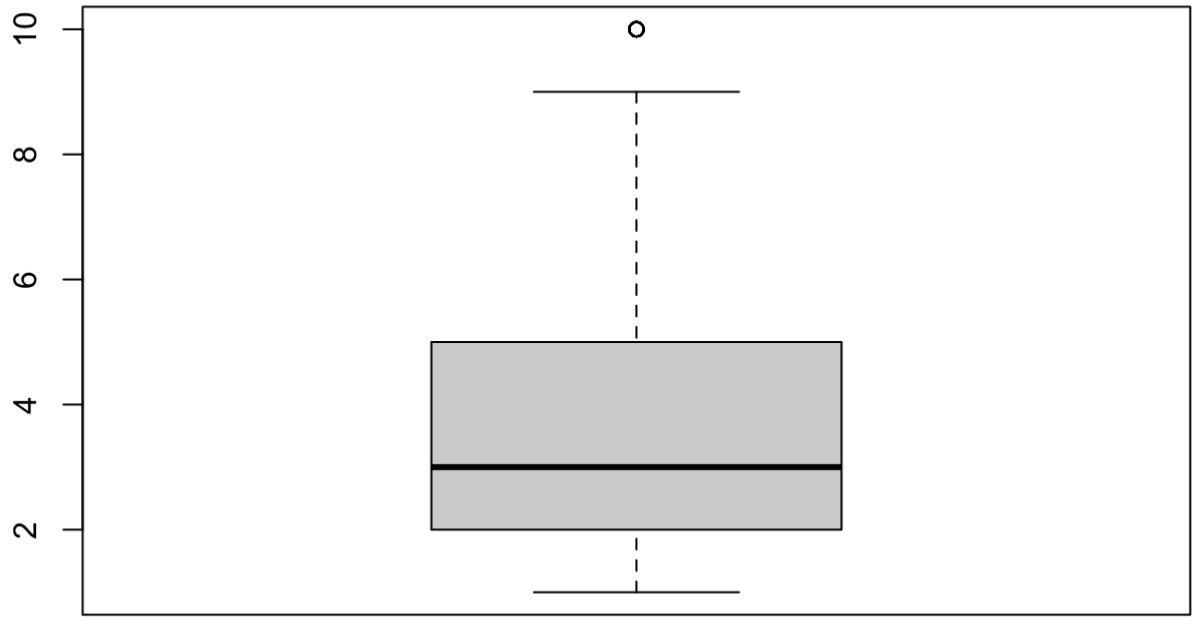


VI. Show historgram box plot for F7 to F9

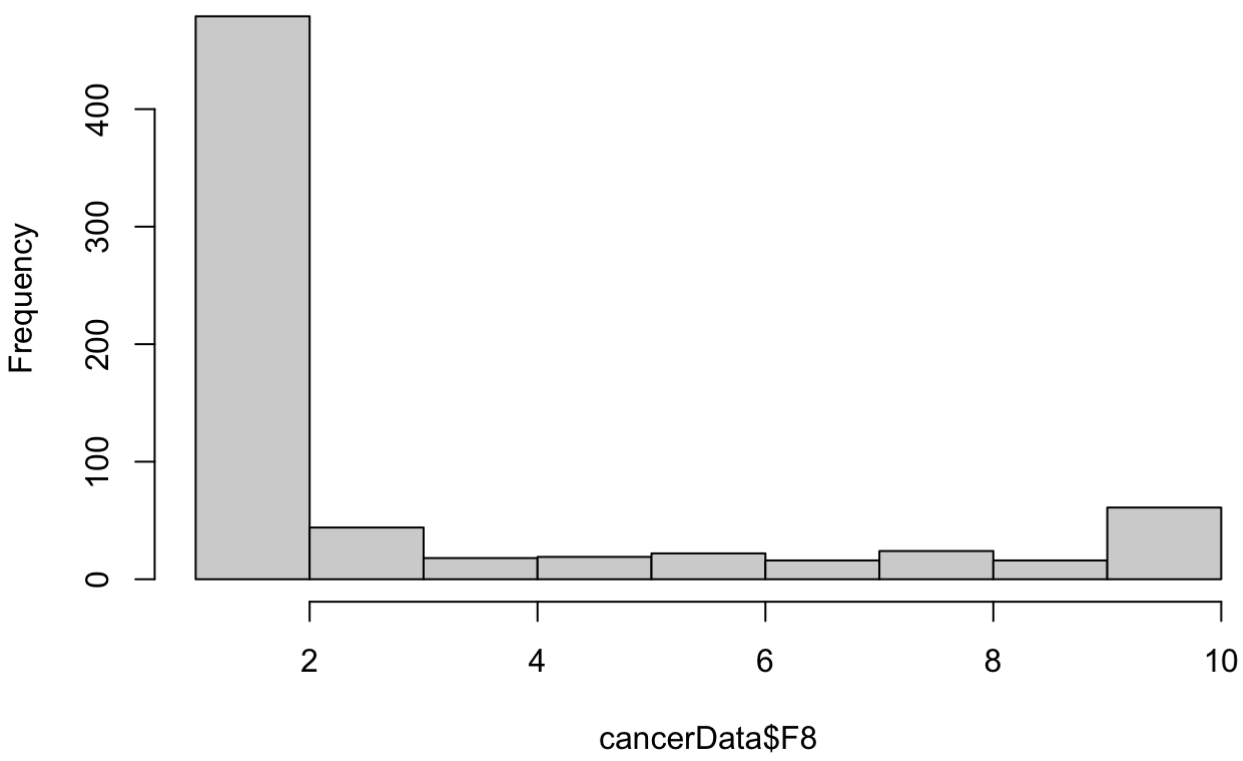
```
hist(cancerData$F7)
```



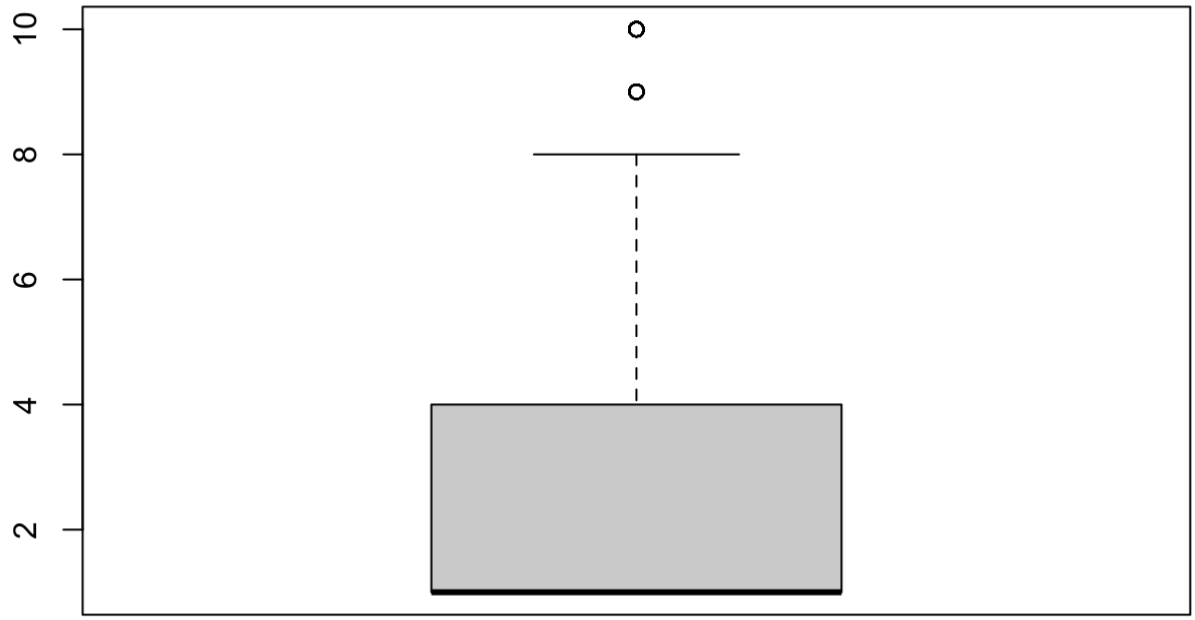

```
boxplot(cancerData$F7)
```



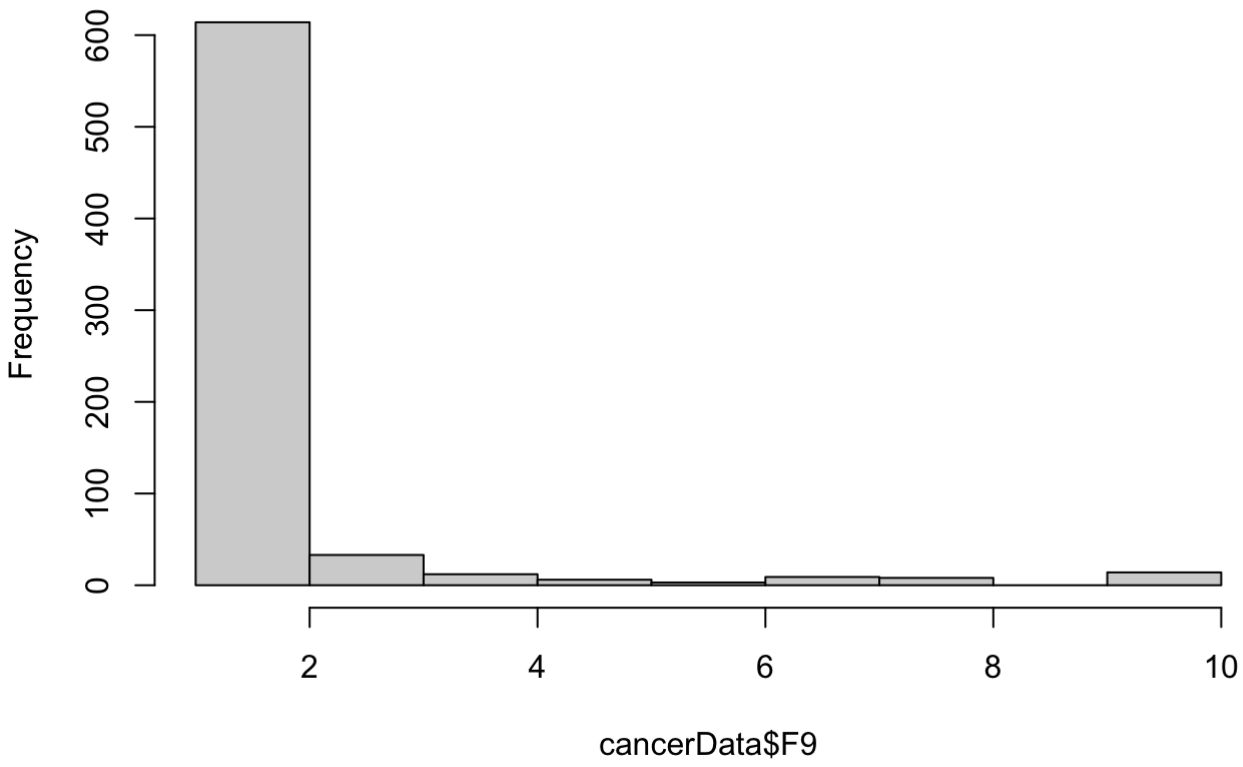

```
hist(cancerData$F8)
```



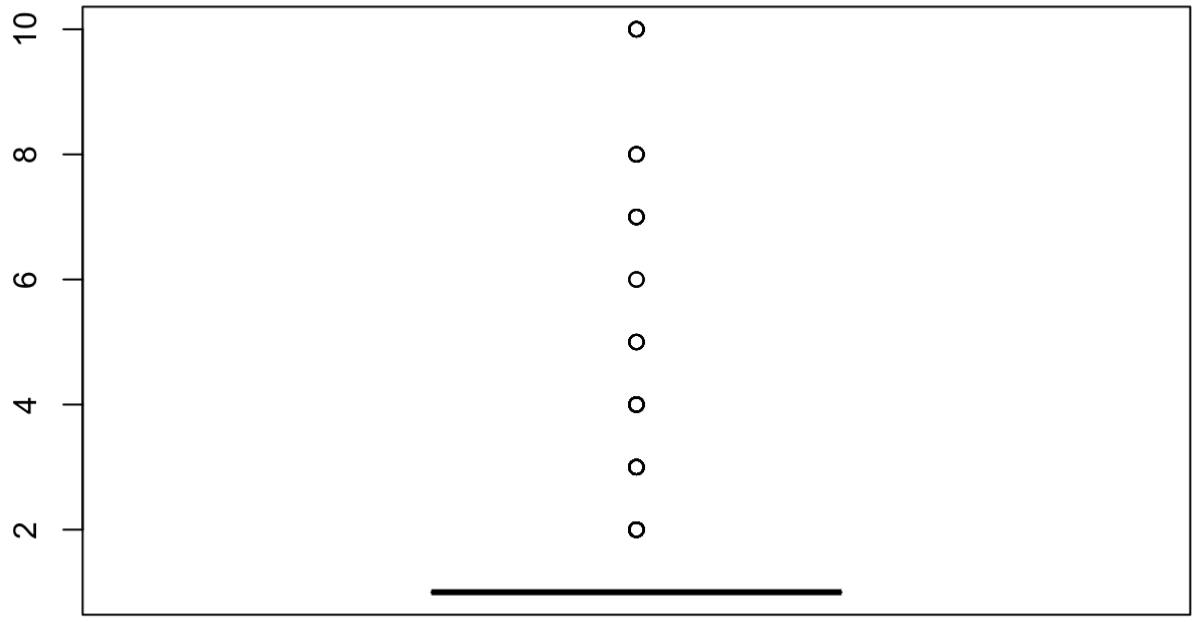

```
boxplot(cancerData$F8)
```




```
hist(cancerData$F9)
```




```
boxplot(cancerData$F9)
```



Delete all Object and Reload Data then remove all column with missing value

```
rm(list=ls())
cancerData=read.csv("/Users/d0f05lt/School/CS513/HW2/breast-cancer-wisconsin.csv", heade
r=TRUE)

cancerData$F6 <- as.numeric(cancerData$F6)
```

```
## Warning: NAs introduced by coercion
```

```
cancerData <- cancerData[rowSums(is.na(cancerData)) == 0,]
```