CS513 HW2: EDA

Aidan Fischer - 10447681

2023-10-07

1)

I)

```
rm(list=ls())
data = read.csv("breast-cancer-wisconsin.csv")
data$F6 <- suppressWarnings(as.numeric(data$F6))
summary(data)
```

```
##      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)

```
data[!complete.cases(data), ]
```

```
##      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)

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

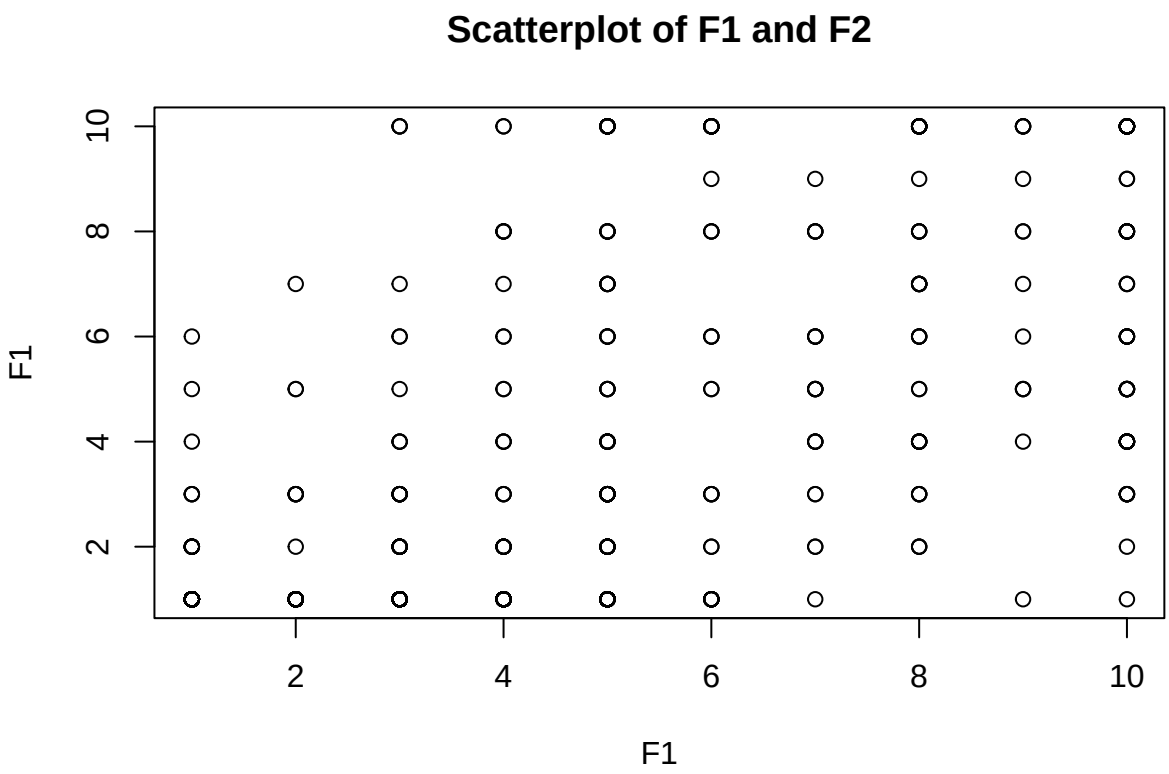
IV)

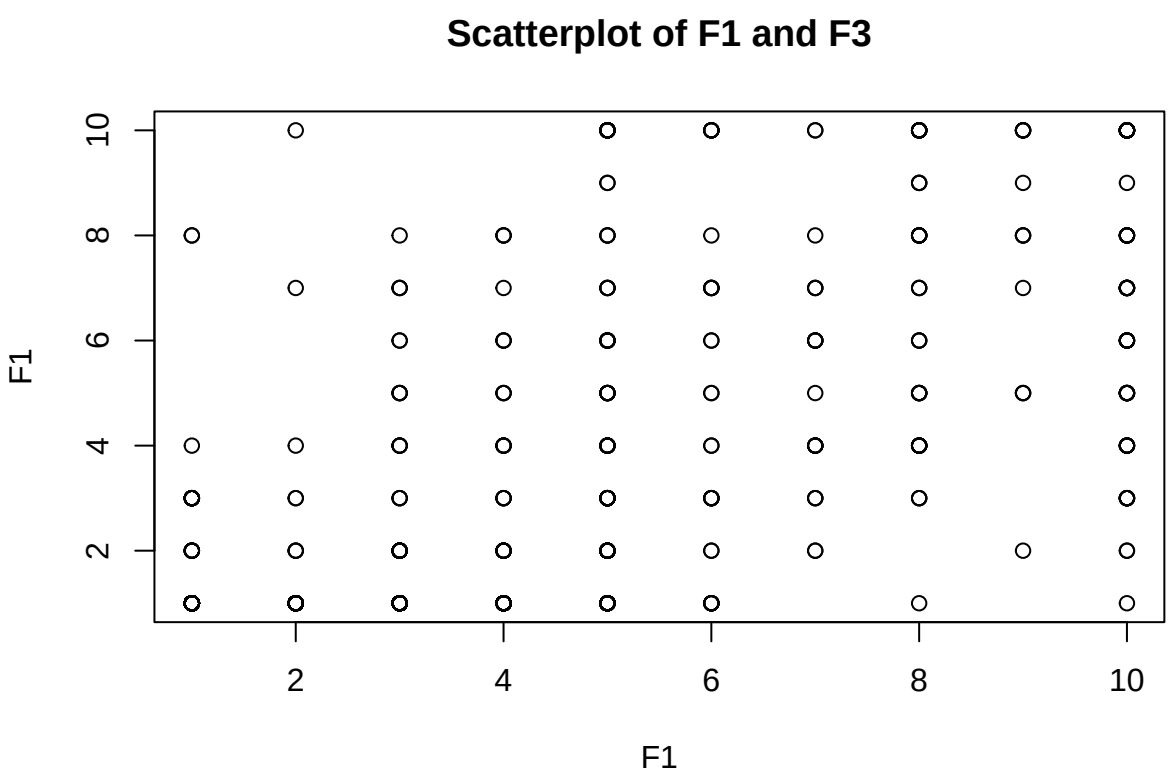
```
table(data$Class, data$F6)
```

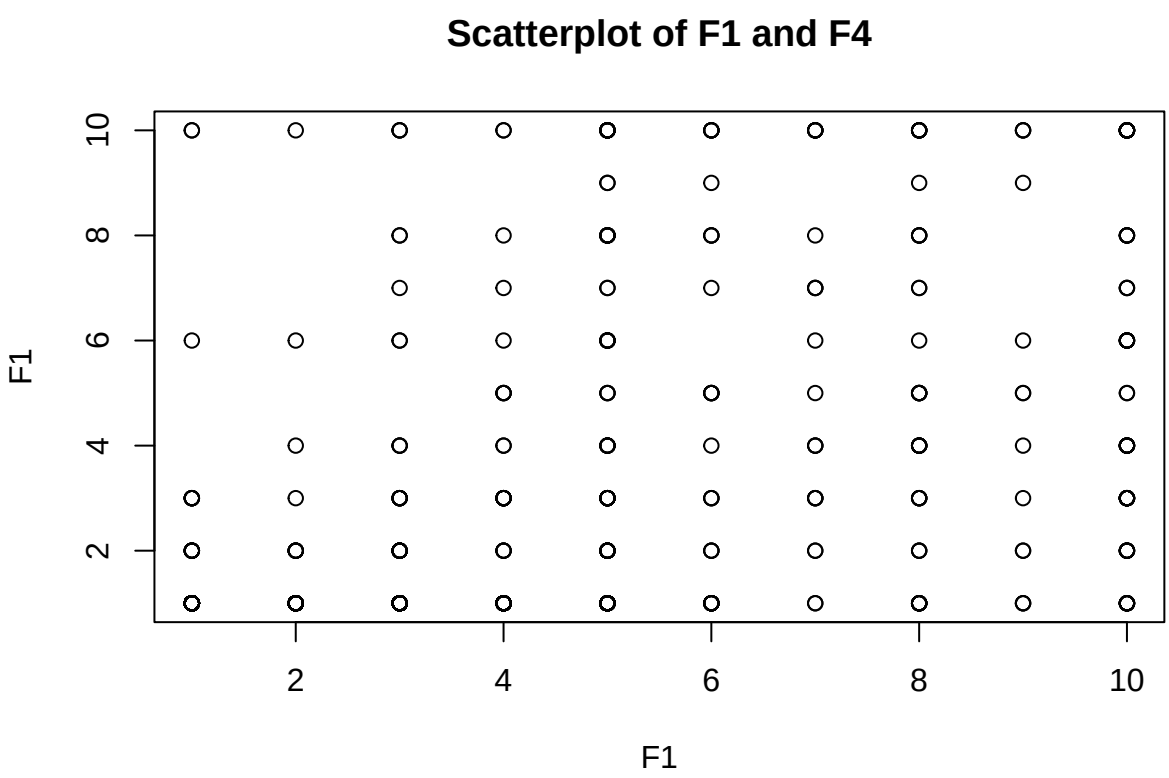
```
##
##       1   2   3 3.54465592972182   4   5   6   7   8   9  10
##   2 387  21  14               14   6  10   0   1   2   0   3
##   4  15   9  14                2  13  20   4   7  19   9 129
```

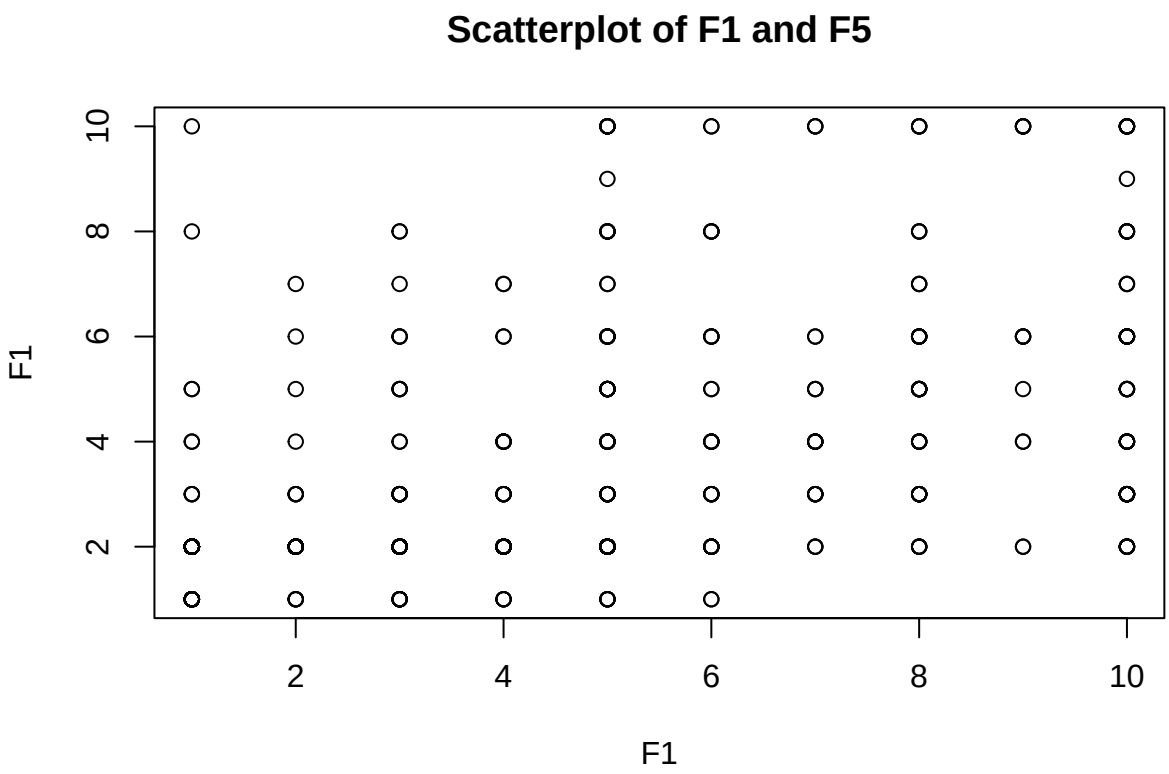
V)

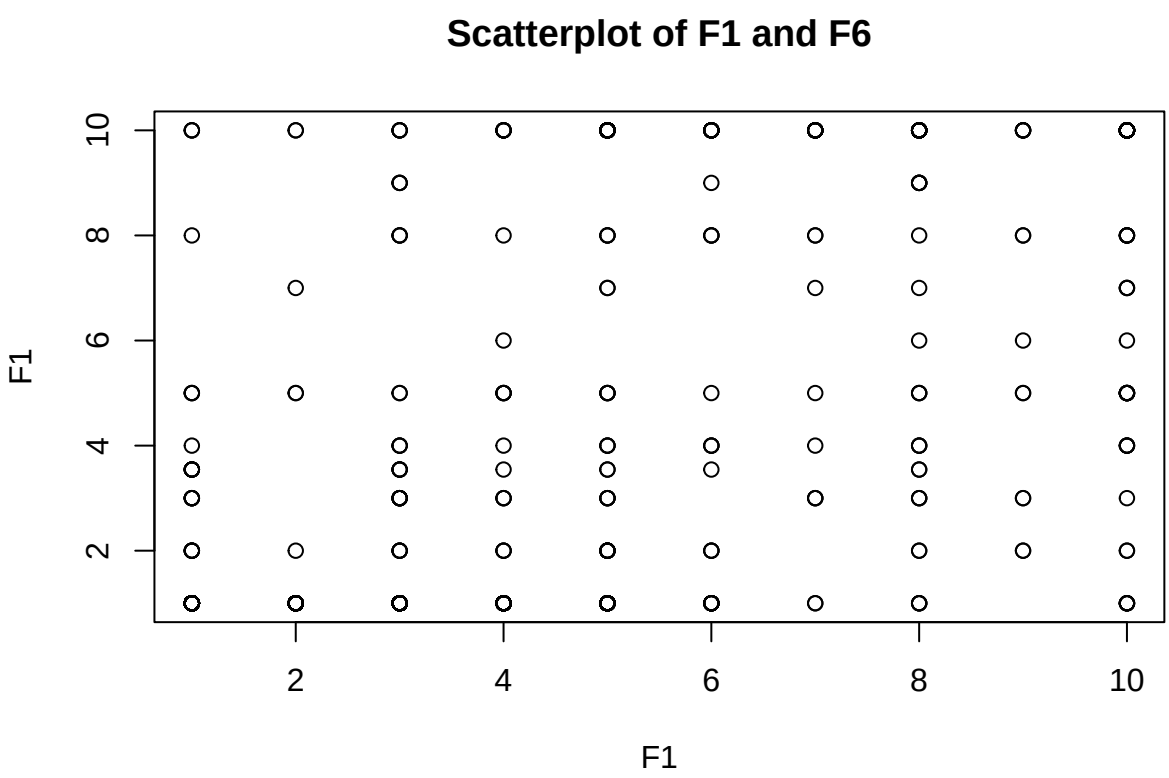
```
for (i in 1:5) {
  for (j in (i+1):6){
    f = paste("F",i,sep='')
    g = paste("F",j,sep='')
    plot(data[[f]],
         data[[g]],
         main=paste("Scatterplot of F",i," and F",j,sep=''),
         xlab=paste("F",i,sep=''),
         ylab=paste("F",i,sep=''))
  }
}
```

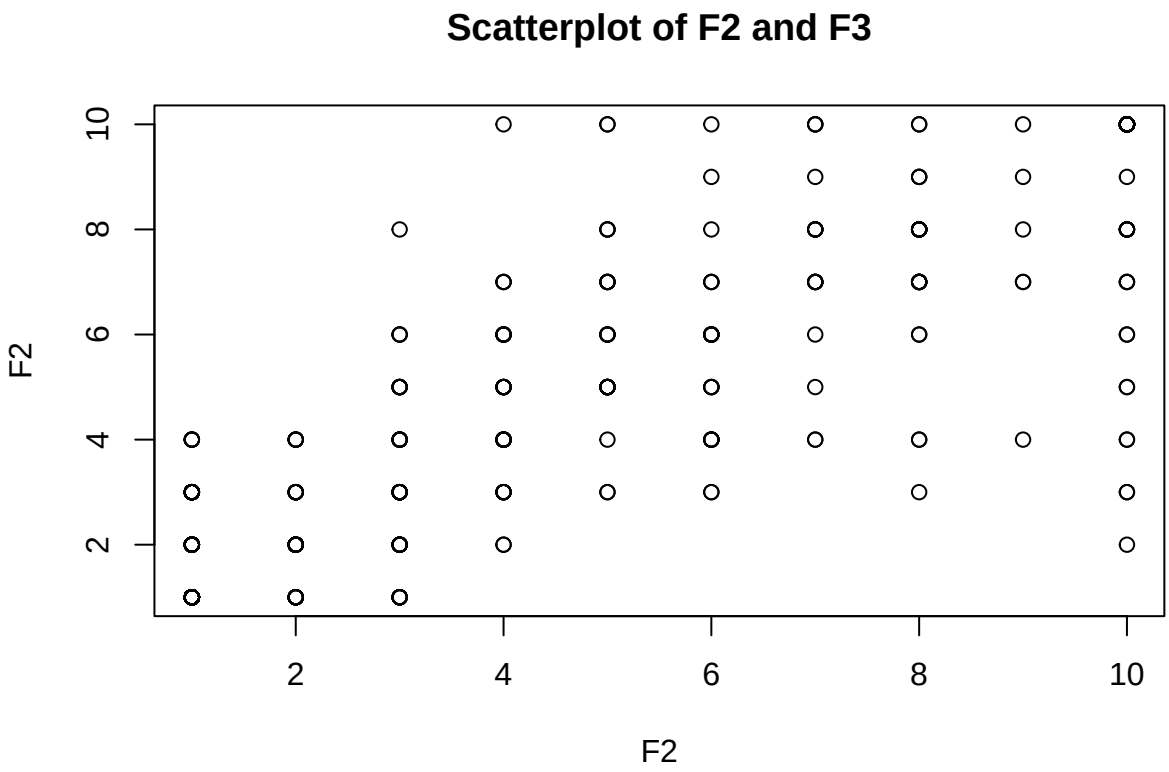

Scatterplot of F1 and F2
F1
F1
2
4
6
8
10
2
4
6
8
10

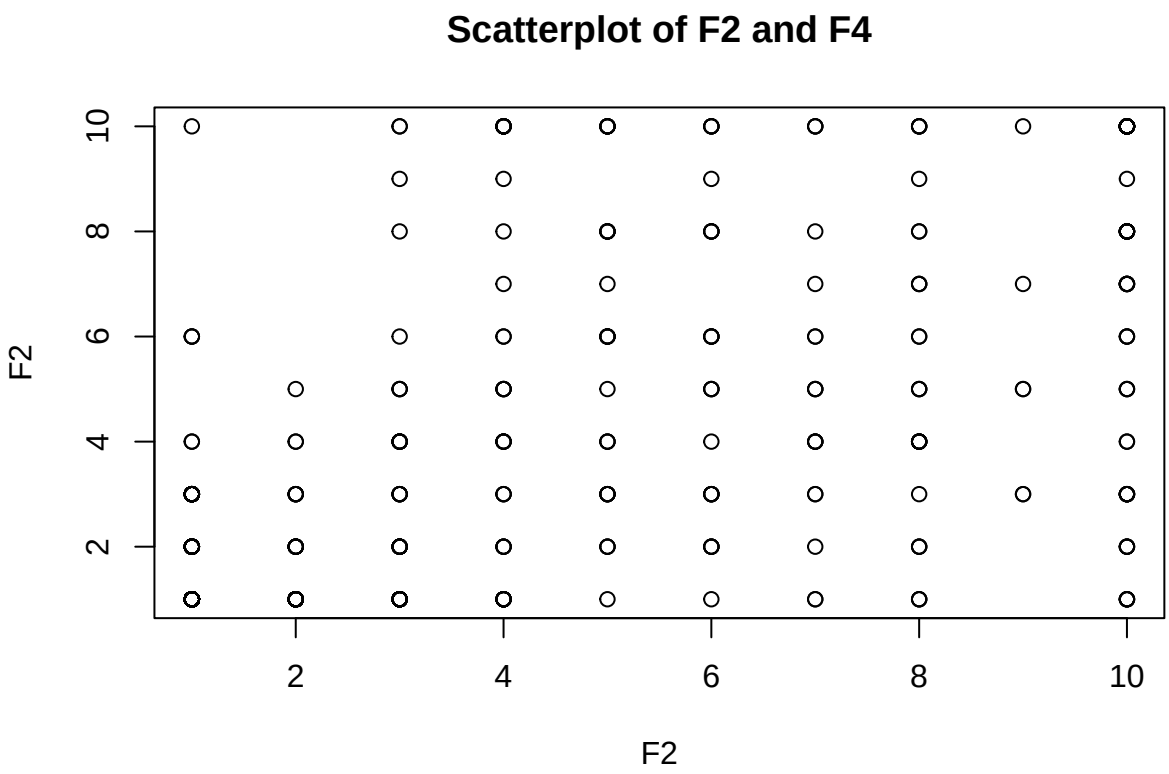

Scatterplot of F1 and F3
F1
F1
2
4
6
8
10
2
4
6
8
10

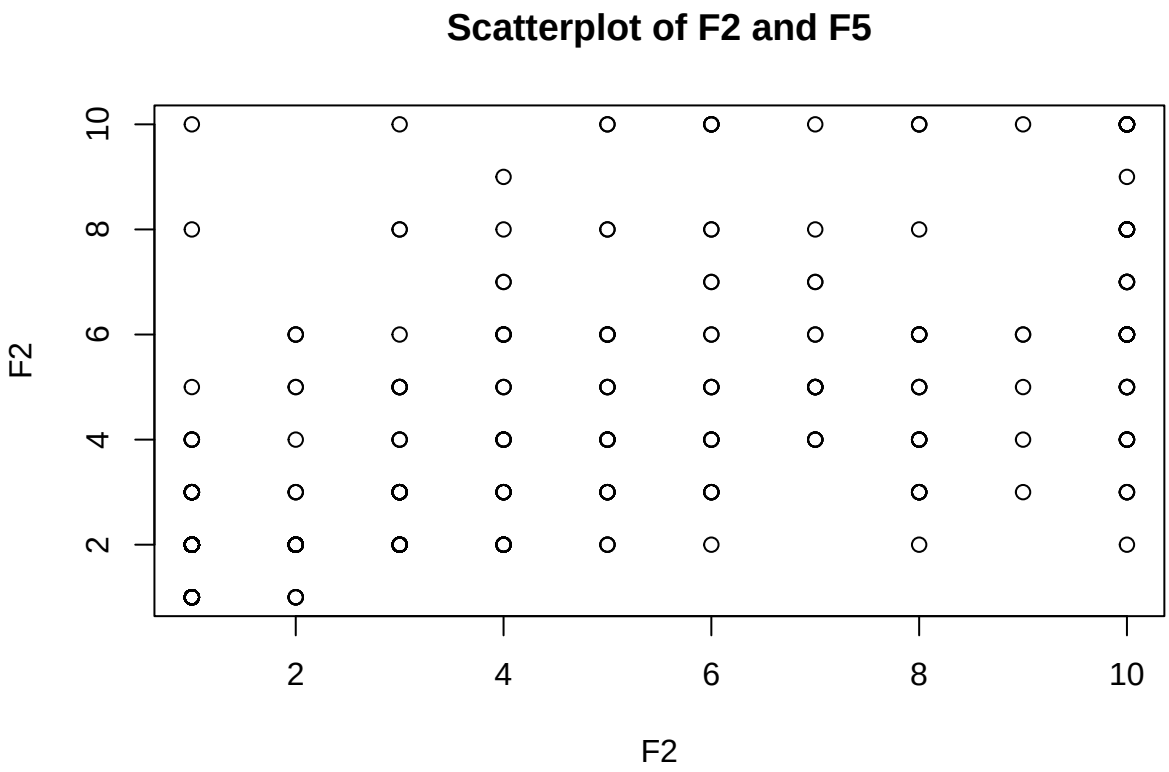

Scatterplot of F1 and F4
F1
F1
2
4
6
8
10
2
4
6
8
10

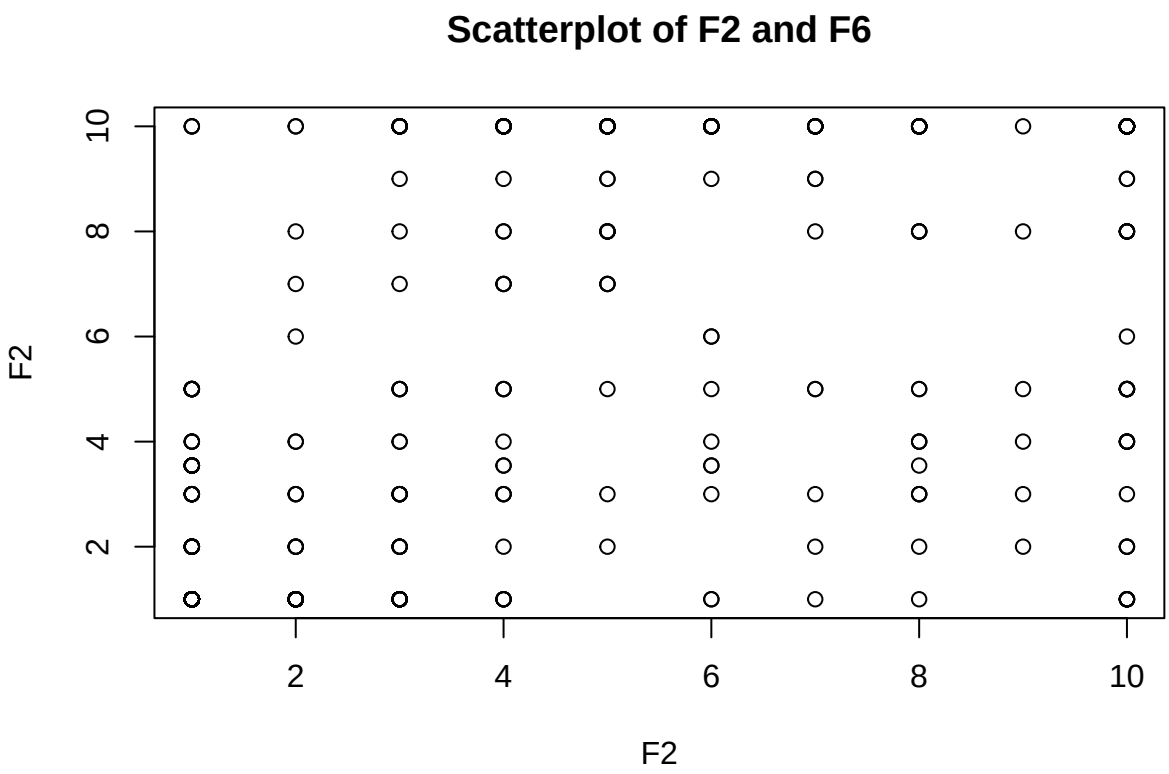

Scatterplot of F1 and F5
F1
F1
2
4
6
8
10
2
4
6
8
10

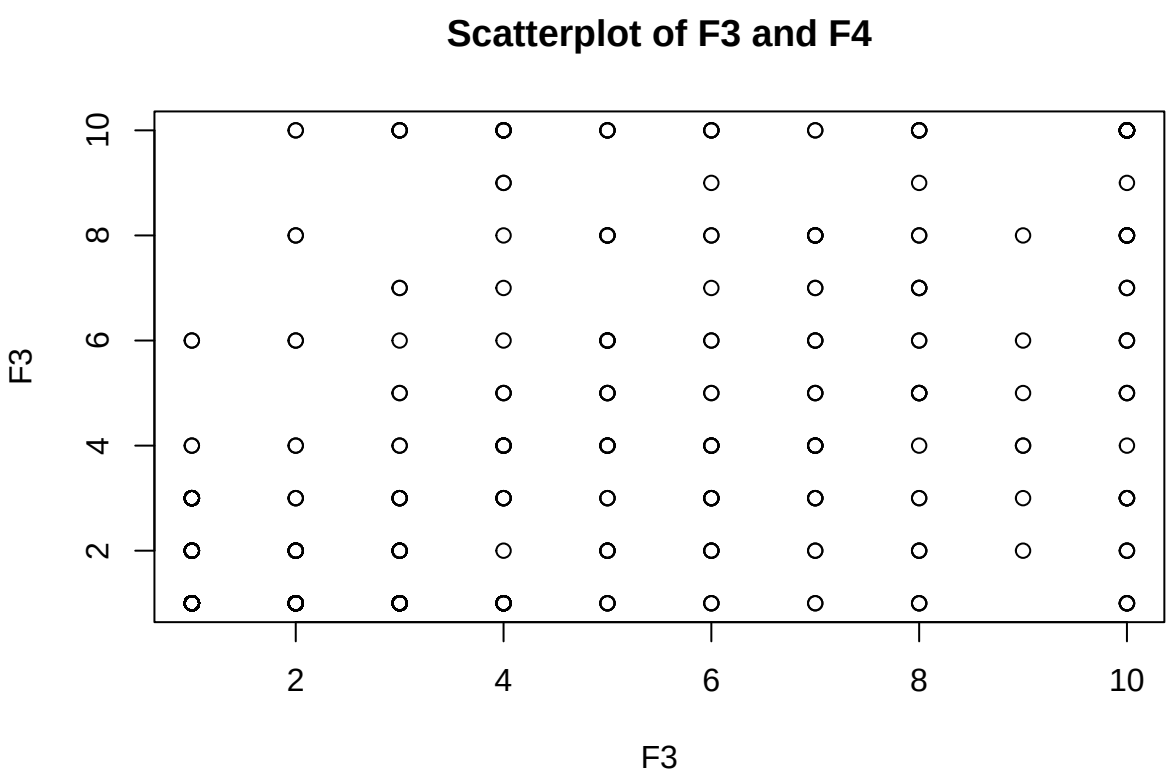

Scatterplot of F1 and F6
F1
F1
2
4
6
8
10
2
4
6
8
10

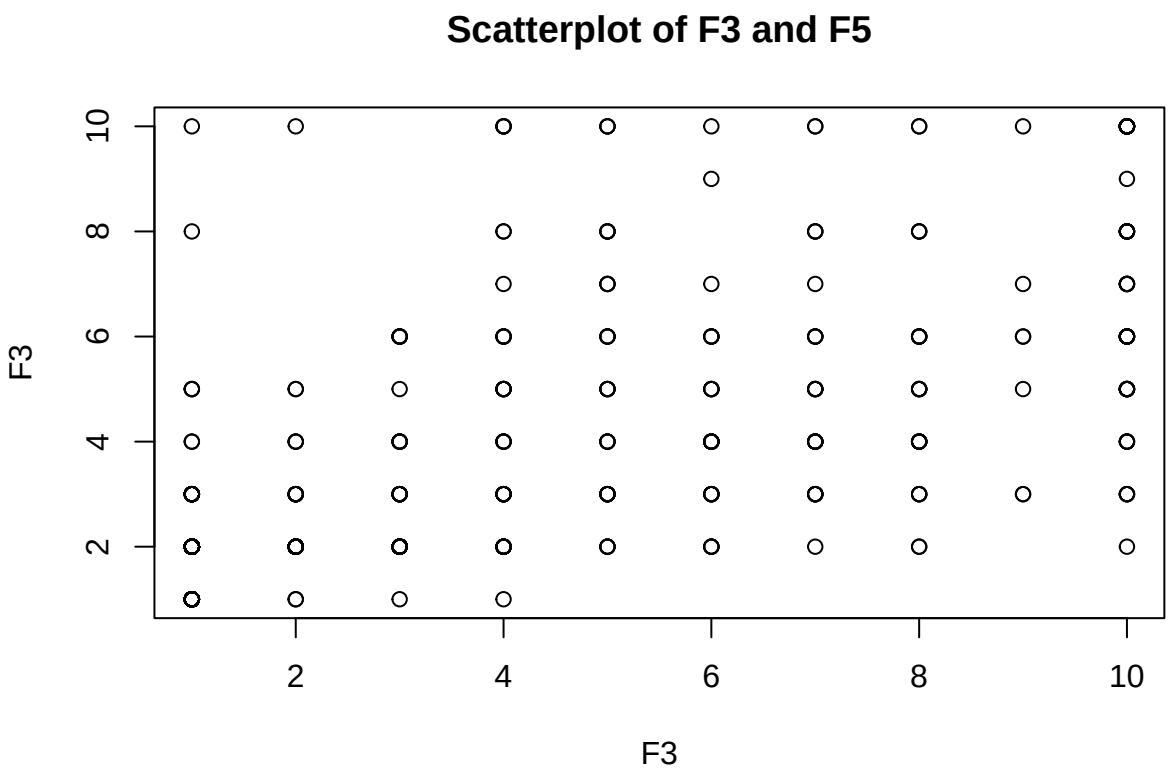

Scatterplot of F2 and F3
F2
F2
2
4
6
8
10
2
4
6
8
10

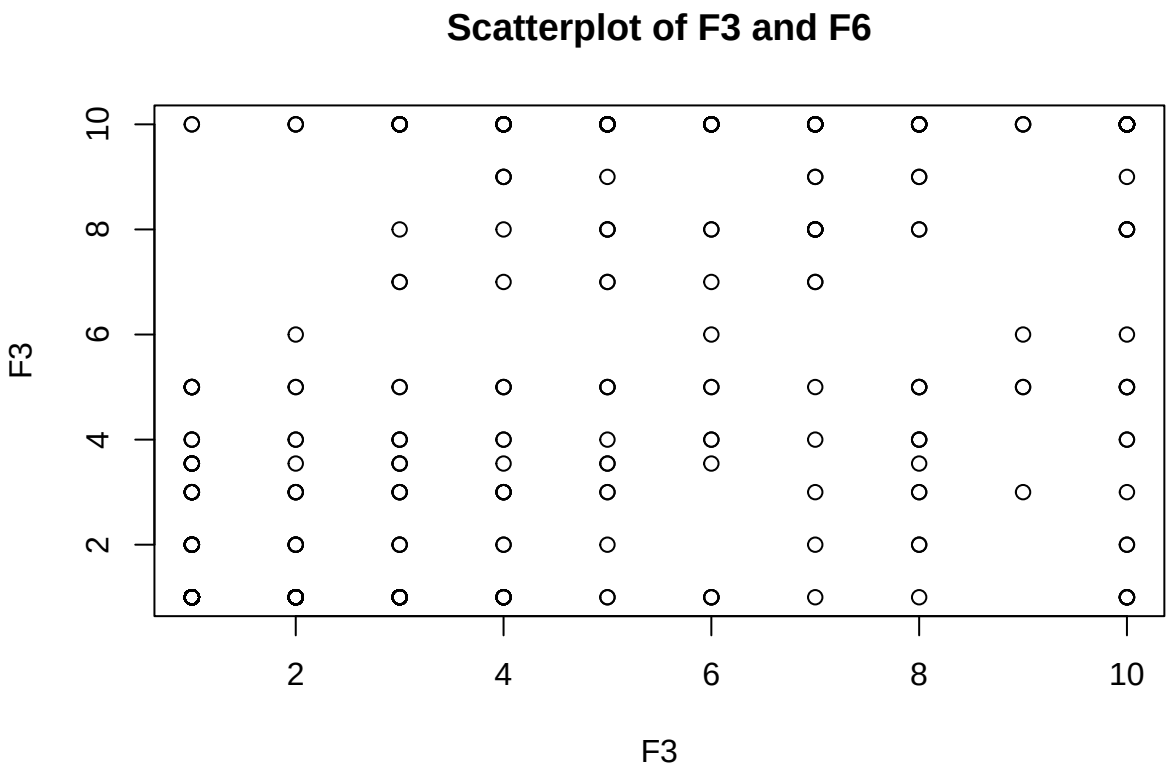

Scatterplot of F2 and F4
F2
2
4
6
8
10
F2
2
4
6
8
10

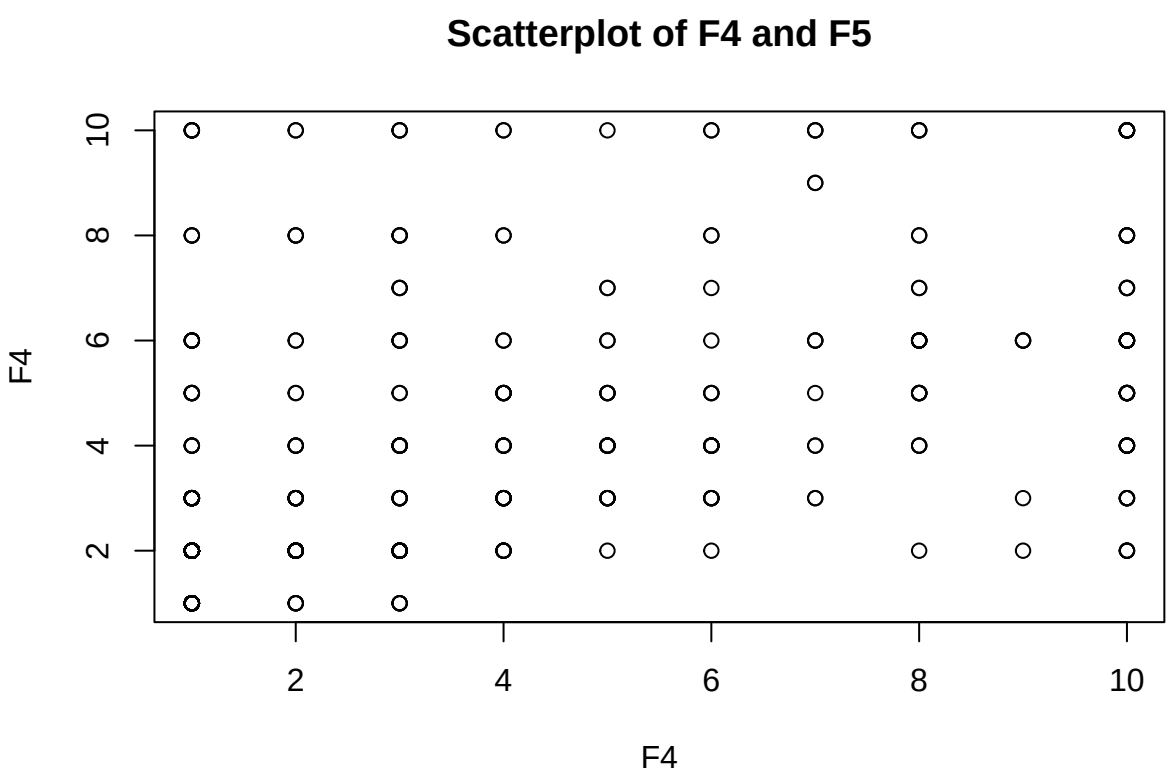

Scatterplot of F2 and F5
F2
10
8
6
4
2
2
4
6
8
10
F2

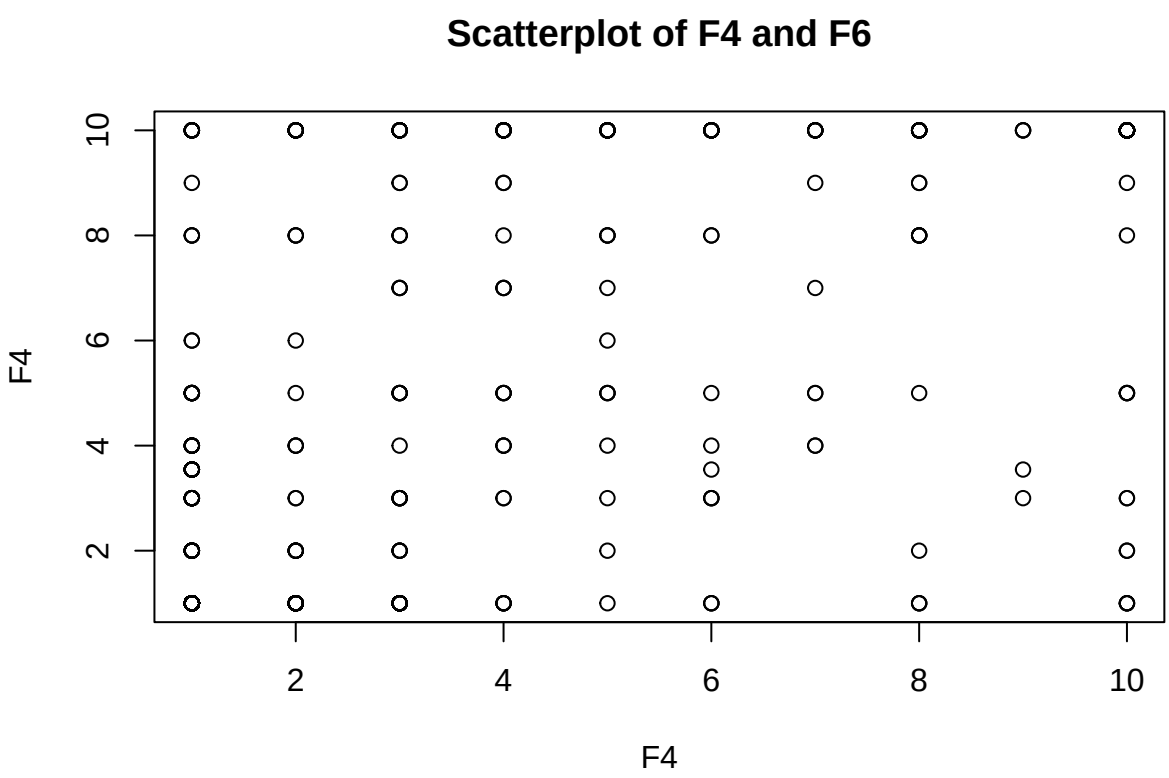

Scatterplot of F2 and F6
F2
2
4
6
8
10
F2
2
4
6
8
10

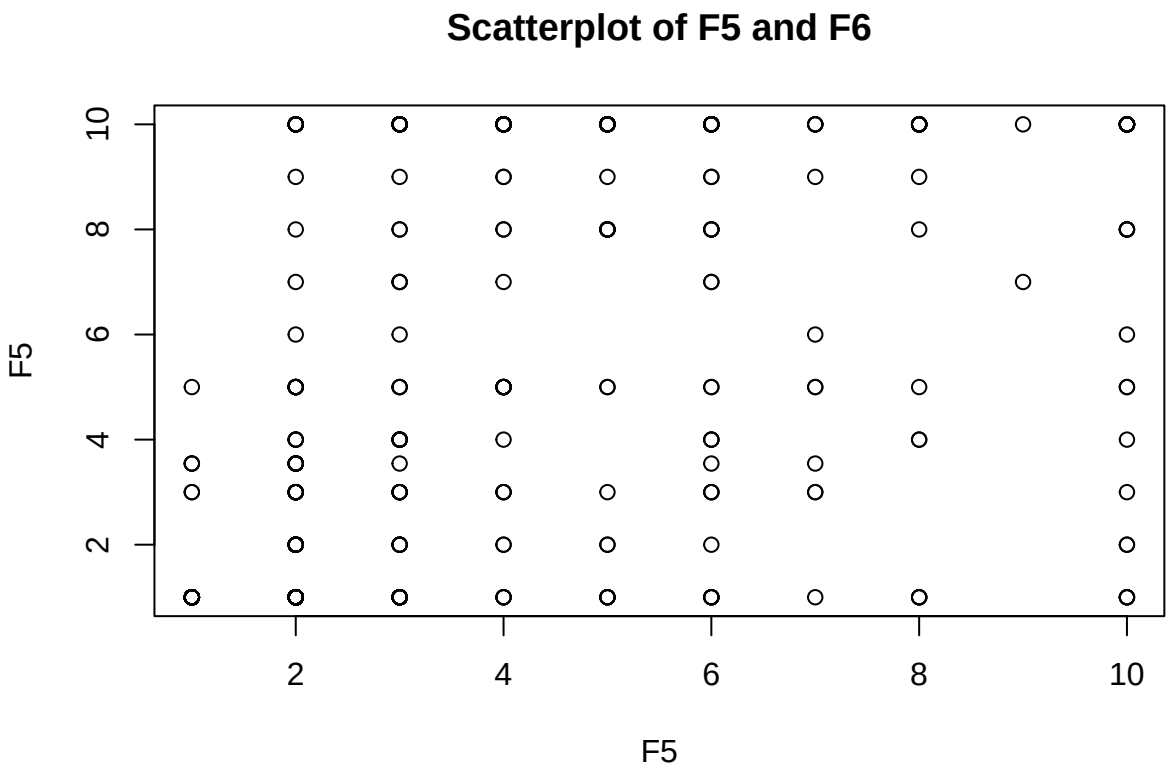

Scatterplot of F3 and F4
F3
2
4
6
8
10
F3
2
4
6
8
10


Scatterplot of F3 and F5
F3
2
4
6
8
10
F3
2
4
6
8
10


Scatterplot of F3 and F6
F3
2
4
6
8
10
F3
2
4
6
8
10

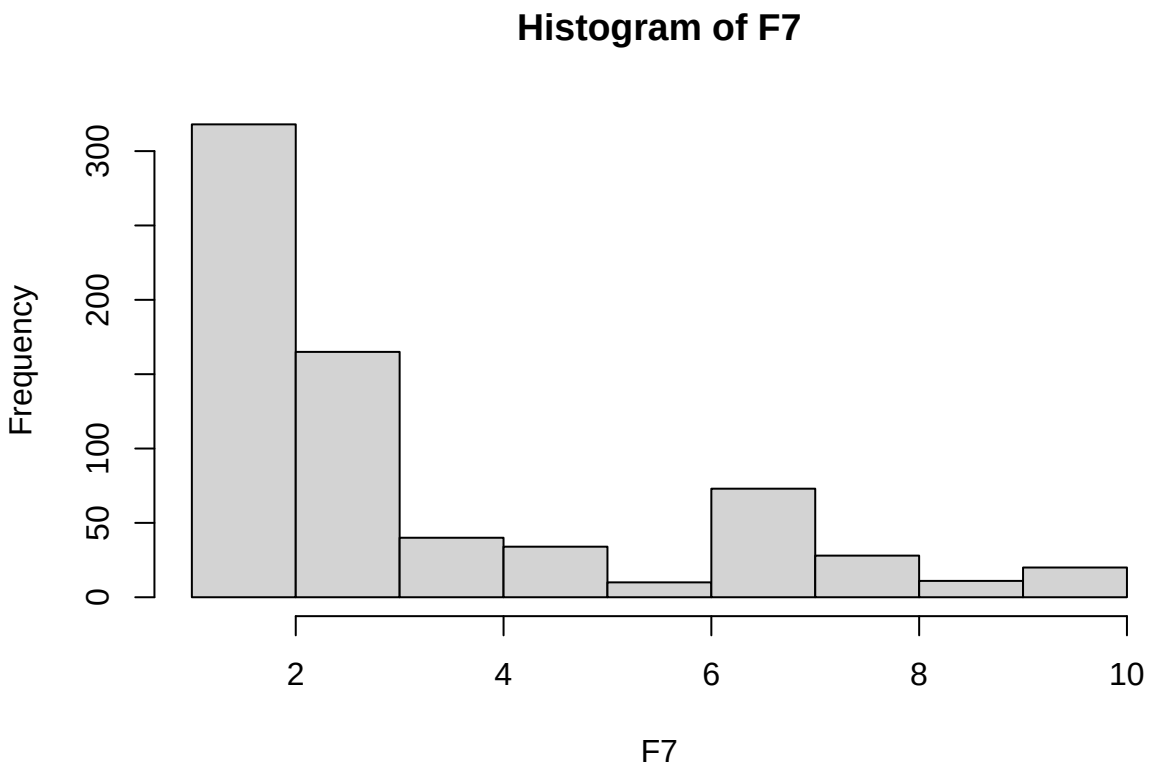

Scatterplot of F4 and F5
F4
F4
2
4
6
8
10
2
4
6
8
10


Scatterplot of F4 and F6
F4
2
4
6
8
10
F4
2
4
6
8
10

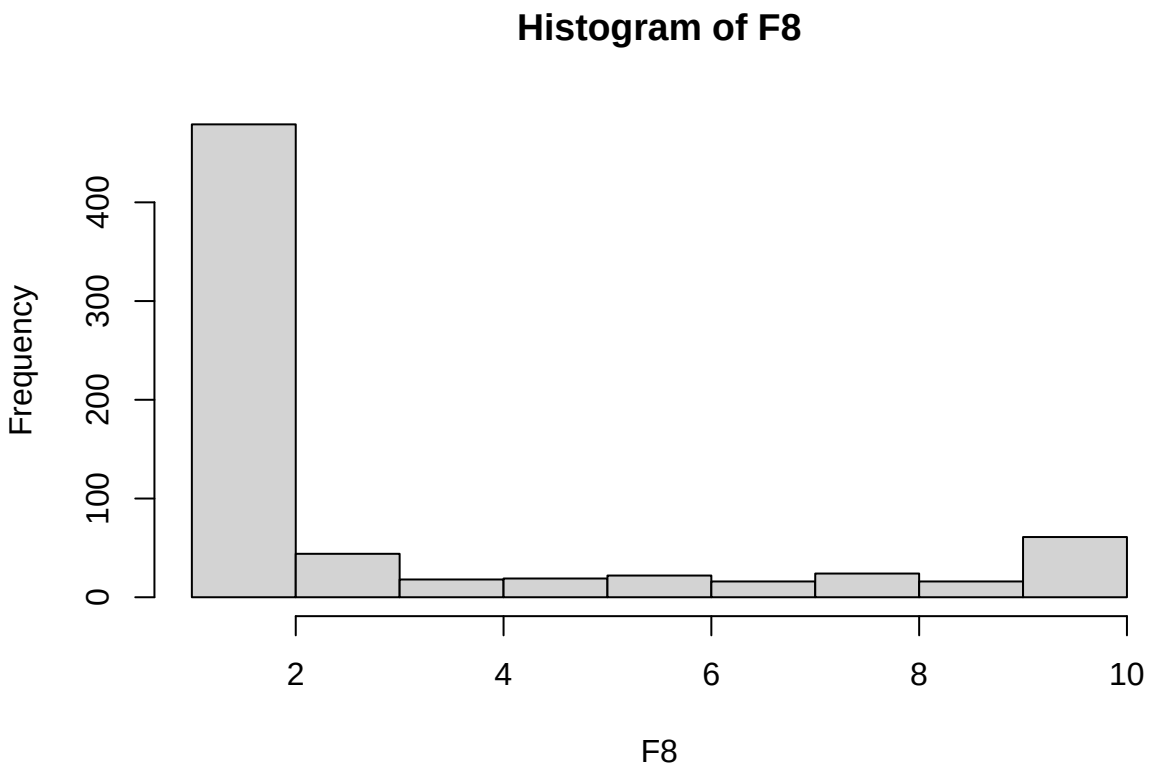


VI)

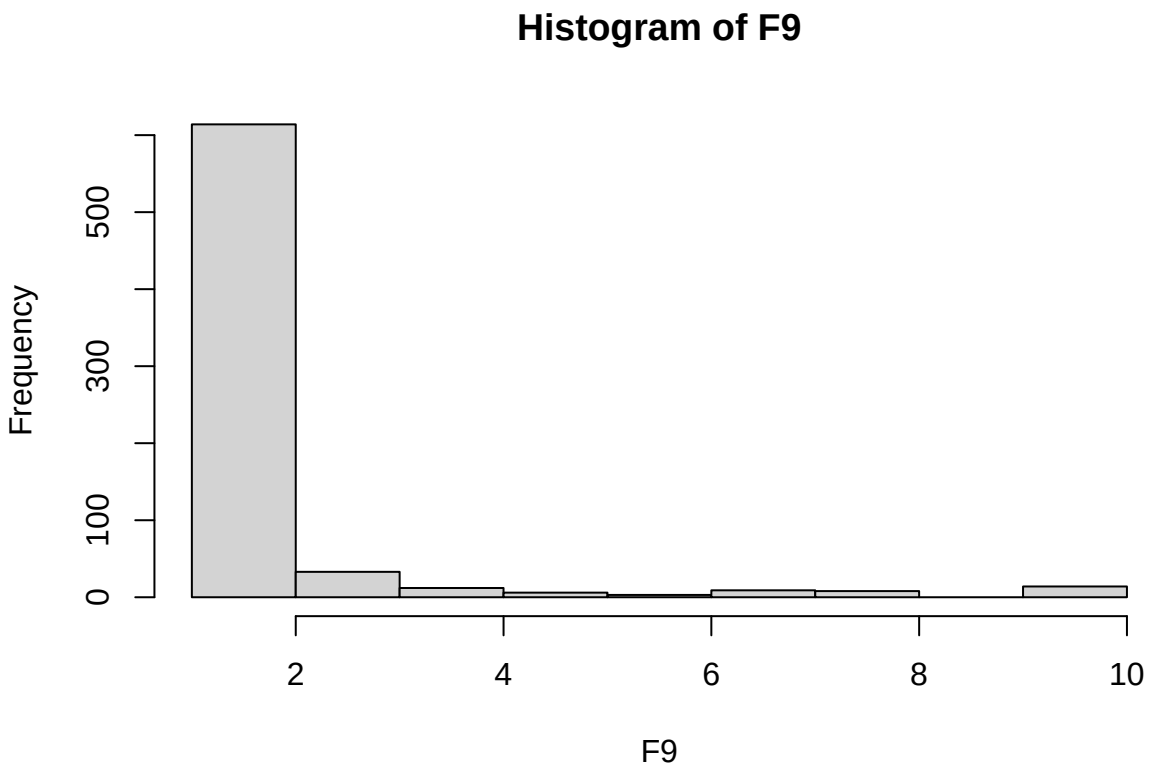
```
hist(data$F7, main="Histogram of F7", xlab="F7")
```




```r
hist(data$F8, main="Histogram of F8", xlab="F8")
```




```r
hist(data$F9, main="Histogram of F9", xlab="F9")
```



2)

```
rm(list=ls())
data = read.csv("breast-cancer-wisconsin.csv")
data = data[complete.cases(data), ]
```