

Lab 8 - STAT 123

Eric Huber

23/03/2022

Question 1:

(1.a)

```
df = read.csv("nba_player_data_2020.csv")
```

(1.b)

```
df = na.omit(df)
row_sub = apply(df, 1, function(row) all(row!=0))
df = df[row_sub,]
```

(1.c)

```
dfc = df[, c(13:16,18:20)]
```

Question 2:

```
colours = c("dodgerblue", "firebrick1", "green3", "orange", "salmon",
"slateblue1")
```

```
cnames = c("2pt makes", "2pt attempts", "2pt %", "FT makes", "Ft attempts",
"FT %")
```

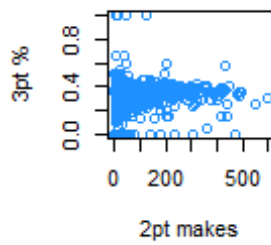
```
par(mfrow= c(2,3))
par(mar = c(5.1, 4.1, 5.1,2.1))
n = dim(dfc)[1]
m = dim(dfc)[2]
```

```
for(i in 2:m){
  ttl = paste("3pt % vs", cnames[i-1])
  plot(dfc[,i], dfc[,1], main = ttl, ylab = "3pt %", col = colours[i-1], xlab
= cnames[i-1])

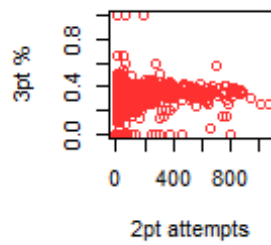
  i = i+1
}
mtext("Correlation Plots", side = 3, line = -1.5, outer = TRUE)
```

Correlation Plots

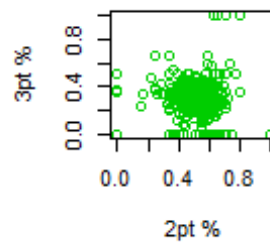
3pt % vs 2pt makes



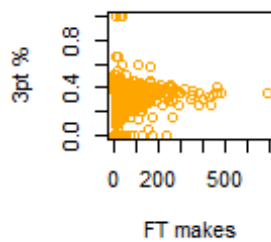
3pt % vs 2pt attempts



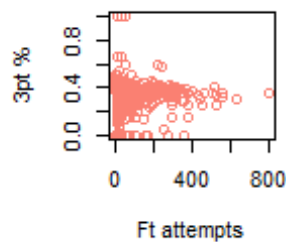
3pt % vs 2pt %



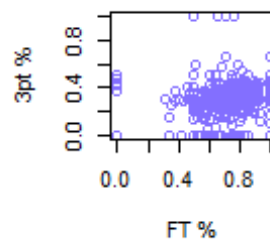
3pt % vs FT makes



3pt % vs Ft attempts



3pt % vs FT %



Question 3:

a

```
cor_vec = numeric()
```

b

```
for(k in 1:m){
  cor_vec[k] = cor(df[,1],df[,k])

  k = k+1
}
```

c

```
names(cor_vec) = c("3pt %", cnames)
```

d

```
cor_vec
```

```
##      3pt %      2pt makes 2pt attempts      2pt %      FT makes  Ft
##      1.0000000  0.09377243  0.11069183 -0.04220238  0.11588018
##      0.09440999
##      FT %
##      0.13846909
```

Question 4:

a

```
easy_way = cor(df_c)[,1]  
names(easy_way) = c("3pt %", cnames)
```

b

```
easy_way  
  
##          3pt %      2pt makes 2pt attempts      2pt %      FT makes  Ft  
attempts  
##  1.00000000  0.09377243  0.11069183 -0.04220238  0.11588018  
0.09440999  
##          FT %  
##  0.13846909
```

(4.c) Besides 3pt % which is of course perfectly and positively correlated, the most positively correlated is the free throw percentage (FT %).