HW4

Kevin Akers

September 18, 2018

Problem 3

According to Peng, the EDA stage acts as a kind of "editing room" for the project. The goals are things like checking for problems with the data and identifying relationships between variables.

Problem 4

```
prob4_data1 <- read.xlsx("HW4_data.xlsx", sheetIndex = 1)
prob4_data2 <- read.xlsx("HW4_data.xlsx", sheetIndex = 2)
hw4 <- data.frame(rbind(prob4_data1, prob4_data2))

#What a nice function!
knitr::kable(summary(hw4), caption = "Data Summary")
```

Table 1: Data Summary

block	depth	phosphate
Min. : 1	Min. :15.56	Min. : 0.01512
1st Qu.: 4	1st Qu.:41.07	1st Qu.:22.56107
Median : 7	Median :52.59	Median :47.59445
Mean : 7	Mean :54.27	Mean :47.83510
3rd Qu.:10	3rd Qu.:67.28	3rd Qu.:71.81078
Max. :13	Max. :98.29	Max. :99.69468

```
#Multipanel plot
par(mfrow = c(2, 2))
hist(hw4$phosphate, main = "Phosphate", xlab = "Phosphate", col = "purple")
hist(hw4$depth, main = "Depth", xlab = "Depth", col = "blue")

#Correlation plot
hw4_corr <- cor(hw4)
corrplot(hw4_corr, method = "circle", bg = "orange")
```
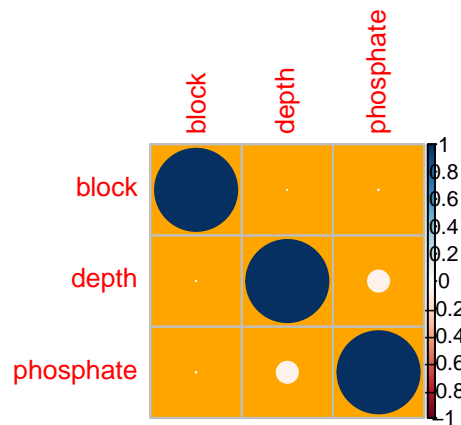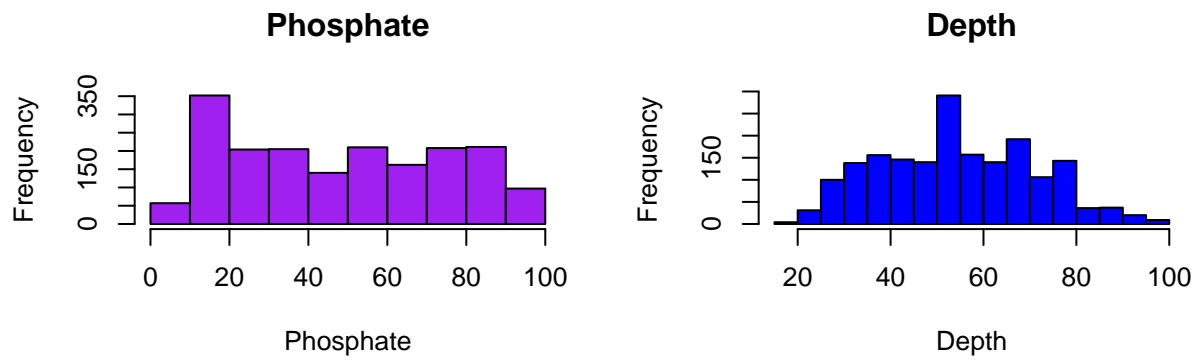
Problem 5

```r
multiplot <- function(x){
  plotgg <- ggplot(x, aes(x[,1], x[,2])) + geom_point() + theme_classic()
  #Really nice function!
  ggMarginal(plotgg, x, type = "histogram", yparams=list(colour="red"))
}
x <- data.frame(cars$speed, cars$dist)
multiplot(x)
```