HW4

Jacob Williams

September 12, 2018

Problem 3

As Roger Peng says "the goal is to show the data, summarize the evidence and identify interesting patterns while eliminating ideas that likely won't pan out.".

Problem 4

```r
library(readxl)
library(knitr)
HW4_data_1 <- read_excel("HW4_data.xlsx",sheet = 1)
HW4_data_2 <- read_excel("HW4_data.xlsx",sheet = 2)
HW4_data_1$day <- rep(1,nrow(HW4_data_1))
HW4_data_2$day <- rep(2,nrow(HW4_data_2))
HW4_data <- rbind(HW4_data_1,HW4_data_2)
HW4_data$block <- as.factor(HW4_data$block)
HW4_data$day <- as.factor(HW4_data$day)
kable(summary(HW4_data[,2:3]))
```

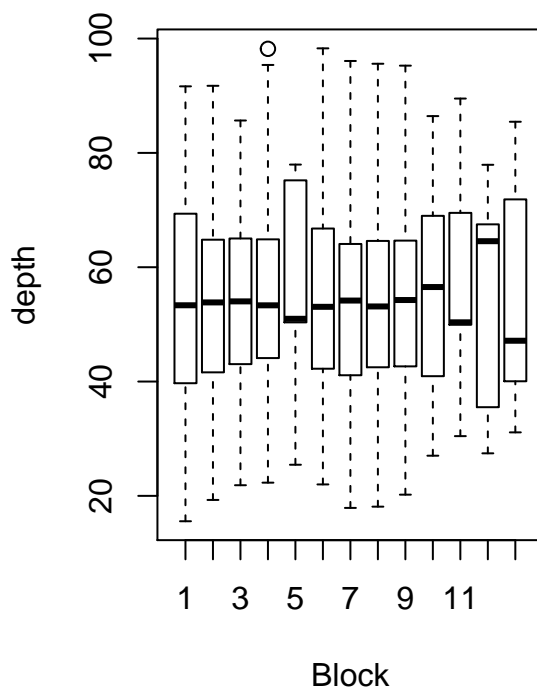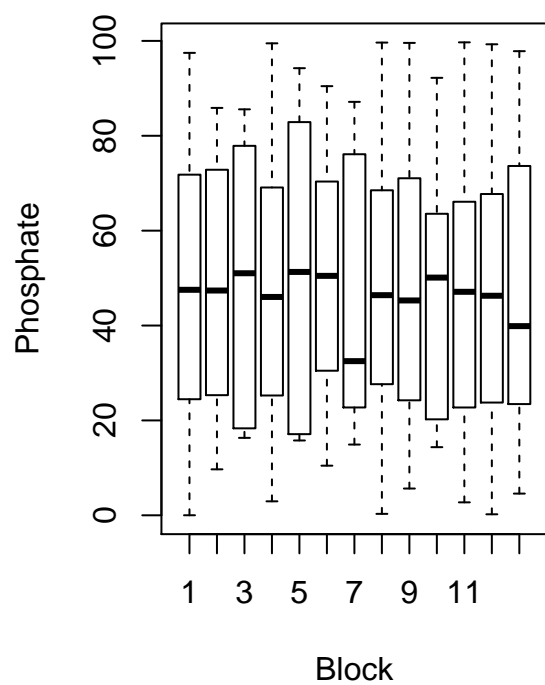| depth	phosphate
Min. :15.56	Min. : 0.01512
1st Qu.:41.07	1st Qu.:22.56107
Median :52.59	Median :47.59445
Mean :54.27	Mean :47.83510
3rd Qu.:67.28	3rd Qu.:71.81078
Max. :98.29	Max. :99.69468

```r
table(HW4_data$block)
```

```
##
##   1   2   3   4   5   6   7   8   9  10  11  12  13
## 142 142 142 142 142 142 142 142 142 142 142 142 142
```
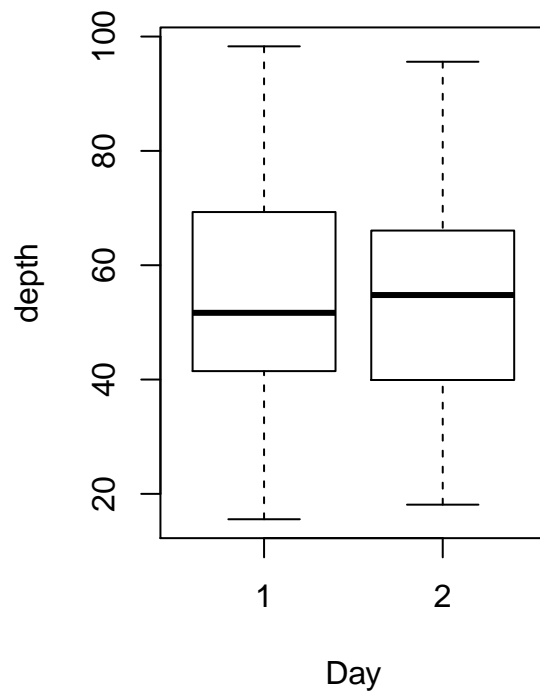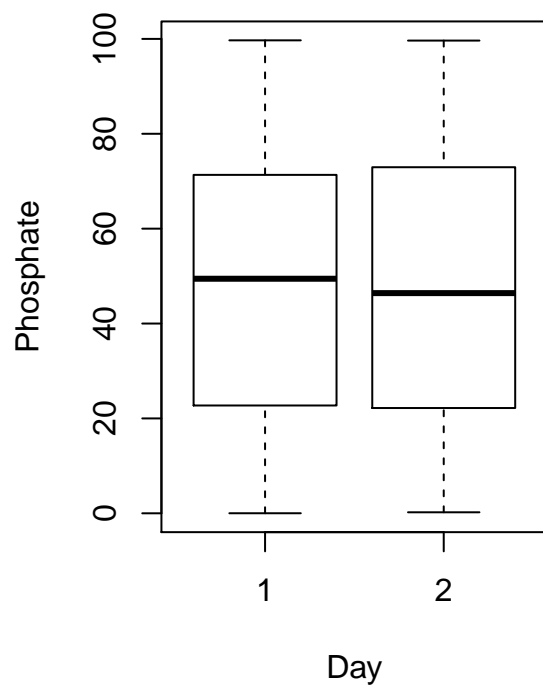
```r
table(HW4_data$day)
```

```
##
##    1    2
## 1136  710
```

```r
par(mfrow=c(1,2))
boxplot(phosphate~block,data = HW4_data, xlab="Block",ylab="Phosphate")
boxplot(depth~block,data = HW4_data, xlab="Block",ylab="depth")
```

```r
par(mfrow=c(1,2))
boxplot(phosphate~day,data = HW4_data, xlab="Day",ylab="Phosphate")
boxplot(depth~day,data = HW4_data, xlab="Day",ylab="depth")
```
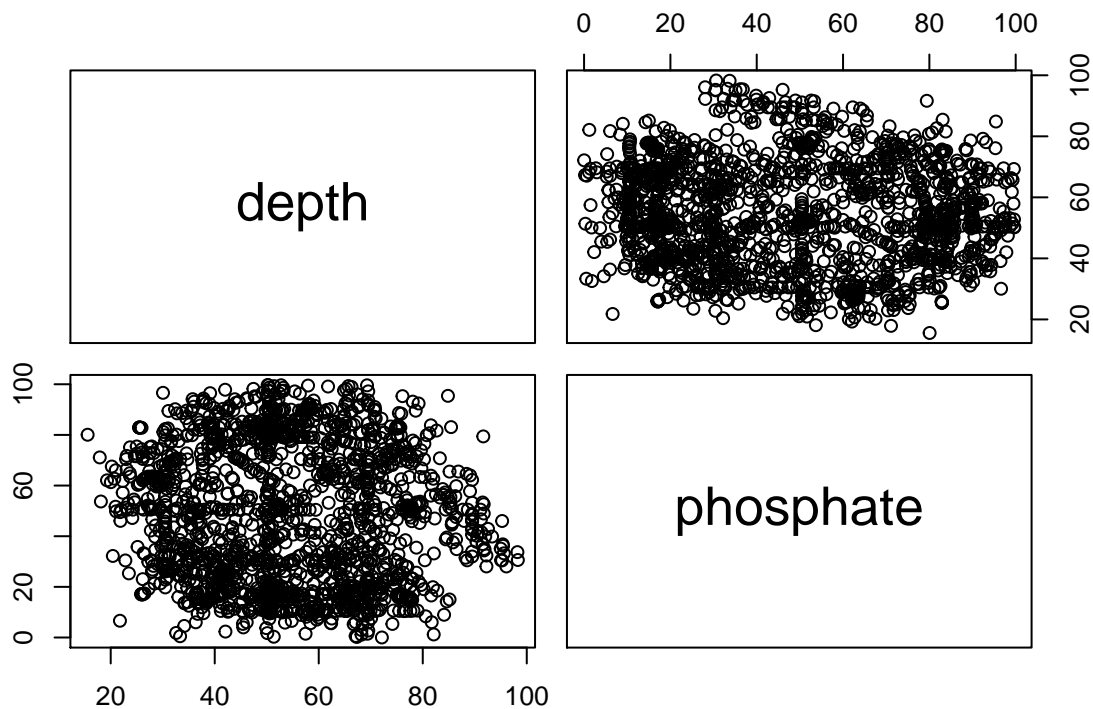
```
par(fig=c(0,0.8,0,0.8), new=TRUE)
plot(HW4_data$depth, HW4_data$phosphate, xlab="Depth",ylab="Phosphate")
par(fig=c(0,0.8,0.45,1), new=TRUE)
boxplot(HW4_data$depth, horizontal=TRUE, axes=FALSE)
par(fig=c(0.6,1,0,0.8),new=TRUE)
boxplot(HW4_data$phosphate, axes=FALSE)
mtext("Multipanel Plot", side=3, outer=TRUE, line=-3)
library(ggplot2)
```

Multipanel Plot
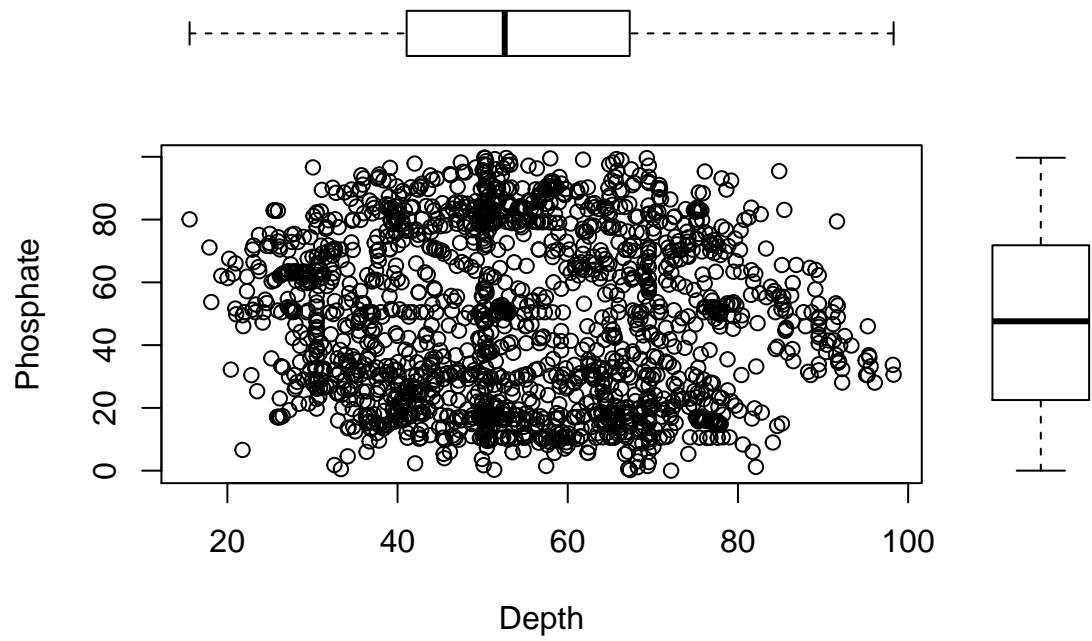
```
pairs(HW4_data[,c(2,3)])
```

```
cor(HW4_data$phosphate,HW4_data$depth)
```

```
## [1] -0.06601891
```

The factors that are present is the block variable and the day, there is 8 different levels of the block variable with 142 replications in each level. The side by side boxplots of the block grouping do not show much variation across the levels of the blocks. There are two different levels of the day variable, day 1 and day 2. There are 1136 observations in day 1 and 710 observations in day 2. The side by side boxplots for the day variable again show little variation across the days for both depth and phosphate. The lesson that can be learned from this dataset is sometimes there are numbers in are dataset that should be treated as a category. These numbers should be treated as factors in R and should be handled carefully when it comes to analysis of the data.

Problem 5

```
scatterBarNorm <- function(x, dcol="blue", lhist=20, num.dnorm=5*lhist, ...){
    ## check input
    stopifnot(ncol(x)==2)
    ## set up layout and graphical parameters
    layMat <- matrix(c(2,0,1,3), ncol=2, byrow=TRUE)
    layout(layMat, widths=c(5/7, 2/7), heights=c(2/7, 5/7))
    ospc <- 0.5 # outer space
    pext <- 4 # par extension down and to the left
    bspc <- 1 # space between scatter plot and bar plots
    par. <- par(mar=c(pext, pext, bspc, bspc),
```

```r
                    oma=rep(ospc, 4)) # plot parameters
    ## scatter plot
    plot(x, xlim=range(x[,1]), ylim=range(x[,2]), pch=20, ...)
    ## 3) determine barplot and height parameter
    ## histogram (for barplot-ting the density)
    xhist <- hist(x[,1], plot=FALSE, breaks=seq(from=min(x[,1]),
            to=max(x[,1]),  length.out=lhist))
    yhist <- hist(x[,2], plot=FALSE, breaks=seq(from=min(x[,2]),
            to=max(x[,2]),  length.out=lhist)) # note: this uses probability=TRUE
    ## determine the plot range and all the things needed for the barplots and lines
    xx <- seq(min(x[,1]), max(x[,1]), length.out=num.dnorm) # evaluation points for the overlaid density
    xy <- dnorm(xx, mean=mean(x[,1]), sd=sd(x[,1])) # density points
    yx <- seq(min(x[,2]), max(x[,2]), length.out=num.dnorm)
    yy <- dnorm(yx, mean=mean(x[,2]), sd=sd(x[,2]))
    ## barplot and line for x (top)
    par(mar=c(0, pext, 0, 0))
    barplot(xhist$density, axes=FALSE, ylim=c(0, max(xhist$density, xy)),
            space=0, col = "grey") # barplot
    lines(seq(from=0, to=lhist-1, length.out=num.dnorm), xy, col=dcol) # line
    ## barplot and line for y (right)
    par(mar=c(pext, 0, 0, 0))
    barplot(yhist$density, axes=FALSE, xlim=c(0, max(yhist$density, yy)),
            space=0, horiz=TRUE, col = "orange") # barplot
    lines(yy, seq(from=0, to=lhist-1, length.out=num.dnorm), col=dcol) # line
    ## restore parameters
    par(par.)
}
scatterBarNorm(iris[,c(1,2)])
```