HW4_Yun_Young

Problem #3 The focus of the EDA stage of an analysis includes, according to Roger Peng, “identifying relationships between variables that are particularly interesting or unexpected, checking to see if there is any evidence for or against a stated hypothesis, checking for problems with the collected data, such as missing data or measurement error, or identifying certain areas where more data need to be collected”. However, this is not the final product or for final decision. At this stage, it is more important to examine the data quickly in order to make decisions about what to do next just like “editing room” of filmmaking.

Problem #4

```
## Loading required package: xlsxjars
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

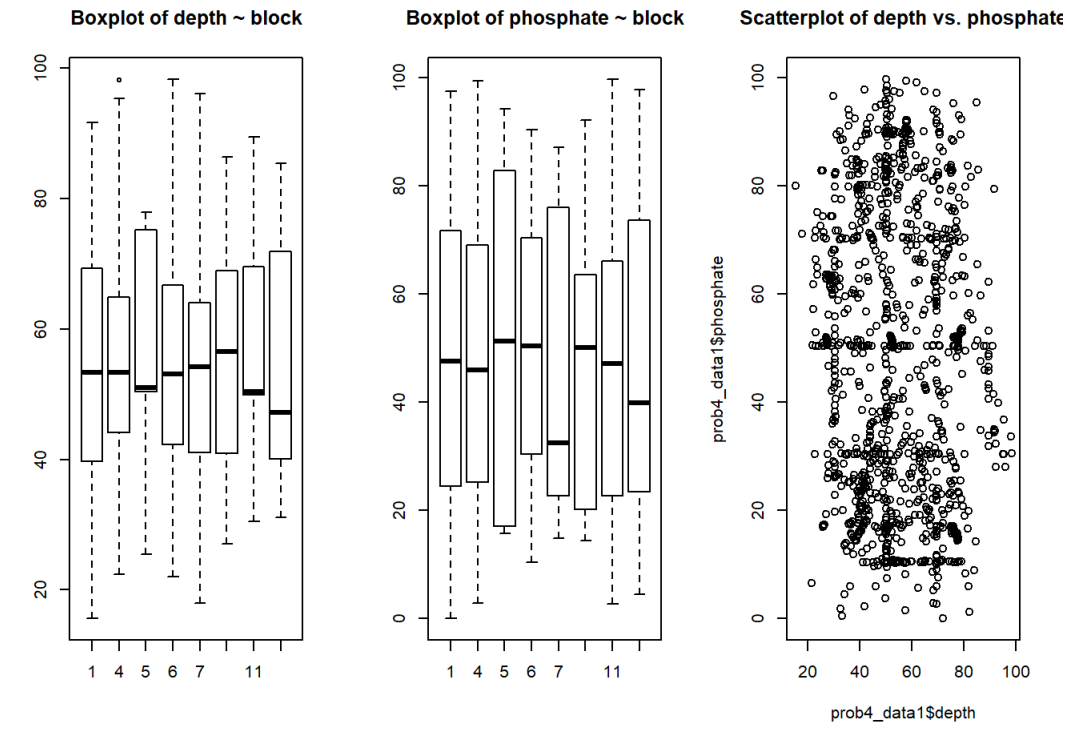
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
##
## Attaching package: 'gridExtra'
```

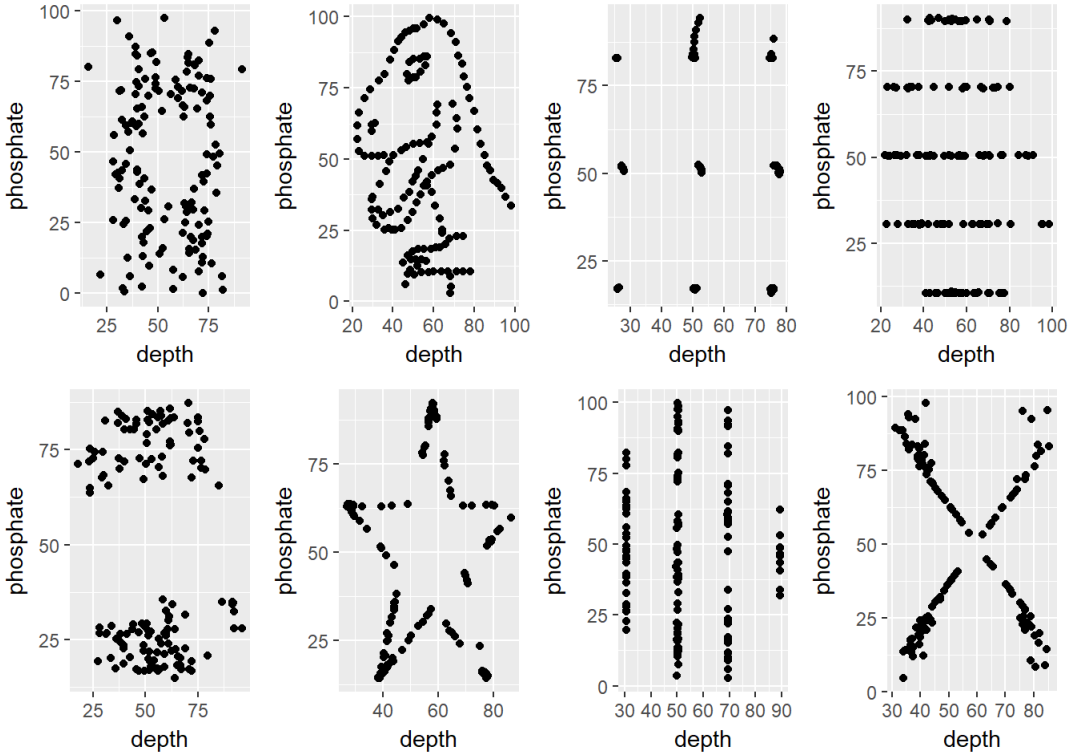
```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
##                Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## block       1.00000000  4.75000  6.50000  7.12500 10.25000 13.00000
## depth      15.56074952 41.47866 51.67204 54.26467 69.26945 98.28812
## phosphate  0.01511933 22.73069 49.44212 47.83609 71.30355 99.69468
```

```
## 'data.frame':    1136 obs. of  3 variables:
##  $ block    : num  4 4 4 4 4 4 4 4 4 4 ...
##  $ depth    : num  55.4 51.5 46.2 42.8 40.8 ...
##  $ phosphate: num  97.2 96 94.5 91.4 88.3 ...
```



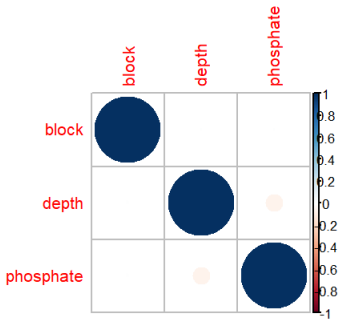

```
## [1] "following is the multipanel plot of depth vs. phosphate by blcok"
```




```
##                  block         depth     phosphate
## block     1.000000e+00  1.908909e-05  6.730804e-05
## depth     1.908909e-05  1.000000e+00 -6.464410e-02
## phosphate 6.730804e-05 -6.464410e-02  1.000000e+00
```

```
## [1] "Considering the summary statistics and the plots, we need to label columns correctly and descriptively in order to work with plots and summary st
atistics that read column names"
```

correlation plot by circle



Problem 5

```
##
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
## [1] "following is 3D scatterplot of prob_4data1 by block, depth, phosphate"
```