

# HW4\_Xu\_Peng

*Peng Xu*

*2017/9/24*

## Question 3

The EDA process provide the “rough cut” for a data analysis. According to Roger Peng, six principles should be summarized as followed. 1) show comparison from different aspects. 2) show causality, mechanism, explanation, and systematic structure. Different factors could be connected to explain one phenomena. 3) show multivariate data. It may uncover unexpected relationships with more features. 4) integrate different evidence to a graphic. 5) describe and document the evidence. A graphic should tell a complete story by itself. 6) present graphics with high-quality content

## Question 4

### Part 1

```
prob4_data1 <- read.xlsx("HW4_data.xlsx", sheetIndex = 1)
prob4_data2 <- read.xlsx("HW4_data.xlsx", sheetIndex = 2)
prob4_RawData <- rbind(prob4_data1,prob4_data2)
summary(prob4_RawData)
```

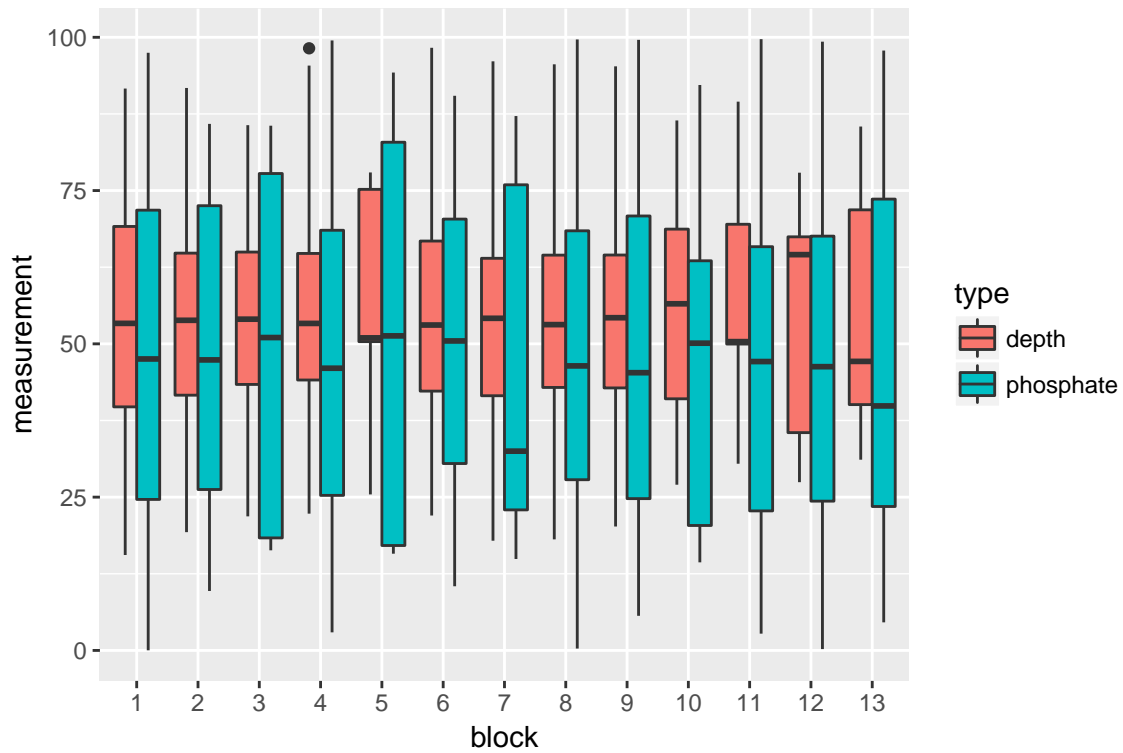
##	block	depth	phosphate
##	Min. : 1	Min. :15.56	Min. : 0.01512
##	1st Qu.: 4	1st Qu.:41.07	1st Qu.:22.56107
##	Median : 7	Median :52.59	Median :47.59445
##	Mean : 7	Mean :54.27	Mean :47.83510
##	3rd Qu.:10	3rd Qu.:67.28	3rd Qu.:71.81078
##	Max. :13	Max. :98.29	Max. :99.69468

### Part 2

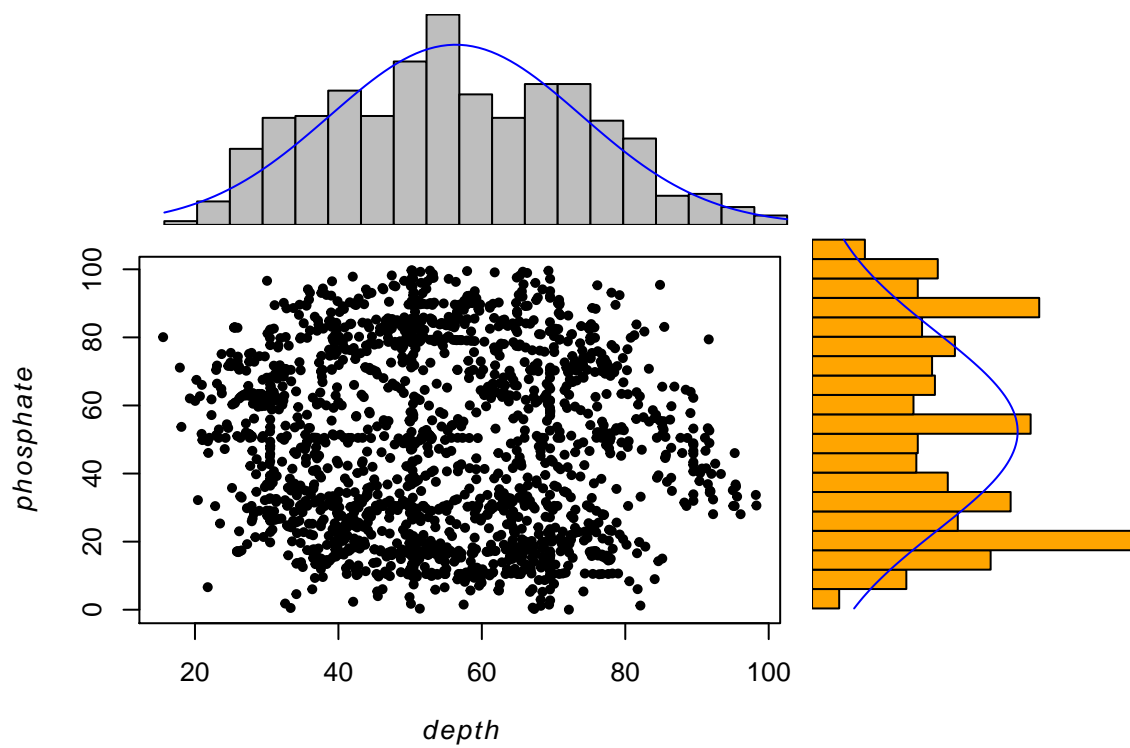
There are three factors: block, depth, phosphate. The block number seems to be the number of different experiment groups, while depth and phosphate are different measurements of each experiment.

### Part 3

```
CleanData <- gather(prob4_RawData,type,value = measurement, depth,phosphate)
CleanData_n <- transform(CleanData, block = as.factor(block))
CleanData_n <- transform(CleanData_n, type = as.factor(type))
ggplot(CleanData_n, aes(x=block, y=measurement, fill=type))+geom_boxplot()
```

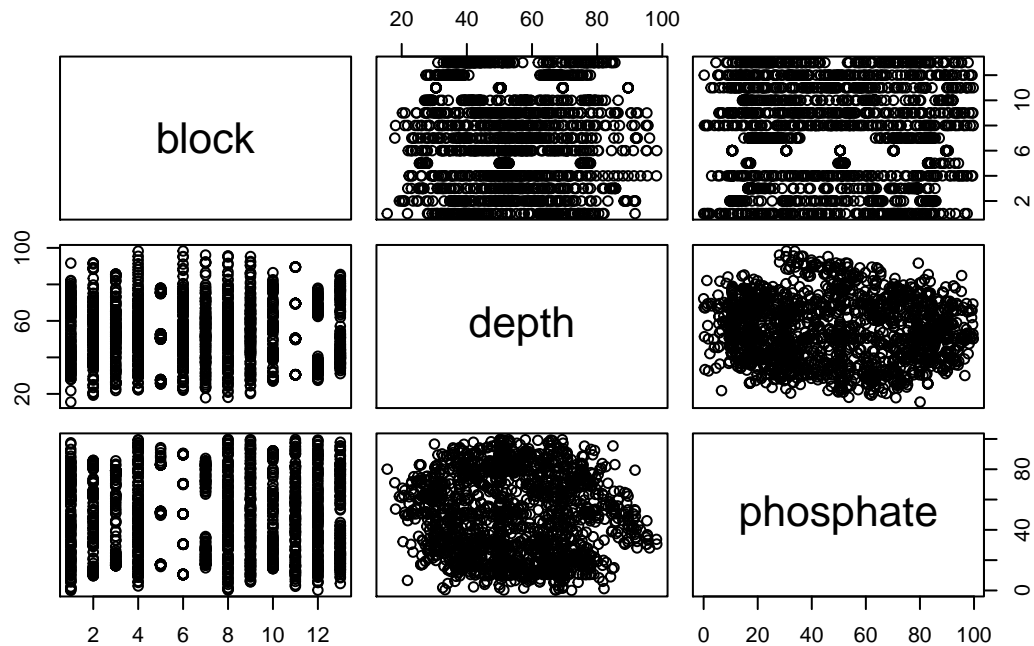


```
Target <- data.frame(cbind(prob4_RawData$depth ,prob4_RawData$phosphate))
scatterBarNorm(Target, xlab=expression(italic(depth)),
                ylab=expression(italic(phosphate)))
```



## Part 4

```
pairs(prob4_RawData)
```



## Part 5

From the summary plot, the block seems to be index number of groups, while depth and phosphate are measurement values. The distributions of depth and phosphate are very similar and almost follow the normal distribution.