STAT 5014 Homework 4

Samantha Sunshine

9/27/17

Problem 3

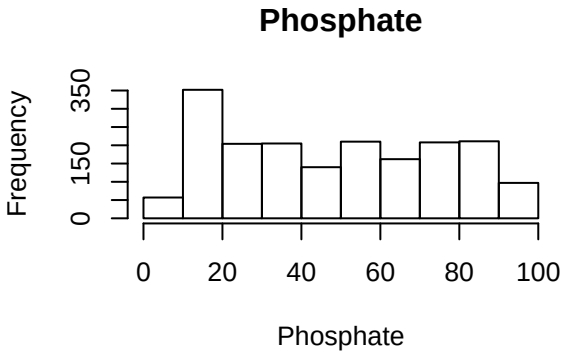
Roger Peng says the focus of exploratory data analysis is to identify relationships between variables, test hypotheses, and to identify and solve problems occurring with data. Another goal would be to summarize data and look for patterns.

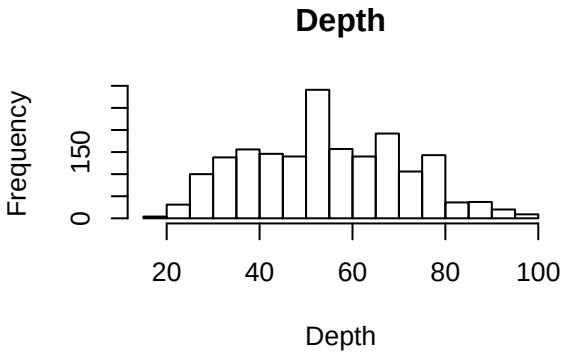
Problem 4

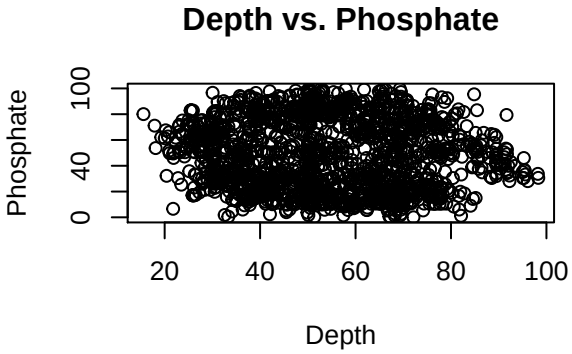
Table 1: Data Summary

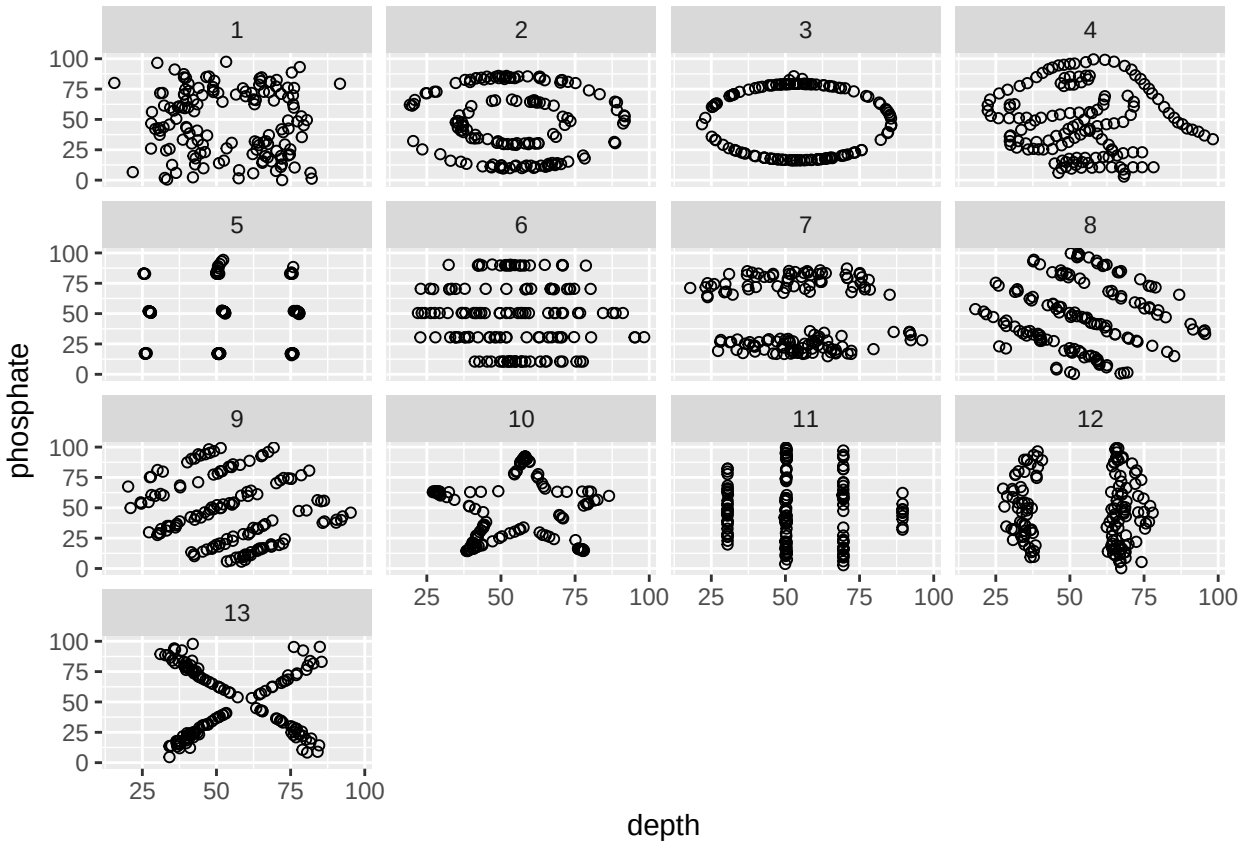
block	depth	phosphate
Min. : 1	Min. :15.56	Min. : 0.01512
1st Qu.: 4	1st Qu.:41.07	1st Qu.:22.56107
Median : 7	Median :52.59	Median :47.59445
Mean : 7	Mean :54.27	Mean :47.83510
3rd Qu.:10	3rd Qu.:67.28	3rd Qu.:71.81078
Max. :13	Max. :98.29	Max. :99.69468

The columns in the dataset, block, depth, and phosphate, are all numeric variables, and it is a 1846x3 matrix.

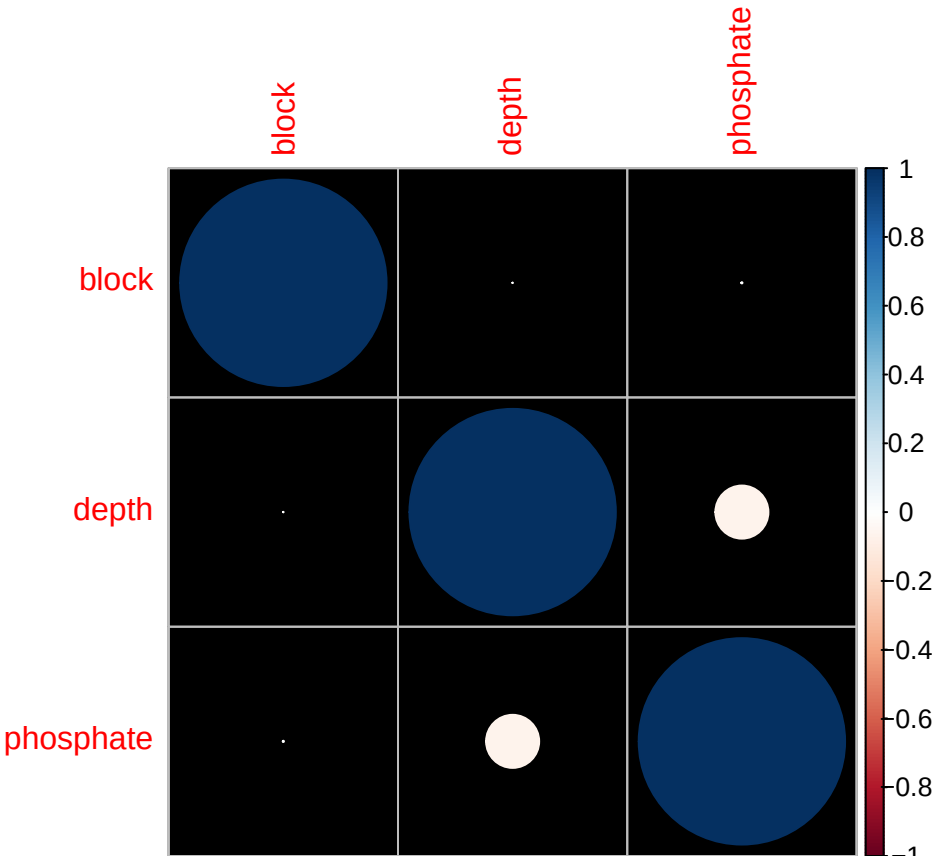

Phosphate
Frequency
0
150
350
0
20
40
60
80
100
Phosphate


Depth
Frequency
0
150
20
40
60
80
100
Depth


Depth vs. Phosphate
Phosphate
0
40
100
20
40
60
80
100
Depth



	block	depth	phosphate
block	1.000000e+00	-5.620472e-06	3.202565e-05
depth	-5.620472e-06	1.000000e+00	-6.601891e-02
phosphate	3.202565e-05	-6.601891e-02	1.000000e+00


block
depth
phosphate
block
depth
phosphate
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8