HW4

Matthew Blevins

September 21, 2017

```
## Warning: package 'dplyr' was built under R version 3.3.3
```

```
##
## 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
```

```
## Warning: package 'tidyr' was built under R version 3.3.3
```

Problem 1

I successfully completed the swirl lessons.

Problem 2

I successfully made this file and put it in the correct location if you are reading this.

Problem 3

Roger Peng provides many reasons for examples to demonstrate the importance of exploratory data analysis. He states that it helps researchers identify the important pieces of information that might be worth further investigating, and he compares the process to film editors making the first cuts of their films. A few of the goals he describes are to examine the validity of the stated hypothesis, determine any issues with the data gathered, assess whether the amount of collected data needs to be increased, and to determine relationships between variables that may be intriguing or surprising.

Problem 4

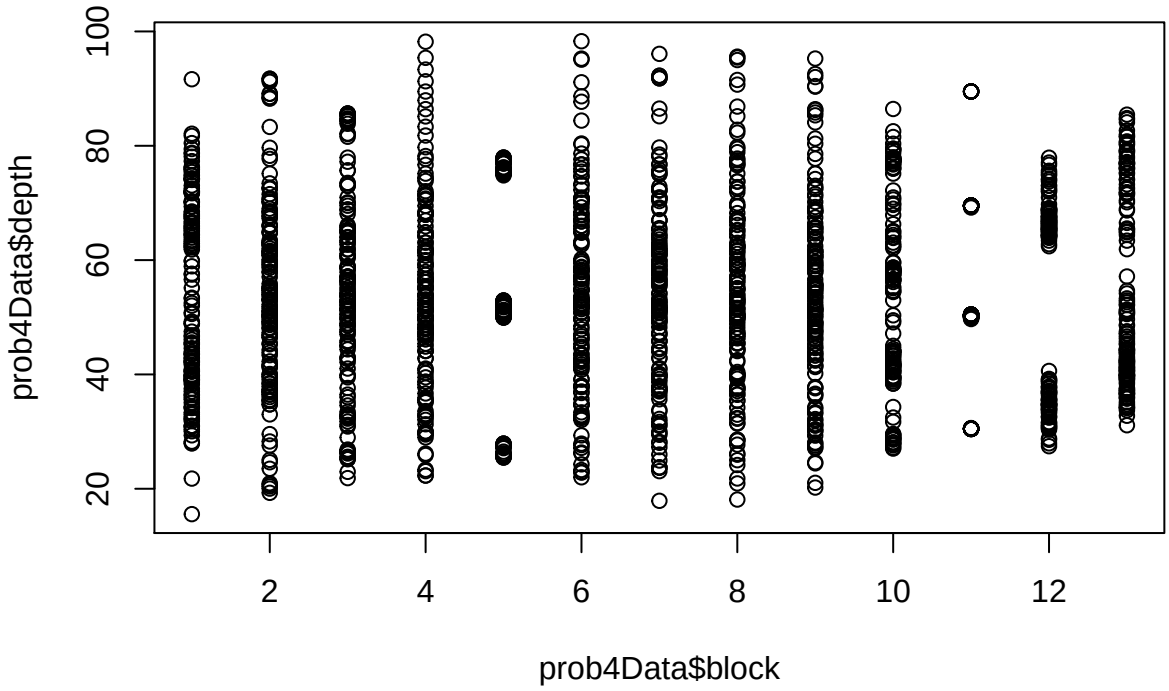
```
## Warning: package 'readxl' was built under R version 3.3.3
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## Warning: package 'lattice' was built under R version 3.3.3
```

```
## Warning: package 'hexbin' was built under R version 3.3.3
## Warning: package 'bindrcpp' was built under R version 3.3.3
```



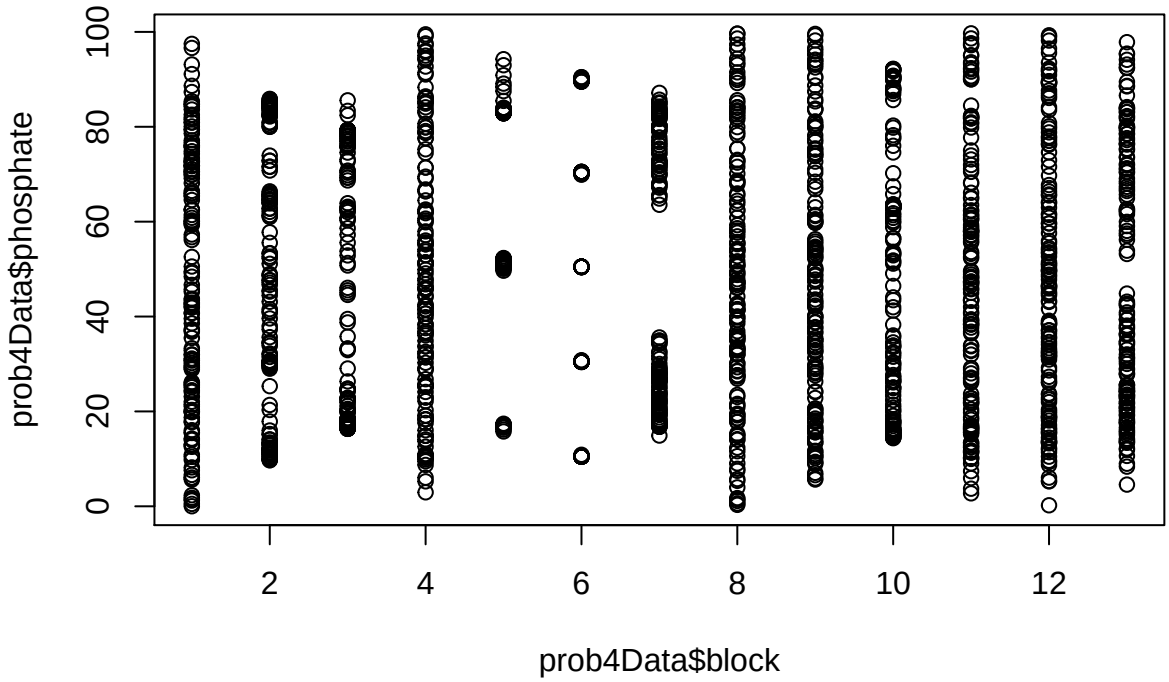


Table 1: Summary Statistics

Block	Depth	Value	Phosphate	Value
1	depth	Min. :15.56	phosphate	Min. : 0.01512
1	depth	1st Qu.:39.72	phosphate	1st Qu.:24.62589
1	depth	Median :53.34	phosphate	Median :47.53527
1	depth	Mean :54.27	phosphate	Mean :47.83472
1	depth	3rd Qu.:69.15	phosphate	3rd Qu.:71.80315
1	depth	Max. :91.64	phosphate	Max. :97.47577
2	depth	Min. :19.29	phosphate	Min. : 9.692
2	depth	1st Qu.:41.63	phosphate	1st Qu.:26.245
2	depth	Median :53.84	phosphate	Median :47.383
2	depth	Mean :54.27	phosphate	Mean :47.831
2	depth	3rd Qu.:64.80	phosphate	3rd Qu.:72.533
2	depth	Max. :91.74	phosphate	Max. :85.876
3	depth	Min. :21.86	phosphate	Min. :16.33
3	depth	1st Qu.:43.38	phosphate	1st Qu.:18.35
3	depth	Median :54.02	phosphate	Median :51.03
3	depth	Mean :54.27	phosphate	Mean :47.84
3	depth	3rd Qu.:64.97	phosphate	3rd Qu.:77.78
3	depth	Max. :85.66	phosphate	Max. :85.58
4	depth	Min. :22.31	phosphate	Min. : 2.949
4	depth	1st Qu.:44.10	phosphate	1st Qu.:25.288
4	depth	Median :53.33	phosphate	Median :46.026
4	depth	Mean :54.26	phosphate	Mean :47.832
4	depth	3rd Qu.:64.74	phosphate	3rd Qu.:68.526

Block	Depth	Value	Phosphate	Value
4	depth	Max. :98.21	phosphate	Max. :99.487
5	depth	Min. :25.44	phosphate	Min. :15.77
5	depth	1st Qu.:50.36	phosphate	1st Qu.:17.11
5	depth	Median :50.98	phosphate	Median :51.30
5	depth	Mean :54.26	phosphate	Mean :47.84
5	depth	3rd Qu.:75.20	phosphate	3rd Qu.:82.88
5	depth	Max. :77.95	phosphate	Max. :94.25
6	depth	Min. :22.00	phosphate	Min. :10.46
6	depth	1st Qu.:42.29	phosphate	1st Qu.:30.48
6	depth	Median :53.07	phosphate	Median :50.47
6	depth	Mean :54.26	phosphate	Mean :47.83
6	depth	3rd Qu.:66.77	phosphate	3rd Qu.:70.35
6	depth	Max. :98.29	phosphate	Max. :90.46
7	depth	Min. :17.89	phosphate	Min. :14.91
7	depth	1st Qu.:41.54	phosphate	1st Qu.:22.92
7	depth	Median :54.17	phosphate	Median :32.50
7	depth	Mean :54.27	phosphate	Mean :47.84
7	depth	3rd Qu.:63.95	phosphate	3rd Qu.:75.94
7	depth	Max. :96.08	phosphate	Max. :87.15
8	depth	Min. :18.11	phosphate	Min. : 0.3039
8	depth	1st Qu.:42.89	phosphate	1st Qu.:27.8409
8	depth	Median :53.14	phosphate	Median :46.4013
8	depth	Mean :54.27	phosphate	Mean :47.8359
8	depth	3rd Qu.:64.47	phosphate	3rd Qu.:68.4394
8	depth	Max. :95.59	phosphate	Max. :99.6442
9	depth	Min. :20.21	phosphate	Min. : 5.646
9	depth	1st Qu.:42.81	phosphate	1st Qu.:24.756
9	depth	Median :54.26	phosphate	Median :45.292
9	depth	Mean :54.27	phosphate	Mean :47.831
9	depth	3rd Qu.:64.49	phosphate	3rd Qu.:70.856
9	depth	Max. :95.26	phosphate	Max. :99.580
10	depth	Min. :27.02	phosphate	Min. :14.37
10	depth	1st Qu.:41.03	phosphate	1st Qu.:20.37
10	depth	Median :56.53	phosphate	Median :50.11
10	depth	Mean :54.27	phosphate	Mean :47.84
10	depth	3rd Qu.:68.71	phosphate	3rd Qu.:63.55
10	depth	Max. :86.44	phosphate	Max. :92.21
11	depth	Min. :30.45	phosphate	Min. : 2.735
11	depth	1st Qu.:49.96	phosphate	1st Qu.:22.753
11	depth	Median :50.36	phosphate	Median :47.114
11	depth	Mean :54.27	phosphate	Mean :47.837
11	depth	3rd Qu.:69.50	phosphate	3rd Qu.:65.845
11	depth	Max. :89.50	phosphate	Max. :99.695
12	depth	Min. :27.44	phosphate	Min. : 0.217
12	depth	1st Qu.:35.52	phosphate	1st Qu.:24.347
12	depth	Median :64.55	phosphate	Median :46.279
12	depth	Mean :54.27	phosphate	Mean :47.832
12	depth	3rd Qu.:67.45	phosphate	3rd Qu.:67.568
12	depth	Max. :77.92	phosphate	Max. :99.284
13	depth	Min. :31.11	phosphate	Min. : 4.578
13	depth	1st Qu.:40.09	phosphate	1st Qu.:23.471
13	depth	Median :47.14	phosphate	Median :39.876

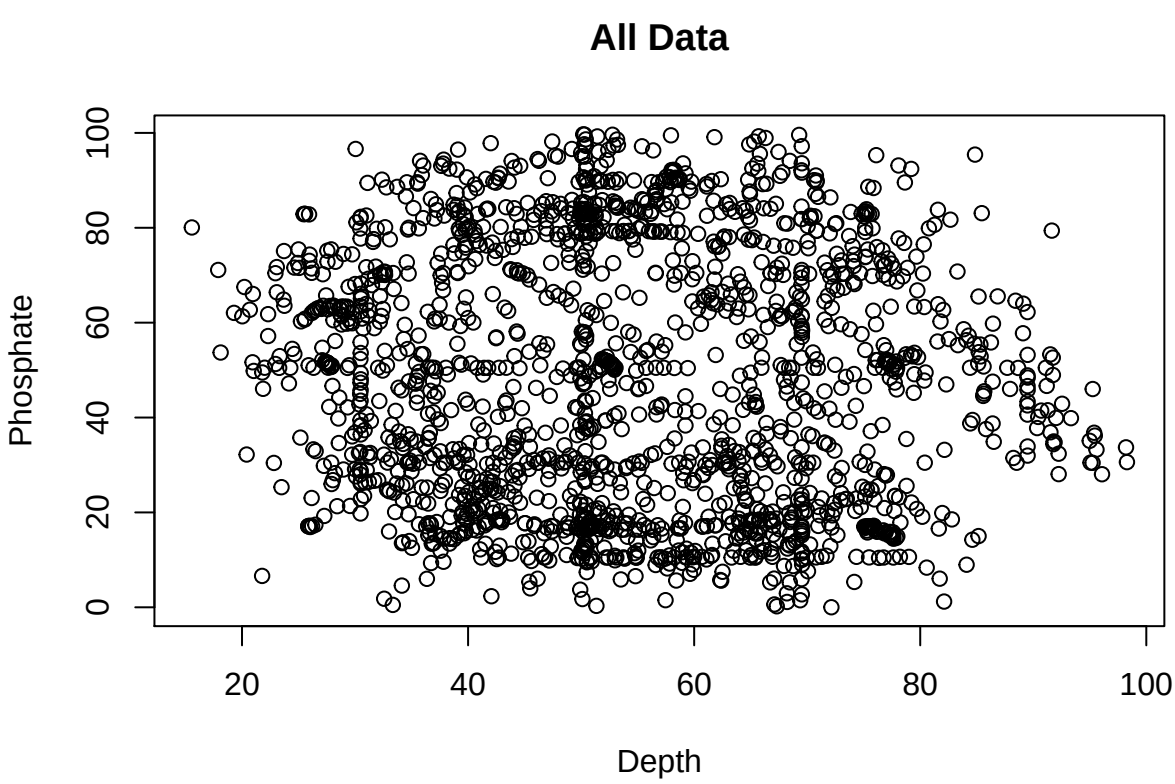
Block	Depth	Value	Phosphate	Value
13	depth	Mean :54.26	phosphate	Mean :47.840
13	depth	3rd Qu.:71.86	phosphate	3rd Qu.:73.610
13	depth	Max. :85.45	phosphate	Max. :97.838

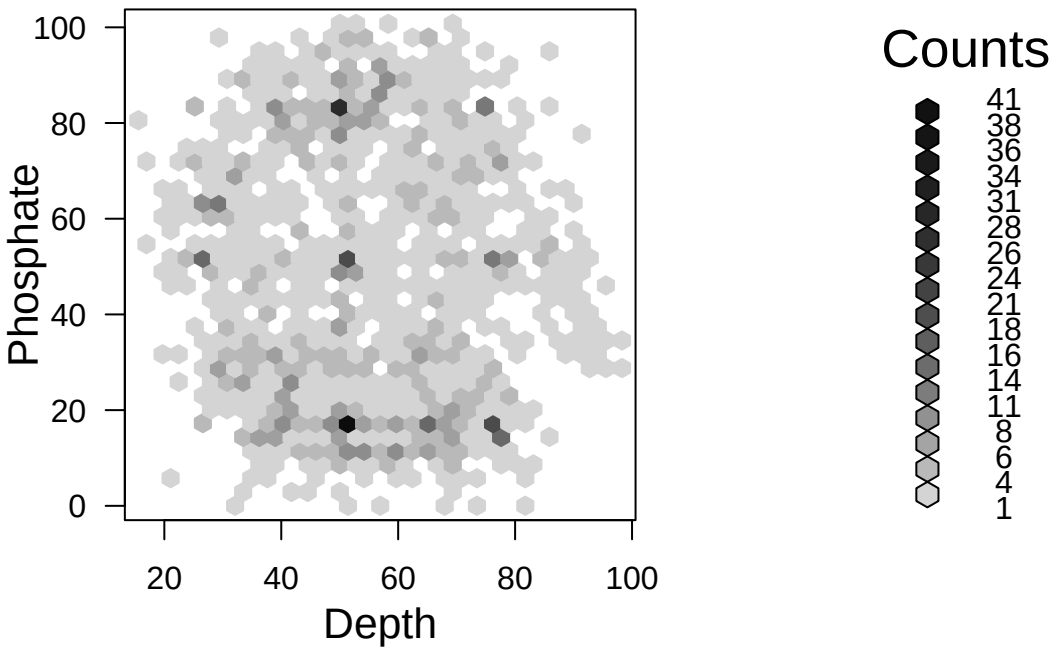
```
## Warning in data.frame(..., check.names = FALSE): row names were found from
## a short variable and have been discarded
```

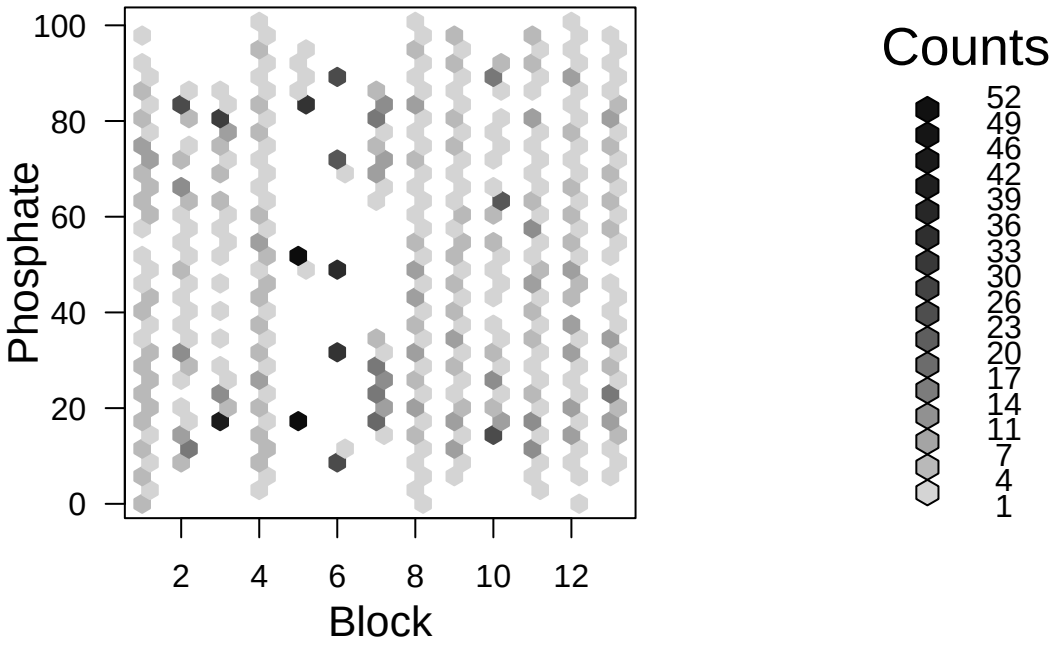
Table 2: Correlation

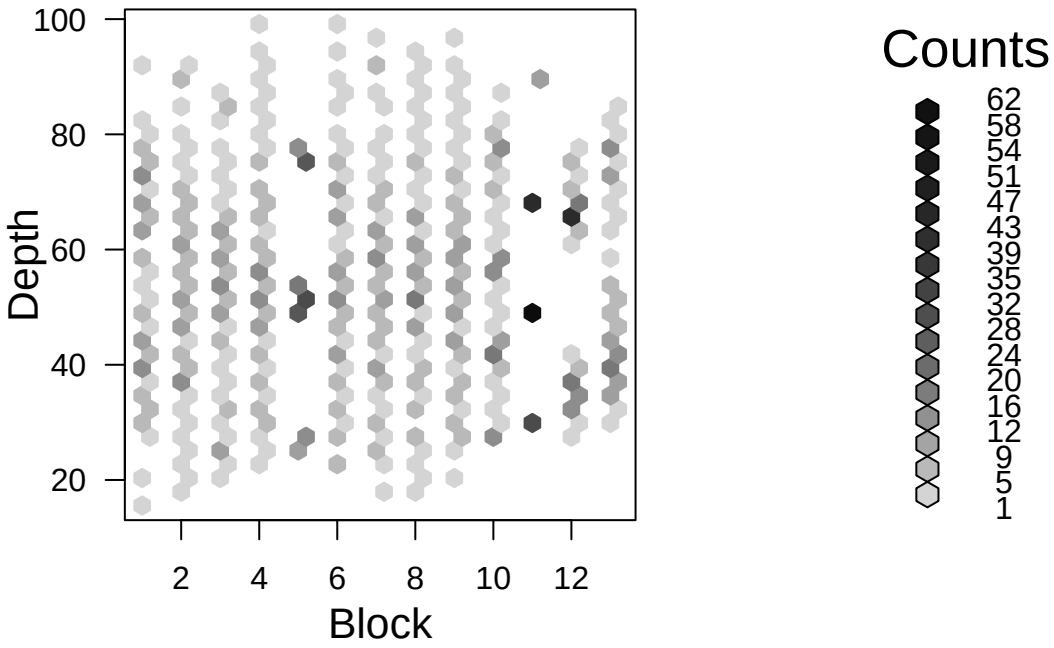
Block	Correlation
1	-0.0641284
1	-0.0685864
1	-0.0683434
1	-0.0644719
1	-0.0603414
1	-0.0617148
2	-0.0685042
2	-0.0689797
2	-0.0686092
2	-0.0629611
2	-0.0694456
2	-0.0665752
3	-0.0655833
3	-0.0641284
3	-0.0685864
3	-0.0683434
3	-0.0644719
3	-0.0603414
4	-0.0617148
4	-0.0685042
4	-0.0689797
4	-0.0686092
4	-0.0629611
4	-0.0694456
5	-0.0665752
5	-0.0655833
5	-0.0641284
5	-0.0685864
5	-0.0683434
5	-0.0644719
6	-0.0603414
6	-0.0617148
6	-0.0685042
6	-0.0689797
6	-0.0686092
6	-0.0629611
7	-0.0694456
7	-0.0665752
7	-0.0655833
7	-0.0641284
7	-0.0685864
7	-0.0683434

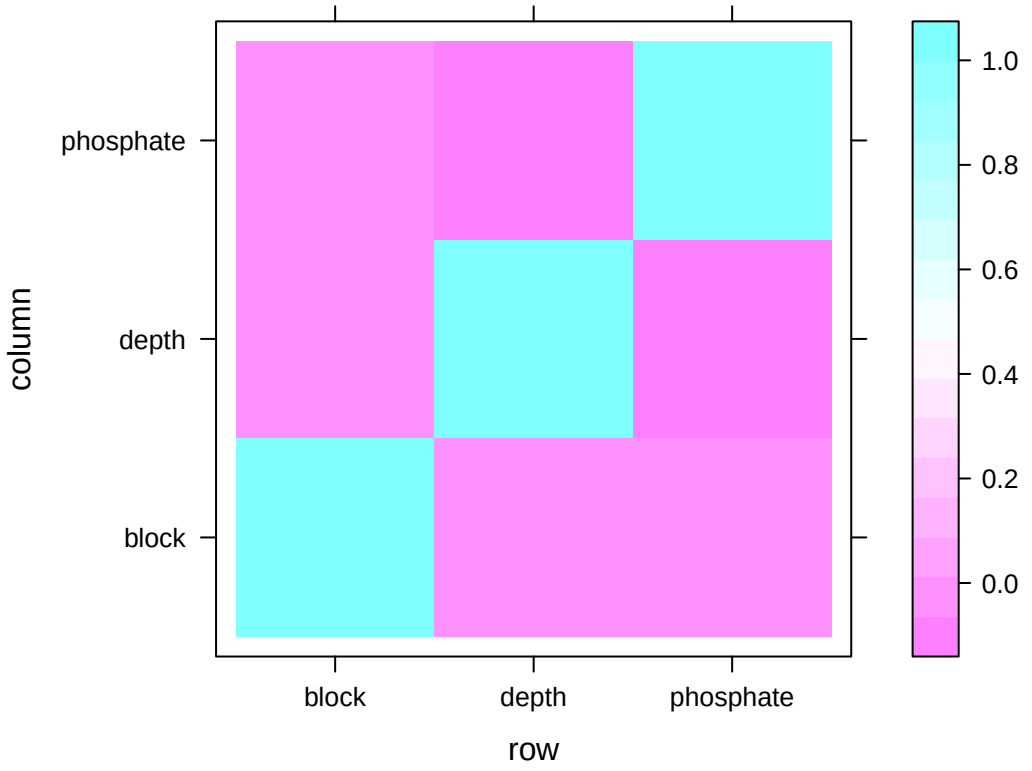
Block	Correlation
8	-0.0644719
8	-0.0603414
8	-0.0617148
8	-0.0685042
8	-0.0689797
8	-0.0686092
9	-0.0629611
9	-0.0694456
9	-0.0665752
9	-0.0655833
9	-0.0641284
9	-0.0685864
10	-0.0683434
10	-0.0644719
10	-0.0603414
10	-0.0617148
10	-0.0685042
10	-0.0689797
11	-0.0686092
11	-0.0629611
11	-0.0694456
11	-0.0665752
11	-0.0655833
11	-0.0641284
12	-0.0685864
12	-0.0683434
12	-0.0644719
12	-0.0603414
12	-0.0617148
12	-0.0685042
13	-0.0689797
13	-0.0686092
13	-0.0629611
13	-0.0694456
13	-0.0665752
13	-0.0655833


All Data
Phosphate
0
20
40
60
80
100
20
40
60
80
100
Depth


Phosphate
Depth
0
20
40
60
80
100
20
40
60
80
100
Counts
41
38
36
34
31
28
26
24
21
18
16
14
11
8
6
4
1


Phosphate
100
80
60
40
20
0
2
4
6
8
10
12
Block
Counts
52
49
46
42
39
36
33
30
26
23
20
17
14
11
7
4
1


Depth
100
80
60
40
20
2
4
6
8
10
12
Block
Counts
62
58
54
51
47
43
39
35
32
28
24
20
16
12
9
5
1



Based on the summary statistics and the plots, this data does not appear to have any significant correlation, and the values of the points vary widely. Utilizing the hexbin plots does show that some points have significant overlap with repeated values. If I were doing research on this data, I would probably start by examining why some of the points in a few specific blocks had so much overlap while the rest of the blocks had their data spread over a large area.