HW5_Porter_Erica

Erica Porter

9/30/2017

Problem 3

In my opinion, a good figure should display the desired insights in a clear manner, while also acknowledging data limitations and discrepancies; a figure/graphic should not lie or mislead audiences to believe something is in the data if it's not. A figure should display all relevant and necessary comparisons without becoming so cumbersome that it can no longer be interpreted or viewed easily at one time. The comparison aspect is key, of course; whether the audience is highly familiar with the background/study or not, it is important to either include multiple groups or references for comparison, or to provide some sort of scale.

However, the guidelines for a good figure may change significantly for different audiences and data; a statistician who is familiar with a large number of graphs, charts, and maps would expect a different level of detail and clarity than a business leader wishing to make quick decisions from key progress indicators. I also prefer figures with a (limited) number of colors, but many companies and individuals have brand guidelines and expectations that may restrict the plot packages and colors possible for use (and as Dr. Franck pointed out, some individuals are color-blind or color-confused, which definitely changes the circumstances).

Figures of course need to use clear fonts, consistent labels, legends where appropriate, consistent and appropriate scales, and proper dimensions for displaying data. While there are some very interesting and cool graphics, I think the graphic should always be chosen according to the scope, potential, and itent of the data. A 3D line graph or layered bar graph is only interesting if it still conveys/summarizes the information as quickly and easily as possible.

Problem 4

This function will vary significantly depending upon the contents of the vector and the criteria for success. For example, if you are flipping a coin and obtaining Heads on a flip is considered a success, the criteria for success is certainly different than when determining the number of passing grades in a vector of exam grades. For the purpose of this problem, I set a success to be a 1, so this function assumes that the user inputs a vector of 0's and 1's.

```
##  [1] 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6 0.6
```

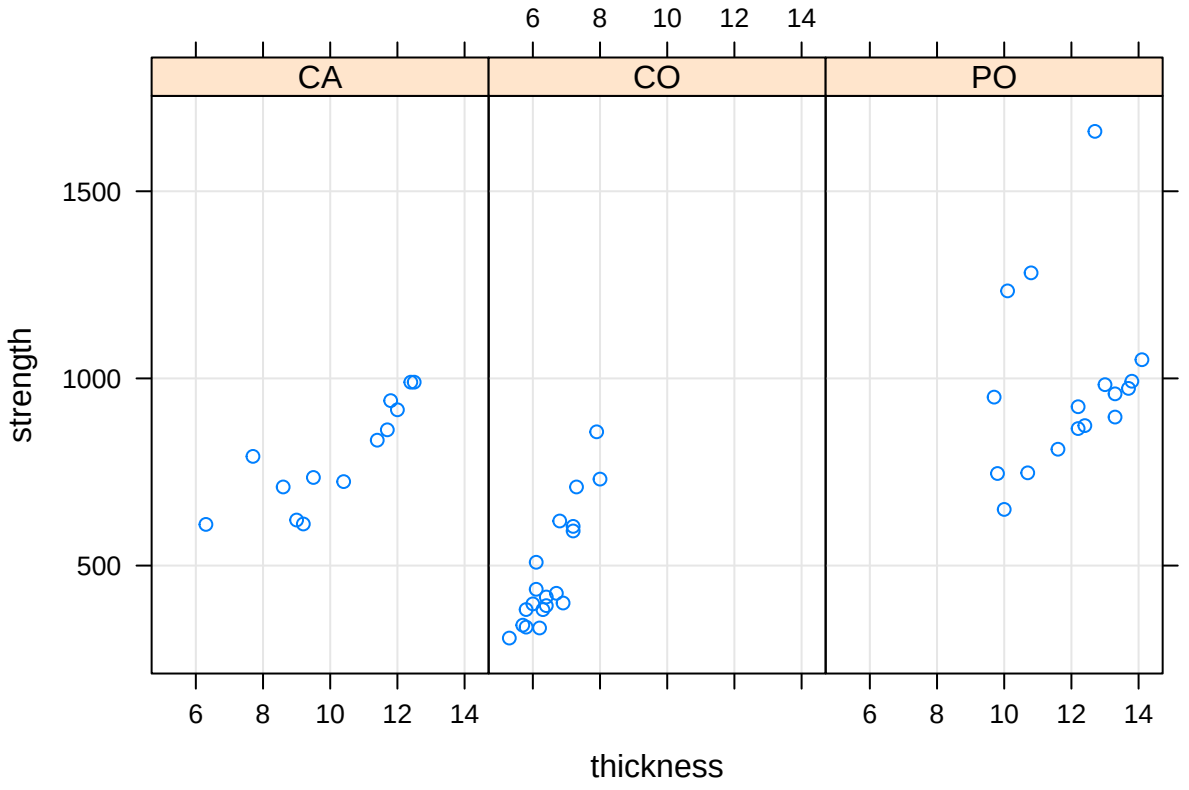
```
##  [1] 1 1 1 1 0 0 0 0 1 1
```

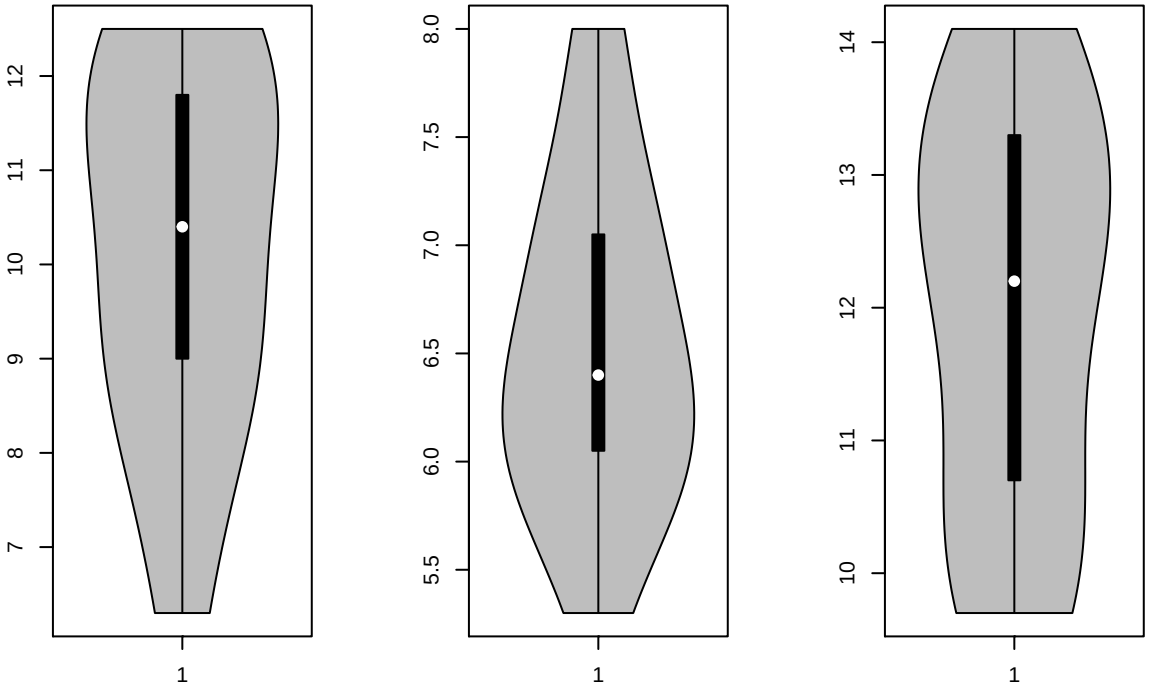
The problem with the matrix P4b_data is that every trial (column) is identical, so it seems that the `rbinom` did not actually use a different probability for each experiment.

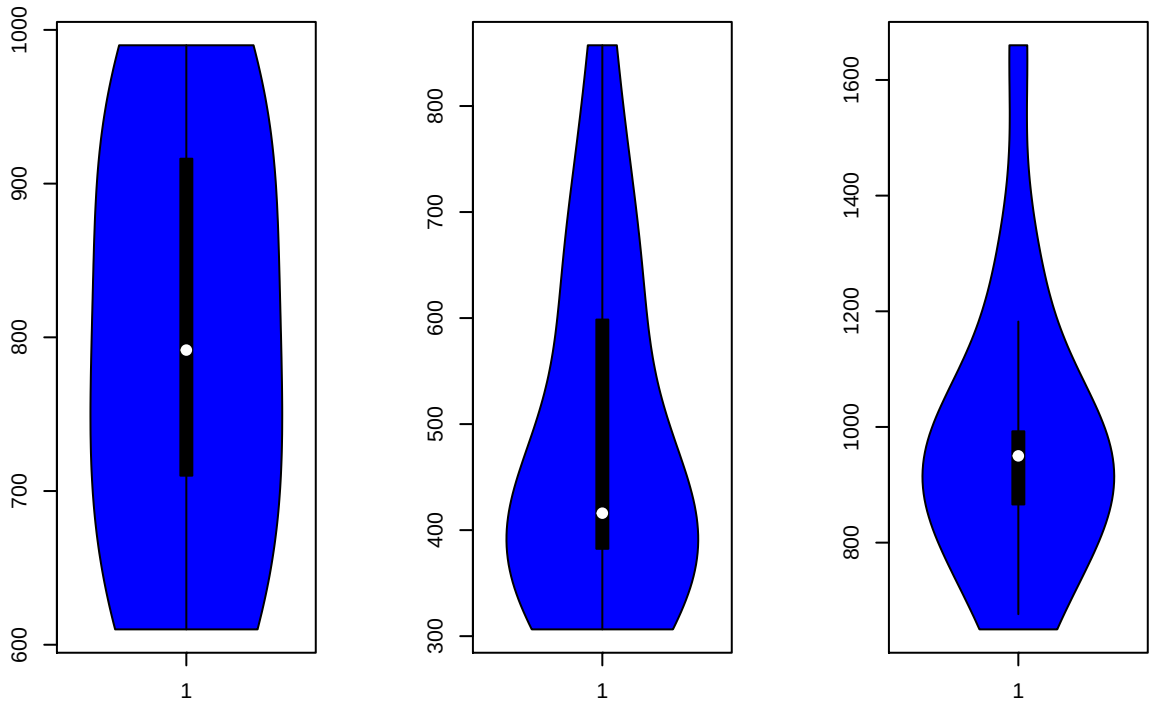
Problem 5

Table 1: Summary by Starch

	Mean Strength	Mean Thickness	SD Strength	SD Thickness	Correlation
CA	795.2923	10.192308	139.0485	1.9708127	0.8403774
CO	482.8263	6.531579	157.5958	0.7409098	0.8881163
PO	976.4294	11.964706	237.7956	1.5136633	0.2069389


CA
CO
PO
strength
thickness
500
1000
1500
6
8
10
12
14


7
8
9
10
11
12
5.5
6.0
6.5
7.0
7.5
8.0
10
11
12
13
14
1
1
1



There appears to be a generally positive correlation between strength and thickness for each type of phosphate; however, each type of phosphate has a fairly different mean, standard deviation, and distribution. “PO” tends to have higher thickness and strength, while “CO” had lower values for each, and the values are more tightly clustered for “CO”. #Problem 6

```
## [1] 1239
```

```
##
##   AK   AL   AR   AZ   CA   CO   CT   DC   DE   FL   GA   HI   IA   ID   IL
##  273  838  709  532 2651  659  438  284   98 1487  972  139 1060  325 1587
##   IN   KS   KY   LA   MA   MD   ME   MI   MN   MO   MS   MT   NC   ND   NE
##  989  756  961  725  703  619  489 1170 1031 1170  533  405 1090  407  620
##   NH   NJ   NM   NV   NY   OH   OK   OR   PA   PR   RI   SC   SD   TN   TX
##  284  733  426  253 2207 1446  774  484 2208  176   91  539  394  795 2650
##   UT   VA   VT   WA   WI   WV   WY
##  344 1238  309  732  898  859  195
```

Table 2: Count of unique city names

state_vec	city_table
NY	2207
PR	176
MA	703
RI	91
NH	284
ME	489
VT	309

state_vec	city_table
CT	438
NJ	733
PA	2208
DE	98
DC	284
VA	1238
MD	619
WV	859
NC	1090
SC	539
GA	972
FL	1487
AL	838
TN	795
MS	533
KY	961
OH	1446
IN	989
MI	1170
IA	1060
WI	898
MN	1031
SD	394
ND	407
MT	405
IL	1587
MO	1170
KS	756
NE	620
LA	725
AR	709
OK	774
TX	2650
CO	659
WY	195
ID	325
UT	344
AZ	532
NM	426
NV	253
CA	2651
HI	139
OR	484
WA	732
AK	273