Stat 511 HW 1

Max Barnes

9/3/2020

I'm still working on figuring out some of the intricacies of R, so please bear with me while I perfect my formatting skills using R Markdown.

1-2

It is difficult to prove sex discrimination mainly due to the difficulty associated with gaining a truly random sample of the population. There are so many factors that effect how people are hired, that even if discrimination does occur, it would be hard to determin so.

1-6

(A) No, a genuine causal relationship cannot be established from from this study. This is because causal relationships can only be established for randomized relationships, of which this study is not.

(B) This study is observational in nature, and could be generalized toward the population of 14-16 year olds, however the inferences made are purely observational.

(C) Potential confounding factors about this study are that the Marijuana gropu was not randomly taken from a random population. The group was drawn from a group of people already enrolled in a drug abuse program (not random by any means). Due to the observational nature of the study, any findings are purely speculative and have little to no statistical significance. This study really only applies to the participants in the study.

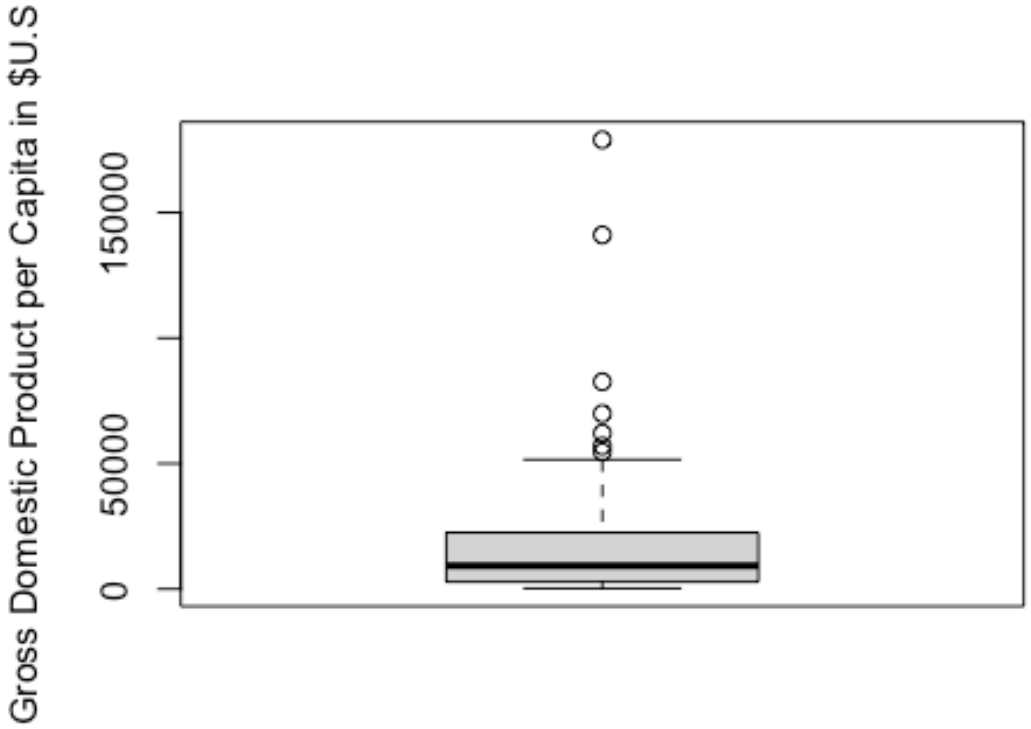
1-11

It can be said that the difference in treatments likely had an impact on the differences in cold rates, however, the study should be carried out again with a tasteless Vitamin C. The participants who correctly guessed that they were not taking the placebo could have used this knowledge to behave differently than they would have if they hadn't known which of the two pills they were taking.

1-12

This study is important given the lack of random selection for the same reasons that the study involving males with Chinese and European heritage was still important. For males with highblood pressure, it appears that fish oil dramatically reduces diastolic pressure levels, just like for men who were studied to understand the effects of a blood pressure reducing drug. The study provides important information to doctors who have patients with highblood pressure.

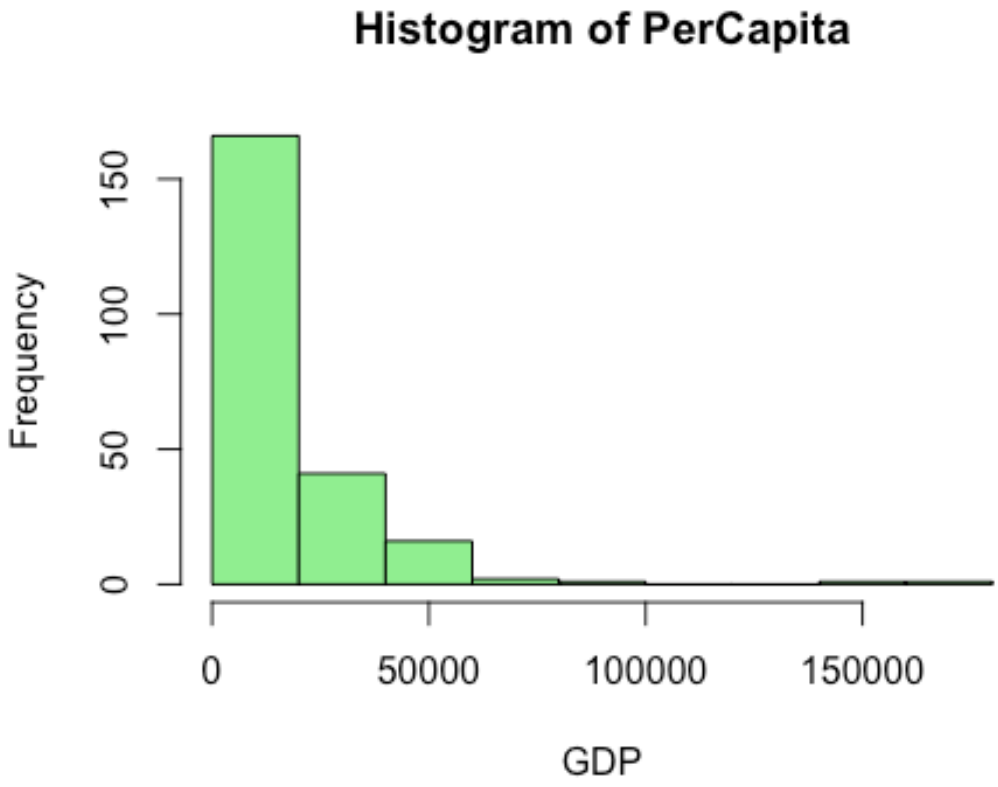
1-16

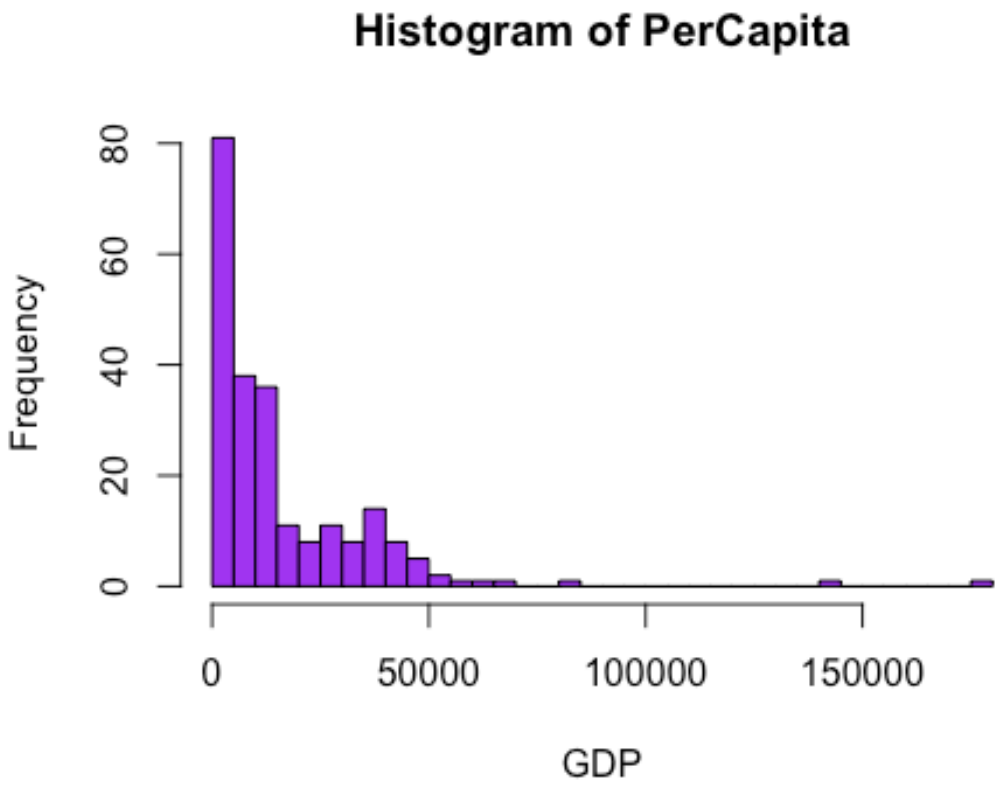


(A)

(B) My display generates and empty box plot for “Rank” and “Country” unless I tell it not to, while Display 1.11 in the book does not. Otherwise, they look fairly similar otherwise.

(C) The bin width is 20,000.





(D)

1-18

```
set.seed(47)
case <-(case0101)
number <-rnorm(47)
df <-data.frame(number)
data <- cbind(df, case)
data1 <- data[with(data, order(number)),]
data1
```

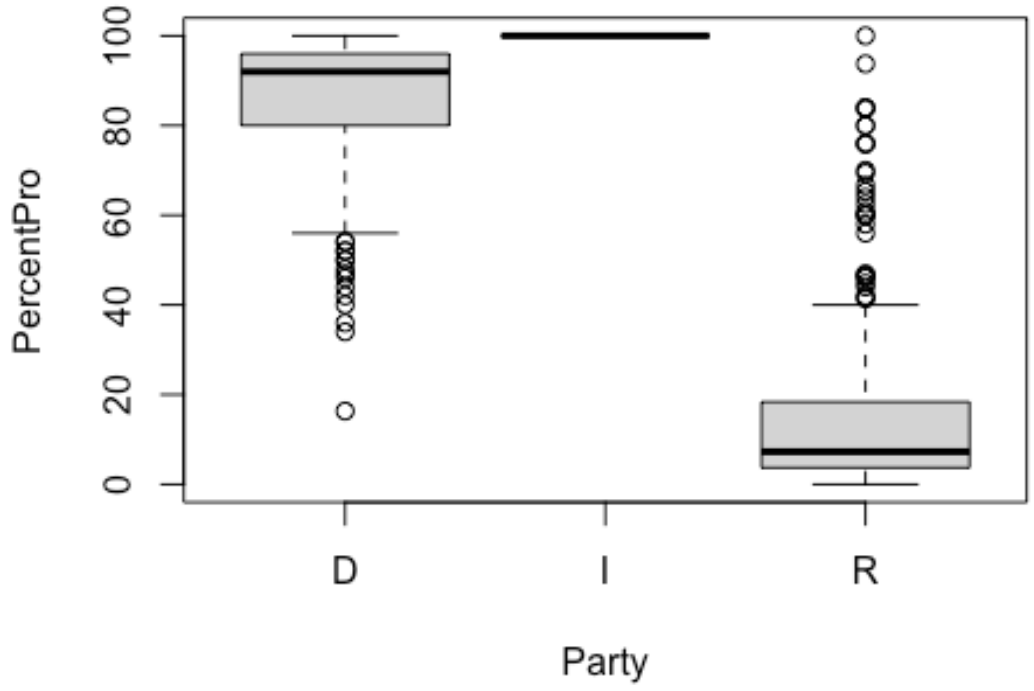
```
##           number Score Treatment
## 28 -2.3223722950  16.6 Intrinsic
## 29 -1.9672191815  17.2 Intrinsic
## 14 -1.8282291682  17.5 Extrinsic
## 27 -1.6081599298  13.6 Intrinsic
## 21 -1.5661620762  21.2 Extrinsic
## 10 -1.4657503001  16.8 Extrinsic
## 33 -1.2004406274  19.3 Intrinsic
## 6  -1.0857374702  12.0 Extrinsic
## 7  -0.9854821582  12.3 Extrinsic
## 11 -0.9224562385  17.2 Extrinsic
## 46 -0.9202418805  26.7 Intrinsic
## 26 -0.8902940235  12.9 Intrinsic
## 19 -0.7033881930  19.5 Extrinsic
```

```
## 37 -0.6932373819  20.6 Intrinsic
## 45 -0.4876963974  24.3 Intrinsic
## 23 -0.3404159885  24.0 Extrinsic
## 25 -0.3264667881  12.0 Intrinsic
## 4  -0.2817650147  10.9 Extrinsic
## 9  -0.2520458977  15.0 Extrinsic
## 43 -0.1546952364  23.1 Intrinsic
## 42 -0.1085646171  22.6 Intrinsic
## 17 -0.0810780515  18.7 Extrinsic
## 20 -0.0405781737  20.7 Extrinsic
## 36  0.0006863592  20.5 Intrinsic
## 8   0.0151308601  14.8 Extrinsic
## 30  0.0275268104  17.5 Intrinsic
## 12  0.0396024331  17.2 Extrinsic
## 41  0.0611627527  22.2 Intrinsic
## 15  0.0914729119  18.5 Extrinsic
## 5   0.1087755466  11.8 Extrinsic
## 32  0.1443376363  19.1 Intrinsic
## 3   0.1854052843   6.1 Extrinsic
## 22  0.2491481681  22.1 Extrinsic
## 38  0.2608364805  21.3 Intrinsic
## 24  0.4171908387  12.0 Intrinsic
## 31  0.4836041107  18.2 Intrinsic
## 13  0.4938201830  17.4 Extrinsic
## 39  0.5066869590  21.6 Intrinsic
## 40  0.5643018376  22.1 Intrinsic
## 16  0.6707792190  18.7 Extrinsic
## 2   0.7111425051   5.4 Extrinsic
## 34  0.8852306473  19.8 Intrinsic
## 35  0.8869350447  20.3 Intrinsic
## 47  0.8979752611  29.7 Intrinsic
## 44  0.9504841749  24.0 Intrinsic
## 18  1.2642410898  19.2 Extrinsic
## 1   1.9946963377   5.0 Extrinsic

first <- data1[1:24, 2:2]
second <- data1[25:47, 2:2]
meanfirst <- mean(first)
meansecond <- mean(second)
meandifference <- meanfirst-meansecond
```

The average of the first 24 scores is **18.1333333**, while the average of the second group of 23 scores is **17.5652174**. The difference between the two averages is **0.5681159**.

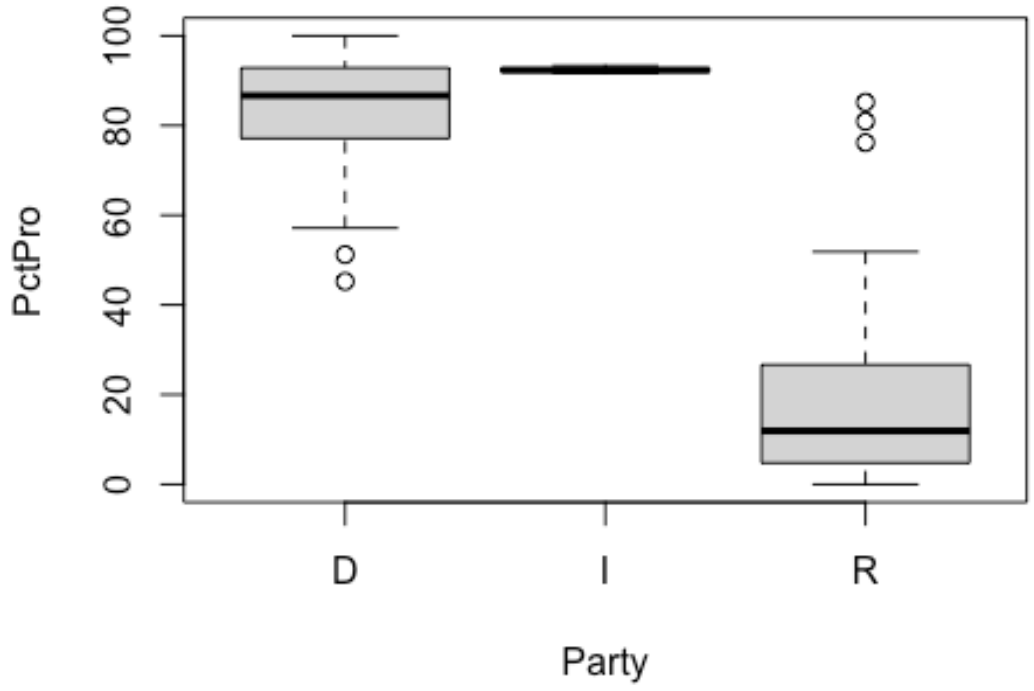
1-26




```
##         D         I         R
##  85.89794 100.00000  16.17260
```

From the boxplot and mean data above, it is relatively easy to see where the median lies for each party. For the Democratic party, the the median lies at about 90%, while the median lies at less than 20% for Republicans. The mean for the Republican party is 16.17%, while the mean for the Democratic party is approximately 85.89%. Democrats more often vote positively for environmental policies than Republicans do. Interestingly, the single Independent voter (Bernie Sanders) always voted positively for environmental policies.

1-27




```
##        D        I        R
## 83.64135 92.43667 18.54579
```

From the boxplot and mean data shown above, it is relatively easy to see where the mean median lie for each party in the Senate. For the Republican party, the median lies below 20% again, while the mean is approximately 18.54% as well, while for the Democrat Party, the median lies at around 90% pro votes on environmental votes, and the mean is 83.64%.