Island EDA

Tyler Honsinger, Ethan Stansfield, Brian Gylland

February 19, 2016

Introduction: This study investigates the effects of physical activity on cognitive ability in 5th year students on the islands. This topic was chosen to help support activities such as recess in schools, NFL Play 60, and other programs that promote physical activity for children. The goal was to provide statistical evidence that would emphasize that these programs improve more than physical health, but cognitive ability as well. The study included 79 students, consisting of 40 girls and 39 boys. All of the participants were given a survey inquiring about how forgetful they felt, followed by a 4 minute difficult arithmetic. Then, 20 of the girls and 20 of the boys were asked to run 5 kilometers each day, for the next 14 days. At this point every participant took the survey and 4 minute arithmetic again. The expectation was that the students who ran each day would see an greater improvement in their arithmetic scores than the students who did not run. The data and results can be seen below.

```
attach(Islands.EDA.Data)
summary(Islands.EDA.Data)
```

```
##             NAME    GENDER  RAN.    INITIAL.TEST.SCORE FINAL.TEST.SCORE
##  Akane Jones  : 1   F:40   No :39   Min.   : 1.00      Min.   : 0.00
##  Alice Morris : 1   M:39   Yes:40   1st Qu.: 8.00      1st Qu.: 7.00
##  Arvid Carlsen: 1                   Median :14.00      Median :14.00
##  AUSTIN HALL  : 1                   Mean   :14.95      Mean   :13.86
##  Austin Moore : 1                   3rd Qu.:20.00      3rd Qu.:19.50
##  Ava Price    : 1                   Max.   :36.00      Max.   :36.00
##  (Other)      :73
##     Initial           Final
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:1.0000   1st Qu.:1.0000
##  Median :1.0000   Median :1.0000
##  Mean   :0.9367   Mean   :0.9114
##  3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :2.0000   Max.   :2.0000
##
```

```
# Quantitative variables
initialTestMean = mean(INITIAL.TEST.SCORE); initialTestMean
```

```
## [1] 14.94937
```

```
finalTestMean = mean(FINAL.TEST.SCORE); finalTestMean
```

```
## [1] 13.86076
```

```
initialTestsd = sd(INITIAL.TEST.SCORE);initialTestsd
```

```
## [1] 8.337279
```

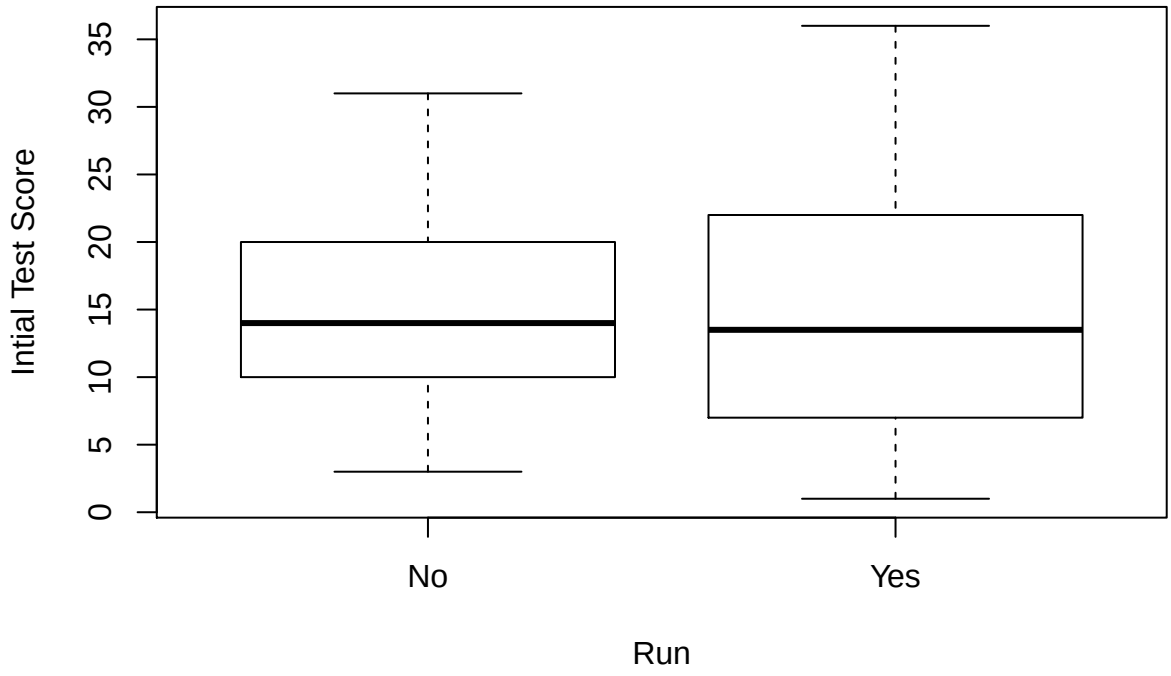
```
iqrInitial = IQR(INITIAL.TEST.SCORE); iqrInitial
```

```
## [1] 12
```

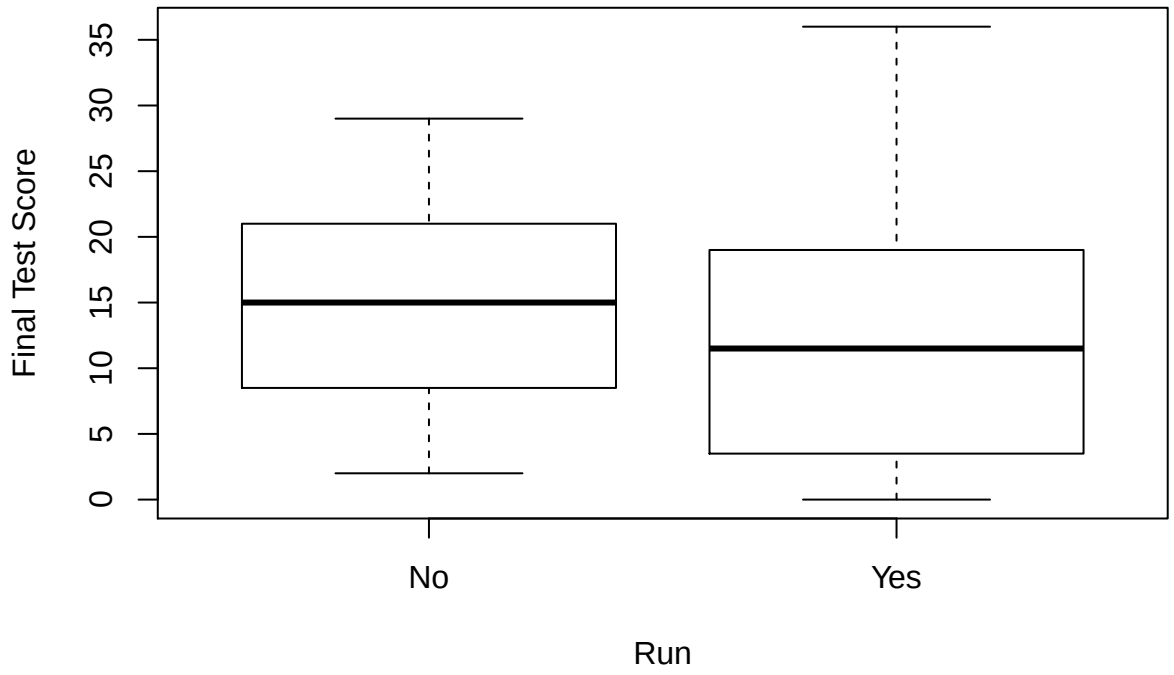
```
iqrFinal = IQR(FINAL.TEST.SCORE); iqrFinal
```

```
## [1] 12.5
```

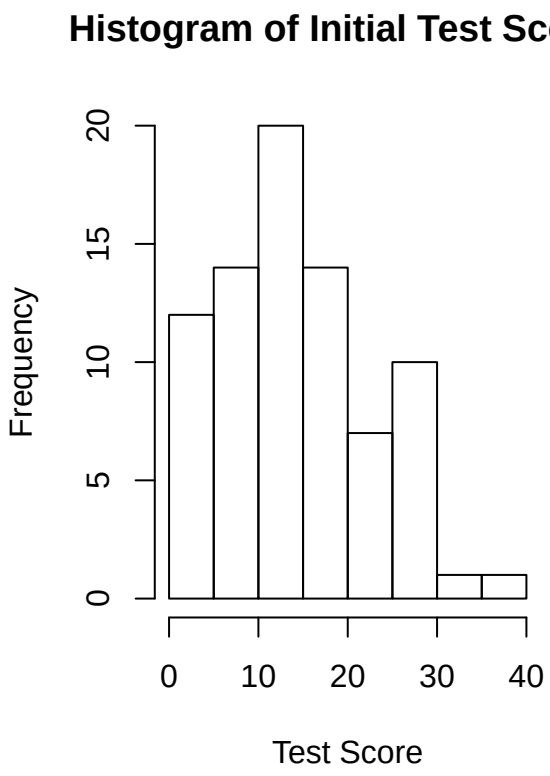
```
initialTestBox = boxplot(INITIAL.TEST.SCORE ~ RAN., xlab = "Run", ylab = "Intial Test Score")
```

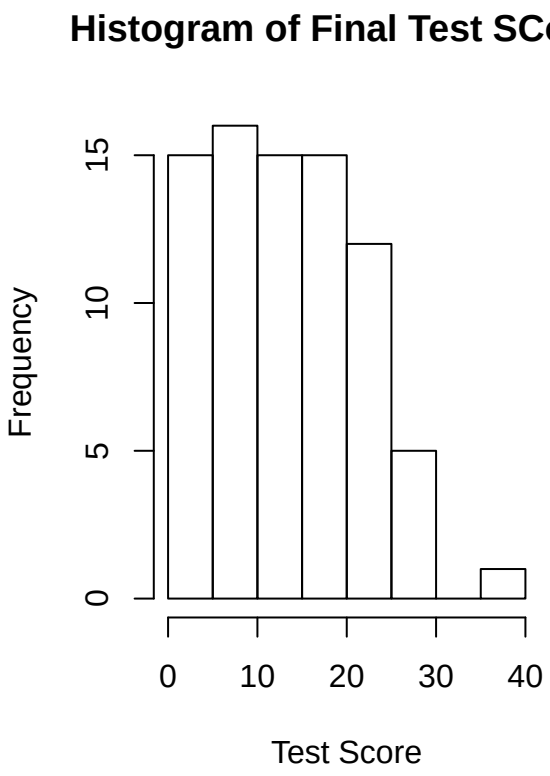



```
finalTestBox = boxplot(FINAL.TEST.SCORE ~ RAN., xlab = "Run", ylab = "Final Test Score")
```




```
par(mfrow=c(1,2))
initialTestHist = hist(INITIAL.TEST.SCORE, main = "Histogram of Initial Test Scores", ylab = "Frequency
finalTestHist = hist(FINAL.TEST.SCORE, main = "Histogram of Final Test SCores", xlab = "Test Score")
```





```r
# Categorical Variables data, 0 = "not at all", 1 = "a little", 2 = "moderately"
initialForgetfulTally = tally(Initial); initialForgetfulTally
```

```
##
##  0  1  2
## 16 52 11
```

```r
finalForgetfulTally = tally(Final); finalForgetfulTally
```

```
##
##  0  1  2
## 13 60  6
```

```r
#Proportion of students who are differing levels of forgetful at start of study
prop.0.initial = initialForgetfulTally[1]/79; prop.0.initial
```

```
##         0
## 0.2025316
```

```r
prop.1.initial = initialForgetfulTally[2]/79; prop.1.initial
```

```
##         1
## 0.6582278
```

```r
prop.2.initial = initialForgetfulTally[3]/79; prop.2.initial
```

```
##         2
## 0.1392405
```

```r
#Proportion of students who are differing levels of forgetful at end of study
prop.0.final = finalForgetfulTally[1]/79; prop.0.final
```

```
##        0
## 0.164557
```

```r
prop.1.final = finalForgetfulTally[2]/79; prop.1.final
```

```
##         1
## 0.7594937
```

```r
prop.2.final = finalForgetfulTally[3]/79; prop.2.final
```

```
##          2
## 0.07594937
```

```
intialForgetfulData = bargraph(~Initial)
finalForgetfulData = bargraph(~Final)
```

Discussion: Our sample data did not support our hypothesis that exercise leads to improvement in cognitive ability. In fact, our data supports the exact opposite. This can be seen in the "initialTestBox" and "finalTestBox". These two plots compare the test scores of the students who did not run against the test scores of the students who did run. When we compare the two plots together, we can see that the spread of the test scores narrowed and the mean lessened for the group of students who did run. For the group of students who did not run, the spread and mean stayed approximately the same. It is important to note here that though the spread changed for the running group, the minimum and maximum range did not. It is also of note that this study was a relatively small size (40 students per group) and that with a larger group it may have been possible to remove some of this variability.

When viewing initialTestHist it appears that the spread of the data is more or less symmetric, or perhaps slightly skewed right. However in finalTestHist, the data is clearly skewed right with the concentration of scores moving lower overall. This movement can also be found in the intialTestMean and the finalTestMean where the mean is more than a whole point lower at the end of the study. The data does have a couple of outliers at the upper end of the data where students scored much higher than all of their peers. This stayed consistent across the study however, and does not affect the final outcome, besides raising the mean score.

The study also looked at whether running had an impact on how forgetful the students felt on a categorical measure. The students could respond that they felt "not at all", "a little", "moderately", or "quite a bit" forgetful. For ease of analysis we changed this data into a numerical scale where 0 is "not at all" and "quite a bit" is a 3. the variables initialForgetfulTally and finalForgetfulTally show the results of this portion of the study. There does not appear to be any statistical significance in the data here. Additionally we have calculated the proportion of students who said they felt differing levels of forgetfulness both before and after the study. This data can be seen in the variables prop.x.intial and prop.x.final. Again, this proportional data does not show any strong correlation between running and levels of forgetfulness.
In order to graphically show this categorical data, we have created two bar graphs, initialForgetfulData and finalForgetfulData. These graphs plot forgetfulness level on the x-axis and number of students who said they were this forgetful on the y axis. There is little visually noticeable difference between these two graphs.

In conclusion, this experiment showed our hypothesis to be incorrect and in fact almost the opposite of what we expected. Perhaps with more students we could make this negative correlation between running and cognitive ability more pronounced statistically.