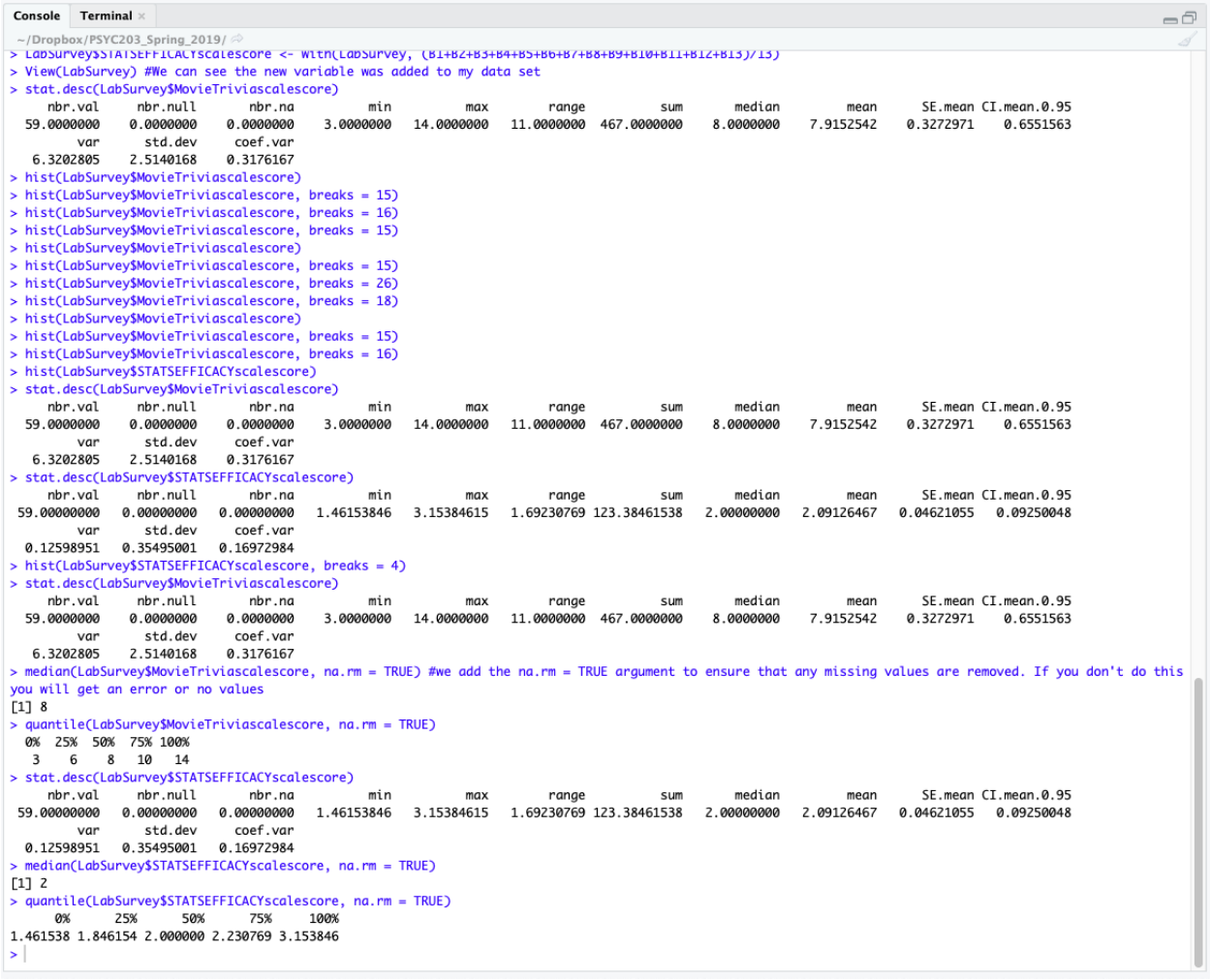
James Dumitru
IIT PSYC 203
Project 2

1&2.) Movie Trivia
Mean: 7.9152542
Median: 8
Mode: 10
Range: 11
Standard Deviation: 2.51
Variance: 6.32

Stats Efficacy
Mean: 2.09126467
Median: 2
Mode: 1.923077
Range:1.692
Standard Deviation: .35495
Variance: 1.2598951



```
Console  Terminal
~/Dropbox/PSYC203_Spring_2019/
> LabSurvey$STATSEFFICACYscalescore <- with(LabSurvey, (B1+B2+B3+B4+B5+B6+B7+B8+B9+B10+B11+B12+B13)/13)
> View(LabSurvey) #We can see the new variable was added to my data set
> stat.desc(LabSurvey$MovieTriviascalescore)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean      SE.mean CI.mean.0.95
  59.0000000    0.0000000    0.0000000    3.0000000   14.0000000   11.0000000  467.0000000    8.0000000    7.9152542    0.3272971    0.6551563
         var      std.dev     coef.var
   6.3202805    2.5140168    0.3176167
> hist(LabSurvey$MovieTriviascalescore)
> hist(LabSurvey$MovieTriviascalescore, breaks = 15)
> hist(LabSurvey$MovieTriviascalescore, breaks = 16)
> hist(LabSurvey$MovieTriviascalescore, breaks = 15)
> hist(LabSurvey$MovieTriviascalescore)
> hist(LabSurvey$MovieTriviascalescore, breaks = 15)
> hist(LabSurvey$MovieTriviascalescore, breaks = 26)
> hist(LabSurvey$MovieTriviascalescore, breaks = 18)
> hist(LabSurvey$MovieTriviascalescore)
> hist(LabSurvey$MovieTriviascalescore, breaks = 15)
> hist(LabSurvey$MovieTriviascalescore, breaks = 16)
> hist(LabSurvey$STATSEFFICACYscalescore)
> stat.desc(LabSurvey$MovieTriviascalescore)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean      SE.mean CI.mean.0.95
  59.0000000    0.0000000    0.0000000    3.0000000   14.0000000   11.0000000  467.0000000    8.0000000    7.9152542    0.3272971    0.6551563
         var      std.dev     coef.var
   6.3202805    2.5140168    0.3176167
> stat.desc(LabSurvey$STATSEFFICACYscalescore)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean      SE.mean CI.mean.0.95
 59.00000000   0.00000000   0.00000000   1.46153846   3.15384615   1.69230769 123.38461538   2.00000000   2.09126467   0.04621055   0.09250048
         var      std.dev     coef.var
  0.12598951   0.35495001   0.16972984
> hist(LabSurvey$STATSEFFICACYscalescore, breaks = 4)
> stat.desc(LabSurvey$MovieTriviascalescore)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean      SE.mean CI.mean.0.95
  59.0000000    0.0000000    0.0000000    3.0000000   14.0000000   11.0000000  467.0000000    8.0000000    7.9152542    0.3272971    0.6551563
         var      std.dev     coef.var
   6.3202805    2.5140168    0.3176167
> median(LabSurvey$MovieTriviascalescore, na.rm = TRUE) #we add the na.rm = TRUE argument to ensure that any missing values are removed. If you don't do this
you will get an error or no values
[1] 8
> quantile(LabSurvey$MovieTriviascalescore, na.rm = TRUE)
  0%  25%  50%  75% 100%
   3    6    8   10   14
> stat.desc(LabSurvey$STATSEFFICACYscalescore)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean      SE.mean CI.mean.0.95
 59.00000000   0.00000000   0.00000000   1.46153846   3.15384615   1.69230769 123.38461538   2.00000000   2.09126467   0.04621055   0.09250048
         var      std.dev     coef.var
  0.12598951   0.35495001   0.16972984
> median(LabSurvey$STATSEFFICACYscalescore, na.rm = TRUE)
[1] 2
> quantile(LabSurvey$STATSEFFICACYscalescore, na.rm = TRUE)
       0%       25%       50%       75%      100%
1.461538 1.846154 2.000000 2.230769 3.153846
>
```

3.) Going back to the meaning of the two, Central Tendency refers to the middle of the distribution, while Variability is about the spread. If I were to report 1 of each, I would use the mean because it's generally preferred unless there is a "bad egg" In the distribution for which I'll use the Median. As for the one measure of variability, I would use the Variance because it's the most likely to be unbiased to the information.

4.) I believe the skew would go away, depending if the "bad" bar went away, if it was close to the tail and but not the end, it would enlarge the skew.

5.)

```
> stat.desc(LabSurvey$MovieTrivia)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean
  58.0000000    0.0000000    0.0000000    3.0000000   14.0000000   11.0000000 462.0000000    8.0000000    7.9655172
     SE.mean CI.mean.0.95          var      std.dev     coef.var
   0.3290397    0.6588906    6.2794918    2.5058914    0.3145924
> stat.desc(LabSurvey$StatsEfficacy)
     nbr.val     nbr.null       nbr.na          min          max        range          sum       median         mean
 58.00000000   0.00000000   0.00000000   1.46153846   3.15384615   1.69230769 121.30769231   2.00000000   2.09151194
     SE.mean CI.mean.0.95          var      std.dev     coef.var
  0.04701360   0.09414312   0.12819618   0.35804494   0.17118953
```

Movie Trivia

- Mean: 7.9655172
- Median: 8
- Mode: 10
- Range: 11
- Standard Deviation: 2.51
- Variance: 6.2794918

Stats Efficacy

- Mean: 2.09151194
- Median: 2
- Mode: 1.923077
- Range:1.692
- Standard Deviation: .35495
- Variance: .12819618

Yes, my prediction came true, the proof is in the increase from the mean and variance. Meaning that the random case created a sort of "hole" that skewed the data even further even if it was slightly.