Practical 2

Trevor Atkins

9/8/2020

Part A

1. Open data

1a

```
Practical1 <- readRDS(file="Practical1.rds")
```

2. First Calculations: Descriptive Statistics

```
table(Practical1$age)
```

```
##
## 14 15 16 17 18 19 20 21 22 23 24 25 26 27
##  1  1  3  2  3  1  3  3  3  2  3  2  1  1
```

The ages that occur most often are tied at 3 times with age 16, 18, 20, 21, 22, and 24.

3. More Descriptives: Exploring Frequencies

3a

```
mean(Practical1$profsc)
```

```
## [1] 69.13793
```

```
median(Practical1$profsc)
```

```
## [1] 68
```

```
range(Practical1$profsc)
```

```
## [1] 33 97
```

```
sd(Practical1$profsc)
```

```
## [1] 16.53508
```

```
which.max(tabulate(Practical1$profsc))
```

```
## [1] 59
```

The mean or average of the Proficiency Score is 69.1. The median of the Proficiency Score is 68. The range of the Proficiency Score is from 33 to 97. The standard deviation of the Proficiency Score is 16.5, which means that the values tend to be close to the mean of the set. The mode of the Proficiency Score is 59. #### 3b

```
min(Practical1$age)
```

```
## [1] 14
```

```
max(Practical1$age)
```

```
## [1] 27
```

```
range(Practical1$age)
```

```
## [1] 14 27
```

```
sd(Practical1$age)
```

```
## [1] 3.5317
```

The minimum age is 14 and the maximum age is 27. The standard deviation of age is 3.53.

3b

The most frequently occuring proficiency score is the mode, which is 59.

4. Getting to Know the Data: Relationships

4a-b

```
plot(Practical1$age, Practical1$profsc, xlab="Age of Acquisition", ylab="Profic
iency Score", main="Age over Proficiency")
```

Figure 1 Scatterplot showing the relationship between Age of Acquisition and Proficiency

4c

At face value, I think there is a relationship between the two variables: as the age of acquisition increases, the proficiency of the learner decreases.

5. Getting to Know the Data: Comparing Groups

5a

```
aggregate(Practical1$profsc~Practical1$gender, Practical1, mean)
```

```
##   Practical1$gender Practical1$profsc
## 1            Female          75.53333
## 2              Male          62.28571
```

5b

The female participants has higher proficiency scores

```
aggregate(Practical1$profsc~Practical1$gender, Practical1, sd)
```

```
##   Practical1$gender Practical1$profsc
## 1            Female          14.24212
## 2              Male          16.51307
```

5c

The male participants scored more homogeneously.

```
boxplot(Practical1$profsc~Practical1$gender, data=Practical1, main="Proficiency
over Gender", xlab="Gender", ylab="Proficiency Score")
```

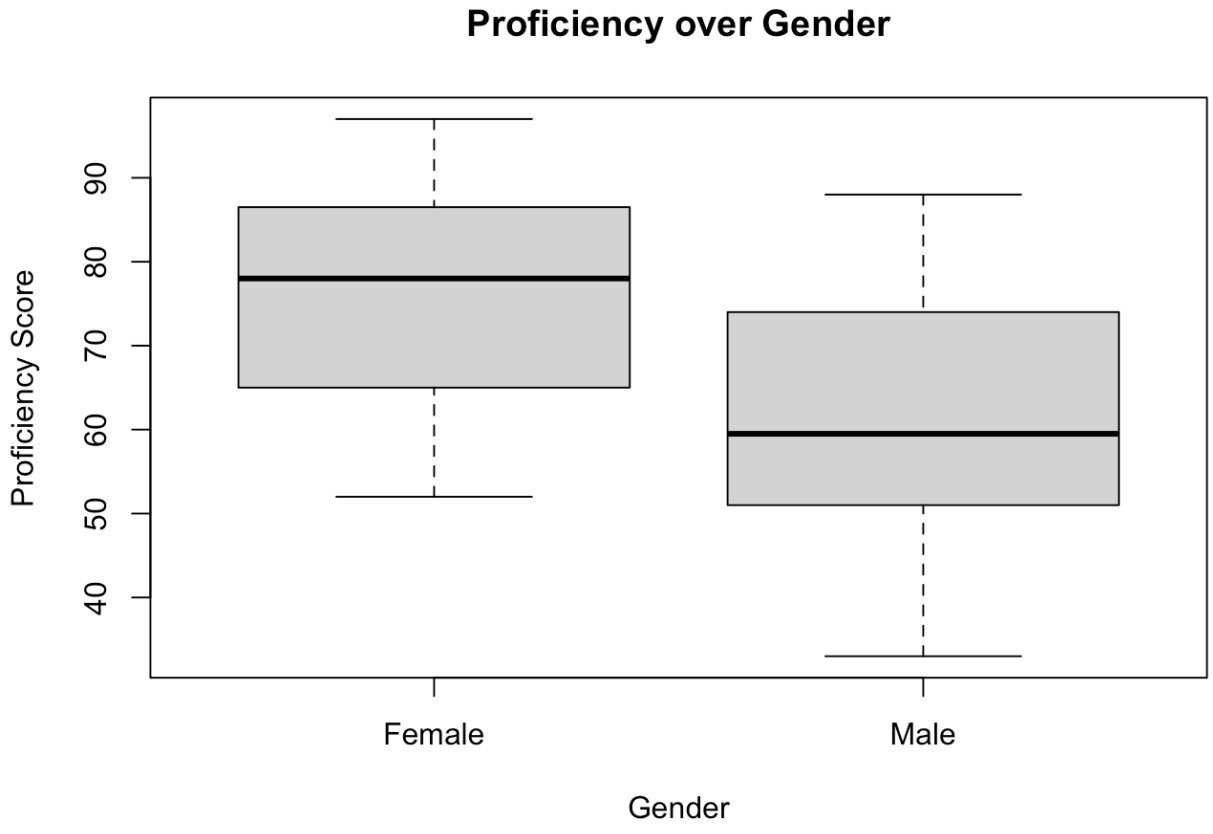


Figure 2 Boxplot showing the comparison of proficiency score between genders.

The median or middle quartile marks the mid-point of the data and is by the bold line that splits the box in half. Half are greater than equal to this value and half are less than equal to this value. Inter-quartile range, the box represents the middle 50% of scores for the group. Upper quartile: 75% of

scores fall below the upper quartile. Lower quartile: 25% of scores fall below the lower quartile.

Part B

1. Entering Data Manually

1a

```
a <- c(3,4,5,6,7,8,9)
b <- c(6,6,6,6,6,6,6)
c <- c(4,4,4,6,7,7,10)
d <- c(1,1,1,4,9,12,14)
install.packages("psych")
```

```
##
## The downloaded binary packages are in
##  /var/folders/_6/4dcygrqd30z3yrk1v612llt40000gn/T//RtmpxzNvyv/downloaded_pac
kages
```

```
library("psych")
mode <- function(z){
  unique_element <- unique(z)
  unique_element[which.max(tabulate(match(z,unique_element)))]
}
describe(a)
```

```
##    vars n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 7    6 2.16      6       6 2.97   3   9     6    0    -1.71 0.82
```

```
mode(a)
```

```
## [1] 3
```

```
describe(b)
```

```
##    vars n mean sd median trimmed mad min max range skew kurtosis se
## X1    1 7    6  0      6       6   0   6   6     0  NaN      NaN  0
```

```
mode(b)
```

```
## [1] 6
```

```
describe(c)
```

```
##    vars n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 7    6 2.24      6       6 2.97   4  10     6 0.54    -1.25 0.85
```

```
mode(c)
```

```
## [1] 4
```

```
describe(d)
```

```
##    vars n mean  sd median trimmed  mad min max range skew kurtosis   se
## X1    1 7    6 5.6      4       6 4.45   1  14    13  0.3    -1.93 2.12
```

```
mode(d)
```

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

1b

No, I do not agree with R's calculation of the mode for variable 'a' because there should be no mode as no number appears more frequently than any other in the list.

Part C

1. Load the File Into R

1a

```
Practical2 <- read.csv(file="Data-Practical2C.csv", head=TRUE, sep=";")
Practical2$Motivation <- as.ordered(Practical2$Motivation)
Practical2$Motivation <- ordered(Practical2$Motivation,
levels = c(1,2,3,4,5),
labels = c("very low", "low", "neutral", "high", "very high"))
saveRDS(Practical2, file="Practical2.rds")
write.csv(Practical2, file="Practical2C.csv")
summary(Practical2)
```

```
##     Subject           Motivation  Proficiency
##  Min.   :  1.00   very low :20   Min.   :13.00
##  1st Qu.: 25.75   low      :24   1st Qu.:34.75
##  Median : 50.50   neutral  :27   Median :46.50
##  Mean   : 50.50   high     :17   Mean   :46.85
##  3rd Qu.: 75.25   very high:12   3rd Qu.:59.00
##  Max.   :100.00                  Max.   :85.00
```

```
str(Practical2)
```

```
## 'data.frame':    100 obs. of  3 variables:
##  $ Subject    : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Motivation : Ord.factor w/ 5 levels "very low"<"low"<..: 5 3 1 2 2 2 1 5
4 3 ...
##  $ Proficiency: int  72 28 26 62 76 40 57 64 66 72 ...
```

2. Descriptive Statistics

2a

```
aggregate(Proficiency~Motivation, Practical2, mean)
```

```
##   Motivation Proficiency
## 1   very low    39.40000
## 2        low    49.29167
## 3    neutral    45.77778
## 4       high    48.23529
## 5  very high    54.83333
```

```
aggregate(Proficiency~Motivation, Practical2, median)
```

```
##   Motivation Proficiency
## 1   very low        41.0
## 2        low        50.0
## 3    neutral        47.0
## 4       high        46.0
## 5  very high        60.5
```

```
aggregate(Proficiency~Motivation, Practical2, sd)
```

```
##   Motivation Proficiency
## 1   very low    15.91887
## 2        low    15.78794
## 3    neutral    16.91911
## 4       high    15.20497
## 5  very high    17.16674
```

```
describe(Practical2$Proficiency)
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   s
e
## X1    1 100 46.85 16.52   46.5   46.51 18.53  13  85    72 0.19    -0.66 1.6
5
```

```
mode(Practical2$Proficiency)
```

```
## [1] 49
```

```
VeryLow <- Practical2$Proficiency[Practical2$Motivation == "very low"]
Low <- Practical2$Proficiency[Practical2$Motivation == "low"]
Neutral <- Practical2$Proficiency[Practical2$Motivation == "neutral"]
High <- Practical2$Proficiency[Practical2$Motivation == "high"]
VeryHigh <- Practical2$Proficiency[Practical2$Motivation == "very high"]
mode(VeryLow)
```

```
## [1] 60
```

```
mode(Low)
```

```
## [1] 40
```

```
mode(Neutral)
```

```
## [1] 28
```

```
mode(High)
```

```
## [1] 66
```

```
mode(VeryHigh)
```

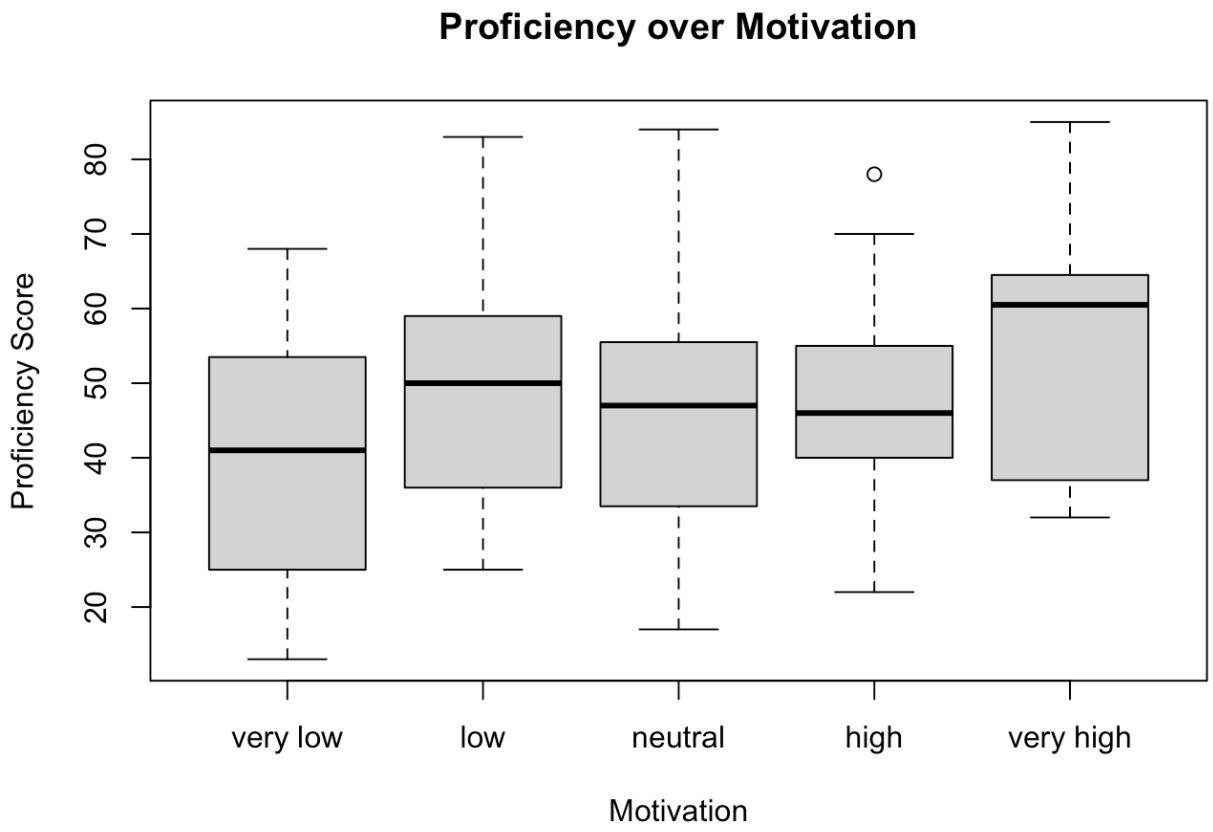
```
## [1] 64
```

Table 1: Shows the Mean, Mode, Median, and Standard Deviation Proficiency Scores for the Overall participants and by Motivation Level

value	Overall	very low	low	neutral	high	very high
mean	46.85	39.4	49.3	45.8	48.2	54.8
mode	49	60	40	28	66	64
median	46.5	41.0	50.0	47.0	46.0	60.5
sd	16.5	15.9	15.8	16.9	15.2	17.2

2b

```
boxplot(Proficiency~Motivation, data=Practical2, main="Proficiency over Motivat
ion", xlab="Motivation", ylab="Proficiency Score")
```

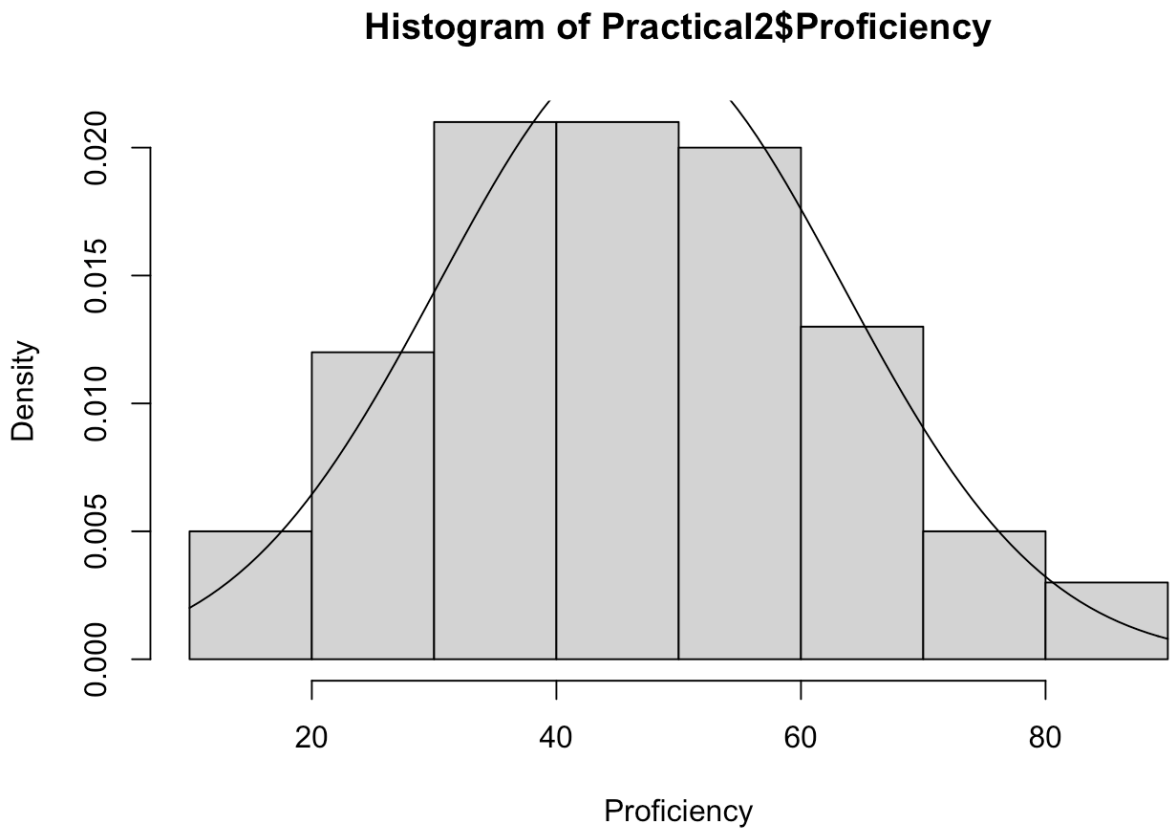


Judging from the boxplot, I think the groups will differ from one another. As the motivation increases, the upper quartile increases and the lower quartile decreases. Which means most of the higher motivation groups will have a higher percentage of people who have higher proficiency scores and a lower percentage of people who have lower proficiency scores.

3. Checking the Normal Distribution

3a

```
hist(Practical2$Proficiency, prob=TRUE, xlab="Proficiency")
curve(dnorm(x, mean=mean(Practical2$Proficiency),sd=sd(Practical2$Proficienc
y)), add=TRUE)
```



3b By looking at the histogram, the data seems to follow a normal bell curve distribution.

3c

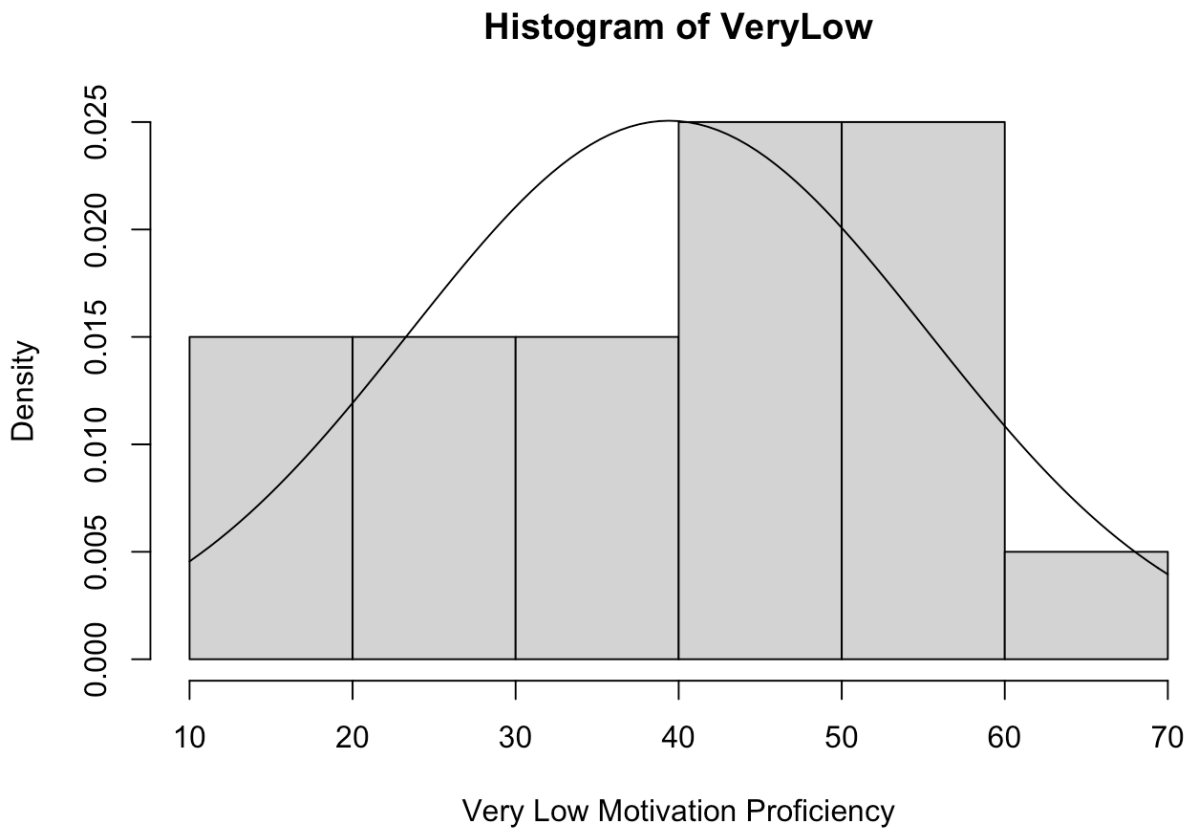
```
describe(Practical2$Proficiency)
```

```
##    vars   n  mean    sd median trimmed   mad min max range skew kurtosis   s
e
## X1    1 100 46.85 16.52   46.5   46.51 18.53  13  85    72 0.19    -0.66 1.6
5
```

Skewness and kurtosis deviate from zero. Skewness is positive and kurtosis is negative. Positive skew shows that the there is a bigger distribution on the right side of the curve and the mean will be greater than the median. A negative kurtosis shows that a distribution is flat and has thin tails.

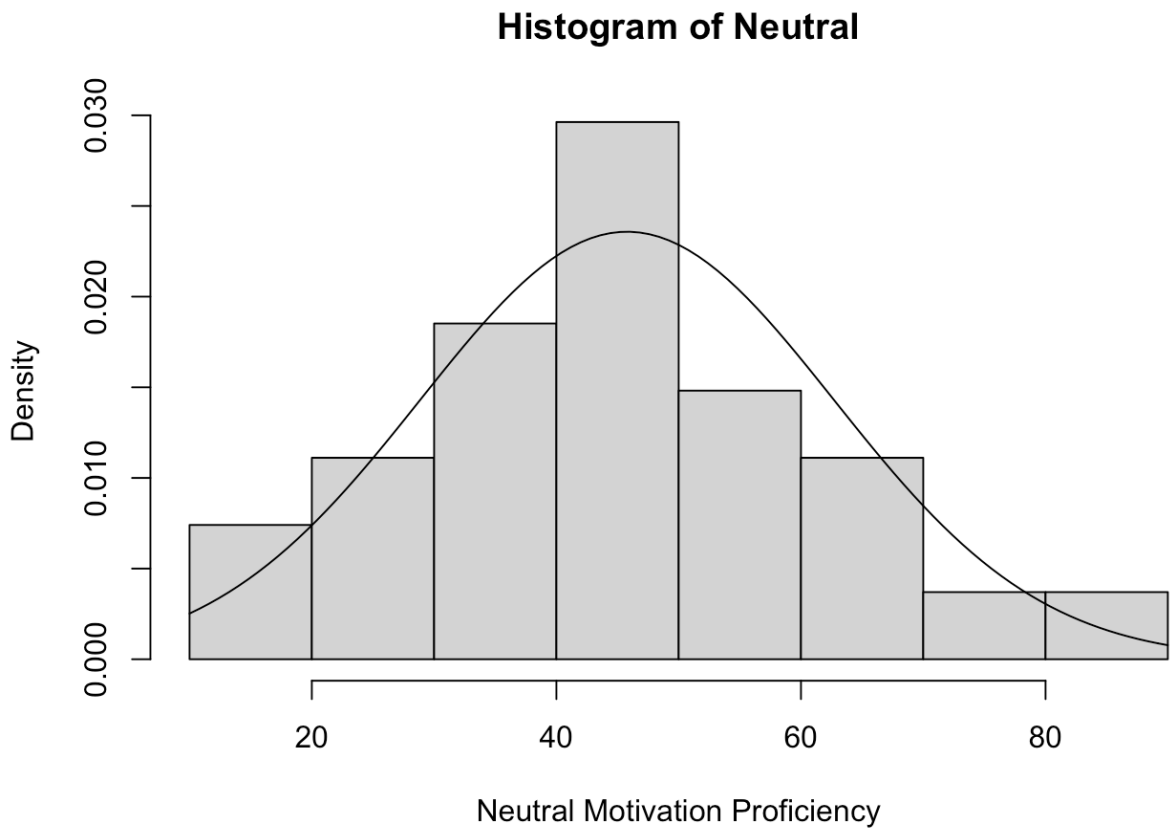
3d

```
hist(VeryLow, prob=TRUE, xlab="Very Low Motivation Proficiency")
curve(dnorm(x, mean=mean(VeryLow),sd=sd(VeryLow)), add=TRUE)
```

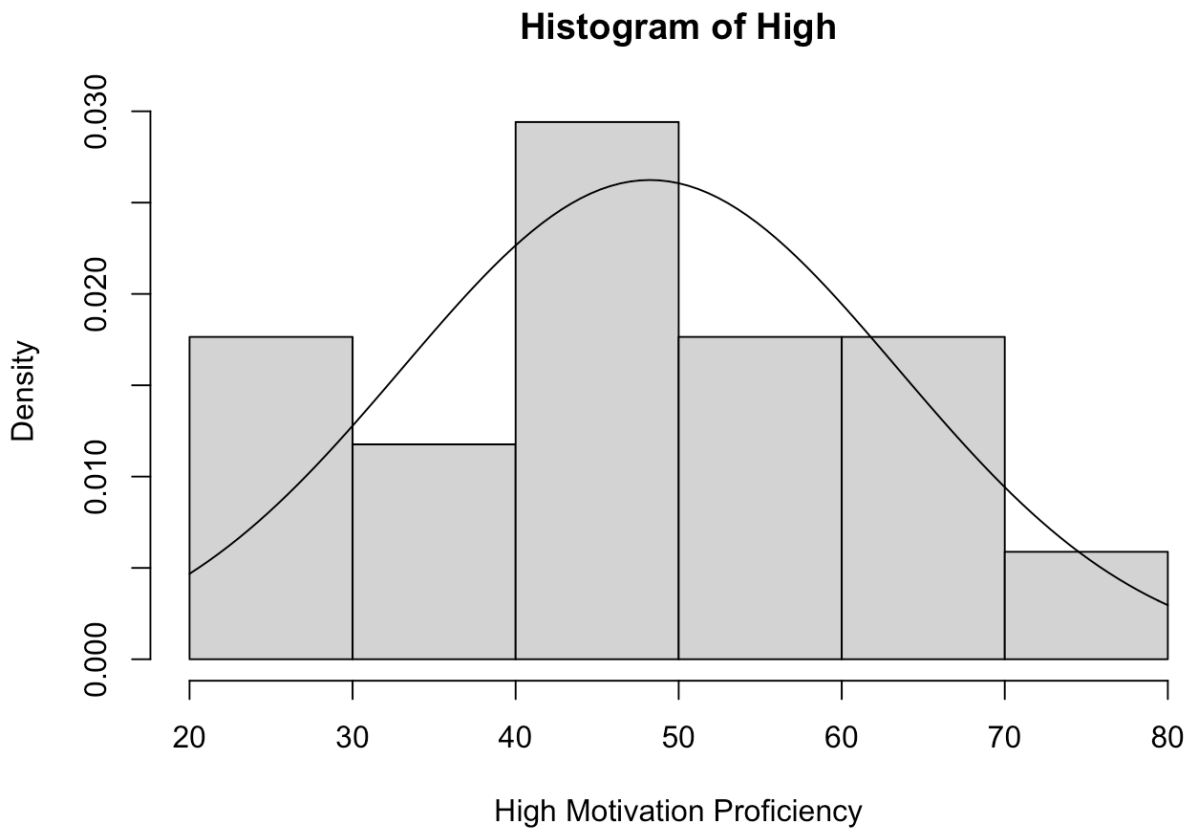



```
hist(Low, prob=TRUE, xlab="Low Motivation Proficiency")
curve(dnorm(x, mean=mean(Low),sd=sd(Low)), add=TRUE)
```

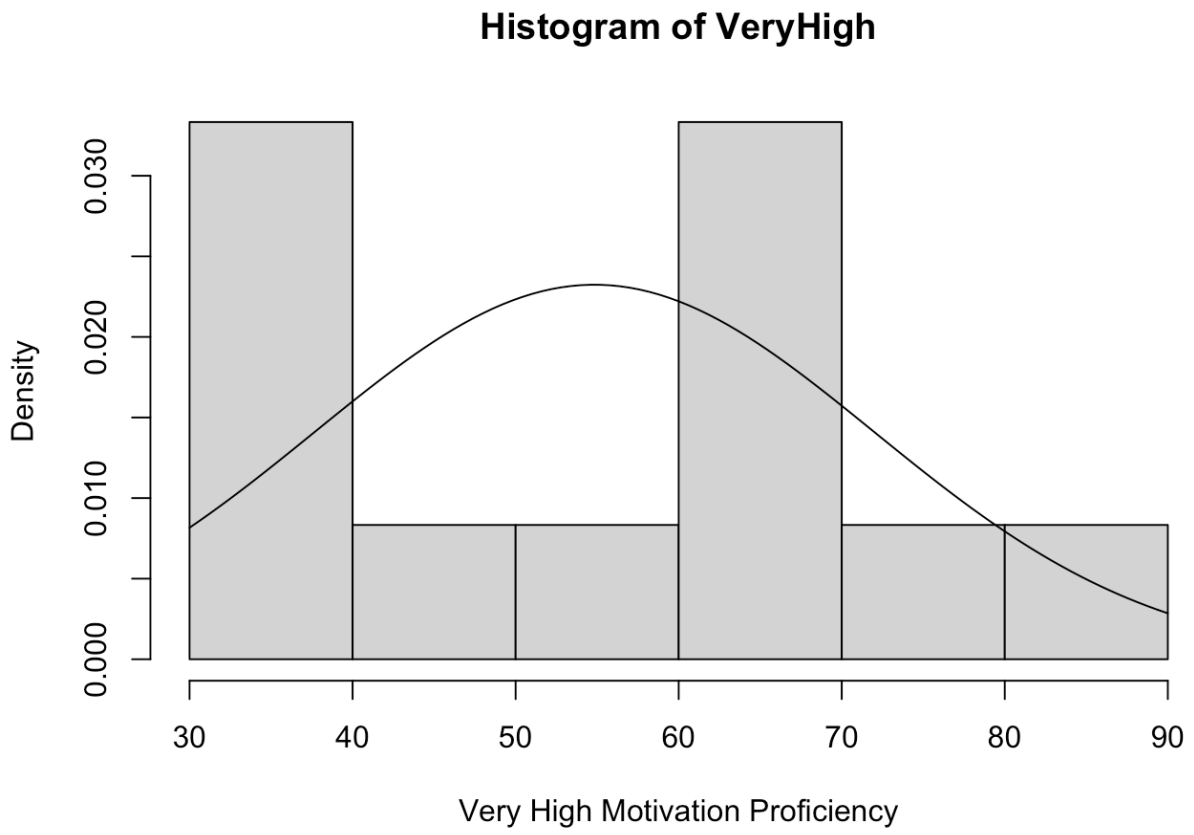

```
hist(Neutral, prob=TRUE, xlab="Neutral Motivation Proficiency")
curve(dnorm(x, mean=mean(Neutral),sd=sd(Neutral)), add=TRUE)
```




```
hist(High, prob=TRUE, xlab="High Motivation Proficiency")
curve(dnorm(x, mean=mean(High),sd=sd(High)), add=TRUE)
```




```
hist(VeryHigh, prob=TRUE, xlab="Very High Motivation Proficiency")
curve(dnorm(x, mean=mean(VeryHigh),sd=sd(VeryHigh)), add=TRUE)
```




```
describe(VeryLow)
```

```
##    vars  n mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 20 39.4 15.92     41   39.25 20.76  13  68    55 0.08     -1.3 3.56
```

```
describe(Low)
```

```
##    vars  n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 24 49.29 15.79     50    48.6 17.05  25  83    58 0.35    -0.92 3.22
```

```
describe(Neutral)
```

```
##    vars  n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 27 45.78 16.92     47   45.39 17.79  17  84    67 0.23    -0.65 3.26
```

```
describe(High)
```

```
##    vars  n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1    1 17 48.24 15.2     46      48 11.86  22  78    56 0.23    -0.87 3.69
```

```
describe(VeryHigh)
```

```
##    vars  n  mean    sd median trimmed   mad min max range skew kurtosis   se
## X1    1 12 54.83 17.17   60.5    54.1 17.79  32  85    53 0.02     -1.4 4.96
```

By looking at the histogram, all of the data groups except for the Neutral Motivation seems to not follow a normal bell curve distribution. Also all of the groups seem to have a positive skew and negative kurtosis.