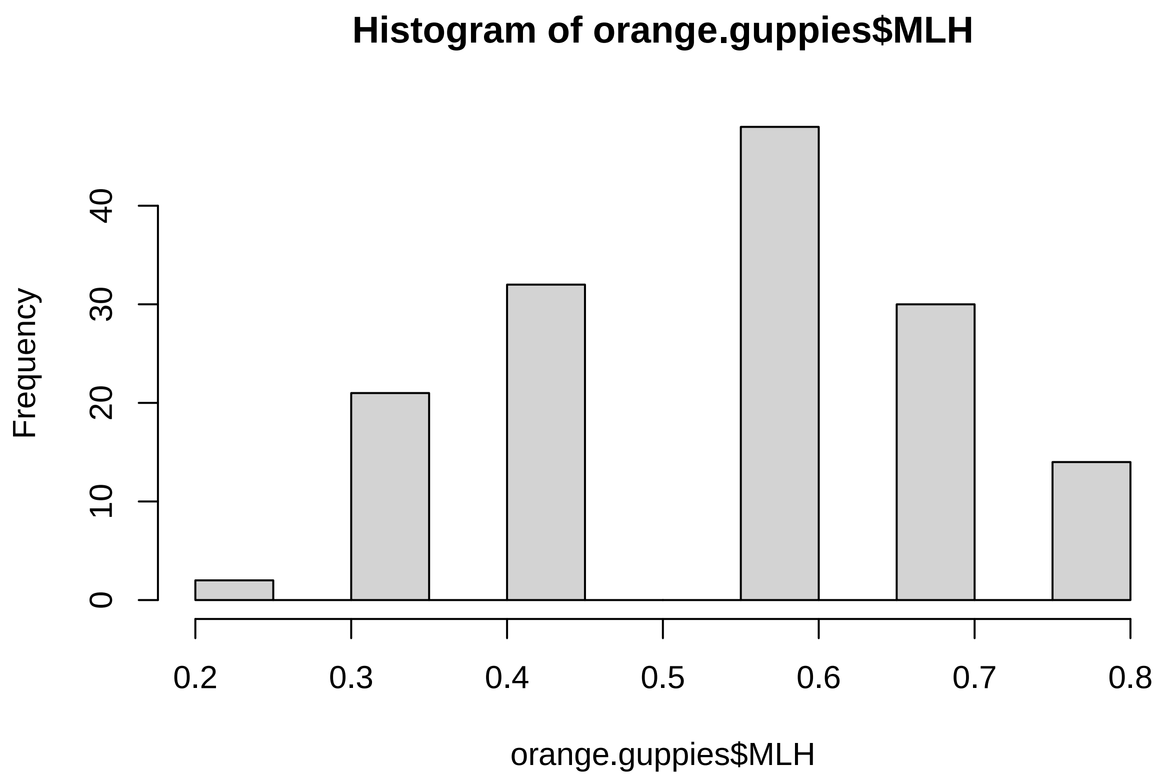
Homework 2 - Public Health 490Z

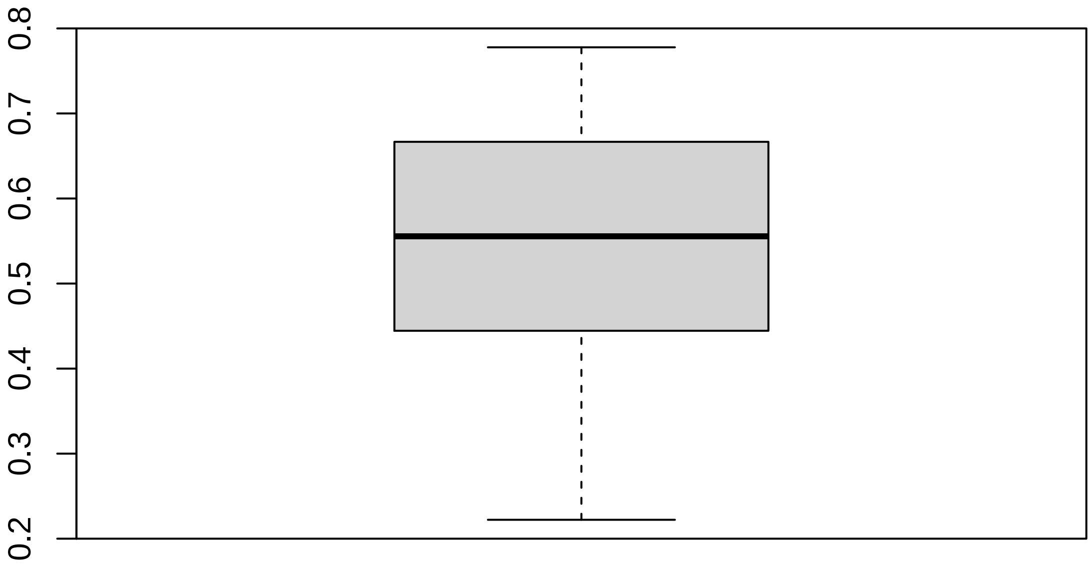
Guppy Coloration

Part A

```
load("C:/Users/aaron/Downloads/Homework 2-20211009/orange.guppies.Rdata")
load("C:/Users/aaron/Downloads/Homework 2-20211009/heartrate.exercise.Rdata")
hist(orange.guppies$MLH)
```




```
boxplot(orange.guppies$MLH)
```



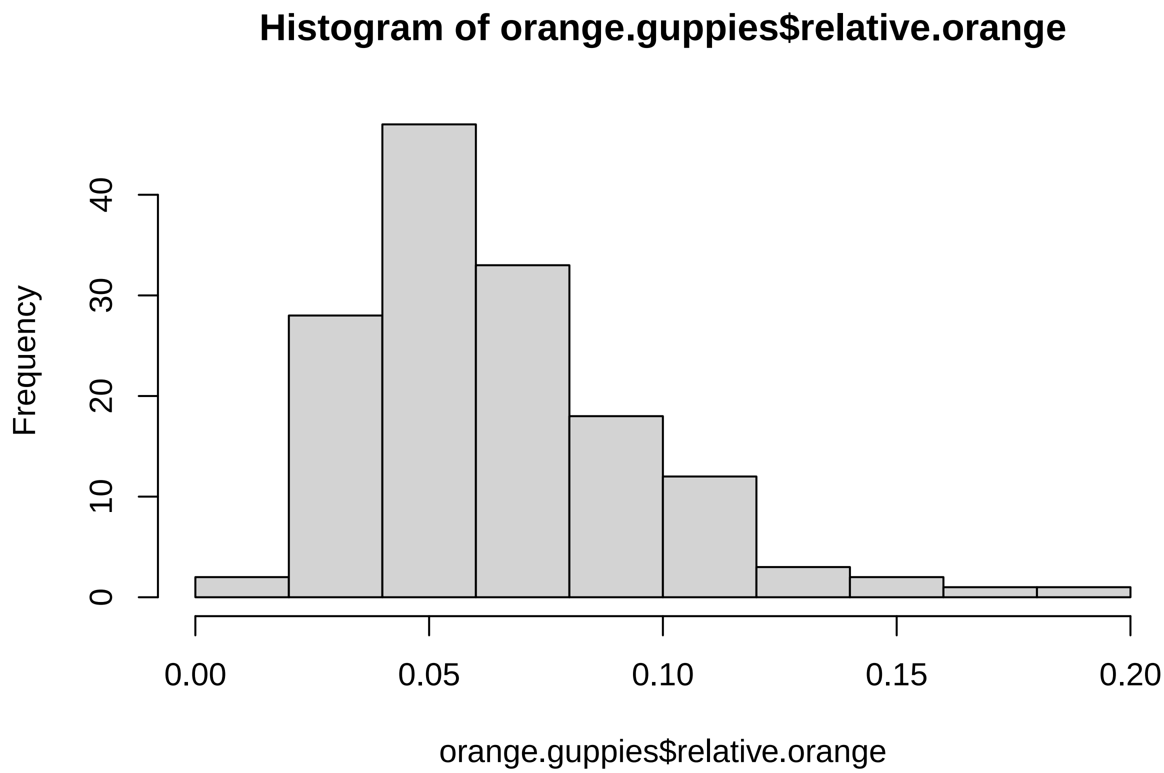

```
summary(orange.guppies$MLH)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2222  0.4444  0.5556  0.5389  0.6667  0.7778
```

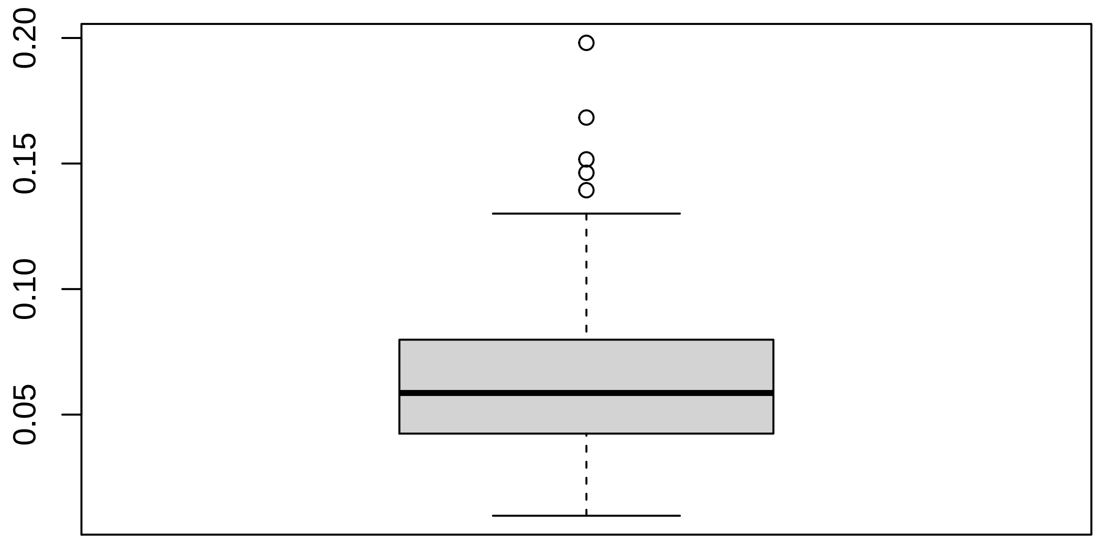
The data is slightly left skewed, with no data points where MLH equals 0.5, because the Median is greater than the Mean. The data is continuous because you can have a scrolling proportion of loci that are heterozygous. It's not discrete because you can have a proportion in-between any 2 random proportions.

Part B

```
orange.guppies$size <- orange.guppies$length * orange.guppies$height
orange.guppies$relative.orange <- orange.guppies$orange.area / orange.guppies$size
hist(orange.guppies$relative.orange)
```




```r
boxplot(orange.guppies$relative.orange)
```



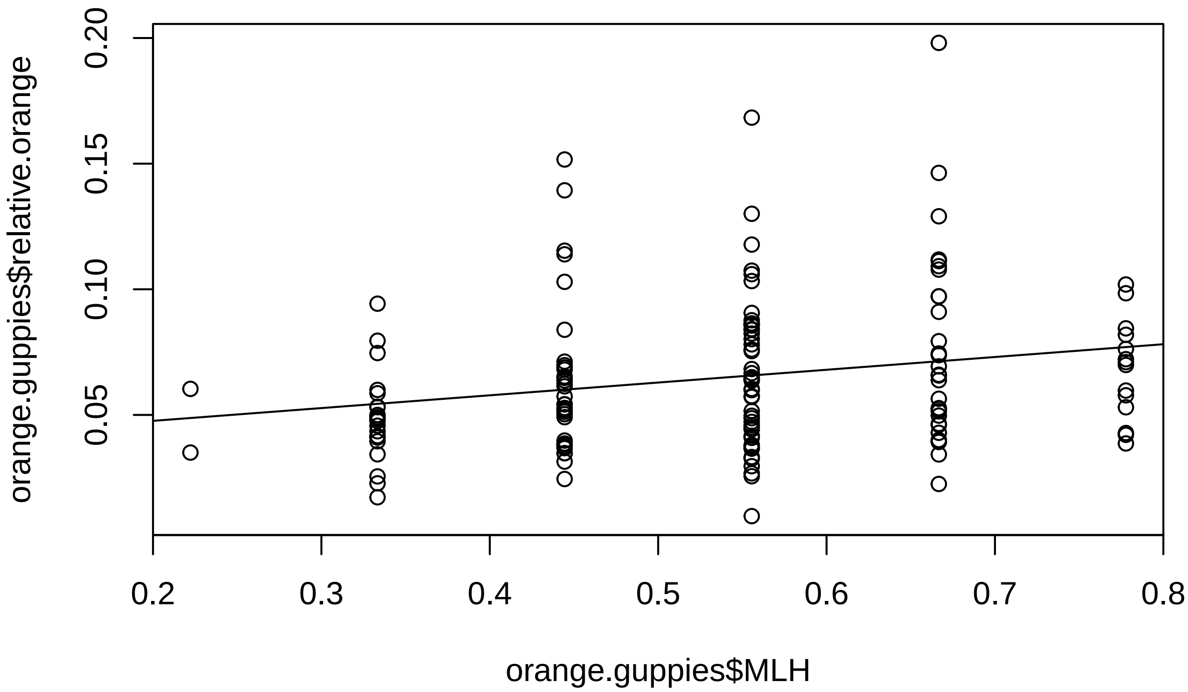

```r
summary(orange.guppies$relative.orange)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00973 0.04244 0.05862 0.06486 0.07985 0.19807
```

The distribution of relative.orange is strongly right skewed because its median is lesser than its mean. There are 5 outliers on the higher end of relative.orange.

Part C

```r
plot(orange.guppies$MLH, orange.guppies$relative.orange)
fit <- lm(orange.guppies$relative.orange ~ orange.guppies$MLH)
abline(fit)
```



The linear model seems to be an okay fit for the data, it represents some of the lowest and highest MLH values well but not so for MLH values around 0.5 or 0.6.

Part D

```r
summary(fit)
```

```
##
## Call:
## lm(formula = orange.guppies$relative.orange ~ orange.guppies$MLH)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.055977 -0.021082 -0.005663  0.013413  0.126711
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.03747    0.01019   3.677 0.000332 ***
## orange.guppies$MLH  0.05082    0.01834   2.771 0.006327 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03 on 145 degrees of freedom
## Multiple R-squared:  0.05028,    Adjusted R-squared:  0.04373
## F-statistic: 7.677 on 1 and 145 DF,  p-value: 0.006327
```

Ho: B = 0 (no association between relative.orange and MLH)

Ha: B does not equal 0 (there is an association between relative.orange and MLH)

Significance Level: a = 0.05

```
t <- 0.05082/0.01834
t
```

```
## [1] 2.770992
```

t = 2.77

p value = 0.015

Our p value is less than our significance level of a = 0.05, which means we reject the null hypothesis that there is no association between relative.orange and MLH.

Resting Heart Rates

Part A

heart rate = 83.277 - 1.566*exercise_hrs

Part B

There is a negative correlation between exercise and heart rate. This model predicts 4.97% of the total variability in heart rate.

Part C

Ho: Exercise is not a statistically significant predictor of heart rate (B = 0)

Ha: Exercise is a statistically significant predictor of heart rate (B does not equal 0)

Significance Level: a = 0.05

```
t <- -1.566/1.070
t
```

```
## [1] -1.463551
```

t = -1.463

p value = 0.120

Our p value is less than our significance level of a = 0.05, which means we reject the null hypothesis that exercise is not a statistically significant predictor of heart rate.

Part D

```
reg <- lm(heartrate_data$heart_rate ~ (heartrate_data$exercise_hrs + heartrate_data$gender))
reg
```

```
##
## Call:
## lm(formula = heartrate_data$heart_rate ~ (heartrate_data$exercise_hrs +
##     heartrate_data$gender))
##
## Coefficients:
##                 (Intercept)  heartrate_data$exercise_hrs
##                      83.816                       -1.508
##   heartrate_data$gendermale
##                      -1.487
```

Interpretation: When exercise_hrs and gendermale equals 0 (meaning that the subject is female), the resting heart rate is 83.816. If the subject is male, their resting heart rate decreases by 1.487. For every additional exercise hour, resting heart rate also decreases by 1.508.

Part E

You could test with simple linear regression. Assume normally distributed, independence, same variance across residual distribution, and linearity. You would have to build a linear regression line between the 2 data and find out if there is a strong relationship between the two variables. Two sample t-test is probably easier.

Food Insecurity Data Challenge

Urban Institute

The COVID-19 pandemic has challenged families in terms of food insecurity.

Likewise, the lack of substantial data has stopped or slowed research on the topic, indirectly resulting in policymakers being unable to interact with communities.

Some suggest to combine different datasets.

But there are different thresholds to food insufficiency, used by different data sets.

The Brookings Institution and the Institute for Policy Research at NWU look at the relationship between food insecurity and food insufficiency, before and during the pandemic.

They combine the following survey datasets: US Census Current Population FSS, COVID Impact, National Health Interviews. These are mapped to the results of the US Census Bureau's HPS.

People still debate on how well food insecurity can be measured in data.

Some researchers have concluded to not combine survey datasets because of how surveys may differ in structure, sample weights, etc.

Others think that the surveys are not structurally different enough to warrant not making a comparison, and reweighted the data with the sample weights to create more consistency.

In the Brookings graph, food insecurity shot up at the start of the pandemic but slightly lowered at the end of summer 2021.

The inconsistency of increments on the X-axis make the drop in food insecurity in summer 2021 appear faster than it was.

The dashed lines suggest a linear connection between datasets, which may not be accurate.

Elaboration about dataset changes could be made clearer.

Their modified graph shows that food insecurity shot up at the start of the pandemic, but has since been slowly decreasing. This graph does not show the most recent results, though.

Their second graph specifies that the two datasets are not comparable and show that in HPS Phase 2 that food insecurity has been decreasing, as well as 2018 values from FSS.

Their graph more correctly represents the data visually and addresses some of the methodological concerns other researchers raised.

USDA

89.5 percent (116.7 million) of U.S. households were food secure throughout 2020.

Unchanged from 89.5 percent in 2019.

Food insecure - uncertain of having/unable to acquire enough food to meet family member need because of insufficient money or resources for food. Includes low food security households.

10.5 percent (13.8 million) of U.S. households were food insecure at some time during 2020.

Unchanged from 10.5 percent in 2019.

Low food security - Enough food to avoid disrupting eating patterns or reducing intake/variety in diet.

6.6 percent (8.6 million) of U.S. households had low food security in 2020.

Essentially unchanged from 6.4 percent in 2019.

Very low food security - Normal eating patterns disrupted because of insufficient money or resources.

3.9 percent (5.1 million) of U.S. households had very low food security at some time during 2020.

Essentially unchanged from 4.1 percent in 2019.

Majority of households with children were food-secure, 85.2%.

More households with Black or Hispanic reference person were food insecure than those who were not.

Households with lower income were more food-insecure than ones who were not.

Households with children had a higher rate of food insecurity (14.8 percent) than those without children.

Single woman households had higher rates of food insecurity than women or men living alone.

Food insecurity rates shot up at the start of the 2008 market crash.

White House

Pandemic relief programs improved food security among households experiencing financial hardship.

Coronavirus dealt a rapid and significant blow to the US economy and workforce, starting a national hunger crisis.

The Federal government has a greater ability to send available resources across the nation than the nonpublic sector.

Pre-pandemic trend graph shows that food insecurity rates shot up at the start of the 2008 market crash.

Mid-April 2021 was the lowest point recorded in food insufficiency rate for households, 8 percent.

The number of households with enough food to eat dropped 4.5 million from December 2020 to June 2021.

Half of this number were households with children.

Weekly food insufficiency increased in the last months of 2020, but slowly decreased in the start of 2021.

In Summer 2021, weekly food insufficieny has slowly started increasing again.

The Consolidated Appropriations Act of 2021 provided $600 cash payments for qualifying adults and children.

The American Rescue Plan Act of 2021 provided $1400.

Research of UMichigan's Poverty Solutions shows that COVID relief bills reduced food insufficiency.

The CAA of 2021 also provided $300 per week unemployment insurance until March 2021, until it was extended to September 2021 by the ARP.

Households on unemployment insurance benefits showed steeper increases in food insufficiency rates after the $600 Federal benefit boost expired in JUly 2020.

Food insecurity rates are again increasing.

More relief efforts are clearly required to further reduce food deprivation caused by the COVID-19 pandemic.