Intelligence of Dogs

Kimberly Cable

05-21-2022

Final Project - Step 2

How to import and clean my data

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Load and read each of the datasets:

- **dog_intelligence.csv**

```
##               Breed Classification obey reps_lower reps_upper
## 1     Border Collie Brightest Dogs  95%          1          4
## 2            Poodle Brightest Dogs  95%          1          4
## 3   German Shepherd Brightest Dogs  95%          1          4
## 4  Golden Retriever Brightest Dogs  95%          1          4
## 5 Doberman Pinscher Brightest Dogs  95%          1          4
## 6 Shetland Sheepdog Brightest Dogs  95%          1          4
```

- **AKC Breed Info.csv**

```
##                  Breed height_low_inches height_high_inches weight_low_lbs
## 1                Akita                26                 28             80
## 2   Anatolian Sheepdog                27                 29            100
## 3 Bernese Mountain Dog                23                 27             85
## 4           Bloodhound                24                 26             80
## 5               Borzoi                26                 28             70
## 6          Bullmastiff                25                 27            100
##   weight_high_lbs
## 1             120
## 2             150
## 3             110
```

```
## 4             120
## 5             100
## 6             130
```

- **Table_4_Heterozygosity_85_breeds.csv**

```
##                   Population Heterozygosity
## 1         Bedlington Terrier       0.312842
## 2     Miniature Bull Terrier       0.321619
## 3                      Boxer       0.343151
## 4            Clumber Spaniel       0.363595
## 5 Greater Swiss Mountain Dog       0.364943
## 6           Airedale Terrier       0.372793
```

- **Table_5_Expected_Heterozygosity_60_breeds.csv**

```
##                   Breed Heterozygosity_x10_4
## 1    Scottish Deerhound               2.0683
## 2         Field Spaniel               2.3165
## 3 Flat-coated Retriever               2.6474
## 4  Bernese Mountain Dog               2.8129
## 5    Standard Schnauzer               2.8129
## 6                 Boxer               3.0611
```

Create New Dataframe from the Intelligence data

```
##               Breed Classification obey reps_lower reps_upper
## 1     Border Collie Brightest Dogs  95%          1          4
## 2            Poodle Brightest Dogs  95%          1          4
## 3   German Shepherd Brightest Dogs  95%          1          4
## 4  Golden Retriever Brightest Dogs  95%          1          4
## 5 Doberman Pinscher Brightest Dogs  95%          1          4
## 6 Shetland Sheepdog Brightest Dogs  95%          1          4
```

Inner Join Breed data to new combined df on key Breed

```
##               Breed Classification obey reps_lower reps_upper
## 1      Border Collie Brightest Dogs  95%          1          4
## 2   Golden Retriever Brightest Dogs  95%          1          4
## 3  Doberman Pinscher Brightest Dogs  95%          1          4
## 4 Labrador Retriever Brightest Dogs  95%          1          4
## 5           Papillon Brightest Dogs  95%          1          4
## 6         Rottweiler Brightest Dogs  95%          1          4
##   height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## 1                19                 21             40              40
## 2                21                 24             55              75
## 3                26                 28             60             100
## 4                21                 24             55              80
## 5                 8                 11              5              10
## 6                22                 27             90             110
```

Inner Join Heterozygosity 4 to new combined df on key Breed = Population

```
##                Breed         Classification obey reps_lower reps_upper
## 1      Border Collie         Brightest Dogs  95%          1          4
## 2   Golden Retriever         Brightest Dogs  95%          1          4
## 3  Doberman Pinscher         Brightest Dogs  95%          1          4
## 4 Labrador Retriever         Brightest Dogs  95%          1          4
## 5         Rottweiler         Brightest Dogs  95%          1          4
## 6         Schipperke Excellent Working Dogs  85%          5         15
##   height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## 1                19                 21             40              40
## 2                21                 24             55              75
## 3                26                 28             60             100
## 4                21                 24             55              80
## 5                22                 27             90             110
## 6                10                 13             12              18
##   Heterozygosity
## 1       0.549583
## 2       0.517779
## 3       0.383763
## 4       0.560590
## 5       0.456510
## 6       0.445437
```

Inner Join Heterozygosity 5 to new combined df on key Breed

```
##                        Breed         Classification obey reps_lower reps_upper
## 1           Golden Retriever         Brightest Dogs  95%          1          4
## 2         Labrador Retriever         Brightest Dogs  95%          1          4
## 3                 Rottweiler         Brightest Dogs  95%          1          4
## 4 German Shorthaired Pointer Excellent Working Dogs  85%          5         15
## 5         Standard Schnauzer Excellent Working Dogs  85%          5         15
## 6       Bernese Mountain Dog Excellent Working Dogs  85%          5         15
##   height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## 1                21                 24             55              75
## 2                21                 24             55              80
## 3                22                 27             90             110
## 4                20                 27             50              80
## 5                17                 19             33              33
## 6                23                 27             85             110
##   Heterozygosity Heterozygosity_x10_4
## 1       0.517779               7.0323
## 2       0.560590               8.4388
## 3       0.456510               4.9640
## 4       0.538761               6.6186
## 5       0.450041               2.8129
## 6       0.399599               2.8129
```

Convert n/a or na to empty cell

Convert obey to numeric

Convert height and weight to numeric

What does the final data set look like?

```
##                        Breed         Classification obey reps_lower reps_upper
```

```
## 1           Golden Retriever         Brightest Dogs 0.95          1          4
## 2         Labrador Retriever         Brightest Dogs 0.95          1          4
## 3                Rottweiler         Brightest Dogs 0.95          1          4
## 4 German Shorthaired Pointer Excellent Working Dogs 0.85          5         15
## 5         Standard Schnauzer Excellent Working Dogs 0.85          5         15
## 6       Bernese Mountain Dog Excellent Working Dogs 0.85          5         15
##   height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## 1                21                 24             55              75
## 2                21                 24             55              80
## 3                22                 27             90             110
## 4                20                 27             50              80
## 5                17                 19             33              33
## 6                23                 27             85             110
##   Heterozygosity Heterozygosity_x10_4
## 1       0.517779               7.0323
## 2       0.560590               8.4388
## 3       0.456510               4.9640
## 4       0.538761               6.6186
## 5       0.450041               2.8129
## 6       0.399599               2.8129
```

What information is not self-evident?

- Initially I do not know exactly what Heterozygosity and Heterozygosity (x10-4) are and the difference between the two columns.

What are different ways you could look at this data?

One could strictly look at the obey percentage without looking at the number of reps a dog can do. You can also just look at the upper and lower reps versus taking the average number of reps a dog can do. Same problem with height and weight if I were to look at if intelligence is strictly by the weight of a breed or how tall a breed is.

How do you plan to slice and dice the data?

- **Add average weight and height to dataframe**

```
##                         Breed         Classification obey reps_lower reps_upper
## 1           Golden Retriever         Brightest Dogs 0.95          1          4
## 2         Labrador Retriever         Brightest Dogs 0.95          1          4
## 3                Rottweiler         Brightest Dogs 0.95          1          4
## 4 German Shorthaired Pointer Excellent Working Dogs 0.85          5         15
## 5         Standard Schnauzer Excellent Working Dogs 0.85          5         15
## 6       Bernese Mountain Dog Excellent Working Dogs 0.85          5         15
##   height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## 1                21                 24             55              75
## 2                21                 24             55              80
## 3                22                 27             90             110
## 4                20                 27             50              80
## 5                17                 19             33              33
## 6                23                 27             85             110
##   Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
```

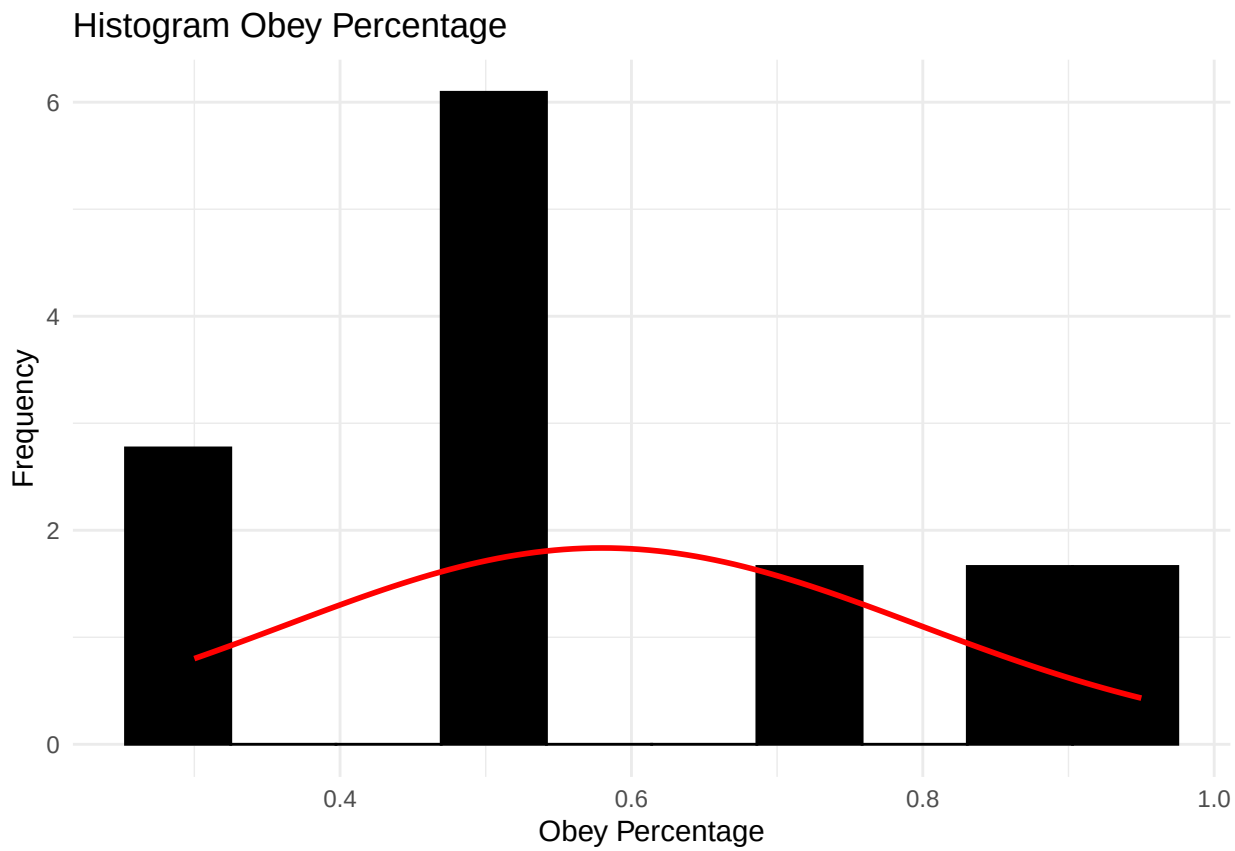
```
## 1       0.517779              7.0323       65.0       22.5
## 2       0.560590              8.4388       67.5       22.5
## 3       0.456510              4.9640      100.0       24.5
## 4       0.538761              6.6186       65.0       23.5
## 5       0.450041              2.8129       33.0       18.0
## 6       0.399599              2.8129       97.5       25.0
```

How could you summarize your data to answer key questions?

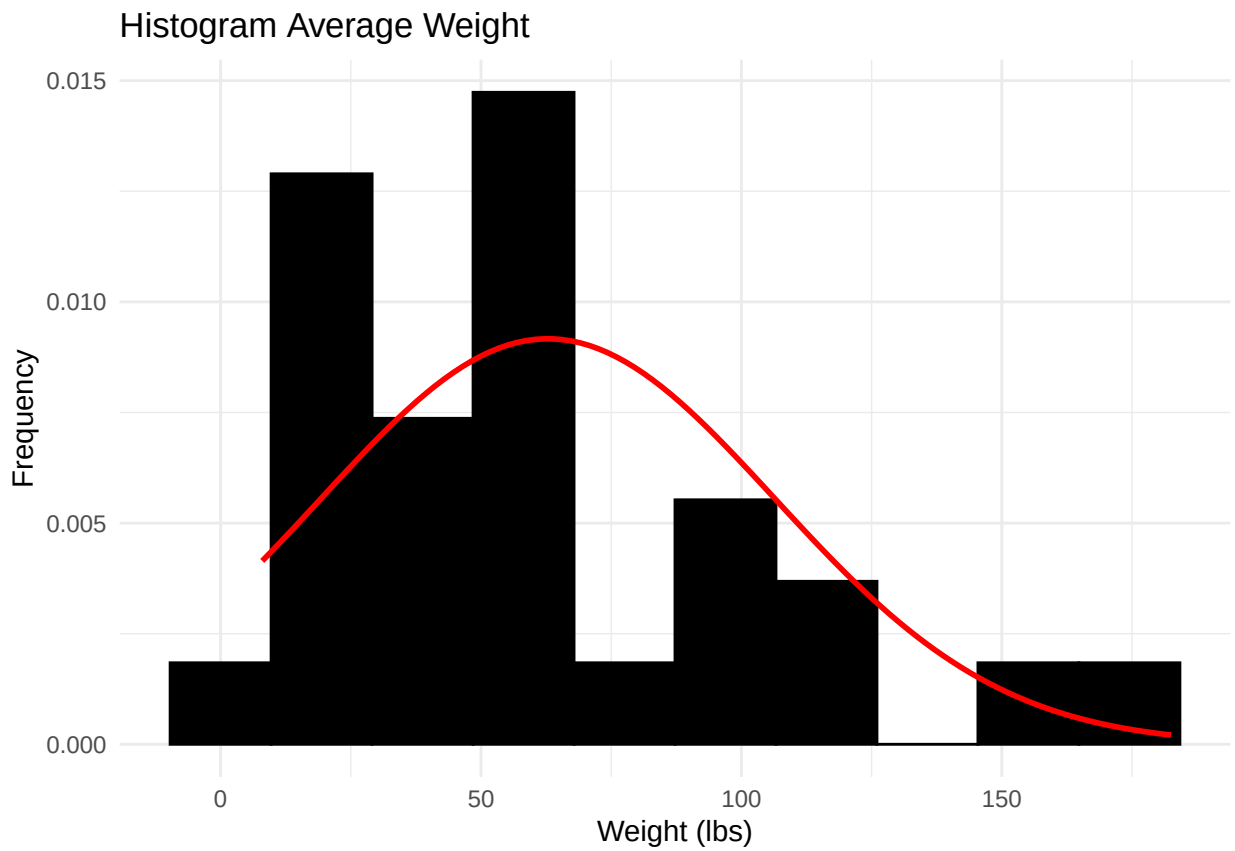
```
##     Breed           Classification          obey        reps_lower
##  Length:29          Length:29          Min.   :0.30   Min.   : 1.00
##  Class :character   Class :character   1st Qu.:0.50   1st Qu.:16.00
##  Mode  :character   Mode  :character   Median :0.50   Median :26.00
##                                        Mean   :0.58   Mean   :30.38
##                                        3rd Qu.:0.70   3rd Qu.:41.00
##                                        Max.   :0.95   Max.   :81.00
##                                        NA's   :4
##    reps_upper     height_low_inches height_high_inches weight_low_lbs
##  Min.   :  4.00   Min.   : 7.00     Min.   :10.00      Min.   :  6.00
##  1st Qu.: 25.00   1st Qu.:14.00     1st Qu.:16.00      1st Qu.: 19.50
##  Median : 40.00   Median :21.00     Median :24.50      Median : 46.00
##  Mean   : 47.31   Mean   :19.05     Mean   :22.12      Mean   : 53.04
##  3rd Qu.: 80.00   3rd Qu.:25.00     3rd Qu.:28.00      3rd Qu.: 72.50
##  Max.   :100.00   Max.   :27.00     Max.   :30.00      Max.   :175.00
##                   NA's   :1         NA's   :1          NA's   :1
##  weight_high_lbs  Heterozygosity   Heterozygosity_x10_4   avg.weight
##  Min.   : 10.00   Min.   :0.3128   Min.   :2.813        Min.   :  8.00
##  1st Qu.: 31.50   1st Qu.:0.4500   1st Qu.:4.550        1st Qu.: 24.75
##  Median : 70.00   Median :0.4879   Median :5.543        Median : 58.75
##  Mean   : 72.64   Mean   :0.4789   Mean   :5.312        Mean   : 62.84
##  3rd Qu.:102.50   3rd Qu.:0.5178   3rd Qu.:6.040        3rd Qu.: 88.12
##  Max.   :190.00   Max.   :0.5630   Max.   :8.439        Max.   :182.50
##  NA's   :1                                              NA's   :1
##    avg.height
##  Min.   : 8.50
##  1st Qu.:15.25
##  Median :22.75
##  Mean   :20.59
##  3rd Qu.:26.00
##  Max.   :28.50
##  NA's   :1
```

What types of plots and tables will help you illustrate the findings to your questions?

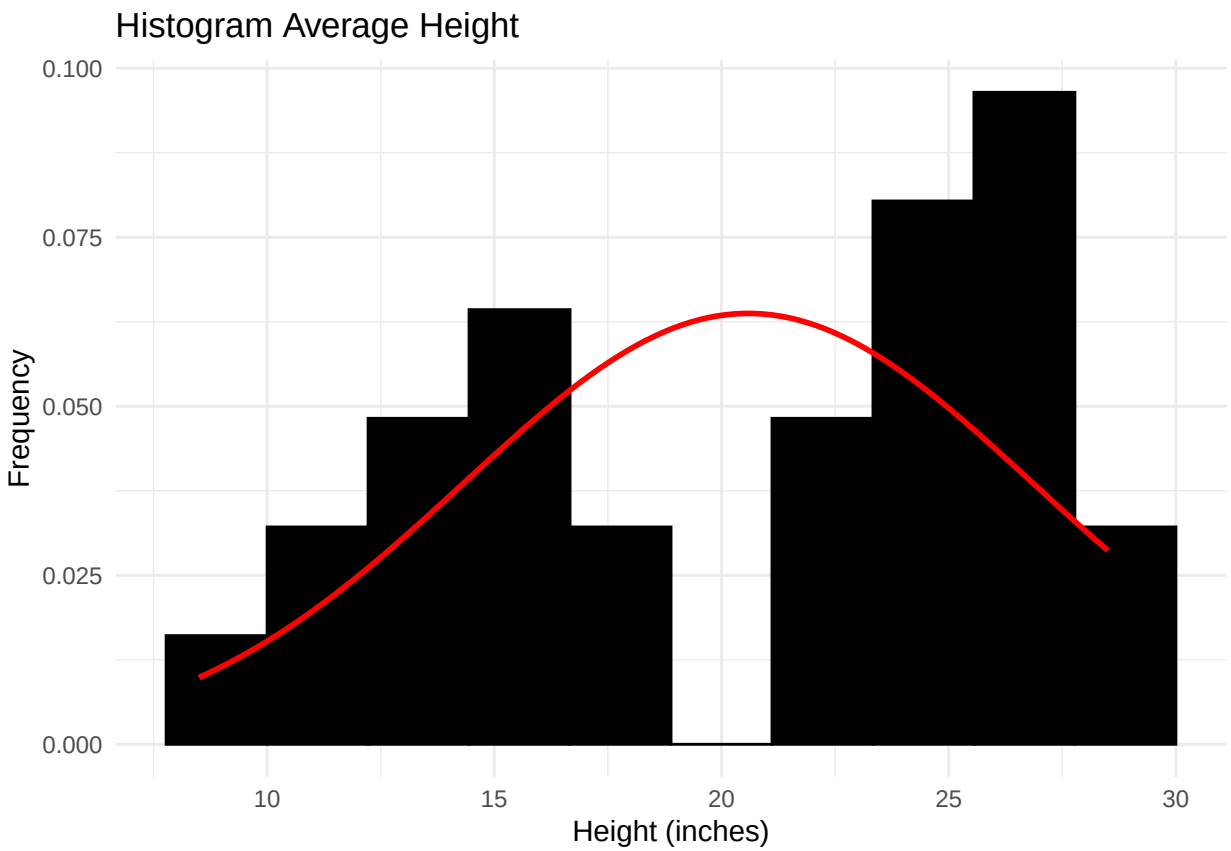
```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```

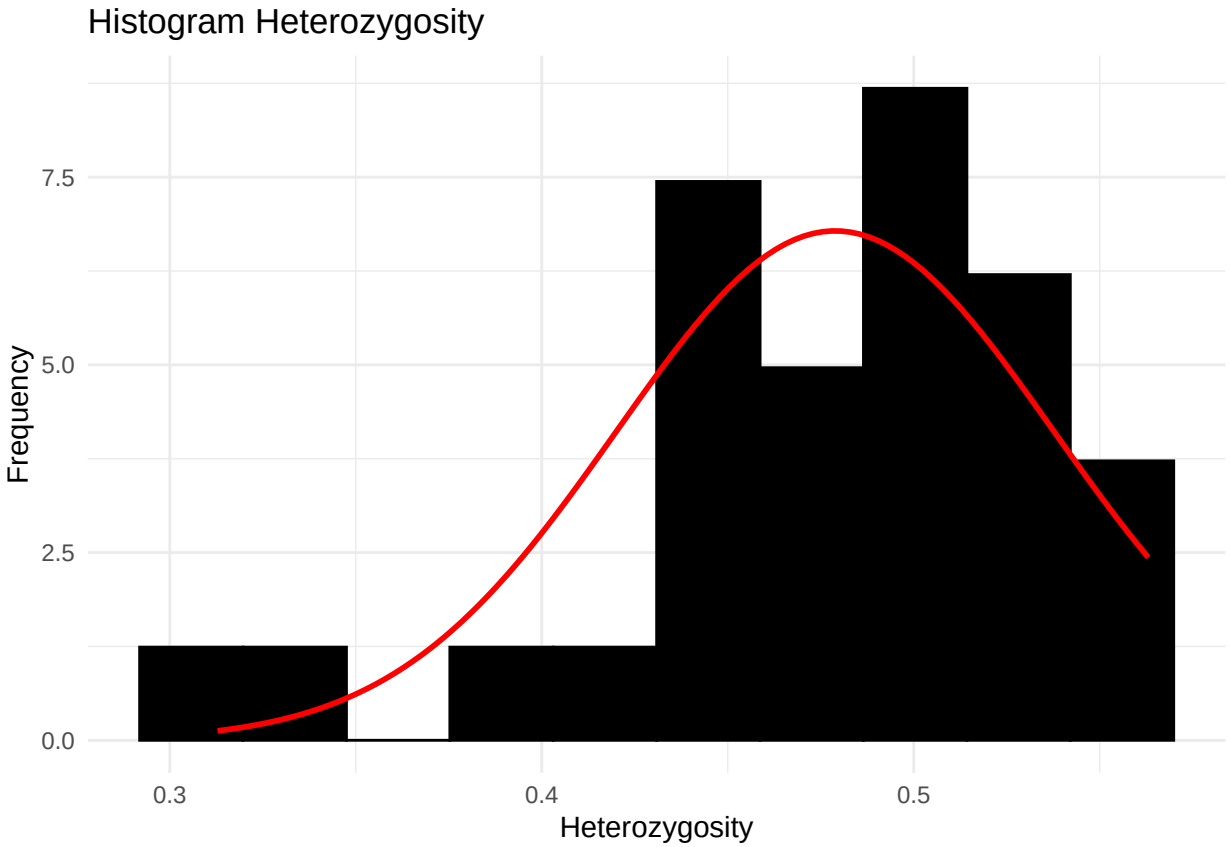


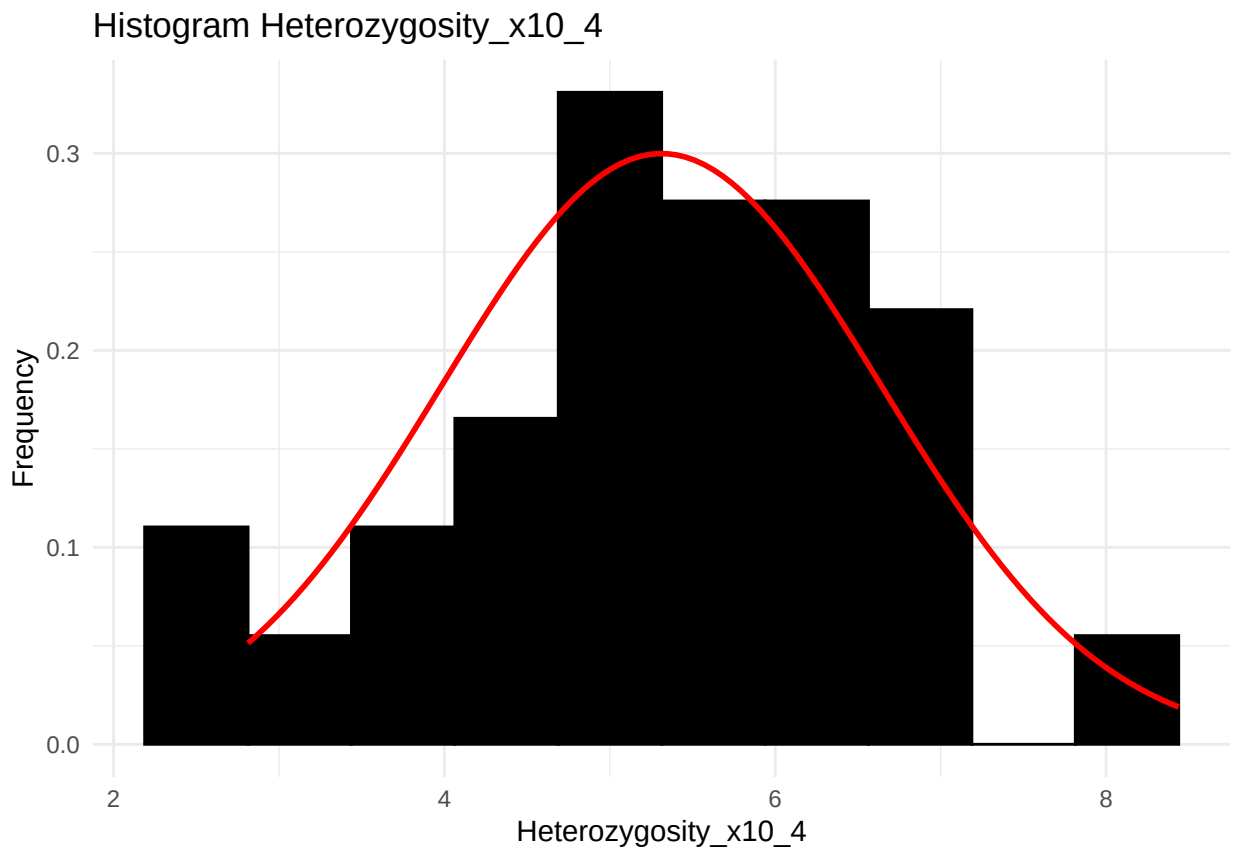

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```




```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```


Histogram Average Height
0.100
0.075
0.050
0.025
0.000
Frequency
10
15
20
25
30
Height (inches)


Histogram Heterozygosity
Frequency
7.5
5.0
2.5
0.0
0.3
0.4
0.5
Heterozygosity

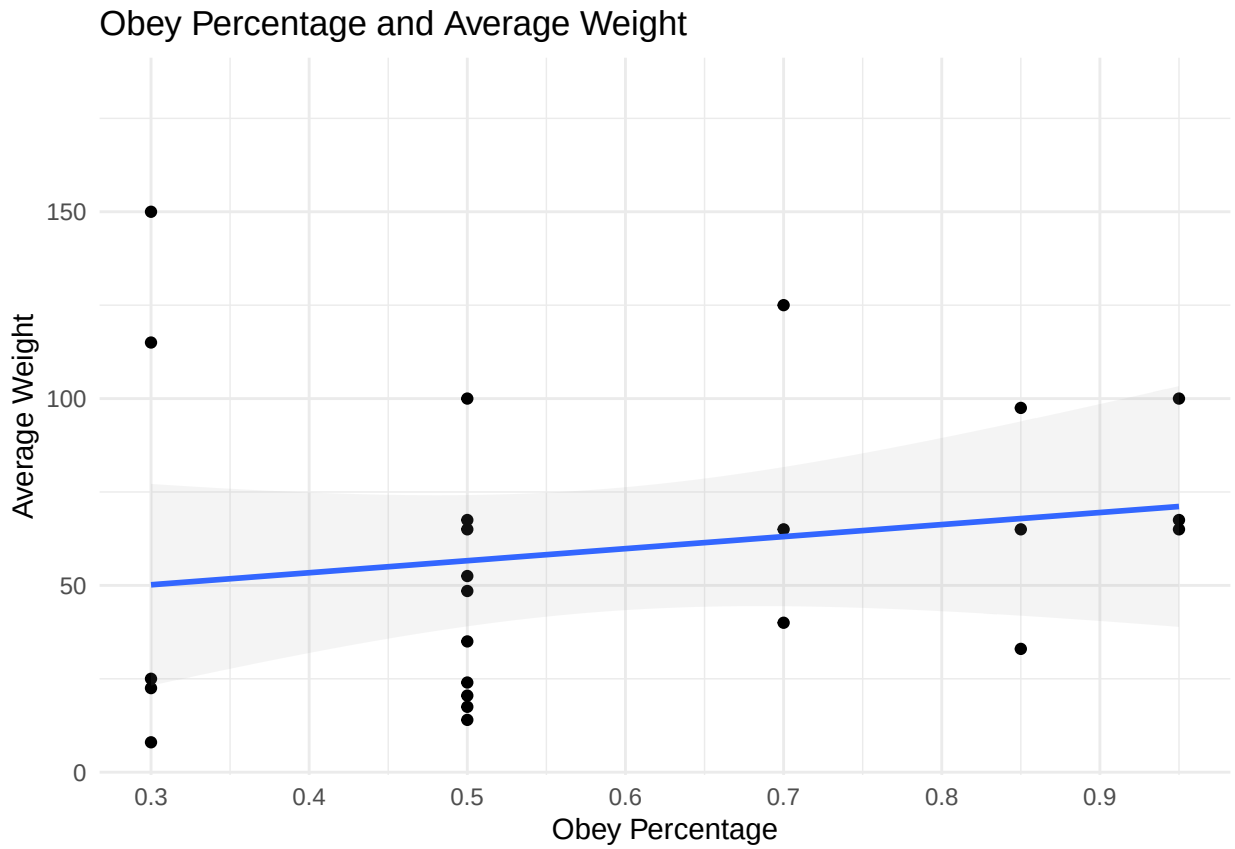


- **Scatter Plot of obey and avg.weight**

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 5 rows containing non-finite values (stat_smooth).

## Warning: Removed 5 rows containing missing values (geom_point).
```

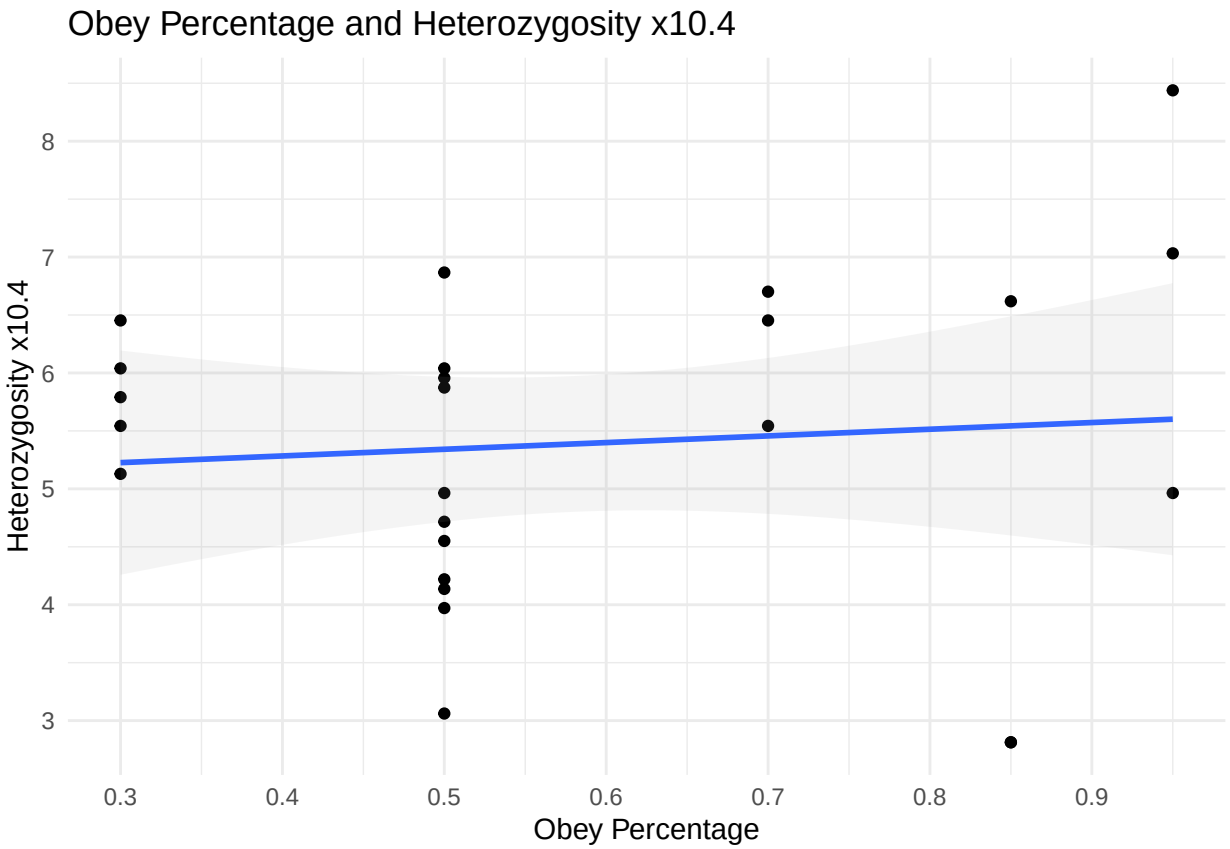


* **Scatter Plot of obey and Heterozygosity_x10_4**

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 4 rows containing non-finite values (stat_smooth).

## Warning: Removed 4 rows containing missing values (geom_point).
```



- **Correlation between obey percentage and avg.weight**

```
##
##  Pearson's product-moment correlation
##
## data:  combined_df$obey and combined_df$avg.weight
## t = 0.88343, df = 22, p-value = 0.3866
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2359190  0.5476023
## sample estimates:
##       cor
## 0.1850928
```

Since the correlation is 0.19 and the p-value is 0.39 we can say that the correlation between the two variables is not significant. Also, the intervals cross 0 so as one goes up the other goes up but then it is reversed.

- **Correlation between obey percentage and Heterozygosity_x10_4**

```
##
##  Pearson's product-moment correlation
##
## data:  combined_df$obey and combined_df$Heterozygosity_x10_4
## t = 0.43369, df = 23, p-value = 0.6686
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.3163255  0.4685203
## sample estimates:
##        cor
## 0.09006233
```

Since the correlation is 0.09 and the p-value is 0.66 we can say that the correlation between the two variables is not significant. Also, the intervals cross 0 so as one goes up the other goes up but then it is reversed.

- **Correlation between avg.weight and Heterozygosity_x10_4**

```
##
##  Pearson's product-moment correlation
##
## data:  combined_df$avg.weight and combined_df$Heterozygosity_x10_4
## t = -0.16629, df = 26, p-value = 0.8692
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4007977  0.3446736
## sample estimates:
##         cor
## -0.03259464
```

Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.

```
##
## Call:
## lm(formula = obey ~ avg.weight + avg.height + Heterozygosity_x10_4,
##     data = combined_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31265 -0.16426 -0.00432  0.14696  0.34899
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.2566674  0.2573047   0.998    0.330
## avg.weight           -0.0008206  0.0020733  -0.396    0.696
## avg.height            0.0142519  0.0129588   1.100    0.284
## Heterozygosity_x10_4  0.0155575  0.0340134   0.457    0.652
##
## Residual standard error: 0.2256 on 20 degrees of freedom
##   (5 observations deleted due to missingness)
## Multiple R-squared:  0.09806,    Adjusted R-squared:  -0.03723
## F-statistic: 0.7248 on 3 and 20 DF,  p-value: 0.549
```

Looking at the Adjusted R-squared of -0.37 and all p-values for the variables are not significant it does not look like any other the variables help with the percentage a dog can obey.

Questions for future steps.

More research would need to be done to find out if any other data can be linked to a dog's intelligence.