

# Intelligence of Dogs

Kimberly Cable

06-04-2022

## Step 1

### Introduction

We all love dogs. Some more than others but all in all, a dog is man's best friend. I think what we enjoy most about our furry companions is their ability to learn. From just being able to fetch a ball and bring it back to helping the police locate missing persons to alerting their companions that they are about to have a seizure. Dogs are amazing animals.

### Research Questions

Studies have been done on which breeds are more intelligent than others, but I was curious if how big a dog was had anything to do with how smart they were.

- Are bigger dogs smarter than smaller dogs?
- Does the classification of a dog really tell their intelligence level?
- Does their heterozygosity (diversity in the genes) have anything to do with their intelligence?
- Within a classification, do the larger breeds fair better than the smaller ones for intelligence?
- Does their heterozygosity influence the number of reps a dog can do?

### Approach

I plan to look at if height, weight and heterozygosity have any affect on how intelligent a dog is based on the percentage of times they can obey a command.

### How your approach addresses (fully or partially) the problem

With my approach I think it would partially answer whether or not how big a dog was and if it plays a part in the how smart they are.

### Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

- dog\_intelligence.csv (Fishman, n.d.b)
- Table\_4\_Heterozygosity\_85\_breeds.csv (Fishman, n.d.b)
- Table\_5\_Expected\_Heterozygosity\_60\_breeds.csv (Fishman, n.d.b)
- AKC Breed Info.csv (Fishman, n.d.a)

## Required Packages

- ggplot2
- dplyr
- magrittr
- Hmisc
- ggm

## Plots and Table Needs

- Histogram
- Scatter Plots
- CDF
- Linear Regression

## Questions for future steps

To begin, I suppose you look at histograms of the different variables and then decide how to proceed.

## 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

##
## Attaching package: 'purrr'

## The following object is masked from 'package:magrittr':
##
##   set_names
```

Load and read each of the datasets:

- dog\_intelligence.csv

```
intelligence_df <- read.csv("Final_Project/data/dog_intelligence.csv",
  header = TRUE, stringsAsFactors = FALSE)
head(intelligence_df)
```

```
##           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_df <- read.csv("Final_Project/data/AKC Breed Info.csv")
head(breed_df)
```

```
##           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

```
heterozygosity_4_df <- read.csv("Final_Project/data/Table_4_Heterozygosity_85_breeds.csv")
head(heterozygosity_4_df)
```

```
##           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

```
heterozygosity_5_df <- read.csv("Final_Project/data/Table_5_Expected_Heterozygosity_60_breeds.csv")
colnames(heterozygosity_5_df)[2] <- "Heterozygosity_x10_4"
head(heterozygosity_5_df)
```

```
##           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

```
combined_df <- intelligence_df
head(combined_df)
```

```
##           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

```
combined_df <- combined_df %>%
  inner_join(breed_df, by = c(Breed = "Breed"))
head(combined_df)
```

```
##           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

```
combined_df <- combined_df %>%
  inner_join(heterozygosity_4_df, by = c(Breed = "Population"))
head(combined_df)
```

```
##           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

```
combined_df <- combined_df %>%
  inner_join(heterozygosity_5_df, by = c(Breed = "Breed"))
head(combined_df)
```

```
## 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

```
combined_df[combined_df == "n/a"] <- ""
combined_df[combined_df == "na"] <- ""
```

### Convert obey to numeric

```
combined_df$obey <- gsub("%", "", as.character(combined_df$obey))  
  
combined_df$obey <- as.numeric(combined_df$obey)/100
```

### Convert height and weight to numeric

```
combined_df$height_low_inches <- as.numeric(combined_df$height_low_inches)  
combined_df$height_high_inches <- as.numeric(combined_df$height_high_inches)  
combined_df$weight_low_lbs <- as.numeric(combined_df$weight_low_lbs)  
combined_df$weight_high_lbs <- as.numeric(combined_df$weight_high_lbs)
```

### What does the final data set look like?

```
head(combined_df)
```

```
##           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

```
combined_df$avg.weight = rowMeans(combined_df[, c("weight_low_lbs",
  "weight_high_lbs")], na.rm = TRUE)
combined_df$avg.height = rowMeans(combined_df[, c("height_low_inches",
  "height_high_inches")], na.rm = TRUE)

head(combined_df)
```

```
##           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?

- Descriptive Statistics on all variables

```
summary(combined_df)
```

```
##      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
```

- Descriptive Statistics on all variables grouped by Classification

```
combined_df %>%
  split(.$Classification) %>%
  map(summary)
```

```
## $'Above Average Working Dogs'
## Breed Classification obey reps_lower reps_upper
## Length:3 Length:3 Min. :0.7 Min. :16 Min. :25
## Class :character Class :character 1st Qu.:0.7 1st Qu.:16 1st Qu.:25
## Mode :character Mode :character Median :0.7 Median :16 Median :25
## Mean :0.7 Mean :16 Mean :25
## 3rd Qu.:0.7 3rd Qu.:16 3rd Qu.:25
## Max. :0.7 Max. :16 Max. :25
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## Min. :16.00 Min. :19.00 Min. : 35.0 Min. : 45.00
## 1st Qu.:20.50 1st Qu.:23.00 1st Qu.: 47.5 1st Qu.: 57.50
## Median :25.00 Median :27.00 Median : 60.0 Median : 70.00
## Mean :22.33 Mean :24.67 Mean : 65.0 Mean : 88.33
## 3rd Qu.:25.50 3rd Qu.:27.50 3rd Qu.: 80.0 3rd Qu.:110.00
## Max. :26.00 Max. :28.00 Max. :100.0 Max. :150.00
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.4467 Min. :5.543 Min. : 40.00 Min. :17.50
## 1st Qu.:0.4603 1st Qu.:5.998 1st Qu.: 52.50 1st Qu.:21.75
## Median :0.4739 Median :6.453 Median : 65.00 Median :26.00
## Mean :0.4704 Mean :6.233 Mean : 76.67 Mean :23.50
## 3rd Qu.:0.4823 3rd Qu.:6.577 3rd Qu.: 95.00 3rd Qu.:26.50
## Max. :0.4906 Max. :6.701 Max. :125.00 Max. :27.00
##
## $'Average Working/Obedience Intelligence'
## Breed Classification obey reps_lower reps_upper
## Length:11 Length:11 Min. :0.5 Min. :26 Min. :40
## Class :character Class :character 1st Qu.:0.5 1st Qu.:26 1st Qu.:40
```



```

## Mode :character Mode :character Median :0.5 Median :26 Median :40
## Mean :0.5 Mean :26 Mean :40
## 3rd Qu.:0.5 3rd Qu.:26 3rd Qu.:40
## Max. :0.5 Max. :26 Max. :40
##
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## Min. : 7.00 Min. :10.00 Min. :10.0 Min. : 18.00
## 1st Qu.:11.25 1st Qu.:15.25 1st Qu.:16.5 1st Qu.: 25.25
## Median :18.00 Median :21.50 Median :30.0 Median : 50.00
## Mean :17.55 Mean :21.05 Mean :36.6 Mean : 52.30
## 3rd Qu.:22.75 3rd Qu.:28.00 3rd Qu.:55.5 3rd Qu.: 70.00
## Max. :27.00 Max. :30.00 Max. :80.0 Max. :120.00
## NA's :1 NA's :1 NA's :1 NA's :1
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.3128 Min. :3.061 Min. : 14.00 Min. : 8.50
## 1st Qu.:0.4557 1st Qu.:4.178 1st Qu.: 21.38 1st Qu.:13.25
## Median :0.5040 Median :4.716 Median : 41.75 Median :19.75
## Mean :0.4742 Mean :4.941 Mean : 44.45 Mean :19.30
## 3rd Qu.:0.5208 3rd Qu.:5.915 3rd Qu.: 61.88 3rd Qu.:25.50
## Max. :0.5630 Max. :6.867 Max. :100.00 Max. :28.50
## NA's :1 NA's :1
##
## $'Brightest Dogs'
## Breed Classification obey reps_lower reps_upper
## Length:3 Length:3 Min. :0.95 Min. :1 Min. :4
## Class :character Class :character 1st Qu.:0.95 1st Qu.:1 1st Qu.:4
## Mode :character Mode :character Median :0.95 Median :1 Median :4
## Mean :0.95 Mean :1 Mean :4
## 3rd Qu.:0.95 3rd Qu.:1 3rd Qu.:4
## Max. :0.95 Max. :1 Max. :4
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## Min. :21.00 Min. :24.0 Min. :55.00 Min. : 75.00
## 1st Qu.:21.00 1st Qu.:24.0 1st Qu.:55.00 1st Qu.: 77.50
## Median :21.00 Median :24.0 Median :55.00 Median : 80.00
## Mean :21.33 Mean :25.0 Mean :66.67 Mean : 88.33
## 3rd Qu.:21.50 3rd Qu.:25.5 3rd Qu.:72.50 3rd Qu.: 95.00
## Max. :22.00 Max. :27.0 Max. :90.00 Max. :110.00
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.4565 Min. :4.964 Min. : 65.00 Min. :22.50
## 1st Qu.:0.4871 1st Qu.:5.998 1st Qu.: 66.25 1st Qu.:22.50
## Median :0.5178 Median :7.032 Median : 67.50 Median :22.50
## Mean :0.5116 Mean :6.812 Mean : 77.50 Mean :23.17
## 3rd Qu.:0.5392 3rd Qu.:7.736 3rd Qu.: 83.75 3rd Qu.:23.50
## Max. :0.5606 Max. :8.439 Max. :100.00 Max. :24.50
##
## $'Excellent Working Dogs'
## Breed Classification obey reps_lower reps_upper
## Length:3 Length:3 Min. :0.85 Min. :5 Min. :15
## Class :character Class :character 1st Qu.:0.85 1st Qu.:5 1st Qu.:15
## Mode :character Mode :character Median :0.85 Median :5 Median :15
## Mean :0.85 Mean :5 Mean :15
## 3rd Qu.:0.85 3rd Qu.:5 3rd Qu.:15
## Max. :0.85 Max. :5 Max. :15
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs

```

```

## Min. :17.0      Min. :19.00      Min. :33.0      Min. : 33.00
## 1st Qu.:18.5     1st Qu.:23.00      1st Qu.:41.5     1st Qu.: 56.50
## Median :20.0     Median :27.00      Median :50.0     Median : 80.00
## Mean :20.0       Mean :24.33       Mean :56.0       Mean : 74.33
## 3rd Qu.:21.5     3rd Qu.:27.00      3rd Qu.:67.5     3rd Qu.: 95.00
## Max. :23.0       Max. :27.00       Max. :85.0       Max. :110.00
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.3996    Min. :2.813       Min. :33.00     Min. :18.00
## 1st Qu.:0.4248  1st Qu.:2.813     1st Qu.:49.00   1st Qu.:20.75
## Median :0.4500  Median :2.813     Median :65.00   Median :23.50
## Mean :0.4628    Mean :4.081       Mean :65.17     Mean :22.17
## 3rd Qu.:0.4944  3rd Qu.:4.716     3rd Qu.:81.25   3rd Qu.:24.25
## Max. :0.5388    Max. :6.619       Max. :97.50     Max. :25.00
##
## $'Fair Working/Obedience Intelligence'
## Breed Classification obey reps_lower reps_upper
## Length:5 Length:5 Min. :0.3 Min. :41 Min. :80
## Class :character Class :character 1st Qu.:0.3 1st Qu.:41 1st Qu.:80
## Mode :character Mode :character Median :0.3 Median :41 Median :80
## Mean :0.3 Mean :41 Mean :80
## 3rd Qu.:0.3 3rd Qu.:41 3rd Qu.:80
## Max. :0.3 Max. :41 Max. :80
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## Min. :11.0 Min. :12.0 Min. : 6.0 Min. : 10.0
## 1st Qu.:12.0 1st Qu.:15.0 1st Qu.: 17.0 1st Qu.: 28.0
## Median :14.0 Median :17.0 Median : 20.0 Median : 30.0
## Mean :17.4 Mean :19.8 Mean : 50.6 Mean : 77.6
## 3rd Qu.:25.0 3rd Qu.:27.0 3rd Qu.:100.0 3rd Qu.:130.0
## Max. :25.0 Max. :28.0 Max. :110.0 Max. :190.0
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.4399 Min. :5.129 Min. : 8.0 Min. :11.5
## 1st Qu.:0.4657 1st Qu.:5.543 1st Qu.: 22.5 1st Qu.:13.5
## Median :0.4688 Median :5.791 Median : 25.0 Median :15.5
## Mean :0.4806 Mean :5.791 Mean : 64.1 Mean :18.6
## 3rd Qu.:0.5092 3rd Qu.:6.040 3rd Qu.:115.0 3rd Qu.:26.0
## Max. :0.5195 Max. :6.453 Max. :150.0 Max. :26.5
##
## $'Lowest Degree of Working/Obedience Intelligence '
## Breed Classification obey reps_lower reps_upper
## Length:4 Length:4 Min. : NA Min. :81 Min. :100
## Class :character Class :character 1st Qu.: NA 1st Qu.:81 1st Qu.:100
## Mode :character Mode :character Median : NA Median :81 Median :100
## Mean :NaN Mean :81 Mean :100
## 3rd Qu.: NA 3rd Qu.:81 3rd Qu.:100
## Max. : NA Max. :81 Max. :100
## NA's :4
## height_low_inches height_high_inches weight_low_lbs weight_high_lbs
## Min. :13.00 Min. :14.0 Min. : 18.00 Min. : 30.0
## 1st Qu.:13.75 1st Qu.:15.5 1st Qu.: 34.50 1st Qu.: 45.0
## Median :20.00 Median :22.0 Median : 55.00 Median : 75.0
## Mean :20.00 Mean :22.0 Mean : 75.75 Mean : 92.5
## 3rd Qu.:26.25 3rd Qu.:28.5 3rd Qu.: 96.25 3rd Qu.:122.5
## Max. :27.00 Max. :30.0 Max. :175.00 Max. :190.0
##

```

```
## Heterozygosity Heterozygosity_x10_4 avg.weight avg.height
## Min. :0.4412 Min. :3.806 Min. : 24.00 Min. :14.00
## 1st Qu.:0.4516 1st Qu.:4.550 1st Qu.: 39.75 1st Qu.:14.38
## Median :0.4715 Median :4.881 Median : 65.00 Median :20.75
## Mean :0.4833 Mean :4.840 Mean : 84.12 Mean :21.00
## 3rd Qu.:0.5032 3rd Qu.:5.171 3rd Qu.:109.38 3rd Qu.:27.38
## Max. :0.5491 Max. :5.791 Max. :182.50 Max. :28.50
##
```

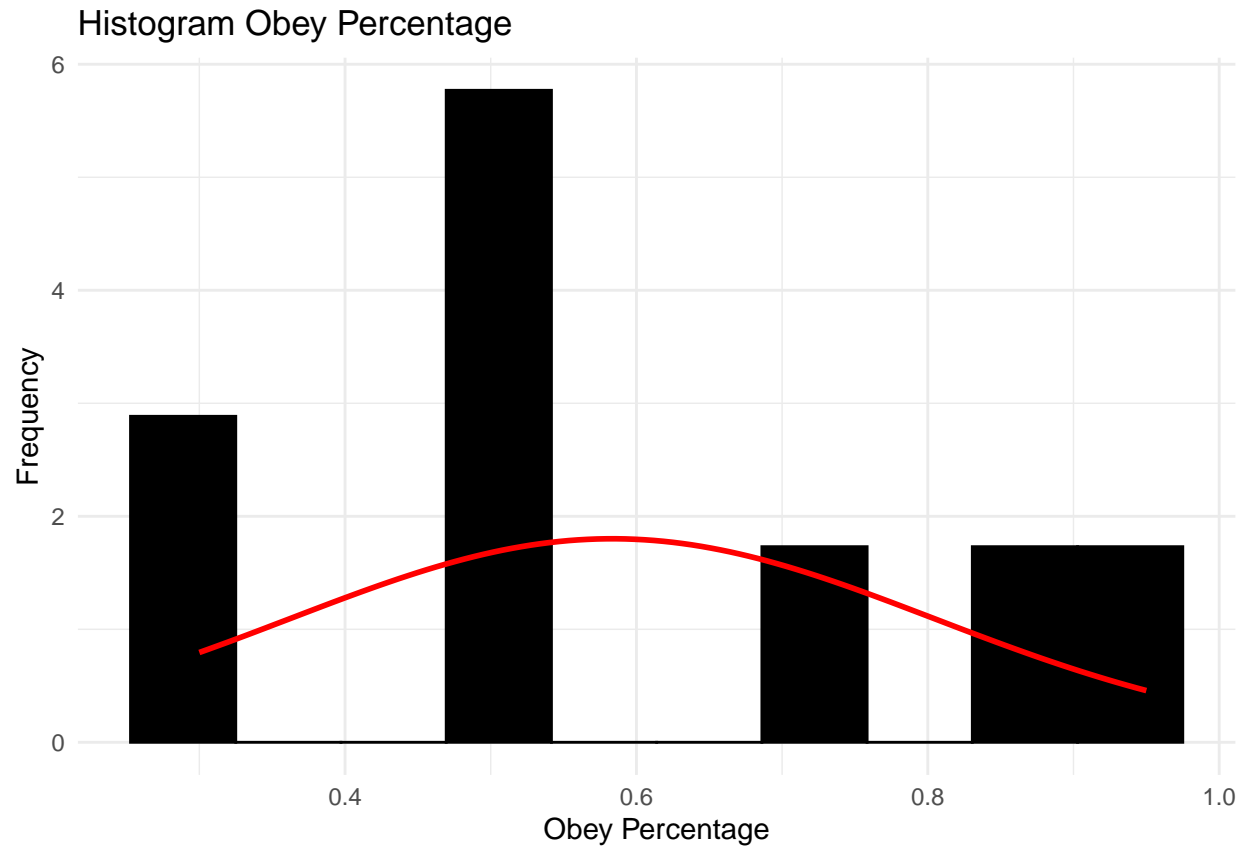
- Remove empty cells from variables for plots

```
combined_complete <- combined_df[complete.cases(combined_df),
]
head(combined_complete)
```

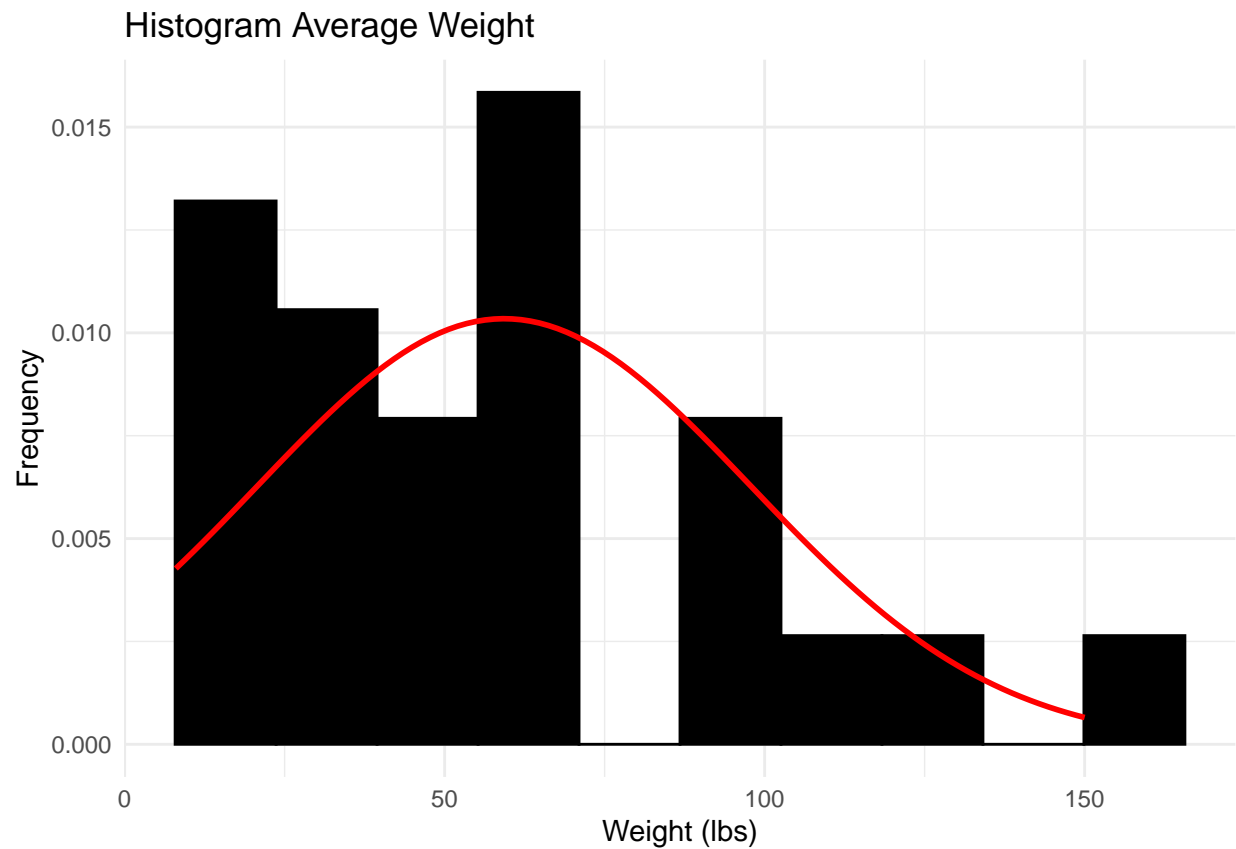
```
## Breed Classification obey reps_lower reps_upper
## 1 Golden Retriever Brightest Dogs 0.95 1 4
## 2 Labrador Retriever Brightest Dogs 0.95 1 4
## 3 Rottweiler Brightest Dogs 0.95 1 4
## 4 German Shorthaired Pointer Excellent Working Dogs 0.85 5 15
## 5 Standard Schnauzer Excellent Working Dogs 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
```

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

```
ggplot(combined_complete, aes(obey)) + labs(title = "Histogram Obey Percentage",
x = "Obey Percentage", y = "Frequency") + geom_histogram(bins = 10,
aes(y = ..density..), color = "black", fill = "black") +
stat_function(fun = dnorm, args = list(mean = mean(combined_complete$obey,
na.rm = TRUE), sd = sd(combined_complete$obey, na.rm = TRUE)),
color = "red", size = 1)
```

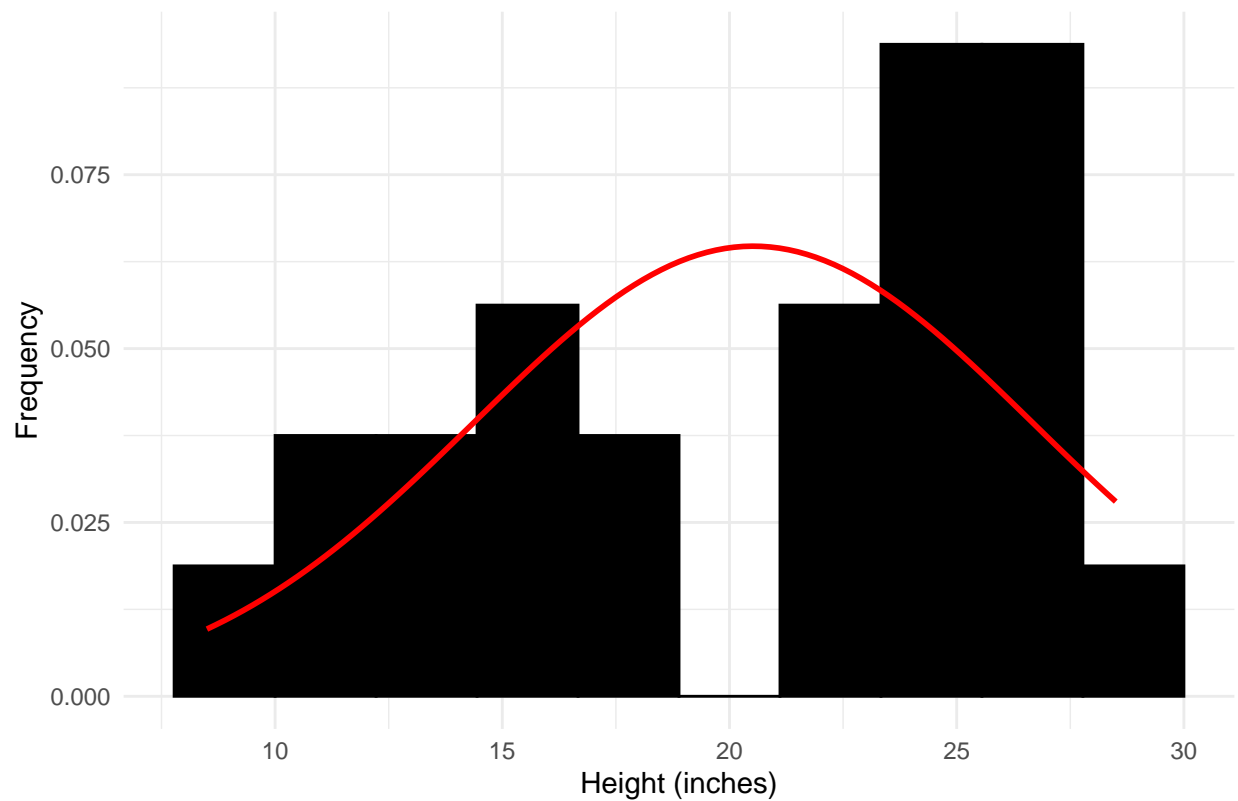


```
ggplot(combined_complete, aes(avg.weight)) + labs(title = "Histogram Average Weight",  
  x = "Weight (lbs)", y = "Frequency") + geom_histogram(bins = 10,  
  aes(y = ..density..), color = "black", fill = "black") +  
  stat_function(fun = dnorm, args = list(mean = mean(combined_complete$avg.weight,  
    na.rm = TRUE), sd = sd(combined_complete$avg.weight,  
    na.rm = TRUE)), color = "red", size = 1)
```



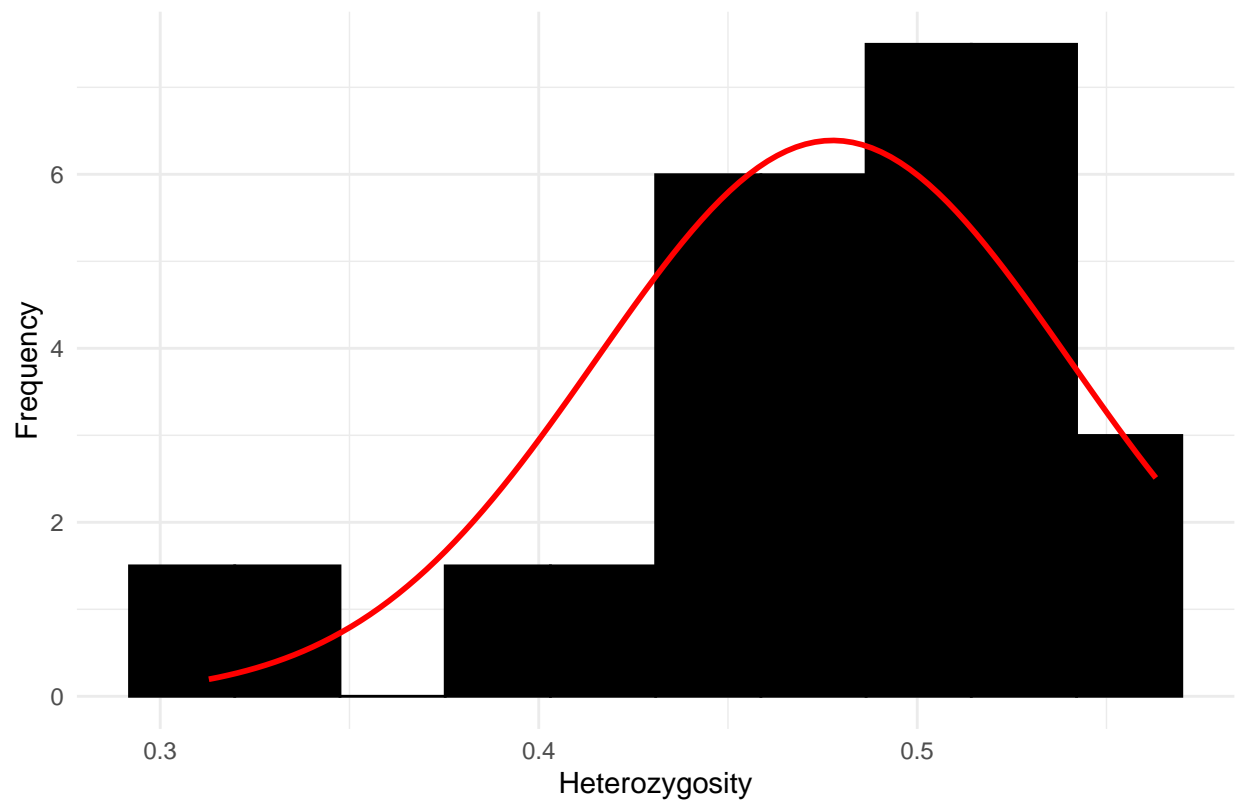
```
ggplot(combined_complete, aes(avg.height)) + labs(title = "Histogram Average Height",  
  x = "Height (inches)", y = "Frequency") + geom_histogram(bins = 10,  
  aes(y = ..density..), color = "black", fill = "black") +  
  stat_function(fun = dnorm, args = list(mean = mean(combined_complete$avg.height,  
    na.rm = TRUE), sd = sd(combined_complete$avg.height,  
    na.rm = TRUE)), color = "red", size = 1)
```

Histogram Average Height

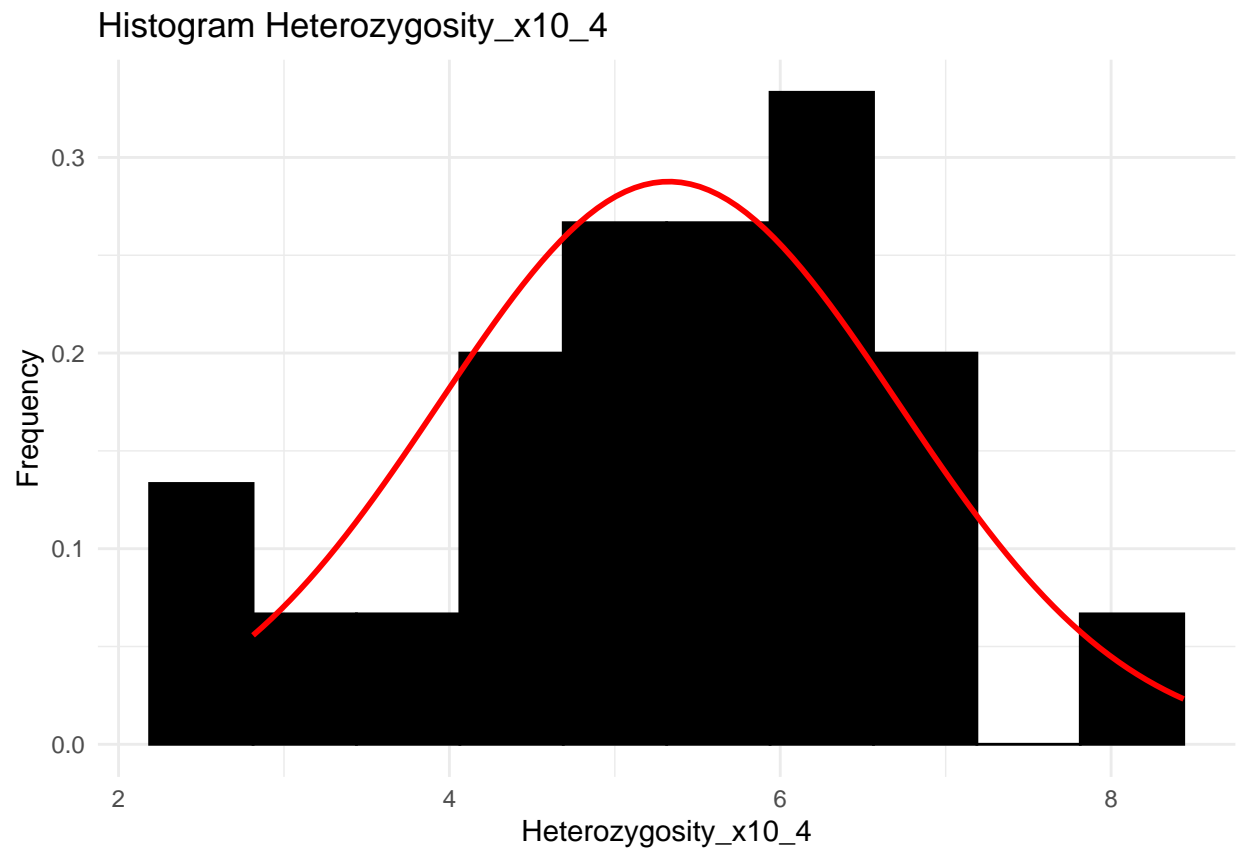


```
ggplot(combined_complete, aes(Heterozygosity)) + labs(title = "Histogram Heterozygosity",  
  x = "Heterozygosity", y = "Frequency") + geom_histogram(bins = 10,  
  aes(y = ..density..), color = "black", fill = "black") +  
  stat_function(fun = dnorm, args = list(mean = mean(combined_complete$Heterozygosity,  
    na.rm = TRUE), sd = sd(combined_complete$Heterozygosity,  
    na.rm = TRUE)), color = "red", size = 1)
```

Histogram Heterozygosity



```
ggplot(combined_complete, aes(Heterozygosity_x10_4)) + labs(title = "Histogram Heterozygosity_x10_4",  
  x = "Heterozygosity_x10_4", y = "Frequency") + geom_histogram(bins = 10,  
  aes(y = ..density..), color = "black", fill = "black") +  
  stat_function(fun = dnorm, args = list(mean = mean(combined_complete$Heterozygosity_x10_4,  
    na.rm = TRUE), sd = sd(combined_complete$Heterozygosity_x10_4,  
    na.rm = TRUE)), color = "red", size = 1)
```

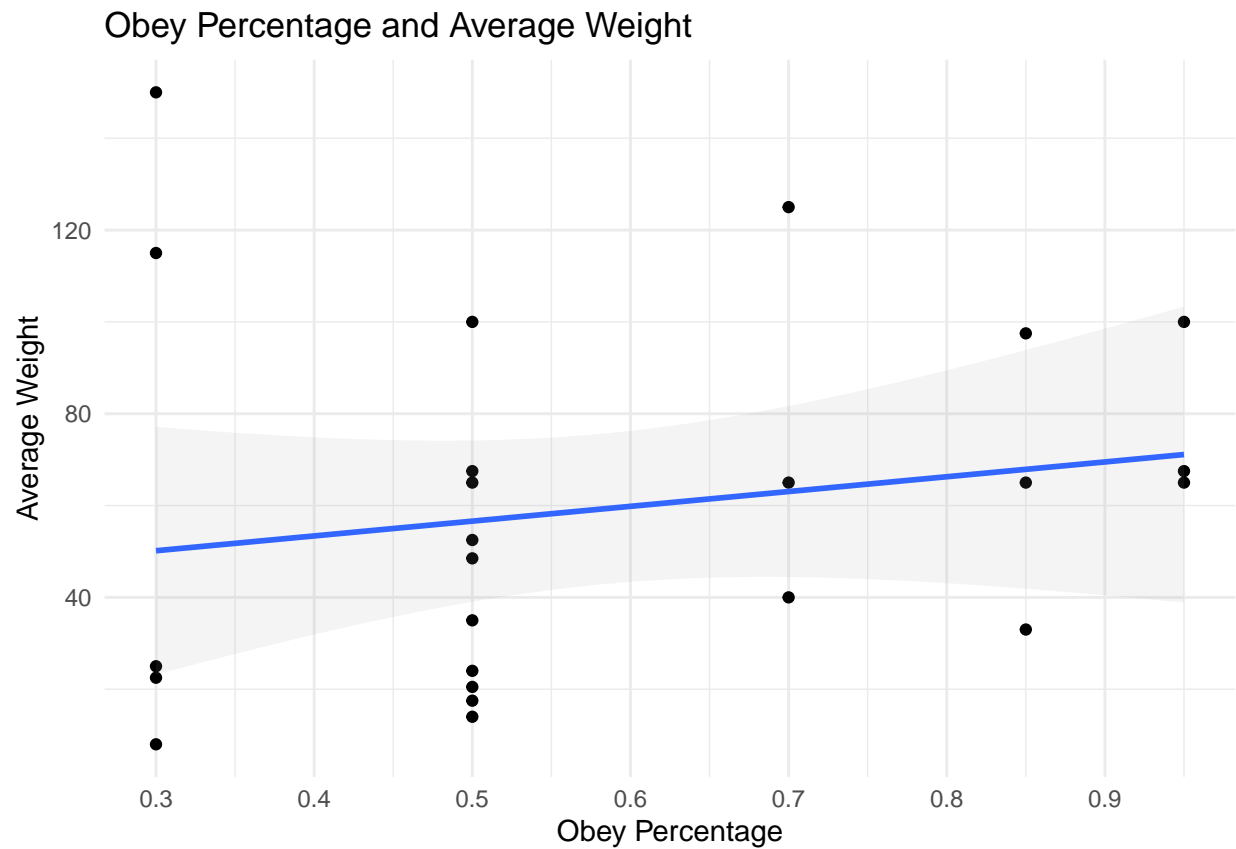


- Scatter Plot of obey and avg.weight

```
scatter <- ggplot(combined_complete, aes(obey, avg.weight))
scatter + geom_point() + scale_x_continuous(n.breaks = 10) +
  geom_smooth(method = "lm", alpha = 0.1) + labs(x = "Obey Percentage",
  y = "Average Weight") + ggtitle("Obey Percentage and Average Weight")
```

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

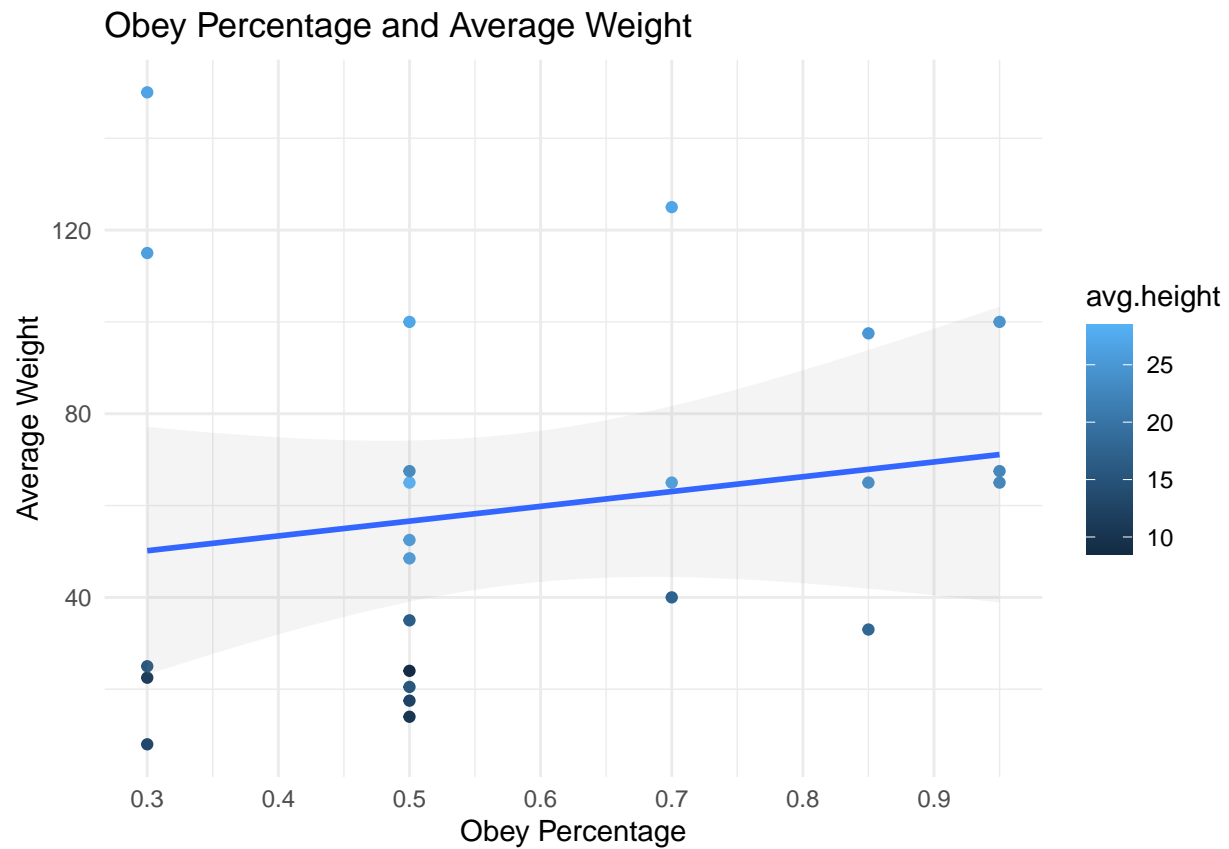




- Scatter Plot of obey and avg.weight colored by height

```
scatter <- ggplot(combined_complete, aes(obey, avg.weight, col = avg.height))
scatter + geom_point() + scale_x_continuous(n.breaks = 10) +
  geom_smooth(method = "lm", alpha = 0.1) + labs(x = "Obey Percentage",
  y = "Average Weight") + ggtitle("Obey Percentage and Average Weight")
```

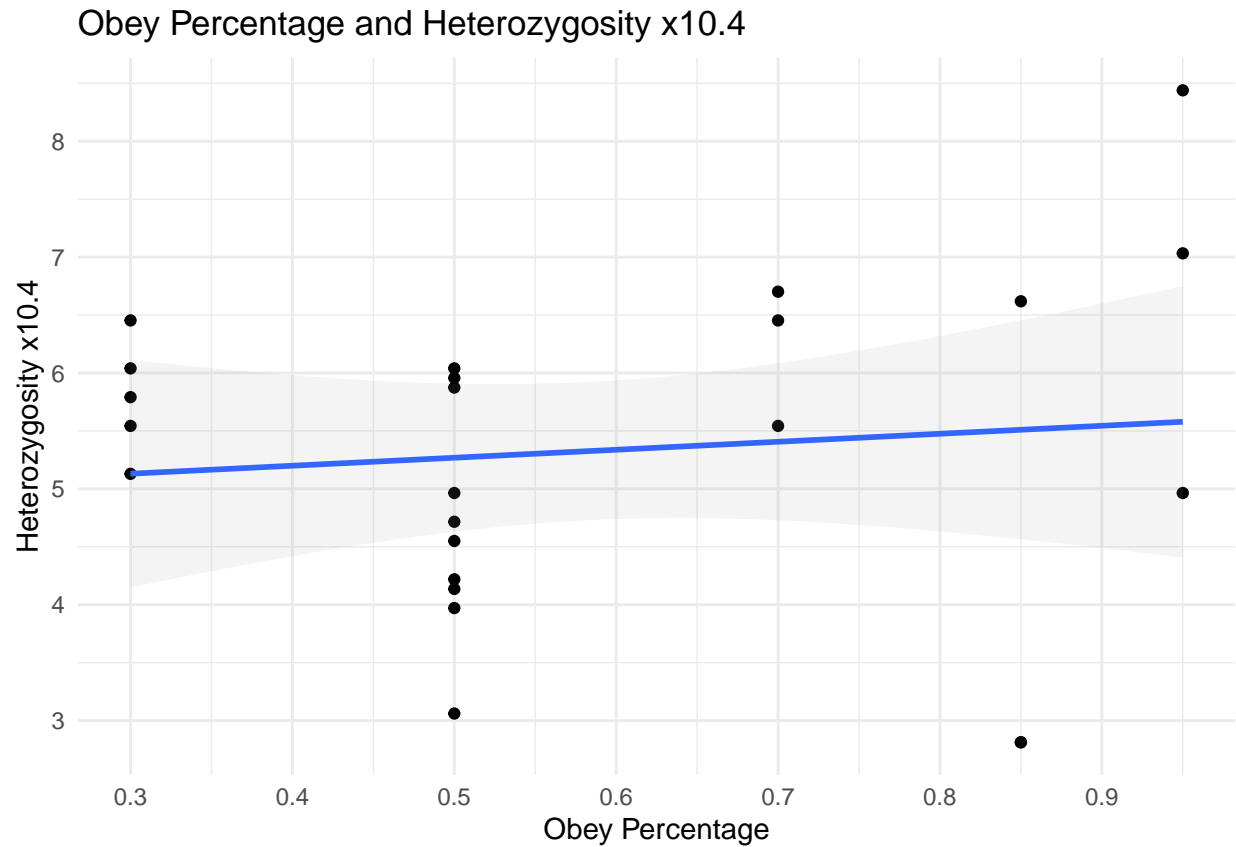
```
## 'geom_smooth()' using formula 'y ~ x'
```



- Scatter Plot of obey and Heterozygosity\_x10\_4

```
scatter <- ggplot(combined_complete, aes(obey, Heterozygosity_x10_4))
scatter + geom_point() + scale_x_continuous(n.breaks = 10) +
  geom_smooth(method = "lm", alpha = 0.1) + labs(x = "Obey Percentage",
  y = "Heterozygosity x10.4") + ggtitle("Obey Percentage and Heterozygosity x10.4")
```

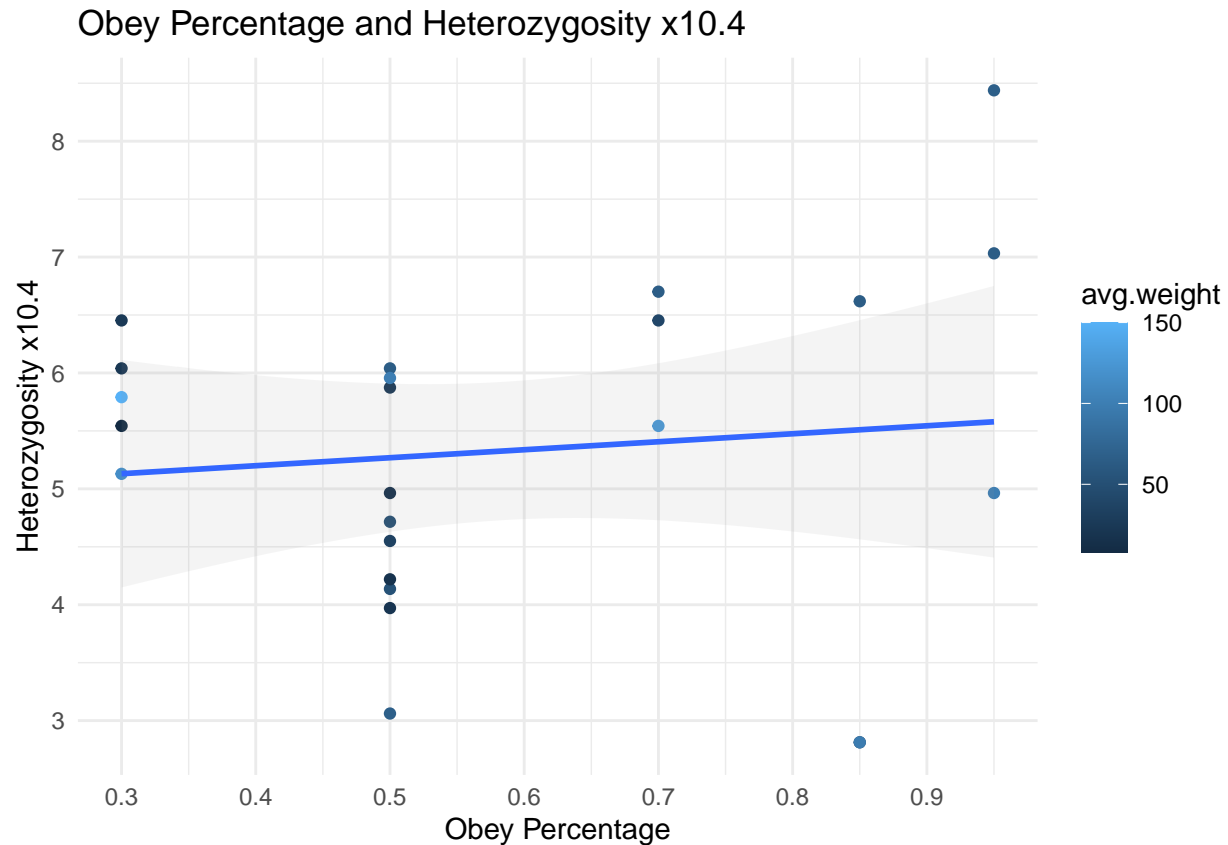
```
## 'geom_smooth()' using formula 'y ~ x'
```



- Scatter Plot of obey and Heterozygosity\_x10\_4 colored by average weight

```
scatter <- ggplot(combined_complete, aes(obey, Heterozygosity_x10_4,
  col = avg.weight))
scatter + geom_point() + scale_x_continuous(n.breaks = 10) +
  geom_smooth(method = "lm", alpha = 0.1) + labs(x = "Obey Percentage",
  y = "Heterozygosity x10.4") + ggtitle("Obey Percentage and Heterozygosity x10.4")
```

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



- Correlation between obey percentage and avg.weight

```
cor.test(combined_df$obey, combined_df$avg.weight, use = "complete.obs")
```

```
##
## 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

```
cor.test(combined_df$obey, combined_df$Heterozygosity_x10_4,
         use = "complete.obs")
```

```
##
## 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

```
cor.test(combined_df$avg.weight, combined_df$Heterozygosity_x10_4,
         use = "complete.obs")
```

```
##
## 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
```

- Correlation between all variables

```
cor(combined_df[, unlist(lapply(combined_df, is.numeric))], use = "complete.obs")
```

```
##              obey  reps_lower  reps_upper height_low_inches
## obey          1.00000000 -0.99564934 -0.95313734      0.26153249
## reps_lower    -0.99564934  1.00000000  0.97197530     -0.24849600
## reps_upper    -0.95313734  0.97197530  1.00000000     -0.23119865
## height_low_inches  0.26153249 -0.24849600 -0.23119865      1.00000000
## height_high_inches  0.30637778 -0.30054392 -0.28580559      0.96909760
## weight_low_lbs     0.23947951 -0.21170210 -0.15476993      0.84736129
## weight_high_lbs    0.14477026 -0.11696117 -0.06131463      0.79255876
## Heterozygosity     0.07193455 -0.05977593 -0.04541667      0.13287772
## Heterozygosity_x10_4 0.11007240 -0.05832246 -0.01860510      0.07461812
## avg.weight        0.18509278 -0.15703279 -0.10032941      0.82261801
## avg.height        0.28690276 -0.27751125 -0.26141002      0.99173281
##
## height_high_inches weight_low_lbs weight_high_lbs
## obey              0.30637778   0.239479506   0.14477026
## reps_lower        -0.30054392   -0.211702105   -0.11696117
## reps_upper        -0.28580559   -0.154769935   -0.06131463
```

```
## height_low_inches      0.96909760    0.847361295    0.79255876
## height_high_inches     1.00000000    0.804873046    0.75482370
## weight_low_lbs         0.80487305    1.000000000    0.96124005
## weight_high_lbs        0.75482370    0.961240047    1.00000000
## Heterozygosity         0.17900053    0.004482911    0.11759281
## Heterozygosity_x10_4   0.04436231    0.034970172    0.10380218
## avg.weight             0.78257009    0.986302832    0.99355113
## avg.height             0.99273957    0.831875415    0.77911756
##
## Heterozygosity Heterozygosity_x10_4 avg.weight
## obey          0.071934550          0.11007240 0.18509278
## reps_lower    -0.059775932         -0.05832246 -0.15703279
## reps_upper    -0.045416668         -0.01860510 -0.10032941
## height_low_inches 0.132877719        0.07461812 0.82261801
## height_high_inches 0.179000532        0.04436231 0.78257009
## weight_low_lbs   0.004482911        0.03497017 0.98630283
## weight_high_lbs  0.117592808        0.10380218 0.99355113
## Heterozygosity   1.000000000        0.54710881 0.07219321
## Heterozygosity_x10_4 0.547108806        1.00000000 0.07648067
## avg.weight       0.072193207        0.07648067 1.00000000
## avg.height       0.157908013        0.05946187 0.80820838
##
## avg.height
## obey          0.28690276
## reps_lower    -0.27751125
## reps_upper    -0.26141002
## height_low_inches 0.99173281
## height_high_inches 0.99273957
## weight_low_lbs   0.83187542
## weight_high_lbs  0.77911756
## Heterozygosity   0.15790801
## Heterozygosity_x10_4 0.05946187
## avg.weight       0.80820838
## avg.height       1.00000000
```

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

```
combined_model <- lm(obey ~ avg.weight + avg.height + Heterozygosity_x10_4,
  data = combined_df)

summary(combined_model)
```

```
##
## 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.

## Step 3

### Introduction

We all love dogs. Some more than others but all in all, a dog is man's best friend. I think what we enjoy most about our furry companions is their ability to learn. From just being able to fetch a ball and bring it back to help the police locate missing persons to alerting their companions that they are about to have a seizure. Dogs are amazing animals.

### The problem statement you addressed

With this analysis, I wanted to determine if a dog's size really determines its intelligence level. Are big dogs smarter than smaller dogs? Are they faster at learning commands or is it based on something else?

How I addressed the problem

1. I looked at 4 datasets to help me determine if the size of a dog determines its intelligence level.
  - a. AKC Breed Information: This dataset contained information such as breed, height and weight
  - b. Dog Intelligence: This dataset contained information such as breed, classification of intelligence, obey percentage, and upper and lower repetitions a dog could handle
  - c. Heterozygosity of 85 breeds: This dataset contained information about the breed and heterozygosity
  - d. Heterozygosity x10\_4 of 60 breeds: This dataset contained information about the breed and heterozygosity x10\_4.
2. I combined all 4 datasets into 1 and began to clean up the data and create new variables for average weight and average height.
3. Did some descriptive statistics
  - a. Obey, Avg Weight, and Avg Height

	Obey	Avg Weight	Avg Height
Min	0.30	8.00	8.5
Median	0.50	58.75	22.75
Mean	0.58	62.84	20.59
Max	0.95	182.50	28.5

b. By Classification: Obey, Avg Weight, and Avg Height

Brightest Dogs

	Obey	Avg Weight	Avg Height
Min	0.95	65.00	22.50
Median	0.95	67.5	22.50
Mean	0.95	77.5	23.17
Max	0.95	83.75	24.50

Excellent Working Dogs

	Obey	Avg Weight	Avg Height
Min	0.85	33.0	18.00
Median	0.85	65.0	23.50
Mean	0.85	65.17	22.17
Max	0.85	97.5	25.00

Above Avg Working Dog

	Obey	Avg Weight	Avg Height
Min	0.70	40.0	17.5
Median	0.70	65.0	26.0
Mean	0.70	76.67	23.5
Max	0.70	95.0	27.0

Average Working/Obedience Intelligence

	Obey	Avg Weight	Avg Height
Min	0.50	14.0	8.5
Median	0.50	41.75	19.75
Mean	0.50	44.45	25.5
Max	0.50	100.0	28.5

Fair Working/Obedience Intelligence

	Obey	Avg Weight	Avg Height
Min	0.30	8.0	11.5
Median	0.30	25.0	15.5



	Obey	Avg Weight	Avg Height
Mean	0.30	64.1	18.6
Max	0.30	150.0	26.5

Lowest Degree of Working/Obedience Intelligence

	Obey	Avg Weight	Avg Height
Min	NA	24.0	14.0
Median	NA	65.0	20.75
Mean	NA	84.12	21.0
Max	NA	182.5	28.5

c. Histograms of obey percentage, average weight, average height, Heterozygosity, and Heterozygosity x104

d. Scatter plots along of obey percentage vs average weight and obey percentage vs Heterozygosity x10.4

e. Correlations:

- i. Obey Percentage vs Average Weight:  
Correlation: 0.185; p-value: 0.387
- ii. Obey Percentage vs Heterozygosity x10.4:  
Correlation: 0.09; p-value: 0.669
- iii. Average Weight vs Heterozygosity x10.4:  
Correlation: -0.32; p-value: 0.869

#### 4. Regression Models

Obey Percentage with Average Weight + Average height + Heterozygosity x10.4:

- R squared: 0.08
- Adj R squared: -0.037
- F-statistic: 0.725
- P-value: 0.549

## Analysis

My preliminary analysis on whether the size of a dog determines its intelligence is as follows:

1. Looking at the mean of each breed and their classification suggests dogs of all weights can fit into each classification of brightest down to lowest.
2. The scatterplot of Obey Percentage and Average Weight does show a slight upwards trend between a dog's obey percentage and their average weight suggesting maybe their weight does play some role in their intelligence.
3. The correlations between the variables all suggested very low relationships between the various variables suggesting that the dog's obey percentage isn't influenced by its size or Heterozygosity.
4. Running a multiple linear regression model to see if average height, average weight, and Heterozygosity x10\_4 have any influence on the obey percentage of a dog suggests that all variables do not influence the obey percentage very much if at all.

## Implications

As the size of a dog does not seem to influence how smart they are I do not see any evil scientist manipulating dog genetics to increase the size of dogs to make them smarter.

## Concluding Remarks

With the limited initial research on if size influences a dog's intelligence, more research would need to be done to find out what, if any, genetics or factors lead to a breeds intelligence level.

## References

Fishman, L. n.d.a. "Dog/Canine Breed Size (AKC)." <https://data.world/len/dog-canine-breed-size-akc>.  
———. n.d.b. "Intelligence of Dogs." <https://data.world/len/intelligence-of-dogs>.