Examining a Relationship Between the Size and Intelligence of Dog Breeds

Julia Beilke
COSC 4931: Intro to Data Science
Dr. Shion Guha, Marquette University

May 2017

1 Introduction

1.1 Project Topic

In this project, I will examine a possible link between dog size and intelligence. I have found two datasets to use for this project. The first includes information about the sizes of 150 distinct dog breeds, including a low and high numeric value for dog height, in inches, and weight, in pounds. The second dataset provides information about the intelligence of 136 distinct dog breeds. In this dataset, the breeds are given a categorical value entitled classification. The classification refers to the dogs' "working/obedience intelligence," the highest value being "brightest dogs" and the lowest value being "lowest degree of working/obedience intelligence."

1.2 Methods for Rating Intelligence

The degree of intelligence for each breed is based upon obedience trials performed by the American Kennel Club and the Canadian Kennel Club per the request of Stanley Coren, a "professor of canine psychology at the University of British Columbia"(2) for his book The Intelligence of Dogs. In these evaluations, the dogs were taught a new command. The low and high number of repetitions it took for dogs of a certain breed to understand a new command, and the probability that they would obey the first command are provided in the dataset as reps_lower, reps_upper, and obey, respectively. As an example of how these values are used to classify dogs, a breed falls into the "Brightest Dogs" category if dogs have a reps_upper value less than or equal to 5, and obey the first command “95% of the time or better” (2). From there, the intelligence rating goes down.

2 Data Organization

2.1 Initial Examination of Data

I began organizing my data by combining the breed intelligence and breed size .csv files into one table. Since there are different breeds in each dataset, quite a few rows were missing either size data or intelligence data. Neither of these would be useful for gathering information, so I eliminated the rows for which data values were missing for any column. I made a plot by taking the values from the height and weight columns for all of the breeds and storing them into arrays using NumPy.

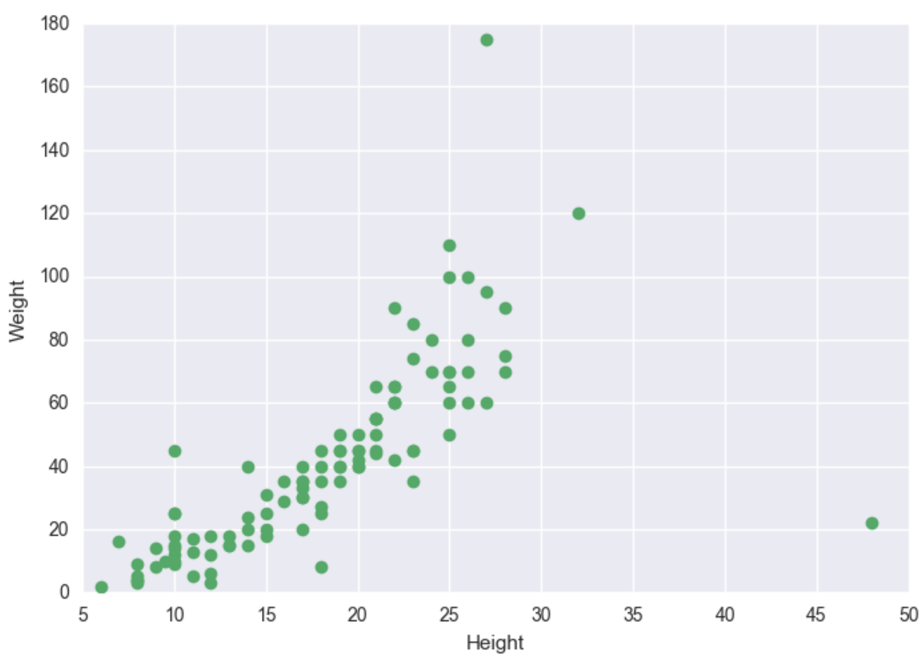

Figure 1: Graph of height vs. weight for all dog breeds for which data is provided.

Next, I used the same plots, except colored each of the points based on their classification.

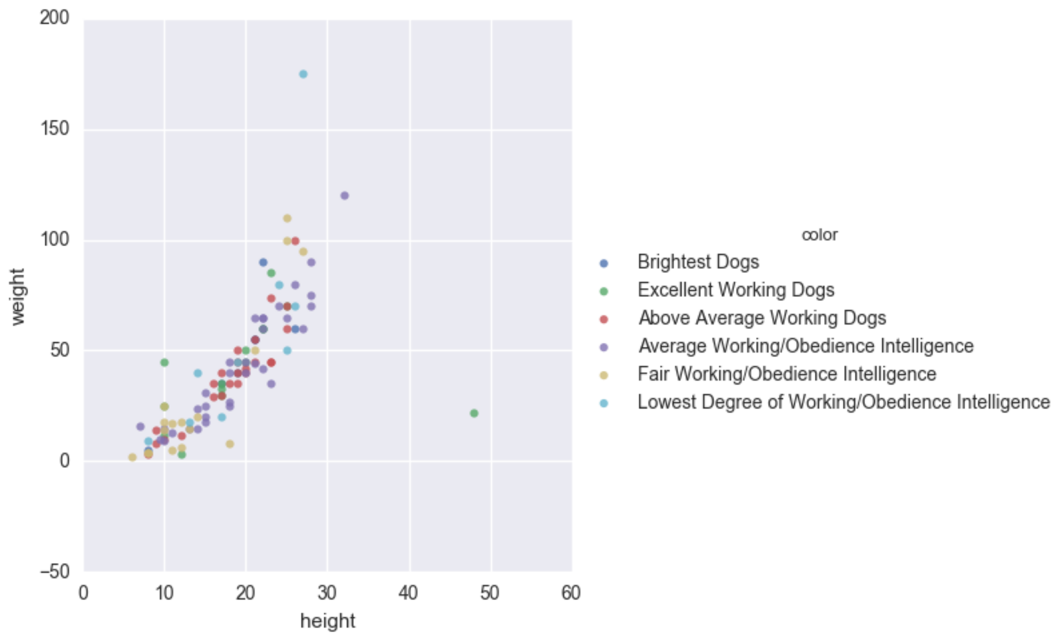

Figure 2: Graph of height vs. weight for all dog breeds for which data is provided, coloring the data points by Intelligence Classification.

Just by looking at this graph, it is hard to tell whether there will be a relationship between breed size and intelligence.

2.2 Reorganizing Data

Moving forward, I began to rearrange the data so that it would be easier to analyze. I took the average of the high and low heights and weights for each breed and put those into an average height and average weight column. I also assigned each intelligence classification a numeric value (1 being the least intelligent group and 6 being the most intelligent). I categorized the dogs based on size as well, grouping them by their weights. I found five groups, ranging from tiny to giant breeds, and assigned each category a numerical value between 1 and 5, 1 being the smallest. I chose to work with weight rather than height to begin with, because that is how I have most often seen dog size referenced as. However, I could do the same for height and see if this yields similar results. The data set that I am working with now looks like the following:

	intelligence_class	size_class	ave_height	ave_weight
Breed				
Border Collie	6	3	20.0	40.0
Golden Retriever	6	4	22.5	65.0
Doberman Pinscher	6	4	27.0	80.0
Labrador Retriever	6	4	22.5	67.5
Papillon	6	1	9.5	7.5
Rottweiler	6	5	24.5	100.0
Australian Cattle Dog	6	3	18.5	40.0
English Springer Spaniel	5	3	20.0	50.0
Schipperke	5	2	11.5	15.0

Figure 3: Sample of reformatted data set, including numeric classifications for intelligence and size, and average heights and weights.

3 Data Analysis Methods

3.1 Original Analysis

I wanted to see if I could justify a correlation between dog intelligence and size, using only the average height of the dogs. I used a function in the SciPy stats library to calculate the Pearson coefficient for the correlation. The "Pearson correlation coefficient measures the linear relationship between two datasets"(SciPy API), and returns a number between -1 and 1. A very high or low coefficient suggests a strong positive or negative correlation, whereas a number close to zero suggests a weak correlation and a value of zero suggests that no correlation exists. The function takes in two vectors, an X and Y, and returns the Pearson coefficient and p-value, assuming the null hypothesis is that no correlation exists. The scatter plot of the data is shown here:

Figure 4: Scatter plot of heights versus intelligence ratings.

Looking at the plot, it seems unlikely that a correlation exists. The Pearson function returned a coefficient of .14137 and a p-value of .15228. The correlation coefficient is very low, and the p-value is significantly over .05, so we fail to reject the null hypothesis. We can conclude from this test that no linear relationship exists.

Moving forward, I wrote functions to calculate the mean, median, and mode intelligence rating for each dog size category described above.

Figure 5: Plot of size rating versus intelligence using the average, median, and mode intelligence rating for each size class.

Based on the plot shown in Figure 5, it seems that there is a positive correlation between dog size and intelligence for tiny, small, and medium sized dogs. For medium, large and giant breeds, there seems to be either a constant or negative correlation.

3.2 Hypothesis Testing

I filtered the data further by size so that I could test for linear relationships between size and intelligence for breeds within the tiny to medium range and the medium to giant range independently. Then, I performed the Pearson test on both of these sections the same way that I used it to test all of the data points together in the previous section.

For dogs with size ratings 1, 2, and 3 (the tiny, small, and medium dogs), the Pearson function yielded a correlation coefficient of roughly .2544 and a p-value near .042. This correlation coefficient, though still weak, is nearly twice as large as the coefficient for all of the data together. A p-value of less than .05

suggests that we reject the null hypothesis in favor of an alternative hypothesis. Therefore, we are able to say that a positive correlation exists for tiny, small, and medium sized dog breads, even though the correlation is not very strong.

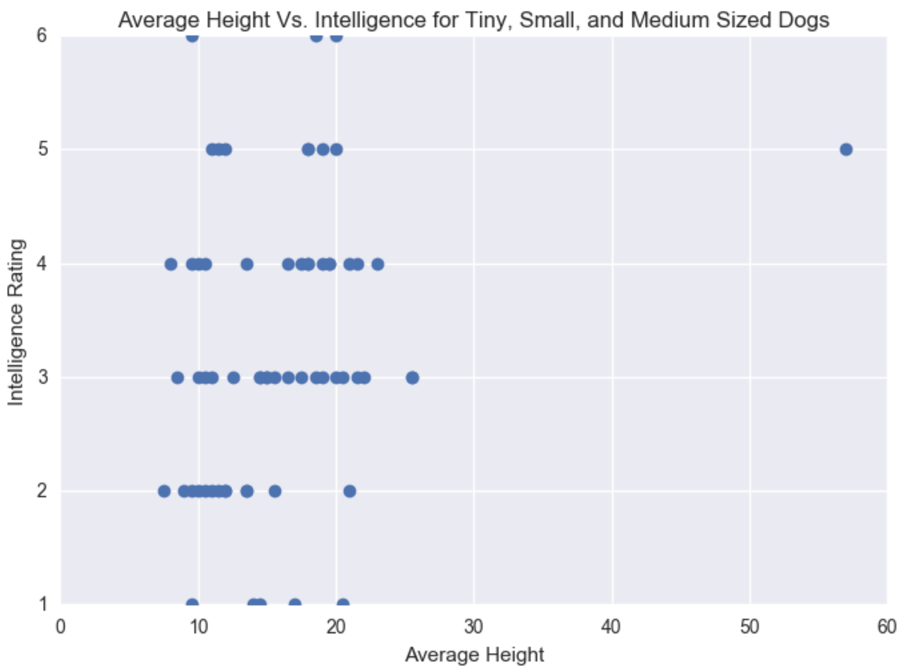

Figure 6: Scatter plot of dog size versus intelligence rating for tiny, small, and medium dog breeds.

For large and giant dog breeds, the Pearson function return a coefficient of -0.3635979 and a p-value of 0.0211. While this is still a weak correlation, it is still noticeably larger in absolute value than the coefficient provided for all data points. The p-value is low, so we can once again reject the null hypothesis and conclude that there is a negative correlation between size and intelligence for large and giant dog breeds.

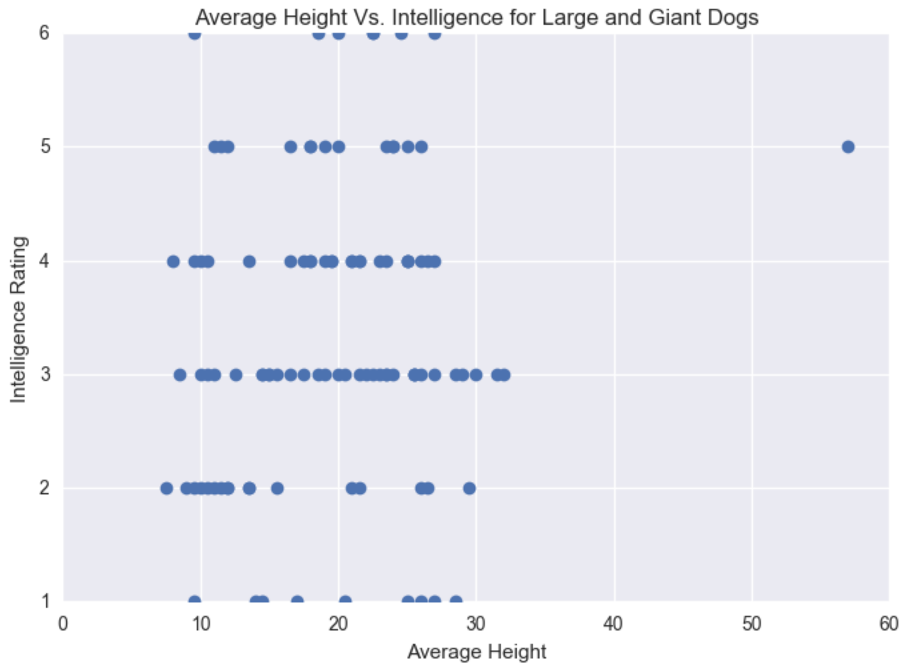

Figure 7: Scatter plot of dog size versus intelligence rating for large and giant dog breeds.

4 Conclusion

With the re-organization and tests that were performed on the datasets, I was able to find a correlation between size and intelligence. However, the correlation was weak, so it is difficult to conclude whether a significant relationship actually exists based on the data that was presented. Even so, this project was a useful educational tool because it required me to think about new ways to look at and analyze data.

4.1 Addition Considerations

Many physical characteristics could be taken into account that are not recognized in the datasets used for this project. Various studies have been done to examine the relationship between intelligence and the shape of a dogs skull, for example. At one point, I considered working with Body Mass Index calculations for breeds. However, I was unable to find a formula that was specifically intended for dogs.

Another direction to consider would be using a different measure of intelligence for dogs. The dataset that I used for this project only factored in a dog's ability to learn new commands. Dogs are bread for a wide variety of purposes. Beagles, for instance, are rated as having the lowest degree of working/obedience intelligence, but they are easily trained to be great hunting dogs.

5 References

GitHub URL

https://github.com/0314beilkej/Dog-Data-Science-Project

Course Webpage
https://www.shionguha.net/cosc4931sp17/

5.1 Works Cited

1. Fishman, L. Is there are relationship between dog size and dog intelligence?, *Data.world.* Retrieved March 9, 2017, from Data.World: https://data.world/len/dog-size-intelligence-linked.

2. "The Intelligence of Dogs", *En.m.wikipedia.org.* Retrieved March 9, 2017, from Wikipedia. https://en.m.wikipedia.org/wiki/The_Intelligence_of_Dogs#cite_refReferenceA_18-0.

3. SciPy API: Pearson Correlation Coefficient
https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.pearsonr.html