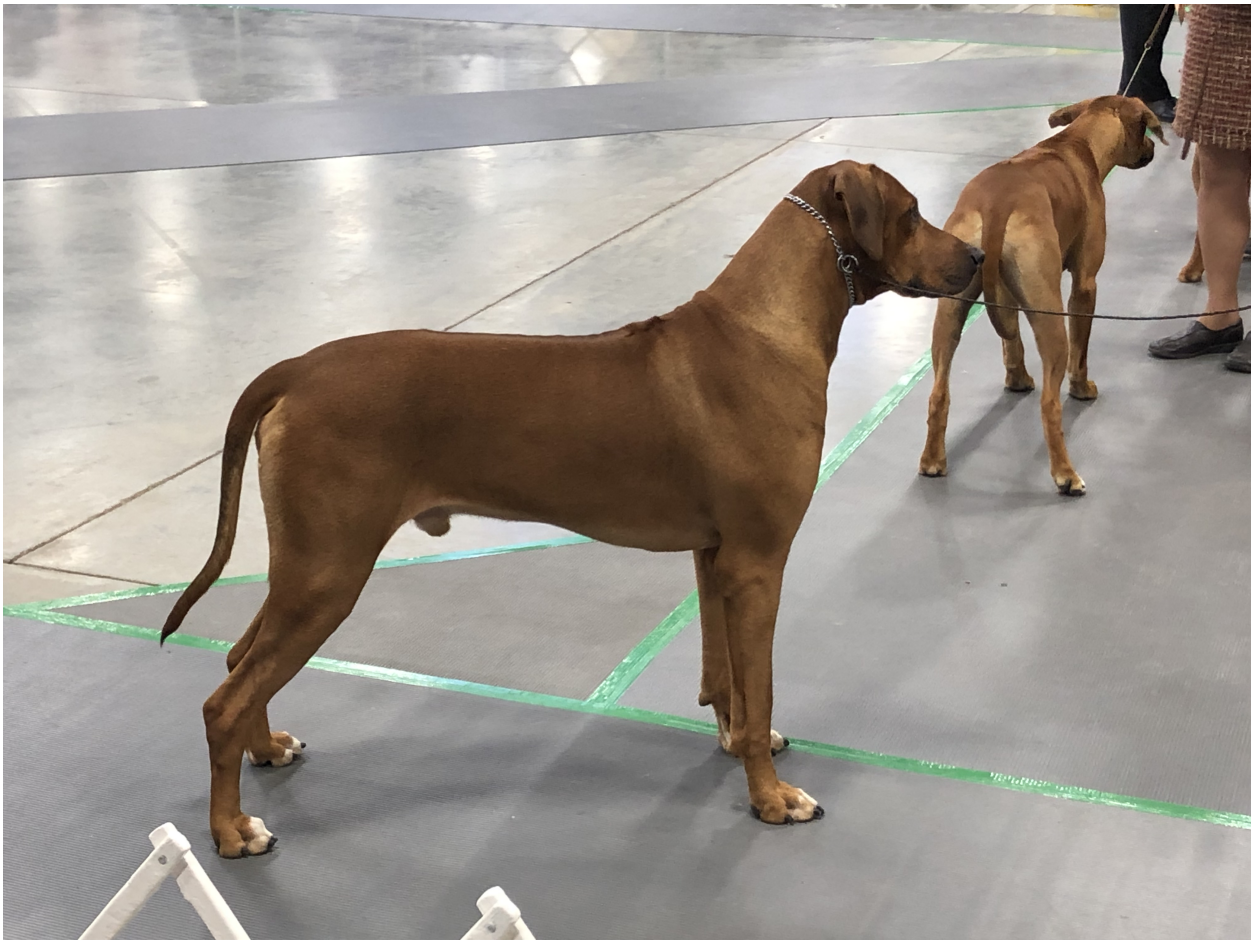
Final Project Final Submission

Cyndie Leary

5/15/2022

Thor, The Rhodesian Ridgeback. Dog of Thunder!



Project Intro

We show my dog, Thor, the Rhodesian Ridgeback in conformation. I would like to know the probability of him winning at any event. Does he have a better chance than other dogs? Once I have his overall probability, I will check what his probability is if the event is outside versus inside. Or if he has a male judge versus a female judge. And combining those, does he do better with a male or female judge when he is inside or outside?

Why I Chose this

I think it would be fun to try and predict if he will win an event that is very subjective. Winning very much depends on the specific dogs that are entered, as well as each judge that is judging. However, since we show locally, we do show against primarily the same dogs for each event.

How

We have historical data that shows how many times he has won (gets Best of Breed) against a given number of dogs. Data would look like: Event1, 5 dogs entered: Result = Lost Event2, 6 dogs entered: Result = Won Thor obviously has the best probability of winning when he is the only dog.

Body

Brief introduction of Dog Shows

A little background about confirmation dog shows. Confirmation compares each dog to how well they conform to the standard of their particular breed. If you are interested, Rhodesian Ridgeback Standard can be found by clicking here

The judging is done by breed. For each breed, dogs are always separated by Male (Dogs) and Female (Bitches). Then after that, they are separated into classes (various puppy classes by age, Bred-by exhibitor, open, etc) to determine the best in each class. Each class is judged separately, but then the winners from each class go into the ring again to decide Winner's Dog or Winner's Bitch. For each event, once all the class dogs and winners have competed, the Winner's Dog and Winner's Bitch go into the ring with all of the Specials (dogs that have previously earned their championship). A dog earns points based on the number of other dogs they beat in the class. Once they have accumulated 25 points (with other stipulations), they become a "Champion" or "Special". These dogs then compete for Best of Breed. There are other awards they can get at this time as well, but the Best of Breed (BOB) winner goes on to the group ring to compete against other BOB winners in the same group (Hound, Working, Sporting, Non-Sporting, Toy, Terrier, and herding). Then the 1st place winner of each of those groups then compete for Best in Show.

For Thor (his full kennel name is **GCHB CH Hilltop's Conquering The World**), he has earned his Bronze Grand Championship. His breed, Rhodesion Ridgeback, is in the hound group.

Data I selected:

For the data, I am only pulling it from when he first became a champion and competed in the Best of Breed class. I have also only counted a WIN as when he was awarded a Best of Breed, although he has earned other awards while in the Best of Breed ring (Best of Opposite Sex, Select Dog, etc). I also did not include any wins he got in the group ring or in the Best of Show rings.

Collecting Data and bringing it into R

I have been manually collecting data in Excel since we started showing dogs (I have always been a data nerd!). So I did parse down the data as explained above and saved it into an Excel file. Then I used the commands below to bring it into R.

First I installed the readxl package, and then called that package to use it.

install.packages("readxl")

```
library("readxl")
```

Then brought in the data:

```
Thor_Data <- read_excel("Thor Data.xlsx")
```

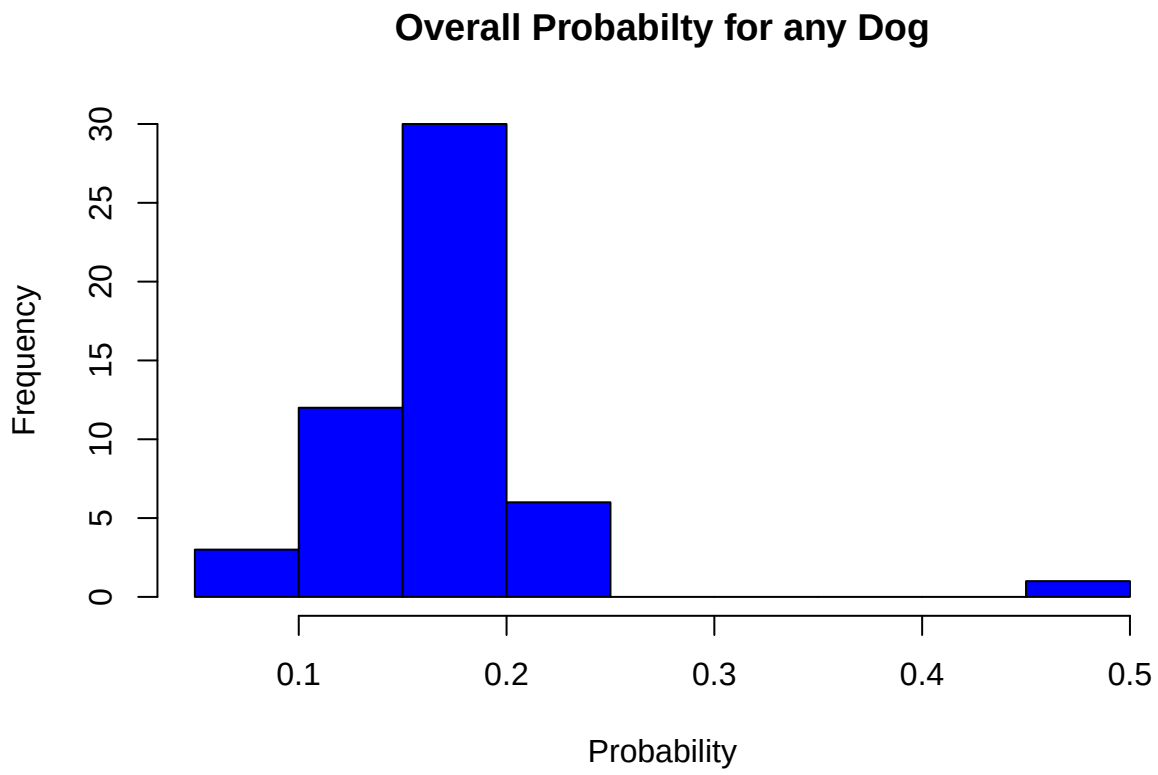
I found this instruction on the following website:

Importing Excel Files into R using readxl package

Here are some graphs to show the data we are working with: This is a graph of the overall probability for each event of any dog winning.

For reference, I used this website to customize my histograms, click for link: techvidvan.com

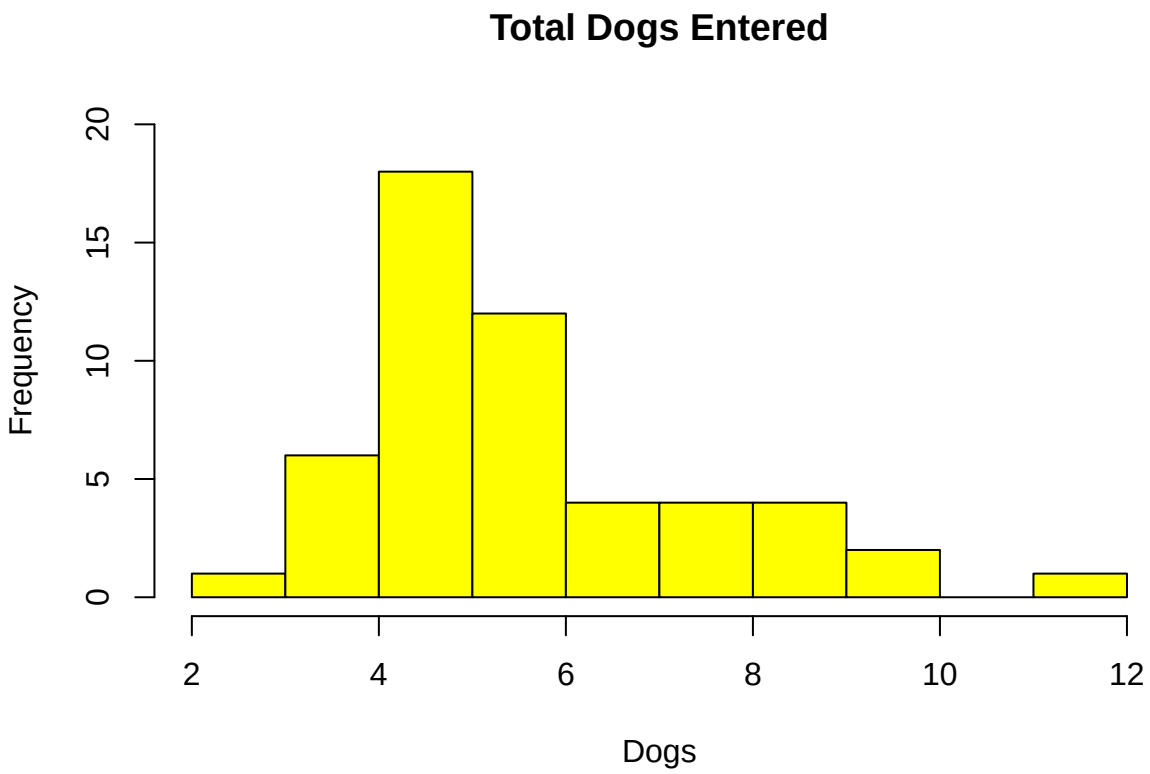
```
hist(Thor_Data$`Overall Probability for any dog`, main = "Overall Probabilty for any Dog",
     breaks = 12, xlab = "Probability", ylab = "Frequency", col = "blue")
```



There looks to be some outliers when there is 2 or less dogs in the event. (50% chance of winning)

I wanted to also graph how many dogs are entered in the Best of Breed ring.

```
hist(Thor_Data$`Dogs in BOB`, breaks = 12, main = "Total Dogs Entered", xlab = "Dogs",
     ylab = "Frequency", ylim=c(0,20),col = "yellow")
```



So, for the majority of the shows, there seems to be about 5 dogs entered.

Thor won 12 times out of 52 shows. Is this more or less than the overall expected value of winning?

Thor's Probability of winning in 52 events is:

```
ProbThorWin <- 12/52
ProbThorWin
```

```
## [1] 0.2307692
```

His probability of winning is about 23%

$$Ho:$$

Thor wins the same amount of times as any other dog in the competition. Or ProbThorWin = Expectmean

$$Ha:$$

Thor wins more times than other dogs. Or ProbThorWin > Expectmean

To determine this, I calculated the Expected number of wins given each probability for each event.

```
ExpectWins <- mean(Thor_Data$`Overall Probability for any dog`)*(dim(Thor_Data)[1])
ExpectWins
```

```
## [1] 9.399206
```

Since the expected wins is approximately 9, and Thor has won 12 times, he seems to be slightly more probable to win. But I will continue to test that out.

Any Dog's Probability of winning in 52 events is:

```
ProbAnyWin <- 9/52
ProbAnyWin
```

```
## [1] 0.1730769
```

Any dog's probability of winning is about 17%, and Thor's is about 23%, but let's continue to test that out.

Calculating Standard Deviation of the probability of each event:

```
OverallStd <- sd(Thor_Data$`Overall Probability for any dog`)
OverallStd
```

```
## [1] 0.06259883
```

And the mean:

```
ExpectMean <- mean(Thor_Data$`Overall Probability for any dog`)
ExpectMean
```

```
## [1] 0.180754
```

Calculating the z-score:

```
ThorzScore <- (ProbThorWin - ExpectMean)/OverallStd
ThorzScore
```

```
## [1] 0.7989808
```

This value shows that Thor's results are less than one standard deviation away from the expected wins mean. So, that seems to close to call. I can not reject the null hypothesis.

Next, I wanted to count the number of times he had a male judge or a female judge.

```
table(Thor_Data$`Judge (M/F)`)
```

```
##
## Female   Male
##     26     26
```

Which, I was surprised at the even split. We will determine which he won with more later.

Also, I wanted to see how many shows where Thor was outside versus inside.

```
table(Thor_Data$`Inside/Outside`)
```

```
##
##  Inside Outside
##      26      26
```

I was also surprised at this result. And we will also calculate how often he won for each as well.

I created a subset of just his wins

```
Thor_WinData <- subset(Thor_Data, Thor_Data$`Won (1=Y/0=N)` >= 1)
```

And then repeated the counts of Judges and Venue location

```
table(Thor_WinData$`Judge (M/F)`)
```

```
##
## Female   Male
##      7      5
```

```
table(Thor_WinData$`Inside/Outside`)
```

```
##
##  Inside Outside
##       7       5
```

Interesting that those counts are balanced in the same way.

So it seems that Thor slightly prefers to show to females and prefers to be inside. But does he win more when both of those are true?

Creating another subset of indoor/outdoor wins.

```
Thor_VenueWin <- subset(Thor_WinData, Thor_WinData$`Inside/Outside` =="Inside")
```

And another count of judges:

```
table(Thor_VenueWin$`Judge (M/F)`)
```

```
##
## Female   Male
##      6      1
```

So this shows, of his 12 total wins, 7 of them were inside, and 6 of those were given to him by a female judge.

I can kind of conclude from this that Thor prefers to be inside and prefers Female judges, in order for him to show his best. Or conversely, more female judges prefer Thor, and see his movement better when he is inside. (He is kind of a Big Diva and doesn't like his feet to be uncomfortable on grass, not to mention if it is wet grass or it is raining!)

But let's look at the probability to verify.

We saw earlier, that Gender of judge or location of event were both split in half, so their probabilities will all be the same.

```
ProbFemale <- 26/52
ProbFemale
```

```
## [1] 0.5
```

```
ProbMale <- 26/52
ProbMale
```

```
## [1] 0.5
```

```
ProbInside <- 26/52
ProbInside
```

```
## [1] 0.5
```

```
ProbOutside <- 26/52
ProbOutside
```

```
## [1] 0.5
```

Using values from above, we can also calculate the conditional probability of Thor winning under each of those situations

```
ProbWinFemale <- 7/52
ProbWinMale <- 5/52
ProbWinInside <- 7/52
ProbWinOutside <- 5/52
ProbWinFemale
```

```
## [1] 0.1346154
```

```
ProbWinMale
```

```
## [1] 0.09615385
```

```
ProbWinInside
```

```
## [1] 0.1346154
```

```
ProbWinOutside
```

```
## [1] 0.09615385
```

Using those numbers as the base numbers, we can now calculate inside/outside (recall they have the same proportions as male female)

```
ProbThorWinFemaleInside <- ProbWinFemale + ProbWinInside
ProbThorWinFemaleOutside <- ProbWinFemale + ProbWinOutside
ProbThorWinMaleInside <- ProbWinMale + ProbWinInside
ProbThorWinMaleOutside <- ProbWinMale + ProbWinOutside
ProbThorWinFemaleInside
```

```
## [1] 0.2692308
```

```
ProbThorWinFemaleOutside
```

```
## [1] 0.2307692
```

```
ProbThorWinMaleInside
```

```
## [1] 0.2307692
```

```
ProbThorWinMaleOutside
```

```
## [1] 0.1923077
```

Topics From Class

R Markdown:

I was introduced to R and R Markdown in this class. I like it quite a bit. I used Juniper Notebooks in my Python class, but that didn't seem to have as much functionality, or we didn't get that far into it. I explored how to add pictures to my R Markdown and was happy that I was able to figure that out! Still need to do some more investigation on how to rotate them or change their size, but I got them to work for this exercise. I also learned how to customize my histograms. I am learning a lot by reading everyone else's projects as well!

Github:

I very briefly used Github in my Python class as well. But mostly for pulling down data. We also explored using it for our final group project, but found it too confusing for many of us who were new to using it. I suspect that as I continue down my Masters degree, I will be using Github far more and look forward to learning it. I have had no problem getting it to work. I do feel that since I had a class in software engineering and am also familiar with using fat clients when building upgrades to ERP systems, that has helped me understand the concept of Github.

Probabilty:

Probability was obviously the basis of my project. Although, I am still confused sometimes with probability versus proportions. But I feel like I have learned a bit more with this project.

Joint Probabilites:

While I was working on this project, I went back to the book, and reviewed Joint Probabilities. I created a table for myself to help check my answers. I was trying to figure out how, without just uploading the data, to get it into R Markdown, but have not succeeded in that yet. But I was glad that my answers were matching my table!

Conditional Probabilities:

Again, going to the book to confirm what I was calculating, I reviewed Conditional Probabilities that shows the outcome A (Thor Winning) give Condition B (Female Judge). Then I took that a step farther, and said the outcome of Thor winning with a female judge, given the event is inside.

Conclusion

With this data, it looks like we should choose shows where Thor can show to a female judge while showing indoors. This is very interesting data to have. It costs about $100 a weekend of showing (2 shows), plus travel expenses, grooming spaces, food, etc and of course, time. So if we can narrow it down to shows with this criteria, we can better use our money to continue to win with Thor. However, the data is really too close to make any hard and fast rules on what to do. Probably proving that it is very subjective!

I did learn a lot with this project, but I struggle with knowing when to use what equations, and how to get them to function in R. I do find R to be very useful though. I want to explore it more. We used SAS and Tableau in other classes and those were both very robust programs that are harder to understand.

I also took initial feedback from Nolan Sawchuk and added in information on the expected # of wins to see if that was helpful. And added graphs per Donna Weber's and Sarah Presse's suggestions. Ellie Byler gave me the great tip on making sure my url's showed up in BLUE in my PDF.

Appendix 1: Full Data Set

How to do this was found here: Printing a Dataframe in R Markdown

	Date	Inside/Outside	Judge (M/F)	# in his Class	Class Dogs	Class Bitches	Special Dog	Special Bitches	Dogs in BOB	Won (1=Y/0=N)	TOTAL DOGS Beat	Dogs	Bitches	Thor P(A) Each Event	Overall Probability for any dog
1	2018-06-21	Outside	Male	NA	4	5	4	0	6	1	5	5	1	0.1666667	0.1666667
2	2018-06-22	Outside	Male	NA	5	7	5	1	8	0	0	6	2	0.0000000	0.1250000
3	2018-06-23	Outside	Male	NA	4	9	6	1	9	0	0	7	2	0.0000000	0.1111111
4	2018-06-24	Outside	Female	NA	4	9	6	0	8	0	0	7	1	0.0000000	0.1250000
5	2018-08-11	Inside	Female	NA	2	2	3	1	6	0	0	4	2	0.0000000	0.1666667
6	2018-08-12	Inside	Male	NA	2	2	3	1	6	0	0	4	2	0.0000000	0.1666667
7	2018-08-25	Outside	Female	NA	4	3	1	2	5	0	0	2	3	0.0000000	0.2000000
8	2018-08-26	Outside	Male	NA	2	3	1	2	5	0	0	2	3	0.0000000	0.2000000
9	2018-11-17	Inside	Female	NA	1	2	3	0	5	1	4	4	1	0.2000000	0.2000000

	Date	Inside/Outside	Judge (M/F)	# in his Class	Class Dogs	Class Bitches	Special Dog	Special Bitches	Dogs in BOB	Won (1=Y/0=N)	TOTAL DOGS Beat	Dogs	Bitches	Thor P(A) Each Event	Overall Probability for any dog
10	2018-11-18	Inside	Male	NA	1	2	3	1	6	0	0	4	2	0.0000000	0.1666667
11	2018-12-15	Inside	Female	NA	2	2	3	1	6	0	0	4	2	0.0000000	0.1666667
12	2018-12-16	Inside	Male	NA	3	1	2	1	5	0	0	3	2	0.0000000	0.2000000
13	2019-01-05	Inside	Female	NA	4	4	3	2	7	0	0	4	3	0.0000000	0.1428571
14	2019-01-06	Inside	Female	NA	4	4	3	2	7	0	0	4	3	0.0000000	0.1428571
15	2019-05-11	Outside	Female	NA	2	2	2	1	5	0	0	3	2	0.0000000	0.2000000
16	2019-05-12	Outside	Male	NA	2	2	1	1	4	1	3	2	2	0.2500000	0.2500000
17	2019-05-25	Outside	Male	NA	2	1	3	1	6	0	0	4	2	0.0000000	0.1666667
18	2019-05-26	Outside	Male	NA	2	1	3	1	6	0	0	4	2	0.0000000	0.1666667
19	2019-06-08	Outside	Female	NA	2	1	1	1	4	1	3	2	2	0.2500000	0.2500000
20	2019-06-09	Outside	Female	NA	2	1	1	1	4	0	0	2	2	0.0000000	0.2500000
21	2019-06-20	Outside	Male	NA	3	4	4	3	9	0	0	5	4	0.0000000	0.1111111
22	2019-06-21	Outside	Female	NA	3	7	5	5	12	0	0	6	6	0.0000000	0.0833333
23	2019-06-22	Outside	Male	NA	4	8	5	3	10	0	0	6	4	0.0000000	0.1000000
24	2019-06-23	Outside	Male	NA	4	4	4	3	9	0	0	5	4	0.0000000	0.1111111
25	2019-07-11	Inside	Female	NA	1	0	3	1	5	0	0	4	1	0.0000000	0.2000000

	Date	Inside/Outside	Judge (M/F)	# in his Class	Class Dogs	Class Bitches	Special Dog	Special Bitches	Dogs in BOB	Won (1=Y/0=N)	TOTAL DOGS Beat	Dogs	Bitches	Thor P(A) Each Event	Overall Probability for any dog
26	2019-07-12	Inside	Male	NA	2	0	2	1	4	1	3	3	1	0.2500000	0.2500000
27	2019-07-13	Inside	Female	NA	2	0	3	1	5	1	4	4	1	0.2000000	0.2000000
28	2019-07-14	Inside	Male	NA	2	0	2	1	4	0	0	3	1	0.0000000	0.2500000
29	2019-08-24	Outside	Male	NA	6	4	5	3	10	1	9	6	4	0.1000000	0.1000000
30	2019-08-25	Outside	Male	NA	4	3	4	3	9	0	0	5	4	0.0000000	0.1111111
31	2019-11-23	Inside	Female	NA	4	3	2	2	6	0	0	3	3	0.0000000	0.1666667
32	2019-11-24	Inside	Female	NA	4	3	2	2	6	0	0	3	3	0.0000000	0.1666667
33	2019-12-14	Inside	Female	NA	1	0	1	0	2	0	0	2	0	0.0000000	0.5000000
34	2019-12-15	Inside	Female	NA	1	1	1	2	5	1	4	2	3	0.2000000	0.2000000
35	2020-01-04	Inside	Female	NA	6	7	3	2	7	0	0	4	3	0.0000000	0.1428571
36	2020-01-05	Inside	Female	NA	6	6	3	2	7	0	0	4	3	0.0000000	0.1428571
37	2020-10-31	Inside	Female	NA	1	1	2	1	5	1	4	3	2	0.2000000	0.2000000
38	2020-11-01	Inside	Female	NA	1	1	2	1	5	0	0	3	2	0.0000000	0.2000000
39	2021-05-01	Inside	Male	NA	1	1	3	0	5	0	0	4	1	0.0000000	0.2000000
40	2021-05-02	Inside	Male	NA	2	1	3	1	6	0	0	4	2	0.0000000	0.1666667
41	2021-06-24	Outside	Male	NA	3	2	2	2	6	0	0	3	3	0.0000000	0.1666667

	Date	Inside/Outside	Judge (M/F)	# in his Class	Class Dogs	Class Bitches	Special Dog	Special Bitches	Dogs in BOB	Won (1=Y/0=N)	TOTAL DOGS Beat	Dogs	Bitches	Thor P(A) Each Event	Overall Probability for any dog
42	2021-06-25	Outside	Female	NA	4	2	2	2	6	0	0	3	3	0.0000000	0.1666667
43	2021-06-26	Outside	Male	NA	4	3	4	2	8	0	0	5	3	0.0000000	0.1250000
44	2021-06-27	Outside	Male	NA	3	8	5	1	8	0	0	6	2	0.0000000	0.1250000
45	2021-07-15	Inside	Female	NA	0	2	3	1	5	1	4	3	2	0.2000000	0.2000000
46	2021-07-16	Inside	Male	NA	0	3	3	1	5	0	0	3	2	0.0000000	0.2000000
47	2021-07-17	Inside	Female	NA	0	3	3	1	5	0	0	3	2	0.0000000	0.2000000
48	2021-07-18	Inside	Female	NA	0	2	4	0	5	1	4	4	1	0.2000000	0.2000000
49	2021-08-21	Outside	Male	NA	0	1	4	0	5	0	0	4	1	0.0000000	0.2000000
50	2021-08-22	Outside	Male	NA	0	0	4	0	4	1	3	4	0	0.2500000	0.2500000
51	2021-08-28	Outside	Female	NA	1	2	3	0	5	0	0	4	1	0.0000000	0.2000000
52	2021-08-29	Outside	Male	NA	1	2	3	0	5	0	0	4	1	0.0000000	0.2000000

Appendix 2: Another picture of Thor!