Mammal Sleep Behavior Analysis

2355170 - Abdullah Burkan Bereketoglu
2502490 - Ismail Cagan Yilmaz
2428852 - Sena Gulizar Aktas 2429330 - Yusuf Turan
2290831 - Zeynep Oz

1/30/2022

Abstract

In this project, we as group_10 analyzed a dataset called Msleep in short, which is Mammal Sleep Behavior Data Set. In this data set, sleep patterns are given with body weight, brain weight, conservation type, genus, order and -vore type that can be said that has effects on the sleeping patterns of the mammals. Sleep patterns in between -vores and other features are analyzed to understand if there exist a pattern at all. These traits then analyzed to find a statistical significance of sleep pattern with different statistical metrics and perspectives, such as linearity testing, mean significance by t-test, correlation analysis, linear regression and variance analysis. Results indicate that there are some significant dependencies between body~brain weight to sleep, furthermore domestication has an effect, but -vore effect is not significant.

Contents

Introduction

In our analysis, we wanted to find a significance of sleep time between different mammal types by looking at several given features of these mammals. By that we wanted to point out the fact that different features that occur on different mammals can change their sleep patterns. To be able to make this analysis we used five different methods and pointed out different features effects on the sleep time of the mammal[1-2]. The metadata that is taken for the data set is from Wikipedia as provider Vincent Arel Bundock mentioned in his website[1-2].

Data Description

In the data set, Mammal Sleep Behavior Data, we have eleven features, except ID, which are; name, genus, -vore, order, conservation type, sleep total, sleep rem, sleep cycle, awake time, brain weight, body weight. Let's look at their descriptions. In the appendices Table-1 there is full description of each data variable with their definitions[2].

To be able to describe the properties that is in Table-1, since Mammal Sleep Behavior Data consisted many NA values, we first needed to omit them or remove them from the data set. Each variable had NA values at different rows, so one

needed to omit NA's for each different analysis they were going to conduct.

Research Questions

We followed certain questions came to our minds to understand the data set in the direction that we desired to follow. And these questions can be seen in the Table-2 in appendices.

By answering these questions we planned on achieveing our aim of the study which we will elaborate in the next part of the study.

Aim of the Study

The overall goal for this study conducted is to see if there is a dependency or correlation of two or more features such as weight, conservation is on effect of the sleep time of mammals. From that we wanted to achieve if there is a statistical significance, then further studies on the case can be conducted to further improve mammal sleep quality, since it is directly correlated with sleep time[3].

Methodology/Analysis

In our study of Mammal Sleep Behavior Data, we used five different approaches on the data set on various features such as body weight, brain weight, conservation, - vore type and analyzed their relation and dependency with sleep time, rem sleep time.

First method that is applied is to check the linearity of the brain weight versus the total sleep time the applied linearity test is QQ Plot which finds linear relationship and visually interprets it, we used QQ method to determine whether a data collection is likely to come from a theoretical distribution, such as normal or exponential[4]. E.g, if we are conducting a statistical study and assume that our dependent variable is normally distributed, we can use a normal QQ chart to validate that assumption[4]. This method is just visual inspection, not fool-proof proof, hence little subjective[4]. However, it allows us to quickly determine if our assumption is credible, and if not, how the assumption is broken and what data points contribute to each breach[4]. Lastly, a scatterplot generated by plotting two sets of quantiles against each other which is the Q-Q plot that is mentioned above[4]. If both sets of quantiles come from the same distribution, the points should form a relatively straight line[4].

Second method is two-tail t-test and mean comparison of brain weights of Car-

nivore and Omnivore Mammals to see if there a significant difference between their means[4]. This method is used to firstly visualize the means of two -vores by boxplot and their outliers, later on comparing the means by t-test, to see if there is a significant difference between the brain weights then from that if there exist a difference we can look for a dependency of sleep time to brain weight[5-6]. By this test we wanted to test the general hypothesis that a significantly smart animal which is omnivore, eventually has a bigger brain and omnivore animals are more developed than carnivore animals, from the result we can also reject or support this hypothesis[6].

Third method that is used is the correlation, and by that we wanted to achieve to find a correlation between body weight and total sleep time of each mammal. If there is such relationship, then we could investigate on how body weight affects the total sleep of each animal[7].

Fourth method that is used is linear modelling on animal's eating behaviors and the corresponding body weights to make a further estimation, if there exist a significant relation that is given as result of the linear model[8-9]. Basically after the correlation, to further investigate and find a predictive function if there is strong enough statistical significance of this effect.

Fifth, and last variance analysis method is applied on the total sleep time ~ -vore types, and total sleep time ~ conservation. This method is conducted to analyze whether there are clustered zones that show low variance for different types of conservation and -vores[10]. By that we can see that for some -vores there indeed visible evidence that -vore type or conservation type is effective on the sleep time of the mammal[10].

Results and Findings

In our linearity test, please refer to Table-2 in appendices, we found out that as it can be also seen from the Figure-2 in appendices, that on the tails of the Q-Q plot dots are showing a significantly non-linear behavior, hence they are not on the straight line. Therefore, it can be deduced that this distribution does not show a normal distribution. Since the points do not form a linearity on the line, it can be said that there is no linear relationship between the total sleep of mammals and their brain weight.On the other hand, if there existed a linear relationship that is visible on the plot, the degree of the relationship can be checked by looking at the Pearson correlation coefficient.

For the second question, as it can be seen from the Figure - 3, we used a box-plot to visualize the means of omnivore and carnivore and later on in the Table - 3 applied a two-tail t-test to test our hypothesis. For two-sided test, our H0(null hypothesis) is that the means of the brain weight of carnivore and omnivore animals are equal and our HA(Alternative hypothesis, omnivore with bigger brains.) is that the means of the brain weight of carnivore and omnivore animals are not equal which means that there is a significant difference[6]. We implemented the two-sided test with 95% confidence interval and found that our p-value (0.4473) is greater than our significance value (0.05) and concluded that H0 can not be re-

jected. With this, we can not claim that there is a significant difference between the brain weights of carnivore and omnivore animals and we can eventually reject the general opinion[6].

For our third question we wanted to answer, whether there is a significant relationship between the animals' total sleep time and their body weight. For this, we performed a correlation test, and we observed that the p value we found was smaller than the significance level[7]. For the results please check Table - 4. In this way, we realized that we could reject the null hypothesis and, as a result, there was a relationship between body weight and total sleep[7]. The correlation coefficient we found is -0.46, which means that there is a negative relationship between the two variables[7]. For the visual result please refer to Figure - 4 (In discussion add can we really say 0.46 correlation high?)

For the fourth question two sub-questions answered. First one "Is there any relationship between Brain weight and rem sleep?" which is actually a continuum question to the third question, here as a result of the Pearson correlation, which can be seen in the appendix as Table - 5, shows us that the p-value of the test is 0.1306, which is greater than the significance level alpha = 0.05. We can conclude that Brain weight and rem sleep are NOT correlated with a correlation coefficient of -0.2213348 and p-value of 0.1306. Second question that is answered in the fourth question is about the estimation of total sleep by linear regression, and this regression is made in between each -vore and their body weights. In the Table - 6 on appendix, the results of the linear model can be seen and it shows that for some of the -vore types sleep time and vore type is not significant at all with 0.76 probability. For insectivore and Carnivore there is at alpha 0.05 level a significance. Also it is seen that sleep time is also related with the body weight. From

the R-squared value we can see that values are not well fit into the model, and from the F-statistic we can clearly deduce that Overall the statistics is significant and perhaps there might be a linear model after all for these relevance. Lastly in Table - 7, one can see the predictions at confidence level 0.95.

In the last question, we conducted a variance analysis to see if there is a significant clustered zones for different -vore types for total sleep time, by that we wanted to conclude with a further study that, having different eating habits affect sleeping behavior significantly. Later on, also wanted to see the results of how domestication as a conservation type affects the sleeping time and it's variance. From the Figure - 5, check appendix, it can be seen that there are no statistical significance for the -vore types and total sleep time, since most of them except omnivores have high variance which shows that there is no significant sleeping cycle for eating habits.

Later on in the second part of the question which includes sleeping time variance analysis with domestication and other types of conservations. We can clearly see that domestication has effect on the variance of the sleeping time. In the Endangered Sleep Spread, which can be seen from Figure - 6 at appendix, one can see that even some mammals sleep twenty and more hours, but when they are domesticated the range gets smaller and between 4-12 which may be an indicator that domestication has an effect on sleep time. Lastly, we wanted to make a one last One Way Anova to the conservation and sleep time comparison, and we clearly saw that from the results of Table - 8, refer to appendix[11]. From the F value we can state that the variance in the conservation(not separated but looked wholly.) shows that variation is not caused by just one variable in the data, but there is a real significant variability, which we saw on the Figure - 6 comparison[11].

Discussion/Conclusion

Since, some of our results for the questions slightly passed the threshold that is put, it actually indicates that our statistical significance are actually not that powerful. For example, in linear model R-squared scores are relatively low and for making predictions one may expect all the values in the linear model summary statistics to be comparably high. Moreover, correlation that we accepted which is -0.46 is actually not that high either, for most scientific studies, this can be said relatively low but promising. Furthermore, on the variance analysis, we saw that there is no significant effect of eating habit type to sleeping cycles, but perhaps there is and we did not include all parameters that are necessary to show that pattern. If there existed more time for the project conducted, we planned on applying non-linear mappings, and clustering methods to see some other patterns, that we might be missing just by looking at the data set. Other than that we also wanted to do a multivariate analysis on the data set to have a better picture of the mammal sleeping behavior.

References

[1] “An updated and expanded version of The Mammals Sleep Dataset,” R. [Online]. Available: https://vincentarelbundock.github.io/Rdatasets/doc/ggplot2/msleep.html. [Accessed: 19-Jan-2022].

[2] “Conservation status,” Wikipedia, 22-Dec-2021. [Online]. Available: https://en.wikipedia.org/wiki/Conservation_status. [Accessed: 19-Jan-2022].

[3] J.-H. Kang and S.-C. Chen, “Effects of an irregular bedtime schedule on sleep quality, daytime sleepiness, and fatigue among university students in Taiwan,” BMC public health, 19-Jul-2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2718885/. [Accessed: 29-Jan-2022].

[4] C. Ford, “University of Virginia Library Research Data Services + Sciences,” Understanding Q-Q Plots, 26-Aug-2015. [Online]. Available: https://data.library.virginia.edu/understanding-q-q-plots/. [Accessed: 20-Jan-2022].

[5] A. Hayes, “What is a two-tailed test?,” Investopedia, 17-Jan-2022. [Online]. Available: https://www.investopedia.com/terms/t/two-tailed-test.asp. [Accessed: 31-Jan-2022].

[6] C. Koch, “Does brain size matter?,” Scientific American, 01-Jan-2016. [Online]. Available: https://www.scientificamerican.com/article/does-brain-size-matter1/. [Accessed: 31-Jan-2022].

[7] Z. Jaadi, “Eveything you need to know about interpreting correlations,” Medium, 16-Oct-2019. [Online]. Available: https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8. [Accessed: 20-Jan-2022].

[8] J. Brown, “Interpreting R output for simple linear regression part 1,” YouTube, 31-Jan-2016. [Digital Video]. Available: https://www.youtube.com/watch?v=u7TxjUI4PRI. [Accessed: 19-Jan-2022].

[9] J. Brown, “Interpreting R output for simple linear regression part 2,” YouTube, 8-Feb-2016. [Digital Video]. Available: https://www.youtube.com/watch?v=sKW2umonEvY. [Accessed: 19-Jan-2022].

[10] “Variance analysis,” Corporate Finance Institute, 31-Jan-2022. [Online]. Available: https://corporatefinanceinstitute.com/resources/knowledge/accounting/variance-analysis/. [Accessed: 31-Jan-2022].

[11] R. Bevans, “ANOVA in R: A step-by-step guide,” Scribbr, 01-Jul-2021. [Online]. Available: https://www.scribbr.com/statistics/anova-in-r/. [Accessed: 31-Jan-2022].

Appendix

Photos of the Authors

Here all the group members who made this project to be successfully accomplished. Please look at the figure-1 in the appendices to see the authors.

Figure-1

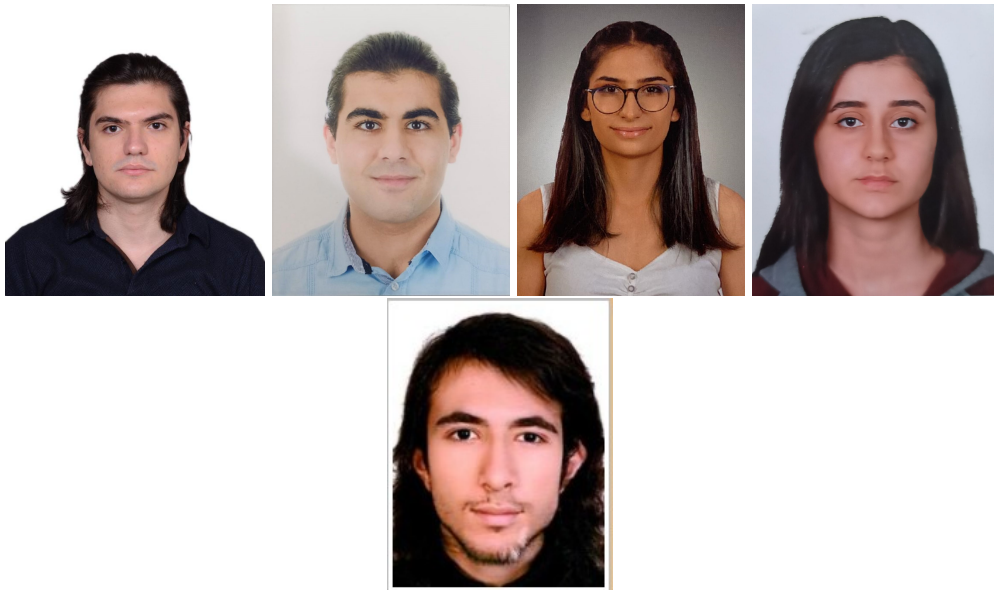

Figure 1: Images of Group Members- Burkan, Yusuf, Zeynep, Sena, Ismail(Left-Up to Right-Down)

Table-1

Features	Definitions
Name	Shows the name of the mammal of the traits, a categorical variable.
Genus	Shows each individual mammals genus (taxonomic group above species) group, a categorical variable.
Vore	Shows each individual mammals -vore group, a categorical variable.
Order	Shows each individual mammals order (taxonomic group above family) group, a categorical variable.
Conserv.	Shows each individual mammals conservation type (Endangered, least concern, etc.), a categorical variable.
Total Sleep	Shows each individual mammals total sleeping time each day, a numeric variable.
Rem Sleep	Shows each individual mammals total time spent on rem sleep each day, a numeric variable.
Sleep Cycle Ratio	Shows each individual mammals total time spent on rem sleep versus total sleep time each day, a numeric (ratio) variable.
Awaken Time	Shows each individual mammals total time of being awake each day, a numeric variable.
Brain weight	Shows each individual mammals brain weight, a numeric variable.
Body Weight	Shows each individual mammals body weight, a numeric variable.

Table-2

Q#	Questions
Q1	Is there a linear dependency that is statistically significant between brain weight and total sleep time?
Q2	Is there any significant difference between the means of brain weight of Carnivore and Omnivore animals?
Q3	Is there any relationship between body weight and total sleep?
Q4	Total sleep estimation of the animal's eating behavior and their body weight with linear model.
Q5	Variance Analysis of Sleep Time for corresponding eating habit and for domestication and other conservation types, to see if there exist a statistically significant difference between Sleep time related to these features.

Figure - 2

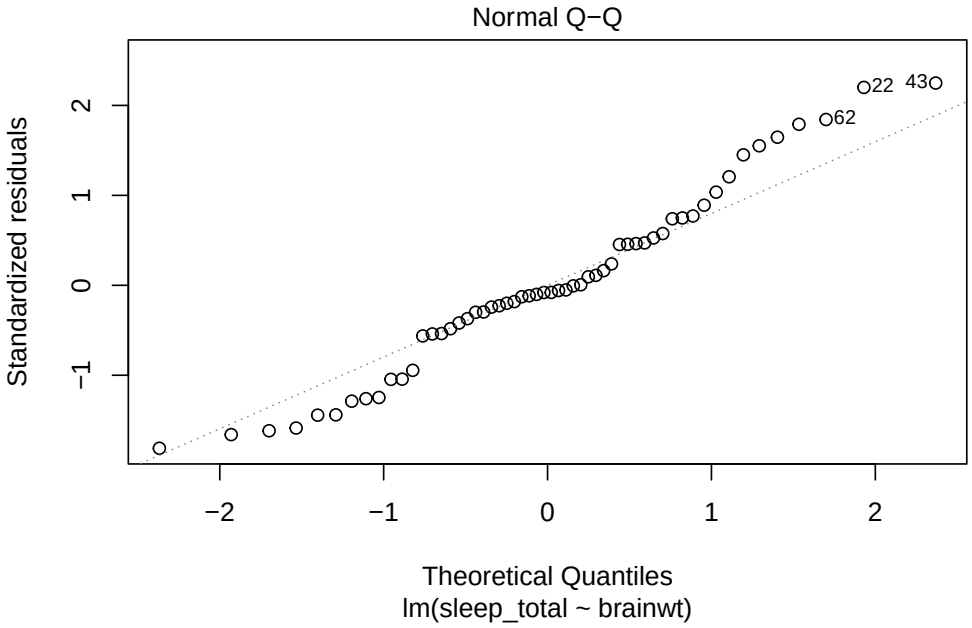

Figure 2: Linearity test Q-Q plot

Figure - 3

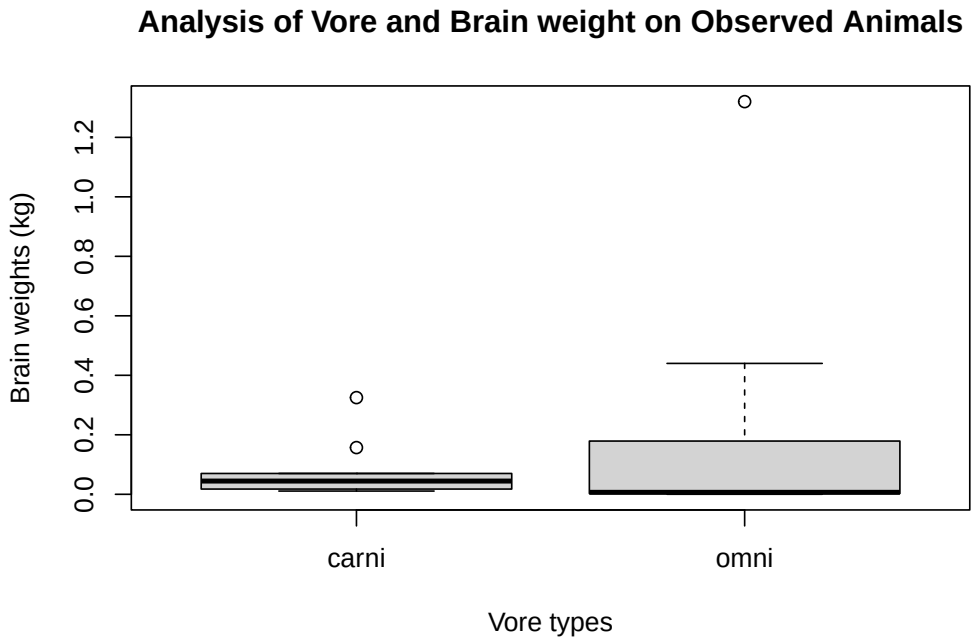

Figure 3: Brain Weight vs Vore Types Boxplot

Table - 3

```
##
##  Welch Two Sample t-test
##
## data:  omni.brainwt$brainwt and carni.brainwt$brainwt
## t = 0.77434, df = 21.114, p-value = 0.4473
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.1119975  0.2449488
## sample estimates:
##  mean of x  mean of y
## 0.14573118 0.07925556
```

Table - 4

```
##
##  Pearson's product-moment correlation
##
## data:  cordata$Total.sleep and cordata$Body.weight
## t = -3.5585, df = 46, p-value = 0.0008787
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.6614249 -0.2079179
## sample estimates:
##       cor
## -0.4646053
```

Figure - 4

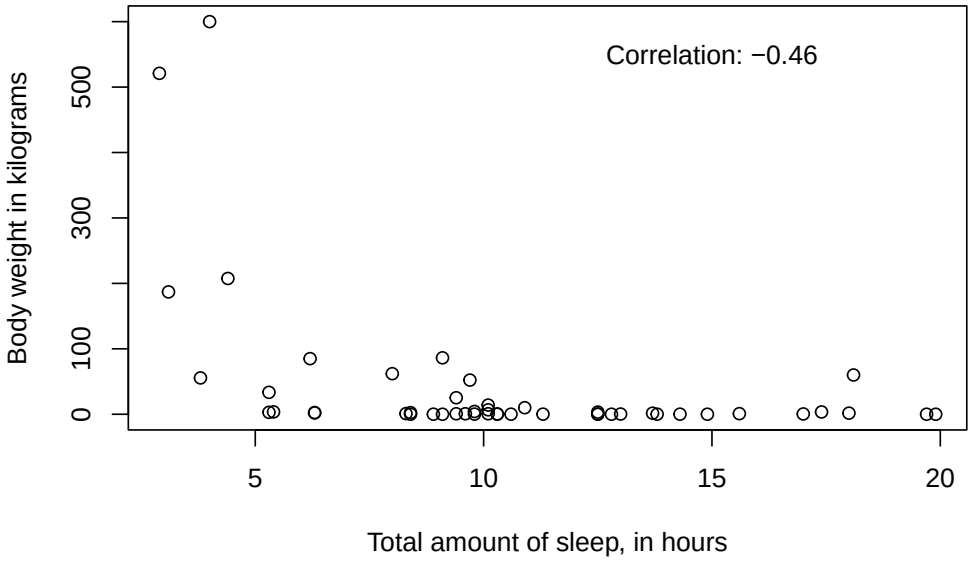

Figure 4: Body weight vs total sleep time correlation plot

Table - 5

```
##
##  Pearson's product-moment correlation
##
## data:  q4$sleep_rem and q4$brainwt
## t = -1.5393, df = 46, p-value = 0.1306
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.47556189  0.06701441
## sample estimates:
##        cor
## -0.2213348
```

Table - 6

```
##
## Call:
## lm(formula = sleep_total ~ vore + bodywt, data = q4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0821 -2.6943 -0.6245  3.9591  8.8740
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.5265817  0.9859437  10.677   <2e-16 ***
## voreherbi   -0.4203690  1.2542630  -0.335   0.7385
## voreinsecti  4.4344392  2.1571566   2.056   0.0435 *
## voreomni     0.4191080  1.3754228   0.305   0.7615
## bodywt      -0.0016268  0.0006155  -2.643   0.0101 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.291 on 71 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.1671, Adjusted R-squared:  0.1202
## F-statistic: 3.562 on 4 and 71 DF,  p-value: 0.0105
```

Table - 7

```
##         fit       lwr      upr
## 1  9.943532  8.396062 11.49100
## 2 10.015112  8.455318 11.57491
## 3 10.525606  8.559731 12.49148
## 4 10.823679  8.909093 12.73827
## 5 14.908963 11.082772 18.73515
```

Figure - 5

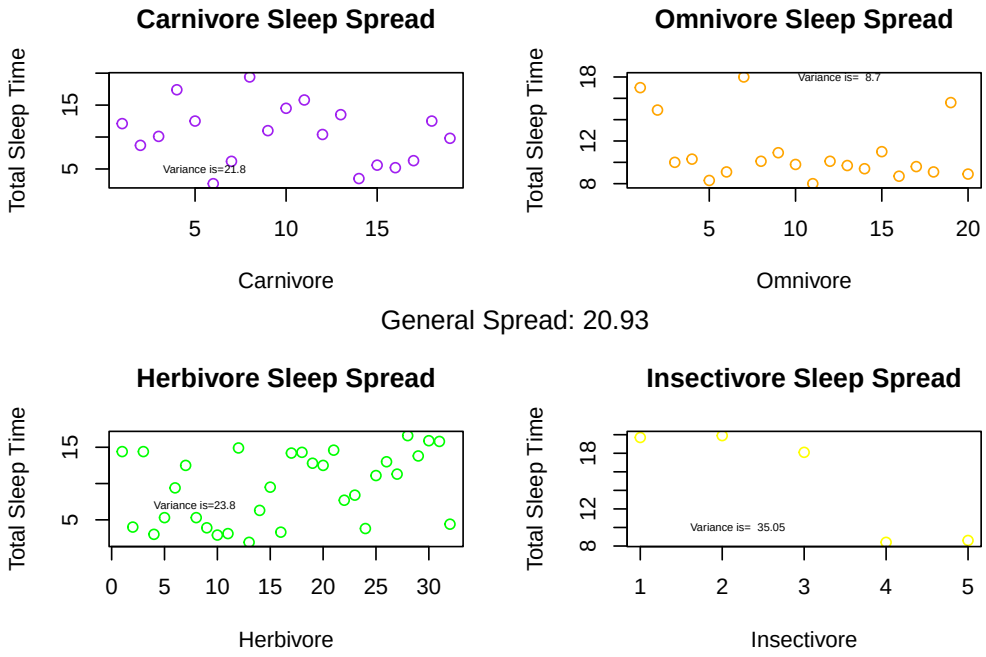

Figure 5: Sleep Spread vs -vore type plot

Figure - 6

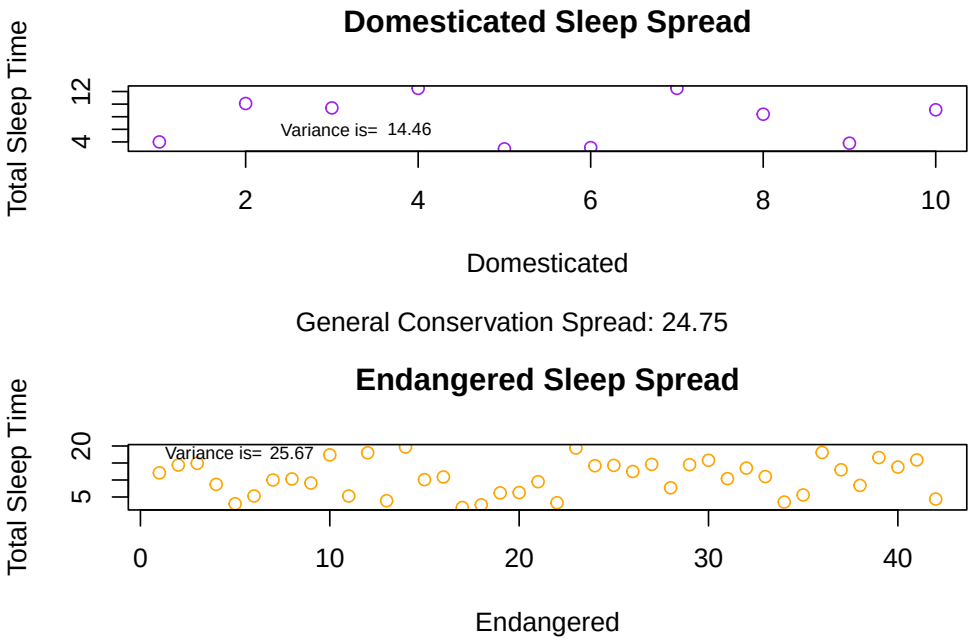

Figure 6: Sleep Spread vs Domesticated or not Domesticated(conservation) type plot

Table - 8

```
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## q4$conservation  5  370.9   74.18     3.9 0.00479 **
## Residuals       48  913.0   19.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 29 observations deleted due to missingness
```