**CSC3060 AIDA – Assignment 3**

##### Dewei Liu

##### 40216004

##### Wednesday, 18 December 2019

# Contents

[Contents 2](#_Toc27421988)

[Introduction 3](#_Toc27421989)

[Section 1 4](#_Toc27421990)

[Question 1.1 4](#_Toc27421991)

[Objective 4](#_Toc27421992)

[Reasoning 4](#_Toc27421993)

[Implementation 4](#_Toc27421994)

[Result 4](#_Toc27421995)

[Question 1.2 6](#_Toc27421996)

[Question 1.3 7](#_Toc27421997)

[Question 1.4 8](#_Toc27421998)

[Question 1.5 9](#_Toc27421999)

[Section 2 10](#_Toc27422000)

[Section 2.1 10](#_Toc27422001)

[Section 3 11](#_Toc27422002)

[Section 3.1 11](#_Toc27422003)

[Conclusions 12](#_Toc27422004)

[References 13](#_Toc27422005)

# Introduction

In this document, please replace [StudentNumber], [FNAME], etc, with the appropriate values. The FirstNme and LastName on the report should match your first name and lastname as it appears on QOL and QSIS.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

# Section 1

## Task 1.1

### Objective

The objective of this task is to differentiate living and non-living things using the feature verticalness.

### Assumption

The critical p-value is set as

### Reasoning

Logistic Regression (LR) will be used as a method in the analysis. LR uses the Sigmoid function (Equation 1), and as a result, it produces values between 0 and 1 (Chandrayan, 2019).

Equation 1 Logistic Regression (Devereux, 2019)

Since the objects are needed to be classifies into two classes, we can set a cut-off value. If the LR model produces a value which is greater than the cut-off value, the object will be identified as a class. Otherwise, it will be identified as the opposite class.

### Implementation

The data frame with two columns verticalness and living is constructed. The values in the column living are Boolean values indicating if the observation is a living thing.

The data then is fit into the LR models. By interpreting the result of the trained model, we can decide if the feature verticalness is a sufficient feature to differentiate living and non-living things.

### Result

The summary of the data fitted into the model is as Table 1.

verticalness living

Min. :0.07534 Min. :0.0

1st Qu.:0.36631 1st Qu.:0.0

Median :0.50616 Median :0.5

Mean :0.51907 Mean :0.5

3rd Qu.:0.61048 3rd Qu.:1.0

Max. :1.27027 Max. :1.0

Table 1 Summary of verticalness ~ logistic living value data

After fitting the model, the result of the model is as Table 1.

Call:

glm(formula = living ~ verticalness, family = "binomial", data = data)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.23105 -1.16944 0.00095 1.17767 1.19216

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.08656 0.37516 -0.231 0.818

verticalness 0.16676 0.65547 0.254 0.799

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 221.81 on 159 degrees of freedom

Residual deviance: 221.74 on 158 degrees of freedom

AIC: 225.74

Number of Fisher Scoring iterations: 3

Table 2 Result of Linear Regression Model

For the Intercept value in the table, the estimate is , which means the model predicts the value of living is given the verticalness value is 0. The z-score is , which is calculated as . It shows the estimate is standard error away from 0. According to the z-score and the degrees of freedom value, p-value of this variable is calculated to be which is larger than the critical p-value. We consider rejecting the hypothesis that intercept value is differ from 0.

For the verticalness value in the table, the estimate is , which means if the verticalness value increases by 1 unit, the predicted value of living will be increased by 0.16676 unit. The z-score is , which is calculated as . It shows the estimate is standard error away from 0. According to the z-score and the degrees of freedom value, p-value of this variable is calculated to be which is larger than the critical p-value. Thus, for the hypothesis, we stick on the hypothesis that the slope value of verticalness value is equal to 0.

According to the result of the model, the coefficient of the estimates of the intercept and verticalness is and , respectively, which derives the Equation 2 and Figure 1.

Equation 2 LR Model

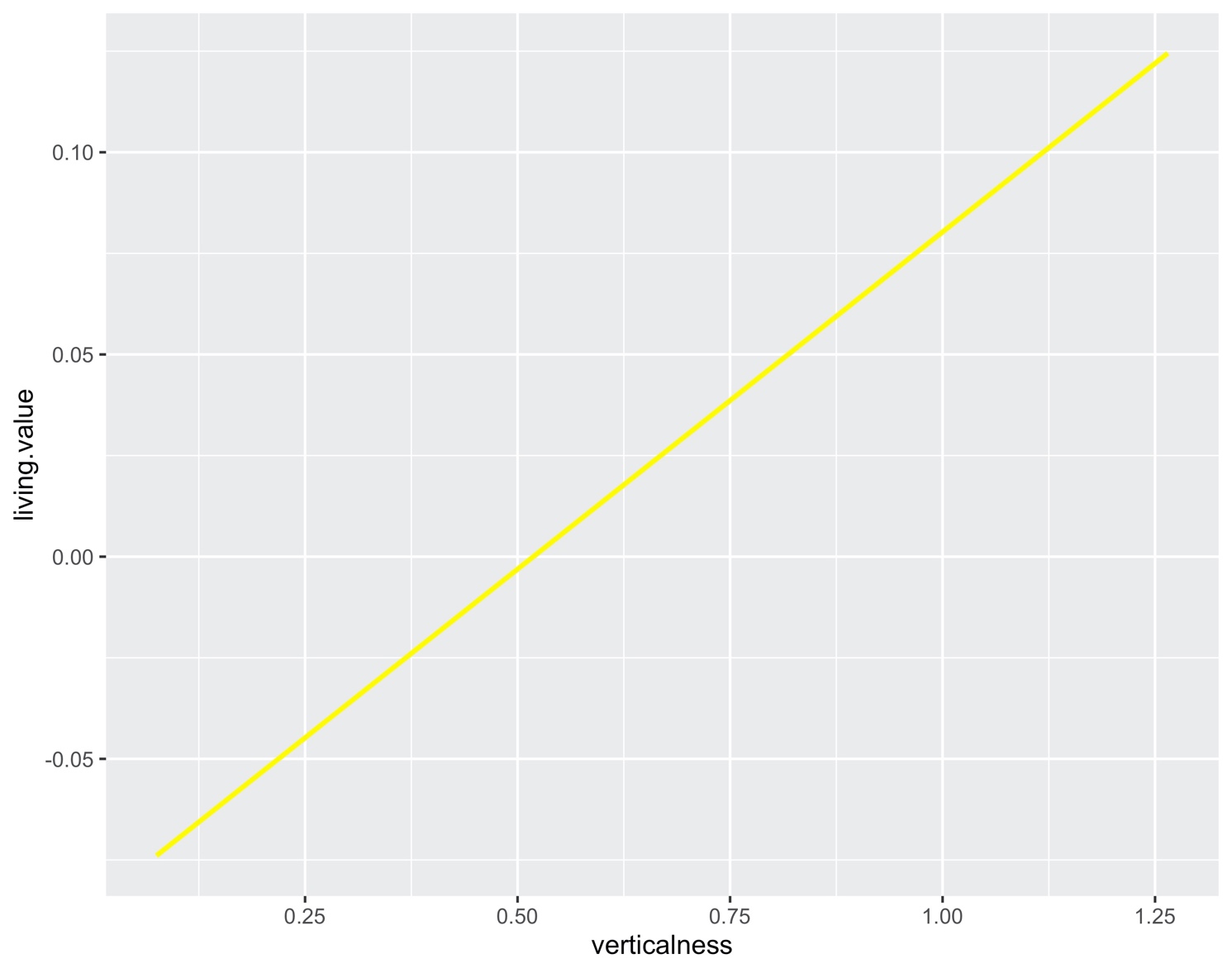


Figure 1 Regression Line living.value ~ verticalness value

## Task 1.2

### Objective

The objective of this task is to create a classifier to differentiate living objects.

### Reasoning

To create the classifier, we need to draw plots to visualise the data, and see how they are distributed according to the verticalness values.

### Implementation

We first draw a histogram to visualise the verticalness distributions of living and non-living objects (Figure 2). We can see there are more living objects with verticalness values between 0.375 and 0.75. However, they do not have clear separation on the feature verticalness.

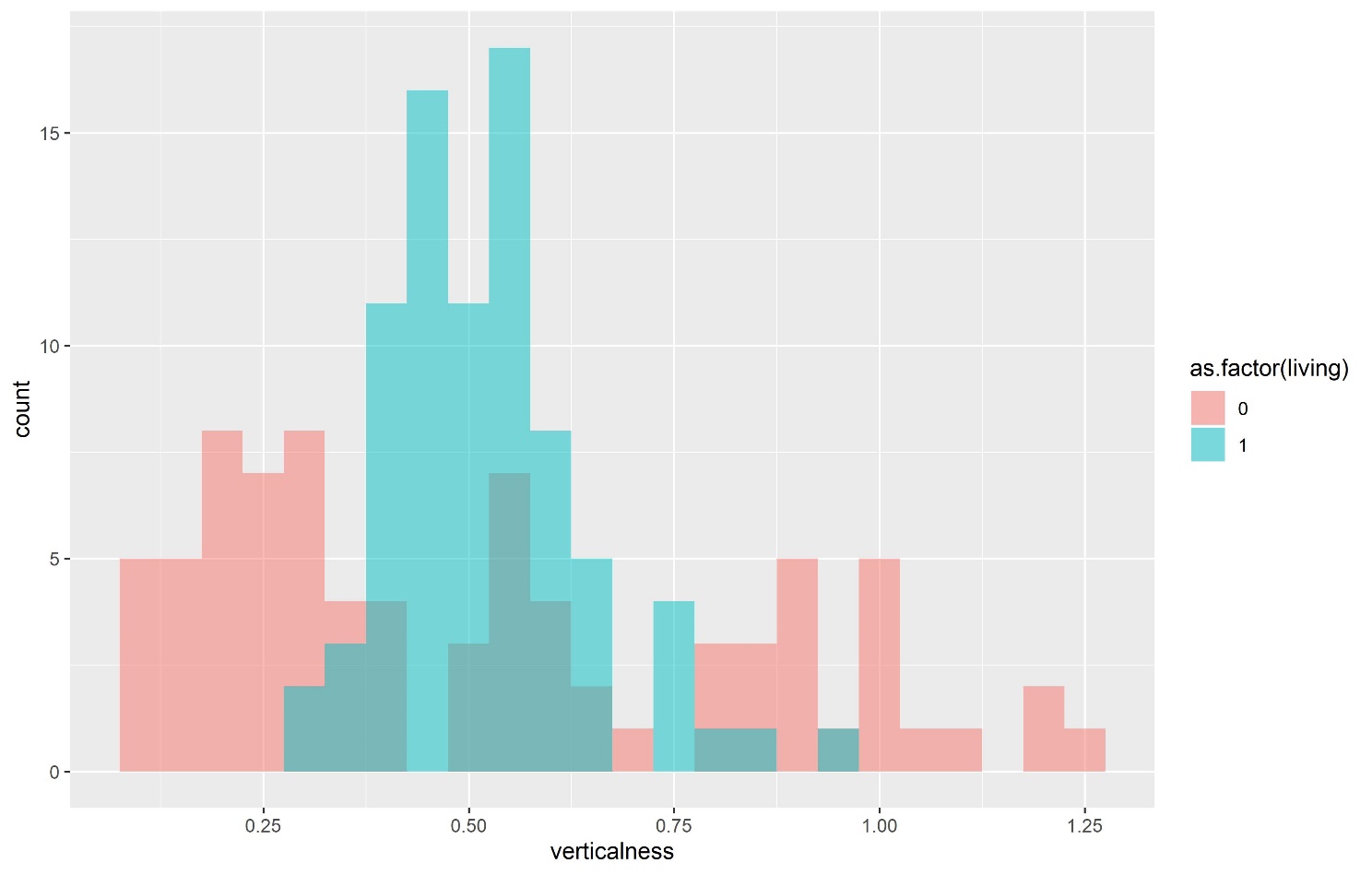


Figure 2 Histogram verticalness distribution

We then plot the training data points and a fitted curve (Figure 3). We can see the slop of the fitted curve is too horizontal, which means there is no strong correlation between these two variables, and it is hard to get valuable information from this figure.

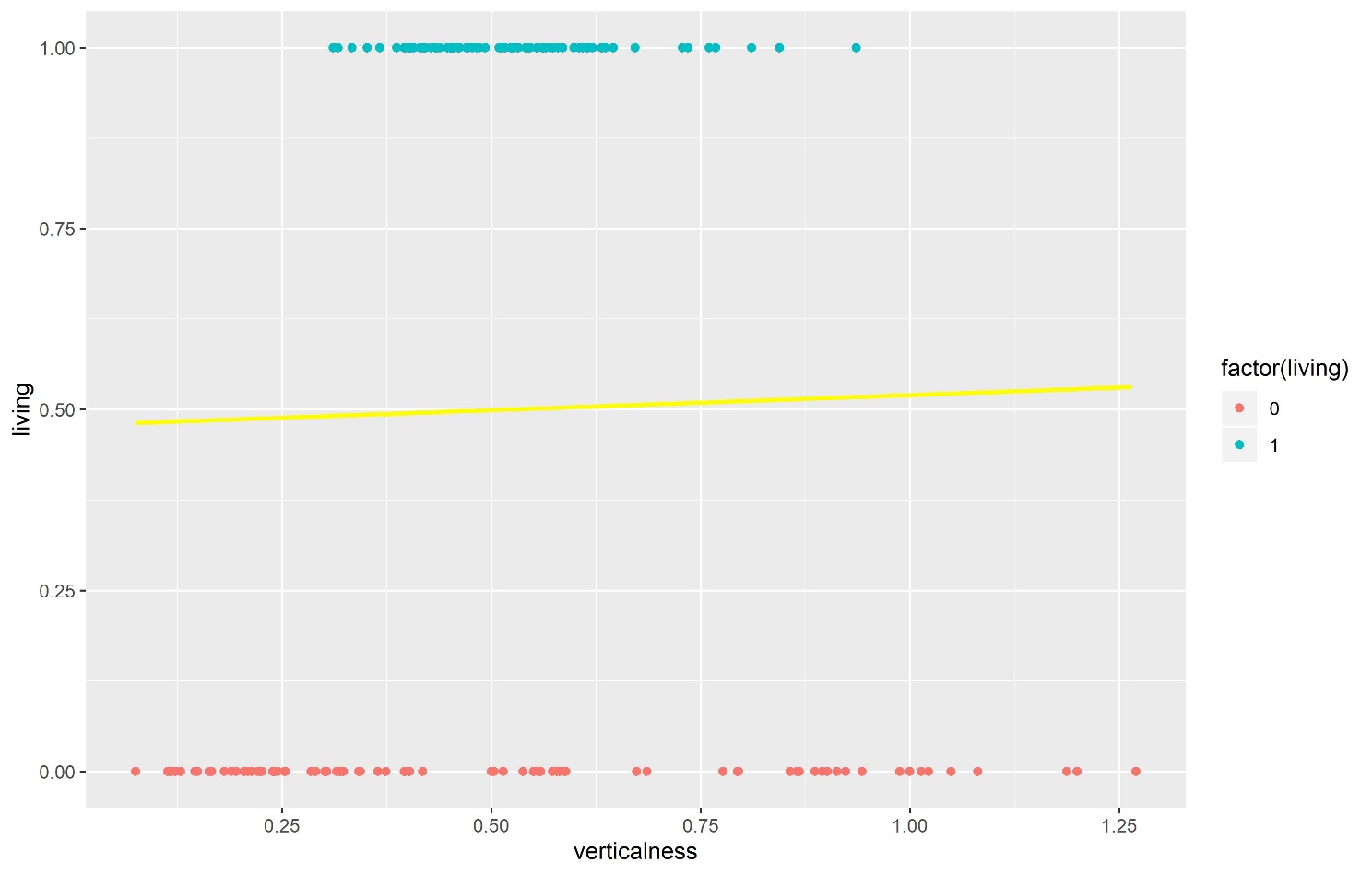


Figure 3 Veticalness ~ living: training data and fitted curve

It is hard to observe in Figure 3, but from Figure 2, we consider observations with verticalness value greater than 0.375 or higher to be living objects. We calculate the cut-off point of the P(X) using the formula.

And we got the cut-off value.

The observations with predicted value higher than the cut-off value are classified as living objects.

The result of the prediction is Table 3 showing 112 predictions were correct and 48 were incorrect. Thus, we got 70% correctness for this classifier.

Mode FALSE TRUE

logical 48 112

Table 3 Summary of Correct Predictions

## Task 1.3

According to Table 4, Table 5 and the summary from assignment 2 (Liu, 2019), we believe the three features *top2tile*, *bottom2tile* and *horizontalness* have the best separations on living and non-living objects among all features.

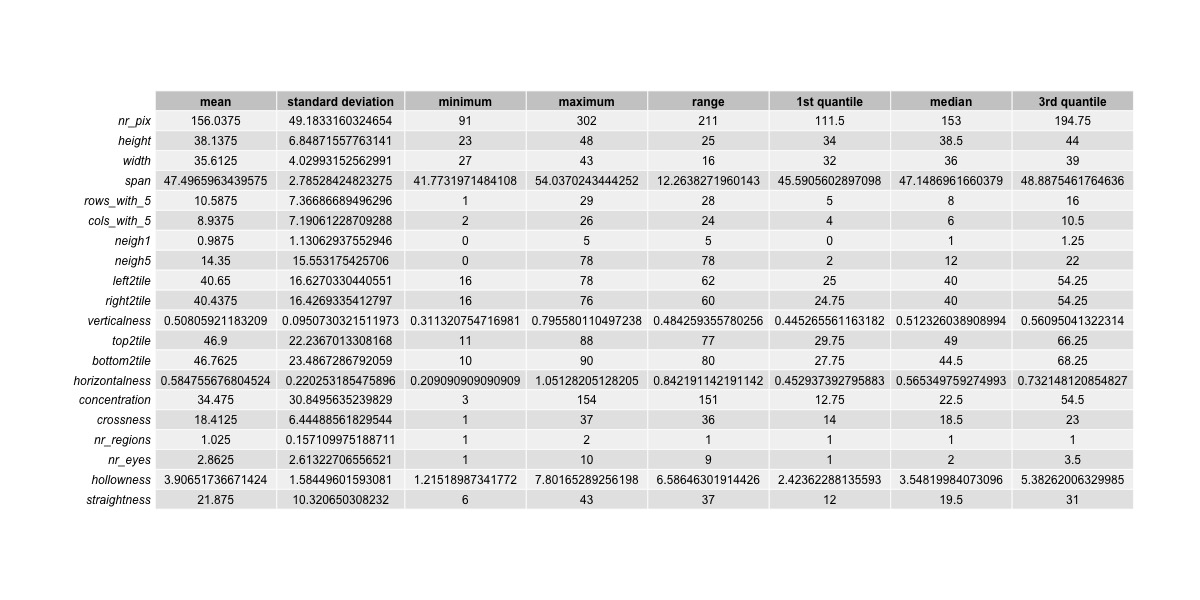


Table 4 Summary statics for living things (Liu, 2019)

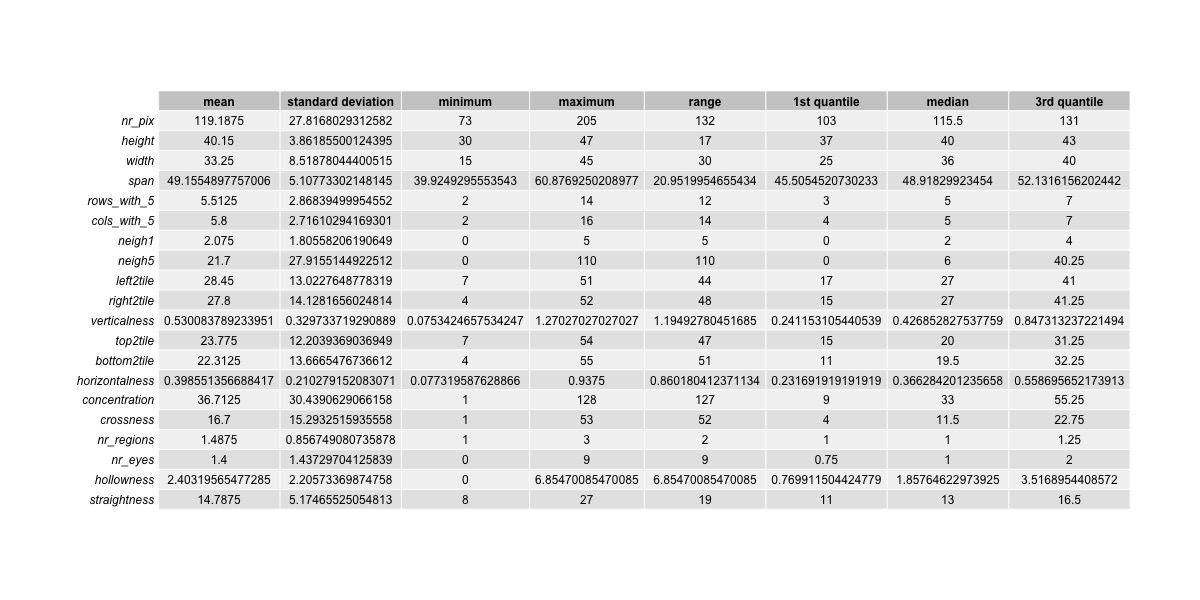


Table 5 Summary statics for non-living things (Liu, 2019)

The following figures imported from assignment 2 (Liu, 2019) shows the separations between living and non-living objects on these three features. Although the separations are not very clear, but we can still consider using these features to perform the logistic regression.

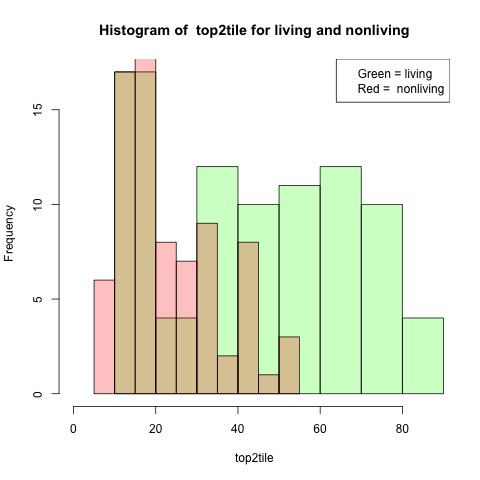


Figure 4 Histogram of top2tile for living and non-living (Liu, 2019)

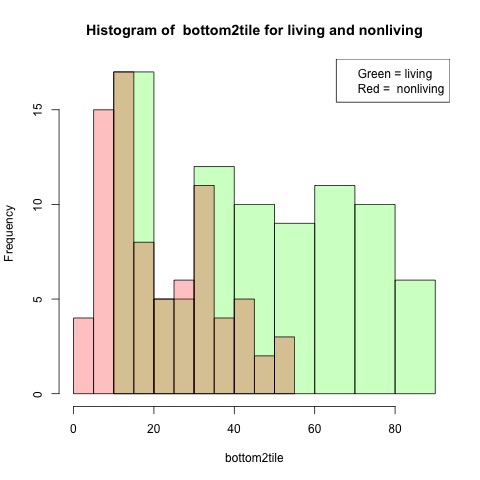


Figure 5 Histogram of bottom2tile for living and non-living (Liu, 2019)

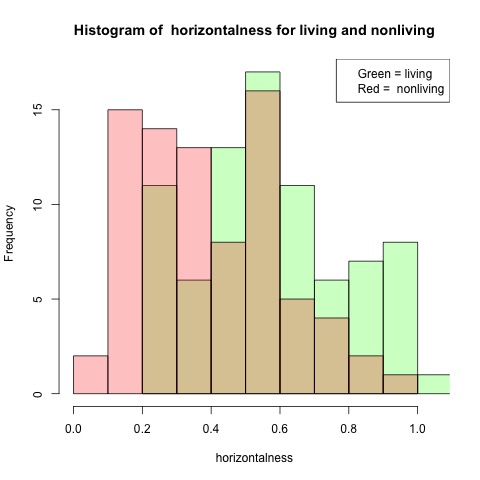


Figure 6 Histogram of horizontalness for living and non-living (Liu, 2019)

## Question 1.4

## Question 1.5

# Section 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## Section 2.1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum

# Section 3

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## Section 3.1

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est labour

# Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

# References

Chandrayan, P. (2019, August 15). *Logistic Regression For Dummies: A Detailed Explanation* . Retrieved from Towardsdatascience: https://towardsdatascience.com/logistic-regression-for-dummies-a-detailed-explanation-9597f76edf46

Devereux, B. (2019, December 16). *160 Topic 16 - Intro to Machine Learning.pptx .* Retrieved from Canvas: https://canvas.qub.ac.uk/courses/8433/files/487509?module\_item\_id=205271