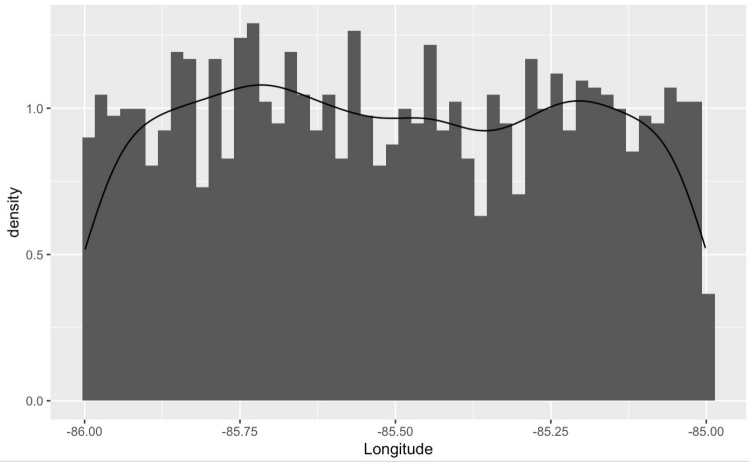
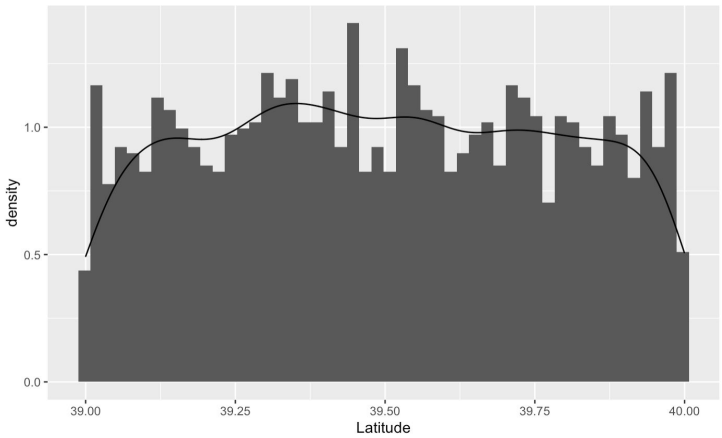
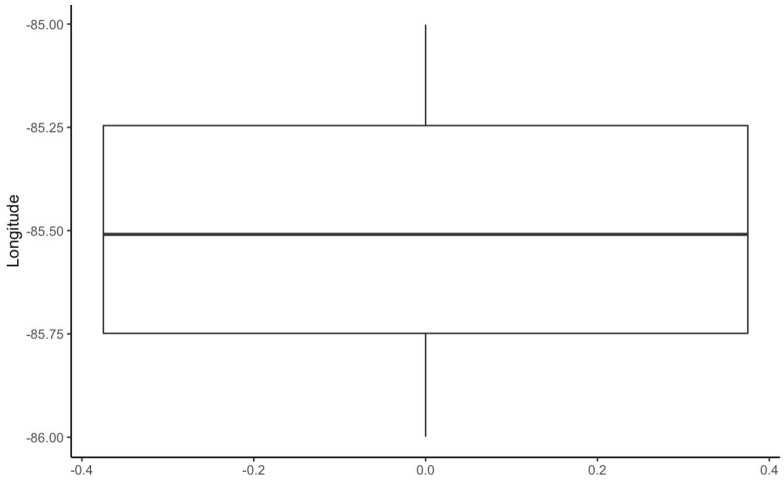
Xtern_Class of 2021
Data Science Assessment

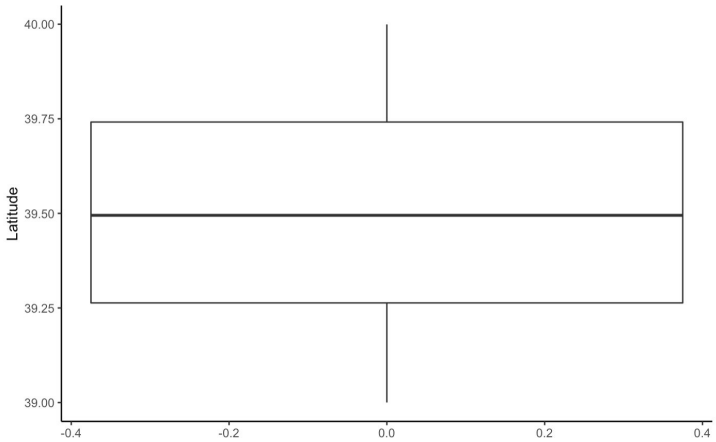
1. **Latitude & Longitude:**

- Using histogram & boxplot to show the density and range of Latitude and Longitude
  -> Conclusion 1: From the graphs, the range of Latitude is from 39.25 to 39.75 while the range of Longitude is from -85.75 to -85.25









2. **Regression between Rating, Votes & Reviews:**

- Multivariable linear regression: Rating ~ Votes & Reviews
  -> Conclusion 2.1: The adjusted R-squared, only 18% of Ratings is explained by Votes & Reviews. Therefore, the correlation among rating, votes and reviews is insignificant.
- Single linear regression: Rating ~ Votes; Rating ~ Reviews
  -> Conclusion 2.2: Both adjusted R-squares are also lower than 20% -> cannot state the regression

```
Call:
lm(formula = Rating ~ Votes + Reviews, data = model1_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.74742 -0.24185  0.03402  0.24947  1.15725

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.538e+00  1.058e-02 334.483  < 2e-16 ***
Votes        3.224e-04  6.300e-05   5.118 3.47e-07 ***
Reviews     -1.532e-06  1.081e-04  -0.014    0.989
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3777 on 1601 degrees of freedom
Multiple R-squared:  0.189,     Adjusted R-squared:  0.188
F-statistic: 186.5 on 2 and 1601 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Rating ~ Votes, data = model1_data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7496 -0.2419  0.0340  0.2494  1.1571

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.538e+00  1.037e-02  341.03   <2e-16 ***
Votes       3.216e-04  1.664e-05   19.32   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3776 on 1602 degrees of freedom
Multiple R-squared:  0.189,     Adjusted R-squared:  0.1885
F-statistic: 373.3 on 1 and 1602 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Rating ~ Reviews, data = model1_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.31334 -0.25554  0.03648  0.24406  1.12159

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.554e+00  1.019e-02  348.88   <2e-16 ***
Reviews     5.319e-04  2.878e-05   18.48   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3806 on 1602 degrees of freedom
Multiple R-squared:  0.1757,    Adjusted R-squared:  0.1752
F-statistic: 341.5 on 1 and 1602 DF,  p-value: < 2.2e-16
```

3. High-quality restaurants has 4.0 (maximum) rating:

- Conclusion 3:

+ There are a total 112 restaurants which have a maximum rating of 4.0. Out of 2019 restaurants in the original dataset, this number accounts for 5%.
+ Among 112 restaurants with 4.0 rating, restaurants serving Indian food account for 35 % (39 restaurants), restaurants serving Chinese food account for 25% (28 restaurants).

4. The popularity of Indian & Chinese Cuisine in Indianapolis

All in all, we can say, Indian and Chinese cuisine are the two main parts of Indianapolis restaurant type, accounting for 16.8% and 12.5% of restaurants in Indy (in total). 10% out of each cuisine restaurant serves high quality food and has good customer service as they have 4.0 (maximum) rating.

Conclusion 4.1: Indian Cuisine

+ There are 339 restaurants serving Indian food in Indianapolis (out of 2019 restaurants). However, we only have 39 high quality (rate 4.0) restaurants accounting for only 11.5%

Conclusion 4.2: Chinese Cuisine

+ There are 252 restaurants serving Chinese food in Indianapolis (out of 2019 restaurants). However, we only have 28 high quality (rate 4.0) restaurants accounting for only 11.1%