

Group_9_Analysis

Brent Strong, Enyu Li, Haotian Wang, Honjin Ren, Mu He

3/7/2022

#Setup and Data Import

Rows: 1,145

Columns: 8

```
$ country_of_origin <chr> "Myanmar", "Uganda", "Ethiopia", "Mexico", "Burun~  
$ aroma <dbl> 7.25, 8.33, 8.42, 7.17, 7.75, 7.92, 7.92, 7.83, 7~  
$ flavor <dbl> 7.42, 7.92, 8.00, 7.08, 7.67, 7.75, 7.83, 7.67, 6~  
$ acidity <dbl> 7.50, 7.92, 8.00, 7.25, 7.50, 7.75, 7.67, 7.58, 7~  
$ category_two_defects <dbl> 4, 1, 7, 3, 5, 0, 1, 2, 2, 1, 0, 8, 0, 2, 0, 0, 2~  
$ altitude_mean_meters <dbl> 1219.20, 1600.00, 1700.00, 1300.00, 1880.00, 1400~  
$ harvested <dbl> 2015, 2013, 2014, 2012, 2012, 2014, NA, 2015, 201~  
$ Qualityclass <chr> "Poor", "Good", "Good", "Poor", "Good", "Good", "~
```

[1] 0.5135371

#Exploratory Data Analysis

Table 1: Summary statistics of continuous variables in the data set.

Variable	Mean	SD	Min.	1st Q.	Median	3rd Q.	Max.
aroma	7.57	0.39	0	7.42	7.58	7.75	8.75
flavor	7.52	0.40	0	7.33	7.58	7.75	8.67
acidity	7.54	0.39	0	7.33	7.50	7.75	8.58
category_two_defects	3.67	5.41	0	0.00	2.00	5.00	55.00
altitude_mean_meters	1850.69	9392.09	1	1100.00	1310.64	1600.00	190164.00
harvested	2013.67	1.81	2010	2012.00	2014.00	2015.00	2018.00

Table 2: Summary statistics of the sepal length by species of irises

country_of_origin	number_of_batch	Proportion_of_good_quality
Brazil	116	0.47
Burundi	2	0.50
China	14	0.64
Colombia	158	0.80
Costa Rica	41	0.56
Cote d'Ivoire	1	0.00
Ecuador	3	0.33
El Salvador	20	0.70
Ethiopia	38	0.92
Guatemala	152	0.50
Haiti	5	0.20
Hawaii	62	0.55
Honduras	48	0.25
India	10	0.50
Indonesia	16	0.56
Japan	1	1.00
Kenya	24	0.92
Laos	2	0.00
Malawi	11	0.09
Mauritius	1	0.00
Mexico	203	0.27
Myanmar	6	0.00
Nicaragua	23	0.22
Panama	4	0.75
Peru	9	0.56
Philippines	5	0.40
Puerto Rico	3	0.33
Taiwan	62	0.42
Tanzania	32	0.50
Thailand	23	0.70
Uganda	32	0.78
United States	9	0.67
Vietnam	8	0.50
Zambia	1	0.00

Table 3: Summary statistics of the sepal length by species of irises

Variable	Qualityclass	n	Mean	SD	Min	Median	Max	IQR
aroma	Good	588	7.76	0.23	7.08	7.75	8.75	0.08
aroma	Poor	557	7.37	0.41	0.00	7.42	8.25	0.16
flavor	Good	588	7.74	0.23	7.00	7.67	8.67	0.16
flavor	Poor	557	7.29	0.42	0.00	7.33	8.08	0.17
acidity	Good	588	7.72	0.25	6.75	7.67	8.58	0.16
acidity	Poor	557	7.34	0.40	0.00	7.33	8.33	0.17
category_two_defects	Good	588	2.87	3.82	0.00	2.00	40.00	2.00
category_two_defects	Poor	557	4.52	6.60	0.00	2.00	55.00	4.00
harvested	Good	588	2013.74	1.90	2010.00	2014.00	2018.00	1.00
harvested	Poor	557	2013.59	1.71	2010.00	2013.00	2018.00	2.00

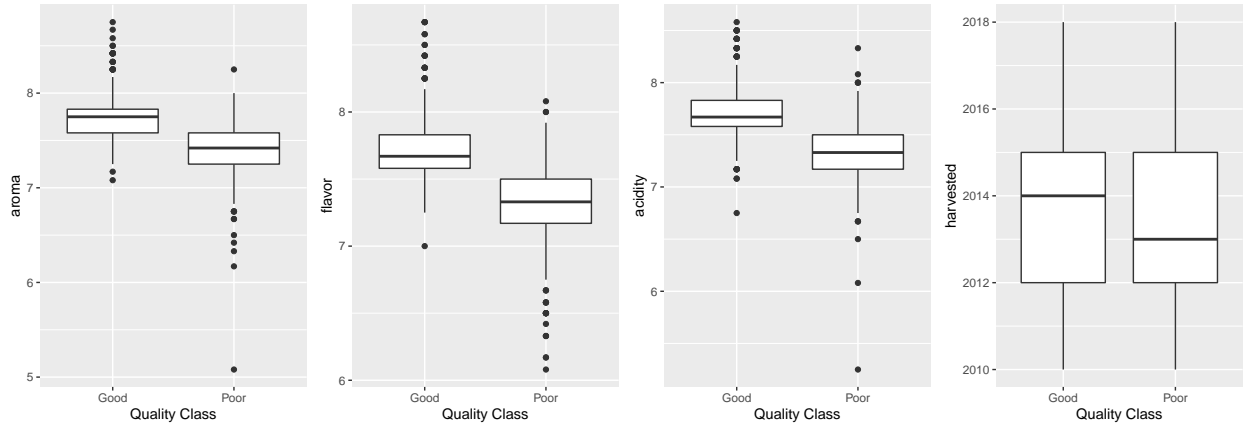


Figure 1: Boxplots of countinuous features on different quality class.

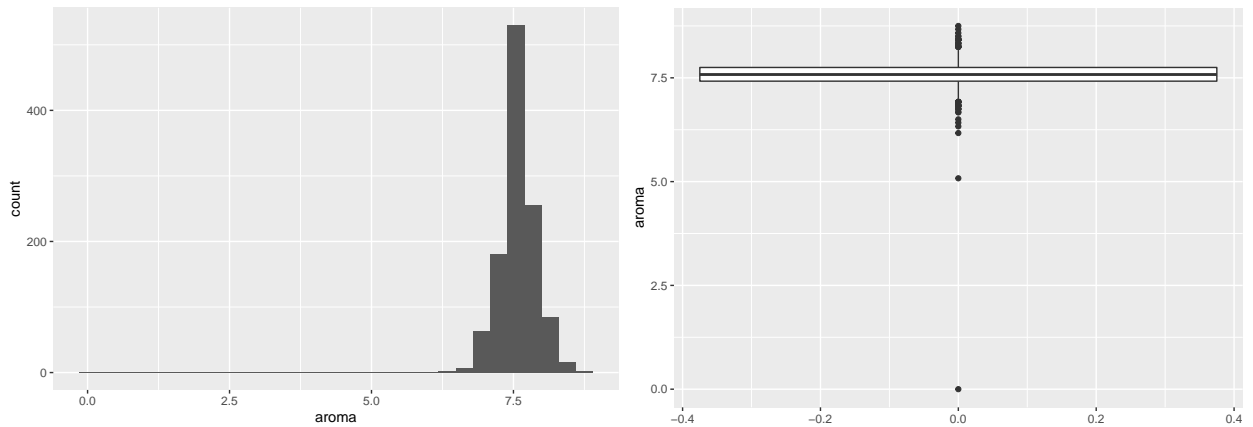


Figure 2: Histogram and boxplot for aroma.

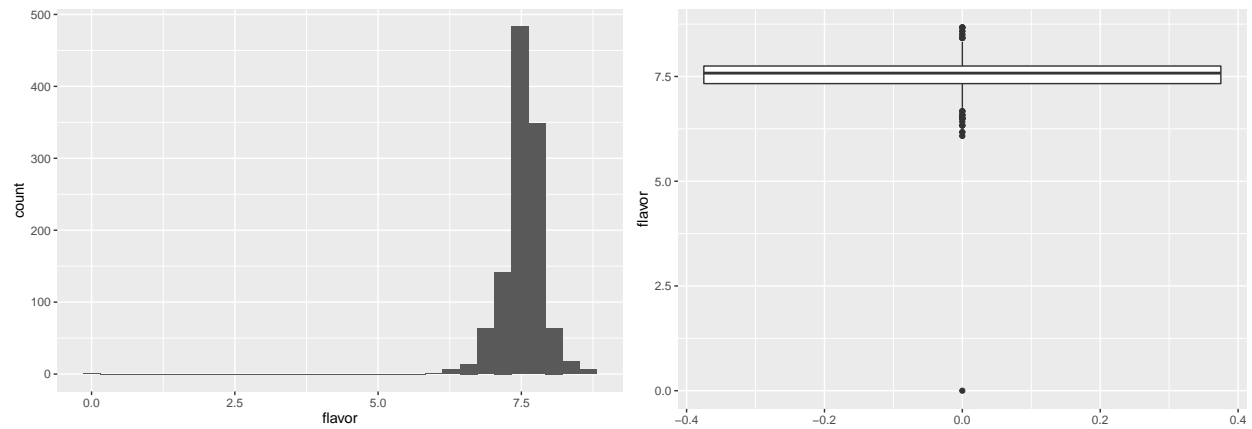


Figure 3: Histogram and boxplot for flavor.

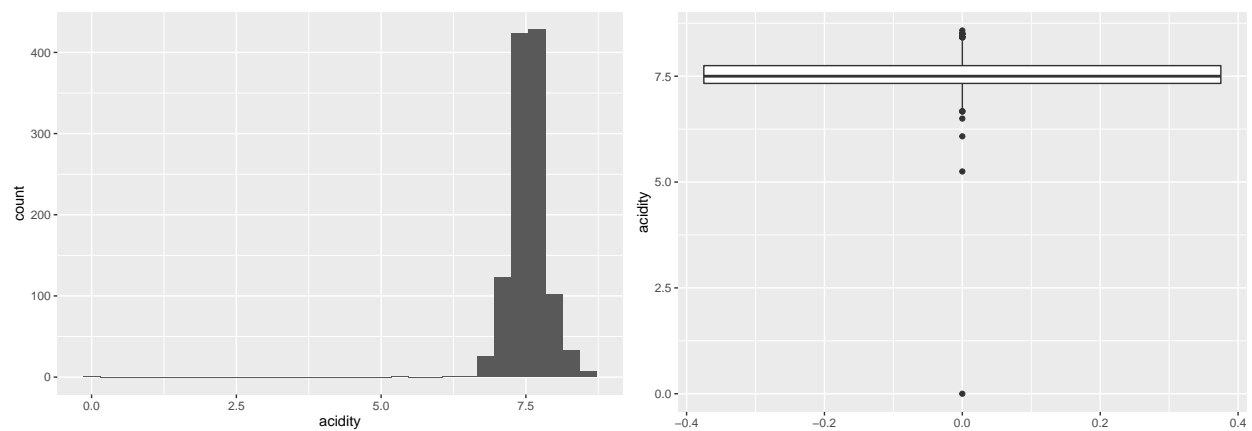


Figure 4: Histogram and boxplot for acidity.

```
# A tibble: 1 x 8
  country_of_origin aroma flavor acidity category_two_defects altitude_mean_met~
  <chr>          <dbl> <dbl> <dbl>          <dbl>          <dbl>
1 Honduras            0      0      0              2          1400
# ... with 2 more variables: harvested <dbl>, Qualityclass <chr>
```

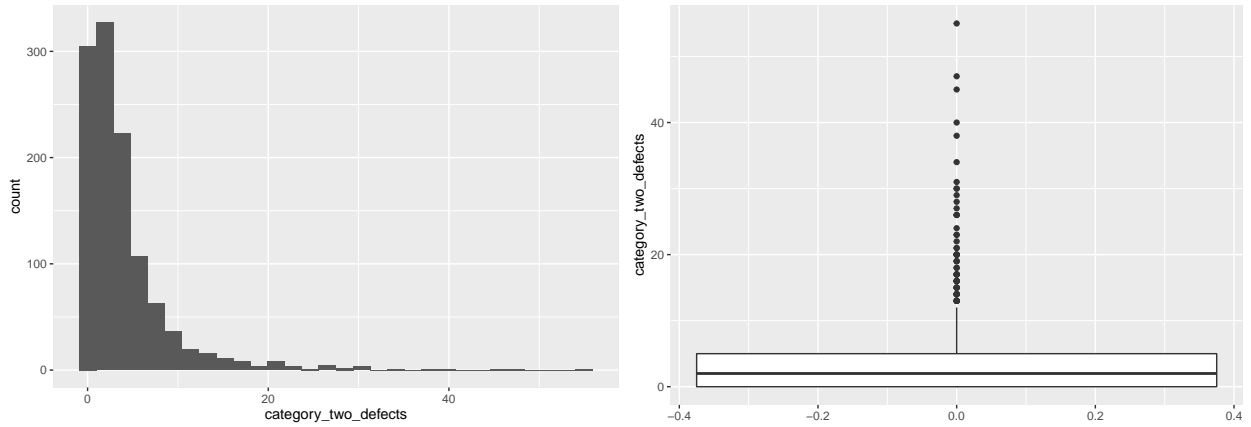


Figure 5: Histogram and boxplot for category two defects.

```

Rows: 1,145
Columns: 9
$ country_of_origin    <chr> "Myanmar", "Uganda", "Ethiopia", "Mexico", "Burun~
$ aroma                <dbl> 7.25, 8.33, 8.42, 7.17, 7.75, 7.92, 7.92, 7.83, 7~
$ flavor              <dbl> 7.42, 7.92, 8.00, 7.08, 7.67, 7.75, 7.83, 7.67, 6~
$ acidity             <dbl> 7.50, 7.92, 8.00, 7.25, 7.50, 7.75, 7.67, 7.58, 7~
$ category_two_defects <dbl> 4, 1, 7, 3, 5, 0, 1, 2, 2, 1, 0, 8, 0, 2, 0, 0, 2~
$ altitude_mean_meters <dbl> 1219.20, 1600.00, 1700.00, 1300.00, 1880.00, 1400~
$ harvested            <dbl> 2015, 2013, 2014, 2012, 2012, 2014, NA, 2015, 201~
$ Qualityclass         <chr> "Poor", "Good", "Good", "Poor", "Good", "Good", "~
$ defects_log          <dbl> 1.6094379, 0.6931472, 2.0794415, 1.3862944, 1.791~

```

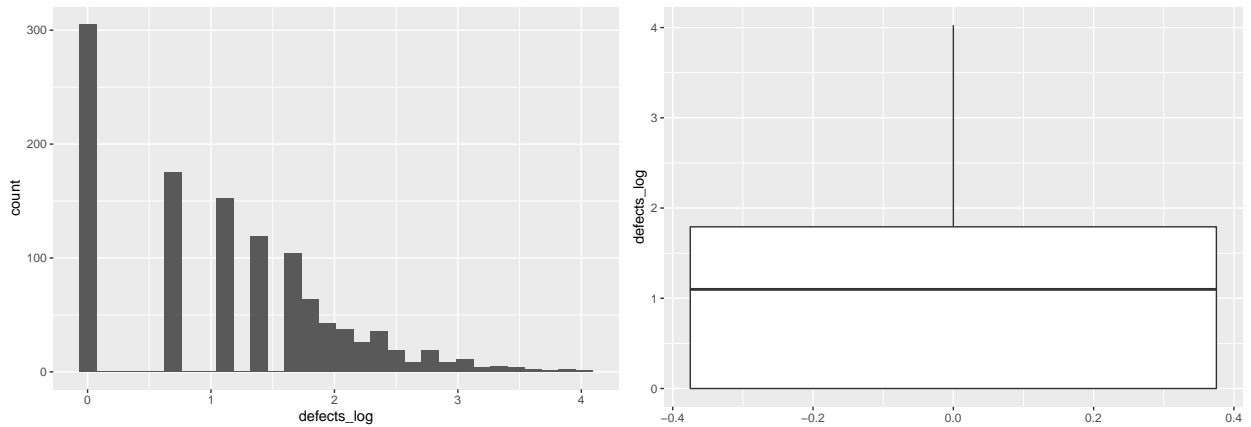


Figure 6: Histogram and boxplot for category two defects after log transformation.

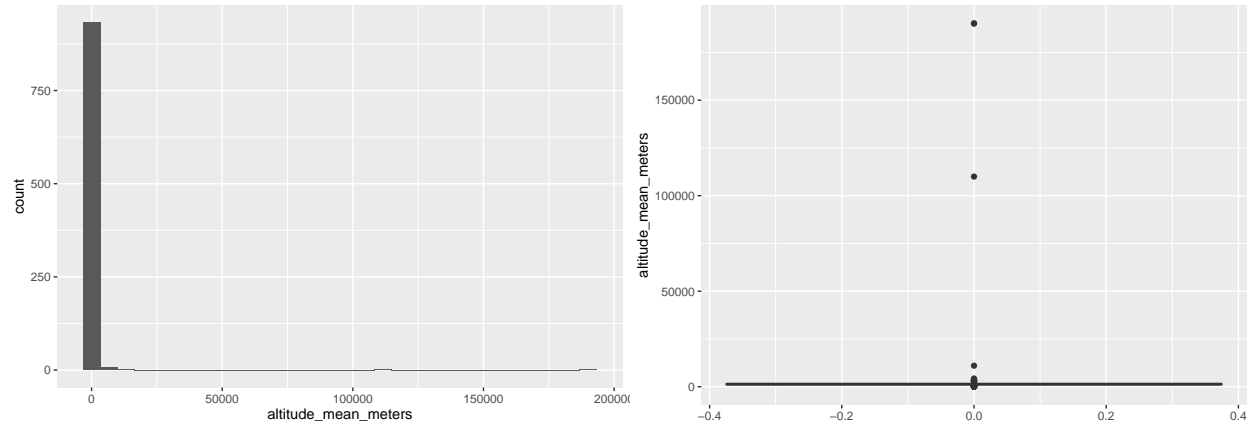


Figure 7: Histogram and boxplot for altitude.

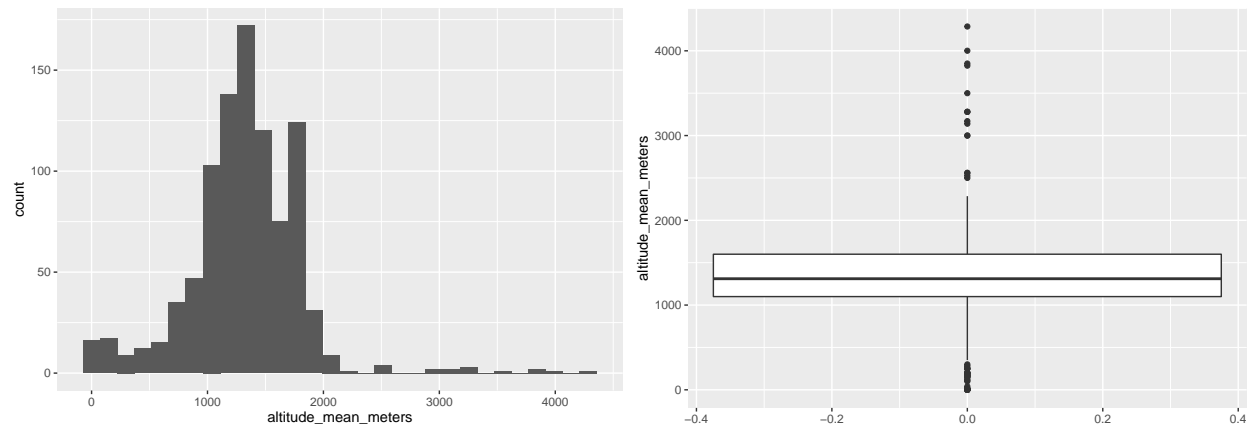


Figure 8: Histogram and boxplot for altitude after removing implausible observations.

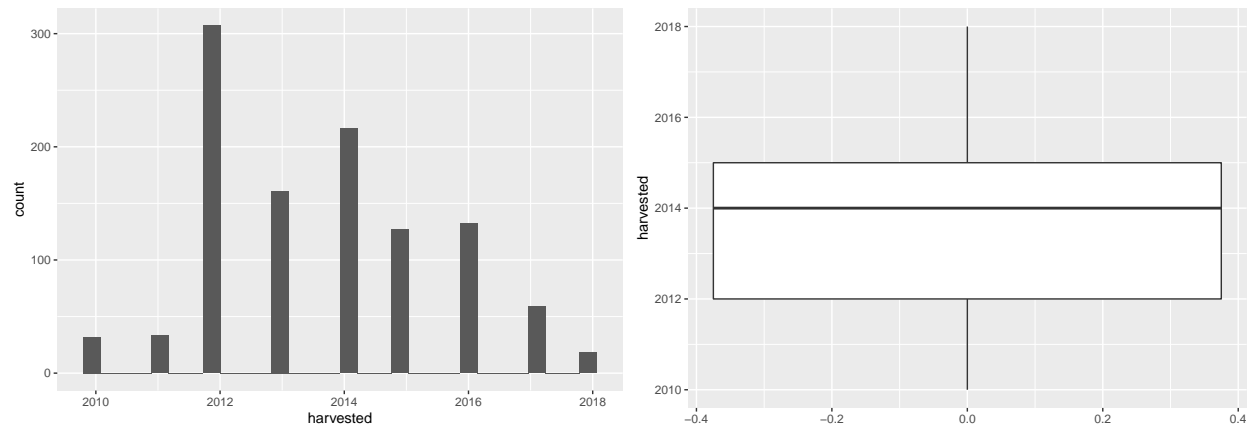
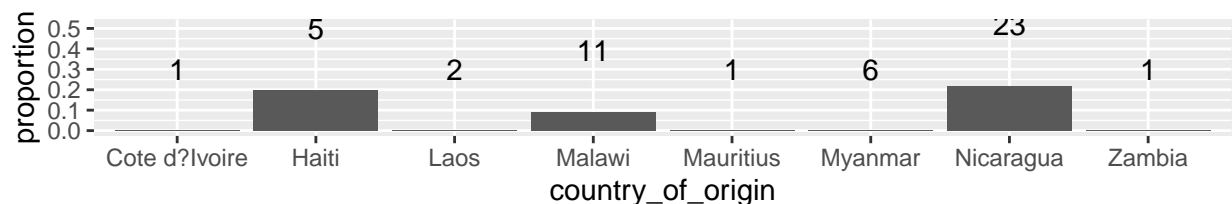
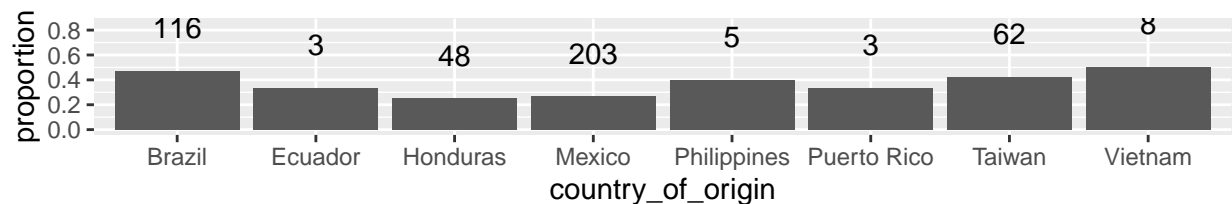
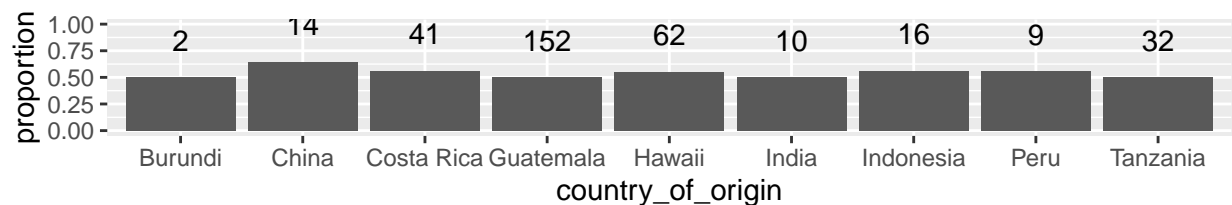
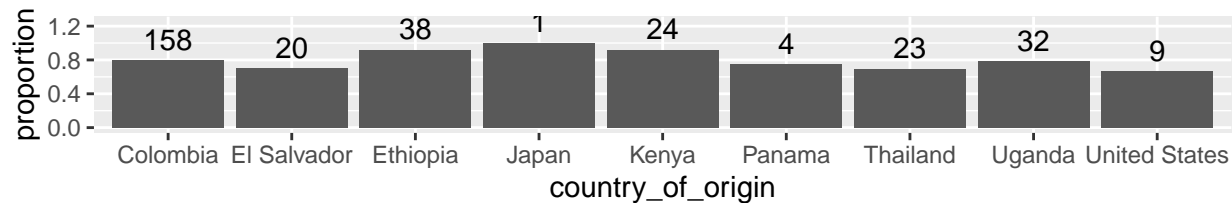


Figure 9: Histogram and boxplot for harvested.



Rows: 1,140

Columns: 8

```
$ country_of_origin <chr> "Myanmar", "Uganda", "Ethiopia", "Mexico", "Burundi"~
$ aroma             <dbl[,1]> <matrix[26 x 1]>
$ flavor            <dbl[,1]> <matrix[26 x 1]>
$ acidity           <dbl[,1]> <matrix[26 x 1]>
$ defects_log       <dbl> 1.6094379, 0.6931472, 2.0794415, 1.3862944, 1.79~
$ year              <fct> 2015, 2013, 2014, 2012, 2012, 2014, NA, 2015, 20~
$ level             <chr> "3", "3", "3", "3", "3", "3", NA, "3", "3", "3",~
$ Qualityclass      <dbl> 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1~
```

A tibble: 10 x 2

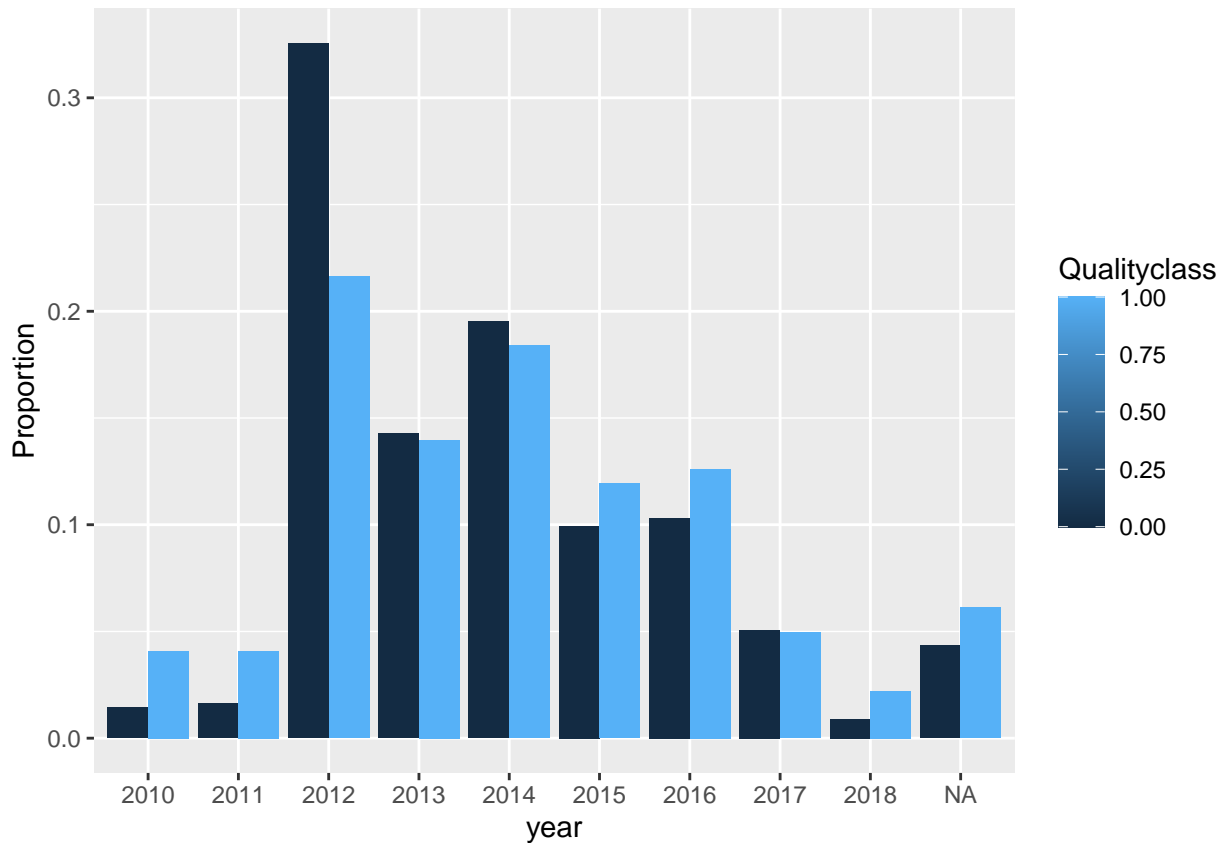
```
  year      n
  <fct> <int>
1 2010     32
2 2011     33
3 2012    307
4 2013    161
5 2014    216
6 2015    125
7 2016    131
8 2017     57
9 2018     18
10 <NA>    60
```

Qualityclass Brazil Burundi China Colombia Costa Rica Cote d'Ivoire

	0	11.0% (61)	0.2% (1)	0.9% (5)	5.8% (32)	3.3% (18)	0.2% (1)
	1	9.2% (54)	0.2% (1)	1.5% (9)	21.5% (126)	3.9% (23)	0.0% (0)
Ecuador	El Salvador	Ethiopia	Guatemala	Haiti	Hawaii	Honduras	
0.4% (2)	1.1% (6)	0.5% (3)	13.4% (74)	0.7% (4)	5.1% (28)	6.3% (35)	
0.2% (1)	2.4% (14)	6.0% (35)	12.9% (76)	0.2% (1)	5.8% (34)	2.0% (12)	
India	Indonesia	Japan	Kenya	Laos	Malawi	Mauritius	Mexico
0.9% (5)	1.3% (7)	0.0% (0)	0.4% (2)	0.4% (2)	1.8% (10)	0.2% (1)	26.9% (149)
0.9% (5)	1.5% (9)	0.2% (1)	3.7% (22)	0.0% (0)	0.2% (1)	0.0% (0)	9.2% (54)
Myanmar	Nicaragua	Panama	Peru	Philippines	Puerto Rico	Taiwan	
1.1% (6)	3.1% (17)	0.2% (1)	0.7% (4)	0.5% (3)	0.4% (2)	6.5% (36)	
0.0% (0)	0.9% (5)	0.5% (3)	0.9% (5)	0.3% (2)	0.2% (1)	4.4% (26)	
Tanzania	Thailand	Uganda	United States	Vietnam	Zambia		
2.9% (16)	1.3% (7)	1.3% (7)	0.5% (3)	0.7% (4)	0.2% (1)		
2.7% (16)	2.7% (16)	4.3% (25)	1.0% (6)	0.7% (4)	0.0% (0)		

Qualityclass	2010	2011	2012	2013	2014	2015
0	1.4% (8)	1.6% (9)	32.5% (180)	14.3% (79)	19.5% (108)	9.9% (55)
1	4.1% (24)	4.1% (24)	21.6% (127)	14.0% (82)	18.4% (108)	11.9% (70)

	2016	2017	2018	NA_
10.3% (57)	5.1% (28)	0.9% (5)	4.3% (24)	
12.6% (74)	4.9% (29)	2.2% (13)	6.1% (36)	



Formal Analysis Using Logistic Regression

Call:


```
glm(formula = Qualityclass ~ level - 1, family = binomial(link = "logit"),
    data = coffee_final)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.294	-1.294	1.065	1.065	1.360

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
level1	-0.41871	0.18209	-2.299	0.021481	*
level2	-0.41376	0.14410	-2.871	0.004087	**
level3	0.26959	0.08158	3.305	0.000951	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1301.7 on 939 degrees of freedom
 Residual deviance: 1276.9 on 936 degrees of freedom
 (201 observations deleted due to missingness)
 AIC: 1282.9

Number of Fisher Scoring iterations: 4

Call:

```
glm(formula = Qualityclass ~ year, family = binomial(link = "logit"),
    data = coffee_final)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6651	-1.1774	0.7585	1.1616	1.3287

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.0986	0.4082	2.691	0.007123	**
year2011	-0.1178	0.5652	-0.208	0.834921	
year2012	-1.4474	0.4244	-3.411	0.000648	***
year2013	-1.0613	0.4376	-2.425	0.015300	*
year2014	-1.0986	0.4303	-2.553	0.010682	*
year2015	-0.8575	0.4462	-1.921	0.054671	.
year2016	-0.8376	0.4447	-1.884	0.059609	.
year2017	-1.0635	0.4867	-2.185	0.028872	*
year2018	-0.1431	0.6660	-0.215	0.829878	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1496.7 on 1079 degrees of freedom
 Residual deviance: 1464.8 on 1071 degrees of freedom
 (60 observations deleted due to missingness)
 AIC: 1482.8

Number of Fisher Scoring iterations: 4

```
Call:
glm(formula = Qualityclass ~ country_of_origin, family = binomial(link = "logit"),
    data = coffee_final)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2534	-1.0427	0.4056	1.0961	2.1899

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.1219	0.1868	-0.652	0.514176
country_of_originBurundi	0.1219	1.4265	0.085	0.931906
country_of_originChina	0.7097	0.5882	1.206	0.227646
country_of_originColombia	1.4924	0.2722	5.483	4.19e-08 ***
country_of_originCosta Rica	0.3670	0.3660	1.003	0.315955
country_of_originCote d'Ivoire	-15.4442	1455.3975	-0.011	0.991533
country_of_originEcuador	-0.5713	1.2389	-0.461	0.644731
country_of_originEl Salvador	0.9692	0.5225	1.855	0.063610 .
country_of_originEthiopia	2.5786	0.6299	4.093	4.25e-05 ***
country_of_originGuatemala	0.1486	0.2482	0.599	0.549414
country_of_originHaiti	-1.2644	1.1335	-1.115	0.264658
country_of_originHawaii	0.3160	0.3163	0.999	0.317681
country_of_originHonduras	-0.9486	0.3832	-2.476	0.013303 *
country_of_originIndia	0.1219	0.6595	0.185	0.853364
country_of_originIndonesia	0.3732	0.5375	0.694	0.487453
country_of_originJapan	15.6880	1455.3975	0.011	0.991400
country_of_originKenya	2.5198	0.7618	3.308	0.000941 ***
country_of_originLaos	-15.4442	1029.1215	-0.015	0.988026
country_of_originMalawi	-2.1807	1.0653	-2.047	0.040660 *
country_of_originMauritius	-15.4442	1455.3975	-0.011	0.991533
country_of_originMexico	-0.8931	0.2452	-3.642	0.000271 ***
country_of_originMyanmar	-15.4442	594.1636	-0.026	0.979263
country_of_originNicaragua	-1.1019	0.5420	-2.033	0.042043 *
country_of_originPanama	1.2205	1.1697	1.043	0.296757
country_of_originPeru	0.3450	0.6964	0.495	0.620259
country_of_originPhilippines	-0.2836	0.9318	-0.304	0.760875
country_of_originPuerto Rico	-0.5713	1.2389	-0.461	0.644731
country_of_originTaiwan	-0.2035	0.3180	-0.640	0.522203
country_of_originTanzania	0.1219	0.3999	0.305	0.760512
country_of_originThailand	0.9486	0.4902	1.935	0.052969 .
country_of_originUganda	1.3949	0.4667	2.989	0.002799 **
country_of_originUnited States	0.8150	0.7314	1.114	0.265113
country_of_originVietnam	0.1219	0.7314	0.167	0.867639
country_of_originZambia	-15.4442	1455.3975	-0.011	0.991533

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1579.4 on 1139 degrees of freedom
Residual deviance: 1352.0 on 1106 degrees of freedom
AIC: 1420

Number of Fisher Scoring iterations: 14

```
Call:
glm(formula = Qualityclass ~ Colombia + Mexico + Honduras + Kenya -
    1, family = binomial(link = "logit"), data = coffee_final)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.2293	-1.1774	0.4172	1.1774	1.6524

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
Colombia	1.3705	0.1980	6.923	4.41e-12	***
Mexico	-1.0150	0.1588	-6.390	1.66e-10	***
Honduras	-1.0704	0.3345	-3.200	0.00137	**
Kenya	2.3979	0.7382	3.248	0.00116	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1580.4 on 1140 degrees of freedom
Residual deviance: 1443.1 on 1136 degrees of freedom
AIC: 1451.1

Number of Fisher Scoring iterations: 4

```
Call:
glm(formula = Qualityclass ~ country_of_origin + year, family = binomial(link = "logit"),
    data = coffee_final)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.2542	-0.9979	0.4052	1.0378	2.1899

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-6.651e-03	5.136e-01	-0.013	0.989669	
country_of_originBurundi	2.191e-01	1.436e+00	0.153	0.878744	
country_of_originChina	7.420e-01	6.232e-01	1.191	0.233757	
country_of_originColombia	1.771e+00	3.042e-01	5.821	5.86e-09	***
country_of_originCosta Rica	4.790e-01	3.910e-01	1.225	0.220603	
country_of_originCote d'Ivoire	-1.550e+01	1.455e+03	-0.011	0.991502	
country_of_originEcuador	-4.227e-01	1.260e+00	-0.336	0.737240	
country_of_originEl Salvador	1.073e+00	5.391e-01	1.989	0.046648	*
country_of_originEthiopia	2.465e+00	6.611e-01	3.729	0.000192	***
country_of_originGuatemala	3.416e-01	2.798e-01	1.221	0.222123	
country_of_originHaiti	-8.781e-01	1.148e+00	-0.765	0.444300	
country_of_originHawaii	5.486e-01	3.821e-01	1.436	0.151063	
country_of_originHonduras	-9.530e-01	4.051e-01	-2.353	0.018641	*
country_of_originIndia	2.318e-01	6.836e-01	0.339	0.734498	
country_of_originIndonesia	3.926e-01	5.598e-01	0.701	0.483101	
country_of_originKenya	2.650e+00	7.824e-01	3.387	0.000708	***
country_of_originLaos	-1.559e+01	1.028e+03	-0.015	0.987904	
country_of_originMalawi	-1.928e+00	1.085e+00	-1.777	0.075511	.
country_of_originMauritius	-1.550e+01	1.455e+03	-0.011	0.991502	

country_of_originMexico	-6.438e-01	2.907e-01	-2.215	0.026779	*
country_of_originMyanmar	-1.559e+01	5.914e+02	-0.026	0.978966	
country_of_originNicaragua	-1.475e+00	6.737e-01	-2.189	0.028621	*
country_of_originPanama	1.445e+00	1.181e+00	1.223	0.221412	
country_of_originPeru	5.849e-01	7.559e-01	0.774	0.439090	
country_of_originPhilippines	-1.290e-01	9.479e-01	-0.136	0.891743	
country_of_originPuerto Rico	4.107e-02	1.260e+00	0.033	0.973987	
country_of_originTaiwan	-6.318e-02	3.494e-01	-0.181	0.856514	
country_of_originTanzania	2.934e-01	4.282e-01	0.685	0.493234	
country_of_originThailand	1.036e+00	5.472e-01	1.894	0.058276	.
country_of_originUganda	1.797e+00	4.983e-01	3.605	0.000312	***
country_of_originUnited States	1.068e+00	7.551e-01	1.414	0.157262	
country_of_originVietnam	3.899e-01	7.499e-01	0.520	0.603108	
country_of_originZambia	-1.519e+01	1.455e+03	-0.010	0.991672	
year2011	3.703e-01	6.341e-01	0.584	0.559245	
year2012	-3.664e-01	4.939e-01	-0.742	0.458170	
year2013	-7.276e-01	5.043e-01	-1.443	0.149064	
year2014	-3.683e-01	4.996e-01	-0.737	0.460985	
year2015	1.082e-01	5.187e-01	0.209	0.834784	
year2016	-5.850e-02	5.123e-01	-0.114	0.909095	
year2017	-2.518e-01	5.579e-01	-0.451	0.651666	
year2018	7.945e-01	7.396e-01	1.074	0.282718	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1496.7 on 1079 degrees of freedom
Residual deviance: 1262.1 on 1039 degrees of freedom
(60 observations deleted due to missingness)
AIC: 1344.1

Number of Fisher Scoring iterations: 14

Call:

```
glm(formula = Qualityclass ~ level + Colombia + Mexico + Honduras +
     Kenya, family = binomial(link = "logit"), data = coffee_final)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-2.2692	-0.9892	0.3981	1.0264	1.9184

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.4603	0.1935	-2.379	0.017356	*
level2	0.3747	0.2458	1.524	0.127456	
level3	0.8263	0.2132	3.876	0.000106	***
Colombia	1.3099	0.2543	5.152	2.58e-07	***
Mexico	-1.2070	0.1856	-6.502	7.91e-11	***
Honduras	-1.3988	0.3520	-3.974	7.06e-05	***
Kenya	2.1294	0.7506	2.837	0.004556	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1301.5 on 938 degrees of freedom
Residual deviance: 1146.5 on 932 degrees of freedom
(201 observations deleted due to missingness)
AIC: 1160.5

Number of Fisher Scoring iterations: 4

Call:

```
glm(formula = Qualityclass ~ level + country_of_origin + year -  
1, family = binomial(link = "logit"), data = coffee_final)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.24878	-0.87349	0.00031	0.98739	2.12047

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
level1	-6.057e-01	6.460e-01	-0.938	0.34842
level2	6.539e-02	6.237e-01	0.105	0.91650
level3	4.660e-01	6.327e-01	0.736	0.46144
country_of_originBurundi	-3.436e-01	1.454e+00	-0.236	0.81321
country_of_originChina	4.626e-01	6.421e-01	0.721	0.47119
country_of_originColombia	1.525e+00	3.612e-01	4.223	2.42e-05 ***
country_of_originCosta Rica	2.346e-01	4.355e-01	0.539	0.59011
country_of_originCote d'Ivoire	-1.606e+01	2.400e+03	-0.007	0.99466
country_of_originEcuador	9.024e-01	1.481e+00	0.609	0.54239
country_of_originEl Salvador	7.797e-01	5.933e-01	1.314	0.18874
country_of_originEthiopia	1.639e+01	4.970e+02	0.033	0.97369
country_of_originGuatemala	-2.782e-03	3.394e-01	-0.008	0.99346
country_of_originHaiti	-7.446e-01	1.185e+00	-0.628	0.52988
country_of_originHawaii	1.717e+01	2.400e+03	0.007	0.99429
country_of_originHonduras	-1.345e+00	4.471e-01	-3.008	0.00263 **
country_of_originIndia	3.083e-01	7.105e-01	0.434	0.66438
country_of_originIndonesia	4.157e-03	6.167e-01	0.007	0.99462
country_of_originKenya	2.400e+00	8.037e-01	2.986	0.00283 **
country_of_originLaos	-1.707e+01	1.696e+03	-0.010	0.99197
country_of_originMalawi	-2.146e+00	1.097e+00	-1.957	0.05035 .
country_of_originMauritius	-1.606e+01	2.400e+03	-0.007	0.99466
country_of_originMexico	-9.546e-01	3.255e-01	-2.933	0.00336 **
country_of_originMyanmar	-1.682e+01	9.735e+02	-0.017	0.98622
country_of_originNicaragua	-1.245e+00	7.120e-01	-1.748	0.08044 .
country_of_originPanama	1.155e+00	1.192e+00	0.969	0.33234
country_of_originPeru	-1.669e+01	2.400e+03	-0.007	0.99445
country_of_originPhilippines	-2.897e-01	9.557e-01	-0.303	0.76183
country_of_originPuerto Rico	6.030e-01	1.289e+00	0.468	0.63987
country_of_originTaiwan	2.949e-01	3.980e-01	0.741	0.45863
country_of_originTanzania	-1.655e-01	4.788e-01	-0.346	0.72962
country_of_originThailand	7.166e-01	6.329e-01	1.132	0.25752
country_of_originUganda	1.292e+00	5.338e-01	2.420	0.01554 *
country_of_originUnited States	8.490e-01	7.818e-01	1.086	0.27750
country_of_originVietnam	4.664e-01	8.165e-01	0.571	0.56787
country_of_originZambia	-1.658e+01	2.400e+03	-0.007	0.99449

year2011	2.581e-01	7.296e-01	0.354	0.72350
year2012	-3.418e-01	6.032e-01	-0.567	0.57092
year2013	-6.905e-01	6.106e-01	-1.131	0.25814
year2014	-4.561e-01	6.076e-01	-0.751	0.45289
year2015	-1.422e-02	6.148e-01	-0.023	0.98154
year2016	9.710e-02	6.157e-01	0.158	0.87469
year2017	-4.201e-01	6.555e-01	-0.641	0.52156
year2018	1.284e+00	8.242e-01	1.557	0.11939

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1289.3 on 930 degrees of freedom
Residual deviance: 1047.4 on 887 degrees of freedom
(210 observations deleted due to missingness)
AIC: 1133.4

Number of Fisher Scoring iterations: 15

Call:

```
glm(formula = Qualityclass ~ aroma + flavor + acidity + country_of_origin +
    defects_log + level + year, family = binomial(link = "logit"),
    data = coffee_final)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.5914	-0.2397	0.0000	0.2843	3.5781

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.52175	1.09510	-1.390	0.16465
aroma	1.62814	0.26590	6.123	9.18e-10 ***
flavor	2.89136	0.36116	8.006	1.19e-15 ***
acidity	1.67688	0.26327	6.369	1.90e-10 ***
country_of_originBurundi	1.88240	5.12830	0.367	0.71357
country_of_originChina	0.49916	1.08844	0.459	0.64652
country_of_originColombia	1.84638	0.57358	3.219	0.00129 **
country_of_originCosta Rica	0.26961	0.76612	0.352	0.72491
country_of_originCote d'Ivoire	-12.11826	6522.63865	-0.002	0.99852
country_of_originEcuador	-1.02265	1.52999	-0.668	0.50388
country_of_originEl Salvador	0.32640	0.96977	0.337	0.73644
country_of_originEthiopia	13.49329	894.76317	0.015	0.98797
country_of_originGuatemala	-0.75268	0.57572	-1.307	0.19108
country_of_originHaiti	2.27451	2.16150	1.052	0.29267
country_of_originHawaii	4.41740	6522.63880	0.001	0.99946
country_of_originHonduras	-0.72501	0.71286	-1.017	0.30913
country_of_originIndia	-2.55120	1.07559	-2.372	0.01770 *
country_of_originIndonesia	-0.38258	1.01141	-0.378	0.70524
country_of_originKenya	0.52684	1.54516	0.341	0.73313
country_of_originLaos	-15.24675	4515.00054	-0.003	0.99731
country_of_originMalawi	-0.65398	1.30094	-0.503	0.61518
country_of_originMauritius	-11.76872	6522.63865	-0.002	0.99856
country_of_originMexico	-0.80196	0.52029	-1.541	0.12323

country_of_originMyanmar	-15.49786	2401.00369	-0.006	0.99485
country_of_originNicaragua	0.53829	1.98308	0.271	0.78605
country_of_originPanama	3.27141	1.79738	1.820	0.06874 .
country_of_originPeru	-14.50164	6522.63864	-0.002	0.99823
country_of_originPhilippines	2.89981	2.57307	1.127	0.25975
country_of_originPuerto Rico	-2.65794	1.78541	-1.489	0.13657
country_of_originTaiwan	1.18951	0.70762	1.681	0.09276 .
country_of_originTanzania	0.91717	0.75964	1.207	0.22729
country_of_originThailand	2.87480	0.99592	2.887	0.00389 **
country_of_originUganda	-1.53625	0.79415	-1.934	0.05306 .
country_of_originUnited States	0.19578	1.52935	0.128	0.89814
country_of_originVietnam	2.24627	1.15874	1.939	0.05256 .
country_of_originZambia	-13.96552	6522.63865	-0.002	0.99829
defects_log	0.33145	0.17162	1.931	0.05345 .
level2	0.52403	0.48450	1.082	0.27943
level3	1.03968	0.48225	2.156	0.03109 *
year2011	-0.22625	1.12956	-0.200	0.84125
year2012	0.03098	0.98109	0.032	0.97481
year2013	0.48471	0.98717	0.491	0.62342
year2014	-0.07904	0.99385	-0.080	0.93661
year2015	-0.14258	0.98571	-0.145	0.88499
year2016	0.78470	1.03677	0.757	0.44913
year2017	0.46753	1.03839	0.450	0.65254
year2018	2.35570	1.32235	1.781	0.07484 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1289.15 on 929 degrees of freedom
Residual deviance: 448.42 on 883 degrees of freedom
(210 observations deleted due to missingness)
AIC: 542.42

Number of Fisher Scoring iterations: 17

Rows: 930

Columns: 12

```
$ country_of_origin <chr> "Myanmar", "Uganda", "Ethiopia", "Mexico", "Burundi"~
$ aroma <dbl[,1]> <matrix[26 x 1]>
$ flavor <dbl[,1]> <matrix[26 x 1]>
$ acidity <dbl[,1]> <matrix[26 x 1]>
$ defects_log <dbl> 1.6094379, 0.6931472, 2.0794415, 1.3862944, 1.79~
$ year <fct> 2015, 2013, 2014, 2012, 2012, 2014, 2015, 2013, ~
$ level <chr> "3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3"~
$ Qualityclass <dbl> 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1~
$ Colombia <dbl> 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0~
$ Mexico <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ Honduras <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0~
$ Kenya <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

Call:

```
glm(formula = Qualityclass ~ aroma + flavor + acidity + Colombia +
     Mexico + Honduras + Kenya + defects_log + level + year, family = binomial(link = "logit"),
```

```
data = coffee_final_nomiss)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.2576	-0.2933	0.0010	0.3296	3.6482

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.2894	0.9358	-1.378	0.1683
aroma	1.3883	0.2328	5.964	2.47e-09 ***
flavor	2.4519	0.3004	8.162	3.30e-16 ***
acidity	1.4874	0.2352	6.324	2.55e-10 ***
Colombia	1.9282	0.4095	4.708	2.50e-06 ***
Mexico	-0.7003	0.3512	-1.994	0.0461 *
Honduras	-0.5767	0.5473	-1.054	0.2920
Kenya	0.8497	1.3961	0.609	0.5427
defects_log	0.3119	0.1509	2.067	0.0387 *
level2	0.4545	0.4188	1.085	0.2779
level3	0.6754	0.3769	1.792	0.0731 .
year2011	-0.1599	1.0447	-0.153	0.8783
year2012	0.1019	0.8983	0.113	0.9097
year2013	0.1358	0.8937	0.152	0.8793
year2014	0.4155	0.8995	0.462	0.6441
year2015	-0.1081	0.9090	-0.119	0.9054
year2016	0.8173	0.9402	0.869	0.3847
year2017	0.2811	0.9682	0.290	0.7716
year2018	2.0529	1.1977	1.714	0.0865 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1289.15 on 929 degrees of freedom
Residual deviance: 493.82 on 911 degrees of freedom
AIC: 531.82

Number of Fisher Scoring iterations: 7

Start: AIC=531.82

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + Kenya + defects_log + level + year

	Df	Deviance	AIC
- year	8	503.11	525.11
- Kenya	1	494.24	530.24
- Honduras	1	494.97	530.97
- level	2	497.14	531.14
<none>		493.82	531.82
- Mexico	1	497.85	533.85
- defects_log	1	498.16	534.16
- Colombia	1	520.03	556.03
- acidity	1	543.95	579.95
- aroma	1	544.44	580.44
- flavor	1	584.68	620.68

Step: AIC=525.11

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + Kenya + defects_log + level

	Df	Deviance	AIC
- level	2	503.99	521.99
- Kenya	1	503.79	523.79
- Honduras	1	504.15	524.15
<none>		503.11	525.11
- defects_log	1	508.43	528.43
+ year	8	493.82	531.82
- Mexico	1	511.95	531.95
- Colombia	1	530.49	550.49
- aroma	1	551.81	571.81
- acidity	1	557.40	577.40
- flavor	1	593.89	613.89

Step: AIC=521.99

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + Kenya + defects_log

	Df	Deviance	AIC
- Honduras	1	504.71	520.71
- Kenya	1	504.78	520.78
<none>		503.99	521.99
+ level	2	503.11	525.11
- defects_log	1	509.68	525.68
- Mexico	1	512.91	528.91
+ year	8	497.14	531.14
- Colombia	1	535.47	551.47
- aroma	1	554.81	570.81
- acidity	1	560.36	576.36
- flavor	1	593.96	609.96

Step: AIC=520.71

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Kenya + defects_log

	Df	Deviance	AIC
- Kenya	1	505.57	519.57
<none>		504.71	520.71
+ Honduras	1	503.99	521.99
- defects_log	1	509.97	523.97
+ level	2	504.15	524.15
- Mexico	1	513.00	527.00
+ year	8	497.78	529.78
- Colombia	1	538.07	552.07
- aroma	1	556.11	570.11
- acidity	1	561.95	575.95
- flavor	1	594.88	608.88

Step: AIC=519.57

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +

defects_log

	Df	Deviance	AIC
<none>		505.57	519.57
+ Kenya	1	504.71	520.71
+ Honduras	1	504.78	520.78
+ level	2	504.92	522.92
- defects_log	1	510.96	522.96
- Mexico	1	514.26	526.26
+ year	8	498.34	528.34
- Colombia	1	538.58	550.58
- aroma	1	556.46	568.46
- acidity	1	563.79	575.79
- flavor	1	597.26	609.26

```
Call: glm(formula = Qualityclass ~ aroma + flavor + acidity + Colombia +  
Mexico + defects_log, family = binomial(link = "logit"),  
data = coffee_final_nomiss)
```

Coefficients:

(Intercept)	aroma	flavor	acidity	Colombia	Mexico
-0.4878	1.3231	2.3850	1.5695	1.8858	-0.8483

defects_log
0.3276

Degrees of Freedom: 929 Total (i.e. Null); 923 Residual

Null Deviance: 1289

Residual Deviance: 505.6 AIC: 519.6

1 2 3
126 201 612

2010 2011 2012 2013 2014 2015 2016 2017 2018
32 33 307 161 216 125 131 57 18

Rows: 1,140

Columns: 12

```
$ country_of_origin <chr> "Myanmar", "Uganda", "Ethiopia", "Mexico", "Burundi"~  
$ aroma <dbl[,1]> <matrix[26 x 1]>  
$ flavor <dbl[,1]> <matrix[26 x 1]>  
$ acidity <dbl[,1]> <matrix[26 x 1]>  
$ defects_log <dbl> 1.6094379, 0.6931472, 2.0794415, 1.3862944, 1.79~  
$ year <dbl> 6, 4, 5, 3, 3, 5, 5, 6, 4, 4, 3, 7, 3, 6, 6, 5, ~  
$ level <chr> "3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3"~  
$ Qualityclass <dbl> 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1~  
$ Colombia <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0~  
$ Mexico <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~  
$ Honduras <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0~  
$ Kenya <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
```

1 2 3 4 5 6 7 8 9
32 33 307 161 276 125 131 57 18

```
Call:
glm(formula = Qualityclass ~ aroma + flavor + acidity + Colombia +
     Mexico + Honduras + Kenya + defects_log + level + year, family = binomial(link = "logit"),
     data = coffee_final_nomiss)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.2576	-0.2933	0.0010	0.3296	3.6482

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.2894	0.9358	-1.378	0.1683
aroma	1.3883	0.2328	5.964	2.47e-09 ***
flavor	2.4519	0.3004	8.162	3.30e-16 ***
acidity	1.4874	0.2352	6.324	2.55e-10 ***
Colombia	1.9282	0.4095	4.708	2.50e-06 ***
Mexico	-0.7003	0.3512	-1.994	0.0461 *
Honduras	-0.5767	0.5473	-1.054	0.2920
Kenya	0.8497	1.3961	0.609	0.5427
defects_log	0.3119	0.1509	2.067	0.0387 *
level2	0.4545	0.4188	1.085	0.2779
level3	0.6754	0.3769	1.792	0.0731 .
year2011	-0.1599	1.0447	-0.153	0.8783
year2012	0.1019	0.8983	0.113	0.9097
year2013	0.1358	0.8937	0.152	0.8793
year2014	0.4155	0.8995	0.462	0.6441
year2015	-0.1081	0.9090	-0.119	0.9054
year2016	0.8173	0.9402	0.869	0.3847
year2017	0.2811	0.9682	0.290	0.7716
year2018	2.0529	1.1977	1.714	0.0865 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1289.15 on 929 degrees of freedom
 Residual deviance: 493.82 on 911 degrees of freedom
 AIC: 531.82

Number of Fisher Scoring iterations: 7

Start: AIC=682.29

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
 Honduras + Kenya + defects_log + level + year

	Df	Deviance	AIC
- level	2	660.94	680.94
- Kenya	1	659.20	681.20
- year	1	659.91	681.91
<none>		658.29	682.29
- defects_log	1	660.50	682.50
- Honduras	1	660.75	682.75
- Mexico	1	664.26	686.26
- Colombia	1	679.44	701.44

- acidity	1	707.71	729.71
- aroma	1	720.74	742.74
- flavor	1	780.07	802.07

Step: AIC=680.94

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + Kenya + defects_log + year

	Df	Deviance	AIC
- Kenya	1	662.03	680.03
- year	1	662.09	680.09
- Honduras	1	662.65	680.65
<none>		660.94	680.94
- defects_log	1	663.63	681.63
+ level	2	658.29	682.29
- Mexico	1	667.72	685.72
- Colombia	1	685.20	703.20
- acidity	1	713.05	731.05
- aroma	1	724.84	742.84
- flavor	1	780.58	798.58

Step: AIC=680.03

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + defects_log + year

	Df	Deviance	AIC
- year	1	663.40	679.40
- Honduras	1	663.88	679.88
<none>		662.03	680.03
- defects_log	1	664.80	680.80
+ Kenya	1	660.94	680.94
+ level	2	659.20	681.20
- Mexico	1	668.97	684.97
- Colombia	1	686.08	702.08
- acidity	1	715.04	731.04
- aroma	1	725.78	741.78
- flavor	1	783.23	799.23

Step: AIC=679.4

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
Honduras + defects_log

	Df	Deviance	AIC
- Honduras	1	664.76	678.76
<none>		663.40	679.40
+ year	1	662.03	680.03
+ Kenya	1	662.09	680.09
- defects_log	1	666.23	680.23
+ level	2	661.06	681.06
- Mexico	1	673.96	687.96
- Colombia	1	686.09	700.09
- acidity	1	716.17	730.17
- aroma	1	725.99	739.99
- flavor	1	785.34	799.34

Step: AIC=678.76

Qualityclass ~ aroma + flavor + acidity + Colombia + Mexico +
defects_log

	Df	Deviance	AIC
<none>		664.76	678.76
- defects_log	1	667.20	679.20
+ Kenya	1	663.36	679.36
+ Honduras	1	663.40	679.40
+ year	1	663.88	679.88
+ level	2	662.94	680.94
- Mexico	1	674.44	686.44
- Colombia	1	689.05	701.05
- acidity	1	718.88	730.88
- aroma	1	727.99	739.99
- flavor	1	786.93	798.93

Call: glm(formula = Qualityclass ~ aroma + flavor + acidity + Colombia +
Mexico + defects_log, family = binomial(link = "logit"),
data = coffee_final_imputed)

Coefficients:

(Intercept)	aroma	flavor	acidity	Colombia	Mexico
-0.2442	1.2827	2.2675	1.2183	1.4034	-0.8231

defects_log
0.1857

Degrees of Freedom: 1139 Total (i.e. Null); 1133 Residual

Null Deviance: 1579

Residual Deviance: 664.8 AIC: 678.8

##Final Model

Call:

glm(formula = Qualityclass ~ aroma + flavor + acidity + Colombia +
Mexico + defects_log, family = binomial(link = "logit"),
data = coffee_final)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8880	-0.3570	0.0092	0.4077	3.2372

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.2442	0.1672	-1.461	0.14404
aroma	1.2827	0.1842	6.964	3.31e-12 ***
flavor	2.2675	0.2399	9.451	< 2e-16 ***
acidity	1.2183	0.1768	6.891	5.56e-12 ***
Colombia	1.4034	0.3017	4.651	3.30e-06 ***
Mexico	-0.8231	0.2683	-3.068	0.00215 **
defects_log	0.1857	0.1192	1.559	0.11903

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1579.36 on 1139 degrees of freedom
Residual deviance: 664.76 on 1133 degrees of freedom
AIC: 678.76

Number of Fisher Scoring iterations: 7

[1] 0.8815789

[1] 0.8904022

[1] 0.8731034

##Sensitivity Analysis

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]

Family: binomial (logit)

Formula: Qualityclass ~ 1 + aroma + flavor + acidity + defects_log + (1 |
country_of_origin)

Data: coffee_final

AIC	BIC	logLik	deviance	df.resid
681.5	711.7	-334.8	669.5	1134

Scaled residuals:

Min	1Q	Median	3Q	Max
-51.267	-0.239	0.004	0.268	13.385

Random effects:

Groups	Name	Variance	Std.Dev.
country_of_origin	(Intercept)	0.6285	0.7928

Number of obs: 1140, groups: country_of_origin, 34

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.2502	0.2443	-1.024	0.3059
aroma	1.3733	0.1950	7.043	1.88e-12 ***
flavor	2.4130	0.2562	9.418	< 2e-16 ***
acidity	1.2662	0.1872	6.765	1.33e-11 ***
defects_log	0.2338	0.1279	1.828	0.0675 .

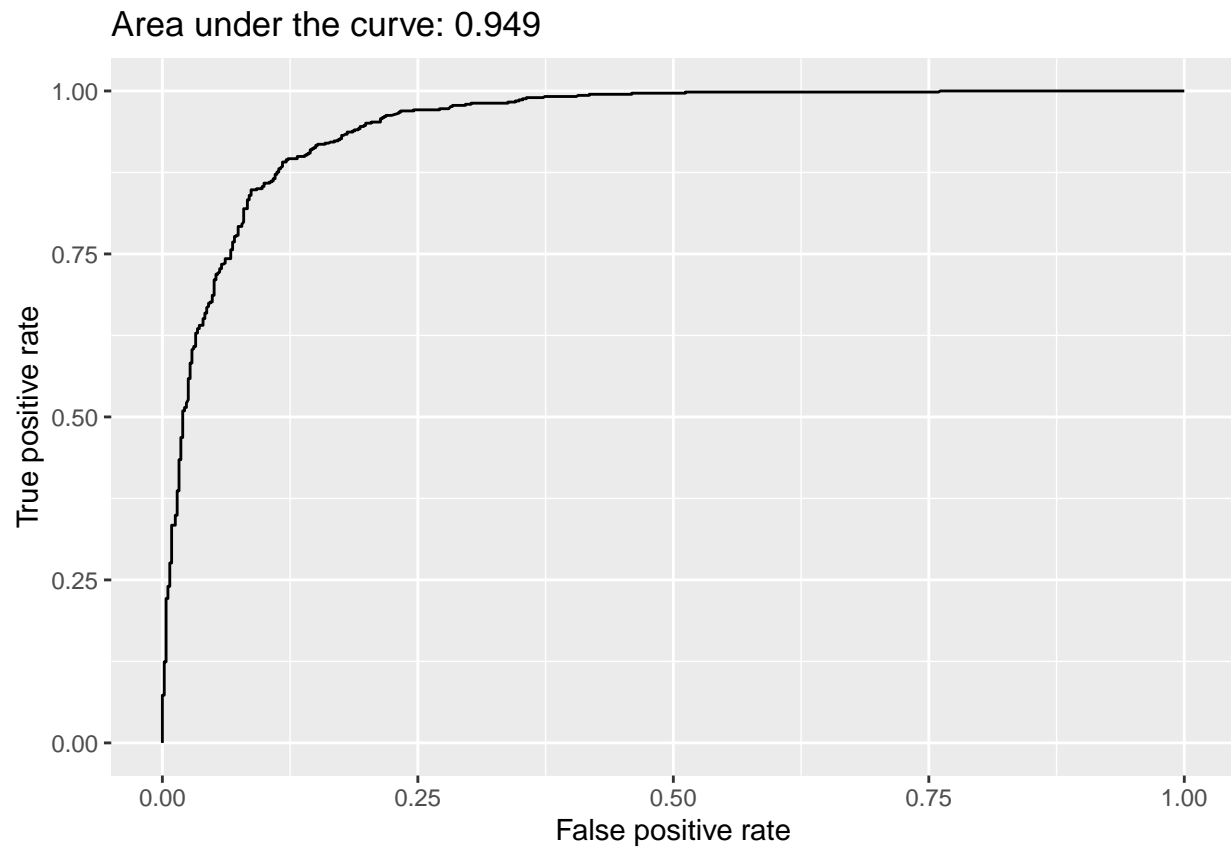
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

[1] 0.8701754

[1] 0.8823189

[1] 0.8592056

Classification boundry



	cut	fpr	tpr	dif
555	0.5091171	0.1175407	0.8909710	0.7734304
559	0.5041421	0.1211573	0.8943782	0.7732209
561	0.5024577	0.1229656	0.8960818	0.7731161
554	0.5132069	0.1175407	0.8892675	0.7717268
556	0.5079264	0.1193490	0.8909710	0.7716220
558	0.5063213	0.1211573	0.8926746	0.7715173

[1] 0.5091171

	cut	fpr	tpr	dif
555	0.5091171	0.1175407	0.8909710	0.7734304
559	0.5041421	0.1211573	0.8943782	0.7732209
561	0.5024577	0.1229656	0.8960818	0.7731161
554	0.5132069	0.1175407	0.8892675	0.7717268
556	0.5079264	0.1193490	0.8909710	0.7716220
558	0.5063213	0.1211573	0.8926746	0.7715173

[1] 0.8798246

[1] 0.8855754

[1] 0.8750265

[1] 0.8824561

[1] 0.8917863

[1] 0.8731034