
OpenDataBC
BC Liquor Store Data Visualization
Rock Chi

SETTING

OpenDataBC had provided a public data on all products currently sold by BC Liquor Stores

This provides a great opportunity to understanding liquor by utilizing data

THE DATA

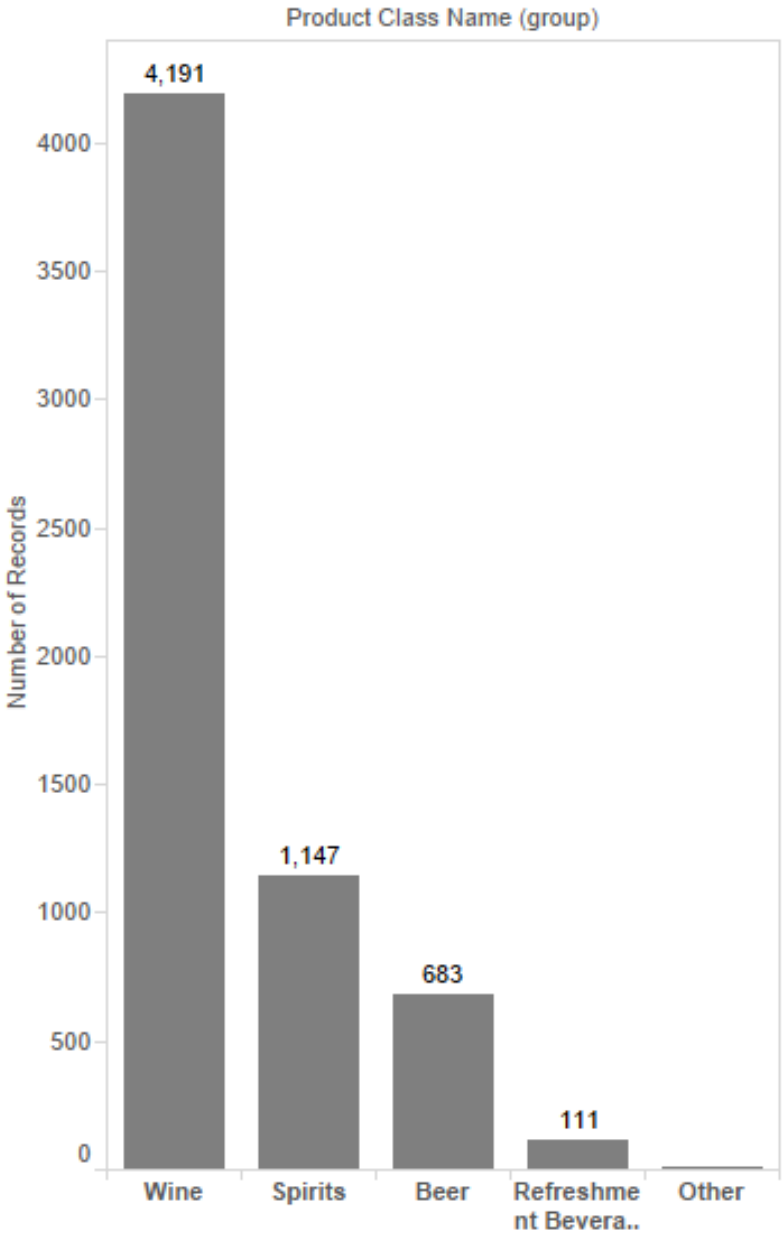
Data provided extracted on January 21, 2012 containing product listing of 6144 items

	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT_LONG_NAME	PRODUCT_	PRODUCT_	PRD_CON	PRODUCT_	CURRENT_	SWEETNESS_CODE
2	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	198267	COPPER MOON - MALBEC	4.82E+10	3	1	14	30.99	0
3	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	305375	DOMAINE D'OR - DRY	4.82E+10	4	1	11.5	32.99	0
4	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	53017	SOMMET ROUGE	5.9E+10	4	1	12	29.99	0
5	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	215525	MISSION RIDGE - PREMIUM DRY WHITE	7.8E+11	4	1	11	33.99	1

PRICE IS RIGHT CHALLENGE

Based on the data given, what are the heuristics used for best-guessing prices currently set by BC Liquor?

WHAT TYPE OF LIQUOR ARE SOLD?

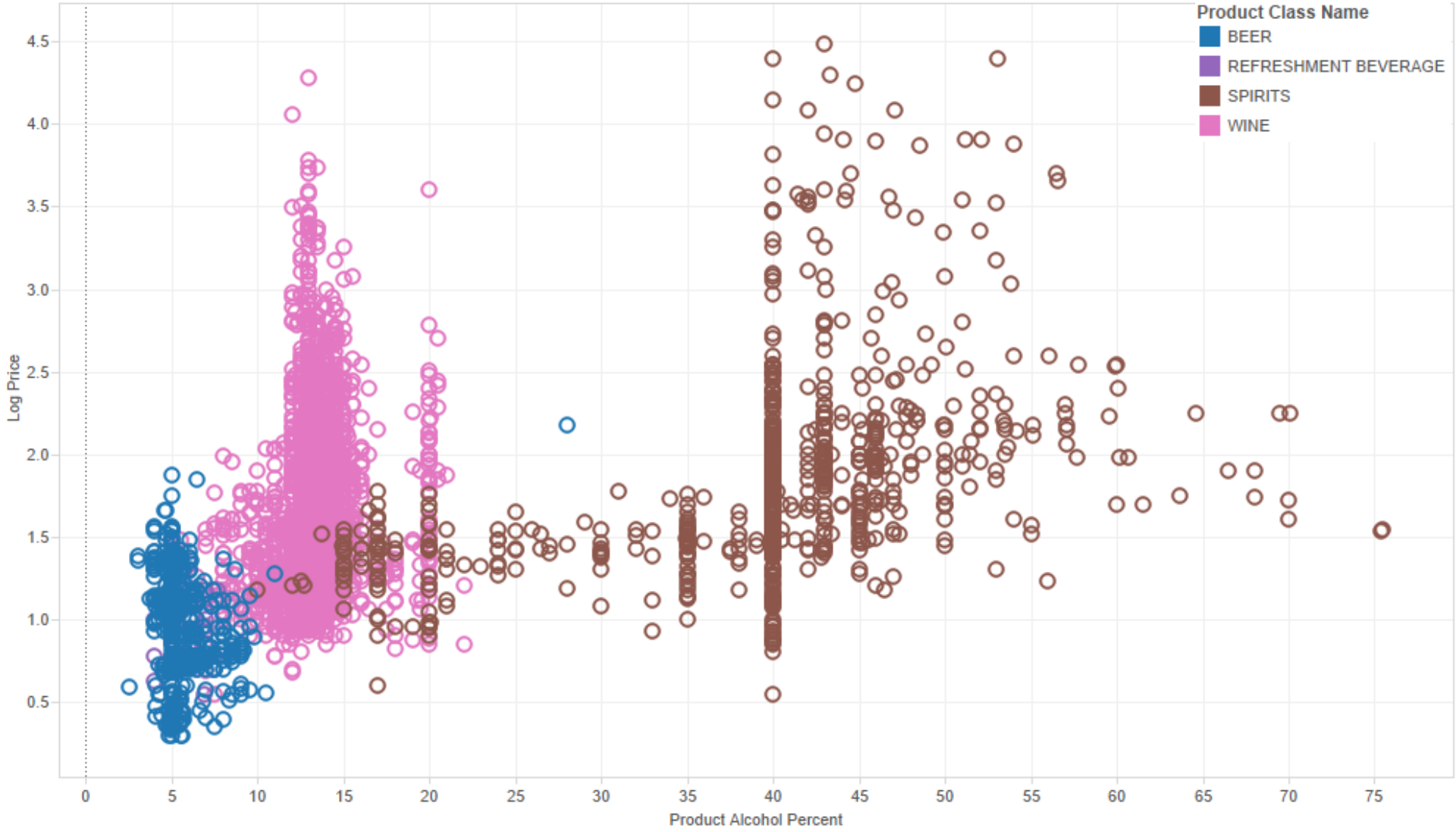


Four major category amongst product in stock

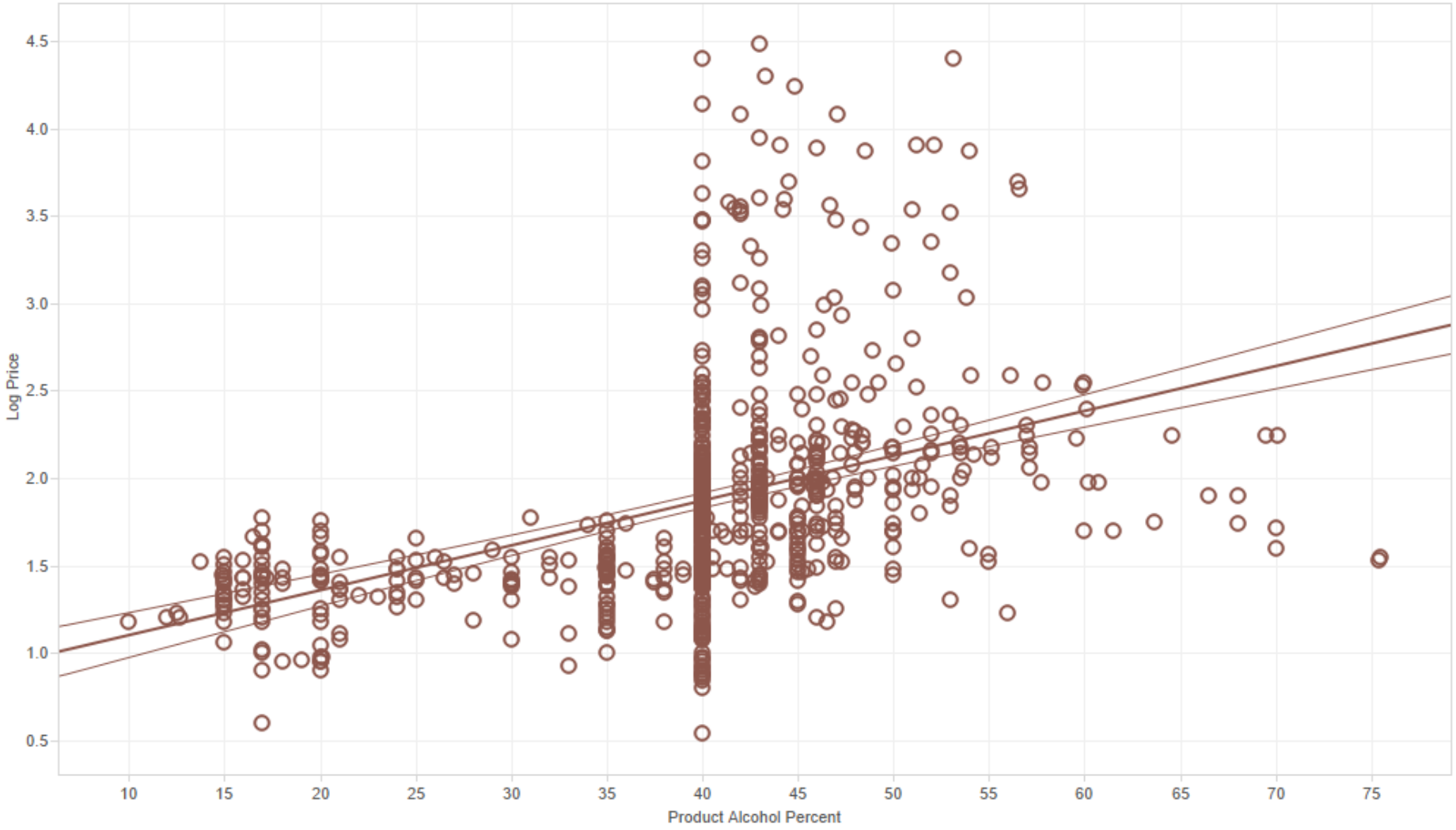
Wine has the greatest product variety with over 4,000 unique products sold.

Other Category contains primarily of de-alcoholized wines and beers

LIQUOR TYPE, ALCOHOL PERCENT AND PRICE



VARIABLE EFFECT ON SPIRITS



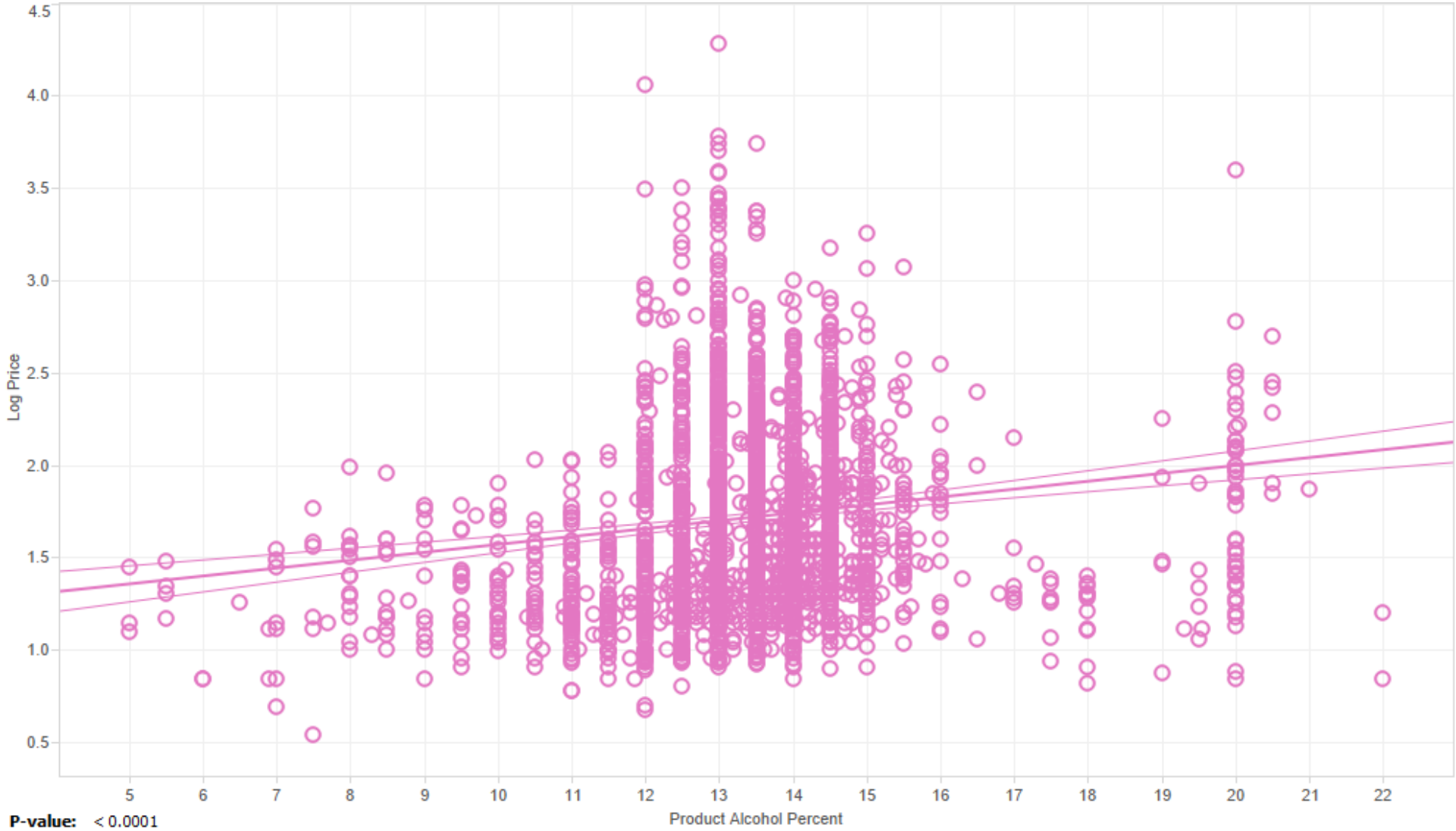
P-value: < 0.0001
Equation: Log Price = 0.0256506*Product Alcohol Percent + 0.844677

Coefficients

Term	Value	StdErr	t-value	p-value
Product Alcohol Percent	0.0256506	0.0020697	12.3932	< 0.0001
intercept	0.844677	0.0850704	9.92916	< 0.0001

Product Class Name
- BEER
- REFRESHMENT BEVERAGE
- SPIRITS
- WINE

VARIABLE EFFECT ON WINES



P-value: < 0.0001

Equation: Log Price = 0.0427722*Product Alcohol Percent + 1.1414

Coefficients

Term	Value	StdErr	t-value	p-value
Product Alcohol Percent	0.0427722	0.0057639	7.42068	< 0.0001
intercept	1.1414	0.0780753	14.6192	< 0.0001

Product Class Name

- BEER
- REFRESHMENT BEVERAGE
- SPIRITS
- WINE

LIQUOR TYPE, ALCOHOL PERCENT AND PRICE

ALCOHOL PERCENT AND LIQUOR TYPE

Liquor with Alcohol Percent greater than 25% can immediately be classified as “Spirit”

Wines generally range from 5% to 22% Alcohol Percentage

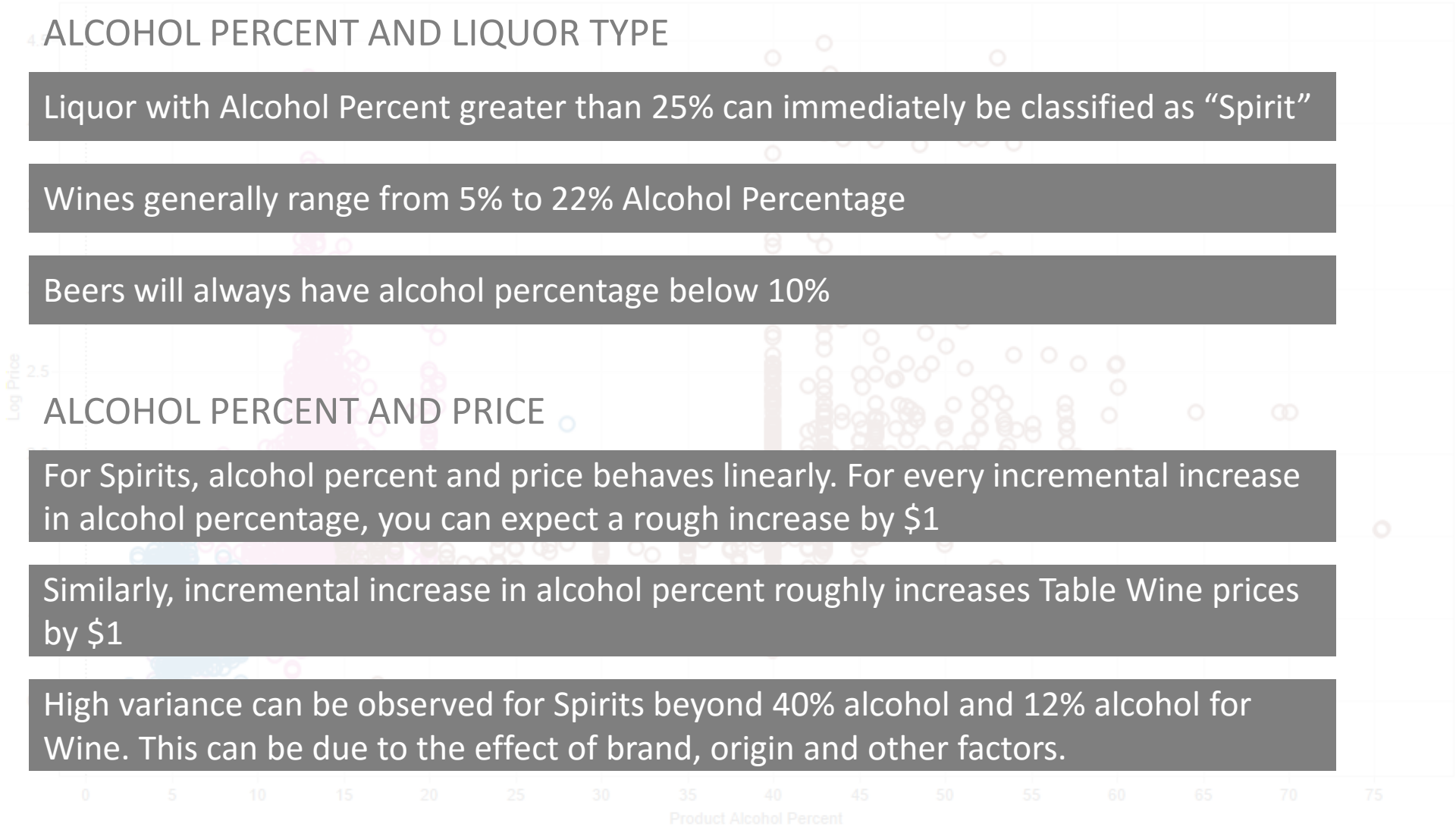
Beers will always have alcohol percentage below 10%

ALCOHOL PERCENT AND PRICE

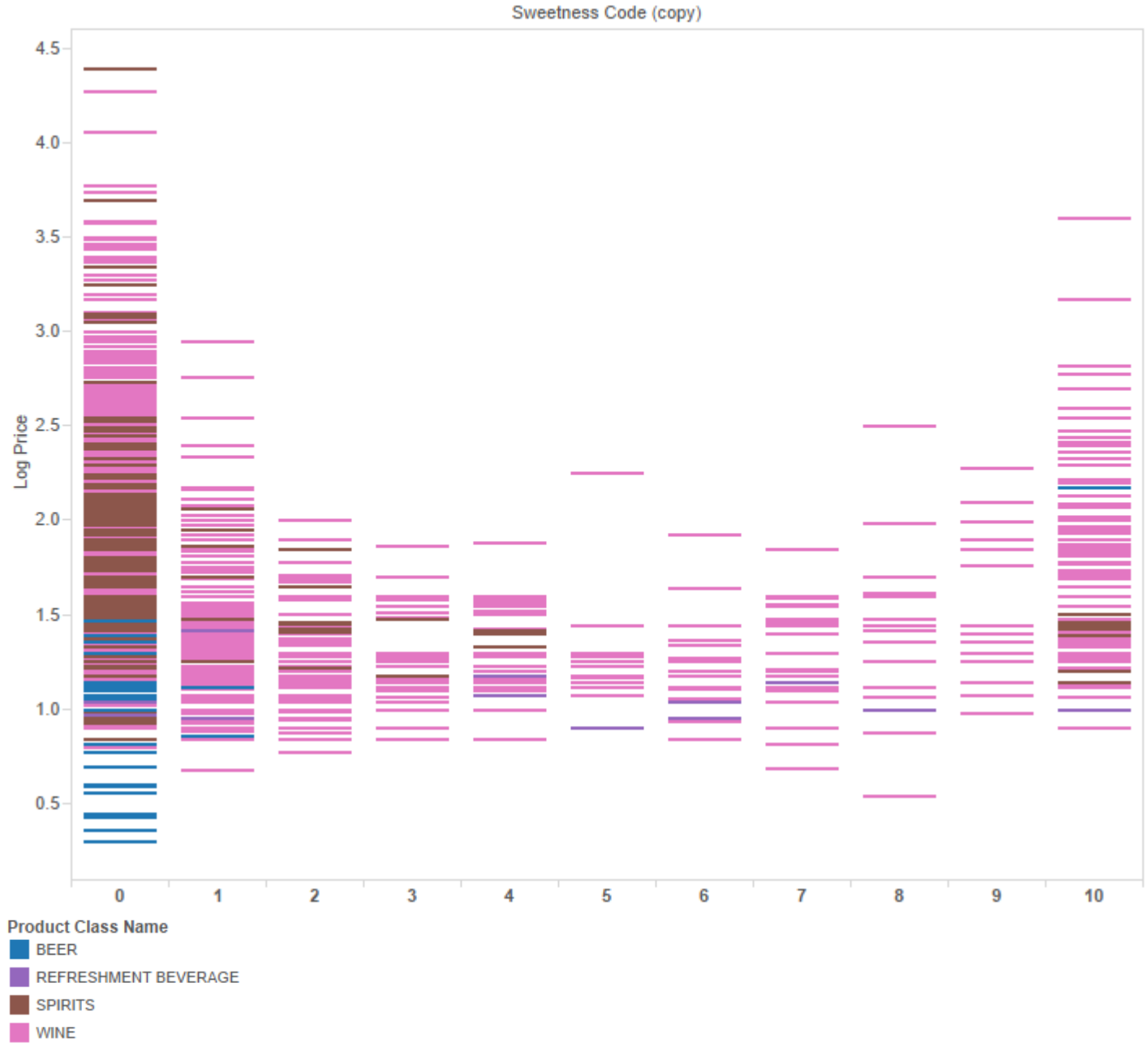
For Spirits, alcohol percent and price behaves linearly. For every incremental increase in alcohol percentage, you can expect a rough increase by $1

Similarly, incremental increase in alcohol percent roughly increases Table Wine prices by $1

High variance can be observed for Spirits beyond 40% alcohol and 12% alcohol for Wine. This can be due to the effect of brand, origin and other factors.



LIQUOR, SWEETNESS AND PRICE



Largest variance in sweetness is seen in Wine Category

Code	Description	Grams Sugar/L
0	Very Dry	0-5
1-2	Off-Dry	5-25
3-4	Medium	25-45
5-6	Sweet	45-65
7-10	Very Sweet	65-105+

Beer, Refreshment and Spirits tend to have low sugar content

TABLE WINE SWEETNESS AND PRICE

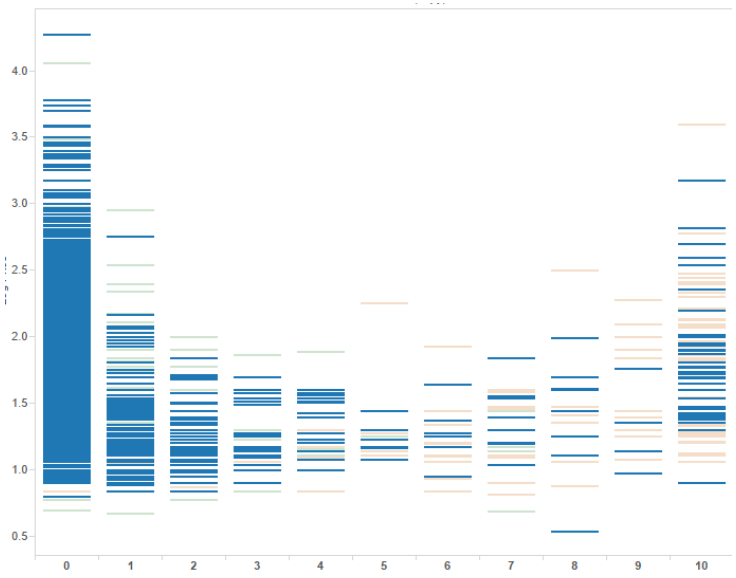
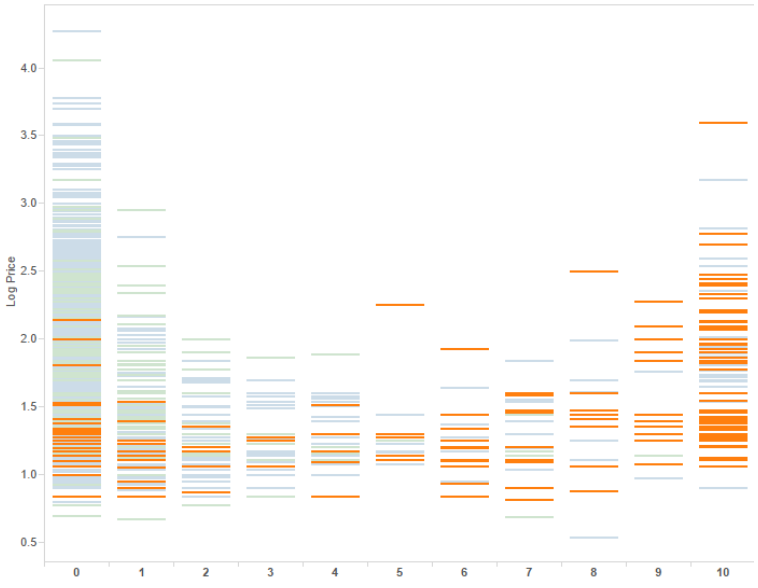
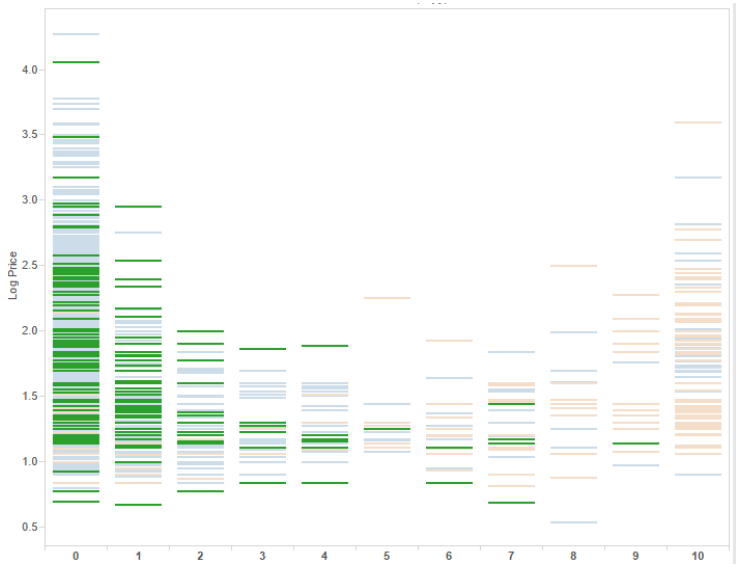


Table Wine



Dessert Wine

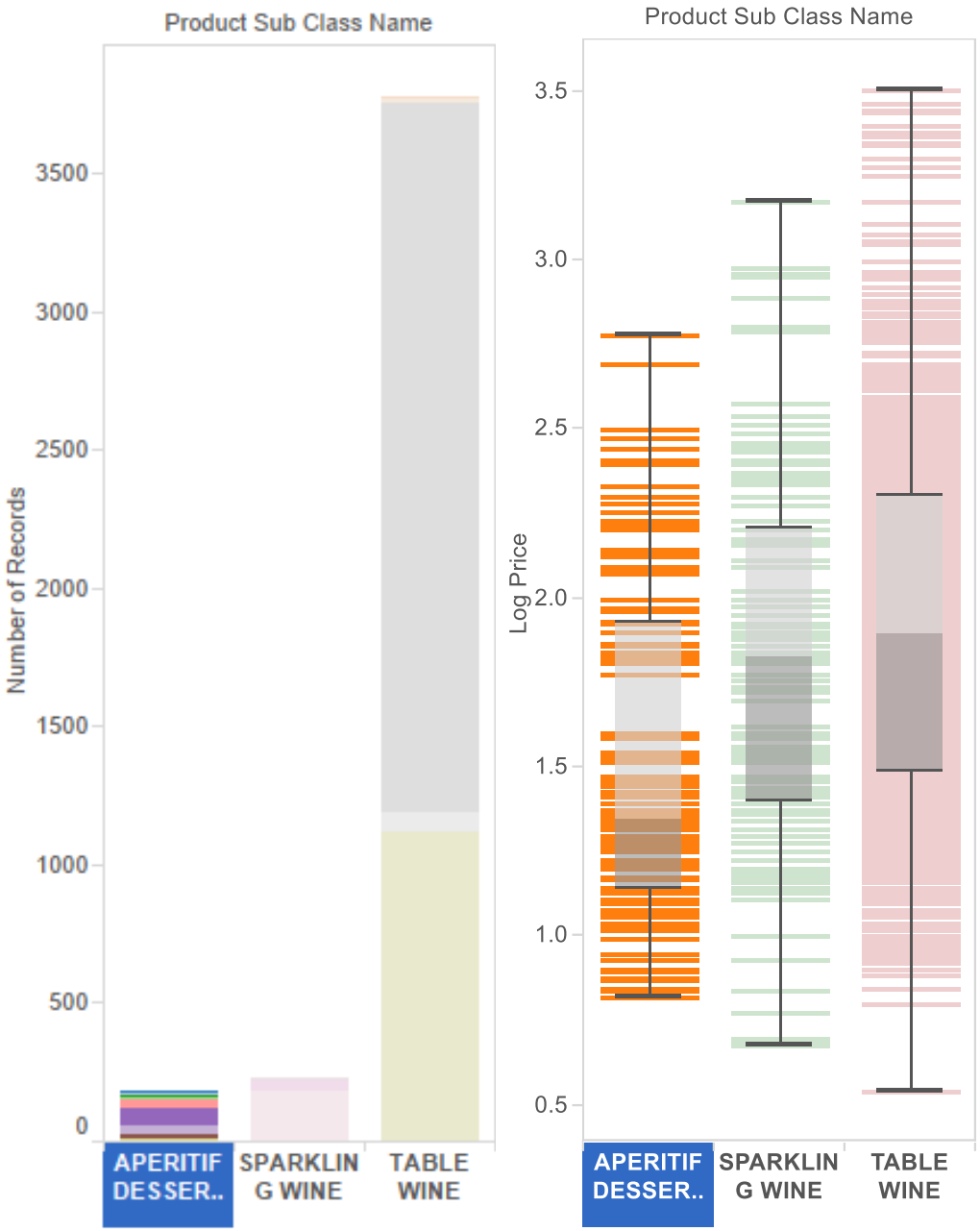


Sparkling Wine

Sweetness has a positive relationship with Price for Dessert Wines

Sweetness has a negative relationship with Price for Sparkling Wines

FURTHER UNDERSTANDING WINE



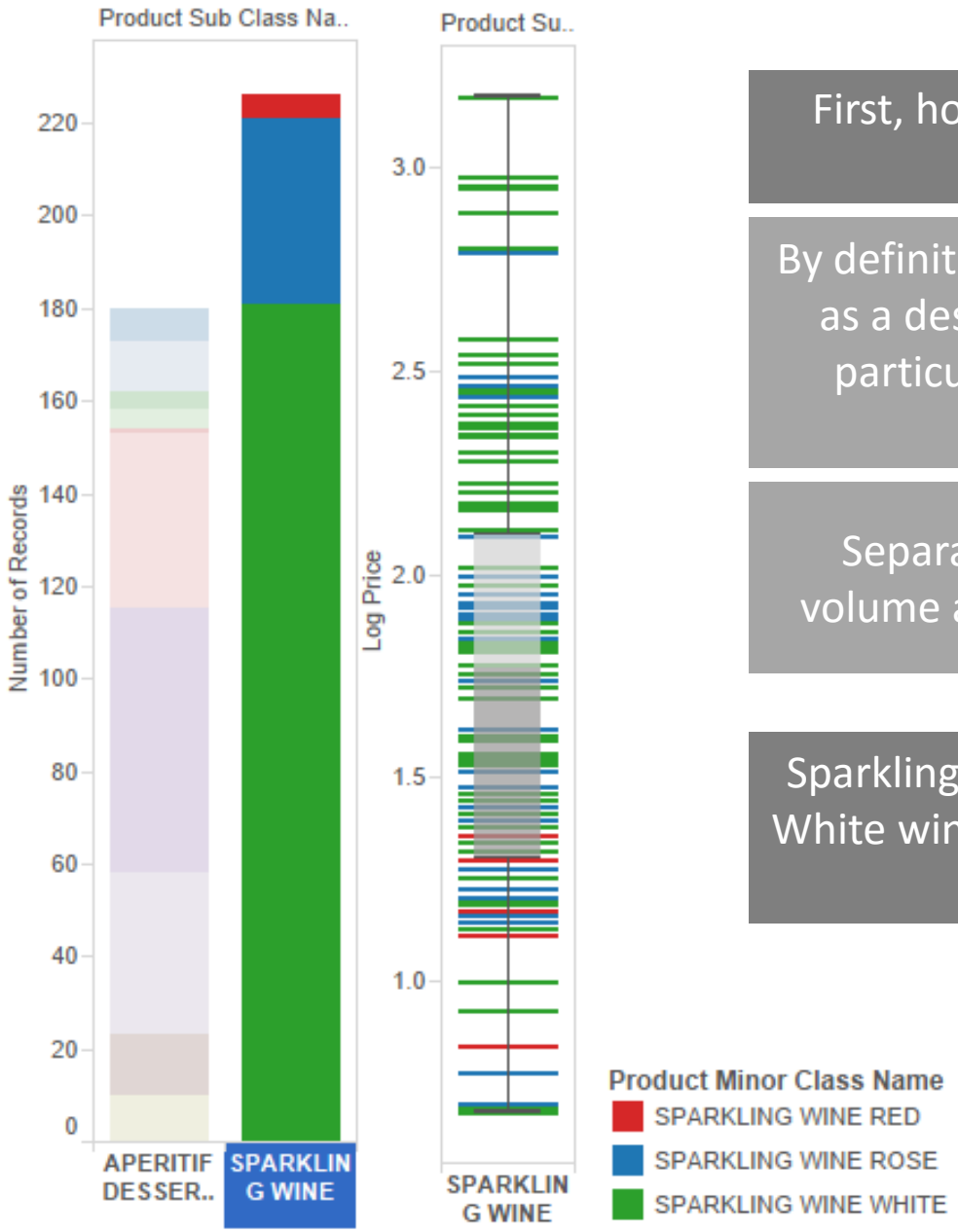
Wine is broken into three classes: Dessert, Sparkling and Table Wine

For three categories of wine exhibit right-skew in price distribution, so log price is used for comparison

Dessert Wine has the largest variety as exhibited by the array of different color-coded minor product type on the left

In examining distribution of products over price points, Dessert Wine tend to be more centralized at ~$18. (log 1.25)

SPARKLING WINE VS. DESSERT WINE



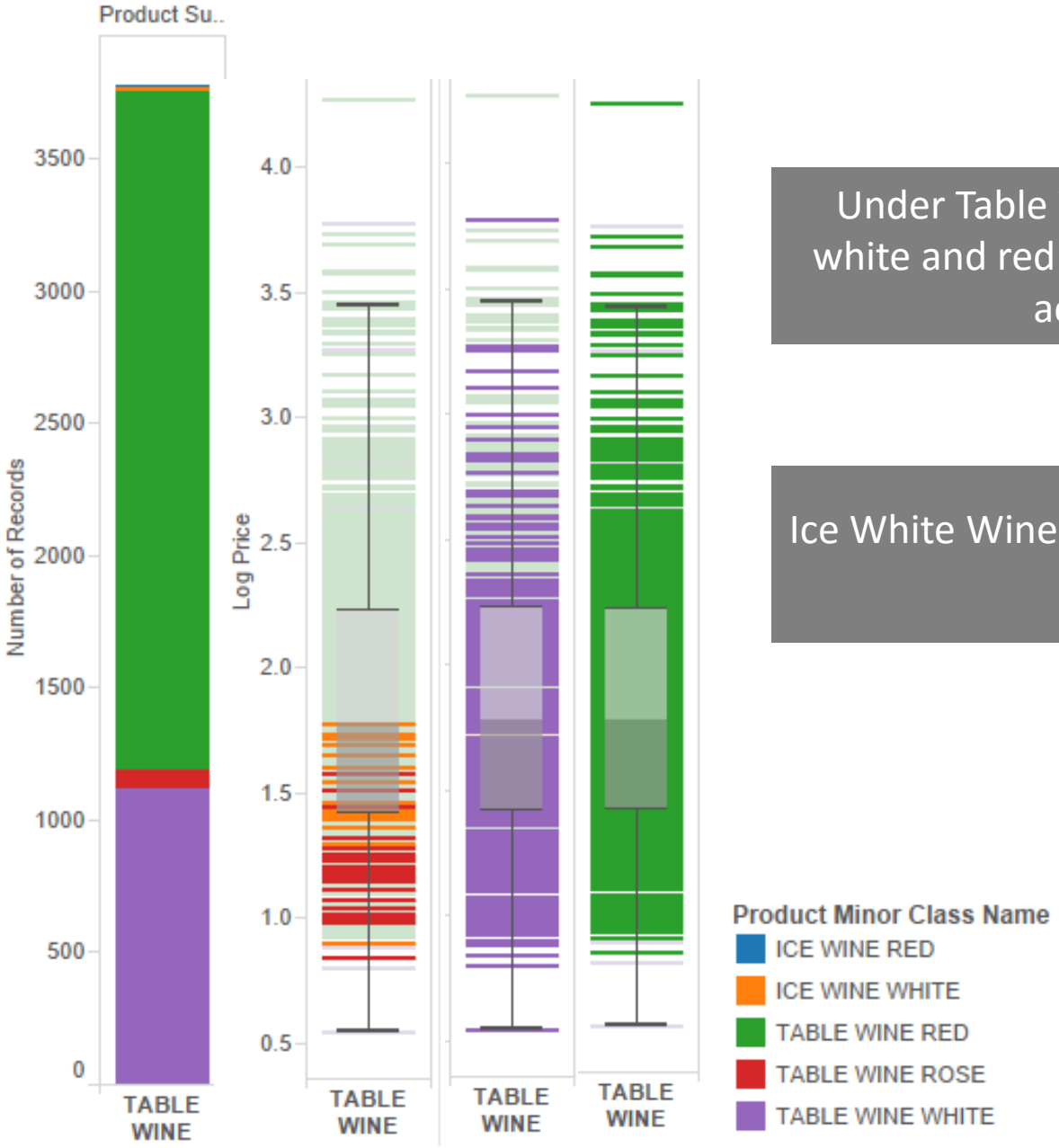
First, how does sparkling wine differ from the dessert wine category?

By definition sparkling wine would be considered as a dessert wine. However, sparking wine in particular is carbonated while the common dessert wine is not

Separate categorization is likely due to the volume and variety of sparkling wine imported

Sparkling wine is broken down by Red, Rose and White wine. However, this factor does not appear to affect price

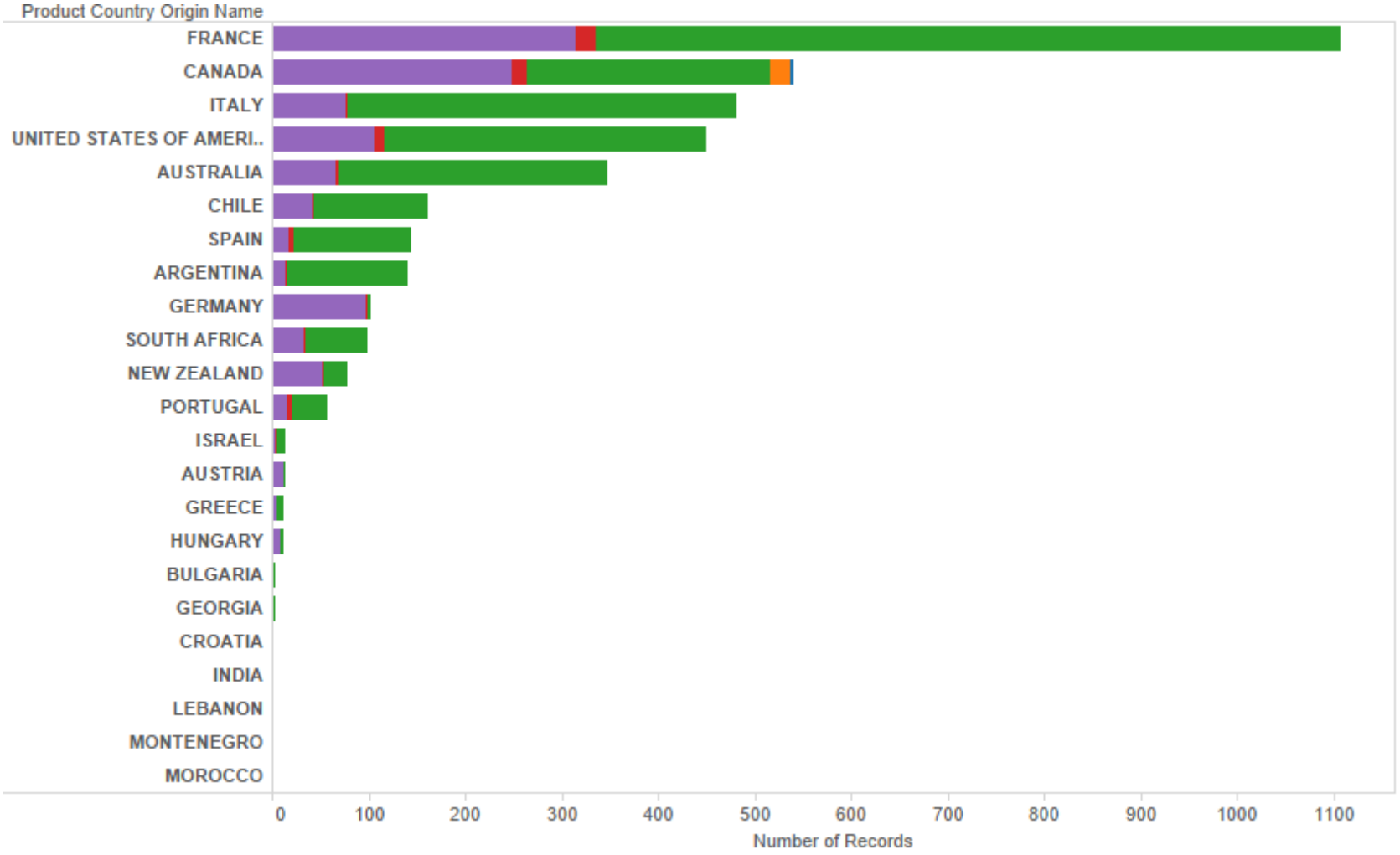
TABLE WINE AND PRICING



Under Table Wine category, we primarily see white and red wine products evenly distributed across all price points.

Ice White Wine and Rose Table Wine tends have a lower priced

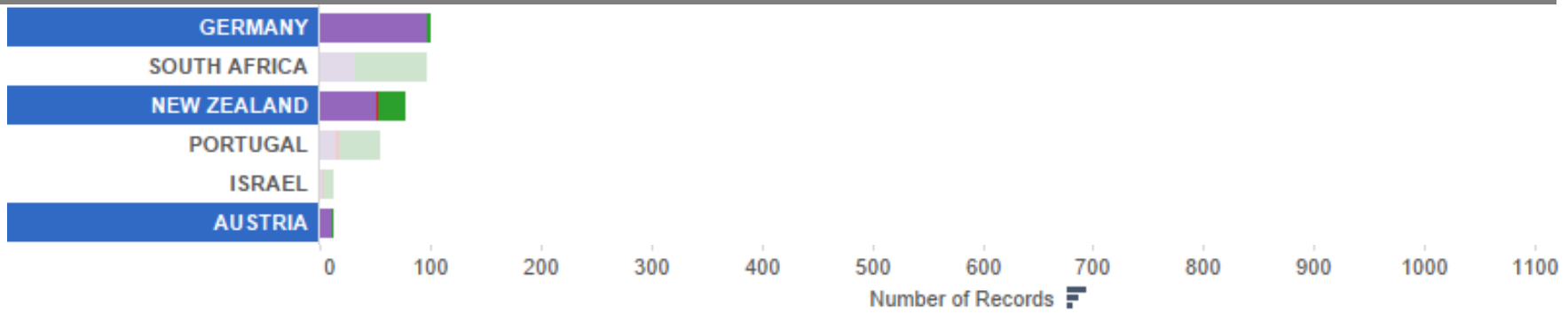
TABLE WINE IMPORT ORIGIN



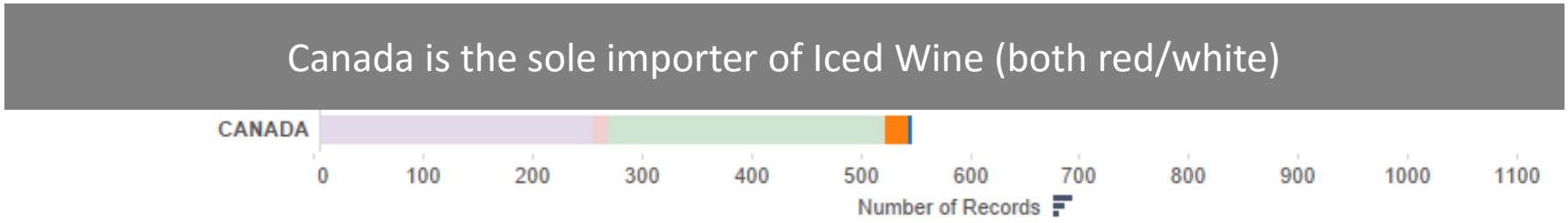
Highest varieties of wine is imported from France

EXCLUSIVE WINE CLASS IMPORTERS

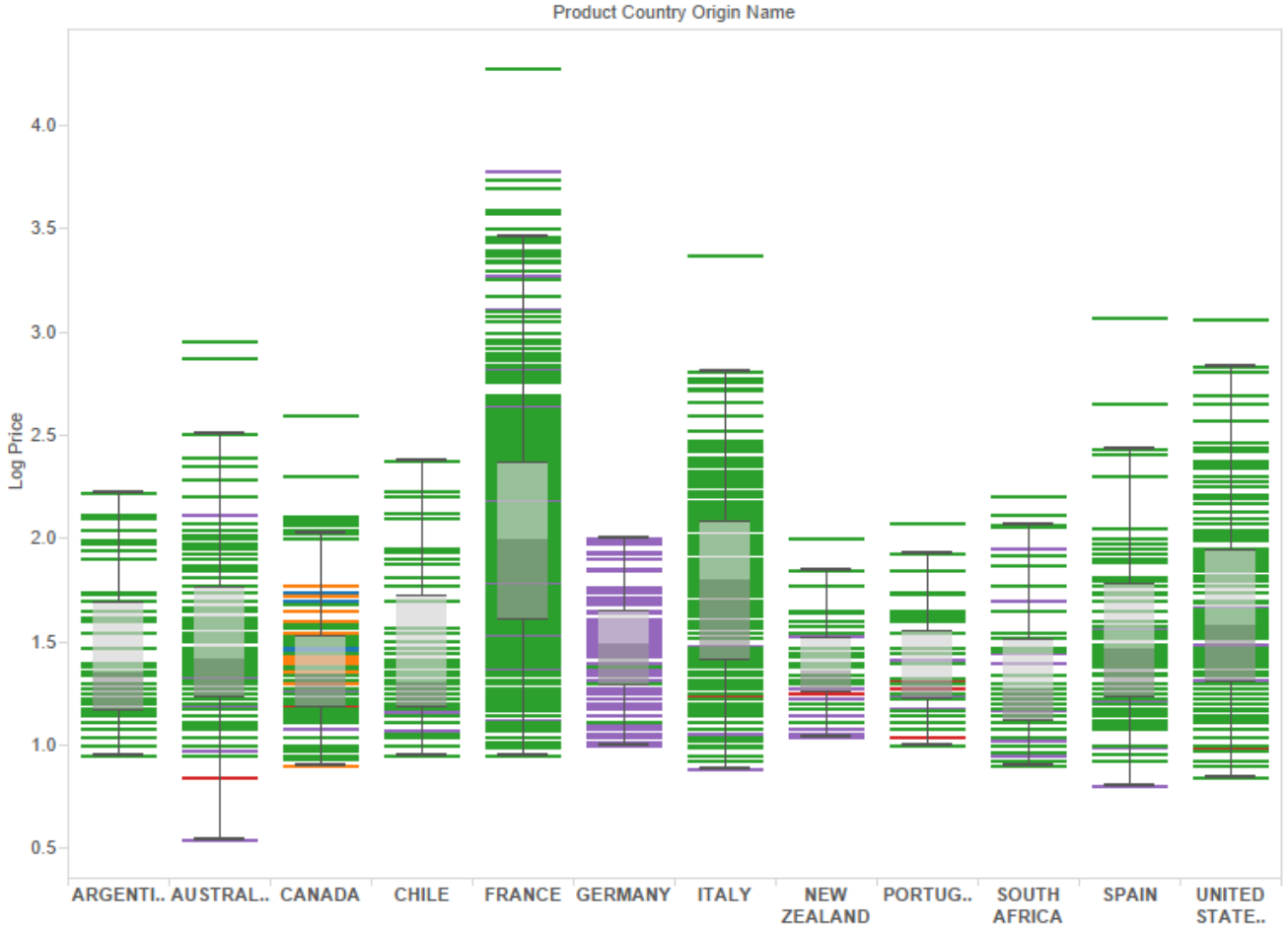
Wine import from Germany, New Zealand and Austria are near exclusively white wine



Canada is the sole importer of Iced Wine (both red/white)



RED WINE IMPORTER BY CONTINENT



Using log 2.5 as cut-off (~$300), wine will generally come from Australia, France, Italy, Spain or United States. (Mainly European importers)

Wine with price greater than $1000 (log 3) comes primarily from France

A QUICK REGRESSION OUTPUT

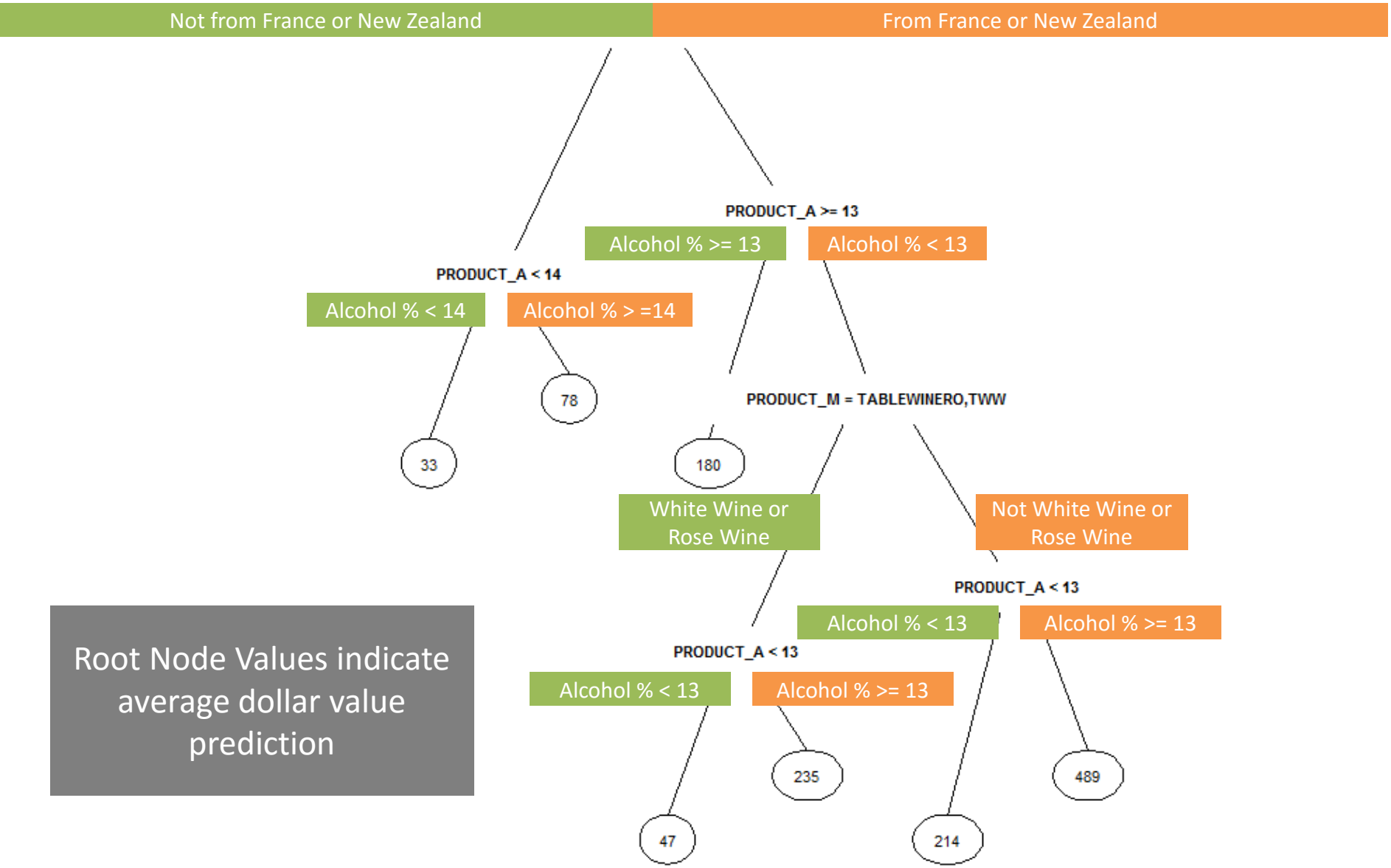
```
Call:
lm(formula = LogPrice ~ SWEETNESS_CODE + PRODUCT_COUNTRY_ORIGIN_NAME +
    PRODUCT_ALCOHOL_PERCENT, data = wine)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1977 -0.5574 -0.1378  0.4182  5.4584

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                                  -0.36147    0.22159  -1.631 0.102914
SWEETNESS_CODE                                0.06121    0.01094   5.596 2.36e-08 ***
PRODUCT_COUNTRY_ORIGIN_NAMEAUSTRALIA          0.15388    0.09157   1.680 0.092951 .
PRODUCT_COUNTRY_ORIGIN_NAMEAUSTRIA            0.63770    0.26579   2.399 0.016479 *
PRODUCT_COUNTRY_ORIGIN_NAMEBULGARIA          -0.04042    0.53383  -0.076 0.939650
PRODUCT_COUNTRY_ORIGIN_NAMECANADA             0.32123    0.08853   3.629 0.000289 ***
PRODUCT_COUNTRY_ORIGIN_NAMECHILE              0.13231    0.10544   1.255 0.209608
PRODUCT_COUNTRY_ORIGIN_NAMECROATIA            0.97515    0.91824   1.062 0.288315
PRODUCT_COUNTRY_ORIGIN_NAMEFRANCE             1.63543    0.08240  19.848  < 2e-16 ***
PRODUCT_COUNTRY_ORIGIN_NAMEGEORGIA           -0.05753    0.53403  -0.108 0.914222
PRODUCT_COUNTRY_ORIGIN_NAMEGERMANY            0.97705    0.12861   7.597 3.82e-14 ***
PRODUCT_COUNTRY_ORIGIN_NAMEGREECE             0.34787    0.26637   1.306 0.191642
PRODUCT_COUNTRY_ORIGIN_NAMEHUNGARY            0.23562    0.26815   0.879 0.379619
PRODUCT_COUNTRY_ORIGIN_NAMEINDIA             -0.06434    0.91824  -0.070 0.944138
PRODUCT_COUNTRY_ORIGIN_NAMEISRAEL             0.18756    0.25648   0.731 0.464658
PRODUCT_COUNTRY_ORIGIN_NAMEITALY              0.90243    0.08778  10.280  < 2e-16 ***
PRODUCT_COUNTRY_ORIGIN_NAMELEBANON            1.33580    0.91825   1.455 0.145832
PRODUCT_COUNTRY_ORIGIN_NAMEMONTENEGRO        -0.16089    0.91825  -0.175 0.860925
PRODUCT_COUNTRY_ORIGIN_NAMEMOROCCO            0.61571    0.91819   0.671 0.502536
PRODUCT_COUNTRY_ORIGIN_NAMENEW ZEALAND        0.30098    0.12944   2.325 0.020117 *
PRODUCT_COUNTRY_ORIGIN_NAMEPORTUGAL           0.34486    0.14511   2.377 0.017523 *
PRODUCT_COUNTRY_ORIGIN_NAMESOUTH AFRICA       0.00977    0.12035   0.081 0.935303
PRODUCT_COUNTRY_ORIGIN_NAMESPAIN              0.36201    0.10842   3.339 0.000849 ***
PRODUCT_COUNTRY_ORIGIN_NAMEUNITED STATES OF AMERICA  0.51163    0.08833   5.792 7.54e-09 ***
PRODUCT_ALCOHOL_PERCENT                       0.23999    0.01507  15.928  < 2e-16 ***


Residual standard error: 0.9149 on 3669 degrees of freedom
  (81 observations deleted due to missingness)
Multiple R-squared:  0.3369, Adjusted R-squared:  0.3325
F-statistic: 77.66 on 24 and 3669 DF,  p-value: < 2.2e-16
```

QUICK REGRESSION TREE PREDICTION



Root Node Values indicate average dollar value prediction

TEXT MINING AND PRODUCT NAME

	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT_	PRODUCT	PRODUCT_	PRODUCT_LONG_NAME	PRODUCT_	PRODUCT_	PRD_CON	PRODUCT_	CURRENT_	SWEETNESS_CODE
2	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	198267	COPPER MOON - MALBEC	4.82E+10	3	1	14	30.99	0
3	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	305375	DOMAINE D'OR - DRY	4.82E+10	4	1	11.5	32.99	0
4	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	53017	SOMMET ROUGE	5.9E+10	4	1	12	29.99	0
5	LIQUOR	WINE	TABLE WI	TABLE WI	CANADA	215525	MISSION RIDGE - PREMIUM DRY WHITE	7.8E+11	4	1	11	33.99	1

In the given dataset, Product Long Name contains features that could likely increase prediction accuracy, especially for wines.

Ideal features include: brand, vineyard name and importantly, grape variety

Challenge 1

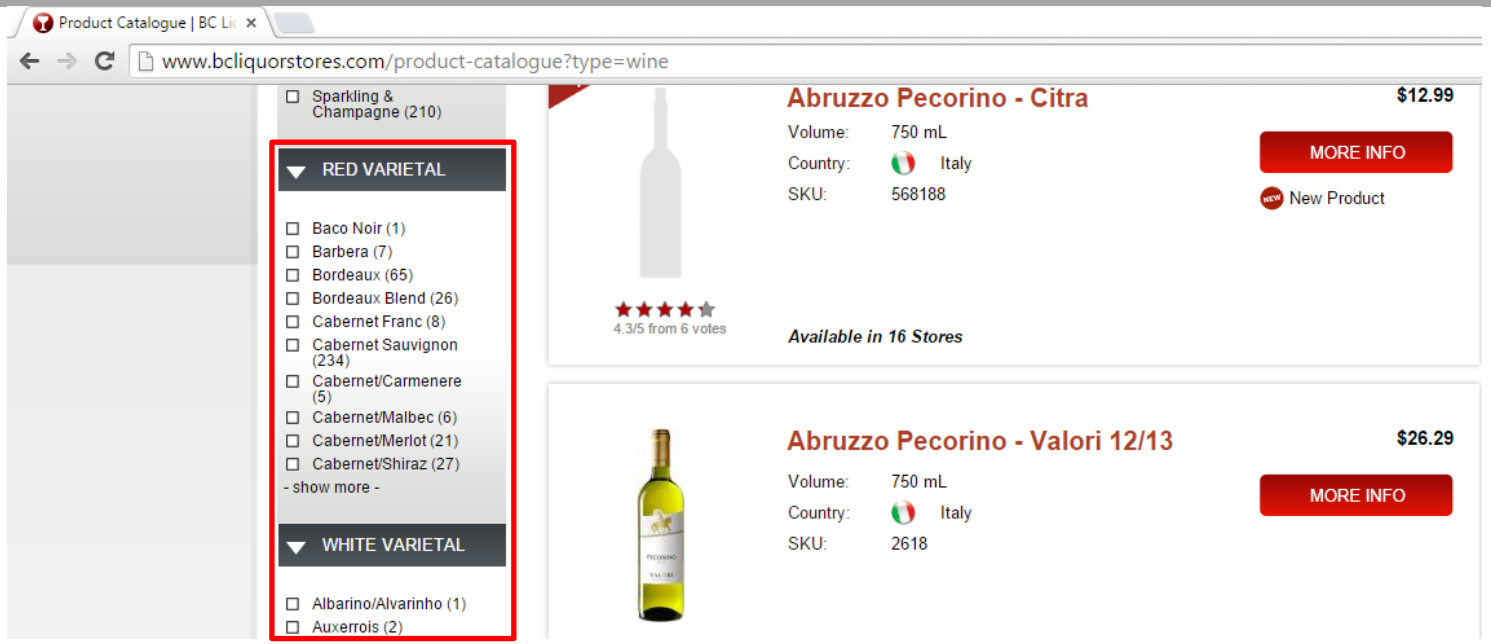
However, not every product label include all of these factors yield large number of records with NULL values

Challenge 2

From text-mining, a term frequency matrix is the output. However, Tableau visualizations require normalized format. To fully visualize text data, custom-scripting will be required

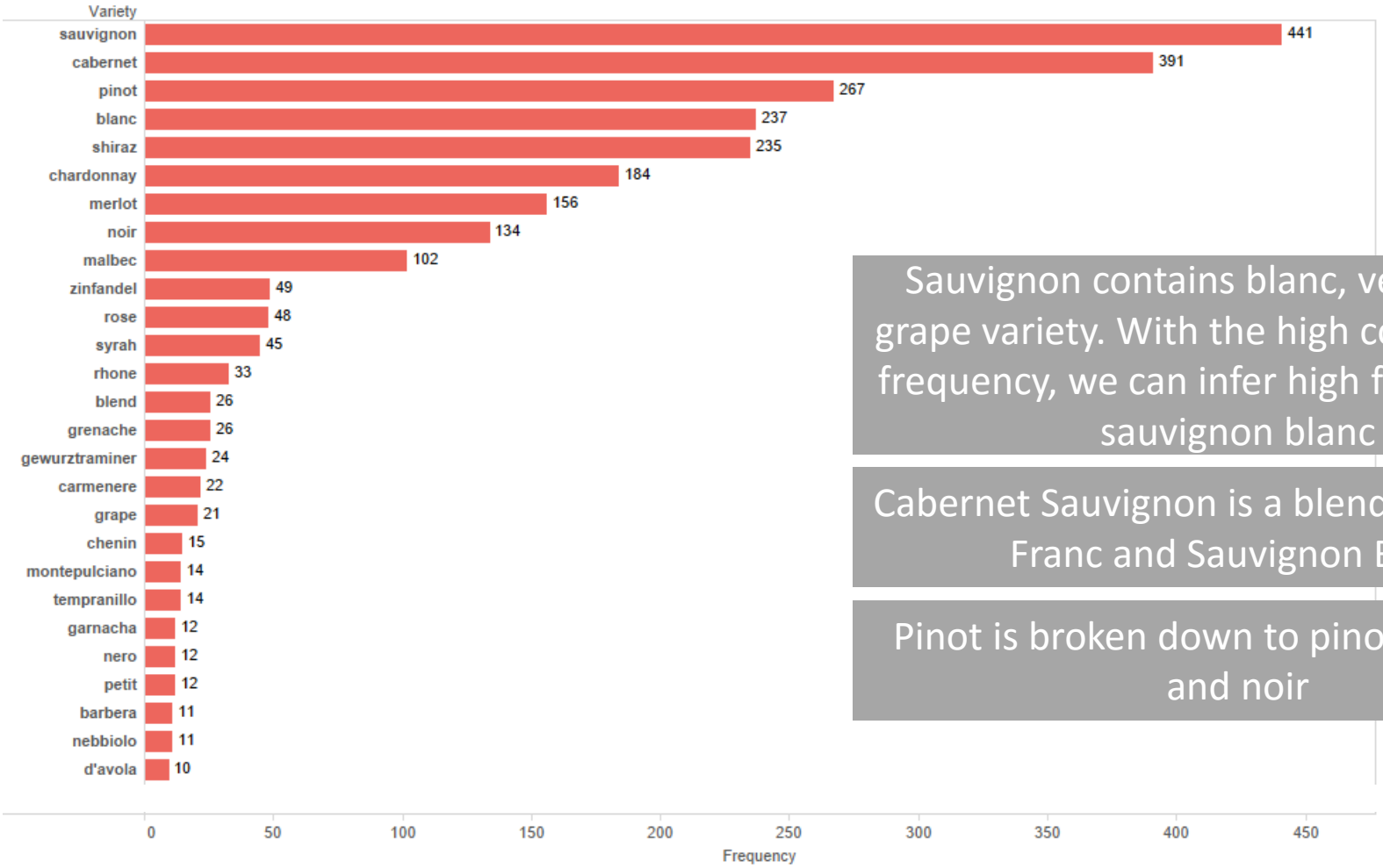
GRAPE VARIETY FREQUENCY

Grape variety keywords are scraped from BC Liquor Website



Of 3775 table wine records, 1955 records do not have grape variety stated within product title

GRAPE VARIETY FREQUENCY



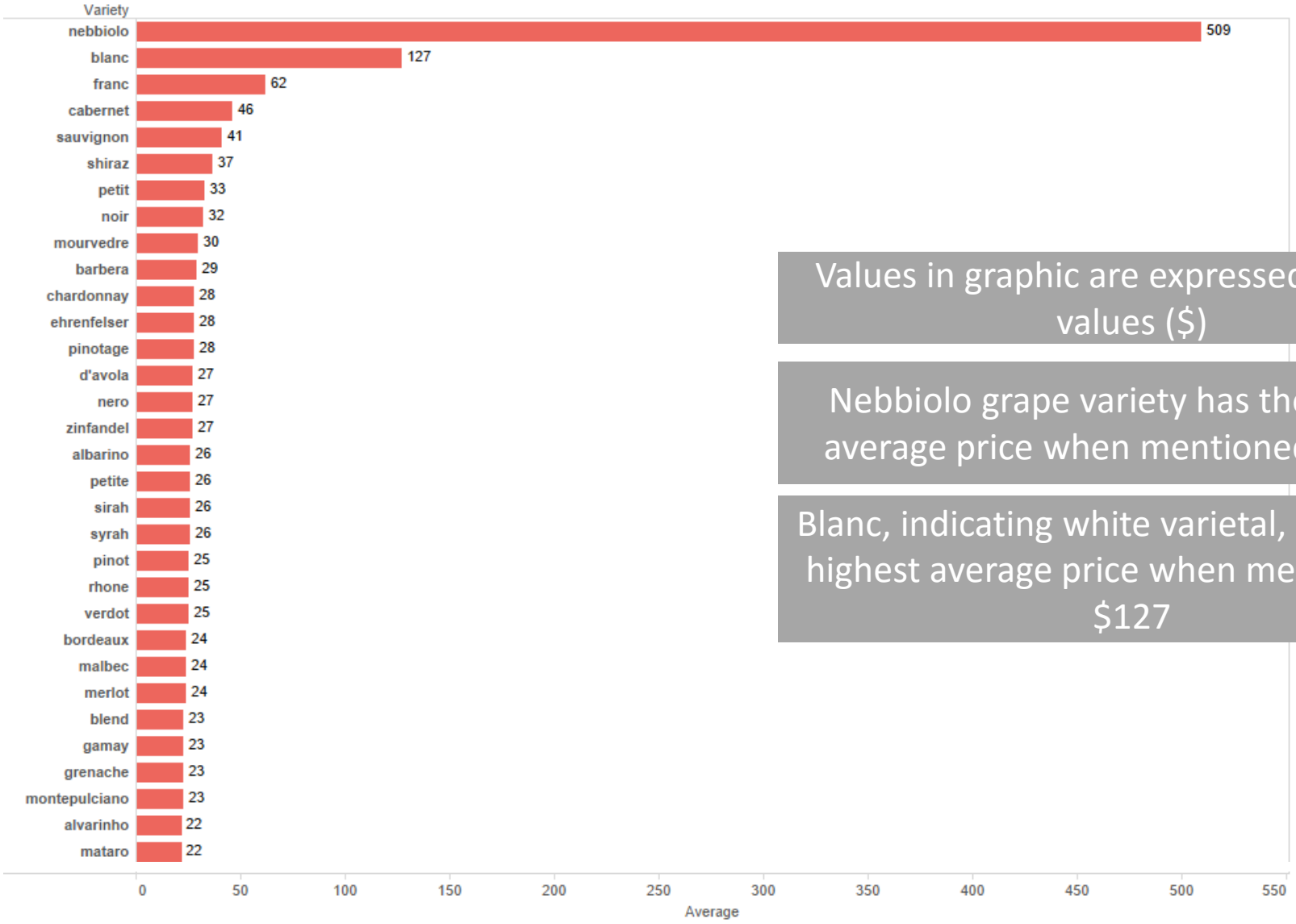
Sauvignon contains blanc, vert and gris grape variety. With the high count in blanc frequency, we can infer high freiquency of sauvignon blanc

Cabernet Sauvignon is a blend of Cabernet Franc and Sauvignon Blanc

Pinot is broken down to pinot blanc, gris and noir

Common grape variety amongst products include Sauvignon Blanc, Cabernet Sauvignon, Pinot Blanc/Noir, blends of Shiraz/Syrah and Chardonnay

GRAPE VARIETY AND PRICE



Values in graphic are expressed as dollar values ($)

Nebbiolo grape variety has the highest average price when mentioned at $509

Blanc, indicating white varietal, has the 2nd highest average price when mentioned at $127

Updated findings for Grape Variety Text-Mining can be found here:
https://public.tableau.com/profile/rock.chi5163#!/vizhome/BCLiquor/Compiled

More of my work could be found at:
https://github.com/rockchi