Capstone Two

A study of Vinho Verde wines and which features correlate with quality.

What is Vinho Verde and Where Did the Data Come From?

-Vinho Verde is a Portuguese wine.

    -Translates to ‘young wine’.

-The Datasets are hosted on both Kaggle and the UCI ML Repository.

    -Consists of two datasets: red, and white wine.

A Look At The Datasets

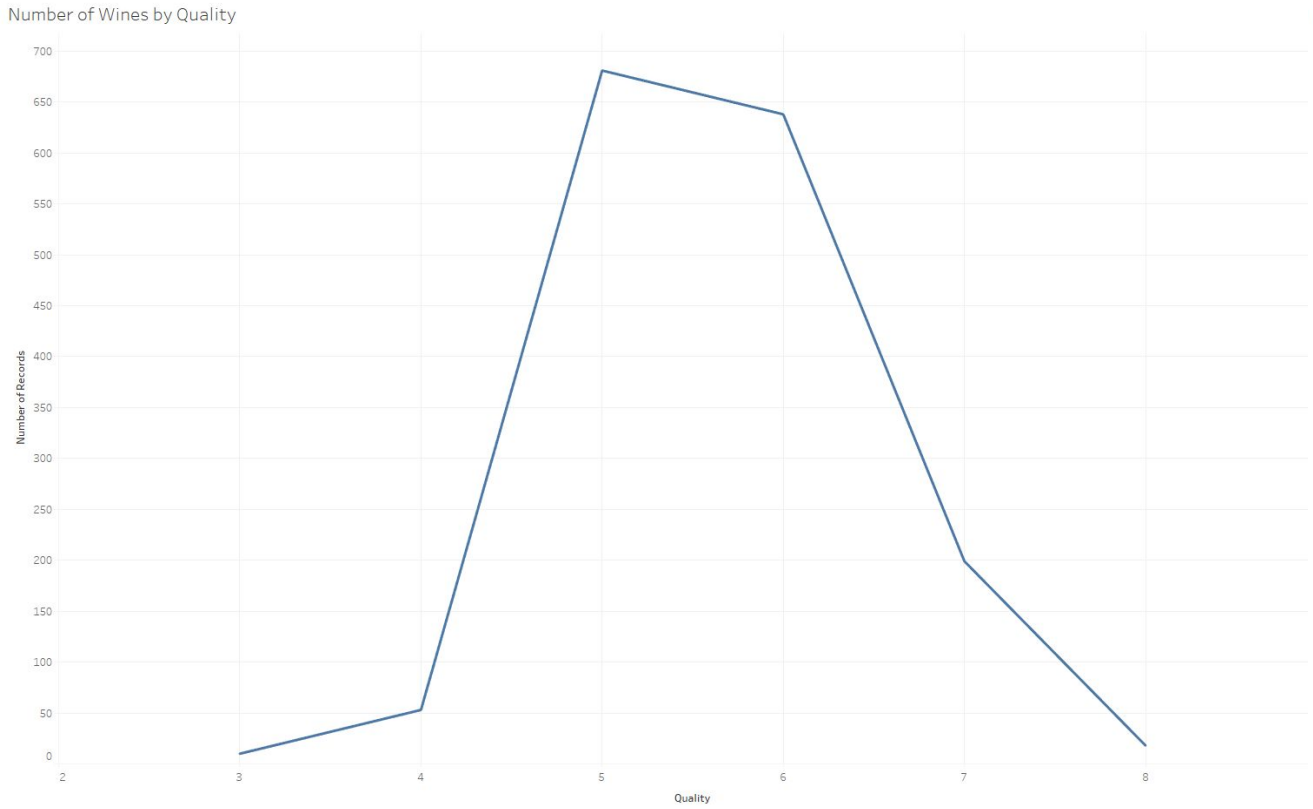
```
[8] df_red.head()
```

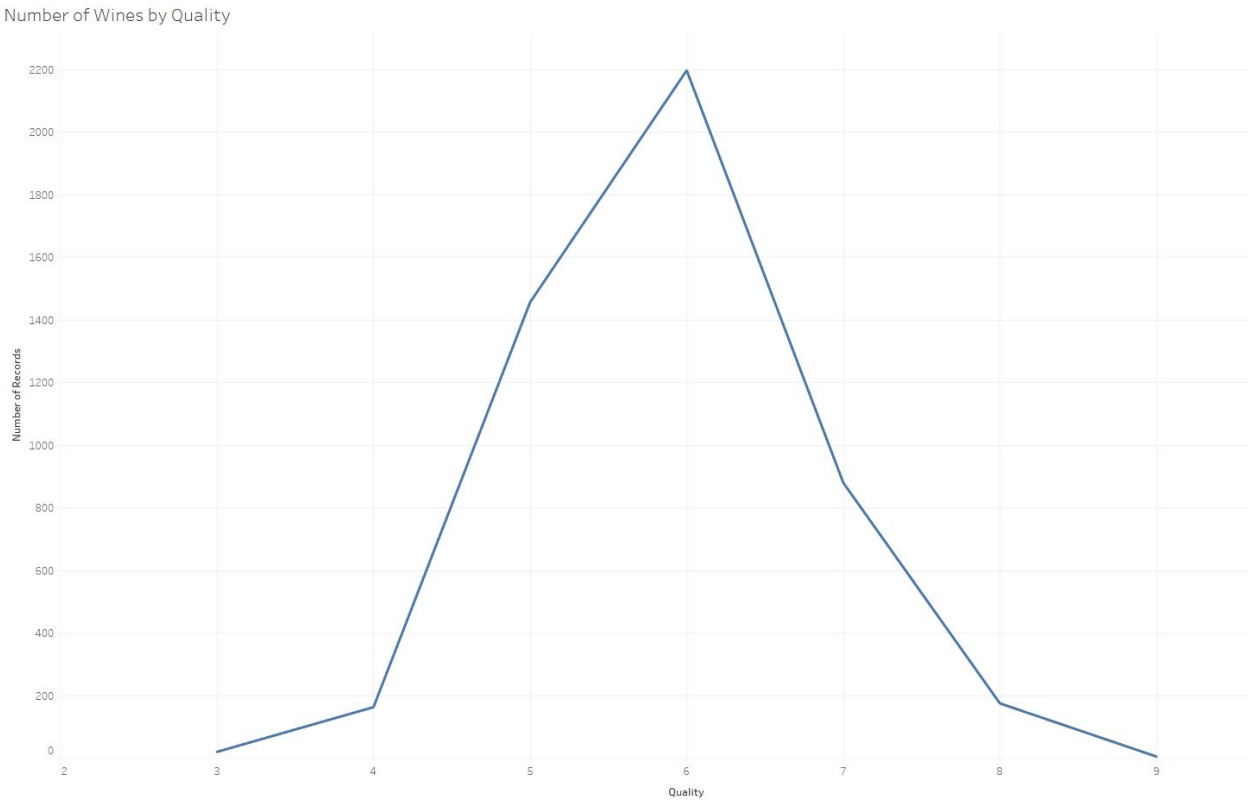
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

```
[9] df_white.head()
```

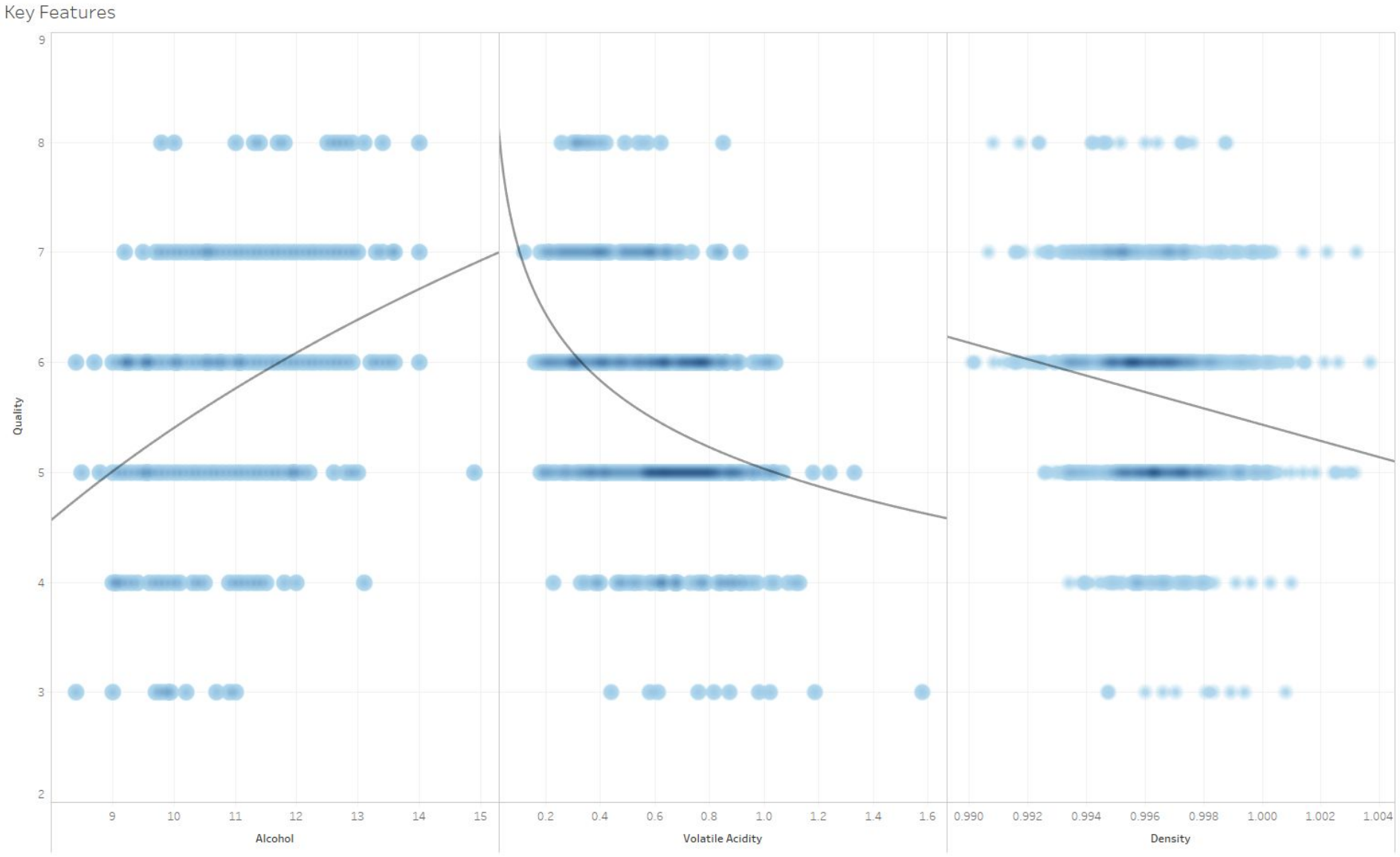
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
1	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
2	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
3	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
4	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6

Number of Wines by Quality: Red (left) and White (right)

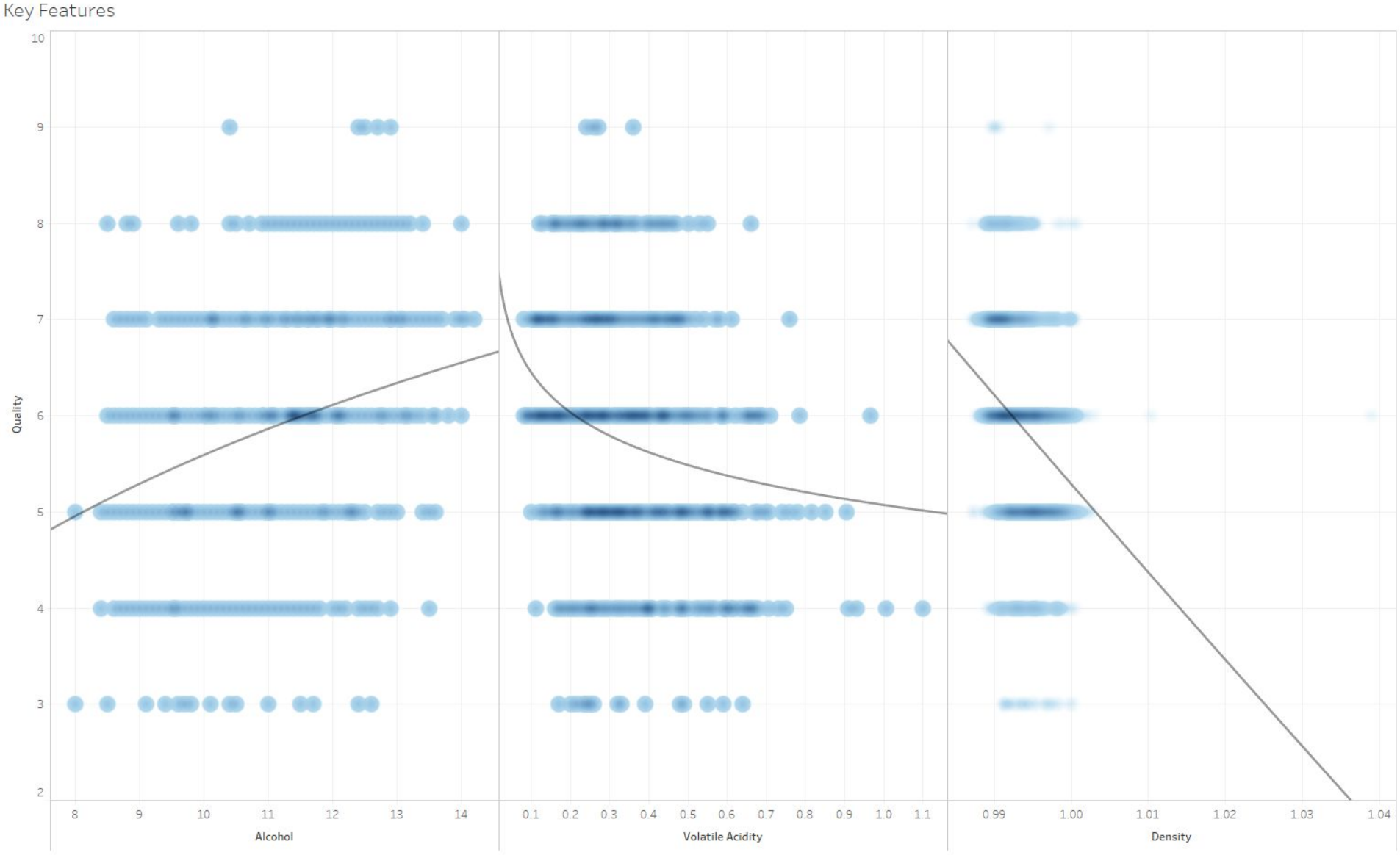




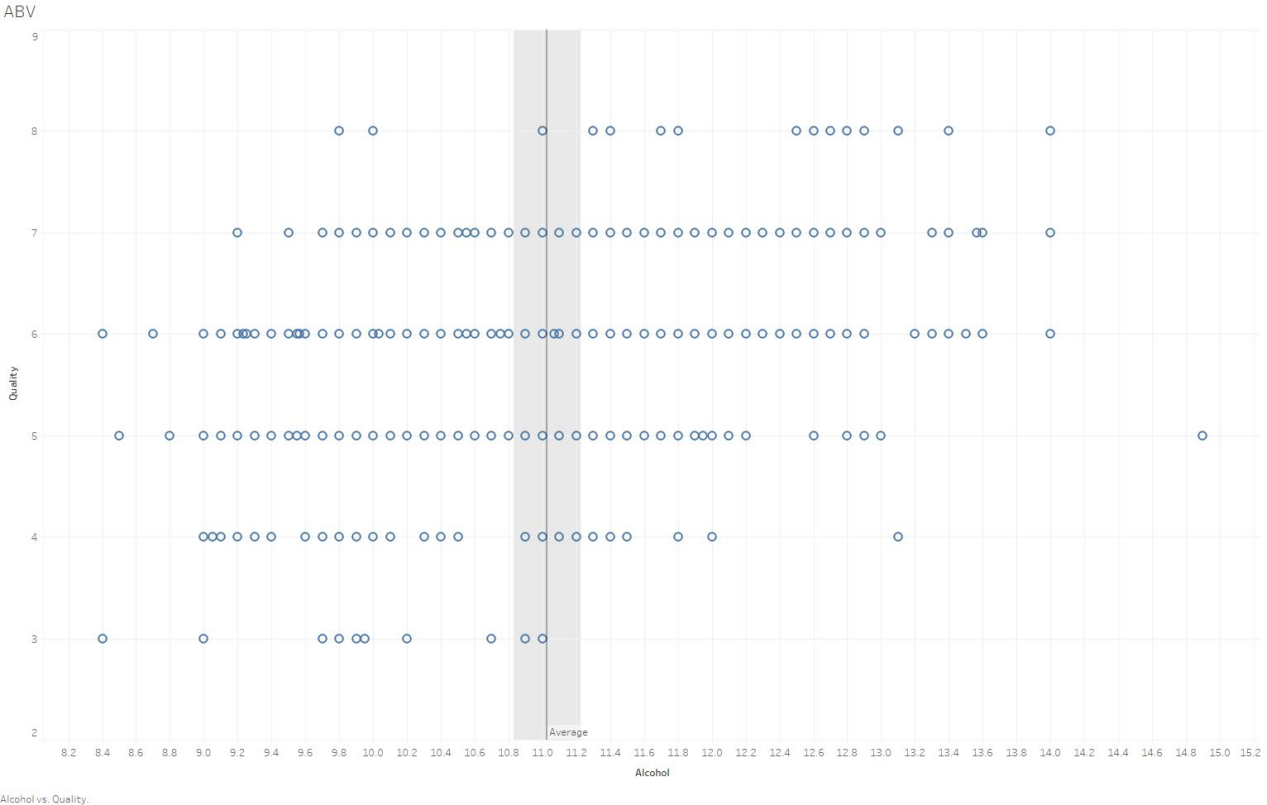
Red Wine Key Features

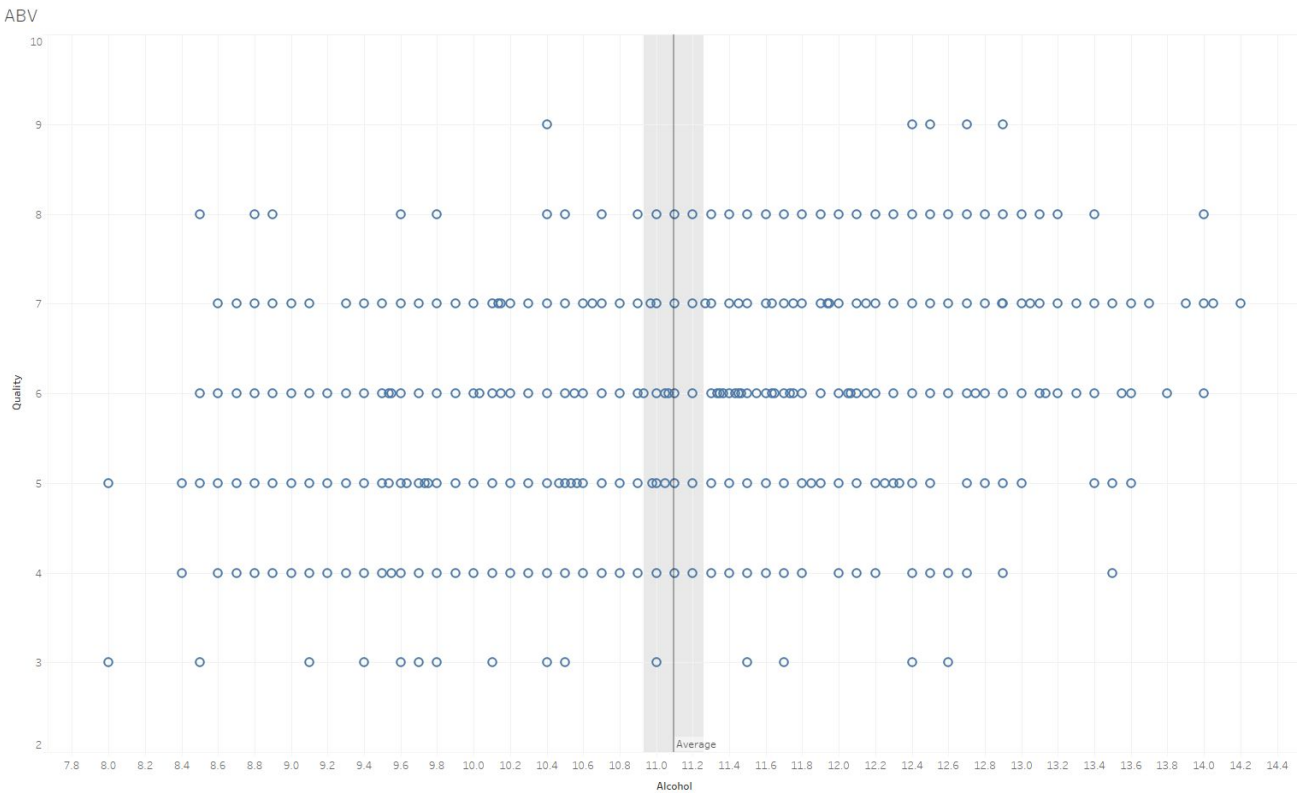


White Wine Key Features



Closer Look at ABV:
Red (left) and White (right)



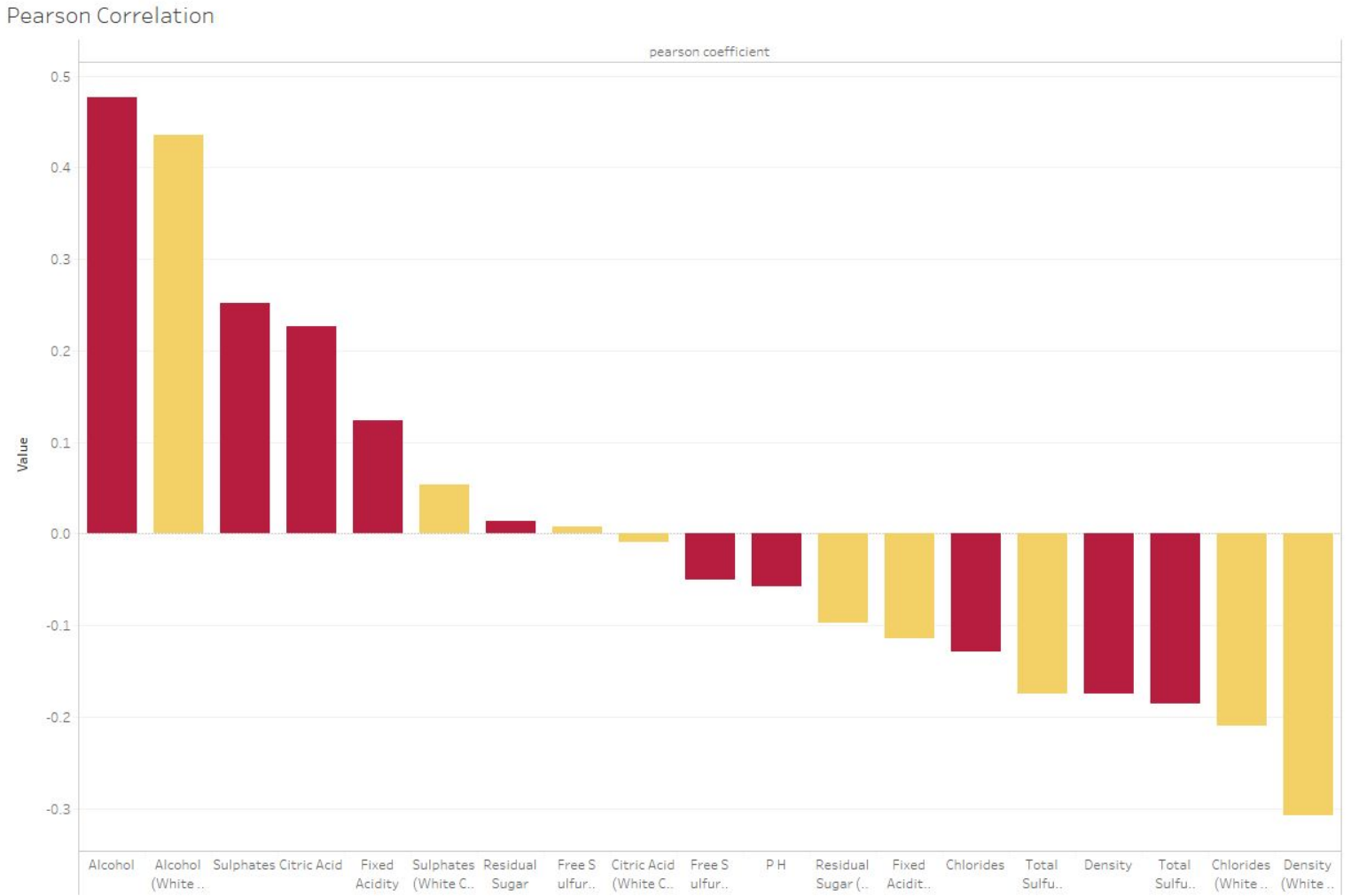


Hypothesis:

H: There is a positive correlation between ABV and quality.

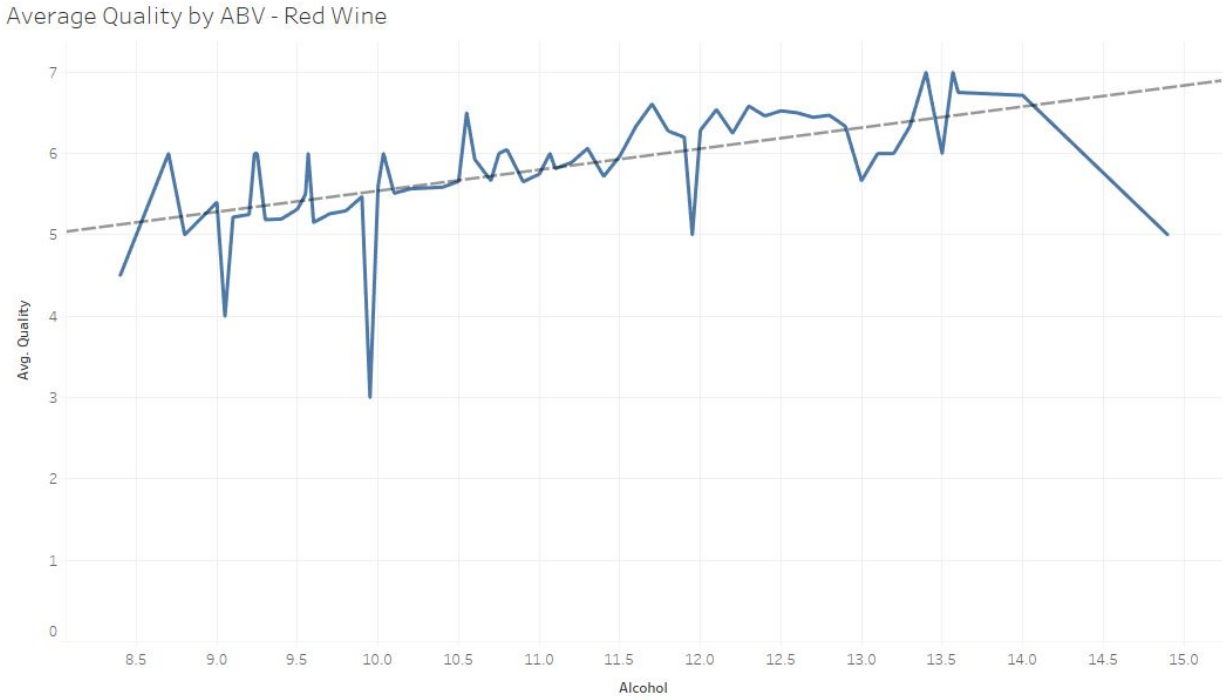
H_0: There is no correlation between quality and ABV.

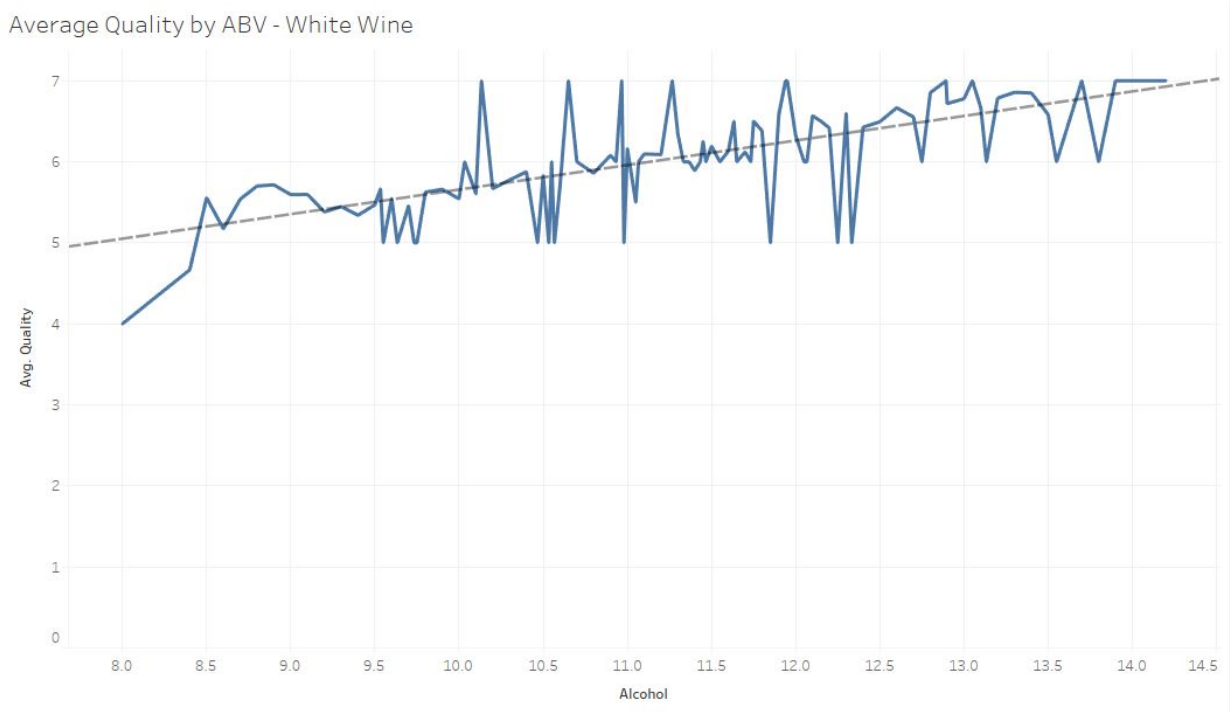
Testing:
Pearson Correlation Coefficient



Conclusion:

The data support the conclusion that brewing higher abv wine will result in a higher quality wine.
For red wine this trend stops at 14% ABV.
For white wine this trend stops at 14.2% ABV.





Future Research:

-Search for a confounding variable for ABV.

-Find if this trend exists in different regions.

-Find out if quality correlates to sales/profits.

Supplemental Graphs

Testing:
Pearson Correlation Coefficient

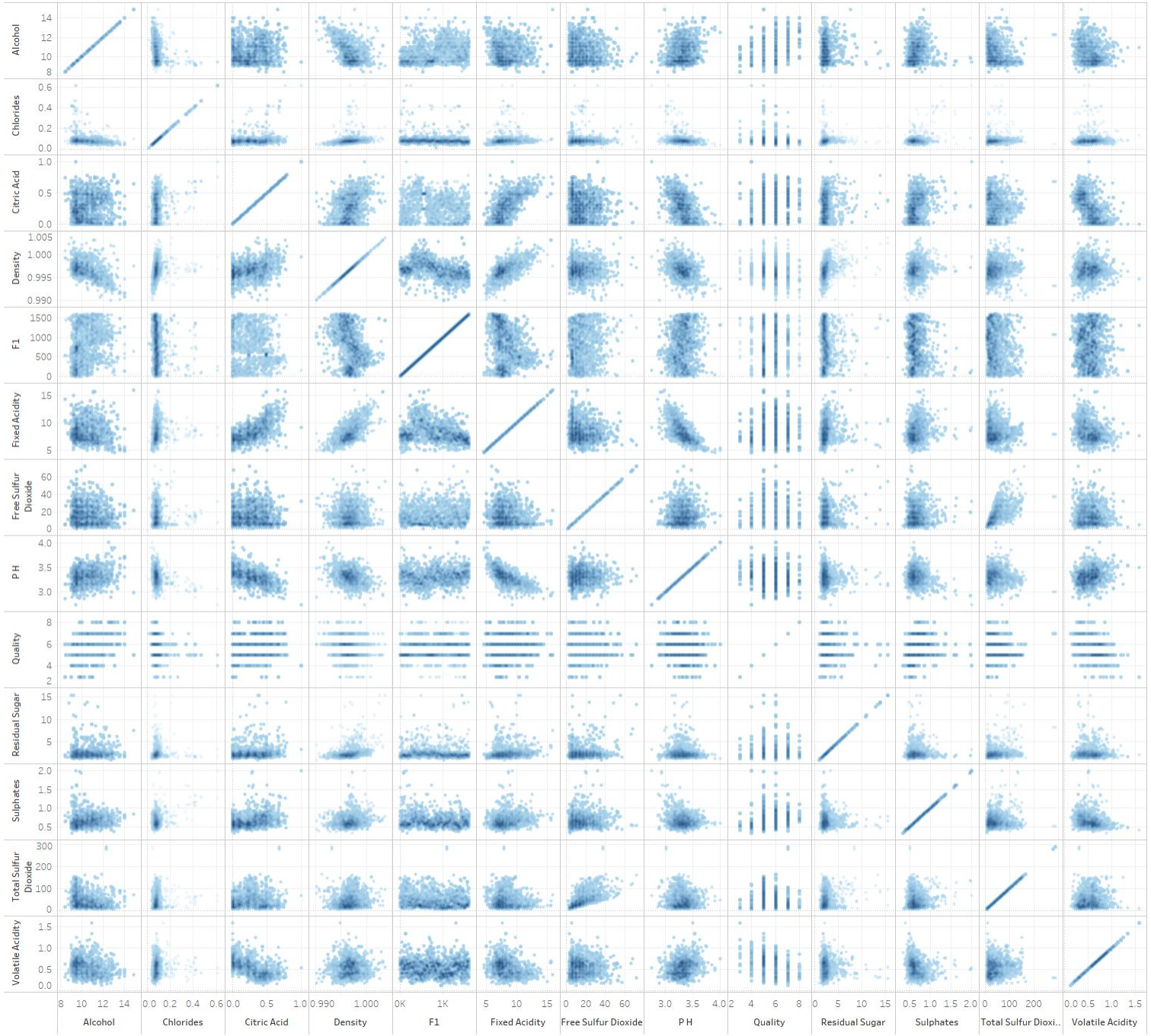
Red

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
pearson coefficient	1.240516e-01	-3.905578e-01	2.263725e-01	0.013732	-1.289066e-01	-0.050656	-1.851003e-01	-1.749192e-01	-0.057731	2.513971e-01	4.761663e-01
p-val	6.495635e-07	2.051715e-59	4.991295e-20	0.583218	2.313383e-07	0.042834	8.621703e-14	1.874957e-12	0.020963	1.802088e-24	2.831477e-91

White

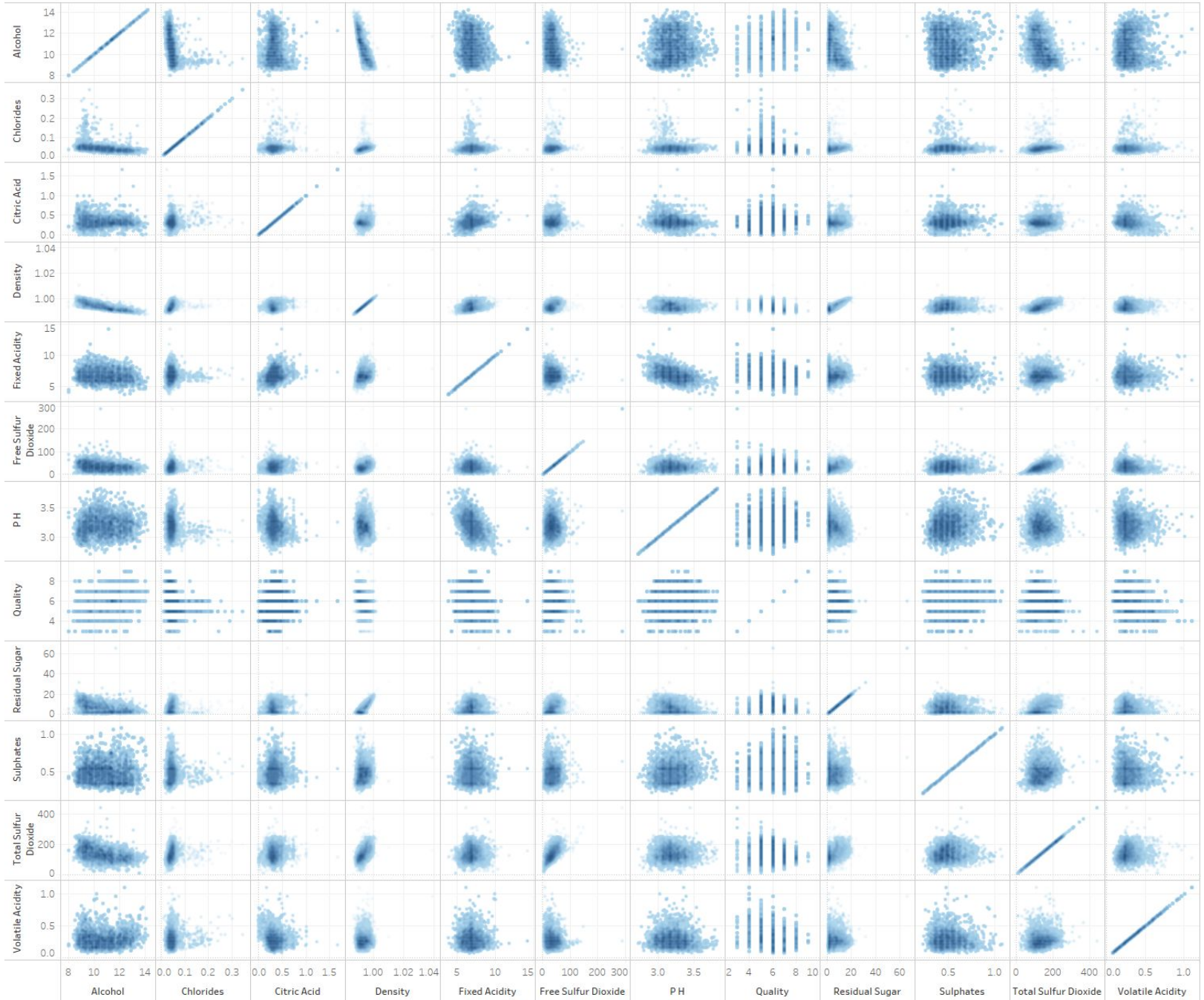
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
pearson coefficient	-1.136628e-01	-1.947230e-01	-0.009209	-9.757683e-02	-2.099344e-01	0.008158	-1.747372e-01	-3.071233e-01	9.942725e-02	0.053678	4.355747e-01
p-val	1.479886e-15	4.673261e-43	0.519346	7.724005e-12	6.506542e-50	0.568127	6.991898e-35	1.727988e-107	3.080613e-12	0.000171	5.614770e-226

Pair Plot



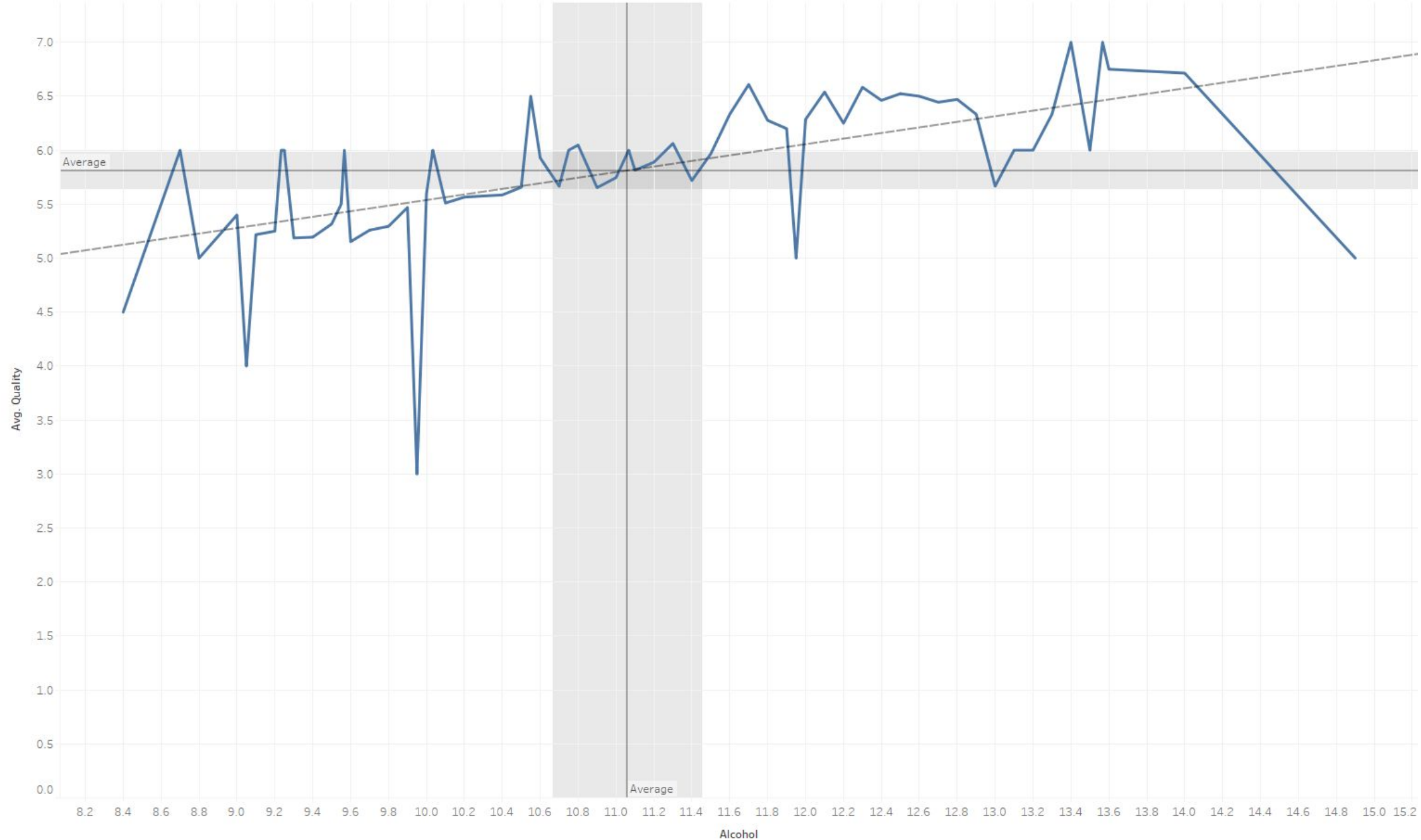
Alcohol, Chlorides, Citric Acid, Density, F1, Fixed Acidity, Free Sulfur Dioxide, P H, Quality, Residual Sugar, Sulphates, Total Sulfur Dioxide and Volatile Acidity vs. Alcohol, Chlorides, Citric Acid, Density, F1, Fixed Acidity, Free Sulfur Dioxide, P H, Quality, Residual Sugar, Sulphates, Total Sulfur Dioxide and Volatile Acidity.

White Pair Plot



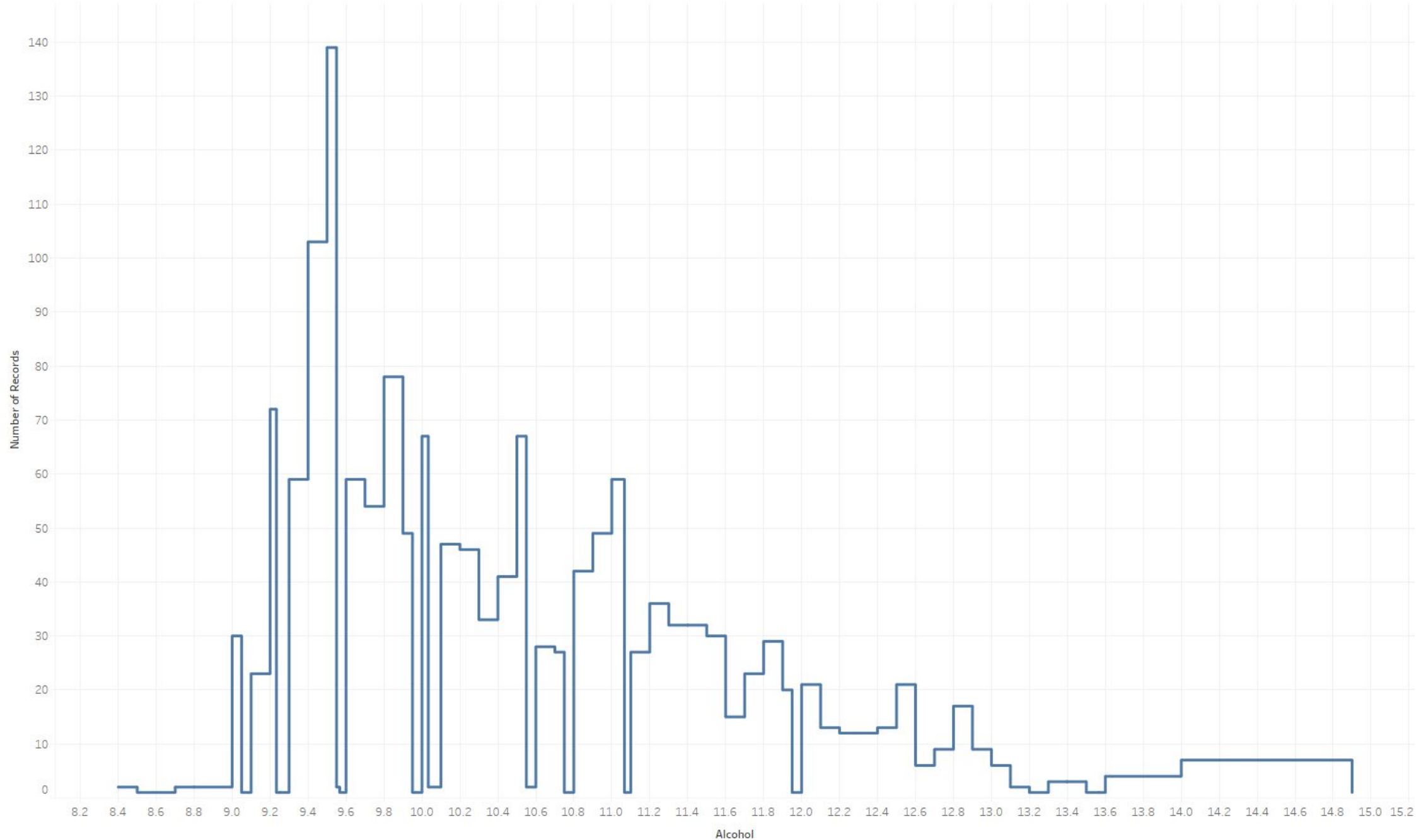
Alcohol, Chlorides, Citric Acid, Density, Fixed Acidity, Free Sulfur Dioxide, P H, Quality, Residual Sugar, Sulphates, Total Sulfur Dioxide and Volatile Acidity vs. Alcohol, Chlorides, Citric Acid, Density, Fixed Acidity, Free Sulfur Dioxide, P H, Quality, Residual Sugar, Sulphates, Total Sulfur Dioxide and Volatile Acidity.

Average Quality by ABV



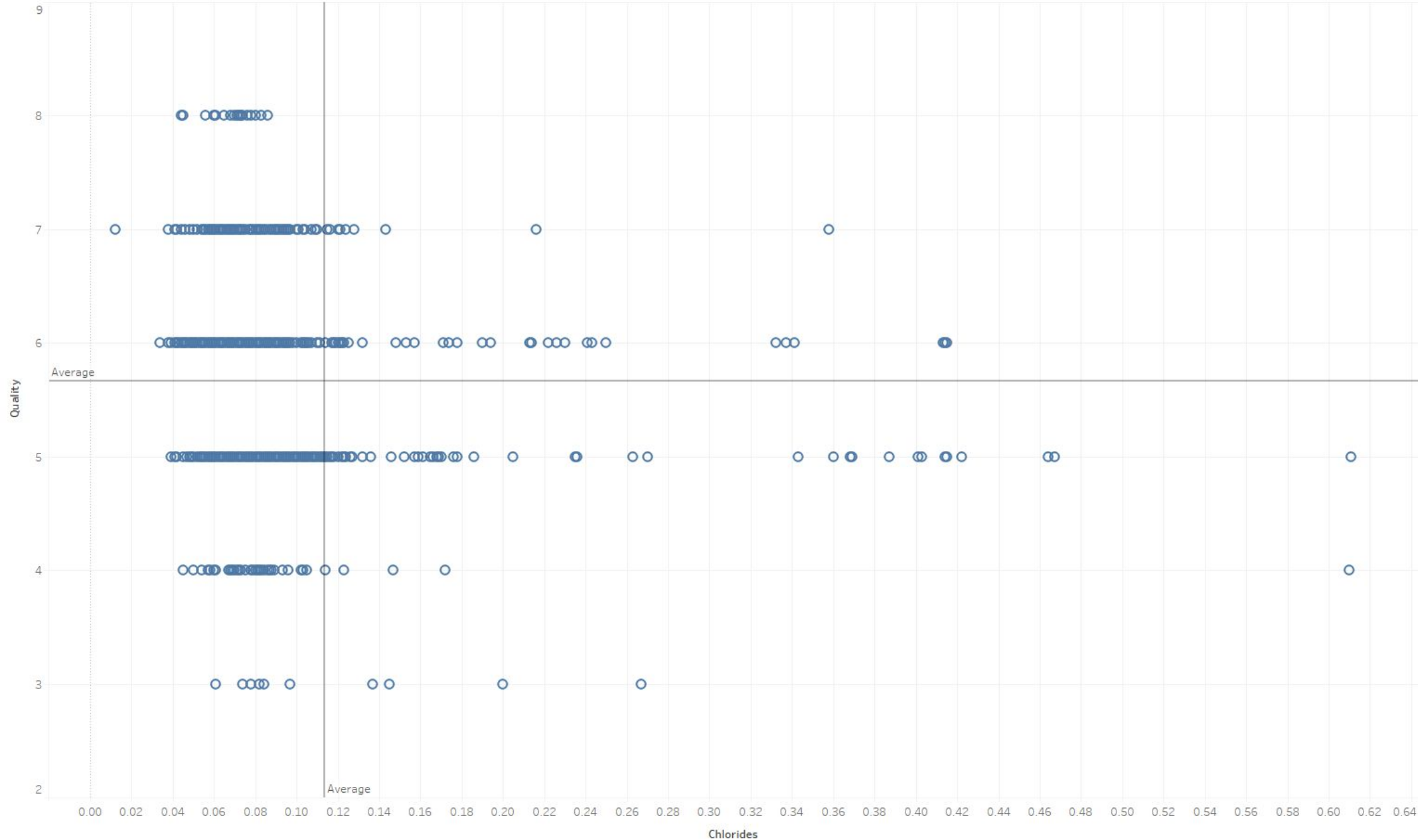
The trend of average of Quality for Alcohol.

Value Count by ABV



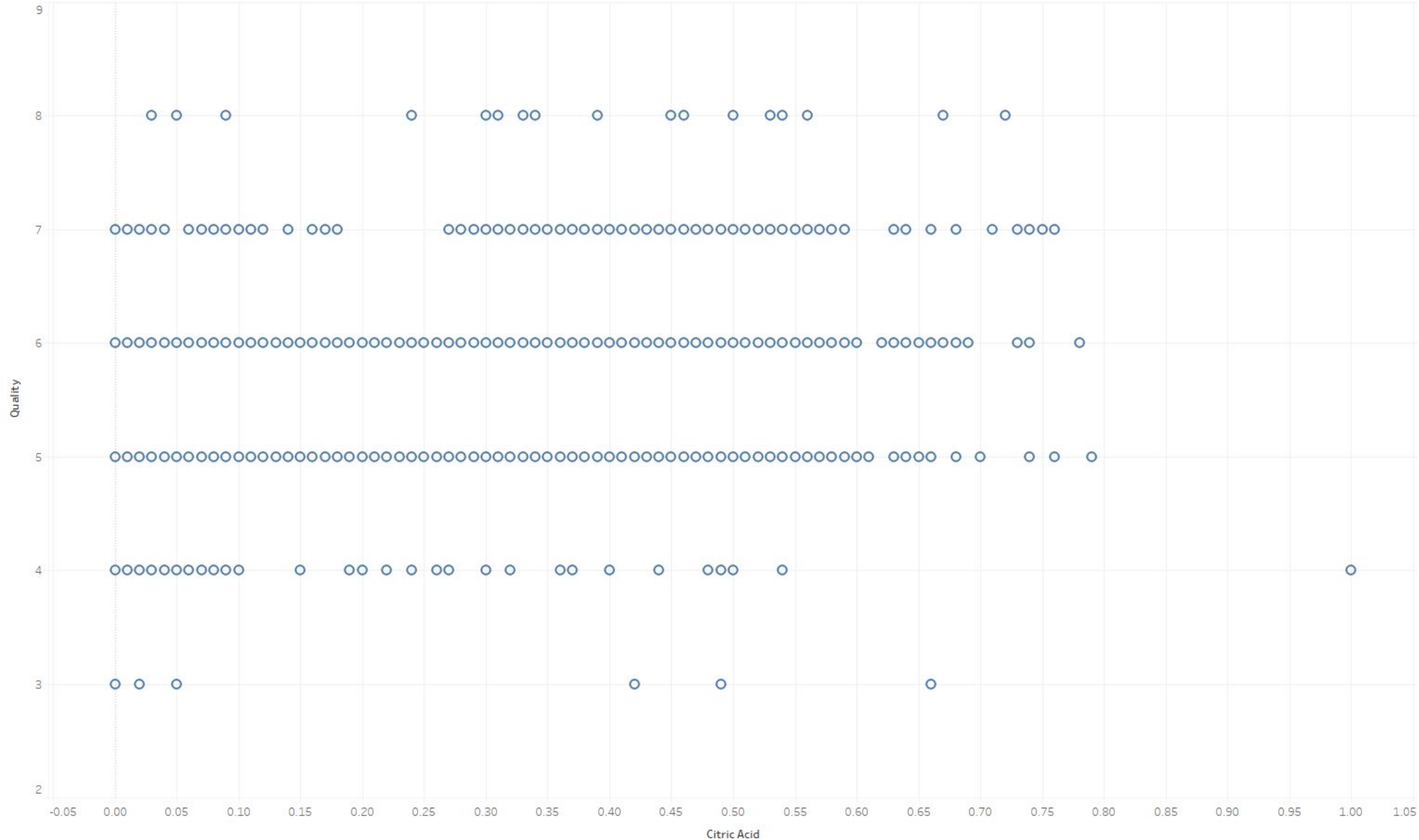
The trend of sum of Number of Records for Alcohol.

Chlorides

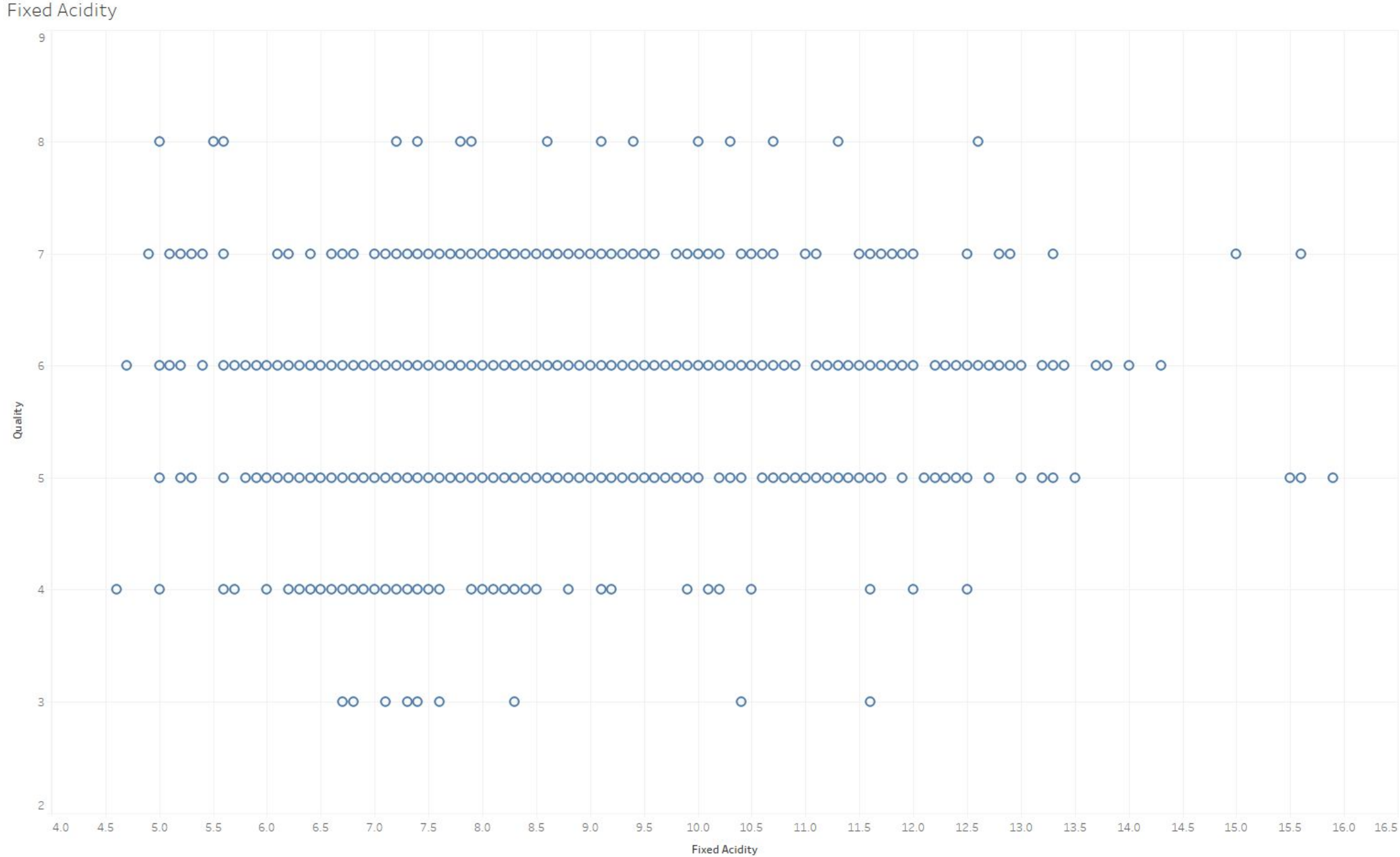


Chlorides vs. Quality.

Citric Acid

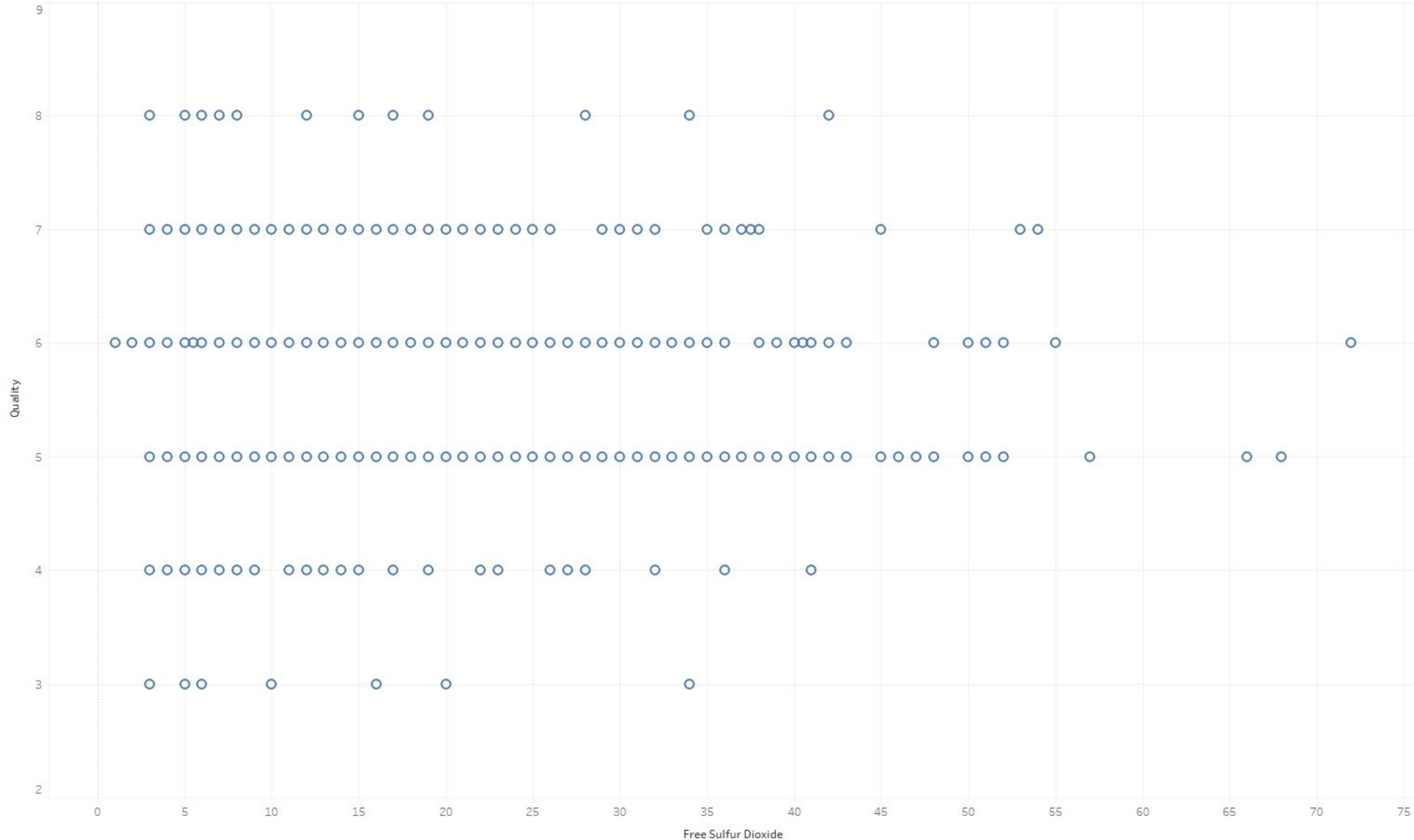


Citric Acid vs. Quality.

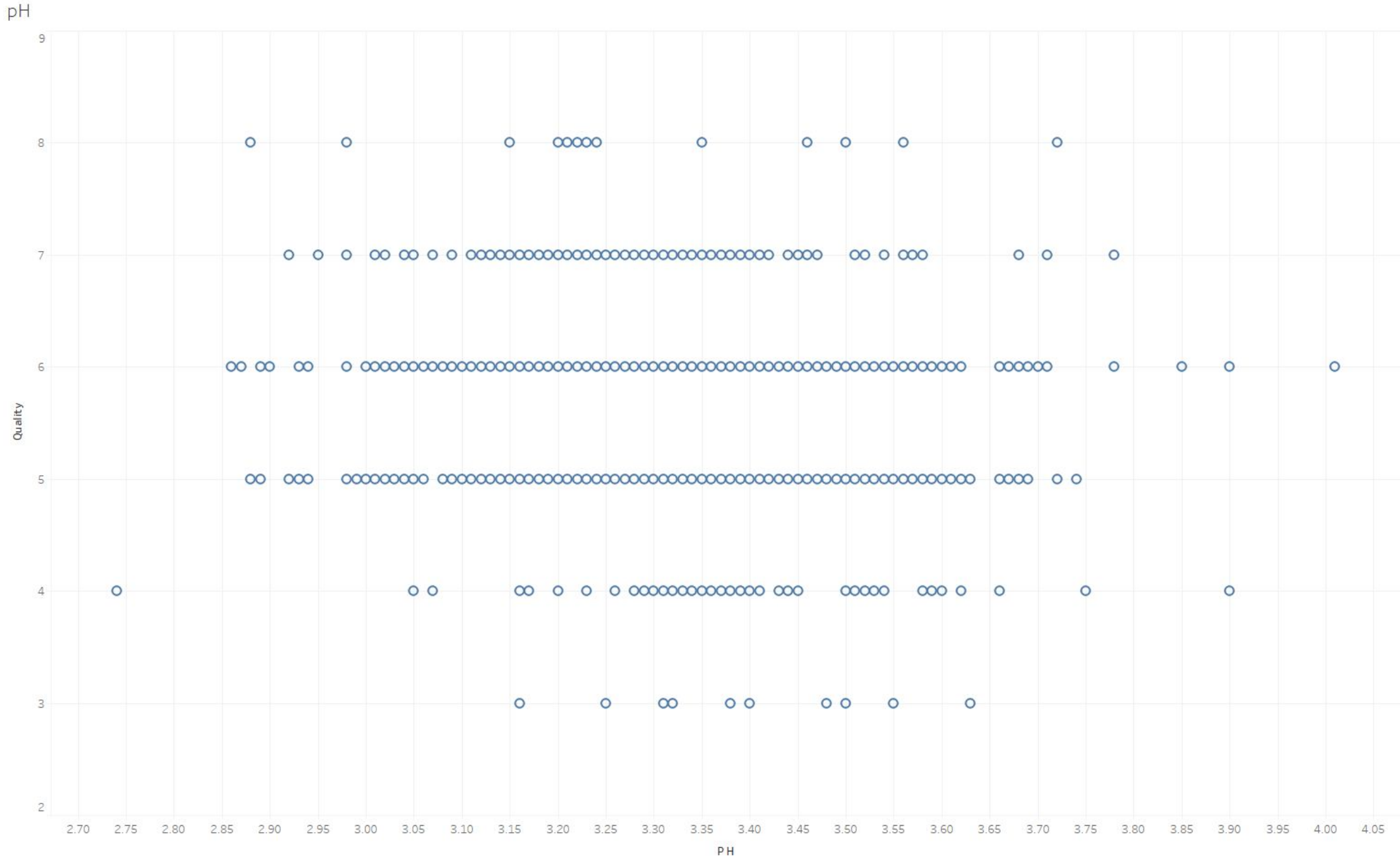


Fixed Acidity vs. Quality.

Free Sulphur Dioxide

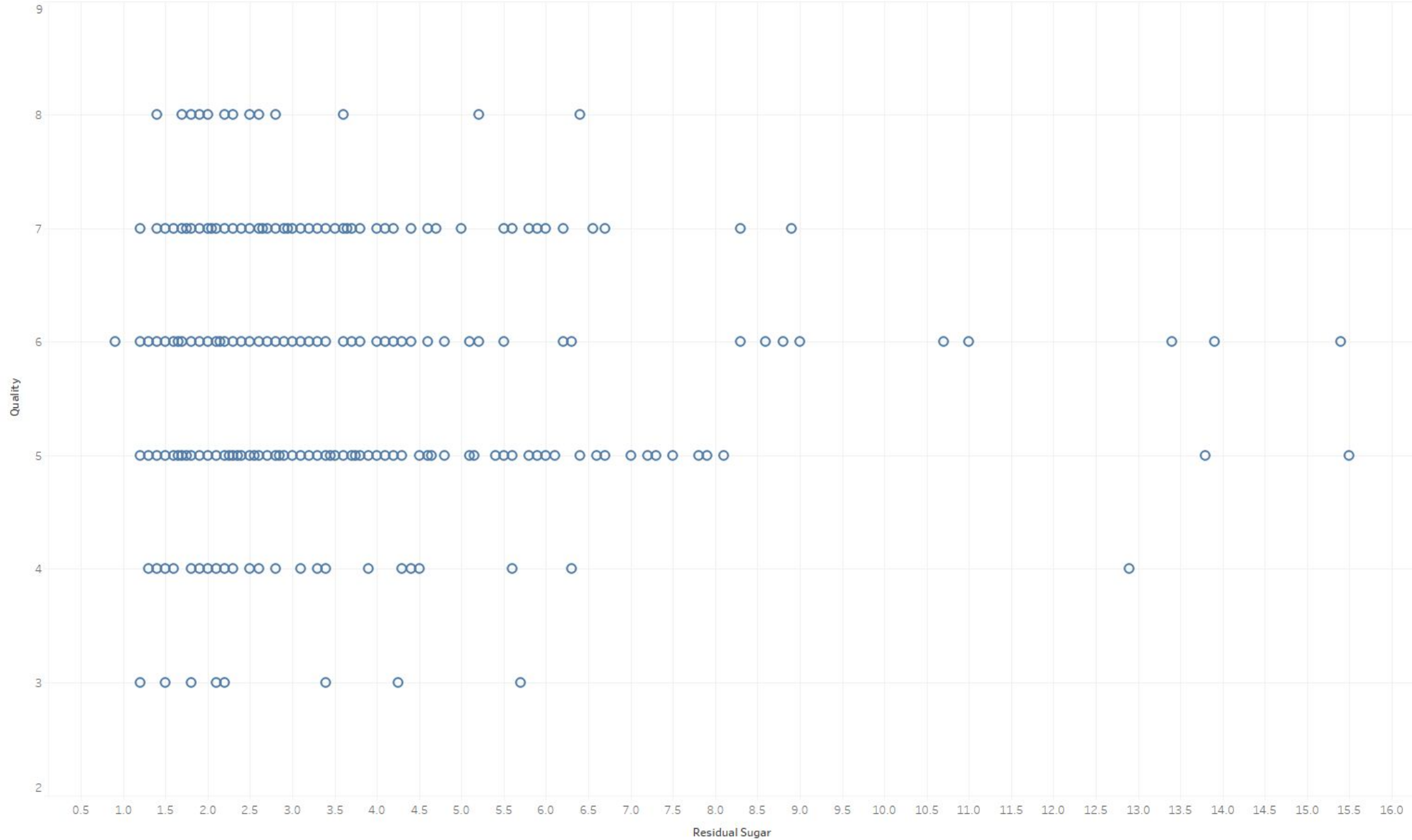


Free Sulfur Dioxide vs. Quality.



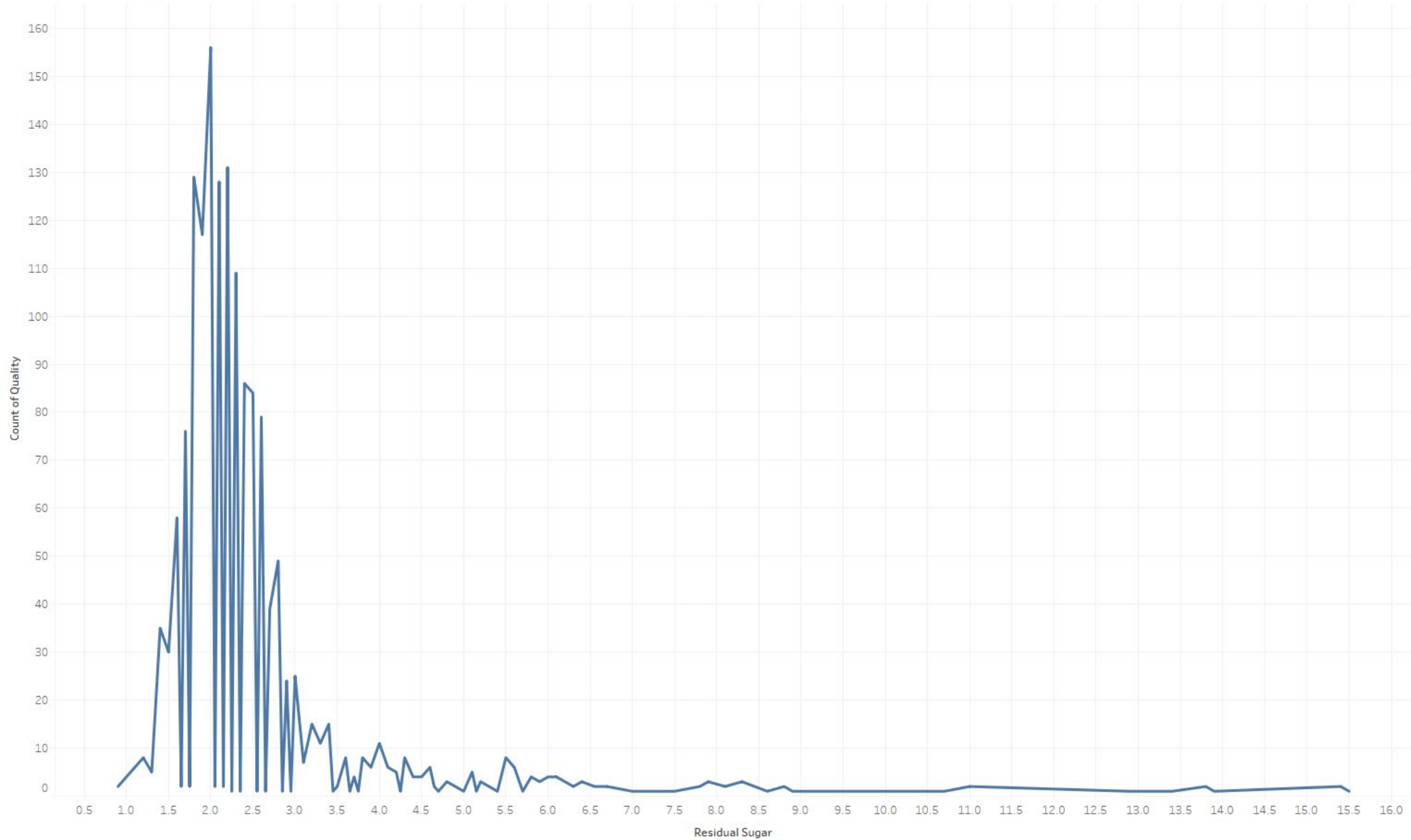
P H vs. Quality.

Residual Sugar



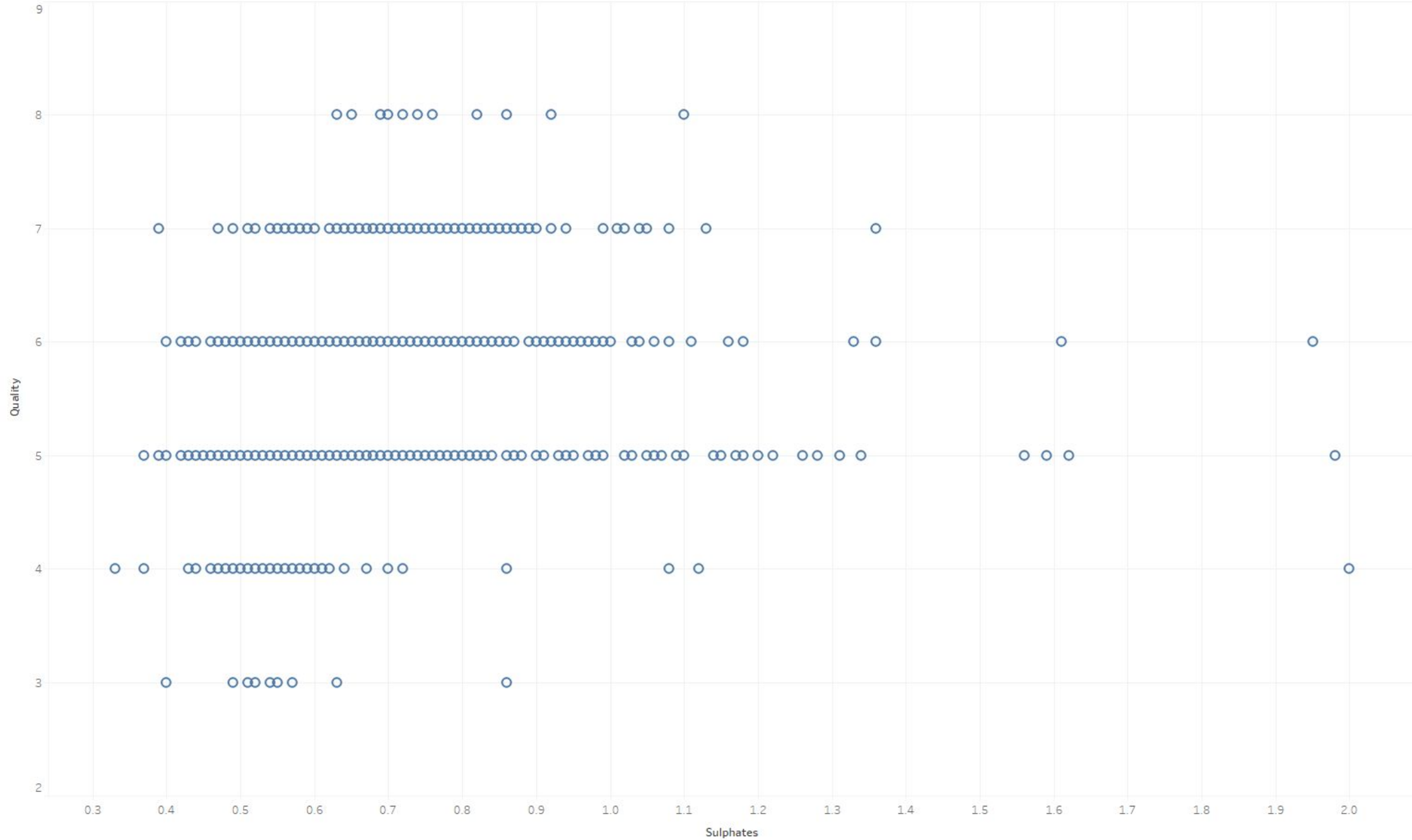
Residual Sugar vs. Quality.

Residual Sugar (2)



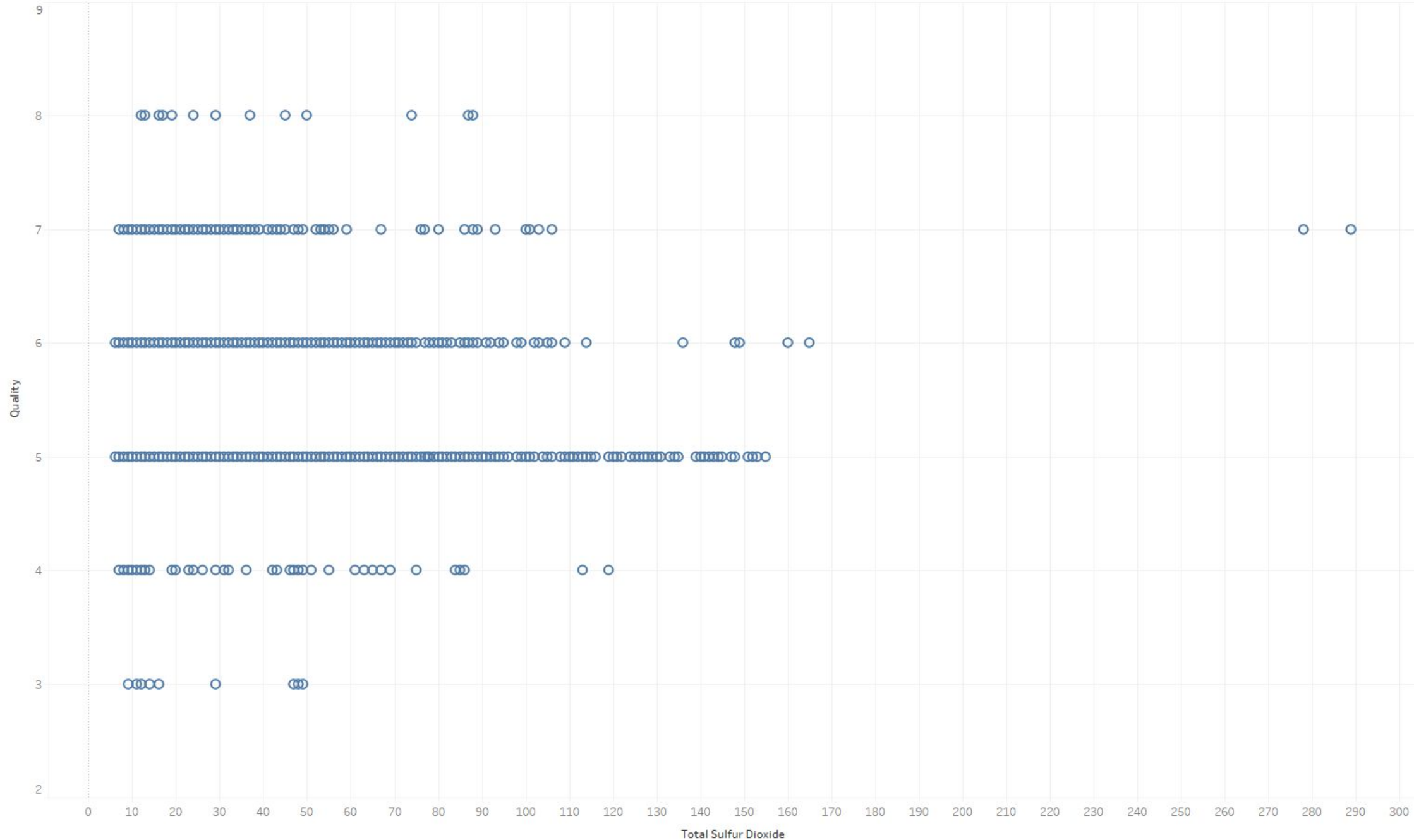
The trend of count of Quality for Residual Sugar.

Sulphates



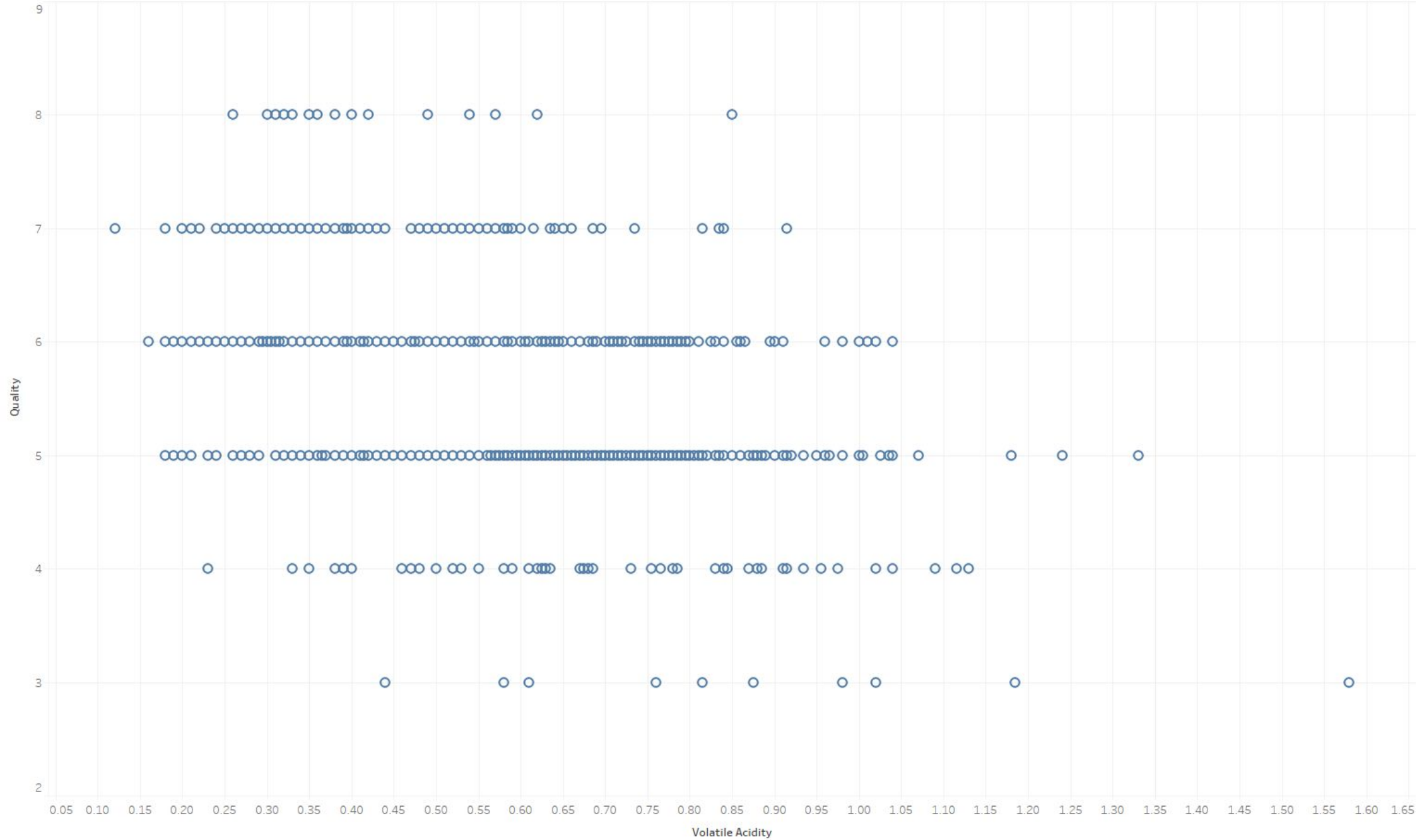
Sulphates vs. Quality.

Total Sulfur Dioxide



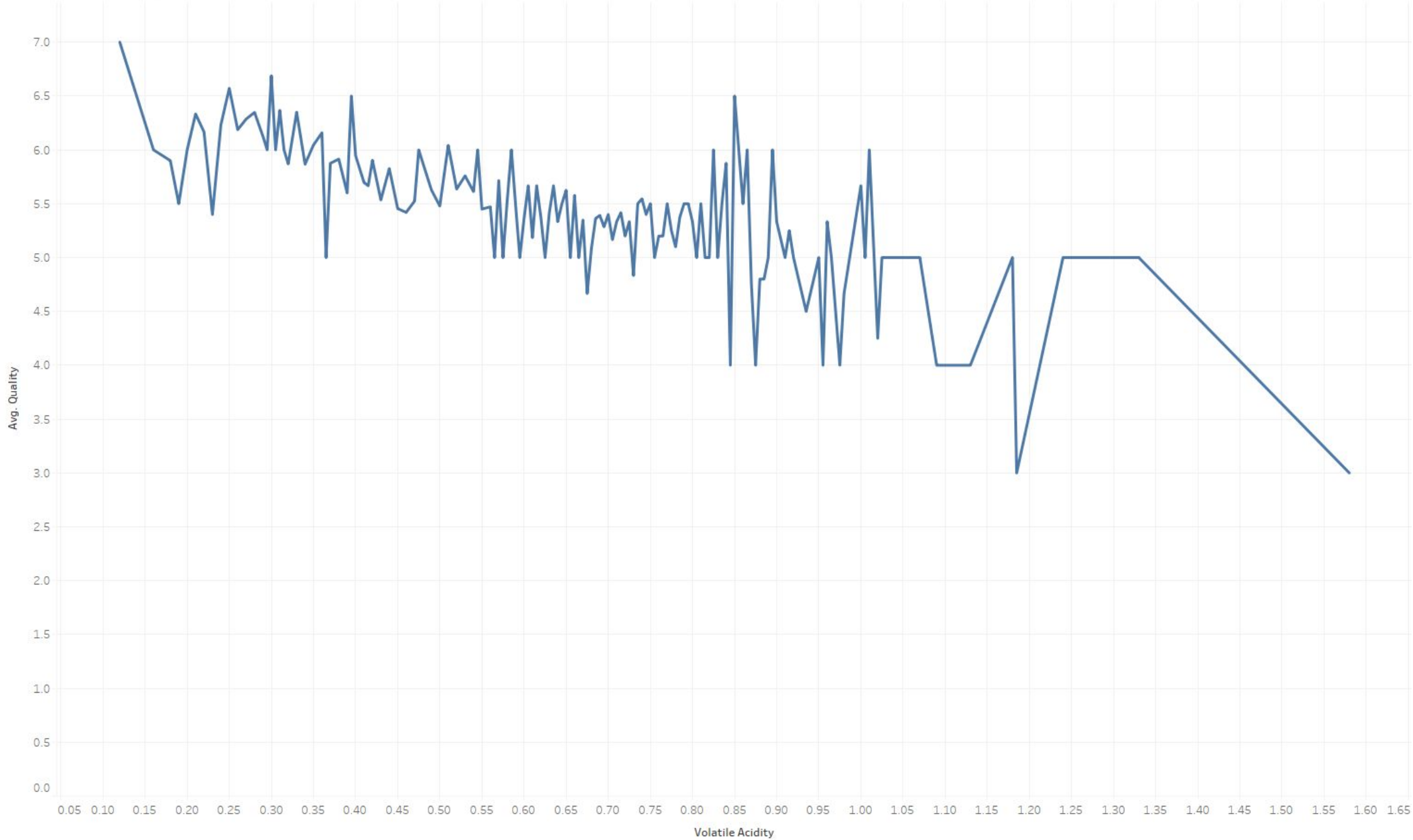
Total Sulfur Dioxide vs. Quality.

Volatile Acidity



Volatile Acidity vs. Quality.

Volatile Acidity (2)



The trend of average of Quality for Volatile Acidity.