

*the in-depth analysis of wine data from Kaggle*

---

# Wines, opinions and prices.

---

by the Drunken Master Team

Danny  
Farai  
Shang  
Stephen

---

---

# Agenda

---

- ❖ Our Dataset
- ❖ How we Prepared
- ❖ What we found
- ❖ Conclusion
- ❖ Limitation and difficulties
- ❖ Q&A

# Our Dataset

```
1 df=pd.read_csv('winemag-data-130k-v2.csv',index_col=0)
2 df.head()
```

	country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety	winery
0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia

❖ 130 K records from a CSV file from Kaggle

## ❖ Key Variables

- Location: Country, Provinces
- sub regions: county-levels
- Price of wines
- wine taster reviewing scores
- wine description
- tasters' twitter account
- winery names

---

# What questions are we exploring

---

- A. What's the top influential factors on wine prices? (Including Region? Weather? Sentiment? Reviewing scores?)
- B. What's the correlation between review point and wine tasters' twitter account popularity / account sentiment?
- D. Other relationship

# How do we prepare

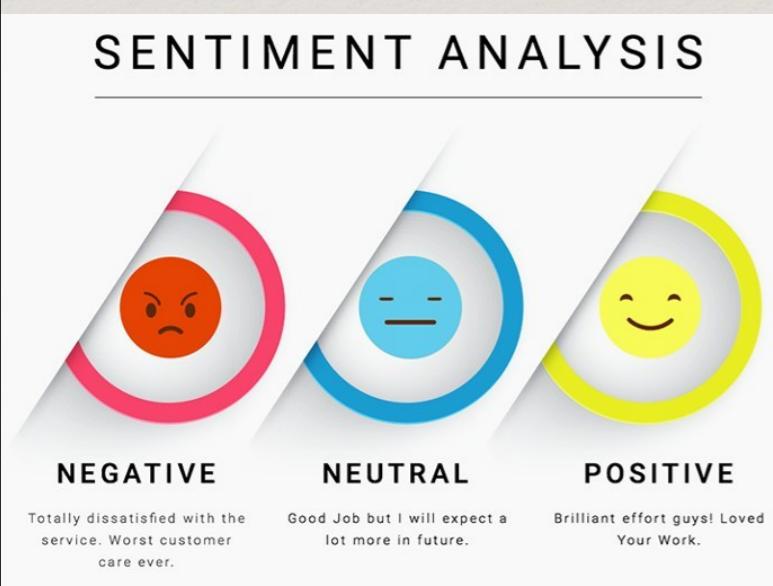
Twitter API



Open Weather API



Sentiment Analysis Lib



Wine Data

Geo Location Libraries

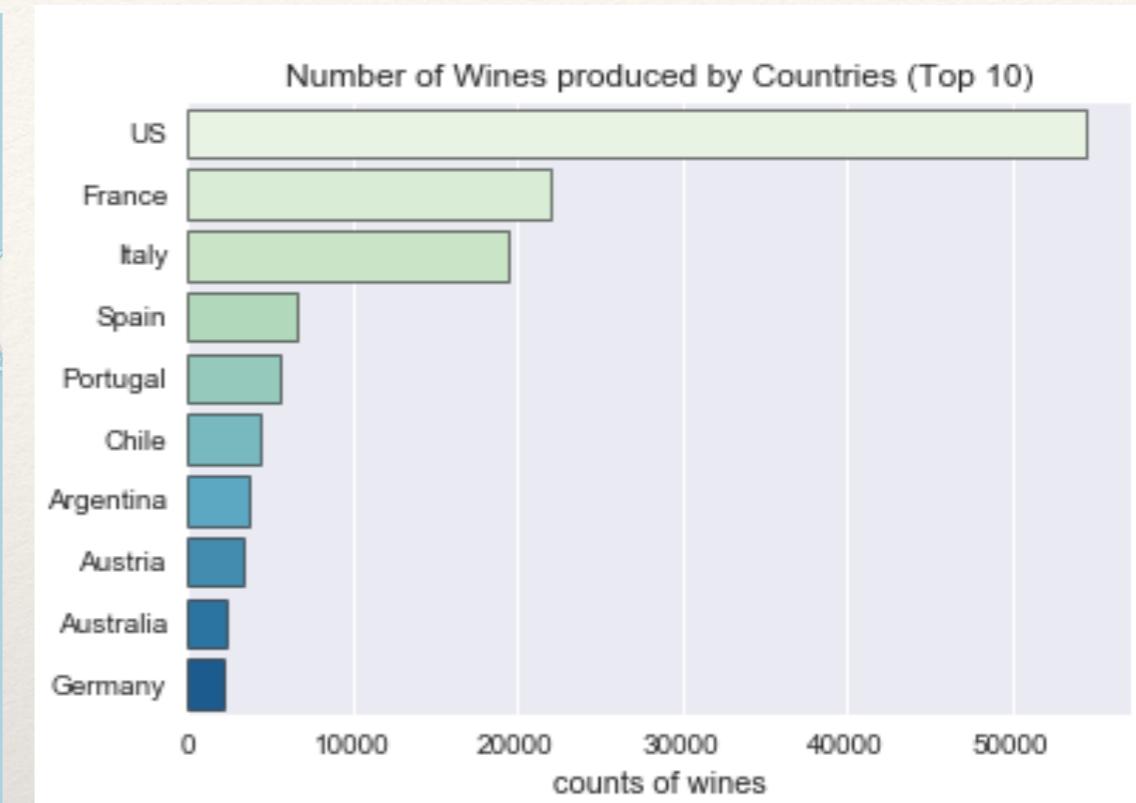
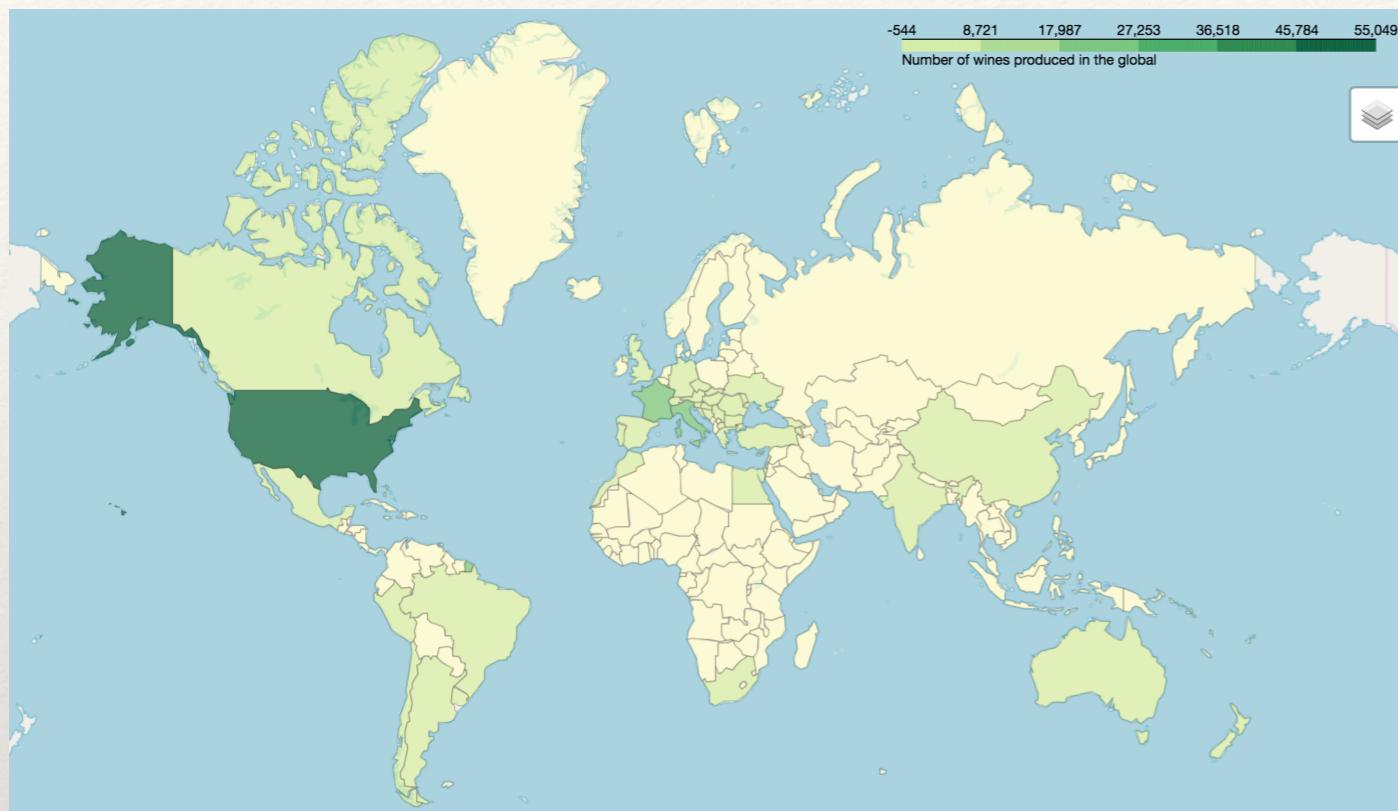


Folium

**What do we find out?**

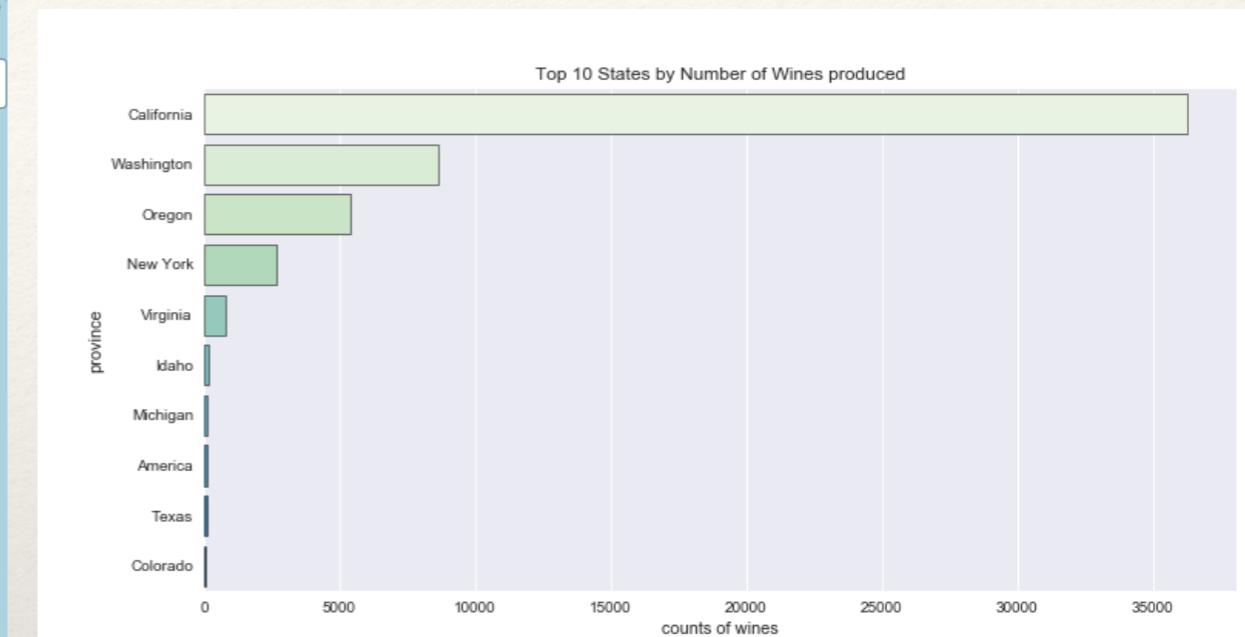
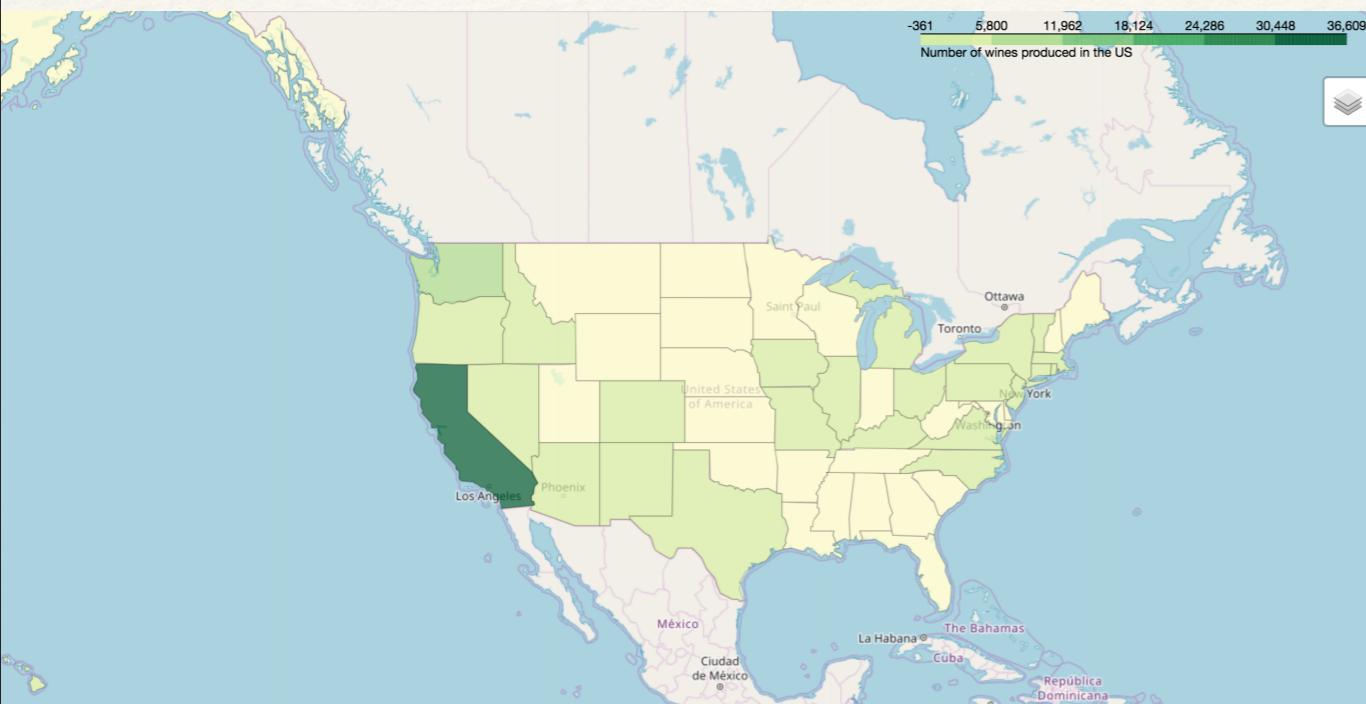
# Demographic Relationship

- ❖ Which country produces the most number of wines?



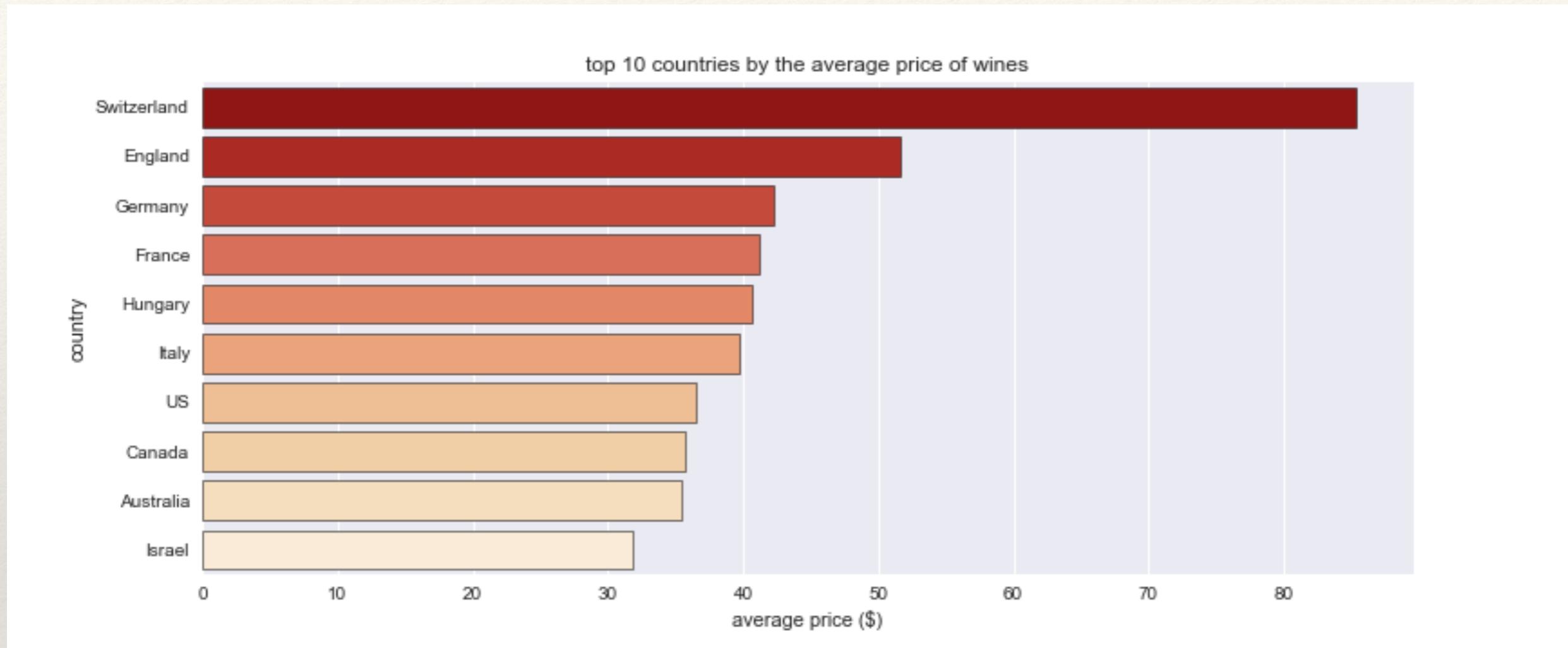
United States, 43.7%

- ❖ Which state is the winner in producing wines in US?



California, 67%

❖ Which country produces the most expensive wines?



Switzerland, \$ 85.3

**Does weather affects wine  
prices?**

climate, weather and soil are the major contributors to wine production after a couple of Google searches.

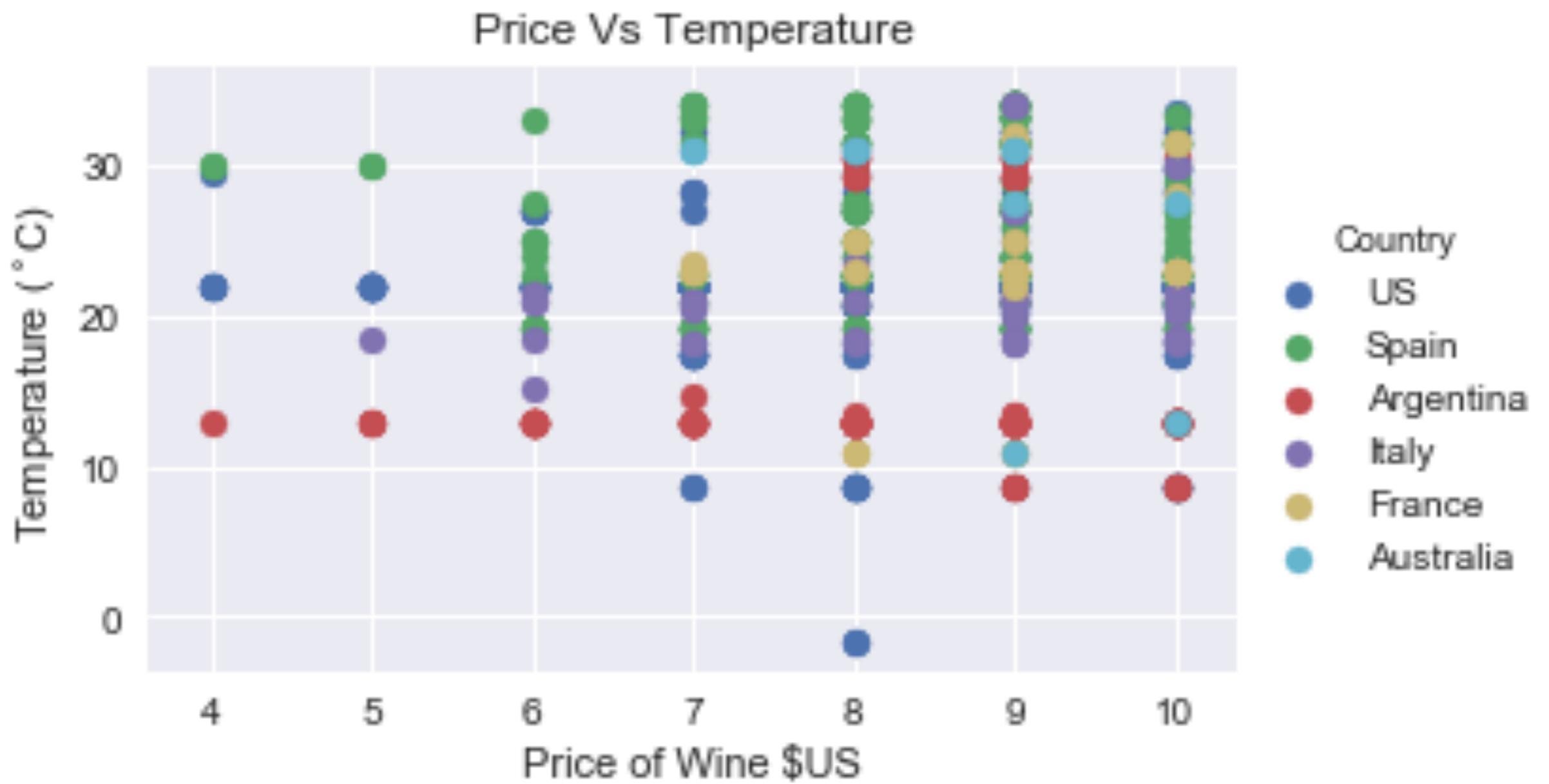
---

# With available sources:

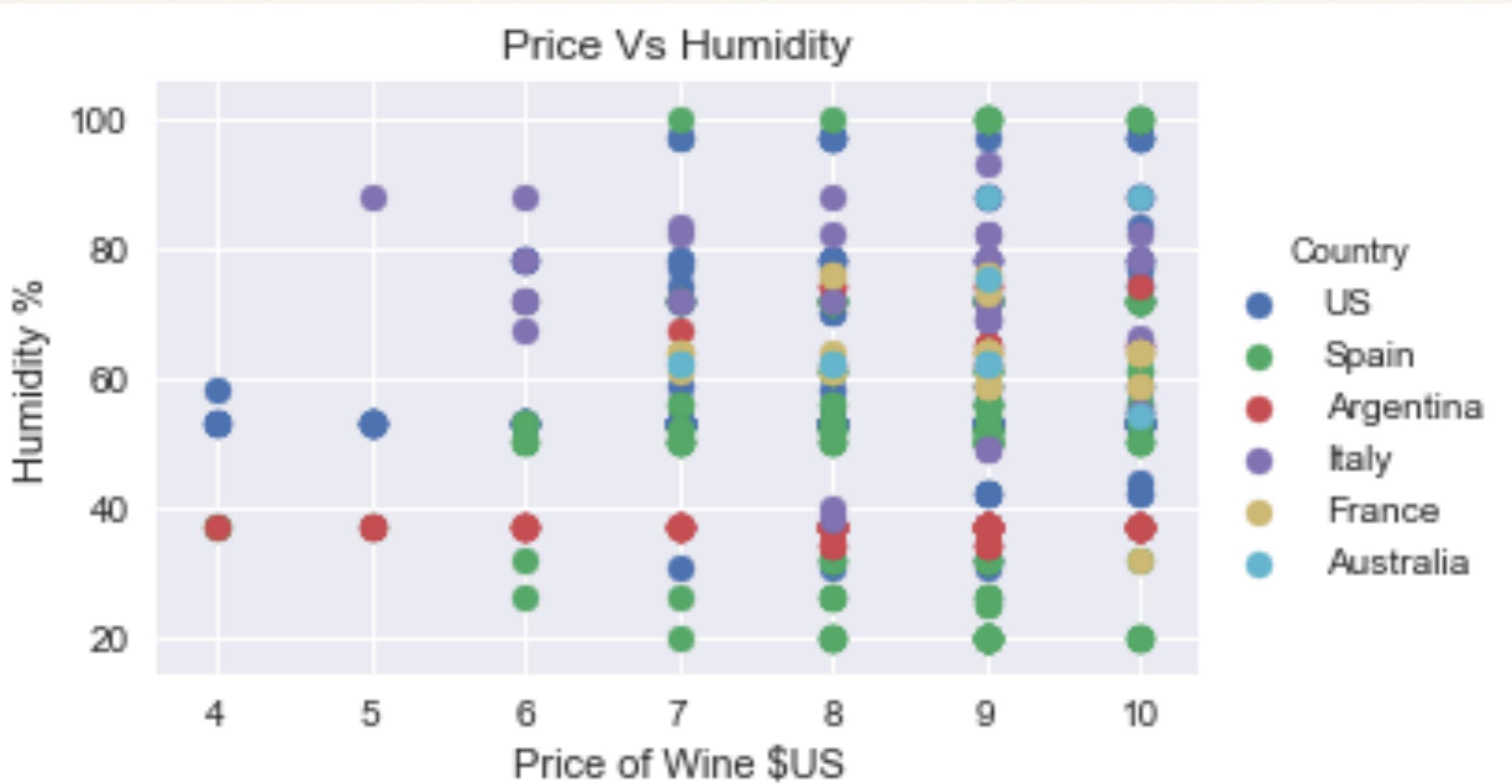
---

- ❖ Wine data from Kaggle: Price, City
- ❖ OpenWeather API: City, Temperature, Humidity
- ❖ The question we could answer is the effects of Temperature and Humidity on the price of wine.

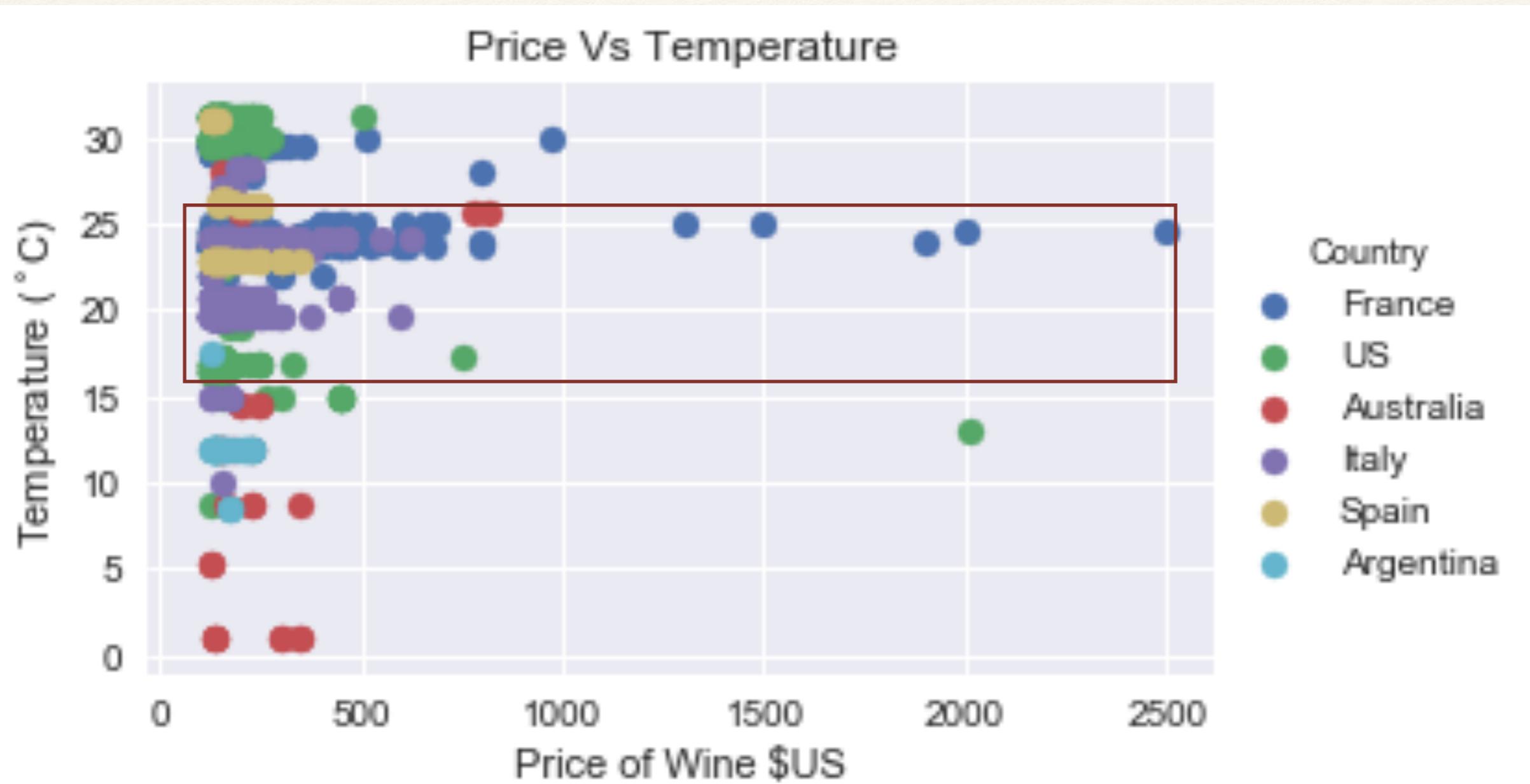
# Temperature VS. Price of the least expensive wines



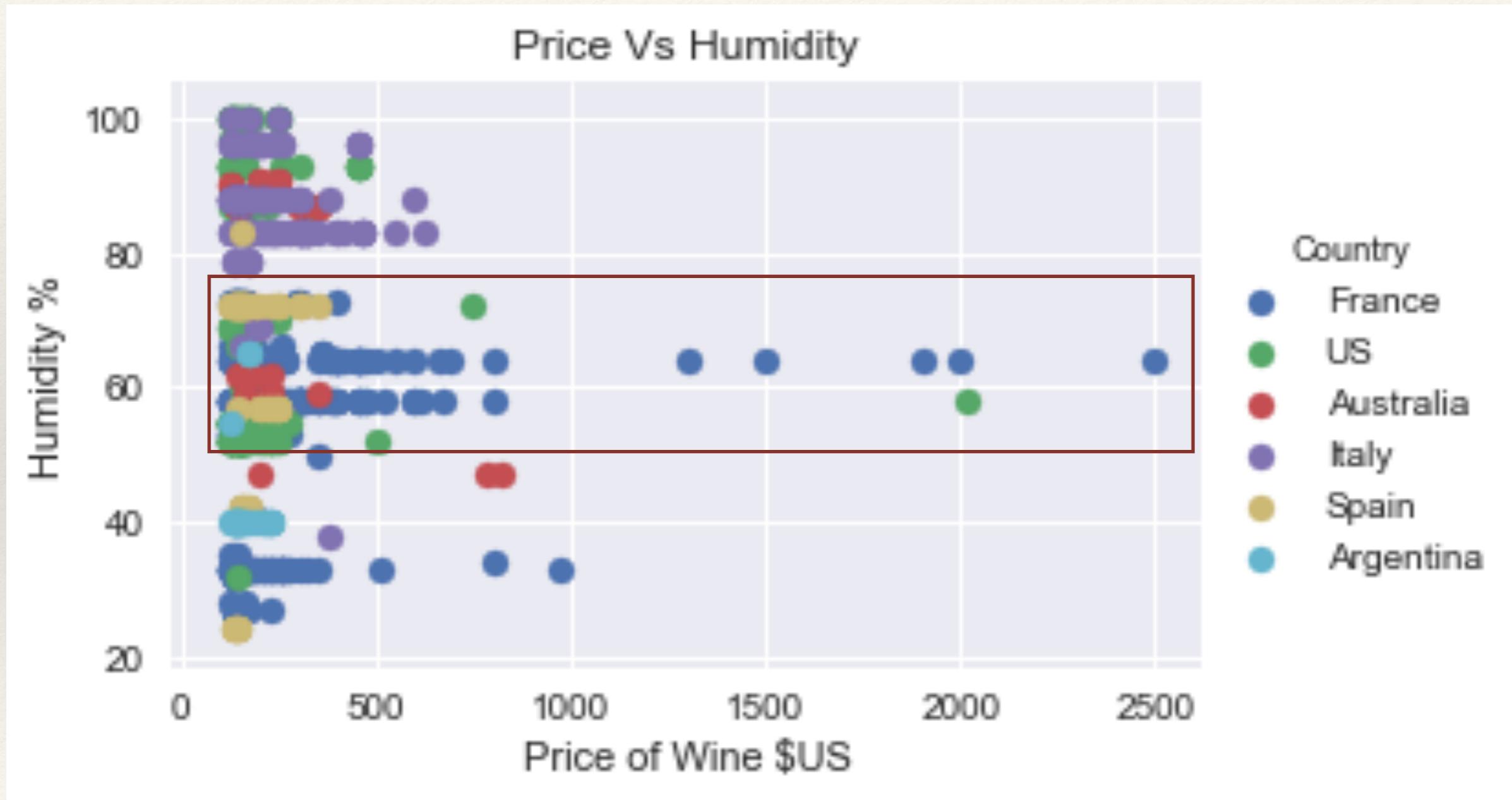
# Humidity VS. Price of least expensive wines



# Temperature vs Price of most expensive wines



# Humidity vs Price most expensive wines



Convergence seems to be key here. Let's narrow down to province to try and find if this is consistent.

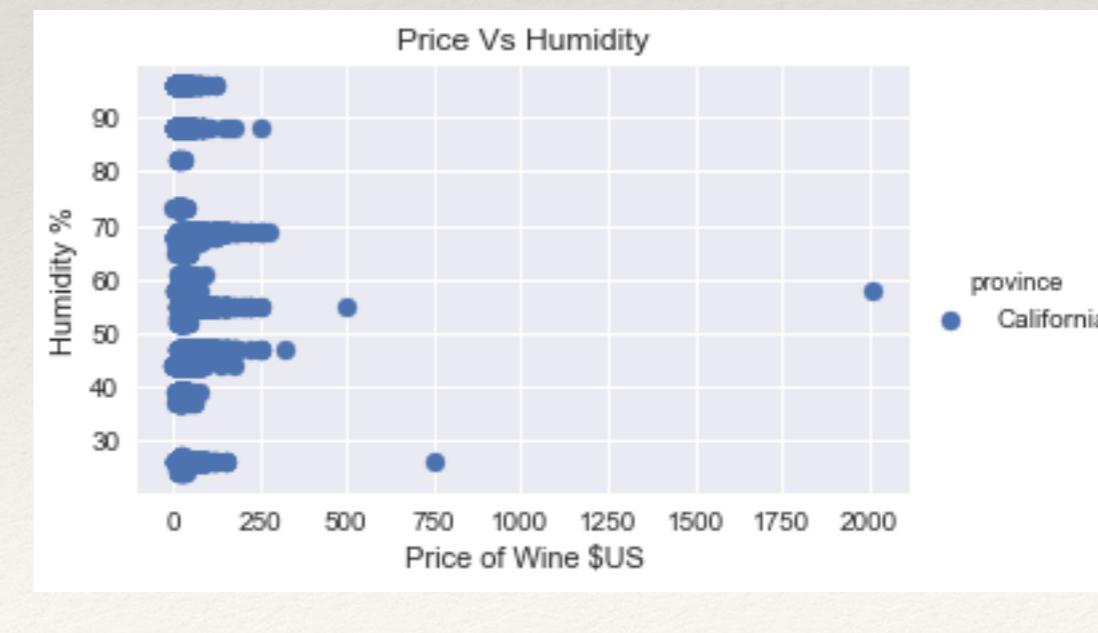
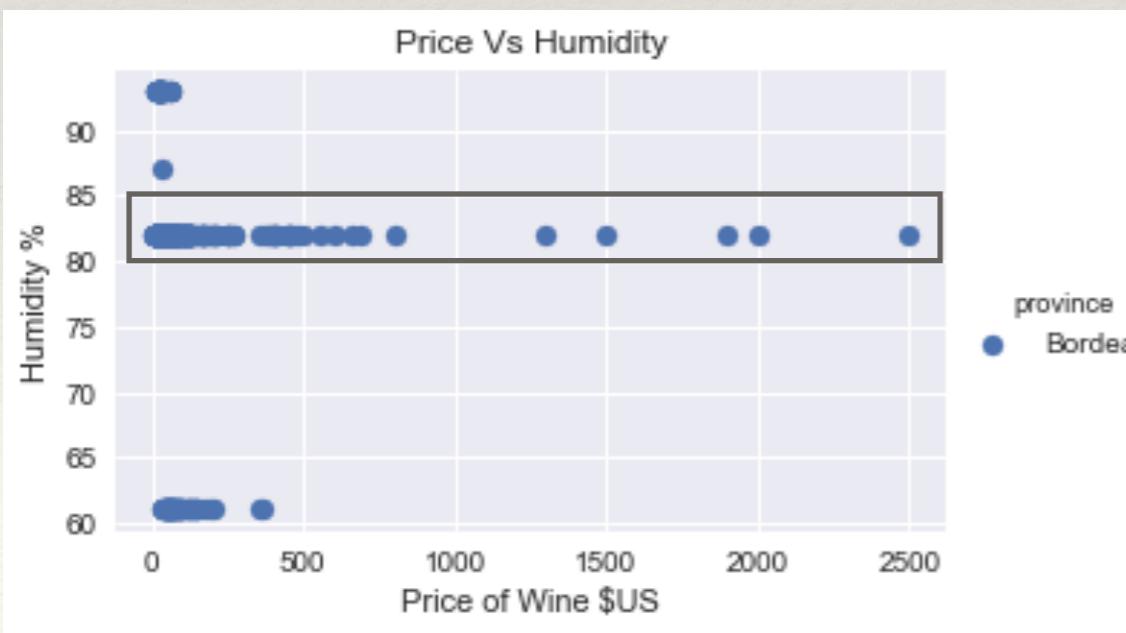
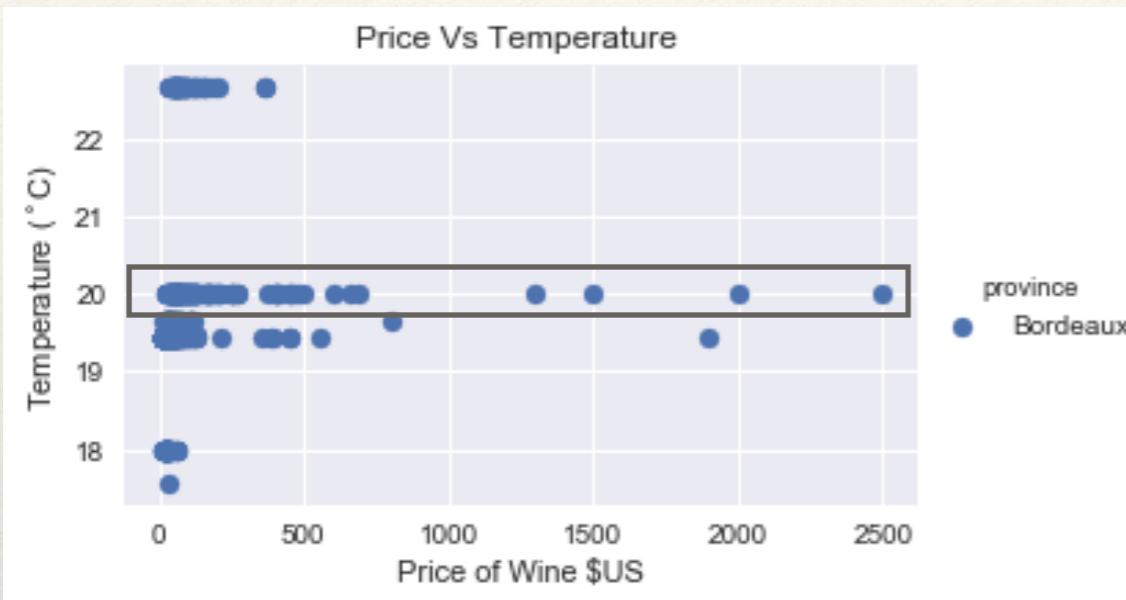
# Top 20 most expensive wines

Country	Points	Price	Province	City	variety	Winery
France	88	3300	Bordeaux	Médoc	Bordeaux-style Red Blend	Château les Ormes Sorbet
France	96	2500	Bordeaux	Pomerol	Bordeaux-style Red Blend	Château Pâtrus
France	96	2500	Burgundy	La Romanée	Pinot Noir	Domaine du Comte Liger-Belair
US	91	2013	California	Arroyo Seco	Chardonnay	Blair
France	97	2000	Bordeaux	Pomerol	Bordeaux-style Red Blend	Château Pâtrus
France	96	2000	Burgundy	La Romanée	Pinot Noir	Domaine du Comte Liger-Belair
France	98	1900	Bordeaux	Margaux	Bordeaux-style Red Blend	Château Margaux
France	100	1500	Bordeaux	Pauillac	Bordeaux-style Red Blend	Château Lafite Rothschild
France	100	1500	Bordeaux	Saint-Émilion	Bordeaux-style Red Blend	Château Cheval Blanc
France	96	1300	Bordeaux	Pauillac	Bordeaux-style Red Blend	Château Mouton Rothschild
France	96	1200	Bordeaux	Pessac-Léognan	Bordeaux-style Red Blend	Château Haut-Brion
France	94	1125	Burgundy	La Romanée	Pinot Noir	Domaine du Comte Liger-Belair
France	97	1100	Bordeaux	Pessac-Léognan	Bordeaux-style Red Blend	Château La Mission Haut-Brion
Austria	94	1100	Wachau		Grüner Veltliner	Emmerich Knoll
Portugal	97	1000	Port		Port	W. & J. Graham's
France	94	1000	Bordeaux	Pessac-Léognan	Bordeaux-style White Blend	Château La Mission Haut-Brion
Portugal	94	980	Port		Port	Kopke
France	95	973	Burgundy	Musigny	Pinot Noir	Domaine Jacques Prieur
France	97	932	Bordeaux	Pessac-Léognan	Bordeaux-style White Blend	Château Haut-Brion
Italy	94	900	Tuscany	Brunello di Montalcino	Sangiovese Grosso	Biondi Santi

# Least Expensive Wines:

country	designation	points	price	province	City	variety	winery
Spain	Flirty Bird	85	4	Central Spain	Vino de la Tierra de Castilla	Syrah	Felix Solis
US		85	4	California	Clarksburg	White Blend	Dancing Coyote
Argentina	Red	84	4	Mendoza Province	Mendoza	Malbec-Syrah	Broke Ass
US		84	4	California	California	Chardonnay	Bandit
US	Unoaked	83	4	California	California	Chardonnay	Pam's Cuties
Spain		84	4	Levante	Yecla	Cabernet Sauvignon	Terrenal
US		86	4	California	California	Merlot	Bandit
US		84	4	California	California	Merlot	Bandit
Spain	Estate Bottled	84	4	Levante	Yecla	Tempranillo	Terrenal
Spain	Flirty Bird	82	4	Central Spain	Vino de la Tierra de Castilla	White Blend	Felix Solis
Romania	UnWineD	86	4	Viile Timisului		Pinot Grigio	Cramele Recas
Portugal	Morgado da Canita	84	5	Alentejano		Portuguese Red	Herdade dos Machados
Italy	Mini	86	5	Veneto	Prosecco	Glera	Anna Spinato
Portugal	Escolha	83	5	Vinho Verde		RosÃ©	Adega Cooperativa Ponte de Barca
France		83	5	France Other	Vin de France	Cabernet Sauvignon	Belle Made For You
France		83	5	France Other	Vin de France	Red Blend	Belle Made For You
Chile	Gato Negro	83	5	Central Valley		Cabernet Sauvignon	San Pedro
US	Organic Grapes	85	5	California	California	Chardonnay	Earth's Harvest
US		84	5	California	California	Red Blend	Mancan
US	Twin Valley	85	5	California	California	Pinot Grigio	Gallo Family Vineyards

# California and Bordeaux Analysis

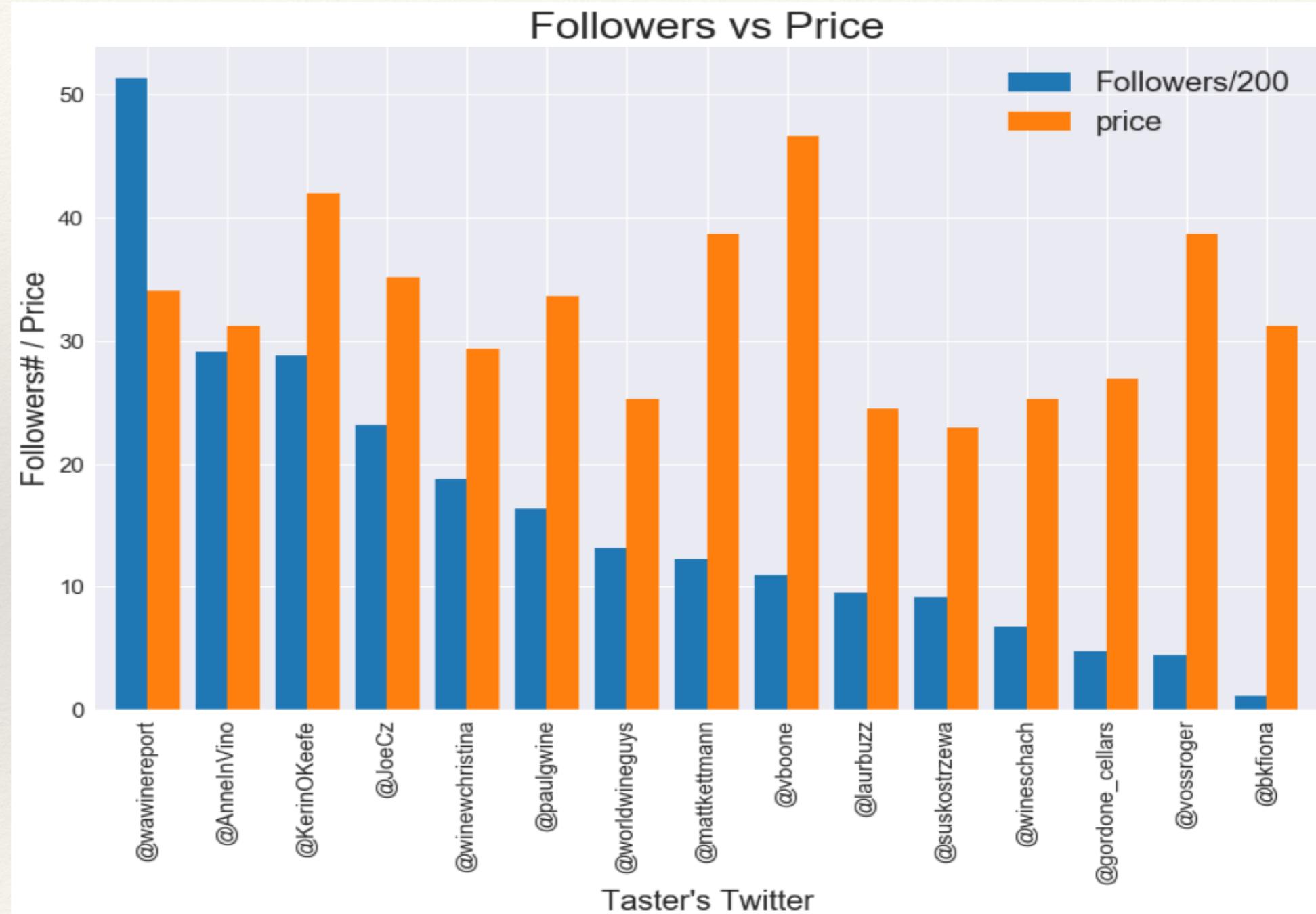


The graphs seem to show a consistency  
of weather affecting the price of wine!!

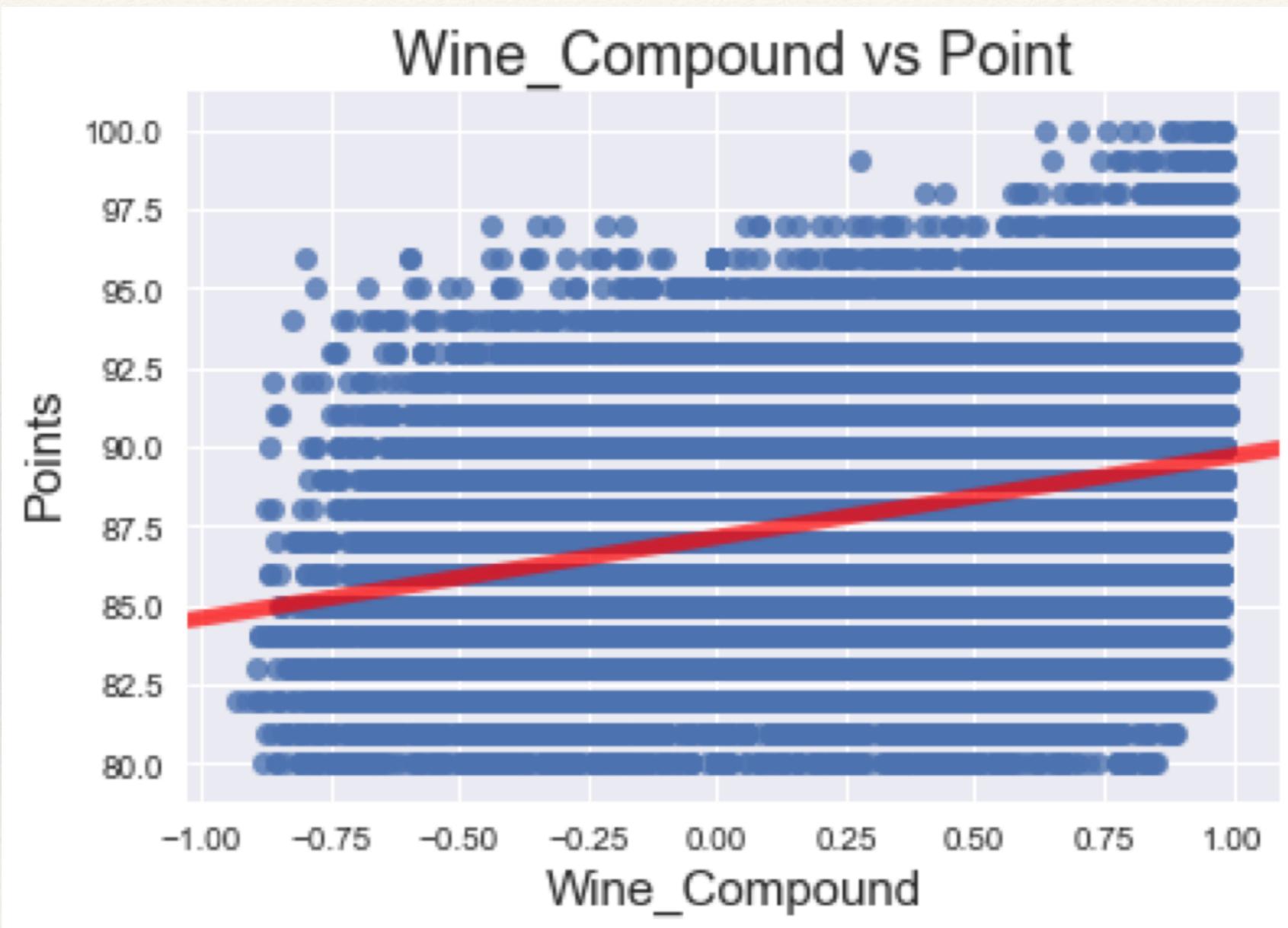
# Sentiment Analysis

# Finding 1: Wines reviewed by popular tasters don't necessarily have Higher mean Higher Price

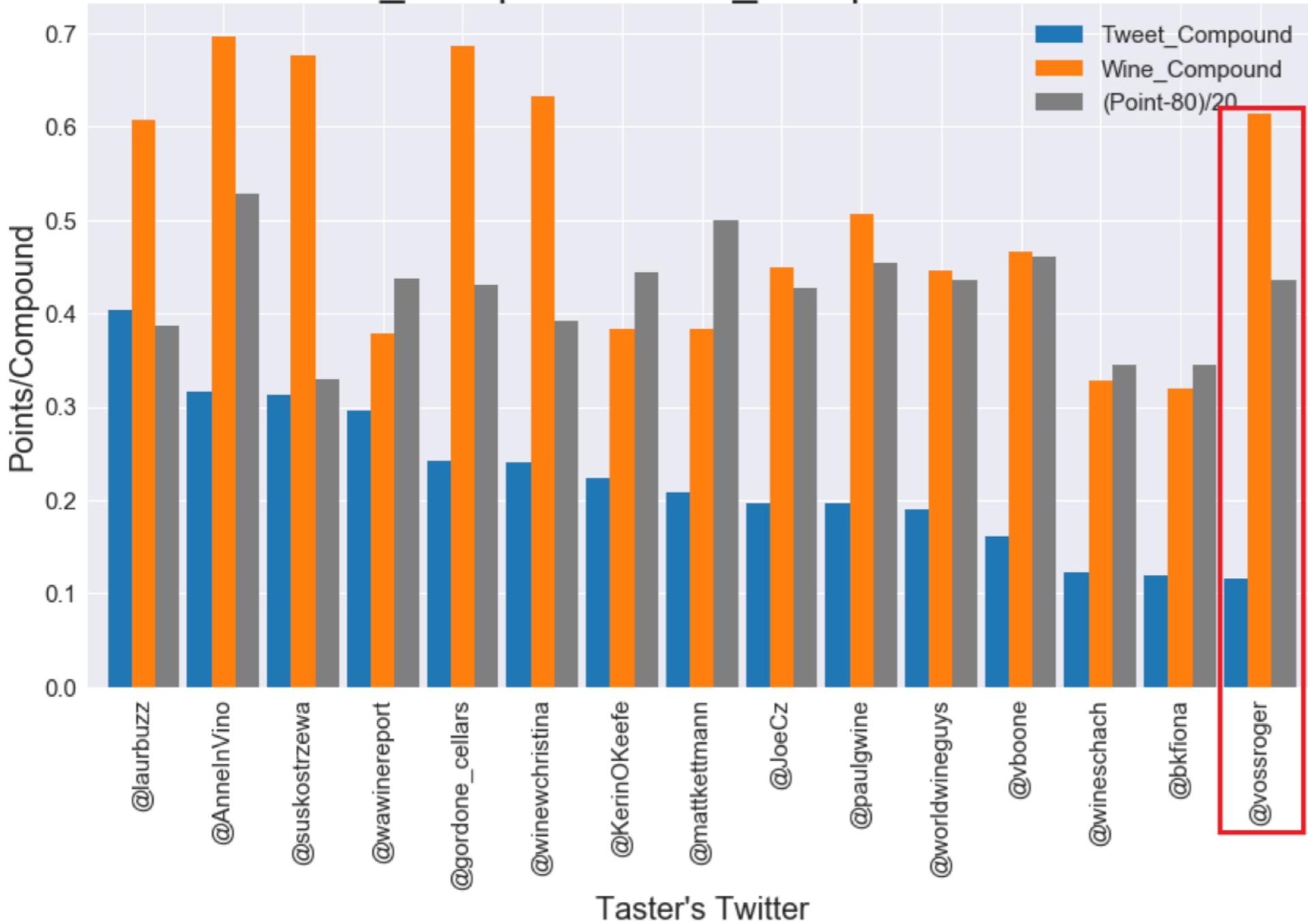
taster_twitter_handle	Followers
@wawinereport	10272
@AnnelnVino	5812
@KerinOKeefe	5763
@JoeCz	4629
@winewchristina	3757
@paulgwine	3256
@worldwineguys	2613
@mattkettmann	2446
@vboone	2175
@laurbuzz	1884
@suskostrzewa	1824
@wineschach	1352
@gordone_cellars	953
@vossroger	880
@bkfiona	218



Finding 2: There is **POSITIVE** correlation between the wine descriptions and review scores.



## Tweet\_Compound/Wine\_Compound vs Point



- All Tasters' **Twitter** have **POSITIVE** sentiment score.
- All Tasters' **Wine Description** sentiment score are even **HIGHER** than tasters' twitter sentiment score.
- For some tasters: Good Wine Descriptions **do NOT always correspond** with Points

# Sentiment Analysis on Tasters' Twitters and Wine Descriptions

## Twitter Posts:



Roger Voss @vossroger · Jul 3

Now there is big hail in Beaune and local flooding. Not good for the vines.

Roger Voss Retweeted

Marc Hilberer @MarcHilberer · May 18

the number of school shootings that have happened since the year 2000 around the world:

ENGLAND: 0

GREECE: 1

NETHERLANDS: 1

SPAIN: 1

INDIA: 1

ARGENTINA: 1

RUSSIA: 1

CHINA: 3

MEXICO: 4

AUSTRALIA: 5

CANADA: 5

GERMANY: 5

SOUTH AFRICA: 5

USA: 213

#Texas



Roger Voss @vossroger · Jan 10

Just tasted a \$110 village #Meursault. Pleasant wine, not worth the price. It's about time @BourgogneWines stopped raising prices because of small production. Drinkers can't afford them.



Roger Voss

@vossroger

European editor (éditeur européen)  
Wine Enthusiast Magazine, Bordeaux,  
Champagne, Burgundy, France, Portugal  
Opinions are my own.

Tweets

636

Following

412

Followers

880

## Wine Descriptions:

- A year in wood and 30 months in bottle before release have allowed this **attractive** wine to fully develop its solid yet **smooth texture**. It has concentration, layers of **bright**
- Somewhat lean, this has a tart, grapefruit character that sits **awkwardly** with the otherwise fresh, green fruit character. Hard-edged.
- Full, **firm** white wine that gives **good** green fruit and pear flavors, lifted with spice from wood. The wine is round, while still retaining **crispness** from grassy acidity, and offers a **good** depth of plum skin structure and flavor.

# Wine Types vs. Reviews

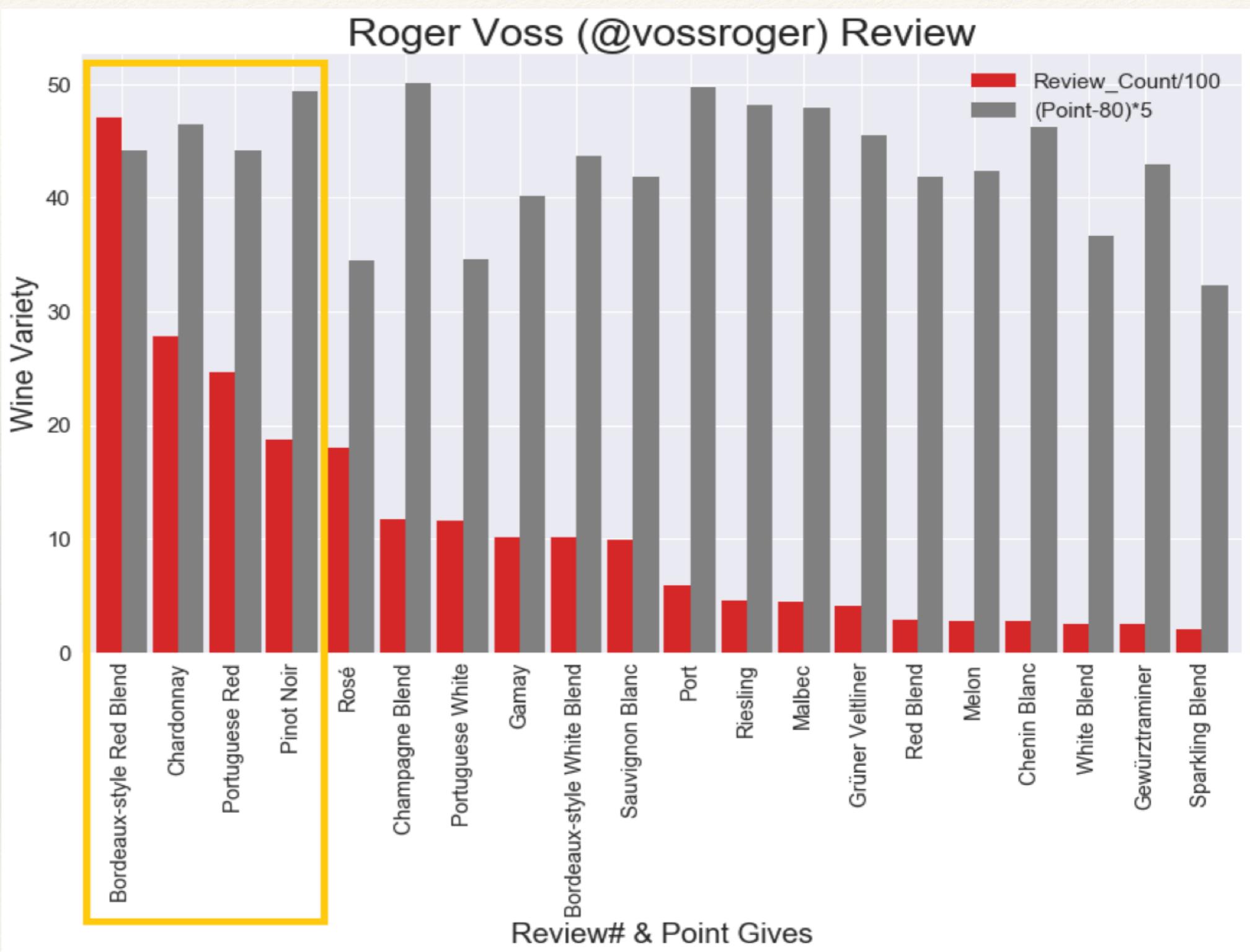


**Roger Voss**

@vossroger

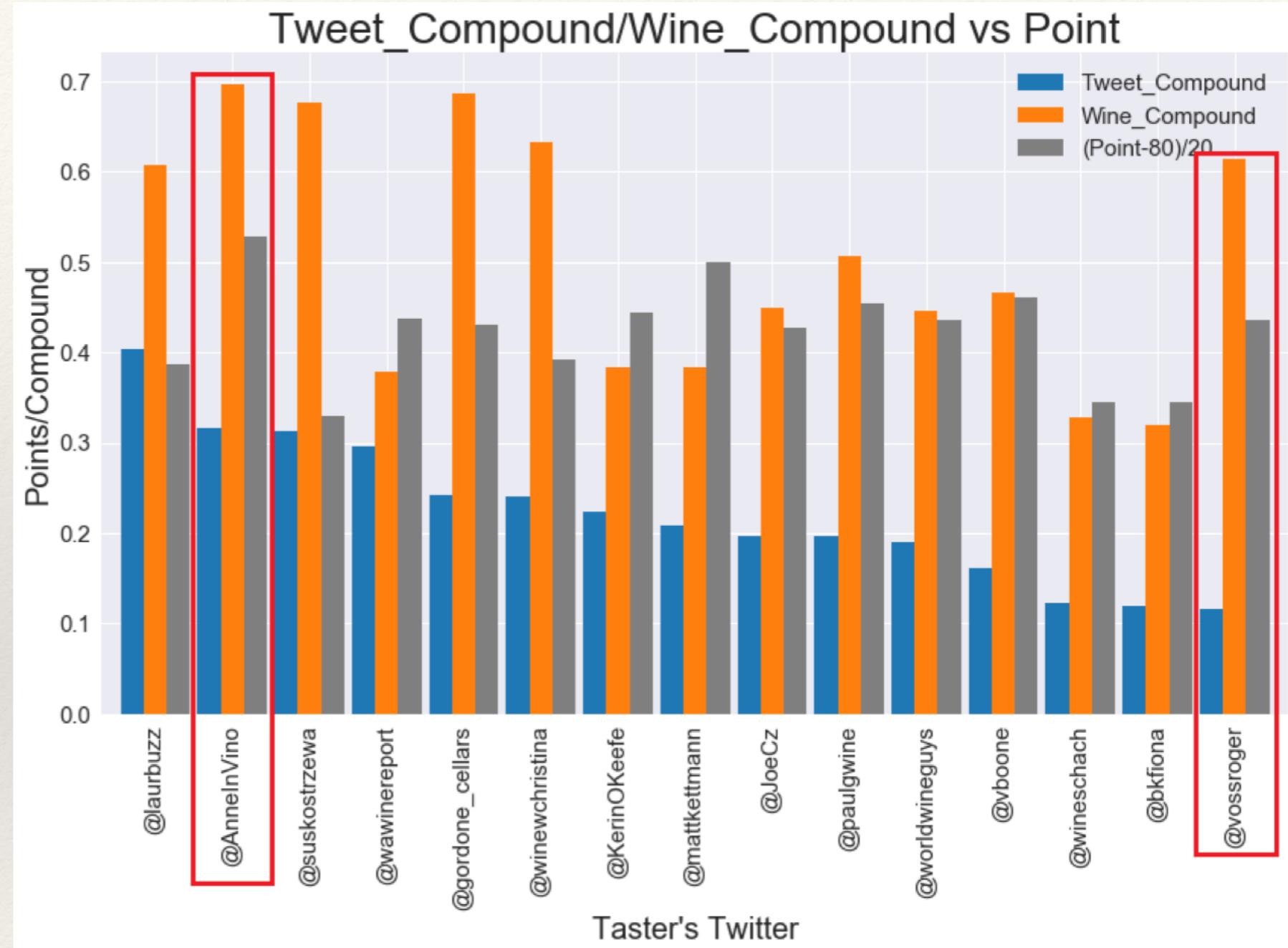
European editor (rééditeur européen)  
Wine Enthusiast Magazine, Bordeaux,  
Champagne, Burgundy, France, Portugal  
Opinions are my own.

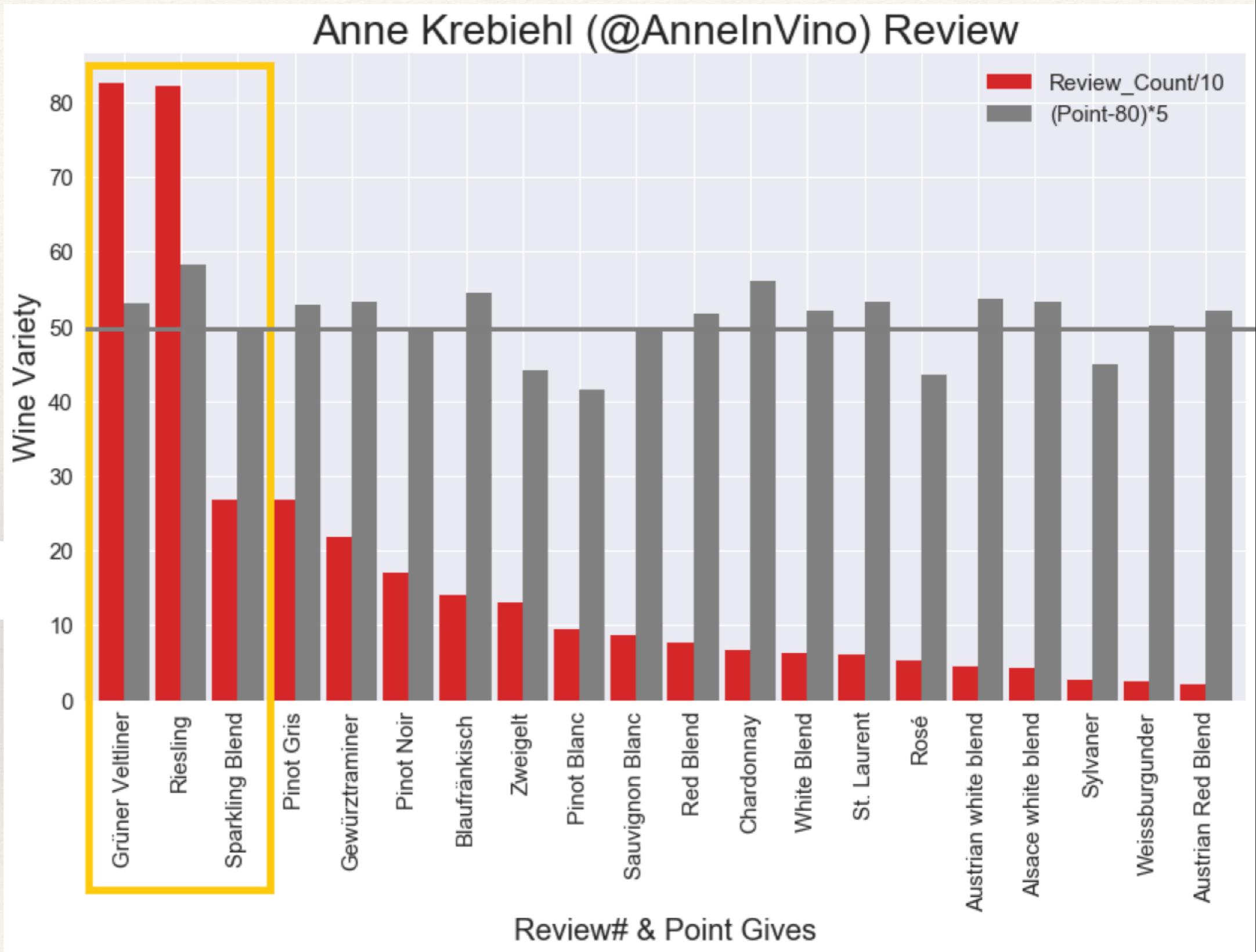
Tweets      Following      Followers  
**636**      412      880



# Who else will be a better taster, if chosen by winery?

- Very Positive Wine Sentiment Score
- High Wine Review Point
- Large Twitter Follower Base



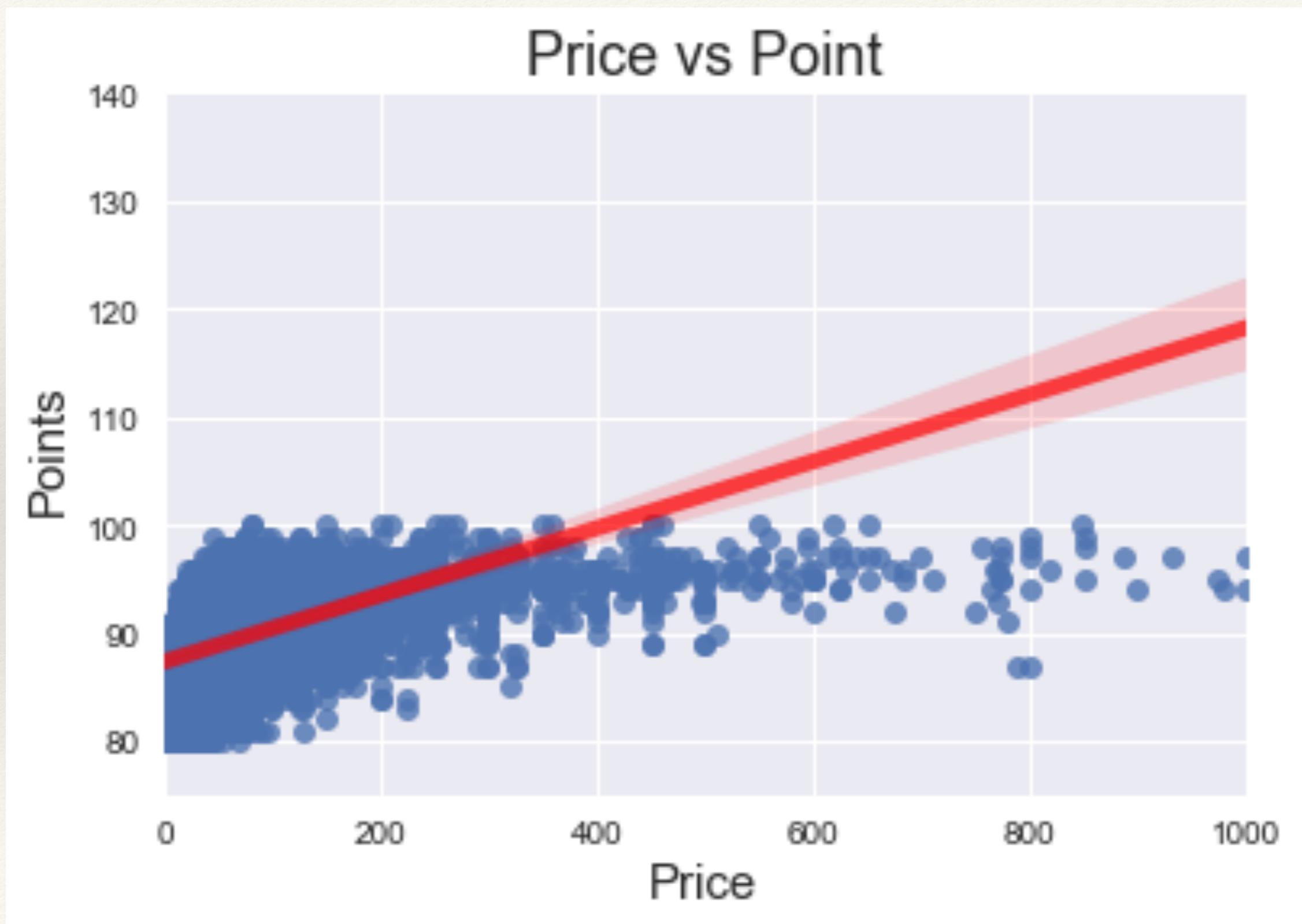


---

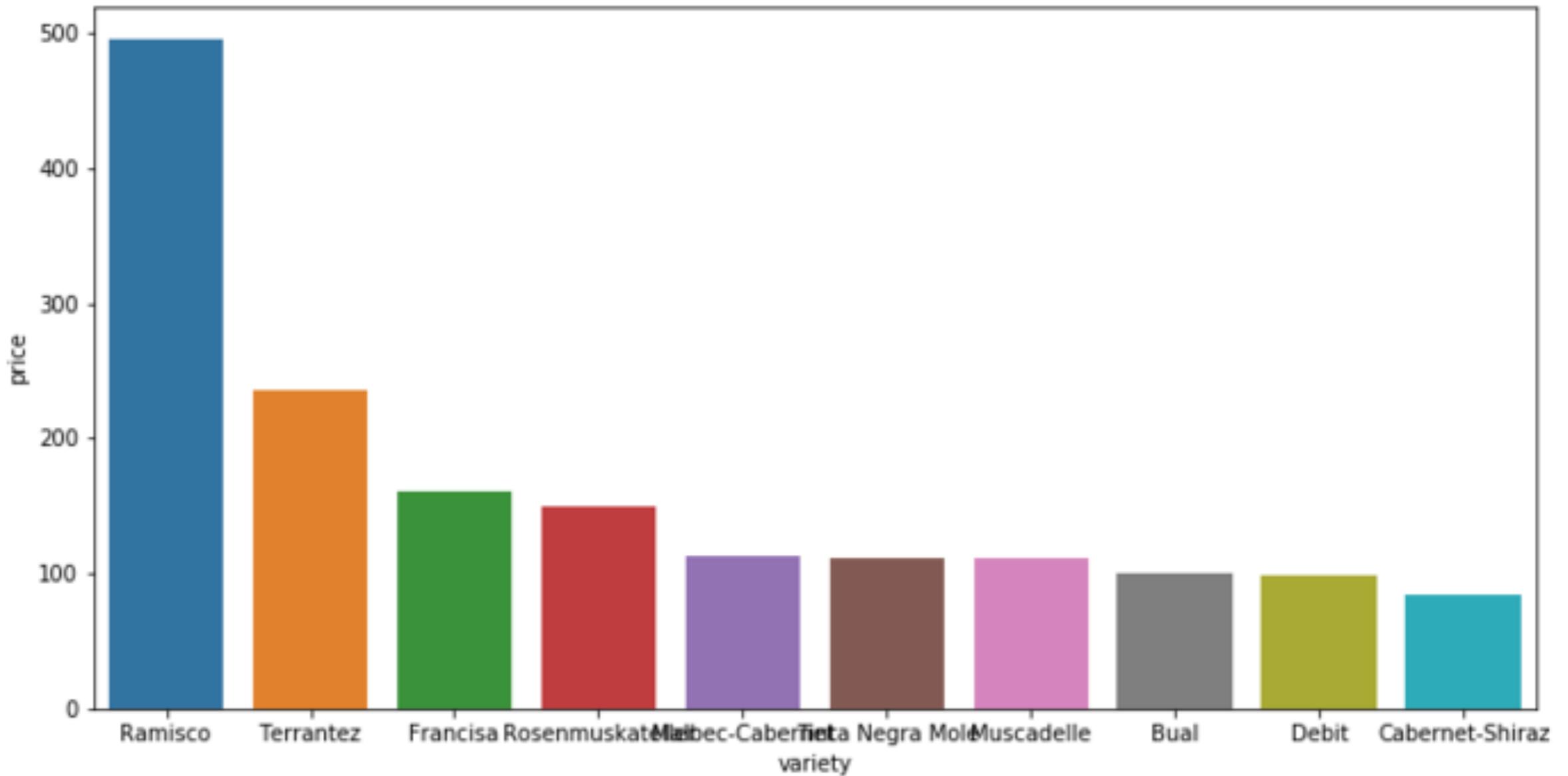
---

# Next Chapter

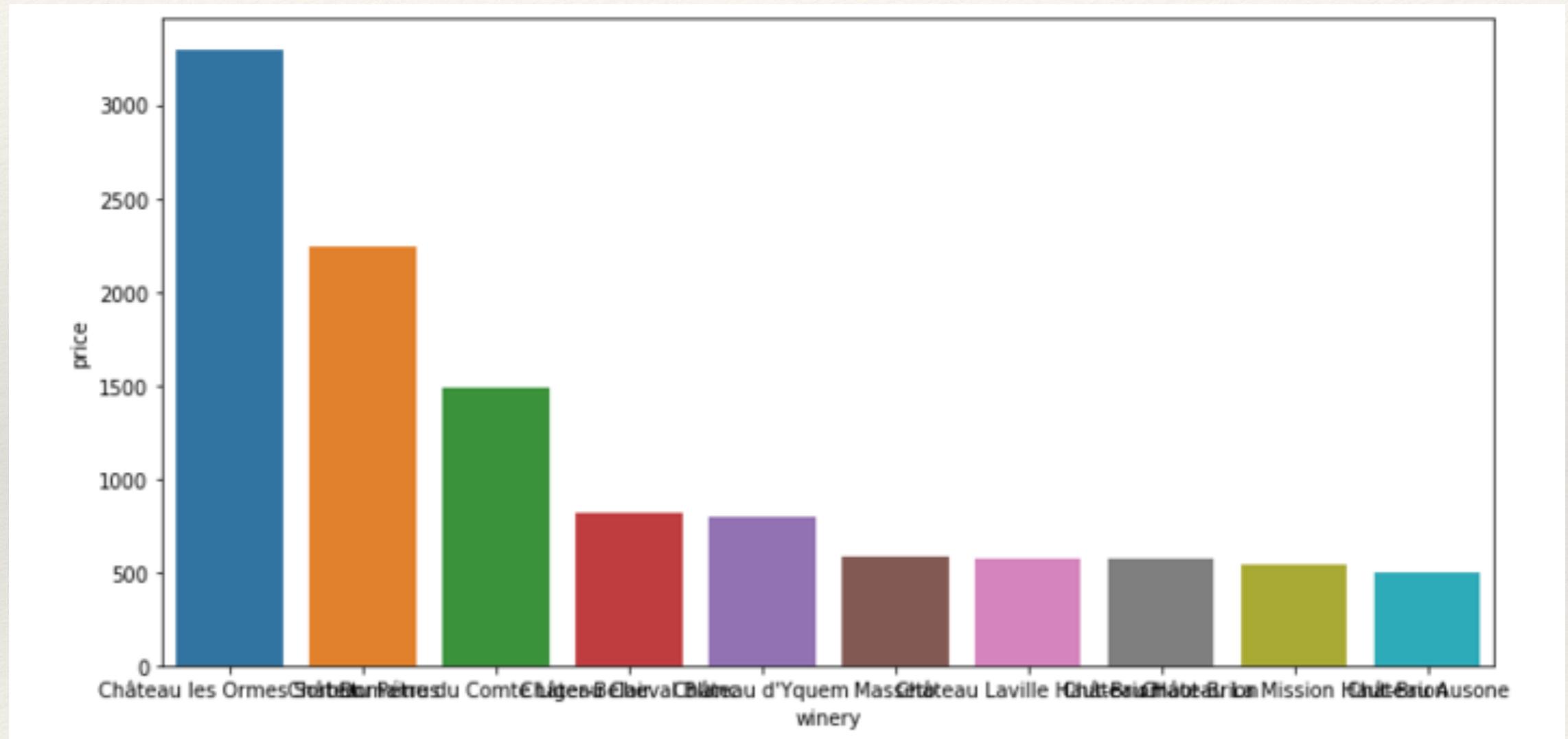
# Review Scores VS Price



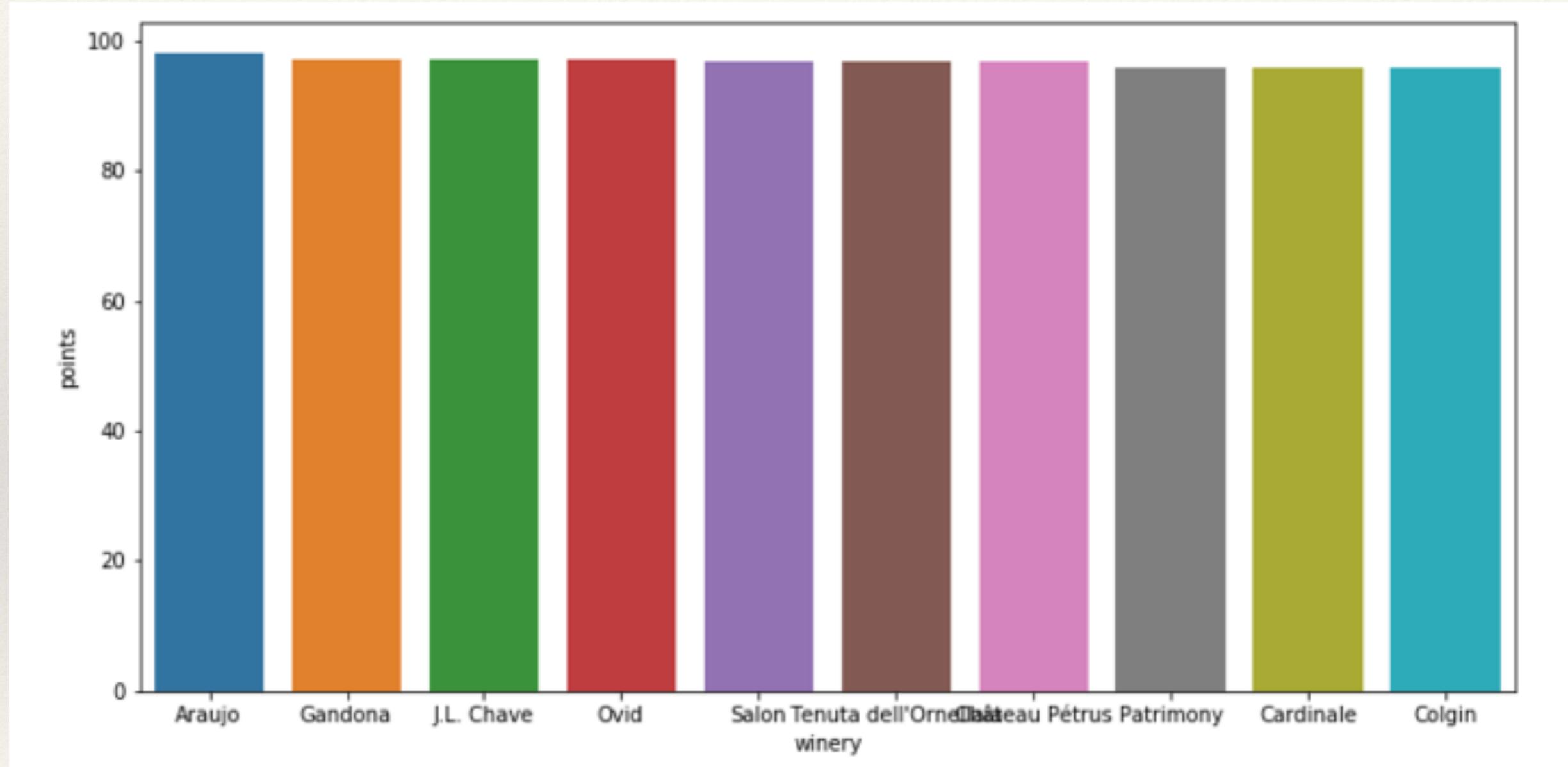
# Wine Type vs Price



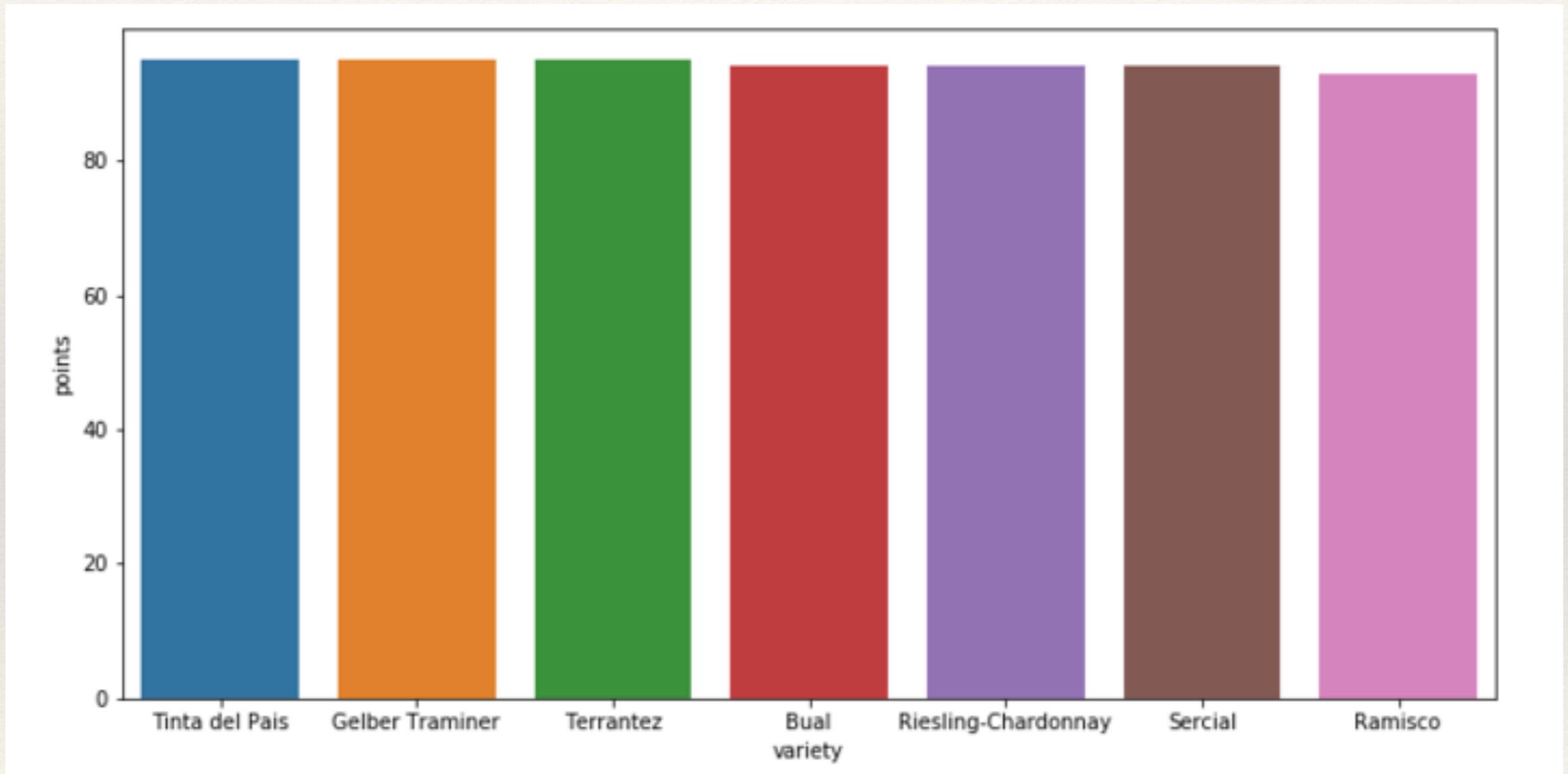
# Wineries vs Price



# Wineries vs Review Scores



# Wine Types vs Review Scores



---

# Conclusions

---

- ❖ Most of wines in our dataset are from United States and so are the least expensive. Most expensive are from France. There is more weather variation in US vineyards than there is in French vineyards.
- ❖ Temperature of 20-30 degrees celsius the is most suitable to produce high-priced wines
- ❖ Certain types of wines are reviewed most by tasters (pinot noir, chardonnay)
- ❖ Positive Correlation between Wine Description and Review Scores
- ❖ Tasters' tend to give more positive Wine Descriptions than personal Twitter posts.

---

# Limits and Difficulties

---

- ❖ We could have used more informative API's to answer other key factors to production of wine but they cost money, lots of money.
- ❖ The weather is based on current weather; we could not acquire historical weather information.
- ❖ The names of some of the cities in the CSV French cities use special characters which are not recognized by the OpenWeather API JSON eg La RomanÃ©e.

# DRUNKEN MASTER

QA

TRAILER