Write-Up/Description of Data

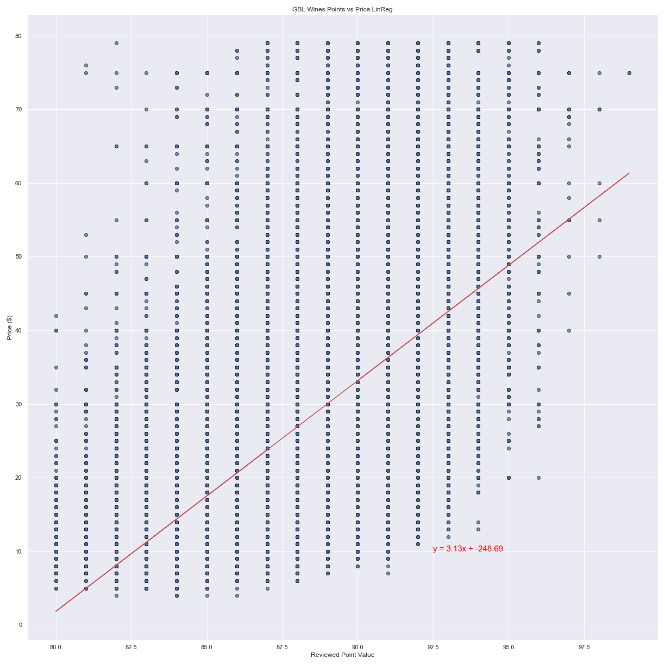
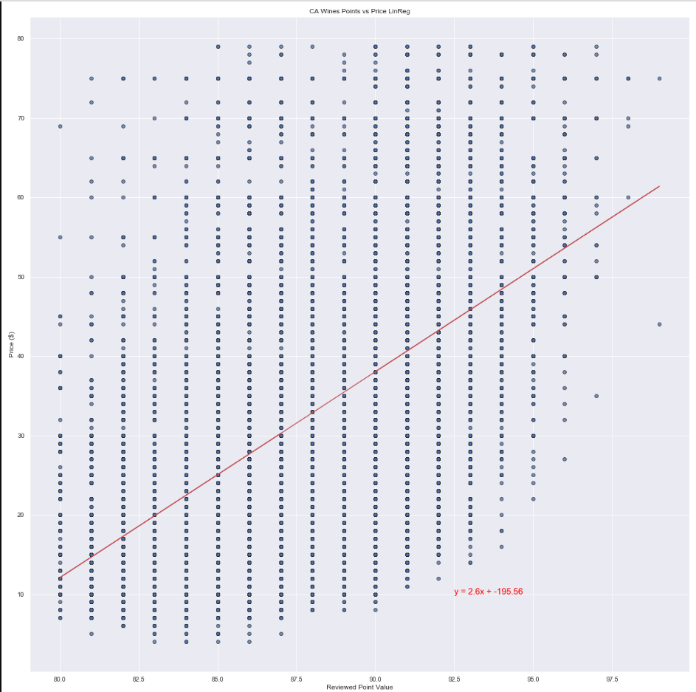
Create a write-up summarizing your major findings. This should include a heading for each “question” you asked of your data and a short description of your findings and any relevant plots.

**California is known for their wines, but do they produce the highest quality (professionally rated) wine compared others?**

Based on the average professional rating (80-100 point scale) of the 33526 reviews of Californian wines and 78011 wines from the rest of the world, Californian wines have a marginally higher average rating of 88.619 points compared to 88.325 points for other wines. From this dataset alone, California produces wines of higher rating than wines elsewhere.

**Does spending more get you better wine?**

From the California wines points vs price scatterplot and global wines (without California wines) points vs price scatterplot, there were moderately positive correlation coefficients of 0.51 and 0.56 respectively. Despite not having a strong correlation, the trend was significant. Generally, the more that is spent on a wine, the greater the quality. Both California wines and wines from the rest of the world had this trend as shown below from the scatterplots below.



**Does California have the best (overall) wineries compared to the rest of the world?**

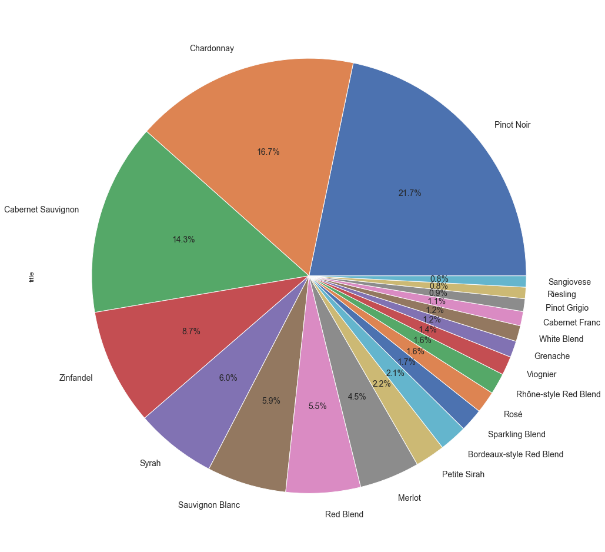
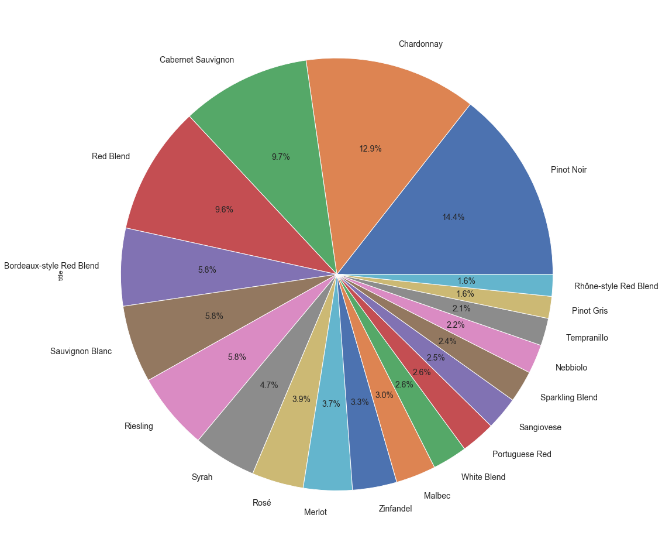
The two metrics that are given to us directly from the dataset used in determining the “best” winery are price and rating. With these two variables in mind, we came up with a new wine metric in order to find a

better balance between the two. This wine metric was calculated by taking the price of the wine and dividing it by the (rating(points)-79). Since the point scale only has a 20 point range from 80-100, the new metric would serve as our overall score to evaluate the quality in terms of what is best. The lower the score of the metric would be the better the wine. We discovered by using pandas that the average overall score of Californian wines was 4.23, while the average overall score of Non-Californian wines was 3.52. This would mean that wines from wineries in California would be more expensive for its rating. From this metric, Californian wineries do not have the best overall wines based on this dataset. Thus, Californian wineries are not the best overall compared to the rest of the world.

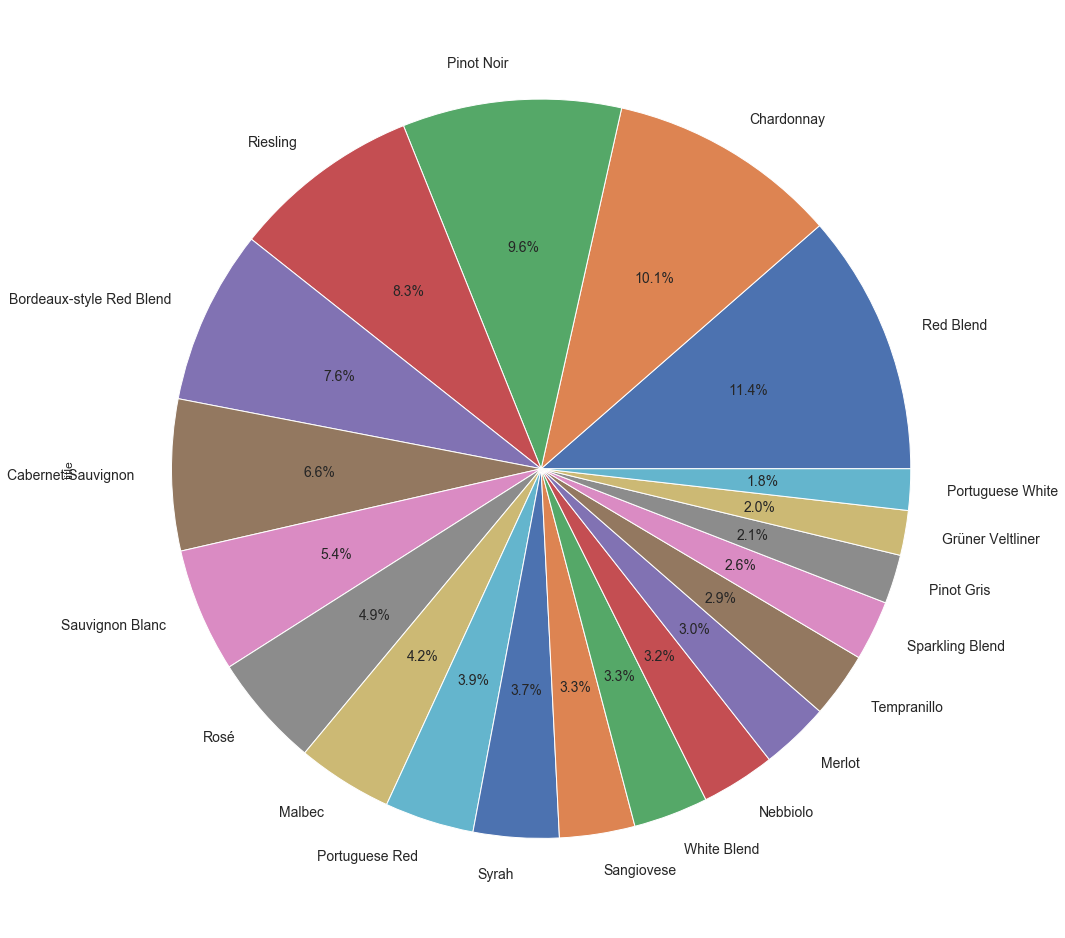
**How does California wine reviews compare to the rest of the world when it comes to the variety of wines reviewed?**

There are 189 varietal types of wine reviewed from California and 686 varietal types of wine reviewed all over the world from this dataset. From the pie plots below, we discovered that California comprised of a significant percentage of the varieties of wine in the world. The top three varietal types (Pinot Noir, Chardonnay, Cabernet Sauvignon) of wine by count in California are also the top three varietal types of wine found globally. These three varietal types of wine form approximately 53% of the varietal types in California, 37 % of the varietal types found globally. When excluding California, this percentage is 26%. Even looking after the top three varietal types, the top 10 most reviewed varietal types in California are among six of the top 10 most reviewed varietal types found elsewhere in the world. From these pie plots, we can suggest that the most reviewed varietal types found in California are also very popular globally.

**Breakdown of 20 varieties of wine in CA by review count Breakdown of top 20 varieties of wine globally by review count**

** **

**Breakdown of top 20 varieties of wine globally (excluding CA) by review count**

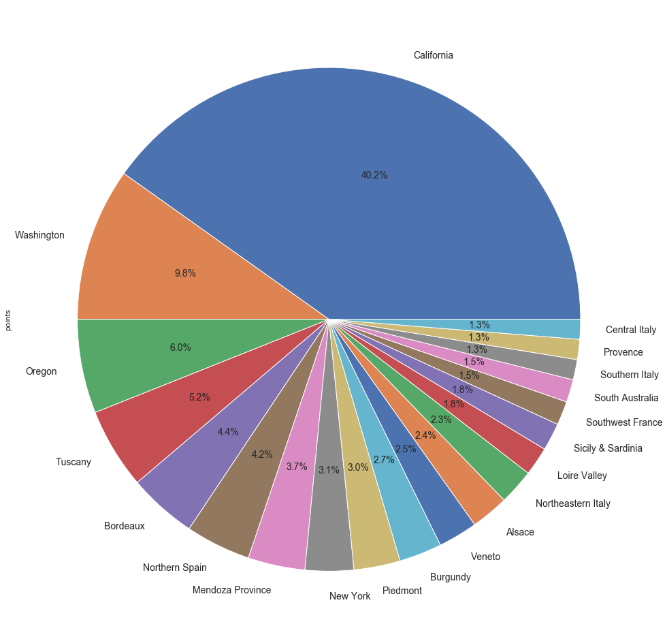
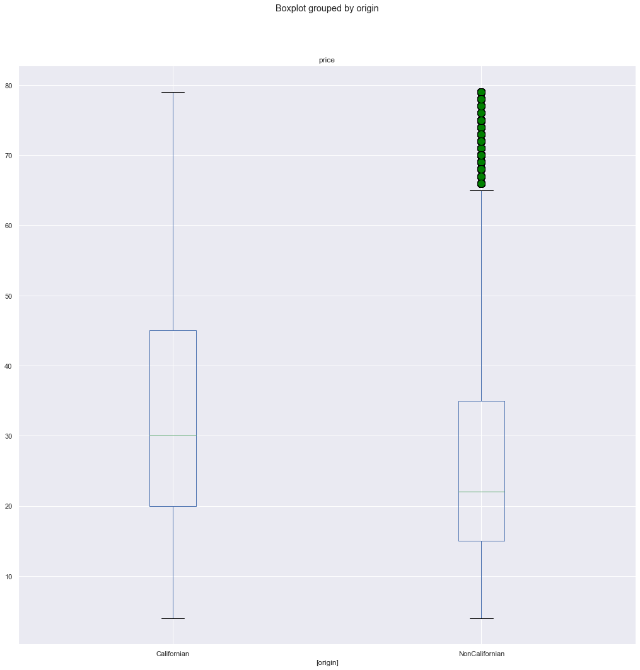


**Does California have the best rated because they have higher diversity of wines?**

While California does have a higher diversity of wines compared with other states/provinces by plurality of reviews and varietal types, we are unable to find any significant relationship between that and review ratings. Californian wines have a marginally higher average rating of approximately 0.3 points higher than that of other wines. To try to highlight the quality of California wines a bit further, we set a cutoff for a “good wine” with any price lower than the lower quartile and rating above the upper quartile. This was illustrated with a bar graph shown below, which displays the absolute frequency of a good wine by state/province.

We also looked at the price distribution for both groups and found that California wines have interquartile range along with a higher median per boxplot shown below. This may suggest a greater distribution of prices. However, this observation was poorly linked to ratings of the wines with the rating difference being very small.

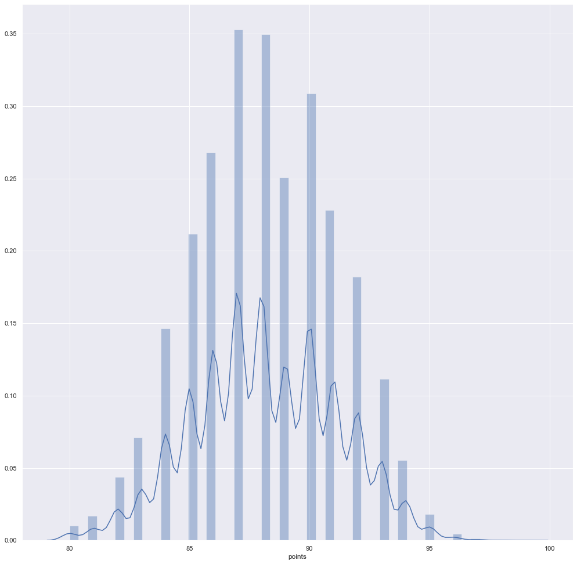
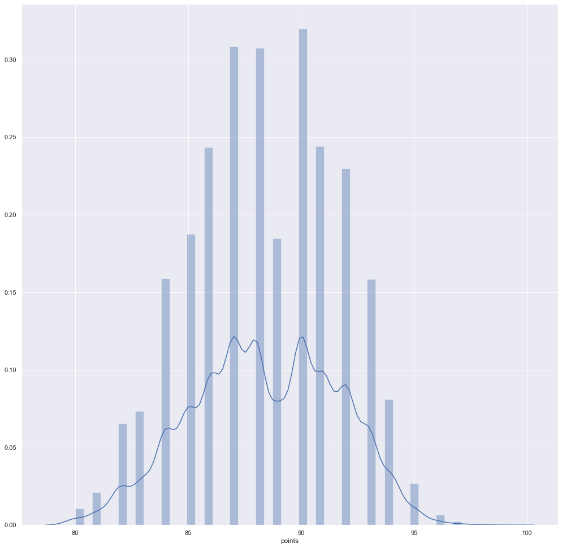
**Boxplot of prices of CA and Non-CA wines** **Breakdown of top 20 states/provinces with the most reviews**



**Do California have the best rated because they have a high number of distributions of reviews(points) by year?**

Despite California having a plurality in the number of reviews compared with other states/provinces, there was no significant difference in the distributions of reviews (points) compared with that of globally. From The histograms below illustrate distributions of both reviews of California wines and global wines. Both histograms have a similar bell curve shaped distribution. However, the global wine distribution is slightly steeper in the middle, which suggests greater relative frequency of wines from 85-90 points. Both distributions also have dip at around 89 points, which may be a simple outlier. Overall, the distributions of reviews are very similar.

**Histogram of Global Review ratings of wines Histogram of CA Review ratings of wines**

As stated before, the average overall score (using our new wine metric) of Californian wines is higher than wines found globally. By using the chi-square test between the observed and expected of California and Non-California wines using our overall score, there is no statistical difference observed as the chi-square value of 0.128 is much lower than the critical value of 3.84. Therefore, we cannot conclude that California has the best wine from this analysis.