Codegnan
Ramavath

ABSTRACT_

I. Introduction

Wine is one of the most popular alcoholic beverage across the world, and it's quality are large affected by multiple factors, such as winery, location, grape type, wine type etc. In order to gather consumer reviews toward different brand, type of wine, many Wine publications created communities for wine lovers to submit their reviews and ratings, such as Wine Enthusiast network by the Wine Magazine.

In a nutshell, in this project, we want to explore:

How different factors impact wine's rating? Key factors include grape variety, winery, country, designation, #2. vineyard, price, and region.
What's the relation between winery and price?
What is the production amount distribution by locations?
Would any specific wine types has/have obvious price advantage?
Is there correlation between wine price and rating (on major wine rating sites)?
Would any specific wineries has/have obvious rate advantage

II. Data Analysis Methods

To complete the research questions, we leveraged several different data visualisations methods with the wine review data, including number chart,scatterplot, box chart, bar chart, density chart, etc. The data are mostly straight forward, but required cleaning before performing analytics and visualisations.

Reading Reference: 1. According to Few (2014), several paragraphs in this article showed bar charts displaying the primary story of the results. Bar charts are the easiest way to read and understand when representing one value, such as countries of wine, wine varieties or prices. Ranking bars help readers compare the values.
2. Both calculation and graphs should contribute to presenting and understanding is as ground rules when choosing method of visualization (Anscombe, 1973).A scatter plot in this article perfectly indicated the purpose and displayed the relationship of two variables, which was also as a choice in our wine research to show diverse price points of wines in different production countries.
Including Plots

You can also embed plots, for example:

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.