Project Proposal - Tennis Trio

Due Friday, October 9, 11:59 PM

Tennis Trio: Hamilton Murrah, Kellyn McDonald, Naima Turbes

```
library(tidyverse)
wine_data <- read_csv("data/winemag-data_first150k.csv")
wine_data
```

```
## # A tibble: 8,000 x 11
##       X1 country description designation points price province region_1 region_2
##    <dbl> <chr>   <chr>       <chr>        <dbl> <dbl> <chr>    <chr>    <chr>
##  1     0 US      This treme~ Martha's V~     96   235 Califor~ Napa Va~ Napa
##  2     1 Spain   Ripe aroma~ Carodorum ~     96   110 Norther~ Toro     <NA>
##  3     2 US      Mac Watson~ Special Se~     96    90 Califor~ Knights~ Sonoma
##  4     3 US      This spent~ Reserve         96    65 Oregon   Willame~ Willame~
##  5     4 France  This is th~ La Brûlade      95    66 Provence Bandol   <NA>
##  6     5 Spain   Deep, dens~ Numanthia       95    73 Norther~ Toro     <NA>
##  7     6 Spain   Slightly g~ San Román       95    65 Norther~ Toro     <NA>
##  8     7 Spain   Lush cedar~ Carodorum ~     95   110 Norther~ Toro     <NA>
##  9     8 US      This re-na~ Silice          95    65 Oregon   Chehale~ Willame~
## 10     9 US      The produc~ Gap's Crow~     95    60 Califor~ Sonoma ~ Sonoma
## # ... with 7,990 more rows, and 2 more variables: variety <chr>, winery <chr>
```

Introduction

We are investigating wine ratings from wine enthusiasts. We chose this data because we understand that people relate to each other and share culture through wine drinking. According to a 2019 study by Forbes, the best ranked wine was from Napa Valley (https://www.forbes.com/sites/eustaciahuen/2019/03/25/wine-3/#6b2719750ed0). We are curious to see how and if this has changed.

Our primary research question is as follows: What is the relationship between the best wines and where these wines from? Our hypothesis is that the best wines are from France and Italy because we know that this is where most wines come from. We also hypothesize that these higher rated wines will be more expensive.

Data Description

User zackthoutt on Kaggle collected this data by scraping data from WineEnthusiast.com.

A description of the columns of our data set are as follows

- `X1`: Number corresponding to the observation. Listed in ascending order from 0
- `country`: The country that the wine is from
- `description`: A few sentences from a sommelier describing the wine's taste, smell, look, feel, etc.
- `designation`: The vineyard within the winery where the grapes that made the wine are from
- `points`: The number of points WineEnthusiast rated the wine on a scale of 1-100 (though they say they only post reviews for wines that score ≥ 80)
- `price`: The cost for a bottle of the wine
- `province`: The province or state that the wine is from
- `region_1`: The wine growing area in a province or state (ie Napa)

- `region_2`: Sometimes there are more specific regions specified within a wine growing area (ie Rutherford inside the Napa Valley), but this value can sometimes be blank
- `variety`: The type of grapes used to make the wine (ie Pinot Noir)
- `winery`: The winery that made the wine

Glimpse of data

```r
glimpse(wine_data)
```

```
## Rows: 8,000
## Columns: 11
## $ X1          <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ country     <chr> "US", "Spain", "US", "US", "France", "Spain", "Spain", ...
## $ description <chr> "This tremendous 100% varietal wine hails from Oakville...
## $ designation <chr> "Martha's Vineyard", "Carodorum Selección Especial Rese...
## $ points      <dbl> 96, 96, 96, 96, 95, 95, 95, 95, 95, 95, 95, 95, 95, 95,...
## $ price       <dbl> 235, 110, 90, 65, 66, 73, 65, 110, 65, 60, 80, 48, 48, ...
## $ province    <chr> "California", "Northern Spain", "California", "Oregon",...
## $ region_1    <chr> "Napa Valley", "Toro", "Knights Valley", "Willamette Va...
## $ region_2    <chr> "Napa", NA, "Sonoma", "Willamette Valley", NA, NA, NA, ...
## $ variety     <chr> "Cabernet Sauvignon", "Tinta de Toro", "Sauvignon Blanc...
## $ winery      <chr> "Heitz", "Bodega Carmen Rodríguez", "Macauley", "Ponzi"...
```