STAT 201: Midterm 1 Practice 1

Your name

In the following code chunk, load your libraries for wrangling, plotting, and making pretty tables. Then run the code chunk:

We have data from over 1000 wines in the dataset `wine_ratings`. Each case in the dataset represents one bottle of wine. The wines included in the dataset are wines that were tasted and reviewed.

- `country`: country of origin

- `variety`: grape type

- `points`: the number of points WineEnthusiast rated the wine on a scale of 1-100 (100 best)

- `price`: price of the wine (USD)

- `title`: title of the wine review, which often contains the year of the wine

- `taster_name`: name of taster/reviewer

- `description`: flavor and taste profile as written by the reviewer

- `year`: year/vintage of the wine, if available in the title of the review

Run the following code to load in the data and take a look at it before continuing!

Exercise 1

First, modify the `wine_ratings` data to remove observations for which we don't have information about price.

Then find the three countries that are most represented in the dataset. Once you have identified those countries, create a new data frame called `wine_ratings_top` that only retains observations from those three countries.

Exercise 2

Obtain the mean and standard deviation of the points and prices of the wines for each of the three countries. Display as an informative, beautiful table. Briefly interpret what you see.
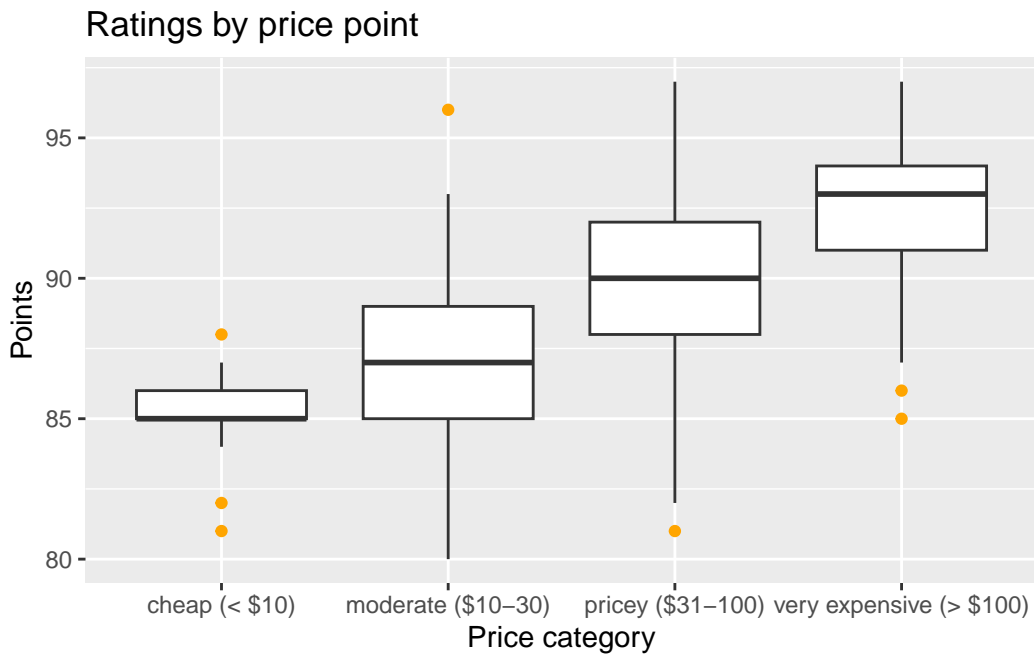
Answer:

Exercise 3

For each of the three countries identified earlier, what proportion of their wines that were produced before 2010 received over 90 points? Display a beautiful table that only shows the country name and the proportion.

Exercise 4

Re-create the following plot and interpret it:

Answer:

Exercise 5

Among all wines in the original dataset that have at least 25 reviews and information about prices, determine which type of wine grape seems to be the worst value. In your answer, briefly describe your methods/reasoning and identify the wine variety.

Answer: