ST557: FINAL PROJECT

Brian Cervantes Alvarez
December 2, 2023

Part A

```
library(tidyverse)


# Read in the wine datasets
redWine <- read_csv("winequality-red.csv")
whiteWine <- read_csv("winequality-white.csv")
description <- read_csv("winequality-info.txt")

# Confirm the structure of the data given the description file
glimpse(redWine)
```

```
Rows: 1,599
Columns: 12
$ `fixed acidity`        <dbl> 7.4, 7.8, 7.8, 11.2, 7.4, 7.4, 7.9, 7.3, 7.8, 7~
$ `volatile acidity`     <dbl> 0.700, 0.880, 0.760, 0.280, 0.700, 0.660, 0.600~
$ `citric acid`          <dbl> 0.00, 0.00, 0.04, 0.56, 0.00, 0.00, 0.06, 0.00,~
$ `residual sugar`       <dbl> 1.9, 2.6, 2.3, 1.9, 1.9, 1.8, 1.6, 1.2, 2.0, 6.~
$ chlorides              <dbl> 0.076, 0.098, 0.092, 0.075, 0.076, 0.075, 0.069~
$ `free sulfur dioxide`  <dbl> 11, 25, 15, 17, 11, 13, 15, 15, 9, 17, 15, 17, ~
$ `total sulfur dioxide` <dbl> 34, 67, 54, 60, 34, 40, 59, 21, 18, 102, 65, 10~
$ density                <dbl> 0.9978, 0.9968, 0.9970, 0.9980, 0.9978, 0.9978,~
$ pH                     <dbl> 3.51, 3.20, 3.26, 3.16, 3.51, 3.51, 3.30, 3.39,~
$ sulphates              <dbl> 0.56, 0.68, 0.65, 0.58, 0.56, 0.56, 0.46, 0.47,~
$ alcohol                <dbl> 9.4, 9.8, 9.8, 9.8, 9.4, 9.4, 9.4, 10.0, 9.5, 1~
$ quality                <dbl> 5, 5, 5, 6, 5, 5, 5, 7, 7, 5, 5, 5, 5, 5, 5, 5,~
```

```
glimpse(whiteWine)
```

```
Rows: 4,898
Columns: 12
$ `fixed acidity`        <dbl> 7.0, 6.3, 8.1, 7.2, 7.2, 8.1, 6.2, 7.0, 6.3, 8.~
$ `volatile acidity`     <dbl> 0.27, 0.30, 0.28, 0.23, 0.23, 0.28, 0.32, 0.27,~
$ `citric acid`          <dbl> 0.36, 0.34, 0.40, 0.32, 0.32, 0.40, 0.16, 0.36,~
$ `residual sugar`       <dbl> 20.70, 1.60, 6.90, 8.50, 8.50, 6.90, 7.00, 20.7~
```

```
$ chlorides              <dbl> 0.045, 0.049, 0.050, 0.058, 0.058, 0.050, 0.045~
$ `free sulfur dioxide`  <dbl> 45, 14, 30, 47, 47, 30, 30, 45, 14, 28, 11, 17,~
$ `total sulfur dioxide` <dbl> 170, 132, 97, 186, 186, 97, 136, 170, 132, 129,~
$ density                <dbl> 1.0010, 0.9940, 0.9951, 0.9956, 0.9956, 0.9951,~
$ pH                     <dbl> 3.00, 3.30, 3.26, 3.19, 3.19, 3.26, 3.18, 3.00,~
$ sulphates              <dbl> 0.45, 0.49, 0.44, 0.40, 0.40, 0.44, 0.47, 0.45,~
$ alcohol                <dbl> 8.8, 9.5, 10.1, 9.9, 9.9, 10.1, 9.6, 8.8, 9.5, ~
$ quality                <dbl> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 5, 5, 5, 7, 5, 7,~
```

```r
# Run a quick summary for both datasets
summary(redWine)
```

```
 fixed acidity   volatile acidity  citric acid    residual sugar
 Min.   : 4.60   Min.   :0.1200   Min.   :0.000   Min.   : 0.900
 1st Qu.: 7.10   1st Qu.:0.3900   1st Qu.:0.090   1st Qu.: 1.900
 Median : 7.90   Median :0.5200   Median :0.260   Median : 2.200
 Mean   : 8.32   Mean   :0.5278   Mean   :0.271   Mean   : 2.539
 3rd Qu.: 9.20   3rd Qu.:0.6400   3rd Qu.:0.420   3rd Qu.: 2.600
 Max.   :15.90   Max.   :1.5800   Max.   :1.000   Max.   :15.500
   chlorides       free sulfur dioxide total sulfur dioxide    density
 Min.   :0.01200   Min.   : 1.00       Min.   :  6.00       Min.   :0.9901
 1st Qu.:0.07000   1st Qu.: 7.00       1st Qu.: 22.00       1st Qu.:0.9956
 Median :0.07900   Median :14.00       Median : 38.00       Median :0.9968
 Mean   :0.08747   Mean   :15.87       Mean   : 46.47       Mean   :0.9967
 3rd Qu.:0.09000   3rd Qu.:21.00       3rd Qu.: 62.00       3rd Qu.:0.9978
 Max.   :0.61100   Max.   :72.00       Max.   :289.00       Max.   :1.0037
       pH          sulphates         alcohol         quality
 Min.   :2.740   Min.   :0.3300   Min.   : 8.40   Min.   :3.000
 1st Qu.:3.210   1st Qu.:0.5500   1st Qu.: 9.50   1st Qu.:5.000
 Median :3.310   Median :0.6200   Median :10.20   Median :6.000
 Mean   :3.311   Mean   :0.6581   Mean   :10.42   Mean   :5.636
 3rd Qu.:3.400   3rd Qu.:0.7300   3rd Qu.:11.10   3rd Qu.:6.000
 Max.   :4.010   Max.   :2.0000   Max.   :14.90   Max.   :8.000
```

```r
summary(whiteWine)
```

```
 fixed acidity    volatile acidity  citric acid     residual sugar
 Min.   : 3.800   Min.   :0.0800   Min.   :0.0000   Min.   : 0.600
 1st Qu.: 6.300   1st Qu.:0.2100   1st Qu.:0.2700   1st Qu.: 1.700
 Median : 6.800   Median :0.2600   Median :0.3200   Median : 5.200
 Mean   : 6.855   Mean   :0.2782   Mean   :0.3342   Mean   : 6.391
 3rd Qu.: 7.300   3rd Qu.:0.3200   3rd Qu.:0.3900   3rd Qu.: 9.900
 Max.   :14.200   Max.   :1.1000   Max.   :1.6600   Max.   :65.800
```

```
  chlorides       free sulfur dioxide total sulfur dioxide    density
Min.   :0.00900   Min.   :  2.00      Min.   :  9.0        Min.   :0.9871
1st Qu.:0.03600   1st Qu.: 23.00      1st Qu.:108.0        1st Qu.:0.9917
Median :0.04300   Median : 34.00      Median :134.0        Median :0.9937
Mean   :0.04577   Mean   : 35.31      Mean   :138.4        Mean   :0.9940
3rd Qu.:0.05000   3rd Qu.: 46.00      3rd Qu.:167.0        3rd Qu.:0.9961
Max.   :0.34600   Max.   :289.00      Max.   :440.0        Max.   :1.0390
      pH          sulphates         alcohol         quality
Min.   :2.720   Min.   :0.2200   Min.   : 8.00   Min.   :3.000
1st Qu.:3.090   1st Qu.:0.4100   1st Qu.: 9.50   1st Qu.:5.000
Median :3.180   Median :0.4700   Median :10.40   Median :6.000
Mean   :3.188   Mean   :0.4898   Mean   :10.51   Mean   :5.878
3rd Qu.:3.280   3rd Qu.:0.5500   3rd Qu.:11.40   3rd Qu.:6.000
Max.   :3.820   Max.   :1.0800   Max.   :14.20   Max.   :9.000
```