

# GPH 2338 Project

Yi Yang

2023-03-13

```
library(haven)
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.2.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
```

```
##
```

```
##      %+%, alpha
```

```
library(psych)
library(pander)
```

```
## Warning: package 'pander' was built under R version 4.2.2
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.2.2
```

```
## corrplot 0.92 loaded
```

```
require(readr)
```

```
## Loading required package: readr
```

```
## Warning: package 'readr' was built under R version 4.2.2
```

```
require(r02pro)
```

```
## Loading required package: r02pro
```

```
## Warning: package 'r02pro' was built under R version 4.2.2
```

```
wine <- read_csv("Wine_Quality_Data.csv")
```

```
## Rows: 6497 Columns: 13
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (1): color  
## dbl (12): fixed_acidity, volatile_acidity, citric_acid, residual_sugar, chlo...  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
head(wine)
```

```
## # A tibble: 6 x 13  
##   fixed_~1 volat~2 citri~3 resid~4 chlor~5 free_~6 total~7 density pH sulph~8  
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 7.4 0.7 0 1.9 0.076 11 34 0.998 3.51 0.56  
## 2 7.8 0.88 0 2.6 0.098 25 67 0.997 3.2 0.68  
## 3 7.8 0.76 0.04 2.3 0.092 15 54 0.997 3.26 0.65  
## 4 11.2 0.28 0.56 1.9 0.075 17 60 0.998 3.16 0.58  
## 5 7.4 0.7 0 1.9 0.076 11 34 0.998 3.51 0.56  
## 6 7.4 0.66 0 1.8 0.075 13 40 0.998 3.51 0.56  
## # ... with 3 more variables: alcohol <dbl>, quality <dbl>, color <chr>, and  
## # abbreviated variable names 1: fixed_acidity, 2: volatile_acidity,  
## # 3: citric_acid, 4: residual_sugar, 5: chlorides, 6: free_sulfur_dioxide,  
## # 7: total_sulfur_dioxide, 8: sulphates
```

Data Preparation

```
sum(wine$color == "white")
```

```
## [1] 4898
```

```
range(wine$fixed_acidity)
```

```
## [1] 3.8 15.9
```

```
range(wine$volatile_acidity)
```

```
## [1] 0.08 1.58
```

```
range(wine$citric_acid)
```

```
## [1] 0.00 1.66
```

```
range(wine$residual_sugar)
```

```
## [1] 0.6 65.8
```

```
range(wine$chlorides)
```

```
## [1] 0.009 0.611
```

```
range(wine$free_sulfur_dioxide)
```

```
## [1] 1 289
```

```
range(wine$total_sulfur_dioxide)
```

```
## [1] 6 440
```

```
range(wine$density)
```

```
## [1] 0.98711 1.03898
```

```
range(wine$pH)
```

```
## [1] 2.72 4.01
```

```
range(wine$sulphates)
```

```
## [1] 0.22 2.00
```

```
range(wine$alcohol)
```

```
## [1] 8.0 14.9
```

```
range(wine$quality)
```

```
## [1] 3 9
```

```
range(wine$color)
```

```
## [1] "red" "white"
```

```
sum(is.na(wine$fixed_acidity))
```

```
## [1] 0
```

```
sum(is.na(wine$fixed_acidity))
```

```
## [1] 0
```

```
sum(is.na(wine$volatile_acidity))
```

```
## [1] 0
```

```
sum(is.na(wine$citric_acid))
```

```
## [1] 0
```

```
sum(is.na(wine$residual_sugar))
```

```
## [1] 0
```

```
sum(is.na(wine$chlorides))
```

```
## [1] 0
```

```
sum(is.na(wine$free_sulfur_dioxide))
```

```
## [1] 0
```

```
sum(is.na(wine$total_sulfur_dioxide))
```

```
## [1] 0
```

```
sum(is.na(wine$density))
```

```
## [1] 0
```

```
sum(is.na(wine$pH))
```

```
## [1] 0
```

```
sum(is.na(wine$sulphates))
```

```
## [1] 0
```

```
sum(is.na(wine$alcohol))
```

```
## [1] 0
```

```
sum(is.na(wine$quality))
```

```
## [1] 0
```