Wine Quality

Daerian Dilkumar, Guan Yu Chen, Derek Frempong, Zongqi Li

2017-11-12

UTorIDs, Student Numbers:

dilkumar, 1000356198

chengu17, 1001591967

lizong4, 1000481196

frempon2, 1000566675

Contents

1 Abstract **3**

2 Introduction **3**

3 Set-Up and Imports **3**

4 Data Analysis **4**

1 Abstract

TODO: FILL

2 Introduction

TODO: FILL

3 Set-Up and Imports

We will begin by importing our data sets and then prepping them to be worked with

```
# Import raw data sets
redWineData = read_delim("winequality-red.csv", delim = ";")
```

```
## Parsed with column specification:
## cols(
##   `fixed acidity` = col_double(),
##   `volatile acidity` = col_double(),
##   `citric acid` = col_double(),
##   `residual sugar` = col_double(),
##   chlorides = col_double(),
##   `free sulfur dioxide` = col_double(),
##   `total sulfur dioxide` = col_integer(),
##   density = col_double(),
##   pH = col_double(),
##   sulphates = col_double(),
##   alcohol = col_double(),
##   quality = col_integer()
## )
```

```
## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)
```

```
## Warning: 2 parsing failures.
## row # A tibble: 2 x 5 col     row                   col               expected actual expected   <int
```

```
whiteWineData = read_delim("winequality-white.csv", delim = ";")
```

```
## Parsed with column specification:
## cols(
##   `fixed acidity` = col_double(),
##   `volatile acidity` = col_double(),
##   `citric acid` = col_double(),
##   `residual sugar` = col_double(),
##   chlorides = col_double(),
##   `free sulfur dioxide` = col_double(),
##   `total sulfur dioxide` = col_double(),
##   density = col_double(),
##   pH = col_double(),
##   sulphates = col_double(),
##   alcohol = col_double(),
##   quality = col_integer()
```

```
## )
# remove NA values
redWineData = (redWineData[complete.cases(redWineData),])
whiteWineData = (whiteWineData[complete.cases(whiteWineData),])
# Prep Sets for merginglibrary
redWineData = redWineData %>% mutate(Type = "Red")
whiteWineData = whiteWineData %>% mutate (Type = "White")

# Merge Data and get Final Dataset
wineData = rbind(redWineData,whiteWineData)

# Glimpse of the Data Set
glimpse(wineData)
```

```
## Observations: 6,495
## Variables: 13
## $ `fixed acidity`        <dbl> 7.4, 7.8, 7.8, 11.2, 7.4, 7.4, 7.9, 7.3...
## $ `volatile acidity`     <dbl> 0.700, 0.880, 0.760, 0.280, 0.700, 0.66...
## $ `citric acid`          <dbl> 0.00, 0.00, 0.04, 0.56, 0.00, 0.00, 0.0...
## $ `residual sugar`       <dbl> 1.9, 2.6, 2.3, 1.9, 1.9, 1.8, 1.6, 1.2,...
## $ chlorides              <dbl> 0.076, 0.098, 0.092, 0.075, 0.076, 0.07...
## $ `free sulfur dioxide`  <dbl> 11, 25, 15, 17, 11, 13, 15, 15, 9, 17, ...
## $ `total sulfur dioxide` <dbl> 34, 67, 54, 60, 34, 40, 59, 21, 18, 102...
## $ density                <dbl> 0.9978, 0.9968, 0.9970, 0.9980, 0.9978,...
## $ pH                     <dbl> 3.51, 3.20, 3.26, 3.16, 3.51, 3.51, 3.3...
## $ sulphates              <dbl> 0.56, 0.68, 0.65, 0.58, 0.56, 0.56, 0.4...
## $ alcohol                <dbl> 9.4, 9.8, 9.8, 9.8, 9.4, 9.4, 9.4, 10.0...
## $ quality                <int> 5, 5, 5, 6, 5, 5, 5, 7, 7, 5, 5, 5, 5, ...
## $ Type                   <chr> "Red", "Red", "Red", "Red", "Red", "Red...
```

4 Data Analysis