

5110_Final_Project

Harshitha Somala

4/25/2022

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr 0.3.4
## v tibble 3.1.6       v dplyr 1.0.7
## v tidyr 1.1.4        v stringr 1.4.0
## v readr 2.1.1        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(modelr)
library(gridExtra)

## Warning: package 'gridExtra' was built under R version 4.1.3

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine

dir <- "C:/Users/harsh/Desktop/NEU/IDMP/Final Project/winequality/winequality"
dir1 <- file.path(dir, "winequality-red.csv")
dir2 <- file.path(dir, "winequality-white.csv")
red_wine_df <- read.csv(dir1, header = TRUE, sep = ";")
white_wine_df <- read.csv(dir2, header = TRUE, sep = ";")
```

Quality Distribution over Red wine

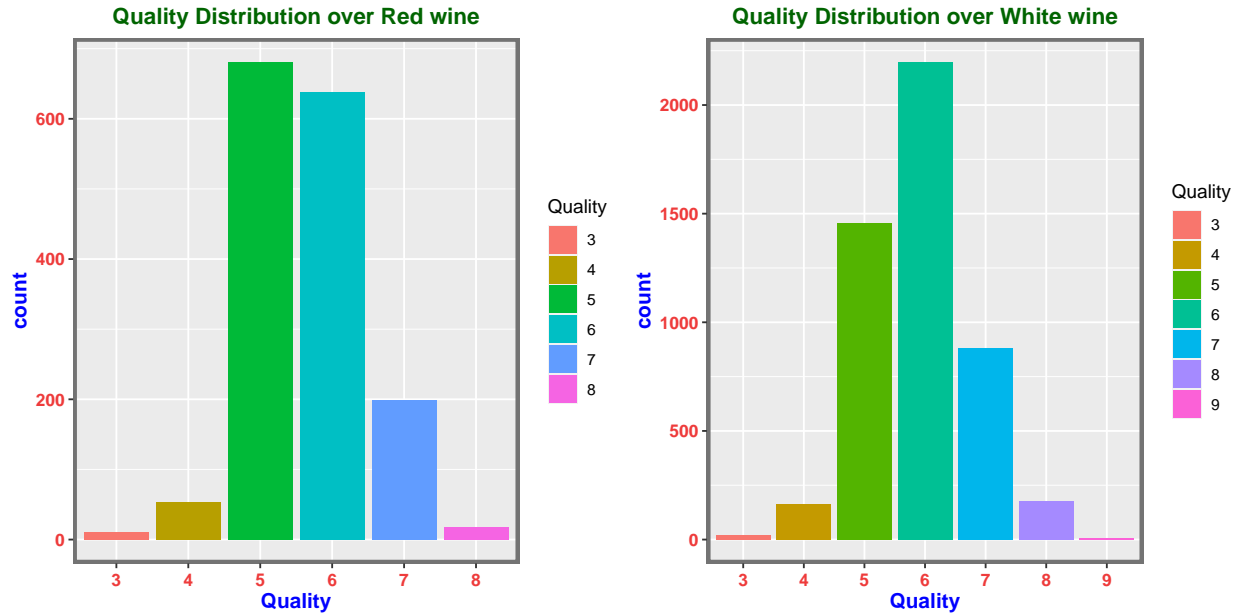
```

qua_red <- ggplot(red_wine_df) +
  geom_bar(aes(x = factor(quality), fill = factor(quality))) +
  labs(title = "Quality Distribution over Red wine",
       x = "Quality",
       fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

qua_white <- ggplot(white_wine_df) +
  geom_bar(aes(x = factor(quality), fill = factor(quality))) +
  labs(title = "Quality Distribution over White wine",
       x = "Quality",
       fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

grid.arrange(qua_red, qua_white, nrow = 1)

```



Fixed Acidity Vs Quality

```
p1 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = fixed.acidity, fill = factor(quality))) +
  labs(title = "Relationship between Fixed Acidity & Quality",
       subtitle = "(Red Wine)",
       x = "Quality",
       y = "Fixed Acidity",
       fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

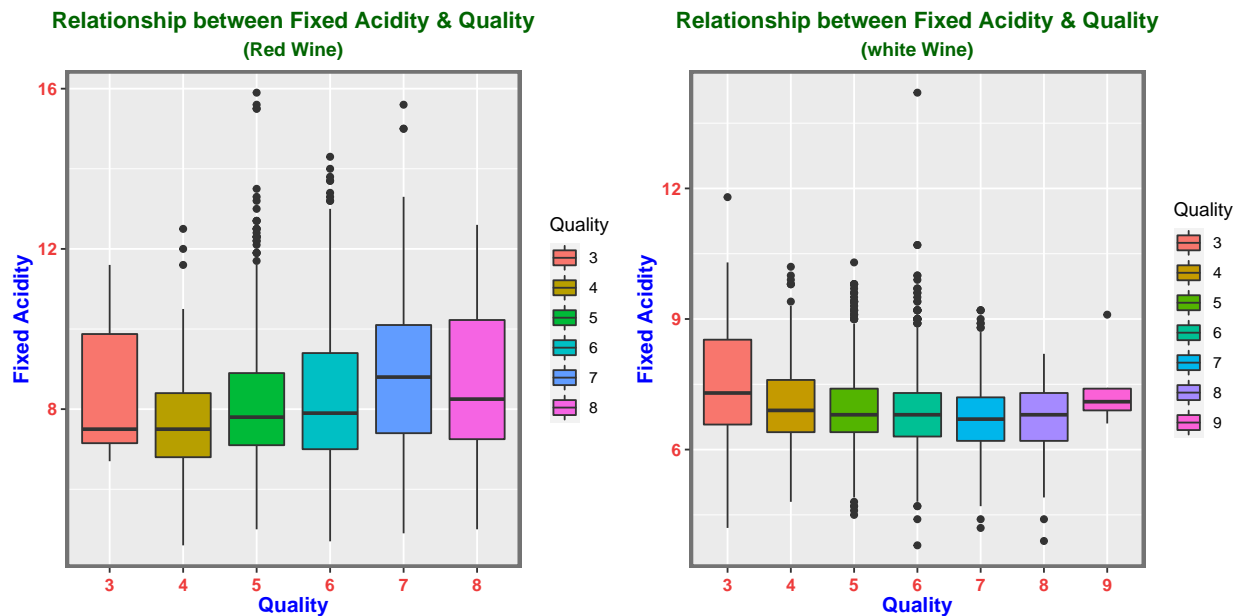
p1w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = fixed.acidity, fill = factor(quality))) +
  labs(title = "Relationship between Fixed Acidity & Quality",
```

```

    subtitle = "(white Wine)",
    x = "Quality",
    y = "Fixed Acidity",
    fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                   vjust = 0.5,
                                   color = "brown2",
                                   face = "bold",
                                   size = 10),
        axis.text.y = element_text(color = "brown2",
                                   face = "bold",
                                   size = 10),
        panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

grid.arrange(p1, p1w, nrow = 1)

```



Volatile Acidity Vs Quality

```

p2 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = volatile.acidity, fill = factor(quality))) +

```

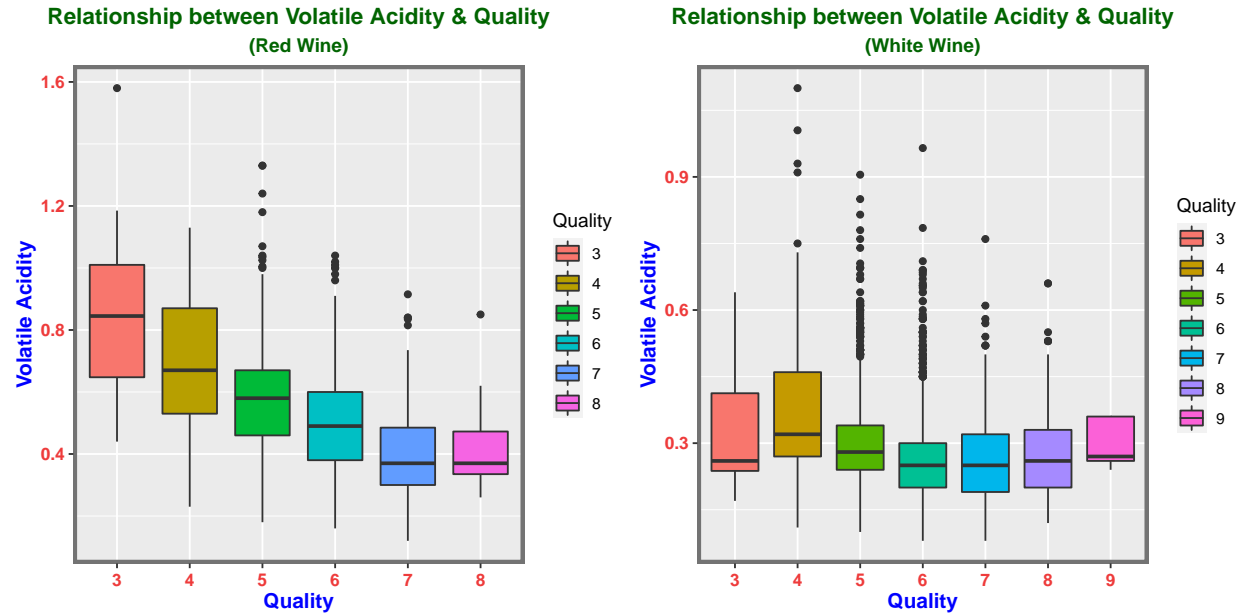
```

labs(title = "Relationship between Volatile Acidity & Quality",
      subtitle = "(Red Wine)",
      x = "Quality",
      y = "Volatile Acidity",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                  fill=NA, size=2))

p2w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = volatile.acidity, fill = factor(quality))) +
  labs(title = "Relationship between Volatile Acidity & Quality",
        subtitle = "(White Wine)",
        x = "Quality",
        y = "Volatile Acidity",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                  color = "darkgreen",
                                  face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                    fill=NA, size=2))

grid.arrange(p2, p2w, nrow = 1)

```



Citric Acid Vs Quality

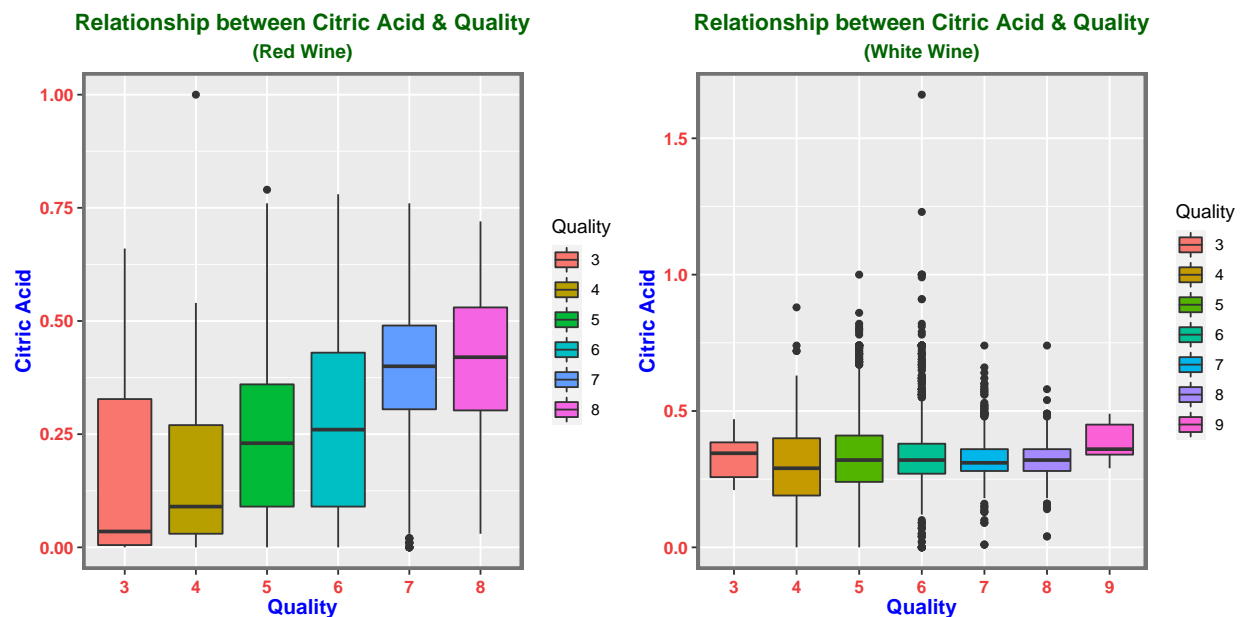
```
p3 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = citric.acid, fill = factor(quality))) +
  labs(title = "Relationship between Citric Acid & Quality",
       subtitle = "(Red Wine)",
       x = "Quality",
       y = "Citric Acid",
       fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                    fill=NA, size=2))

p3w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = citric.acid, fill = factor(quality))) +
  labs(title = "Relationship between Citric Acid & Quality",
       subtitle = "(White Wine)",
       x = "Quality",
```

```

y = "Citric Acid",
fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                    color = "darkgreen",
                                    face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                  fill=NA, size=2))
grid.arrange(p3, p3w, nrow = 1)

```



Residual Sugar Vs Quality

```

p4 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = residual.sugar, fill = factor(quality))) +
  labs(title = "Relationship between Residual Sugar & Quality",
        subtitle = "(Red Wine)",
        x = "Quality",

```

```

    y = "Residual Sugar",
    fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

p4w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = residual.sugar, fill = factor(quality))) +
  labs(title = "Relationship between Residual Sugar & Quality",
       subtitle = "(White Wine)",
       x = "Quality",
       y = "Residual Sugar",
       fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

grid.arrange(p4, p4w, nrow = 1)

```




Chlorides Vs Quality

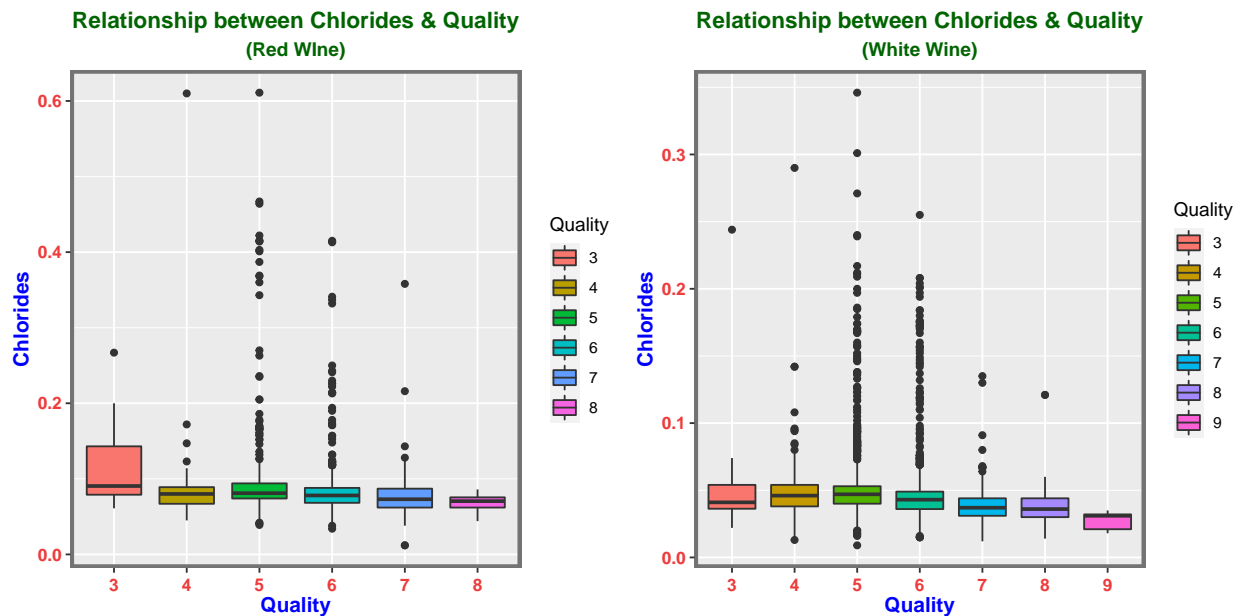
```
p5 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = chlorides, fill = factor(quality))) +
  labs(title = "Relationship between Chlorides & Quality",
        subtitle = "(Red Wine)",
        x = "Quality",
        y = "Chlorides",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                    fill=NA, size=2))

p5w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = chlorides, fill = factor(quality))) +
```

```

labs(title = "Relationship between Chlorides & Quality",
      subtitle = "(White Wine)",
      x = "Quality",
      y = "Chlorides",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))
grid.arrange(p5, p5w, nrow = 1)

```



Fixed Acidity Vs Quality

```

p6 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = free.sulfur.dioxide, fill = factor(quality))) +

```

```

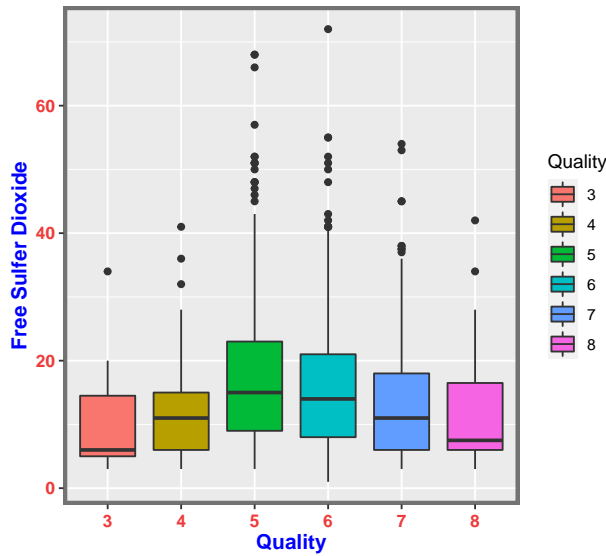
labs(title = "Relationship between Free Sulfur Dioxide & Quality",
      subtitle = "(Red Wine)",
      x = "Quality",
      y = "Free Sulfur Dioxide",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                    color = "darkgreen",
                                    face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

p6w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = free.sulfur.dioxide, fill = factor(quality))) +
  labs(title = "Relationship between Free Sulfur Dioxide & Quality",
        subtitle = "(White Wine)",
        x = "Quality",
        y = "Free Sulfur Dioxide",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                  color = "darkgreen",
                                  face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

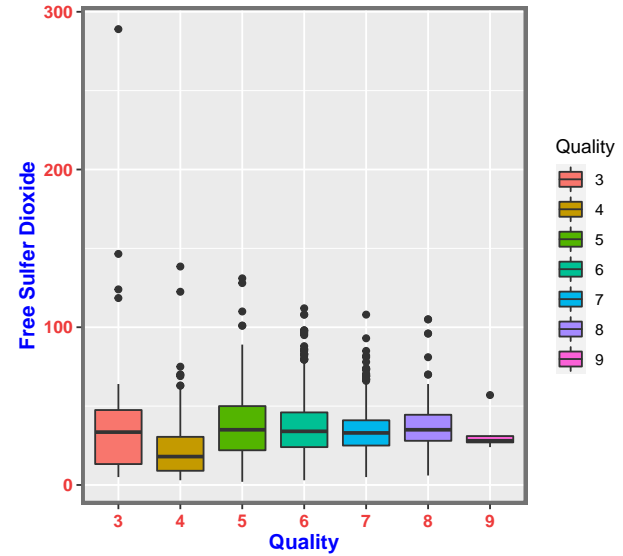
grid.arrange(p6,p6w, nrow = 1)

```

Relationship between Free Sulfur Dioxide & Quality
(Red Wine)



Relationship between Free Sulfur Dioxide & Quality
(White Wine)



Total Sulfur Dioxide Vs Quality

```
p7 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = total.sulfur.dioxide, fill = factor(quality))) +
  labs(title = "Relationship between Total Sulfur Dioxide & Quality",
        subtitle = "(Red Wine)",
        x = "Quality",
        y = "Total Sulfur Dioxide",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                    fill=NA, size=2))

p7w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = total.sulfur.dioxide, fill = factor(quality))) +
  labs(title = "Relationship between Total Sulfur Dioxide & Quality",
```

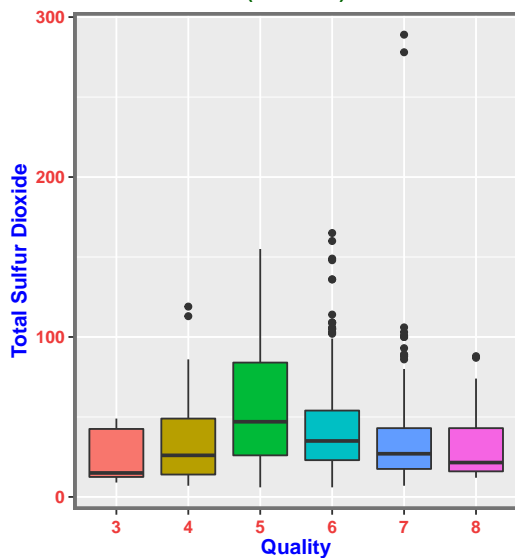
```

    subtitle = "(White Wine)",
    x = "Quality",
    y = "Total Sulfur Dioxide",
    fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

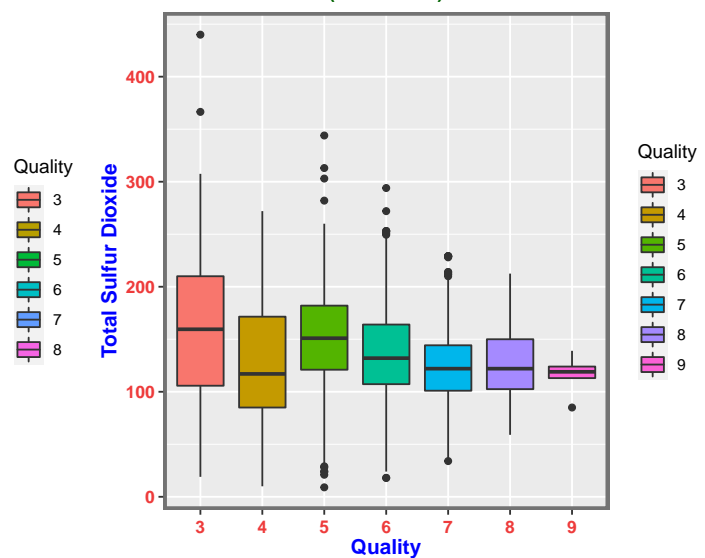
grid.arrange(p7,p7w, nrow = 1)

```

Relationship between Total Sulfur Dioxide & Quality
(Red Wine)



Relationship between Total Sulfur Dioxide & Quality
(White Wine)



Density Vs Quality

```

p8 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = density, fill = factor(quality))) +

```

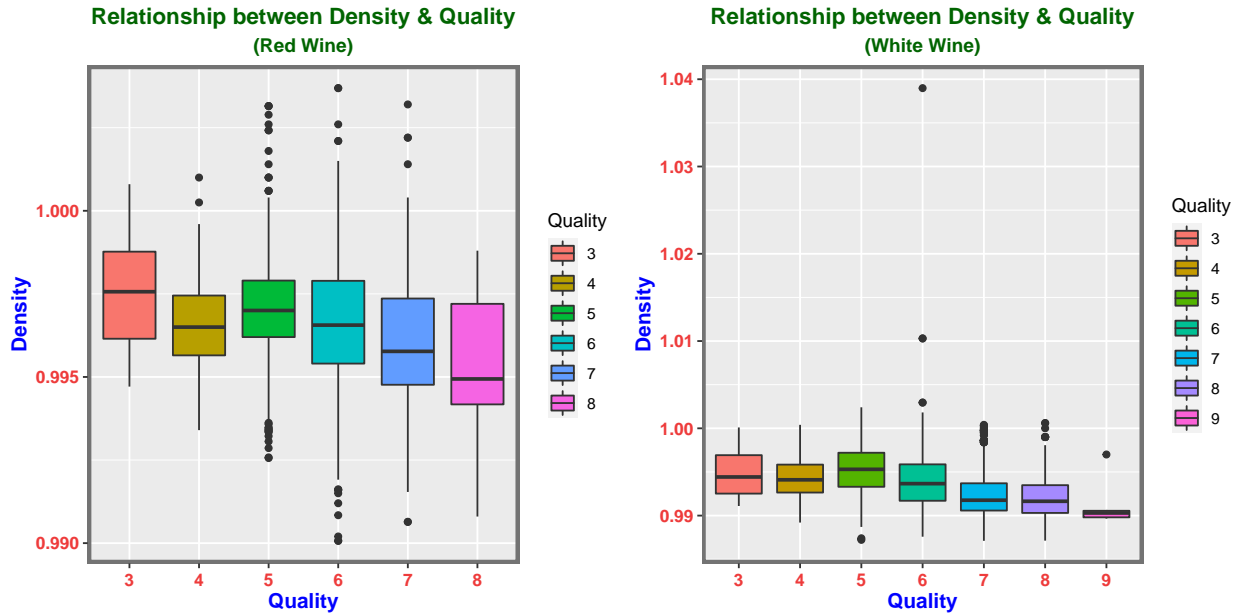
```

labs(title = "Relationship between Density & Quality",
      subtitle = "(Red Wine)",
      x = "Quality",
      y = "Density",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                    color = "darkgreen",
                                    face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

p8w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = density, fill = factor(quality))) +
  labs(title = "Relationship between Density & Quality",
        subtitle = "(White Wine)",
        x = "Quality",
        y = "Density",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                  color = "darkgreen",
                                  face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

grid.arrange(p8,p8w, nrow = 1)

```



pH Vs Quality

```
p9 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = pH, fill = factor(quality))) +
  labs(title = "Relationship between pH & Quality",
        subtitle = "(Red Wine)",
        x = "Quality",
        y = "pH",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                   vjust = 0.5,
                                   color = "brown2",
                                   face = "bold",
                                   size = 10),
        axis.text.y = element_text(color = "brown2",
                                   face = "bold",
                                   size = 10),
        panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

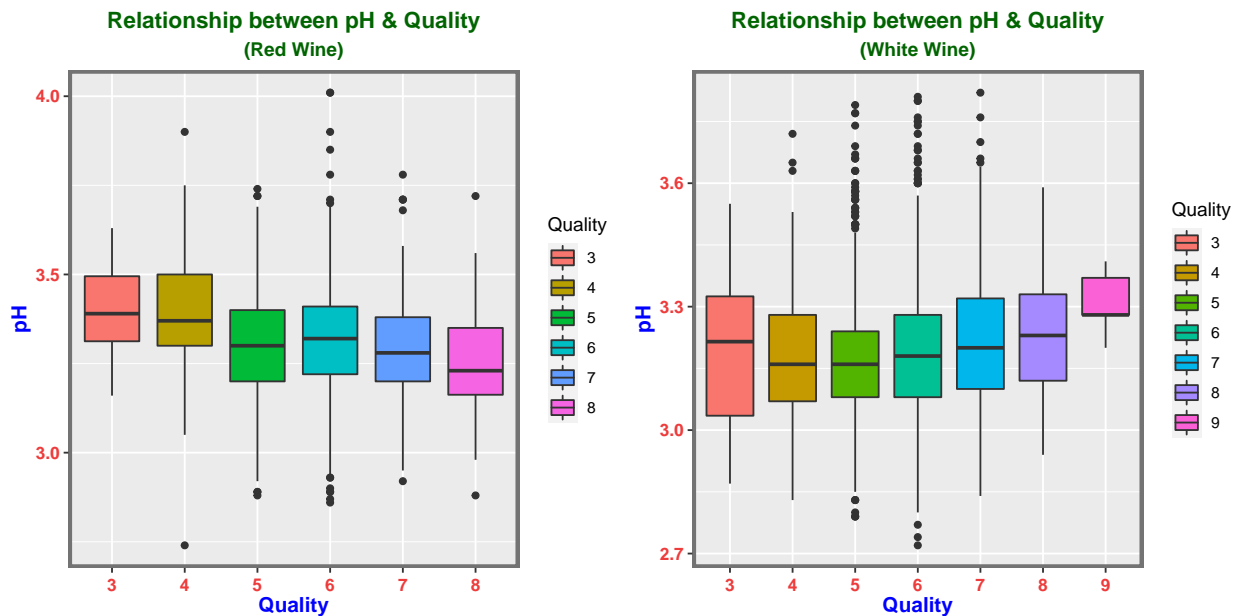
p9w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = pH, fill = factor(quality))) +
  labs(title = "Relationship between pH & Quality",
```

```

    subtitle = "(White Wine)",
    x = "Quality",
    y = "pH",
    fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                       color = "darkgreen",
                                       face = "bold"),
        axis.title = element_text(color = "blue",
                                    face = "bold",
                                    size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

grid.arrange(p9, p9w, nrow = 1)

```



Sulphates Vs Quality

```

p10 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = sulphates, fill = factor(quality))) +

```



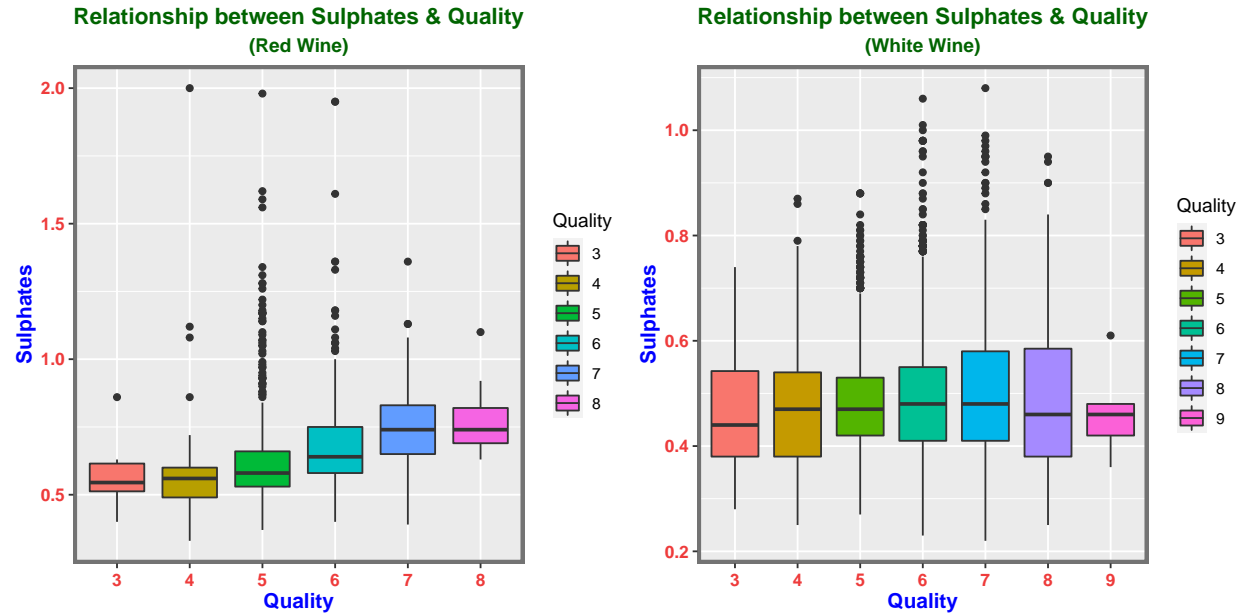
```

labs(title = "Relationship between Sulphates & Quality",
      subtitle = "(Red Wine)",
      x = "Quality",
      y = "Sulphates",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                    color = "darkgreen",
                                    face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

p10w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = sulphates, fill = factor(quality))) +
  labs(title = "Relationship between Sulphates & Quality",
        subtitle = "(White Wine)",
        x = "Quality",
        y = "Sulphates",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                  color = "darkgreen",
                                  face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                    vjust = 0.5,
                                    color = "brown2",
                                    face = "bold",
                                    size = 10),
        axis.text.y = element_text(color = "brown2",
                                    face = "bold",
                                    size = 10),
        panel.border = element_rect(colour = "grey45",
                                     fill=NA, size=2))

grid.arrange(p10, p10w, nrow = 1)

```



Alcohol Vs Quality

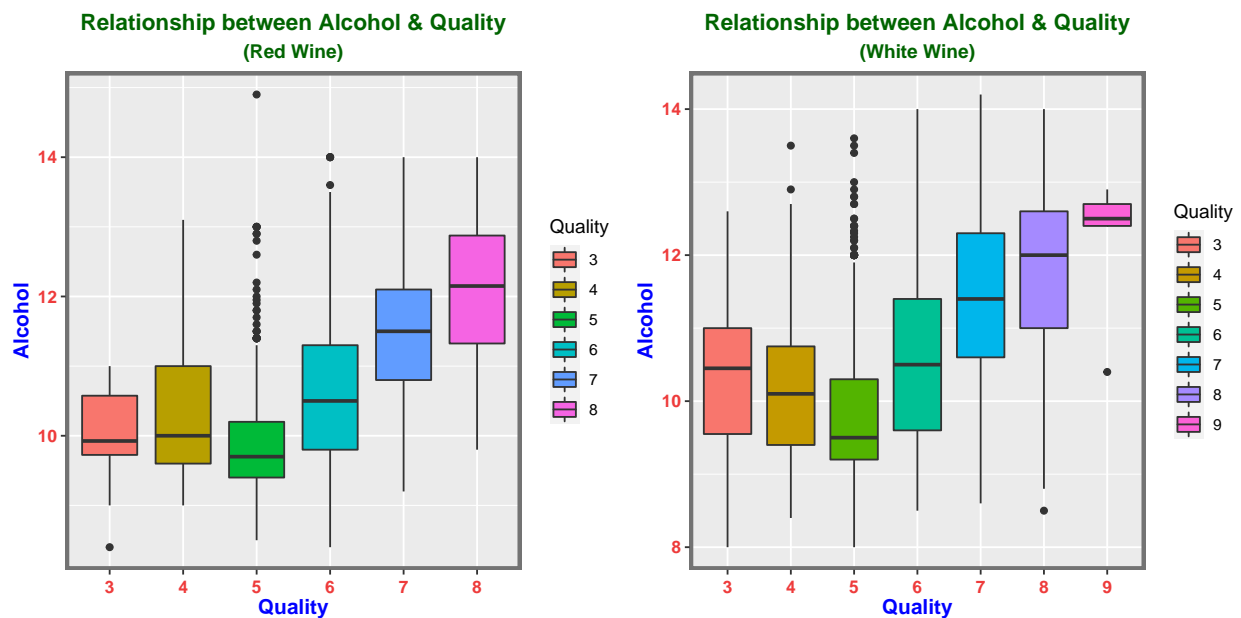
```
p11 <- ggplot(red_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = alcohol, fill = factor(quality))) +
  labs(title = "Relationship between Alcohol & Quality",
        subtitle = "(Red Wine)",
        x = "Quality",
        y = "Alcohol",
        fill = "Quality")+
  theme(plot.title = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
        plot.subtitle = element_text(hjust = 0.5,
                                      color = "darkgreen",
                                      face = "bold"),
        axis.title = element_text(color = "blue",
                                   face = "bold",
                                   size = 12),
        axis.text.x = element_text(hjust = 0.5,
                                   vjust = 0.5,
                                   color = "brown2",
                                   face = "bold",
                                   size = 10),
        axis.text.y = element_text(color = "brown2",
                                   face = "bold",
                                   size = 10),
        panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))

p11w <- ggplot(white_wine_df) +
  geom_boxplot(aes(x = factor(quality), y = alcohol, fill = factor(quality))) +
```

```

labs(title = "Relationship between Alcohol & Quality",
      subtitle = "(White Wine)",
      x = "Quality",
      y = "Alcohol",
      fill = "Quality")+
theme(plot.title = element_text(hjust = 0.5,
                                color = "darkgreen",
                                face = "bold"),
      plot.subtitle = element_text(hjust = 0.5,
                                   color = "darkgreen",
                                   face = "bold"),
      axis.title = element_text(color = "blue",
                                 face = "bold",
                                 size = 12),
      axis.text.x = element_text(hjust = 0.5,
                                  vjust = 0.5,
                                  color = "brown2",
                                  face = "bold",
                                  size = 10),
      axis.text.y = element_text(color = "brown2",
                                  face = "bold",
                                  size = 10),
      panel.border = element_rect(colour = "grey45",
                                   fill=NA, size=2))
grid.arrange(p11, p11w, nrow = 1)

```



Stepwise Model Selection for red wine

```

#Partition of data
red_wine_std <- data.frame(scale(red_wine_df))

```

```
set.seed(10)
partition_rw <- resample_partition(red_wine_std,
                                   p=c(train=0.5,
                                       valid=0.25,
                                       test=0.25))
```

```
#Function to calculate the RMSE of the predictors
step <- function(response, predictors, candidates, partition)
{
  rhs <- paste0(paste0(predictors, collapse="+"), "+", candidates)
  formulas <- lapply(paste0(response, "~", rhs), as.formula)
  rmses <- sapply(formulas,
                  function(fm) rmse(lm(fm, data=partition$train),
                                     data=partition$valid))

  names(rmses) <- candidates
  attr(rmses, "best") <- rmses[which.min(rmses)]
  rmses
}
```

```
model_rw <- NULL

preds <- "1"
cands <- c("fixed.acidity", "volatile.acidity", "citric.acid", "residual.sugar",
          "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density",
          "pH", "sulphates", "alcohol")
s1 <- step("quality", preds, cands, partition_rw)

model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      fixed.acidity    volatile.acidity    citric.acid
##      1.0544658        0.9726777        1.0399684
##      residual.sugar    chlorides    free.sulfur.dioxide
##      1.0622978        1.0491001        1.0565806
## total.sulfur.dioxide    density    pH
##      1.0403927        1.0373038        1.0591219
##      sulphates    alcohol
##      1.0273706    0.9097212
## attr("best")
##      alcohol
## 0.9097212
```

```
preds <- "alcohol"
cands <- c("fixed.acidity", "volatile.acidity", "citric.acid", "residual.sugar",
          "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density",
          "pH", "sulphates")
s1 <- step("quality", preds, cands, partition_rw)

model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      fixed.acidity    volatile.acidity    citric.acid
```

```
##          0.8948661          0.8547690          0.8934924
##      residual.sugar      chlorides  free.sulfur.dioxide
##          0.9097878          0.9092339          0.9089740
## total.sulfur.dioxide      density          pH
##          0.9049693          0.9067079          0.8901634
##          sulphates
##          0.8903749
## attr("best")
## volatile.acidity
##          0.854769
```

```
preds <- c("alcohol","volatile.acidity")
cands <- c("fixed.acidity","citric.acid","residual.sugar",
          "chlorides","free.sulfur.dioxide","total.sulfur.dioxide","density",
          "pH","sulphates")
s1 <- step("quality", preds, cands, partition_rw)

model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      fixed.acidity      citric.acid      residual.sugar
##      0.8517647          0.8547965          0.8558405
##      chlorides  free.sulfur.dioxide total.sulfur.dioxide
##      0.8547562          0.8532196          0.8498207
##      density          pH          sulphates
##      0.8535269          0.8516224          0.8479916
## attr("best")
## sulphates
## 0.8479916
```

```
preds <- c("alcohol","volatile.acidity","sulphates")
cands <- c("fixed.acidity","citric.acid","residual.sugar",
          "chlorides","free.sulfur.dioxide","total.sulfur.dioxide","density",
          "pH")
s1 <- step("quality", preds, cands, partition_rw)

model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      fixed.acidity      citric.acid      residual.sugar
##      0.8459167          0.8478441          0.8491646
##      chlorides  free.sulfur.dioxide total.sulfur.dioxide
##      0.8459167          0.8457725          0.8413798
##      density          pH
##      0.8481381          0.8459547
## attr("best")
## total.sulfur.dioxide
##      0.8413798
```

```
preds <- c("alcohol","volatile.acidity","sulphates","total.sulfur.dioxide")
cands <- c("fixed.acidity","citric.acid","residual.sugar",
          "chlorides","free.sulfur.dioxide","density",
          "pH")
```

```
s1 <- step("quality", preds, cands, partition_rw)
```

```
model_rw <- c(model_rw, attr(s1, "best"))
```

```
s1
```

```
##      fixed.acidity      citric.acid      residual.sugar      chlorides
##      0.8401288         0.8415037         0.8427016         0.8389203
## free.sulfur.dioxide      density      pH
##      0.8428014         0.8415675         0.8390964
## attr("best")
## chlorides
## 0.8389203
```

```
preds <- c("alcohol", "volatile.acidity", "sulphates", "total.sulfur.dioxide", "chlorides")
```

```
cands <- c("fixed.acidity", "citric.acid", "residual.sugar",
           "free.sulfur.dioxide", "density",
           "pH")
```

```
s1 <- step("quality", preds, cands, partition_rw)
```

```
model_rw <- c(model_rw, attr(s1, "best"))
```

```
s1
```

```
##      fixed.acidity      citric.acid      residual.sugar free.sulfur.dioxide
##      0.8371915         0.8390372         0.8402443         0.8395613
##      density      pH
##      0.8391711         0.8347072
## attr("best")
##      pH
## 0.8347072
```

```
preds <- c("alcohol", "volatile.acidity", "sulphates", "total.sulfur.dioxide", "chlorides", "pH")
```

```
cands <- c("fixed.acidity", "citric.acid", "residual.sugar",
           "free.sulfur.dioxide", "density")
```

```
s1 <- step("quality", preds, cands, partition_rw)
```

```
model_rw <- c(model_rw, attr(s1, "best"))
```

```
s1
```

```
##      fixed.acidity      citric.acid      residual.sugar free.sulfur.dioxide
##      0.8348389         0.8342898         0.8361165         0.8356013
##      density
##      0.8355847
## attr("best")
## citric.acid
## 0.8342898
```

```
preds <- c("alcohol", "volatile.acidity", "sulphates", "total.sulfur.dioxide",
           "chlorides", "pH", "citric.acid")
```

```
cands <- c("fixed.acidity", "residual.sugar",
           "free.sulfur.dioxide", "density")
```

```
s1 <- step("quality", preds, cands, partition_rw)
```

```
model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      fixed.acidity      residual.sugar free.sulfur.dioxide      density
##      0.8340361          0.8359434          0.8350586          0.8343719
## attr("best")
## fixed.acidity
##      0.8340361
```

```
preds <- c("alcohol", "volatile.acidity", "sulphates", "total.sulfur.dioxide",
           "chlorides", "pH", "citric.acid", "fixed.acidity")
cands <- c("residual.sugar", "free.sulfur.dioxide", "density")
s1 <- step("quality", preds, cands, partition_rw)

model_rw <- c(model_rw, attr(s1, "best"))
s1
```

```
##      residual.sugar free.sulfur.dioxide      density
##      0.8357948          0.8348426          0.8348930
## attr("best")
## free.sulfur.dioxide
##      0.8348426
```

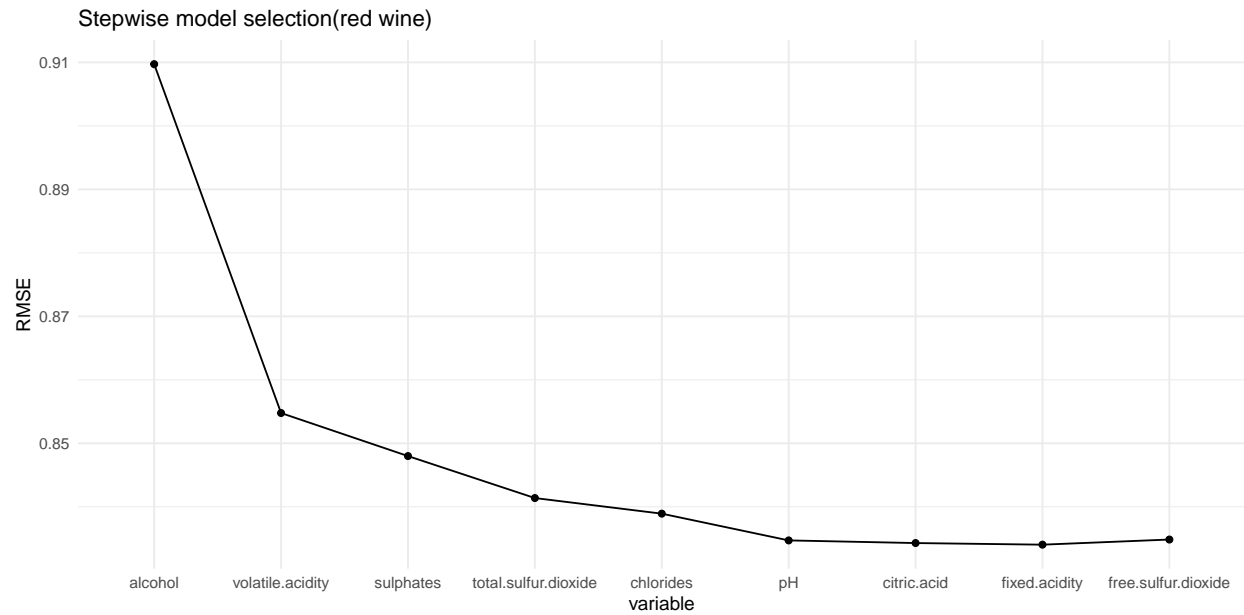
Model stopped improving at:

- fit = quality ~ alcohol + volatile acidity + sulphates + total sulfur dioxide + chlorides + pH + citric acid + fixed acidity

Visualizing how adding each variable affects the RMSE.

```
step_model <- tibble(index=seq_along(model_rw),
                     variable=factor(names(model_rw), levels=names(model_rw)),
                     RMSE=model_rw)

ggplot(step_model, aes(y=RMSE)) +
  geom_point(aes(x=variable)) +
  geom_line(aes(x=index)) +
  labs(title="Stepwise model selection(red wine)") +
  theme_minimal()
```



visualizing how each variable(including the predictors) affects the RMSE for Red wine

```
temp_model_rw <- model_rw
preds <- c("alcohol","volatile.acidity","sulphates","total.sulfur.dioxide",
           "chlorides","pH","citric.acid", "fixed.acidity", "free.sulfur.dioxide")
cands <- c("residual.sugar","density")
s1 <- step("quality", preds, cands, partition_rw)

temp_model_rw <- c(temp_model_rw, attr(s1, "best"))
s1
```

```
## residual.sugar      density
##      0.8363852      0.8356519
## attr("best")
##      density
## 0.8356519
```

```
preds <- c("alcohol","volatile.acidity","sulphates","total.sulfur.dioxide",
           "chlorides","pH","citric.acid", "fixed.acidity", "free.sulfur.dioxide",
           "density")
cands <- c("residual.sugar")
s1 <- step("quality", preds, cands, partition_rw)

temp_model_rw <- c(temp_model_rw, attr(s1, "best"))
s1
```

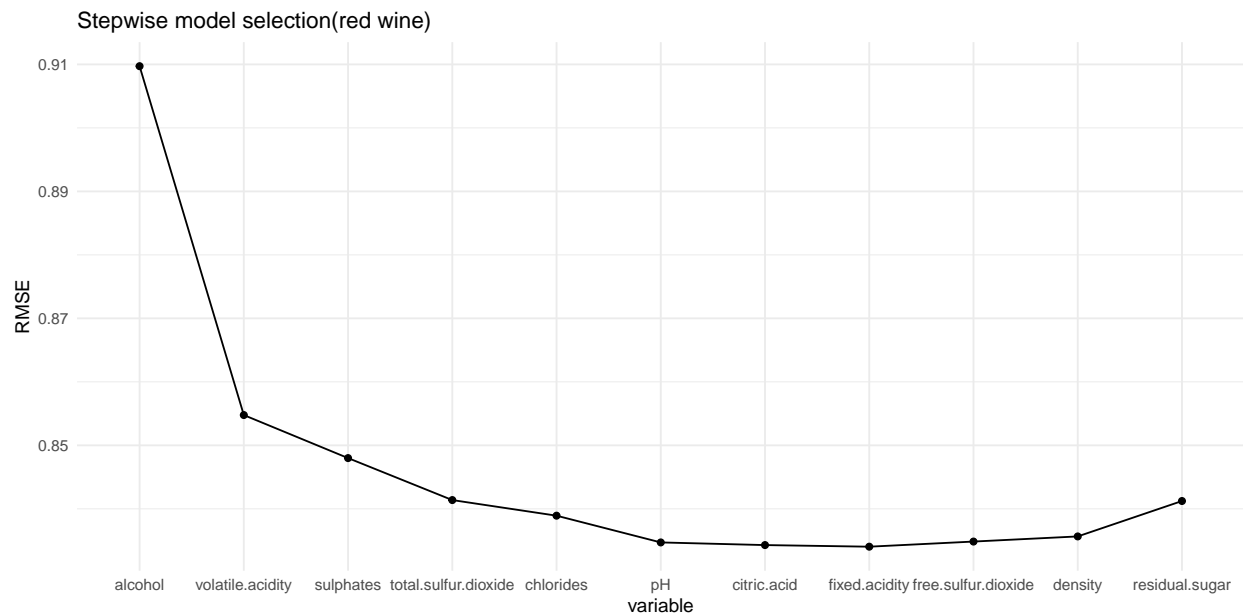
```
## residual.sugar
##      0.8412307
## attr("best")
```



```
## residual.sugar
##      0.8412307
```

```
#Representation of RMSE for all the variables
step_model <- tibble(index=seq_along(temp_model_rw),
                     variable=factor(names(temp_model_rw), levels=names(temp_model_rw)),
                     RMSE=temp_model_rw)

ggplot(step_model, aes(y=RMSE)) +
  geom_point(aes(x=variable)) +
  geom_line(aes(x=index)) +
  labs(title="Stepwise model selection(red wine)") +
  theme_minimal()
```



Consider the following fits and extract the best fit model:

- fit1 <- quality ~ alcohol + volatile acidity + sulphates + total sulfur dioxide + chlorides + pH
- fit2 <- quality ~ alcohol + volatile acidity + sulphates + total sulfur dioxide + chlorides + pH + citric acid
- fit3 <- quality ~ alcohol + volatile acidity + sulphates + total sulfur dioxide + chlorides + pH + citric acid + fixed acidity

Cross Validation

```
set.seed(2020)
partition_rw_train <- red_wine_std[-partition_rw$test$idx,]

redwine_cv <- crossv_kfold(partition_rw_train, 5)

redwine_cv
```

```
## # A tibble: 5 x 3
##   train          test          .id
##   <named list>    <named list>    <chr>
## 1 <resample [959 x 12]> <resample [240 x 12]> 1
## 2 <resample [959 x 12]> <resample [240 x 12]> 2
## 3 <resample [959 x 12]> <resample [240 x 12]> 3
## 4 <resample [959 x 12]> <resample [240 x 12]> 4
## 5 <resample [960 x 12]> <resample [239 x 12]> 5
```

#Calculating RMSE for each fold of data

```
cv_rw <- redwine_cv %>%
  mutate(fit = purrr::map(train,
    ~ lm(quality ~ alcohol + volatile.acidity + sulphates +
      total.sulfur.dioxide + chlorides + pH, data = .)),
    rmse = purrr::map2_dbl(fit, test, ~ rmse(.x, .y)))

cv_rw
```

```
## # A tibble: 5 x 5
##   train          test          .id  fit          rmse
##   <named list>    <named list>    <chr> <named list> <dbl>
## 1 <resample [959 x 12]> <resample [240 x 12]> 1    <lm>         0.829
## 2 <resample [959 x 12]> <resample [240 x 12]> 2    <lm>         0.778
## 3 <resample [959 x 12]> <resample [240 x 12]> 3    <lm>         0.815
## 4 <resample [959 x 12]> <resample [240 x 12]> 4    <lm>         0.811
## 5 <resample [960 x 12]> <resample [239 x 12]> 5    <lm>         0.770
```

#Average of RMSEs

```
mean(cv_rw$rmse)
```

```
## [1] 0.8005269
```

Comparing models using CV

```
do_redwine_cv <- function(formula) {
  redwine_cv %>%
    mutate(fit = map(train,
      ~ lm(formula, data = .)),
      rmse = map2_dbl(fit, test, ~ rmse(.x, .y))) %>%
    summarize(cv_rmse = mean(rmse)) %>%
    pull(cv_rmse)
}
```

Calling the function

```
fit1_rmse <- do_redwine_cv(quality ~ alcohol + volatile.acidity + sulphates +
  total.sulfur.dioxide + chlorides + pH )

fit2_rmse <- do_redwine_cv(quality ~ alcohol + volatile.acidity + sulphates +
  total.sulfur.dioxide + chlorides + pH + citric.acid)
```

```
fit3_rmse <- do_redwine_cv(quality ~ alcohol + volatile.acidity + sulphates +
  total.sulfur.dioxide + chlorides + pH + citric.acid + fixed.acidity)

fit1_rmse
```

```
## [1] 0.8005269
```

```
fit2_rmse
```

```
## [1] 0.8000535
```

```
fit3_rmse
```

```
## [1] 0.8022745
```

Best fit model:

- $\text{quality} \sim \text{alcohol} + \text{volatile.acidity} + \text{sulphates} + \text{total.sulfur.dioxide} + \text{chlorides} + \text{pH}$

Stepwise Model Selection for white wine

```
#Partition of data
white_wine_std <- data.frame(scale(white_wine_df))
set.seed(10)
partition_ww <- resample_partition(white_wine_std,
  p=c(train=0.5,
    valid=0.25,
    test=0.25))
```

```
model_ww <- NULL
```

```
preds <- "1"
cands <- c("fixed.acidity", "volatile.acidity", "citric.acid", "residual.sugar",
  "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density",
  "pH", "sulphates", "alcohol")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
##      fixed.acidity    volatile.acidity    citric.acid
##      1.0215179      1.0208275      1.0292701
##      residual.sugar    chlorides    free.sulfur.dioxide
##      1.0256987      1.0061436      1.0292602
## total.sulfur.dioxide    density    pH
##      1.0125231      0.9869927      1.0236879
##      sulphates    alcohol
##      1.0272717      0.9206781
## attr("best")
##      alcohol
## 0.9206781
```

```

preds <- "alcohol"
cands <- c("fixed.acidity","volatile.acidity","citric.acid","residual.sugar",
          "chlorides","free.sulfur.dioxide","total.sulfur.dioxide","density",
          "pH","sulphates")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2

```

```

##      fixed.acidity      volatile.acidity      citric.acid
##      0.9179510         0.9048555         0.9198725
##      residual.sugar      chlorides  free.sulfur.dioxide
##      0.9141193         0.9194475         0.9134829
## total.sulfur.dioxide      density      pH
##      0.9205878         0.9193175         0.9191888
##      sulphates
##      0.9179753
## attr("best")
## volatile.acidity
##      0.9048555

```

```

preds <- c("alcohol","volatile.acidity")
cands <- c("fixed.acidity","citric.acid","residual.sugar",
          "chlorides","free.sulfur.dioxide","total.sulfur.dioxide","density",
          "pH","sulphates")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2

```

```

##      fixed.acidity      citric.acid      residual.sugar
##      0.9017757         0.9058093         0.8921348
##      chlorides  free.sulfur.dioxide total.sulfur.dioxide
##      0.9042788         0.8983406         0.9035253
##      density      pH      sulphates
##      0.8999125         0.9039963         0.9026877
## attr("best")
## residual.sugar
##      0.8921348

```

```

preds <- c("alcohol","volatile.acidity","residual.sugar")
cands <- c("fixed.acidity","citric.acid",
          "chlorides","free.sulfur.dioxide","total.sulfur.dioxide","density",
          "pH","sulphates")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2

```

```

##      fixed.acidity      citric.acid      chlorides
##      0.8882746         0.8933910         0.8919934
## free.sulfur.dioxide total.sulfur.dioxide      density

```

```
##          0.8886907          0.8920658          0.8903775
##          pH          sulphates
##          0.8898555          0.8896877
## attr("best")
## fixed.acidity
##          0.8882746
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity")
cands <- c("citric.acid","chlorides","free.sulfur.dioxide",
          "total.sulfur.dioxide","density","pH","sulphates")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
##          citric.acid          chlorides  free.sulfur.dioxide
##          0.8892712          0.8881038          0.8854824
## total.sulfur.dioxide          density          pH
##          0.8881527          0.8889947          0.8876715
##          sulphates
##          0.8858385
## attr("best")
## free.sulfur.dioxide
##          0.8854824
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity",
          "free.sulfur.dioxide")
cands <- c("citric.acid","chlorides",
          "total.sulfur.dioxide","density","pH","sulphates")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
##          citric.acid          chlorides total.sulfur.dioxide
##          0.8864740          0.8852411          0.8849268
##          density          pH          sulphates
##          0.8865579          0.8849405          0.8832588
## attr("best")
## sulphates
## 0.8832588
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity",
          "free.sulfur.dioxide","sulphates")
cands <- c("citric.acid","chlorides",
          "total.sulfur.dioxide","density","pH")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
##          citric.acid          chlorides total.sulfur.dioxide
```

```
##          0.8843311          0.8830244          0.8822735
##          density          pH
##          0.8838580          0.8830231
## attr("best")
## total.sulfur.dioxide
##          0.8822735
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity",
           "free.sulfur.dioxide","sulphates", "total.sulfur.dioxide")
cands <- c("citric.acid","chlorides",
           "density","pH")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
## citric.acid  chlorides    density    pH
##  0.8832933  0.8820711  0.8835013  0.8818766
## attr("best")
##          pH
##  0.8818766
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity",
           "free.sulfur.dioxide","sulphates", "total.sulfur.dioxide","pH")
cands <- c("citric.acid","chlorides","density")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
## citric.acid  chlorides    density
##  0.8828781  0.8817222  0.8876671
## attr("best")
## chlorides
##  0.8817222
```

```
preds <- c("alcohol","volatile.acidity","residual.sugar","fixed.acidity",
           "free.sulfur.dioxide","sulphates", "total.sulfur.dioxide",
           "pH","chlorides")
cands <- c("citric.acid","density")
s2 <- step("quality", preds, cands, partition_ww)

model_ww <- c(model_ww, attr(s2, "best"))
s2
```

```
## citric.acid    density
##  0.8827247  0.8882242
## attr("best")
## citric.acid
##  0.8827247
```

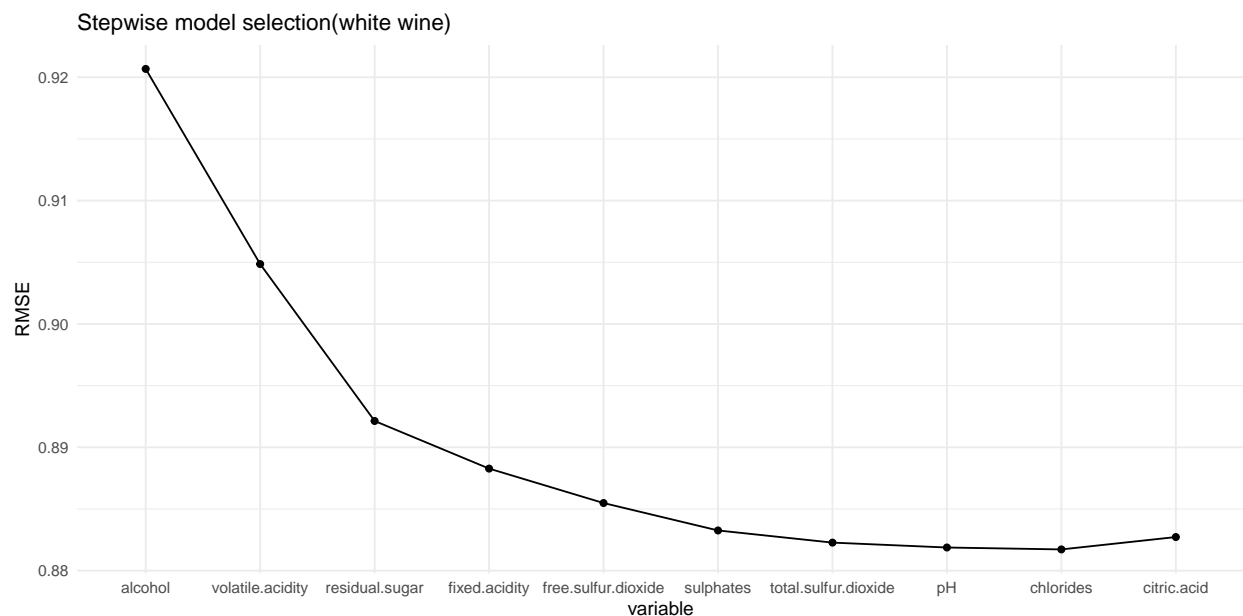
Model stopped improving at:

- $\text{fit} = \text{quality} \sim \text{alcohol} + \text{volatile acidity} + \text{residual sugar} + \text{fixed acidity} + \text{free sulfur dioxide} + \text{sulphates} + \text{total sulfur dioxide} + \text{pH} + \text{chlorides}$

Visualizing how adding each variable affects the RMSE.

```
step_model_ww <- tibble(index=seq_along(model_ww),
                        variable=factor(names(model_ww), levels=names(model_ww)),
                        RMSE=model_ww)

ggplot(step_model_ww, aes(y=RMSE)) +
  geom_point(aes(x=variable)) +
  geom_line(aes(x=index)) +
  labs(title="Stepwise model selection(white wine)") +
  theme_minimal()
```



visualizing how each variable(including the predictors) affects the RMSE for Red wine

```
temp_model_ww <- model_ww
preds <- c("alcohol", "volatile.acidity", "residual.sugar", "fixed.acidity",
          "free.sulfur.dioxide", "sulphates", "total.sulfur.dioxide",
          "pH", "chlorides", "citric.acid")
cands <- c("density")
s2 <- step("quality", preds, cands, partition_ww)

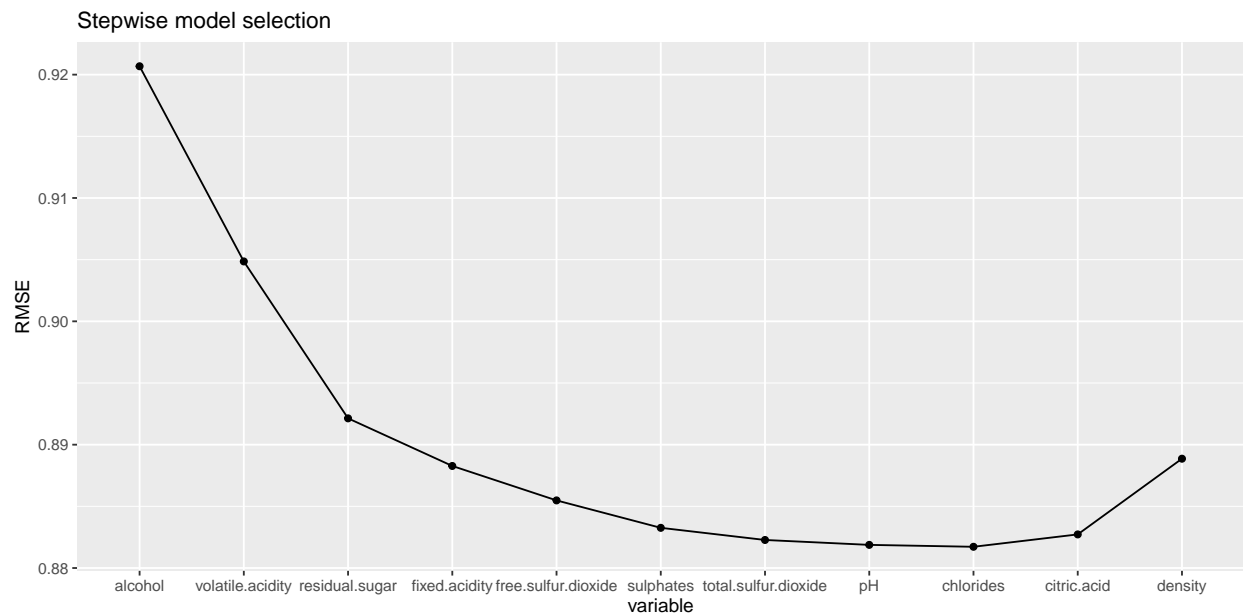
temp_model_ww <- c(temp_model_ww, attr(s2, "best"))
s2
```

```
## density
## 0.8888631
```

```
## attr("best")
## density
## 0.8888631
```

```
step_model_ww <- tibble(index=seq_along(temp_model_ww),
                        variable=factor(names(temp_model_ww), levels=names(temp_model_ww)),
                        RMSE=temp_model_ww)

ggplot(step_model_ww, aes(y=RMSE)) +
  geom_point(aes(x=variable)) +
  geom_line(aes(x=index)) +
  labs(title="Stepwise model selection")
```



Consider the following fits and extract the best fit model:

- fit1 <- quality ~ alcohol + volatile.acidity + residual.sugar + fixed.acidity + free.sulfur.dioxide + sulphates + total.sulfur.dioxide + pH
- fit2 <- quality ~ alcohol + volatile.acidity + residual.sugar + fixed.acidity + free.sulfur.dioxide + sulphates + total.sulfur.dioxide + pH + chlorides

Cross Validation

```
set.seed(2020)
partition_ww_train <- white_wine_std[-partition_ww$test$idx,]

whitewine_cv <- crossv_kfold(partition_ww_train, 5)

whitewine_cv
```

```
## # A tibble: 5 x 3
```



```
##   train                test                .id
##   <named list>         <named list>         <chr>
## 1 <resample [2,938 x 12]> <resample [735 x 12]> 1
## 2 <resample [2,938 x 12]> <resample [735 x 12]> 2
## 3 <resample [2,938 x 12]> <resample [735 x 12]> 3
## 4 <resample [2,939 x 12]> <resample [734 x 12]> 4
## 5 <resample [2,939 x 12]> <resample [734 x 12]> 5

#Calculating RMSE for each fold of data
cv_rw <- whitewine_cv %>%
  mutate(fit = purrr::map(train,
    ~ lm(quality ~ alcohol + volatile.acidity + residual.sugar +
      fixed.acidity + free.sulfur.dioxide +
      sulphates + total.sulfur.dioxide + pH, data = .)),
    rmse = purrr::map2_dbl(fit, test, ~ rmse(.x, .y)))

cv_rw
```

```
## # A tibble: 5 x 5
##   train                test                .id  fit          rmse
##   <named list>         <named list>         <chr> <named list> <dbl>
## 1 <resample [2,938 x 12]> <resample [735 x 12]> 1    <lm>         0.861
## 2 <resample [2,938 x 12]> <resample [735 x 12]> 2    <lm>         0.913
## 3 <resample [2,938 x 12]> <resample [735 x 12]> 3    <lm>         0.887
## 4 <resample [2,939 x 12]> <resample [734 x 12]> 4    <lm>         0.870
## 5 <resample [2,939 x 12]> <resample [734 x 12]> 5    <lm>         0.807
```

```
#Average of RMSEs
mean(cv_rw$rmse)
```

```
## [1] 0.8674873
```

Comparing models using CV

```
do_whitewine_cv <- function(formula) {
  whitewine_cv %>%
    mutate(fit = map(train,
      ~ lm(formula, data = .)),
      rmse = map2_dbl(fit, test, ~ rmse(.x, .y))) %>%
    summarize(cv_rmse = mean(rmse)) %>%
    pull(cv_rmse)
}
```

Calling the function

```
fit1_rmse_ww <- do_whitewine_cv(quality ~ alcohol + volatile.acidity + residual.sugar +
  fixed.acidity + free.sulfur.dioxide +
  sulphates + total.sulfur.dioxide + pH)

fit2_rmse_ww <- do_whitewine_cv(quality ~ alcohol + volatile.acidity + residual.sugar +
```

```
fixed.acidity + free.sulfur.dioxide +  
sulphates + total.sulfur.dioxide + pH + chlorides)
```

```
fit1_rmse_ww
```

```
## [1] 0.8674873
```

```
fit2_rmse_ww
```

```
## [1] 0.8675702
```

Best fit Model

- $\text{quality} \sim \text{alcohol} + \text{volatile.acidity} + \text{residual.sugar} + \text{fixed.acidity} + \text{free.sulfur.dioxide} + \text{sulphates} + \text{total.sulfur.dioxide} + \text{pH}$