1. Use read.csv to open both data sets and set them as df1 and df2. Pay attention to what the separator (sep) is.
   (Hint: use "," for comma delimited for a . csv file)
   Example: df1 <-read.csv("/Users/kamelliareshadi/Desktop/CSV/winequality-red.csv",sep=';')

```
df1 <- read.csv('C:/Users/shrey/Documents/winequality-red.csv',sep=';')
df1


df2 <- read.csv('C:/Users/shrey/Documents/winequality-white.csv',sep=';')
df2
```

2. Now add a label column to both df1 and df2 indicating a label 'red' or 'white'.

```
df1$label <- sapply(df1$pH,function(x){'red'})
df2$label <- sapply(df2$pH,function(x){'white'})
```

3. Check the head of df1 and df2

```
head(df1)
head(df2)
```

4. Combine df1 and df2 into a single data frame called wine.
   str(wine)

```
wine <- rbind(df1,df2)
str(wine)
```

5. Create a Histogram of residual sugar from the wine data. Color by red and white wines.

```
install.packages('ggplot2')
library(ggplot2)
pl <- ggplot(wine,aes(x=residual.sugar)) +
geom_histogram(aes(fill=label),color='black',bins=50)
pl
# Optional adding of fill colors
pl + scale_fill_manual(values = c('#ae4554','#faf7ea')) + theme_bw()
```

6. Create a Histogram of citric.acid from the wine data. Color by red and white wines.

```
pl <- ggplot(wine,aes(x=citric.acid)) +
geom_histogram(aes(fill=label),color='black',bins=50)
# Optional adding of fill colors
pl + scale_fill_manual(values = c('#ae4554','#faf7ea')) + theme_bw()
```

7. Create a Histogram of alcohol from the wine data. Color by red and white wines.

```
pl <- ggplot(wine,aes(x=alcohol)) +
geom_histogram(aes(fill=label),color='black',bins=50)
# Optional adding of fill colors
```

```
pl + scale_fill_manual(values = c('#ae4554','#faf7ea')) + theme_bw()
```

8. Create a scatterplot of residual.sugar versus citric.acid, color by red and white wine.

```
pl <- ggplot(wine,aes(x=citric.acid,y=residual.sugar)) +
geom_point(aes(color=label),alpha=0.2)
pl + scale_color_manual(values = c('#ae4554','#faf7ea')) +theme_dark()
```

9. Create a scatterplot of volatile.acidity versus residual.sugar, color by red and white wine.

```
pl <- ggplot(wine,aes(x=volatile.acidity,y=residual.sugar)) +
geom_point(aes(color=label),alpha=0.2)
pl + scale_color_manual(values = c('#ae4554','#faf7ea')) +theme_dark()
```

10. Grab the wine data without the column 'label' and call it clus.data, and check the head of clus.data

```
clus.data <- wine[,1:12]
head(clus.data)
```

11. Call the kmeans function on clus.data and assign the results to wine.cluster.

```
wine.cluster <- kmeans(wine[1:12],2)
wine.cluster
```

12. Print out the wine.cluster Cluster Means and explore the information.

```
print(wine.cluster$centers)
```

13. Use the table() function to compare your cluster results to the real results. Which is easier to correctly group, red or white wines?

```
table(wine$label,wine.cluster$cluster)
```