Wine.R

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(viridis)
```

```
## Loading required package: viridisLite
```

```r
library(hrbrthemes)


#set string values as characters
#load the greek characters
wine = read.csv("wine.csv", stringsAsFactors = FALSE, encoding = 'UTF-8')


#remove descriptive columns from dataset
wine = wine[,-c(1,3)]


#Which countries have the highest count of wines in the dataset?
topten <-wine %>% group_by(country) %>% summarize(count=n()) %>% arrange(desc(count))

View(topten)

toptencount <- data.frame(
  Country=c("US","Italy","France","Spain","Chile","Argentina","Portugal","Australia","New Zealan
d","Austria"),
  Count=c(62397,23478,21098,8258,5816,5631,5322,4957,3320,3057)
)

View(toptencount)

# Barplot
ggplot(toptencount, aes(x=Country, y=Count)) +
  geom_bar(stat = "identity",fill="blue",) +
  scale_fill_hue(c = 40) +
  theme(legend.position="none")
```
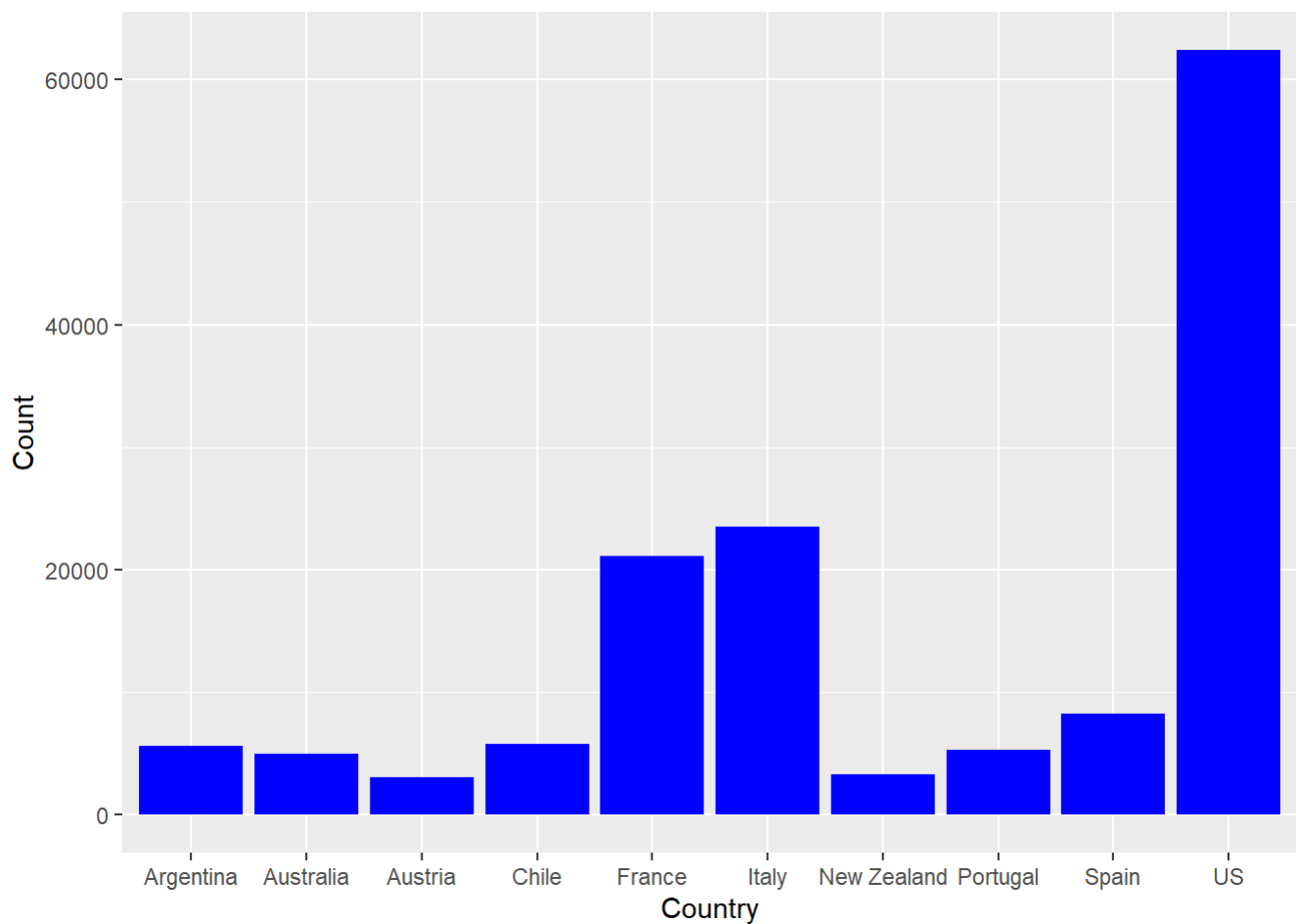
```
#What are the top 10 countries?
selected_countries = wine %>% group_by(country) %>% summarize(count=n()) %>% arrange(desc(count)) %>% top_n(10) %>% select(country)
```

```
## Selecting by count
```

```
selected_countries
```

```
## # A tibble: 10 x 1
##    country
##    <chr>
##  1 US
##  2 Italy
##  3 France
##  4 Spain
##  5 Chile
##  6 Argentina
##  7 Portugal
##  8 Australia
##  9 New Zealand
## 10 Austria
```

```
#changing the format from data fram to vector as.character referencing the country
selected_countries = as.character(selected_countries$country)
class(selected_countries)
```

```
## [1] "character"
```

```
#subsetting data selecting top ten countires and their points from wine
select_points = wine %>% filter(country %in% selected_countries) %>% select(country, points) %>%
arrange(country)


#What is the Distribution of the top 10 Wine producing countries?
#What countries have the highest rated wine
ggplot(select_points, aes(x = reorder(country,points,median), y = points)) +
  geom_boxplot(aes(fill = country)) +
  xlab("Country") +
  ylab("Points") +
  ggtitle("Distribution of Top 10 Countries") +
  theme(plot.title = element_text(hjust = 0.5))
```
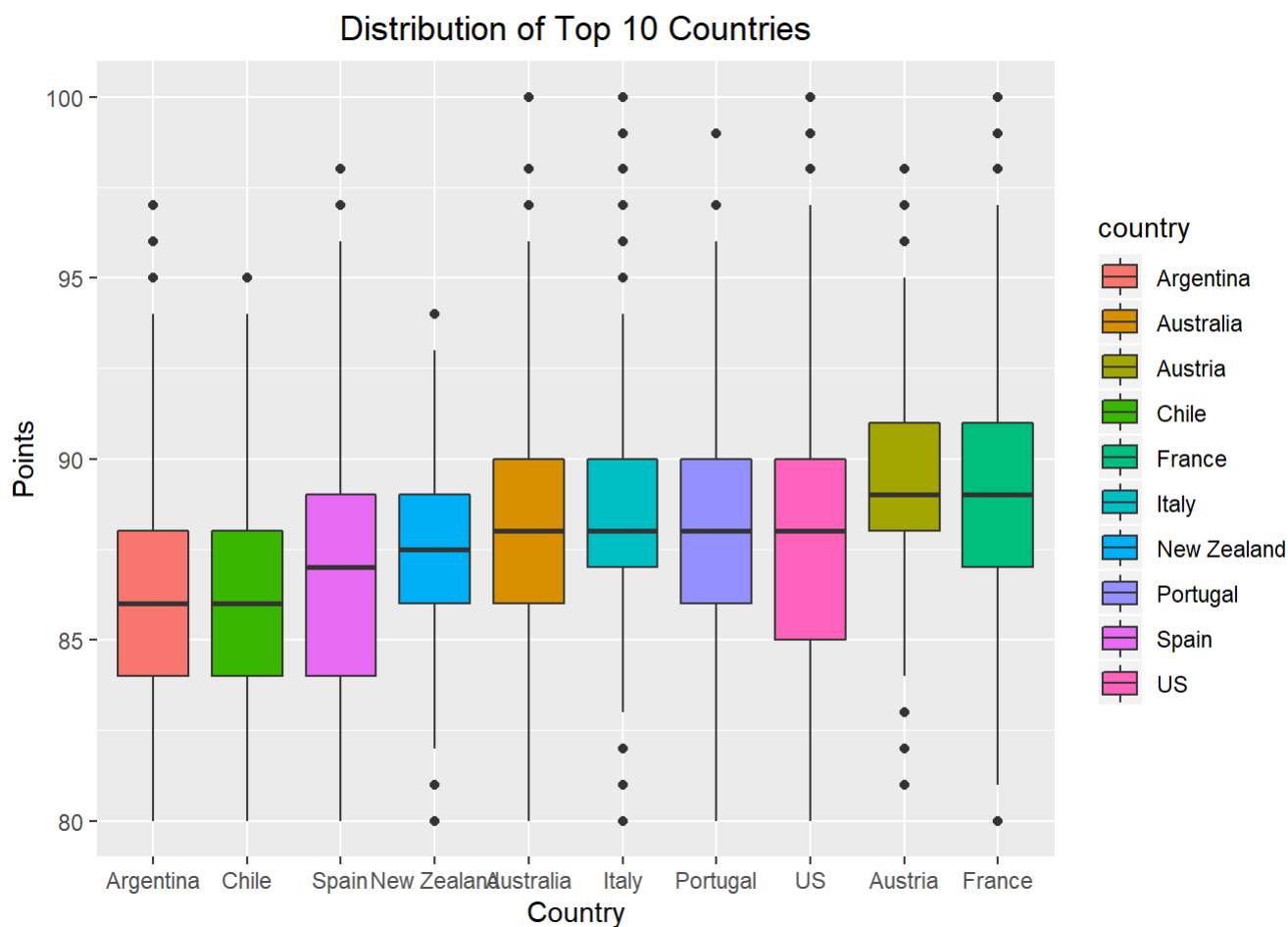
```r
#What is the best value wine?
#Use intersect to find top 15% cheapest wines with high rating
top15percent=wine %>%
  arrange(desc(points)) %>%
  filter(points > quantile(points, prob = 0.85))

cheapest15percent=wine %>%
  arrange(price) %>%
  head(nrow(top15percent))

goodvalue = intersect(top15percent,cheapest15percent)
goodvalue
```

```
##      country            designation points price          province
## 1     France         Aydie l'Origine     93    12 Southwest France
## 2         US                             93    12        Washington
## 3   Portugal     Seleção do Enólogo     92    11              Dão
## 4     France                             92    12          Bordeaux
## 5      Italy              Villachigi     92    13           Tuscany
## 6   Portugal            Dona Sophia     92    13              Tejo
## 7   Portugal Picos do Couto Reserva     92    11              Dão
## 8         US                             92    12            Oregon
## 9     France        Château Labrande     92    13 Southwest France
## 10  Portugal              Alvarinho     92    13             Minho
## 11        US                             92    11        Washington
## 12        US        Moscato d'Andrea     92    12        California
## 13        US                             92    11        Washington
## 14   Austria                  Andau     92    13        Burgenland
## 15  Portugal             Grand'Arte     92    13            Lisboa
## 16        US                             92    12        California
##               region_1           region_2               variety
## 1              Madiran                     Tannat-Cabernet Franc
## 2  Columbia Valley (WA)   Columbia Valley Johannisberg Riesling
## 3                                             Portuguese Red
## 4          Haut-Médoc                  Bordeaux-style Red Blend
## 5              Chianti                          Sangiovese
## 6                                             Portuguese Red
## 7                                             Portuguese Red
## 8     Willamette Valley Willamette Valley            Pinot Gris
## 9               Cahors                              Malbec
## 10                                              Alvarinho
## 11 Columbia Valley (WA)   Columbia Valley             Riesling
## 12         Napa Valley              Napa         Muscat Canelli
## 13 Columbia Valley (WA)   Columbia Valley             Riesling
## 14                                              Zweigelt
## 15                                         Touriga Nacional
## 16         Napa Valley              Napa        Sauvignon Blanc
##                   winery
## 1         Château d'Aydie
## 2          J. Bookwalter
## 3          Pedra Cancela
## 4  Château Devise d'Ardilley
## 5          Chigi Saracini
## 6   Quinta do Casal Branco
## 7        Quinta do Serrado
## 8                  Lujon
## 9         Jean-Luc Baldès
## 10               Aveleda
## 11           Pacific Rim
## 12         Robert Pecota
## 13              Bridgman
## 14          Scheiblhofer
## 15            DFJ Vinhos
## 16           Honker Blanc
```