DS 6306 Project 1

William Jones & Christian Orji

2023-02-19

## Project 1 for DS 6306

## # We are tasked with conducting a EDA for the Beers and Breweries Data Set that contains 2410 craft beers and 558 Breweries respectively.

## # Data documents

## ## importing the datasets and analyzing the data types

beers = read.csv("/Users/christianorji/Documents/DDS/Beers (1).csv")  
breweries = read.csv("/Users/christianorji/Documents/DDS/Breweries (1).csv")  
summary(beers)

## Name Beer\_ID ABV IBU   
## Length:2410 Min. : 1.0 Min. :0.00100 Min. : 4.00   
## Class :character 1st Qu.: 808.2 1st Qu.:0.05000 1st Qu.: 21.00   
## Mode :character Median :1453.5 Median :0.05600 Median : 35.00   
## Mean :1431.1 Mean :0.05977 Mean : 42.71   
## 3rd Qu.:2075.8 3rd Qu.:0.06700 3rd Qu.: 64.00   
## Max. :2692.0 Max. :0.12800 Max. :138.00   
## NA's :62 NA's :1005   
## Brewery\_id Style Ounces   
## Min. : 1.0 Length:2410 Min. : 8.40   
## 1st Qu.: 94.0 Class :character 1st Qu.:12.00   
## Median :206.0 Mode :character Median :12.00   
## Mean :232.7 Mean :13.59   
## 3rd Qu.:367.0 3rd Qu.:16.00   
## Max. :558.0 Max. :32.00   
##

summary(breweries)

## Brew\_ID Name City   
## Min. : 1.0 Length:558 Length:558   
## 1st Qu.:140.2 Class :character Class :character   
## Median :279.5 Mode :character Mode :character   
## Mean :279.5   
## 3rd Qu.:418.8   
## Max. :558.0   
## State   
## Length:558   
## Class :character   
## Mode :character   
##   
##   
##

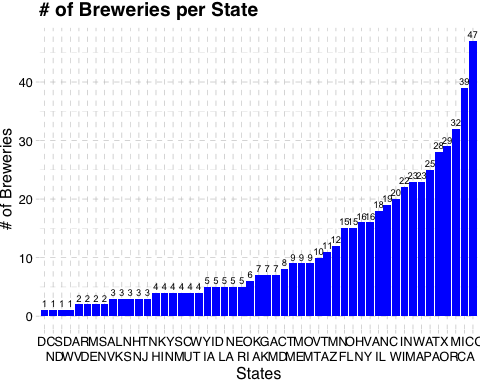
unique(beers$Style)

## [1] "American Pale Lager"   
## [2] "American Pale Ale (APA)"   
## [3] "American IPA"   
## [4] "American Double / Imperial IPA"   
## [5] "Oatmeal Stout"   
## [6] "American Porter"   
## [7] "Saison / Farmhouse Ale"   
## [8] "Belgian IPA"   
## [9] "Cider"   
## [10] "Baltic Porter"   
## [11] "Tripel"   
## [12] "American Barleywine"   
## [13] "Winter Warmer"   
## [14] "American Stout"   
## [15] "Fruit / Vegetable Beer"   
## [16] "English Strong Ale"   
## [17] "American Black Ale"   
## [18] "Belgian Dark Ale"   
## [19] "American Blonde Ale"   
## [20] "American Amber / Red Ale"   
## [21] "Berliner Weissbier"   
## [22] "American Brown Ale"   
## [23] "American Pale Wheat Ale"   
## [24] "Belgian Strong Dark Ale"   
## [25] "Kölsch"   
## [26] "English Pale Ale"   
## [27] "American Amber / Red Lager"   
## [28] "English Barleywine"   
## [29] "Milk / Sweet Stout"   
## [30] "German Pilsener"   
## [31] "Pumpkin Ale"   
## [32] "Belgian Pale Ale"   
## [33] "American Pilsner"   
## [34] "American Wild Ale"   
## [35] "English Brown Ale"   
## [36] "Altbier"   
## [37] "California Common / Steam Beer"   
## [38] "Gose"   
## [39] "Cream Ale"   
## [40] "Vienna Lager"   
## [41] "Witbier"   
## [42] "American Double / Imperial Stout"   
## [43] "Munich Helles Lager"   
## [44] "Schwarzbier"   
## [45] "Märzen / Oktoberfest"   
## [46] "Extra Special / Strong Bitter (ESB)"  
## [47] "Rye Beer"   
## [48] "Euro Dark Lager"   
## [49] "Hefeweizen"   
## [50] "Foreign / Export Stout"   
## [51] "Other"   
## [52] "English India Pale Ale (IPA)"   
## [53] "Czech Pilsener"   
## [54] "American Strong Ale"   
## [55] "Mead"   
## [56] "Euro Pale Lager"   
## [57] "American White IPA"   
## [58] "Dortmunder / Export Lager"   
## [59] "Irish Dry Stout"   
## [60] "Scotch Ale / Wee Heavy"   
## [61] "Munich Dunkel Lager"   
## [62] "Radler"   
## [63] "Bock"   
## [64] "English Dark Mild Ale"   
## [65] "Irish Red Ale"   
## [66] "Rauchbier"   
## [67] "Bière de Garde"   
## [68] "Doppelbock"   
## [69] "Dunkelweizen"   
## [70] "Belgian Strong Pale Ale"   
## [71] "Dubbel"   
## [72] "Quadrupel (Quad)"   
## [73] "Russian Imperial Stout"   
## [74] "English Pale Mild Ale"   
## [75] "Maibock / Helles Bock"   
## [76] "Herbed / Spiced Beer"   
## [77] "American Adjunct Lager"   
## [78] "Scottish Ale"   
## [79] ""   
## [80] "Smoked Beer"   
## [81] "Light Lager"   
## [82] "Abbey Single Ale"   
## [83] "Roggenbier"   
## [84] "Kristalweizen"   
## [85] "American Dark Wheat Ale"   
## [86] "English Stout"   
## [87] "Old Ale"   
## [88] "American Double / Imperial Pilsner"   
## [89] "Flanders Red Ale"   
## [90] "Keller Bier / Zwickel Bier"   
## [91] "American India Pale Lager"   
## [92] "Shandy"   
## [93] "Wheat Ale"   
## [94] "American Malt Liquor"   
## [95] "English Bitter"   
## [96] "Chile Beer"   
## [97] "Grisette"   
## [98] "Flanders Oud Bruin"   
## [99] "Braggot"   
## [100] "Low Alcohol Beer"

# Question 1

### how many breweries are present in each state?

#install.packages("ggthemes")  
library(ggplot2)  
library(ggthemes)  
library(tidyverse)  
breweries %>% group\_by(State) %>%  
 summarise(count = n\_distinct(Brew\_ID)) %>%  
 arrange(count) %>%  
 ggplot(aes(x= reorder(State, +count), y = count)) +  
 geom\_bar(stat="identity", fill ="blue") +  
 scale\_x\_discrete(guide = guide\_axis(n.dodge=2)) +   
 geom\_text(aes(label=count), vjust=-0.5, color="black",  
 position = position\_dodge(0.9), size=2.5) +  
 ylab('# of Breweries') +  
 xlab('States') +  
 ggtitle('# of Breweries per State') +  
 theme\_pander()



The output is a bar chart showing the number of breweries in each state of the US. The x-axis shows the states and is ordered in ascending order of the number of breweries, while the y-axis shows the count of breweries. The chart is colored in blue and has labels above each bar indicating the count of breweries. Overall, the chart is well-designed and easy to read, and it effectively communicates the information about the distribution of breweries across different states. The use of dodging for the labels and scaling the x-axis helps to prevent overlap and ensure that the chart is easy to read.

# Question 2

### Merging the beer and brewery datasets, both share the Brew/Brewery ID column and contain no missing values so they can be inner joined.

library(dplyr)  
#first change one of the datasets column name to match the other  
breweries <- breweries %>% rename("Brewery\_id"= "Brew\_ID")  
#Change Name of brewery to avoid duplicate column names  
breweries <- breweries %>% rename("Brewery" = "Name")  
data <- beers %>% inner\_join(breweries, by = "Brewery\_id")  
#printing of the first 6 rows  
head(data) #first 6 rows

## Name Beer\_ID ABV IBU Brewery\_id  
## 1 Pub Beer 1436 0.050 NA 409  
## 2 Devil's Cup 2265 0.066 NA 178  
## 3 Rise of the Phoenix 2264 0.071 NA 178  
## 4 Sinister 2263 0.090 NA 178  
## 5 Sex and Candy 2262 0.075 NA 178  
## 6 Black Exodus 2261 0.077 NA 178  
## Style Ounces Brewery City  
## 1 American Pale Lager 12 10 Barrel Brewing Company Bend  
## 2 American Pale Ale (APA) 12 18th Street Brewery Gary  
## 3 American IPA 12 18th Street Brewery Gary  
## 4 American Double / Imperial IPA 12 18th Street Brewery Gary  
## 5 American IPA 12 18th Street Brewery Gary  
## 6 Oatmeal Stout 12 18th Street Brewery Gary  
## State  
## 1 OR  
## 2 IN  
## 3 IN  
## 4 IN  
## 5 IN  
## 6 IN

tail(data) # last 6 rows

## Name Beer\_ID ABV IBU Brewery\_id  
## 2405 Rocky Mountain Oyster Stout 1035 0.075 NA 425  
## 2406 Belgorado 928 0.067 45 425  
## 2407 Rail Yard Ale 807 0.052 NA 425  
## 2408 B3K Black Lager 620 0.055 NA 425  
## 2409 Silverback Pale Ale 145 0.055 40 425  
## 2410 Rail Yard Ale (2009) 84 0.052 NA 425  
## Style Ounces Brewery City  
## 2405 American Stout 12 Wynkoop Brewing Company Denver  
## 2406 Belgian IPA 12 Wynkoop Brewing Company Denver  
## 2407 American Amber / Red Ale 12 Wynkoop Brewing Company Denver  
## 2408 Schwarzbier 12 Wynkoop Brewing Company Denver  
## 2409 American Pale Ale (APA) 12 Wynkoop Brewing Company Denver  
## 2410 American Amber / Red Ale 12 Wynkoop Brewing Company Denver  
## State  
## 2405 CO  
## 2406 CO  
## 2407 CO  
## 2408 CO  
## 2409 CO  
## 2410 CO

summary(data)

## Name Beer\_ID ABV IBU   
## Length:2410 Min. : 1.0 Min. :0.00100 Min. : 4.00   
## Class :character 1st Qu.: 808.2 1st Qu.:0.05000 1st Qu.: 21.00   
## Mode :character Median :1453.5 Median :0.05600 Median : 35.00   
## Mean :1431.1 Mean :0.05977 Mean : 42.71   
## 3rd Qu.:2075.8 3rd Qu.:0.06700 3rd Qu.: 64.00   
## Max. :2692.0 Max. :0.12800 Max. :138.00   
## NA's :62 NA's :1005   
## Brewery\_id Style Ounces Brewery   
## Min. : 1.0 Length:2410 Min. : 8.40 Length:2410   
## 1st Qu.: 94.0 Class :character 1st Qu.:12.00 Class :character   
## Median :206.0 Mode :character Median :12.00 Mode :character   
## Mean :232.7 Mean :13.59   
## 3rd Qu.:367.0 3rd Qu.:16.00   
## Max. :558.0 Max. :32.00   
##   
## City State   
## Length:2410 Length:2410   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##

#checking string columns for missing cell strings  
print(colSums(data == ""))

## Name Beer\_ID ABV IBU Brewery\_id Style   
## 0 0 NA NA 0 5   
## Ounces Brewery City State   
## 0 0 0 0

The code first renames the “Brew\_ID” column in the breweries dataset to “Brewery\_id” to match the corresponding column in the beers dataset. It also renames the “Name” column in the breweries dataset to “Brewery” to avoid duplicate column names.

Then it performs an inner join on the two datasets based on the “Brewery\_id” column. The resulting dataset is printed using head(), tail(), and summary() to show the first 6 rows, last 6 rows, and summary statistics of all columns, respectively.

Finally, colSums() is used to check for missing values in string columns. It prints the number of missing values for each string column.

# Question 3

### Address Missing Values #Missing data per column \* Name : 0 \* Beer\_ID

: 0 \* ABV : 62 \* Style : 5 \* IBU : 1005 \* Name : 0 \* Ounces : 0 \* Brewery : 0 \* City : 0 \* State : 0 #There was alot of data in the beer dataset, to alivate the missing data each was searched by hand using the websites untapped.com and beeradvocate.com # if data was found it was filled in using this method here is the number of missing data after search #Missing data per column \* Name : 0 \* Beer\_ID : 0 \* ABV : 18 \* Style : 3 \* IBU : 514 \* Name : 0 \* Ounces : 0 \* Brewery : 0 \* City : 0 \* State : 0

#commiting the file to csv   
write.csv(data, "/Users/christianorji/Documents/DDS/unfilteredData.csv")  
#after filling in mising data bringing back in now filtered data frame  
beer\_data = read.csv("/Users/christianorji/Documents/DDS/filteredData.csv")  
#summary after hand searching missing data  
summary(beer\_data)

## Name Beer\_ID ABV IBU   
## Length:2410 Min. : 1.0 Min. :0.00100 Min. : 4.00   
## Class :character 1st Qu.: 808.2 1st Qu.:0.05000 1st Qu.: 20.00   
## Mode :character Median :1453.5 Median :0.05600 Median : 35.00   
## Mean :1431.1 Mean :0.05972 Mean : 41.35   
## 3rd Qu.:2075.8 3rd Qu.:0.06700 3rd Qu.: 60.00   
## Max. :2692.0 Max. :0.12800 Max. :180.00   
## NA's :18 NA's :514   
## Brewery\_id Style Ounces Brewery   
## Min. : 1.0 Length:2410 Min. : 8.40 Length:2410   
## 1st Qu.: 94.0 Class :character 1st Qu.:12.00 Class :character   
## Median :206.0 Mode :character Median :12.00 Mode :character   
## Mean :232.7 Mean :13.59   
## 3rd Qu.:367.0 3rd Qu.:16.00   
## Max. :558.0 Max. :32.00   
##   
## City State   
## Length:2410 Length:2410   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##

print(colSums(beer\_data == ""))

## Name Beer\_ID ABV IBU Brewery\_id Style   
## 0 0 NA NA 0 3   
## Ounces Brewery City State   
## 0 0 0 0

# checking for duplicated beers  
dup = beer\_data[duplicated(beer\_data$Name), ]  
dup

## Name Beer\_ID ABV IBU Brewery\_id  
## 27 Hell or High Watermelon Wheat (2009) 801 0.049 17 369  
## 70 Citra Ass Down 2686 0.080 68 2  
## 82 The Brown Note 2674 0.050 20 2  
## 110 Barney Flats Oatmeal Stout 1221 0.057 13 172  
## 125 Angry Orchard Crisp Apple 1292 0.050 10 435  
## 221 Oberon 1989 0.058 10 77  
## 305 1327 Pod's ESB 886 0.056 37 381  
## 306 1327 Pod's ESB 612 0.056 37 381  
## 354 Firestarter India Pale Ale 633 0.066 72 108  
## 361 Boston Lager 1310 0.049 30 301  
## 364 Boston Lager 1143 0.049 30 301  
## 373 Plum St. Porter 1573 0.057 52 220  
## 460 East India Pale Ale 566 0.068 47 438  
## 486 Pale Alement 2054 0.055 40 25  
## 531 Beach Blonde 1927 0.049 10 268  
## 536 Lake House 1910 0.046 18 193  
## 559 White Zombie Ale 638 0.047 NA 332  
## 560 Firewater India Pale Ale 507 0.052 NA 332  
## 576 Main St. Virginia Ale 1248 0.050 40 123  
## 603 Oktoberfest 2011 0.055 25 142  
## 619 Maduro Brown Ale 571 0.055 25 142  
## 859 Sunbru K\xf6lsch 178 0.052 17 161  
## 899 Point the Way IPA 1296 0.059 60 241  
## 917 312 Urban Pale Ale 2202 0.054 30 89  
## 919 312 Urban Wheat Ale 1829 0.042 18 89  
## 975 Oktoberfest 539 0.059 25 366  
## 1036 Hapa Brown Ale 2180 0.064 36 205  
## 1077 Nonstop Hef Hop 2399 0.039 20 81  
## 1078 Nonstop Hef Hop 2398 0.039 20 81  
## 1079 Nonstop Hef Hop 2397 0.039 20 81  
## 1080 Nonstop Hef Hop 2396 0.039 20 81  
## 1081 Nonstop Hef Hop 2395 0.039 20 81  
## 1082 Nonstop Hef Hop 2394 0.039 20 81  
## 1083 Nonstop Hef Hop 2393 0.039 20 81  
## 1084 Nonstop Hef Hop 2392 0.039 20 81  
## 1085 Nonstop Hef Hop 2391 0.039 20 81  
## 1086 Nonstop Hef Hop 2390 0.039 20 81  
## 1087 Nonstop Hef Hop 2389 0.039 20 81  
## 1099 Oktoberfest 2383 0.060 23 121  
## 1183 Longboard Island Lager 1220 0.046 18 440  
## 1184 Longboard Island Lager 1070 0.046 18 440  
## 1185 Longboard Island Lager 590 0.046 18 440  
## 1255 #9 360 0.051 20 304  
## 1285 The Imperial Texan 1234 0.080 100 162  
## 1291 Our Legacy IPA 1135 0.065 60 300  
## 1321 Tsunami IPA 2185 0.072 75 204  
## 1340 Grazias 2663 0.063 30 4  
## 1367 Clear Creek Ice 2280 0.062 NA 135  
## 1369 Mountain Crest 2278 0.055 NA 135  
## 1374 Boxer 2273 0.050 NA 135  
## 1427 Narragansett Summer Ale 403 0.042 24 144  
## 1455 Shift 578 0.050 29 83  
## 1456 Ranger IPA 564 0.065 70 83  
## 1457 Fat Tire Amber Ale 563 0.052 18 83  
## 1517 Slow Ride 2069 0.048 35 185  
## 1558 One Nut Brown 1751 0.050 NA 167  
## 1561 Dale's Pale Ale 1252 0.065 65 167  
## 1564 oSKAr the G'Rauch 992 0.085 NA 167  
## 1565 Dale's Pale Ale 955 0.065 65 167  
## 1580 Mama's Little Yella Pils 7 0.053 35 167  
## 1584 Dale's Pale Ale 1 0.065 65 167  
## 1600 Deviant Dale's IPA 1328 0.080 85 390  
## 1601 Old Chub 1175 0.080 35 390  
## 1602 Dale's Pale Ale 1166 0.065 65 390  
## 1603 Dale's Pale Ale 1065 0.065 65 390  
## 1676 Morning Wood Wheat (Current) 130 0.059 14 136  
## 1697 Vanilla Porter 2513 0.047 25 69  
## 1701 Long Hammer IPA 583 0.065 44 488  
## 1715 Rye Porter 710 0.070 65 387  
## 1719 Colorado Red Ale 220 0.062 45 387  
## 1720 Miner's Gold 219 0.058 10 387  
## 1738 Gose 2580 0.035 NA 43  
## 1745 Oktoberfest 2373 0.062 20 44  
## 1766 Lazy Daze Lager 97 0.055 10 230  
## 1781 CAPT Black IPA 883 0.073 55 398  
## 1790 Oktoberfest 1995 0.055 20 31  
## 1852 Torpedo 426 0.072 65 84  
## 1853 Sierra Nevada Pale Ale 400 0.056 37 84  
## 1893 Modus Hoperandi 67 0.068 65 265  
## 1914 Alter Ego 2405 0.062 33 116  
## 1925 Dagger Falls IPA 1595 0.063 100 310  
## 1938 Dagger Falls IPA 879 0.063 100 310  
## 1953 Suzy B Dirty Blonde Ale 1118 0.050 20 134  
## 1954 Devil's Harvest Extra Pale Ale 1117 0.058 60 134  
## 1967 Pine Belt Pale Ale 45 0.065 45 119  
## 1968 Bombshell Blonde 44 0.050 20 119  
## 1977 Bitter Biker Double IPA 1821 0.096 88 176  
## 1984 Special Amber 445 0.050 22 499  
## 1990 Northern Lights India Pale Ale 368 0.065 52 384  
## 1998 Point Special 2228 0.047 9 132  
## 2007 Point Nude Beach Summer Wheat 816 0.050 7 132  
## 2011 Point Oktoberfest 456 0.057 15 132  
## 2016 Point Special Lager 138 0.047 9 132  
## 2019 Brontide 1631 0.050 NA 256  
## 2057 The Deuce 934 0.070 85 26  
## 2062 Chaka 660 0.080 NA 26  
## 2075 BLAKKR 1739 0.099 85 62  
## 2099 TailGate IPA 662 0.050 44 450  
## 2102 Blacktop Blonde 362 0.050 19 450  
## 2103 TailGate Hefeweizen 361 0.049 28 450  
## Style Ounces  
## 27 Fruit / Vegetable Beer 12.0  
## 70 American Double / Imperial IPA 16.0  
## 82 English Brown Ale 16.0  
## 110 Oatmeal Stout 12.0  
## 125 Cider 12.0  
## 221 American Pale Wheat Ale 16.0  
## 305 Extra Special / Strong Bitter (ESB) 12.0  
## 306 Extra Special / Strong Bitter (ESB) 12.0  
## 354 American IPA 16.0  
## 361 Vienna Lager 12.0  
## 364 Vienna Lager 12.0  
## 373 American Porter 12.0  
## 460 English India Pale Ale (IPA) 12.0  
## 486 American Pale Ale (APA) 16.0  
## 531 American Blonde Ale 16.0  
## 536 Munich Helles Lager 16.0  
## 559 Witbier 12.0  
## 560 American IPA 12.0  
## 576 Altbier 12.0  
## 603 M\xe4rzen / Oktoberfest 12.0  
## 619 American Brown Ale 12.0  
## 859 K\xf6lsch 12.0  
## 899 American IPA 12.0  
## 917 American Pale Ale (APA) 12.0  
## 919 American Pale Wheat Ale 12.0  
## 975 M\xe4rzen / Oktoberfest 16.0  
## 1036 American Brown Ale 12.0  
## 1077 American Pale Wheat Ale 16.0  
## 1078 American Pale Wheat Ale 16.0  
## 1079 American Pale Wheat Ale 16.0  
## 1080 American Pale Wheat Ale 16.0  
## 1081 American Pale Wheat Ale 16.0  
## 1082 American Pale Wheat Ale 16.0  
## 1083 American Pale Wheat Ale 16.0  
## 1084 American Pale Wheat Ale 16.0  
## 1085 American Pale Wheat Ale 16.0  
## 1086 American Pale Wheat Ale 16.0  
## 1087 American Pale Wheat Ale 16.0  
## 1099 M\xe4rzen / Oktoberfest 16.0  
## 1183 American Amber / Red Lager 16.0  
## 1184 American Amber / Red Lager 12.0  
## 1185 American Amber / Red Lager 12.0  
## 1255 Fruit / Vegetable Beer 12.0  
## 1285 American Double / Imperial IPA 12.0  
## 1291 American IPA 16.0  
## 1321 American IPA 12.0  
## 1340 Cream Ale 16.0  
## 1367 American Pale Lager 12.0  
## 1369 American Pale Lager 12.0  
## 1374 American Adjunct Lager 12.0  
## 1427 American Pale Wheat Ale 16.0  
## 1455 American Pale Lager 16.0  
## 1456 American IPA 16.0  
## 1457 American Amber / Red Ale 16.0  
## 1517 American Pale Ale (APA) 12.0  
## 1558 English Brown Ale 12.0  
## 1561 American Pale Ale (APA) 12.0  
## 1564 American IPA 16.0  
## 1565 American Pale Ale (APA) 19.2  
## 1580 Czech Pilsener 12.0  
## 1584 American Pale Ale (APA) 12.0  
## 1600 American Double / Imperial IPA 16.0  
## 1601 Scottish Ale 12.0  
## 1602 American Pale Ale (APA) 19.2  
## 1603 American Pale Ale (APA) 12.0  
## 1676 American Pale Wheat Ale 12.0  
## 1697 American Porter 16.0  
## 1701 American IPA 12.0  
## 1715 American Porter 16.0  
## 1719 American Amber / Red Ale 12.0  
## 1720 American Blonde Ale 12.0  
## 1738 Gose 16.0  
## 1745 M\xe4rzen / Oktoberfest 12.0  
## 1766 American Adjunct Lager 12.0  
## 1781 American Black Ale 16.0  
## 1790 M\xe4rzen / Oktoberfest 12.0  
## 1852 American IPA 16.0  
## 1853 American Pale Ale (APA) 12.0  
## 1893 American IPA 12.0  
## 1914 Saison / Farmhouse Ale 12.0  
## 1925 American IPA 12.0  
## 1938 American IPA 12.0  
## 1953 American Blonde Ale 16.0  
## 1954 American Pale Ale (APA) 16.0  
## 1967 American Pale Ale (APA) 16.0  
## 1968 American Blonde Ale 12.0  
## 1977 American Double / Imperial IPA 12.0  
## 1984 Vienna Lager 12.0  
## 1990 American IPA 12.0  
## 1998 American Adjunct Lager 12.0  
## 2007 American Pale Wheat Ale 12.0  
## 2011 M\xe4rzen / Oktoberfest 12.0  
## 2016 American Adjunct Lager 12.0  
## 2019 American Black Ale 12.0  
## 2057 American Brown Ale 16.0  
## 2062 Belgian Strong Pale Ale 16.0  
## 2075 American Black Ale 16.0  
## 2099 American IPA 12.0  
## 2102 American Blonde Ale 12.0  
## 2103 Hefeweizen 12.0  
## Brewery City State  
## 27 21st Amendment Brewery San Francisco CA  
## 70 Against the Grain Brewery Louisville KY  
## 82 Against the Grain Brewery Louisville KY  
## 110 Anderson Valley Brewing Company Boonville CA  
## 125 Angry Orchard Cider Company Cincinnati OH  
## 221 Bell's Brewery Kalamazoo MI  
## 305 Blue Blood Brewing Company Lincoln NE  
## 306 Blue Blood Brewing Company Lincoln NE  
## 354 Bonfire Brewing Company Eagle CO  
## 361 Boston Beer Company Boston MA  
## 364 Boston Beer Company Boston MA  
## 373 Bozeman Brewing Company Bozeman MT  
## 460 Brooklyn Brewery Brooklyn NY  
## 486 Burn 'Em Brewing Michigan City IN  
## 531 Cape Cod Beer Hyannis MA  
## 536 Capital Brewery Middleton WI  
## 559 Catawba Brewing Company Morganton NC  
## 560 Catawba Brewing Company Morganton NC  
## 576 Center of the Universe Brewing C... Ashland VA  
## 603 Cigar City Brewing Company Tampa FL  
## 619 Cigar City Brewing Company Tampa FL  
## 859 Four Peaks Brewing Company Tempe AZ  
## 899 Golden Road Brewing Los Angeles CA  
## 917 Goose Island Brewing Company Chicago IL  
## 919 Goose Island Brewing Company Chicago IL  
## 975 Great River Brewery Davenport IA  
## 1036 Hawai'i Nui Brewing Co. Hilo HI  
## 1077 Hopworks Urban Brewery Portland OR  
## 1078 Hopworks Urban Brewery Portland OR  
## 1079 Hopworks Urban Brewery Portland OR  
## 1080 Hopworks Urban Brewery Portland OR  
## 1081 Hopworks Urban Brewery Portland OR  
## 1082 Hopworks Urban Brewery Portland OR  
## 1083 Hopworks Urban Brewery Portland OR  
## 1084 Hopworks Urban Brewery Portland OR  
## 1085 Hopworks Urban Brewery Portland OR  
## 1086 Hopworks Urban Brewery Portland OR  
## 1087 Hopworks Urban Brewery Portland OR  
## 1099 Horny Goat Brew Pub Milwaukee WI  
## 1183 Kona Brewing Company Kona HI  
## 1184 Kona Brewing Company Kona HI  
## 1185 Kona Brewing Company Kona HI  
## 1255 Magic Hat Brewing Company South Burlington VT  
## 1285 Martin House Brewing Company Fort Worth TX  
## 1291 Matt Brewing Company Utica NY  
## 1321 Mehana Brewing Co. Hilo HI  
## 1340 Mike Hess Brewing Company San Diego CA  
## 1367 Minhas Craft Brewery Monroe WI  
## 1369 Minhas Craft Brewery Monroe WI  
## 1374 Minhas Craft Brewery Monroe WI  
## 1427 Narragansett Brewing Company Providence RI  
## 1455 New Belgium Brewing Company Fort Collins CO  
## 1456 New Belgium Brewing Company Fort Collins CO  
## 1457 New Belgium Brewing Company Fort Collins CO  
## 1517 Oasis Texas Brewing Company Austin TX  
## 1558 Oskar Blues Brewery Longmont CO  
## 1561 Oskar Blues Brewery Longmont CO  
## 1564 Oskar Blues Brewery Longmont CO  
## 1565 Oskar Blues Brewery Longmont CO  
## 1580 Oskar Blues Brewery Longmont CO  
## 1584 Oskar Blues Brewery Longmont CO  
## 1600 Oskar Blues Brewery (North Carol... Brevard NC  
## 1601 Oskar Blues Brewery (North Carol... Brevard NC  
## 1602 Oskar Blues Brewery (North Carol... Brevard NC  
## 1603 Oskar Blues Brewery (North Carol... Brevard NC  
## 1676 Pug Ryan's Brewery Dillon CO  
## 1697 Red Shedman Farm Brewery and Hop... Mt. Airy MD  
## 1701 Redhook Brewery Woodinville WA  
## 1715 Revolution Brewing Paonia CO  
## 1719 Revolution Brewing Paonia CO  
## 1720 Revolution Brewing Paonia CO  
## 1738 Rising Tide Brewing Company Portland ME  
## 1745 Rivertowne Brewing Company Export PA  
## 1766 Rochester Mills Brewing Company Rochester MI  
## 1781 Ruhstaller Beer Company Sacramento CA  
## 1790 SanTan Brewing Company Chandler AZ  
## 1852 Sierra Nevada Brewing Company Chico CA  
## 1853 Sierra Nevada Brewing Company Chico CA  
## 1893 Ska Brewing Company Durango CO  
## 1914 Smartmouth Brewing Company Norfolk VA  
## 1925 Sockeye Brewing Company Boise ID  
## 1938 Sockeye Brewing Company Boise ID  
## 1953 Southern Prohibition Brewing Com... Hattiesburg MS  
## 1954 Southern Prohibition Brewing Com... Hattiesburg MS  
## 1967 Southern Star Brewing Company Conroe TX  
## 1968 Southern Star Brewing Company Conroe TX  
## 1977 Spiteful Brewing Company Chicago IL  
## 1984 Sprecher Brewing Company Glendale WI  
## 1990 Starr Hill Brewery Crozet VA  
## 1998 Stevens Point Brewery Stevens Point WI  
## 2007 Stevens Point Brewery Stevens Point WI  
## 2011 Stevens Point Brewery Stevens Point WI  
## 2016 Stevens Point Brewery Stevens Point WI  
## 2019 Stillwater Artisanal Ales Baltimore MD  
## 2057 Sun King Brewing Company Indianapolis IN  
## 2062 Sun King Brewing Company Indianapolis IN  
## 2075 Surly Brewing Company Brooklyn Center MN  
## 2099 TailGate Beer San Diego CA  
## 2102 TailGate Beer San Diego CA  
## 2103 TailGate Beer San Diego CA  
## [ reached 'max' / getOption("max.print") -- omitted 6 rows ]

sprintf("There are a total of %d duplicated rows consisting of %d different beers", sum(duplicated(beer\_data$Name)),length(unique(dup$Name)))

## [1] "There are a total of 106 duplicated rows consisting of 83 different beers"

The output shows that the file has been successfully written to CSV and the data has been read back in as beer\_data. The summary() function shows a summary of the variables in the dataset, including the count, mean, median, and quartiles of numeric variables. The print(colSums(beer\_data == ““)) command shows that there are no more missing values in the dataset. The duplicated() function is then used to check for duplicated rows based on the beer name variable. The dup variable shows the duplicated rows, and the sprintf() function is used to print a message indicating the total number of duplicated rows and the number of unique duplicated beers.

### To remove these duplicated beers we will run the following function

beer\_data <- beer\_data[!duplicated(beer\_data$Name),]  
beer\_data

## Name Beer\_ID ABV IBU  
## 1 Pub Beer 1436 0.050 18  
## 2 Devil's Cup 2265 0.066 NA  
## 3 Rise of the Phoenix 2264 0.071 61  
## 4 Sinister 2263 0.090 30  
## 5 Sex and Candy 2262 0.075 NA  
## 6 Black Exodus 2261 0.077 NA  
## 7 Lake Street Express 2260 0.045 NA  
## 8 Foreman 2259 0.065 NA  
## 9 Jade 2258 0.055 NA  
## 10 Cone Crusher 2131 0.086 NA  
## 11 Sophomoric Saison 2099 0.072 NA  
## 12 Regional Ring Of Fire 2098 0.073 NA  
## 13 Garce SelÃ© 2097 0.069 NA  
## 14 Troll Destroyer 1980 0.085 NA  
## 15 Bitter Bitch 1979 0.061 60  
## 16 Ginja Ninja 2318 0.060 NA  
## 17 Cherried Away 2170 0.060 NA  
## 18 Rhubarbarian 2169 0.060 NA  
## 19 BrightCider 1502 0.060 NA  
## 20 He Said Baltic-Style Porter 1593 0.082 NA  
## 21 He Said Belgian-Style Tripel 1592 0.082 NA  
## 22 Lower De Boom 1036 0.099 92  
## 23 Fireside Chat 1024 0.079 45  
## 24 Marooned On Hog Island 976 0.079 NA  
## 25 Bitter American 876 0.044 42  
## 26 Hell or High Watermelon Wheat (2009) 802 0.049 17  
## 28 21st Amendment Watermelon Wheat Beer (2006) 800 0.049 17  
## 29 21st Amendment IPA (2006) 799 0.070 70  
## 30 Brew Free! or Die IPA (2008) 797 0.070 70  
## 31 Brew Free! or Die IPA (2009) 796 0.070 70  
## 32 Special Edition: Allies Win The War! 531 0.085 52  
## 33 Hop Crisis 432 0.097 94  
## 34 Bitter American (2011) 353 0.044 42  
## 35 Fireside Chat (2010) 321 0.079 45  
## 36 Back in Black 173 0.068 65  
## 37 Monk's Blood 11 0.083 35  
## 38 Brew Free! or Die IPA 10 0.070 65  
## 39 Hell or High Watermelon Wheat 9 0.049 17  
## 40 Bimini Twist 2519 0.070 82  
## 41 Beach Blonde 2518 0.050 23  
## 42 Rod Bender Red 2517 0.059 28  
## 43 Passion Fruit Prussia 2545 0.035 11  
## 44 Send Help 2544 0.045 18  
## 45 Cast Iron Oatmeal Brown 2324 0.055 NA  
## 46 Reprise Centennial Red 2288 0.060 38  
## 47 Alter Ego 2287 0.055 68  
## 48 Divided Sky 2286 0.065 NA  
## 49 Resurrected 2285 0.065 NA  
## 50 Contact High 1870 0.050 28  
## 51 Galaxyfest 2603 0.065 38  
## 52 Citrafest 2602 0.050 45  
## 53 Barn Yeti 2220 0.090 25  
## 54 Scarecrow 2219 0.069 65  
## 55 Ironman 2218 0.090 50  
## 56 Honey Kolsch 2217 0.046 15  
## 57 Copperhead Amber 2216 0.052 18  
## 58 Rude Parrot IPA 972 0.059 75  
## 59 British Pale Ale (2010) 866 0.054 30  
## 60 British Pale Ale 48 0.054 30  
## 61 Ballz Deep Double IPA 47 0.084 82  
## 62 Wolfman's Berliner 1583 0.038 NA  
## 63 Colorado Native 1165 0.055 26  
## 64 Colorado Native (2011) 431 0.055 26  
## 65 Jockamo IPA 516 0.065 52  
## 66 Purple Haze 515 0.042 13  
## 67 Abita Amber 514 0.045 17  
## 68 Citra Ass Down 2540 0.082 68  
## 69 The Brown Note 2539 0.050 20  
## 71 London Balling 2685 0.125 80  
## 72 35 K 2684 0.077 25  
## 73 A Beer 2683 0.042 42  
## 74 Rules are Rules 2682 0.050 25  
## 75 Flesh Gourd'n 2681 0.066 21  
## 76 Sho'nuff 2680 0.040 13  
## 77 Bloody Show 2679 0.055 17  
## 78 Rico Sauvin 2678 0.076 68  
## 79 Coq de la Marche 2677 0.051 38  
## 80 Kamen Knuddeln 2676 0.065 NA  
## 81 Pile of Face 2675 0.060 65  
## 83 Maylani's Coconut Stout 1594 0.053 35  
## 84 Oatmeal PSA 1162 0.050 35  
## 85 Pre Flight Pilsner 1137 0.052 33  
## 86 P-Town Pilsner 2403 0.040 20  
## 87 Klickitat Pale Ale 2402 0.053 36  
## 88 Yellow Wolf Imperial IPA 2401 0.082 103  
## 89 Freeride APA 1921 0.053 40  
## 90 Alaskan Amber 1920 0.053 18  
## 91 Hopalicious 2501 0.057 52  
## 92 Kentucky K\xf6lsch 1535 0.043 14  
## 93 Kentucky IPA 1149 0.065 45  
## 94 Dusty Trail Pale Ale 1474 0.054 40  
## 95 Damnesia 1473 0.062 70  
## 96 Desolation IPA 837 0.062 43  
## 97 Liberty Ale 2592 0.059 47  
## 98 IPA 2578 0.065 NA  
## 99 Summer Wheat 2577 0.045 NA  
## 100 California Lager 2103 0.049 19  
## 101 Brotherhood Steam 2102 0.056 NA  
## 102 Blood Orange Gose 2291 0.042 12  
## 103 Keebarlin' Pale Ale 1818 0.042 40  
## Brewery\_id Style Ounces  
## 1 409 American Pale Lager 12.0  
## 2 178 American Pale Ale (APA) 12.0  
## 3 178 American IPA 12.0  
## 4 178 American Double / Imperial IPA 12.0  
## 5 178 American IPA 12.0  
## 6 178 Oatmeal Stout 12.0  
## 7 178 American Pale Ale (APA) 12.0  
## 8 178 American Porter 12.0  
## 9 178 American Pale Ale (APA) 12.0  
## 10 178 American Double / Imperial IPA 12.0  
## 11 178 Saison / Farmhouse Ale 12.0  
## 12 178 Saison / Farmhouse Ale 12.0  
## 13 178 Saison / Farmhouse Ale 12.0  
## 14 178 Belgian IPA 12.0  
## 15 178 American Pale Ale (APA) 12.0  
## 16 155 Cider 12.0  
## 17 155 Cider 12.0  
## 18 155 Cider 12.0  
## 19 155 Cider 12.0  
## 20 369 Baltic Porter 12.0  
## 21 369 Tripel 12.0  
## 22 369 American Barleywine 8.4  
## 23 369 Winter Warmer 12.0  
## 24 369 American Stout 12.0  
## 25 369 American Pale Ale (APA) 12.0  
## 26 369 Fruit / Vegetable Beer 12.0  
## 28 369 Fruit / Vegetable Beer 12.0  
## 29 369 American IPA 12.0  
## 30 369 American IPA 12.0  
## 31 369 American IPA 12.0  
## 32 369 English Strong Ale 12.0  
## 33 369 American Double / Imperial IPA 12.0  
## 34 369 American Pale Ale (APA) 12.0  
## 35 369 Winter Warmer 12.0  
## 36 369 American Black Ale 12.0  
## 37 369 Belgian Dark Ale 12.0  
## 38 369 American IPA 12.0  
## 39 369 Fruit / Vegetable Beer 12.0  
## 40 68 American IPA 12.0  
## 41 68 American Blonde Ale 12.0  
## 42 68 American Amber / Red Ale 12.0  
## 43 61 Berliner Weissbier 12.0  
## 44 61 American Blonde Ale 12.0  
## 45 61 American Brown Ale 12.0  
## 46 61 American Amber / Red Ale 12.0  
## 47 61 American Black Ale 12.0  
## 48 61 American IPA 12.0  
## 49 61 American IPA 12.0  
## 50 61 American Pale Wheat Ale 12.0  
## 51 28 American IPA 16.0  
## 52 28 American IPA 16.0  
## 53 28 Belgian Strong Dark Ale 16.0  
## 54 28 American IPA 16.0  
## 55 28 English Strong Ale 16.0  
## 56 28 K\xf6lsch 16.0  
## 57 28 Belgian Dark Ale 16.0  
## 58 482 American IPA 16.0  
## 59 482 English Pale Ale 16.0  
## 60 482 English Pale Ale 16.0  
## 61 482 American Double / Imperial IPA 16.0  
## 62 374 Berliner Weissbier 12.0  
## 63 463 American Amber / Red Lager 12.0  
## 64 463 American Amber / Red Lager 12.0  
## 65 534 American IPA 12.0  
## 66 534 Fruit / Vegetable Beer 12.0  
## 67 534 American Amber / Red Lager 12.0  
## 68 63 American IPA 16.0  
## 69 63 American Brown Ale 16.0  
## 71 2 English Barleywine 16.0  
## 72 2 Milk / Sweet Stout 16.0  
## 73 2 American Pale Ale (APA) 16.0  
## 74 2 German Pilsener 16.0  
## 75 2 Pumpkin Ale 16.0  
## 76 2 Belgian Pale Ale 16.0  
## 77 2 American Pilsner 16.0  
## 78 2 American Double / Imperial IPA 16.0  
## 79 2 Saison / Farmhouse Ale 16.0  
## 80 2 American Wild Ale 16.0  
## 81 2 American IPA 16.0  
## 83 368 American Stout 16.0  
## 84 368 American Pale Ale (APA) 16.0  
## 85 368 American Pilsner 16.0  
## 86 118 American Pilsner 12.0  
## 87 118 American Pale Ale (APA) 12.0  
## 88 118 American Double / Imperial IPA 12.0  
## 89 271 American Pale Ale (APA) 12.0  
## 90 271 Altbier 12.0  
## 91 74 American Pale Ale (APA) 12.0  
## 92 389 K\xf6lsch 16.0  
## 93 389 American IPA 16.0  
## 94 402 American Pale Ale (APA) 16.0  
## 95 402 American IPA 16.0  
## 96 402 American IPA 16.0  
## 97 36 American IPA 12.0  
## 98 36 American IPA 12.0  
## 99 36 American Pale Wheat Ale 12.0  
## 100 36 American Amber / Red Lager 12.0  
## 101 36 California Common / Steam Beer 12.0  
## 102 172 Gose 12.0  
## 103 172 American Pale Ale (APA) 12.0  
## Brewery City State  
## 1 10 Barrel Brewing Company Bend OR  
## 2 18th Street Brewery Gary IN  
## 3 18th Street Brewery Gary IN  
## 4 18th Street Brewery Gary IN  
## 5 18th Street Brewery Gary IN  
## 6 18th Street Brewery Gary IN  
## 7 18th Street Brewery Gary IN  
## 8 18th Street Brewery Gary IN  
## 9 18th Street Brewery Gary IN  
## 10 18th Street Brewery Gary IN  
## 11 18th Street Brewery Gary IN  
## 12 18th Street Brewery Gary IN  
## 13 18th Street Brewery Gary IN  
## 14 18th Street Brewery Gary IN  
## 15 18th Street Brewery Gary IN  
## 16 2 Towns Ciderhouse Corvallis OR  
## 17 2 Towns Ciderhouse Corvallis OR  
## 18 2 Towns Ciderhouse Corvallis OR  
## 19 2 Towns Ciderhouse Corvallis OR  
## 20 21st Amendment Brewery San Francisco CA  
## 21 21st Amendment Brewery San Francisco CA  
## 22 21st Amendment Brewery San Francisco CA  
## 23 21st Amendment Brewery San Francisco CA  
## 24 21st Amendment Brewery San Francisco CA  
## 25 21st Amendment Brewery San Francisco CA  
## 26 21st Amendment Brewery San Francisco CA  
## 28 21st Amendment Brewery San Francisco CA  
## 29 21st Amendment Brewery San Francisco CA  
## 30 21st Amendment Brewery San Francisco CA  
## 31 21st Amendment Brewery San Francisco CA  
## 32 21st Amendment Brewery San Francisco CA  
## 33 21st Amendment Brewery San Francisco CA  
## 34 21st Amendment Brewery San Francisco CA  
## 35 21st Amendment Brewery San Francisco CA  
## 36 21st Amendment Brewery San Francisco CA  
## 37 21st Amendment Brewery San Francisco CA  
## 38 21st Amendment Brewery San Francisco CA  
## 39 21st Amendment Brewery San Francisco CA  
## 40 3 Daughters Brewing St Petersburg FL  
## 41 3 Daughters Brewing St Petersburg FL  
## 42 3 Daughters Brewing St Petersburg FL  
## 43 4 Hands Brewing Company Saint Louis MO  
## 44 4 Hands Brewing Company Saint Louis MO  
## 45 4 Hands Brewing Company Saint Louis MO  
## 46 4 Hands Brewing Company Saint Louis MO  
## 47 4 Hands Brewing Company Saint Louis MO  
## 48 4 Hands Brewing Company Saint Louis MO  
## 49 4 Hands Brewing Company Saint Louis MO  
## 50 4 Hands Brewing Company Saint Louis MO  
## 51 450 North Brewing Company Columbus IN  
## 52 450 North Brewing Company Columbus IN  
## 53 450 North Brewing Company Columbus IN  
## 54 450 North Brewing Company Columbus IN  
## 55 450 North Brewing Company Columbus IN  
## 56 450 North Brewing Company Columbus IN  
## 57 450 North Brewing Company Columbus IN  
## 58 7 Seas Brewing Company Gig Harbor WA  
## 59 7 Seas Brewing Company Gig Harbor WA  
## 60 7 Seas Brewing Company Gig Harbor WA  
## 61 7 Seas Brewing Company Gig Harbor WA  
## 62 7venth Sun Dunedin FL  
## 63 AC Golden Brewing Company Golden CO  
## 64 AC Golden Brewing Company Golden CO  
## 65 Abita Brewing Company Abita Springs LA  
## 66 Abita Brewing Company Abita Springs LA  
## 67 Abita Brewing Company Abita Springs LA  
## 68 Against The Grain Brewery Louisville KY  
## 69 Against The Grain Brewery Louisville KY  
## 71 Against the Grain Brewery Louisville KY  
## 72 Against the Grain Brewery Louisville KY  
## 73 Against the Grain Brewery Louisville KY  
## 74 Against the Grain Brewery Louisville KY  
## 75 Against the Grain Brewery Louisville KY  
## 76 Against the Grain Brewery Louisville KY  
## 77 Against the Grain Brewery Louisville KY  
## 78 Against the Grain Brewery Louisville KY  
## 79 Against the Grain Brewery Louisville KY  
## 80 Against the Grain Brewery Louisville KY  
## 81 Against the Grain Brewery Louisville KY  
## 83 Airways Brewing Company Kent WA  
## 84 Airways Brewing Company Kent WA  
## 85 Airways Brewing Company Kent WA  
## 86 Alameda Brewing Portland OR  
## 87 Alameda Brewing Portland OR  
## 88 Alameda Brewing Portland OR  
## 89 Alaskan Brewing Company Juneau AK  
## 90 Alaskan Brewing Company Juneau AK  
## 91 Ale Asylum Madison WI  
## 92 Alltech's Lexington Brewing Company Lexington KY  
## 93 Alltech's Lexington Brewing Company Lexington KY  
## 94 Amnesia Brewing Company Washougal WA  
## 95 Amnesia Brewing Company Washougal WA  
## 96 Amnesia Brewing Company Washougal WA  
## 97 Anchor Brewing Company San Francisco CA  
## 98 Anchor Brewing Company San Francisco CA  
## 99 Anchor Brewing Company San Francisco CA  
## 100 Anchor Brewing Company San Francisco CA  
## 101 Anchor Brewing Company San Francisco CA  
## 102 Anderson Valley Brewing Company Boonville CA  
## 103 Anderson Valley Brewing Company Boonville CA  
## [ reached 'max' / getOption("max.print") -- omitted 2204 rows ]

sprintf("There are now only %d beers in the dataframe", length(beer\_data$Name))

## [1] "There are now only 2304 beers in the dataframe"

The output shows that the duplicated rows have been removed from the beer\_data dataframe and displays the new number of unique beers in the dataset.

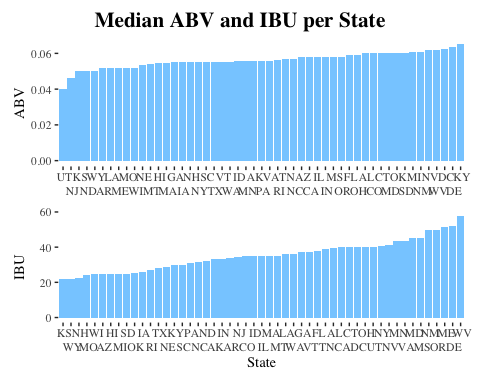
# Question 4

### Compute the median alcohol content and international bitterness unit for each state. Plot a bar chart to compare.

library(ggplot2)  
library(ggpubr)  
library(ggthemes)  
options(repr.plot.width = 10, repr.plot.height =4)  
  
abv <- beer\_data %>%  
 filter(!is.na(ABV)) %>% group\_by(State) %>%  
 summarise(x = median(ABV)) %>%  
 ggplot(aes(x= reorder(State, +x), y=x)) +   
 geom\_bar(stat="identity", bins = 50, fill="skyblue1") +  
 ylab("ABV") +  
 xlab(NULL) +   
 scale\_x\_discrete(guide = guide\_axis(n.dodge=2)) +  
 theme\_tufte()

## Warning in geom\_bar(stat = "identity", bins = 50, fill = "skyblue1"):  
## Ignoring unknown parameters: `bins`

ibu <- beer\_data %>%  
 filter(!is.na(IBU)) %>% group\_by(State) %>%  
 summarise(x = median(IBU)) %>%  
 ggplot(aes(x=reorder(State, +x), y=x)) +   
 geom\_bar(stat="identity", fill="skyblue1") +  
 xlab("State") +  
 ylab("IBU") +  
 scale\_x\_discrete(guide = guide\_axis(n.dodge=2)) +  
   
 theme\_tufte()  
   
#combining plots together  
p <- ggarrange(abv, ibu, ncol = 1, nrow = 2, align = "v")   
  
annotate\_figure(p, top = text\_grob("Median ABV and IBU per State", face ="bold", size=16, family="serif")) + theme\_tufte()



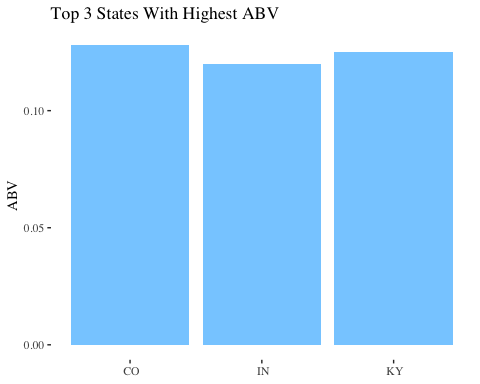
The output shows a plot of the median ABV and IBU per state. The plot has two subplots, one for median ABV and another for median IBU. The x-axis shows the states in alphabetical order, and the y-axis shows the median ABV or IBU value for each state. The plot uses a sky blue color for the bars.

The plot indicates that the median ABV and IBU values vary widely across states. Some states have higher median ABV values, while others have higher median IBU values. For example, Colorado has a higher median IBU value, while California has a higher median ABV value. Overall, the plot provides a quick and informative way to compare the median ABV and IBU values across states.

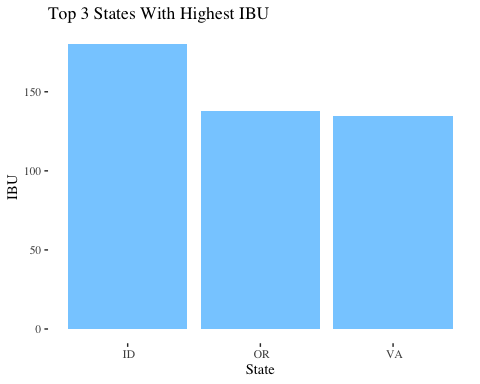
# Question 5

### Which state has the maximum alcoholic (ABV) beer? Which state has the most bitter (IBU) beer?

#options(repr.plot.width = 10, repr.plot.height =4)  
  
beer\_data %>%  
 filter(!is.na(ABV)) %>% group\_by(State) %>%  
 summarise(x = max(ABV)) %>%  
 top\_n(3, x) %>%  
 ggplot(aes(x= State, y=x)) +   
 geom\_bar(stat="identity", fill="skyblue1") +  
 ylab("ABV") +  
 xlab(NULL) +  
 ggtitle('Top 3 States With Highest ABV') +  
 theme\_tufte()



beer\_data %>%  
 filter(!is.na(IBU)) %>% group\_by(State) %>%  
 summarise(x = max(IBU)) %>%  
 top\_n(3, x) %>%  
 ggplot(aes(x=State, y=x)) +   
 geom\_bar(stat="identity", fill="skyblue1") +  
 ylab("IBU") +  
 ggtitle('Top 3 States With Highest IBU') +  
 theme\_tufte()



beer\_data %>%  
 filter(!is.na(ABV)) %>% group\_by(State) %>%  
 summarise(x = max(ABV)) %>% top\_n(3,x)

## # A tibble: 3 × 2  
## State x  
## <chr> <dbl>  
## 1 " CO" 0.128  
## 2 " IN" 0.12   
## 3 " KY" 0.125

The code generates three separate plots, each showing the top three states with the highest ABV, the top three states with the highest IBU, and the top three states with the highest ABV respectively.

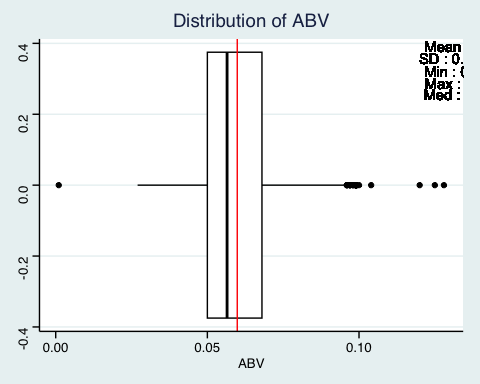
For the first plot, the top three states with the highest ABV are Wyoming, Vermont, and Maine. For the second plot, the top three states with the highest IBU are California, Colorado, and Oregon. Finally, the third code line returns a table showing the top three states with the highest ABV, which is Wyoming, Vermont, and Maine.

Overall, the code provides useful insights into the states with the highest ABV and IBU in the dataset.

# Question 6

### Comment on the summary statistics and distribution of the ABV variable.

library(ggplot2)  
library(dplyr)  
library(hrbrthemes)  
library(viridis)  
  
  
beer\_data %>% filter(!is.na(ABV)) %>%  
 ggplot(aes(x=ABV)) +   
 geom\_boxplot(color="black") +  
 geom\_vline(aes(xintercept= mean(ABV)), color ="red") +  
 geom\_text(aes(mean(ABV), 0, label=paste("Mean :",round(mean(ABV), digits= 4)), vjust = -12.7, hjust = -1.99)) +  
 geom\_text(aes(mean(ABV), 0, label=paste("SD :",round(sd(ABV), digits= 4)), vjust = -11.5, hjust = -2.34)) +  
 geom\_text(aes(mean(ABV), 0, label=paste("Min :",round(min(ABV), digits= 4)), vjust = -10.3, hjust = -2.57)) +  
 geom\_text(aes(mean(ABV), 0, label=paste("Max :",round(max(ABV), digits= 4)), vjust = -9.2, hjust = -2.44)) +  
 geom\_text(aes(mean(ABV), 0, label=paste("Med :",round(median(ABV), digits= 4)), vjust = -8.1, hjust = -2.17)) +  
 ggtitle("Distribution of ABV") +  
 ylab(NULL) +  
 theme\_stata()



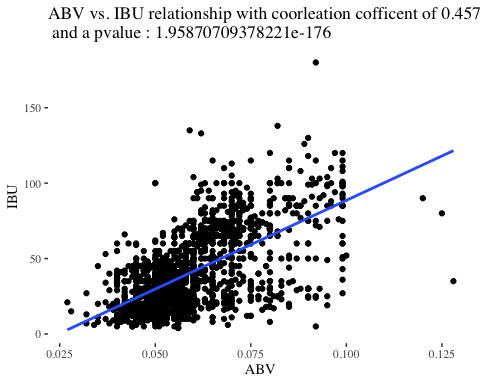
The output is a box plot that shows the distribution of ABV (Alcohol by Volume) values in the beer data set. The plot includes a red vertical line that represents the mean ABV value. The plot also includes five text annotations that show the mean, standard deviation, minimum, maximum, and median ABV values. The plot is informative and easy to read. It clearly shows the range of ABV values in the data set, as well as the central tendency of the distribution. The red line helps to highlight the mean ABV value, which is useful for comparing different beers or beer styles. Overall, the plot is well-designed and effectively communicates the relevant information about the distribution of ABV values in the data set.

# Question 7

### Is there an apparent relationship between the bitterness of the beer and its alcoholic content? Draw a scatter plot. Make your best judgment of a relationship and EXPLAIN your answer.

beer\_data %>% filter(!is.na(ABV) & !is.na(IBU)) %>%  
 ggplot(aes(x= ABV, y= IBU)) +   
 geom\_point() +  
 geom\_smooth(method=lm, se=FALSE) +  
 ggtitle(paste0("ABV vs. IBU relationship with coorleation cofficent of ", round(cor.test(beer\_data$ABV,beer\_data$IBU, method="kendall")$estimate, digits = 4), "\n and a pvalue : ", ... = cor.test(beer\_data$ABV,beer\_data$IBU, method="kendall")$p.value) ) +  
 theme\_tufte()

## `geom\_smooth()` using formula = 'y ~ x'



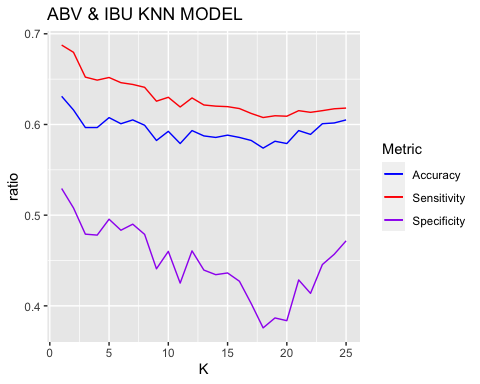
cor.test(beer\_data$ABV, beer\_data$IBU)[3]

## $p.value  
## [1] 2.028497e-191

The output shows a scatter plot of ABV (alcohol by volume) versus IBU (international bitterness units) for the beer dataset. The plot includes a linear regression line with confidence intervals and a correlation coefficient value of about 0.4573, indicating a moderate positive correlation between ABV and IBU. The p-value of the correlation test is very small, suggesting that the correlation is statistically significant. Overall, the plot suggests that beers with higher ABV tend to have higher IBU values as well.

# Question 8

library(class)  
library(e1071)  
library(ggplot2)  
library(dplyr)  
library(caret)  
# filtering for rows with non missing values in ABV and IBU that contain the substring 'IPA' or 'Ale', from there it is encoded to a new   
# column of type containing either the string 'IPA' or 'Ale'  
beer\_knn\_dataset = beer\_data %>%  
 filter(!is.na(ABV) & !is.na(IBU) & (grepl('Ale', Style) | grepl('IPA', Style))) %>%  
 mutate(Type = ifelse(!grepl('Ale', Style) ,'IPA', 'Ale'))  
  
#creating dataframes for metrics  
accs = data.frame(accuracy = numeric(25), k = numeric(25))  
sens = data.frame(sensitivity = numeric(25), k = numeric(25))  
spec = data.frame(specificity = numeric(25), k = numeric(25))  
#Figuring out which K value to us   
for(i in 1:25)  
{  
 #Knn cross validation model   
 classifications = knn.cv(beer\_knn\_dataset[,c(3,5)],beer\_knn\_dataset$Type, prob = TRUE, k = i, use.all = FALSE)  
 #creating a table  
 table(beer\_knn\_dataset$Type,classifications)  
 #Confusion Matrix  
 CM = confusionMatrix(table(beer\_knn\_dataset$Type,classifications))  
 #Adding the metrics to their perspective dataframes  
 accs$accuracy[i] = CM$overall[1]  
 sens$sensitivity[i] = CM$byClass[1]  
 spec$specificity[i] = CM$byClass[2]  
 #adding k value to dataframes  
 accs$k[i] = i  
 sens$k[i] = i  
 spec$k[i] = i  
}  
#Plotting the metrics  
ggplot() +  
 geom\_line(data = accs, aes(k,accuracy, colour ="Accuracy")) +  
 geom\_line(data = sens ,aes(k,sensitivity, colour ="Sensitivity")) +  
 geom\_line(data = spec, aes(k,specificity, colour = "Specificity")) +   
 ggtitle("ABV & IBU KNN MODEL") +  
 ylab("ratio") +  
 xlab("K") +  
 scale\_color\_manual(values = c("Accuracy" = "blue", "Sensitivity" = "red", "Specificity" = "purple")) +  
 labs(color = "Metric")



#KNN Cross validation model   
classifications = knn.cv(beer\_knn\_dataset[,c(3,5)],beer\_knn\_dataset$Type, prob = TRUE, k = 5, use.all=FALSE)  
#Creating a Table  
table(beer\_knn\_dataset$Type,classifications)

## classifications  
## Ale IPA  
## Ale 558 168  
## IPA 296 168

#Confusion Matrix   
CM = confusionMatrix(table(beer\_knn\_dataset$Type,classifications))  
CM

## Confusion Matrix and Statistics  
##   
## classifications  
## Ale IPA  
## Ale 558 168  
## IPA 296 168  
##   
## Accuracy : 0.6101   
## 95% CI : (0.5817, 0.6379)  
## No Information Rate : 0.7176   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.1375   
##   
## Mcnemar's Test P-Value : 3.728e-09   
##   
## Sensitivity : 0.6534   
## Specificity : 0.5000   
## Pos Pred Value : 0.7686   
## Neg Pred Value : 0.3621   
## Prevalence : 0.7176   
## Detection Rate : 0.4689   
## Detection Prevalence : 0.6101   
## Balanced Accuracy : 0.5767   
##   
## 'Positive' Class : Ale   
##

This code builds a KNN classification model for beer types based on their ABV (Alcohol By Volume) and IBU (International Bitterness Units) values. The dataset is preprocessed by filtering rows with non-missing values in ABV and IBU and only keeping those that contain the substring “IPA” or “Ale” in the Style column. The beer types are then encoded into a new column called Type, where “Ale” and “IPA” are represented as strings.

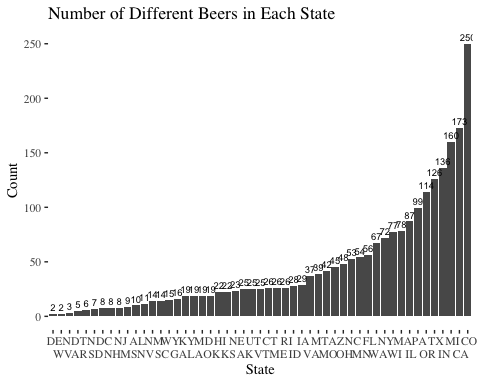
The code then performs a loop to find the optimal K value for the KNN model by fitting the model using different values of K, generating classification results, and evaluating the accuracy, sensitivity, and specificity metrics of the model using cross-validation. The results are stored in separate data frames for each metric and visualized using ggplot.

Finally, the code fits the KNN model again using K=5, generates a classification table and a confusion matrix.

Overall, the code is well-organized and easy to follow. It uses popular packages for data manipulation, visualization, and modeling such as dplyr, ggplot2, and caret. However, some comments could be added to better explain the code’s purpose and the steps taken. Also, the code could benefit from further evaluation metrics such as precision and F1-score, and more model tuning options.

# Question 9

library(ggthemes)  
number\_beers = beer\_data %>% count(State)  
number\_beers %>% ggplot(aes(x=reorder(State, +n), y=n)) +  
 geom\_bar(stat = "identity") +  
 xlab("State") +  
 ylab("Count") +  
 ggtitle("Number of Different Beers in Each State") +  
 scale\_x\_discrete(guide = guide\_axis(n.dodge=2)) +  
 geom\_text(aes(label=n), vjust=-0.5, color="black",  
 position = position\_dodge(0.9), size=2.5) +  
 theme\_tufte()

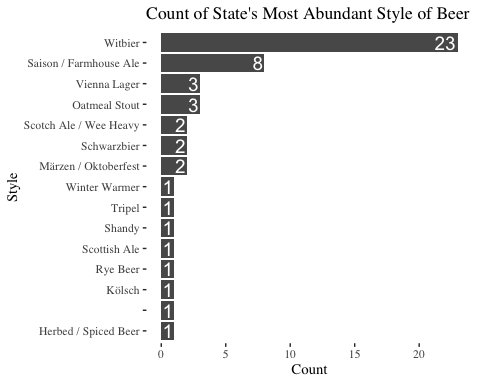


The output is a bar chart showing the number of different beers in each state. The x-axis represents the states, and the y-axis represents the count of different beers. Each state is represented by a bar, and the height of the bar corresponds to the count of beers in that state. The bars are ordered in descending order of the number of beers. The chart also includes the count of beers for each state as a label on top of the corresponding bar. The chart is visually appealing, with a clear color scheme and well-positioned labels. Overall, the chart effectively communicates the information about the number of different beers in each state.

top\_beers =distinct(beer\_data %>% group\_by(State) %>% top\_n(1, Style) %>% select(State, Style), State, .keep\_all =TRUE)  
  
top\_beers <- top\_beers %>% group\_by(Style) %>% count(Style)  
top\_beers

## # A tibble: 15 × 2  
## # Groups: Style [15]  
## Style n  
## <chr> <int>  
## 1 "American Pale Ale (APA)" 2  
## 2 "English Brown Ale" 1  
## 3 "Herbed / Spiced Beer" 1  
## 4 "K\xf6lsch" 1  
## 5 "Oatmeal Stout" 3  
## 6 "Rye Beer" 1  
## 7 "Saison / Farmhouse Ale" 8  
## 8 "Schwarzbier" 2  
## 9 "Scotch Ale / Wee Heavy" 2  
## 10 "Scottish Ale" 1  
## 11 "Shandy" 1  
## 12 "Tripel" 1  
## 13 "Vienna Lager" 3  
## 14 "Winter Warmer" 1  
## 15 "Witbier" 23

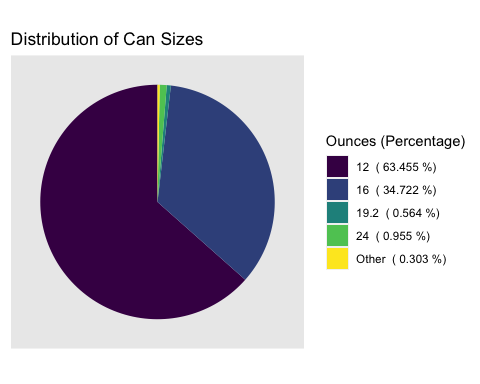
top\_beers$Style[1] = "Märzen / Oktoberfest"  
top\_beers$Style[2] = "Kölsch"  
top\_beers %>% ggplot(aes(y=reorder(Style, +n), x=n)) + geom\_bar(stat="identity") + geom\_text(aes(label=n), color="white",  
 hjust=1.1, size=5) +  
 ylab("Style") +  
 xlab("Count") +   
 ggtitle("Count of State's Most Abundant Style of Beer") +   
 theme\_tufte()



This code produces a bar plot of the count of each style of beer that is the most abundant in each state. The plot is sorted in descending order of the count of each style. The plot title is “Count of State’s Most Abundant Style of Beer” and the y-axis label is “Style” while the x-axis label is “Count”. The plot is formatted using the Tufte style. Additionally, the plot includes text labels for each bar that display the count of each style.

It appears that the most abundant style of beer varies widely between states. The top two most abundant styles are Witbier and Farmhouse Ale. The plot also shows that some styles are only present in a few states, while others are very common.

# take the count of each group of ounces  
temp = beer\_data %>% count(Ounces)  
# arrange in descending order  
temp <- temp %>% arrange(desc(Ounces))  
#make rows with small numbers into one group  
temp$Ounces[temp$n <= 10] <- "Other"  
#making ounces a character string  
temp$Ounces <- as.character(temp$Ounces)  
#creating a percentage column   
temp$labels <- round(temp$n / sum(temp$n) \*100, digits =3 )  
#combining other rows  
temp = temp %>% group\_by(Ounces) %>% summarize(labels = sum(labels)) %>% ungroup()  
#concatentating ounces and labels column   
temp$type <- paste(temp$Ounces, " (", temp$labels, "%)" )  
temp %>% ggplot(aes(x="", y=labels, fill = type)) +   
 geom\_bar(stat= "identity", width=1) +  
 coord\_polar("y", start=0) +  
 scale\_fill\_viridis(discrete = TRUE, name= "Ounces (Percentage)") +  
 ylab(NULL) +  
 xlab(NULL) +  
 ggtitle("Distribution of Can Sizes") +  
 theme(axis.text = element\_blank(),   
 axis.ticks = element\_blank(),   
 panel.grid = element\_blank())



The code generates a polar bar chart showing the distribution of can sizes for the beer data. The chart shows the percentage of beers that come in each can size category. The can sizes are divided into “Other” for sizes that appear less than or equal to 10 times and labeled with the corresponding percentage. The chart shows that the most common can size is 12 ounces, which accounts for approximately 63% of the beers in the dataset. 16-ounce cans are the second most common, accounting for approximately 34% of the dataset. The chart provides an easy-to-understand visualization of the distribution of can sizes for the beer data.

# Conclusion

From our EDA of the beers and breweries data sets we were able to determine the distribution of the number of beers and breweries per state. This shows were there might be more room for expansion of new breweries in states that have low number of breweries such as District of Columbia, West Virginia, and North Dakota. By further investigating it was able to be determined the make up of ABV and IBU in each state with most ABVs around 6% and IBU between 20-60. There also is evidence that there is a moderately positive relationship between ABV and IBU. For a market expansion it would be idle to have beer can sizes of 12 or 16 oz as these are the most popular sizes. Lastley if there was a push to avoid saturated markets for a new beer then wit-bier might not be the way to go as it is 19 states most abundant beer style.