Project 1

Akbar Thobani

2/19/2020

```
#loading libraries
library(dplyr)
library(tidyverse)
library(ggplot2)
library(data.table)
```

```
#Reading in the data
breweries <- read.csv("Breweries.csv",header=TRUE)
beer <- read.csv("Beers.csv",header=TRUE)
```

```
#count of breweries by state
breweries %>% count(State)
```

```
## # A tibble: 51 x 2
##    State     n
##    <fct> <int>
##  1 " AK"     7
##  2 " AL"     3
##  3 " AR"     2
##  4 " AZ"    11
##  5 " CA"    39
##  6 " CO"    47
##  7 " CT"     8
##  8 " DC"     1
##  9 " DE"     2
## 10 " FL"    15
## # ... with 41 more rows
```

```
#assessing merge
head(beer)
```

```
##                   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
## 1            American Pale Lager     12
```

```
## 2        American Pale Ale (APA)     12
## 3                    American IPA     12
## 4 American Double / Imperial IPA     12
## 5                    American IPA     12
## 6                  Oatmeal Stout     12
```

```
head(breweries)
```

```
##   Brew_ID                      Name          City State
## 1       1        NorthGate Brewing    Minneapolis    MN
## 2       2 Against the Grain Brewery    Louisville    KY
## 3       3  Jack's Abby Craft Lagers    Framingham    MA
## 4       4 Mike Hess Brewing Company     San Diego    CA
## 5       5   Fort Point Beer Company San Francisco    CA
## 6       6     COAST Brewing Company    Charleston    SC
```

```
str(beer)
```

```
## 'data.frame':    2410 obs. of  7 variables:
##  $ Name      : Factor w/ 2305 levels "#001 Golden Amber Lager",..: 1638 577 1704 1842 1819 268 1160 1
##  $ Beer_ID   : int  1436 2265 2264 2263 2262 2261 2260 2259 2258 2131 ...
##  $ ABV       : num  0.05 0.066 0.071 0.09 0.075 0.077 0.045 0.065 0.055 0.086 ...
##  $ IBU       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Brewery_id: int  409 178 178 178 178 178 178 178 178 178 ...
##  $ Style     : Factor w/ 100 levels "","Abbey Single Ale",..: 19 18 16 12 16 80 18 22 18 12 ...
##  $ Ounces    : num  12 12 12 12 12 12 12 12 12 12 ...
```

```
str(breweries)
```

```
## 'data.frame':    558 obs. of  4 variables:
##  $ Brew_ID: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Name   : Factor w/ 551 levels "10 Barrel Brewing Company",..: 355 12 266 319 201 136 227 477 59 49
##  $ City   : Factor w/ 384 levels "Abingdon","Abita Springs",..: 228 200 122 299 300 62 91 48 152 136
##  $ State  : Factor w/ 51 levels " AK"," AL"," AR",..: 24 18 20 5 5 41 6 23 23 23 ...
```

```
#merge by Brewery ID key
beerbrew <- left_join(beer, breweries, by = c("Brewery_id" = "Brew_ID"))
```

```
#print top 6 & bottom 6
head(beerbrew, n=6)
```

```
##                Name.x 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                     Name.y 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(beerbrew, n=6)
```

```
##                           Name.x 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                   Name.y   City State
## 2405           American Stout     12 Wynkoop Brewing Company Denver    CO
## 2406              Belgian IPA     12 Wynkoop Brewing Company Denver    CO
## 2407 American Amber / Red Ale     12 Wynkoop Brewing Company Denver    CO
## 2408              Schwarzbier     12 Wynkoop Brewing Company Denver    CO
## 2409  American Pale Ale (APA)     12 Wynkoop Brewing Company Denver    CO
## 2410 American Amber / Red Ale     12 Wynkoop Brewing Company Denver    CO
```

```
#missing values - INCOMPLETE
is.na(beerbrew$Name)
```

```
## logical(0)
```

```
summary(beerbrew)
```

```
##                     Name.x        Beer_ID            ABV
##  Nonstop Hef Hop       :  12   Min.   :   1.0   Min.   :0.00100
##  Dale's Pale Ale       :   6   1st Qu.: 808.2   1st Qu.:0.05000
##  Oktoberfest           :   6   Median :1453.5   Median :0.05600
##  Longboard Island Lager:   4   Mean   :1431.1   Mean   :0.05977
##  1327 Pod's ESB        :   3   3rd Qu.:2075.8   3rd Qu.:0.06700
##  Boston Lager          :   3   Max.   :2692.0   Max.   :0.12800
##  (Other)               :2376                    NA's   :62
##       IBU           Brewery_id                                Style
##  Min.   :  4.00   Min.   :  1.0   American IPA                  : 424
##  1st Qu.: 21.00   1st Qu.: 94.0   American Pale Ale (APA)       : 245
##  Median : 35.00   Median :206.0   American Amber / Red Ale      : 133
##  Mean   : 42.71   Mean   :232.7   American Blonde Ale           : 108
##  3rd Qu.: 64.00   3rd Qu.:367.0   American Double / Imperial IPA: 105
##  Max.   :138.00   Max.   :558.0   American Pale Wheat Ale       :  97
##  NA's   :1005                     (Other)                      :1298
```

```
##      Ounces                             Name.y              City
##  Min.   : 8.40   Brewery Vivant            :  62   Grand Rapids:  66
##  1st Qu.:12.00   Oskar Blues Brewery       :  46   Portland    :  64
##  Median :12.00   Sun King Brewing Company  :  38   Chicago     :  55
##  Mean   :13.59   Cigar City Brewing Company:  25   Indianapolis:  43
##  3rd Qu.:16.00   Sixpoint Craft Ales       :  24   San Diego   :  42
##  Max.   :32.00   Hopworks Urban Brewery    :  23   Boulder     :  41
##                  (Other)                   :2192   (Other)     :2099
##      State
##   CO    : 265
##   CA    : 183
##   MI    : 162
##   IN    : 139
##   TX    : 130
##   OR    : 125
##  (Other):1406
```

```
str(beerbrew)
```

```
## 'data.frame':    2410 obs. of  10 variables:
##  $ Name.x    : Factor w/ 2305 levels "#001 Golden Amber Lager",..: 1638 577 1704 1842 1819 268 1160 1
##  $ Beer_ID   : int  1436 2265 2264 2263 2262 2261 2260 2259 2258 2131 ...
##  $ ABV       : num  0.05 0.066 0.071 0.09 0.075 0.077 0.045 0.065 0.055 0.086 ...
##  $ IBU       : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Brewery_id: int  409 178 178 178 178 178 178 178 178 178 ...
##  $ Style     : Factor w/ 100 levels "","Abbey Single Ale",..: 19 18 16 12 16 80 18 22 18 12 ...
##  $ Ounces    : num  12 12 12 12 12 12 12 12 12 12 ...
##  $ Name.y    : Factor w/ 551 levels "10 Barrel Brewing Company",..: 1 2 2 2 2 2 2 2 2 2 ...
##  $ City      : Factor w/ 384 levels "Abingdon","Abita Springs",..: 32 131 131 131 131 131 131 131 131
##  $ State     : Factor w/ 51 levels " AK"," AL"," AR",..: 38 16 16 16 16 16 16 16 16 16 ...
```

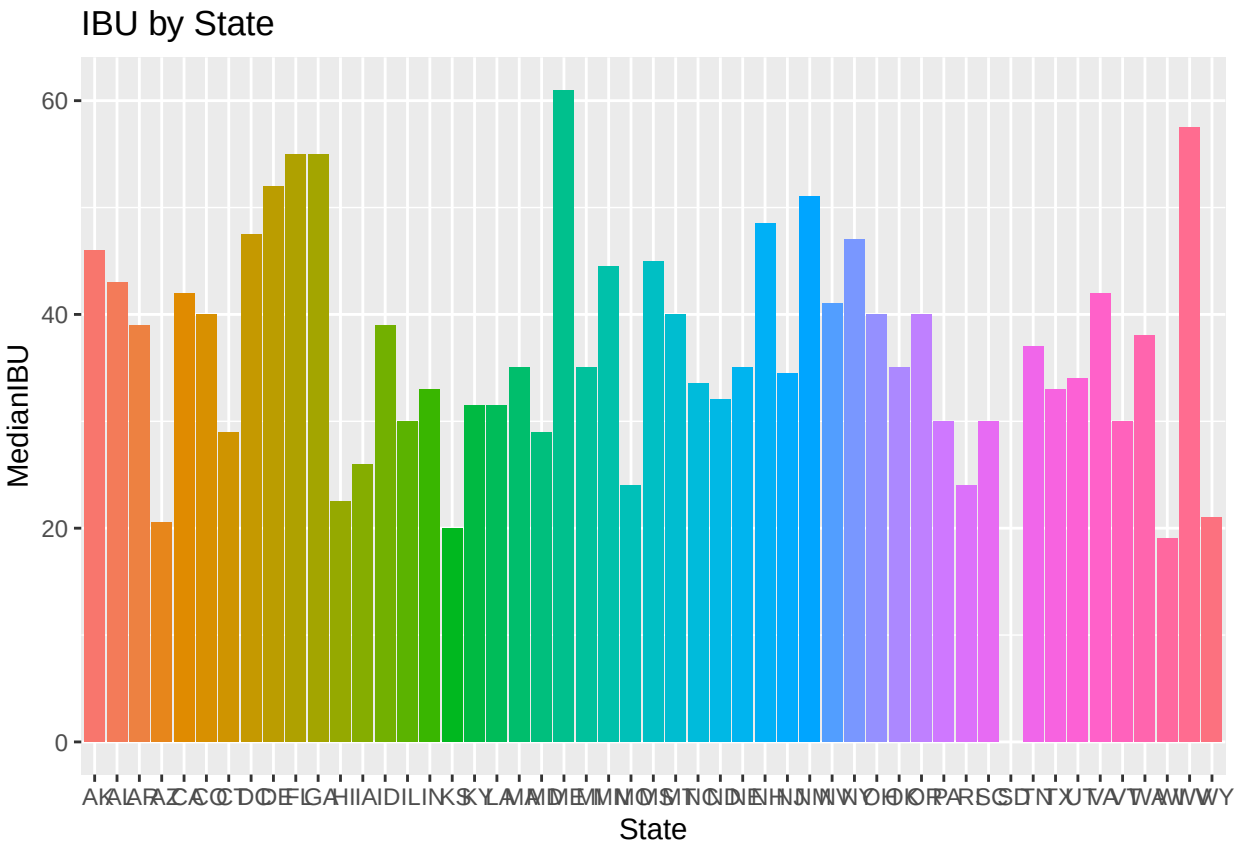
```
#barplot of alcohol content and international bitterness by state - INCOMPLETE
#finding median values
beerbrew %>% group_by(State) %>% summarize(median(ABV, na.rm=TRUE))
```

```
## # A tibble: 51 x 2
##    State `median(ABV, na.rm = TRUE)`
##    <fct>                       <dbl>
##  1 " AK"                       0.056
##  2 " AL"                       0.06
##  3 " AR"                       0.052
##  4 " AZ"                       0.055
##  5 " CA"                       0.058
##  6 " CO"                       0.0605
##  7 " CT"                       0.06
##  8 " DC"                       0.0625
##  9 " DE"                       0.055
## 10 " FL"                       0.057
## # ... with 41 more rows
```

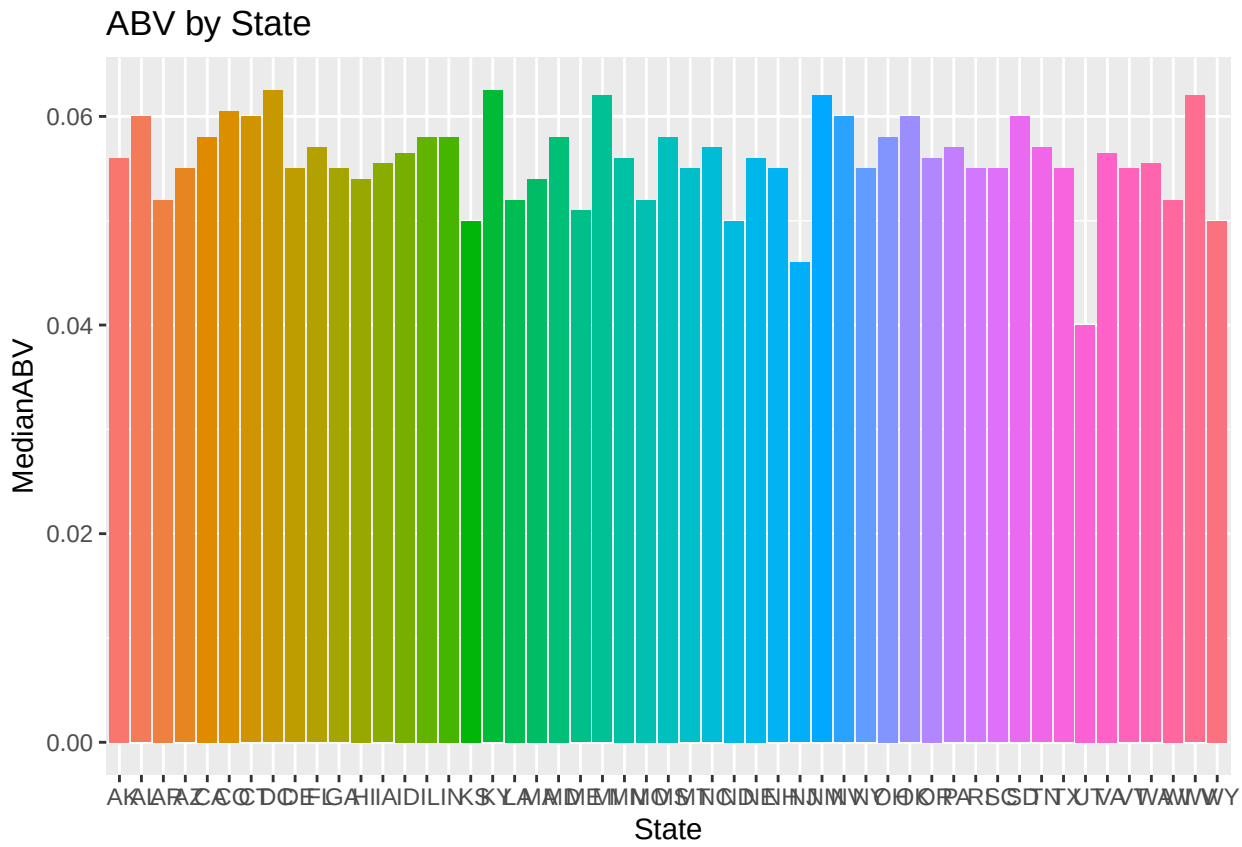
```
medians <- setDT(beerbrew)[,list(MedianABV=as.numeric(median(ABV, na.rm=TRUE)),MedianIBU=as.numeric(medi

ggplot(data=medians, mapping = aes(x=State, y=MedianIBU, fill=State)) + geom_bar(stat = "identity") + gg
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```




```
ggplot(data=medians, mapping = aes(x=State, y=MedianABV, fill=State)) + geom_bar(stat = "identity") + g
```




```
#finding state with max ABV and IBU
which.max(beerbrew$ABV)
```

```
## [1] 2279
```

```
beerbrew[2279,]
```

```
##                                                Name.x Beer_ID   ABV IBU
## 1: Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale    2565 0.128  NA
##    Brewery_id            Style Ounces                  Name.y    City
## 1:         52 Quadrupel (Quad)   19.2 Upslope Brewing Company Boulder
##    State
## 1:    CO
```

```
which.max(beerbrew$IBU)
```

```
## [1] 148
```

```
beerbrew[148,]
```

```
##                       Name.x Beer_ID   ABV IBU Brewery_id
## 1: Bitter Bitch Imperial IPA     980 0.082 138        375
##                            Style Ounces                  Name.y    City
```

```
## 1: American Double / Imperial IPA     12 Astoria Brewing Company Astoria
##    State
## 1:    OR
```

```
#summary statistics of ABV
summary(beerbrew$ABV)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.00100 0.05000 0.05600 0.05977 0.06700 0.12800      62
```

```
#Scatterplot of ABV and Internationl bitterness to assess correlation
ggplot(data=beerbrew, mapping = aes(x=ABV, y=IBU)) + geom_point(mapping = aes(color = State)) + geom_sm
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 1005 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1005 rows containing missing values (geom_point).
```