Gautam Kapila Case Study 1

G.Kapila

7/5/2019

United States Brewery Market and Beer Composition Statistics

Introduction

The work summarizes following aspects of US brewery market and beer composition per state

- Number of breweries present per state
- Acoholic content of craft beers per state
- Bitterness of the beers per state
- State with beer having maximum alchoholic content and bitterness
- Relationship between beer alcoholic content and beer bitterness

It is envisaged that this work could be extended in future, to come up with proposals on where to start a new brewery venture and target beer composition

Session and Library Information

```
sessionInfo()
```

```
## R version 3.5.3 (2019-03-11)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.5.3  magrittr_1.5    tools_3.5.3     htmltools_0.3.6
##  [5] yaml_2.2.0      Rcpp_1.0.1      stringi_1.4.3   rmarkdown_1.12
##  [9] knitr_1.22      stringr_1.4.0   xfun_0.6        digest_0.6.19
## [13] evaluate_0.13
```

```
library(tidyr)
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
```

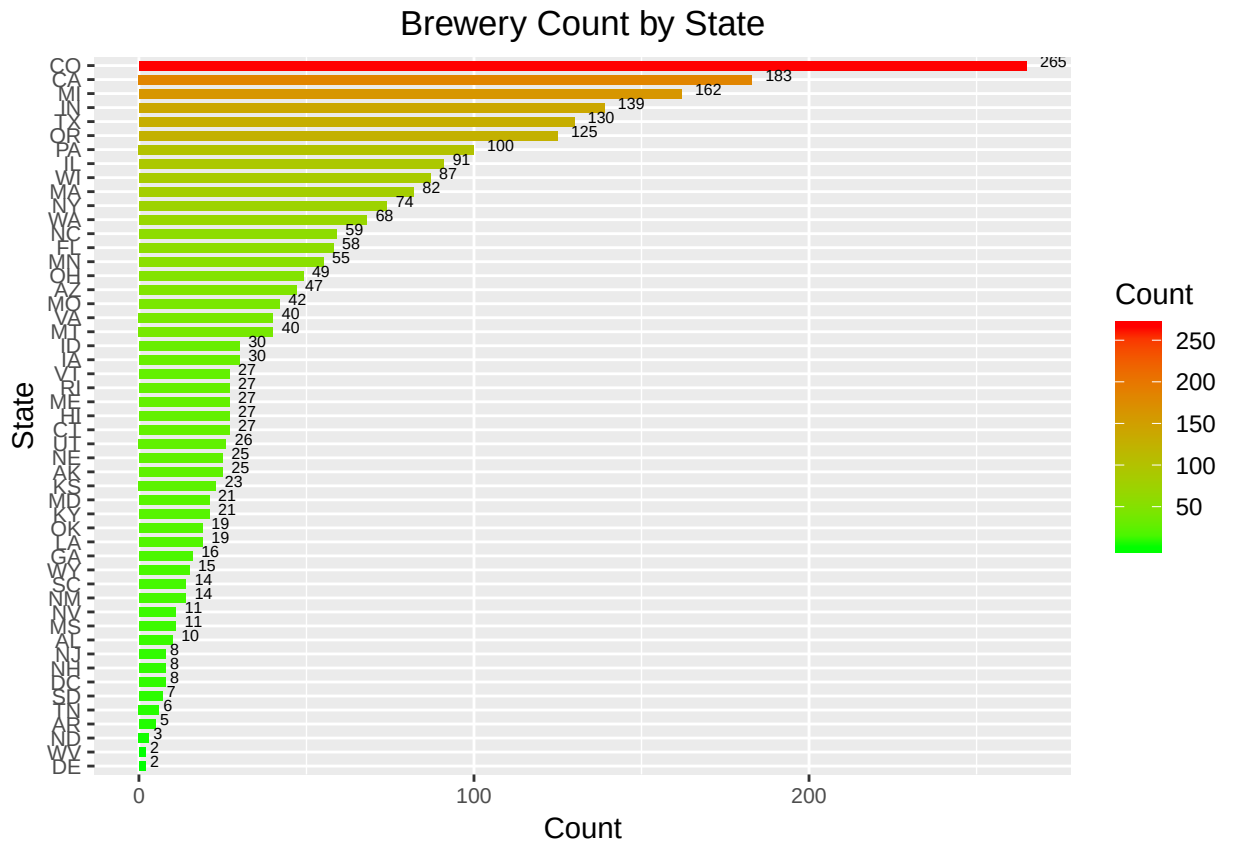
```
library(stringr)
library(ggplot2)
```

Data Preparation

```
beers <- read.csv(file = "CaseStudy1_2_2_2\\beers.csv",sep = ",")
breweries <- read.csv(file = "CaseStudy1_2_2_2\\breweries.csv",sep = ",")
bd    <- merge(beers, breweries, by.x = "Brewery_id", by.y = "Brew_ID")
bd$State <- gsub('\\s+', '', bd$State)
colnames(bd) <- c("BreweryID","BeerName","BeerID","ABV","IBU","Style","Oz","BreweryName","City","State")
write.csv(bd,file = "bd.csv",row.names = FALSE)
```

Breweries present in each state

```
BDist <- bd %>% dplyr::count(State,sort = TRUE, name = "Count")
ggplot(data = BDist, aes(x=reorder(State,Count),y=Count,fill = Count)) +
    geom_bar(stat = "identity",width = 0.7) +
    coord_flip() +
    geom_text(aes(label=Count), hjust = -0.5, vjust = 0.1, size=2)+
    ggtitle('Brewery Count by State') +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab("State")+
    theme(axis.text = element_text(size = 8)) +
    scale_fill_gradient(low = "green",high = "red")
```



Printing first 6 and last 6 data from merged file

```
rbind(head(bd,6),tail(bd,6))
```

```
##      BreweryID                BeerName BeerID   ABV IBU
## 1            1             Get Together   2692 0.045  50
## 2            1            Maggie's Leap   2691 0.049  26
## 3            1              Wall's End   2690 0.048  19
## 4            1                  Pumpion   2689 0.060  38
## 5            1               Stronghold   2688 0.060  25
## 6            1             Parapet ESB   2687 0.056  47
## 2405       556            Pilsner Ukiah     98 0.055  NA
## 2406       557 Heinnieweisse Weissebier     52 0.049  NA
## 2407       557          Snapperhead IPA     51 0.068  NA
## 2408       557       Moo Thunder Stout      50 0.049  NA
## 2409       557       Porkslap Pale Ale      49 0.043  NA
## 2410       558 Urban Wilderness Pale Ale     30 0.049  NA
##                                 Style Oz              BreweryName
## 1                         American IPA 16           NorthGate Brewing
## 2                  Milk / Sweet Stout 16           NorthGate Brewing
## 3                   English Brown Ale 16           NorthGate Brewing
## 4                         Pumpkin Ale 16           NorthGate Brewing
## 5                      American Porter 16           NorthGate Brewing
## 6    Extra Special / Strong Bitter (ESB) 16           NorthGate Brewing
```

```
## 2405                     German Pilsener 12         Ukiah Brewing Company
## 2406                          Hefeweizen 12       Butternuts Beer and Ale
## 2407                        American IPA 12       Butternuts Beer and Ale
## 2408                  Milk / Sweet Stout 12       Butternuts Beer and Ale
## 2409              American Pale Ale (APA) 12       Butternuts Beer and Ale
## 2410                    English Pale Ale 12 Sleeping Lady Brewing Company
##               City State
## 1      Minneapolis    MN
## 2      Minneapolis    MN
## 3      Minneapolis    MN
## 4      Minneapolis    MN
## 5      Minneapolis    MN
## 6      Minneapolis    MN
## 2405         Ukiah    CA
## 2406 Garrattsville    NY
## 2407 Garrattsville    NY
## 2408 Garrattsville    NY
## 2409 Garrattsville    NY
## 2410     Anchorage    AK
```

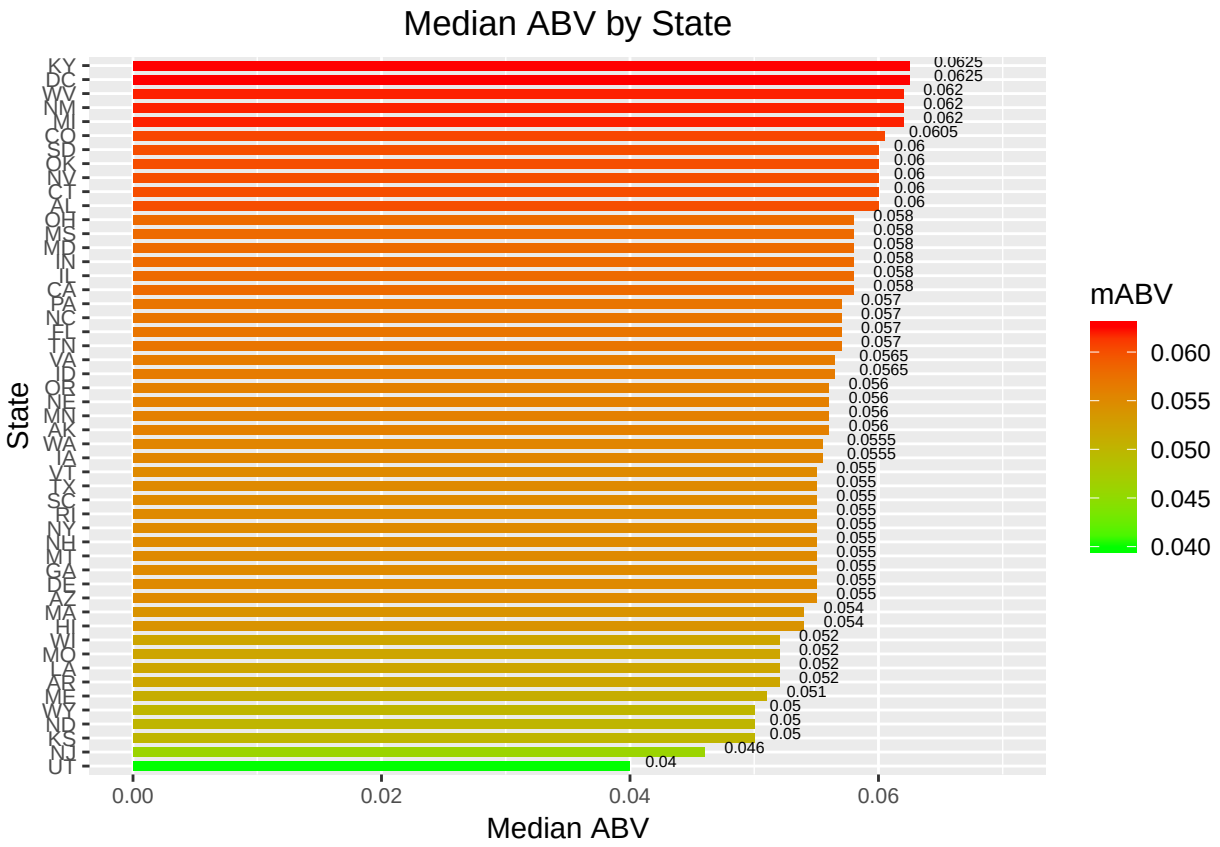
Printing NA's count per column

```
sapply(bd, function(x) sum(length(which(is.na(x)))))
```

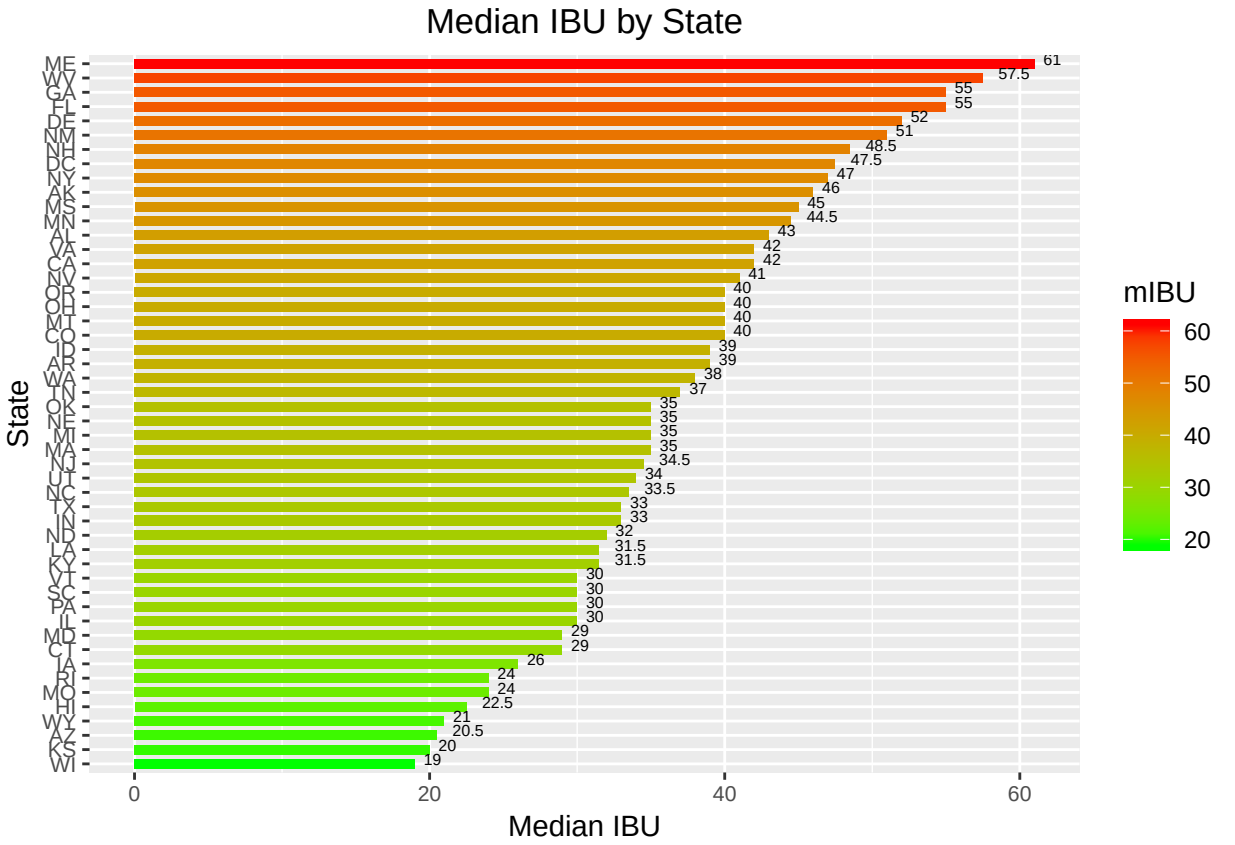
```
##   BreweryID    BeerName      BeerID         ABV         IBU       Style
##           0           0           0          62        1005           0
##          Oz BreweryName        City       State
##           0           0           0           0
```

Median Alcohol Content and IBU for each state

```
mABVbyState <- bd %>% drop_na(ABV) %>% group_by(State) %>% summarize(mABV = median(ABV))
ggplot(data = mABVbyState, aes(x=reorder(State,mABV),y=mABV,fill = mABV)) +
    geom_bar(stat = "identity",width = 0.7) +
    coord_flip() +
    geom_text(aes(label=mABV), hjust = -0.5, vjust = 0.1, size=2)+
    ggtitle('Median ABV by State') +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab("State")+
    ylab("Median ABV")+
    theme(axis.text = element_text(size = 8)) +
    scale_fill_gradient(low = "green",high = "red") +
    ylim(0,0.07)
```




```
mIBUbyState <- bd %>% drop_na(IBU) %>% group_by(State) %>% summarize(mIBU = median(IBU))
ggplot(data = mIBUbyState, aes(x=reorder(State,mIBU),y=mIBU,fill = mIBU)) +
    geom_bar(stat = "identity",width = 0.7) +
    coord_flip() +
    geom_text(aes(label=mIBU), hjust = -0.5, vjust = 0.1, size=2)+
    ggtitle('Median IBU by State') +
    theme(plot.title = element_text(hjust = 0.5)) +
    xlab("State")+
    ylab("Median IBU")+
    theme(axis.text = element_text(size = 8)) +
    scale_fill_gradient(low = "green",high = "red")
```



State with Max Alcohol Content and Max IBU

State with Max Alcoholic Beer

```
maxABVbyState <- bd %>% drop_na(ABV) %>% group_by(State) %>% summarize(mABV = max(ABV))
maxABVState <- as.character(maxABVbyState[order(-maxABVbyState$mABV),][1,1])
maxABVState
```

```
## [1] "CO"
```

State with most bitter bear (max IBU)

```
maxIBUbyState <- bd %>% drop_na(IBU) %>% group_by(State) %>% summarize(mIBU = max(IBU))
maxIBUState <- as.character(maxIBUbyState[order(-maxIBUbyState$mIBU),][1,1])
maxIBUState
```

```
## [1] "OR"
```

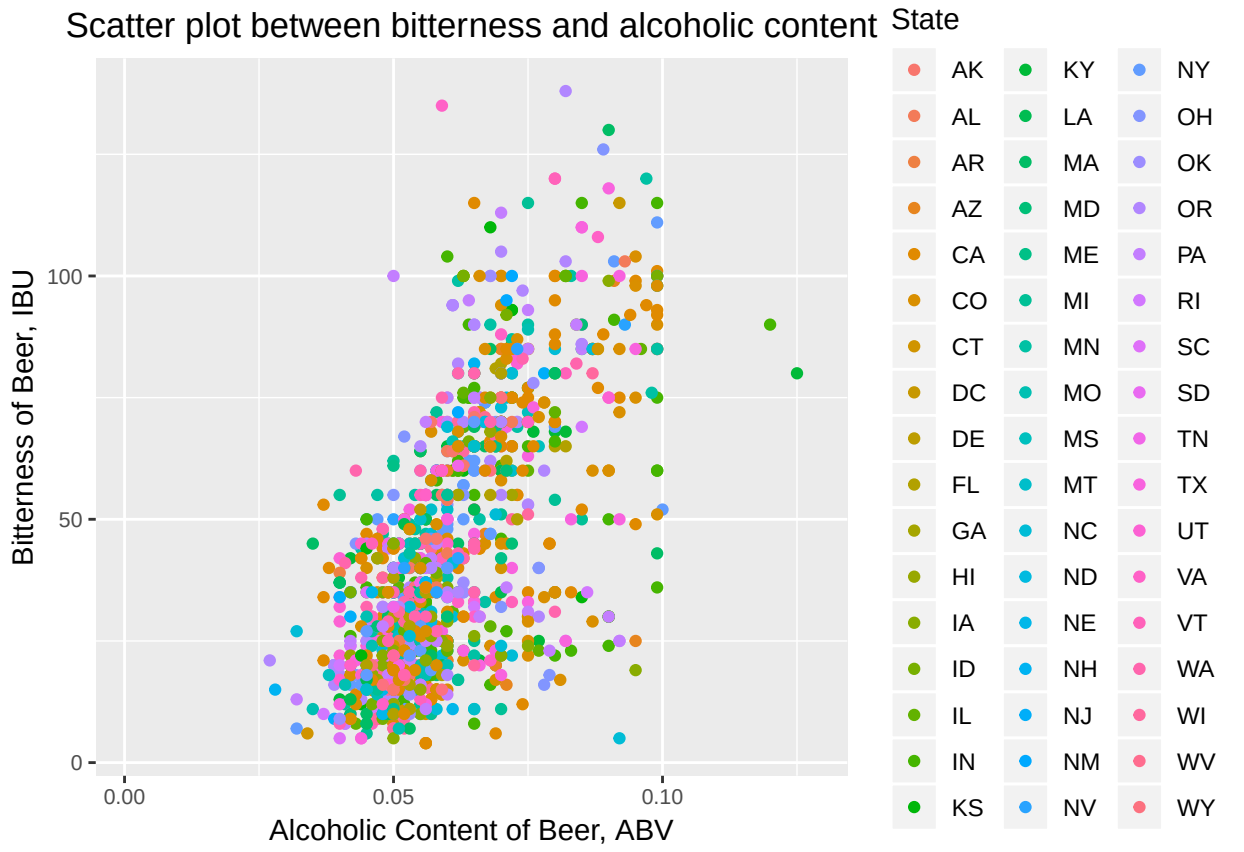
Summary Statistics for ABV variable

```
summary(beers$ABV)
```

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

Relationship between Beer Bitterness and Alcoholic Content

```
ggplot(bd, aes(x=ABV, y=IBU, color=State)) +
  geom_point() +
  xlab("Alcoholic Content of Beer, ABV")+
  ylab("Bitterness of Beer, IBU") +
  theme(axis.text = element_text(size = 8)) +
  ggtitle('Scatter plot between bitterness and alcoholic content') +
  theme(plot.title = element_text(hjust = 0.5))
```



Following aspects of the relationship are evident

1. In general, higher bitterness tracks with higher alcholic content.

2. It appears that one can have a range of bitterness for a given alcoholic content, and vice versa.

There is more to getting the right bitterness (IBU) for a beer than just it's alcoholic content (ABV)

Conclusions

- Colorado by far has the largest number of breweries.
  - CA, MI, IN, TX are other states with high density of breweries.
  - *Targeting these states for additional opening of brewery should be considered.*
- Co ranks 5th in list of states with highest alcholol content producing states.
  - However, it produces beer with highest alcholic content.
  - *Perhaps there is additional apetite in market to absorb higher alcholic content beer.*
- Clearly bitterness and alcoholic content track each other.
  - However, beers with different amount of bitterness can be produced at any given alcoholic content.
  - *One can be flexible in tailoring beer composition to suite whats popular in local market*

GitHub Reference

https://github.com/gkapila07/msds_homeworks/tree/master/dds