Case Study 01

Alexy Morris

10/11/2021

With the Beer and Breweries data set provided by you, the CEO and CFO of Budweiser, our team was able to answer the questions you presented and have even addressed some additional questions that arose. As a disclaimer we are working on a relatively local scale so all information should only be generalized the the United States of America, our team would not recommend utilizing the same information abroad. Overall our team firmly believes that the following work will reflect the current and future trends in the craft beer market.

1-How many breweries are present in each state?

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

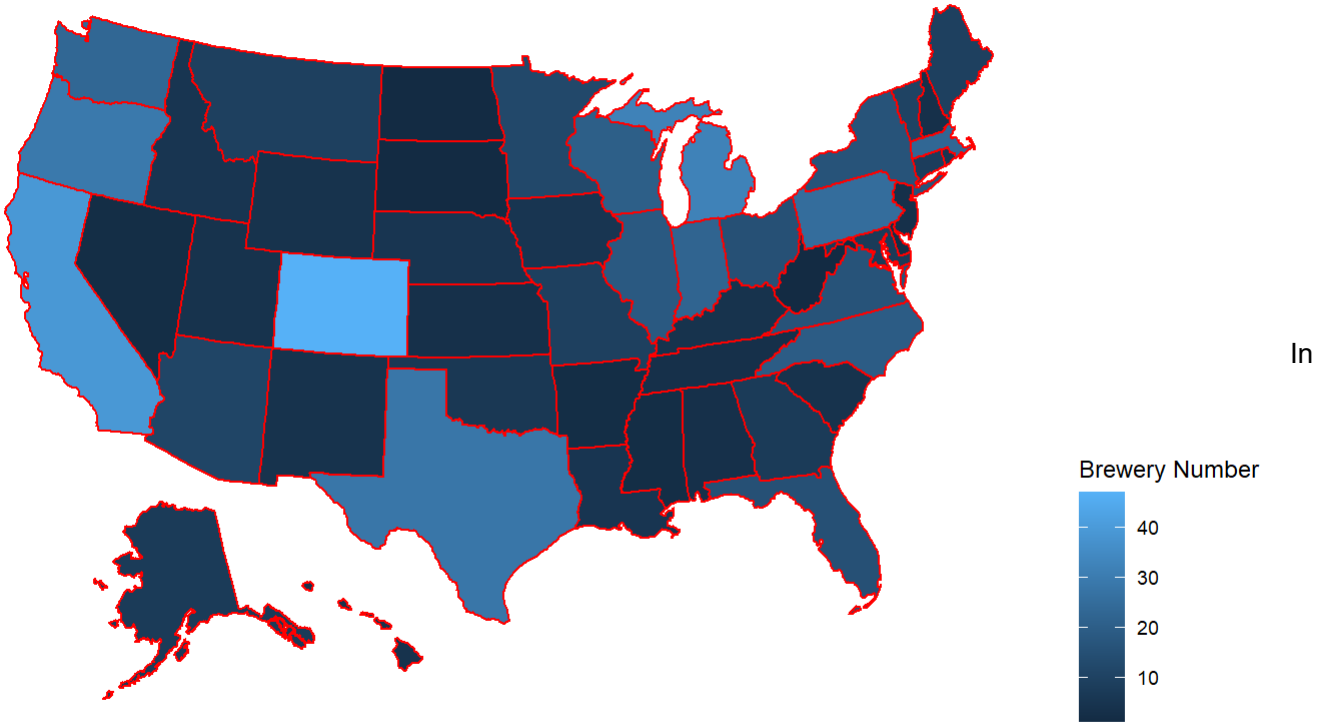
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
## [1] "C:/Users/Alex M/Documents"
```

```
## # A tibble: 51 x 2
##    State     n
##    <chr> <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
```

```
## # A tibble: 51 x 2
##    State           n
##    <chr>       <int>
##  1 Alaska          7
##  2 Alabama         3
##  3 Arkansas        2
##  4 Arizona        11
##  5 California     39
##  6 Colorado       47
##  7 Connecticut     8
##  8 <NA>            1
##  9 Delaware        2
## 10 Florida        15
## # ... with 41 more rows
```



In this section, we automated the counting process using the 51 states as categories and then counted the number of times said category appears in our data.

Through the 50 states we see that the highest number of breweries is in Colorado with 47 and one of the lowest is in South Dakota with 1. That said most states are in the middle with between about 10 to 30.

2-Merge beer data with the breweries data. Print the first 6 observations and the last six observations to check the merged file.

```
##   Brew_ID     Beer_Name Beer_ID   ABV IBU                             Style Ounces
Name        City State
## 1       1  Get Together    2692 0.045  50                       American IPA     16 NorthGat
e Brewing Minneapolis    MN
## 2       1 Maggie's Leap    2691 0.049  26                  Milk / Sweet Stout     16 NorthGat
e Brewing Minneapolis    MN
## 3       1    Wall's End    2690 0.048  19                   English Brown Ale     16 NorthGat
e Brewing Minneapolis    MN
## 4       1       Pumpion    2689 0.060  38                        Pumpkin Ale     16 NorthGat
e Brewing Minneapolis    MN
## 5       1    Stronghold    2688 0.060  25                     American Porter     16 NorthGat
e Brewing Minneapolis    MN
## 6       1   Parapet ESB    2687 0.056  47 Extra Special / Strong Bitter (ESB)     16 NorthGat
e Brewing Minneapolis    MN
```

```
##      Brew_ID                Beer_Name Beer_ID   ABV IBU                   Style Ounces
Name          City
## 2405     556             Pilsner Ukiah      98 0.055  NA         German Pilsener     12
Ukiah Brewing Company         Ukiah
## 2406     557  Heinnieweisse Weissebier      52 0.049  NA              Hefeweizen     12
Butternuts Beer and Ale Garrattsville
## 2407     557           Snapperhead IPA      51 0.068  NA            American IPA     12
Butternuts Beer and Ale Garrattsville
## 2408     557         Moo Thunder Stout      50 0.049  NA      Milk / Sweet Stout     12
Butternuts Beer and Ale Garrattsville
## 2409     557         Porkslap Pale Ale      49 0.043  NA American Pale Ale (APA)     12
Butternuts Beer and Ale Garrattsville
## 2410     558 Urban Wilderness Pale Ale      30 0.049  NA        English Pale Ale     12 Sleep
ing Lady Brewing Company     Anchorage
##      State
## 2405    CA
## 2406    NY
## 2407    NY
## 2408    NY
## 2409    NY
## 2410    AK
```

For convenience we combined the 2 dataframes given by matching up the brewery id numbers. We also took the time to rename a few columns for easy identification.

3-Address the missing values in each column

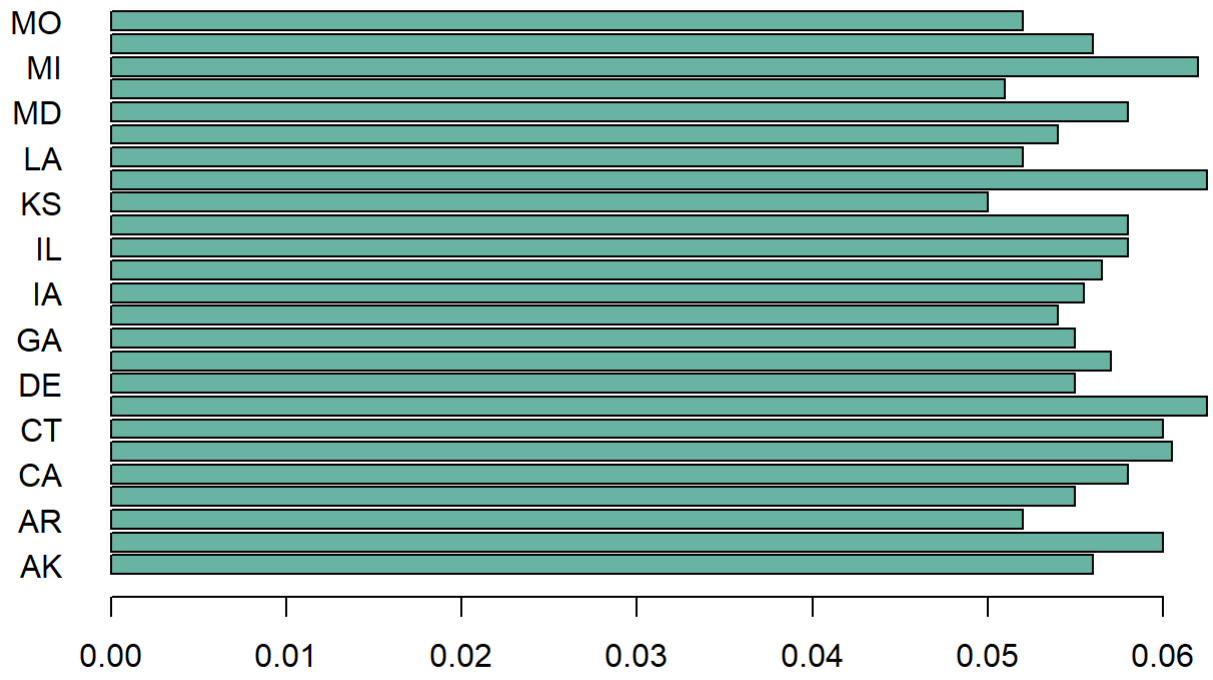
```
##   Brew_ID Beer_Name   Beer_ID       ABV       IBU     Style    Ounces      Name      City
State
##         0         0         0        62      1005         5         0         0         0
0
```

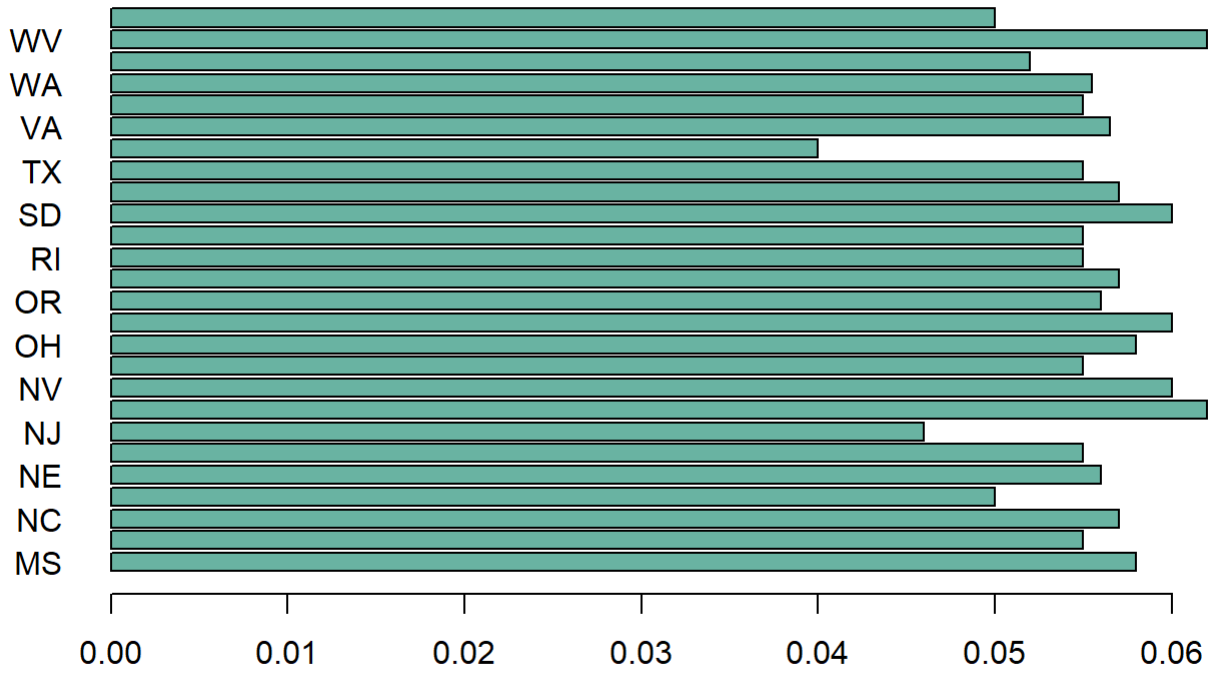
In this sections code we simply told our program to count/sum every missing value in each column.

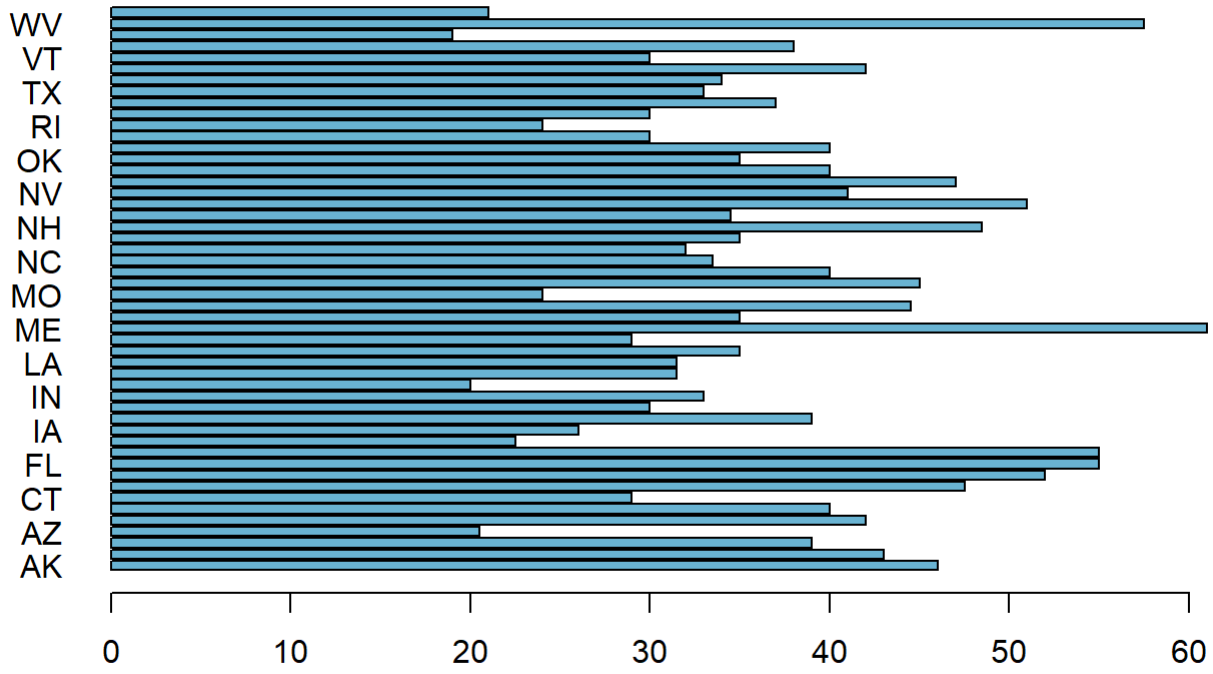
While giving the data a cursory glance our team noticed a number of missing values specifically in the IBU and ABV columns; unfortunately upon closer examination we see there are 1000+ missing values across our columns. The volume of missing vales means that it would be incredibly labor intensive on our teams part to fill the blanks

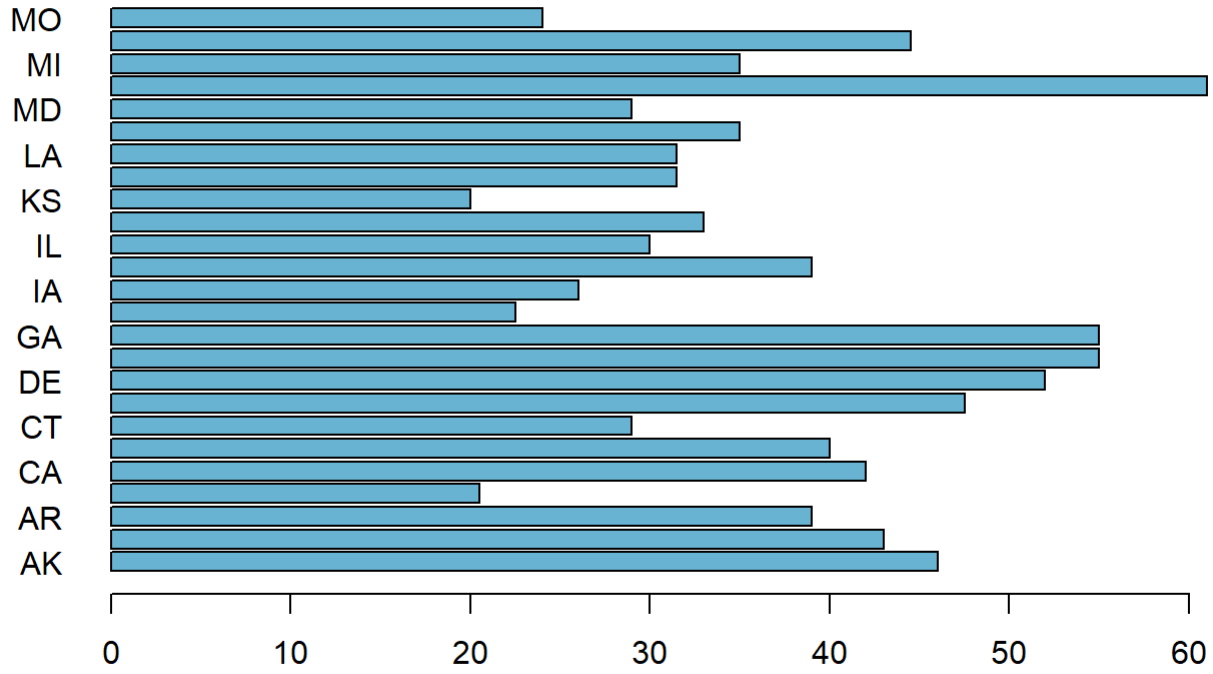
manually.

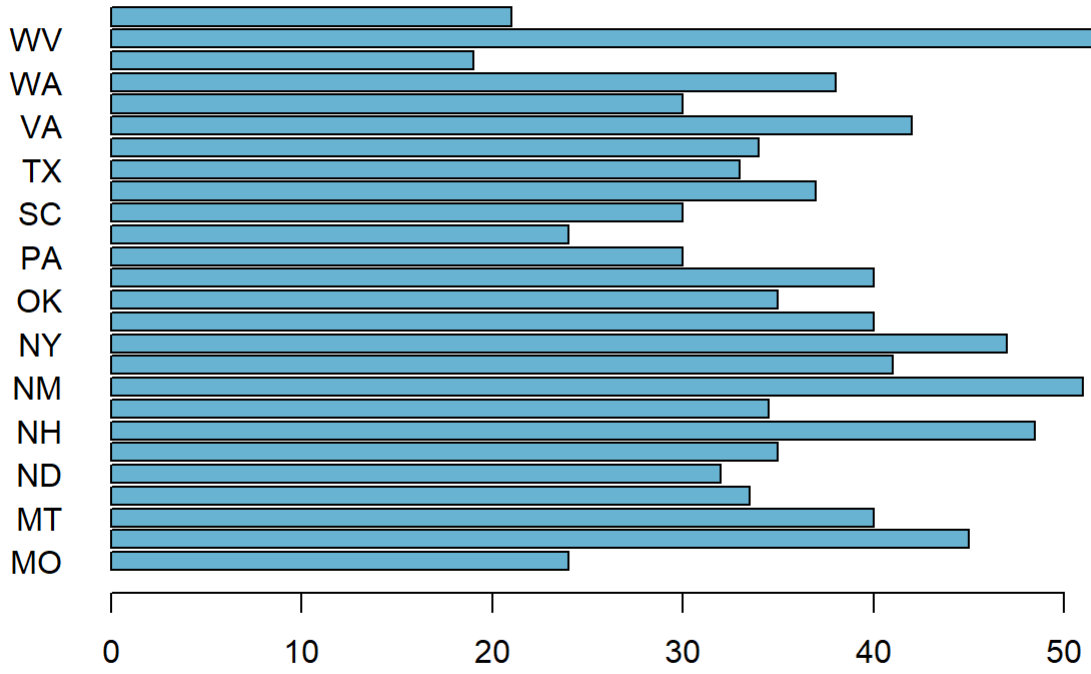
4-Compute the median alcohol content and international bitterness unit for each state. Plot a bar chart to compare.






WV
VT
TX
RI
OK
NV
NH
NC
MO
ME
LA
IN
IA
FL
CT
AZ
AK
0
10
20
30
40
50
60


MO
MI
MD
LA
KS
IL
IA
GA
DE
CT
CA
AR
AK
0
10
20
30
40
50
60



To

compute the median ABV and IBU level for each state we first broke down the larger given dataframe to only the needed variables in order to save on computation times. Then we got to the relevant calculations and lastly broke the medians into two data frames.

Our bar graphs revealed that, in regards to median, ABV is mostly consistent across state lines but on the other hand there was a lot of variety when it came to IBU.

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

```
##   Brew_ID                                          Beer_Name Beer_ID   ABV IBU
Style Ounces
## 1      52 Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale    2565 0.128  NA       Quadru
pel (Quad)   19.2
## 2       2                                   London Balling    2685 0.125  80     English
Barleywine   16.0
## 3      18                                             Csar    2621 0.120  90 Russian Impe
rial Stout   16.0
## 4      52     Lee Hill Series Vol. 4 - Manhattan Style Rye Ale    2564 0.104  NA
Rye Beer   19.2
## 5      47                                           4Beans    2574 0.100  52          Bal
tic Porter   12.0
##                         Name       City State
## 1   Upslope Brewing Company    Boulder    CO
## 2 Against the Grain Brewery Louisville    KY
## 3   Tin Man Brewing Company Evansville    IN
## 4   Upslope Brewing Company    Boulder    CO
## 5       Sixpoint Craft Ales   Brooklyn    NY
```

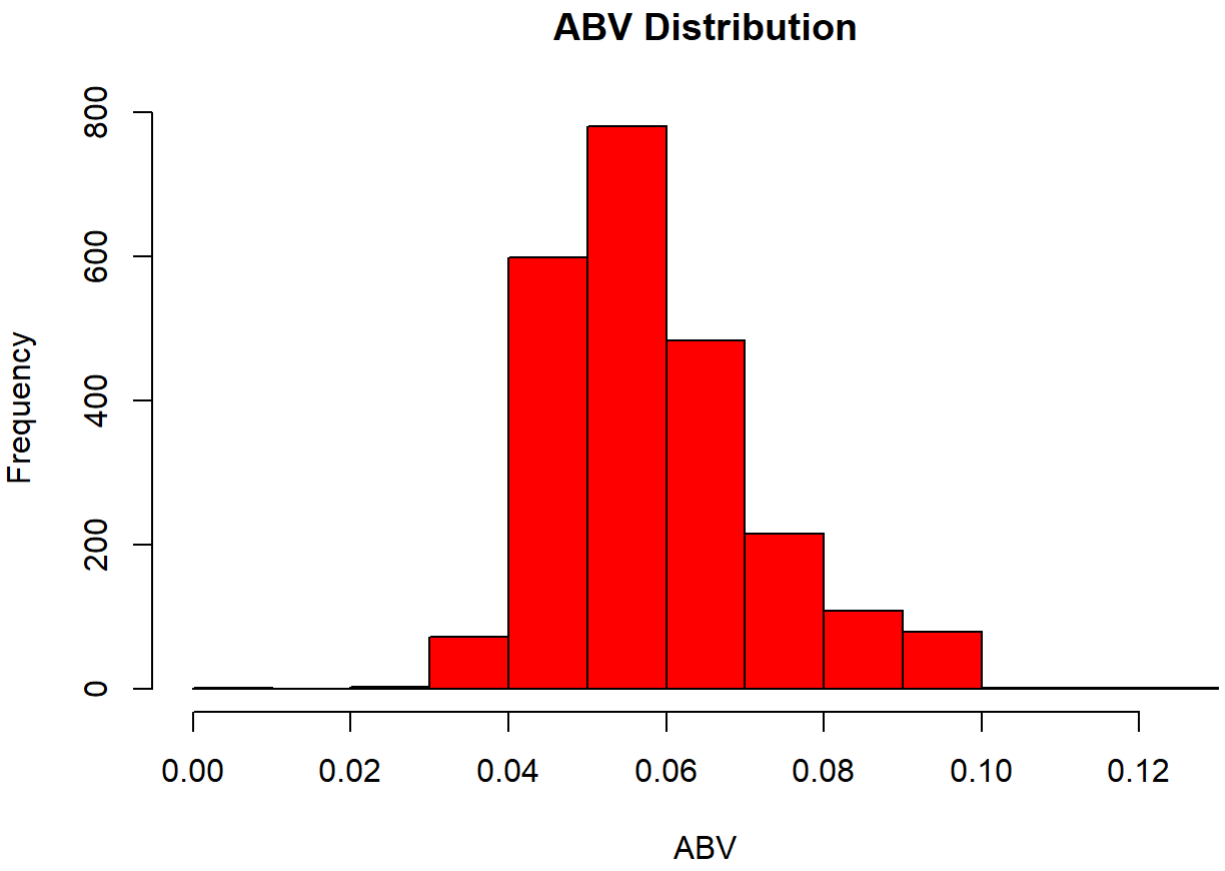
```
##   Brew_ID                      Beer_Name Beer_ID   ABV IBU                         Style Ou
nces                              Name
## 1     375       Bitter Bitch Imperial IPA     980 0.082 138 American Double / Imperial IPA
12            Astoria Brewing Company
## 2     345              Troopers Alley IPA    1676 0.059 135                  American IPA
12         Wolf Hills Brewing Company
## 3     231                   Dead-Eye DIPA    2067 0.090 130 American Double / Imperial IPA
16           Cape Ann Brewing Company
## 4     100 Bay of Bengal Double IPA (2014)    2440 0.089 126 American Double / Imperial IPA
12 Christian Moerlein Brewing Company
## 5      62                    Abrasive Ale      15 0.097 120 American Double / Imperial IPA
16              Surly Brewing Company
## 6     273                    Heady Topper    1111 0.080 120 American Double / Imperial IPA
16                      The Alchemist
## 7     273                    Heady Topper     379 0.080 120 American Double / Imperial IPA
16                      The Alchemist
##              City State
## 1         Astoria    OR
## 2        Abingdon    VA
## 3      Gloucester    MA
## 4      Cincinnati    OH
## 5 Brooklyn Center    MN
## 6       Waterbury    VT
## 7       Waterbury    VT
```

For the fifth question we simply reorganized the graphs from greatest ABV to least ABV and again from greatest to least for IBU.

The state with the highest ABV is Colorado and the state with the highest IBU is Oregon.

6-Comment on the summary statistics and distribution of the ABV variable.

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



By

using the baseline functions of R we get a range of standard statistical values and even a histogram based off of our data.

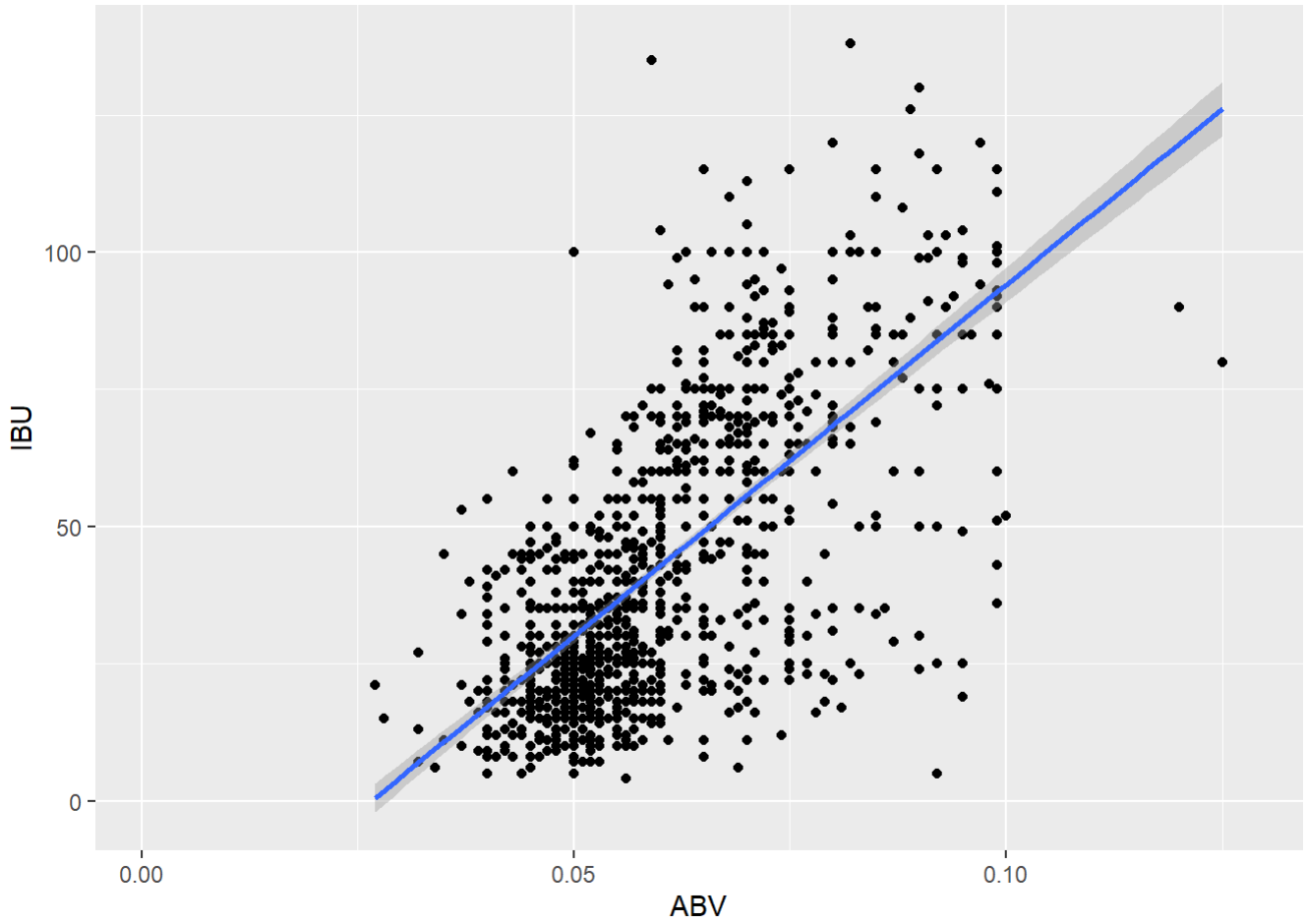
Both of graph and our summary statistics paint a picture of a normal distribution. Our mean and median are very close together and our 1st and 3rd quadrilles appear to be inline with the expectations of a normal distribution.

7- Is there an apparent relationship between the bitterness of the beer and its alcoholic content? Draw a scatter plot.

```
## `geom_smooth()` using formula 'y ~ x'
```

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

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



To address the relationship between ABV and IBU we have graphed IBU with respect to ABV and created a linear model to highlight the direct relationship between the two.

We notice that there does appear to be a direct positive relationship between ABV and IBU.

8-Budweiser would also like to investigate the difference with respect to IBU and ABV between IPAs (India Pale Ales) and other types of Ale (any beer with “Ale” in its name other than IPA). You decide to use KNN classification to investigate this relationship. Provide statistical evidence one way or the other. You can of course assume your audience is comfortable with percentages … KNN is very easy to understand conceptually.

```
## -- Attaching packages ---------------------------------------------------------------
----------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4     v purrr   0.3.4
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------------------------------------
------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
## [1] TRUE
## attr(,"prob")
## [1] 0.8571429
## Levels: FALSE TRUE
```

```
## [1] TRUE
## attr(,"prob")
## [1] 0.8888889
## Levels: FALSE TRUE
```

For our KNN classification we have taken the essentially the same scatter plot from above and plotted different test points to examine the likely of a beer with our same ABV and IBU being a IPA or an Ale.

Form a number of test we see that with our given test we were much more likely to get IPA as True in the case of a high IBU with the ABV seemingly not mattering.

In conclusion, through our research into the beer/brewery data frames we have uncovered many answer to questions both requested by you, our client, and observed from notable trends. We hope these results provided the desired conclusions and in the event your company wants for a deeper analyzation we hope you will keep our team in mind.