Study of US Craft Beer and Breweries

David Stroud, Cory Thigpen, Korey MacVittie, and Stacy Conant

February 2018

```
# Read in beer and breweries data set using _csv for more tidy output
Beers <- read_csv('Beers.csv', col_types = cols())
Breweries <- read_csv('Breweries.csv', col_types = cols())
```

```
# 1.How many breweries are present in each state?
table(Breweries$State)
```

```
##
## AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO
##  7  3  2 11 39 47  8  1  2 15  7  4  5  5 18 22  3  4  5 23  7  9 32 12  9
## MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV
##  2  9 19  1  5  3  3  4  2 16 15  6 29 25  5  4  1  3 28  4 16 10 23 20  1
## WY
##  4
```

```
# Renamed Brewery_id to Brew_ID to satisfy merging requirement
Beers <- rename(Beers, Brew_ID = Brewery_id)
```

```
# 2. Merge beer data with the breweries data.
BrewPub <- full_join(Beers, Breweries, by="Brew_ID")
#(dplyr::tbl_df(BrewPub))
```

```
# 2.a Print the first six observations
print(BrewPub %>% head())
```

```
## # A tibble: 6 x 10
##   Name.x   Beer_ID    ABV   IBU Brew_ID Style   Ounces Name.y  City  State
##   <chr>      <int>  <dbl> <int>   <int> <chr>    <dbl> <chr>   <chr> <chr>
## 1 Pub Beer    1436 0.0500    NA     409 Americ~   12.0 10 Bar~ Bend  OR
## 2 Devil's~    2265 0.0660    NA     178 Americ~   12.0 18th S~ Gary  IN
## 3 Rise of~    2264 0.0710    NA     178 Americ~   12.0 18th S~ Gary  IN
## 4 Sinister    2263 0.0900    NA     178 Americ~   12.0 18th S~ Gary  IN
## 5 Sex and~    2262 0.0750    NA     178 Americ~   12.0 18th S~ Gary  IN
## 6 Black E~    2261 0.0770    NA     178 Oatmea~   12.0 18th S~ Gary  IN
```

```
# 2. b Print the last six observations
print(BrewPub %>% tail())
```

```
## # A tibble: 6 x 10
##   Name.x   Beer_ID    ABV   IBU Brew_ID Style   Ounces Name.y  City  State
##   <chr>      <int>  <dbl> <int>   <int> <chr>    <dbl> <chr>   <chr> <chr>
## 1 Rocky M~    1035 0.0750    NA     425 Americ~   12.0 Wynkoo~ Denv~ CO
## 2 Belgora~     928 0.0670    45     425 Belgia~   12.0 Wynkoo~ Denv~ CO
## 3 Rail Ya~     807 0.0520    NA     425 Americ~   12.0 Wynkoo~ Denv~ CO
## 4 B3K Bla~     620 0.0550    NA     425 Schwar~   12.0 Wynkoo~ Denv~ CO
## 5 Silverb~     145 0.0550    40     425 Americ~   12.0 Wynkoo~ Denv~ CO
## 6 Rail Ya~      84 0.0520    NA     425 Americ~   12.0 Wynkoo~ Denv~ CO
```

```
# 3. Report the number of NA's in each column.
MissingValues <- sapply(BrewPub, function(x)sum(is.na(x)))
```

```
print(MissingValues)
```

```
##  Name.x Beer_ID     ABV     IBU Brew_ID   Style  Ounces  Name.y    City
##       0       0      62    1005       0       5       0       0       0
##   State
##       0
```

```
# 4. Compute the median alcohol content unit for each state.
Bitter <- BrewPub %>%
  na.omit() %>%
group_by(State) %>%
summarise(Median = median(ABV)) %>%
arrange(Median)
Bitter %>% tbl_df %>% print(n=50)
```

```
## # A tibble: 50 x 2
##    State Median
##    <chr>  <dbl>
##  1 AR    0.0400
##  2 UT    0.0400
##  3 NJ    0.0460
##  4 NH    0.0465
##  5 KS    0.0500
##  6 MO    0.0500
##  7 ND    0.0500
##  8 SC    0.0500
##  9 WI    0.0510
## 10 LA    0.0510
## 11 WY    0.0510
## 12 HI    0.0520
## 13 RI    0.0525
## 14 MA    0.0540
## 15 AZ    0.0550
## 16 DE    0.0550
## 17 NV    0.0550
## 18 TN    0.0550
## 19 TX    0.0550
## 20 VT    0.0550
## 21 MN    0.0555
## 22 IA    0.0560
## 23 MI    0.0560
## 24 NE    0.0560
## 25 OR    0.0560
## 26 WA    0.0560
## 27 MD    0.0565
## 28 AK    0.0570
## 29 IL    0.0570
## 30 IN    0.0570
## 31 MT    0.0570
## 32 PA    0.0570
## 33 VA    0.0570
## 34 KY    0.0575
## 35 OH    0.0575
## 36 CA    0.0580
```

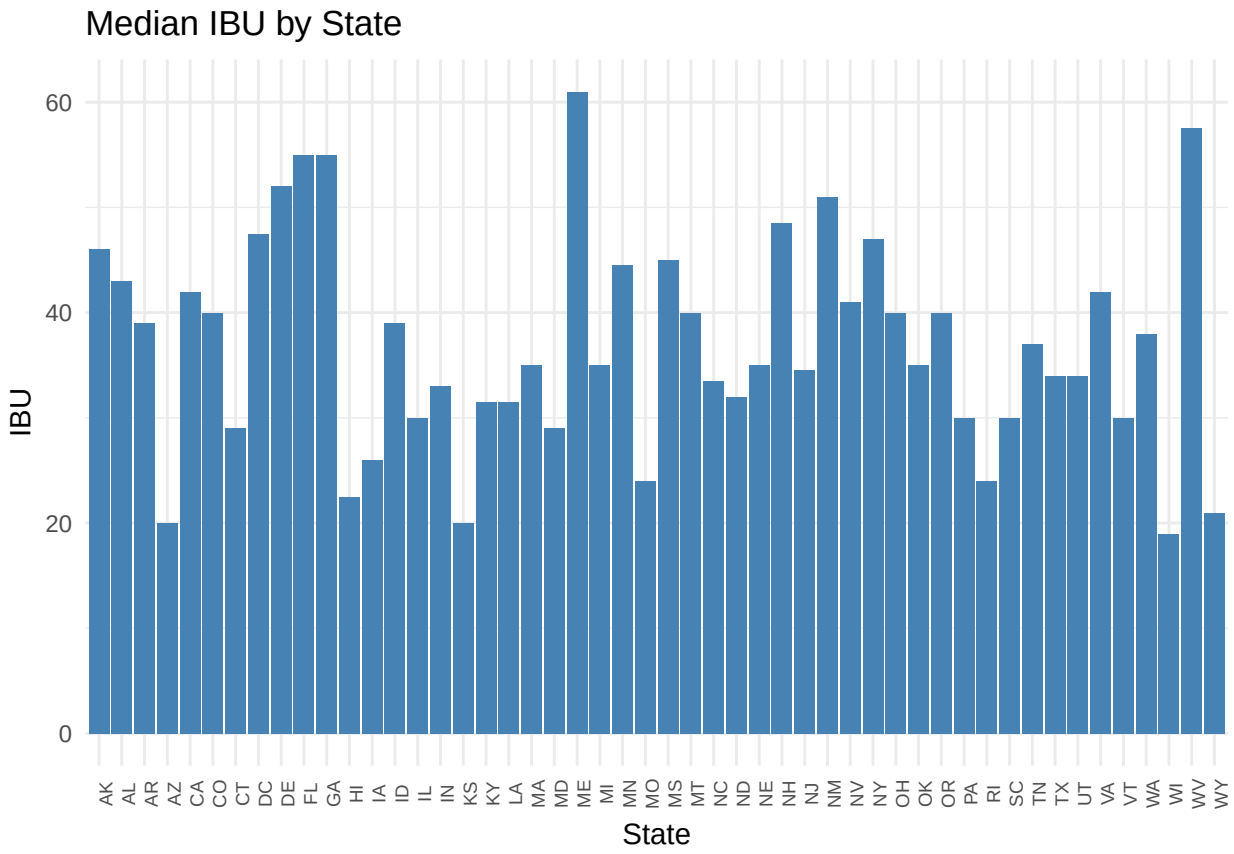
```
## 37 ID    0.0580
## 38 MS    0.0580
## 39 DC    0.0590
## 40 NY    0.0595
## 41 AL    0.0600
## 42 CT    0.0610
## 43 NC    0.0610
## 44 NM    0.0610
## 45 FL    0.0620
## 46 GA    0.0620
## 47 WV    0.0620
## 48 OK    0.0630
## 49 CO    0.0650
## 50 ME    0.0670
```

```
# 4. a Compute the median alcohol content unit for each state.
FireWater <- BrewPub %>%
  na.omit() %>%
group_by(State) %>%
summarise(Median = median(IBU)) %>%
arrange(Median)
FireWater %>% tbl_df %>% print(n=50)
```

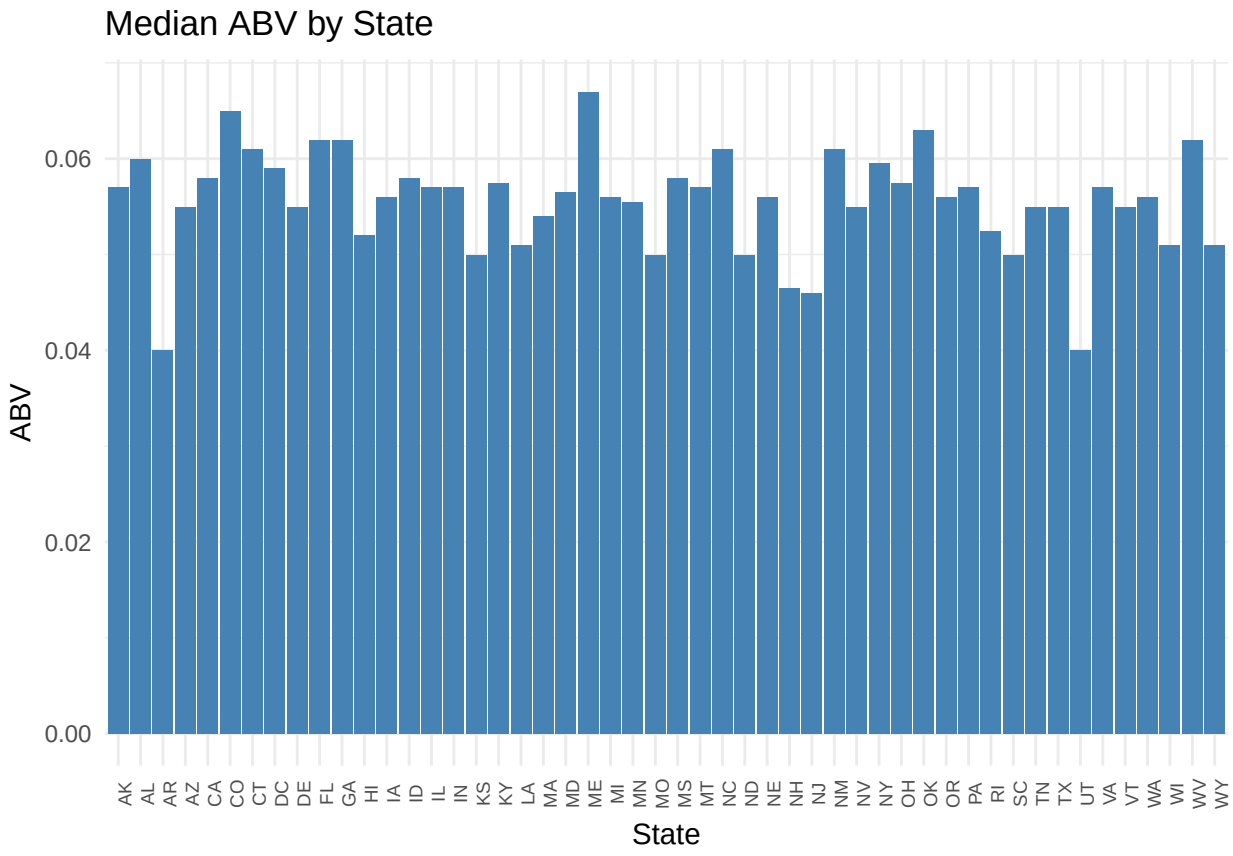
```
## # A tibble: 50 x 2
##    State Median
##    <chr>  <dbl>
##  1 WI      19.0
##  2 AZ      20.0
##  3 KS      20.0
##  4 WY      21.0
##  5 HI      22.5
##  6 MO      24.0
##  7 RI      24.0
##  8 IA      26.0
##  9 CT      29.0
## 10 MD      29.0
## 11 IL      30.0
## 12 PA      30.0
## 13 SC      30.0
## 14 VT      30.0
## 15 KY      31.5
## 16 LA      31.5
## 17 ND      32.0
## 18 IN      33.0
## 19 NC      33.5
## 20 TX      34.0
## 21 UT      34.0
## 22 NJ      34.5
## 23 MA      35.0
## 24 MI      35.0
## 25 NE      35.0
## 26 OK      35.0
## 27 TN      37.0
## 28 WA      38.0
## 29 AR      39.0
```

```
## 30 ID      39.0
## 31 CO      40.0
## 32 MT      40.0
## 33 OH      40.0
## 34 OR      40.0
## 35 NV      41.0
## 36 CA      42.0
## 37 VA      42.0
## 38 AL      43.0
## 39 MN      44.5
## 40 MS      45.0
## 41 AK      46.0
## 42 NY      47.0
## 43 DC      47.5
## 44 NH      48.5
## 45 NM      51.0
## 46 DE      52.0
## 47 FL      55.0
## 48 GA      55.0
## 49 WV      57.5
## 50 ME      61.0
```

```r
# 4. b Plot a bar chart to compare IBU by state
ggplot(data=FireWater, aes(x=State, y=Median)) +
  geom_bar(stat="identity", fill="steelblue")+
  theme_minimal() +
  theme(axis.text.x=element_text(size=rel(0.8), angle=90)) +
  ggtitle("Median IBU by State") +
  labs(x="State",y="IBU")
```




```
# 4. c Plot a bar chart to compare ABV by state
ggplot(data=Bitter, aes(x=State, y=Median)) +
  geom_bar(stat="identity", fill="steelblue")+
  theme_minimal() +
  theme(axis.text.x=element_text(size=rel(0.8), angle=90))+
  ggtitle("Median ABV by State") +
  labs(x="State",y="ABV")
```




```
# 5. Which state has the maximum alcoholic (ABV) beer?
print(BrewPub[which.max(BrewPub$ABV),])
```

```
## # A tibble: 1 x 10
##   Name.x     Beer_ID   ABV   IBU Brew_ID Style  Ounces Name.y  City  State
##   <chr>        <int> <dbl> <int>   <int> <chr>   <dbl> <chr>   <chr> <chr>
## 1 Lee Hill ~    2565 0.128    NA      52 Quadr~   19.2 Upslop~ Boul~ CO
```

```
# 5. a Which state has the most bitter (IBU) beer?
print(BrewPub[which.max(BrewPub$IBU),])
```

```
## # A tibble: 1 x 10
##   Name.x   Beer_ID    ABV   IBU Brew_ID Style   Ounces Name.y  City  State
##   <chr>      <int>  <dbl> <int>   <int> <chr>    <dbl> <chr>   <chr> <chr>
## 1 Bitter ~     980 0.0820   138     375 Americ~   12.0 Astori~ Asto~ OR
```

```
# 6. Summary Statistics for ABV variable
summary(BrewPub$ABV)
```

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

```
# 7. Draw a scatter plot to compare relationship between beer
# bitterness and alcohol content
ggplot(BrewPub, aes(x=IBU, y= ABV)) +
  geom_point(shape=1) +
   geom_smooth(method=lm) + # add linear regression line
  theme(axis.text.x=element_text(size=rel(1.0)))+
ggtitle("Correlation between IBU and ABV ") +
```

```
labs(x="IBU",y="ABV")
```