

LIS 4317 Final Project

Patricia McGee

2022-04-26

The Assignment

The Project outline

Each student in the class will create his/her visual project. The student will share the link on Canvas. The grade for the final project will be scored based on your instructor and your classmates peer review comments and grades.

Objectives

The purpose of this project is to apply what you've learned in this class: design and statistical analysis apparatus. Your project needs to capture any type of data based on your interest and utilize plot.ly, Tableau or Rstudio. The software you picked has to be covered in class in order to tell a story of your data analysis you selected. You can employ any of the techniques and practices we outline in the class. From Time series analysis, part to whole, ranking, deviation analysis, distribution, correlation, multivariate, and even promising trends as discussed in Chapter 13 in our textbook.

Step 1. Begin searching for a dataset that captivates your interest. Your dataset should contain at least 50 observations and 5 to 20 variables.

Step 2. You will provide a written documentation of your project objectives and accomplishments that will include a general research question and objectives. Those questions will contain your underlying theme of your project as outline below.

Short summary:

Problem description: Give a brief but precise description or definition of the problem or hypothesis you set to evaluate. Related work: How does this problem and the method relate to the problems/methods have addressed in the Provide example of existing visual analytics that you liked. Solution: How did you solve the problem? Describe the technical approach. Tell us what methodology you will take: Time series analysis, part to whole, ranking, deviation analysis, distribution, correlation, multivariate, and even promising trends.

Step 3. Create a visualization of your data and discussion to your output. You can use any of the apps we covered in the class including: plot.ly, Tableau, RStudio and Python.

My Information and Code

I have decided to use the Beer Profile and Ratings Dataset from Kaggle: <https://www.kaggle.com/ruthgn/beer-profile-and-ratings-data-set>

I will read in the csv file "beer_profile_and_ratings.csv" and save this as beers.

```
beers <- read.csv("beer_profile_and_ratings.csv")
head(beers)
```

```
##           i..Name  Style
## 1           Amber Altbier
## 2           Double Bag Altbier
## 3           Long Trail Ale Altbier
## 4           Doppelsticke Altbier
## 5 Sleigh'r Dark DoÃ¼ble Alt Ale Altbier
## 6           Sticke Altbier
##
##           Brewery
## 1           Alaskan Brewing Co.
## 2           Long Trail Brewing Co.
## 3           Long Trail Brewing Co.
## 4 Uerige ObergÃ¼rige Hausbrauerei GmbH / Zum Uerige
## 5           Ninkasi Brewing Company
## 6 Uerige ObergÃ¼rige Hausbrauerei GmbH / Zum Uerige
##
##           Beer.Name..Full.
## 1           Alaskan Brewing Co. Alaskan Amber
## 2           Long Trail Brewing Co. Double Bag
## 3           Long Trail Brewing Co. Long Trail Ale
## 4 Uerige ObergÃ¼rige Hausbrauerei GmbH / Zum Uerige Uerige Doppelsticke
## 5           Ninkasi Brewing Company Sleigh'r Dark DoÃ¼ble Alt Ale
## 6           Uerige ObergÃ¼rige Hausbrauerei GmbH / Zum Uerige Uerige Sticke
##
## 1
## 2 Notes:This malty, full-bodied double alt is also known as âStickebierâ German
## 3
## 4
## 5
## 6
##  ABV Min.IBU Max.IBU Astringency Body Alcohol Bitter Sweet Sour Salty Fruits
## 1 5.3      25      50           13  32      9      47      74      33      0      33
## 2 7.2      25      50           12  57     18     33     55     16      0      24
## 3 5.0      25      50           14  37      6     42     43     11      0      10
## 4 8.5      25      50           13  55     31     47    101     18      1      49
## 5 7.2      25      50           25  51     26     44     45      9      1      11
## 6 6.0      25      50           22  45     13     46     62     25      1      34
##  Hoppy Spices Malty review_aroma review_appearance review_palate review_taste
## 1   57      8   111      3.498994      3.636821      3.556338      3.643863
## 2   35     12    84      3.798337      3.846154      3.904366      4.024948
## 3   54      4    62      3.409814      3.667109      3.600796      3.631300
## 4   40     16   119      4.148098      4.033967      4.150815      4.205163
## 5   51     20    95      3.625000      3.973958      3.734375      3.765625
## 6   60      4   103      4.007937      4.007937      4.087302      4.192063
##  review_overall number_of_reviews
## 1      3.847082           497
## 2      4.034304           481
## 3      3.830239           377
## 4      4.005435           368
## 5      3.817708            96
## 6      4.230159          315
```

```
summary(beers)
```

```
##      i..Name          Style          Brewery      Beer.Name..Full.
## Length:3197      Length:3197      Length:3197      Length:3197
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## Description          ABV          Min.IBU          Max.IBU
## Length:3197      Min.   : 0.000      Min.   : 0.00      Min.   : 0.00
## Class :character  1st Qu.: 5.000      1st Qu.:15.00      1st Qu.: 25.00
## Mode  :character  Median : 6.000      Median :20.00      Median : 35.00
##                      Mean   : 6.527      Mean   :21.18      Mean   : 38.99
##                      3rd Qu.: 7.600      3rd Qu.:25.00      3rd Qu.: 45.00
##                      Max.   :57.500      Max.   :65.00      Max.   :100.00
##
## Astringency        Body          Alcohol          Bitter
## Min.   : 0.00      Min.   : 0.00      Min.   : 0.00      Min.   : 0.00
## 1st Qu.: 9.00      1st Qu.: 29.00      1st Qu.: 6.00      1st Qu.: 17.00
## Median :14.00      Median : 40.00      Median : 11.00      Median : 31.00
## Mean   :16.52      Mean   : 46.13      Mean   : 17.06      Mean   : 36.36
## 3rd Qu.:21.00      3rd Qu.: 58.00      3rd Qu.: 22.00      3rd Qu.: 52.00
## Max.   :81.00      Max.   :175.00      Max.   :139.00      Max.   :150.00
##
## Sweet              Sour          Salty          Fruits
## Min.   : 0.00      Min.   : 0.00      Min.   : 0.000      Min.   : 0.00
## 1st Qu.: 33.00      1st Qu.: 11.00      1st Qu.: 0.000      1st Qu.: 12.00
## Median : 54.00      Median : 22.00      Median : 0.000      Median : 29.00
## Mean   : 58.27      Mean   : 33.15      Mean   : 1.017      Mean   : 38.53
## 3rd Qu.: 77.00      3rd Qu.: 42.00      3rd Qu.: 1.000      3rd Qu.: 60.00
## Max.   :263.00      Max.   :284.00      Max.   :48.000      Max.   :175.00
##
## Hoppy              Spices          Malty          review_aroma
## Min.   : 0.00      Min.   : 0.00      Min.   : 0.00      Min.   :1.510
## 1st Qu.: 18.00      1st Qu.: 4.00      1st Qu.: 45.00      1st Qu.:3.423
## Median : 33.00      Median : 10.00      Median : 73.00      Median :3.720
## Mean   : 40.92      Mean   : 18.35      Mean   : 75.33      Mean   :3.639
## 3rd Qu.: 56.00      3rd Qu.: 23.00      3rd Qu.:103.00      3rd Qu.:3.978
## Max.   :172.00      Max.   :184.00      Max.   :239.00      Max.   :5.000
##
## review_appearance  review_palate      review_taste      review_overall
## Min.   :1.571      Min.   :1.286      Min.   :1.214      Min.   :1.136
## 1st Qu.:3.605      1st Qu.:3.470      1st Qu.:3.500      1st Qu.:3.567
## Median :3.833      Median :3.742      Median :3.792      Median :3.830
## Mean   :3.754      Mean   :3.660      Mean   :3.702      Mean   :3.748
## 3rd Qu.:4.000      3rd Qu.:3.966      3rd Qu.:4.033      3rd Qu.:4.033
## Max.   :4.667      Max.   :5.000      Max.   :5.000      Max.   :5.000
##
## number_of_reviews
## Min.   : 1.0
## 1st Qu.: 23.0
## Median : 93.0
## Mean   : 233.3
## 3rd Qu.: 284.0
## Max.   :3290.0
```

Exploring and Cleaning the Data

First, we need to narrow this dataset down for our purposes. There are currently 3197 observations.

We remove some fields that we don't need. We will remove Description and the Min.IBU and Max.IBU.

Then we will filter this to only select Styles. I have chosen 8 beer styles that I enjoy. They are: Blonde Ale - Belgian Cream Ale Lager - American Pumpkin Beer Red Ale - Irish Stout - Irish Dry Stout - Sweet / Milk Wheat Beer - Hefeweizen

This brings us down to 248 observations.

```
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

beers <- beers[,-7:-8]
beers <- beers[,-5]

beers <- beers %>%
  rename(Name = i..Name, FullBeerName = Beer.Name..Full.)

beers_filtered <- beers %>%
  filter(Style %in% c("Blonde Ale - Belgian", "Cream Ale", "Lager - American",
                     "Pumpkin Beer", "Red Ale - Irish", "Stout - Irish Dry",
                     "Stout - Sweet / Milk", "Wheat Beer - Hefeweizen"))

top_beers <- beers_filtered %>%
  arrange(desc(review_overall))%>%
  top_n(n = 10, review_overall)

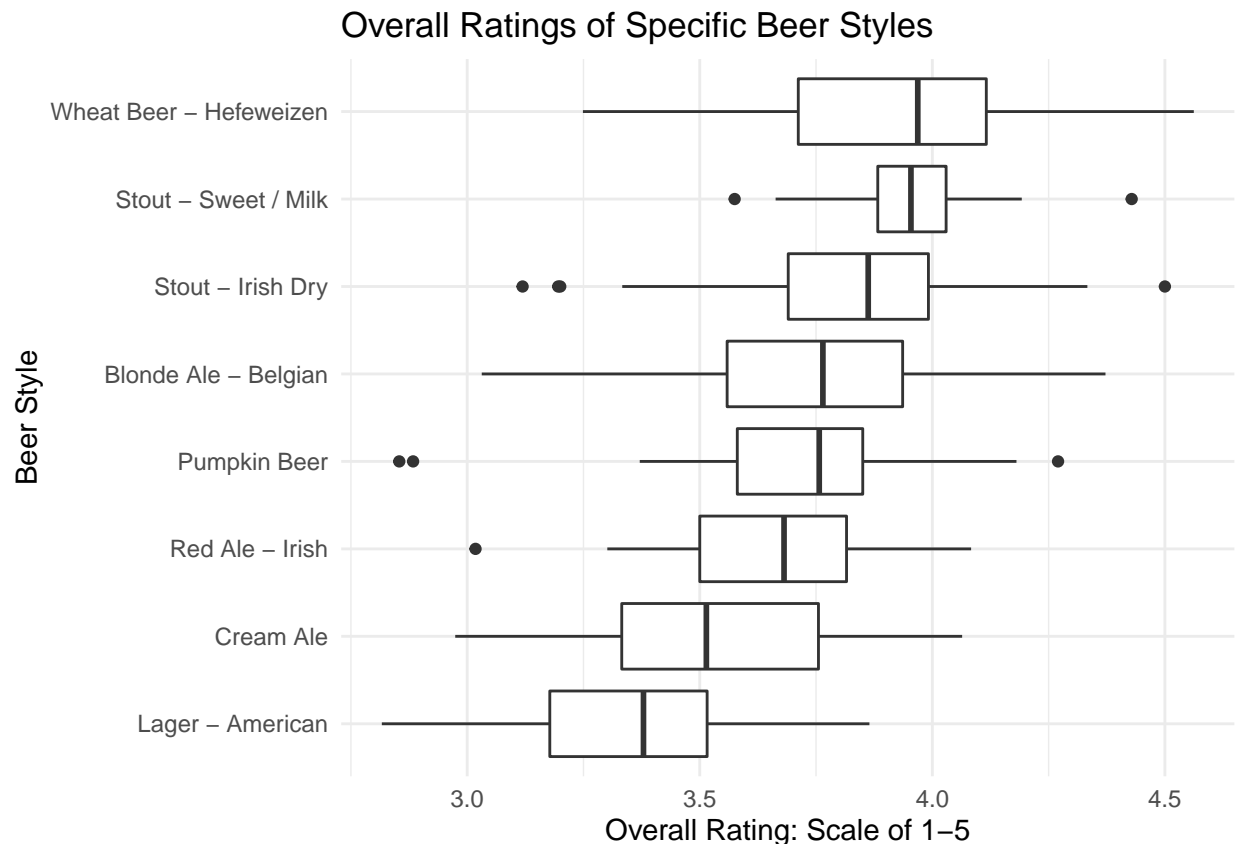
topABVbeers <- beers_filtered %>%
  arrange(desc(ABV))%>%
  top_n(n = 10, ABV)
```

Plots

I chose to create plots showing the top 10 beers by overall rating and the top 10 beers by ABV. I also created a box plot showing the range of the overall rating by beer style for the 8 selected styles and a linear regression line showing the ABV for each beer style.

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

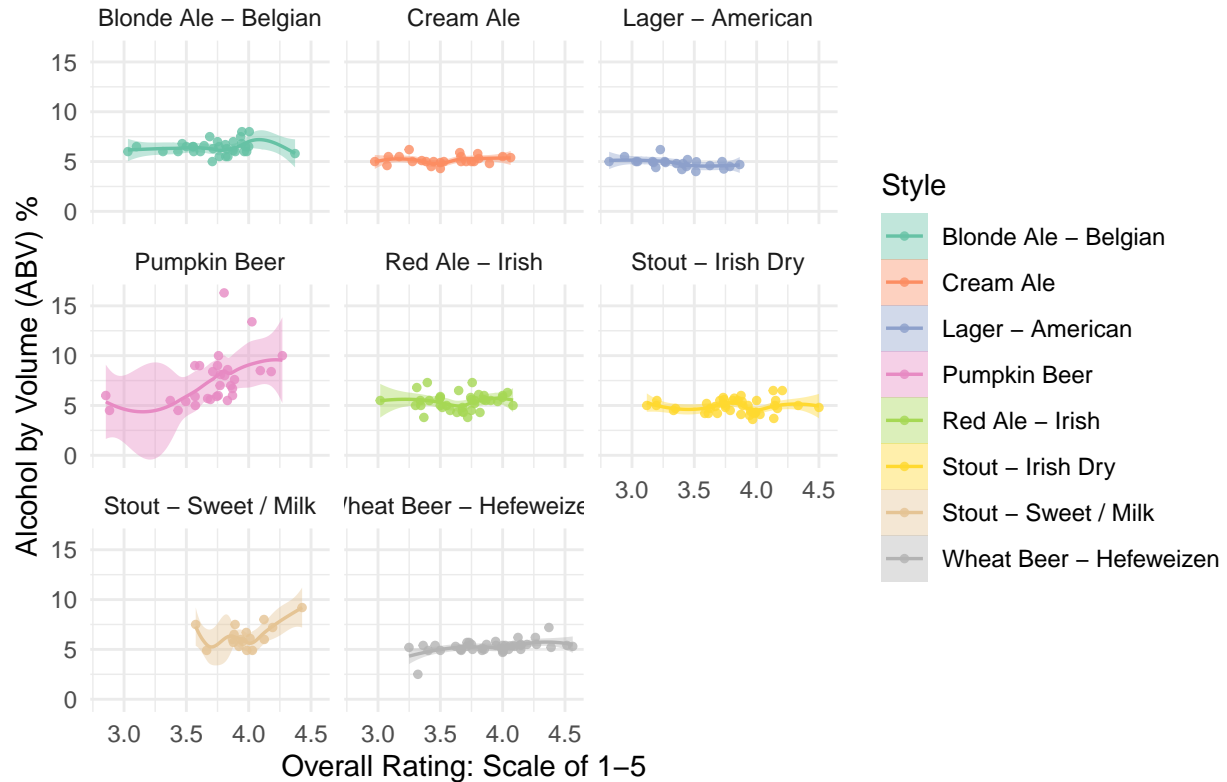
```
beers_filtered %>%
  mutate(BeerStyle = fct_reorder(Style, review_overall))%>%
  ggplot(aes(BeerStyle, review_overall))+
  geom_boxplot()+
  coord_flip()+
  theme_minimal()+
  labs(title = "Overall Ratings of Specific Beer Styles", x = "Beer Style",
       y = "Overall Rating: Scale of 1-5")
```



```
ggplot(beers_filtered, aes(review_overall, ABV))+
  geom_point(size = 1, alpha = 0.8, aes(color = Style))+
  geom_smooth(size = 0.6, aes(color = Style, fill = Style))+
  theme_minimal()+
  scale_color_brewer(palette = "Set2")+
  scale_fill_brewer(palette = "Set2")+
  facet_wrap(vars(Style))+
  labs(title = "Alcohol by Volume (ABV) and Overall Rating of Specific Beer Styles",
       x = "Overall Rating: Scale of 1-5", y = "Alcohol by Volume (ABV) %")
```

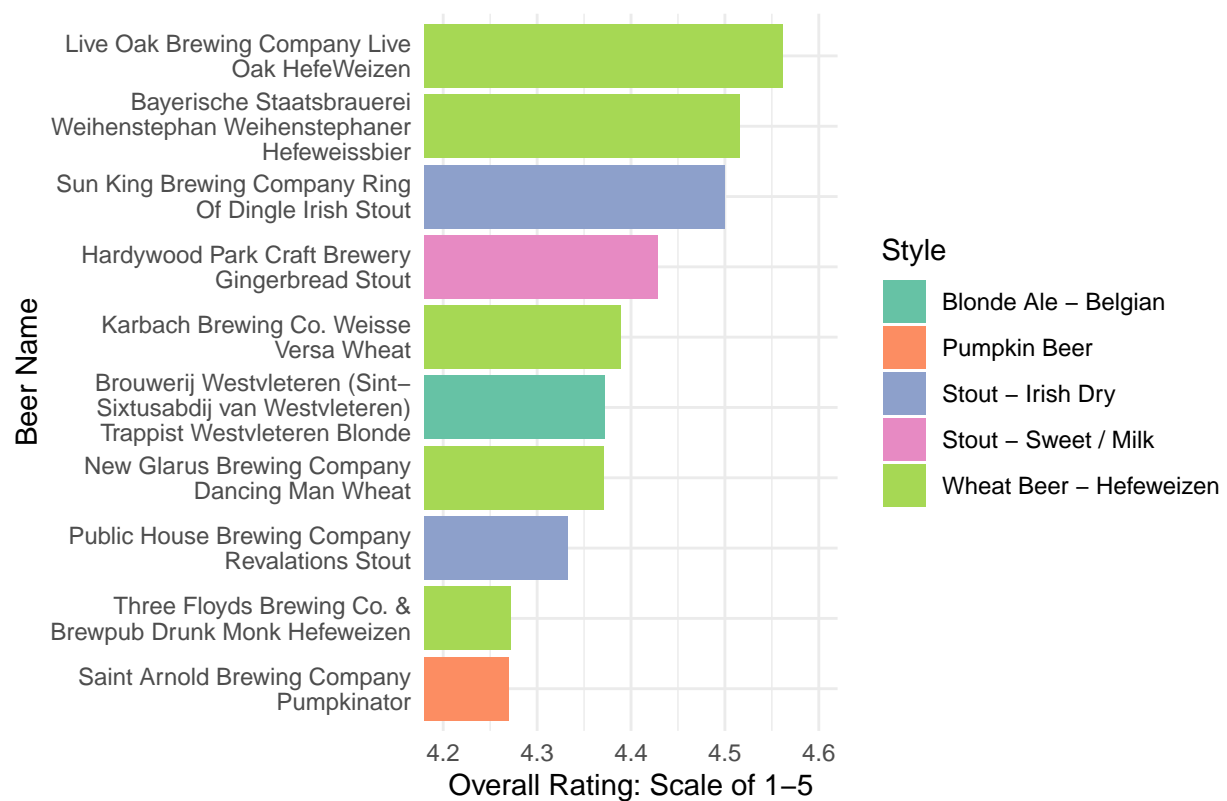
'geom_smooth()' using method = 'loess' and formula 'y ~ x'

Alcohol by Volume (ABV) and Overall Rating of Specific Beer Styles



```
top_beers %>%
  mutate(Beer = fct_reorder(FullBeerName, review_overall)) %>%
  ggplot(aes(review_overall, Beer)) +
  geom_col(aes(fill = Style)) +
  coord_cartesian(xlim = c(4.2, 4.6)) +
  scale_fill_brewer(palette = "Set2") +
  scale_y_discrete(labels = function(y) str_wrap(y, width = 30)) +
  labs(title = "Top 10 Rated Beers of Selected Styles", y = "Beer Name",
       x = "Overall Rating: Scale of 1-5") +
  theme_minimal()
```

Top 10 Rated Beers of Selected Styles



```
topABVbeers %>%
  mutate(Beer = fct_reorder(FullBeerName, ABV)) %>%
  ggplot(aes(Beer, ABV)) +
  geom_col(aes(fill = Style)) +
  coord_flip() +
  scale_fill_brewer(palette = "Set2") +
  scale_x_discrete(labels = function(x) str_wrap(x, width = 30)) +
  labs(title = "Top 10 Beers by ABV with Beer Style", x = "Beer Name", y = "% ABV") +
  theme_minimal()
```

Top 10 Beers by ABV with Beer Style

