

Chocolate N' Tea

Chigo

2/16/2022

Installing and loading the necessary packages

```
install.packages("tidyverse")
install.packages("skimr")
install.packages("janitor")
install.packages('ggplot2')
library(ggplot2)
library(tidyverse)
library(skimr)
library(janitor)
```

Importing the data and creating a dataframe

```
flavors_df <- read_csv("flavors_of_cacao.csv")

## Rows: 1795 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (6): Company, BeanOrigin, CocoaPercent, CompanyLocation, BeanType, Broad...
## dbl (3): REF, ReviewDate, Rating
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
show_col_types = FALSE
```

Inspect the dataframe

```
head(flavors_df)

## # A tibble: 6 x 9
##   Company BeanOrigin REF ReviewDate CocoaPercent CompanyLocation Rating
##   <chr>    <chr>    <dbl>    <dbl> <chr>          <chr>          <dbl>
## 1 A. Morin Agua Grande 1876      2016 63%      France        3.75
## 2 A. Morin Kpime      1676      2015 70%      France        2.75
## 3 A. Morin Atsane      1676      2015 70%      France        3
## 4 A. Morin Akata      1680      2015 70%      France        3.5
## 5 A. Morin Quilla      1704      2015 70%      France        3.5
## 6 A. Morin Carenero    1315      2014 70%      France        2.75
## # ... with 2 more variables: BeanType <chr>, `Broad BeanOrigin` <chr>
colnames(flavors_df)
```

```
## [1] "Company"          "BeanOrigin"      "REF"             "ReviewDate"
## [5] "CocoaPercent"     "CompanyLocation" "Rating"           "BeanType"
## [9] "Broad BeanOrigin"

glimpse(flavors_df)

## Rows: 1,795
## Columns: 9
## $ Company          <chr> "A. Morin", "A. Morin", "A. Morin", "A. Morin", "A.~
## $ BeanOrigin        <chr> "Agua Grande", "Kpime", "Atsane", "Akata", "Quilla"~
## $ REF               <dbl> 1876, 1676, 1676, 1680, 1704, 1315, 1315, 1315, 131~
## $ ReviewDate        <dbl> 2016, 2015, 2015, 2015, 2015, 2014, 2014, 2014, 201~
## $ CocoaPercent      <chr> "63%", "70%", "70%", "70%", "70%", "70%", "70%", "7~
## $ CompanyLocation   <chr> "France", "France", "France", "France", "France", "~
## $ Rating            <dbl> 3.75, 2.75, 3.00, 3.50, 3.50, 2.75, 3.50, 3.7~
## $ BeanType          <chr> "Unknown", "Unknown", "Unknown", "Unknown", "Unknow~
## $ `Broad BeanOrigin` <chr> "Sao Tome", "Togo", "Togo", "Togo", "Peru", "Venezu~
```

Selecting specific columns

```
trimmed_flavors_df <- flavors_df %>%
  select(Rating, CocoaPercent, Company, CompanyLocation)
head(trimmed_flavors_df)
```

```
## # A tibble: 6 x 4
##   Rating CocoaPercent Company CompanyLocation
##   <dbl> <chr>         <chr>      <chr>
## 1   3.75 63%          A. Morin France
## 2   2.75 70%          A. Morin France
## 3    3   70%          A. Morin France
## 4   3.5 70%          A. Morin France
## 5   3.5 70%          A. Morin France
## 6   2.75 70%          A. Morin France
```

Check for the maximum rating

```
trimmed_flavors_df %>%
  summarize(highest_rating=max(Rating))
```

```
## # A tibble: 1 x 1
##   highest_rating
##   <dbl>
## 1           5
```

Filtering by CocoaPercent and Rating

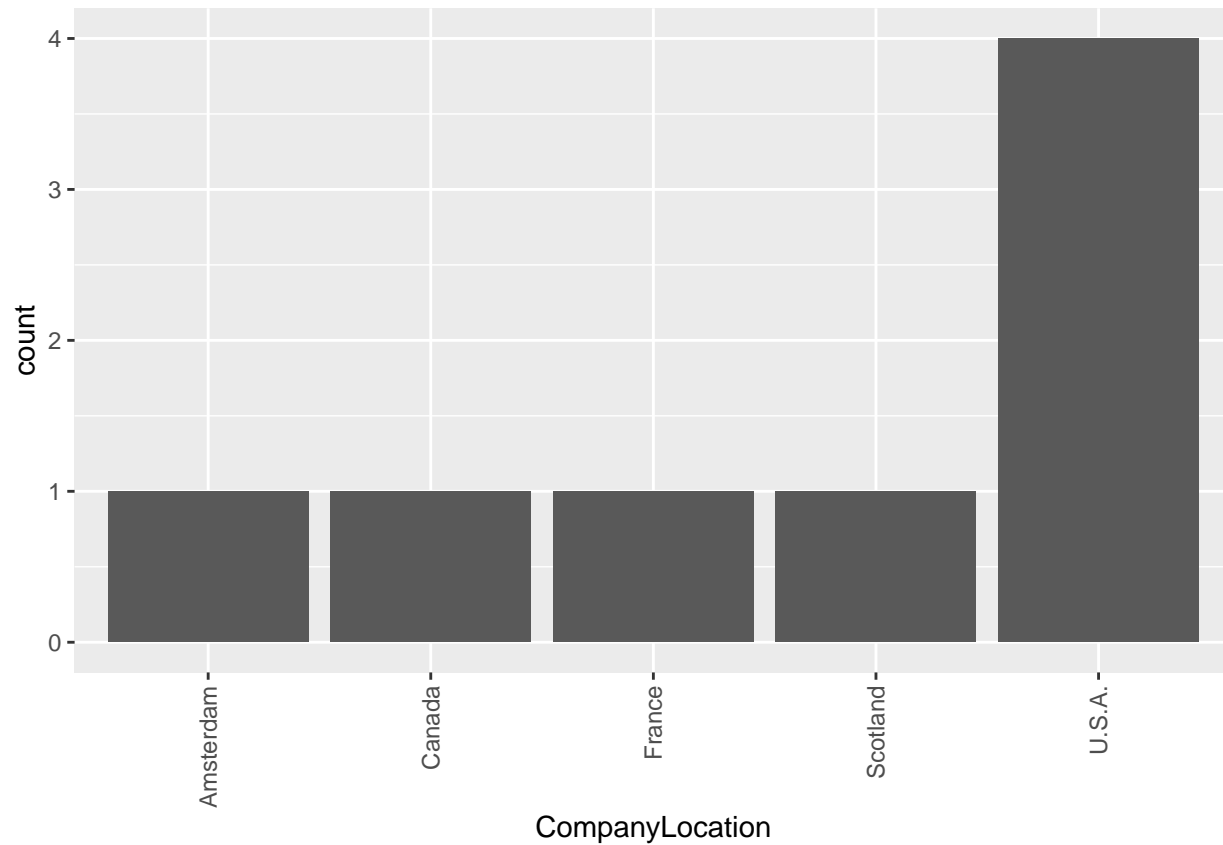
```
best_trimmed_flavors_df <- trimmed_flavors_df %>%
  filter(CocoaPercent >= 80, Rating >= 3.75)
head(best_trimmed_flavors_df)
```

```
## # A tibble: 6 x 4
##   Rating CocoaPercent Company CompanyLocation
##   <dbl> <chr>         <chr>      <chr>
## 1   3.75 80%          Chocolate Makers Amsterdam
## 2   3.75 80%          Chocolate Tree, The Scotland
```

##	3	3.75	80%	Ethereal	U.S.A.
##	4	3.75	82%	Potomac	U.S.A.
##	5	4	80%	Pralus	France
##	6	3.75	80%	Rogue	U.S.A.

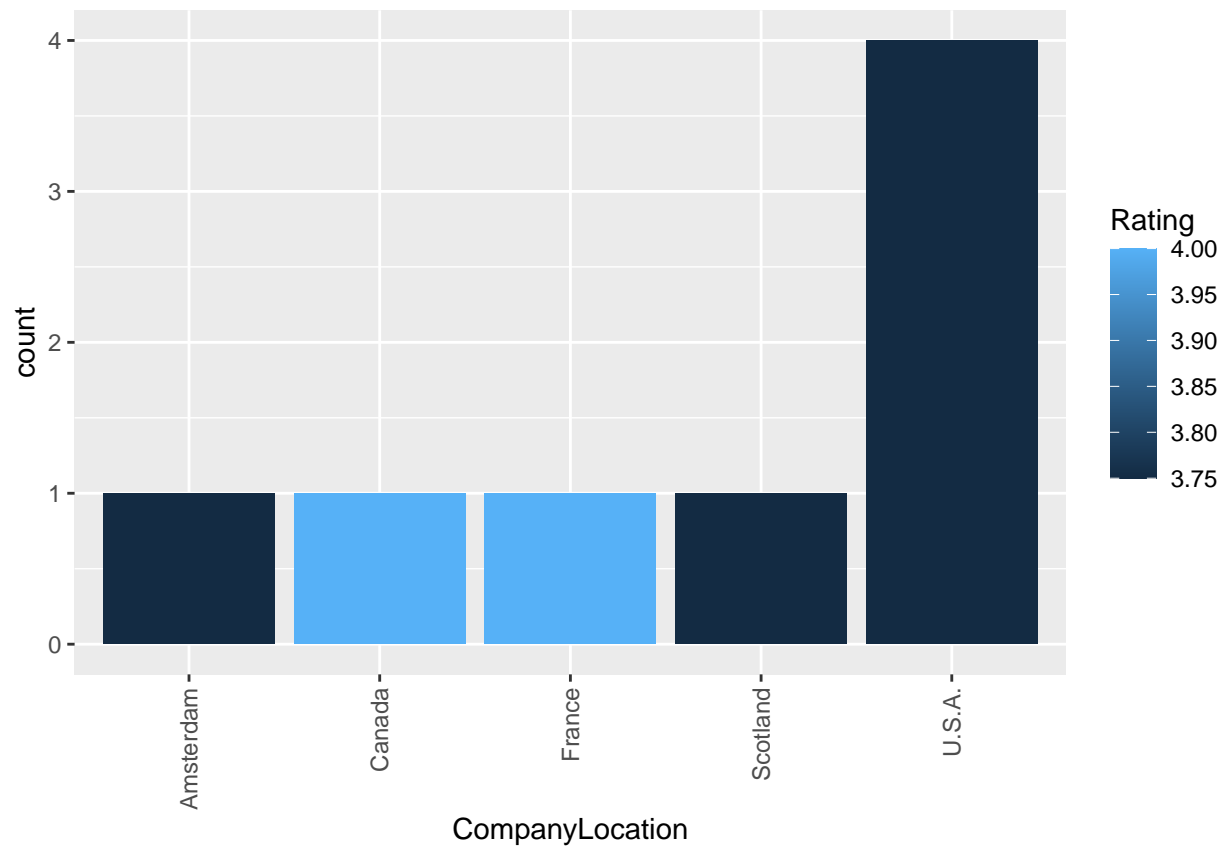
Spread of companies(grouped by location) that produce the best rated flavors

```
ggplot(data = best_trimmed_flavors_df) +
  geom_bar(mapping = aes(x=CompanyLocation)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



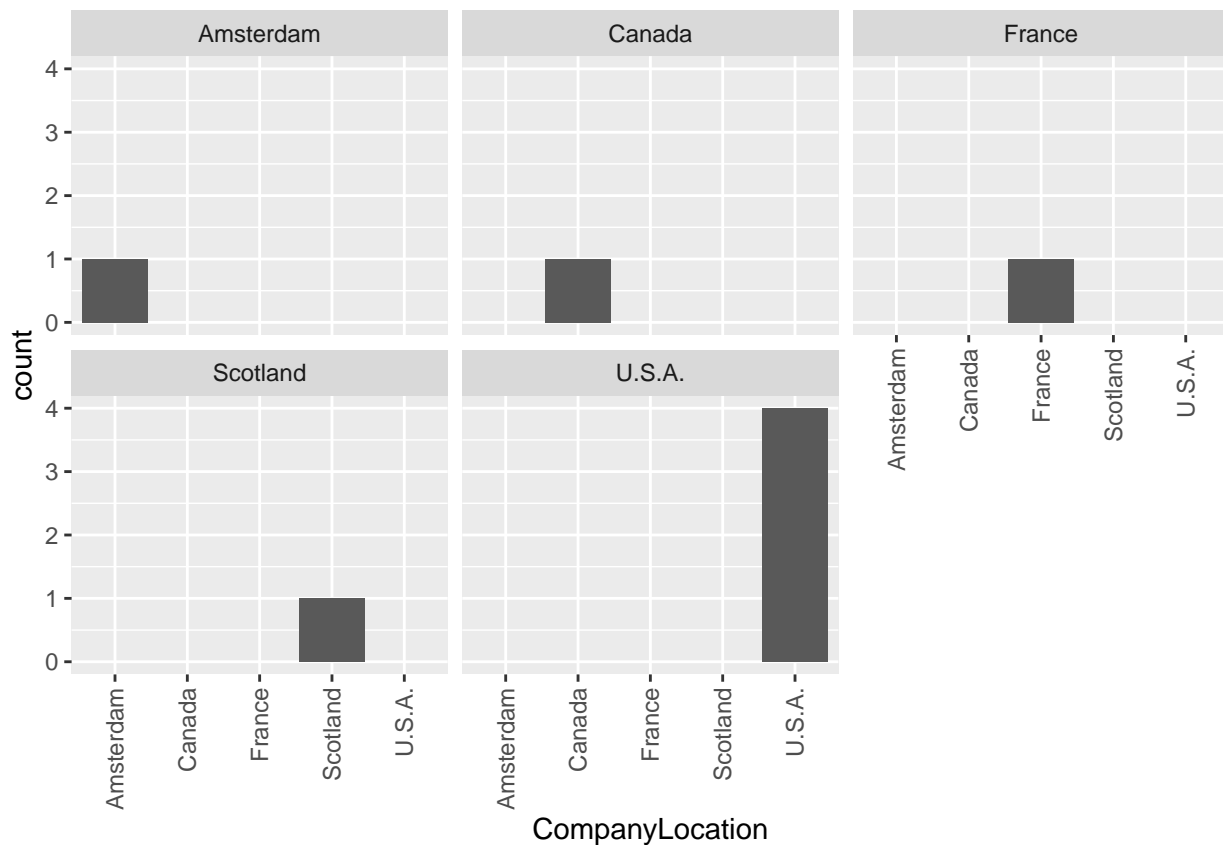
Locations that produce the flavors with the highest rating

```
ggplot(data = best_trimmed_flavors_df) +
  geom_bar(mapping = aes(x=CompanyLocation, fill = Rating)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



Facet wrap

```
ggplot(data = best_trimmed_flavors_df) +  
  geom_bar(mapping = aes(x=CompanyLocation)) +  
  facet_wrap(~CompanyLocation) +  
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



Selecting specific columns

```
new_trimmed_flavors_df <- flavors_df %>%
  select(Rating, CocoaPercent, Company, CompanyLocation, BeanOrigin, BeanType, `Broad BeanOrigin`)
head(new_trimmed_flavors_df)
```

```
## # A tibble: 6 x 7
##   Rating CocoaPercent Company CompanyLocation BeanOrigin BeanType
##   <dbl> <chr>         <chr>    <chr>          <chr>    <chr>
## 1   3.75 63%          A. Morin France      Agua Grande Unknown
## 2   2.75 70%          A. Morin France      Kpime      Unknown
## 3    3    70%          A. Morin France      Atsane     Unknown
## 4   3.5  70%          A. Morin France      Akata      Unknown
## 5   3.5  70%          A. Morin France      Quilla     Unknown
## 6   2.75 70%          A. Morin France      Carenero   Criollo
## # ... with 1 more variable: `Broad BeanOrigin` <chr>
```

Filtering by CocoaPercent

```
cocoa_content_df <- new_trimmed_flavors_df %>%
  filter(CocoaPercent >= 70)
head(cocoa_content_df)
```

```
## # A tibble: 6 x 7
##   Rating CocoaPercent Company CompanyLocation BeanOrigin BeanType
##   <dbl> <chr>         <chr>    <chr>          <chr>    <chr>
## 1   2.75 70%          A. Morin France      Kpime      Unknown
```

```
## 2 3 70% A. Morin France Atsane Unknown
## 3 3.5 70% A. Morin France Akata Unknown
## 4 3.5 70% A. Morin France Quilla Unknown
## 5 2.75 70% A. Morin France Carenero Criollo
## 6 3.5 70% A. Morin France Cuba Unknown
## # ... with 1 more variable: `Broad BeanOrigin` <chr>
```

Best Chocolate scatterplot

```
ggplot(data = cocoa_content_df) +
  geom_point(mapping = aes(x = CocoaPercent, y = Rating)) +
  labs(title = 'CocoaPercent Vs Rating')
```

