Dark Chocolate Ratings

Paul Lazarescu

2022-10-16

Business Objective

Our project manager is looking for which countries produce the highest-rated bars of super dark chocolate (a high percentage of cocoa). This data will help them create their next chocolate bar menu.

Preview

First, we start by checking our working directory. A working directory simply means the folder within which our files are stored. If the wd is incorrect or changes, roll back to base ("~/"), then set to correct folder location. Afterwards, we load the 'tidyverse' and 'rlang' packages for our analysis, as well as our dataset.

We run a few pieces of code that show us the data we are working with.

```
colnames(chocolate_df)
```

```
## [1] "Company \n(Maker-if known)"        "Specific Bean Origin\nor Bar Name"
## [3] "REF"                               "Review\nDate"
## [5] "Cocoa\nPercent"                    "Company\nLocation"
## [7] "Rating"                            "Bean\nType"
## [9] "Broad Bean\nOrigin"
```

```
head(chocolate_df) #vertical
```

```
## # A tibble: 6 x 9
##   `Company \n(Maker-if known)` Specific Bean Origin\nor B~1   REF `Review\nDate`
##   <chr>                        <chr>                       <dbl>          <dbl>
## 1 A. Morin                     Agua Grande                  1876           2016
## 2 A. Morin                     Kpime                        1676           2015
## 3 A. Morin                     Atsane                       1676           2015
## 4 A. Morin                     Akata                        1680           2015
## 5 A. Morin                     Quilla                       1704           2015
## 6 A. Morin                     Carenero                     1315           2014
## # i abbreviated name: 1: `Specific Bean Origin\nor Bar Name`
## # i 5 more variables: `Cocoa\nPercent` <chr>, `Company\nLocation` <chr>,
## #   Rating <dbl>, `Bean\nType` <chr>, `Broad Bean\nOrigin` <chr>
```

Before we trim down the dataset. Let's see what the highest rating possible is, and let's see how many countries we are working with.

```
print(max(trimmed_flavors_df$Rating))
```

```
## [1] 5
```

```
n_distinct(trimmed_flavors_df$`Company
Location`)
```

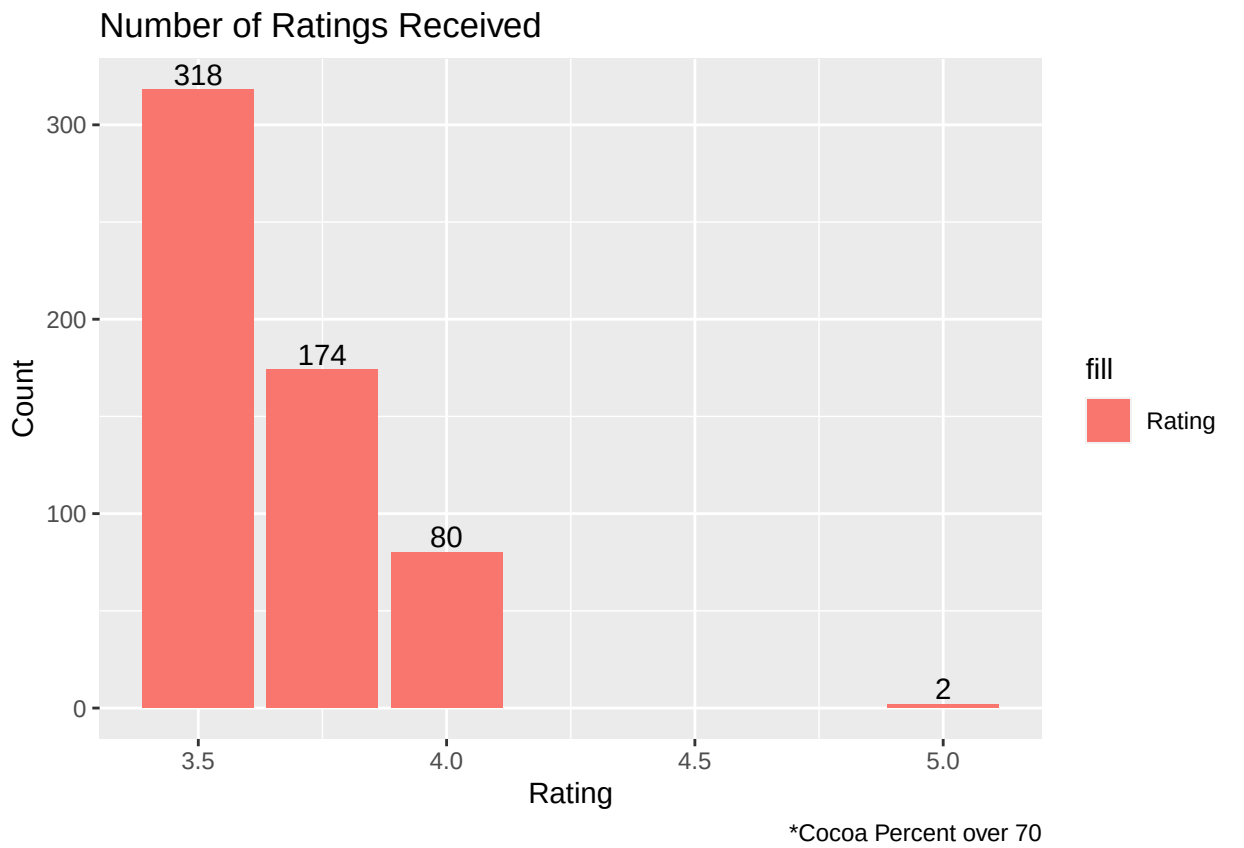
```
## [1] 60
```

The highest rating attainable is a five, and the number of cocoa source locations is sixty.

Preparation

In order to answer the business objective, we must filter for the data we are going to be using. We choose 3.5 because that represents a relatively high rating, and a cocoa percent of at-least 70%, as that is considered "dark" chocolate within our industry. For the purpose of ad-hoc analysis, this work perfectly.

```
best_trimmed_flavors_df <- trimmed_flavors_df %>%
  filter(Rating >= "3.5", `Cocoa
Percent` >= "70%")
view(best_trimmed_flavors_df)
```



Above is a visualization that shows the distribution of ratings. What it shows us is that there are very few dark chocolates that receive perfect ratings, but there are quite a lot between 3.5 to 4.0. This gives us ample sample size to work with. *Additional exploratory analysis also showed that the majority of ratings fall below 4.0, irrespective of cocoa percentage.*

Analysis

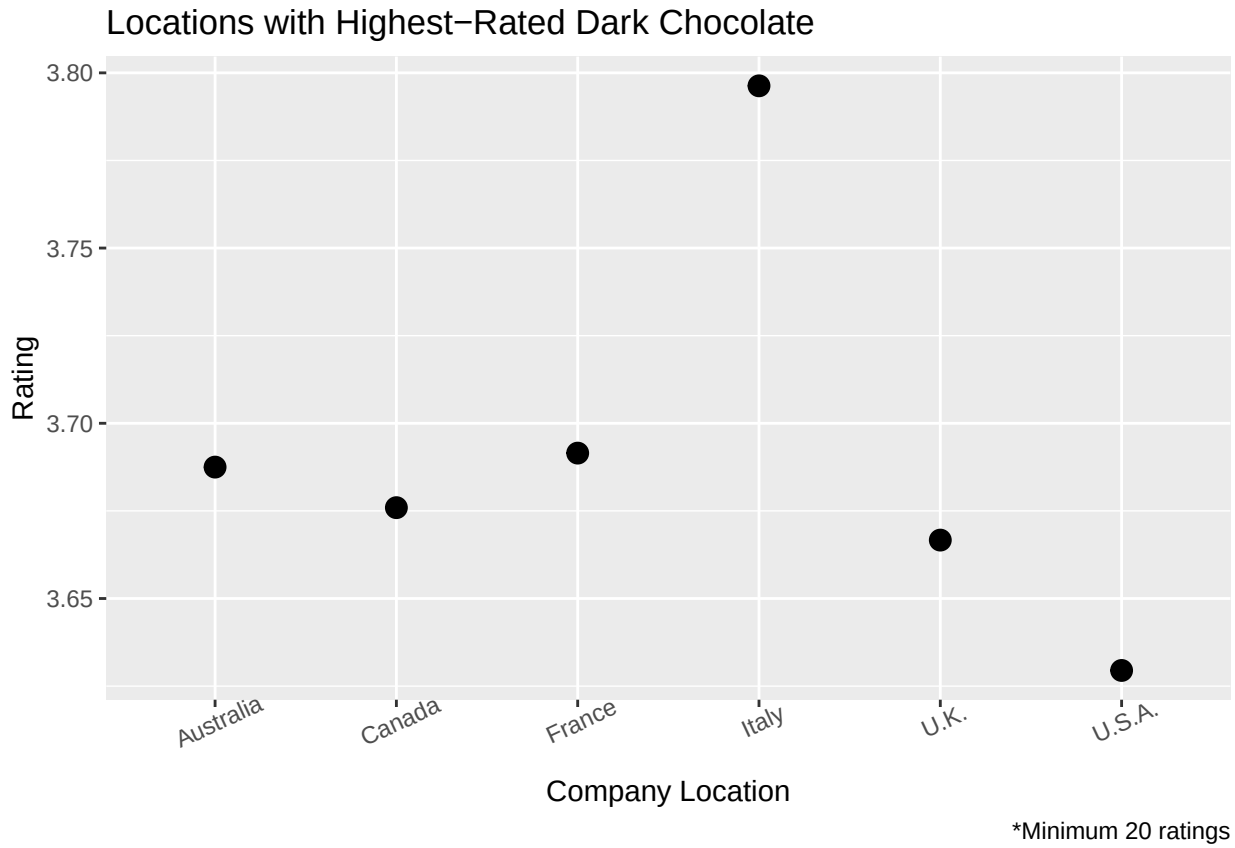
We want to see which countries produce the highest-rated dark chocolate. We have established a minimum cocoa percentage of 70%, a lowest rating of 3.5, and will be using a baseline (n) count of 20 ratings to give us an accurate sample size.

```
highest_rated3 <- filter(summarise(group_by(best_cleaned, company_location), count=n(), mean(rating)), c
```

```
print(highest_rated3[order(highest_rated3$rating, decreasing = TRUE), ])
```

```
## # A tibble: 6 x 3
##   company_location count_bars rating
##   <chr>                <int>  <dbl>
## 1 Italy                   27   3.80
## 2 France                  47   3.69
## 3 Australia               20   3.69
## 4 Canada                  54   3.68
## 5 U.K.                    24   3.67
## 6 U.S.A.                 224   3.63
```

Our table, sorted by highest ratings. Let's create a graph to visualize these findings.



According to the graph, the location that provides us with the highest-rated bars of super dark chocolate is **Italy**. This analysis may be used as the basis for sourcing a larger quantity of our dark chocolate from the country of Italy.

Conclusion

To surmise, our stakeholder was looking to find out which country produced the highest rated dark-chocolate. According to the findings, that country was: Italy, followed by France and then Australia. We can use this information to source more of our inventory from the most economically-viable source.

Further analysis can be done by varying the levels of cocoa in the chocolate, joining a table containing shipping data and costs, finding the top-producing companies in a given region, or looking at specific bean origin.