

Project 5

Read in the dataset you will be working with:

```
chocolate <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2022/2022-01-18/chocolate.csv')
```

Information about the dataset

The dataset provides information about various chocolate bars, including details about their manufacturers, origins, ingredients, characteristics, taste profiles, and ratings.

This dataset is valuable for chocolate enthusiasts, manufacturers, and researchers interested in exploring the characteristics and origins of chocolate bars, as well as analyzing how various factors, such as cocoa percentage, specific bean origin, and manufacturer location, influence the quality and characteristics of the chocolate bars. Analysis and visualization of this data can provide insights into regional preferences, quality trends, and potential relationships between different attributes in the chocolate industry.

Question: How do the cocoa percentages vary across different regions or countries of bean origin?

Introduction:

This dataset provides a comprehensive overview of chocolate bars, offering insights into their origins, manufacturing details, ingredients, and quality assessments. Within this document, we aim to explore the relationship between the geographic aspects of chocolate production, such as the country of bean origin and the manufacturer's location, and their impact on the quality and characteristics of chocolate bars. To answer the question of how cacao percentages vary across countries of origin, we will use columns which provide the cocoa percent, as well as the corresponding country of origin to visualize geo-spatial data. The columns that will be used to execute the analysis are: country_of_bean_origin, cocoa_percent, most_memorable_flavors, and geometry.

Approach:

This project will involve geospatial analysis, as the approach for this analysis involves exploring how cocoa percentages vary across different regions or countries of bean origin in the chocolate dataset. We will begin by preparing the data, selecting the essential columns, and calculating summary statistics to gain an initial understanding of the cocoa percentage distribution. Subsequently, we will use `ne_countries()` to visualize by gradient the mean cocoa percentage by country. Additionally, I will provide a table with mean cocoa percent and the corresponding flavors (most_memorable_flavors).

Analysis:

```
#convert percentage to numeric decimal
chocolate1 <- chocolate %>%
  mutate(cocoa_percent1 = as.numeric(sub("%", "", cocoa_percent)))

#get mean cocoa percentage by country
mean_perc <- chocolate1 %>%
  filter(country_of_bean_origin != 'Blend') %>%
  group_by(country_of_bean_origin) %>%
  summarize(
    mean_percent = mean(cocoa_percent1, na.rm = TRUE),
    most_memorable_characteristics = first(most_memorable_characteristics)
  ) %>%
  rename("sovereignty" = "country_of_bean_origin")

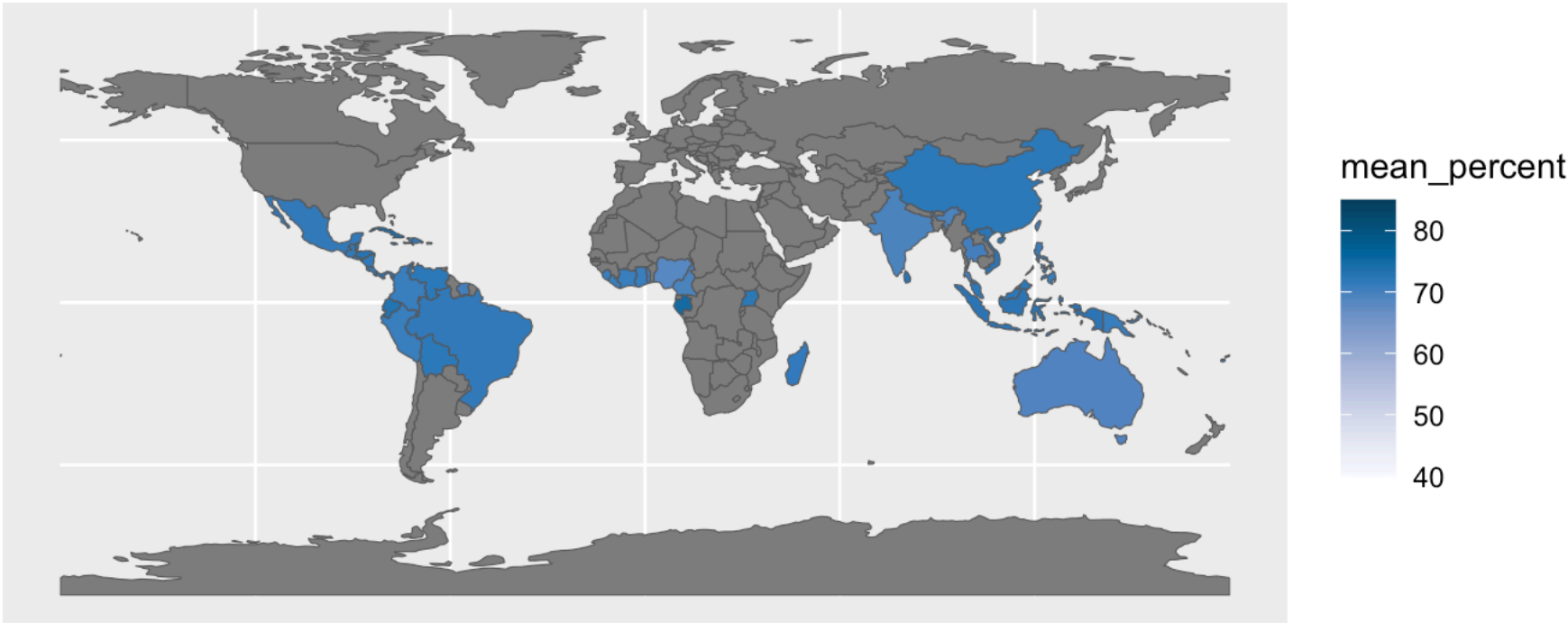
#get map data
sf_world <- ne_countries(returnclass= 'sf')

#color countries by average cocoa percentage
sf_world2 <- left_join(sf_world, mean_perc) %>%
  select(sovereignty, mean_percent, geometry)

## Joining with `by = join_by(sovereignty)`

ggplot(sf_world2) + geom_sf(aes(fill = mean_percent)) +
  scale_fill_continuous_sequential(
    palette = "PuBu", rev = TRUE,
    limits = c(40, 85)

  )
```



```
char_table <- mean_perc %>%
  select(mean_percent, most_memorable_characteristics) %>%
  arrange(mean_percent)

char_table

## # A tibble: 61 x 2
##   mean_percent most_memorable_characteristics
##   <dbl> <chr>
## 1      42 gritty, sweet, hot cocoa
## 2      65 creamy, rich, dairy, nutty
## 3      66 nuts, fruit, cocoa
## 4      68 fatty, herbal, earthy
## 5     68.3 dirt, floral, sweet
## 6     68.3 candy like off flavor
## 7      69 vanilla, banana, citrus
## 8     69.2 creamy, nutty, fruity
## 9     69.3 earthy
## 10    69.4 fruity with coffee notes
## # i 51 more rows
```

Discussion:

As we can see from the map data, it appears that the countries that the countries with the smallest mean percent of cocoa reside in Africa, while those with the highest average cocoa percentage are in the Americas or Asia. I suspect that this is most likely due to different varieties of cacao, as well as different climate conditions for growing within each area. From the tibble, the chocolates with the higher percentage of cacao tend to be more bitter tasting.