Good Plot, Bad Plot

Kyle Bosworth

2024-10-29

Contents

Load Libraries

```
library(here)
library(tidyverse)
```

Read in data: 2019 - Ramen Ratings

```
ramen_ratings<- readr::read_csv("https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/da
glimpse(ramen_ratings)
```

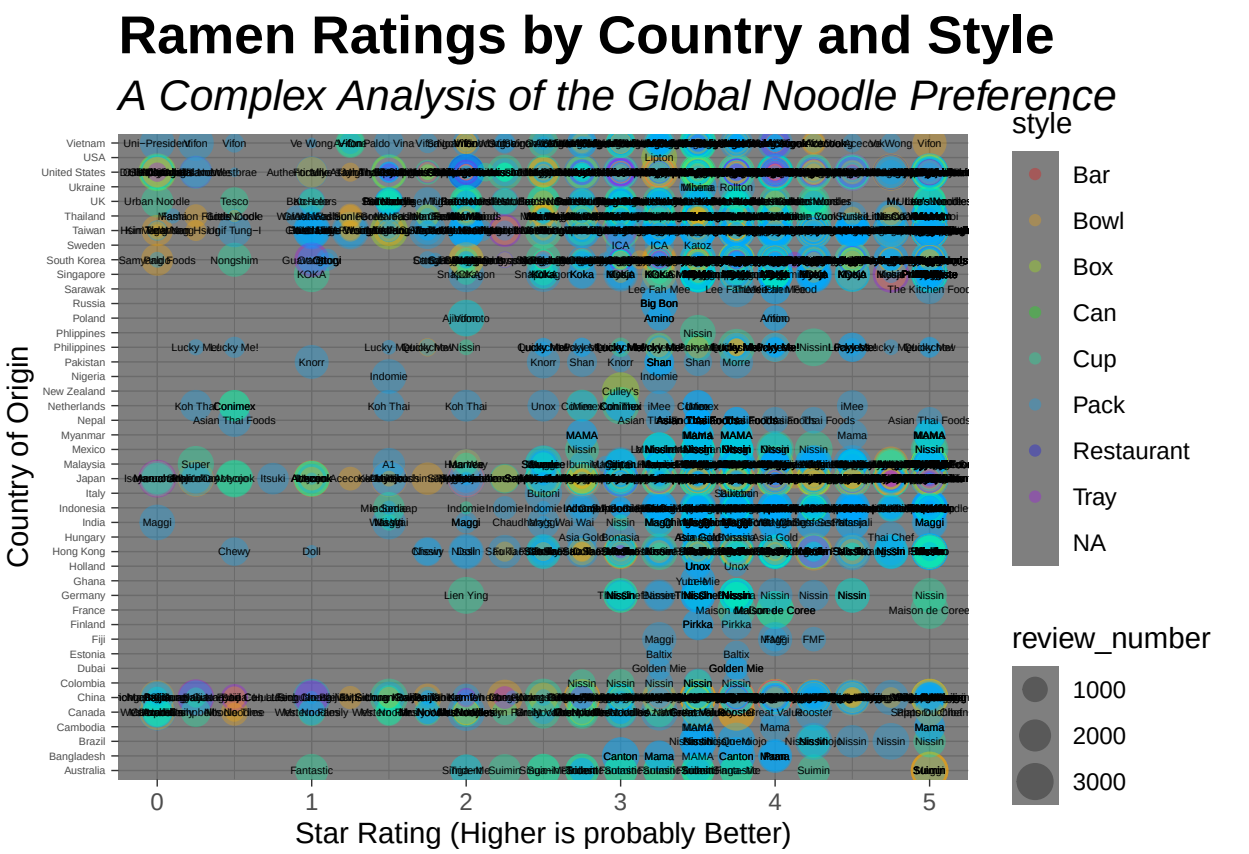
```
## Rows: 3,180
## Columns: 6
## $ review_number <dbl> 3180, 3179, 3178, 3177, 3176, 3175, 3174, 3173, 3172, 31~
## $ brand         <chr> "Yum Yum", "Nagatanien", "Acecook", "Maison de Coree", "~
## $ variety       <chr> "Tem Tem Tom Yum Moo Deng", "tom Yum Kung Rice Vermicell~
## $ style         <chr> "Cup", "Pack", "Cup", "Cup", "Tray", "Cup", "Pack", "Pac~
## $ country       <chr> "Thailand", "Japan", "Japan", "France", "Japan", "Japan"~
## $ stars         <dbl> 3.75, 2.00, 2.50, 3.75, 5.00, 3.50, 3.75, 5.00, 3.50, 4.~
```

```
View(ramen_ratings)
```

Figure 1. Bad Plot, Ramen Ratings by Country and Style

```
ggplot(ramen_ratings, aes(x = stars,
                          y = country)) +
  geom_point(aes(color = style, #points to plot, colored by style, size based on review number, added t
                 size = review_number), alpha = 0.3) +
  geom_text(aes(label = brand), size = 1.5) + #added text to each point (brand names), text is tiny and
  theme_dark() +
  scale_color_manual(values = rainbow(length(unique(ramen_ratings$style)))) + #color scale is rainbow,
  labs(title = "Ramen Ratings by Country and Style",
       subtitle = "A Complex Analysis of the Global Noodle Preference",
```

```
      x = "Star Rating (Higher is probably Better)",
      y = "Country of Origin") +
  theme(axis.text.y = element_text(size = 4),
        plot.title = element_text(face = "bold",
                                  size = 20),
        plot.subtitle = element_text(face = "italic",
                                     size = 16))
```



Bad plot characteristics: 1. Over-plotting: super dense scatter plot with overlap, obscures data points and any recognizable patterns. 2. Poor color choice: rainbow scheme is not uniform and may be misleading for colorblind viewers. 3. Illegible text: small, unreadable, cluttered 4. Inconsistent visual representation of data: lacks clear focus, does not highlight and specific pattern, disproportionate, overload of data 5. Poor contrast: dark theme with transparency

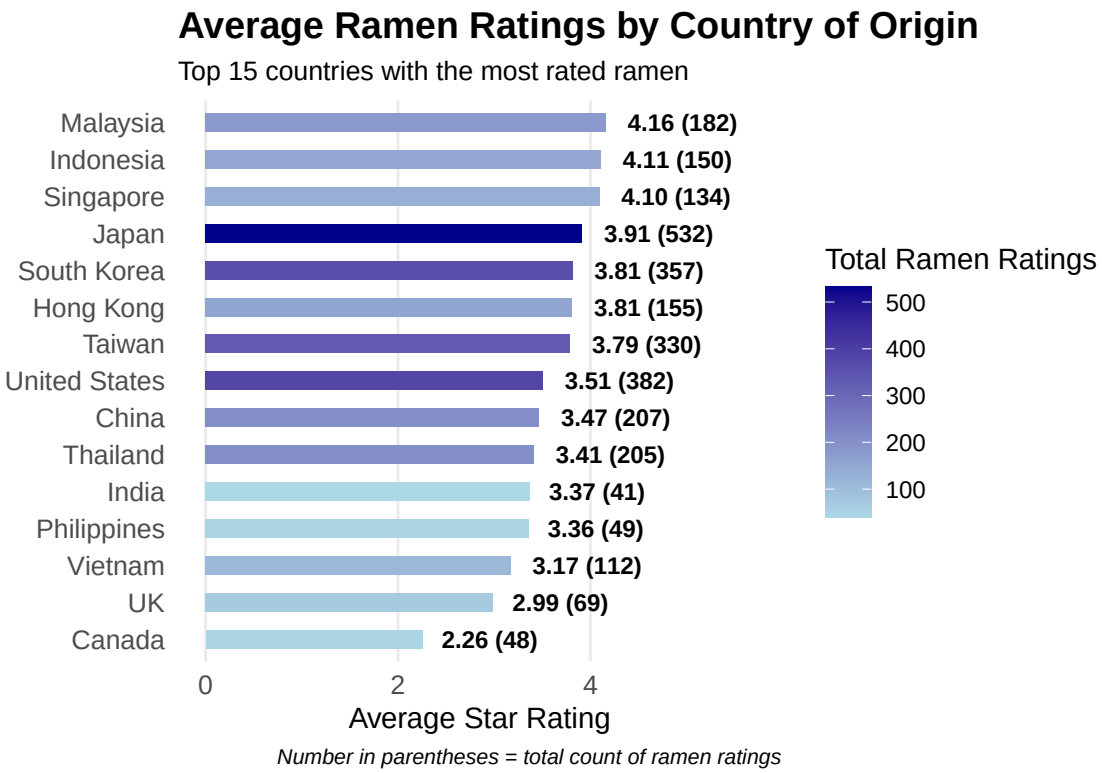
Figure 2. Good Plot, Ramen Ratings by Country and Style

```
country_avg<- ramen_ratings %>%
  group_by(country) %>% #goups by country
  summarise(avg_stars = mean(stars, na.rm = TRUE), #calculate mean and ignore the na values
            count = n()) %>% #counts number of rating/each country
  arrange(desc(avg_stars)) %>% #sorts country by average star in decreasing order
  top_n(15, count)  #select top 15 countries based on the number of total ramen ratings thay have

ggplot(country_avg, aes(x = reorder(country, avg_stars),
```

```
                        y = avg_stars)) +
  geom_col(aes(fill = count), width = 0.5) +  #bars colored by rating count, set a width
  geom_text(aes(label = sprintf("%.2f (%d)", avg_stars, count)), #function for rounding the average sta
#           `.2f` = floating wth 2 decimal places, `%d` is the integer.

            hjust = -0.2, size = 3, fontface = "bold") +  #bold the ratings labels to make them pop out
  coord_flip() + #horizontal barchart
  scale_fill_gradient(low = "lightblue", high = "darkblue") +  #using a two-color blue gradient
  labs(title = "Average Ramen Ratings by Country of Origin",
       subtitle = "Top 15 countries with the most rated ramen",
       caption = "Number in parentheses = total count of ramen ratings",
       x = "", #assuming you can tell that they are countries
       y = "Average Star Rating",
       fill = "Total Ramen Ratings") +
  theme_minimal() +
  theme(
    axis.text = element_text(size = 10),
    plot.title = element_text(face = "bold", size = 14),
    plot.subtitle = element_text(size = 10),
    plot.caption = element_text(size = 8, face = "italic"),
    panel.grid.major.y = element_blank(), #removes horizontal grid line to make cleaner
    panel.grid.minor.x = element_blank(), #removes some vertical grid lines to make cleaner
    legend.position = "right",
    plot.margin = margin(t = 20, r = 20, b = 20, l = 20, unit = "pt")  #add margin for more space
  ) +
  expand_limits(y = 5.7)  #makes the y-axis a little longer to fit labels
```



Good plot characteristics: 1. Clear data presentation: displays ramen ratings by country ,easy to interpret, no overlap and good use of space. 2. Appropriate color scheme: easy and accessible to read, colorblind friendly color gradient scale, informative of data 3. Clear titles an captions 4. minimalist theme, informative labels that are legible