The Ramen Report

Keerthana

15/06/2021

In this report i have analysed the ramen-ratings dataset form the kaggle. The link for the dataset is given below: ramen-ratings.csv

Now let us import the necessay libraries for the analysis

```
library(tidyr)
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
```

```
library(ggplot2)
```

Here we would be using the Tidyr package which helps us clean and format the dataframe. We would be using dplyr package to perform data manipulation operations and we would also be using the ggplot2 package that helps us visualize our insights which we derived out of the data set.

Now let us read the data and store it in a data frame thus we can work on the created dataframe without affecting the source dataframe.

```
ramen <- read.csv("ramen-ratings.csv")
```

Lets have a closer look into the dataframe by looking into the column names, thus learning its structure. Following we can have a look into the first few rows of the records just to get an insight of the dataframe.

```
colnames(ramen)
```

```
## [1] "Review.." "Brand"    "Variety"  "Style"    "Country"  "Stars"    "Top.Ten"
```

```
head(ramen)
```

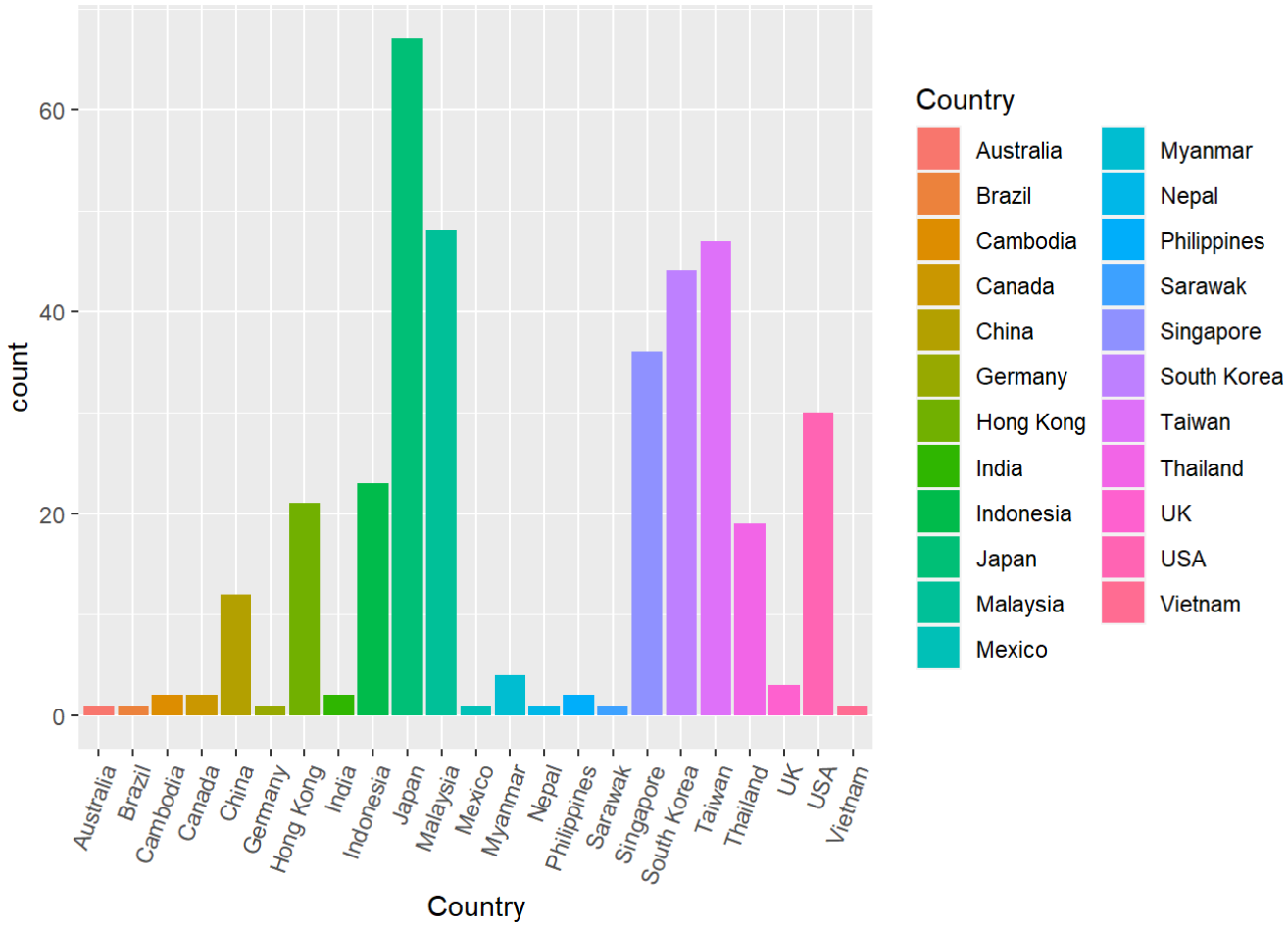
```
##   Review..          Brand
## 1     2580      New Touch
## 2     2579       Just Way
## 3     2578         Nissin
## 4     2577        Wei Lih
## 5     2576 Ching's Secret
## 6     2575  Samyang Foods
##                                                    Variety Style     Country
## 1                                T's Restaurant Tantanmen    Cup       Japan
## 2 Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles  Pack      Taiwan
## 3                             Cup Noodles Chicken Vegetable   Cup         USA
## 4                             GGE Ramen Snack Tomato Flavor  Pack      Taiwan
## 5                                          Singapore Curry  Pack       India
## 6                                 Kimchi song Song Ramen  Pack South Korea
##   Stars Top.Ten
## 1  3.75
## 2     1
## 3  2.25
## 4  2.75
## 5  3.75
## 6  4.75
```

According to the description of the data, it is given that the ratings of the ramen are given by the ‘Stars’ column whose value ranges between 0 to 5. Let us work on finding the country that sells the best large number of ramen. Thus let us extract the dataframe to the ramen which holds 5 star rating and arrange it in descending order thus we can get to view from the highest rating to lower.

```
top_ramen<- ramen %>%
  filter(Stars==5) %>%
  arrange(desc(Stars))
```

Lets plot the ‘top_ramen’ dataset into a bar graph so that we can easily get to know which country serves the most number of ramen whose rating is 5 stars

```
ggplot(data = top_ramen)+
  geom_bar(mapping = aes(x=Country,fill=Country))+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
```



We can see that Japan has the largest number of 5 star ramen. To be more precise let us count the number of 5 star ramen available in Japan.

```
nrow(subset(top_ramen,Country == "Japan"))
```

```
## [1] 67
```

This code shows the count of 5 star ramen available in japan. Now once again look into the source dataframe.

```
colnames(ramen)
```

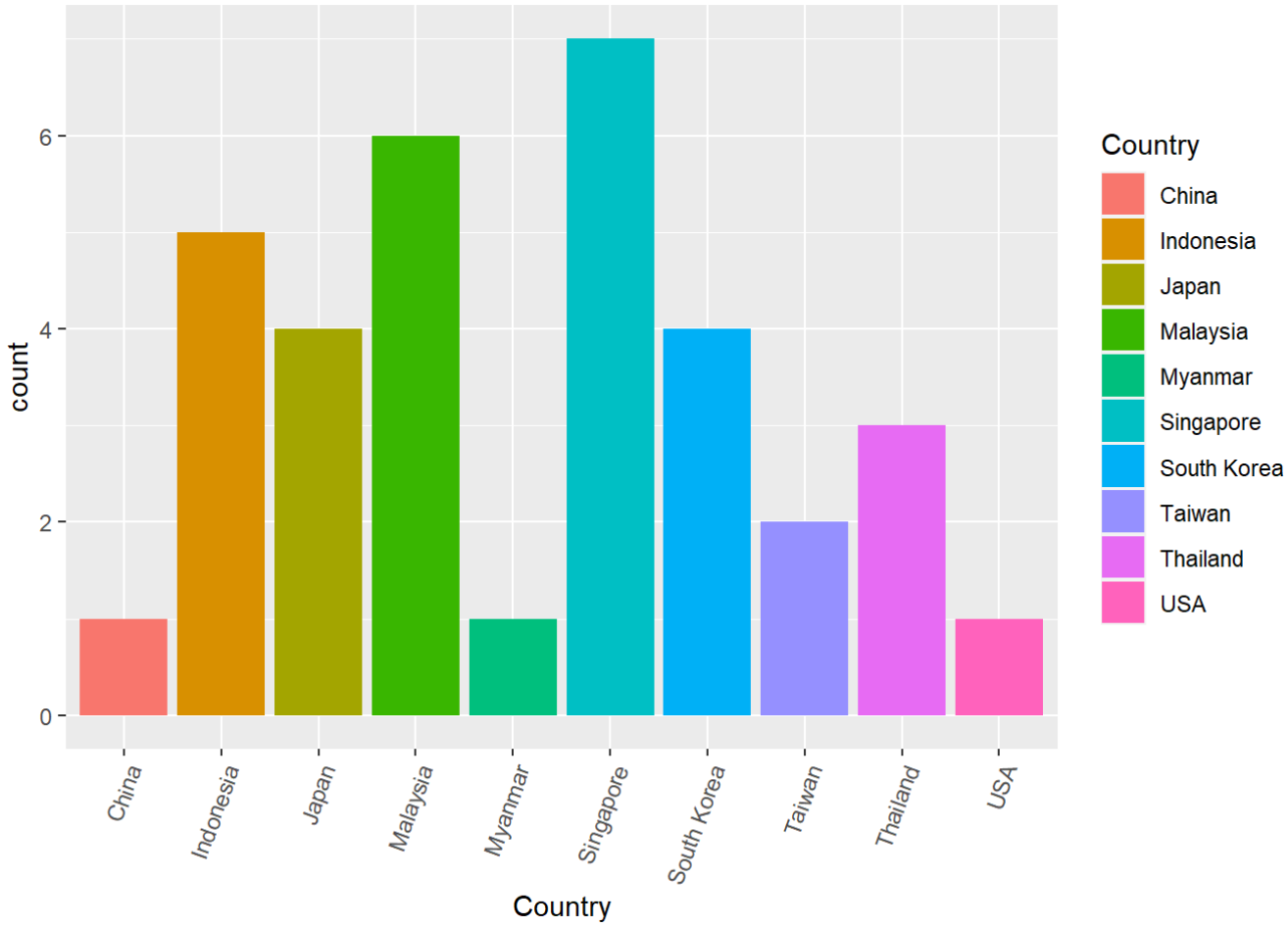
```
## [1] "Review.." "Brand"    "Variety"  "Style"    "Country"  "Stars"    "Top.Ten"
```

The column ‘Top.Ten’ is a special column and only sleected ramen has entry into that. These ramen with an entry in the column has set a record into the top ten best ramen of the year. Thus let us now find the best out of the best record holding ramen. For that we need to extract from the source dataframe, only the netries with a record in ‘Top.Ten’ column. It can be done as follows

```
elite_ramen <- subset(top_ramen, Top.Ten != "")
```

Now let us plot this ‘elite_ramen’ dataset into a bar graph and find out which country has the highest number of record holding ramen. You might be guessing that it would obviously be Japan. But the result turns out to be,

```
ggplot(data = elite_ramen)+
  geom_bar(mapping = aes(x=Country,fill=Country))+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
```



Thus we get to know that though japan has a large number of 5 star ramen , Singapore hold the greater number of record holding ramen. Like this we can work on this dataset to derive new conclusions and derivations. It has no limits. Thus keep exploring :)