

CS1__OmarShatrat__AliKamel

Omar Shatrat

September 25, 2018

```
load("C:\\Users\\Omar S\\Downloads\\ramen.RData")
```

```
length(unique(ramen$Brand))
```

```
## [1] 355
```

1. 355 different brands are reviewed in the dataset. I recognize brands such as Nissin and Maruchan.

```
library(stringi)
```

```
library(stringr)
```

```
chrTopTen <- as.character(ramen$Top.Ten)
```

```
topTenYears <- str_sub(chrTopTen[chrTopTen != ""], stri_locate(chrTopTen[chrTopTen != ""],  
                                                                pattern = "20", fixed = T)[,2]-2, -4)
```

```
unique(stri_trim(topTenYears))
```

```
## [1] "2016" "2015" "2013" "2014" "2012"
```

2. We have Top Ten Data from 2012 - 2016

```
usa <- ramen$Brand[ramen$Country == "USA"]
```

```
unique(usa)
```

```
## [1] Nissin Yamachan
## [3] Jackpot Teriyaki Lipton
## [5] Pringles Myojo
## [7] Daifuku Dream Kitchen
## [9] Dr. McDougall's Shirakiku
## [11] Mama Pat's Goku-Uma
## [13] Gefen Farmer's Heart
## [15] Nongshim Maruchan
## [17] Roland Koyo
## [19] IbuRamen Fortune
## [21] Thai Smile Sapporo Ichiban
## [23] Crystal Noodle Authentically Asian
## [25] One Dish Asia Thai Pavilion
## [27] Osaka Ramen Annie Chun's
## [29] Snapdragon Miracle Noodle
## [31] Lotus Foods Sakura Noodle
## [33] Thai Kitchen Komforte Chockolates
## [35] Tasty Bite Star Anise Foods
## [37] Tradition Sun Noodle
```

```
## [39] S&S                Right Foods
## [41] Hosooniyi          Mexi-Ramen
## [43] Chikara            US Canning
## [45] Tayho              Fu Chang Chinese Noodle Company
## [47] Teriyaki Time      Smack
## [49] Westbrae
## 355 Levels: 1 To 3 Noodles 7 Select 7 Select/Nissin ... Zow Zow
```

3. there are 49 Ramen brands from the USA.

```
ramenNum1 <- str_detect(ramen$Top.Ten, pattern = "#1") & !str_detect(ramen$Top.Ten, pattern = "#10")
ramen$Brand[ramenNum1]

## [1] Prima Taste MyKuali      MyKuali      Prima Taste Indomie
## 355 Levels: 1 To 3 Noodles 7 Select 7 Select/Nissin ... Zow Zow
```

4. MyKuali and Prima Taste have both won #1 multiple times.

```
country_rating <- data.frame(ramen)

avg_ratings <- data.frame(ramen_brand = unique(ramen$Brand), avg_rating = NA)

for(i in avg_ratings$ramen_brand){avg_ratings[avg_ratings$ramen_brand == i, "avg_rating"] <-
  mean(country_rating[country_rating$Brand == i, "Stars"], na.rm = T)
}

cat("Maximum Average Rating:\t")

## Maximum Average Rating:
max(avg_ratings$avg_rating)

## [1] 5
cat("\n")

cat("Brands With Maximum Average Rating:\n\n")

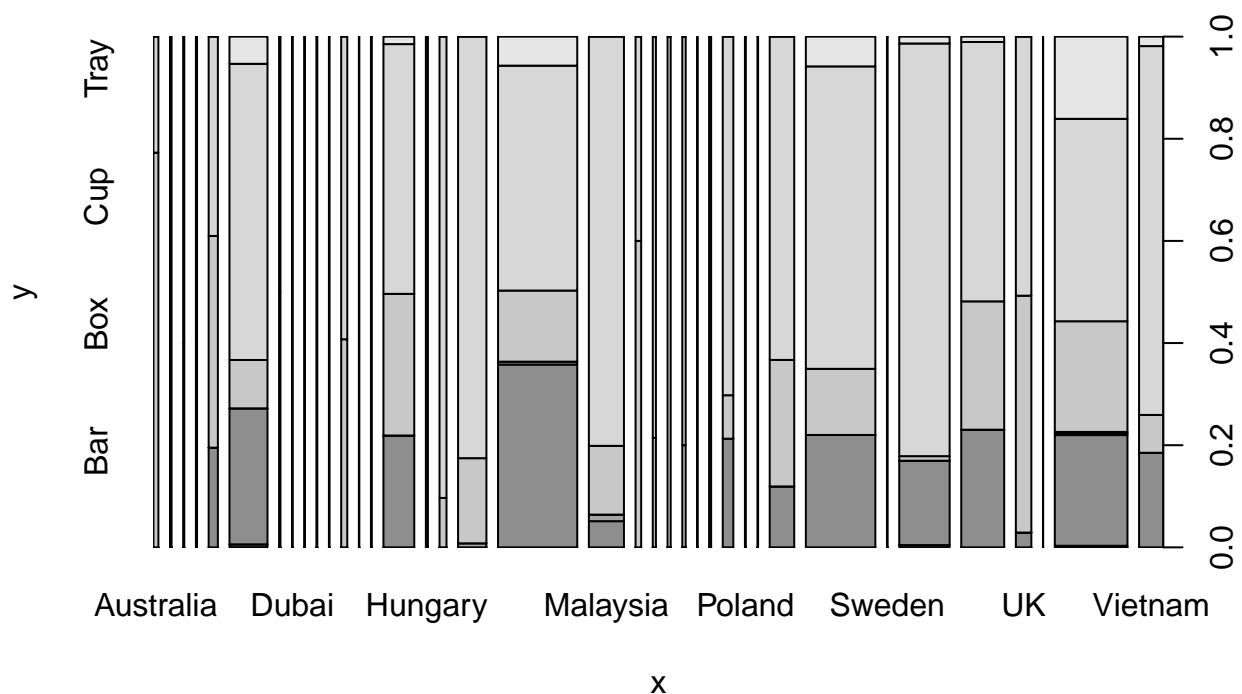
## Brands With Maximum Average Rating:
fiveStar <- avg_ratings[avg_ratings$avg_rating == "5", "ramen_brand"]
fiveStar

## [1] Tao Kae Noi      Jackpot Teriyaki    Prima
## [4] Takamori         The Ramen Rater Select Kiki Noodle
## [7] Daifuku          Patanjali          Nyor Nyar
## [10] ORee Garden      Prima Taste        MyOri
## [13] Seven & I        Torishi            ChoripDong
## [16] The Bridge       Plats Du Chef      Kimura
```

```
## [19] Peyang Komforte Chockolates Daddy
## [22] Song Hak Foodmon Higashi
## 355 Levels: 1 To 3 Noodles 7 Select 7 Select/Nissin ... Zow Zow
```

5. There are 24 brands that have the highest average rating of 5 stars.

```
plot(ramen$Country, ramen$Style)
```



6. From the plot, we notice that packaging proportions for countries do not vary by significant amounts. There are a few countries which rely more heavily on certain packaging types, but for the most part, the distribution is similar.

```
cat("USA Min, Max, Avg.\n")
```

```
## USA Min, Max, Avg.
```

```
ratingsUSA <- ramen$Stars[ramen$Country == "USA"]
```

```
min(ratingsUSA)
```

```

## [1] 0
max(ratingsUSA)

## [1] 5
mean(ratingsUSA)

## [1] 3.457043
cat("\n")
cat("Singapore Min, Max, Avg.\n")

## Singapore Min, Max, Avg.
ratingsSingapore <- ramen$Stars[ramen$Country == "Singapore"]
min(ratingsSingapore)

## [1] 2
max(ratingsSingapore)

## [1] 5
mean(ratingsSingapore)

## [1] 4.126147
cat("\n")
cat("Top Ten Appearances\n")

## Top Ten Appearances
topTenUSA <- ramen$Top.Ten[ramen$Country == "USA"]
countTopTenUSA <- str_count(topTenUSA, pattern = "201")
cat("USA")

## USA
sum(countTopTenUSA)

## [1] 1
topTenSingapore <- ramen$Top.Ten[ramen$Country == "Singapore"]
countTopTenSingapore <- str_count(topTenSingapore, pattern = "201")
cat("Singapore")

## Singapore
sum(countTopTenSingapore)

## [1] 7

```

7. The minimum, maximum, and average ratings respectively for the US are 0, 5, and 3.46 stars.

The minimum, maximum, and average ratings respectively for Singapore are 2, 5, and 4.13 stars.

I expect Singapore to be in the Top Ten more often because they have a higher average and a higher minimum rating.

I am correct because Singapore appears 7 times in the Top Ten while the US only appears once.

```
cat("Production by Country:\n")
```

```
## Production by Country:
```

```
cat("\n")
```

```
summary(ramen$Country)
```

##	Australia	Bangladesh	Brazil	Cambodia	Canada
##	22	7	5	5	41
##	China	Colombia	Dubai	Estonia	Fiji
##	169	6	3	2	4
##	Finland	Germany	Ghana	Holland	Hong Kong
##	3	27	2	4	137
##	Hungary	India	Indonesia	Japan	Malaysia
##	9	31	126	352	156
##	Mexico	Myanmar	Nepal	Netherlands	Nigeria
##	25	14	14	15	1
##	Pakistan	Philippines	Poland	Sarawak	Singapore
##	9	47	4	3	109
##	South Korea	Sweden	Taiwan	Thailand	UK
##	309	3	224	191	69
##	United States	USA	Vietnam		
##	1	323	108		

```
cat("\n")
```

```
mostProduction <- which.max(summary(ramen$Country))
```

```
cat("Country with Largest Production:\n")
```

```
## Country with Largest Production:
```

```
mostProduction
```

```
## Japan
```

```
## 19
```

```
cat("\n")
```

```
cat("Country that made the Top 10 most frequently:\n ")
```

```
## Country that made the Top 10 most frequently:
##
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

countryTopTen <- ramen$Country[str_detect(ramen$Top.Ten, pattern = "#")]
Mode(countryTopTen)

## [1] Singapore
## 38 Levels: Australia Bangladesh Brazil Cambodia Canada China ... Vietnam
cat("\n")

cat("Number of times said country made the Top 10:\n")

## Number of times said country made the Top 10:
#Finds the number of times the Country Made the Top Ten
numTimesTopTen <- table(ramen$Country[str_detect(ramen$Top.Ten, pattern = "#")])
maxNumTimesTopTen <- max(numTimesTopTen)
maxNumTimesTopTen

## [1] 7
```

8. The country that makes the most ramen is Japan. There are index number 19 in the table and appeared 352 times, the most of any country.

We defined best as which country's ramen appeared in the Top Ten the most. We feel that this is better than taking the average rating because the average is more a measure of consistency. Finding the number of Top Ten appearances shows ramen ratings for countries at their peak.

Singapore makes the best ramen. They appeared in the Top Ten more times than any other country (7)

```
variety <- as.character(ramen$Variety)
spicy <- str_detect(variety, "spicy") | str_detect(variety, "Spicy")

spicyLen <- length(spicy[spicy == "TRUE"])
notSpicyLen <- length(spicy[spicy == "FALSE"])

cat("Total Spicy Ramen:\t")

## Total Spicy Ramen:
```

```

spicyLen

## [1] 270
cat("Total Non-Spicy Ramen\t")

## Total Non-Spicy Ramen
notSpicyLen

## [1] 2310
cat("\n")

cat("Mean rating for spicy ramen:\t")

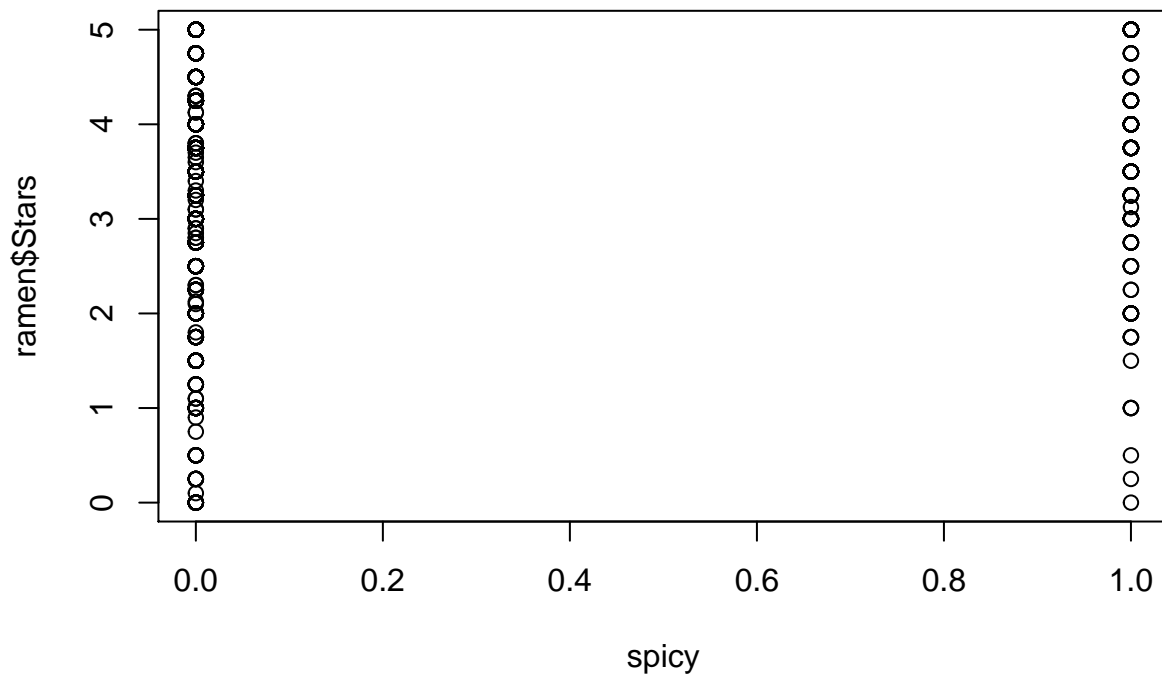
## Mean rating for spicy ramen:
mean(ramen$Stars[spicy == "TRUE"])

## [1] 3.688426
cat("Mean rating for non-spicy ramen:\t")

## Mean rating for non-spicy ramen:
mean(ramen$Stars[spicy == "FALSE"], na.rm = "TRUE")

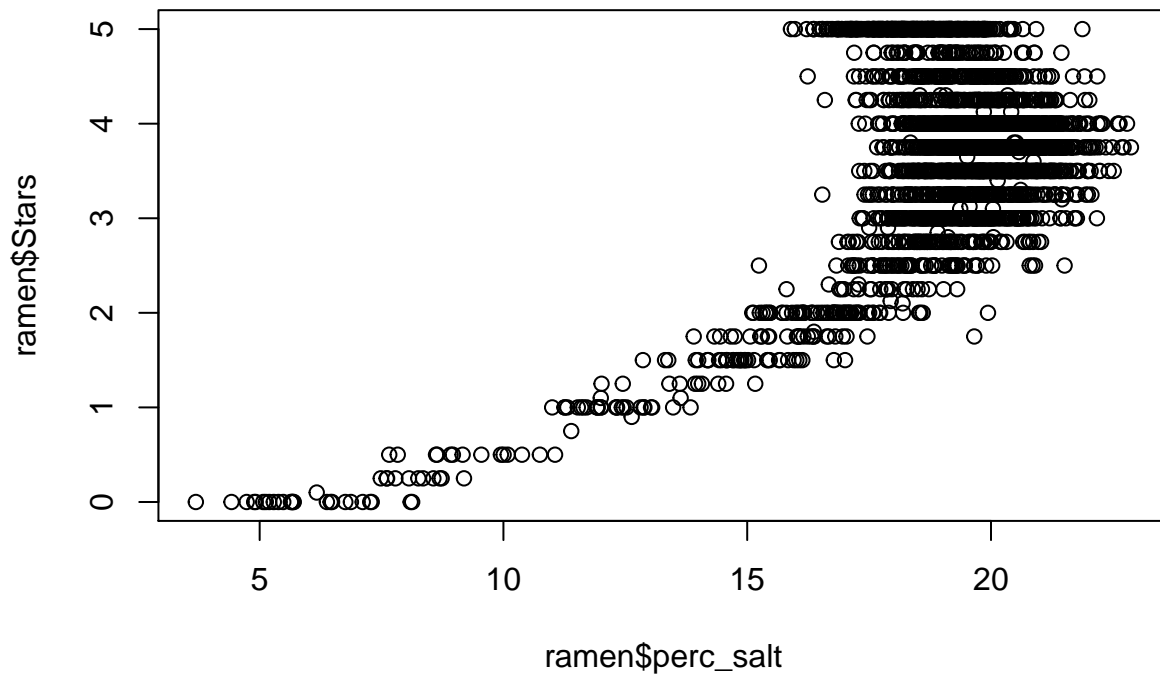
## [1] 3.650726
plot(spicy, ramen$Stars)

```



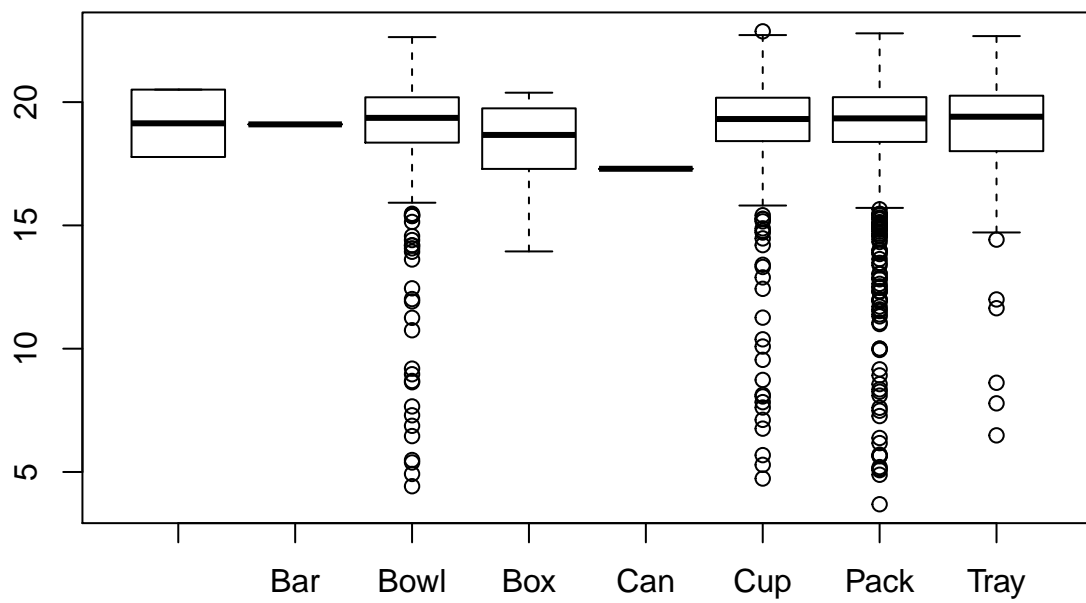
9. 270 types of Ramen are considered spicy. The average rating for spicy ramen is 3.69 stars while the average rating for non-spicy ramen is 3.65 stars. The plot shows non-spicy ramen to be more closely clustered within the 3-4 star range while spicy ramen tends to be distributed more equally accross the rating spectrum. Spicy ramen does not tend to be rated much higher than non-spicy ramen.

```
plot(ramen$perc_salt, ramen$Stars)
```

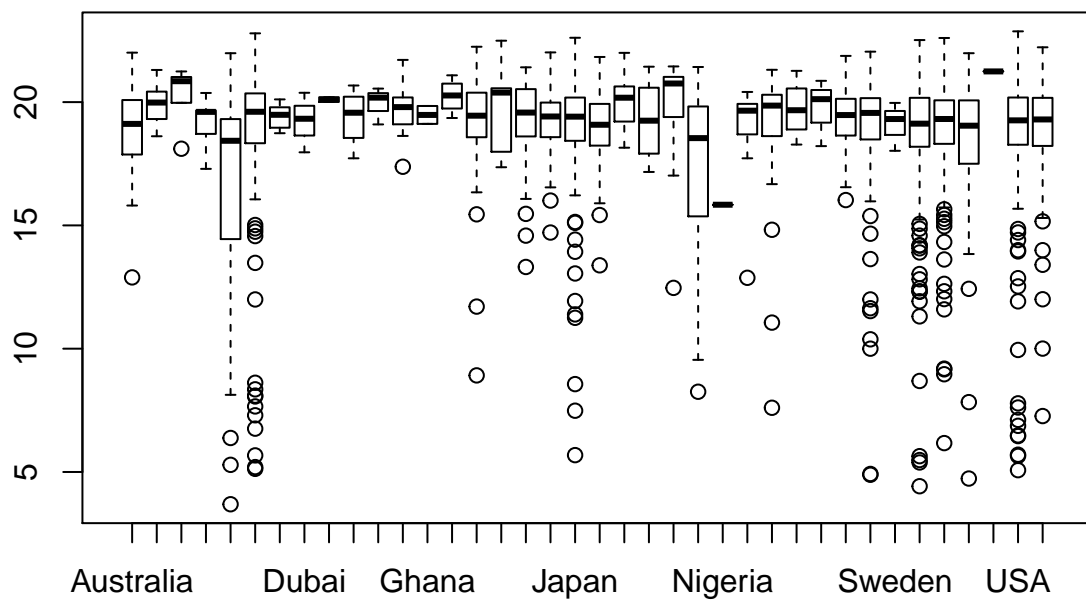


10. As the salt content goes up, the ramen ratings tend to increase.

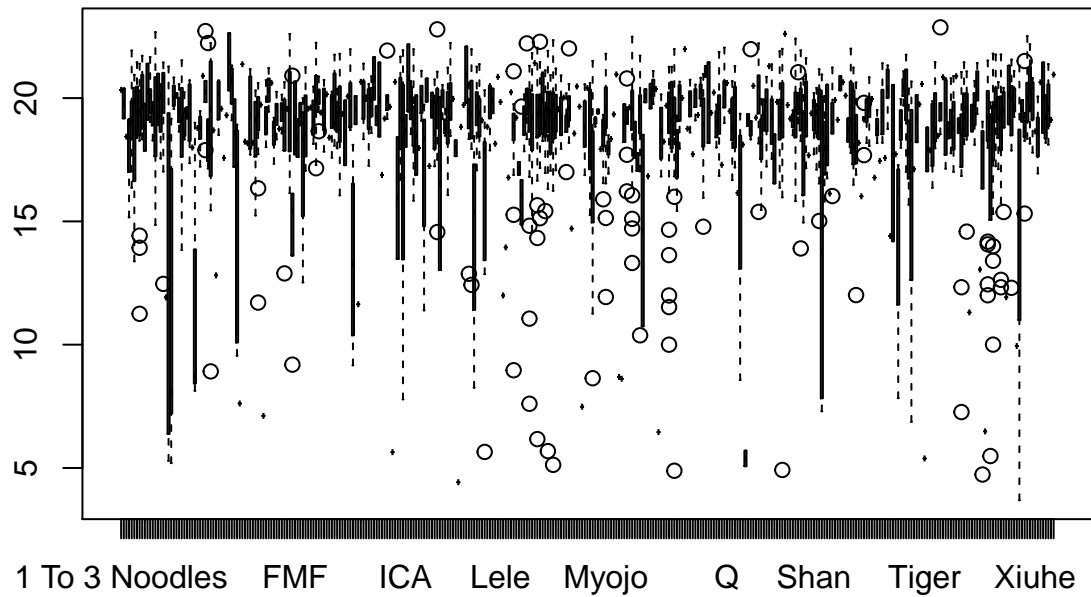
```
plot(ramen$Style, ramen$perc_salt)
```

```
plot(ramen$Country, ramen$perc_salt)
```



```
plot(ramen$Brand, ramen$perc_salt)
```



```
plot(ramen$Variety, ramen$perc_salt)
```

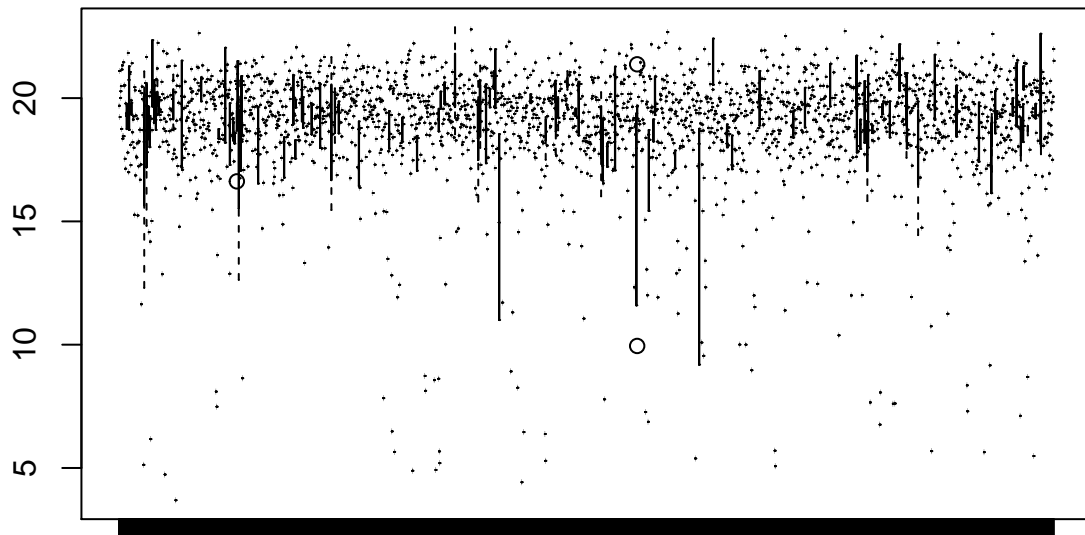
```
## Warning in axis(side = 1, at = 1:2413, labels = c("\A\" Series Artificial
## Chicken", : font width unknown for character 0x81

## Warning in axis(side = 1, at = 1:2413, labels = c("\A\" Series Artificial
## Chicken", : font width unknown for character 0x9d

## Warning in axis(side = 1, at = 1:2413, labels = c("\A\" Series Artificial
## Chicken", : font width unknown for character 0x90

## Warning in axis(side = 1, at = 1:2413, labels = c("\A\" Series Artificial
## Chicken", : font width unknown for character 0x9d

## Warning in axis(side = 1, at = 1:2413, labels = c("\A\" Series Artificial
## Chicken", : font width unknown for character 0x1d
```



"A" Series Artificial Chicken Duck Itnok Miso Pizza Shio Tom Yum

11. Style has nearly no affect on the salt content present in ramen. There are many outliers for Bowl, Cup, and Pack, but the boxplots are nearly identical in terms of their median and quartiles.

A ramen's country does not appear to have a huge effect on its salt content. Similar to Style, there are many outliers but the overwhelming proportion of salt is similar accross all countries.

Most Ramen brands use large amounts of salt. There is relatively low variance.

Despite the type or flavor of ramen you choose, there is likely a high salt content.

```
cat("0 to 5%\n\n")
```

```
## 0 to 5%
```

```

length(ramen$perc_salt[ramen$perc_salt <= 5])

## [1] 5
mean(ramen$Stars[ramen$perc_salt <= 5], na.rm = TRUE)

## [1] 0
zeroToFiveCountry <- ramen$Country[ramen$perc_salt <= 5]
which.max(summary(zeroToFiveCountry))

## South Korea
##          31
max(table(zeroToFiveCountry))

## [1] 2
cat("\n")
cat("5 to 10%\n\n")

## 5 to 10%
length(ramen$perc_salt[ramen$perc_salt <= 10 & ramen$perc_salt >= 5])

## [1] 43
mean(ramen$Stars[ramen$perc_salt <= 10 & ramen$perc_salt >= 5], na.rm = TRUE)

## [1] 0.175
fiveToTenCountry <- ramen$Country[ramen$perc_salt <= 10 & ramen$perc_salt >= 5]
which.max(summary(fiveToTenCountry))

## China
##          6
max(table(fiveToTenCountry))

## [1] 10
cat("\n")
cat("10 to 15%\n\n")

## 10 to 15%
length(ramen$perc_salt[ramen$perc_salt <= 15 & ramen$perc_salt >= 10])

## [1] 73
mean(ramen$Stars[ramen$perc_salt <= 15 & ramen$perc_salt >= 10], na.rm = TRUE)

## [1] 1.20625
tenToFifteenCountry <- ramen$Country[ramen$perc_salt <= 15 & ramen$perc_salt >= 10]
which.max(summary(tenToFifteenCountry))

## Taiwan
##          33
max(table(tenToFifteenCountry))

```

```

## [1] 13
cat("\n")

cat("15 to 20%\n\n")

## 15 to 20%
length(ramen$perc_salt[ramen$perc_salt <= 20 & ramen$perc_salt >= 15])

## [1] 1677
mean(ramen$Stars[ramen$perc_salt <= 20 & ramen$perc_salt >= 15], na.rm = TRUE)

## [1] 3.806548
fifteenToTwentyCountry <- ramen$Country[ramen$perc_salt <= 20 & ramen$perc_salt >= 15]
which.max(summary(fifteenToTwentyCountry))

## Japan
##      19
max(table(fifteenToTwentyCountry))

## [1] 238
cat("\n")

cat("20-inf%\n\n")

## 20-inf%
length(ramen$perc_salt[ramen$perc_salt >= 20])

## [1] 783
mean(ramen$Stars[ramen$perc_salt >= 20], na.rm = TRUE)

## [1] 3.764866
twentyAndAboveCountry <- ramen$Country[ramen$perc_salt >= 20]
which.max(summary(twentyAndAboveCountry))

## Japan
##      19
max(table(twentyAndAboveCountry))

## [1] 105

```

12. We decided to separate ramen based off of its salt content. Our intervals are 0-5%, 5-10%, 10-15%, 15-20%, and 20-inf%.

There are 5 ramens with a salt content under 5%. The average rating was 0 stars. South Korea produced the largest quantity of these ramen (2/5).

There are 43 ramen with a salt content between 5 and 10%. The average rating was .175 stars. China produced the largest quantity of these ramen (10/43).

There are 73 ramen with a salt content between 10 and 15%. The average rating was 1.21 stars. Taiwan produced the largest quantity of these ramen (13/73).

There are 1677 ramen with a salt content between 15 and 20%. The average rating was 3.81 stars, Japan produced the largest quantity of these ramen (238/1677).

There are 783 ramen with a salt content above 20%. The average rating was 3.76 stars. Japan produced the largest quantity of these ramen (105/783).

From this we can get a general overview of the ramen market. Most ramen seems to be produced by Asian countries, with Japan shouldering the biggest load. There are very few ramen with a salt content below 15%, and they all tend to receive horrible ratings. The large majority of ramen contain above 15% of salt; their ratings are dramatically higher. It is as though high salt is a prerequisite for receiving high ratings. After that, more variables such as flavor and taste come into play.

```
cat("Mean Star Rating for Ramen in Top 10:\t")

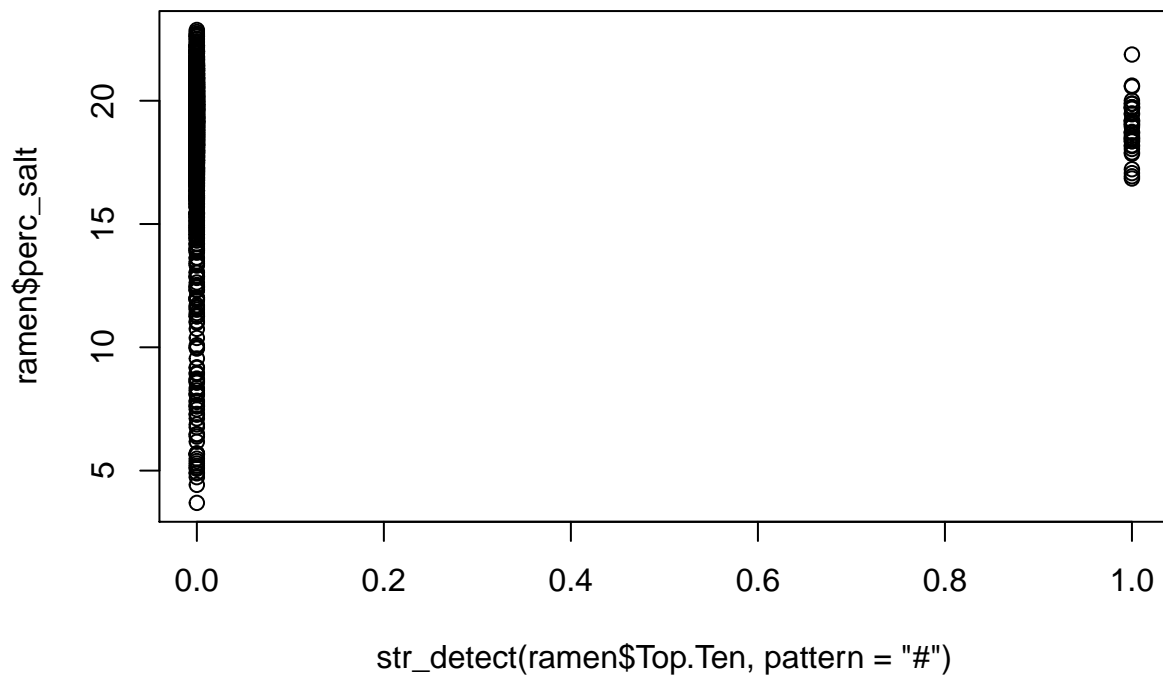
## Mean Star Rating for Ramen in Top 10:
mean(ramen$Stars[str_detect(ramen$Top.Ten, pattern = "#")])

## [1] 4.918919
```

```
cat("Min Star Rating for Ramen in Top 10:\t")

## Min Star Rating for Ramen in Top 10:
min(ramen$Stars[str_detect(ramen$Top.Ten, pattern = "#")])

## [1] 4
plot(str_detect(ramen$Top.Ten, pattern = "#"), ramen$perc_salt)
```



```
cat("Mean Rating for Japanese Ramen:\t")

## Mean Rating for Japanese Ramen:
mean(ramen$Stars[ramen$Country == "Japan"])

## [1] 3.981605
cat("Mean Rating for Singaporean Ramen")

## Mean Rating for Singaporean Ramen
mean(ramen$Stars[ramen$Country == "Singapore"])

## [1] 4.126147
cat("\n")

curry <- str_detect(variety, "curry") | str_detect(variety, "Curry")
curryLen <- length(curry[curry == "TRUE"])
cat("Total Ramen Containing Curry:\t")
```



```

## Total Ramen Containing Curry:
curryLen

## [1] 126
notCurryLen <- length(curry[curry == "FALSE"])
cat("Total Ramen Not Containing Curry:\t")

## Total Ramen Not Containing Curry:
notCurryLen

## [1] 2454
cat("Mean Rating for Ramen Containing Curry:\t")

## Mean Rating for Ramen Containing Curry:
mean(ramen$Stars[curry == "TRUE"])

## [1] 4.121032
cat("Mean Rating for Ramen Not Containing Curry:\t")

## Mean Rating for Ramen Not Containing Curry:
mean(ramen$Stars[curry == "FALSE"], na.rm = "TRUE")

## [1] 3.630702
cat("\n")

chicken <- str_detect(variety, "chicken") | str_detect(variety, "Chicken")
chickenLen <- length(chicken[chicken == "TRUE"])
cat("Total Ramen Containing Chicken:\t")

## Total Ramen Containing Chicken:
chickenLen

## [1] 328
notChickenLen <- length(chicken[chicken == "FALSE"])
cat("Total Ramen Not Containing Chicken:\t")

## Total Ramen Not Containing Chicken:
notChickenLen

## [1] 2252
cat("Mean Rating for Ramen Containing Chicken:\t")

## Mean Rating for Ramen Containing Chicken:
mean(ramen$Stars[chicken == "TRUE"])

## [1] 3.394588
cat("Mean Rating for Ramen Not Containing Chicken:\t")

## Mean Rating for Ramen Not Containing Chicken:

```

```

mean(ramen$Stars[chicken == "FALSE"], na.rm = "TRUE")

## [1] 3.692608
cat("\n")

beef <- str_detect(variety, "beef") | str_detect(variety, "Beef")
beefLen <- length(beef[beef == "TRUE"])
cat("Total Ramen Containing Beef:\t")

## Total Ramen Containing Beef:
beefLen

## [1] 232
notBeefLen <- length(chicken[chicken == "FALSE"])
cat("Total Ramen Not Containing Beef:\t")

## Total Ramen Not Containing Beef:
notBeefLen

## [1] 2252
cat("Mean Rating for Ramen Containing Beef:\t")

## Mean Rating for Ramen Containing Beef:
mean(ramen$Stars[beef == "TRUE"])

## [1] 3.509591
cat("Mean Rating for Ramen Not Containing Beef:\t")

## Mean Rating for Ramen Not Containing Beef:
mean(ramen$Stars[beef == "FALSE"], na.rm = "TRUE")

## [1] 3.66903
cat("\n")

vegetable <- str_detect(variety, "vegetable") | str_detect(variety, "Vegetable")
vegetableLen <- length(vegetable[vegetable == "TRUE"])
cat("Total Ramen Containing Vegetables:\t")

## Total Ramen Containing Vegetables:
vegetableLen

## [1] 66
notVegetableLen <- length(vegetable[vegetable == "FALSE"])
cat("Total Ramen Not Containing Vegetables:\t")

## Total Ramen Not Containing Vegetables:
notVegetableLen

## [1] 2514
cat("Mean Rating for Ramen Containing Vegetables:\t")

## Mean Rating for Ramen Containing Vegetables:

```

```

mean(ramen$Stars[vegetable == "TRUE"])

## [1] 3.295455
cat("Mean Rating for Ramen Not Containing Vegetables:\t")

## Mean Rating for Ramen Not Containing Vegetables:
mean(ramen$Stars[vegetable == "FALSE"], na.rm = "TRUE")

## [1] 3.664118
cat("\n\n")

cat("Number of Times Keyword Shows Up in Top 10:\n\n")

## Number of Times Keyword Shows Up in Top 10:
cat("Curry:\n")

## Curry:
table(curry[str_detect(ramen$Top.Ten, "#"))

##
## FALSE TRUE
##    31    6
cat("\n")

cat("Spicy:\n")

## Spicy:
table(spicy[str_detect(ramen$Top.Ten, "#"))

##
## FALSE TRUE
##    35    2
cat("\n")

cat("Chicken:\n")

## Chicken:
table(chicken[str_detect(ramen$Top.Ten, "#"))

##
## FALSE TRUE
##    33    4
cat("\n")

cat("Beef:\n")

## Beef:
table(beef[str_detect(ramen$Top.Ten, "#"))

##
## FALSE TRUE
##    36    1

```

```

cat("\n")

cat("Vegetable:\n")

## Vegetable:
table(vegetable[str_detect(ramen$Top.Ten, "#")])

##
## FALSE
##      37
cat("\n")

lowFat <- str_detect(variety, "Low")
lowFatLen <- length(lowFat[lowFat == "TRUE"])
cat("Number of Ramen That Are Low Fat:\t")

## Number of Ramen That Are Low Fat:
lowFatLen

## [1] 4
cat("Mean Rating for Low Fat Ramen:\t")

## Mean Rating for Low Fat Ramen:
mean(ramen$Stars[lowFat == "TRUE"])

## [1] 2.1875

```

13. Because of our results for #10, we strongly recommend that Roamin' Ramen uses a high amount of salt to maximize their chances of getting a favorable rating. Our plot shows that every single Top Ten winner has used a salt percentage above 15.

Among 37 Top Ten winners, the minimum rating was 4 stars. The average was 4.91 stars.

We also recommend for Roamin' Ramen to base their operations out of Singapore or possibly shadow Singaporean brands in order to observe why they score in the Top Ten so frequently. The average rating for a Singaporean ramen was 4.12 stars. Keep in mind that no rating under 4 stars made the top ten and the overwhelming majority of ratings that did were rated much higher than just 4 stars. Japan, the country with the highest level of ramen production, averaged a rating of just under 4 stars.

We recommend for Roamin' Ramen to make their ramen curry-flavored, or at least put "curry" in the title. Out of 126 ramens containing the phrase "curry", the average rating was 4.13 stars. The counter-average (average for ramens not containing the phrase "curry") was 3.63 stars.

Upon further analysis, we recommend for Roamin Ramen to not use the terms "chicken", "beef", or "vegetable" when describing their ramen, as these ramens have an average rating lower than their counter-average rating. Among 328 ramens containing the phrase "chicken", the average rating was 3.4 stars. For beef, it was 3.51 stars. For vegetable, it was 3.3 stars. The respective counter-averages were 3.7, 3.67, and 3.66 stars.

It should be noted that among the 37 ramens that did make the Top Ten, 6 contained the phrase "curry". 2 contained "spicy", 4 contained "chicken", 1 contained "beef", and 0 contained "vegetable". It clearly seems as though ramen curries are more favorable when it comes to rankings.

Our final recommendation for Roamin' Ramen is to avoid taking the healthy route. Among the 4 ramens that were described as "Low Fat", the average rating was 2.19 stars. Roamin' Ramen should focus more on flavor in order to make the Top 10.