

# Ramen

*Devin Adams, Fiona Almeyda Wijaya, Natalia Jones*

*September 27, 2018*

## Question 1

```
#loading libraries
library(stringi)

library(stringr)

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)

library(dplyr)

library(formattable)

load("/Users/fionalmeyda/Downloads/ramen.Rdata")

uniqueRamen <- unique(tolower(ramen$Brand))

summary(uniqueRamen)

##      Length      Class      Mode
##      353 character character
```

## Question 2

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

results <- stri_sub(topTen[str_detect(topTen, "#")])

results <- unique(stri_sub(results, 1, 4))

results
```

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

### Question 3

```
ramenCountries <- subset(ramen, ramen$Country == "USA" | ramen$Country == "United States", select = c("Brand", "Top.Ten"))
unique(tolower(ramenCountries$Brand))
```

```
## [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] "hosoonyi" "mexi-ramen"
## [43] "chikara" "us canning"
## [45] "tayho" "fu chang chinese noodle company"
## [47] "teriyaki time" "smack"
## [49] "westbrae"
```

### Question 4

```
ramenRep <- ramen[grep("#1", ramen$Top.Ten), c("Brand", "Top.Ten")]
ramenRep <- ramenRep[!grepl("#10", ramenRep$Top.Ten), c("Brand", "Top.Ten")]
ramenRep$Brand <- tolower(ramenRep$Brand)
ramenRep <- ramenRep %>% add_count(Brand)
ramenRep
```

```
## # A tibble: 5 x 3
##   Brand      Top.Ten    n
##   <chr>      <fct>  <int>
## 1 prima taste 2016 #1      2
```

```
## 2 mykuali      2015 #1      2
## 3 mykuali      2014 #1      2
## 4 prima taste 2013 #1      2
## 5 indomie      2012 #1      1
```

## Question 5

```
sumRamen <- aggregate(ramen[,5], list(tolower(ramen$Brand)), mean)

names(sumRamen) <- c("Brand", "Stars")

sumRamen <- sumRamen[sumRamen$Stars == max(sumRamen$Stars, na.rm = T) & !is.na(sumRamen$Stars), c("Brand", "Stars")]

sumRamen
```

```
##              Brand Stars
## 40          choripdong    5
## 48              daddy    5
## 49          daifuku    5
## 68          foodmon    5
## 100          higashi    5
## 118    jackpot teriyaki    5
## 128          kiki noodle    5
## 131          kimura    5
## 137    komforte chockolates    5
## 184              myori    5
## 198          nyor nyar    5
## 205          oree garden    5
## 214          patanjali    5
## 216          peyang    5
## 218    plats du chef    5
## 225          prima    5
## 226    prima taste    5
## 260          seven & i    5
## 271          song hak    5
## 287          takamori    5
## 289          tao kae noi    5
## 300          the bridge    5
## 302 the ramen rater select    5
## 309          torishi    5
```

## Question 6

```
ramenPack <- ramen %>% add_count(Country, Style)

colnames(ramenPack)[8] <- ("Total")

ramenPack <- ramenPack[!duplicated(ramenPack[c("Country", "Style")]), c("Country", "Style", "Total")]

ramenPack <- group_by(ramenPack, Country) %>% mutate(Percent = round(100 * Total / sum(Total), digits = 0))
```

```
ramenPack <- ramenPack[order(ramenPack$Country),]
```

```
ramenPack
```

```
## # A tibble: 89 x 4
## # Groups:   Country [38]
##   Country   Style Total Percent
##   <fct>     <fct> <int>   <dbl>
## 1 Australia Cup      17      77
## 2 Australia Pack      5      23
## 3 Bangladesh Pack      7     100
## 4 Brazil    Cup      2      40
## 5 Brazil    Pack      3      60
## 6 Cambodia  Pack      5     100
## 7 Canada    Cup     17      41
## 8 Canada    Pack     16      39
## 9 Canada    Bowl      8      20
## 10 China    Bowl     45      27
## # ... with 79 more rows
```

## Question 7

```
ramenUS <- ramen[ramen$Country == "United States" | ramen$Country == "USA", c("Brand", "Variety", "Country", "Stars")]
```

```
ramenUsResults <- data.frame("US", min(ramenUS$Stars), mean(ramenUS$Stars), max(ramenUS$Stars))
```

```
colnames(ramenUsResults) <- c("Country", "Min", "Average", "Max")
```

```
ramenUsResults
```

```
##   Country Min Average Max
## 1      US  0 3.457948   5
```

```
ramenSingapore <- ramen[ramen$Country == "Singapore", c("Brand", "Variety", "Country", "Stars")]
```

```
ramenSingaporeResults <- data.frame("Singapore", min(ramenSingapore$Stars), mean(ramenSingapore$Stars), max(ramenSingapore$Stars))
```

```
colnames(ramenSingaporeResults) <- c("Country", "Min", "Average", "Max")
```

```
ramenSingaporeResults
```

```
##   Country Min Average Max
## 1 Singapore  2 4.126147   5
```

```
ramenUsSing <- ramen[grep("#", ramen$Top.Ten), c("Country", "Top.Ten")]
```

```
ramenUsSing <- ramenUsSing[ramenUsSing$Country == "USA" | ramenUsSing$Country == "United States" | ramenUsSing$Country == "Singapore", ]
```

```
ramenUsSing
```

```
##   Country Top.Ten
## 634 Singapore 2016 #1
## 656 Singapore 2016 #8
```

```
## 674 Singapore 2016 #5
## 1310 Singapore 2014 #8
## 1589 Singapore 2013 #1
## 1590 Singapore 2013 #2
## 1647 USA 2013 #4
## 1973 Singapore 2012 #10
```

## Question 8

```
new_table<- table(ramen$Country)

new_dataframe<-as.data.frame(new_table)

names(new_dataframe)[1]='ramencountry'

max(new_dataframe$Freq)
```

```
## [1] 352
```

```
new_dataframe$ramencountry[new_dataframe$Freq==352]
```

```
## [1] Japan
```

```
## 38 Levels: Australia Bangladesh Brazil Cambodia Canada China ... Vietnam
```

```
max(ramen$Stars,na.rm=T)
```

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

```
country_star<-as.data.frame(tapply(ramen$Stars, ramen$Country, mean))
```

```
names(country_star)[1]= "AverageRating"
```

```
country_star
```

```
##           AverageRating
## Australia      3.138636
## Bangladesh     3.714286
## Brazil         4.350000
## Cambodia       4.200000
## Canada         2.243902
## China          3.421893
## Colombia       3.291667
## Dubai          3.583333
## Estonia        3.500000
## Fiji           3.875000
## Finland        3.583333
## Germany        3.638889
## Ghana          3.500000
## Holland        3.562500
## Hong Kong      3.801825
## Hungary        3.611111
## India          3.395161
## Indonesia      4.067460
## Japan          3.981605
```

```
## Malaysia          NA
## Mexico             3.730000
## Myanmar            3.946429
## Nepal              3.553571
## Netherlands        2.483333
## Nigeria            1.500000
## Pakistan           3.000000
## Philippines        3.329787
## Poland             3.625000
## Sarawak            4.333333
## Singapore          4.126147
## South Korea        NA
## Sweden             3.250000
## Taiwan             3.665402
## Thailand           3.384817
## UK                 2.997101
## United States      3.750000
## USA                3.457043
## Vietnam            3.187963
```

## Question 9

```
spicy <- ramen[str_detect(tolower(ramen$Variety), "spicy"), c("Variety", "Stars")]
str(spicy)
```

```
## 'data.frame':   270 obs. of  2 variables:
## $ Variety: Factor w/ 2413 levels "\A\" Series Artificial Chicken",...: 1446 2230 363 528 718 1216 6
## $ Stars : num  1 2.5 4 5 3.25 5 3.75 3.75 3.5 4.75 ...
```

```
mild <- ramen[!str_detect(tolower(ramen$Variety), "spicy"), c("Variety", "Stars")]
str(mild)
```

```
## 'data.frame':   2310 obs. of  2 variables:
## $ Variety: Factor w/ 2413 levels "\A\" Series Artificial Chicken",...: 2193 456 721 1955 1110 2056
## $ Stars : num  3.75 2.25 2.75 3.75 4.75 4 3.75 0.25 5 5 ...
```

## Question 10

```
summary(ramen$perc_salt)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  3.691  18.372  19.340  18.951  20.198  22.870
```

```
firstq <- ramen[ramen$perc_salt <= 18.372, c("Variety", "Stars")]
```

```
secondq <- ramen[ramen$perc_salt > 18.372 & ramen$perc_salt <= 19.340, c("Variety", "Stars")]
```

```
thirdq <- ramen[ramen$perc_salt > 19.340 & ramen$perc_salt <= 20.198, c("Variety", "Stars")]
```

```
fourthq <- ramen[ramen$perc_salt > 20.198 & ramen$perc_salt <= 22.870, c("Variety", "Stars")]
```

```
mean(firstq[["Stars"]], na.rm = TRUE)
```

```
## [1] 3.117691
mean(secondq[["Stars"]], na.rm = TRUE)

## [1] 3.915969
mean(thirdq[["Stars"]], na.rm = TRUE)

## [1] 3.824728
mean(fourthq[["Stars"]], na.rm = TRUE)

## [1] 3.759031
```

## Question 11

```
##Question #11

#Country
ramenSaltCountry <- ramen %>% add_count(Country)

colnames(ramenSaltCountry)[8] <- ("Total")

ramenSaltCountry2 <- group_by(ramenSaltCountry, Country) %>% mutate(AvgSalt = round(sum(perc_salt) / Total, 2))
ramenSaltCountry3 <- ramenSaltCountry2[!duplicated(ramenSaltCountry2[c("Country")]), c("Country", "Total", "AvgSalt")]
ramenSaltCountry3

## # A tibble: 38 x 3
## # Groups:   Country [38]
##   Country      Total AvgSalt
##   <fct>      <int>   <dbl>
## 1 Japan        352    19.2
## 2 Taiwan       224    18.6
## 3 USA          323    18.8
## 4 India         31    19.1
## 5 South Korea  309    19.1
## 6 Singapore   109    19.4
## 7 Thailand    191    18.9
## 8 Hong Kong   137    19.4
## 9 Vietnam     108    18.9
## 10 Ghana        2    19.5
## # ... with 28 more rows

#Style
ramenSaltStyle <- ramen %>% add_count(Style)

colnames(ramenSaltStyle)[8] <- ("Total")

ramenSaltStyle2 <- group_by(ramenSaltStyle, Style) %>% mutate(AvgSalt = round(sum(perc_salt) / Total, 2))
```

```

ramenSaltStyle3 <- ramenSaltStyle2[!duplicated(ramenSaltStyle2[c("Style")]), c("Style", "Total", "AvgSalt")]
ramenSaltStyle3

```

```

## # A tibble: 8 x 3
## # Groups:   Style [8]
##   Style Total AvgSalt
##   <fct> <int>   <dbl>
## 1 Cup    450    18.9
## 2 Pack  1531    19.0
## 3 Tray   108    18.8
## 4 Bowl   481    18.9
## 5 Box     6    18.1
## 6 Can     1    17.3
## 7 Bar     1    19.1
## 8 ""      2    19.1

```

*#Brand*

```

ramenSaltBrand <- ramen %>% add_count(Brand)

colnames(ramenSaltBrand)[8] <- ("Total")

ramenSaltBrand2 <- group_by(ramenSaltBrand, Brand) %>% mutate(AvgSalt = round(sum(perc_salt) / Total , 6))

ramenSaltBrand3 <- ramenSaltBrand2[!duplicated(ramenSaltBrand2[c("Brand")]), c("Brand", "Total", "AvgSalt")]

ramenSaltBrand3

```

```

## # A tibble: 355 x 3
## # Groups:   Brand [355]
##   Brand      Total AvgSalt
##   <fct>      <int>   <dbl>
## 1 New Touch      9    18.9
## 2 Just Way       2    16.5
## 3 Nissin      381    19.5
## 4 Wei Lih       15    19.1
## 5 Ching's Secret  4    19.9
## 6 Samyang Foods  52    19.3
## 7 Acecook       15    18.5
## 8 Ikeda Shoku    2    20.8
## 9 Ripe'n'Dry     3    15.2
## 10 KOKA         25    19.3
## # ... with 345 more rows

```

*#Variety*

```

ramenSaltSpicy <- ramen[str_detect(tolower(ramen$Variety),"spicy"), c("Variety", "Stars", "perc_salt")]

ramenSaltMild <- ramen[str_detect(tolower(ramen$Variety),"mild"), c("Variety", "Stars", "perc_salt")]

ramenSaltNeither <- ramen[!str_detect(tolower(ramen$Variety),"spicy") & !str_detect(tolower(ramen$Variety),"mild"), c("Variety", "Stars", "perc_salt")]

(round(sum(ramenSaltSpicy$perc_salt) / length(ramenSaltSpicy$Variety) , digits = 6))

```



```
## [1] 18.9502
(round(sum(ramenSaltMild$perc_salt) / length(ramenSaltMild$Variety) , digits = 6))

## [1] 19.43494
(round(sum(ramenSaltNeither$perc_salt) / length(ramenSaltNeither$Variety) , digits = 6))

## [1] 18.9479
```

## Question 12

```
not_topten <- subset(ramen,!str_detect(ramen$Top.Ten,"#"))

summary(not_topten$Stars)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##    0.000   3.250   3.750   3.636   4.250   5.000         3

first_quStar <- subset(not_topten, not_topten$Stars <= 3.250 | is.na(not_topten$Stars), c("Variety", "Country"))
second_quStar <- subset(not_topten, not_topten$Stars > 3.250 & not_topten$Stars <= 3.750, c("Variety", "Country"))
third_quStar <- subset(not_topten, not_topten$Stars > 3.750 & not_topten$Stars <= 4.250, c("Variety", "Country"))
fourth_quStar <- subset(not_topten, not_topten$Stars > 4.250 & not_topten$Stars <= 5.000, c("Variety", "Country"))
topten_onlyStar<- subset(ramen, str_detect(ramen$Top.Ten,"#"),c("Variety","Country"))
```

## Question 13

```
#Country
country_star
```

```
##           AverageRating
## Australia           3.138636
## Bangladesh          3.714286
## Brazil              4.350000
## Cambodia            4.200000
## Canada              2.243902
## China               3.421893
## Colombia            3.291667
## Dubai               3.583333
## Estonia             3.500000
## Fiji                3.875000
## Finland             3.583333
## Germany             3.638889
## Ghana               3.500000
## Holland             3.562500
## Hong Kong           3.801825
## Hungary             3.611111
## India               3.395161
```

```
## Indonesia      4.067460
## Japan          3.981605
## Malaysia       NA
## Mexico         3.730000
## Myanmar        3.946429
## Nepal          3.553571
## Netherlands    2.483333
## Nigeria        1.500000
## Pakistan       3.000000
## Philippines    3.329787
## Poland         3.625000
## Sarawak        4.333333
## Singapore      4.126147
## South Korea     NA
## Sweden         3.250000
## Taiwan         3.665402
## Thailand       3.384817
## UK             2.997101
## United States  3.750000
## USA            3.457043
## Vietnam        3.187963
```

```
#Style
```

```
ramenRatingStyle <- ramen %>% add_count(Style)
```

```
colnames(ramenRatingStyle)[8] <- ("Total")
```

```
ramenRatingStyle2 <- group_by(ramenRatingStyle, Style) %>% mutate(AvgStar = round(sum(Stars, na.rm = TRUE)/n(), 2))
```

```
ramenRatingStyle3 <- ramenRatingStyle2[!duplicated(ramenRatingStyle2[c("Style")])], c("Style", "Total", "AvgStar")
```

```
ramenRatingStyle3
```

```
## # A tibble: 8 x 3
## # Groups:   Style [8]
##   Style Total AvgStar
##   <fct> <int>   <dbl>
## 1 Cup    450    3.50
## 2 Pack  1531    3.69
## 3 Tray   108    3.55
## 4 Bowl   481    3.67
## 5 Box     6    4.29
## 6 Can     1    3.5
## 7 Bar     1    5
## 8 ""      2    3.38
```

```
# Salt Content
```

```
mean(firstq[["Stars"]], na.rm = TRUE)
```

```
## [1] 3.117691
```

```
mean(secondq[["Stars"]], na.rm = TRUE)
```

```
## [1] 3.915969
```

```

mean(thirdq[["Stars"]], na.rm = TRUE)

## [1] 3.824728
mean(fourthq[["Stars"]], na.rm = TRUE)

## [1] 3.759031
#Variety

ramenStarSpicy <- ramen[str_detect(tolower(ramen$Variety),"spicy"), c("Variety", "Stars")]
ramenStarMild <- ramen[str_detect(tolower(ramen$Variety),"mild"), c("Variety", "Stars")]
ramenStarNeither <- ramen[!str_detect(tolower(ramen$Variety),"spicy") & !str_detect(tolower(ramen$Variety),"mild"), c("Variety", "Stars")]

(round(sum(ramenStarSpicy$Stars) / length(ramenStarSpicy$Variety) , digits = 6))

## [1] 3.688426
(round(sum(ramenStarMild$Stars) / length(ramenStarMild$Variety) , digits = 6))

## [1] 3.583333
(round(sum(ramenStarNeither$Stars, na.rm = TRUE) / length(ramenStarNeither$Variety) , digits = 6))

## [1] 3.646785

```

1. There are 353 values
  2. 2016, 2015, 2013, 2014 & 2012
  3. 49 brands are from the United States
  4. prima taste & mykuali both got the #1 spot more than once
  5. There are 24 brands that tie with an average score of 5
  6. Yes there is a correlation between country and how it is packaged
  - 7.
  - A) Based on the results I would expect that Singapore would show in the top ten more often
  - B) This is confirmed by the results
  8. Japan makes the most ramen with 352 ramens and Brazil makes the best ramen. The country that makes the most and the country that makes the best are different
  9. There are 270 brands that are described as spicy, those that are spicy have a higher average rating
  10. Overall the rating goes up with the higher amounts of salt but there is some decrease in ratings do start to decrease slightly when you get into the higher salt contents quartile 3 and 4
  - 11.
- Country: The average salt per product ranges for the high in the US at 21.24% and the low at 15.83 in Nigeria but it does appear there can be a significant difference based on the country
- Styles: Looking at the average salt content between the different styles there is only a small variance between each style, there does not appear to be a correlation
- Brand: There is a significant difference in the amount of salt contained in the different brands, ranging from 22.61 to 4.42

Variety: I choose three different types of varieties brands that are listed as Spicy, Mild and brands that have neither in their title

12. For question 12 we broke down the varieties into 5 different sections, ramen with star ratings in the 1st quartile, 2nd quartile, 3rd quartile, 4th quartile and lastly those that ranked in the top ten. As an additional note we included those that had N/A ratings in the first quartile to make sure they were represented.
13. Looking at the different variables provided, Roamin' Ramen would produce the highest average rating by:
  - a. Producing the ramen in Brazil, this country has the highest average rating
  - b. Making the Ramen spicy, Ramen with Spicy in the title has a higher average rating than those with mild or with neither in the title
  - c. The style should be either a bar, box or pack. Bar and box having the highest rating but having a small enough sample size this may be an outlier, pack having the highest rating with large sample size
  - d. The salt content should be in the second quartile, greater than 18.372% and less than or equal to 19.340% varieties with these salt contents have the highest average rating