Case Study 1

Chase Darlington

October 3, 2018

```
#for MAC
#load("~/Downloads/ramen.Rdata")

#for PC
load("C:\\Users\\Chase Darlington\\Downloads\\ramen.RData")

#ramen
#names(ramen)
```

1) (3 pts) How many different brands are reviewed in the data set? Do you recognize any of them?

355 brands are reviewed in the data set.

```
#rawbrands <- ramen$Brand
#brands <- data.frame(rawbrands[!duplicated(rawbrands)])
#brands
#nrow(brands)
#nrow(ramen$Brand[!duplicated(ramen$Brand)])

#or

length(ramen$Brand[!duplicated(ramen$Brand)])
```

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

2) (3 pts) What years do we have "top ten" data from? (Hint: This will require turning the Top Ten column into a string column and using some of our string functions)

We have top ten information from 2012-2016

```
ramen$Top.Ten[ramen$Top.Ten==""] <- NA
Top.Ten <- na.omit(ramen)
#Top.Ten

seperatedTop.Ten <- data.frame(do.call('rbind', strsplit(as.character(Top.Ten$Top.Ten), " #", fixed=TRU
#seperatedTop.Ten
levels(seperatedTop.Ten$X1)
```

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

3) (3 pts) Which ramen brands are from the United States (hint: beware of abbreviations!)?

Refer below for the 49 brands available in the United States

```
#levels(ramen$Country)
sort(ramen$Country[!duplicated(ramen$Country)])
```

```
##  [1] Australia    Bangladesh    Brazil      Cambodia      Canada
##  [6] China        Colombia      Dubai       Estonia       Fiji
## [11] Finland      Germany       Ghana       Holland       Hong Kong
## [16] Hungary      India         Indonesia   Japan         Malaysia
## [21] Mexico       Myanmar       Nepal       Netherlands   Nigeria
```

```
## [26] Pakistan        Philippines    Poland         Sarawak        Singapore
## [31] South Korea  Sweden         Taiwan         Thailand       UK
## [36] United States USA            Vietnam
## 38 Levels: Australia Bangladesh Brazil Cambodia Canada China ... Vietnam
```

```r
#Note: there exists a USA and United States
#So, reassign values
ramen$Country[ramen$Country=="United States"] <- "USA"
#Check 3. For some reason USA is still listed as a level...
#levels(ramen$Country)
#...so, Check 2
ramen[ramen$Country=="United States"]
```

```
## data frame with 0 columns and 2580 rows
```

```r
ramen$Country[ramen$Country=="United States"]
```

```
## factor(0)
## 38 Levels: Australia Bangladesh Brazil Cambodia Canada China ... Vietnam
```

```r
#reprint all the countries; no duplicates
#sort(ramen$Country[!duplicated(ramen$Country)])
#levels(ramen$Country)




#Then, Dubai isn't a country, nor Sarawak: correct to UAE, and Malaysia; Hong Kong is a special adminis
ramen$Country[ramen$Country=="Dubai"] <- "UAE"
```

```
## Warning in `[<-.factor`(`*tmp*`, ramen$Country == "Dubai", value =
## structure(c(19L, : invalid factor level, NA generated
```

```r
ramen$Country[ramen$Country=="Sarawak"] <- "Malaysia"
#sort(ramen$Country[!duplicated(ramen$Country)])
```

```r
#notably, all United States entries were converted to USA
USbrands <- (ramen[ramen$Country=="USA",])
USbrands <- (USbrands[!duplicated(USbrands$Brand),])
length(USbrands$Brand)
```

```
## [1] 50
```

4) (3 pts) Has any brand of ramen won the #1 slot in the Top Ten Ramens list more than once? Which brand(s)?

MyKuali ranked #1 2014 to 2015, 2 years in a row

```r
#reapply NA,just in case
ramen$Top.Ten[ramen$Top.Ten==""] <- NA
Top.Ten <- na.omit(ramen)
#Top.Ten

seperatedTop.Ten <- data.frame(do.call('rbind', strsplit(as.character(Top.Ten$Top.Ten), " #", fixed=TRU
class(seperatedTop.Ten)
```

```
## [1] "data.frame"
```

```r
colnames(seperatedTop.Ten) <- c("Year","Rank")
#seperatedTop.Ten
Top.Ten <- cbind(Top.Ten, seperatedTop.Ten)
```

```
Top.Ten <- Top.Ten[order((Top.Ten$Year)),]
Top.Ten <- Top.Ten[order((Top.Ten$Rank)),]
Top.Ten <- Top.Ten[order(stringr::str_length(Top.Ten$Rank)),]
#Top.Ten

Top.Ten[Top.Ten$Rank==1,]
```

```
##          Brand                          Variety Style   Country Stars
## 2476     Indomie          Special Fried Curly Noodle  Pack Indonesia     5
## 1589 Prima Taste             Singapore Laksa La Mian  Pack Singapore     5
## 1494     MyKuali         Penang White Curry Noodle  Pack  Malaysia     5
## 1081     MyKuali    Penang Red Tom Yum Goong Noodle  Pack  Malaysia     5
## 634  Prima Taste Singapore Laksa Wholegrain La Mian  Pack Singapore     5
##       Top.Ten perc_salt Year Rank
## 2476 2012 #1   19.48722 2012    1
## 1589 2013 #1   18.36248 2013    1
## 1494 2014 #1   18.48394 2014    1
## 1081 2015 #1   18.93488 2015    1
## 634  2016 #1   16.84561 2016    1
```
```
#MyKuali ranked #1 2014 to 2015
```

5) (3 pts) Which brand has the highest average star rating? If there is a tie, report multiple brands.

24 brands tie in at 5 stars (ChoripDong, Daddy, Daifuku...). Refer to the data frame below.

```
AvgStarRatings <- aggregate( ramen$Stars ~ ramen$Brand, ramen, mean )
AvgStarRatings[AvgStarRatings$`ramen$Stars`==5.0,]
```

```
##               ramen$Brand ramen$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
## 138   Komforte Chockolates           5
## 186                  MyOri           5
## 200              Nyor Nyar           5
## 207             ORee Garden           5
## 216              Patanjali           5
## 218                 Peyang           5
## 220          Plats Du Chef           5
## 227                  Prima           5
## 228            Prima Taste           5
## 262              Seven & I           5
## 273               Song Hak           5
## 289               Takamori           5
## 291           Tao Kae Noi           5
## 302             The Bridge           5
## 304 The Ramen Rater Select           5
## 311                Torishi           5
```

6) (5 pts) Is the way a ramen is packaged related to which country it is from?

There is no correlation between packaging style and a country's level of cultural diversity or development, as observed in the plot below. But, it is obvious that pack, cup, and bowl are widely pertinent in all countries (in descending order of significance). Notably, Canada and Australia use less packs than cups.

```r
#names(ramen)
#head(ramen)
#Style indicates packaging
#data.frame(sort(unique(ramen$Style)))
#Bar, bowl, box, can, cup, pack, tray

#is packaging correlated to country of origin?
#install.packages("ggpubr")
library("ggpubr")
```

```
## Loading required package: ggplot2
```

```
## Loading required package: magrittr
```

```r
#ggscatter(ramen, x = "Style", y = "Country",
          #add = "reg.line", conf.int = TRUE,
          #cor.coef = TRUE, cor.method = "pearson",
          #xlab = "Country of Origin", ylab = "Packaging Style")

#another way, too

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

```
##
##         Australia Bangladesh Brazil Cambodia Canada China Colombia Dubai
##                 0          0      0        0      0     1        0     0
##   Bar           0          0      0        0      0     0        0     0
##   Bowl          0          0      0        0      8    45        0     0
##   Box           0          0      0        0      0     0        0     0
##   Can           0          0      0        0      0     0        0     0
##   Cup          17          0      2        0     17    16        3     0
##   Pack          5          7      3        5     16    98        3     0
##   Tray          0          0      0        0      0     9        0     0
##
##         Estonia Fiji Finland Germany Ghana Holland Hong Kong Hungary India
##               0    0       0       0     0       0         0       0     0
##   Bar         0    0       0       0     0       0         0       0     0
##   Bowl        0    0       0       0     0       0        30       0     0
##   Box         0    0       0       0     0       0         0       0     0
##   Can         0    0       0       0     0       0         0       0     0
##   Cup         0    0       0      11     0       0        38       0     3
##   Pack        2    4       3      16     2       4        67       9    28
##   Tray        0    0       0       0     0       0         2       0     0
##
##         Indonesia Japan Malaysia Mexico Myanmar Nepal Netherlands Nigeria
##                 0     0        0      0       0     0           0       0
##   Bar           0     0        0      0       0     0           0       0
##   Bowl          0   126        8      0       0     0           0       0
##   Box           1     2        2      0       0     0           0       0
##   Can           0     0        0      0       0     0           0       0
##   Cup          21    49       21     15       3     0           3       0
##   Pack        104   155      128     10      11    14          12       1
```

```
##     Tray            0      20         0        0        0        0              0         0
## 
##            Pakistan Philippines Poland Sarawak Singapore South Korea Sweden
##                   0           0      0       0         0           0      0
##     Bar           0           0      0       0         0           0      0
##     Bowl          0          10      0       0        13          68      0
##     Box           0           0      0       0         0           0      0
##     Can           0           0      0       0         0           0      0
##     Cup           0           4      0       0        27          40      0
##     Pack          9          33      4       0        69         183      3
##     Tray          0           0      0       0         0          18      0
## 
##            Taiwan Thailand  UK United States USA Vietnam
##                 1        0   0             0   0       0
##     Bar         0        0   0             0   1       0
##     Bowl       37       44   2             0  70      20
##     Box         0        0   0             0   1       0
##     Can         0        0   0             0   1       0
##     Cup         2       48  32             0  70       8
##     Pack      181       97  35             0 129      78
##     Tray        3        2   0             0  52       2
```

7) (5 pts) What is the average, maximum, and minimum rating for ramens from the US? What about for ramens from Singapore? Based on your findings, which country do you expect to land in the Top Ten more often? Check your prediction.

```r
mean(ramen[ramen$Country == "USA", "Stars"],  na.rm = T )
```

```
## [1] 3.457948
```

```r
max(ramen[ramen$Country == "USA", "Stars"],  na.rm = T )
```

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

```r
min(ramen[ramen$Country == "USA", "Stars"],  na.rm = T )
```

```
## [1] 0
```

```r
mean(ramen[ramen$Country == "Singapore", "Stars"],  na.rm = T )
```

```
## [1] 4.126147
```

```r
max(ramen[ramen$Country == "Singapore", "Stars"],  na.rm = T )
```

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

```r
min(ramen[ramen$Country == "Singapore", "Stars"],  na.rm = T )
```

```
## [1] 2
```

```r
###Based on my findings Singapore would appear in Top Ten rankings more often because their ramen's mea
```

```r
(ramen[ramen$Country == "Singapore", "Top.Ten"])
```

```
##    [1] <NA>      <NA>      <NA>      <NA>      <NA>      <NA>      <NA>
##    [8] <NA>      <NA>      <NA>      <NA>      <NA>      <NA>      <NA>
##   [15] <NA>      <NA>      <NA>      <NA>      <NA>      <NA>      <NA>
##   [22] <NA>      <NA>      <NA>      <NA>      <NA>      <NA>      <NA>
##   [29] <NA>      <NA>      <NA>      <NA>      <NA>      <NA>      <NA>
##   [36] <NA>      <NA>      <NA>      <NA>      <NA>      2016 #1   2016 #8
```

```
## [43] <NA>     2016 #5  <NA>     <NA>     <NA>     <NA>     <NA>
## [50] <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [57] <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [64] <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [71] <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [78] <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [85] <NA>     <NA>     <NA>     <NA>     <NA>     2014 #8  <NA>
## [92] <NA>     2013 #1  2013 #2  <NA>     <NA>     <NA>     <NA>
## [99] 2012 #10 <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [106] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## 38 Levels:  2012 #1 2012 #10 2012 #2 2012 #3 2012 #4 2012 #5 ... 2016 #9
```

```
(ramen[ramen$Country == "USA", "Top.Ten"])
```

```
##   [1] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##   [9] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [17] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [25] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [33] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [41] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [49] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [57] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [65] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [73] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [81] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [89] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
##  [97] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [105] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [113] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [121] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [129] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [137] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [145] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [153] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [161] <NA>    <NA>     <NA>     <NA>     <NA>     2013 #4  <NA>     <NA>
## [169] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [177] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [185] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [193] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [201] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [209] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [217] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [225] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [233] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [241] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [249] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [257] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [265] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [273] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [281] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [289] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [297] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [305] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [313] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
## [321] <NA>    <NA>     <NA>     <NA>     <NA>     <NA>     <NA>
```

```
## 38 Levels:  2012 #1 2012 #10 2012 #2 2012 #3 2012 #4 2012 #5 ... 2016 #9
```
###My prediction is true. Singapore's ramen appears in the top ten 7 times while USA's appears only onc

8) (5 pts) Which country makes the most ramen? Which country makes the best ramen? How did you define "best"? Why? Was the country that makes the best ramen also the country that makes the most ramen?

```r
sort(table(ramen$Country), decreasing=TRUE)
```

```
## 
##         Japan        USA   South Korea      Taiwan      Thailand
##           352        324           309         224           191
##         China     Malaysia    Hong Kong   Indonesia     Singapore
##           169        159           137         126           109
##       Vietnam         UK   Philippines      Canada         India
##           108         69            47          41            31
##       Germany     Mexico     Australia  Netherlands       Myanmar
##            27         25            22          15            14
##         Nepal     Hungary      Pakistan  Bangladesh      Colombia
##            14          9             9           7             6
##        Brazil    Cambodia          Fiji      Holland        Poland
##             5          5             4           4             4
##       Finland     Sweden       Estonia        Ghana       Nigeria
##             3          3             2           2             1
##         Dubai     Sarawak United States
##             0          0             0
```

#Japan make the most ramen

```r
w <- table(ramen$Country, ramen$Stars==5)
#w
sort(w[,2], decreasing = TRUE)
```

```
##         Japan     Malaysia        Taiwan  South Korea     Singapore
##            74          55            48          44            36
##           USA    Indonesia     Hong Kong      Thailand         China
##            31          23            22          19            12
##       Myanmar         UK       Cambodia        Canada         India
##             4          4             2           2             2
##   Philippines    Australia        Brazil       Germany        Mexico
##             2          1             1           1             1
##         Nepal      Vietnam    Bangladesh      Colombia         Dubai
##             1          1             0           0             0
##       Estonia         Fiji       Finland         Ghana       Holland
##             0          0             0           0             0
##       Hungary  Netherlands       Nigeria      Pakistan        Poland
##             0          0             0           0             0
##       Sarawak     Sweden  United States
##             0          0             0
```

#Japan has the most 5 start rating
In conclusion Japan make the most ramen and the best ramen

9) (5 pts) How many ramens are considered spicy as part of their variety (hint, don't forget that "spicy"" and "Spicy"" are both spicy)? Are spicy ramens typically rated higher than non-spicy ramens?

```r
library(stringr)
sum(str_count(ramen$Variety, "Spicy"))+
  sum(str_count(ramen$Variety, "spicy"))
```
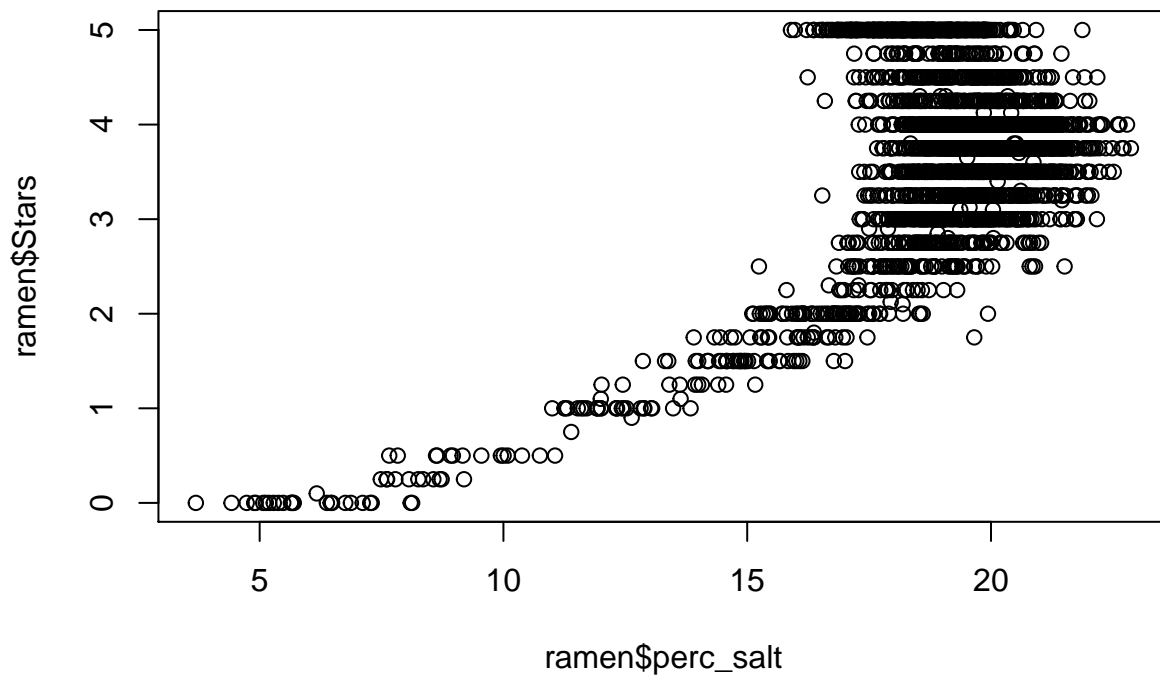
```
## [1] 277
```

#277

10) (5 pts) Based on this data set, what effect does saltiness have on ramen ratings?

Higher percentage salt content is significantly correlated with ramen ratings. The plot below illustrates and the correlation test confirms; note the low p-value and >0.5 correlation coefficient.

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



```r
cor.test(ramen$perc_salt, ramen$Stars)
```

```
##
##  Pearson's product-moment correlation
##
## data:  ramen$perc_salt and ramen$Stars
## t = 35.014, df = 2575, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5411804 0.5935225
## sample estimates:
##       cor
## 0.5679254
```

11) (5 pts) Are there certain styles, countries, brands, or flavors of ramen that tend to use more or less
    salt? Feel free to be creative in how you parse "variety" into flavors.

```
mean(ramen$perc_salt)
```

```
## [1] 18.95061
```

```
SaltRatingsbyStyle <- aggregate(ramen$perc_salt ~ ramen$Style, ramen, mean )
SaltRatingsbyStyle[SaltRatingsbyStyle$`ramen$perc_salt`>19,]
```

```
##   ramen$Style ramen$perc_salt
## 1                    19.14078
## 2         Bar        19.09752
## 7        Pack        19.00416
```

```
#The countries that use salt more than the mean
SaltRatingsbyStyle[SaltRatingsbyStyle$`ramen$perc_salt`<19,]
```

```
##   ramen$Style ramen$perc_salt
## 3        Bowl        18.88129
## 4         Box        18.11725
## 5         Can        17.29685
## 6         Cup        18.90007
## 8        Tray        18.76751
```

```
#The countries that use salt less than the mean

SaltRatingsbyCountry <- aggregate( ramen$perc_salt ~ ramen$Country, ramen, mean )

SaltRatingsbyCountry[SaltRatingsbyCountry$`ramen$perc_salt`>19,]
```

```
##     ramen$Country ramen$perc_salt
## 2      Bangladesh        19.91238
## 3          Brazil        20.23857
## 4        Cambodia        19.12907
## 7        Colombia        19.42893
## 8         Estonia        20.10291
## 9            Fiji        19.38411
## 10        Finland        19.94231
## 11        Germany        19.64901
## 12          Ghana        19.48006
## 13        Holland        20.24594
## 14      Hong Kong        19.35348
## 15        Hungary        19.69600
## 16          India        19.07243
## 17      Indonesia        19.30272
## 18          Japan        19.16634
## 19       Malaysia        19.05508
## 20         Mexico        20.00075
## 21        Myanmar        19.23389
## 22          Nepal        19.76184
## 27         Poland        19.72265
## 28      Singapore        19.41049
## 29    South Korea        19.13972
## 30         Sweden        19.10296
```

```r
#The countries that use salt more than the mean
SaltRatingsbyCountry[SaltRatingsbyCountry$`ramen$perc_salt`<19,]
```

```
##     ramen$Country ramen$perc_salt
## 1       Australia        18.83627
## 5          Canada        16.17303
## 6           China        18.61621
## 23    Netherlands        16.83779
## 24        Nigeria        15.83863
## 25       Pakistan        18.74806
## 26     Philippines       18.98782
## 31         Taiwan        18.61788
## 32       Thailand        18.86414
## 33             UK        18.25107
## 34            USA        18.79352
## 35        Vietnam        18.87525
```

```r
#The countries that use salt less than the mean

SaltRatingsbyBrand <- aggregate( ramen$perc_salt ~ ramen$Brand, ramen, mean )
head(SaltRatingsbyBrand[SaltRatingsbyBrand$`ramen$perc_salt`>19,])
```

```
##        ramen$Brand ramen$perc_salt
## 1    1 To 3 Noodles        20.33477
## 2          7 Select        19.78582
## 5   A-Sha Dry Noodle       19.08609
## 7               ABC        19.62949
## 9             Adabi        19.85185
## 11         Ajinatori        20.26037
```

```r
#The Brands that use salt more than the mean
head(SaltRatingsbyBrand[SaltRatingsbyBrand$`ramen$perc_salt`<19,])
```

```
##            ramen$Brand ramen$perc_salt
## 3        7 Select/Nissin        18.43610
## 4                  A-One        18.06370
## 6                     A1        18.07266
## 8                 Acecook        18.48431
## 10                Ah Lai        18.81920
## 18 Authentically Asian        11.91938
```

```r
#The Brands that use salt more than the mean


#unique(ramen$Variety)
#install.packages("ggpubr")
#library(ggpubr)
ramen$flavor<-NA
#ramen
ramen$Variety <- str_to_lower(ramen$Variety)

ramen$flavor[grep(c("curry"), ramen$Variety)] <- "curry"
ramen$flavor[grep(c("tomato"), ramen$Variety)] <- "tomato"
ramen$flavor[grep(c("sesame"), ramen$Variety)] <- "sesame"
ramen$flavor[grep(c("toona"), ramen$Variety)] <- "toona"
ramen$flavor[grep(c("udon"), ramen$Variety)] <- "udon"
```

```
ramen$flavor[grep(c("kimchi"), ramen$Variety)] <- "kimchi"
ramen$flavor[grep(c("beef"), ramen$Variety)] <- "beef"
ramen$flavor[grep(c("chicken"), ramen$Variety)] <- "chicken"
ramen$flavor[grep(c("shoyu"), ramen$Variety)] <- "shoyu"
ramen$flavor[grep(c("miso"), ramen$Variety)] <- "miso"
ramen$flavor[grep(c("laksa"), ramen$Variety)] <- "laksa"
ramen$flavor[grep(c("mushroom"), ramen$Variety)] <- "mushroom"
ramen$flavor[grep(c("tempura"), ramen$Variety)] <- "tempura"
ramen$flavor[grep(c("yakisoba"), ramen$Variety)] <- "yakisoba"
ramen$flavor[grep(c("seafood", "prawn", "prawns", "shrimp", "shrimps"), ramen$Variety)] <- "seafood"
```

```
## Warning in grep(c("seafood", "prawn", "prawns", "shrimp", "shrimps"), ramen
## $Variety): argument 'pattern' has length > 1 and only the first element
## will be used
```

```
#ramen

SaltRatingsbyFlavor <-aggregate( ramen$perc_salt ~ ramen$flavor, ramen, mean )
SaltRatingsbyFlavor[SaltRatingsbyFlavor$`ramen$perc_salt`>19,]
```

```
##     ramen$flavor ramen$perc_salt
## 1          beef        19.13457
## 4        kimchi        19.44451
## 5         laksa        19.14599
## 8       seafood        19.34624
## 9        sesame        19.27228
## 11      tempura        19.54899
## 12       tomato        19.09747
## 14         udon        19.52899
## 15     yakisoba        19.16808
```

```
#The Flavor that use salt more than the mean
SaltRatingsbyFlavor[SaltRatingsbyFlavor$`ramen$perc_salt`<19,]
```

```
##     ramen$flavor ramen$perc_salt
## 2       chicken        18.95588
## 3         curry        18.91829
## 6          miso        18.93953
## 7      mushroom        17.78168
## 10        shoyu        18.73322
## 13        toona        18.62559
```

```
#The Flavor that use salt less than the mean
```

```
mean(ramen$perc_salt)
```

```
## [1] 18.95061
```

```
AvgSaltRatings <- aggregate( ramen$perc_salt ~ ramen$Country, ramen, mean )

#there are certain ramen from certain country, brand that uses more salt than other.


#plot(ramen$perc_salt, ramen$Country)
#plot(data.frame(ramen$Country, ramen$perc_salt))
#plot(data.frame(ramen$Brand, ramen$perc_salt))
```

12) [Open Ended] (10 pts) You have been hired by a new ramen start up, Roamin' Ramen. They want to better understand the ramen market and ask you to break down ramen into 5 collections of "similar" ramens. How do you go about it? Discuss what you find.

Key varieties: curry, tomato, sesame, toona, udon, kimchi, beef, chicken, seafood/prawns/shrimps, tempura, mushroom, shoyu, laksa, yakisoba Main 5 Collections: Chicken, Beef, Seafood, Curry, Udon

There are 5 main kinds of ramen. The japanese would normally say: Shoyu, Shio, Miso, Curry. But, globally, the distinct kinds are actually chicken, beef, seafood, curry, and udon. The code below parses varieties for distinct flavors and counts them. Refer below for the actual results.

```
#unique(ramen$Variety)

sort(table(ramen$flavor), decreasing=TRUE)
```

```
##
##  chicken     beef  seafood    curry     udon mushroom yakisoba   sesame
##      312      227      110       96       56       52       50       47
##     miso   kimchi    shoyu    laksa   tomato  tempura    toona
##       41       40       38       34       25       12        2
#1 - chicken, 2 - beef, 3 - seafood, 4 - curry, 5 - udon
```

13) [Open Ended] (10 pts) Roamin' Ramen wants to make the next Top Ten Ramen winner. What recommendations do you have for them based on this data set? Be sure to use the columns provided, and generate your own ideas by examining the Variety column. Support your assertions with statistics (like mean, min, max or others you prefer) and at least one plot.

Recommendations: Pack –> Laksa –> Low sodium; Learn from the Japanese and Singaporean ramen houses, relocate to those countries, and/or partner therein

```
ramen$Top.Ten[ramen$Top.Ten==""] <- NA
Top.Ten <- na.omit(ramen)
addtop <- data.frame(do.call('rbind', strsplit(as.character(Top.Ten$Top.Ten), " #", fixed=TRUE)))
colnames(addtop)<-c("Year", "Rank")
Top.Ten <- cbind(Top.Ten, addtop)
Top.Ten$Rank<-as.double(Top.Ten$Rank)
#sort((Top.Ten$Rank), decreasing=FALSE)
#Top.Ten[order(as.numeric(Top.Ten$Rank), decreasing=FALSE),]
#Top.Ten

tab1 <- table(Top.Ten$Rank, Top.Ten$Style)
tab1 <- rbind(tab1, Total = colSums(tab1))
tab1
```

```
##          Bar Bowl Box Can Cup Pack Tray
## 1         0    0   0   0   0    3    0
## 2         0    0   0   0   0    1    0
## 3         0    0   0   0   0    2    0
## 4         0    0   0   0   0    0    1
## 5         0    0   0   0   0    1    1
## 6         0    0   0   0   0    1    0
## 7         0    0   0   0   0    0    1
## 8         0    0   0   0   0    2    0
## 9         0    0   0   0   0    2    0
## 10        0    0   0   0   0    2    1
## Total 0    0    0   0   0   0   14    4
```

```
#14 of the recently ranked Top Ten used "Pack" styles, while there were no "Cup" styles and only 4 "Tra

tab2 <- table(Top.Ten$Rank, Top.Ten$flavor)
tab2 <- rbind(tab2, Total = colSums(tab2))
tab2
```

```
##        beef chicken curry laksa tomato yakisoba
## 1         0       0     1     2      0        0
## 2         0       0     0     0      1        0
## 3         0       1     1     0      0        0
## 4         0       0     0     0      0        1
## 5         0       0     0     1      0        1
## 6         0       0     1     0      0        0
## 7         0       0     0     0      0        1
## 8         1       0     0     1      0        0
## 9         0       1     1     0      0        0
## 10        0       2     0     0      0        1
## Total     1       4     4     4      1        4
```

```
#Yakisoba, Laksa, Curry, and Chicken are all strong flavors aong the Top.Ten; Laksa has 2 placements fo

tab3 <- table(Top.Ten$Rank, Top.Ten$Country)
tab3 <- rbind(tab3, Total = colSums(tab3))
tab3[11,]
```

```
##     Australia    Bangladesh        Brazil      Cambodia        Canada
##             0             0             0             0             0
##         China      Colombia         Dubai       Estonia          Fiji
##             1             0             0             0             0
##       Finland       Germany         Ghana       Holland     Hong Kong
##             0             0             0             0             1
##       Hungary         India     Indonesia         Japan      Malaysia
##             0             0             1             4             3
##        Mexico       Myanmar         Nepal   Netherlands       Nigeria
##             0             0             0             0             0
##      Pakistan   Philippines        Poland       Sarawak     Singapore
##             0             0             0             0             5
##   South Korea        Sweden        Taiwan      Thailand            UK
##             1             0             1             1             0
## United States           USA       Vietnam
##             0             0             0
```

```
#We would recommend learning from, associating with, or partnering with a Japanese or Singaporean ramen

plot(Top.Ten$Rank, Top.Ten$perc_salt, abline(lm(Top.Ten$perc_salt ~ Top.Ten$Rank)))
```
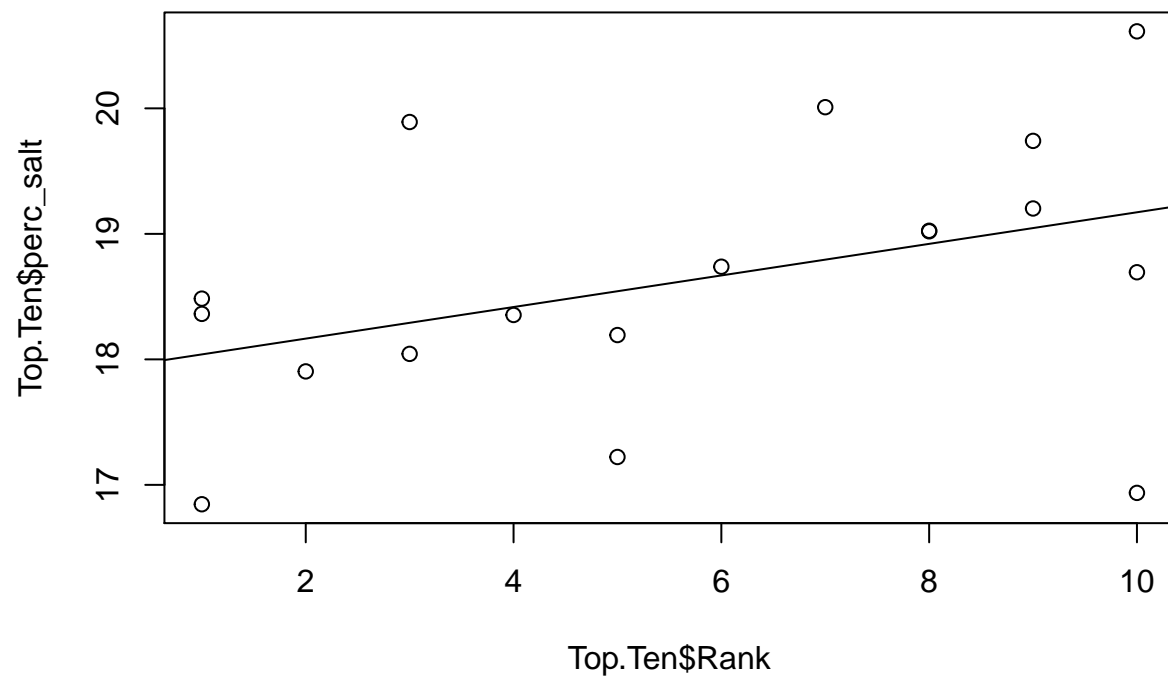
```
cor(Top.Ten$perc_salt, Top.Ten$Rank)
```

```
## [1] 0.4003134
```

```
#Finally, saltiness will have a 40% impact on the rating, and there is positive correlation between hig
```