R Notebook

Code ▾

2/22/2020

Hide

```
ramrate <- read.csv("ramen-ratings-1.csv")
str(ramrate)
```

```
'data.frame':   2580 obs. of  6 variables:
 $ Review : int  2580 2579 2578 2577 2576 2575 2574 2573 2572 2571 ...
 $ Brand  : Factor w/ 355 levels "1 To 3 Noodles",..: 193 122 195 339 38 252 8 110 236 137 ...
 $ Variety: Factor w/ 2413 levels "\"A\" Series Artificial Chicken",..: 2195 1448 458 723 1957 1
112 2058 1375 816 2232 ...
 $ Style  : Factor w/ 8 levels "","Bar","Bowl",..: 6 7 6 7 7 7 6 8 7 7 ...
 $ Country: Factor w/ 38 levels "Australia","Bangladesh",..: 19 33 37 33 17 31 19 19 19 30 ...
 $ Stars  : Factor w/ 43 levels "0","0.1","0.25",..: 34 7 16 19 34 41 36 34 3 18 ...
```

Hide

```
ramrate$Review <- as.character(ramrate$Review)
ramrate$Brand <- as.character(ramrate$Brand)
ramrate$Variety <- as.character(ramrate$Variety)

ramrate$Style <- as.factor(ramrate$Style)
ramrate$Country <- as.factor(ramrate$Country)

ramrate$Stars <- as.numeric(as.character(ramrate$Stars))
```

```
NAs introduced by coercion
```

Hide

```
ramrate <- subset(ramrate,!is.na(ramrate$Stars)) #removing "NA" values from column Stars.


colnames(ramrate) <- c("Review", "Brand", "Flavor", "Style", "Country","Stars")

ramrate
```

	Revi... <chr>	Brand <chr>	Flavor <chr>
1	2580	New Touch	T's Restaurant Tantanmen
2	2579	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles
3	2578	Nissin	Cup Noodles Chicken Vegetable
4	2577	Wei Lih	GGE Ramen Snack Tomato Flavor

	Revi... <chr>	Brand <chr>	Flavor <chr>
5	2576	Ching's Secret	Singapore Curry
6	2575	Samyang Foods	Kimchi song Song Ramen
7	2574	Acecook	Spice Deli Tantan Men With Cilantro
8	2573	Ikeda Shoku	Nabeyaki Kitsune Udon
9	2572	Ripe'n'Dry	Hokkaido Soy Sauce Ramen
10	2571	KOKA	The Original Spicy Stir-Fried Noodles

1-10 of 2,577 rows | 1-4 of 6 columns   Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
summary(ramrate)
```

```
    Review              Brand               Flavor              Style              Country
 Length:2577         Length:2577         Length:2577         Pack   :1528      Japan      : 352
 Class :character    Class :character    Class :character    Bowl   : 481      USA        : 323
 Mode  :character    Mode  :character    Mode  :character    Cup    : 450      South Korea: 307
                                                             Tray   : 108      Taiwan     : 224
                                                             Box    :   6      Thailand   : 191
                                                                    :   2      China      : 169
                                                             (Other):   2      (Other)    :1011

     Stars
 Min.   :0.000
 1st Qu.:3.250
 Median :3.750
 Mean   :3.655
 3rd Qu.:4.250
 Max.   :5.000
```
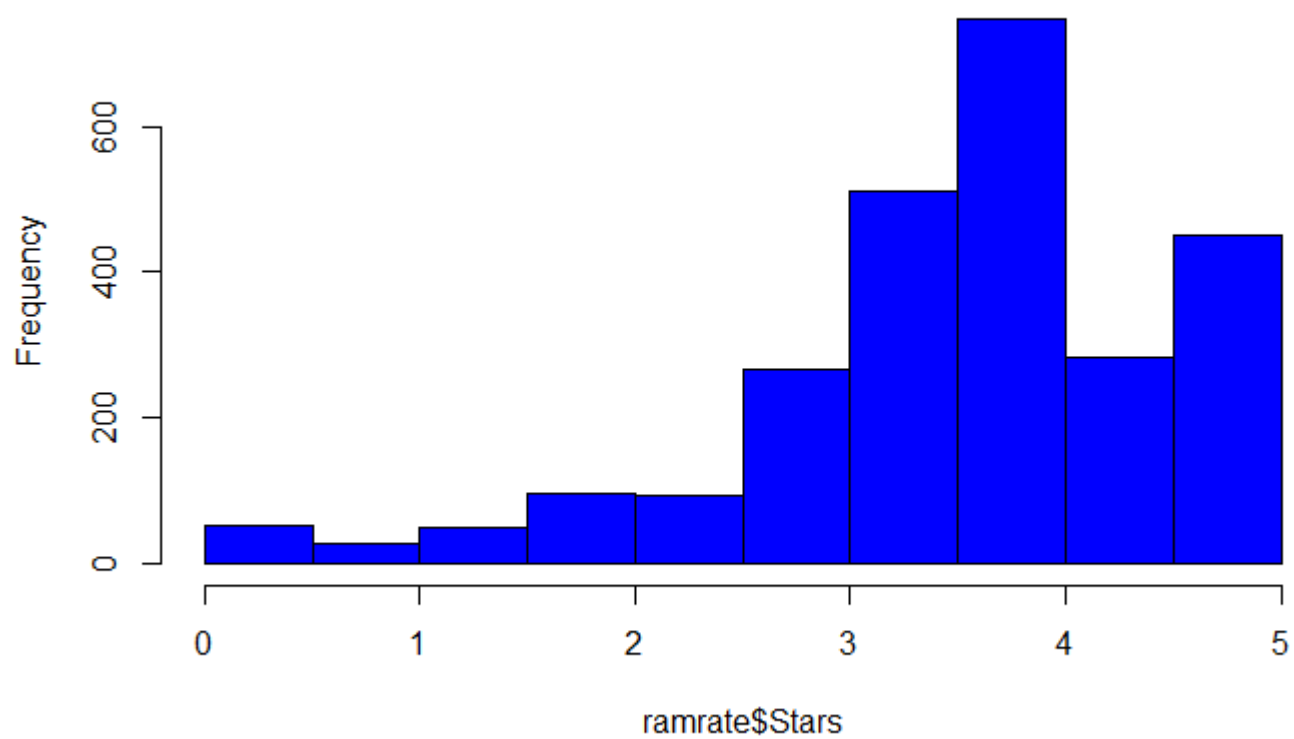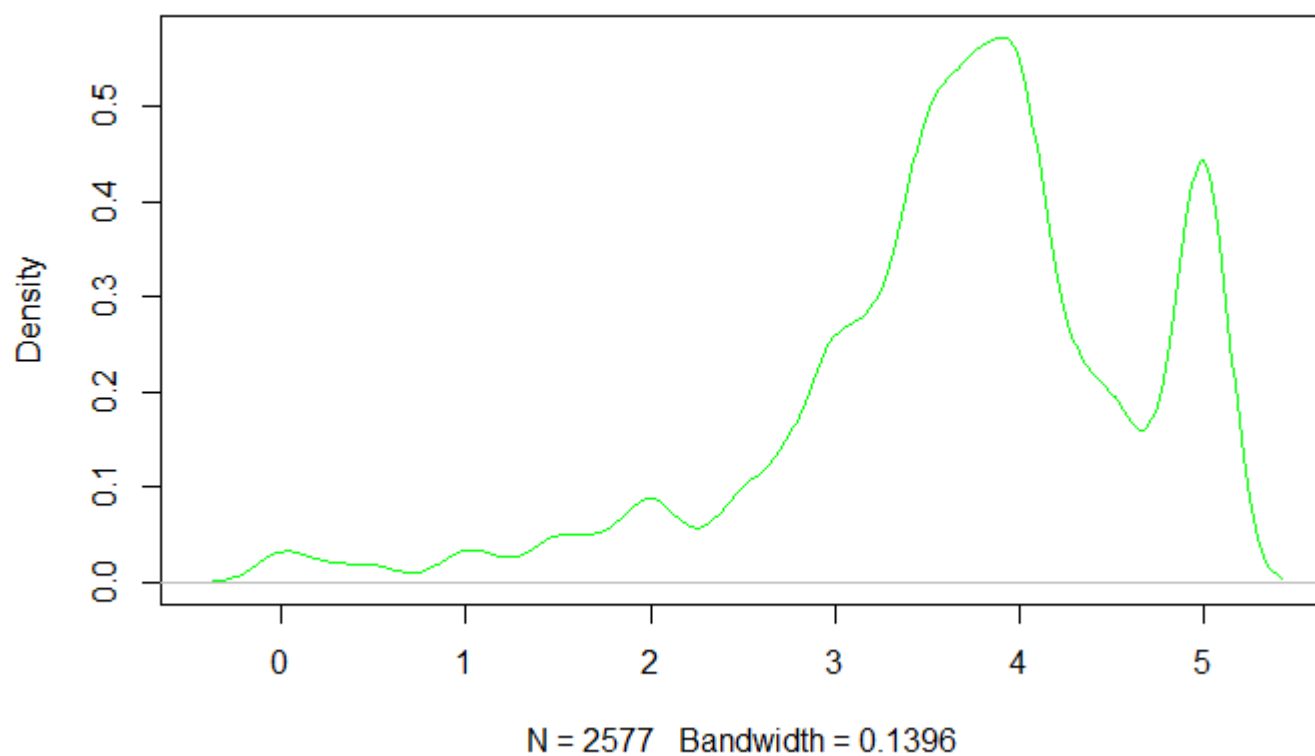
Hide

```
##############################plots:
hist(ramrate$Stars,main = "Histogram of Stars",col = "Blue")
```

Histogram of Stars


```
d <- density(ramrate$Stars)
plot(d, main="Density plot of Stars", col ="Green")
```

Hide

Density plot of Stars

N = 2577   Bandwidth = 0.1396

Hide

```
boxplot(ramrate$Stars, main="Boxplot of Stars", col = "Red")
```

Boxplot of Stars
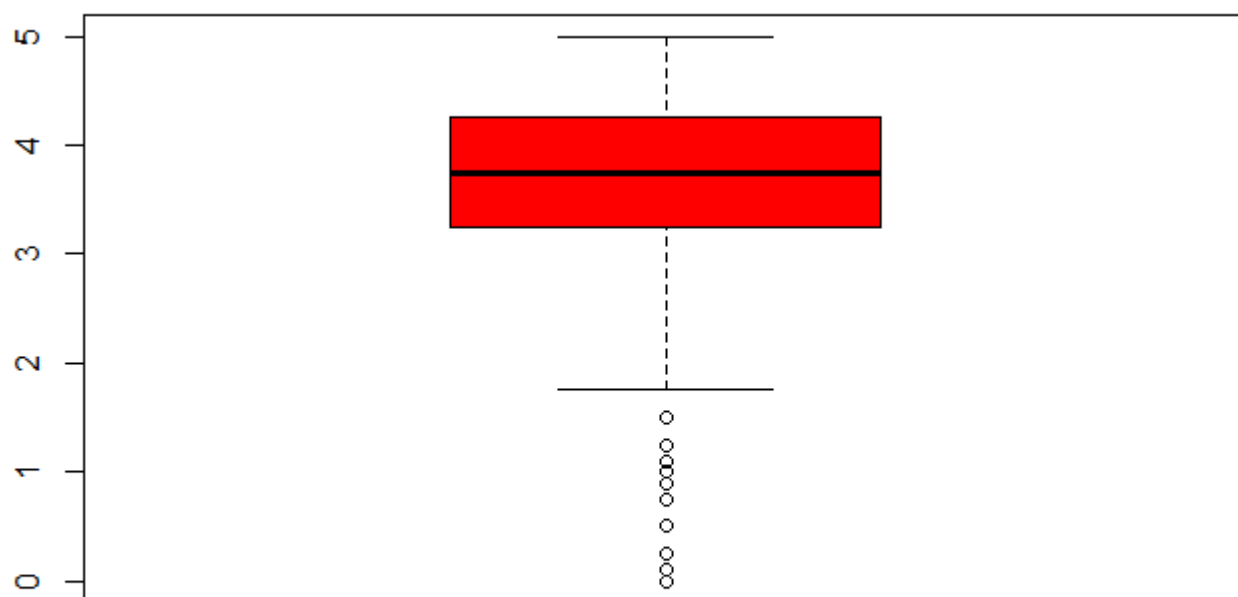
Hide

```
NA
NA
NA
```

*Comments: Looking at the graphs above, I don't believe the Stars data follows a normal distribution.

Hide

```
#function:
ratey <- function() {

        #creating new column and populating with numerical values form Stars:
        Rating <- c(ramrate$Stars)

        #appending to ramrate dataframe:
        ramrate <- data.frame(ramrate$Review, ramrate$Brand, ramrate$Flavor, ramrate$Style,ramr
ate$Country, ramrate$Stars, Rating)

        #renaming columns so there is no "ramrate." in each column name:
        colnames(ramrate) <- c("Review", "Brand", "Flavor", "Style", "Country","Stars","Rating"
)

        #replacing numeric values for factor values:
        ramrate$Rating[which(ramrate$Rating >= 0 & ramrate$Rating <= 1)] <- "Poor"
        ramrate$Rating[which(ramrate$Rating > 1 & ramrate$Rating <= 2)] <- "Below Average"
        ramrate$Rating[which(ramrate$Rating > 2 & ramrate$Rating <= 3)] <- "Average"
        ramrate$Rating[which(ramrate$Rating > 3 & ramrate$Rating <= 4)] <- "Good"
        ramrate$Rating[which(ramrate$Rating > 4 & ramrate$Rating <= 5)] <- "Excellent"

        ramrate$Rating <- as.factor(ramrate$Rating)

        ramrate

}

#setting function to my dataframe:
ramrate <- ratey()
ramrate
```

Revi... <fctr>	Brand <fctr>	Flavor <fctr>
2580	New Touch	T's Restaurant Tantanmen
2579	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles
2578	Nissin	Cup Noodles Chicken Vegetable
2577	Wei Lih	GGE Ramen Snack Tomato Flavor
2576	Ching's Secret	Singapore Curry

Revi… <fctr>	Brand <fctr>	Flavor <fctr>
2575	Samyang Foods	Kimchi song Song Ramen
2574	Acecook	Spice Deli Tantan Men With Cilantro
2573	Ikeda Shoku	Nabeyaki Kitsune Udon
2572	Ripe'n'Dry	Hokkaido Soy Sauce Ramen
2571	KOKA	The Original Spicy Stir-Fried Noodles

1-10 of 2,577 rows | 1-3 of 7 columns   Previous **1** 2 3 4 5 6 … 100 Next

Hide

NA

Hide

```
# Simple Pie Chart

#counters:
poorC = 0
bavgC = 0
avgC = 0
goodC = 0
exclC = 0
#for loop for counting:
for(i in 1:length(ramrate$Rating)) {

  if(ramrate[i,7] == "Poor") {
    poorC <- poorC + 1
  }

  if(ramrate[i,7] == "Below Average") {
    bavgC <- bavgC + 1
  }

  if(ramrate[i,7] == "Average") {
    avgC <- avgC + 1
  }

  if(ramrate[i,7] == "Good") {
    goodC <- goodC + 1
  }

  if(ramrate[i,7] == "Excellent") {
    exclC <- exclC + 1
  }
}

slices <- c(poorC, bavgC, avgC, goodC, exclC)
lbls <- c("Poor", "Below Average", "Average", "Good", "Excellent")
pie(slices, labels = lbls, main="Pie Chart of Ramen Ratings")
```
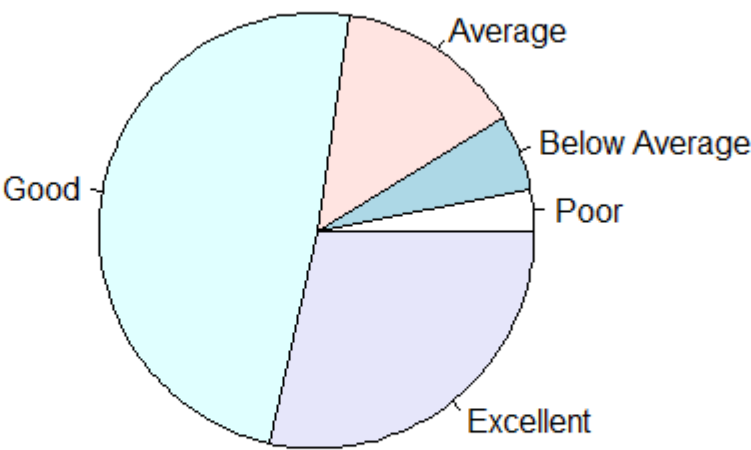
Pie Chart of Ramen Ratings

Hide

```
ramrate[order(ramrate$Country),]
```

	Revi... <fctr>	Brand <fctr>	Flavor <fctr>
230	2349	Fantastic	Noodles Chicken & Corn Flavour
271	2308	Fantastic	Noodles Chicken Chow Mein Flavour
278	2301	Suimin	Noodle With Oriental Chicken Flavour
283	2296	Suimin	Noodles With Chicken Flavour
294	2285	Fantastic	Noodles Beef Flavour
305	2274	Suimin	Noodles With Braised Beef Flavour
326	2253	Singa-Me	Instant Noodles Mi Goreng Flavour
348	2231	Fantastic	Noodle Chicken Flavour
360	2219	Fantastic	Noodles Crispy Bacon Flavour
391	2188	Singa-Me	Instant Noodles Chicken Flavour

1-10 of 2,577 rows | 1-5 of 7 columns   Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
ramrate[order(ramrate$Flavor, decreasing =TRUE),]
```

	Revi… <fctr>	Brand <fctr>	Flavor <fctr>
209	2370	Samyang Foods	Zzaldduck Buldak Bokkeummyun Snack
1635	943	Amino	Zurek
752	1827	Nongshim	Zha Wang ((Jjawang) Noodles With Chajang Sauce
850	1729	Vedan	Zan Vegetarian Instant Noodle Series Shiitake Mushroom Flavour
664	1915	Itsuki	Yuzu Sesame Sauce Udon
147	2432	Fashion Food	Yummy Spicy Rice Soup With Seafood
1551	1027	Samyang Foods	Yukgaejang Beef Ramyun
1177	1401	Myojo	Yomise No Yakisoba Shiodare Flavor With Black Pepper Mayonnaise
161	2418	Myojo	Yomise No Yakisoba Karashi Mentaiko Flavor
592	1987	Nissin	Yokohama Walker Iekei Ramen

1-10 of 2,577 rows | 1-4 of 7 columns   Previous   **1**   2   3   4   5   6   …   100   Next

Hide

```
subset(ramrate, Stars >= 2.5 & Style == "Cup")
```

	Revi… <fctr>	Brand <fctr>	Flavor <fctr>
1	2580	New Touch	T's Restaurant Tantanmen
7	2574	Acecook	Spice Deli Tantan Men With Cilantro
16	2565	KOKA	Mushroom Flavour Instant Noodles
21	2560	Nissin	Cup Noodles Laksa Flavour
23	2558	KOKA	Creamy Soup With Crushed Noodles Hot & Sour Fish Flavor
37	2543	KOKA	Curry Flavour Instant Noodles
39	2541	Nissin	Cup Noodles Very Veggie Spicy Chicken Flavor Ramen Noodle Soup
41	2539	Nissin	Demae Iccho Tonkotsu Flavour Instant Noodle
42	2538	Nissin	Cup Noodles Very Veggie Beef Flavor Ramen Noodle Soup
43	2537	Nissin	Cup NoodlesTom Yam Seafood (More Shrimptacular)

1-10 of 396 rows | 1-4 of 7 columns   Previous **1** 2 3 4 5 6 … 40 Next

Hide

```
subset(ramrate, Brand == "Nissin" | Country != "Japan")
```

	Revi… <fctr>	Brand <fctr>	Flavor <fctr>
2	2579	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles
3	2578	Nissin	Cup Noodles Chicken Vegetable
4	2577	Wei Lih	GGE Ramen Snack Tomato Flavor
5	2576	Ching's Secret	Singapore Curry
6	2575	Samyang Foods	Kimchi song Song Ramen
10	2571	KOKA	The Original Spicy Stir-Fried Noodles
11	2570	Tao Kae Noi	Creamy tom Yum Kung Flavour
12	2569	Yamachan	Yokohama Tonkotsu Shoyu
13	2568	Nongshim	Mr. Bibim Stir-Fried Kimchi Flavor
14	2567	Nissin	Deka Buto Kimchi Pork Flavor

1-10 of 2,336 rows | 1-4 of 7 columns   Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
subset(ramrate, select = c(Flavor, Country, Rating))
```

	Flavor <fctr>	Country <fctr>	R <f
1	T's Restaurant Tantanmen	Japan	G
2	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles	Taiwan	Pe
3	Cup Noodles Chicken Vegetable	USA	Av
4	GGE Ramen Snack Tomato Flavor	Taiwan	Av
5	Singapore Curry	India	G
6	Kimchi song Song Ramen	South Korea	E:
7	Spice Deli Tantan Men With Cilantro	Japan	G
8	Nabeyaki Kitsune Udon	Japan	G
9	Hokkaido Soy Sauce Ramen	Japan	Pe

	Flavor <fctr>	Country <fctr>	R <f
10	The Original Spicy Stir-Fried Noodles	Singapore	A

1-10 of 2,577 rows   Previous **1** 2 3 4 5 6 … 100 Next

[Hide]

```
ramrate <- subset(ramrate, select = c(Brand, Flavor, Style, Country, Stars, Rating, Review))
ramrate
```

	Brand <fctr>	Flavor <fctr>
1	New Touch	T's Restaurant Tantanmen
2	Just Way	Noodles Spicy Hot Sesame Spicy Hot Sesame Guan-miao Noodles
3	Nissin	Cup Noodles Chicken Vegetable
4	Wei Lih	GGE Ramen Snack Tomato Flavor
5	Ching's Secret	Singapore Curry
6	Samyang Foods	Kimchi song Song Ramen
7	Acecook	Spice Deli Tantan Men With Cilantro
8	Ikeda Shoku	Nabeyaki Kitsune Udon
9	Ripe'n'Dry	Hokkaido Soy Sauce Ramen
10	KOKA	The Original Spicy Stir-Fried Noodles

1-10 of 2,577 rows | 1-4 of 7 columns   Previous **1** 2 3 4 5 6 … 100 Next

[Hide]

```
NA
NA
```

[Hide]

```
UJKdata <- subset(ramrate, Country == "USA" | Country == "Japan" | Country == "South Korea")

countryRating <- table(droplevels(UJKdata$Country),UJKdata$Rating)
countryRatingProp <- prop.table(countryRating)
countryRatingProp <- addmargins(countryRatingProp)

countryRatingProp
```

	Average	Below Average	Excellent	Good	Poor	Sum
Japan	0.036659878	0.009164969	0.161914460	0.143584521	0.007128310	0.358452138
South Korea	0.030549898	0.011201629	0.093686354	0.172097760	0.005091650	0.312627291
USA	0.052953157	0.026476578	0.070264766	0.167006110	0.012219959	0.328920570
Sum	0.120162933	0.046843177	0.325865580	0.482688391	0.024439919	1.000000000

Hide

```
###########################################################################################
##########
goodCountryProp <- countryRatingProp[1:3,4]/countryRatingProp[4,4]
goodCountryProp
```

```
     Japan South Korea         USA
 0.2974684   0.3565401   0.3459916
```

Hide

```
barplot(goodCountryProp[order(goodCountryProp ,decreasing = TRUE)], main="Proportion of Countrie
s with a Good Rating", ylim = c(0,0.7))
```


Proportion of Countries with a Good Rating
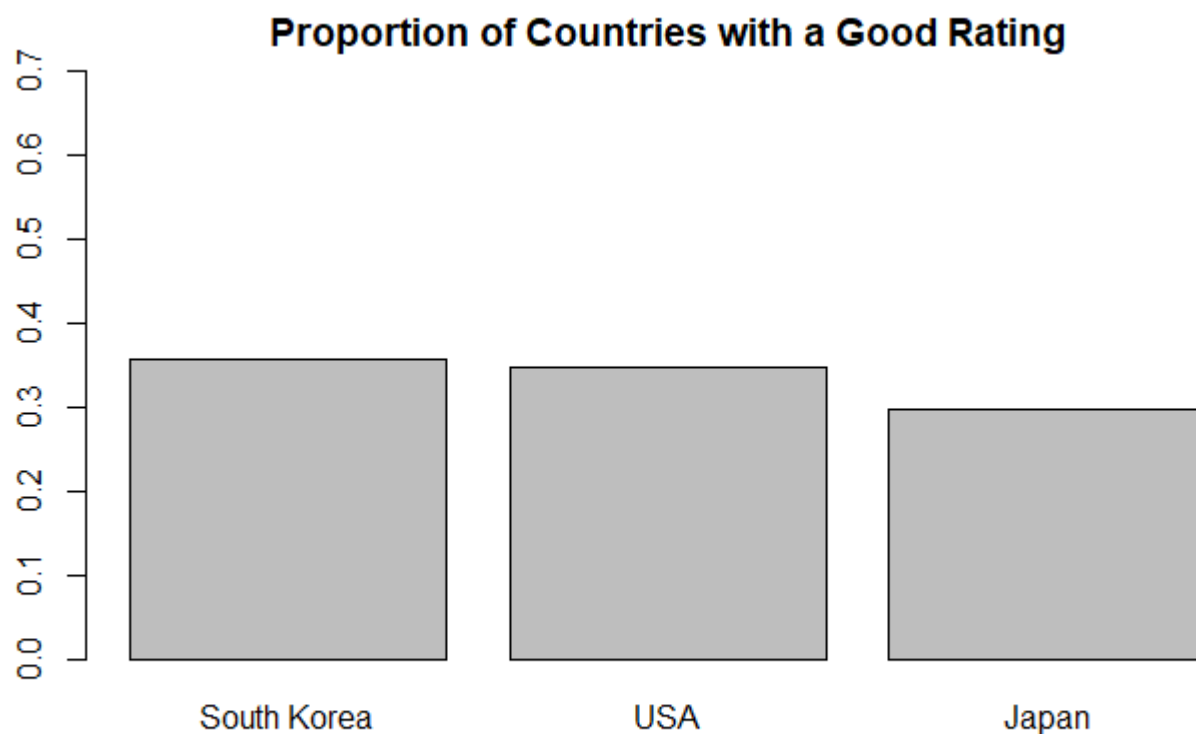
Hide

```
######################################################################################
##########

belowAvgCountryProp <- countryRatingProp[1:3,2]/countryRatingProp[4,2]
belowAvgCountryProp
```

```
      Japan South Korea         USA
  0.1956522   0.2391304   0.5652174
```

Hide

```
barplot(belowAvgCountryProp[order(belowAvgCountryProp ,decreasing = TRUE)], main="Proportion of
 Countries with a Below Average Rating", ylim = c(0,0.7))
```

Proportion of Countries with a Below Average Rating
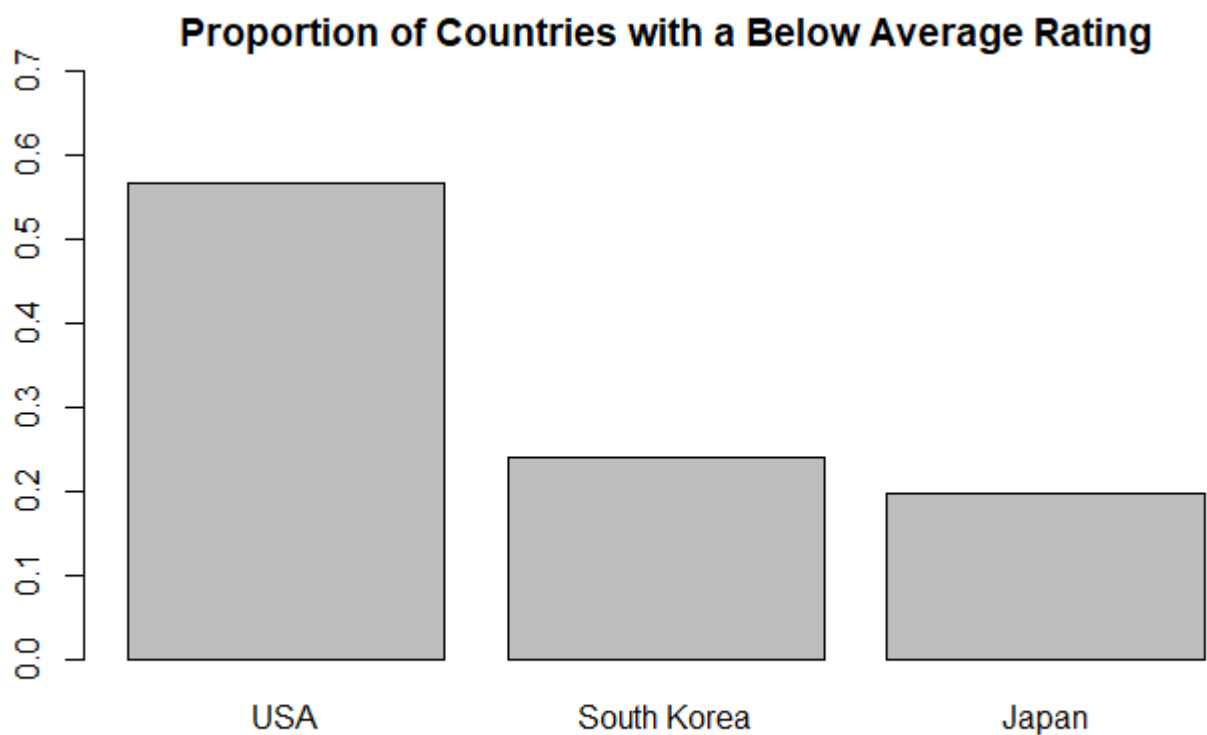
Hide

```
NA
NA
```

*Comments:

Based on the Pareto Charts above, it looks like South Korea just barely has more good ramen places than USA with Japan not too far behind in third place.
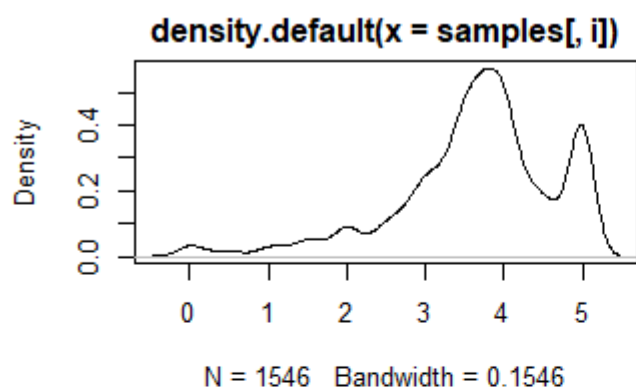
When looking at which country has the most below average restuarants, USA takes the cake by far with South Korea in second and Japan in last place again.
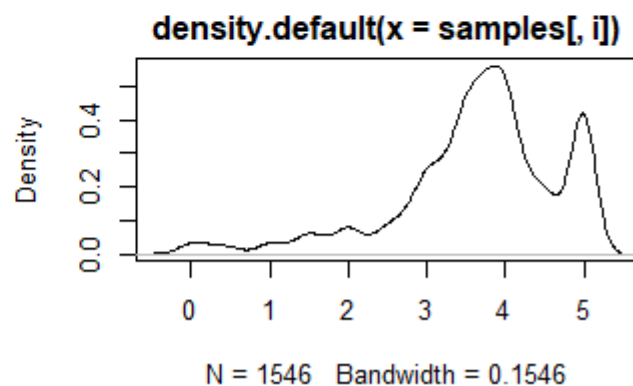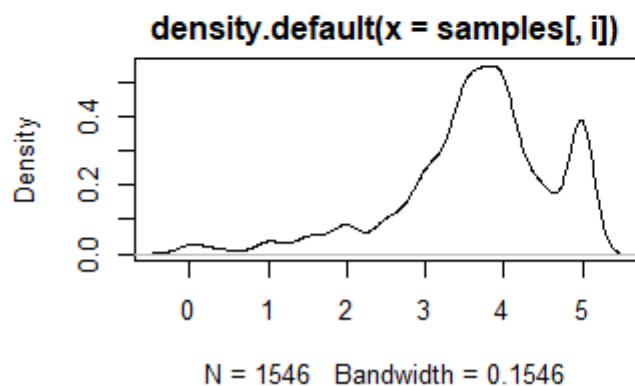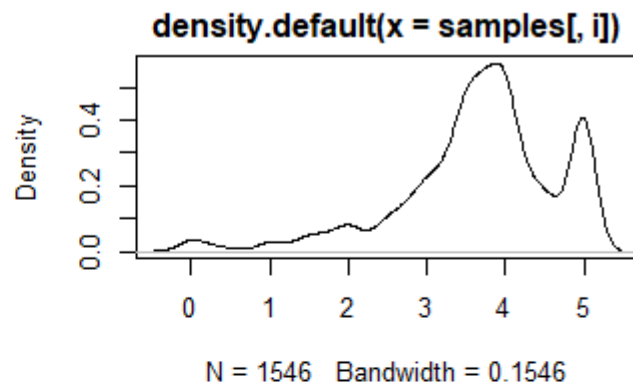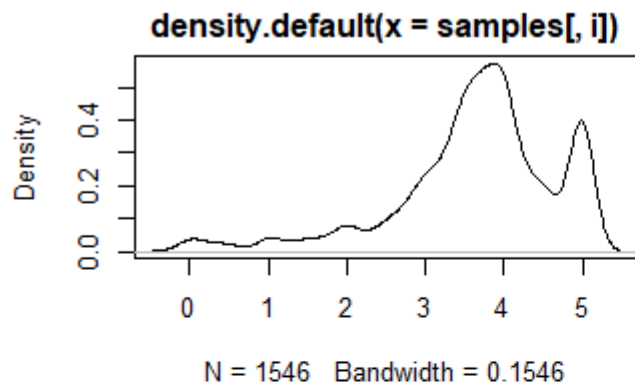
Hide

```
samples <- replicate(n=8, sample(ramrate$Stars, size=.60*length(ramrate$Stars)))

par(mfrow=c(2,2))

for(i in 1:8){

  den <- density(samples[,i])
  plot(den)
}
```


density.default(x = samples[, i])

density.default(x = samples[, i])

density.default(x = samples[, i])

density.default(x = samples[, i])

density.default(x = samples[, i])

N = 1546   Bandwidth = 0.1546

density.default(x = samples[, i])

N = 1546   Bandwidth = 0.1546

density.default(x = samples[, i])

N = 1546   Bandwidth = 0.1546

density.default(x = samples[, i])

N = 1546   Bandwidth = 0.1546
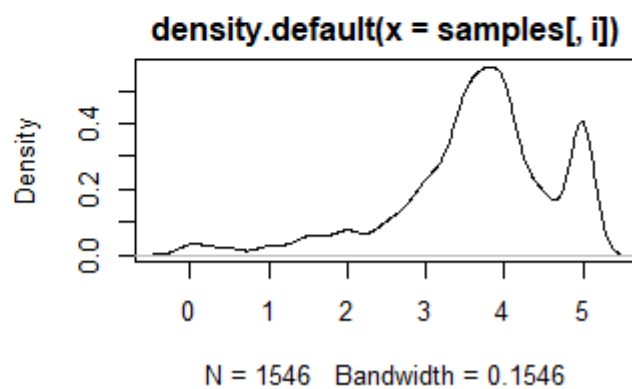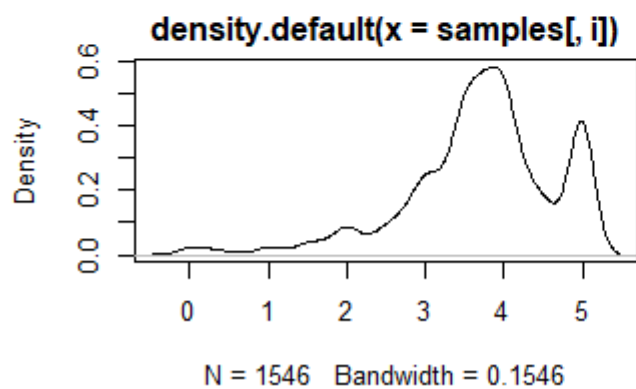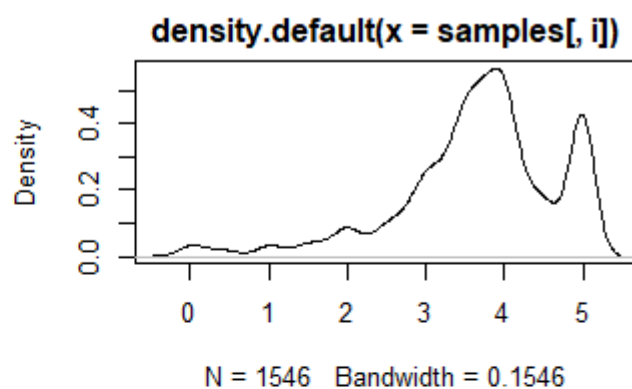
Hide

```
#The Resulting graphs are NOT normallly distributed!!!

##################################################################################


mnsamples1 <- replicate(n=8, mean(sample(ramrate$Stars, size=.60*length(ramrate$Stars))))
Starmean1 <- density(mnsamples1)
plot(Starmean1)

#I observe that after taking the mean of the samples...I find that the distribution looks more n
ormal! In fact, as the numebr of samples goes to infinity the mean of those samples will display
a perfectly normal distribution.
##################################################################################


mnsamples2 <- replicate(n=800, mean(sample(ramrate$Stars, size=.60*length(ramrate$Stars))))
Starmean2 <- density(mnsamples2)
plot(Starmean2)
```
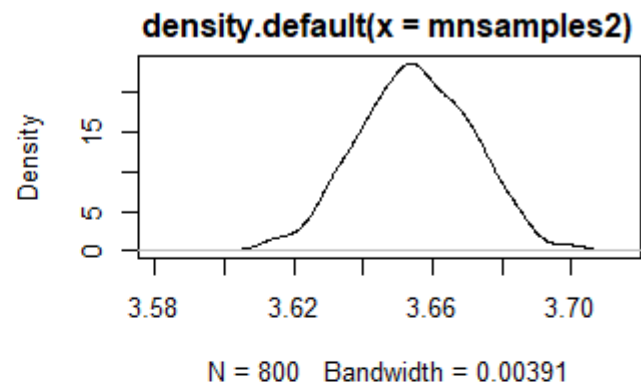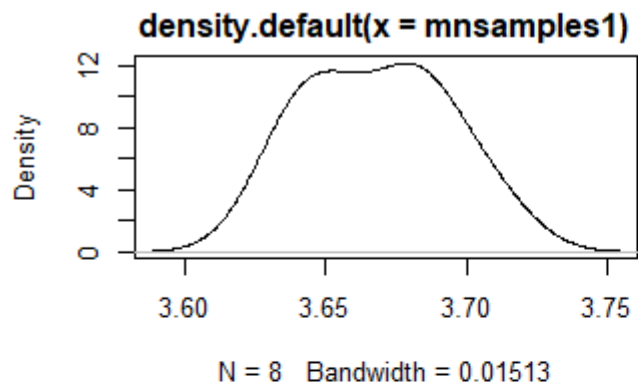
density.default(x = mnsamples1)

N = 8   Bandwidth = 0.01513

density.default(x = mnsamples2)

N = 800   Bandwidth = 0.00391

Hide

```
# This plot shows a more normal distribution just as predicted.
```

*Comment: The 8 individual samples dont appear to be normally distributed!

After 8 samples are averaged, it looks like it is getting more normally distributed.

After 800 samples are averaged, it looks like it is definitely approaching a normal distribution!