

PG-3 Test Buritos Review Data

Rishabh

November 16, 2018

1. Create a variable called ritos and store your data in it.

```
ritos <- read.csv("Burritos_Data.csv")
```

2. Print the first 6 rows of ritos

```
ritos[1:6,]
```

```
##           Location      Burrito      Date Neighborhood
## 1  graciela's taco shop  adobada  2/14/2016  chula vista
## 2  graciela's taco shop  barbacoa 2/14/2016  chula vista
## 3  cortez mexican food  california 10/7/2016  bonsall
## 4  el pueblo mexican food  california 8/10/2016  cardiff
## 5  pollos maria  california 1/27/2016  carlsbad
## 6  senor grubby's  california 4/24/2016  carlsbad
##           Address  Yelp  Google  Chips  Cost  Hunger  Length  Circum
## 1  5047 Central Ave  4.0    4.5    No  5.99    3    NA    NA
## 2  5047 Central Ave  4.0    4.5    No  5.99    2    NA    NA
## 3  5517 Mission Rd  4.2    4.0    No  6.25    4  22.5  18.0
## 4  820 Birmingham Dr 4.0    4.7    No  4.99    4  19.0  21.0
## 5  3055 Harding St  4.0    3.8    Yes 6.59    4    NA    NA
## 6  377 Carlsbad Village Dr 4.0    4.1    No  9.00    2  19.0  21.5
##  Volume Tortilla Temp Meat Fillings Meat.filling Uniformity Salsa Synergy
## 1  NA      4.0  4.0  3.0      3.5      4.0      4.5  4.0  4.0
## 2  NA      3.5  4.0  3.5      NA      4.0      NA  4.0  4.0
## 3  0.58     3.5  4.0  2.5      3.0      1.5      2.5  2.5  2.8
## 4  0.67     4.5  4.5  3.5      4.0      4.5      5.0  2.5  4.5
## 5  NA      4.0  5.0  4.0      3.5      4.5      5.0  2.5  4.5
## 6  0.70     2.0  3.5  3.0      1.5      1.0      1.0  2.5  1.5
##  Wrap Taste  Rec Reviewer
## 1  4.5  4.0 <NA>  Scott
## 2  1.5  4.0 <NA>  Emily
## 3  5.0  3.2  no  Scott
## 4  5.0  4.3  yes  Scott
## 5  4.0  4.2 <NA>  Scott
## 6  3.5  1.5  no  Scott
##
##                                     Notes
## 1                                     Salsa verde spicy
## 2
## 3
## 4
## 5
## 6 The ends of the burrito were decent, but the middle was terrible without much meat
##  Unreliable NonSD Beef Pico Guac Cheese Fries Sour.cream Pork Chicken
## 1          0  NA  NA  NA  NA  NA  NA  NA  1  NA
## 2          0   1  NA  NA  NA  NA  NA  NA  NA  NA
```

```
## 3      0      1      1      1      1      1      1      NA      NA
## 4      0      1      1      NA      1      1      NA      NA      NA
## 5      0      1      1      NA      1      1      NA      NA      NA
## 6      0      1      1      NA      1      1      NA      NA      NA
##   Shrimp Fish Rice Beans Lettuce Tomato Bell.peper Carrots Cabbage Sauce
## 1      NA   NA   NA      1      NA      NA      NA      NA      NA   NA
## 2      NA   NA   NA      NA      NA      NA      NA      NA      NA   NA
## 3      NA   NA   NA      NA      NA      NA      NA      NA      NA   NA
## 4      NA   NA   NA      NA      NA      NA      NA      NA      NA   NA
## 5      NA   NA   NA      NA      NA      NA      NA      NA      NA   NA
## 6      NA   NA   NA      NA      NA      NA      NA      NA      NA   NA
##   Salsa.1 Cilantro Onion Taquito Pineapple Ham Chile.relleno Nopales
## 1      1      1      1      NA      NA      NA      NA      NA
## 2      NA      1      1      NA      NA      NA      NA      NA
## 3      NA      NA      NA      NA      NA      NA      NA      NA
## 4      NA      NA      NA      NA      NA      NA      NA      NA
## 5      NA      NA      NA      NA      NA      NA      NA      NA
## 6      NA      NA      NA      NA      NA      NA      NA      NA
##   Lobster Queso Egg Mushroom Bacon Sushi Avocado Corn Zucchini
## 1      NA   NA   NA      NA      NA      NA      NA      NA      NA
## 2      NA   NA   NA      NA      NA      NA      NA      NA      NA
## 3      NA   NA   NA      NA      NA      NA      NA      NA      NA
## 4      NA   NA   NA      NA      NA      NA      NA      NA      NA
## 5      NA   NA   NA      NA      NA      NA      NA      NA      NA
## 6      NA   NA   NA      NA      NA      NA      NA      NA      NA
```

3. Print the dimension of ritos

```
dim(ritos)
```

```
## [1] 339  63
```

4. Filter the data using variable nonSD == 0, store it in ritos and print the dimension of this dataset

```
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
```

```
ritos <- filter(ritos, NonSD == 0)
dim(ritos)
```

```
## [1] 332  63
```

5. Print the number of unique values for Burrito variable

```
unique(ritos$Burrito)
```

```
## [1] adobada                barbacoa
## [3] california             california
## [5] carne asada            carnitas
## [7] chicken                custom
## [9] al pastor              shredded beef
## [11] chicken nopalito       deborah's special
## [13] adobada                california + guac + sour cream
## [15] california surf        chicken shawarma
## [17] fish                   mahi
## [19] philly                 quesa
## [21] quesaburro             surf & turf
## [23] chicken asada          cali diablo
## [25] california chipotle    pork california
## [27] chile verde pork       especial
## [29] machaca                ranchero steak
## [31] arizona                bomb
## [33] chimichanga beef       bitchin california
## [35] pollo adobado          tijuana
## [37] 2 in 1                 bean and cheese
## [39] mauna lani             mixed
## [41] pokirrito classic      shrimp
## [43] breakfast              california (only cheese)
## [45] california everything  california everything mini
## [47] chile relleno and carnitas hot cheetos
## [49] spicy a la diabla      surfin california
## [51] tgunz                  al pastor
## [53] bacon breakfast        battered fish
## [55] campeon                chile relleno
## [57] dave's california      el rusio
## [59] fajitas                bandido
## [61] california - chicken   california - pork adobada
## [63] california - steak     colimas burrito
## [65] hawaiian               holy moly
## [67] nutty                  veg out
## [69] vegetarian             california burrito
## [71] chicken avocado        el hawaiiano
## [73] fusion                 local
## [75] addiction              california chicken
## [77] oaxacalifornia         carne adobada
## [79] pollo asado            cabeza
## [81] azteca                 lobster
## [83] 619 burrito original   baja monster
## [85] california breakfast   chicken and rice
## [87] monster california     shrimp with guac
## [89] tejano                 tilapia
## [91] asada                  bean and rice grande size
## [93] pollo california
## 95 Levels: 2 in 1 619 burrito original addiction adobada ... vegetarian
```

6. Using the Burrito variable, filter data for california and print the number of unique values for the same.

Hint : Use grepl to filter 'california'

```
a<-filter(ritos, grepl('california', ritos$Burrito))
length(unique(a$Burrito))
```

```
## [1] 21
```

7. Select the variables Location and Yelp and print the first 6 rows

```
b <- ritos[c(1,6)]
b[1:6,]
```

```
##           Location Yelp
## 1  graciela's taco shop 4.0
## 2  graciela's taco shop 4.0
## 3  cortez mexican food 4.2
## 4  el pueblo mexican food 4.0
## 5           pollos maria 4.0
## 6      senor grubby's 4.0
```

8. Select columns with Burrito ratings i.e. from Tortilla to Wrap and print the first 6 rows

```
c <- select(ritos, Tortilla:Wrap)
c[1:6,]
```

```
##   Tortilla Temp Meat Fillings Meat.filling Uniformity Salsa Synergy Wrap
## 1      4.0  4.0  3.0      3.5          4.0        4.5  4.0    4.0  4.5
## 2      3.5  4.0  3.5      NA          4.0        NA  4.0    4.0  1.5
## 3      3.5  4.0  2.5      3.0          1.5        2.5  2.5    2.8  5.0
## 4      4.5  4.5  3.5      4.0          4.5        5.0  2.5    4.5  5.0
## 5      4.0  5.0  4.0      3.5          4.5        5.0  2.5    4.5  4.0
## 6      2.0  3.5  3.0      1.5          1.0        1.0  2.5    1.5  3.5
```

9. Create a new data frame df from ritos, select only columns 'Taste', 'Rec', 'Cost', 'Volume'

```
df <- select(ritos, Taste, Rec, Cost, Volume)
```

10. Create a new column('Cost_vol') in the same dataframe df using the formula(Cost/Volume), print the first 6 rows of the df

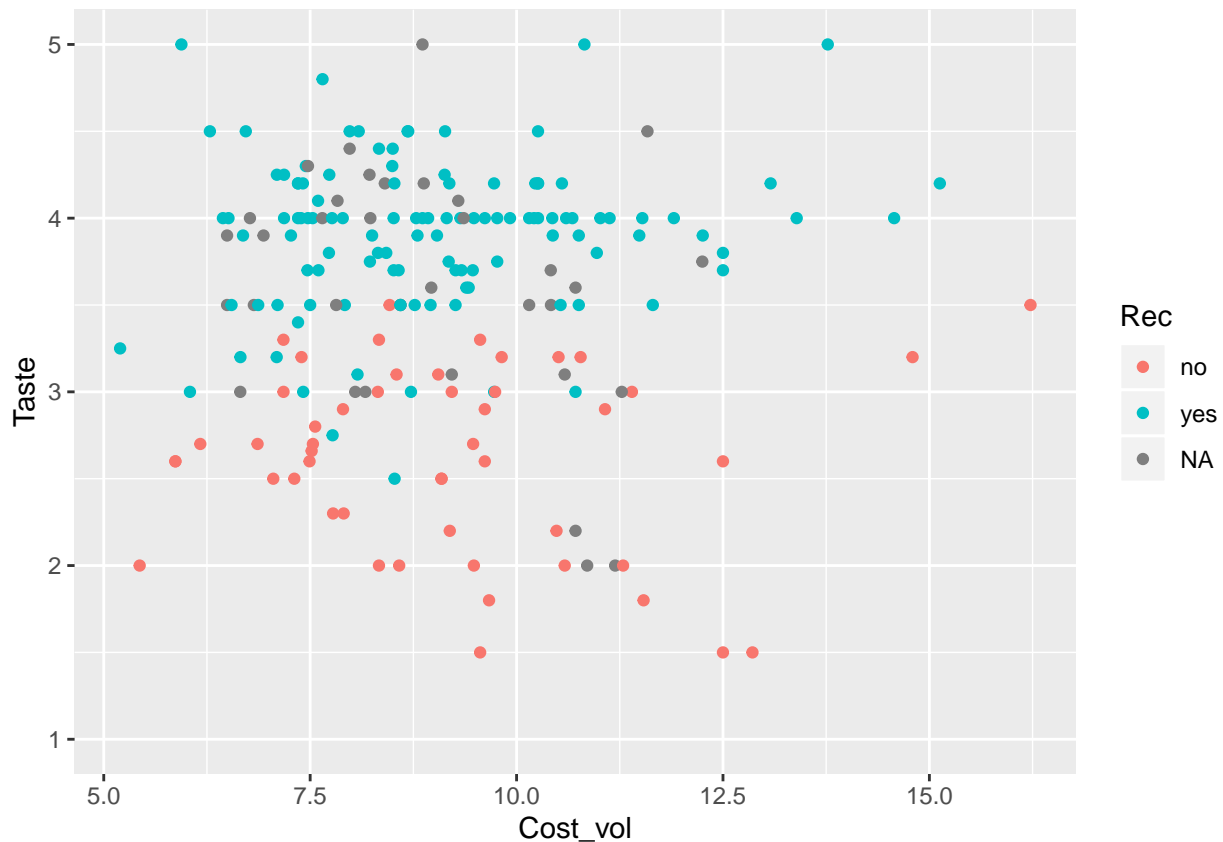
```
Cost_vol <- ritos$Cost/ritos$Volume
d <- cbind(Cost_vol, df)
d[1:6,]
```

```
##   Cost_vol Taste  Rec Cost Volume
## 1      NA   4.0 <NA> 5.99     NA
## 2      NA   4.0 <NA> 5.99     NA
## 3 10.775862   3.2   no 6.25  0.58
## 4  7.447761   4.3  yes 4.99  0.67
## 5      NA   4.2 <NA> 6.59     NA
## 6 12.857143   1.5   no 9.00  0.70
```

11. Plot(dotplot) Taste on y axis and Cost_vol on x axis for df and colour code by variable Rec

```
library(ggplot2)
ggplot(df, aes(x = Cost_vol, y = Taste))+
  geom_point(aes(colour = Rec))
```

```
## Warning: Removed 132 rows containing missing values (geom_point).
```



12. Select burrito rating variable and recommendation(select variable rec and Tortilla to Taste), store this in a new variable called ing and print the first 6 rows

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
ing <- select(ritos, c(Rec, Tortilla: Taste))
head(ing)
```

```
##      Rec Tortilla Temp Meat Fillings Meat.filling Uniformity Salsa Synergy
## 1 <NA>      4.0  4.0  3.0      3.5      4.0      4.5  4.0    4.0
## 2 <NA>      3.5  4.0  3.5      NA      4.0      NA  4.0    4.0
## 3  no      3.5  4.0  2.5      3.0      1.5      2.5  2.5    2.8
## 4  yes      4.5  4.5  3.5      4.0      4.5      5.0  2.5    4.5
## 5 <NA>      4.0  5.0  4.0      3.5      4.5      5.0  2.5    4.5
## 6  no      2.0  3.5  3.0      1.5      1.0      1.0  2.5    1.5
##      Wrap Taste
## 1  4.5    4.0
## 2  1.5    4.0
## 3  5.0    3.2
## 4  5.0    4.3
## 5  4.0    4.2
## 6  3.5    1.5
```

13. Convert the above data to long format and store in variable `df.melt`; use 'Rec' and 'Taste' as your id variables, print the structure of this dataset

```
#library(UsingR)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:data.table':
##
##      dcast, melt
```

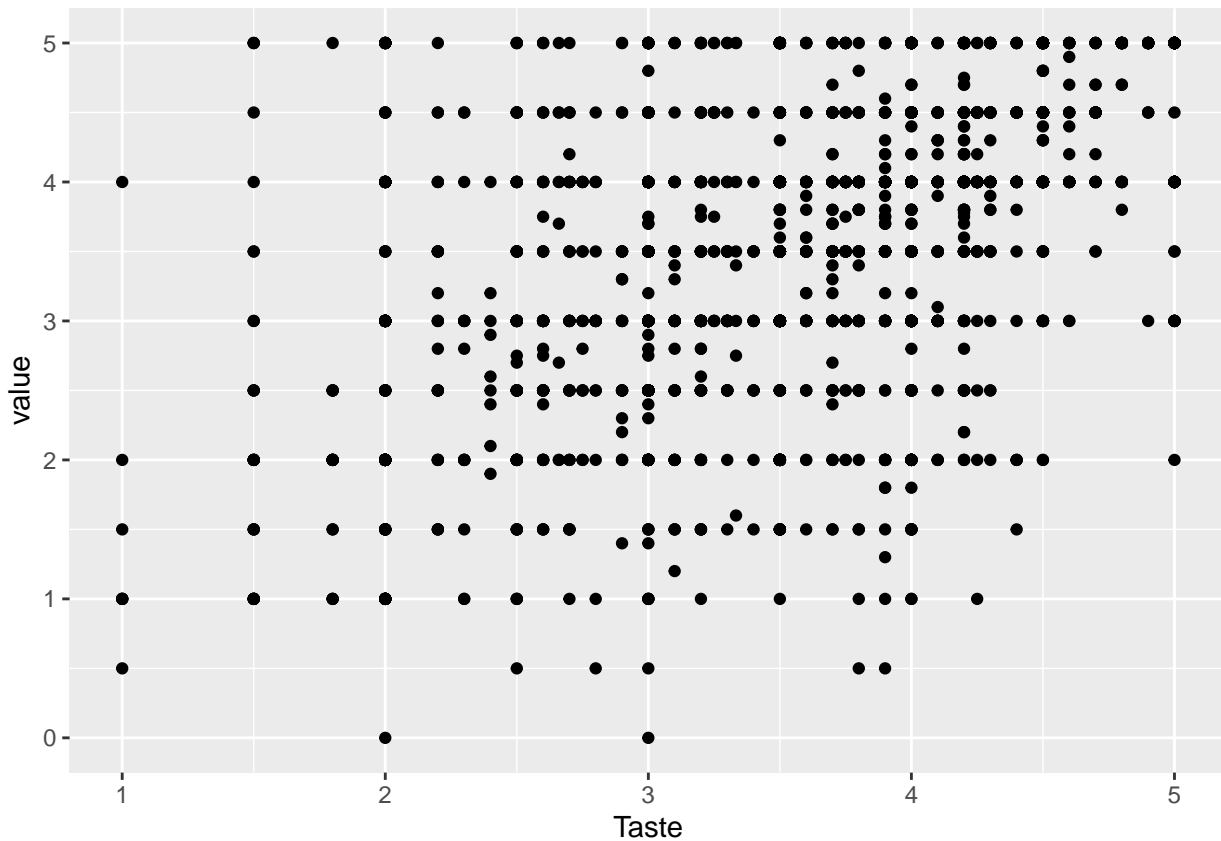
```
df.melt<-melt(ing, id.vars=c('Rec','Taste'))
str(df.melt)
```

```
## 'data.frame':    2988 obs. of  4 variables:
## $ Rec      : Factor w/ 2 levels "no","yes": NA NA 1 2 NA 1 NA 1 2 NA ...
## $ Taste    : num  4 4 3.2 4.3 4.2 1.5 3 3.5 2.75 3.2 ...
## $ variable: Factor w/ 9 levels "Tortilla","Temp",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ value    : num  4 3.5 3.5 4.5 4 2 2.5 3.5 2.5 3 ...
```

14. Plot(ggplot and `geom_point`) the above dataset using x variable 'Taste' and y variable 'Value'

```
ggplot(df.melt, aes(x = Taste, y = value)) + geom_point()
```

```
## Warning: Removed 83 rows containing missing values (geom_point).
```



15. Create a `facet_wrap` of `df.melt` use x-axis="Taste", y-axis="Value" and colour = "variable", use `geom_point`, add a smoother(lm) using `geom_smooth`, colour the smoother in "black", give the x axis label as 'Burrito Qulaity' and y-axis label as 'score'

```
ggplot(df.melt, aes(x = Taste, y = value, colour = variable)) +
  geom_point() + geom_smooth(method='lm', color='black') + facet_wrap(~variable) +
  labs(x='Burrito Qulaity', y='Score')
```

```
## Warning: Removed 83 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 83 rows containing missing values (geom_point).
```

