Data Visualization: Cookie Analysis

Andrea Lambert: #215035587
Shayla Schwartz: #216564577
Sidrah Mirza: #216707044

PSYC 3031A

Dr. Monique Herbert

November 4, 2020

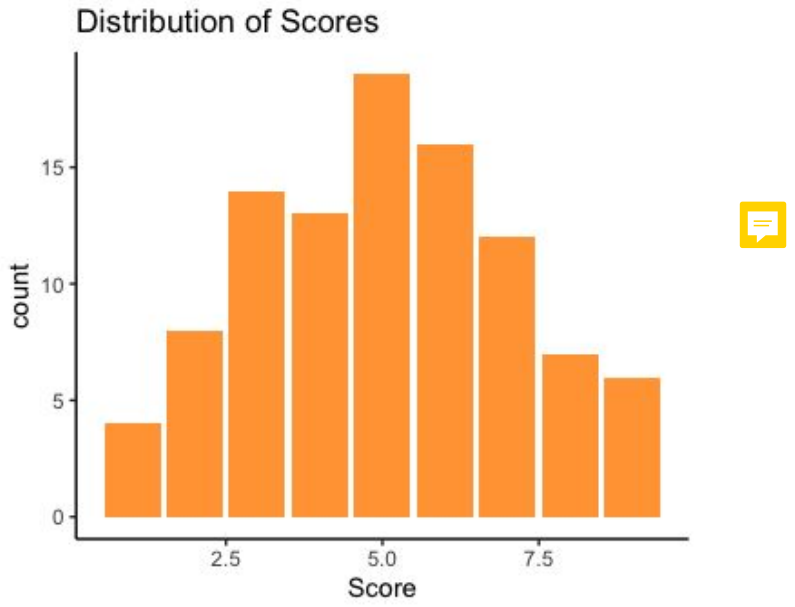
Research Question

The purpose of this report is to examine how an individual measures the tastiness of different cookies and whether it is impacted by recipe or personal preference. We will be dividing our analysis into two primary areas; recipe and taster. The purpose of these sections is to examine variance within the data. If recipe is deemed the more important factor we should see a large difference in recipe means and sums of ingredients with consistent means for tasters. Alternatively if personal taster is deemed the more important factor we should see consistent means for recipe and ingredients and a large variance in taster means.

Description of Data

Our data "Cookies" was retrieved from the website *The Data and Story Library.* Eleven tasters tested nine different cookies and gave each a rating from one to ten (one was the worst and ten was the best.) The cookie recipes varied the amount of sugar (0.25, 0.375 and 0.5 cups) the type of oil (Canola, Vegetable and Olive) and type of chocolate chip (Milk, Semisweet and dark) to determine the best recipe.

	Score	Sugar	Oil	Chip	Taster
1	5	0.25	Canola	Milk	11
2	7	0.375	Canola	Semisweet	11
3	2	0.5	Canola	Dark	11
4	5	0.25	Vegetable	Dark	11
5	7	0.375	Vegetable	Milk	11
6	4	0.5	Vegetable	Semisweet	11
7	3	0.25	Olive	Semisweet	11
8	5	0.375	Olive	Dark	11
9	4	0.5	Olive	Milk	11
10	8	0.25	Canola	Milk	7
11	9	0.375	Canola	Semisweet	7
12	2	0.5	Canola	Dark	7
13	4	0.25	Vegetable	Dark	7
14	1	0.375	Vegetable	Milk	7

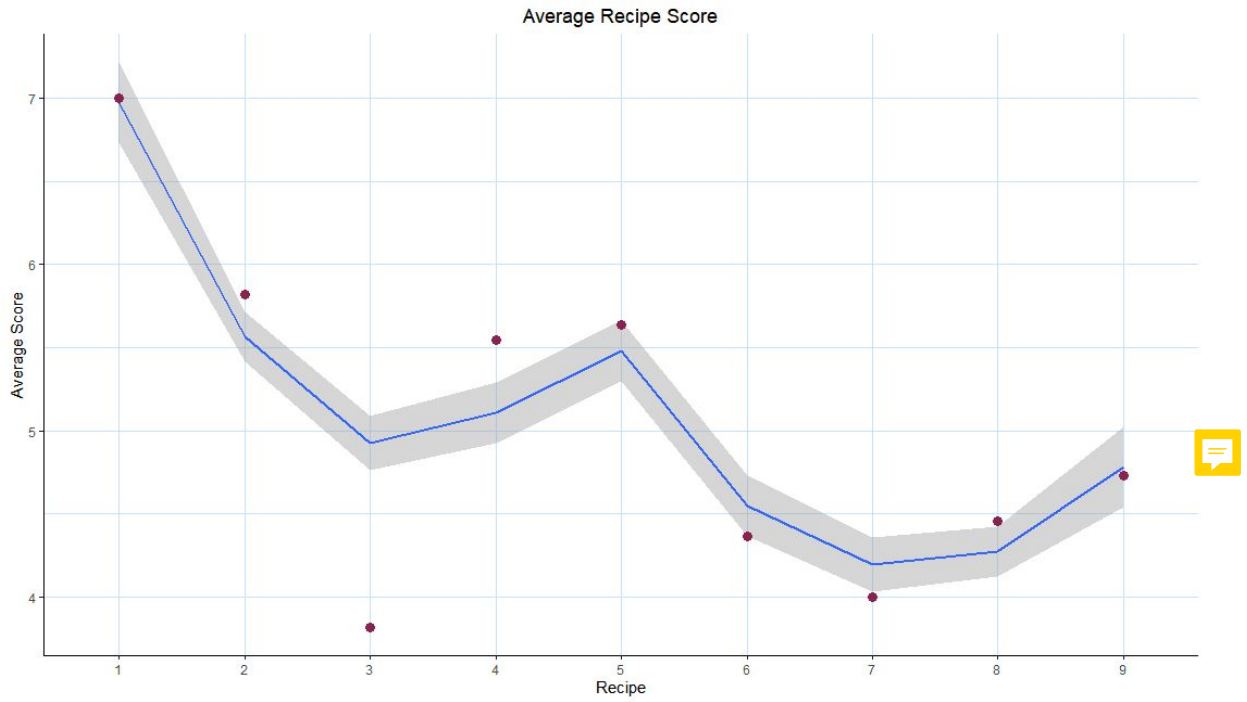


This graph displays the distribution of scores. The X axis displays the scores from one to 10 and the Y axis displays the number of times the score appears in the data.

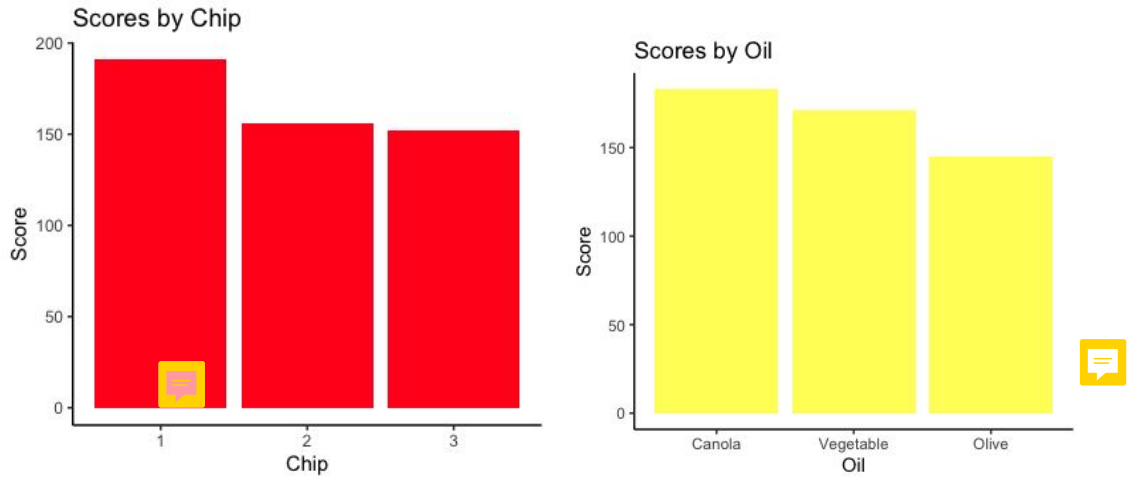
Recipe Analysis

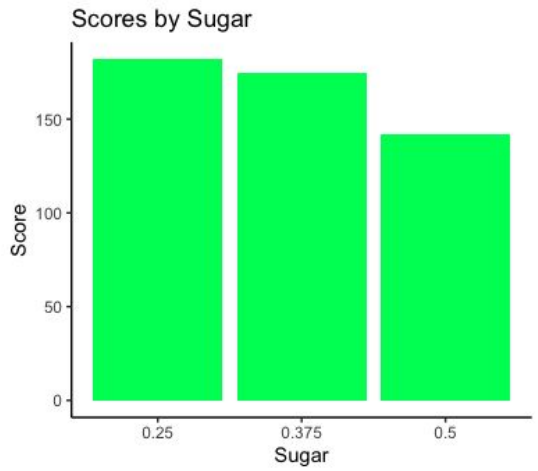
	Recipe	Sugar	Oil	Chip	Recipe_Avg
1	1	0.25	Canola	Milk	7.00
2	2	0.375	Canola	Semisweet	5.81
3	3	0.5	Canola	Dark	3.81
4	4	0.25	Vegetable	Dark	5.54
5	5	0.375	Vegetable	Milk	5.63
6	6	0.5	Vegetable	Semisweet	4.36
7	7	0.25	Olive	Semisweet	4.00
8	8	0.375	Olive	Dark	4.45
9	9	0.5	Olive	Milk	4.72

The above dataset examines the nine different recipes; their ingredients and their average rating.



This graph measures the average scores of each of the individual recipes. The purple dots refer to the means, the blue line is the line of best fit and the grey area represents standard error. Examining this graph we see moderate fluctuations amongst recipe means The most popular recipe is #1 with a mean of 7.00. This recipe contains 0.25 cups of sugar, canola oil and milk chocolate chips. The least popular recipe is #3 (mean = 3.81) which contains 0.5 cups of sugar, oil and dark chocolate. It is worth noting that despite fluctuations the majority of means fall within the standard error.





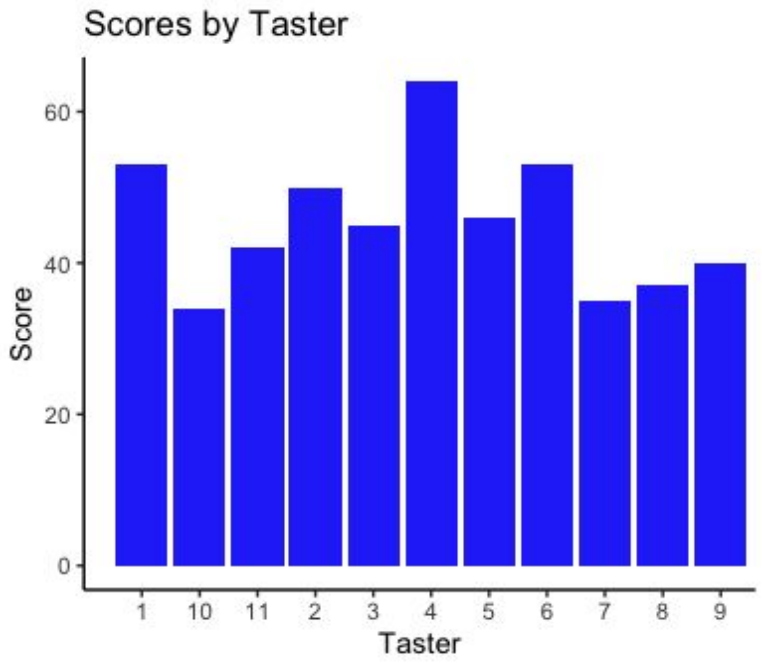
The graphs above illustrate the sum totals for the different types of chocolate chips, oil and amount of sugar. Chocolate chips seem to have the largest effect on scores with milk chocolate having the largest sum total. Variability amongst sugar and oil scores

Taster Analysis

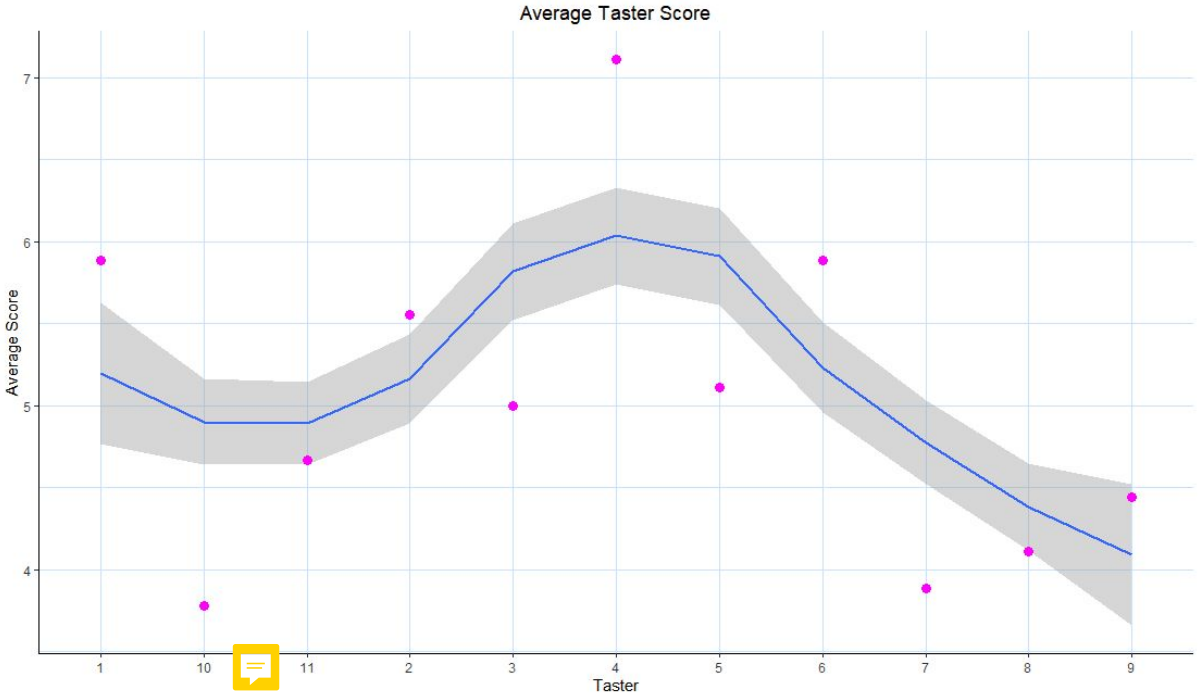
In this section we examine how the individual taster's scores fluctuate.

	Taster	Recipe1	Recipe2	Recipe3	Recipe4	Recipe5	Recipe6	Recipe7	Recipe8	Recipe9	Taster_Avg
1	1	9	7	3	4	6	8	6	5	5	5.89
2	2	6	3	4	5	6	3	6	8	9	5.56
3	3	8	6	4	7	6	3	5	4	2	3.83
4	4	9	8	5	8	9	5	6	7	7	7.11
5	5	6	6	6	4	6	5	5	5	3	5.11
6	6	8	7	2	9	6	5	4	5	7	5.89
7	7	8	9	2	4	1	5	3	1	2	3.89
8	8	5	5	6	3	7	2	4	2	3	4.11
9	9	7	3	3	5	6	4	1	4	7	4.44
10	10	6	3	5	7	2	4	1	3	3	3.78
11	11	5	7	2	5	7	4	3	5	4	4.67

The above dataset examines how each taster rated the different recipes and their mean score.



This graph displays the distribution of data. The Y axis shows the sum total of the tasters and the X axis shows the tasters



This graph measures the averages of the individual tasters The purple dots refer to the means, the blue line is the line of best fit and the grey area represents standard error. Using this graph we can see that there is a large fluctuation amongst tasters. Taster 10 achieved the lowest average rating with a mean of 3.78, a maximum score of 7 and a mode of 3. Conversely taster 4 achieved the highest average with a mean of 7.11 a maximum of nine and a minimum of 4. Additionally, 7 out of 10 participants fall outside of the expected standard error while the remaining three fall around its edge.

Limitations

Our retrieved data lacked a classification scale. The amount of sugar was recorded on a scale from zero to two and the types of oil and chocolate was recorded as one, two or three. There was no specification which amount/ type corresponded with each score. We assumed Sugar: 0 = 0.25, 1 = 0.375 2 = 0.5 cups , Oil: 0 = Canola, 1 = Vegetable, 2 = Olive, Chocolate Chips: 1 = Milk, 2 = Semisweet, 3 = Dark.

Additionally, the experiment only tested a total of nine cookie recipes. To truly determine the best cookie tasters should have tried 27 cookies, combining every level of each of the three variables. However, that is just too many cookies so it's understandable why they only tried 9.

We were unable to organize the taster in numerical order in the graphs. However, since this is a categorical variable it should not affect the results.

Discussion

Both recipe and individual preference seem to have an impact on perceived tastiness of a cookie. However, it appears that personal preference has the most impact of the two variables. Examination of the two graphs reveals greater variance in taster means compared to recipe means with most taster means falling outside of the expected standard error. Further examination of the data is needed to determine if these differences are significant.

R Code

```
#this is the data-visualization group project for group J
#install and load these packages to use later
install.packages("tidyverse")
install.packages("here")
library(tidyverse)
library(here)

#import data using read_csv function
cookieData <- read_csv(file = here("Data-V project", "Data", "cookieData.csv"))

# I constructed a new dataframe using the data.frame function and saved it as an object
# I set the columns equal to the recipe numbers (and avg) and used the combine function to
insert the ratings
# Taster I indicated the scores from one to 11
# I then used the view function to check my work
cookies_summary_taster <- data.frame(Taster = c(1:11),
                                     Recipe1 = c(9,6,8,9,6,8,8,5,7,6,5),
                                     Recipe2 = c(7,3,6,8,6,7,9,5,3,3,7),
                                     Recipe3 = c(3,4,4,5,6,2,2,6,3,5,2),
```

```
                        Recipe4 = c(4,5,7,8,4,9,4,3,5,7,5),
                        Recipe5 = c(6,6,6,9,6,6,1,7,6,2,7),
                        Recipe6 = c(8,3,3,5,5,5,5,2,4,4,4),
                        Recipe7 = c(6,6,5,6,5,4,3,4,1,1,3),
                        Recipe8 = c(5,8,4,7,5,5,1,2,4,3,5),
                        Recipe9 = c(5,9,2,7,3,7,2,3,7,3,4),
                        Taster_Avg =
c(5.89,5.56,3.83,7.11,5.11,5.89,3.89,4.11,4.44,3.78,4.67))


view(cookies_summary_taster)

# I constructed a new dataframe using the data.frame function and saved it as an object
# I set the columns equal to the ingredient (and avg) and used the combine function to insert
the ratings
# I used the rep function and indicated which numbers should be repeated and how many
times
# I then used the view function to check my work
cookies_summary_recipe <- data.frame(Recipe = c(1:9),
                        Sugar = rep(0:2, time = 3),
                        Oil = c(rep(0, time = 3), rep(1, time = 3), rep(2, time=3)),
                        Chip = c(1,2,3,3,1,2,2,3,1),
                        Recipe_Avg = c(7,5.81,3.81,5.54,5.63,4.36,4,4.45,4.72))
#using mutate function to recode the categorical variables into character and change the
level names
cookies_summary_recipe <- mutate(cookies_summary_recipe,
        Sugar = as.character(Sugar),
        Sugar = fct_recode(Sugar,
                    "0.25" = "0",
                    "0.375" = "1",
                    "0.5" = "2"),
        Oil = as.character(Oil),
        Oil = fct_recode(Oil,
                  "Canola" = "0",
                  "Vegetable" = "1",
                  "Olive" = "2"),
        Chip = as.character(Chip),
        Chip = fct_recode(Chip,
                    "Milk" = "1",
                    "Semisweet" = "2",
                    "Dark" = "3"))
view(cookies_summary_recipe)

#now using the original cookieData dataframe
#use the spec function to see what types the variables are showing up as
spec(cookieData)
```

```
#recode the categorical variables from numeric to character (Sugar, Oil, Chip and Taster)
and rename levels
#save the mutation to the original data using "cookieData <-"
#use the mutate function and within that function use the dataframe cookieData
#convert the chosen categorical variable into character using as.character
#recode the levels of the chosen variable using fct_recode
#repeat the last two comments with each categorical variable
cookieData <- mutate(cookieData,
                     Sugar = as.character(Sugar),
                     Sugar = fct_recode(Sugar,
                                        "0.25" = "0",
                                        "0.375" = "1",
                                        "0.5" = "2"),
                     Taster = as.character(Taster),
                     Taster = fct_recode(Taster,
                                         "1" = "1",
                                         "2" = "2",
                                         "3" = "3",
                                         "4" = "4",
                                         "5" = "5",
                                         "6" = "6",
                                         "7" = "7",
                                         "8" = "8",
                                         "9" = "9",
                                         "10" = "10",
                                         "11" = "11"),
                     Oil = as.character(Oil),
                     Oil = fct_recode(Oil,
                                      "Canola" = "0",
                                      "Vegetable" = "1",
                                      "Olive" = "2"),
                     Chip = as.character(Chip),
                     Chip = fct_recode(Chip,
                                       "Milk" = "1",
                                       "Semisweet" = "2",
                                       "Dark" = "3"))

#graph the distribution of scores using ggplot with the aesthetic argument of x with the
variable of score
#add the geom_bar function to turn the ggplot into a bar graph with the argument of fill =
"orange" to make the bars orange
#add the theme functions to remove the grey grid backround while keeping the x and y axis
lines black
#add the title distribution of scores using the labs function
score_distribution <- cookieData %>%
  ggplot(aes(x = Score)) +
  geom_bar(fill = "orange") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
```

```
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(title = "Distribution of Scores")
#print the graph
score_distribution

#create a graph showing the scores for the different levels of the chip variable
#save as the object score_chip
#grab cookieData and use a piping operator to feed it into a ggplot
#use the ggplot function with the aesthetic arguments of x equals the chip variable and y
equals the score variable
#use the geom_col function to turn the ggplot into a bar graph and include the argument fill =
red to turn the bars red
#use the theme function to remove the grey grid while keeping the x and y axis lines black
#use the labs function to change the title
score_chip <- cookieData %>%
  ggplot(aes(x = Chip, y = Score)) +
  geom_col(fill = "red") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
          panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(title = "Scores by Chip")

#print the graph
score_chip

#repeat with the variable of oil instead of chip and change the colour of the bars to yellow
score_oil <- cookieData %>%
  ggplot(aes(x = Oil, y = Score)) +
  geom_col(fill = "yellow") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(title = "Scores by Oil")

#print the graph
score_oil

#repeat with the variable of sugar and change the colour of the bars to green
score_sugar <- cookieData %>%
  ggplot(aes(x = Sugar, y = Score)) +
  geom_col(fill = "green") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(title = "Scores by Sugar")

#print the graph
score_sugar

#repeat with the variable of taster and change the colour of the bars to blue
#add the point of 0 to the x axis
```

```
score_taster <- cookieData %>%
  ggplot(aes(x = Taster, y = Score)) +
  geom_col(fill = "blue") +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  labs(title = "Scores by Taster") + expand_limits(x = 0)

#print the graph
score_taster

#create a new column with an id for each recipe
cookieData$Recipe = (rep(1:9, time=11))

#recode the recipe column to name the levels and ensure that it is a character type
cookieData<-mutate(cookieData,
                   Recipe = as.character(Recipe),
                   Recipe = fct_recode(Recipe,
                                       "1" = "1",
                                       "2" = "2",
                                       "3" = "3",
                                       "4" = "4",
                                       "5" = "5",
                                       "6" = "6",
                                       "7" = "7",
                                       "8" = "8",
                                       "9" = "9")

#create a new column for the cookieData data frame containing the average score for each
recipe

RecipeMns <- cookieData %>% group_by(Recipe) %>% mutate(RecipeMns = mean(Score))

# Use the ggplot function to create a smooth line graph
# Within the aes function the x and y axis are specified and the group function is required for
the smooth line
# use the geom_point function to add points to the graph
# within geom_point change point size to 3 and colour to violet
# The theme function is used to remove the grey grid and keep the axis lines black
ggplot(data = RecipeMns, mapping = aes(x = Recipe, y = RecipeMns, group = 1)) +
      geom_smooth()+
      geom_point(size = 3, colour = "violetred4")+
      labs(title = "Average Recipe Score",
                  x = "Recipe",
                  y= "Average Score")+
      theme(plot.title = element_text(hjust = 0.5), panel.grid.major = element_line(colour =
"slategray1"), panel.grid.minor = element_line(colour = "slategray1"),
            panel.background = element_blank(), axis.line = element_line(colour = "black"))
```

```
# create a new column on the cookieData dataframe which contains the average score per
taster

  TasterMns <- cookieData %>% group_by(Taster) %>% mutate(TasterMns = mean(Score))

# The ggplot function maps the data
# aes specifies the x and y axis and changes the colour and slides
# points are added and the point size and colour are changed within geom_point
# The titles background and line are adjusted
  ggplot(data = TasterMns, mapping = aes(x = Taster, y = TasterMns, group = 1))+
        geom_smooth()+
        geom_point(size = 3, colour = "magenta")+
        labs(title = "Average Taster Score",
                          x = "Taster",
                          y= "Average Score")+
        theme(plot.title = element_text(hjust = 0.5), panel.grid.major = element_line(colour =
"slategray1"), panel.grid.minor = element_line(colour = "slategray1"),
              panel.background = element_blank(), axis.line = element_line(colour = "black"))
```