```r
library(ggplot2)
library(dplyr)
```

```
##
##     'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

load data

```r
data <- read.csv('C:/Users/555/Desktop/602project/FastFoodNutritionMenuV3.csv')
head(data)
```

```
##      Company                                Item Calories Calories.from.Fat
## 1 McDonald’s                           Hamburger      250                80
## 2 McDonald’s                       Cheeseburger      300               110
## 3 McDonald’s                  Double Cheeseburger      440               210
## 4 McDonald’s                            McDouble      390               170
## 5 McDonald’s        Quarter Pounder® with Cheese      510               230
## 6 McDonald’s Double Quarter Pounder® with Cheese      740               380
##   Total.Fat..g. Saturated.Fat..g. Trans.Fat..g. Cholesterol..mg. Sodium...mg.
## 1             9               3.5           0.5                25          520
## 2            12                 6           0.5                40          750
## 3            23                11           1.5                80         1150
## 4            19                 8             1                65          920
## 5            26                12           1.5                90         1190
## 6            42                19           2.5               155         1380
##   Carbs..g. Fiber..g. Sugars..g. Protein..g. Weight.Watchers.Pnts
## 1        31         2          6          12                247.5
## 2        33         2          6          15                  297
## 3        34         2          7          25                  433
## 4        33         2          7          22                  383
## 5        40         3          9          29                  502
## 6        40         3          9          48                  720
```

check the data

```r
glimpse(data)
```

```
## Rows: 1,147
## Columns: 14
## $ Company              <chr> "McDonald’s", "McDonald’s", "McDonald’s", "McDona~
## $ Item                 <chr> "Hamburger", "Cheeseburger", "Double Cheeseburger~
## $ Calories             <chr> "250", "300", "440", "390", "510", "740", "540", ~
## $ Calories.from.Fat    <chr> "80", "110", "210", "170", "230", "380", "260", "~
```

```
## $ Total.Fat..g.        <chr> "9", "12", "23", "19", "26", "42", "29", "24", "2~
## $ Saturated.Fat..g.    <chr> "3.5", "6", "11", "8", "12", "19", "10", "8", "11~
## $ Trans.Fat..g.        <chr> "0.5", "0.5", "1.5", "1", "1.5", "2.5", "1.5", "1~
## $ Cholesterol..mg.     <chr> "25", "40", "80", "65", "90", "155", "75", "70", ~
## $ Sodium...mg.         <chr> "520", "750", "1150", "920", "1190", "1380", "104~
## $ Carbs..g.            <chr> "31", "33", "34", "33", "40", "40", "45", "37", "~
## $ Fiber..g.            <chr> "2", "2", "2", "2", "3", "3", "3", "3", "3", "4",~
## $ Sugars..g.           <chr> "6", "6", "7", "7", "9", "9", "9", "8", "8", "13"~
## $ Protein..g.          <chr> "12", "15", "25", "22", "29", "48", "25", "24", "~
## $ Weight.Watchers.Pnts <chr> "247.5", "297", "433", "383", "502", "720", "534"~
```

```
data[, 3:14] <- lapply(data[, 3:14], as.numeric)
```

```
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
## Warning in lapply(data[, 3:14], as.numeric):      NA
```

```
head(data)
```

```
##      Company                                Item Calories Calories.from.Fat
## 1 McDonald's                           Hamburger      250                80
## 2 McDonald's                        Cheeseburger      300               110
## 3 McDonald's                  Double Cheeseburger      440               210
## 4 McDonald's                            McDouble      390               170
## 5 McDonald's        Quarter Pounder® with Cheese      510               230
## 6 McDonald's Double Quarter Pounder® with Cheese      740               380
##   Total.Fat..g. Saturated.Fat..g. Trans.Fat..g. Cholesterol..mg. Sodium...mg.
## 1             9               3.5           0.5                25          520
## 2            12               6.0           0.5                40          750
## 3            23              11.0           1.5                80         1150
## 4            19               8.0           1.0                65          920
## 5            26              12.0           1.5                90         1190
## 6            42              19.0           2.5               155         1380
##   Carbs..g. Fiber..g. Sugars..g. Protein..g. Weight.Watchers.Pnts
## 1        31         2          6          12                247.5
## 2        33         2          6          15                297.0
## 3        34         2          7          25                433.0
## 4        33         2          7          22                383.0
## 5        40         3          9          29                502.0
## 6        40         3          9          48                720.0
```

```
miss_value <- sum(is.na(data))
miss_value
```

```
## [1] 1298
```

cause too many columns have missing values and three are different products, so i think we should not use any mid,mean to replace the NA value.

check the correlation

```
correlation_matrix <- cor(data[, 3:14], use = "complete.obs")

calories_correlation <- correlation_matrix["Calories", ]
print(calories_correlation)
```

```
##             Calories    Calories.from.Fat        Total.Fat..g.
##            1.0000000            0.8564877            0.8560343
##    Saturated.Fat..g.       Trans.Fat..g.     Cholesterol..mg.
##            0.8538338            0.6584781            0.6672815
##         Sodium...mg.           Carbs..g.            Fiber..g.
##            0.7211498            0.6885597            0.4848605
##           Sugars..g.         Protein..g. Weight.Watchers.Pnts
##            0.3151264            0.7945189            0.9905870
```

Calories vs Total Fat

```
ggplot(data, aes(x = Total.Fat..g., y = Calories)) +
  geom_point(alpha = 0.5, color = "blue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Scatter Plot of Calories vs Total Fat (Filtered Data)", x = "Total Fat (g)", y = "Calori
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 68 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 68 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

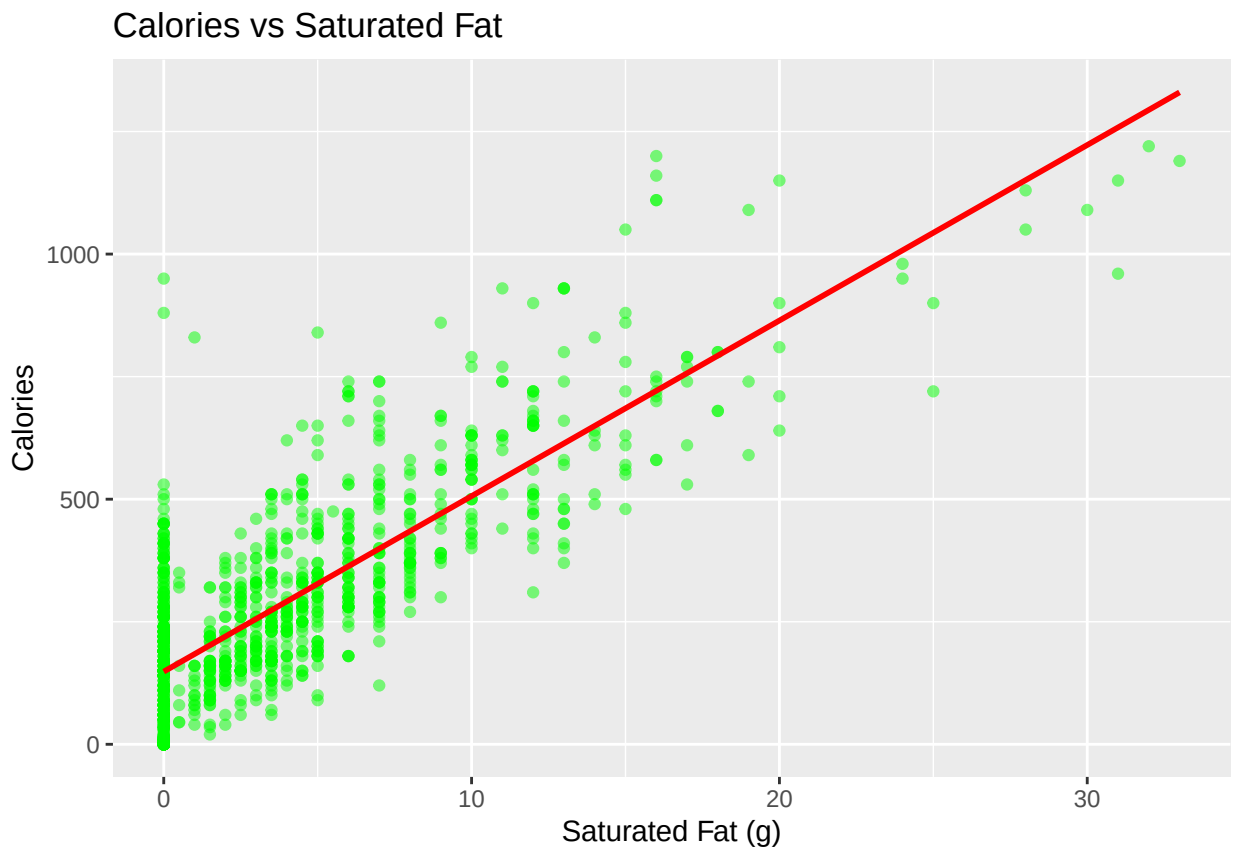
Calories vs Saturated Fat

```
ggplot(data, aes(x = Saturated.Fat..g., y = Calories)) +
  geom_point(alpha = 0.5, color = "green") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Calories vs Saturated Fat", x = "Saturated Fat (g)", y = "Calories")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 68 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 68 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



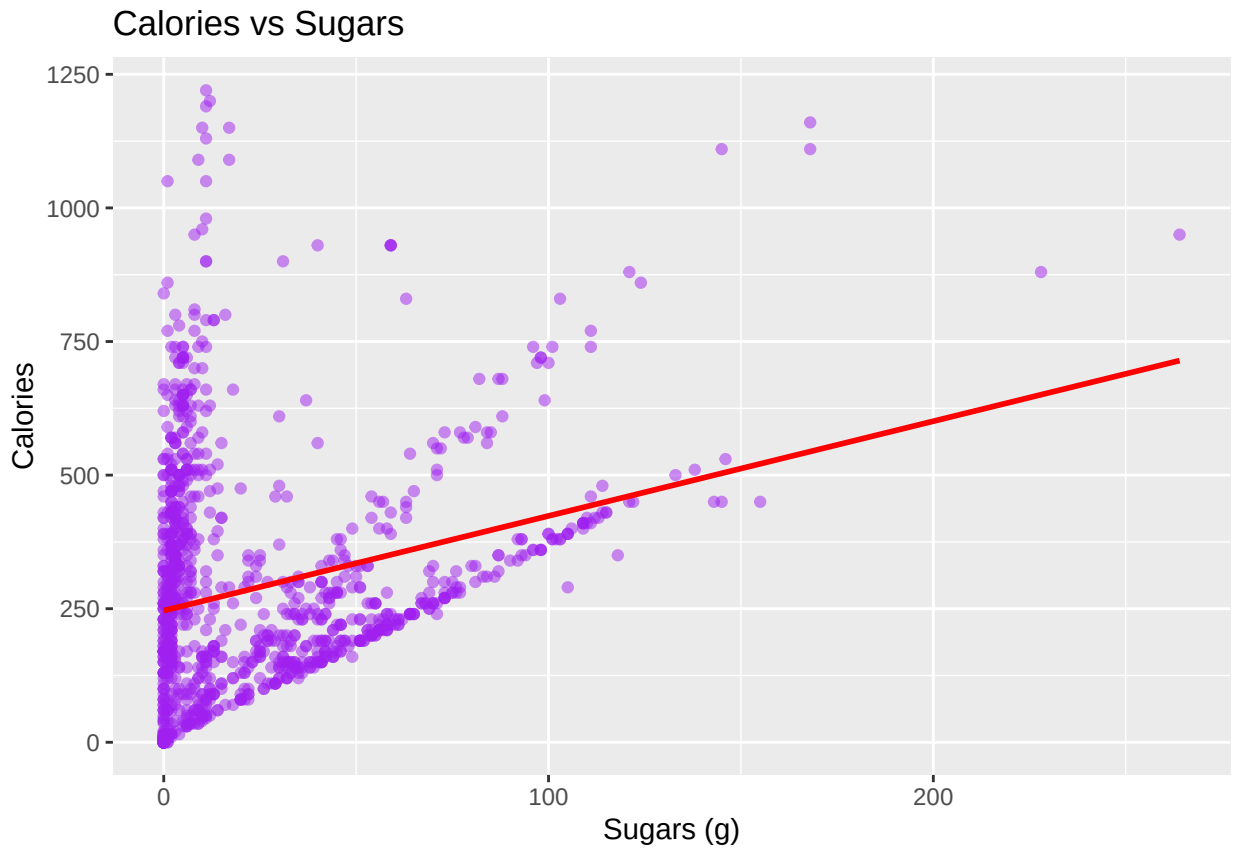
Calories vs Sugars

```
ggplot(data, aes(x = Sugars..g., y = Calories)) +
  geom_point(alpha = 0.5, color = "purple") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Calories vs Sugars", x = "Sugars (g)", y = "Calories")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 29 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 29 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Calories vs Protein

```
ggplot(data, aes(x = Protein..g., y = Calories)) +
  geom_point(alpha = 0.5, color = "orange") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Calories vs Protein", x = "Protein (g)", y = "Calories")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 68 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 68 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

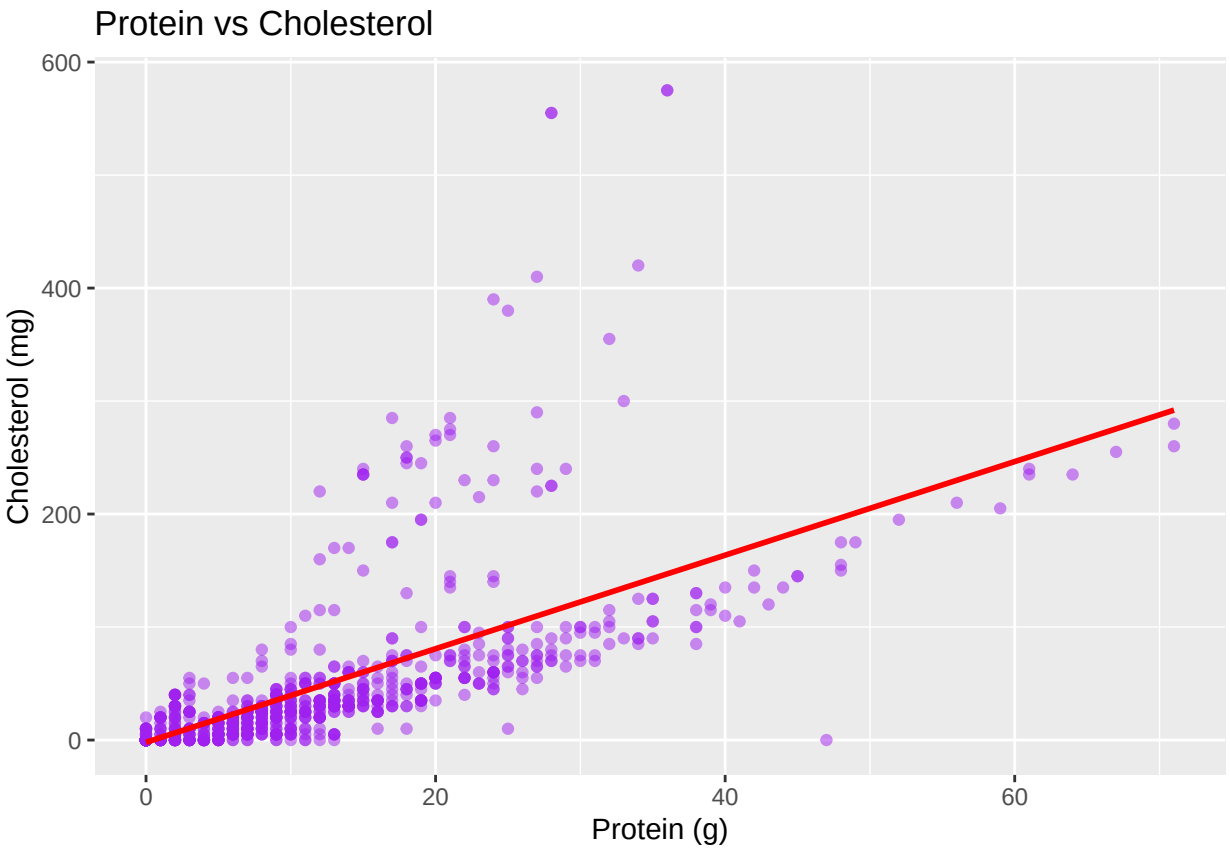
Protein vs Cholesterol

```
ggplot(data, aes(x = Protein..g., y = Cholesterol..mg.)) +
  geom_point(alpha = 0.5, color = "purple") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Protein vs Cholesterol", x = "Protein (g)", y = "Cholesterol (mg)")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 82 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 82 rows containing missing values or values outside the scale range
## (`geom_point()`).
```


Protein vs Cholesterol
600
400
200
0
Cholesterol (mg)
0
20
40
60
Protein (g)