MATH 3070 Lab Project 6

Prachi Aswani

October 3, 2024

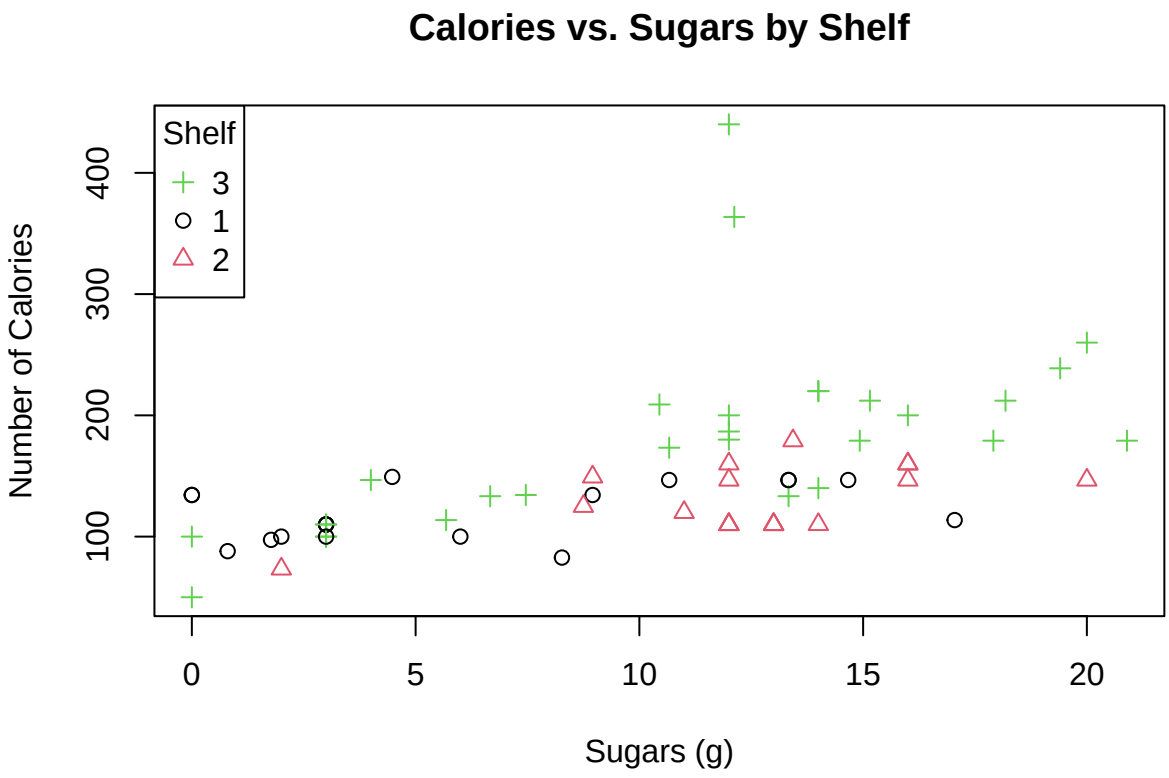
Contents

Remember: I expect to see commentary either in the text, in the code with comments created using **#**, *or (preferably) both!* ***Failing to do so may result in lost points!***

Problem 1 (Verzani problem 5.1)

For the `UScereal` *(**MASS**) data set, create a scatter plot of* `calories` *modeled by* `sugars` *using the* `shelf` *variable to create different plot characters. Add a legend to indicate the shelf number. Is there any patterns? (Use base R plotting for this problem.)*

```
# Load the necessary library and dataset
library(MASS)
data("UScereal")

# Create a scatter plot of calories vs. sugars, using 'shelf' to differentiate plot characters
with(UScereal, {
  plot(sugars, calories,
       pch = as.numeric(shelf),  # Set different plot characters based on shelf
       col = as.numeric(shelf),  # Color points by shelf too
       xlab = "Sugars (g)",
       ylab = "Number of Calories",
       main = "Calories vs. Sugars by Shelf"
  )

  # Add a legend to identify which shelf corresponds to each plot character
  legend("topleft",
         legend = unique(shelf),
         pch = unique(as.numeric(shelf)),
         col = unique(as.numeric(shelf)),
         title = "Shelf")
})
```



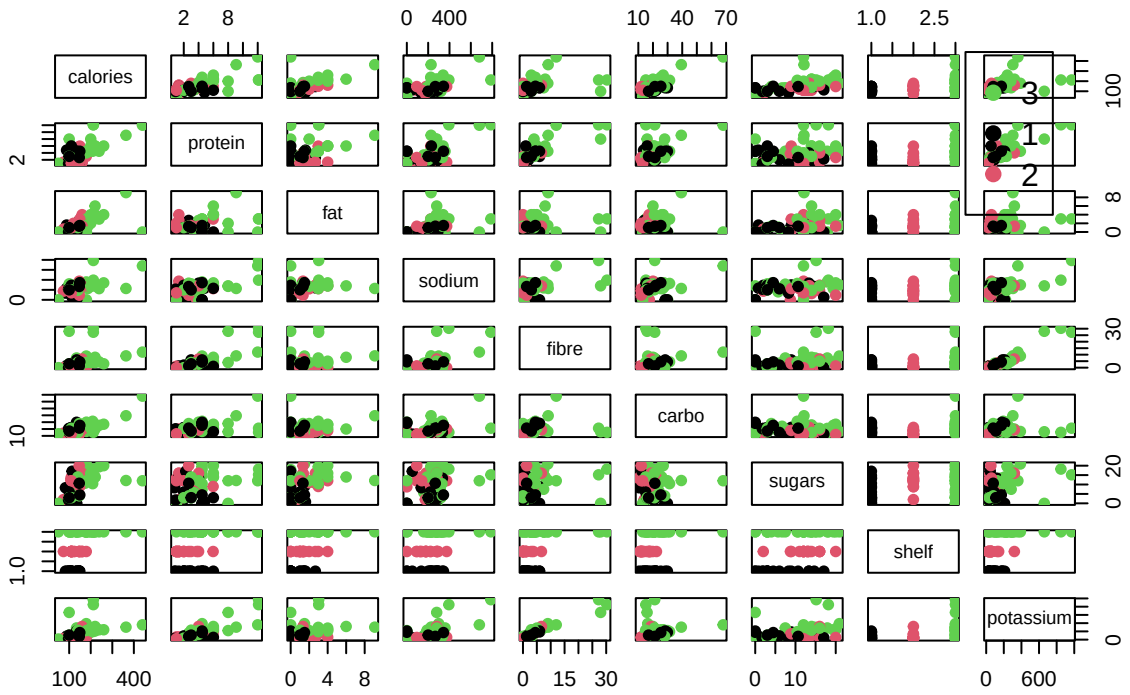

```
# It looks as though this graph tells the story that the more sugar a ceral has,the more likely
# it is to have slightly higher calories
```

Problem 2 (Verzani problem 5.3)

For the data set `UScereal` *(**MASS**) make a pairs plot of the numeric variables. Which correlation looks larger: fat and calories or fat and sugars?*

```
# Load the necessary library and dataset
library(MASS)
data("UScereal")

# Select only the numeric columns from the UScereal dataset
numeric_vars <- UScereal[sapply(UScereal, is.numeric)]

# Create the pairs plot, coloring points based on the shelf variable
pairs(numeric_vars,
      main = "Pairs Plot of Numeric Variables in UScereal (Colored by Shelf)",
      col = UScereal$shelf,  # Color points by the shelf variable
      pch = 19)              # Use solid points for clarity

# Add a legend to indicate which color corresponds to which shelf
legend("topright",
       legend = unique(UScereal$shelf),  # Unique shelf numbers
```

```
       col = unique(UScereal$shelf),      # Matching colors for each shelf
       pch = 19)                          # Solid points in legend
```




```
# It seems that there is some positive relationship between fat and calories.
# As the fat content increases, the calories tend to increase as well.
#The spread of points generally moves upward as fat increases.

# There appears to be little to no clear relationship between fat and sugars.
# The points are spread widely and don't seem to form any noticeable pattern,
# suggesting a weak or non-existent correlation.

# Fat and Calories have a stronger correlation compared to Fat and Sugars.
# There is a more consistent upward trend between fat and calories,
# whereas the relationship between fat and sugars is much more scattered and unclear.
```

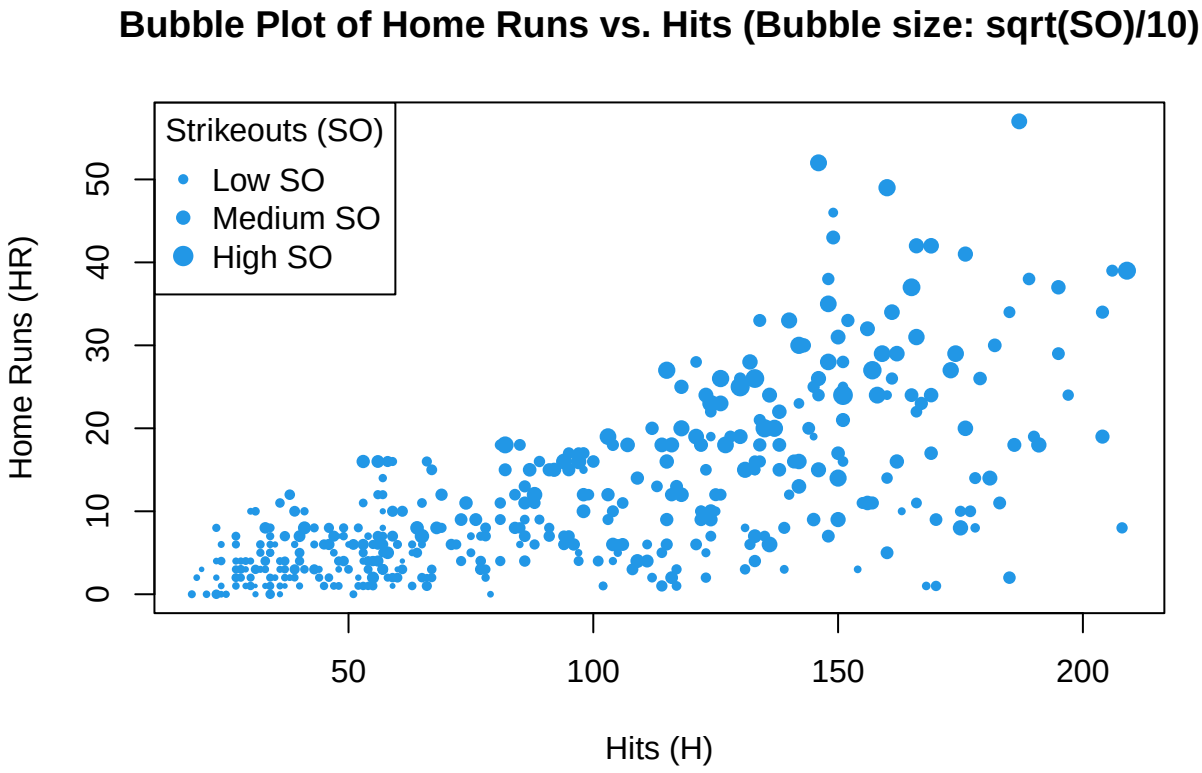
Problem 3 (Verzani problem 5.4)

For the data set `batting` (***UsingR***) *make a bubble plot of home runs hit (*`HR`*) modeled by hits (*`H`*) where the scale factor for each point is given by* `sqrt(SO)/10`*. Is there any story to be told by the size of the points? (You must use base R plotting for this problem.)*

```
# Load the necessary library and dataset
library(UsingR)
```

```
## Warning: package 'UsingR' was built under R version 4.3.3

## Loading required package: HistData

## Warning: package 'HistData' was built under R version 4.3.3

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
data("batting")

# Create the bubble plot using the formula syntax
plot(HR ~ H, data = batting, pch = 16, col = 4,
     cex = sqrt(batting$SO)/10,                    # Bubble size based on sqrt(SO)/10
     xlab = "Hits (H)", ylab = "Home Runs (HR)",   # Axis labels
     main = "Bubble Plot of Home Runs vs. Hits (Bubble size: sqrt(SO)/10)")  # Title

# Add a legend for bubble sizes (approximate scaling values)
legend("topleft",
       legend = c("Low SO", "Medium SO", "High SO"),
       pch = 16,
       pt.cex = c(sqrt(50)/10, sqrt(100)/10, sqrt(200)/10),  # Approximate SO values
       col = 4,
       title = "Strikeouts (SO)")
```




```
#It looks like the story to be told is that the more hits
# a player gets,correlates to the higher number of homeruns he will hit
```