advanced_visualization_with_r_part_1_exercises

Exercise 1

==========

Question 1

Read the fast_food_data.csv into a dataset named "fast_food_data".

Set both the `header` and `stringsAsFactors` arguments equal to TRUE.

Subset the data set to be named "fast_food_subset" and include columns 3, 5, 6, 10, 11.

Then rename those columns "type", "calories","totfat","carbs", & "sugars".

Answer:

```
fast_food_data = read.csv("data/fast_food_data.csv",header = TRUE, stringsAsFactors = TRUE)
fast_food_subset = fast_food_data[,c(3,5,6,10,11)]
head(fast_food_subset)
```

```
##     Type Calories Total.Fat..g. Carbs..g. Sugars..g.
## 1 Burger      240             8        32          6
## 2 Burger      290            11        33          7
## 3 Burger      530            27        47          9
## 4 Burger      520            26        41         10
## 5 Burger      720            40        51         14
## 6 Burger      750            43        42         10
```

```
colnames(fast_food_subset) <- c("type", "calories","totfat","carbs","sugars")
```

Question 2

Create a dataset `fast_food_num`, which consists of only the numeric variables from fast_food_subset. Hint: Drop the 'type' column.

Then retrieve the number of columns from `fast_food_num` and store it in the variable `num_col`.

Sample `num_col` number of colors from `colors` and store it in `color_sam`.

Make sure to set the seed to 2 before sampling.

What four colors did sample choose?

Answer: May vary. Our answer is "lightgray", "lavenderblush4", "grey12", and "grey88"

```
fast_food_num = fast_food_subset[,-c(1)]
num_col = ncol(fast_food_num)
set.seed(2)
color_sam = sample(colors(),size = num_col)
color_sam
```

```
## [1] "lightgray"      "lavenderblush4" "grey12"         "grey88"
```

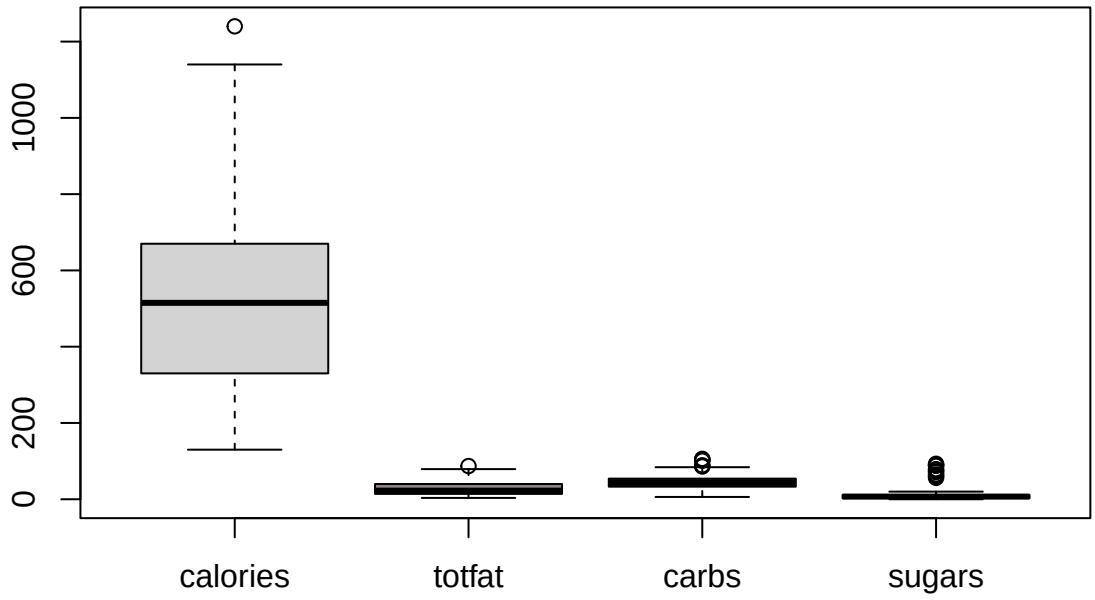
Question 3

Make a boxplot of the variables in `fast_food_num` using the colors stored in `color_sam`.

Which variable has the widest range?

Answer: Calories

```
boxplot(fast_food_num, col = color_sam)
```



Question 4

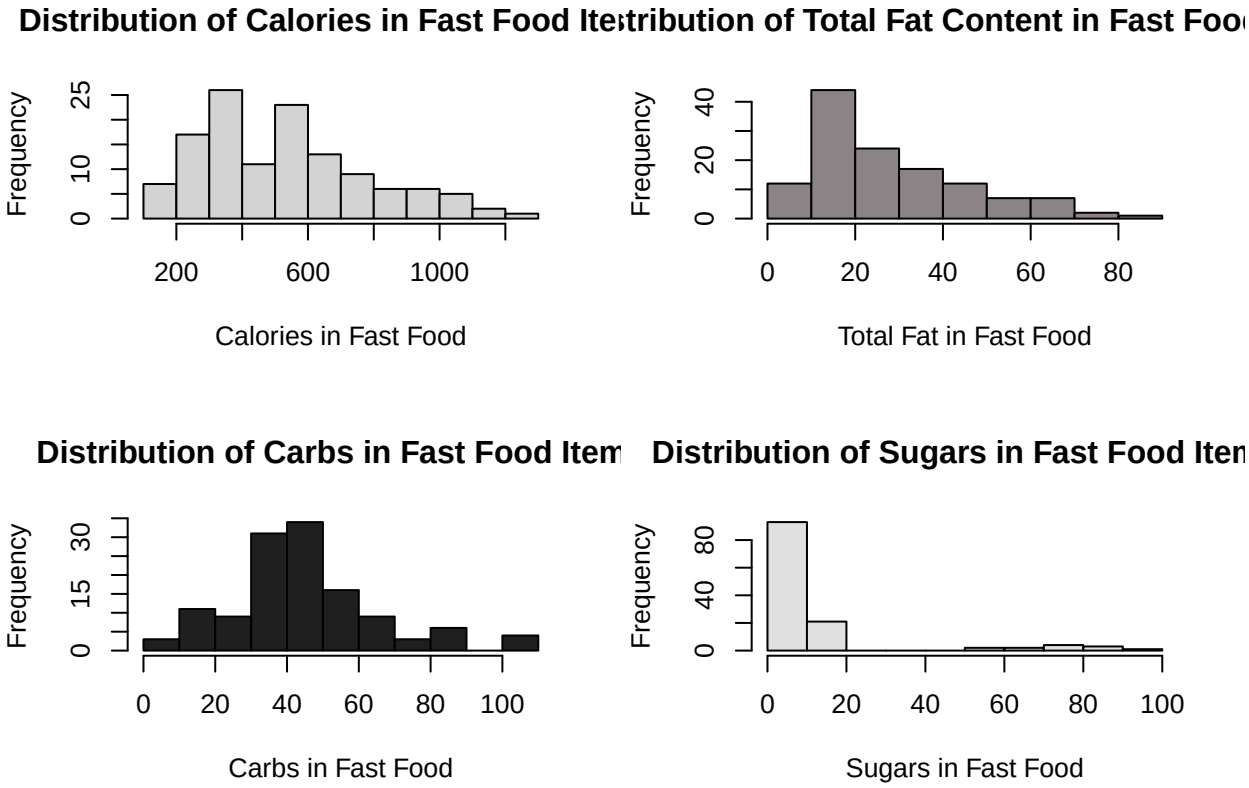
Create a 2x2 grid of histograms of all 4 variables in `fast_food_num`, using colors in `color_sam`.

appropriately labeled with xlabel and title for each.

Answer:

```
par(mfrow = c(2, 2))
hist(fast_food_num$calories,
     col = color_sam[1],
     xlab = "Calories in Fast Food",
     main = "Distribution of Calories in Fast Food Items")
hist(fast_food_num$totfat,
     col = color_sam[2],
     xlab = "Total Fat in Fast Food",
     main = "Distribution of Total Fat Content in Fast Food Items")
hist(fast_food_num$carbs,
     col = color_sam[3],
     xlab = "Carbs in Fast Food",
     main = "Distribution of Carbs in Fast Food Items")
hist(fast_food_num$sugars,
```

```
     col = color_sam[4],
     xlab = "Sugars in Fast Food",
     main = "Distribution of Sugars in Fast Food Items")
```



Question 5

Reset the grid by running par(mfrow = c(1, 1)).

Begin with plotting total fat against carbohydrate. Have total fat be on the x-axis and carbohydrate on the y-axis.

Use the column index to specify the variables. Create appropriate labels for x and y axes and title.

Fill in with triangle symbol(set pch to 17) and color "salmon2".

Answer:

```
par(mfrow = c(1, 1))
plot(x = fast_food_num$totfat,
     y = fast_food_num$carbs,
     xlab = "Total Fat",
```

```
     ylab = "Total Carb",
     main = "Total fat against carbohydrate",
     pch = 17,
     col = "salmon2")
```