

V506 Fall 24 Lab 5 In Class Exercise

Solutions to the Class Exercise

1. Load the required packages for the analysis: dplyr, ggplot, descriptr. Set your working directory or create a project. Either way, confirm that you are working from your working directory for this project.

```
# Clean the environment
rm(list=ls())

# Set your working directory
setwd("/Users/luisenriquenavarro/Library/CloudStorage/OneDrive-IndianaUniversity/V506/Fall24/Lab5")

# Packages Required for the session
library(pacman)
p_load(dplyr, ggplot2, rmarkdown, rio, here, descriptr)
```

2. Load the fastfood data and save it as a tibble object in R.

```
fastfood <- rio::import(file = "fastfood.csv", header = TRUE) %>% tibble()
fastfood %>% head(n=10)
```

```
## # A tibble: 10 x 17
##   restaurant item      calories cal_fat total_fat sat_fat trans_fat cholesterol
##   <chr>      <chr>      <int>  <int>    <int>  <dbl>    <dbl>        <int>
## 1 Mcdonalds Artisan ~    380    60      7      2      0          95
## 2 Mcdonalds Single B~    840   410     45     17     1.5        130
## 3 Mcdonalds Double B~   1130   600     67     27      3         220
## 4 Mcdonalds Grilled ~    750   280     31     10     0.5        155
## 5 Mcdonalds Crispy B~    920   410     45     12     0.5        120
## 6 Mcdonalds Big Mac     540   250     28     10      1         80
## 7 Mcdonalds Cheesebu~    300   100     12      5     0.5         40
## 8 Mcdonalds Classic ~    510   210     24      4      0         65
## 9 Mcdonalds Double C~    430   190     21     11      1         85
## 10 Mcdonalds Double Q~    770   400     45     21     2.5        175
## # i 9 more variables: sodium <int>, total_carb <int>, fiber <int>, sugar <int>,
## #   protein <int>, vit_a <int>, vit_c <int>, calcium <int>, salad <chr>
```

3. Explore the Data

517 observations from 17 variables. Each observation represents an item from the menu of a fast food restaurant.

```
fastfood %>% str()
```

```
## tibble [515 x 17] (S3: tbl_df/tbl/data.frame)
## $ restaurant : chr [1:515] "Mcdonalds" "Mcdonalds" "Mcdonalds" "Mcdonalds" ...
## $ item       : chr [1:515] "Artisan Grilled Chicken Sandwich" "Single Bacon Smokehouse Burger" "Doul
## $ calories   : int [1:515] 380 840 1130 750 920 540 300 510 430 770 ...
## $ cal_fat    : int [1:515] 60 410 600 280 410 250 100 210 190 400 ...
## $ total_fat  : int [1:515] 7 45 67 31 45 28 12 24 21 45 ...
## $ sat_fat    : num [1:515] 2 17 27 10 12 10 5 4 11 21 ...
## $ trans_fat  : num [1:515] 0 1.5 3 0.5 0.5 1 0.5 0 1 2.5 ...
## $ cholesterol: int [1:515] 95 130 220 155 120 80 40 65 85 175 ...
## $ sodium     : int [1:515] 1110 1580 1920 1940 1980 950 680 1040 1040 1290 ...
## $ total_carb : int [1:515] 44 62 63 62 81 46 33 49 35 42 ...
## $ fiber      : int [1:515] 3 2 3 2 4 3 2 3 2 3 ...
## $ sugar      : int [1:515] 11 18 18 18 18 9 7 6 7 10 ...
## $ protein    : int [1:515] 37 46 70 55 46 25 15 25 25 51 ...
## $ vit_a      : int [1:515] 4 6 10 6 6 10 10 0 20 20 ...
## $ vit_c      : int [1:515] 20 20 20 25 20 2 2 4 4 6 ...
## $ calcium    : int [1:515] 20 20 50 20 20 15 10 2 15 20 ...
## $ salad     : chr [1:515] "Other" "Other" "Other" "Other" ...
```

```
fastfood %>% summary()
```

```
##      restaurant      item      calories      cal_fat
## Length:515      Length:515      Min.   : 20.0      Min.   :  0.0
## Class :character Class :character 1st Qu.: 330.0      1st Qu.: 120.0
## Mode  :character Mode  :character Median : 490.0      Median : 210.0
##                                     Mean  : 530.9      Mean  : 238.8
##                                     3rd Qu.: 690.0      3rd Qu.: 310.0
##                                     Max.   :2430.0      Max.   :1270.0
##
##      total_fat      sat_fat      trans_fat      cholesterol
## Min.   :  0.00      Min.   : 0.000      Min.   :0.000      Min.   :  0.00
## 1st Qu.: 14.00      1st Qu.: 4.000      1st Qu.:0.000      1st Qu.: 35.00
## Median : 23.00      Median : 7.000      Median :0.000      Median : 60.00
## Mean   : 26.59      Mean   : 8.153      Mean   :0.465      Mean   : 72.46
## 3rd Qu.: 35.00      3rd Qu.:11.000      3rd Qu.:1.000      3rd Qu.: 95.00
## Max.   :141.00      Max.   :47.000      Max.   :8.000      Max.   :805.00
##
##      sodium      total_carb      fiber      sugar
## Min.   : 15      Min.   :  0.00      Min.   : 0.000      Min.   :  0.000
## 1st Qu.: 800      1st Qu.: 28.50      1st Qu.: 2.000      1st Qu.:  3.000
## Median :1110      Median : 44.00      Median : 3.000      Median :  6.000
## Mean   :1247      Mean   : 45.66      Mean   : 4.137      Mean   :  7.262
## 3rd Qu.:1550      3rd Qu.: 57.00      3rd Qu.: 5.000      3rd Qu.:  9.000
## Max.   :6080      Max.   :156.00      Max.   :17.000      Max.   :87.000
##
##                                     NA's   :12
##      protein      vit_a      vit_c      calcium
## Min.   :  1.00      Min.   :  0.00      Min.   :  0.00      Min.   :  0.00
## 1st Qu.: 16.00      1st Qu.:  4.00      1st Qu.:  4.00      1st Qu.:  8.00
## Median : 24.50      Median : 10.00      Median : 10.00      Median : 20.00
## Mean   : 27.89      Mean   : 18.86      Mean   : 20.17      Mean   : 24.85
## 3rd Qu.: 36.00      3rd Qu.: 20.00      3rd Qu.: 30.00      3rd Qu.: 30.00
```

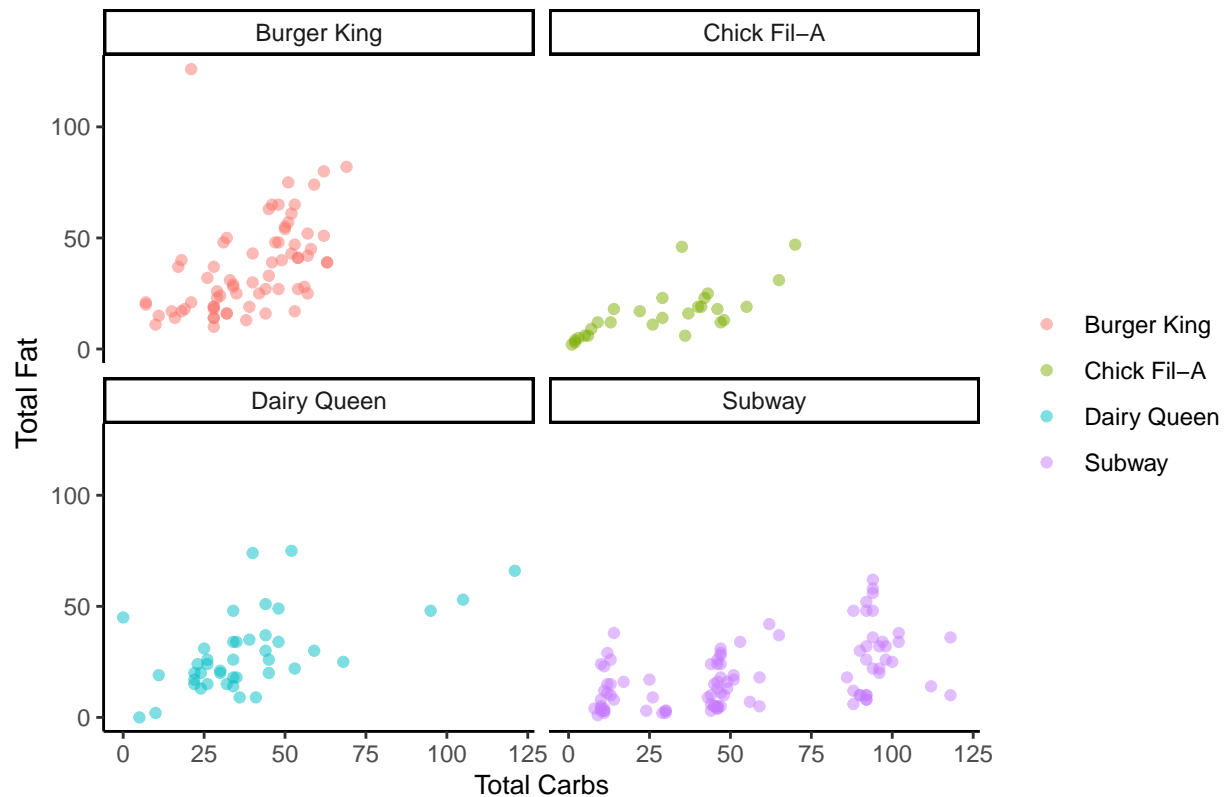
```
## Max.      :186.00   Max.      :180.00   Max.      :400.00   Max.      :290.00
## NA's      :1       NA's      :214      NA's      :210      NA's      :210
## salad
## Length:515
## Class :character
## Mode  :character
##
##
##
##
```

```
#ds_auto_summary_stats(fastfood)
```

4 Scatter Plot Carbs and Fat

```
fastfood %>% filter(restaurant == "Chick Fil-A" | restaurant == "Burger King" | restaurant == "Subway")
  ggplot(mapping = aes(x = total_carb, y = total_fat, fill = restaurant, color = restaurant))
    geom_point(alpha = 0.5)+
    facet_wrap(~restaurant)+
    theme_classic()+
    labs(x = "Total Carbs", y = "Total Fat", title = "Correlation of Carbs and Fat in t
    theme(plot.title = element_text(angle = 0, size = 12, face = "bold"))+
    theme(axis.title.x = element_text(angle = 0, size = 10))+
    theme(axis.title.y = element_text(angle = 0, size = 10))+
    theme(legend.title = element_blank())
```

Correlation of Carbs and Fat in the Menu



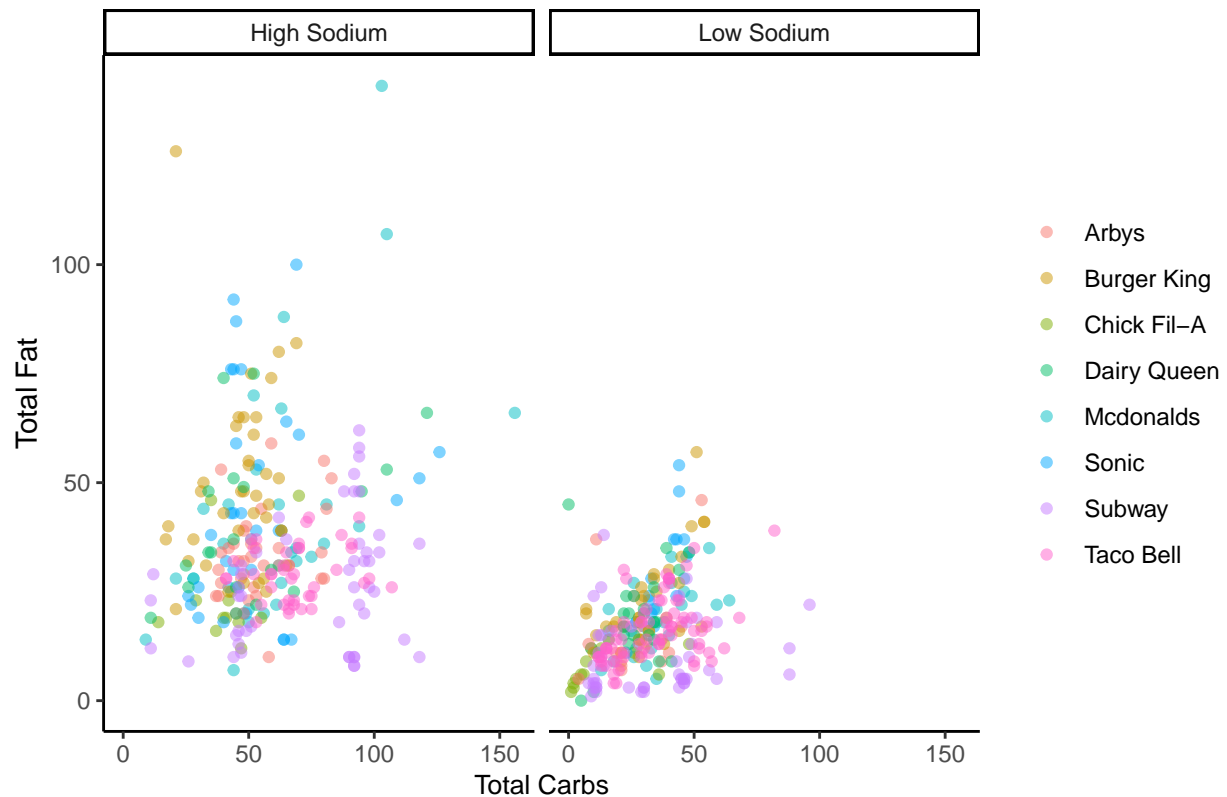
5. Scatter Sodium

```
median_sodium <- median(fastfood$sodium, na.rm = TRUE)

fastfood1 <- fastfood %>% mutate(high_sodium = ifelse(sodium >= median_sodium, "High Sodium", "Low Sodium"))

fastfood1 %>% ggplot(mapping = aes(x = total_carb, y = total_fat, fill = restaurant, color = restaurant)) +
  geom_point(alpha = 0.5) +
  facet_wrap(~high_sodium) +
  theme_classic() +
  labs(x = "Total Carbs", y = "Total Fat", title = "Correlation of Carbs and Fat in the Menu") +
  theme(plot.title = element_text(angle = 0, size = 12, face = "bold")) +
  theme(axis.title.x = element_text(angle = 0, size = 10)) +
  theme(axis.title.y = element_text(angle = 0, size = 10)) +
  theme(legend.title = element_blank())
```

Correlation of Carbs and Fat in the Menu



6. Hypothesis Test Total Fat

```
data1 <- fastfood1 %>% filter (restaurant == "Chick Fil-A") %>% select(total_fat)
data2 <- fastfood1 %>% filter (restaurant == "Mcdonalds") %>% select(total_fat)

t.test(data1, data2)
```

```
##
## Welch Two Sample t-test
##
## data: data1 and data2
## t = -3.9983, df = 81.992, p-value = 0.0001389
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -23.449841 -7.867898
## sample estimates:
## mean of x mean of y
## 16.14815 31.80702
```

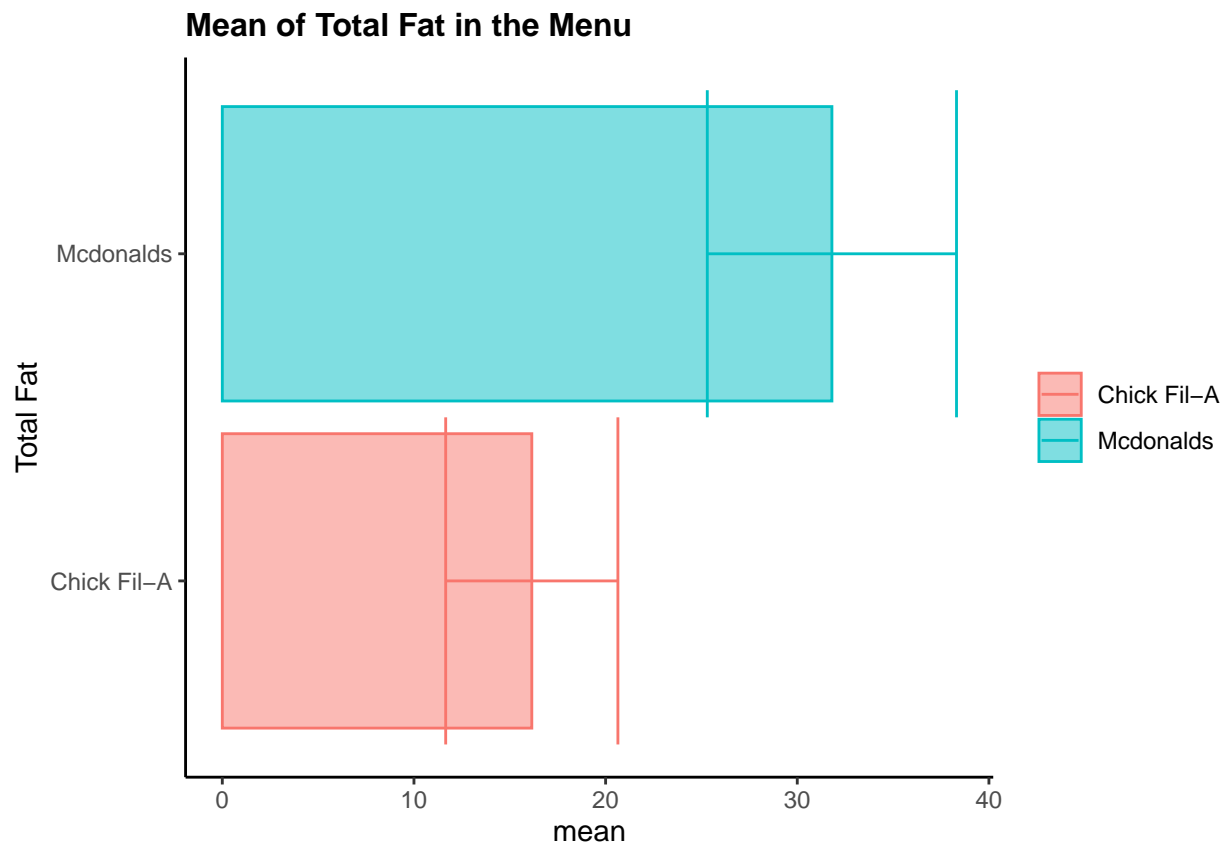
```
total_fat_restaurant <- fastfood1 %>% filter(restaurant == "Chick Fil-A" | restaurant == "Mcdonalds") %>%
  group_by(restaurant) %>%
  summarize(mean = mean(total_fat, na.rm = TRUE),
            sd = sd(total_fat, na.rm = TRUE),
```

```

      obs = n()) %>%
      mutate(mean_se = sd/sqrt(obs)) %>%
      mutate(margin_error = qt(0.975,df=obs)*mean_se) %>%
      mutate(ci_low = mean - margin_error) %>%
      mutate(ci_high = mean + margin_error)

total_fat_restaurant %>% ggplot(mapping = aes(x = reorder(restaurant, mean), y = mean, fill = restaurant)) +
  geom_bar(stat = "identity", alpha = 0.5) +
  geom_errorbar(aes(ymin = ci_low, ymax = ci_high), width=1)+
  coord_flip()+
  labs(x = "Total Fat", title = "Mean of Total Fat in the Menu") +
  theme_classic()+
  theme(plot.title = element_text(angle = 0, size = 12, face = "bold"))+
  theme(legend.title = element_blank())

```



7. Hypothesis Test Sodium

```

data1 <- fastfood1 %>% filter (high_sodium == "High Sodium") %>% select(cholesterol)
data2 <- fastfood1 %>% filter (high_sodium == "Low Sodium") %>% select(cholesterol)

t.test(data1, data2)

```

##

```
## Welch Two Sample t-test
##
## data: data1 and data2
## t = 11.381, df = 326.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  46.59045 66.06264
## sample estimates:
## mean of x mean of y
## 100.34615  44.01961
```

```
cholesterol_restaurant <- fastfood1 %>% group_by(high_sodium) %>%
  summarize(mean = mean(cholesterol, na.rm = TRUE),
            sd = sd(cholesterol, na.rm = TRUE),
            obs = n()) %>%
  mutate(mean_se = sd/sqrt(obs)) %>%
  mutate(margin_error = qt(0.975,df=obs)*mean_se) %>%
  mutate(ci_low = mean - margin_error) %>%
  mutate(ci_high = mean + margin_error)

cholesterol_restaurant %>% ggplot(mapping = aes(x = reorder(high_sodium, mean), y = mean, fill = high_sodium)) +
  geom_bar(stat = "identity", alpha = 0.5) +
  geom_errorbar(aes(ymin = ci_low, ymax = ci_high), width=1)+
  coord_flip()+
  labs(x = "Cholesterol", title = "Mean of Cholesterol in the Menu") +
  theme_classic()+
  theme(plot.title = element_text(angle = 0, size = 12, face = "bold"))+
  theme(legend.title = element_blank())
```

