

Data 211 Final Project

Patrick Meade

2025-11-16

```
#Libraries
```

```
library(ggplot2)
library(readxl)

## Warning: package 'readxl' was built under R version 4.5.2

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.5.2

## Warning: package 'tidyr' was built under R version 4.5.2

## Warning: package 'readr' was built under R version 4.5.2

## Warning: package 'purrr' was built under R version 4.5.2

## Warning: package 'forcats' was built under R version 4.5.2

## Warning: package 'lubridate' was built under R version 4.5.2

## — Attaching core tidyverse packages ————— tidyverse
2.0.0 —
## ✓ dplyr     1.1.4      ✓ readr     2.1.5
## ✓ forcats   1.0.1      ✓ stringr   1.5.2
## ✓ lubridate 1.9.4      ✓ tibble    3.3.0
## ✓ purrr     1.1.0      ✓ tidyr    1.3.1
## — Conflicts —
tidyverse_conflicts() —
## X dplyr::filter() masks stats::filter()
## X dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
#Load data into a new dataset
```

```
Calories<-read_excel("C:/Users/Patrick/Desktop/Prog/R/Data 211 Final Project
Data Collection .xlsx")
```

```
#Change format for graph and table
```

```
NewCalories<-Calories%>%
  rename(`My Average` = `Ultra-Processed`)%>%
  mutate(`Day of Data`=c(1:14))
```

Initial Statistical Summary

```
summary(Calories)
```

```
## Day of Data Total Calories Ultra-Processed Unprocessed
## Min. :2025-11-02 00:00:00 Min. :1935 Min. : 774 Min. : 790
## 1st Qu.:2025-11-05 06:00:00 1st Qu.:2116 1st Qu.: 920 1st Qu.:1012
## Median :2025-11-08 12:00:00 Median :2216 Median :1068 Median :1194
## Mean   :2025-11-08 12:00:00 Mean   :2242 Mean   :1063 Mean   :1179
## 3rd Qu.:2025-11-11 18:00:00 3rd Qu.:2392 3rd Qu.:1214 3rd Qu.:1259
## Max.   :2025-11-15 00:00:00 Max.   :2580 Max.   :1315 Max.   :1610
## Percent of Ultra-Processed Calories
## Min. :36.00
## 1st Qu.:40.50
## Median :47.00
## Mean   :47.64
## 3rd Qu.:53.75
## Max.   :60.00

sd(Calories$`Percent of Ultra-Processed Calories`)

## [1] 7.830975

range(Calories$`Percent of Ultra-Processed Calories`)

## [1] 36 60
```

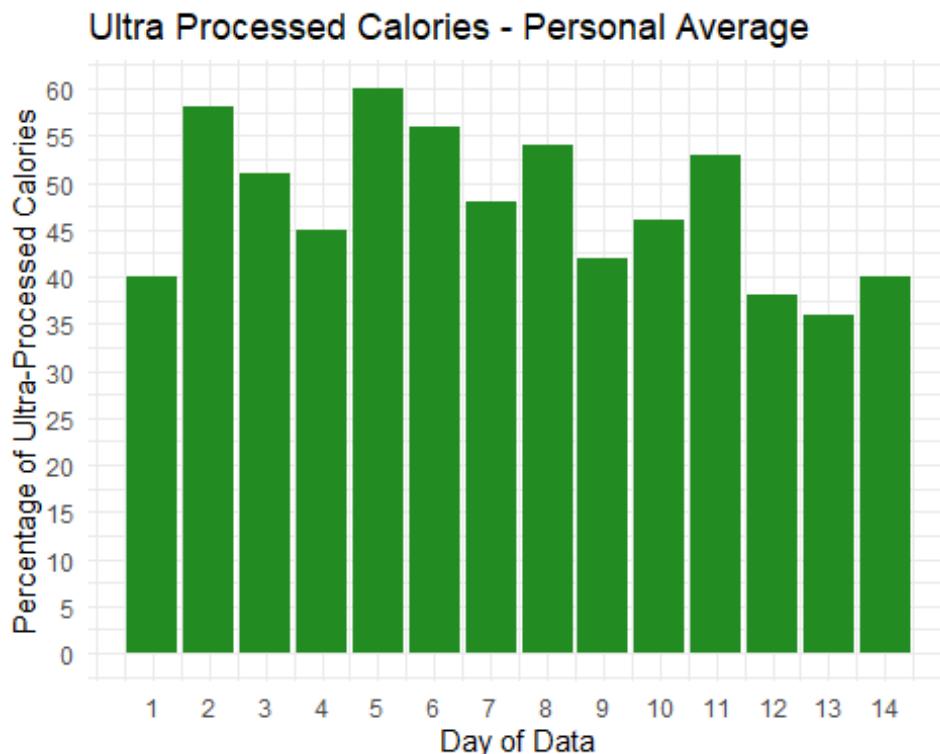
Table

```
PercentTable<-NewCalories%>%
  select(`Day of Data`, `Percent of Ultra-Processed Calories`)
print(PercentTable)

## # A tibble: 14 × 2
##   `Day of Data` `Percent of Ultra-Processed Calories`
##       <int>                <dbl>
## 1 1                    40
## 2 2                    58
## 3 3                    51
## 4 4                    45
## 5 5                    60
## 6 6                    56
## 7 7                    48
## 8 8                    54
## 9 9                    42
## 10 10                  46
## 11 11                  53
## 12 12                  38
## 13 13                  36
## 14 14                  40
```

Visualization

```
ggplot(data=NewCalories,aes(x=`Day of Data`,y=`Percent of Ultra-Processed Calories`))+  
  geom_bar(stat="identity",fill="forestgreen") +  
  scale_x_continuous(breaks=c(1:14)) +  
  scale_y_continuous(breaks=seq(0,60,by=5)) +  
  labs(title="Ultra Processed Calories - Personal Average",  
       x="Day of Data",  
       y="Percentage of Ultra-Processed Calories") +  
  theme_minimal()
```



Statistical Analysis

p_m :percentage of my ultra-processed calories.

p_a :percentage of American average ultra-processed calories.

$H_0: p_m = p_a$ vs. $H_1: p_m < p_a$

```
t_test_result<-t.test(NewCalories$`Percent of Ultra-Processed Calories`,mu=55,alternative="less")  
print(t_test_result)  
  
##  
## One Sample t-test
```

```
##  
## data: NewCalories$`Percent of Ultra-Processed Calories`  
## t = -3.5153, df = 13, p-value = 0.001901  
## alternative hypothesis: true mean is less than 55  
## 95 percent confidence interval:  
##       -Inf 51.34927  
## sample estimates:  
## mean of x  
## 47.64286
```

$p - value \approx 0.0019 < \alpha = 0.05$. Reject H_0 . So at 5% significance level, we have enough evidence to conclude that the percentage of my ultra-processed calories is less than the American average of 55%.