Untitled

Vrinda

2024-10-12

```r
library('ggvis')
library('tidyverse')
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4      v readr     2.1.5
## v forcats   1.0.0      v stringr   1.5.1
## v ggplot2   3.5.1      v tibble    3.2.1
## v lubridate 1.9.3      v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter()      masks stats::filter()
## x dplyr::lag()         masks stats::lag()
## x ggplot2::resolution() masks ggvis::resolution()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
library('ggplot2')
library('stats')
library('dplyr')
library('corrplot')
```

```
## corrplot 0.94 loaded
```

```r
library('mosaic')
```

```
## Registered S3 method overwritten by 'mosaic':
##    method                           from
##    fortify.SpatialPolygonsDataFrame ggplot2
##
## The 'mosaic' package masks several functions from core packages in order to add
## additional features.  The original behavior of these functions should not be affected by this.
##
## Attaching package: 'mosaic'
##
## The following object is masked from 'package:Matrix':
##
##     mean
##
## The following objects are masked from 'package:dplyr':
##
##     count, do, tally
```

```
## 
## The following object is masked from 'package:purrr':
## 
##      cross
## 
## The following object is masked from 'package:ggplot2':
## 
##      stat
## 
## The following objects are masked from 'package:ggvis':
## 
##      prop, props
## 
## The following objects are masked from 'package:stats':
## 
##      binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##      quantile, sd, t.test, var
## 
## The following objects are masked from 'package:base':
## 
##      max, mean, min, prod, range, sample, sum
```

```r
# read the file
fast_food= read.csv("C:\\Users\\Vrinda\\Downloads\\FastFoodNutritionScraped20241011V2.csv", header=T, na
```

State the hypothesis

1. Null Hypothesis(Ho) - Mean Trans Fat of the two companies is equal. i.e. $\mu1 = \mu2$

2. Alternative Hypothesis(H1) - Mean Trans fat of the two companies are unequal i.e. $\mu1 \neq \mu2$

```r
#visualisation to see how trans fat is different based on compnay
violin_plot <- ggplot(fast_food, aes(x = Company, y = Trans.Fat..g., fill = Company)) +
geom_violin(trim = FALSE, draw_quantiles = c(0.25, 0.5, 0.75), alpha = 0.7) +
geom_boxplot(width = 0.1, fill = "white", color = "black", outlier.shape = NA) +
labs(title = "Violin Plot of Trans Fat by Restaurant",
x = "Company",
y = "Trans Fat",
fill = "Company") +
theme_minimal()
print(violin_plot)
```

Violin Plot of Trans Fat by Restaurant
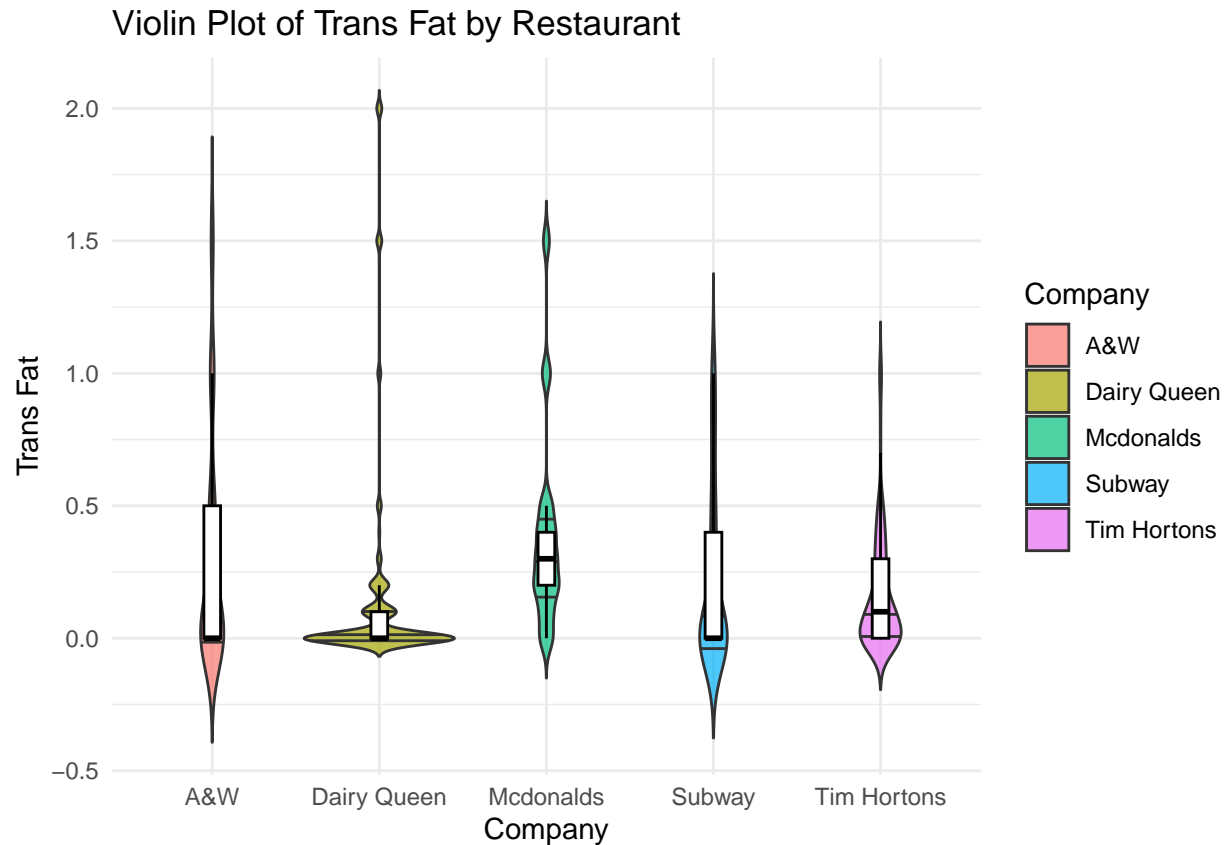

```
#splitting the data based on company wise
company1_data <- fast_food$Trans.Fat..g.[fast_food$Company == 'A&W']
company2_data <- fast_food$Trans.Fat..g.[fast_food$Company == 'Dairy Queen']
company3_data <- fast_food$Trans.Fat..g.[fast_food$Company == 'Mcdonalds']
company4_data <- fast_food$Trans.Fat..g.[fast_food$Company == 'Subway']
company5_data <- fast_food$Trans.Fat..g.[fast_food$Company == 'Tim Hortons']
```

AS the data is left skewed we can't use T-test as it is not normally distributed , so we are going with Bootstraping Distribution

```
#bootstrap test as the data is not normally distributed
n = 1000
b = do(n)*(mean(resample(company1_data,replace = TRUE))- mean(resample(company4_data,replace = TRUE)))

quantile(b$result,c(0.025,0.975))
```

```
##        2.5%       97.5%
## -0.03306907  0.20492643
```

As we can see from the confidence interval we have the lower quantile in negative which implies that the mean Of Trans Fat is greater in Subway rather than A&W.