Pedestrian vs Car Wait Time Analysis

Terra Pahlke

2022-11-02

Contents

Introduction

As a university student who commutes daily, I frequently encounter the challenge of navigating busy intersections as both a pedestrian and a driver. Over time, I observed that, as a pedestrian, my wait times at traffic signals often seemed significantly longer compared to those of vehicles. This observation led me to question whether pedestrians truly experience longer delays, or if the perception of extended wait times was influenced by factors such as traffic patterns or signal timings.

For my project, I aimed to investigate whether there is an actual difference in the duration of wait times for pedestrians versus vehicles at this intersection. Specifically, I sought to test whether pedestrians indeed experience longer delays when waiting for traffic signals to change, or if it is simply a subjective perception based on the time of day, traffic flow, or signal cycle.

Overview

Question: Does the duration of the traffic light cycle differ depending on whether a pedestrian presses the button or a car pulls up?

Variable of interest: Wait times at the chosen intersection.

Measuring: Using phone stopwatch. The time begins at either the push of the button or the stopping of the car over the pressure plate. The time ends when the light turns green (and walking is permitted)

- Treatment:

    - Pushing button
    - Car activating pressure plate

- Blocking factors:

    - Day of the week (Monday - Friday)
    - Time of day (Morning, Afternoon, Evening)

Assumptions:

- Right turn makes no difference

- Left turn activates the pressure plate and affects the lights

- Orthogonality between the types of days <- Assuming but will be checking

- Weather has no impact

- All data points in the week of November, if there is variability due to time of year/season we will not be seeing it

- If the walk button is not pushed the walk signal will NEVER activate, even when it should/could. Thus when a car activates the light (and the walk button is not pushed) the time measurement is until the light turns green and it 'would' be safe for a pedestrian to walk and had the button been pushed the signal would have activated.

Model:

This is a generalized complete block design (CBD) with 2 blocks (time and day), where time has 3 factors and day has 5 factors, and the treatment is button or car activated.

Model: $y_{ij} = \mu + \beta_i + \tau_j + \gamma_k + (\beta\tau)_{ij} + (\beta\gamma)_{ik} + (\tau\gamma)_{jk} + (\beta\tau\gamma)_{ijk} + \epsilon_{ijk}$

where: y_{ijk} represents the response for the kth treatment and i, jth block

μ represents the overall mean

β represents the blocking effect for experimental unit i, in this case the day of the week

γ represents the blocking effect for the experimental unit j, in this case the time of day

τ represents the treatment effect for experimental unit k, in this case push or car

$(\beta\tau)_{ij}$ represents the interaction effect between the day and time

$(\beta\gamma)_{ik}$ represents the interaction effect between the day and treatment

$(\tau\gamma)_{jk}$ represents the interaction effect between the time and treatment

$(\beta\tau\gamma)_{ijk}$ represents the interaction effect between the time, day and treatment

ϵ represents the experimental error

Number of collected data points: 30

Sources of variability:

- If I am conducting the push treatment and a car pulls up to make a left turn I can no longer use that data as it cannot be classified to push or car alone.

Affect: I have to redo that treatment

- Sometimes the cars do not pull up to the light correctly, thus missing the pressure plate and not properly activating the light change.

Affect: Wait time for the light to change will take longer than it 'should' have.

Control: If I were the car I could have controlled for them as I could be placing the car exactly where needed. I will not be since I do not have a car.

Hypotheses

H_0: All τ are the same vs. H_A : At least one τ is different

General set up of data collection:

Time	Mon	Tue	Wed	Thur	Fri
7:20am					
2:30pm					
9:30pm					

I will be collecting data for 5 days, 3 times a day with 2 treatments (randomly assigned order).

Design plan

Table 2: Design Plan

Day	Time	treatment 1	2	duration 1	2
Mon	Morning	1	2	NA	NA
Mon	Afternoon	2	1	NA	NA
Mon	Evening	2	1	NA	NA
Tue	Morning	2	1	NA	NA
Tue	Afternoon	2	1	NA	NA
Tue	Evening	2	1	NA	NA
Wed	Morning	1	2	NA	NA
Wed	Afternoon	1	2	NA	NA
Wed	Evening	1	2	NA	NA
Thu	Morning	1	2	NA	NA
Thu	Afternoon	2	1	NA	NA
Thu	Evening	2	1	NA	NA
Fri	Morning	1	2	NA	NA
Fri	Afternoon	2	1	NA	NA
Fri	Evening	1	2	NA	NA

Table 3: Collected data

Day	Time	Treatment	Duration
Mon	Morning	push	65.10
Mon	Morning	car	50.92
Mon	Afternoon	push	101.62
Mon	Afternoon	car	21.21
Mon	Evening	push	23.02
Mon	Evening	car	18.88
Tue	Morning	push	59.96
Tue	Morning	car	60.32
Tue	Afternoon	push	63.34
Tue	Afternoon	car	40.52
Tue	Evening	push	19.13
Tue	Evening	car	15.79
Wed	Morning	push	20.45
Wed	Morning	car	15.89
Wed	Afternoon	push	70.21
Wed	Afternoon	car	65.25
Wed	Evening	push	19.20
Wed	Evening	car	20.80
Thu	Morning	push	51.68
Thu	Morning	car	29.31
Thu	Afternoon	push	55.88
Thu	Afternoon	car	30.89
Thu	Evening	push	18.89
Thu	Evening	car	19.06
Fri	Morning	push	79.95
Fri	Morning	car	10.22
Fri	Afternoon	push	78.97

Day	Time	Treatment	Duration
Fri	Afternoon	car	44.35
Fri	Evening	push	18.79
Fri	Evening	car	20.87

Analysis

Assumption checking

Treatment vs. Duration(sec)
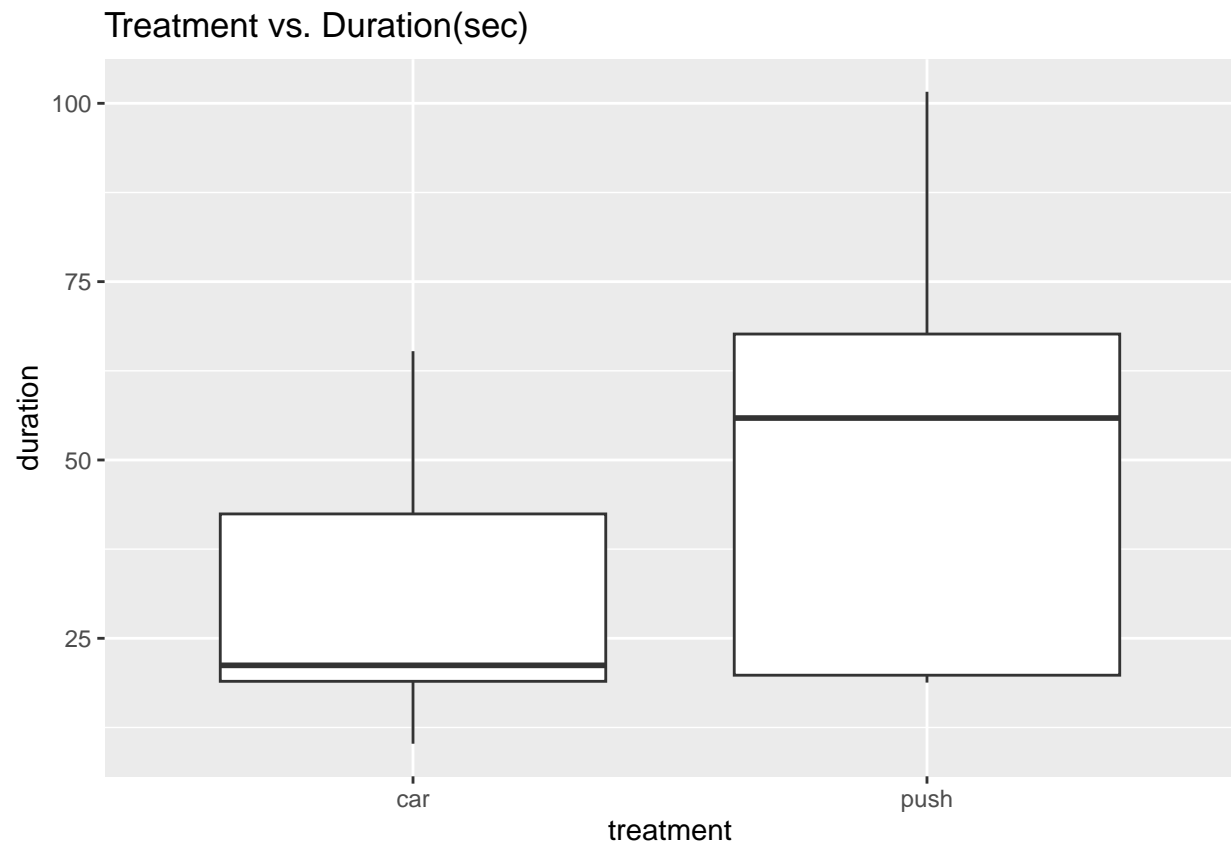
Model assumptions:

- independence(between and within): yes

- normality: yes, random

- constant variance: sort of (good enough)

- balanced: yes, there are the same number of data points in each section

Fit the model

```r
# entire model
mod1 <- aov(duration~.^3, data = data1)
summary(mod1)
```

```
##              Df Sum Sq Mean Sq
## day           4    691     173
## time          2   7381    3690
## treatment     1   2649    2649
```

```
## day:time            8   2331      291
## day:treatment       4   1211      303
## time:treatment      2   1388      694
## day:time:treatment  8   1978      247
```

```r
# double interaction model
mod1.1 <- aov(duration~.^2, data = data1)
summary(mod1.1)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## day                4    691     173   0.699 0.6141
## time               2   7381    3690  14.924 0.0020 **
## treatment          1   2649    2649  10.713 0.0113 *
## day:time           8   2331     291   1.178 0.4112
## day:treatment      4   1211     303   1.225 0.3728
## time:treatment     2   1388     694   2.808 0.1192
## Residuals          8   1978     247
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# main effect model
mod1.2 <- aov(duration~., data = data1)
summary(mod1.2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## day           4    691     173   0.550 0.700912
## time          2   7381    3690  11.752 0.000337 ***
## treatment     1   2649    2649   8.436 0.008220 **
## Residuals    22   6908     314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Due to the number of observations we are unable to test for a 3 way interaction, thus the best model for this data would be

Model: $y_{ij} = \alpha + \beta_i + \tau_j + \gamma_k + (\beta\tau)_{ij} + (\beta\gamma)_{ik} + (\tau\gamma)_{jk} + \epsilon_{ijk}$

However, it can be seen that there is no significance from any of the interactions, thus a more appropriate model would be

Model: $y_{ij} = \alpha + \beta_i + \tau_j + \gamma_k + \epsilon_{ijk}$

Analysis of variance table

```r
mod1.3 <- lm(duration ~ ., data = data1)
anova(mod1.3)
```

```
## Analysis of Variance Table
##
## Response: duration
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

7

```
## day         4  691.0   172.7  0.5501 0.7009123
## time        2 7380.8  3690.4 11.7520 0.0003374 ***
## treatment   1 2649.1  2649.1  8.4361 0.0082201 **
## Residuals  22 6908.4   314.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It can be seen from the anova table above that there is no statistical significance to the day (Monday - Friday), however there is a very strong statistical significance for the time of day (p-value $= 0.0003374 < \alpha = 0.05$) and treatment (p-value $= 0.0082201 < \alpha = 0.05$). Thus we can conclude that there is enough statistical evidence to reject the null hypothesis (τ's are the same) in favor of the alternative (τ's are not the same) and say that there is a difference in the duration for the traffic lights to change if I push the button or if a car pulls up.
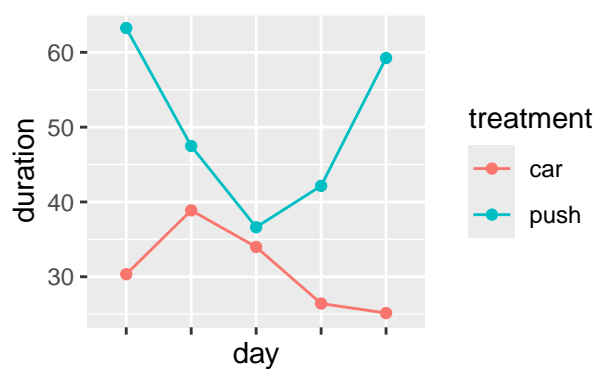
Interaction plots

A Interaction plot of Day and Time
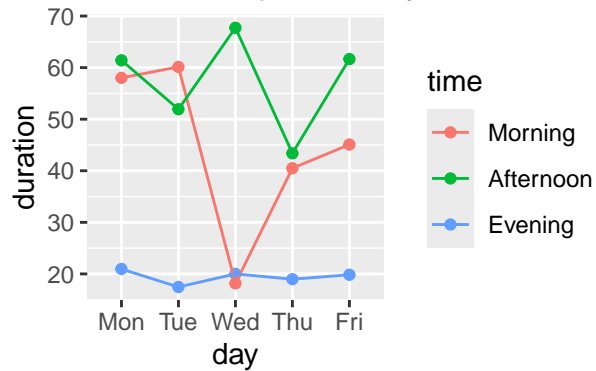
B Interaction plot of Time and Treatm
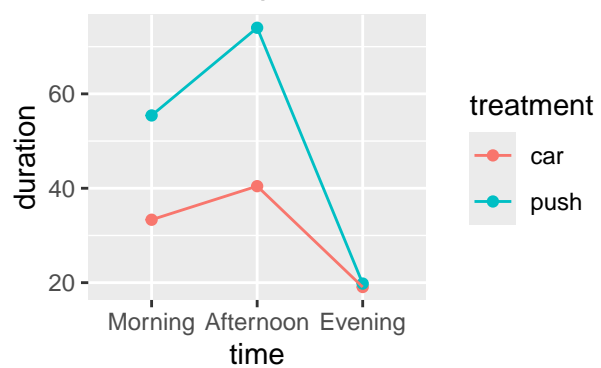
C Interaction plot of Day and Treatment

From the interaction plots above it can be seen that all interactions are interesting and should be examined as there are no parallel lines. The most interesting interactions are between the time of day and treatment.
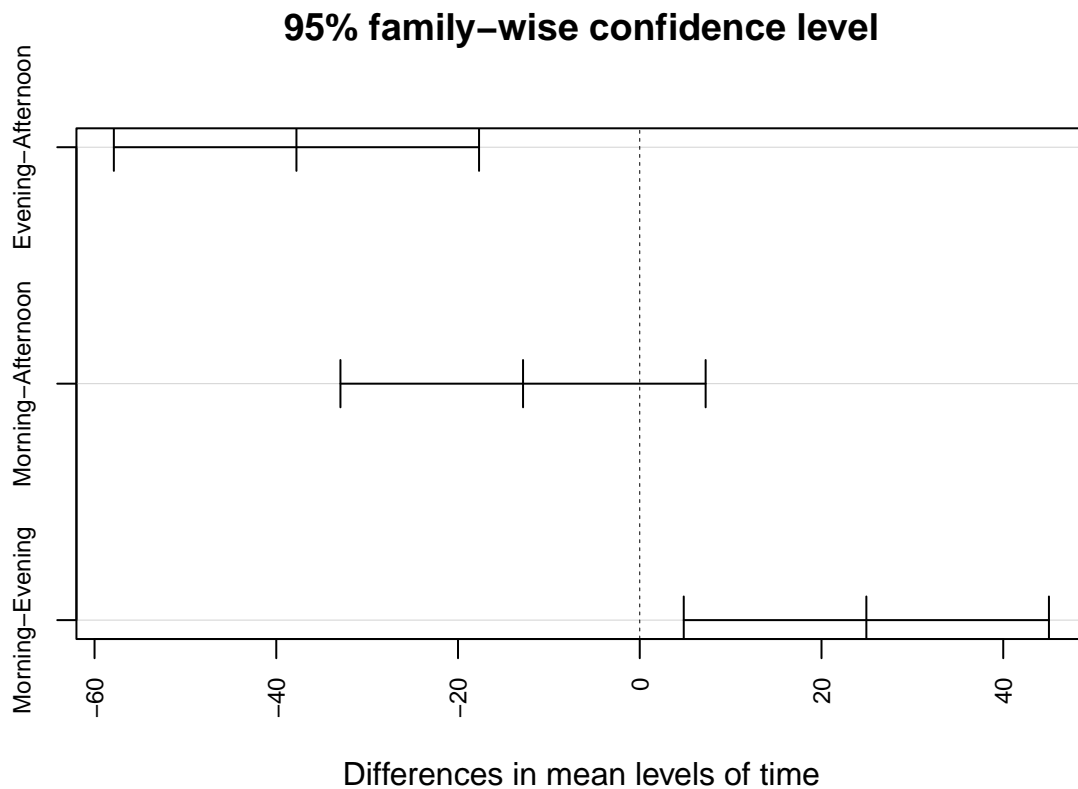
95% confidence intervals for all pairwise comparisons, and between treatments.

Interesting interactions:

- Difference between times of day

$\gamma_1 - \gamma_2$, $\gamma_1 - \gamma_3$, $\gamma_2 - \gamma_3$

95% family–wise confidence level

Differences in mean levels of time

It can be seen from the contrast above that there is a statistically significant difference between evening-afternoon and morning-evening, we will conduct a further analysis between the afternoon/morning and evening times.

$\frac{1}{2}(\gamma_1 + \gamma_2) - \gamma_3$

```
Estimate Std. Error t value   Pr(>|t|) lower CI
time c=(0.5 -1 0.5)   31.359   6.863167 4.569173 0.0001502331 17.12566
upper CI
time c=(0.5 -1 0.5) 45.59234
```

Thus we can conclude that there is a statistically significant weighted average difference between the evening wait times versus the morning and afternoon wait times. We are 95% confident that the true weighted average difference for wait time in the morning/afternoon and the evening is between (17.12566, 45.59234)

- Difference between day of week

$$\beta_1 - \beta_2,\ \beta_1 - \beta_3,\ \beta_1 - \beta_4,\ \beta_1 - \beta_5,\ \beta_2 - \beta_3,\ \beta_2 - \beta_4,\ \beta_2 - \beta_5,\ \beta_3 - \beta_5,\ \beta_4 - \beta_5$$

## 95% family–wise confidence level
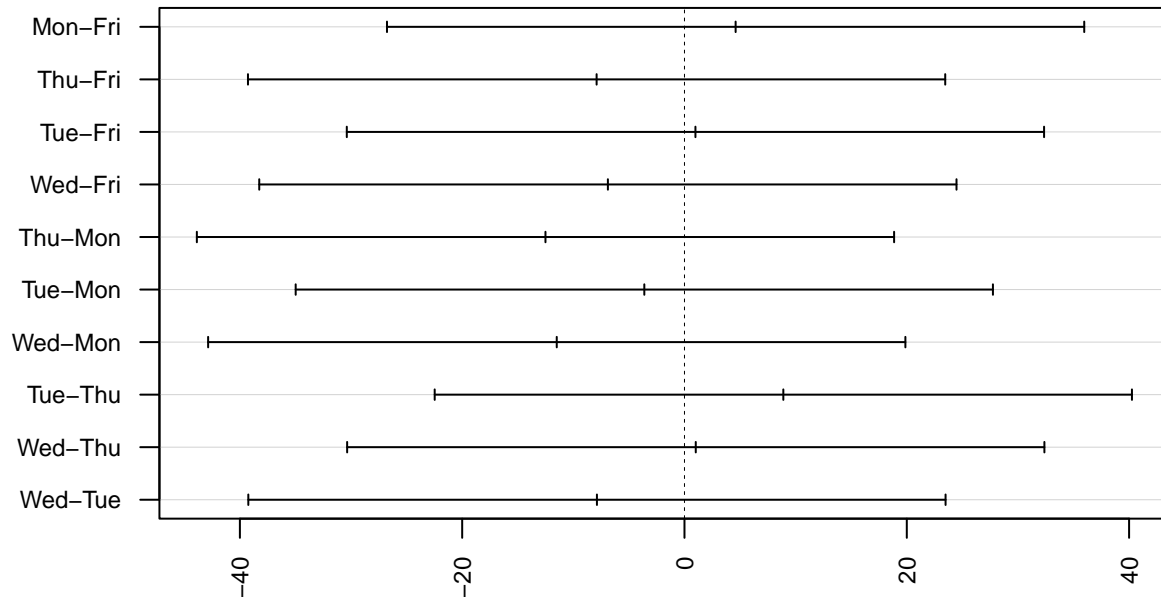


Differences in mean levels of day

It can be seen from the contrast above that there is no statistically significant differences between the days of the week, thus we can conclude that the wait times do no vary depending on the day of the week.

- Difference between treatments

$$\tau_1 - \tau_2$$

```
Tukey multiple comparisons of means
95% family-wise confidence level
##
Fit: aov(formula = duration ~ .^2, data = data1)
##
$treatment
diff   lwr   upr   p adj
push-car 18.794 5.552981 32.03502 0.0113039
```

It can be seen from the TukeyHSD above that there is a statistically significant difference between treatments push and car where on average the push treatment has a longer wait time than the car treatment. We are 95% confident that the true mean difference between treatments is between (5.552981, 32.03502)

## Conclusion

From the analysis above, it can be seen that there is enough statistically significant evidence to reject the null hypothesis. Thus there is a difference in wait times for cars versus pedestrians. We are 95% confident that the true mean difference between treatments is between (5.552981, 32.03502) and can conclude that the lights take longer to change for pedestrians than they do for cars. It was further found that there is no difference between the day of the week (Mon-Fri), however there is a difference between the time of day (Morning/Afternoon vs. Evening). We are 95% confident that the true weighted average difference for wait time in the morning/afternoon and the evening is between (17.12566, 45.59234) and can conclude that the lights take longer to change in the morning/afternoon than they do in the evening for both pedestrians and cars.

# Further research

I contacted the city for further information about the functioning of the lights at this intersection (special thanks to Todd from the city). It was found that the lights work on two variations of cycles. The first cycle is between 7am - 7pm where they work under a coordinated operation of 100s cycle length. Within this 100s cycle length the longest possible duration of each light is shown in the following pie chart.
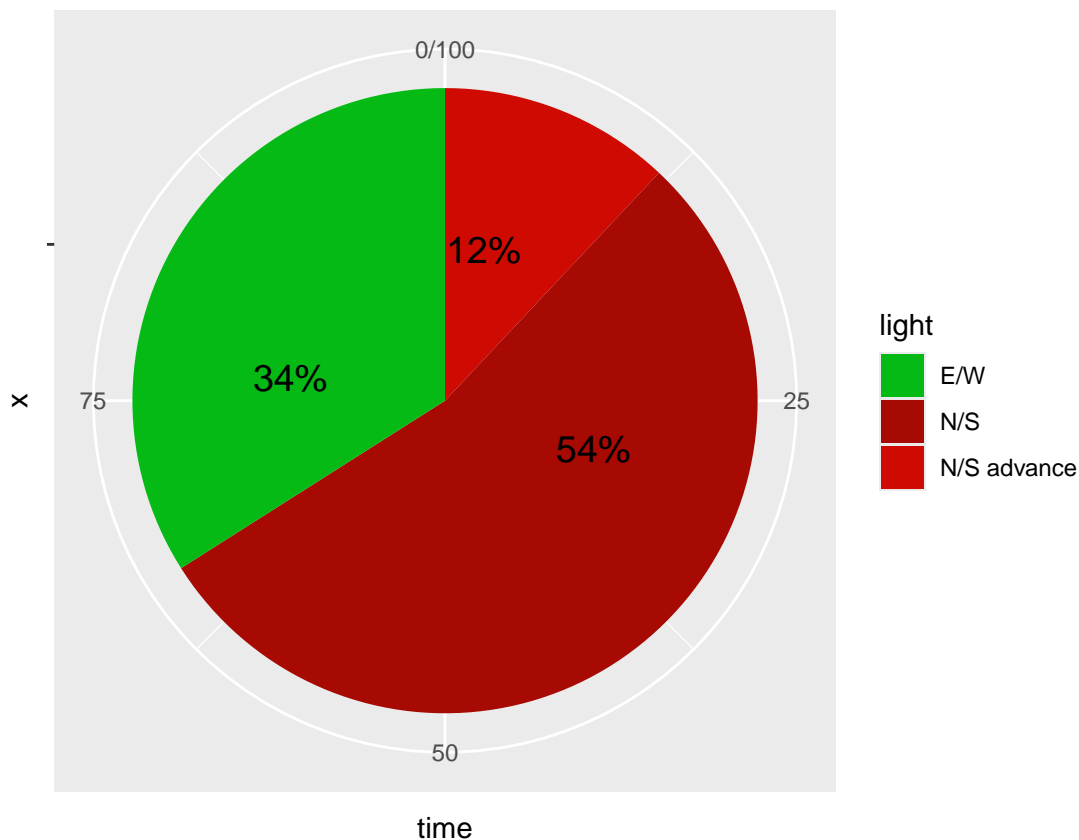


This pie chart shows the percentage (which is equivalent to the time in seconds) of each light set times of the 100s cycle. It is important to note that there is not enough time for the pedestrian walk cycle to complete after the E/W green light has begun, thus if a pedestrian pushes the walk button while the light is green they will have to wait the entirety of the cycle for the walk signal to appear. In this study the pedestrian walk button was not pushed during a green E/W light (only red).

In the data it can be seen that one of the points associated with Monday afternoon for the push treatment took a total of 101.62s. This brings up an important question of who is wrong? Since the button was only pushed during a red E/W (thus 66% of the pie chart equaling 66 seconds of maximum wait time) how is it that this pedestrian (me) timed a wait time of 35s more than the longest possible wait time?

Besides this single point, every other point falls within reasonable wait times (according to the city) for both treatments.

The second cycle runs between 7pm - 7am where the lights work under free-mode operation where the lights will remain red E/W until either a pedestrian or car shows up where it will immediately activate at the same rate for both.

The city's description of the two cycle variations lines up with this studies findings as it was found that there is a statistically significant difference between the morning/afternoon times (between 7am-7pm) and the evening times (7pm-7am). It also shows that there is no difference between day of week as this cycle runs the same for each day. This brings us to the treatments. In this study it was found that there is enough statistical evidence to show that the average wait time for pedestrians is longer than the average wait time for cars. The city claims that the car does not activate the lights faster, it just depends when in the cycle the signal (pedestrian or car) is activated. It is possible that sometimes the pushing of the pedestrian button was directly after the lights changed to red it was often activated at the start of the cycle thus warranting a longer wait time. Whereas the cars arrived at various times in the cycle, thus some arrived at the beginning and had to wait just as long as the pedestrians and some arrived near the end of the cycle where they did not have to wait long at all.

Overall, the city's claims support the study's findings that ultimately pedestrian wait times are longer on average then car wait times.

# Appendix

```r
knitr::opts_chunk$set(echo = TRUE, fig.pos='!h')
set.seed(4)
f <- factor(c("push", "car")) #treatments
b1t <- sample(f, 2) #block 1, treatment order
b2t <- sample(f, 2)
b3t <- sample(f, 2)
b4t <- sample(f, 2)
b5t <- sample(f, 2) #5 blocks
b6t <- sample(f, 2)
b7t <- sample(f, 2)
b8t <- sample(f, 2)
b9t <- sample(f, 2)
b10t <- sample(f, 2)
b11t <- sample(f, 2)
b12t <- sample(f, 2)
b13t <- sample(f, 2)
b14t <- sample(f, 2)
b15t <- sample(f, 2) # 15 blocks
t2 <- rbind(b1t, b2t, b3t, b4t, b5t, b6t, b7t, b8t, b9t, b10t, b11t, b12t, b13t, b14t, b15t)
block1 <- factor(rep(c("Mon", "Tue", "Wed", "Thu", "Fri"), each = 3)) #day of week
block2 <- factor(rep(c("Morning", "Afternoon", "Evening"), 5)) #time of day

plan   <- data.frame(day = block1, time = block2, treatment = t2, y1 = NA, y2 = NA)
rownames(plan) <- NULL
knitr::kable(plan, col.names = c("Day", "Time", "treatment 1", "2", "duration 1", "2"), caption = "Desig
#1 is push, 2 is car
# data
y1 <- c(65.1, 101.62, 23.02, 59.96, 63.34, 19.13, 20.45, 70.21, 19.20, 51.68,
        55.88, 18.89, 79.95, 78.97, 18.79) #push data
y2 <- c(50.92, 21.21, 18.88, 60.32, 40.52, 15.79, 15.89, 65.25, 20.80, 29.31,
        30.89, 19.06, 10.22, 44.35, 20.87) #car activation

y <- c(rbind(y1,y2))

data1 <- data.frame(day = rep(block1, each = 2), time = rep(block2, each = 2),
                    treatment = f, duration = y)
knitr::kable(data1, col.names = c("Day", "Time", "Treatment", "Duration"), caption = "Collected data")
library(tidyverse)
#boxplot
data1 %>%
  ggplot() +
  aes(x = treatment, y = duration) +
  labs(title = 'Treatment vs. Duration(sec)') +
  geom_boxplot()
# entire model
mod1 <- aov(duration~.^3, data = data1)
summary(mod1)
# double interaction model
mod1.1 <- aov(duration~.^2, data = data1)
summary(mod1.1)
# main effect model
```

13

```r
mod1.2 <- aov(duration~., data = data1)
summary(mod1.2)
mod1.3 <- lm(duration ~ ., data = data1)
anova(mod1.3)
library(tidyverse)
library(ggpubr)
# re-order data for interaction plot
order_data <- data1 %>%
  mutate(
    day = day %>% factor(levels=c("Mon", "Tue", "Wed", "Thu", "Fri")), #to keep in proper order
    time = time %>% factor(levels = c("Morning", "Afternoon", "Evening")),
    treatment = treatment)

#day x time
dt <- order_data %>%
  ggplot() +
  aes(x = day, color = time, group = time, y = duration) +
  stat_summary(fun = mean, geom = "point") +
  labs(title = 'Interaction plot of Day and Time') +
  stat_summary(fun = mean, geom = "line")

#time x treatment
ttr <- order_data %>%
  ggplot() +
  aes(x = time, color = treatment, group = treatment, y = duration) +
  stat_summary(fun = mean, geom = "point") +
  labs(title = 'Interaction plot of Time and Treatment') +
  stat_summary(fun = mean, geom = "line")

#day x treatment
dtr <- order_data %>%
  ggplot() +
  aes(x = day, color = treatment, group = treatment, y = duration) +
  stat_summary(fun = mean, geom = "point") +
  labs(title = 'Interaction plot of Day and Treatment') +
  stat_summary(fun = mean, geom = "line")

ggarrange(dt, ttr, dtr + rremove("x.text"),
          labels = c("A", "B", "C", "D"),
          ncol = 2, nrow = 2)
library(gmodels)
#only need to do TukeyHSD for a pairwise comparison
tukey_time <- TukeyHSD(mod1.1, which = "time", conf.level = 0.95)
plot(TukeyHSD(mod1.1, "time"), las = 3, cex.axis = 0.75)
con1 <- fit.contrast(mod1.3, "time", c(1/2, -1, 1/2), conf.int = 0.95) #between afternoon/morning - eve
con1
tukey_day <- TukeyHSD(mod1.1, which = "day", conf.level = 0.95)
plot(TukeyHSD(mod1.1, "day"), las = 2, cex.axis = 0.75)
TukeyHSD(mod1.1, which = "treatment", conf.level = 0.95)
ns_a <- 12
ns <- 54
ew <- 34
total <- sum(ns_a, ns, ew)
```

```r
# pie chart of light times
library("formattable")
data2 <- data.frame(light = c("N/S advance", "N/S", "E/W"), time = c(12, 54, 34))

data2 %>%
  ggplot() +
  aes(x="", y = time, fill = light) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start=0)+
  scale_fill_manual(values = c("#06ba15", "#a60a03", "#cf0a02"))+
  geom_text(aes(y = time/3 + c(0, cumsum(time)[-length(time)]),
            label = percent(time/100, 0)), size=5)
```