STAT321_FinalReport

2023-06-05

FINAL Report: Investigating The Relationship Between Land Use Diversity and Pedestrian Activity in Seattle

Authors: Finch Brown and JP Lopez

Load Libraries

install.packages("tidyverse") library(tidyverse) install.packages("dplyr") library(dplyr) install.packages("ggplot2") library(ggplot)

loading datasets

people_moving_df <- read.csv("Public_Life.csv") View(people_moving_df)

locations_df <- read.csv("locations.csv") View(locations_df)

Section One: Introduction

An introduction that introduces the research question and hypothesis, explaining why it is interesting and worth studying

This report aims to investigate the impact of land use diversity on pedestrian activity in Seattle. The focus of this study emanates from the premise that land use diversity might be an essential element influencing pedestrian movement. Our hypothesis suggests that higher land use diversity stimulates increased pedestrian activity. The relevance of this investigation lies in its potential to inform urban development and planning policies, subsequently fostering active, sustainable, and vibrant urban environments.

Section Two: Data

A data section describing the source of the data, measurement, and uses plots to summarize the dependent variable

The datasets used in this report were sourced from the Public Life Data repository by Seattle's Department of Transportation (link: https://data.seattle.gov/Transportation/Public-Life-Data-People-Moving/7rx6-5pgd). The primary dataset, 'Public Life Data: People Moving,' comprises pedestrian counts from different Seattle locations. The second dataset, 'Locations,' contains various land use types (e.g., Residential, Mixed, Commercial, etc.).

To illustrate the dependent variable, i.e., pedestrian movement, a boxplot is presented, demonstrating pedestrian counts in single-use and mixed-use spaces.

ggplot(new_df, aes(x = land_use, y = total_count, fill = land_use), na.rm = TRUE) + geom_boxplot() + xlab("Land Use Diversity") + ylab ("Number of People Moving") + theme(legend.position = "none")

Section Three: Results

A results section containing a scatterplot of the main relationship iof interest and output for the main regression

To illustrate the main relationship of interest, a scatterplot is provided below. The scatterplot showcases the distribution of pedestrian counts relative to land use diversity.

remove rows with NAs

new_df_clean <- na.omit(new_df)

generate the scatterplot

ggplot(new_df_clean, aes(x = land_use, y = total_count, color = land_use)) + geom_jitter(width = 0.3) + xlab("Land Use") + ylab("Pedestrians Through the Space")

plot(new_df_clean$classification, new_d f_c lean$total_count, xlab = "Classification (0 for Single Use, 1 for Mixed Use)", ylab = "Pedestrians Through the Space", col = ifelse(new_df_clean$classification == 1, "green", "purple"))

The principal regression analysis was performed with the 'total_count' as the outcome variable and 'classification' (land use diversity) as the explanatory variable. The regression output is as follows: model <- lm(total_count ~ classification, data = new_df) summary(model)

The model indicates a significant positive relationship between land use diversity and pedestrian activity. This relationship suggests that a higher level of land use diversity corresponds to an increase in pedestrian movement.

Conclusion

A brief conclusion summarizing the results, assesses the extent to which you found support for your hypothesis, and describes limitations of your analysis and threats to inference

The analysis provides evidence supporting our hypothesis that land use diversity positively influences pedestrian activity in Seattle. This finding has significant implications for urban development strategies and the promotion of pedestrian-friendly environments.

However, these results should be interpreted with some limitations in mind. The study is observational and, as such, cannot confirm causality. Additionally, the model explains a relatively small portion of the variance in pedestrian activity, indicating that other factors beyond land use diversity likely contribute significantly to pedestrian movement.

Moving forward, research may consider incorporating more variables into the analysis, such as population density, available amenities, and public transportation facilities, to gain a more comprehensive understanding of pedestrian activity determinants.

Despite the limitations, this analysis provides valuable insights into urban design and planning, hinting that land use diversity can foster more vibrant, pedestrian-friendly environments. The results of this study provide a stepping stone for future research exploring urban environments and pedestrian activity

Initial Analysis of Pedestrian Movement Through Single-Use and Mixed-Use Spaces

by Finch Brown and JP Lopez

Introduction

In this project, we will examine the relationship between land use diversity and pedestrian activity using the Seattle Public Life Data: People Moving dataset linked here: (https://data.seattle.gov/Transportation/Public-Life-Data-People-Moving/7rx6-5pgd). The tentative research question we present is: "How does land use diversity impact pedestrian activity in Seattle?" Our group hypothesizes that a higher level of land use diversity will increase pedestrian activity.

Research Design

To measure land use diversity (explanatory variable), we categorized each location based on the street classifications from the Gehl Institute's Public Life Data Protocol, which were used in these studies. We classified the 10 categories of streets used in the studies as either single-use or mixed-use areas. We measured pedestrian activity (outcome variable) based on the total counts of moving pedestrians taken for each entry in the dataset. If the data analysis reveals a positive relationship between land use diversity and pedestrian activity, this finding will support our hypothesis, suggesting that diverse land use contributes to more pedestrian movement.

Exploratory Data Analysis

Our initial analysis visually shows a positive relationship between mixed-used areas and more pedestrians moving through the space, with a statistically significant p-value. We had some issues with data visualizations and will continue to work on cleaning the data and interpreting summary statistics in the final report.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.1     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- 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
```

```
# loading data
people_moving_df <- read.csv("Public_Life.csv")

locations_df <- read.csv("locations.csv")

unique(locations_df$location_character)
```

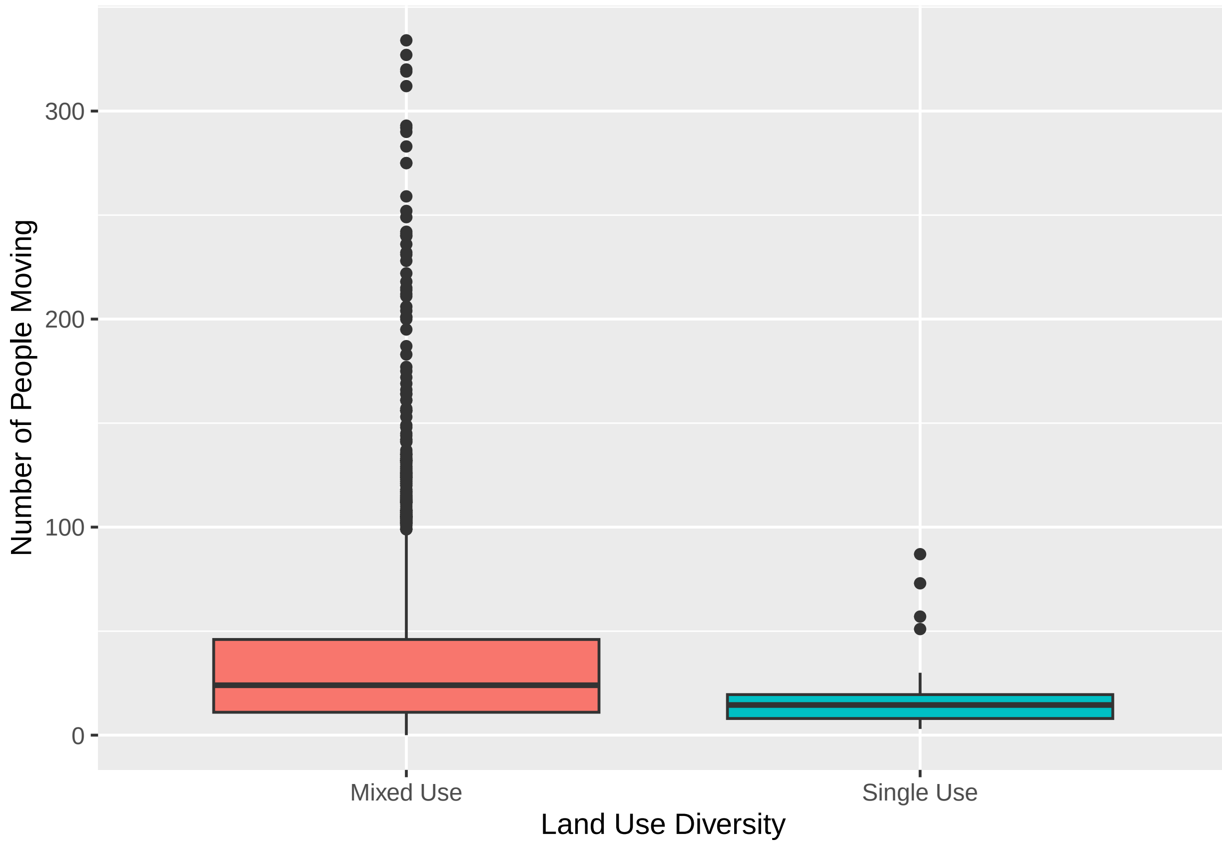
```
## [1] "Commercial"      "Mixed"           "Infrastructural" "CBD"
## [5] "Residential"     "Medical"         "Recreational"    "Civic"
```

```
# isolating important rows
small_loc_df <- data.frame(locations_df$location_id, locations_df$location_line_typology_vehicular)

small_move_df <- data.frame(people_moving_df$unique_moving_id,  people_moving_df$moving_row_total, peopl

# changing column names
colnames(small_loc_df) <- c("location", "typology")

colnames(small_move_df) <- c("id", "total_count", "start", "end", "location")

# combining dataframes
new_df <- left_join(small_loc_df, small_move_df,
                             by = join_by(location),
                             multiple = "all")

# creating column to categorize neighborhood types

unique(new_df$typology)
```

```
##  [1] "Neighborhood Corridor"             "Urban Village Neighborhood"
##  [3] "Urban Village Neighborhood Access" "Downtown Neighborhood"
##  [5] "Downtown Neighborhood Access"      "Downtown"
##  [7] "Urban Center Connector"            "Urban Village Main"
##  [9] "Neighborhood Yield Street"         "Industrial Access"
```

```
new_df$classification <- ifelse(new_df$typology %in% c("Neighborhood Corridor", "Urban Village Neighbor

# creating column for count per unit of time

new_df$start <- as.POSIXlt(new_df$start, format="%d/%m/%Y %I:%M:%S %p", tz = "UTC", na.rm = T)

new_df$end <- as.POSIXlt(new_df$end, format="%d/%m/%Y %I:%M:%S %p", tz = "UTC", na.rm = T)

new_df$duration <- as.numeric((new_df$end - new_df$start) / 60)

new_df$count_per_minute <- new_df$total_count / new_df$duration

# making a boxplot

new_df$land_use <- ifelse(new_df$classification == 1, "Mixed Use", "Single Use")

boxplot <- ggplot(new_df, aes(x = land_use,
                          y = total_count,
                          fill = land_use), na.rm = TRUE) +
  geom_boxplot() +
  xlab("Land Use Diversity") +
  ylab ("Number of People Moving") +
  theme(legend.position = "none")
boxplot
```

```
## Warning: Removed 4 rows containing non-finite values (`stat_boxplot()`).
```

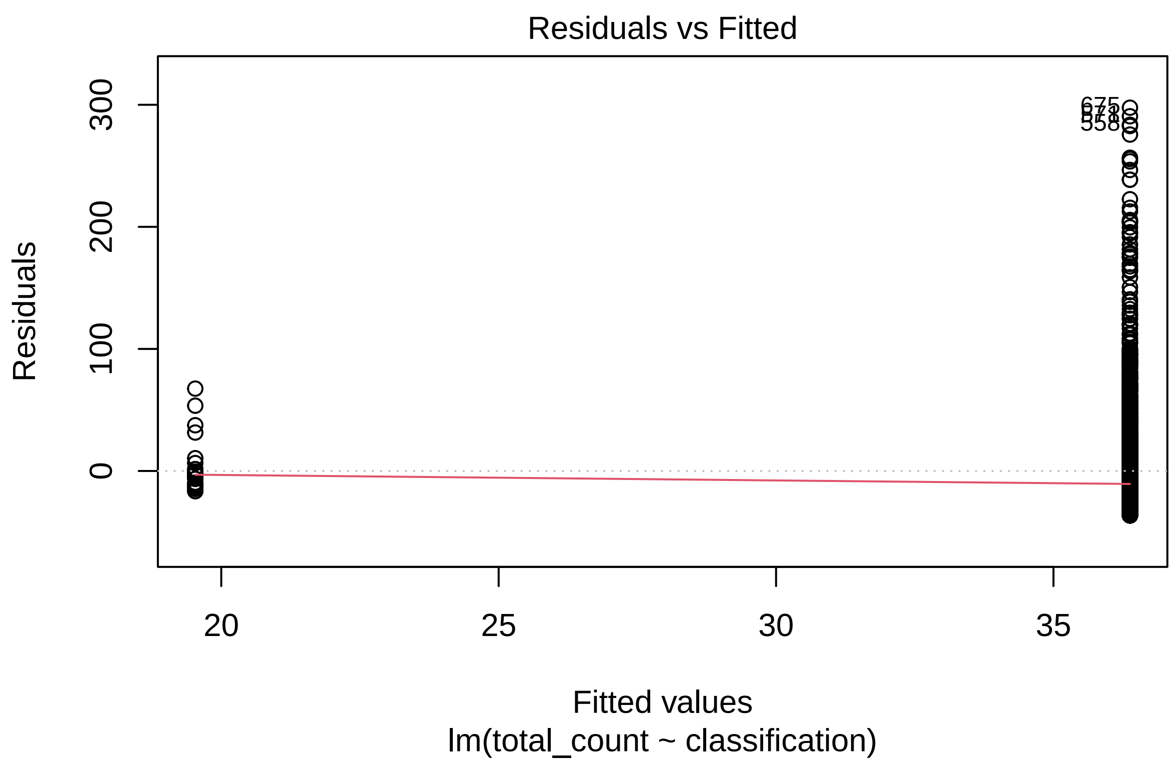


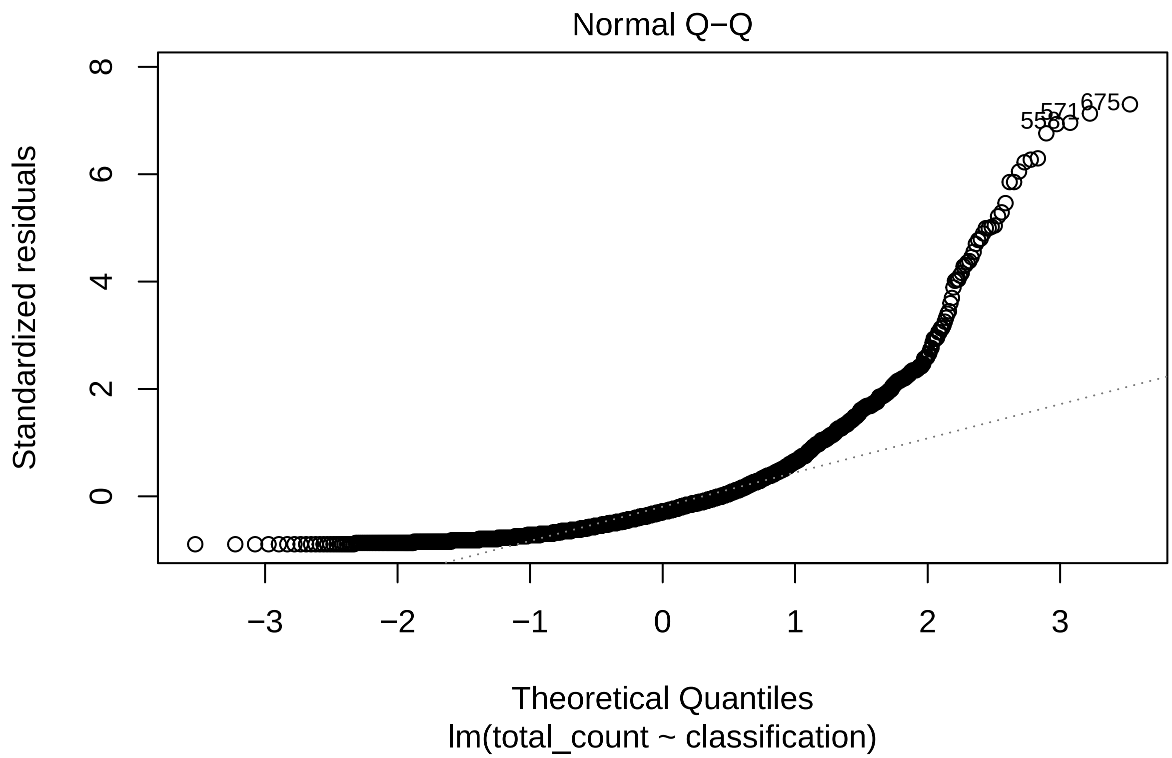

```
# making a scatterplot
# plot(new_df$land_use, df$total_count, na.rm = T,
 #     xlab = "Land Use",
 #    ylab = "Pedestrians Through the Space",
 #    col = ifelse(new_df$land_use == 1, "green", "purple"))

# conducting main regression

model <- lm(total_count ~ classification, data = new_df)

summary(model)
```

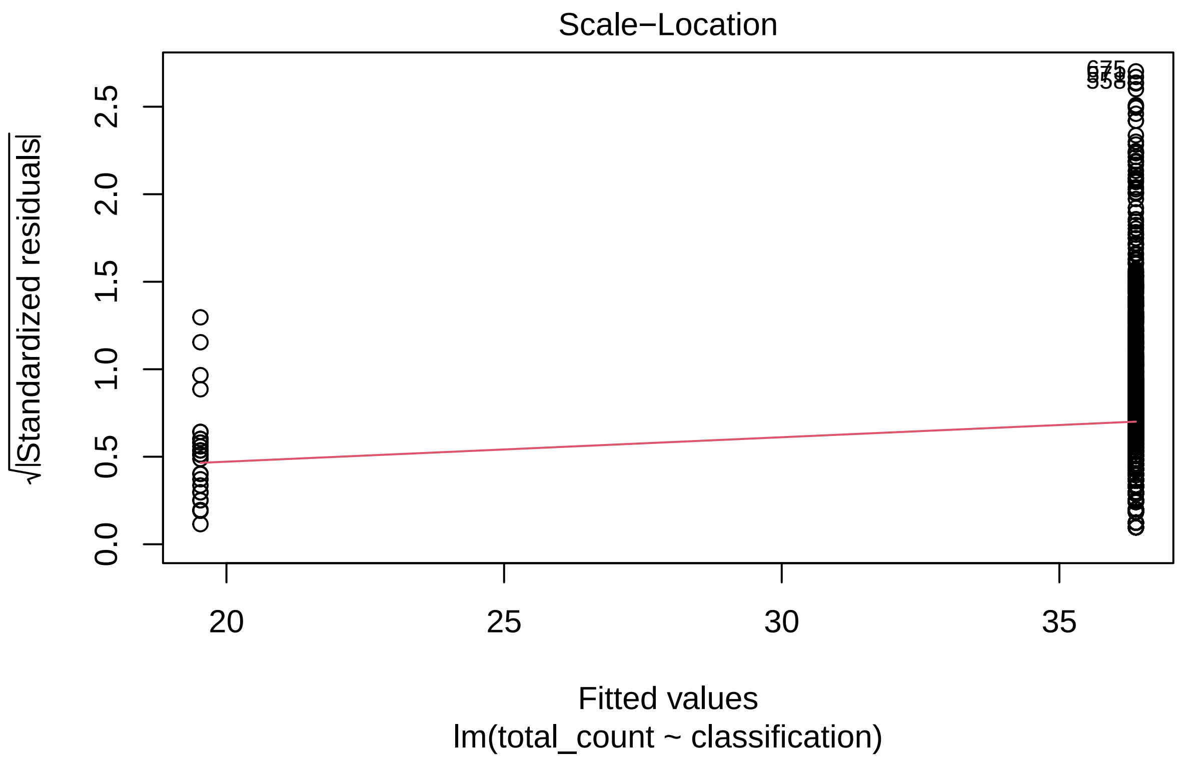
```
##
## Call:
## lm(formula = total_count ~ classification, data = new_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.379 -25.379 -11.379   9.621 297.621
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      19.531      7.207   2.710  0.00678 **
## classification   16.848      7.256   2.322  0.02032 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

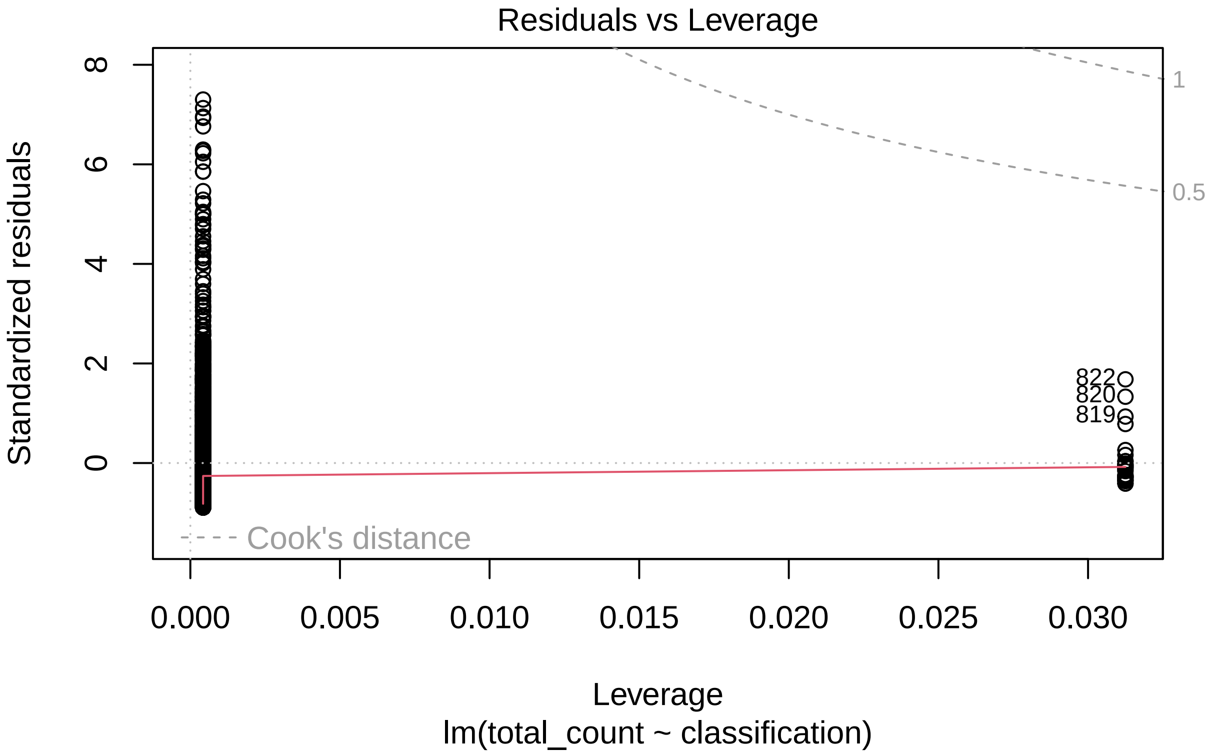
```
##
## Residual standard error: 40.77 on 2376 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.002264,   Adjusted R-squared:  0.001844
## F-statistic: 5.391 on 1 and 2376 DF,  p-value: 0.02032
```

```
plot(model)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(total_count ~ classification)
675
571
558
−3
−2
−1
0
1
2
3
8
6
4
2
0


Scale−Location
675
958
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
20
25
30
35
Fitted values
lm(total_count ~ classification)


Residuals vs Leverage
Standardized residuals
8
6
4
2
0
1
0.5
822
820
819
Cook's distance
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Leverage
lm(total_count ~ classification)