

# From Point A to Point B: E-Scooters as a Mode of Transportation in Chicago

Sarah Scott

9/20/2021

In 2020, Chicago piloted an E-Scooter program to test how city residents would respond to this new mode of transportation. The city's full report can be found here:

<https://www.chicago.gov/content/dam/city/depts/cdot/Misc/EScooters/2021/2020%20Chicago%20E-scooter%20Evaluation%20-%20Final.pdf> (<https://www.chicago.gov/content/dam/city/depts/cdot/Misc/EScooters/2021/2020%20Chicago%20E-scooter%20Evaluation%20-%20Final.pdf>)

The goal of my analysis is to look at the origin and end points of scooter rides. Where were these novel wheels popular and where did they sit growing dust and rust on the sidewalk? I am using the raw data provided on the City of Chicago website, which can be found here:

<https://data.cityofchicago.org/Transportation/E-Scooter-Trips-2020/3rse-fbp6> (<https://data.cityofchicago.org/Transportation/E-Scooter-Trips-2020/3rse-fbp6>)

## Setting Up The Proverbial Road

```
#Load packages (need to run every session); note that I needed to install the packages lubridate, ggmap, and gtsummary.

#install.packages("lubridate")
#install.packages("ggmap")
#install.packages('gtsummary')
library(lubridate)
library(tidyverse)
library(haven)
library(ggmap)
library(gtsummary)

#Clear R Environment.
rm(list = ls())

# Set working directory to location of data
setwd("~/Desktop/UChicago/Coding Camp/Final Project")

#Load Scooter Data csv file
scooter_data_raw <- read_csv('E-Scooter_Trips_-_2020.csv')
```

## Data Exploration and Manipulation

In the original data set, "Trip Duration" was given in seconds, "Trip Distance" in meters, the naming conventions of "Start Community" and "End Community" were overly complicated, and during my initial data import, the date and time column was read into R as a character column instead of a date-time column. The various code blocks below rename and reformat my data.

```

#Renaming Start and End Community columns; converting trip duration from seconds to mins; converting trip distance from meters to miles
scooter_data <- scooter_data_raw %>%
  rename(Start_Community = `Start Community Area Name`,
         End_Community = `End Community Area Name`) %>%
  mutate(Trip_Duration_Mins = `Trip Duration`/60,
         Trip_Distance_Miles = `Trip Distance`/1609)

#Create a new column in scooter_data that lists the starting community and the ending community separated by the symbol "->".
#https://www.marsja.se/how-to-concatenate-two-columns-or-more-in-r-stringr-tidyr/

scooter_data$start_end_community <- paste(scooter_data$Start_Community, "->",
scooter_data$End_Community)

#Convert Start Time and End Time columns to date time format.
#Citation: https://stackoverflow.com/questions/4310326/convert-character-to-class-date

scooter_data$Start_date_time <- mdy_hms(scooter_data$`Start Time`)
scooter_data$End_date_time <- mdy_hms(scooter_data$`End Time`)

#Separate Start Time and End Time columns, which originally included the date and time of the ride, into separate columns where the Date is by itself in one column, and the time is by itself in another.
#file:///Users/scotty/Downloads/lubridate%20(2).pdf
#https://www.marsja.se/how-to-extract-time-from-datetime-in-r-with-examples/

scooter_data$Start_Date <- date(scooter_data$Start_date_time)
scooter_data$Start_Time <- format(scooter_data$Start_date_time, format = "%H:%M:%S")
scooter_data$Start_Time <- hms(scooter_data$Start_Time)

scooter_data$End_Date <- date(scooter_data$End_date_time)
scooter_data$End_Time <- format(scooter_data$End_date_time, format = "%H:%M:%S")
scooter_data$End_Time <- hms(scooter_data$End_Time)

```

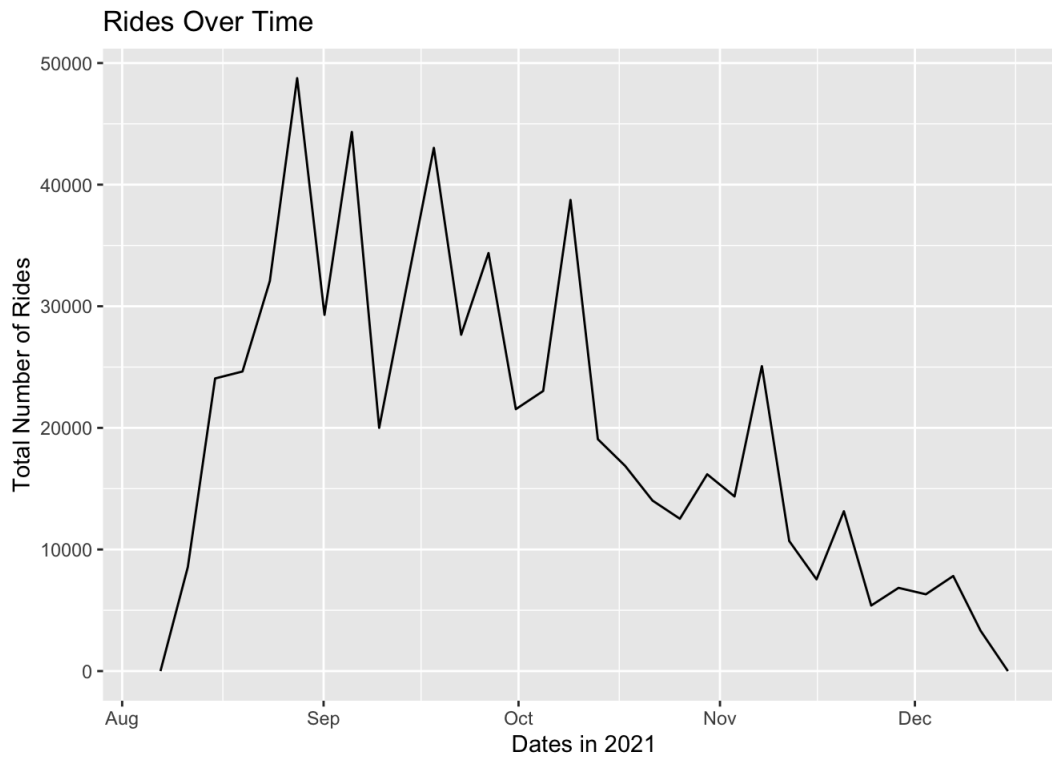
## Graph of Number of Trips over Time (Months)

The pilot program was scheduled to last from mid-August 2020 through December 2020. Below is a basic graph plotting the number of rides taken over this time period. I have included this visual to give you a sense of how many rides occurred, as well as the frequency, during the pilot. From mid-August to early September, we can see that residents quickly learned of the program and began utilizing the scooters. The number of rides generally decreases over time which is congruent with the expected drop in temperatures as the weather drops throughout the fall.

```

#Graph number of scooter trips over time (months)
ggplot(data = scooter_data,
       mapping = aes(x = Start_Date)) +
  geom_freqpoly() +
  labs(title = "Rides Over Time", x = "Dates in 2021", y = "Total Number of Rides")

```



## Defining Different Types of Rides

There are two types of rides that I am interested in - Neighborhood Rides, which are defined as rides that start and end in the same neighborhood, and Crosstown\_Rides, which are defined as rides that start and end in different neighborhoods. In the table below, we can see that most rides were Neighborhood Rides (about 70%).

Crosstown Rides, on average, were 62.5% longer (in minutes) and 127% longer (in miles) than Neighborhood Rides. This appears to be a significant difference. However, the raw numbers indicate that riders utilized the scooters for shorter rides overall, regardless of what neighborhood they started/ended in. On average, Neighborhood Rides lasted for 8 minutes and covered less than 1 mile. Crosstown rides were only 5 minutes longer on average and still covered under 2 miles.

Finally, we can see that Lime Bikes were more popular than either the Bird Bikes or Spin Bikes individually for both Crosstown and Neighborhood Rides.

```
#Create a column called "ride_type" that identifies if the ride is a Neighborhood Ride or a Crosstown Ride.
scooter_data <- scooter_data %>%
  mutate(ride_type = ifelse(Start_Community == End_Community, "Neighborhood Ride", "Crosstown Ride"))

#Create a Summary Table looking at Crosstown vs. Neighborhood Rides
scooter_data %>%
  select(ride_type, Trip_Duration_Mins, Trip_Distance_Miles, Vendor) %>%
  tbl_summary(by = ride_type) %>%
  add_overall() %>% add_n() %>%
  modify_header(label ~ "***Variable***") %>%
  modify_caption("***Table 1. Comparison of Crosstown Rides vs. Neighborhood Rides***") %>%
  bold_labels()
```

**Table 1. Comparison of Crosstown Rides vs. Neighborhood Rides**

Variable	N	Overall, N = 629,175 <sup>1</sup>	Crosstown Ride, N = 188,732 <sup>1</sup>	Neighborhood Ride, N = 440,443 <sup>1</sup>
<b>Trip_Duration_Mins</b>	629,175	10 (5, 19)	13 (9, 22)	8 (4, 16)
<b>Trip_Distance_Miles</b>	629,175	1.16 (0.51, 2.26)	1.95 (1.19, 3.11)	0.86 (0.30, 1.73)

<sup>1</sup> Median (IQR); n (%)

Variable	N	Overall, N = 629,175 <sup>1</sup>	Crosstown Ride, N = 188,732 <sup>1</sup>	Neighborhood Ride, N = 440,443 <sup>1</sup>
Vendor	629,175			
bird		181,019 (29%)	57,607 (31%)	123,412 (28%)
lime		278,665 (44%)	90,323 (48%)	188,342 (43%)
spin		169,491 (27%)	40,802 (22%)	128,689 (29%)

<sup>1</sup> Median (IQR); n (%)

## Map of Ride Start and End Points

In order to visualize where most of the rides started, I have created one map that plots where rides start and another that shows where rides end.

### Mapping Start Locations of Rides

In the map below, we can see that rides start in various neighborhoods in Chicago. Larger and lighter blue circles indicate that more rides start in those neighborhoods.

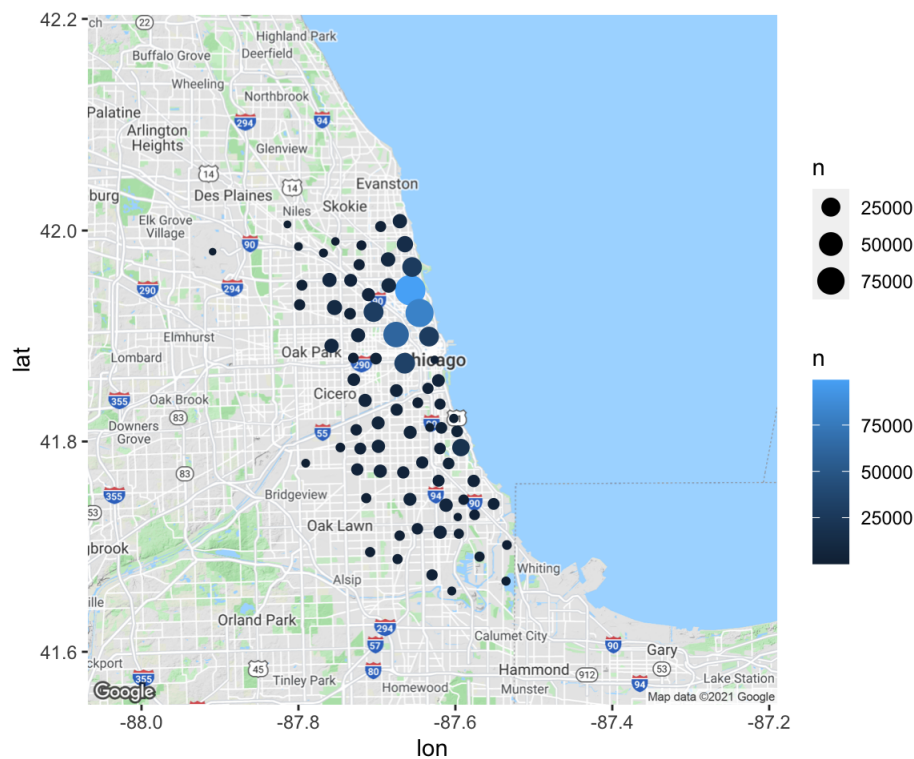
```
#https://cran.r-project.org/web/packages/ggmap/ggmap.pdf
#https://github.com/dkahle/ggmap/issues/262
#https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf
#https://www.youtube.com/watch?v=SdvGzbOZ-Qs

#register_google(key = "###") In order to get your Google API, enter ?register_google in the console.

#Count the number of rides that started in a given community.

start_community_with_lat_long <- scooter_data %>%
  group_by(Start_Community, `Start Centroid Latitude`, `Start Centroid Longitude`) %>%
  summarise(n=n())

#Map scooter rides based on their starting location
(map_rides_start_community <- ggmap(get_googlemap("Chicago, IL",
  zoom = 10,
  maptype = 'terrain',
  color = 'color',)) +
  geom_point(data = start_community_with_lat_long, aes(x = `Start Centroid Longitude`,
    y = `Start Centroid Latitude`,
    color = n,
    size = n,
  )))
```



If we zoom in on the larger blue circles, we can see that most rides started in these three neighborhoods:

Lake View

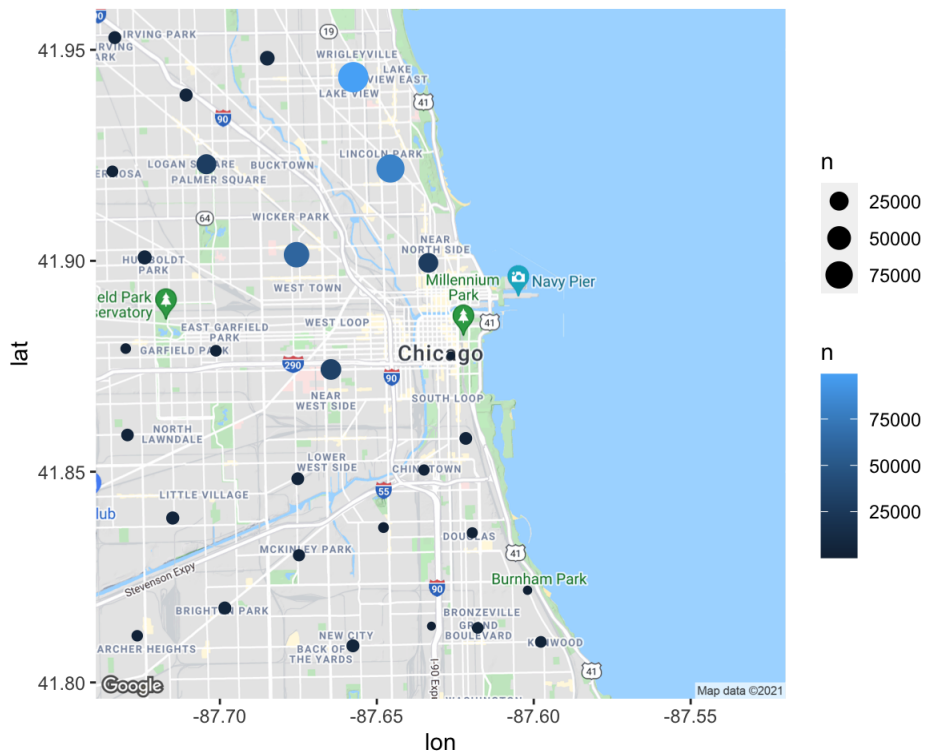
Lincoln Park

West Town

These findings are confirmed by sorting the neighborhoods by the number of trips (n) that began in those communities.

*#Zoom in on communities where the majority of rides started.*

```
(map_rides_start_community <- ggmap(get_googlemap("Chicago, IL",
  zoom = 12,
  maptype = 'terrain',
  color = 'color',)) +
  geom_point(data = start_community_with_lat_long, aes(x = `Start Centroid Longitude`,
    y = `Start Centroid Latitude`,
    color = n,
    size = n,
  )))
```



```
#Confirm the neighborhoods where most rides start
start_community_with_lat_long %>%
  arrange(desc(n)) %>%
  select(Start_Community, n)
```

```
## # A tibble: 78 × 3
## # Groups:   Start_Community, Start_Centroid Latitude [78]
##   `Start_Centroid Latitude` Start_Community      n
##   <dbl> <chr> <int>
## 1 41.9 LAKE VIEW 99318
## 2 41.9 LINCOLN PARK 80704
## 3 41.9 WEST TOWN 61413
## 4 41.9 NEAR WEST SIDE 32739
## 5 41.9 LOGAN SQUARE 29330
## 6 41.9 NEAR NORTH SIDE 28374
## 7 42.0 UPTOWN 27864
## 8 41.8 HYDE PARK 18863
## 9 42.0 EDGEWATER 14554
## 10 41.9 BELMONT CRAGIN 11041
## # ... with 68 more rows
```

## Mapping End Locations of Rides

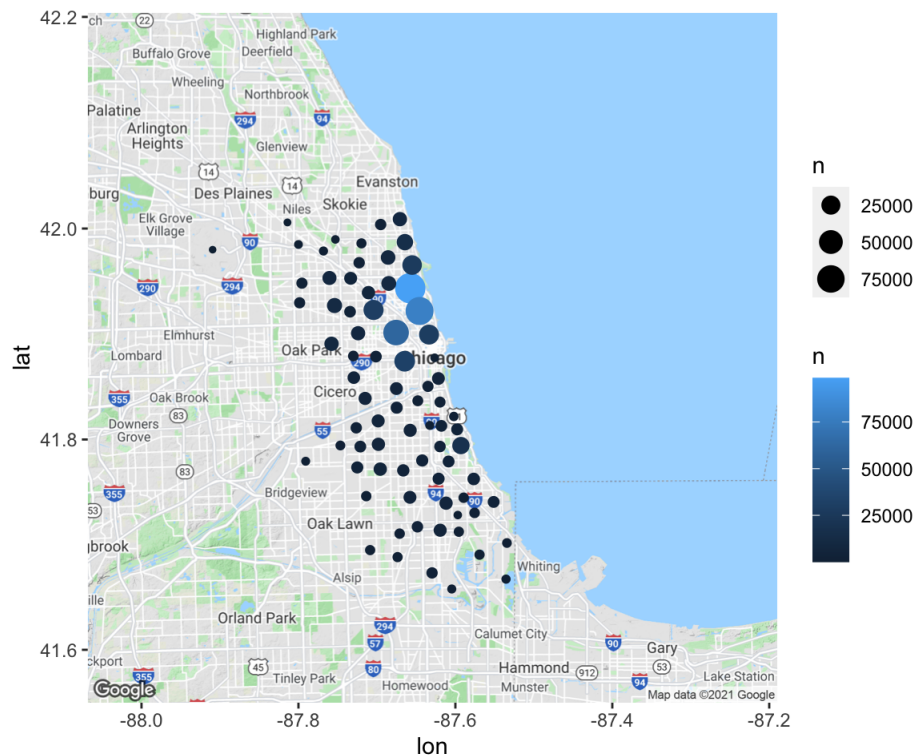
The map below shows the location that rides ended in Chicago. As before, larger and lighter blue circles indicate that more rides start in those neighborhoods.

```
#Count the number of rides that ended in a given community.
```

```
end_community_with_lat_long <- scooter_data %>%
  group_by(End_Community, `End Centroid Latitude`, `End Centroid Longitude`) %>%
  summarise(n=n())
```

```
#Map scooter rides based on their ending location
```

```
(map_rides_end_community <- ggmap(get_googlemap("Chicago, IL",
  zoom = 10,
  maptype = 'terrain',
  color = 'color',)) +
  geom_point(data = end_community_with_lat_long, aes(x = `End Centroid Longitude`,
    y = `End Centroid Latitude`,
    color = n,
    size = n,
  )))
```



If we zoom in on the larger blue circles, we can see that most rides ended in these three neighborhoods:

Lake View

Lincoln Park

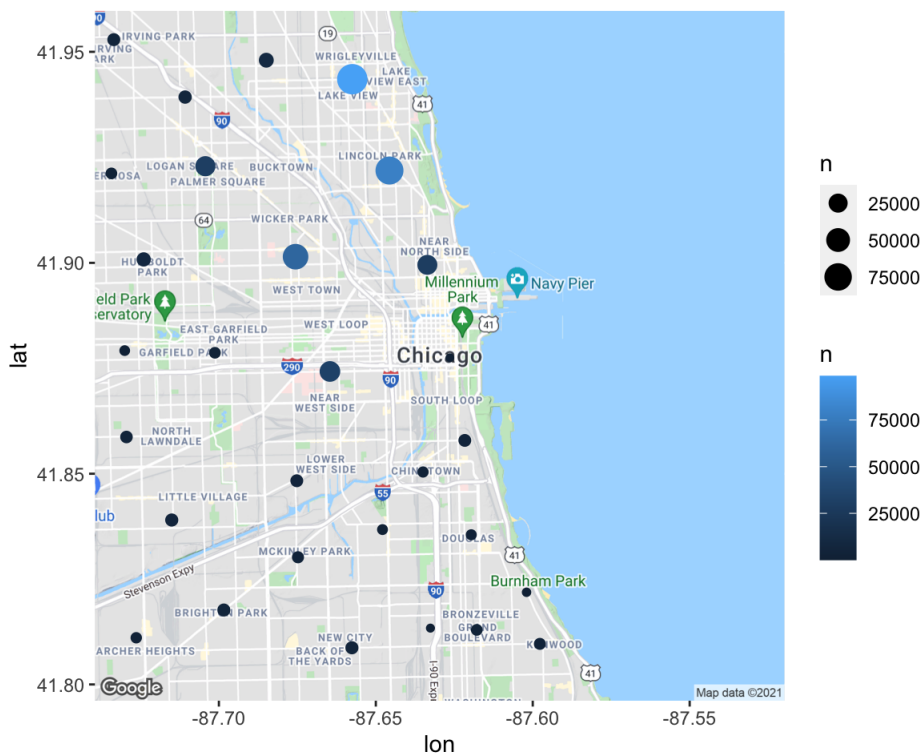
West Town

These findings are confirmed by sorting the neighborhoods by the number of trips (n) that ended in those communities.

##Zoom in on communities where the majority of rides ended.

```
end_community_with_lat_long <- scooter_data %>%
  group_by(End_Community, `End Centroid Latitude`, `End Centroid Longitude`) %>%
  summarise(n=n())

(map_rides_end_community <- ggmap(get_googlemap("Chicago, IL",
  zoom = 12,
  maptype = 'terrain',
  color = 'color',)) +
  geom_point(data = end_community_with_lat_long,
    aes(x = `End Centroid Longitude`,
        y = `End Centroid Latitude`,
        color = n,
        size = n,
        )))
```



#Confirm the neighborhoods where most rides end

```
end_community_with_lat_long %>%
  arrange(desc(n)) %>%
  select(End_Community, n)
```

```
## # A tibble: 78 × 3
## # Groups:   End_Community, End Centroid Latitude [78]
##   `End Centroid Latitude` End_Community      n
##   <dbl> <chr> <int>
## 1 41.9 LAKE VIEW 98568
## 2 41.9 LINCOLN PARK 78872
## 3 41.9 WEST TOWN 61391
## 4 41.9 NEAR WEST SIDE 32470
## 5 41.9 LOGAN SQUARE 29756
## 6 41.9 NEAR NORTH SIDE 28892
## 7 42.0 UPTOWN 27214
## 8 41.8 HYDE PARK 18148
## 9 42.0 EDGEWATER 14754
## 10 41.9 BELMONT CRAGIN 10906
## # ... with 68 more rows
```



# Implications

After looking at this data, we know that scooters were well utilized in neighborhoods in the north of Chicago. While more data is needed to understand why people in southern neighborhoods might not choose to travel by bike, one conjecture we could make is that people in northern Chicago might live and work within a two block radius. In contrast, people in southern Chicago might travel farther to get to work and therefore traveling by scooter is a less attractive option.

Moving forward, Chicago should capitalize on the demand in Lake View, Lincoln Park, and West Town and put the majority of e-scooters/e-scooter charging stations there. Simultaneously, they should survey neighborhoods in the Southside to determine future demand.