Assesing Demand for Transit Accessesible Apartments in Chicago

Noah Anderson

Project Overview

This project uses the methodology laid out in Chapter 1 of Public Policy Analytics: Code & Context for Data Science in Government [3]. This project applies Steif's methodology for indicator variable analysis to transit demand in the City of Chicago. This project utilizes open transit data from the City of Chicago and wrangling it so that it is compatible with American Community Survey data. This project will use indicator variables to asses transit demand in the City of Chicago.

Introduction

The City of Chicago has been encouraging Transit Orientated Development (TOD) since 2013 following a series of ordinances adopted by City Council refereed to as the Equitable TOD (ETOD) plan [1]. The plan encourages compact and mixed use development near Chicago Transit Authority (CTA) and Metra rail stations. The goals are to increase walkablity and transit ridership to reduce greenhouse gas emissions and create a more stable tax base. The ETOD plan seeks to enable all people regardless of income, race, ethnicity, age, gender, or immigration status to experience the benefits of walkable pedestrianized areas near transit stops.

The goal of this project is to use indicator variables from the American Community Survey to asses demand for transit access. Additionally I am analyzing demographic data to take an informal analysis on the equity of TOD in Chicago. The variables analyzed are the percentage of white residents, percentage of residents with bachelors, the percentage below the poverty line, and the mean rent. By breaking the ACS tracts up into TOD and non-TOD, I compare differences between these variables within the two geographic groups. Data is pulled from the 2005-2009 and 2014-2019 Five Year ACS to observe changes in the selected indicator variables before and after the City of Chicago has enacted the ETOD plan.

Pulling ACS Data

The code chunk below utilizes the tidycensus package to pull Five Year ACS data for the 2005-2009 and 2015-2019 surveys. I could have used the 2017-2021 ACS is also available, but I though it would be best to avoid data that includes 2020 due to the pandemics impact on transit ridership and popularity. The two data frames created below specifically pull data from Cook County IL.

```
# Gets data for 2005-2009 Five Year ACS for Cook County and wrangles it
cook09 <-
  get_acs(geography = "tract", variables = c("B25026_001E","B02001_002E","B15001_050E",
                                             "B15001_009E","B19013_001E","B25058_001E",
                                             "B06012_002E"), # Selects target variables
          year=2009, state=17, county=031, geometry=T, output="wide", cache_table = TRU
E) %>%
  st_transform('ESRI:102728') %>% # Converts to a standard coordinate system

  # Renames variables
  rename(TotalPop = B25026_001E, Whites = B02001_002E,
         FemaleBachelors = B15001_050E, MaleBachelors = B15001_009E,
         MedHHInc = B19013_001E, MedRent = B25058_001E,
         TotalPoverty = B06012_002E) %>%
  dplyr::select(-NAME, -starts_with("B")) %>%

  # Converts variables to percentages
  mutate(pctWhite = ifelse(TotalPop > 0, Whites / TotalPop,0),
         pctBachelors = ifelse(TotalPop > 0, ((FemaleBachelors + MaleBachelors) / TotalP
op),0),
         pctPoverty = ifelse(TotalPop > 0, TotalPoverty / TotalPop, 0),
         year = "2009", county = "Cook") %>%
  dplyr::select(-Whites, -FemaleBachelors, -MaleBachelors, -TotalPoverty)
```

```
## Getting data from the 2005-2009 5-year ACS
```

```
## Downloading feature geometry from the Census website.  To cache shapefiles for use in
future sessions, set `options(tigris_use_cache = TRUE)`.
```

```
# Repeats steps for 2014-2019 ACS
cook19 <-
  get_acs(geography = "tract", variables = c("B25026_001E","B02001_002E","B15001_050E",
                                             "B15001_009E","B19013_001E","B25058_001E",
                                             "B06012_002E"),
          year=2019, state=17, county=031, geometry=T, output="wide", cache_table = TRU
E) %>%
  st_transform('ESRI:102728') %>%
  rename(TotalPop = B25026_001E, Whites = B02001_002E,
         FemaleBachelors = B15001_050E, MaleBachelors = B15001_009E,
         MedHHInc = B19013_001E, MedRent = B25058_001E,
         TotalPoverty = B06012_002E) %>%
  dplyr::select(-NAME, -starts_with("B")) %>%
  mutate(pctWhite = ifelse(TotalPop > 0, Whites / TotalPop,0),
         pctBachelors = ifelse(TotalPop > 0, ((FemaleBachelors + MaleBachelors) / TotalP
op),0),
         pctPoverty = ifelse(TotalPop > 0, TotalPoverty / TotalPop, 0),
         year = "2019", county = "Cook") %>%
  dplyr::select(-Whites, -FemaleBachelors, -MaleBachelors, -TotalPoverty)
```

```
## Getting data from the 2015-2019 5-year ACS
## Downloading feature geometry from the Census website.  To cache shapefiles for use in
future sessions, set `options(tigris_use_cache = TRUE)`.
```

Wrangling Transit Data and Setting Boundaries

The ACS data pulled is for all of Cook County, so a necessary first step is limiting the tracts to Chicago only. The chunk below reads in a geojson shapefile of Chicago Neighborhood boundaries from the City of Chicago website data.cityofchicago.org. After a few transformations, this shapefile can be used to limit the ACS data from all of Cook County, to the City of Chicago only.

```
chicago_neighborhoods <- st_read("https://raw.githubusercontent.com/noahba65/chicago-TO
D/main/DATA/chicago_boundary.geojson") %>%
  st_transform('ESRI:102728') %>% st_sf()
```

```
## Reading layer `chicago_boundary' from data source
##   `https://raw.githubusercontent.com/noahba65/chicago-TOD/main/DATA/chicago_boundary.
geojson'
##   using driver `GeoJSON'
## Simple feature collection with 98 features and 4 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.94011 ymin: 41.64454 xmax: -87.52414 ymax: 42.02304
## Geodetic CRS:  WGS 84
```

```
# Erases internal boundaries between neighborhoods to create a Chicago boundary
chicago_boundary <- st_union(chicago_neighborhoods) %>%
  st_sf()

# Creating Chicago only ACS tract data for 2009 and 2019 using the st_intersection funci
ton to only select data and geometries within the Chicago boundary.
chicago09 <- st_intersection(chicago_boundary, cook09)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```
chicago19 <- st_intersection(chicago_boundary,cook19)
```

```
## Warning: attribute variables are assumed to be spatially constant throughout
## all geometries
```

```
allTracts_chicago <- rbind(chicago09,chicago19)
```

The next step is to wrangle transit data for CTA stops. Using data from Chicago City Scapes Github page, a geojson file containing the locations of each CTA L entrance is read. Then a quarter mile buffer around each L stop is created. This buffer zone will be vital in selecting which ACS tracts will be designated as TOD and non-TOD. All tracts that have a centroid within the quarter-mile buffer are counted as TOD. This chunk visualizes which tracts are selected in the map below.

```
cta_stops <- st_read("https://raw.githubusercontent.com/ChicagoCityscape/gis-data/maste
r/stations_cta/cta_entrances.geojson") %>%
  st_transform('ESRI:102728') %>% st_sf()
```

```
## Reading layer `cta_entrances' from data source
##   `https://raw.githubusercontent.com/ChicagoCityscape/gis-data/master/stations_cta/ct
a_entrances.geojson'
##   using driver `GeoJSON'
## Simple feature collection with 381 features and 3 fields
## Geometry type: POINT
## Dimension:     XY
## Bounding box:  xmin: -87.904 ymin: 41.72156 xmax: -87.6058 ymax: 42.07362
## Geodetic CRS:  WGS 84
```

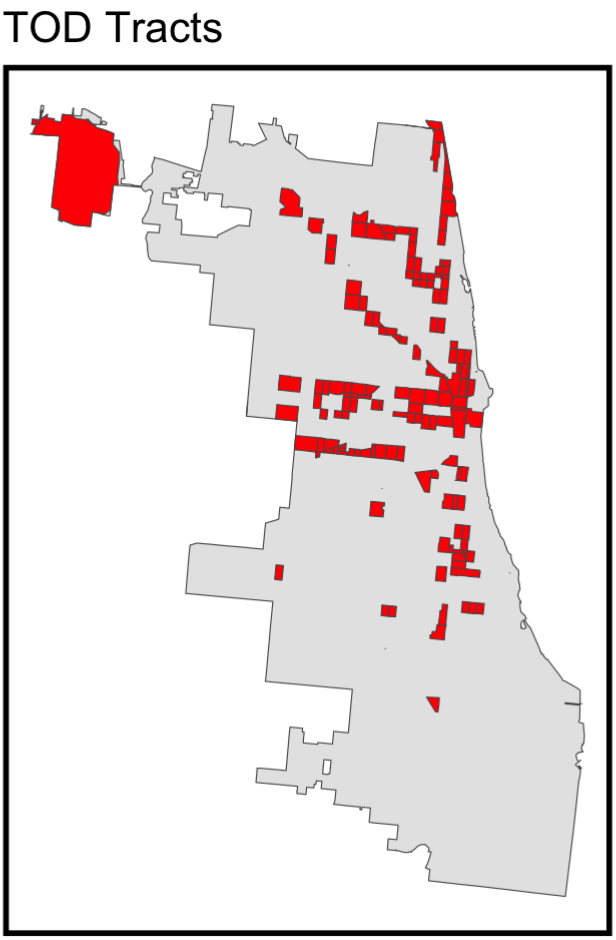
```
# Creating CTA buffer
cta_buffer <- st_union(st_buffer(cta_stops,5280/4)) %>%
  st_sf()


# Uses the st_centroid function to select which tracts are TOD
tract_buffer_chicago <- st_centroid(chicago09)[cta_buffer,] %>%
  st_drop_geometry() %>%
  left_join(dplyr::select(chicago09,GEOID)) %>%
  st_sf()
```

```
## Warning: st_centroid assumes attributes are constant over geometries
```

```
## Joining with `by = join_by(GEOID)`
```

```
ggplot() +
  geom_sf(data = st_union(chicago09)) +
  geom_sf(data = tract_buffer_chicago, fill = "red")  +
  ggtitle("TOD Tracts") + mapTheme()
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## ℹ Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Visualizing Indicator Variables

Now we can group the 2009 and 2019 Chicago data by TOD and non-TOD tracts for comparison. The graphs shown below visualize the differences in the indicators from 2009 to 2019.

```
# Groups the AllTracts Chicago data frame by TOD and non-TOD
allTracts_chicago.group <-
  rbind(
    st_centroid(allTracts_chicago)[tract_buffer_chicago,] %>%
      st_drop_geometry() %>%
      left_join(allTracts_chicago) %>%
      st_sf() %>%
      mutate(TOD = "TOD"),
    st_centroid(allTracts_chicago)[tract_buffer_chicago, op = st_disjoint] %>%
      st_drop_geometry() %>%
      left_join(allTracts_chicago) %>%
      st_sf() %>%
      mutate(TOD = "Non-TOD")) %>%
  mutate(MedRent.inf = ifelse(year == "2009", MedRent * 1.2, MedRent))
```

```
## Warning: st_centroid assumes attributes are constant over geometries
```

```
## Joining with `by = join_by(GEOID, TotalPop, MedHHInc, MedRent, pctWhite,
## pctBachelors, pctPoverty, year, county)`
```

```
## Warning: st_centroid assumes attributes are constant over geometries
```

```
## Joining with `by = join_by(GEOID, TotalPop, MedHHInc, MedRent, pctWhite,
## pctBachelors, pctPoverty, year, county)`
```

```
# 2009 rent is adjusted for inflation

# Summarizes the means of the indicator variables
allTracts_chicago.Summary <-
  st_drop_geometry(allTracts_chicago.group) %>%
    group_by(year, TOD) %>%
    summarize(Rent = mean(MedRent.inf, na.rm = T),
              Population = mean(TotalPop, na.rm = T),
              Percent_White = mean(pctWhite, na.rm = T),
              Percent_Bach = mean(pctBachelors, na.rm = T),
              Percent_Poverty = mean(pctPoverty, na.rm = T))
```

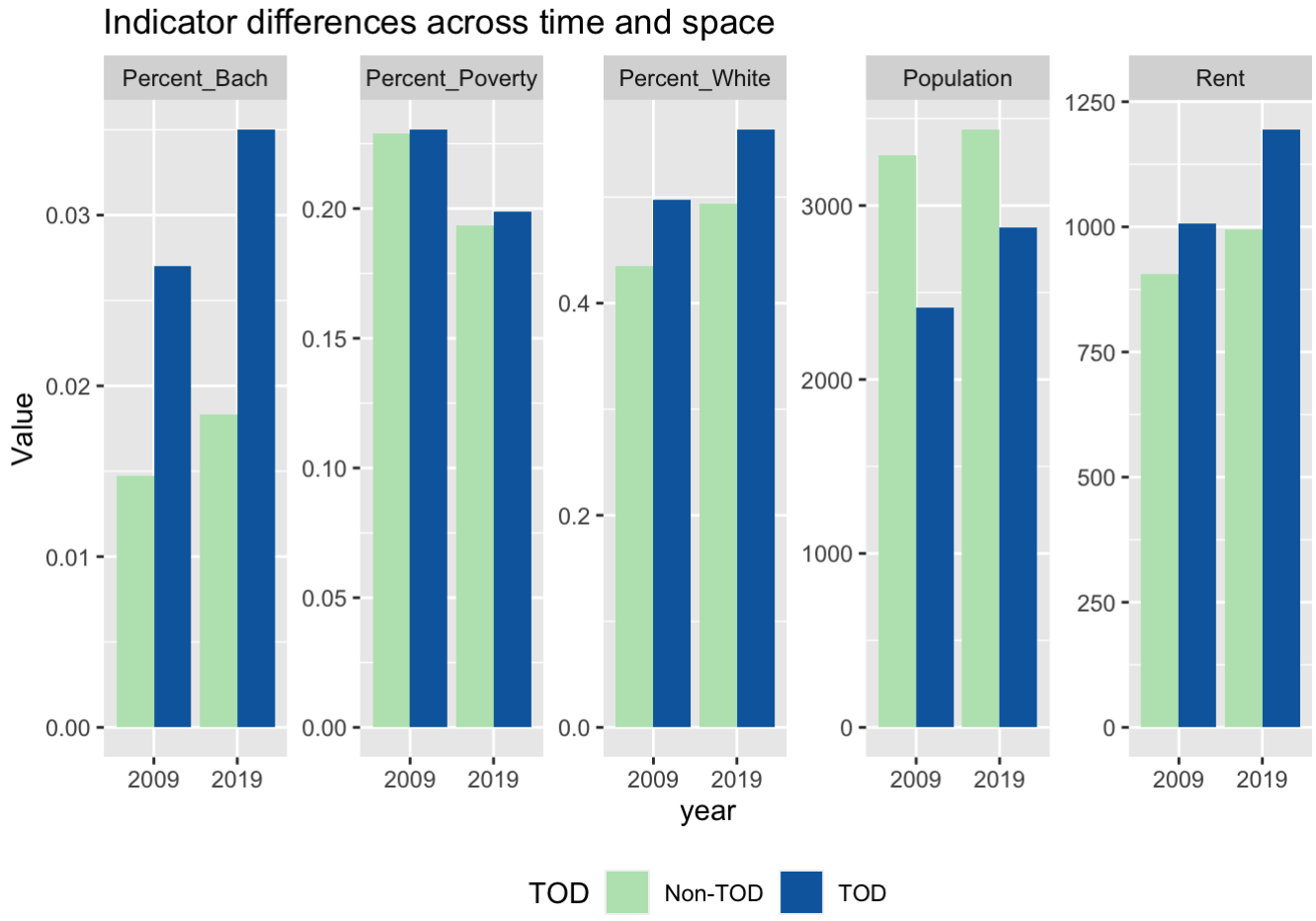
```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
# Creates a cleaner table for summary statistics
kable(allTracts_chicago.Summary) %>%
  kable_styling() %>%
  footnote(general_title = "\n",
           general = "Table 1.2")
```

year	TOD	Rent	Population	Percent_White	Percent_Bach	Percent_Poverty
2009	Non-TOD	905.5855	3288.106	0.4351066	0.0147097	0.2287846
2009	TOD	1006.9391	2415.240	0.4974378	0.0270132	0.2303972
2019	Non-TOD	995.0058	3436.045	0.4936661	0.0183328	0.1934790
2019	TOD	1194.0496	2876.618	0.5635410	0.0350023	0.1989863

Table 1.2

```
# Visualizes the changes in indicators with bar charts
allTracts_chicago.Summary %>%
  gather(Variables, Value, -year, -TOD) %>%
  ggplot(aes(year, Value, fill = TOD)) +
    geom_bar(stat = "identity", position = "dodge") +
    facet_wrap(~Variables, scales = "free", ncol=5) +
    scale_fill_manual(values = c("#bae4bc", "#0868ac")) +
    labs(title = "Indicator differences across time and space")  + theme(legend.position
="bottom")
```



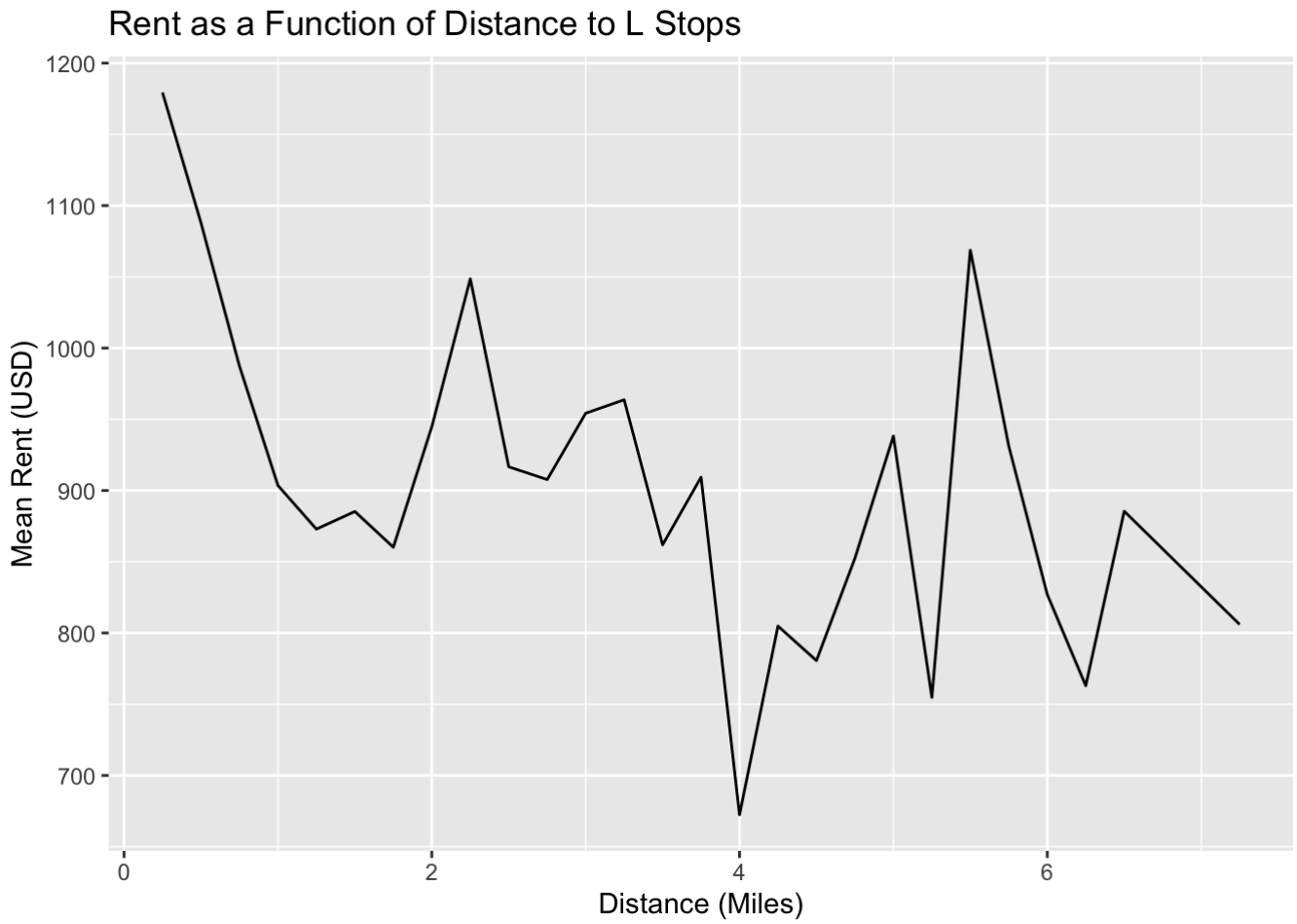
The overall percentage in poverty were similar in both years with an overall reduction in 2019. The growth in the difference for bachelor percentage is stark from 2009 to 2019 with TOD tracts apparently attracting higher percentages of residents with bachelors. The overall percentage in poverty were similar in both years with an overall reduction in 2019. The difference in percentage of white residents stayed roughly the same at approximately 7%. The charts above show that difference in rent prices in TOD and non-TOD tracts grew significantly from 2009 to 2019. In this 10 year time frame the difference in TOD and non-TOD rent prices doubled.

To further explore the relationship between rent prices and distance to train stops, a line plot is created recording the rent prices at quarter mile intervals. This is done using the multiple rings function defined in the source document from Ken Steif read to the document in with the first chunk.

```
allTracts_chicago.rings <-
  st_join(st_centroid(dplyr::select(chicago19, GEOID, year)),
          multipleRingBuffer(st_union(cta_stops), 47520, 5280/4)) %>% # Creates quarter
mile rings around transit stops
  st_drop_geometry() %>%
  left_join(dplyr::select(allTracts_chicago, GEOID, MedRent, year),
            by=c("GEOID"="GEOID", "year"="year")) %>%
  st_sf() %>%
  mutate(distance = distance / 5280) #convert to miles
```

```
## Warning: st_centroid assumes attributes are constant over geometries
```

```
medRent_Distance_chicago <-  allTracts_chicago.rings %>%
  group_by(distance) %>%
  summarize(mean_rent = mean(MedRent, na.rm = TRUE))


ggplot() +

  geom_line(data = medRent_Distance_chicago, aes(x = distance, y = mean_rent)) +
  xlab("Distance (Miles)") + ylab("Mean Rent (USD)") + ggtitle("Rent as a Function of Di
stance to L Stops")
```



The above plot shows that rent prices increase dramatically as you get within a mile of L stations indicating high demand for TOD in Chicago. The relationship becomes more chaotic after a mile with the effect distance to stops has on rent price diminishing after it is no longer a reasonable walk.

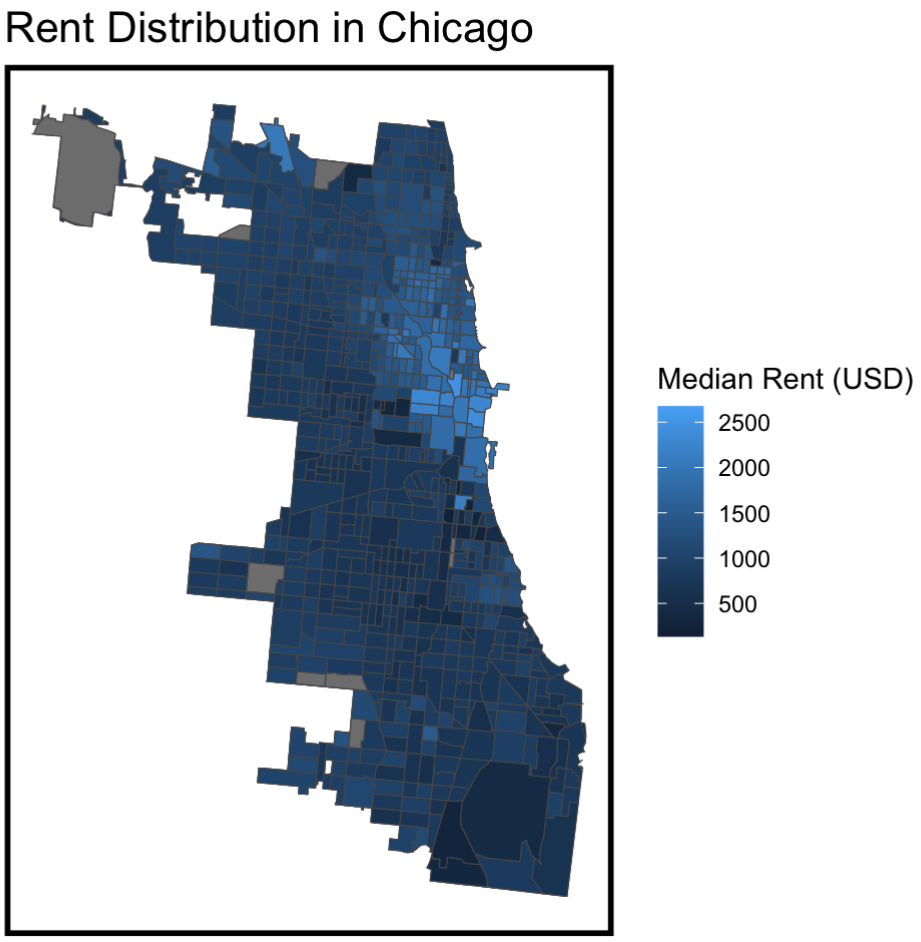
Discussion

While analysis of indicator variables are not conclusive, they would seem to suggest the City of Chicago's ETOD plan has achieved significant success in encouraging the growth of TOD. The population and rent price increasing in TOD tracts in relation to non-TOD tracts indicates increasing demand for TOD in Chicago. While the indicator variables suggest success for TOD in Chicago whether this development is equitable or not could be called into question. White residents continue to be over-represented in TOD tracts suggesting racial inequalities for transit access. While the rising rent prices near transit stations and the increased number highly educated residents indicate success in increasing demand for TOD and perceived attraction to this development style, it presents unique problems in ensuring all socioeconomic groups can benefit from access to public transportation. The City of Chicago should place special emphasis on their efforts to preserve and facilitate the development of affordable housing within TOD zones as laid out in their plan by persevering existing affordable housing and modifying City code to include preference for affordable housing in TOD zones (policy page 12)

Limitations

One key limitation is that indicator variables simply look at snapshots in time rather than an advanced statistical analysis factoring in variance and statistical significance. Another limitation is geographical bias that can be hidden in the ACS data. For instance,looking at a Rent map, we can see that the area near Downtown has much higher rents. This could influence of some of the high value associated with TOD regions since Downtown has much more to offer than simple transit access. Whether Chicagoans are gravitating towards TOD areas for other reasons like better job markets or increased amenities is difficult to discern. Regardless of the reason though, it is apparent that TOD is becoming increasingly popular and cities across the nation should look to Chicago as an example for successful TOD.

```
ggplot() +
  geom_sf(data = allTracts_chicago, aes(fill = MedRent)) +
  labs(fill = "Median Rent (USD)") + ggtitle("Rent Distribution in Chicago") +
  mapTheme()
```



References

[1] EQUITABLE TRANSIT-ORIENTED DEVELOPMENT. Chicago.gov, City of Chicago, chicago.gov/city/en/sites/equitable-transit-oriented-development/home.htmlchicago.gov/city/en/sites/equitable-transit-oriented-development/home.html.

[2] EQUITABLE TRANSIT-ORIENTED DEVELOPMENT (ETOD) POLICY PLAN. City of Chicago.

[3] Steif, K. (2021). Public Policy Analytics: Code and Context for Data Science in Government. CRC Press.