---
title: "Census Mapping DMV"
author: "Tyler Sanders"
date: "November 28, 2018"
output:
  pdf_document: default
  html_document: default

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

###DATA
##Prerequisites
```{r}
library(modelr)
library(tidyverse)
library(leaflet)
library(tigris)
library(acs)
library(ggplot2)
library(IDPmisc)
library(plotly)
api.key.install("80d68bbbdc19c70a4c2c9f188638956df52d8e87")
```

##Data Tidying for Commuting in DC

###Mapping with Leaflet & Tigris
##Public Transit Use for Commuting DMV(variable = B08134_061)
```{r}
MD <- tracts(state = 'MD')
plot(MD)
my_states <- c('MD', 'VA', 'DC')
dmv <- counties(state = my_states)
plot(dmv)


where_i_want <- c("MD", "DC", "VA")

##my_counties <- c("51107", 51177 51179 24003 51600 51510 24017 24009 24031 51013 51153 24037 51069 51610 24021 24510 24027 51059 ##24033 24005
plot(dmv)
dmv <- tracts(state = my_states)

dmv_ptu_data <- acs.fetch(endyear = 2012,
                          geography = geo.make(state = (my_states),
                                               county = "*"),
                           variable ="B08134_061")

dmv_ptu_df <- data.frame(paste0(as.character(dmv_ptu_data@geography$state),
                                      as.character(dmv_ptu_data@geography$county),
                                 dmv_ptu_data@geography$tract),
                         dmv_ptu_data@estimate)

colnames(dmv_ptu_df) <- c("GEOID", "use.transit")

dmv_ptu_merged <- geo_join(dmv, dmv_ptu_df, "GEOID", "GEOID")

pal <- colorQuantile("Greens", NULL, n = 4)

popup <- paste0("Number of residents who commute using Public Transit: ", as.character(dmv_ptu_df$use.transit))

map_dmv_ptu <- leaflet() %>%
  addProviderTiles("CartoDB.Positron") %>%
  addPolygons(data = dmv_ptu_merged,
              fillColor = ~pal(dmv_ptu_df$use.transit),
              fillOpacity = 0.7,
              weight = 0.2,
              smoothFactor = 0.2,
              popup = popup) %>%
  addLegend(pal = pal,
            values = dmv_ptu_df$use.transit,
            position = "bottomright",
            title = "DMV Public Transit Use for Commuting")
```

##Public Transit Use for Commuting Maryland (variable = B08134_061)
```{r}

MD <- tracts(state = 'MD')
plot(MD)

md_ptu_data <- acs.fetch(endyear = 2012,
                         geography = geo.make(state = "MD",
                                              county = "*",
                                              tract = "*"),
                         variable ="B08134_061")

md_ptu_df <- data.frame(paste0(as.character(md_ptu_data@geography$state),
                               as.character(md_ptu_data@geography$county),
                               md_ptu_data@geography$tract),
                        md_ptu_data@estimate)

colnames(md_ptu_df) <- c("GEOID", "use.transit")

md_ptu_merged <- geo_join(MD, md_ptu_df, "GEOID", "GEOID")

pal <- colorNumeric("Greens", NULL, n = 3)

popup <- paste0("Number of residents who commute using Public Transit: ", as.character(md_ptu_df$use.transit))

map_md_ptu <- leaflet() %>%
  addProviderTiles("CartoDB.Positron") %>%
  addPolygons(data = md_ptu_merged,
              fillColor = ~pal(md_ptu_df$use.transit),
```

```
              fillOpacity = 0.7,
              weight = 0.2,
              smoothFactor = 0.2,
              popup = popup) %>%
  addMarkers(lng = -77.017834, lat = 38.975532) %>%
  addLegend(pal = pal,
            values = md_ptu_df$use.transit,
            position = "bottomright",
            title = "Maryland Public Transit Use")

map_md_ptu



```


##Public Transit Use for Commuting Virginia (variable = B08134_061)
```{r}

VA <- tracts(state = 'VA')
plot(VA)

va_ptu_data <- acs.fetch(endyear = 2012,
                         geography = geo.make(state = "VA",
                                              county = "*",
                                              tract = "*"),
                         variable ="B08134_061")

va_ptu_df <- data.frame(paste0(as.character(va_ptu_data@geography$state),
                               as.character(va_ptu_data@geography$county),
                               va_ptu_data@geography$tract),
                        va_ptu_data@estimate)

colnames(va_ptu_df) <- c("GEOID", "use.transit")

va_ptu_merged <- geo_join(VA, va_ptu_df, "GEOID", "GEOID")

pal <- colorNumeric("Greens", NULL, n = 3)

popup <- paste0("Number of residents who commute using Public Transit: ", as.character(va_ptu_df$use.transit))

#map_va_ptu <- leaflet() %>%
#  addProviderTiles("CartoDB.Positron") %>%
#  addPolygons(data = va_ptu_merged,
#              fillColor = ~pal(va_ptu_df$use.transit),
#              fillOpacity = 0.7,
#              weight = 0.2,
#              smoothFactor = 0.2,
#              popup = popup) %>%
#  addProviderTiles(providers$Esri.NatGeoWorldMap) %>%
#  addCircleMarkers(lat = metro_stops$Y, lng = metro_stops$X, radius = 5, weight = .1, opacity = .1,
#                   label = paste0("Metro Stop: ", as.character(metro_stops$NAME),
#                                  '\r\n-----\r\n',  "\r\n Line: ",
#                                  as.character(metro_stops$LINE))) %>%
#  addLegend(pal = pal,
#            values = va_ptu_df$use.transit,
#            position = "bottomleft",
#            title = "Virginia Public Transit Use")



```


###DC Metro Data
```{r}

#metro_stops <- Metro_Station_Entrances_Regional

#metro_bus <- Metro_Bus_Line


#tmap_md_ptu <- leaflet() %>%
#  addProviderTiles("CartoDB.Positron") %>%
#  addPolygons(data = md_ptu_merged,
#              fillColor = ~pal(md_ptu_df$use.transit),
#              fillOpacity = 0.7,
#              weight = 0.2,
#              smoothFactor = 0.2,
#              popup = popup) %>%
#  addProviderTiles(providers$Esri.NatGeoWorldMap) %>%
#  addCircleMarkers(lat = metro_stops$Y, lng = metro_stops$X, radius = 5, weight = .1, opacity = .1,
#                   label = paste0("Metro Stop: ", as.character(metro_stops$NAME),
#                                  '\r\n-----\r\n',  "\r\n Line: ",
#                                  as.character(metro_stops$LINE))) %>%
#  addLegend(pal = pal,
#            values = md_ptu_df$use.transit,
#            position = "bottomleft",
#            title = "Maryland Public Transit Use")

# tmap_md_ptu


?addCircleMarkers

```

```{r}


##building dataset

testadd_data <- acs.fetch(endyear = 2012,
                          geography = geo.make(state = (my_states),
                                               county = "*"),
```

```
                            variable ="B19013_001") #hhincome

testadd_df <- data.frame(paste0(as.character(testadd_data@geography$state),
                                      as.character(testadd_data@geography$county),
                                 testadd_data@geography$tract),
                         testadd_data@estimate)

colnames(testadd_df) <- c("GEOID", "hhincome")



dmv_ptu_data

test2 <- left_join(dmv_ptu_df, testadd_df, "GEOID")

colnames(test2) <- c("GEOID", "usetransit", "hhincome")

full_dmv_data <- test2

#Buiding Data
#add new variable repeated


#B02001_001 total population
#B02009_001 Black pop. total
#B03001_003 hispanic or latino total

#B19013_001 household income
#B05010_002 below poverty line

#B08136_001 total average travel time
#B08136_007 avg travel time pub trans
#B08134_061 use public transit




#DP03_0021 #percent commuting to work
#DP05_0065  percent latino


##Geode

va <- (51)
plot(va)

#va_tracts <- counties(51, county = "Fairfax County")
plot(va_tracts)
?tracts
silver_counties <- c(059, 013, 107, 510)

va_s <-  tracts(state = 51, county = silver_counties)



##Total Population



add_data <- acs.fetch(endyear = 2012,
                            geography = geo.make(state = 51,
                                                 county = silver_counties,
                                                 tract = "*"),
                             variable ="B02001_001") #total population

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                      as.character(add_data@geography$county),
                                 add_data@geography$tract),
                           add_data@estimate)

colnames(add_df) <- c("GEOID", "pop12")
hold <- add_df
hold <- left_join(va_s, add_df, "GEOID")
data <- hold

###
add_data <- acs.fetch(endyear = 2014,
                            geography = geo.make(state = 51,
                                                 county = silver_counties,
                                                 tract = "*"),
                             variable ="B02001_001") #total population

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                      as.character(add_data@geography$county),
                                 add_data@geography$tract),
                           add_data@estimate)

colnames(add_df) <- c("GEOID", "pop14")
hold <- left_join(data, add_df, "GEOID")
data <- hold

add_data <- acs.fetch(endyear = 2016,
                            geography = geo.make(state = 51,
                                                 county = silver_counties,
                                                 tract = "*"),
                             variable ="B02001_001") #total population

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                      as.character(add_data@geography$county),
                                 add_data@geography$tract),
                           add_data@estimate)

colnames(add_df) <- c("GEOID", "pop16")
hold <- left_join(data, add_df, "GEOID")
data <- hold
```

```
### use transit ###

add_data <- acs.fetch(endyear = 2012,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B08134_061") #use transit

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "transit12")
hold <- left_join(data, add_df, "GEOID")
data <- hold

###

add_data <- acs.fetch(endyear = 2014,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B08134_061")

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "transit14")
hold <- left_join(data, add_df, "GEOID")
data <- hold

###

add_data <- acs.fetch(endyear = 2016,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B08134_061")

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "transit16")
hold <- left_join(data, add_df, "GEOID")
data <- hold

###hhincome###
add_data <- acs.fetch(endyear = 2012,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B19013_001") #hhincome

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "hhincome12")
hold <-  left_join(data, add_df, "GEOID")
data <- hold

##

add_data <- acs.fetch(endyear = 2014,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B19013_001")

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "hhincome14")
hold <- left_join(data, add_df, "GEOID")
data <- hold

##

add_data <- acs.fetch(endyear = 2016,
                          geography = geo.make(state = 51,
                                               county = silver_counties,
                                               tract = "*"),
                           variable ="B19013_001")

add_df <- data.frame(paste0(as.character(add_data@geography$state),
                                    as.character(add_data@geography$county),
                               add_data@geography$tract),
                         add_data@estimate)

colnames(add_df) <- c("GEOID", "hhincome16")
hold <- left_join(data, add_df, "GEOID")
data <- hold
```

```
###pcttransit###

test <- mutate(hold, pcttransit12 = transit12 / pop12)
data <- test
test <- mutate(data, pcttransit14 = transit14 / pop14)
data <- test
test <- mutate(data, pcttransit16 = transit16 / pop16)
data <- test
data$incomegrowth <- ((data$hhincome16 - data$hhincome12) / data$hhincome12)
silver_income <- data
###gather and tidy###

gt <- test %>% gather(`pcttransit12`, `pcttransit14`, `pcttransit16`, key = "year", value = "pcttransit") #this worked first try!!

ready <- na.exclude(gt) #14 obs. excluded with NA in pctransit as a result of a pop listing of 0.
ready <- na.exclude(silver_income)

silver_income_change$before <- ifelse(silver_income_change$year == "pcttransit12" | silver_income_change$year == "pctusetransit14", "before",
"after") #after silver line construction == 1
ready$silver <- ifelse(ready$GEOID == "5159471201" | ready$GEOID == "5159482100" | ready$GEOID == "5159480202" | ready$GEOID == "5159482202" |
ready$GEOID == "5159471202" | ready$GEOID == "5159481900" | ready$GEOID == "5159482203" |
ready$GEOID == "5159481202" | ready$GEOID == "5159482201" | ready$GEOID == "5159482302" | ready$GEOID == "5159481400" | ready$GEOID == "5159470600"
| ready$GEOID == "5159470500" | ready$GEOID == "5159480201" | ready$GEOID == "5159482301" | ready$GEOID == "5159482303" | ready$GEOID ==
"5159480300" | ready$GEOID == "5159482400" | ready$GEOID == "5159470800" | ready$GEOID == "5159482001", 1, 0)
##https://www.fairfaxcounty.gov/demographics/interactive-map-block-groups##

ready$growth <- ((ready$pcttransit16 - ready$pcttransit12) / ready$pcttransit12)

silver_test <- silver

silver_test$growth <- ((silver_test$transit16 - silver_test$transit12) / silver_test$transit12)

silver_growth <- silver_test

silver_growth <- ready

silver_income <- ready

write.csv(silver_growth, 'silver_growth.csv')




saveRDS(silver, "silver.R")

load("silver.R")
write.csv(silver,'silver.csv')



silver <- read.csv("silver.csv") #final data for first regression
silver_growth <- read.csv("silver_growth.csv") #final data for growth regression


silver_income_change <- read.csv("silver_income.csv")

silver_income_change <- silver_income_change %>% gather(`pcttransit12`, `pcttransit14`, `pcttransit16`, key = "year", value = "pcttransit")

silver_income_change$income <- ((silver_income_change$hhincome16 - silver_income_change$hhincome12) / silver_income_change$hhincome12 *100)

silver_income_change$after <- ifelse(silver_income_change$year == "pcttransit16", 1, 0)


### heterogenious treatment effects ###



lm <- lm(pcttransit ~ silver + after + silver * after, data = silver)
silver <- silver %>% add_predictions(lm)
summary(lm) # intercept and silver variables significant at 99%

lm_silver <- lm

ggplot(silver, aes(lm$fitted.values)) +
  geom_point() +
  geom_abline(aes(lm))

ggplot(data = silver, aes(pop16, pcttransit)) +
  abline(lm(pcttransit ~ silver + after + silver * after, data = silver))

ggplot(data = silver, aes(pcttransit, pred)) +
  geom_smooth(method='lm')




###growth regression###

silver_growth <- NaRV.omit(silver_growth)
sna.omit(silver_growth)

growth_lm <- lm(growth ~ silver + after + silver*after, data = silver_growth)
silver_growth <- silver_growth %>% add_predictions(growth)
summary(growth)

ggplot(data = silver, aes(pcttransit, pred)) +
  geom_smooth(method='lm') +
  geom_point()

plot(lm)
```

```
###income growth regression###

silver_income_lm <- lm(income ~ silver + after + silver*after, data = silver_income_change)

income_growth <- lm(income ~ silver, data = silver_income)
summary(income_growth)

silver_income <- silver_income %>% add_predictions(income_growth)

write.csv(silver_income, "silver_income.csv")


###Linear Regressions###

summary(lm_silver)

summary(growth_lm)


summary(silver_income_lm)



```

```{r}

plot(silver$pcttransit, silver$silver, xlim = c(0, 1.05), ylim = c(-1,1),
     col = ifelse(silver$silver == "1", "red", "blue"),
     xlab = "X Axis",
     ylab = "Y Axis",
     main = "Title") +
abline(lm_silver) + #add regression line
abline(v = 0, lty = "dashed")



p <- ggplot(silver_income_change, aes(pcttransit, silver, colour=silver)) +
  geom_point() +
  geom_jitter()
p

p + labs(aesthetic='custom text')


p + scale_fill_discrete(name= "Test")

color= ifelse(silver_income_change$silver,'Next to Silver Line', 'Control Population')



ggplot(silver_income_change, aes(pcttransit, before, col = silver, "red")) +
  geom_point(col = "black") +
  geom_line(col = "red") +
  geom_jitter() +
  xlab("Percentage of Residents Who Commute with Transit") +
  ylab("Difference in Differences post Silver Line Opening") +
  ggtitle("Measuring the Impact of Silver Line Phase 1 on Commuter Behavior"))


ggplot(silver_income_change, aes(pcttransit, before, col = silver, col = "red")) +
  geom_point() +
  geom_line(col = "red") +
  geom_jitter() +
  xlab("Percentage of Residents Who Commute with Transit") +
  ylab("Difference in Differences post Silver Line Opening") +
  ggtitle("Measuring the Impact of Silver Line Phase 1 on Commuter Behavior")
```