

Hi!

I'm Aseem Deodhar.

In Fall 2018 I took the same class you are taking right now!

I graduated in May 2020 with a Master's degree in Urban Informatics.

I work at the MAPC - Metropolitan Area Planning Council as a Research Analyst in the Data Services Department.



## Data View

Map

## Table

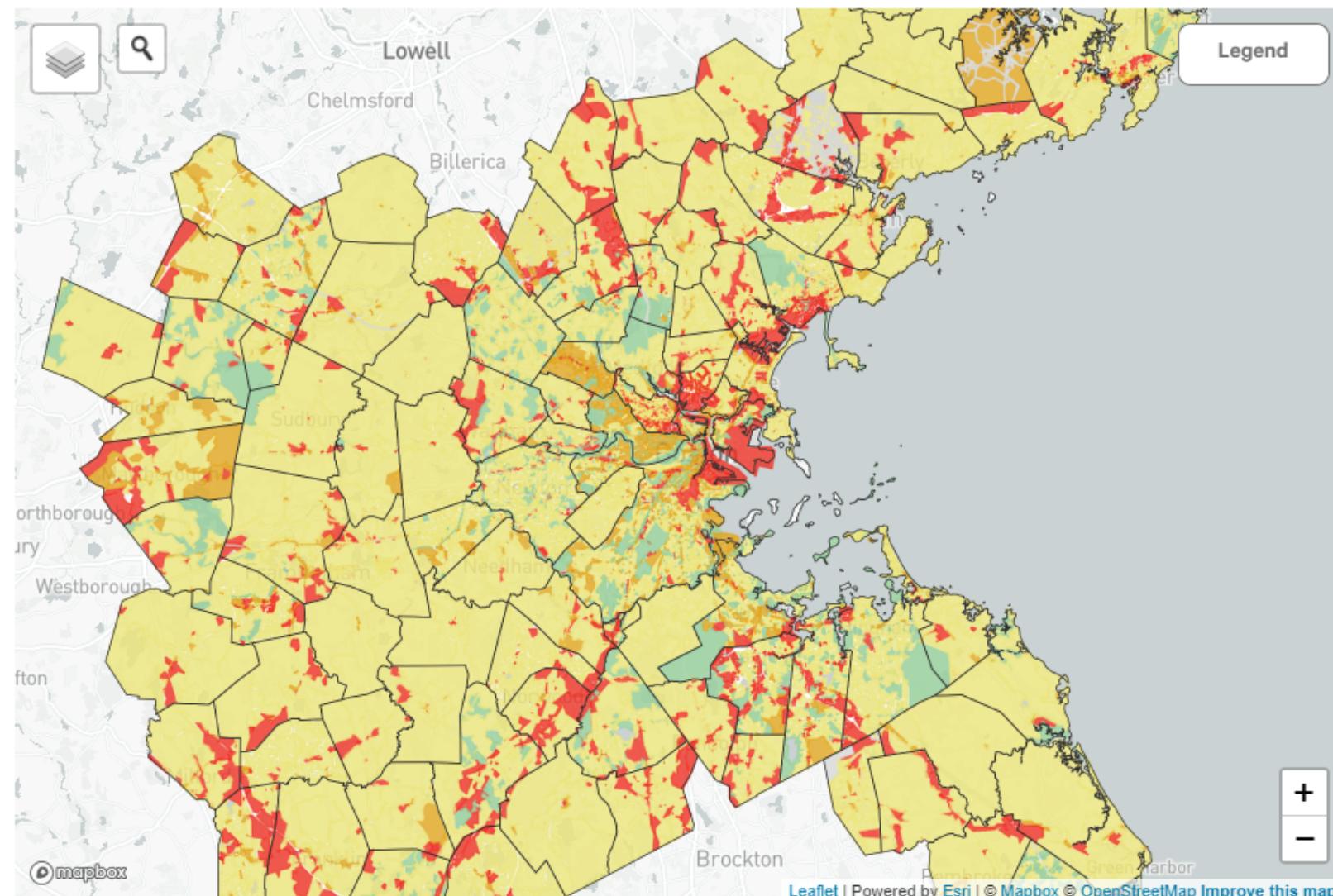
## Export

.CSV

.shp

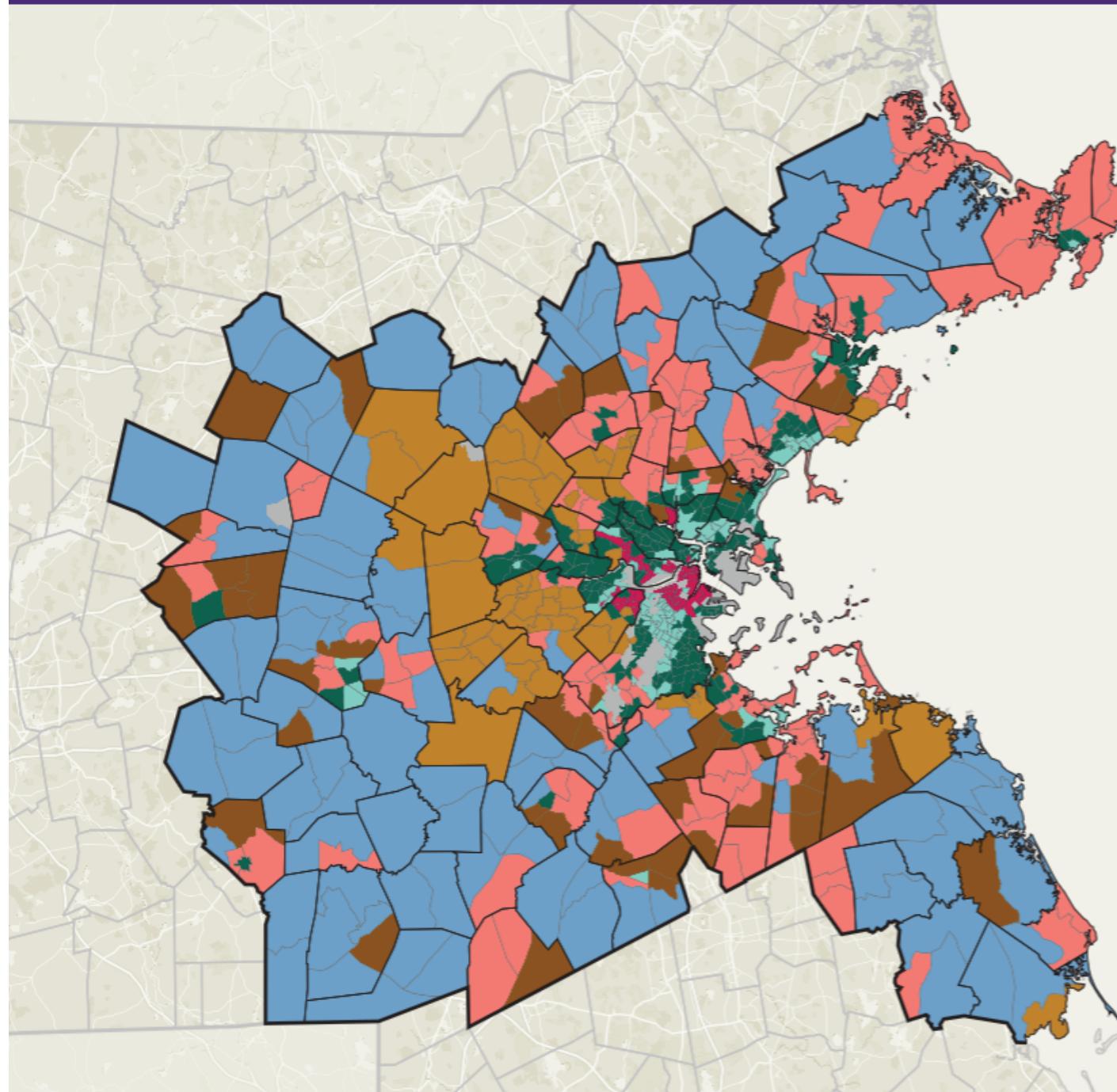
## Filters

Municipality		0
Zone Use Type	Type here to search filters by town/city	
Multifamily Housing (2+ Units)		
Lot Details		
	Acton	+
	Arlington	+
	Ashland	+
	Bedford	+
	Bellingham	+
	Belmont	+
	Beverly	+
	Bolton	+
	Boston	+
	Boxborough	+
	Braintree	+
	Brookline	+
	Burlington	+
	Cambridge	+
	Canton	+
	Carlisle	+
	Chelsea	+
	Cohasset	+



## Contribute





## Submarket 1

High-Density Urban, High Prices

### Key Highlights

Premium-priced

Increasing home prices and rents

Declining numbers of renters

[View Full Profile >>](#)

1

2

3

4

5

6

7

[Read Policy Strategy >>](#)



# Administrative Data

Data generally collected and maintained by **government entities**.

Used for administrative record keeping, policy development, making statistical estimates, etc..

Administrative data are **inherently spatial** and will always be associated with some geography.

## Examples:

Property Tax Records

Birth and Death records

Census Data

Employment and Labor Records

Land Parcel Assessment

Building Permits Data

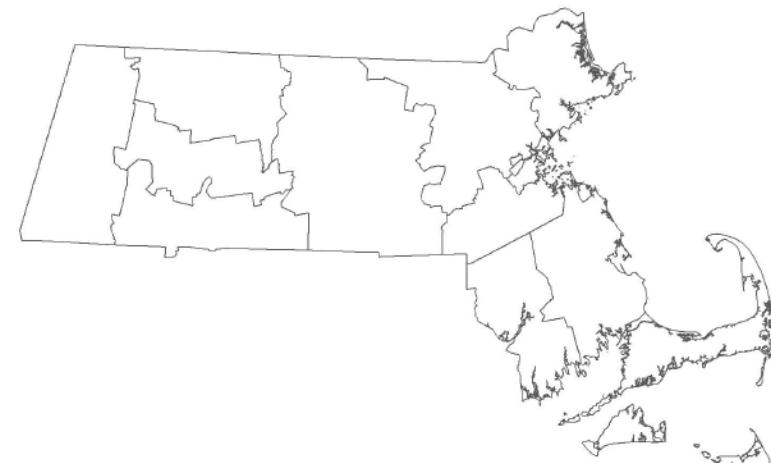
# Administrative Geographies

## Political Geographies

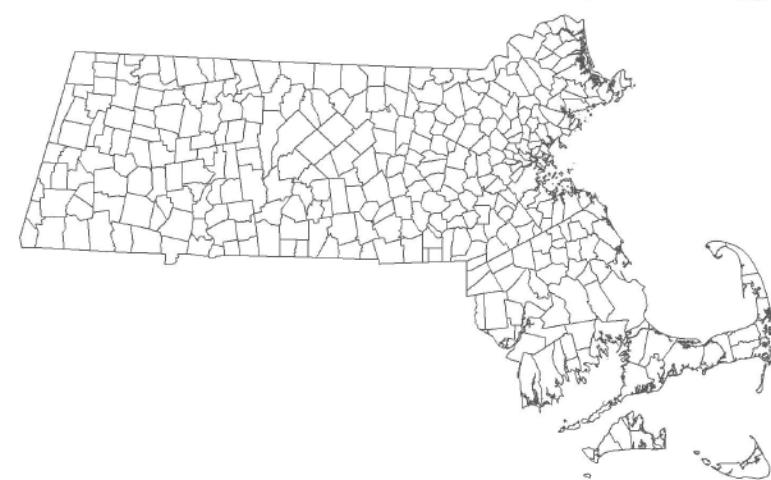
**State**  
FIPS Code: 2 digit  
eg: MA - 25



**County**  
FIPS Code: 3 digit  
eg: Suffolk County - 025



**County Subdivision**  
FIPS Code: 5 digit  
eg: Boston City - 07000



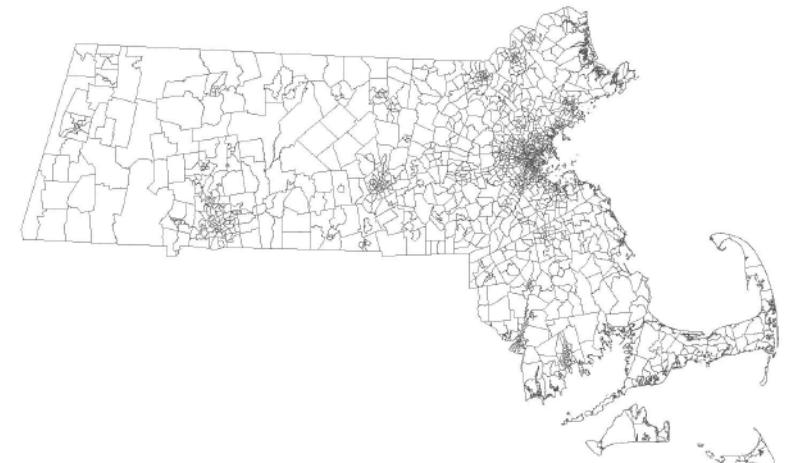
**Boston GEOID:**  
2502507000

## Census Geographies

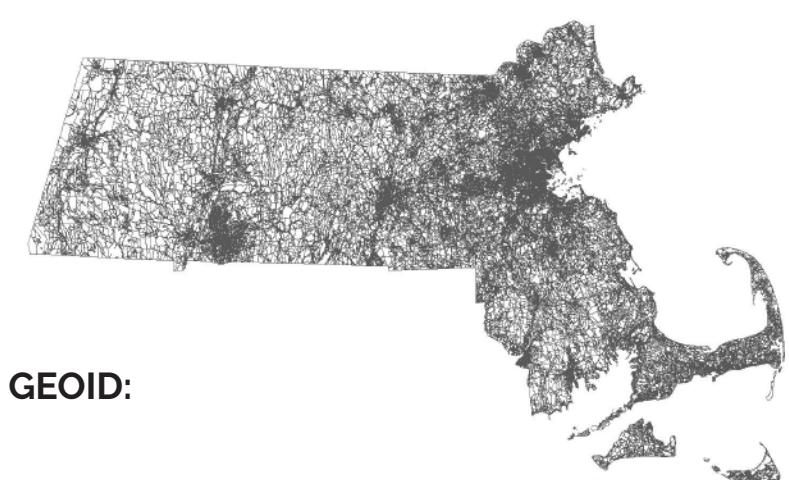
**State**  
FIPS Code: 2 digit



**Census Tract**  
4 digits CT code +  
2 digit decimal suffix  
eg: Northeastern  
University is in 0104.05



**Census Block**  
4 digits Census Block code  
eg: Snell Library is in 1001



**Snell Library Census Block GEOID:**  
250250104051001

# Administrative Data Sources

## ANSI FIPS Codes

Standardized numeric or alphabetical codes for identifying geographies

## MAPC

Metropolitan Boston data sources and analyses

## Census Data

Decennial Census & ACS (1 & 5 year)

## Bureau of Labor Statistics

Employment, Labor Market rates and counts

## LEHD LODES Data

Longitudinal Employer-Household Dynamics  
Mobility and employment data

## MBTA GTFS Feed

Transit systems schedules and related resources

## MassGIS and other local resources

Repository of state and local spatial and statistical data for Massachusetts

<https://www.census.gov/library/reference/code-lists/ansi.html>

<https://www.mapc.org/learn/data-tools/>  
<https://datacommon.mapc.org/>

<https://data.census.gov/cedsci/>

<https://www.bls.gov/data/tools.htm>

<https://lehd.ces.census.gov/data/>

<https://www.mbta.com/developers/gtfs>

<https://www.mass.gov/orgs/massgis-bureau-of-geographic-information>

# Continuous Data

Does not have abrupt changes from value to value.

Can have an infinite number of values between two numbers.

Always numeric. Remember that date/time data are technically numeric.

## Examples:

Population Density

Average Lot size



# Categorical Data

Contains a finite number of distinct data groups or *categories*.

Generally cannot be ranked.

Note: categorized continuous data can possibly be ranked.

Can have numeric *value*, but is treated non-numerically.

## Examples:

Municipality

Street Name

MBTA Bus Route



# Categorizing Continuous Data

Continuous data is divided into groups of ranges.

Groups determined by:

- a. Number of desired groups
- b. Quantities represented in each group.

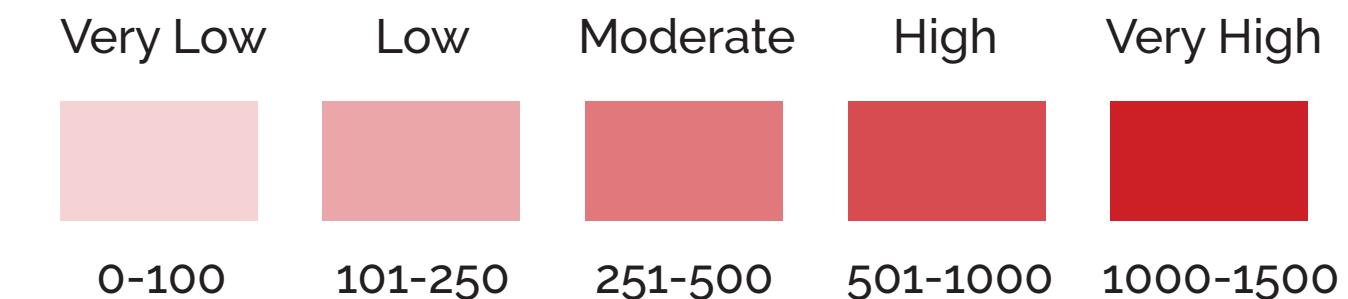
Have a set low and high value.

Groups can be ranked.

## Examples:

Population Density

Average Lot size



# Spatial File Structure

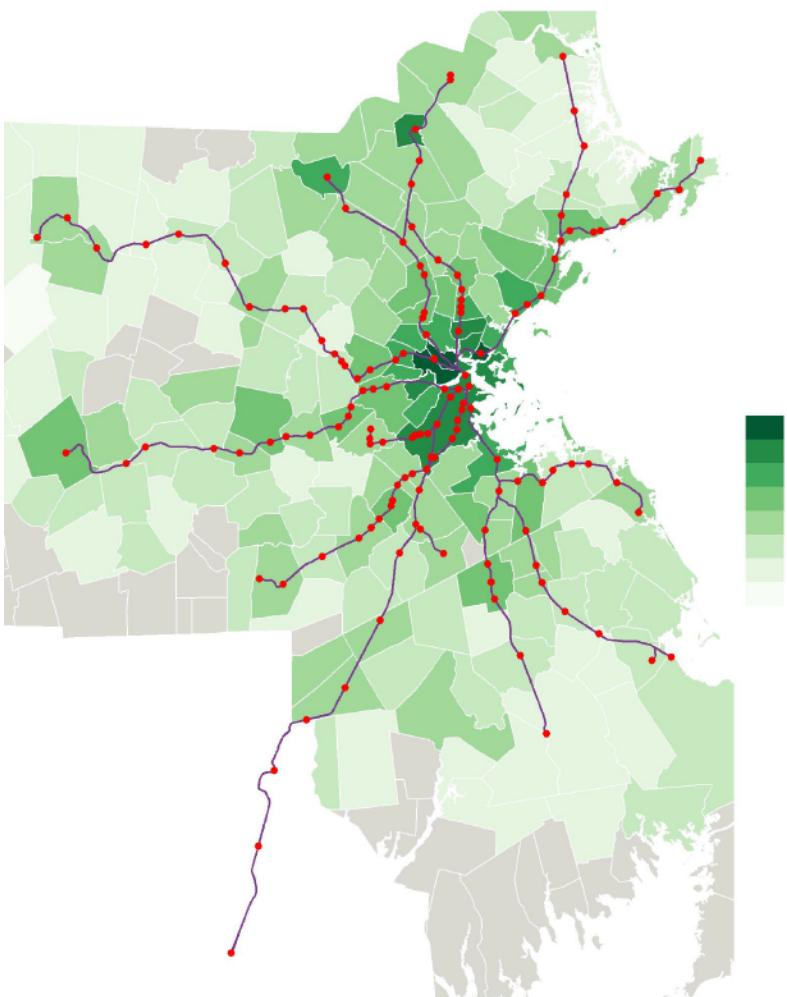
```
> muni_shp
Simple feature collection with 351 features and 11 fields
geometry type: MULTIPOLYGON
dimension: XY
bbox: xmin: 33869.7 ymin: 777514.6 xmax: 330838.8 ymax: 959743.2
projected CRS: NAD83 / Massachusetts Mainland
# A tibble: 351 x 12
  town_id town2  coussubfp10 fips_stco county pop2010 hu2010 area_acres sq_miles hh_dens pop_dens_sqmile geometry
    <int> <chr> <chr>      <int> <chr>     <dbl>   <dbl>      <dbl>   <dbl>      <dbl>      <dbl> <MULTIPOLYGON [m]>
 1    104 Aquinn~ 01585       25007 DUKES      311     503    3692.     5.77     0.618      53.9 (((257628 789763.3, 257707.3 789759.6, 257797~
 2     10 Arling~ 01605       25017 MIDDLE~    42844   19974    3515.     5.49     2.14     7800. (((228132.1 909534.6, 228330.2 909532, 228495~
 3     11 Ashbur~ 01885       25027 WORCES~    6081    2599    26219.    41.0     2.34     148. (((169429.9 937325.2, 169690.4 937076.9, 1700~
 4     12 Ashby   01955       25017 MIDDLE~    3074    1191    15417.    24.1     2.58     128. (((174978.6 939866.5, 175519.2 939842.3, 1756~
 5     13 Ashfie~ 02095       25011 FRANKL~    1737    877    25762.    40.3     1.98     43.2 (((98731.23 924309.3, 98741.76 924304.9, 9874~
 6     14 Ashland 02130       25017 MIDDLE~    16593    609     8235.    12.9     2.51     1290. (((202083.8 892526.1, 202273.5 892427.2, 2022~
 7     15 Athol   02480       25027 WORCES~    11584   5231    21346.    33.4     2.21     347. (((145506.6 931588.9, 145513.4 931544.8, 1459~
 8     16 Attleb~ 02690       25005 BRISTOL   43593   18022    17795.    27.8     2.42     1568. (((222987.8 853069.6, 224212.9 850986.9, 2230~
 9     17 Auburn  02760       25027 WORCES~    16188   6840    10503.    16.4     2.37     986. (((174524.8 886444.1, 174651.8 886058.7, 1746~
10     18 Avon    02935       25021 NORFOLK   4356    1769    2904.     4.54     2.46     960. (((236385 872675.3, 236290.7 872626.8, 236304~
# ... with 341 more rows
> muni_shp %>% st_transform(crs = 4326)
Simple feature collection with 351 features and 11 fields
geometry type: MULTIPOLYGON
dimension: XY
bbox: xmin: -73.50814 ymin: 41.23796 xmax: -69.92798 ymax: 42.88679
geographic CRS: WGS 84
# A tibble: 351 x 12
  town_id town2  coussubfp10 fips_stco county pop2010 hu2010 area_acres sq_miles hh_dens pop_dens_sqmile geometry
    <int> <chr> <chr>      <int> <chr>     <dbl>   <dbl>      <dbl>   <dbl>      <dbl>      <dbl> <MULTIPOLYGON [°]>
 1    104 Aquinn~ 01585       25007 DUKES      311     503    3692.     5.77     0.618      53.9 ((((-70.81138 41.3559, -70.81043 41.35587, -70~
 2     10 Arling~ 01605       25017 MIDDLE~    42844   19974    3515.     5.49     2.14     7800. ((((-71.1581 42.43583, -71.15569 42.4358, -71.~
 3     11 Ashbur~ 01885       25027 WORCES~    6081    2599    26219.    41.0     2.34     148. ((((-71.87301 42.68591, -71.86982 42.68369, -7.~
 4     12 Ashby   01955       25017 MIDDLE~    3074    1191    15417.    24.1     2.58     128. ((((-71.80542 42.70899, -71.79882 42.70879, -7.~
 5     13 Ashfie~ 02095       25011 FRANKL~    1737    877    25762.    40.3     1.98     43.2 ((((-72.73329 42.56275, -72.73316 42.56271, -7.~
 6     14 Ashland 02130       25017 MIDDLE~    16593    609     8235.    12.9     2.51     1290. ((((-71.47473 42.28321, -71.47244 42.28232, -7.~
 7     15 Athol   02480       25027 WORCES~    11584   5231    21346.    33.4     2.21     347. ((((-72.16437 42.63296, -72.16428 42.63257, -7.~
 8     16 Attleb~ 02690       25005 BRISTOL   43593   18022    17795.    27.8     2.42     1568. ((((-71.22285 41.92764, -71.20816 41.90885, -7.~
 9     17 Auburn  02760       25027 WORCES~    16188   6840    10503.    16.4     2.37     986. ((((-71.8086 42.22804, -71.80704 42.22457, -71.~
10     18 Avon    02935       25021 NORFOLK   4356    1769    2904.     4.54     2.46     960. ((((-71.06011 42.10365, -71.06125 42.10321, -7.~
# ... with 341 more rows
> |
```

# Spatial File Structure

```
> muni_shp
Simple feature collection with 351 features and 11 fields
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: 33869.7 ymin: 777514.6 xmax: 330838.8 ymax: 959743.2
projected CRS: NAD83 / Massachusetts Mainland
# A tibble: 351 x 12
  town_id town2  cousubfp10 fips_stco county  pop2010 hu2010 area_acres sq_miles hh_dens pop_dens_sqmile
    <int> <chr>   <chr>     <chr>    <chr>    <dbl>   <dbl>     <dbl>    <dbl>    <dbl>    <dbl>
1     104 Aquinn~ 01585
2      10 Arling~ 01605
3     11 Ashbur~ 01885
4     12 Ashby   01955
5    13 Ashfie~ 02095
6     14 Ashland 02130
7     15 Athol   02480
8    16 Attleb~ 02690
9     17 Auburn  02760
10    18 Avon    02935
# ... with 341 more rows
> muni_shp %>% st_transform(crs = 4326)
Simple feature collection with
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: -73.50811 ymin: 41.35587 xmax: -70.81138 ymax: 42.43583
geographic CRS: WGS 84
# A tibble: 351 x 12
  town_id town2  cousubfp10 fips_stco county  pop2010 hu2010 area_acres sq_miles hh_dens pop_dens_sqmile
    <int> <chr>   <chr>     <chr>    <chr>    <dbl>   <dbl>     <dbl>    <dbl>    <dbl>    <dbl>
* 1     104 Aquinn~ 01585
2      10 Arling~ 01605
3     11 Ashbur~ 01885
4     12 Ashby   01955
5    13 Ashfie~ 02095
6     14 Ashland 02130
7     15 Athol   02480
8    16 Attleb~ 02690
9     17 Auburn  02760
10    18 Avon    02935
# ... with 341 more rows
>
```

The figure displays a map of Massachusetts with a grid overlay. The grid is defined by dashed lines representing bounding boxes for each town in the 'muni\_shp' dataset. The x-axis is labeled 'xmin' at 33869.7 and 'xmax' at 330838.8. The y-axis is labeled 'ymin' at 777514.6 and 'ymax' at 959743.2. The map shows the state's coastline and internal town boundaries.

# Mapping in R

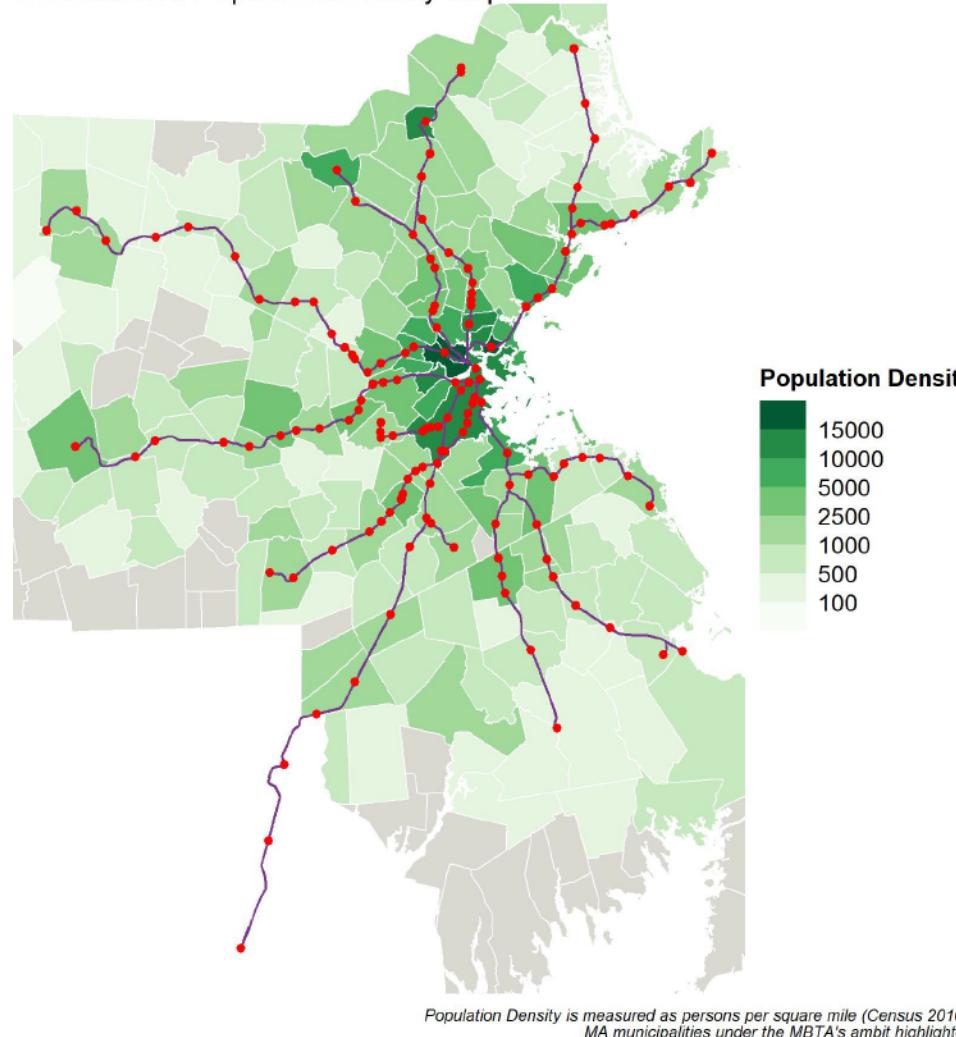


Mystery  
map!

# Mapping in R

## MBTA Commuter Rail Network

Overlaid on a Population Density Map



```
library(tidyverse)
library(sf)
library(ggspatial)

ggplot( )+
  geom_sf(data = mbta_muni_shp,
          aes(geometry = geometry),
          color = 'white',
          fill = '#D9D9D2',
          size = 0.25)+
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),
          aes(geometry = geometry,
              fill = pop_dens_sqmile),
          color = 'white',
          size = 0.25)+
  geom_sf(data = mbta_cr_lines,
          aes(geometry = geometry),
          color = '#7B388C')+
  geom_sf(data = mbta_cr_stops,
          aes(geometry = geometry),
          size = 1,
          color = 'red')+
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),
                       name="Population Density",
                       palette = 'Greens',
                       direction = 1)+

  coord_sf(xlim = c(171357.0, 271642.8),
            ylim = c(814457.5, 950071.0),
            expand = TRUE)+

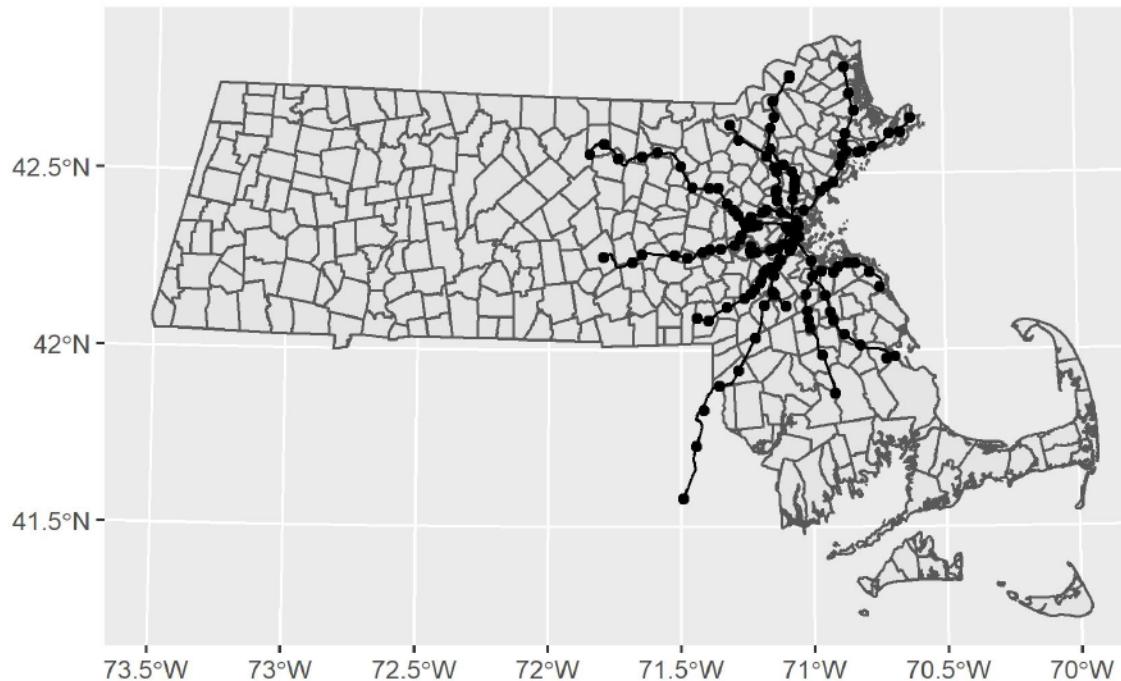
  theme_void()+
  labs(title = 'MBTA Commuter Rail Network',
       subtitle = "Overlaid on a Population Density Map",
       caption = 'Population Density is measured as persons per square mile (Census 2010)\nMA municipalities under the MBTA's ambit highlighted')+


theme(plot.title = element_text(face = 'bold',
                                size = 14),
      plot.subtitle = element_text(face = 'plain',
                                  size = 11),
      plot.caption = element_text(face = 'italic',
                                  size = 6),
      plot.caption.position = 'plot',
      legend.title = element_text(face = 'bold',
                                  size = 9))+

ggsave('mbta_commrail_network.jpeg',
       height = 7,
       width = 5,
       units = 'in',
       dpi = 'retina')
```

# Mapping in R

Identify and plot the spatial components of your map.

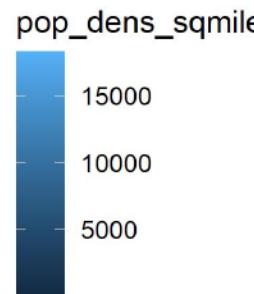
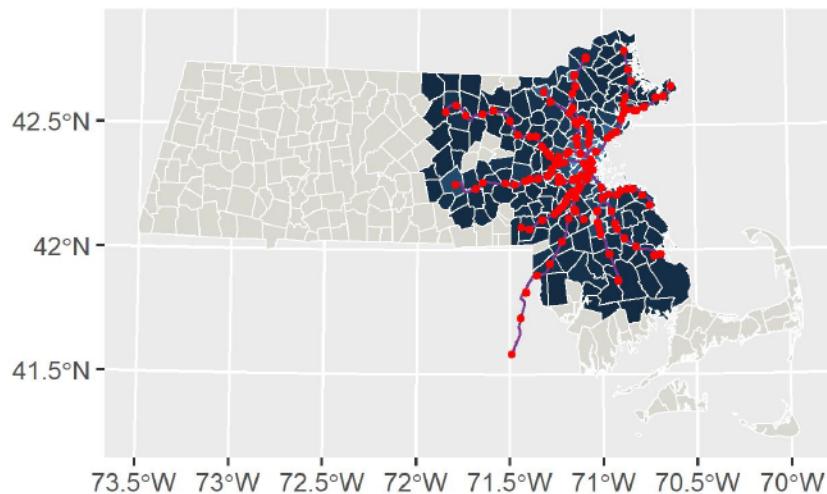


```
library(tidyverse)
library(sf)
library(ggspatial)

ggplot() +
  geom_sf(data = mbta_muni_shp,
          aes(geometry = geometry),
          color = 'white',
          fill = '#D9D9D2',
          size = 0.25) +
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),
          aes(geometry = geometry,
              fill = pop_dens_sqmile),
          color = 'white',
          size = 0.25) +
  geom_sf(data = mbta_cr_lines,
          aes(geometry = geometry),
          color = '#7B388C') +
  geom_sf(data = mbta_cr_stops,
          aes(geometry = geometry),
          size = 1,
          color = 'red') +
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),
                       name="Population Density",
                       palette = 'Greens',
                       direction = 1) +
  coord_sf(xlim = c(171357.0, 271642.8),
           ylim = c(814457.5, 950071.0),
           expand = TRUE) +
  theme_void() +
  labs(title = 'MBTA Commuter Rail Network',
       subtitle = "Overlaid on a Population Density Map",
       caption = 'Population Density is measured as persons per square mile (Census 2010)\n\nMA municipalities under the MBTA\\'s ambit highlighted') +
  theme(plot.title = element_text(face = 'bold',
                                  size = 14),
        plot.subtitle = element_text(face = 'plain',
                                    size = 11),
        plot.caption = element_text(face = 'italic',
                                   size = 6),
        plot.caption.position = 'plot',
        legend.title = element_text(face = 'bold',
                                   size = 9)) +
  ggsave('mbta_commrail_network.jpeg',
         height = 7,
         width = 5,
         units = 'in',
         dpi = 'retina')
```

# Mapping in R

## Adding object aesthetics.

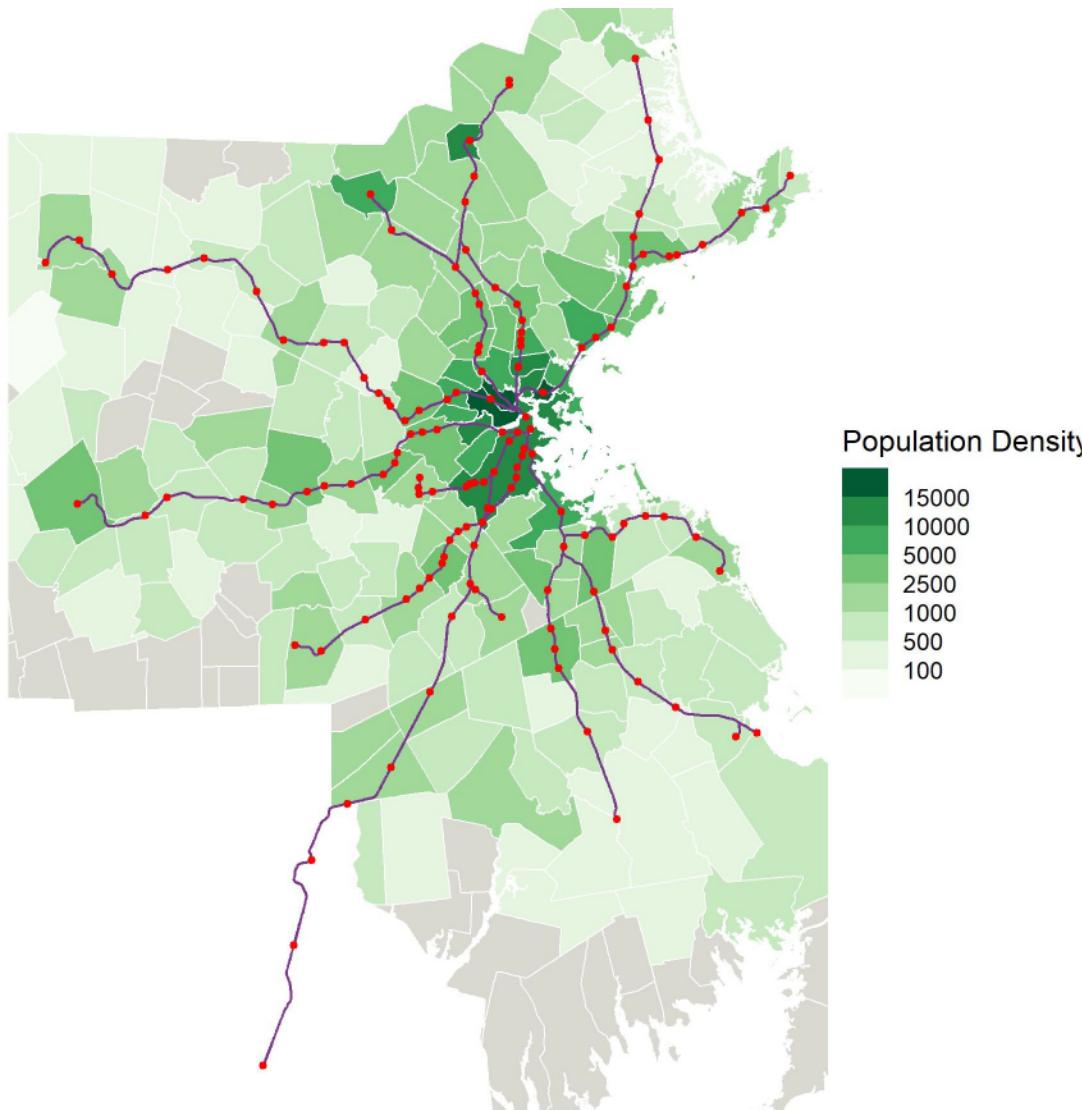


```
library(tidyverse)
library(sf)
library(ggspatial)

ggplot() +
  geom_sf(data = mbta_muni_shp,
          aes(geometry = geometry),
          color = 'white',
          fill = '#D9D9D2',
          size = 0.25) +
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),
          aes(geometry = geometry,
              fill = pop_dens_sqmile),
          color = 'white',
          size = 0.25) +
  geom_sf(data = mbta_cr_lines,
          aes(geometry = geometry),
          color = '#7B388C') +
  geom_sf(data = mbta_cr_stops,
          aes(geometry = geometry),
          size = 1,
          color = 'red') +
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),
                       name="Population Density",
                       palette = 'Greens',
                       direction = 1) +
  coord_sf(xlim = c(171357.0, 271642.8),
           ylim = c(814457.5, 950071.0),
           expand = TRUE) +
  theme_void() +
  labs(title = 'MBTA Commuter Rail Network',
       subtitle = "Overlaid on a Population Density Map",
       caption = 'Population Density is measured as persons per square mile (Census 2010)\n\nMA municipalities under the MBTA\'s ambit highlighted') +
  theme(plot.title = element_text(face = 'bold',
                                  size = 14),
        plot.subtitle = element_text(face = 'plain',
                                    size = 11),
        plot.caption = element_text(face = 'italic',
                                    size = 6),
        plot.caption.position = 'plot',
        legend.title = element_text(face = 'bold',
                                    size = 9)) +
  ggsave('mbta_commrail_network.jpeg',
         height = 7,
         width = 5,
         units = 'in',
         dpi = 'retina')
```

# Mapping in R

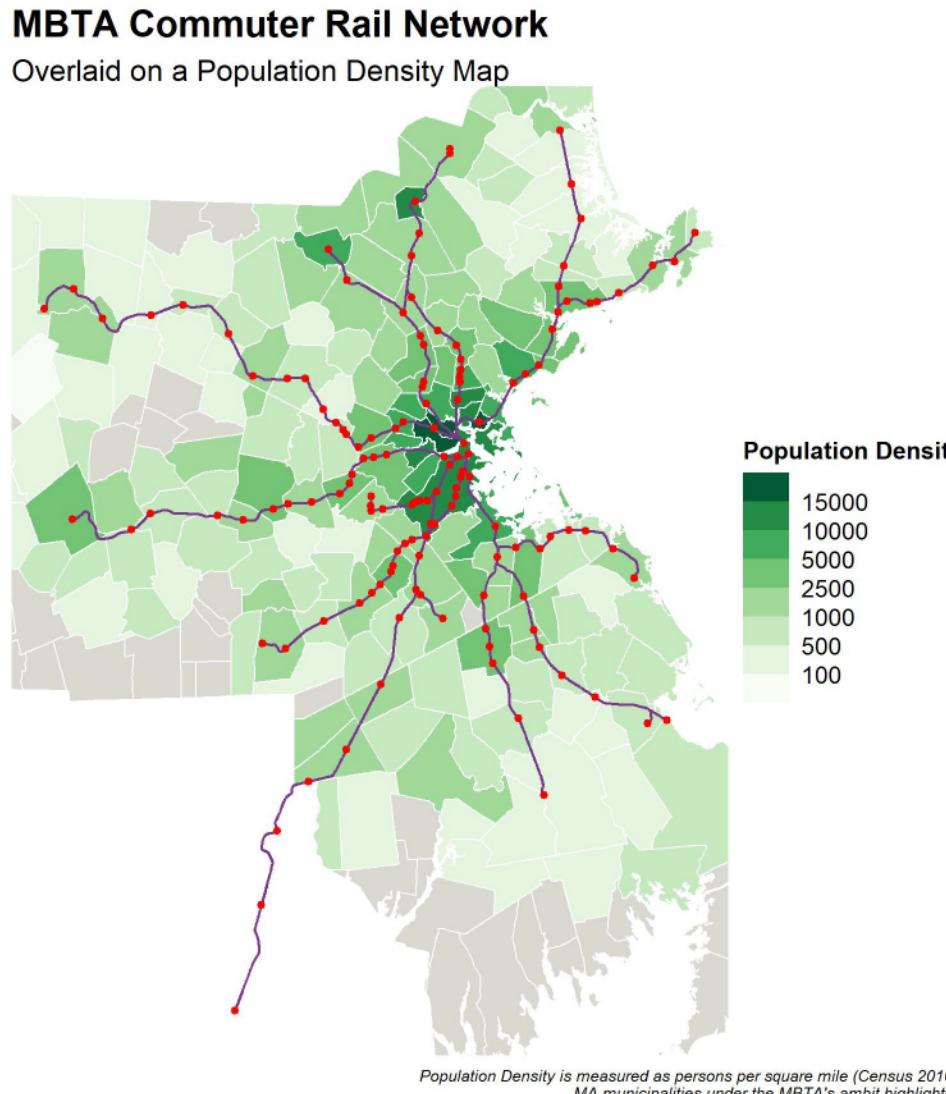
## Rationalizing aesthetics and map extents



```
● ● ●  
library(tidyverse)  
library(sf)  
library(ggspatial)  
  
ggplot() +  
  
  geom_sf(data = mbta_muni_shp,  
          aes(geometry = geometry),  
          color = 'white',  
          fill = '#D9D9D2',  
          size = 0.25) +  
  
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),  
          aes(geometry = geometry,  
              fill = pop_dens_sqmile),  
          color = 'white',  
          size = 0.25) +  
  
  geom_sf(data = mbta_cr_lines,  
          aes(geometry = geometry),  
          color = '#7B388C') +  
  
  geom_sf(data = mbta_cr_stops,  
          aes(geometry = geometry),  
          size = 1,  
          color = 'red') +  
  
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),  
                       name = "Population Density",  
                       palette = 'Greens',  
                       direction = 1) +  
  
  coord_sf(xlim = c(171357.0, 271642.8),  
           ylim = c(814457.5, 950071.0),  
           expand = TRUE) +  
  
  theme_void() +  
  
  labs(title = 'MBTA Commuter Rail Network',  
       subtitle = "Overlaid on a Population Density Map",  
       caption = 'Population Density is measured as persons per square mile (Census 2010)  
       \n MA municipalities under the MBTA\\'s ambit highlighted') +  
  
  theme(plot.title = element_text(face = 'bold',  
                                    size = 14),  
        plot.subtitle = element_text(face = 'plain',  
                                    size = 11),  
        plot.caption = element_text(face = 'italic',  
                                    size = 6),  
        plot.caption.position = 'plot',  
        legend.title = element_text(face = 'bold',  
                                    size = 9)) +  
  
  ggsave('mbta_commrail_network.jpeg',  
         height = 7,  
         width = 5,  
         units = 'in',  
         dpi = 'retina')
```

# Mapping in R

## Labelling and titling the map

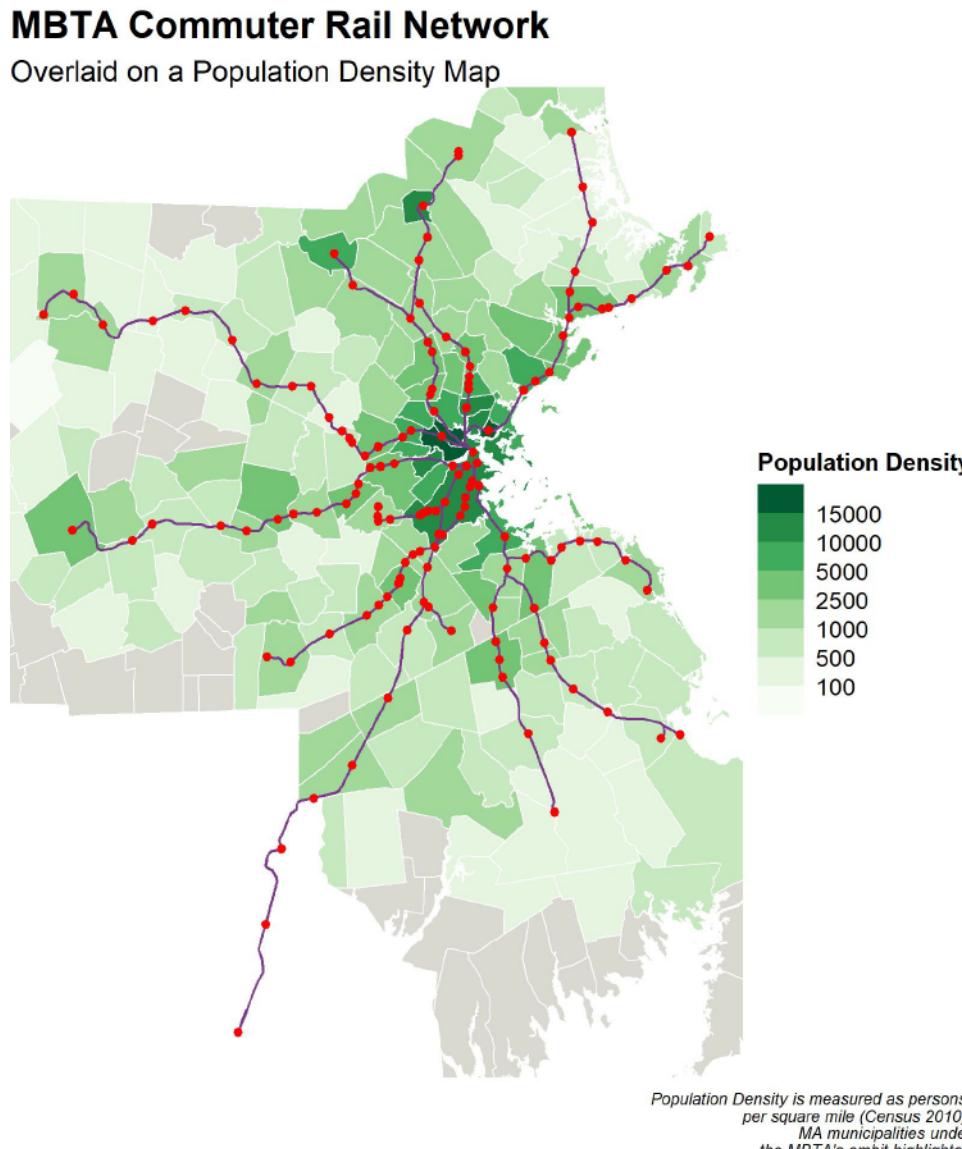


```
library(tidyverse)
library(sf)
library(ggspatial)

ggplot() +
  geom_sf(data = mbta_muni_shp,
          aes(geometry = geometry),
          color = 'white',
          fill = '#D9D9D2',
          size = 0.25) +
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),
          aes(geometry = geometry,
              fill = pop_dens_sqmile),
          color = 'white',
          size = 0.25) +
  geom_sf(data = mbta_cr_lines,
          aes(geometry = geometry),
          color = '#7B388C') +
  geom_sf(data = mbta_cr_stops,
          aes(geometry = geometry),
          size = 1,
          color = 'red') +
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),
                       name="Population Density",
                       palette = 'Greens',
                       direction = 1) +
  coord_sf(xlim = c(171357.0, 271642.8),
           ylim = c(814457.5, 950071.0),
           expand = TRUE) +
  theme_void() +
  labs(title = 'MBTA Commuter Rail Network',
       subtitle = "Overlaid on a Population Density Map",
       caption = 'Population Density is measured as persons per square mile (Census 2010)\nMA municipalities under the MBTA\'s ambit highlighted') +
  theme(plot.title = element_text(face = 'bold',
                                  size = 14),
        plot.subtitle = element_text(face = 'plain',
                                    size = 11),
        plot.caption = element_text(face = 'italic',
                                    size = 6),
        plot.caption.position = 'plot',
        legend.title = element_text(face = 'bold',
                                    size = 9)) +
  ggsave('mbta_commrail_network.jpeg',
         height = 7,
         width = 5,
         units = 'in',
         dpi = 'retina')
```

# Mapping in R

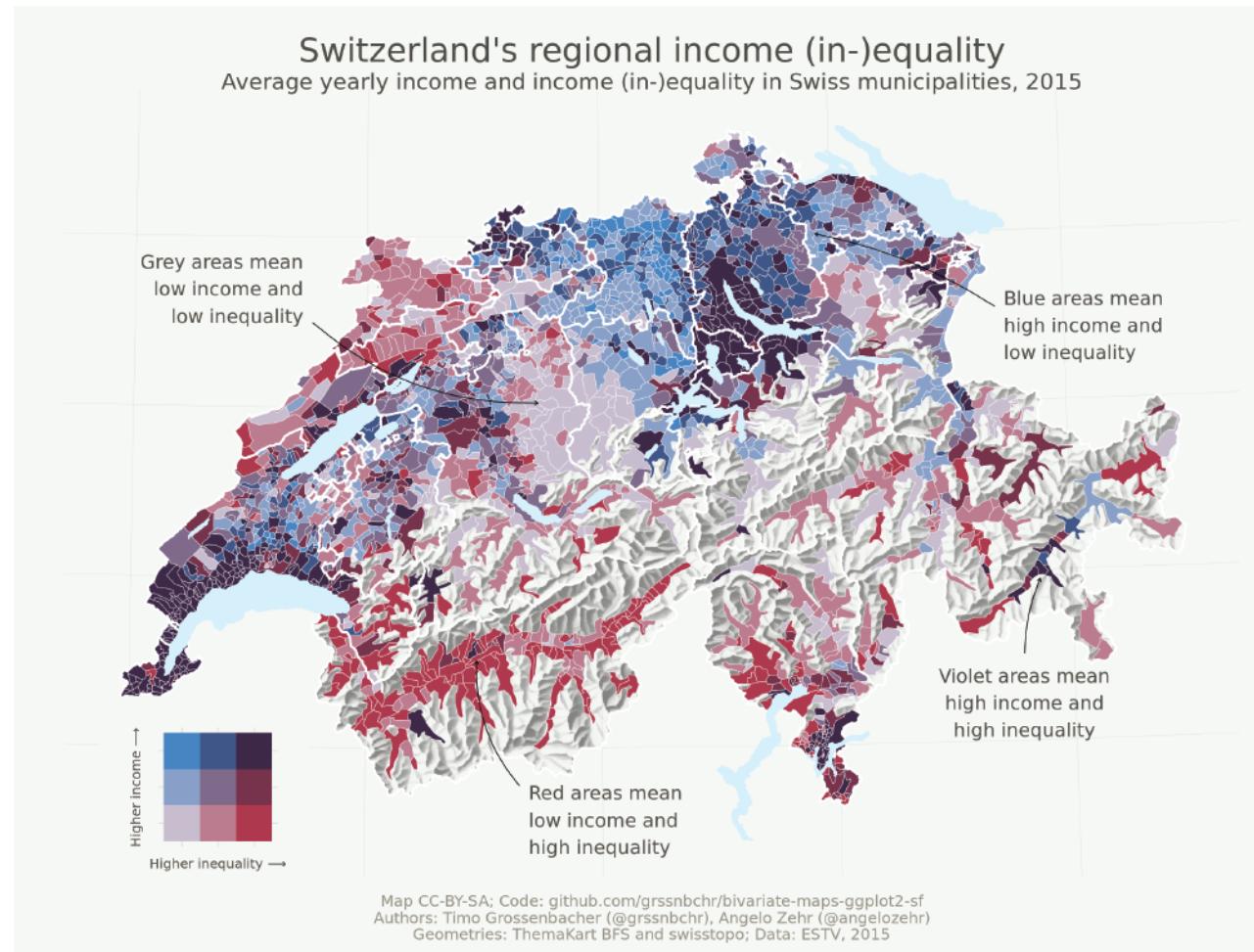
## Saving the map



```
library(tidyverse)
library(sf)
library(ggspatial)

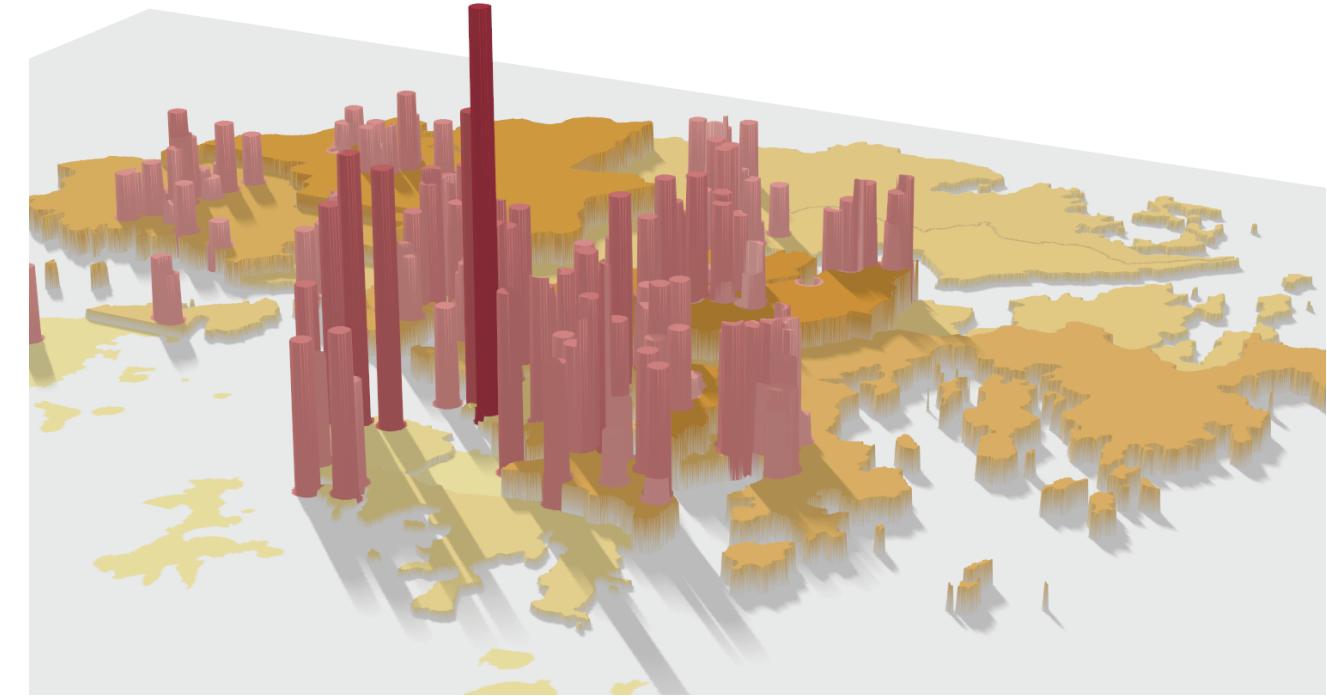
ggplot() +
  geom_sf(data = mbta_muni_shp,
          aes(geometry = geometry),
          color = 'white',
          fill = '#D9D9D2',
          size = 0.25) +
  geom_sf(data = mbta_muni_shp %>% filter(MBTA_ASSESS == 1),
          aes(geometry = geometry,
              fill = pop_dens_sqmile),
          color = 'white',
          size = 0.25) +
  geom_sf(data = mbta_cr_lines,
          aes(geometry = geometry),
          color = '#7B388C') +
  geom_sf(data = mbta_cr_stops,
          aes(geometry = geometry),
          size = 1,
          color = 'red') +
  scale_fill_fermenter(breaks = c(0, 100, 500, 1000, 2500, 5000, 10000, 15000, 20000),
                       name="Population Density",
                       palette = 'Greens',
                       direction = 1) +
  coord_sf(xlim = c(171357.0, 271642.8),
           ylim = c(814457.5, 950071.0),
           expand = TRUE) +
  theme_void() +
  labs(title = 'MBTA Commuter Rail Network',
       subtitle = "Overlaid on a Population Density Map",
       caption = 'Population Density is measured as persons per square mile (Census 2010)\nMA municipalities under the MBTA's ambit highlighted') +
  theme(plot.title = element_text(face = 'bold',
                                  size = 14),
        plot.subtitle = element_text(face = 'plain',
                                    size = 11),
        plot.caption = element_text(face = 'italic',
                                    size = 6),
        plot.caption.position = 'plot',
        legend.title = element_text(face = 'bold',
                                    size = 9)) +
  ggsave('mbta_commrail_network.jpeg',
         height = 7,
         width = 5,
         units = 'in',
         dpi = 'retina')
```

# {sf} & {ggplot2} Examples



<https://timogrossenbacher.ch/2019/04/bivariate-maps-with-ggplot2-and-sf/>

Tim Grossenbacher  
Apr 19, 2019



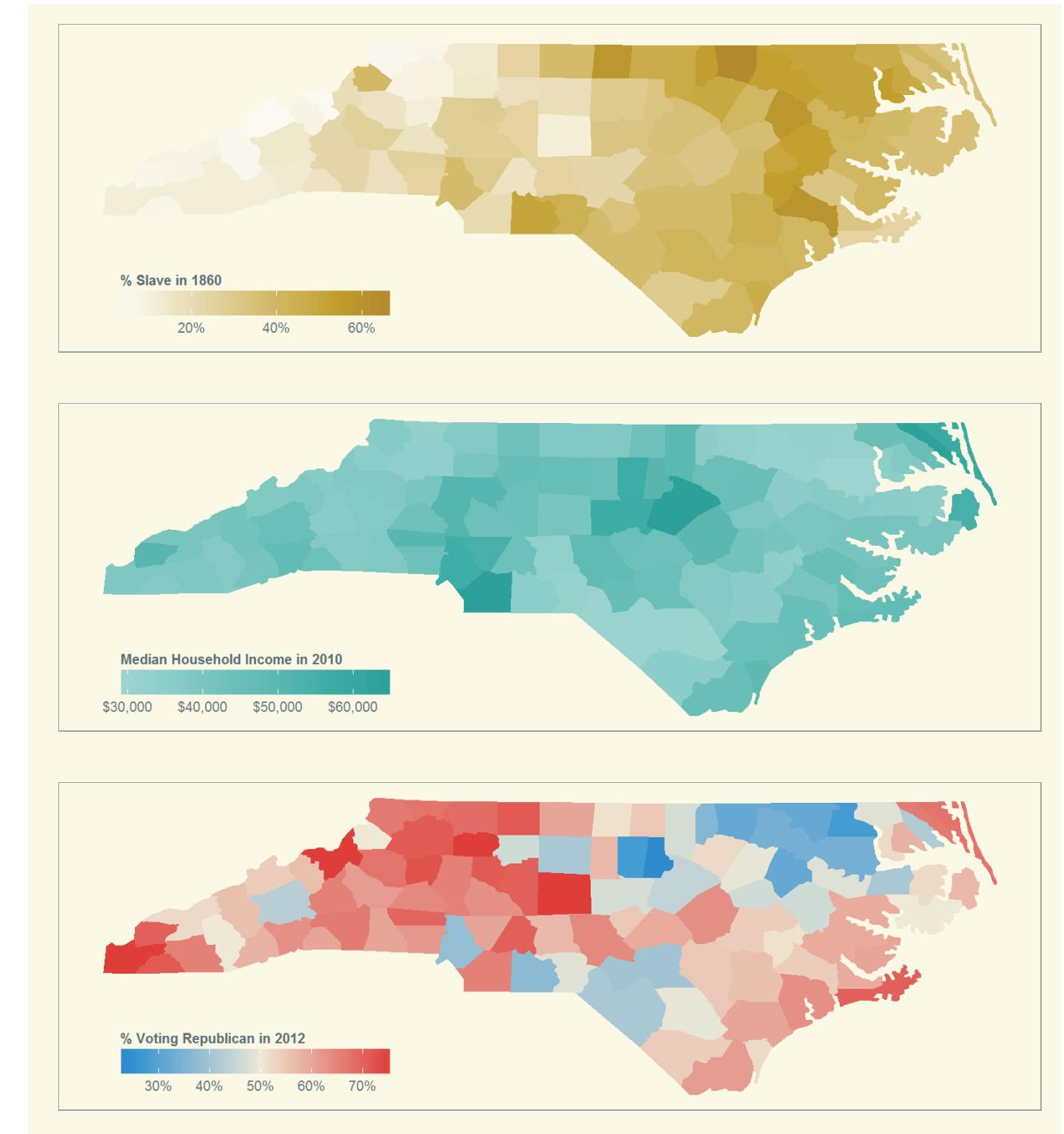
<https://towardsdatascience.com/introducing-3d-ggplots-with-rayshader-r-c61e27c6f0e9>

Carrie Lo  
May 23, 2020

# {sf} & {ggplot2} Examples



<https://jcheshire.com/r-spatial-data-hints/great-maps-ggplot2/>  
James Cheshire  
2020



<http://www.peterhaschke.com/r/2013/12/05/NCmaps.html>  
Peter Haschke  
Dec 05, 2013

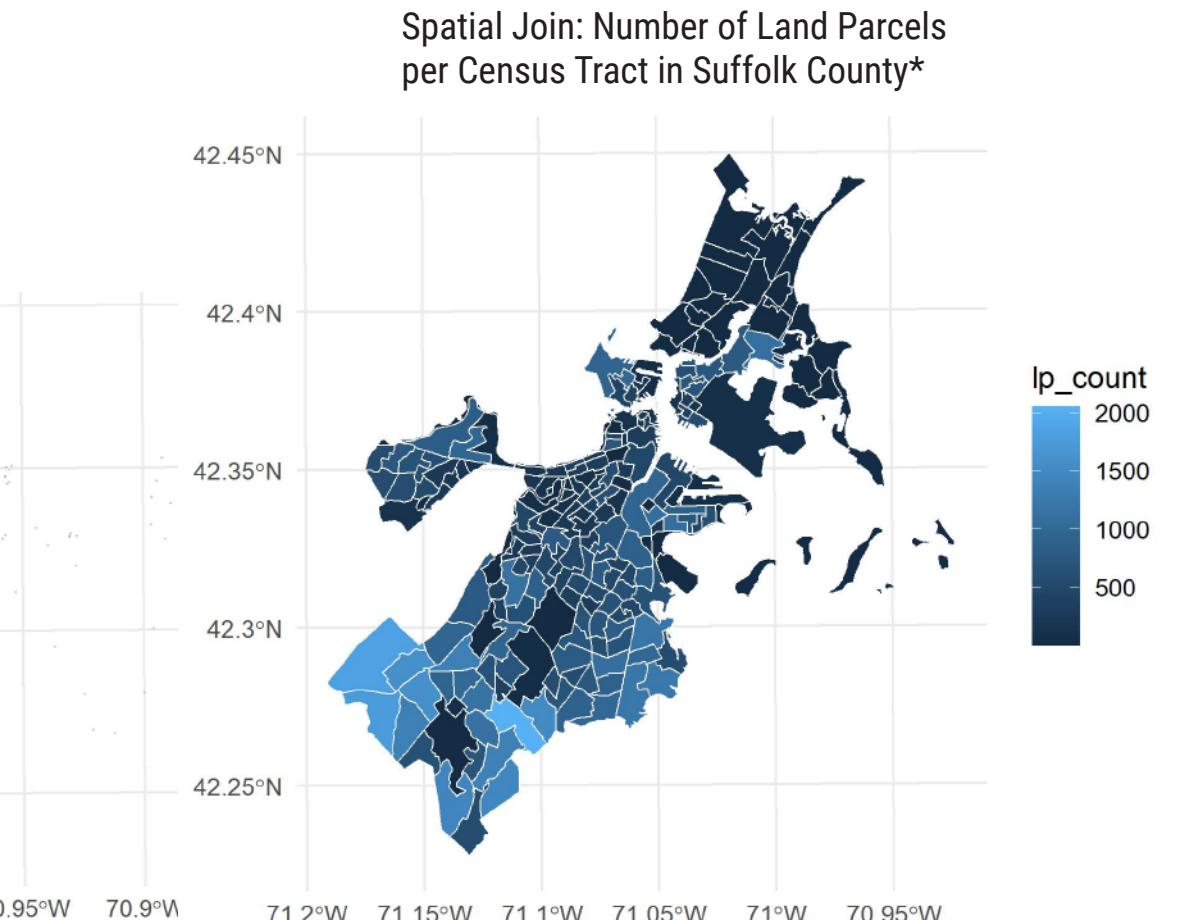
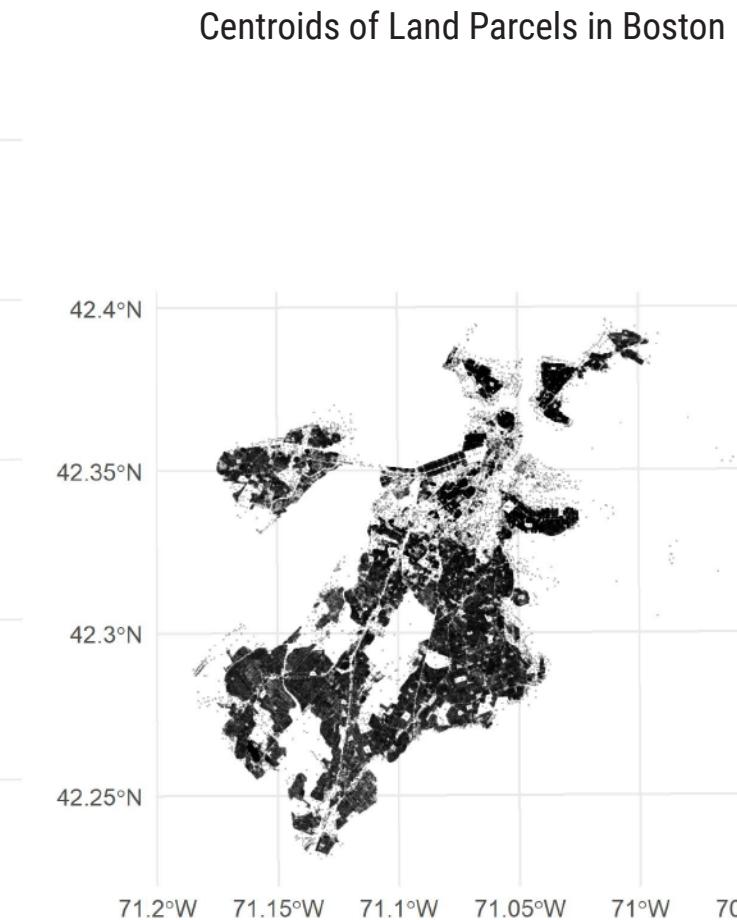
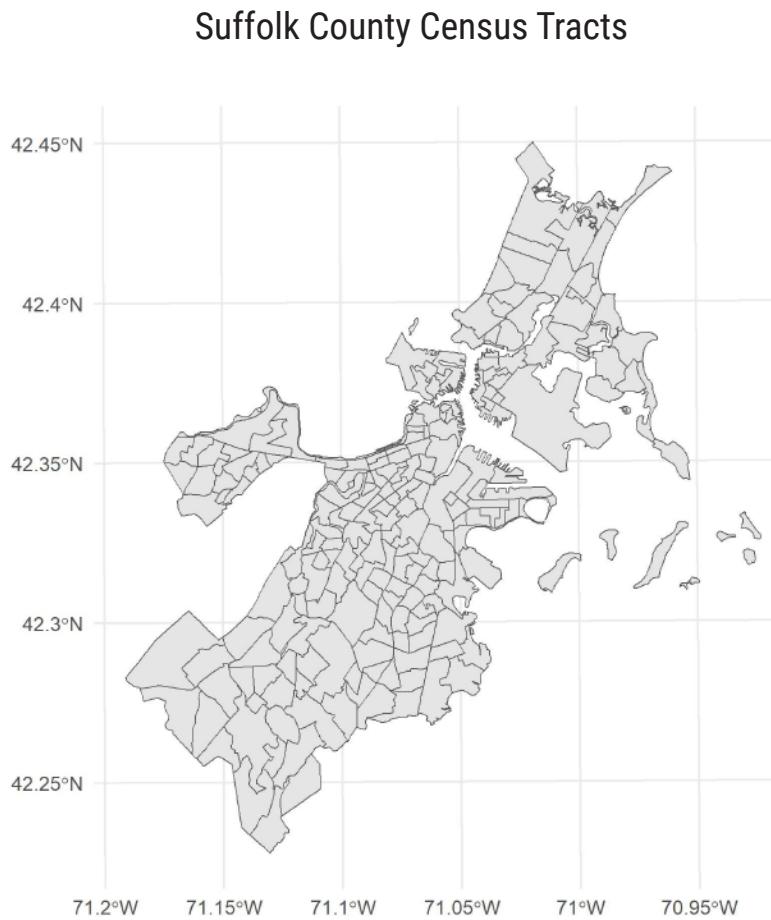
# Some {sf} Tools in R

## Spatial Join: joining two spatial files

### st\_join

Code Snippet:

```
lp_count_ct <- st_join(bos_censustracts,  
                      bos_landparcels_centroids) %>%  
  group_by(GEOID10, NAME10) %>%  
  summarise(lp_count = n())
```



# Some `{sf}` Tools in R

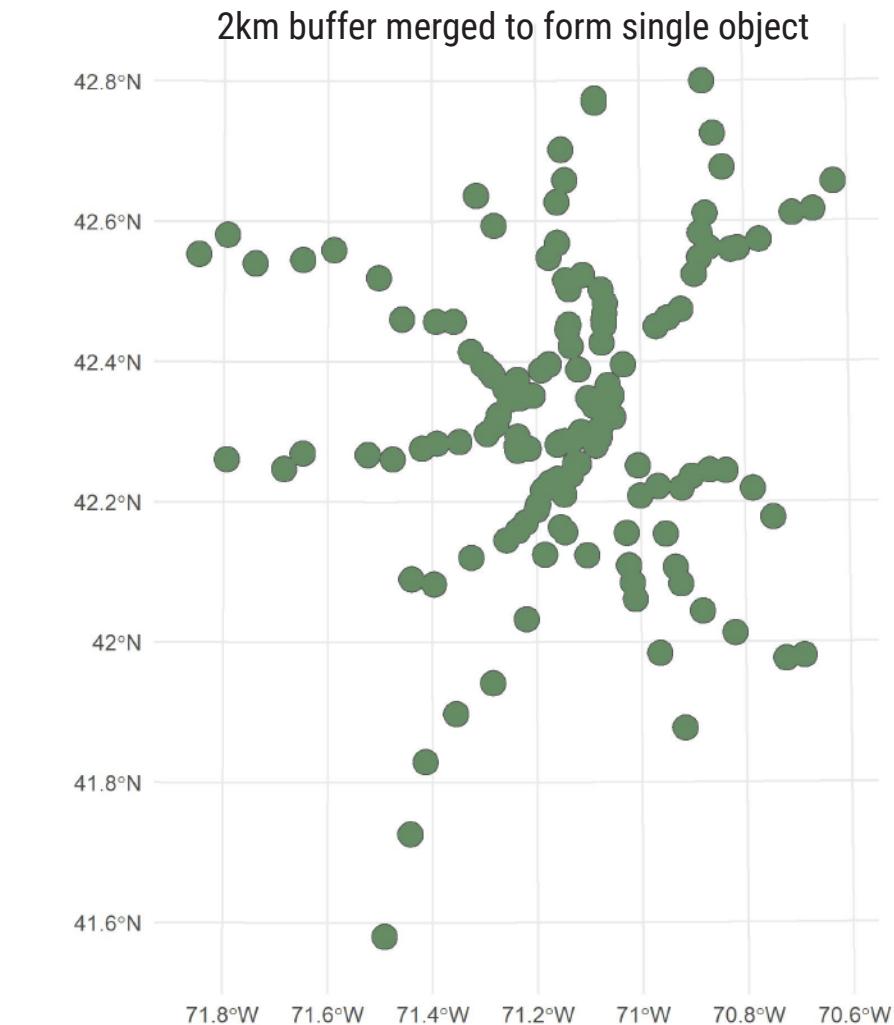
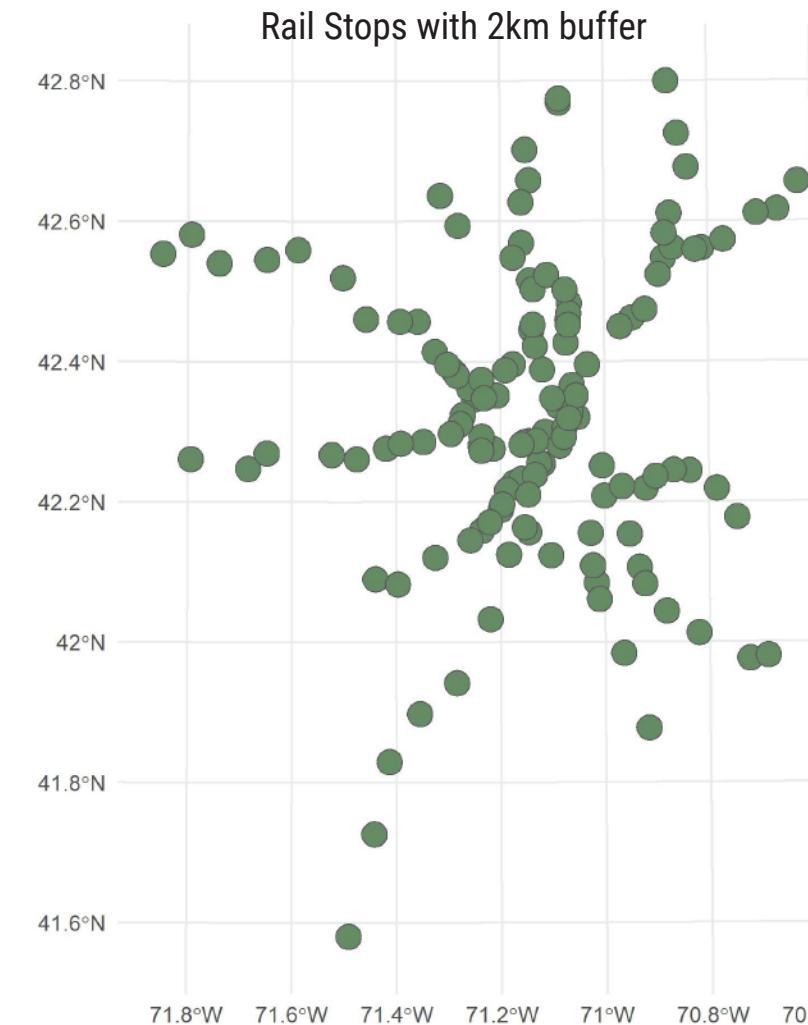
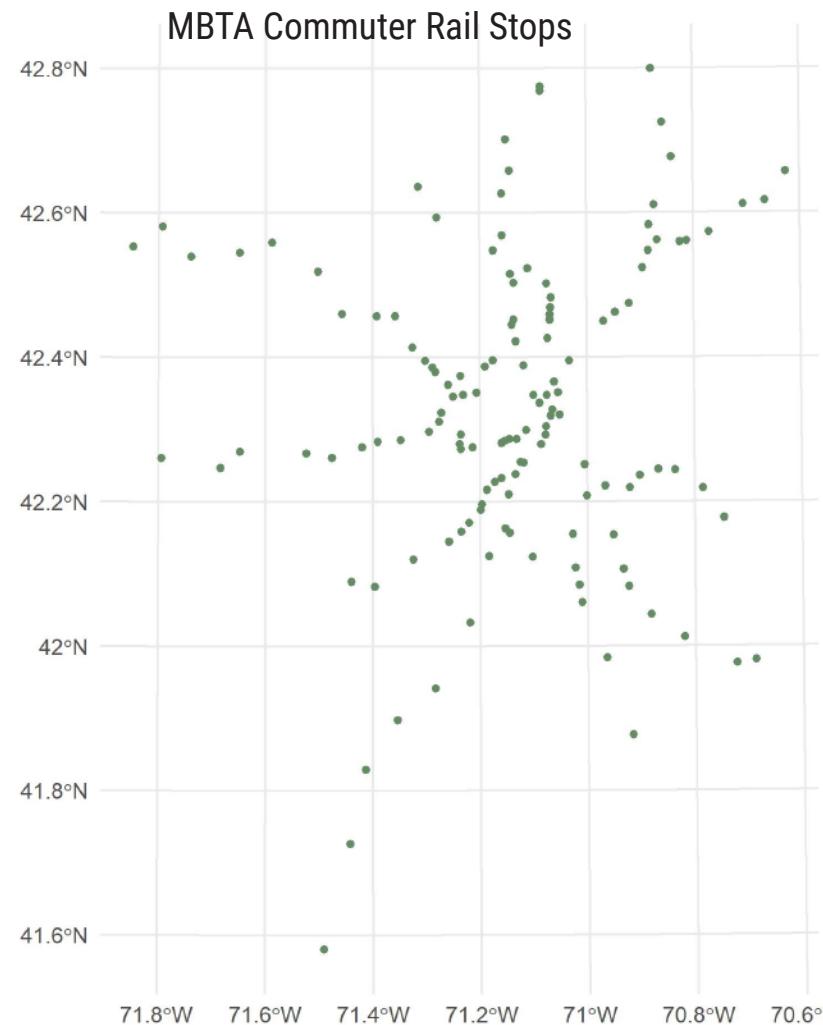
**Buffer:** generate polygon around object

`st_buffer`

Code Snippet:

```
# generate a 2km wide buffer around Commuter Rail stops
twokm_commrail <- st_buffer(mbta_cr_stops,
                             dist = 2000)

# Since sf objects are data frames, we can perform all regular tibble operations.
# Summarizing the sf object will 'melt' the geometry at the specified group.
# If no grouping is present, it will merge into a single geometry.
twokm_commrail %>%
  group_by() %>%
  summarise()
```



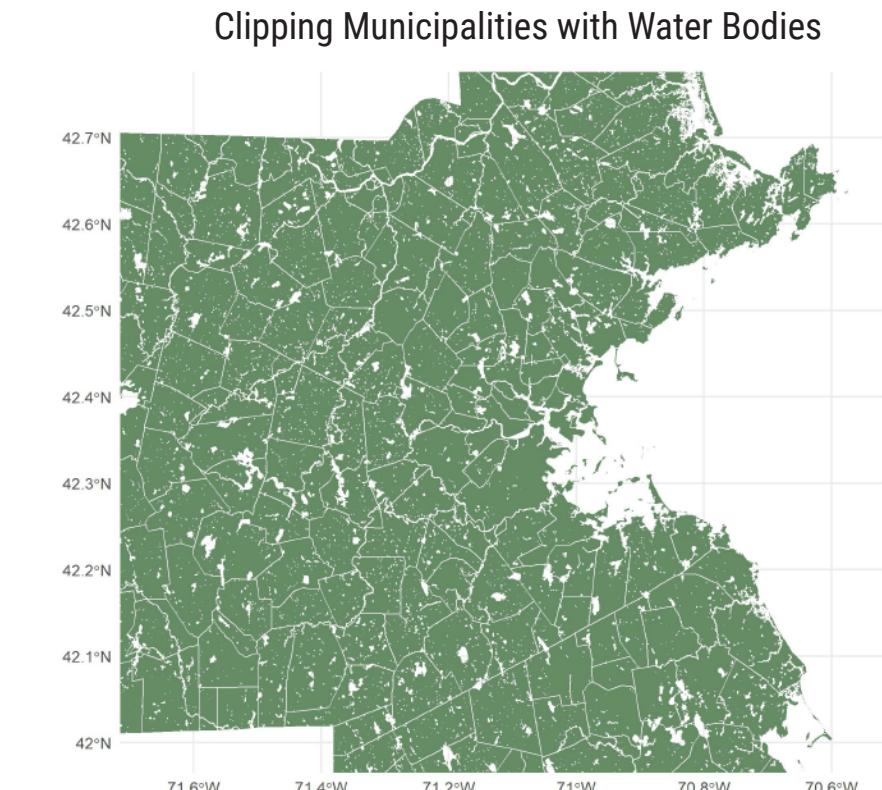
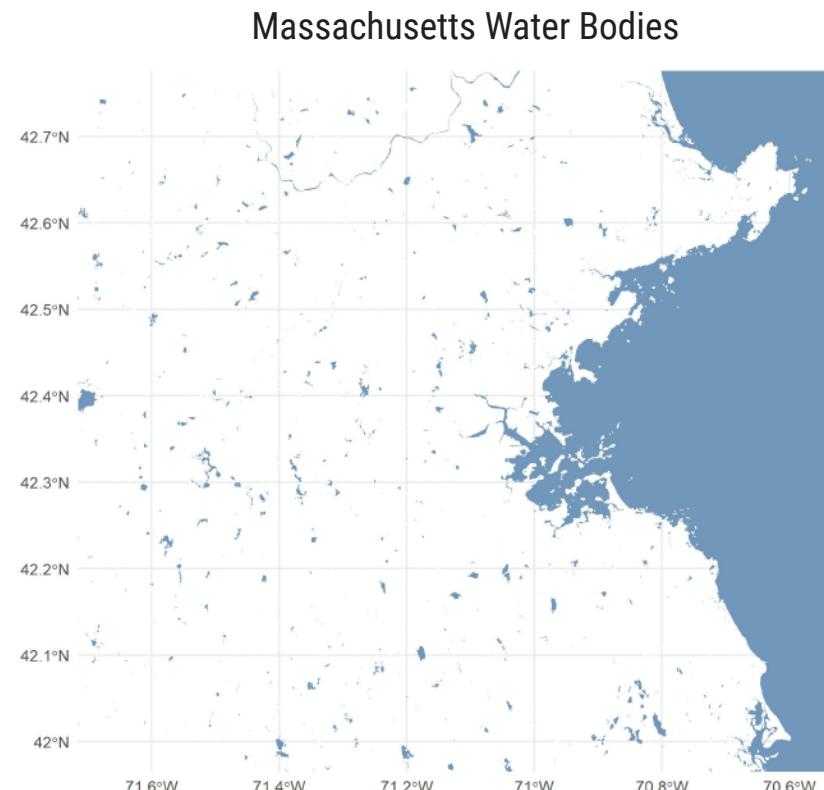
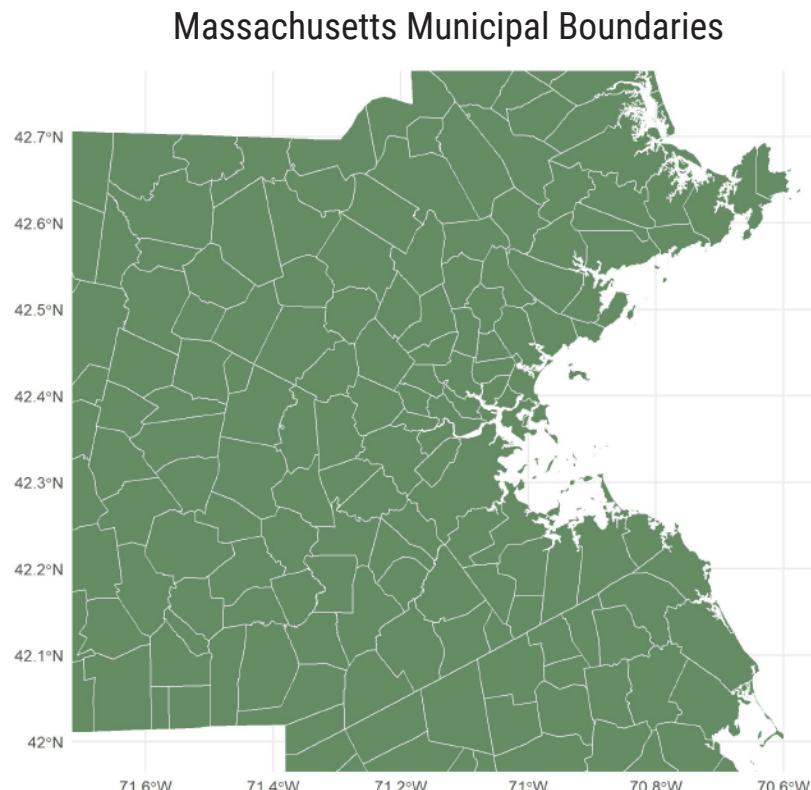
# Some {sf} Tools in R

## Clipping: Cut geometry with other geometry

### `st_difference`

Code Snippet:

```
# st_difference removes the portion of x overlapped by y
muni_hydro_clip <- st_difference(muni_shp, # municipalities shapefile
                                    ma_hydro) # hydrological bodies shapefile
```





## Aseem Deodhar

Research Analyst I  
[adeodhar@mapc.org](mailto:adeodhar@mapc.org)  
Pronouns: *he, him, his*



Metropolitan Area Planning Council  
60 Temple Place, Boston, MA 02111  
[www.mapc.org](http://www.mapc.org)