Advanced Data Visualizations

West - Week 3

AUTHOR
Jeff Goseland

PUBLISHED
November 22, 2025

Week 3 Homework

Part A

```r
#loads needed libraries
suppressMessages({
  library(tidyverse)
  library(RColorBrewer)
  library(sf)
  library(ggmap)
  library(leaflet)
  library(keyring)
  library(htmltools)
  library(Polychrome)
})
```

1. Load spacial data from a shapefile into R

```r
filename <- '../data/School_Boundaries.shp'

school_boundaries_shp <- st_read(filename, stringsAsFactors = FALSE)
```

```
Reading layer `School_Boundaries' from data source
  `/Users/jeffgoseland/Library/CloudStorage/GoogleDrive-jeff.goseland@gmail.com/My Drive/Notre
Dame /Advanced Data Visualizations/advanced_data_viz/data/School_Boundaries.shp'
  using driver `ESRI Shapefile'
```

```
Simple feature collection with 57 features and 3 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -86.3219 ymin: 41.61083 xmax: -86.1971 ymax: 41.71744
Geodetic CRS:  WGS 84
```

```
summary(school_boundaries_shp)
```

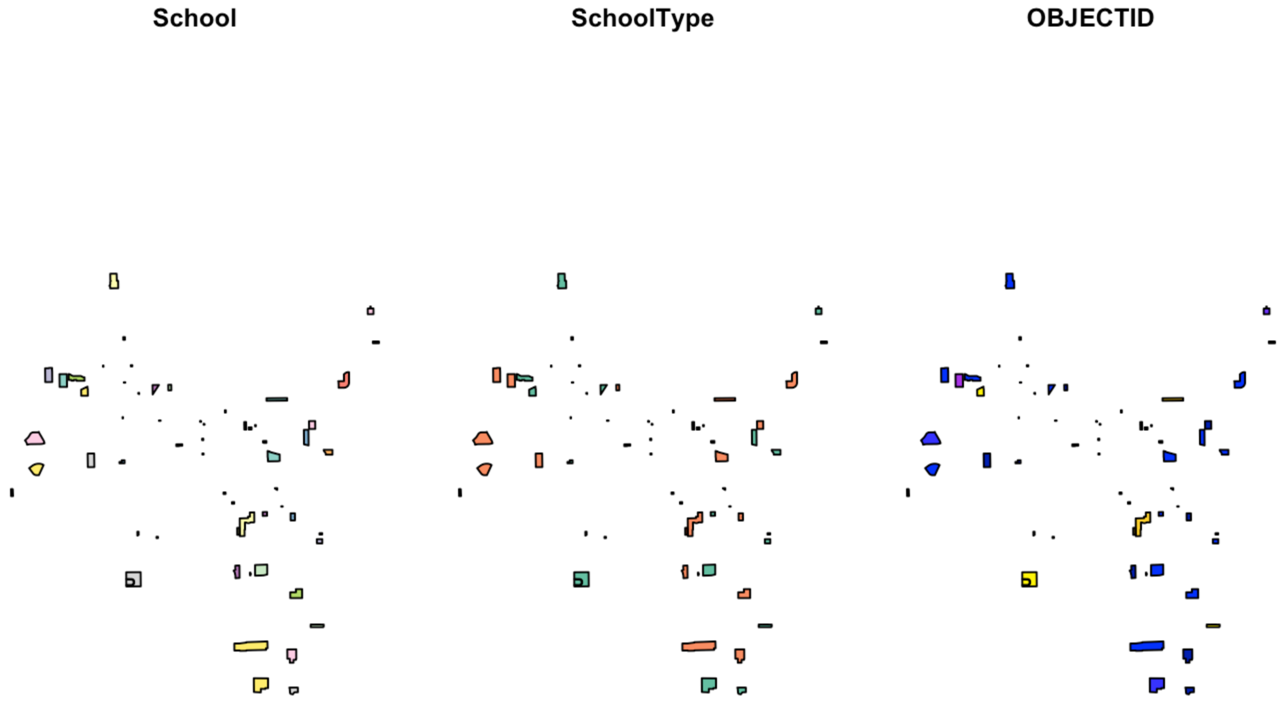
```
    School           SchoolType           OBJECTID              geometry
 Length:57          Length:57          Min.   :  6.0   MULTIPOLYGON :57
 Class :character   Class :character   1st Qu.: 67.0   epsg:4326    : 0
 Mode  :character   Mode  :character   Median :118.0   +proj=long...: 0
                                       Mean   :141.1
                                       3rd Qu.:165.0
                                       Max.   :548.0
```

```
school_boundaries_shp[1,0]
```

	geometry
	<s_MULTIP>
1	<s_MULTIP>

1 row

```
plot(school_boundaries_shp)
```

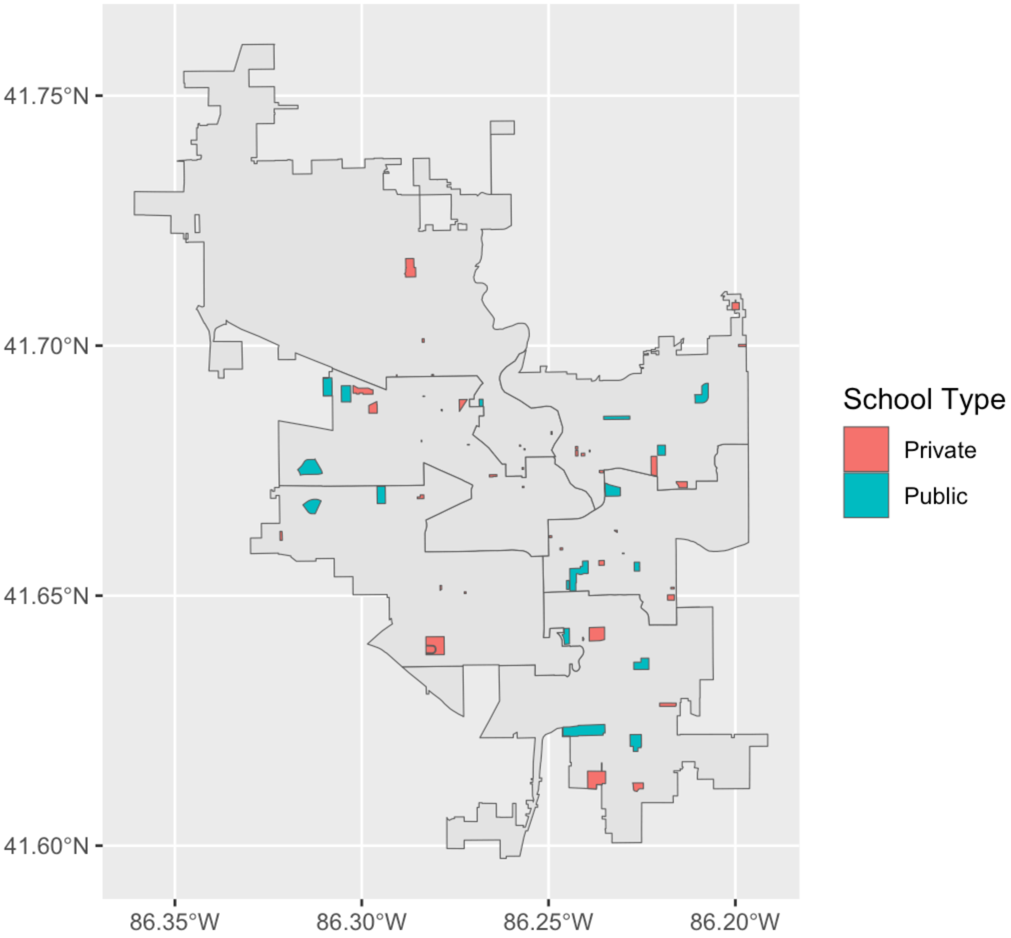


```
# I don't know if we were supposed to do this but the data looks way better
# with the district latyer underneath it for context
districts <- st_read("../data/City_Council_Districts.shp", stringsAsFactors = FALSE)
```

```
Reading layer `City_Council_Districts' from data source
  `/Users/jeffgoseland/Library/CloudStorage/GoogleDrive-jeff.goseland@gmail.com/My Drive/Notre
Dame /Advanced Data Visualizations/advanced_data_viz/data/City_Council_Districts.shp'
  using driver `ESRI Shapefile'
Simple feature collection with 6 features and 14 fields
Geometry type: MULTIPOLYGON
```

```
Dimension:     XY
Bounding box:  xmin: -86.36085 ymin: 41.59745 xmax: -86.19133 ymax: 41.76027
Geodetic CRS:  WGS 84
```

```
ggplot() +
   geom_sf(data = districts) +
  geom_sf(data = school_boundaries_shp$geometry,
          aes(fill = school_boundaries_shp$SchoolType)) +
  labs(fill = "School Type")
```



2. Create a spatial data frame from a table of latitude and longitudes

- Include the code you used to load the csv;
- Include the code you used to convert it to a spatial data frame;
- Include the output of the summary() command run on the spatial data frame; and
- Create a static map with the colors of the features reflecting some variable.

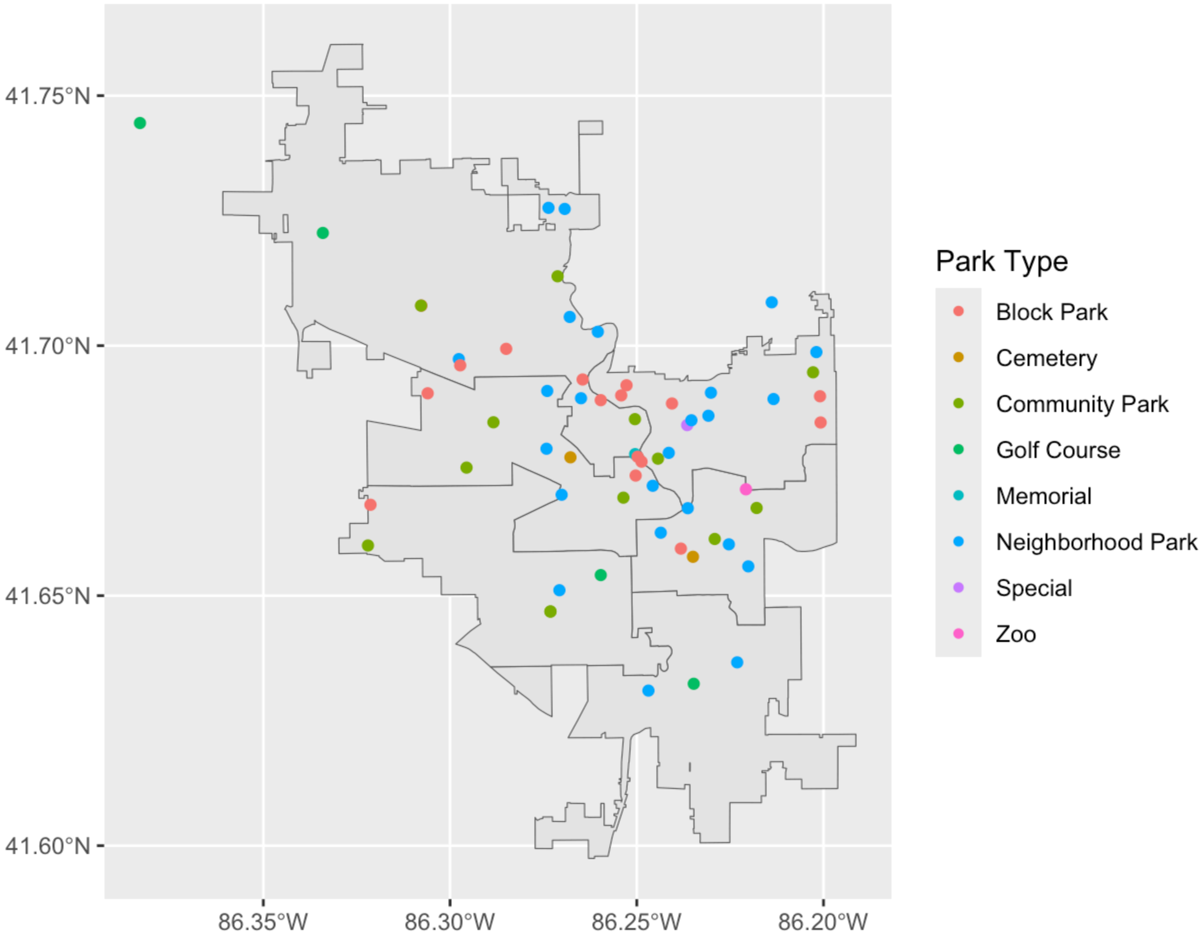
```r
# loading the data
features_points <- read.csv("../data/Parks_Locations_and_Features.csv")


# getting the lat/long
features_spacial <- features_points %>%
                    st_as_sf(coords = c("Lon","Lat")) %>%
                    st_set_crs(value = 4326)

summary(features_spacial)
```

```
 Park_Name          Park_Type            Zip_Code     Aqua_Feat__Pool
Length:62          Length:62          Min.   :46601   Min.   :1  
Class :character   Class :character   1st Qu.:46614   1st Qu.:1  
Mode  :character   Mode  :character   Median :46616   Median :1  
                                      Mean   :46617   Mean   :1  
                                      3rd Qu.:46619   3rd Qu.:1  
                                      Max.   :46635   Max.   :1  
                                                      NA's   :60  
Aqua_Feat__Spray Backstop__Practice   Ballfield       Basketball
Min.   :1        Min.   :1          Min.   :1.000   Min.   :0.50  
1st Qu.:1        1st Qu.:1          1st Qu.:1.000   1st Qu.:0.50  
Median :1        Median :1          Median :1.000   Median :1.00  
Mean   :1        Mean   :1          Mean   :1.769   Mean   :1.25  
3rd Qu.:1        3rd Qu.:1          3rd Qu.:2.000   3rd Qu.:2.00  
Max.   :1        Max.   :1          Max.   :7.000   Max.   :4.00  
NA's   :57       NA's   :59         NA's   :49      NA's   :46  
   Blueway   Complex__Ballfield Complex__Tennis   Concessions    Disk_Golf
Min.   :1    Min.   :1          Min.   :1       Min.   :1     Min.   :1  
1st Qu.:1    1st Qu.:1          1st Qu.:1       1st Qu.:1     1st Qu.:1  
```

```
Median :1    Median :1          Median :1       Median :1    Median :1  
Mean   :1    Mean   :1          Mean   :1       Mean   :1    Mean   :1  
3rd Qu.:1    3rd Qu.:1          3rd Qu.:1       3rd Qu.:1    3rd Qu.:1  
Max.   :1    Max.   :1          Max.   :1       Max.   :1    Max.   :1  
NA's   :61   NA's   :61         NA's   :61      NA's   :60   NA's   :61  
Driving_Range Educational_Experience  Event_Space Fitness_Course
Min.   :1     Min.   :1            Min.   :1    Min.   :1     
1st Qu.:1     1st Qu.:1            1st Qu.:1    1st Qu.:1     
Median :1     Median :1            Median :1    Median :1     
Mean   :1     Mean   :1            Mean   :1    Mean   :1     
3rd Qu.:1     3rd Qu.:1            3rd Qu.:1    3rd Qu.:1     
Max.   :1     Max.   :1            Max.   :1    Max.   :1     
NA's   :61    NA's   :54           NA's   :57   NA's   :61    
Garden__Community Garden__Display      Golf        Hockey__Ice   Loop_Walk   
Min.   :1         Min.   :1.000   Min.   :0.500   Min.   :1    Min.   :1.00  
1st Qu.:1         1st Qu.:1.500   1st Qu.:0.875   1st Qu.:1    1st Qu.:1.00  
Median :1         Median :2.000   Median :1.000   Median :1    Median :1.00  
Mean   :1         Mean   :1.667   Mean   :0.875   Mean   :1    Mean   :1.25  
3rd Qu.:1         3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:1    3rd Qu.:1.00  
Max.   :1         Max.   :2.000   Max.   :1.000   Max.   :1    Max.   :3.00  
NA's   :59        NA's   :59      NA's   :58      NA's   :61   NA's   :54    
MP_Field__Large MP_Field__Multiple MP_Field__Small Multiuse_Court
Min.   :1.000   Min.   :5          Min.   :1       Min.   :1     
1st Qu.:1.000   1st Qu.:5          1st Qu.:1       1st Qu.:1     
Median :1.000   Median :5          Median :1       Median :1     
Mean   :1.455   Mean   :5          Mean   :1       Mean   :1     
3rd Qu.:1.500   3rd Qu.:5          3rd Qu.:1       3rd Qu.:1     
Max.   :4.000   Max.   :5          Max.   :1       Max.   :1     
NA's   :51      NA's   :61         NA's   :57      NA's   :61    
 Natural_Area   Open_Turf    Open_Water    Other___Active Other_Passive
Min.   :1     Min.   :1    Min.   :1.000   Min.   :1      Min.   :1    
1st Qu.:1     1st Qu.:1    1st Qu.:1.000   1st Qu.:1      1st Qu.:1    
Median :1     Median :1    Median :1.000   Median :1      Median :1    
Mean   :1     Mean   :1    Mean   :1.118   Mean   :1      Mean   :1    
3rd Qu.:1     3rd Qu.:1    3rd Qu.:1.000   3rd Qu.:1      3rd Qu.:1    
Max.   :1     Max.   :1    Max.   :2.000   Max.   :1      Max.   :1    
NA's   :52    NA's   :45   NA's   :45      NA's   :61     NA's   :61    
 Passive_Node   Picnic_Grounds   Playground__Destination Playground__Local
```

```
Min.   :1.000   Min.   : 1.000   Min.   :1.000           Min.   :1.000
1st Qu.:1.000   1st Qu.: 1.000   1st Qu.:1.000           1st Qu.:1.000
Median :1.000   Median : 1.000   Median :1.000           Median :1.000
Mean   :1.636   Mean   : 2.222   Mean   :1.167           Mean   :1.138
3rd Qu.:1.500   3rd Qu.: 1.000   3rd Qu.:1.000           3rd Qu.:1.000
Max.   :4.000   Max.   :12.000   Max.   :2.000           Max.   :2.000
NA's   :51      NA's   :53       NA's   :56              NA's   :33
  Public_Art    Shelter     Shelter__Group    Skate_Park Sledding_Hill
Min.   :1    Min.   :1.00   Min.   :1.000   Min.   :1    Min.   :1
1st Qu.:1    1st Qu.:1.00   1st Qu.:1.000   1st Qu.:1    1st Qu.:1
Median :1    Median :1.00   Median :1.000   Median :1    Median :1
Mean   :1    Mean   :1.25   Mean   :1.538   Mean   :1    Mean   :1
3rd Qu.:1    3rd Qu.:1.25   3rd Qu.:2.000   3rd Qu.:1    3rd Qu.:1
Max.   :1    Max.   :2.00   Max.   :4.000   Max.   :1    Max.   :1
NA's   :60   NA's   :58     NA's   :49      NA's   :61   NA's   :60
  Structure       Tennis       Trail__Primitive Volleyball
Min.   :1.0   Min.   : 2.000   Min.   :1        Mode:logical
1st Qu.:1.0   1st Qu.: 2.000   1st Qu.:1        NA's:62
Median :1.0   Median : 2.000   Median :1
Mean   :1.2   Mean   : 3.615   Mean   :1
3rd Qu.:1.0   3rd Qu.: 4.000   3rd Qu.:1
Max.   :2.0   Max.   :15.000   Max.   :1
NA's   :57    NA's   :49       NA's   :61
Water_Access__Developed Water_Access__General Water_Feature   Address
Min.   :1.00            Min.   :1             Min.   :1     Length:62
1st Qu.:1.00            1st Qu.:1             1st Qu.:1     Class :character
Median :1.00            Median :1             Median :1     Mode  :character
Mean   :1.25            Mean   :1             Mean   :1
3rd Qu.:1.25            3rd Qu.:1             3rd Qu.:1
Max.   :2.00            Max.   :1             Max.   :1
NA's   :54              NA's   :55            NA's   :61
         geometry
POINT        :62
epsg:4326    : 0
+proj=long...: 0
```

```r
ggplot() +
  geom_sf(data = districts) +
  geom_sf(data = features_spacial,
          aes(col = Park_Type)) +
  guides(col = guide_legend(override.aes = list(shape = 16))) +
  labs(col = "Park Type")
```



Part B

1. Use the ggmaps package to geocode your subset. Use the output = “more” option.

- Alternatively, you may use the tidygeocoder package.

```
#registering gkey
# key_set_with_value("google_api", password = "har-har-har")
register_google(key = key_get("google_api"))

# loading the data
business_licenses <- read.csv("../data/Goseland_Business_Licenses_9.csv")
```

2. Save the final data frame as a .csv (with the field names lat, lon, and loctype appended onto the original data). Submit this .csv.

```
# get geocode using 'more' option
print("I'm starting.....")
```

```
[1] "I'm starting....."
```

```
geo_data <- business_licenses %>%
            mutate_geocode(Full_Address, output = "more")
print("I'm finished.....")
```

```
[1] "I'm finished....."
```

```
write_csv(geo_data, "./goseland_geocoded_output.csv")
```

3. Submit a PDF or HTML file with the code you wrote to do the geocoding and a short justification of the steps you took.

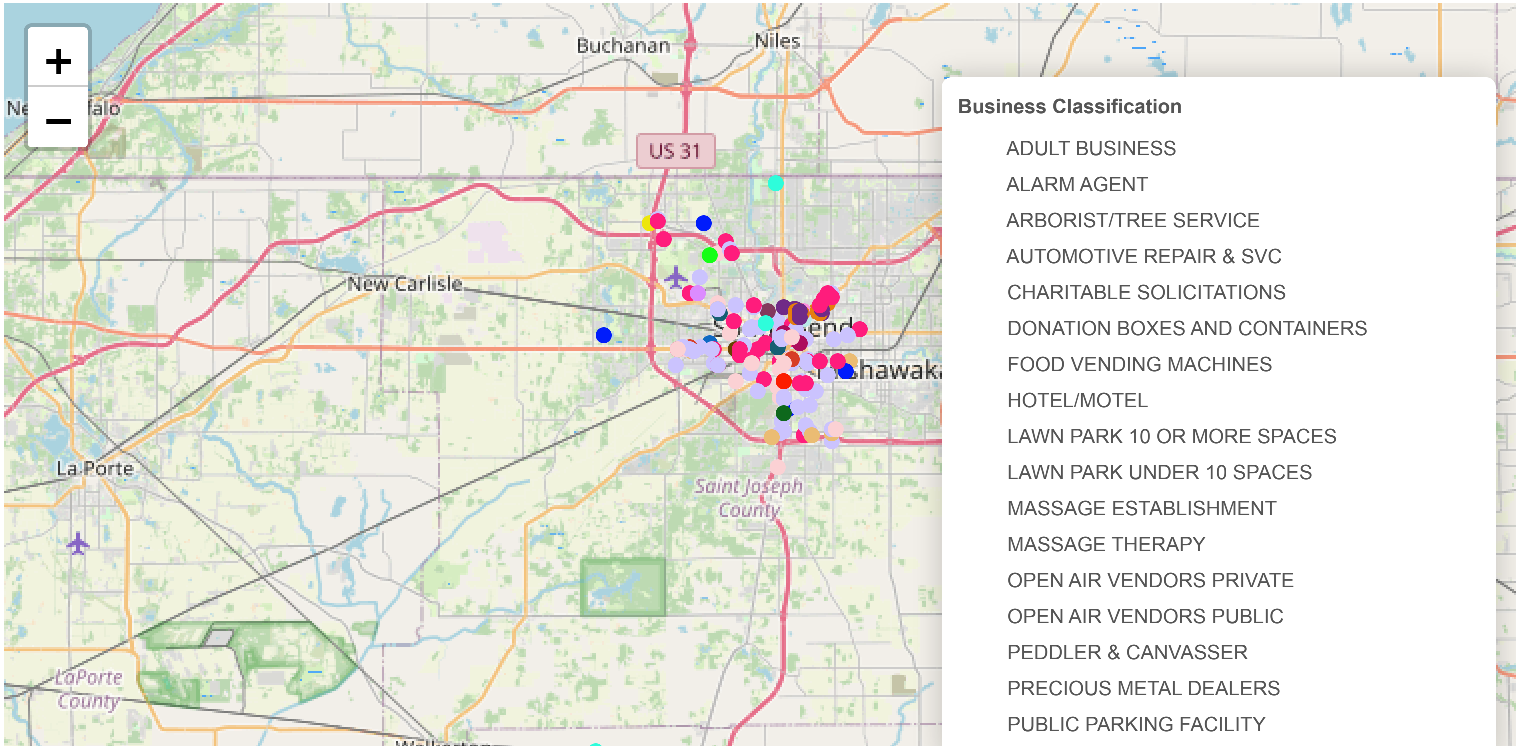
Justification:

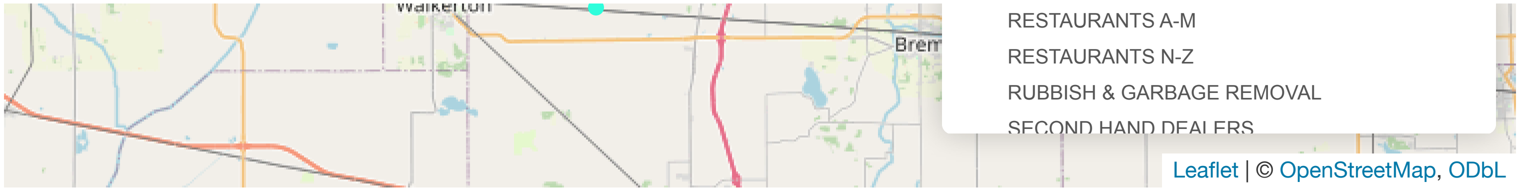
- I registered the key securely with keyring so I can do it once and refer to it again later
- I use ggplot to geocode since its simple and the full address was already concatonated for me in the Excel sheet (thank you)
- I stored the geo_data in a separate value since the geocoding process preserves the original columns
- simple save to csv

Part C

Using the examples in the earlier parts of this assignment and in the courseware, use R to make a Leaflet map showing at least one of the datasets from Part A.

```
# with so many categories I need to a way to make distinct colors
# and this worked well -- used google here to find a way
distinct_colors <- createPalette(length(unique(geo_data$Classification_Description)),
                                 c("#ff0000", "#00ff00", "#0000ff"))

# this part I had to do some google and claude help
# I tried to do it inline and had lots of failures
color_palette <- colorFactor(palette = distinct_colors,
                             domain = geo_data$Classification_Description)

leaflet(geo_data) %>%
    addTiles(group = 'Street Map') %>%
    addCircleMarkers(lng = ~lon, lat = ~lat,
                     color = ~color_palette(Classification_Description), # here
                     fillOpacity = 1,
                     radius = 4,
                     stroke = FALSE,
                     weight = 1,
                     popup = ~paste0(
                       "<b>", Business_Name, "</b><br/>",
                       Classification_Description, "<br/>",
                       Street_Address),
                     label = ~Business_Name) %>%
  addLegend( position = "bottomright",
             pal = color_palette, #here
             values = ~Classification_Description,
             title = "Business Classification",
             opacity = 1) %>%
   # the legend was too big so I used Claude to help me find a way to
   # make it smaller
   htmlwidgets::onRender("
      function(el, x) {
        var legend = el.querySelector('.legend');
        if(legend) {
          legend.style.fontSize = '10px';
          legend.style.maxHeight = '400px';
          legend.style.overflowY = 'auto';
        }
      }
    ")
```





1. Write a paragraph explaining what you are showing, why you made particular design choices, and how it could be useful for the audience/user.

> This is a map of the South Bend area with the businesses and types of businesses plotted by classification. There were so many classifications that I found a way to make very distinct colors as all of the other palettes were creating gradients and hues of the same colors making distinction very hard.
>
> Also because of the number of classifications, I found a way to make the legend smaller. I think both of the choices on the classifications makes it easier for the map consumer to use it.

2. Include the Leaflet code used to generate the map in a PDF or HTML file.