sprint_1

2023-04-15

Section #01: Benchmarking Data Exploration

Data source for this section

- 2016 LL33 Data Disclosure for CY2015 reporting, Government of New York City. https://www.nyc.gov/site/buildings/codes/benchmarking.page

The purpose of this section is to understand the dataset that we are going to use as the input of building energy consumption intensity - 2015 building energy consumption benchmarking data collected under NYC Local Law 84/133 Energy Benchmarking, which requires owners and managers of buildings larger than 50,000 square (25,000 after 2016) to report their building's energy usage to the City of New York on a yearly basis.

The column representing the energy consumption intensity is Site_EUI, and there are many factors impacting this value, like Water, ELectricity, building age

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.2.1     v dplyr   1.1.1
## v tidyr   1.3.0     v stringr 1.5.0
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

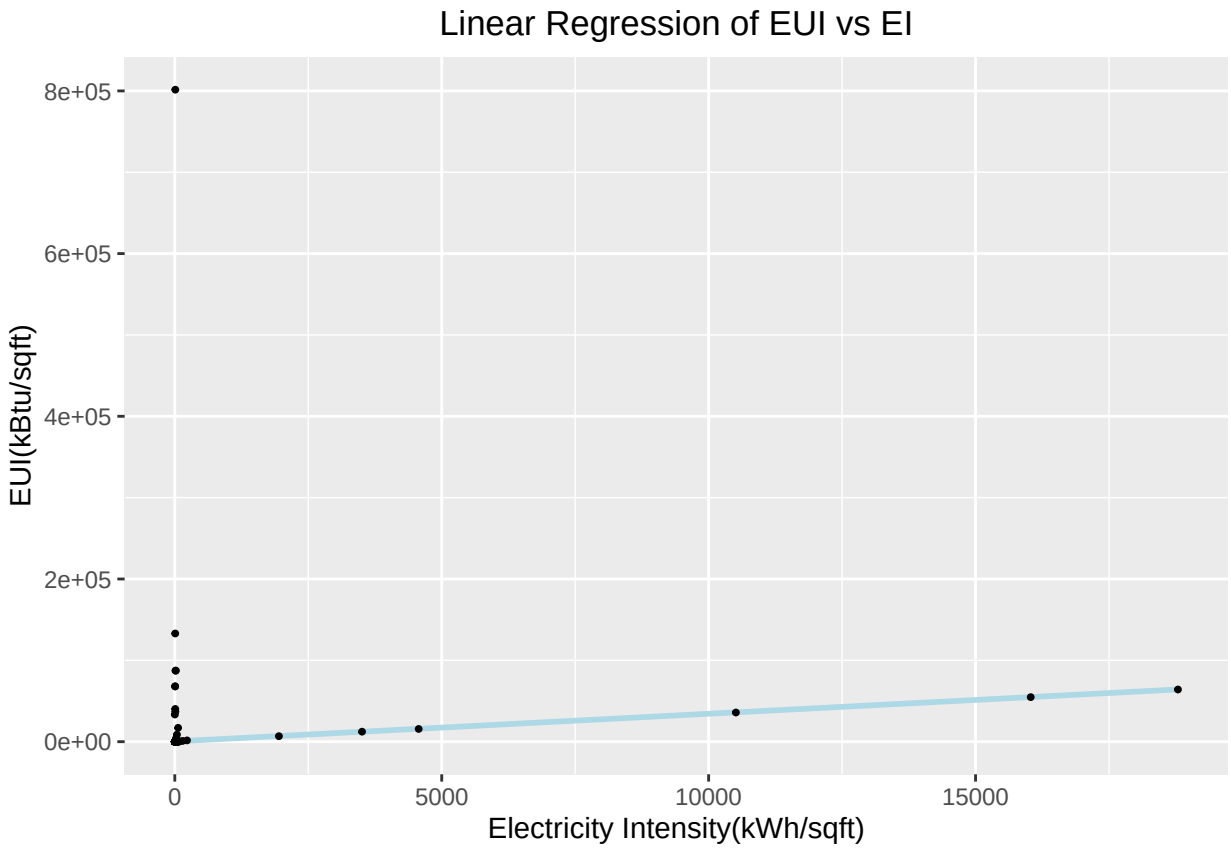
```
library(ggplot2)

# Load raw data
file_path <- file.path(dirname(rstudioapi::getSourceEditorContext()$path), "dataset/NYCBuildingEnergyUs
raw_data <- read_csv(file_path)
```

```
## Rows: 13223 Columns: 57
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (30): Coreported_BBL_Status, BBLs_Coreported, Reported_BINs, Property_Na...
## dbl (27): Record_Number, Order, BBL, Street_Number, Zip_Code, Largest_Proper...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Pre-processing
my_data <- raw_data %>%
  filter(DOF_Benchmarking_Submission_Status == "In Compliance") %>% # filter valid Benchmarking Submiss
  select(Record_Number,
         Site_EUI_kBtu_per_sqft,
         Weather_Normalized_Site_Electricity_Intensity_kWh_per_sqft,
         Weather_Normalized_Site_Natural_Gas_Intensity_therms_per_sqft,
```

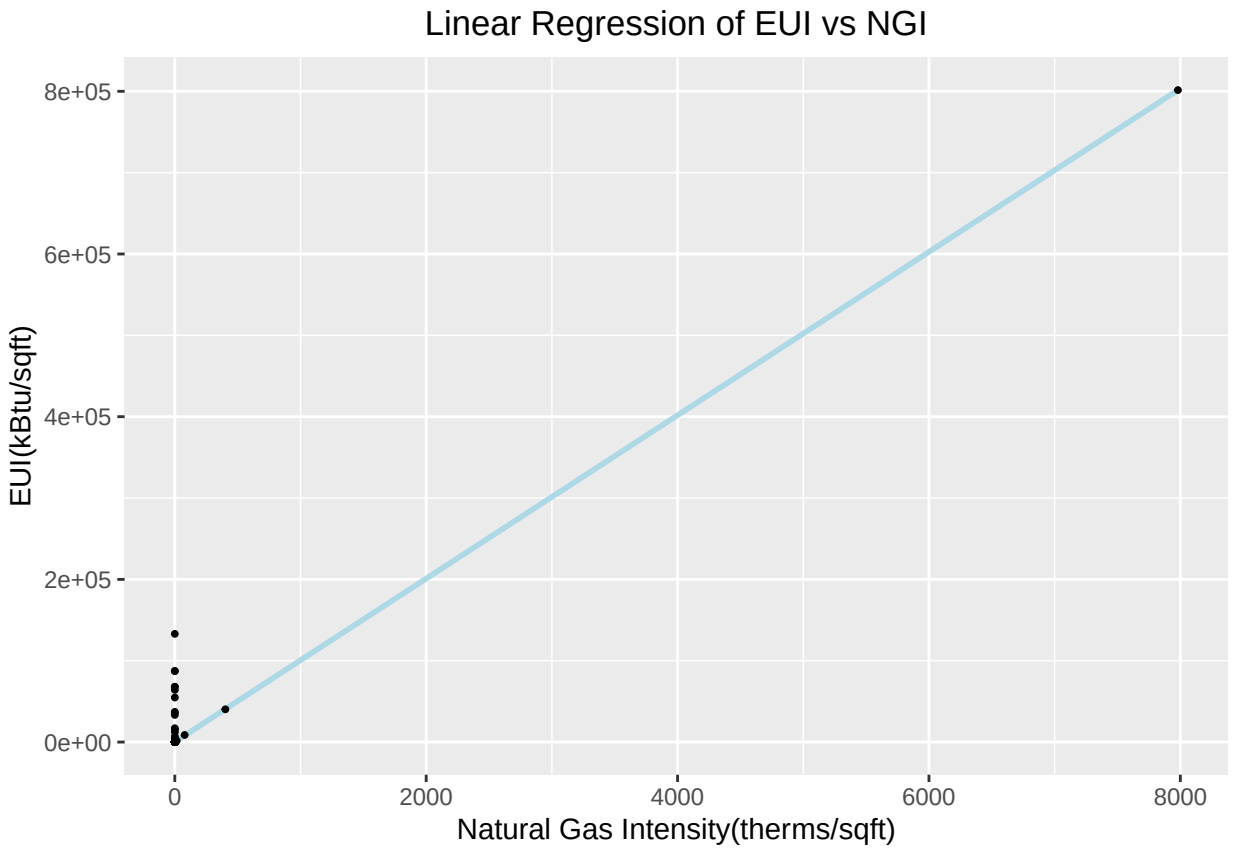
```
        Total_GHG_Emissions_Metric_Tons_CO2e,
        Municipally_Supplied_Potable_Water_Indoor_Intensity_gal_per_sqft
        ) %>%
  na.omit() %>% # quit properties with missing values
  rename(Record = Record_Number, # rename as the column names are too long
        EUI = Site_EUI_kBtu_per_sqft,
        EI = Weather_Normalized_Site_Electricity_Intensity_kWh_per_sqft,
        NGI = Weather_Normalized_Site_Natural_Gas_Intensity_therms_per_sqft,
        GHG = Total_GHG_Emissions_Metric_Tons_CO2e,
        WI = Municipally_Supplied_Potable_Water_Indoor_Intensity_gal_per_sqft)

# Fit in linear model with different X variables
# Create a scatterplot with regression line
ggplot(data = my_data, aes(x = EI, y = EUI)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightblue") +
  geom_point(size = 0.7) +
  labs(x = "Electricity Intensity(kWh/sqft)", y = "EUI(kBtu/sqft)", title = "Linear Regression of EUI vs
  theme(plot.title = element_text(hjust = 0.5)) # make title in the middle
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



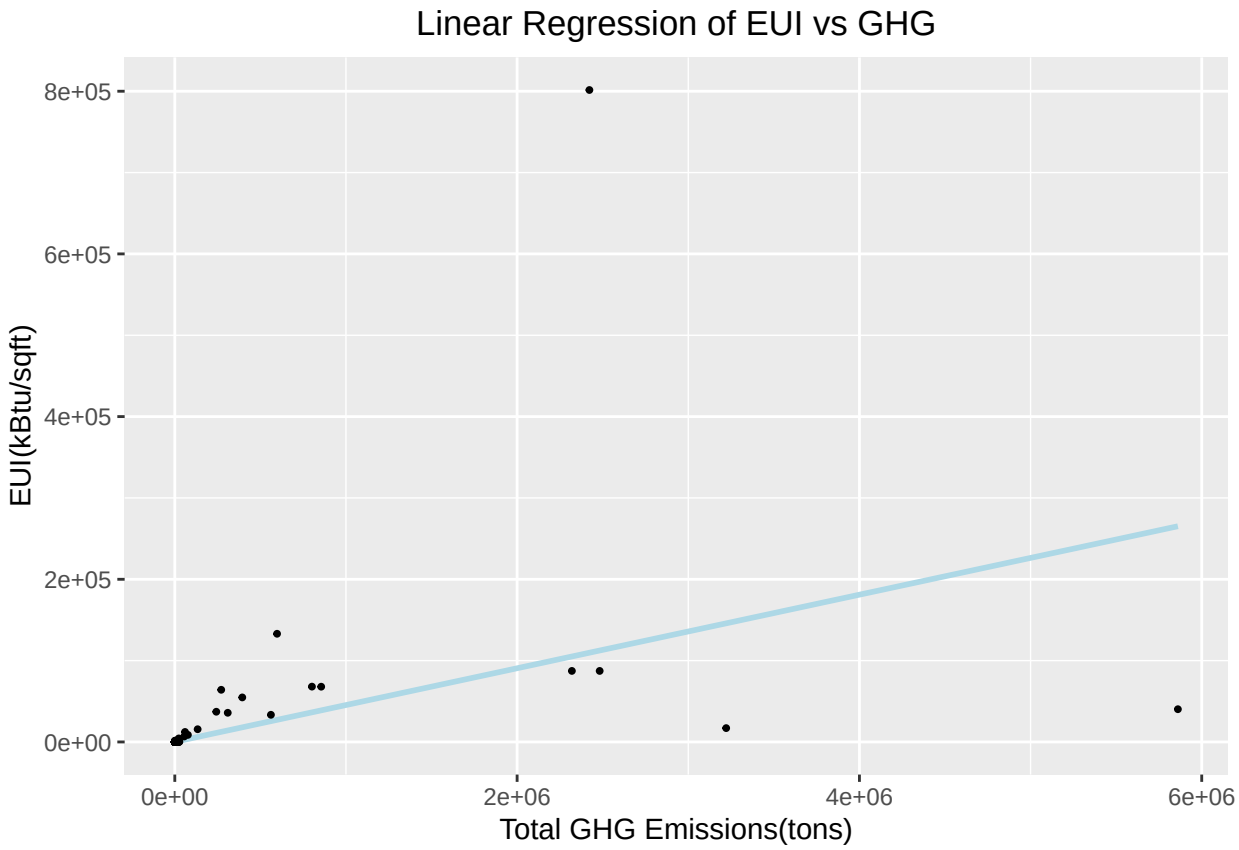

```
ggplot(data = my_data, aes(x = NGI, y = EUI)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightblue") +
  geom_point(size = 0.7) +
  labs(x = "Natural Gas Intensity(therms/sqft)", y = "EUI(kBtu/sqft)", title = "Linear Regression of EU
  theme(plot.title = element_text(hjust = 0.5))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



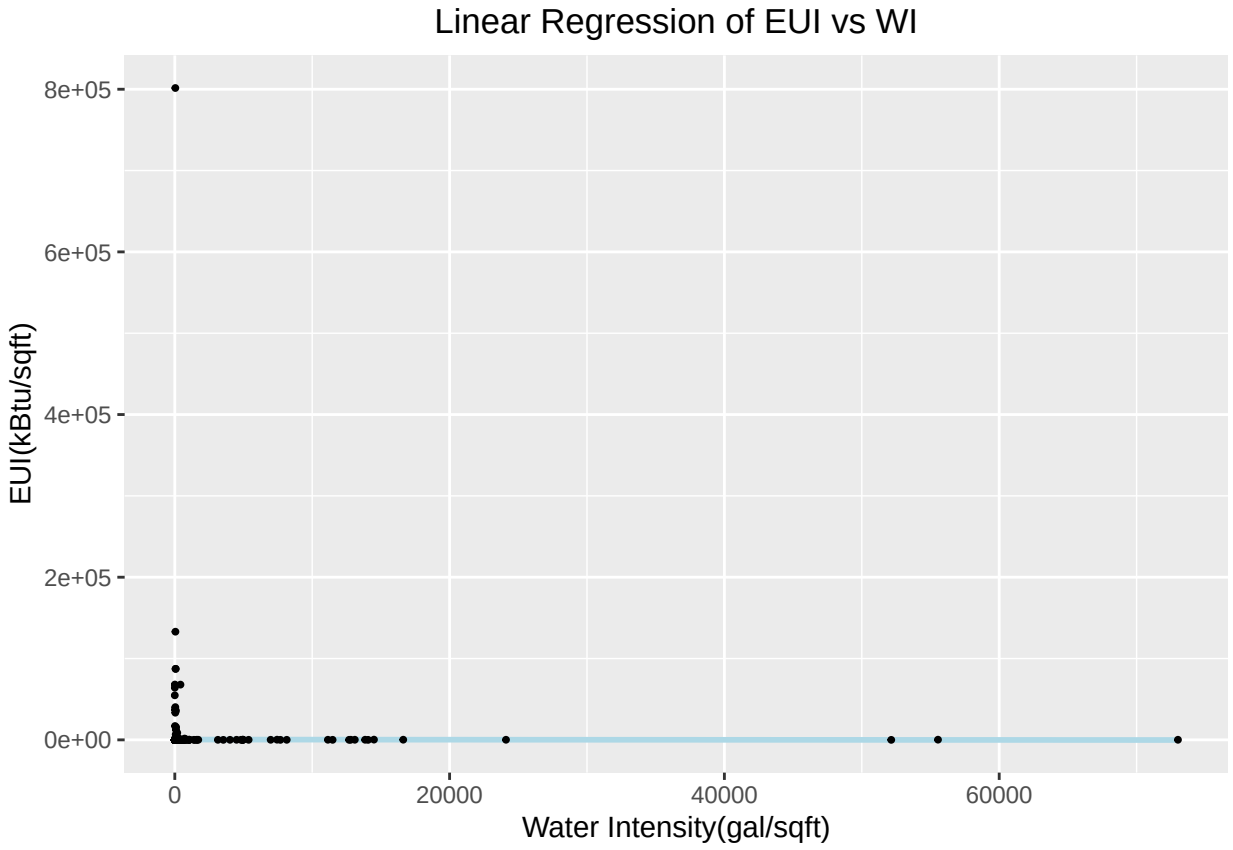

```
ggplot(data = my_data, aes(x = GHG, y = EUI)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightblue") +
  geom_point(size = 0.7) +
  labs(x = "Total GHG Emissions(tons)", y = "EUI(kBtu/sqft)", title = "Linear Regression of EUI vs GHG")
  theme(plot.title = element_text(hjust = 0.5))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggplot(data = my_data, aes(x = WI, y = EUI)) +
  geom_smooth(method = "lm", se = FALSE, color = "lightblue") +
  geom_point(size = 0.7) +
  labs(x = "Water Intensity(gal/sqft)", y = "EUI(kBtu/sqft)", title = "Linear Regression of EUI vs WI") +
  theme(plot.title = element_text(hjust = 0.5))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Section #02: Spatial Join

The purpose of the spatial join operation is to join the data from American Community Survey(ACS) collected by Census Bureau to the 2015 building energy consumption benchmarking data collected under NYC Local Law 84/133 Energy Benchmarking.

Data source for spatial join

- 2016 LL33 Data Disclosure for CY2015 reporting, Government of New York City. https://www.nyc.gov/site/buildings/codes/benchmarking.page
- 2010 Shapefiles of NYC census blocks, United States Census Bureau. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
- 2023 Shapefiles of NYC Tax Lot (BBL), NYC Open Data. https://www.nyc.gov/site/planning/data-maps/open-data.page#pluto

Step #01: shapefile importing and visualization

Import the .shp files to and visualize the polygons in Arcgis Pro

Step #02: spatial join using Tax Lot data and Census block data

Spatial join with options: - Target features: Tax Lot data - Join features: Census block data - Join operation: One to one - Match option: Within - Fields to join: GeoId (the only required feature for joining ACS data)

This generate a new Tax Lot data table with a new column of the Census Block it belongs to. This means we have BBL number and Census Block geoID in each row.

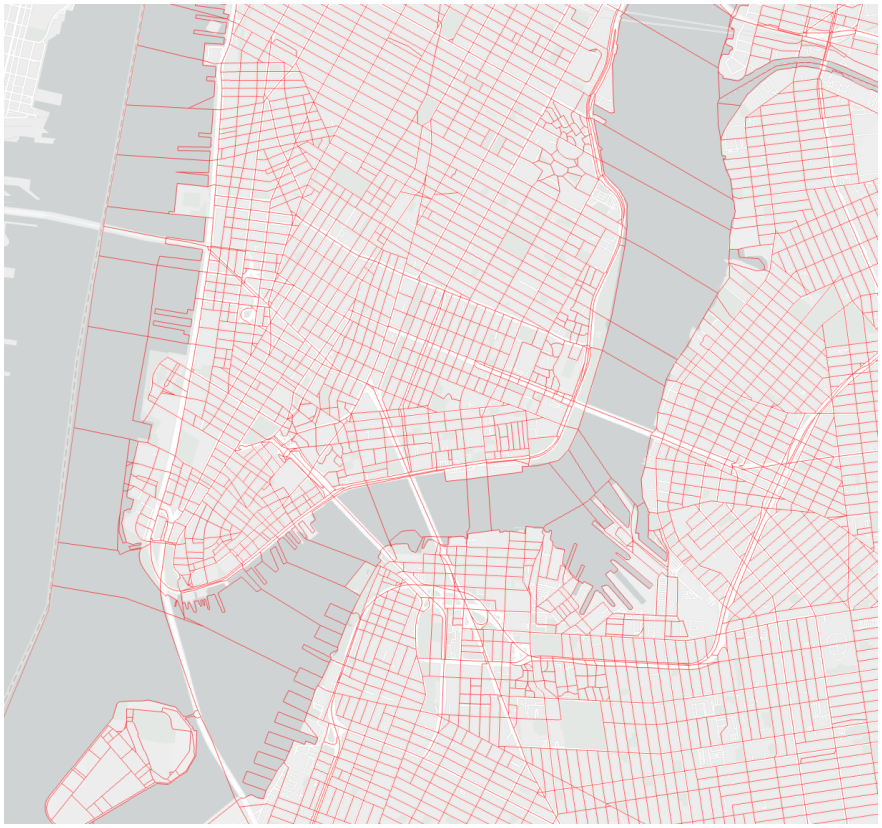


Figure 1: census block boundaries

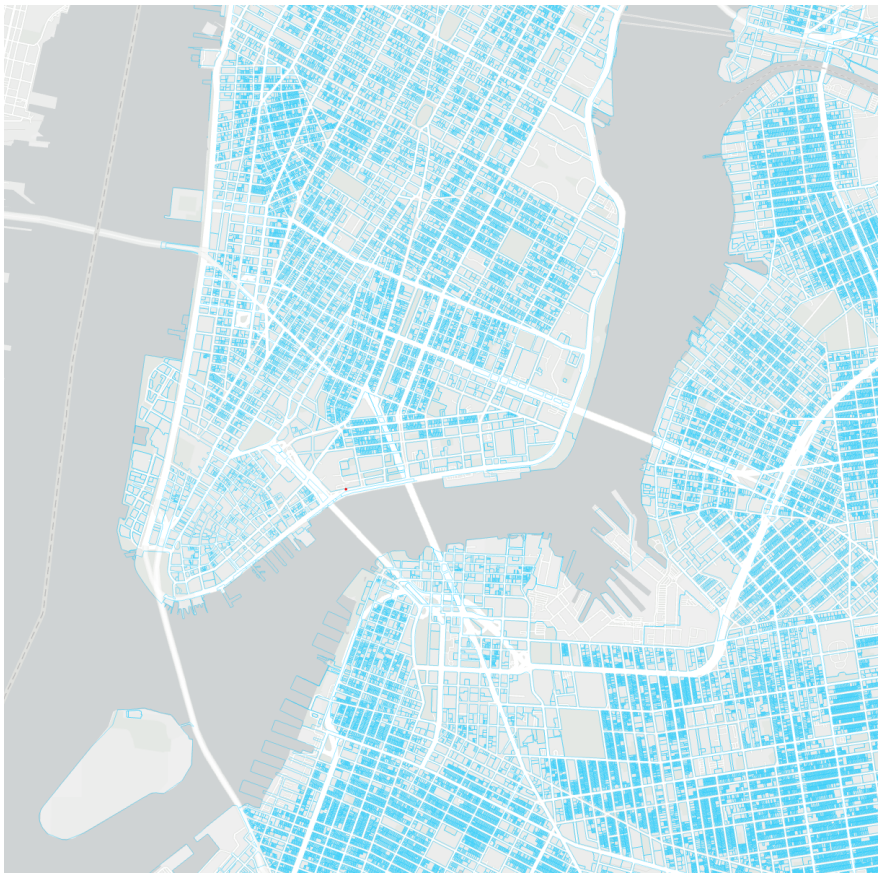


Figure 2: tax lot boundaries

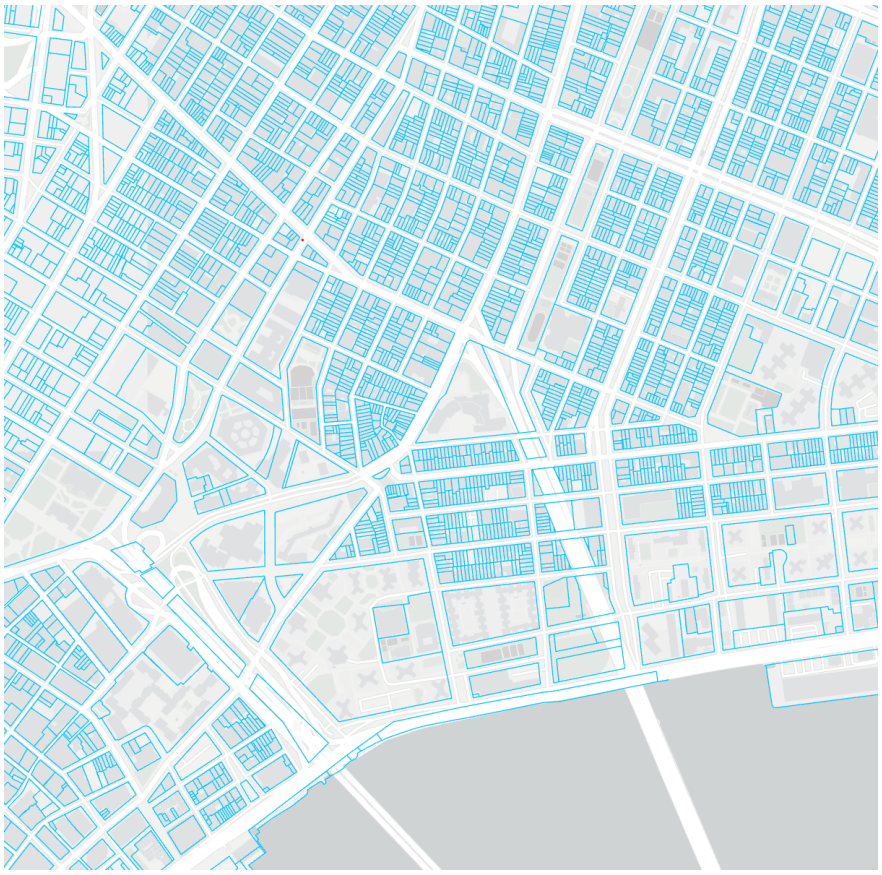


Figure 3: tax lot boundaries zoomed

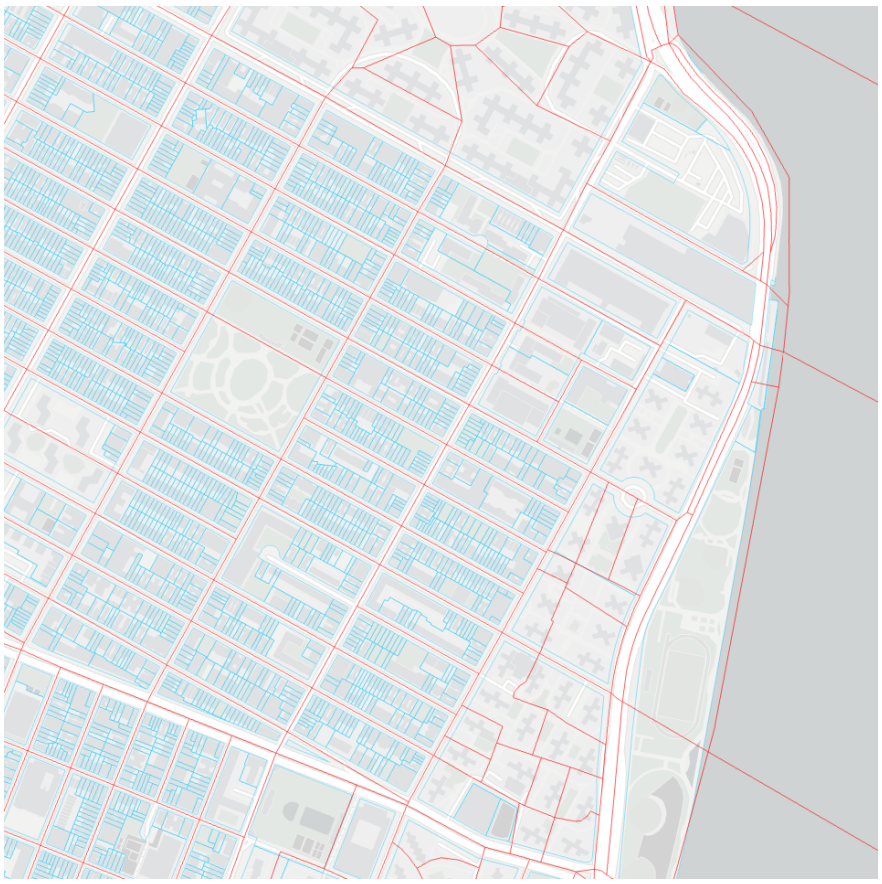


Figure 4: overlay

Step #03: join building energy data

Join the geoID data to the building energy dataset using the field of BBL. This results in a new building energy use data table with a new column of Census Block geoID information it belongs to.

```
# load benchmarking table
BR <- read.csv(file = paste0(file.path(dirname(rstudioapi::getSourceEditorContext()$path)),
                             "/dataset/NYC_Building_Energy_with_GEOID/table.csv"),
               header=TRUE) %>%
  data.frame()

head(BR)
```

```
##   OID_ Record_Number Order_        BBL Coreported_BBL_Status BBLs_Coreported
## 1   NA            NA  12515 2034327503
## 2   NA       2666985   6733 2036000004
## 3   NA       2639029   6751 2036720001
## 4   NA            NA  12519 2036040001
## 5   NA       2682362   6722 2035630005
## 6   NA       4402905   5381 2023090001
##             Reported_BINs                        Property_Name
## 1
## 2                 2092718           731-755 White Plains Road
## 3                 2022645       1921-1965 Lafayette Park Lane
## 4
## 5 2116665;2819754;2819749          Gold - 669 White Plains Rd
## 6                2093995  Carnegie Management: 112 Lincoln Ave
##                   Parent_Property_Id                Parent_Property_Name
## 1
## 2 Not Applicable: Standalone Property Not Applicable: Standalone Property
## 3 Not Applicable: Standalone Property Not Applicable: Standalone Property
## 4
## 5 Not Applicable: Standalone Property Not Applicable: Standalone Property
## 6 Not Applicable: Standalone Property Not Applicable: Standalone Property
##   Street_Number          Street_Name Zip_Code Borough
## 1           329 ADMIRAL LANE            10473   Bronx
## 2          1850 LAFAYETTE AVENUE        10473   Bronx
## 3          1965 LAFAYETTE AVENUE        10473   Bronx
## 4           700 WHITE PLAINS ROAD       10473   Bronx
## 5           669 WHITE PLAINS ROAD       10473   Bronx
## 6           112 LINCOLN AVENUE          10454   Bronx
##   DOF_Benchmarking_Submission_Status Primary_Property_Type
## 1                       In Violation
## 2                     In Compliance   Multifamily Housing
## 3                     In Compliance   Multifamily Housing
## 4                       In Violation
## 5                     In Compliance   Multifamily Housing
## 6                     In Compliance   Multifamily Housing
##                                    List_of_All_Property_Use_Types_at_Property
## 1
## 2 Medical Office, Multifamily Housing, Retail Store, Supermarket/Grocery Store
## 3                                                          Multifamily Housing
## 4
## 5                                    Multifamily Housing, Office, Parking
## 6                                                          Multifamily Housing
```

```
##   Largest_Property_Use_Type Largest_Property_Use_Type_Gross_Floor_Area_sqft
## 1                                                                        NA
## 2       Multifamily Housing                                          606627
## 3       Multifamily Housing                                          400933
## 4                                                                        NA
## 5       Multifamily Housing                                           50604
## 6       Multifamily Housing                                           90000
##   X2nd_Largest_Property_Use_Type
## 1
## 2                Medical Office
## 3                 Not Available
## 4
## 5                       Parking
## 6                 Not Available
##   X2nd_Largest_Property_Use_Type_Gross_Floor_Area_sqft
## 1
## 2                                                36180
## 3                                        Not Available
## 4
## 5                                                 8635
## 6                                        Not Available
##   X3rd_Largest_Property_Use_Type
## 1
## 2     Supermarket/Grocery Store
## 3                 Not Available
## 4
## 5                        Office
## 6                 Not Available
##   X3rd_Largest_Property_Use_Type_Gross_Floor_Area_sqft Year_Built
## 1                                                           NA
## 2                                                 9638       1977
## 3                                        Not Available       1969
## 4                                                           NA
## 5                                                  334       2009
## 6                                        Not Available       1920
##   Number_of_Buildings_Self_reported Occupancy Metered_Areas_Energy
## 1                                NA        NA
## 2                                 0       100       Whole Building
## 3                                 1       100       Whole Building
## 4                                NA        NA
## 5                                 1       100       Whole Building
## 6                                 1       100       Whole Building
##   Metered_Areas_Water ENERGY_STAR_Score Site_EUI_kBtu_per_sqft
## 1                                   NA                     NA
## 2       Not Available               94                   70.9
## 3       Not Available               61                   89.0
## 4                                   NA                     NA
## 5       Not Available               NA                   78.1
## 6       Not Available               74                   46.2
##   Weather_Normalized_Site_EUI_kBtu_per_sqft
## 1                                        NA
## 2                                      70.9
## 3                                        NA
## 4                                        NA
```

```
## 5                                    79.0
## 6                                    46.1
##   Weather_Normalized_Site_Electricity_Intensity_kWh_per_sqft
## 1                                                         NA
## 2                                                        3.1
## 3                                                        4.6
## 4                                                         NA
## 5                                                        7.4
## 6                                                        7.0
##   Weather_Normalized_Site_Natural_Gas_Intensity_therms_per_sqft
## 1                                                            NA
## 2                                                           0.6
## 3                                                           0.6
## 4                                                            NA
## 5                                                           0.5
## 6                                                           0.2
##   Source_EUI_kBtu_per_sqft Weather_Normalized_Source_EUI_kBtu_per_sqft
## 1                       NA                                         NA
## 2                     97.6                                       96.5
## 3                    126.6                                         NA
## 4                       NA                                         NA
## 5                    135.7                                      136.0
## 6                     98.5                                       98.3
##   Fuel_Oil_1_Use__kBtu Fuel_Oil_2_Use__kBtu Fuel_Oil_4_Use__kBtu
## 1
## 2       Not Available       Not Available       Not Available
## 3       Not Available           4124640.6       Not Available
## 4
## 5       Not Available       Not Available       Not Available
## 6       Not Available       Not Available       Not Available
##   Fuel_Oil_5_6_Use__kBtu Diesel_2_Use_kBtu District_Steam_Use_kBtu
## 1
## 2         Not Available    Not Available          Not Available
## 3         Not Available    Not Available          Not Available
## 4
## 5         Not Available    Not Available          Not Available
## 6         Not Available    Not Available          Not Available
##   District_Hot_Water_Use_kBtu District_Chilled_Water_Use_kBtu
## 1
## 2              Not Available                  Not Available
## 3              Not Available                  Not Available
## 4
## 5              Not Available                  Not Available
## 6              Not Available                  Not Available
##   Natural_Gas_Use_kBtu Weather_Normalized_Site_Natural_Gas_Use_therms
## 1                   NA                                            NA
## 2             39383800                                      396617.5
## 3             25111336                                      255164.1
## 4                   NA                                            NA
## 5              2669935                                       27288.8
## 6              2008800                                       19935.2
##   Electricity_Use_Grid_Purchase_kBtu Weather_Normalized_Site_Electricity_kWh
## 1                                 NA                                      NA
## 2                            7284016                               2042628.6
```

```
## 3                              6438830                                1845628.4
## 4                                   NA                                      NA
## 5                              1308722                                 379132.1
## 6                              2151972                                 630706.9
##   Total_GHG_Emissions_Metric_Tons_CO2e Direct_GHG_Emissions_Metric_Tons_CO2e
## 1                                   NA                                    NA
## 2                               2696.0                                2091.9
## 3                               2173.9                                1639.9
## 4                                   NA                                    NA
## 5                                250.4                                 141.8
## 6                                285.2                                 106.7
##   Indirect_GHG_Emissions_Metric_Tons_CO2e DOF_Property_Floor_Area_sqft
## 1                                      NA                       110952
## 2                                   604.1                      1021752
## 3                                   534.0                       400932
## 4                                      NA                        78347
## 5                                   108.5                        58234
## 6                                   178.5                        89275
##   Property_GFA_Self_reported_sqft Water_Use_All_Water_Sources_kgal
## 1                              NA                               NA
## 2                          657996                               NA
## 3                          400933                               NA
## 4                              NA                               NA
## 5                           50938                             3821
## 6                           90000                               NA
##   Municipally_Supplied_Potable_Water_Indoor_Intensity_gal_per_sqft
## 1                                                              NA
## 2                                                              NA
## 3                                                              NA
## 4                                                              NA
## 5                                                           75.01
## 6                                                              NA
##      Release_Date DEP_Provided_Water_Use_kgal
## 1                                          NA
## 2 6/17/2016 11:59                          NA
## 3 5/27/2016 10:47                          NA
## 4                                          NA
## 5 7/31/2016 17:10                        3821
## 6 4/27/2016 10:08                          NA
##   Automatic_Water_Benchmarking_Eligible Reported_Water_Method      BBL_1
## 1                                                             2034327503
## 2                                                             2036000004
## 3                                                             2036720001
## 4                                                             2036040001
## 5                              Eligible                   ABS 2035630005
## 6                                                             2023090001
##      GEOID10 Shape_Length Shape_Area
## 1 3.6005e+14    1545.1905  125824.45
## 2 3.6005e+14    4510.0241  167236.53
## 3 3.6005e+14    1533.9754  113607.34
## 4 3.6005e+14     812.9707   41278.00
## 5 3.6005e+14    1014.5318   61452.82
## 6 3.6005e+14     846.4841   23530.14
```

Section #03: Next Step - Join ACS dataset to building energy dataset using the BR above