

GIS Lab 10

Camille Okonkwo

Contents

Load Packages	2
Read in shapefile	2
Scatter plots	3
Multiple Regression	7
Geographically-Weighted Regression	9
Mapping Results	12
Local R-Square	12
Percent 65 years and over	14
Percent High School Diploma	17

Load Packages

```
library(sf)
library(tmap)
library(tidyverse)
library(spgwr)
library(jtools)
library(tmapttools)
library(broom.mixed)
```

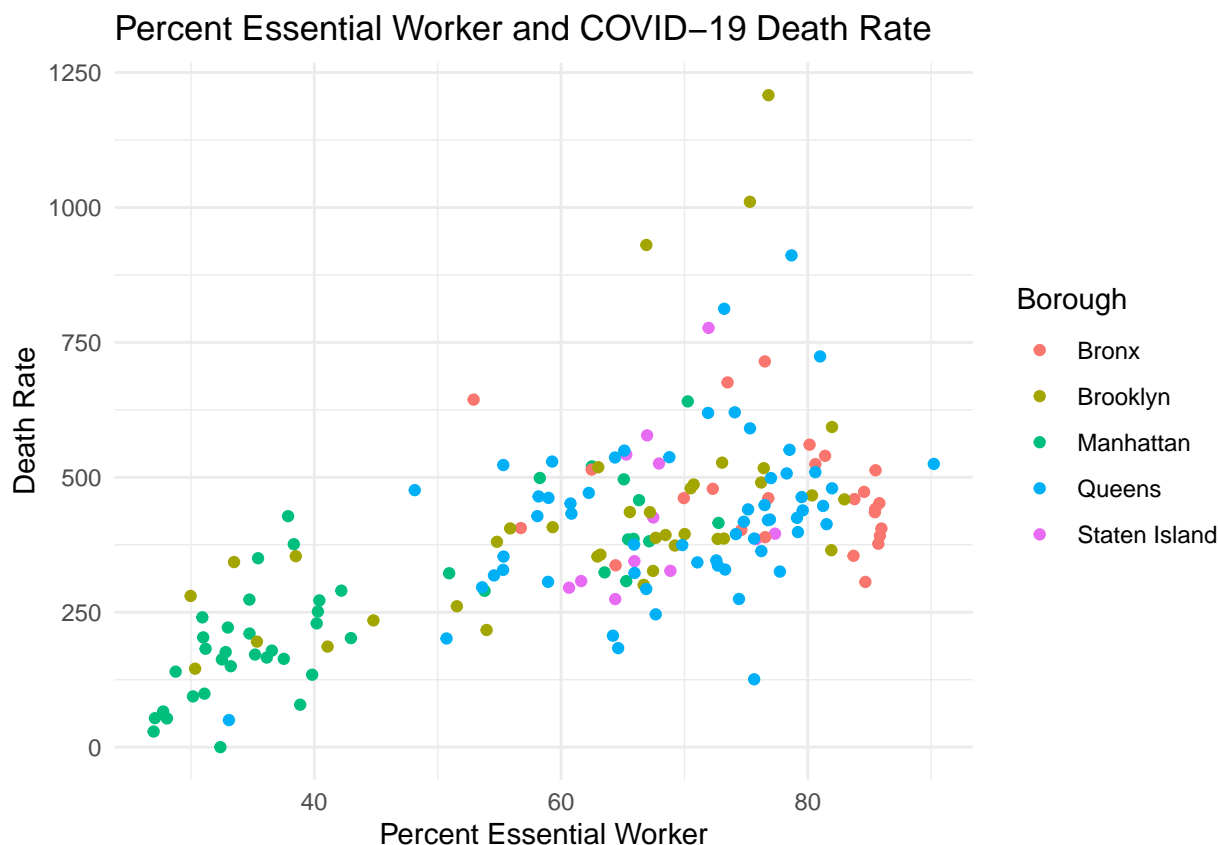
Read in shapefile

```
data = st_read ("data/NYC_zip_COVID.shp")

## Reading layer `NYC_zip_COVID' from data source
##   `/Users/camilleokonkwo/Desktop/Spring 2024/GIS/GIS_lab10/data/NYC_zip_COVID.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 177 features and 13 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 913176 ymin: 120122 xmax: 1067382 ymax: 272844
## Projected CRS: NAD83 / New York Long Island (ftUS)
```

Scatter plots

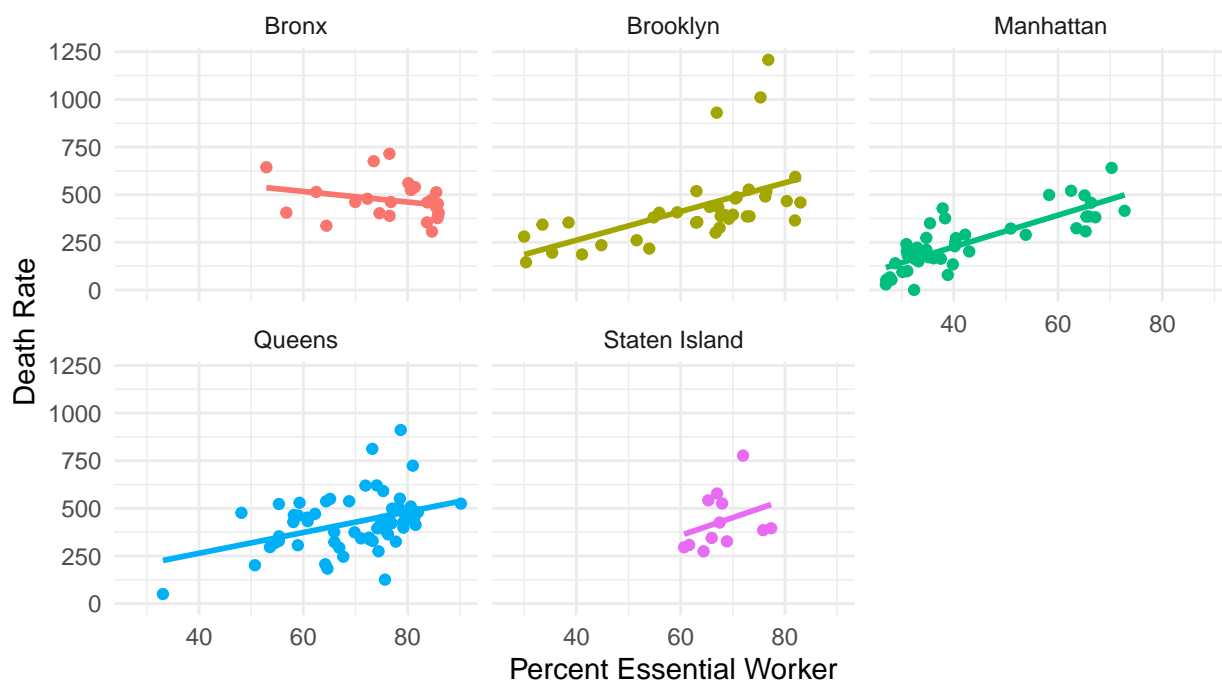
```
# Essential Workers
data |>
  ggplot(aes(x = per_ssn, y = deth_rt, color = CovidZip_2)) +
  geom_point() +
  labs(title = "Percent Essential Worker and COVID-19 Death Rate",
       x = "Percent Essential Worker",
       y = "Death Rate",
       color = "Borough") +
  theme_minimal()
```



```
# Add linear fit and facet
data |>
  ggplot(aes(x = per_ssn, y = deth_rt, color = CovidZip_2)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  labs(title = "Percent Essential Worker and COVID-19 Death Rate",
       x = "Percent Essential Worker",
       y = "Death Rate",
       color = "Borough") +
  facet_wrap(~CovidZip_2) +
  theme_minimal() +
  theme(legend.position="bottom")
```

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

Percent Essential Worker and COVID-19 Death Rate

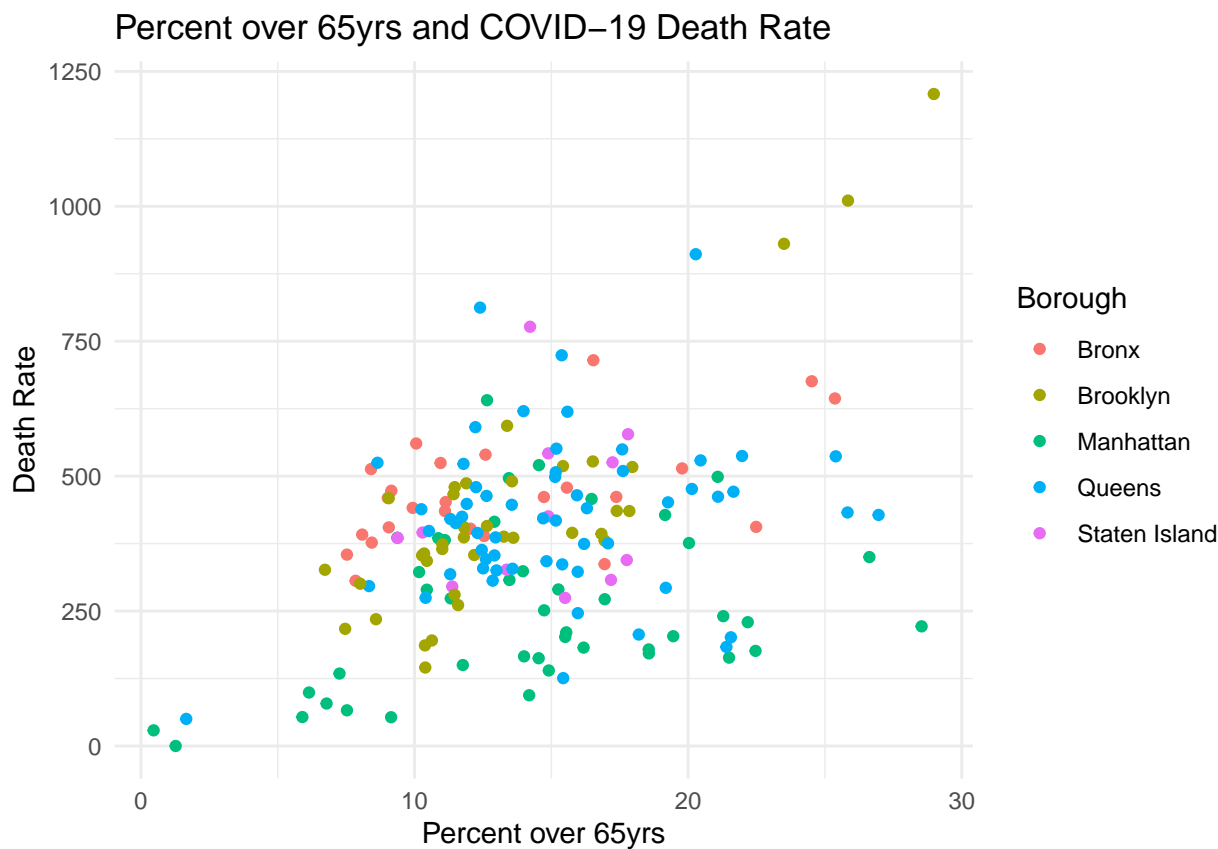


Borough —●— Bronx —●— Brooklyn —●— Manhattan —●— Queens —●— Staten Island

Percent 65yrs and above

data |>

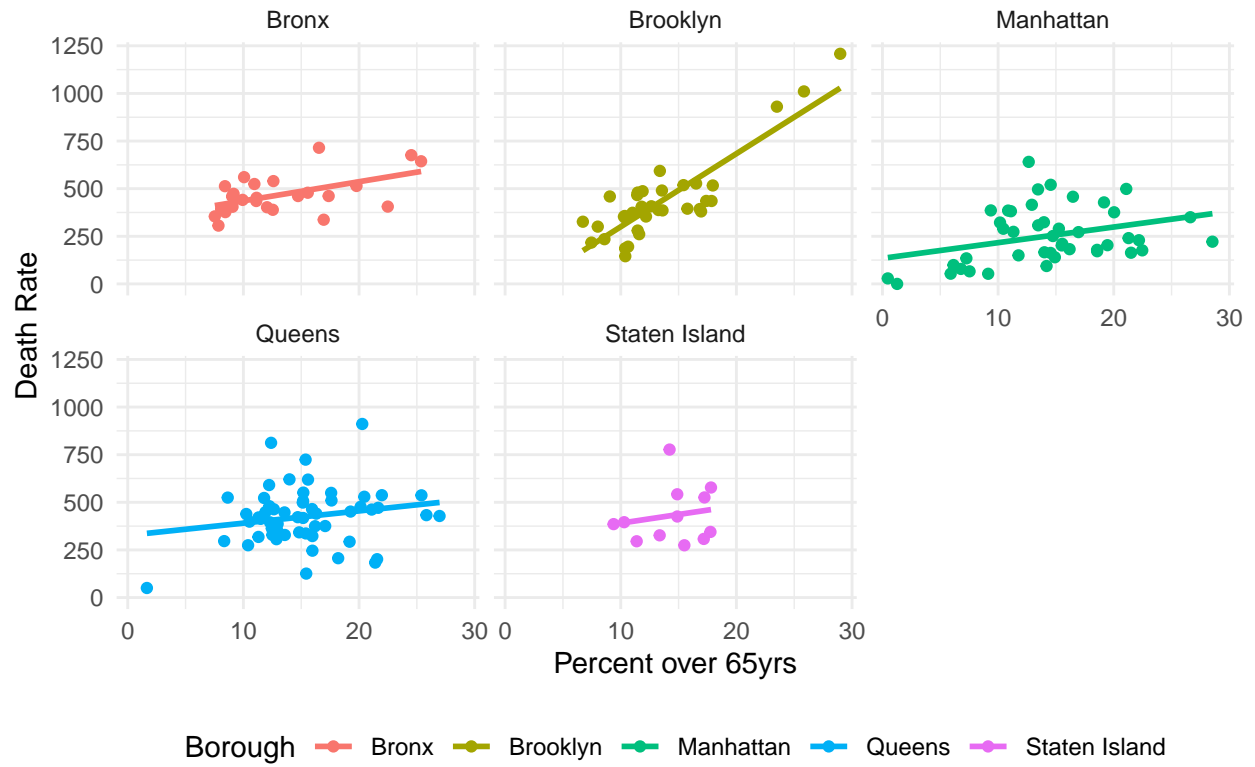
```
ggplot(aes(x = pr_pp65, y = deth_rt, color = CovidZip_2 )) +  
  geom_point() +  
  labs(title = "Percent over 65yrs and COVID-19 Death Rate",  
        x = "Percent over 65yrs",  
        y = "Death Rate",  
        color = "Borough") +  
  theme_minimal()
```



```
# Add linear fit and facet
data |>
  ggplot(aes(x = pr_pp65, y = deth_rt, color = CovidZip_2 )) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  labs(title = "Percent over 65yrs and COVID-19 Death Rate",
       x = "Percent over 65yrs",
       y = "Death Rate",
       color = "Borough") +
  facet_wrap(~CovidZip_2) +
  theme_minimal() +
  theme(legend.position="bottom")
```

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

Percent over 65yrs and COVID-19 Death Rate



Multiple Regression

```
mreg = lm(deth_rt ~ pr_pp65 + per_ssn + pr_hghd + med_inc + pr_pbtr,
          data = data)
```

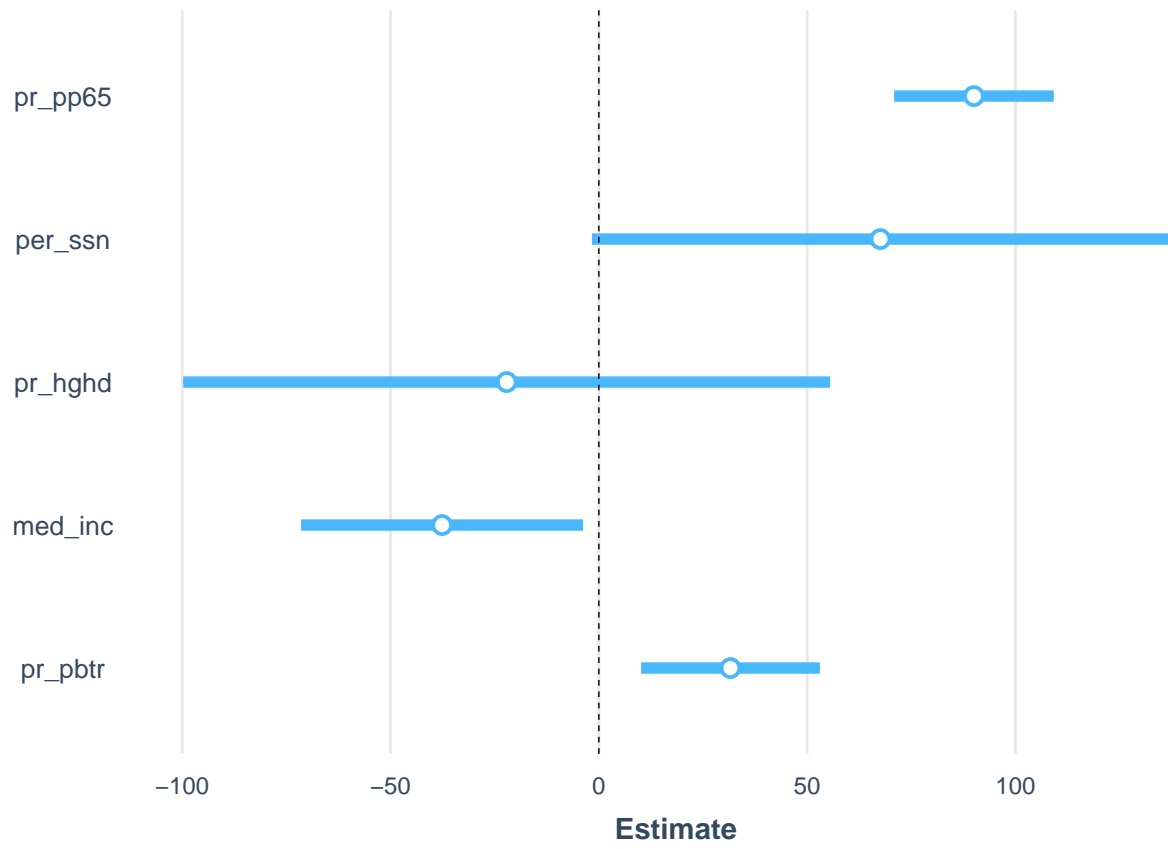
```
summary(mreg)
```

```
##
## Call:
## lm(formula = deth_rt ~ pr_pp65 + per_ssn + pr_hghd + med_inc +
##      pr_pbtr, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -265.24  -57.94  -17.50   38.47  411.91
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.015e+02  2.023e+02  -0.502  0.61645
## pr_pp65      1.787e+01  1.925e+00   9.283 < 2e-16 ***
## per_ssn      3.952e+00  2.051e+00   1.927  0.05569 .
## pr_hghd     -1.019e+00  1.811e+00  -0.563  0.57445
## med_inc     -1.094e-03  4.985e-04  -2.195  0.02948 *
## pr_pbtr      2.067e+00  7.113e-01   2.906  0.00414 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 108.6 on 171 degrees of freedom
## Multiple R-squared:  0.6339, Adjusted R-squared:  0.6232
## F-statistic: 59.22 on 5 and 171 DF,  p-value: < 2.2e-16
```

```
AIC(mreg)
```

```
## [1] 2169.66
```

```
plot_summs(mreg, scale = TRUE, inner_ci_level = .95)
```



Geographically-Weighted Regression

```
# Fixed Bandwidth
class(data)

## [1] "sf"          "data.frame"

data = as_Spatial(data)

fixed_bw = gwr.sel(deth_rt ~ pr_pp65 + per_ssn + pr_hghd + med_inc + pr_pbtr,
  data = data,
  adapt=FALSE,
  gweight = gwr.Gauss,
  verbose = TRUE)

## Bandwidth: 78496.56 CV score: 2141539
## Bandwidth: 126883.3 CV score: 2203255
## Bandwidth: 48591.91 CV score: 2010631
## Bandwidth: 30109.82 CV score: 1775233
## Bandwidth: 18687.26 CV score: 1679980
## Bandwidth: 11627.73 CV score: 1742741
## Bandwidth: 19925.84 CV score: 1659436
## Bandwidth: 23815.77 CV score: 1644578
## Bandwidth: 22638.03 CV score: 1634948
## Bandwidth: 22302.59 CV score: 1634677
## Bandwidth: 22387.45 CV score: 1634638
## Bandwidth: 22390.51 CV score: 1634638
## Bandwidth: 22390.66 CV score: 1634638
## Bandwidth: 22390.66 CV score: 1634638
## Bandwidth: 22390.66 CV score: 1634638
## Bandwidth: 22390.66 CV score: 1634638
## Bandwidth: 22390.66 CV score: 1634638

fbw_gwr = gwr(deth_rt ~ pr_pp65 + per_ssn + pr_hghd + med_inc + pr_pbtr,
  data = data,
  bandwidth = fixed_bw,
  gweight=gwr.Gauss,
  hatmatrix=TRUE,
  se.fit=TRUE)

fbw_gwr

## Call:
## gwr(formula = deth_rt ~ pr_pp65 + per_ssn + pr_hghd + med_inc +
##      pr_pbtr, data = data, bandwidth = fixed_bw, gweight = gwr.Gauss,
##      hatmatrix = TRUE, se.fit = TRUE)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 22390.66
## Summary of GWR coefficient estimates at data points:
##           Min.      1st Qu.      Median      3rd Qu.      Max.
## X.Intercept. -9.1573e+02 -3.0161e+02 -2.5626e+01  2.2154e+02  7.0328e+02
## pr_pp65       1.1584e+01  1.5056e+01  2.0538e+01  2.3903e+01  2.9517e+01
## per_ssn       -8.8985e-01  7.0495e-01  2.4654e+00  4.6193e+00  1.3216e+01
## pr_hghd       -5.3821e+00 -4.3866e+00 -3.4063e+00 -1.5629e+00  1.1298e+01
## med_inc       -8.0191e-03 -1.2836e-03 -6.5646e-04 -2.2671e-04 -5.6039e-05
```

```
## pr_pbtr      -2.5012e+00  1.9214e+00  2.3330e+00  3.4876e+00  6.3473e+00
##              Global
## X.Intercept. -101.4925
## pr_pp65      17.8703
## per_ssn      3.9521
## pr_hghd      -1.0188
## med_inc      -0.0011
## pr_pbtr      2.0671
## Number of data points: 177
## Effective number of parameters (residual: 2traceS - traceS'S): 29.4049
## Effective degrees of freedom (residual: 2traceS - traceS'S): 147.5951
## Sigma (residual: 2traceS - traceS'S): 86.659
## Effective number of parameters (model: traceS): 22.45424
## Effective degrees of freedom (model: traceS): 154.5458
## Sigma (model: traceS): 84.68785
## Sigma (ML): 79.13398
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 2104.117
## AIC (GWR p. 96, eq. 4.22): 2072.143
## Residual sum of squares: 1108407
## Quasi-global R2: 0.7988258

names(fbw_gwr)

## [1] "SDF"          "lhat"         "lm"           "results"      "bandwidth"    "adapt"
## [7] "hatmatrix"    "gweight"      "gTSS"         "this.call"    "fp.given"     "timings"

# Grab results
results = fbw_gwr$SDF

names(results)

## [1] "sum.w"          "X.Intercept."  "pr_pp65"
## [4] "per_ssn"        "pr_hghd"       "med_inc"
## [7] "pr_pbtr"        "X.Intercept._se" "pr_pp65_se"
## [10] "per_ssn_se"     "pr_hghd_se"    "med_inc_se"
## [13] "pr_pbtr_se"     "gwr.e"         "pred"
## [16] "pred.se"        "localR2"       "X.Intercept._se_EDF"
## [19] "pr_pp65_se_EDF" "per_ssn_se_EDF" "pr_hghd_se_EDF"
## [22] "med_inc_se_EDF" "pr_pbtr_se_EDF" "pred.se.1"

summary(results)

## Object of class SpatialPolygonsDataFrame
## Coordinates:
##      min      max
## x 913176 1067382
## y 120122  272844
## Is projected: TRUE
## proj4string :
## [+proj=lcc +lat_0=40.1666666666667 +lon_0=-74 +lat_1=41.0333333333333
## +lat_2=40.6666666666667 +x_0=300000 +y_0=0 +datum=NAD83 +units=us-ft
## +no_defs]
## Data attributes:
##      sum.w      X.Intercept.      pr_pp65      per_ssn
## Min.   : 3.674   Min.   :-915.73   Min.   :11.58   Min.   : -0.8899
## 1st Qu.:32.226   1st Qu.: -301.61   1st Qu.:15.06   1st Qu.: 0.7050
```

```

## Median :48.148      Median : -25.63      Median :20.54      Median : 2.4654
## Mean   :44.141      Mean   : -78.60      Mean   :19.80      Mean   : 3.2588
## 3rd Qu.:56.595      3rd Qu.: 221.54      3rd Qu.:23.90      3rd Qu.: 4.6193
## Max.   :68.115      Max.   : 703.28      Max.   :29.52      Max.   :13.2165
##      pr_hghd      med_inc      pr_pbtr      X.Intercept._se
## Min.   : -5.382      Min.   : -8.019e-03      Min.   : -2.501      Min.   : 197.1
## 1st Qu.: -4.387      1st Qu.: -1.284e-03      1st Qu.: 1.921      1st Qu.: 211.4
## Median : -3.406      Median : -6.565e-04      Median : 2.333      Median : 231.6
## Mean   : -1.957      Mean   : -1.121e-03      Mean   : 2.579      Mean   : 275.6
## 3rd Qu.: -1.563      3rd Qu.: -2.267e-04      3rd Qu.: 3.488      3rd Qu.: 286.6
## Max.   :11.298      Max.   : -5.604e-05      Max.   : 6.347      Max.   :1510.0
##      pr_pp65_se      per_ssn_se      pr_hghd_se      med_inc_se
## Min.   : 1.777      Min.   : 1.950      Min.   : 1.715      Min.   :0.0004281
## 1st Qu.: 1.930      1st Qu.: 2.109      1st Qu.: 1.868      1st Qu.:0.0004603
## Median : 2.158      Median : 2.325      Median : 2.079      Median :0.0005301
## Mean   : 2.665      Mean   : 2.759      Mean   : 2.470      Mean   :0.0006876
## 3rd Qu.: 2.901      3rd Qu.: 2.892      3rd Qu.: 2.637      3rd Qu.:0.0007616
## Max.   :14.336      Max.   :15.427      Max.   :15.761      Max.   :0.0041352
##      pr_pbtr_se      gwr.e      pred      pred.se
## Min.   :0.7330      Min.   : -173.1870      Min.   : -54.62      Min.   : 9.415
## 1st Qu.:0.8390      1st Qu.: -50.8168      1st Qu.:307.01      1st Qu.:16.141
## Median :0.8862      Median : -8.9492      Median :417.24      Median :19.483
## Mean   :1.0221      Mean   : 0.5365      Mean   :389.21      Mean   :22.854
## 3rd Qu.:1.0579      3rd Qu.: 33.5278      3rd Qu.:467.63      3rd Qu.:24.599
## Max.   :5.4145      Max.   : 321.1994      Max.   :950.32      Max.   :82.169
##      localR2      X.Intercept._se_EDF      pr_pp65_se_EDF      per_ssn_se_EDF
## Min.   :0.5621      Min.   : 201.7      Min.   : 1.819      Min.   : 1.995
## 1st Qu.:0.6807      1st Qu.: 216.3      1st Qu.: 1.975      1st Qu.: 2.158
## Median :0.7769      Median : 237.0      Median : 2.208      Median : 2.380
## Mean   :0.7560      Mean   : 282.0      Mean   : 2.727      Mean   : 2.823
## 3rd Qu.:0.8400      3rd Qu.: 293.3      3rd Qu.: 2.968      3rd Qu.: 2.959
## Max.   :0.8604      Max.   :1545.2      Max.   :14.669      Max.   :15.786
##      pr_hghd_se_EDF      med_inc_se_EDF      pr_pbtr_se_EDF      pred.se.1
## Min.   : 1.755      Min.   :0.0004381      Min.   :0.7500      Min.   : 9.634
## 1st Qu.: 1.912      1st Qu.:0.0004710      1st Qu.:0.8585      1st Qu.:16.517
## Median : 2.127      Median :0.0005424      Median :0.9069      Median :19.937
## Mean   : 2.527      Mean   :0.0007037      Mean   :1.0459      Mean   :23.386
## 3rd Qu.: 2.699      3rd Qu.:0.0007794      3rd Qu.:1.0825      3rd Qu.:25.171
## Max.   :16.128      Max.   :0.0042314      Max.   :5.5405      Max.   :84.082

```

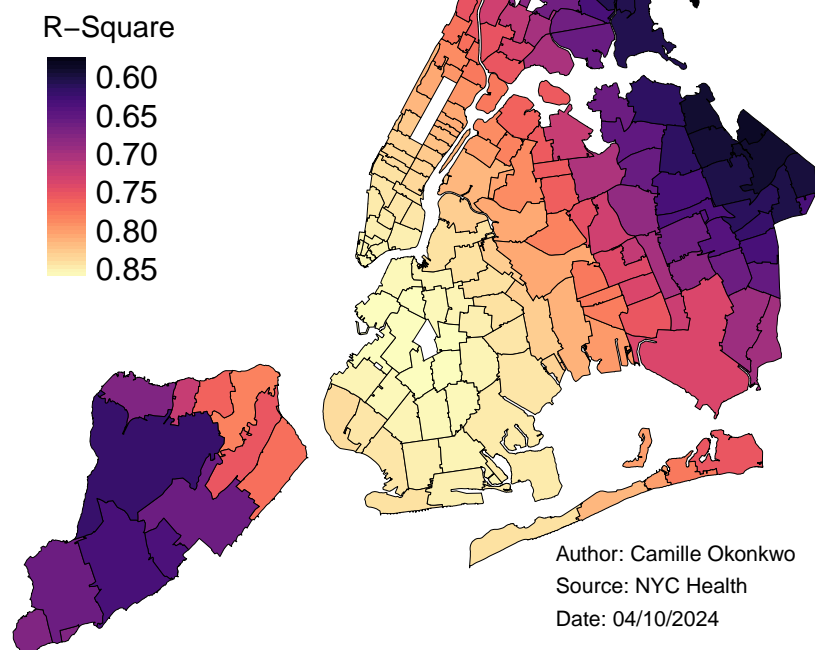
Mapping Results

Local R-Square

```
# Local R Square
r2 =
  tm_shape(results) +
  tm_polygons(col = "localR2",
    palette = "magma",
    style = "cont",
    border.col = "black",
    lwd = .5,
    title = "R-Square") +
  tm_legend(legend.position = c(0.05, 0.450),
    legend.text.size = 1) +
  tm_layout(title = "Local R-Square",
    title.size = 2,
    title.position = c(0.0, .89),
    frame = FALSE) +
  tm_credits("Author: Camille Okonkwo \nSource: NYC Health \nDate: 04/10/2024",
    size = .7,
    position = c(.66,.04))
```

r2

Local R-Square



```
tmap_save(r2,
  dpi = 300,
  width = 6,
  height = 6,
```

```
filename = "local_r2.png")
```

```
## Map saved to /Users/camilleokonkwo/Desktop/Spring 2024/GIS/GIS_lab10/local_r2.png
```

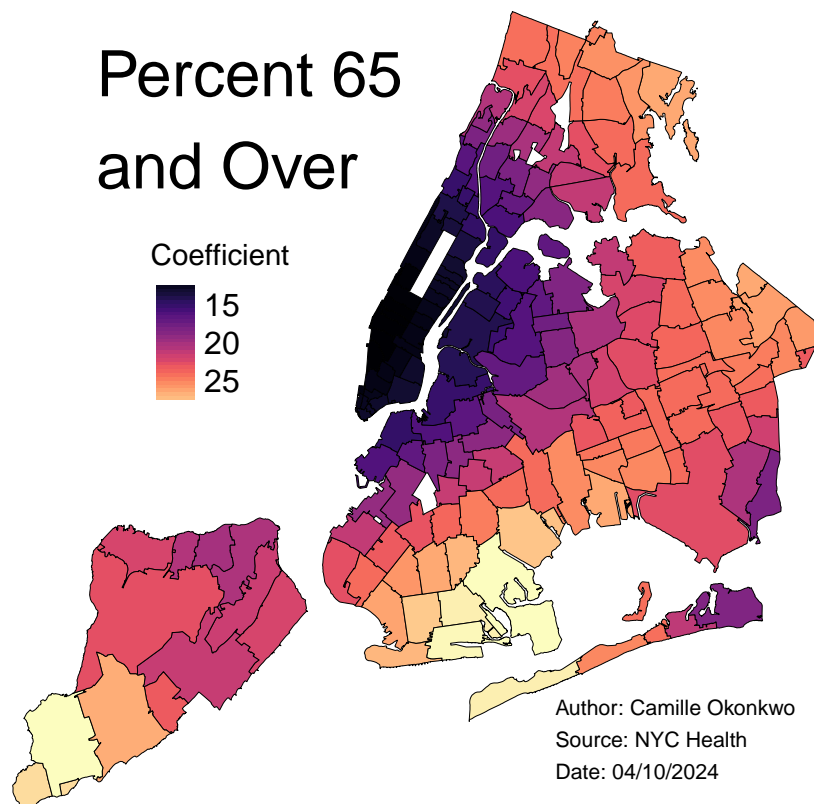
```
## Resolution: 1800 by 1800 pixels
```

```
## Size: 6 by 6 inches (300 dpi)
```

Percent 65 years and over

```
coef_65yrs =
  tm_shape(results) +
  tm_polygons(col = "pr_pp65",
    palette = "magma",
    style = "cont",
    border.col = "black",
    lwd = .5,
    title = "Coefficient") +
  tm_legend(legend.position = c(0.18, 0.50),
    legend.text.size = 1) +
  tm_layout(title = "Percent 65 \nand Over",
    title.size = 2,
    title.position = c(0.12, .85),
    frame = FALSE) +
  tm_credits("Author: Camille Okonkwo \nSource: NYC Health \nDate: 04/10/2024",
    size = .7,
    position = c(.66,.04))
```

```
coef_65yrs
```



```
tmap_save(coef_65yrs,
  dpi = 300,
  width = 6,
  height = 6,
  filename = "65yr_coef.png")
```

```
## Map saved to /Users/camilleokonkwo/Desktop/Spring 2024/GIS/GIS_lab10/65yr_coef.png
```

```
## Resolution: 1800 by 1800 pixels
```

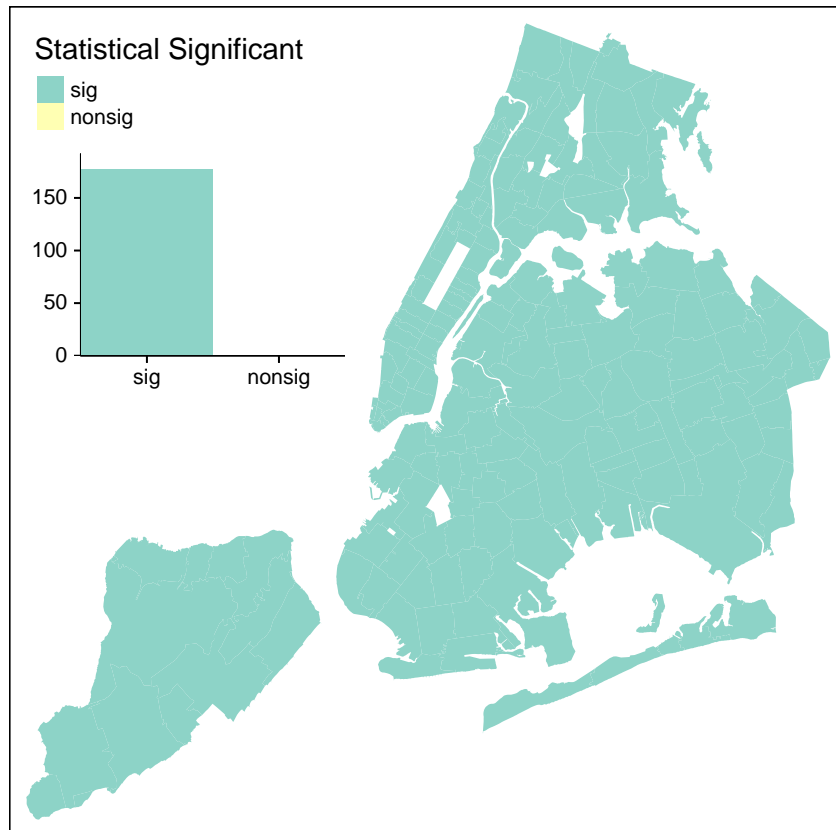
```
## Size: 6 by 6 inches (300 dpi)
```

Significance of 65 years and over

```
results$tval_65 = results$pr_pp65/results$pr_pp65_se

results$t_65_cat = cut(results$tval_65,
                        breaks=c(min(results$tval_65), -2, 2,
                                max(results$tval_65)),
                        labels=c("sig", "nonsig", "sig"))

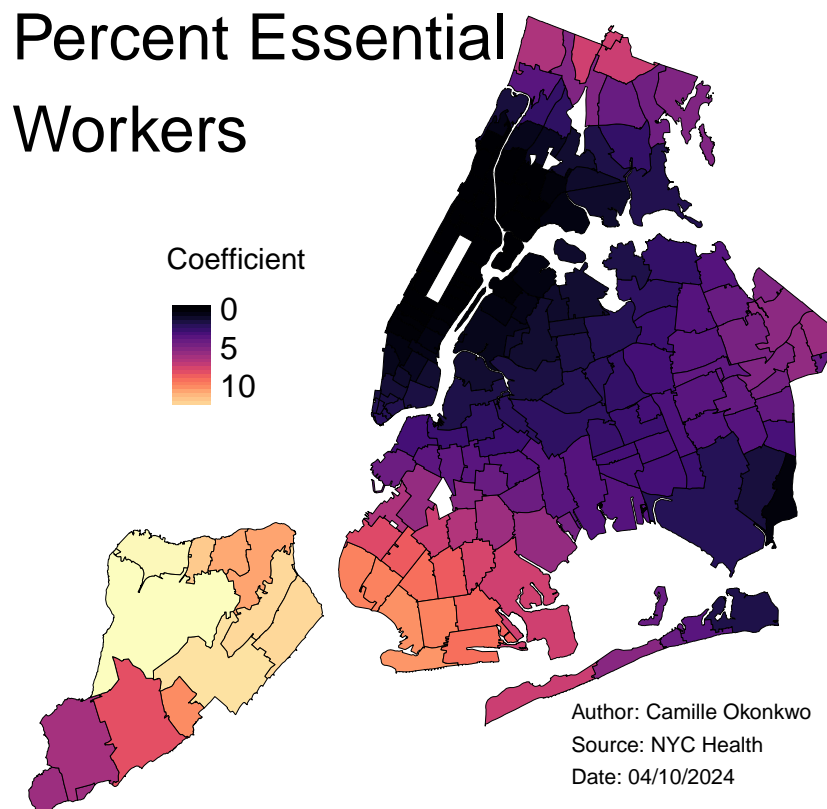
tm_shape(results) +
  tm_fill(col = "t_65_cat",
          title = "Statistical Significant",
          legend.hist = TRUE)
```



```
## Percent Essential Workers
```

```
essen = tm_shape(results) +
  tm_polygons(col = "per_ssn",
    palette = "magma",
    style = "cont",
    border.col = "black",
    lwd = .5,
    title = "Coefficient") +
  tm_legend(legend.position = c(0.18, 0.50),
    legend.text.size = 1) +
  tm_layout(title = "Percent Essential \nWorkers",
    title.size = 2,
    title.position = c(0.0, .9),
    frame = FALSE) +
  tm_credits("Author: Camille Okonkwo \nSource: NYC Health \nDate: 04/10/2024",
    size = .7,
    position = c(.66,.04))
```

```
essen
```



```
tmap_save(essen,
  dpi = 300,
  width = 6,
  height = 6,
  filename = "essential_coef.png")
```

```
## Map saved to /Users/camilleokonkwo/Desktop/Spring 2024/GIS/GIS_lab10/essential_coef.png
```

```
## Resolution: 1800 by 1800 pixels
```



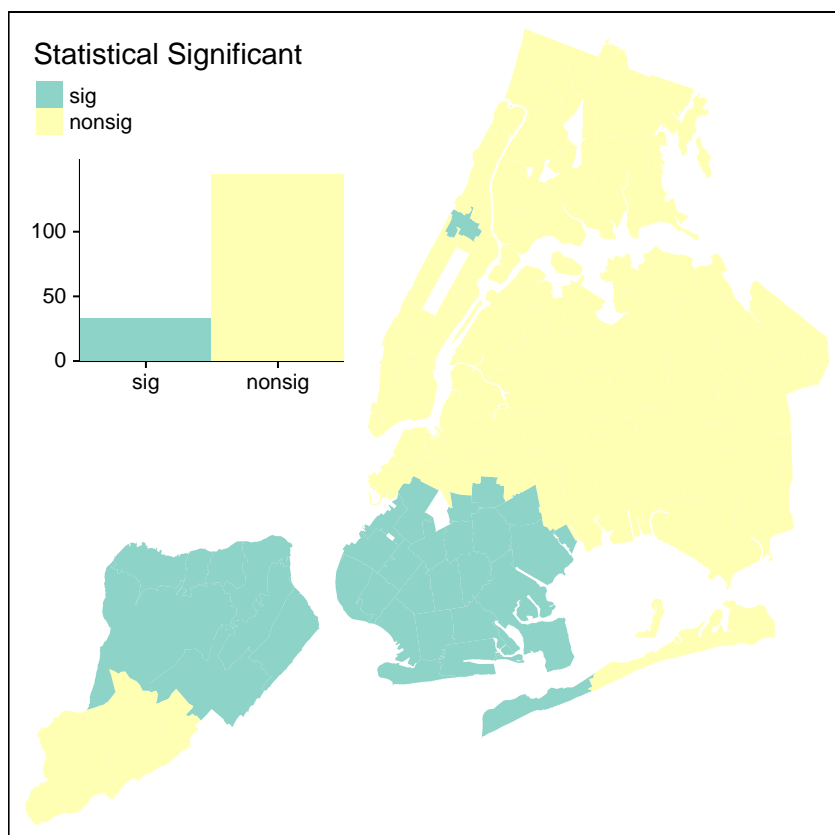
```
## Size: 6 by 6 inches (300 dpi)
```

Significance of Essential Workers

```
results$tval_essen = results$per_ssn/results$per_ssn_se

results$t_essen_cat = cut(results$tval_essen,
                           breaks=c(min(results$tval_essen), -2, 2, max(results$tval_essen)),
                           labels=c("sig","nonsig", "sig"))

tm_shape(results) +
  tm_fill(col = "t_essen_cat",
          title = "Statistical Significant",
          legend.hist = TRUE)
```



Percent High School Diploma

```
hs_diploma = tm_shape(results) +
  tm_polygons(col = "pr_hghd",
              palette = "magma",
              style = "cont",
              midpoint = 0,
              border.col = "black",
              lwd = .5,
              title = "Coefficient") +
  tm_legend(legend.position = c(0.0, 0.45),
```

```

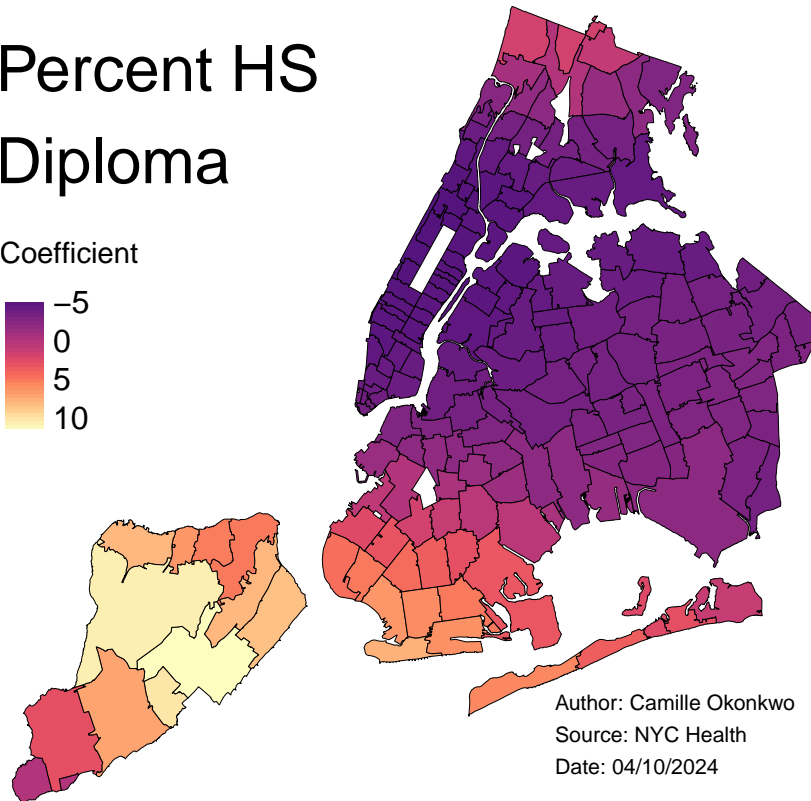
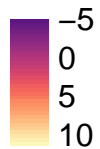
    legend.text.size = 1) +
tm_layout(title = "Percent HS \nDiploma",
           title.size = 2,
           title.position = c(0., .85),
           frame = FALSE) +
tm_credits("Author: Camille Okonkwo \nSource: NYC Health \nDate: 04/10/2024",
           size = .7,
           position = c(.66,.04))

```

hs_diploma

Percent HS Diploma

Coefficient



```

tmap_save(hs_diploma,
          dpi = 300,
          width = 6,
          height = 6,
          filename = "hs_diploma.png")

```

Map saved to /Users/camilleokonkwo/Desktop/Spring 2024/GIS/GIS_lab10/hs_diploma.png

Resolution: 1800 by 1800 pixels

Size: 6 by 6 inches (300 dpi)

Significance of High School Diploma

```

results$tval_hs = results$pr_hghd/results$pr_hghd_se

results$t_hs_cat = cut(results$tval_hs,
                       breaks=c(min(results$tval_hs), -2, 2, max(results$tval_hs)),

```

```
labels=c("sig","nonsig", "sig"))

tm_shape(results) +
  tm_fill(col = "t_hs_cat",
          title = "Statistical Significant",
          legend.hist = TRUE)
```

