Linear Regression and Morans I test

```r
# Import the needed packages
library(rgdal)
```

```
## Loading required package: sp
```

```
## rgdal: version: 1.3-4, (SVN revision 766)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
##  Path to GDAL shared files: C:/Users/Robert/Documents/R/win-library/3.5/rgdal/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: C:/Users/Robert/Documents/R/win-library/3.5/rgdal/proj
##  Linking to sp version: 1.3-1
```

```r
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0      v purrr   0.2.5
## v tibble  1.4.2      v dplyr   0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(sf)
```

```
## Linking to GEOS 3.6.1, GDAL 2.2.3, proj.4 4.9.3
```

```r
library(ggplot2)
library(broom)
library(tmap)
library(spdep)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
##
##      expand
```

```
## Loading required package: spData
```

```
## To access larger datasets in this package, install the spDataLarge
## package with: `install.packages('spDataLarge',
## repos='https://nowosad.github.io/drat/', type='source'))`
```

```
library(sp)


# Import shapefile of census tracts that has all our key data attached to it
DCSF <- st_read("Census_Tract_Combine_Final.shp")
```

```
## Reading layer `Census_Tract_Combine_Final' from data source `C:\Users\Robert\Desktop\GIS\Fina
l_Part3\GIS_3\Census_Tract_Combine_Final.shp' using driver `ESRI Shapefile'
## Simple feature collection with 179 features and 27 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: -77.11976 ymin: 38.79165 xmax: -76.9094 ymax: 38.99581
## epsg (SRID):    4326
## proj4string:    +proj=longlat +datum=WGS84 +no_defs
```

```
# Remove the census tracts that will be left out of analysis (tracts 6202 and 6804)
DCSF <- DCSF[-c(73,79), ]

# Create scatter plot to view the data - Diversity score (independent variable) and Change in Me
dian Household Income 2010-2015 (dependent variable)
qplot(Diversity, Change_201, data = DCSF, geom = "point") + stat_smooth(method="lm", se=FALSE, s
ize=1)
```
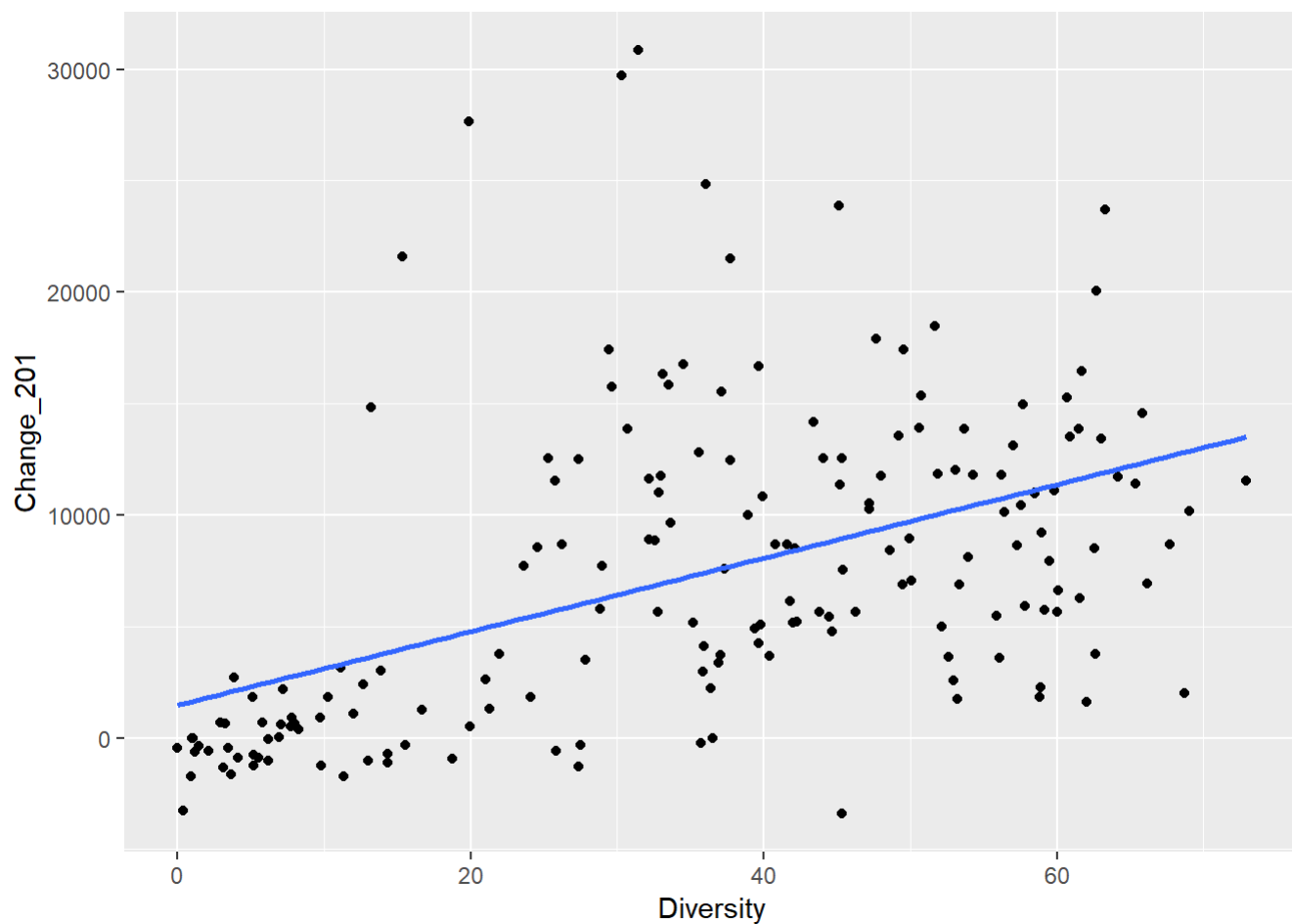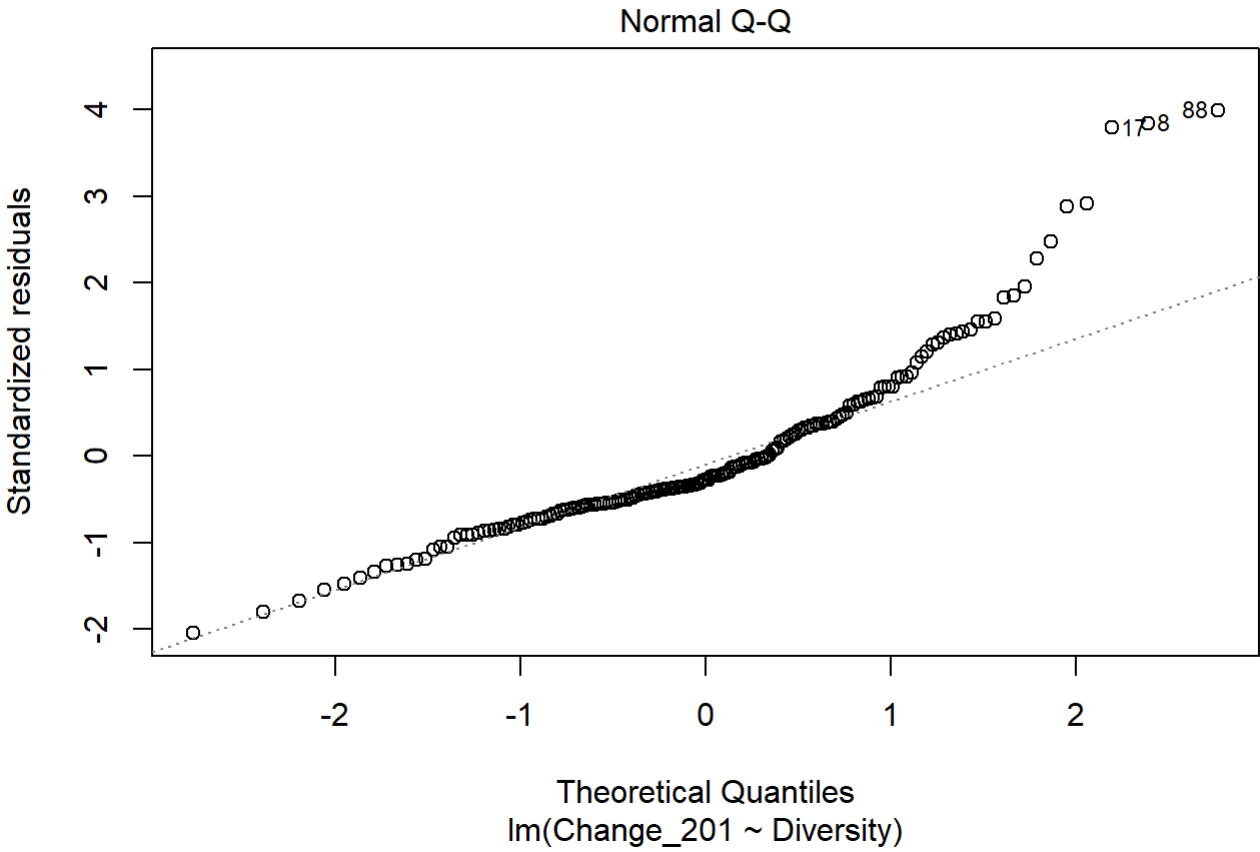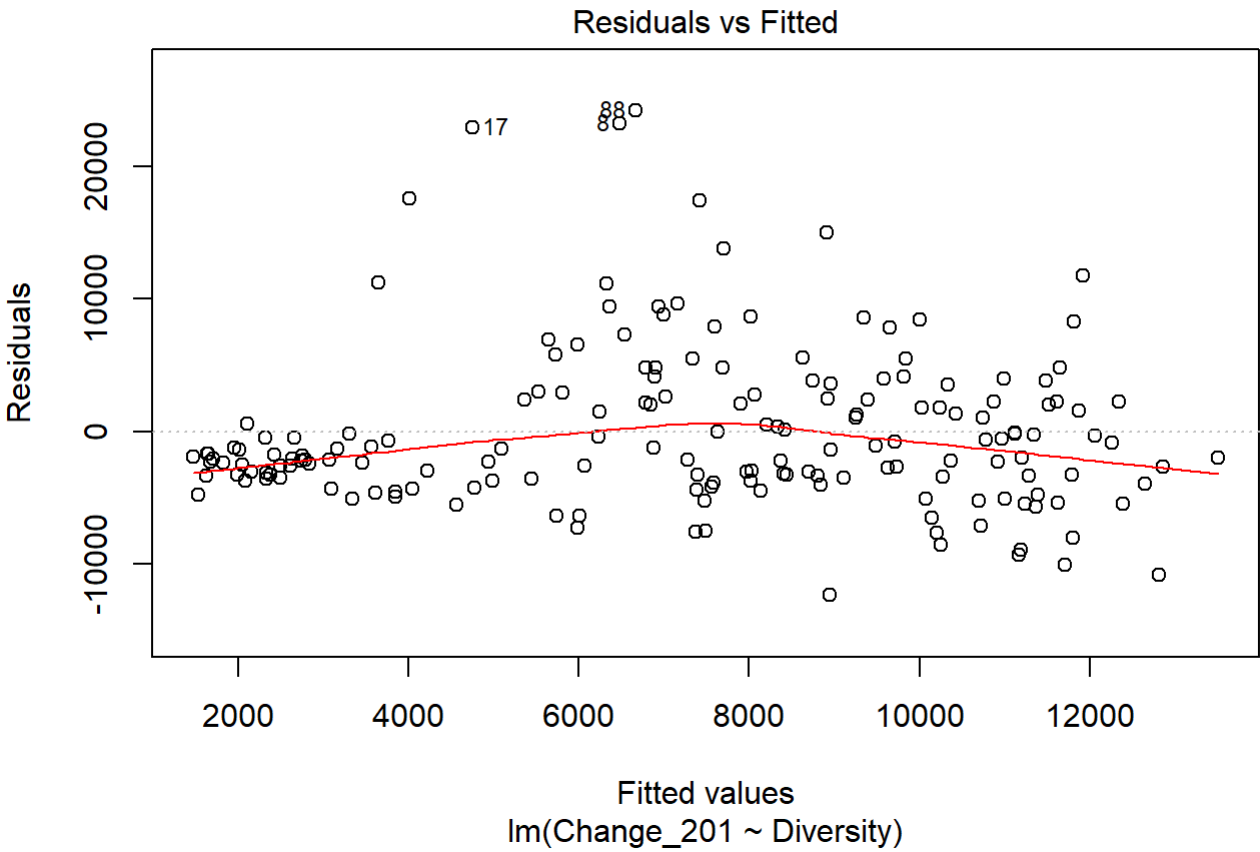
```
# Use function to fit to linear regrssion model
model1 <- lm(Change_201 ~ Diversity, data = DCSF)

# Write the regression results into a dataframe
model1_res <- tidy(model1)

# Examine the results
summary(model1)
```
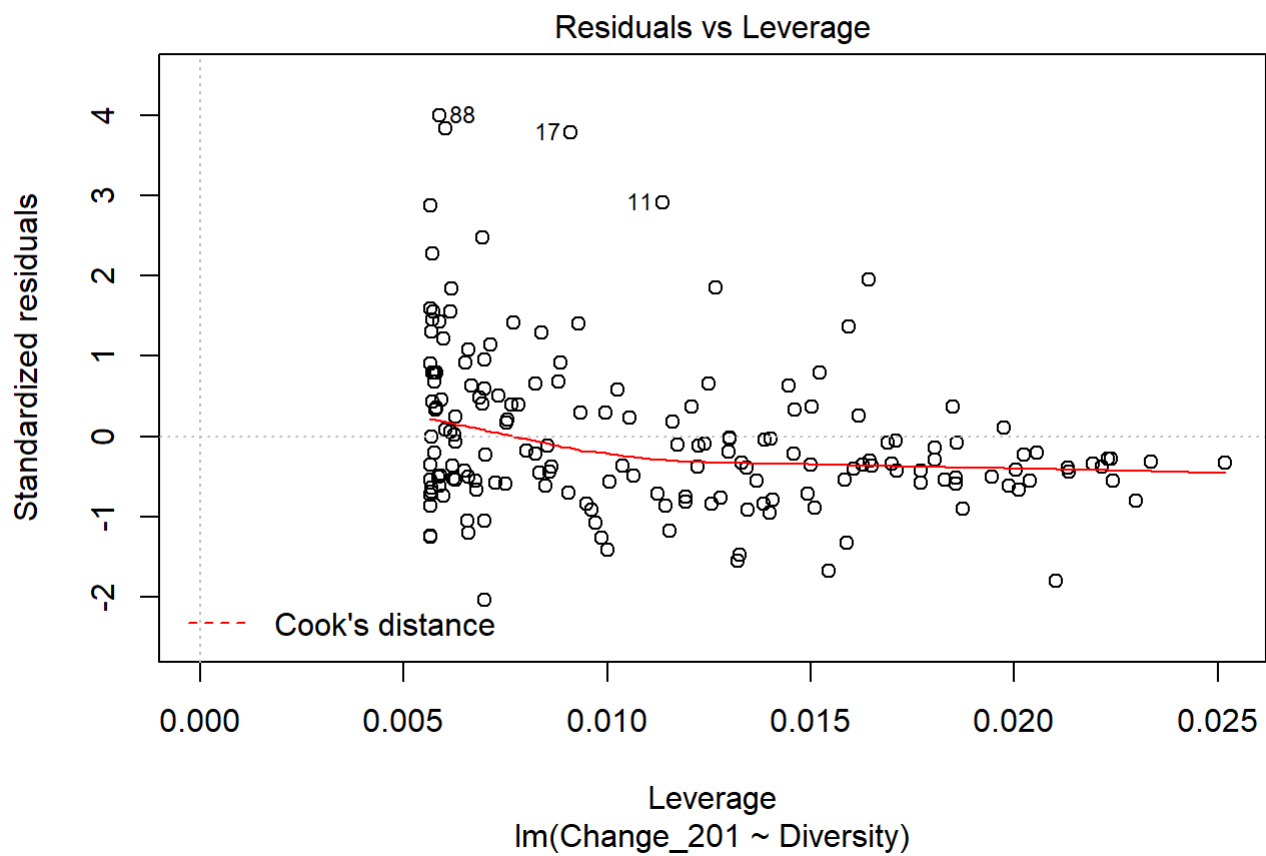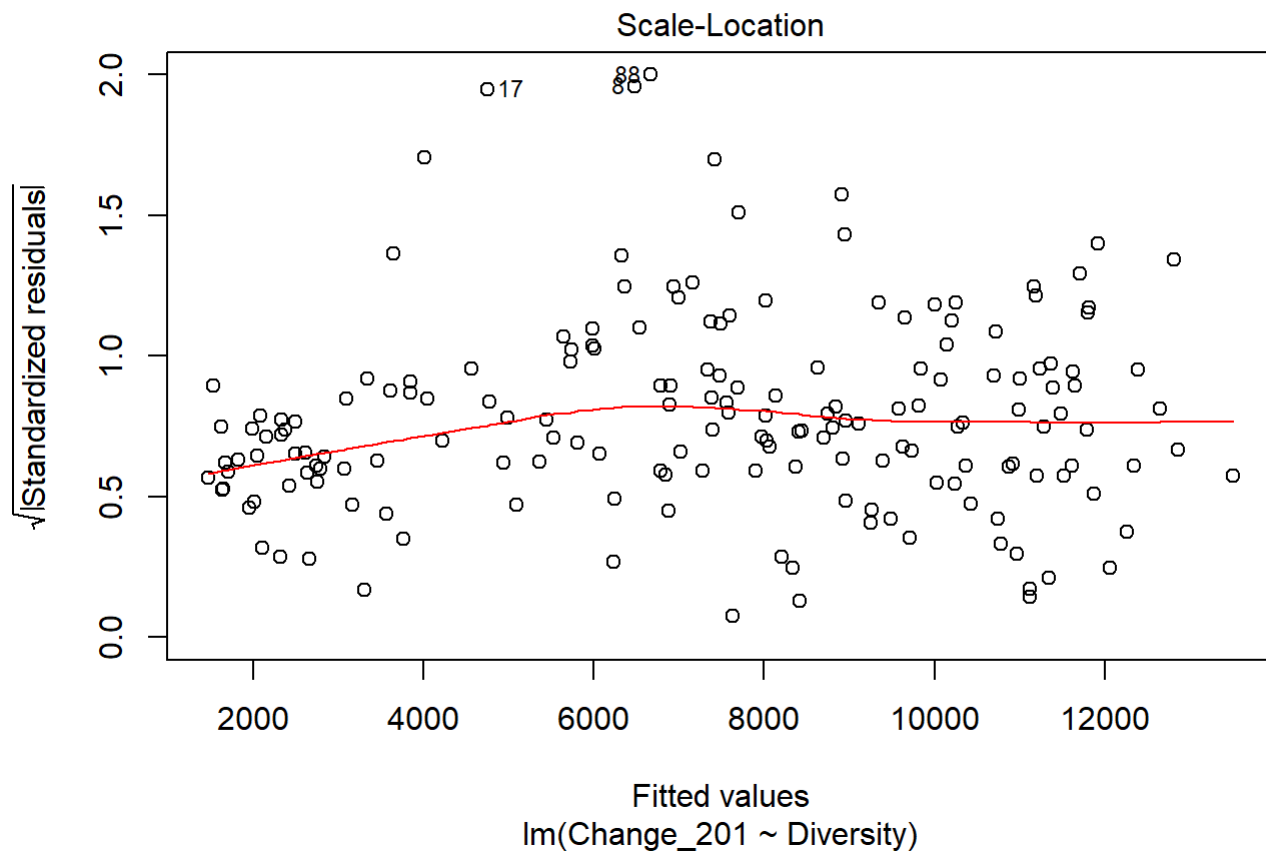
```
##
## Call:
## lm(formula = Change_201 ~ Diversity, data = DCSF)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -12368  -3522  -1666   2370  24194
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1465.90     927.68   1.580    0.116
## Diversity     165.12      22.71   7.272 1.13e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6071 on 175 degrees of freedom
## Multiple R-squared:  0.2321, Adjusted R-squared:  0.2277
## F-statistic: 52.89 on 1 and 175 DF,  p-value: 1.134e-11
```

```
# Examine the diagnostic plots to see if there is any patterning in the residuals
plot(model1)
```

Residuals vs Fitted

Fitted values
lm(Change_201 ~ Diversity)

Normal Q-Q

Theoretical Quantiles
lm(Change_201 ~ Diversity)

Scale-Location

√|Standardized residuals|

Fitted values
lm(Change_201 ~ Diversity)

Residuals vs Leverage

Standardized residuals

Cook's distance

Leverage
lm(Change_201 ~ Diversity)

```
# Save the residuals into to the dataframe
DCSF$model1_resids <- model1$residuals

# Plot the residuals
tmap_mode("view")
```
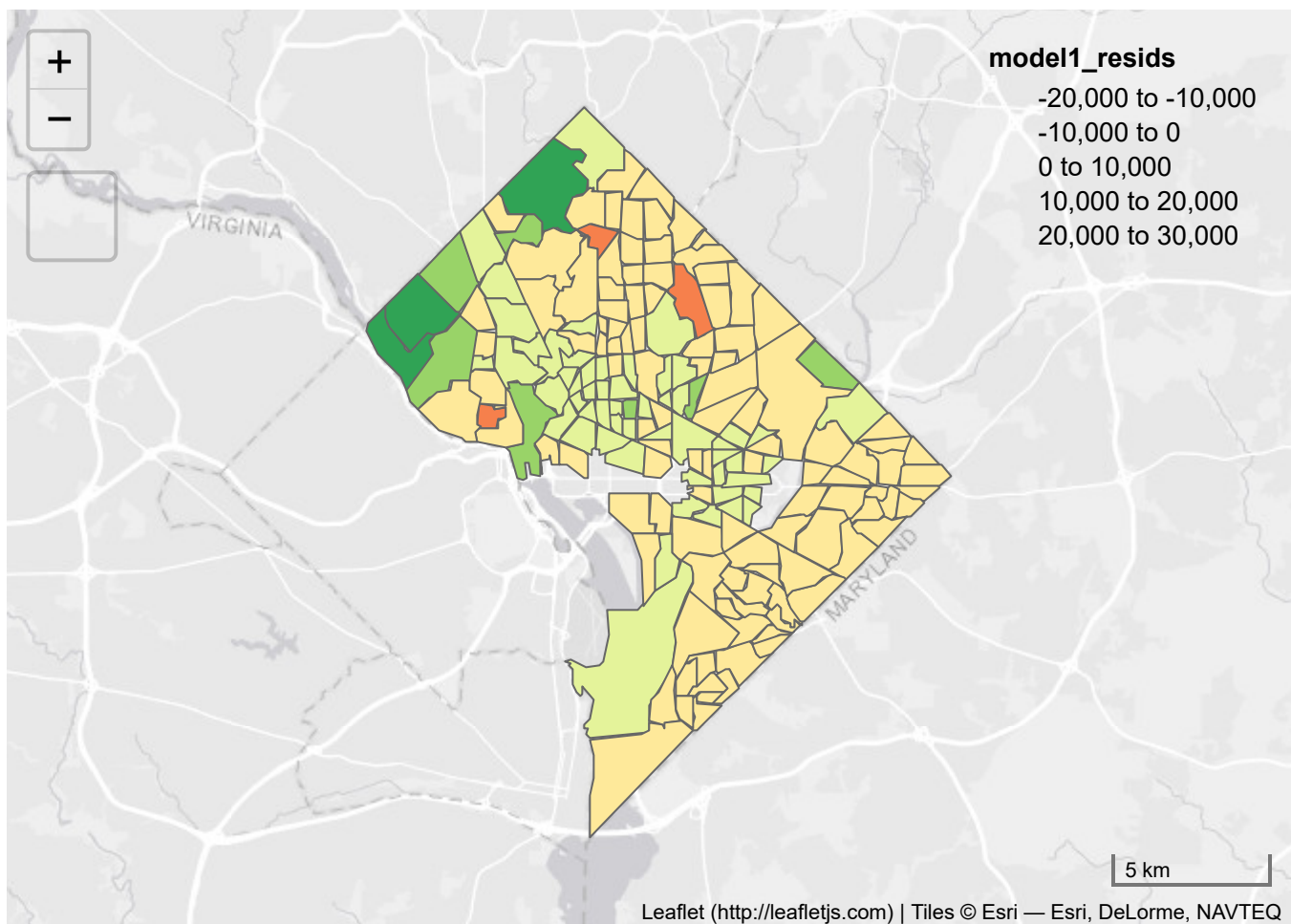
```
## tmap mode set to interactive viewing
```

```
qtm(DCSF, fill = "model1_resids")
```

```
## Variable "model1_resids" contains positive and negative values, so midpoint is set to 0. Set
midpoint = NA to show the full spectrum of the color palette.
```
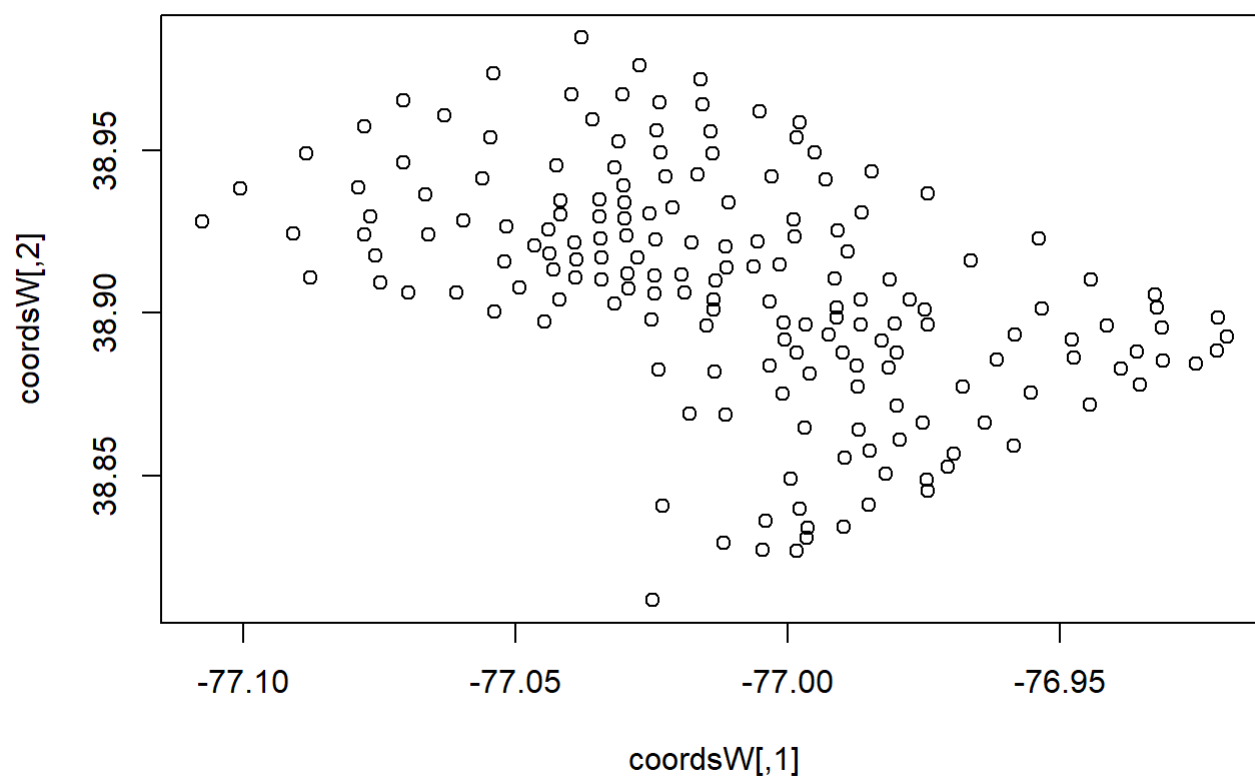


```
# Conduct a Moran's I test to see, statistically, if there is any autocorrelation of the residua
ls. First step is converting the dataframe to a spatial object.
DCSP <- as(DCSF,"Spatial")

# Next, calculate the centroids of all census tracts
coordsW <- coordinates(DCSP)
plot(coordsW)
```
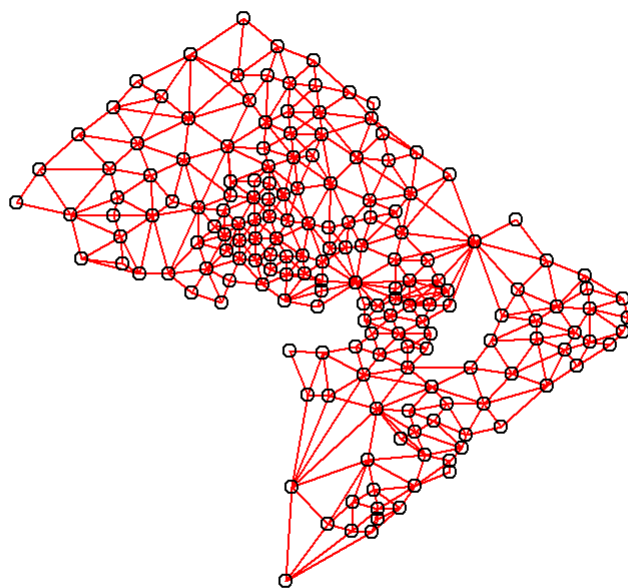
```
# Generate a spatial weights matrix of Queen's case neighbours
DC_nb <- poly2nb(DCSP, queen=T)

# Plot them to check accuracy
plot(DC_nb, coordinates(coordsW), col="red")
```

```
# Create a spatial weights object from these weights
DC.lw <- nb2listw(DC_nb, style="C")

# Run Moran's I test on the residuals
moran.test(DCSP@data$model1_resids, DC.lw)
```

```
##
##   Moran I test under randomisation
##
## data:  DCSP@data$model1_resids
## weights: DC.lw
##
## Moran I statistic standard deviate = 7.403, p-value = 6.655e-14
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      0.311490876       -0.005681818       0.001835575
```