Socially Vulnerable Populations and Susceptibility to Damages

Spatial Autocorrelation

Combine Census Tract Shapefiles with FEMA Data

```
# Read census spatial data as SpatialPolygonsDataFrame
pr_tracts <- readOGR(dsn='../data/census-tract/shapefiles',
                     layer = 'cb_2017_72_tract_500k')
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\rossm\Documents\MIDS\W210\main\data\census-tract\shapefiles", layer: "cb_
## with 908 features
## It has 9 fields
## Integer64 fields read as strings:  ALAND AWATER
```

```
# Convert the GEOID to character
pr_tracts@data$GEOID <- as.character(pr_tracts@data$GEOID)
```

```
# Read FEMA data
d_fema <- fread('../data/open-fema/FEMA-Large-Tract-Demographics-WindSpeed-PR.csv',
                encoding = 'UTF-8')

# Convert variables names to lower case
names(d_fema) <- tolower(names(d_fema))

# Select required fields
d_fema <- select(d_fema, censustractid, county, below_poverty_rate, bachelors_degree_rate,
                 unemployed_labor_rate, owner_occupied_rate, built_1979_or_earlier_rate,
                 waterlevel, pwg_mph, pwg_saffir_simpson, damageamount_cap)

# Convert tractid to characrter
d_fema$censustractid <- as.character(d_fema$censustractid)
```

```
# Join the SpatialPolygonsDataFrame with the FEMA dataframe
pr_tracts_demo <- geo_join(pr_tracts, d_fema, by_sp='GEOID',
                           by_df='censustractid', how='inner')
nrow(pr_tracts_demo)
```

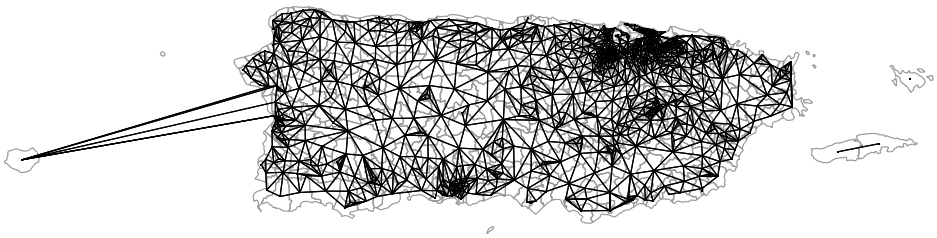
```
## [1] 886
```

```r
# Create neighbors
nbq <- poly2nb(pr_tracts_demo)
nblist <- nb2listw(nbq, zero.policy = T, style='W')
summary(nbq)
```

```
## Neighbour list object:
## Number of regions: 886
## Number of nonzero links: 5136
## Percentage nonzero weights: 0.6542708
## Average number of links: 5.79684
## 1 region with no links:
## 201
## Link number distribution:
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  15
##   1   4  20  74 129 189 162 142  94  41  18   5   6   1
## 4 least connected regions:
## 44 387 565 738 with 1 link
## 1 most connected region:
## 844 with 15 links
```

Visualize Census Tract Connections

```r
# Plot using the plot function
plot(pr_tracts_demo, border='grey60')
plot(nbq, coordinates(pr_tracts_demo), add=TRUE, pch='.')
```



Check Moran's I

```r
# Below poverty rate
moran.test(pr_tracts_demo$below_poverty_rate, nblist,
           zero.policy = T, randomisation = FALSE, na.action = na.omit)
```

```
##
##  Moran I test under normality
##
## data:  pr_tracts_demo$below_poverty_rate
## weights: nblist
## omitted: 325, 828 n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = 18.358, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      0.3748643701     -0.0011337868      0.0004194792
```

```r
# Owner occupied rate
moran.test(pr_tracts_demo$owner_occupied_rate, nblist,
           zero.policy = T, randomisation = FALSE, na.action = na.omit)
```

```
##
##  Moran I test under normality
##
## data:  pr_tracts_demo$owner_occupied_rate
## weights: nblist
## omitted: 325, 828 n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = 18.75, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      0.3828815366     -0.0011337868      0.0004194792
```

```r
# Peak wind gust (saffir-simpson)
moran.test(pr_tracts_demo$pwg_saffir_simpson, nblist,
           zero.policy = T, randomisation = FALSE, na.action = na.omit)
```

```
##
##  Moran I test under normality
##
## data:  pr_tracts_demo$pwg_saffir_simpson
## weights: nblist  n reduced by no-neighbour observations
```

```
##
##
## Moran I statistic standard deviate = 47.546, p-value < 2.2e-16
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##      0.9713245747     -0.0011312217      0.0004183317
```

Linear Regression and Spatial Dependencies

```r
# OLS regression
lm_model <- lm(formula=log1p(damageamount_cap) ~ pwg_saffir_simpson + log1p(waterlevel)
               + below_poverty_rate + bachelors_degree_rate + unemployed_labor_rate
               + built_1979_or_earlier_rate + owner_occupied_rate, data = pr_tracts_demo)

summary(lm_model)
```

```
##
## Call:
## lm(formula = log1p(damageamount_cap) ~ pwg_saffir_simpson + log1p(waterlevel) +
##     below_poverty_rate + bachelors_degree_rate + unemployed_labor_rate +
##     built_1979_or_earlier_rate + owner_occupied_rate, data = pr_tracts_demo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0422 -0.7892  0.1041  0.9325  4.0922
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  3.33347    0.75109   4.438 1.02e-05 ***
## pwg_saffir_simpson           0.15168    0.04848   3.129  0.00181 **
## log1p(waterlevel)            0.82970    0.11387   7.286 7.11e-13 ***
## below_poverty_rate           1.76462    0.62637   2.817  0.00495 **
## bachelors_degree_rate        0.51728    0.98829   0.523  0.60082
## unemployed_labor_rate       -1.76356    1.16039  -1.520  0.12892
## built_1979_or_earlier_rate   0.27926    0.26899   1.038  0.29947
## owner_occupied_rate          0.63751    0.45423   1.404  0.16082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.437 on 876 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.0775, Adjusted R-squared:  0.07013
## F-statistic: 10.51 on 7 and 876 DF,  p-value: 9.886e-13
```

```r
# AIC
lm_model$AIC <- AIC(lm_model)
paste('AIC:', round(lm_model$AIC,1))
```

```
## [1] "AIC: 3159.6"
```

```r
# Normality of residuals
shapiro.test(lm_model$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  lm_model$residuals
## W = 0.96477, p-value = 8.933e-14
```

```r
# Moran test for spatial correlation in residuals
(moran_model <- lm.morantest(lm_model, listw = nblist,
                             zero.policy = T, alternative = 'two.sided'))
```

```
##
##  Global Moran I for regression residuals
##
## data:
## model: lm(formula = log1p(damageamount_cap) ~ pwg_saffir_simpson +
## log1p(waterlevel) + below_poverty_rate + bachelors_degree_rate +
## unemployed_labor_rate + built_1979_or_earlier_rate +
## owner_occupied_rate, data = pr_tracts_demo)
## weights: nblist
##
## Moran I statistic standard deviate = 6.3022, p-value = 2.935e-10
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation         Variance
##     0.1235112823    -0.0045123367     0.0004126652
```

Visualizing OLS Residuals

```r
# Census tracts with NAs
remove_tracts <- d_fema[rowSums(is.na(d_fema)) > 0, ]
paste('Census tracts removed:', nrow(remove_tracts))
```

```
## [1] "Census tracts removed: 2"
```

```r
# Remove NAs from spatial
pr_tracts_demo_clean <- pr_tracts_demo[!(pr_tracts_demo@data$GEOID %in% c('72037160100', '7211
```

```r
# OLS - NA removed
lm_model_clean <- lm(formula=log1p(damageamount_cap) ~ pwg_saffir_simpson + log1p(waterlevel)
                     + below_poverty_rate + bachelors_degree_rate + unemployed_labor_rate
                     + built_1979_or_earlier_rate + owner_occupied_rate, data = pr_tracts_demo)

# Add residuals
pr_tracts_demo_clean@data$residuals <- residuals(lm_model_clean)
```

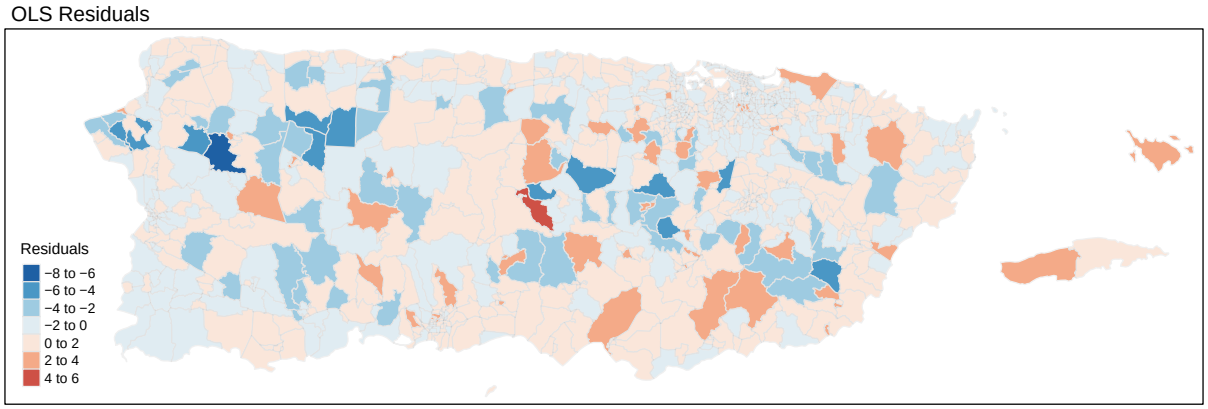
```r
# Adjust bounding box
bbox_new <- st_bbox(pr_tracts_demo_clean) # current bounding box

# Range of values
xrange <- bbox_new$xmax - bbox_new$xmin # range of x values
yrange <- bbox_new$ymax - bbox_new$ymin # range of y values

# Options for box adjustments
bbox_new[1] <- bbox_new[1] + (0.21 * xrange) # xmin - left
bbox_new[3] <- bbox_new[3] - (0.01 * xrange) # xmax - right

# Make bounding box a sf polygon
bbox_new <- bbox_new %>%
  st_as_sfc()
```

```r
# Plot residuals
tm_shape(pr_tracts_demo_clean, bbox = bbox_new) +
  tm_fill('residuals', title = 'Residuals', palette = '-RdBu', midpoint = 0) +
  tm_borders(alpha = 0.1) +
  tm_layout(main.title = 'OLS Residuals', main.title.size = 1,
            legend.position = c('left', 'bottom'), legend.title.size = 0.9)
```



Spatial Tests - LaGrange Multiplier Tests

```
# Run LaGrange multiplier tests
lmtests <- lm.LMtests(lm_model, nblist, zero.policy = T,
                      test = c('LMerr','LMlag','RLMerr','RLMlag','SARMA'))
summary(lmtests)
```

```
##  Lagrange multiplier diagnostics for spatial dependence
## data:
## model: lm(formula = log1p(damageamount_cap) ~ pwg_saffir_simpson +
## log1p(waterlevel) + below_poverty_rate + bachelors_degree_rate +
## unemployed_labor_rate + built_1979_or_earlier_rate +
## owner_occupied_rate, data = pr_tracts_demo)
## weights: nblist
##
##         statistic parameter   p.value
## LMerr  36.2124450         1 1.769e-09 ***
## LMlag  33.2013597         1 8.309e-09 ***
## RLMerr  3.0183146         1   0.08233 .
## RLMlag  0.0072294         1   0.93224
## SARMA  36.2196744         2 1.365e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Both LMerr and LMlag are significant and so must check robust versions
# RLMerr is not significant (p=0.082), indicating spatial error should be used
```

Spatial Regression

```
# Spatial regression - error model
sem_model <- errorsarlm(formula = log1p(damageamount_cap) ~ pwg_saffir_simpson + log1p(waterlev
                        + below_poverty_rate + bachelors_degree_rate + unemployed_labor_rate
                        + built_1979_or_earlier_rate + owner_occupied_rate,
                        data = pr_tracts_demo, listw = nblist, zero.policy = T)

summary(sem_model)
```

```
##
## Call:errorsarlm(formula = log1p(damageamount_cap) ~ pwg_saffir_simpson +
##     log1p(waterlevel) + below_poverty_rate + bachelors_degree_rate +
##     unemployed_labor_rate + built_1979_or_earlier_rate + owner_occupied_rate,
##     data = pr_tracts_demo, listw = nblist, zero.policy = T)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -5.705704 -0.793370  0.092459  0.893803  4.357808
##
## Type: error
## Regions with no neighbours included:
##  201
## Coefficients: (asymptotic standard errors)
##                              Estimate Std. Error z value  Pr(>|z|)
## (Intercept)                  3.463077   0.764696  4.5287 5.935e-06
## pwg_saffir_simpson           0.121148   0.062895  1.9262   0.05408
## log1p(waterlevel)            0.799730   0.123656  6.4674 9.971e-11
## below_poverty_rate           1.513384   0.640014  2.3646   0.01805
## bachelors_degree_rate        0.737640   0.992470  0.7432   0.45734
## unemployed_labor_rate       -1.496280   1.237912 -1.2087   0.22677
## built_1979_or_earlier_rate   0.469329   0.281660  1.6663   0.09565
## owner_occupied_rate          0.526739   0.474949  1.1090   0.26741
##
## Lambda: 0.30884, LR test value: 35.15, p-value: 3.0531e-09
## Asymptotic standard error: 0.047217
##     z-value: 6.541, p-value: 6.1122e-11
## Wald statistic: 42.784, p-value: 6.1122e-11
##
## Log likelihood: -1553.228 for error model
## ML residual variance (sigma squared): 1.9309, (sigma: 1.3896)
## Number of observations: 884
## Number of parameters estimated: 10
## AIC: 3126.5, (AIC for lm: 3159.6)
```

```
Hausman.test(sem_model)
```

```
##
##  Spatial Hausman test (asymptotic)
##
## data:  NULL
## Hausman test = 14.388, df = 8, p-value = 0.07219
```

Table 1: Social Vulnerability and Susceptibility to Damages

	Dependent variable:	
	Damage Amount per Capita (Log)	
	OLS	Spatial Error
	(1)	(2)
Peak Wind Gusts	0.152***	0.121*
	(0.048)	(0.063)
Water Level (log)	0.830***	0.800***
	(0.114)	(0.124)
Below Poverty Rate	1.765***	1.513**
	(0.626)	(0.640)
Bachelor Degree Rate	0.517	0.738
	(0.988)	(0.992)
Unemployed Labor Rate	−1.764	−1.496
	(1.160)	(1.238)
Homes built 1979 or Earlier Rate	0.279	0.469*
	(0.269)	(0.282)
Owner Occupied Rate	0.638	0.527
	(0.454)	(0.475)
Constant	3.333***	3.463***
	(0.751)	(0.765)
Observations	884	884
Akaike Inf. Crit.	3,159.605	3,126.455
Note:		*p<0.1; **p<0.05; ***p<0.01