Fire & Flood Risks in EJ Communities

David Amanfu, Chayan Jani, Kelsie Roberton

April 2022

```
# Set your working directory
getwd()
```

```
## [1] "/Users/davidamanfu/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_ENV872_E
```

```
#knitr::opts_knit$set(root.dir = '/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_
knitr::opts_chunk$set(echo = FALSE)
getwd()
```

```
## [1] "/Users/davidamanfu/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_ENV872_E
```

Contents

Filtering Out Missing Data 6
Simple Correlative Graphs 6
    Coastal Flooding Tests 14

List of Tables

List of Figures

##Initializing Setup

#Data Setup ##Data Retrieval

For our analysis, we used data collected from the Environmental Protection Agency (EPA) EJ Screen to look at pre-1960 housing and low-income housing populations. For reference, EJScreen is an environmental justice screening and mapping tool that provides the EPA with datasets on a national-scale that combines environmental and geographic indicators (epa.gov). All EJScreen indicators are publicly available data on the EPA website. For our analysis, data was downloaded as an Excel file here. Our analysis used 2020 data from EJScreen as opposed to 2021 data in order to avoid any data inconsistencies that may have not been resolved in the 2021 data. As well, the Census Bureau's data we utilized was from 2020, and for consistency sake we continued with 2020 data. We are not aware of any significant changes between the two otherwise. To nicely overlay the information from the census tracts onto a geometry, we included matching census tract information from the US Census Bureau's Gazetteer Files database. The Gazetteer files, downloadable here, include geographic identifying information including tract tags, county names, and importantly, representative latitude and longitude coordinates. To gather data for wildfire and flood risk, we utilized FEMA's National Risk Index (NRI) data. The NRI data contains information regarding a geographic boundary's exposure to and risk implications of 18 natural hazards. The data is made available at both the county and census tract level. We used the csv for all census tracts, downloadable here, due to the large size of the shapefile. We combined this data with US Census shapefiles at the census-tract level, with Cartographic Boundary Files, downloadable [here]. ## Load Datasets

#Data Wrangling We started our analysis by selecting the following columns from the raw (name of data file) GEOID, NAMESLAD, NAMESLADCO, ALAND, AWATER, geometry for each state (NC, FL and CA). Next, we filtered for each state that has to do with fire and floods. Following this, EJScreen data came in census block groups. We used the 'group by' summarize function to aggregate into census tracts. Following this, we can match these census tracts by the NRI risk data files (by state).

Because EJ Screen data does not come with geometric coordinates, we had to import and merge the census gazetteer files to at least give them a corresponding point within the census. Following this, we attempted to merge both of the filtered files for each state's fire and flood data. Finally, we were able to create three datasets per shape: fire-risks, flood-risks, and EJScreen demographic. We then proceeded with the exploratory analysis, which can be shown below

Given the extensive nature of all of the datasets, we decided to remove California from our analyses and focus on North Carolina and Florida. ##Filter Steps ### Initial Filter By State

###Filter NRI Dataset For Risks of Interest

##Adding Location Data ###Shape Files from Census Data Add in Shape Files https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html

```
## Reading layer 'cb_2020_06_tract_500k' from data source
##   '/Users/davidamanfu/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_ENV872_EDA
##   using driver 'ESRI Shapefile'
## Simple feature collection with 9109 features and 13 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -124.4096 ymin: 32.53444 xmax: -114.1312 ymax: 42.00948
## Geodetic CRS:  NAD83

## Reading layer 'cb_2020_37_tract_500k' from data source
##   '/Users/davidamanfu/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_ENV872_EDA
##   using driver 'ESRI Shapefile'
## Simple feature collection with 2660 features and 13 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
```

```
## Bounding box:  xmin: -84.32187 ymin: 33.84232 xmax: -75.46062 ymax: 36.58812
## Geodetic CRS:  NAD83

## Reading layer 'cb_2020_12_tract_500k' from data source
##   '/Users/davidamanfu/Desktop/Duke MPP/Environ Data /872 Final Project/AmanfuJaniRoberton_ENV872_EDA
##   using driver 'ESRI Shapefile'
## Simple feature collection with 5122 features and 13 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -87.63494 ymin: 24.5231 xmax: -80.03136 ymax: 31.00089
## Geodetic CRS:  NAD83
```

Using Point Data for Longitude/Latitude Coordinates from US Census Gazetteer Files https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2020.html

###Filtering With Shape Geometry

###Merging NRI Data with Tract Shapes, EJScreen with Tract Coordinates

#Data Exploration ##Simple Mapping

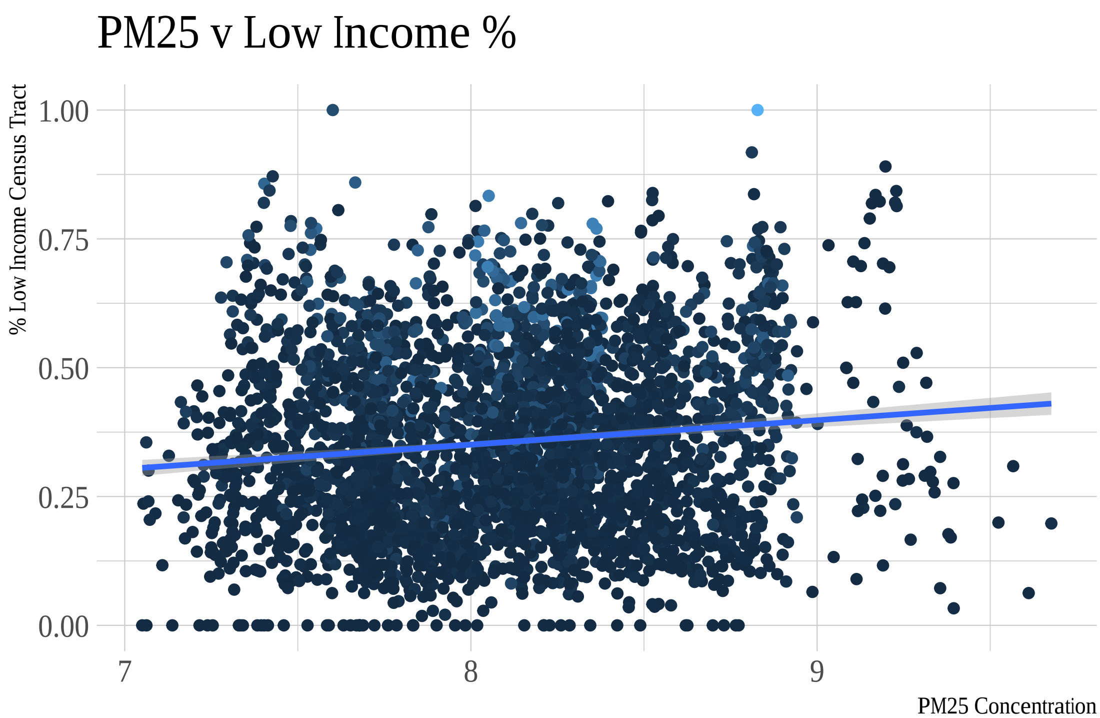
```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, pleas
```

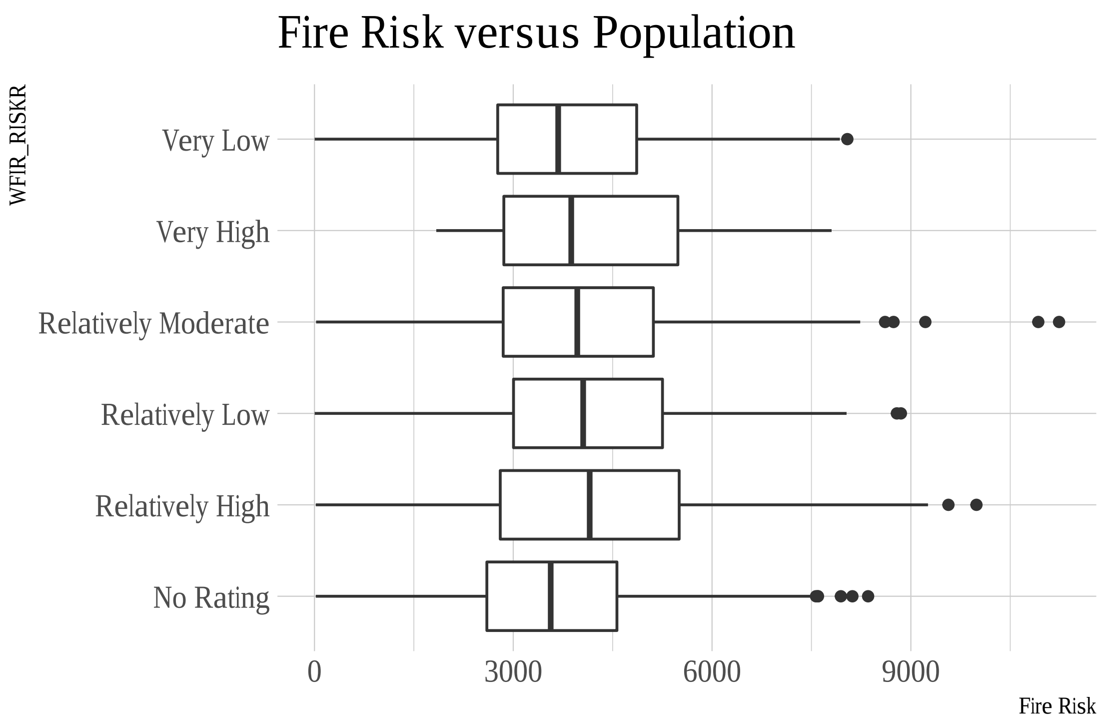
Filtering Out Missing Data

#Data Analysis

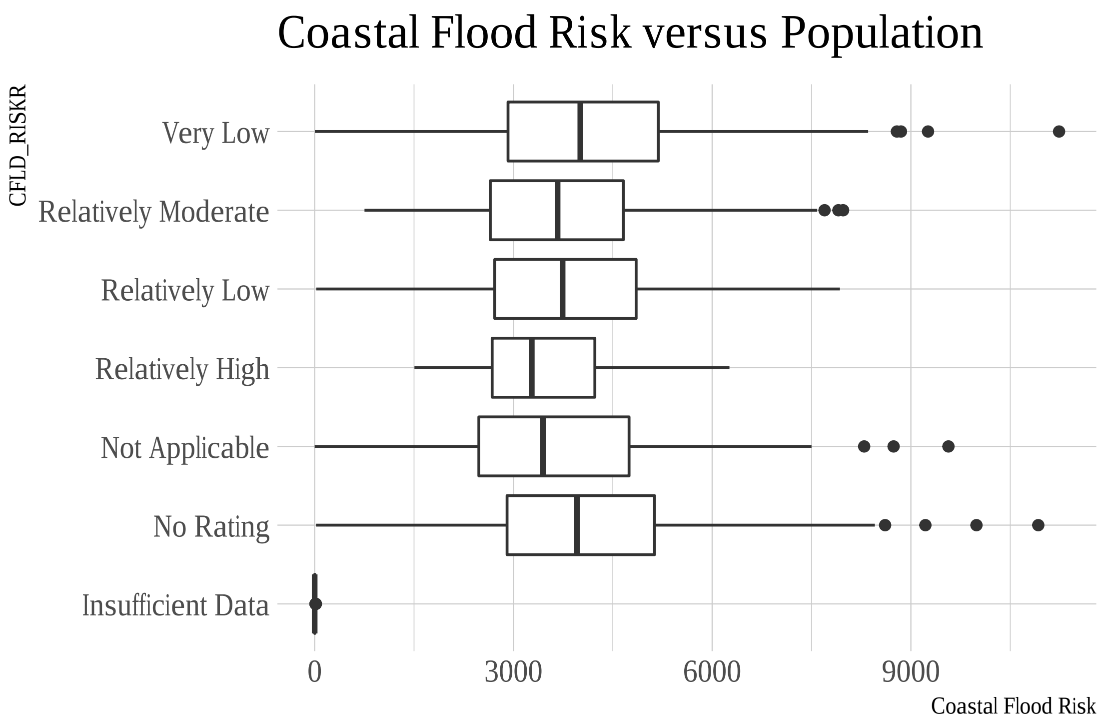
Simple Correlative Graphs

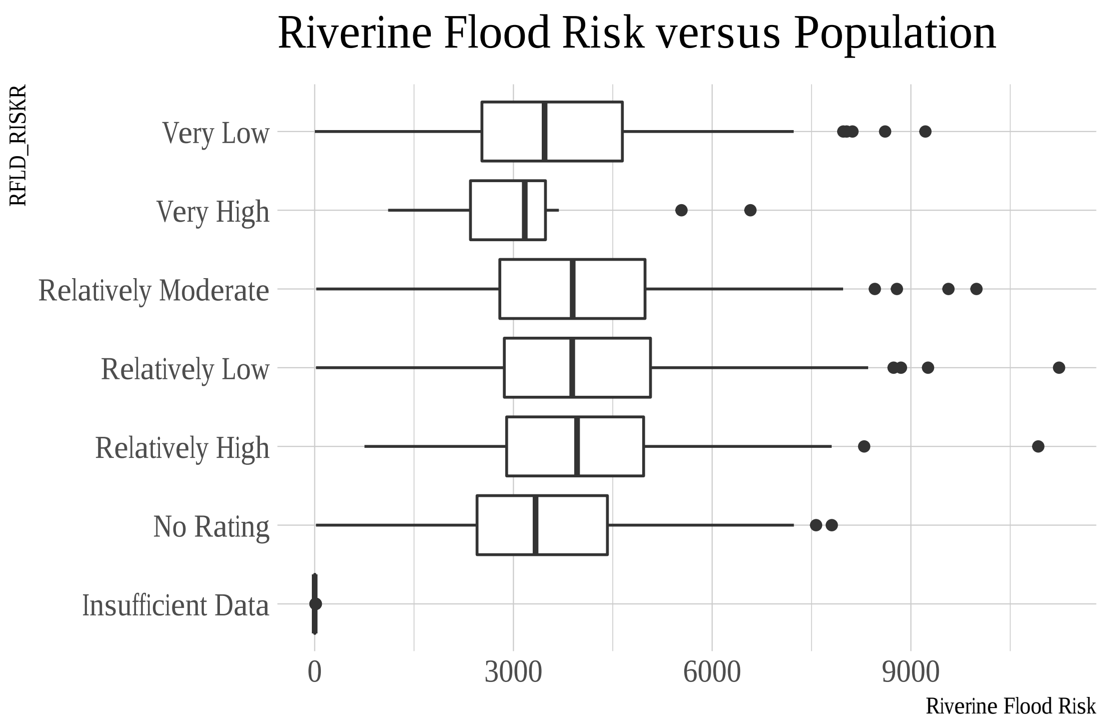
```
## 'geom_smooth()' using formula 'y ~ x'

## Warning: Removed 1859 rows containing non-finite values (stat_smooth).

## Warning: Removed 1859 rows containing missing values (geom_point).
```


PM25 v Low Income %
% Low Income Census Tract
1.00
0.75
0.50
0.25
0.00
7
8
9
PM25 Concentration

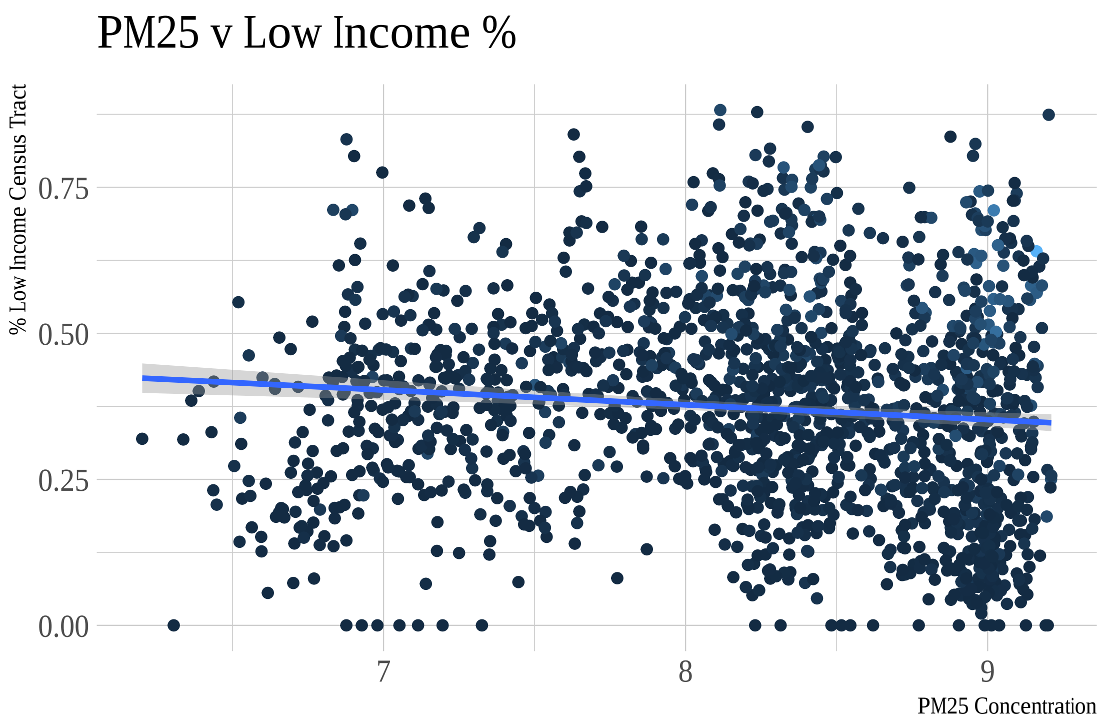


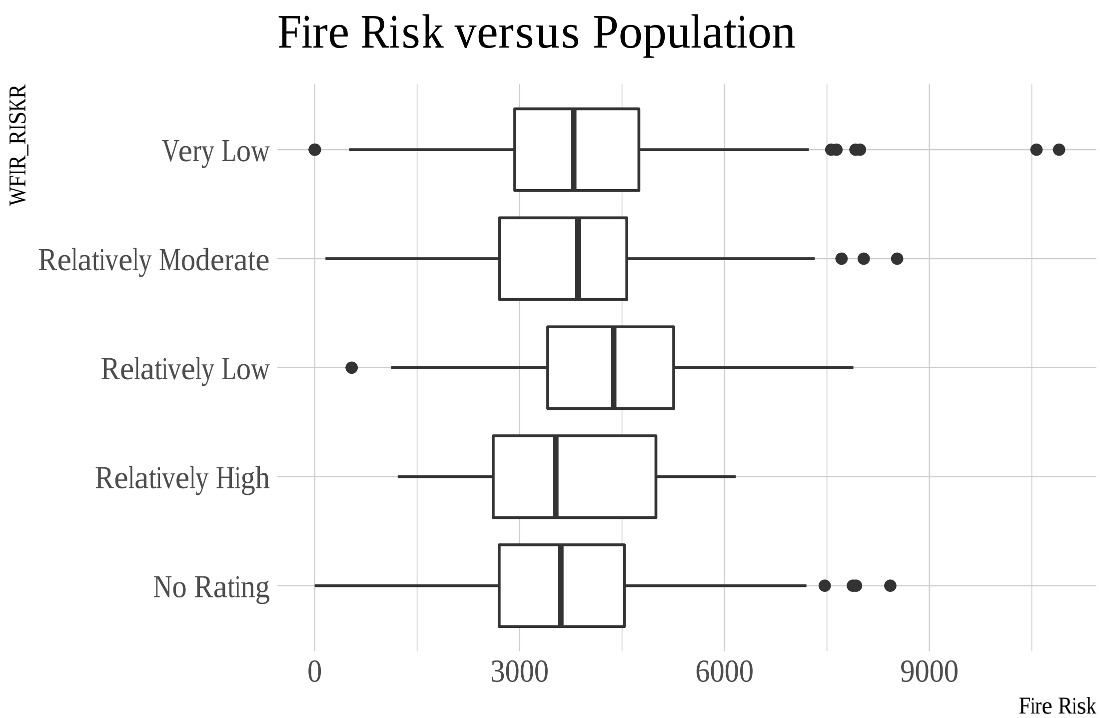
Fire Risk versus Population
WFIR_RISKR
Very Low
Very High
Relatively Moderate
Relatively Low
Relatively High
No Rating
0
3000
6000
9000
Fire Risk

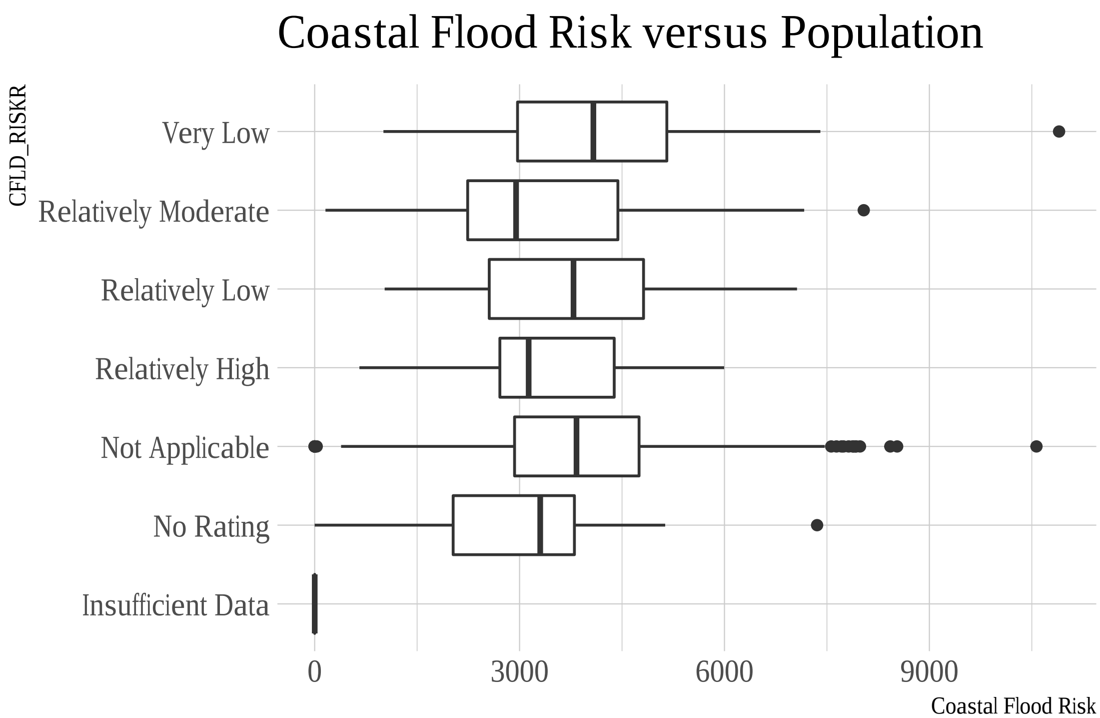

Coastal Flood Risk versus Population
CFLD_RISKR
Very Low
Relatively Moderate
Relatively Low
Relatively High
Not Applicable
No Rating
Insufficient Data
0
3000
6000
9000
Coastal Flood Risk

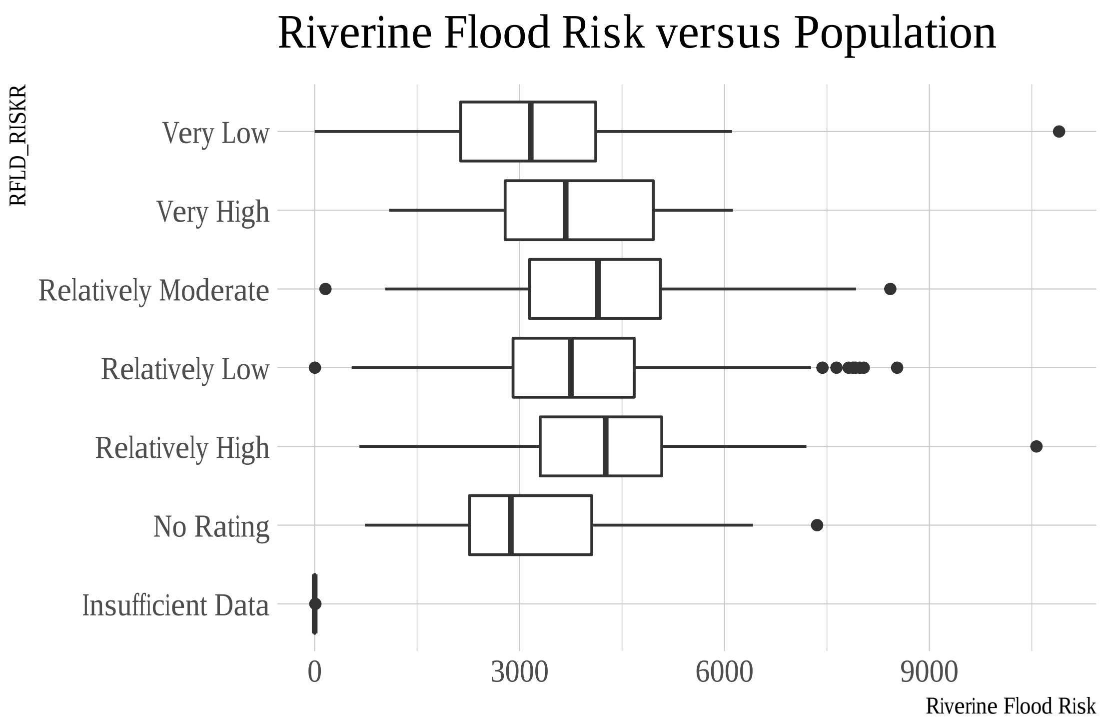



```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 953 rows containing non-finite values (stat_smooth).

## Warning: Removed 953 rows containing missing values (geom_point).
```


PM25 v Low Income %
% Low Income Census Tract
0.75
0.50
0.25
0.00
7
8
9
PM25 Concentration


Fire Risk versus Population
WFIR_RISKR
Very Low
Relatively Moderate
Relatively Low
Relatively High
No Rating
0
3000
6000
9000
Fire Risk


Coastal Flood Risk versus Population
CFLD_RISKR
Very Low
Relatively Moderate
Relatively Low
Relatively High
Not Applicable
No Rating
Insufficient Data
0
3000
6000
9000
Coastal Flood Risk



##Visualized Mapped Graphics

##Statistical Analyses ### Merged Datasets

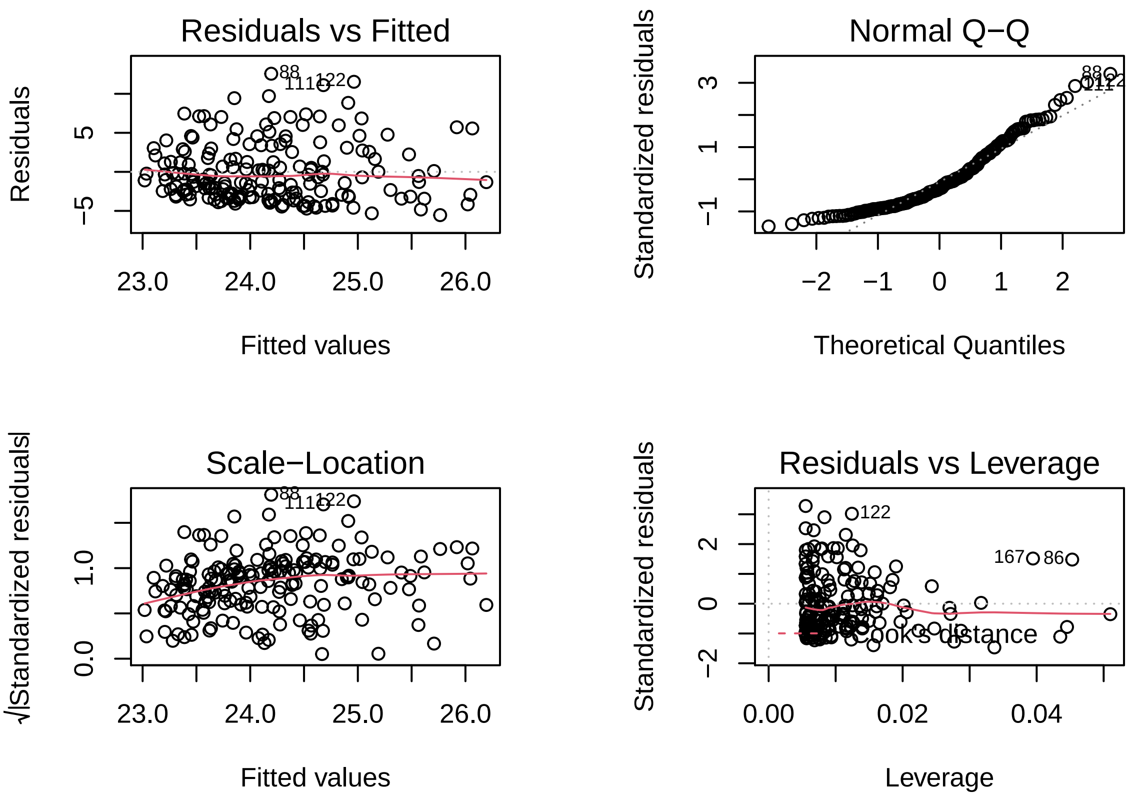
Coastal Flooding Tests

The p-value of the test is 0.01496, which is less than the significance level alpha = 0.05. We can conclude that low income population and coastal flood risks are significantly correlated with a correlation coefficient of 0.1806238 and p-value of 0.01496.It has a positive correlation which means flood risk is impacting a significant %of low income population.

```
##
## Call:
## lm(formula = LI_CF_risk$CFLD_RISKS ~ LI_CF_risk$LOWINCPCT)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.543 -2.982 -1.099  2.348 12.572
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           22.7073     0.6638  34.209   <2e-16 ***
## LI_CF_risk$LOWINCPCT   4.7471     1.9321   2.457    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 3.846 on 179 degrees of freedom
## Multiple R-squared:  0.03262,    Adjusted R-squared:  0.02722
## F-statistic: 6.037 on 1 and 179 DF,  p-value: 0.01496
```

```
##
##  Pearson's product-moment correlation
##
## data:  LI_CF_risk$CFLD_RISKS and LI_CF_risk$LOWINCPCT
## t = 2.457, df = 179, p-value = 0.01496
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.03570679 0.31810110
## sample estimates:
##       cor
## 0.1806238
```

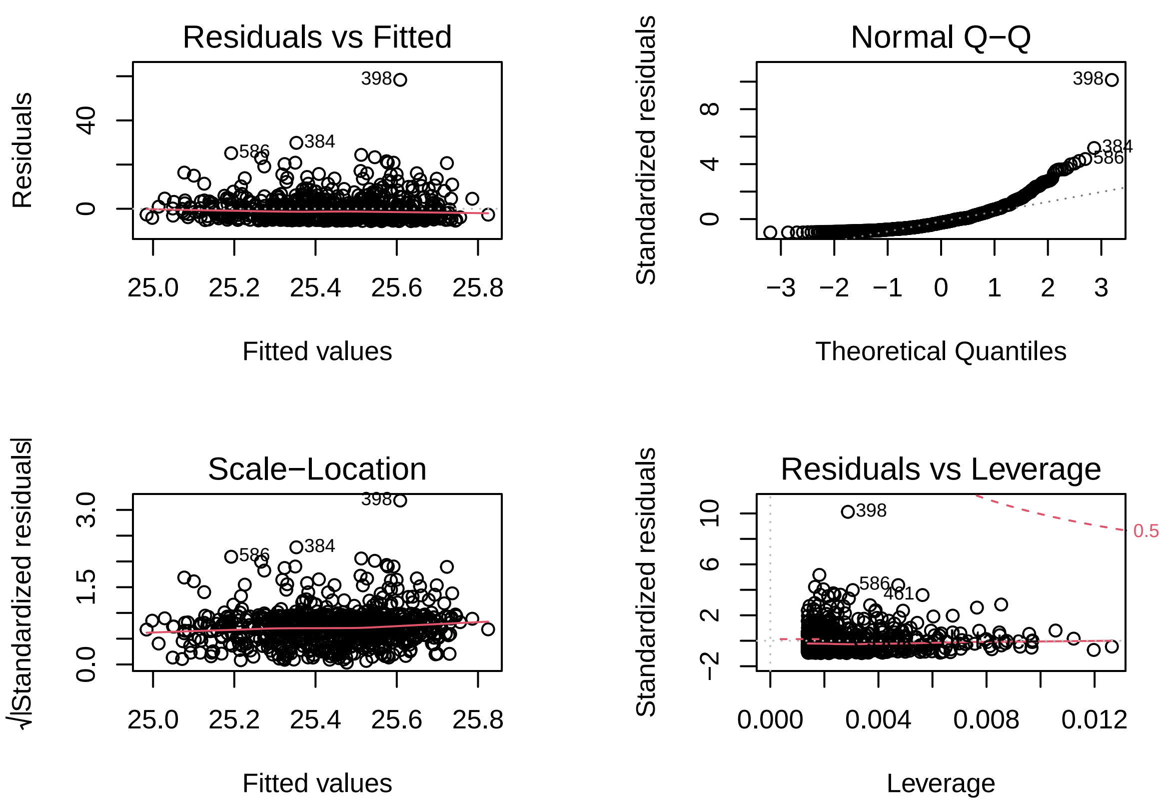


###Riverine Flooding Tests The p-value of the test is 0.4548, which is greater than the significance level alpha = 0.05. From this We can conclude that low income population and river flood risks are not significantly correlated and correlation coefficient is -0.02787792, negative implying inverse relationship.This might be due to the states selected and the overall chances of river flooding. This also means that the population at risk from river floods is not mostly low income.

```
##
## Call:
## lm(formula = LI_RF_risk$RFLD_RISKS ~ LI_RF_risk$LOWINCPCT)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.614 -3.837 -1.512  1.743 58.369
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           25.8249     0.5552  46.515   <2e-16 ***
## LI_RF_risk$LOWINCPCT  -0.9649     1.2903  -0.748    0.455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.774 on 719 degrees of freedom
## Multiple R-squared:  0.0007772,  Adjusted R-squared:  -0.0006126
## F-statistic: 0.5592 on 1 and 719 DF,  p-value: 0.4548
```

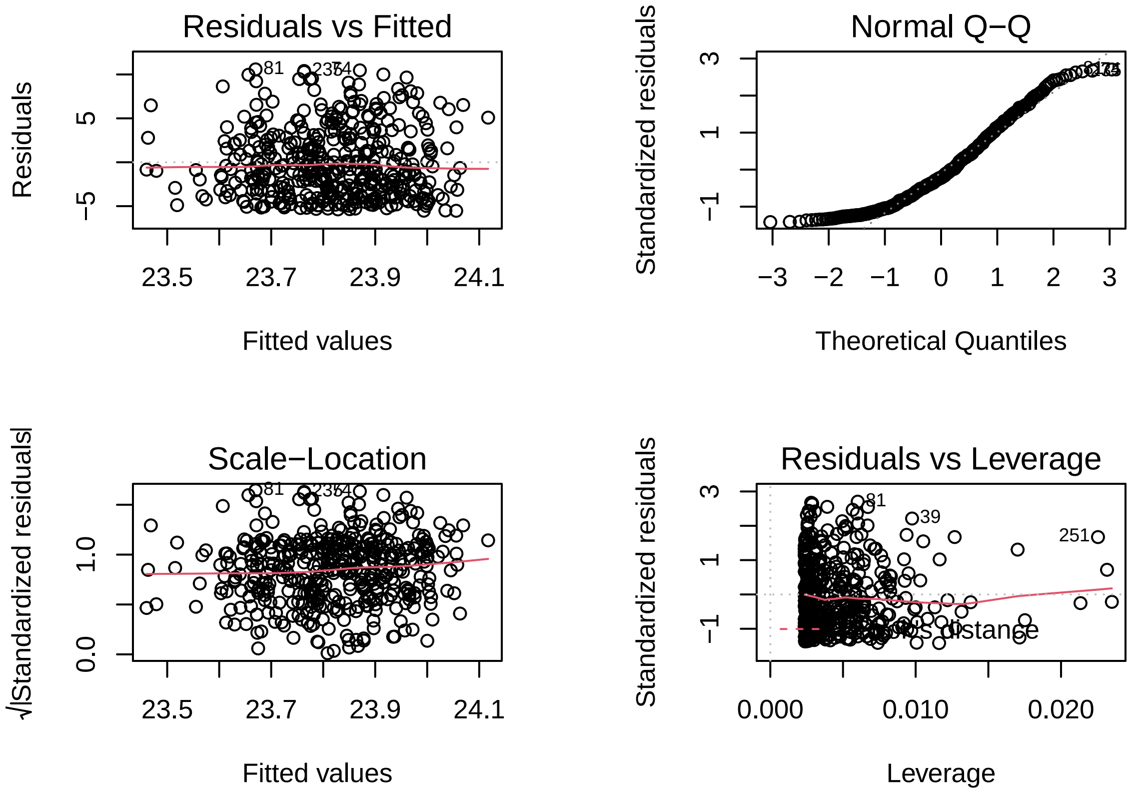
```
##
##  Pearson's product-moment correlation
##
## data:  LI_RF_risk$RFLD_RISKS and LI_RF_risk$LOWINCPCT
## t = -0.74781, df = 719, p-value = 0.4548
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.10068801  0.04522918
## sample estimates:
##         cor
## -0.02787792
```



###Wildfire Risk Tests The p-value of the test is 0.53, which is greater than the significance level alpha = 0.05. From this we can conclude that low income population and wildfire risks are not significantly correlated and correlation coefficient is -0.03065284 , negative implying inverse relationship.This might be due to the states selected and the overall chances of wildfire risk. This also means that the population at risk from wildfire is not mostly low income.

```
##
## Call:
## lm(formula = LI_WF_risk$WFIR_RISKS ~ LI_WF_risk$LOWINCPCT)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.526 -3.190 -0.857  2.554 10.575
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           24.1172     0.5118  47.126   <2e-16 ***
## LI_WF_risk$LOWINCPCT  -0.8640     1.3747  -0.628     0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.923 on 420 degrees of freedom
## Multiple R-squared:  0.0009396,  Adjusted R-squared:  -0.001439
## F-statistic: 0.395 on 1 and 420 DF,  p-value: 0.53
```

```
##
##  Pearson's product-moment correlation
##
## data:  LI_WF_risk$WFIR_RISKS and LI_WF_risk$LOWINCPCT
## t = -0.62849, df = 420, p-value = 0.53
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1257439  0.0649963
## sample estimates:
##         cor
## -0.03065284
```



#Conclusions The project aimed to look at the Environment Justice component from natural hazards. While there were several EJ communities, for our final report we focused on the % of the low income population. We looked at the National risk index for three natural hazard categories: coastal flooding, river flooding and wildfire. We looked at these parameters for two states: Florida and North Carolina

On analyzing the counties with high or relatively high natural hazard risk and comparing it with the % of low income population in that county we found that: - In all three states where there was a high risk of coastal flooding, there was a significant percentage of low income population. - In all three states where there was high risk of river flooding and wildfire risk there was not a significant percentage of low income population.

##Limitations The datasets we decided to use were very dense, making it difficult for R to adequately download the values. Unfortunately, a lot of filtering for the original EJScreen dataset was required to actually upload, push, and pull the data from GitHub to R and vice versa. This left a lot of precious time that could have been spent exploring air quality, which was another natural hazard risk that we initially planned to analyze.

Furthermore, there was a lot of data for the state of California compared to North Carolina and Florida. That said, it made it difficult for us to understand and identify a trend.