Evaluation of the relationship between socioeconomic status and flood risk in coastal New Jersey

GEO 327G/386G

GIS & GPS Applications in the Earth Sciences

Fall 2019

Dr. Mark Helper

Semester Project

Prepared by: Megan Bushlow

December 9, 2019

1 Background

Flooding associated with tropical storms or hurricanes can cause widespread damage to infrastructure, disrupt daily life, and displace families. One of the most damaging storms to hit the northeastern United States in recent years was Hurricane Sandy. In October of 2012, Hurricane Sandy caused extensive damage along the coast of New Jersey, which has approximately 130 miles of coastline along the Atlantic Ocean.[1] Hurricane Sandy displaced thousands of families from their homes, shut-down public transportation systems for days, and caused an estimated $70 billion in damage across the Northeast.[2] Across the United States, lower income families are often more likely to be impacted by disasters because they are more likely to live within the floodplain. When they are impacted, lower income families are more likely to have significant damages because they are more likely to rent rather than own property, to have jobs that are not accommodating to missed work, and to live in neighborhoods with more vulnerable infrastructure.[3] Lower income families also tend to have less political power, which puts them at a disadvantage during recovery periods.[3]

2 Problem Formulation

The purpose of this project is to investigate the relationship between socioeconomic status and flood risk in coastal New Jersey. Are more vulnerable socioeconomic groups more likely to reside in the floodplain in New Jersey? Due to the larger impacts of flooding on groups of certain socioeconomic status, an understanding of the flood risk for different groups is valuable to predict the impacts of flooding after a storm and to guide the allocation of resources during recovery.

The relationship between flood risk and socioeconomic status was evaluated within the eight coastal counties of New Jersey, in the area three miles within the coastline. The area was limited to three miles from the coastline to reduce file size for analyses. As described in Section 3.1, data on socioeconomic status was obtained from census data records, and data on flood risk was obtained using Federal Emergency Management Agency (FEMA) Flood Insurance Rate Maps (FIRM). These FEMA flood maps are the standard maps used to determine flood insurance rates in the US.

To evaluate the relationship between socioeconomic status and flood risk, a variety of operations were performed in ESRI's ArcGIS Desktop and Microsoft Excel. The outputs of the analyses performed for this project are a set of charts and tables presenting the socioeconomic composition (income and race) of different flood risk categories.

3 Methods

This section presents the data collection, data pre-processing, data processing, and data analysis methods that were used to evaluate the relationship between flood risk and socioeconomic status in coastal New Jersey.

3.1 Data Collection

As stated in Section 2, the two primary data sets used in the analyses for this project were census data obtained from the New Jersey Department of Environmental Protection (NJDEP) Bureau of GIS and flood map data from FEMA.

Definitions of the flood zone areas included in a FIRM are provided in FEMA's "How to Read a Flood Insurance Rate Map Tutorial" document.[4] A summary of the flood zone area definitions relevant to this project is included below:

- Zone X: area outside of 100-year floodplain
- Zone D: area with possible flood hazards
- Zone A: area within 100-year floodplain, determined by approximate methods
- Zone AE: area within 100-year floodplain, determined by detailed methods

- Zone AH: area of 100-year shallow flooding, where average depths of ponding range between 1 and 3 feet, determined by detailed methods
- Zone AO: area of 100-year shallow flooding, where average depths of sheet flow range between 1 and 3 feet, determined by detailed methods
- Zone VE: area within 100-year floodplain with additional coastal hazards, determined by detailed methods

For purposes of this project, all zone categories beginning with "A" were grouped together. The order of increasing flood risk was considered to be X, D, A, and VE.

Income data in spreadsheet format from the US Census Bureau was downloaded from the American FactFinder site. Although census data is available at the block level, income data was found only at the census tract level, so all analyses were performed at the census tract level. Table 1 provides information on these data sets.

Several supplemental data sets were obtained to support the analysis. Table 2 provides information on these supplemental data sets.

3.2 Data Pre-processing

The data files obtained needed to be projected into the same projected coordinate system, as applicable. FIRM shapefiles for the eight coastal counties in New Jersey (Atlantic, Bergen, Cape May, Hudson, Middlesex, Monmouth, Ocean, and Union) were displayed in ArcGIS and merged into a single shapefile using the "Merge" tool. The merged shapefile was projected into the New Jersey State Plane coordinate system using the "Project" tool. The merged, projected shapefile is displayed with the New Jersey counties shapefile in Figure 1, showing the flood zone areas present in these counties. All shapefiles other than the FIRM were downloaded already having the appropriate New Jersey State Plane coordinate system projection.

3.3 Data Processing and Analysis

The socioeconomic composition of the flood zone areas in the coastal counties of New Jersey was evaluated using two approaches: a polygon overlay analysis and raster map algebra. The two approaches are described in the following sections.

3.3.1 Polygon Overlay Analysis

To limit the file size for analyses, only areas within three miles of the coastline were included. A three-mile buffer of the *coast* shapefile was created using the "Buffer" tool with the *coast* shapefile as the input feature, three miles as the linear unit, side type of "full," end type of "round," the "planar" method, and dissolve type of "all." A screenshot of the resulting buffer area is shown in Figure 2.

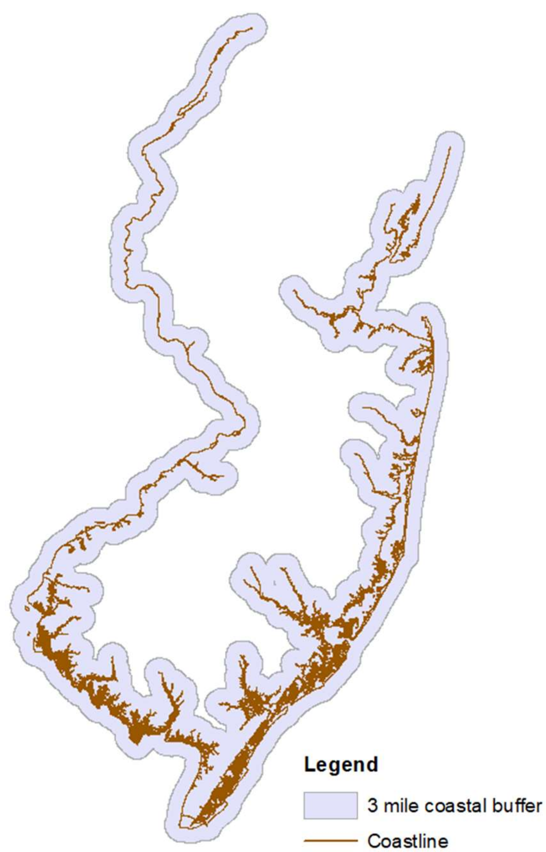


Figure 2. Three-mile buffer of New Jersey coastline

Table 1. Primary data sets used						
Data	**Name**	**Source**	**Description**	**Type**	**Projection**	**Notes**
2010 Census data	Govt_census_tract_2010	NJDEP Bureau of GIS	2010 US Census data for New Jersey at the census tract level. Attributes include an identifier for each census tract (GEOID10), the associated county and state, ethnic breakdown, age and gender breakdown, family size, and shape (census tract) area in square feet	Polygon shapefile	NAD 1983 New Jersey State Plane Coordinates (uses Transverse Mercator projection)	Metadata (including description of attributes) provided in text file included with download
Income data	NJ_Income_Data	US Census Bureau American FactFinder	2017 income data for New Jersey at the census tract level	CSV	N/A	Metadata provided in text file and separate CSV included in download
Flood risk maps	S_FLD_HAZ_AR_[COUNTY]	FEMA National Flood Hazard Layer (NFHL) Inventory site	Current FIRM utilized by FEMA. Files were downloaded individually for each of the eight counties included in this project	Polygon shapefile	Unprojected with GCS NAD 1983 Geographic Coordinate System	Various other shapefiles were included in the download. Only the layer showing special flood hazard areas (SFHA) was used. Metadata for each FIRM (by county) was provided in an XML file included in the download

Table 2. Supplemental data sets used						
Data	**Name**	**Source**	**Description**	**Type**	**Projection**	**Notes**
New Jersey counties	New_Jersey_Counties	NJDEP Bureau of GIS - OpenData	Counties in New Jersey	Polygon shapefile	NAD 1983 New Jersey State Plane Coordinates (uses Transverse Mercator projection)	Metadata (including description of attributes) provided in XML file included with download
New Jersey rivers	ir_river_all2004	NJDEP Bureau of GIS	2004 data on rivers in New Jersey	Polyline shapefile	NAD 1983 New Jersey State Plane Coordinates (uses Transverse Mercator projection)	Metadata (including description of attributes) provided in text file included with download
New Jersey coastline	coast	NJDEP Bureau of GIS - OpenData	Coastline in New Jersey, last updated 2009	Polyline shapefile	NAD 1983 New Jersey State Plane Coordinates (uses Transverse Mercator projection)	Metadata (including description of attributes) provided in text file included with download

Figure 1. Flood zone areas in eight coastal counties of New Jersey

Legend

Counties

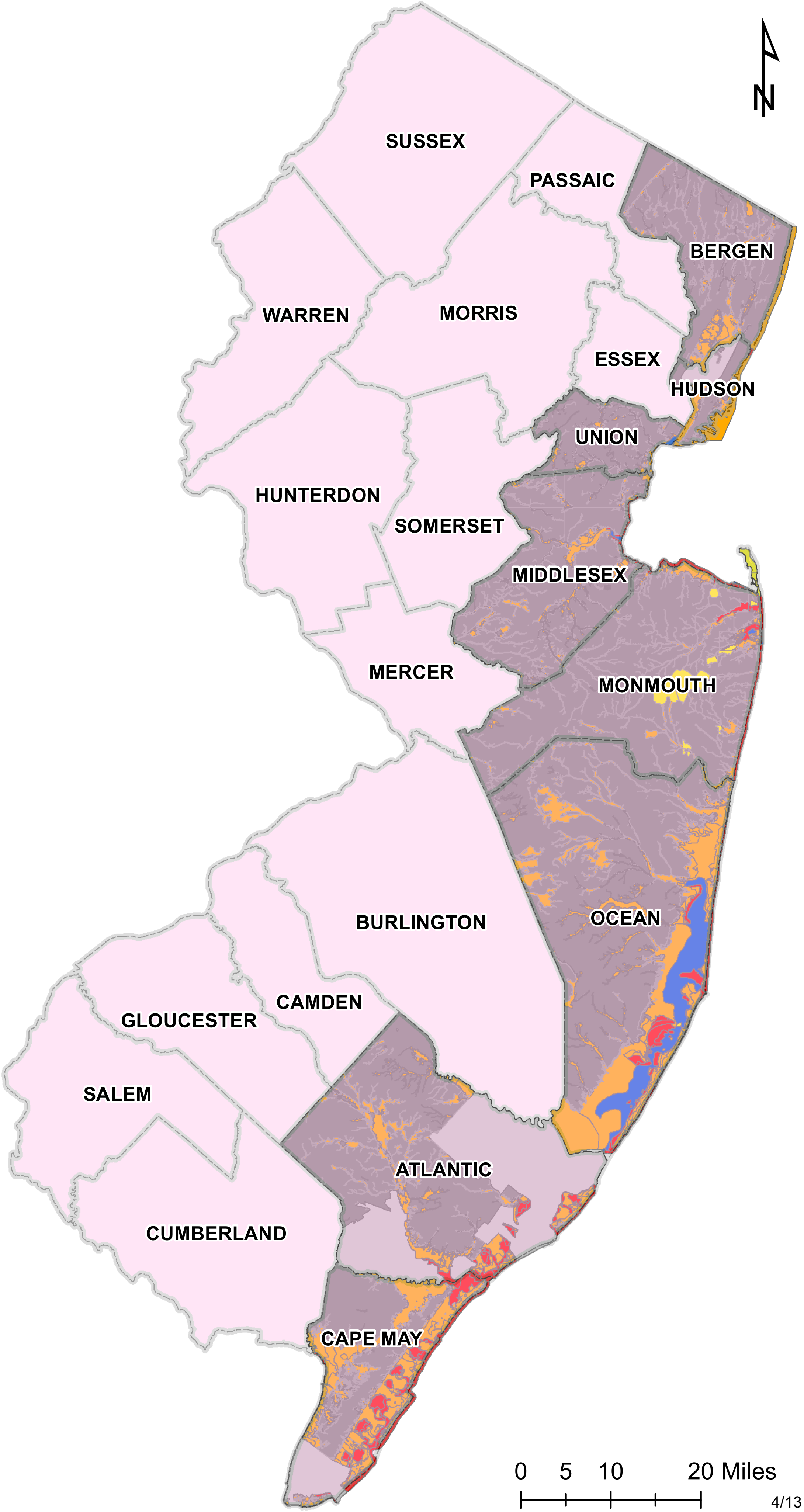
FIRM (Projected)

Flood Zone

- A; AE; AH; AO
- AREA NOT INCLUDED
- D
- OPEN WATER
- VE
- X

Datum: NAD 1983
Projection: NJ State Plane Coordinate System

Sources:
NJDEP Bureau of GIS
FEMA NFHL Inventory



Prepared by: Megan Bushlow
Date: December 9, 2019

After the buffer was created, the merged flood zone layer and the census tract layer were clipped to the buffer area using the “Clip” tool. The “Intersect” tool, within the Overlay Analysis toolset, was then used on the clipped flood zone and census tract layers. The output of this tool was a shapefile of polygons divided into smaller portions having the same flood zone category and census tract. Only polygons that were the intersection of both layers were included in the output shapefile; therefore, any of the census tracts that were included in the buffer area along the west side of the state were eliminated using this operation.

The income data obtained from the American FactFinder site was joined to the attribute table resulting from the intersection operation; this was completed using the GEOID identifier for each census tract as the field on which to base the join. The resulting attribute table was exported to Microsoft Excel for further analysis.

The US Census Data is broken into sixteen income categories; three aggregated income categories were defined for purposes of this project, as listed in Table 3. The total number of households in each census tract/flood zone combination, in each of the three income categories, was calculated by summing the sixteen categories appropriately. Pivot tables were used to create summary tables of percentages for the overall area and for individual counties. Resulting tables are presented in Section 4.

Table 3. Income categories for this project

Category	**Income Range ($)**
Low income	< 40,000
Middle income	40,000 - 125,000
High income	>= 125,000

3.3.2 Raster Map Algebra

An evaluation of the relationship between flood risk and socioeconomic status was also performed using rasterized data. The merged, clipped, flood zone shapefile was rasterized using the “Polygon to Raster” tool, with *FLD_ZONE* as the value field and 100 ft as the cell size.

In Excel, an average income value was assigned to each census tract by weighting the income for the number of households in each of the sixteen income categories. A table of the income values for each census tract was joined to the census data shapefile (which was clipped to the buffer area) using the GEOID identifier as the field on which to base the join. The shapefile was then rasterized using the “Polygon to Raster” tool, with *Income* as the value field and 100 ft as the cell size. The “Extract by Mask” tool was used to extract the portion of the income raster that matched the flood risk raster.

The goal of the raster operations was to determine the portion of cells with different flood categories and income categories. To do this, map algebra was used to assign values to each cell based on the assignment rule presented in Table 4.

The values in Table 4 were assigned such that the lowest value, 1, could be considered the lowest vulnerability (not within the 100-year floodplain and high income) and the highest value, 12, could be considered the highest vulnerability (within the 100-year floodplain with additional coastal flood risks and low income).

Table 4. Assigned values for income and flood zone combined categories

Income Category	Flood Zone Category	Assigned Value
High	X	1
Middle	X	2
Low	X	3
High	D	4
Middle	D	5
Low	D	6
High	A/AE/AH/AO	7
Middle	A/AE/AH/AO	8
Low	A/AE/AH/AO	9
High	VE	10
Middle	VE	11
Low	VE	12

The following syntax was used to perform the map algebra operation:

```
Con(("income_extr" < 40000) & ("FldZone_grd" == 2), 3, Con(("income_extr" < 40000)  & (("FldZone_grd"
== 1) | ("FldZone_grd" == 3) | ("FldZone_grd" == 6) | ("FldZone_grd" == 9)), 9, Con(("income_extr" <
40000) & ("FldZone_grd" == 8), 6, Con(("income_extr" < 40000) & ("FldZone_grd" == 4), 12,
Con(("income_extr" >= 40000)  & ("income_extr" < 125000)  & ("FldZone_grd" == 2), 2,
Con(("income_extr" >= 40000)  & ("income_extr" < 125000)  & (("FldZone_grd" == 1) | ("FldZone_grd" ==
3) | ("FldZone_grd" == 6) | ("FldZone_grd" == 9)), 8, Con(("income_extr" >= 40000)  & ("income_extr" <
125000)  & ("FldZone_grd" == 8), 5, Con(("income_extr" >= 40000)  & ("income_extr" < 125000)  &
("FldZone_grd" == 4), 11, Con(("income_extr" >= 125000) & ("FldZone_grd" == 2), 1, Con(("income_extr"
>= 125000) & (("FldZone_grd"  == 1) | ("FldZone_grd" == 3) | ("FldZone_grd" == 6) | ("FldZone_grd" ==
9)), 7, Con(("income_extr" >= 125000) & ("FldZone_grd" == 8), 4, Con(("income_extr" >= 125000) &
("FldZone_grd" == 4), 10, 0))))))))))))
```

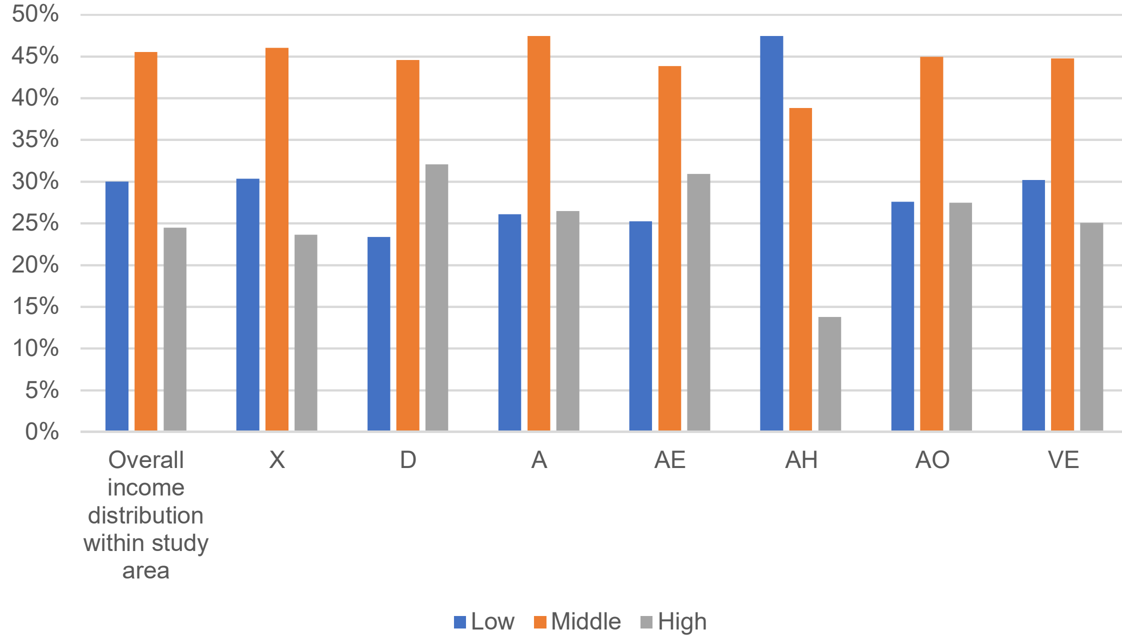
4 Results

This section presents the results of the two approaches for evaluating the relationship between socioeconomic status and flood risk.

4.1 Polygon Overlay Analysis

The results of the polygon overlay analysis are in the form of compositions of the various flood zone areas by socioeconomic group. As shown in Figure 3, for the study area overall (area in eight coastal counties within three miles of the coastline), 30% of the households were low income, 46% of the households were middle income, and 24% of the households were high income. The income distributions within the various flood zone categories closely matched the overall income distribution, indicating that there was not a strong relationship between flood risk and income. However, the flood zone category of AH did show a larger percentage of low income households, with 47% of households in AH in the low income category and only 14% of households in AH in the high income category. The flood zone category of D had a moderately larger percentage of high income households, relatively, with only 23% of households in D in the low income category, and 32% in the high income category.

Figure 3. Income composition by flood zone for entire study area



Figures 4 through 11 present the income distributions of flood zones for each of the eight coastal counties studied. These figures show that, even when evaluated by county, the income distributions of each flood zone category generally match the income distributions of the counties overall.

Figure 4. Income composition by flood zone in Atlantic County

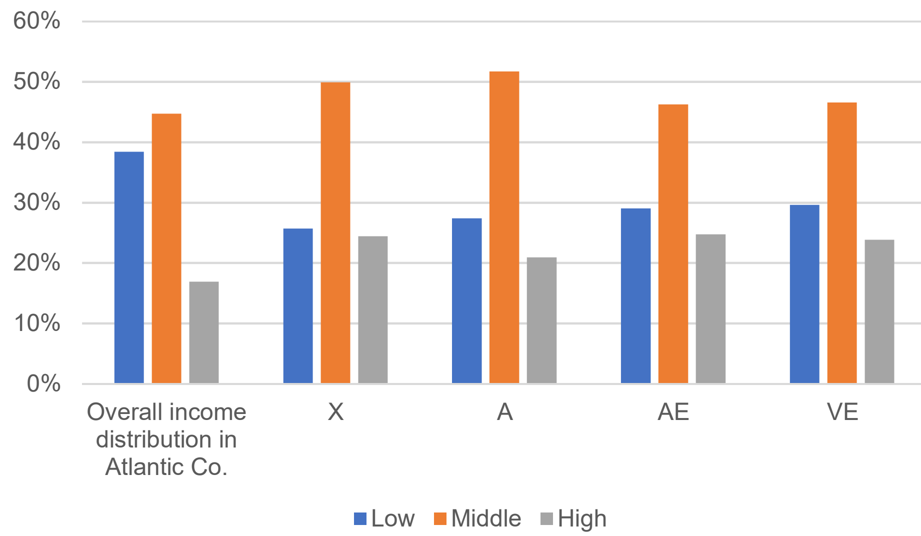


Figure 5. Income composition by flood zone in Bergen County

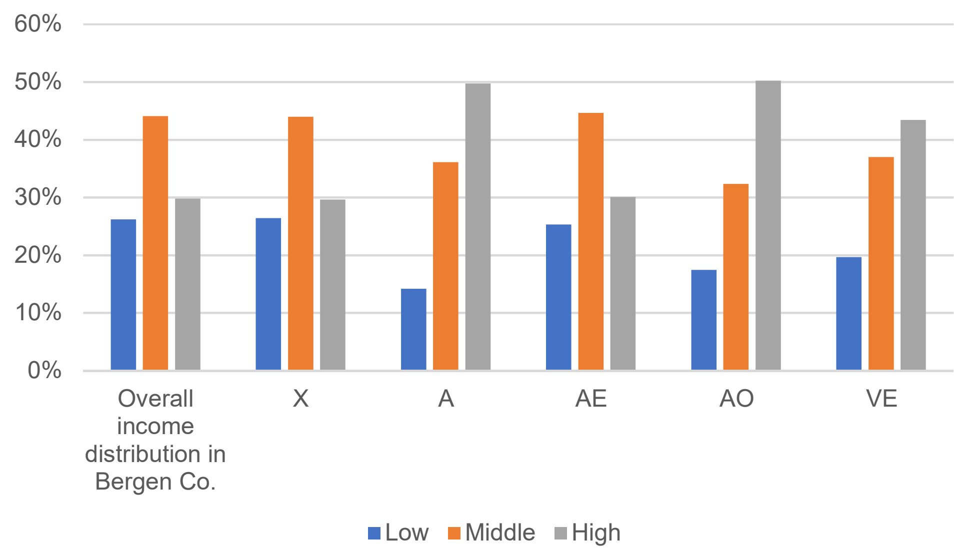


Figure 6. Income composition by flood zone in Cape May County

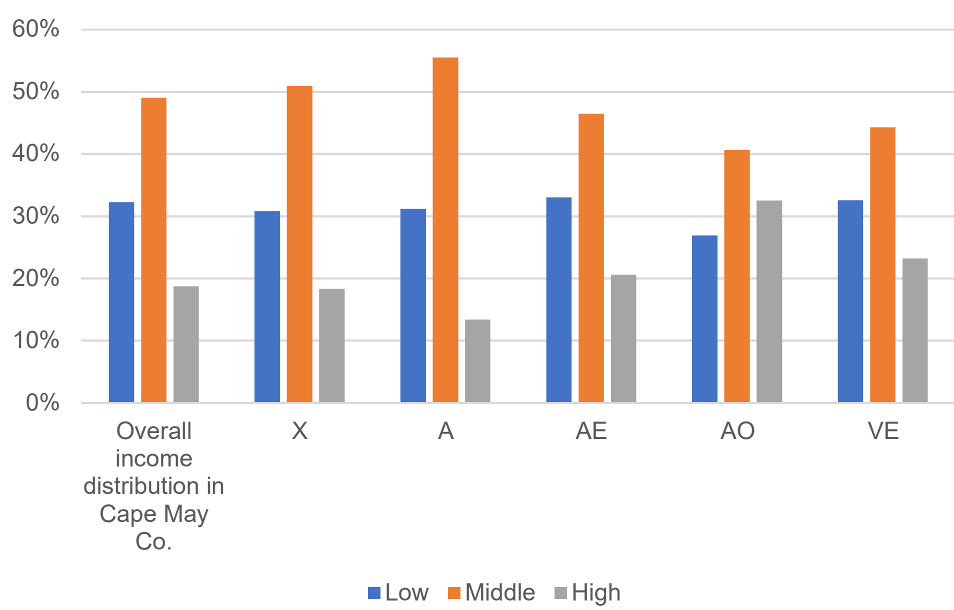


Figure 7. Income composition by flood zone in Hudson County

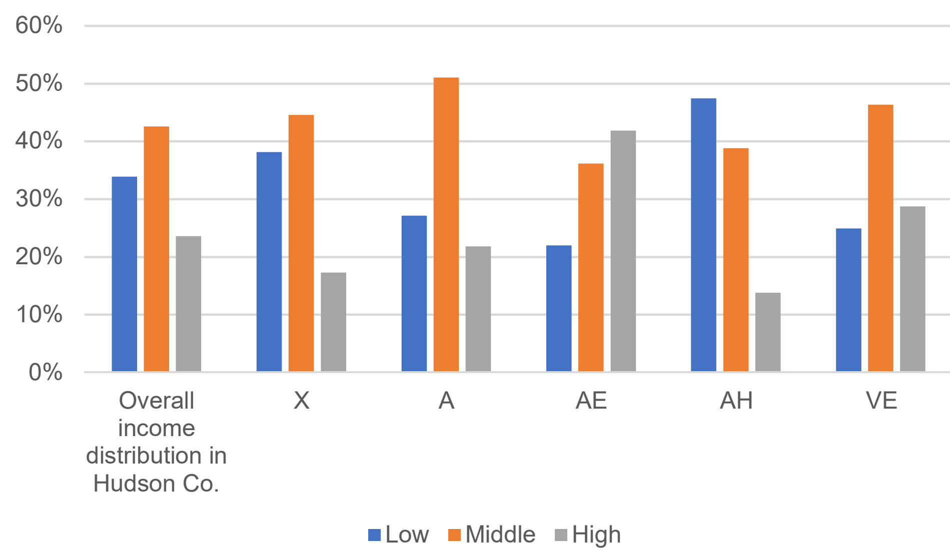


Figure 8. Income composition by flood zone in Middlesex County

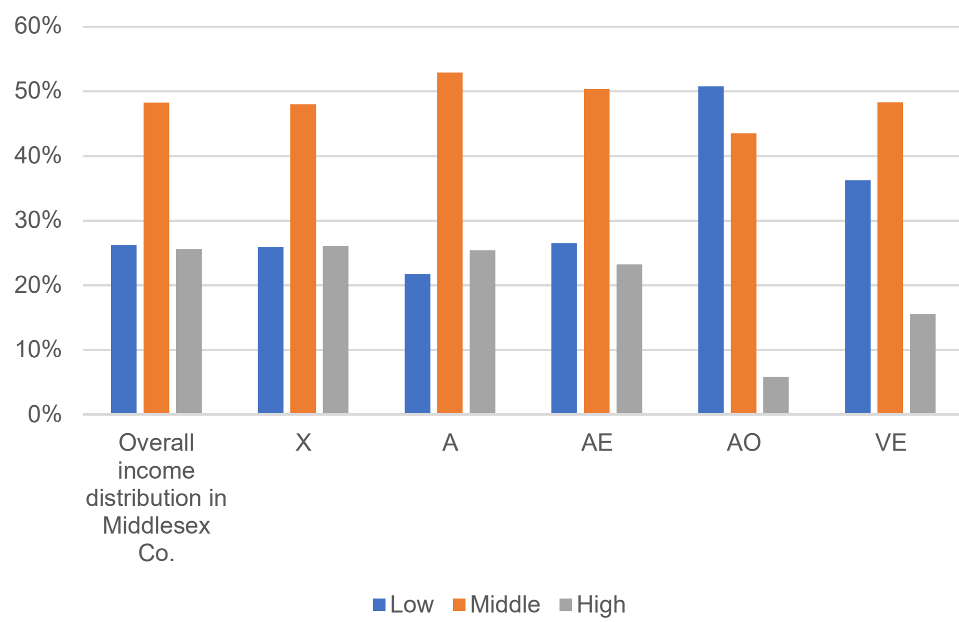


Figure 9. Income composition by flood zone in Monmouth County

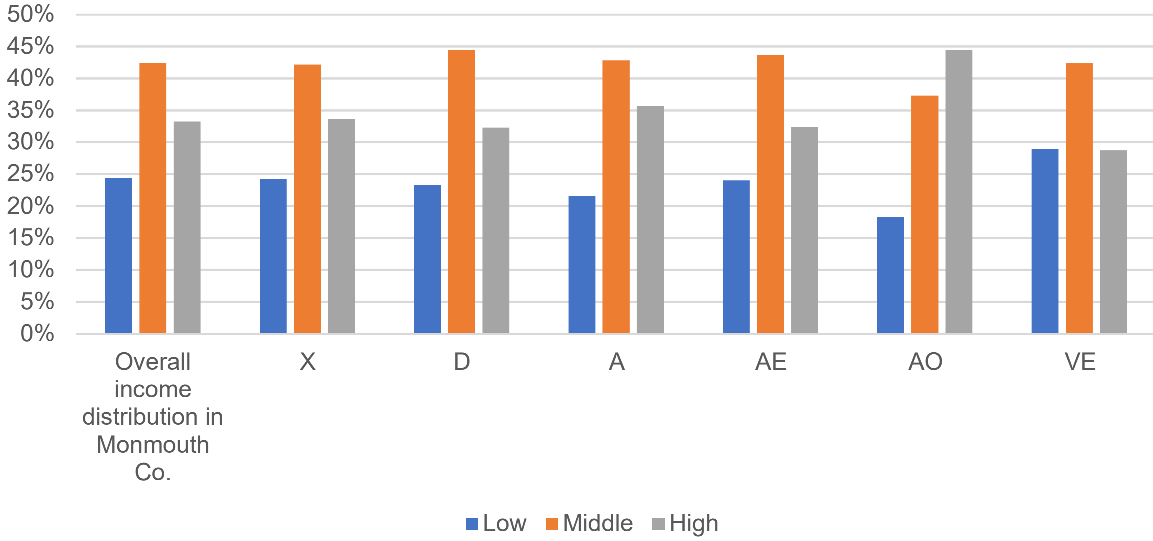


Figure 10. Income composition by flood zone in Ocean County

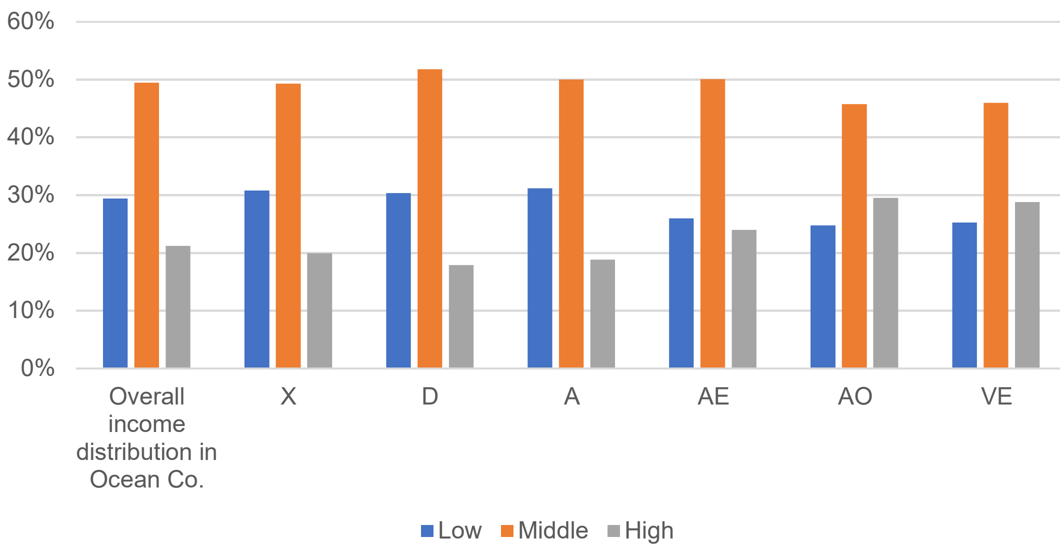
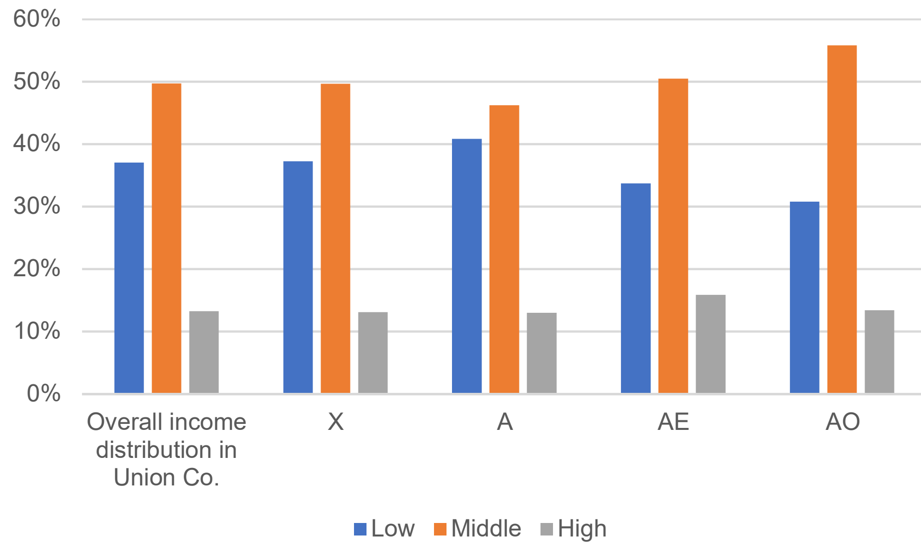
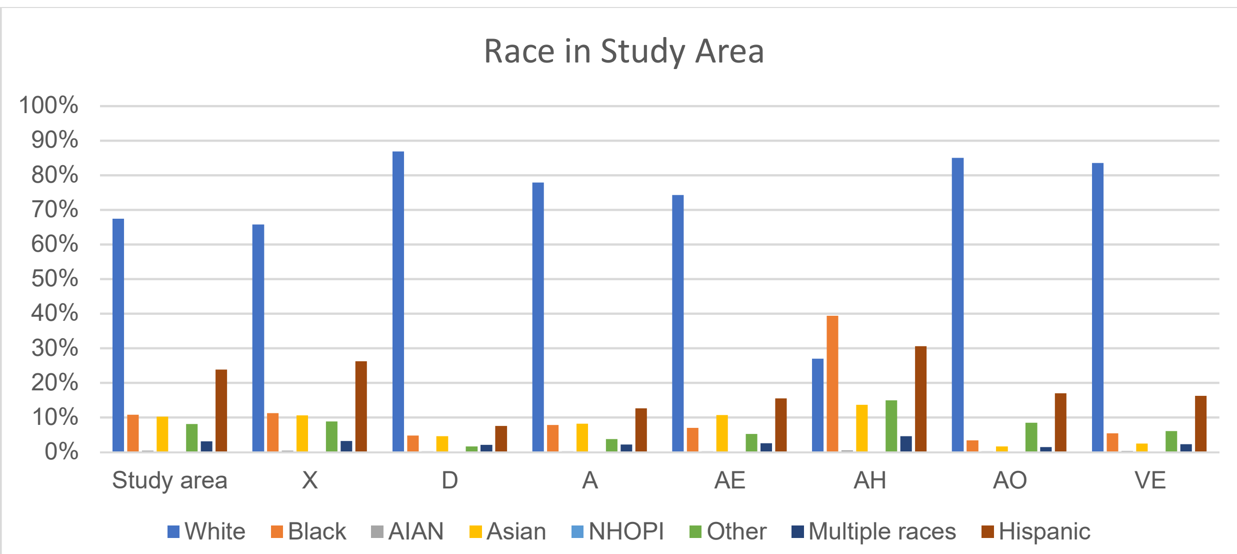


Figure 11. Income composition by flood zone in Union County



Race was another variable studied in relation to flood risk. As shown in Figure 12, for the study area overall, the race composition was 67% white, 11% black, 10% Asian, 8% other, and 3% multiple races, with 24% of the population Hispanic. Figure 12 indicates that the race compositions of the various flood zones generally match the race composition of the study area overall. However, the AH category does show a lower portion of white residents and higher portion of other races as compared to the overall distribution.

Figure 12. Race composition by flood zone in Union County



4.2 Raster Map Algebra

Figure 13 shows the flood vulnerability within the study area resulting from the map algebra described in Section 3.3.2. Table 5 lists the portions of the study area covered by each of the combined categories. As shown in Table 5, approximately 50% of the study area was not within the 100-year floodplain. Approximately 28% of the study area was characterized by middle income households within the 100-

Figure 13. Flood vulnerability within study area

Legend

Counties

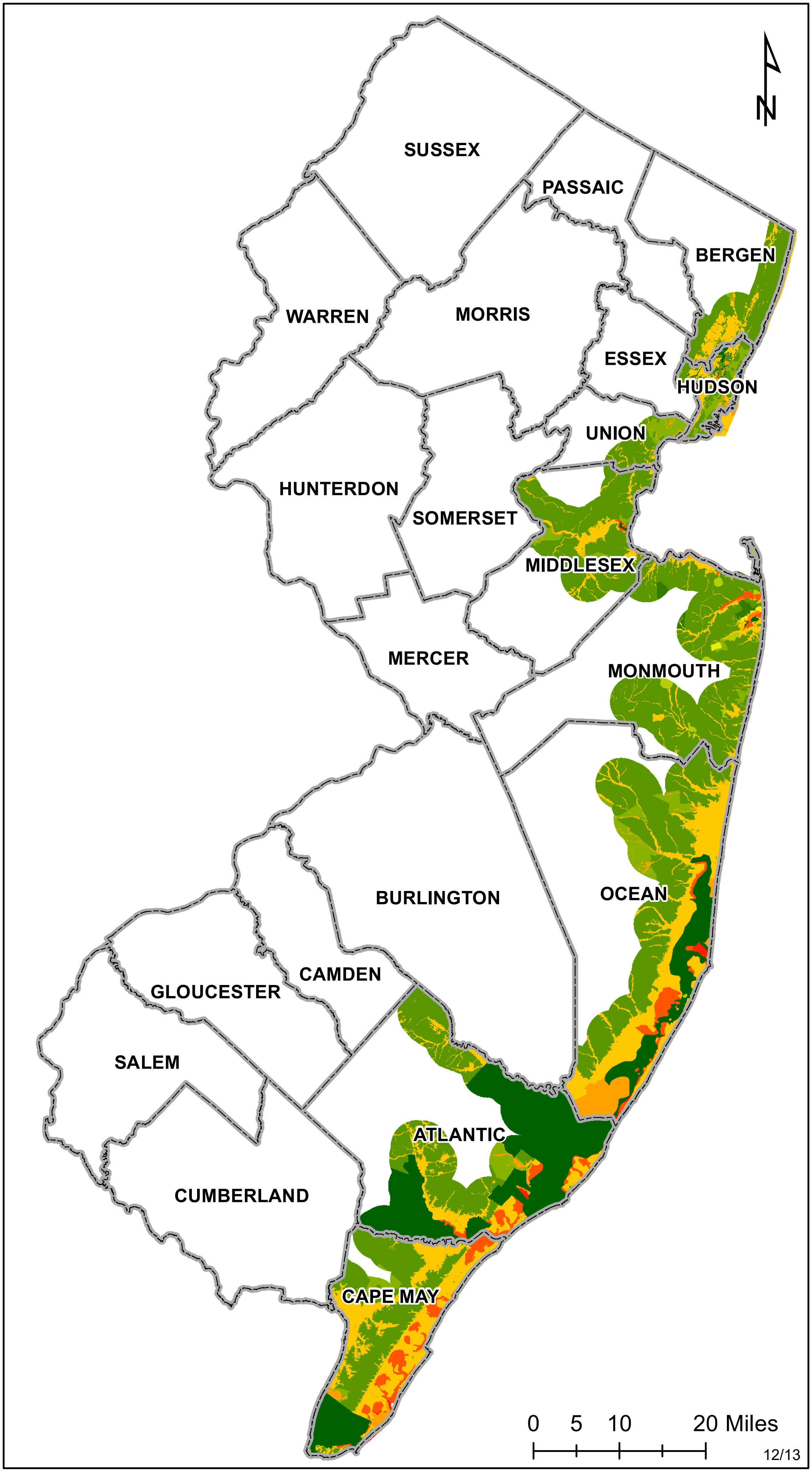
Assigned Value

0
1
2
3
5
6
7
8
9
10
11
12

Datum: NAD 1983
Projection: NJ State Plane Coordinate System

Sources:
NJDEP Bureau of GIS

Prepared by: Megan Bushlow
Date: December 9, 2019



year floodplain, while only 3% and 0% of the study area was characterized by low and high income households, respectively, within the 100-year floodplain.

Table 5. Portions of study area falling within combined income/flood risk categories

Description	Value	Percent of Study Area
High income, X	1	0%
Middle income, X	2	47%
Low income, X	3	4%
High income, D	4	0%
Middle income, D	5	1%
Low income, D	6	0%
High income, A	7	0%
Middle income, A	8	22%
Low income, A	9	2%
High income, VE	10	0%
Middle income, VE	11	5%
Low income, VE	12	1%
None	0	17%

5 Conclusions and Summary

Two approaches, polygon overlay analysis and raster map algebra, were used to evaluate the relationship between flood risk and socioeconomic status in coastal New Jersey. The analysis indicated that there is not a strong relationship between flood risk and the socioeconomic factors of income and race in coastal New Jersey. Several factors may have contributed to these results, including the categories selected for analysis and the nature of coastal areas in New Jersey. The results showed that a larger percentage of the study area had middle income households, and that the highest percentage of households within the floodplain were middle income; if the definitions for low, middle, and high incomes were changed, the analysis may yield different results. The coastal areas in New Jersey are largely composed of urban areas (such as Hoboken and Jersey City in the north) or summer beach towns that have high levels of seasonal tourism. As a result, it is possible that larger portions of coastal neighborhoods, which have higher flood risks, are middle to high income compared to other areas of high flood risk in the country.

There were several limitations of the analyses performed for this project. Census tract level income data was used, so incomes were not known at a very local scale. Flood risk was characterized by FEMA flood maps, which may be out of date and may not accurately represent flood risk for all areas. Future analysis could use alternative methods for characterizing flood risk. The analysis was limited to the area within three miles of the Atlantic coastline. However, the analysis performed for this project resulted in a general understanding of the economic and race characteristics within coastal New Jersey.

6 Sources

1. https://stockton.edu/coastal-research-center/njbpn/geologic-hist.html
2. https://www.nationalgeographic.com/environment/natural-disasters/reference/hurricane-sandy/
3. https://www.governing.com/commentary/col-disasters-disadvantaged-climate-justice.html
4. https://www.fema.gov/media-library-data/20130726-1550-20490-1950/ot_firm.pdf