Urban Flood Risk and Social Vulnerability Assessment in Wilmington, Delaware: A Comparative Study of Machine Learning Techniques

SPPA667 Urban Evidence Based Policy
Project Final Presentation

Kyungmin Lee

Dec 18, 2020

Contents

1. Introduction

Urban flood

Minority, Vulnerability, Justice

2. Background

Project Overview

Urban Evidence Based Policy Approach

- The Evidence Based Policy is to make more defensibly policy decisions based on conscientious, explicit, and judicious use of scientific evidence by **using science (Big Data and machine learning) as evidence in public policy** (Straf et al., 2012).

- In this way it is useful for policy makers by **providing insights into which policy interventions are most likely to lend desirable outcomes** (Androutsopoulou et al., 2018).

2. Background

Project Overview

Research Purpose

This research aims to find out the circumstances of the socioeconomic and demographic status is in the flood zone and suggest policy implications.

Research Question

Are vulnerable people living in the Wilmington city exposed to flood risk?

3. Data Description

Raw Data	Year	Type	Source
Census Block Groups	2016	Shapefile (.shp)	U.S. Census Bureau, Department of Commerce
Flood Depth Grid	2014	Raster	Federal Emergency Management Agency (FEMA)
Census data: Socioeconomic and Demographic Status (SED)	2014 (ACS5)	CSV	U.S. Census Bureau

3. Data Description

Census Block Groups - Delaware State

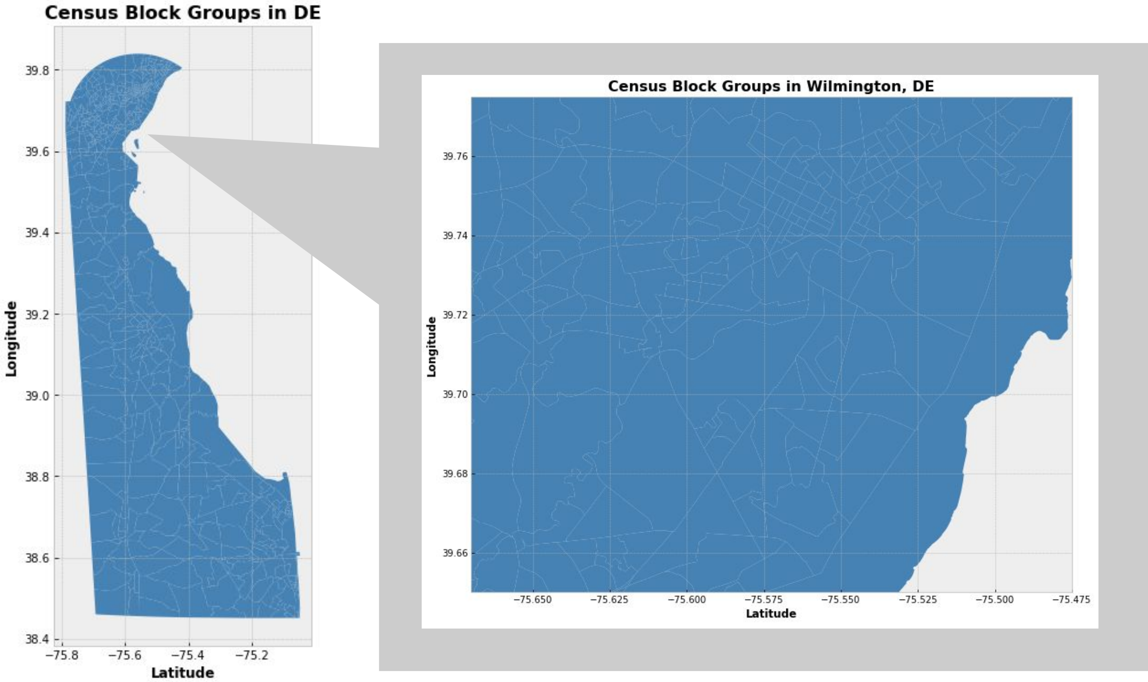


Fig 2. Census Block Groups in New Castle County. This figure is based on 2016 shapefile.The raw data shapefile includes water area, where they are generally in territorial seas, coastal water, and Great Lakes water areas.

3. Data Description

Flood Depth Grid (CstDpth_01pct) - New Castle County

* Flood hazard is defined by a relation between depth of flooding and the annual chance of inundation greater than that depth (FEMA, 2014).

* CstDpth_01pct [Unit: Feet]: Coastal flood depth for the 1 percent annual chance flood event.

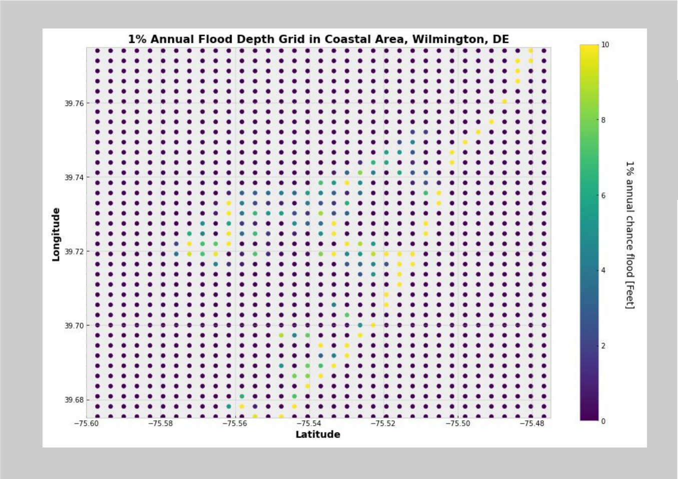
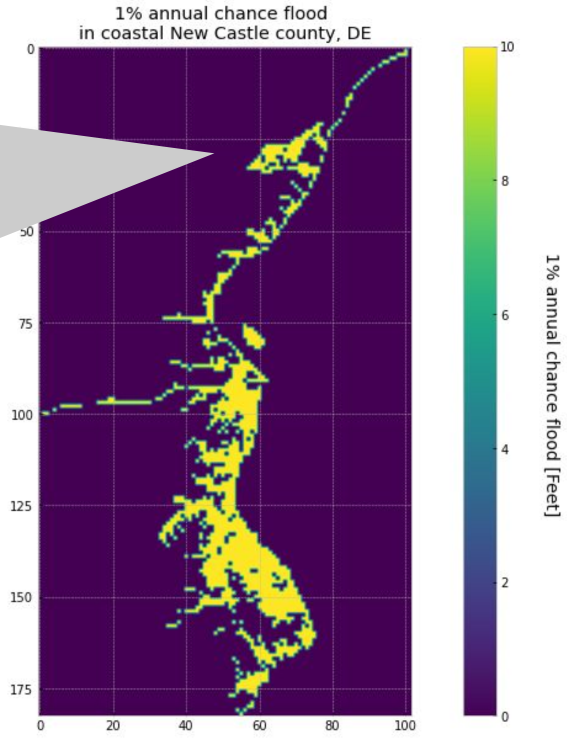


Fig 3. Flood Depth Grid with Latitude and Longitude
The raw data is based on the 2014 raster data from the FEMA. It shows that the local area surrounded by the Delaware River has the flood risk.



3. Data Description

Census Data: Socioeconomic and Demographic (SED) data

	lat	lon	dep	dep_cl	mincome	per_nonwhite	per_below_povlev	population	no_school	no_school_frac	under18	over65	has_water_int	log_dep
count	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000	203.000000
mean	39.750172	-75.553137	0.516256	0.516256	55261.177340	57.237379	18.446325	967.147783	727.669951	0.756534	200.970443	135.093596	0.285714	0.066267
std	0.011163	0.016557	1.620714	1.620714	30641.633629	32.408620	17.069917	315.344881	232.317797	0.087995	128.113460	98.604802	0.452871	0.203469
min	39.724676	-75.586413	0.000000	0.000000	10177.000000	3.810976	0.000000	389.000000	280.000000	0.417404	21.000000	12.000000	0.000000	0.000000
25%	39.741115	-75.565193	0.000000	0.000000	32206.000000	20.620621	2.692308	719.500000	576.000000	0.701105	99.000000	64.000000	0.000000	0.000000
50%	39.752075	-75.554583	0.000000	0.000000	47976.000000	66.979362	16.071429	962.000000	745.000000	0.762509	187.000000	108.000000	0.000000	0.000000
75%	39.760295	-75.540436	0.000000	0.000000	69403.500000	88.115942	28.991597	1141.000000	876.000000	0.793226	261.000000	178.000000	1.000000	0.000000
max	39.771255	-75.519216	10.000000	10.000000	162143.000000	100.000000	77.600000	1792.000000	1179.000000	1.022071	673.000000	390.000000	1.000000	1.000000

4. Methodology

4.1. Data summary

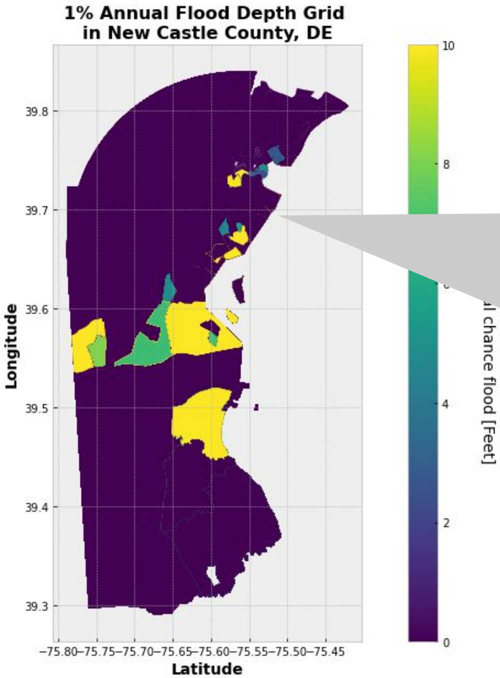
4.2. Geospatial analysis

4.3. Machine learning analysis

1) Random Forest Classifier
2) Logistic Regression Classification
3) Random Forest Regressor

4. Methodology

Geospatial Analysis



1) Spatial join

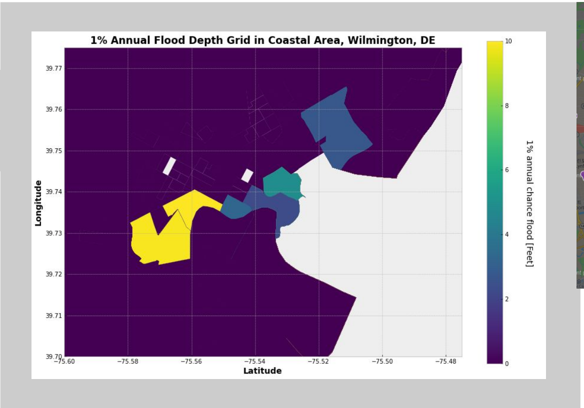
- Flood depth grid & Census block groups

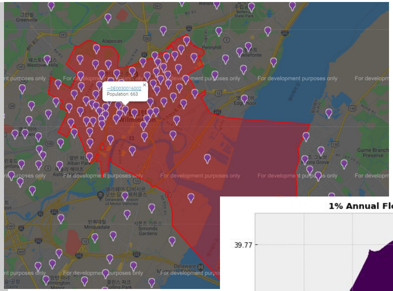
2) Merge

- Shapefile to the SED data

3) Subset

- New Castle county to Wilmington city





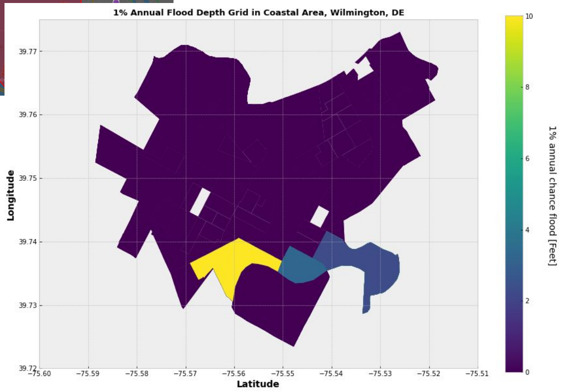
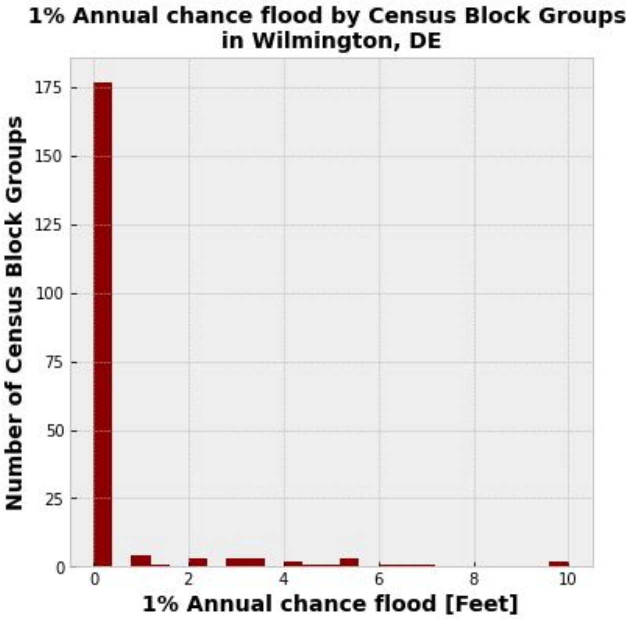


Fig 4. Flood risk in the Wilmington City

Flood depth grid based on census block groups shapefile

5. Analysis Results

5.1. Data summary



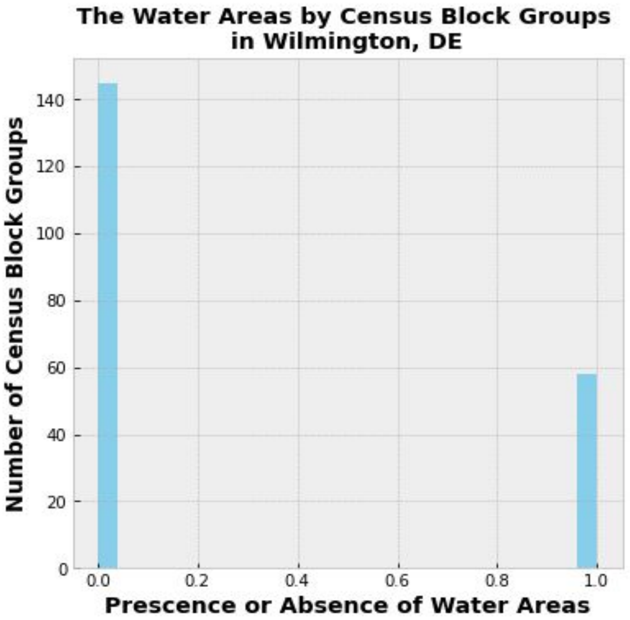


Fig 4. The Flood Depth & Water Areas

The flood risk status by census block groups in Wilmington, DE. The 1% annual chance flood is multilevel data set, while the water area is binary dataset.

5. Analysis Results

5.1. Data summary

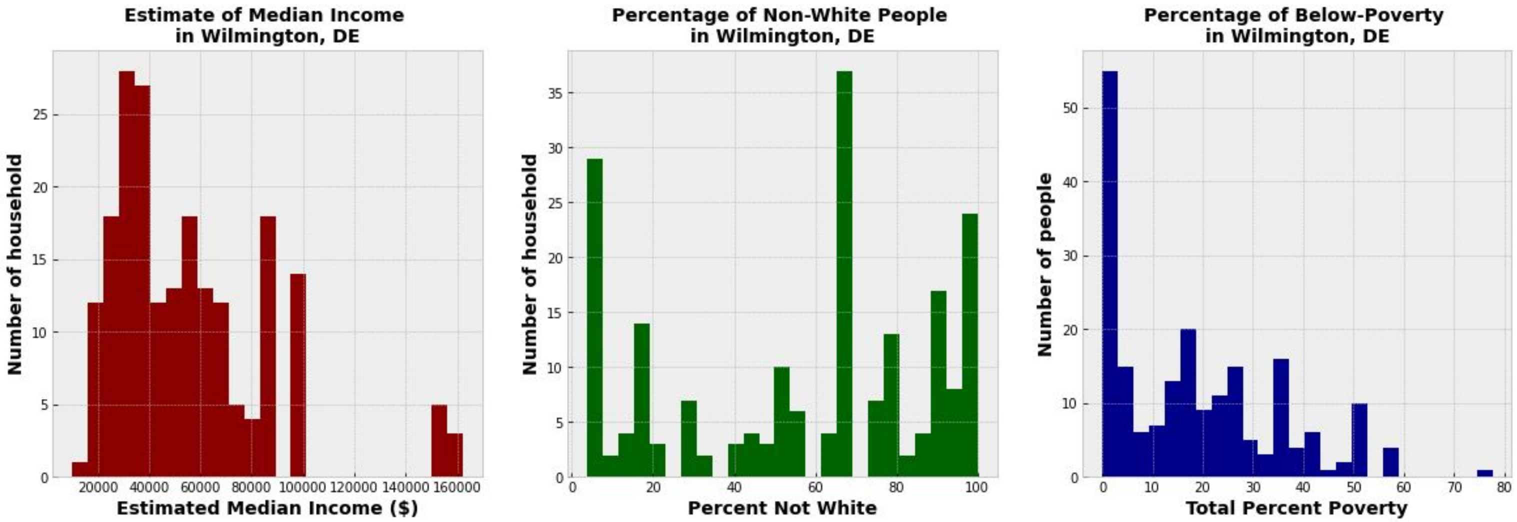


Fig 4. The socioeconomic demographic status in Wilmington, DE. The data is estimated from 2014 to 2018 and the data source is from the U.S. Census Bureau. Based on the census block groups, the mean of the median household income is $55,261, the percentage of not white people is 57.24%, and the rate of the total percentage of people under poverty line is 18.45%.

5. Analysis Results

5.1. Data summary

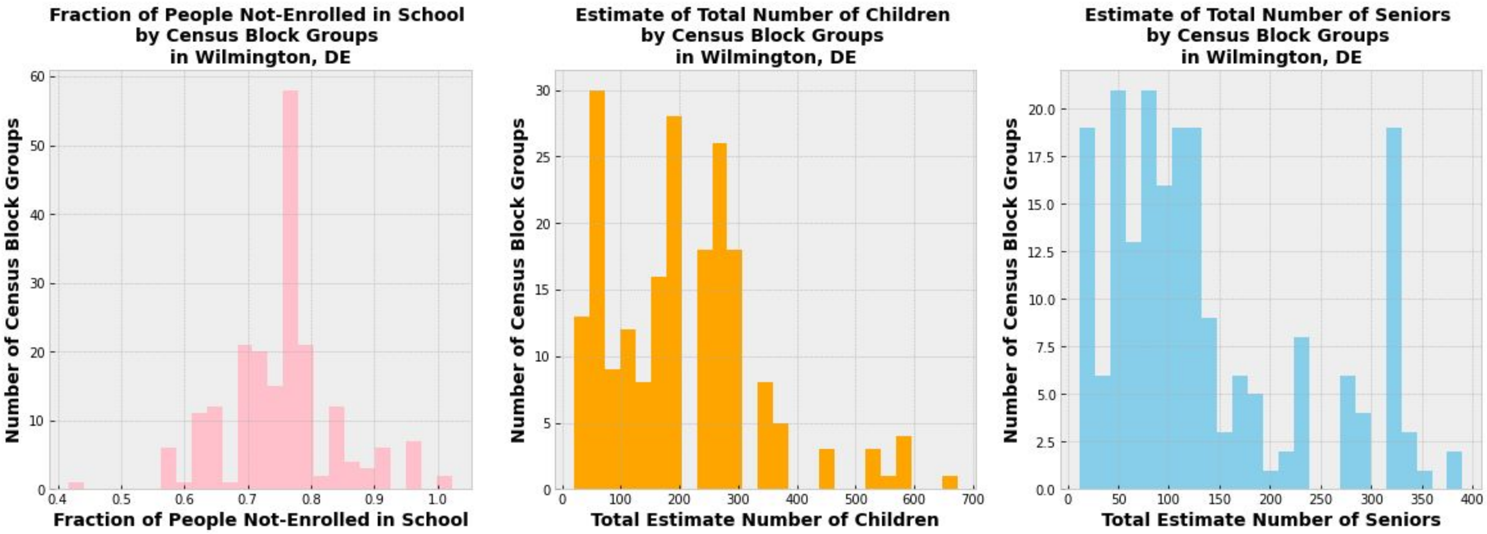


Fig 4. The socioeconomic demographic status in Wilmington, DE. The data is estimated from 2014 to 2018 and the data source is from the U.S. Census Bureau. Based on the census block groups, the mean of the fraction of people not enrolled in school is 0.76 (75.65%), the total estimate number of children (age under 18) is 201, and the total estimate number of seniors (age over 65) is 135.

5. Analysis Results

5.1. Data summary

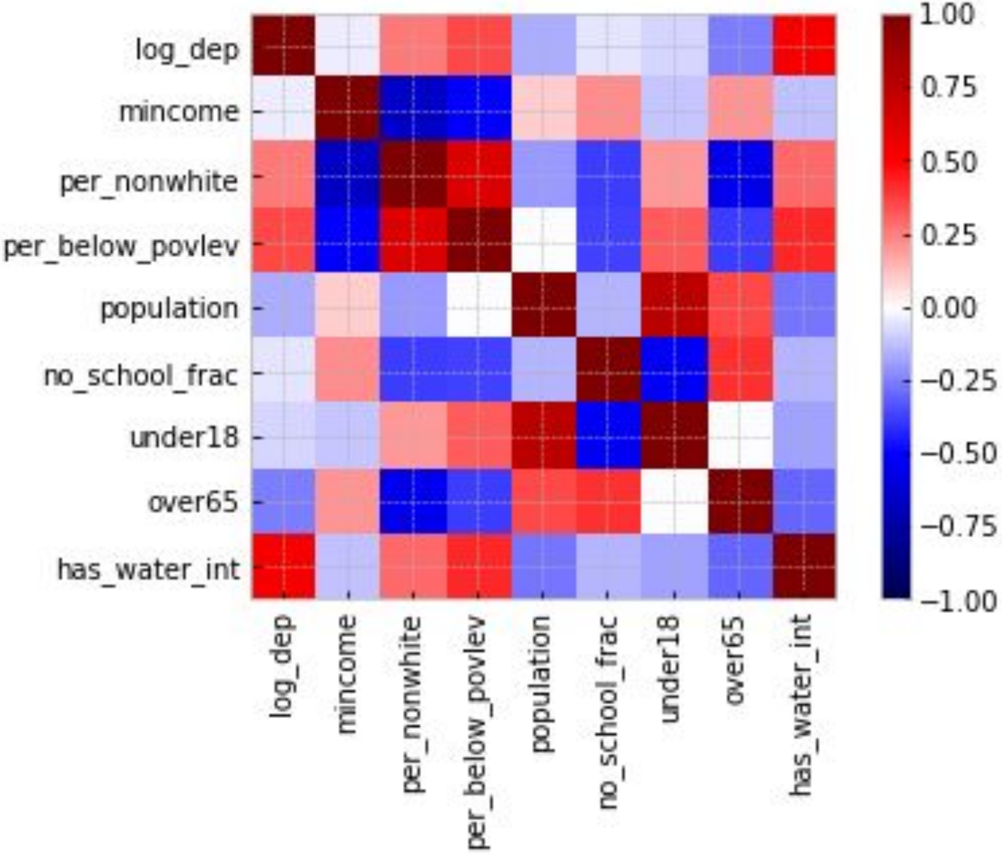


Fig 5. Correlation between features

The correlation between the 1% annual chance flood risk and percentage of people not white, and the percentage of the people below poverty line are highly correlated.

5. Analysis Results

5.2. Geospatial analysis result

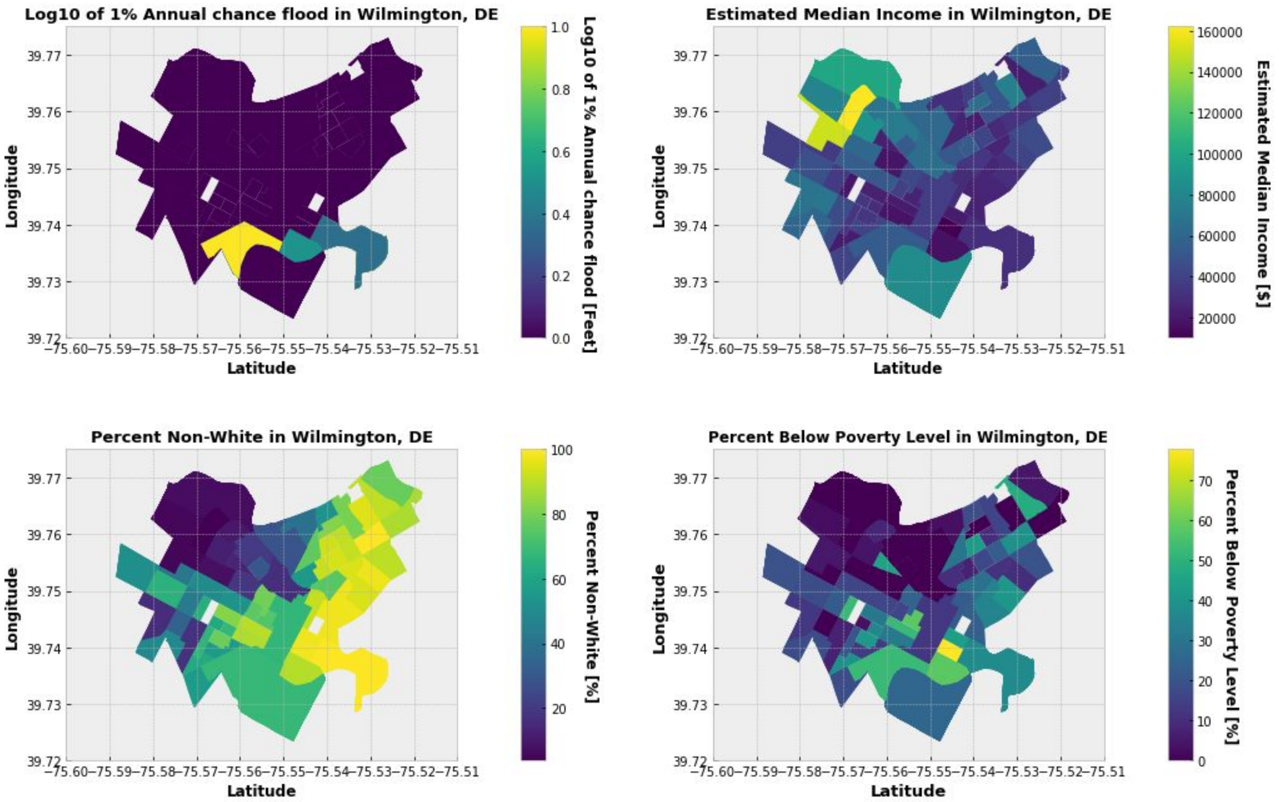


Fig 6-1. Geospatial comparison between flood risk and SED data

Based on the geospatial analysis, people living in the low median income household, people not white, and people under poverty line are living in the flood hazard zone in the City of Wilmington, DE.

5. Analysis Results

5.2. Geospatial analysis result

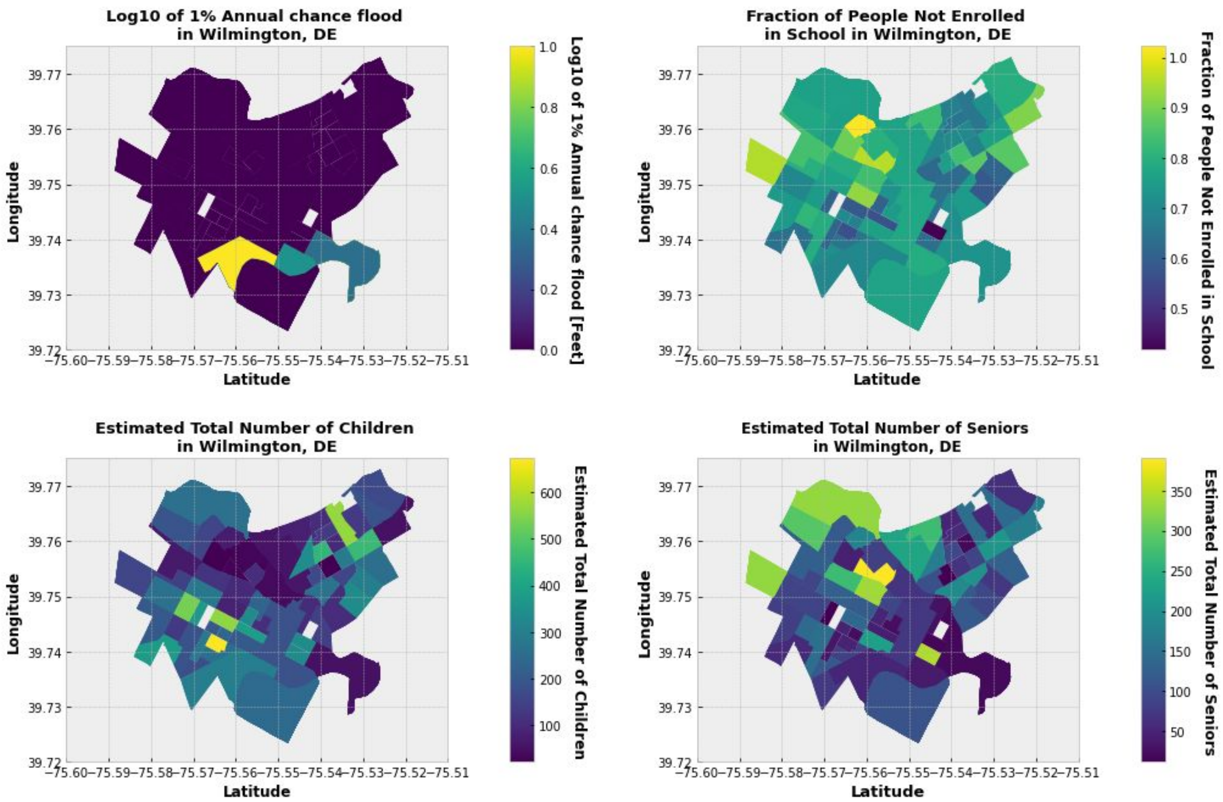


Fig 6-2. Geospatial comparison between flood risk and SED data

Based on the geospatial analysis, people not enrolled in school, children (people under 18 years old), and seniors (people over 65 years old) are not highly close to the flood hazard zone in the City of Wilmington, DE..

5. Analysis Results

5.3. Machine Learning

1) Random Forest Classifier

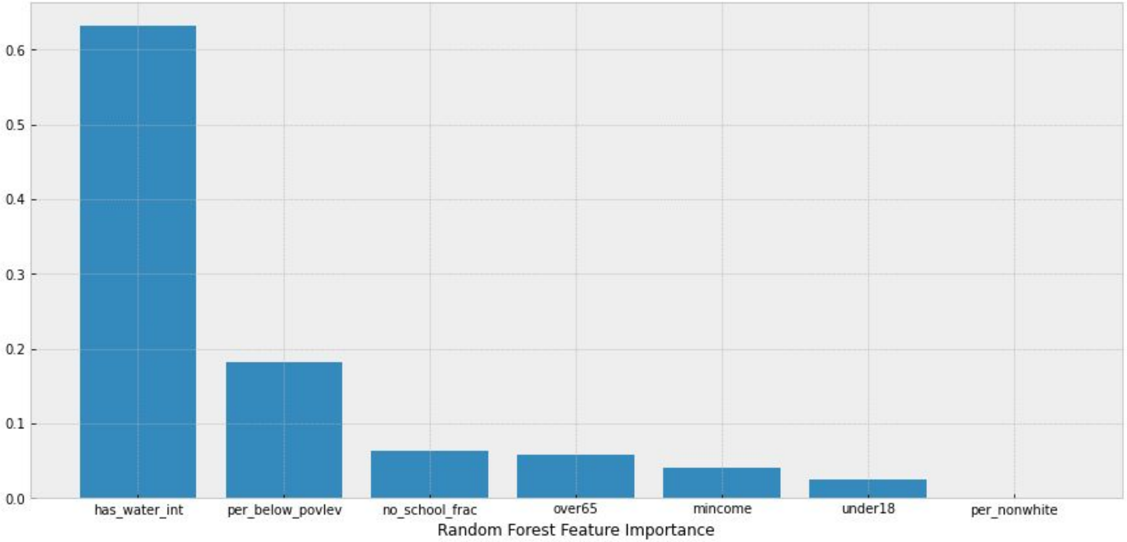
- Target: Presence or absence of flood risk (Binary)
- Objects: Census block groups
- Features:
    - Estimated median household income
    - Percent of population that is not white
    - Percent of people below the poverty line
    - Fraction of people not enrolled in school
    - Estimate number of children (Age under 18)
    - Estimate number of senior (Age 0ver 65)
    - Presence or absence of water areas (where the water areas is over 0.05%)

5. Analysis Results

5.3. Machine Learning

1) Random Forest Classifier - Result

	RFC Model (All features included)
Training accuracy	1.000
Testing accuracy	1.000



5. Analysis Results

5.3. Machine Learning

2) Logistic Regression Classification

- Target: Presence or absence of flood risk (Binary)
- Objects: Census block groups
- Features:
    - Estimated median household income
    - Percent of population that is not white
    - Percent of people below the poverty line
    - Fraction of people not enrolled in school
    - Estimate number of children (Age under 18)
    - Estimate number of senior (Age 0ver 65)
    - Presence or absence of water areas (where the water areas is over 0.05%)

5. Analysis Results

5.3. Machine Learning

2) Logistic Regression Classification - Result

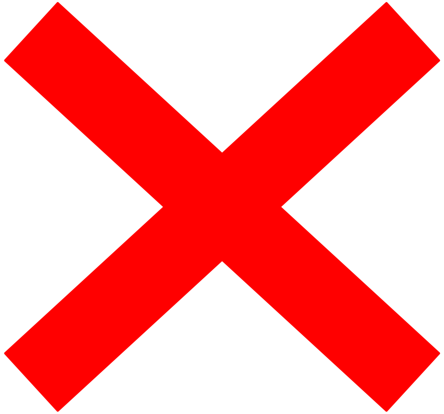
	LRC Model (All features included)
Training accuracy	1.000
Testing accuracy	1.000

Features	**Coefficient**
Percentage of people under poverty line	0.4707
Percentage of people not white	-0.1711
Estimated total number of seniors (Over 65)	-0.0925
Estimated total number of children (Under 18)	-0.046
Presence or absence of water areas (binary)	0.018
Fraction of people not enrolled in school	-0.028
Estimated household median income	0.003

5. Analysis Results

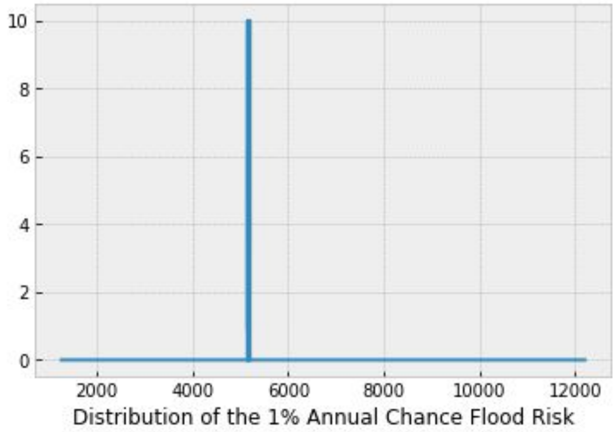
5.3. Machine Learning

1) **Random Forest Classifier**
2) **Logistic Regression Classification**



It is because….

target variable is too easy to be predicted!



What should I do next?

1) Increase feature variables (people not enrolled in school, children, seniors, etc.)
2) Add water areas as input variable
3) **Change the model**

5. Analysis Results

5.3. Machine Learning

3) Random Forest Regressor

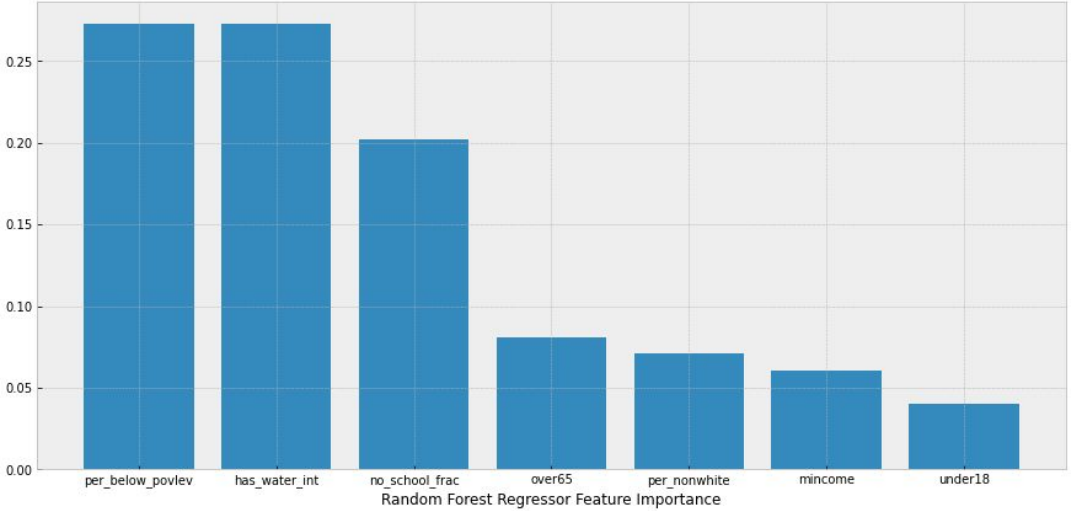
- Target (1): 1% Annual chance flood risk (Multilevel)
- Objects: Census block groups
- Features (7):
  - Estimated median household income
  - Percent of population that is not white
  - Percent of people below the poverty line
  - Fraction of people not enrolled in school
  - Estimate number of children (Age under 18)
  - Estimate number of senior (Age 0ver 65)
  - Presence or absence of water areas (where the water areas is over 0.05%)

5. Analysis Results

5.3. Machine Learning

3) Random Forest Regressor - Result

	MSE (All features included)	MSE/var (All features included)
Training accuracy	0.051	0.206
Testing accuracy	0.049	0.206



6. Discussion and Conclusion

- **The vulnerable people are exposed to flood risk in Wilmington, Delaware.**
    - According to the geospatial analysis result, people in the low median income household, people not-white, and people under poverty line are exposed to the flood risk.

- **During the policy-making decision process, policy makers can consider it as a factor for the policy intervention.**
    - The state government officers, policy makers, and researchers may have to consider the vulnerable people and flood risk in **the local climate adaptation policy.**

6. Discussion and Conclusion

- **Limitation**
  - In terms of the middle size of the urban city, it is limited to get diversified dataset of the flood depth grid from the open data source platform.
  - Since the accuracy of the RFR machine learning model is very low, the model should be more adjusted.
  - The flood risk depth raster data has limited information of how it is collected.

- **Further study**
  - Adjust the machine learning model to increase model accuracy
  - (Expected) case study: New York City

Thank You

Kyungmin Lee

GitHub: @KLeeDE

Email: kmlee@udel.edu