Fall 2018
GEO 4938/6938: GIS & Hazard

Final Project Paper: Integrated Vulnerability Analysis: Case Studies of Tropical Storm Allison and Hurricane Harvey

Xiaoxuan Li (Shawn)

1. Introduction

The concept of vulnerability, first introduced in the 1970s by O'Keefe (O' Keefe et al. 1976), has taken an essential role in hazard assessment research for more than one decade. Vulnerability refers to the disability of human society to withstand adverse impacts. It provides a way to understand the relationship between the environment and social systems. Furthermore, it helps to assess a society's capacity and abilities to handle disasters. However, to understand the characteristics of hazard impact areas more specifically, an integrated vulnerability analysis should be implemented by combining multiple dimensionalities like hazard, economic, social and demographic characteristics. Generally, an integrated vulnerability assessment is commonly used in physical hazard and human domains to develop conceptual frameworks for measuring underlying hazard risks and to make possible decisions for urban planners and other users to minimize hazard impacts. However, most of the previous researches regarding integrated vulnerability analysis have focused on only one case or two cases neglecting vulnerability and hazard exposure changing over time within a specific area. Apparently, by comparing the results of integrated vulnerability analysis based on two cases mentioned above, more detailed social, economic and physical hazard-related information (people movement, hazard hot-spot identification, social and economic pattern changes and hazard exposure distribution variation) can be obtained to make further analysis, which will help to make a more reliable risk assessment.

In this study, a composite method combining hazard exposure and social vulnerability was developed to examine how vulnerability and hazard exposure patterns have changed in Harris County between two tropical cyclone cases: tropical storm Allison (2001) (**Figure 1**) and hurricane Harvey (2017) (**Figure 2**) showing as follows. Also, **Figure 3** shows the total population distribution of Harris county within census tracts in 2000 and 2016. From the figure, it is evident that the total population in Harris County has increased since 2000. Indeed, from the report of Harris County Budget Management estimates, Harris County is one of the fastest population growth counties with 38% growth since 2000. Another tip should be mentioned in **Figure 3** is: the census tracts within Harris county are not the same in 2000 and 2016. The number of census tracts in 2016 is 100 bigger than the number in 2000. In this case, the datasets used in my analysis are all disaggregated into two different census tract areal units. Also, by using NHGIS datasets as Dr. Ash commented during my presentation, it is possible to compare these two cases and do quantitative analysis within the same census tract. Data aggregation would also be a further improvement in this study.

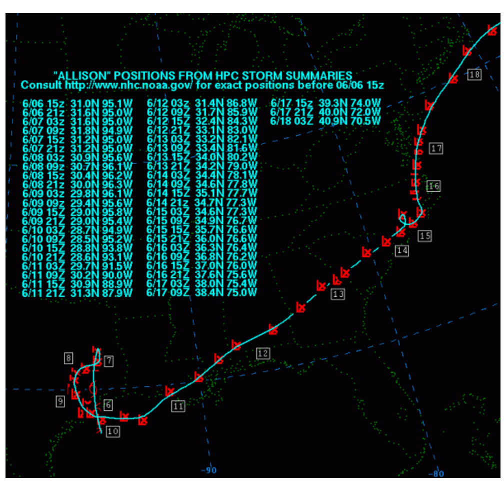


Figure 1. Study case: tropical storm Allison

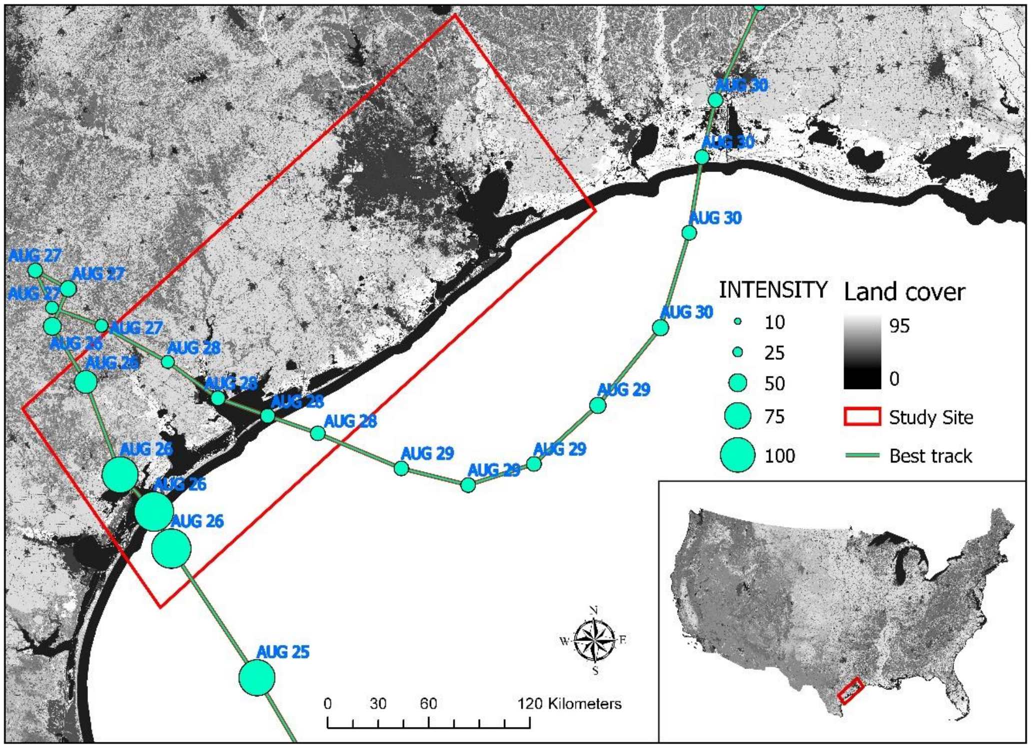


Figure 2. Study case: Hurricane Harvey

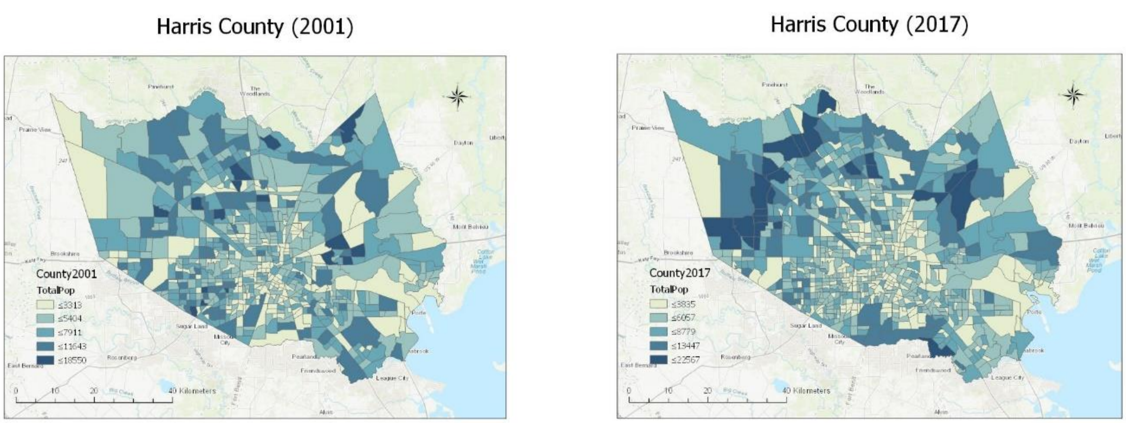


Figure 3. Study site: Harris County

2. Literature Review

The purpose of the literature review is to present some literature on integrated vulnerability assessment, hazard exposure and social vulnerability that have overlapping research concepts and methods with my study. There are three main sections: 1) Demographic data visualization and limitation; 2) SoVI and SVI; 3) Integrated vulnerability analysis; 4) Conceptual framework.

2.1 Demographic data visualization and limitation in vulnerability analysis

Wong et al. (2013) developed a comprehensive procedure to introduce, manipulate and visualize American Community Survey (ACS) data using GIS extensions. He first talked about some challenges existing to the use of ACS data in GIS and the way to handle and analyze these challenges. The biggest challenge using ACS data in GIS is to offer high accuracy of estimates to the users. By applying for multiple GIS-based extensions, Wong and his co-authors make it possible to collect, process and visualize the ACS data. The ArcGIS extensions successfully tackle some challenges mentioned above. In my opinion, Wong has done an excellent job examining the possibility of using popular survey data to analyze demography issues. He also has a deep sight in the potential and limitations of processing multiple population-based survey data. From this paper, I learned some basic characteristics of demographic data and some limitations while mapping them in GIS. Wong's study helps me deal with my 2016 ACS data.

2.2 Evaluation of SoVI and SVI (Cutter et al. 2003, Flanagan et al. 2011)

Both Dr. Cutter's and Dr. Flanagan's papers use factor analytic approaches to calculate the social vulnerability index. The social vulnerability index in these two papers is expected to help government agencies ensure the safety and high-quality life of their residents. Also, the index can be combined with hazard frequency data and economic losses to examine the cause and effect that triggers huge dollar losses. As mentioned above, for both papers, similar social vulnerability index calculation is employed. However, the methods are sort of different based on their models. For Susan's paper, their SoVI calculation model is originated from the hazard-place model of vulnerability modified from Susan's work in 1996. Specifically, in this article, they examine only the social vulnerability portion of the model. When it comes to Barry's paper, he mainly introduced an essential disaster management risk model based on SVI scores with the sake of improving disaster management cycle. I think these two papers all evaluate social vulnerability process from different perspectives. They successfully examined the possibility of applying social vulnerability to the model of hazard vulnerability assessment based on divergent datasets and spatial scales. I think further researches about social vulnerability should be focusing on combining the social vulnerability data with other data to make a complete risk model for understanding and predicting disaster event. I choose SoVI for my primary vulnerability index, and I have achieved to combine SoVI scores with physical hazard exposure.

2.3 Integrated vulnerability analysis (Koks et al. 2015)

Koks combined physical and social vulnerability in risk assessment studies for a comprehensive study of the feasibility of risk reduction strategies in flood risk management (FRM). In his paper, he combined hazard data and social & physical vulnerability indices with high-quality demographic data name PC6, which provides an explicit depiction of the population at the zip code level. He applied the composition and creation of the social vulnerability index (SVI) to his analysis. Then, the SVI indices were overlapped with flood hazard data to capture the potential risk area. A cluster analysis was also employed to differentiate vulnerable groups, which can meet specific requirements within different regions. The result shows promising spatial cluster patterns emphasizing a robust spatial variation between the vulnerability characteristics of people. The risk assessment presented in this paper helped me understand the concept of integrated vulnerability analysis and eventually made me choose integrated vulnerability analysis as my final topic.

2.4 Conceptual framework for vulnerability analysis (Frazier et al. 2014)

Dr. Frazier developed a Spatially Explicit Resilience Vulnerability (SERV) model at sub-county in Sarasota County. Three domains were included in this analysis: exposure, sensitivity and adaptive capacity. The final vulnerability score equals the combination of hazard exposure and sensitivity subtracted by adaptive capacity. There are many highlights in this paper: many indicators were considered for both sensitivity and adaptive capacity domains; good mapping depiction for vulnerability and all sub-domains (hazard exposure, sensitivity, and adaptive capacity); sub-county level data aggregation; very detailed PCA process and excellent weighting scheme. To be honest, the conceptual framework of my vulnerability analysis was mostly inspired by Dr. Frazier’s work.

3. Data

To accomplish more precisely vulnerability analysis, multiple types of datasets were used in this study (**Table 1**).

Table 1. Datasets description

Datasets Name	Data Resources	Type/Resolution	Usage
NEXRAD Level II Radar Datasets	weather.gov	Radar Reflectivity 500m	2D Rainfall data
Shuttle Radar Topography Mission Digital Elevation Models	cr.usgs.gov	Raster 10m	Watershed generation/ Surface volume calculation
Landcover dataset	water.usgs.gov	Polygon	Manning coefficient variables extraction
Gage station observations	harriscountyfws.org	Point Survey level	1D Rainfall observation for validation
2000 socioeconomic data	American Community Survey	Excel Census-tract level	SoVI indicators for Allison case
2016 socioeconomic data	American Community Survey	Excel Census-tract level	Sovi indicators for Harvey case

County boundary dataset	gis-txdot.opendata.arcgis.com	Polygon	Census-tract disaggregation

3.1 Primary Data

NEXRAD Level II Radar Dataset is a primary dataset for rainfall estimation and flood simulation in this study. According to the definition of National Weather Service, the Weather Surveillance Radar-1988 Doppler (WSR-88D), also known as "NEXRAD," is a pulsed Doppler weather radar deployed more than 150 covering the United States to measure meteorological and hydrological phenomena. Radar reflectivity information was derived from NEXRAD Level II data using python & C standalone applications distributed by Dr. Tang and Dr. Matyas in Hurricane Research Group. The temporal resolution is 15 minutes for Allison case (June 5~9 in 2001), and Harvey case (August 25~29 in 2017) and the spatial resolution is 500 m in Harris County.

Shuttle Radar Topography Mission (SRTM) Digital Elevation Model is a primary dataset for watershed generation and surface volume calculation. The spatial resolution of DEM is 10 m.

Socioeconomic datasets are original datasets for calculating SoVI scores. Typically, there are four significant domains: 1) Human domain; 2) Financial domain; 3) Household domain; 4) Education domain. (**Table 2**)

Table 2. Domains and indicators in socioeconomic datasets

Human domain	TotalPop
	PFemale
	PAge14
	PAge65
	Black
	Hispanic
	Asian
Financial domain	PCI (+)
	PUnemployment
	QPoverty
	MedianIncome (+)
	QRich20K (+)
Household domain	QRenter
	QBlackAlone
	Qmobile
	QDense5
	QUNOCCHU
	QFHH
Education domain	QEducation
	QDisability
	QEnglish

For indicator in each domain:

1) Human domain:

Indicator Name	Description (for both Allison & Harvey case)
TotalPop	Total population
PFemale	Percent Female
PAge14	Percent Age smaller than 14 years old
PAge65	Percent Age greater than 65 years old
Black	Black population
Hispanic	Hispanic population
Asian	Asian population

2) Financial domain

Indicator Name	Description (for both Allison & Harvey case)
PCI (+)	Per capita income (+)
PUnemployment	Percent unemployment population
QPoverty	Percent poverty population
MedianIncome (+)	Median income (+)
QRich20K (+)	Percent income greater than 20,000 dollars (+)

3) Household domain

Indicator Name	Description (for both Allison & Harvey case)
QRenter	Percent renter
QBlackAlone	Percent black woman who live alone
Qmobile	Percent mobile houses
QDense5	Percent more than 5 people living in one unit
QUNOCCHU	Percent unoccupied housing unit
QFHH	Percent Families with female-head households with no spouse

4) Education domain

Indicator Name	Description (for both Allison & Harvey case)
QEducation	Percent education is at or smaller than elementary level
QDisability	Percent disability
QEnglish	Percent English deficiency

3.2 Ancillary Data

County boundary dataset was used to clip features derived from other datasets like DEM, Radar raster file, river boundary, and watershed. Count boundary dataset, along with census tract boundary dataset, were collected from Texas County boundaries website.

Landcover dataset was used to extract Manning's coefficient values from land use types defined by USGS. The Manning's values then inputted into flow accumulation as a type of weighting raster for estimating the flooding flow volume. Landcover dataset can be obtained in USGS website.

Gage station observations were collected from Harris County Flood Warning System for validating the rainfall outputs from radar reflectivity transformation. The resolution of gage station observations is at a survey level (two decimal point accuracy), which means that they are the best datasets used to estimate rainfall during hurricane season.

4. Methods (Figure 6)

4.1 Z-R transformation

The NWS has a very limit number of gage stations across the United States, which makes it challenging to estimate rainfall using gage station data solely. Basically, radar reflectivity value Z equals the number of drops n_i of a certain diameter D_i multiply by drop diameter D_i (**Equation 1**).

$$Z = \sum n_i \times D_i^6 \quad (1)$$

However, we cannot directly obtain the drop size and its amount using radar. In this case, previous researchers (Marshall et al. 1947) developed the Z-R relationship (**Equation 2**) to address this problem by correlating radar reflectivity values (Z) with rainfall rates (R).

$$Z_e = \frac{P_r \times R^2}{const} \quad (2)$$

Where Z_e is the equivalent reflectivity, P_r is power returned, and R is rainfall rate.

When we try to use a log operation to make Z_e function well in **Equation 3**,

$$dBZ_e = 10 \times \log Z_e \quad (3)$$

We finally got an empirical relationship to estimate rainfall rate using transformed Z-R relationship in Equation 4,

$$Z_e = a \times R^b \quad (4)$$

Where a and b are all empirical variables. According to NOAA's Z-R relationship table, for hurricane season case, we define that variable a equals to 250 and variable b equals to 1.2 (Equation 5).

$$Z_e = 250 \times R^{1.2} \quad (4)$$

After theoretically find this Z-R relationship, thousands of radar reflectivity files for Allison case, and Harvey case were obtained using R & C standalone application. KHGX radar station records were collected every 15 minutes. Then, a transformation from radar reflectivity (dBZ) to 5-day total rainfall (mm) was achieved using Python code showing below.

```
import arcpy
import os
from arcpy.sa import *
out = 0
arcpy.env.workspace = r"E:\radar\radarcode\radar-grid-release\Allison11"
ras_names = arcpy.ListRasters()
```

```
for ras_name in ras_names:
    ras = arcpy.Raster(os.path.join(arcpy.env.workspace, ras_name))
    outCon = Con(IsNull(ras), 0, ras)
    out = out + (10**(outCon / 10) / 250)**(1 / 1.2) * 0.25
out.save("final")
```

4.2 Watershed delineation

The Arc Hydro Tool was used to delineate the watershed in Harris County. While generating flow accumulation, a Manning's coefficient raster was applied for a more accurate water volume in the specific basin. After this, a watershed was generated in Harris County. The overlapping DEM raster file will be transformed to points and fit into each basin (sub-watershed). Also, for each basin, using a spatial join, the amount of water volume can also be obtained.

4.3 Surface volume

From the definition of ArcGIS Help, surface Volume calculates the projected area, surface area, and volume of a surface relative to a given base height, or reference plane. **Figure 5** shows what surface volume analysis is operated. Basically, we need at least two inputs for estimating the flood height:

The first one is a high-resolution DEM raster surface split by basins in Harris County; another one is the relative reference plane. The idea of obtaining flood plane is: by using python code, hundred of planes will be looped to calculate the volume below the plane. Among these surface volumes, there must be a volume that is closest to the rainfall volume in the specific basin. In this case, it is able to find the flood height. Then, for each basin, any DEM points whose values are smaller than the value of flood height calculated would be considered as a flooded DEM point. Finally, all DEM points will be merged and export as raster file with a spatial resolution of 10 m and output as flood zone map. The python code used to operate surface volume is showing below.

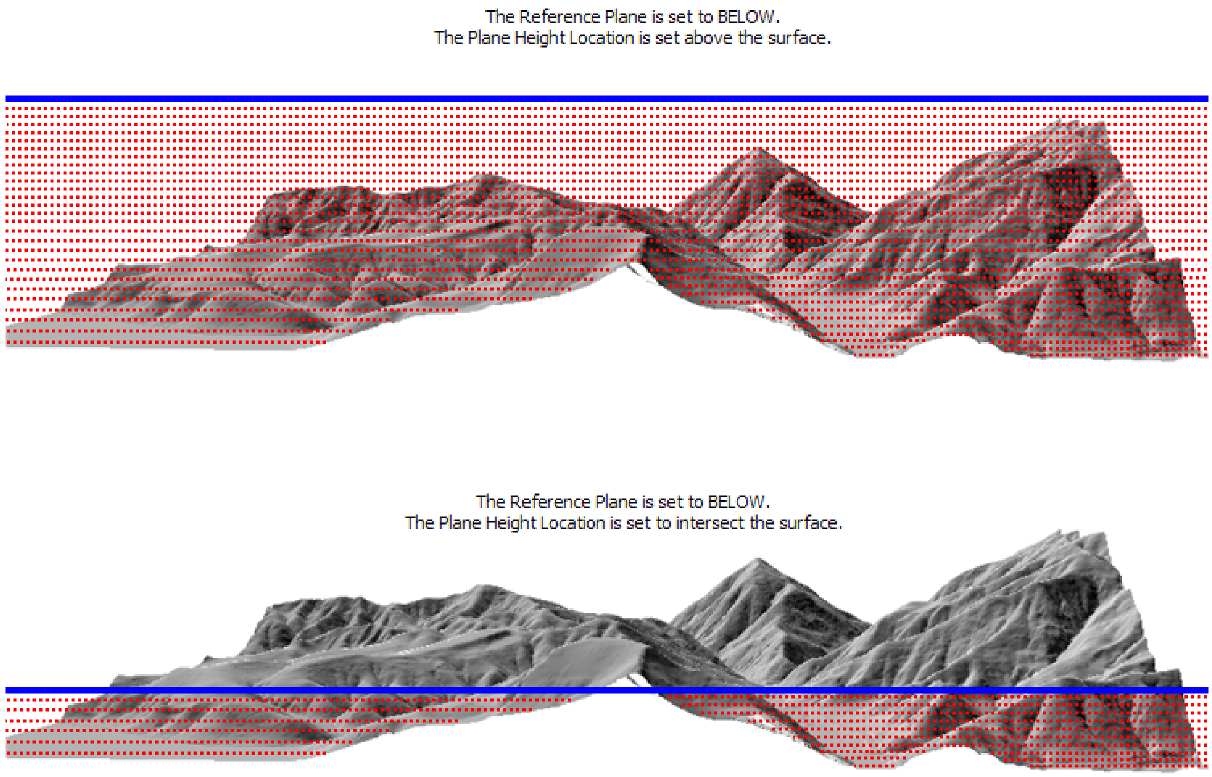


Figure 5. The map interpretation of surface volume mechanism

Surface volume main body

```
import arcpy
from tkinter.filedialog import askdirectory
import os
import sys
arcpy.env.workspace = r"E:\temp\NewIdea\Hazard\final\ArcGIS\SurfaceVolume\Stat"
raster = arcpy.Raster("DEM.tif")
b = 48
for i in range(100):
    arcpy.MakeFeatureLayer_management("Basin.shp", "layer" + str(i), ' "FID" = ' + str(i))
    arcpy.Clip_management(raster, "#", "clip" + str(i) +" .tif","layer" + str(i), "0", "ClippingGeometry")
    b = b - 1
    arcpy.SurfaceVolume_3d("clip.tif", "Surface" + str(i) + ".txt", reference_plane="Below", base_z= b)
```

Extract and transform surface volume information into text

```
Folder = askdirectory()
MaxName = os.path.join(Folder, "Surface"+".txt")
if os.path.exists(MaxName):
    os.remove(MaxName)
d = []
for filename in os.listdir(Folder):
    if filename.endswith(".txt"):
        a = 0
        b = []
```

```
        FileName = os.path.join(Folder, filename)
        with open(FileName) as datafile:
            for i, line in enumerate(datafile):
                for Spline in datafile:
                    a = str(filename) + " " + Spline.split()[2] + " " + Spline.split()[7]
                    b.append(a)
        c = str(max(b))
        d.append(c)
with open(MaxName, 'w') as f:
    for item in d:
        print(item)
        f.write("{}\n".format(item))
sys.exit()
```

Point extraction code:

```
for i in range(17):
    arcpy.MakeFeatureLayer_management("Basin.shp", "layer" + str(i), ' "FID" = ' + str(i))
    arcpy.Clip_analysis("Point100_Clip_Project.shp", "layer" + str(i), "Allison" + str(i) + ".shp")
```

4.4 Principal Component Analysis (PCA)

To determine to dominate sensitivity indicators for Harris County, a PCA in R was conducted on the list of social vulnerability indicators based on the SoVI standard. PCA is a data-reduction technique that identifies groups of variables that are inter-correlated and reduces the number of variables in the analysis (Johnston, R. J. 1978). By applying PCA, the number of social vulnerability indicators was reduced from 21 to 15 for two cases. Furthermore, based on the factors calculated from PCA (eigenvalue greater than 1, factor value greater than 0.5 or smaller than -0.5), 6 (5) profiles were determined for Allison (Harvey) case. Based on these profiles, social vulnerability characteristics of each census tract area were examined. The R code below shows the process of PCA.

```
library(psych)

library(xlsx)

data<-read.table("E:\\temp\\NewIdea\\Hazard\\final\\Social\\PCA_2017.csv", sep=",", header=T)

a <- scale(data)

data.pca <- principal(a, nfactors = 10, residuals = FALSE,rotate="varimax",n.obs=NA, covar=FALSE,

     scores=TRUE,missing=FALSE,impute="median",oblique.scores=TRUE,method="regression")

b <- data.pca$values

c <- data.pca$scores

print(b)

print(c)

print(data.pca)

summary(data.pca)
```

```
write.table(c, "C:\\Users\\lxiao\\Desktop\\lab9.csv", sep="\t")
```

4.5 Integration process

In this part, flood zone map was generated based on flood height and population density map was created derived from the population information in socioeconomic data. Then, flood zone map and population density map were both classified into three categories: low, medium and high and combined into one map similarly to bivariate map showing the hazard exposure distribution. For social vulnerability, after assessing each profile in census tract area, social vulnerability scores were also classified into three categories based on standard deviation. In the end, an integrated vulnerability map was created by combining hazard exposure and social vulnerability together.

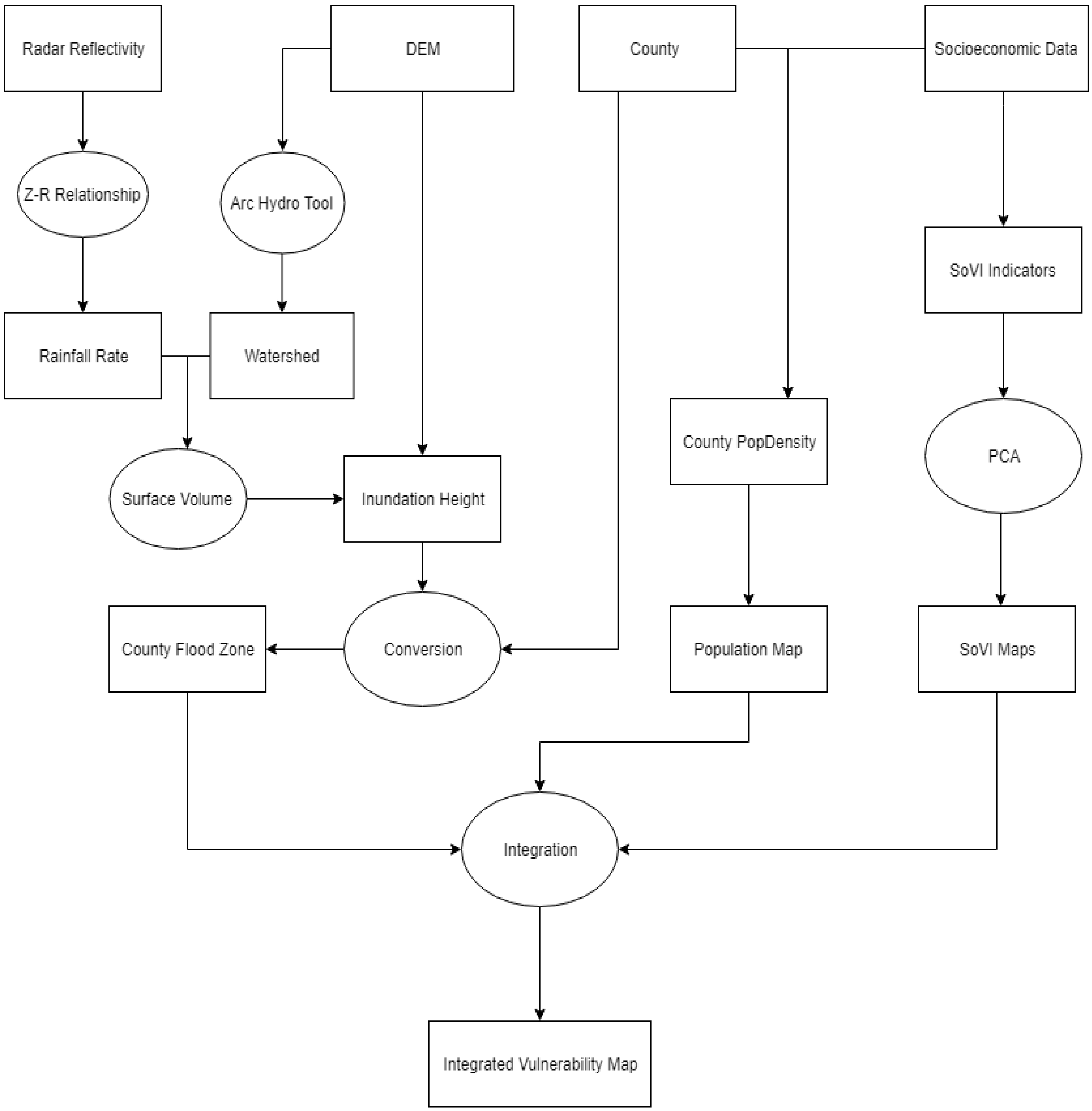


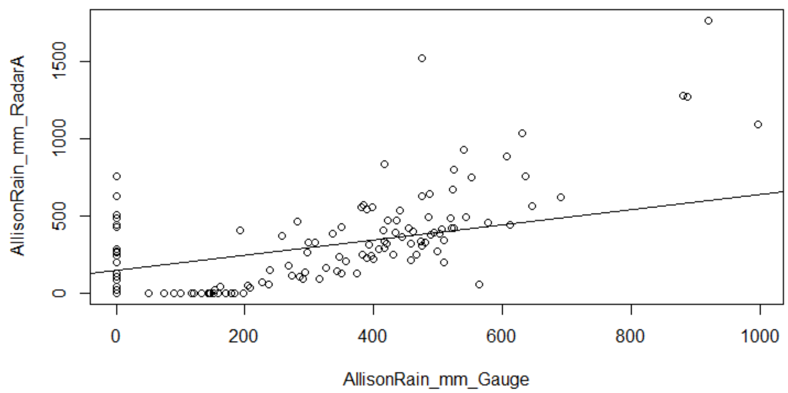
Figure 6. Flowchart of methodology developed

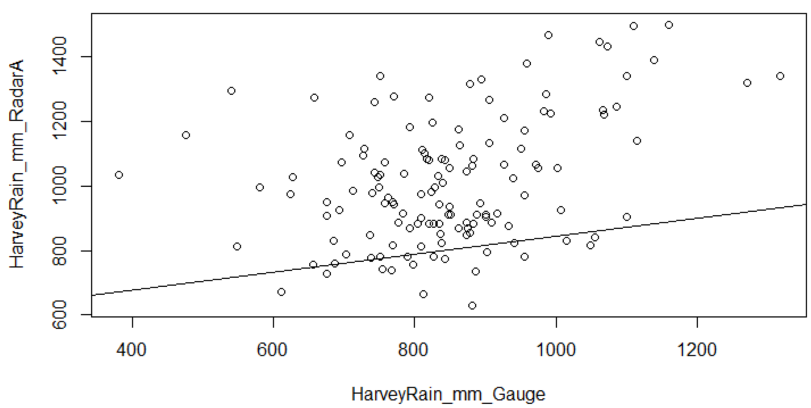
5. Results

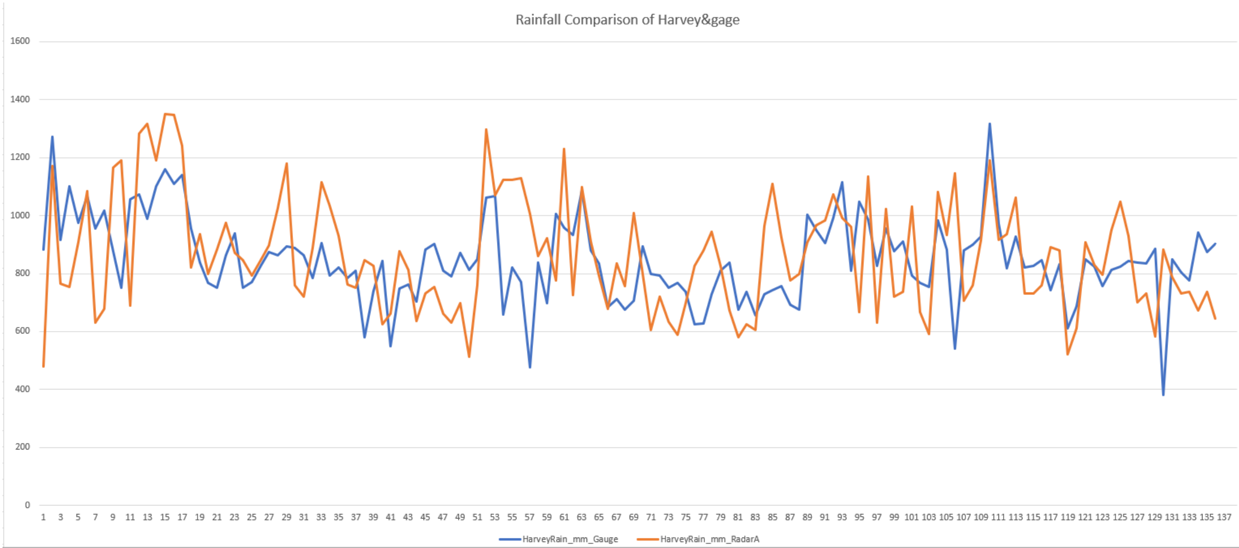
5.1 Rainfall data validation

In this part, a linear regression model was applied to both cases. From **Table 3**, it seems that Allison's fit line is more "significant." However, while the datasets were put into scatterplots in **Figure 7**, it seems that Harvey's curve line is more "accurate." From my point of view, I would conclude that the linear regression model results were primarily influenced by the rainfall value. For Harvey case, almost all the rainfall values are 500~800 greater than that of Allison case. This is also the reason why the accuracy of Harvey case cannot increase significantly even though I have applied data adjustment. Despite the linear slope value, when we talk about the significance of these two fit lines made in R, it is confident to say they are all significant because the p-value in each case is all smaller than 0.05. Furthermore, for Allison case, the Pearson's correlation reached nearly 0.7, which indicates that the data in Allison case is relatively more consistent with gage station data.

Table 3. Linear regression models for Allison & gage and Harvey & gage







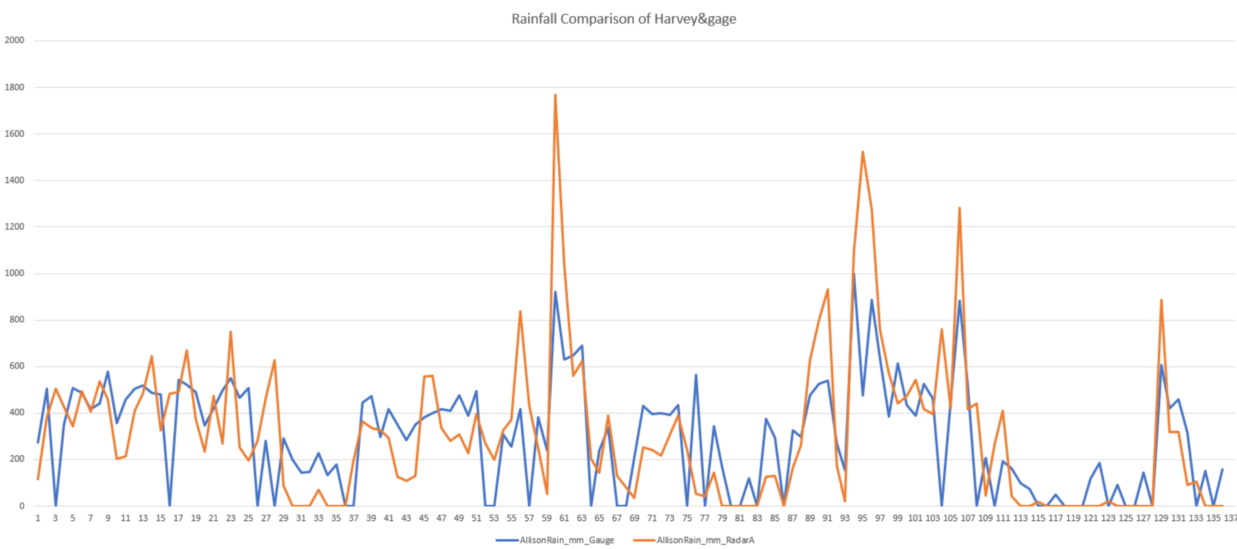


Figure 7. Rainfall comparison of Allison & gage and Harvey & gage

```
Call:
lm(formula = Data$HarveyRain_mm_Gauge ~ Data$HarveyRain_mm_RadarA)

Residuals:
    Min      1Q  Median      3Q     Max 
-471.86  -74.67    2.01   75.92  380.60 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               565.23404   61.41859   9.203 6.10e-16 ***
Data$HarveyRain_mm_RadarA   0.27746    0.05946   4.666 7.34e-06 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 134.9 on 134 degrees of freedom
Multiple R-squared:  0.1398,	Adjusted R-squared:  0.1334 
F-statistic: 21.77 on 1 and 134 DF,  p-value: 7.342e-06
```

```
Call:
lm(formula = Data$AllisonRain_mm_Gauge ~ Data$AllisonRain_mm_RadarA)

Residuals:
    Min      1Q  Median      3Q     Max 
-522.12 -144.00   23.88  120.49  392.35 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)    
(Intercept)                144.93167   19.97281   7.256 2.89e-11 ***
Data$AllisonRain_mm_RadarA   0.49623    0.04468  11.105  < 2e-16 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 166.8 on 134 degrees of freedom
Multiple R-squared:  0.4793,	Adjusted R-squared:  0.4754 
F-statistic: 123.3 on 1 and 134 DF,  p-value: < 2.2e-16
```

```
> res <- cor(Data)
> round(res, 2)
                      OBJECTID HarveyRain_mm_Gauge HarveyRain_mm_RadarA AllisonRain_mm_Gauge AllisonRain_mm_RadarA
OBJECTID                  1.00               -0.24                -0.22                -0.25                 -0.15
HarveyRain_mm_Gauge      -0.24                1.00                 0.37                 0.15                  0.28
HarveyRain_mm_RadarA     -0.22                0.37                 1.00                 0.10                  0.19
AllisonRain_mm_Gauge     -0.25                0.15                 0.10                 1.00                  0.69
AllisonRain_mm_RadarA    -0.15                0.28                 0.19                 0.69                  1.00
HarveyRain_mm_RadarAAAA  -0.22                0.37                 1.00                 0.10                  0.19
                      HarveyRain_mm_RadarAAAA
OBJECTID                                -0.22
HarveyRain_mm_Gauge                      0.37
HarveyRain_mm_RadarA                     1.00
AllisonRain_mm_Gauge                     0.10
AllisonRain_mm_RadarA                    0.19
HarveyRain_mm_RadarAAAA                  1.00
```

Linear regression statistics & Pearson’s correlation for Allison & gage and Harvey & gage

After validating the total rainfall data, we compared the rainfall distribution with the map captured from NOAA website (**Figure 8**). Unfortunately, there are no available rainfall maps for either Allison case or Harvey case. In this case, there is no way to do the statistical analysis but just have a simple comparison. In Allison’s map, it is obvious to find that in northeastern Houston city, the total rainfall record has peaked at 35 inches, which equals approximately 900 mm. Compared with my map, it seems that my records are much bigger than 900 mm. The reason is that, in my Allison’s rainfall data, there are three outliers whose total rainfall records are extremely high. Besides these three points, others are all below 900 mm. In this case, I would say the error is acceptable.

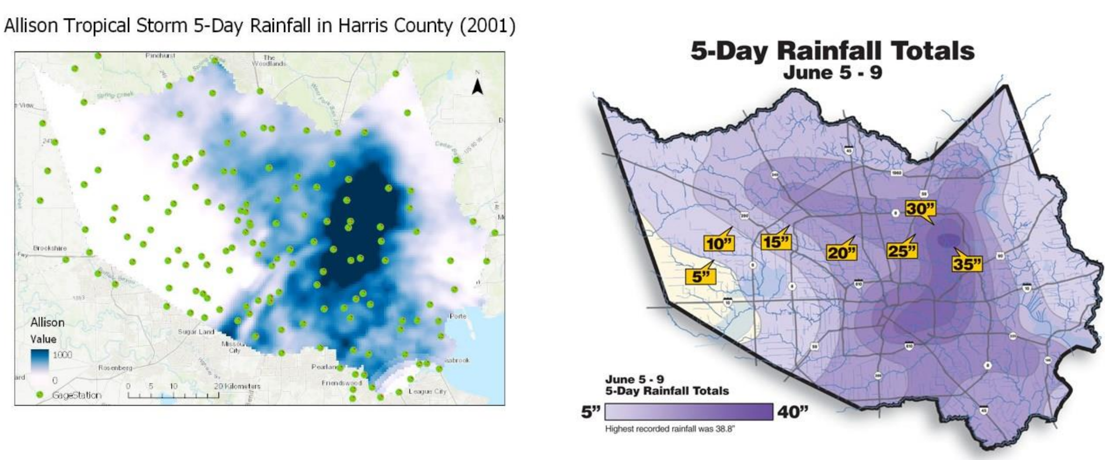


Figure 8. 5-day total rainfall derived from Radar datasets for Allison case

For Harvey case, the NOAA’s map on the left side of **Figure 9** cannot be used to compare directly because this map is based on 4-day rainfall. However, even though this is a 4-day rainfall map, some cities in the half southern part of Harris county still have total rainfall records greater than 30 inches. According to the NPR report titled “Harvey The “Most” Significant Tropical Cyclone Rainfall Event in the US History” wrote by Merrit Kennedy, the maximum rainfall record reached 60.58 inches, which equals to 1524 mm. Compared with my map, the maximum records between these maps are similar enough (only 24 mm difference). Also, in this report, the number of official river gauge stations in Harris county broke the previous tropical rainfall records, which indicates that hurricane Harvey has brought a devastating rainfall to this poor county. In fact, floods induced by Harvey rainfall has already been considered as a 1000-year flood.

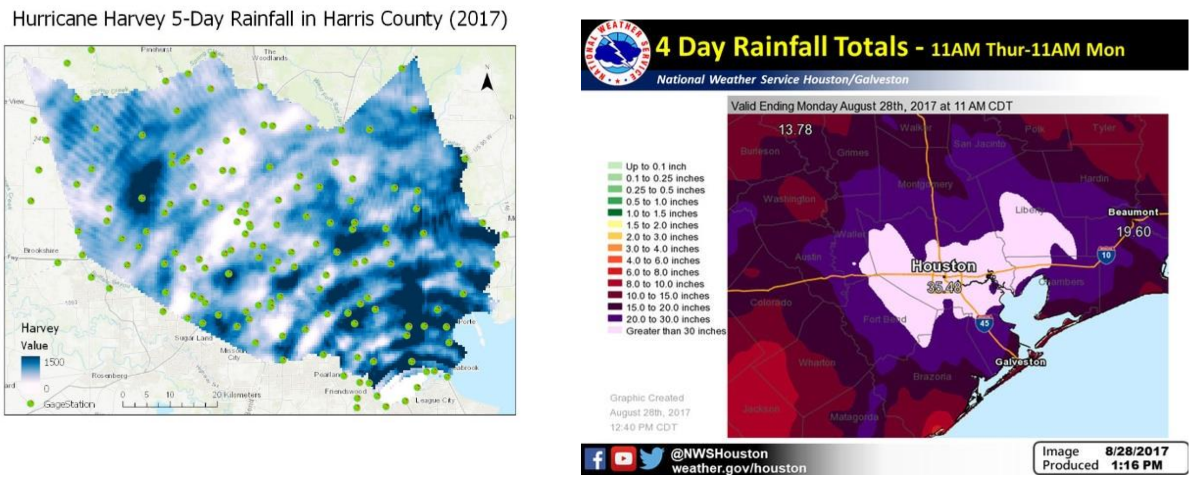


Figure 9. 5-day total rainfall derived from Radar datasets for Harvey case

5.2 Hazard exposure map

5.2.1 Watershed Delineation

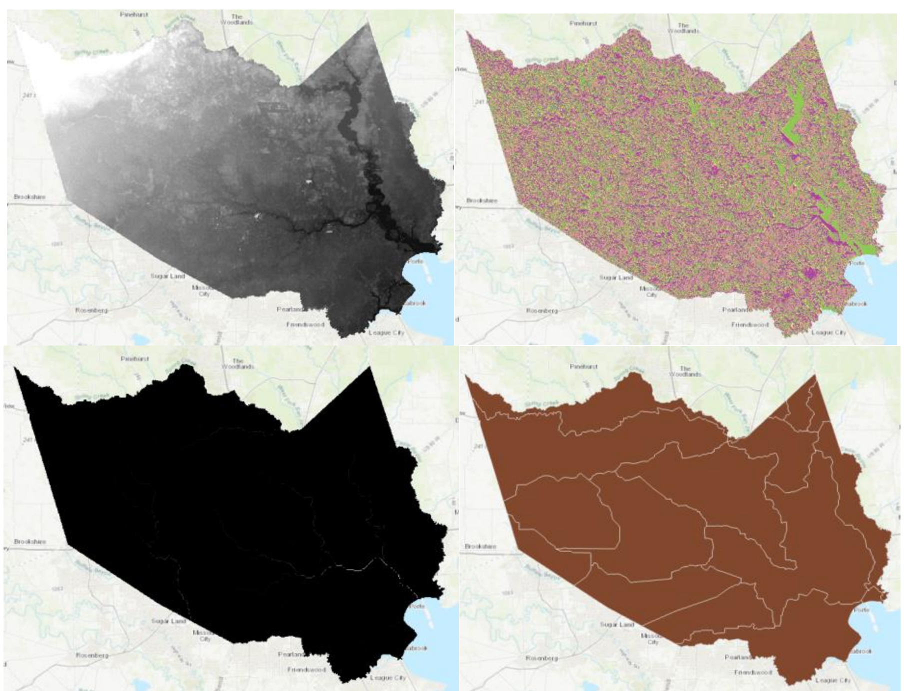


Figure 10. DEM Fill (left top), Flow Direction (right top), Flow Accumulation (left bottom) and Watershed basin (right bottom)

Harris County Watershed Basin

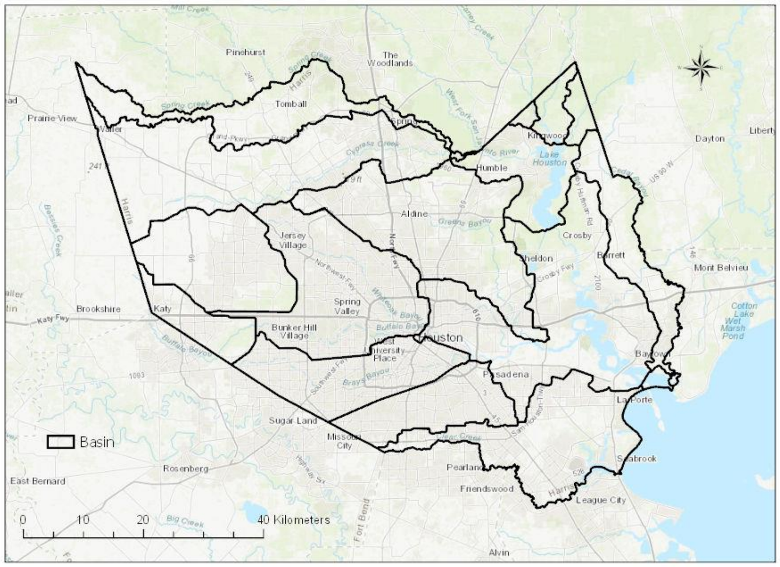


Figure 11. Watershed (Basin)

5.2.2 Z-R statistics

Table 4 shows the result of Z-R transformation and rainfall volume in each basin derived from 5.2.1. The rainfall raster file calculated from python code in 4.1 was transformed into point whose cell size is about 500 m. Then, for each basin, spatial join was applied to the basin and all rainfall points inside it. In this case, the total amount of rainfall would be calculated in each basin. The average rainfall values were inherently obtained. For Harvey case, it is easy to see that the average rainfall in each basin is much bigger than that of Allison case except basin 0, 10,11 and 12, which are all located in the northeastern of Houston city. In other words, Rainfall-induced by Allison in northeastern Houston city is so high that it can even tie to Harvey's intensity. That may be the reason why its name has been retired even though it is just a tropical storm.

Table 4. Results of ArcHydro process and Z-R transformation

Watershed	Allison_ Average_ Rainfall (mm)	Average_ Harvey_ Rainfall (mm)	Rainfall_ Cell_ Count (500*500)	Watershed_ Volume (m^3)	Allison_ Rainfall_ Volume (m^3)	Harvey_ Rainfall_ Volume (m^3)
0	874.9510309	1120.622423	776	4271470732	169740500	217400750
1	249.8385047	1033.757757	2675	30540095843	167079500	691325500
2	348.890785	986.587884	1172	12143327408	102225000	289070250
3	285.1363636	1070.022727	44	103811823.1	3136500	11770250
4	151.0810811	1099.141892	148	650195658.5	5590000	40668250
5	176.9380531	1077.181416	226	1069571277	9997000	60860750
6	0	993.5622407	482	2996841250	0	119724250

7	3.61416185	1107.940751	1384	10241820527	1250500	383347500
8	343.8656566	933.2535354	1980	27138269068	170213500	461960500
9	489.9750831	1048.761628	1204	15319176698	147482500	315677250
10	984.1357466	1087.643665	884	4293370585	217494000	240369250
11	1165.316467	1023.061178	2174	30695791278	633349500	556033750
12	949.3479152	1200.479836	2926	54412628418	694448000	878151000
13	219.6113161	1265.214022	813	4322183437	44636000	257154750
14	209.3333333	1325.039216	51	24313277.87	2669000	16894250
15	751.5741206	1157.844849	1592	26907353546	299126500	460822250

5.2.3 Surface volume result

In 5.2.2, Rainfall volume for each basin has been obtained using python code. These rainfall volume records will be regarded as references which are essential to determine the flood height. As mentioned in section 4, a surface volume analysis was implemented to calculate the flood height for each basin. In **Table 5**, The lists named "Plane_Height_Maximum," "Plane_Height_Allison," "Plane_Height_Harvey" showing in Table 5 represent the maximum height in the specific basin, the selected plane height for Allison case and the selected plane height for Harvey case, respectively. These height records were all obtained using python code based on watershed characteristics and rainfall volumes.

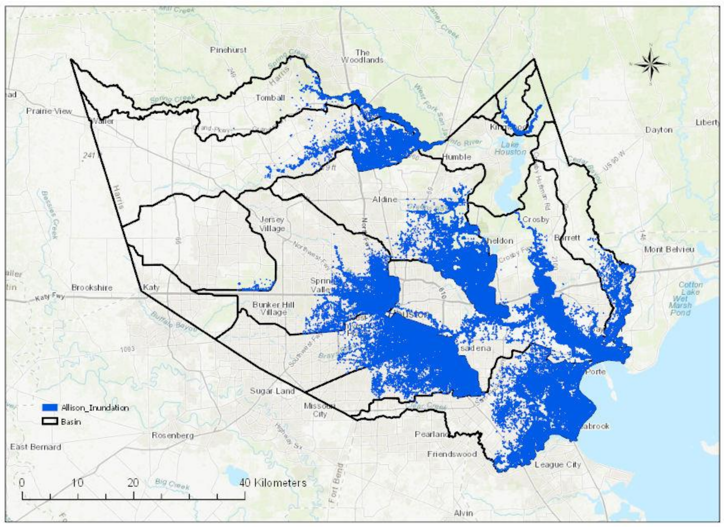
Table 5. The result of the surface volume process

Watershed	Plane_ Height_ Maximum (m)	Plane_ Height_ Allison (m)	Plane_ Height_ Harvey (m)	Watershed_ Volume (m^3)	Allison_ Rainfall_ Volume (m^3)	Harvey_ Rainfall_ Volume (m^3)
0	47	33	36	4271470732	169740500	217400750
1	99	37	42	30540095843	167079500	691325500
2	104	42	48	12143327408	102225000	289070250
3	38	24	27	103811823.1	3136500	11770250
4	51	24	25	650195658.5	5590000	40668250
5	51	20	28	1069571277	9997000	60860750
6	64	0	34	2996841250	0	119724250
7	73	29	40	10241820527	1250500	383347500
8	86	23	26	27138269068	170213500	461960500
9	75	18	21	15319176698	147482500	315677250
10	37	16	16	4293370585	217494000	240369250
11	85	21	20	30695791278	633349500	556033750
12	97	6	6	54412628418	694448000	878151000
13	40	11	15	4322183437	44636000	257154750
14	17	7	14	24313277.87	2669000	16894250
15	83	9	10	26907353546	299126500	460822250

5.2.4 Inundation map

Based on the water height calculated in 5.2.3 section, all DEM points within each basin were evaluated. If their elevation values are not bigger than the water height, these DEM points would be considered as flooded DEM points. Finally, all flooded DEM points were transformed into a raster with a spatial resolution of 10 m, which shows in Figure 12 and 13.

Inundation Map of Allison Case in Harris County



Inundation Map of Harvey Case in Harris County

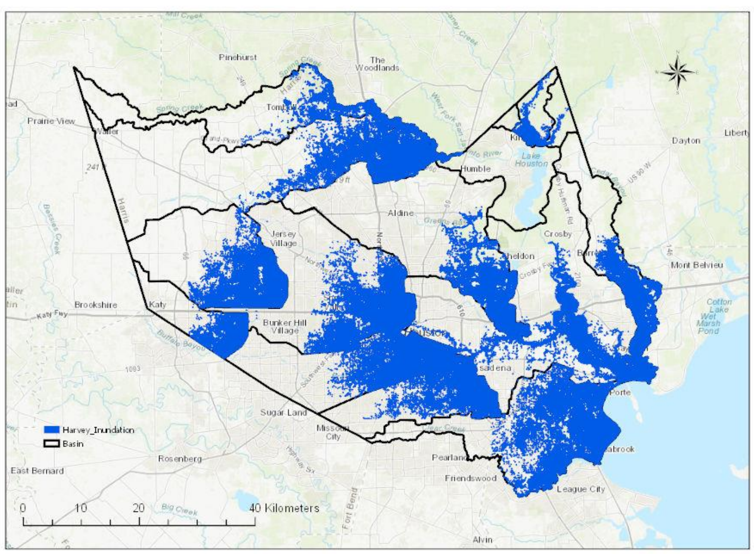
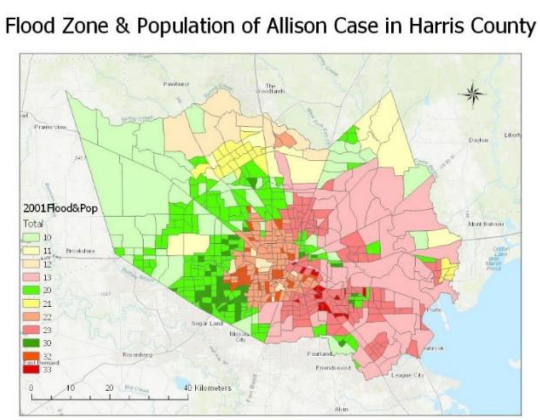


Figure 12, 13 Flood zones in Harris county for both cases

5.2.5 Result: hazard exposure map

Figure 14 and 15 depict the hazard exposure distribution within Harris county for both cases. For those census tract areas that overlap with flood zone would be selected as flooded census tract areas. According to the intensity of flood zone, these flooded areas are classified into three categories in flood zone symbology (Low, Medium and High), which represent low flooding impacts, medium flooding impacts, and high flooding impacts, respectively. Like the classification of the flood zone, population intensity for these census tract areas was also classified into three parts representing low population density, medium population densit, and high population density, respectively. After the symbology process introduced above, hazard exposure information was obtained by combining flood zone maps and population density maps together. In order to maintain the similarity of values and corresponding legends, these two maps have the same symbology classification method and the same symbology colors for each category. The specific legend notations are listed in **Table 6**. Additionally, because the intensity of population and hazard are all based on Allison case when making symbology for Harvey case, the colors would all appear "dark." The reason is, in Harvey case, the population and hazard intensity are all bigger than that of Allison case. In **Figure 14**, it is evident that the cities located in the north and west of Harris county tend to have less possibility of encountering flood impacts. Houston city, on the contrary, is the most vulnerable city that may be impacted by floods. In **Figure 15**, even though Houston city is still the most vulnerable city all over the county, most of the other cities are not in good condition, either. The reason is apparent: hurricane Harvey is just too horrible and devastating. However, there are still some areas that are lucky enough to escape from this catastrophic event due to high elevation. For example, some areas in the southwestern part of Harris County.



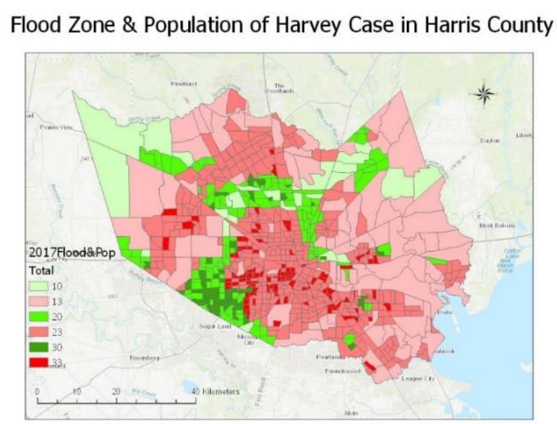


Figure 14, 15. Hazard exposure map combining flood zone and population density (Allison case)

Table 6. Legend notation of figures above

Legend Index	Population density	Flood Intensity
10	Low	No flood
11	Low	Low
12	Low	Medium
13	Low	High

20	Medium	No flood
21	Medium	Low
22	Medium	Medium
23	Medium	High
30	High	No flood
32	High	Medium
33	High	High

5.3 Result: Social vulnerability score calculation

In **Table 7 and 10**, for those eigenvalues greater than 1 would be selected as a factor. In Allison (Harvey) case, the number of factors was 6 (5) in total. In **Table 8 and 11**, all the indicators would be evaluated and interpreted for Allison and Harvey case, respectively. **Table 9 and Table 12** shows top and bottom 5 highest SoVI scores and bottom 5 lowest SoVI scores and their corresponding dominant factors for both cases. **Figure 16 and 17** depict the spatial distribution of SoVI in Harris County for Allison and Harvey case, respectively.

Table 7. Eigenvalue and cumulative variance for Allison case

Index	Eigenvalue	% of Cumulative Variance
1	6.11252080	0.2
2	2.99552433	0.36
3	2.46968049	0.52
4	1.84721986	0.61
5	1.44786620	0.69
6	1.16216057	0.77
7	0.88455300	0.83

Table 8. Factor analysis for Allison case

Indicators	PC1	PC2	PC3	PC4	PC5	PC6
TotalPop	0.10	0.00	0.05	-0.06	0.03	-0.17
PFemale	-0.11	0.06	0.17	-0.03	0.90	0.19
PAge14	0.69	-0.05	0.22	-0.18	0.40	-0.37
PAge65	-0.12	0.18	0.04	-0.03	0.21	0.88

Black	-0.09	-0.18	0.92	0.02	0.08	0.12
Hispanic	0.81	-0.38	-0.25	0.01	-0.19	-0.02
Asian	-0.15	0.07	-0.05	-0.01	0.05	-0.11
PCI(+)	-0.36	0.83	-0.25	0.04	0.02	0.16
PUnemployment	0.12	-0.18	0.17	0.12	0.03	-0.01
QPoverty	0.58	-0.32	0.53	0.15	0.00	0.12
MedianIncome(+)	-0.24	0.86	-0.30	-0.12	0.10	-0.08
QRich20K(+)	-0.07	0.94	-0.08	0.01	-0.02	0.12
QRenter	-0.11	0.05	-0.07	0.93	0.08	-0.09
QBlackAlone	-0.04	-0.14	0.87	-0.02	0.11	-0.02
Qmobile	0.09	-0.10	-0.08	0.08	-0.02	-0.05
QDense5	0.85	-0.17	0.08	-0.21	0.01	-0.08
QUNOCCHU	-0.04	-0.12	0.15	0.89	-0.16	0.11
QFHH	0.25	-0.32	0.64	0.06	0.40	-0.32
QEducation	0.66	0.16	0.15	-0.14	0.41	-0.02
QDisability	0.45	-0.40	0.46	0.13	-0.09	0.45
QEnglish	0.82	-0.25	-0.12	0.08	-0.24	0.02

Profile 1: P14, Hispanic, QDense5, QEnglish

Profile 1 interpretation: for those areas whose dominant profile is profile 1, they may have a more Hispanic population in-house unit. They may have more children smaller than 14 years old. It seems like these clustered people have difficulty communicating using English.

Profile 2 (+): PCI(+), MedianIncome(+), QRich20K(+)

Profile 2 interpretation: this profile is a totally “positive” profile. For those areas who have more factor values in this profile, it is safe to say that more rich people live in these areas. They are rich from all kinds of aspects.

Profile 3: Black, QBlackAlone, QPoverty, QFHH

Profile 3 interpretation: in this profile, people tend to be black women who live alone. People may also have the poor financial condition.

Profile 4: QRenter, QUNOCCHU

Profile 4 interpretation: in this profile, the houses are more likely rental houses. Also, in this area, the unoccupied house rate is still high. For those two situations, they will all increase the vulnerability while occurring floods.

Profile 5: PFemale, QEducation

Profile 5 interpretation: for those areas whose dominant profile is profile 5, people tend to be women. Also, the people living in these areas may not get good education, which may cause some communication issues while occurring a hazard or may not prepare well due to lack of necessary hazard preparedness and mitigation knowledge.

Profile 6: PAge65, QDisability

Profile 6 interpretation: people are more likely old and disabled people in profile 6.

Score = PC1-PC2+PC3+PC4+PC5+PC6

Table 9. Top 5 and Bottom 5 SoVI scores of 2000 census tract areas and dominant factors

OBJECTID	TRAC	PC1	PC2	PC3	PC4	PC5	PC6	SoVI
92	5507	0.170903	-0.53779	0.078398	6.978978	-0.07704	-0.65771	7.031317
11	5552	0.400713	-0.45783	1.870476	0.626915	0.429732	2.398673	6.184343
249	2112	0.673354	-0.35163	3.304389	0.058086	3.365266	-1.83368	5.919042
13	2517	0.478848	-0.49893	2.00513	0.80485	0.275352	1.469826	5.532936
90	5410	-0.38513	-0.35091	1.755063	1.091986	0.760291	1.900398	5.473511
199	2529	-2.99739	-0.77209	0.897166	-2.56778	-10.2331	-3.24398	-17.3729
2	2516	-2.12404	0.346906	1.847066	0.68635	-10.1475	-2.53059	-12.6157
194	2304	-3.4183	-1.85366	-0.15492	-2.19653	-4.5764	-1.8218	-10.3143
23	2513	-2.53391	-1.78937	-0.47371	-2.14067	-4.47441	-2.34909	-10.1824
330	4507	0.983591	7.918559	0.803204	-0.18151	-1.16435	0.408176	-7.06945

Table 10. Eigenvalue and cumulative variance for Harvey case

Index	Eigenvalue	% of Cumulative Variance
1	6.73941215	0.22
2	3.36174639	0.38
3	1.97359926	0.52
4	1.63408362	0.6

5	1.18377354	0.68
6	0.93618568	0.76

Table 11. Factor analysis for Harvey case

Indicators	PC1	PC2	PC3	PC4	PC5
TotalPop	0.01	-0.05	0.16	-0.02	-0.16
PFemale	-0.04	0.15	0.05	-0.11	0.04
PAge14	0.13	0.11	0.72	0.25	-0.31
PAge65	-0.33	-0.05	-0.23	0	0.77
Black	0.15	0.92	-0.06	-0.01	0.07
Hispanic	0.58	-0.3	0.54	0.26	-0.15
Asian	-0.15	-0.12	-0.15	-0.08	-0.11
PCI(+)	-0.84	-0.27	-0.33	-0.08	0.05
PUnemployment	0.3	0.66	0.31	-0.02	0.14
QPoverty	0.51	0.34	0.27	0.37	-0.16
MedianIncome(+)	-0.89	-0.27	-0.07	-0.09	0.05
QRich20K(+)	-0.93	-0.19	-0.08	-0.01	-0.02
QRenter	0.47	0.2	-0.32	0.28	-0.43
QBlackAlone	0.15	0.83	-0.26	0.01	-0.03
QMobile	0.26	-0.06	0.87	0.1	0.04
QDense5	0.1	0.22	-0.02	0.04	0.07
QUNOCCHU	0.09	-0.07	0.16	0.03	-0.02
QFHH	0.49	0.54	0.36	0.2	-0.03
QEducation	0.09	0.04	0.17	0.93	0.03
QDisability	0.26	0.42	-0.03	0.01	0.69
QEnglish	0.45	-0.21	0.4	0.5	-0.21

5 Profiles:

Profile 1: PCI(-), MedianIncome(-), QRich20K(-), Hispanic

Profile 1 interpretation: contrast to the situation in Allison case, this profile is “negative” profile, which will increase the vulnerability because of poverty and Hispanic.

Profile 2: Black, PUnemployment, QBlackAlone, QFHH

Profile 2 interpretation: people are living at a relatively low unemployment area. They may be the black women who lives alone without any spouse.

Profile 3: PAge14, Hispanic, QMobine

Profile 3 interpretation: people seem to be Hispanic. They may have more children. People may also live in mobile houses.

Profile 4: QEducation, QEnglish

Profile 4 interpretation: people in this profile are lack of education and English proficiency.

Profile 5: Page65, QDisability

Profile 5 interpretation: people in this profile may be old and disabled.

Score = PC1+PC2+PC3+PC4+PC5

Table 12. Top 5 and Bottom 5 SoVI scores of 2016 census tract areas and dominant factors

OBJECTID	TRAC	PC1	PC2	PC3	PC4	PC5	SoVI
275	5312	-0.10617	6.293936	2.068121	-0.0872	0.677508	8.84619
56	2319	-0.2707	2.250024	-0.88242	3.164615	3.039151	7.300675
282	5342	-0.01909	3.036203	0.92175	1.216773	1.735146	6.890784
70	3214	-0.18102	3.158901	0.555184	1.106412	1.988715	6.628196
281	5340	-0.35585	2.284371	0.18377	1.36446	3.114021	6.59077
364	3506	-0.52085	-0.87902	-2.37181	-1.48693	-1.24619	-6.50481
566	5422	-1.71621	-0.38792	-1.39902	-0.79013	-2.19262	-6.48589
359	3124	-0.89136	-0.561	-2.33436	-1.03449	-1.33941	-6.16061
362	4306	-1.10642	-1.10447	-1.73695	-0.69585	-1.00908	-5.65276
377	2305	0.523248	-0.71168	-2.26582	-1.52708	-1.58317	-5.56451

From Figure 16 and 17, it is evident that in many areas located in northern Harris County, the SoVI value has reduced significantly in 2017 compared with 2001. In northern Houston city, the SoVI values decreased, two. However, in southern Houston city, the SoVI values increased. Also, in some areas of southeastern Harris County, the SoVI also increased since 2001. While considering high social vulnerability areas, the dominant factors are factor 1 (low age level, Hispanic clustered) and factor 3 (poverty, Black clustered) in northwestern Harris county and factor 6 (disabled old people) and factor 5 (low education level, women clustered) in northeastern Harris County. In 2017 SoVI map, the dominant factors in western Houston city and some areas in southeastern Harris county are mainly factor 2 (high unemployment rate, black).

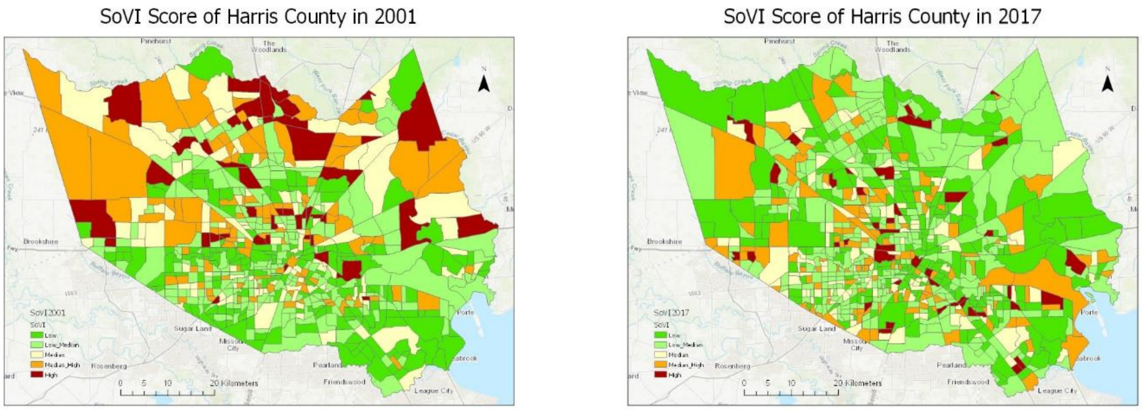


Figure 16, 17 Comparison between SoVI maps of Harris county in 2001 and 2017

5.4 Result: Integration vulnerability analysis (hazard exposure + social vulnerability)

In this section, composite scores were calculator based on **equation 5** below:

$$Composite\ Score = SoVI \times 100 + PopDen \times 10 + F \qquad (4)$$

Where SoVI, PopDen, F represents SoVI, population density and flood impacts reclassified as 1, 2, 3 categories in each census tract area, respectively.

Figure 18 and 19 are the composite maps combining hazard exposure and social vulnerability together. The legend and color of hazard exposure in the maps are all the same as **Figure 14 and 15** (copy and paste).

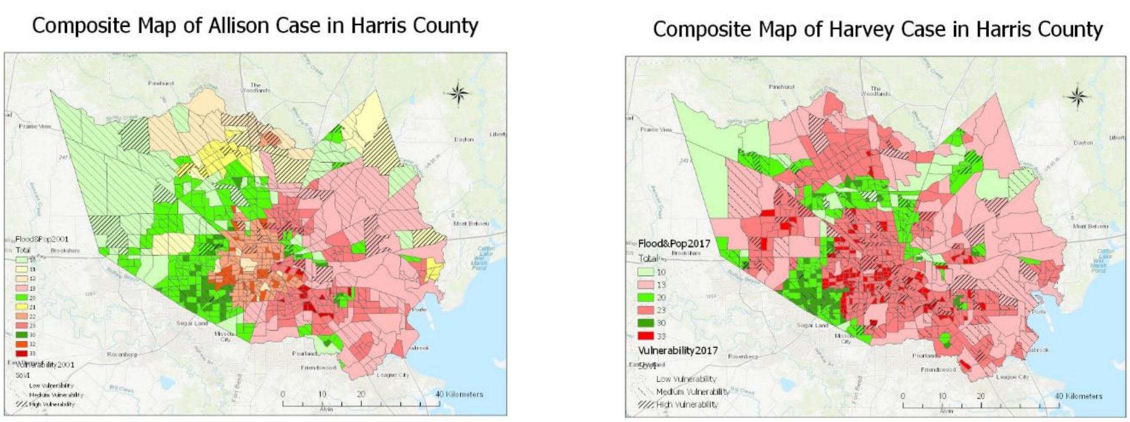


Figure 18,19. Integration map combining hazard exposure and social vulnerability for both cases

Table 13 and 14 listed the detailed top composite scores (most vulnerable) in Harris County for Allison and Harvey cases. Explicitly, in **Figure 20**, the spatial distribution of top 7 composite score areas in Harris County for both cases was depicted. It is interesting that the most vulnerable areas are clustered in Houston city.

Table 13. Composite score evaluation combining flood impacts, population density and social vulnerability for Allison case

Tract Number (Top 7)	Flood Impacts	Population Density	Social Vulnerability	Composite Score
2214	Medium	High	High	332
5321	No impact	High	High	330
5212	No impact	High	High	330
5211	No impact	High	High	330
4229	No impact	High	High	330
3321	High	High	High	333
3331	High	High	High	333

Table 14. Composite evaluation combining flood impacts, population density and social vulnerability for Harvey case

Tract Number (Top 7)	Flood Impacts	Population Density	Social Vulnerability	Composite Score
3301	High	High	High	333
3106	High	High	High	333
3206	High	High	High	333
5223	High	High	High	333
3331	High	High	High	333
4514	No impact	High	High	330
4551	No impact	High	High	330

Most Vulnerable areas in Harris County for both cases

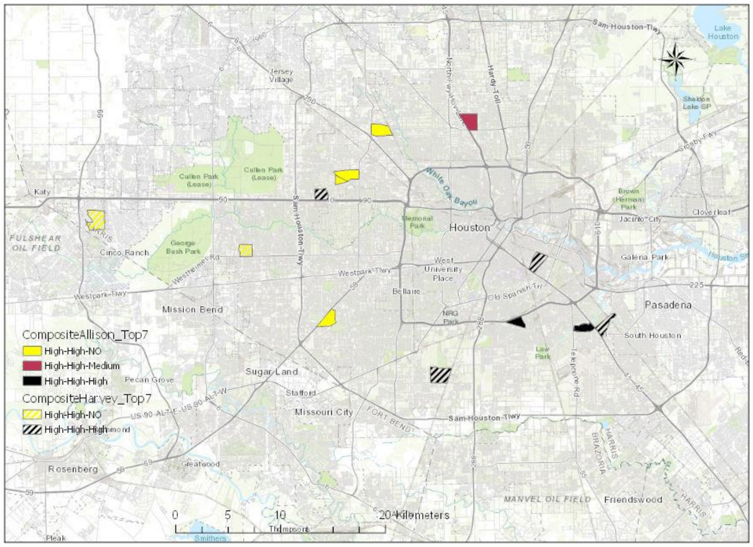


Figure 20. Spatial distribution of top 7 composite score areas in Harris county for both cases

6. Conclusions

In this study, an integrated vulnerability analysis was successfully achieved. A composite method combining hazard exposure and social vulnerability was conducted to examine the spatial pattern change of vulnerability and hazard exposure from 2001 to 2017 in Harris County. The population result shows that the overall population density of Harris County has increased since 2001. Notably, people in Harris county would like to move their houses to the north for some reasons (Hurricanes? More jobs?). After mapping and validating Harris flood zones, the spatial patterns of floods induced by Allison and Harvey can be interpreted. During tropical storm Allison, an extreme flash flood occurred in the southeastern part of Harris County with a 1000 mm rainfall record. The extreme rainfall lasted around almost two days until Allison was gone. During Hurricane Harvey, there was also an extreme flood occurred. Compared with Allison, Harvey brought 5-day rainfall with a historical record of 1500 mm over most of Harris county regions (worst in Houston city). This devastating rainfall event developed a 1000-year extreme flood that has never happened in the United States before. For SoVI map, it is safe to say that 2017 SoVI of Harris County is decreased since 2000. Compared with 2000 SoVI, some high-SoVI areas in the northern part of Harris County are classified as low-SoVI ones. One of the reason might be, with the increasing of the overall population in northern Harris County, the percentage of minority races, poverty, and less education has decreased since 2000. In all, by combining hazard exposure information and social vulnerability information, a composite score can be calculated to examine detailed vulnerability changes quantitatively. In Figure 20, the most vulnerable areas were depicted based on two hazards in Harris County. It is interesting to observe that most of the vulnerable areas are within Houston city. In fact, according to some previous reports distributed by Houston City Data, due to the high population density, high social vulnerability, and poor drainage system, rainfall and flash floods have become the biggest threat in Houston city. In this case, some political policies should be established in order to reduce the unemployment rate and improve the living condition of poor black citizens (reduce factor 2 value). Also, urban planners should pay more attention on the drainage system in Houston city. some necessary maintenance and technical improvement should also be fully considered.

7. Reference

(1) Social Vulnerability and Hazard Analysis for Hurricane Harvey. (n.d.). Retrieved from https://www.oxfamamerica.org/static/media/files/hurricane-harvey-background.pdf

(2) WEATHER RADAR PRINCIPLES - Severe Weather Information Centre. (n.d.). Retrieved from http://severe.worldweather.wmo.int/TCFW/RAIV_Workshop2016/14_Radar_StaceySteward.pdf

(3) Manning's n Values for Various Land Covers To Use for Dam Breach Analyses by NRCS in Kansas. Retrieved from https://www.wcc.nrcs.usda.gov/ftpref/wntsc/H&H/HecRAS/NEDC/lectures/docs/Manning%92s%20n values%20for%20Kansas%20Dam%20Breach%20Analyses%20-%20Adopted%20071216.pdf

(4) Spectroscopy of Urban Vulnerability: Annals of the ... (n.d.). Retrieved from http://www.tandfonline.com/doi/full/10.1080/00045608.2012.702485

(5) Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social Vulnerability to Environmental Hazards*. Social Science Quarterly, 84(2), 242-261. doi:10.1111/1540-6237.8402002

(6) Cutter, S. L. (2012). Hazards Vulnerability and Environmental Justice. doi:10.4324/9781849771542

(7) Frazier, T. G., Thompson, C. M., & Dezzani, R. J. (2014). A framework for the development of the SERV model: A Spatially Explicit Resilience-Vulnerability model. Applied Geography, 51, 158-172. doi:10.1016/j.apgeog.2014.04.004

(8) Kappes, M. S., Keiler, M., Elverfeldt, K. V., & Glade, T. (2012). Challenges of analyzing multi-hazard risk: A review. Natural Hazards, 64(2), 1925-1958. doi:10.1007/s11069-012-0294-2

(9) Koks, E., Jongman, B., Husby, T., & Botzen, W. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. Environmental Science & Policy, 47, 42-52. doi:10.1016/j.envsci.2014.10.013

(10) Flanagan, B. E., Gregory, E. W., Hallisey, E. J., Heitgerd, J. L., & Lewis, B. (2011). A Social Vulnerability Index for Disaster Management. Journal of Homeland Security and Emergency Management, 8(1). doi:10.2202/1547-7355.1792

(11) Wong, D. W., & Sun, M. (2013). Handling Data Quality Information of Survey Data in GIS: A Case of Using the American Community Survey Data. Spatial Demography, 1(1), 3-16. doi:10.1007/bf03354884

(12) Marshall et al. (1947). A real time colour display and rainfall measurement system for meteorological radar. doi:10.5353/th_b3123010