A.Paladino, R. Ballard, M. York

STAT 6289 Spatial Final

4/30/25

Project 3: Flood Risk Mismatch in the Mid-Atlantic

# Introduction

Flooding is among the most frequent and costly natural disasters in the United States, with profound consequences for infrastructure, ecosystems, and human safety. While the Federal Emergency Management Agency (FEMA) designates official flood zones to guide mitigation and planning efforts, a growing body of evidence suggests that damaging flood events increasingly occur outside of these historically defined boundaries. This evolving hazard landscape raises critical questions about the adequacy of existing flood risk maps in the context of climate change, urbanization, and shifting weather patterns.

To address these concerns, this study integrates multiple authoritative datasets—including FEMA flood zone maps, OpenStreetMap geospatial basemaps, U.S. Census polygonal data, and historical storm event annotations from the National Oceanic and Atmospheric Administration (NOAA)—to produce a series of interactive visualizations. These visualizations aim to enhance public understanding and policy evaluation by revealing where flooding is occurring relative to official flood zone designations, and how the frequency and severity of these events have changed over time.

Preliminary analysis reveals a clear and troubling trend: both the number and intensity of flood events have increased in absolute terms, with a notable rise in impactful incidents outside designated flood zones. This observation underscores the need for dynamic, data-driven flood risk assessments that move beyond static maps and better reflect contemporary and future realities. By combining spatial, temporal, and demographic data, our approach offers a holistic view of flood risk and its societal implications.

# Data

Data was extracted from Census Bureau, FEMA, Esri, OpenStreetMap, and NOAA-NWS endpoints. NOAA Data included severe weather events parsed from the National Centers For Environmental Information (NCEI) Storm Events Database. FEMA data included National Flood Hazard Layer (NFHL/layer 28) data as was made available via database search results and the FEMA Esri REST endpoint. Maps were implemented using OpenStreetMap in the leaflet framework.

## NOAA NCEI

NOAA event details data was extracted from NOAA National Centers for Environmental Information (NCEI) Events Database Bulk CSV SFTP server. The SFTP html was scraped to identify target URLs for annual event archives and data was iteratively extracted with a variable wait time to avoid rate-limiting.

These annual archives were then collated row-wise, filtered to include only DC, MD, and VA State FIPS, (11, 24, and 51 respectively), event\_type was filtered to include only “flash flood”, “flood”, and “coastal flood” and columns were subset to include only those germane to the analysis.

Compressed 1950-2024 annual events datasets total 316MB, compressed processed event data for DMV total 7.94MB. Processed data was saved as a pipe delimited CSV and loaded as string-typed then fields type-mutated as needed for portability.

Columns included in compressed NOAA-NCEI Events base table storm\_details\_dmv.csv.gz:

| "EVENT\_ID" | ID assigned by NWS for each individual storm event contained within a storm episode; links the record with the same event in the storm\_event\_details, storm\_event\_locations and storm\_event\_fatalities tables (Primary database key field). |
| --- | --- |
| "BEGIN\_YEARMONTH" | The year and month that the event began |
| "BEGIN\_DAY" | The day of the month that the event began |
| "BEGIN\_TIME" | The time of day that the event began |
| "STATE" | The state name where the event occurred (no State ID’s are included here; State Name is spelled out in ALL CAPS). |
| "STATE\_FIPS" | A unique number (State Federal Information Processing Standard) assigned to the county by the National Institute for Standards and Technology (NIST). |
| "YEAR" | The four digit year for the event in this record. |
| "MONTH\_NAME" | The name of the month for the event in this record (spelled out; not abbreviated). |
| "EVENT\_TYPE" | The only events permitted in Storm Data are listed in Table 1 of Section 2.1.1 of NWS Directive 10-1605 at http://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf.  The chosen event name should be the one that most accurately describes the meteorological event leading to fatalities, injuries, damage, etc. However, significant events, such as tornadoes, having no impact or causing no damage, should also be included in Storm Data. |
| "CZ\_FIPS" | The county FIPS number is a unique number assigned to the county by the National Institute for Standards and Technology (NIST) or NWS Forecast Zone Number |
| "CZ\_NAME" | County/Parish, Zone or Marine Name assigned to the county FIPS number or NWS Forecast Zone. |
| "BEGIN\_DATE\_TIME" | MM/DD/YYYY hh:mm:ss (24 hour time usually in LST) |
| "INJURIES\_DIRECT" | The number of injuries directly caused by the weather event. |
| "INJURIES\_INDIRECT" | The number of injuries indirectly caused by the weather event. |
| "DEATHS\_DIRECT" | The number of deaths directly caused by the weather event. |
| "DEATHS\_INDIRECT" | The number of deaths indirectly caused by the weather event. |
| "DAMAGE\_PROPERTY" | The estimated amount of damage to property incurred by the weather event (e.g. 10.00K =  $10,000; 10.00M = $10,000,000) |
| "DAMAGE\_CROPS" | The estimated amount of damage to crops incurred by the weather event (e.g. 10.00K = $10,000; 10.00M = $10,000,000) |
| "FLOOD\_CAUSE" | Reported or estimated cause of the flood. |
| "BEGIN\_LOCATION" | The name of city, town or village from which the range is calculated and the azimuth is determined. |
| "BEGIN\_LAT" | The latitude in decimal degrees of the begin point of the event or damage path. |
| "BEGIN\_LON" | The longitude in decimal degrees of the begin point of the event or damage path. |

Full Data Dictionary in Appendix

## 

## FEMA NFHL (Database)

NFHL shapefiles were programmatically extracted from the FEMA National Flood Hazard Layer (NFHL) Database in a similar fashion to the NOAA NCEI bulk csv extract. The query front-end html was extracted and parsed for target URLs with DFIRM6 IDs prefaced with “11”,”24”,”51” (DC,MD,VA).

A dataframe was constructed with target URLs and other shapefile metadata and shapefile .zips were iteratively downloaded. These shapefile .zips were then iteratively unpacked into a temp directory. The flood layer, S\_FLD\_HAZ\_AR.shp, from each shapefile was loaded and DFIRM and floodzone designation counts by polygon (census tract) for the DFIRM6 were extracted.

These DFIRM6’s were then mapped to FIPS5, and the data was reshaped to have flood zone designation as a flattened field with counts. These values could then be normalized to tabulate % of DFIRM6/FIPS5 tracts designated as A,AE,X,etc.

Compressed DMV DFIRM6 shapefiles total 5.38GB, tabular DFIRM-GEOID Floodzone “outputs.csv” total 9.7KB

| "GEOID" | DFIRM6-FIPS5 mapping for county |
| --- | --- |
| “FLD\_ZONE” | FEMA Flood Zone designation:  "A", "AE", "X", "VE", "AO", "OPEN WATER", "AREA NOT INCLUDED", "AH", "D" |
| “COUNT” | Number of census tracts of a given FLD\_ZONE designation for a GEOID |

FEMA data reference in Appendix

## 

## FEMA NFHL (REST)

The NCEI FEMA flood dataset used to fetch the DMV in bulk was pulled from an ESRI Rest Endpoint located [here](https://services.arcgis.com/P3ePLMYs2RVChkJx/ArcGIS/rest/services/USA_Flood_Hazard_Reduced_Set_gdb/FeatureServer/0/query). While the dataset is hosted on ESRI, it was created and maintained by FEMA in effort to be compliant with the National Flood Insurance Rate Program. The program exists as a reference to those who live within mapped hazard zones, as each habitant is required to purchase flood insurance for their home or assets. The Endpoint was normalized and turned into a front end interface where a user can simply enter a bounding box in WKT format and extract hazard zones that fall within the defined region. The frontend will output a normalized table that users can export. The frontend, while stored locally, was utilized in this project for FEMA flood hazard zones as well as census attribution and geometries. The only step required to manipulate this dataset was a simple import into our R environment .

**FEMA NFHL (REST) Attribution**

| DFIRM\_ID | Digital Flood Insurance Rate Map identifier for the panel the feature belongs to. |
| --- | --- |
| FLD\_AR\_ID | Unique ID for the flood hazard area polygon. |
| STUDY\_TYP | Type of study used (e.g., NP = Not Profiled, indicates no detailed hydraulic study). |
| FLD\_ZONE | Flood Zone designation (e.g., X, AE, VE). Describes flood risk based on annual chance. |
| ZONE\_SUBTY | Subtype of flood zone (e.g., “0.2 Percent Annual Chance Flood Hazard” = 500-year flood zone). |
| SFHA\_TF | Special Flood Hazard Area True/False: T = In SFHA (high risk), F = Not in SFHA. |
| STATIC\_BFE | Base Flood Elevation (in feet), if static. May be blank for areas with no detailed study. |
| V\_DATUM | Vertical datum used (e.g., NAVD88). Often omitted if not applicable. |
| DEPTH | Depth of flooding in feet for shallow flooding zones (like Zone AO) |
| LEN\_UNIT | Unit for depth or elevation (usually “Feet”). |
| VELOCITY | Flood velocity (ft/s), used in coastal flood zones (e.g., VE). |
| VEL\_UNIT | Unit for velocity (usually “ft/s”). |
| DUAL\_ZONE | Indicates if the feature belongs to overlapping flood zones (e.g., insurance zone differs from regulatory zone). |
| SOURCE\_CIT | Source citation or study for the data. |
| GFID | Global Flood ID – a unique global identifier for the feature. |

## Census Geometries & Census Demographics

We were also able to use the same front end interface used to pull FEMA Flood Hazard Zones to pull Census data, where we could simply select an output type (census geometries vs. attribution) as well as select a census level (state, county, tract, or block group). As mentioned, the front end sits on top of a multitude of API endpoints, including [one](https://tigerweb.geo.census.gov/arcgis/rest/services/TIGERweb) for census data geometries and another for census [demographics](https://api.census.gov/data/2023/pdb/blockgroup). We used both the block group and county level geometries for this project, in addition to block group level demographics (socioeconomic data). Similar to our FEMA Hazard data, we imported these front end exports into our R environment for further manipulation and analysis.

Census Block Group Geometry Attribution

| MTFCC | MAF/TIGER Feature Class Code – Code describing the geographic feature type (e.g., G5030 = Census Block Group). |
| --- | --- |
| OID | Object Identifier – Internal row index (may be software-specific or database primary key). |
| GEOID | Geographic Identifier – Concatenated unique ID for the block group: STATE + COUNTY + TRACT + BLKGRP (e.g., 540339401001). |
| STATE | State FIPS Code – 2-digit code representing the U.S. state (e.g., 54 = West Virginia). |
| COUNTY | County FIPS Code – 3-digit code for the county within the state. |
| TRACT | Census Tract Code – Subdivision of a county used for statistical analysis. |
| BLKGRP | Block Group Code – 1-digit identifier of the block group within the tract. |
| BASENAME | Short or base name of the block group (often just the block group number). |
| NAME | Human-readable name, typically in the form "Block Group 1, Census Tract 9401, County X, State Y". |
| LSADC | Legal/Statistical Area Description Code – Classifies the entity type (e.g., BG = Block Group). |
| FUNCSTAT | Functional Status – Indicates if the geography is active: |
| AREALAND | Land area in square meters. |
| AREAWATER | Water area in square meters. |
| STGEOMETRY | Geometry column – actual polygon or multipolygon shape for mapping. |
| CENTLAT | Latitude of the centroid of the geometry. |
| CENTLON | Longitude of the centroid. |
| INTPTLAT | Latitude of the internal point (used for labeling or map placement). |
| INTPTLON | Longitude of the internal point. |

Census County Geometry Attribution - Same as BG table with exception of columns highlights below

| MTFCC | MAF/TIGER Feature Class Code – Describes the feature type. For counties, this is usually G4020. |
| --- | --- |
| OID | Object Identifier – A sequential ID, usually internal to your GIS system (e.g., ArcGIS), used for data indexing. |
| GEOID | Geographic Identifier – 5-digit concatenated code: STATE (2-digit FIPS) + COUNTY (3-digit FIPS). E.g., 54033 = WV, Harrison County. |
| STATE | State FIPS Code – 2-digit code identifying the U.S. state. |
| COUNTY | County FIPS Code – 3-digit code identifying the county within the state. |
| COUNTYNMS | Full name of the county (e.g., "Harrison County"). |
| BASENAME | Base name of the county without “County” or equivalent suffix (e.g., "Harrison"). |
| NAME | Full descriptive name including "County" or equivalent designation. |
| LSADC | Legal/Statistical Area Description Code – Describes the legal/statistical nature of the area (e.g., 06 = County or Equivalent) |
| FUNCSTAT | Functional Status Code – Indicates operational status: |
| COUNTYCC | County Change Code – Tracks county boundary changes (e.g., splits, merges). Often H1 for historical consistency or left blank. | |

Census Block Group Socioeconomic Attribution - only a couple columns included. Full attribution can be found in the index.

| State\_name | State Name |
| --- | --- |
| County\_name | County Name |
| GIDBG | Block Group ID (Block Group GEOID |
| Med\_HHD\_Inc\_BG\_ACS\_17\_21 | Median Household Income |
| Med\_House\_Value\_BG\_ACS\_17\_21 | Median Household Income |
| pct\_Mobile\_Homes\_ACS\_17\_21 | % of population living in Mobile Homes |
| pct\_PUB\_ASST\_INC\_ACS\_17\_21 | % of population on public assistance income |
| pct\_Prs\_Blw\_Pov\_Lev\_ACS\_17\_21 | % of population below poverty level |
| pct\_College\_ACS\_17\_21 | % of people 25 or older with college degree |

# Methods

## Event Time Series Analysis

Using the NOAA-NCEI event details table a series of pivot tables were created tabulating total: Event Count, Estimated Damage (in log-dollars), and Casualties (Injuries and Deaths) both per-FIPS5 State-County per decade, and total in DMV per decade.

Further cleaning the event dataset and preparatory steps included parsing string scalars in CROP\_DAMAGE and PROPERTY\_DAMAGE and totalling, creating FIPS5 as a string concatenation of STATE\_FIPS and CZ\_FIPS, totalling INJURIES\_DIRECT, INJURIES\_INDIRECT, DEATHS\_DIRECT, DEATHS\_INDIRECT, and implementing dplyr calls to pivot, then creating panelled visualizations in the form of multiboxplots for FIPS decade totals and bar charts for DMV decade totals.

## Event Floodzone Mismatch Analysis

To examine the mismatch between flood events and designated flood hazard zones, the latitude and longitude of NOAA-NCEI flood events were spatially joined with FEMA flood zone polygons. A variable named in\_flood\_zone was created to classify events: those intersecting with “AE” or “A” zones were labeled as “In” the flood hazard zone, while those intersecting with “X” zones were labeled as “Out.” Flood events were then mapped by their geographic coordinates and color-coded according to their in\_flood\_zone classification. An interactive visualization was developed to display these events over time, including summary statistics on the occurrence of floods inside and outside of the flood zones.

To explore this mismatch more closely within Washington, D.C., NOAA flood events were filtered to include only those occurring in the region. These events were again intersected and joined with FEMA flood zone polygons, generating a marker\_id variable representing the specific polygon each event intersected. As in the previous analysis, each event was assigned an “In” or “Out” designation based on whether it fell within “AE” or “A” (hazard zones) or “X” (non-hazard) classifications. The resulting data were aggregated by marker\_id, providing summary statistics of flood event counts by event type (flood, flash flood, and coastal flood) and overall total. A detailed visualization was created for the Washington, D.C. area, mapping both flood zone polygons and the geographic location of flood events. This interactive tool included features such as pop-up summaries of aggregated flood counts per polygon and the ability to toggle point coloring based on flood zone inclusion or year of occurrence.

## Census Enrichment Analysis & SocioEconomic Impact

After understanding the interaction between flood events and mapped flood hazard zones, we decided to test the interaction between flood event counts and census geometries. In order to get census spatial data, we relied on a pre-existing pipeline used to hit an ESRI Rest Census API Endpoint to fetch and normalize block group geometries for the DMV into a usable geospatial format. After successful fetch and export, we imported our Block Groups into our R environment, and concatenated boundaries for the Virgina, Maryland, and DC. The result was one object representing block group geometries for the entire DMV region to which we then used to compute a spatial join (inner and within) with the flood events dataset mentioned in prior sections. The join provided us with a large table consisting of both flood events AND census attribution then used to compute multiple aggregations. Due to the rich attribution in our census tables, we were able to aggregate flood event counts across multiple census levels, each painting a different story about localized vs. regional impacts to severe weather events.

As mentioned in prior sections, there was a large disparity between total storm events and storm events related specifically to flooding, the latter to which we found no attribution until the 1990’s. Despite this small caveat, it was interesting to compare aggregate total flood events per census geography to aggregate flood specific events. We found that Montgomery County, Maryland, and Fairfax County, VA, were top two in both categories; Montgomery County taking the lead in total storm events, and Fairfax county in flood specific events. This enforced the narrative that these two counties were subject to a massive amount of natural disaster from 1950 - present. Because we had aggregate event counts per census geometry, it was easy to compute time series analysis and generate time series plots highlighting anomalies at multiple levels. This led us to examine correlations across levels during similar years.

We took our analysis a step further by generating choropleth maps highlighting flood event counts. We generated maps for both the block group (most granular) AND county levels. The county level map required us to use the same census pipeline to fetch county-specific boundaries and essentially go through the same enrichment steps. As mentioned, the comparison between these maps painted a very different story, one at macro vs. micro levels. These maps, as interesting as they were, only painted a fraction of the picture.

In alignment with the overall purpose of the project, We took this analysis yet another step by creating logic to essentially generate an aggregate table consisting of block group ids, block group geometries, total event counts per block group, and total mapped FEMA flood hazard layer per block group. We were able to use this table, import it into ArcGIS Online, and generate the Bivariate Map highlighting disparities between flood event counts and mapped flood zone areas. As mentioned, and because the map painted half the picture, we decided to compute a vulnerability metric. The goal of the metric was to compare block groups that experienced a lot of flooding events but had minimally mapped hazard zones (most vulnerable) to areas with minimal flooding events with a lot of mapped hazard areas. This vulnerability index alone was highlighted by the bivariate map.

Finally, and on the topic of flood vulnerability, we decided to conduct one final join on a census attribution layer to examine correlations, if any, between highly vulnerable zones and socioeconomic factors like median income, etc.

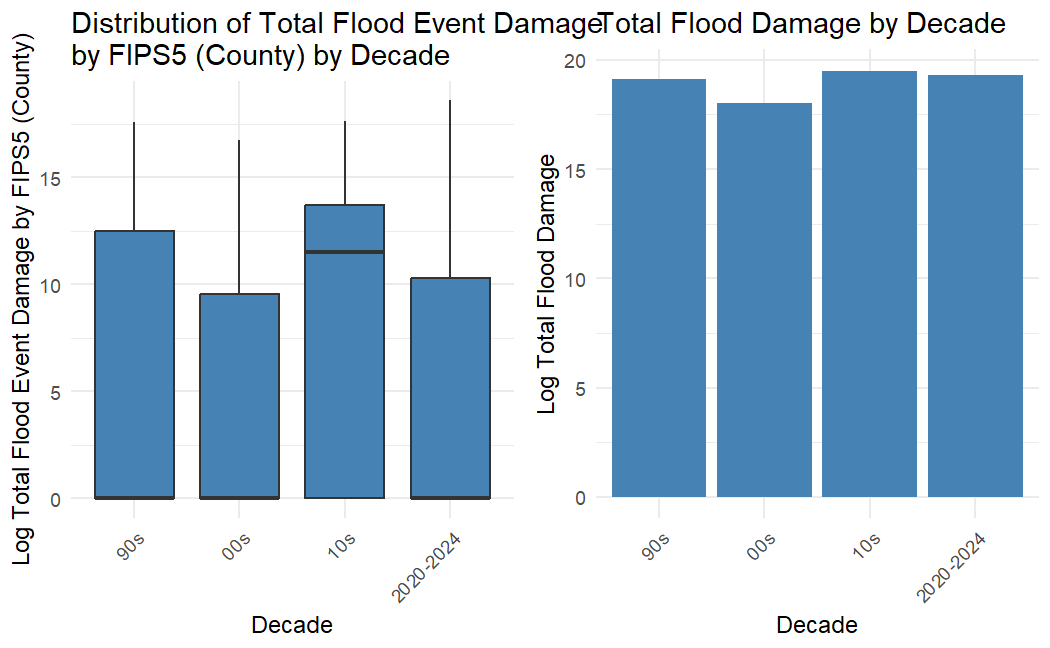
# Results

*(NOTE: The Results section must present findings clearly, using well-formatted and labeled tables and figures. All visuals must be original (no screenshots). Each result must be described and interpreted in full sentences.)*

## Event Time Series Analysis

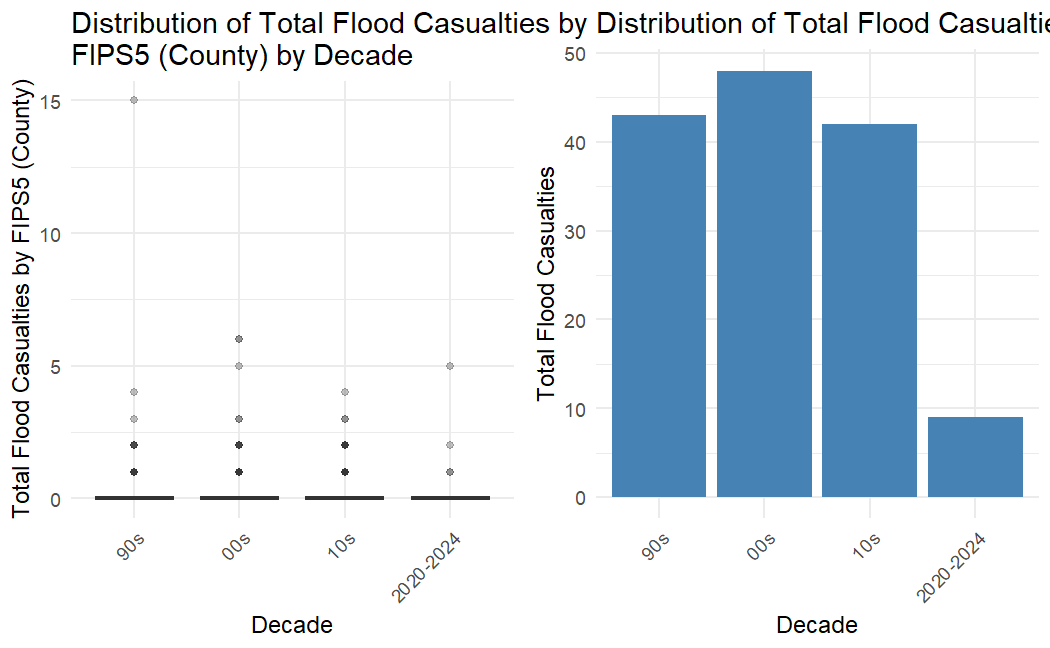
(Multiboxplot depicting per-decade event totals by FIPS and bar chart depicting total DMV events by decade)

The above panelled charts depict decade total event counts for 1990s-2024 in the DMV. The multiboxplot on the left visualizes total events by FIPS5 (State-County level), and the box plot on the right depicts DMV event totals by decade. We can see a clear increasing trend in the total events both at the FIPS5 level and in aggregate for the region. We also see an increased outlier effect, with flood-prone counties experiencing more extreme numbers of flood events as time progresses. Of note is that 2020-2024 totals include 40% of the timescale compared to full decades, so one could extrapolate that full decade totals will continue established trends.



(Multiboxplot depicting per-decade log-dollar event damage totals by FIPS and bar chart depicting total DMV log-dollar event damage by decade)

For the majority of events (~56%) no estimated damage is reported, thus a log+1 transformation is applied to help with interpretability of the aggregated FIPS5 damage totals and associated visualization. With this transformation applied we can see that damage distributions in the 00s skewed and trended downwards when compared to the 90s, however median FIPS total estimated damage increased quite a bit in the 10s. As with event counts the first 4 years of the 20s are on track to eclipse previous decades in terms of estimated total damage per DMV county and overall for the region.



(Multiboxplot depicting total casualties by FIPS and bar chart depicting total DMV log-dollar event damage by decade)

Here is what appears to be a rare positive story in the decade time series analysis. Decade to-date casualty totals for 2020s appear to be lower than previous decades. This could be due to many factors and warrants additional investigation. To speculate, there could be additional advance early warning, improved resiliency, or behavioral changes that account for the decrease in casualties.

## Event Floodzone Mismatch Analysis

\* Insert RShiny time series map here \*

The interactive time series map of flood event occurrences, colored by spatial mismatch with FEMA flood hazard zones, provides insight into the changes in event distribution within hazard zones as the occurrences of these events increases over time. The visual is a testament to the increasing occurrences of floods over time in the DMV region, as well as a display of the shortcomings of the current FEMA flood hazard zones. Across the timeseries, it can be observed that the percentage of floods occurring in the hazard zones fluctuates around 10%, a statistic reflecting the weak accuracy of the zoning. The lack of noticeable change in percentage of events occurring inside the flood zones implies the geographic distribution of flood events relative to hazard zones is not changing with their increase. While not indicating significant geographic shift in location of flood occurrences, the increasing flood occurrences and constant ratio of those within the flood zone means that each year, an increasing number of flood events are occurring outside of the flood zone. By proxy, not adjusting the FEMA flood zones would breach the definition of the different layers as the probability of a region experiencing a flood event increases.

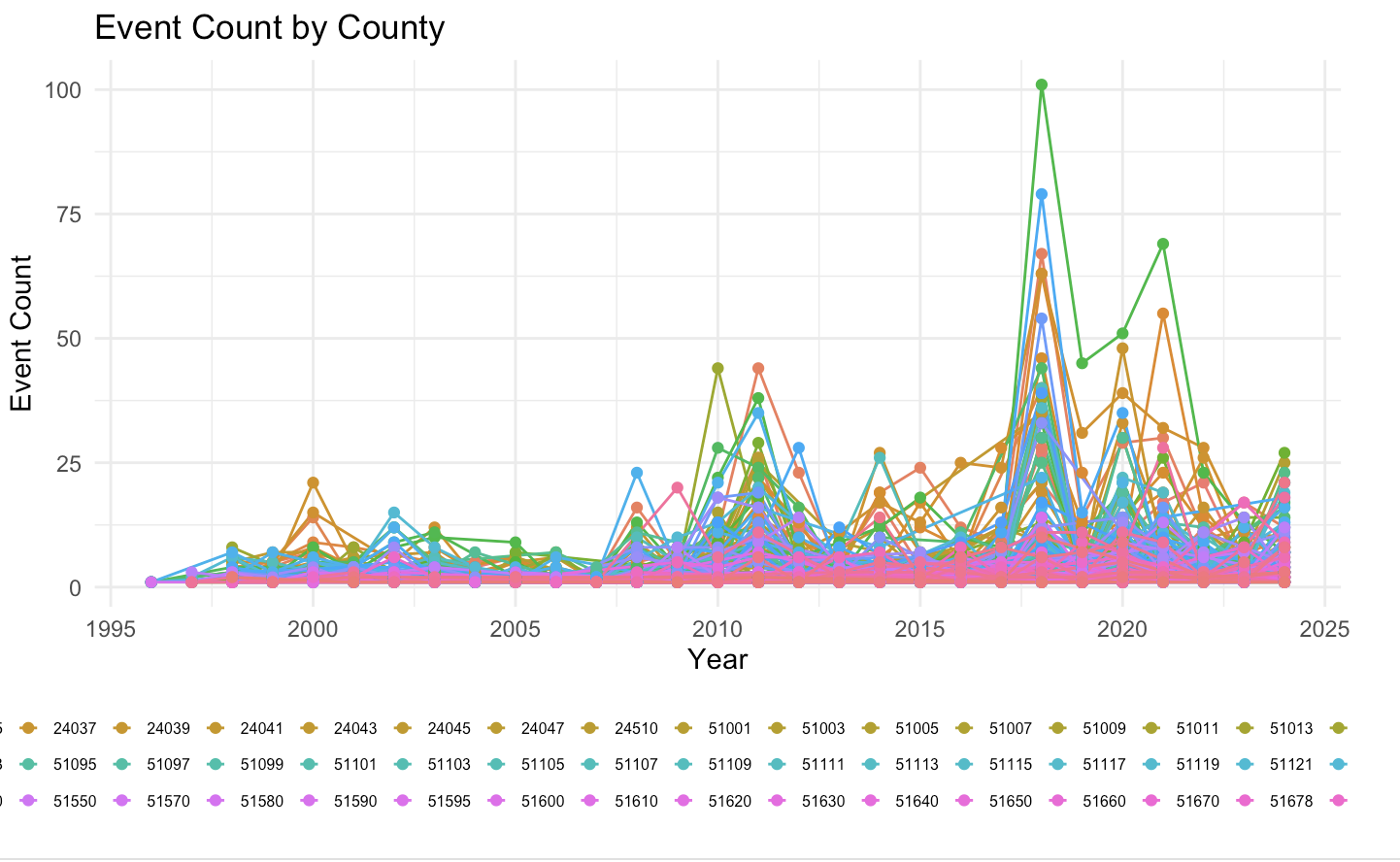
Interactive map of DC Mismatch Events:

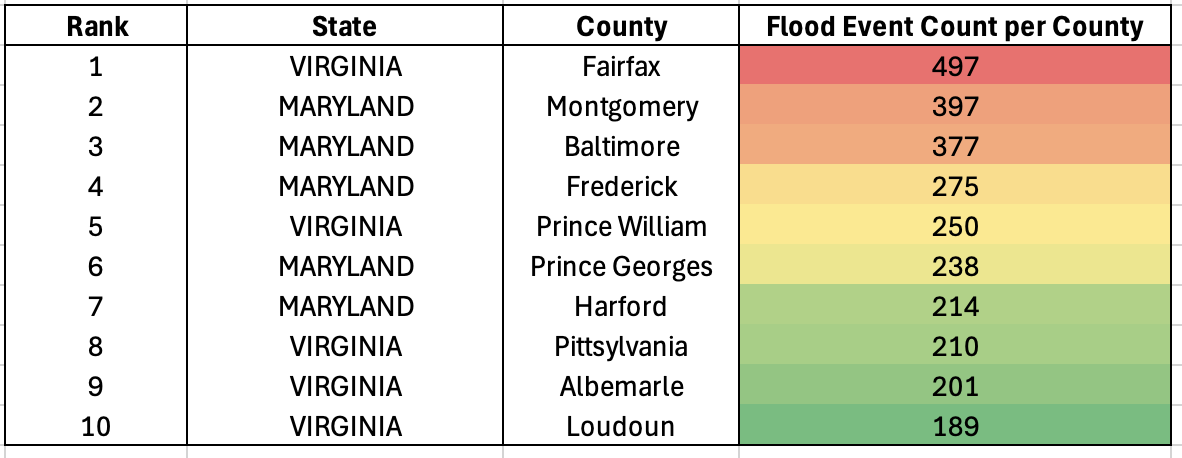
Visualization found here: <https://marleemooskie.github.io/flood_maps/>

HTML code for visualization found here: <https://github.com/marleemooskie/flood_maps>

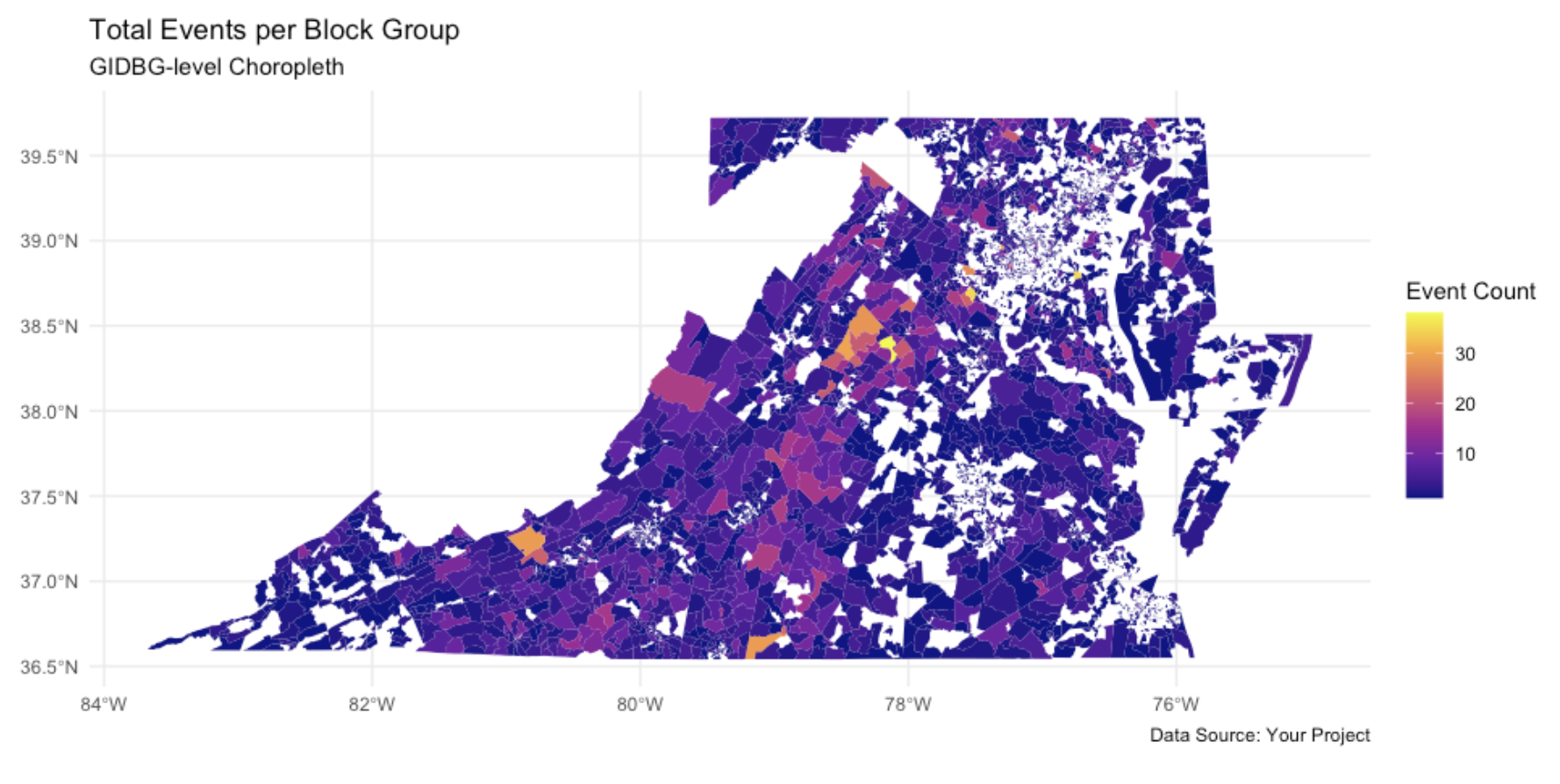
Building on the timeseries mismatch analysis of the DMV, the interactive flood event map of Washington, DC reveals the clear spatial mismatch between recorded flood events and FEMA flood hazard zones and echoes the results of the mismatch time series in more geographic detail. The map reveals that several flood events cluster within areas not designated as flood hazard zones, such as along Rhode Island Avenue NE. This explicitly reveals the underestimation of flood risk in FEMA’s current zoning. The ability to toggle between flood zone classification and event year highlights the recurring events outside mapped hazard zones across multiple years, emphasizing the need reassessment of existing floodplain designations in Washington, DC.

## Census Enrichment & Analysis

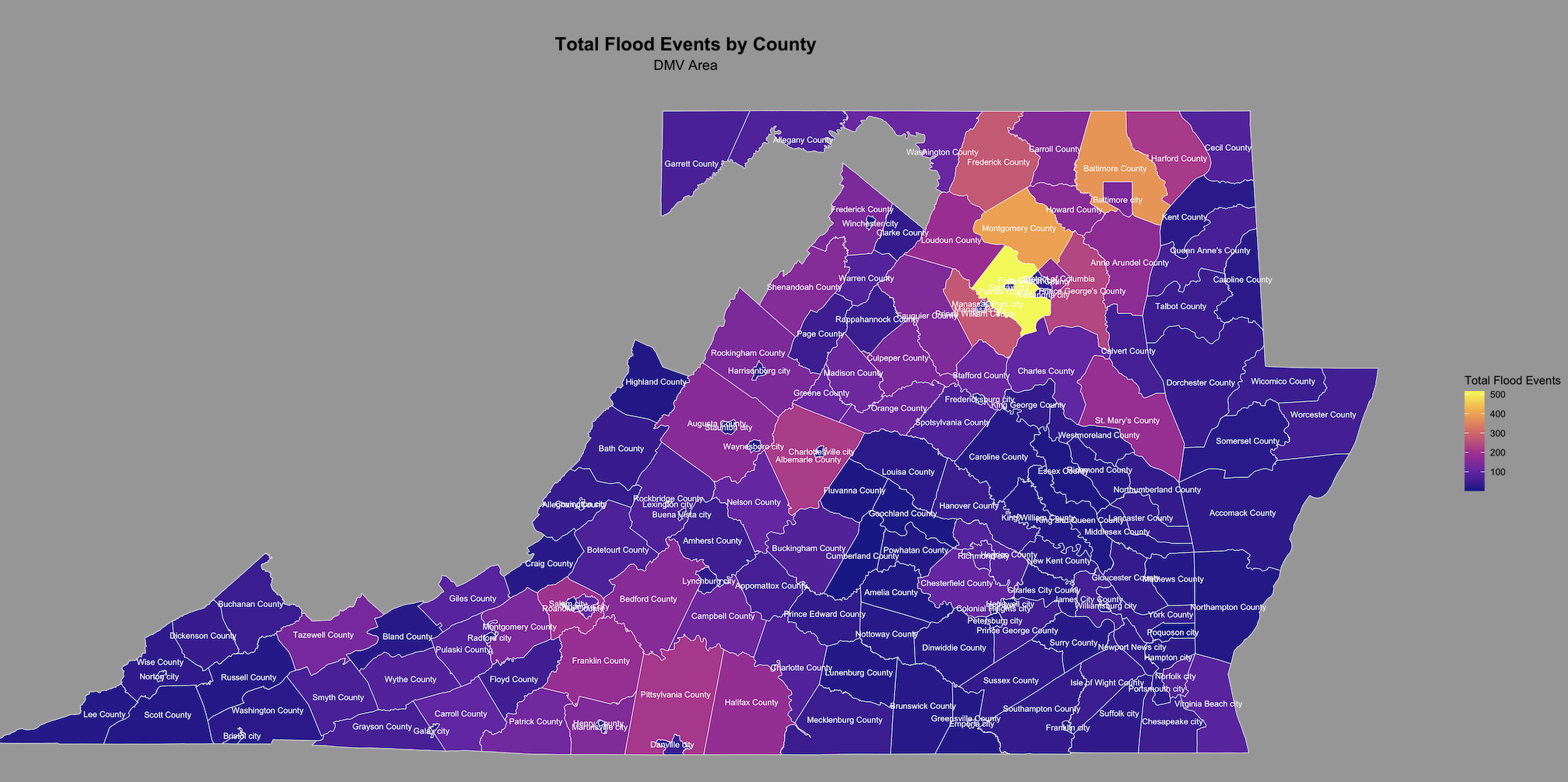


The above line plot shows flood event counts at the county level between 1995 and 2024. Immediately, we notice that event counts are doubling and almost tripling every decade when compared to the decade prior. Not only do we notice some anomalous years, we notice similar anomalous years across counties, indicating specific years where flood event counts were particularly high. The year 2018 saw a profound increase in flood event counts across the DMV, more so than any other recorded year in our dataset. Light research tells us that 2018 broke the yearly average precipitation totals by nearly 2 feet in DC alone ([WUSA9](https://www.wusa9.com/article/weather/weather-blog/dcs-top-weather-events-of-2018/65-624010286)). While the line plot is hard to interpret on a county by county basis, it provides a general summary of increasing flood events over time. 

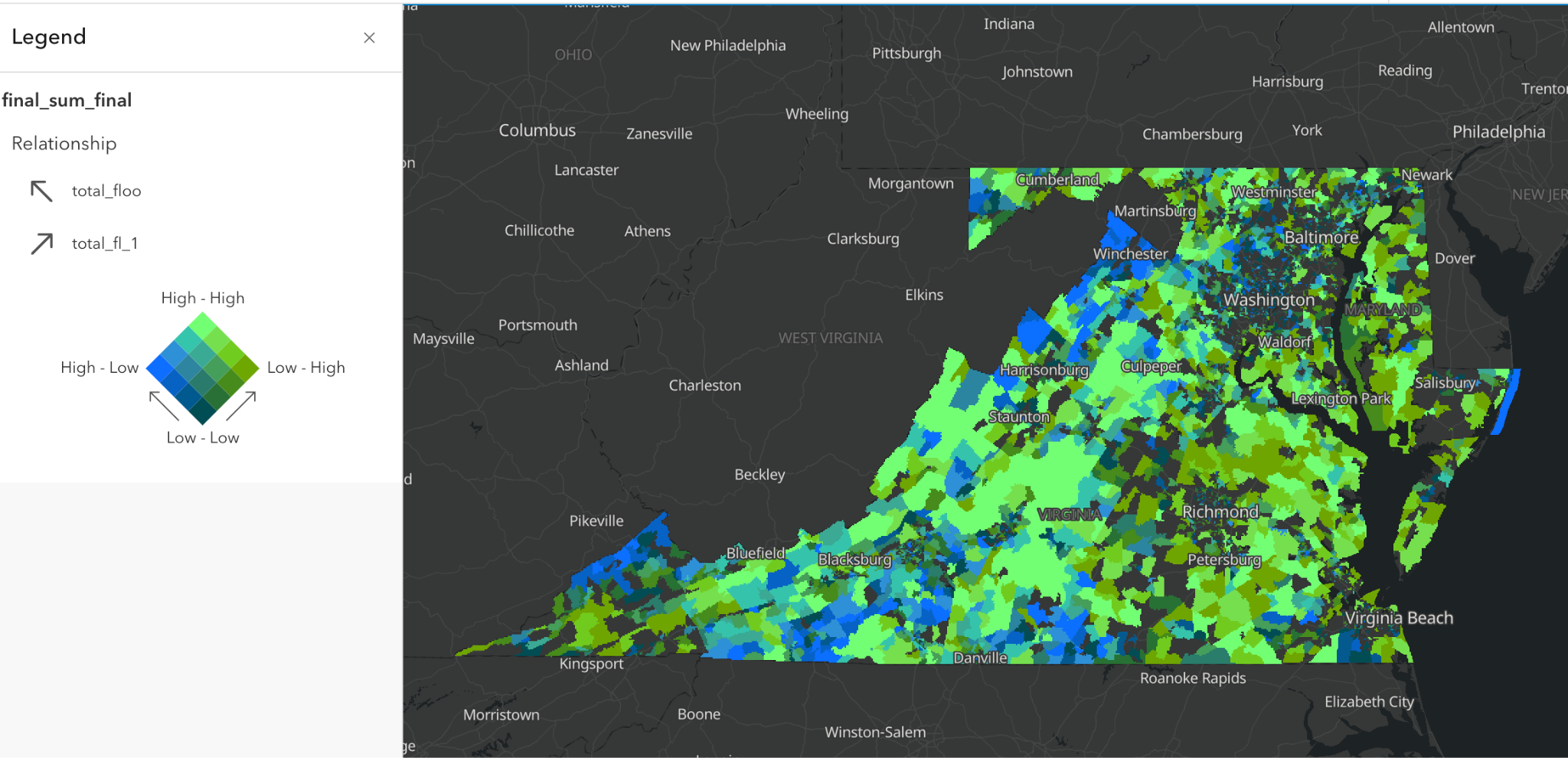
The above chart ranks the top 10 flood events by county from 1995 - 2024. While the line plots illustrate trends over time, the table shows us which areas were most impacted on a numerical basis. Because the attribution was so rich in our census table, we were able to duplicate the above tables for each census level. These tables can be found in our PowerPoint presentation. We found it interesting to compare storm event counts to flood specific event counts over time. In either case, we found both Montgomery County and Fairfax County being the two most impacted counties, Fairfax county by total storm event counts, and Montgomery County by flood-specific event counts. This evidence is also further supported by additional charts in our PowerPoint presentation. While there was significant evidence supporting flood event counts in DC, it still ranked much lower than several VA and MD counties.



In much of our analysis, we took the more granular approach modeling event counts at the lowest census level: block groups. We found however, this approach being tough to represent on 1 plot for the entire DMV, as block group boundaries are very small and change in size depending on the population represented. As you get closer to major population centers (DC and Arlington in this case), we notice it harder to physically see mapped block groups at smaller scales, as each block group attempts to adhere to somewhat strict population maxs and mins. This small caveat is why we notice empty unmapped data in the above plot. That is not to say that areas with smaller block groups were unimportant. In fact, we found the opposite to be true, as large population centers with smaller block groups in Fairfax county actually accounted for the largest event counts. Seeing that this was hard to represent at the block group level, we decided to zoom out.

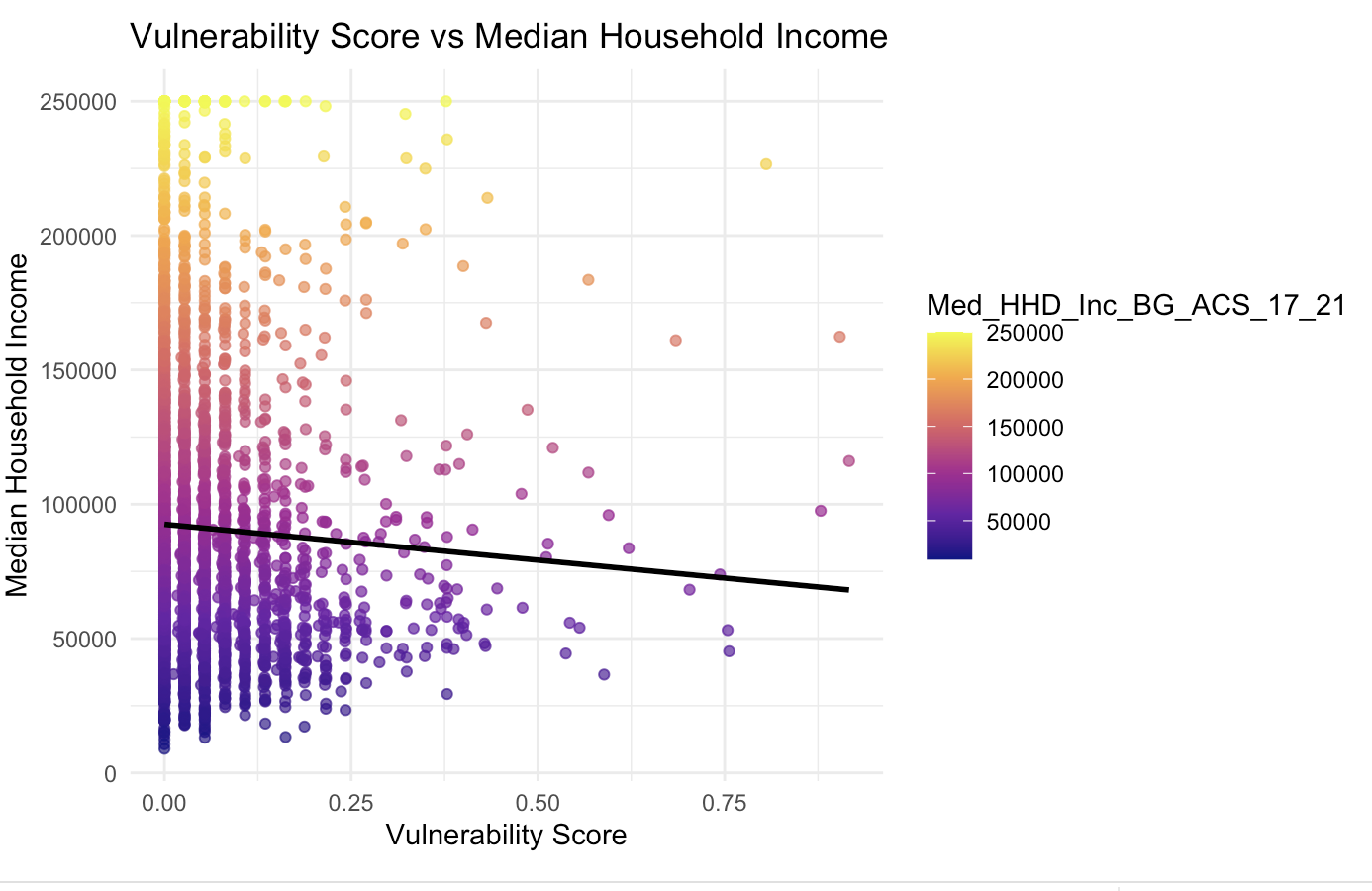


The above map,while not as granular, is a better static representation of events counts in the DMV, as we now clearly see counties such as Fairfax, Montgomery, and Baltimore come into play. As mentioned during the presentation, we also see some geographical correlation between event counts and the Shenandoah Mountains and the Shenandoah River Watershed, perhaps indicating that areas in those physical features are more subject to flood events. Aside from this finding, what stands out most is Northern VA and Maryland, although our analysis did not provide a clear window as to why these regions were most affected. The above map only paints so much of the picture. It shows the results of a simple spatial join between flooding events and census geometries. What happens when we invite FEMA Hazard Zones into our Findings?



<https://xyzeus.maps.arcgis.com/home/item.html?id=326be42372544820af02b717a5c430b3>

We had to get somewhat creative with the above join. The end goal was to show the total event counts AND total mapped flood zone hazard area (square meters) per block group geometry. The analysis would help us identify areas where flooding was common but mapped hazards were lacking; a total mismatch analysis, if you will. Because we found trouble conducting this workflow in R, we decided to write logic in python, where we would write logic to make the join, and then create an aggregate table table representing block groups, total flood event counts per block group, total mapped hazard area per block group, and block group geometries. Because we wanted to model not one, but now two geospatial phenomena in our aggregate table, we turned our sites to ArcGIS Online, where we could employ a Bivariate Map to sufficiently highlight disparities between event counts and total mapped hazard areas. Areas in blue represent block groups experience a major amount of flood events, but are lacking in mapped hazard zones. These can be interpreted as areas most vulnerable or severely lacking in flood planning. Areas in forest green represent areas that see minimal flood events but have adequately mapped hazard areas. Finally, areas in light green represent areas that see a ton of flooding AND have adequately mapped hazard areas. Areas in light green can be interpreted as areas where hazard mapping is sufficient.



We took insights from our bivariate map and decided to compute a vulnerability score. More vulnerable areas were areas with more events but minimally mapped hazard areas. Less vulnerable areas were areas that saw minimal flood impact but had adequately mapped hazard zones. We took our analysis 1 step further and incorporated census attribution, which essentially required one final join against our final agg table with census attribution, a dataset that we got from the same ESRI Rest Endpoint as the block groups. After computing the join, we were able to start analyzing correlations between our vulnerability score, and socioeconomic factors such as median household income (as seen in the plot above). We did find several positive relationships between high vulnerability and low socioeconomic status.

# Discussion

## Limitations:

The structure and information provided by collected datasets set limitations on this project. In terms of scope, the NOAA flood events database does not include sufficient information on those floods occurring prior to the 1990’s. This project, as a result, was based on patterns observed after this time period and into the present. With connections to the effects of climate change, the ability to observe these patterns in flood events over a longer period of time would prove beneficial in understanding broader flood trends. In addition to the limited time series, the location of the flood events themselves were designated using latitude and longitude coordinates. Presumably, this is not how the locality of floods occurs, which typically span over a region and cannot be consolidated into one starting and ending point. This limits our ability to understand the patterns of flood occurrences inside and outside of flood zones, as they likely expanded past the initial latitude and longitude they began at.

Aside from limited factors within the information provided by the flood datasets, the structure of the FEMA flood zone layer dataset also provided its own limitations. The size of this layer, when spanning across the entire DMV, was too large to successfully map across the entire region. As a result, only the calculations of its intersections could be used to visualize flood mismatch, without including the layers themselves. This removes the information on locality of flood events in comparison to hazard zones obtained through visually mapping flood zones alongside flood events.

## Improvements:

To address the limitations of the latitude and longitude structure of the flood event data, as previously mentioned, extra steps can be taken to estimate the affected flood region. This study could have been improved by spatially representing the floods using a radius of some predicted length around the latitude and longitude point, as compared to the point alone. This could better estimate the region affected by the flood, giving more accurate insight into its overlap with flood hazard zones and different census tracts. Our study could have been additionally improved by more thorough investigation into the socioeconomic associations with flood occurrences and their fallout. This includes expanding past our analysis of median household income to tell a more cohesive socioeconomic story including variables such as diversity and land area usage (eg rural or urban). Additionally, expanding our analysis past flood zone occurrences to include how its impacts, such as casualties and damage, interplay with socioeconomic trends can improve our study,

## Potential directions:

Moving forward with this project, there are multiple potential directions to facilitate a more in depth understanding of flood event occurrences in the greater DMV area. As previously mentioned in the improvements, there is capacity to explore a greater understanding of the socioeconomic relationship between flood events and their impact on different communities in the region. This could include expanding to include analysis of the interactions between diversity, land use, flood casualties, flood damage, and more.

Other opportunities for continued analysis include a deeper investigation into the drivers of our current results. This could include exploring specific extremely damaging events in greater detail. For example, the hurricane event observed in 2018 is associated with an anomalous increase in flood events. Investigating the geographic and socioeconomic distribution of years with hurricanes, such as 2018, could provide insight into how patterns of flooding in relation to communities and flood hazard zones changes between typical flood events and those resulting from large hurricane events.

# Team Member Work Attribution

| **Work Item** | **Team Member(s) Responsible** |
| --- | --- |
| NCEI Data Extraction and processing | Russ |
| NFHL Extraction (Database Search Endpoint) and processing | Russ |
| Event Timeseries Analysis | Russ |
| Project organization | Russ, Marlee, Andy |
| Mismatch analysis - design wireframing | Marlee, Russ |
| Census Extract, Cleaning, Processing | Andy |
| FEMA Flood Zone Extraction | Andy, Marlee |
| Census Enrichment Analysis + SocioEconomic Impact + Visualizations | Andy |
| Bivariate Map (ESRI) showing Flood events vs. Total Mapped Flood Hazard Area per BG | Andy |
| FEMA Flood Zone and NOAA Flood Event Mismatch Analysis | Marlee |
| Interactive Flood Zone Map | Marlee |
| QA & Source Control | Russ |

# Appendix

## NOAA NCEI Data Dictionary

https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/Storm-Data-Bulk-csv-Format.pdf

**Event Details File** (named StormEvents\_details-ftp\_v1.0\_d2019\_c20200219.csv):

Where d = data year and c = creation date

**begin\_yearmonth** Ex: 201212 (YYYYMM format)

The year and month that the event began

**begin\_day** Ex: 31 (DD format)

The day of the month that the event began

**begin\_time** Ex: 2359 (hhmm format)

The time of day that the event began

**end\_yearmonth** Ex: Ex: 201301 (YYYYMM format)

The year and month that the event ended

**end\_day** Ex: 01 (DD format)

The day of the month that the event ended

**end\_time** Ex: 0001 (hhmm format)

The time of day that the event ended

**episode\_id** Ex: 61280, 62777, 63250

ID assigned by NWS to denote the storm episode; Episodes may contain multiple Events.

The occurrence of storms and other significant weather phenomena having sufficient intensity

to cause loss of life, injuries, significant property damage, and/or disruption to commerce.

**event\_id** Ex: 383097, 374427, 364175

ID assigned by NWS for each individual storm event contained within a storm episode; links

the record with the same event in the storm\_event\_details, storm\_event\_locations and

storm\_event\_fatalities tables (Primary database key field).

**state** Ex: GEORGIA, WYOMING, COLORADO

The state name where the event occurred (no State ID’s are included here; State Name is

spelled out in ALL CAPS).

**state\_fips** Ex: 45, 30, 12

A unique number (State Federal Information Processing Standard) assigned to the county by

the National Institute for Standards and Technology (NIST).

**year** Ex: 2000, 2006, 2012

The four digit year for the event in this record.

**month\_name** Ex: January, February, March

The name of the month for the event in this record (spelled out; not abbreviated).

**event\_type** Ex: Hail, Thunderstorm Wind, Snow, Ice (spelled out; not abbreviated)

The only events permitted in Storm Data are listed in Table 1 of Section 2.1.1 of NWS Directive

10-1605 at http://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf.

The chosen event name should be the one that most accurately describes the meteorological

event leading to fatalities, injuries, damage, etc. However, significant events, such as tornadoes,

having no impact or causing no damage, should also be included in Storm Data.

From Section 2.1.1 of NWS Directive 10-1605:

Event Name Designator (County or Zone) Event Name Designator (County or Zone)

Astronomical Low Tide Z

Avalanche Z

Blizzard Z

Coastal Flood Z

Cold/Wind Chill Z

Debris Flow C

Dense Fog Z

Dense Smoke Z

Drought Z

Dust Devil C

Dust Storm Z

Excessive Heat Z

Extreme Cold/Wind Chill Z

Flash Flood C

Flood C

Freezing Fog Z

Frost/Freeze Z

Funnel Cloud C

Hail C

Heat Z

Heavy Rain C

Heavy Snow Z

High Surf Z

High Wind Z

Hurricane (Typhoon) Z

Ice Storm

Z

Lake-Effect Snow Z

Lakeshore Flood Z

Lightning C C

Marine Hail M

Marine High Wind M

Marine Strong Wind M

Marine Thunderstorm Wind M

Rip Current Z

Seiche Z

Sleet Z

Storm Surge/Tide Z

Strong Wind Z

Thunderstorm Wind C

Tornado C

Tropical Depression Z

Tropical Storm Z

Tsunami Z

Volcanic Ash Z

Waterspout M

Wildfire Z

Winter Storm Z

Winter Weather Z

**cz\_type** Ex: C, Z , M

Indicates whether the event happened in a (C) County/Parish, (Z) NWS Public Forecast Zone

or (M) Marine.

**cz\_fips** Ex: 245, 003, 155

The county FIPS number is a unique number assigned to the county by the National Institute

for Standards and Technology (NIST) or NWS Forecast Zone Number (See addendum)

**cz\_name** Ex: AIKEN, RICHMOND, BAXTER

County/Parish, Zone or Marine Name assigned to the county FIPS number or NWS Forecast

Zone.

**wfo** Ex: CAE, BYZ, GJT

The National Weather Service Forecast Office’s area of responsibility (County Warning Area)

in which the event occurred.

**begin\_date\_time** Ex: 04/1/2012 20:48:00

MM/DD/YYYY hh:mm:ss (24 hour time usually in LST)

**cz\_timezone** Ex: EST-5, MST-7, CST-6

Time Zone for the County/Parish, Zone or Marine Name. Eastern Standard Time (EST),

Central Standard Time (CST), Mountain Standard Time (MST), etc.

**end\_date\_time** Ex: 04/1/2012 21:03:00

MM/DD/YYYY hh:mm:ss (24 hour time usually in LST)

**injuries\_direct** Ex: 1, 0, 56

The number of injuries directly caused by the weather event.

**injuries\_indirect** Ex: 0, 15, 87

The number of injuries indirectly caused by the weather event.

**deaths\_direct** Ex: 0, 45, 23

The number of deaths directly caused by the weather event.

**deaths\_indirect** Ex: 0, 4, 6

The number of deaths indirectly caused by the weather event.

**damage\_property** Ex: 10.00K, 0.00K, 10.00M

The estimated amount of damage to property incurred by the weather event (e.g. 10.00K =

$10,000; 10.00M = $10,000,000)

**damage\_crops** Ex: 0.00K, 500.00K, 15.00M

The estimated amount of damage to crops incurred by the weather event (e.g. 10.00K =

$10,000; 10.00M = $10,000,000).

**source** Ex: Public, Newspaper, Law Enforcement, Broadcast Media, ASOS, Park and Forest

Service, Trained Spotter, CoCoRaHS, etc.

The source reporting the weather event (can be any entry; isn’t restricted in what’s allowed)

**magnitude** Ex: 0.75, 60, 0.88, 2.75

The measured extent of the magnitude type ~ only used for wind speeds (in knots) and hail size

(in inches to the hundredth).

**magnitude\_type** Ex: EG, MS, MG, ES

EG = Wind Estimated Gust; ES = Estimated Sustained Wind; MS = Measured Sustained Wind;

MG = Measured Wind Gust (no magnitude is included for instances of hail).

**flood\_cause** Ex: Ice Jam, Heavy Rain, Heavy Rain/Snow Melt

Reported or estimated cause of the flood.

category

Unknown (During the time of downloading this particular file, NCEI has never seen anything

provided within this field.)

**tor\_f\_scale** Ex: EF0, EF1, EF2, EF3, EF4, EF5

Enhanced Fujita Scale describes the strength of the tornado based on the amount and type of

damage caused by the tornado. The F-scale of damage will vary in the destruction area;

therefore, the highest value of the F-scale is recorded for each event.

EF0 – Light Damage (40 – 72 mph)

EF1 – Moderate Damage (73 – 112 mph)

EF2 – Significant damage (113 – 157 mph)

EF3 – Severe Damage (158 – 206 mph)

EF4 – Devastating Damage (207 – 260 mph)

EF5 – Incredible Damage (261 – 318 mph)

tor\_length Ex: 0.66, 1.05, 0.48

Length of the tornado or tornado segment while on the ground (in miles to the tenth).

tor\_width Ex: 25, 50, 1760, 10

Width of the tornado or tornado segment while on the ground (in whole yards).

tor\_other\_wfo Ex: DDC, ICT, TOP,OAX

Indicates the continuation of a tornado segment as it crossed from one National Weather

Service Forecast Office to another. The subsequent WFO identifier is provided within this

field.

tor\_other\_cz\_state Ex: KS, NE, OK

The two-character representation for the state name of the continuing tornado segment as it

crossed from one county or zone to another. The subsequent 2-Letter State ID is provided

within this field.

**tor\_other\_cz\_fips** Ex: 41, 127, 153

The FIPS number of the county entered by the continuing tornado segment as it crossed from

one county to another. The subsequent FIPS number is provided within this field.

**tor\_other\_cz\_name** Ex: DICKINSON, NEMAHA, SARPY

The FIPS name of the county entered by the continuing tornado segment as it crossed from one

county to another. The subsequent county or zone name is provided within this field in ALL

CAPS.

**begin\_range** Ex: 0.59, 0.69, 4.84, 1.17 (in miles)

The distance to the nearest tenth of a mile, to the location referenced below.

**begin\_azimuth** Ex: ENE, NW, WSW, S

16-point compass direction from the location referenced below.

**begin\_location** Ex: PINELAND, CENTER, ORRS, RUSK

The name of city, town or village from which the range is calculated and the azimuth is

determined.

**end\_range** see begin\_range

**end\_azimuth** see begin\_azimuth

**end\_location** see begin\_location

**begin\_lat** Ex: 29.7898

The latitude in decimal degrees of the begin point of the event or damage path.

**begin\_lon** Ex: -98.6406

The longitude in decimal degrees of the begin point of the event or damage path.

**end\_lat** Ex: 29.7158

The latitude in decimal degrees of the end point of the event or damage path. Signed negative (-)

if in the southern hemisphere.

**end\_lon** Ex: -98.7744

The longitude in decimal degrees of the end point of the event or damage path. Signed negative

(-) if in the eastern hemisphere.

**episode\_narrative** Ex: A strong upper level system over the southern Rockies lifted northeast

across the plains causing an intense surface low pressure system and attendant warm front to

lift into Nebraska.

The episode narrative depicting the general nature and overall activity of the episode. The

National Weather Service creates the narrative.

**event\_narrative** Ex: Heavy rain caused flash flooding across parts of Wilber. Rainfall of 2 to

3 inches fell across the area.

The event narrative provides descriptive details of the individual event. The National Weather

Service creates the narrative.

Storm Data Location File

(named StormEvents\_locations-ftp\_v1.0\_d1972\_c20181029.csv.gz)

Where dyyyy = data year and cyyyymmdd = file creation date

**episode\_id** Ex: 61280, 62777, 63250

ID assigned by NWS to d

## FEMA NFHL Database Download

Metadata and documentation:

<https://www.fema.gov/sites/default/files/documents/fema_rm-firm-database-technical-reference-nov-2023.pdf>

## 

## FEMA NFHL Rest Endpoint

* [Metadata and documentation](https://services.arcgis.com/P3ePLMYs2RVChkJx/ArcGIS/rest/services/USA_Flood_Hazard_Reduced_Set_gdb/FeatureServer)

## Census Geometries & Census Demographics

* [Census Block Group Metadata and Documentation](https://tigerweb.geo.census.gov/arcgis/rest/services/TIGERweb/tigerWMS_Census2020/MapServer/8)
* [Census County Metadata and Documentation](https://tigerweb.geo.census.gov/arcgis/rest/services/TIGERweb/tigerWMS_Census2020/MapServer/82)
* [Census Demographic Metadata and Documentation](https://api.census.gov/data/2023/pdb/blockgroup/variables.html)