Kristin Martin

Boston university | met cs 689 spring 2 2020

FEMA IHP Analysis

Project Summary & Report Findings

Table of Contents

[Project Scope 2](#_Toc39017940)

[Introduction 2](#_Toc39017941)

[A Note on Changes in Dataset 2](#_Toc39017942)

[Key Business Questions 2](#_Toc39017943)

[Design 3](#_Toc39017944)

[Dimensions 3](#_Toc39017945)

[Date Dimension 3](#_Toc39017946)

[Location Dimension 3](#_Toc39017947)

[Disaster Dimension 4](#_Toc39017948)

[Facts 4](#_Toc39017949)

[Data Warehouse Schema 5](#_Toc39017950)

[ETL Process 5](#_Toc39017951)

[Visualizations 6](#_Toc39017952)

[References 9](#_Toc39017953)

# Project Scope

## Introduction

FEMA, or the Federal Emergency Management Agency, oversees the federal response to disasters and emergencies in the United States. FEMA allocates federal resources to municipalities, businesses, and individuals facing disasters and disaster recovery efforts. Although much of FEMA’s effort is focused on large disaster recovery efforts, it also provides aid directly to households affected by disasters through its Individuals and Households Program (IHP). The IHP, according to FEMA documentation, “provides financial and direct services to eligible individuals and households affected by a disaster who have uninsured or underinsured necessary expenses and serious needs” (FEMA, 2019).

This data warehouse seeks to provide meaningful insight into the data collected from households participating in the IHP program, so FEMA (or another third party) can ascertain which areas have been hit hardest, which disasters had the biggest impact, if demographic factors play a role in aid, along with other geographic and demographic insights.

## A Note on Changes in Dataset

This dataset has been very tumultuous to work with due to the file size and frequent updates. After attempting to a load individual registrations for FEMA’s IHP program and failing, and making a very laborious attempt at FEMA’s hazard mitigation program dataset but coming up quite short on unique rows, I was able to locate a pre-aggregated set of individual data that met the project requirements. A link to this and all data sources used are located in the References section herein.

I also submitted the ETL package that I implemented for the Hazard Mitigation Program – it is not represented in the final visualizations, but was used in the initial presentation so I thought it important to include.

## Key Business Questions

1. In the last [time span], what [geographic area] has been awarded the most IHP relief dollars?
2. How many disasters were declared in [geographic area] during [timeframe]?
3. In which [time span] does FEMA award the most IHP relief dollars?
4. What percentage of individuals requesting aid are granted it, and how does that compare with the volume of requests?
5. What proportion of dollars awarded were allocated for Other Needs as compared to Housing Assistance in the last [time span]?
6. What are the most awarded disaster types? Is time of year a factor?
7. What proportion of dollars awarded were allocated for Other Needs as compared to Housing Assistance in the last [time span]?
8. Is more aid awarded to Other Needs, or Housing Assistance and has it changed over time?

# Design

## Dimensions

### Date Dimension

The purpose of the Date dimension is to describe time data of the business fact. It is hierarchical, meaning that the same data is represented at different granularities. For example, the day, the month, and year are all individually stored attributes in the date dimension.

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| dateId (PK) | Surrogate key to uniquely identify record in DW |
| calendarDate | The date in ‘MM-DD-YYYY’ format, the date the disaster was declared |
| calendarDayOfWeek | The day of the week |
| calendarDayOfMonth | The number day on the calendar |
| calendarMonthOfYear | The number of the month |
| calendarWeekOfYear | The number of the week in the year |
| calendarYear | The calendar year |
| fiscalQuarter | The number of the fiscal quarter (1,2,3,4) |
| fiscalYear | The fiscal year |

### Location Dimension

The purpose of the Location dimension is to describe the geography of the business fact. Like the Date dimension, Location is also hierarchical. Attributes in Location like the zip code, county, and region illustrate its hierarchy.

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| locationId (PK) | Surrogate key to uniquely identify record in DW |
| femaRegion | FEMA’s assigned region number (note: unable to implement this with the final dataset, would need to join regions from another source) |
| locRegion | Geographic region (e.g., “Southwest”) |
| locPriorRegion | Prior version of location region (slowly changing dimension type 3, decision makers could change the way regions are in the future, but likely won’t change it again) |
| locState | The state of the location |
| locCounty | The county of the location |
| locCity | The city of the location |
| locZipCode | The zip code of the location |
| locationIndex | Synthetic key populated in python to ensure ease when mapping |

### Disaster Dimension

The purpose of the Disaster dimension is to describe the disaster relating to the relief that was given. Because the nature of emergency management, records in the disaster table will likely change, and it is important to log each record with the appropriate version of data. Therefore, the Disaster dimension will use a Type 6 slowly changing dimension to ensure that all changes in the disaster dimension are tracked.

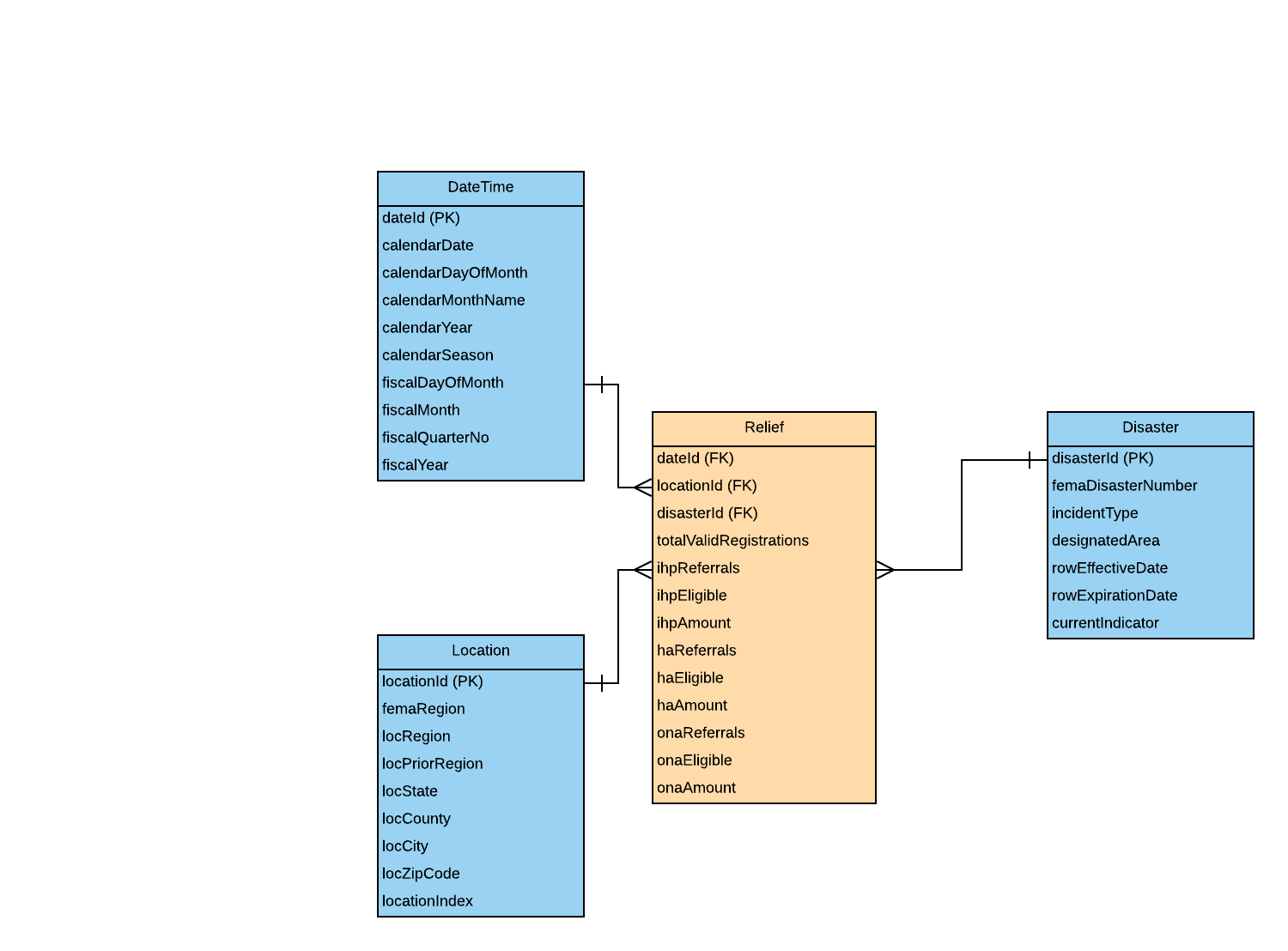
|  |  |
| --- | --- |
| **Attribute** | **Description** |
| disasterId (PK) | Surrogate key to uniquely identify record in DW |
| femaDisasterNumber | FEMA’s designated id number for the disaster |
| incidentType | Category of disaster (hurricane, flood, etc) |
| designatedArea | Geographic region encompassing disaster |
| rowEffectiveDate | The date that this record came into effect |
| rowExpirationDate | The date the record is no longer effective |
| currentIndicator | Marked as “current” if the row is the current detail |

## Facts

The purpose of the Relief table is to measure the facts surrounding the relief efforts of the IHP at FEMA.

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| disasterId (FK) | Foreign key to Disaster dimension |
| locationId (FK) | Foreign key to location dimension |
| dateTimeId (FK) | Foreign key to datetime dimension |
| totalValidRegistrations | The number of individuals who successfully registered with IHP |
| IhpReferrals | The number of referrals for the entire IH program |
| IhpEligible | The number of eligible individuals for the entire IH program |
| IhpAmount | Dollars awarded for IH program |
| haReferrals | Number of referrals for the housing assistance portion of IHP |
| haEligible | Number of eligible individuals for the housing assistance portion of IHP |
| haAmount | Dollars awarded for housing assistance portion of IHP |
| onaReferrals | Number of referrals for other needs portion of IHP |
| onaEligible | The number of eligible individuals for other needs portion of IHP |
| onaAmount | Dollars awarded for other needs portion of IHP |

## Data Warehouse Schema



# ETL Process

The ETL process encompassed the vast majority of implementation time and was performed iteratively to ensure good visual representation. The initial table creation was done directly in MS SQL Server, but all other actionable ETL items occurred within Tableau and the Jupyter IDE. The datasets were extracted directly from the government sources at FEMA.gov (links are in the References section).

The dataset was relatively complete, but represented at different levels of granularity or format, so most of the ETL was spent trying to identify unique rows and maintain the volume of data to ensure it was properly represented in the visualizations and any analysis. The ETL process was heavily documented in a Jupyter Notebook titled ‘etlIHP.ipynb’, which was submitted with this report. The Python Pandas library was consulted extensively and the following ETL systems were implemented:

* Data Profiling
  + Disaster Dimensions
  + Location Dimension
* Deduplication, Conforming
* Surrogate Key Generator (Sequences in SQL Server)
* Slowly Changing Dimensions
  + Type 2: Disaster
  + Type 3: Region
* Change Data Capture
  + Not implemented, but handled by maintaining ‘id’ and ‘lastRefresh’ fields from data sources as audit columns (see ETL document for more detail)

# Visualizations

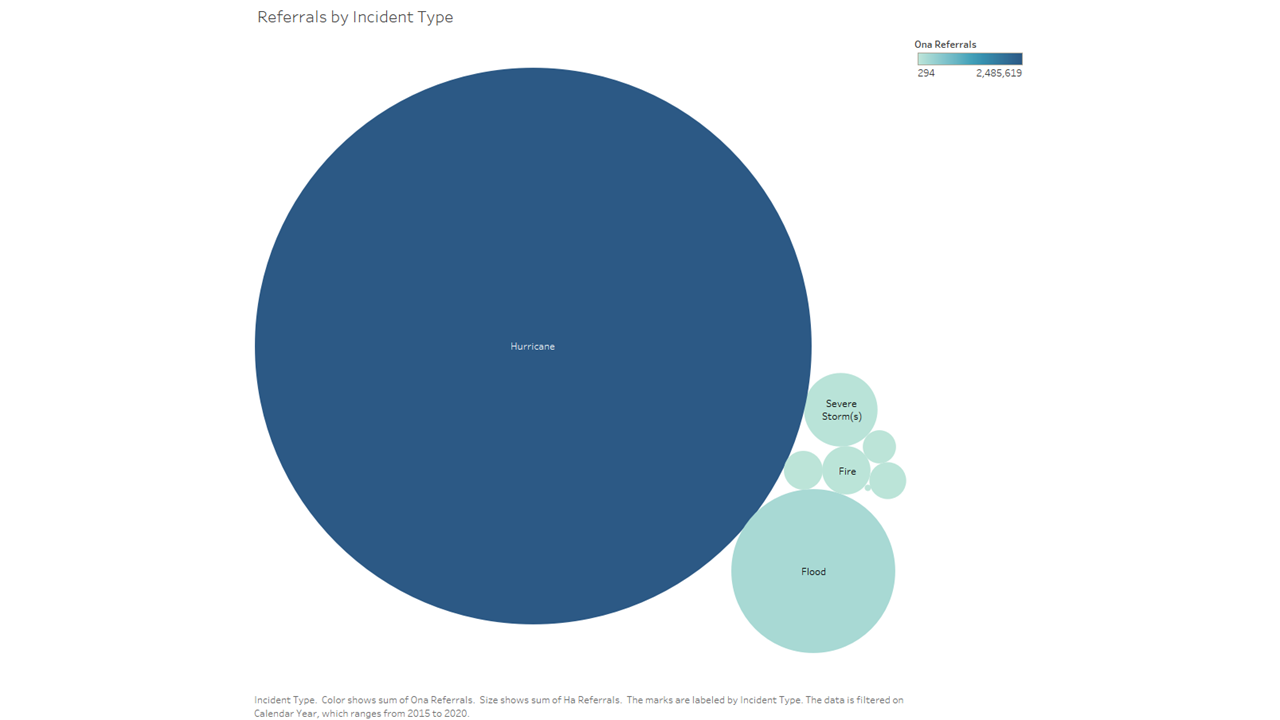


Figure 1 illustrates the number of referrals given by disaster type and proves that hurricanes account for the vast majority of referrals to the program.

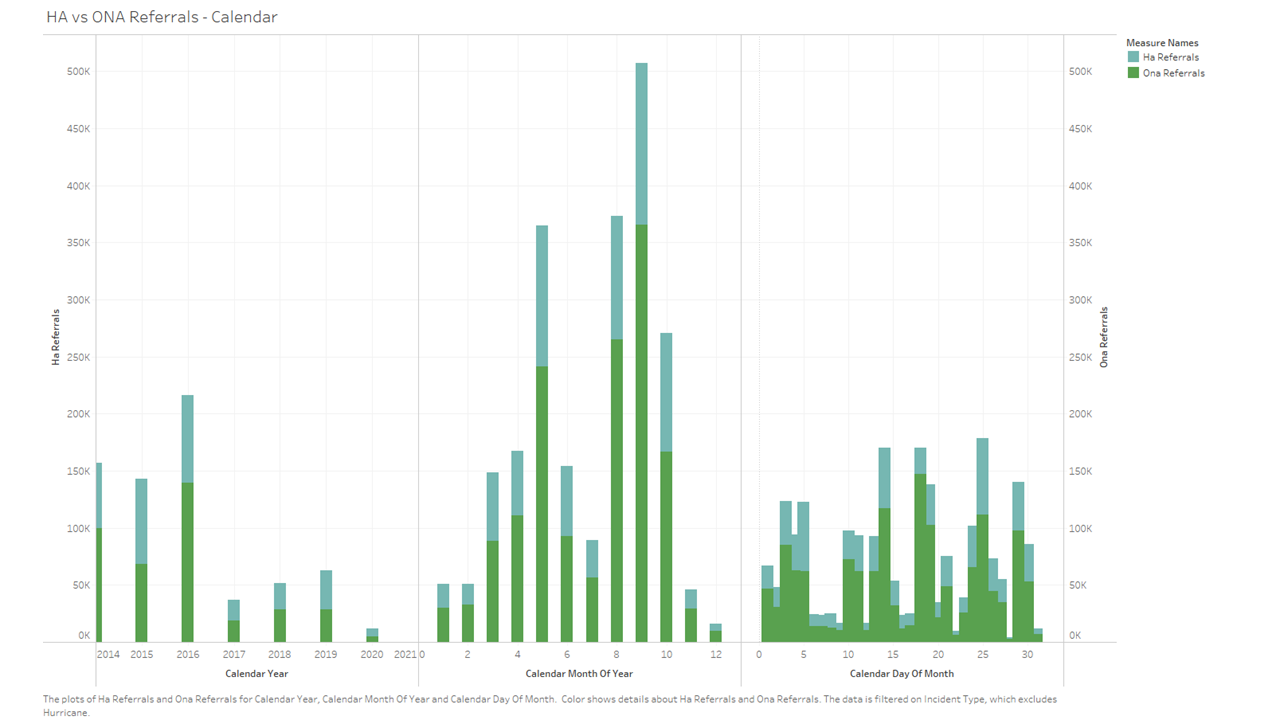


Figure 2 illustrates the trend in referrals over years, months, and days, also comparing the overall share of HA referrals (blue) and ONA referrals (green). The entire bar represents the total number of requests (filtered 2015-present)

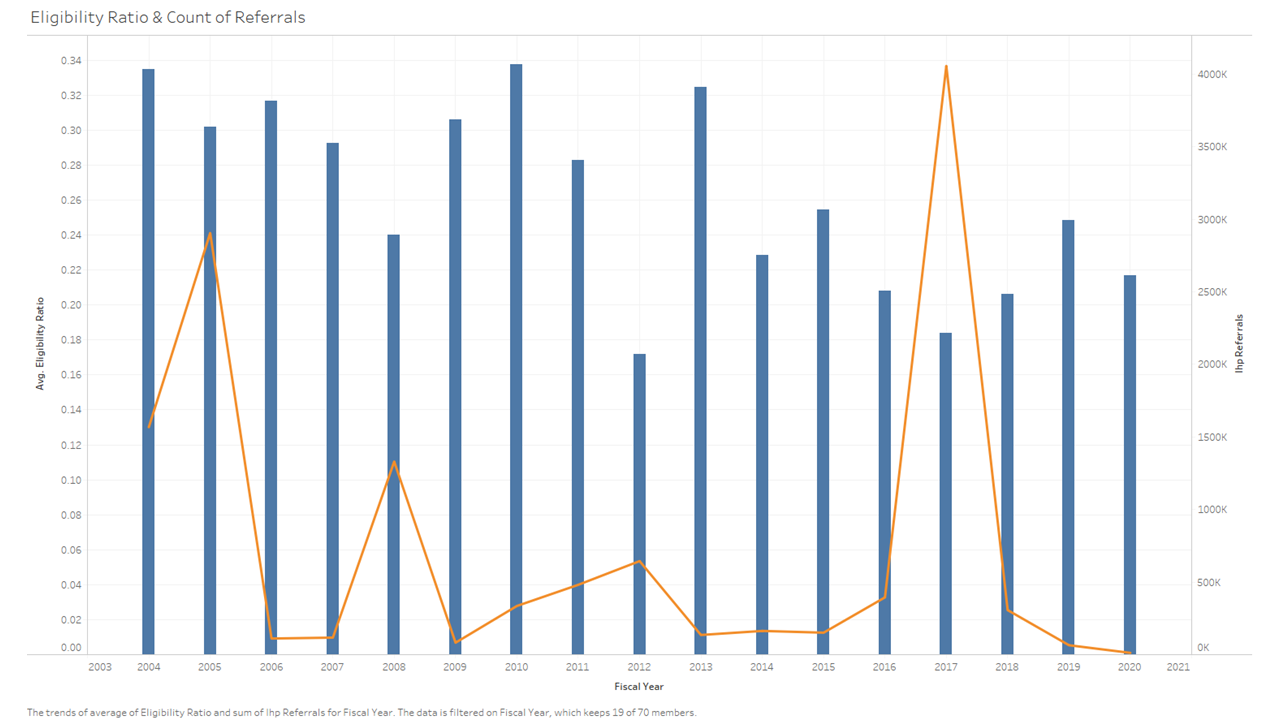


Figure 3 compares the number of referrals (requests for aid) – orange, and the average ratio of approved requests (blue). Ratio was calculated within Tableau by adding the total number of referrals and awards and dividing it by the number of awards.

Figure 4

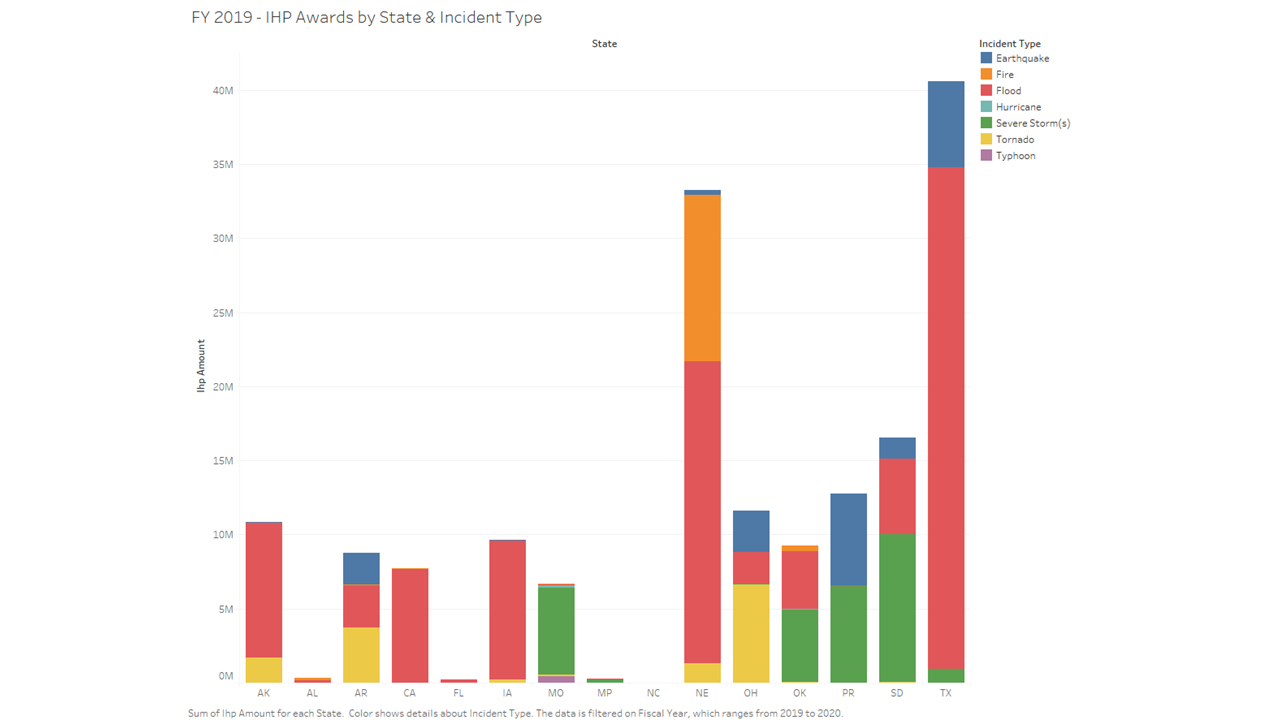


Figure 5 illustrates the total awards given for the IHP program by state in the last fiscal year. Each incident type is represented.

# References

FEMA. (n.d.-a). *OpenFEMA Dataset: Disaster Declarations Summaries—V2 | FEMA.gov*. Retrieved April 26, 2020, from <https://www.fema.gov/openfema-dataset-disaster-declarations-summaries-v2>

FEMA. (n.d.-b). *OpenFEMA Dataset: Hazard Mitigation Assistance Mitigated Properties—V2 | FEMA.gov*. Retrieved April 26, 2020, from <https://www.fema.gov/openfema-dataset-hazard-mitigation-assistance-mitigated-properties-v2>

FEMA. (n.d.-c). *OpenFEMA Dataset: Registration Intake and Individuals Household Program—V1 | FEMA.gov*. Retrieved April 28, 2020, from <https://www.fema.gov/openfema-dataset-registration-intake-and-individuals-household-program-v1>

*List of State Abbreviations (Download CSV, JSON)*. (n.d.). Retrieved April 26, 2020, from <https://worldpopulationreview.com/states/state-abbreviations/>

McKinney, W. (2017). Python for Data Analysis, 2nd Edition. OReilly Media, Inc.

Pandas. (n.d.-a). *pandas.DataFrame.drop\_duplicates—Pandas 1.0.3 documentation*. Retrieved April 29, 2020, from <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.drop_duplicates.html>

Pandas. (n.d.-b). *Series—Pandas 1.0.3 documentation*. Retrieved April 29, 2020, from <https://pandas.pydata.org/pandas-docs/stable/reference/series.html>

Pandas. (n.d.-c). *Time series / date functionality—Pandas 1.0.3 documentation*. Retrieved April 25, 2020, from <https://pandas.pydata.org/pandas-docs/stable/user_guide/timeseries.html?highlight=datetime#time-date-components>