

Data Description and Breakdown

There are two datasets being using in this project:

- [covid19-newjersey](#) : An *ongoing* data collection of the number of COVID-19 cases per New Jersey municipalities from March 25, 2020 to Present.
- [municipal.csv](#) : Basic population statistics and which county they are a part of for each NJ municipality.

At the moment, these are the two main datasets being used. I may incorporate median income for each township if I am able to find a good dataset describing this.

Let's dive into these datasets!

COVID-19 New Jersey Data

Overview

The scripts for automating and scraping the designated articles for this data is located in [another repository](#). More information about the collection, automation, and notes can be found there.

The data is collected and recorded by different counties in NJ, which is then compiled and posted in articles by [NJ.com](#).

Data Observations and Attributes

After scraping the data from the article, it is stored in a CSV file format for an ongoing compilation of the number of cases per municipality in New Jersey. The structure of the file is as such:

Date	NJ Municipality	Number of Cases
march-25-2020	allendale	4
march-25-2020	alpine	1
...

The "Date" column is structured as {month name}-{day}-{year} . The "Municipal" variable contains the official municipality name as posted in [nj_municipals.txt](#) . The "Cases" column is simply an integer indicating the number of cases.

Notes

As of April 26, all counties have been reporting township case numbers. However, some counties began disclosing the data later than other counties (eg. Atlantic and Mercer County), creating a discord in the data.

There was no publication of data for May 1, 2020 on NJ.com.

Most (if not all) counties with high rates of infection have been posting daily, but other counties have been updating a variable amount of days.

I recently found a bug of duplicate municipality names and am actively working on fixing it.

New Jersey Municipality Data

Overview

This data contains some basic information for each municipality in New Jersey.

The data is grabbed from this [Wikipedia Article](#) and stored it in CSV file named `municipial.csv`.

Data Observations and Attributes

After scraping the data from the Wikipedia entry, it is stored into a CSV file as such:

Municipal	County	Population	Type
newark	esssex	285154	city
edison	middlesex	102450	township
somerville	somerset	12418	borough
...

The attributes for this dataset is the "Municipal", "County", "Population", and "Type". The "Population" column is the estimated population of the municipality in 2017. The "Type" column contains whether this municipality is a "city", "township", "borough", etc.

Notes

I hope to expand this dataset and include a median income column for each municipality to further explore the analysis of COVID-19 cases in NJ.

Data Scripts

There are three main Python Notebook files that have code that has been used to aggregate or analyze the data above.

- [Analytics.ipynb](#) : Contains the preloaded data frames of the described datasets from above.
- [NJMunicipalFullData.ipynb](#) : Scrapes the Wikipedia entry for NJ Municipality data.
- [NJCovid19Data.ipynb](#) : Scrapes NJ.com article for COVID-19 cases for one day. I have a seperate script located in [this repository](#) to automate the collection of data daily.

More details can be found in those specific files.