**Overview of Contents of Kathryn Haglich’s Research Summer 2020**

***GitHub Link:*** [*https://github.com/hagk17/Summer2020\_Covid\_Vulnerability\_Analysis*](https://github.com/hagk17/Summer2020_Covid_Vulnerability_Analysis)

***Link to Application:***[*https://haglichk.shinyapps.io/CovidApp/*](https://haglichk.shinyapps.io/CovidApp/)

**Data**

This file contains the raw data also available in the Vulnerability Analysis folder.

For all files structured **TOWN\_CovidCases\_2020mmdd.csv**

-merged results of all the Covid cases since Jan1, 2020

● *Conf<DATE>*:

-Covid19 case count by town (the original data downloaded from Massgov, used for labelling purpose)

-For populations <50,000, <5 cases are reported as such or suppressed for confidentiality purposes

-This is updated with data up to mm/dd

● *Case<DATE>:*

-Covid19 case count by town

-Identical to Conf<DATE> except changing <5 to -999 for mapping purposes

-This is updated with data up to mm/dd

● *CRate<Date>*:

-Case rate by town

-Similar to original from Massgov except changed \* to -999 for mapping purpose

-This is updated with data up to July 8, 2020

**TOWN\_CovidCases\_20200909.csv**

-The data set described above updated with data up to September 9, 2020

**CT\_ACD2014\_2018\_vars.csv**

-Census tract level vulnerability data for Massachusetts.

-Currently only contains variables from ACS 2014-2018.

-Please refer to the README\_CT\_vars.csv when using this data.

**CT\_TOWN\_linkage.csv**

-The linkage file between census tract ID and town ID/name for Massachusetts

-Can be used for aggregating census tract level variables to town level.

*-***Note**: some tracts have multiple towns, thus these towns will have the same census tract ID, you may want to aggregate it using a better method rather than the file provided.

**TOWN\_pop\_density\_sqmi.csv**

-Population, area in square miles, and population density for all MA towns

**daily\_cases.csv**

-daily confirmed Covid infections for 40 towns March 23, 2020 - June 1, 2020

**daily\_deaths.csv**

-daily confirmed Covid deaths for 40 towns March 23, 2020 - June 1, 2020

**clean\_mmdd\_data.xlsx**

-Excel file that is the cleaned and organized output of Code\_to\_Clean\_Data.R

-Sheet 1: Aggregated daily cases and death data into one table

-Sheet 2: 40 towns’ population, min, eng gathered from daily data files

-Sheet 3: Calculated toals and proportions for town level socio-economic data

-Sheet 4: Time series representation of first dataframe

-Sheet 5: Time series representation of Covid data from third data frame

-Sheet 6: Town level socio-economic data combined with recent total covid cases

**clean\_0909\_data.xlsx**

-Data set described above updated with data up to September 9, 2020

**app\_mmdd\_data.xlsx**

-Data necessary to run the application

-Sheets 4-6 of clean\_mmdd\_data.xlsx

**app\_0909\_data.xlsx**

-Data set described above updated with data up to September 9, 2020

**Cleaning\_Function.R**

-assistant function to read, clean, aggregate, and organize the raw data to prepare for

use in analysis and application

-*Data Inputs*: TOWN\_CovidCases\_2020mmdd.csv, CT\_ACD2014\_2018\_vars.csv, CT\_TOWN\_linkage.csv, TOWN\_pop\_density\_sqmi.csv, daily\_cases.csv, daily\_deaths.csv

-*Outputs:* a list of 6 data frames (as described above in clean\_mmdd\_data.xlsx)

-can be stored in an excel spreadsheet as specified above in Code\_to\_Clean\_Data.R

-unless there are changes to the input, there is no need to edit this file

**Code\_to\_Clean\_Data.R**

-R code that runs the Cleaning\_Function and allows for input to update weekly covid cases data

-last updated with data from September 9, 2020

-*Outputs:*

-clean\_mmdd\_data.xlsx

-app\_mmdd\_data.xlsx

**fullApp.R**

-R code to run the full shiny application.

-**NOTE**: This code is an exact replica of the app code on the app’s github, but this is **NOT** the published code. In order to update the published application, use the code available on the app’s github

-*Input*: app\_mmdd\_data.xlsx

-Last updated with data from August 19, 2020.

**Correlation Exploration.R**

-R code to create correlation matrices and plots.

-*Outputs:*

-correlation\_matrices.xlsx

-PropCorrs.png

-RawCorrs.png

**correlation\_matrices.xlsx**

-Correlation matrices for all census variables and most recent covid totals MA towns

-Last updated with data from August 19, 2020.

-Sheet 1: Correlation matrix for raw variable

-Sheet 2: Correlation matrix for proportional variables

**correlation\_matrices\_colored.xlsx**

-Same as correlation\_matrices.xlsx except correlations are color coded

-Note: Colors were added by hand.

**PropsCorrs.png**

-Picture of correlation between proportional census & total covid cases variables.

**RawCorrs.png**

-Picture of correlation between raw census and total covid cases variables.

**InitialModelExploration.R**

-Rough experimental models for the census and covid variables.

-Mostly linear regressions with logistic transformations.

**GitHub and Application Instructions**

-Instructions on how to upload/download from GitHub, and to update the application.

**File Overview.docx**

-File overview of data originally from Vulnerability Analysis.