Boston Area Research Initiative ● Northeastern & Harvard Universities
E-Mail: bari@northeastern.edu ● Web: www.bostonarearesearchinitiative.net

# COVID-19 Weekly and Daily Cases by Town in Massachusetts (v. January 2021)

## Overview

This document describes the structure of the COVID-19 Weekly and Daily Cases by Town in Massachusetts data set, which contains counts of weekly and daily new positive infection counts and cumulative sum since the start of the pandemic for each one of the 351 in Massachusetts. The data span is April 14th, 2020 to January 21st, 2021 for the weekly cases and April 1st, 2020 to January 19th, 2021 for the daily cases (Note: The dates represent the publishing date for reports released at https://www.mass.gov/info-details/covid-19-response-reporting). The datasets included are *Weekly_Town_COVID19_MA.csv* and *Daily_Town_COVID19_MA.csv*.

The data released by the Department of Public Health (DPH) is in *.doc format between April 14th and May 20th, followed by *.pdf format until the present time (January 21st is the last date released in this dataset). We have extracted this content and reorganized it in datasets that are in a *.csv format to be more accessible to researchers and analysts.

The original data from the DPH included only cumulative weekly counts and rates before May 20th. As of May 27th, DPH released a new Weekly Public Health Report, which included the City/Town data and more detailed information, including count, rate, total persons tested, tested rate, percent positivity. They have included further variables since then as well. This release, however, includes only the original variables (i.e., counts and rates).

To create the daily town measures, we are using the *Weekly_Town_COVID19_MA.csv* and the daily COVID-19 data per county in Massachusetts (raw data in *xlsx format from https://www.mass.gov/info-details/archive-of-covid-19-cases-in-massachusetts) with an imputation method detailed herein. Because the state began publishing town-level data on April 14th, an extension of this technique is used to have complete data starting at April 1st.

## Table of Contents

1. COVID dataset information ..... 3
1.1. Description of Contents ..... 3
1.2. Summary of variables ..... 4

# 1. COVID dataset information

## 1.1. Description of Contents

The weekly data for the 351 towns in Massachusetts is contained in a dataset *Weekly_Town_COVID19_MA.csv* consisting of 14,391 rows (i.e., 351 towns * 41 weeks) and four columns, during the period April 14th, 2020 to January 21st, 2021. The daily data is organized in a dataset *Daily_Town_COVID19_MA.csv* consisting of 102,141 rows (i.e., 351 towns * 291 days) and five columns during the period April 1st, 2020 to January 19th, 2020.

*Note:* There are situations where the weekly count is negative (191 of 14,391 rows). This happens due to various errors in reporting. It is important to note that the state changed the way of counting total numbers of infections according to the newest released epidemiological information on COVID-19.

To create the daily town measures, we are using weekly infection cases at the town level together with daily infection cases at the county level. The steps considered are the following:

- Calculate the weekly sum of infected cases per county (from the weekly cases data per town);
- Calculate percentages of cases per town from the sum per county (e.g., if county X includes two towns, A and B, and town A has 100 infections per week April 14th – April 21st, while town B has 900 infections per week April 14th – April 21st, then town A has 10% of the infections while town B has 90%);
- Extract the daily infected cases per town from the daily infected cases per county by using the percentages from the weekly counts (e.g., If county X has 200 infections on Day 14 April, then town A will have 20 infections and town B will have 80 infections);
- To have a full month of analysis in April, the percentages from April 8th – April 13th (from the Department of Public Health (DPH) reporting on 14 April) are used for April 1st – April 7th.

Because we do not have data to calculate values before April 1st, the daily cases column will have the same value as the cumulative sum column.

*Note:* There are situations where the daily count is negative (4,216 rows from a total of

102,141 rows). This happens due to various errors in reporting. It is important to notice that the state changed the way of counting total numbers of infections according to the newest released epidemiological information on COVID-19. This can also happen due to the calculations and imputations of town daily values from the county daily values.

### 1.2. Summary of variables

Weekly data:

- *Town* is the town on which the covid cases are calculated.
- *weekd* is the start of the week for which the cases are calculated.
- *total_cases* is the cumulative sum of cases (e.g. starting from the first week until the mentioned week).
- *weekly_cases* is the count of covid cases per week (7 days).

Daily data:

- *Date* is the day for which the cases are calculated.
- *County* is the county on which the covid cases are calculated.
- *Town* is the town on which the covid cases are calculated.
- *covid_cases_cum* is the cumulative sum of cases (e.g. starting from the first day until the mentioned day).
- *daily_cases* is the count of covid cases per day.