

Exploring DC's COVID-19 Data

Copyright Thomas H. Woteki, All Rights Reserved

May 15, 2020

Introduction

This report explores the incidence of the COVID-19 disease in Washington DC and surrounding counties using two publicly available sources of data. One source is the DC Government. All the data pertaining to DC is from DC's official corona virus web site. The other source is data published by the NY Times and used in some of its reporting. The last section of the report provides more information about the sources of data.

Section 1 of the report compares the incidence of COVID-19 in DC to the VA and MD counties which surround it. Section 2 discusses testing and positive cases in DC as a whole. Section 3 summarizes the incidence of COVID-19 by Ward. Section 4 describes the sources of data.

Some highlights of this report are:

- The total number of cases in DC is comparable to Montgomery and Fairfax counties, all three of which are substantially lower than Prince George's. However, when adjusted for population DC has a substantially higher number of cases per 100 hundred thousand population than all counties except Prince George's.
- As of the date of this report, DC has reported that 35,532 tests have been performed of which 7,042 returned positive. DC does not report whether any individuals may have been tested more than once.
- The data indicate that the number of tests conducted per day has been increasing but we cannot say whether this trend will continue.
- While the number of tests per day has been increasing, the number of positive cases shows a favorable downward trend as does the positive rate per day. The positive rate is the number of reported positives per day divided by the number of reported tests per day.
- The number of reported positives in Ward 6 is substantially lower than the positives reported for Wards 1, 4, 5, 7 and 8 and about 200 hundred more than reported for Wards 2 and 3.

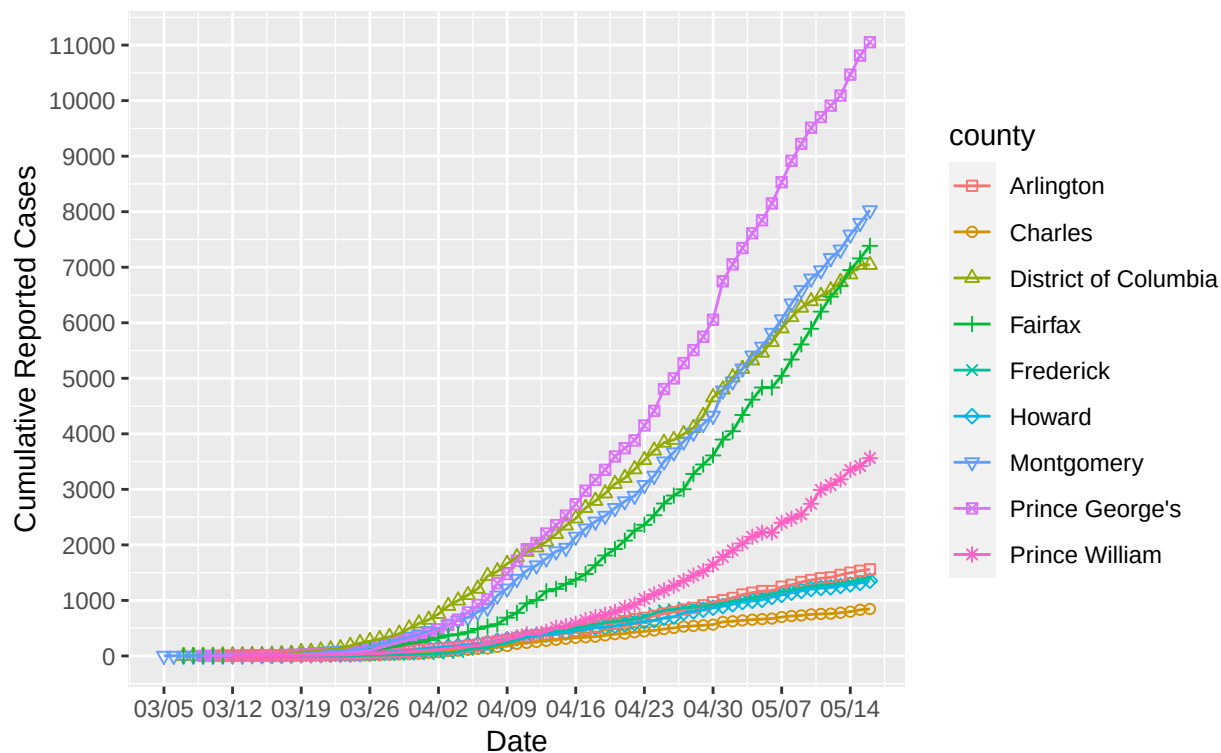
1. DC and Surrounding Counties

This section provides some context for what's happening in DC by comparing it to the surrounding counties in VA and MD. The DC numbers are from DC. The county numbers are from the NYTimes source.

The first graph shows the growth in the number of reported positive cases of COVID-19. In terms of absolute numbers DC and its surrounding counties seem to break into four groups. However, the plot should be interpreted cautiously. It does not take into account either population or testing. More populous counties are likely to have a greater number of positive cases. Similarly, the greater the rate of testing in a county the greater the number cases likely to be found.

Cumulative COVID-19 Cases, DC Area Counties

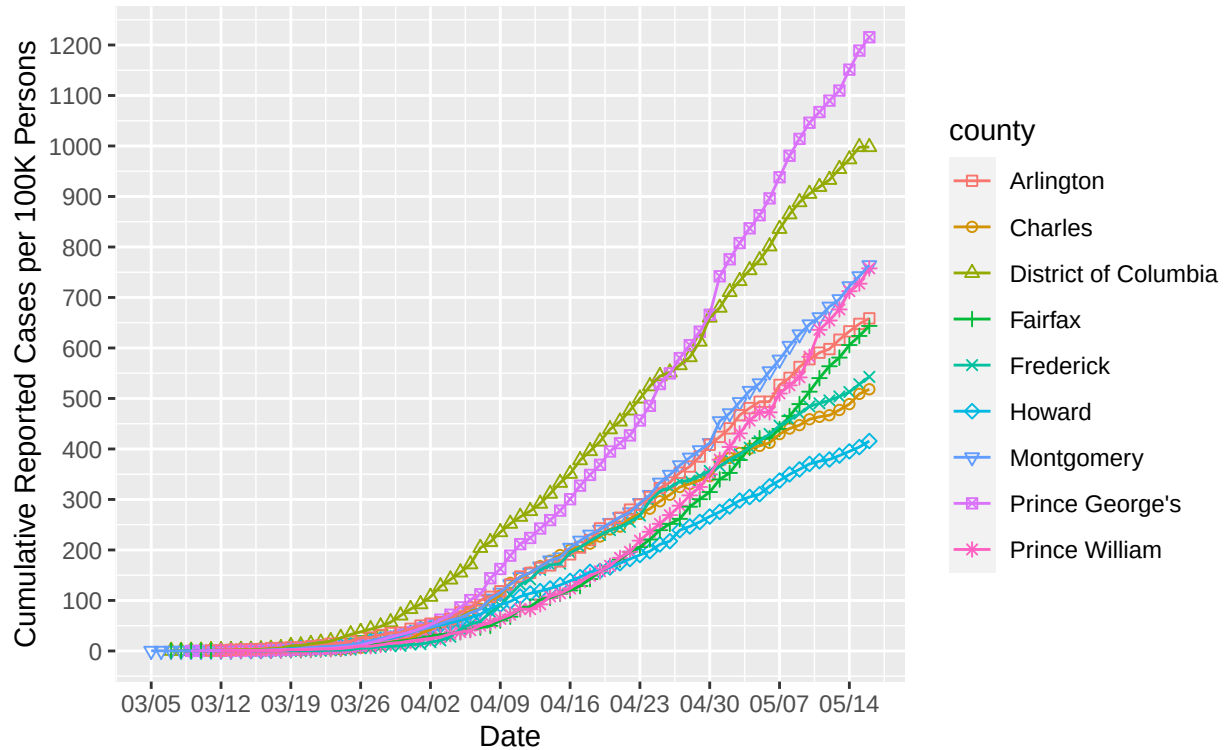
Data Sources: DC Govt for DC, NYTimes for Counties



The following graph adjusts for differences in population using reported cases per 100 thousand population. The population data are taken from official open source US Census Bureau data. DC and Prince George's have around 1,000 to 1,200 cases per 100 thousand, respectively, followed by the remaining counties ranging from around 400 to 750 cases per 100 thousand population. As above, this chart does not account for testing.

Cumulative COVID-19 Cases Per 100K Pop., DC Area Counties

Data Sources: DC Govt for DC, NYTimes for Counties



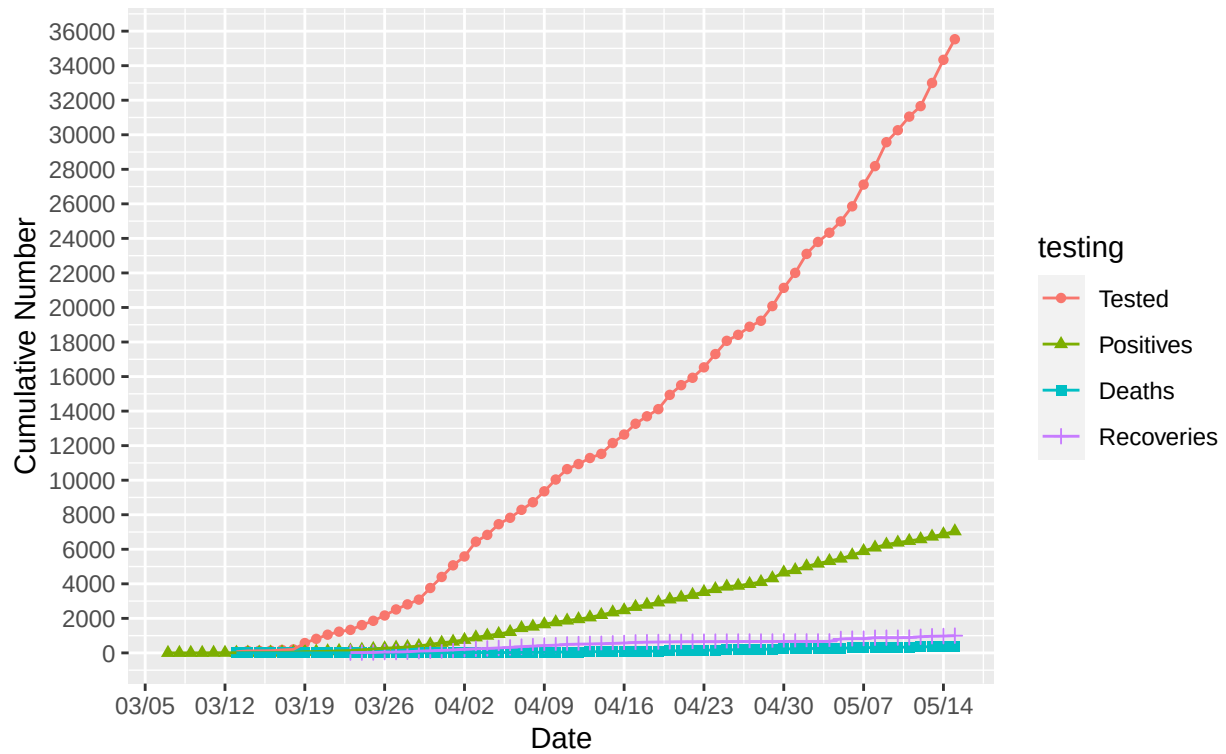
2. DC Overview

This section focuses on DC at large using official DC Gov't data as the only source.

The following chart effectively duplicates a chart that appears on the DC Gov't website. It shows the cumulative number of tests that have been performed, the positive cases based on those tests, and reported deaths and recoveries. As of the latest date on the chart, there have been approximately 35,532 tests administered of which approximately 7,042 have resulted positive.

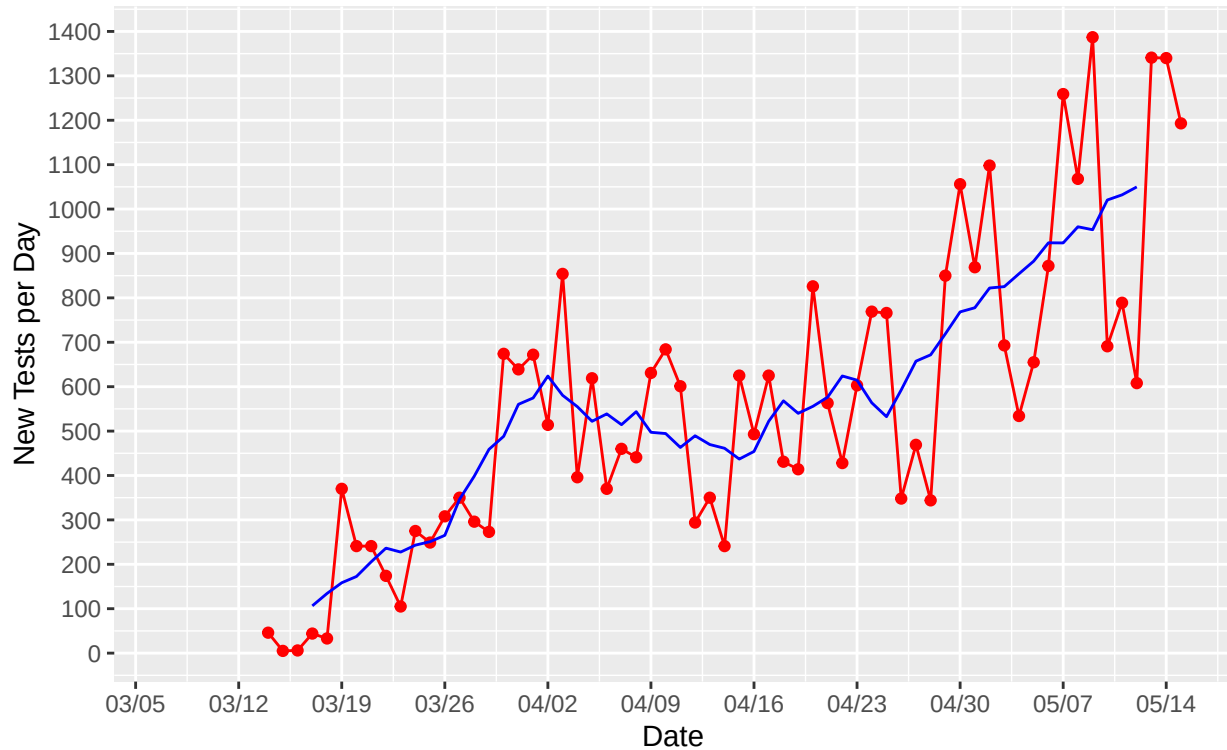
DC COVID-19 Overview: Cumulative Numbers

Source: Official DC Gov't Data



Rather than cumulative numbers, the following graphs respectively show the number of tests and the number of positives per day.

DC: New Tests per Day
Source: Official DC Gov't Data

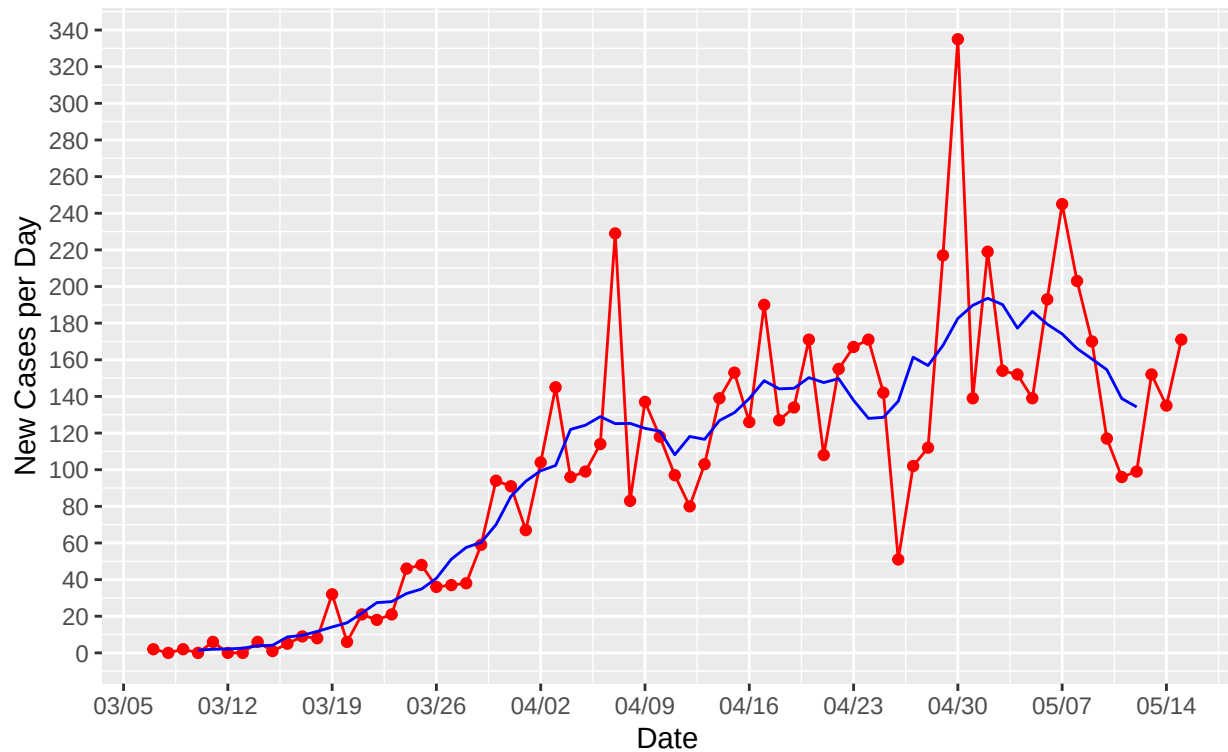


Viewing the above, the data indicate that starting in the last week of April the number of tests conducted per day has been increasing. The blue line is a 7-day moving average. April 12th, 19th and 26th were Sundays suggesting there is a drop in testing around the weekends.

In line with the increased incidence of testing, the following graph, which also includes a moving 7-day moving average, suggests an increasing number of positives per day until about the beginning of May when the trend is moving slightly downward. This is also seen in the other plot below.

DC: New Positive Cases per Day

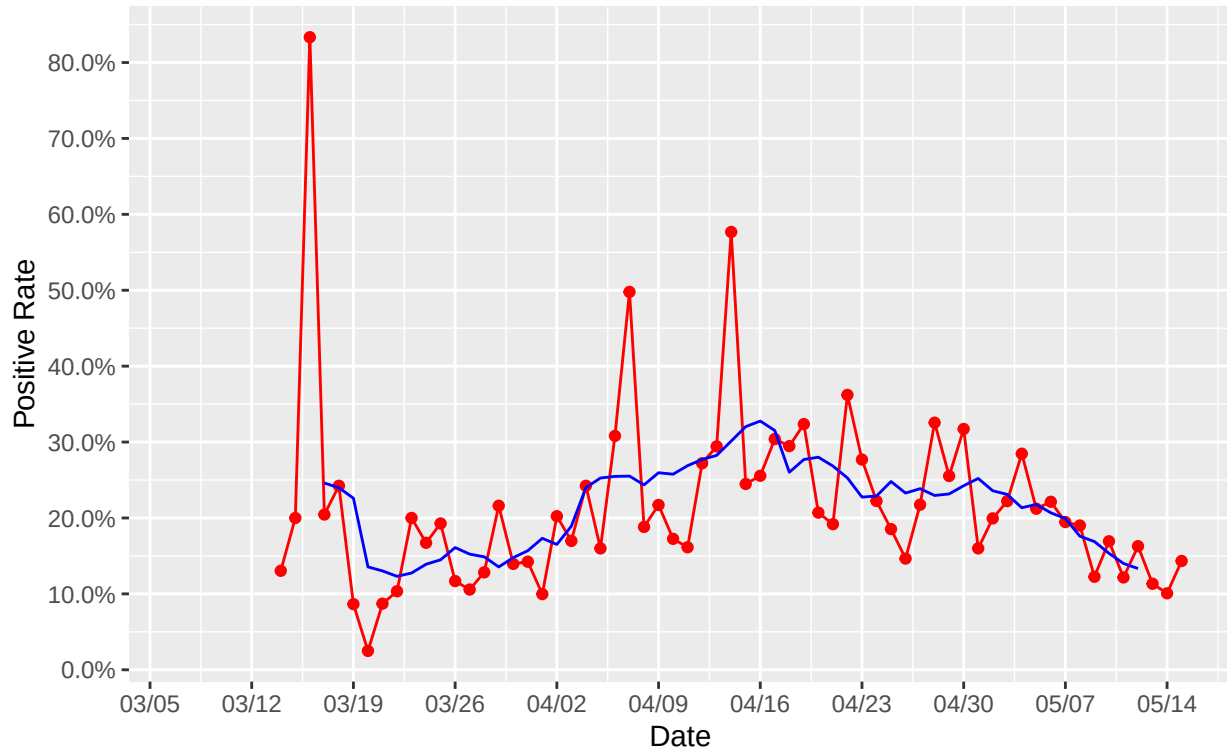
Source: Official DC Gov't Data



The following graph shows the the positive rate, that is the number of reported positives per day divided by the number of reported tests per day. This calculation is an approximation of the true positive rate.

DC: Positive Rate per Day

Source: Official DC Gov't Data



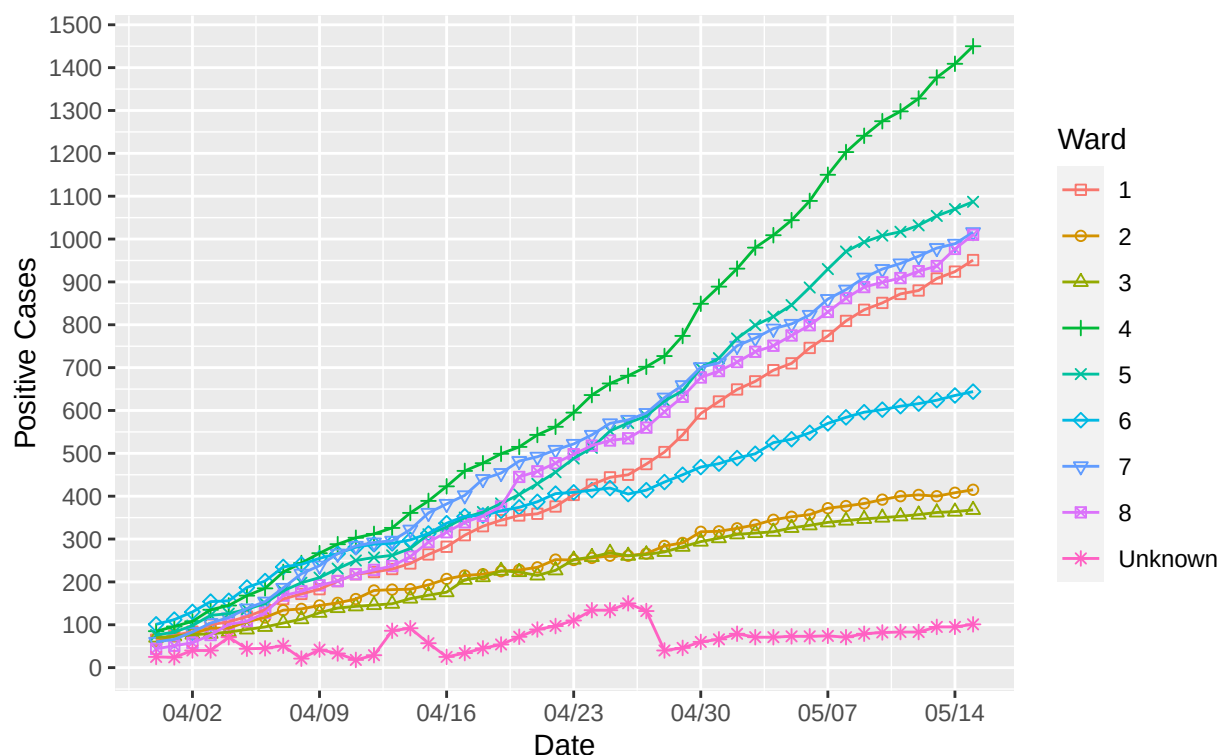
On March 16th 6 tests were performed of which 5 were positive. On April 7th 460 tests were performed of which 229 were positive. And on the 14th there were 242 tests with 139 positives. Aside from these seemingly unusual peaks, for which I do not have additional information, there are indications in the latter part of the 7-day moving average, from around April 16th to the last date on the chart, that the positive rate is declining. However, without additional information as to who is being tested and how representative they are of the DC population, this seeming favorable trend should be interpreted very cautiously.

3. DC Wards

The following graph shows the cumulative number of positive cases per DC Ward. (DC does not provide number of tests performed by Ward in downloadable form.)

Cumulative Positive Cases Reported per DC Ward

Source: Official DC Govt.Data



4. Data Sources

The DC Govt. data are downloadable from <https://coronavirus.dc.gov/page/coronavirus-data> as a Microsoft Excel file. The DC data includes data on testing: number of individuals tested, number of positives etc. Parts of the data are broken down by Ward, race, ethnicity and some other factors. The DC site presents a number of graphs and charts related to the testing numbers.

The NYTimes data are available at <https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv> as a comma separated file that can be opened in Excel. The NYTimes has data at the state and county level including the number of reported cases and deaths for each state and US territory including DC. The NYTimes data are acquired from a variety of sources, typically a state's web site via "web scraping". I have compared the NYTimes data to the official DC data and they appear to be consistent. The NYTimes data is typically a day behind the official DC data. Hence I assume the NYTimes state and county level data are pretty reliable. I use the county level data to compare DC to the VA and MD counties that surround it.

All of the data relating to DC presented in this report are based on the official DC Gov't data.