Exploring DC's COVID-19 Data

Copyright Thomas H. Woteki, All Rights Reserved

May 5, 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. The other is data published by the NY Times and used in some of its reporting.

The DC Govt. data are downloadable from https://coronavirus.dc.gov/page/coronavirus-data (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 (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[1]. 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.

This report uses a combination of DC and NYTimes data to complement the presentation on the DC site. All of the data relating to DC presented here are based on the official DC Gov't data.

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 state 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 three groups: Prince George's county followed by DC and the other MD counties and thirdly the VA counties. 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
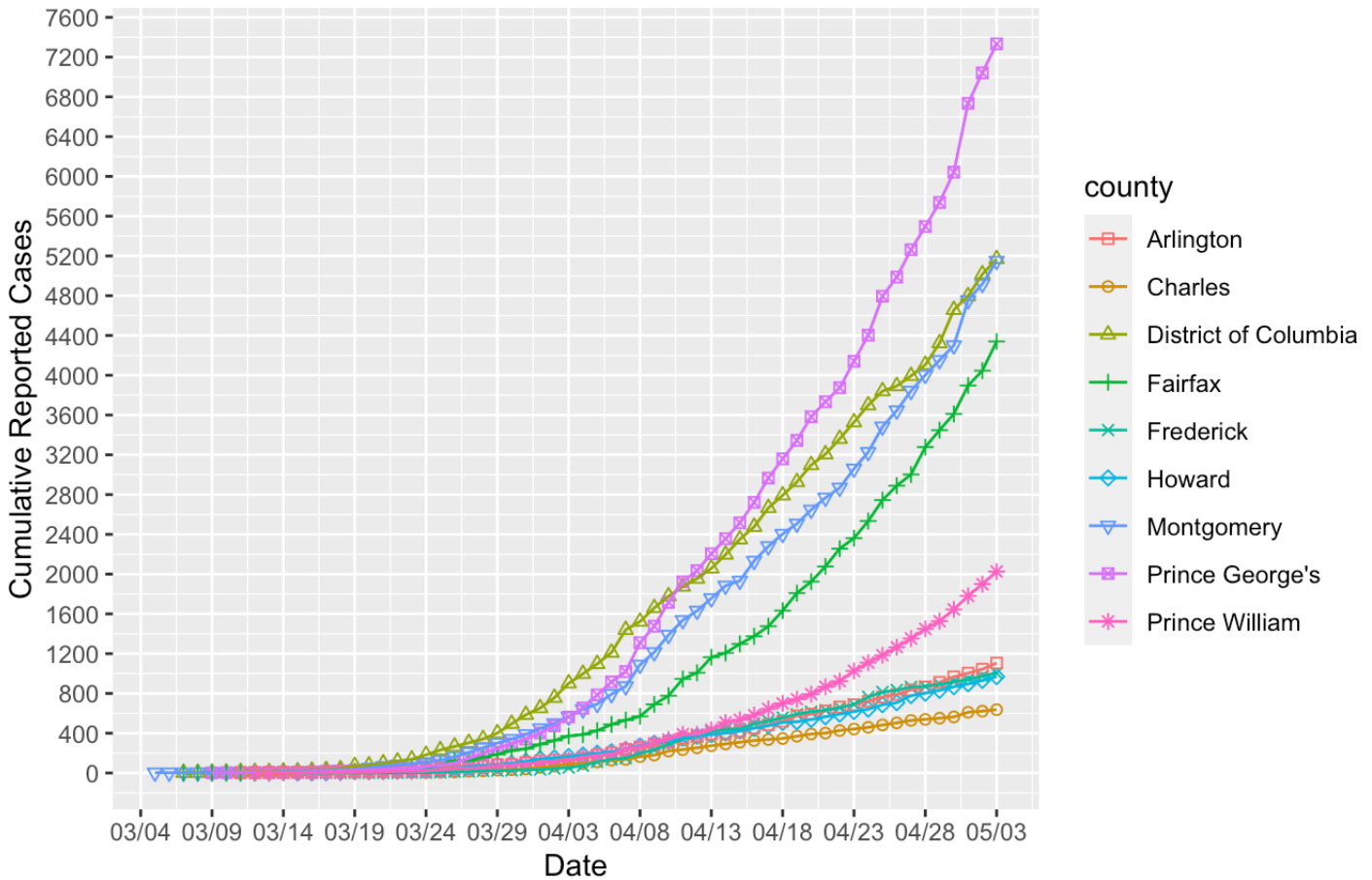
The following graph adjusts for differences in population using reported cases per 100 thousand population. The population data are taken from official open source US Census Bureau data. Unlike the previous chart, the counties now appear to be in two groups: DC and Prince George's with 700 to 800 cases per 100 thousand population followed by the remaining counties ranging from 300 to 500 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
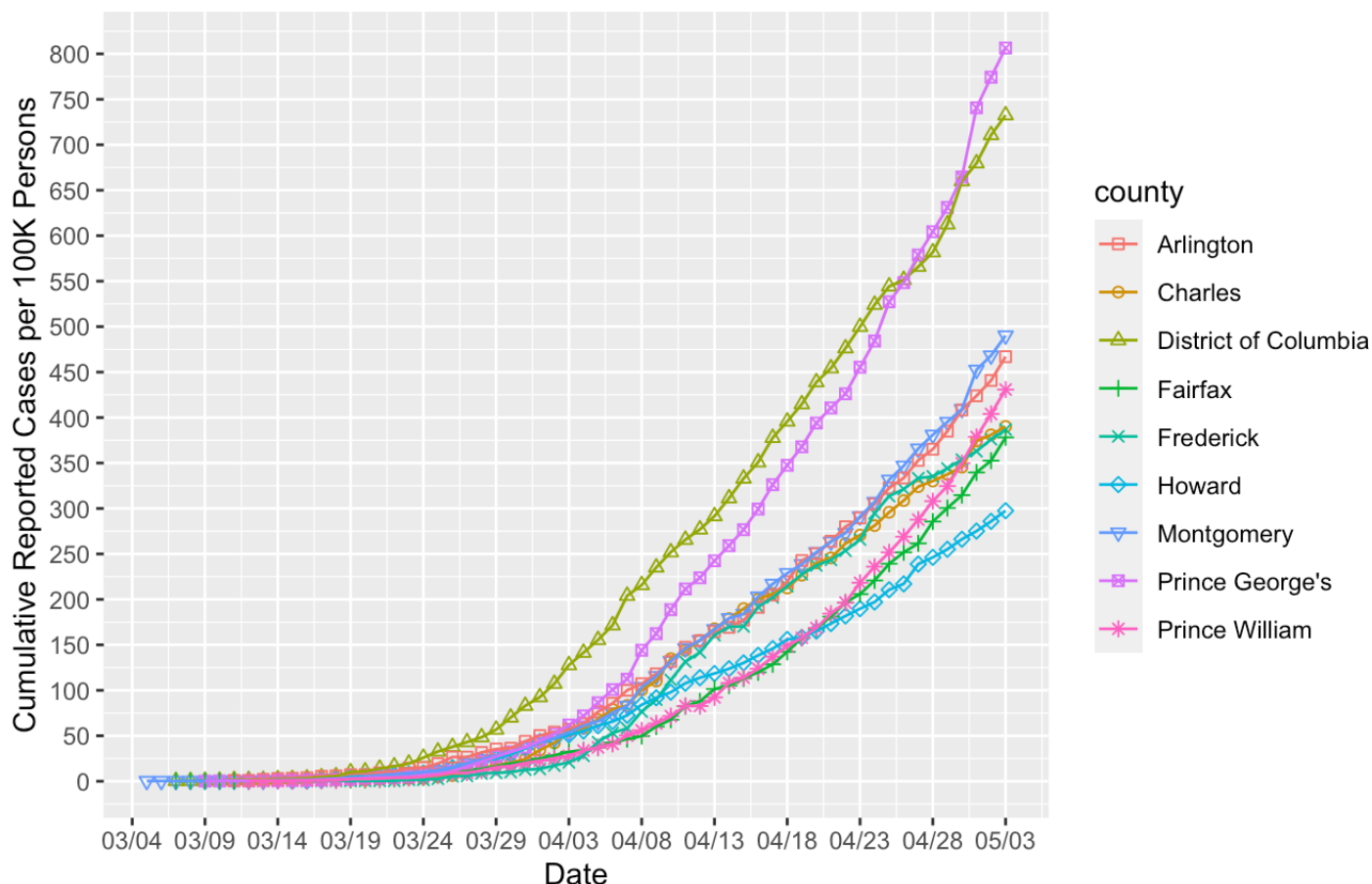
DC Overview

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 the resulting deaths and recoveries. As of the latest date on the chart, there have been approximately 23,795 tests administered of which approximately 5,170 have resulted positive.

DC COVID-19 Overview: Cumulative Numbers
Source: Official DC Gov't Data
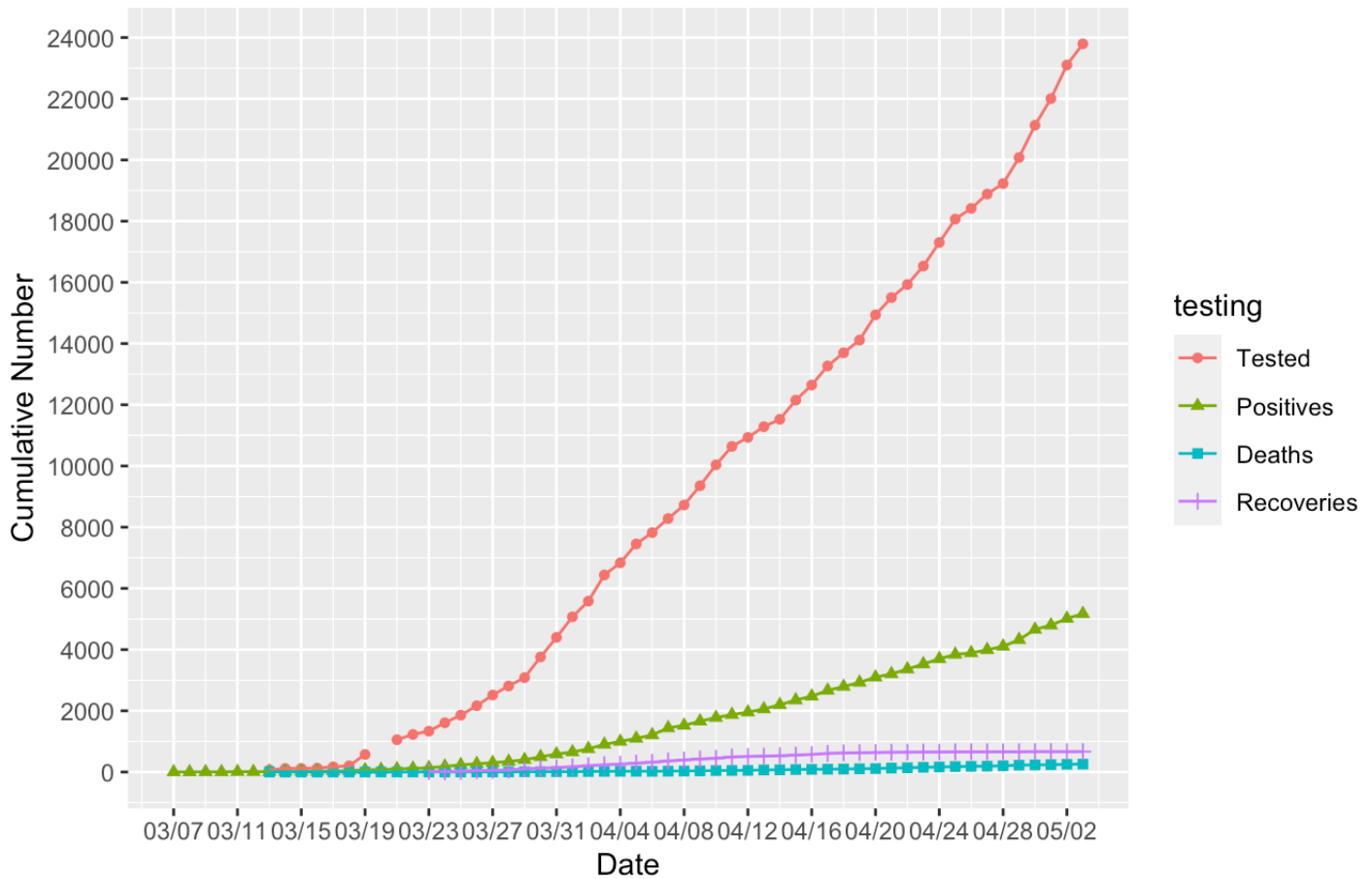
Rather than cumulative numbers, the following graphs respectively show the number of tests and the number of positives per day. Viewing the first graph, the data indicate that the number of tests being performed per day is increasing. April 12th, 19th and 26th were Sundays suggesting there is a drop in testing around the weekends.

DC: New Tests per Day
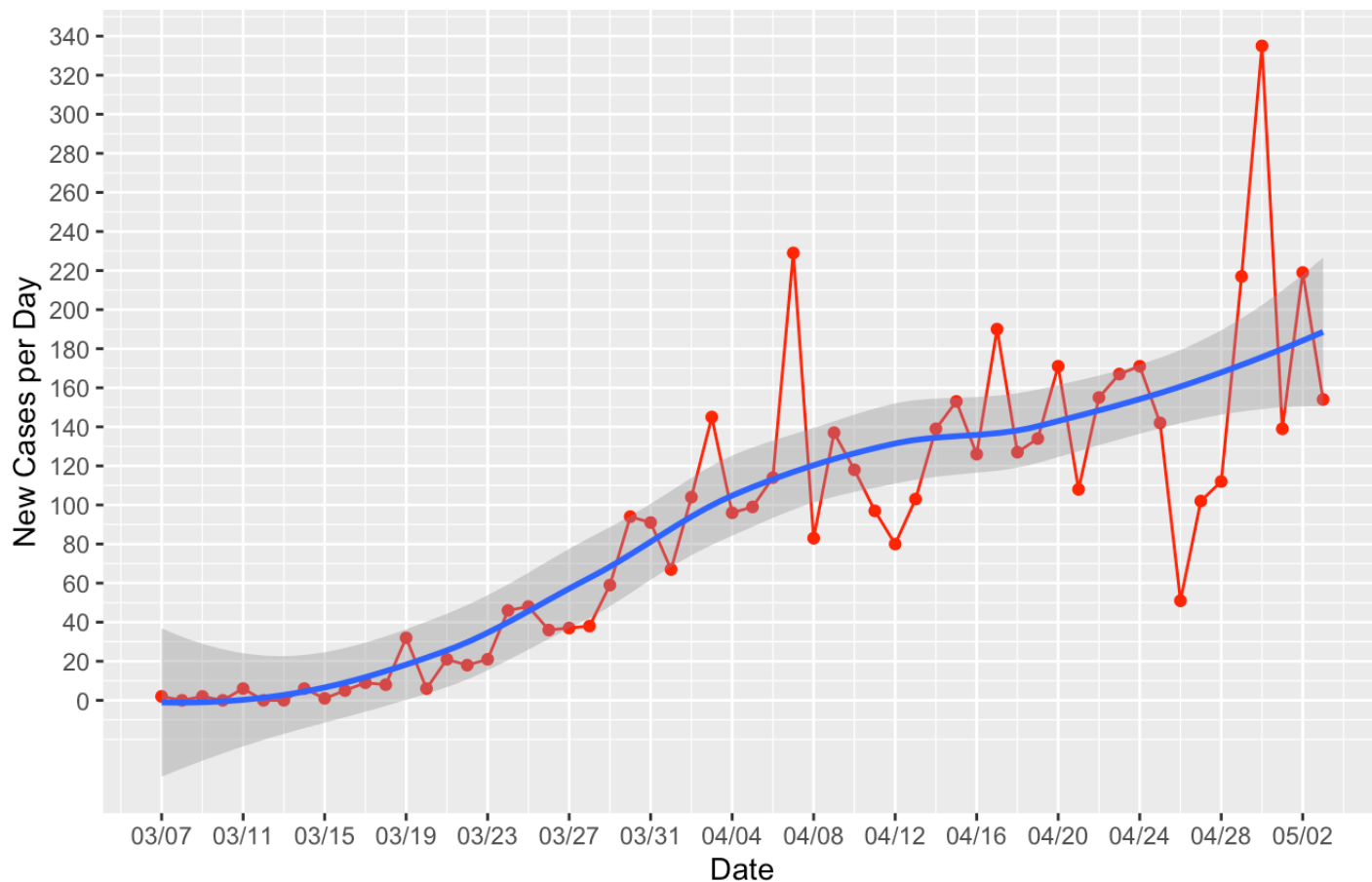Source: Official DC Gov't Data
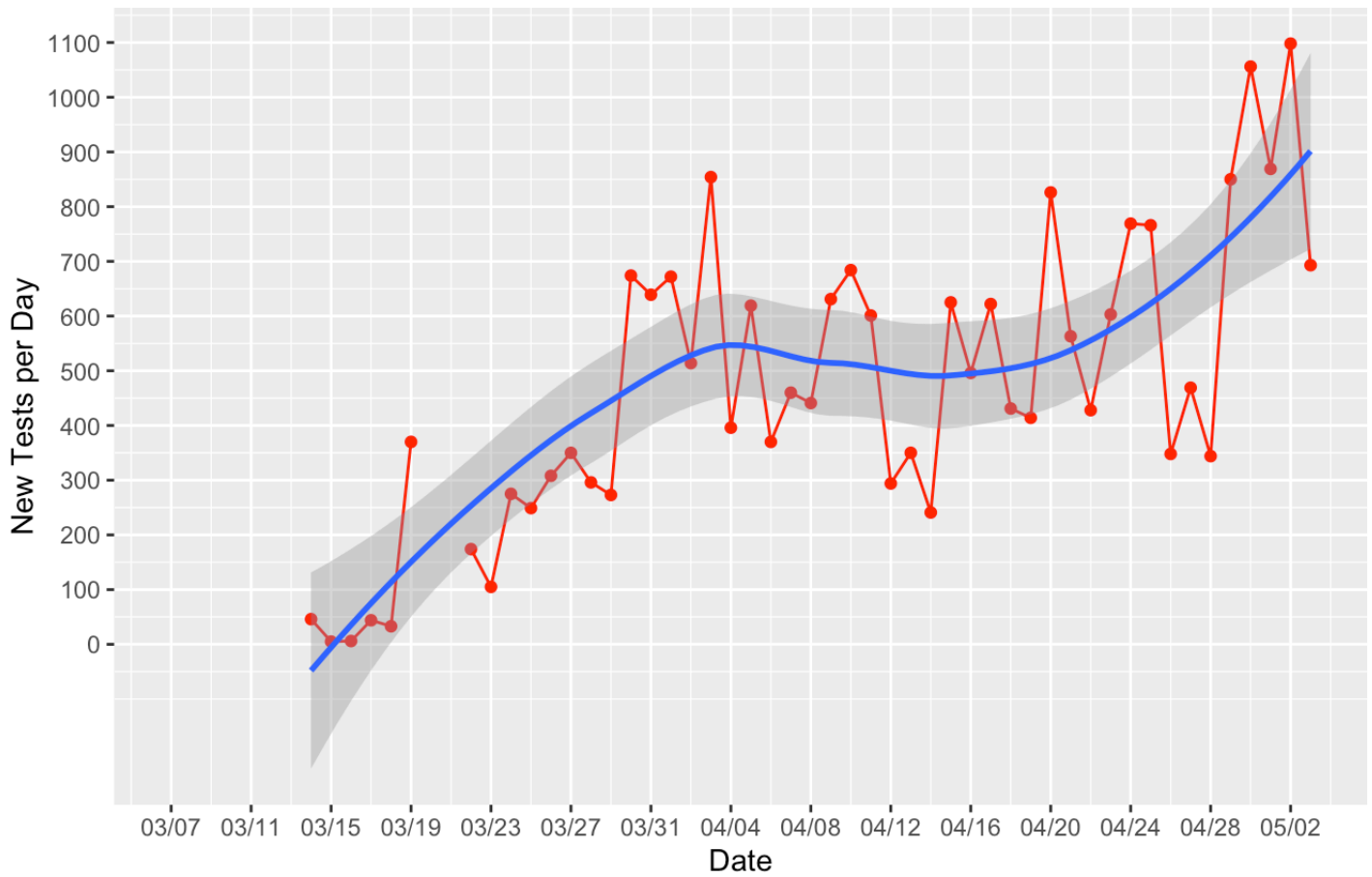
In line with the increased incidence of testing, the following graph shows an increasing number of positives per day.
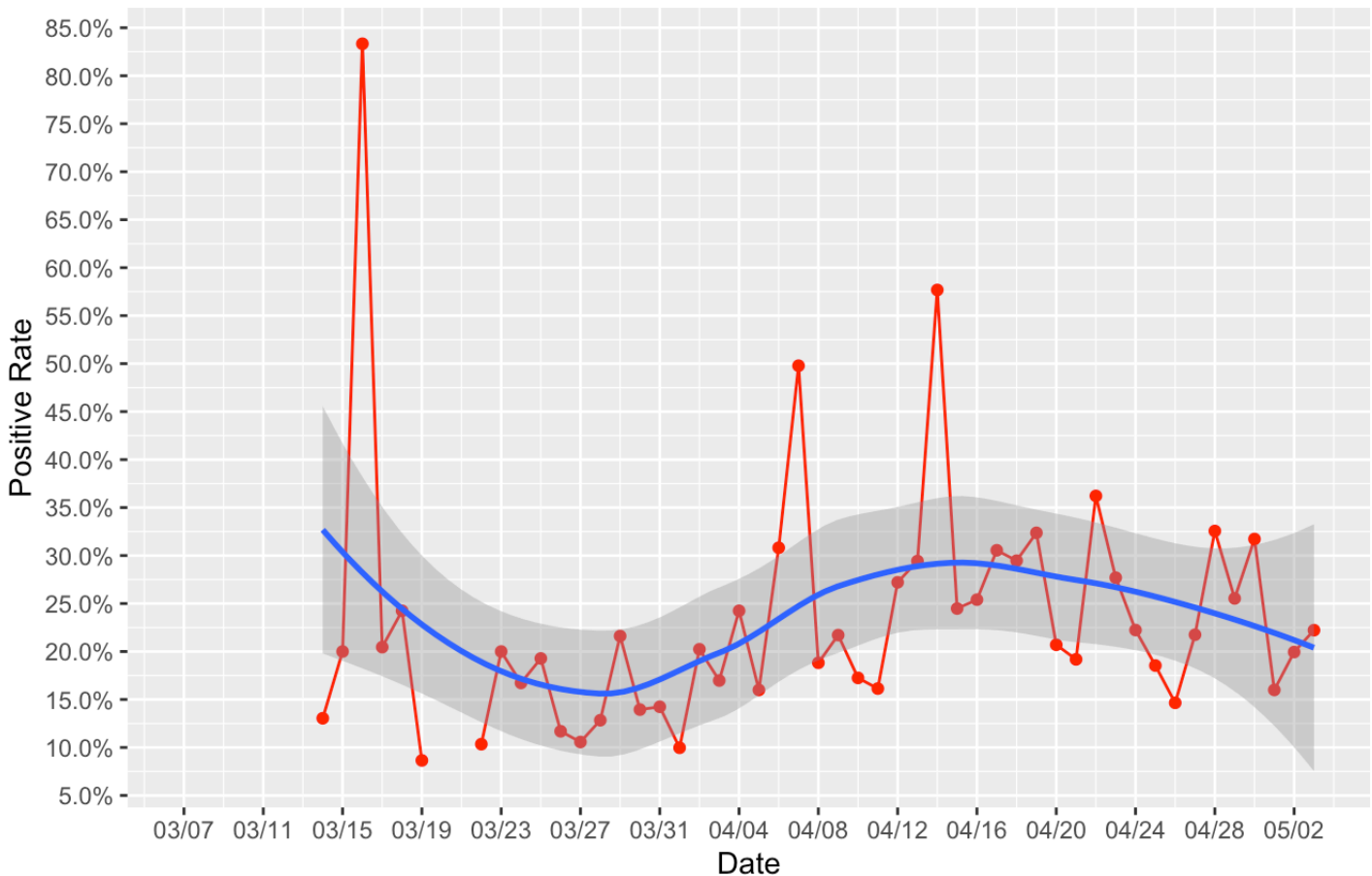
DC: New Positive Cases per Day
Source: Official DC Gov't Data

The following graph shows the the positive rate, that is the proportion of tests that result in a positive.

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 curve, from around April 16th to May 3rd, 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.

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
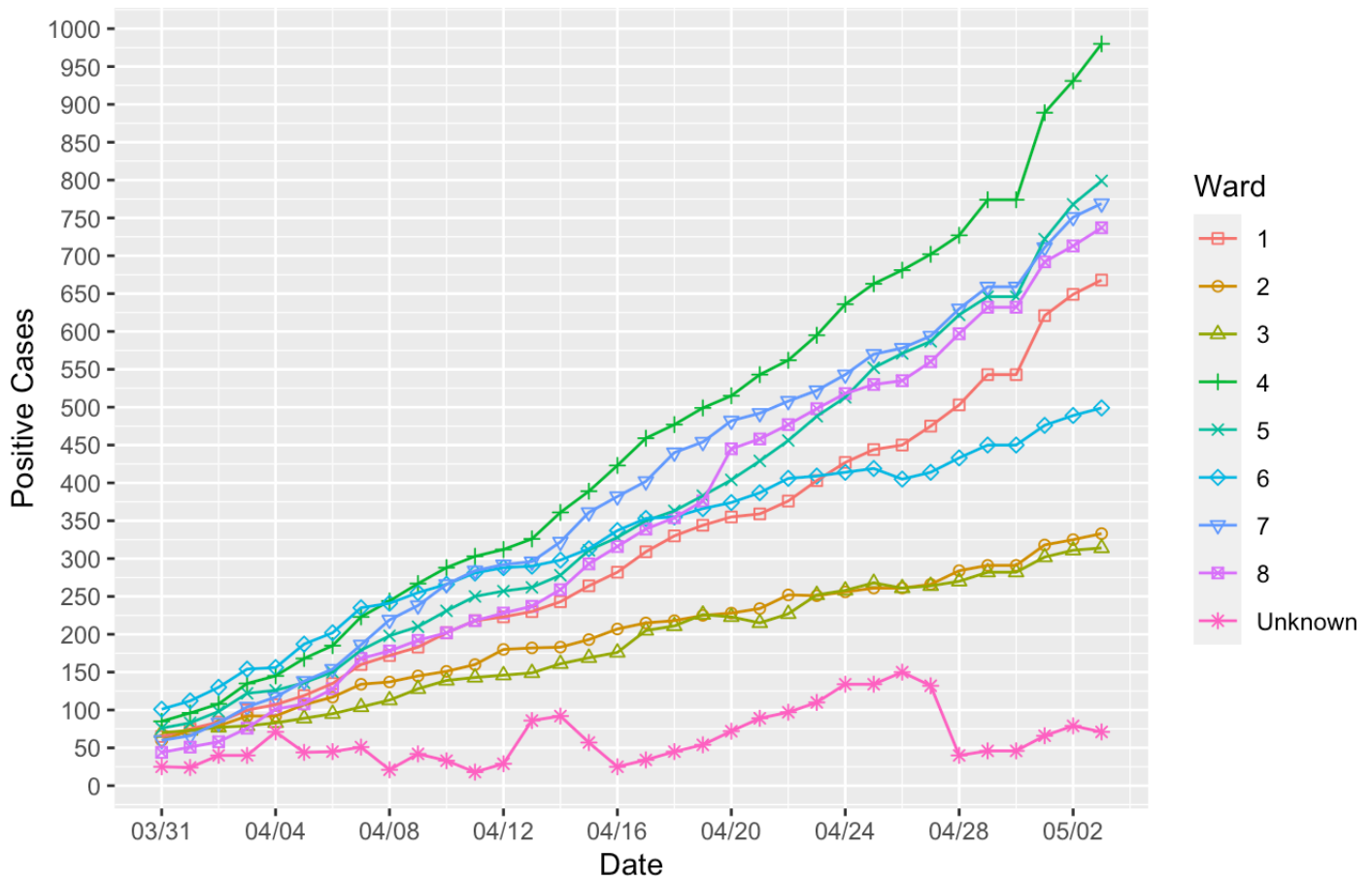
1. I have not included the comparison in this report.↵