

An Visual Analysis of Covid-19 Cases Ontario

Alexei Marcilio

April 9, 2020

The government of Ontario provides compiled data from public health units on confirmed positive cases of COVID-19. I am using that data to produce a quick visualization of the number of cases and fatalities in Ontario. There's only a few columns available, including an age range, the outcome, and the reporting city so there's no value in doing any machine learning analysis.

The data can be found here:

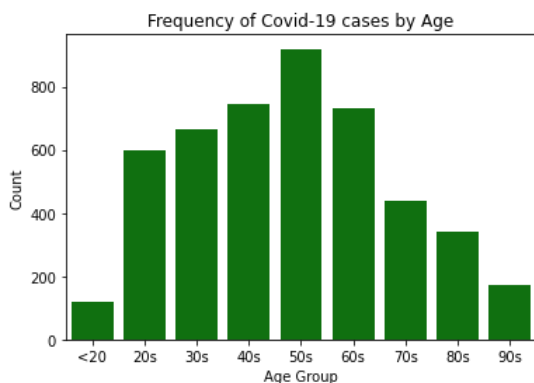
<https://data.ontario.ca/dataset/confirmed-positive-cases-of-covid-19-in-ontario/resource/455fd63b-603d-4608-8216-7d8647f43350>
<https://data.ontario.ca/dataset/confirmed-positive-cases-of-covid-19-in-ontario/resource/455fd63b-603d-4608-8216-7d8647f43350>

```
1 # Libraries
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import matplotlib.dates as mdates
5 import matplotlib.ticker as ticker
6 import numpy as np
7 # required python3 -m pip install seaborn
8 import seaborn as sns

1 # Import data
2 data = pd.read_csv("covid.csv")
```

Cases by age

```
1 # pandas dataframe
2 df = pd.DataFrame(data)
3 isknown = df['Age_Group'] != 'Unknown'
4 # Sort Age Groups ~ categorical
5 df['Age_Group'] = pd.Categorical(
6     df['Age_Group'],
7     ['<20', '20s', '30s', '40s', '50s', '60s', '70s', '80s', '90s'])
8 df = df[isknown].sort_values(['Age_Group'])
9 df['Age_Group'] = df['Age_Group'].astype('category')
10 fig = sns.countplot(df['Age_Group'], color='green')
11 plt.xlabel("Age Group")
12 plt.ylabel("Count")
13 plt.title("Frequency of Covid-19 cases by Age");
```



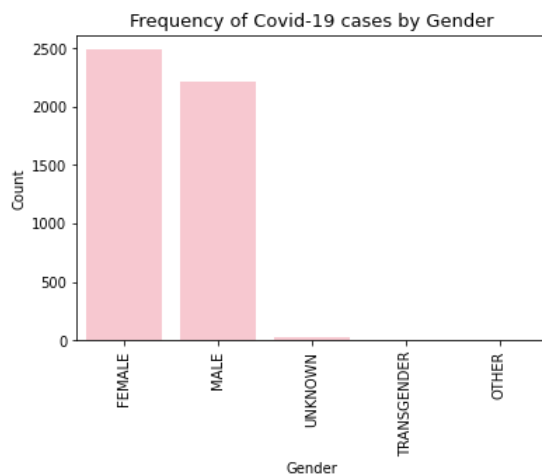
If we look at cases by age we can see that the distribution roughly follows the age distribution of Ontario residents. The median age in Ontario is about 40, and it looks like that's approximately the median age of people with covid-19. This might indicate it does discriminate about who gets infected.

Cases by Gender

```

1 fig = sns.countplot(df['CLIENT_GENDER'], color='pink', \
2                     order = df['CLIENT_GENDER'].value_counts().index)
3 plt.xlabel("Gender")
4 plt.ylabel("Count")
5 plt.title("Frequency of Covid-19 cases by Gender", fontsize=13)
6 plt.xticks(rotation=90);

```



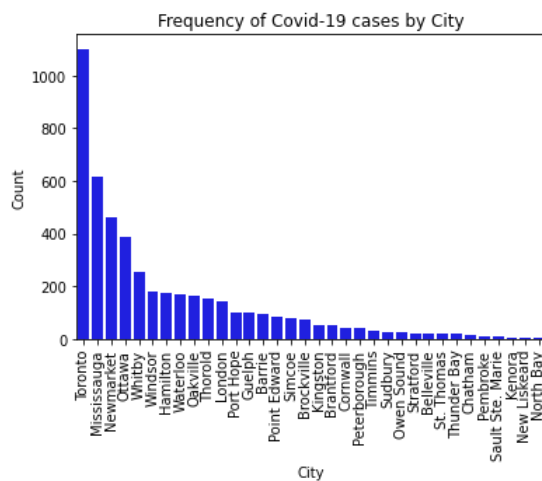
The number of cases are pretty evenly distributed between males and females.

Cases by City

```

1 fig = sns.countplot(df['Reporting_PHU_City'], color='blue', \
2                     order = df['Reporting_PHU_City'].value_counts().index)
3 plt.xlabel("City")
4 plt.ylabel("Count")
5 plt.title("Frequency of Covid-19 cases by City")
6 plt.xticks(rotation=90);

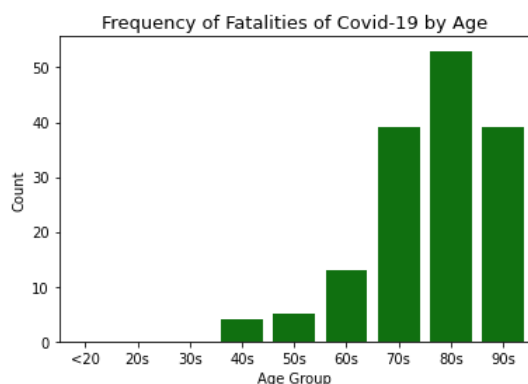
```



It's no surprise that Toronto has the most Covid-19 cases, but it does not nearly have the most cases per capita. Newmarket has the third most number of cases, more than Ottawa, a city that is ten times its size.

Fatalities by Age

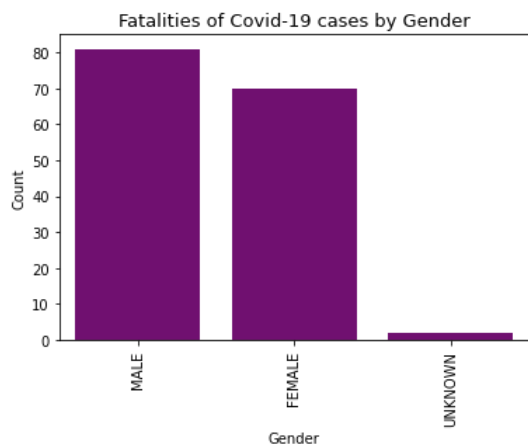
```
1 is_fatal = df['OUTCOME1'] == 'Fatal'
2 dfFatal = df[is_fatal]
3 fig = sns.countplot(dfFatal['Age_Group'], color='green')
4 plt.xlabel("Age Group")
5 plt.ylabel("Count")
6 plt.title("Frequency of Fatalities of Covid-19 by Age", fontsize=13);
```



The vast majority of deaths are of people over 70. There have been no fatalities in Ontario for those under 40.

Fatalities by Gender

```
1 fig = sns.countplot(dfFatal['CLIENT_GENDER'], color='purple', \
2                     order = dfFatal['CLIENT_GENDER'].value_counts().index)
3 plt.xlabel("Gender")
4 plt.ylabel("Count")
5 plt.title("Fatalities of Covid-19 cases by Gender", fontsize=13)
6 plt.xticks(rotation=90);
```



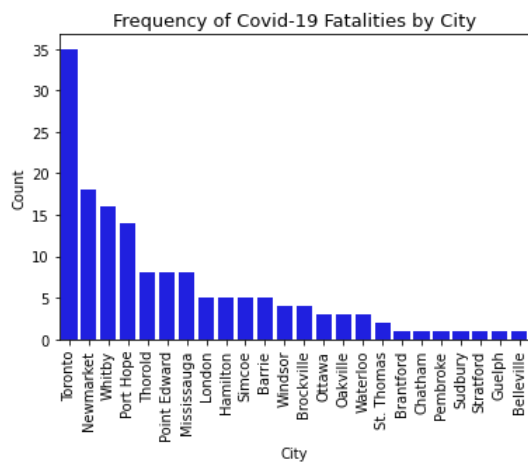
We can see that slightly more males died from Covid-19 than females despite there being more female cases.

Fatalities by City

```

1 fig = sns.countplot(dfFatal['Reporting_PHU_City'], color='blue', \
2                     order = dfFatal['Reporting_PHU_City'].value_counts().index)
3 plt.xlabel("City")
4 plt.ylabel("Count")
5 plt.title("Frequency of Covid-19 Fatalities by City", fontsize=13)
6 plt.xticks(rotation=90);

```



Toronto has the most fatalities, but certainly not the most per capita. Newmarket is a hotspot with the second highest number of fatalities in Ontario despite having less than 100,000 people.

Cases over Time

```

1 # Convert string to date
2 df['ACCURATE_EPISODE_DATE'] = pd.to_datetime(df['ACCURATE_EPISODE_DATE'])

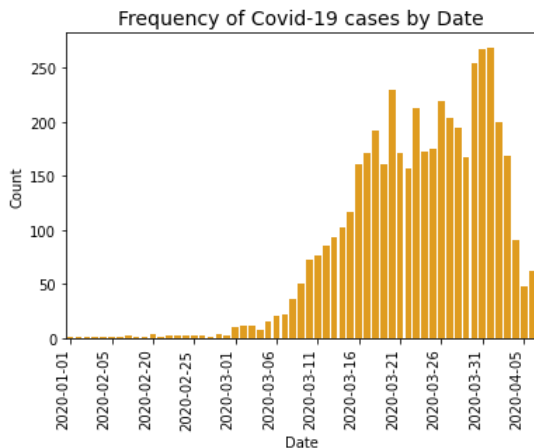
```

Cases by Date

```

1 fig = sns.countplot(df['ACCURATE_EPISODE_DATE'], color='orange')
2 plt.xlabel("Date")
3 plt.ylabel("Count")
4 plt.title("Frequency of Covid-19 cases by Date", fontsize=14)
5 plt.xticks(rotation=90)
6 x_dates = df['ACCURATE_EPISODE_DATE'].dt.strftime(
7     '%Y-%m-%d').sort_values().unique()
8 fig.set_xticklabels(labels=x_dates, rotation=90, ha='right')
9 new_ticks = [i.get_text() for i in fig.get_xticklabels()]
10 plt.xticks(range(0, len(new_ticks), 5), new_ticks[::5]);

```



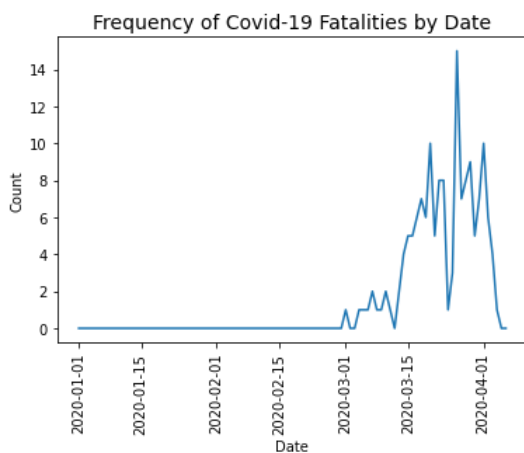
It's hard to say if the number of cases are leveling off. The steep decline in numbers is probably due to a backlog in updated the government data website.

Fatalities by Date

```

1 # Add column for fatal: 1,0
2 df['Fatal'] = df['OUTCOME1'].apply(lambda x: 1 if x == 'Fatal' else 0)
3
4 df['ACCURATE_EPISODE_DATE'] = pd.to_datetime(df['ACCURATE_EPISODE_DATE']).dt.date
5
6 g = df.groupby(['ACCURATE_EPISODE_DATE'])['Fatal'].sum().plot()
7 plt.xlabel("Date")
8 plt.ylabel("Count")
9 plt.title("Frequency of Covid-19 Fatalities by Date", fontsize=14)
10 plt.xticks(rotation=90);

```



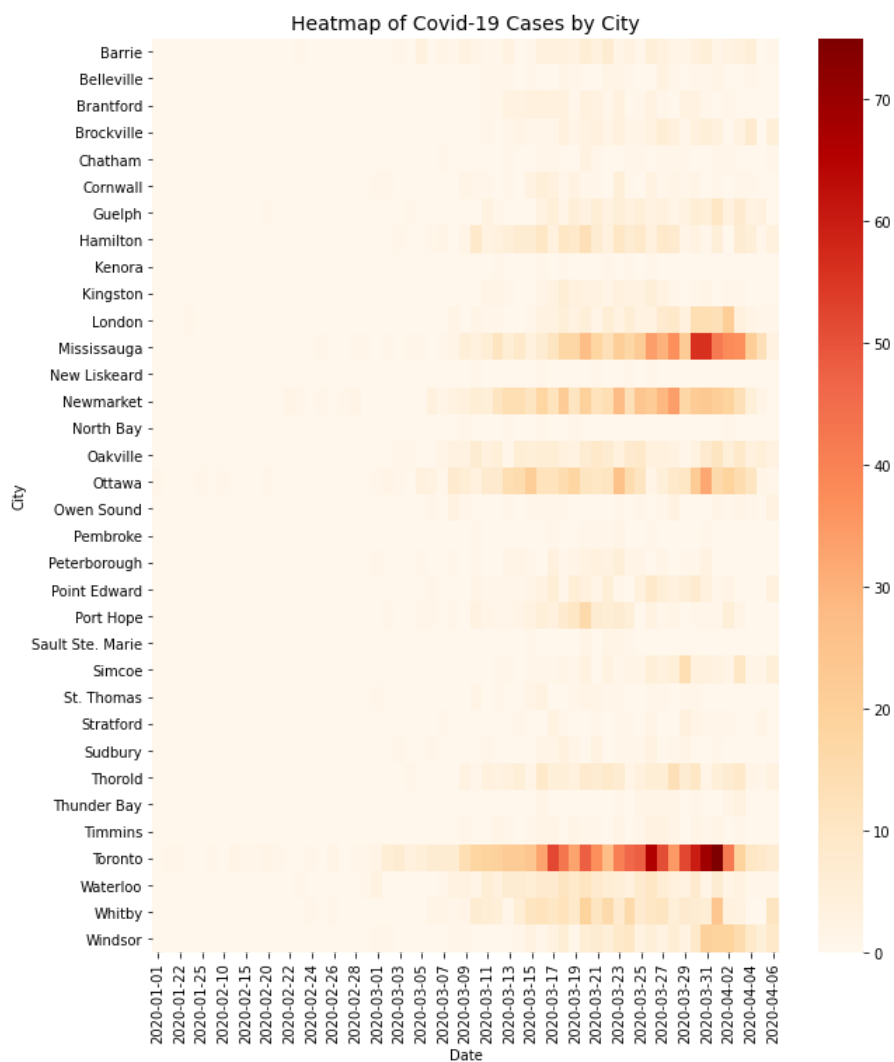
The spike in fatalities levels off, which is probably due to latency in the government data. There has also been news about a backlog of tests, and this could be reason for this apparent decline.

Heatmap of Cases by City over Time

```

1 grouped_cases = df.groupby(['Reporting_PHU_City', 'ACCURATE_EPISODE_DATE'
2                             ]).agg({'ROW_ID': ['count']})
3
4 grouped_cases.columns = ['TotalCases']
5 grouped_cases = grouped_cases.reset_index()
6
7 # Build Heatmap
8 grouped_cases.columns = ['City', 'Date', 'TotalCases']
9
10 #Pivot table for heatmap
11 heatmap1_data = pd.pivot_table(grouped_cases,
12                                 values='TotalCases',
13                                 index=['City'],
14                                 columns='Date',
15                                 fill_value=0)
16
17 #heatmap1_data['Date'] = pd.heatmap1_data(df['Date']).dt.date
18
19 #heatmap1_data['Date'] = heatmap1_data.Date.astype('string')
20 plt.subplots(figsize=(10, 12))
21 ax = sns.heatmap(heatmap1_data,
22                 annot=False,
23                 fmt="d",
24                 linewidths=0,
25                 cmap="OrRd")
26
27 ax.set_title('Heatmap of Covid-19 Cases by City', fontsize=14);

```



There's a clear increase in cases over time, with some cities such as Toronto and Mississauga showing large increases more recently.

It's likely that the decrease in cases in the most recent days are due to a delay by the government in posting the data.

Heatmap of Fatalities by City over Time

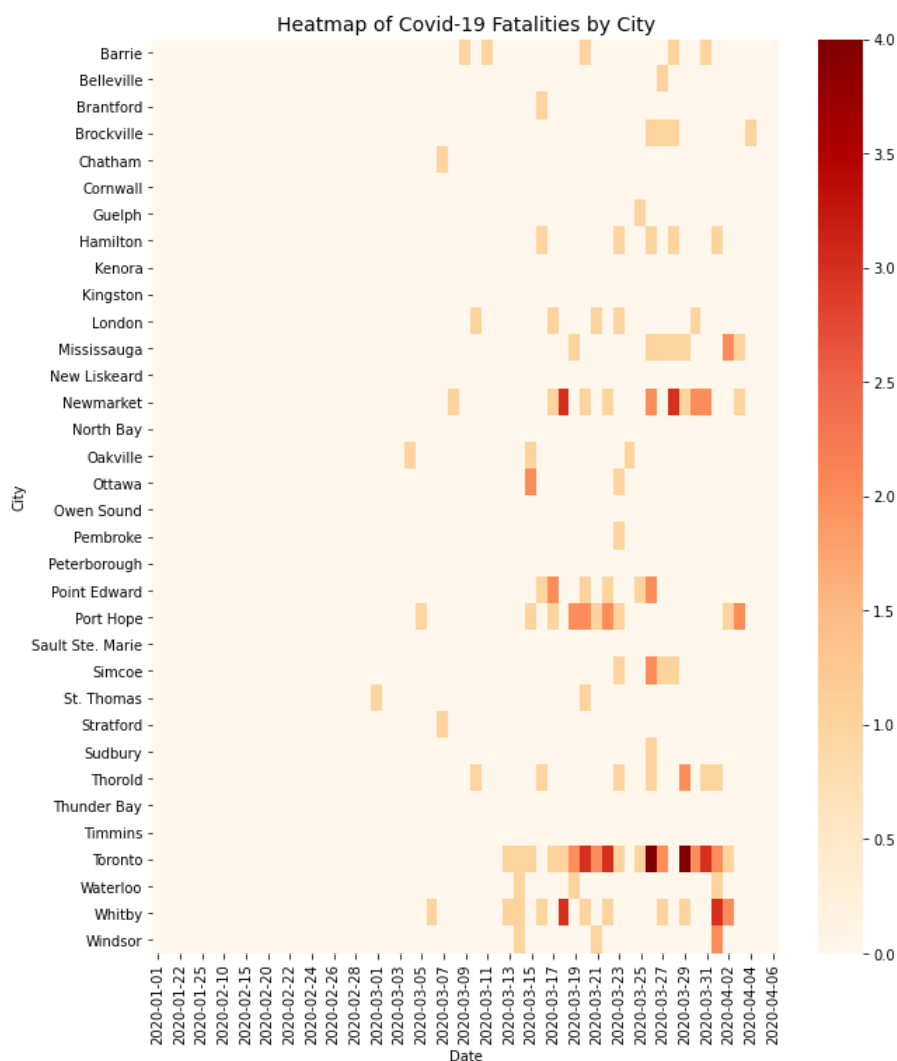
```

1 #g = df.groupby(['Reporting_PHU_City', 'ACCURATE_EPISODE_DATE'])['Fatal'].sum()

1 # Aggregate columns
2 grouped_multiple = df.groupby(['Reporting_PHU_City', 'ACCURATE_EPISODE_DATE'
3                               ]).agg({'Fatal': ['sum']})
4 grouped_multiple.columns = ['totalFatal']
5 grouped_multiple = grouped_multiple.reset_index()

1 # Rename columns
2 grouped_multiple.columns = ['City', 'Date', 'Fatalities']
3 #Pivot table for heatmap
4 heatmap1_data = pd.pivot_table(grouped_multiple,
5                                values='Fatalities',
6                                index=['City'],
7                                columns='Date',
8                                fill_value=0)
9 plt.subplots(figsize=(10, 12))
10 #ax.set_xticklabels(heatmap1_data['Date'].dt.strftime('%Y-%m-%d'))
11
12 ax = sns.heatmap(heatmap1_data,
13                 annot=False,
14                 fmt="d",
15                 linewidths=0,
16                 cmap="OrRd")
17 ax.set_title('Heatmap of Covid-19 Fatalities by City', fontsize=14);

```



The heatmap above shows the progression of fatalities over time.

We can see that some cities had a fatality, and then no more followed, such as Stratford, and other cities such as Newmarket show an increase in fatalities.