

Bioinformatics Project Presentation

COVID19-N20

Sicheng Zhou

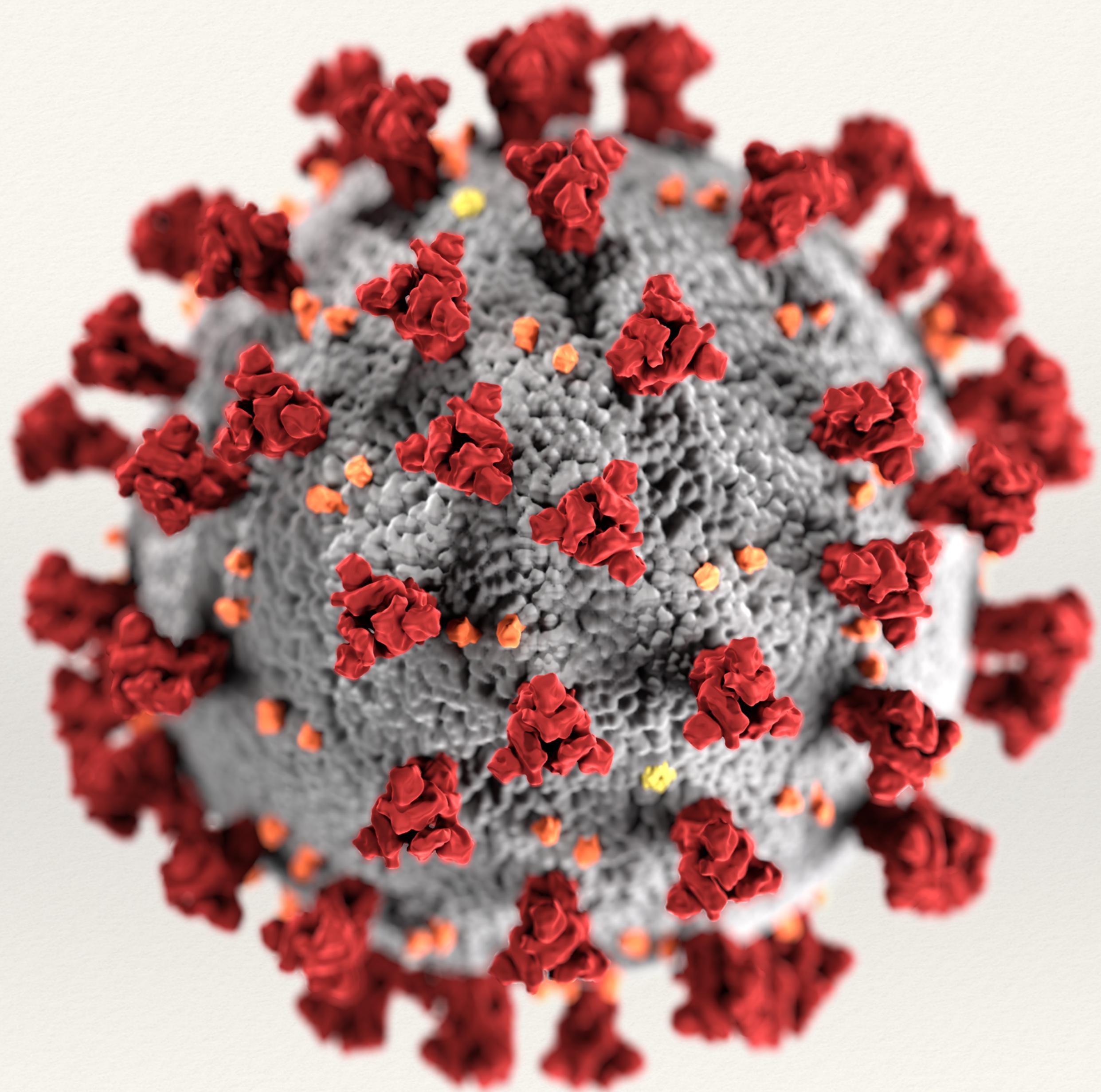
Instructor: Pf. Patricia Francis-Lyon

- ❖ Background
- ❖ John Hopkins University Data Set
- ❖ COVID-19N20 Data Set
- ❖ Exploratory Data Analysis
- ❖ Logistic Model
- ❖ RNN Model
- ❖ Future Plan
- ❖ References

Background

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, the capital of China's Hubei province, and has since spread globally, resulting in an ongoing pandemic. As of 5 May 2020, more than 3.65 million cases have been reported across 187 countries and territories, resulting in more than 256,000 deaths. More than 1.19 million people have recovered.

From: Wikipedia



John Hopkins University Data Set

<https://github.com/CSSEGISandData/COVID-19>

- [!\[\]\(13b6bdd0ca077c333d50231f1443cb1d_img.jpg\) time_series_covid19_confirmed_US.csv](#)
- [!\[\]\(5dbedd4e1e8871e3a0e67053ad2f9701_img.jpg\) time_series_covid19_confirmed_global.csv](#)
- [!\[\]\(d4749465acb9b53e115af1f9ce82539c_img.jpg\) time_series_covid19_deaths_US.csv](#)
- [!\[\]\(3e3001313d495ec87b5a6a5de6205728_img.jpg\) time_series_covid19_deaths_global.csv](#)
- [!\[\]\(e26df985e6d3e053d2593dc7b93b41cf_img.jpg\) time_series_covid19_recovered_global.csv](#)

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20
	Afghanistan	33.0	65.0	0	0	0
	Albania	41.1533	20.1683	0	0	0
	Algeria	28.0339	1.6596	0	0	0
	Andorra	42.5063	1.5218	0	0	0
	Angola	-11.2027	17.8739	0	0	0
	Antigua and Barbuda	17.0608	-61.7964	0	0	0
	Argentina	-38.4161	-63.6167	0	0	0
	Armenia	40.0691	45.0382	0	0	0
Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0
New South Wales	Australia	-33.8688	151.2093	0	0	0
Northern Territory	Australia	-12.4634	130.8456	0	0	0
Queensland	Australia	-28.0167	153.4	0	0	0
South Australia	Australia	-34.9285	138.6007	0	0	0
Tasmania	Australia	-41.4545	145.9707	0	0	0

COVID-19N20 Data Set

- ❖ Based on JHK Data Set
- ❖ Calculate new confirmed cases per day
- ❖ Calculate new death cases per day
- ❖ Tag the day as 1 when there are at least 20 cases for “country” (10 for “state”)
- ❖ For all the countries and all US states listed in JHK Data Set
- ❖ <https://github.com/secregister01/COVID19-N20>

country	date	confirmed	deaths	new_confirmed	new_deaths	tagged_day
US	2/24/20	51	0	36	0	1
US	2/25/20	51	0	0	0	2
US	2/26/20	57	0	6	0	3
US	2/27/20	58	0	1	0	4
US	2/28/20	60	0	2	0	5
US	2/29/20	68	1	8	1	6
US	3/1/20	74	1	6	0	7
US	3/2/20	98	6	24	5	8
US	3/3/20	118	7	20	1	9
US	3/4/20	149	11	31	4	10
US	3/5/20	217	12	68	1	11
US	3/6/20	262	14	45	2	12
US	3/7/20	402	17	140	3	13
US	3/8/20	518	21	116	4	14
US	3/9/20	583	22	65	1	15
US	3/10/20	959	28	376	6	16
US	3/11/20	1281	36	322	8	17
US	3/12/20	1663	41	382	5	18
US	3/13/20	2179	49	516	8	19
US	3/14/20	2727	58	548	9	20

Country Area

1	Afghanistan,251,830 mi ²
2	Albania,11,100 mi ²
3	Algeria,919,600 mi ²
4	Andorra,180.7 mi ²
5	Angola,481,400 mi ²
6	Antigua and Barbuda,169.9 mi ²
7	Argentina,1.074 million mi ²
8	Armenia,11,484 mi ²
9	Australia,2.97 million mi ²
10	Austria,32,386 mi ²
11	Azerbaijan,33,436 mi ²
12	Bahamas,5,358 mi ²
13	Bahrain,295.5 mi ²
14	Bangladesh,56,977 mi ²
15	Barbados,166.4 mi ²
16	Belarus,80,153 mi ²
17	Belgium,11,849 mi ²
18	Belize,-1
19	Benin,44,310 mi ²
20	Bhutan,14,824 mi ²
21	Bolivia,424,200 mi ²
22	Bosnia and Herzegovina,19,767 mi ²
23	Botswana,231,804 mi ²
24	Brazil,3.286 million mi ²
25	Brunei,-1
26	Bulgaria,42,855 mi ²
27	Burkina Faso,105,869 mi ²
28	Burma,261,227 mi ²
29	Burundi,10,747 mi ²
30	Cabo Verde,1,557 mi ²
31	Cambodia,69,898 mi ²
32	Cameroon,183,569 mi ²
33	Canada,3.855 million mi ²
34	Central African Republic,240,534 mi ²
35	Chad,495,800 mi ²
36	Chile,292,260 mi ²
37	China,3.705 million mi ²
38	Colombia,440,800 mi ²
39	Comoros,641 mi ²
40	Congo (Brazzaville),132,047 mi ²
41	Congo (Kinshasa),905,400 mi ²
42	Costa Rica,19,730 mi ²
43	Cote d'Ivoire,124,504 mi ²
44	Croatia,21,851 mi ²
45	Cuba,42,426 mi ²
46	Cyprus,3,572 mi ²
47	Czechia,30,450 mi ²
48	Denmark,16,577 mi ²
49	Diamond Princess,-1
50	Djibouti,8,958 mi ²
51	Dominica,-1
52	Dominican Republic,18,704 mi ²

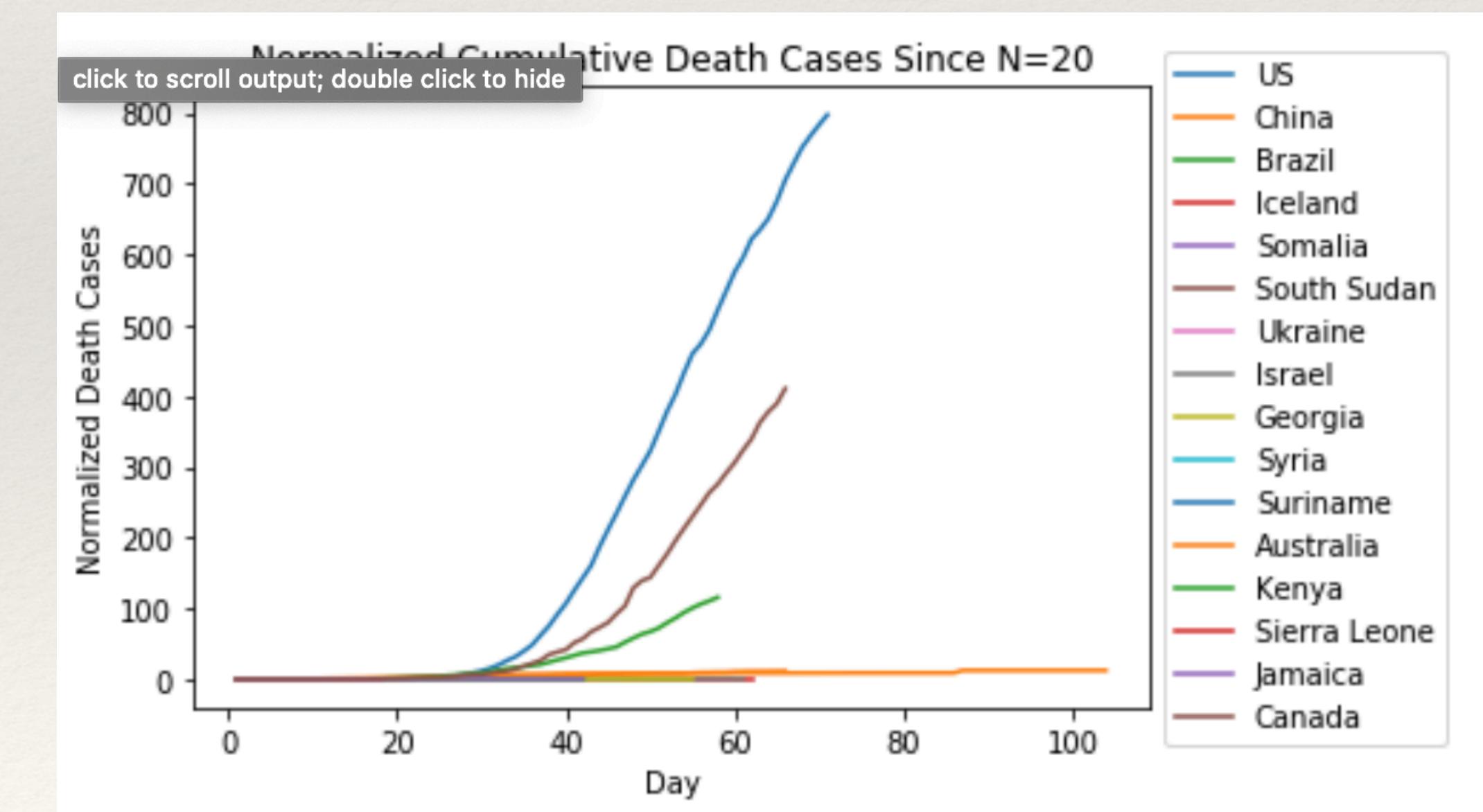
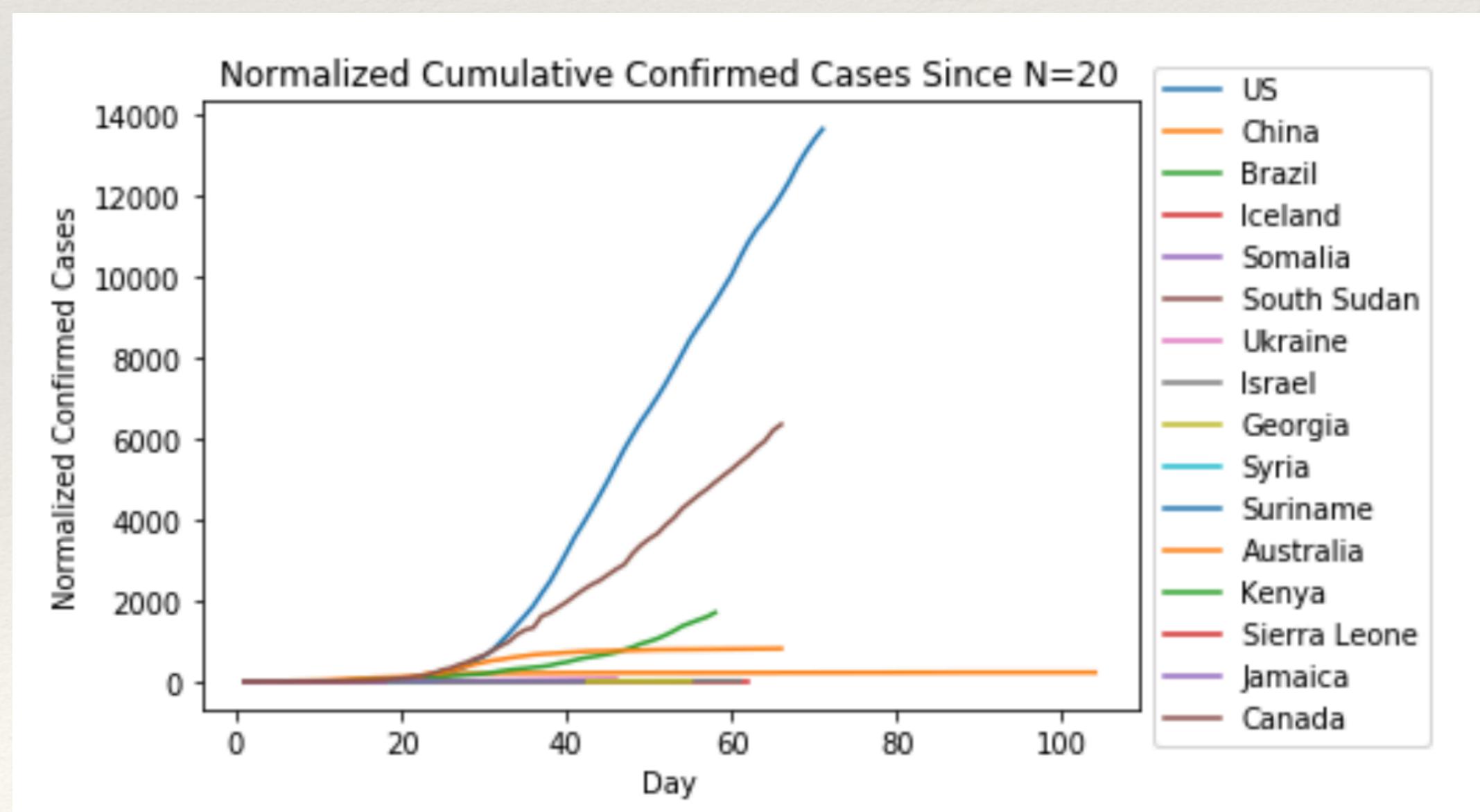
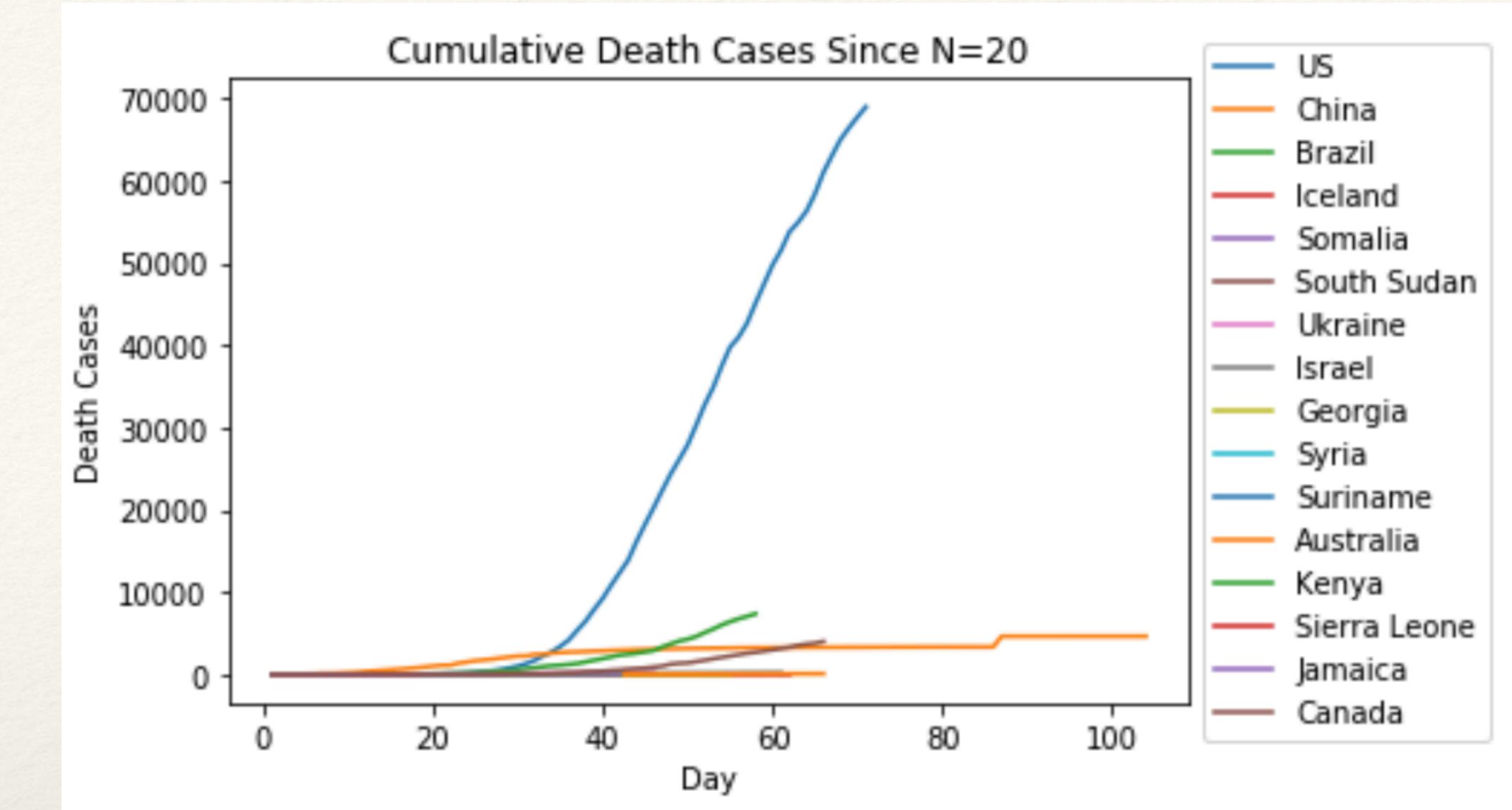
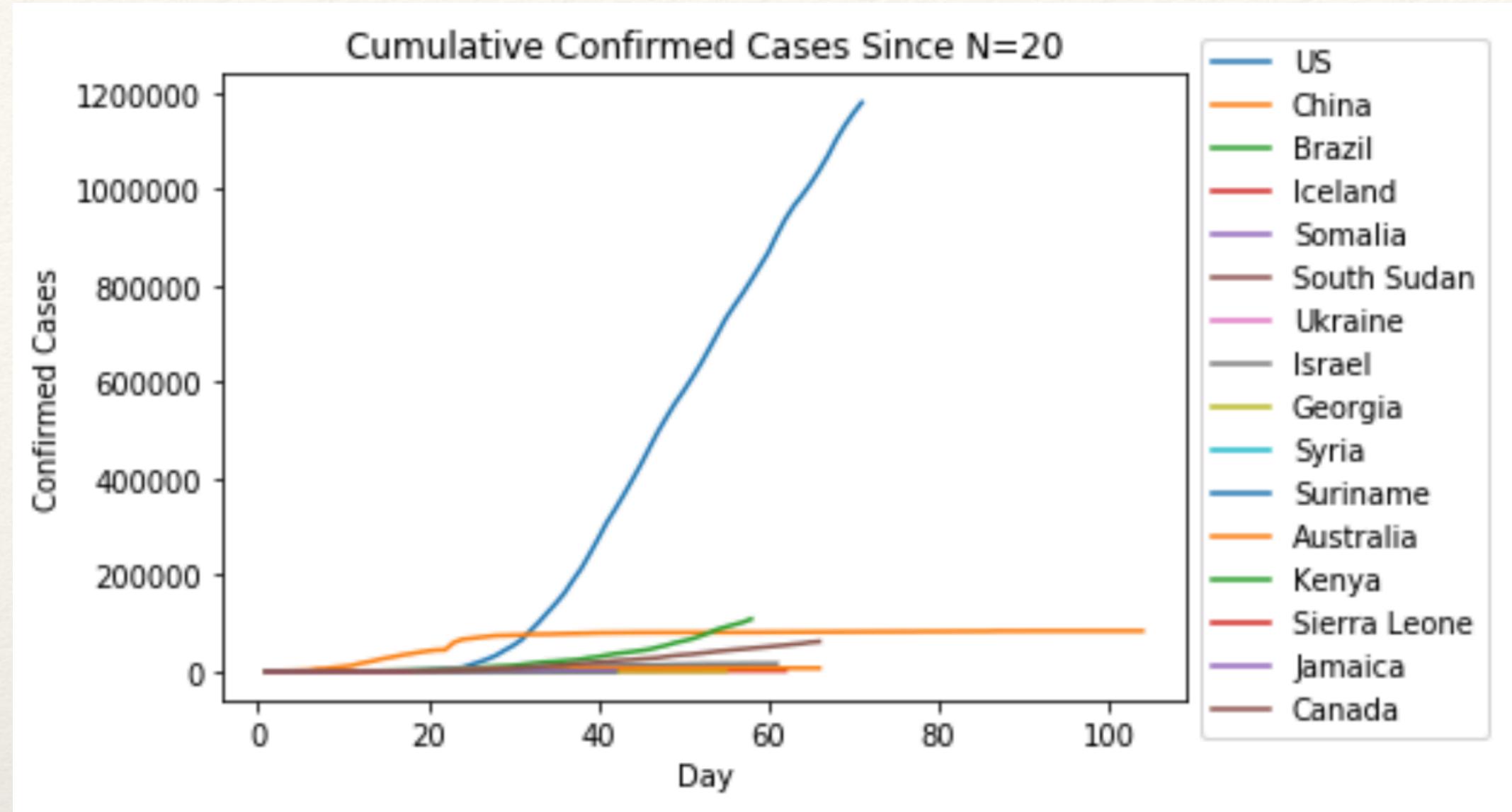
Country GDP Per Capita

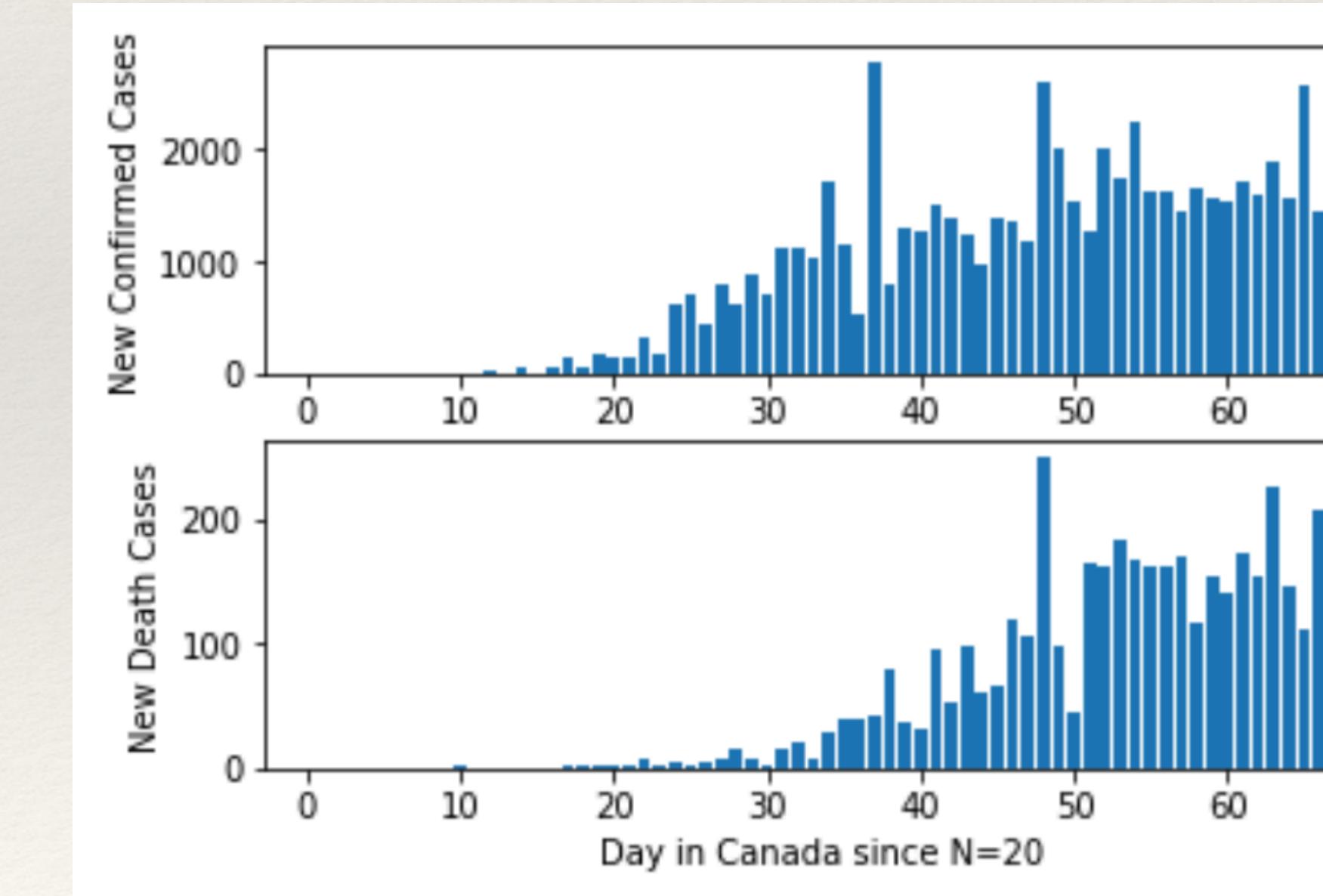
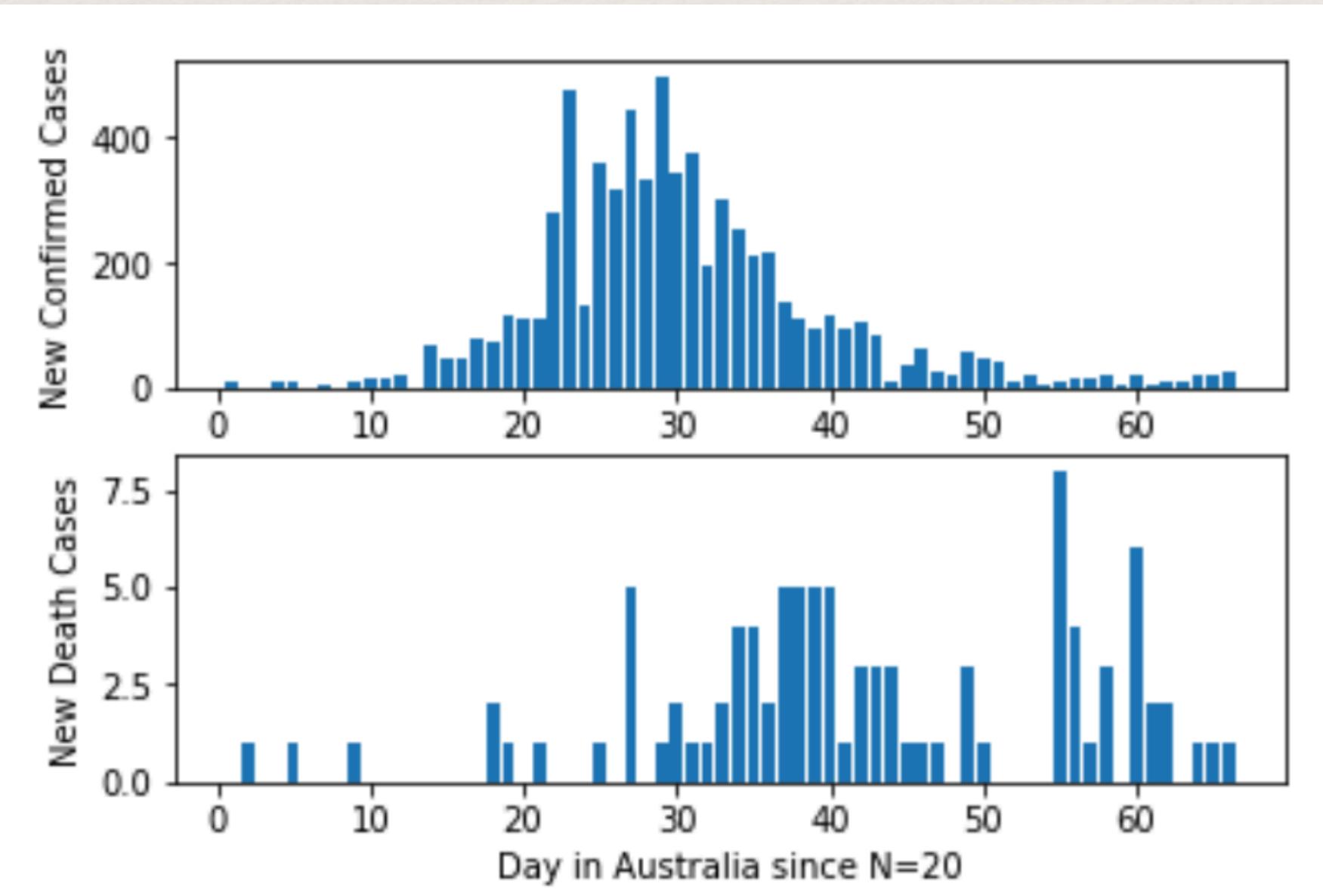
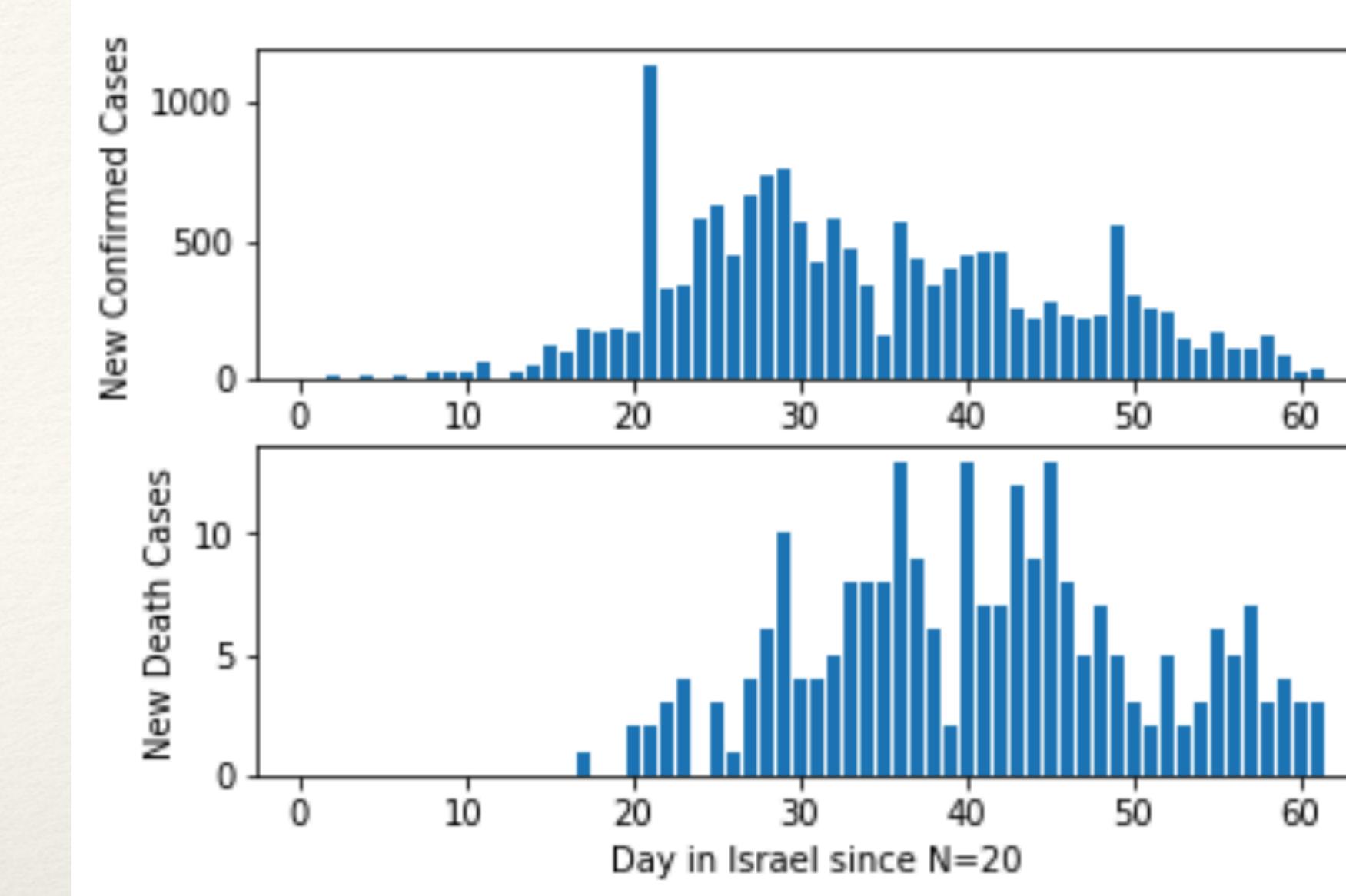
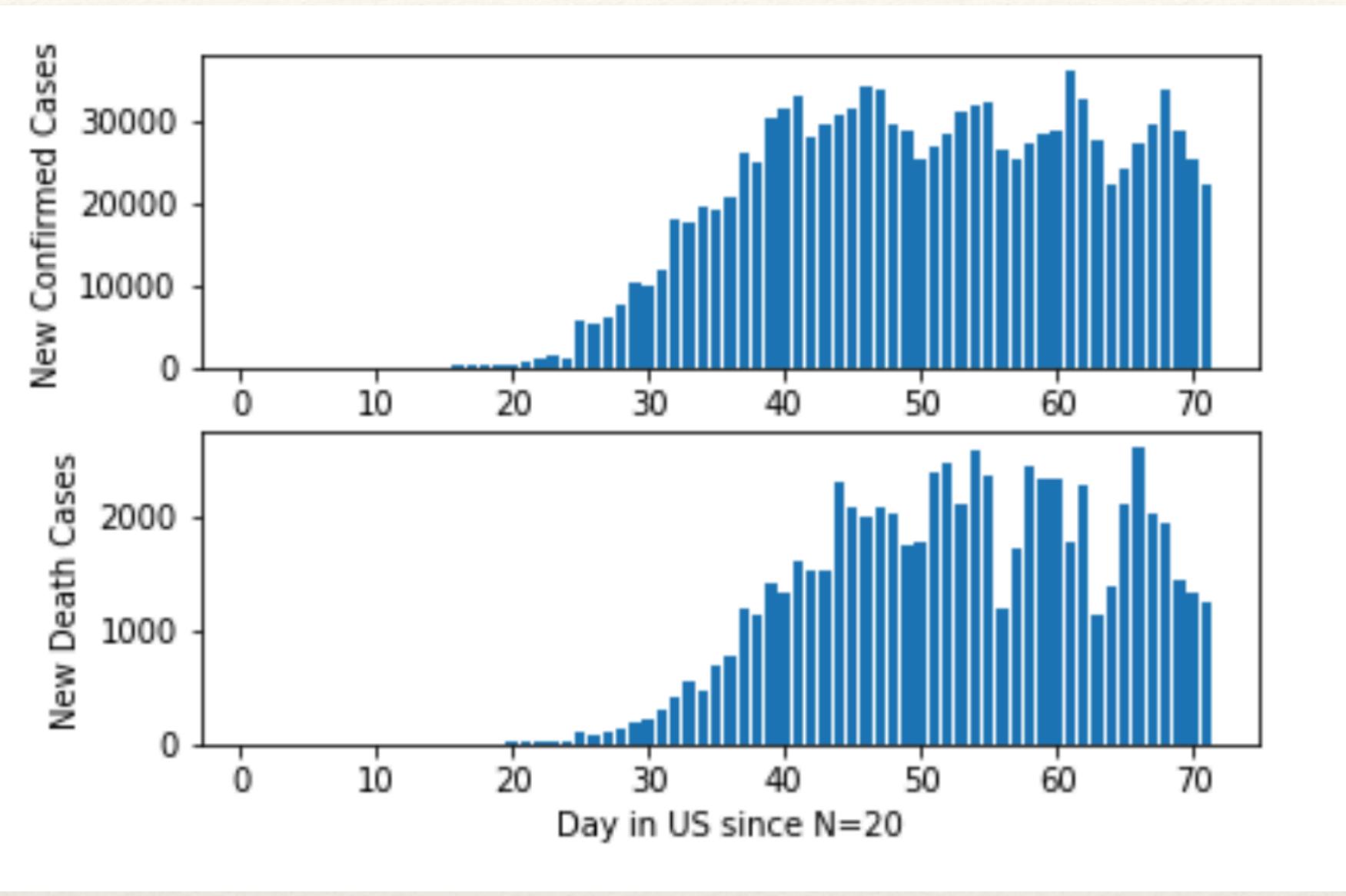
1	Afghanistan,520.90 USD (2018)
2	Albania,5,268.85 USD (2018)
3	Algeria,4,114.72 USD (2018)
4	Andorra,42,029.76 USD (2018)
5	Angola,3,432.39 USD (2018)
6	Antigua and Barbuda,16,726.98 USD (2018)
7	Argentina,11,683.95 USD (2018)
8	Armenia,4,212.07 USD (2018)
9	Australia,57,373.69 USD (2018)
10	Austria,51,461.95 USD (2018)
11	Azerbaijan,4,721.18 USD (2018)
12	Bahamas,32,217.87 USD (2018)
13	Bahrain,24,050.76 USD (2018)
14	Bangladesh,1,698.26 USD (2018)
15	Barbados,17,949.28 USD (2018)
16	Belarus,6,289.94 USD (2018)
17	Belgium,47,518.64 USD (2018)
18	Belize,4,884.74 USD (2018)
19	Benin,901.54 USD (2018)
20	Bhutan,3,243.23 USD (2018)
21	Bolivia,3,548.59 USD (2018)
22	Bosnia and Herzegovina,6,065.67 USD (2018)
23	Botswana,8,258.64 USD (2018)
24	Brazil,8,920.76 USD (2018)
25	Brunei,31,628.33 USD (2018)
26	Bulgaria,9,272.63 USD (2018)
27	Burkina Faso,715.12 USD (2018)
28	Burma,1,325.95 USD (2018)
29	Burundi,271.75 USD (2018)
30	Cabo Verde,3,635.41 USD (2018)
31	Cambodia,1,510.32 USD (2018)
32	Cameroon,1,533.74 USD (2018)
33	Canada,46,232.99 USD (2018)
34	Central African Republic,475.72 USD (2018)
35	Chad,728.34 USD (2018)
36	Chile,15,923.36 USD (2018)
37	China,9,770.85 USD (2018)
38	Colombia,6,667.79 USD (2018)
39	Comoros,1,415.26 USD (2018)
40	Congo (Brazzaville),2,147.77 USD (2018)
41	Congo (Kinshasa),561.78 USD (2018)
42	Costa Rica,12,027.37 USD (2018)
43	Cote d'Ivoire,1,715.53 USD (2018)
44	Croatia,14,909.69 USD (2018)
45	Cuba,8,821.82 USD (2018)
46	Cyprus,28,159.30 USD (2018)
47	Czechia,23,078.57 USD (2018)
48	Denmark,61,350.35 USD (2018)
49	Diamond Princess,-1
50	Djibouti,3,082.54 USD (2018)
51	Dominica,7,691.35 USD (2018)
52	Dominican Republic,8,050.63 USD (2018)

Country Population

1	Afghanistan,37.17 million (2018)
2	Albania,2.846 million
3	Algeria,42.23 million (2018)
4	Andorra,77,006 (2018)
5	Angola,30.81 million (2018)
6	Antigua and Barbuda,96,286 (2018)
7	Argentina,44.49 million (2018)
8	Armenia,2.965 million (2019)
9	Australia,24.99 million (2018)
10	Austria,8.859 million (2019)
11	Azerbaijan,9.981 million (2019)
12	Bahamas,385,640 (2018)
13	Bahrain,1.569 million (2018)
14	Bangladesh,161.4 million (2018)
15	Barbados,286,641 (2018)
16	Belarus,9.485 million (2018)
17	Belgium,11.46 million (2019)
18	Belize,383,071 (2018)
19	Benin,11.49 million (2018)
20	Bhutan,754,394 (2018)
21	Bolivia,11.35 million (2018)
22	Bosnia and Herzegovina,3.324 million (2018)
23	Botswana,2.254 million (2018)
24	Brazil,209.5 million (2018)
25	Brunei,428,962 (2018)
26	Bulgaria,7 million (2019)
27	Burkina Faso,19.75 million (2018)
28	Burma,53.71 million (2018)
29	Burundi,11.18 million (2018)
30	Cabo Verde,543,767 (2018)
31	Cambodia,16.25 million (2018)
32	Cameroon,25.22 million (2018)
33	Canada,37.59 million (2019)
34	Central African Republic,4.666 million (2018)
35	Chad,15.48 million (2018)
36	Chile,18.73 million (2018)
37	China,1.393 billion (2018)
38	Colombia,49.65 million (2018)
39	Comoros,832,322 (2018)
40	Congo (Brazzaville),5.244 million (2018)
41	Congo (Kinshasa),84.07 million (2018)
42	Costa Rica,4.999 million (2018)
43	Cote d'Ivoire,25.07 million (2018)
44	Croatia,4.076 million
45	Cuba,11.34 million (2018)
46	Cyprus,1.17 million
47	Czechia,10.65 million
48	Denmark,5.806 million (2019)
49	Diamond Princess,-1
50	Djibouti,958,920 (2018)
51	Dominica,71,625 (2018)
52	Dominican Republic,10.63 million (2018)

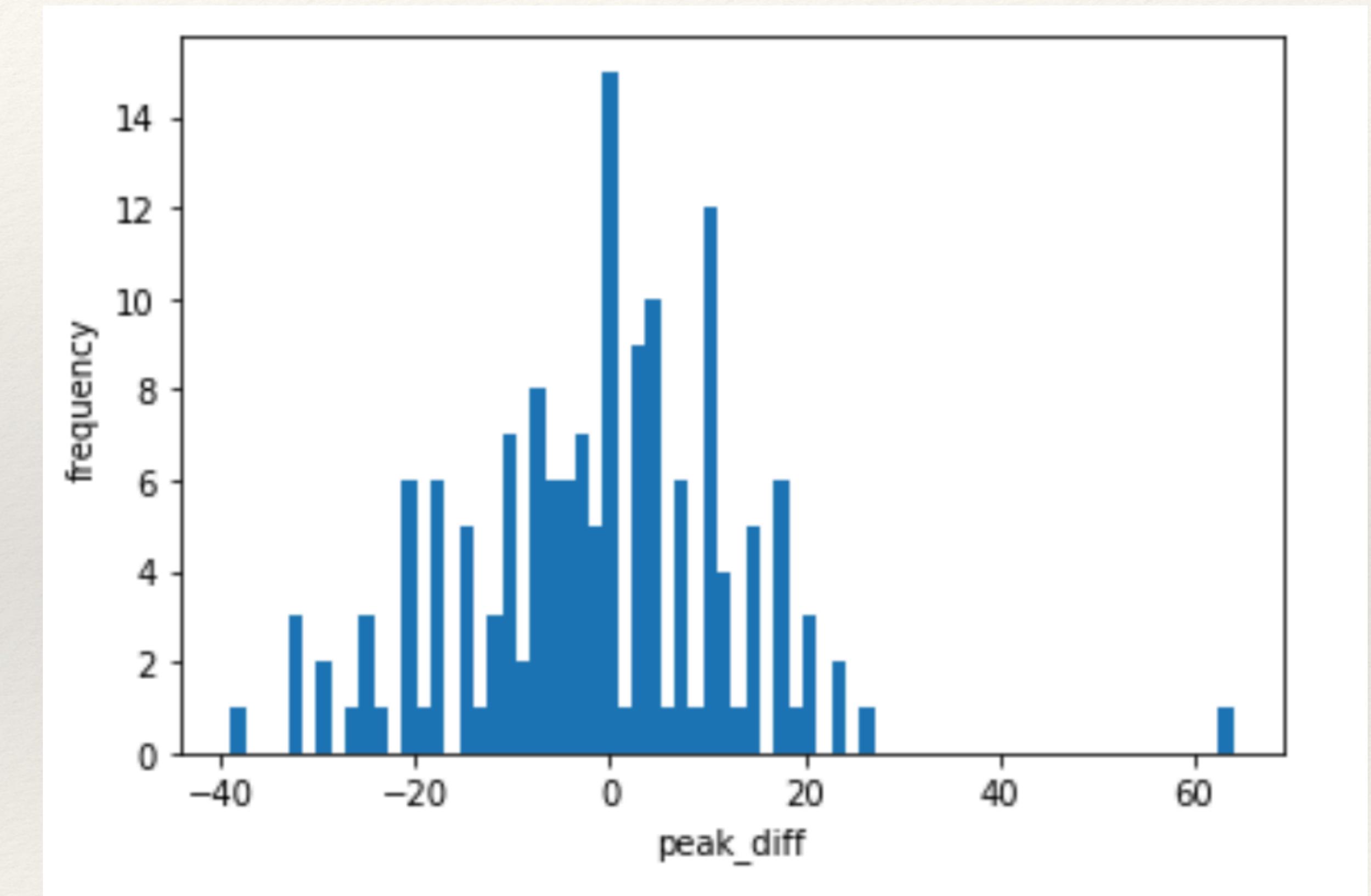
Exploratory Data Analysis



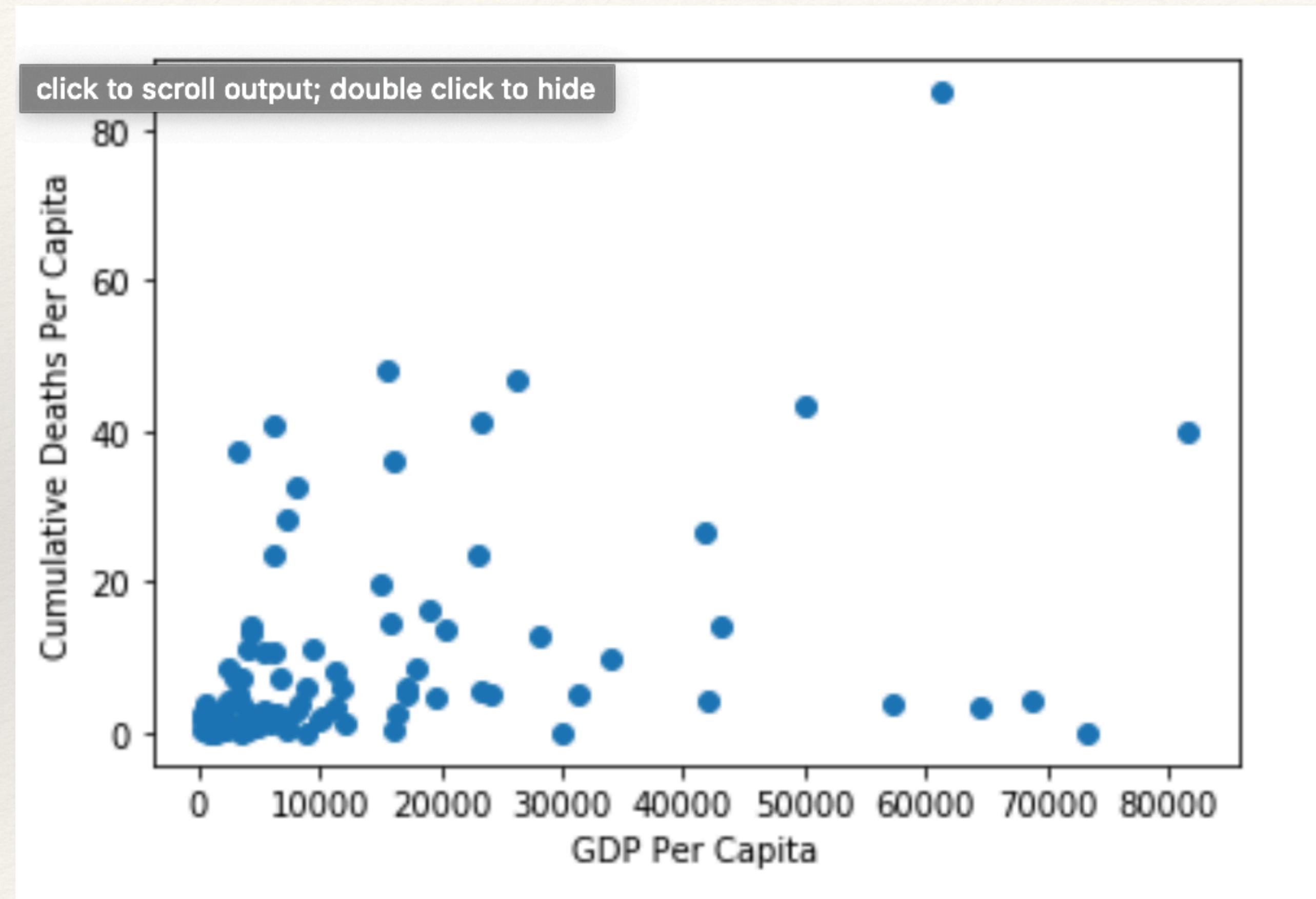


Difference of “Peaks”

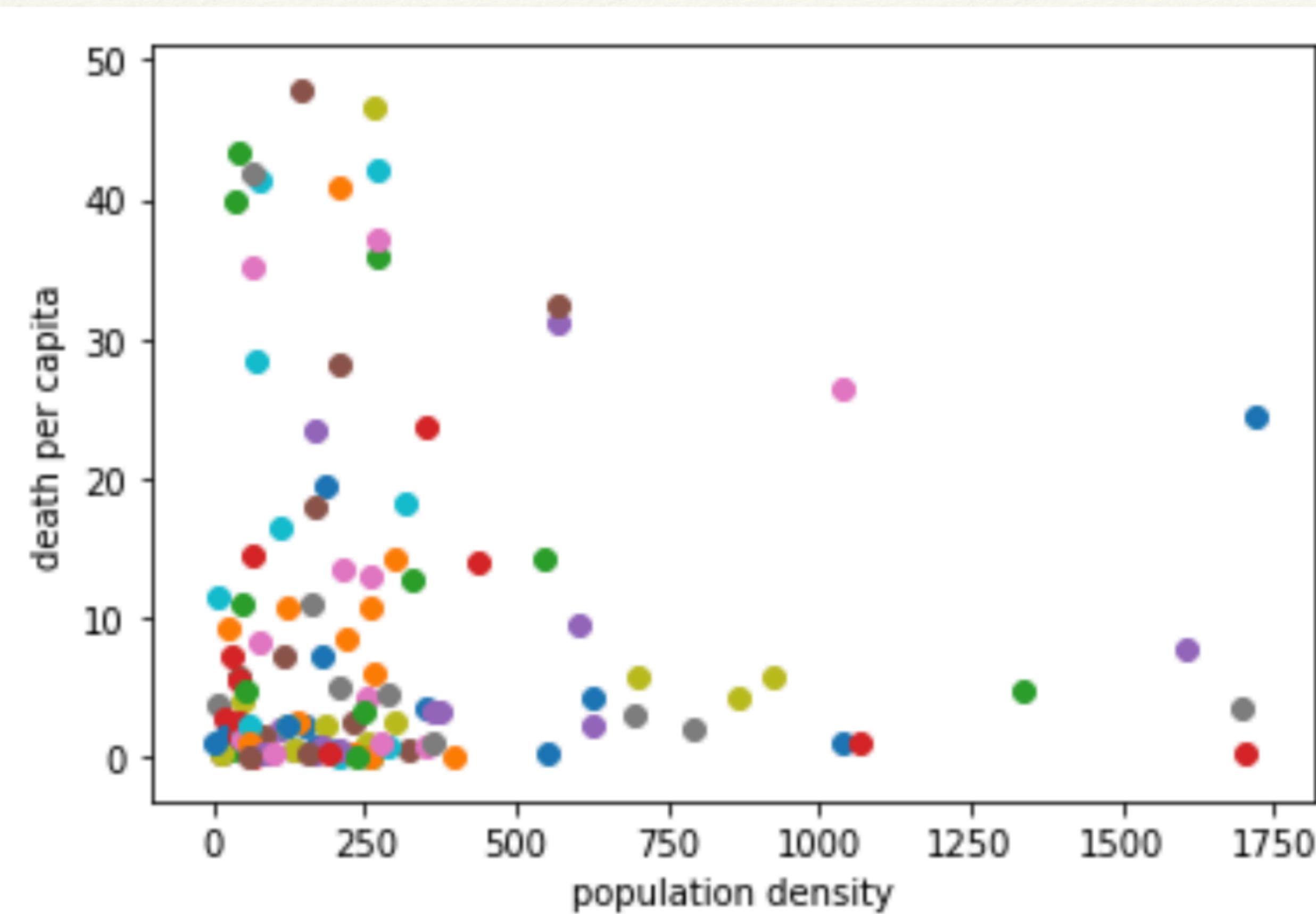
- ❖ Between new daily confirmed cases and new death cases
- ❖ For all the countries listed
- ❖ Normalized distribution



Cumulative Deaths vs GDP



Cumulative Deaths vs Population Density

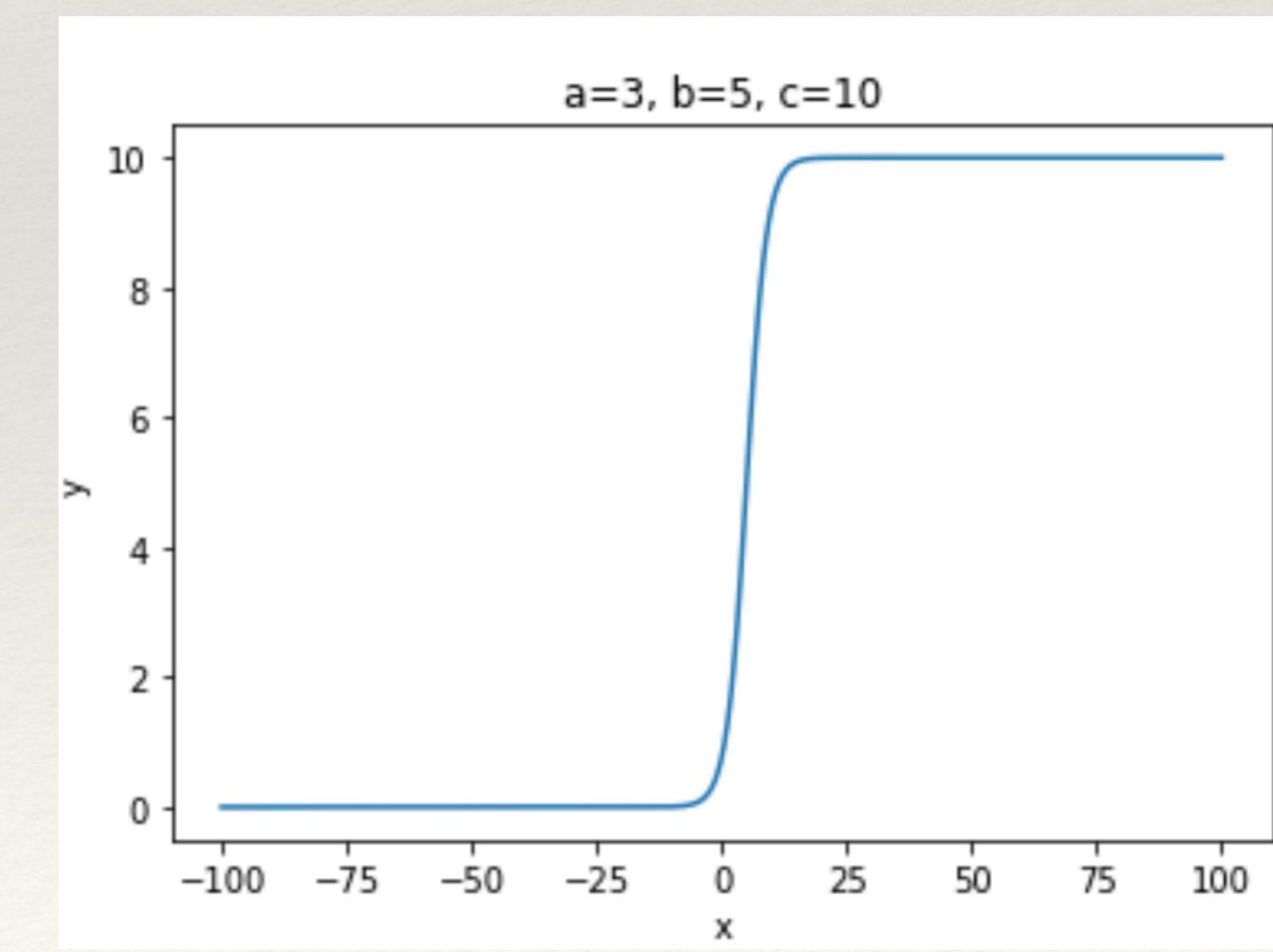
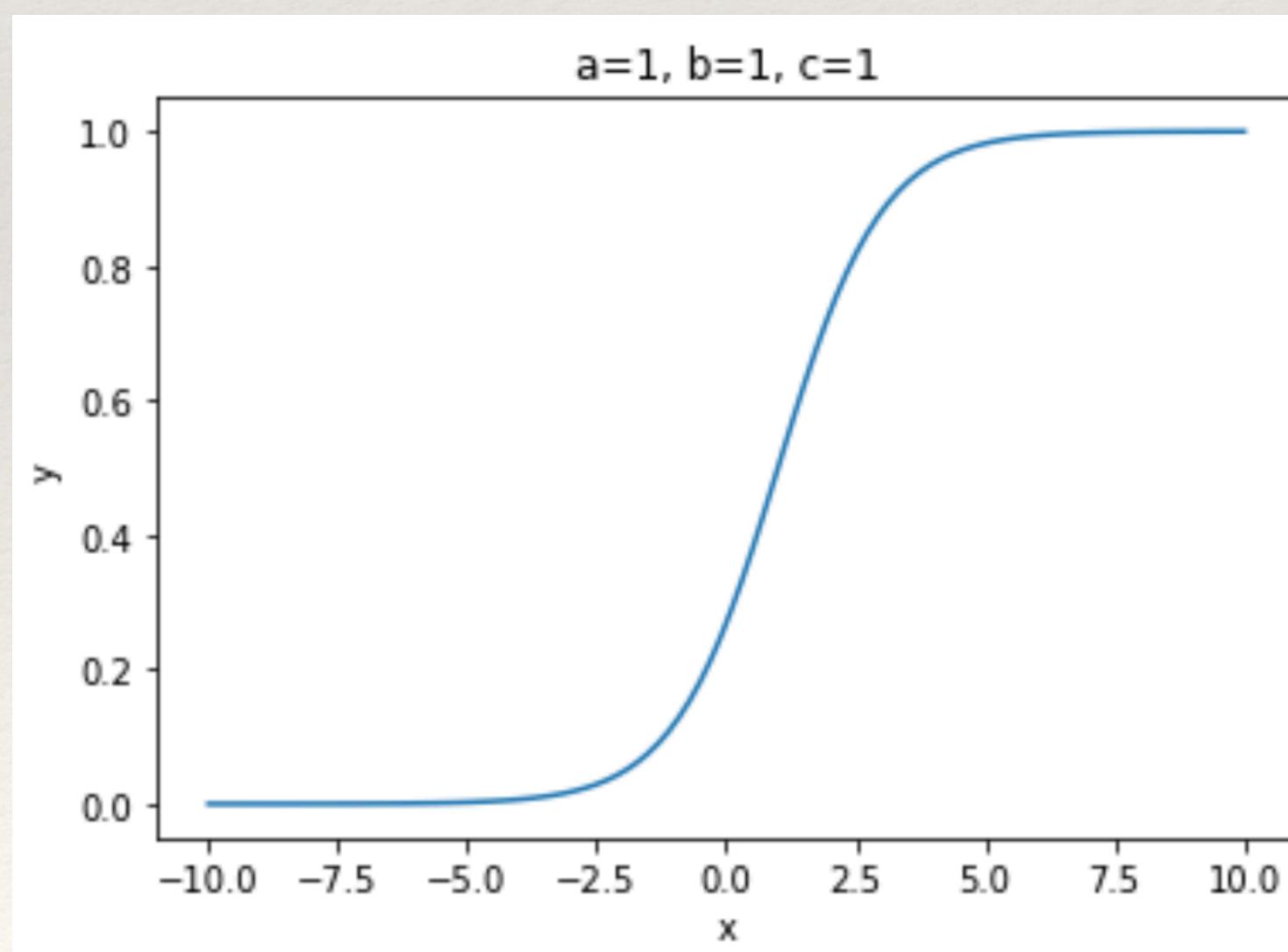
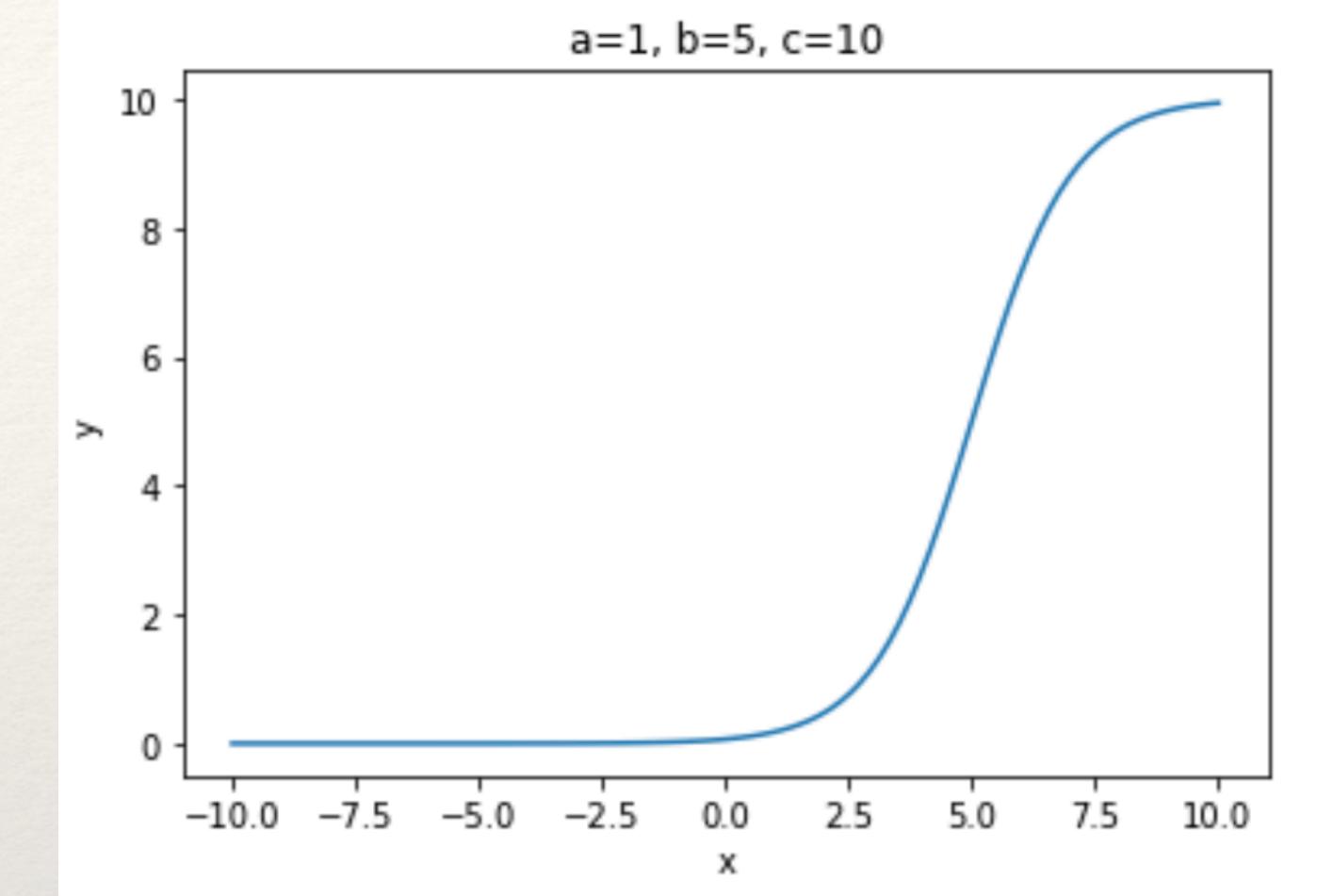


Logistic Model

Fitting Logistic Curve

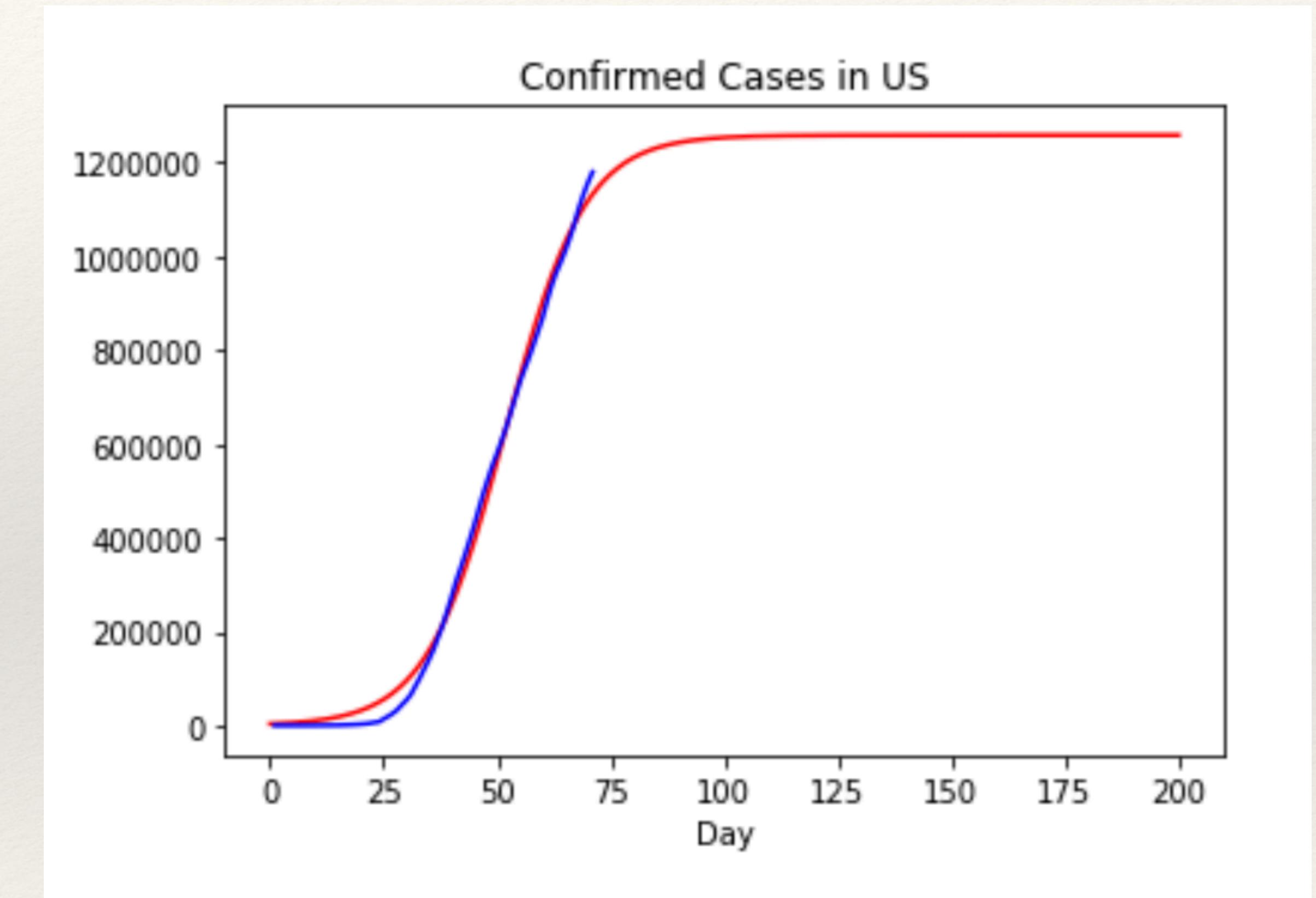
- ❖ Equation
- ❖ Parameter
- ❖ Plot

```
def logistic_model(x, a, b, c):  
    return c/(1 + np.exp(-(x-b)/a))
```



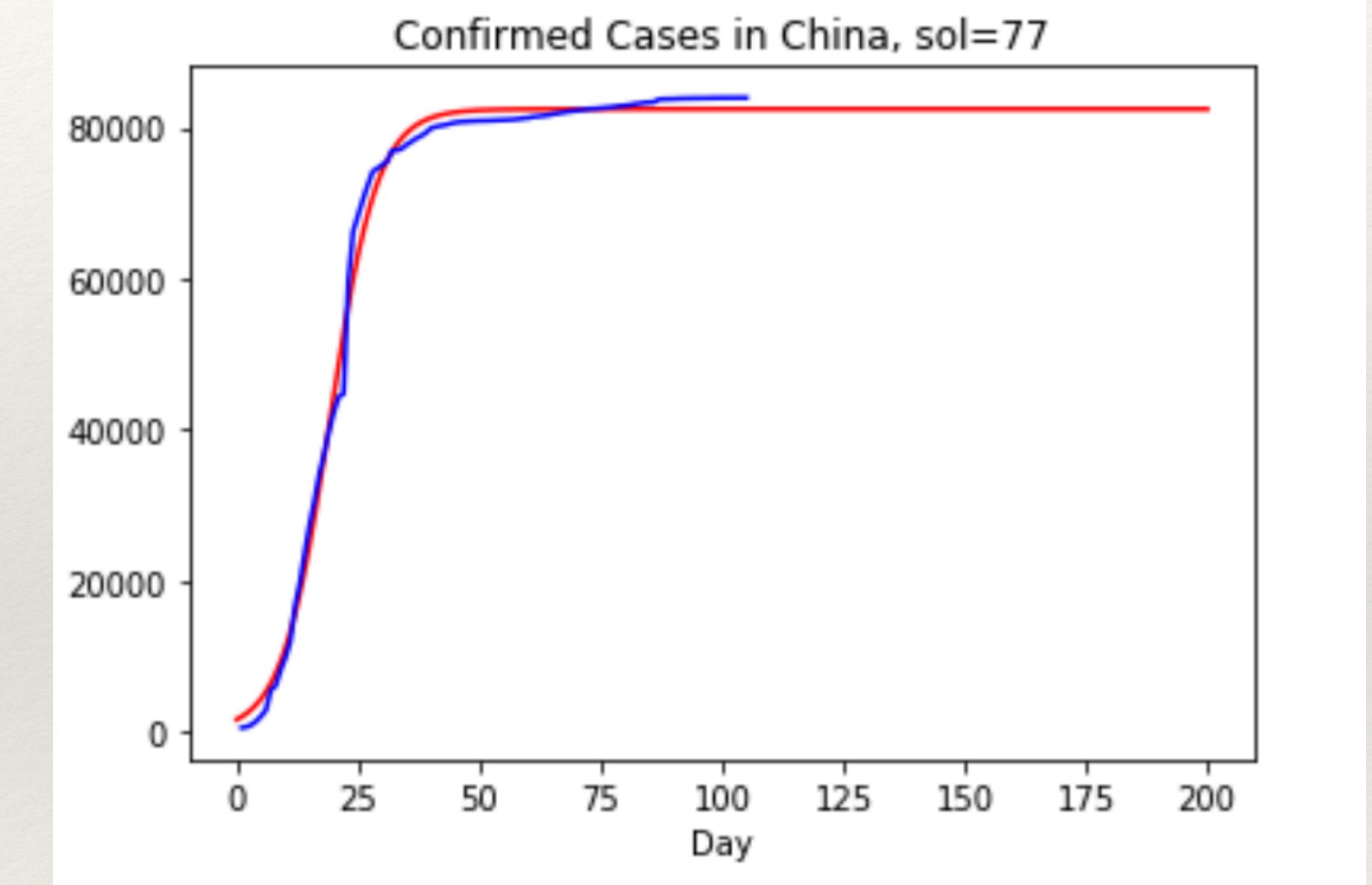
US Data

- ❖ $a = 8.72494646930103$
- ❖ $b = 51.8378166069162$
- ❖ $c = 1258912.3372332316$
- ❖ Test size = 0.21
- ❖ tagged_day_zero = 183
- ❖ train_error = 29.14
- ❖ test_error = 28.07

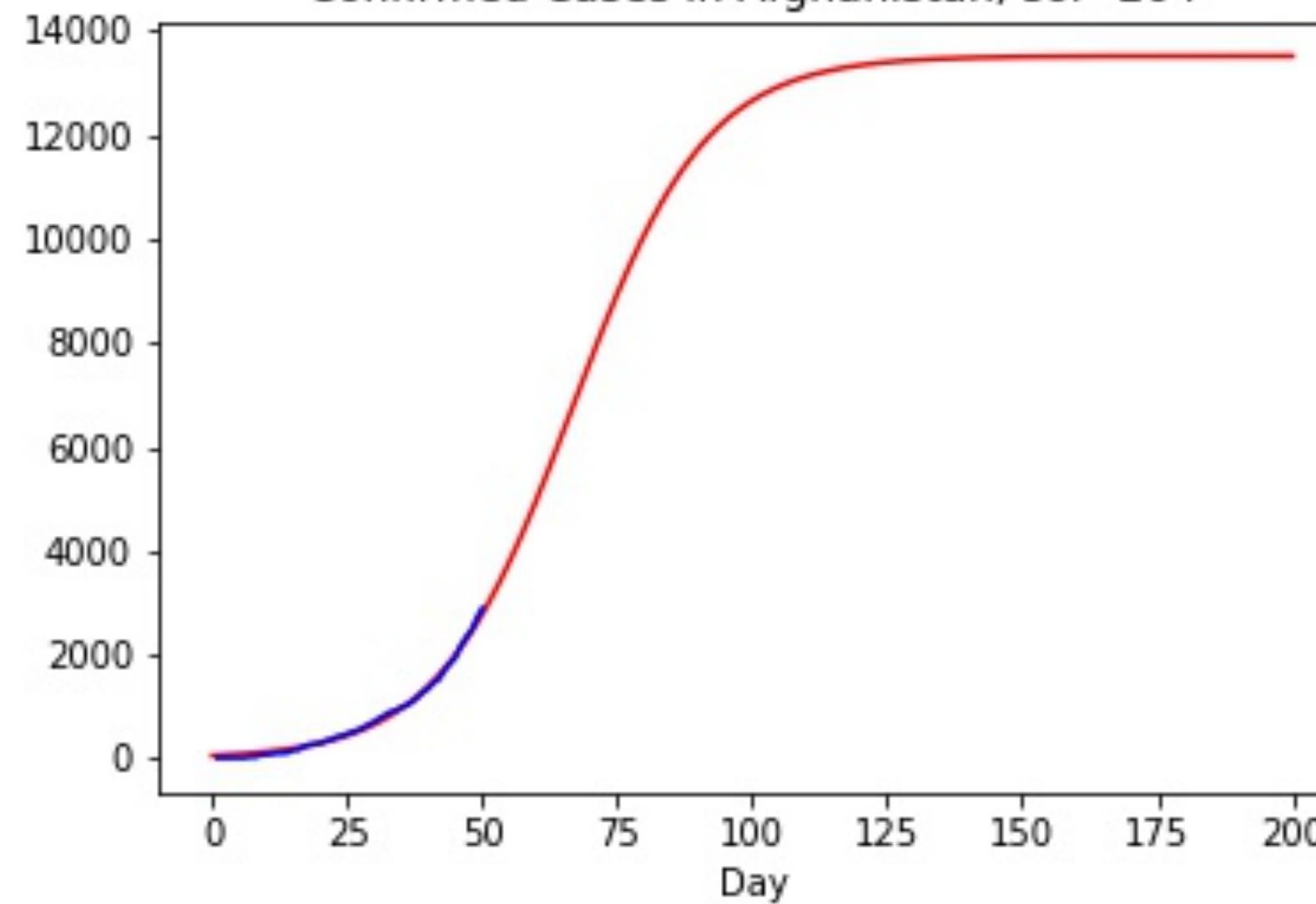


China Data

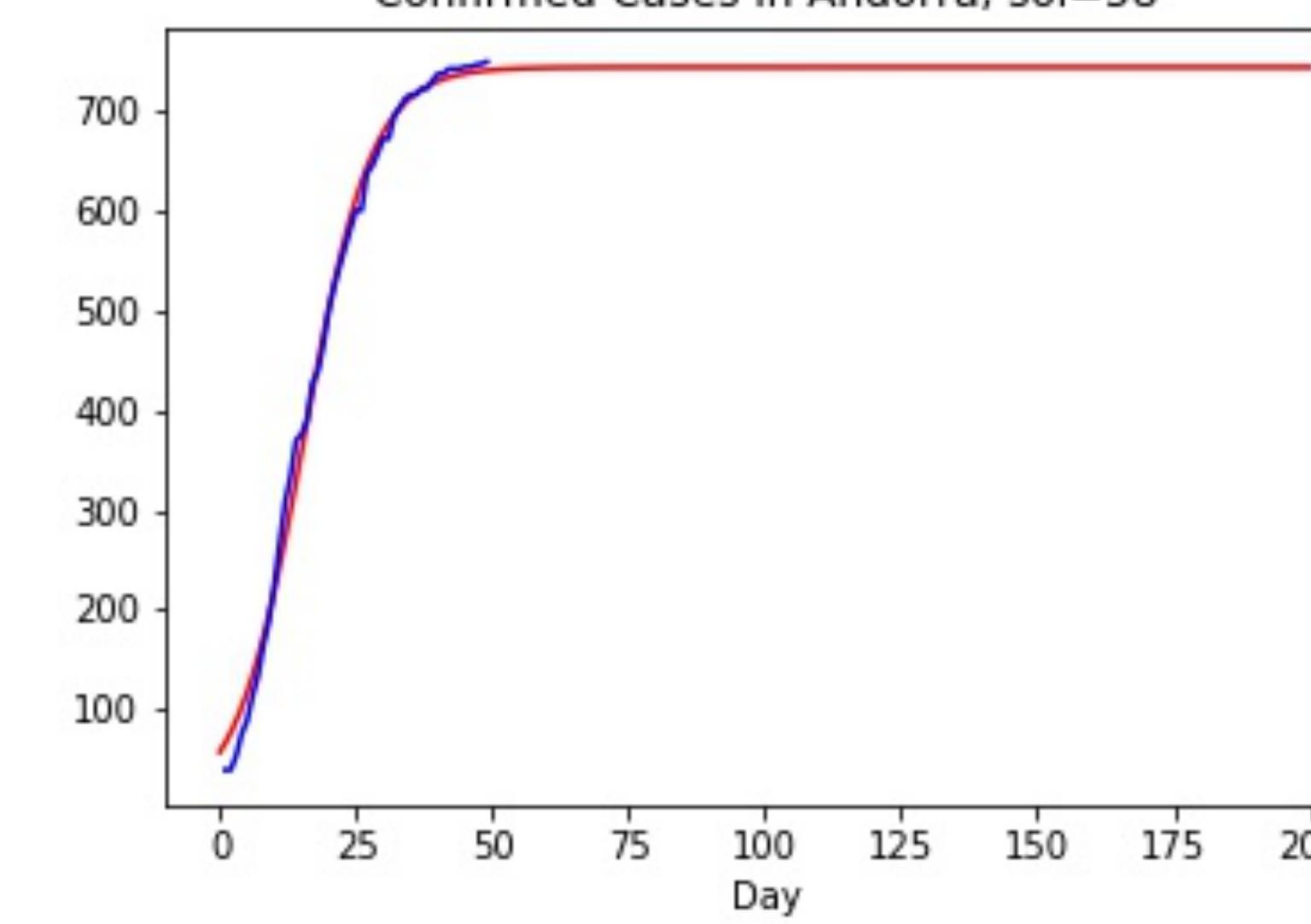
- ❖ $a = 4.913204856494007$
- ❖ $b = 19.158678456846218$
- ❖ $c = 82483.53508770172$
- ❖ Test size = 0.42
- ❖ tagged_day_zero = 77
- ❖ train_error = 0.5001
- ❖ test_error = 0.5179



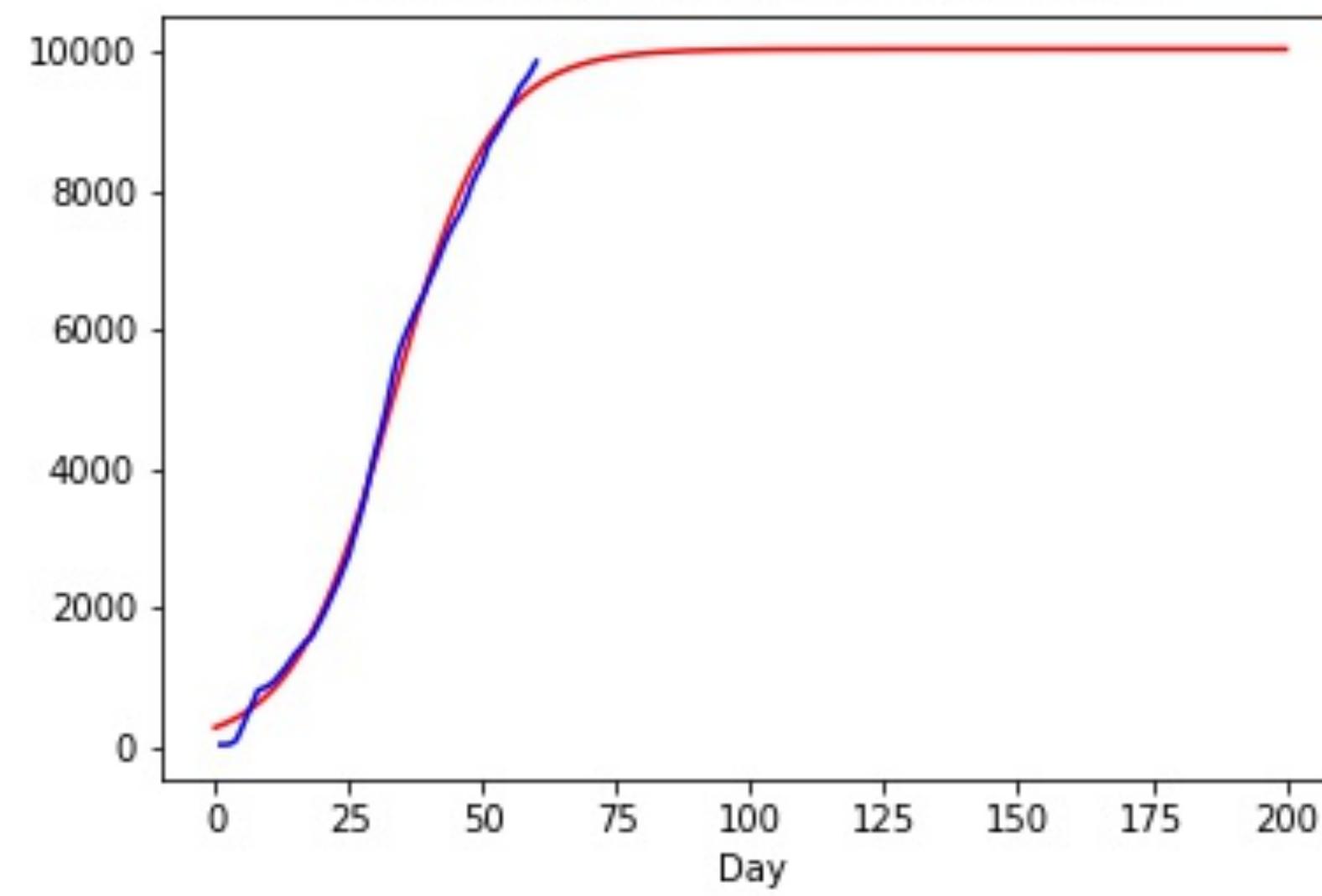
Confirmed Cases in Afghanistan, sol=204



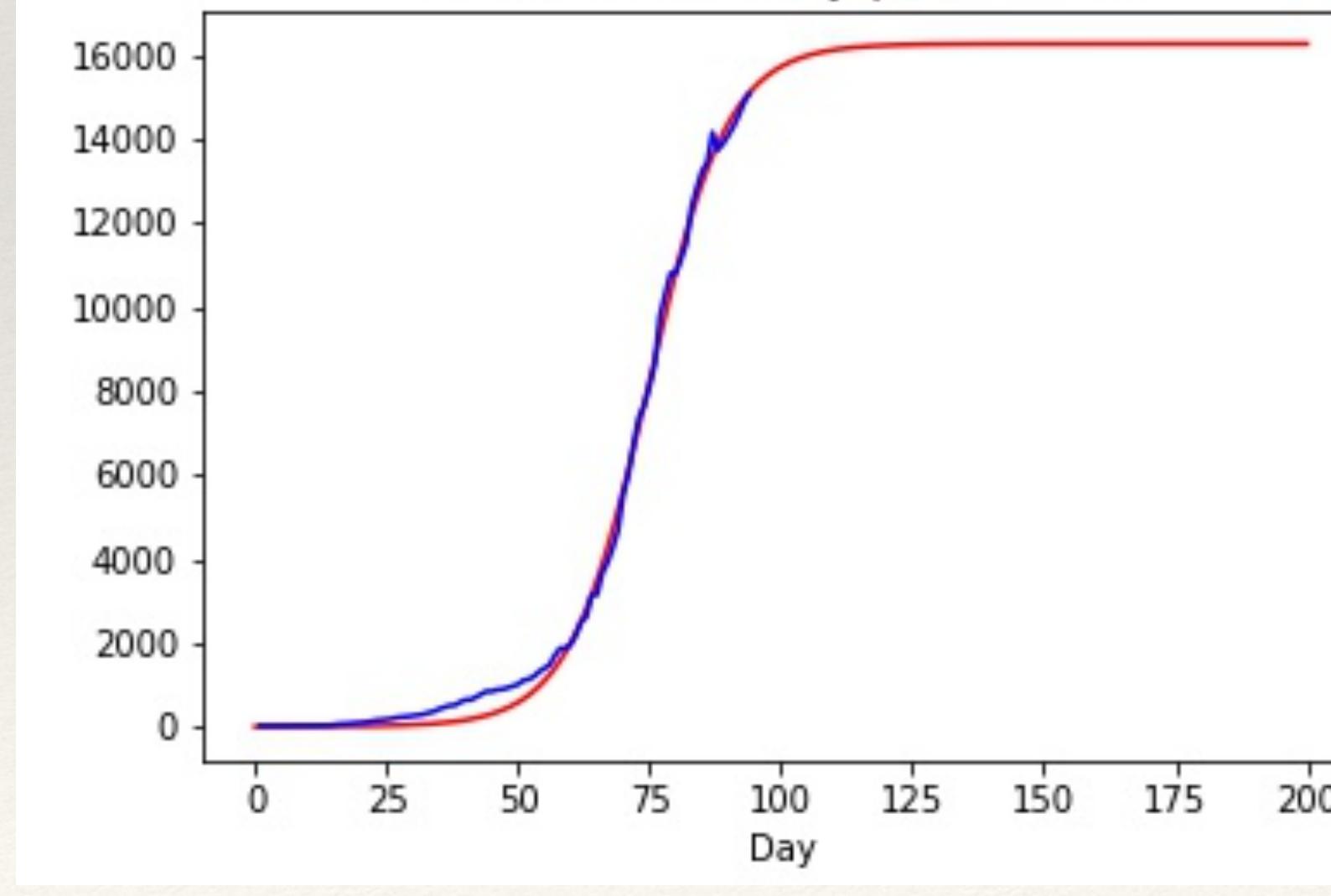
Confirmed Cases in Andorra, sol=58



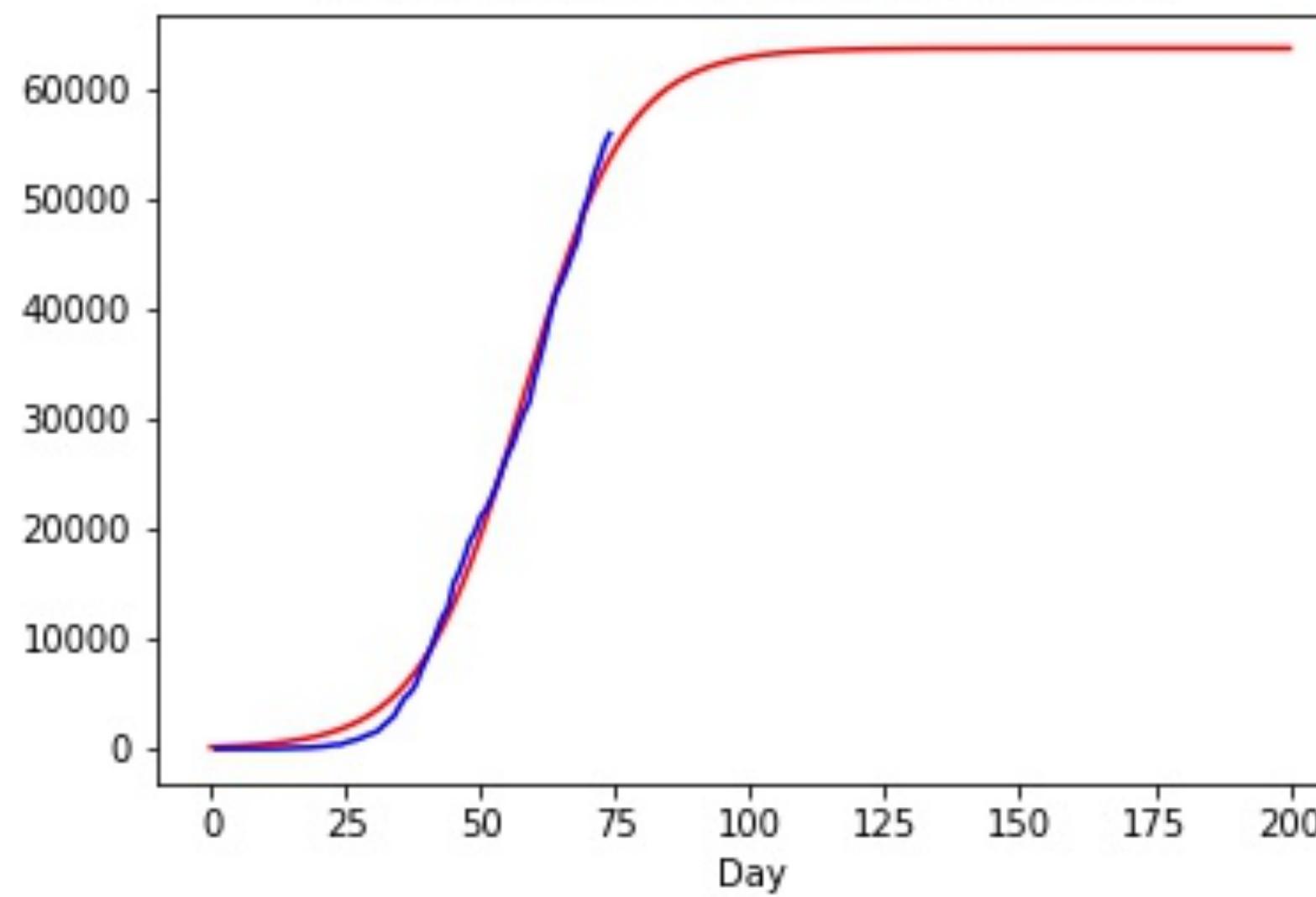
Confirmed Cases in Denmark, sol=120



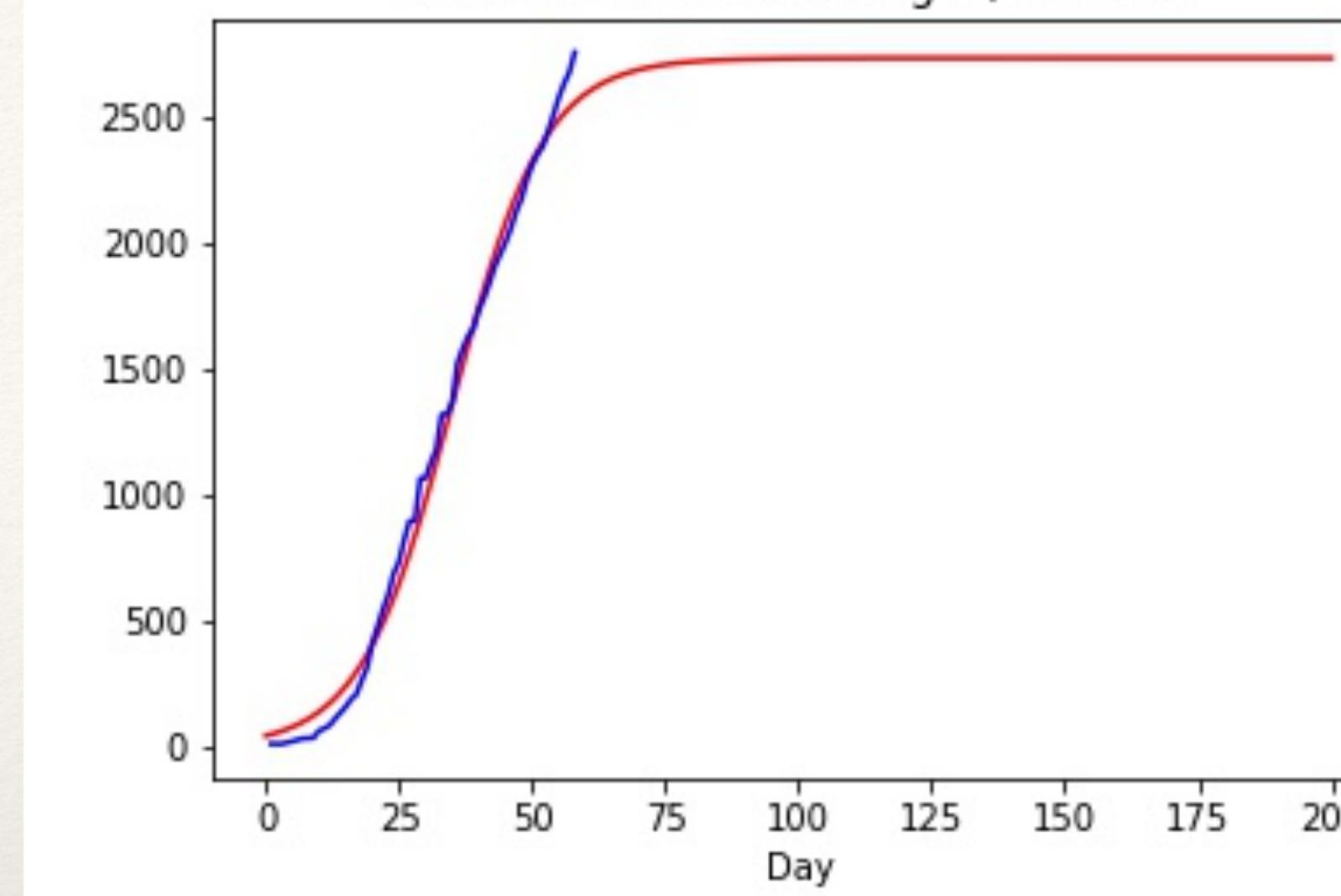
Confirmed Cases in Japan, sol=150



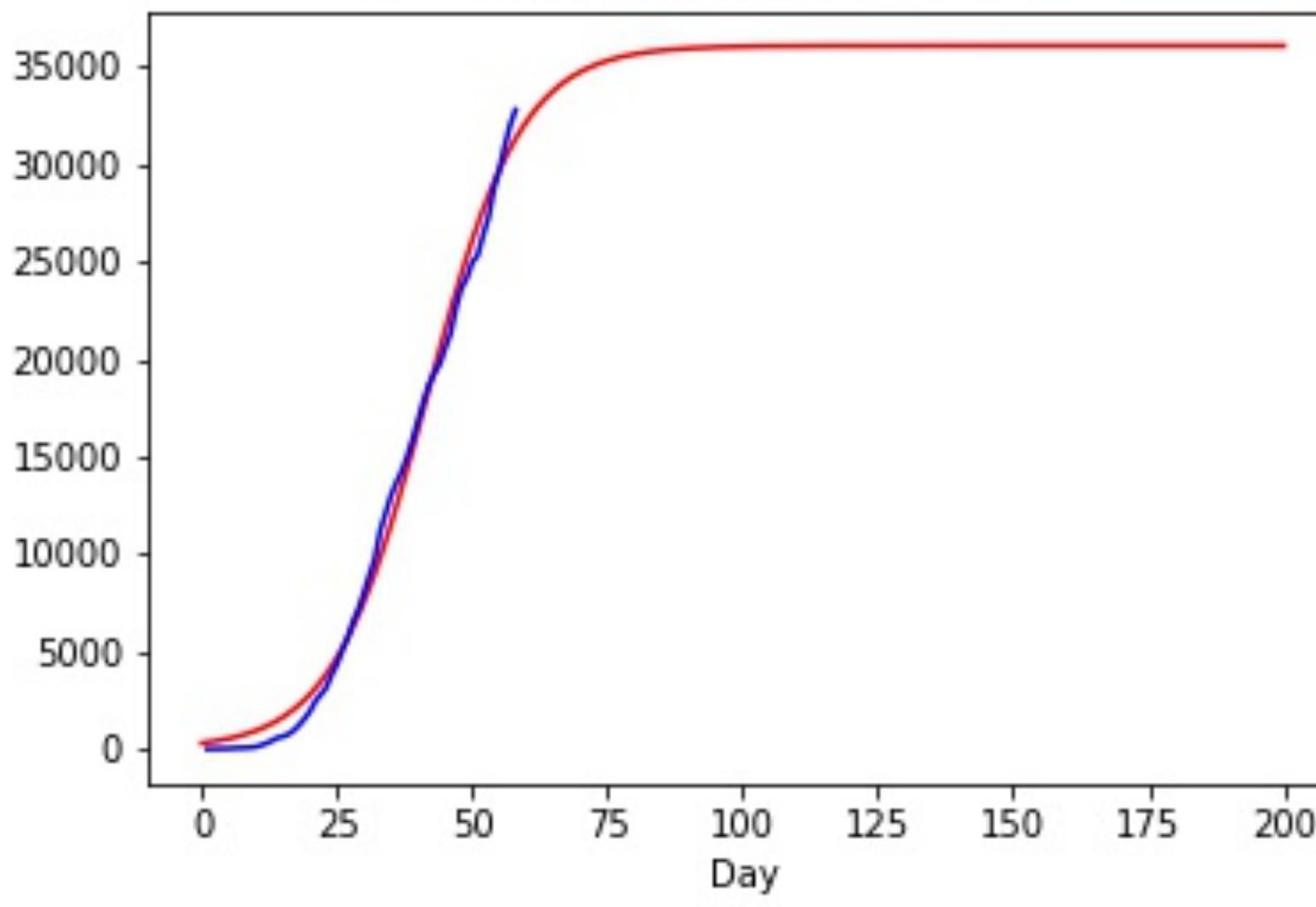
Confirmed Cases in California, sol=168



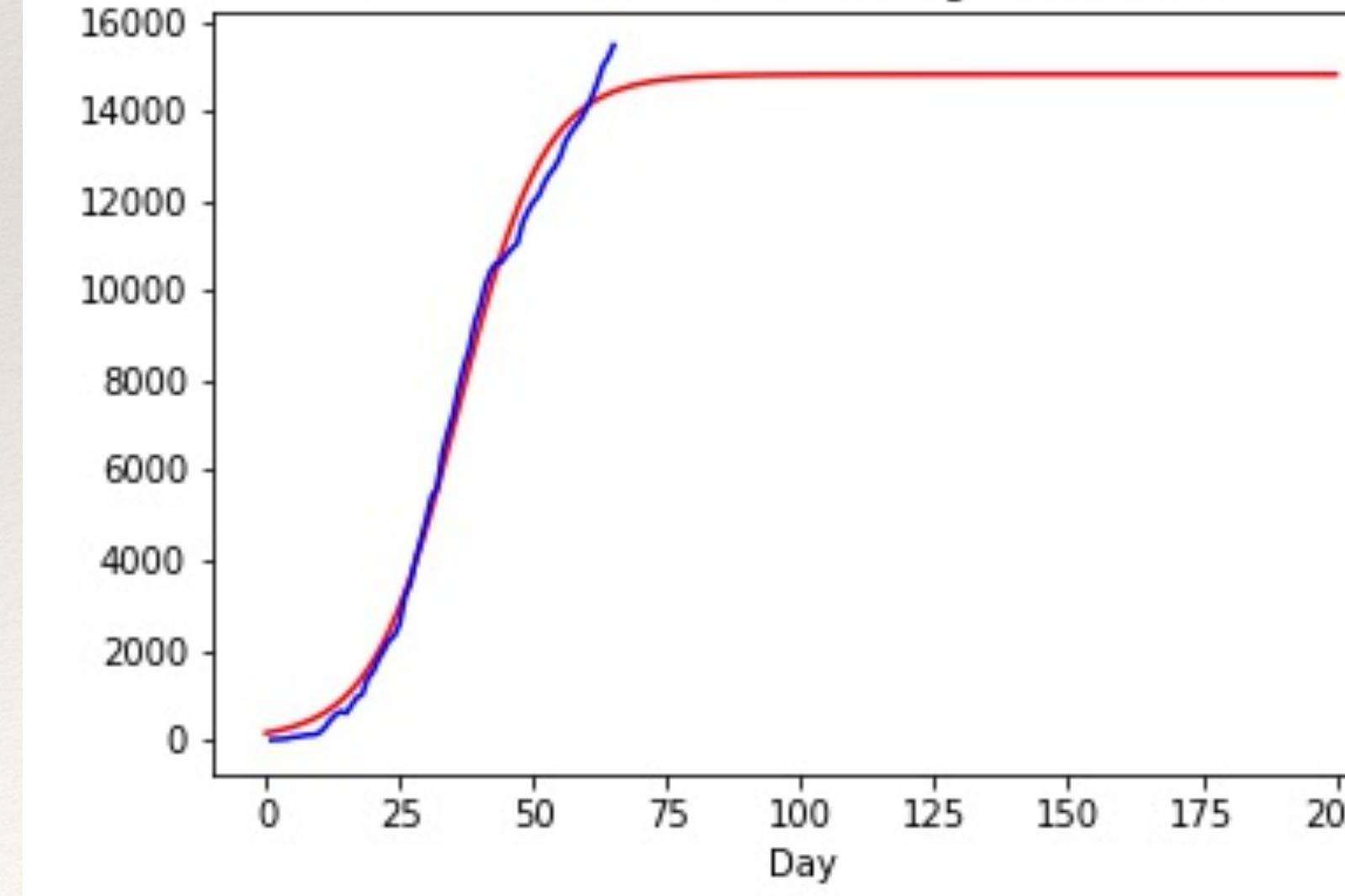
Confirmed Cases in Oregon, sol=125



Confirmed Cases in Texas, sol=137



Confirmed Cases in Washington, sol=114



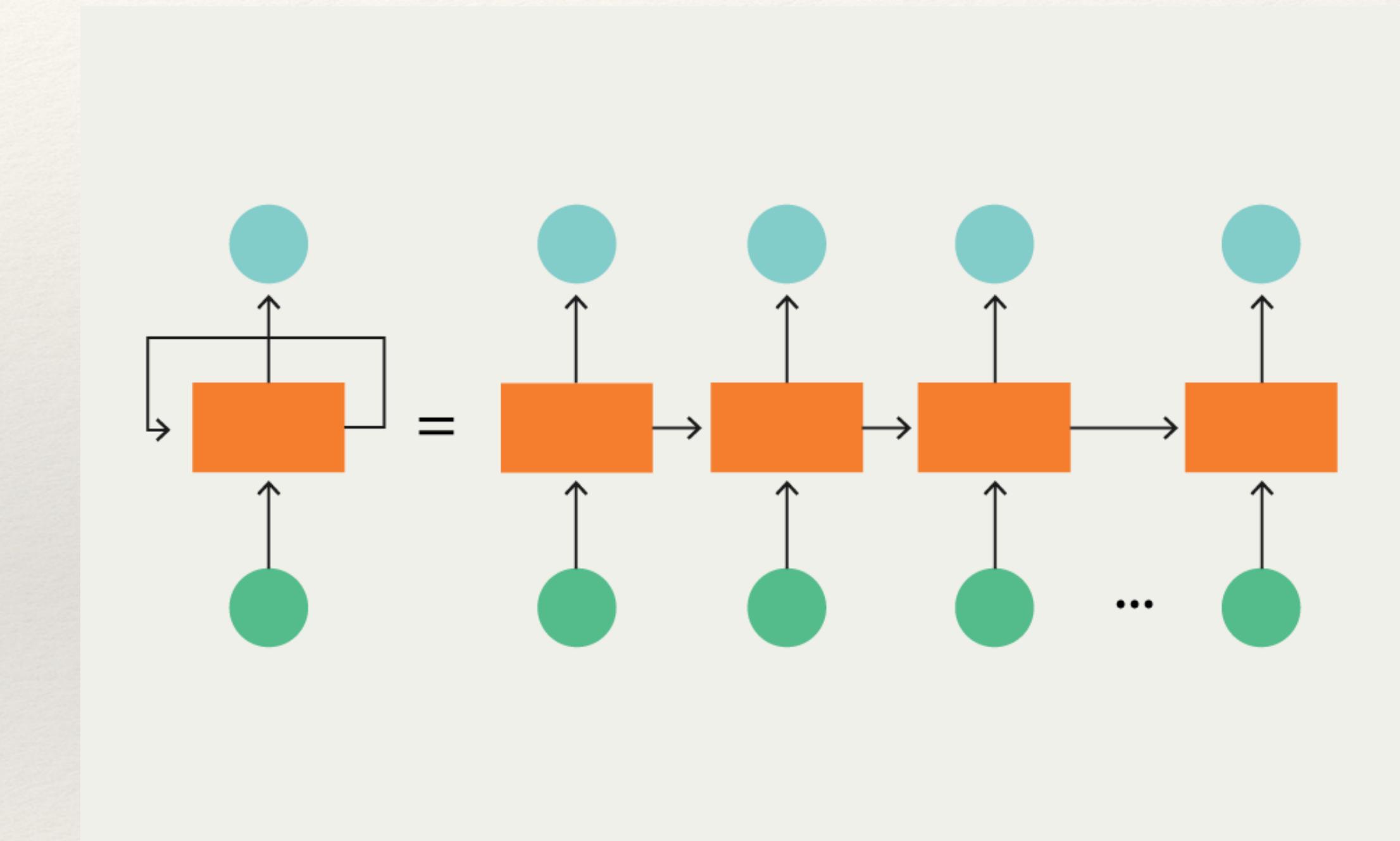
Suggested Test Size

1	country	tagged_day_zero	train_error	test_error	suggested_test_size
2	Afghanistan	182	0.9281744422054777	0.9281725737747879	0.6400000000000005
3	Albania	103	0.22850386959603874	0.22160830188194117	0.6700000000000005
4	Algeria	161	1.9567736143374257	1.9564264904966011	0.4300000000000027
5	Andorra	55	0.19526707736490015	0.1974430394852956	0.4700000000000003
6	Angola	934	0.05919058904525142	0.059145905595821564	0.4100000000000025
7	Antigua and Barbuda	47	0.01915029554631809	0.019220062503140522	0.3700000000000002
8	Argentina	139	1.3814485063659232	1.3808679357851963	0.3500000000000002
9	Armenia	184	1.108819286330219	1.0892054130529674	0.16
10	Australia	79	0.15564767819348307	0.15744579423522528	0.3000000000000016
11	Austria	69	0.6198626533516718	0.5720792001318318	0.6700000000000005
12	Azerbaijan	82	0.3302919427505849	0.33213923486689856	0.2200000000000006
13	Bahamas	74	0.05073747782377197	0.04901079866242805	0.17
14	Bahrain	188	0.19639889999072355	0.19563464409064138	0.5700000000000004
15	Bangladesh	97	0.6305113884402688	0.6235124157975255	0.2700000000000013
16	Barbados	32	0.049679947928361216	0.05046892503018488	0.2700000000000013
17	Belarus	120	0.8842622013095455	0.8830771145736999	0.2900000000000015
18	Belgium	137	3.0608129475097736	3.7596868375107624	0.6500000000000005
19	Belize	-1	-1	-1	-1

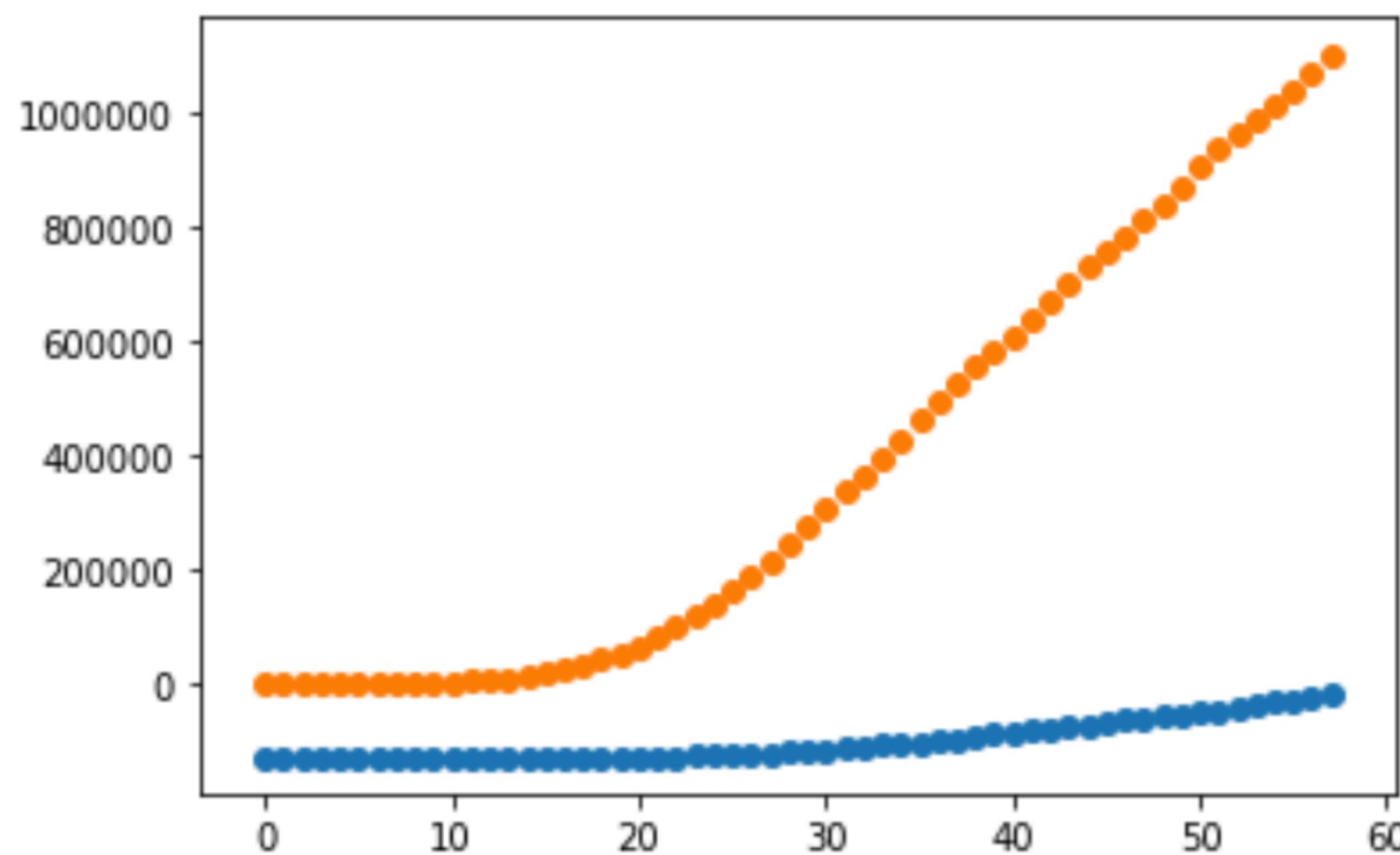
RNN Model

Do Not Suggest to Use Neural Networks

- ❖ Logistic Curve fits the data very well
- ❖ Hard to preprocess the data to feed RNN
- ❖ Not enough data to feed LSTM



Cumulative Cases



Day

Real

Prediction

Future Plan

- ❖ Command line tool
- ❖ Website
- ❖ Maintain COVID19-N20 Data Set



References

- ❖ https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Tennessee
 - ❖ https://en.wikipedia.org/wiki/Coronavirus_disease_2019
 - ❖ <https://github.com/CSSEGISandData/COVID-19>
 - ❖ <https://github.com/secregister01/COVID19-N20>
- ❖

Thank you!