Forecasting the Future of COVID-19 with Social Distancing

Jacob Budnick
Jbudnick5@gmail.com

About Me

- Mechanical / Electrical Engineering Background
- Problem solver
- Branching out, broadening horizons

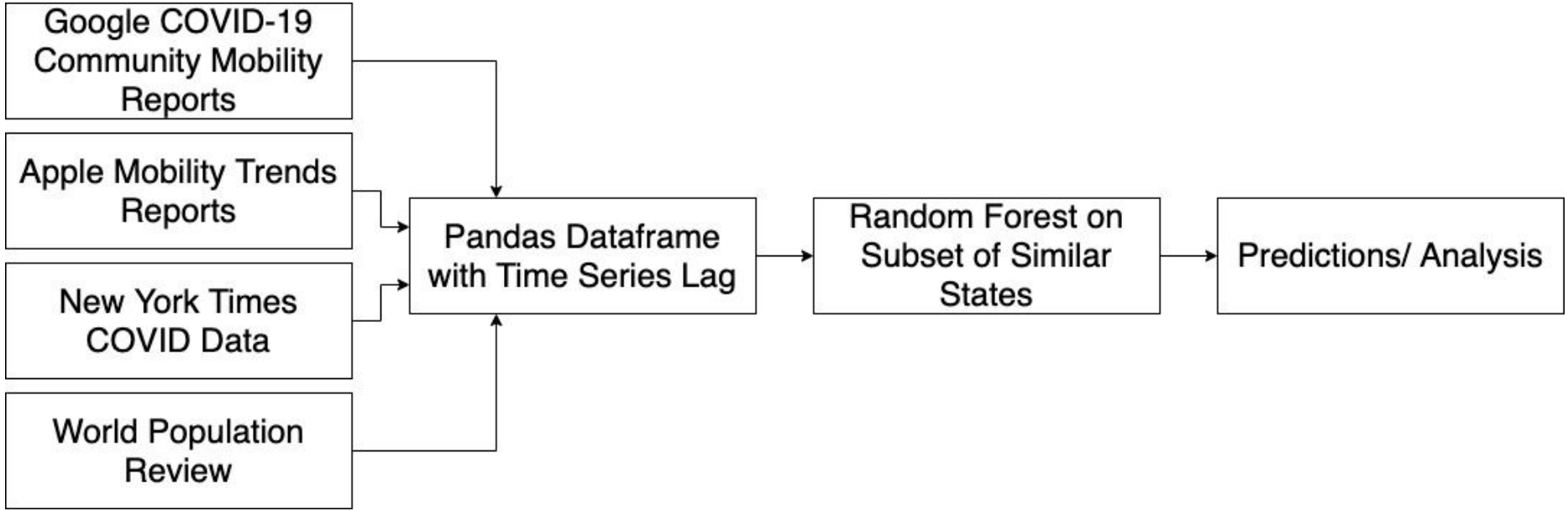
Motivation

What level of social distancing must be maintained so that we see a consistent drop in the number of new cases?

- COVID - 19 Pandemic Effects

- Possibility of Second Spike in New Infections

Data Sources / Process



Forecasting Model

- Days Elapsed from First Confirmed Cases
- Apple/Google Social Distancing Parameters (7 day moving average applied)
- Number of Daily New Cases over 21 days (7 day moving average applied)
- State Population Densities

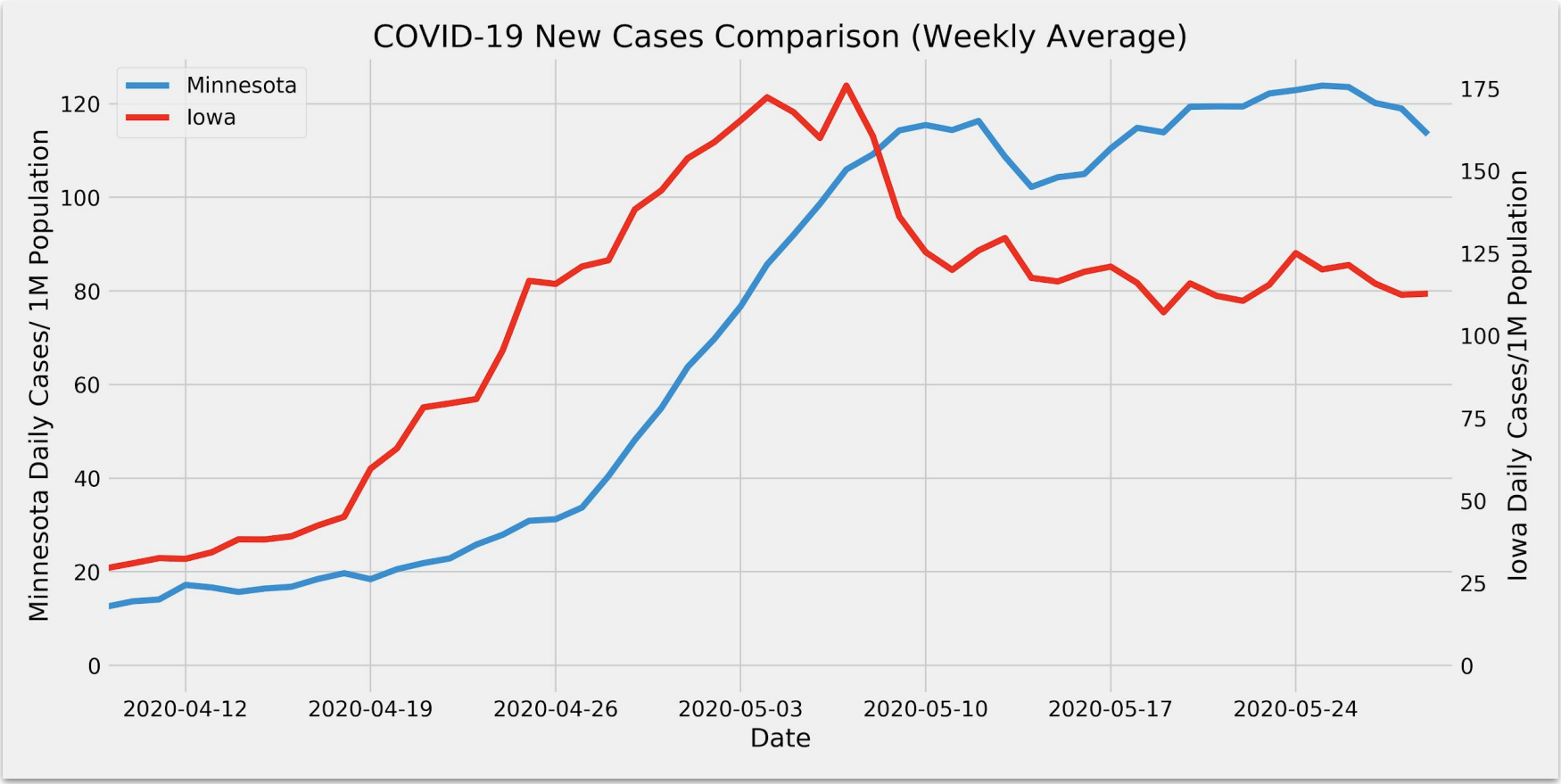
Total 30 features

Assumptions

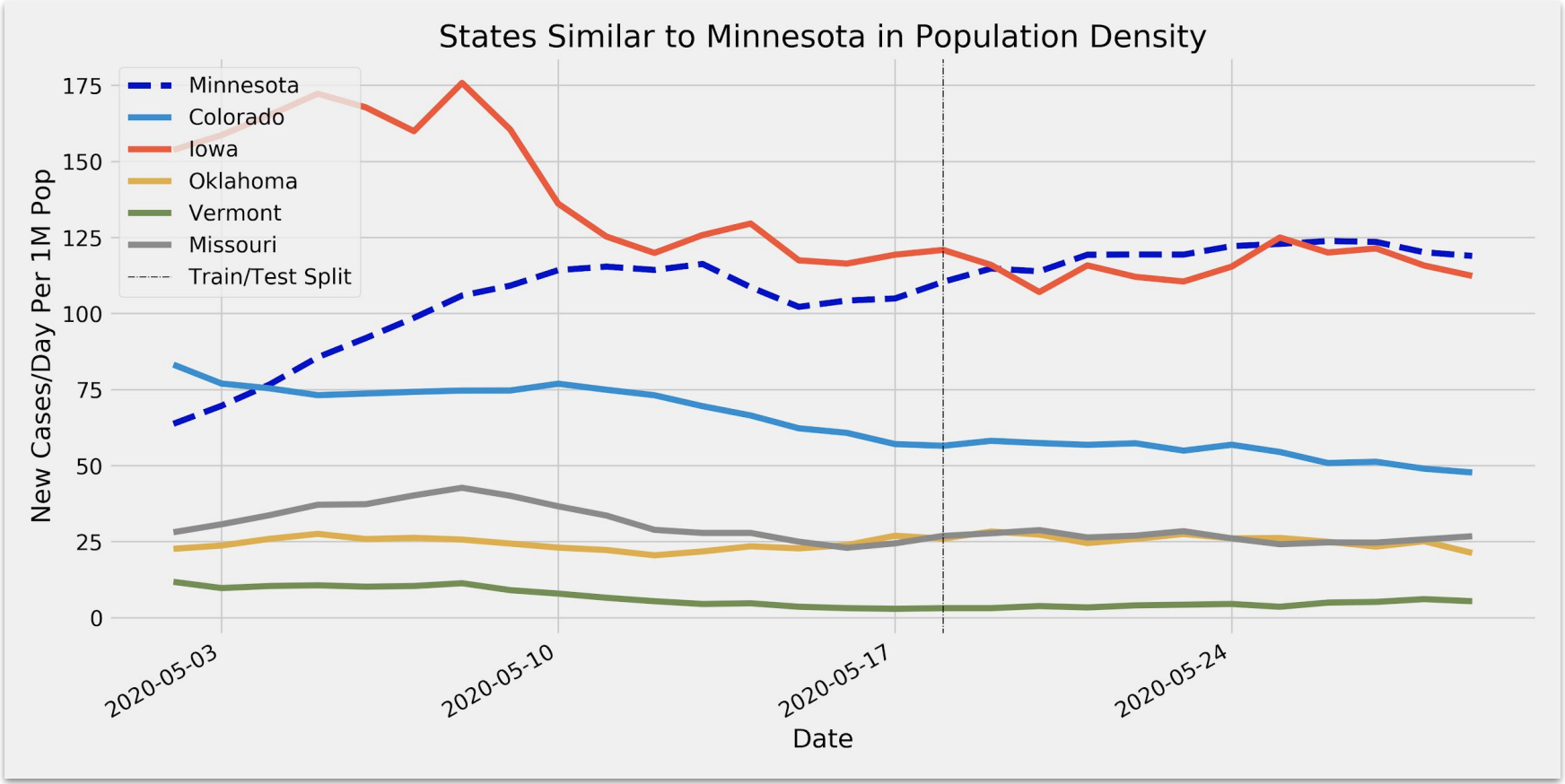
- Medical knowledge/awareness of virus is correlated with time since first infections

- Population distribution throughout each individual state is uniform (for simplicity)

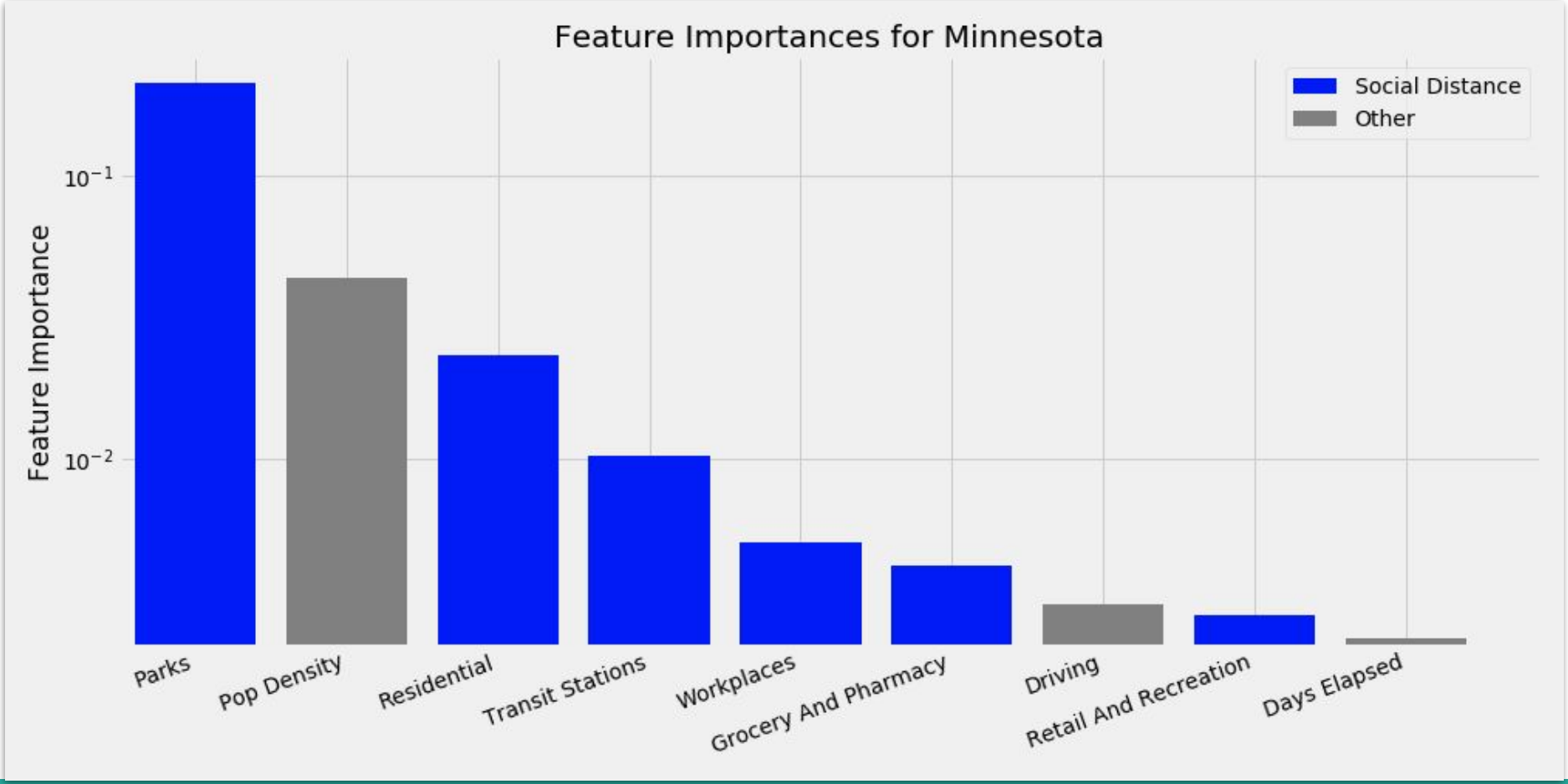
Training Focus



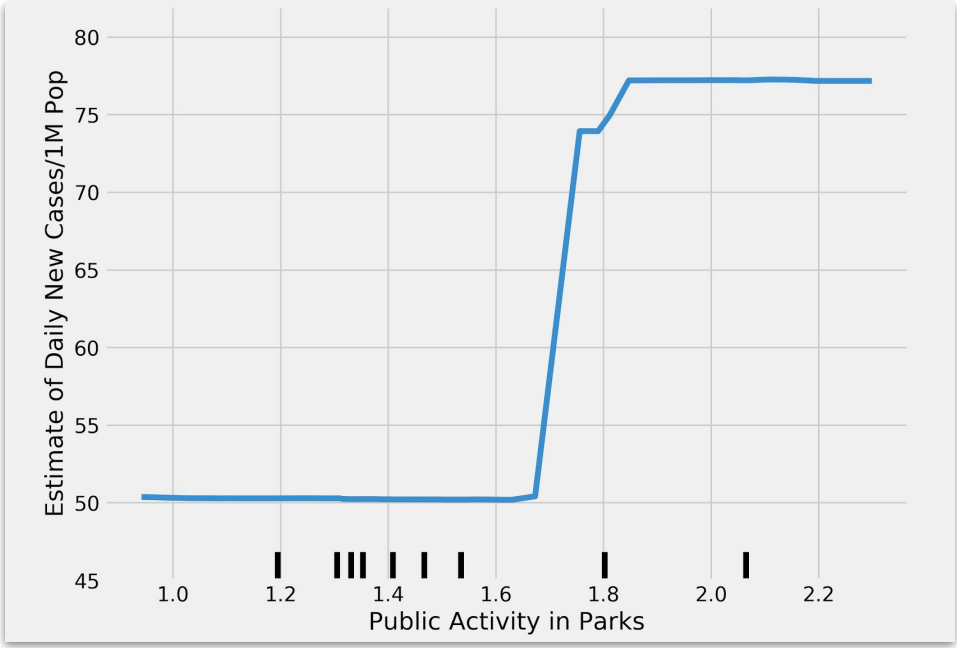
Training Focus - Random Forest Regression Model

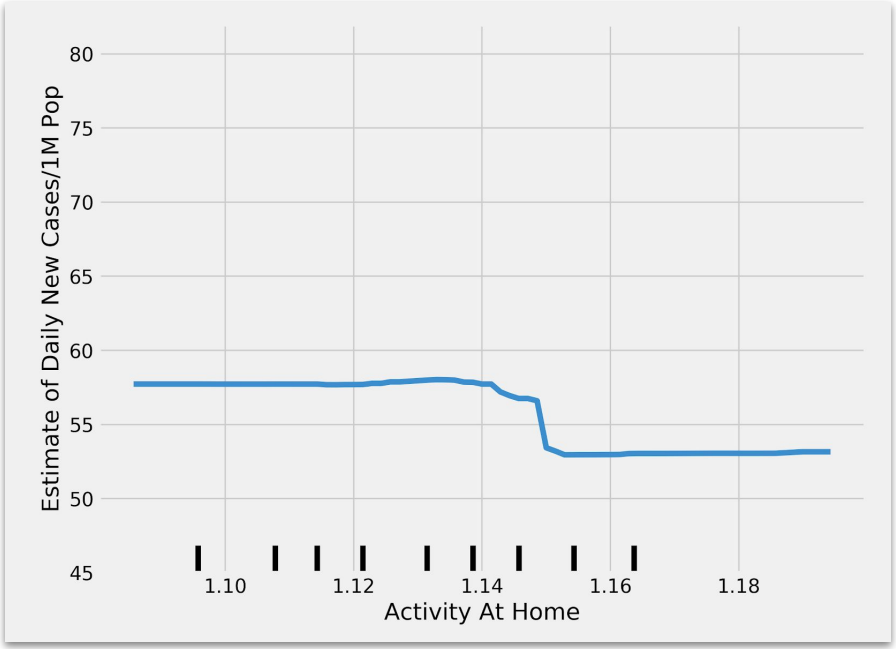


Feature Importances

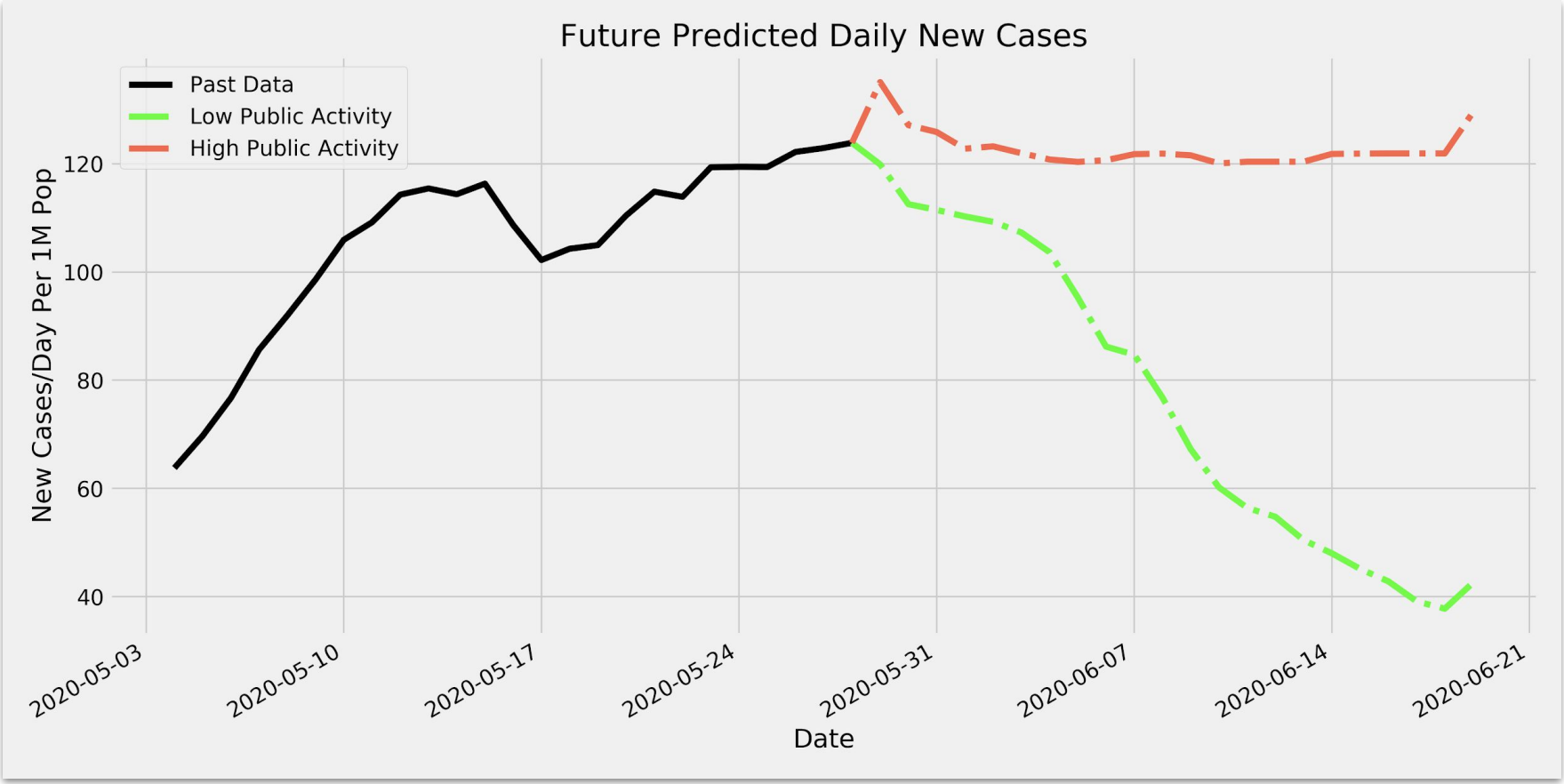


Feature Importances - Partial Dependency Plots





Minnesota Forecast



Minnesota Recommendations

- Promote social distancing at parks and public areas; limit attendance if possible
- Continue to encourage people to stay home, exercise caution when outside
- Continue to monitor and discourage public activity

Thank you for your attention. Questions?

Jacob Budnick

Jbudnick5@gmail.com

Github: /Jbudnick

LinkedIn: /JacobBudnick

Pandas

scikits learn

machine learning in Python

Sources

https://www.xifin.com/resources/blog/202004/covid-19-maintaining-momentum-during-volume-recessions (Header Image)

https://towardsdatascience.com/exploratory-data-analysis-with-pandas-508a5e8a5964 (pandas image)

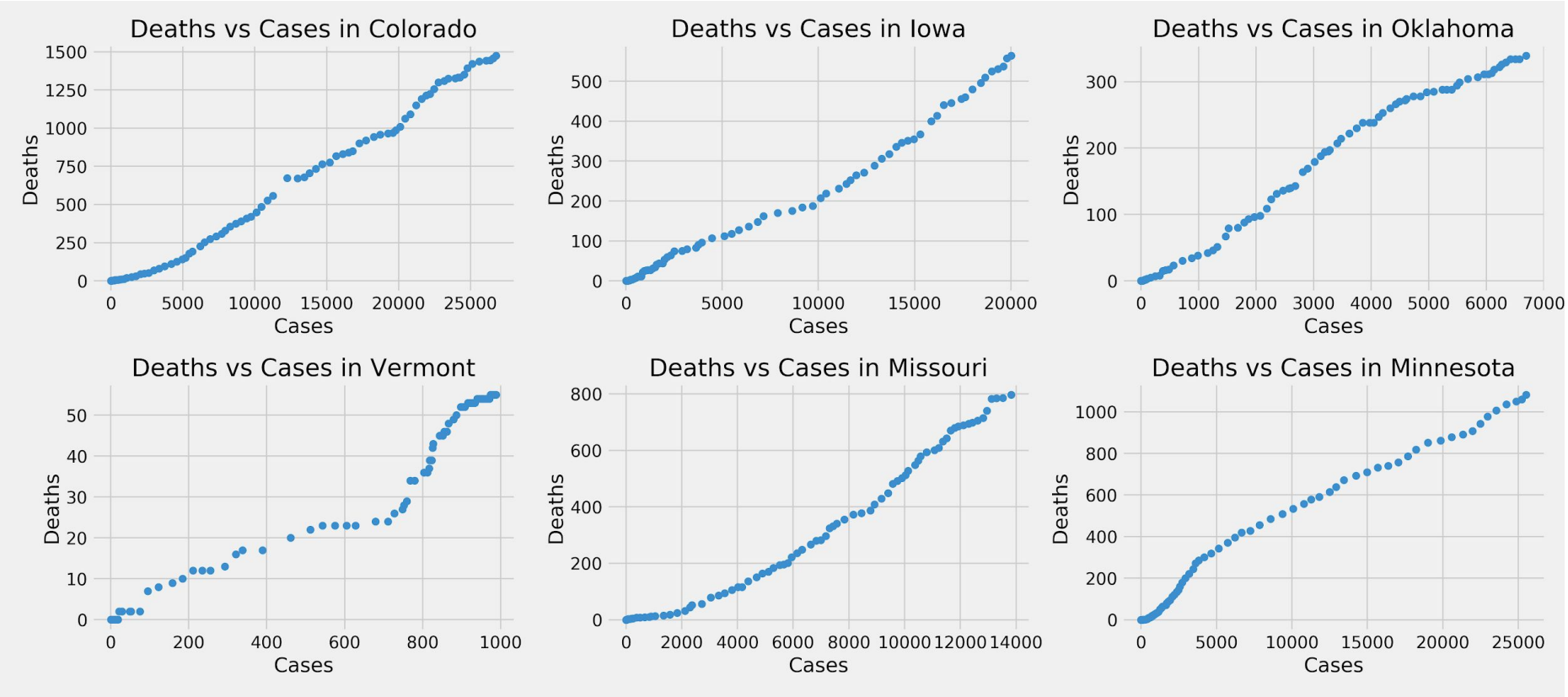
https://www.google.com/covid19/mobility/ (Google Data)

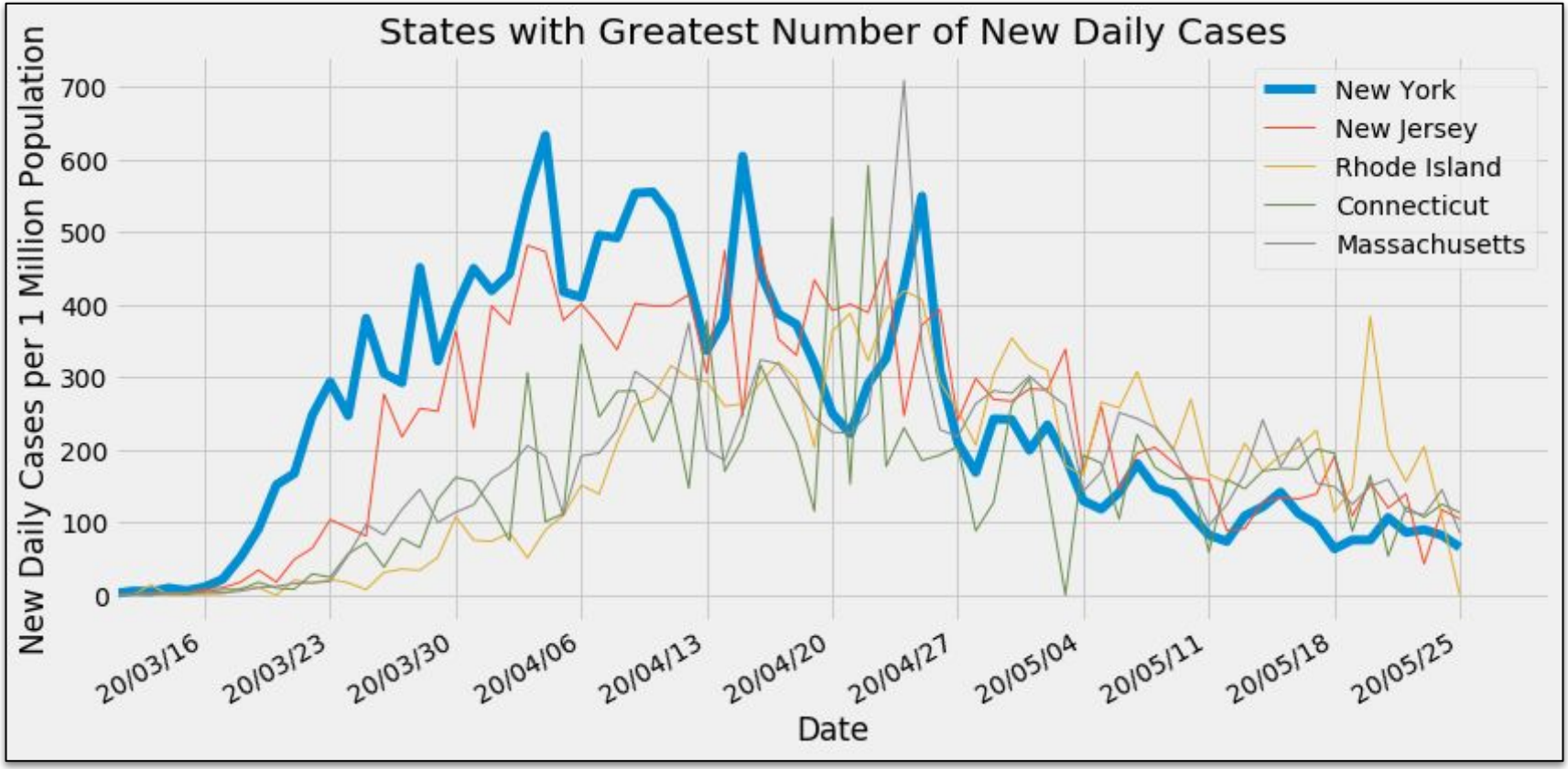
https://github.com/nytimes/covid-19-data/blob/master/us-states.csv

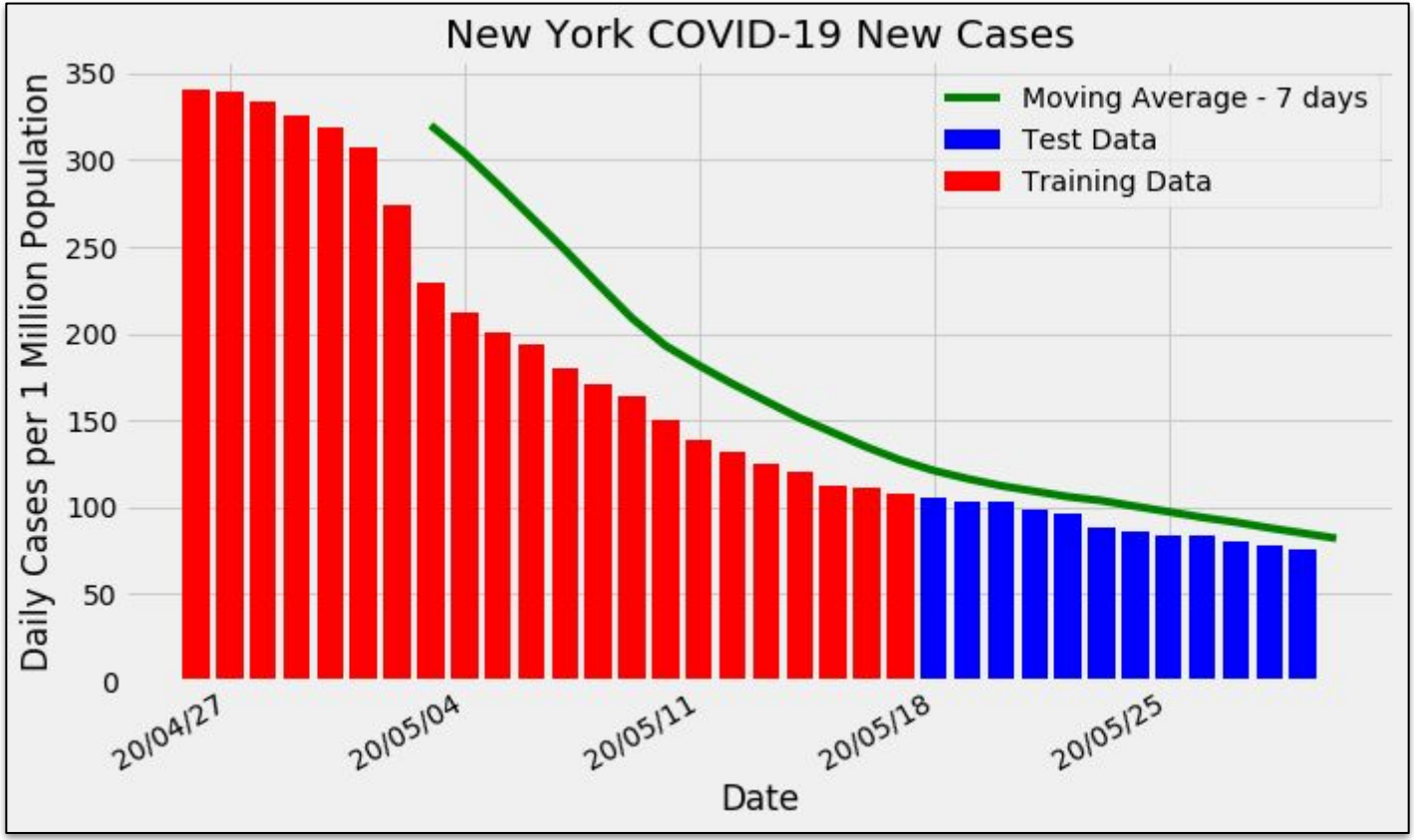
https://www.apple.com/covid19/mobility

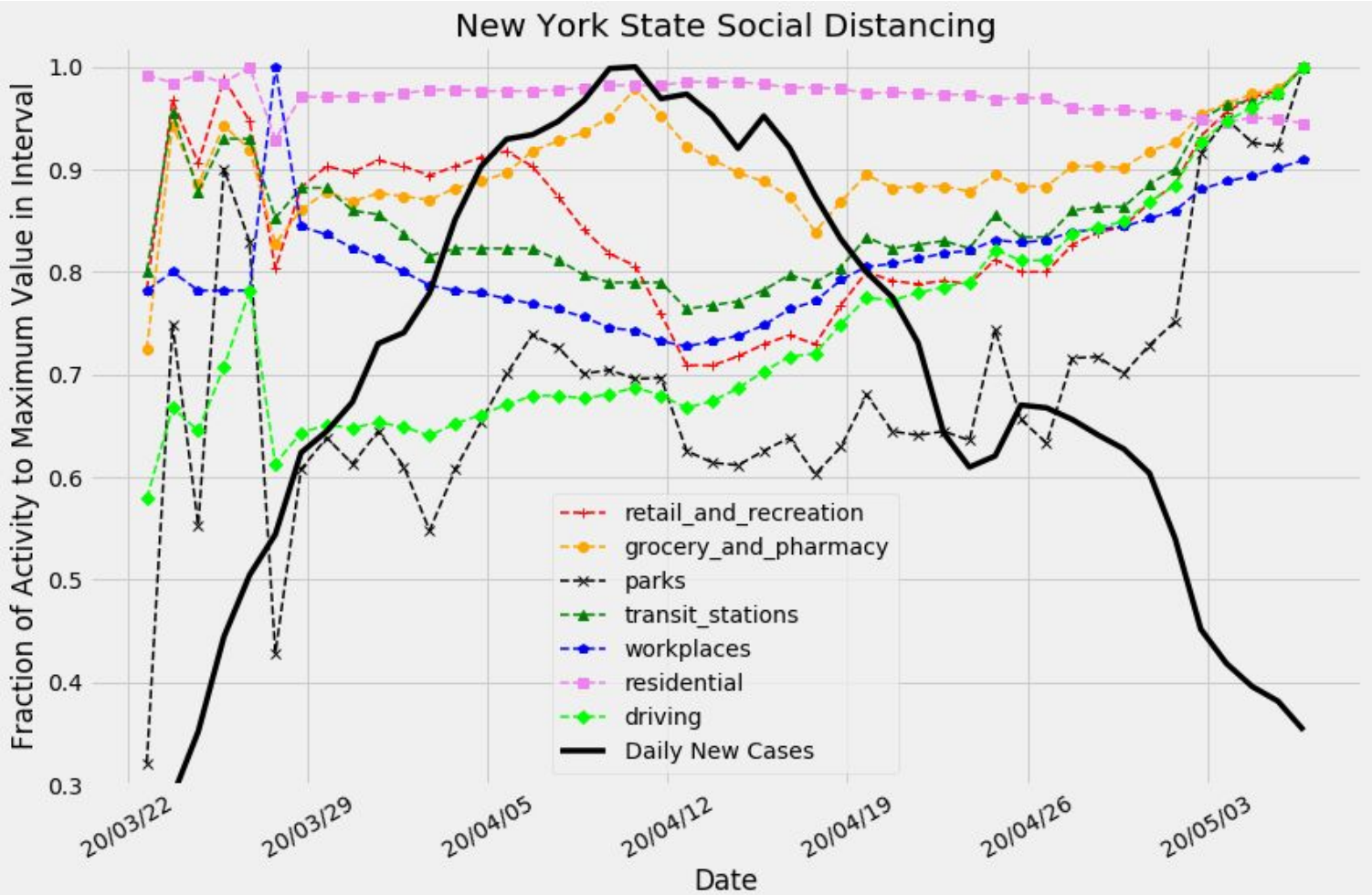
https://worldpopulationreview.com/states/

Appendix

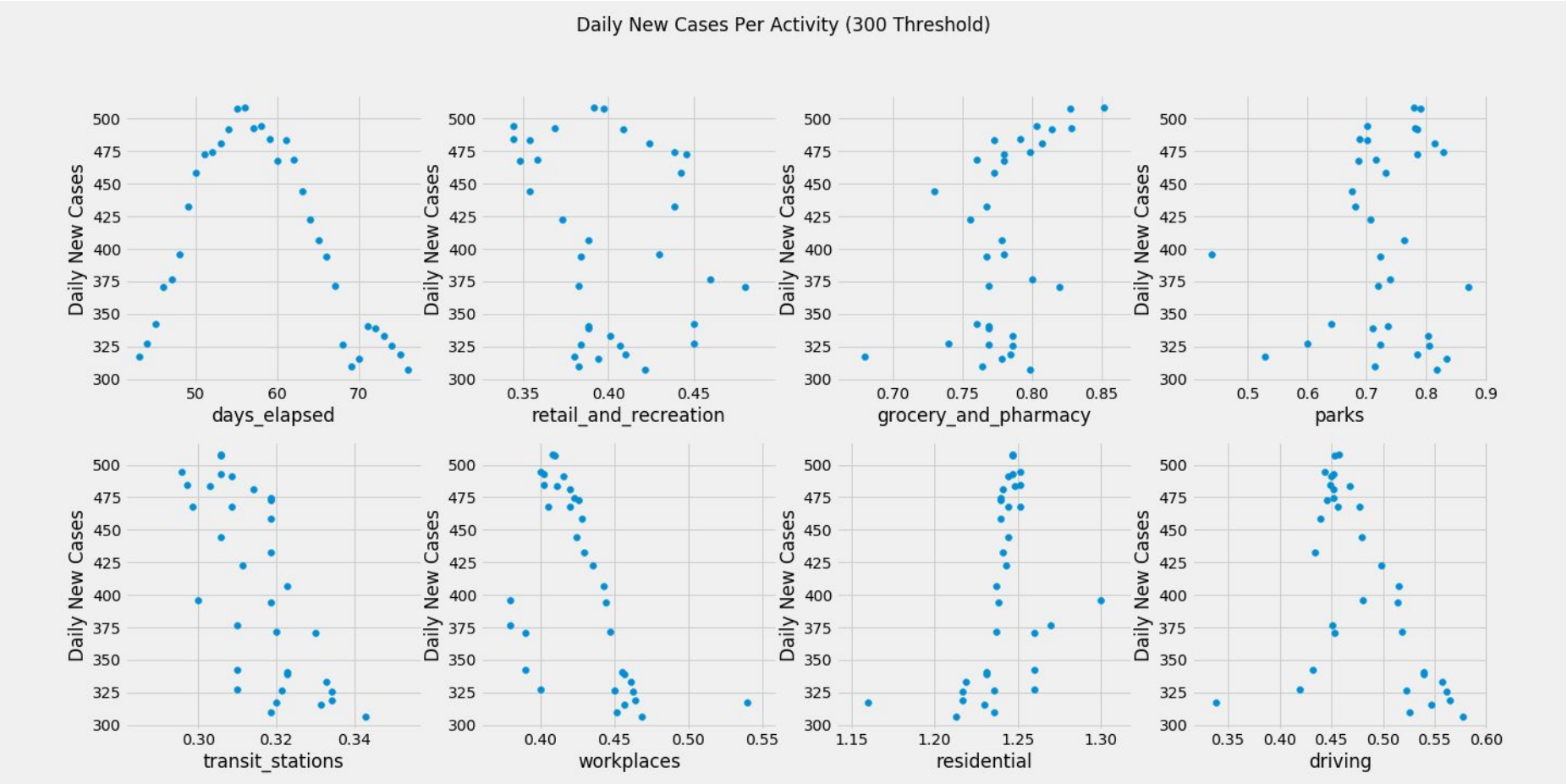

Deaths vs Cases in Colorado
Deaths
Cases
0
250
500
750
1000
1250
1500
0
5000
10000
15000
20000
25000
Deaths vs Cases in Iowa
Deaths
Cases
0
100
200
300
400
500
0
5000
10000
15000
20000
Deaths vs Cases in Oklahoma
Deaths
Cases
0
100
200
300
0
1000
2000
3000
4000
5000
6000
7000
Deaths vs Cases in Vermont
Deaths
Cases
0
10
20
30
40
50
0
200
400
600
800
1000
Deaths vs Cases in Missouri
Deaths
Cases
0
200
400
600
800
0
2000
4000
6000
8000
10000
12000
14000
Deaths vs Cases in Minnesota
Deaths
Cases
0
200
400
600
800
1000
0
5000
10000
15000
20000
25000


States with Greatest Number of New Daily Cases
New Daily Cases per 1 Million Population
700
600
500
400
300
200
100
0
20/03/16
20/03/23
20/03/30
20/04/06
20/04/13
20/04/20
20/04/27
20/05/04
20/05/11
20/05/18
20/05/25
Date
New York
New Jersey
Rhode Island
Connecticut
Massachusetts


New York COVID-19 New Cases
Daily Cases per 1 Million Population
350
300
250
200
150
100
50
0
Moving Average - 7 days
Test Data
Training Data
20/04/27
20/05/04
20/05/11
20/05/18
20/05/25
Date


New York State Social Distancing
Fraction of Activity to Maximum Value in Interval
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
20/03/22
20/03/29
20/04/05
20/04/12
20/04/19
20/04/26
20/05/03
Date
retail_and_recreation
grocery_and_pharmacy
parks
transit_stations
workplaces
residential
driving
Daily New Cases

NY Data


Daily New Cases Per Activity (300 Threshold)
Daily New Cases
days_elapsed
retail_and_recreation
grocery_and_pharmacy
parks
transit_stations
workplaces
residential
driving

Raw Data Sample

country_region_code	country_region	sub_region_1	date	retail_and_recreation_percent_change_from_baseline	grocery_and_pharmacy_percent_change_from_baseline	parks_percent_change_from_baseline	transit_stations_percent_change_from_baseline	workplaces_percent_change_from_baseline	residential_percent_change_from_baseline
US	United States	New York	2020-02-15	3	-4	-2	-4	1	0
US	United States	New York	2020-02-16	7	-3	13	0	0	-1
US	United States	New York	2020-02-17	7	-2	23	-22	-37	9
US	United States	New York	2020-02-18	-2	-5	-4	-5	-10	3
US	United States	New York	2020-02-19	3	-2	11	-3	-9	2
US	United States	New York	2020-02-20	2	-3	0	-3	-9	2
US	United States	New York	2020-02-21	-1	-5	-8	-6	-9	3
US	United States	New York	2020-02-22	8	0	29	3	4	-1
US	United States	New York	2020-02-23	9	2	54	3	3	-1
US	United States	New York	2020-02-24	3	0	25	-1	4	0
US	United States	New York	2020-02-25	-1	-2	-1	-1	3	0
US	United States	New York	2020-02-26	4	3	-1	0	3	0
US	United States	New York	2020-02-27	-1	-2	-4	0	0	1
US	United States	New York	2020-02-28	1	0	-9	0	3	1
US	United States	New York	2020-02-29	7	1	-5	-1	4	0
US	United States	New York	2020-03-01	10	5	9	0	3	-1

ses_per_pop(t-1)	days_elapsed(t)	retail_and_recreation(t)	grocery_and_pharmacy(t)	parks(t)	transit_stations(t)	workplaces(t)	residential(t)	driving(t)	pop_density(t)
54.980996	77.0	0.615714	0.875714	1.601429	0.470000	0.561429	1.184286	0.680129	71.592186
63.776953	78.0	0.618571	0.877143	1.654286	0.474286	0.562857	1.181429	0.693671	71.592186
69.716104	79.0	0.621429	0.881429	1.708571	0.475714	0.562857	1.178571	0.704314	71.592186
76.707762	80.0	0.625714	0.888571	1.745714	0.480000	0.564286	1.175714	0.711814	71.592186
85.654077	81.0	0.631429	0.892857	1.787143	0.484286	0.565714	1.172857	0.717729	71.592186
91.969123	82.0	0.644286	0.904286	1.930000	0.494286	0.567143	1.170000	0.730271	71.592186
98.609946	83.0	0.642857	0.901429	1.860000	0.490000	0.565714	1.161429	0.729257	71.592186
105.952440	84.0	0.645714	0.905714	1.774286	0.490000	0.567143	1.161429	0.732543	71.592186
109.185142	85.0	0.657143	0.917143	1.871429	0.498571	0.567143	1.157143	0.748671	71.592186
114.297322	86.0	0.665714	0.930000	1.877143	0.507143	0.568571	1.154286	0.768214	71.592186
115.450068	87.0	0.672857	0.941429	1.905714	0.515714	0.570000	1.151429	0.782971	71.592186
114.372501	88.0	0.682857	0.954286	1.955714	0.525714	0.575714	1.148571	0.797771	71.592186
116.327158	89.0	0.685714	0.962857	1.890000	0.527143	0.578571	1.147143	0.808171	71.592186
108.683948	90.0	0.700000	0.984286	1.987143	0.540000	0.582857	1.152857	0.830943	71.592186
102.218543	91.0	0.705714	0.992857	2.051429	0.545714	0.585714	1.151429	0.845371	71.592186
104.298499	92.0	0.705714	0.998571	1.995714	0.544286	0.588571	1.152857	0.849700	71.592186
104.975111	93.0	0.707143	1.002857	1.975714	0.542857	0.590000	1.152857	0.858729	71.592186
110.438127	94.0	0.705714	1.005714	1.871429	0.532857	0.591429	1.155714	0.857000	71.592186

State	People/Square Mile	Recovery Factor
Colorado	56.4	2.0
Iowa	56.9	1.6
Oklahoma	57.7	1.4
Vermont	68.1	12.1
Missouri	89.7	1.6
Minnesota	71.6	1.0

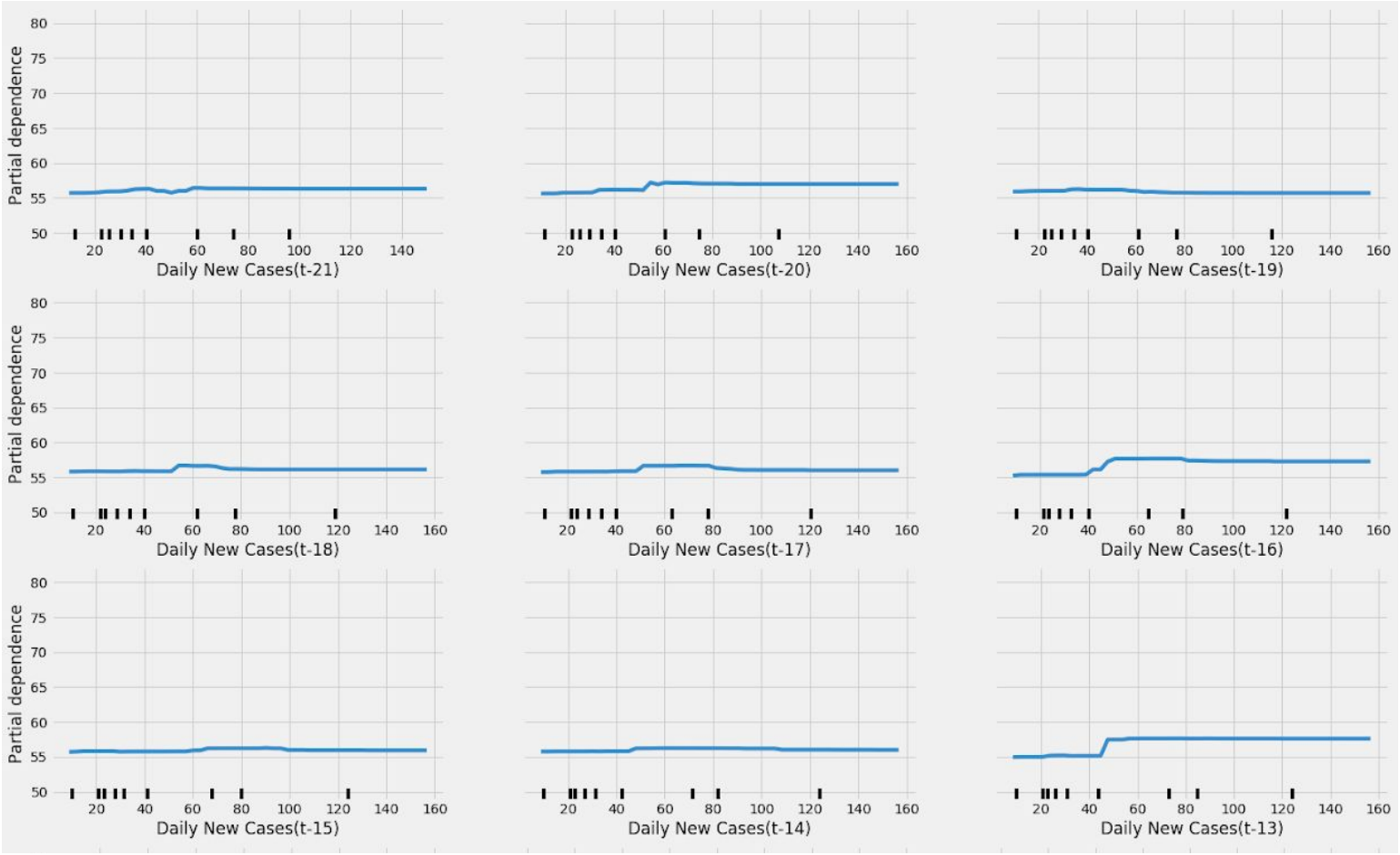
State	Population Density	Recovery Factor
Alaska	1.29	3.14
Wyoming	5.84	2.19
Montana	7.47	17.17
North Dakota	11.04	2.18
South Dakota	11.91	1.55
New Mexico	17.29	1.35
Idaho	22.10	1.75
Nebraska	25.42	1.63
Nevada	28.60	1.26
Kansas	35.60	3.02
Utah	39.94	1.11
Maine	43.63	1.18
Oregon	44.81	2.04
Colorado	56.40	2.02
Iowa	56.93	1.56
Oklahoma	57.65	1.37
Arkansas	58.40	1.18
Mississippi	63.71	1.00
Arizona	64.95	1.16
Vermont	68.14	12.08
Minnesota	71.59	1.04

West Virginia	73.97	1.02
Missouri	89.75	1.59
Alabama	96.92	1.05
Louisiana	107.52	3.93
Wisconsin	108.05	1.00
Texas	112.82	1.23
Kentucky	113.96	1.53
Washington	117.33	1.63
New Hampshire	153.16	1.54
Tennessee	167.27	1.29
South Carolina	173.32	1.03

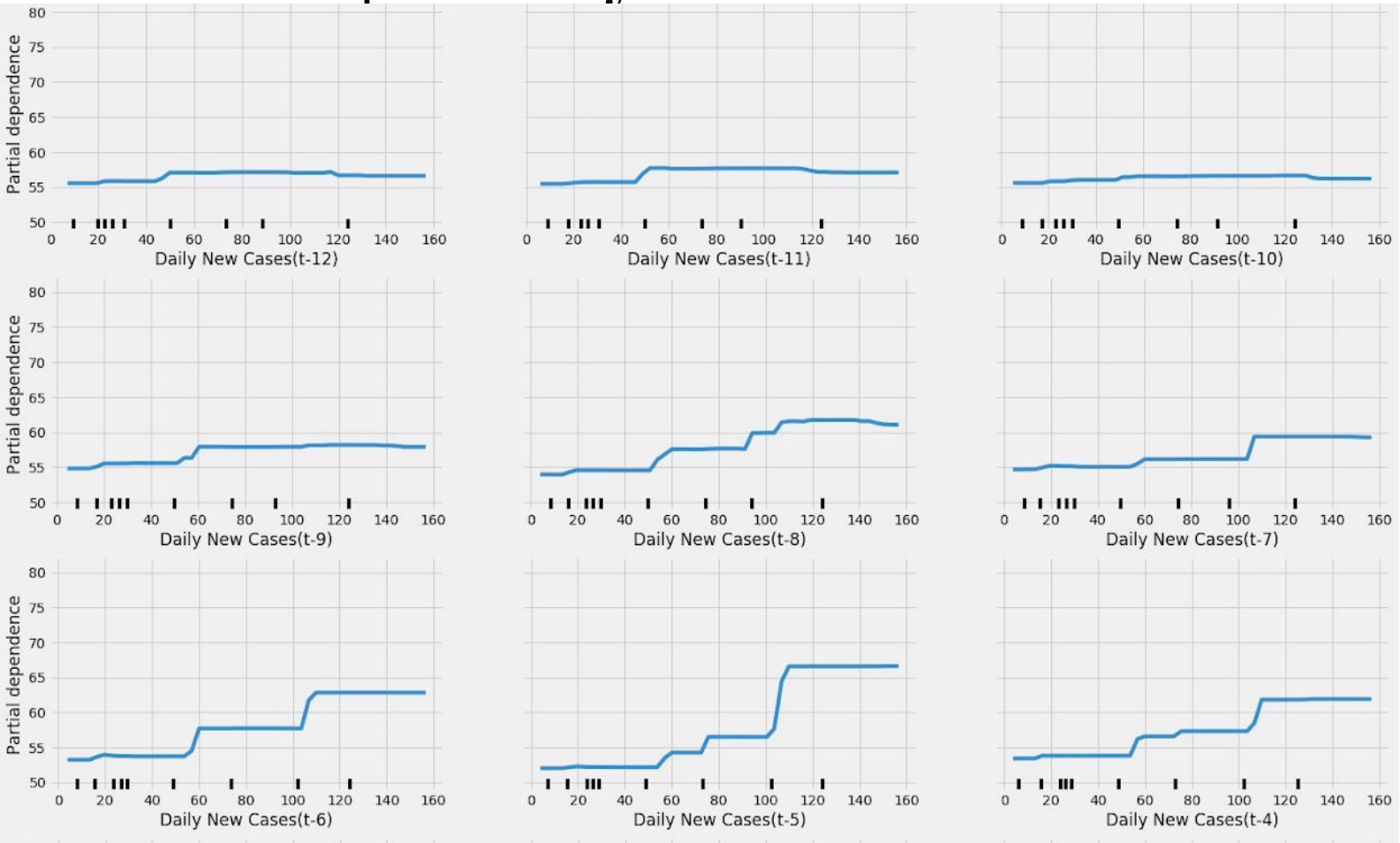
Michigan	177.67	4.03
Georgia	186.67	1.34
Indiana	188.28	1.59
North Carolina	218.27	1.04
Virginia	218.44	1.01
Hawaii	219.94	1476.42
Illinois	228.02	1.37
California	256.37	1.02
Pennsylvania	286.54	2.31
Ohio	287.50	1.72
Florida	410.13	1.67

New York	412.52	6.75
Delaware	504.31	2.09
Maryland	626.67	1.20
Connecticut	735.87	3.28
Massachusetts	894.44	3.60
Rhode Island	1021.43	2.96
New Jersey	1215.20	3.92

MN Partial Dependency Plots

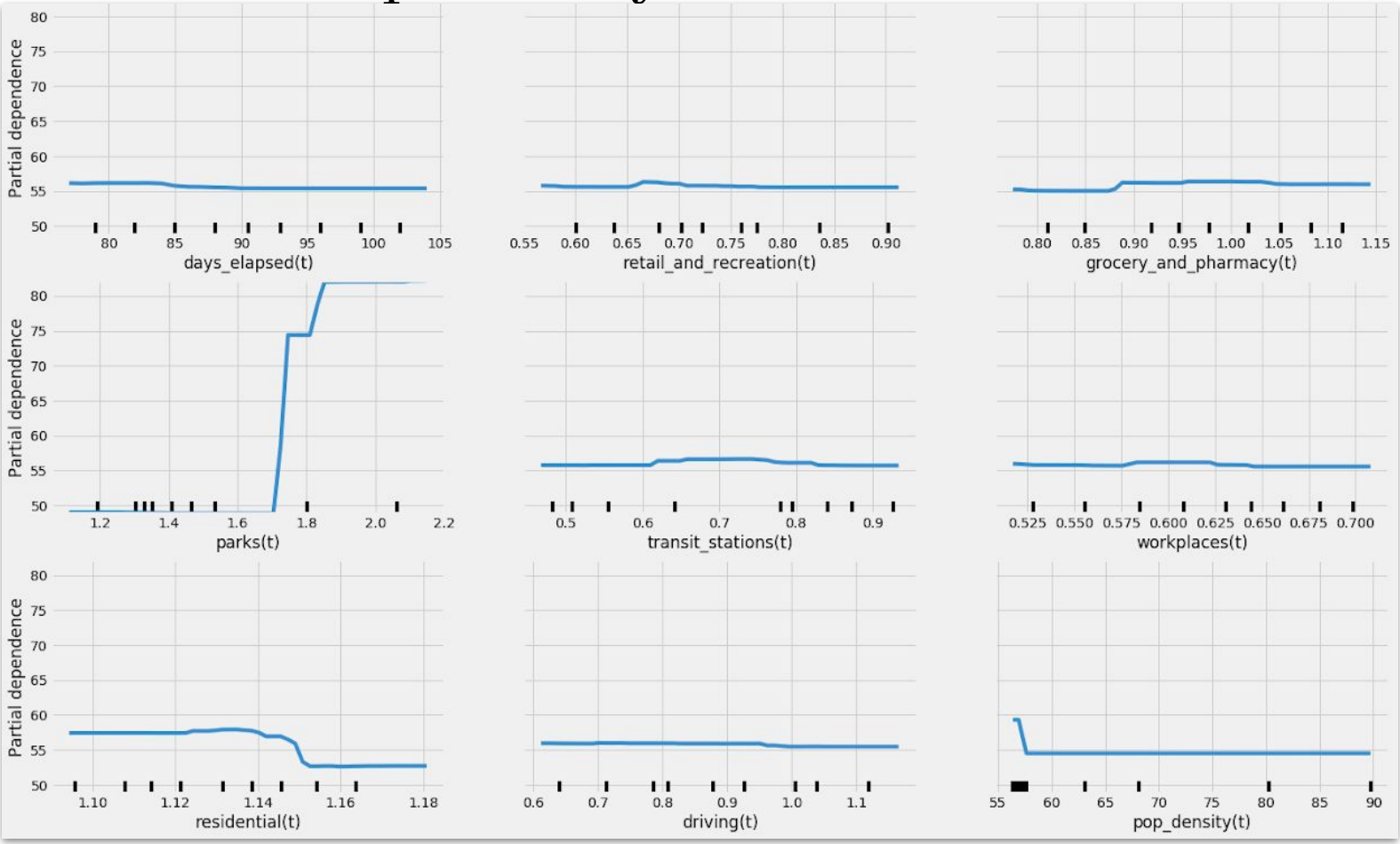


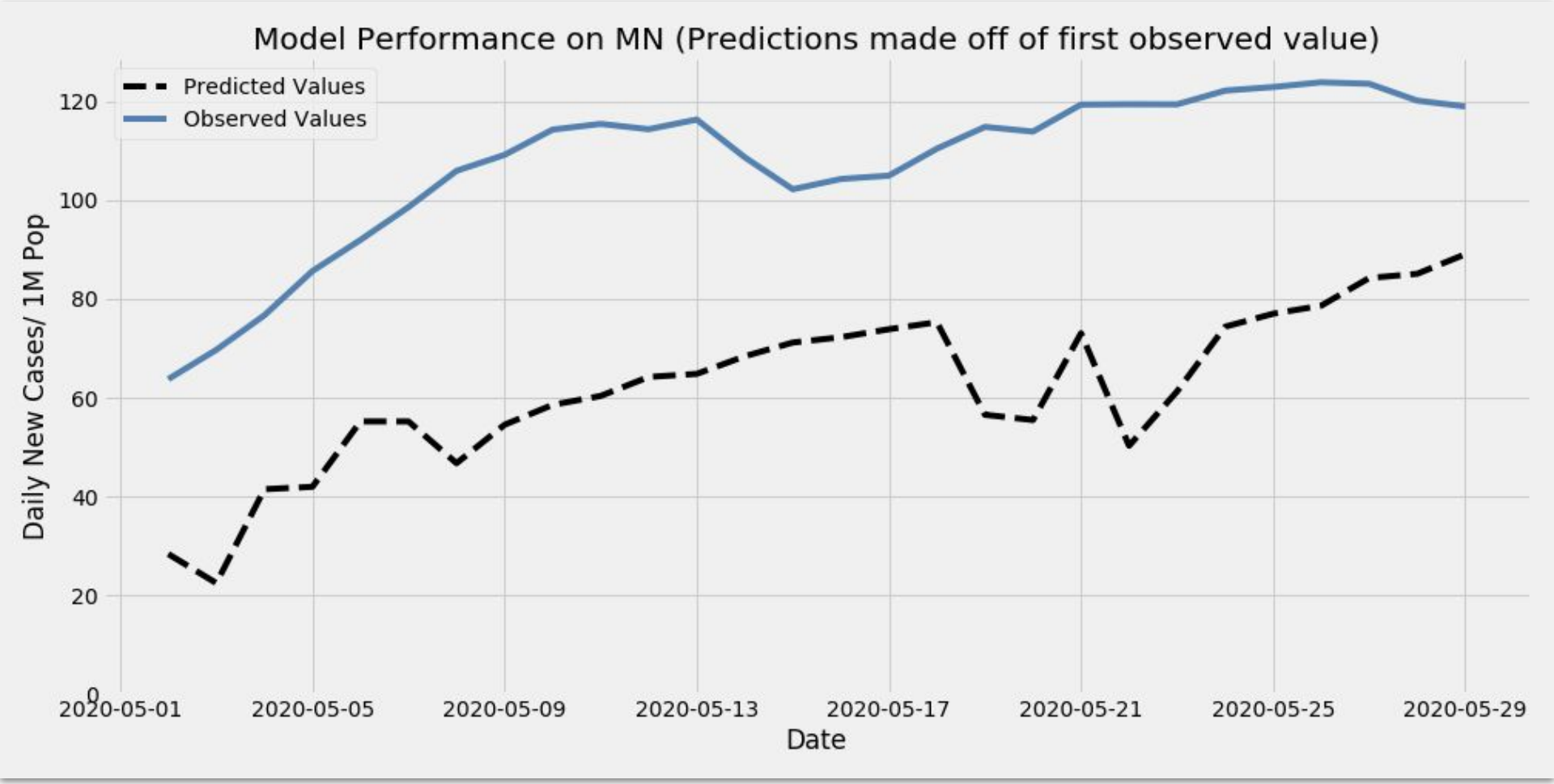
MN Partial Dependency Plots

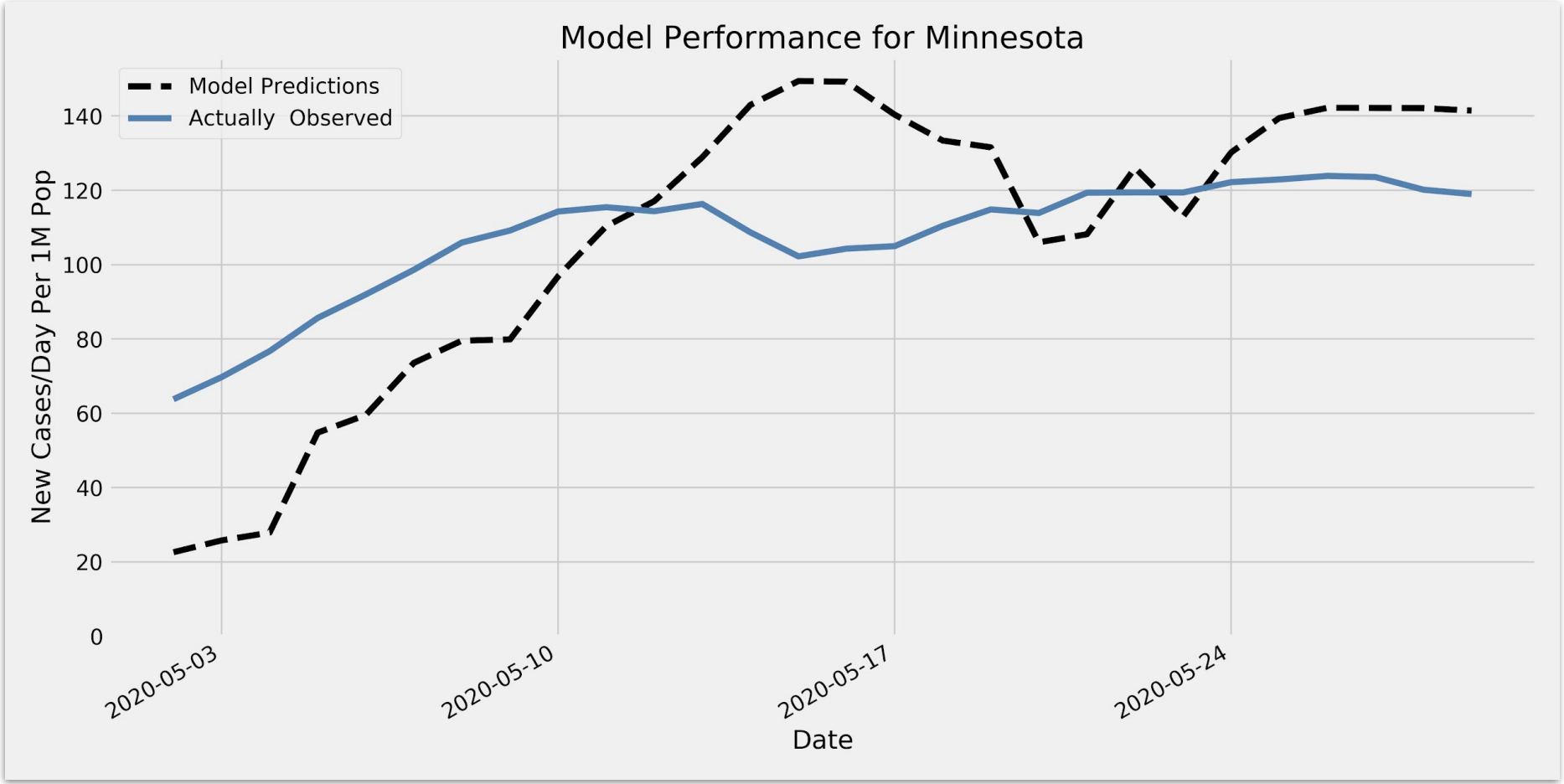


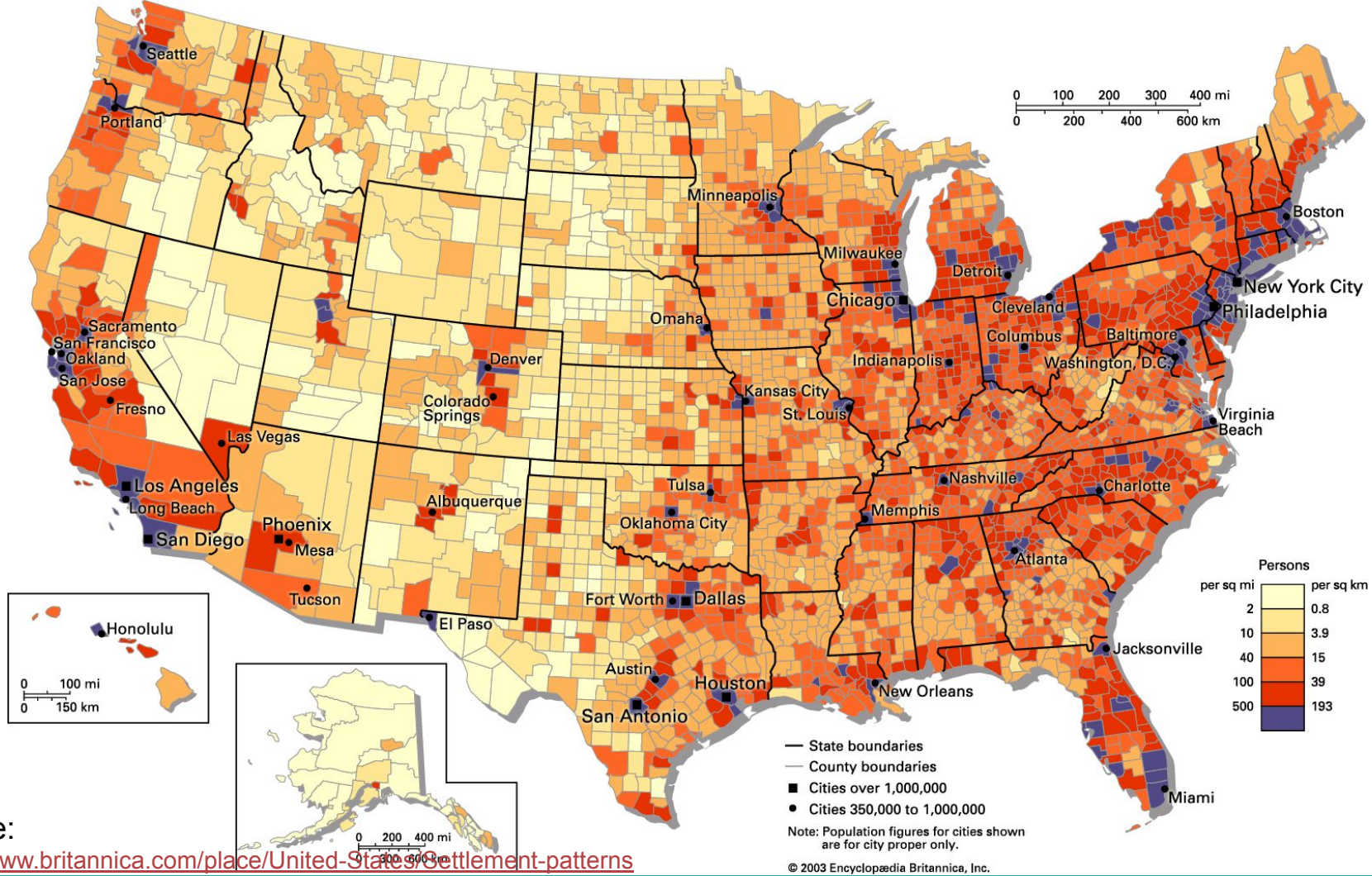
MN Partial Dependency Plots

MN Partial Dependency Plots

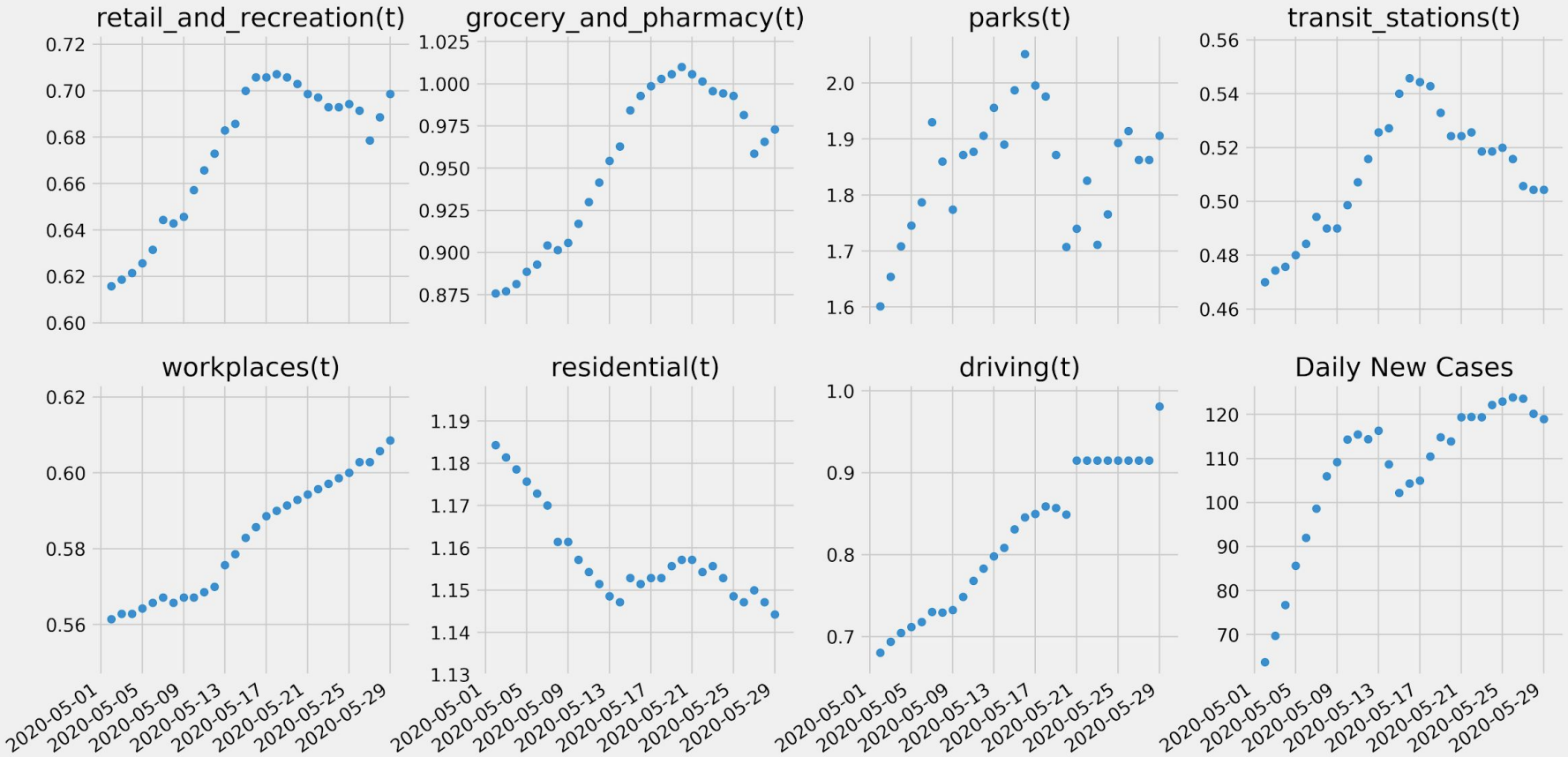


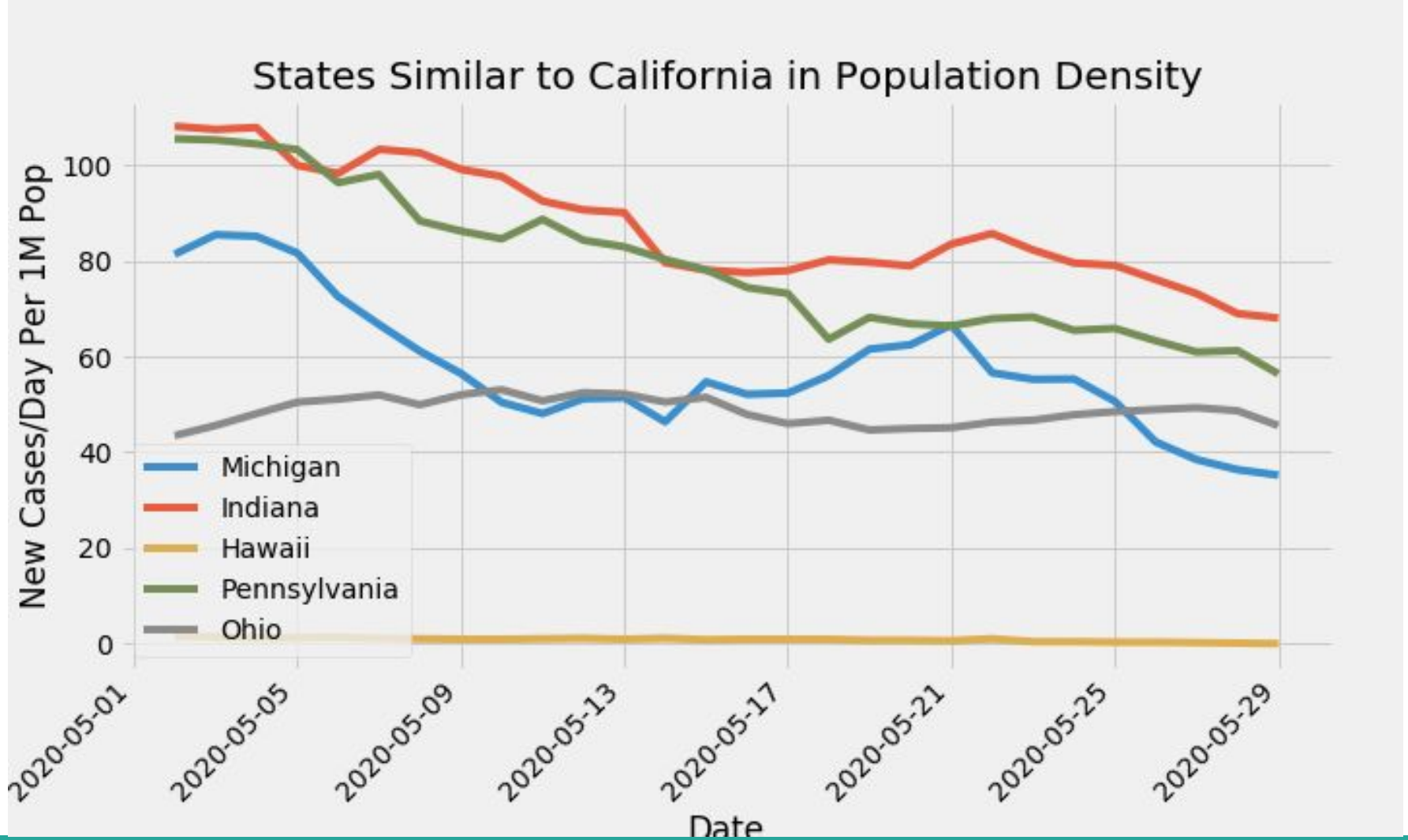

Model Performance on MN (Predictions made off of first observed value)
Predicted Values
Observed Values
Daily New Cases/ 1M Pop
0
20
40
60
80
100
120
2020-05-01
2020-05-05
2020-05-09
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
Date

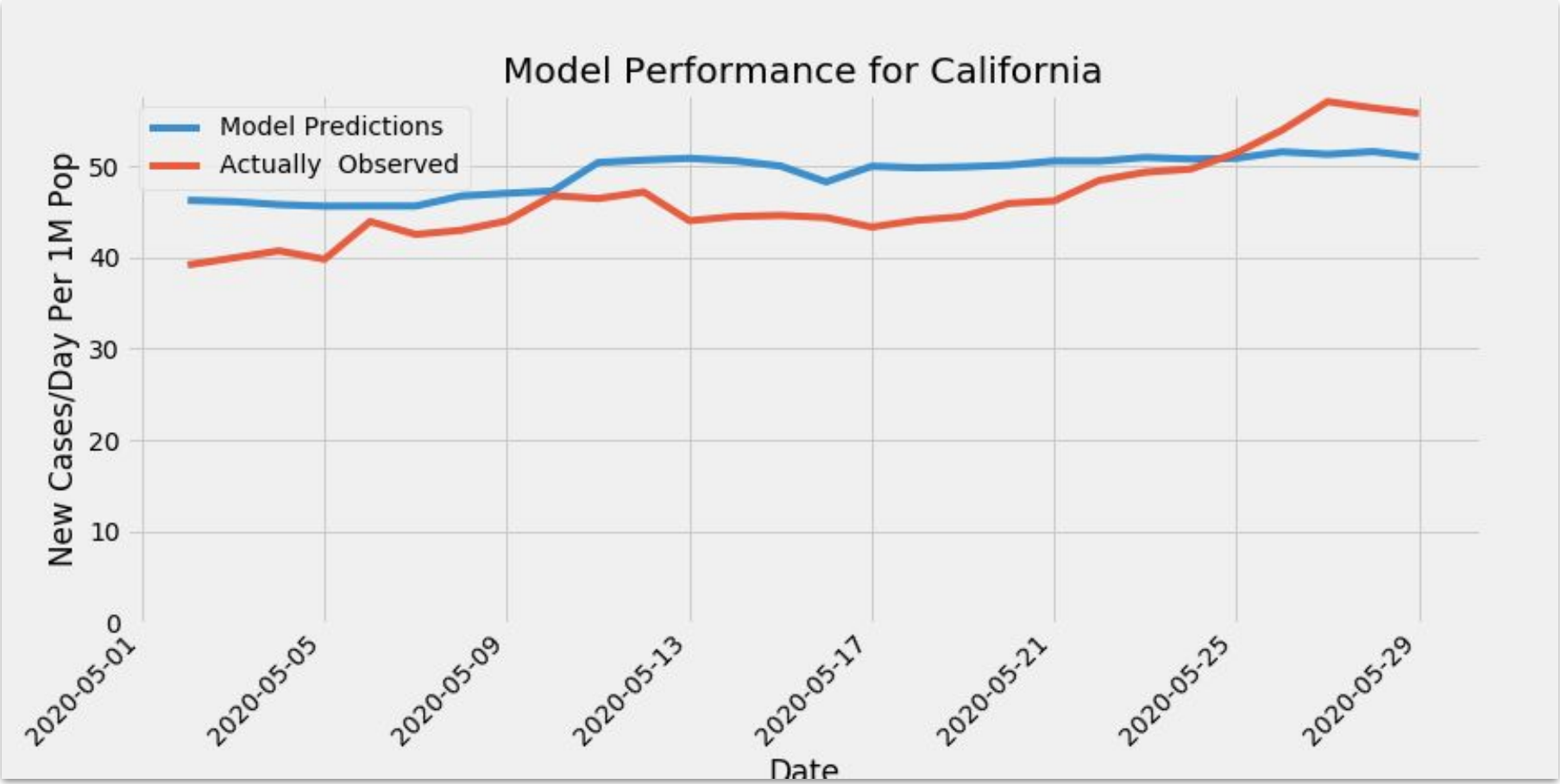

Model Performance for Minnesota
Model Predictions
Actually Observed
New Cases/Day Per 1M Pop
0
20
40
60
80
100
120
140
2020-05-03
2020-05-10
2020-05-17
2020-05-24
Date



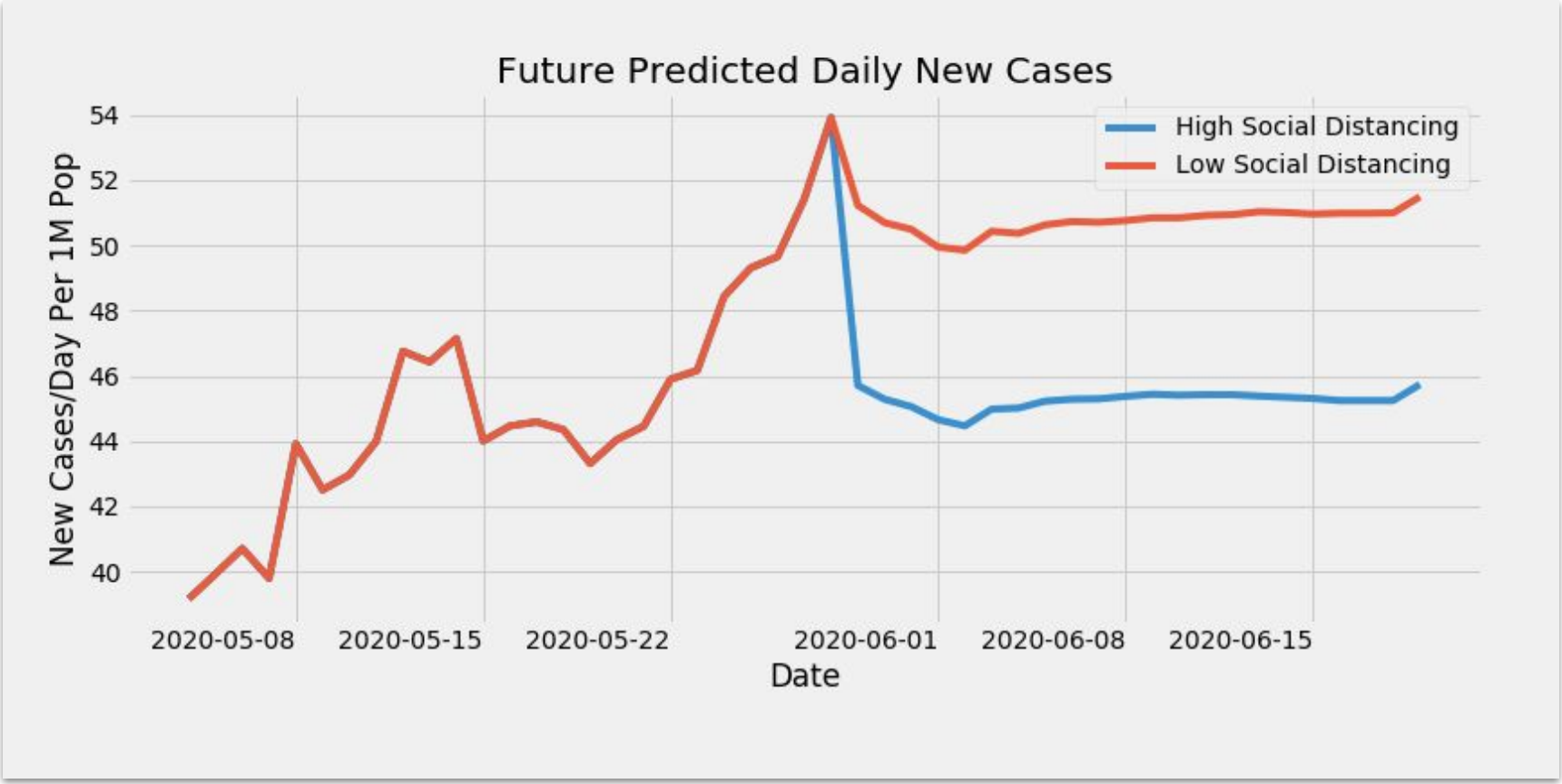
Source:
https://www.britannica.com/place/United-States/Settlement-patterns

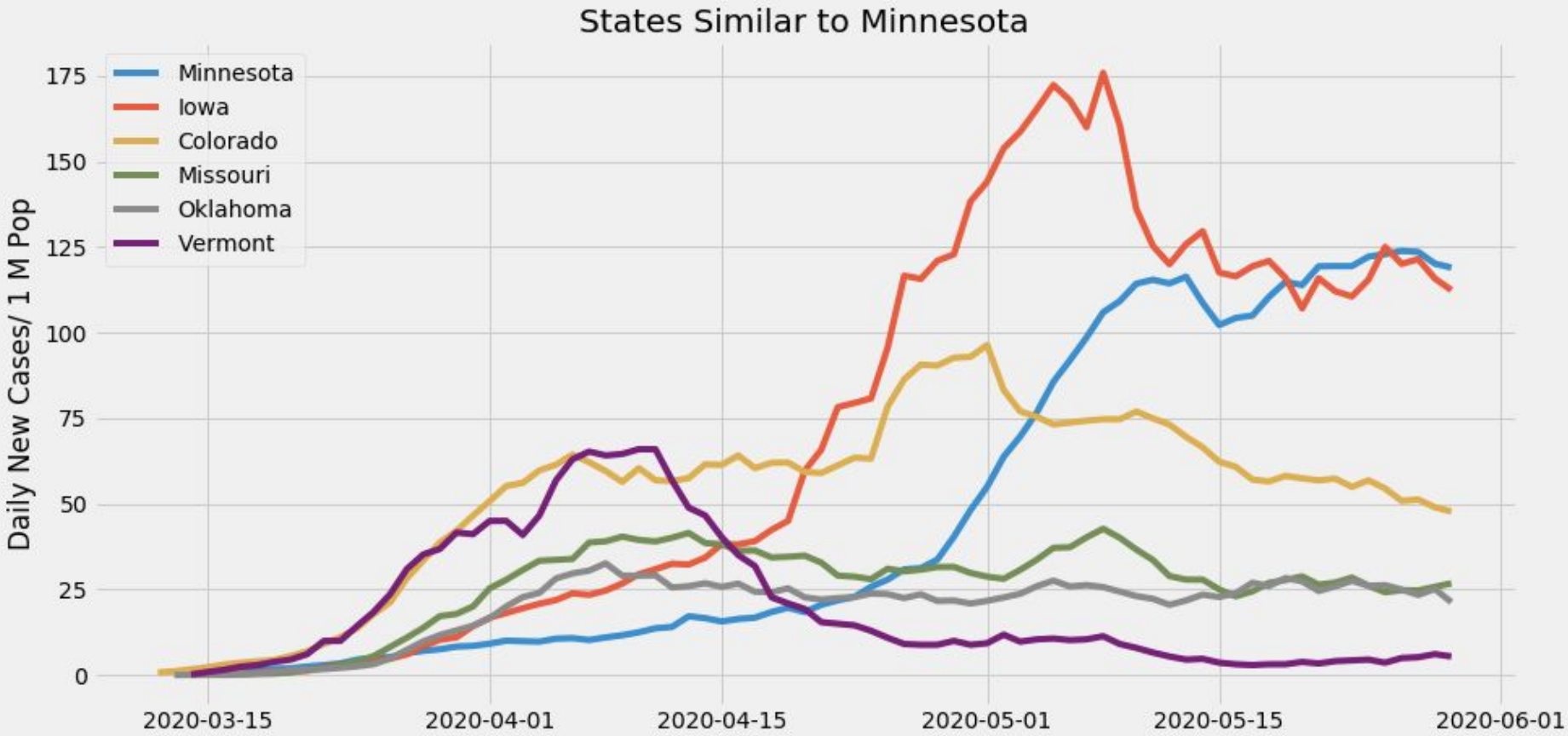

MN Social Distancing Features
retail_and_recreation(t)
grocery_and_pharmacy(t)
parks(t)
transit_stations(t)
workplaces(t)
residential(t)
driving(t)
Daily New Cases
2020-05-01
2020-05-05
2020-05-09
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29

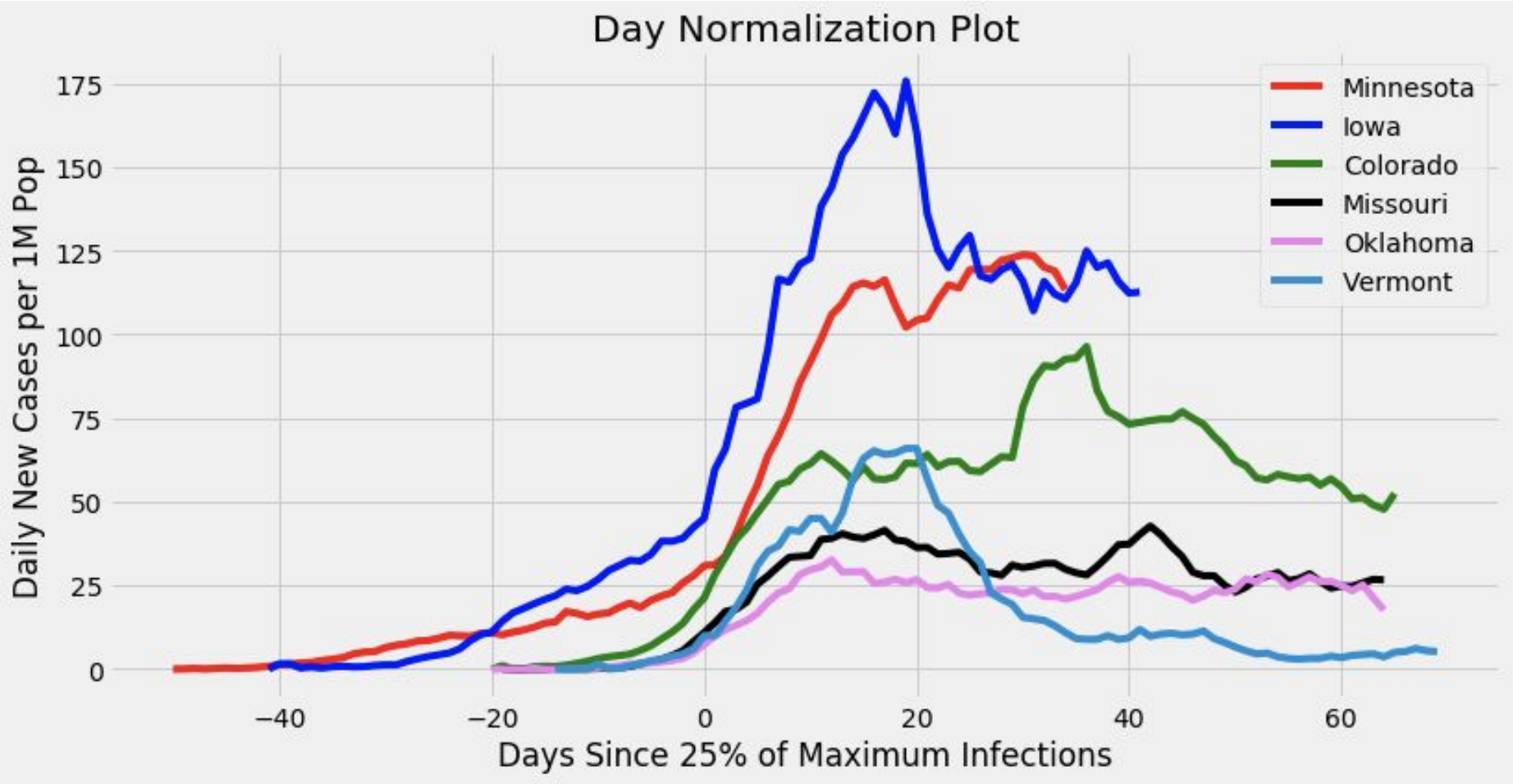

States Similar to California in Population Density
New Cases/Day Per 1M Pop
100
80
60
40
20
0
Michigan
Indiana
Hawaii
Pennsylvania
Ohio
2020-05-01
2020-05-05
2020-05-09
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
Date


Model Performance for California
Model Predictions
Actually Observed
New Cases/Day Per 1M Pop
0
10
20
30
40
50
2020-05-01
2020-05-05
2020-05-09
2020-05-13
2020-05-17
2020-05-21
2020-05-25
2020-05-29
Date

California

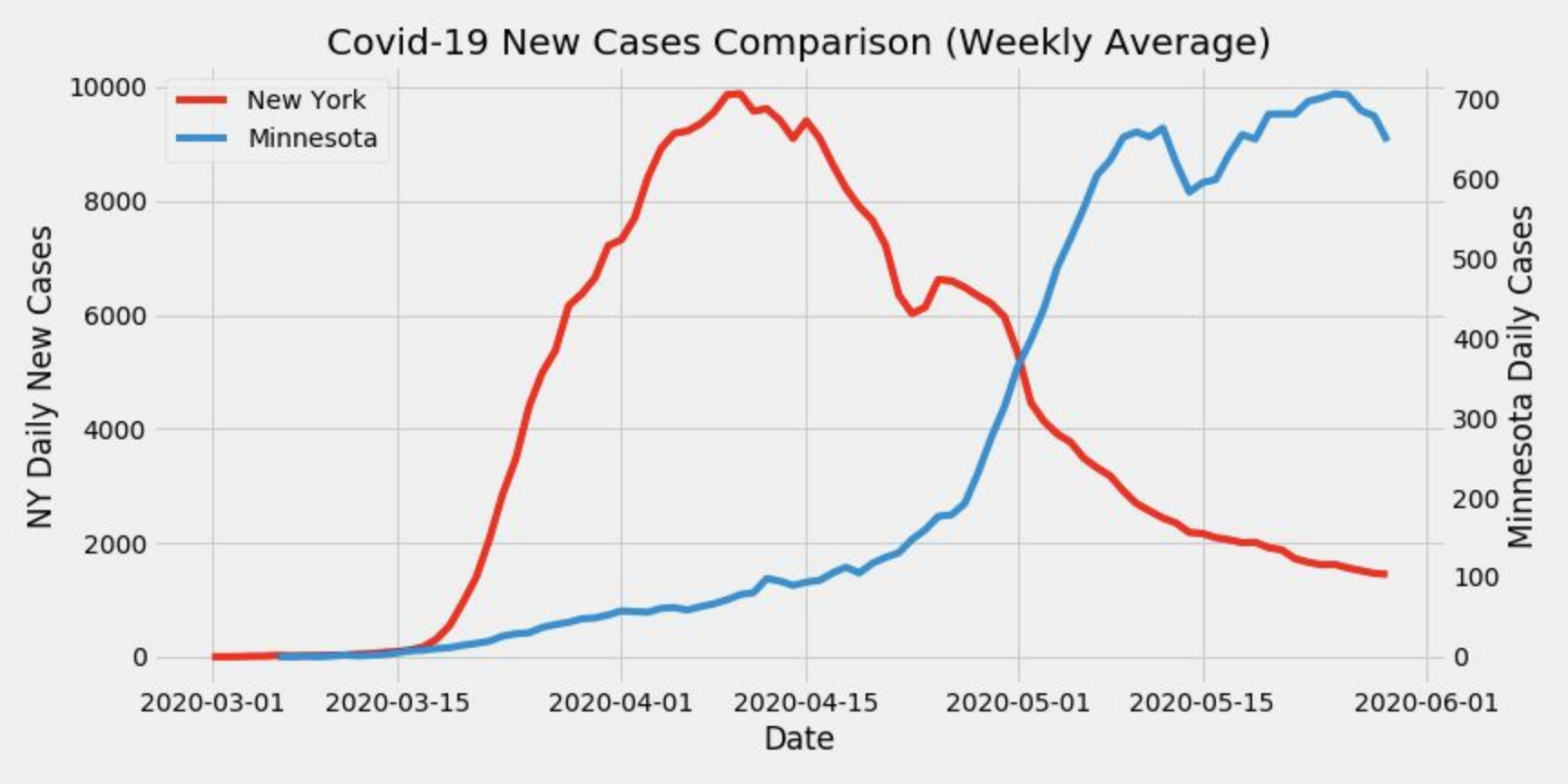

Future Predicted Daily New Cases
High Social Distancing
Low Social Distancing
New Cases/Day Per 1M Pop
54
52
50
48
46
44
42
40
2020-05-08
2020-05-15
2020-05-22
2020-06-01
2020-06-08
2020-06-15
Date


States Similar to Minnesota
Minnesota
Iowa
Colorado
Missouri
Oklahoma
Vermont
Daily New Cases/ 1 M Pop
175
150
125
100
75
50
25
0
2020-03-15
2020-04-01
2020-04-15
2020-05-01
2020-05-15
2020-06-01


Day Normalization Plot
Daily New Cases per 1M Pop
175
150
125
100
75
50
25
0
−40
−20
0
20
40
60
Days Since 25% of Maximum Infections
Minnesota
Iowa
Colorado
Missouri
Oklahoma
Vermont

Minnesota High/Low Social Distancing (% of January 2020 levels)

	Retail/Recreation %	Grocery/Pharmacy %	Parks %	Transit Stations %	Workplaces %	Residential %	Driving %
High	52.7	74.0	94.1	43.3	50.1	119.4	55.8
Low	93.9	117.0	229.9	96.1	72.7	108.6	124.7


Covid-19 New Cases Comparison (Weekly Average)
New York
Minnesota
NY Daily New Cases
Minnesota Daily Cases
Date
10000
8000
6000
4000
2000
0
700
600
500
400
300
200
100
0
2020-03-01
2020-03-15
2020-04-01
2020-04-15
2020-05-01
2020-05-15
2020-06-01