San Diego COVID-19 Risk Zones Analysis using Time-series Data

Research Team:

Dr. Ming-Hsiang Tsou, mtsou@sdsu.edu, Director of the Center for Human Dynamics in the Mobile Age, Professor, Department of Geography , San Diego State University
Dr. Chii-Dean (Joey) Lin, Associate Chair, Department of Mathematics & Statistics.
Dr. Bo-Wen Shen, Associate Professor, Department of Mathematics & Statistics.

Morgan Daniels (SDSU Alumni)
Jian Xu (SDSU Doctoral Student).

How can we categorize them into **different groups** based on their graph patterns and the "**similarity**".
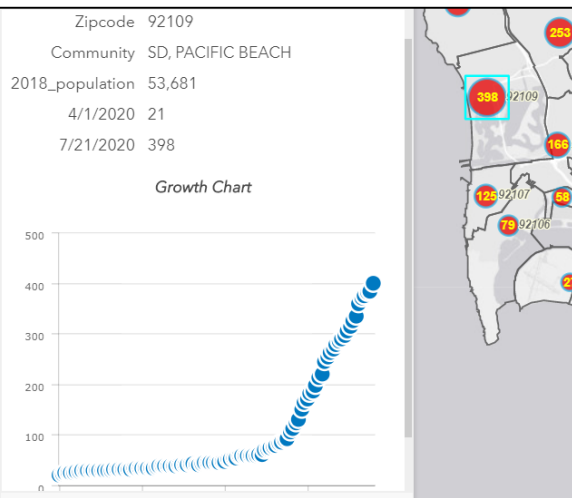
Pacific Beach
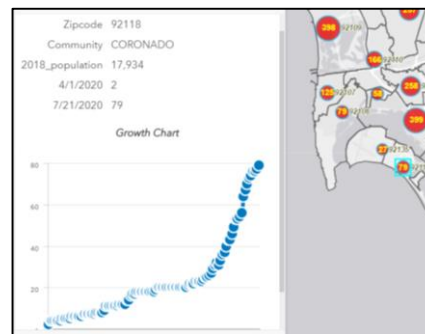
Coronado
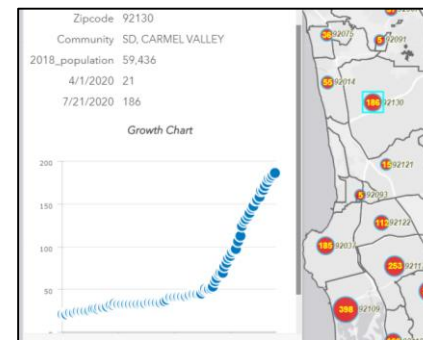
Carmel Valley
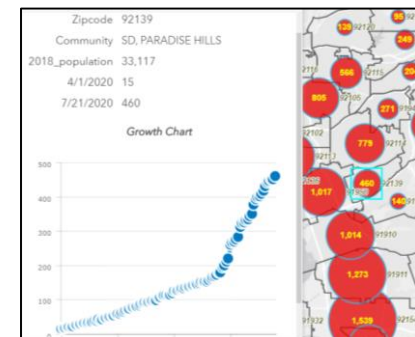
Paradise Hills
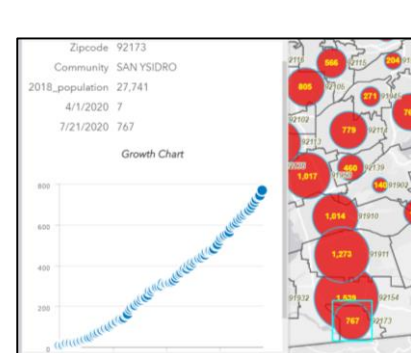
San Ysidro
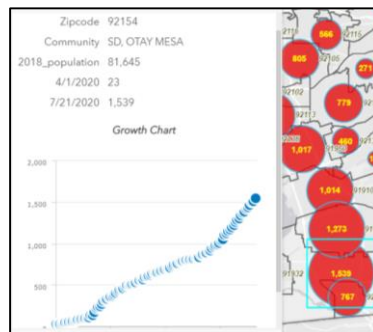
Otay Mesa
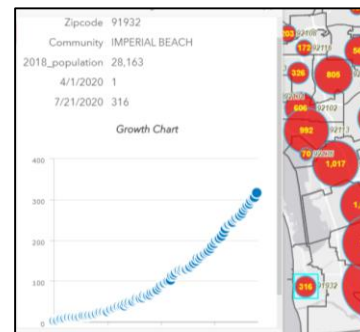
Imperial Beach
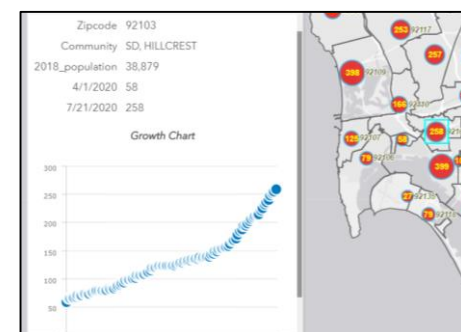
Hillcrest
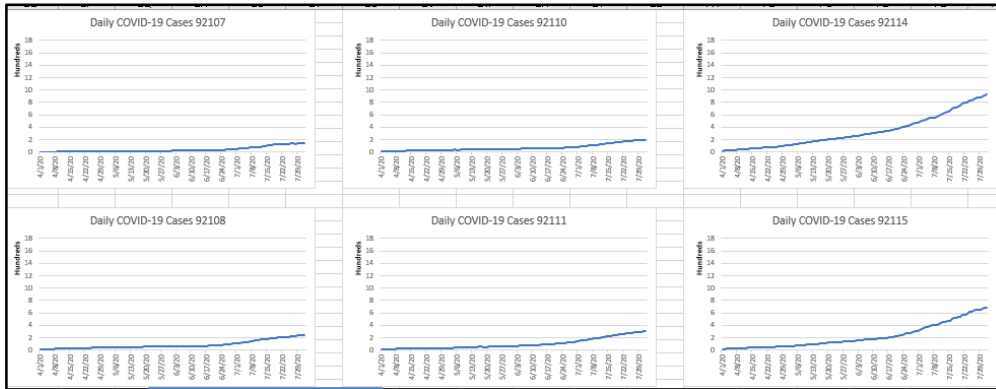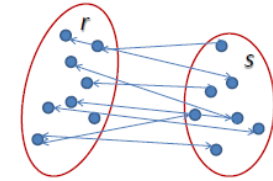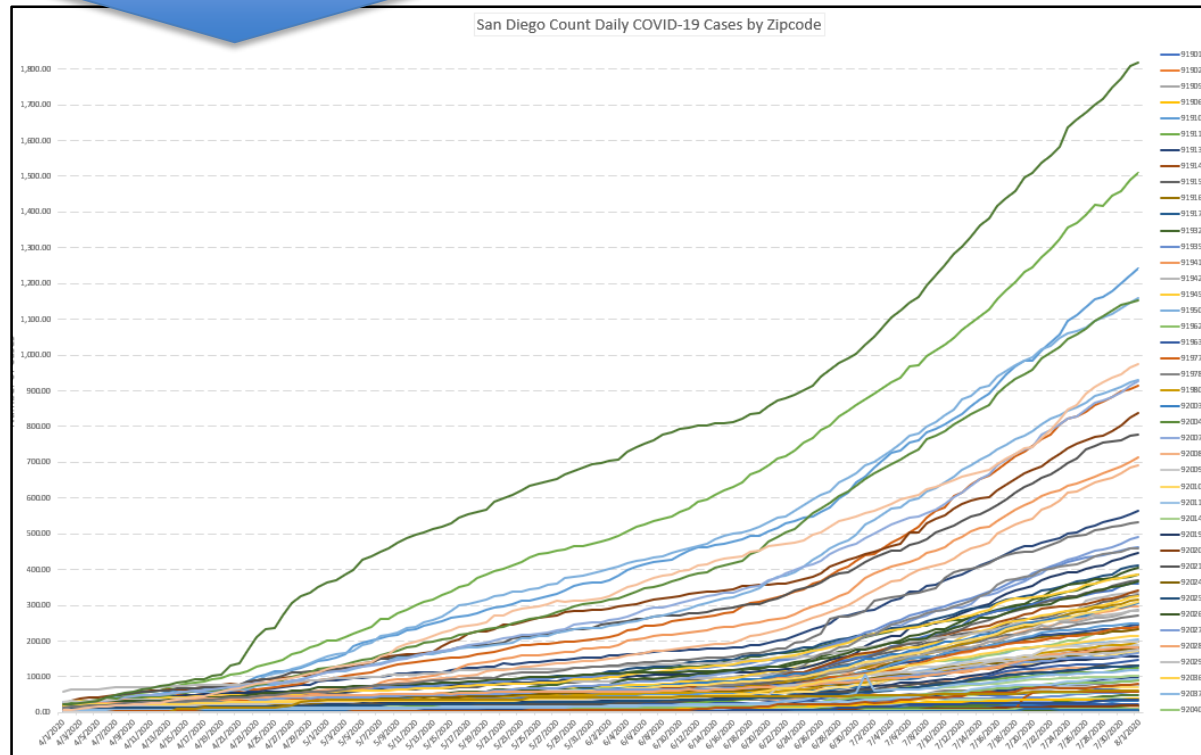
$$L(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

Average Linkage Clustering

Results:

We testing the cluster number = 8, 12, and 20.

The number = 8 is the best choice for this cluster analysis.

Based on their mean value in each cluster. We can label these cluster zones as:

LOW RISK (few COVID19 cases)

MEDIUM RISK

HIGH RISK (very large numbers of COVID19 cases)

Type-A, Type-B, Type C show different growing patterns within each risk level

8 clusters

The MEANS Procedure

| CLUSTER	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum	LABEL
1	37	_4_1_2020	37	2.892	3.446	0.000	13.000	Low Risk
		_5_1_2020	37	7.919	7.776	0.000	35.000	TypeA
		_6_1_2020	37	11.865	10.398	0.000	41.000	
		_7_1_2020	37	27.946	23.828	1.000	106.000	Areas
		_8_1_2020	37	60.811	44.398	5.000	146.000	
2	33	_4_1_2020	33	11.091	7.125	1.000	29.000	Low Risk –
		_5_1_2020	33	30.121	11.205	12.000	61.000	Type B
		_6_1_2020	33	52.636	14.521	30.000	88.000	
		_7_1_2020	33	113.879	28.728	68.000	167.000	Areas
		_8_1_2020	33	243.818	66.068	158.000	371.000	
3	10	_4_1_2020	10	19.100	16.387	1.000	58.000	Medium Risk
		_5_1_2020	10	52.900	20.658	27.000	93.000	Type A
		_6_1_2020	10	94.500	21.104	53.000	128.000	
		_7_1_2020	10	202.100	17.597	177.000	223.000	
		_8_1_2020	10	409.900	59.955	283.000	492.000	
4	4	_4_1_2020	4	15.750	2.500	13.000	19.000	Medium Risk
		_5_1_2020	4	68.250	14.773	59.000	90.000	Type B
		_6_1_2020	4	152.750	22.111	129.000	182.000	
		_7_1_2020	4	318.250	43.965	283.000	376.000	
		_8_1_2020	4	625.250	90.090	532.000	713.000	
5	6	_4_1_2020	6	17.833	6.676	7.000	25.000	Medium Risk
		_5_1_2020	6	112.667	10.172	94.000	123.000	Type C
		_6_1_2020	6	258.500	39.088	208.000	318.000	
		_7_1_2020	6	479.500	53.773	422.000	557.000	
		_8_1_2020	6	894.167	72.112	778.000	974.000	
6	3	_4_1_2020	3	17.667	5.774	11.000	21.000	High Risk
		_5_1_2020	3	143.000	17.059	124.000	157.000	Type A
		_6_1_2020	3	357.333	43.016	311.000	396.000	
		_7_1_2020	3	668.667	19.655	654.000	691.000	
		_8_1_2020	3	1185.333	50.003	1154.000	1243.000	
7	1	_4_1_2020	1	16.000	.	16.000	16.000	High Risk
		_5_1_2020	1	202.000	.	202.000	202.000	Type B
		_6_1_2020	1	478.000	.	478.000	478.000	
		_7_1_2020	1	875.000	.	875.000	875.000	
		_8_1_2020	1	1511.000	.	1511.000	1511.000	
8	1	_4_1_2020	1	23.000	.	23.000	23.000	High Risk
		_5_1_2020	1	363.000	.	363.000	363.000	Type C
		_6_1_2020	1	697.000	.	697.000	697.000	

Most zip codes are Low risk areas (37 + 33) with green color.

Medium risk areas (10 + 4 + 6) are yellow and light brown areas.

High risk areas (3 + 1 + 1) are red zones.
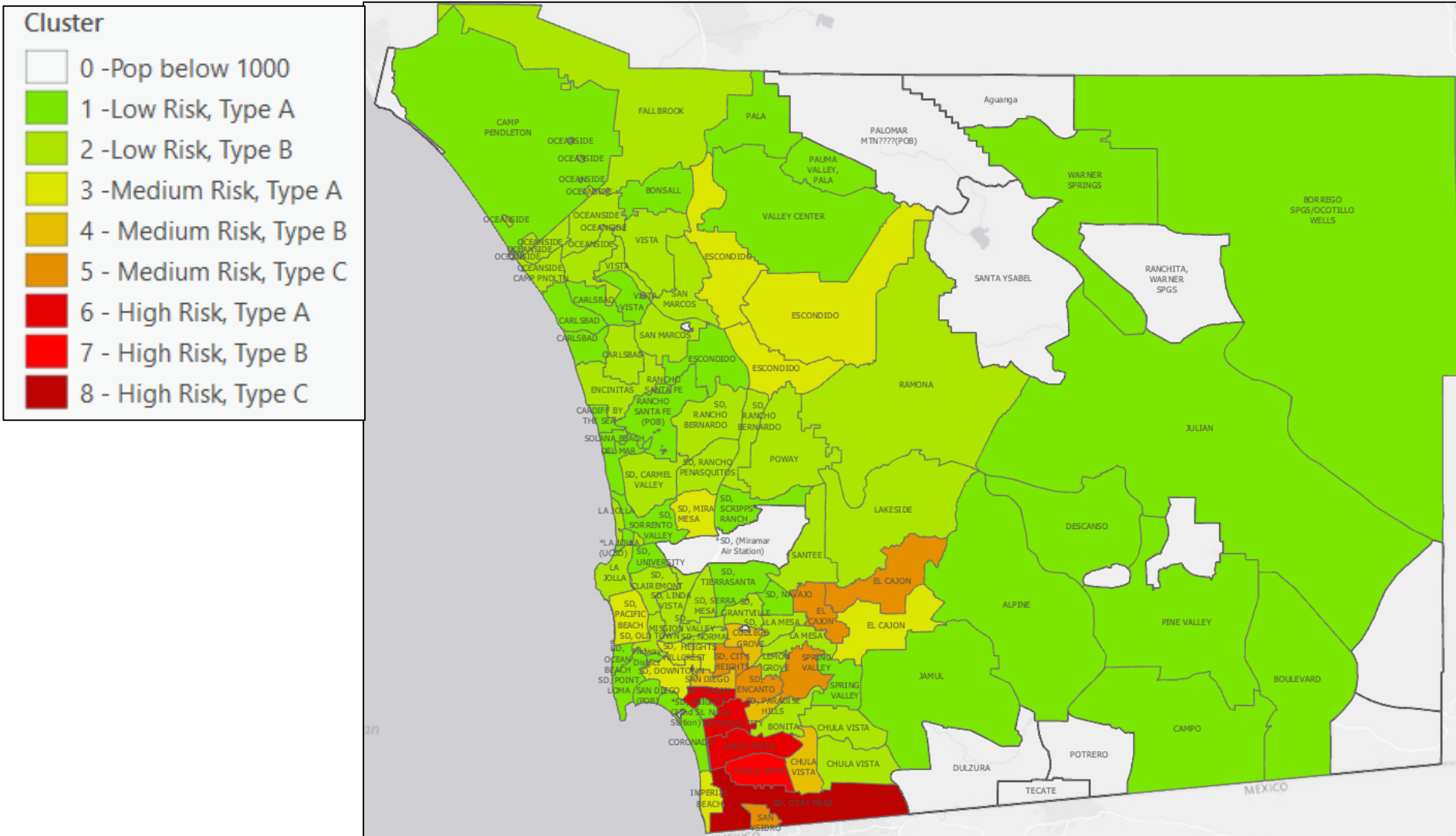
Medium risk areas (10 + 4 + 6) are yellow and light brown areas.
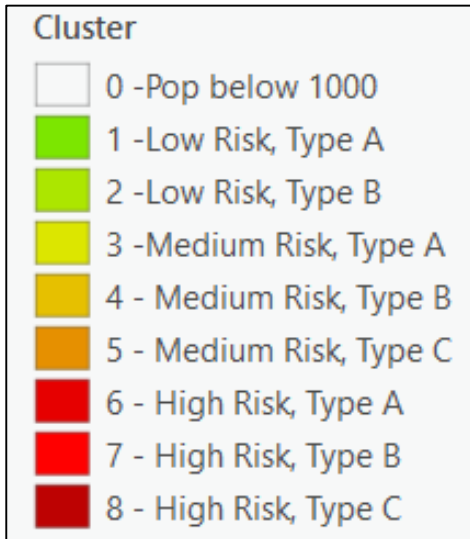
High risk areas (3 + 1 + 1) are Logan Heights, National City, Chula Vista (west), Chula Vista (south), and Otay Mesa.

Second Cluster Method: Use the Change Ratio Value
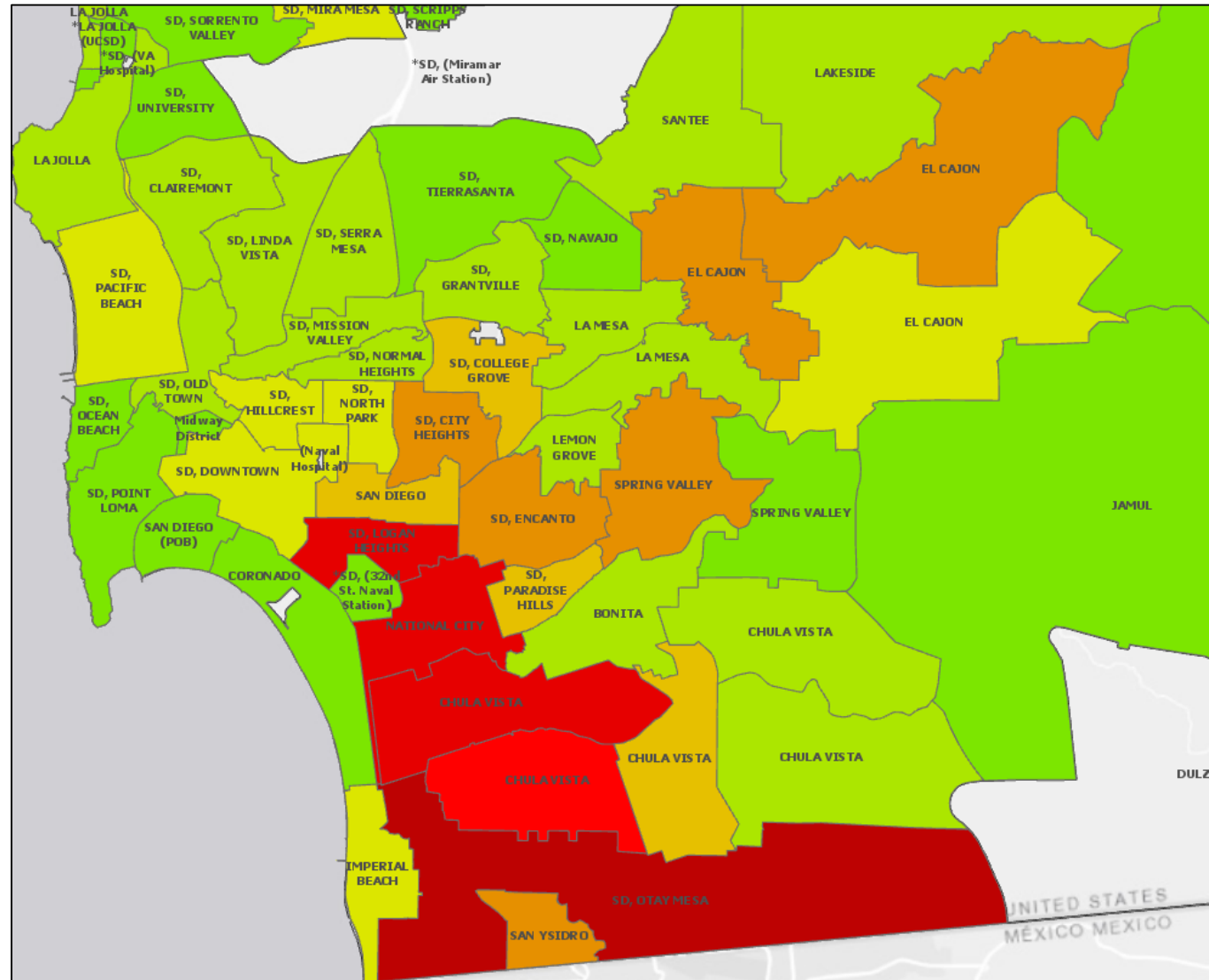
7/24/2020	7/25/2020	7/26/2020	7/27/2020	7/28/2020	7/29/2020	7/30/2020	7/31/2020	8/1/2020	Change Ratio (Time Series)	ZipCode
0	0	0	0.000123	0	0	6.13E-05	6.13E-05	0		92010
0.000126	0.000167	0	4.19E-05	0.000126	0.000167	0.000126	4.19E-05	0		92011
0	0	0	0.000145	0	0	0	0	0		92014
0.000113	2.26E-05	0.000226	0.000181	2.26E-05	0.000293	0.000135	0.000248	0.000158		92019
0.00021	0.000226	0.000178	0.000145	6.47E-05	0.000194	0.000242	0.000356	0.00021		92020
0.000265	0.000177	0.000368	0.000162	0.000132	2.94E-05	5.89E-05	0.000206	2.94E-05		92021
3.54E-05	3.54E-05	1.77E-05	1.77E-05	0	7.08E-05	0	5.31E-05	5.31E-05		92024
0.000116	0.000214	7.77E-05	0.000136	9.71E-05	7.77E-05	0.000291	0.000136	7.77E-05		92025
0.000179	7.16E-05	5.37E-05	0.000125	3.58E-05	5.37E-05	0.000233	0.000161	8.95E-05		92026
0.000187	0.000131	0.000187	0.000131	9.33E-05	0.000112	0.000149	0.00028	0.000112		92027
0.000166	2.07E-05	0.000166	0.000207	0.000104	8.29E-05	0.000145	8.29E-05	4.14E-05		92028
0	0	4.89E-05	4.89E-05	4.89E-05	0.000147	4.89E-05	0	0		92029
0	0	0	0.00023	0	0	0	0	0		92036
4.7E-05	4.7E-05	4.7E-05	2.35E-05	0	7.04E-05	4.7E-05	7.04E-05	4.7E-05		92037
0.000182	2.27E-05	0.000205	0.000205	0.000136	0.000136	9.09E-05	0.000182	4.55E-05		92040
2.38E-05	4.77E-05	9.54E-05	9.54E-05	7.15E-05	2.38E-05	4.77E-05	0.000167	7.15E-05		92054
0	0	2.56E-05	0	0	2.56E-05	2.56E-05	0	0		92055
9.19E-05	7.35E-05	9.19E-05	3.68E-05	5.52E-05	5.52E-05	0.00011	9.19E-05	5.52E-05		92056
0.000107	0.000124	5.33E-05	0.00016	3.56E-05	0.000107	5.33E-05	0.000178	1.78E-05		92057
0.000285	0	4.08E-05	0.000122	4.08E-05	4.08E-05	0.000245	0.000163	8.16E-05		92058

The cluster analysis for change ratio could detect whether zip codes have similar patterns of change in growth rate.

20 _clusters_ _complete_ linkage

CLUSTER	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum	Wave Pattern (Time Series)
1	52	_4_2_2020	52	4.36204442	5.93056289	-10.10610000	21.65440000	Cluster 1: Consistent and mild rate change pattern
		_5_1_2020	52	0.86470192	2.59806479	-8.96060000	11.15200000	
		_6_1_2020	52	2.92330385	6.53557284	0.00000000	43.30880000	
		_7_1_2020	52	8.48551538	6.49072694	0.00000000	21.27850000	
		_8_1_2020	52	3.42419038	3.84257956	0.00000000	13.36630000	
2	28	_4_2_2020	28	4.16146964	4.05089349	0.00000000	14.72210000	Cluster 2: Increasing rate changes in late stage (07/01)
		_5_1_2020	28	5.16941071	4.24751331	0.00000000	14.18000000	
		_6_1_2020	28	3.76519286	3.93041877	0.00000000	12.32290000	
		_7_1_2020	28	21.74252500	11.37151188	3.58080000	51.22120000	
		_8_1_2020	28	12.92041429	8.23620292	0.00000000	27.93880000	
3	4	_4_2_2020	4	7.15900000	3.67308607	2.44970000	10.23790000	Cluster 3: Relative large rate change overall and a dramatic change in late June
		_5_1_2020	4	9.37985000	6.96595694	3.01960000	17.06310000	
		_6_1_2020	4	6.26630000	5.15542892	2.44970000	13.65040000	
		_7_1_2020	4	33.22082500	8.65090064	21.13720000	41.61880000	
		_8_1_2020	4	14.10315000	5.11887404	9.05880000	20.47570000	
4	3	_4_2_2020	3	0.00000000	0.00000000	0.00000000	0.00000000	Cluster 4: Very low rate change but with a dramatic rate change in late June
		_5_1_2020	3	0.00000000	0.00000000	0.00000000	0.00000000	
		_6_1_2020	3	2.54356667	4.40558670	0.00000000	7.63070000	
		_7_1_2020	3	30.18223333	30.01344836	0.00000000	60.02400000	
		_8_1_2020	3	6.40336667	5.88645788	0.00000000	11.57940000	
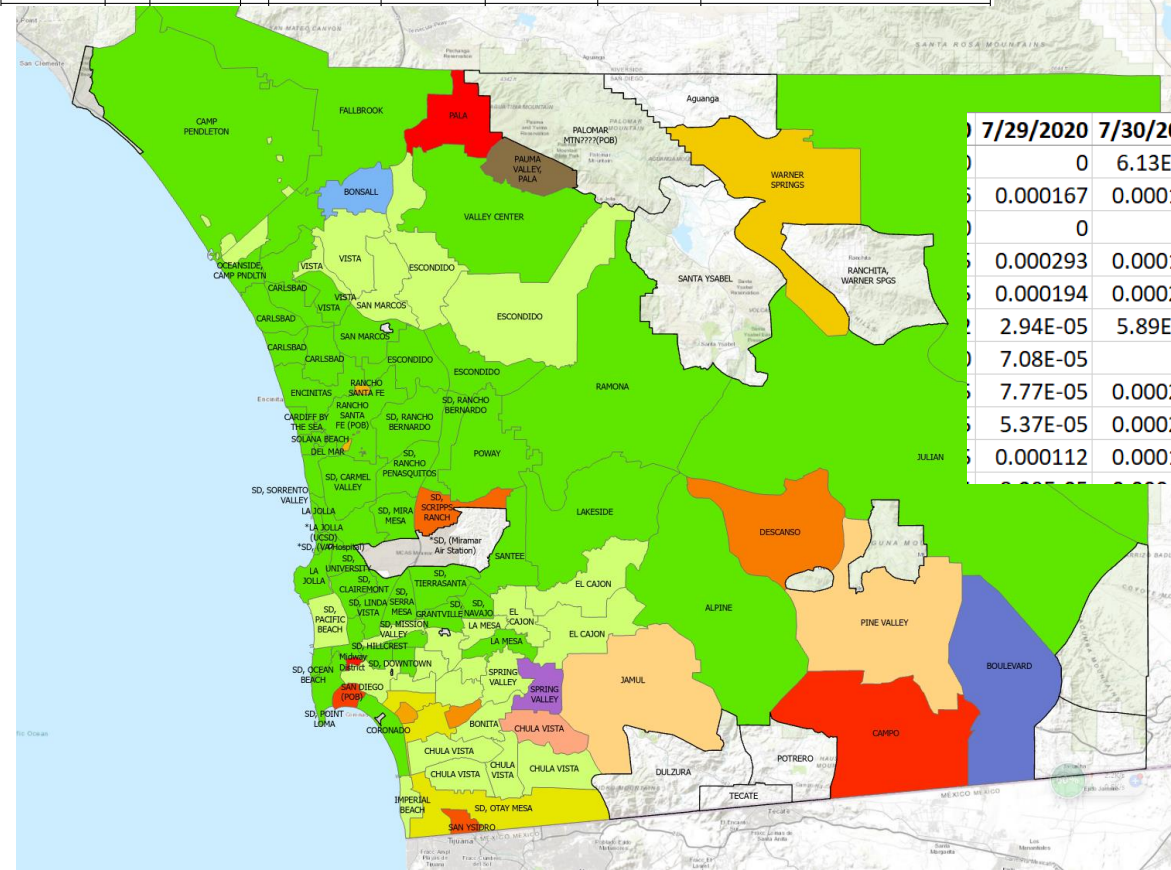
7/29/2020	7/30/2020	7/31/2020	8/1/2020	Change Ratio (Time Series)	ZipCode
0	6.13E-05	6.13E-05	0		92010
0.000167	0.000126	4.19E-05	0		92011
0	0	0	0		92014
0.000293	0.000135	0.000248	0.000158		92019
0.000194	0.000242	0.000356	0.00021		92020
2.94E-05	5.89E-05	0.000206	2.94E-05		92021
7.08E-05	0	5.31E-05	5.31E-05		92024
7.77E-05	0.000291	0.000136	7.77E-05		92025
5.37E-05	0.000233	0.000161	8.95E-05		92026
0.000112	0.000149	0.00028	0.000112		92027

GIS Mapping Results (ZOOM IN).
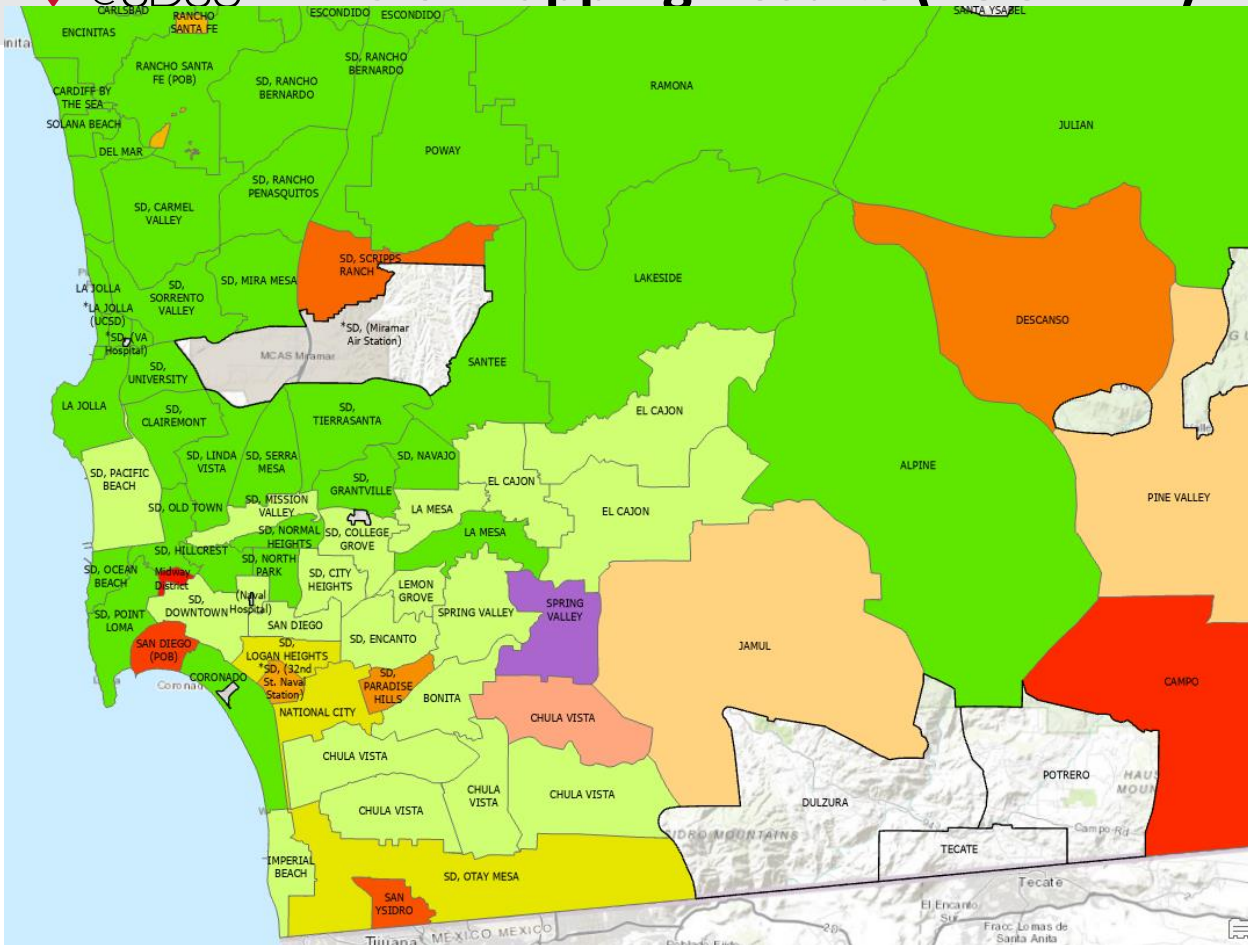
Possible explanation for each cluster:

Cluster 1 (Change pattern 1): Consistent and mild rate change pattern

Cluster 2 (Change pattern 2): Increasing rate changes in late stage (07/01)

Cluster 3 (Change pattern 3): Large relative rate change overall and a dramatic change in late June

Cluster 4 (Change pattern 4): Very low rate change but with a dramatic rate change in late June

After using the **Piecewise** Regression Analysis, all but 2 zip codes have **TWO Growth Rate Stages (with significant statistic analysis).** Zip codes without two stages of growth rate are 91962 (Pine Valley) and 92086 (Warner Springs).

Ho:	There are not 2 stages	0	
Ha:	There are two significant stages .	1	alpha= 0.05

Summarize the zips into two groups: one with two stages and another one only has one stage.

Model:
$Z = 0$ if $x < 90$
$Z = 1$ if $x \geq 90$
$E(Y|z = 1) = B0 - 90B2 + X(B1 + B2)$
$E(Y|z = 0) = B0 + B1X$

	2 Stages	Inconclusive
	0.97894737	0.02105263

Ratio of Number of Stages of COVID-19 Cases in each San Diego Zipcode through August 1, 2020

Community	CORONADO
2018_population	17,934
4/1/2020	2
7/21/2020	79

Growth Chart
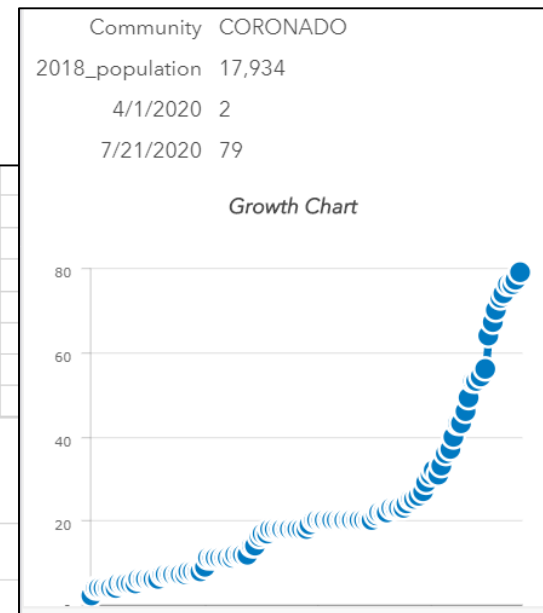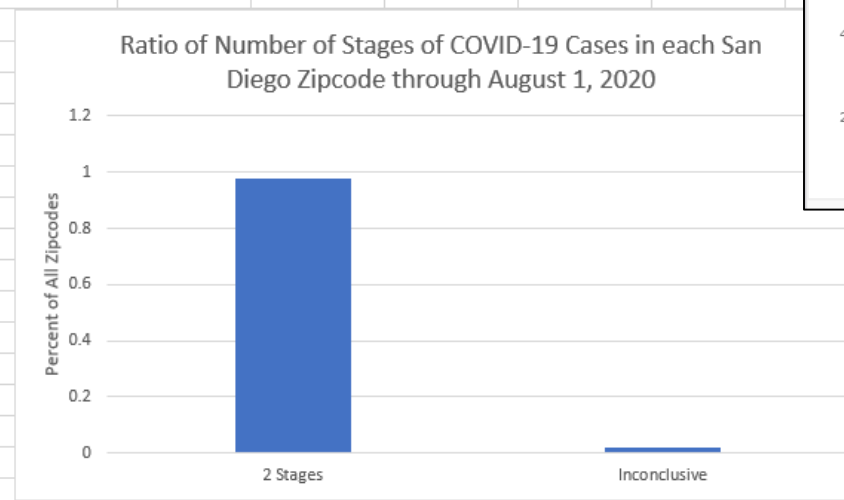
There are **55 Zip codes with a second growth stage** starting on **June 20**, 2020, while the second growth stage for another 9 Zipcode areas starts on June 21, 2020…..

Bins	Frequency
20-Jun	55
21-Jun	9
22-Jun	6
23-Jun	3
24-Jun	6
25-Jun	2
26-Jun	2
27-Jun	2
28-Jun	0
29-Jun	3
30-Jun	5

Knot Count for Significant Piecewise Regression Models of COVID-19 Cases in San Diego County
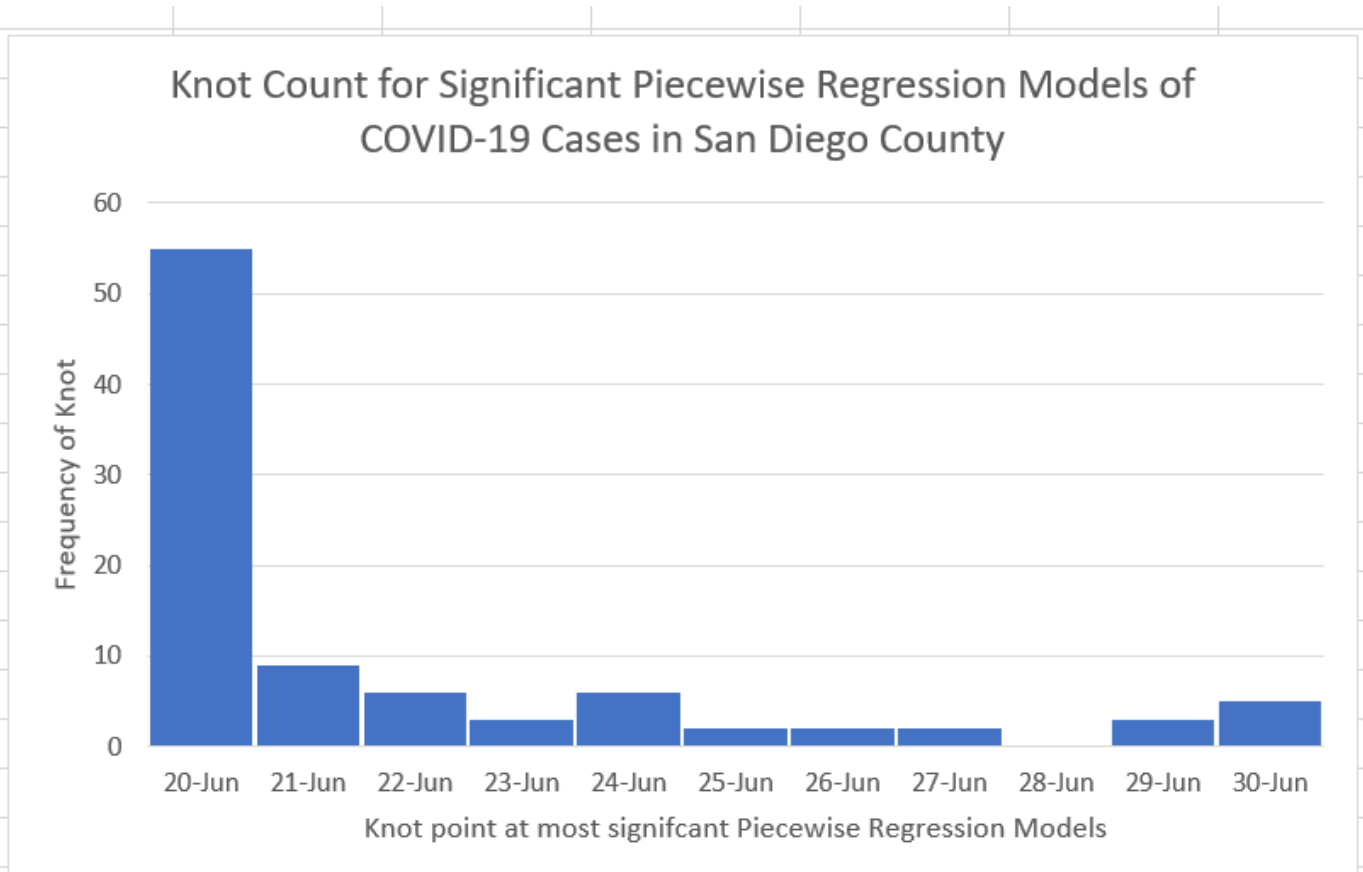
What Happen before 6/20/2020?

Major Events in San Diego before 6/20/2020:

6/12 The Re-Opening of Gyms, Bars, Movie Theaters in San Diego County.

Date	Event	Link	Source
5/20/2020	Face coverings to be required when restaraunts, stores open	https://www.county	San Diego County
5/20/2020	State approved the County of San Diego to allow retail businesses to have customers in stores with restrictions.	https://www.sandie	San Diego County
5/26/2020	Places of Worship, Hair Salons and Barbershops Can Open with Modifications	https://www.county	San Diego County
6/2/2020	Beach activities can restart with guidelines	https://www.county	San Diego County
6/5/2020	City to Reopen Lakes for Recreational Use	https://www.sandie	San Diego Mayor
6/9/2020	Beach parking lots can reopen	https://www.county	San Diego County
6/12/2020	Indoor movie theatres can reopen June 12	https://www.county	San Diego County
6/12/2020	Gyms, Hotels, Bars, Other Businesses to Open June 12	https://www.county	San Diego County
6/18/2020	Mayor Introduces Proposal to Create More Outdoor Dining and Retail Space	https://www.sandie	San Diego Mayor
6/18/2020	Gov. Newsom orders mask mandate	https://www.cdph.c	Gov. Newsom
6/19/2020	Personal care services (nail salons, tattoo parlors, body waxing) may reopen	https://www.sdsher	San Diego County
7/1/2020	San Diego Restaurants to Close Nightly at 10 PM	https://www.county	San Diego County

20 – 12 = 8 days → The average COVID-19 Incubation period is 5 days.
The incubation period for COVID-19 is thought to extend to 14 days, with a median time of 4-5 days from exposure to symptoms onset. https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html

Note: COVID-19 timeline tables for 16 major U.S. cities and California. Includes: Dallas, New York City, Los Angeles, Chicago, Houston, Washington D.C., Boston, Denver, Detroit, Seattle, Las Vegas, San Francisco, Minneapolis, Miami, Phoenix, San Diego, California, US, and Global.
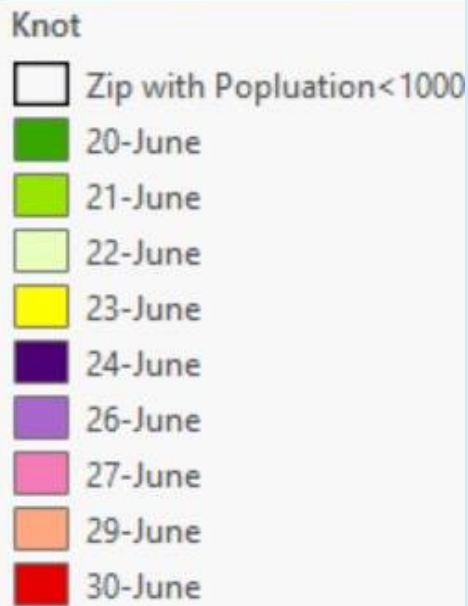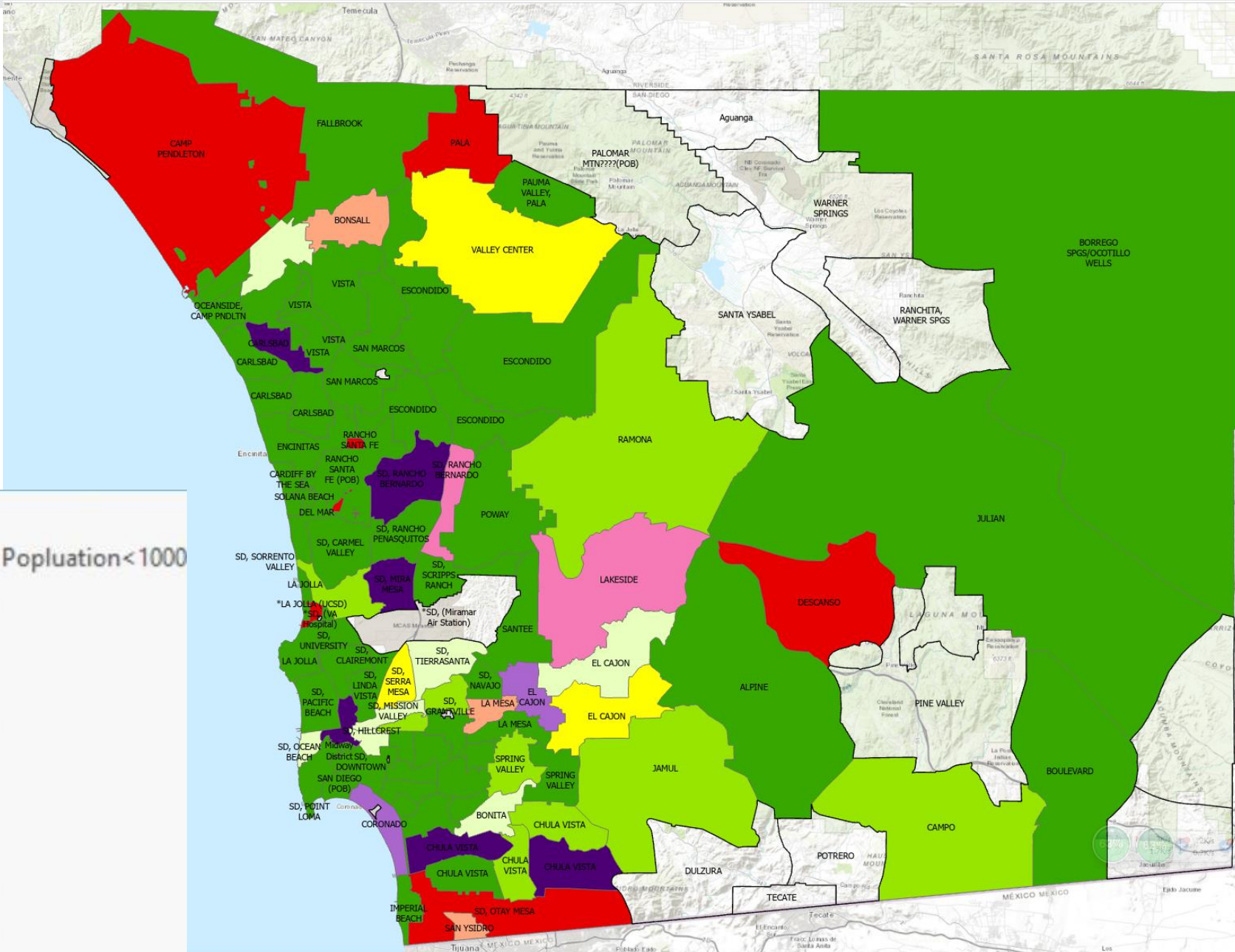
Major findings:

1. The high risk areas are located in South Bay Area (**Logan Heights, National City, Chula Vista (west), Chula Vista (south), and Otay Mesa**.). Their growth patterns (wave) are also unique compared to other areas.
2. Some zip codes have unique **growth rate change** patterns, including Otay Mesa, Coronado, San Ysidro, Spring Valley, Paradise Hills, National City, Scripps Ranch.
3. Most clearly show **two stages of growth patterns** with significant differences.
4. Most zip codes started the second growth rate change on 6/20/2020.

Q & A ?