One Health & Outbreak Surveillance Online Symposium 2021

November 29, 2021 (Online ZOOM Symposium)

Big Data Analytics for Outbreak Surveillance:

Using Social Media, Smart Phones, Census, and Surveys, to Monitor Flu and COVID-19 Outbreaks

Prof. Ming-Hsiang (Ming) Tsou （鄒明祥）

Twitter **@mingtsou** mtsou@mail.sdsu.edu,

Director of **the Center for Human Dynamics in the Mobile Age**

Director of **the Big Data Analytics Program (Master of Science)**

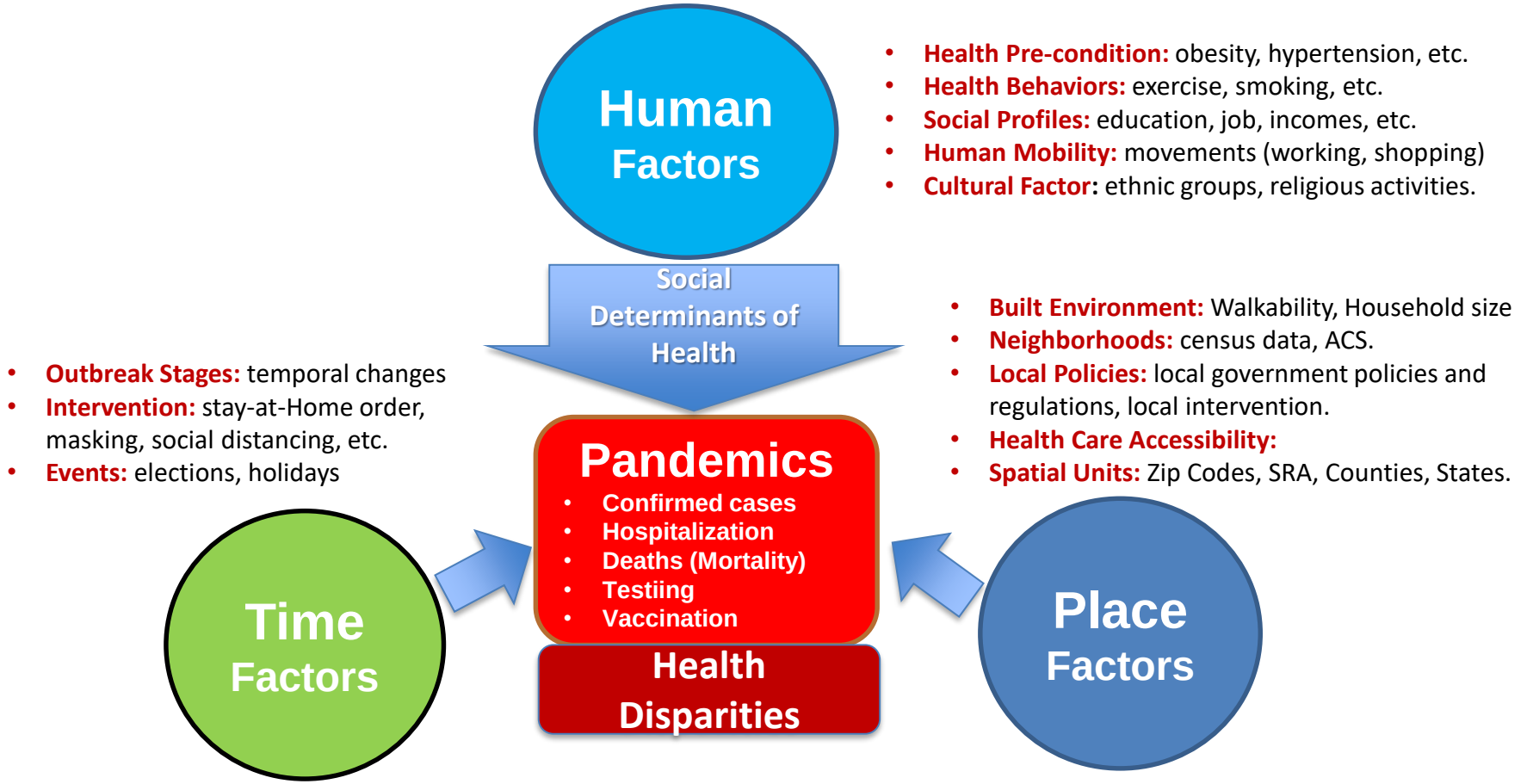
Professor, Department of Geography , San Diego State University

HDMA @SDSU | THE CENTER FOR HUMAN DYNAMICS IN THE MOBILE AGE

"Pandemics" are not only medical disease outbreaks, but also a serious problem of social inequality and health disparity.

An Integrated Big Data Analytics Approach

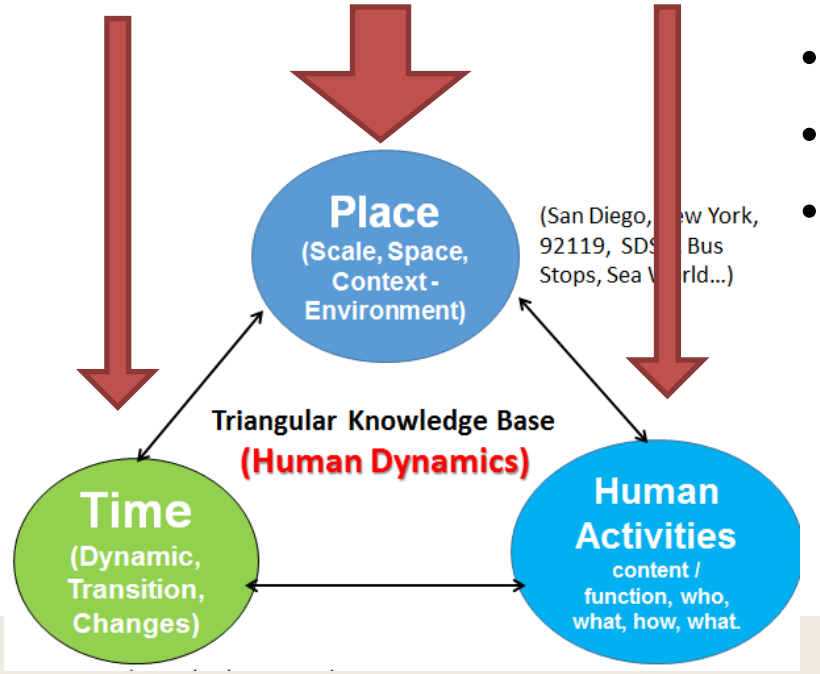
To monitor and mitigate pandemics, we need to develop an integrated Big Data Analytics Approach to track and analyze the spatiotempoal patterns of pandemics using the Human, Place, Time (HPT) framework.



Big Data + Mobile Technology

Enabling New Research Opportunity and Innovations

- GPS tracking devices, Wireless Sensors
- Wi-Fi Tracking, Cellular signal tracking,
- Smart Phone Apps and Tablets/iPad Apps
- Built-in Camera, Voice recorders.



Near-Real Time Monitoring Pandemics !

Image source: apple.com

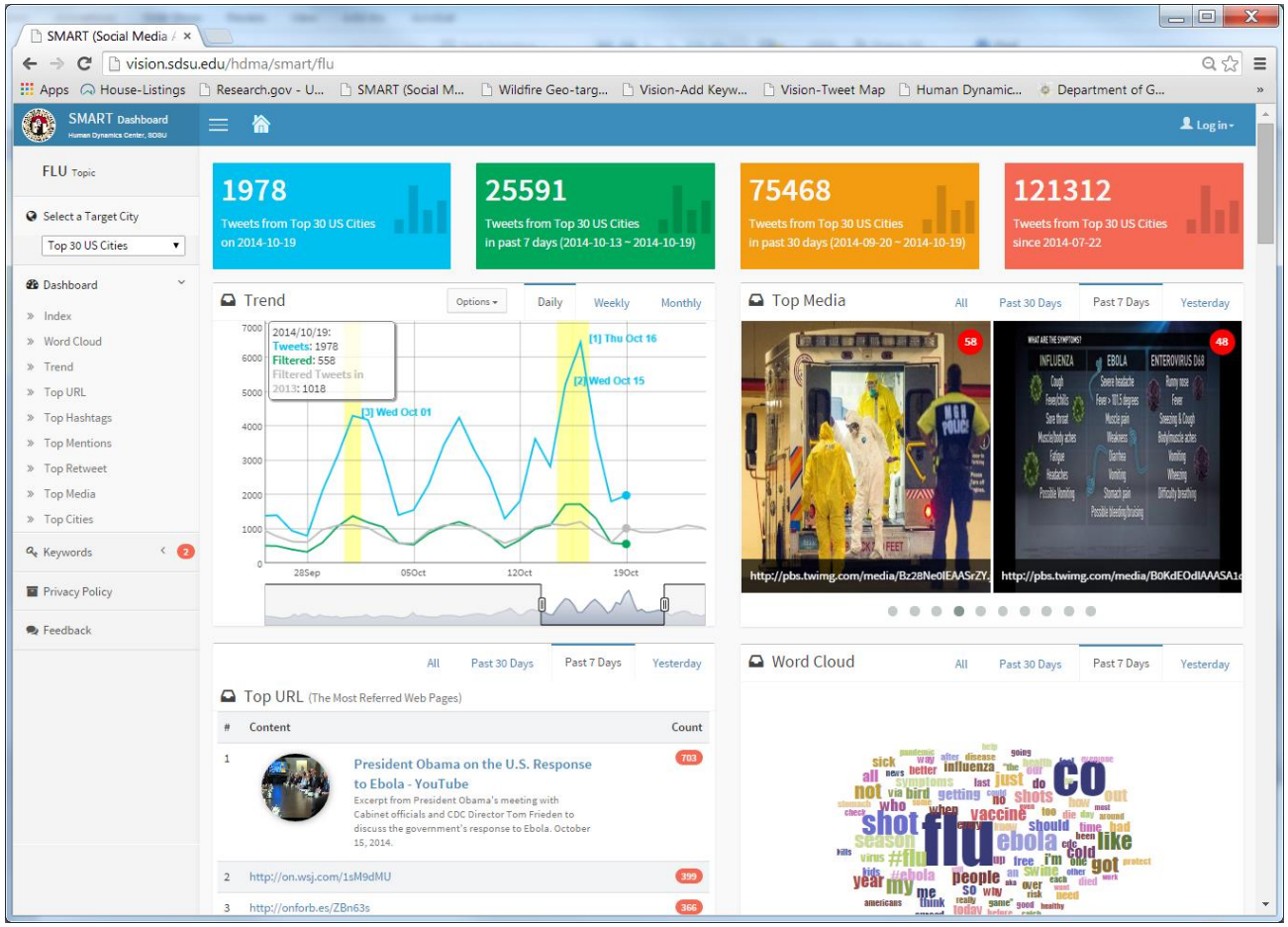
Research Showcase #1:

Geo-Targeted Social Media (Twitter) Analytics for Tracking Flu Outbreaks in U.S. (2014)

SMART Dashboard

[YouTube Video](https://www.youtube.com/watch?v=NaUhMZJysHM)
https://www.youtube.com/watch?v=NaUhMZJysHM

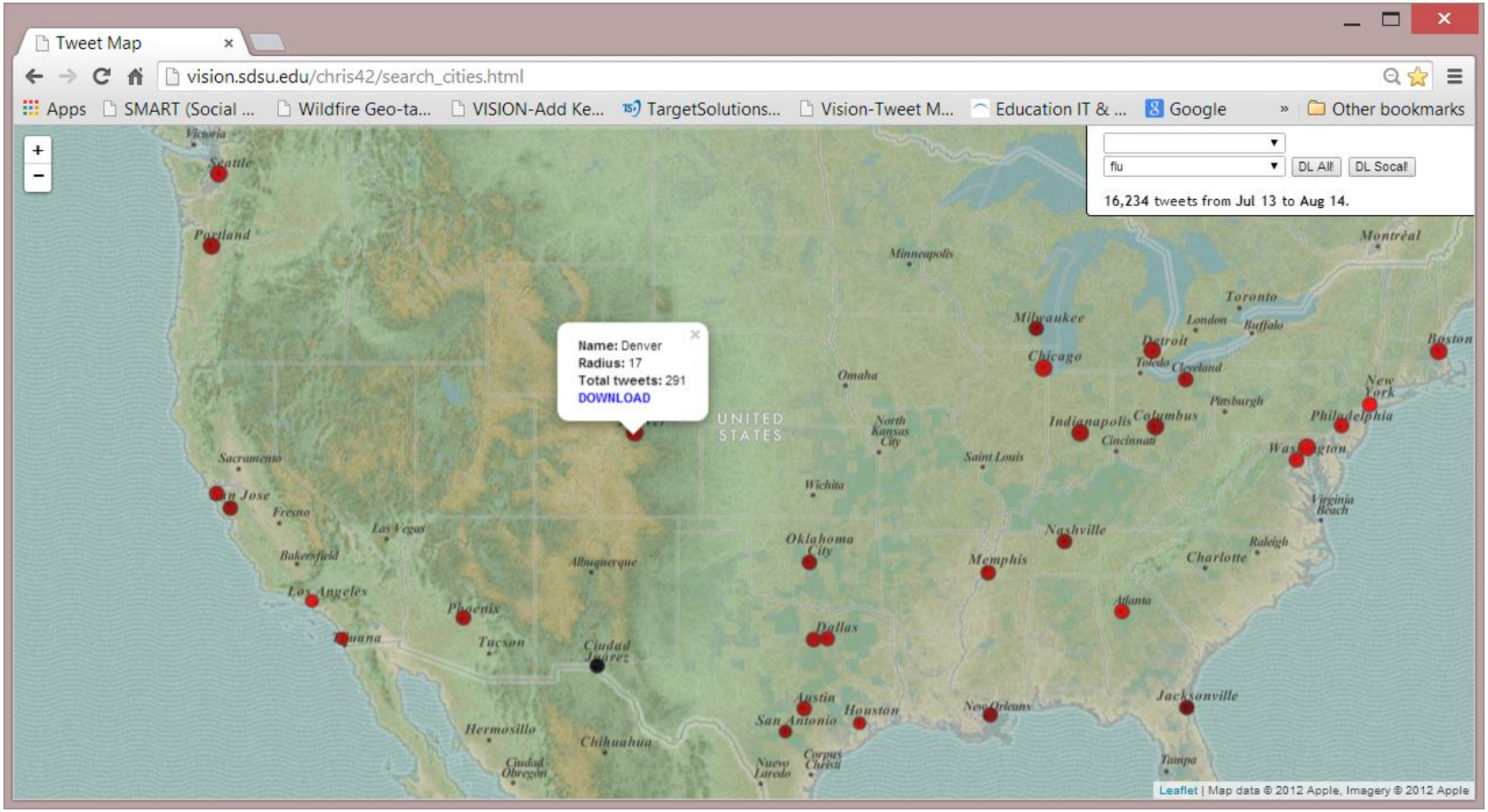
Social Media Analytic and Research Testbed
http://vision.sdsu.edu/hdma/smart/



Real-time social media analytics (Trend Analysis, Word Clouds, Top URL, web pages, Top Hashtags/Mentions/Stories).

Collect Tweets from Top 31 U.S. Cities (17 miles radius)

31 different cities across the United States (chosen based on their population sizes): *Atlanta, Austin, Baltimore, Boston, Chicago, Cleveland, Columbus, Dallas, Denver, Detroit, El Paso, Fort Worth, Houston, Indianapolis, Jacksonville, Los Angeles, Memphis, Milwaukee, Nashville-Davidson, New Orleans, New York, Oklahoma City, Philadelphia, Phoenix, Portland, San Antonio, San Diego, San Francisco, San Jose, Seattle, and Washington, D.C.* **Using "flu" or "Influenza" keywords.**



Real-Time Monitoring of Flu Outbreaks in U.S.

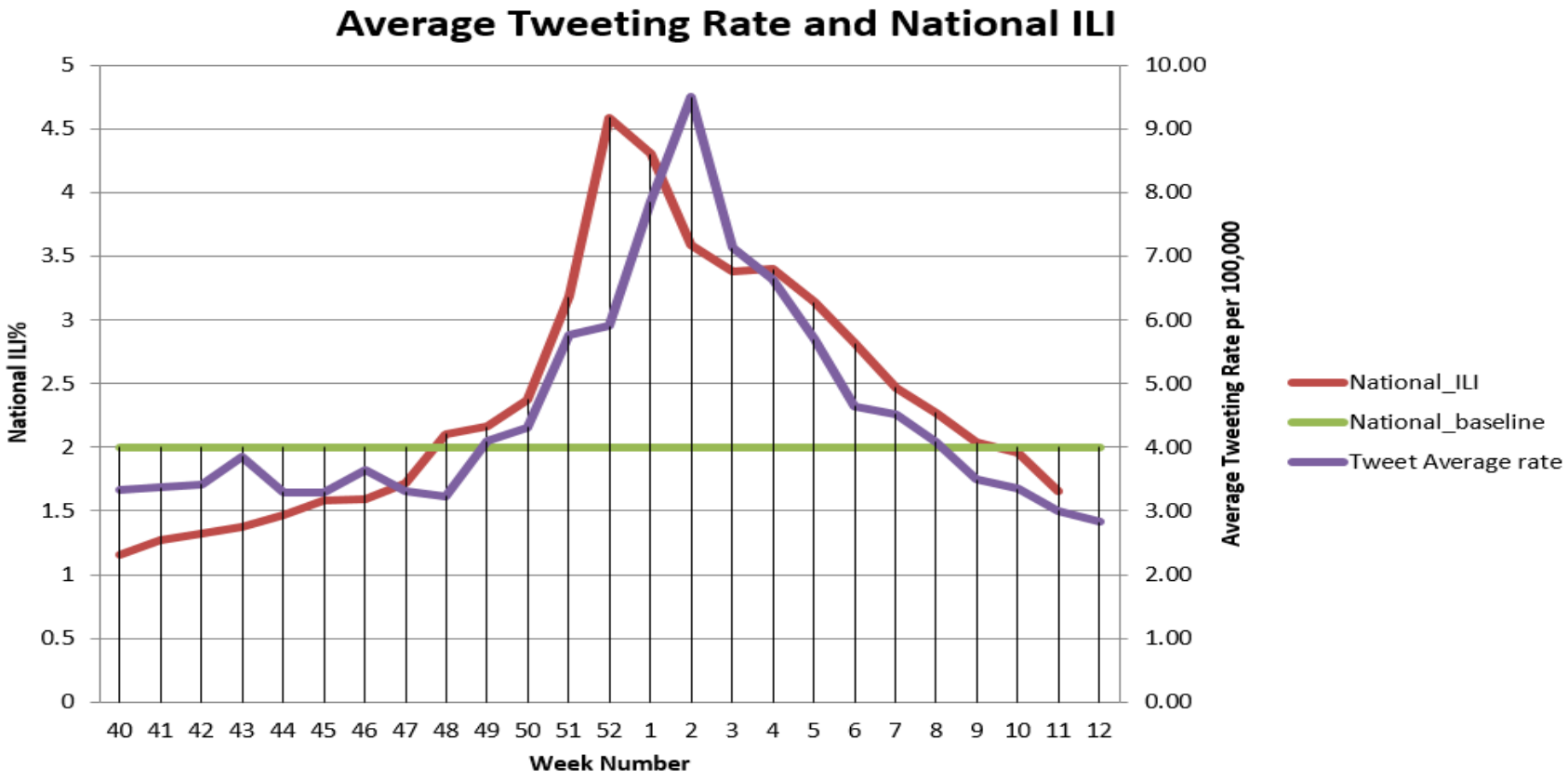
(National Scale – combined 31 Cities), 2013 – 2014 flu season

RED Line: National ILI (provided by CDC)

Purple Line: Weekly Tweeting Rate

(R) value = **0.8494**

ILI: Influenza-like Illness



CDC in Atlanta

Two research papers in the

Journal of Medical Internet Research

Original Paper

The Complex Relationship of Realspace Events and Messages in Cyberspace: Case Study of Influenza and Pertussis Using Tweets

Anna C Nagel[1], MPH; Ming-Hsiang Tsou[2], PhD; Brian
Dipak K Gupta[5], PhD; Jiue-An Yang[2], MA; Su Han[2], M
Mark H Sawyer[6,7], MD

[1]Graduate School of Public Health, San Diego State University, San Dieg
[2]Department of Geography, San Diego State University, San Diego, CA, U
[3]School of Communication, San Diego State University, San Diego, CA,
[4]Department of Linguistics, San Diego State University, San Diego, CA,
[5]Department of Political Science, San Diego State University, San Diego,
[6]Division of Pediatrics Infectious Diseases, University of California San I
[7]Epidemiology and Immunization Services Branch, City of San Diego He

Corresponding Author:
Ming-Hsiang Tsou, PhD
Department of Geography
San Diego State University
Storm Hall #326
5500 Campanile Dr

2013

2014

Original Paper

The Reliability of Tweets as a Supplementary Method of Seasonal Influenza Surveillance

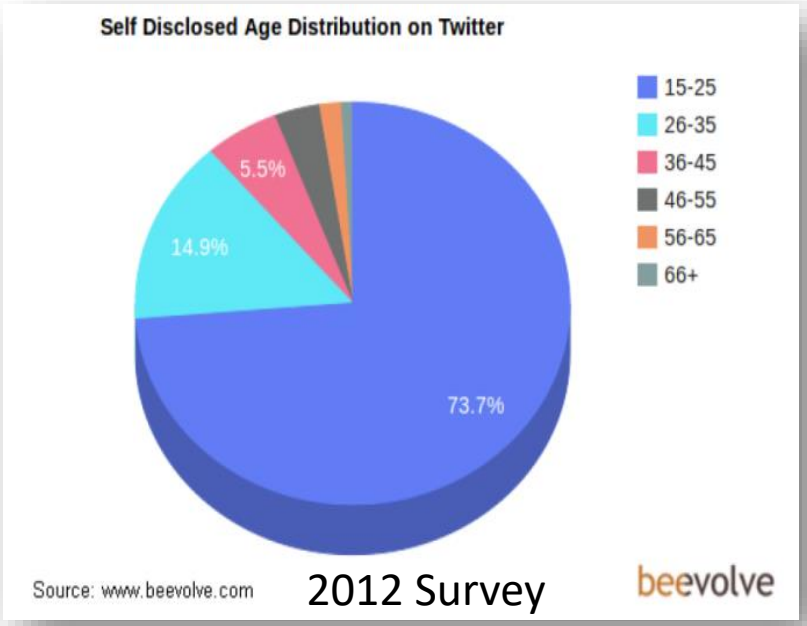
Anoshe Amber Aslam[1], MPH; Ming-Hsiang Tsou[2], PhD; Brian H Spitzberg[3], PhD; Li An[2], PhD; J Mark Gawron[4], PhD; Dipak K Gupta[5], PhD; K Michael Peddecord[1], PhD; Anna C Nagel[1], MPH; Christopher Allen[2], MA; Jiue-An Yang[2], MA; Suzanne Lindsay[1], PhD

[1]Graduate School of Public Health, San Diego State University, San Diego, CA, United States
[2]Department of Geography, San Diego State University, San Diego, CA, United States
[3]School of Communication, San Diego State University, San Diego, CA, United States
[4]Department of Linguistics, San Diego State University, San Diego, CA, United States
[5]Department of Political Science, San Diego State University, San Diego, CA, United States

Corresponding Author:
Ming-Hsiang Tsou, PhD
Department of Geography
San Diego State University
Storm Hall 313C
5500 Campanile Drive
San Diego, CA, 92115

The Limitations and Challenges of Social Media

Social Media messages can NOT represent all population, but it can provide **warning signals** and **real-time updates**.



Twitter Users are

- **Young** (75% are between 15 – 25 years old).
- More **Urban** residents than rural
- Higher adoption% in African Americans
- Many Journalists and **Mass Media** staff.
- **30% are not real "human beings" (bots):** many advertisement and marketing activities.

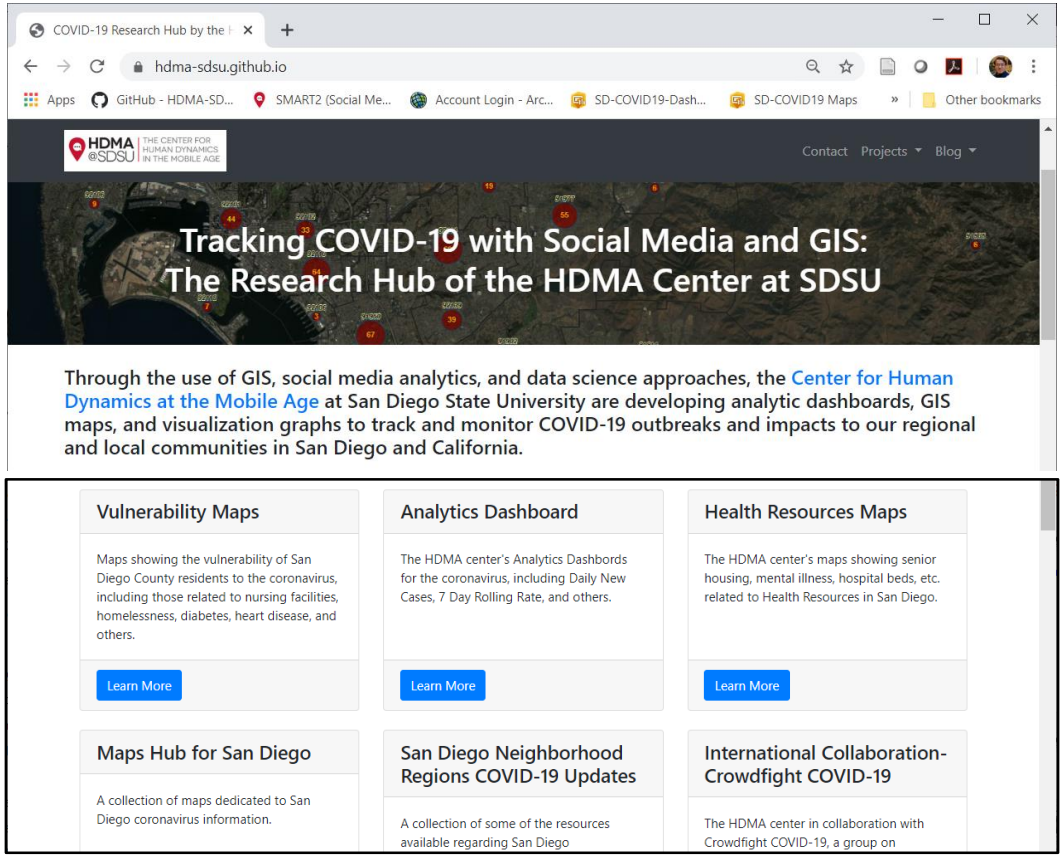
Using Different Keywords can get different demographic groups:

- **#Healthcare: include more senior people (Very few teenagers will tweet about "healthcare"). (We need more background study).**
- **"Keywords" as a sampling tool for social media users.**

Research Showcase #2:

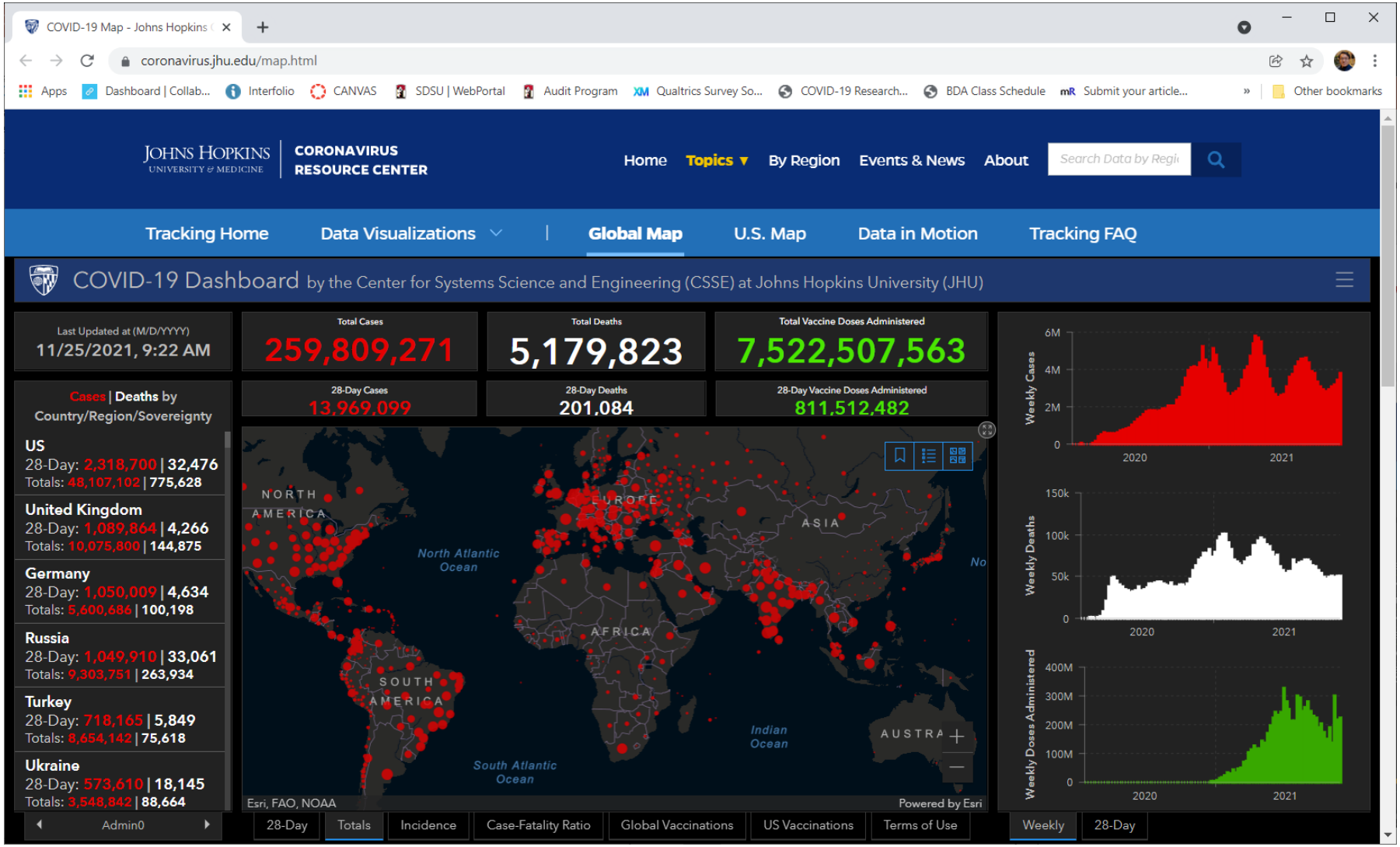
Tracking COVID-19 Pandemic in San Diego and U.S. cities.

https://hdma-sdsu.github.io/



The Famous COVID-19 Map developed by Johns Hopkins University

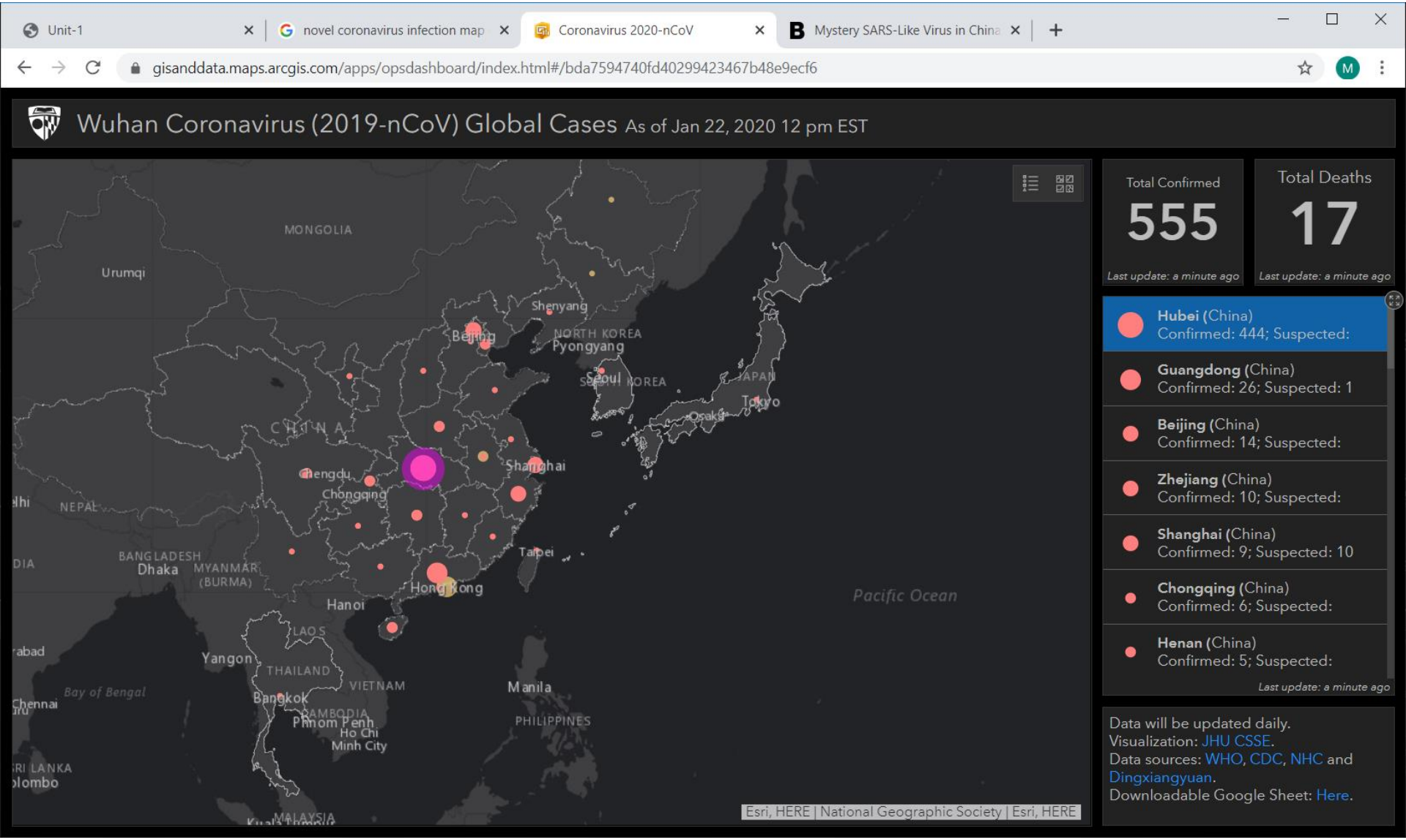
(using ESRI's ArcGIS Online tool, called "Operation Dashboard" (11/25/2021).



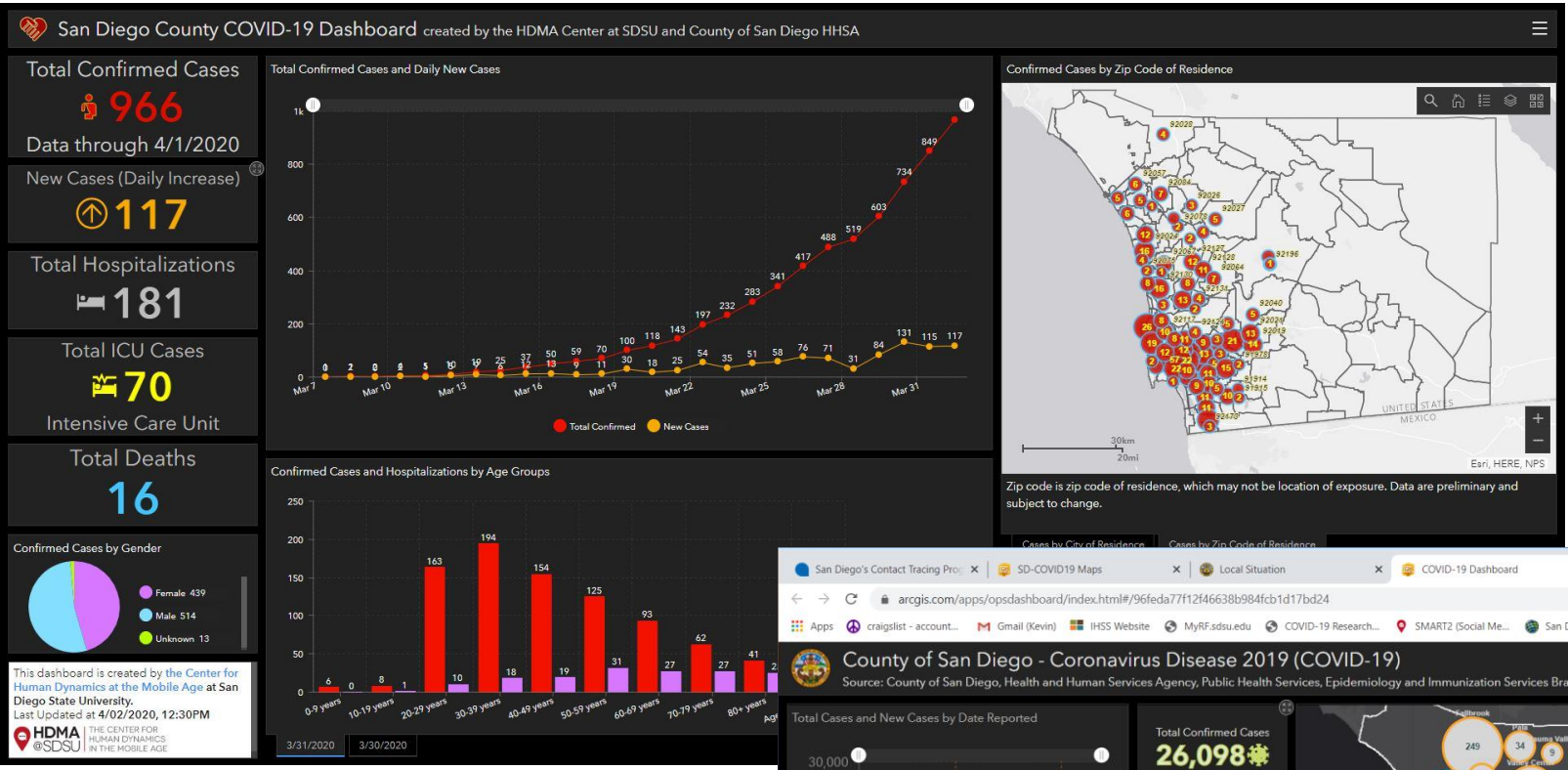
https://coronavirus.jhu.edu/map.html (What's today's number?)

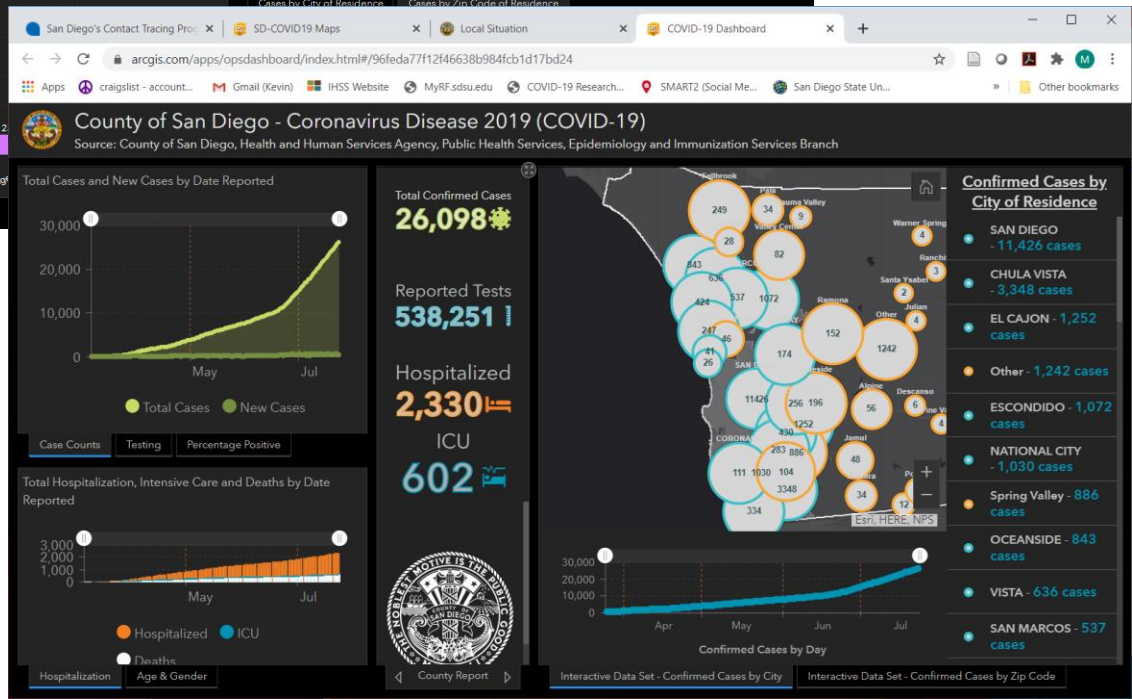
The early version (Jan 22, 2020): Track the spread of novel coronavirus, Johns Hopkins University. 555 total confirmed cases, 17 deaths.

https://www.pbs.org/newshour/health/track-the-spread-of-novel-coronavirus-with-this-map
https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6



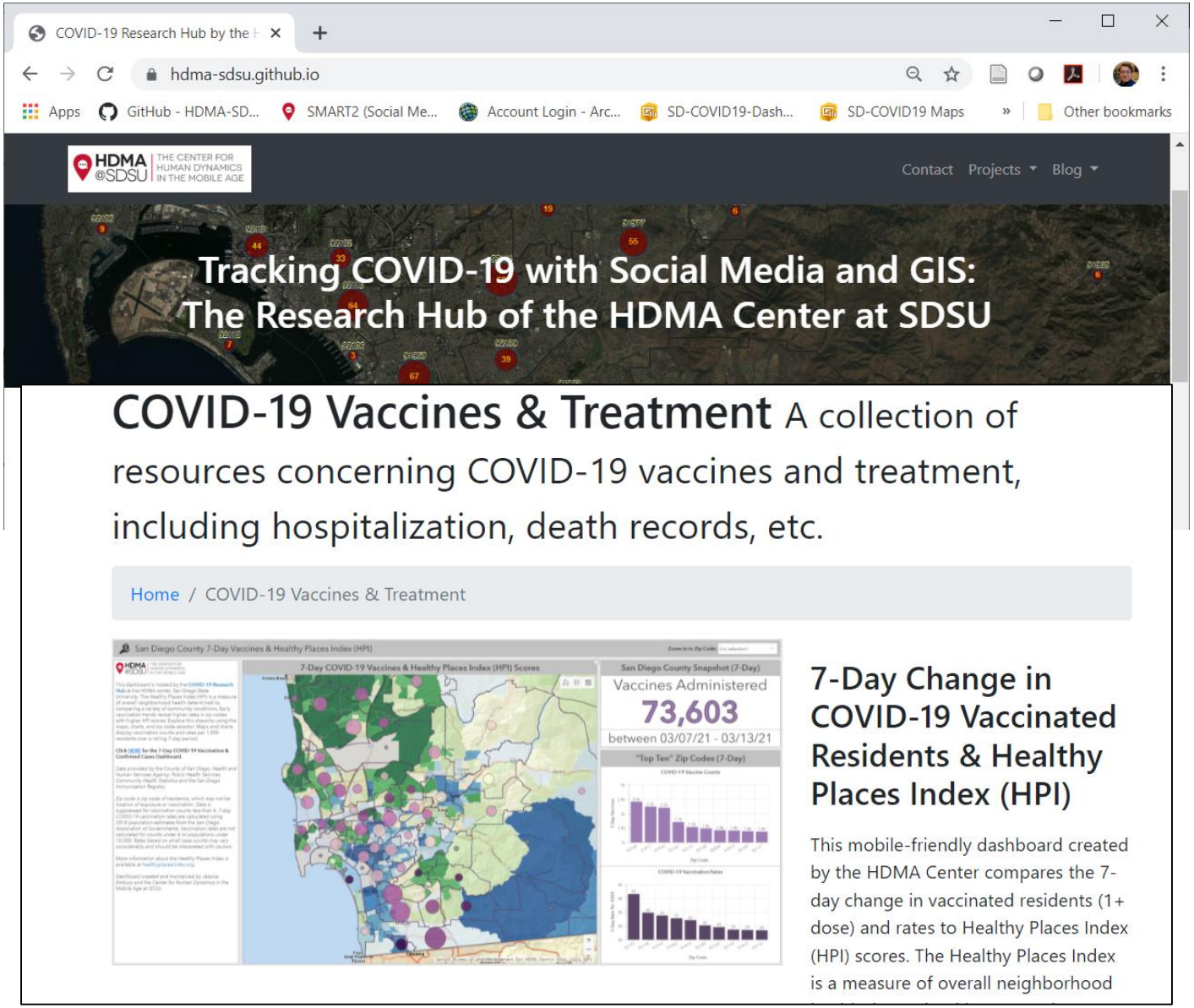
The first COVID-19 Dashboard for San Diego County (4/02/2020) – created by Dr. Ming-Hsiang Tsou and his students at the HDMA Center.





San Diego County GIS team created a new version dashboard using ESRI operation dashboard after 4/09/2020 ….

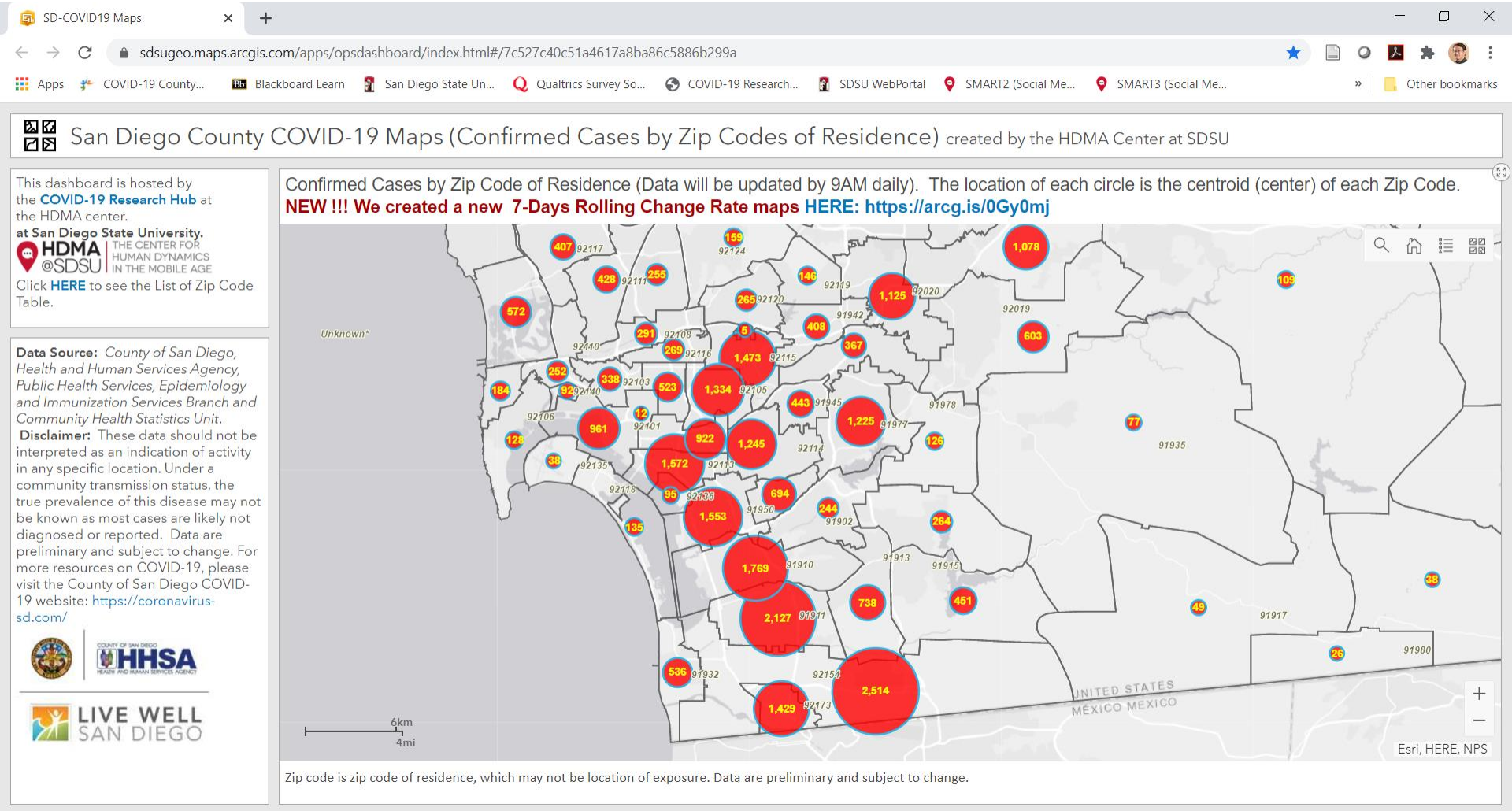
Research Hub (4/27/2020):
https://hdma-sdsu.github.io/



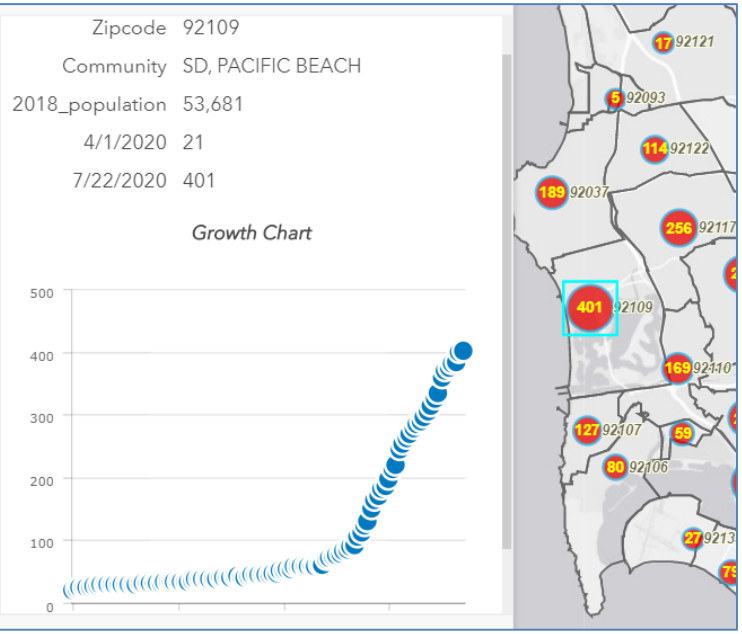
COVID-19 Confirmed Cases at ZipCode Level with Growth Charts

(Daily Updated - Started on April 01, 2020) https://arcg.is/1zXq1m

(weekly updated AFTER July 01, 2021 in San Diego County)

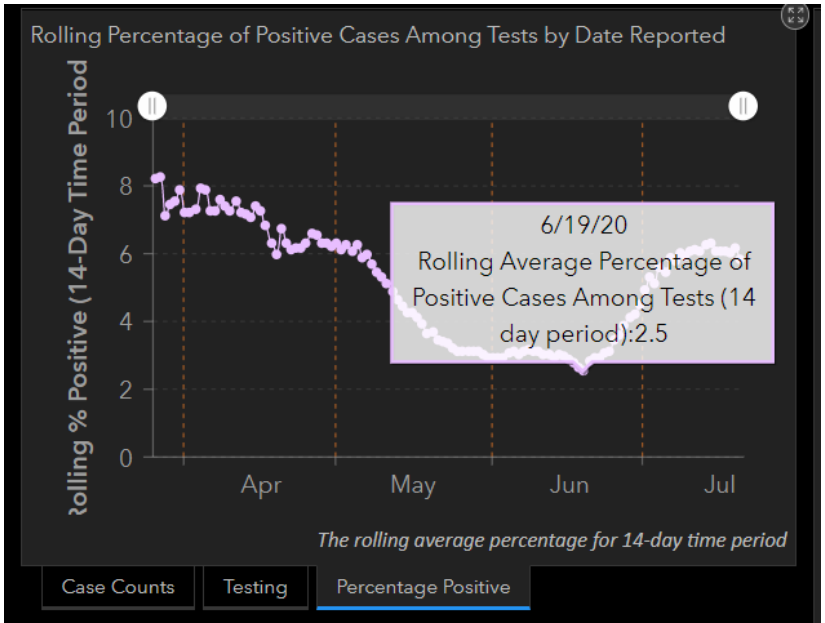


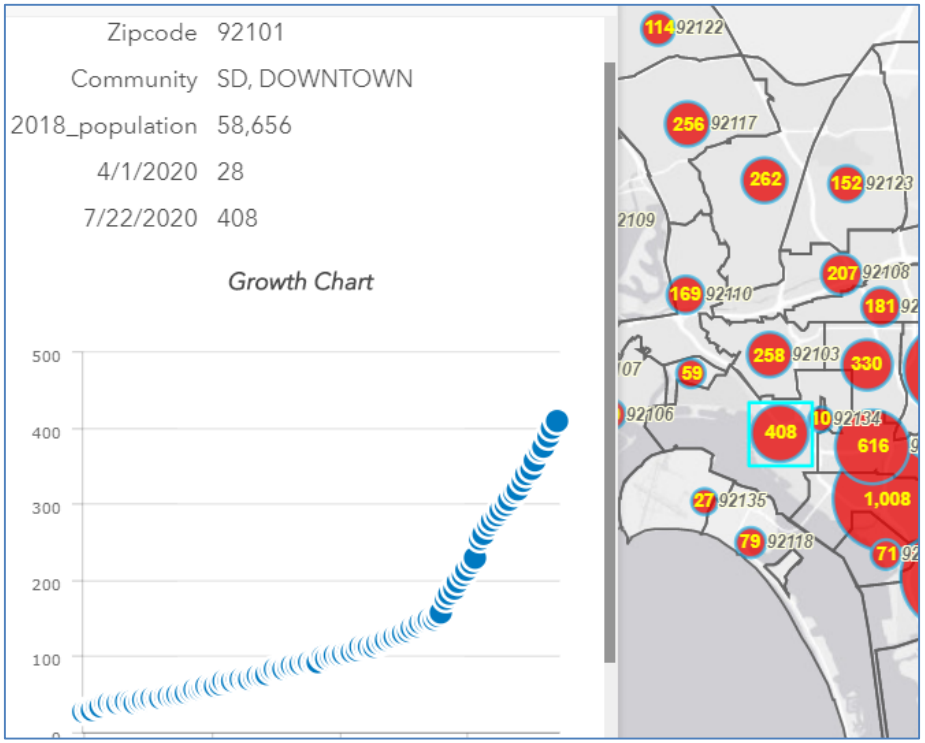
San Diego Outbreak Event #2: The First Wave (Growth Stage) Outbreaks Date: After 6/20/2020 (start to increase rapidly).



Pacific Beach: from 40 cases(5/19) to **90 (6/19, 2.25x) to 401 cases (7/22, 4.5x).**

SD Downtown: from 85 cases (5/19) to 144 (6/19, **1.7x**), to 408 cases (7/19, **2.8x**).





What Happen before 6/20/2020?

Major Events in San Diego before 6/20/2020:
Our HDMA Team has created the **time table** (excel sheet) for 16 U.S. cities, including San Diego). URL:
https://www.dropbox.com/sh/gk3dguqrosmiab1/AADf6QFfn04tj2k6mB15PHjMa?dl=0

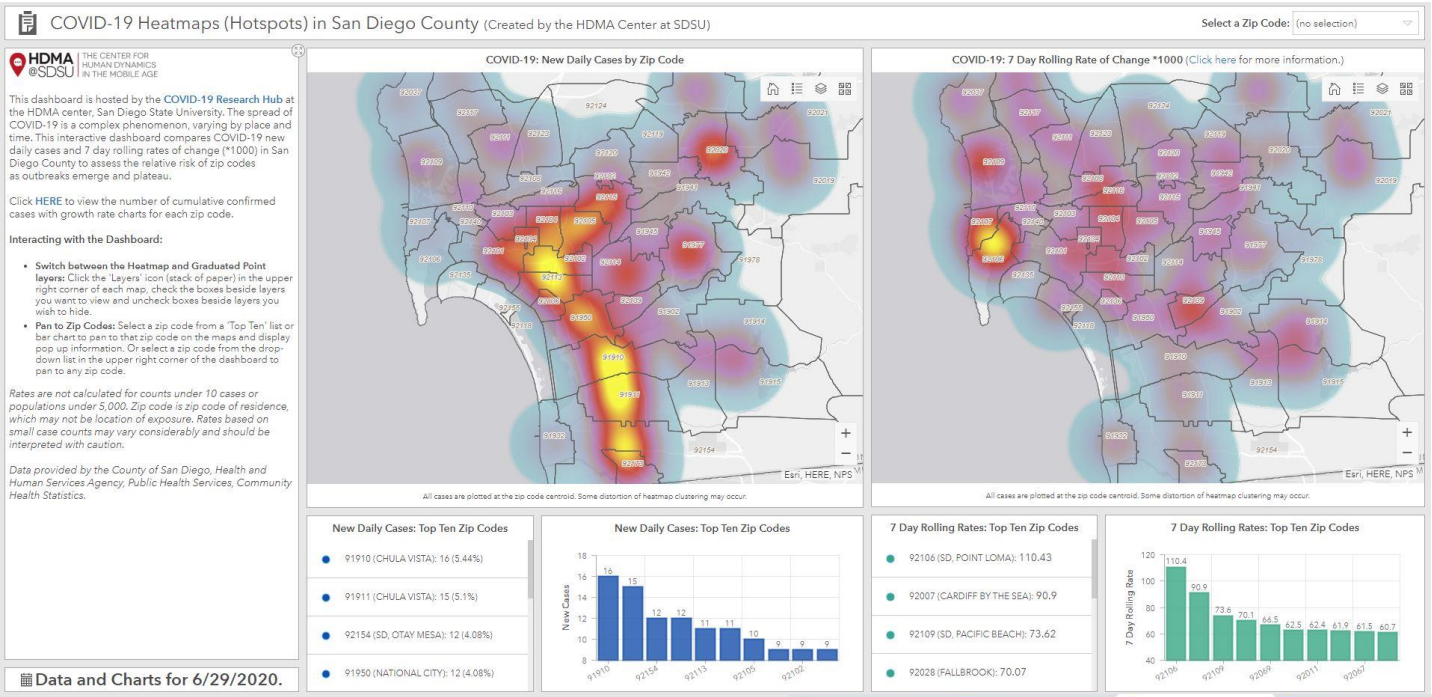
4/27/2020	Limited reopenings of beaches, bays, oceans	https://www.sandiego.gov/mayor/ne	San Diego Mayor
5/5/2020	Local businesses can begin to reopen with plan	https://www.countynewscenter.com	San Diego County
5/12/2020	San Diego County allows more businesses to open	https://www.kpbs.org/news/2020/ma	San Diego County
5/20/2020	State approved the County of San Diego to allow retail businesses to have	https://www.sandiegocounty.gov/co	San Diego County
5/27/2020	City of San Diego to Distribute $1M in Funding to Support Artists Strugglin	https://www.sandiego.gov/mayor/ne	San Diego Mayor
5/26/2020	Barbershops and hair salons can reopen with guidelines	https://covid19.ca.gov/pdf/guidance	Gov. Newsom
5/27/2020	Places of worship can reopen with guidelines	https://covid19.ca.gov/pdf/guidance	Gov. Newsom
6/5/2020	City to Reopen Lakes for Recreational Use	https://www.sandiego.gov/mayor/ne	San Diego Mayor
6/12/2020	San Diego County gyms, bars, movie theaters reopening	https://www.10news.com/news/corc	Gov. Newsom
6/18/2020	Mayor Introduces Proposal to Create More Outdoor Dining and Retail Spa	https://www.sandiego.gov/mayor/ne	San Diego Mayor
6/19/2020	Personal care services (nail salons, tattoo parlors, body waxing) may reop	https://www.sdsheriff.net/newsroom	Gov. Newsom

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.

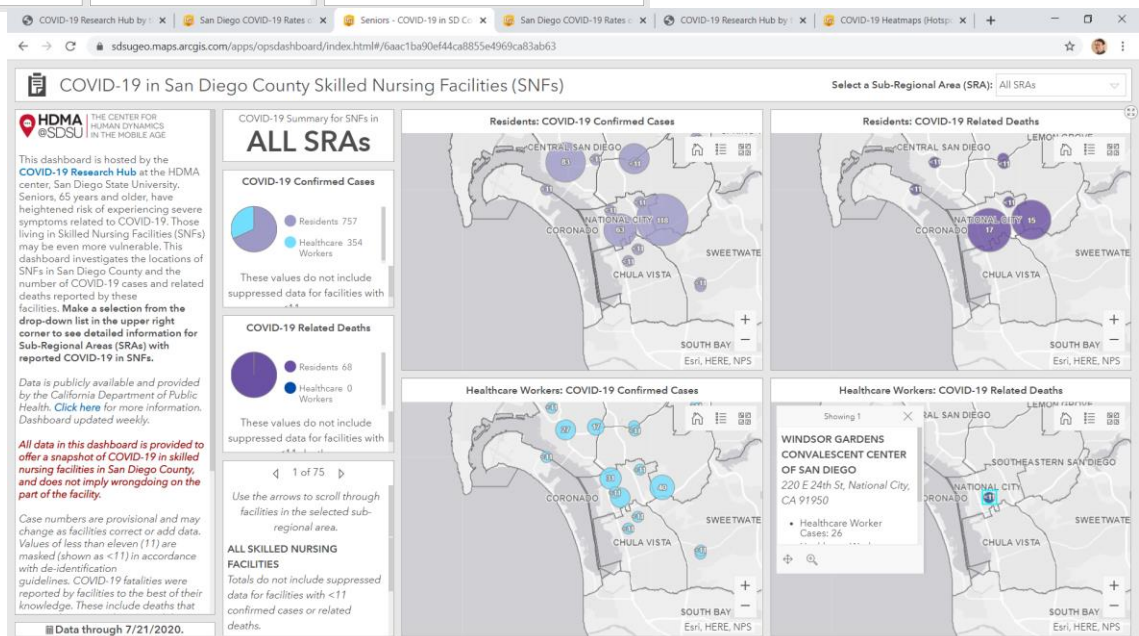
Heatmap (daily increased cases + 7 days rolling). (6/29/2020)

https://arcg.is/1br8v9

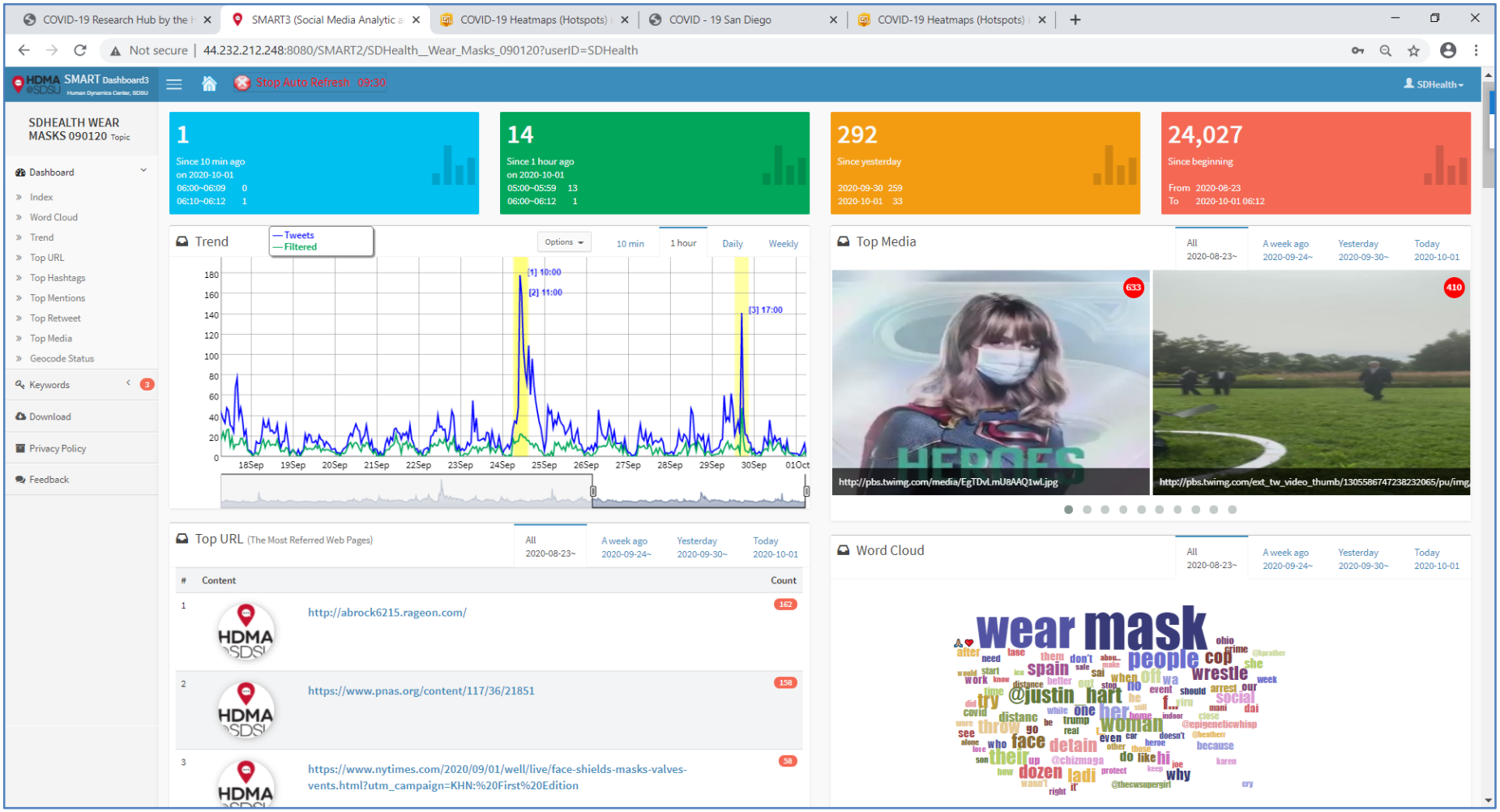


Residents in the Nursing Facility
COVID-19 Related Deaths

COVID-19 in **Skilled Nursing Facilities** in San Diego https://arcg.is/0bjCe1



Use SMART dashboard and Twitter data to analyze near real-time Public Opinions in Wearing Masks and Social Distancing in San Diego



July 6-12, 2020, Masks Topic (n=3074)

Main Themes:

- Masks are effective and should be worn to prevent COVID spread (80%)

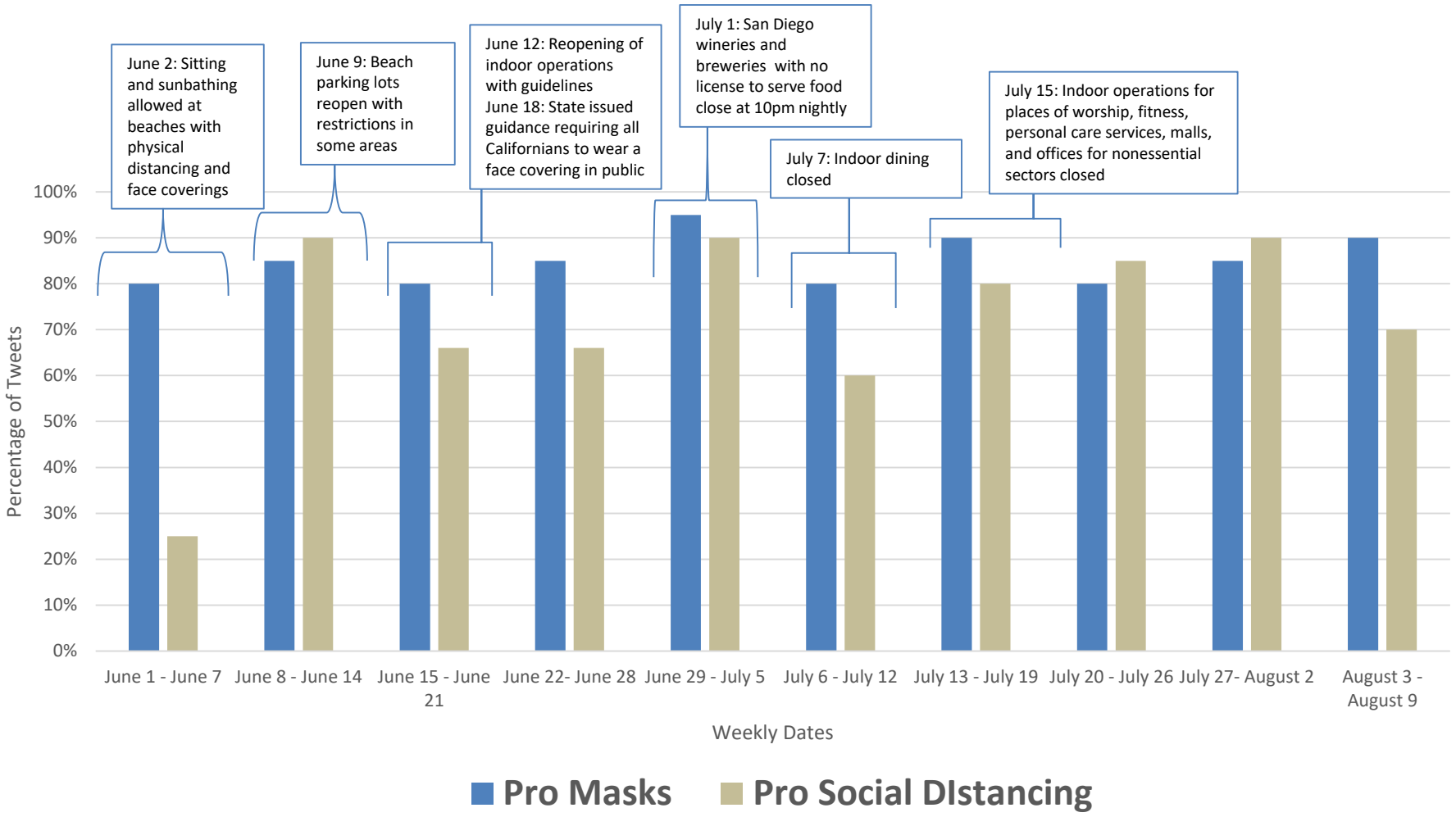
Emerging Themes (Anti-Masks):

- My body, my choice should apply to masks as well (liberty)
- Trump finally put on a mask
- Gavin Newsom is responsible for bad policies

Top Retweet: (665 RT)

- @CHIZMAGA I'm pretty sure after today the Left will never ask Trump to wear a Mask ever again. Boss!

Masks and Social Distancing Twitter Sentiment and Policy Timeline

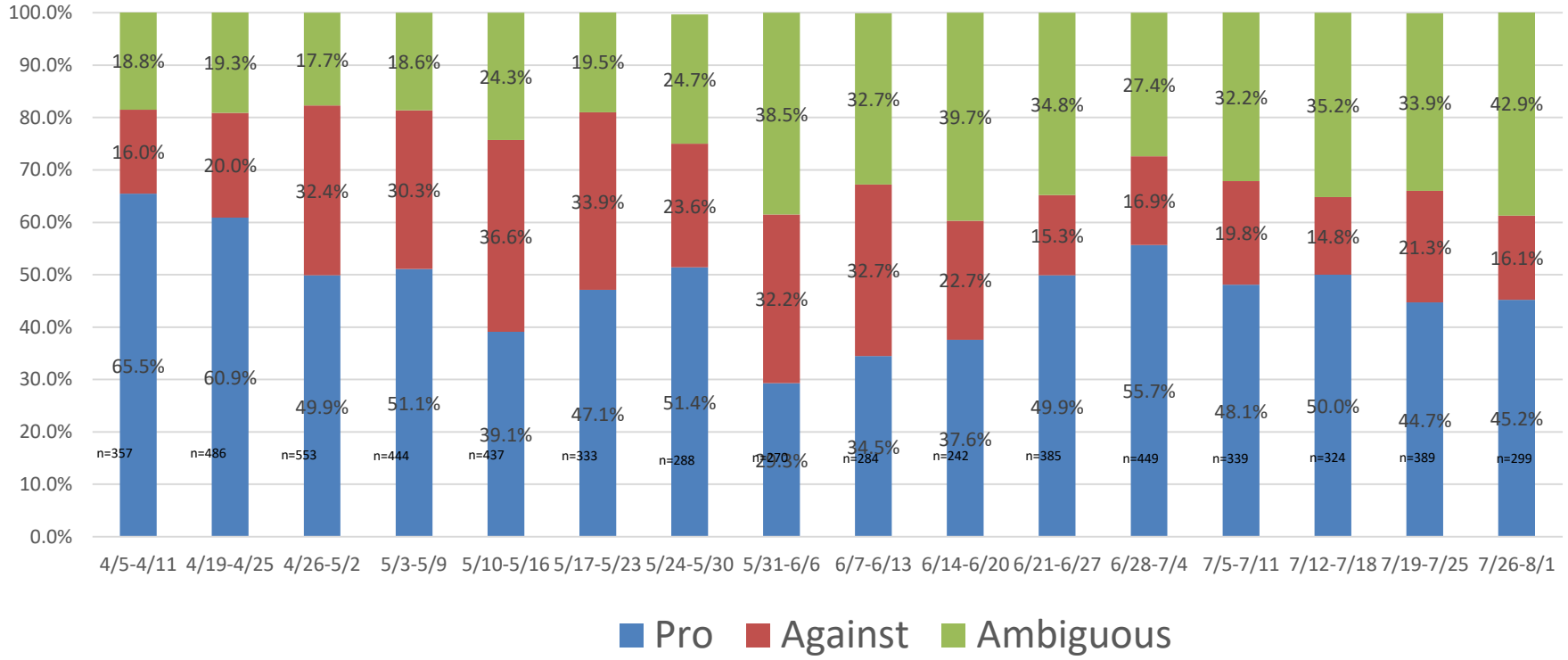


Stay-At-Home Policy Perceptions (%)

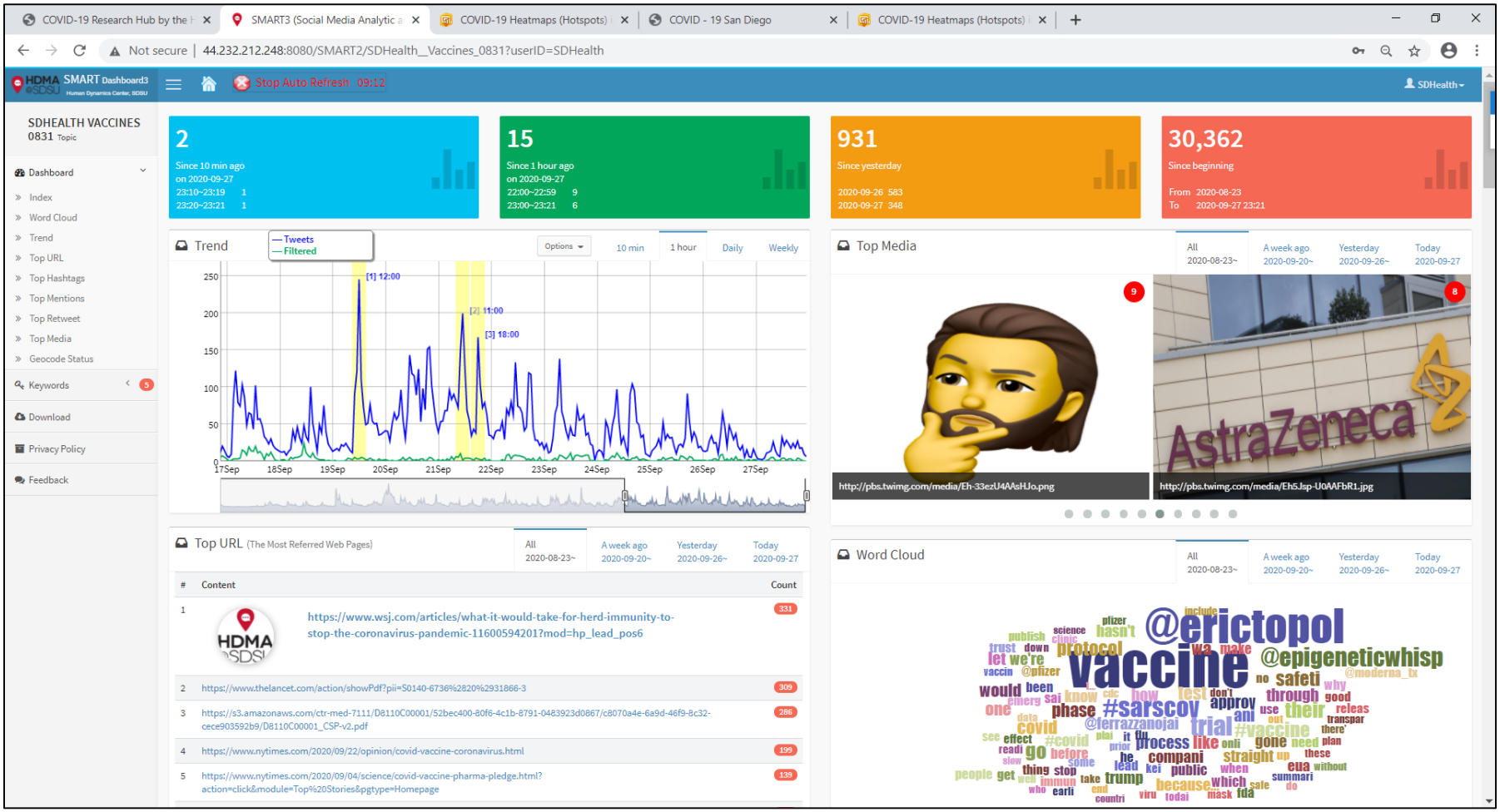
April 5-11 through July 25 - August 1 2020

(n=5,522)

These information can help health agencies to decide how to implement second or third “stay-at-home” policy in San Diego.



Analyzing COVID-19 Vaccine Issue



- **Identify option leaders in both Pro-Vaccine and Anti Vaccine groups**
- **Analyze the “fake news” and “misinformation”.**

8/3/2020 – 8/9/2020 Vaccines: (n=543)

Main Themes

- **Anti vaccine (~55% original tweets)**
  - Forced vaccine
  - Not effective because the virus is mutating
  - There is a political agenda with election coming up (distrust)
  - Why should I get a vaccine if people my age are asymptomatic? Or those who have died have had pre existing conditions?
  - Big Pharma has $ agenda

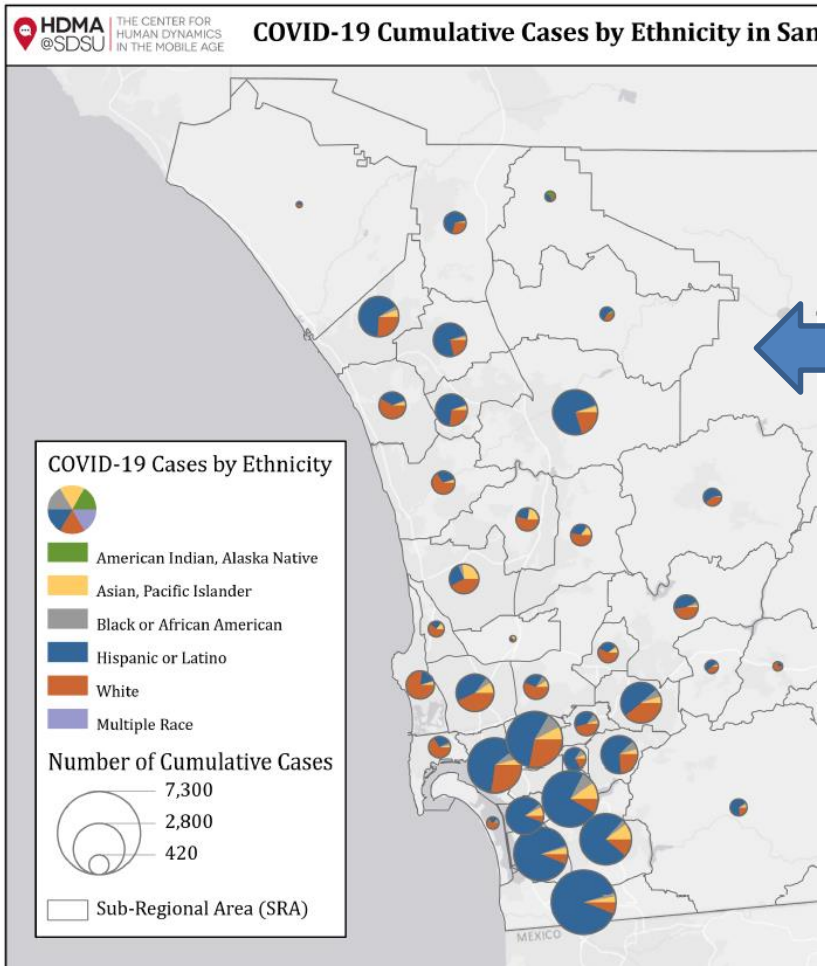
Top Retweets

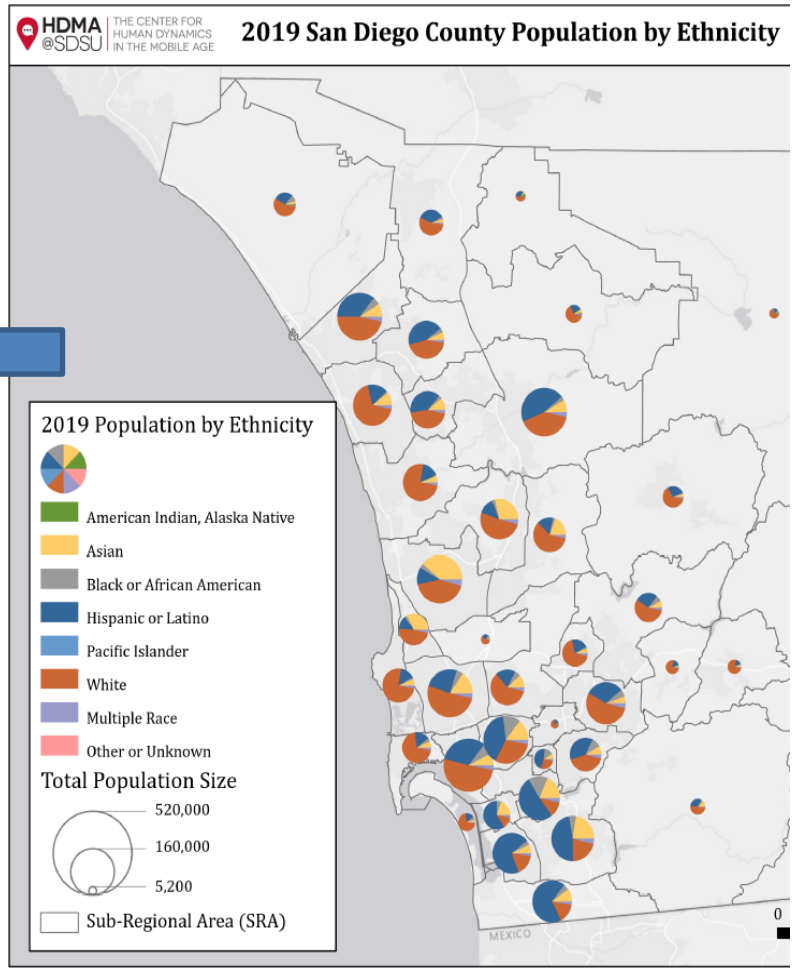
- (696 RT) @CHIZMAGA: “America doesn’t need a vaccine, it needs to find a cure for liberalism.”
- (420 RT) @CHIZMAGA: “Donald Trump is the vaccine to save America.”

COVID-19 Health Disparity Problems:

Hispanic Population got 3 times (3X) numbers of COVID-19 confirmed cases comparing to white ethnic group. (10/17/2020)

Hispanic Population: Blue, White Population: Brown.

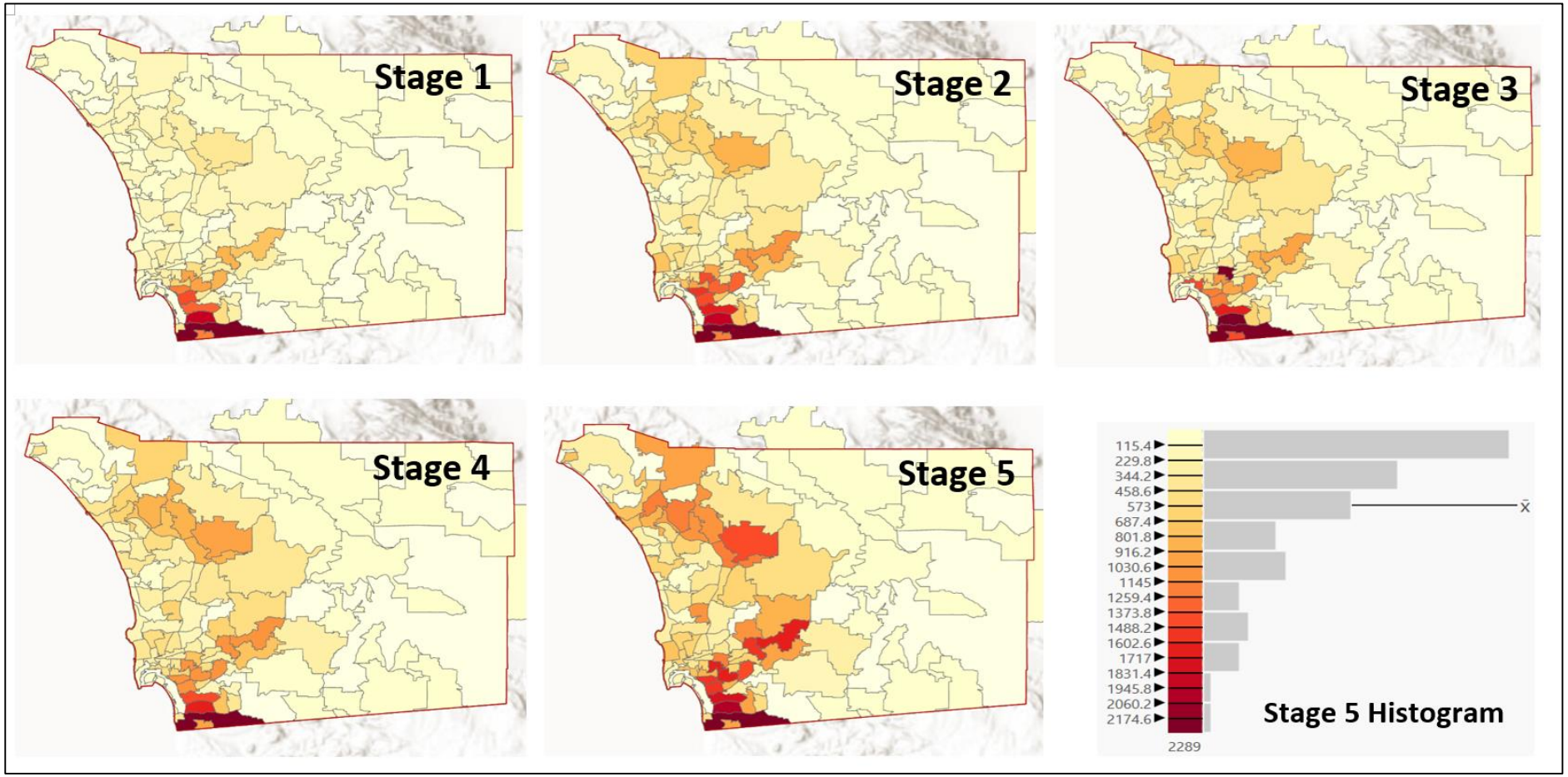




https://hdma-sdsu.github.io/img/SD-COVID19-Ethnicity-Maps-101720.pdf

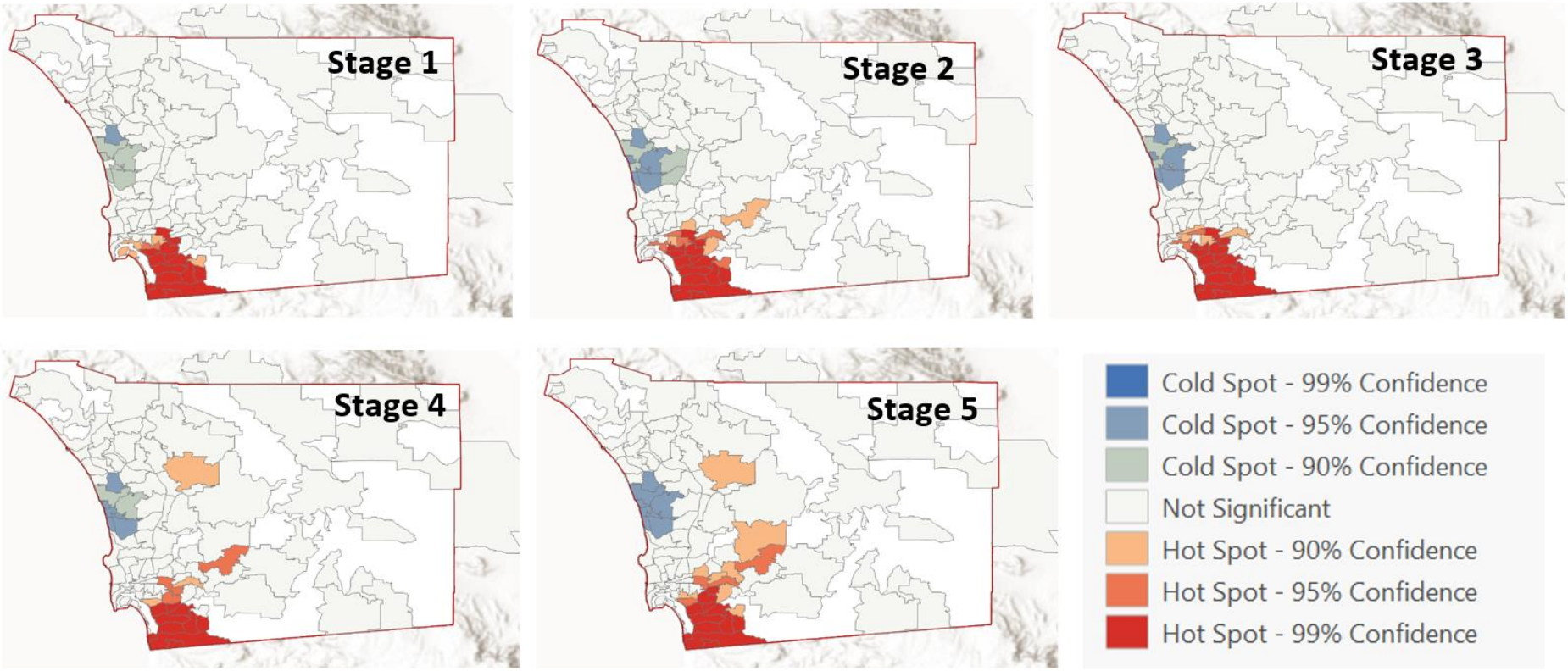
(The Border does not play an important role in the Health Disparity issues.)

The Spatial Cluster Patterns of COVID-19 in San Diego (4/01/2020 – 4/31/2021)



The Spatial Cluster Patterns of COVID-19 in San Diego during the five stages of outbreaks (using equal interval classification (20 classes) for the total confirmed cases in each ZIP Code units during each period). **South Bay areas = Hispanic communities**

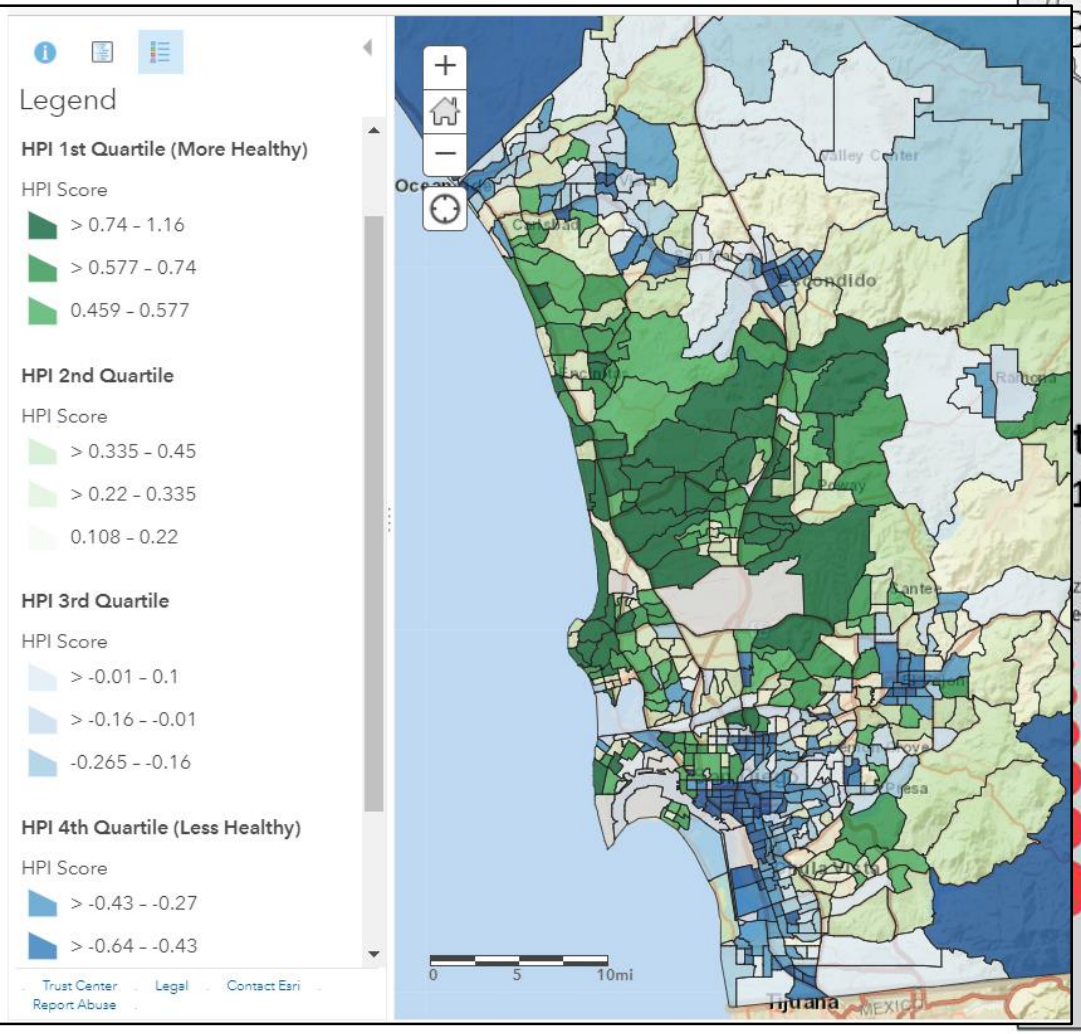
The Hot Spot (red) and Cold Spot (blue) Analysis of COVID-19 in San Diego (4/01/2020 – 4/31/2021)

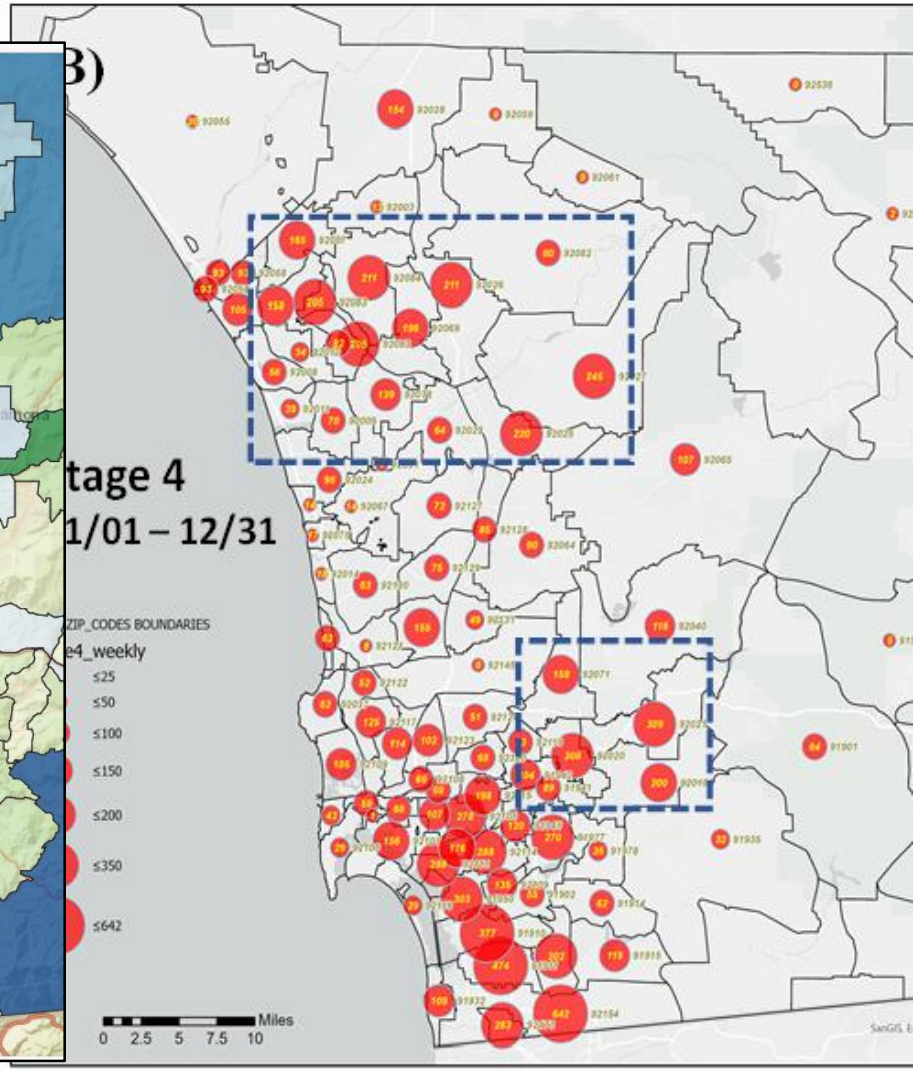


The Hot Spot (red) and Cold Spot (blue) Analysis of COVID-19 in San Diego during the five stages of outbreaks (using **Getis-Ord Gi* statistics** with ArcGIS Pro). The three levels of color shades indicate the p-Values with 99%, 95%, and 90% Confidence levels.

California's Healthy Places Index (HPI) San Diego, 2019

Low HPI value areas → High COVID-19 Rates (2020)
(Least Healthy, Poor Neighborhood)





Correlation between Social Factor and COVID-19 Cases at ZIP CODE level.

→ Social Determinants of Health

Table 2. The Pearson's correlation coefficient results (N=94) in the County of San Diego (94 Zipcode units) at Stages 1, 2, 3, 4 and 5.

Social Variables (Most are "per capita")	Stage1 coefficent	Stage2 coefficent	Stage3 coefficent	Stage4 coefficent	Stage5 coefficent	All Stages coefficent
Total_pop (with original data, not normalized).	0.328*	0.194	0.4***	0.389***	0.388***	0.394***
Population Density	0.335***	0.188	0.289	0.138	0.102	0.162
Ethnicity						
white	-0.147	-0.413***	-0.223*	-0.294**	-0.294**	-0.325**
asian	-0.064	-0.097	-0.092	-0.155	-0.155	-0.159
american_Indian	-0.1	0.489***	0.106	0.27**	0.27**	0.288**
Total_hispanic	**0.73*** **	**0.601*** **	**0.693*** **	**0.808*** **	**0.725*** **	**0.82*** **
other_race_one	0.403***	0.407***	0.425***	0.587***	0.587***	0.602***
non_hisp_white	**-0.602*** **	**-0.628*** **	**-0.607*** **	**-0.703*** **	**-0.703*** **	**-0.728*** **
non_hisp_black	0.386***	0.305**	0.308**	0.239*	0.239*	0.312**
Economic						
HouseholdwithCash (assistance)	**0.773*** **	**0.531*** **	**0.533*** **	**0.715*** **	**0.689*** **	**0.724*** **
professinal	**-0.428*** **	**-0.413*** **	**-0.275*** **	**-0.546*** **	**-0.546*** **	**-0.546*** **
Married	-0.452***	-0.419***	-0.398***	-0.213*	-0.295**	-0.319**
unemployedCvilian	0.479***	0.422***	0.523***	0.552***	0.45***	0.558***
popBelowpoverty	0.279**	0.235*	0.424***	0.321**	0.321**	0.345***
popUninsurance	**0.31****	**0.688*** **	**0.521*** **	**0.671*** **	**0.652*** **	**0.69*** **

Key SES Variables: Speaking Languages, Age, Education, Household size (Stage 4 and 5 only)

Social Variables (Most are "per capita")	Stage1 coefficent	Stage2 coefficent	Stage3 coefficent	Stage4 coefficent	Stage5 coefficent	All Stages coefficent
Speaking Language						
Pop5andOlderEnglish	**-0.672*** **	**-0.488*** **	**-0.597*** **	**-0.684*** **	**-0.684*** **	**-0.685*** **
Pop5andOlderSpanish	**0.782*** **	**0.559*** **	**0.684*** **	**0.776*** **	**0.776*** **	**0.789*** **
Age						
popAge15_24	**0.325** **	**0.145**	**0.014**	**-0.085**	**-0.085**	**0.006** **
popAge25_44	0.077	0.104	0.309**	0.155	0.155	0.173**
popAge65	-0.309**	-0.16	-0.254*	-0.219*	-0.219*	-0.25**
Education						
pop25OlderLess9	**0.64*** **	**0.545*** **	**0.633*** **	**0.782*** **	**0.782*** **	**0.784*** **
pop25Older9_12grade	0.43***	0.569***	0.503***	0.683***	0.683***	0.681***
pop25OlderBachelor	-0.495***	-0.418***	-0.289**	-0.525***	-0.525***	-0.533***
pop25OlderMaster	**-0.539*** **	**-0.514*** **	**-0.408*** **	**-0.64*** **	**-0.64*** **	**-0.655*** **
Other Factors						
medianIcome	-0.532***	-0.398***	-0.348***	-0.456***	-0.544***	-0.502***
foreignborn	0.436***	0.346***	0.435***	0.475***	0.475***	0.474***
ave_house_size	**0.17**	**0.254***	**0.342*** **	**0.594*** **	**0.531*** **	**0.528*** **

"*" means P-value < 0.05 "**" means P-value < 0.01 "***" means p-value < 0.001

Numbers without "*" means P-value > 0.05 (not significant).

HDMA @SDSU

Health Disparity in COVID-19 Testing Results. (accumulated on 2/27/2021) Per ZIP CODE

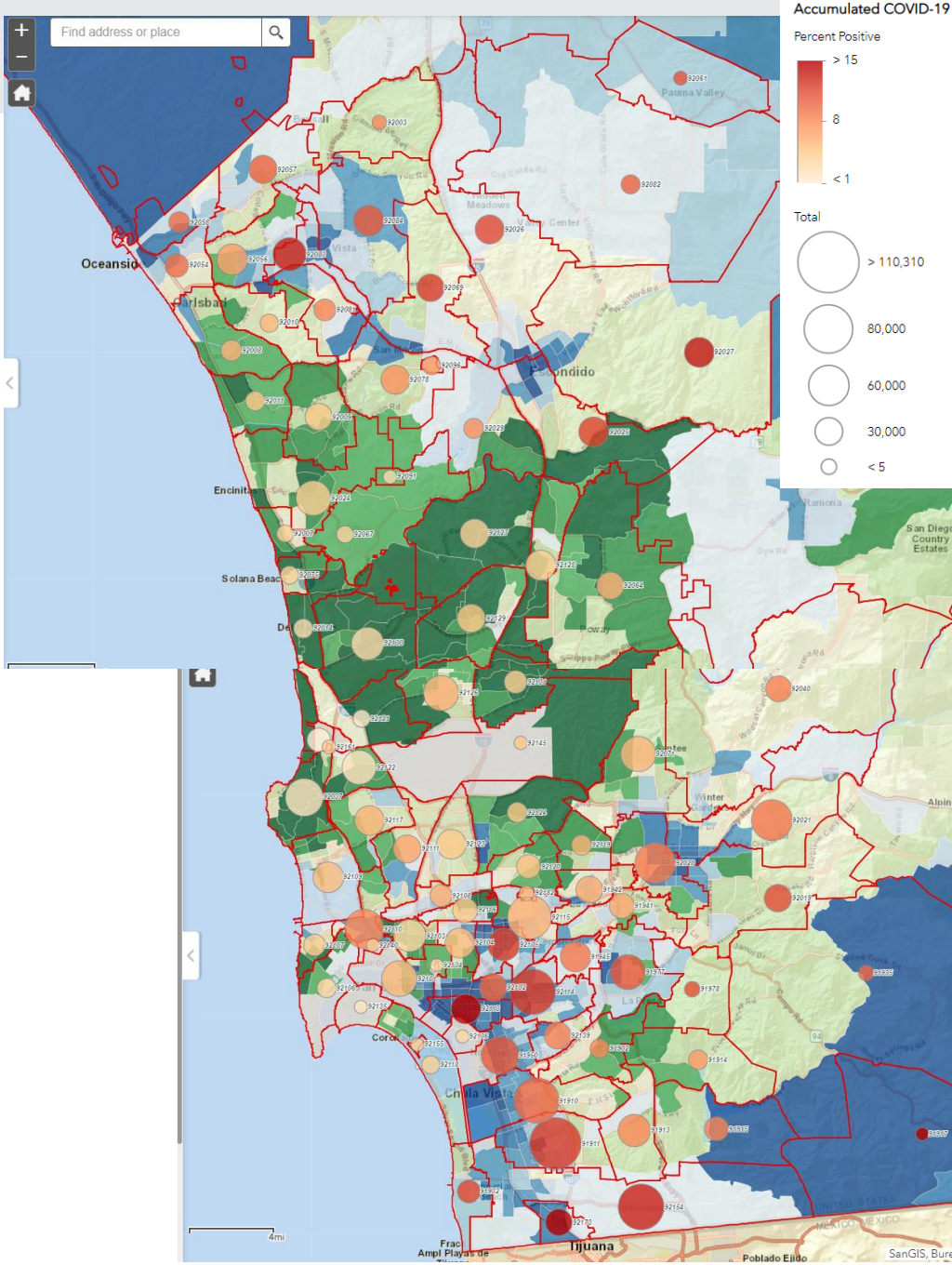
https://sdsugeo.maps.arcgis.com/apps/webappviewer/index.html?id=a24b1294ba88431aac1114e9e4600847

•Accumulated COVID-19 Test Results (on 2/27/2021)

•Overlay with HPI layers (green: healthier areas, blue: less healthy).

•Healthy Areas → Very low percentage of Positive cases (light circles)

• Low HPI Areas (unhealthy) → High percentage of positive cases. (dark circles) – **Need more testing sites!!!**

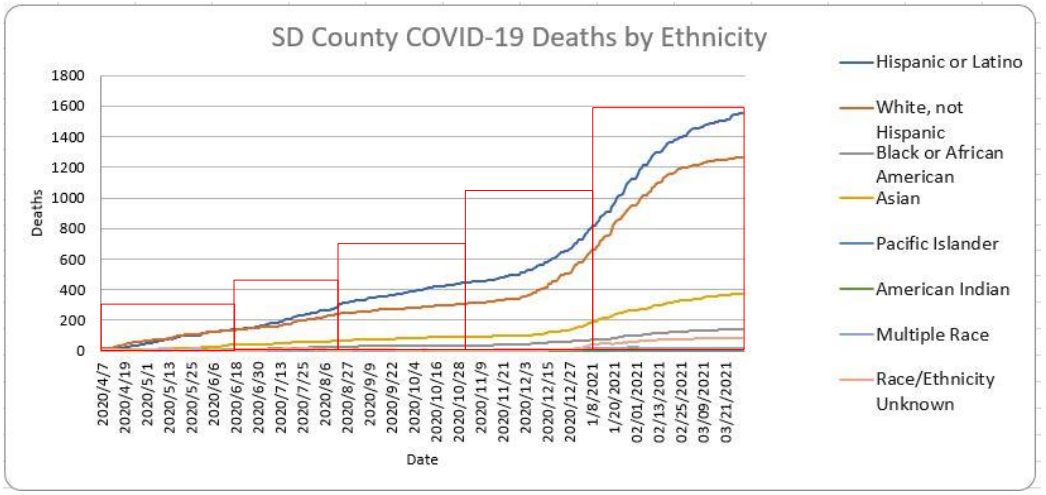


COVID-19 Death Cases (Mortality) by Ethnicity (March 2021)

COVID-19 Death Cases (Mortality) by Ethnicity **(Original Daily Death Numbers)**

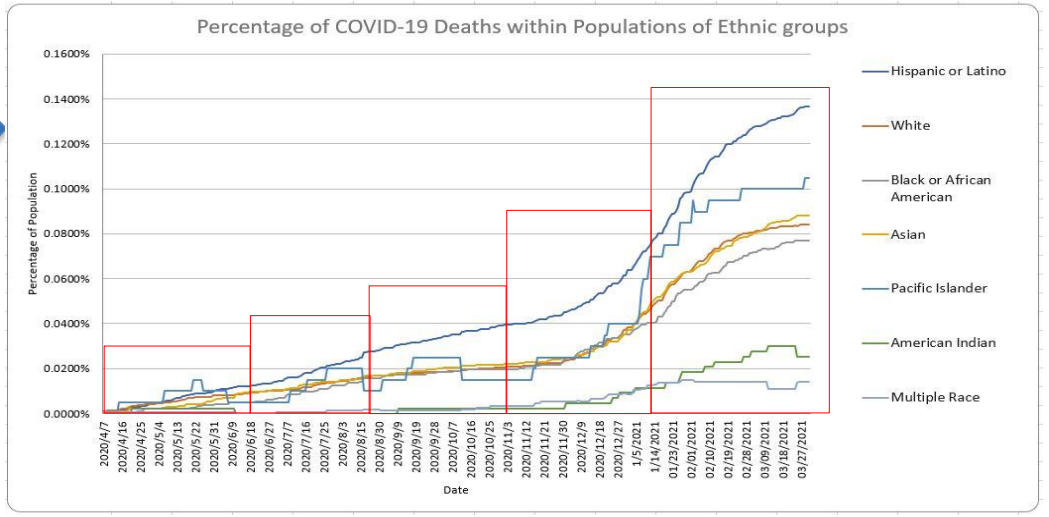
Why the White Ethnic Group has high death rate?

(More % of Elder Population in white ethnic group cases?)

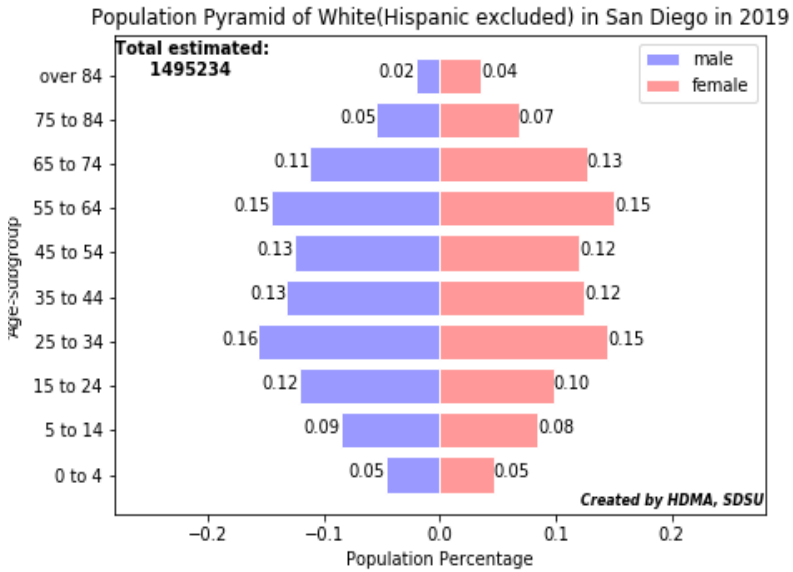


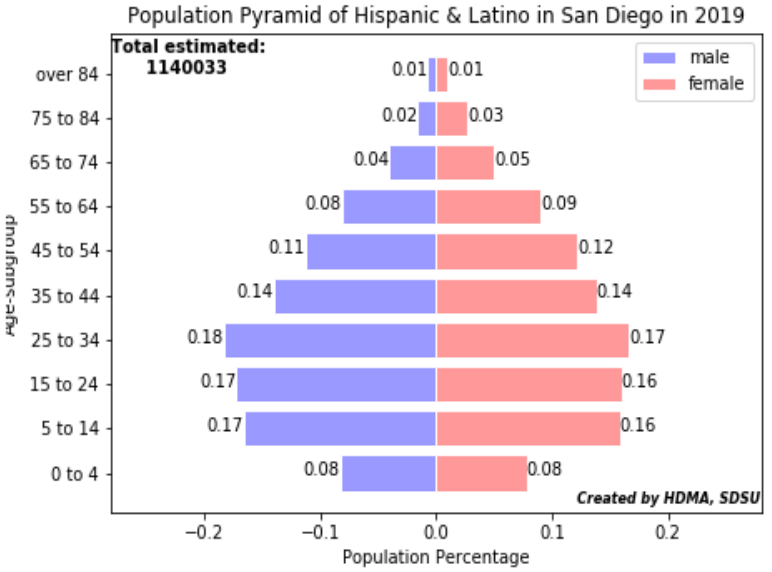
COVID-19 Death Cases (Mortality) by Ethnicity **(Normalized by each Ethnicity Group population)**

Hispanic > Pacific Islander> White > **Asian** > Black > Multiple Race > American Indian.



San Diego Population Pyramid (Whites vs. Hispanics)

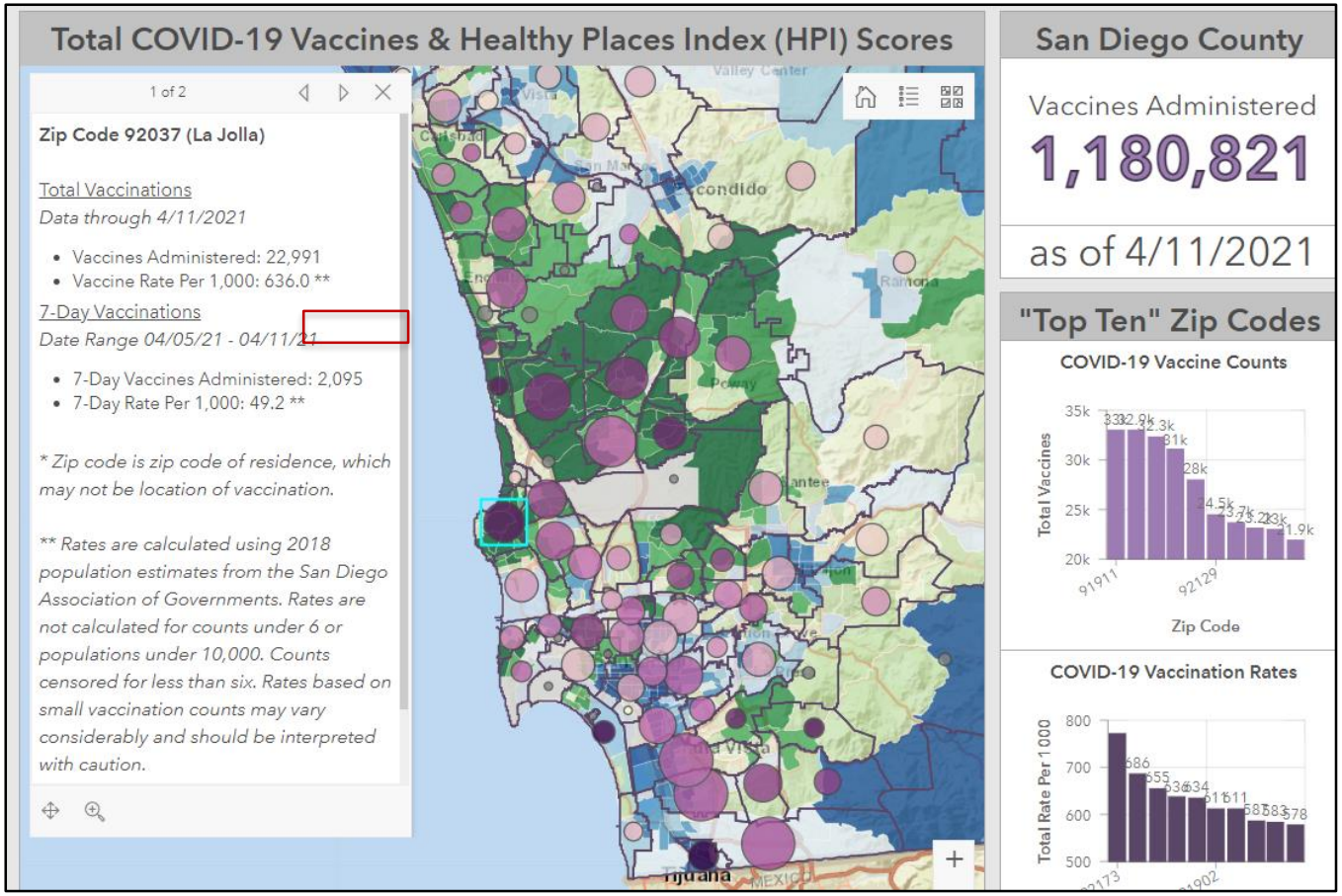




https://hdma-sdsu.github.io/sharing_data.html

Health Disparity in COVID-19 Vaccination Rates (Per Zip Code) (Date updated on 4/11/2021)

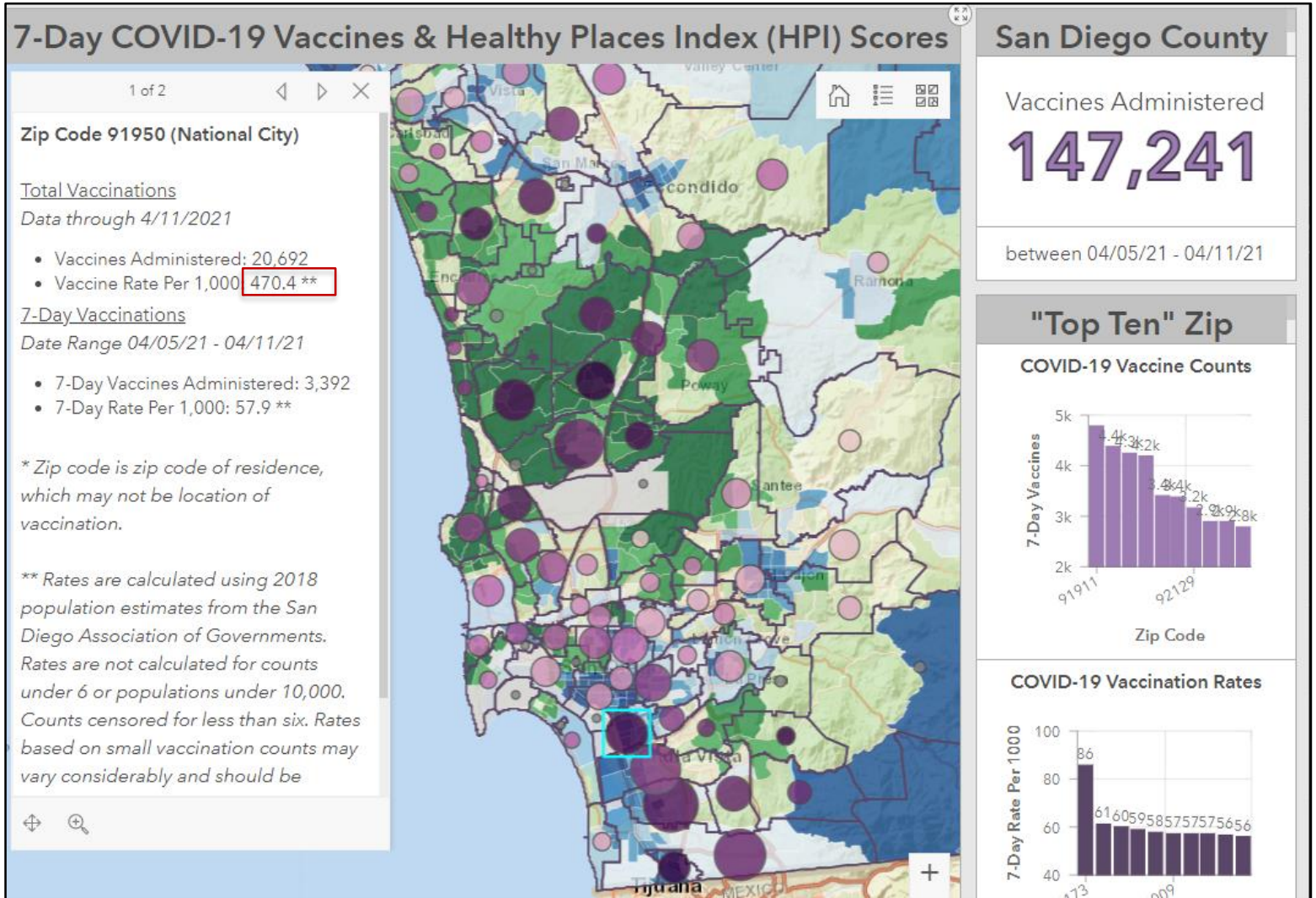
Accumulated Vaccination Rates are much higher in High HPI regions (Darker circles).
Rural and South bay areas have relative low percentage of Vaccination rate. (**Lighter circles**)
https://hdma-sdsu.github.io/COVID19%20Treatment.html



Wealthy communities have higher vaccination rates, poor communities have lower vaccination rates (but very high COVID-19 infected cases!!!)

LAST 7 Days of COVID-19 Vaccination Rates (Per Zip Code) (Date updated on 4/11/2021)

Last 7 Days of Vaccination Rates are still higher in High HPI regions (Carlsbad, Carmel Valley). **Rural and South bay areas** have better vaccination rate now. (Thank to the **SAVE project** effort).
https://hdma-sdsu.github.io/COVID19%20Treatment.html



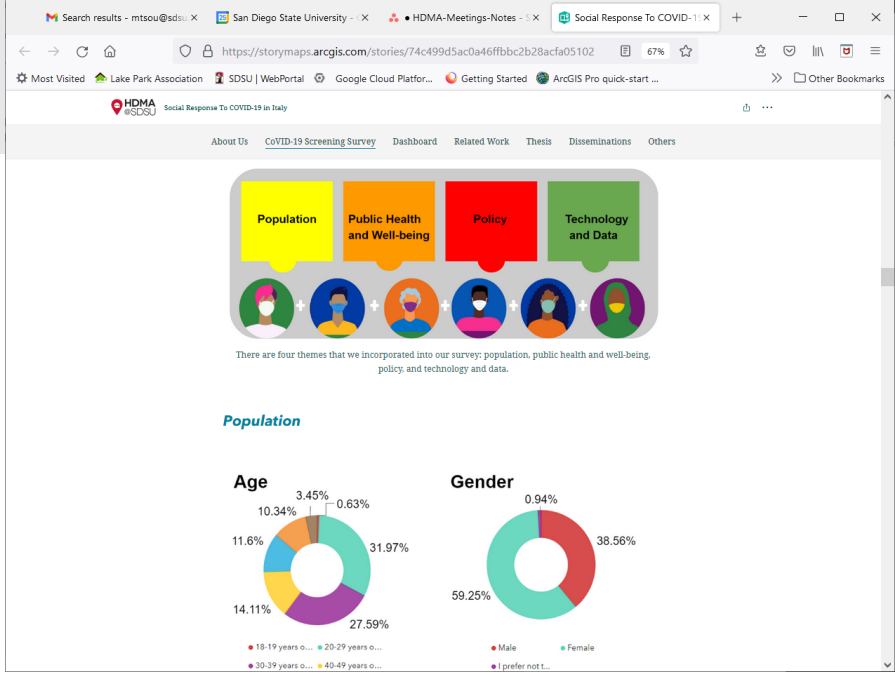
Project SAVE in San Diego

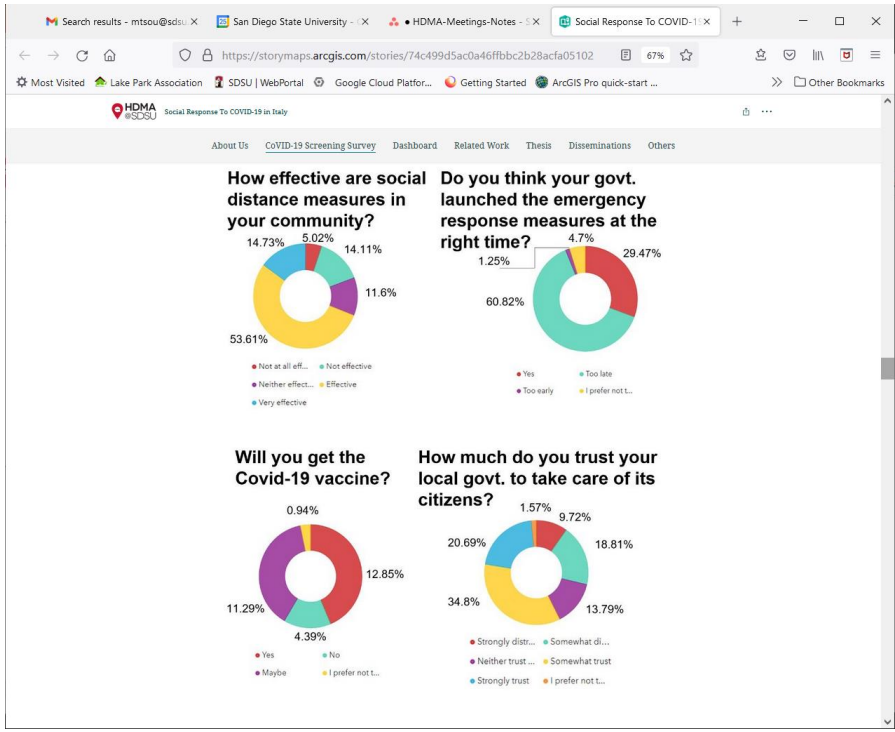
Project SAVE (Scheduling Assistance for Vaccine Equity) assists those facing technology, language, or other healthcare barriers. Working in partnership with local government entities and trusted partners, Project SAVE Community Health Workers (CHWs) and Promotores facilitate access to vaccine appointments and **provide referrals to no-appointment vaccination** events.
https://www.sandiegocounty.gov/content/sdc/hhsa/programs/phs/community_epidemiology/dc/2019-nCoV/vaccines/Project-SAVE.html

CoVID-19 Screening Survey

Track IT.CoVID-19 International Screening Survey

Track IT. Covid-19 Screening Tool @SDSU focuses on collecting data from a large-scale survey covering respondents from all over the world. The data will be collected via snowball sampling through a survey instrument. Data collection is ongoing and may be complemented with representative samples in the future. Track IT. Covid-19 Screening Tool is a screening tool initiative built by researchers at the **Center for Human Dynamics in the Mobile Age (HDMA)** at **San Diego State University** in San Diego, California, USA to provide a clearer picture of how people across the world experience the crisis caused by the coronavirus.





International Survey for COVID-19 Pandemic

https://storymaps.arcgis.com/stories/74c499d5ac0a46ffbbc2b28acfa05102

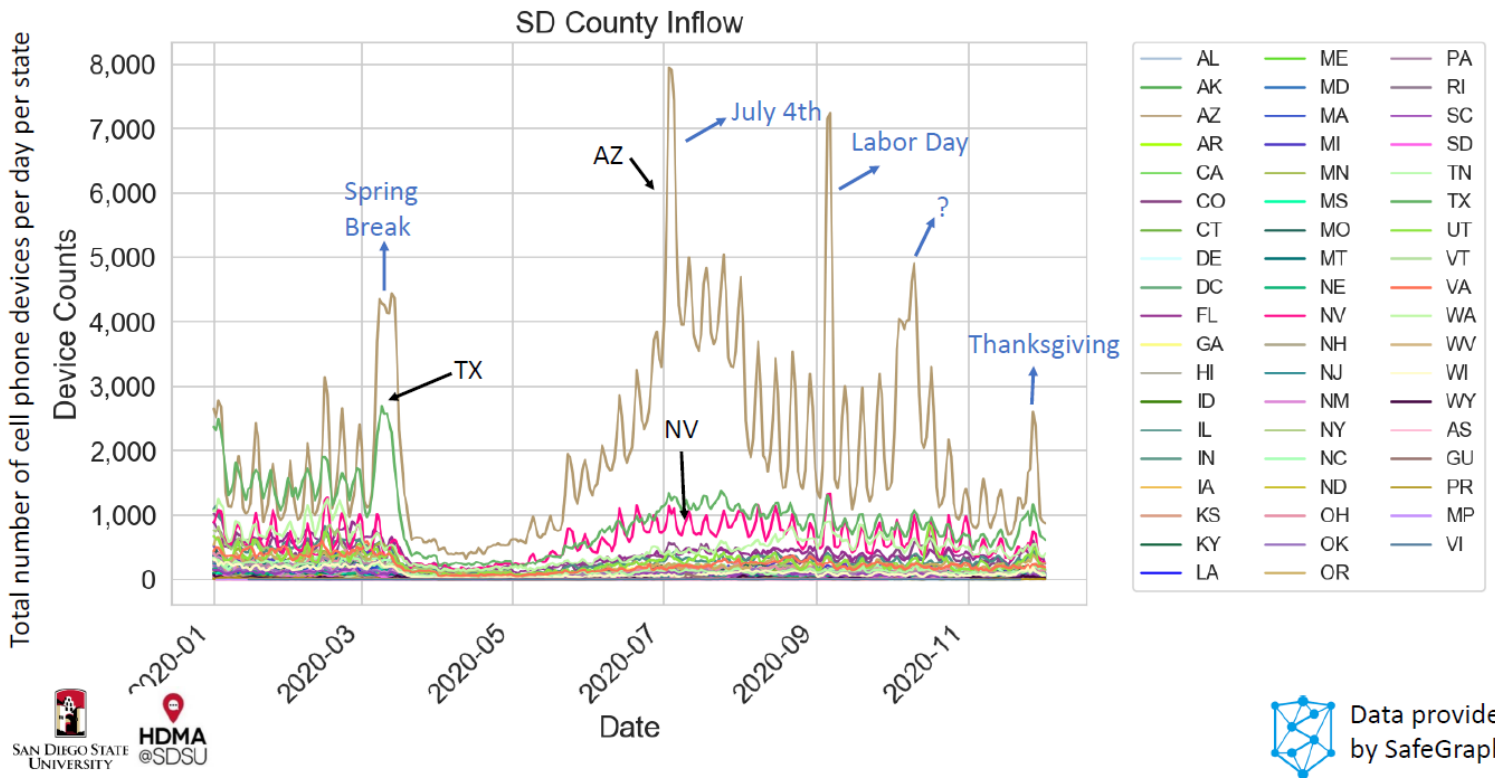
Human Mobility Analysis (by Dr. Atsushi Nara)

SafeGraph Social Distancing Metrics (from Smart phones and mobile devices).

- Count the number of mobile devices, Daily data at a **census block group (CBG)** level
- The device count is based on the number of devices originated from a home CBG to a destination CBG, Home is defined as the common nighttime (6pm-7am) location for the device over a 6-week period., •Data Range: January 1, 2020 –December 2, 2020

https://docs.safegraph.com/docs/social-distancing-metrics

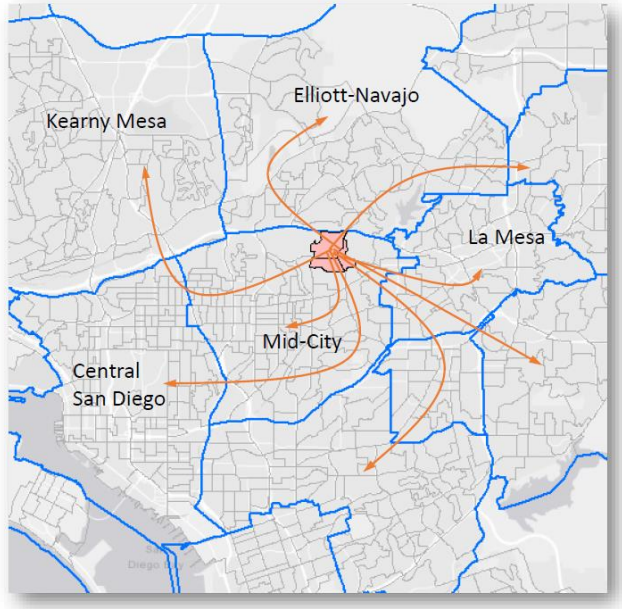
SD County Inflow (State-level aggregation)



HDMA @SDSU

SDSU student movements (Outflow):

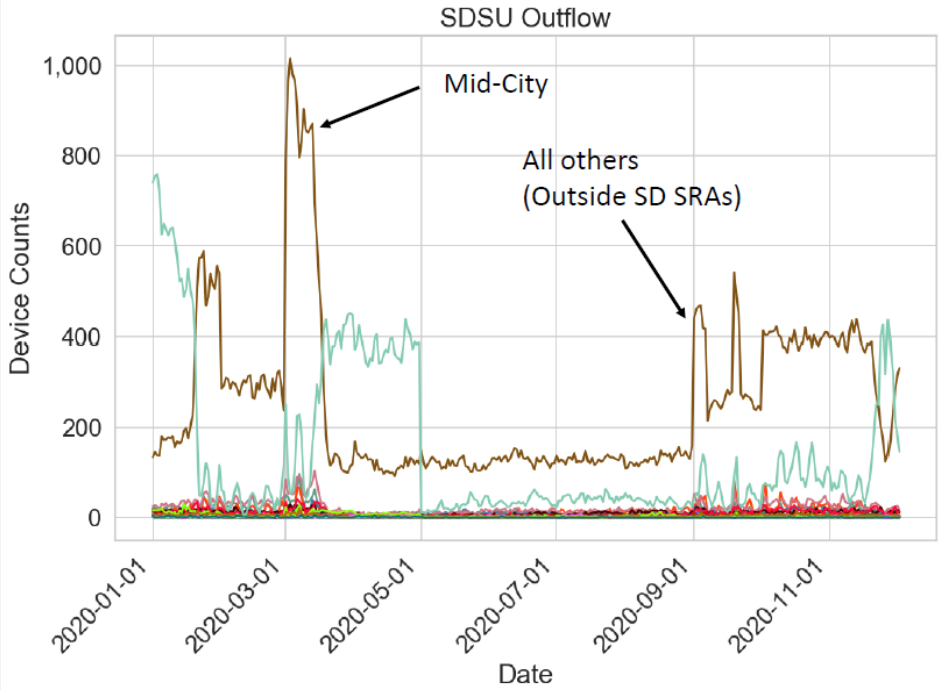
Decreased during the stay-at-home order. Increased after the Fall 2020 semester started.



San Diego State University HDMA @SDSU

Data provided by SafeGraph

SDSU Outflow (SRA-level aggregation)



San Diego State University HDMA @SDSU

Data provided by SafeGraph

Challenges and Opportunity

for Monitoring COVID-19 Outbreaks)

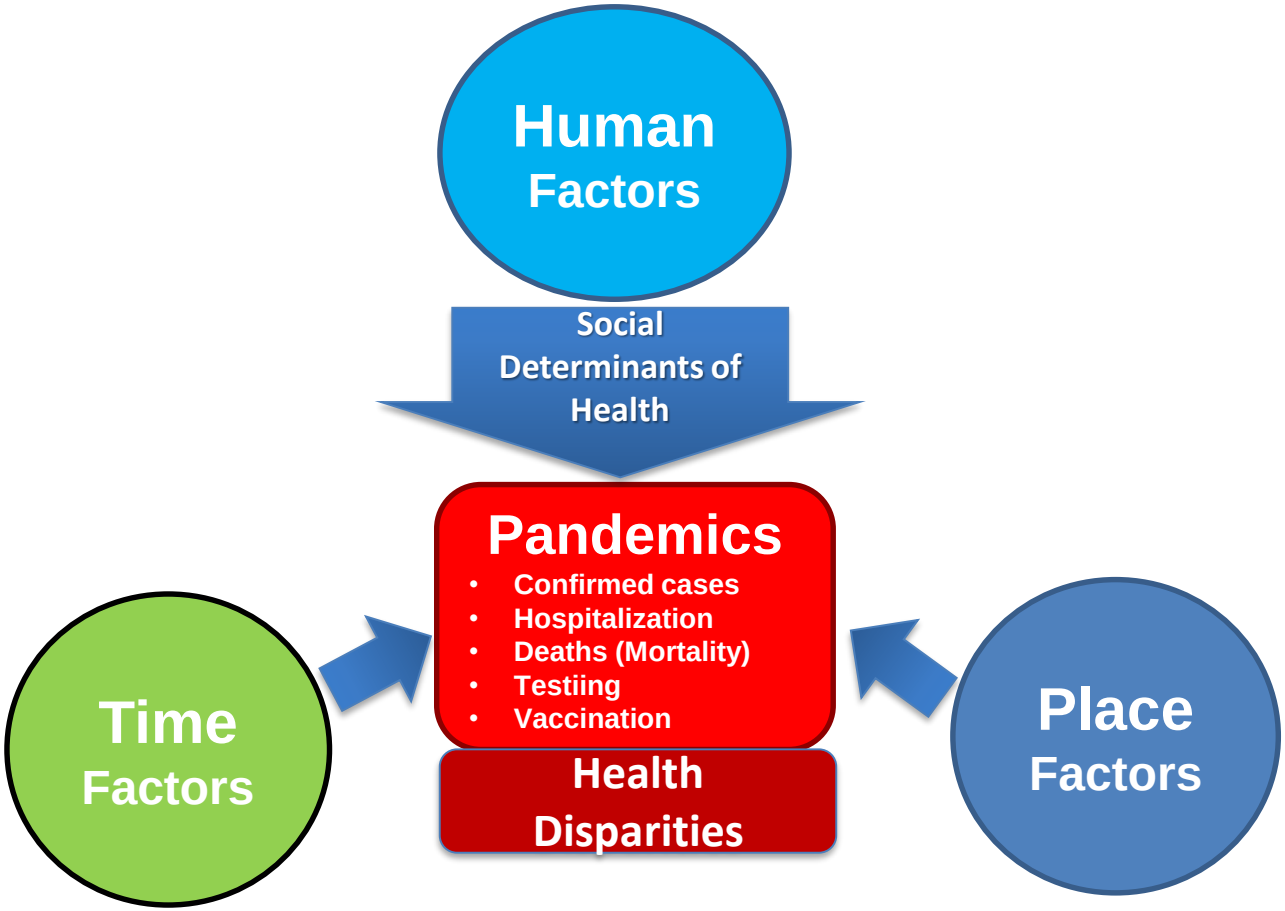
- We need to develop more **effective and easy-to-create mapping and visualization approaches.** (with innovative cartographic design and data visualization methods)…

- **Data Privacy vs Data Analytics** (We can only get the zipcode level COVID-19 data or SRA level, **not census tract** or census block group levels…). San Diego County can not provide **cluster outbreak** location information (but L.A. County did provide). Mortality data → no zipcode-level.

- Develop **new geographic knowledge or new models** for analyzing COVID-19 impacts for vulnerable groups (human mobility/dynamics models, disease outbreaks models, decision support systems, etc.). (Traditional models SEIR needs some improvements..)

Suggestions for Public Health Agencies in San Diego:

1. Put more medical resources on Testing sites and Vaccination sites **in the low HPI areas (South Bay, Rural Areas and Escondido).**
2. Add more supports to **Minority Groups (Black and Hispanic communities), low-income families, Spanish-Speaking families, less education groups.**
3. The Health Disparity Issue in San Diego is prevailing everywhere (not just in South Bay Areas) at **neighborhood** level.



https://hdma-sdsu.github.io

Thank you!

Q & A

Dr. Ming-Hsiang (Ming) Tsou, San Diego State University
mtsou@sdsu.edu
Twitter **@mingtsou**

Funded by
- SDSU **RADx UP Project: Community Fighting COVID. (NIH funding).**