COVID-19 VACCINATION STATUS
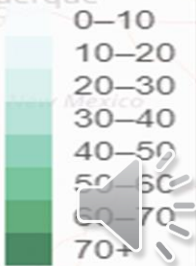
CALIFORNIA COUNTIES

Vaccination Status
Hover over a county

Project – 3
Data Visualization

Prepared by:
TIKRAM SUBEDY

0—10
10—20
20—30
30—40
40—50
50—60
60—70
70+

PURPOSE OF THE PROJECT:

Create an interactive visualization dashboard for database on Covid-19 vaccination in California counties created during ETL project using

- plots,
- data, and
- map

Data Sources, Collection and Processing:

- Data Sources:

  - California department of public health

    - https://data.chhs.ca.gov/dataset/vaccine-progress-dashboard

  - World Population Review

    - https://worldpopulationreview.com/us-counties/states/ca

  - Github:

    - https://github.com/codeforgermany/click_that_hood/blob/main/public/data/california-counties.geojson

Data Sources, Collection and Processing:

- Data Extraction:
  - Web scraping
    - Splinter
  - Downloading datafiles

```
1  # California Population data by county (web scrapping)
2  # setting up splinter
3
4  executable_path = {'executable_path': ChromeDriverManager().install()}
5  browser = Browser('chrome', **executable_path, headless=False)
6
7  url = 'https://worldpopulationreview.com/us-counties/states/ca'
8  browser.visit(url)
9
10 table = pd.read_html(url)
11 county_population = table[0]
12
13 # saving the datafile as a csv file
14 county_population.to_csv('./data/county_population2021_data.csv', index = False)
15 county_population.head()
16
```

```
====== WebDriver manager ======
Current google-chrome version is 94.0.4606
Get LATEST driver version for 94.0.4606
Get LATEST driver version for 94.0.4606
Trying to download new driver from https://chromedriver.storage.googleapis.com/94.0.4606.61/chromedriver_win32.zip
Driver has been saved in cache [C:\Users\tsube\.wdm\drivers\chromedriver\win32\94.0.4606.61]
```

	Name	2021 Population	Growth Since 2010	Density (mi²)
0	Los Angeles County	9969510	1.49%	13488.85
1	San Diego County	3347270	7.86%	4528.89
2	Orange County	3175130	5.31%	4295.98
3	Riverside County	2520060	14.47%	3409.67
4	San Bernardino County	2206750	8.13%	2985.76

Data Sources, Collection and Processing:

- Data Processing:
  - Python Pandas

  (fixed inconsistent county
  names among the datafiles)

```python
1  # cleaning county names (removed the last part containing "County" to make it identical with other datafiles for relation)
2  county = []
3  for i in county_name:
4      i = i[:-7]
5      county.append(i)
6
7  county_population = pd.DataFrame({'county': county, 'population': population, 'density': density})
8
9  # save the cleaned data as csv
10 county_population.to_csv('./data/county_population.csv', index = False)
11 county_population
```

	county	population	density
0	Los Angeles	9969510	13488.85
1	San Diego	3347270	4528.89
2	Orange	3175130	4295.98
3	Riverside	2520060	3409.67
4	San Bernardino	2206750	2985.76
5	Santa Clara	1918880	2596.26
6	Alameda	1680480	2273.71
7	Sacramento	1578680	2135.97

Data Sources, Collection and Processing:

Data Processing:

– Python Pandas

(unnecessary data rows and 'NA's removed)

```python
1   # Selecting only the residence county as reported in the data (elemination method is used)
2   ca_vaccin_data = csv_data.query('county !="All CA Counties"')
3   ca_vaccin_data = ca_vaccin_data.query('county !="All CA and Non-CA Counties"')
4   ca_vaccin_data = ca_vaccin_data.query('county !="Outside California"')
5   ca_vaccin_data = ca_vaccin_data.query('county !="Unknown"')
6   ca_vaccin_data = ca_vaccin_data.query('county !="California"')
7   # getting new dataframe for total vaccination (fully, partially vaccinated) by counties
8
9   county_vac_data = pd.DataFrame(ca_vaccin_data, columns = ['county', 'administered_date',\
10                                            'partially_vaccinated', 'fully_vaccinated'])
11
12  county_vac_data['month'] = pd.DatetimeIndex(county_vac_data['administered_date']).month_name()
13
14  # Creating a new dataframe with months as variable to be utilized for ploting by months
15  county_vac_data_cleaned = pd.DataFrame(county_vac_data, columns=['county', 'month','administered_date','fully_vaccinated','p
16
17  county_vac_data_cleaned.to_csv('./data/county_vac_data_cleaned.csv', index = False)
18  county_vac_data_cleaned
```
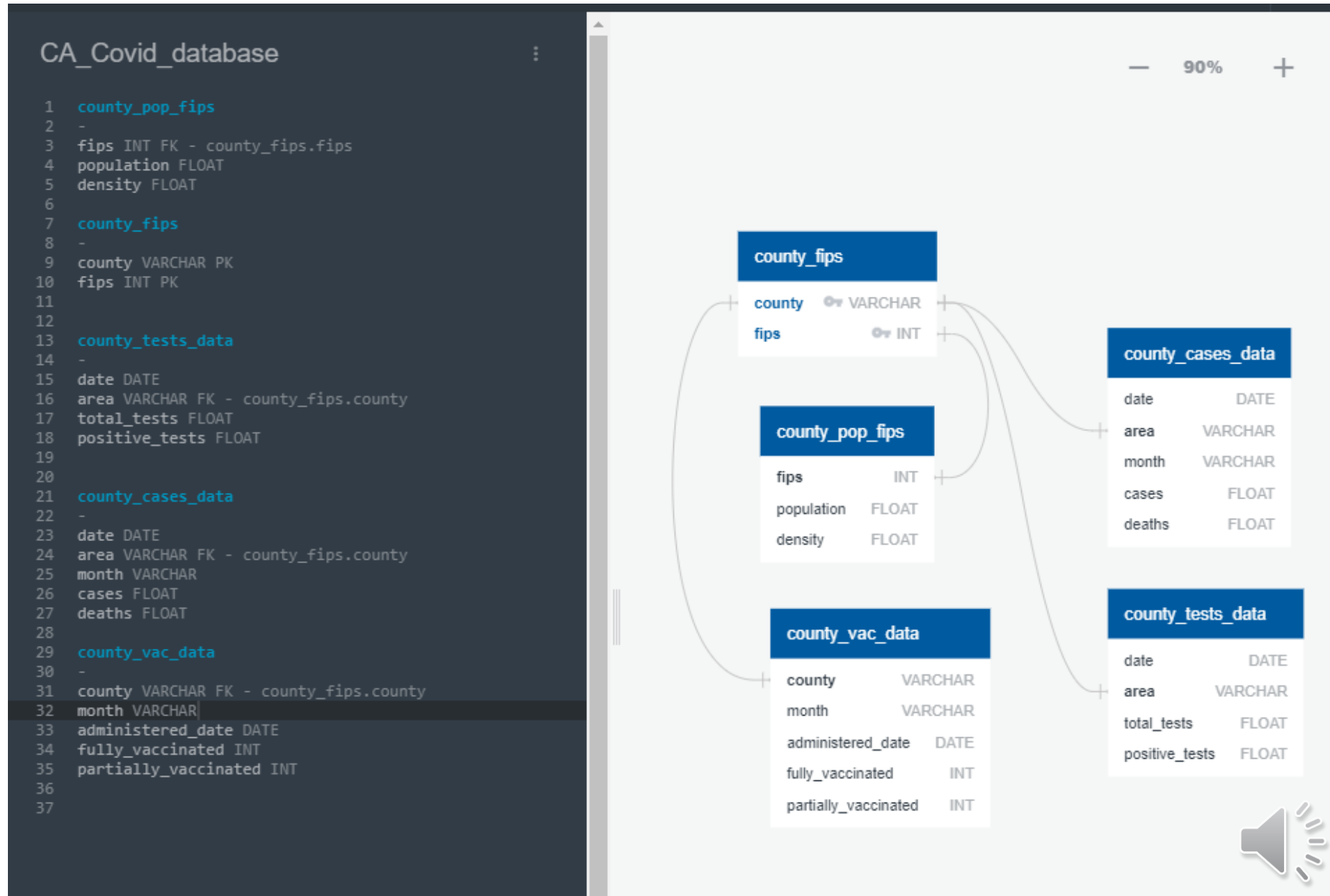
	county	month	administered_date	fully_vaccinated	partially_vaccinated
0	Alameda	July	2020-07-27	0	0
1	Alameda	July	2020-07-28	0	0
2	Alameda	July	2020-07-29	0	0
3	Alameda	July	2020-07-30	0	0
4	Alameda	July	2020-07-31	0	0
...	...	...	...	...	...
27337	Yuba	October	2021-10-07	78	45
27338	Yuba	October	2021-10-08	88	63

Database and Data Query:

- ## ER Diagram:

  (Established relation with primary and foreign keys)

Database and Data Query:

SQL Queries:

Postgres,
PgAdmin

Database and Data Query:

Loading Database:

SQLAlchemy
(selecting SQL views to obtain
the final database)

```python
1  from sqlalchemy import create_engine
2  engine = create_engine(f'postgresql://{username}:{password}@localhost:5432/county_covid_database')
3  connection = engine.connect()
4
```

```python
# Loading combined dataset (using sql view 'county_wise_combined_data')

combined_data = pd.read_sql("SELECT * from county_wise_combined_data", connection)
combined_data.to_csv('./data/combined_data.csv', index = False)
combined_data.head()
```

	county	population	tests	positive	cases	deaths	total_full	total_partial	percentFullVax	percentCases	percentDeaths
0	Alameda	1680480.0	3590879.0	130133.0	107950.0	1230.0	1151524	1128245	68.52	6.42	0.07
1	Alpine	1209.0	2151.0	46.0	98.0	0.0	705	790	58.31	8.11	0.00
2	Amador	40446.0	149585.0	5485.0	4745.0	58.0	17826	19639	44.07	11.73	0.14
3	Butte	196880.0	289588.0	18739.0	17542.0	228.0	96451	100901	48.99	8.91	0.12
4	Calaveras	46319.0	55933.0	3768.0	3270.0	65.0	20647	21932	44.58	7.06	0.14
5	Colusa	21805.0	23090.0	2199.0	2343.0	16.0	10478	11059	48.05	10.75	0.07
6	Contra Costa	1159540.0	2141866.0	102018.0	89960.0	882.0	804166	805267	69.35	7.76	0.08
7	Del Norte	27956.0	127754.0	3646.0	3210.0	29.0	10578	11220	37.84	11.48	0.10
8	El Dorado	197037.0	245215.0	14841.0	14052.0	124.0	100822	104156	51.17	7.13	0.06
9	Fresno	1013400.0	1605085.0	144932.0	121531.0	1857.0	477528	517126	47.12	11.99	0.18
10	Glenn	29245.0	31724.0	3152.0	3038.0	24.0	12613	12727	43.13	10.39	0.08
11	Humboldt	134186.0	193795.0	8638.0	7500.0	50.0	75391	76411	56.18	5.59	0.04
12	Imperial	180500.0	275217.0	25266.0	20027.0	740.0	118000	128800	65.20	16.62	0.41

Database and Data Query:

Loading Database:

Converting into Json data file

```
let data = [{"county":"Alameda","population":1680480.0,"tests":3590879.0,"positive":130133.0,"cases":107950.0,"deaths":1230.0,
"total_full":1151524,"total_partial":1128245,"percentFullVax":68.52,"percentCases":6.42,"percentDeaths":0.07},{"county":"Alpine",
"population":1209.0,"tests":2151.0,"positive":46.0,"cases":98.0,"deaths":0.0,"total_full":705,"total_partial":790,"percentFullVax":58.31,
"percentCases":8.11,"percentDeaths":0.0},{"county":"Amador","population":40446.0,"tests":149585.0,"positive":5485.0,"cases":4745.0,"deaths":58.0,
"total_full":17826,"total_partial":19639,"percentFullVax":44.07,"percentCases":11.73,"percentDeaths":0.14},{"county":"Butte","population":196880.
0,"tests":289588.0,"positive":18739.0,"cases":17542.0,"deaths":228.0,"total_full":96451,"total_partial":100901,"percentFullVax":48.99,
"percentCases":8.91,"percentDeaths":0.12},{"county":"Calaveras","population":46319.0,"tests":55933.0,"positive":3768.0,"cases":3270.0,
"deaths":65.0,"total_full":20647,"total_partial":21932,"percentFullVax":44.58,"percentCases":7.06,"percentDeaths":0.14},{"county":"Colusa",
"population":21805.0,"tests":23090.0,"positive":2199.0,"cases":2343.0,"deaths":16.0,"total_full":10478,"total_partial":11059,"percentFullVax":48.
05,"percentCases":10.75,"percentDeaths":0.07},{"county":"Contra Costa","population":1159540.0,"tests":2141866.0,"positive":102018.0,
"cases":89960.0,"deaths":882.0,"total_full":804166,"total_partial":805267,"percentFullVax":69.35,"percentCases":7.76,"percentDeaths":0.08},
{"county":"Del Norte","population":27956.0,"tests":127754.0,"positive":3646.0,"cases":3210.0,"deaths":29.0,"total_full":10578,
"total_partial":11220,"percentFullVax":37.84,"percentCases":11.48,"percentDeaths":0.1},{"county":"El Dorado","population":197037.0,
"tests":245215.0,"positive":14841.0,"cases":14052.0,"deaths":124.0,"total_full":100822,"total_partial":104156,"percentFullVax":51.17,
"percentCases":7.13,"percentDeaths":0.06},{"county":"Fresno","population":1013400.0,"tests":1605085.0,"positive":144932.0,"cases":121531.0,
"deaths":1857.0,"total_full":477528,"total_partial":517126,"percentFullVax":47.12,"percentCases":11.99,"percentDeaths":0.18},{"county":"Glenn",
"population":29245.0,"tests":31724.0,"positive":3152.0,"cases":3038.0,"deaths":24.0,"total_full":12613,"total_partial":12727,"percentFullVax":43.
13,"percentCases":10.39,"percentDeaths":0.08},{"county":"Humboldt","population":134186.0,"tests":193795.0,"positive":8638.0,"cases":7500.0,
"deaths":50.0,"total_full":75391,"total_partial":76411,"percentFullVax":56.18,"percentCases":5.59,"percentDeaths":0.04},{"county":"Imperial",
"population":180599.0,"tests":375217.0,"positive":35366.0,"cases":30027.0,"deaths":740.0,"total_full":118090,"total_partial":128800,
"percentFullVax":65.39,"percentCases":16.63,"percentDeaths":0.41},{"county":"Inyo","population":18225.0,"tests":21922.0,"positive":1355.0,
"cases":1597.0,"deaths":38.0,"total_full":8992,"total_partial":9337,"percentFullVax":49.34,"percentCases":8.76,"percentDeaths":0.21},
{"county":"Kern","population":913090.0,"tests":1396151.0,"positive":128279.0,"cases":111923.0,"deaths":1421.0,"total_full":366141,
"total_partial":395875,"percentFullVax":40.1,"percentCases":12.26,"percentDeaths":0.16},{"county":"Kings","population":156056.0,"tests":462446.0,
"positive":32161.0,"cases":27628.0,"deaths":280.0,"total_full":51145,"total_partial":56331,"percentFullVax":32.77,"percentCases":17.7,
"percentDeaths":0.18},{"county":"Lake","population":64524.0,"tests":92463.0,"positive":6699.0,"cases":5609.0,"deaths":84.0,"total_full":29859,
"total_partial":31361,"percentFullVax":46.28,"percentCases":8.69,"percentDeaths":0.13},{"county":"Lassen","population":30483.0,"tests":179553.0,
"positive":7171.0,"cases":6050.0,"deaths":23.0,"total_full":6732,"total_partial":6468,"percentFullVax":22.08,"percentCases":19.85,
"percentDeaths":0.08},{"county":"Los Angeles","population":9969510.0,"tests":29303282.0,"positive":1702841.0,"cases":1324628.0,"deaths":24128.0,
"total_full":6010757,"total_partial":6285936,"percentFullVax":60.29,"percentCases":13.29,"percentDeaths":0.24},{"county":"Madera",
"population":158217.0,"tests":333610.0,"positive":22335.0,"cases":19641.0,"deaths":252.0,"total_full":66615,"total_partial":70982,
"percentFullVax":42.1,"percentCases":12.41,"percentDeaths":0.16},{"county":"Marin","population":257154.0,"tests":638851.0,"positive":18076.0,
"cases":16093.0,"deaths":226.0,"total_full":198248,"total_partial":199579,"percentFullVax":77.09,"percentCases":6.26,"percentDeaths":0.09},
```

Database and Data Query:

Loading Database:

Converting into Json data file

```python
1  import json
2  with open("./data/countyGeoJson.js") as f:
3      geodata = json.load(f)
4  with open("./data/combined_data.js") as f:
5      data = json.load(f)
```

```python
1  for x in data:
2      for y in geodata["features"]:
3          if y["properties"]["name"] == x["county"]:
4              y["properties"]["data"] = x
5  countyGeoData = geodata
6  countyGeoData
```

```
{'type': 'FeatureCollection',
 'features': [{'type': 'Feature',
   'properties': {'name': 'Alameda',
    'cartodb_id': 1,
    'created_at': '2015-07-04T21:04:58Z',
    'updated_at': '2015-07-04T21:04:58Z',
    'data': {'county': 'Alameda',
     'population': 1680480.0,
     'tests': 3590879.0,
     'positive': 130133.0,
     'cases': 107950.0,
     'deaths': 1230.0,
     'total_full': 1151524,
     'total_partial': 1128245,
     'percentFullVax': 68.52,
     'percentCases': 6.42,
     'percentDeaths': 0.07}},
   'geometry': {'type': 'MultiPolygon',
    'coordinates': [[[[-122.312934, 37.897333],
```

```python
1  with open('countyGeoData.js', 'w') as fp:
2      json.dump(countyGeoData, fp, sort_keys=False, indent=4)
3
```

Data Visualization:
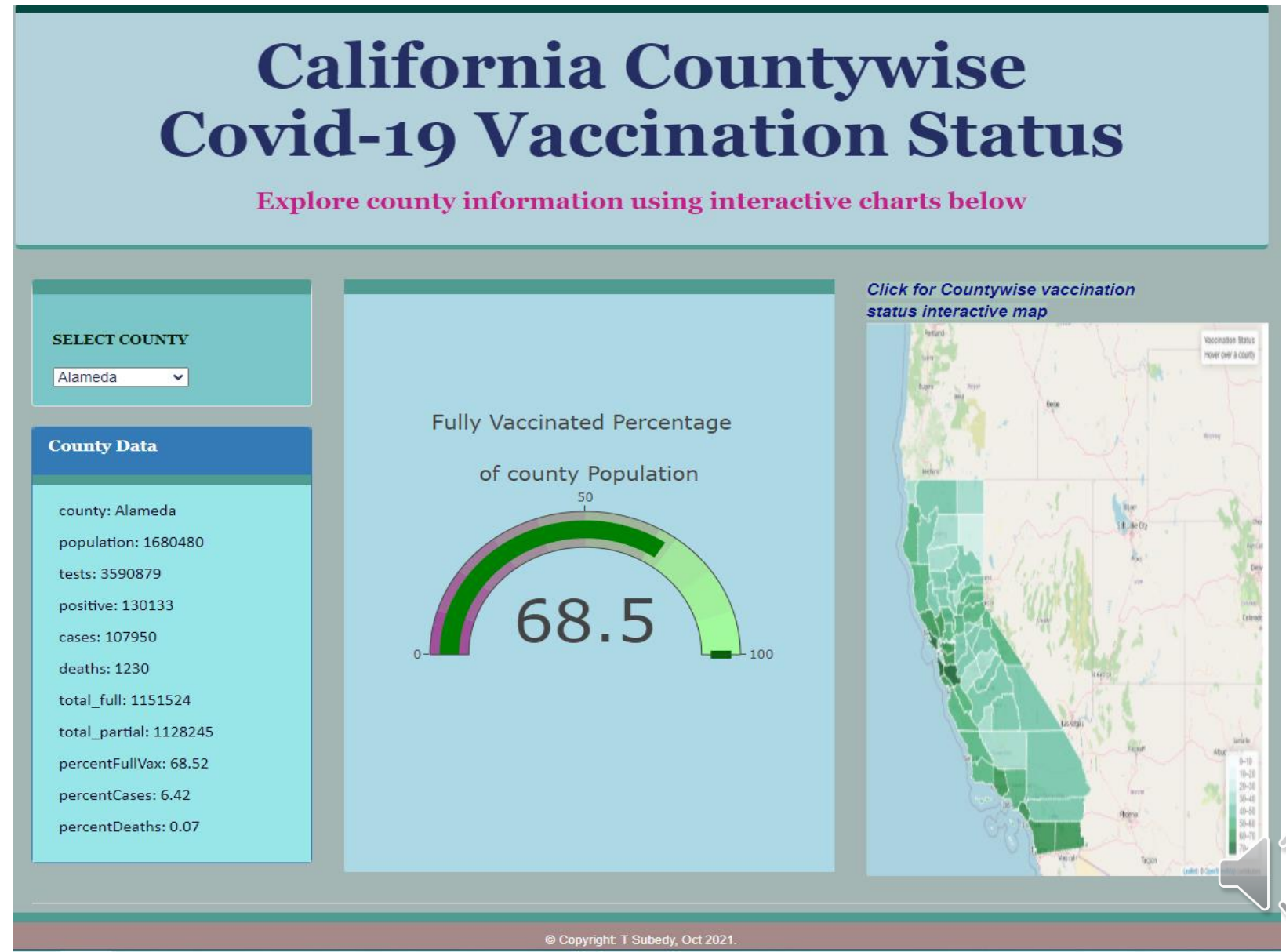
- Javascript,
- Plotly,
- Leaflets
- HTML, CSS

Dashboard and Data Visualization:

- Dashboard:
  - Javascript
  - Plotly
  - Html/ CSS
  - Mapping
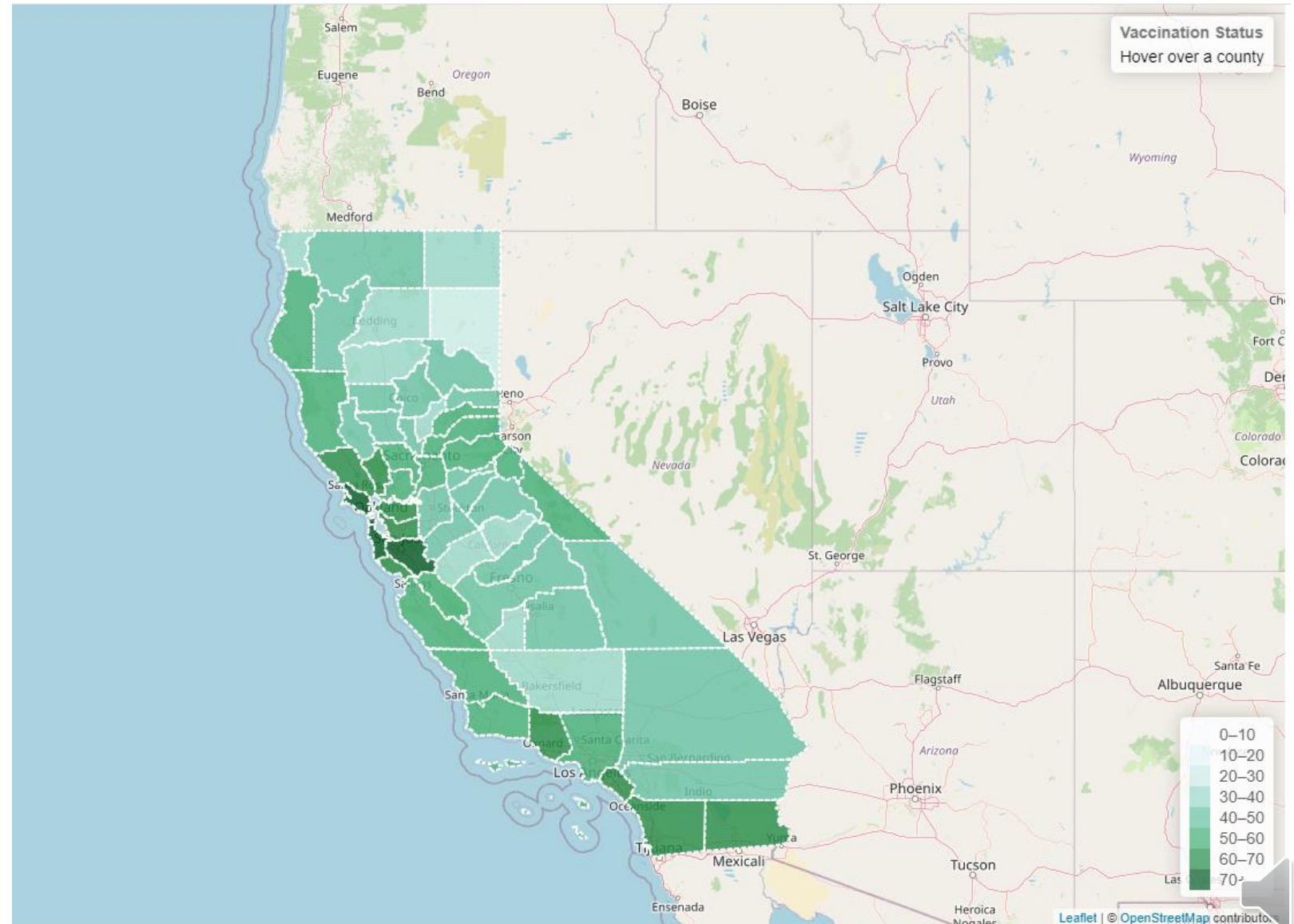 - Leaflets

https://tsubedy.github.io/Project_3/

Dashboard and Data Visualization:

Dashboard:

Map using Leaflets

- Displays data on hoovering over the map and clicking on counties

- Legend colors are based data values)

https://tsubedy.github.io/Project_3/

Dashboard and Data Visualization:

Dashboard:

Interactive bar chart (with legends on/off)

https://tsubedy.github.io/Project_3/

Limitations:

- The project is a part of the assignments from the class of Data Analytics Bootcamp and is limited to demonstrate the technical skills learned so far in the class.
- Some of data used for this project are not up to date as they were downloaded in the form of csv files directly from the source sites.
- Data is limited to 58 California Counties.
- The visualizations show only the status of the California county vaccination.
- It is not providing any statistical tests or analyses of the data.

California Countywise Covid-19 Vaccination Status

Explore county information using interactive charts below

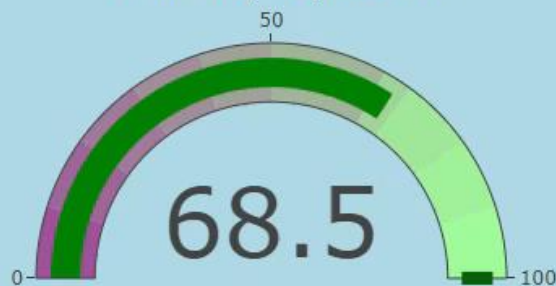**Click for Countywise vaccination status interactive map**
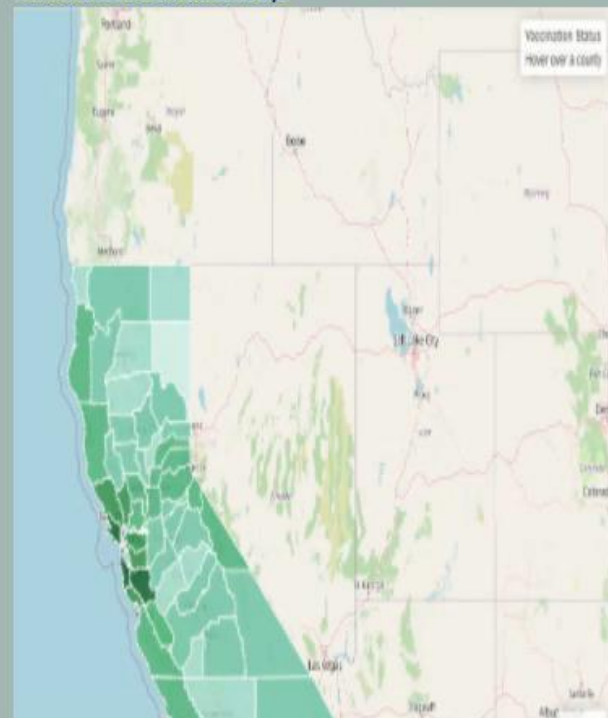
SELECT COUNTY

Alameda

County Data

county: Alameda

population: 1680480

tests: 3590879

positive: 130133

cases: 107950

deaths: 1230

total_full: 1151524

total_partial: 1128245

Fully Vaccinated Percentage

of county Population

50

0

68.5

100

Vaccination Status
Hover over a county

Questions:

? ? ?

THANK YOU !!!