COVID-19 Study

Questions

❖   What states have the highest COVID related deaths and do these high death rates have any relationship to the social characteristics apparent across the different states ?

❖   Is there a relationship between the education level of a state and the number of reported cases of COVID for that state?

Data Set Selection

	Province_State	Country_Region	Last_Update	Lat	Long_	Confirmed	Deaths	Recovered	Active	FIPS	Incident_Rate	Total_Test_Results
1												
2	Alabama	US	2021-01-02 05:30:44	32.3182	-86.9023	365747	4872	202137.0	158738.0	1.0	7459.375895463867	1885216.0
3	Alaska	US	2021-01-02 05:30:44	61.3707	-152.4044	47019	206	7165.0	39615.0	2.0	6422.844801071706	1275750.0
4	American Samoa	US	2021-01-02 05:30:44	-14.271	-170.132	0	0		0.0	60.0	0.0	2140.0
5	Arizona	US	2021-01-02 05:30:44	33.7298	-111.4312	530267	9015	76934.0	444318.0	4.0	7285.171274003372	5155330.0
6	Arkansas	US	2021-01-02 05:30:44	34.9697	-92.3731	229442	3711	199247.0	26484.0	5.0	7602.945718144717	2051488.0
7	California	US	2021-01-02 05:30:44	36.1162	-119.6816	2365024	26504		2309915.0	6.0	5912.603297465698	33058311.0
8	Colorado	US	2021-01-02 05:30:44	39.0598	-105.3111	337161	4873	18102.0	314186.0	8.0	5854.774381044729	4444206.0

Primary:
COVID-19 Cases

Secondary:
American Community Survey - Social Characteristics

Label	Estimate	Margin of Error	Percent	Percent Margin of Error
⌄ HOUSEHOLDS BY TYPE				
⌄ Total households	1,867,893	±5,799	1,867,893	(X)
⌄ Married-couple family	883,521	±5,502	47.3%	±0.2
With own children of the householder under 18 years	313,139	±4,409	16.8%	±0.2
⌄ Cohabiting couple household	78,846	±1,841	4.2%	±0.1
With own children of the householder under 18 years	27,962	±1,355	1.5%	±0.1
⌄ Male householder, no spouse/partner present	331,361	±3,369	17.7%	±0.2
With own children of the householder under 18 years	21,350	±1,103	1.1%	±0.1
⌄ Householder living alone	244,446	±3,161	13.1%	±0.2
65 years and over	66,999	±1,614	3.6%	±0.1
⌄ Female householder, no spouse/partner present	574,165	±4,264	30.7%	±0.2
With own children of the householder under 18 years	119,948	±2,621	6.4%	±0.1
⌄ Householder living alone	311,963	±3,850	16.7%	±0.2
65 years and over	151,059	±2,336	8.1%	±0.1
Households with one or more people under 18 years	560,887	±5,198	30.0%	±0.2
Households with one or more people 65 years and over	565,454	±2,531	30.3%	±0.1

Staging and Modeled Tables

```
In [11]:  dataset_id = "fda_staging"

In [45]:  !bq --location=US mk --dataset {dataset_id}

          BigQuery error in mk operation: Dataset 'still-primer-302701:fda_staging'
          already exists.

In [13]:  !bq --location=US load --autodetect --skip_leading_rows=1 --allow_jagged_rows=true --source_format=CSV {dataset_id}.
          US_COVID_Cases \
          gs://covid_milestone/us_covid/*.csv

          Waiting on bqjob_r13c8987ede88be60_00000178e300bfe5_1 ... (1s) Current status: DONE

In [2]:   dataset_id = "uscb_staging"

In [3]:   bucket_name = "covid_milestone"

In [4]:   !bq --location=US mk --dataset {dataset_id}

          BigQuery error in mk operation: Dataset 'still-primer-302701:uscb_staging'
          already exists.

In [9]:   !bq --location=US load --autodetect --skip_leading_rows=2 --source_format=CSV {dataset_id}.Social_Characteristics \
          gs://{bucket_name}/state_char/*.csv

          Waiting on bqjob_r5471224cd7e4b7ea_00000178e2fa88ae_1 ... (4s) Current status: DONE
```

Staging and Modeled Tables

```
In [3]:  %%bigquery
         CREATE TABLE datamart.States AS
         select  DISTINCT u.Geographic_Area_Name as state, u.id as id, 'US' as country from uscb_staging.Social_Characteristi
         cs u
         join fda_staging.US_COVID_Cases c on c.Province_State = u.Geographic_Area_Name;

Out[3]:  □
```

```
In [4]:  %%bigquery
         CREATE TABLE datamart.Cases AS
         SELECT Province_State as state,
         CAST(Last_Update as DATE) as last_Update,
         concat(Province_State, cast(last_Update as string)) as id,
         Total_Test_Results as total_test_results,
         Confirmed as confirmed, Deaths as deaths,
         Recovered as recovered, Active as active,
         Case_Fatality_Ratio as case_fatality_ratio,
         Testing_Rate as testing_rate
         FROM fda_staging.US_COVID_Cases;

Out[4]:  □
```

```
In [5]:  %%bigquery
         CREATE TABLE datamart.Location AS
         SELECT DISTINCT Province_State as state, Lat as latitude, Long_ as longitude FROM fda_staging.US_COVID_Cases;

Out[5]:  □
```

Beam Pipelines

```python
import logging
import apache_beam as beam
from apache_beam.io import WriteToText
from apache_beam.io.gcp.bigquery import ReadFromBigQuery, WriteToBigQuery


class FormatName(beam.DoFn):
    def process(self, element):
        state = element['state']
        latitude = element['latitude']
        longitude = element['longitude']

        if state not in states:
            states.append(state)
            if latitude != None:
                record = {'state': state, 'latitude': latitude, 'longitude': longitude}
                return [record]

states = []
```
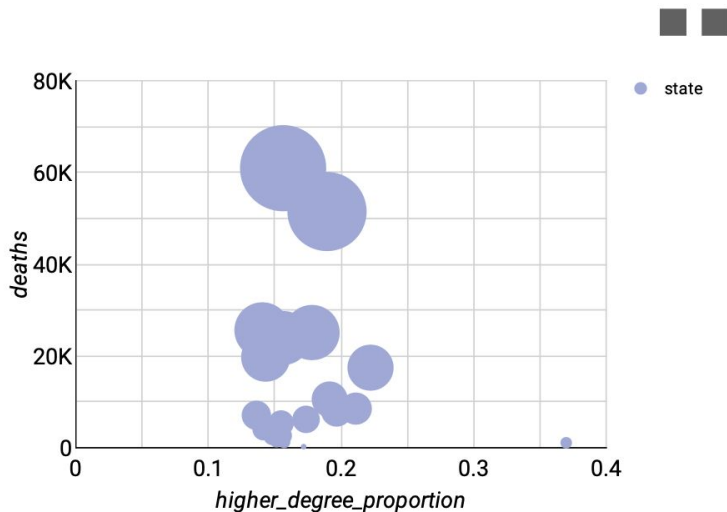
Cross-dataset queries & Data Visualisations

Number of Deaths Due To COVID in each U.S. State that ranked Top 20 for Having Highest Higher Degree Proportion Among the Adult Population Who Graduated High School


Number of Deaths Due to COVID in each U.S. State that ranked Top 20 for Having High Proportion of Elders Who Are 65 Years Old or Older Among the Entire State Population Represented By the Circles that Has a Radius Corresponding to Confirmed Cases in Each State
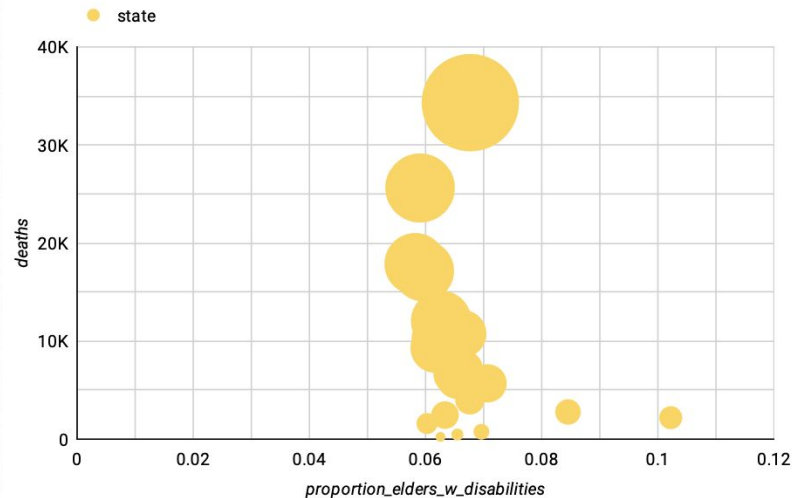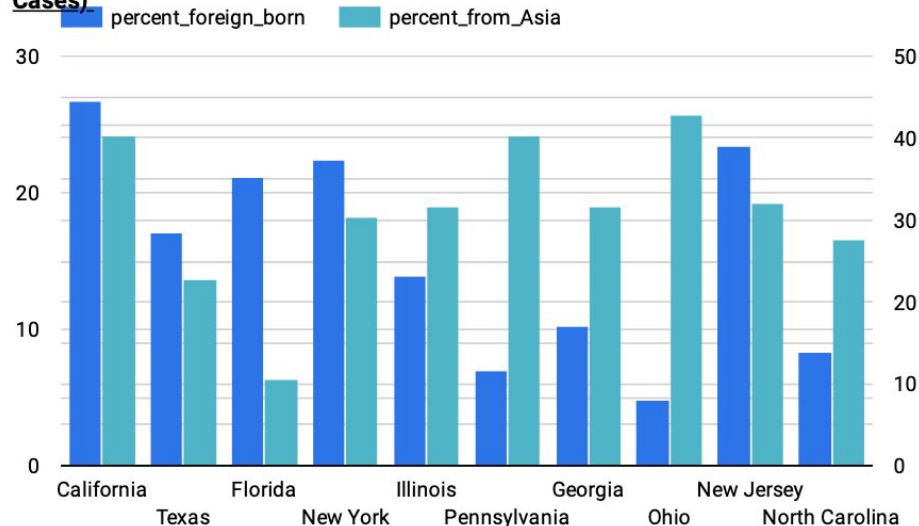
Data Visualisations

Number of Deaths and Confirmed Cases of COVID-19 in each U.S. State that ranked Top 10 for Having the Most Confirmed Cases on the Eastern Half of the United States
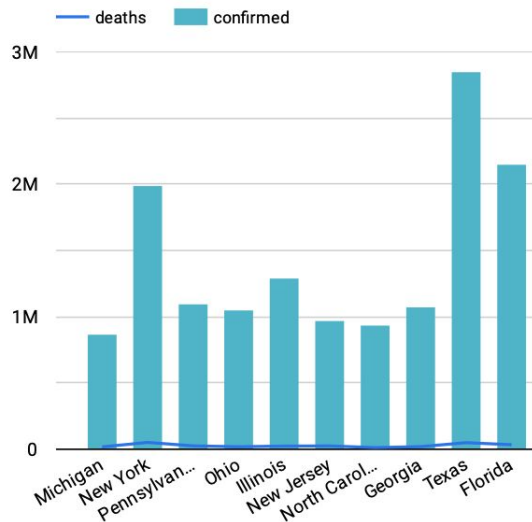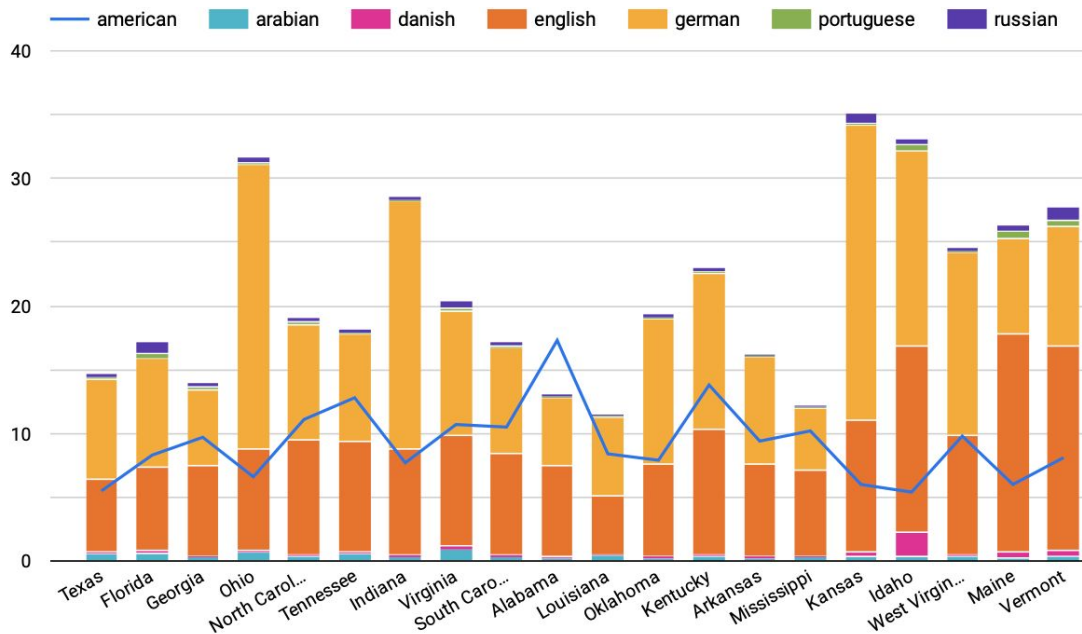

Percentage of Foreign Born Residents and Asian Citizens in Each State that Ranked Top 10 for Having the Most Confirmed Cases (Ranked by Most to Least Confirmed Cases)

Data Visualisations

Ancestry Distribution For the Top 20 States with Highest Percentage of American Heritage in its Population Ordered by Highest to Lowest Number of Confirmed Cases

Improvements

- Financial aspects of each state
- Compare with different countries (socioeconomic aspect)
- Percentage vs Count