Into the Corona-Verse

The Impact of COVID on the American Society

The COVID American Database Project

Database Phase 1

Daniel Caley

IST-659

February 7, 2021

Contents

Project Summary:

A non for profit wants to understand the impact of COVID related cases and deaths across America. Currently they have daily csv files by county of the number of COVID cases and deaths starting from May 2020 until present. They are finding that they need all this data in one place and simply aggregating the daily csv files into a single excel file is not scalable to perform analysis. In addition to their COVID data, they want to bring in additional Census Bureau data to help make their analysis even richer. The data includes population, household median income, and race by ZCTA. They will also need the ability to map ZCTA to Counties, Cities, and State. The final outcome for this project will allow workers of the non-profit to have all of their data in one place in order to perform analysis.

Stakeholders:

The stakeholders in this project are the non-profit who is interested in the impact of COVID on race and socioeconomic status in America. Inside of the non-profit the analysts working there will use the data to create presentations and research material surrounding the effects of the Coronavirus on different segments of the US population. Ultimately to help inform businesses, local municipalities, and news organizations that are in need of additional funding and to help combat the coronavirus impact on America.

High Level Rules:

- COVID cases and deaths are broken out daily by County. They only report daily, not twice a day.
- Cases and deaths might not be reported if there was no cases and death on that day for that county. The CSV by the non- profit will fill in the missing days with 0 cases and deaths by county.
- Census Bureau data is created at the Zip Code Tabulation Areas (ZCTA) level. This is different from postal Zip Code. There can be many Zip Codes underlying a ZCTA. We are not working with Zip Code in this dataset so we will not have to worry. Just need to know that these are 2 different ways to break out County level data.
- A Geo County State map will be required to connect the County to ZCTA.
    - The Geo County State map will include a county id and State.
    - The hierarchy for the multiple geographical attributes in this table are the following, from lowest level to the top:
        1. ZCTA
        2. City
        3. County

4. State
    - Said differently there can be many ZCTA underlying a City, there can be many Cities underlying a County, and there can be many Counties underlying a State.

- The Census Bureau provides population by age group and by ZCTA.
    - Each ZCTA will have multiple age groups and consist of 6 groups.
- The Census Bureau also provides Household Median Income by ZCTA.
- The Census Bureau also has race information by ZCTA. The race information provided in a race category. The Race categories consist of White, Asian, Black or African American, and Other.

Details of data needed and maintenance:

The following data will be needed:

1. COVID data will be needed and is tracked by the non-profit. They will be responsible for creating the CSV files. They will also be responsible for making sure that all counties exist in the CSV file with a 0 value for cases and deaths in the event there are none. The data engineer, myself, will be responsible for uploading the data into the database on a daily basis.
2. The Census Bureau data is ingested from an API. This data is updated yearly at the end of Quarter 1 or the beginning of Quarter 2. It's the responsibility of the Data Engineer to swap out the previous year's data with the current.
3. The non for profit does not want stale data from the Census Bureau and only wants the new data.

Glossary

A **ZCTA** stands for Zip Code Tabulation Areas and are generalized areal representations of United States Postal Service Zip Code service areas. ZCTA are trademarked by the U.S. Census Bureau. The hierarchy is that one or many Zip Codes can underly a ZCTA.

A **City** an inhabited place of greater size, population or importance. In the hierarchy many zip codes can exist inside of a city.

A **County** is political and administrative division of a state, providing certain local governmental services. There are many cities to a county.

A **State** is a territory considered as an organized political community under one government. There are many counties to a single State.

Household Median Income is the income level earned by a given household. The formal definition is the income cut-off where half of the households earn ore, and half earn less. In this case Household Median Income is a widely used metric when defining socioeconomic areas in business and non for profits.

Population is all inhabitants of a particular Zip Code, ZCTA, city, county, or state.

Race is a grouping of people who identify with each other on the basis of shared attributes that distinguish them from other groups such as a common set of traditions, ancestry, language, history, society or skin color. For the census data Race consists of 6 groups. They are the following:

1. White alone
2. Black or African American alone

3. American Indian and Alaska Native alone
4. Asian alone
5. Native Hawaiian and Other Pacific Islander alone
6. Other Race alone

Age Group consist of 23 groups. They are the following:

1. Under 5 years
2. 5 to 9 years
3. 10 to 14 years
4. 15 to 17 years
5. 18 and 19 years
6. 20 years
7. 21 years
8. 22 to 24 years
9. 25 to 29 years
10. 30 to 34 years
11. 35 to 39 years
12. 40 to 44 years
13. 45 to 49 years
14. 50 to 54 years
15. 55 to 59 years
16. 60 and 61 years
17. 62 to 64 years
18. 65 and 66 years

19. 67 to 69 years
20. 70 to 74 years
21. 75 to 79 years
22. 80 to 84 years
23. 85 years and over

Cases refer to the number of COVID related cases.

Deaths refer to the number of COVID related deaths.

COVID a highly contagious respiratory disease caused by the SARS-CoV-2 virus. Also known for causing a pandemic across the world that has disrupt economies, impacted a way of life, and caused many deaths.

Data Questions

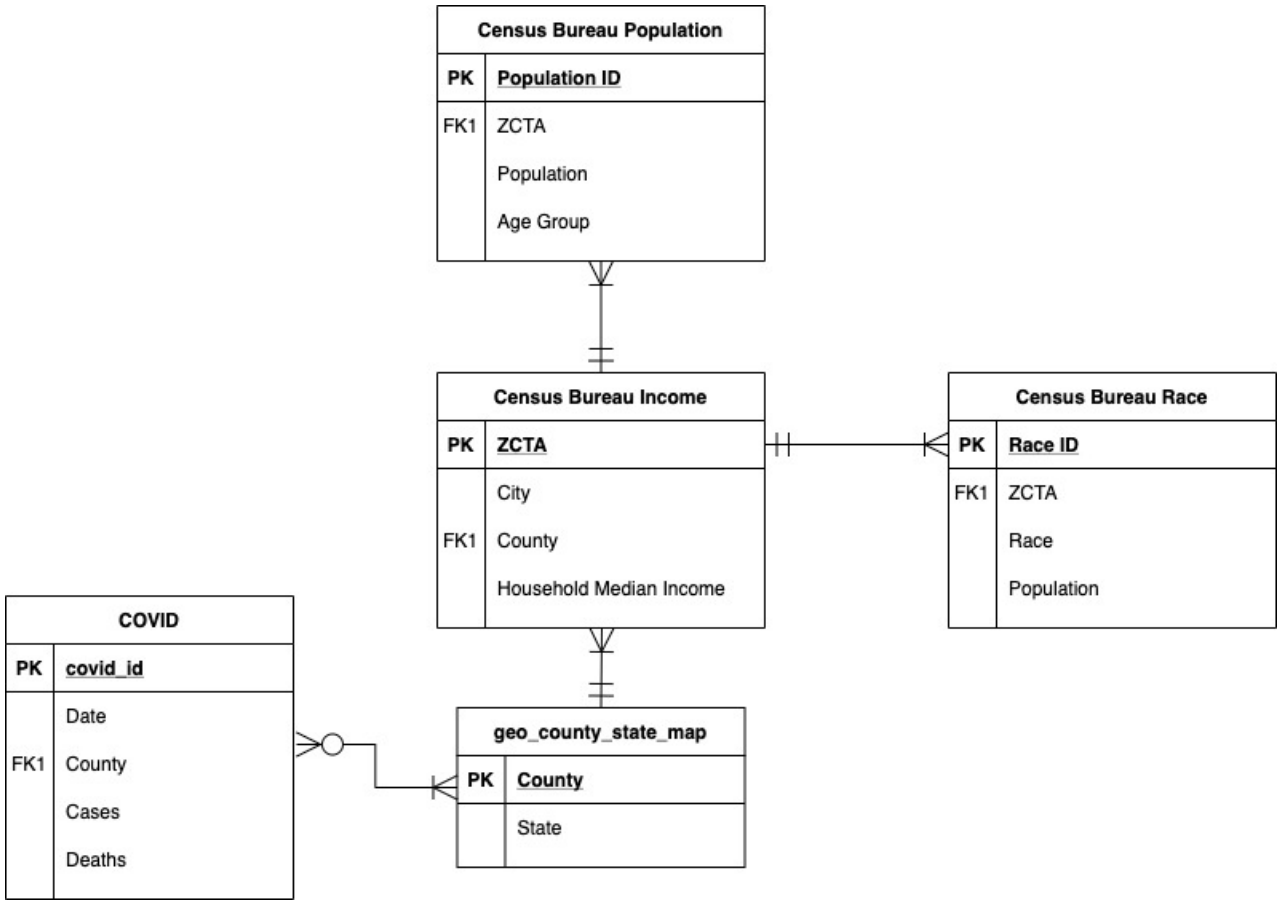
1. What is the trailing 7-day COVID cases and deaths from May 2020 until current?
    - Calculate the trailing 7-day cases divided by population.
    - Calculate the trailing 7-day deaths divided by population.
2. What is the Population and Household Median Income for the ZCTA 85203, Mesa AZ.
    - This should be by ZCTA, County, & State
    - What Percentage Lives in 85203, Maricopa vs the state?
3. What is the total population By County, Include State as a column?
    - What is the top 25 populated counties?
    - What is the trailing 7 days COVID cases divided by population for the top 25 counties?
4. What is the weighted average household median income by County?

- What's the top 25 highest weighted average household median income by population?
- What is the total population of these counties?
- What's the bottom 25 weighted average household median income by population?
- What is the trailing 7-day COVID cases divide by population?
- What does the top 25 highest weighted average household median income look compare to the bottom 25?

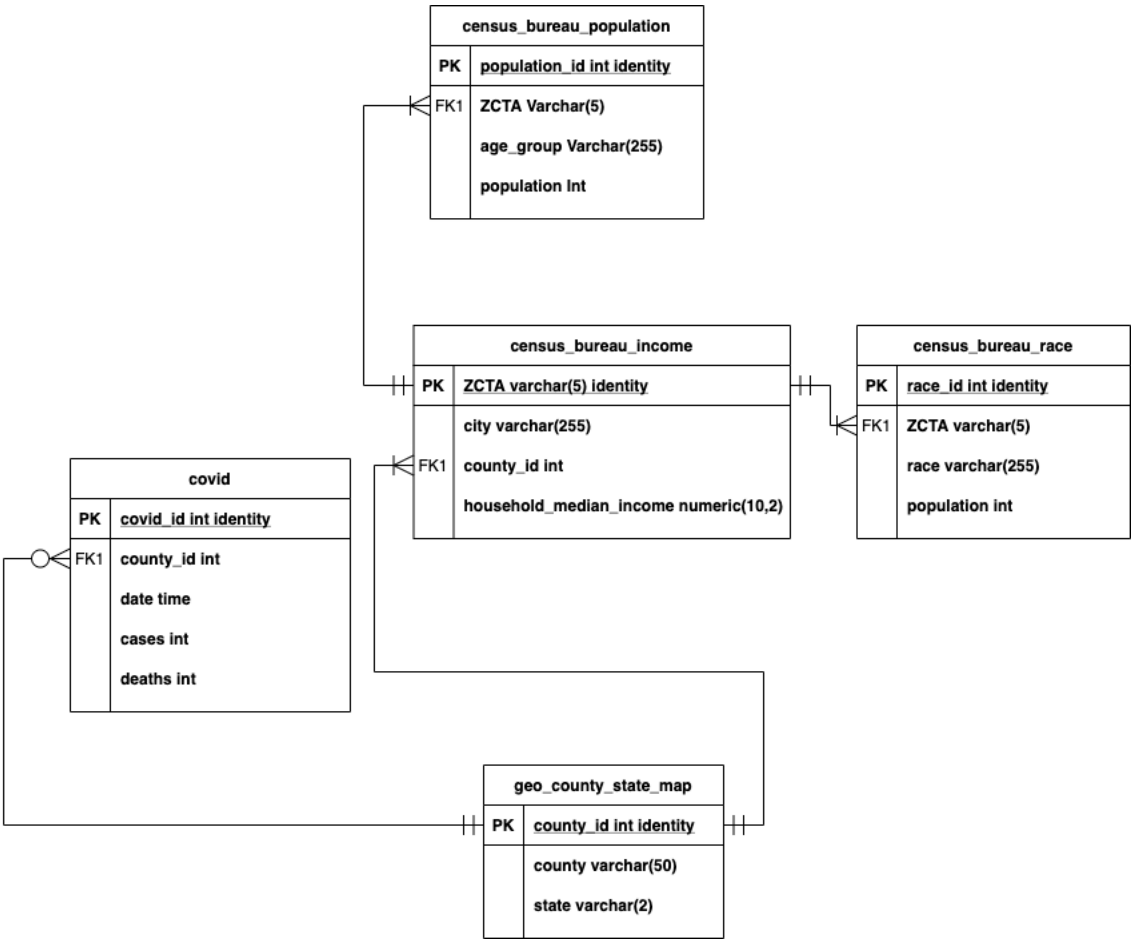
5. For ZCTA 85203, Arizona, what is the population broken out by age group?
    - How does that compare to the County?
    - How does that compare to the State?
6. What where the top 25 Counties who had the highest deaths by percentage of population?
    - The population must be over 1000 people.
    - Include their household median income.
    - Not looking at New York
7. Group the age groups that are considered the high-risk age categories for COVID. That is age groups over 65 and under 5.
    - Make this a calculated field that high risk or low risk.
    - Total the population by high risk and low risk by county.
    - Any counties with a population that is greater than 30% of the total population what is the number of COVID cases and deaths by total population vs low risk counties.
    - Group these counties into 2 rows, high risk and low risk. How do they total cases and deaths compare to each other?
8. How many Deaths Vs. Total Population

- Compare 2 rows below 30% and above 30% and how do the total cases and deaths by population compare to each other?

Conceptual Model



Logical Model



Physical Database Design

```
-- Creating Tables, Views, procedure, and functions.
-- Creating Drops

--Drop Tables
Drop table if exists dc_census_bureau_population
go

drop table if exists dc_census_bureau_race
go

Drop table if exists dc_covid
go

Drop table if exists dc_census_bureau_income
go

Drop table if exists dc_geo_county_state_map
go



-- Creating database tables
-- Creating daily covid & death table
CREATE TABLE dc_geo_county_state_map(
	-- Creating Columns
	county_id varchar(5),
	county varchar(255) NOT NULL,
	state varchar(2) NOT NULL,
	-- Applying Constraints
	CONSTRAINT PK_dc_geo_county_state_map PRIMARY KEY (county_id)
	)
go


CREATE TABLE dc_covid (
	-- Creating Columns
	covid_id varchar(10),
	county_id varchar(5) NOT NULL,
	date date NOT NULL,
	cases int NOT NULL,
	deaths int NOT NULL,
	-- Applying Constraints
	CONSTRAINT PK_dc_covid PRIMARY KEY (covid_id),
	CONSTRAINT FK1_dc_covid FOREIGN KEY (county_id) REFERENCES
		dc_geo_county_state_map(county_id)
	)
go

CREATE TABLE dc_census_bureau_income (
	-- Creating Columns
	zcta varchar(5),
```

```
        county_id varchar(5) NOT NULL,
        household_median_income numeric(10,2),
        -- Applying Constraints
        CONSTRAINT PK_dc_census_bureau_income PRIMARY KEY (zcta),
        CONSTRAINT FK_dc_census_bureau_income FOREIGN KEY (county_id) REFERENCES
        dc_geo_county_state_map(county_id)
        )
go

CREATE TABLE dc_census_bureau_race (
        -- Creating Columns
        race_id varchar(8),
        zcta varchar(5),
        race varchar(255),
        population int,
        -- Applying Constraints
        CONSTRAINT PK_dc_census_bureau_race PRIMARY KEY (race_id),
        CONSTRAINT FK_dc_census_bureau_race FOREIGN KEY (zcta) REFERENCES
        dc_census_bureau_income(zcta)
        )
go

CREATE TABLE dc_census_bureau_population (
        -- Creating Columns
        population_id varchar(7),
        zcta varchar(5),
        age_group varchar(255),
        population int
        -- Applying Constraints
        CONSTRAINT PK_dc_bureau_population PRIMARY KEY (population_id),
        CONSTRAINT FK_dc_bureau_population FOREIGN KEY (zcta) REFERENCES
        dc_census_bureau_income(zcta)
        )
go


--- Inserting records into tables.  Total recoreds is over 1 million.
--- Therefore I will load the records through a bulk upload process.
--- The below will only show 5 records for each table.

--- Inserting Records for geo_county_state_map
INSERT INTO dc_geo_county_state_map(county_id, county, state)
VALUES  ('01001', 'Autauga County',  'AL'),
              ('01003', 'Baldwin County',  'AL'),
              ('01005', 'Barbour County',  'AL'),
              ('01007', 'Bibb County',  'AL'),
              ('01009', 'Blount County',  'AL')
go
```

```
-- Look to see if the insert worked
select * from dc_geo_county_state_map
```

Results | Messages

	county_id	county	state
1	01001	Autauga County	AL
2	01003	Baldwin County	AL
3	01005	Barbour County	AL
4	01007	Bibb County	AL
5	01009	Blount County	AL

```
--- Inserting Records for census_bureau_income
INSERT INTO dc_census_bureau_income(zcta, county_id, household_median_income)
VALUES  ('36003', '01001',  '37000'),
              ('36480', '01003',  '27461'),
              ('36005', '01005',  '49722'),
              ('35034', '01007',  '39087'),
              ('35013', '01009',  '0')
go

-- Look to see if the insert worked
select * from dc_census_bureau_income
```

	zcta	county_id	household_median_income
1	35013	01009	0.00
2	35034	01007	39087.00
3	36003	01001	37000.00
4	36005	01005	49722.00
5	36480	01003	27461.00

```
--- Inserting Records for covid
INSERT INTO dc_covid(covid_id, county_id, date, cases, deaths)
VALUES  ('0000009481', '01001',  '2020-03-24', '1', '0'),
              ('0000001935', '01003',  '2020-03-14', '1', '0'),
              ('0000028399', '01005',  '2020-04-03', '1', '0'),
              ('0000019658', '01007',  '2020-03-30', '2', '0'),
              ('0000010838', '01009',  '2020-03-25', '1', '0')
go

-- Look to see if the insert worked
select * from dc_covid
```

	covid_id	county_id	date	cases	deaths
1	0000001935	01003	2020-03-14	1	0
2	0000009481	01001	2020-03-24	1	0
3	0000010838	01009	2020-03-25	1	0
4	0000019658	01007	2020-03-30	2	0
5	0000028399	01005	2020-04-03	1	0

```
--- Inserting Records for population
INSERT INTO dc_census_bureau_population(population_id, zcta, age_group, population)
VALUES  ('360031', '36003', 'Under 5 years',  '111'),
              ('364801', '36480', 'Under 5 years',  '28'),
              ('360051', '36005', 'Under 5 years',  '57'),
              ('350341', '35034', 'Under 5 years',  '547'),
              ('350131', '35013', 'Under 5 years',  '0')
go


-- Look to see if the insert worked
select * from dc_census_bureau_population
```

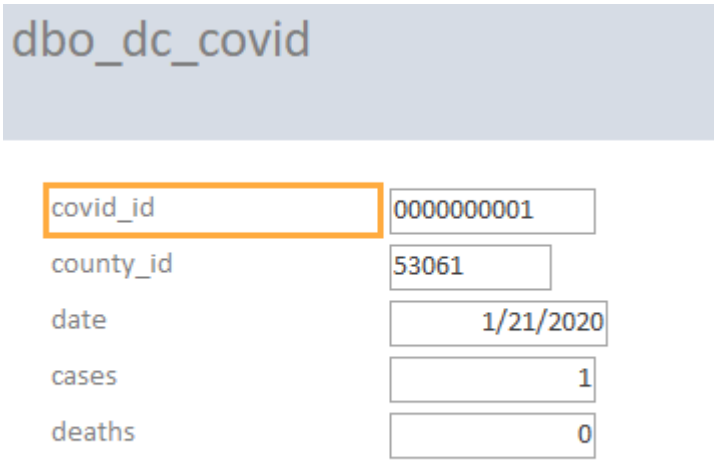
	population_id	zcta	age_group	population
1	350131	35013	Under 5 years	0
2	350341	35034	Under 5 years	547
3	360031	36003	Under 5 years	111
4	360051	36005	Under 5 years	57
5	364801	36480	Under 5 years	28

```
--- Inserting Records for race
INSERT INTO dc_census_bureau_race(race_id, zcta, race, population)
VALUES  ('36003101', '36003', 'Black or African American alone',  '1102'),
              ('36480100', '36480', 'White alone',  '1389'),
              ('36005102', '36005', 'American Indian and Alaska Native alone',  '0'),
              ('35034100', '35034', 'White alone',  '2942'),
              ('35013105', '35013', 'Other Race alone',  '0')
go

-- Look to see if the insert worked
select * from dc_census_bureau_race
```

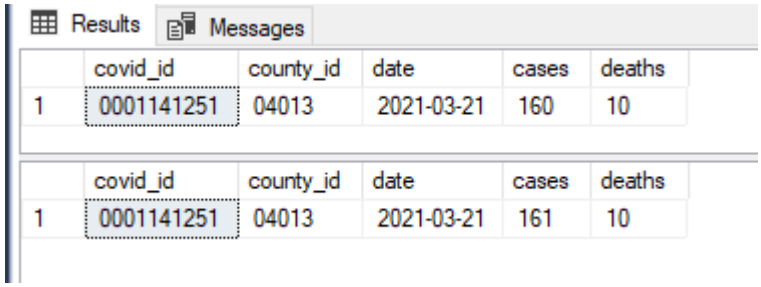
	race_id	zcta	race	population
1	35013105	35013	Other Race alone	0
2	35034100	35034	White alone	2942
3	36003101	36003	Black or African American alone	1102
4	36005102	36005	American Indian and Alaska Native alone	0
5	36480100	36480	White alone	1389

Utilizing Access & Creating Procedures:

```
--- Form made from Access to help fix issues in the database
```




```
--- Creating a procedure to update a COVID cases based from the county id and date.
--- The first parameter is the user name for the user to change
--- The second is the new email address

CREATE PROCEDURE dc_ChangeCovidCases(@covid_cases int,@county_id varchar(5), @covid_date
date)
AS
BEGIN
	UPDATE dc_covid SET cases = @covid_cases
	WHERE county_id = @county_id and date = @covid_date
END
GO
```



Creating Views and Functions:

```
-- Before answering the questions going to create a series of views and functions to help
answer the questions.
```

```
--- Finding total covid cases and deaths by county.
drop view if exists dc_covid_county

go
CREATE VIEW dc_covid_county as
select dc_covid.county_id,
          dc_geo_county_state_map.county,
          dc_geo_county_state_map.state,
          sum(cases) total_cases,
          sum(deaths) total_deaths
from dc_covid
join dc_geo_county_state_map
on dc_covid.county_id = dc_geo_county_state_map.county_id
group by dc_covid.county_id,
              dc_geo_county_state_map.county,
              dc_geo_county_state_map.state
go
--- Checking to see if the view worked
select * from dc_covid_county
```

	county_id	county	state	total_cases	total_deaths
1	01009	Blount County	AL	6383	130
2	01011	Bullock County	AL	1194	39
3	01023	Choctaw County	AL	581	24
4	01029	Cleburne County	AL	1446	41
5	01049	DeKalb County	AL	8674	179
6	01053	Escambia County	AL	3839	74
7	01055	Etowah County	AL	13604	339
8	01063	Greene County	AL	897	33
9	01065	Hale County	AL	2147	72
10	01073	Jefferson County	AL	74079	1442
11	01087	Macon County	AL	1520	47
12	01099	Monroe County	AL	1692	39
13	01127	Walker County	AL	7015	269
14	01131	Wilcox County	AL	1247	26
15	02020	Anchorage Municipality	AK	27440	160
16	02050	Bethel Census Area	AK	3685	20
17	02068	Denali Borough	AK	78	0
18	02158	Kusilvak Census Area	AK	1217	3
19	02188	Northwest Arctic Borough	AK	595	2
20	02261	Valdez-Cordova Census Area	AK	521	2
21	04005	Coconino County	AZ	16952	323
22	04012	La Paz County	AZ	2430	74
23	05007	Benton County	AR	27946	399
24	05009	Boone County	AR	3713	78

```
--- Finding daily covid cases nation wide from May 2020 until current.
drop view if exists dc_daily_covid
go
CREATE VIEW dc_daily_covid as
```

```
	with daily_covid as (
		Select county_id,
			date,
			sum(cases) daily_cases,
			sum(deaths) daily_deaths
		from dc_covid
		-- We have to go 7 days before May as we want a rolling 7 days.
		-- SQL will begin the sum but the first 6 records won't be a rolling 7 days
		where date between '20200424' and GETDATE()
		group by county_id, date
	),
	trailing_covid as (
		select *,
			sum(daily_cases) over (partition by county_id order by date rows
between 6 preceding and current row) t7d_cases,
			sum(daily_deaths) over (partition by county_id order by date rows
between 6 preceding and current row) t7d_deaths
		from daily_covid
	)
		select  * from  trailing_covid
		where date >= '20200501'
	go

--- Checking our view
select * from dc_daily_covid order by county_id, date
```

	county_id	date	daily_cases	daily_deaths	t7d_cases	t7d_deaths
1	01001	2020-05-01	0	-1	6	1
2	01001	2020-05-02	3	0	8	1
3	01001	2020-05-03	3	0	11	1
4	01001	2020-05-04	5	0	14	0
5	01001	2020-05-05	0	0	13	-1
6	01001	2020-05-06	5	0	15	-1
7	01001	2020-05-07	3	0	19	-1
8	01001	2020-05-08	6	1	25	1
9	01001	2020-05-09	1	0	23	1
10	01001	2020-05-10	6	0	26	1
11	01001	2020-05-11	10	0	31	1
12	01001	2020-05-12	7	0	38	1
13	01001	2020-05-13	2	0	35	1
14	01001	2020-05-14	11	0	43	1
15	01001	2020-05-15	-1	0	36	0
16	01001	2020-05-16	7	0	42	0
17	01001	2020-05-17	0	0	36	0
18	01001	2020-05-18	10	0	36	0
19	01001	2020-05-19	7	0	36	0
20	01001	2020-05-20	9	-1	43	-1
21	01001	2020-05-21	11	0	43	-1
22	01001	2020-05-22	2	0	46	-1
23	01001	2020-05-23	6	0	45	-1
24	01001	2020-05-24	4	0	49	-1

```
-- creating a zcta, county, state view with total population and median income
drop view if exists dc_geo_population
go
CREATE VIEW dc_geo_population as

	Select dc_census_bureau_income.zcta,
		dc_census_bureau_income.county_id,
		   dc_geo_county_state_map.county,
		   dc_geo_county_state_map.state,
		   sum(dc_census_bureau_population.population) population,
		   dc_census_bureau_income.household_median_income
	from dc_census_bureau_income
		join dc_geo_county_state_map
			on dc_census_bureau_income.county_id =
dc_geo_county_state_map.county_id
		join dc_census_bureau_population
			on dc_census_bureau_income.zcta = dc_census_bureau_population.zcta
	group by dc_census_bureau_income.zcta,
			dc_census_bureau_income.county_id,
			dc_geo_county_state_map.county,
			dc_geo_county_state_map.state,
			dc_census_bureau_income.household_median_income
go
```

```sql
-- Checking our view
select * from dc_geo_population
where state = 'NY' order by zcta
```

	zcta	county_id	county	state	population	household_median_income
1	06390	36103	Suffolk County	NY	125	61125.00
2	10001	36061	New York County	NY	24117	92840.00
3	10002	36061	New York County	NY	74479	36982.00
4	10003	36061	New York County	NY	53977	118161.00
5	10004	36061	New York County	NY	3335	190223.00
6	10005	36061	New York County	NY	8701	189702.00
7	10006	36061	New York County	NY	3092	179044.00
8	10007	36061	New York County	NY	7408	224063.00
9	10009	36061	New York County	NY	58293	63717.00
10	10010	36061	New York County	NY	35906	132988.00
11	10011	36061	New York County	NY	49949	138272.00
12	10012	36061	New York County	NY	23318	106467.00
13	10013	36061	New York County	NY	28799	113191.00
14	10014	36061	New York County	NY	30344	133501.00
15	10016	36061	New York County	NY	52886	126628.00
16	10017	36061	New York County	NY	15846	131045.00
17	10018	36061	New York County	NY	8806	122484.00
18	10019	36061	New York County	NY	45498	103792.00
19	10020	36061	New York County	NY	0	0.00
20	10021	36061	New York County	NY	44280	122169.00
21	10022	36061	New York County	NY	31130	150718.00
22	10023	36061	New York County	NY	62541	132605.00
23	10024	36061	New York County	NY	58102	143623.00
24	10025	36061	New York County	NY	92251	91624.00

```sql
-- Creating a view for age groups that includes a covid_age_type
-- and attributes like county and state
drop view if exists dc_age_group

go
create view dc_age_group as

select
	dc_census_bureau_population.zcta,
		case
			when age_group in ('Under 5 years','65 and 66 years',
					'67 to 69 years','70 to 74 years',
					'75 to 79 years','80 to 84 years',
					'85 years and over')
				then 'High Risk Age Group'
			else 'Low Risk Age Group'
				end as covid_age_group,
```

```
		dc_census_bureau_population.age_group,
		dc_geo_county_state_map.county_id,
		dc_geo_county_state_map.county,
		dc_geo_county_state_map.state,
		dc_census_bureau_population.population,
		case
			when age_group = '85 years and over'
				then 85
			else left(ltrim(right(age_group,8)),2)* 1
		end
			as sort_key
from dc_census_bureau_population
	left join dc_census_bureau_income
		on dc_census_bureau_population.zcta = dc_census_bureau_income.zcta
	left join dc_geo_county_state_map
		on dc_census_bureau_income.county_id = dc_geo_county_state_map.county_id
go

-- Checking our view

select * from dc_age_group
order by zcta, sort_key
```

	zcta	covid_age_group	age_group	county_id	county	state	population	sort_key
1	00601	High Risk Age Group	Under 5 years	72001	Adjuntas Municipio	PR	803	5
2	00601	Low Risk Age Group	5 to 9 years	72001	Adjuntas Municipio	PR	835	9
3	00601	Low Risk Age Group	10 to 14 years	72001	Adjuntas Municipio	PR	1270	14
4	00601	Low Risk Age Group	15 to 17 years	72001	Adjuntas Municipio	PR	652	17
5	00601	Low Risk Age Group	18 and 19 years	72001	Adjuntas Municipio	PR	479	19
6	00601	Low Risk Age Group	20 years	72001	Adjuntas Municipio	PR	216	20
7	00601	Low Risk Age Group	21 years	72001	Adjuntas Municipio	PR	253	21
8	00601	Low Risk Age Group	22 to 24 years	72001	Adjuntas Municipio	PR	671	24
9	00601	Low Risk Age Group	25 to 29 years	72001	Adjuntas Municipio	PR	1125	29
10	00601	Low Risk Age Group	30 to 34 years	72001	Adjuntas Municipio	PR	1012	34
11	00601	Low Risk Age Group	35 to 39 years	72001	Adjuntas Municipio	PR	825	39
12	00601	Low Risk Age Group	40 to 44 years	72001	Adjuntas Municipio	PR	1100	44
13	00601	Low Risk Age Group	45 to 49 years	72001	Adjuntas Municipio	PR	1138	49
14	00601	Low Risk Age Group	50 to 54 years	72001	Adjuntas Municipio	PR	1227	54
15	00601	Low Risk Age Group	55 to 59 years	72001	Adjuntas Municipio	PR	1110	59
16	00601	Low Risk Age Group	60 and 61 years	72001	Adjuntas Municipio	PR	555	61
17	00601	Low Risk Age Group	62 to 64 years	72001	Adjuntas Municipio	PR	731	64
18	00601	High Risk Age Group	65 and 66 years	72001	Adjuntas Municipio	PR	498	66
19	00601	High Risk Age Group	67 to 69 years	72001	Adjuntas Municipio	PR	640	69
20	00601	High Risk Age Group	70 to 74 years	72001	Adjuntas Municipio	PR	654	74
21	00601	High Risk Age Group	75 to 79 years	72001	Adjuntas Municipio	PR	582	79
22	00601	High Risk Age Group	80 to 84 years	72001	Adjuntas Municipio	PR	393	84
23	00601	High Risk Age Group	85 years and over	72001	Adjuntas Municipio	PR	344	85
24	00602	High Risk Age Group	Under 5 years	72003	Aguada Municipio	PR	1492	5

```
Drop function if exists dbo.dc_GetPopulation

go
CREATE FUNCTION dbo.dc_GetPopulation(@countyid varchar(5))
RETURNS INT AS
BEGIN
	DECLARE @Popcounty int

	SELECT @Popcounty = SUM(population) from dc_geo_population where county_id =
@countyid

	RETURN @Popcounty
END
Go
--- Checking to see if the function works. Maricopa County is 04013
select top 1 county_id, dbo.dc_GetPopulation(county_id) county_population
from dc_covid
where county_id = '04013'
```

	county_id	county_population
1	04013	4401060

```
--- Create a function that, when given the zcta, returns the total population by County
Drop function if exists dbo.dc_ZCTA_CountyPop

go
CREATE FUNCTION dbo.dc_ZCTA_CountyPop(@zcta varchar(5))
RETURNS INT AS
BEGIN
	DECLARE @Countypop int

	SELECT @Countypop = SUM(population) from dc_geo_population where county_id =
(select county_id from dc_geo_population where zcta =  @zcta)

	RETURN @Countypop
END
Go

-- Checking to see if the function works. Using ZCTA 85203
select zcta, dbo.dc_ZCTA_CountyPop(zcta) county_population
from dc_geo_population
where zcta = '85203'
```

	zcta	county_population
1	85203	4401060

```
--- Create a function that, when given the zcta, returns the total population by State
Drop function if exists dbo.dc_ZCTA_StatePop

go
CREATE FUNCTION dbo.dc_ZCTA_StatePop(@zcta varchar(5))
RETURNS INT AS
BEGIN
	DECLARE @Statepop int
```

```
	SELECT @Statepop = SUM(population) from dc_geo_population where state = (select
state from dc_geo_population where zcta =  @zcta)

	RETURN @Statepop
END
Go

-- Checking to if the function works
Select zcta, dbo.dc_ZCTA_StatePop(zcta) state_population
from dc_geo_population
where zcta = '85203'
```

	zcta	state_population
1	85203	7057786

```
--- Create a function that, when given the zcta, returns the total Household Median
Income by County
Drop function if exists dbo.dc_County_MedianIncome

go
CREATE FUNCTION dbo.dc_County_MedianIncome(@zcta varchar(5))
RETURNS INT AS
BEGIN
	DECLARE @CountyIncome numeric(10,2)

	SELECT @CountyIncome = SUM(household_median_income*population) / sum(population)
from dc_geo_population where county_id = (select county_id from dc_geo_population where
zcta =  @zcta)

	RETURN @CountyIncome
END
Go

-- Checking to if the function works
Select zcta, household_median_income, dbo.dc_County_MedianIncome(zcta)
County_MedianIncome
from dc_geo_population
where zcta = '85203'
```

	zcta	household_median_income	County_MedianIncome
1	85203	54919.00	68403

```
--- Create a function that, when given the zcta, returns the Household Median Income by
State
Drop function if exists dbo.dc_State_MedianIncome

go
CREATE FUNCTION dbo.dc_State_MedianIncome(@zcta varchar(5))
RETURNS INT AS
BEGIN
	DECLARE @StateIncome numeric(10,2)
```

```
	SELECT @StateIncome = SUM(household_median_income*population) / sum(population)
from dc_geo_population where state = (select state from dc_geo_population where zcta =
@zcta)

	RETURN @StateIncome
END
Go

-- Checking to if all these median income function works
Select zcta, household_median_income, dbo.dc_State_MedianIncome(zcta) State_MedianIncome
from dc_geo_population
where zcta = '85203'
```

	zcta	household_median_income	State_MedianIncome
1	85203	54919.00	62285

```
Drop function if exists dbo.dc_County_AgeGroup

go
CREATE FUNCTION dbo.dc_County_AgeGroup(@zcta varchar(5), @agegroup varchar(255))
RETURNS INT AS
BEGIN
	DECLARE @Countypop int

	SELECT @Countypop = sum(population) from dc_age_group where county =
                                                                    (select
county
                                                                        from
dc_age_group
                                                                        where
zcta =  @zcta and age_group = @agegroup)
                                                                    and age_group
= @agegroup

	RETURN @Countypop
END
Go

-- Checking to if all the age group county function works
Select zcta, population, age_group, dbo.dc_County_AgeGroup(zcta, age_group)
County_Population
from dc_age_group
where zcta = '85203'
order by sort_key
```

	zcta	population	age_group	County_Population
1	85203	3241	Under 5 years	282759
2	85203	3264	5 to 9 years	291566
3	85203	2903	10 to 14 years	307546
4	85203	1760	15 to 17 years	180807
5	85203	1225	18 and 19 years	115921
6	85203	548	20 years	59131
7	85203	258	21 years	59573
8	85203	1762	22 to 24 years	174934
9	85203	3654	25 to 29 years	330686
10	85203	2885	30 to 34 years	308067
11	85203	2455	35 to 39 years	291103
12	85203	2401	40 to 44 years	282633
13	85203	2504	45 to 49 years	282690
14	85203	2370	50 to 54 years	272665
15	85203	2423	55 to 59 years	262802
16	85203	769	60 and 61 years	100101
17	85203	1196	62 to 64 years	141200
18	85203	606	65 and 66 years	87136
19	85203	953	67 to 69 years	124696
20	85203	986	70 to 74 years	170966
21	85203	721	75 to 79 years	121053
22	85203	512	80 to 84 years	77887
23	85203	401	85 years and over	75138

```
--- Create a function that, when given the zcta and age group, returns the States total
population
Drop function if exists dbo.dc_State_AgeGroup

go
CREATE FUNCTION dbo.dc_State_AgeGroup(@zcta varchar(5), @agegroup varchar(255))
RETURNS INT AS
BEGIN
	DECLARE @Statepop int

	SELECT @statepop = sum(population) from dc_age_group where state =
(select state
	from
dc_age_group
	where
zcta =  @zcta and age_group = @agegroup)
and age_group
= @agegroup

	RETURN @statepop
END
Go

-- Checking to if all the age group state function works
```

```
Select zcta, population, age_group, dbo.dc_State_AgeGroup(zcta, age_group)
State_Population
from dc_age_group
where zcta = '85203'
order by sort_key
```

	zcta	population	age_group	State_Population
1	85203	3241	Under 5 years	434649
2	85203	3264	5 to 9 years	451200
3	85203	2903	10 to 14 years	472429
4	85203	1760	15 to 17 years	279414
5	85203	1225	18 and 19 years	194500
6	85203	548	20 years	102784
7	85203	258	21 years	100655
8	85203	1762	22 to 24 years	285223
9	85203	3654	25 to 29 years	501843
10	85203	2885	30 to 34 years	464121
11	85203	2455	35 to 39 years	444010
12	85203	2401	40 to 44 years	423540
13	85203	2504	45 to 49 years	425403
14	85203	2370	50 to 54 years	422843
15	85203	2423	55 to 59 years	427914
16	85203	769	60 and 61 years	171524
17	85203	1196	62 to 64 years	248113
18	85203	606	65 and 66 years	159150
19	85203	953	67 to 69 years	227293
20	85203	986	70 to 74 years	314562
21	85203	721	75 to 79 years	229963
22	85203	512	80 to 84 years	145588
23	85203	401	85 years and over	131065

Answering Data Questions:

```
-- 1. What is the daily and trailing 7-day COVID cases and deaths from May 2020 until
current?
--            Calculate the trailing 7-day cases divided by population.
--            Calculate the trailing 7-day deaths divided by population.


-- 1. What is the daily and trailing 7-day COVID cases and deaths May 2020 until current?
--            Going to create 2 temporay views, commit the tables to daily cases and
deaths view, and then drop the first 2 views


-- Droping views to not muddy up the database.  Was only using them as suedo temp table.

--            Calculate the trailing 7-day cases divided by population.
--            Calculate the trailing 7-day deaths divided by population.



-- 2.  What is the Population and Household Median Income for the ZCTA 85203.
--                  This should be by ZCTA, County, & State
--                  What Percentage Lives in 85203, Maricopa vs the state
Select zcta,
              county,
              state,
              population ZCTA_Population,
              dbo.dc_ZCTA_CountyPop(zcta) County_Population,
              dbo.dc_ZCTA_StatePop(zcta) State_Population,
              household_median_income ZCTA_MedianIncome,
              dbo.dc_County_MedianIncome(zcta) County_MedianIncome,
              dbo.dc_State_MedianIncome(zcta) State_MedianIncome
from dc_geo_population
where zcta = '85203'


Select zcta,
              county,
              state,
              population * 1.0 / dbo.dc_ZCTA_StatePop(zcta) ZCTA_perc_vs_state,
              dbo.dc_ZCTA_CountyPop(zcta) * 1.0 / dbo.dc_ZCTA_StatePop(zcta)
County_perc_vs_state
from dc_geo_population

where zcta = '85203'
```

	zcta	county	state	ZCTA_Population	County_Population	State_Population	ZCTA_MedianIncome	County_MedianIncome	State_MedianIncome
1	85203	Maricopa County	AZ	39797	4401060	7057786	54919.00	68403	62285

	zcta	county	state	ZCTA_perc_vs_state	County_perc_vs_state
1	85203	Maricopa County	AZ	0.005638737133	0.623575155154

```sql
--- 3. What is the total population By County?
---                  What is the top 25 populated counties?
---                  What's their Weighted Average Household Median Income by population?
---                  What is the trailing 7 days COVID cases divided by population?

select state,
             county,
             SUM(population) total_population
from dc_geo_population
group by state, county

select top 25 state,
             county,
             SUM(population) total_population
from dc_geo_population
group by state, county
order by total_population desc

select top 25 state,
          county,
          SUM(population) total_population
from dc_geo_population
group by state, county
order by total_population desc

go
with top_25_county as (
                                        select top 25 state,
                                                  county,
                                                  county_id,
                                                  SUM(population) total_population
                                        from dc_geo_population
                                        group by state, county, county_id
                                        order by total_population desc)

select date,
          state,
          county,
          top_25_county.county_id,
          daily_cases,
          t7d_cases,
          t7d_deaths,
       t7d_deaths * 1.0 / t7d_cases deaths_cases_perc,
          t7d_cases * 1.0 / total_population cases_by_population
from top_25_county
       left join dc_daily_covid
       on top_25_county.county_id = dc_daily_covid.county_id
order by total_population desc, date
```

	state	county	total_population
1	CA	Los Angeles County	10294506
2	IL	Cook County	5508302
3	TX	Harris County	4405610
4	AZ	Maricopa County	4401060
5	CA	San Diego County	3260294
6	CA	Orange County	3075155
7	TX	Dallas County	2741458
8	FL	Miami-Dade County	2699439
9	NY	Kings County	2589974
10	CA	Riverside County	2536415
11	NY	Queens County	2255738
12	WA	King County	2215875
13	NV	Clark County	2179948
14	TX	Bexar County	2030193
15	CA	San Bernardino County	1969461
16	FL	Broward County	1926205
17	CA	Santa Clara County	1913644
18	CA	Alameda County	1732726
19	TX	Tarrant County	1670120
20	NY	New York County	1621398
21	MA	Middlesex County	1615535
22	MI	Wayne County	1588978
23	PA	Philadelphia County	1546122
24	CA	Sacramento County	1530796
25	FL	Hillsborough County	1508389

	state	county	total_population
1	CA	Los Angeles County	10294506
2	IL	Cook County	5508302
3	TX	Harris County	4405610
4	AZ	Maricopa County	4401060
5	CA	San Diego County	3260294
6	CA	Orange County	3075155
7	TX	Dallas County	2741458
8	FL	Miami-Dade County	2699439
9	NY	Kings County	2589974
10	CA	Riverside County	2536415
11	NY	Queens County	2255738
12	WA	King County	2215875
13	NV	Clark County	2179948
14	TX	Bexar County	2030193
15	CA	San Bernardino County	1969461
16	FL	Broward County	1926205
17	CA	Santa Clara County	1913644
18	CA	Alameda County	1732726
19	TX	Tarrant County	1670120
20	NY	New York County	1621398
21	MA	Middlesex County	1615535
22	MI	Wayne County	1588978
23	PA	Philadelphia County	1546122
24	CA	Sacramento County	1530796
25	FL	Hillsborough County	1508389

	date	state	county	county_id	daily_cases	t7d_cases	cases_by_population
1	2020-05-01	CA	Los Angeles County	06037	1033	5670	0.000550779221
2	2020-05-02	CA	Los Angeles County	06037	679	5787	0.000562144506
3	2020-05-03	CA	Los Angeles County	06037	768	6134	0.000595851806
4	2020-05-04	CA	Los Angeles County	06037	555	5800	0.000563407316
5	2020-05-05	CA	Los Angeles County	06037	1598	6839	0.000664334937
6	2020-05-06	CA	Los Angeles County	06037	829	6159	0.000598280286
7	2020-05-07	CA	Los Angeles County	06037	783	6245	0.000606634257
8	2020-05-08	CA	Los Angeles County	06037	869	6081	0.000590703429
9	2020-05-09	CA	Los Angeles County	06037	901	6303	0.000612268330
10	2020-05-10	CA	Los Angeles County	06037	480	6015	0.000584292242
11	2020-05-11	CA	Los Angeles County	06037	581	6041	0.000586817861
12	2020-05-12	CA	Los Angeles County	06037	922	5365	0.000521151767
13	2020-05-13	CA	Los Angeles County	06037	1248	5784	0.000561853089
14	2020-05-14	CA	Los Angeles County	06037	901	5902	0.000573315514
15	2020-05-15	CA	Los Angeles County	06037	930	5963	0.000579241004
16	2020-05-16	CA	Los Angeles County	06037	1044	6106	0.000593131909
17	2020-05-17	CA	Los Angeles County	06037	671	6297	0.000611685495
18	2020-05-18	CA	Los Angeles County	06037	477	6193	0.000601583019
19	2020-05-19	CA	Los Angeles County	06037	1122	6393	0.000621010857
20	2020-05-20	CA	Los Angeles County	06037	1284	6429	0.000624507868
21	2020-05-21	CA	Los Angeles County	06037	1180	6708	0.000651609703
22	2020-05-22	CA	Los Angeles County	06037	1015	6793	0.000659866534
23	2020-05-23	CA	Los Angeles County	06037	1003	6752	0.000655883827
24	2020-05-24	CA	Los Angeles County	06037	933	7014	0.000681334296
25	2020-05-25	CA	Los Angeles County	06037	1030	7567	0.000735052269
26	2020-05-26	CA	Los Angeles County	06037	1804	8249	0.000801201198

```
--4.   What is the weighted average household median income by County?
--                  What's the top 25 highest weighted average household median income
by population?
--                  What is the total population of these counties?
--                  What's the bottom 25 weighted average household median income by
population?
--                  What is the trailing 7-day COVID cases divide by population?
--                  What does the top 25 highest weighted average household median
income look compare to the bottom 25?


select top 25 state,
          county,
          SUM(population) total_population,
          coalesce(sum(population * household_median_income) /
nullif(sum(population),0),0) weighted_avg_income
from dc_geo_population
group by state, county
order by weighted_avg_income desc
```

```
select top 25 state,
          county,
          SUM(population) total_population,
          coalesce(sum(population * household_median_income) /
nullif(sum(population),0),0) weighted_avg_income
from dc_geo_population
where state != 'PR'
group by state, county
having coalesce(sum(population * household_median_income) / nullif(sum(population),0),0)
> 0
order by weighted_avg_income


go
with top_25_county as (
                                        select top 25 state,
                                                  county,
                                                  county_id,
                                                  SUM(population) total_population,
                                                  coalesce(sum(population *
household_median_income) / nullif(sum(population),0),0) weighted_avg_income
                                        from dc_geo_population
                                        where state != 'PR'
                                        group by state, county, county_id
                                        order by weighted_avg_income desc),

        top_25_cases as (
                                        select date,
                                                  state,
                                                  county,
                                                  top_25_county.county_id,
                                                  total_population,
                                                  weighted_avg_income,
                                                  daily_cases,
                                                  t7d_cases,
                                                  t7d_cases * 1.0 / total_population
cases_by_population
                                        from top_25_county
                                               left join dc_daily_covid
                                               on top_25_county.county_id =
dc_daily_covid.county_id
                                        ),
       top_25_totaled as (
                                        select 'Top 25 Counties' as covid_impact,
                                               date,
                                               sum(total_population) total_population,
                                               sum(total_population *
weighted_avg_income) / sum(total_population) weighted_avg_income,
                                               sum(t7d_cases) t7d_cases,
                                               sum(t7d_cases) * 1.0 /
sum(total_population) cases_by_population
                                        from top_25_cases
                                        group by date
       ),
       bottom_25_county as (
                                        select top 25 state,
                                                  county,
                                                  county_id,
```

```
                                                SUM(population) total_population,
                                                coalesce(sum(population *
household_median_income) / nullif(sum(population),0),0) weighted_avg_income
                                        from dc_geo_population
                                        where state != 'PR'
                                        group by state, county, county_id
                                        order by weighted_avg_income),

        bottom_25_cases as (
                                        select date,
                                                state,
                                                county,
                                                bottom_25_county.county_id,
                                                total_population,
                                                weighted_avg_income,
                                                daily_cases,
                                                t7d_cases,
                                                t7d_cases * 1.0 / total_population
cases_by_population
                                        from bottom_25_county
                                               left join dc_daily_covid
                                               on bottom_25_county.county_id =
dc_daily_covid.county_id
                                        ),
       bottom_25_totaled as (
                                        select 'Bottom 25 Counties' as covid_impact,
                                               date,
                                               sum(total_population) total_population,
                                               sum(total_population *
weighted_avg_income) / sum(total_population) weighted_avg_income,
                                               sum(t7d_cases) t7d_cases,
                                               sum(t7d_cases) * 1.0 /
sum(total_population) cases_by_population
                                        from bottom_25_cases
                                        group by date
       )
       select top_25_totaled.date,
                 top_25_totaled.cases_by_population cases_by_population_top_25_income,
                 bottom_25_totaled.cases_by_population
cases_by_population_bottom_25_income,
                 top_25_totaled.cases_by_population -
bottom_25_totaled.cases_by_population perc_delta

       from top_25_totaled
       join bottom_25_totaled

       on top_25_totaled.date = bottom_25_totaled.date
```

	state	county	total_population	weighted_avg_income
1	VA	Loudoun County	390359	143618.815556
2	VA	Arlington County	263721	135089.083656
3	TX	Hays County	27052	129645.794728
4	MD	Howard County	250942	128820.279686
5	VA	Alexandria city	66334	128628.040522
6	CA	San Mateo County	786170	127817.613326
7	CA	Santa Clara County	1913644	125876.116410
8	VA	Fairfax County	1245820	123120.462702
9	NJ	Morris County	550333	121818.730881
10	NM	Los Alamos County	18625	121324.000000
11	CO	Douglas County	313457	120753.046197
12	NY	Nassau County	1369611	119087.739613

	state	county	total_population	weighted_avg_income
1	SC	McCormick County	1437	6367.867084
2	VA	Sussex County	2666	8064.580645
3	KY	Wolfe County	327	14977.000000
4	KY	Monroe County	1722	16956.422764
5	SC	Barnwell County	147	17009.000000
6	TN	Putnam County	2948	17611.311397
7	KY	Owsley County	520	18917.000000
8	TN	Hancock County	1125	20065.706666
9	MS	Jefferson County	5290	20886.000000
10	GA	Clay County	5785	21748.852722
11	LA	East Carroll Parish	8687	22733.230804
12	AR	Phillips County	14553	23812.452209

	date	cases_by_population_top_25_income	cases_by_population_bottom_25_income	perc_delta
1	2020-05-05	0.001046083190	0.012347796261	-0.011301713071
2	2020-05-07	0.000971479211	0.014105642256	-0.013134163045
3	2020-05-14	0.000758538884	0.025595952666	-0.024837413782
4	2020-05-16	0.000713294761	0.026367689933	-0.025654395172
5	2020-05-23	0.000671500909	0.018646146040	-0.017974645131
6	2020-05-25	0.000668245936	0.022536044987	-0.021867799051
7	2020-05-30	0.000629772156	0.024675378898	-0.024045606742
8	2020-06-01	0.000654900547	0.022629535301	-0.021974634754
9	2020-06-03	0.000626647382	0.037660222955	-0.037033575573
10	2020-06-08	0.000466958412	0.030061375307	-0.029594416895
11	2020-06-10	0.000430307417	0.011273015740	-0.010842708323
12	2020-06-17	0.000337019893	0.013986889900	-0.013649870007
13	2020-06-19	0.000327971069	0.012400317314	-0.012072346245
14	2020-06-24	0.000376535264	0.014237401361	-0.013860866097
15	2020-06-26	0.000433236892	0.017243538891	-0.016810301999
16	2020-06-28	0.000463182643	0.019832157321	-0.019368974678
17	2020-07-03	0.000445801088	0.022295520020	-0.021849718932
18	2020-07-05	0.000426531648	0.022379023840	-0.021952492192
19	2020-07-12	0.000456672697	0.023339317773	-0.022882645076
20	2020-07-14	0.000489808321	0.027973779800	-0.027483971479
21	2020-07-19	0.000541171794	0.033109264748	-0.032568092954
22	2020-07-21	0.000541432191	0.032775249467	-0.032233817276
23	2020-07-23	0.000553150094	0.032107218905	-0.031554068811
24	2020-07-28	0.000526394217	0.046511627906	-0.045985233689
25	2020-07-30	0.000547616640	0.047012650828	-0.046465034188
26	2020-08-06	0.000559920438	0.039956578013	-0.039396657575

```
--5.   For ZCTA 85203, Arizona, what is the population broken out by age group?
--                  How does that compare to the State?

--- Heres the answere by population but this doesn't really tell us how 85203 compares.
We will have to normalize.
--- We can normalize this by taking the population of the repective location and divide
by the total population.
--- ie. zcta age group population divide by zcta total population.

select zcta,
         age_group,
         population zcta_popualtion,
         dbo.dc_County_AgeGroup(zcta,age_group) county_population,
         dbo.dc_State_AgeGroup(zcta, age_group) state_population
from dc_age_group
where zcta = '85203'
```

```
order by sort_key

-- Normalizing the value per the above.
--- ie. zcta age group population divide by zcta total population.
              select zcta,
                           age_group,
                           population * 1.0/ sum(population) over () zcta_perc,
                           dbo.dc_County_AgeGroup(zcta,age_group) * 1.0/
sum(dbo.dc_County_AgeGroup(zcta,age_group)) over () county_perc,
                           dbo.dc_State_AgeGroup(zcta, age_group) * 1.0/
sum(dbo.dc_State_AgeGroup(zcta, age_group)) over () state_perc
              from dc_age_group
              where zcta = '85203'
              order by sort_key

go
```

	zcta	age_group	zcta_popualtion	county_population	state_population
1	85203	Under 5 years	3241	282759	434649
2	85203	5 to 9 years	3264	291566	451200
3	85203	10 to 14 years	2903	307546	472429
4	85203	15 to 17 years	1760	180807	279414
5	85203	18 and 19 years	1225	115921	194500
6	85203	20 years	548	59131	102784
7	85203	21 years	258	59573	100655
8	85203	22 to 24 years	1762	174934	285223
9	85203	25 to 29 years	3654	330686	501843
10	85203	30 to 34 years	2885	308067	464121
11	85203	35 to 39 years	2455	291103	444010
12	85203	40 to 44 years	2401	282633	423540
13	85203	45 to 49 years	2504	282690	425403
14	85203	50 to 54 years	2370	272665	422843
15	85203	55 to 59 years	2423	262802	427914
16	85203	60 and 61 years	769	100101	171524
17	85203	62 to 64 years	1196	141200	248113
18	85203	65 and 66 years	606	87136	159150
19	85203	67 to 69 years	953	124696	227293
20	85203	70 to 74 years	986	170966	314562
21	85203	75 to 79 years	721	121053	229963
22	85203	80 to 84 years	512	77887	145588
23	85203	85 years and over	401	75138	131065

	zcta	age_group	zcta_perc	county_perc	state_perc
1	85203	Under 5 years	0.081438299369	0.064247931180	0.061584326869
2	85203	5 to 9 years	0.082016232379	0.066249040003	0.063929396555
3	85203	10 to 14 years	0.072945196874	0.069879983458	0.066937280331
4	85203	15 to 17 years	0.044224439028	0.041082602827	0.039589468992
5	85203	18 and 19 years	0.030781214664	0.026339336432	0.027558217265
6	85203	20 years	0.013769882151	0.013435626871	0.014563207215
7	85203	21 years	0.006482900721	0.013536057222	0.014261554544
8	85203	22 to 24 years	0.044274694072	0.039748151581	0.040412531635
9	85203	25 to 29 years	0.091815966027	0.075137807709	0.071104876231
10	85203	30 to 34 years	0.072492901474	0.069998364030	0.065760140644
11	85203	35 to 39 years	0.061688066939	0.066143838075	0.062910663485
12	85203	40 to 44 years	0.060331180742	0.064219301713	0.060010320516
13	85203	45 to 49 years	0.062919315526	0.064232253139	0.060274284315
14	85203	50 to 54 years	0.059552227554	0.061954392805	0.059911564334
15	85203	55 to 59 years	0.060883986230	0.059713341785	0.060630061608
16	85203	60 and 61 years	0.019323064552	0.022744747856	0.024302805440
17	85203	62 to 64 years	0.030052516521	0.032083179961	0.035154508793
18	85203	65 and 66 years	0.015227278438	0.019798866636	0.022549564410
19	85203	67 to 69 years	0.023946528632	0.028333174280	0.032204575202
20	85203	70 to 74 years	0.024775736864	0.038846550603	0.044569500973
21	85203	75 to 79 years	0.018116943488	0.027505419149	0.032582880807
22	85203	80 to 84 years	0.012865291353	0.017697327462	0.020627998638
23	85203	85 years and ...	0.010076136392	0.017072705211	0.018570271187

```sql
--     6.     What Counties had the highest deaths by percentage of population?
--                   The population must be over 1000 people.
--                   Include their household median income.

with population_counties as (
                                                select county_id,
                                                          county,
                                                          state,
                                                          sum(population) population,
                                                          sum(population *
household_median_income) / nullif(sum(household_median_income),0) household_median_income
                                                from dc_geo_population
                                                group by county_id, county, state
                                                )
       select top 25
                 population_counties.county_id,
                 population_counties.county,
                 population_counties.state,
                 population_counties.population,
                 population_counties.household_median_income,
                 total_deaths,
                 total_deaths * 1.0 / population_counties.population death_perc
       from population_counties
       join dc_covid_county
       on population_counties.county_id = dc_covid_county.county_id
       where population >= 10000
       and population_counties.county_id != 36061
       order by death_perc desc
```

	county_id	county	state	population	household_median_income	total_deaths	death_perc
1	18177	Wayne County	IN	10383	1801.440481	200	0.019262255610
2	01069	Houston County	AL	16062	2721.598410	269	0.016747603038
3	55135	Waupaca County	WI	10791	2277.563581	160	0.014827170790
4	01121	Talladega County	AL	12667	3387.807250	167	0.013183863582
5	01123	Tallapoosa County	AL	10956	5597.036792	144	0.013143483023
6	39141	Ross County	OH	10888	2465.719700	140	0.012858192505
7	45035	Dorchester County	SC	16814	3589.557216	201	0.011954323777
8	13277	Tift County	GA	11027	6392.198741	123	0.011154439104
9	13275	Thomas County	GA	10905	10905.000000	116	0.010637322329
10	18109	Morgan County	IN	13205	2966.830320	138	0.010450586898
11	04017	Navajo County	AZ	49546	3048.477072	517	0.010434747507
12	13299	Ware County	GA	16237	15441.000000	157	0.009669273880
13	13129	Gordon County	GA	11352	3775.749680	106	0.009337561663
14	19139	Muscatine County	IA	10073	2137.119821	94	0.009331877295
15	46103	Pennington County	SD	20799	3760.130833	190	0.009135054569
16	13297	Walton County	GA	26516	26516.000000	240	0.009051138934
17	08087	Morgan County	CO	10345	1510.783390	92	0.008893185113
18	48209	Hays County	TX	27052	9704.878311	235	0.008686973236
19	47179	Washington County	TN	27866	13522.094551	235	0.008433216105
20	13313	Whitfield County	GA	27557	20235.023611	231	0.008382625104
21	01055	Etowah County	AL	42335	7891.882736	339	0.008007558757
22	54029	Hancock County	WV	12278	2748.813009	95	0.007737416517
23	35031	McKinley County	NM	58951	3662.833996	454	0.007701311258
24	16083	Twin Falls County	ID	16476	4404.069994	126	0.007647487254
25	39077	Huron County	OH	14399	3676.020854	110	0.007639419404

```
---7.  Group the age groups that are considered the high-risk age categories for COVID.
That is age groups over 65 and under 5.
---                  Make this a calculated field that is a Boolean field being either 1
or 0; high risk or low risk.
---                  Total the population by high risk and low risk by county.
---                  Any counties that are greater than 30% being high risk what is the
number of COVID cases and deaths by total population vs low risk counties.
---                  Group these counties into 2 rows, high risk and low risk.  How do
they total cases and deaths compare to each other?
go
with risk_counties as (
                                        select dc_age_group.county,
                                                covid_age_group,
                                                SUM(dc_age_group.population)
population_age_group,
                                                coalesce(SUM(dc_age_group.population)
* 1.0 /

      nullif(SUM(SUM(dc_age_group.population)) over (Partition by
dc_age_group.county),0),0) perc_population
                                        from dc_age_group
                                        group by dc_age_group.county,
```

```
                                                    covid_age_group),
        county_population as (select county_id,
                                                    county,
                                                    sum(population) total_population
                                                    from dc_geo_population
                                                    group by county_id,
                                                                county
                                                    )

        select 'High Risk' as risk_type,
                  sum(total_population) total_population,
                  sum(total_cases) * 1.0 / sum(total_population) perc_cases,
                  sum(total_deaths) * 1.0 / sum(total_population) perc_deaths
        from county_population
        left join dc_covid_county
        on county_population.county_id = dc_covid_county.county_id
        where county_population.county in (select distinct county from risk_counties where
perc_population > 0.30 and covid_age_group = 'High Risk Age Group')

        union

        select 'Low Risk' as risk_type,
                  sum(total_population) total_population,
                  sum(total_cases) * 1.0 / sum(total_population) perc_cases,
                  sum(total_deaths) * 1.0 / sum(total_population) perc_deaths
        from county_population
        left join dc_covid_county
        on county_population.county_id = dc_covid_county.county_id
        where county_population.county in (select distinct county from risk_counties where
perc_population > 0.90 and covid_age_group = 'Low Risk Age Group')
```

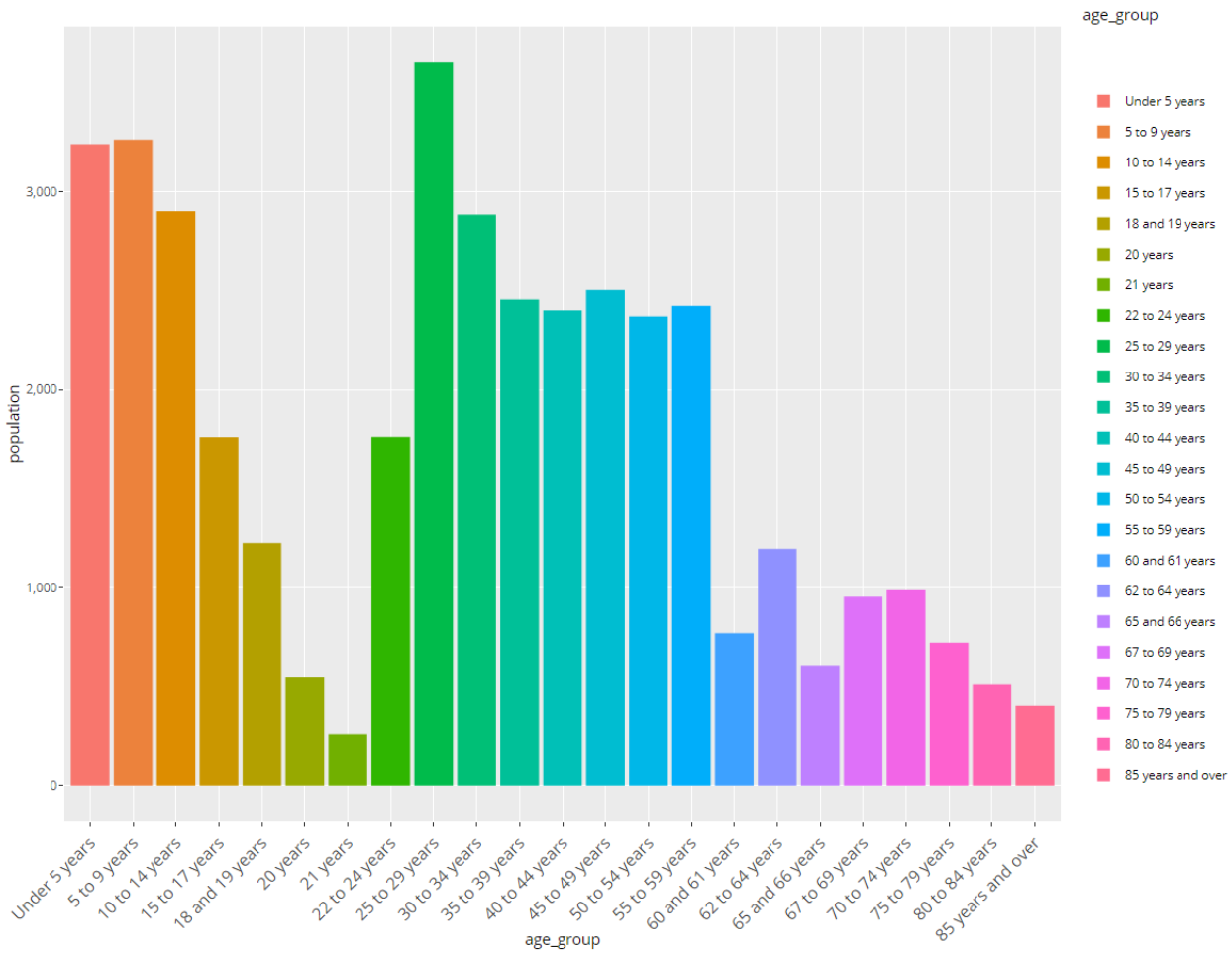
	risk_type	total_population	perc_cases	perc_deaths
1	High Risk	5205626	0.077999456741	0.001918693352
2	Low Risk	15538	1.066289097695	0.018856995752

```
--  8. How many Deaths Vs. Total Population
go
with county_population as (
                                                    select county_id,
                                                              sum(population) population
                                                    from dc_geo_population
                                                    group by county_id
                                                    )

select SUM(total_deaths) total_deaths,
          SUM(population) total_population,
          SUM(total_deaths) * 1.0 / SUM(population) death_perc
from dc_covid_county
join county_population

on dc_covid_county.county_id = county_population.county_id
```

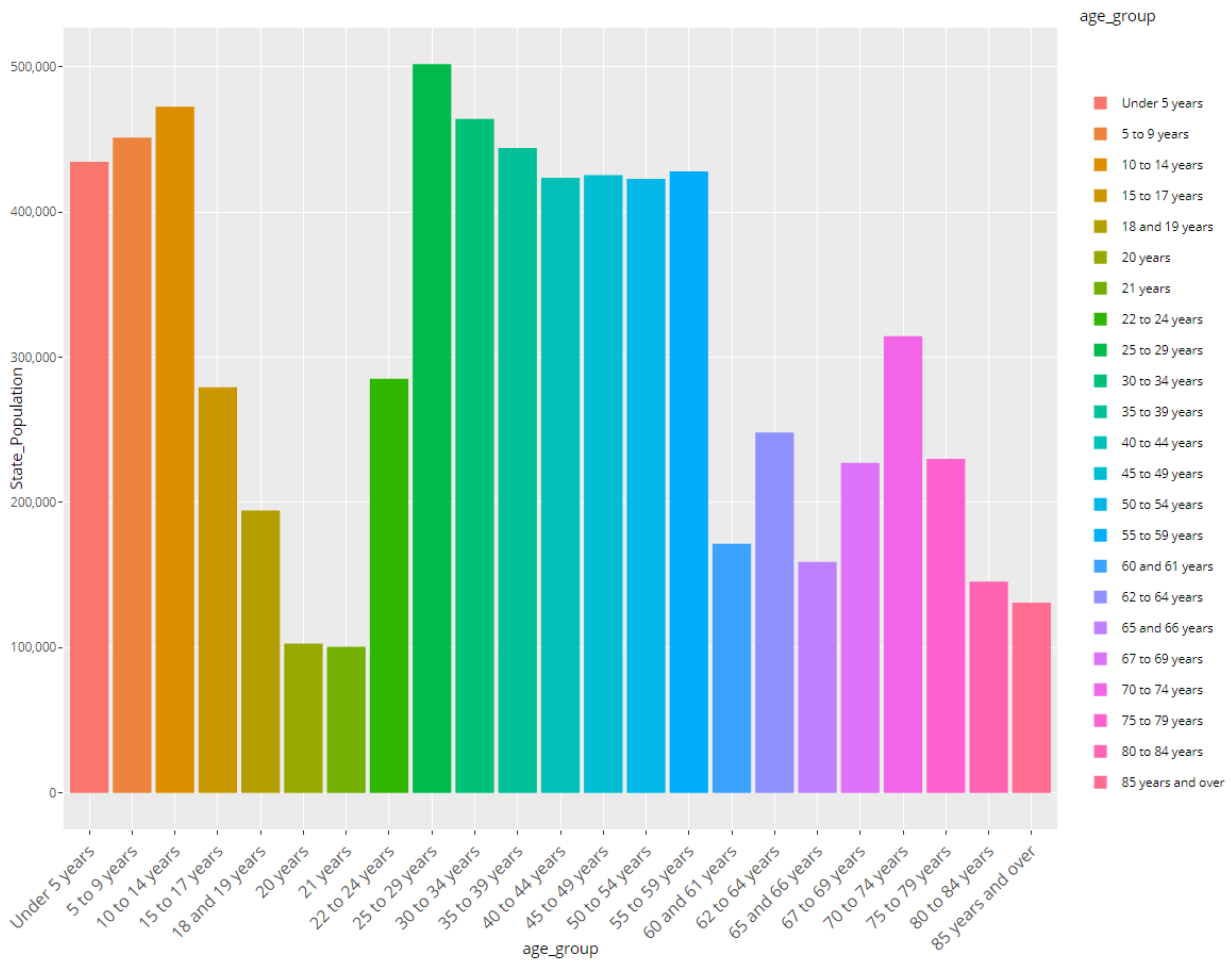
	total_deaths	total_population	death_perc
1	534385	321182638	0.001663804131

Data Visualizations:

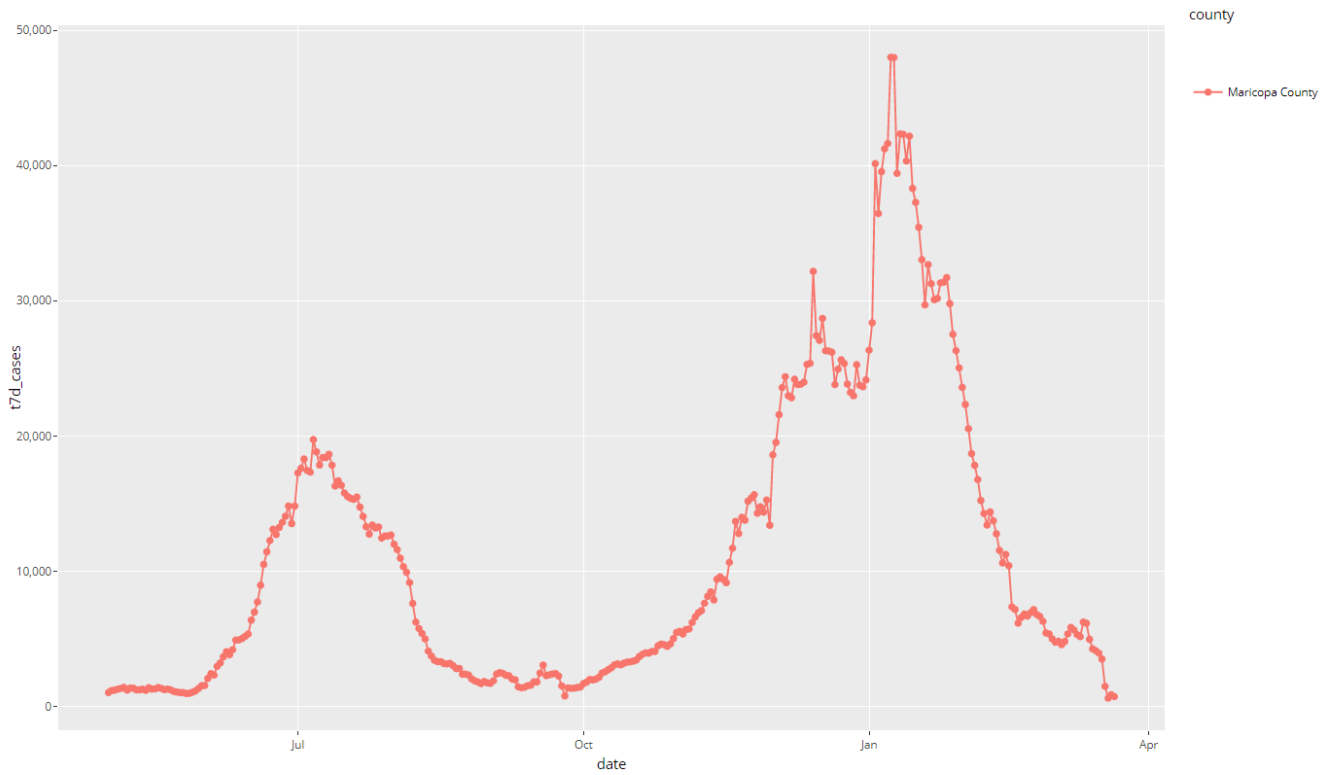
Population by ZCTA and Age Group



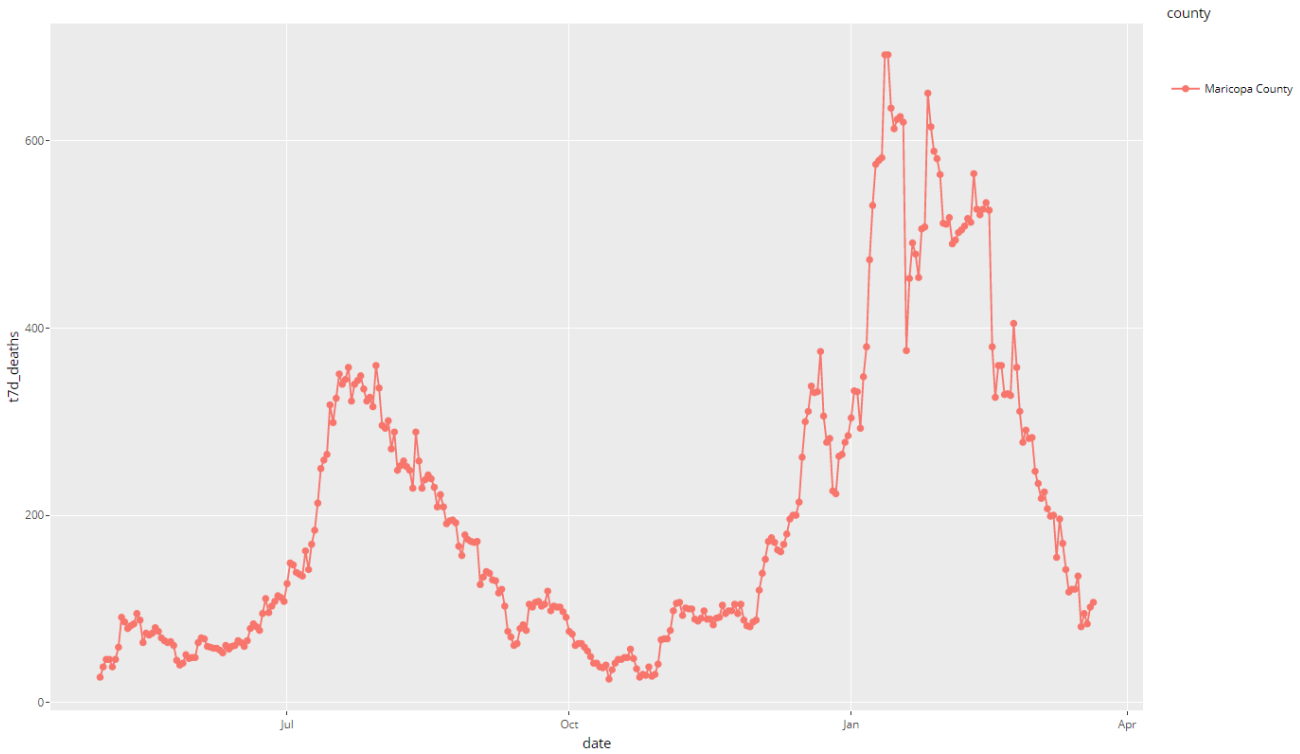
Population by State and Age Group



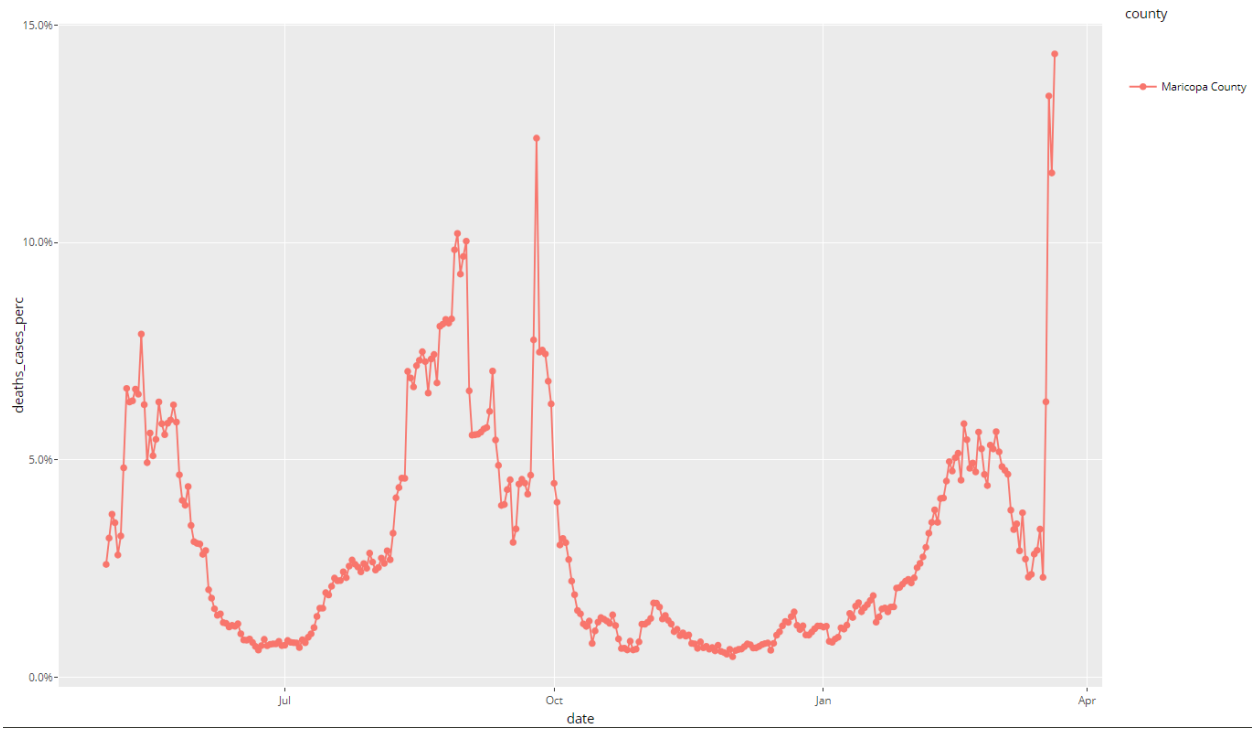
Daily COVID Cases for Maricopa County



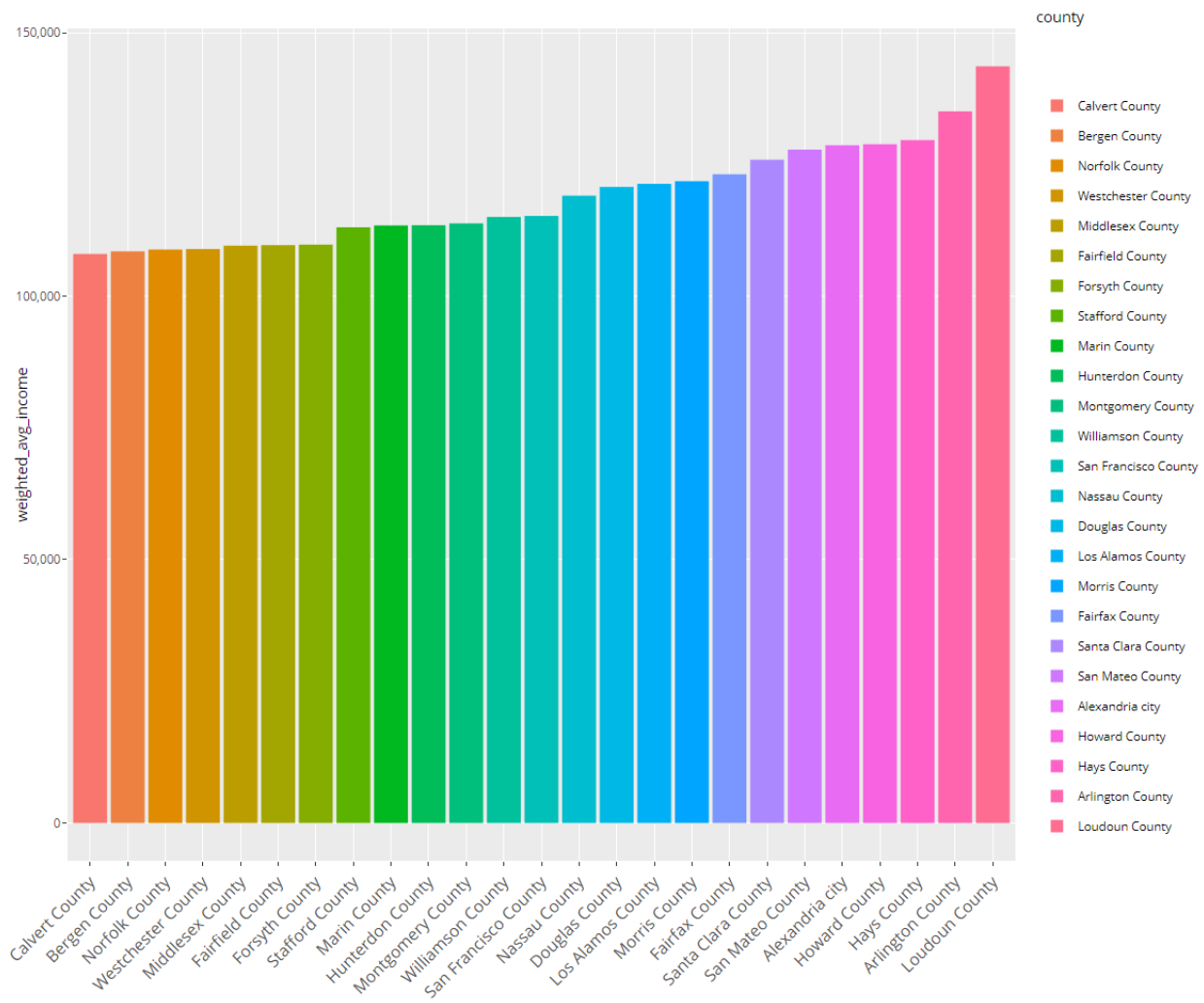
Daily COVID Deaths for Maricopa County



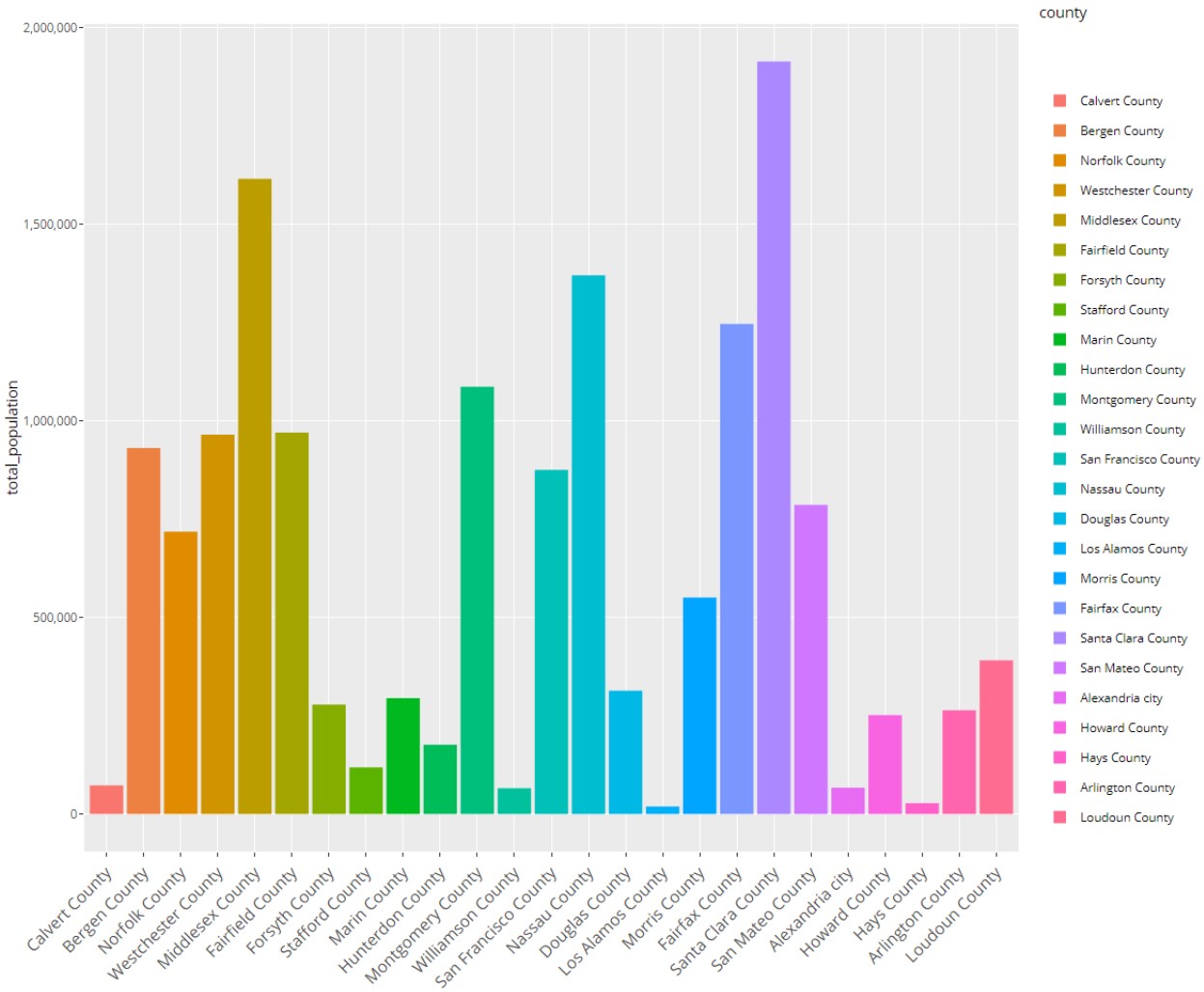
Daily COVID Deaths per cases for Maricopa County



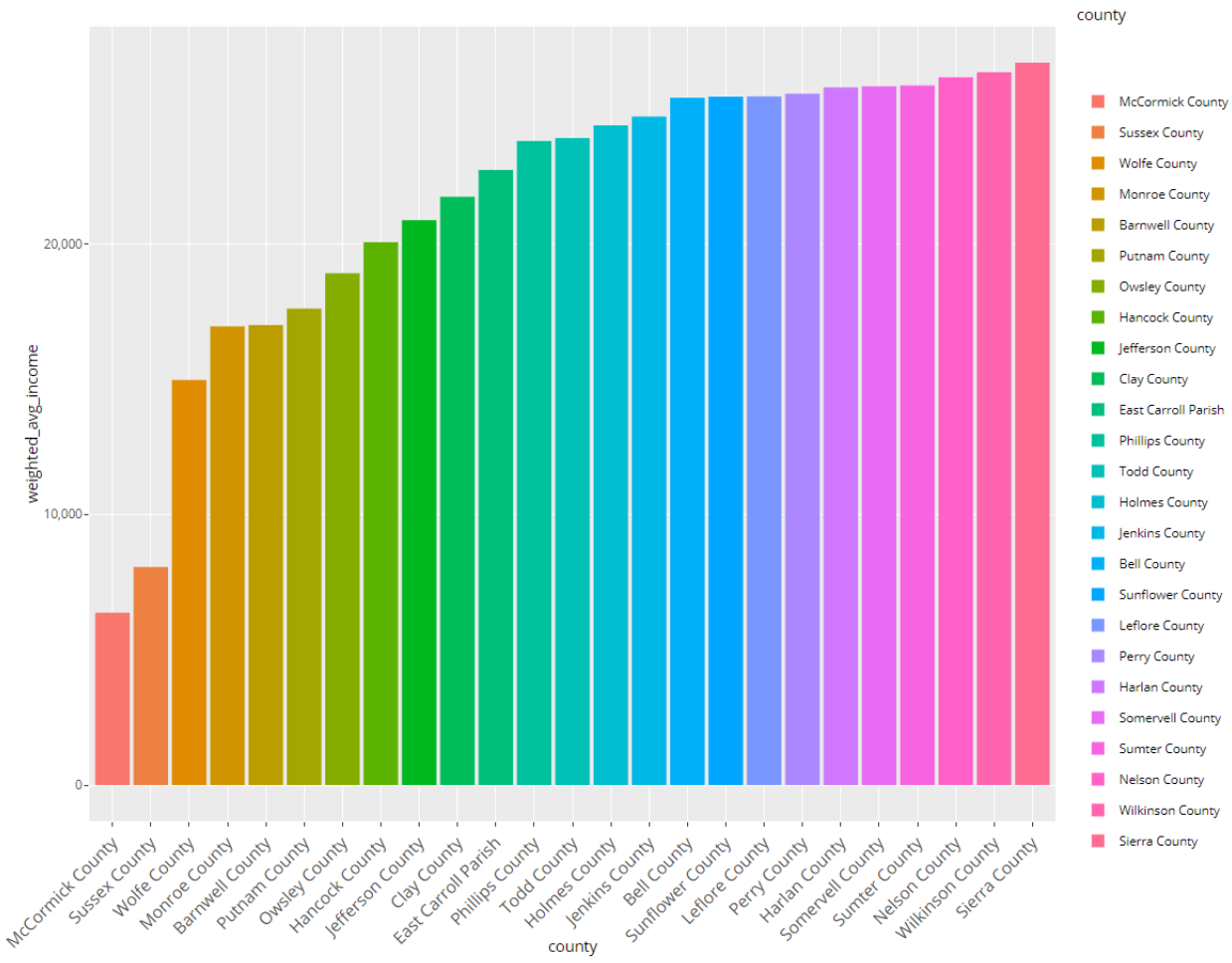
Top 25 Household Median Income by County (Image 1) & Population (Image 2)



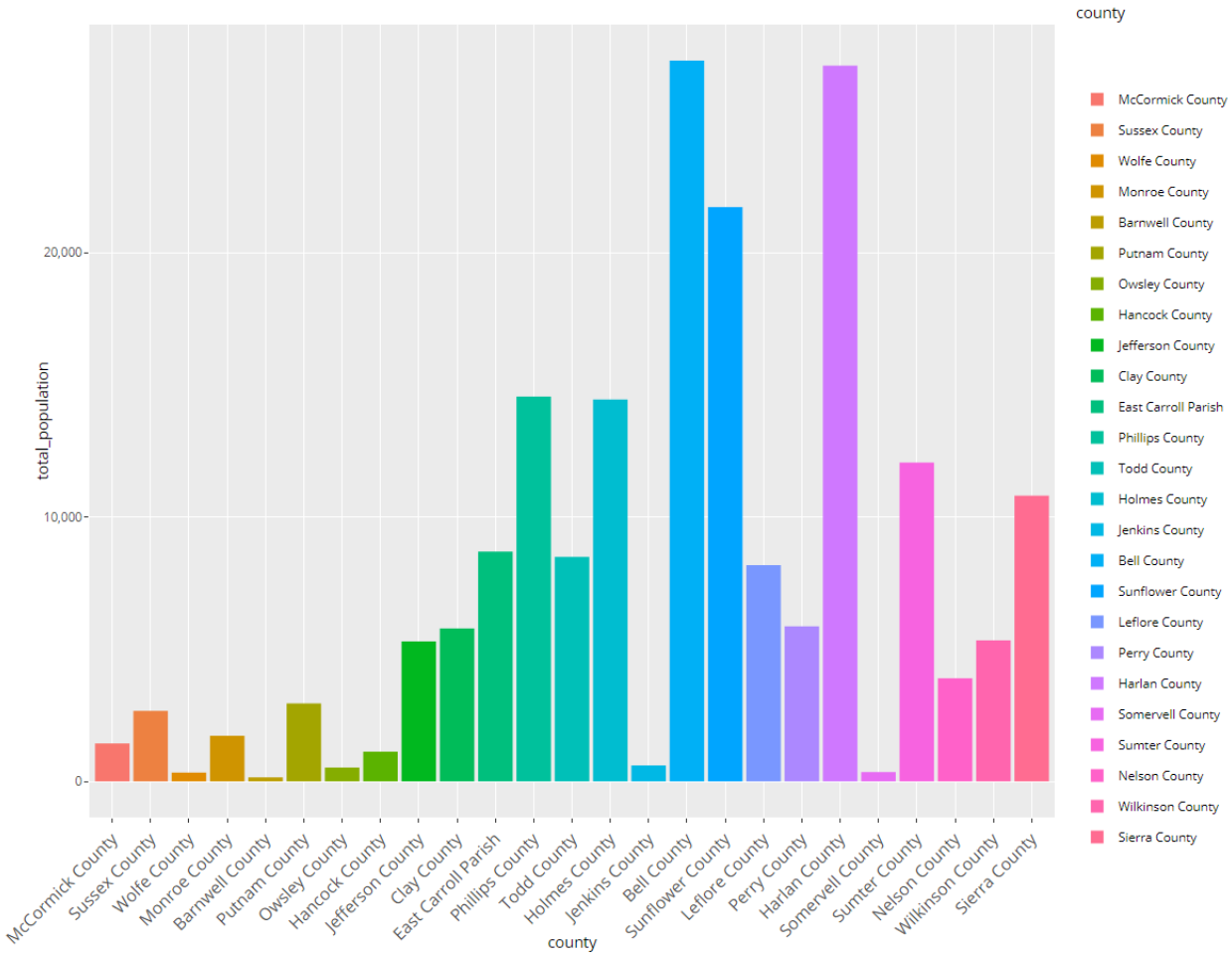
The Population of Top 25 Household Median Income sorted by income.



Bottom 25 Household Median Income by County (Image 1) & Population (Image 2)



The Population of Top 25 Household Median Income sorted by income.



Reflection:

I have worked with databases now for 9 years and at first, I was under the impression that I would only get a little out of this class. In short, I was wrong. From Logical and Conceptual Models to Normal Form and then creating functions and procedures, I'm able to really approach database tasks and problems more efficiently and intelligently. For example, I would have skipped the modeling phase, dodged the normal form exercise and begin grabbing data to put into tables prior to this class. I would have ended up spending an exorbitant amount of time fixing my table, making mistakes when doing analytics, and had to update multiple tables with the same data. Something I could have done better is understanding the COVID data. What I discovered later on is that the cases and deaths where a cumulative number. I had to transform that data, drop the old and insert the new records after discovering this. How I built my database made it easy to perform this task. The COVID data was by county but broke it's own rule by combining New York City Counties into one County called New York City. I could have created an additional table that stored the county codes associate to New York County, then joined up on this table to bring back to the census data. The last thing I could have done better, which I started to do, was create more views. For example, I created a zcta that had county, state as the dimensions and household median income as the metrics. I also did this for age group. I should have created a few more views. One for Race and then an aggregated view on county and state. This view would have made it easy to perform analytics on. Rather than building long queries users can just query these tables.

Summary:

In building this database I had an objective to use census data to help answer questions surrounding the impact of COVID on our society. The lenses I was exploring was from a socioeconomics, age, and race perspective. I thought by weighting cases and deaths by population I could discern what the impact looked like for each of these groups. I will say there's probably a more intelligent way to discern this information and I'm excited to continue my data discovery exercise outside of this class. I'll probably implement the views I discussed in my reflection to make the analysis easier to perform. Deploy functions to aggregate metrics in a more straightforward method. Grow my understanding of procedures and implement them to help manage my data. I cannot help to be thankful of everything I learned in this class and ready to apply my new skills.