

Stage 1: Analysis on US COVID-19 Cases and Deaths by State

Three tables are used by this source (<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>) and the variable dictionaries for each table detailed below

covid_confirmed_usafacts:

This dataset keeps track of the number of confirmed COVID-19 cases logged on a particular day using countyFIPS as the UID. countyFIPS ranges from 1000 to 56045 as of the time of last capture.

Name	Definition	Data Type	Possible Values	Required
countyFIPS	UID for each county	Integer, [1000, 56045]	1000, 56045, 32503	Yes
countyName	Name of county	String	Burke, Buncombe, Guilford	Yes
State Name	Abbreviation of State	String	NC, SC, CA, MI	Yes
StateFIPS	UID for each state	Integer, [1,56]	1-56	Yes
Date (xx/xx/xxxx)	Number of cases reported on this day; month/day/year	Integer, non-negative	0, 45, 100	Yes

covid_county_population:

This simple dataset presents the population of each count at one moment in time. The data captured in this project is projected from February 9, 2022. countyFIPS is the UID which ranges from 1000 to 56045 as of the moment of this data capture.

Name	Definition	Data Type	Possible Values	Required
countyFIPS	UID for each county	Integer, [1000, 56405]	1000, 3500, 56045	Yes
County Name	Name of county	String	Burke County, Buncombe County, Guilford County	Yes
State	Abbreviation of State	String	FL, MI, TX, CA	Yes
population	Total population of a county	Integer, non-negative	0, 2705, 32303	Yes

covid_deaths_usafacts:

This dataset keeps track of the number of confirmed COVID-19 deaths logged on a particular day using countyFIPS as the UID. countyFIPS ranges from 1000 to 56045 as of the time of last capture.

Name	Definition	Data Type	Possible Values	Required
countyFIPS	UID for each county	Integer, [1000, 56045]	1000, 56045, 21086	Yes
County Name	Name of county	String	Guilford, Wake, Harnett	Yes
State	Abbreviation of State	String	FL, GA, NY, KY	Yes
StateFIPS	UID for each state	Integer, [1,56]	1-56	Yes

Date (xx/xx/xxxx)	Number of deaths reported per day	Integer, non-negative	0, 10, 102	Yes
----------------------	---	--------------------------	------------	-----

Preliminary intuitions:

It is hypothesized that as time continues the number of confirmed cases and deaths will increase. We can also guess that areas with high population will have higher cases and deaths as well. However as we hit the time where vaccinations are introduced, the correlation between cases and deaths may fall depending on the area.

Enrichment Data Sets

Provided below are data dictionaries of supporting data sets that detail characteristics of US Counties. All of these sets will be used in conjunction with the ones above to draw conclusions on Covid 19 in the United States.

ACS Demographic and Housing:

This data details demographic data regarding counties in the US. This includes data regarding racial classification, age and sex classification, and housing costs.

Name	Definition	Data Type	Possible Values	Required
GEO_ID	The geographic id for every county present in the sample	int	0500000US01003	yes
NAME	Name of the county in the form county,state	string	Baldwin County, Alabama	yes
Demographic info (DP05_XXXXX)	Numbers regarding the counties demographic	int	16875	no

How to merge:

A possible approach could be masking the GEO_ID to the last 5 numbers, which is the FIPS. This would need to be a left join containing all the FIPS from the merged covid dataset and only the matching of the ACS Demographic set as this file does not contain the same counties.

Analysis potential:

This data could allow us to find patterns regarding a specific demographic and covid cases. A potential hypothesis could be, Do states with higher percentages of 19-24 year olds have higher positive case rates?

Presidential Elections by County:

This data lists the total number of votes for Presidential candidates per county. It also makes note of the candidate's political party and if they went on to win the Presidency.

Name	Definition	Data Type	Possible Values	Required
state	This field contains the name of the State in which the vote is being recorded.	string	North Carolina, West Virginia, Nevada, Florida	yes
county	This field contains the name of the county in which the vote is being recorded.	string	Wake county, Guilford county, Orange county	yes
candidate	This is the name of the Presidential candidate.	string	Joe Biden, Donald Trump, Rush Limbaugh	No
party	This is the associated candidate's political party affiliation.	string	DEM, REP, GOP, LIB, GRN	yes
total_votes	This number is	integer	1, 200, 3693,	yes

	the total number of votes for the associated candidate.		NaN	
won	This indicates whether or not the associated candidate went on to win the Presidential race.	boolean	True, False	no

How to Merge:

The most secure way to merge this data set is to add a column for all state abbreviations. I would do this by using the map function to add the corresponding abbreviation to each appropriate row. Next I will drop the 'candidate' and 'won' columns because they are unnecessary information. I will then unstack the party column and have each row contain the complete list of possible candidates with their total votes under their respective columns. After the data preparations are complete I will merge on state abbreviations and counties to ensure that the data doesn't lose integrity.

Analysis Potential:

This data could shed light on if peoples' political affiliations have an effect of Covid-19 trends. An early hypothesis would be, Did counties that voted overwhelmingly for the republican candidate have had higher Covid cases and deaths?

ACS Social, Economic, and Housing:

Social:

This dataset provides estimates about the social statuses of people living in the US by county in 2019.

Name	Definition	Data type	Possible Values	Required?
Label	Name of county and state	String	Baldwin County, Alabama, Calhoun County, Alabama	Yes
HOUSEHOLDS BY TYPE	Number of households by type per county	Integer	82325, 44636	No
RELATIONSHIP	Population in households by relationship per county	Integer	220247, 110627	No
MARITAL STATUS	Population of county by marital status	Integer	88707, 96712	No
FERTILITY	Number of women in the county who had a birth in the past 12 months	Integer	1402, 1508	No
GRANDPARENTS	Number of grandparents in the county	Integer	6891, 2595	No

SCHOOL ENROLLMENT	Number of students in the county by school enrollment level	Integer	47583, 26109	No
EDUCATIONAL ATTAINMENT	Population of county by educational attainment level	Integer	159717, 79084	No
VETERAN STATUS	Number of veterans in the county	Integer	176331, 87525	No
DISABILITY STATUS OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION	Number of people with a disability in the county	Integer	220911, 111075	No
RESIDENCE 1 YEAR AGO	Population of county by type of residence 1 year ago	Integer	221737, 112208	No
PLACE OF BIRTH	Population of county by place of birth	Integer	223234, 113605	No
U.S. CITIZENSHIP STATUS	Population of county by U.S. citizenship status	Integer	11002, 1906	No

YEAR OF ENTRY	Population of county by year of entry	Integer	12720, 2787	No
WORLD REGION OF BIRTH OF FOREIGN BORN	Number of foreign born in the county by world region of birth	Integer	668288, N	No
LANGUAGE SPOKEN AT HOME	Population of county by language spoken at home	Integer	212618, N	No
ANCESTRY	Population of county by ancestry	Integer	223234, 113605	No
COMPUTERS AND INTERNET USE	Population of county by computers and internet use	Integer	82325, 44636	No

Economic:

This dataset provides estimates about the economic statuses of people living in the US by county in 2019.

Name	Definition	Data type	Possible values	Required?
Label	Name of county and state	String	Baldwin County, Alabama, Calhoun County, Alabama	Yes

EMPLOYMENT STATUS	Population of county by employment status	Integer	183875, 101561	No
COMMUTING TO WORK	Population of county by type of commute	Integer	94867, 46169	No
OCCUPATION	Population of county by occupation	Integer	96012, 45641	No
INDUSTRY	Population of county by industry	Integer	96012, 45641	No
CLASS OF WORKER	Population of county by class of worker	Integer	96012, 45641	No
INCOME AND BENEFITS (IN 2019 INFLATION-ADJUSTED DOLLARS)	Population of county by income and benefits	Integer	82325, 53962	No
HEALTH INSURANCE COVERAGE	Population of county by health insurance coverage	Integer	220911, 49101	No

PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL	Percentage of people whose income in the past 12 months was below the poverty level	Double	(X)	No
---	---	--------	-----	----

Housing:

This dataset provides estimates about the housing accommodations of people living in the US by county in 2019.

Name	Definition	Data type	Possible values	Required?
Label	Name of county and state	String	Baldwin County, Alabama, Calhoun County, Alabama	Yes
HOUSING OCCUPANCY	Number of housing units by occupancy per county	Integer	119425, 53809	No
UNITS IN STRUCTURE	Number of housing units by structure per county	Integer	119425, 53809	No
YEARS STRUCTURE BUILT	Number of housing units	Integer	119425, 53809	No

	by year built per county			
ROOMS	Number of housing units by rooms per county	Integer	119425, 53809	No
BEDROOMS	Number of housing units by bedrooms per county	Integer	119425, 53809	No
HOUSING TENURE	Number of occupied housing units by housing tenure per county	Integer	82325, 44636	No
YEAR HOUSEHOLDER MOVED INTO UNIT	Number of occupied housing units by year householder moved into unit per county	Integer	82325, 44636	No
VEHICLES AVAILABLE	Number of occupied housing units by vehicles available per county	Integer	82325, 44636	No

HOUSE HEATING FUEL	Number of occupied housing units by house heating fuel per county	Integer	39761, N	No
SELECTED CHARACTERISTICS	Number of occupied housing units by selected characteristics per county	Integer	82325, 44636	No
OCCUPANTS PER ROOM	Number of occupied housing units by occupants per room per county	Integer	39761, N	No
VALUE	Number of owner-occupied units by value per county	Integer	65297, 4191	No
MORTGAGE STATUS	Number of owner-occupied units by mortgage status per county	Integer	65297, 31479	No
SELECTED MONTHLY OWNER COSTS (SMOC)	Number of housing units by SMOC per county	Integer	36074, 29223	No

SELECTED MONTHLY OWNER COSTS AS A PERCENTAGE OF HOUSEHOLD INCOME (SMOCAP)	Number of housing units by SMOCAP per county	Integer	36074, 28696	No
GROSS RENT	Number of occupied units paying rent by gross rent per county	Integer	12317, N	No
GROSS RENT AS A PERCENTAGE OF HOUSEHOLD INCOME (GRAP)	Number of occupied units paying rent by GRAP per county	Integer	14380, 2241	No

How to merge:

I am going to merge my enrichment datasets with our COVID-19 dataset by using the county name and state name as the index. Before I merge, I would have to convert the state abbreviations to their full names in the COVID-19 dataset.

Analysis potential:

My enrichment data could be helpful by bringing insight into the social, economic, and housing statuses of people in the US and seeing if there is a correlation with the number of COVID-19 cases. For example, I could test if counties with higher numbers of occupants per housing unit are correlated with higher numbers of COVID-19 cases.

Employment:

Below is a variable dictionary from the provided csv regarding employment information within different counties. The given dataset focused on the level of employment as well as potential earnings by county.

Name	Definition	Data type	Possible values	Required?
Area Code	UID for area	Integer	1000, 5000, 7500	Yes
St	UID for state	Integer [1-56]	1, 25, 50	Yes
Cnty	UID for county	Integer	1, 500, 7500	Yes
Own	Ownership code	Integer [0-5]	1, 3, 5	Yes
NAICS	UID for industry	Integer	10, 1000, 5000	Yes
Year	Year when data was recorded	Integer, non-negative	2020	Yes
Qtr	Quarter data was recorded	String	A	Yes
Area Type	Defines the type of area where data was collected	String	Nation, State, County	Yes
St Name	Name of State	String	North Carolina, Texas, Arizona	Yes
Area	Name of County	String	Wake, Rowan, Mecklenburg	Yes
Ownership	Ownership title associated with ownership code	String	Federal, State, Private	Yes

Industry	Industry title associated with industry code	String	Healthcare, Education, Construction	Yes
Annual Average Status Code	Status code or disclosure code (related to industry)	String	Healthcare, Education, Construction	Yes
Annual Average Establishment Count	Quarterly establishment counts	Integer	100, 500, 1000	Yes
Annual Average Employment	Annual average of monthly employment levels	Integer	5000, 10000, 30000	Yes
Annual Total Wages	Sum of four quarterly total wages	Integer	5000, 10000, 50000	Yes
Annual Average Weekly Wage	Average weekly wage based on 12-month employment levels	Integer	500, 1000, 3000	Yes
Annual Average Pay	Average pay each year	Integer	500, 1000, 3000	Yes
Employment Location Quotient Relative to U.S.	Comparing concentration in employment in one area compared to another	Float	1.0 3.0 5.0	Yes

Total Wage Location Quotient Relative to U.S.	Comparing concentration in wages in one area compared to another	Float	1.0 3.0 5.0	Yes
---	--	-------	-------------	-----

How can you merge the data with the primary COVID-19 dataset. Identify the individual variable which map between the datasets.

- A potential solution would be mapping a connection between the county name as well as the state name between the two data sets.

Describe how enrichment data can help in the analysis of COVID-19 spread.

- The provided enrichment dataset can help determine if COVID-19 spreads more easily among those that are in white/blue collar jobs. It can also help determine if cases were higher in counties with those that have more manual labor jobs compared to office jobs, where workers were allowed to work from home.

Pose initial hypothesis questions.

- Because many white collared workers were allowed to work from home during the pandemic, were blue collar workers more susceptible to COVID-19 infection?
- Did more rural counties (where there are less office jobs) see higher COVID-19 cases compared to (where there are more office jobs and larger employers) urban counties?

Hospital Beds Dataset

The Hospital Utilization dataset (<https://protect-public.hhs.gov/pages/hospital-utilization>) contains two main tables: "Reported State Hospital Capacity and COVID19 Patient Impact", as well as "Weekly Hospital Capacity." For the purposes of this project, we will exclude the "...COVID19 Patient Impact" because it is too granular and does not cover specific counties.

Name	Definition	Data Type	Possible Values	Required
X	Longitudinal measurement of East-West	Float	100.555555555555	No

Y	Latitudinal measurement of North-South	Float	-34.4892097189374	No
Hospital PK	Unique identifier of a particular hospital	String of 6 digits	610663, 010024, 522158	Yes
Collection Week	Date of last data collection	Datetime formatted "m/d/y, 0:00 AM/PM"	10/14/2021, 9:00 PM	No
State	Abbreviation of Hospital's state	String	CA, MI, TX	No
CCN	Number used to verify Medicare/Medicaid certification for survey and certification, assessment-related activities and communications (Center for Medicare & Medicare Services)	Integer	10118, 49433, 19222	No
Hospital Name	Name of Hospital	String	COMMUNITY HOSPITAL INC,	No
Address	Hospital's street address	String	509 WILSON AVENUE	No
City	Hospital's city	String	MORGANTON, CHARLOTTE	No
Hospital Subtype	Describes the hospital's domain of work	String	Short Term, Childrens Hospital, Long Term	No

FIPS Code	ID of the hospital's county	Integer	1001, 50000, 12055	No
Is Metro Micro	Details the size of the hospital	Boolean	True, False	No
Total Beds 7 Day Average	Average number of beds regularly maintained and staffed over a 7 day average (OECD Glossary of Statistical Terms)	Integer	6, 7, 0120	No
Average of total number of all staffed inpatient and outpatient beds		Integer		No
Average of all staffed inpatient and outpatient adult beds		Integer		No
Average of total number of staffed inpatient adult beds		Integer		No
Average of total number of staffed inpatient beds that are occupied		Integer		No
Average number of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed or suspected COVID19		Integer		No
Average number of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19		Integer		No

Average number of patients currently hospitalized in a pediatric inpatient bed who are suspected or laboratory-confirmed positive for COVID-19		Integer		No
Average number of patients currently hospitalized in a pediatric inpatient bed who have laboratory-confirmed COVID-19		Integer		No
Average number of total number of staffed inpatient beds		Integer		No
Average of total staffed inpatient ICU beds		Integer		No
Average of total number of staffed adult ICU beds		Integer		No
Average of total number of staffed adult ICU beds that are occupied		Integer		No
Average of total number of staffed inpatient adult ICU beds that are occupied		Integer		No
Average number of patients currently hospitalized in a designated adult ICU bed who have suspected or laboratory-confirmed COVID-19		Integer		No
Average number of patients currently		Integer		No

hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19				
Average number of patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed influenza		Integer		No
Average of patients (all ages) currently hospitalized in a designated ICU bed with laboratory-confirmed influenza		Integer		No
Average number of patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed COVID-19 and laboratory-confirmed influenza reported in the 7-day period.		Integer		No
Sum of reports of total number of all staffed inpatient and outpatient beds in the hospital		Integer		No
Sum of reports of all staffed inpatient and outpatient adult beds in the hospital		Integer		No

Sum of reports of all staffed inpatient adult beds in the hospital		Integer		No
Sum of reports of total number of staffed inpatient beds that are occupied		Integer		No
Sum of reports of total number of staffed inpatient adult beds that are occupied		Integer		No
Sum of reports of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed or suspected COVID19		Integer		No
Sum of reports of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19		Integer		No
Sum of reports of patients currently hospitalized in a pediatric inpatient bed who are suspected or laboratory-confirmed-positive for COVID-19		Integer		No
Sum of reports of patients currently hospitalized in a pediatric inpatient bed who have laboratory-confirmed COVID-19		Integer		No

Sum of reports of total number of staffed inpatient beds		Integer		No
Sum of reports of total number of staffed inpatient ICU beds		Integer		No
Sum of reports of total number of staffed inpatient adult ICU beds		Integer		No
Sum of reports of total staffed inpatient ICU beds occupied		Integer		No
Sum of reports of total number of staffed inpatient adult ICU beds that are occupied		Integer		No
Sum of reports of patients currently hospitalized in a designated adult ICU bed who have suspected or laboratory-confirmed COVID-19		Integer		No
Sum of reports of patients currently hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19		Integer		No
Sum of reports of patients currently hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19. Including patients who have both		Integer		No

laboratory-confirmed COVID-19 and laboratory-confirmed influenza				
Sum of reports of patients (all ages) currently hospitalized in a designated ICU bed with laboratory-confirmed influenza		Integer		No
Sum of reports of patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed COVID-19 and laboratory-confirmed influenza		Integer		No
Number of times in the 7 day period that the facility reported total number of all staffed inpatient and outpatient beds in your hospital		Integer		No
Number of times in the 7-day period that the facility reported total number of all staffed inpatient and outpatient adult beds		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient adult beds		Integer		No

Number of times in the 7-day period that the facility reported total number of staffed inpatient beds that are occupied		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient adult beds that are occupied		Integer		No
Number of times in the 7-day period that the facility reported patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed or suspected COVID19		Integer		No
Number of times in the 7-day period that the facility reported patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19		Integer		No
Number of times in the 7-day period that the facility reported Patients currently hospitalized in a pediatric inpatient bed who are suspected or laboratory-confirmed-positive for COVID-19		Integer		No
Number of times in the 7-day period that the		Integer		No

facility reported patients currently hospitalized in a pediatric inpatient bed who have laboratory-confirmed COVID-19				
Number of times in the 7-day period that the facility reported total number of staffed inpatient beds		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient ICU beds		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient adult ICU beds		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient ICU beds		Integer		No
Number of times in the 7-day period that the facility reported total number of staffed inpatient adult ICU beds that are occupied		Integer		No
Number of times in the 7-day period that the facility reported patients currently hospitalized in a designated adult ICU bed who have suspected		Integer		No

or laboratory-confirmed COVID-19				
Number of times in the 7-day period that the facility reported patients currently hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19		Integer		No
Number of times in the 7-day period that the facility reported patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed influenza		Integer		No
Number of times in the 7-day period that the facility reported patients (all ages) currently hospitalized in a designated ICU bed with laboratory-confirmed influenza		Integer		No
Number of times in the 7-day period that the facility reported patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed COVID-19 and laboratory-confirmed influenza		Integer		No
Sum of number of patients who were		Integer		No

admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19				
Sum of number of patients age 18-19 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age 20-29 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age 30-39 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age 40-49 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age 50-59 who were admitted to an adult inpatient bed on the previous calendar		Integer		No

day who had confirmed COVID-19				
Sum of number of patients age 60-69 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age 70-79 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients 80 or older who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of patients age unknown who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No
Sum of number of pediatric patients who were admitted to an inpatient bed on the previous calendar day who had confirmed COVID-19		Integer		No

Sum of total number of ED visits who were seen on the previous calendar day who had a visit related to COVID-19		Integer		No
Sum of number of patients who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 18-19 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 20-29 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 30-39 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 40-49 who were admitted to an adult inpatient bed on the previous calendar		Integer		No

day who had suspected COVID-19				
Sum of number of patients age 50-59 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 60-69 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age 70-79 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients 80 or older who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of number of patients age unknown who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No

Sum of number of pediatrics patients who were admitted to an inpatient bed on the previous calendar day who had suspected COVID-19		Integer		No
Sum of total number of patient visits to the ED who were seen on the previous calendar day regardless of reason for visit		Integer		No
Sum of number of patients (all ages) who were admitted to an inpatient bed on the previous calendar day who had laboratory-confirmed influenza		Integer		No
zip_code		String of 5 digits	28655, 27412, 27553	No
Last Updated		Datetime formatted "m/d/y, 0:00 AM/PM"	9/6/2021, 8:00 PM , 2/15/2022, 3:06 PM , 10/14/2021, 9:00 PM	No

[Merging these datasets](#)

The primary key formulated in Team 2's Primary COVID-19 dataset, labeled "UID", uses the state abbreviation appended to the "CountyFIPS". This is to provide a way to identify each county uniquely. In order to merge the super-table to these smaller tables one must be able to find commonalities between the datasets.

[Merging Weekly Hospital Capacity with Primary COVID-19 Dataset](#)

The primary key of the Weekly Hospital Capacity dataset is called "Hospital PK." However, the Primary COVID-19 dataset does not include a reference to a particular

hospital, just each county. The common variables between the two datasets is "State," the abbreviated state name, as well as the county FIPS code ("countyFIPS_cases" in Primary COVID-19 dataset; "FIPS Code" in Weekly Hospital Capacity Dataset). Combining these two datasets emerges a possible correlation between the recent hospital trends/conditions with the total deaths/cases count of the last week. This could lead to correlating the population total of a given area with the specific hospital conditions of the last recorded week.