COVID-19 Dataset description (web).

We will utilize the data from usfacts.org. The dataset contains daily county-level tracker of COVID-19 cases. This makes it easy to follow COVID-19 cases on a granular level, as does the ability to break down infections per 100,000 people (with the population data). The underlying data is available for download below the US county map and has helped government agencies like the Centers for Disease Control and Prevention in its nationwide efforts.

COVID-19 Spread Map tracks the hotspots, cases, and deaths of the disease across America. We can set the state data to focus on or look at national statistics. On the summary page it tracks and graphs total cases updated to the most recent date (approx. 2-day gap to current day). It also tracks first and full vaccination status, delivering a total and a national percentage. There are then 4 other tabs after summary: cases and deaths, vaccination progress, impact and recovery, and related articles.

On cases and deaths it gives a total stat, a recent daily stat, and a 7-day average. It can be viewed in 1, 3, and 6 month intervals. You can also view hotspots, per 100k stat, and total. The vaccination progress tab accounts for distributed doses and doses used, and see one dose vs fully fax status percentage by state. It also shows demographic data by sex, age, and race/ethnic group. On impact and recovery we are given stats for various stats for economic, standard of living, health, and government growth and decline. In the economic domain are stats for unemployment claims, inflation rate, S&P 500, air travel passengers, retail sales, and unemployment rate. In the standard of living domain we have data for household spending, personal income, food scarcity, housing insecurity, and saving rate. In the health domain there are stats for new and old cases and deaths. In the government domain we have stats for federal govt spending, federal reserve assets, state tax revenue, and federal deficit. For each, the claims, per 100k stat, and % change are detailed. Related articles have national related articles. Otherwise on the site you have links for social issues and government processing related to COVID.

The COVID number of cases dataset csv

Identical to number of deaths dataset.

The COVID number of deaths dataset csv

The columns are as follows: CountyFIP, County Name, State, StateFIP, then date. Date starts at 1/2/20. When there is a 0 in CountyFIP, the row is statewide. The data values increase as time goes on. There are 3138 counties accounted for. There are 56 StateFIPS. Federal Information Processing Standard is FIPS. It is a code assigned to county and state.

The COVID death dataset csv

The columns are as follows: countyFIP, County Name, State, Population. A "0" for countyFIP indicates statewide statistics.

VARIABLES DICTIONARY
Column Name
Description
Type

State
A state of USA, or abbreviation
Plain Text

7-Day Average Cases
Average of all cases for the last week per state
Number

7-Day Average Deaths
Average of all covid related deaths for the last week per state
Number

Cases
Sum of cases per county
Number

Deaths
Sum of covid related deaths per county
Number

Population
Sum of inhabitants per county
Number

#########(Date)
Numerical identifier of a specific day
Date & Time

CountyName
Name of a county, a political and administrative division of a state, providing certain local governmental services.
Plain Text

countyFIPS
Federal InformationProcessing System(FIPS) codes are numbers **which uniquely identify geographic areas**. The number of digits in FIPS codes vary depending on the level of geography. State-level FIPS codes have two digits, county-level FIPS codes have

five digits of which the first two are the FIPS code of the state to which the county belongs.
Number

StateFIPS

Federal InformationProcessing System(FIPS) codes are numbers **which uniquely identify geographic areas**. State-level FIPS codes have two digits.
Number

VARIABLES

The variables in the datasets are the following: time, county, state, cases, deaths, population. (Is it just cases, deaths, and population? Can I consolidate state and county into "location"?)

PRELIMINARY INTUITIONS:

The data comes in modification of these variables.

Cases and deaths have positive correlation. Deaths are always less than cases.

Cases and deaths have positive correlation with population.

Cases and deaths have positive correlation with time.

The dataset does not account for population changes across time.

High county population is caused by big populous cities within in it.

Hospital Utilization Dataset (Aka'sha)

HHS Protect Public Data hub is the source. Hospital utilization is reported to the HHS Protect by all hospitals registered with Centers for Medicaid and Medicare Services(CMS) as of June 1, 2020. It contains data for utilization and capacity status of hospitals in the US. COVID ICU bed utilization values include both confirmed and suspected cases of COVID.
The visualizations on this page display state-aggregated data for hospital utilization. These data are derived from reports with facility-level granularity across two main sources: HHS TeleTracking, and reporting provided directly to HHS Protect by state/territorial health departments on behalf of their healthcare facilities.
The data uses the latest values reported by each facility within the last four days. This allows for a more comprehensive picture of the hospital utilization within a state by ensuring a hospital is represented, even if they miss a single day of reporting. No statistical analysis is applied to account

for non-response and/or to account for missing data. HHS publicly tracks the frequency hospitals report this data on the Hospital Reporting page.

This data contains information on the following: hospital utilization, hospital reporting, therapeutics, and national testing. Under hospitalization we can find data on hospital utilization by facility, inpatient bed utilization by state, and ICU bed utilization by state. Under hospital reporting we can find hospital reporting frequency, percentage of hospitals reporting by state, and other important info on this data at https://protect-public.hhs.gov/pages/hospital-reporting. Info on therapeutics distribution data at https://protect-public.hhs.gov/pages/therapeutics-distribution. Under national testing we have COVID-19 Diagnostic Laboratory Testing, national testing trends, and more info on this data at https://protect-public.hhs.gov/pages/national-testing.

The hospital dataset is found here: Reported Patient Impact and Hospital Capacity.

MAPPABLE VARIABLES: date(collection_week), state, city<county>,

   The following variables can map to total CASES:

   Total_adult_patients_hospitalized_confirmed_and_suspected_covid_7_day_sumtotal_pe diatric_patients_hospitalized_confirmed_and_suspected_covid_7_day_sum
      Staffed_icu_adult_patients_confirmed_and_suspected_covid_7_day_sum
Could we map deaths to new openings? Is there anything that maps to population?

A simple left join would suffice for merging this with the main COVID-19 dataset.

This enrichment data can help in the analysis of the COVID-19 spread by detailing the intensity of the pandemic and its impact on public health, by supporting accurate case and confirmation numbers, by detailing spread that happens at the hospitals. (Aka'sh)

(Aka'sh) Hospital Utilization Dataset Documentation (ctrl+f DONE>NEXT to move to the next section.)
https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw 7u/data

Column Name
Description
Type

hospital_pk
This unique key for the given hospital that will match the ccn column if it exists, otherwise, it is a derived unique key.
Plain Text

collection_week
This date indicates the start of the period of reporting (the starting Friday).

Date & Time

state
[FAQ - 1. d)] The two digit state/territory code for the hospital.
Plain Text

ccn
[FAQ - 1. b)] CMS Certification Number (CCN) of the given facility
Plain Text

hospital_name
[FAQ - 1. a)] The name of the facility reporting.
Plain Text

address
The address of the facility reporting.
Plain Text

city
The city of the facility reporting.
Plain Text

zip
The 5-digit zip code of the facility reporting.
Plain Text

hospital_subtype
The sub-type of the facility reporting. Valid values are: Children's Hospitals, Critical Access Hospitals, Long Term, Psychiatric, Rehabilitation & Short Term. Some facilities are not designated with this field.
Plain Text

fips_code
The Federal Information Processing Standard (FIPS) code of the location of the hospital.
Plain Text

is_metro_micro
This is based on whether the facility serves a Metropolitan or Micropolitan area. True if yes, and false if no.

Plain Text

total_beds_7_day_avg
[FAQ - 2. a)] Average of total number of all staffed inpatient and outpatient beds in the hospital, including all overflow, observation, and active surge/expansion beds used for inpatients and for outpatients (including all ICU, ED, and observation) reported during the 7-day period.
Number

all_adult_hospital_beds_7_day_avg
[FAQ - 2. b)] Average of all staffed inpatient and outpatient adult beds in the hospital, including all overflow and active surge/expansion beds for inpatients and for outpatients (including all ICU, ED, and observation) reported during the 7-day period.
Number

all_adult_hospital_inpatient_beds_7_day_avg
[FAQ - 3. b)] Average of total number of staffed inpatient adult beds in the hospital including all overflow and active surge/expansion beds used for inpatients (including all designated ICU beds) reported during the 7-day period.
Number

inpatient_beds_used_7_day_avg
[FAQ - 4. a)] Average of total number of staffed inpatient beds that are occupied reported during the 7-day period.
Number

all_adult_hospital_inpatient_bed_occupied_7_day_avg
[FAQ - 4. b)] Average of total number of staffed inpatient adult beds that are occupied reported during the 7-day period.
Number

total_adult_patients_hospitalized_confirmed_and_suspected_covid_7_day_avg
[FAQ - 9. a)] Average number of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed or suspected COVID19, including those in observation beds reported during the 7-day period.
Number

total_adult_patients_hospitalized_confirmed_covid_7_day_avg

[FAQ - 9. b)] Average number of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19, including those in observation beds. This average includes patients who have both laboratory-confirmed COVID-19 and laboratory-confirmed influenza.
Number

total_pediatric_patients_hospitalized_confirmed_and_suspected_covid_7_day_avg
[FAQ - 10. a)] Average number of patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who are suspected or laboratory-confirmed-positive for COVID-19. This average includes those in observation beds reported in the 7-day period.
Number

total_pediatric_patients_hospitalized_confirmed_covid_7_day_avg
[FAQ - 10. b)] Average number of patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who have laboratory-confirmed COVID-19. Including those in observation beds. Including patients who have both laboratory-confirmed COVID-19 and laboratory confirmed influenza in this field reported in the 7-day period.
Number

inpatient_beds_7_day_avg
[FAQ - 3. a)] Average number of total number of staffed inpatient beds in your hospital including all overflow, observation, and active surge/expansion beds used for inpatients (including all ICU beds) reported in the 7-day period.
Number

total_icu_beds_7_day_avg
[FAQ - 5. a)] Average number of total number of staffed inpatient ICU beds reported in the 7-day period.
Number

total_staffed_adult_icu_beds_7_day_avg
[FAQ - 5. b)] Average number of total number of staffed inpatient adult ICU beds reported in the 7-day period.
Number

icu_beds_used_7_day_avg

[FAQ - 6. a)] Average number of total number of staffed inpatient ICU beds that are occupied reported in the 7-day period.
Number

staffed_adult_icu_bed_occupancy_7_day_avg
[FAQ - 6. b)] Average of total number of staffed inpatient adult ICU beds that are occupied reported in the 7-day period.
Number

staffed_icu_adult_patients_confirmed_and_suspected_covid_7_day_avg
[FAQ - 12. a)] Average number of patients currently hospitalized in a designated adult ICU bed who have suspected or laboratory-confirmed COVID-19 reported in the 7-day period.
Number

staffed_icu_adult_patients_confirmed_covid_7_day_avg
[FAQ - 12. b)] Average number of patients currently hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19. Including patients who have both laboratory-confirmed COVID-19 and laboratory-confirmed influenza in this field reported in the 7-day period.
Number

total_patients_hospitalized_confirmed_influenza_7_day_avg
[FAQ - 33] Average number of patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed influenza. Including those in observation beds reported in the 7-day period.
Number

icu_patients_confirmed_influenza_7_day_avg
[FAQ - 35] Average of patients (all ages) currently hospitalized in a designated ICU bed with laboratory-confirmed influenza in the 7-day period.
Number

total_patients_hospitalized_confirmed_influenza_and_covid_7_day_avg
[FAQ - 36] 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.
Number

total_beds_7_day_sum
[FAQ - 2. a)] Sum of reports of total number of all staffed inpatient and outpatient beds in the hospital, including all overflow, observation, and active surge/expansion beds used for inpatients and for outpatients (including all ICU, ED, and observation) reported during the 7-day period.
Number

all_adult_hospital_beds_7_day_sum
[FAQ - 2. b)] Sum of reports of all staffed inpatient and outpatient adult beds in the hospital, including all overflow and active surge/expansion beds for inpatients and for outpatients (including all ICU, ED, and observation) reported during the 7-day period.
Number

all_adult_hospital_inpatient_beds_7_day_sum
[FAQ - 3. b)] Sum of reports of all staffed inpatient and outpatient adult beds in the hospital, including all overflow and active surge/expansion beds for inpatients and for outpatients (including all ICU, ED, and observation) reported during the 7-day period.
Number

inpatient_beds_used_7_day_sum
[FAQ - 4. a)] Sum of reports of total number of staffed inpatient beds that are occupied reported during the 7-day period.
Number

all_adult_hospital_inpatient_bed_occupied_7_day_sum
[FAQ - 4. b)] Sum of reports of total number of staffed inpatient adult beds that are occupied reported during the 7-day period.
Number

total_adult_patients_hospitalized_confirmed_and_suspected_covid_7_day_sum
[FAQ - 9. a)] Sum of reports of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed or suspected COVID19. Including those in observation beds reported during the 7-day period.
Number

total_adult_patients_hospitalized_confirmed_covid_7_day_sum
[FAQ - 9. b)] Sum of reports of patients currently hospitalized in an adult inpatient bed who have laboratory-confirmed COVID-19. Including those in observation beds. Including patients who

have both laboratory-confirmed COVID-19 and laboratory confirmed influenza in this field during the 7-day period.
Number

total_pediatric_patients_hospitalized_confirmed_and_suspected_covid_7_day_sum
[FAQ - 10. a)] Sum of reports of patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who are suspected or laboratory-confirmed-positive for COVID-19. Including those in observation beds reported in the 7-day period.
Number

total_pediatric_patients_hospitalized_confirmed_covid_7_day_sum
[FAQ - 10. b)] Sum of reports of patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who have laboratory-confirmed COVID-19. Including those in observation beds. Including patients who have both laboratory-confirmed COVID-19 and laboratory confirmed influenza in this field reported in the 7-day period.
Number

inpatient_beds_7_day_sum
[FAQ - 3. a)] Sum of reports of total number of staffed inpatient beds in your hospital including all overflow, observation, and active surge/expansion beds used for inpatients (including all ICU beds) reported in the 7-day period.
Number

total_icu_beds_7_day_sum
[FAQ - 5. a)] Sum of reports of total number of staffed inpatient ICU beds reported in the 7-day period.
Number

total_staffed_adult_icu_beds_7_day_sum
[FAQ - 5. b)] Sum of reports of total number of staffed inpatient adult ICU beds reported in the 7-day period.
Number

icu_beds_used_7_day_sum
[FAQ - 6. a)] Sum of reports of total number of staffed inpatient ICU beds reported in the 7-day period.
Number

staffed_adult_icu_bed_occupancy_7_day_sum
[FAQ - 6. b)] Sum of reports of total number of staffed inpatient adult ICU beds that are occupied reported in the 7-day period.
Number

staffed_icu_adult_patients_confirmed_and_suspected_covid_7_day_sum
[FAQ - 12. a)] Sum of reports of patients currently hospitalized in a designated adult ICU bed who have suspected or laboratory-confirmed COVID-19 reported in the 7-day period.
Number

staffed_icu_adult_patients_confirmed_covid_7_day_sum
[FAQ - 12. b)] Sum of reports of patients currently hospitalized in a designated adult ICU bed who have laboratory-confirmed COVID-19. Including patients who have both laboratory-confirmed COVID-19 and laboratory-confirmed influenza in this field reported in the 7-day period.
Number

total_patients_hospitalized_confirmed_influenza_7_day_sum
[FAQ - 33] Sum of reports of patients (all ages) currently hospitalized in an inpatient bed who have laboratory-confirmed influenza. Including those in observation beds reported in the 7-day period.
Number

icu_patients_confirmed_influenza_7_day_sum
[FAQ - 35] Sum of reports of patients (all ages) currently hospitalized in a designated ICU bed with laboratory-confirmed influenza reported in the 7-day period.
Number

total_patients_hospitalized_confirmed_influenza_and_covid_7_day_sum
[FAQ - 36] Sum of reports 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.
Number

total_beds_7_day_coverage

[FAQ - 2. a)] Number of times in the 7 day period that the facility reported total number of all staffed inpatient and outpatient beds in your hospital, including all overflow, observation, and active surge/expansion beds used for inpatients and for outpatients (including all ICU, ED, and observation).
Number

all_adult_hospital_beds_7_day_coverage
[FAQ - 2. b)] Number of times in the 7-day period that the facility reported total number of all staffed inpatient and outpatient adult beds in your hospital, including all overflow and active surge/expansion beds for inpatients and for outpatients (including all ICU, ED, and observation).
Number

all_adult_hospital_inpatient_beds_7_day_coverage
[FAQ - 3. b)] Number of times in the 7-day period that the facility reported total number of staffed inpatient adult beds in your hospital including all overflow and active surge/expansion beds used for inpatients (including all designated ICU beds).
Number

inpatient_beds_used_7_day_coverage
[FAQ - 4. a)] Number of times in the 7-day period that the facility reported total number of staffed inpatient beds that are occupied.
Number

all_adult_hospital_inpatient_bed_occupied_7_day_coverage
[FAQ - 4. b)] Number of times in the 7-day period that the facility reported total number of staffed inpatient adult beds that are occupied.
Number

total_adult_patients_hospitalized_confirmed_and_suspected_covid_7_day_coverage
[FAQ - 9. a)] 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. Including those in observation beds.
Number

total_adult_patients_hospitalized_confirmed_covid_7_day_coverage
[FAQ - 9. b)] 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. Including those

in observation beds. Including patients who have both laboratory-confirmed COVID-19 and laboratory confirmed influenza in this field.
Number

total_pediatric_patients_hospitalized_confirmed_and_suspected_covid_7_day_coverage
[FAQ - 10. a)] Number of times in the 7-day period that the facility reported Patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who are suspected or laboratory-confirmed-positive for COVID-19. Including those in observation beds.
Number

total_pediatric_patients_hospitalized_confirmed_covid_7_day_coverage
[FAQ - 10. b)] Number of times in the 7-day period that the facility reported patients currently hospitalized in a pediatric inpatient bed, including NICU, PICU, newborn, and nursery, who have laboratory-confirmed COVID-19. Including those in observation beds. Including patients who have both laboratory-confirmed COVID-19 and laboratory confirmed influenza in this field.
Number

inpatient_beds_7_day_coverage
[FAQ - 3. a)] Number of times in the 7-day period that the facility reported total number of staffed inpatient beds in your hospital including all overflow, observation, and active surge/expansion beds used for inpatients (including all ICU beds).
Number

total_icu_beds_7_day_coverage
[FAQ - 5. a)] Number of times in the 7-day period that the facility reported total number of staffed inpatient ICU beds.
Number

total_staffed_adult_icu_beds_7_day_coverage
[FAQ - 5. b)] Number of times in the 7-day period that the facility reported total number of staffed inpatient adult ICU beds.
Number

icu_beds_used_7_day_coverage
[FAQ - 6. a)] Number of times in the 7-day period that the facility reported total number of staffed inpatient ICU beds.
Number

staffed_adult_icu_bed_occupancy_7_day_coverage
[FAQ - 6. b)] Number of times in the 7-day period that the facility reported total number of staffed inpatient adult ICU beds that are occupied.
Number

staffed_icu_adult_patients_confirmed_and_suspected_covid_7_day_coverage
[FAQ - 12. a)] Number of times in the 7-day period that the facility reported patients currently hospitalized in a designated adult ICU bed who have suspected or laboratory-confirmed COVID-19.
Number

staffed_icu_adult_patients_confirmed_covid_7_day_coverage
[FAQ - 12. b)] 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. Including patients who have both laboratory-confirmed COVID-19 and laboratory-confirmed influenza in this field.
Number

total_patients_hospitalized_confirmed_influenza_7_day_coverage
[FAQ - 33] 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. Including those in observation beds.
Number

icu_patients_confirmed_influenza_7_day_coverage
[FAQ - 35] 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.
Number

total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage
[FAQ - 36] 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.
Number

previous_day_admission_adult_covid_confirmed_7_day_sum

[FAQ - 17. a)] Sum of number of patients who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19 at the time of admission reported in the 7-day period.
Number

previous_day_admission_adult_covid_confirmed_18-19_7_day_sum
[FAQ - 17. a1)] 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 at the time of admission reported in the 7-day period.
Number

previous_day_admission_adult_covid_confirmed_20-29_7_day_sum
[FAQ - 17. a2)] 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 at the time of admission reported in the 7-day period.
Number

previous_day_admission_adult_covid_confirmed_30-39_7_day_sum
[FAQ - 17. a3)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_40-49_7_day_sum
[FAQ - 17. a4)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_50-59_7_day_sum
[FAQ - 17. a5)] Sum of number of patients age 50-59 who were admitted to an adult inpatient bed on the previous calendar day who had confirmed COVID-19 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_60-69_7_day_sum

[FAQ - 17. a6)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_70-79_7_day_sum
[FAQ - 17. a7)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_80+_7_day_sum
[FAQ - 17. a8)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_confirmed_unknown_7_day_sum
[FAQ - 17. a9)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_pediatric_covid_confirmed_7_day_sum
[FAQ - 18. a)] Sum of number of pediatric patients who were admitted to an inpatient bed, including NICU, PICU, newborn, and nursery, on the previous calendar day who had confirmed COVID-19 at the time of admission.
Number

previous_day_covid_ED_visits_7_day_sum
[FAQ - 20] Sum of total number of ED visits who were seen on the previous calendar day who had a visit related to COVID-19 (meets suspected or confirmed definition or presents for COVID diagnostic testing – do not count patients who present for pre-procedure screening) reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_7_day_sum

[FAQ - 17. b)] Sum of number of patients who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_18-19_7_day_sum
[FAQ - 17. b1)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_20-29_7_day_sum
[FAQ - 17. b2)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_30-39_7_day_sum
[FAQ - 17. b3)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_40-49_7_day_sum
[FAQ - 17. b4)] Sum of number of patients age 40-49 who were admitted to an adult inpatient bed on the previous calendar day who had suspected COVID-19 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_50-59_7_day_sum
[FAQ - 17. b5)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_60-69_7_day_sum

[FAQ - 17. b6)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_70-79_7_day_sum
[FAQ - 17. b7)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_80+_7_day_sum
[FAQ - 17. b8)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_adult_covid_suspected_unknown_7_day_sum
[FAQ - 17. b9)] 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 at the time of admission reported in 7-day period.
Number

previous_day_admission_pediatric_covid_suspected_7_day_sum
[FAQ - 18. b)] Sum of number of pediatrics patients who were admitted to an inpatient bed, including NICU, PICU, newborn, and nursery, on the previous calendar day who had suspected COVID-19 at the time of admission reported in 7-day period.
Number

previous_day_total_ED_visits_7_day_sum
[FAQ - 19] Sum of total number of patient visits to the ED who were seen on the previous calendar day regardless of reason for visit. Including all patients who are triaged even if they leave before being seen by a provider reported in the 7-day period.
Number

previous_day_admission_influenza_confirmed_7_day_sum

[FAQ - 34] Sum of number of patients (all ages) who were admitted to an inpatient bed on the previous calendar day who had laboratory-confirmed influenza at the time of admission reported in 7-day period.
Number

geocoded_hospital_address
The geocoded latitude and longitude of the facility reporting.
Point

hhs_ids
As a part of an effort to improve the granularity of spatial data, unique identifiers (named "HHS IDs" in the datasets) have been assigned to each individual facility. These unique identifiers are provided so data users can reference each individual "brick and mortar" facility that is reporting data to HHS, even in cases when multiple facilities report under the same CMS Certification Number (CCN). Additional datasets and further details related to HHS IDs will be released at a later date.
Plain Text

previous_day_admission_adult_covid_confirmed_7_day_coverage
[FAQ - 17a] Number of times in the 7 day period that the facility reported the number of patients admitted to adult inpatient bed on previous calendar day with confirmed COVID-19
Number

previous_day_admission_pediatric_covid_confirmed_7_day_coverage
[FAQ - 18a] Number of times in the 7 day period that the facility reported the number of pediatric patients admitted to an inpatient bed on previous calendar day with confirmed COVID-19
Number

previous_day_admission_adult_covid_suspected_7_day_coverage
[FAQ - 17b] Number of times in the 7 day period that the facility reported the number of patients admitted to adult inpatient bed on previous calendar day with suspected COVID-19
Number

previous_day_admission_pediatric_covid_suspected_7_day_coverage
[FAQ - 18b] Number of times in the 7 day period that the facility reported the number of pediatric patients admitted to inpatient bed on previous calendar day with suspected COVID-19
Number

previous_week_personnel_covid_vaccinated_doses_administered_7_day
[FAQ - 41] The number reported of COVID-19 vaccination doses administered in the previous week regardless of whether it is a first or second dose in a series. (Collected on Wednesdays only)
Number

total_personnel_covid_vaccinated_doses_none_7_day
[FAQ - 42] The number reported of healthcare personnel who have not yet received a single vaccine dose. (Collected on Wednesdays only)
Number

total_personnel_covid_vaccinated_doses_one_7_day
[FAQ - 43] The current number reported of healthcare personnel who have received at least one dose of COVID-19 vaccination that is administered in a multi-dose series. This field is meant to represent those who have begun but not completed the vaccination process. Does not include those who received a single-dose vaccine in this field. (Collected on Wednesdays only)
Number

total_personnel_covid_vaccinated_doses_all_7_day
[FAQ - 44] The current number reported of healthcare personnel who have received a complete series of a COVID-19vaccination.Includes those who have received all doses in a multi-dose series as well as those who received a single-dose vaccine. (Collected on Wednesdays only)
Number

previous_week_patients_covid_vaccinated_doses_one_7_day
[FAQ - 46] The number reported of patients who received the first dose of a COVID-19 vaccine that is administered in a multi-dose series. (Collected on Wednesdays only)
Number

previous_week_patients_covid_vaccinated_doses_all_7_day
[FAQ - 47] The number reported of patients who received the final dose in a COVID-19 vaccination series. (Collected on Wednesdays only)
Number

is_corrected
If a fields has been altered from it's submitted form to smooth out data anomalies, this flag will be set to true
Checkbox

<each of our enrichment dataset info goes here>

Political Leanings dataset (Joseph)

Description: The U.S. presidential election results dataset contains information about voting data at a county-by county level. This includes information about political leanings of voters, total votes cast, and candidates' political information. This has important implications for election day covid data as well as provides a good idea about several confounding variables from the primary data set.

The dataset features multiple CSV files that break down the governor, house, senate, and presidential races. Among those each category has 2 CSV files, 1 identifying the votes and 1 identifying the candidate. The data set finally features additional CSVs containing data about statewide results for each race. For each county race data set the voting information is broken down into 5 columns: state, county, current_votes, total_votes, and percent. The ones containing candidate information include state, county, candidate, party, and votes. Votes and percent are expressed as numerical values and state, county, party, and candidate are text while votes is an outlying Boolean value. The dataset should be noted for containing thorough, organized data ideal for the project, however unnecessary variables and redundancies not needed may be removed.

Variable dictionary:

Variable Name: state

Variable Type: Text

Description: Variable detailing what state a vote was cast in.

Variable Name: county

Variable Type: Text

Description: Describes what county a vote was cast in.

Variable Name: candidate

Variable Type: Text

Description: The name of the politician running for a given political race.

Variable Name: party

Variable Type: Text

Description: Identifies the officially declared political party of the candidate.

Variable Name: current_votes

Variable Type: Number

Description: Keeps track of report votes cast per county. This number is separate and less than the Total Votes value.

Variable Name: total_votes

Variable Type: Number

Description: All votes total identified per county.

Variable Name: votes

Variable Type: Number

Description: Number of votes reported for a candidate on a statewide level.

Variable Name: percent

Variable Type: Number

Description: Percentage of votes that were cast in each county based on total and reported voting.

Variable Name: won

Variable Type: Boolean

Description: Boolean valuing indicating whether a candidate won their given race or not.

How to map the datasets:

My simplest intuition is the point to the county based data and say that if we can align the values in these datasets with the counties from the main dataset we have a good starting point to bring the data into one location.

How can this enrichment data help in the analysis of the spread of covid?:

There's some amount of important information in a lot of different aspects of this data. For instance, we can analyze whether having more condensed voting places (or voting restrictions in general) increased the spread of covid since more people would have to wait in line with others and risk getting sick. The best example of how this can help is in seeing if a counties voting persuasions resulted in an increase or decrease in cases and we can discuss if policy concerning covid increased or decreased the spread. The most logical question to ask is: Does a blue county (or a red or purple county) tend to have a lower or higher spread and can we identify any bumps or spikes in cases close to policy that might indicate the policy of a given state or county had an impact on the cases.

(Joseph Brannon end)

Employment Dataset Description (Chandra Shekhar):
The employment dataset is a Quarterly Census of Employment and Wages (QCEW) data classified using North American Industry Classification System (NAICS). This dataset provides the employment details of whole US on a quarterly and annual basis. It is classified based on year starting from 1975 to 2021. The data is available in a single downloadable csv file for every year. In this project we are considering a single csv file dataset for the year 2021. The dataset contains the columns as follow:

1. Area FIPS(area_fips): It is a 5 character code with unique values to each county. All the counties are labelled with a 5 digit unique code called FIPS code. The annual data of the USA is defined with an id of characters and digits.

   Data Type: Text

2. Ownership Code(own_code): It is a single digit number that represents the ownership group id.

   Data Type:Text

3. Industry Code (industry_code): It is a 6 digit number that represents the superset industry code of the data.

   Data Type: Text

4. Aggregation Level Code(agglvl_code): It is a 2 character value that indicates the aggregation level of the data summarised on the record.

   Data Type: Text

5. Size Code(size_code): It is a single digit value that indicates the size of an organisation based on the employee strength.

   Data Type: Text

6. Year: This is a 4 digit number which represents the data collected for the year.

   Data Type: Text

7. Quarter (Qtr): This column represents the data collected for the quarter of the year. For annual data the quarter value is represented with 'A'.

   Data Type: Text

8. Disclosure Code(disclosure_code): This is a single character data that represents if the data is disclosed or not. Possible values are blank- for disclosed and 'N'- for not disclosed.

   Data Type: Text

9. Area Title(area_title):It is the name of an area associated with the "area code" that identifies the geographic area that the data represents.

   Data Type: Text

10. Owner Title(own_title): It is an ownership title associate with the own code

Data Type: Text

11. Industry Title(industry_title): It is an Industry title associated with the industry code.

    Data Type: Text

12. Aggregation level Title(agglvl_title): It is the title of the aggregation level of the data associated with the aggregation level code.

    Data Type: Text

13. Size Title(size_title): It is the title of industry size which is associated with the size code.

    Data Type: Text

14. Annual Average Establishment Count: It is an integer value which states the annual establishment counts for a given year.

    Data Type: Numeric

15. Annual Average Employment Level: It is an integer value that gives the annual average employment level for a given year.

    Data Type: Numeric

16. Annual Total Wages(total_annual_wages): It is an integer value that denotes the total annual wage levels for a given year.

    Data Type: Numeric

17. Taxable Annual Wages: It is the sum of 4 quarterly total taxable wages in a given year.

    Data Type: Numeric

18. Annual Contribution: t is the sum of 4 quarterly contributions given in a year.

    Data Type: Numeric

19. Annual Average Weekly Wage: It is an integer value that gives the Average weekly wage based on the 12-monthly employment levels.

    Data Type: Numeric

20. Annual Average Pay: It is an integer value that denotes the average annual wage based on the employment level.

    Data Type: Numeric

21. Location Quotient Disclosure(lq_disclosure_code): It is a single character value that defines the location quotient disclosure code.

    Data Type: Text

22. Location Quotient Annual Establishment (lq_annual_avg_estabs): It is a location quotient of average annual establishment count relative to US.

    Data Type: Numeric

23. Location Annual Employment Level Quotient(lq_annual_avg_emplvl): This is a floating value that denotes the average annual employment quotient of the area compared to the total US employment.

    Data Type: Numeric

24. Annual Wage level Location Quotient(lq_total_annual_wages): This is a floating value that denotes the average annual wage quotient of the area compared to the total US employment.

    Data Type: Numeric

25. Annual Taxable Wage level Location Quotient(lq_taxable_annual_wages): This is a floating value that denotes the average annual taxable wage quotient of the area compared to the total US employment.

    Data Type: Numeric

26. Annual Contribution Quotient (lq_annual_contributions): This is a floating value that denotes the average annual contribution of the area compared to the total US employment.

    Data Type: Numeric

27. Annual Weekly Wage Location Quotient(lq_annual_avg_wkly_wage): This is a floating value that denotes the average weekly wage quotient of the area compared to the total US employment.

    Data Type: Numeric

28. Average Annual Pay Location Quotient(lq_avg_annual_pay): This is a floating value that denotes the average annual pay quotient of the area compared to the total US employment.

    Data Type: Numeric

29. Over the year Disclosure Code(oty_disclosure_code): It is a single character data that represents the disclosure code over the year. The values are either blank or 'N'.

    Data Type: Text

30. Annual establishment change(oty_annual_avg_estabs_chg): It is numeric data that represents the value for establishment change over the year for a given year.

    Data Type: Numeric

31. Annual establishment percentage Change(oty_annual_avg_estabs_pct_chg): It is numeric data that represents the value in percentage for establishment change over the year for a given year.

    Data Type: Numeric

32. Annual Employment level change(oty_annual_avg_emplvl_chg): It is numeric data that represents the value of change in the employment level over the year for a given year.

    Data Type: Numeric

33. Annual Employment level percentage change(oty_annual_avg_emplvl_pct_chg): It is numeric data that represents the value of percentage change in the employment level over the year for a given year.

    Data Type: Numeric

34. Annual Wage level change(oty_total_annual_wages_chg): It is a numeric data that represents the value of change in the wage level over the year for a given year.

    Data Type: Numeric

35. Annual Wage level percentage Change(oty_total_annual_wages_pct_chg): It is a numeric data that represents the value of percentage change in the wage level over the year for a given year.

    Data Type: Numeric

36. Annual Taxable Wage level change(oty_taxable_annual_wages_chg): It is numeric data that represents the value of change in the taxable wage level over the year for a given year.

    Data Type: Numeric

37. Annual Taxable Wage level percentage change(oty_taxable_annual_wages_chg): It is numeric data that represents the value of percentage change in the taxable wage level over the year for a given year.

Data Type: Numeric

38. Annual Contribution change(oty_annual_contributions_chg): It is numeric data that represents the value of change in the contribution level over the year for a given year.

    Data Type: Numeric

39. Annual Contribution percentage change(oty_annual_contributions_chg): It is numeric data that represents the value of change in the percentage of contribution level over the year for a given year.

    Data Type: Numeric

40. Annual Average weekly Wage level change(oty_annual_avg_wkly_wage_chg): It is numeric data that represents the value of change in the average weekly wage level over the year for a given year.

    Data Type: Numeric

41. Annual Average Weekly Wage level percentage Change (oty_annual_avg_wkly_wage_pct_chg): It is numeric data that represents the value of change in the percentage of average weekly wage level over the year for a given year.

    Data Type: Numeric

42. Annual Average Pay level change(oty_avg_annual_pay_chg): It is numeric data that represents the value of change in the average pay level over the year for a given year.

    Data Type: Numeric

43. Annual Average Pay level percentage Change(oty_avg_annual_pay_pct_chg): It is a numeric data that represents the value of change in the percentage of average pay level over the year for a given year.

    Data Type: Numeric

Merging Employment Dataset with primary COVID-19 Dataset:
The primary COVID-19 Dataset is formed by merging the Number of Cases, Number of Deaths and Population by County datasets. These datasets are merged based on the 'CountyFIPS code' and 'State Columns'.
The Employment dataset is a tricky dataset. The county codes for this dataset are defined as 'area_fips'. The data type of this column is 'Objective'. It contains the values of both digits and strings. Since the employment dataset represents the data for all the counties and metropolitan areas, it does not match with the 'CountyFIPS' from the primary COVID-19 dataset. Hence, all the rows where the data of 'area_fips' is not matching with 'CountyFIPS' should be removed. Even after removing the rows that represent metropolitan areas and unknown areas the merge

will fail, as every county in the Employment has multiple occurrences based on the 'own_code' and 'industry_code'. The combinations of 'area_fips', 'own_code' and 'industry_code' are considered as unique records in the employment dataset. So, each value in 'area_fips' will be repeated for the combination of 'own_code' and 'industry_code' values in that county. To overcome this, a combination of 'own_code' and 'industry_code' should be decided and respective data for each county from the above filtered table should be collected. The primary dataset and Employment dataset can be merged based on the 'area_fips' of the new filtered dataset and 'CountyFIPS' from the primary dataset.

Analysis of COVID-19 spread using Enrichment Dataset:
COVID-19 impacted many employees and businesses. The employment dataset provides changes in employment and wages value. These values can be used to find if the changes in employment, pay, wages or taxable pay had any impact on the growth or decline in the number of cases in a county. One possible analysis is to check how trade, transportation and utilities industry reflected the changes on covid cases in a county. In this section industry code 1021(Trade, transportation, and utilities) and ownership code 5(Private) will be considered.

Initial Hypothesis Questions:

1. Does change in employment in private organisations affect the number of cases?
2. Does a county with negative value in employment level change quotient have more cases?
3. How are the Over-the-year change in values related to covid cases?
4. Does an increase in the number of employees in a quarter result in an increase of covid cases?
5. Does the change in weekly-wages of state government owned data have an impact on covid cases?   **(Chandra Shekhar end)**

ACS Demographics and Housing Estimates (Chandan):

The American Community Survey is the latest survey conducted by the US Census Bureau wherein it uses monthly samples to produce annual estimates demographics for small areas. The Census Bureau also produces 5 year and 1-year estimates. The population data set includes people living in both housing units and group quarters. ACS is conducted both in America and Puerto Rico (where it's called Puerto Rico Community Survey (PRCS)).

The present data set is a 5 - year data estimate of 2019 county-wise in United States and Puerto Rico.

The data set has various categories marked by different measures (Estimate, Margin of Error, Percent, Percent Margin of Error).

1. Estimate (E) – Sum total of observation. Data type – Integer.

2.   Margin of Error (MOE) – numeric difference between estimate and upper and lower confidence bounds of a population. Data Type: Float

3.   Percent (P) – (available estimate in a category/ total population estimate) * 100. Data Type: Float

4.   Percent Margin of Error (PMOE) – level of confidence intervals. Data Type: Float

The data set has categories like below:

1.   Sex and Age

2.   Race

3.   Race alone or in combination with one or more races.

4.   Hispanic/Latino and Race

5.   Total Housing Units.

6.   Citizen, Voting age population.

Now these categories have subcategories within them as shown below:

1.   Sex and Age

    1.1   Total Population. Data type : Integer.

        1.1.1   Male. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

        1.1.2   Female. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

        1.1.3   Sex Ratio (Males/100 Females) Data Type for {Estimate, Margin of Error, Percent, Percent Margin of error} – Float

        1.1.4   Age Frames – Ranging between 0 to 65 and above with interval size of 10. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

        1.1.5   Median Age. Data Type for {Estimate, Margin of Error, Percent, Percent Margin of error} – Float

1.1.6   Above/Under certain age frames (<18,>16,>18,>21,>62,>65). . Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

1.1.7   Above 18 years

   1.1.7.1   Male. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

   1.1.7.2   Female. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

   1.1.7.3   Sex Ratio(Males/100 Females). Data Type for {Estimate, Margin of Error, Percent, Percent Margin of error} – Float

1.1.8   Above 65 years

   1.1.8.1   Male. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

   1.1.8.2   Female. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

   1.1.8.3   Sex Ratio(Males/100 females). Data Type for {Estimate, Margin of Error, Percent, Percent Margin of error} – Float

2.   Race

   2.1 Total Population. Data Type : Integer.

      2.1.1   One race. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

         2.1.1.1 White

         2.1.1.2 Black or African American.

         2.1.1.3 American Indian and Alaska Native.

               2.1.1.3.1   Cherokee Tribal Grouping.

               2.1.1.3.2   Chippewa Tribal Grouping.

               2.1.1.3.3   Navajo Tribal Grouping.

               2.1.1.3.4   Sioux Tribal Grouping.

         2.1.1.4 Asian.

2.1.1.4.1   Asian Indian.

2.1.1.4.2   Chinese.

2.1.1.4.3   Filipino.

2.1.1.4.4   Japanese.

2.1.1.4.5   Korean.

2.1.1.4.6   Vietnamese.

2.1.1.4.7   Other Asian.

2.1.1.5   Native Hawaiian and other Pacific Islander.

2.1.1.5.1   Native Hawaiian

2.1.1.5.2   Guamanian or Chamorro

2.1.1.5.3   Samoan

2.1.1.5.4   Other Pacific Islander

2.1.1.6   Some other race. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

2.1.1.7   Two or more races. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

2.1.1.7.1   White and Black or African American.

2.1.1.7.2   White and American Indian and Alaska Native.

2.1.1.7.3   White and Asian.

2.1.1.7.4   Black or African American and American Indian and Alaska Native.

3.   Race alone or in combination with one or more races.

3.1   Total Population. Data Type : Integer.

3.1.1   White. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float,   Percent Margin of error – Float)

3.1.2   Black or African American. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

3.1.3   American Indian and Alaska Native. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

3.1.4   Asian. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float,   Percent Margin of error – Float)

3.1.5   Native Hawaiian and other Pacific Islander. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

3.1.6   Some other race. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float)

4.   Hispanic/Latino and Race

   4.1 Total Population. Data Type : Integer.

      4.1.1   Hispanic/Latino (of any Race). Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

         4.1.1.1 Mexican

         4.1.1.2 Puerto Rican

         4.1.1.3 Cuban

         4.1.1.4 Other Hispanic or Latino

      4.1.2   Not Hispanic/Latino. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

         4.1.2.1 White Alone

         4.1.2.2 Black or African American Alone

         4.1.2.3 American Indian and Alaska Native alone

         4.1.2.4 Asian alone

         4.1.2.5 Native Hawaiian and other Pacific Islander alone

         4.1.2.6 Some other race alone.

      4.1.3   Two or more races. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

         4.1.3.1 Two races including some other race

4.1.3.2   Two races excluding some other race, and three or more other races.

5.   Total Housing Units. Data Type: Integer.

6.   Citizen, Voting age population.

6.1   Citizen, 18 and over population. Data type: Integer.

6.1.1   Male. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

6.1.2   Female. Data Type for {Estimate – Integer, Margin of Error – Integer, Percent – float, Percent Margin of error – Float).

All measures (Estimate, Margin of Error, Percent, Percent of Margin of Error) have above mentioned categories and sub-categories, which are column – names in the provided data set.

For e.g.,   [Measure] à   [Category] à [Sub-Category]

Estimate is a measure of the Category, Sex and Age and it has total population as its sub-category.

The column name will hence be named as " Estimate!! SEX AND AGE!!Total Population".

The data set has 2 unique columns   i.e.,

1) ID(Code of the considered County), data type is String.

2) Geographical Area Name – Consists of county and state name, data type is String.

3) The column names are {DP05_0001E,DP05_0001M, ……… , DP05_0089PE,DP05_0089PM } .

How can I merge the data with the primary COVID – 19 datasets?

è By making use of Geo ID, County Name from enrichment ACS data set and County FIPS from COVID confirmed cases data set.

How can Enrichment data help analyze Covid-19 Spread? Initial Hypothesis questions:

i) What percentage of total estimate per category is most probable to acquire/spread Covid-19?

ii) How can categorical interdependencies affect the spread of Covid-19?

iii) One county can talk about how each category within it (for, e.g.Within Race, Blacks and whites) is affected differently, which can help determine spread trends for other counties. Cumulatively, it speaks about a state, and subsequently spread trends for the entire country can be established.

DONE>NEXT.

Sytiva Wheeler

ACS Social, Economic, and Housing

Variable Dictionary

Housing Characteristics

Variable	Description	Datatype
HOUSING OCCUPANCY	Information on the amount of occupied or unoccupied housing units.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
UNITS IN STRUCTURE	Different types of housing separated by **amount** of units in one structure, includes mobile homes, RV's, boats, vans, etc.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
YEAR STRUCTURE BUILT	Year the structure was constructed.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
ROOMS	Total rooms a housing unit has, includes the median number of rooms among all housing units.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
BEDROOMS	Total number of bedrooms a housing unit has.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
HOUSING TENURE	Amount of owner and renter occupied housing units,	Estimate - Integer Margin of Error - Integer

	includes average household size of each type.	Percent - Float Percent Margin of Error - Float
YEAR HOUSEHOLDER MOVED INTO UNIT	Amount of people who moved into a unit during a specific year.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
VEHICLES AVAILABLE	Amount of vehicles available to the people housing unit.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
HOUSE HEATING FUEL	Amount of housing units who have a specific type of heating.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
SELECTED CHARACTERISTICS	Number of housing units that are lacking specific facilities (Plumbing, telephone service, etc.)	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
OCCUPANTS PER ROOM	Amount of housing units with a specific number of occupants per room.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
VALUE	Amount of housing units that have a specific value. Only out of owner-occupied units.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
MORTGAGE STATUS	Amount of units that have and do not have a mortgage.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
SELECTED MONTHLY OWNER COSTS (SMOC)	Amount of units out of those with a mortgage whose monthly costs fall in a certain range.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float

SELECTED MONTHLY OWNER COSTS AS A PERCENTAGE OF HOUSEHOLD INCOME (SMOCAPI)	Amount of units out of those without a mortgage whose monthly costs fall in a certain range.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
GROSS RENT	Amount of units whose rent falls within a certain range.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
GROSS RENT AS A PERCENTAGE OF HOUSEHOLD INCOME	Amount of units that pay a certain percentage of household income	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float

Social Characteristics

Variable	Description	Datatype
HOUSEHOLDS BY TYPE	Amount of households with specific individuals/ families living in them (Married couple family, Male householder, female householder, etc.)	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
RELATIONSHIP	Relationship each individual in a household is to the householder.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
MARITAL STATUS	Amount of people with specific marital statuses.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
FERTILITY	Amount of women, age 15-50 within the past 12 months who have given birth and are either married or unmarried. Number is per 1000 women.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
GRANDPARENTS	Amount of grandparents	Estimate - Integer

	living with their grandchildren who are responsible for them, for a certain amount of years, and who are a specific sex	Margin of Error - Integer Percent - Float Percent Margin of Error - Float
SCHOOL ENROLLMENT	People enrolled in school, separated by different school levels (nursery, elementary, middle, etc.)	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
EDUCATIONAL ATTAINMENT	Highest level of education of people 25 or older.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
VETERAN STATUS	Amount of veterans in 18 or over population	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
DISABILITY STATUS OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION	Amount of people with disabilities in specific range or years. (under 18, 18 - 64 years, etc.)	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
RESIDENCE 1 YEAR AGO	Amount of people who, within a year, have or have not changed their homes and if so, changed to one in another place.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
PLACE OF BIRTH	Amount of people who were born in certain places.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
U.S. CITIZENSHIP STATUS	Amount of people either born in the U.S. or not.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
YEAR OF ENTRY	Amount of people who entered the U.S. before or after 2010	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error -

		Float
WORLD REGION OF BIRTH OF FOREIGN BORN	Amount of people born in certain countries outside of the U.S.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
LANGUAGE SPOKEN AT HOME	Amount of people who speak specific other languages and if they speak english less than fluently.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
ANCESTRY	Amount of people who are of a certain ancestry.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
COMPUTERS AND INTERNET USE	Amount of households who have some kind of internet access.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float

Economic Characteristics

Variables	Description	Datatype
EMPLOYMENT STATUS	Amount of employment status of people who are a specific age, gender, or have children.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
COMMUTING TO WORK	Amount of people who commute to work using specific means and are 16 and over.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
OCCUPATION	Amount of people 16 and over in specific types of occupations.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float

INDUSTRY	Amount of people 16 and over who work in a specific industry.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
CLASS OF WORKER	Amount of people 16 and older who work in a specific working class.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
INCOME AND BENEFITS (IN 2019 INFLATION-ADJUSTED DOLLARS)	Amount of households who fall in a certain range of income including government benefits. Includes median and mean income among total households.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
HEALTH INSURANCE COVERAGE	Amount of noninstitutionalized people in certain age groups and statuses of employment who do or do not have health insurance.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float
PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL	Percentage of families/people among different categories of families and ages of people, whose income within the past year has been below the poverty level.	Estimate - Integer Margin of Error - Integer Percent - Float Percent Margin of Error - Float

Enrichment Dataset Description

These enrichment datasets contain information about the United States population in regards to housing, economic and social characteristics. Social characteristics include specific information regarding the occupants of a household. This includes things such as household type, marital status, birth place, citizenship status, ancestry, spoken language used at home, year of U.S. entry, internet access available to the household, etc. Housing characteristics would include more specific and intimate information related to the type of housing one may have. This includes occupancy of housing units, housing structure and year it was built, rooms in the structure, choice of heating type (coal, electricity, kerosene, etc), occupants per room, mortgage status, etc. Economic characteristics would include things related to a person or household's financial status, including their employment, class of employment, health insurance, industry of employment, etc.

Data Merging

I believe this data can be merged with the COVID-19 dataset by using the information we have in the housing,social and economic datasets for each county since the COVID-19 dataset is separated by county as well.

How it helps COVID-19 spread analysis

Having data on the amount of people per household can be used to show how much more or less likely a household is able to spread the disease. Knowing the jobs many people have might show who is more or less likely to get the virus based on profession. For example, maybe since teachers interact with many children, there are more cases of the virus among people in that profession than many others. My hypothesis is that we can find housing,economic,and social aspects that people have that will correlate with the amount of cases and/or deaths which may be able to show us which aspects of someone's housing, economic or social status can make them more susceptible to contracting COVID-19.

DONE>NEXT.

Task 3: (50 pts)

- Team: (20 pts)
    - Create a team notebook to read in the COVID-19 data (cases, deaths, and population) using pandas and display the dataframe in a notebook.
    - Merge all the three variables (cases, deaths, and population) to create a super COVID-19 datafame. Export it to a csv format.
- Member: (30 pts)
    - Calculate COVID-19 data trends for last week of the data. Are the cases increasing, decreasing, or stable? Each student chooses a state to analyze.
    - Each student member creates notebooks to read the Enrichment data and displays them on a notebook.
    - Each student member performs initial merges with the COVID-19 data using the variables in the Enrichment data.

Deliverable Each member creates separate notebooks for member tasks. Upload all notebooks to Github Repository.

‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡‡

COVID TREND ANALYSIS:

(**Aka'sh**) I will analyze the state of Pennsylvania. Covid related deaths have been increasing steadily with recent deaths occurring in every county with covid deaths above 100. Confirmed cases are also on the steady rise with more heavily impacted counties having a more drastic increase in the last week.

(**Joseph**) I will analyze the state of Wyoming. Cases in Wyoming are on a rapid rise where we saw 400 cases on average in August quickly became 500 new cases daily. This fast spread could be due to a variety of variables including that Wyoming is more rural than some other states and where some saw spikes Wyoming lagged a few weeks behind. Covid deaths have been staying relatively steady yet saw a recent spike, increasing from 25 to 40 in one day. Neither death nor cases appear to be dipping (beyond the fact that Wyoming is the least populous state in America) and will likely increase based on the given trends and information gathered.

(**Chandra Shekhar Malgari**) I have analyzed the state of Indiana. During the last week the number of confirmed cases and deaths in Indiana state were increasing. The cases and deaths curves are almost flat on the 5th and 6th day. On the last day both the cases and deaths increased.

(**Chandan Kumar Chunduru**) I did my analysis for the state of Texas .For the last week in the dataset the number of confirmed cases have been increasing steadily.Coming to the death count, these are significantly high and have been increasing steadily over last week.

(**Sytiva Wheeler**) After analyzing the state of New York's cases of Covid-19 to see whether they were increasing, decreasing or constant throughout the past week, it seems that the cases are steadily increasing over the course of the week. Starting on September 13th to the 20th, about 5000 cases were reported on the 13th and over the week it fluctuated steadily, but constantly increased all the way to the 20th.

DATA MERGE METHODOLOGY:

The cases and deaths csv have the same columns with the same data, except for the dates, where the columns are the same but the data is different. Simply extending the data would just increase the data, and there would be no indicator of a shift in data type from cases per date to deaths per date. There would just be numbers. The states have dynamic data that is tracked across an increasing vector via dates. My solution is to extend the index and add data to the row beneath the date for each index. If that is not possible I would edit the columns such that there we be two versions of a certain date, one that holds data for cases, and one that holds data for deaths.