

# COVID Check3

Chen, Xiangnan & Cui, Augustine

2020/10/20

## Introduction

COVID-19 is an infectious disease caused by the most recently discovered coronavirus. The disease spreads primarily from person to person through small droplets from the nose or mouth when people with COVID-19 cough, sneeze or speak [1]. The current coronavirus disease COVID-19 pandemic is hitting the globe unprecedentedly. Lives have been taken, and economic activities have been stagnated. Thus, it is crucial for people to understand better this global pandemic's current situation and future development to be better prepared to solve this global crisis. In this report, we are using COVID-19 data gathered in the United States to check if there is a linear relationship between date and cumulative cases.

## Background

Among all the studies that focus on COVID-19, we find three sources interesting. The first is CDC's weekly report that gives an overview of the COVID-19 related hospitalizations and deaths. The data from Utah analyzes the relationship between COVID-19 cases, hospitalization, and testing with the deprivation level in different areas within Utah. The third research focuses on the Age distribution of the COVID-19 pandemic.

The CDC report focuses on the relationship between COVID-19 related hospitalizations and mortality. It also includes categorical variables like age and race. The overall cumulative COVID-19 hospitalization rate is 174.8 per 100,000, with the highest rates in people aged 65 years and older (472.3 per 100,000) and 50–64 years (261.5 per 100,000).[2] The second report analyzes COVID-19 Data from the Utah Department of Health. It examines the relationship between cases and hospitalization, the relationship between testing and the level of deprivation. The report shows that the infection number of Utah's high-deprivation areas is three times higher than the lower-deprivation areas, so does hospitalization and testing rates[3]. The third research focuses on the relationship between COVID-19 and ages. The source researches the relationship of incidence between each age group. It also calculates the weekly median age of people with COVID-19. The research shows that the distribution centers at age group from 20-39 years, which means that younger adults are more likely to contribute to community transmission of COVID-19[4].

Our approach to COVID-19 is to analyze the relationship between the overall cumulative number of cases and the CDC report date. Different from the previous studies, we are using the cumulative number of cases. We believe that using the cumulative number can offer better interpretability when constructing a model with date (Time series data).

## Data

There are three quantitative variables 'cdc\_report\_dt', 'pos\_spec\_dt' and 'onset\_dt'. There are eight categorical variables 'current\_status', 'sex', 'age\_group', 'Race.and.ethnicity.combined', 'hosp\_yn', 'icu\_yn', 'death\_yn' and 'medcond\_yn'. Description of each variable is in the Appendix.

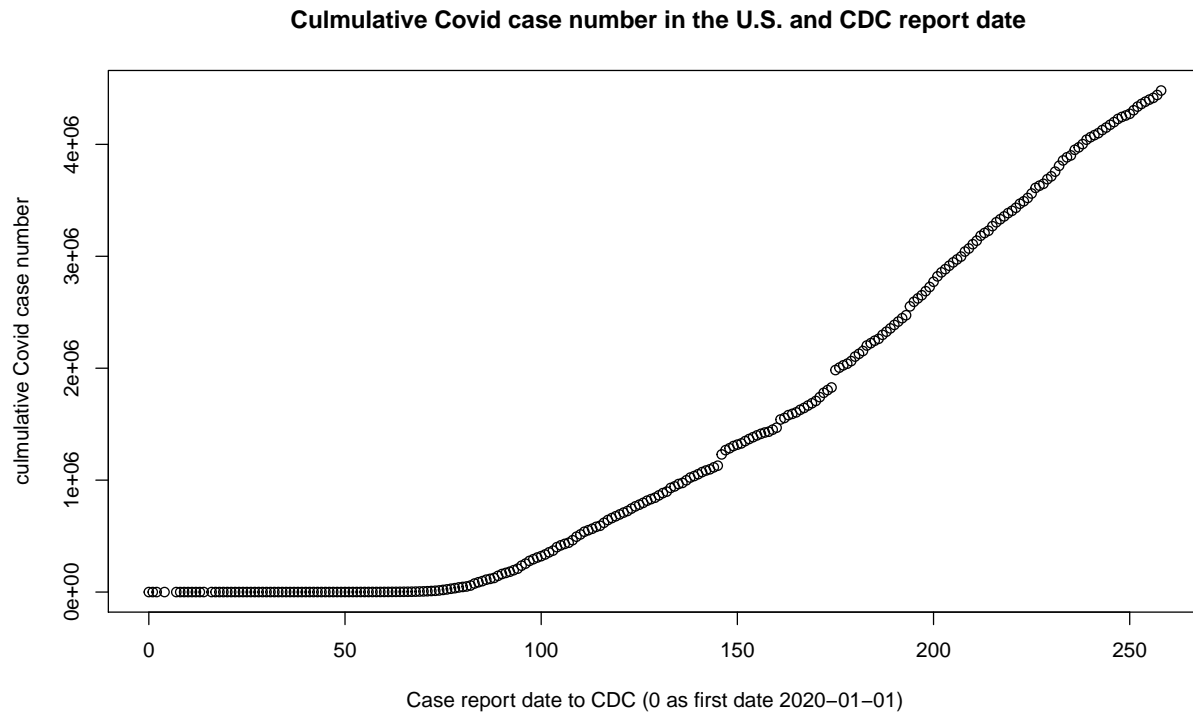
There are observations and each observation contains variables.

```
cov <- read.csv("COVID-19_Case_Surveillance_Public_Use_Data.csv")

reportDate <- unique(cov$cdc_report_dt)
case_date <- count(cov, cov$cdc_report_dt)
```

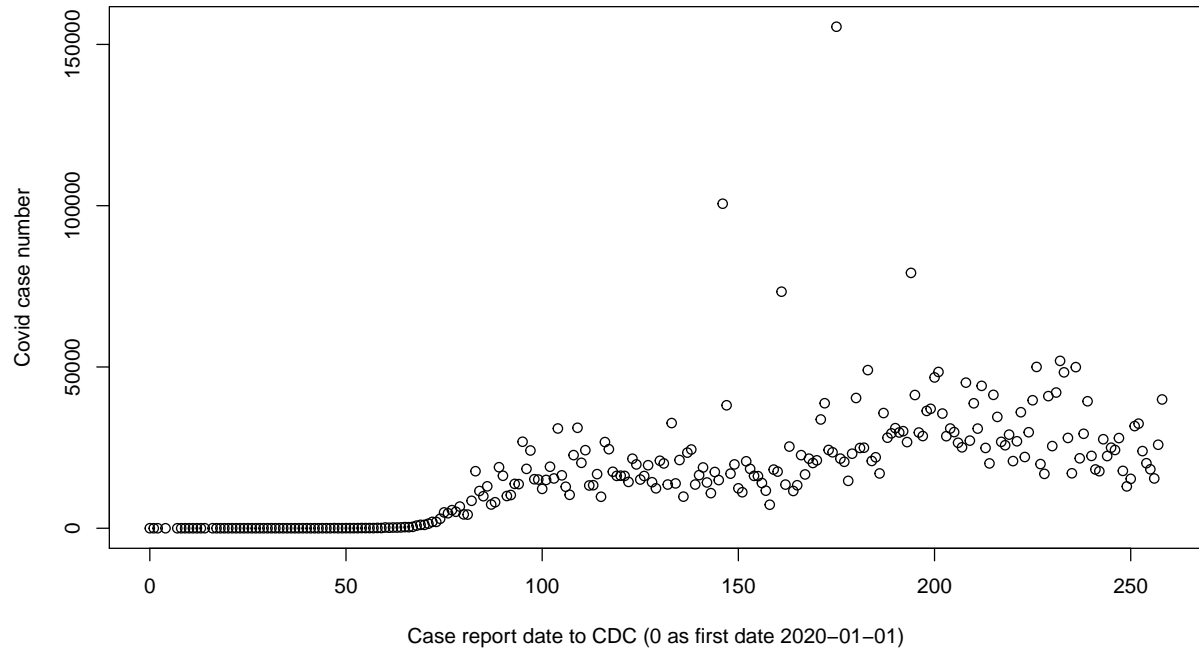
```
case_date$date <- as.Date(case_date$cov$cdc_report_dt`)
case_date$date <- case_date$date - as.Date("2020-01-01")
case_date$cum_case <- cumsum(case_date$n)
```

```
plot(case_date$date, case_date$cum_case, xlab = "Case report date to CDC (0 as first date 2020-01-01)",
      ylab = "culmulative Covid case number", main = "Culmulative Covid case number in the U.S. and CDC")
```



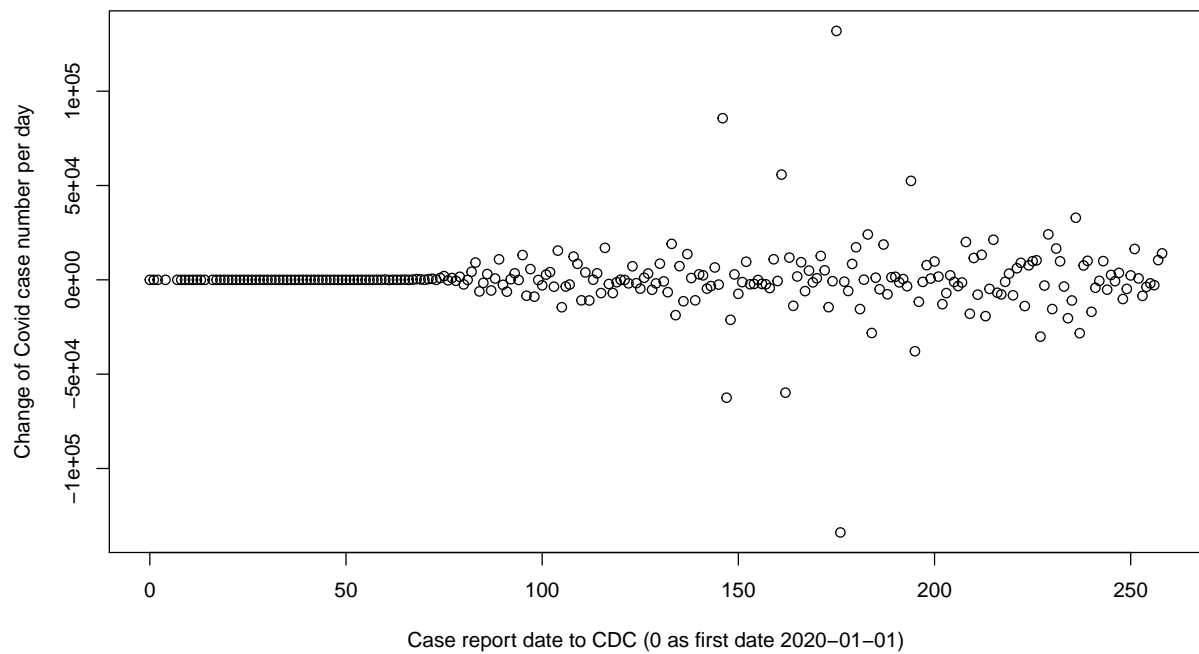
```
plot(case_date$date, case_date$n, xlab = "Case report date to CDC (0 as first date 2020-01-01)",
      ylab = "Covid case number", main = "Covid case number in the U.S. and CDC report date")
```

**Covid case number in the U.S. and CDC report date**



```
plot(case_date$date, c(0, diff(case_date$n)), xlab = "Case report date to CDC (0 as first date 2020-01-01)",
      ylab = "Change of Covid case number per day", main = "Change of Covid case number in the U.S. and CDC report date")
```

**Change of Covid case number in the U.S. and CDC report date**

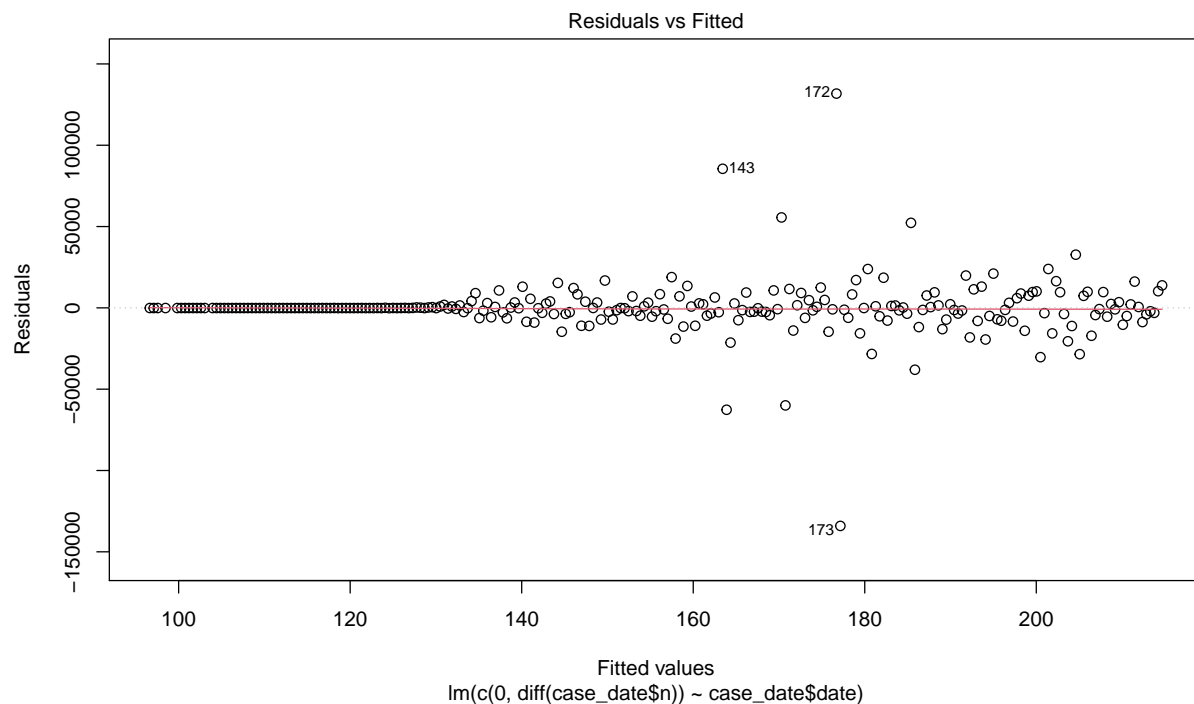


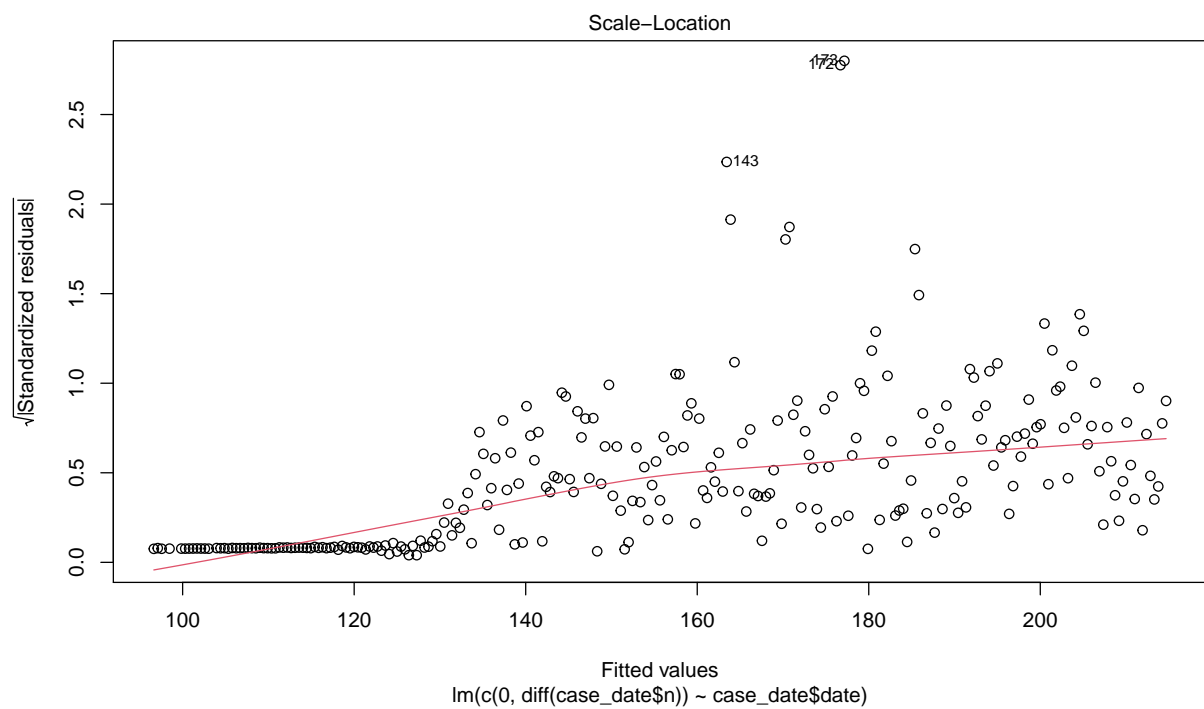
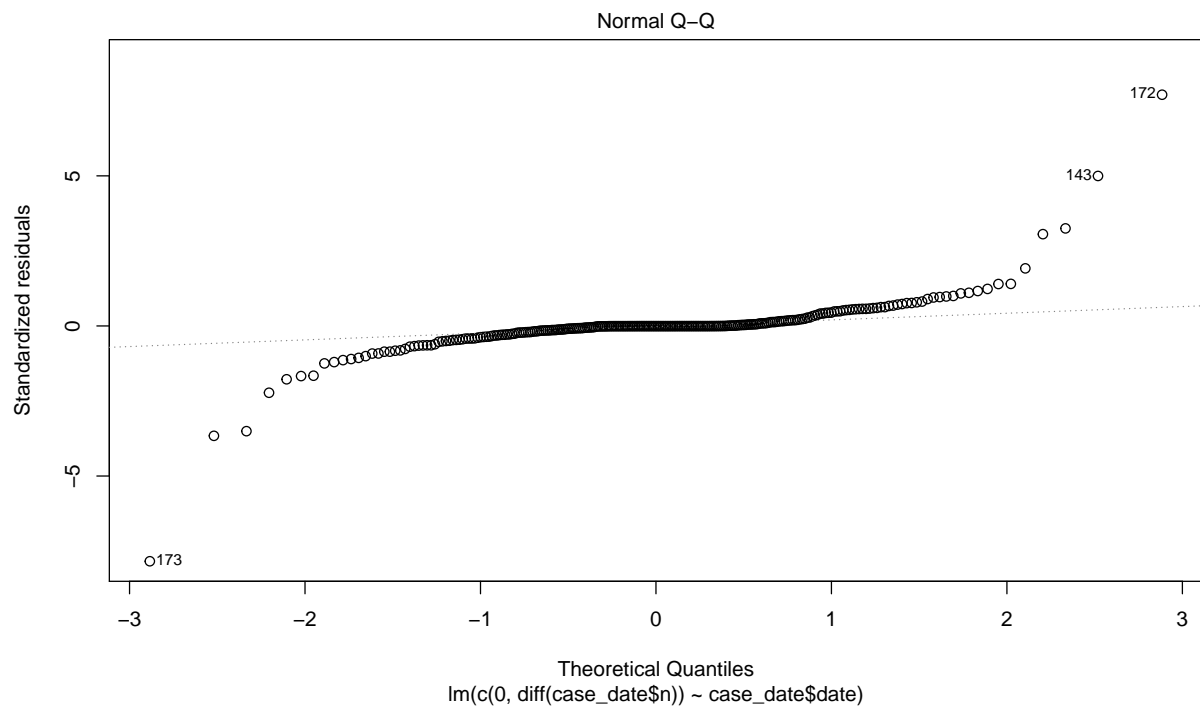
There is a strong positive nonlinear relationship between the CDC report date and cumulative Covid case

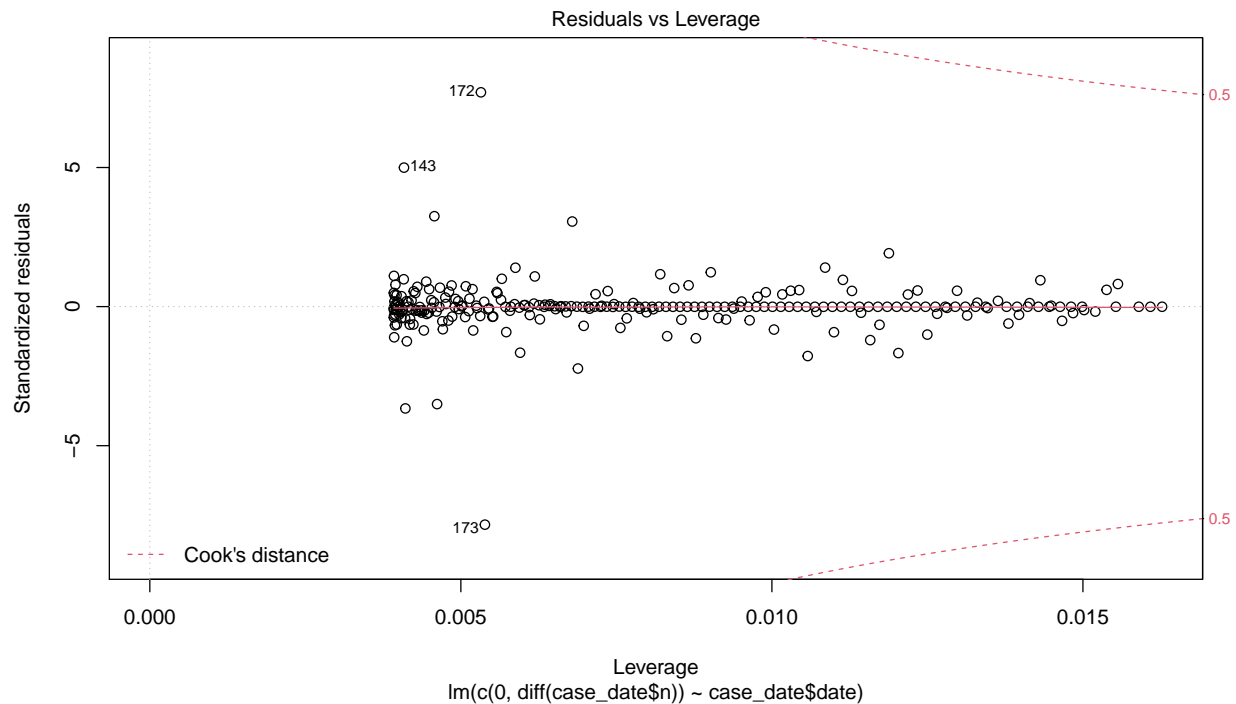
number. We can see that among the first 80 dates—there are very few cases reported to CDC. After the 100th day, there is a strong linear relationship between the CDC report date and cumulative Covid case number. To our surprise, there is not an obvious trend showing an increase in slope. In contrast, it is rather linear after the 100th day.

We are not using date of first positive specimen collection ('pos\_spec\_dt') and date of symptom onset ('onset\_dt'), because we specifically choose to use the CDC report date as our predictor variable. We are currently not using any categorical variables within our model because we want first to see the relationship between the case number and the CDC report date.

```
changeMod <- lm(c(0, diff(case_date$n)) ~ case_date$date)
plot(changeMod)
```







## References

- [1] Q&A on coronaviruses (COVID-19) <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses>
- [2] “COVIDView: A Weekly Surveillance Summary of U.S. COVID-19 Activity,” Centers for Disease Control and Prevention. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>. [Accessed: 28-Sep-2020].
- [3] Lewis NM, Friedrichs M, Wagstaff S, et al. Disparities in COVID-19 Incidence, Hospitalizations, and Testing, by Area-Level Deprivation — Utah, March 3–July 9, 2020. MMWR Morb Mortal Wkly Rep 2020;69:1369–1373. DOI: [http://dx.doi.org/10.15585/mmwr.mm6938a4external icon](http://dx.doi.org/10.15585/mmwr.mm6938a4external%20icon)
- [4] Boehmer TK, DeVies J, Caruso E, et al. Changing Age Distribution of the COVID-19 Pandemic — United States, May–August 2020. MMWR Morb Mortal Wkly Rep. ePub: 23 September 2020. DOI: [http://dx.doi.org/10.15585/mmwr.mm6939e1external icon](http://dx.doi.org/10.15585/mmwr.mm6939e1external%20icon).
- [5] Centers for Disease Control and Prevention. COVID-19 Case Surveillance Public Use Data, September 30, 2020 <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf>

## Appendix

Column.names	Description	Type
cdc_report_dt	Initial case report date to CDC	Quantative: Date & Time
pos_spec_dt	Date of first positive specimen collection	Quantative: Date & Time
onset_dt	Symptom onset date, if symptomatic	Quantative: Date & Time
current_status	Case Status: Laboratory-confirmed case; Probable case	Categorical: Text
sex	Sex: Male; Female	Categorical: Text
age_group	Age Group: 0 - 9 Years; 10 - 19 Years; 20 - 39 Years; 40 - 49 Years; 50 - 59 Years; 60 - 69 Years; 70 - 79 Years; 80 + Years	Categorical: Text
Race and ethnicity (combined)	Race and ethnicity (combined): Hispanic/Latino; American Indian / Alaska Native, Non-Hispanic; Asian, Non-Hispanic; Black, Non-Hispanic; Native Hawaiian / Other Pacific Islander, Non-Hispanic; White, Non-Hispanic; Multiple/Other, Non-Hispanic	Categorical: Text
hosp_yn	Hospitalization status: Yes/No	Categorical: Text
icu_yn	ICU admission status: Yes/No	Categorical: Text
death_yn	Death status: Yes/No	Categorical: Text
medcond_yn	Presence of underlying comorbidity or disease: Yes/No	Categorical: Text