

CSC343 Phase 4
Shi Tang, Yuxin Yang
23, November 2020

Decisions made when cleaning the data:

- For table Unemployment: When importing data to table Unemployment, the original dataset has a different format than our schema. We want the State abbreviation code as the primary key, but the dataset only contains the full name instead of the abbreviation code of each state. Therefore, we need to create a temporary table to store the dataset, then join the temporary table with table StateOverview to get the State code.
- For table StateOverview and Cases: The attribute State code (sCode) in table Cases references the state code in table StateOverview, but the data of table Cases originally contained a few State codes that are not in table StateOverview, which are Guam, Northern Marianas, Virgin Islands and American Samoa. We decided to remove these states since we have implemented a foreign key constraint that Cases.sCode refers to StateOverview.sCode. Therefore, we created a temporary table to store the original data, then used a *where* clause to filter out those states we don't want:
 - *create table Cases as(select * from temp2 where sCode != 'AU' and sCode != 'GU' and sCode != 'MP' and sCode != 'PR' and sCode != 'VT' and sCode != 'AS')*
- For table MaskMandate and EnforceMask: The data for these two tables was originally in the same large dataset. The original dataset records the date when each state mandated face mask wearing of either all individuals or employees only. The date is left blank if the state did not enforce any mask mandate. We decide to add one new attribute enforcementLevel in the table MaskMandate, indicating whether or not the state mandated face mask wearing. We include only the states that have the mandate in the table EnforceMask, and record the type of people applicable to the mandate (all individuals/employees only), the enforcement measures (by fines/by criminal charge/both/none), and the enforcement date. The original table records the enforcement date separately for different types of people. We decided to integrate the dates into one attribute *dateMandate* by recording the later one of the two dates.
- For table QuarantineMandate: The original table includes all the states with some states not enforcing the quarantine mandate, so some of the cells of the enforcement date are blank. We decided to record only the states that implement the quarantine mandate with the enforcement date in the table QuarantineMandate.