INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

Introduction:

For our team project, we wanted to research and record information that was current and relevant which is why we decided to look at data surrounding the COVID-19 pandemic. We knew that we could find a lot of data relating to COVID-19 because it had such an impact on everyone. We were looking at current news which was full of debate over whether lockdowns and vaccine mandates were effective and necessary for preventing deaths. Our hypothesis is that lockdowns and vaccine mandates do in fact prevent deaths, but due to there being so much controversy around this subject, we wanted to investigate it by collecting data related to COVID-19. We wanted to see how much of an impact COVID-19 had on the United States. To answer these questions, we wanted to document case rates, mortality rates, and vaccine rates in the United States by country, state, county, and select cities, so we decided to use those in our database. As college students, we also chose to include data from colleges in the United States. By collecting this data, we hoped to create SQL Views and Queries which could prove our hypothesis.
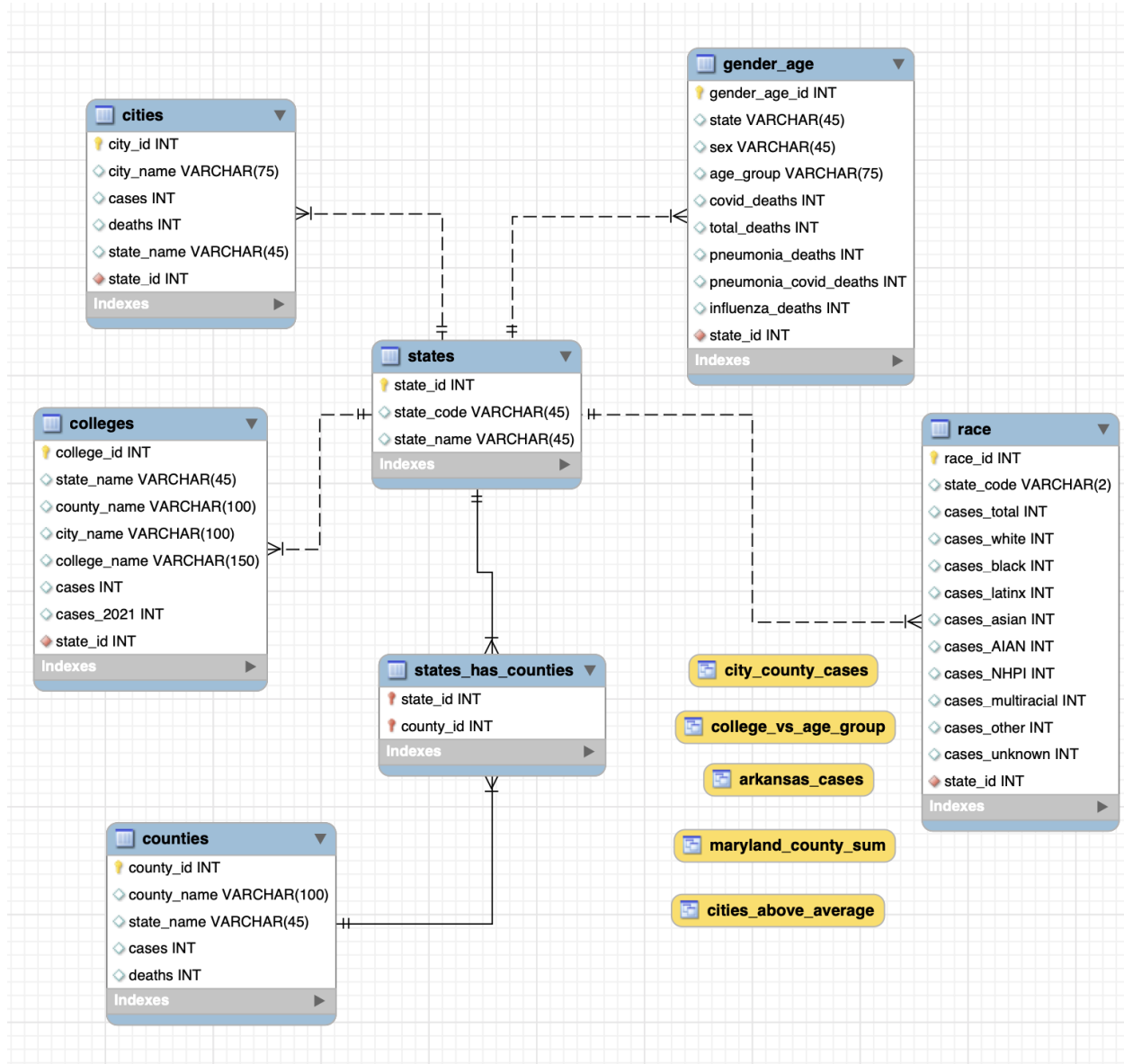
Database Description:
Logical Design (include a PNG of your ERD in this section):

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

Our database includes 7 tables. For the logical design of our database, we wanted the states table to include COVID-19 data for the United States and also act as the main linking table. The rest of our tables branch out from the main states table and they include the following: counties, cites, colleges, and social demographics like race and gender. We chose this order because we wanted the main focus of our table to be COVID-19 information for each state

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

within the United States but still wanted to include detailed information about specific counties and cities.

One of our main priorities for the database was to include relevant COVID-19 information for each of our tables. Since we did not want users to deal with information overload, we focused on the most important information to provide the broader picture. For example, we broadly focused on the number of COVID-19 cases and the number of COVID-19 deaths for each county or city, but we did not include certain information like vaccination rate for each company or school within the state. We wanted to make our information as concise as possible while still maintaining the overall comprehensiveness of our data and structure.

Sample Data:

Our data for this database came from a couple different sources. The counties and colleges data came from a New York Times database and the gender/age and race data came from the CDC Covid Data Tracker. Finally, we created the cities data ourselves. Our tables have data for the number of covid cases, covid deaths, and total deaths. Below is a portion of our gender/age table which contains data on covid deaths vs total deaths across numerous different age groups and sexes for all of the states.

gender_age_id	state	sex	age_group	covid_deaths	total_deaths	state_id
1	Alabama	All Sexes	All Ages	15877	121980	1
2	Alabama	All Sexes	Under 1 year	0	756	1
3	Alabama	All Sexes	0-17 years	0	1353	1
4	Alabama	All Sexes	1-4 years	0	167	1
5	Alabama	All Sexes	5-14 years	0	212	1
6	Alabama	All Sexes	15-24 years	34	1404	1
7	Alabama	All Sexes	18-29 years	108	2355	1
8	Alabama	All Sexes	25-34 years	211	2739	1
9	Alabama	All Sexes	30-39 years	332	3492	1
10	Alabama	All Sexes	35-44 years	551	4467	1
11	Alabama	All Sexes	40-49 years	858	5898	1
12	Alabama	All Sexes	45-54 years	1295	8284	1

Views / Queries:

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

View Name	Req. A	Req. B	Req. C	Req. D	Req. E
city_county_ca ses	A	B	C	D	
maryland_cou nty_sum	A	B	C	D	
cities_above_a verage		B	C		E
college_vs_age _group	A	B	C		
arkansas_case s	A	B			

1) **City_county_cases:** Combines the covid cases numbers of counties and cities for a state
2) **Maryland_county_sum:** Sums all of the counties cases of maryland
3) **Cities_above_average:** Averages all of the covid cases for cities and then finds those that are above that average number
4) **College_vs_age_group:** Combines the sum of all cases for all for all the colleges in the state with the death of people in that age group within the state
5) **Arkansas_cases:** Combines cases from all colleges and races for the state of Arkansas

Changes from original design:

      When we began writing our project proposal, we had multiple ideas that we wanted to include in our COVID-19 database, however many of them didn't make it into the final product due to various different reasons. Our original plan for the scope of our database was to observe

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

how COVID-19 and vaccination rates have impacted the world, with a focus on more populated areas, such as the United States, China, countries in Europe, etc. We planned to pull the top 20 countries, as well as data from each state in the United States, and use data from them to build our database. However, we decided that this was too large a feat, and would take too long. So we made the decision to focus only on the United States. In our original plan, we also intended to include hospitalization rates, unemployment rates, and prison data in our database to get a better understanding of each location. However, we already had a large amount of data to manage through, and made the decision that attempting to include other information that wasn't directly related or didn't have too much of an impact on understanding the effects of COVID-19 would be too difficult in tandem with everything else.

One key change we made to our database was to provide COVID-19 information related to social demographic factors. So, we added race and age into our database because these are factors that have a substantial impact on COVID-19 rates and provide better context as to why each location was affected to the extent that it was. Furthermore, this information helps us understand the impact of COVID-19 on different demographic groups within the United States. This information was relatively easy to find, so we had no issues deciding to include it in our database

Lessons Learned:

Our group learned a lot from this final project. We had to make changes to our project plan because we realized that it was going to become difficult to find all the data for the tables we want. If it was difficult to find the data online then we had to possibly change the tables around or think of something else. We realized how complex and complicated data collection is. We had to be more flexible in terms of changing our tables to adapt to new information we found that we knew would work better. We originally were going to include data from countries around the world, but we ended up having to just focus on the United States and coming up with tables that only had U.S data relating to COVID-19. Reaching out to our TA when we were having

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

doubts on our project was something that ended up being really helpful. They gave us suggestions and answered any questions we had.

In addition to the data design and data collection process, we also developed stronger communication skills. Communication is really important when working in a group to successfully finish a project. From the beginning we learned that we needed to communicate in our group chat about any assignments relating to our final project well in advance so we would avoid waiting until the last minute to finish them. All of us had busy schedules so it was hard sometimes to find a time for a meeting that worked for everyone so we ended up having to divide substantial portions of work for assignments like our progress report. However, we all still reviewed and edited each other's sections to polish our final product and gain stronger familiarity with each aspect of our project.

Potential Future Work:

Our original plan was to look at COVID-19 data from multiple countries around the world, but because of mainly time and how complex our database would be, we had to adjust and just focus on the United States. Our potential future work on this project would be expanding our research on finding data from other countries such China, India, countries in Europe, etc. We also wanted to record data overtime. For example, having a table for monthly county increases over the course of the pandemic to track the progress of Covid. We also can include more columns with COVID-19 data that we couldn't include in our tables now. Data such as hospitalizations, unemployment, and other factors that had an impact during the pandemic. We are still in the pandemic so there will be new data that we can possibly add to our existing tables or even make new tables. There is a new booster shot for the COVID-19 vaccine, so slowly data will be collected for that, which we can then add to our database.

INST327 0203
Project Report: 12/15/21
Team 4
Avani Badugu, John Spurrier, Joanna Song, Renee Paulraj, Noah Caple

Works Cited

"CDC Covid Data Tracker." *Centers for Disease Control and Prevention*, Centers for Disease
Control and Prevention, https://covid.cdc.gov/covid-data-tracker/#datatracker-home.

Nytimes. "Nytimes/COVID-19-Data: An Ongoing Repository of Data on Coronavirus Cases and
Deaths in the U.S." *GitHub*, https://github.com/nytimes/covid-19-data.