1

Datasets to use  
- John Hokpins: Total cases, total deaths, total deaths per 100,000   
(population), case fatality ratio (how many people die per hundred cases),   
active cases, recovered cases  
- Population  
- Vaccinations (if time doesn’t allow for using 3 datasets, we can get rid of   
this one?)  
Steps to do  
1. Design database schema ERD – Hesh (& Megan if needed)

Megan will: -

* add “boosted-per100-people” to the Vaccination Table,
* complete the ERD diagram,
* generate the SQL create tables code,
* create the Database and Tables,
* load to “Project-2/Megan” on Github for team review.

1. One database – name it covid\_case\_db

I suggest we name it “Integrated-Covid-View\_db” team to reflect that is not just another view of Covid but an integrated one – see what you think -Mike

1. Tables: (ensure there are minimal columns)  
   i. JHU Covid  
   •(columns)  
   •(Mike will fill these) COMPLETE – READY FOR REVIEW)  
   ii. Population  
   •(columns)  
   iii. Vaccinations  
   •(columns)  
   2. Extract data - use Mike's script (won’t need to add much to this) –   
   Already done – Mike will tidy this  
   3. Transform data to tables - use Mike's script as basis  
   a. Use new functions to clean/transform data (from session 13.1) –   
   Mike DONE  
   i. E.g. drop duplicates, missing/no, pandas profiling  
   b. Develop new tables (based on schema)  
   c. Could join tables together  
   4. Remove visualisation script – Mike DONE  
   5. Load tables to database - this will use ETL activity - Megan  
   a. Connect to database  
   b. Load tables to database  
   6. Write report (ensure there is acknowledgement of data, and whoever   
   cleaned data – leverage Project 1 README) – Group effort (Monday)  
   7. README (make extensive – similar to Mike’s project 1 README) – Mike?  
   8. Presentation (use README)

Hesh – we didn’t discuss this at today’s zoom call, but could I suggest that you start thinking about how best to present our “Integrated Covid Data Base”? It would be good to have some preliminary ideas in place when we meet on Monday.