My thoughts on getting the data Team,

Based on Uday’s feedback and my tutor’s feedback option 2 is the most elegant solution and I think we run with that. Let’s see how it goes and review tomorrow in class.

Getting the data – option 1: -

1. Run the “create-covid-db” Jupyter Notebook to get the latest Covid, Vaccination and Population data (say, at March 15th for example) and load it to the PostrgeSQL DB,
2. Use Flask to extract each of the DB tables as per the following:-

<https://jasonrhaas.com/2017/07/17/adding-a-simple-api-to-your-postgres-database.html>

1. Extract the most recent JHU data – the Notebook downloads all the time series data for the entire Covid pandemic – a bit over two years.
2. Extract the Vaccination and Population data.
3. Convert the data to CSV format.
4. Merge the most recent Covid data with the Vaccination and Population data based on Country-id

(I’m not sure at this stage how best to easily merge these – I’ll get some advice from my tutor this afternoon.)

1. Use the merge CSV file to drive our visualisations.

OR

Getting the data – option 2: -

1. Run the “create-covid-db” Jupyter Notebook to get the latest Covid, Vaccination and Population data (say, at March 15th for example) and load it to the PostrgeSQL DB,
2. Use PostgreSQL to extract the most recent JHU data and merge it with the Vaccination and Population data
3. Use Flask to extract the merged data frame as per the following:-

<https://jasonrhaas.com/2017/07/17/adding-a-simple-api-to-your-postgres-database.html>

1. Use the merged data frame and D3.json to drive our visualisations.