Sept 17th:

* Group created, all met each other
* Lab session with vocabulary game in order to come up with a bunch of potential project ideas / potential data.
* This process was exhausted until we had plenty of ideas

Sept 24th:

* Started investigating data sets around the ideas we came up with last week, and fast prototyping to get a quick idea on their feasibility.
* We made a soft decision that using data related to covid/vaccinations would be the goal of our project.
* Evaluated certain datasets to see which would work, decided against ones like the industry profiler which only had data from 2016-2020

Sept 30th:

- Group meetup to formalize project ideas and relevant data sets.

- Set up Gitlab for Group project

- Uploaded necessary content into Gitlab

Oct 1st:

* Lab session with prototyping carried out with NZ covid-19 dataset
* Reporting any difficulties encountered
* Refinement of project idea / questions to be asked
* Looking for ways to tie our data together and which datasets will allow this
* Main difficulty was mapping national regions and DHB regions to make sure that they align properly

Oct 7th:

* Group meetup to work in tandem on our project
* Figuring out what is the best way to scrape our information
* Discussing further underlying variables to relate our data across
* Figuring out how to directly download files from github within R
* Getting up to speed on where we are at

Oct 8th

- Entire group lab session

- Review session with peer group, discussing tactics, future plans, current problems

- Continuing our problem solving regarding relating datasets together, our regions for datasets have not lined up properly. Possible solution found regarding scraping DHB population data directly from their website

- Found a potential 3rd variable in the wellbeing measure dataset by means of generalised trust in health system per region

Oct 12th

* Group meet up
* Discuss work done on our own time
* Collaborating to discuss whether it is worth using the wellbeing dataset regarding joining issues
* Wellbeing dataset very messy and proving difficult to extract the information we want in an effective and efficient way
* Figuring out how to combine columns and remove N/A’s in the process - one column of variables was stored in multiple columns in our dataset and we had to find a way around this
* Problem solving searching for possible solutions

Oct 15th

* Lab session with entire group attendance
* Working directory error on some people’s laptops and not on others
* Seeking help from tutors
* Aligning regions to match with monthly job database and population database in order to produce a per capita plot
* Error with wellbeing function on September quarter dataset, very confusing

Our datasets and why we chose them:

Covid-19 Dataset:

This was a large dataset with lots and lots of information on societal, socioeconomic factors etc. all of which would be useful and are related to covid-19. By region as well, so useful for our project.

Wellbeing Dataset:

This dataset had interesting measurements on wellbeing by region which fits perfectly into our project idea. Moreover it it messy, so we thought we could display our skills by wrangling with this dataset.

Border crossings:

This dataset allows for another factor that would be affected by covid-19 and is examinable by region. It will proudly display lockdowns effects on the amount of people moving around, in and out of the country.

Population dataset:

This dataset would allow us to create per capita plots, to filter out the discrepancy of Auckland being much larger than any other region in New Zealand, meaning results will be more representative of a model population rather than actual population which would be disproportionate.