Group Log

Date	Decision(s)/Contribution(s)	Issue(s) Encountered	Person(s) Responsible
19/09/2022	During today's lab tutorial session, our team discussed possible project ideas or topics that each of us is interested in. By the second half of the tutorial, we decided to go with Henri's suggestion, which is how agricultural activities affect rates of illnesses. Furthermore, we decided to use GitHub to facilitate collaboration and version control. Lastly, we settled our meeting schedules for tomorrow, Thursday, and Friday.	The main issue we encountered today is the topic on which we want to investigate and do data wrangling.	All
20/09/2022	We started our meeting with HanByeol walking us through Github, which we will use to ensure our work is consistent. After that task, we jump into the main objective of our meeting. We initially finalised our chosen topic by formulating the question we wanted to answer and focus on. Then we discuss the indicators or variables we need to answer the question. We identified four possible indicators, which we believe can be broken down into more variables. Henri provided helpful information that enabled us to understand the topic further	Identifying the indicators or the data we wanted to use to answer our topic.	All

	and tips for the indicators. I (Alzen) found some possible data sources that provided a good place to gather datasets related to our topic. Lastly, each of us chose the indicators we wanted to start working on.		
22/09/2022	Today we started by organizing the meeting schedules for next week. After that, Henri, HanByeol, and I (Alzen) began discussing the problem with the indicators the group identified in our last meeting. We decided to discuss it in detail. Instead of thinking about the individual variables, we started thinking about the data we needed as datasets, each containing information about a specific subject. Then, Henri and I (Alzen) constructed a dataset about the agricultural density (or activity) based on the data we had already found and Henri found during our discussion. Moses joined us halfway, and then HanByeol walked us through Github again by demonstrating the routine that we will do throughout our project. By the end of our meeting, we decided that Moses and Henri would work with the agricultural density dataset.	Identification of appropriate variables that will be included in the agricultural density dataset.	All
22/09/2022	Later that night, I (Moses) was looking through the datasets gathered earlier this afternoon. I decided to start manipulating a little on those datasets by removing	Understanding of why there were NAs in those datasets. The NAs were found to be caused by censorship for	Moses and Henri

	NAs.	protecting business interests and farming practices.	
23/09/2022	Continued working through the datasets relating to agricultural activity. Removed irrelevant data. Standardised date format. Made two different datasets relating to horticultural and livestock activity, joinable into one data frame.	Identifying the relevant functions available in R to use for cleaning the data.	Moses and Henri
23/09/2022	HanByeol and I (Alzen) worked on the water quality dataset. We started by going through the LAWA website and looking for existing groundwater, river, and lake quality datasets. We are joining these datasets with the ones that we already found. However, we can't find any, so we plan to scrape the web pages containing the necessary information. We started by looking for some patterns in the HTML structure to create helper functions to automate scraping.	Finding patterns in the HTML structure of the web pages and a way to scrape the web pages we selected.	Alzen and HanByeol
24/09/2022	Search and explore LAWA website to see what kind of data we can get and how to scrape those.	Realise the data range LAWA has is different and there are many regions does not have enough data we expect.	HanByeol
25/09/2022	Scraping the LAWA website appears to be tedious. So, I decided to look for another option, which is converting the latitude and longitude to address. I was able to do	Understanding the reverse_geocode function.	Alzen

	it using the "tidygeocoder" package. This allows us to join the tables (data frames) we have by region. However, there is a new problem: the lack of relevant data for the analysis and visualisation phase later.		
27/09/2022	Continued working through the datasets relating to agricultural activity. Created keys with two of the datasets using the interaction function. Then transformed the combined datasets from long to wide format dataset and did some basic plotting.	Using the suitable variables when creating keys, figuring out the right _join() function	Moses and Henri
27/09/2022	The team went to the lab session to get the lecturer's opinion about the current progress of our work and advice on how to deal with the problem that we encounter. Based on the lecturer's opinion and advice we decided to isolate the dataset about groundwater quality and compare only specific years due to missing data. However, we decided to still look for some data that can fill up the gaps in the current one. Additionally, we wrote down some alternative topics if the current topic will not pan out and started planning how to make our current work reproducible.	Finding additional data about groundwater quality and thinking of alternative topics.	All
28/09/2022	Worked on a groundwater quality dataset to make data tidy using interaction functions. Transform the dataset to a wide	Dataset's quality is not high enough as we want data in an individual year.	HanByeol

	format.		
29/09/2022	The team had a short meeting to discuss the current problem about the groundwater quality dataset. Initially it is decided that we are proceeding with the plan to compare only certain years, but since we were able to find some data about it we proceeded with the original plan of having a dataset that will help us with comparison by year and region. We also decided to consider adding a new dataset about E. coli and bowel cancer where we need to scrape news sites about it.		Alzen, Henri, and HanByeol
30/09/2022	HanByeol and I (Alzen) started working on the data about groundwater quality. We did some initial analysis about the dataset to find some anomalies that we can fix such as NAs. We then manipulate the dataset by mutating the dataset for readability, selecting the relevant columns, get the rows that satisfy the condition, and add up the measurements. Next, we converted the dataset to a wide format for presentation. Lastly, we assigned each other with the datasets that we are going to manipulate next.		Alzen and HanByeol
03/10/2022	Based on the comments from the TAs I started looking at some possible ways to	Project reproducibility as I am struggling to grasp the concept	Alzen

	visualise and present our data. I found a dashboarding tool using RMarkdown and the team considered using it for presentation and visualisation. Additionally, I am still wrangling the groundwater quality dataset.	of it and how we can execute it as a team.	
03/10/2022	Started working on the data about river-quality(E.coli)using relative functions. Transform long-range datasets into wide format.	There is an issue that I did not decide which column should be use as key to spread data set.	HanByeol
03/10/2022	Worked on data regarding farm area and farm counts. Removed irrelevant data and cleaned it to match the format of the table regarding farm activity. Next step is to join the two tables by area and year to have a single table that records agriculture extent and agriculture activity.	Sharing the data using Github	Moses and Henri
04/10/2022	Created a google doc for our project report		All
04/10/2022	Joined the data regarding agricultural activity together using a key made up of observation "Area" and observation "Year". Made the code reproducible by using URLs to load csv, rather than locally stored csv's which would require downloading csv files separately.	URLs from government sources were unique to each webpage session and were deleted every few minutes, rendering our code unusable after the URLs were deleted. Resolved by hosting the csv's on github, and using github permalinks.	Moses and Henri
04/10/2022	I started exploring and reading about	Understanding the components	Alzen

	flexdashboard. I created some plots using different plotting packages such as ggplot, plotly, and highcharter where two of them have some built-in interactivity embedded to them. I also looked up about shiny package for plot rendering and interactivity-adding functions.	and functions of the packages I am using.	
06/10/2022 to 07/20/2022	I am still exploring the flexdashboard by adding interactivity to the plots that I have made and adding ways for users to interact with the dashboard.	Incorporating interactivity to the plots and finding viable ways to visualise our data is complicated	Alzen
08/10/2022	The team initially discussed and prioritised the deliverables that we need to submit. Then we proceeded to create more visualisations based on the data we have and restructuring our code. By the end of the meeting we made sure that our GitHub is up to date.	Incorporating the visualisations we have to flexdashboard.	All
10/10/2022 to 13/10/2022	I worked on our dashboard for four days trying different available components using shiny with flexdashboard. I created maps using leaflet; however, specific features such as valuebox and gauge are not working as planned. Furthermore, as I delved into it, I'm encountering complicated situations to some extent where I found myself stuck.	Adding shiny components and understanding complicated situations.	Alzen
13/10/2022	Made tidy water quality data frame, and	We wanted to combine E.coli	HanByeol

	generated mock-up line plots for groundwater quality and river quality that are about average E.Coli count and nitrate nitrogen amount in NZ.	and nitrate nitrogen line plots in one plot sheet however E.coli measurements are too large to make line plots of nitrate nitrogen unrecognized.	
13/10/2022	Found a method to scrape data using APIs. Scraped data relating to agricultural activity area extent from the Ministry for Environment (MfE). We are now planning to apply that method to other data sets to make the code reproducible, while showing we are able to web-scrape. Created a single function to web scrape data from MfE	Parsing the XML and converting it into a dataframe; figuring out which node within the XML contains the relevant data.	Henri
13/10/2022	Added additional columns that describe agricultural activity area extent, broken down by sub industry. E.g: area for dairy cattle in hectares	Wrangling the data in general. Standardising the naming convention for region names. Keeping track of which columns are dbl, chr, and int.	Henri and Moses
13/10/2022	Carried on trying different visualisations	Using flex dashboard. Deciding on the level of interactiveness. Using different visualisations to figure out which are the most intuitive	All
14/10/2022	Decided to abandon flex dashboard. Worked using ggplot uploaded to ArcGIS storymaps instead.	Figuring out a consistent format for graphs to be made in order for storymap sidecar displays to work	All

Date	Activity	Issues/Notes	Person
14/10/2022	Started using ArcMap to create choropleth maps of water quality indicators across New Zealand.	Working out which tools are relevant. Tried using intersect and spatial join to generate stats for mean and median. Found that there is no inbuilt tool to calculate median. Also found that means were not being calculated correctly due to numbers being character strings. Solved by changing the column type to dbl.	Henri
15/10/2022	Managed to work out a method to map water quality across NZ, using layer joining in the context menu.		Henri
15/10/2022	Created data frames to find the proportions of sites that exceeded the thresholds I found from government websites.		Alzen
16/10/2022	Finalised the maps and exported them as a shapefile to upload to ArcGIS storymaps.	Briefly lost track of file names and file locations	Henri
16/10/2022	Created a time-series line graph for an average population of Agriculture Livestock Activities throughout the years.	Understanding the relevant functions used from highcharter library for graphing	Moses and Alzen
16/10/2022	Completed the story map for the Agricurlture activities		Moses
16/10/2022	Completed the story map for the water		Alzen

	quality.		
17/10/2022	Found that our preparation for the group presentation was in the wrong direction. Redid our presentation prep by creating diagrams and creating a slideshow of our data wrangling process.	We had focused too much time on making end-user products for analysis rather than working on making a cohesive presentation geared towards the process of our data wrangling	All
19/10/2022	I completed the Jupyter notebook for the water quality.		Alzen
19/10/2022	I completed the Jupyter notebook for the Agriculture Activity		Moses
20/10/2022	I successfully merged the notebooks for agricultural activity and water quality using the nbmerge and addressed the issues after joining the two notebooks.		Alzen
21/10/2022	The team had a quick discussion about the other deliverables we have to submit and to ensure the git branches we are currently working on are up-to-date before merging all of the changes to the master.		All
21/10/2022	I combined the dataset documentation for agricultural activity and water quality to create a README file for our GitHub repository. I also made a relational data schema using dbdiagram (a relational		Alzen

	database diagram design tool online).		
23/10/2022	Worked with ipynb file, comment all codes about 'highcharter' and use links of GitHub png files to import to ipynb file.		HanByeol
24/10/2022	I have completed the sections of the project report assigned to me.		Alzen
25-26/10/2022	Worked on the project report. Completed the sections assigned to me.		Henri
26/10/2022	Added the link to our GitHub repository and comments to our Jupyter Notebook. Edited the dataset schema for agriculture in our presentation slides.		Alzen
26/10/2022	Finalised the group report		All