DATA 201

Group Project Diary

Patrick Kearney
William Pattison
Lucrezia Ma

Purpose of this diary :

> every day you work on the project, take a detailed note of who-did-what in a text file. Even when you work by yourself, take note of what you did and then report that into the group diary

Sep 19th

We were starting to search for a dataset in the government database website (https://data.govt.nz/) and we discussed some interesting topics, such as politics, economy and health.

Sep 26th

We had a short meeting before the lab to share ideas about the project. We were still thinking about the suitable topic to do wrangling.

The google doc was created for diary and report separately, and added each member to the group.

Oct 1st

Created a GitHub project so we can work together on the same document.

Oct 3rd

We came up with the idea of wrangling data sets on transport, for example, the amount of crashes in NZ and data around that. We aimed to study the trend of injuries and death during these 30 years of car crashes in NZ and comparing it with the global dataset. Also, we were interested in car crashes during the public holiday, such as Christmas and New Year, etc.

On the feedback session
I reported our dataset and possible questions that we could ask. Unfortunately, Thomas suggested our study questions (such as how the holiday and road type effect a car crash) seem too simple and obvious. He suggested we should go for something more complicated

and ask deeper questions for the project, such as how this project can help reduce the car crash. Otherwise we can't get good mark.

Oct 5th

Email Giulio and Thomas about the confusing part of project, such as the main goal of this project and how to present the results.

This project is focusing on data wrangling (obtaining the data, tidy format, data presentation) instead of analysis of data. There is no marking down due to using wrong statistical model.

A complex data-set is recommended because it has the potential to answer some complex questions where the answer is not currently in a tidy and obvious form.

Oct 9th

We change the project topic to study vehicle fleet in New Zealand from 2000 to 2015. The aim is to investigate the amount of vehicles and which factors may have an influence on that in New Zealand from 2000 to 2015. This data came from New Zealand transport government website (https://www.transport.govt.nz).

We started to do wrangling by joining tables and finding interesting trends and factors.

Plot of Vehicles (excluding 2 wheelers) Per Thousand People in New Zealand was created.

Oct 10th

We are trying to find some interseting things in the dataset:

- Looking for the trend in the amount of vehicles for the past 15 years (2000 - 2015)
  Vehicle per thousands people is steeply increasing by years from 2000 to 2007 but deceased by years from 2008 to 2011. It is increasing again after 2012.

- To investigate the consumption of the fuel during these years
  By plotting the trend of fuel consumption per person from 2001 to 2016. This may indicate the efficiency of fuel, fuel price, type of cars(SUV, hybrid cars), conditions of cars (engine size), and car sales.

- To looking for car safey in recent years
  car crashes by year

- To explore any relationship between lung disease and the amount of cars

  We are trying to find a dataset but there is no specific source about this. So far, the closest dataset we found is air pollution by vehicle. However, this research was published in 2004 by Ministry of Transport and dataset is not available.

Interesting find:
We have two plots, Figure 1, *Light_passagers(amount of cars) by Year,* and Figure 2, *NZ_liters(fuel consumption) by Year,* in the same data frame. They show a similar pattern. In this case, we are starting to find the relationship between them.

Oct 14th

We decided to seperate data into three parts.

- Looking for the trend in the amount of vehicles for the past 15 years (2000 - 2015)

  Comparing the trend of ownership rates with different types of vehicles

- To investigate the consumption of the fuel during these years
  - fuel usage
  - fuel price
  - enginze size
- To looking for car safey in recently years

  car crashes by year

Oct 15th

Finishing the coding and moving on to do the power point for the presentation.
We created a google slide document and assigned each person to work on one part of the topic.

Oct 19th

We are working on getting the report done for submission.