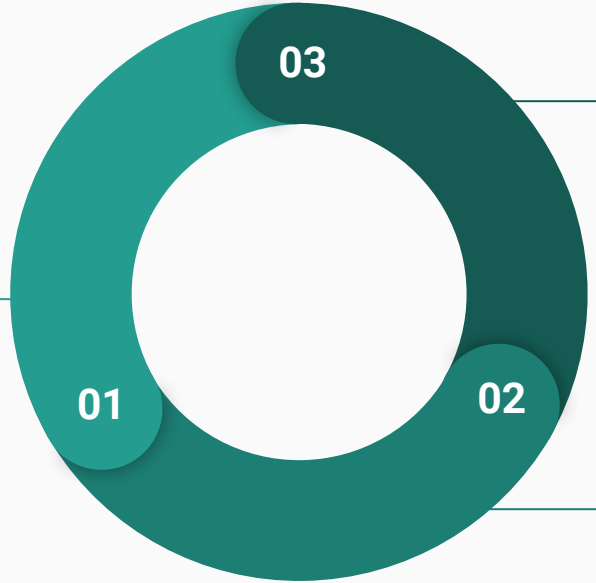
How does the severity of earthquakes affect migration of residents and their mental health?

BY DATA**M**AC

Starting our project/identifying data sources



Allocation of tasks and start

Tasks allocated, regular group meetings.

Fast prototyping

Generate ideas - Vocabulary Game

Group discussion

Deciding topic, datasets.

Cleaning the Population Dataset

```
#change_rate clean up
change_rate <- read.csv(file="change rate.csv", skip = 1)
change_rate <- change_rate %>% rename(year = X, Population_change = Population.change)
change_rate <- change_rate %>% slice(1:16)
change_rate
write.csv(change_rate, file = "change_rate_1.csv")
```

```
population <- population %>% mutate(year = c(2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016, 2017,2018
population_change_rate <- merge(population, change_rate)
population_change_rate <- population_change_rate %>% rename(total_population = Total)
write.csv(population_change_rate, file = "population_change_rate.csv")
```

```
egion_population <- read.csv(file="region population.csv", skip = 1)
egion_population <- region_population %>% rename(year = X)
egion_population <- region_population %>% slice(1:16)
select the city we want to Performance
egion_population2 <- region_population %>% select(year, Northland.Region, Auckland.Region, Wellington.Region, Cante
ename(Northland = Northland.Region, Auckland = Auckland.Region, Wellington = Wellington.Region, Canterbury = Canter
egion_population3 <- gather(region_population2, city, Total.Pop, Northland:Canterbury)%>%
                      mutate(Pop.change = if_else(city == lag(city), Total.Pop - lag(Total.Pop), Total.Pop-Total.P
                             Pop.change = na_if(Pop.change, 0),
                             city = if_else(grepl("Northland",city), 'Auckland', city)) %>%
                      group_by(year, city) %>% summarise(Total.Pop = sum(Total.Pop), Pop.change = sum(Pop.change))
rite.csv(region_population3, "FINAL_population_by_city.csv") # Save csv file for each city by year
```

Cleaning the Earthquake Dataset

Earthquake Dataset - Wrangling Function

```r
wrangle_earthquake_data <- function(dataset, city) {
    # step one - select only the desired columns
    filtered_data <- dataset %>% select(eventtype, origintime, longitude, latitude, magnitude, depth)

    # step two - filter only the eventtype == "earthquake"
    filtered_data <- filtered_data %>% filter(eventtype == "earthquake")

    # step three - filter out insiginificant earthquakes
    filtered_data <- filtered_data %>% filter(magnitude >= 4.0)

    # step four - add two columns - one with year and one with city
    filtered_data <- filtered_data %>% mutate(year = origintime %>% substr(1, 4))
    filtered_data <- filtered_data %>% mutate(city = city)

    # final step - return filtered data
    filtered_data <- filtered_data %>% select(city, year, origintime, longitude, latitude, magnitude, depth)
    return(filtered_data)
}
```

Cleaning the Mental Health Dataset

Cleaning mental health dataset - each year

```
path <- glue('Mental health data/{file}')
df <- read_excel(path, sheet = sheetnum, skip = 3) %>%
select(1:3)

names(df) <- c('DHB', 'sex', 'total') # Rename columns
yr <- file %>%substr(6, 9) # Extract year from filename

df <- df %>% mutate(DHB = if_else(is.na(DHB), lag(DHB), DHB), year = yr) %>%
        fill(DHB) %>%
        filter(grepl('Northland|Waitemata District|Waitematā|Auckland District|Manukau District
        mutate(DHB = if_else(grepl("Waitematā",DHB), 'Waitemata District Health Board', DHB))
```

Cleaning mental health dataset - binding

	DHB	sex	total	year
	<chr>	<chr>	<chr>	<chr>
	Northland District Health Board	Total	3306	2005
	Northland District Health Board	Male	1827	2005
	Northland District Health Board	Female	1479	2005
	Waitemata District Health Board	Total	15028	2005
	Waitemata District Health Board	Male	8437	2005
	Waitemata District Health Board	Female	6591	2005

Cleaning mental health dataset - grouping

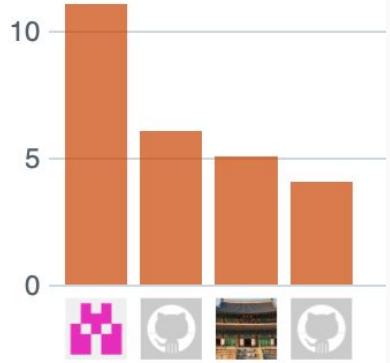
```
df<- df %>% mutate(DHB = if_else(grepl("Northland|Waitemata|Auckland|Manukau",DHB), 'Auckland', DHB)) %>%
      mutate(DHB = if_else(grepl("Capital|Wairarapa|Hutt",DHB), 'Wellington', DHB)) %>%
      mutate(DHB = if_else(grepl("Canterbury",DHB), 'Canterbury', DHB)) %>%
      mutate(total = total %>% as.integer) %>%
      group_by(DHB, year, sex) %>% summarise(total = sum(total))

colnames(df)[1] = 'city'

return(df)
```

Using git

Excluding merges, **4 authors** have pushed **26 commits** to main and **26 commits** to all branches. On main, **62 files** have changed and there have been **110,942 additions** and **1,020 deletions**.



Combining the three datasets

```
joined_table <- eq_summary_df %>% full_join(mh_df, by=c('city','year')) %>%
        full_join(pop_df, by=c('city','year')) %>%
        filter(sex=='Total') #%>% na.omit()
joined_table %>% head()
```

Creating an R package

```
install.packages("remotes")
remotes::install_github("chenthih/nzmentalquake")
```

Documentation

`?nzmentalquake::population_dataset`

Format

An object of class `data.frame` with 48 rows and 5 columns.

Details

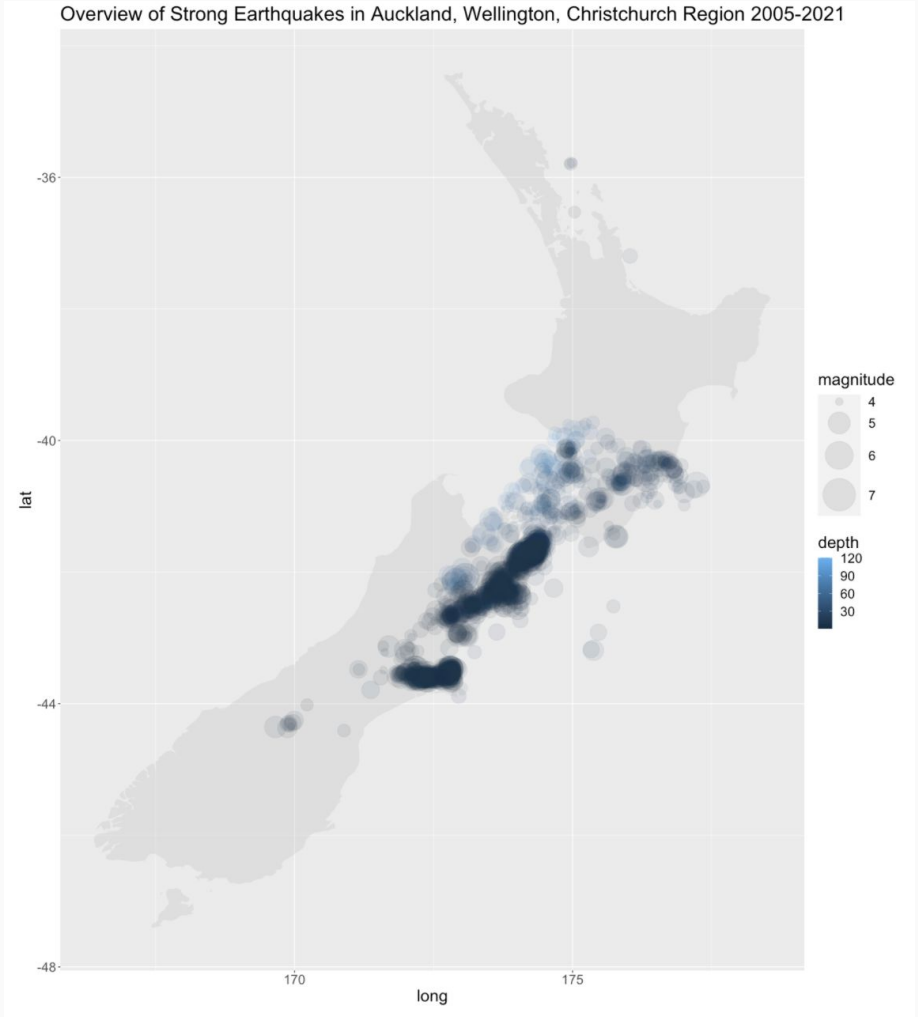
@format A tibble with 48 rows and 4 variables:

year	int year of the recorded total, ranging from 2005 to 2020
City	chr city of recorded population, either Auckland, Canterbury or Wellington
Total.Pop	int the total population of the city in the given year
Pop.change	int the difference in population from the previous year in the given city

Source

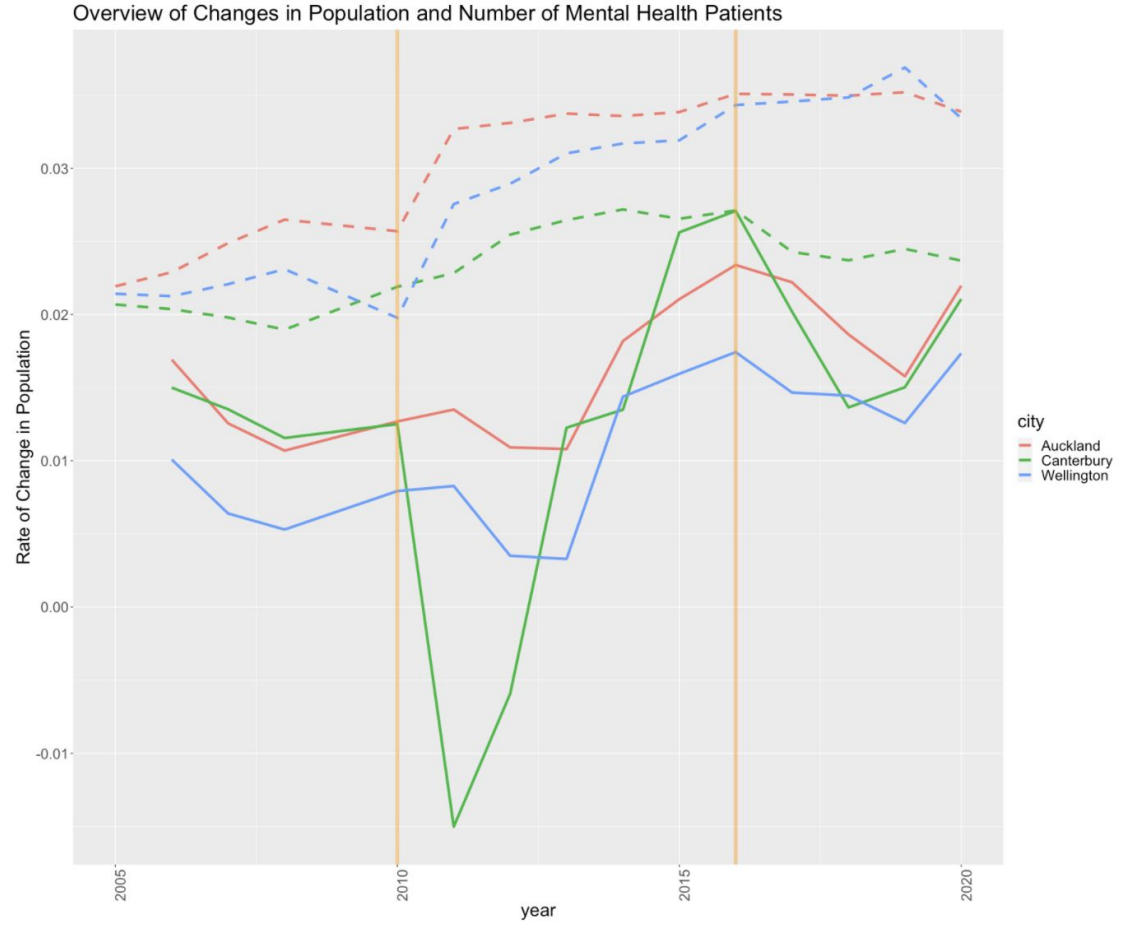
http://infoshare.stats.govt.nz/Default.aspx#

Plotting our data - Quake Map



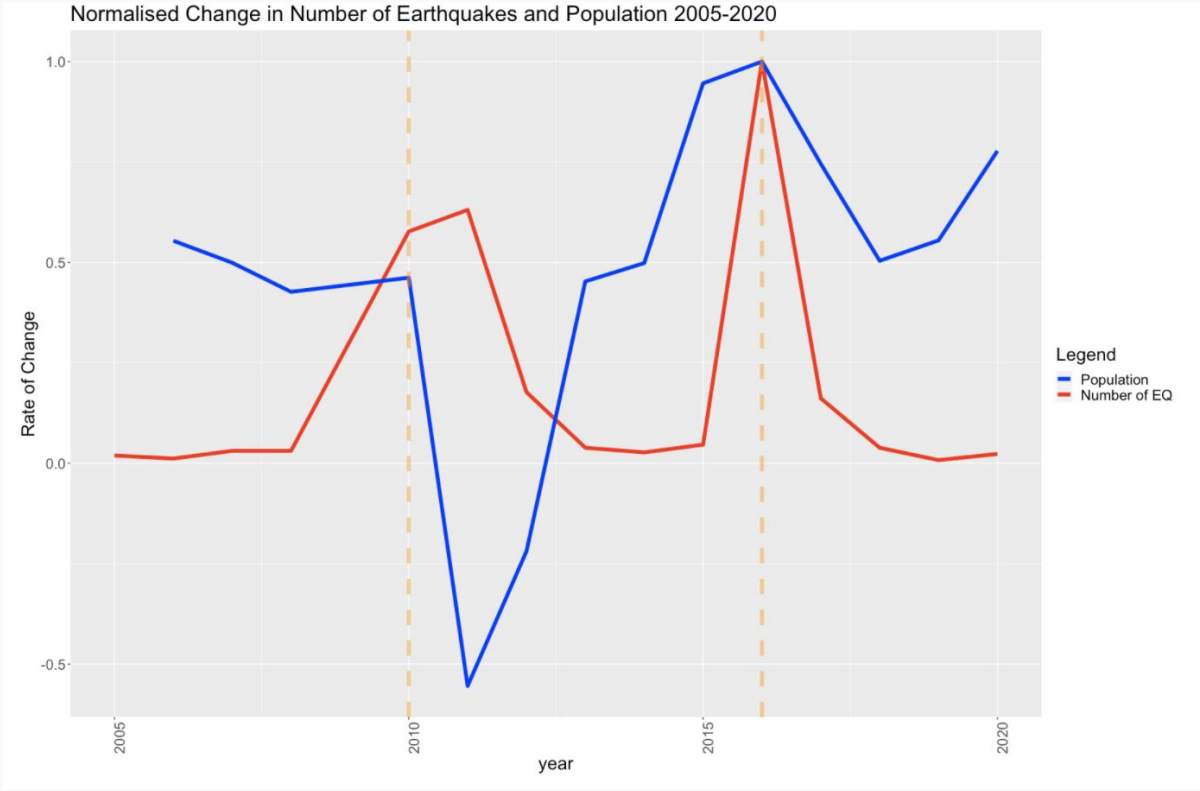
Plotting our data -

Overview of Population and Number of Mental Health Patients Change



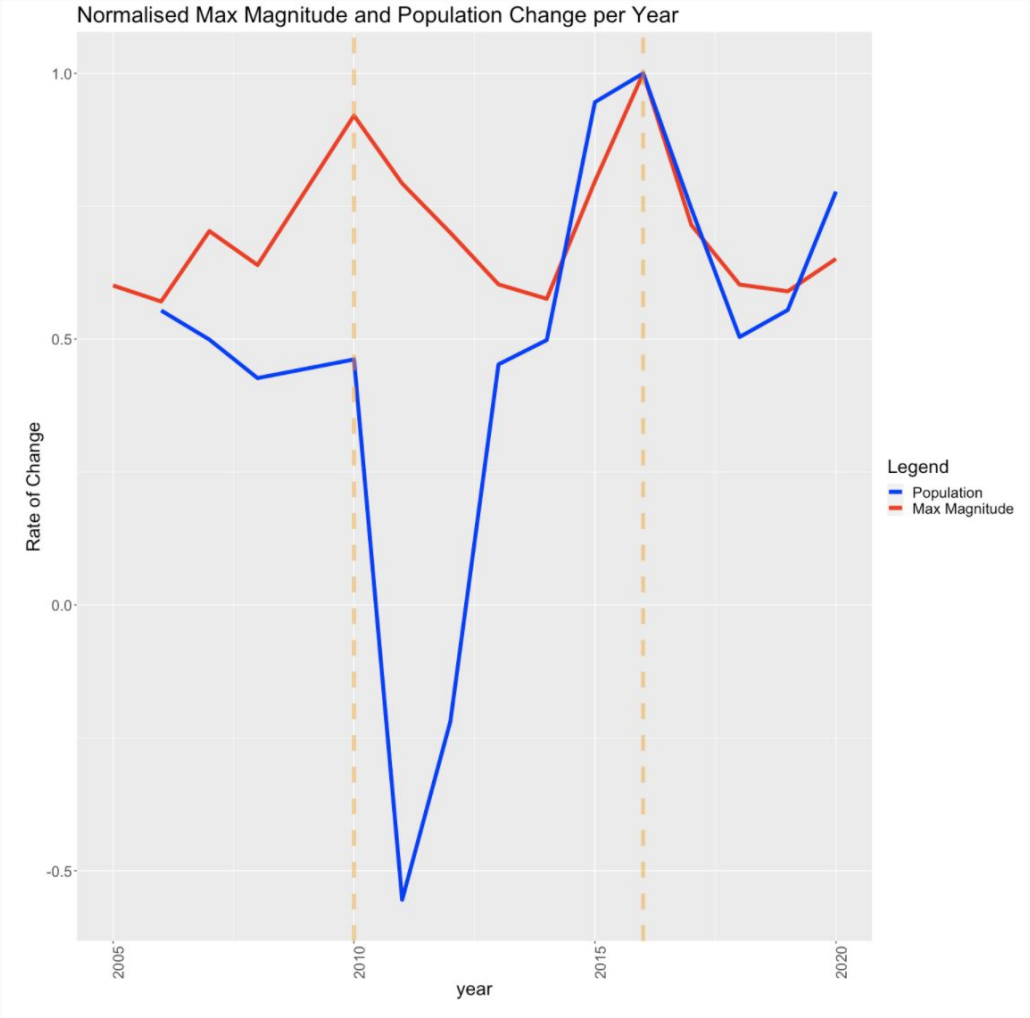
Plotting our data -

Number of Earthquakes / Population



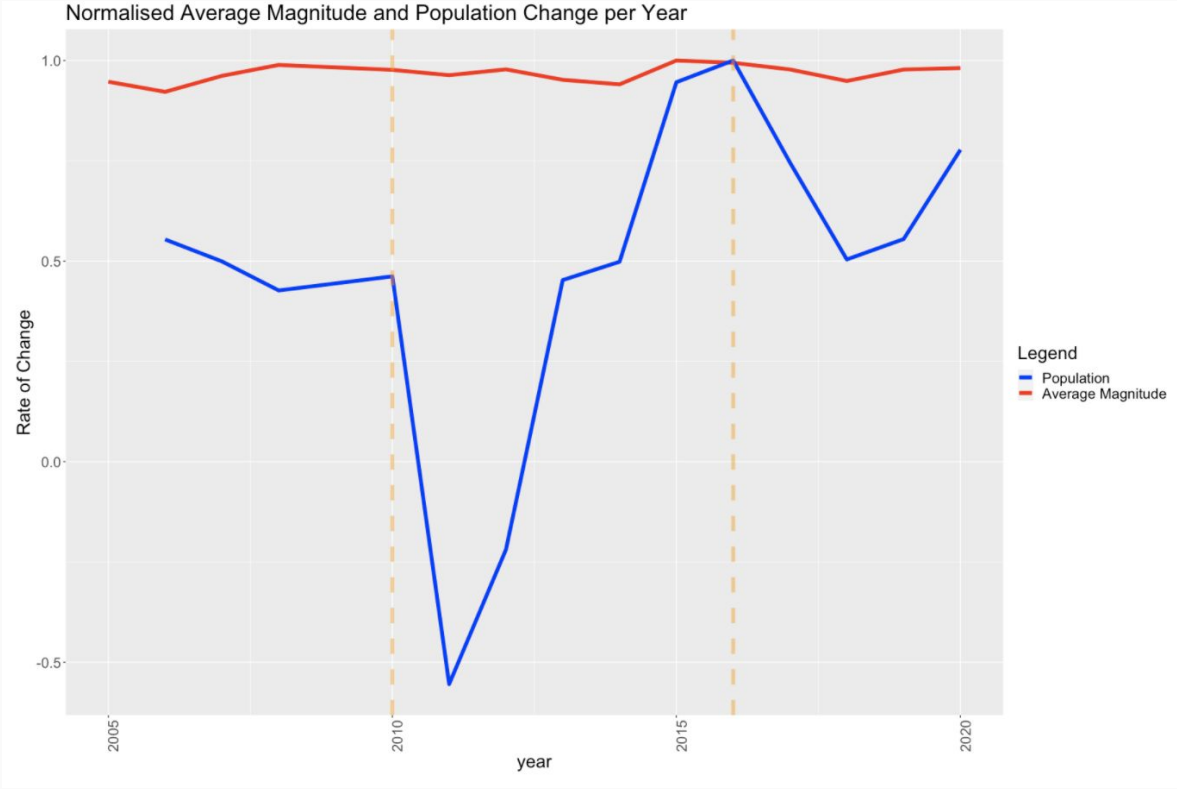
Plotting our data -

Max Magnitude/ Population



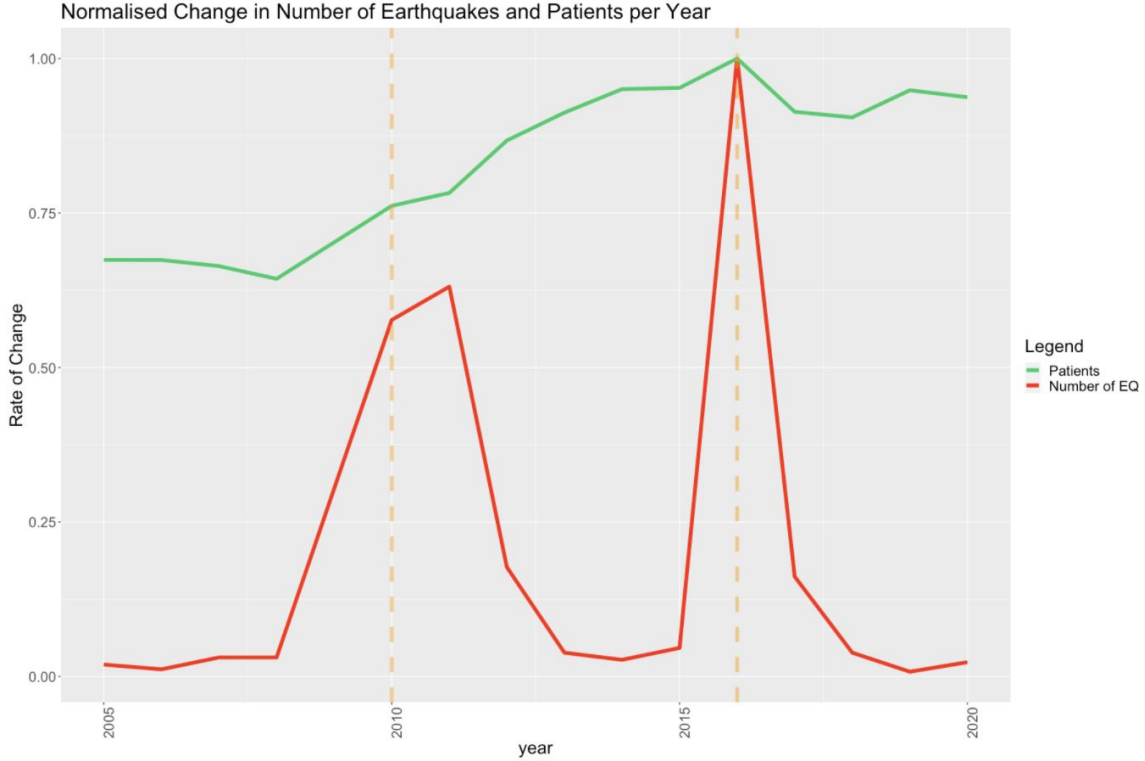
Plotting our data -

Average Magnitude/ Population



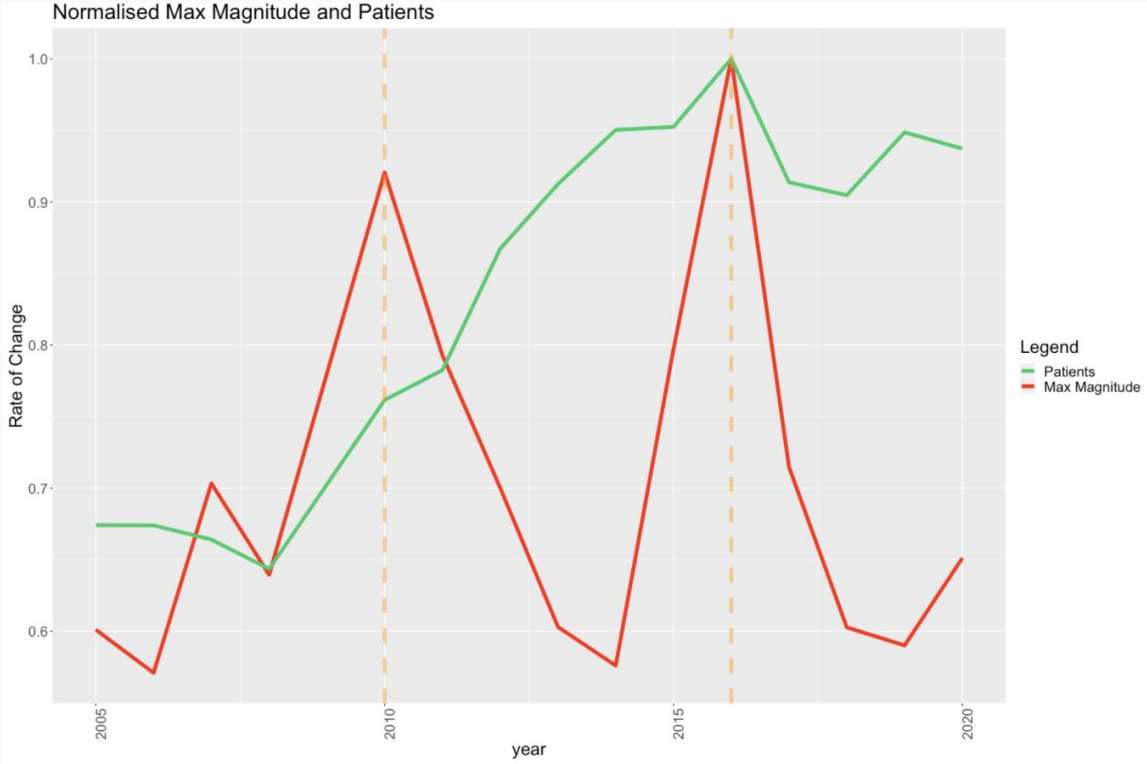
Plotting our data -

Number of Earthquakes / Mental Health



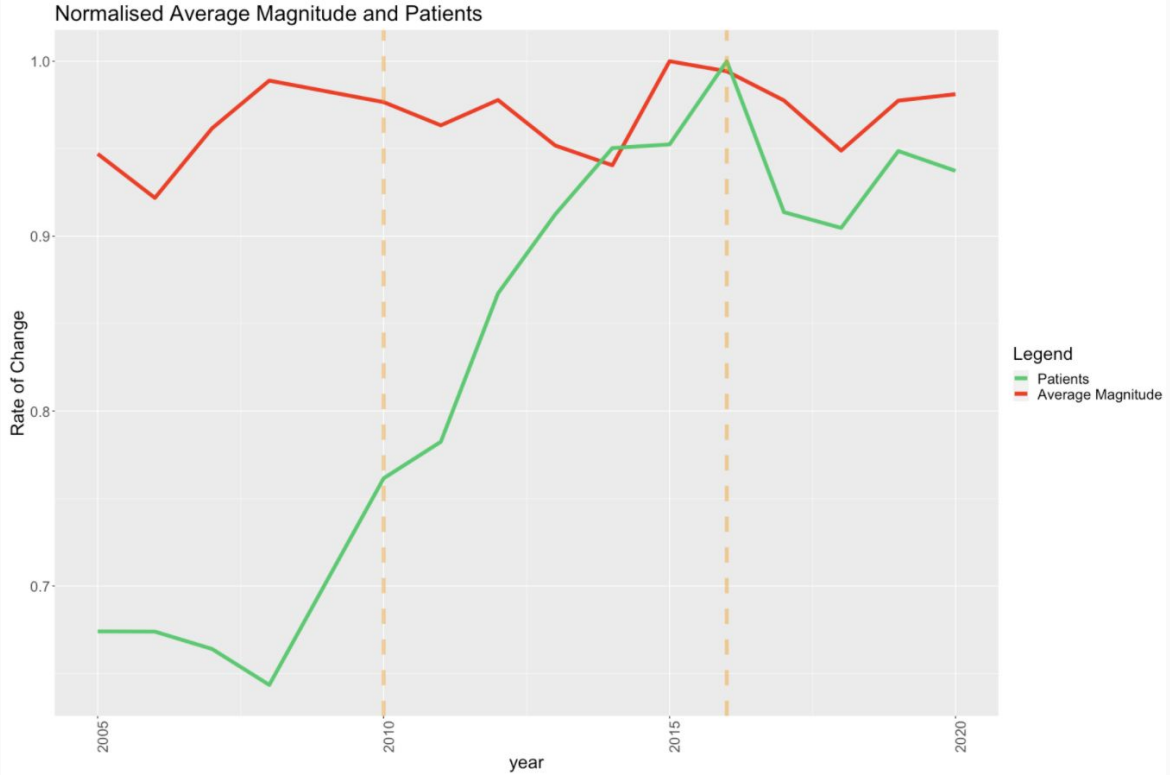
Plotting our data -

Max Magnitude/ Mental Health



Plotting our data -

Average Magnitude/ Mental Health



Conclusion

There are some clear trends in the different datasets we have wrangled.

Relationship between each is unclear