# Earthquakes Database in a Historical Perspective

*Project for Database and Analytics Programming course

Juan Pablo Andres Palma Bustos
*Master in Science Data Analytics (MSCDAD_B)*
*National College of Ireland*
Dublin, Ireland
x20225776@student.ncirl.ie

***Abstract*—The paper aims to explore how Earthquakes can affect the trajectory of the countries' economic development over the years. The database was built using the scrapping technique and programming language to process, clean, and analyse the data. In conclusion, the paper shows how something like natural disaster does not exist, instead of experiencing natural phenomena with bad human management of these conditions.**

***Index Terms*—Earthquake, Programming, Database, Web scraping, Python, R.**

## I. Introduction

This paper is about obtaining a historical database of Earthquakes in many countries of the world. Coming from one of the most seismic countries, the Earthquakes and their consequences inspire research about the social context, economic detriments, and how different countries can deal with complex natural conditions.

Using the programming language of Python for scraping the web and R for later analysis, the paper aims to contribute as an exploratory project for integrating historical information when we research natural events.

The database was obtained from the website Wikipedia. However, due to the imprecise nature of the information that we tried to get, the web page offers us available semi-structured data, free for use and retrieving and legally scraping the webpage.

The website declares that the dataset's natural events were selected by different criteria like the high magnitude of the Earthquake, the consequences after the event, and many others. Still, it does not follow a single standard. Therefore as a resume, it is not the most depured website, but it is an effort to consolidate all the information available in one place.

Finally, after all these considerations above, we will explore and respond to how the Earthquakes affect the economic performance of the countries in Indonesia, Mexico, and Chile.

## II. Related Work

The question about the economic consequences of the natural phenomenon as Earthquakes is not new in the literature. However, the evidence is not conclusive. Even when metanalysis studies [1] confirmed the negative and significant effect, there had a focus in developing countries.

Often the concept "natural disaster" is used for describing an Earthquake and another natural phenomenon. But, unfortunately, this concept drives us to a misconception about these events. It does not exist 'Natural Disaster', and often it corresponds to badly faced or foreseen natural phenomena [2]. We live on a constantly moving planet, and these releases of energy are part of the cycle and consequence of inhabiting the Earth.

This perspective changes the focus to human decisions and responsibility on these events. Social conditions are essential and often can make a difference between of affected population affected faced the recovery process. So there is when the research questions which consequence can be expected and nuanced after these natural events.

Research as Elham Zandian and others [3] will support the adverse long-term effects, specifically on human capital and wealth and the importance of considering the territorial context, culture, and country-level development.

Results like J. Paudel, H. Ryu [4] show that the consequences had a significative social class and genre bias by time perspective. It means Earthquakes (like any other Natural event) can affect lower class and Female people worst due to previous inequalities.

Furthermore, the recovery process can be misunderstood if the politicians and international organisations do not consider all the factors. For instance, the Haiti case [5] showed the limitation of the Humanitarian help if it does not consider the different rural and city or poverty levels of the affected zone.

On the other hand, previous researches have documented that Earthquakes are inevitable, sudden, often and variable [1]. However, a long-term perspective evaluates how countries react and learn for the future.

Historical research in Chile by documentary analysis

[6] focused on how the consequences in the associative relation of the population and other cultural and institutional demonstrations like (religion or government) also have to be considered in a comprehensive study. In addition, they found similarities in an earthquake in 1657 (on the least researched in the colonies) with the big one in 2010, which gives important insight into how seismic cycles can affect a country and its development.

Finally, metropolises and big cities are attractive for many services and make life easier. Still, the consequences of climate change and our systemic and interdependent lifestyle can have worst effects than before, as we have seen in recent events [1].

## III. Methodology

### A. Description of the dataset

The project consists of the integration and joined-use of two data sets. Each of them had a different challenge in extracting the data and later cleaning.

The Earthquake historical data set was a web scraping process of obtaining information from five articles on the Wikipedia websites and integrating 34 tabulated lists of Earthquakes from 1652 BC until 2021, with 98 countries and 1275 cases (See the top 20 in Table I).

TABLE I
Top 20 - Total cases by Country

| N | Country | Obs | N | Country | Obs |
|---|---|---|---|---|---|
| 1 | Indonesia | 130 | 11 | Philippines | 34 |
| 2 | China | 123 | 12 | Mexico | 33 |
| 3 | United States of America | 112 | 13 | New Zealand | 31 |
| 4 | Japan | 70 | 14 | Guinea | 28 |
| 5 | Turkey | 63 | 15 | Russia | 25 |
| 6 | Iran | 51 | 16 | Greece | 21 |
| 7 | India | 43 | 17 | Pakistan | 20 |
| 8 | Chile | 43 | 18 | Australia | 19 |
| 9 | Peru | 37 | 19 | Papua New Guinea | 19 |
| 10 | Italy | 35 | 20 | Afghanistan | 19 |

On the other hand, we have obtained data from the World Bank, which had an API (Application Programming Interface) for the unrestricted use of the users. This database is simply the 'Gross Domestic Product per capita in current US dollar' (GDP PCAP CD).

The GDP (PCAP CD) is valid for this comparative process because it divides by the population, which estimates the country's economy size proportional to the people. So the country's economy cannot be measured as an absolute value if we would like to compare them.

The objective is to use these databases to display visualization, which allows us to see the trajectory of the countries' economic growth and how it changes when a Natural Phenomenon as Earthquakes occurs.

### B. Technologies implemented

During the programming process of the paper, we have used the programming language Python to extract the information and later processing and cleaning. As a multiparadigm language, Python allows us to access the website, retrieve and wrangle information, connect with users services like API.

It also implies a challenge to put in action the contents learnt during the course and the new perspective about how and from where we can get information.

The technologies implemented are described as follows.

- Numpy [7] is a powerful package for dealing with multidimensional array objects and others such as matrices. It also includes numerical and mathematical operation, data manipulation, linear algebra, and many other mathematical expression and implementations.
- Pandas [8] is the package that allows work with DataFrame objects for data manipulation with integrated indexing. It was essential during the cleaning process and for the data format.
- Requests [9] is a library used for making HTTP requests in Python. We have used it to have access to the code structure of the website as text and get the information displayed.
- BeatifulSoup [10] is a Python library for pulling HTML and XML files data. It gives access to a specific section of the website parse tree and navigates through, searching and many other functionalities.
- Regular expressions [11] are a highly specialized programming language that permits finding, selecting, changing, and removing some specific piece of text or pattern. For instance, we have used it for cleaning strings and identifying words for building new variables from a part of text information.

### C. Data processing

We have chosen the Earthquake historical data set because we would like to build a data set from many information sources. Also, we would like to visualise the economic trajectory of the countries and Earthquake events in the same graphic. Moreover, it can be a capstone for future work to rebuild a historical frame of this natural phenomenon.

Many studies have worked with data from the 20th century, many others from the 19th century, but it is not easy to see a data set that integrates them. This effort can be extended

TABLE II
TABLE EXTRACTION DESCRIPTION

| Name period | Tables N |
|---|---|
| Pre-11th century | 1 |
| 11th–18th centuries | 1 |
| 19th century | 1 |
| 1901–1910 | 1 |
| 1911–1920 | 1 |
| 1921–1930 | 1 |
| 1931–1940 | 1 |
| 1941–1950 | 1 |
| 1951–1960 | 1 |
| 1961–1970 | 1 |
| 1971–1980 | 1 |
| 1981–1990 | 1 |
| 1991–2000 | 1 |
| 2001-2010 | 10 |
| 2011-2020 | 10 |
| 2021-2030 | 1 |
| **Total** | **34** |

and complement the data set with more information, but it remains for future work in the field.

*1) Data extraction:* We implemented a coding program for the Earthquakes historical data to get information in 5 pages in Wikipedia and give format to 34 different tables (See Table II).

The first step was to get the column names such as Date, Time, Place, Latitude, Longitude, Fatalities, Magnitude, Comments, Sources.

The next step was to get the exact information and give the place in a formatted table with the column names.

Finally, using the Pandas method of concatenation, we consolidate in one database the 34 tables created in the process.

The World Bank data set was more straightforward because the API is easy to use and extract information, so we selected the indicators we wanted to get. Then, it retrieved a data frame with the solicited information.

*2) Data cleaning:* The data cleaning process was divided into two activities. Firstly, we correct all imprecise or wrong information from the extracting process. Secondly, create new variables from the raw table to proper columns.

The cleaning process is an iterative development that is highly time-consuming. In total, 38 cases of the 1,275 presented problems. Most of these problems correspond to merged columns or rows in the original tables, which produce that the information retrieved in this case's column was incorrect.

To solve the problem, we identified five patterns that make the observation incorrect, and we implemented a code to correct each of them.

On the other hand, when we tried to create new variables like Country, we realised many other problems that had to be solved. The most difficult one was to clean the information that we had got from the column Place.

We explored different ways, but an efficient solution was creating the variable Country getting the last word from the variable Place and then fixing the cases where the solution does not retrieve a valid case.

These remained cases of Country we solved, developing two dictionaries that allowed us to look for information in other original columns and give a valid, correct observation.

Finally, we created new variables such as Year, Magnitude and Deaths, mainly using a similar process that combined regular expressions for solved the problems.

*3) Data analysis:* The Data Analysis activity was coded in R, using R Studio. Firstly, we have opened the Earthquakes Historical Database and displayed some information not restricted to a specific period.

This section used the characteristics of the data set and tried to display information about Earthquakes from another point of view.

Secondly, we developed a code that merges both databases with more programming to visualise some data using the advantages of two data sets.

Finally, we displayed a graph with double-Y axes connecting the historical GDP (per capita in the Current US dollar) and the Magnitude of Earthquakes for three Countries. A lower-middle-income like Indonesia, a middle-income economy like Mexico and a high-income economy like Chile.

It is important to note that this paper focused on building the database, especially the Earthquake historical database, and all the analyses had an exploratory focus.

## IV. RESULTS AND EVALUATION

One of the advantages of the database is that we can look into a more extensive range when we try to describe the seismic activity in the world. For example, Table III displays a "Top 10 of Earthquakes with a Magnitude over 9.0".

It displayed that the Earthquake of Valdivia in Chile is the highest in history. However, many other big Earthquakes before the 20th century. For example, Indonesia had a natural event in 1833, the United States of America in 1964 and 1585, Japan in 869 and Canada in 1700.

It is crucial to notice that this information can be imprecise, especially with older events. However, we have

TABLE III
TOP 10 EARTHQUAKES WITH MAGNITUDE OVER 9.0

| | Country | Year | Mag | Deaths |
|---|---|---|---|---|
| 1 | Chile | 1960 | 9.5 | 5,700 |
| 2 | United States of America | 1964 | 9.2 | 125 |
| 3 | Indonesia | 1833 | 9.2 | No information |
| 4 | Indonesia | 2004 | 9.1 | 227,898 |
| 5 | Japan | 2011 | 9.1 | 18184.0 |
| 6 | United States of America | 1585 | 9.2 | No information |
| 7 | Japan | 869 | 9.0 | 1000 |
| 8 | Canada | 1700 | 9.0 | No information |
| 9 | Chile | 1868 | 9.0 | 693 |
| 10 | Russia | 1952 | 9.0 | 2,336 |

met the project's purpose if we open a field to a historical perspective about these events.

Secondly, the consequences after Earthquake are material losses in the best case. Many families lost all their belongings and all their economic achievements.

But these events were fatal for many people. So often evaluate the consequences like a singular event, it is challenging to realise the magnitude. For this reason, in Table IV, it is possible to compare the average of deaths that countries had suffered in events related to Earthquakes, this table consider countries that have had more than 15 Earthquakes registered in the database.

TABLE IV
AVERAGE OF DEATHS BY COUNTRY

| N | Country | Mean | N | Country | Mean |
|---|---|---|---|---|---|
| 1 | China | 17,805 | 11 | Chile | 1,787 |
| 2 | Iran | 15,979 | 12 | Afghanistan | 541 |
| 3 | Italy | 8,811 | 13 | Mexico | 464 |
| 4 | Pakistan | 7,628 | 14 | Philippines | 207 |
| 5 | Turkey | 5,875 | 15 | Russia | 200 |
| 6 | Japan | 4,942 | 16 | Guinea | 105 |
| 7 | Greece | 4,760 | 17 | United States of America | 74 |
| 8 | Peru | 3,482 | 18 | New Zealand | 17 |
| 9 | Indonesia | 2,086 | 19 | Papua New Guinea | 7 |
| 10 | India | 1,992 | 20 | Australia | 3 |

Thirdly, a common belief is that when the Earthquake is more potent will be related to higher rates of fatalities. However, Table V suggests that the statistical relation between Magnitude and Deaths is not strong. The association is 0.12 (all data set) and 0.15 for the top 20 countries (with more than 15 events), respectively.

It opened a research question about the conditions for having some countries where the correlation between Magnitude and Fatalities is higher. As the literature suggests, it can be due to the social context and development level, as the quality of constructions and institutional strength.

### A. Combined Analysis

Exploring a combined use of the two databases is possible to Plot double-Y axes with GDP (per capita in Current USDollar) and Magnitude of the Earthquake during the years. //

TABLE V
CORRELATION BETWEEN MAGNITUDE OF THE EARTHQUAKE AND DEATHS

| N | Country | Correlation | N | Country | Correlation |
|---|---|---|---|---|---|
| 1 | Italy | 0.50 | 11 | Japan | 0.22 |
| 2 | Iran | 0.50 | 12 | Indonesia | 0.22 |
| 3 | Pakistan | 0.43 | 13 | Papua New Guinea | 0.20 |
| 4 | India | 0.38 | 14 | Peru | 0.19 |
| 5 | China | 0.37 | 15 | United States of America | 0.16 |
| 6 | Philippines | 0.32 | 16 | Turkey | 0.14 |
| 7 | Chile | 0.32 | 17 | Afghanistan | 0.10 |
| 8 | Russia | 0.28 | 18 | New Zealand | -0.07 |
| 9 | Mexico | 0.26 | 19 | Guinea | -0.08 |
| 10 | Greece | 0.23 | 20 | Australia | -0.17 |

The correlation for all the database is 0.12
The correlation for top 20 countries is 0.15

We have reduced the perspective of the period to 1960-2020 because of the data available in the World Bank. However, we explored some insights using three examples: lower-income, middle-income, and high-income countries like Indonesia, Mexico, and Chile. //

About Indonesia (see Figure 1), it is possible to see a flatter curve of development in GDP, and even when one of the most devastating Tsunamis and Earthquakes in 2004, it is not possible to determinate a detriment in the economic performance. //

However, looking closer, it is possible to suggest that around 2004, after the Tsunami, the economy rises for a period of 5 or 6 years. For instance, from 2004 to 2011, the economy of Indonesia increased twice their GPD, which is very significant. //

In general, a different tendency is seen in Mexico (See Figure 2), viewing the country's economy as developing well during the second half of the 20th century. Still, some detriments in the performance are clear and visible when Earthquakes occur.

The first one is noticeable in 1980 when the economy grew five years. Later, it decreased almost in the same amount, and another slight reduction closer to 1985 when another Earthquake occurred.

The same tendency is possible in the middle of the '90s and all the period. Therefore, it is possible to propose the hypothesis that for some reason, Mexico seems to suffer more the consequences of these natural events, and always when an earthquake occurs is possible to see a detriment in their economic performance.

Finally, the case of Chile is a noticeable difference. The economic performance is clearly with a tendency of increase, especially from the end of the dictatorships in the early 90s.

However, a detriment at the end of the 90s could drive the reader to a wrong conclusion because it was mainly due to the Asian financial crisis in 1997, whose consequences significantly affected Chile.

Furthermore, it is crucial to highlight that in 2010 the country experienced an earthquake of a magnitude of 8.8. Even when it is possible to see that effect in the curve, the economic recovery was substantial.

## V. Conclusion and Future Work

This exploratory analysis's first and obvious conclusion is that only economic development does not mean better performance managing the consequences of a natural event.

Some factors make Mexico more dependent and have affection any time an earthquake occurs. The literature suggests the Natural disaster concept drives us to misconception because the social order, the institutions, quality of the policy, politician, culture and economic performance will affect how well or bad a country can deal with the consequences of these events.

We can not predict Earthquakes, but we can learn more about what we were already experienced, and that is why this effort for a project that tried to rebuild and consolidate information is required.

As a challenge for future work, firstly, it is important to work harder to consolidate a database with better information, maybe integrate description available on the web every single event that this database was able to get, or other options has to be explored.

Secondly, even when we have remarked on the importance of the time perspective, it will also be essential to develop research that can consider the local context of each event. For example, an Earthquake will affect mainly a local place, where the epicentre is located. Also, it is necessary to apply a more specific focus on local consequences. Maybe we should drop the category of "country" when analysing and researching with a territorial focus.

## References


[1] F. Gonzale, Natural Disasters and Economic Growth: a synthesis of empirical evidence in Revista de Ciencias Sociales Volume 31, Number 61, enero-junio 2022. ISSN-P: 0188-9834 ISSN-E: 2395-8669. DOI: http://dx.doi.org/10.20983/noesis.2022.1.8.

[2] M. Vega-Centeno, "Los terremotos, el crecimiento económico y el desarrollo" (Earthquakes, economic growth and develeppment0. Economía Vol. XXXIV, N° 67, semestre enero-junio 2011, pp. 57-80 / ISSN 0254-4415 Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[3] Zandian E, Rimaz S, Holakouie Naieni K, Nedjat S, Naderimagham S, Larijani B, Farzadfar F. Economic Effects of 1978 Tabas Earthquake (Iran). Arch Iran Med. 2016; 19(6): 409 – 413.

[4] J. Paudel, H. Ryu, Natural disasters and human capital: The case of Nepal's earthquake, Elsevier, World Development 111 (2018) 1–12.

[5] E. Ferris, S. Ferro-Ribeiro, PDisasters, 2012, 36(S1).2012 Published by Blackwell Publishing, 9600 Garsington Road, Oxford, OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

[6] D. Steward, "EL TERREMOTO DE 1657 EN CONCEPCIÓN, CHILE. UN ANÁLOGO COLONIAL DEL TERREMOTO EN MAULE 2010" (The Earthquake of 2657 in Concepcion, Chile. An similar colonial earthquake in Maule 2010) CUADERNOS DE HISTORIA 55 DEPARTAMENTO DE CIENCIAS HISTÓRICAS UNIVERSIDAD DE CHILE - DICIEMBRE 2021: 191-221.

[7] Numpy Documentation: numpy.org/devdocs/

[8] Pandas Documentation: andas.pydata.org/docs/user_guide/index.html

[9] Requests Documentation: pypi.org/project/requests/

[10] Beautiful Soup Documentation: www.crummy.com/software/BeautifulSoup/bs4/doc/

[11] Regular Expressions docs.python.org/3/library/re.html

Fig. 1. GDP Per Capita and Magnitude of Earthquakes in Indonesia

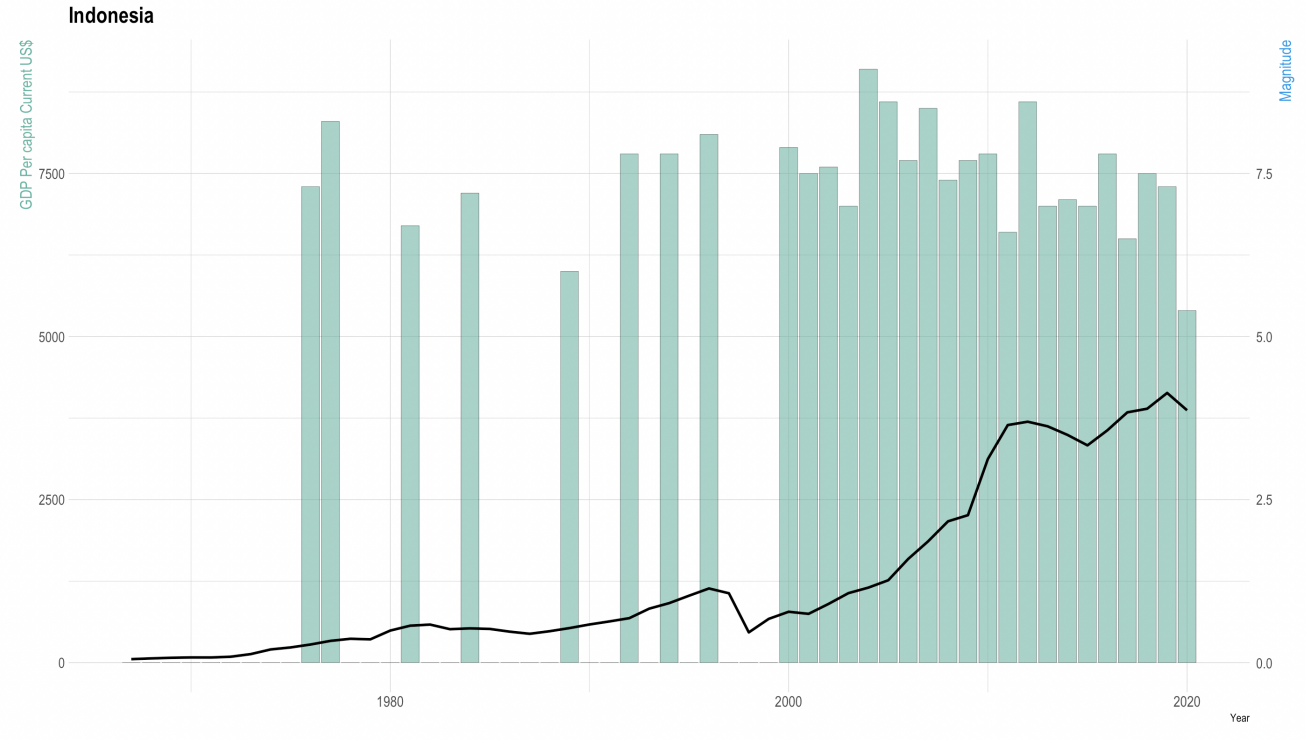


Fig. 2. GDP Per Capita and Magnitude of Earthquakes in Mexico

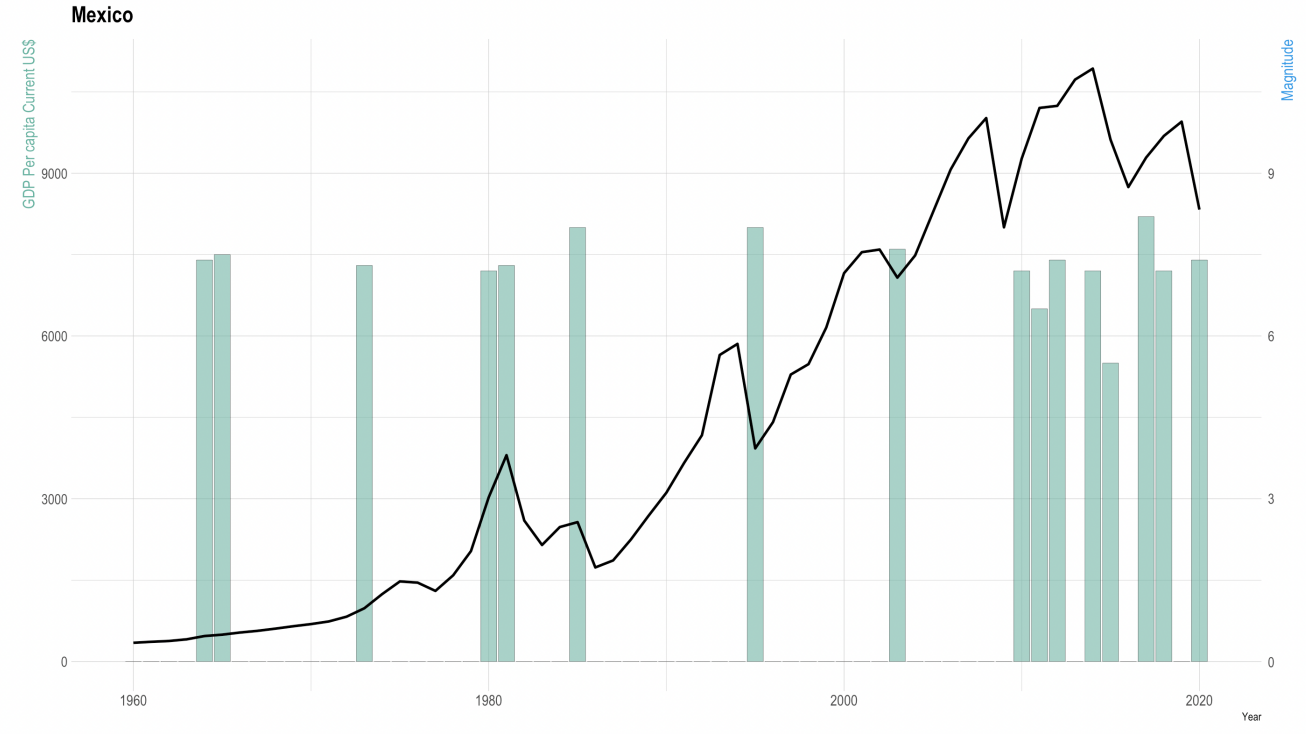

Fig. 3. GDP Per Capita and Magnitude of Earthquakes in Chile