4th October 2021

CANTERBURY Earthquake DATASET datasheet

# motivation

## For what purpose was the dataset created? Canterbury Earthquake was created to provide earthquake information (magnitude, depth, time etc.) to study the problem statement: How does the severity of earthquakes in Canterbury, New Zealand affect migration of residents and their mental health. The dataset was created for the specific task to be combined with two other datasets namely Canterbury Population and Canterbury Mental Health Consultation Count for the purpose of the study. The initial dataset is obtained from Geonet, an organisation that runs on the purpose of providing public access to hazards information, including earthquake reports and Volcanic Alert Bulletins. It also allows the retrieval of fundamental data sets, such as GPS Rinex files, earthquake hypocentres and instrument waveform data. These data are made freely available to the research community.

## Who created this dataset? The initial data was extracted from Geonet using the Quakesearch tool on [*https://quakesearch.geonet.org.nz/*](https://quakesearch.geonet.org.nz/) . There are no specific names of individuals credited for the data. Geonet is the result of a partnership between the Earthquake Commission (EQC), GNS Science, and Land Information New Zealand (LINZ). Using the dataset provided publicly by Geonet, the DATAMAC team consisting of Chenthi Heer, Maple (Jeong Im) Lee, Zhenyuan He and Ada (Rabbiatul Adawiyah) Mohd Izhair of the DATAMAC group have cleaned up the original Geonet dataset for specific use of the DATA201 project for the University of Canterbury.

## Who funded the creation of this dataset? This dataset did not require any direct funding to the DATAMAC team as all data extracted from Geonet was data open for public access for free without any costs. Geonet is funded by several sponsors listed below: i. The Earthquake Commission (EQC) ii. Land Information New Zealand (LINZ) iii. Ministry of Business, Innovation and Employment (MBIE) iv. National Emergency Management Agency (NEMA)

# Composition

## What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Each instance is information on each occurring earthquake within the Canterbury region, consisting of 7 columns: *eventtype, origintime, longitude, latitude, magnitude, depth and severity*.

## How many instances are there in total? There are 24,215 instances within the dataset, dating from the 1st of January 2005 until 30th September 2021.

## Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable). The dataset contains all possible instances as it is a record of every occurring earthquake within the Canterbury, Auckland, and Wellington region. However, there may be missing values that are not known to us, and this would be difficult to verify with the resources available to the team.

## What data does each instance consist of? “Raw” data (e.g., unprocessed text or images)or features? In either case, please provide a description. Each instance is of raw text: *eventtype, origintime, longitude, latitude, magnitude, depth and severity*.

## Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text. This dataset contains no missing or NULL values from individual instances of earthquake data.

## Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)? If so, please describe how these relationships are made explicit. As the instances in the dataset are of every occurring earthquake in the Canterbury region, we assume that there is a relationship between each individual instance. Research indicates that earthquakes can systematically trigger other ones, with bigger magnitude earthquakes being more likely to do so (O’Malley et. al., 2018). Instances in this dataset may also consist of 'aftershocks' rather than principal earthquakes. Aftershocks result from sudden change in stress occurring between rocks and the previous release of stress brought on by the principal earthquake. Therefore there is a clear possible relationship between each instance in the dataset, but it may also not be the sole necessary relationship between each of the instances.

# collection process

## How was the data associated with each instance acquired? The data is directly observed using technology and machinery specific for earthquake detection.

## What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated? Sensitive monitoring devices called seismometers are used to measure the magnitude and location of earthquakes. The New Zealand Seismometer Network is made up of over 50 main stations spread throughout the country. Sensing equipment is used alongside computer programmes which can record and analyse data. Geonet continuously collects data using a Global Navigation Satellite System (GNSS). GNSS receivers determine very precisely positions on the surface of the earth. Using information on positions of the surface, Geonet is able to track deformation. This system monitors the slow movement of tectonic plates over time to determine the strain point and if the direction of movement in certain areas is changing.

## Who was involved in the data collection process (e.g., students, crowdworkers, contractors). The data collection process mainly involves geohazards analysts who work at the National Geohazard Monitoring Centre.

## Over what timeframe was the data collected? From the 1st of January 2005 until the 30th of September 2021.

## Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation. From the DATAMAC team, no ethical issues were found as these dataset values are of earthquake information generated from machines, obtained by an official organisation: Geonet.

# USES

## Has the dataset been used for any tasks already? If so, please provide a description. Aligned with the purpose of this dataset being created as stated in section A, this dataset has been used to investigate the relationship of earthquake severity with number of mental health consultations and migration of residents in the relevant region. The dataset is included in an R package that can be installed by public through the "remotes" package and the github repository "chenthih".

## Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point. Yes, on github. [*https://github.com/adaizh/datamac*](https://github.com/adaizh/datamac)

## What (other) tasks could the dataset be used for? The dataset can be used for any task that involves the investigation of earthquakes and the different earthquake attributes/details that may be significant to one's study.

# Distribution

## Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description. No, the dataset will not be distributed to third parties outside the entity.

## How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)? The dataset does not have a digital object identifier, but is accessible via github under the DATAMAC's group project repository "adaizh/datamac".

## When will the dataset be distributed? The dataset is available on the github repository as of this writing

**REFERENCES**

## Oregon State University. (2018, August 2). Earthquakes can systematically trigger other ones on opposite side of Earth. ScienceDaily. Retrieved October 13, 2021 from www.sciencedaily.com/releases/2018/08/180802102352.htm

## O’Malley, R., Mondal, D., Goldfinger, C., & Behrenfeld, M. (2018). Evidence of systematic triggering at teleseismic distances following large earthquakes. Scientific Reports, 8(1). doi: 10.1038/s41598-018-30019-2