

MC Erdbeben

Ausgangslage

Für eine Versicherung soll ein News-Feed erstellt werden, welcher kontinuierlich die Erdbebendaten von der Internetseite herunter lädt, über die Pipeline aufbereitet und in einem Dashboard anzeigt.

Die Daten werden von der Website <https://earthquake.usgs.gov/earthquakes/feed/v1.0/geojson.php> geladen

```
# function to check if package is present
install_if_not_present <- function(pkg_name){
  if(!requireNamespace(pkg_name, quietly = TRUE)){
    install.packages(pkg_name)
  }
}

install_if_not_present("tidyverse")
install_if_not_present("geojsonio")

## Registered S3 method overwritten by 'geojsonsf':
##   method      from
##   print.geojson geojson

install_if_not_present("geojsonR")
install_if_not_present("sf")
install_if_not_present("tidyjson")
install_if_not_present("jsonlite")
install_if_not_present("lubridate")
install_if_not_present("ggplot2")
install_if_not_present("leaflet")

library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.4      v tibble     3.2.1
## v lubridate  1.9.3      v tidyr      1.3.0
## v purrr      1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(geojsonio)

##
## Attache Paket: 'geojsonio'
##
## Das folgende Objekt ist maskiert 'package:base':
```

```
##
##      pretty
library(geojsonR)
library(sf)

## Linking to GEOS 3.11.2, GDAL 3.6.2, PROJ 9.2.0; sf_use_s2() is TRUE
library(tidyjson)

##
## Attache Paket: 'tidyjson'
##
## Das folgende Objekt ist maskiert 'package:stats':
##
##      filter
library(jsonlite)

##
## Attache Paket: 'jsonlite'
##
## Das folgende Objekt ist maskiert 'package:tidyjson':
##
##      read_json
##
## Das folgende Objekt ist maskiert 'package:purrr':
##
##      flatten
library(lubridate)
library(ggplot2)
library(leaflet)
```

Data Frames enthält. Die verschiedenen Levels können mittels `$`-Zeichen angesprochen und so auch extrahiert werden.

Daten aufbereiten

Mit dem JSON Objekt werden Metadaten mitgeliefert, die in die Variable “metadata” gespeichert werden.

```
metadata <- past_hours_js$metadata
metadata

## $generated
## [1] 1.698584e+12
##
## $url
## [1] "https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_hour.geojson"
##
## $title
## [1] "USGS All Earthquakes, Past Hour"
##
## $status
## [1] 200
##
## $api
## [1] "1.10.3"
##
## $count
## [1] 7
```

Beim Eintrag “generated” sieht man, dass das Format nicht dem eines Datums entspricht. Dies muss entsprechend angepasst werden, was zu einem späteren Zeitpunkt erfolgt.

Das JSON-Objekt enthält Features als Data Frame, welche wiederum Data Frames enthält.

```
features <- past_hours_js$features
features

##      type properties.mag      properties.place properties.time
## 1 Feature          0.74    10 km WNW of The Geysers, CA  1.698584e+12
## 2 Feature          1.23  18 km NNW of Ocotillo Wells, CA  1.698583e+12
## 3 Feature          5.00   Kermadec Islands, New Zealand  1.698582e+12
## 4 Feature          1.01     8 km NW of The Geysers, CA  1.698582e+12
## 5 Feature          2.50    30 km ESE of Mapleton, Utah  1.698582e+12
## 6 Feature          2.60    66 km SE of Kaktovik, Alaska  1.698582e+12
## 7 Feature          1.31     2 km NNE of The Geysers, CA  1.698581e+12
##      properties.updated properties.tz
## 1      1.698584e+12      NA
## 2      1.698583e+12      NA
## 3      1.698584e+12      NA
## 4      1.698583e+12      NA
## 5      1.698583e+12      NA
## 6      1.698583e+12      NA
## 7      1.698583e+12      NA
##
##      properties.url
## 1 https://earthquake.usgs.gov/earthquakes/eventpage/nc73953270
## 2 https://earthquake.usgs.gov/earthquakes/eventpage/ci40590680
## 3 https://earthquake.usgs.gov/earthquakes/eventpage/us700017d1
## 4 https://earthquake.usgs.gov/earthquakes/eventpage/nc73953260
```

```

## 5 https://earthquake.usgs.gov/earthquakes/eventpage/us700017dh
## 6 https://earthquake.usgs.gov/earthquakes/eventpage/us700017di
## 7 https://earthquake.usgs.gov/earthquakes/eventpage/nc73953250
##                                     properties.detail
## 1 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/nc73953270.geojson
## 2 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/ci40590680.geojson
## 3 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017dl.geojson
## 4 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/nc73953260.geojson
## 5 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017dh.geojson
## 6 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017di.geojson
## 7 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/nc73953250.geojson
##  properties.felt properties.cdi properties.mmi properties.alert
## 1          NA          NA          NA          NA
## 2          NA          NA          NA          NA
## 3          NA          NA          NA          NA
## 4           1           0          NA          NA
## 5          NA          NA          NA          NA
## 6          NA          NA          NA          NA
## 7          NA          NA          NA          NA
##  properties.status properties.tsunami properties.sig properties.net
## 1      automatic              0              8          nc
## 2      automatic              0             23          ci
## 3      reviewed              0            385          us
## 4      automatic              0             16          nc
## 5      reviewed              0             96          us
## 6      reviewed              0            104          us
## 7      automatic              0             26          nc
##  properties.code properties.ids properties.sources
## 1      73953270      ,nc73953270,              ,nc,
## 2      40590680      ,ci40590680,              ,ci,
## 3      700017dl      ,us700017dl,              ,us,
## 4      73953260      ,nc73953260,              ,nc,
## 5      700017dh      ,us700017dh,              ,us,
## 6      700017di      ,us700017di,              ,us,
## 7      73953250      ,nc73953250,              ,nc,
##                                     properties.types properties.nst
## 1                                     ,nearby-cities,origin,phase-data,          7
## 2      ,nearby-cities,origin,phase-data,scitech-link,          8
## 3                                     ,origin,phase-data,          42
## 4      ,dyfi,nearby-cities,origin,phase-data,scitech-link,          10
## 5                                     ,origin,phase-data,          41
## 6                                     ,origin,phase-data,          32
## 7      ,nearby-cities,origin,phase-data,scitech-link,          16
##  properties.dmin properties.rms properties.gap properties.magType
## 1      0.03371      0.02      110          md
## 2      0.39050      0.18      100          ml
## 3      1.42000      0.83       79          mb
## 4      0.01523      0.01       79          md
## 5      0.27300      0.61       45          ml
## 6      0.46900      0.57      189          ml
## 7      0.01251      0.03       90          md
##  properties.type      properties.title geometry.type
## 1      earthquake      M 0.7 - 10 km WNW of The Geysers, CA      Point
## 2      earthquake      M 1.2 - 18 km NNW of Ocotillo Wells, CA      Point

```

```
## 3      earthquake      M 5.0 - Kermadec Islands, New Zealand      Point
## 4      earthquake      M 1.0 - 8 km NW of The Geysers, CA      Point
## 5      earthquake      M 2.5 - 30 km ESE of Mapleton, Utah      Point
## 6      earthquake      M 2.6 - 66 km SE of Kaktovik, Alaska      Point
## 7      earthquake      M 1.3 - 2 km NNE of The Geysers, CA      Point
##          geometry.coordinates      id
## 1 -122.85400, 38.82367, 2.41000 nc73953270
## 2 -116.18067, 33.29767, 0.42000 ci40590680
## 3  -178.1653, -30.6704, 35.0000 us700017dl
## 4 -122.81834, 38.83133, 1.60000 nc73953260
## 5   -111.2791, 39.9884, 9.9380 us700017dh
## 6   -142.3912, 69.7115, 6.1180 us700017di
## 7   -122.7475, 38.7925, 1.0100 nc73953250
```

Das Data Frame geometry enthält die Koordinaten. Diese werden in einzelne Spalten aufgeteilt. Die Werte werden in einer Matrix gespeichert.

```
geometry <- features$geometry
coordinates <- geometry$coordinates
coordinates <- do.call(rbind, coordinates)
colnames(coordinates) <- c("lon", "lat", "depth")
```

Daten zu einem Dataframe zusammenführen

Nun kann das weitere Data Frame properties mit den Koordinaten-Matrix verbunden und in einem Dataframe für die Exploration gespeichert werden.

```
properties_df <- cbind(features$properties, coordinates)
head(properties_df)
```

```
##      mag      place      time      updated tz
## 1 0.74    10 km WNW of The Geysers, CA 1.698584e+12 1.698584e+12 NA
## 2 1.23   18 km NNW of Ocotillo Wells, CA 1.698583e+12 1.698583e+12 NA
## 3 5.00    Kermadec Islands, New Zealand 1.698582e+12 1.698584e+12 NA
## 4 1.01      8 km NW of The Geysers, CA 1.698582e+12 1.698583e+12 NA
## 5 2.50     30 km ESE of Mapleton, Utah 1.698582e+12 1.698583e+12 NA
## 6 2.60     66 km SE of Kaktovik, Alaska 1.698582e+12 1.698583e+12 NA
##                                     url
## 1 https://earthquake.usgs.gov/earthquakes/eventpage/nc73953270
## 2 https://earthquake.usgs.gov/earthquakes/eventpage/ci40590680
## 3 https://earthquake.usgs.gov/earthquakes/eventpage/us700017dl
## 4 https://earthquake.usgs.gov/earthquakes/eventpage/nc73953260
## 5 https://earthquake.usgs.gov/earthquakes/eventpage/us700017dh
## 6 https://earthquake.usgs.gov/earthquakes/eventpage/us700017di
##                                     detail
## 1 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/nc73953270.geojson
## 2 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/ci40590680.geojson
## 3 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017dl.geojson
## 4 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/nc73953260.geojson
## 5 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017dh.geojson
## 6 https://earthquake.usgs.gov/earthquakes/feed/v1.0/detail/us700017di.geojson
##      felt cdi mmi alert      status tsunami sig net      code      ids sources
## 1    NA  NA  NA    NA automatic      0   8  nc 73953270 ,nc73953270,      ,nc,
## 2    NA  NA  NA    NA automatic      0  23  ci 40590680 ,ci40590680,      ,ci,
## 3    NA  NA  NA    NA reviewed      0 385  us 700017dl ,us700017dl,      ,us,
## 4     1   0  NA    NA automatic      0  16  nc 73953260 ,nc73953260,      ,nc,
```

```

## 5  NA NA NA  NA reviewed      0 96 us 700017dh ,us700017dh,      ,us,
## 6  NA NA NA  NA reviewed      0 104 us 700017di ,us700017di,      ,us,
##                                     types nst      dmin rms gap
## 1                                     ,nearby-cities,origin,phase-data,  7 0.03371 0.02 110
## 2                                     ,nearby-cities,origin,phase-data,scitech-link,  8 0.39050 0.18 100
## 3                                     ,origin,phase-data, 42 1.42000 0.83 79
## 4 ,dyfi,nearby-cities,origin,phase-data,scitech-link, 10 0.01523 0.01 79
## 5                                     ,origin,phase-data, 41 0.27300 0.61 45
## 6                                     ,origin,phase-data, 32 0.46900 0.57 189
##  magType      type                                     title      lon
## 1      md earthquake      M 0.7 - 10 km WNW of The Geysers, CA -122.8540
## 2      ml earthquake      M 1.2 - 18 km NNW of Ocotillo Wells, CA -116.1807
## 3      mb earthquake      M 5.0 - Kermadec Islands, New Zealand -178.1653
## 4      md earthquake      M 1.0 - 8 km NW of The Geysers, CA -122.8183
## 5      ml earthquake      M 2.5 - 30 km ESE of Mapleton, Utah -111.2791
## 6      ml earthquake      M 2.6 - 66 km SE of Kaktovik, Alaska -142.3912
##      lat depth
## 1 38.82367 2.410
## 2 33.29767 0.420
## 3 -30.67040 35.000
## 4 38.83133 1.600
## 5 39.98840 9.938
## 6 69.71150 6.118

```

Entfernen von Spalten

Für unsere Problemstellung benötigen wir nicht alle gelieferten Spalten. Um entscheiden zu können, welche Spalten für die Weiterarbeit weggelassen werden, schauen wir uns als erstes den Datenbeschrieb an.

Datenbeschrieb

mag: Werte von -1 bis 10 Die angegebene Stärke ist diejenige, die der U.S. Geological Survey für dieses Erdbeben als offiziell ansieht, und war die beste verfügbare Schätzung der Größe des Erdbebens zum Zeitpunkt der Erstellung dieser Seite. Andere Magnituden, die mit den von hier aus verlinkten Webseiten verbunden sind, wurden zu verschiedenen Zeitpunkten nach dem Erdbeben mit verschiedenen Arten von seismischen Daten ermittelt. Obwohl es sich dabei um legitime Schätzungen der Stärke handelt, betrachtet der U.S. Geological Survey sie nicht als die bevorzugte "offizielle" Stärke für das Ereignis.

Die Erdbebenstärke ist ein Maß für die Größe eines Erdbebens an seinem Ursprung. Es handelt sich um ein logarithmisches Maß. Bei gleichem Abstand vom Erdbeben ist die Amplitude der seismischen Wellen, aus denen die Magnitude bestimmt wird, bei einem Erdbeben der Magnitude 5 etwa zehnmal so groß wie bei einem Erdbeben der Magnitude 4. Die Gesamtenergiemenge, die durch das Erdbeben freigesetzt wird, steigt in der Regel um einen größeren Faktor: Bei vielen gebräuchlichen Magnitudenarten steigt die Gesamtenergie eines durchschnittlichen Erdbebens um einen Faktor von etwa 32 für jede Einheit, um die die Magnitude zunimmt.

Es gibt verschiedene Möglichkeiten, die Magnitude aus Seismogrammen zu berechnen. Die verschiedenen Methoden sind für unterschiedliche Erdbebengrößen und unterschiedliche Entfernungen zwischen der Erdbebenquelle und der Aufzeichnungsstation geeignet. Die verschiedenen Magnituden-Typen sind in der Regel so definiert, dass die Magnitudenwerte für Erdbeben in einem mittleren Bereich der aufgezeichneten Erdbebengrößen innerhalb weniger Zehntel einer Magnitudeneinheit übereinstimmen, aber die verschiedenen Magnituden-Typen können Werte aufweisen, die sich für sehr große und sehr kleine Erdbeben sowie für einige spezifische Klassen seismischer Quellen um mehr als eine Magnitudeneinheit unterscheiden. Dies ist darauf zurückzuführen, dass Erdbeben in der Regel komplexe Ereignisse sind, bei denen im Zuge des Verwerfungs- oder Bruchprozesses Energie in einem breiten Frequenzbereich und in unterschiedlicher Stärke freigesetzt wird. Die verschiedenen Arten von Magnituden messen unterschiedliche Aspekte der seismischen Strahlung

(z. B. niederfrequente Energie gegenüber hochfrequenter Energie). Die Beziehung zwischen den Werten der verschiedenen Magnitudenarten, die einem bestimmten seismischen Ereignis zugeordnet werden, kann dem Seismologen ein besseres Verständnis der Prozesse im Brennpunkt des seismischen Ereignisses ermöglichen. Die verschiedenen Magnituden-Typen sind nicht alle gleichzeitig für ein bestimmtes Erdbeben verfügbar.

Manchmal werden vorläufige Größenordnungen auf der Grundlage unvollständiger, aber schnell verfügbarer Daten geschätzt und gemeldet. Beispielsweise berechnen die Tsunami-Warnzentren eine vorläufige Magnitude und den Ort eines Ereignisses, sobald genügend Daten für eine Schätzung vorliegen. In diesem Fall ist die Zeit von entscheidender Bedeutung, um eine Warnung auszusenden, wenn durch das Ereignis wahrscheinlich Tsunami-Wellen ausgelöst werden. Solche vorläufigen Magnituden werden durch verbesserte Schätzungen der Magnitude ersetzt, sobald mehr Daten zur Verfügung stehen.

Bei großen Erdbeben der heutigen Zeit ist die Magnitude, die letztlich als bevorzugte Magnitude für die Berichterstattung an die Öffentlichkeit ausgewählt wird, in der Regel eine Momentmagnitude, die auf dem skalaren seismischen Moment eines Erdbebens basiert, das durch Berechnung des seismischen Momententensors bestimmt wird, der den Charakter der vom Erdbeben erzeugten seismischen Wellen am besten wiedergibt. Das skalare seismische Moment, ein Parameter des seismischen Momententensors, kann auch über das multiplikative Produkt Steifigkeit des gestörten Gesteins \times Bruchfläche \times durchschnittliche Verwerfungsverschiebung während des Erdbebens geschätzt werden.

rms: root-mean-square (RMS) Der quadratische Mittelwert (RMS) der Reisezeitresiduen in Sekunden unter Verwendung aller Gewichte ist ein Maß für die Übereinstimmung zwischen den beobachteten Ankunftszeiten und den vorhergesagten Ankunftszeiten an diesem Ort. Kleinere Werte deuten auf eine bessere Anpassung der Daten hin. Der Wert ist abhängig von der Genauigkeit des verwendeten Geschwindigkeitsmodells zur Berechnung der Erdbebenposition, der Gewichtung der Qualität der Ankunftszeitdaten und dem angewandten Verfahren zur Lokalisierung des Erdbebens. Typische Daten bewegen sich zwischen 0.13 und 1.39

nst: Die Anzahl der Seismostationen, die zur Bestimmung des Standorts eines Erdbebens verwendet werden.

dmin: Horizontale Entfernung in Grad vom Epizentrum zur nächstgelegenen Station. 1 Grad entspricht ungefähr 111,2 Kilometern. Im Allgemeinen gilt: Je geringer dieser Wert ist, desto zuverlässiger ist die berechnete Tiefe des Erdbebens. Typische Werte bewegen sich zwischen 0.4 und 7.1

gap: Der größte azimutale Abstand zwischen benachbarten Stationen in Grad beeinflusst die Zuverlässigkeit der berechneten horizontalen Position des Erdbebens. Eine geringere Distanz bedeutet im Allgemeinen höhere Genauigkeit. Größere azimutale Lücken als 180 Grad zeigen in der Regel große Unsicherheiten in Bezug auf die Lage und Tiefe des Erdbebens. Typischer Wertebereich: 0-180

magType: Die Methode oder der Algorithmus, der zur Berechnung der bevorzugten Größenordnung des Ereignisses verwendet wird. Werte: "Md", "MI", "Ms", "Mw", "Me", "Mi", "Mb", "MLg" Hier werden diese Typen genauer beschrieben: <https://www.usgs.gov/programs/earthquake-hazards/magnitude-types>

tz: Zeitonenabweichung von der UTC in Minuten am Epizentrum des Ereignisses.

net: Die ID eines Datenlieferanten. Kennzeichnet das Netz, das als bevorzugte Informationsquelle für dieses Ereignis gilt. Werte: ak, at, ci, hv, ld, mb, nc, nm, nn, pr, pt, se, us, uu, uw

sig: Eine Bewertungszahl, die die Bedeutung eines Ereignisses anzeigt. Je höher die Zahl, desto bedeutender das Ereignis. Dieser Wert wird anhand einer Reihe von Faktoren wie Ausmaß, maximaler MMI, berichteten Empfindungen und geschätzten Auswirkungen ermittelt.

ids: Eine durch Kommata getrennte Liste von Ereignis-IDs, die mit einem Ereignis verknüpft sind.

code: Ein von der entsprechenden Quelle für das Ereignis zugewiesener - und eindeutiger - Identifizierungscode.

sources: Eine durch Kommata getrennte Liste von Netzwerkteilnehmern.

tsunami: Dieses Flag wird bei großen Ereignissen in ozeanischen Regionen auf "1" gesetzt und ansonsten auf "0". Das Vorhandensein oder der Wert dieses Flags sagt nichts darüber aus, ob tatsächlich ein Tsunami aufgetreten ist oder auftreten wird. Wenn das Flag den Wert "1" hat, enthält das Ereignis einen Link zur NOAA Tsunami-Website für Tsunami-Informationen.

felt: Die Gesamtzahl der an das DYFI-System übermittelten Spürmeldungen. Did You Feel It? (DYFI) sammelt Informationen von Menschen, die ein Erdbeben gespürt haben, und erstellt Karten, die zeigen, was die Menschen erlebt haben und wie groß die Schäden sind. Werte zwischen 44 und 843

cdi: Die höchste gemeldete Intensität für das Ereignis. Berechnet von DYFI. Werte von 0-10

mmi: Die maximale geschätzte instrumentelle Intensität für das Ereignis. Berechnet von ShakeMap. ShakeMap ist ein Produkt des USGS Earthquake Hazards Program in Verbindung mit den regionalen seismischen Netzen. ShakeMaps liefern nahezu in Echtzeit Karten der Bodenbewegungen und der Erschütterungsintensität nach schweren Erdbeben. Diese Karten werden von öffentlichen und privaten Organisationen auf Bundes-, Landes- und Kommunalebene für die Reaktion und den Wiederaufbau nach einem Erdbeben, für öffentliche und wissenschaftliche Informationen sowie für Bereitschaftsübungen und Katastrophenplanung verwendet. Werte von 0-10

alert: Die Alarmstufe der PAGER-Skala für Erdbebenauswirkungen. Mögliche Werte: "green", "yellow", "orange", "red".

status: Zeigt an, ob das Ereignis von einem Menschen überprüft wurde. Der Status ist entweder automatisch oder geprüft. Automatische Ereignisse werden direkt von automatischen Verarbeitungssystemen gebucht und wurden nicht von einem Menschen überprüft oder geändert. Überprüfte Ereignisse wurden von einem Menschen geprüft. Der Grad der Überprüfung kann von einer schnellen Gültigkeitsprüfung bis hin zu einer sorgfältigen Neuanalyse des Ereignisses reichen. Mögliche Werte: "automatic", "reviewed", "deleted"

type: Art des seismischen Ereignisses."Erdbeben", "Steinbruch" Typische Werte: "earthquake", "quarry"

lon: Längengrad im WGS84-System angegeben. Die Werte bewegen sich zwischen -180.0 und 180.0.

lat: Breitengrad im WGS84-System angegeben. Die Werte bewegen sich zwischen -90.0 und 90.0.

depth: Tiefe des Ereignisses von der Erdoberfläche gemessen in Kilometer. Werte von 0 bis 1000.

Aus diesem Datenbeschrieb geht hervor, dass wir die Spalten *mag*, *place*, *time*, *status*, *updated*, *type*, *longitude*, *latitude*, *depth*, *tsunami*, *sig*, *code* behalten werden und alle anderen Spalten aus dem Data Frame entfernen.

```
df <- properties_df %>% select(mag, place, time, status, updated, type, lon, lat, depth, tsunami, sig, code)
```

Zeitstempel konvertieren

Wie wir bereits gesehen haben, müssen die Spalten mit den Zeitstempeln (*time* und *updated*) noch konvertiert werden. Der Zeitstempel wird in Millisekunden seit der Epoche (1970-01-01T00:00:00.000Z), angegeben, ohne Berücksichtigung von Schaltsekunden.

Dazu wird für die Berechnung das Basisdatum addiert mit dem vorhandenen Wert, welcher wiederum durch 1000 geteilt wird.

```
# properties_df <- properties_df %>%
basedate <- as.POSIXct("1970-01-01 00:00:00", tz = "UTC")

df <- df %>%
  mutate(time = basedate + (time/ 1000)) %>%
  mutate(updated = basedate + (updated/1000))
```

Falls in den Spalten *mag*, *place*, *time*, *type*, *longitude*, *latitude*, *depth*, *code* fehlende Werte vorhanden sind, ist eine Verwendung dieser Observationen nicht sinnvoll. Deshalb werden diese Observationen aus dem Datensatz entfernt.

```
df <- df %>%
  drop_na(mag, place, time, type, lon, lat, depth, code)
```


Überprüfung der zulässigen Werte

Wie im Datenbeschrieb zu entnehmen, dürfen sich die Werte für den Längengrad im Bereich von -180 bis 180, für den Breitengrad von -90 bis 90 und für die Tiefe von 0 bis 1000 bewegen. Dies wird anhand einer Abfrage überprüft. Dazu wird eine zusätzliche Spalte für die jeweilige Variable erstellt und mit einem booleanschen Operator versehen. Der Wert True steht jeweils für einen Wert, der sich im zulässigen Bereich befindet.

```
df <- df %>%  
  mutate(lon_inrange = between(lon, -180, 180),  
         lat_inrange = between(lat, -90, 90),  
         depth_inrange = between(depth, 0, 1000),  
         mag_inrange = between(mag, -1, 10))
```

Werte, welche sich nicht im zulässigen Bereich befinden, in einem neuen Data Frame anzeigen lassen.

```
not_inrange <- df %>%  
  filter(lon_inrange == FALSE,  
         lat_inrange == FALSE,  
         depth_inrange == FALSE,  
         mag_inrange == FALSE)  
not_inrange
```

```
## [1] mag           place           time           status          updated  
## [6] type           lon            lat            depth           tsunami  
## [11] sig            code           lon_inrange    lat_inrange     depth_inrange  
## [16] mag_inrange  
## <0 Zeilen> (oder row.names mit Länge 0)
```

Falls sich Werte nicht im zulässigen Bereich befinden, werden diese Observationen entfernt.

```
df <- df %>%  
  filter(lon_inrange == TRUE,  
         lat_inrange == TRUE,  
         depth_inrange == TRUE,  
         mag_inrange == TRUE)  
df
```

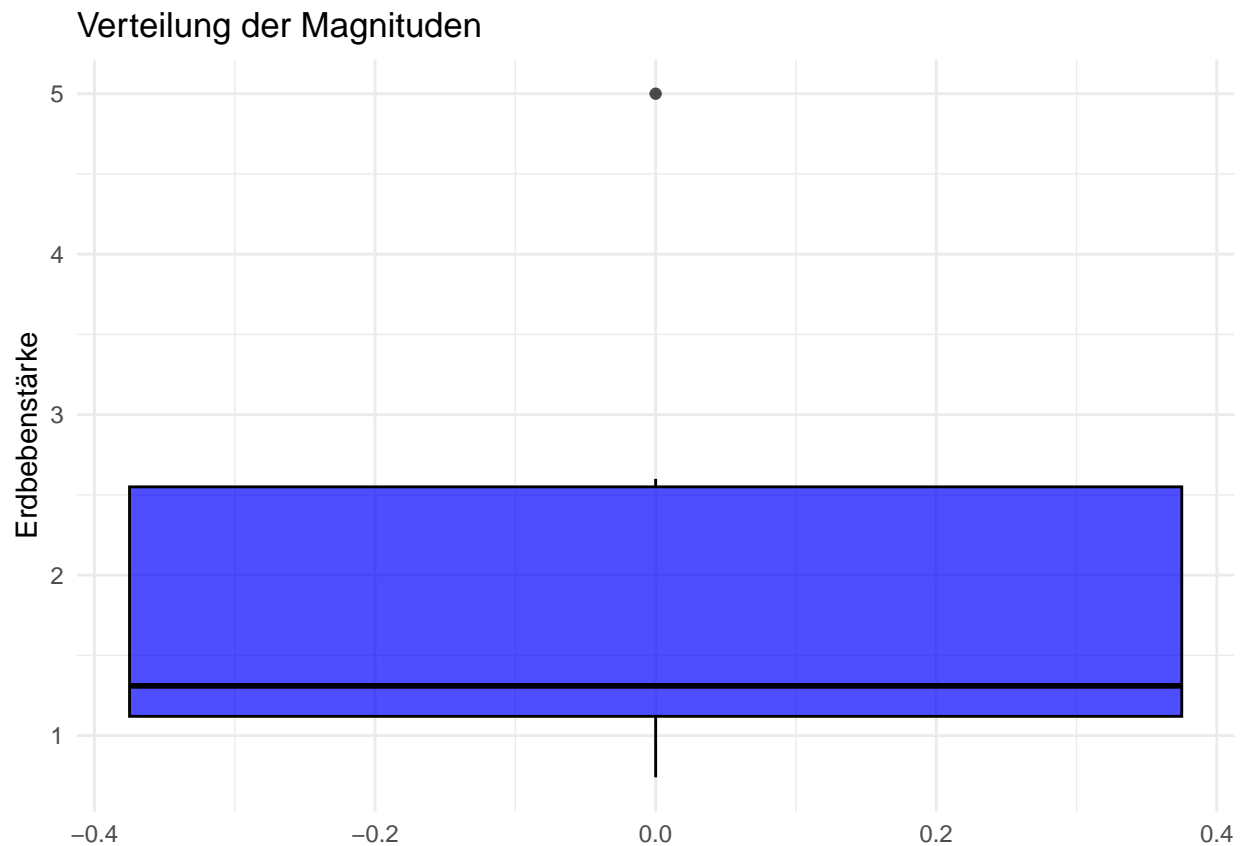
```
##      mag           place           time      status  
## 1 0.74    10 km WNW of The Geysers, CA 2023-10-29 12:50:56 automatic  
## 2 1.23   18 km NNW of Ocotillo Wells, CA 2023-10-29 12:36:43 automatic  
## 3 5.00    Kermadec Islands, New Zealand 2023-10-29 12:28:17 reviewed  
## 4 1.01      8 km NW of The Geysers, CA 2023-10-29 12:27:51 automatic  
## 5 2.50     30 km ESE of Mapleton, Utah 2023-10-29 12:15:41 reviewed  
## 6 2.60     66 km SE of Kaktovik, Alaska 2023-10-29 12:12:38 reviewed  
## 7 1.31      2 km NNE of The Geysers, CA 2023-10-29 12:07:18 automatic  
##      updated      type      lon      lat      depth tsunami sig  
## 1 2023-10-29 12:52:30 earthquake -122.8540 38.82367 2.410      0      8  
## 2 2023-10-29 12:40:29 earthquake -116.1807 33.29767 0.420      0     23  
## 3 2023-10-29 12:47:58 earthquake -178.1653 -30.67040 35.000      0    385  
## 4 2023-10-29 12:43:21 earthquake -122.8183 38.83133 1.600      0     16  
## 5 2023-10-29 12:38:50 earthquake -111.2791 39.98840 9.938      0     96  
## 6 2023-10-29 12:30:25 earthquake -142.3912 69.71150 6.118      0    104  
## 7 2023-10-29 12:44:10 earthquake -122.7475 38.79250 1.010      0     26  
##      code lon_inrange lat_inrange depth_inrange mag_inrange  
## 1 73953270      TRUE      TRUE      TRUE      TRUE  
## 2 40590680      TRUE      TRUE      TRUE      TRUE  
## 3 700017d1      TRUE      TRUE      TRUE      TRUE
```

## 4	73953260	TRUE	TRUE	TRUE	TRUE
## 5	700017dh	TRUE	TRUE	TRUE	TRUE
## 6	700017di	TRUE	TRUE	TRUE	TRUE
## 7	73953250	TRUE	TRUE	TRUE	TRUE

Explorative Analyse

Verteilung der Erdbebenstärken (Magnituden) im Datenset. Dazu wird eine Visualisierung erstellt.

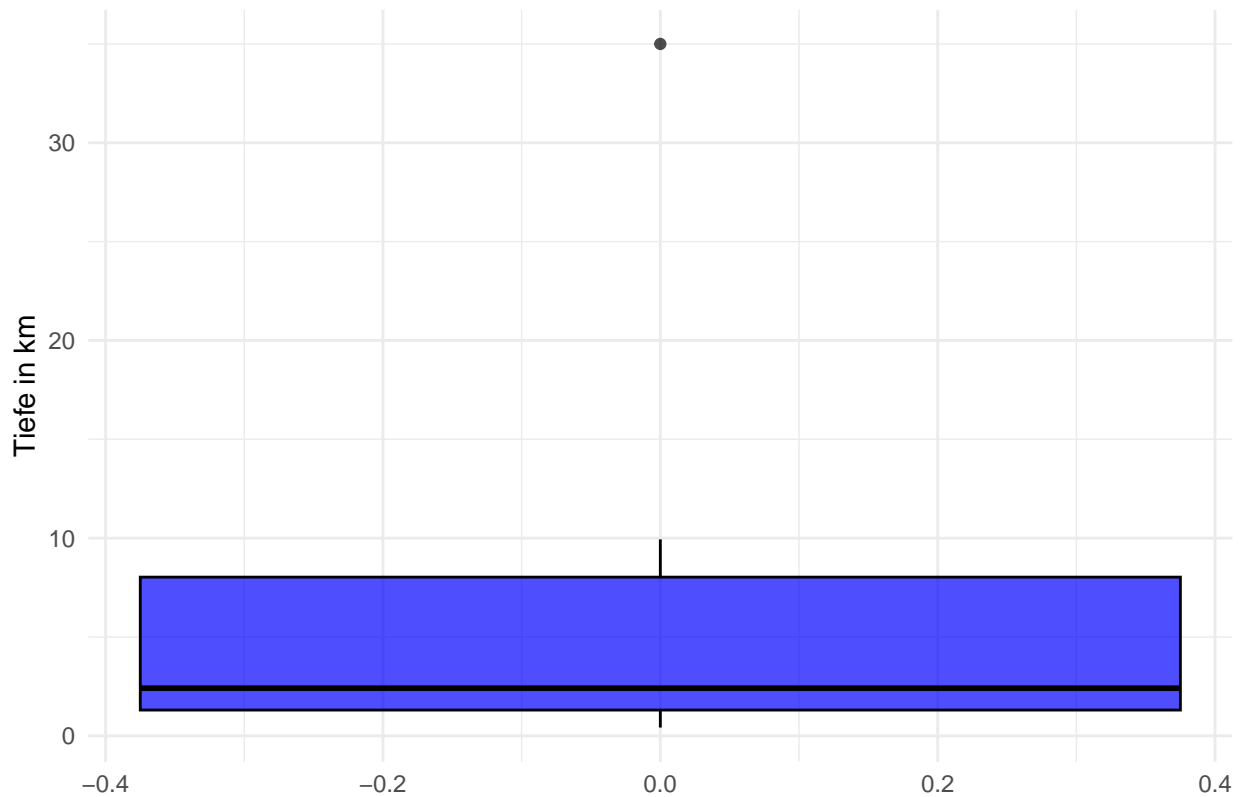
```
ggplot(df, aes(y = mag)) +  
  geom_boxplot(fill="blue", color="black", alpha=0.7) +  
  theme_minimal() +  
  labs(title="Verteilung der Magnituden", y="Erdbebenstärke")
```



Nun wollen wir schauen, wie die Tiefen der Epizentren verteilt sind

```
ggplot(df, aes(y = depth)) +  
  geom_boxplot(fill="blue", color="black", alpha=0.7) +  
  theme_minimal() +  
  labs(title="Verteilung der Tiefen der Epizentren", y="Tiefe in km", y="")
```

Verteilung der Tiefen der Epizentren



Sind Tsunami-Warnungen ausgegeben worden?

```
df %>% filter(tsunami != 0)
```

```
## [1] mag      place      time      status      updated
## [6] type      lon        lat        depth      tsunami
## [11] sig       code       lon_inrange lat_inrange depth_inrange
## [16] mag_inrange
## <0 Zeilen> (oder row.names mit Länge 0)
```

Visualisierung auf der Weltkarte

Für die Visualisierung auf einer Weltkarte kann die Bibliothek “leaflet” verwendet werden.

```
map <- leaflet(df) %>%
  addTiles() %>%
  addCircleMarkers(
    lat = ~lat,
    lng = ~lon,
    radius = ~mag * 2, # Größe der Punkte nach Magnitude
    fillColor = ~colorQuantile("YlOrRd", mag)(mag), # Farbschema
    color = "#000000", # Umrandungsfarbe
    weight = 1, # Umrandungsdicke
    opacity = 1, # Umrandungstransparenz
    fillOpacity = 0.7, # Fülltransparenz
    popup = ~paste("Ort:", place, "<br>Magnitude:", round(mag, 4), "<br>Tiefe:", round(df$depth, 2))
  )
print(map)
```

Für einen Kartenausschnitt, in dem alle vorhandenen Erdbeben ersichtlich sind, kann die bbox verwendet werden. Die Daten werden im JSON-Objekt mitgeliefert. Für die Verwendung in einem Leaflet werden diese Informationen in ein Data Frame gespeichert.

```
bbox <- past_hours_js$bbox
```

```
bbox <- data.frame(bbox, coord = c("lon_min", "lat_min", "depth_min", "lon_max", "lat_max", "depth_max"))
```

```
bbox
```

##		bbox	coord
## 1	-178.1653	lon_min	
## 2	-30.6704	lat_min	
## 3	0.4200	depth_min	
## 4	-111.2791	lon_max	
## 5	69.7115	lat_max	
## 6	35.0000	depth_max	