Capstone\_Earthquakes

Steph Loney

2023-04-01

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

#Import the Earthquake dataset

df = read.csv("C:\\Users\\Steph\\Documents\\Assignments\\Capstone\\all\_month\_earthquakes.csv")

#Duplicate “Place” column

df$state = df$place

#Remove everything before the comma in the new “state” variable to identify states (as well as non-US locations)

df$state = sub(".\*,", "", df$state)  
head(df$state)

## [1] " Oklahoma" " CA" " Alaska" " CA" " CA" " CA"

#Find out each of the possible values for “place” now that it’s been simplified

unique(df$state)

## [1] " Oklahoma"   
## [2] " CA"   
## [3] " Alaska"   
## [4] " Chile"   
## [5] " California"   
## [6] " Tonga"   
## [7] " Hawaii"   
## [8] " Puerto Rico"   
## [9] "Kermadec Islands region"   
## [10] " Nevada"   
## [11] " India"   
## [12] " Guam"   
## [13] "Central Alaska"   
## [14] " Japan region"   
## [15] " Wyoming"   
## [16] " New Zealand"   
## [17] " Idaho"   
## [18] " Mexico"   
## [19] "southern Idaho"   
## [20] " Fiji"   
## [21] " Indonesia"   
## [22] " U.S. Virgin Islands"   
## [23] " Oregon"   
## [24] " Nicaragua"   
## [25] " Arkansas"   
## [26] " Montana"   
## [27] " Argentina"   
## [28] " Japan"   
## [29] " Tajikistan"   
## [30] " Greece"   
## [31] " Papua New Guinea"   
## [32] " Washington"   
## [33] "Banda Sea"   
## [34] "south of the Fiji Islands"   
## [35] "southeast of the Loyalty Islands"   
## [36] " China"   
## [37] " Solomon Islands"   
## [38] " Tennessee"   
## [39] " Texas"   
## [40] " Timor Leste"   
## [41] " El Salvador"   
## [42] " Portugal"   
## [43] " Canada"   
## [44] "Vanuatu region"   
## [45] "south of Tonga"   
## [46] " Italy"   
## [47] " Philippines"   
## [48] " New Mexico"   
## [49] " Afghanistan"   
## [50] " NV"   
## [51] " Panama"   
## [52] "southern Mid-Atlantic Ridge"   
## [53] " Anguilla"   
## [54] " MX"   
## [55] "Southern Alaska"   
## [56] "Reykjanes Ridge"   
## [57] "Fiji region"   
## [58] "Greenland Sea"   
## [59] " Utah"   
## [60] "Greece-Albania border region"   
## [61] "Wyoming"   
## [62] " Russia"   
## [63] " South Carolina"   
## [64] "Washington"   
## [65] " Vanuatu"   
## [66] " Pakistan"   
## [67] "central Mongolia"   
## [68] "western Texas"   
## [69] " Colombia"   
## [70] " Peru"   
## [71] "Dominican Republic region"   
## [72] "Nevada"   
## [73] " Missouri"   
## [74] "Tristan da Cunha region"   
## [75] "Puerto Rico region"   
## [76] " Costa Rica"   
## [77] "Kuril Islands"   
## [78] "south of the Kermadec Islands"   
## [79] "Oklahoma"   
## [80] "Alaska Peninsula"   
## [81] "south of Panama"   
## [82] "Bismarck Sea"   
## [83] " Turkey"   
## [84] " Guatemala"   
## [85] "Near the coast of southern Peru"   
## [86] " Cuba"   
## [87] " Syria"   
## [88] " Bolivia"   
## [89] " Iran"   
## [90] " Wallis and Futuna"   
## [91] "Pitcairn Islands region"   
## [92] "Mid-Indian Ridge"   
## [93] "central East Pacific Rise"   
## [94] " Arizona"   
## [95] " New York"   
## [96] " Maine"   
## [97] "northern Mid-Atlantic Ridge"   
## [98] " New Caledonia"   
## [99] " Kansas"   
## [100] "California-Nevada border region"   
## [101] "South Sandwich Islands region"   
## [102] "east of the Philippine Islands"   
## [103] " Dominican Republic"   
## [104] "Kyrgyzstan-Tajikistan border region"   
## [105] " Florida"   
## [106] " Taiwan"   
## [107] " Venezuela"   
## [108] "Utah"   
## [109] " Haiti"   
## [110] "Western Montana"   
## [111] "Southern Texas"   
## [112] " Virginia"   
## [113] "Idaho-Montana border region"   
## [114] " Myanmar"   
## [115] "Southwest Indian Ridge"   
## [116] " Colorado"   
## [117] "north of Ascension Island"   
## [118] "Myanmar"   
## [119] "Tyrrhenian Sea"   
## [120] "Chagos Archipelago region"   
## [121] " Morocco"   
## [122] " Northern Mariana Islands"   
## [123] " Kyrgyzstan"   
## [124] " Barbados"   
## [125] "South Atlantic Ocean"   
## [126] "north of Svalbard"   
## [127] "Mariana Islands region"   
## [128] "Pacific-Antarctic Ridge"   
## [129] "west of Macquarie Island"   
## [130] "south of Africa"   
## [131] " Thailand"   
## [132] " Romania"   
## [133] " South Africa"   
## [134] "southeast Indian Ridge"   
## [135] " Ecuador"   
## [136] "west of the Galapagos Islands"   
## [137] "Solomon Islands"   
## [138] "central Mid-Atlantic Ridge"   
## [139] "Balleny Islands region"   
## [140] "Owen Fracture Zone region"   
## [141] "east of the Kuril Islands"   
## [142] " Mauritania"   
## [143] "South Indian Ocean"   
## [144] "Prince Edward Islands region"   
## [145] "Southeastern Alaska"   
## [146] " Iowa"   
## [147] "Easter Island region"   
## [148] "off the coast of Oregon"   
## [149] "Northern California"   
## [150] "southern East Pacific Rise"   
## [151] " Jamaica"   
## [152] " Kiribati region"   
## [153] "Afghanistan-Tajikistan-Pakistan region"  
## [154] " Maryland"   
## [155] " Georgia"   
## [156] " Namibia"   
## [157] " Australia"   
## [158] " South Sudan"   
## [159] " Kazakhstan"   
## [160] " France"   
## [161] "southeast of Easter Island"   
## [162] " Montenegro"

#Export dataset to work in Excel. This will better allow me to manipulate this data to delete rows that pertain non-US locations (not pertinent to this project) and clean up the locations as well (update “CA” to “California,” “Central Alaska” to “Alaska” etc.). It should be noted that the two US border regions have not been collapsed down into one state as these straddle state lines.

write.csv(df, "C:\\Users\\Steph\\Documents\\Assignments\\Capstone\\state\_eq.csv", row.names=FALSE)

#Open updated dataset

eq = read.csv("C:\\Users\\Steph\\Documents\\Assignments\\Capstone\\state\_eq(edit).csv")

#Create a California-specific dataframe (this does not contain the CA/NV border numbers as this lies in a “limbo” area). Keep only earthquake entries (delete “quarry blast” and “sonic boom”)

ca = eq[(eq$state %in% "California"),]  
ca2 = subset(ca, type!="quarry blast" & type!="sonic boom")  
table(ca2$type)

##   
## earthquake   
## 5049

#Check numbers from original df dataset

table(df$type)

##   
## earthquake experimental explosion explosion   
## 12564 1 46   
## ice quake other event quarry blast   
## 25 1 125   
## sonic boom   
## 1

df2 = subset(df, type!="experimental explosion" & type!="explosion" & type!="ice quake" & type!="other event" & type!="quarry blast" & type!="sonic boom")  
table(df2$type)

##   
## earthquake   
## 12564

table(df$state)

##   
## Afghanistan Alaska   
## 8 2448   
## Anguilla Argentina   
## 6 17   
## Arizona Arkansas   
## 3 6   
## Australia Barbados   
## 1 1   
## Bolivia CA   
## 4 3840   
## California Canada   
## 1279 22   
## Chile China   
## 64 18   
## Colombia Colorado   
## 10 7   
## Costa Rica Cuba   
## 1 3   
## Dominican Republic Ecuador   
## 12 1   
## El Salvador Fiji   
## 6 10   
## Florida France   
## 1 1   
## Georgia Greece   
## 1 18   
## Guam Guatemala   
## 9 7   
## Haiti Hawaii   
## 1 860   
## Idaho India   
## 139 9   
## Indonesia Iowa   
## 97 1   
## Iran Italy   
## 16 2   
## Jamaica Japan   
## 1 46   
## Japan region Kansas   
## 14 7   
## Kazakhstan Kiribati region   
## 1 1   
## Kyrgyzstan Maine   
## 1 1   
## Maryland Mauritania   
## 1 1   
## Mexico Missouri   
## 13 4   
## Montana Montenegro   
## 934 1   
## Morocco MX   
## 1 16   
## Myanmar Namibia   
## 2 1   
## Nevada New Caledonia   
## 428 5   
## New Mexico New York   
## 156 2   
## New Zealand Nicaragua   
## 13 8   
## Northern Mariana Islands NV   
## 3 190   
## Oklahoma Oregon   
## 405 52   
## Pakistan Panama   
## 4 11   
## Papua New Guinea Peru   
## 33 16   
## Philippines Portugal   
## 28 2   
## Puerto Rico Romania   
## 298 1   
## Russia Solomon Islands   
## 18 13   
## South Africa South Carolina   
## 1 3   
## South Sudan Syria   
## 1 1   
## Taiwan Tajikistan   
## 15 12   
## Tennessee Texas   
## 18 293   
## Thailand Timor Leste   
## 1 10   
## Tonga Turkey   
## 24 4   
## U.S. Virgin Islands Utah   
## 7 126   
## Vanuatu Venezuela   
## 23 1   
## Virginia Wallis and Futuna   
## 1 6   
## Washington Wyoming   
## 207 87   
## Afghanistan-Tajikistan-Pakistan region Alaska Peninsula   
## 1 4   
## Balleny Islands region Banda Sea   
## 1 9   
## Bismarck Sea California-Nevada border region   
## 1 7   
## Central Alaska central East Pacific Rise   
## 3 4   
## central Mid-Atlantic Ridge central Mongolia   
## 1 1   
## Chagos Archipelago region Dominican Republic region   
## 4 1   
## east of the Kuril Islands east of the Philippine Islands   
## 1 1   
## Easter Island region Fiji region   
## 2 11   
## Greece-Albania border region Greenland Sea   
## 1 1   
## Idaho-Montana border region Kermadec Islands region   
## 3 15   
## Kuril Islands Kyrgyzstan-Tajikistan border region   
## 7 1   
## Mariana Islands region Mid-Indian Ridge   
## 3 4   
## Myanmar Near the coast of southern Peru   
## 1 1   
## Nevada north of Ascension Island   
## 6 1   
## north of Svalbard Northern California   
## 3 1   
## northern Mid-Atlantic Ridge off the coast of Oregon   
## 5 1   
## Oklahoma Owen Fracture Zone region   
## 5 2   
## Pacific-Antarctic Ridge Pitcairn Islands region   
## 2 1   
## Prince Edward Islands region Puerto Rico region   
## 2 11   
## Reykjanes Ridge Solomon Islands   
## 1 1   
## South Atlantic Ocean South Indian Ocean   
## 1 1   
## south of Africa south of Panama   
## 2 1   
## south of the Fiji Islands south of the Kermadec Islands   
## 26 2   
## south of Tonga South Sandwich Islands region   
## 2 15   
## southeast Indian Ridge southeast of Easter Island   
## 3 1   
## southeast of the Loyalty Islands Southeastern Alaska   
## 7 3   
## Southern Alaska southern East Pacific Rise   
## 2 1   
## southern Idaho southern Mid-Atlantic Ridge   
## 2 2   
## Southern Texas Southwest Indian Ridge   
## 2 2   
## Tristan da Cunha region Tyrrhenian Sea   
## 1 1   
## Utah Vanuatu region   
## 5 1   
## Washington west of Macquarie Island   
## 3 1   
## west of the Galapagos Islands Western Montana   
## 1 4   
## western Texas Wyoming   
## 14 24

#Cut out all non “earthquake” type variables and check numbers from eq dataset

eq2 = eq[(eq$type %in% "earthquake"),]  
table(eq2$state)

##   
## Alaska Arizona   
## 2425 3   
## Arkansas California   
## 6 5049   
## California-Nevada border region Colorado   
## 7 7   
## Georgia Hawaii   
## 1 860   
## Idaho Idaho-Montana border region   
## 141 3   
## Iowa Kansas   
## 1 7   
## Maine Maryland   
## 1 1   
## Missouri Montana   
## 4 915   
## Nevada New Mexico   
## 624 156   
## New York Oklahoma   
## 2 379   
## Oregon South Carolina   
## 41 3   
## Tennessee Texas   
## 18 309   
## Utah Virginia   
## 131 1   
## Washington Wyoming   
## 192 111

#Check ca dataset numbers

table(ca2$state)

##   
## California   
## 5049

#Question 1

#Find the average depth of earthquakes of magnitude greater than 3.0 in California

mean(ca2$depth[ca2$mag > 3.0], na.rm=TRUE)

## [1] 5.685645

#Question 2

#Calcualate average magnitude of an earthquake in CA

mean(ca2$mag, na.rm=TRUE)

## [1] 1.151634

#Create non-CA US dataset

nus = eq2[!(eq2$state %in% "California"),]

#Calcualate average magnitude of an earthquake in US

mean(nus$mag)

## [1] 1.349052

#Question 3

#Calculate earthquakes that occurred in CA

table(ca2$type)

##   
## earthquake   
## 5049

#Calculate total earthquakes in dataset

table(df2$type)

##   
## earthquake   
## 12564

#California earthquakes make up 5049 of the total 12564 in the dataset, or just about 40%.

5049/12564

## [1] 0.4018625

#Question 4

#Bins for each full number of magnitude and visualize this.

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

breaks = c(-2, -1, 0, 1, 2, 3, 4, 5, 6, 7)  
  
ggplot(df2, aes(x = mag)) +  
 geom\_histogram(color = "white", fill = "lightblue", breaks = breaks) +  
 scale\_x\_continuous(breaks = breaks) +  
 geom\_text(  
 stat = "bin", aes(label = after\_stat(count)),  
 vjust = -1, breaks = breaks)

## Warning: Removed 1 rows containing non-finite values (`stat\_bin()`).

## Warning: Removed 1 rows containing non-finite values (`stat\_bin()`).

