Earthquake Research

Leona Pierce, Matt Collins

Intro

Set up

The first step before beginning the investigation of the data is to set up our environment with the necessary packages and import the data. This is done in the code block below.

```
library(tidyverse)
library(janitor)
library(dplyr)
library(knitr)
library(ggplot2)
earthquake <- read.csv("./data/earthquake_data.csv")
```

To make the data easier to use, it is also important to tidy the data. In this case, the data itself is tidy, but the column names could be shorter and more direct to make it easier to reference them within the code.

```
names(earthquake)[1] <- "general_worry"
names(earthquake)[2] <- "big_one_worry"
names(earthquake)[3] <- "big_one_occur"
names(earthquake)[4] <- "experienced_earthquake"
names(earthquake)[5] <- "taken_preacautions"
names(earthquake)[6] <- "san_andreas_familiar"
names(earthquake)[7] <- "yellowstone_familiar"
names(earthquake)[8] <- "age"
names(earthquake)[9] <- "gender"
names(earthquake)[10] <- "household_income"
names(earthquake)[11] <- "region"

earthquake <- earthquake
```

Research Questions

In order to explore this data, we have selected 4 research questions to investigate.

Question 1

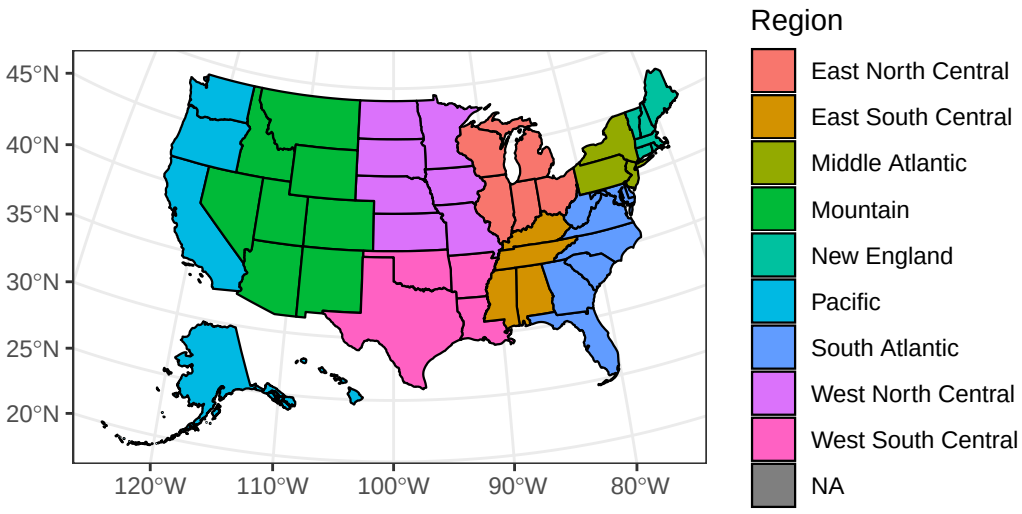
The first question we want to answer is where in the United States are people most worried about earthquakes and where in the United States are people most prepared for earthquakes? This question can be reinterpreted as the relationship between region and worriedness or preparedness.

First, let's look at where most of the responses in this data are coming from by creating a table showing the response counts of each region and the corresponding percentage of the whole.

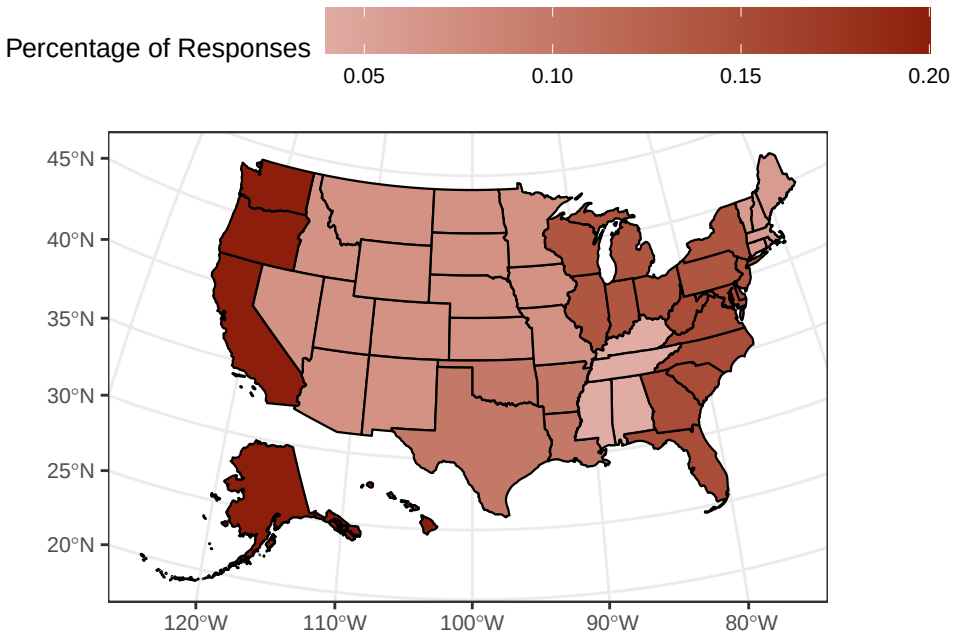
Region	Count	Percentage
	35	0.03
East North Central	140	0.14
East South Central	40	0.04
Middle Atlantic	137	0.14
Mountain	67	0.07
New England	63	0.06
Pacific	206	0.20
South Atlantic	155	0.15
West North Central	71	0.07
West South Central	99	0.10

The table shows that the responses are mostly spread out across the whole of the United States, although there are noticeably the fewest responses in the East South Central region. The blank row at the top of the table shows responses that did not give a region, which will be ignored in our analysis.

An important thing to define before we continue with our analysis is what states are contained in each region. To show maps of the US in this analysis, I will use the "usmap" package. This first map will give each region (or state in that region) a different color, and I am using the US Census Bureau's definition of the states in each region.



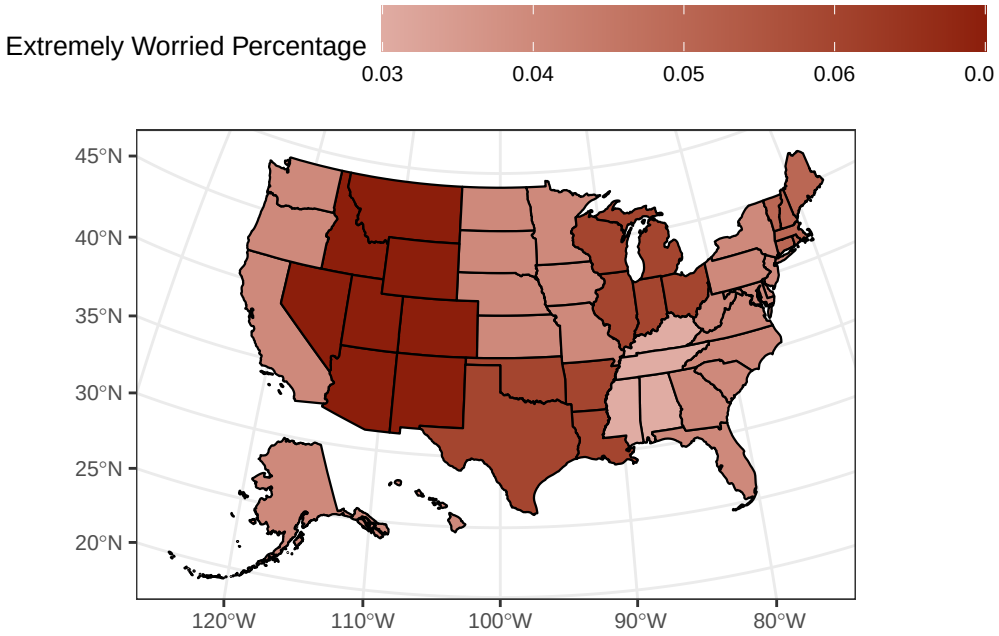
Now, let's recreate the table of regions and their respective count of responses in map form. An important thing to note about the maps in this analysis is that since we were not given an exact US state for the response, all states in a region will have identical counts or percentages.

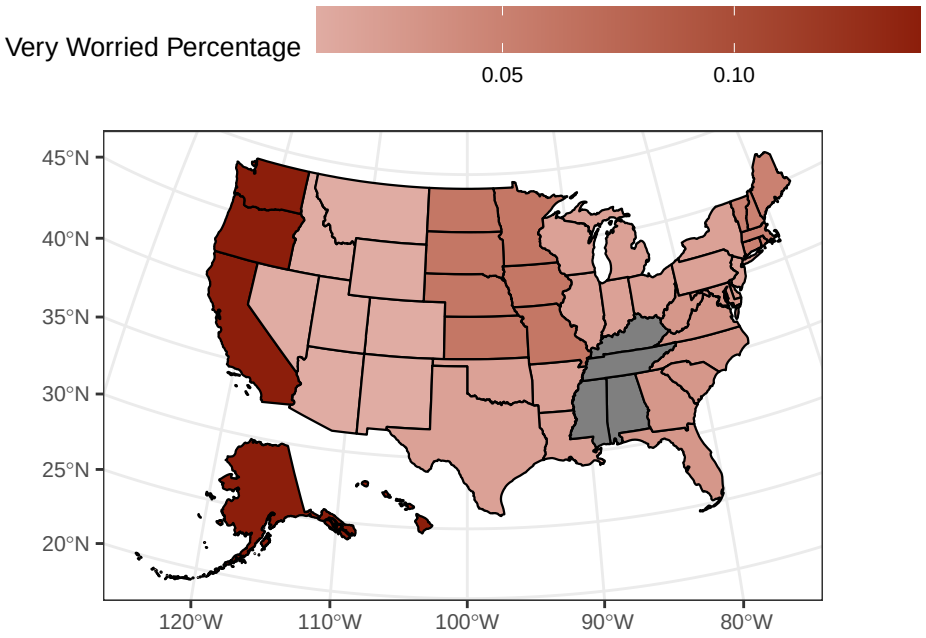


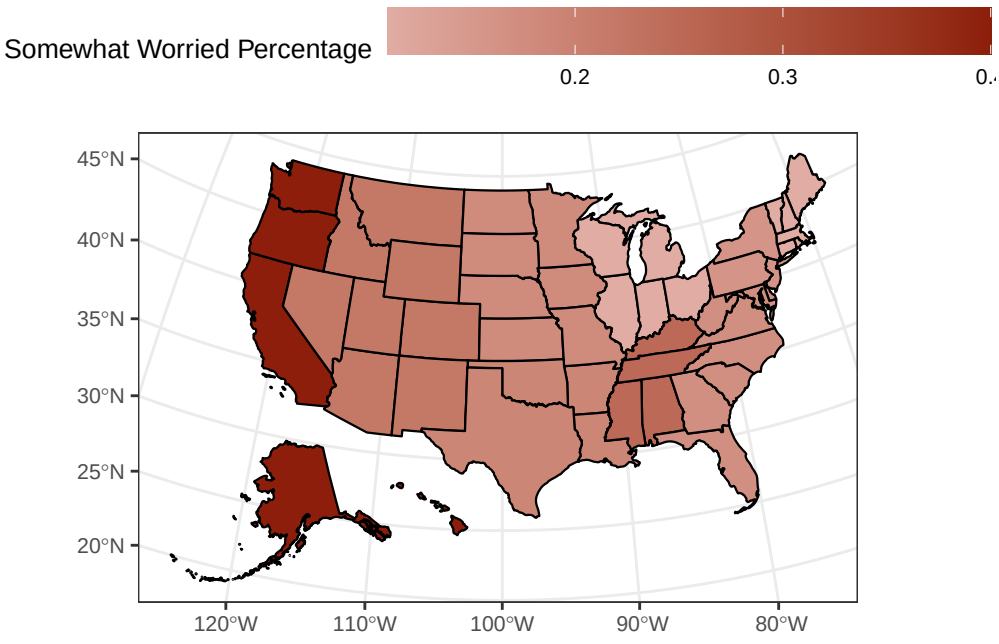
Now I will move into answering the first part of our research question: where in the US are people most worried about earthquakes? This will be accomplished using the "general_worry" column in our dataset. First, I will create a table where each row is a combination of US state and level of worry (of 5 possible choices).

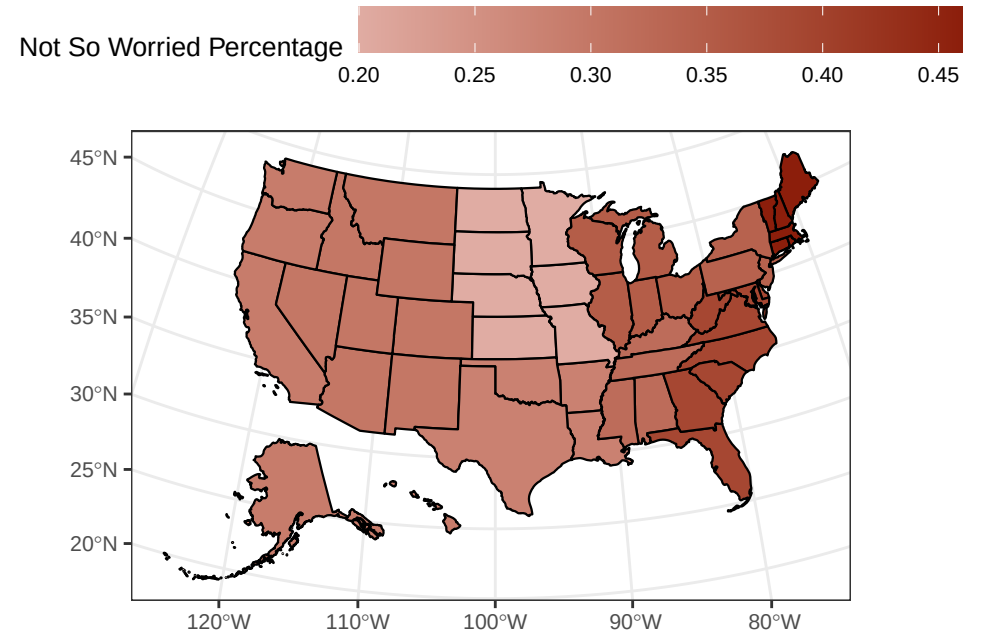
```
# A tibble: 6 x 7
  Region             state general_worry     count total_percentage region_count
  <chr>              <chr> <chr>             <int>            <dbl>        <int>
1 East North Central IL    Extremely worried     8             0.01          140
2 East North Central IL    Not at all worri~    64             0.06          140
3 East North Central IL    Not so worried       49             0.05          140
4 East North Central IL    Somewhat worried     16             0.02          140
5 East North Central IL    Very worried          3             0             140
6 East North Central IN    Extremely worried     8             0.01          140
# i 1 more variable: region_percentage <dbl>
```

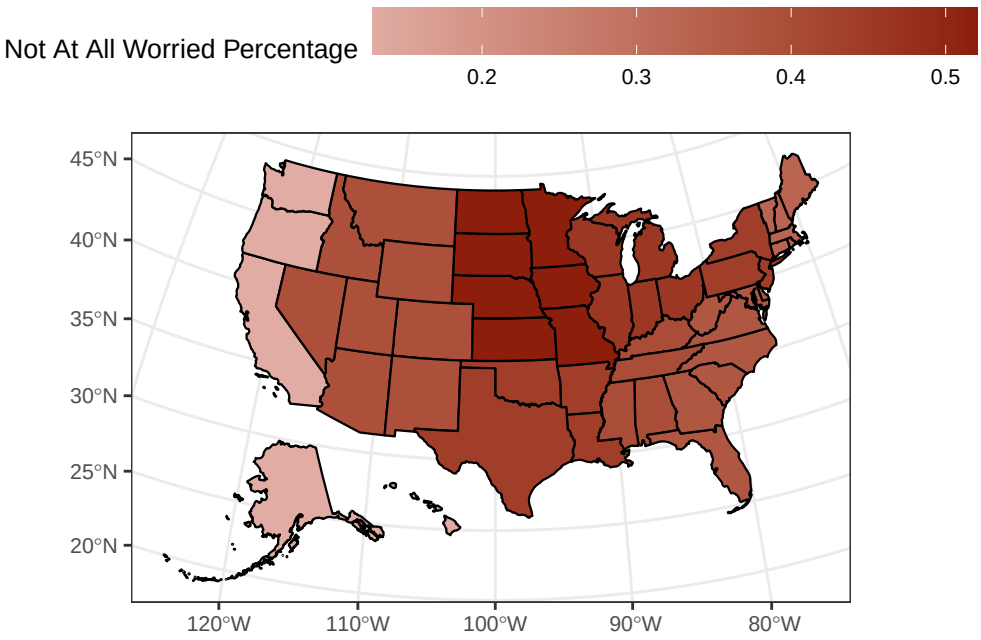
Then, we can plug this table into the same process for our response count map and get 5 new maps showing the percentage of different levels of worry, with one map for each level of worry.




Very Worried Percentage
0.05
0.10
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W


Somewhat Worried Percentage
0.2
0.3
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W





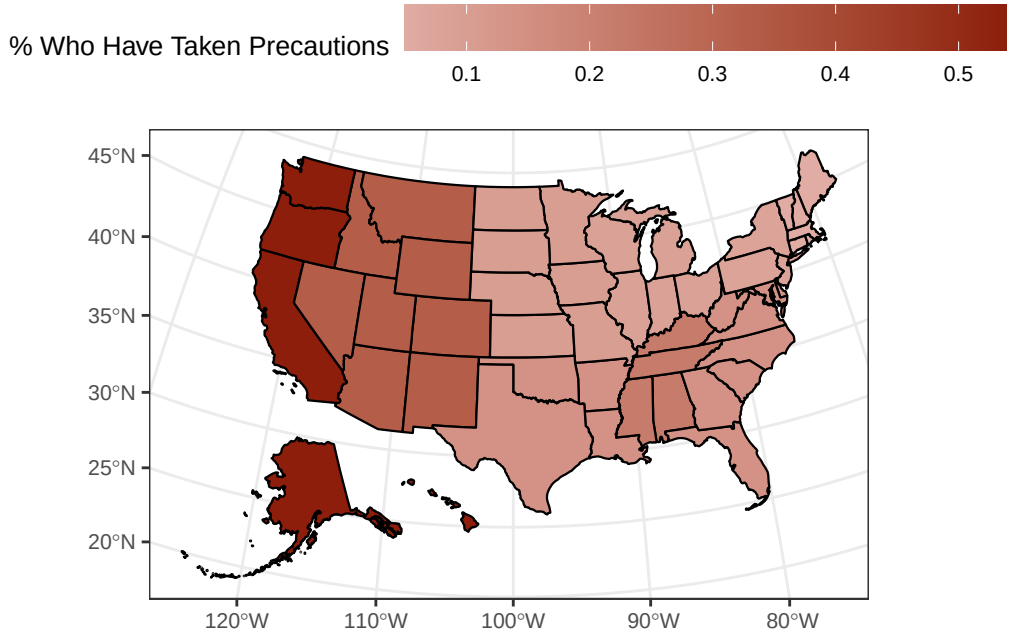
As we might have expected, the states/regions that are most worried about earthquakes are the western ones, particularly the Pacific and Mountain regions, where earthquakes are the most common. It is important to pay attention to the scale at the top of these maps because although the Mountain region is a darker shade of red for the "Extremely Worried" graph than

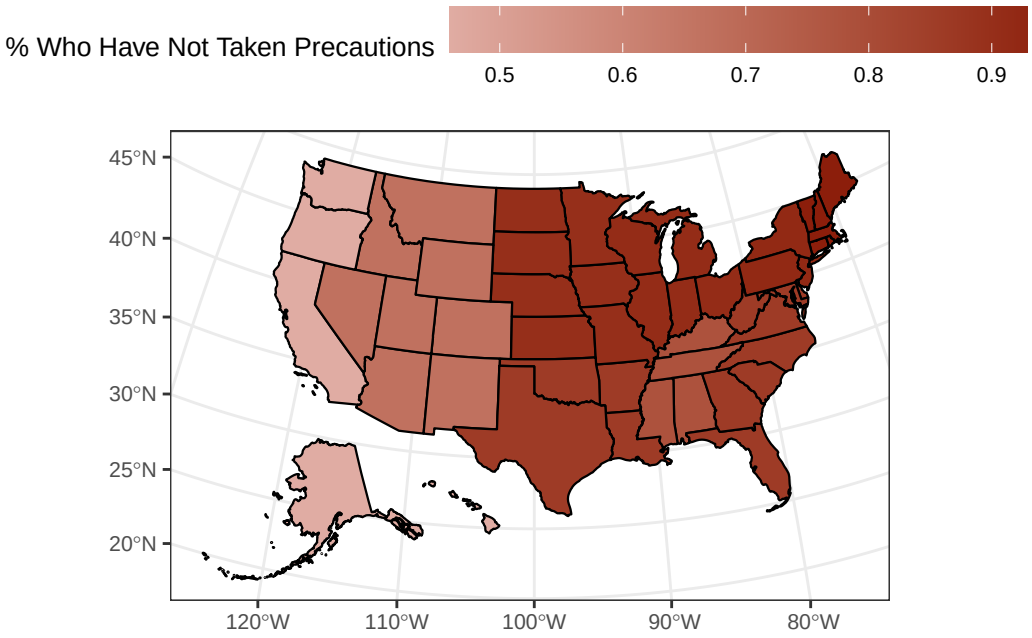
it is for the "Somewhat Worried" graph, the scales show that a higher percentage of responses are "Somewhat Worried". Another side note is that the East South Central region is gray in the "Very Worried" graph because that region had no responses with that level of worry. From these maps, we can conclude that the western portion of the United States is generally more worried about earthquakes.

Now, we will follow this exact same process for determining where in the US people are most likely to have prepared for earthquakes, using the "taken_precautions" column. Below is the table that we will use to create the maps.

```
# A tibble: 6 x 7
  Region            state taken_preacautions count total_percentage region_count
  <chr>             <chr> <chr>               <int>            <dbl>        <int>
1 East North Centr~ IL    No                    128             0.13          140
2 East North Centr~ IL    Yes                    12             0.01          140
3 East North Centr~ IN    No                    128             0.13          140
4 East North Centr~ IN    Yes                    12             0.01          140
5 East North Centr~ OH    No                    128             0.13          140
6 East North Centr~ OH    Yes                    12             0.01          140
# i 1 more variable: region_percentage <dbl>
```

Below are the maps, one for the "Yes" response and one for the "No" response. Because it is a binary response, the colors of the two maps will be inverse of each other.





Just like the graphs displaying the level of worry, these graphs show that the Western part of the United States is more likely to have taken precautions for earthquakes. In fact, it almost appears to be a gradient going from west to east. Logically, this tracks because a person is more likely to have taken precautions for earthquakes if earthquakes happen relatively often. In other words, you won't prepare for something extremely unlikely to happen to you.

The following visualizations have allowed us to answer where in the US people are most worried and prepared for earthquakes, the answer being the Western part of the United States where earthquakes occur more often.

Question 2

Does Knowledge affect how worried someone is?

Moving beyond how worried Americans are generally about earthquakes, we wanted to look specifically at how Americans feels about the "Big One". One question we had in regard to the "Big One" is how knowledge about both the San Andreas fault line and the Yellowstone supervolcano (which are possible causes of the "Big One") may affect how worried one is. On the one hand, more knowledge may result in a greater understanding of how earthquakes occur and their effects, so there would be less irrational worry. But on the other hand, learning of the existence of these possible causes may heighten one's worry as they becoming more aware of a realistic cause of the "Big One".

To investigate this question, we decided to visualize the data as two way tables containing the proportions of “Yes” and “No” responses to the question “Do you think the”Big One” will occur in your lifetime?” compared to their knowledge of the fault line and super volcano in @san-andreas-worry and @yellowstone-worry respectively.

```
san_andreas_table <- earthquake %>%
  tabyl(san_andreas_familiar, big_one_occur) %>%
  adorn_totals(where = c("row", "col") ) %>%
  adorn_percentages(denominator = "all") %>%
  adorn_pct_formatting(digits = 2) %>%
  adorn_title(
  placement = "combined",
  row_name = "Knowledge level",
  col_name = "Worry level")

san_andreas_formatNs <- attr(san_andreas_table, "core") %>%
adorn_totals(where = c("row", "col")) %>%
mutate(
across(where(is.numeric), format, big.mark = ",")
)
```

```
Warning: There was 1 warning in `mutate()`.
i In argument: `across(where(is.numeric), format, big.mark = ",")`.
Caused by warning:
! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
Supply arguments directly to `.fns` through an anonymous function instead.

  # Previously
  across(a:b, mean, na.rm = TRUE)

  # Now
  across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
san_andreas_FreqTab <- san_andreas_table %>%
adorn_ns(position = "front", ns = san_andreas_formatNs)

san_andreas_FreqTab %>% kable(digits = c(0, 0, 2, 2, 2))
```

Table 2: Worry of 'Big One' occuring in one's liftime compared to knowledge of the San Andreas fault line

Knowledge level/Worry level	No	Yes	Total
	6 (0.59%)	6 (0.59%)	12 (1.18%)
Extremely familiar	55 (5.43%)	78 (7.70%)	133 (13.13%)
Not at all familiar	72 (7.11%)	34 (3.36%)	106 (10.46%)
Not so familiar	78 (7.70%)	36 (3.55%)	114 (11.25%)
Somewhat familiar	229 (22.61%)	168 (16.58%)	397 (39.19%)
Very familiar	137 (13.52%)	114 (11.25%)	251 (24.78%)
Total	577 (56.96%)	436 (43.04%)	1,013 (100.00%)

```
yellowstone_table <- earthquake %>%
  tabyl(yellowstone_familiar, big_one_occur) %>%
  adorn_totals(where = c("row", "col") ) %>%
  adorn_percentages(denominator = "all") %>%
  adorn_pct_formatting(digits = 2) %>%
  adorn_title(
  placement = "combined",
  row_name = "Knowledge level",
  col_name = "Worry level")

yellowstone_formatNs <- attr(yellowstone_table, "core") %>%
adorn_totals(where = c("row", "col")) %>%
mutate(
across(where(is.numeric), format, big.mark = ",")
)
yellowstone_FreqTab <- yellowstone_table %>%
adorn_ns(position = "front", ns = yellowstone_formatNs)

yellowstone_FreqTab %>% kable(digits = c(0, 0, 2, 2, 2))
```

Table 3: Worry of 'Big One' occuring in one's liftime compared to knowledge of the Yellowstone supervolcano

Knowledge level/Worry level	No	Yes	Total
	6 (0.59%)	6 (0.59%)	12 (1.18%)
Extremely familiar	40 (3.95%)	53 (5.23%)	93 (9.18%)
Not at all familiar	171 (16.88%)	99 (9.77%)	270 (26.65%)
Not so familiar	142 (14.02%)	77 (7.60%)	219 (21.62%)

Knowledge level/Worry level	No	Yes	Total
Somewhat familiar	152 (15.00%)	128 (12.64%)	280 (27.64%)
Very familiar	66 (6.52%)	73 (7.21%)	139 (13.72%)
Total	577 (56.96%)	436 (43.04%)	1,013 (100.00%)

Based on both tables, in most cases, more people believe that the "Big One" is not going to occur in their life time. However, the one exception is that when someone is extremely familiar with either the San Andreas fault line or the Yellowstone supervolcano, they are more likely than not to believe that the "Big One" will be in their lifetime.

This data supports the claim that more knowledge is related to more worry, but it in unclear what the underlying nature of the relationship is. We can't clearly say that more knowledge causes more worry because there are several possible other explanations. For example, the relationship could actually be the opposite such that worry causes one to seek out knowledge rather than knowledge causing one to worry.

Question 3

Question 4

Conclusion