

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.

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.

Research Questions

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

How does worry differ across the United States?

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 Table 1 and Table 2 respectively.

Table 1: Worry of ‘Big One’ occurring in one’s lifetime 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%)

Table 2: Worry of ‘Big One’ occurring in one’s lifetime 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%)
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 is 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.

Does age and income relate to one's worry?

In the previous section, we looked at if there are patterns between one's knowledge about fault lines and super volcanoes and their worry about the “Big One”. In this section we will look at another set of variables that could have a pattern in relation to worry. To be specific, we will be looking at age and income.

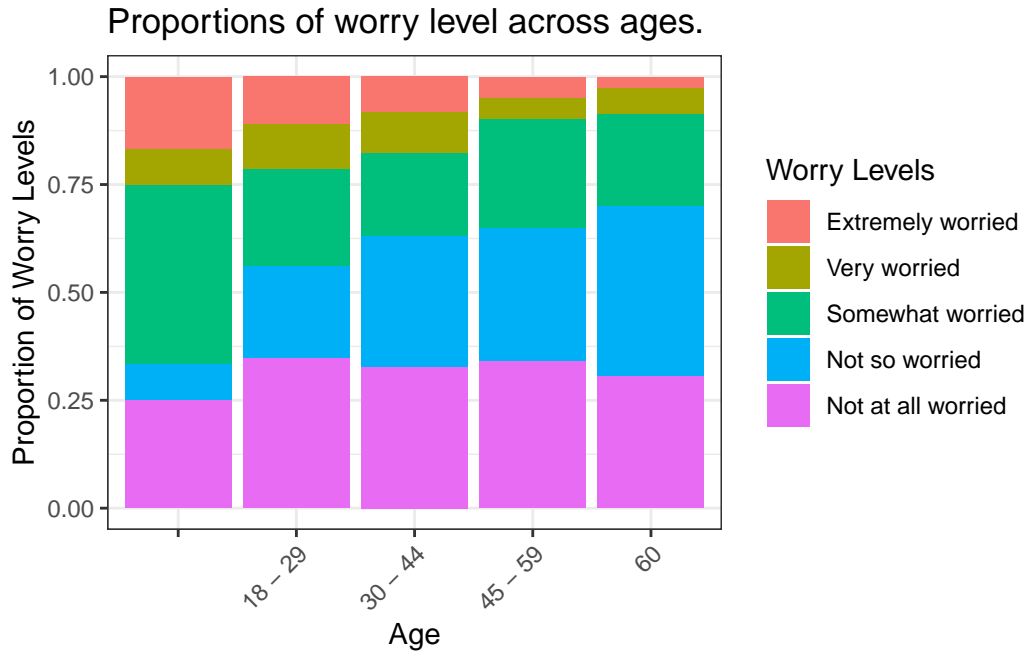


Figure 1: Proportions of worry level across ages.

In Figure 1, there are a few patterns of interest. The first is that the proportion of people who are “Not at all worried” does not appear to change very much with age and stays relatively consistent. This suggests that age is not a factor in one having very low worry levels about the “Big One”.

The second pattern is that the proportion of people that are “Extremely worried” decreases as age increases while the proportion of people that are “Not so worried” increases as age increases. This in some ways makes sense because in order for proportionally less people to have one level of worry, there needs to be proportionally more people to have other levels of worry to account for this change.

Another interesting thing of note in regards to the second pattern is that the “Very worried” and “Somewhat worried” are approximately the same proportion across all age groups. This is likely the result of one of two changes. The first is that the decrease in “Extremely” and increase in “Not so” is a direct shift from the first to the second. The second explanation is that there is a shift where many people decrease their worry across “Extremely”, “Very”,

“Somewhat” and “Not so”, but the proportions of these changes were equivalent such that as a net change it merely appears as a decrease in “Extremely” and an increase in “Not so”. In order to more accurately determine this, we would need to conduct this survey multiple times across the same group of people to see how individual answers change with age.

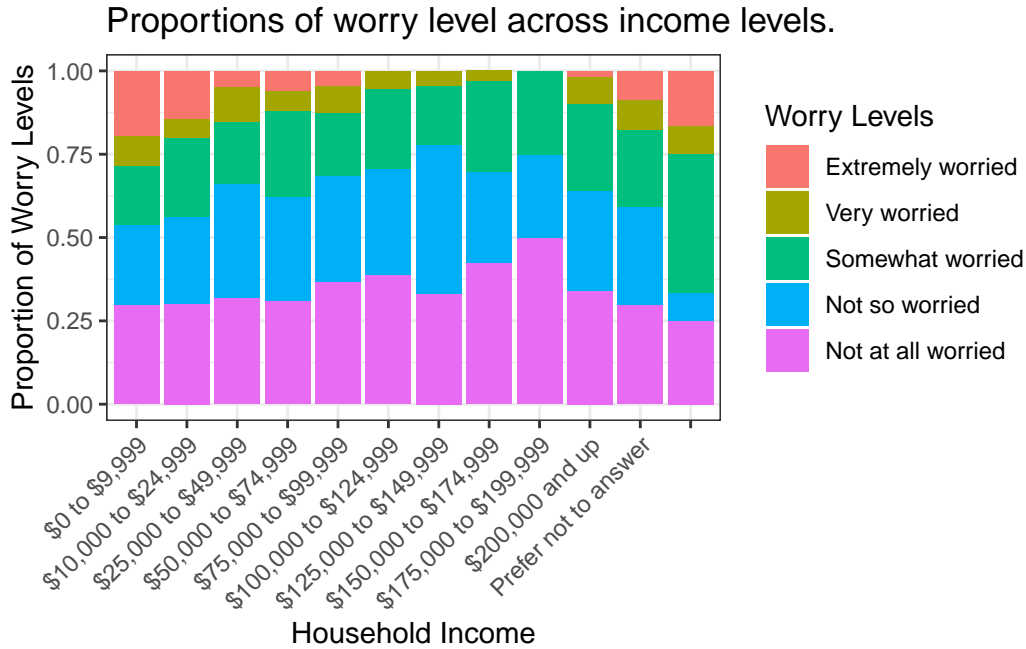


Figure 2: Proportions of worry level across income levels.

As compared to Figure 1, the patterns in Figure 2 are less apparent. One initial pattern that can be identified is that there seems to be an overall decrease in worry as income increases, but the \$200,000 and up income range does not follow this trend. However, this could be the result of outliers due to the fact that this income bracket is uncapped.

Beyond this initial trend, any other conclusions would likely require greater information and more data gathering; however, another possible route for further investigation would be an analysis of how these same demographic attributes affect another element of worriedness about earthquakes. Specifically, we will look at the effects on preparations and precautions in regards to earthquakes.

Who is prepping for earthquakes?

Conclusion

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

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 = ",")
  )
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))

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))
age_worry_proportion <- earthquake %>%
  group_by(age, big_one_worry) %>%
  summarise(count = n()) %>%
  mutate(proportion = count / sum(count))

worry_order <- c("Extremely worried",
                 "Very worried",
                 "Somewhat worried",
                 "Not so worried",
                 "Not at all worried")

age_worry_proportion$big_one_worry <- factor(age_worry_proportion$big_one_worry, levels = worry_order)

ggplot(age_worry_proportion, aes(x = age, y = proportion, fill = big_one_worry)) +
  geom_bar(stat = "identity", position = "stack") +
  labs(x = "Age",
       y = "Proportion of Worry Levels",
       fill = "Worry Levels",
       title = "Proportions of worry level across ages.") +

```

```

theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
income_worry_proportion <- earthquake %>%
  group_by(household_income, big_one_worry) %>%
  summarise(count = n()) %>%
  mutate(proportion = count / sum(count))

worry_order <- c("Extremely worried",
                 "Very worried",
                 "Somewhat worried",
                 "Not so worried",
                 "Not at all worried")

income_worry_proportion$big_one_worry <- factor(income_worry_proportion$big_one_worry, levels = worry_order)

x_order <- c("$0 to $9,999",
             "$10,000 to $24,999",
             "$25,000 to $49,999",
             "$50,000 to $74,999",
             "$75,000 to $99,999",
             "$100,000 to $124,999",
             "$125,000 to $149,999",
             "$150,000 to $174,999",
             "$175,000 to $199,999",
             "$200,000 and up",
             "Prefer not to answer",
             "")

income_worry_proportion$household_income <- factor(income_worry_proportion$household_income, levels = x_order)

ggplot(income_worry_proportion, aes(x = household_income, y = proportion, fill = big_one_worry)) +
  geom_bar(stat = "identity", position = "stack") +
  labs(x = "Household Income",
       y = "Proportion of Worry Levels",
       fill = "Worry Levels",
       title = "Proportions of worry level across income levels.") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```