**Data:** https://raw.githubusercontent.com/mmkenton/stat436/refs/heads/main/bear_attacks.csv
**R Script:** https://raw.githubusercontent.com/mmkenton/stat436/refs/heads/main/hw2.R
**Shiny app:** https://mkenton.shinyapps.io/homework2/

**Problem Formulation**

As someone who spends a lot of time backcountry hiking and camping, I've encountered countless written and verbal warnings to "watch out for bears." I've been very fortunate to have only encountered one bear (grizzly, female) in close proximity while hiking, but I became curious about how common bear attacks actually are. I wanted to know where bear attacks are happening, who is getting attacked, which bears are most likely to attack people, and how bear attack frequency has changed throughout the years.

**Design Choices**

I knew I wanted to start with a map of North America to easily visualize where the attacks were happening. For each attack, I chose to include the specific location and year of the attack, which appears when a user toggles over the data point. Each point is colored by bear species to further convey demographic data in a small area. Next, I wanted to include a visualization centered around the victims themselves, so I created a scatter plot showing the gender of the victim and the year of the attack. I also encoded the size of each data point to correspond to the age of the victim. For user inputs, it seemed intuitive to add a slider for year range and checkboxes for bear species. I also chose to include a table that gives more information about each attack, which also helps relate the visualizations together. When I created the table, I modified the original data frame to exclude columns that were not particularly relevant or interesting.

To create the interface, I took an incremental approach. I started with a skeleton of a Shiny app, focusing on getting the title and supporting text to appear on the screen. Once the app interface was working, I added the slider and checkbox inputs. I wrote code to create simple versions of the graphical outputs and tested them before wrapping them into functions to be called by the server. After that, I added the outputs to the server. At this point, I had skeletons of each of the elements, but they were not connected. I wrote the reactive function so each output (map, graph, and table) would display only the data points selected by the user inputs. Once I had the mechanics all implemented, I spent a while adjusting aesthetics and titles to make the output more visually appealing. I tried to focus on ways to include as much relevant information as possible without it becoming overwhelming or unnecessary.

**Discussion**

Prior to making the visualization, I had a few predictions about what patterns I would see. First, I predicted that bear attacks would be most concentrated in the northern and western areas of the maps. I was partially correct, although many bear attacks occurred in the northeast, which surprised me. In regards to victim demographics, I was expecting there to be a pretty even split

in the gender of bear attack victims. However, I was surprised to see that many more men were victims of bear attacks than women. This could be because the dataset goes back to 1901, when it was less socially acceptable for women to be out hiking or camping alone. I was most surprised by the wide age range of bear attack victims. I expected victims to be in their 20s and 30s, but there were victims as young as less than a year old, and as old as 93.

Another pattern I predicted was that the most common type of bear attack would be a black bear. This prediction turned out to be correct. Before making the visualization, I also thought that bear attacks would have decreased in recent years. However, this surprisingly turned out to be false. The number of bear attacks has stayed fairly constant from 1980 to 2018. However, there is no data for the last 6 years, so it is possible that attacks have decreased during that time.