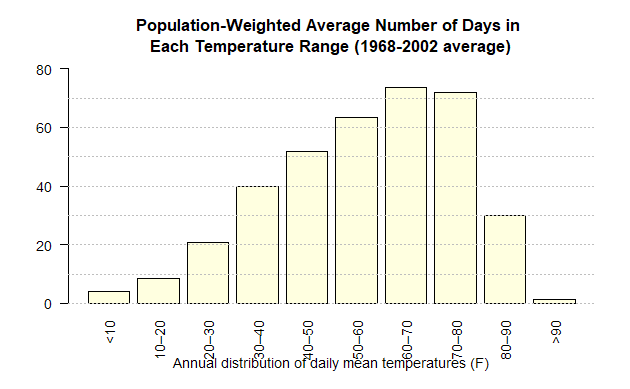
Amartya Sinha  
PPHA 39930  
Assignment 2  
January 31, 2024

**Q1.** Refer to attached working document for code behind all figures and plots

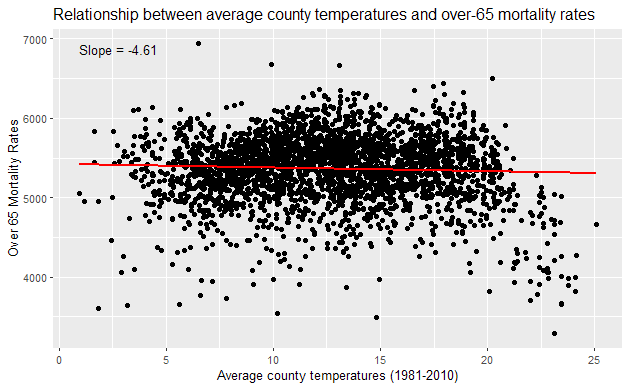
**Q1. a)** 

**Q1. b)** Based on the data provided, we see that the average number of days above 90⁰F per year expected as a population weighted average across the US is just over a day at **1.21 days**.

**Q1. c)** **Yuma county** in Arizona has the highest number of days above 90⁰F, at almost 100 days (99.54 days to be precise). On average over the sample period of 1968-2002 in the dataset, the **number of counties that experienced no days above 90⁰F per year are 115 counties** out of a total of 2988, or about 3.85%.

**Q2.** **a)** As of 2021, the age-adjusted death rate in the United States was 879.7 deaths per 100,000 population. This represents an increase of 5.3% from the previous year, which had a death rate of 835.4 deaths per 100,000 population. [Data is from the CDC](https://www.cdc.gov/nchs/products/databriefs/db456.htm).   
Based on the our data, **the national average over-65 mortality rate is 5365.07 deaths per 100,000 population**. This is **about 6 times higher than the 2021 age-adjusted death rate**.

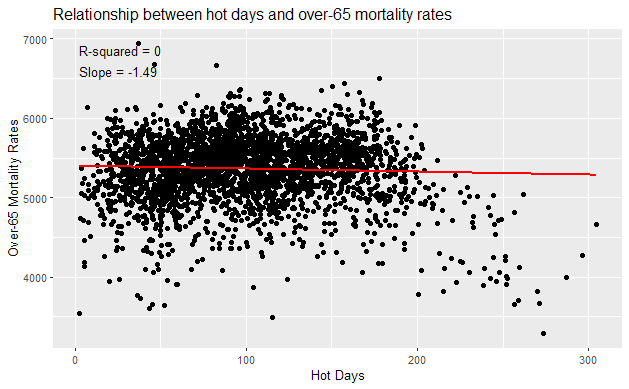
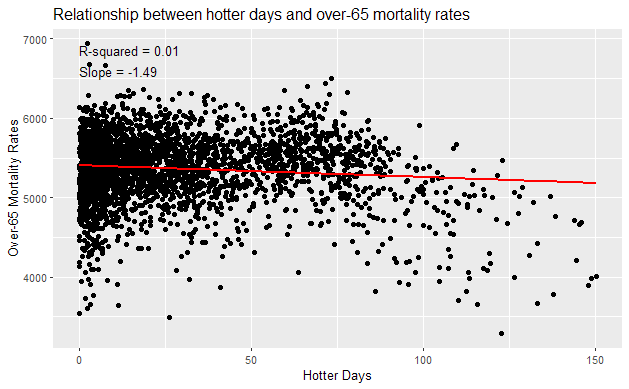
**Q2. b)** The total number of deaths recorded from 1968-2002 in the data is **50,830,306, or about 50.8 million deaths**.

**Q3. a)** 

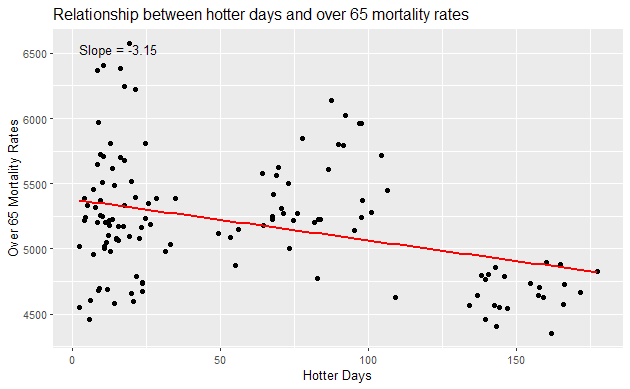
We see that the slope is -4.61. The pattern that emerges is that there is a slight downward trend as the average county temperature increases. This means that with an increase in temperature by 1⁰F, the over-65 mortality rate decreases by 4.61 deaths per 100,000 population.

**Q3. b)** Based on the graphs for hot and hotter days, we see that the mortality rate is negatively correlated with increasing number of both hot and hotter days. However, most of the points on the scatter plot are concentrated around a relatively small set of values. Based on the R-squared values, we see that there is really no correlation in either of the graphs.

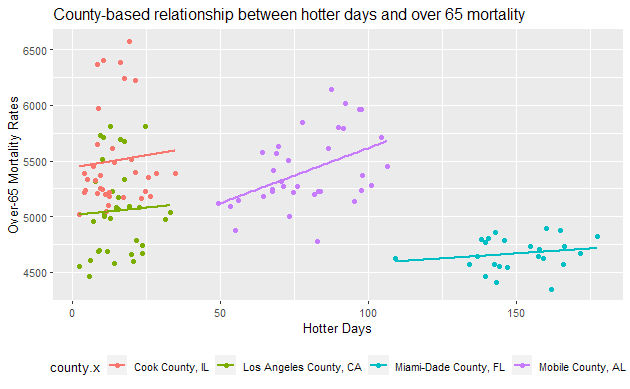
The concentration of scatter points between 0 and 75 days for the hotter days graph shows that most people over 65 live in counties which do not experience temperatures higher than 80⁰F for most of the year. Similarly, the concentration of scatter points between 0 and about 200 shows that out of people aged over 65, most of them live in counties which experience days with temperatures higher than 70⁰F from anywhere between less than a month to most of the year.

**Q3. c)** The patterns observed above might occur because people over 65 might prefer to live in places with more comfortable temperatures. As a result, the number of points in the hotter days graph seems to suggest that fewer people live in places with hotter days than those who live in places with hot days. This is also evident from the scale of X-axis in both graphs, where the range for the hot days graph is more than 300, compared to only around 150 for hotter days. The decrease in over-65 mortality rate in both cases is not really correlated due to the extremely low R-squared value, but the decreasing slope is likely explained by people over 65 preferring to live in places with less extreme temperatures. Hence, the decrease in slope is likely more indicative of the fact that there are fewer numbers of people over-65 per 100,000 in places with extreme temperatures, rather than indicating that hotter temperatures reduce mortality rate.

**Q4. a)** 

The scatter plot shows a negative correlation between the number of hotter days and the over-65 mortality rates in the selected counties, as indicated by the negative slope of -3.15. The relationship one can conclude from this is that as the number of hotter days (temperature hotter than 80°F) increases, the over-65 mortality rate tends to decrease. However, the data points are quite scattered, indicating a high degree of variability and suggesting that other factors may also be influencing the mortality rate. However, without the r-squared value and taking into account other factors, it would be erroneous to make the above conclusion.

**Q4. b)** 

The relationship is much more clear now, since the points on the scatter plot have been grouped by county type. In each county, it is obvious that as the number of hotter days increase, the mortality rate of people over 65 also increases. I would be much more confident in drawing this conclusion from this plot than the conclusion from the plot above because it did not take into account the factor of location when accounting for mortality rate plotted against number of hotter days. In other words, this graph shows us how each county’s over-65 mortality rate is affected by an increase in number of hotter days, rather than lumping together likely disproportionate ratios of population aged over-65 to total population.

**Q5. a)**