Part 2

At first glance, this histogram looks relatively normal and somewhat evenly distributed; but after looking at it for a longer period of time and having time to digest the claims, I am in a better position to discuss what I think. My biggest issues with the claims are: the representative nature of the claim, the data generation process backing up the claims, and using the mean in the claims.

Firstly, I do not really understand the comparison that the author makes. They argue that on average people from this distribution live less than people from the United States. The data comes from World Bank retirees and when they died, but the author is comparing these results to the broader US population. I am not sure that the ages of when retirees from the World Bank died are very representative of the greater US population. The author also cites that people reaching 65 in the United States generally live to the ages of 84.4 and 86.6, as evidence that these results are comparable. However, the author does not discuss if this is the case for their distribution as well. They present a blanket statement, where the evidence provided does not support their claim.

Additionally, I know that the World Bank has a very diverse work force that come from different cultures and societies. Considering that the life expectancy of people in different cultures can vary dramatically, it does not make a lot of sense to generalize these results to another homogenous population (the US population). Furthermore, the data collected was over the period of the previous 6 months. While it is a decent amount of time, it could be said that the short time frame of the data collection process could introduce bias into the analysis and interpretation.

Finally, the author uses the phrase, “wide distribution” in the opening line of the email and then goes on to describe some outliers (retirees who dies in their 50s and 100s) and the average. In general averages can be misleading and outliers tend to affect the average more. Without proper explanation of the data these statistics could be misleading. They could do a better job of talking about the distribution itself, other descriptive statistics (median, interquartile ranges, etc.), and even perhaps the tails of the distribution. My final comment is regarding the bin sizes used in the histogram. Bin selection is very important during this process and I would like to know what that process was like for this histogram.