Uma Vaidyanathan - Technical Exercise: Part 2

The email that was provided attempts to make sense of the numbers provided in the histogram. Unfortunately, it draws some misleading conclusions about the figure including the following:

1.) It confuses a mere difference in means with a statistical difference in means:

The author's conclusion that World Bank retirees must not be living as long as the US population at large does not appear to have been made from a statistical test but rather a verbal one by merely looking at the absolute values of the two means. In other words, this difference could be just due to chance.

2.) It is unclear how representative of the World Bank population the sample cited in the email is:

Population statistics (such as the US average age of death) tend to be based on carefully chosen samples that are representative of the entire population in question. The World Bank sample of retirees on the other hand, is based on data collected in the past 6 years, which may or may not be representative of the larger World Bank population of retirees.

3.) It is unclear whether the US population of retirees is the appropriate comparison group:

World Bank retirees are arguably from more diverse ethnic and racial backgrounds than US retirees. As such, it is unclear whether the US population of retirees may even be the appropriate comparison in the first place.

4.) A deeper dive into reasons for retirement is required:

The author notes that age of death of some of the subjects in the World Bank sample skewed as young as 50. This suggests that some of the people in this sample may have had to, or perhaps voluntarily chosen to retire earlier than the average population (e.g., circumstances such as terminal illness). In fact, a quick look at the distribution of the age of death in the World Bank sample shows that it is slightly left skewed with a small bump at the lower end suggesting perhaps a different set of causes for people who died earlier than 65 than the rest of the sample, which conforms more to a normal distribution.

To enable a more equitable comparison, it might be better to restrict both the World Bank sample and US population to the subjects living at least till the age of 65 before retiring. Further, a deeper dive into specific causes of death at various ages could also shed further light on the issues that the author of the email raises.

In sum, while this is an interesting preliminary dataset, it requires analyses using a much stronger statistical basis, with a deeper understanding of experimental design to make some of the conclusions the author of the email attempts to.