Analytics Foundations: Problem Set 2

1. You've been asked to explore differences between rates of death from Covid in the east vs. the west. You download some data (*covidPerCapita)* from the CDC that has Covid death rates per capita (**covidDeathsPerCapita**) by zip code (**zip**). Not all zip codes are reporting, and even handling the data from those that *do* report is difficult so you select a random sample of zip codes in the east and the west.

    a. Conduct a t-test for the difference of means in per capita deaths from Covid in the east **region** vs the west **region.** Don't forget to verify assumptions. Note anything of interest to you, including the results of any tests performed, violations of assumptions, etc.

    b. Does there appear to be a difference?

    c. Can you conclude from this analysis that individuals in one region have had higher death rates than individuals in the other region? Reflect carefully on the analysis we've done. Discuss any problems you find with this methodology and anything that you'd like to do differently.