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 (**coviddata**) 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. Take a look at the distributions of death from Covid within each region. What do you see? Would you assume approximate Normality? Equal variances?

    b. Conduct the correct test to see if there is a significant difference in per capita deaths from Covid in the east **region** vs the west **region.** Discuss assumptions and why you chose the method you did.

    c. Write out the hypothesis test (include null and alternative hypothesis, test statistic, p-value and conclusion….conclusion should be in terms of the problem).

    d. Can you conclude from this analysis that individuals in one region have had higher death rates than individuals in the other region? Why or why not?