1. **Problem Micromap (Commentary)**

Question: Provide a few sentences of commentary about the micromap to answer ALL of the following questions:

1. Which geographical areas are experiencing the largest increase in death rates over the 5-year period? Why should we care about this?

**The three states with the highest increases in death rates during the previous five years are Louisiana, Tennessee, and Oregon. Death rates rising needs attention. The area where the death rate is highest and the trend should interest us since they may tell why the death rate is rising and suggest potential solutions.**

1. . Which geographical areas are experiencing little to no change in death rates over the 5-year period? Why should we care about this?

**The three states that have seen the least to no change in death rates during the past five years are Montana, Ohio, and Alabama. Similar to the previous answer, knowing which areas have little to no variation in mortality rates may be helpful in figuring out how to keep a death rate stable.**

1. Which geographical areas have a low age-adjusted death rate, but are experiencing a large increase in death rates over the 5-year period? Why should we care about this?

**Three states that have below-median age-adjusted death rates and are seeing a significant rise in death rates are Louisiana, Tennessee, and Oregon. This is the most extreme situation in which the trend deviates significantly from the current pace. Because it is the most alarming death rate trend and needs to be understood, we should be concerned in this instance.**

1. Are certain data elements missing? Why should we care about this?

**Yes, there are missing data elements. The warning that appears when the micromapST result is produced indicates that 'NA' values have been added to the data, although it is not clear which are missing. It appears that several states do not have recorded values for the confidence intervals for the most recent 5-year trend. The entire uteruscancer 1 .csv file can contain other missing values. We should be concerned since a NA value could give a false impression of the data as the ones presenting it. Wyoming, Alaska, and North Dakota in this instance do not have confidence intervals for the most recent values of the 5-year trend**

**Problem 2 (Probability Distributions in R)**

Set your seed (pick a number)

**set.seed(5678)**

1. Create a vector `x' containing 10000 random samples from a chi-square distribution with 10 degrees of freedom (hint:?rchisq)

**x <- rchisq(n=10000, df=10)**

**head(x)**

**#Answer - [1] 7.795379 7.668076 5.437041 15.274104 19.933838 5.029184**

1. Find the value of the chi-square distribution with 10 degrees of freedom corresponding to the 99th percentile (hint: ?qchisq)

**y <- qchisq(p = .99, df = 10)**

**y**

**#Answer - [1] 23.20925**

1. What percentage of observations in your sample fall above this value? Is it exactly 1%? Should it be? Discuss.

**ct <- length(which(x > y))**

**ct #Answer - [1] 102**

**(ct / length(x > y)) \* 100 #Answer - [1] 1.02**

**Since I am using random chisq, there is a chance that the percentile won't be exactly 1%**

**because of the degree of freedom.**

1. Using ggplot, display a density histogram of the vector 'x' (with 50 bins and your choice of colors for the boundaries and fill) and overlay the histogram with the density plot of the vector (size of the curve should greater than or equal to 1.5 and adjust value should be 1.5).

**library(ggplot2)**

**library(dplyr)**

**a1 <- ggplot() +**

**geom\_histogram(aes(x), bins = 50, size = 1.5)**

**fill <- c("blue")**

**a1 <- a1 + labs(x="x Values", y="y values") + ggtitle("Histogram with density plot")**

**a1**

**ggplot() + geom\_histogram(aes(x), bins=50,size=1.5)**