Stat 451

Fall 2020 Homework 1

Due Friday, Sept. 4 by midnight. Problem 1 is to be done by hand (not typed) and problem 2 in R Markdown. Make a pdf scan of problem 1 (either with a scanner or a smartphone app like Genius Scan, Cam Scan, or Tap Scanner; please don't submit a photo). Submit the R Markdown (Rmd) file (not the html output file) for problem 2 in the separate link in Moodle.

1. Using the data in problem 78 of Chapter 2 on p. 61 of the text (marijuana usage among 16-year olds in Europe):

a) Construct a neat handwritten stemplot (with a straight vertical line and uniform sized numerals). Use the tens digit as the stems, but split each stem into two parts (0-4,5-9) (as in the example on p.26 of the text) and put the leaves in order.
b) Compute the five-number summary indicating how you determined the positions of the medians and quartiles (use the formula in the text for the quartiles).
c) Determine whether there are any outliers by the 1.5 IQR criterion. If there are, identify them (by name as well as value).
d) Describe the distribution of the variable in a two or three sentences (see guidelines on pp. 37-8 of the textbook). Be concise and remember context.

2. To be done in RStudio/RMarkdown. Submit only the R Markdown file. See the R Markdown guidelines, particularly item 6. The file GrowthRates.csv contains the estimated population growth rates (in percent) for various time periods for 235 countries and territories (source: United Nations through statisticstimes.com). Consider only the estimated growth rate for 2015 to 2020 (Rate2015to20). Create an R Markdown report with the following in this order. You do not need to label each part.

a) Create a histogram and a density plot and put them side by side (see item 19 of the R Markdown guidelines for how to do this). Include an appropriate label on the x-axis (remember units) for both plots.
b) Describe the distribution of growth rates in two or three sentences. Remember context.
c) Compute and report the median and IQR and the mean and standard deviation of Rate2015to20 using inline R code (item 20 of the R Markdown guidelines), like "The mean growth rate is …" Say which measures of center and spread would be most appropriate (if either) for this variable and why.
d) There is also a categorical variable Continent. Create side-by-side boxplots of growth rate for the 6 continents, including an appropriate label for the y axis. Which continent tended to have the highest growth rates? Which tended to have the lowest? Which continent had the most variability and which the least?