Download file `Babies.data` and store the information in an object named `BABIES` using the function `read.table()`

Description of the variables:

`bwt` : Birth weight in ounces (999 unknown)

`gestation` : Length of pregnancy in days (999 unknown)

`parity` : 0 = first born, 9 = unknown

`age` : Mother's age in years

`height` : Mother's height in inches (99 unknown)

`weight` : Mother's prepregnancy weight in pounds (999 unknown)

`smoke` : Smoking status of mother (0 = not now, 1 = yes now, 9 = unknown)

These data are a subset from a much larger study dealing with child health and development.

a. Create a "clean" data set that removes subjects if any observations on the subject are "unknown". Note that `bwt, gestation, parity, height, weight,` and `smoke` use values of 999, 999, 9, 99, 999, and 9, respectively, to denote "unknown". Store the modified data set in an object name `CLEAN`.
b. Use the information in `CLEAN` to create a density histogram of the birth weights of babies whose mothers have never smoked (`smoke=0`) and another histogram placed directly below the first in the same graphics device for the birth weights of babies whose mothers currently smoke (`smoke=1`). Make the range of the x-axis 30 to 180 (ounces) for both histograms. Superimpose a density curve over each histogram.
c. What is the mean weight difference between babies of smokers and non-smokers? Can you think of any reasons not to use the mean as a measure of center to compare birth weights in this problem?
d. Create side-by-side boxplots to compare the birth weights of babies whose mother's never smoked and those who currently smoke. Use traditional graphics (`boxplot()`) as well as Trellis/lattice graphs to create the boxplot (`bwplot()`).
e. What is the median weight difference between babies who are firstborn and those who are not?
f. Create a single graph of the densities for pre-pregnancy `weight` for mothers who have never smoked and for mothers who currently smoke. Make sure both densities appear on the same graphics device and place a color coded legend in the top right corner of the graph.
g. What is the mean pre-pregnancy weight difference between mothers who do not smoke and those who do? Can you think of any reasons not to use the mean as a measure of center to compare pre-pregnancy weights in this problem?
h. Compute the body weight index (BWI) for each mother in `CLEAN`. Recall that BWI is defined as kg/m^2 (0.0254 m = 1 in., and 0.45359 kg = 1 lb.).Add the variables weight in kg, height in m, and BWI to `CLEAN` and store the result in `CLEANP`.
i. Group pregnant mothers according to their BWI quartile. Find the mean and standard deviation for baby birth weights in each quartile for mothers who have never smoked and those who

currently smoke. Find the median and IQR for baby birth weights in each quartile for mothers who have never smoked and those who currently smoke. Based on your answers, would you characterize birth weight in each group as relative symmetric or skewed?

j. Create side-by-side boxplots of `bwt` based on whether the mother smokes conditioned on BWI quartiles.

k. Does it appear that BWI is related to the birth weight of a baby? Create a scatterplot of birth weight (`bwt`) versus BWI while conditioning on BWI quartiles and whether the mother smokes to help smokes to help answer the question.

l. Replace baby birth weight (`bwt`) with gestation length (`gestation`) and answer questions (i), (j), (k).

m. Create a scatterplot of `bwt` versus `gestation` conditioned on BWI quartiles and whether the mother smokes. Fit straight lines to the data using `lm()`, `lqs()`, and `rlm()`; and display the lines in the scatterplots. What do you find interesting about the resulting graphs?

n. Create a table of smoke by `parity`. Display the numerical results in a graph. What percent of mothers did the smoke during the pregnancy of their first child?