1) In an observational study, the treatment groups should be closely balanced on causally-relevant background variables to mitigate the effect of confounding factors. Identify background characteristics in the data set that might affect gestation length and birth weight. Compare the distributions of those background variables for the smokers and non-smokers. Describe succinctly any *substantial* differences you see between the group of smokers and the group of non-smokers, reporting means and SDs as support for your comparisons:

Which variables did you examine that proved to be similar for both groups?
2) A premature birth is defined as one that occurs before a gestational age of 270 days (about 9 months). Do the data provide evidence that supports the Surgeon General’s assertion that smoking during pregnancy increases the rate of premature delivery? Write a concise answer, reporting what variables and statistics you used to reach your conclusion.

3) The Surgeon General also claims that newborns of mothers who smoke have smaller birth weights at every gestational age (number of days into pregnancy when child is born). Perform a statistical analysis that allows you to answer the question “Do the data support the Surgeon General’s assertion?” Write 2 – 3 sentences explaining your analysis and conclusion, including relevant numerical or graphical evidence. Examine the sensitivity of your conclusion to the effects of outliers, and consider that there may be some gestational ages for which the data don’t provide enough evidence to make claims either way.