"Some students seemed confused by the instruction to 'Make sure that at least two of these questions involve at least three variables.' One of the variables should (most likely) be the response variable, which means the question should ask how the response is affected by two or more of the explanatory variables.(Suggestion: 2 numeric variables, 1 categorical variable)"

"Your research questions should not be one sentence questions like 'How do race and smoking affect the birth weight of a child?'. Your questions should be more specific than this and they should be at least 2-3 sentences long. For example,

"It 's believed that smoking during pregnancy can lead to lower birth weights in infants. We are interested in whether the effect of smoking on birthweight is the same for individuals of different races. If the effect of smoking on birth weight is different across the racial groups in our data, we are interested in exploring the potential causes of these differences."

"Lastly, here is an example of how to do side by side box plot for multiple factors, in this case birth weight by race and smoking."

```
ggplot(aes(y = bwt, x = factor(race), fill = factor(smoke)), data = birthwt) + geom_boxplot()
```