Analytics Foundations: Lab 13

1. Data were collected in an effort to relate the safety of certain vehicles to different aspects of those vehicles. To obtain this data, submit the following code:

```
safety <- read.csv('https://raw.githubusercontent.com/IAA-Faculty/statistical_foundations/master/safety.csv')
```

This dataset has the following variables:

- **Unsafe**: binary safety designation (1 = below average (unsafe), 0 = average or above average (safe))
- **Type**: type of car (Large, Medium, Small, Sport/Utility, Sports)
- **Region**: manufacturing region (Asia, N America)
- **Weight**: integer value for car weight ranging from 1 to 6
- **Size**: size of car corresponding to **Type** (1 = Small/Sports, 2 = Medium, 3 = Large or Sport/Utility)

Do not create a training and test set and just use the whole dataset for the following analysis.

Perform the following analysis:

a. Build a logistic regression, predicting **Unsafe** using the variables **Region, Weight,** and **Size**. Treat **Weight** as a *continuous* variable. Treat **Region** and **Size** as *categorical*. Make sure to use the `factor` function for **Size**.
    a. Which variables were significant at the 0.05 level?
    b. What is the concordance proportion for this model?

b. Remove variables **one at a time** that have a p-value above 0.05.
    a. What variables are left in the model?
    b. What is the proportion of concordance with your final model?
    c. What is the interpretation of the Size variable for comparing categories 1 to 3?