EDA Homework 7

Due: Friday, March 24, 5pm.

In this assignment, we will be investigating the differences between red and white wine. We have a dataset that has chemical measurements of a large number of different red and white wines. The original purpose of this dataset was to build a predictive model of wine quality, but here we will just be interested in the differences between red and white wines. The dataset has a large number of variables aside from wine color, but a best subsets procedure tells us that the best two-predictor model uses `total_sulfur_dioxide` and `density` and so we will focus on those for the analysis.

To get the data, you can install the `ucidata` package

```
install.packages(devtools)
devtools::install_github("coatless/ucidata")
```

Once you have done this, you can load the data using

```
library(ucidata)
data(wine)
```

Assignment:

1. There are a couple of outliers in density, remove them.
2. Make a scatterplot of wine type as a function of total sulfur dioxide. Add a loess smoother and a logistic regression smoother. Comment on similarities or differences.

   Note: To do this, you will need to change wine type from white/red to 0/1.
3. Make a scatterplot of wine type as a function of density. Add a loess smoother and a logistic regression smoother. Comment on similarities or differences.
4. Fit two logistic regression models: one with no interaction between density and total sulfur dioxide and one with an interaction between density and total sulfur dioxide. Make two coplots for each model showing the fits. For each model, you should have one coplot of wine type as a function of density conditioned on total sulfur dioxide, and one with wine type as a function of total sulfur dioxide conditioned on density.
5. Comment briefly on what these plots tell you about the relationships among the three variables.

Submit a pdf with the answers to the questions and an Rmd or R file with the code you used.