EDA Homework 3

Due: Friday, September 14 at 5pm.

Work in groups of up to 3, and make one submission for each group. Submit exactly two files: (i) a PDF/HTML file with your write-up and graphs and (ii) a .r/.txt/.Rmd file with code to reproduce your graphs.

We'll try transfomations and modeling of the food web data, described in Cleveland starting on page 58.

It is available in the `lattice.RData` file, which you can download from
http://ml.stat.purdue.edu/stat695t/R/lattice.RData.

The data set is called `food.web`, and it contains two variables

- `mean.length`, a positive, continuous variable giving the average length of the food web, and
- `dimension`, a categorical variable taking values `Mixed`, `Two`, and `Three`, giving the type of food web.

We will be interested in how `mean.length` varies with `dimension`.

1. Make a Q-normal plot of `mean.length`, faceted out by `dimension`. Are the data well described by a normal distribution?
2. Let's try some transformations of `mean.length`. Make Q-normal plots of log-transformed `mean.length` and power-transformed `mean.length`, with power -1 (so the inverse transformation), faceted out by `dimension`. Do either of these transformations make `mean.length` look more normally distributed?
3. Let's model the inverse of `mean.length` as a function of dimension. Create a linear model with the inverse of `mean.length` as the response (y variable) and `dimension` as the predictor (x variable). Describe the dependence of the inverse of `mean.length` on `dimension`. Make a Q-normal plot of the pooled residuals to check for normality.