AMS 241: Bayesian Nonparametric Methods (Fall 2015)

Homework problem on curve fitting with DP mixture models
(optional, extra credit problem)

Consider the data set `ozone` from the "ElemStatLearn" R package. The data set includes measurements of ozone concentration in parts per billion, wind speed in miles per hour, temperature in degrees Fahrenheit, and radiation in langleys, recorded over 111 days from May to September of 1973 in New York.

Develop a Dirichlet process mixture of multivariate normals model for the joint distribution of the four variables included in this problem. You can consider location mixing only (such that the kernel covariance matrix is diagonal), as well as the general version of the model mixing on both the kernel mean vector and covariance matrix. Implement the model(s) using an appropriate posterior simulation method. Develop inference for bivariate densities for specific pairs of variables, as well as inference for conditional relationships between the variables. You can consider ozone concentration as the response variable, but also explore further conditional relationships of interest.