A data set taken from an unpublished master’s thesis contains data from 39 sections of large highways in the state of Minnesota. The data are in ‘highway.dat’ and a SAS program to read in the data is in ‘highway.sas’. The following variables are included in highway.dat:

1. **CASENUMB** = Case Number (1 to 39)
2. **Y = RATE** = 1973 accident rate per million vehicle miles
3. **X1 = Len** = length of the segment in miles
4. **X2 = ADT** = average daily traffic counts in thousands
5. **X3 = TRKS** = truck volume as a percent of the total volume
6. **X4 = SLIM** = speed limit
7. **X5 = LWID** = lane width in feet
8. **X6 = SHLD** = width in feet of outer shoulder on the roadway
9. **X7 = ITG** = number of freeway-type interchanges per mile in the segment
10. **X8 = SIGS** = number of signalized interchanges per mile in the segment
11. **X9 = ACPT** = number of access points per mile in the segment
12. **X10 = LANE** = total number of lanes of traffic in both directions
13. **X11 = FAI** = 1 if federal aid interstate highway, 0 otherwise
14. **X12 = PA** = 1 if principal arterial highway, 0 otherwise
15. **X13 = MA** = 1 if major arterial highway, 0 otherwise

Write a report that addresses the following:

1. Which of the independent variables seem related to **RATE**?
2. Find a model that is parsimonious in number of variables and yet gives good predictions of **RATE**.
3. Should one work with **RATE** as the response for rather some transformation of **RATE** such as the log or square root? What, if any, would be the advantages of a transformation here?
4. What types of collinearities are found among the data?
5. Are there influential observations and outliers? If so, what effects do they have upon your conclusions?

Your report should be at most eight pages including figures, tables, and computer output. If you include figures, tables, or output be sure to number them (Figure 1, etc) and provide captions. Refer to figures, tables, and output by number in your text.