Stat ST465/665, Project 1

Be sure to label all plots and include your code in the indicated place.

Problems

1. **(12 points)** This project has to do with the data of measured values of environmental variables contained in cutdown_environment.xlsx . The reference is *(A. Rencher and W. Christensen, Methods of Multivariate Analysis, Table 7.2, Wiley, New York, 2012.)*.

   The variables are AAT = avg air temp, AST = avg soil temp, AHC = avg humidity, Wind = total wind, Evap = evaporation

   (a) Create a matrix scatter plot for the data.

   (b) Does the plot indicate a dependency between AAT and AST? If there is a dependency, describe it. If not, explain why.

   (c) Does the plot indicate a dependency between EVAP and AHC? If there is a dependency, describe it. If not, explain why.

   (d) Does the plot indicate a dependency between AAT and WIND? If there is a dependency, describe it. If not, explain why.

   (e) Show your code.

2. **(16 points)** This project involves data related to the yield of a corn field contained in baker_corn_field_data.txt The reference is *(Calvin, T.S., et al, Yield variability in a central Iowa field, Transactions of the ASAE, 40 (1997), 883-889.)*.

   The first two variables (columns) are locations in the 16-hectar Baker corn field, The remaining variables are various iron (Fe), potassium (K), and magnesium (Mg) variables in yield and additional soil quality measurements. Refer to the variables as A, B, C, D, ..., M.

   (a) Using the second column (second location variable B) to group the data, create a 3x3 matrix scatter plot using the data for the variables in the third (C), seventh (G), and eight columns (H) with different color points for the data in the different groups.

   (b) List groups that appear to be relatively more tightly clustered than other groups.

   (c) List groups that appear to be relatively widely dispersed. For each group listed, indicate at least one of the scatter plots to justify the choice.

   (d) Show your code.