EXAMPLE ASSIGNMENT
(For illustration only)

The following illustrative assignment is based on the **California Rainfall Data** discussed in the first lecture. The main purpose of this illustration is to give you some idea of the kind of analysis and presentation that I expect to see. Your submission should be in the form of a short report on the problem, complete with tables and graphics where appropriate. One of the main objectives of this course is to give you experience in presenting analytical results in a clear and coherent manner. You should endeavor to master such skills, since they are bound to serve you well in the future. Don't be alarmed if you do not understand the all details of the questions or the answer. Both will involve methods of analysis that will be presented later in the course. So in reading the example report, concentrate on the **form** of the presentation rather than the specific content. However, it would be useful to look at Section IV in the NOTEBOOK on the class web page. In particular, look at the sections: "Opening ARCMAP" and "Opening JMPIN". These give you general instructions on how to access the software for the class and set up appropriate paths to the class directory inside the software.

Before doing this Assignment, look at the "California Rain" reference in the Reference Materials.

(1) Open ARCMAP, and then open the file **calif_rain.mxb** that appears in the class directory **F:\sys502\arcview\ projects\california**.

(a) Right click on the data frame **Rainfall Levels** and select **Activate** (the title of the data frame should now be **bold**, indicating that it is activated). The colored dots denote rainfall levels in a selection of California cities, and the contoured surface denotes elevation levels. (The names of these cities can be seen by activating the data frame **California Cities**.) Next, right click on the layer, **Calif_Cities**, and open its Attribute Table. Here you will see a number of attributes listed for each city. The main objective of this exercise is to study the relation between Rainfall Levels (PERCIP) and the three attributes ALTITUDE, LATITUDE, and DISTANCE (from the Pacific Coastline).

1. By visually comparing Rainfall Levels with their corresponding Elevation Levels on the map, can you see any sort of relation between these values? Does this relation seem reasonable, given what you know about climate? Be explicit.

2. Next make the same types of comparisons between Rainfall Levels and the two attributes, Latitude and Distance to the Pacific Coast.

(b) Now activate the data frame **California Cities**, and find the cities of **Salinas** and **St. Peidras**. Re-activate the data frame **Rainfall Levels** and examine the above attributes for these two cities. (The numerical values of these attributes for each city can be accessed directly by first clicking the **Identify** icon on the vertical tool bar bordering the map, and then clicking on map location of the city.)

1. Does the lower level of rainfall in Salinas versus St. Piedras seem reasonable, given their relative Altitude, Latitude, and Distance values? Explain.

2. By examining the locations of Salinas and St. Piedras relative to the topography of California shown on the map, can you think of any other factors that might account for the lower rainfall in Salinas? Be explicit.

(2) Next you will analyze these relations statistically by using multiple regression. To do so, leave ARCMAP open, and next open JMPIN. Inside JMPIN open the data file **Calif_rain.jmp** in the class directory **F:\sys502\jmpin**. You will see that this data file looks very much like the Attribute File in ARCMAP (and in fact was imported from ARCMAP).

(a) To regress Rainfall (**Percip**) on the attributes (**Alt**, **Lat**, **Dist**), click **Analyze** $\rightarrow$ **Fit Model**, and in the window that opens set the dependent variable Y to **Percip,** by first clicking on **Percip** in the left column, and then clicking on 'Y'. Similarly, set the independent (explanatory) variables to (**Alt**, **Lat**, **Dist**), by click on these three variables (with **Ctrl** held down) and then clicking 'Add'). Now click 'Run Model'.

1. In the 'Fit Least Squares' window that opens, scroll down to the **Parameter Estimates** table and check the estimated beta coefficient ('Estimate') and P-value ('P>|t|') for each explanatory variable. Do the signs of these coefficients and their associated P-values agree with your expectations as expressed above? Be explicit.

2. Next scroll up to **Summary of Fit** and look at the adjusted R-square value (**RSquare Adj**). What does this tell you about the overall adequacy of this model?

3. To learn more, scroll down to the **Residual-by-Predicted Plot** and observe that there are two rather extreme outliers. By touching the mouse to each, you will see that their row numbers are 19 and 29, which correspond to the cities, **Tule Lake** and **Crescent City**, in the data table.

4. Locate these cities in ARCMAP. Do you see any common features of these two points? Do their values seem reasonable?

(b) To see what happens if we remove these two outliers, click on the row numbers 19 and 29 in the data table (with **Ctrl** held down) and in the Main Menu click **Rows** $\rightarrow$ **Exclude/Unexclude**. You will now see small red markers next to these rows, indicating that they have been temporarily excluded from the data set (they can be added back in by clicking **Rows** $\rightarrow$ **Exclude/Unexclude** once more).

1. Now repeat the above regression analysis with these two data point excluded.

2. By looking at the resulting beta estimates, P-values, adjusted R-square value, and Residual-by-Predicted Plot, what can you conclude about this new regression relative to the one above? Be explicit in your discussion. Don't simply state how the values differ. Try to interpret their meaning.

3. As a final step in this analysis, you will save the regression for the original regression (including the two possible outliers) as a new data set. To do so, right click on the title of the **Parameter Estimates** table and then click **Save Columns** $\rightarrow$ **Residuals**. You will see that a new column has been added to the data table labeled **Residual Percip**.

(3) These regression residuals can be exported back into ARCMAP where they can be analyzed **spatially**. This has already been done. Activate the data frame **Residuals_1** in ARCMAP, open the Attribute Table for the **Residuals** layer, and you will find the appropriate residuals listed as RES_1. (Notice that this data table has only 28 rows, since **Tule Lake** and **Crescent City** have been omitted.) These residual values are now displayed as the colored dots on this map.

(a) To analyze these residuals spatially, first consider the residual for **Salinas**. Is this value explainable in terms of your earlier observations about Salinas? (Remember that a **negative** residual means that the observed rainfall in Salinas is **less** than that predicted by the regression model.)

(b) Next find the three cities Susanville, Bishop, and Daggett and observe that all of their residuals are very negative. Notice also that all of these cities are located on the Eastern slopes of mountains (away from the coast). This suggests that there may be a significant "Rain Shadow" effect that is not accounted for in the above explanatory variables.

1. If you now activate the data frame **Rain Shadow**, your will see that six cities have been selected (on the basis of more detailed topographic data) as possible candidates for Rain Shadow effects (including Salinas as well as the three cities mentioned above). This effect can be incorporated into the regression analysis by adding a 'dummy variable' with value '1' for Rain Shadow cities and '0' elsewhere. This variable, designated as **Shadow**, has already been included in the JMPIN data table.

2. Now re-run your last regression (excluding the two outliers) with **Shadow** added to the list of explanatory variables. By examining the new beta estimates, P-values, adjusted R-square value, and Residual-by-Predicted Plot, what conclusions can you draw about this revised regression?

3. Finally, the residuals for this regression have also been exported to ARCMAP, and can be seen by activating the data frame **Residuals_2** in ARCMAP (where they appear as RES_2 in the Attribute Table for the **Residuals** layer). Compare these spatial residuals with those above and comment on their implications for the final regression analysis.