**Lesson 5 Homework**

**Reminder**: You must recreate the **PG1** library so you can access your practice files..

1)

The **pg1.np\_codelookup** table is primarily used to look up a park name or park code. However, the table also includes columns for the park type and park region. Use this table to analyze the frequency of park types by the various regions.

1. Create a new program. Write a PROC FREQ step to analyze rows from **pg1.np\_codelookup**.
   * Generate a two-way frequency table for **Type** by **Region**.
   * Exclude any park type that contains the word *Other*.
   * The levels with the most rows should come first in the order.
   * Suppress the display of column percentages.
   * Use **Park Types by Region**as the report title.
   * Submit the program and review the results.
2. What are the top three park types based on total frequency count?
3. Modify the PROC FREQ step to make the following changes:  
   1. Limit the park types to the three that were determined in the previous step.
   2. In addition to suppressing the display of column percentages, use the CROSSLIST option to display the table.
   3. Add a frequency plot that groups the bars by the row variable, displays row percentages, and has a horizontal orientation.  
      **Note**: Use SAS documentation to learn how the GROUPBY=, SCALE=, and ORIENT= options can be used to control the appearance of the plot.
   4. Use **Selected Park Types by Region** as the report title.
   5. Submit the program and review the results.
4. Which **Region** has the highest Row Percent value?

**2)**

**Creating an Output Table with Custom Columns**

The **pg1.np\_westweather** table contains weather-related information for four national parks: Death Valley National Park, Grand Canyon National Park, Yellowstone National Park, and Zion National Park. Use the MEANS procedure to analyze the data in this table.

**a.** Create a new program. Write a PROC MEANS step to analyze rows from **pg1.np\_westweather** where values for **Precip** are ***not*** equal to zero. Analyze precipitation amounts grouped by **Name** and **Year**. Create only an output table, named **rainstats**, with columns for the N and SUM statistics. Name the columns **RainDays** and **TotalRain** respectively. Keep only those rows that are the combination of **Year** and **Name**.

* 1. **b.** Write a PROC PRINT step to print the **rainstats** table. Suppress the printing of observation numbers, and display column labels. Display the columns in the following order: **Name**, **Year**, **RainDays**, and **TotalRain**. Label **Name** as **Park Name**, **RainDays** as **Number of Days Raining**, and **TotalRain** as **Total Rain Amount (inches)**. Use **Rain Statistics by Year and Park** as the report title.
  2. **c.** Run the program and review the results.