Creating a Custom Format from a Table

The **pg2.np_species** table provides a detailed species list for selected national parks. Create a format that categorizes park codes into regions (for example, Northeast or Intermountain). Use the **pg2.np_codelookup** table to create a custom format.

1. Open **p204p05.sas** from the **practices** folder. Modify the first DATA step to create the **np_lookup** table that will be used to build a custom format.
    - Add a RETAIN statement to create the **FmtName** column with a value of *$RegLbl*.
    - Add a RENAME= data set option to the SET statement to rename the **ParkCode** column to **Start**.
    - Add conditional statements to create the **Label** column.
        - The **Label** column is equal to the **Region** column unless the region is missing. In that case, the **Label** column is equal to a value of *Unknown*.
    - Add a KEEP statement to include the **Start**, **Label**, and **FmtName** columns.
    - Submit the first DATA step and examine the output data.

```
data np_lookup;
    retain FmtName '$RegLbl';
    set pg2.np_codeLookup(rename=(ParkCode=Start));
    if Region ne ' ' then Label=Region;
    else Label='Unknown';
    keep Start Label FmtName;
run;
```

2. How many columns are in the **np_lookup** table? What is the value of **FmtName** and **Label** in the first row?

    There are three columns in the **np_lookup** table. **FmtName** is *$regLbl* and **Label** is *Southeast* in the first row.

3. Modify the PROC FORMAT step to read in the **np_lookup** table.

```
proc format cntlin=np_lookup;
run;
```

4. Modify the second DATA step.
    - Create a new column named **Region**. Use the PUT function to create the new column based on using the $RegLbl format on the **ParkCode** column.
    - Submit the program and confirm the results in the PROC FREQ output.

```
data np_endanger;
    set pg2.np_species;
    where Conservation_Status='Endangered';
    Region=put(ParkCode,$RegLbl.);
run;
```

5. How many rows have a **Region** value of *Unknown*?

There are four rows with a **Region** value of *Unknown*.