Practice

If you restarted your SAS session, open and submit the **libname.sas** program in the course files.

Level 1

4. Creating a Custom Format from a Table

The **pg2.np_monthlyTraffic** table contains monthly traffic counts at locations in national parks. Create a format that categorizes park codes into their type (for example, National Park, National Seashore, and so on). The **pg2.np_codeLookup** table contains park codes and the associated park types.

a. Open **p204p04.sas** from the **practices** folder. Highlight the PROC MEANS step and run the selected code. Review the output. Notice that the traffic statistics are listed by a four-letter park code.

b. Open the **pg2.np_codeLookup** table. Notice that **ParkCode** contains the four-letter park code and **Type** contains the type of park.

	Name_Code	ParkName	ParkCode	Region	Type
1	Abraham Lincol...	Abraham Lincol...	ABLI	Southeast	National Historical Park
2	Acadia National...	Acadia National...	ACAD	Northeast	National Park
3	Accounting Op...	Accounting Op...	AOC		
4	Adams National...	Adams National...	ADAM	Northeast	National Historical Park
5	Adams National...	Adams National...	ADNM		

c. Modify the DATA step.

1) Add a RENAME= data set option to the SET statement to rename the **ParkCode** column to **Start** and the **Type** column to **Label**.

2) Add a RETAIN statement before the SET statement to create the **FmtName** column with a value of *$TypeFmt* (without a period at the end).

d. In the PROC FORMAT statement, add a CNTLIN= option to build a format from the **type_lookup** table.

e. In the PROC MEANS step, add a FORMAT statement so that the $TypeFmt format is applied to the **ParkCode** column.

f. Run the program and review the results. Verify that the data is grouped by park types.

Traffic Statistics

The MEANS Procedure

Analysis Variable : Count			
Name	**Month**	**Mean**	**Sum**
National Park	1	9136	2220012
	2	10529	2558538
	3	12073	2933676
	4	13316	3235715
	5	16060	3822339
	6	21039	5007282
	7	25274	6015158
	8	22655	5391849
	9	21473	5110551
	10	18656	4440049
	11	12662	3013589
	12	11443	2723550
National Seashore	1	6792	319220

Level 2

5. 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.

a. 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.

1) Add a RETAIN statement to create the **FmtName** column with a value of *$RegLbl*.

2) Add a RENAME= data set option to the SET statement to rename the **ParkCode** column to **Start**.

3) 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*.

4) Add a KEEP statement to include the **Start**, **Label**, and **FmtName** columns.

b. Highlight the first DATA step and run the selected code. Verify the output table.

FmtName	Start	Label
$RegLbl	ABLI	Southeast
$RegLbl	ACAD	Northeast
$RegLbl	AOC	Unknown
$RegLbl	ADAM	Northeast
$RegLbl	ADNM	Unknown

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

d. In 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. Run the program and confirm the results in the PROC FREQ output.

Number of Endangered Species by Region

The FREQ Procedure

Region	Frequency	Percent
Alaska	31	8.29
Intermountain	70	18.72
Midwest	16	4.28
Northeast	7	1.87
Pacific West	183	48.93
Southeast	63	16.84
Unknown	4	1.07

Challenge

6. Updating a Custom Format by Using the CNTLOUT= Option

The **pg2.np_summary** table contains public use statistics from the National Park Service. The values of the **Type** column represent the park type as a code. A format is applied to display descriptive values for the park types.

a. Open **p204p06.sas** from the **practices** folder. Run the program and review the results. Notice that some of the park types are still displayed as codes because the custom format does not include a label for those values.

b. Write a PROC FORMAT step that uses the CNTLOUT= option to create a table named **typfmtout** from the existing $TypCode format. Run the step and view the output table. The **typfmtout** table contains several extra columns, but the critical columns for this practice are **FmtName**, **Start**, and **Label**. Notice that the values for **FmtName** do not include the $ as a prefix.

c. Open the **pg2.np_newcodes** table. Notice that it contains the format name, the **Type** values, and the labels in the **FmtName**, **Start**, and **Label** columns.

d. Write a DATA step that creates a table named **typfmt_update** by concatenating the output table from PROC FORMAT and the **pg2.np_newcodes** table. Change the values of **FmtName** to *$TypCode* and keep only the **FmtName**, **Start**, and **Label** columns.

e. Write a PROC FORMAT that re-creates the $TypCode format using the CNTLIN= option to read the new table that contains the updated format values.

f. Run the PROC FREQ step again and verify that all **Type** codes are displayed with labels.

Park Frequencies by Type

The FREQ Procedure

Type	Frequency	Percent
National Monument	63	46.67
National Park	52	38.52
National Preserve	7	5.19
National Seashore	10	7.41
National River	3	2.22

End of Practices