REPORT WRITING: A CASE STUDY
Ann Lehman, SAS Institute Inc.
Tim Lehman, SAS Institute Inc.

The following paper is a modified version of the case study presented in the SAS® Processing course. This study was designed to illustrate to intermediate level students the wide range of SAS capabilities that may be needed in a report writing assignment. A complex INPUT program is used to read multiple raw files; SAS procedures are used to verify and summarize the data; SAS data management techniques shape a data set to fit a report design; a tailored report is written.

What is The Task?
You have obtained a tape of twenty raw data files. The files are composed of records of daily weather data from all reporting weather stations in the United States for years 1960 to 1979. You have been asked to design a report that presents a concise summary of weather for the state of Washington during the years 1977 to 1979.

Example of Output Report.
Tables have been designed to present average maximum and minimum temperatures for selected months during each year and also overall years. Each reporting station is summarized with a single table with two tables to be printed per page. The pages will be numbered at the bottom.

Output from Report Program
Average MAX and MIN temperatures
State of Washington
1977-1979

<table>
<thead>
<tr>
<th>STATION</th>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAKOMA</td>
<td>1977</td>
<td>44.7</td>
<td>63.2</td>
<td>75.2</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>max</td>
<td>34.1</td>
<td>43.9</td>
<td>55.0</td>
<td>45.4</td>
</tr>
<tr>
<td></td>
<td>min</td>
<td>34.1</td>
<td>43.9</td>
<td>55.0</td>
<td>45.4</td>
</tr>
<tr>
<td></td>
<td>1978</td>
<td>48.7</td>
<td>52.2</td>
<td>76.2</td>
<td>62.1</td>
</tr>
<tr>
<td></td>
<td>max</td>
<td>40.0</td>
<td>42.5</td>
<td>53.4</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td>min</td>
<td>40.0</td>
<td>42.5</td>
<td>53.4</td>
<td>46.5</td>
</tr>
<tr>
<td></td>
<td>1979</td>
<td>43.4</td>
<td>58.7</td>
<td>70.7</td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>max</td>
<td>32.2</td>
<td>42.9</td>
<td>56.1</td>
<td>47.6</td>
</tr>
<tr>
<td></td>
<td>min</td>
<td>32.2</td>
<td>42.9</td>
<td>56.1</td>
<td>47.6</td>
</tr>
</tbody>
</table>

How is The Tape Constructed?
The data tape contains twenty files named US1960, US1961, ..., US1979. Each file contains daily weather data for the United States for one year. One record contains the data for all reporting stations for one day. The number of reporting stations each day varies and is not known but ranges from 225 to 245. Each station reports 21 numeric variables and, if a reporting station has missing values for any variable, the variable is given a missing value code. Each file contains a record for each day of the year beginning with January 1, but the date is on the records. Examination of the documentation reveals that the station code numbers 24281, 24233, 24157, and 94240 identify the stations Seattle, Takoma, Spokane, and Quillayute, respectively. You need to read the weather information for these four stations from three raw files named US1977, US1978, and US1979. The variables to be read are station number, maximum and minimum temperature, rainfall, and snowfall. The files were written using unformatted FORTRAN WRITE statements in 4-byte integer arrays. The stations are recorded in ascending order on each record.

Define the SAS Data Set to Be Created
You have decided to create 8 numeric SAS variables called STATION, DATE, MONTH, YEAR, MAXTEMP, MINTEMP, PRECIP, and SNOW. Space can be saved by reducing the length of the variables STATION, DATE, MONTH, and YEAR to 4 bytes. The approximate space needed for the data set can be computed or derived by running a SAS job that creates these variables with the statement OPTION OBS=O; in effect. Under MVS the SAS log will display the number of observations stored per track. The data set will contain 365 days X 3 years X 4 stations = 4380 observations.

Read the Raw File
Three SAS steps are used to provide formats for the data, read the raw files, and sort and store the resulting data set. The beginning SAS date is created using the MDY function and is incremented each time a new record is read. The LENGTH= option must be used on the INFILE statement. The record length is determined when the first INPUT is executed. The number of reporting stations, NSTA, is computed and used as the stop value on the iterative DO that reads the station codes. Each reporting station in the ACCPSTA format list is located; the SAS observation is created and written to the data set. The variables are labeled, and the data set is sorted and labeled for processing and documentation.

SAS® is the registered trademark of SAS Institute Inc., Cary, NC, USA
**Read The Raw Data**

```
PROC FORMAT DDNAME = SASLIB;
VALUE ACCPSTA 24281 24233, 24157, 94240 = 'YES';
VALUE WA 24281 = 'SPOKANE' 24233 = 'TAKOMA'
24157 = 'SEATTLE' 94240 = 'QUILLAYUTE';
%LET BEGIN = 1977;

DATA WEATHER;
LENGTH STATION DATE YEAR MONTH 4;
ARRAY VARS MAXTEMP MINTEMP PRECIP SNOW;
IF _N_ = 1 THEN DATE MDY(I,I,&BEGIN) 1;
INFILE DATAIN LENGTH = L;
INPUT @;
NSTA L/84;
SKIP = (NSTA 1),', 4;
DO I = 1 TO NSTA*4 BY 4;
  INPUT @I STATION IB4. @;
  IF PUT(STATION,AGCPSTA.) = 'YES' THEN DO;
    INPUT (MAXTEMP MINTEMP PREGIP SNOW) (+SKIP IB4.)
    DO OVER VARS;
    IF VARS >= 9999998 THEN VARS END;
    OUTPUT;
  END;
END;
KEEP STATION--SNOW;
FORMAT DATE DATE DATE7.;
LABEL MAXTEMP 'MAXIMUM DAILY TEMP IN DEG F'
MINTEMP 'MINIMUM DAILY TEMP IN DEG F'
PRECIP 'RAIN IN .001 INCHES'
SNOW 'SNOWFALL IN .01 INCHES';
PROC SORT DATA WEATHER
OUT = SASDATA.WEATHER
(LABEL = WASHINGTON STATE, 1977-1979);
BY STATION;
```

**The Data Set WEATHER is Examined, Verified, and Documented**

```
The Data Set WEATHER Is Examined, Verified, and Documented
```

**Summarize and Reshape**

- **Obtain averages with PROC SUMMARY. Also request the N statistic to check for missing values.**

```
PROC SUMMARY DATA=SASDATA.WEATHER;
   CLASS STATION YEAR MONTH;
   VARIABLES MINTEMP MAXTEMP;
   OUTPUT OUT=SASDATA.SUMMARY
      MEAN=MINTEMP MAXTEMP
      N=NMIN NMAX;
```

- **In order to understand the result of PROC SUMMARY, select one observation for each value of the variable _TYPE_ and examine the CLASS and _TYPE_ variables.**

```
   DATA SHOWTYPE;
   SET SASDATA.SUMMARY;
   BY _TYPE_; IF FIRST._TYPE_; PROC PRINT DATA=SHOWTYPE;
   VAR STATION--TYPE_;
   FORMAT MAXTEMP MINTEMP 5.1;

   OBS STATION YEAR MONTH _TYPE_;
   1 . . 0
   2 . . 1 1
   3 . 1977 2
   4 . 1977 1 3
   5 24157 . . 4
   6 24157 . 1 5
   7 24157 1977 . 6
   8 24157 1977 1 7
```

- **Create two SAS data sets from the type 5 and type 7 observations in the output summary data set. Subset to obtain only the quarterly months January, April, July, and October. Create a new variable called MONTH to have a three-character month value.**

```
   DATA SASDATA.TYPE5(DROP=YEAR) SASDATA.TYPE7;
   SET SASDATA.SUMMARY
   (RENAME=(MONTH=M) DROP=NMIN NMAX _FREQ_);
   IF _TYPE_ = 5 THEN OUTPUT TYPES;
   IF _TYPE_ = 7 THEN OUTPUT TYPE7;
   DROP _TYPE_;
```

- **Transpose the TYPES data set so that the months become variables. Use MONTH as the ID variable to define the new variable names. Station is used as a BY variable. Each level of the BY variable is transposed, and the resulting subsets are concatenated by the TRANSPOSE procedure. The transposed data set contains a variable called _NAME_ that has the old variable names as values. To understand the result of PROC TRANSPOSE, examine both the input and output SAS data sets.**

```
   PROC TRANSPOSE DATA=SASDATA.TYPE5
   DOUT=SASDATA.TPOSE5;
   ID M; BY STATION;
   VAR MAXTEMP MINTEMP;
   PROC TRANSPOSE DATA=SASDATA.TYPE7
   DOUT=SASDATA.TPOSE7;
   ID M; BY STATION YEAR;
   VAR MAXTEMP MINTEMP;
   PROC PRINT DATA=SASDATA.TYPES;
   FORMAT STATION WA. MAXTEMP MINTEMP 3.;
   PROC PRINT DATA=SASDATA.TPOSE5;
   FORMAT STATION WA. JAN--OCT 3.;
```

1177
Observe the two transposed data sets by station. Notice that the year will not be in standard collating sequence. A final DATA step is used to add the total observations to the data set as a variable. This is needed to compute the total pages for the page count on the report. The values of _NAME_ are modified to fit the report structure.

A variable named OFFSET is created for the PUT statement to offset the lines of ‘max’ and ‘min’ values. This final data set has one line for each line in the report.

Interleave the two transposed data sets by station. Notice that the year will not be in standard collating sequence. A final DATA step is used to add the total observations to the data set as a variable. This is needed to compute the total pages for the page count on the report. The values of _NAME_ are modified to fit the report structure. A variable named OFFSET is created for the PUT statement to offset the lines of ‘max’ and ‘min’ values. This final data set has one line for each line in the report.

```
DATA REPORT;
  SET SASDATA.TPOSES SASDATA.TPOSE7;
  BY STATION;
  DATA SASDATA.REPORT1;
    DO I = 1 TO N;
      SET REPORT POINT = I NOBS = N;
      TOTOS = N;
      IF _NAME_ = 'MAXTEMP' THEN DO;
        _NAME_ = 'max';
        OFFSET = 0;
      END;
      ELSE DO;
        _NAME_ = 'min';
        OFFSET = 2;
      END;
      OUTPUT;
    END;
  END;
  STOP;
  /*PLEASE! WHEN YOU USE THE POINT= OPTION, DON'T FORGET TO STOP*/
  PROC PRINT DATA=SASDATA.REPORT1 (OBS=10);
  ID STATION;
```

Final Data Set From Which the Report Is Written

```
STATION YEAR _NAME_ JAN APR JUL OCT TOTOS OFFSET
SPOKANE 1977 max 27 65 80 57 32 0
SPOKANE 1977 min 17 37 54 36 32 2
SPOKANE 1978 max 32 55 81 60 32 0
SPOKANE 1978 min 23 36 75 33 32 2
SPOKANE 1979 max 18 56 85 62 32 0
SPOKANE 1979 min 3 34 56 40 32 2
SPokane 1977 . 26 59 82 60 32 0
SPOKANE 1977 . 14 36 55 36 32 2
TAKOMA 1977 max 45 63 75 59 32 0
TAKOMA 1977 min 34 44 55 45 32 2
```
### Average MAX and MIN Temperatures

**State of Washington**

**1977-1979**

#### STATION: SPOKANE

<table>
<thead>
<tr>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>26.6</td>
<td>65.0</td>
<td>80.5</td>
<td>57.3</td>
</tr>
<tr>
<td></td>
<td>17.2</td>
<td>36.7</td>
<td>53.5</td>
<td>45.7</td>
</tr>
<tr>
<td>1978</td>
<td>32.3</td>
<td>55.2</td>
<td>81.2</td>
<td>60.3</td>
</tr>
<tr>
<td></td>
<td>22.8</td>
<td>35.1</td>
<td>55.3</td>
<td>32.8</td>
</tr>
<tr>
<td>1979</td>
<td>10.3</td>
<td>56.4</td>
<td>84.6</td>
<td>62.2</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>34.6</td>
<td>56.2</td>
<td>40.0</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>25.9</td>
<td>59.8</td>
<td>82.1</td>
<td>59.9</td>
</tr>
<tr>
<td></td>
<td>14.3</td>
<td>35.8</td>
<td>55.0</td>
<td>36.2</td>
</tr>
</tbody>
</table>

#### STATION: TAKOMA

<table>
<thead>
<tr>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>44.7</td>
<td>63.2</td>
<td>75.2</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>34.1</td>
<td>43.9</td>
<td>55.0</td>
<td>45.4</td>
</tr>
<tr>
<td>1978</td>
<td>48.7</td>
<td>57.3</td>
<td>78.2</td>
<td>62.1</td>
</tr>
<tr>
<td></td>
<td>40.0</td>
<td>42.5</td>
<td>55.6</td>
<td>46.5</td>
</tr>
<tr>
<td>1979</td>
<td>43.4</td>
<td>58.7</td>
<td>78.7</td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>32.3</td>
<td>42.9</td>
<td>56.1</td>
<td>47.6</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>45.6</td>
<td>59.7</td>
<td>76.7</td>
<td>60.6</td>
</tr>
<tr>
<td></td>
<td>35.4</td>
<td>43.1</td>
<td>55.5</td>
<td>46.5</td>
</tr>
</tbody>
</table>

#### STATION: SEATTLE

<table>
<thead>
<tr>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>44.5</td>
<td>61.4</td>
<td>71.3</td>
<td>59.5</td>
</tr>
<tr>
<td></td>
<td>34.8</td>
<td>33.1</td>
<td>58.6</td>
<td>46.9</td>
</tr>
<tr>
<td>1978</td>
<td>40.5</td>
<td>58.0</td>
<td>73.9</td>
<td>61.9</td>
</tr>
<tr>
<td></td>
<td>38.6</td>
<td>43.2</td>
<td>56.8</td>
<td>47.8</td>
</tr>
<tr>
<td>1979</td>
<td>42.1</td>
<td>58.3</td>
<td>75.4</td>
<td>61.3</td>
</tr>
<tr>
<td></td>
<td>20.3</td>
<td>63.0</td>
<td>57.2</td>
<td>48.6</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>45.1</td>
<td>59.2</td>
<td>73.5</td>
<td>60.9</td>
</tr>
<tr>
<td></td>
<td>34.6</td>
<td>43.4</td>
<td>56.2</td>
<td>47.1</td>
</tr>
</tbody>
</table>

#### STATION: QUILLAVUCE

<table>
<thead>
<tr>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>48.1</td>
<td>57.1</td>
<td>66.8</td>
<td>59.2</td>
</tr>
<tr>
<td></td>
<td>37.3</td>
<td>36.9</td>
<td>47.3</td>
<td>39.3</td>
</tr>
<tr>
<td>1978</td>
<td>48.5</td>
<td>56.1</td>
<td>69.4</td>
<td>63.0</td>
</tr>
<tr>
<td></td>
<td>37.3</td>
<td>38.9</td>
<td>50.3</td>
<td>40.3</td>
</tr>
<tr>
<td>1979</td>
<td>43.3</td>
<td>56.1</td>
<td>69.5</td>
<td>61.6</td>
</tr>
<tr>
<td></td>
<td>27.3</td>
<td>38.3</td>
<td>50.4</td>
<td>43.3</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>46.6</td>
<td>56.5</td>
<td>68.8</td>
<td>61.3</td>
</tr>
<tr>
<td></td>
<td>31.8</td>
<td>38.0</td>
<td>49.3</td>
<td>40.9</td>
</tr>
</tbody>
</table>

---

### PROC TABULATE as an Alternative

The original SAS data set consisted of three years of daily weather data for four reporting stations. The following execution of the TABULATE procedure on the daily data set produces the same tables as the previous set of programs:

```sas
PROC TABULATE DATA=WEATHER FORMAT=6.1;
CLASS STATION YEAR MONTH;
FORMAT MONTH MFMT. STATION WA.);
VAR MINTEMP MAXTEMP;
TABLE STATION,
YEAR*(MAXTEMP='max' MINTEMP='min') MEAN='MEAN'!
ALL*(MAXTEMP='max' MINTEMP='min' ) MEAN='MEAN', MONTH;
```

---

One of Four Tables from PROC TABULATE

**Average MAX and MIN Temperatures**

**State of Washington**

**1977-1979**

STATION: SPOKANE

<table>
<thead>
<tr>
<th>YEAR</th>
<th>JAN</th>
<th>APR</th>
<th>JUL</th>
<th>OCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>26.8</td>
<td>65.0</td>
<td>80.5</td>
<td>57.3</td>
</tr>
<tr>
<td></td>
<td>17.2</td>
<td>36.7</td>
<td>53.5</td>
<td>45.7</td>
</tr>
<tr>
<td>1978</td>
<td>32.3</td>
<td>55.2</td>
<td>81.2</td>
<td>60.3</td>
</tr>
<tr>
<td></td>
<td>22.8</td>
<td>35.1</td>
<td>55.3</td>
<td>32.8</td>
</tr>
<tr>
<td>1979</td>
<td>10.3</td>
<td>56.4</td>
<td>84.6</td>
<td>62.2</td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td>34.6</td>
<td>56.2</td>
<td>40.0</td>
</tr>
<tr>
<td><strong>ALL</strong></td>
<td>25.9</td>
<td>59.8</td>
<td>82.1</td>
<td>59.9</td>
</tr>
<tr>
<td></td>
<td>14.3</td>
<td>35.8</td>
<td>55.0</td>
<td>36.2</td>
</tr>
</tbody>
</table>