Paper 231-2009

Enhancements to Maps using the Annotate Dataset

David Corliss, Marketing Associates, Detroit, MI

Abstract

The Annotate Dataset in SAS® is employed to enhance maps. Records are created for the display of list of landmarks and railroad routes using latitude and longitude coordinates. Data for Alaska and Hawaii are projected with a separate location for each state and a reduced scale for Alaska, allowing the display of points in the preferred USCOUNTY map format.

Introduction

The annotate data provides a powerful tool for the display of the graphical output from SAS procedures but also presents a number of challenges. This provides working examples of two of the most common enhancements found on maps: landmarks and routes. All the enhancements are plotted by latitude and longitude using the SAS USCOUNTY map dataset. This map dataset supports data for all 50 states by placing Alaska and Hawaii in the lower left corner of the map.

Figure 1
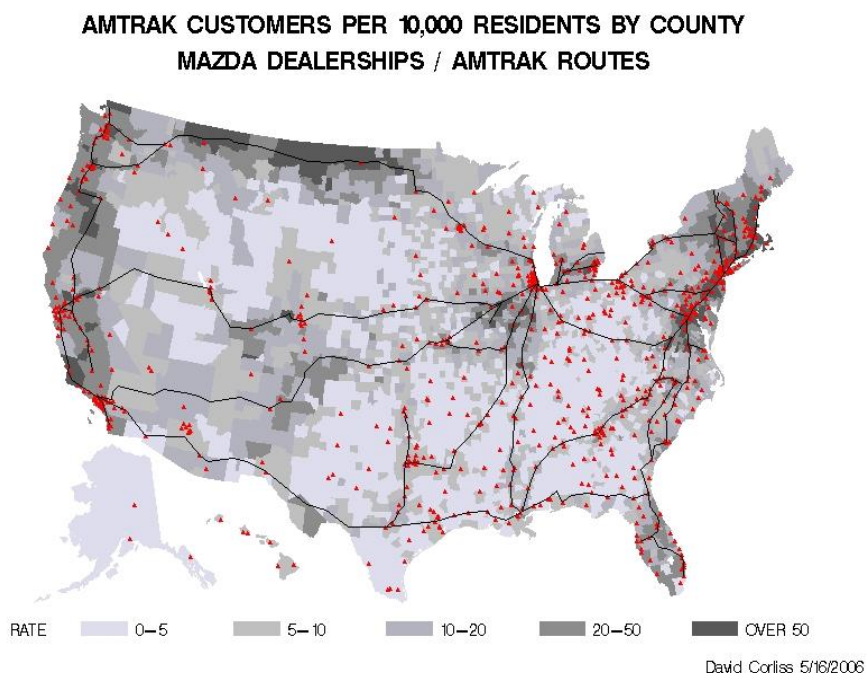
Landmarks

Figure 1 provides the location of all Mazda automobile dealerships in the United States overlaid with AMTRAK train routes. Plotting the dealerships may be accomplished by one of two means. The latitude and longitude of each point can be typed into a card statement. This works well where only a few points are needed. In this situation, with hundreds of dealerships, a more efficient means is preferable. A list of the cities of all the dealerships is read into a SAS dataset and merged with the USCITY map dataset provided by SAS. The latitude and longitude for cities on the AMTRAK routes are found is the same way.

1

```
**** LAT/LONG DATA FOR US CITIES ****;

DATA WORK.CITIES;
   SET MAPS.USCITY END=LAST;

   KEEP X Y CITY LONG LAT CITY_FLAG STATE;
   X=ATAN(1)/45 * LONG;
   Y=ATAN(1)/45 * LAT;

   CITY = UPCASE(CITY);
   CITY_FLAG=1;
RUN;

PROC SORT DATA=WORK.CITIES NODUPKEY;
   BY STATE CITY;
RUN;

DATA WORK.ZIP_CITIES;
   INFILE "C:\FMCC\ZIP_LAT_LONG.TXT"
   DSD DLM='09'X LRECL=80 TRUNCOVER;
   INPUT
      ZIP_CODE        :$5.
      ST_ABBR_CD      :$2.
      LAT             :11.3
      LONG            :11.3
      CITY            :$40.
      STATE_TEXT      :$40.;

   STATE = STFIPS(ST_ABBR_CD);

   X=ATAN(1)/45 * LONG;
   Y=ATAN(1)/45 * LAT;

   CITY = UPCASE(CITY);
   CITY_FLAG=1;

   KEEP X Y CITY LONG LAT CITY_FLAG
   STATE ZIP_CODE;
RUN;

**** MAZDA DEALERSHIPS ****;

DATA WORK.MAZDA;
   INFILE 'C:\MAZDA.TXT' DSD DLM='09'X;
   INPUT
      CITY        :$80.
      ST_ABBR_CD  :$2.;

   STATE = STFIPS(ST_ABBR_CD);
   CITY = UPCASE(CITY);
   LENGTH MAKE $10.;
   MAKE = 'MAZDA';
RUN;


PROC SORT DATA=WORK.MAZDA NODUPKEY;
   BY STATE CITY;
RUN;
```

```
DATA WORK.MAZDA WORK.ALASKA WORK.HAWAII;
   MERGE WORK.MAZDA (IN=A) WORK.CITIES;
   BY STATE CITY;
   IF A=1;

   IF STATE = 2 THEN OUTPUT WORK.ALASKA;
   ELSE IF STATE = 15 THEN
      OUTPUT WORK.HAWAII;
   ELSE IF STATE NOT = . THEN
      OUTPUT WORK.MAZDA;
RUN;

DATA WORK.MAZDA;
   SET WORK.MAZDA;
   DROP STATE ST_ABBR_CD;
RUN;

**** CITIES ON AMTRAK ROUTES ****;

DATA WORK.ROUTES;
   INFILE 'C:\ROUTES.TXT' DSD DLM='09'X;
   INPUT
      ROUTE          :$3.
      CITY           :$80.
      ST_ABBR_CD     :$2.;
   ORDER = _N_;
   STATE = STFIPS(ST_ABBR_CD);
   CITY = UPCASE(CITY);
RUN;

PROC SORT DATA=WORK.ROUTES;
   BY STATE CITY;
RUN;



DATA WORK.ROUTES;
   MERGE WORK.ROUTES (IN=A) WORK.CITIES;
   BY STATE CITY;
   IF A=1;
RUN;

PROC SORT DATA=WORK.ROUTES;
   BY ORDER;
RUN;

DATA WORK.ROUTES;
   SET WORK.ROUTES;
   DROP STATE ST_ABBR_CD;
RUN;
```

Projection

Once every landmark and route point has been assigned a location on the (round) surface of
the earth, PROC GPROJECT is used to calculate the location on a flat map. Alaska and Hawaii
are projected separately. The data for Alaska is set to a smaller scale and moved to the lower
left corner of the map. Hawaii retains the same scale as the main map but is moved to a

position between Alaska and Texas. The landmark data is projected simultaneously with the map boundary data, ensuring they end up together in the end.

```
**** CONSOLIDATE DATA AND PROJECT ***;

DATA WORK.COMBO;
   SET MAPS.STATES(WHERE=(STATE NOT
   IN(2 15))) WORK.MAZDA WORK.ROUTES;
RUN;

PROC GPROJECT DATA=WORK.COMBO
   OUT=WORK.REGION DUPOK;
   ID STATE;
RUN; QUIT;

**** CREATE SEPARATE DATASETS FOR CITY AND GEOGRAPHIC REGION DATA ****;

DATA WORK.REG WORK.CTY WORK.RT;
   SET WORK.REGION;
   LONG = LONG * -1;
   IF CITY_FLAG = 1 THEN OUTPUT WORK.CTY;
   IF ROUTE NOT = '' THEN OUTPUT WORK.RT;
   ELSE OUTPUT WORK.REG;
RUN;

**** PROJECT, SCALE AND TRANSLATE DATA FOR ALASKA  AND HAWAII ****;

DATA WORK.ALASKA;
   SET WORK.ALASKA MAPS.COUNTY;
   WHERE STATE = 2;
RUN;

PROC GPROJECT DATA=WORK.ALASKA OUT=WORK.ALASKA DUPOK PROJECT=LAMBERT
PARALEL1=59 PARALEL2=67;
  ID STATE;
RUN; QUIT;

DATA WORK.ANNO_ALASKA WORK.AKCOUNTY;
   SET WORK.ALASKA;
   X = (X * 0.43) - 0.29;
   Y = (Y * 0.43) - 0.16;
   IF CITY_FLAG = 1 THEN OUTPUT
   WORK.ANNO_ALASKA;
   IF CITY_FLAG NOT = 1 THEN OUTPUT
   WORK.AKCOUNTY;
RUN;

DATA WORK.HAWAII;
   SET WORK.HAWAII MAPS.COUNTY;
   WHERE STATE = 15;
RUN;

PROC GPROJECT DATA=WORK.HAWAII OUT=WORK.HAWAII DUPOK PROJECT=ALBERS
PARALEL1=19 PARALEL2=21;
  ID STATE;
RUN; QUIT;

DATA WORK.ANNO_HAWAII WORK.HICOUNTY;
```

```
   SET WORK.HAWAII;
   X = (X * 1.00) - 0.17;
   Y = (Y * 1.00) - 0.16;
   IF CITY_FLAG = 1 THEN OUTPUT
   WORK.ANNO_HAWAII;
   IF CITY_FLAG NOT = 1 THEN OUTPUT
   WORK.HICOUNTY;
RUN;
```

Description of Markers for Landmarks

The following code specifies all the characteristics of the markers to be placed on the map – shape, size, color, etc. The function variable is used here to specify that a fixed symbol is to be placed on the map.

```
**** ANNOTATE DATA SET ****;

DATA WORK.CTYLBL;
   SET WORK.CTY WORK.ANNO_ALASKA
       WORK.ANNO_HAWAII;

   LENGTH FUNCTION COLOR STYLE $8
          TEXT $25;

   RETAIN XSYS YSYS '2' WHEN 'a';

   IF CITY_FLAG = 1;

   RANDOM = RANUNI(0);
   FUNCTION='SYMBOL';
   STYLE='SPECIAL';
   IF MAKE = 'MAZDA' THEN DO;
      COLOR='RED';
      TEXT='L'; /* SHAPE = STAR */
   END;
   POSITION='5';
   SIZE=0.7;
   OUTPUT;

RUN;


DATA WORK.ANNO;
LENGTH FUNCTION COLOR STYLE $8 TEXT $25;
  RETAIN XSYS YSYS '2' WHEN 'a';
  KEEP FUNCTION COLOR POSITION SIZE
       STYLE TEXT XSYS YSYS WHEN X Y
       MAKE RANDOM STATE;
  SET WORK.CTYLBL;
RUN;

PROC SORT DATA=WORK.CTYLBL;
   BY RANDOM;
RUN;

PROC SORT DATA=WORK.ANNO;
   BY RANDOM;
```

```
RUN;

PROC SORT DATA=WORK.RT;
   BY ROUTE ORDER;
RUN;
```

Routes

In SAS, geographic boundaries are simply lists of consecutive points. One record in the annotate dataset is created for each point, with the stipulation that a line be drawn to the next point. At the end, a command is given to close the polygon by drawing a line segment to the first point on the list. If this last step is omitted, the result is a line connecting the series of points – a route. All of the AMTRAK routes have been placed in a SAS dataset as a succession of cities with a number for each route. Here, the annotate function MOVE is used to move to the first point at the start of a route and DRAW is used to draw lines to all the successive points in a route.

```
**** ANNOTATE DATA FOR THE ROUTES ****;

DATA ANNO_ROUTES;
   LENGTH FUNCTION COLOR $8;
   RETAIN XSYS YSYS '2' WHEN 'a' COLOR
          'BLACK' SIZE 1 XSAVE YSAVE;
   DROP XSAVE YSAVE;

   SET WORK.RT;
      BY ROUTE;

   IF FIRST.ROUTE THEN DO;
      FUNCTION='MOVE';
      XSAVE=X;
      YSAVE=Y;
      OUTPUT;
   END;

   ELSE DO;
      FUNCTION='DRAW';
      OUTPUT;
   END;
RUN;
```

Final Creation of the Map

The annotate data for the landmarks and the routes are combined in to a single dataset. Alaska and Hawaii are deleted from the map boundary data and replaced with the data projected along with the annotate data. PROC GMAP draws on the map boundary data, analytical data for each state and the annotate dataset. Titles are footnotes are assigned and the map is drawn.

```
DATA WORK.ANNO;
   SET WORK.ANNO WORK.ANNO_ROUTES;
RUN;

**** DRAW COUNTY LEVEL MAP ****;

TITLE1  J=C HEIGHT=1.5 FONT=SWISSB
```

```
    'MAZDA CUSTOMERS PER 10,000 RESIDENTS
     BY COUNTY';
TITLE2   J=C HEIGHT=1.5 FONT=SWISSB
    'MAZDA DEALERSHIPS / AMTRAK ROUTES';
FOOTNOTE J=R HEIGHT=1.0 FONT=SWISS
    'David Corliss 5/17/2006';


DATA WORK.USCOUNTY;
    SET MAPS.USCOUNTY;
    IF STATE NOT IN (2,15);
RUN;

DATA WORK.USCOUNTY;
    SET
       WORK.USCOUNTY
       WORK.AKCOUNTY
       WORK.HICOUNTY;
RUN;



PROC GMAP MAP=WORK.USCOUNTY
DATA=WORK.CUSTOMER_COUNTY ANNO=WORK.ANNO;
    ID STATE COUNTY;
    CHORO RATE / DISCRETE ANNO=WORK.ANNO
       COUTLINE=SAME;
    FORMAT RATE RATE.;
RUN; QUIT;
```

Conclusion

While creating a simple map without any annotation is fairly straightforward (for SAS GRAPH), annotation of the map can be extremely complex. These examples of landmarks, routes, and annotating in all 50 states can some provide direction for the most basic and most common map enhancements through the use of the SAS annotate dataset.

Contact Information

The author greatly values your thoughts and input.

David Corliss
Marketing Associates
777 Woodward Avenue
Suite 500
Detroit, MI 48226
Phone: (313) 202 - 6323 dcorliss@marketingassociates.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.