SAS, North American Public Opinion and the 49th Parallel

Timothy B. Gravelle
Vice President, Advanced Analytics Research Strategy Group
Research on Canadian and American public opinion toward the other country is *asymmetrical*.

- Significant research exists on Canadian attitudes toward the U.S. and Canada–U.S. relations.
- Comparatively little research exists on American attitudes toward (or relations with) Canada.
- Existing research pays little attention to *space* – that is, proximity to/distance from the other country.
- There are several public opinion data sources – many publicly available and without cost – that have been under-explored with respect to this topic.
research questions

- What do the Canadian and American publics think about Canada–U.S. relations?

- What roles do political party identification (and ideology) and proximity to the Canada–U.S. border play in shaping attitudes toward the other country? And how do they interact?
The survey data need to be geocoded (latitude–longitude coordinates appended using available geographic indicators).

Distance to the Canada–U.S. border needs to be determined.

The data contain missing values due to non-response; these need to be imputed.

The data were collected using complex sample designs; this needs to be accounted for.

The regression models contain interaction effects involving a non-linear (logged) variable; effect plots are essential to depicting and interpreting them correctly, but ODS does not accommodate these as a matter of course.
key PROCs, functions and programs

- PROC GEOCODE
- GEODIST function
- IveWare (SAS-callable program for multiple imputation)
- PROC SURVEYREG
- PROC SURVEYLOGISTIC
- PROC MIANALYZE
- PROC SGPLOT
“Do you think Canada’s ties with the United States should be much closer, somewhat closer, about the same as now, somewhat more distant, or much more distant?”
“Please indicate the extent to which you would like U.S. ties to be closer or more distant with each of the following countries – Canada.”

- Much closer: 31%
- Somewhat closer: 27%
- About the same as now: 38%
- Somewhat more distant: 1%
- Much more distant: 1%
“Geocoding” refers to the appending of geodetic (latitude–longitude) coordinates using auxiliary geographic information (e.g., street address, postal code/ZIP code, city, area code/exchange).

PROC GEOCODE is a SAS/GRAPH procedure first introduced in SAS 9.2 and provided a number of geocoding methods: ZIP code, ZIP+4, city/state or IP address. Street level (address-based) geocoding was added in the third maintenance release of SAS 9.2 (TS2M3).
Geocoding requires two datasets:

- An input dataset containing geographic data (e.g., address, postal code/ZIP).
- A lookup dataset containing reference geographic data and geodetic (latitude-longitude) coordinates.

The default lookup data set for ZIP and CITY geocoding is SASHELP.ZIPCODE. This dataset is provided with Base SAS and is updated for each SAS release.

U.S. address and ZIP datasets are routinely updated by SAS and are freely available online at support.sas.com (search for “maps online”).
geocoding

- Canadian data: Performed by matching the postal code, forward sortation area (FSA) or federal electoral district (FED) from the survey data to latitude–longitude coordinates created from the Statistics Canada Postal Code Conversion File.

- American data: Performed using PROC GEOCODE (street-level and ZIP code geocoding), with secondary matching using the Area Code World Gold database (telephone area code/exchange).
geocoding (Canada)

```sas
DATA data_08_1a;
MERGE data_08_1 FED2003_geocodes;
BY FED2003;
IF CASE_ID~=.;
IF LAT~=. AND LON~=. THEN GEO_MATCH="FED - 2003 Representation Order - PCCF (March 2009)";
ELSE GEO_MATCH="No match";
RUN;

DATA data_06_1a;
MERGE data_06_1 FSA_06_geocodes;
BY FSA;
IF CASE_ID~=.;
IF LAT~=. AND LON~=. THEN GEO_MATCH="FSA - PCCF (Sept. 2006)"; ELSE GEO_MATCH="No match";
RUN;
```
PROC GEOCODE DATA=data_1 OUT=data_1a
  METHOD=ADDRESS NOZIP NOCITY
  LOOKUPSTREET=geocode.usm
  ADDRESSCITYVAR=CITY
  ADDRESSSTATEVAR=STATE_ABBR
  ADDRESSZIPVAR=ZIP
  ADDRESSVAR=ADDRESS
  ATTRIBUTEVAR=CountyFp;
RUN;

DATA geo_1b (DROP=M_CITY--Y);
SET geo_1a;
IF _MATCHED_="None";
RUN;
PROC GEOCODE DATA=data_1b OUT=data_1c METHOD=ZIP
   LOOKUP=geocode.zipcode_11q1_unique
   ADDRESSCITYVAR=CITY
   ADDRESSSTATEVAR=STATE_ABBR
   ADDRESSZIPVAR=ZIP
   LOOKUPCITYVAR=CITY
   LOOKUPSTATEVAR=STATECODE
   LOOKUPZIPVAR=ZIP
   LOOKUPXVAR=X
   LOOKUPYVAR=Y
   ATTRIBUTEVAR=(STATE COUNTY);
RUN;
calculating border distance/proximity
Geocoding of Canada–U.S. and U.S.–Mexico border crossings was done manually using Google Earth.
Geocoding of Canada–U.S. and U.S.–Mexico border crossings was done manually using Google Earth.
distance to the Canada–U.S. border

- Given two latitude–longitude coordinates, the GEODIST function can calculate the distance between them (accounting for the curvature of the earth).
- The input coordinates can be in decimal degrees or radians (decimal degrees are the default).
- The distance output can be in kilometres or miles (kilometres are the default).
- Example:
  \[ \text{DIST1}=\text{GEODIST}(\text{lat1}, \text{lon1}, \text{lat2}, \text{lon2}, 'K') \]
distance to the Canada–U.S. border

/* Macro for calculating geodetic distances */
%MACRO DISTANCE (datain=, nbr=, lat1=, long1=, 
                   lat2=, long2=);
   DATA &datain;
   SET &datain;
   DIST&nbr=GEODIST(&lat1, &lon1, &lat2, &lon2);
RUN;
%MEND DISTANCE;
distance to the Canada–U.S. border

\%DISTANCE (datain=data_2a, nbr=1, lat1=LAT, lon1=LON, lat2=62.616460, lon2=\-141.004228);
\%DISTANCE (datain=data_2a, nbr=2, lat1=LAT, lon1=LON, lat2=64.086649, lon2=\-141.001485);
\%DISTANCE (datain=data_2a, nbr=3, lat1=LAT, lon1=LON, lat2=59.451303, lon2=\-136.359501);

[...]

\%DISTANCE (datain=data_2a, nbr=125, lat1=LAT, lon1=LON, lat2=43.833677, lon2=\-66.123543);
distance to the Canada–U.S. border

```sas
DATA data_2b;
SET data_2a (DROP=DIST6 DIST13 DIST 14 DIST125);
DISTANCE_CAN_US_BORDER=MIN(of DIST:);
LN_DISTANCE_USA=LOG(DISTANCE_CAN_US_BORDER);
RUN;
```
distance to the Canada–U.S. border

- With Canadian data, distance to the border is highly right-skewed, but *logged* distance to the border is approximately normally distributed.

Source: Canadian Election Studies, 1997–2008
distance to the Canada–U.S. border

- With American data, logging distance to the border introduces some left-skew, but this is still preferred on substantive theoretical grounds.

Source: Gallup Poll World Affairs studies, 2001–2011
PROC SURVEYLOGISTIC DATA=data_7;
  BY _Imputation_;
  MODEL CANADA_TIES_US=
    MALE LN_AGE EDU_UNIV EDU_COLLEGE
    INCOME_LT_20K INCOME_20_40K INCOME_60_80K
    INCOME_80_100K INCOME_100K_PL
    PROVINCE_NL PROVINCE_NS PROVINCE_PE PROVINCE_NB
    PROVINCE_QC PROVINCE_MB PROVINCE_SK PROVINCE_AB
    PROVINCE_BC
    INTEREST_POLITICS ECONOMY_BETTER ECONOMY_WORSE
    INTLTRADE_CR_JOBS
regression model specification (CES)

```plaintext
POST_PARTYCONS POST_PARTY_NDP POST_PARTY_BQ
POST_PARTY_OTHER POST_NO_PARTY LEFT_RIGHT
LN_DISTANCE_USA
LN_DIST_USA_CONS LN_DIST_USA_NDP LN_DIST_USA_BQ
LN_DIST_USA_OTH_PARTY LN_DIST_USA_NO_PARTY
LN_DIST_USA_LEFT_RIGHT
/LINK=CLOGIT RRSQUARE;
STRATA YEAR PROVINCE;
WEIGHT WEIGHT_RENORM;
ODS OUTPUT NObs=n;
ODS OUTPUT ParameterEstimates=ParmEst_1;
ODS OUTPUT GlobalTests=ChiSqDF;
ODS OUTPUT OddsRatios=ORs;
ODS OUTPUT FitStatistics=ModelFit;
ODS OUTPUT RRSquare=RSq;
RUN;
```
### results (CES)

The MIANALYZE Procedure

| Parameter                | Estimate | Std Error | Pr > |t| |
|--------------------------|----------|-----------|------|---|
| Intercept_1              | -2.295   | 0.073     | <.0001 |
| Intercept_2              | -0.991   | 0.071     | <.0001 |
| Intercept_3              | 1.355    | 0.071     | <.0001 |
| POST_PARTY_CONS          | 0.382    | 0.052     | <.0001 |
| POST_PARTY_NDP           | -0.479   | 0.067     | <.0001 |
| POST_PARTY_BQ            | -0.395   | 0.081     | <.0001 |
| POST_PARTY_OTHER         | -0.340   | 0.150     | 0.0245 |
| POST_NO_PARTY            | 0.039    | 0.059     | 0.5118 |
| LEFT_RIGHT               | 0.074    | 0.013     | <.0001 |
| LN_DISTANCE_USA          | 0.026    | 0.032     | 0.4279 |
| LN_DIST_USA_CONS         | -0.088   | 0.042     | 0.0342 |
| LN_DIST_USA_NDP          | 0.086    | 0.057     | 0.1300 |
| LN_DIST_USA_BQ           | -0.072   | 0.088     | 0.4173 |
| LN_DIST_USA_OTH_PTY      | 0.316    | 0.150     | 0.0354 |
| LN_DIST_USA_NO_PTY       | -0.112   | 0.049     | 0.0219 |
| LN_DIST_USA_L_R          | -0.004   | 0.012     | 0.7092 |
plotting interaction effects (CES)

DATA plot_0 (DROP=i);
DISTANCE_CAN_US_BORDER=0.1; OUTPUT;
DISTANCE_CAN_US_BORDER=0.5; OUTPUT;
DO i=1 TO 2500;
DISTANCE_CAN_US_BORDER=i; OUTPUT; END;
RUN;
plotting interaction effects (CES)

```sas
DATA plot_1;
RETAIN ID;
SET plot_0;
ID=_n_;  
LN_MEAN_DIST=4.65307;
LN_DISTANCE_USA=(LOG(DISTANCE_CAN_US_BORDER));
LN_DISTANCE_USA_CTR=
   (LOG(DISTANCE_CAN_US_BORDER)) - LN_MEAN_DIST;
```
/* P(Y <=1), cumulative probability for "Much Closer" */
/* Liberals */
REG_EQN_1_LIB=
-2.295142 + /* Intercept_1 */
  0.381830 * 0 + /* POST_PARTY_CONS */
-0.478545 * 0 + /* POST_PARTY_NDP */
-0.395198 * 0 + /* POST_PARTY_BQ */
-0.340100 * 0 + /* POST_PARTY_OTHER */
  0.038511 * 0 + /* POST_NO_PARTY */
  0.025694 * LN_DISTANCE_USA_CTR + /* DISTANCE USA*/
-0.088088 * 0 * LN_DISTANCE_USA_CTR + /* CONS */
  0.086099 * 0 * LN_DISTANCE_USA_CTR + /* NDP */
-0.071803 * 0 * LN_DISTANCE_USA_CTR + /* BQ */
  0.316425 * 0 * LN_DISTANCE_USA_CTR + /* OTH */
-0.112061 * 0 * LN_DISTANCE_USA_CTR; /* NO PARTY */
CP_1_LIB=CDF('LOGISTIC',REG_EQN_1_LIB);

plotting interaction effects (CES)
plotting interaction effects (CES)

/* P(Y <=1), cumulative probability for "Much Closer" */
/* Conservatives*/
REG_EQN_1_LIB=
  -2.295142 + /* Intercept_1 */
  0.381830 * 1 + /* POST_PARTY_CONS */
-0.478545 * 0 + /* POST_PARTY_NDP */
-0.395198 * 0 + /* POST_PARTY_BQ */
-0.340100 * 0 + /* POST_PARTY_OTHER */
  0.038511 * 0 + /* POST_NO_PARTY */
  0.025694 * LN_DISTANCE_USA_CTR + /* DISTANCE USA*/
-0.088088 * 1 * LN_DISTANCE_USA_CTR + /* CONS */
  0.086099 * 0 * LN_DISTANCE_USA_CTR + /* NDP */
-0.071803 * 0 * LN_DISTANCE_USA_CTR + /* BQ */
  0.316425 * 0 * LN_DISTANCE_USA_CTR + /* OTH */
-0.112061 * 0 * LN_DISTANCE_USA_CTR; /* NO PARTY */
CP_1_CONS=CDF('LOGISTIC',REG_EQN_1_LIB);
plotting interaction effects (CES)

[...And repeat for the NDP, Bloc Québécois; and also repeat for the cumulative probabilities of the other response categories for all parties, changing the Intercept.]

RUN;
plotting interaction effects (CES)

PROC FORMAT;
VALUE LN_DISTANCE_USA_CTR
   -4.65307 = "1"
   -3.95992 = "2"
   -3.04363 = "5"
   -2.35048 = "10"
   -1.65734 = "20"
   -0.74105 = "50"
   -0.04790 = "100"
    0.64525 = "200"
    1.56154 = "500"
    2.25469 = "1000"
    2.94783 = "2000"
    3.17098 = "2500"
;
RUN;
plotting interaction effects (CES)

ods graphics on /border=off height=3.5in width=4.5in;
ods listing image_dpi=600 style=journal sge=off
   gpath="c:\documents\can-us public opinion\canada election studies";

proc sgpplot data=plot_1;
series y=cp_1_lib x=ln_distance_usa_ctr
   /lineattrs=(thickness=1 pattern=solid color=black)
   legendlabel="Liberal, P(Y) <= 1";
series y=cp_2_lib x=ln_distance_usa_ctr
   /lineattrs=(thickness=1 pattern=dash color=black)
   legendlabel="Liberal, P(Y) <= 2";
series y=cp_3_lib x=ln_distance_usa_ctr
   /lineattrs=(thickness=1 pattern=dot color=black)
   legendlabel="Liberal, P(Y) <= 3";

plotting interaction effects (CES)

KEYLEGEND /POSITION=BOTTOM LOCATION=OUTSIDE
   ACROSS=1 DOWN=3;
YAXIS MIN=0 MAX=1 LABEL="Cumulative Probability";
XAXIS MIN=-5.34622 /* 0.5 */ MAX= 3.17098 /* 2500 */
   VALUES=(-3.95992 -3.04363 -2.35048 -1.65734
   -0.74105 -0.04790 0.645250 1.56154 2.25469
   2.94783)
   LABEL="Distance from Canada-U.S. Border (km)";
FORMAT LN_DISTANCE_USA_CTR LN_DISTANCE_USA_CTR .;
RUN;
ODS GRAPHICS OFF;

[...And repeat for the Conservatives, NDP and Bloc
Québécois plots, changing the Y-axis variables and
the legend labels.]
effect plots (CES)
PROC SURVEYLOGISTIC DATA=data_7;
   BY _Imputation_ ;
   MODEL US_TIES_CANADA=
      MALE LN_AGE EDU_SOME_COLLEGE EDU_COLLEGE
      INCOME_LT_35K INCOME_35_50K INCOME_75_100K
      INCOME_100_150K INCOME_150K_PL
      RACE_BLACK RACE_OTHER HISPANIC
      DIVISION_NE DIVISION_MA DIVISION_ENC DIVISION_WNC
      DIVISION_ESC DIVISION_WSC DIVISION_MTN
      DIVISION_PAC
      SAT_US_ECON SAT_AREA_ECON FOLLOW_NEWS_IR
      PARTY_REPUBLICAN PARTY_INDEPENDENT IDEOLOGY
      LN_DISTANCE_CAN
      LN_DIST_CAN_REP LN_DIST_CAN_IND LN_DIST_CAN_IDEOL
   /LINK=CLOGIT RSQUARE;
   OUTPUT OUT=data_1c PREDPROBS=I;
regression model specification (Gallup Panel)

```
STRATA STATE_GEOCODE;
CLUSTER DEMO_RECRUIT_CASE_ID;
WEIGHT WEIGHT_RENORM;
ODS OUTPUT NObs=n;
ODS OUTPUT ParameterEstimates=ParmEst_1;
ODS OUTPUT GlobalTests=ChiSqDF;
ODS OUTPUT OddsRatios=ORs;
ODS OUTPUT FitStatistics=ModelFit;
ODS OUTPUT RSquare=RSq;
RUN;
```
# results (Gallup Panel)

The MIANALYZE Procedure

| Parameter                  | Estimate | Std Error | Pr > |t| |
|----------------------------|----------|-----------|------|---|
| Intercept_1                | -1.104   | 0.087     | < .0001 |
| Intercept_2                | 0.047    | 0.086     | 0.5804 |
| PARTY_REPUBLICAN           | -0.093   | 0.060     | 0.1207 |
| PARTY_INDEPENDENT          | -0.039   | 0.083     | 0.6356 |
| IDEOLOGY                  | 0.124    | 0.028     | < .0001 |
| LN_DISTANCE_CAN           | -0.002   | 0.044     | 0.9667 |
| LN_DIST_CAN_REP           | -0.017   | 0.065     | 0.7970 |
| LN_DIST_CAN_IND           | 0.006    | 0.091     | 0.9470 |
| LN_DIST_CAN_IDEOL         | 0.060    | 0.030     | 0.0459 |
plotting interaction effects (Gallup Panel)

```plaintext
DATA plot_0 (DROP=i);
DISTANCE_CAN_US_BORDER=0.1; OUTPUT;
DISTANCE_CAN_US_BORDER=0.5; OUTPUT;
DO i=1 TO 2500;
DISTANCE_CAN_US_BORDER=i; OUTPUT; END;
RUN;
```
plotting interaction effects (Gallup Panel)

DATA plot_1;
RETAIN ID;
SET plot_0;
    ID=_n_;  
LN_MEAN_DIST=6.44027;
LN_DISTANCE_CAN=(LOG(DISTANCE_CAN_US_BORDER));
LN_DISTANCE_CAN_CTR=(LOG(DISTANCE_CAN_US_BORDER))-LN_MEAN_DIST;

/* P(Y <=1), cumulative probability for "Much Closer" */
/* Liberals */
REG_EQN_1_LIBERAL=
    -1.10427 /* Intercept_1 */
    ( 0.12407 * (1-2.78125)) /* IDEOLOGY */
    (-0.00184 * LN_DISTANCE_CAN_CTR) /* DISTANCE CAN */
    ( 0.05955 * ((1-2.78125) * LN_DISTANCE_CAN_CTR));
    /* IDEOLOGY * LN_DISTANCE_CAN */
CP_1_LIBERAL=CDF('LOGISTIC',REG_EQN_1_LIBERAL);
/* Moderates */
REG_EQN_1_MODERATE=
-1.10427 + /* Intercept_1 */
( 0.12407 * (3-2.78125)) + /* IDEOLOGY */
(-0.00184 * LN_DISTANCE_CAN_CTR + /* DISTANCE CAN */
( 0.05955 * ((3-2.78125) * LN_DISTANCE_CAN_CTR));
/* IDEOLOGY * LN_DISTANCE_CAN */
CP_1_MODERATE=CDF('LOGISTIC',REG_EQN_1_MODERATE);
plotting interaction effects (Gallup Panel)

/* Conservatives*/
REG_EQN_1_CONSERVATIVE=
-1.10427 + /* Intercept_1 */
(0.12407 * (5-2.78125)) + /* IDEOLOGY */
(-0.00184 * LN_DISTANCE_CAN_CTR + /* DISTANCE CAN */
(0.05955 * ((5-2.78125) * LN_DISTANCE_CAN_CTR));
/* IDEOLOGY * LN_DISTANCE_CAN */
CP_1_CONSERVATIVE=CDF('LOGISTIC',REG_EQN_1_CONSERVATIVE);

[...And repeat for the cumulative probabilities of the other response categories for all ideologies, changing the intercept.]

RUN;
plotting interaction effects (Gallup Panel)

PROC FORMAT;
VALUE LN DISTANCE CAN CTR
-6.44027 = "1"
-5.74712 = "2"
-4.83083 = "5"
-4.13767 = "10"
-3.44454 = "20"
-2.52825 = "50"
-1.83510 = "100"
-1.14195 = "200"
-0.22566 = "500"
 0.46748 = "1000"
 1.16063 = "2000"
 1.38377 = "2500"
;
RUN;
plotting interaction effects (Gallup Panel)

```plaintext
ODS GRAPHICS ON /BORDER=OFF HEIGHT=3.5IN WIDTH=4.5IN;
ODS LISTING IMAGE_DPI=600 STYLE=JOURNAL SGE=OFF
  GPATH="C:\Documents\CAN-US public opinion\Canada
  Election Studies"

PROC SGPLOT DATA=plot_1;
SERIES Y=CP_1_LIBERAL X=LN_DISTANCE_CAN_CTR
  /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=BLACK)
  LEGENDLABEL="Liberal, P(Y) <= 1"
SERIES Y=CP_2_LIBERAL X=LN_DISTANCE_CAN_CTR
  /LINEATTRS=(THICKNESS=1 PATTERN=DASH COLOR=BLACK)
  LEGENDLABEL="Liberal, P(Y) <= 2"
SERIES Y=CP_1_MODERATE X=LN_DISTANCE_CAN_CTR
  /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=DARKGRAY)
  LEGENDLABEL="Moderate, P(Y) <= 1"
SERIES Y=CP_2_MODERATE X=LN_DISTANCE_CAN_CTR
  /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=DARKGRAY)
  LEGENDLABEL="Moderate, P(Y) <= 2"
```
SERIES Y=CP_1_CONSERVATIVE X=LN_DISTANCE_CAN_CTR /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=LIGHTGRAY) LEGENDLABEL="Conservative, P(Y) <= 1";
SERIES Y=CP_2_CONSERVATIVE X=LN_DISTANCE_CAN_CTR /LINEATTRS=(THICKNESS=1 PATTERN=SOLID COLOR=LIGHTGRAY) LEGENDLABEL="Conservative, P(Y) <= 2";
KEYLEGEND /POSITION=BOTTOM LOCATION=OUTSIDE ACROSS=2 DOWN=3;
YAXIS MIN=0 MAX=0.8 LABEL="Cumulative Probability";
XAXIS MIN=-7.13342 /* 0.5 */ MAX=1.38377 /* 2500 */ VALUES=(-5.74712 -4.83083 -4.13767 -3.44454 -2.52825 -1.83510 -1.14195 -0.22566 0.46748 1.16063) LABEL="Distance from Canada-U.S. Border (km)";
FORMAT LN_DISTANCE_CAN_CTR LN_DISTANCE_CAN_CTR.
RUN;
ODS GRAPHICS OFF;
results (Gallup Panel)
conclusions

- In both countries, general attitudes toward Canada–U.S. relations are shaped by political factors (party identification or ideology).
  - But border proximity has no main effect.
  - Rather, the effects of party ID and ideology are contingent upon (moderated by) proximity to the border.
- SAS has very good built-in tools for geocoding and calculating geodetic distances that can serve as a practical alternative to a full-blown GIS.
- Effect plots created by ODS are useful, but aren’t always available or “camera-ready.” So it pays to really know the underlying regression equations so you can create plots that meet your exact requirements.
conclusions

- Survey data (and our respondents) come from *somewhere*. In studying public opinion and policy attitudes, we ignore this at our peril. Fortunately, SAS provides the tools to ensure we don’t.


Thank you!

Timothy B. Gravelle
Vice President, Advanced Analytics
Research Strategy Group

111 Queen Street East
Toronto, ON M5C 1S1

tgravelle@rsginc.net

647-426-6241