

KML Macro:

Integrating SAS® and Google® API

and Its Application in

Mapping Manitoba's Health Data

on Google Earth and Google Map

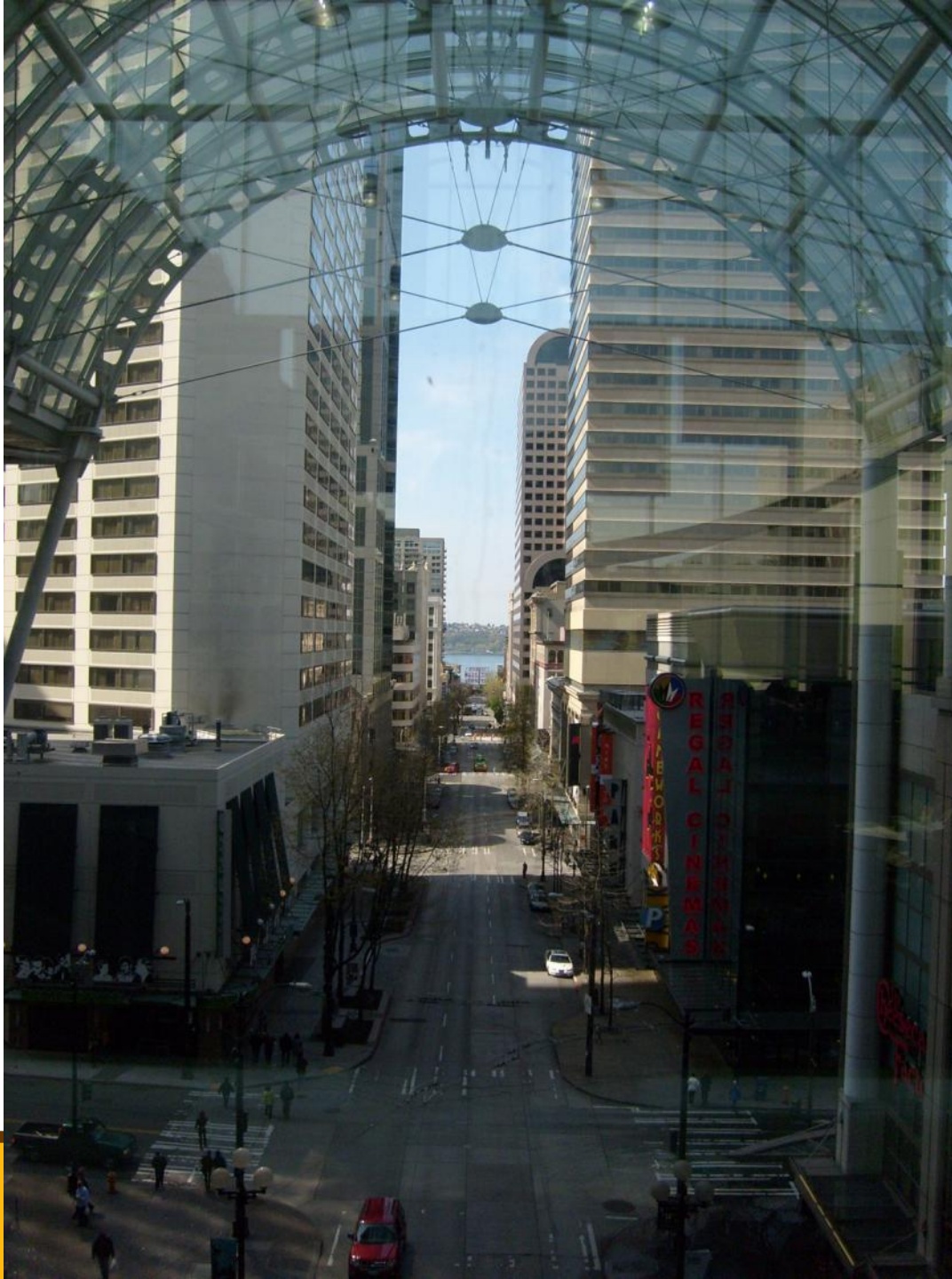
Presented and selected as “Honorable mention Paper” in
2010 SAS Global Forum, Seattle

Mahmoud Azimaee
MCHP Data Acquisition Officer





CONVENTION & TRADE CENTER







Alliance Café/Academic Sponsors

eWise PUTNUT

IBM

Gold and Silver Sponsors

redhat N HATEZZA

IBM

Premier Sponsors

hp TERADATA

IBM

Gold and Silver Sponsors

redhat N HATEZZA

IBM

Alliance Café/Academic Sponsors

eWise PUTNUT

IBM

sas

Follow us on Twitter @SASInstitute They are #SAS10

sas







PUBLIC MARKET



FARMERS MARKET

LASALLE HOTEL







ROUS ROYCE
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99

COLD SEAFOOD
SUMMER!
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99

WINGLESS
CRAB
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99

FRESH
WHOLE
ALASKAN
HALIBUT
\$9.99/lb.

FRESH
WHOLE
ALASKAN
KING
SALMON
\$19.99/lb.

FRESH
WHOLE
KING
SALMON
\$19.99/lb.

FRESH
WHOLE
KING
SALMON
\$19.99/lb.

FRESH
WHOLE
KING
SALMON
\$19.99/lb.

WE DELIVER
TO ALL
DOWNTOWN
HOTELS
AND TO THE
AIRPORT!

AIRLINE
TRAVEL
PACKS!
24 HOUR PACK
\$5.99
48 HOUR PACK
\$10.99

1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99

FISH MARKET
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99
1/2 lb. 3.99
1/4 lb. 2.99



FILLET
OF YOUR
CE
A CHARG

TODAYS
SPECIAL
COOK
SHR
IN S
WE ALSO PROVIDE
FREE DELIVERY TO
ALL MAJOR DOWNTOWN

FRESH MEDIUM
RED KING SALMON
\$8.99
POUND

FRESH
JUMBO
GRILLING
SCALLOPS

WE PACK FISH
TO GO FOR
48 HOURS

FANCY
GRILLING
SCAMPI'S

KING CRAB
SECTIONS
WE PACK
TO GO

FRESH
LOCAL
MUSSELS
\$4.49

FANCY
JUMBO
GRILLING
LOBSTER
TAILS


FRESH
MONK
FISH
\$16.99
POUND

FRESH
ALASKAN
HALIBUT
CHUCKS
\$19.99





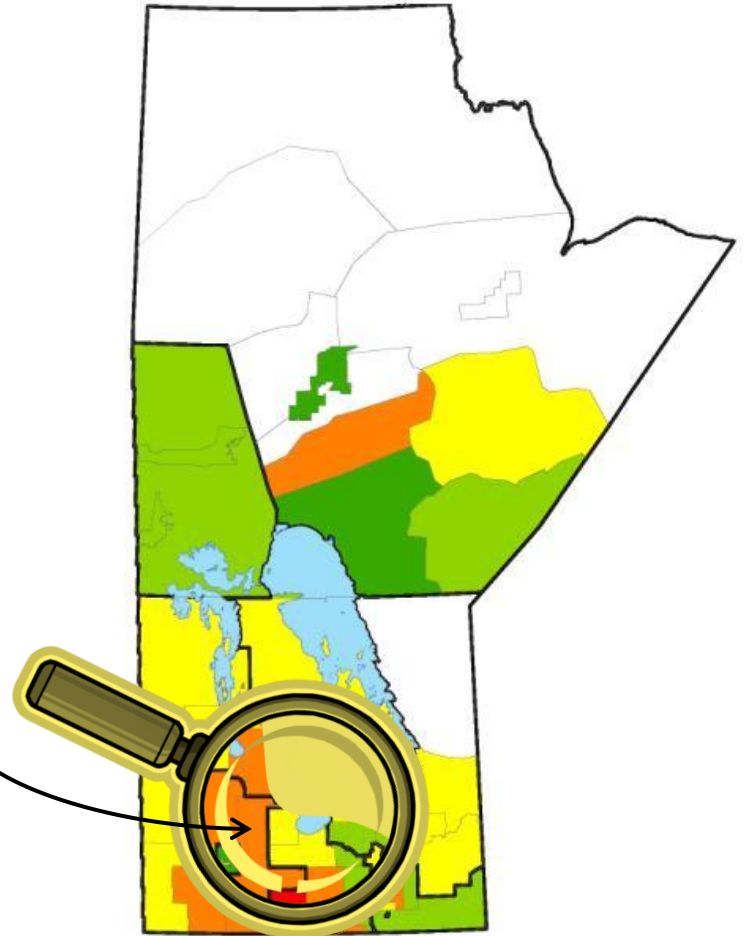
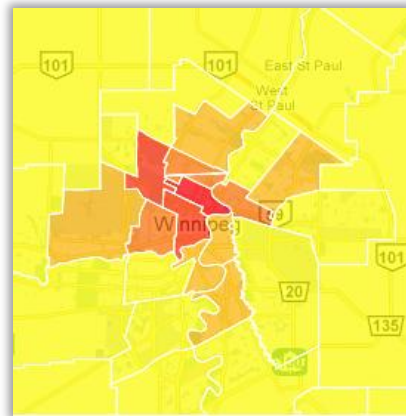
401.1.1
coming in
a) to be
b) to be
c) to be
d) to be



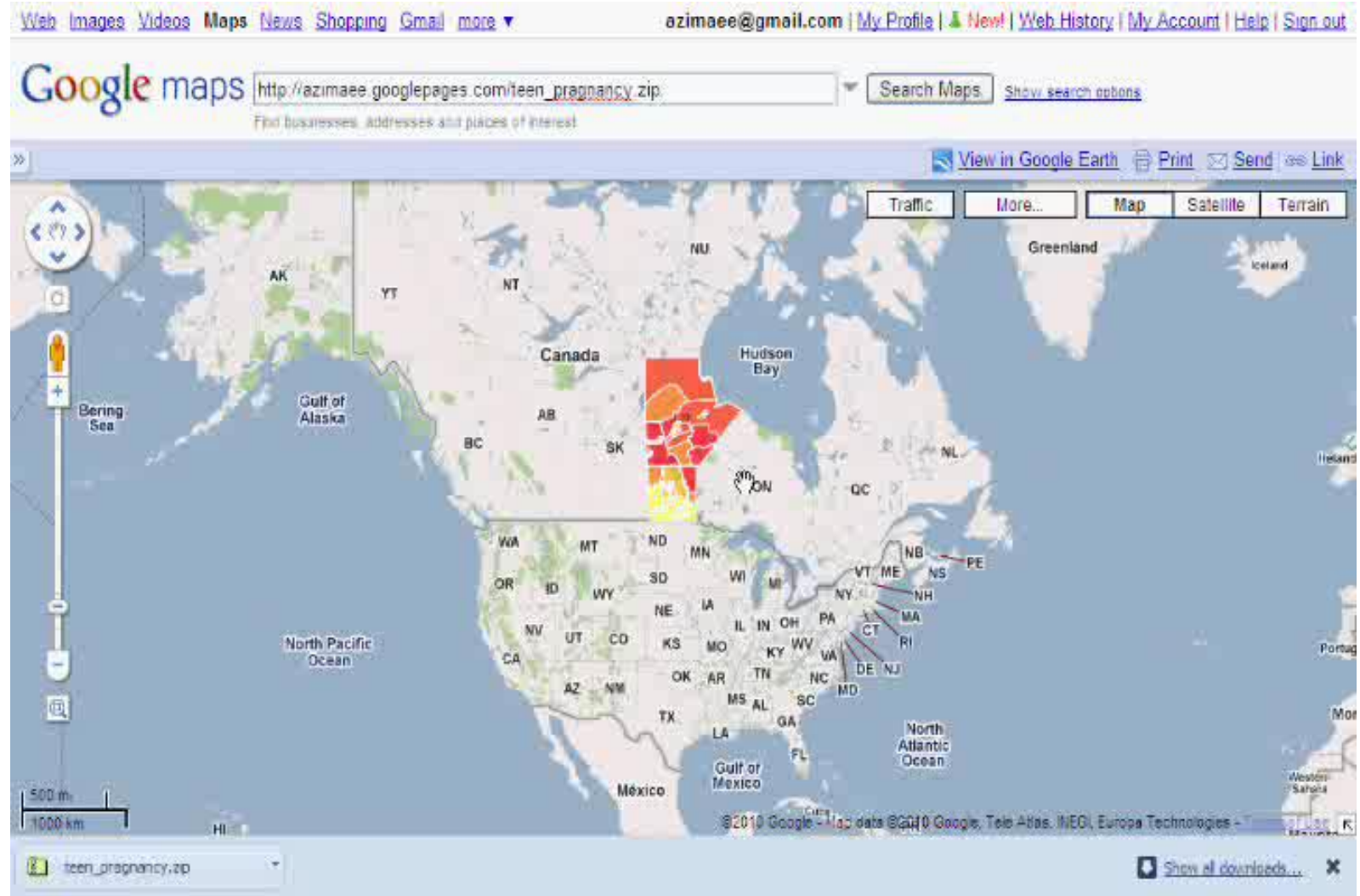


Mapping Health Data

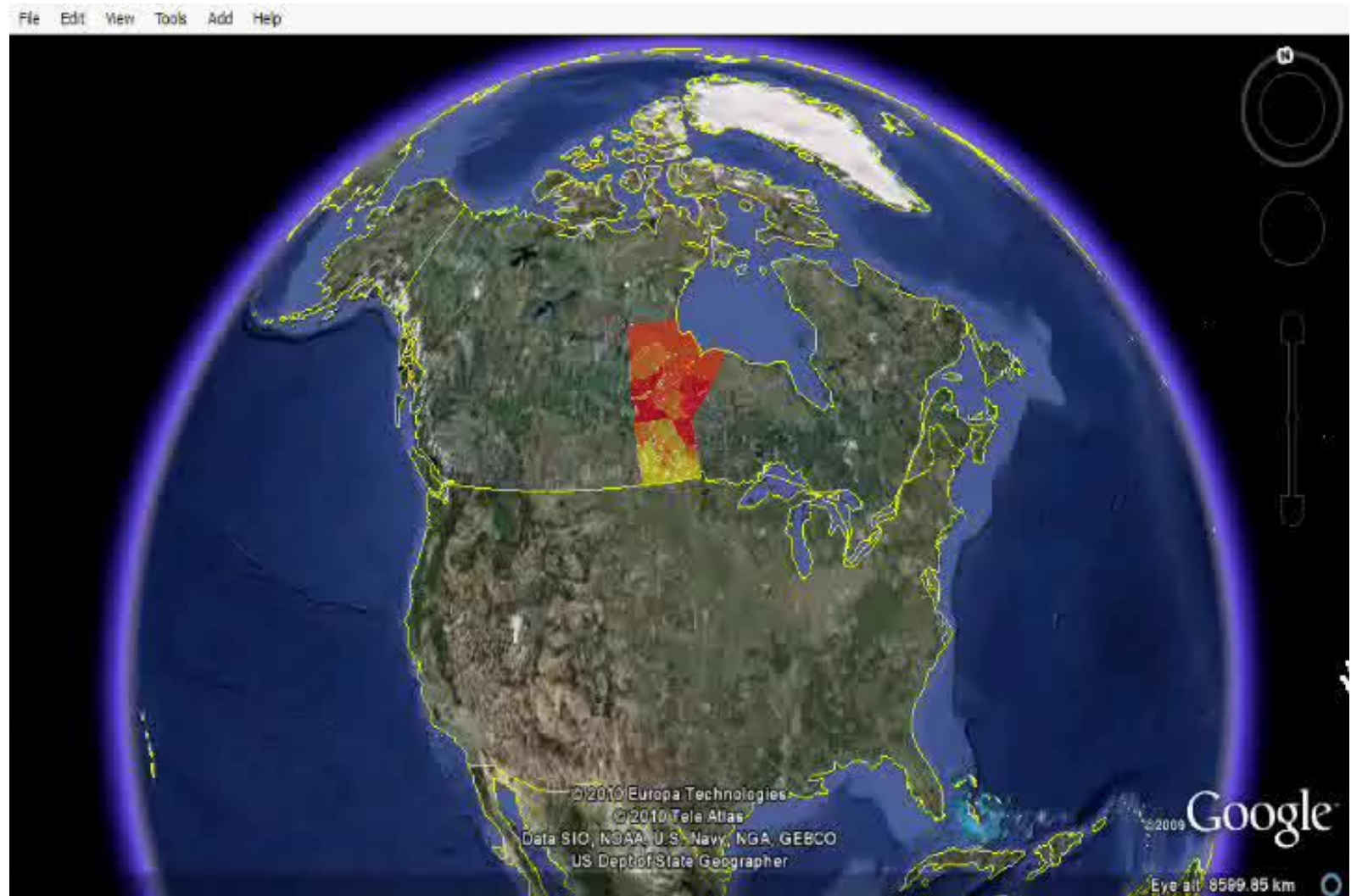
- Using common GIS software's such as ArcView
 - Static maps like:



Why not using the Google Map?



Or even Google Earth?



Background

- Using SAS® and Google Earth to Access and Display Air Pollution Data
 - By: Josh Drukenbrod and David Mintz
 - SAS Global Forum 2008
 - Great works and inspiring , but:
- Not fully automated process
- Requires three separate steps

Definitions and Concepts

- Google API
 - Application programming interface
 - The API provides a number of utilities for manipulating maps and adding content to the map through a variety of services, allowing you to create robust map applications on your website. (From: Google API Documentations)

Definitions and Concepts

- KML
 - Keyhole Markup Language
 - Submitted by Google to the Open Geospatial Consortium (OGC)
 - KML is a file format used to display geographic data in an Earth browser such as Google Earth, Google Maps, and Google Maps for mobile. KML uses a tag-based structure with nested elements and attributes and is based on the XML standard.
 - Many applications display KML, including Google Earth, Google Maps, Google Maps for mobile, NASA WorldWind, ESRI ArcGIS Explorer, Adobe PhotoShop, AutoCAD, and Yahoo! Pipes.

Objective

- Using SAS to create a KML file for a specific geographical region by assigning an estimated measure to each sub-region
- Requirements:
 - Two data sets:
 1. SAS Perimeter Coordinates Data
 2. Rate Data
 - A SAS Macro

SAS Perimeter Coordinates Data

- Available through Governments websites in common GIS formats:
 - Shapefile (SHP, SHX, DBF), AutoCAD (DXF), Geography Markup Language (GML)
 - Need to be converted to an appropriate SAS data structure
- SAS Map Library also can be used
 - Need some modifications

SAS Perimeter Coordinates Data

Variable	Description	Is variable name reserved?	Required/Optional	Type
latitude	geographical latitude	yes	required	numeric
longitude	geographical longitude	yes	required	numeric
rhad	region's short name (code)	no	required	character
seg	region's segments	no	required	numeric
name	region's full name	yes	optional	character
pop	region's population in 2008	yes	optional	numeric

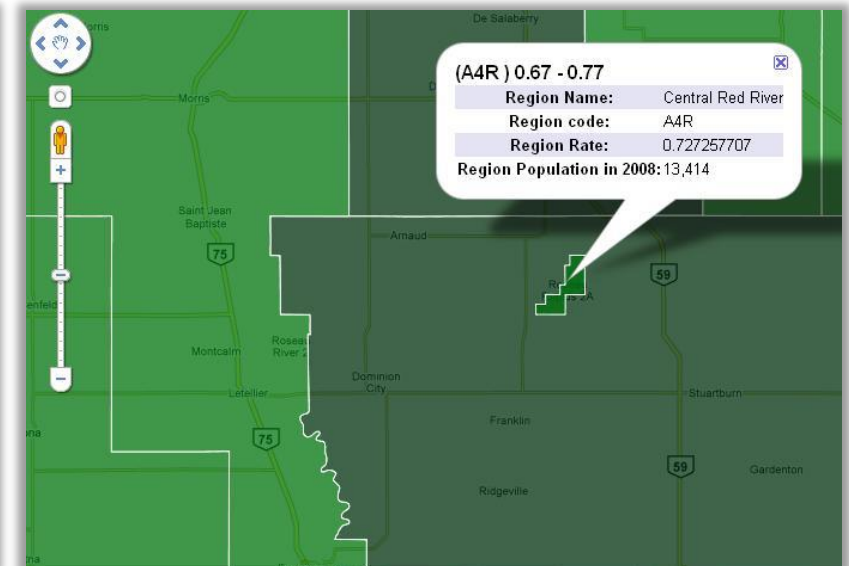
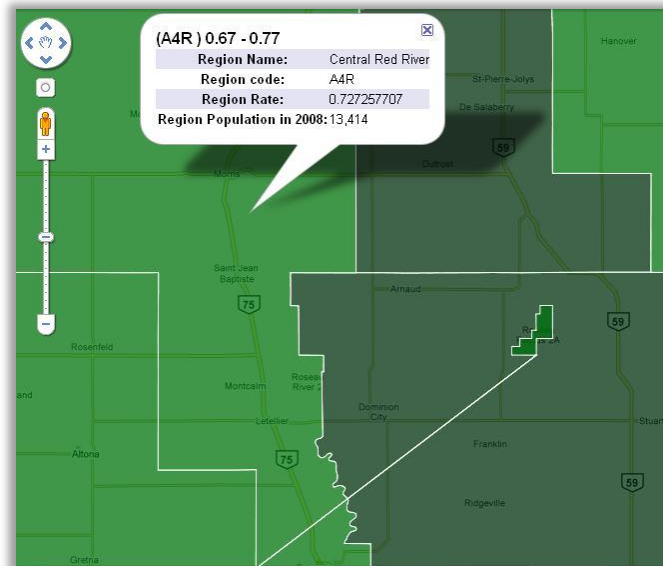
Note: This data set has multiple records per region

SAS Perimeter Coordinates Data

- Google draws a polygon for each region
- Polygons will be drawn tracing counter-clockwise over perimeters of each region
 - Make sure that your data is sorted in a proper order by coordinates in each region.
- If some regions need to be made by more than one polygon, they must be marked with additional segments (seg)

SAS Perimeter Coordinates Data

- what would happen if a segment value remain intact while there are more than one polygon in a region?



SAS Perimeter Coordinates Data

- How it looks?

	rhad	name	pop	latitude	longitude	seg
1574	A4R	Central Red River	13414	49.0004481	-97.21484018	1
1575	A4R	Central Red River	13414	49.00046116	-97.24576519	1
1576	A4R	Central Red River	13414	49.00046146	-97.26162183	1
1577	A4R	Central Red River	13414	49.00046715	-97.2836282	1
1578	A4R	Central Red River	13414	49.00048652	-97.30562724	1
1579	A4R	Central Red River	13414	49.00047259	-97.32475551	1
1580	A4R	Central Red River	13414	49.192295	-96.93912505	2
1581	A4R	Central Red River	13414	49.19241335	-96.97270967	2
1582	A4R	Central Red River	13414	49.20013813	-96.97308338	2
1583	A4R	Central Red River	13414	49.20005025	-96.96173093	2
1584	A4R	Central Red River	13414	49.20746235	-96.96161652	2
1585	A4R	Central Red River	13414	49.20729447	-96.94513699	2
1586	A4R	Central Red River	13414	49.21470644	-96.94524385	2
1587	A4R	Central Red River	13414	49.21466444	-96.93927769	2
1588	A4R	Central Red River	13414	49.22926712	-96.93955995	2
1589	A4R	Central Red River	13414	49.22941587	-96.9338684	2
1590	A4R	Central Red River	13414	49.23692321	-96.93383794	2
1591	A4R	Central Red River	13414	49.2368965	-96.91680144	2
1592	A4R	Central Red River	13414	49.224041	-96.9168625	2
1593	A4R	Central Red River	13414	49.20725631	-96.91655735	2
1594	A4R	Central Red River	13414	49.20717236	-96.93894946	2
1595	A4R	Central Red River	13414	49.192295	-96.93912505	2
1596	BN1	Blue Water	8225	51.05959517	-95.15365386	1

Rate Data

- Must have at least two variables: rate , rhad
- Rate: The quantity you want to map
 - Such as Prevalence/Incidence Rate, Population, or any absolute frequency number
 - Can have any name
 - Is numeric
- Rhad: Regions
 - Must have the exact same name, properties and values in your perimeter data
- Rate Data Set is a single record per Region

CALLING THE MACRO

- The KML macro has been written as a "Statement-style macro" by using STMT option which requires the IMPLMAC system option to be in effect. It can be turned on through SAS setting or simply by adding the following command before calling the KML macro:

```
Option implmac=1;  
_kml <option>;
```

MACRO OPTIONS

- **rate_file=[]**
 - A two level SAS data name for the Rate data set
- **rate=[rate]**
 - Rate variable name
- **region=[rhad]**
 - Region variable name
- **xy_file=[]**
 - A two level SAS data name for the Perimeter coordinates data set
- **path=[]**
 - Full path for the output .kml file.

MACRO OPTIONS

- **title=['Map Created by SAS KML Macro']**
 - Choose a title for your Map
- **opacity=[70]**
 - Percentage of opacity in the final map: Insert an integer number between 0 (being completely transparent) to 100 (being completely opaque).
- **color=['red']**
 - Color theme for the map:
 - red
 - green
 - red2green
 - blue

MACRO OPTIONS

- **method=['quintiles']**
 - Classification method
 - Equal
 - Quintiles
 - k-mean
 - sd
- **showrate=[1]**
 - Control showing the rate values off or on
 - By default the exact rate value in each region will be shown in an info-bubble by clicking on the region. If there are concerns about the confidentiality of the rate values, assign value of 0 to this option to prevent showing of the exact values.

CLASSIFICATION METHODS

- **Equal Intervals**
 - Divides the range of the rate values into 5 equal classes.
- **Quintiles Method**
 - Distributes the rate values into 5 groups that contain an equal number of rates. These five groups are defined by calculating 20th, 40th, 60th, and 80th percentiles.
- **K-mean Clustering Method**
 - Classifies observations to 5 groups in which each observation belongs to the class with the nearest mean. A numerical iterative method is used by Proc FASTCLUS

CLASSIFICATION METHODS

- **Standard Deviation Method**
 - Finds the mean value and then defines 0.5 standard deviation above and below the mean as the middle class. The two other class breaks will be defined by one standard deviation above and below the mean. The values beyond \pm one standard deviation from the mean, will be classified as the first and last classes.

Example 1

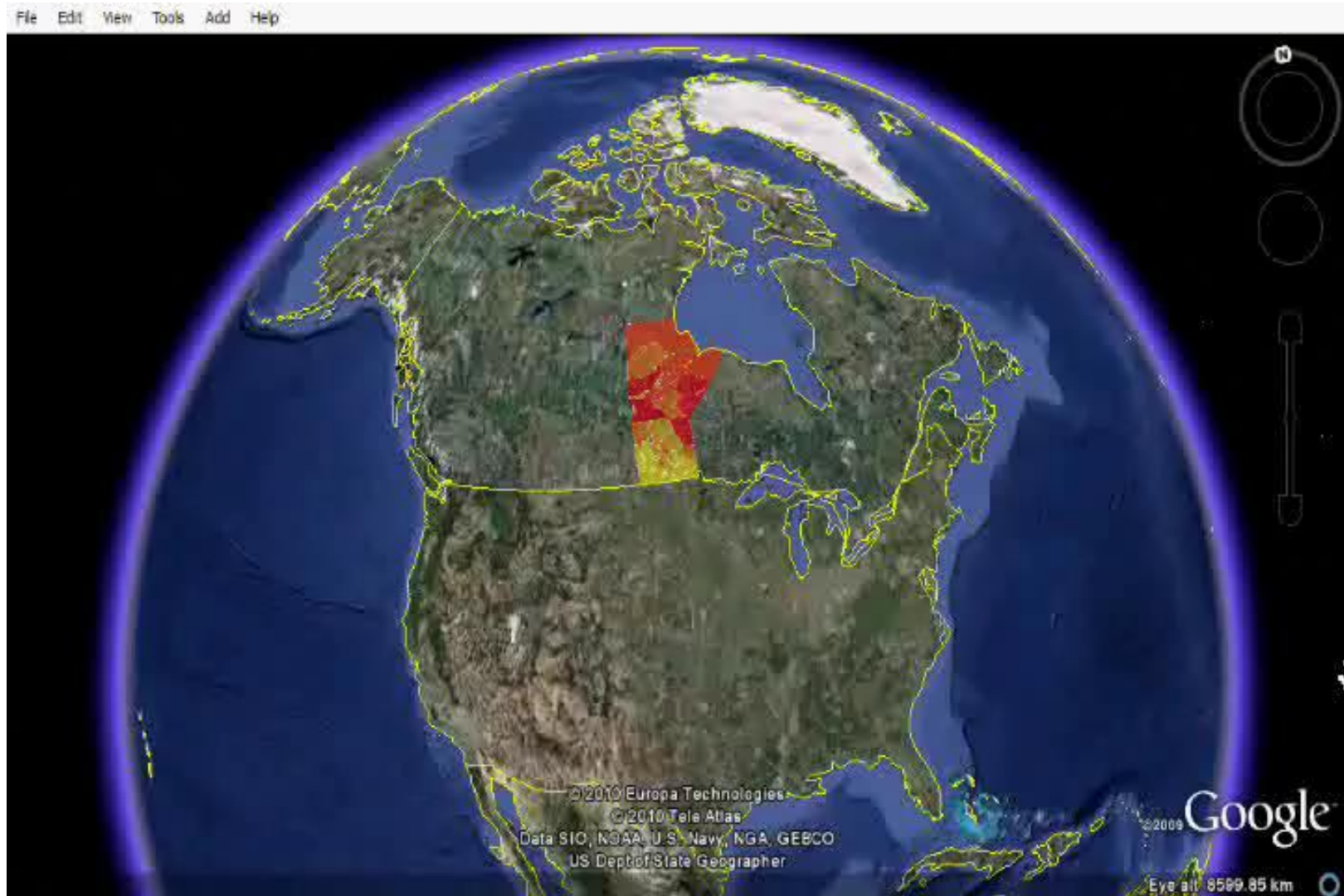
- Teen Pregnancy Rate in Manitoba from 1996/97-2003/04
- Data by: Manitoba Centre for Health Policy
- Rates are age-adjusted per 1000 females age 15-19
- Regional health authority districts (RHAD) break the province down to 80 individual districts
- SAS perimeter coordinates data has been created by the author and is available through Email.

Example 1

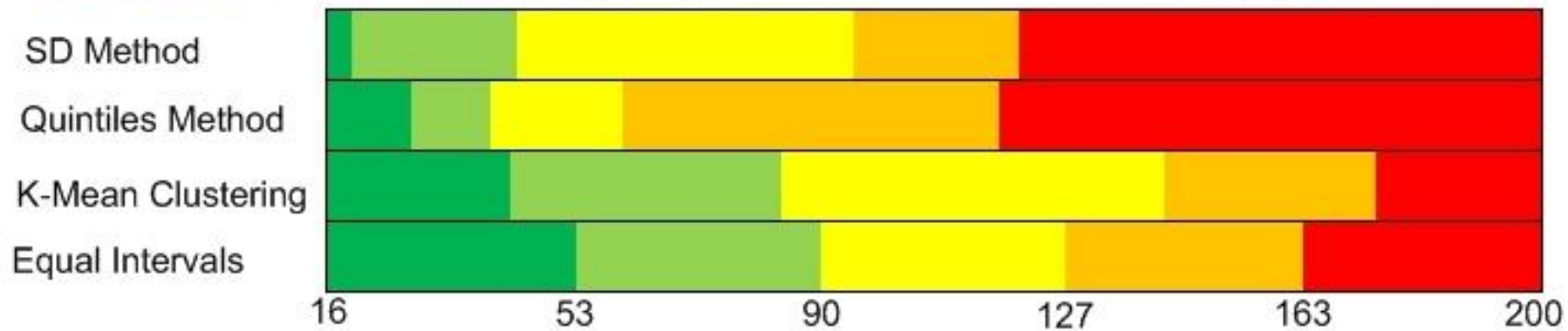
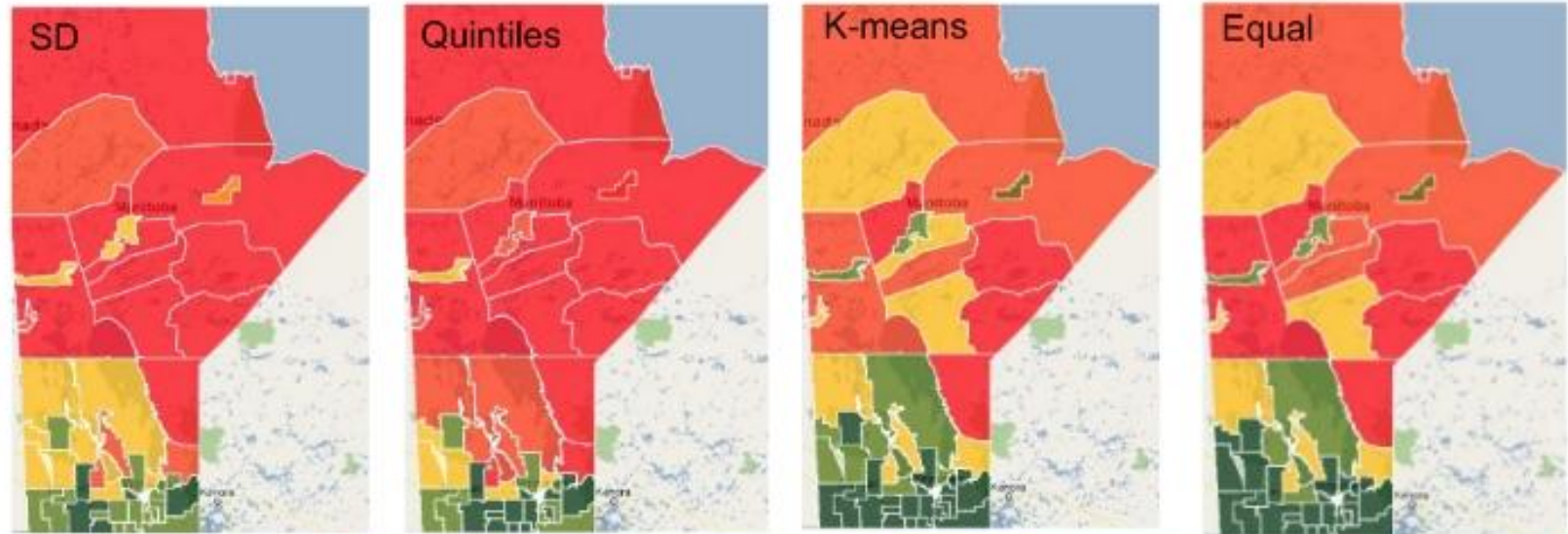
- Calling the Macro:

```
libname mymaps 'C:\My Maps\';  
%include 'C:\My Maps\_kml.sas';  
option implmac=1 ;  
_kml  
    rate_file=mymaps.MCHP_Data  
    rate=pregnancy  
    xy_file=mymaps.MB_by_districts  
    path='C:\My Maps\teen_pregnancy.kml'  
    title='Teen Pregnancy Rates 1996/97-2003/04, (Age-adjusted rate of  
    teen pregnancies per 1,000 females age 15-19)'  
    method='equal';  
Run;
```


Example 1



Comparing all 4 classification methods



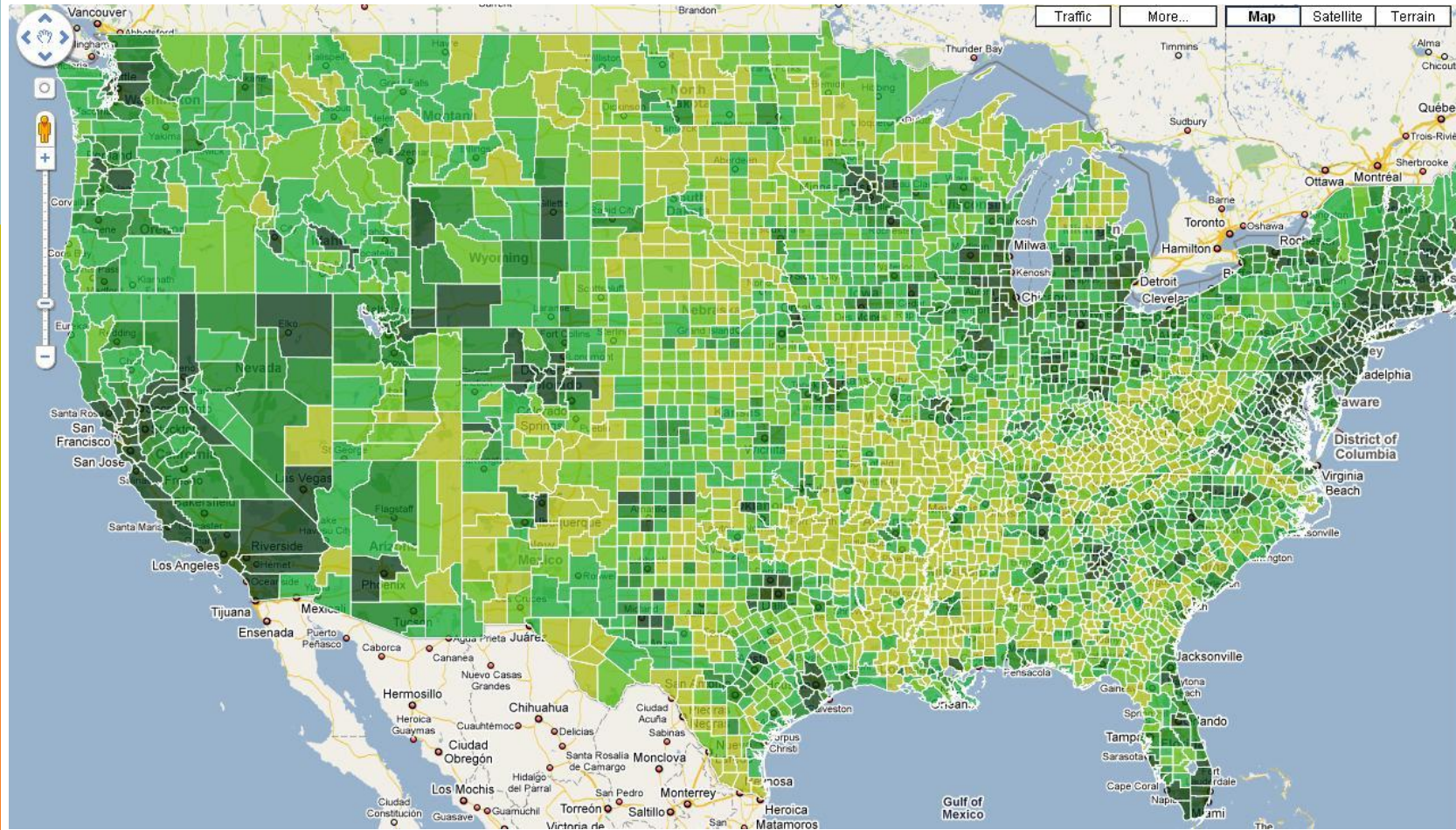
Example 2 (Using SAS MAP Library)

```
libname mymaps 'C:\My Maps\';
data mymaps.USA;
    *** Get names and Average Household Income of the counties;
    merge maps.counties(in=datain)    maps.usaac(in=namein keep=state county cntyname
    AHI89);
    by state county;
    if datain;
    rename cntyname=name;
    pi = gamma(0.5)**2;          *** Convert coordinates from radian to degree;
    longitude = x * -(180/pi);
    latitude = y * (180/pi);
    cnty=state||county;    * Re-define counties;
    if longitude=. | latitude=. then delete;
    seg=County || segment; * Re-define segments;
    if density < 4 ;        * Only keep data for a moderate resolution;
run;
```

Example 2 (Using SAS MAP Library)

```
*** Call KML Macro;  
%include 'C:\My Maps\_kml.sas';  
option implmac=1 ;  
_kml  
    rate_file=mymaps.USA  
    rate=AH189  
    region=cnty  
    xy_file=mymaps.USA  
    path='C:\My Maps\USA.kml'  
    title='Average Household Income in USA by Counties, '  
    opacity=75  
    color='green'  
run;
```

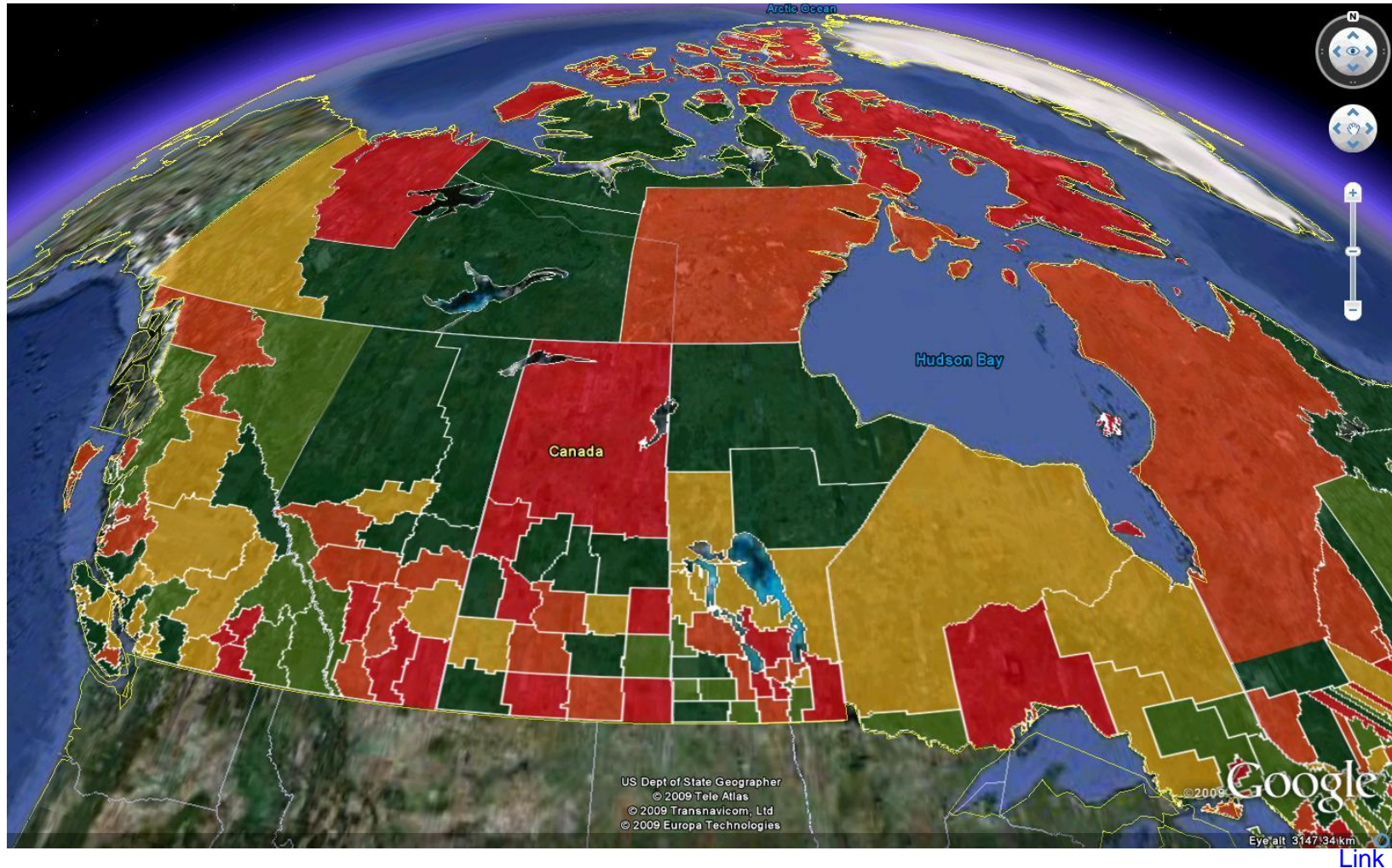
Example 2 (Using SAS MAP Library)



Example 3 (Using SAS MAP Library)

```
libname mymaps 'C:\My Maps\';
data canada;
    set maps.canada3;
    pi = gamma(0.5)**2;          *** Convert coordinates from radian to degree;
    longitude = x * -(180/pi);
    latitude = y * (180/pi);
    subdiv=province||CDCCODE;   * Re-define sub-divisions;
    if longitude=. | latitude=. then delete;
    seg=CDCCODE || segment;     * Re-define segments;
    if density < 4 ;            * Only keep data for a moderate resolution;
run;
data mymaps.canada; *** Add names of sub-divisions;
    merge canada(in=datain) maps.cancens(in=namein);
    by province cdcode;
    if datain;
run;
```

Example 2 (Using SAS MAP Library)

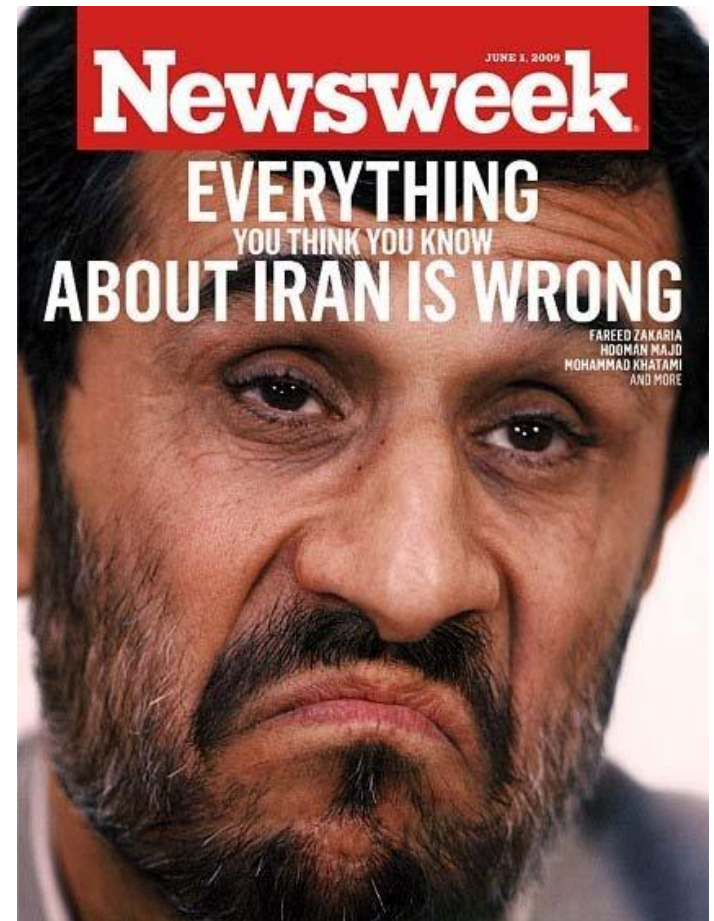


HOW TO USE KML FILE?

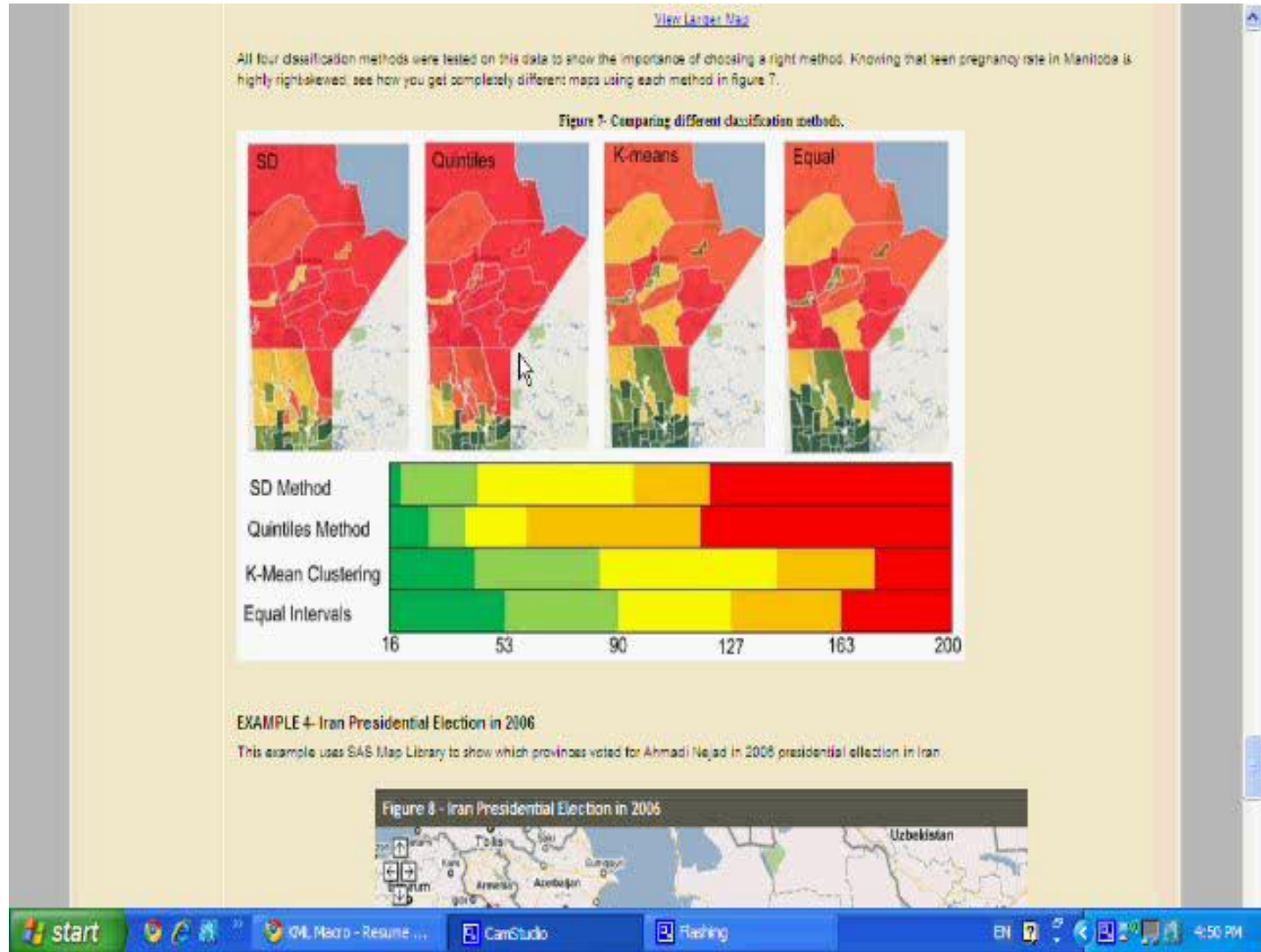
- Upload kml file to a URL address then in maps.google.com put that URL on the search.
- Sign in with a Google account in maps.google.com Import your file.
- If you have Google Earth software installed on your computer, just click on the KML file and it will be opened by Google Earth.
- There is also free Google Earth plug-in available online for Firefox and Internet Explorer that make you able to view your KML file 3-D within your browser.

An Example of Embedded Map

- Iran Presidential Election in 2006
- Who voted for this guy:
- [Iran's map by provinces](#)



Iran Presidential Election in 2006



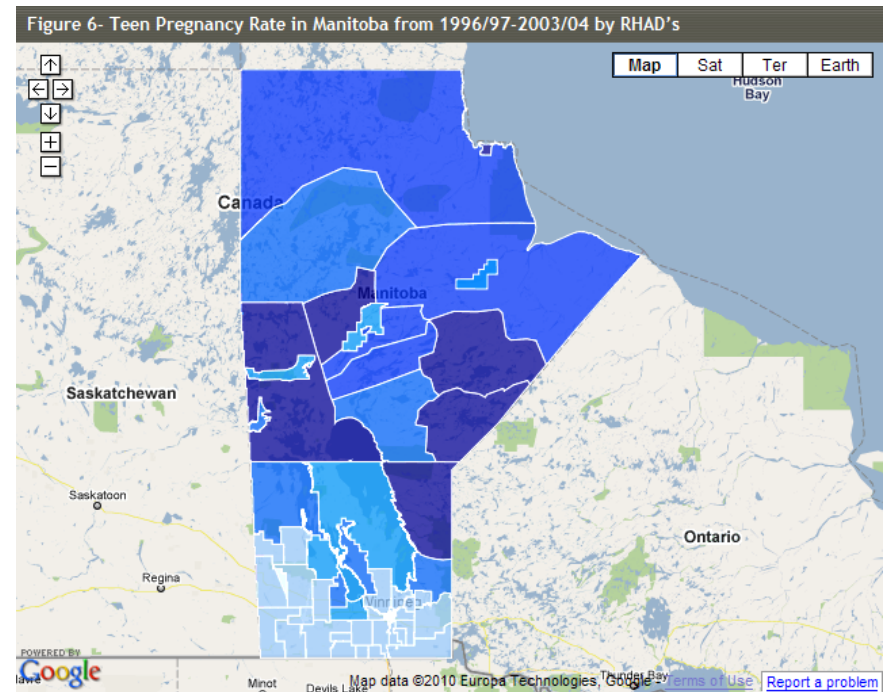
BODY OF THE MACRO

- Sort the coordinate data set by region
- Run Proc Univariate for the rate data set to calculate Percentiles, Mean, and Standard Deviation of the rates
- If user has chosen “K-Mean” classification Method then Run Proc FastClus
- Based on the user choice of Color Theme and rate class of each region assign hexadecimal color codes to each region
- Convert opacity choice of user from percentage to hexadecimal and concatenate it with the color code
- Start writing all these information in XML language in a file on disk

What is New in Version 1.2

Comparing to version presented in SGF 2010

- The latest KML version 2.2 for namespace declaration
 - No need to obtain a Google API Key anymore!
 - One extra color theme : blue



Conclusion

- Using KML macro is the best solution for web based research reports as the maps created by this macro can be easily embedded to any webpage, while all the zooming and navigating features still can be reserved.

Acknowledgment

- I would like to thank Charles Burchill (Manager, Program and Analysis Systems, and Privacy Officer at the Manitoba Centre for Health Policy) for his great advice, comments and encouragement in writing and testing this macro.



APRIL 11–14
SEATTLE
WASHINGTON

CONTACT INFORMATION

Mahmoud Azimae

Manitoba Centre for Health Policy

University of Manitoba

Work Phone: (204) 789-3668

E-mail: azimae@gmail.com

Web: www.dastneveshteha.com

