SAS® GLOBAL FORUM 2016

IMAGINE. CREATE. INNOVATE.

Creating a better SAS Visual Analytics user experience under HIPAA data restrictions

Authors: Margaret Burgess, Catherine Olson, Jessica Wegner

#SASGF

Creating a better SAS Visual Analytics user experience under HIPAA data restrictions

Margaret Burgess, Catherine Olson, Jessica Wegner

Optum

ABSTRACT

The HIPAA Privacy Rule can restrict geographic data used in healthcare analytics. This poster presents a variety of data visualizations an analyst could use when describing a healthcare population and challenged with these restrictions. It explores the different options SAS Visual Analytics (VA) offers and provides tips on data preparation prior to using SAS VA Designer.

BACKGROUND

To be compliant with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, protected health information (PHI) must be de-identified to reduce privacy risks to individuals.[1] Identifiable information can be used if informed consent is obtained from the individuals (see Section 46.116 of the Privacy Rule) but it is often time consuming and costly.[2] De-identification is determined by a qualified expert or removal of specific identifiers.

A qualified expert reviews data to determine the level of risk that the data could be used alone or with other available data to identify individuals. In order to reduce the risk of identification, the use of certain attributes may be limited. In the case of geographic data, only high level geographic subdivisions may be allowed (such as census region or division) when used with other potential identifiers.

Removal of identifiers (also known as Safe Harbor) is a frequently used method to de-identify a dataset. For geographic locations, Safe Harbor limits to only the state level or the first three digits of a ZIP code (the first three digits of a ZIP code must be for an area of more than 20,000 people).

DATA PREPARATION

When constructing the dataset prior to using SAS VA Designer, consider the following:

Dataset Summary Level
- When creating the dataset for SAS VA, create one record per item that contains all the attributes to be analyzed. For example, if the VA report is going to provide metrics about patients, create your dataset with one record per patient.
- Each patient record should contain all of the attributes for that patient.

DATA PREPARATION CONTINUED

Metadata
- The SAS dataset created should be self-documenting; it should contain as much metadata as possible. Include any selection criteria or pertinent information about the file that uses VA report titles and headings. This includes criteria used to create the SAS dataset. For example, if the patient file includes only Medicare subjects, make sure that this attribute is included on each patient record. It is important to have this information as the SAS VA report library grows to keep track of the exact contents of each report.
- Include LABELS in the dataset. SAS VA uses labels as the default when creating the report. This results in a better user experience.

METHODS

Four different options for displaying geographic data with various levels of detail were explored to determine which provided the best user experience.

- **Bar Chart & Tree Map -** When limited to the use of broad geographic areas, such as census division, this option displayed the number of members in a division using bar chart and tree map objects.
- **Geo Region Map –** When allowed use of more specific geographic areas, such as state, this option displayed the number of members in each state using the geo region map object. Three different interactions with the map were created using button bar objects that represented age, business line, and gender data categories. These filtered the map display depending on the button selected.
- **Geo Bubble Map with Longitude/Latitude –** When limited to the use of broad geographic areas, such as region, this option displayed the number of members in each region using the geo region map object along with the creation of a custom geographic data category for longitude and latitude. Three different interactions with the map were created using button bar objects that represented age, business line, and gender data categories. These filtered the map display depending on the button selected.
- **Cross Tab –** When allowed use of more specific geographic areas, such as three digit zip code, this option displayed the number of members in each three digit zip code viewable by age, gender and business line. Three different interactions with the cross tab were created using drop down objects that represented age, business line, and gender data categories. These filtered the cross tab display depending on the categories selected.

Creating a better SAS Visual Analytics user experience under HIPAA data restrictions

Margaret Burgess, Catherine Olson, Jessica Wegner

Optum

METHODS CONTINUED

The four different options for displaying geographic data used key technical components of SAS VA:

- **Company Standard Colors -** Company standard colors were applied to the data objects by using appropriate red, blue, and green hexadecimal scales.
- **Custom Data Categories -** Custom data categories were used to rename elements for gender and business line. Gender values 'M' and 'F' were renamed to 'Male' and 'Female' and business line values of 'MCD', 'MCR', and 'COM' were renamed to 'Medicaid', 'Medicare', and 'Commercial'.
- **Filtering -** The filtering option was used to exclude invalid data elements from the state data category and missing values from business line, gender and age categories.
- **Interactions -** Interactions were used to filter the geo map object, geo bubble map, and cross tab objects by gender, business line, and age category.
- **Geography Custom Categories -** Geography custom categories were used to create longitude and latitude points to plot on the geo bubble map for region.
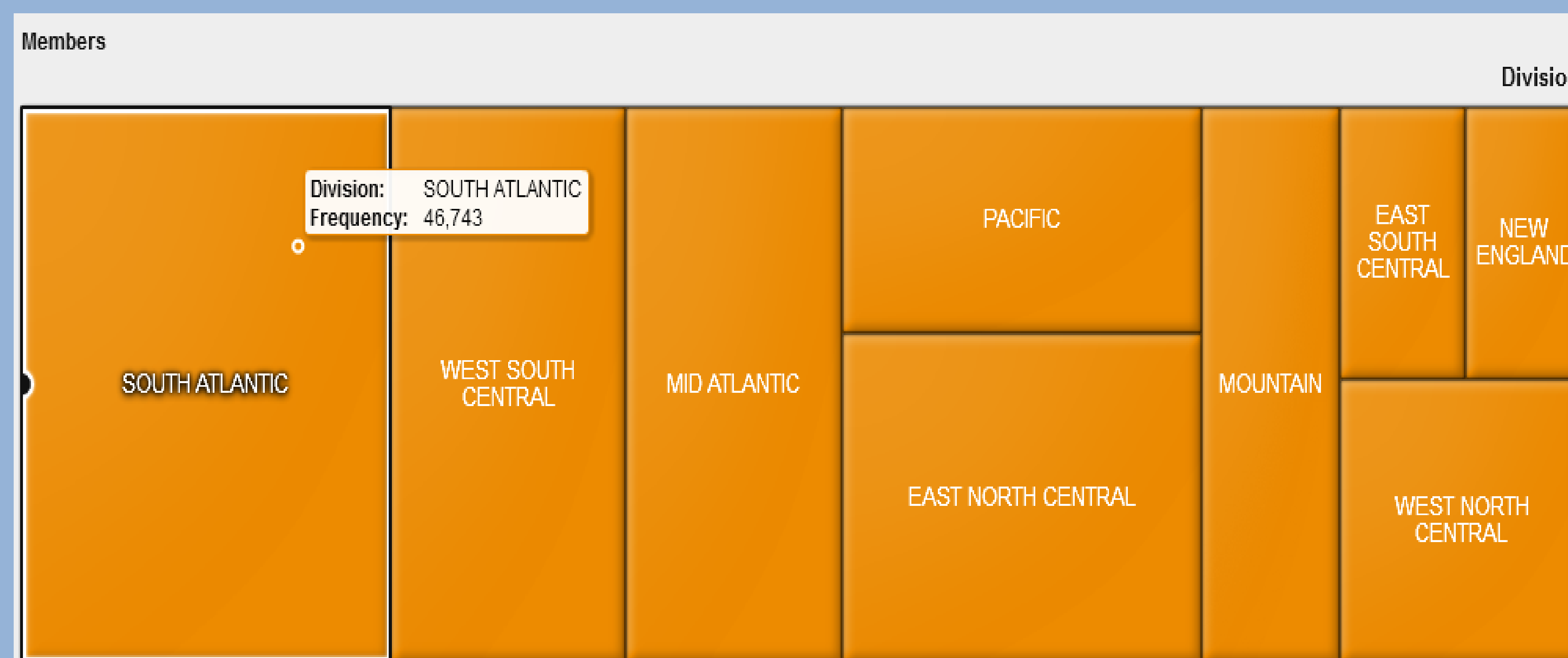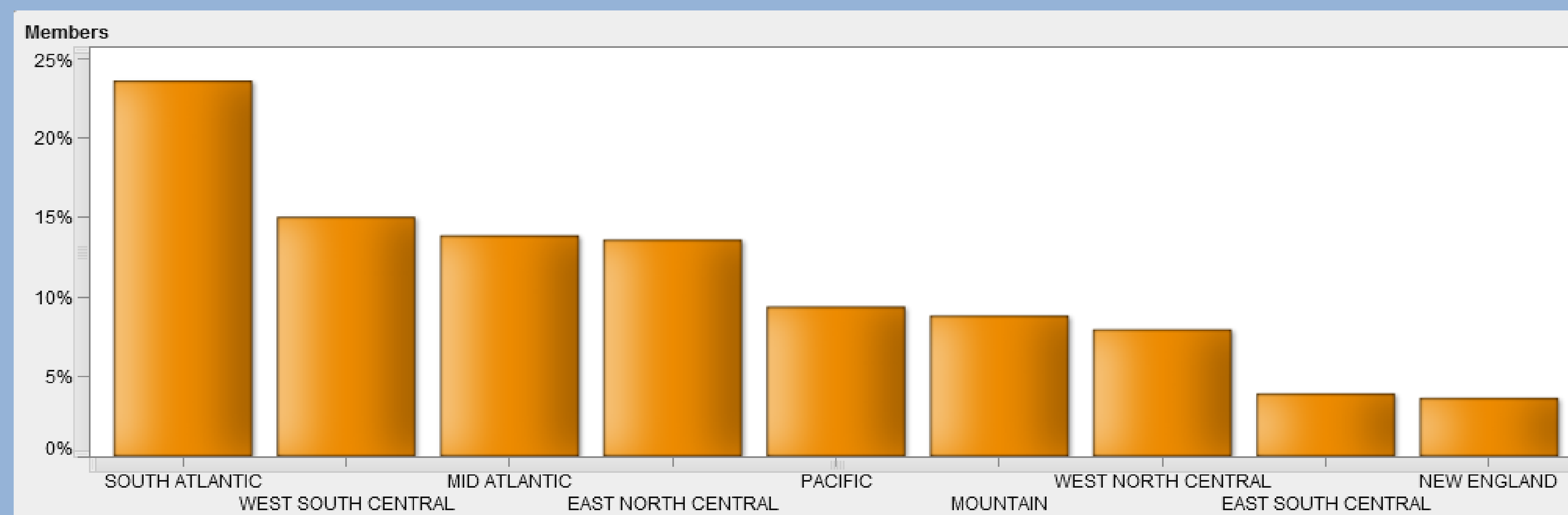
RESULTS

Each of the four different options for displaying a healthcare population's limited geographic data are shown below.

Bar Chart & Tree Map

A company standard orange was used for the "data fill" along with matte for the "data skin" in both the bar chart and tree map objects. An example illustration describes the South Atlantic division. The bar chart shows that there are approximately 24% of the members in the South Atlantic division. The tree map shows that there are 46,743 members in the South Atlantic division.

RESULTS CONTINUED

Creating a better SAS Visual Analytics user experience under HIPAA data restrictions
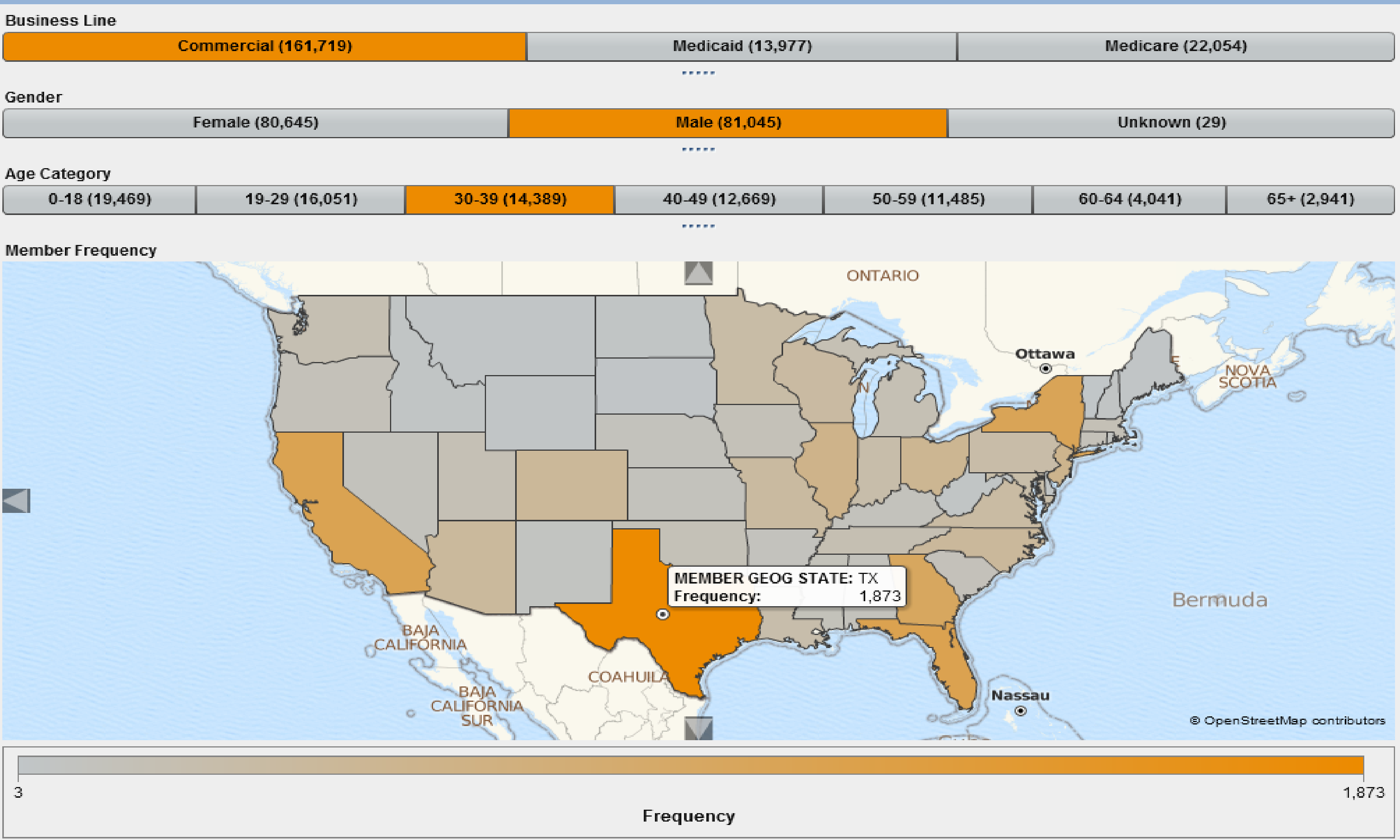
Margaret Burgess, Catherine Olson, Jessica Wegner
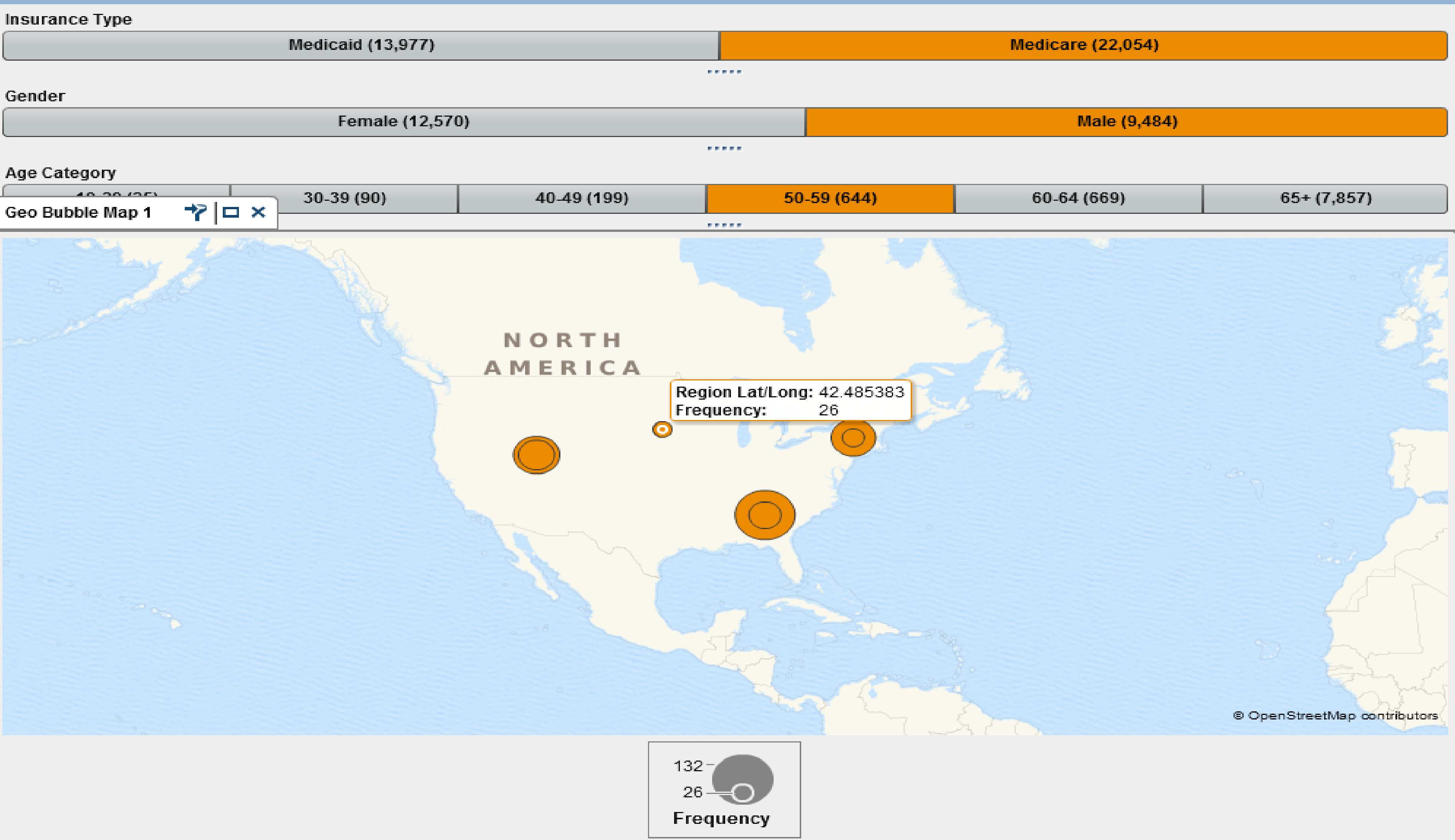
Optum

RESULTS CONTINUED

Geo Region Map

The state geo map used a company standard orange for the greatest gradient color and company standard grey for the lowest gradient color. For the gender, age category and business line button bars, the company standard orange was used for the selected button and the company standard grey was used for buttons that are not selected. Interactions were created between business line and gender button bars, gender and age category button bars, and the age category button bar and state geo map. An example illustration shows that there are a total of 161,719 commercial members. Of those commercial members, 81,045 are male. Of those male members, 14,389 are in the 30-39 age category. Of those in the 30-39 age category, 1,873 live in Texas.

RESULTS CONTINUED

Geo Bubble Map with Longitude/Latitude

Company standard colors and interactions were used similar to the Geo Region Map option described previously. The geography custom category was used to represent latitude and longitude points for Midwest, Northeast, South, and West. An example illustration shows that there are 22,054 Medicare members. Of the Medicare members, 9,484 are male. Of the males, 644 are between 50-59 years old. Of those between 50-59 years old, 26 live in the Midwest.

Creating a better SAS Visual Analytics user experience under HIPAA data restrictions
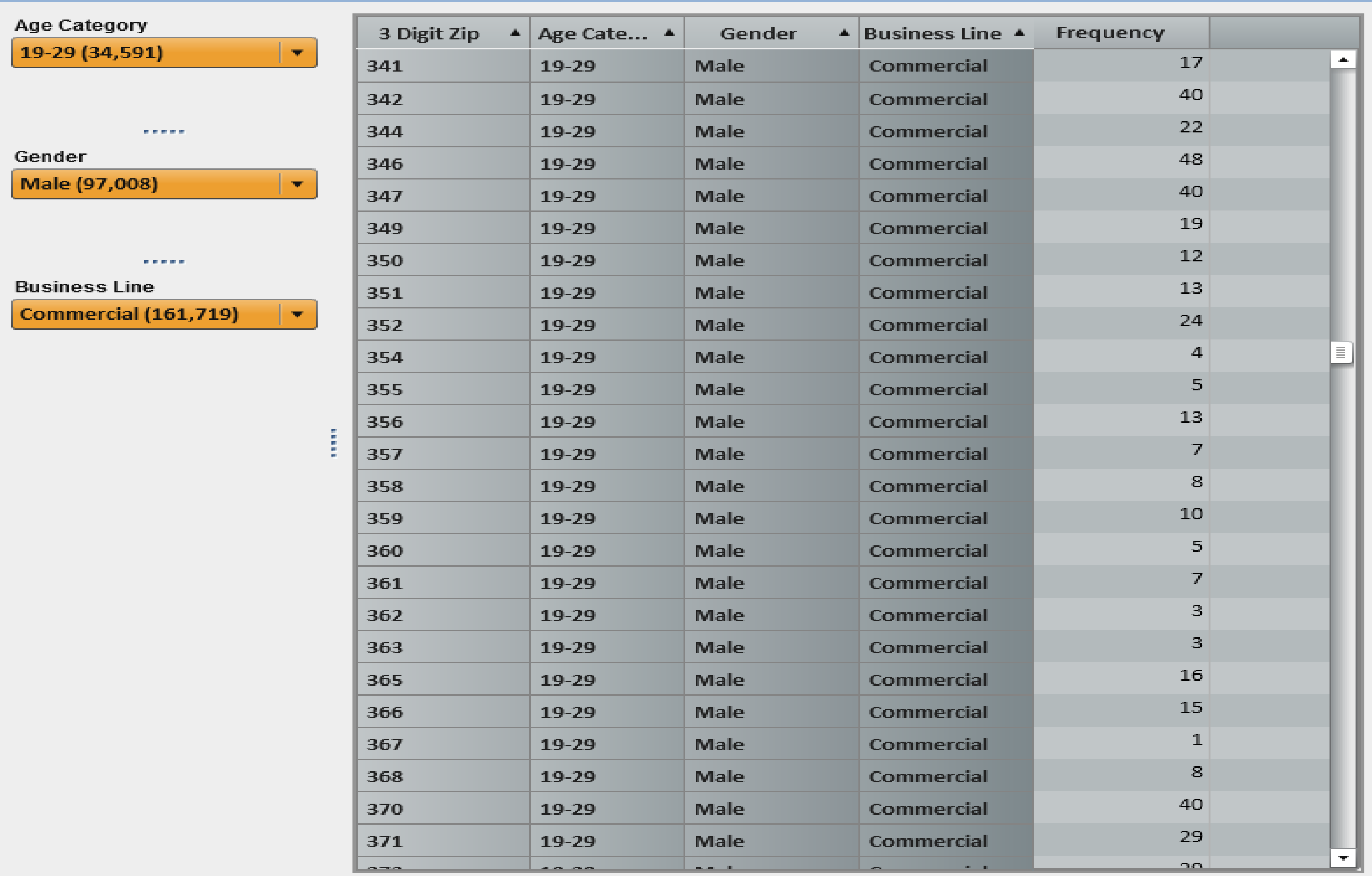
Margaret Burgess, Catherine Olson, Jessica Wegner

Optum

RESULTS CONTINUED

Cross Tab

Company standard orange was used for the drop down objects and company standard grey for the crosstab. Interactions were created between the age category drop down and the cross tab, the business line drop down and the cross tab, and the gender drop down and the cross tab. An example illustration is that there are a total of 34, 591 members who are between 19-29 years old, 97,008 who are male, and 161,719 who have commercial insurance. There are 17 members who live in the 341 three digit zip code, are 19-29 years old, are male and have commercial insurance.

Age Category
19-29 (34,591)

Gender
Male (97,008)

Business Line
Commercial (161,719)

3 Digit Zip	Age Cate...	Gender	Business Line	Frequency
341	19-29	Male	Commercial	17
342	19-29	Male	Commercial	40
344	19-29	Male	Commercial	22
346	19-29	Male	Commercial	48
347	19-29	Male	Commercial	40
349	19-29	Male	Commercial	19
350	19-29	Male	Commercial	12
351	19-29	Male	Commercial	13
352	19-29	Male	Commercial	24
354	19-29	Male	Commercial	4
355	19-29	Male	Commercial	5
356	19-29	Male	Commercial	13
357	19-29	Male	Commercial	7
358	19-29	Male	Commercial	8
359	19-29	Male	Commercial	10
360	19-29	Male	Commercial	5
361	19-29	Male	Commercial	7
362	19-29	Male	Commercial	3
363	19-29	Male	Commercial	3
365	19-29	Male	Commercial	16
366	19-29	Male	Commercial	15
367	19-29	Male	Commercial	1
368	19-29	Male	Commercial	8
370	19-29	Male	Commercial	40
371	19-29	Male	Commercial	29

CONCLUSIONS

The HIPAA Privacy Rule can limit the type of geographic data used when describing a healthcare population. However, multiple options exist in SAS VA to display limited geographic healthcare data in a way that is still informative and useful. The cross tab option was crucial for displaying the more detailed three digit zip code data to capture all the attributes of the multiple three digit zip codes. The geo-map and the geo bubble map with longitude/latitude were attractive options to visually display the members by state and region. Bar graphs, tree maps and other visual analytics objects also are attractive options when there are broad geographic areas with a fewer number of categories, such as census divisions.

REFERENCES

[1]Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule Web Site. http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html Accessed October, 1 2015.

[2]Faden, R, Beauchamp T, and Kass N. Informed Consent, Comparative Effectiveness, and Learning Health Care. N Engl J Med. 2014; 370: 766-68.