Bridging the Gap: Importing Health Indicators Warehouse Data into SAS® Visual Analytics Using SAS® Stored Processes and APIs

Li-Hui Chen, PhD. Consumer Product Safety Commission[1]; Manuel Figallo, SAS Institute Inc.

ABSTRACT

The Health Indicators Warehouse (HIW) is part of the US Department of Health and Human Services' (DHHS) response to make federal data more accessible. Through it, users can access data and metadata for over 1,200 indicators from approximately 180 federal and nonfederal sources. The HIW also supports data access by applications such as SAS® Visual Analytics through the use of an application programming interface (API). An API serves as a communication interface for the integration of two separate systems. As a result of the API, HIW data consumers using SAS Visual Analytics can avoid difficult manual data processing. This paper provides detailed information about how to access HIW data with SAS Visual Analytics in order to produce easily understood visualizations with minimal effort through a methodology that automates HIW data processing. This paper also shows how to run SAS® macros inside a stored process to make HIW data available in SAS Visual Analytics for exploration and reporting via API calls; the SAS macros are provided. Use cases involving dashboards are also examined in order to demonstrate the value of streaming data directly from the HIW. Both IT professionals and population health analysts will benefit from understanding how to import HIW data into SAS Visual Analytics using SAS® Stored Processes, macros, and APIs. This can be very helpful to organizations that want to lower maintenance costs associated with data management while gaining insights into health data with visualizations. This paper provides a starting point for any organization interested in deriving full value from SAS® Visual Analytics while augmenting their work with HIW data.

INTRODUCTION

Through both government and private sector efforts, more and more data have become freely accessible by way of the Internet. Some of those data can be accessed through the use of an Application Programming Interface (API), which allows users to gather data though computer to computer communication. Applications, such as SAS® Visual Analytics (SAS® VA), can handle and disseminate data to meet users' needs. The National Center for Health Statistics' (NCHS) Health Indicators Warehouse (HIW) (National Center for Health Statistics 2015) is one of several federal projects providing access to data through an API.

This paper provides detailed information on how to access HIW data with SAS® VA in order to produce easily understood health statistics visualizations. It guides readers through a process to automate data access to the HIW and then shows the reader how to use SAS® VA to present health information via a web browser. As a result of reading this paper, SAS users will learn to build reports in SAS® VA using data from the HIW's API. The following are covered in the method's section of this paper:

---

[1] Dr. Li Hui Chen is now with the U.S. Consumer Product Safety Commission. The majority of the work presented here was done while she was employed by the Centers for Disease Control and Prevention's National Center for Health Statistics.

1. Process Overview. Overview of the process from gathering information from the HIW to building reports in SAS® VA.

2. Overview of the HIW-API. Finding the necessary information for the API calls from the HIW.

3. How to access data through HIW-API calls in order to load the data into SAS® Visual Analytics for reporting. Instructions and SAS macros to access data through HIW API calls are provided.

4. Build SAS® Visual Analytics Report using HIW-API Data. Use cases to illustrate possible scenarios and demonstrate the value of using SAS® VA and stored processes to feed data directly in SAS® Visual Analytics using the HIW API.

BACKGROUND

HEALTH INDICATORS WAREHOUSE (HIW)

Brief overview

The National Center for Health Statistics' (NCHS) Health Indicators Warehouse (HIW), available at www.healthindicators.gov, is part of the Department of Health and Human Services' (DHHS) response to the Open Government Initiative to make federal data more accessible to all users. Through the HIW, users can view and download data and metadata for over 1,200 indicators on health status, outcomes, and determinants from more than 180 different federal and nonfederal sources. HIW also provides access to data through the use of an Application Programming Interface (API), which is a communication interface that allows machine to machine communications. Applications such as SAS® Visual Analytics can use the HIW's API to access the data repository.
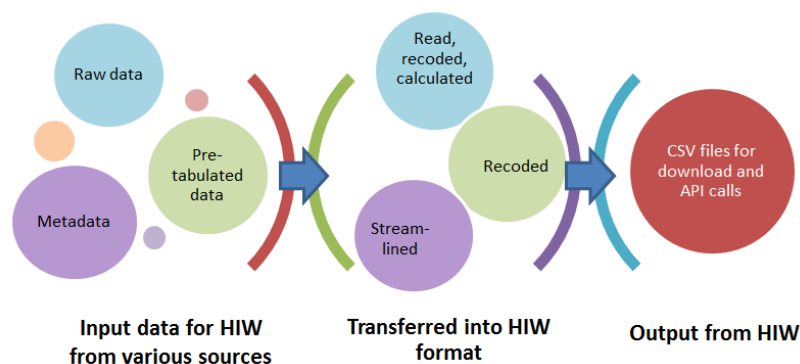

Figure 1. This figure illustrates how data are transformed into organized data and available to the general public through HIW-API calls and csv files.

As shown in Figure 1, the HIW team gathers data from either raw untreated data or pre-tabulated data, as well as all sources of metadata. The data are then transferred into the HIW format through different methods, e.g., recoding for pre-tabulated data, streamlining for metadata. Once the data and metadata are transferred, they are made available to the general public through the HIW website. Users can either download CSV files or gather the data through API calls.

HIW Web Services

The HIW provides programmatic access to the underlying data and metadata of over 1,200 indicators through the use of its API. The data can be queried in multiple ways by using available web services for over 100 data types, or data structures, in the database. For example, HIW users can use the API to get a list of all data sources. The API has been designed to allow application developers maximum flexibility when implementing a solution that utilizes the HIW-API. In most cases, however, there are only a handful of data types or structures which would likely interest developers. Detailed information required to access the underlying services can be found in the documentation section of the HIW for developer site (http://developers.healthindicators.gov/). The data which are accessible via the HIW services can be used to develop diverse applications.

SAS® VISUAL ANALYTICS: BRIEF OVERVIEW

SAS® Visual Analytics is a web-based visualization tool which allows analysts to explore huge volumes of data to identify patterns and trends for further analysis. SAS® Visual Analytics can also run stored processes or batch programs to trigger APIs calls that load HIW health indicators into the SAS® Visual Analytics environment.

APIS: BRIEF DESCRIPTION OF APIS

All of the data and metadata in the HIW are available for use and in this way are considered "open data." The term "open data" refers to publicly available data which are structured in a way which enables the data to be fully discoverable and usable by end users. Open data are best distributed by way of APIs.

APIs are connectors between two disparate systems, much like power outlets that connect buildings to a larger electrical grid. In the federal government, as in the private sector, there has been a tremendous growth in APIs in recent years.

BENEFITS OF APIS AND VISUAL ANALYTICS

Users can bring in data from the HIW using APIs and stored processes that are run in SAS® Visual Analytics (or other software with that capacity) in order to explore and analyze data quickly and easily. Analysts can accelerate analytic computations to derive value from large amounts of diverse HIW indicators.

Figure 2. APIs "bridge the gap" between two disparate systems to effectively produce a single integrated solution. The two systems in this case are the HIW (for data) and SAS® Visual Analytics (for visualizations).

Through the use of visualizations tools such as SAS® Visual Analytics, users can produce graphs that reveal the complex relationship among multiple indicators, making an unwieldy amount of data easily digestible. Users have the ability to interact with the visualizations to change the nature of the display, filter out what's not relevant, drill into lower levels of details, and highlight subsets of data across multiple graphs simultaneously.
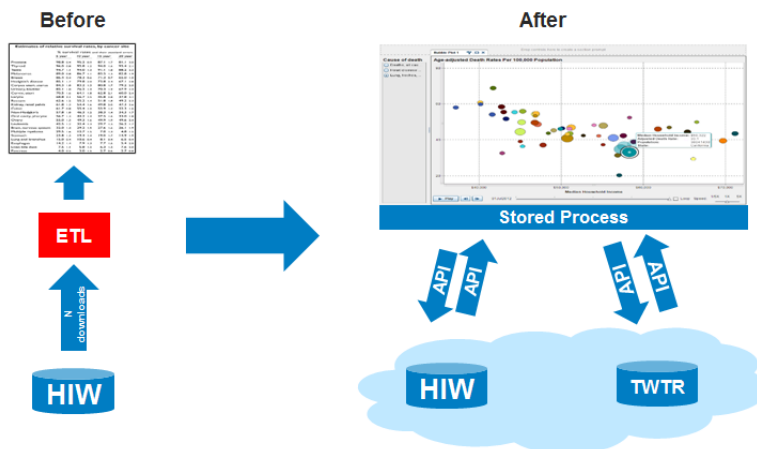

Figure 3. Gap Analysis. Current state (left) without API's and future state (right) with API's reveals the elimination of manual ETL processes (red box) associated with moving data. SAS developers, further, can write stored processes, stored in SAS Metadata, that make API calls to the HIW and other API-based data sources, such as Twitter.

API's not only integrate systems, but also mitigate against the burden of manually moving data from one system to another as shown in Figure 3. Removing manual ETL processing through API services between systems is a boon for organizations since this reduces barriers to entry as well as reduce costs, time, and errors associated with data management.

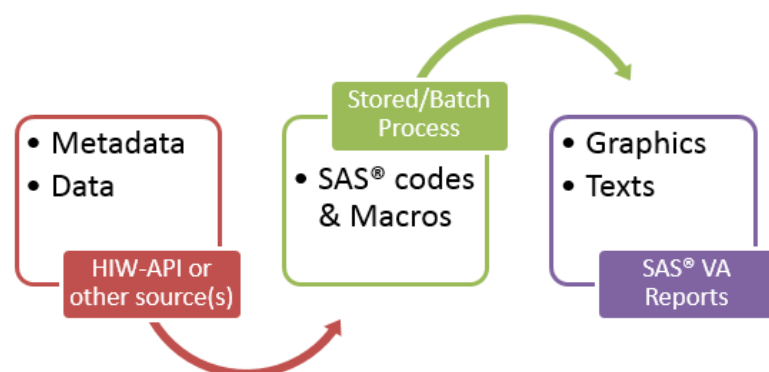## METHOD: BUILD REPORTS IN SAS® VA USING DATA FROM HIW-API SERVICES

PROCESS OVERVIEW

Figure 4. This figure illustrates how organized data such as the data available through the HIW-API service can be used in SAS® Visual Analytics reports by way of stored processes.

Figure 4 shows what is needed to generate API calls that will extract indicator data and metadata from the HIW website. SAS® codes and macros were written to make these API calls and transfer the data into the SAS® Visual Analytics server in order to build SAS® Visual Analytics reports. The whole process can be handled by a stored process which enables the creation of SAS® Visual Analytics reports. Visual Analytics reports include custom graphs such as population pyramids with textual descriptions, all of which utilize the data retrieved via the API call; this is done in an automated manner.

OVERVIEW OF THE HIW-API

The API which provides access to the underlying data within the HIW is built to be flexible so the user can access data through a variety of ways. We first recommend that users identify the indicators that they will require. Find indicators to use the HIW-API

There are multiples methods to find target indicators in the HIW. Details regarding how to find indicators in the HIW can be found in the Health Indicators Warehouse User's Guide (http://www.healthindicators.gov/App_Resources/Documents/HIW%20User's%20Guide.pdf) .

Getting started with the HIW-API

The HIW-API lets developers access the underlying data and metadata from over 1,200 health indicator descriptions. Detailed information can be found in the HIW Developers Site (http://developers.healthindicators.gov) and by following the examples provided (http://developers.healthindicators.gov/Examples/RESTIntro).

The first step required to access data through the HIW-API is to get an API Key. The key can be obtained for free by completing this short registration form: http://developers.healthindicators.gov/ApiKey. Once the key has been obtained, it can be used in the SAS macros provided in this article to access HIW data.

The documentation which covers every method and data type which the HIW-API provides access to can be found in the documentation section (http://developers.healthindicators.gov/Documentation). The reader will need to use this resource in order to apply the SAS macros provided in this article. One such macro that is key to extracting HIW data through the API is depicted in Figure 5.

```
%extractResource(RESTQuery="http://healthindicators.gov?indic=71&time=2011&fips=0", FSOutput="C:\hiw_indic71.xml"
```
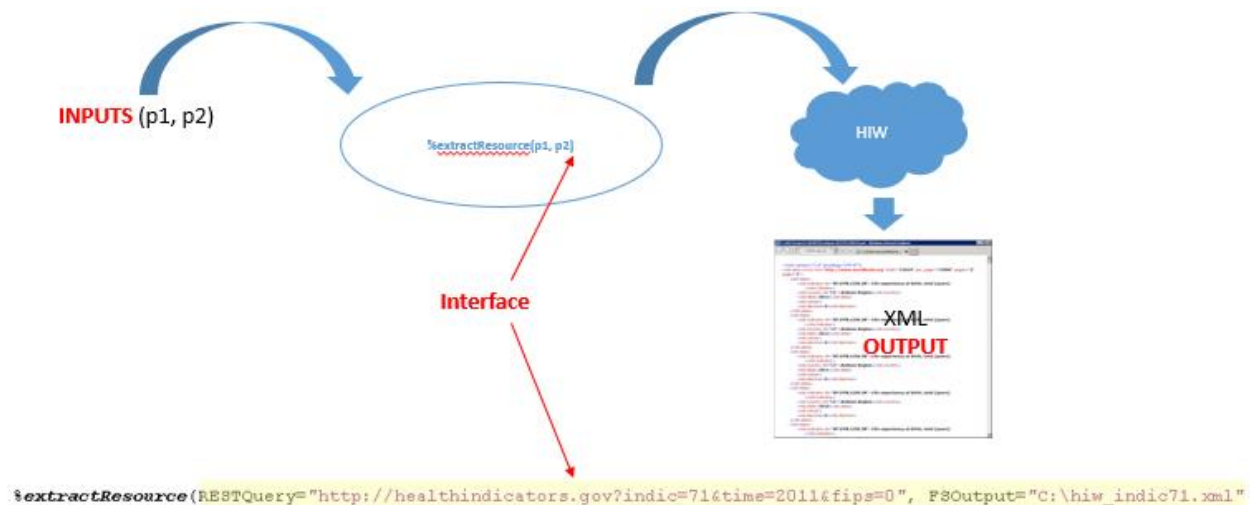
Figure 5. The %extractResource macro (or for that matter %extractHIWResource macro) takes two input parameters, RESTQuery and FSOutput, to make a REST API call to the HIW and produce filesystem (FS) output in the form of an XML file.

All of the other SAS macros discussed in this paper are available at this location:

http://tinyurl.com/SGF2016

Once the macros are unzipped in C:\, readers can review detailed instructions on how to use the macros that will make the API calls.

To get started, follow these steps after downloading the macros:

1. Go to C:\SAS\code\macros\v0_0_1_paM0\Examples and review the README.docx file.

2. Open the Driver Program (e.g., extractHIWResource_Driver_v1.sas or extractManyHIWResources_Driver_v1_0.sas). This calls the macro that will make the API calls.

3. Run the Driver Program.

In addition to extractHIWResource and extractManyHIWResources, the stored processes and batch processes discussed later, use the %convertXML2SAS, %convertManyXML2SAS, %blendDatasets, %transposeLong2Wide, and %loadDataset2LASR macros. Each of these macros contain a header with detailed descriptions, and a sample project that orchestrates all of the aforementioned macros into one single application is available in C:\SAS\code\macros\v0_0_1_paM0\Examples\Projects\Project1. The next section describes how they were used to ultimately generate a report.

ACCESS DATA THROUGH THE HIW-API AND LOAD TO SAS® VISUAL ANALYTICS TO GENERATE REPORT

This section contains a general description and a conceptual diagram of the stored and batch processes used to extract indicator data from the HIW and load the data into Visual Analytics for analysis.
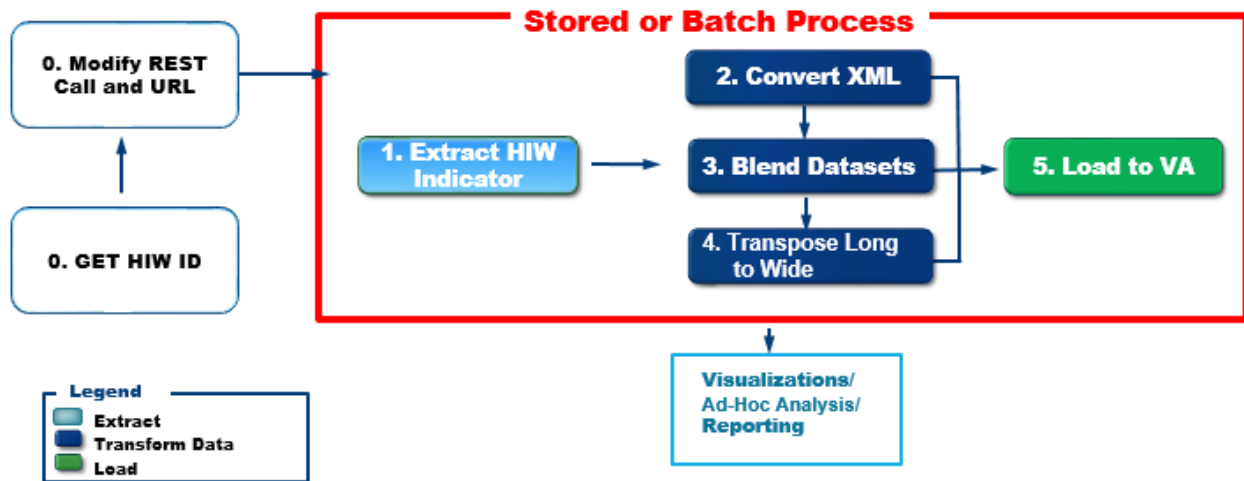
Figure 6. Flow chart for the steps involved in accessing HIW data through an API call that loads selected indicators directly into the SAS® Visual Analytics server for analysis. Note that each step also denotes a macro.

Sequential vs. parallel processing

In this section, two different approaches for running the macro that extracts HIW macros are compared: 1. Sequential; and, 2. Parallel (Figure 6)
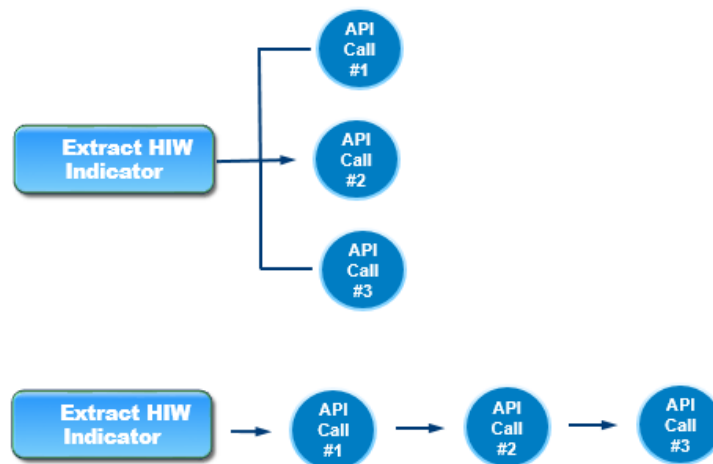

Figure 7. Sequential (top) vs. parallel or asynchronous processing (bottom).

If there are sufficient system resources for an asynchronous or parallel run of the macro used to extract HIW data through an API call, then that is the recommended approach.

Metrics showing that asynchronous runs of API calls are much faster than sequential runs are presented in Figure 8.

6
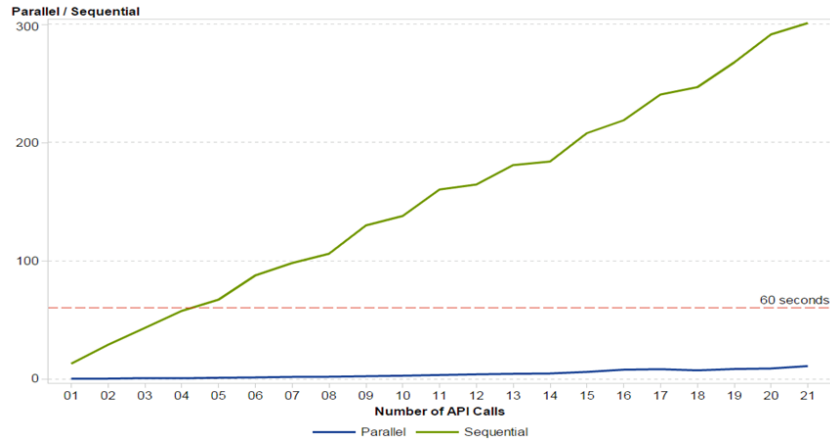
Figure 8. Notional and preliminary metrics showing that asynchronous runs of API calls are much faster than sequential runs. The x-axis in the figure shows the number of API calls made by the SAS macro used to extract HIW data, and the y-axis shows the number of seconds they took.

Stored and Batch Processes

There are two options for collecting indicator data from the HIW and loading the data into SAS Visual Analytics: a stored process and a batch process. In other words, to access HIW data via API calls, it is necessary to run a set of SAS macros either through a stored process (when the data size is relatively small) or by way of a batch process (for several indicators especially at the U.S. county-level).

These steps will make the indicators and associated data available in Visual Analytics for exploration and reporting as shown in Figure 6. It's worth emphasizing that the macro used to make the API call can be run asynchronously (i.e., in parallel) or sequentially depending on the system resources made available for the execution of the SAS macros.

The steps to run SAS codes and macros inside a stored process to generate the necessary API calls to the HIW are described in the link below.
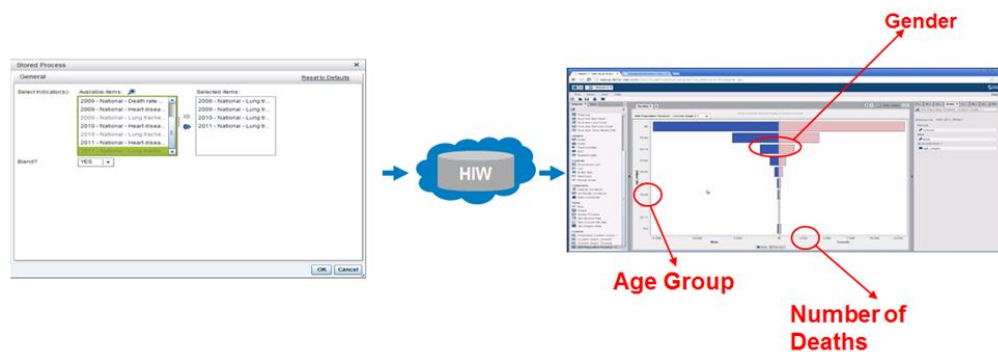
http://tinyurl.com/SGF2016

Figure 9. Running a stored process from SAS® Visual Analytics provides immediate access to the HIW data for use in SAS® Visual Analytics visualizations, such as a population pyramid to analyze age groups, gender, and number of deaths. The end-user who downloads a copy of the HIW data may prefer a stored process when response time for data access is short (e.g., less than one minute).

A stored process is code, such as macros, run in a browser by an end-user. Whether the code is run asynchronously or sequentially, end-users typically should not wait more than 1 minute for results. In cases where the API call takes much longer due to the number of records retrieved, a batch process is recommended, as shown in Display 1.

7
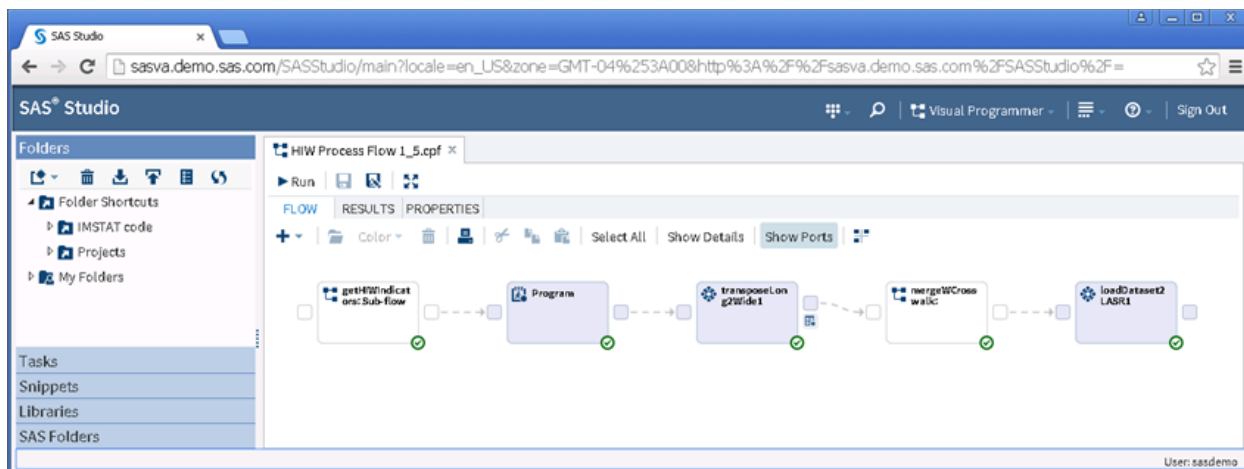
Display 1. If transferring HIW data through an API call takes a long time (because of the amount of data retrieved), a batch process in SAS® Studio is recommended.

The tinyurl link above also provides instructions on running SAS macros for HIW-API access in a batch process using SAS® Studio.

RESULTS

BUILD SAS® VISUAL ANALYTICS REPORT USING HIW-API DATA

The previous sections provided all of the building blocks for SAS® Visual Analytics reports using data from the HIW-API call. This section illustrates three possible scenarios (Figure 10) and their respective use cases to demonstrate the usefulness of SAS® Visual Analytics and stored processes to stream data directly from the HIW using an API call.
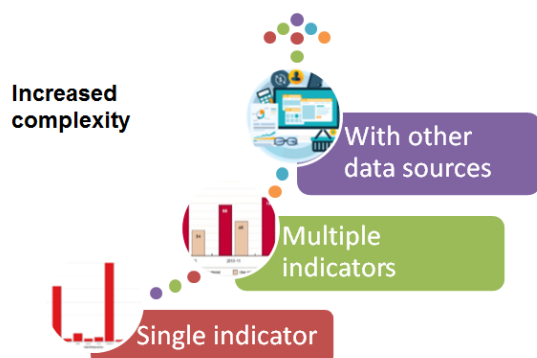

Figure 10. Three possible scenarios on using HIW-API data to build SAS® VA reports.

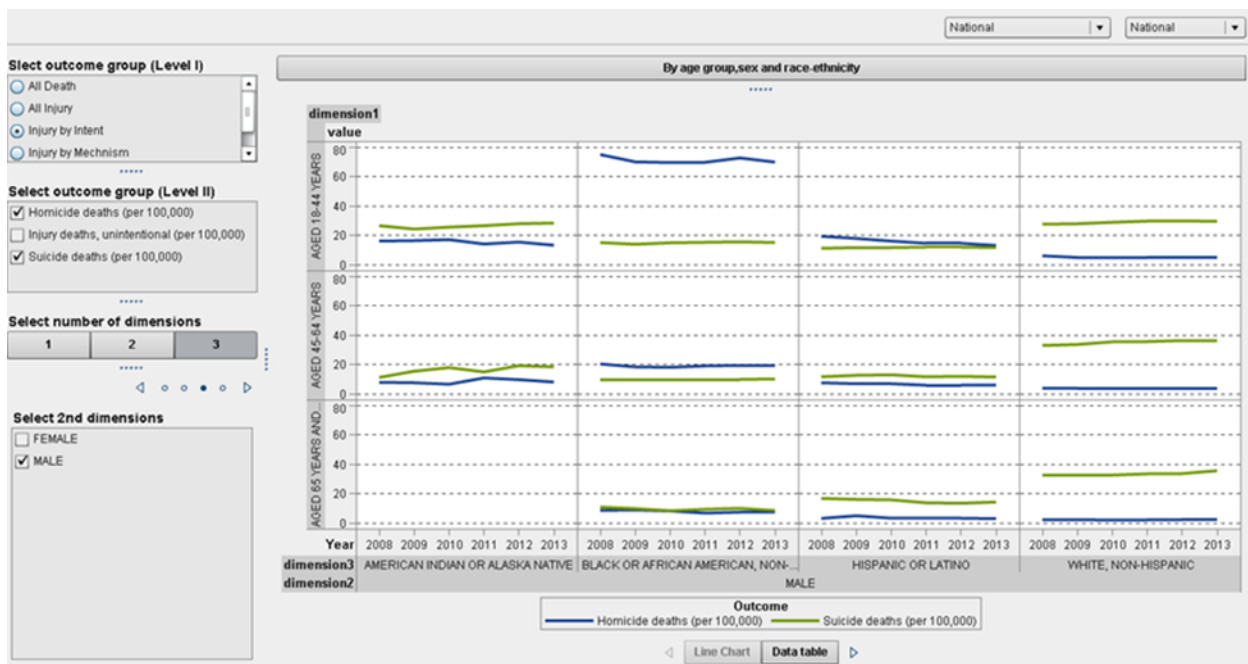### Scenario 1: Build Report using single health indicators from the HIW

Reports can be built using a single health indicator (Output 1). In this example, drug poisoning death rates in the U.S. are compared for two time periods. Statistics by age, sex, and race/ethnicity for a specific state in the time-period specified in the map can be checked by clicking on a U.S. state on the map. The report can be easily adopted by another indicator with the same data structure--e.g., motor vehicle crash related deaths—by reusing the same macros. When new data are available through the HIW-API, updates can be automated using the macros; instructions to download and update a single indicator are provided in the following link: http://tinyurl.com/SGF2016
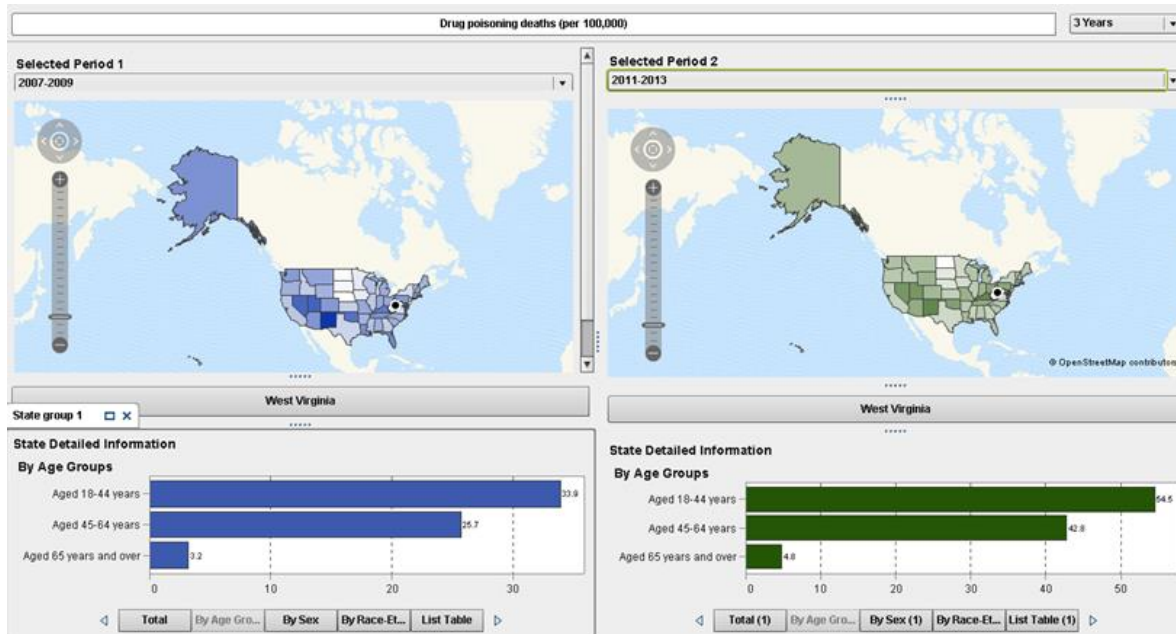
Output 1. Comparing statistics and maps for one indicator.

Scenario 2: Build Report using multiple health indicators from the HIW

Reports can be built using multiple health indicators. In this example, multiple indicators describing injury mortalities were grouped together to generate an interactive report which allowed users to explore the statistics (for example, output 2-3). When new data for any of the indicators used for the report are available through the HIW-API – updates can be automated using the macros; instructions to download and update multiple indicators are provided in the following link: http://tinyurl.com/SGF2016

Output 2. Identifying trends among groups.
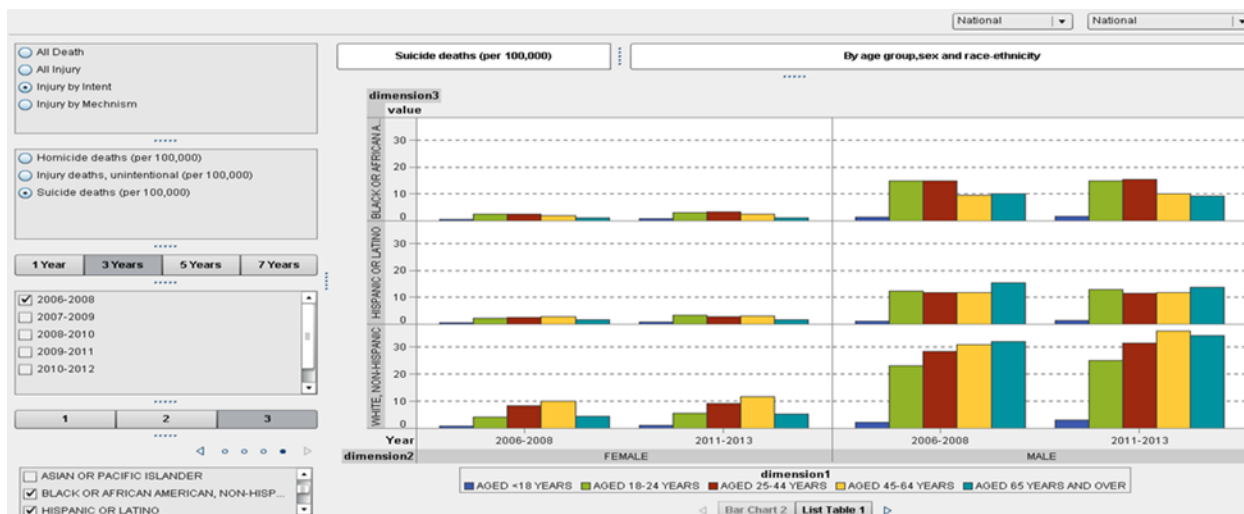
**Output 3. Comparing statistics among groups.**

Scenario 3: Build Report using health indicators from the HIW and other data sources

Reports can be built using data from HIW and other data sources. In this example, another government data, National Electronic Injury Surveillance System (NEISS) (Consumer Product Safety Commission 2016), is used with HIW data to illustrate how HIW data through API services can be used to build reports with other data sources.

### *Data from National Electronic Injury Surveillance System (NEISS)*

The Consumer Product Safety Commission's National Electronic Injury Surveillance System (NEISS) is a surveillance system for consumer products-related injuries treated in hospital emergency rooms. The system is based on a probability sample of hospital emergency department (ED) in the United States and its territories. Data collected from the NEISS sample is weighted based on the sample design to produce national estimates of product-related injuries treated in hospital emergency rooms. In addition, the NEISS data provide a source for follow-up investigations of product-related injuries. Information from NEISS is available to the public online at www.cpsc.gov/en/Research--Statistics/NEISS-Injury-Data/. Data from NEISS was downloaded through http://www.cpsc.gov/cgibin/NEISSQuery/Home.aspx and transfer to SAS data files (sas7bdat).

In the example, categories of consumer products involved in injuries were ranked using the estimated number of ED visits by age groups. The estimated rates of injuries associated with the top three product categories   are .shown in the figure by year along with the data table (output 4).
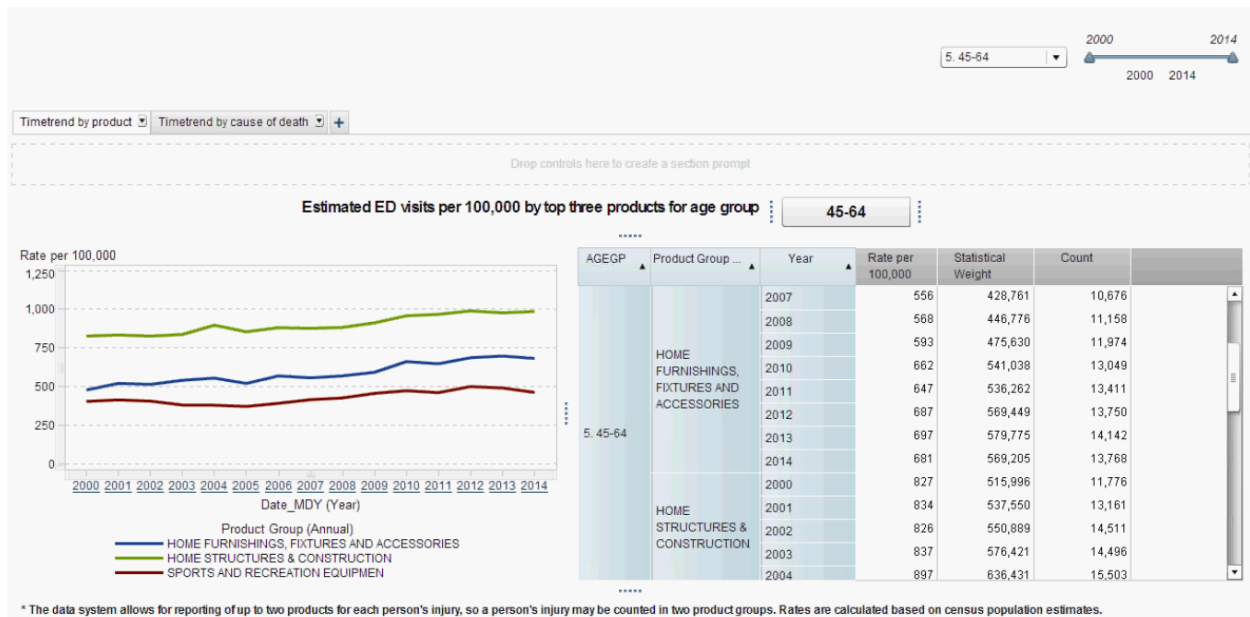
Output 4. Product Category related ED visits per 100,000 for age 45-64 NEISS, 2000-2014.

Data from HIW were brought in to provide background information about the national trends on injury death (Output 5). The two sessions were linked by the report level control keys, in this example: age groups and year. When 2015 data are released, the NEISS part of data can be downloaded through the website and the HIW part of data updates can be automated using the macros.
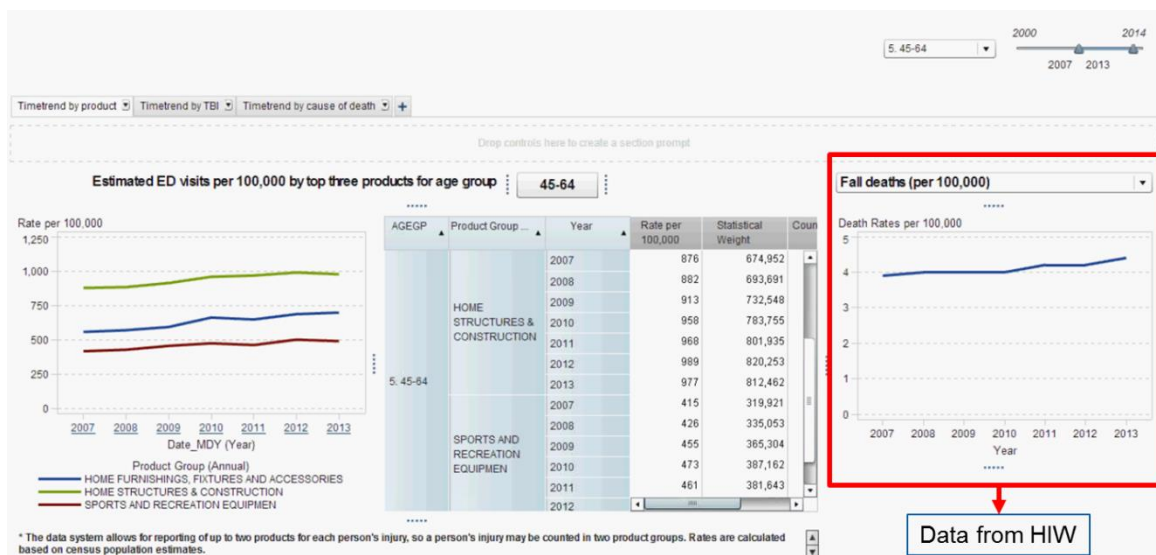
Output 5. Data from HIW added to provide information about unintentional fall death rate per 100,000.

CONCLUSION

This paper describes how stored processes and batch processes can enable integration between API data sources, such as the HIW, and SAS® Visual Analytics. We have also shown different ways of improving the efficiency of these processes by better utilizing system resources and generating metrics; more specifically, SAS macros inside a VA Stored Process were run sequentially and then run in parallel and results were compared. Not surprisingly, HIW-API calls inside a VA Stored Process are much faster when done in parallel. In addition, we show how HIW data can be used to generalize powerful

visualizations in SAS® Visual Analytics that make a large amount of data easily accessible; this is done through HIW-APIs that make data available on-demand.

Both IT professionals and population health analysts will benefit from understanding how to import HIW data into SAS® Visual Analytics using Stored Processes and APIs. These tools allow them to serve their audiences and customers. This paper provides a starting point for organizations interested in using HIW data to augment their analysis related to population health. Accessing and manipulating HIW open data with SAS® Visual Analytics can be very important to organizations that want to streamline their data management processes and lower high maintenance costs associated with data extraction and access while gaining insights into health data. Health analysts will also benefit from this paper through several use cases demonstrating the value of population health data accessed through an API with SAS® Visual Analytics.

REFERENCES

Abousalh-Neto, Nascif. The Forest and the Trees: See it All with SAS® Visual Analytics Explorer (2013). http://support.sas.com/resources/papers/proceedings13/058-2013.pdf

Consumer Product Safety Commission. National Electronic Injury Surveillance System (NEISS). 2016. http://www.cpsc.gov/en/Research--Statistics/NEISS-Injury-Data/ (accessed February 24, 2016).

Figallo-Monge, Manuel. "Macro Design and Usage in a Multi-Tier Architecture for ETL and Google Visualization API Integration". http://support.sas.com/resources/papers/proceedings12/003-2012.pdf

National Center for Health Statistics. Health Indicators Warehouse. 2016. www.healthindicators.gov (accessed February 24, 2016).

SAS Publishing. SAS 9.1 Macro Language Reference. SAS Publishing, 2004.

SAS® 9.4 Stored Processes. http://support.sas.com/documentation/cdl/en/stpug/67499/HTML/default/viewer.htm#n180km1hadyuton13

gx4bzlqhpwr.htm

SAS® Visual Analytics 7.3 -- User's Guide. http://support.sas.com/documentation/cdl/en/vaug/68648/PDF/default/vaug.pdf

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Li Hui Chen
CPSC
LChen@cpsc.gov

Manuel Figallo
SAS
Manuel.figallo@sas.com
https://www.linkedin.com/in/mfigallo