

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

Li-Hui Chen, National Center for Health Statistics, Centers for Disease Control and Prevention;

Manuel Figallo, SAS Institute Inc.;

Josh McCullough, Social & Scientific Systems Inc.

ABSTRACT

The National Center for Health Statistics' Health Indicators Warehouse (HIW) 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 it, users can view and download data and metadata for over 1,200 indicators on health status, outcomes, and determinants from approximately 180 different federal and nonfederal sources. HIW also provides access to data through the use of an Application Programming Interface (API). An API is a communication interface that applications such as SAS® Visual Analytics (SAS® VA) can use to access data from HIW and other data repositories.

This paper provides detailed information on how to access HIW data with SAS® VA in order to produce easily understood health statistics visualizations with minimal effort. It will guide readers through a process and methodology to automate data access to the HIW and demonstrate the use of SAS® VA to present health information via a web browser.

This paper also shows how to run SAS macros inside a stored process to generate the required API calls to the HIW in order to make the indicators and associated data and metadata available in SAS® VA for exploration and reporting; the macro codes are provided. Use cases are also explored in order to demonstrate the value of using SAS® VA and stored processes to stream data directly from the HIW. Dashboards, for instance, are created to visually summarize results gained from exploring the data.

Both IT professionals and population health analysts will benefit from understanding how to import HIW data into SAS® VA using Stored Processes and APIs. This paper ultimately provides a starting point for any organization interested in using HIW data to augment their analysis related to population health. Integrating HIW with SAS® VA can be very helpful 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 with visualizations. Analysts will also benefit from this paper through the use cases, which demonstrate the value of population health data accessed through an API with SAS® VA.

INTRODUCTION

Through both government and private sector efforts, more and more data are 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 through 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) 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 API. The following are covered in the methods:

- i. Process Overview. Overview of the process from gathering information from the HIW to building the report in SAS® VA.
- ii. Overview of the HIW API. Finding the necessary information for the API calls from the HIW.
- iii. Access Data through HIW APIs and Load to SAS® Visual Analytics to Generate Report. Instructions and SAS codes and macros to access data through APIs from the HIW to make the indicators and associated data available in Visual Analytics for exploration and reporting.
- iv. 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 stream data directly from 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.

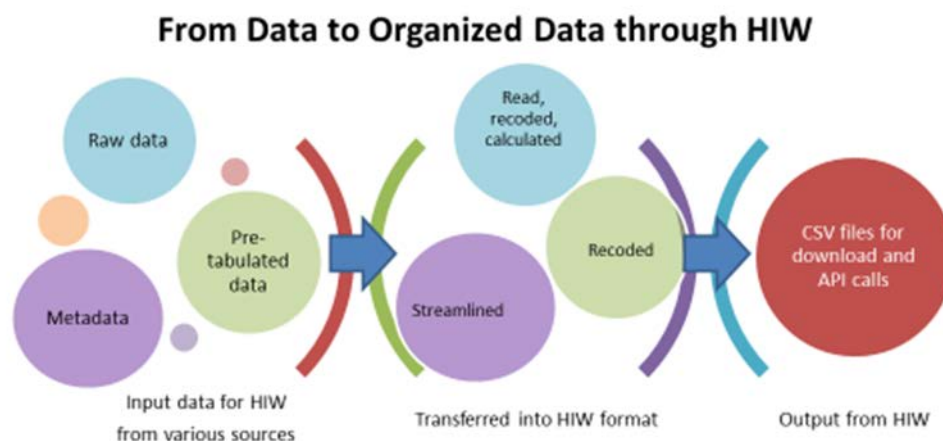


Figure 1. This figure illustrates how data are transformed into organized data and available to the general public through HIW API calls and csv files.

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 for over 1,200 indicators through the use of its API. The data can be queried in multiple ways by using available services for the 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 types in which developers are likely to be interested. 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 that trigger APIs that distribute HIW indicators.

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 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.

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 will 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.

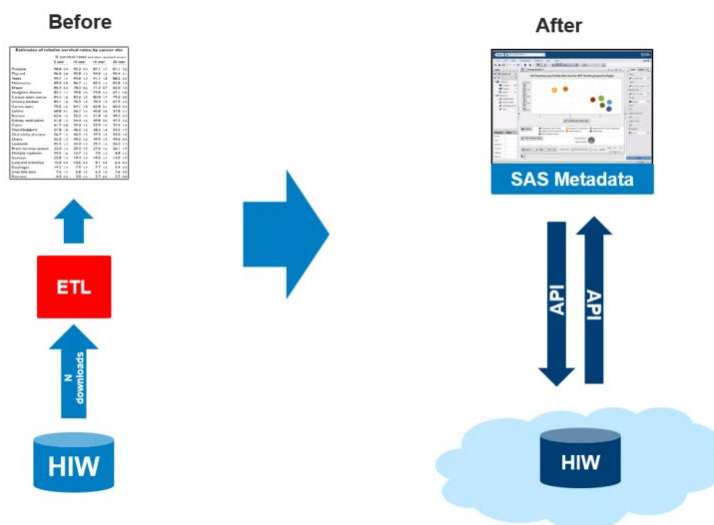


Figure 3. Gap Analysis. Current state (left) without API's and future state (right) with API's reveals the elimination of manual ETL processes (red box) associated with moving data. SAS developers, further, can write stored processes, stored in SAS Metadata, that use the HIW API.

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

i. Process Overview

From Organized Data to Information/Knowledge through SAS® VA

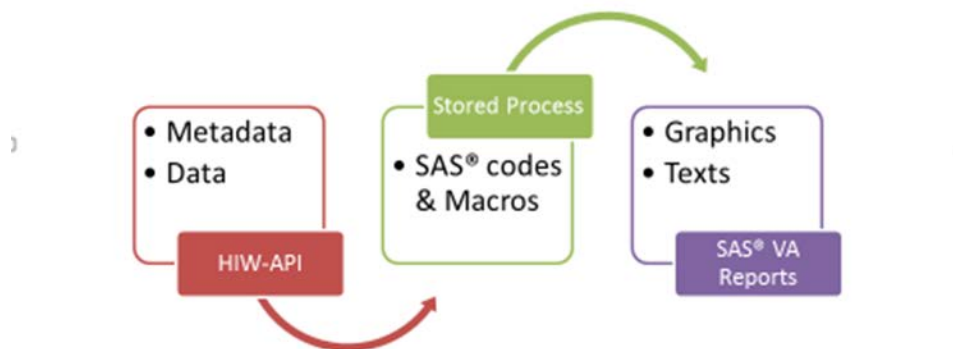


Figure 4. This figure illustrates how organized data such as the data available through the HIW-API service can be used in SAS® Visual Analytics reports by way of stored processes.

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 Visual Analytics server in order to build SAS® Visual Analytics reports. The whole process is handled by a stored process and enables the creation of SAS® Visual Analytics reports. Visual Analytics

reports include custom graphs such as population pyramids with textual descriptions, all of which are populated with the data retrieved via the API call; this is done in an automated manner.

ii. **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 to download HIW data through the API.

iii. **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.

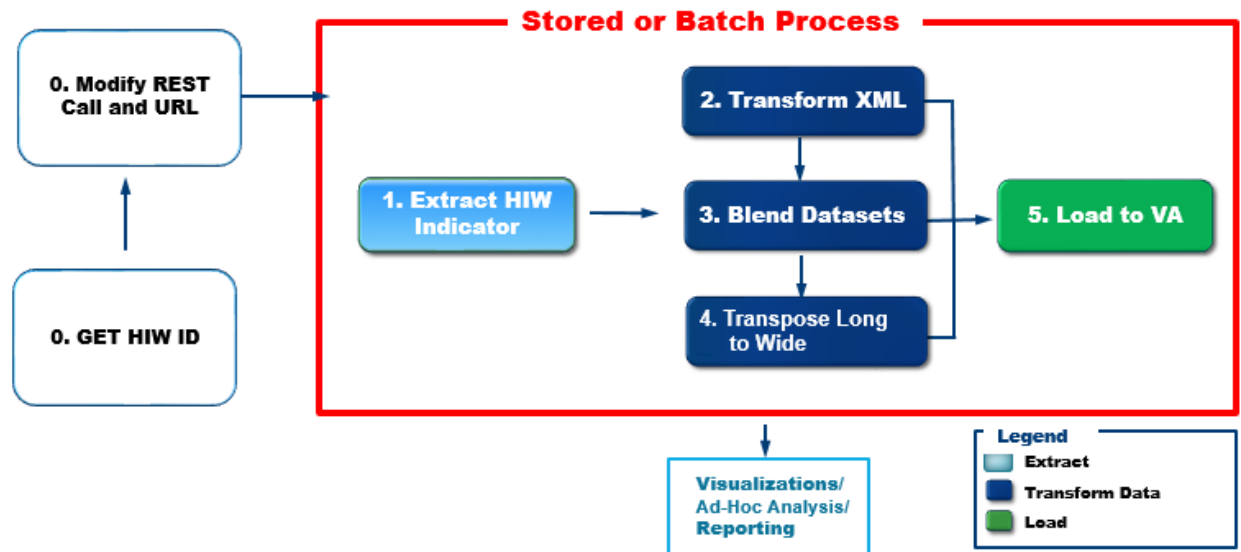


Figure 5. This figure provides a flow chart for the steps involved in accessing HIW data through an API call that loads selected indicators directly into the SAS® Visual Analytics server for analysis.

Sequential vs. parallel processing

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

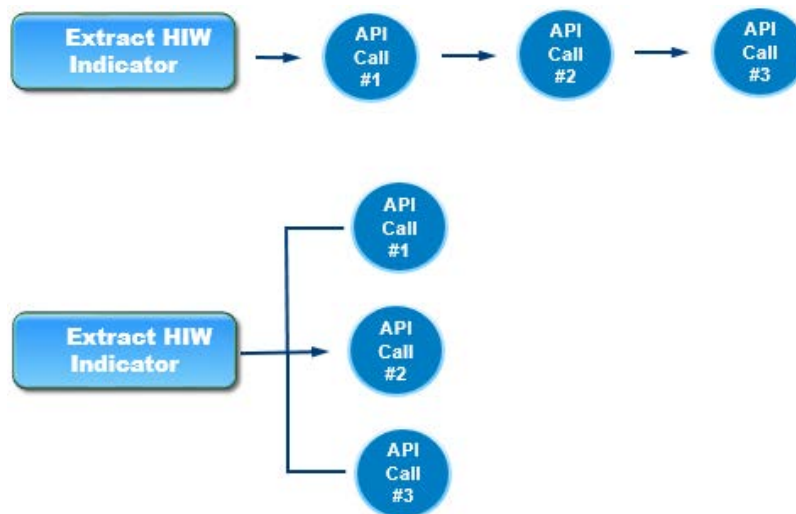


Figure 6. Sequential (top) vs. parallel or asynchronous processing (bottom).

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 7.

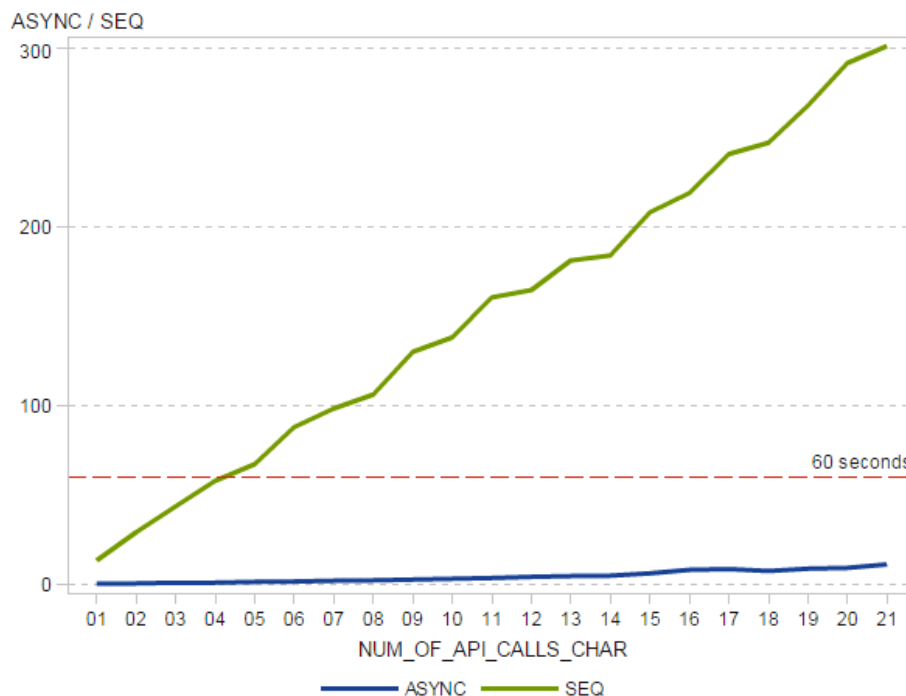


Figure 7. Metrics showing that asynchronous runs of API calls are much faster than sequential runs. The x-axis in the figure shows the number of API calls made by the SAS macro used to extract HIW data, and the y-axis shows the number of seconds they took.

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 8. 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/SESUG2015>

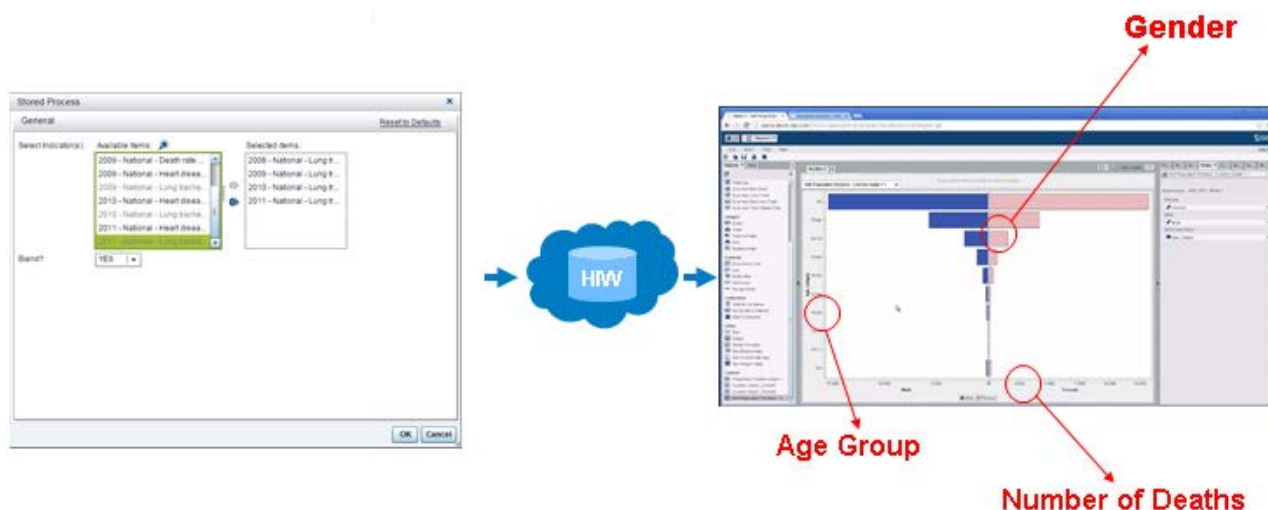


Figure 8. Running a stored process from SAS® Visual Analytics provides immediate access to the HIW data for use in SAS® Visual Analytics visualizations, such as a population pyramid to analyze age groups, gender, and number of deaths. The end-user who downloads a copy of the HIW data may prefer a stored 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 Figure 9.

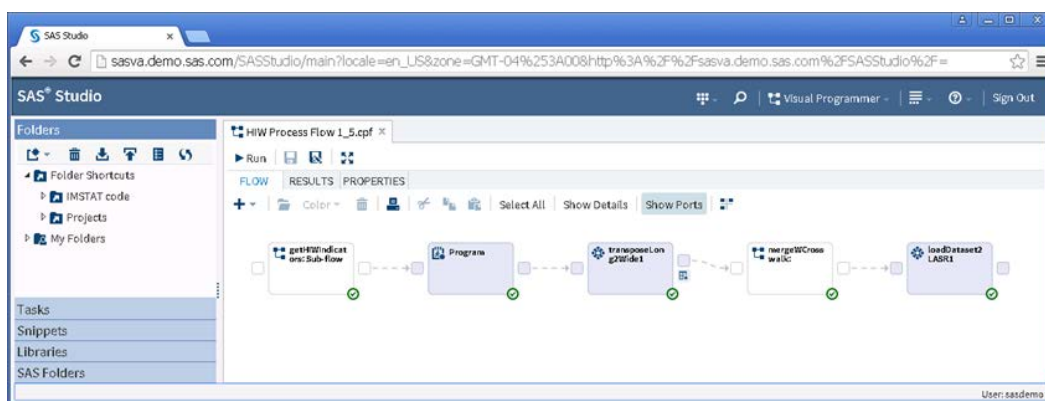


Figure 9. If transferring HIW data through an API call takes a long time (because of the amount of data retrieved), a batch process is recommended.

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

iv. **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 will illustrate 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.

Scenarios on Using HIW-API data to Build SAS®VA Reports

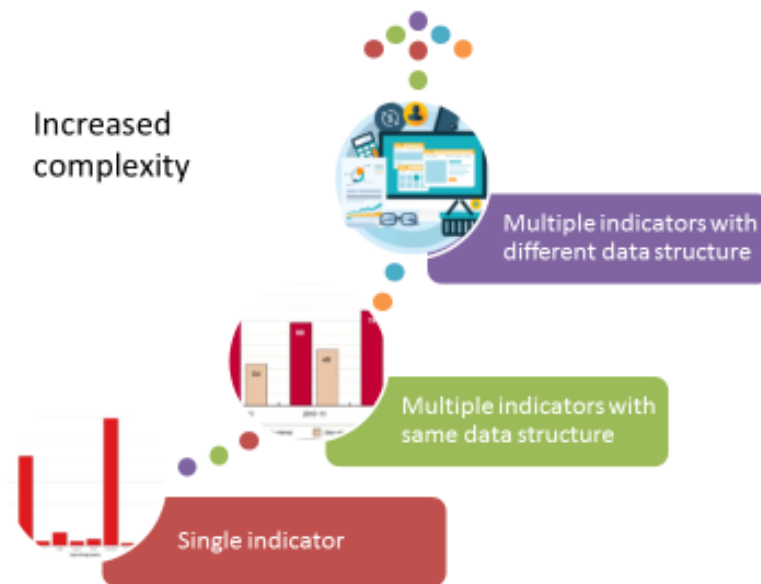


Figure 10. Three possible scenarios on using HIW API data to build SAS® VA reports

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

Reports can be built using a single health indicator (Figure 11). 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/SESUG2015>

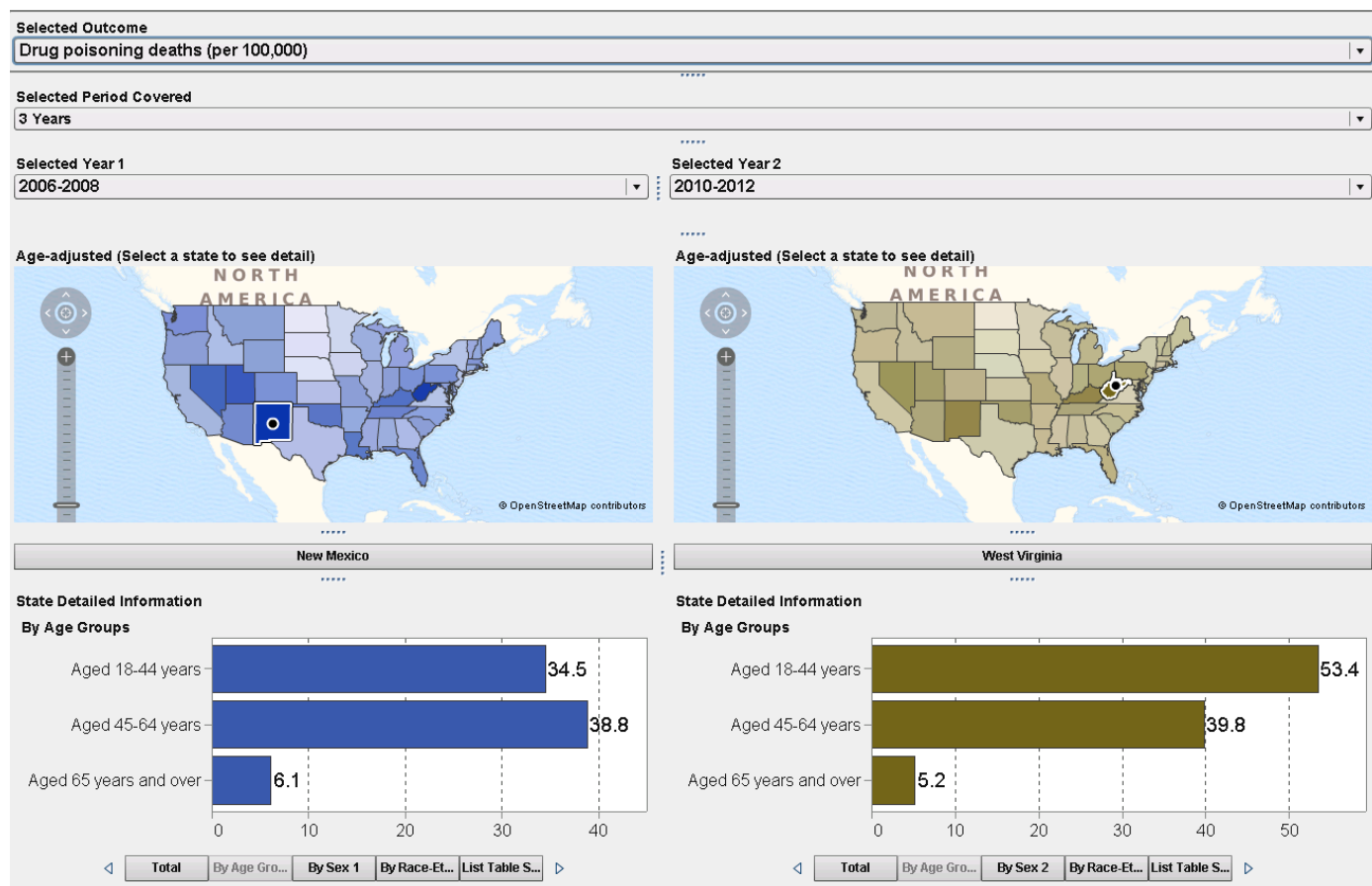


Figure 11. Comparing statistics and maps for one indicator

Scenario 2: Build Report using multiple health indicators which share the same data structure from the HIW

Reports can be built using multiple health indicators that share the same data structure. 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, see figures 12-14). 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/SESUG2015>

Select Outcome Group (Level I)

☐ All Causes
 ☒ Detailed Causes

Select Outcome Group (Level II)

☐ All Injury
 ☒ Injury by Intent
 ☐ Injury by Mechanism

Select Year Groups

1 Year

3 Years

5 Years

7 Years

Select Year to Compare

☒ 2006-2008
 ☐ 2007-2009
 ☐ 2008-2010
 ☐ 2009-2011
 ☐ 2010-2012

Select Numbers of Dimensions

1

2

3

Select Dimensions

Select 3rd Dimension

☐ American Indian or Alaska Native
 ☐ Asian or Pacific Islander
 ☒ Black or African American, non-Hispanic

Select Outcome Group (Level III)

Homicide deaths (per 100,000)

Injury deaths, unintentional (per 100,000)

Other Intent

Suicide deaths (per 100,000)

Suicide deaths (per 100,000)

By age group, sex and race-ethnicity

Dimensions

Black or African American

Hispanic or Latino

White, non-Hispanic

2006-2008

2010-2012

2006-2008

2010-2012

Female

Male

Dimension1

☒ Aged <18 years
 ☐ Aged 18-24 years
 ☐ Aged 25-44 years
 ☐ Aged 45-64 years
 ☐ Aged 65 years and over

12



Figure 14. Focusing on one sub-group (e.g., female)

Scenario 3: Build Report using multiple health indicators with different data structures from HIW

Reports can be built using multiple health indicators with different data structures. If the purpose of the visualization is to compare the indicators, then the first step would be to understand what characteristics the indicators shared and use the filter function to make the indicators comparable. Figure 15 is an example: since one of the indicators only includes data for those aged 65 and over, the other indicators were filtered to only include data for those 65 and over.

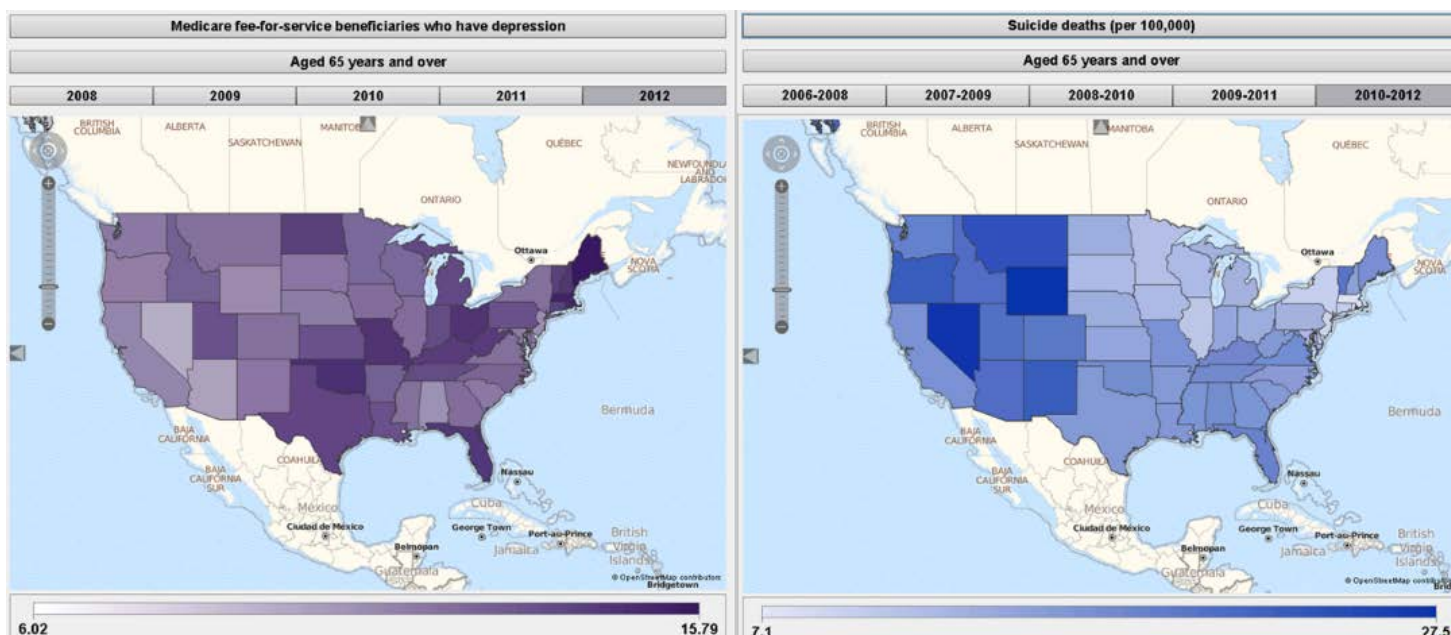


Figure 15. Comparing statistics with two indicators with different structures

Since SAS® Visual Analytics has the ability to bring in multiple data sources to build the visualization, indicators with different structures can be brought in to build the report. Figure 16 is an example, which brought in four different indicators to build the report. Additional indicators were brought in to add information and do not need to be filtered.

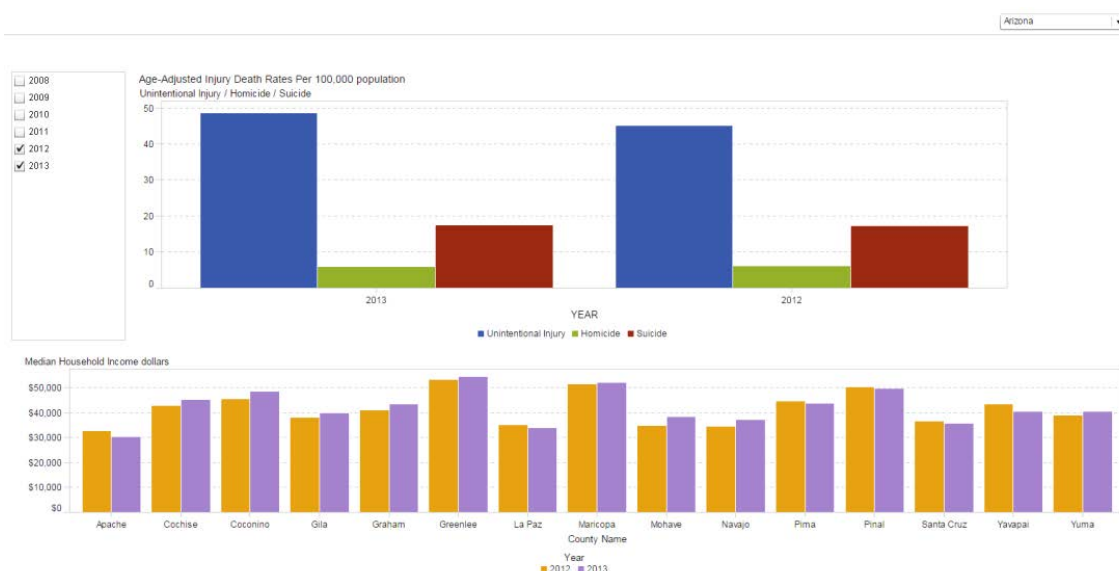


Figure 16. Present statistics with different structure.

When new data for any of the indicators used for the report are available through 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/SESUG2015>

Since data with different structures are used to build the report, some modification will be needed. Depending on the differences between the data structures, the programs will need to be adjusted accordingly. However, once the stored processes are built, the same routine can be reapplied.

RESULTS

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.

CONCLUSION

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.

ACKNOWLEDGEMENTS

We would also like to thank all of our colleagues at the Centers for Disease Control and Prevention, SAS Federal Government Business Unit, and Social & Scientific Systems Inc. In particular we would like to thank Jim Craver, Kate Brett, Larry Bowers, Derek Britton, Esmeralda Stuk, Nicholas Cardamone, and Roy Fleischer, all of whom provided invaluable feedback and assistance.

CONTACT INFORMATION

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

Li-Hui Chen
Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Rd, Hyattsville, MD 20782
eyx5@cdc.gov

Manuel Figallo
SAS Federal
1530 Wilson Blvd, Suite 800
Arlington, VA 22201
Manuel.Figallo@sas.com
<http://www.linkedin.com/mfigallo>

Josh McCullough
Social & Scientific Systems Inc.
8757 Georgia Avenue
Silver Spring, MD 20910
JMcCullough@s-3.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.

Other brand and product names are trademarks of their respective companies.