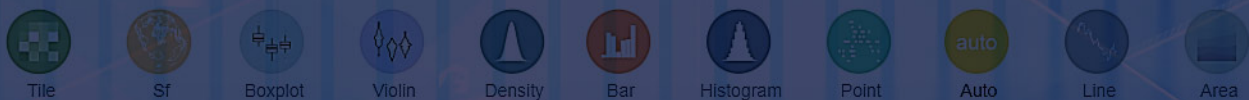


BioSense Platform User Manual for the Data Quality Dashboard

February 2020



Technical Assistance: support.syndromicsurveillance.org

The National Syndromic Surveillance Program (NSSP) promotes and advances development of the cloud-based BioSense Platform, a secure integrated electronic health information system that hosts standardized analytic tools and facilitates collaborative processes. The BioSense Platform is a product of the Centers for Disease Control and Prevention (CDC).

CONTENTS

1 Overview, Navigation, and Functionality, 1

- 1.1 Feeds, 1*
- 1.2 Filters, 2*
- 1.3 Navigation, 3*
- 1.4 Graph Functionality, 5*

2 Home Page, 8

3 Data Flow Visualizations, 10

- 3.1 Introduction to Data Flow Page, 10*
- 3.2 Data Flow Banner, 10*
- 3.3 Data Flow Overview, 11*
- 3.4 Visit Count Comparison, 12*
- 3.5 Filter Reasons, 12*
- 3.6 Exception Reasons, 13*

4 Timeliness, 15

- 4.1 Timeliness Banner, 15*
- 4.2 Overall Timeliness, 16*
- 4.3 Timeliness Over Time, 17*

5 Completeness, 19

- 5.1 Completeness Banner, 20*
- 5.2 Completeness by Element, 20*
- 5.3 Overall Completeness Over Time, 21*
- 5.4 Completeness by Element Over Time, 21*

6 Validity, 22

- 6.1 Validity Banner, 23*
- 6.2 Overall Validity, 23*
- 6.3 Overall Invalidity, 24*
- 6.4 Missing, Valid, Invalid by Element, 25*
- 6.5 Overall Validity Over Time, 26*

7 Plot Builder Overview, 27

8 Plot Builder Reference Guide, 29

- 8.1 Step 1: Select the Dataset, 29*
- 8.2 Step 2: Drag and Drop the Variables, 29*
- 8.3 Step 3: Select Graph Type, 30*
- 8.4 Step 4: Label the Graph, 30*
- 8.5 Step 5: Change the Graph's Theme, 31*
- 8.6 Step 6: Subset Data, 32*
- 8.7 Step 7: Export Graph, 32*
- 8.8 Download Dashboard Data, 33*

9 Map View, 34

BioSense Platform User Manual for the Data Quality (DQ) Dashboard



Prerequisites

Data Quality (DQ) Dashboard is best viewed with the Microsoft Edge or Google Chrome browser. Other browsers (e.g., Internet Explorer or Safari) may not display data, graphics, and reports as intended. Here is the URL:

<https://dashboards.syndromicsurveillance.org/app/dqdb>

BioSense Platform Site Administrator Account: Site administrators have access to the DQ Dashboard. If site administrators want to grant access to other users within their sites, submit a ticket to the NSSP Service Desk <http://support.syndromicsurveillance.org> with information on who should be granted access. Please include the user's username and email address.

Users will be able to access the DQ Dashboard via a link in the Home tab in the Access & Management Center (AMC).

1. Overview, Navigation, and Functionality

The DQ Dashboard is an interactive tool developed by the National Syndromic Surveillance Program (NSSP) to help site administrators identify potential data processing issues. The DQ Dashboard helps to ensure useful syndromic data by measuring the timeliness, completeness, and validity of data that is processed on the BioSense Platform.

Note that data processing issues, timeliness, completeness, and validity are assessed at the *primary facility* level. Data sent from associated facilities are rolled up to their primary facility's C_BioSenseFacility_ID as registered in the Master Facility Table (MFT). When data are subset (e.g., via filters) to a primary facility, data from its associated facilities are included.

1.1 Feeds

The NSSP BioSense Platform receives syndromic surveillance information provided by sites. The NSSP groups facilities under a single *administrative* authority called a site. A site is a public health department (local, state, or county) that serves as the administrative hub for any facility with which it has made agreements to provide data from that facility to the BioSense Platform. Each site is responsible for one or more data feeds that transmit data, and each feed can contain data from one or more facilities.

Initially, feed names are created and added to the MFT by the BioSense Platform Onboarding Team in coordination with each organization's site administrator *before* site onboarding. Feeds can be associated directly with one facility or may act as a central feed for several facilities. New feeds may be added after onboarding, and feeds may be renamed in some situations.

When the dashboard is displayed for a site, **all feeds associated with that site in the MFT are displayed by default**. The dashboard is populated by underlying tables that contain metrics (data flow, timeliness, completeness, or validity) by facility level by day (either arrived_date or c_visit_date).

1.2 Filters

The facility field dropdown is filtered when another variable (feed, parent organization, vendor, or facility type) is selected. Only facilities with the attribute(s) selected by other filters (as recorded in the MFT) are shown. If data are filtered by single or multiple facilities, the dashboard will display metrics for the selected facilities. If one or multiple feeds are selected, only facilities associated with selected feeds will be available for selection.

If a feed, vendor, parent organization, or facility type filter value is selected, the front end identifies all active facilities associated with the selected value in the MFT and displays all data from those facilities. For example, if a user filters on Feed A, the MFT is referenced to identify what active facilities are registered to Feed A and displays all data for facilities registered in the MFT under Feed A.

In the future, an option to allow users to review existing data for feeds that have been renamed is being planned.

► Filters and Filtered Data

Filtered data only include information about the *feed*. There is no information that relates to the facility or its characteristics. Therefore, only the feed filter applies to filtered data.

► Filters and Exceptioned Data

Depending on the specific exception reasons, the exceptioned messages could be missing facility information or contain facility information not registered in the MFT. Therefore, use of facility-related filters such as Facility Name, Parent Organization, Vendor, or Facility Type will only include exceptioned data for which this level of information is available.

QUICK
TIPS



For questions about data processing, data quality issues, or enhancement requests, please submit a ticket to the NSSP Service Desk <http://support.syndromicsurveillance.org>.

1.3 Navigation

Use Figure 1 below to map each section of the Home page to its description on the following page.

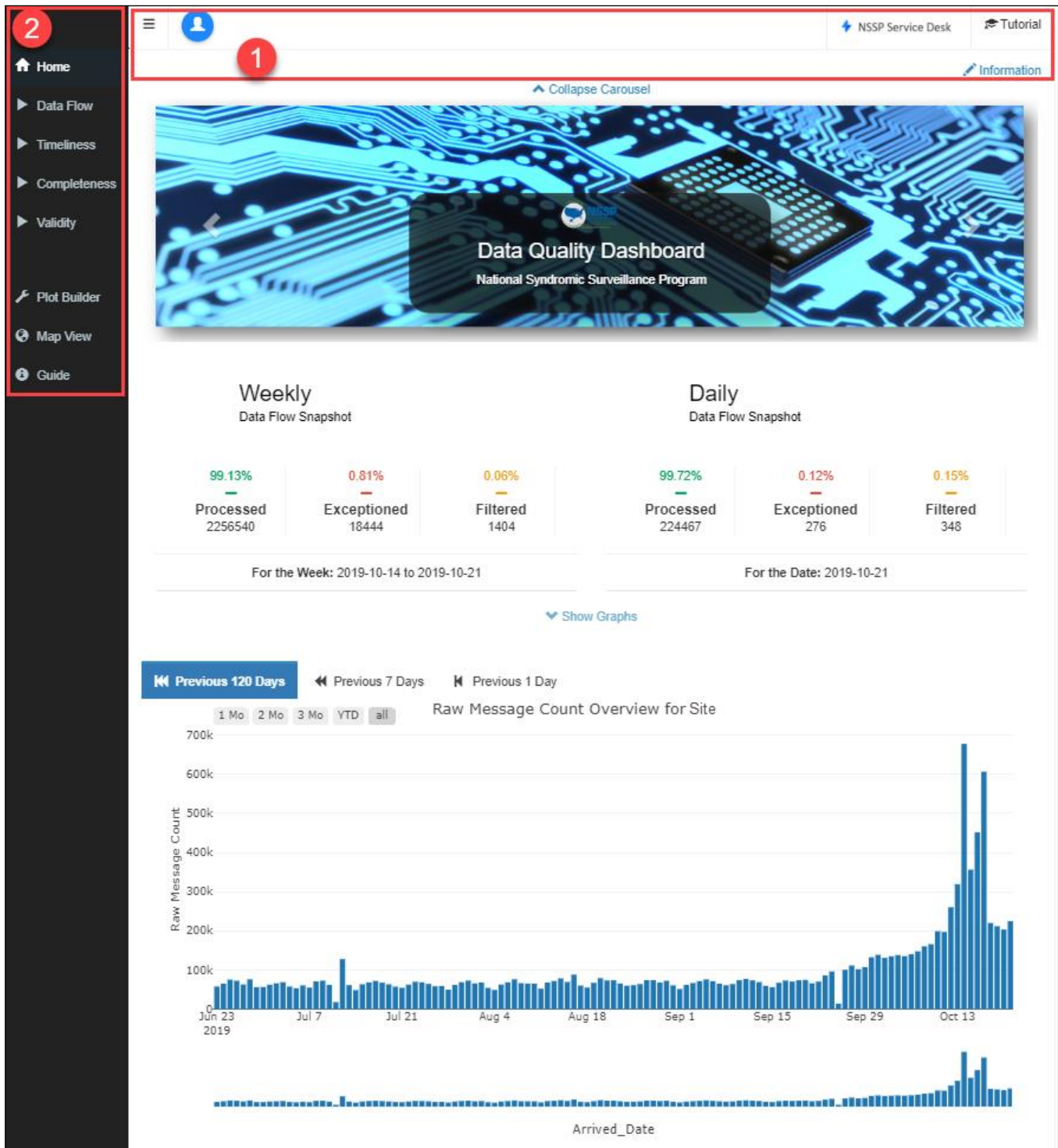


Figure 1. Data Quality Dashboard Home Page

1. **Assistance Bar:** The assistance bar at the top of every page contains options for getting information.
 - a. **Information:** Augments the Web page being viewed by defining terms (i.e., processed, exceptions, filtered, timeliness, completeness, and validity) and describes what data are available (including timeframe and when data are refreshed).
 - b. **Service Desk:** Provides a direct link to the NSSP Service Desk for questions about data processing status, data quality issues, and enhancement requests.
 - c. **Tutorial:** Provides high-level guidance on dashboard functionality.
2. **Navigation Pane:** The navigation pane on the left side of every page allows you to navigate to the following pages:
 - a. **Home:** Returns to the Data Quality Dashboard Home page from other pages.
 - b. **Data Flow:** Provides an overview of your site's data processing, including information about processed, filtered, and exceptioned message counts and percentages.
 - c. **Timeliness:** Provides an overview of your site's data timeliness, including the number and percent of visits received within 24 hours, between 24 and 48 hours, and after 48 hours.
 - d. **Completeness:** Provides an overview of your site's data completeness, including the percent of visits that contain at least one non-null value for specified data elements.
 - e. **Validity:** Provides an overview of your site's data validity, including the percent of messages that contain a valid value for specified data elements.
 - f. **Plot Builder:** Enables all users to create their own charts.
 - g. **Map View:** Displays a geographical representation of facilities registered in the Master Facility Table (MFT), their current facility status values (Planned, Not Planned, Onboarding, Active, Inactive), and respective visit counts.
 - h. **Guide:** Shows an overview of the dashboard and explains its navigation and functionality.

Note: The assistance bar and navigation panes are available in every dashboard page.

1.4 Graph Functionality

Graphs may include tabs, filters, radio buttons, toggle switches, and other interactive functionality to control data displayed in each visualization (Figure 2). You may subset data to improve analysis or identify concerns affecting data processing, timeliness, completeness, and validity.

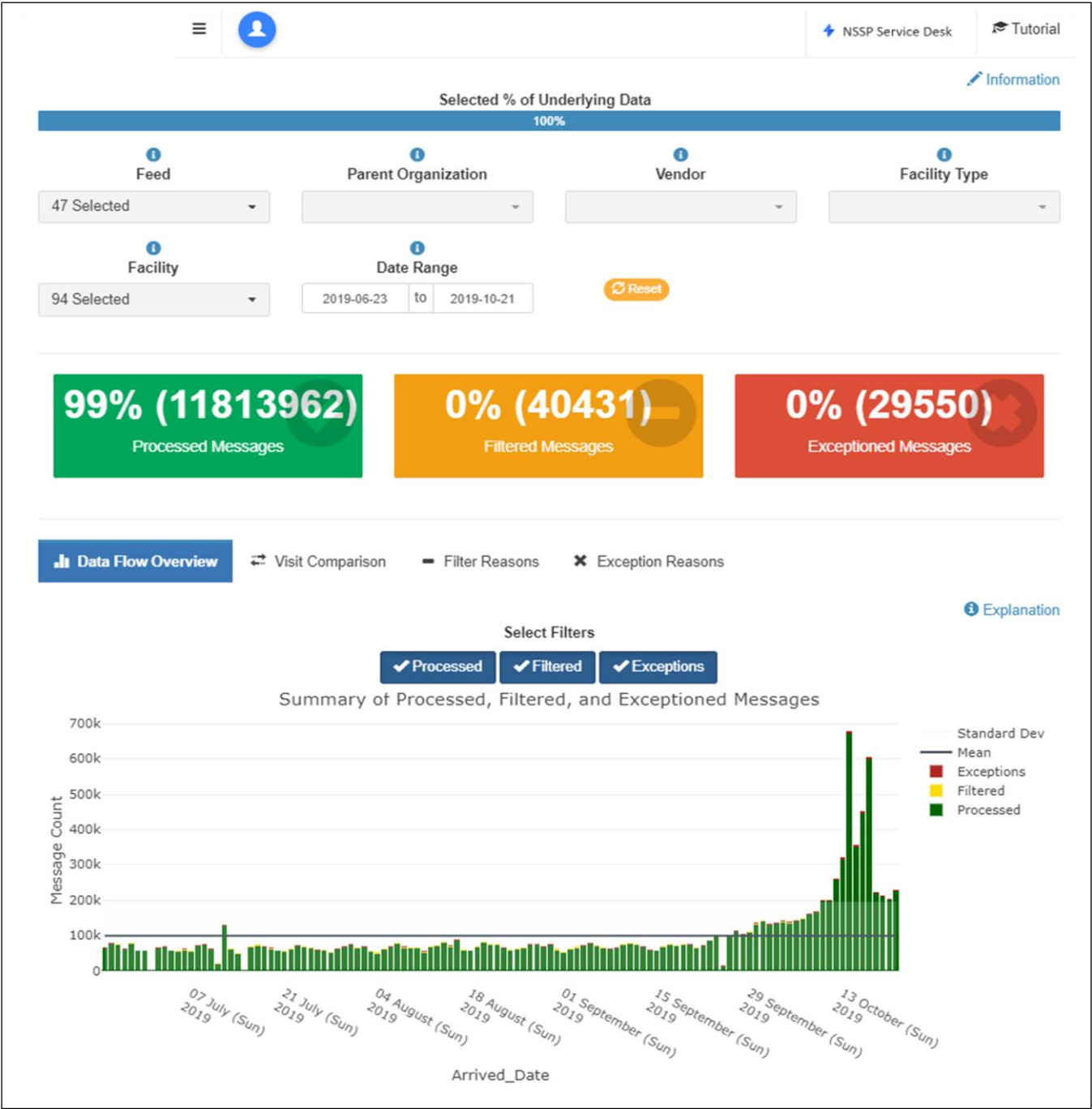



Figure 2. Graph Functionality

1. **Filters:** All pages other than Home include filters for Feed, Parent Organization, Vendor, Facility Type, Facility, and Date Range that you can use to subset data in your visualizations.

In addition, the Completeness and Validity pages each include filters for data elements both individually and in groups. By default, the data elements in the group named *Priority 1* will display.

Note: To reset filters, click the **Reset** button.

- a. **Infotips:** Each filter has information describing it and how it can be used to subset your data. Click on the  icon to view. Click again to hide it.
- b. **Percent Bar:** The bar above the filters displays the percent of underlying data that are included in your visualizations based on your filter selections. You may view, at most, 120 days of data (i.e., when the bar indicates 100%, 120 days of data are being displayed).
- c. **Radio Buttons:** The Timeliness page includes radio buttons below the filter bar that allow you to aggregate data by facility, feed, vendor, or parent organization.

Note that data processing issues, timeliness, completeness, and validity are assessed at the primary facility level. If data are subset to a primary facility via the Facility filter, data from associated facilities will be included. The Facility Type filter is associated with the primary facility's facility type and will include all associated facilities registered to the primary facility.

2. **Tabs:** Tabs are available below the filters on the Data Flow, Timeliness, Completeness, and Validity pages. These tabs allow you to view different visualizations within each page.
3. **Toggle Switches:** The Timeliness, Completeness, and Validity pages provide toggle switches. Use these toggle switches to subset data by predetermined timeliness and validity thresholds.
4. **Interactive Functionality:** All graphs are dynamic and interactive. You may subset the data displayed in the graphs by clicking the legend or tiles to add or remove data. To reset legend selections, double click anywhere on the legend. All the visualizations have interactive capabilities as indicated on the mode bar in the top right corner. Note: For the mode bar to display, you must hover over the graph.

In Figure 3, from left to right, the mode bar offers the following functionality:



Figure 3. Mode Bar

- a. **Download Plot as PNG:** Click to download the plot in PNG format.
- b. **Zoom:** Click and hold your mouse, moving across the region, to zoom in on a region of a graph. Double-click anywhere on the plot to return to the original view.
- c. **Pan:** Select to move horizontally and vertically around regions of your graph. Click and hold your mouse to explore data. Double-click anywhere on the plot to return to the original view.
- d. **Box Select:** Click to subset the data displayed in the graphs by selecting parts of the graph in a box.
- e. **Lasso:** Click to subset the data displayed in the graphs by selecting parts of the graph.
- f. **Zoom In:** Click on the plus sign (+) button to zoom in. Axes labels and annotations will remain the same size to preserve readability. Axes labels will automatically optimize as you zoom in.

- g. **Zoom Out:** Click on the minus sign (-) button to zoom out. Axes labels and annotations will remain the same size to preserve readability. Axes labels will automatically optimize as you zoom out.
 - h. **Auto Scale:** Click to return the plot to an optimized setting that includes all the viewable data regardless of the Axes Range setting.
 - i. **Reset Axes:** Click to include your Axes Range, if this has been set. If it has not been set, the graph will return to an optimized setting that includes all viewable data (i.e., auto scale).
 - j. **Toggle Spike Lines:** Click to display lines on the graph indicating the x-axis and y-axis coordinates.
 - k. **Show Closest Data on Hover:** Click to display data for just the one point under the cursor.
 - l. **Compare Data on Hover:** Click to show data for all points with the same x-value.
 - m. **3D Charts:** Click for additional options for controlling rotations and toggling between the default view and your last saved view.
5. **Explanation Button:** Explains the information being conveyed in the visualization, how to isolate or subset data in the graph, and available data (i.e., timeframe, when data are refreshed). A modal is available for each visualization in the upper-right corner.

2. Home Page

The Home page (Figure 4) allows users to review BioSense Platform announcements and view high-level processing metrics.

The banner in the center of the Home page contains platform announcements that could affect data processing or data quality. If data processing or data quality issues are noted, please keep that in mind as you review the various dashboard pages: Data Flow, Timeliness, Completeness, and Validity.

The Home page also contains high-level processing metrics (i.e., processed, exceptioned, and filtered percentages and counts) for messages that arrived the previous day and week. The *previous day* is defined as 00:01 UTC to 00:00 UTC the previous day. The *previous week* is defined as a rolling 7-day period as of 00:00 UTC the previous day to 7 days prior. Please review the Data Flow page Calculation Information for more details on how processed, filtered, and exceptioned messages are defined.

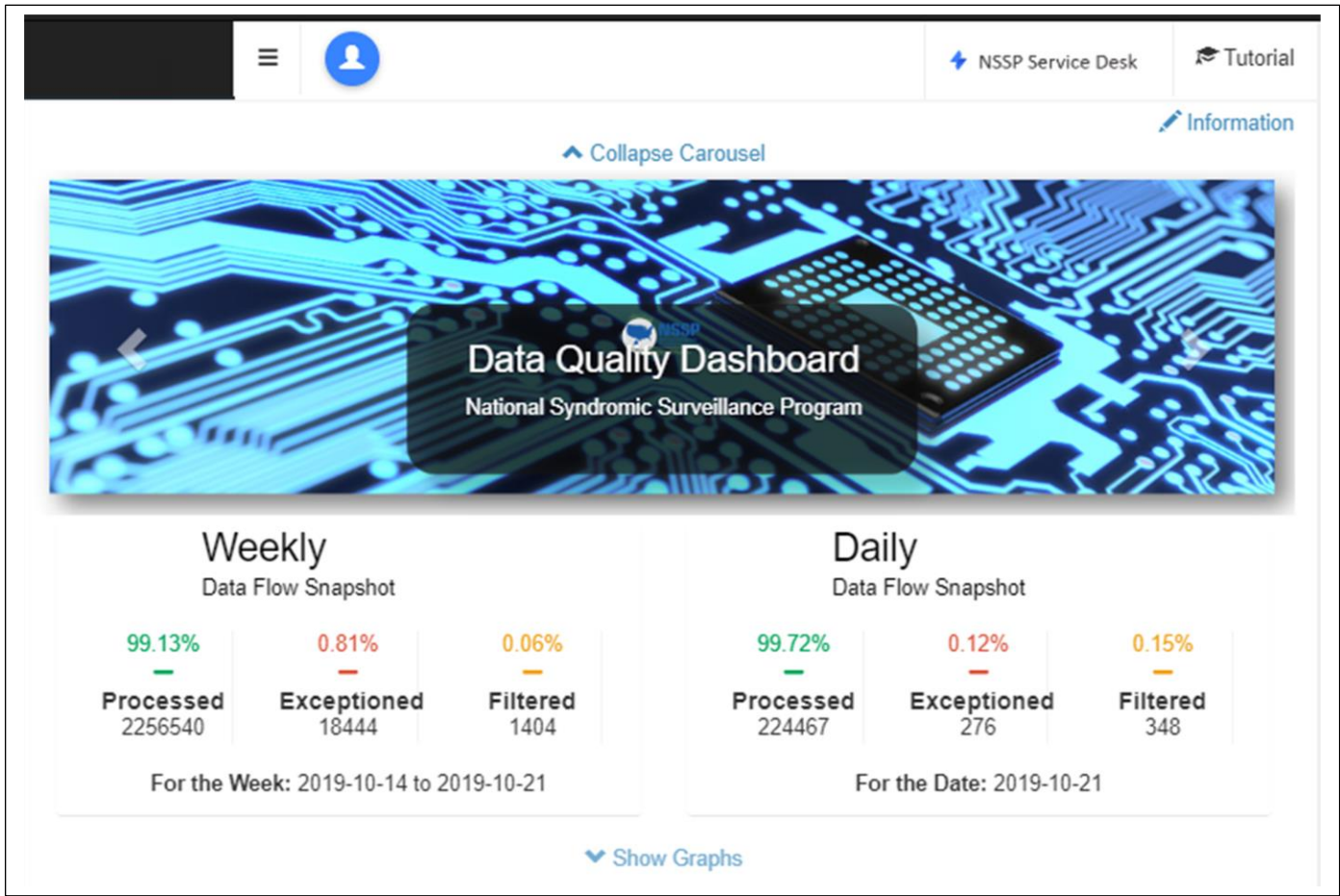


Figure 4. Home Page Processing Metrics

Click **Show Graphs** on the Home page to display the Raw Message Count time series (Figure 5). The Raw Message Count graph displays the number of messages that arrived on the platform by arrived_date. The y-axis shows the message count, and the x-axis shows the arrived_date. The timeline below the graph allows you to control the timeframe of data depicted.

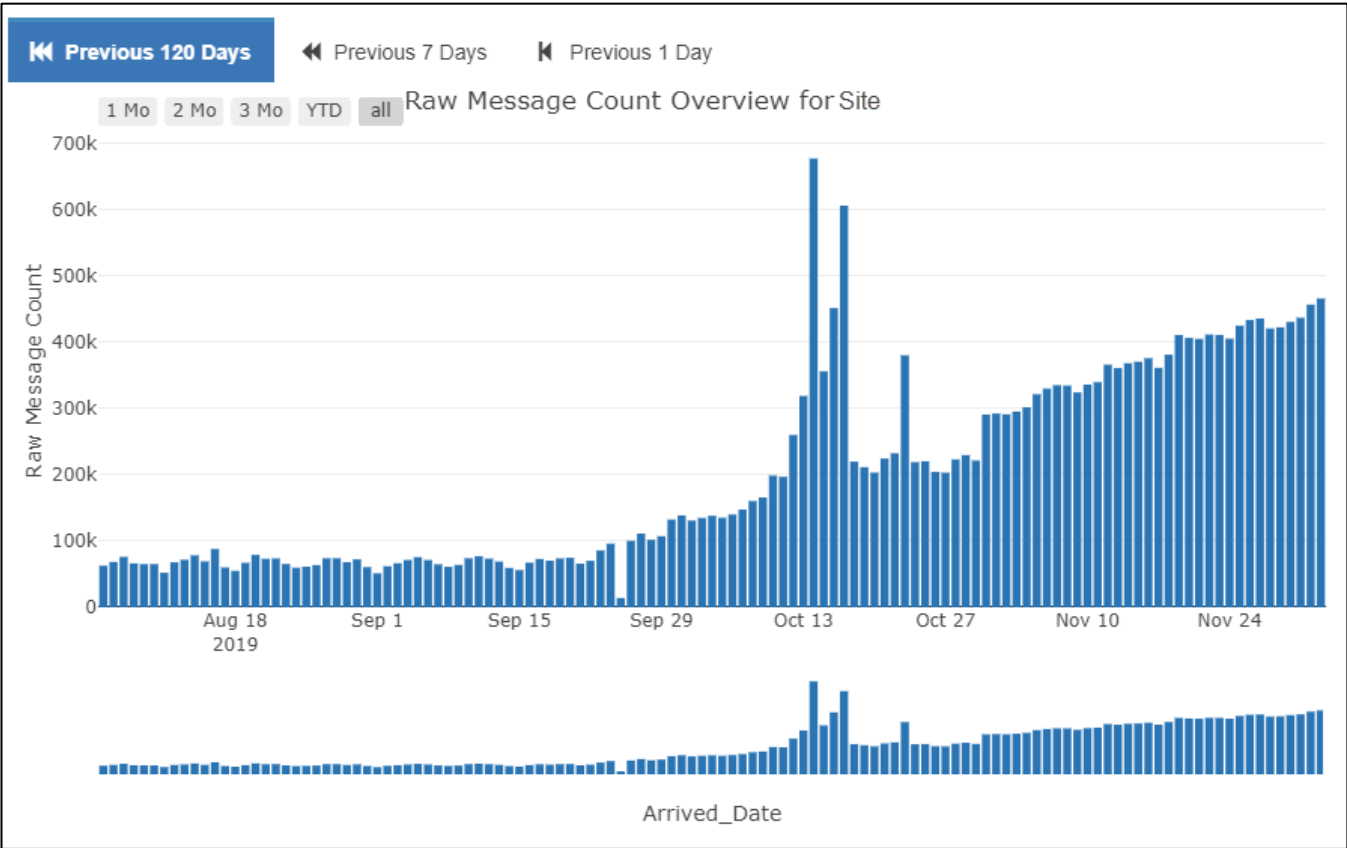


Figure 5. Home Page—Show Graphs

3. Data Flow Visualizations

3.1 Introduction to Data Flow Page

The Data Flow page allows users to identify processing issues within their site by visualizing the total message count and percentage of messages that were processed, filtered, or exceptioned.

As incoming messages are processed into the Raw table, the BioSense Platform checks the messages to make sure each contains the minimum information required to process data to the next level, which is the Processed table. If one or more checks fail, the message is flagged with a status of “filtered” and is not processed further. Filtered Messages are those that do not meet basic HL7 requirements. As non-filtered data progress from Raw to Processed tables, the BioSense Platform applies additional checks to assess validity of the Facility Identifier, Patient Identifier, and Visit Date. These critical data elements form a unique BioSense Identifier. If one or more checks fail, the message is triaged to the Exceptions table and is not processed further.

For more detail on Filter and Exception reason codes and descriptions, refer to the *Data Dictionary* on the [NSSP Technical Resource Center](#).

Data Refresh: Data Flow graphics are updated daily between 06:00:00 UTC and 10:00:00 UTC and are based on data that arrived on the platform by 00:00:00 UTC. Underlying calculation data include the previous 120 days of data (based on arrived_date).

Note: Filtered messages and some Exceptioned messages do not contain facility information. Therefore, the Facility, Parent Organization, Vendor, and Facility Type filters will not subset filtered message counts or percentages in the visualizations. Filtering on one or more facilities, parent organizations, vendors, or facility types will yield no filtered message information in the Data Flow page visualizations.

3.2 Data Flow Banner

The Data Flow banner (Figure 6) contains counts and percentages of processed messages, filtered messages, and exceptioned messages. Processed messages are indicated in green, filtered messages are indicated in orange, and exceptioned messages are indicated in red.

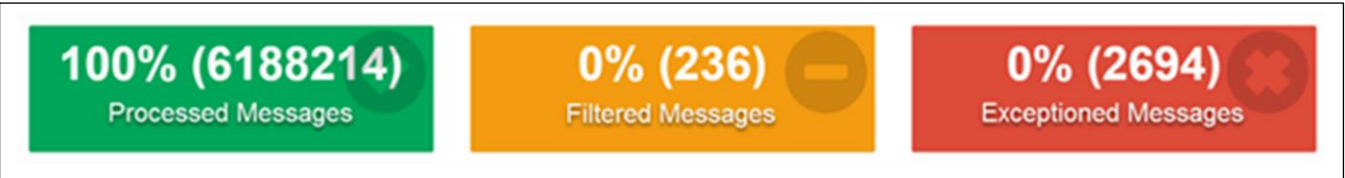


Figure 6. Data Flow Banner

- The percent of processed messages is calculated as the number of messages processed into the Processed table over total messages processed into the Raw table.
- The percent of filtered messages is calculated as the number of filtered messages over total messages processed into the Raw table.
- The percent of exceptioned messages is calculated as the number of exceptioned messages over the total messages processed into the Raw table.

3.3 Data Flow Overview

The Summary of Processed, Filtered, and Exceptioned Messages graph (Figure 7) is a stacked bar chart of these message counts by arrived_date. The x-axis shows the arrived_date and the y-axis shows the message count.

You can click on the **Processed**, **Filtered**, and **Exceptions** buttons above the graph to isolate selected message type(s). The horizontal grey line indicates the mean count of the selected message types (Processed, Filtered, or Exceptions). The translucent bar indicates one standard deviation above and below the mean.

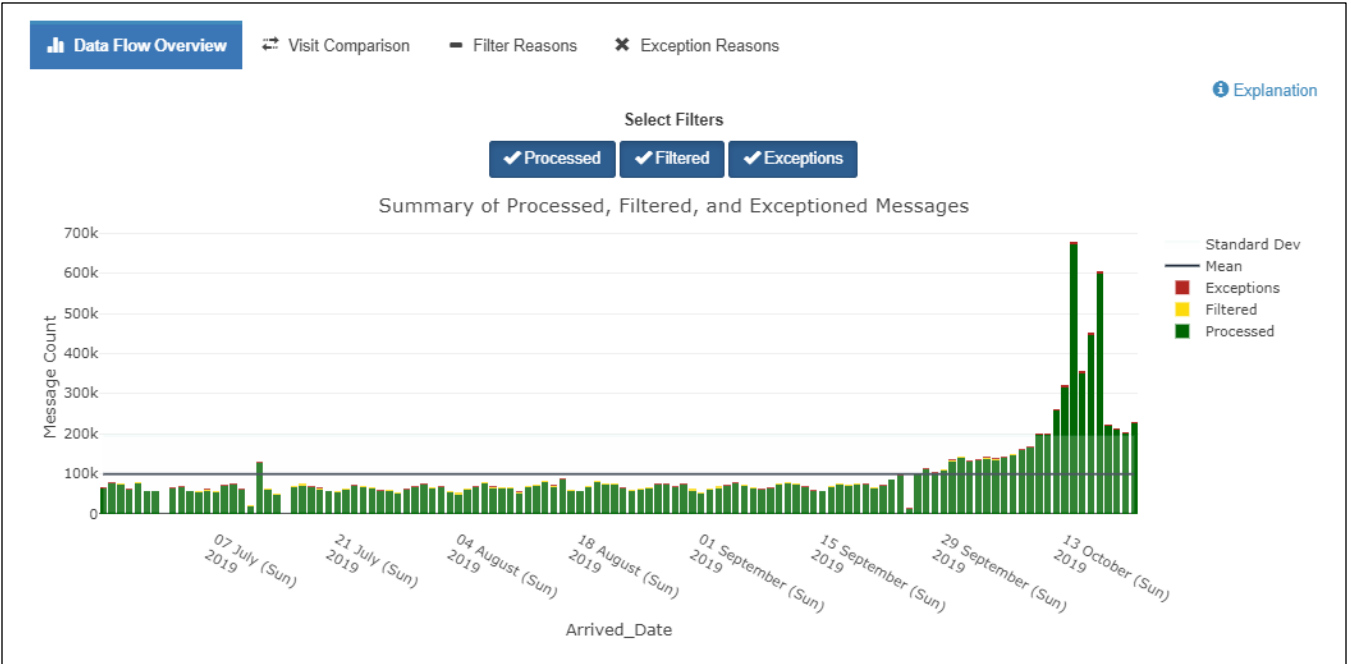


Figure 7. Data Flow Overview—Summary of Processed, Filtered, and Exceptioned Messages

3.4 Visit Count Comparison

The Visit Comparison graph (Figure 8) allows users to identify ESSENCE backlogs or other processing issues. The graph displays your site's Processed message counts, by arrived_date, and your site's Processed and ESSENCE visit counts, by c_visit_date. The x-axis shows the arrived_date, and the y-axis shows the message count. The processed message counts (bars) are calculated by summing the total messages (by arrived_date) and summing the Processed and ESSENCE visit counts (by c_visit_date). You can click on **Processed Messages**, **Processed Visits**, and **ESSENCE Visits** in the legend to isolate the selected messages or visit types.

Note: The ESSENCE visit count will likely be unequal to the Processed visit count for yesterday's c_visit_date due to the timing of replication. This does not indicate a processing issue unless the counts remain unequal over time.

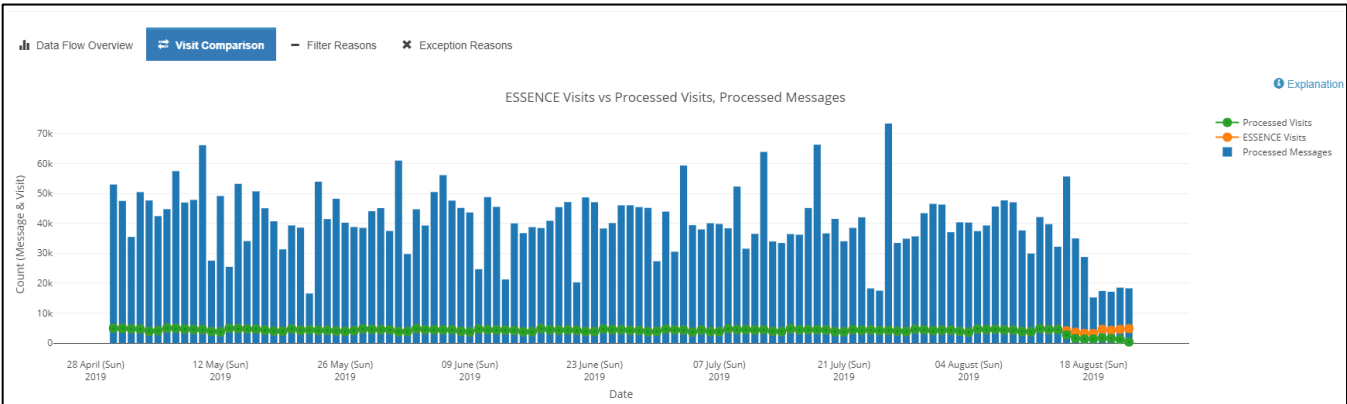


Figure 8. Visit Comparison Graph

3.5 Filter Reasons

The Filter Reasons Over Time graph (Figure 9) displays your site's filtered message counts by arrived_date and is broken down by filter reason. The x-axis shows the arrived_date, and the y-axis shows the message count. Filter reasons will only appear in the legend if your site has messages filtered for that reason. You can click on the different filter reasons in the legend to isolate the selected filtered messages.

The horizontal grey line indicates the mean count of filtered messages.

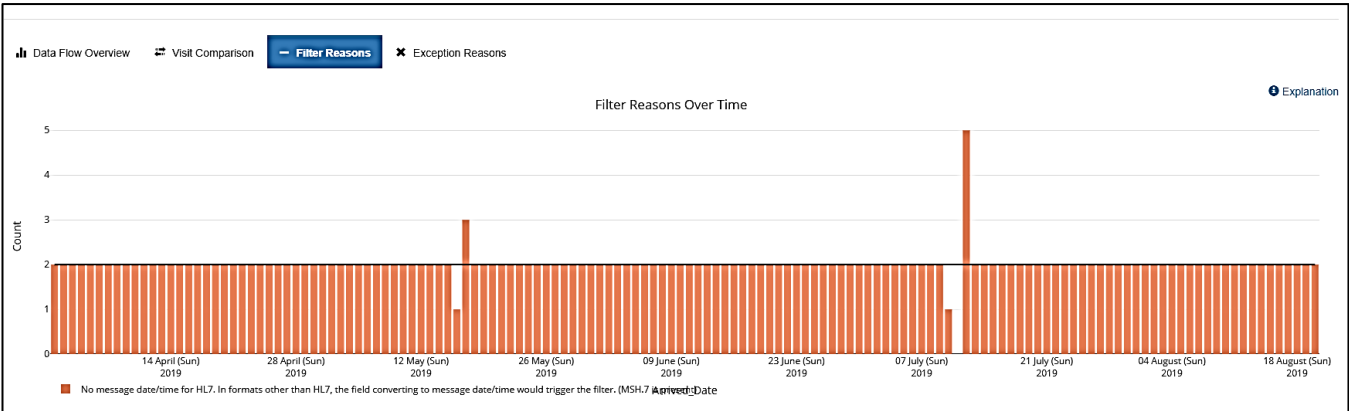


Figure 9. Filter Reasons Over Time Graph

The Filtered Messages table displays your site’s 5,000 most recent (by arrived date) filtered message details including:

- Message ID
- Filter Reason Description
- Message
- Filter Reason Code
- Arrived Date
- Message Type
- Channel Name
- Message Status
- File Name

You can select which columns to include by clicking **Column Visibility** in the upper-left corner of the table (Figure 10). You can also filter the table by searching for specific text, sorting the columns, or by downloading the table in Excel or CSV (Comma Separated Values) format.



Figure 10. Filtered Messages Table

3.6 Exception Reasons

The Exception Reasons Over Time graph (Figure 11) displays your site’s Excepted message counts by arrived_date and exception reason. The x-axis shows the arrived_date, and the y-axis shows the message count. (The message count bars are segmented by Exception Reason.) Note: Exception Reasons will only appear in the legend if your site has messages with an exception for that reason. You can click on the different exception reasons in the legend to exclude them from the graph.

The horizontal grey line indicates the mean count of all excepted messages.

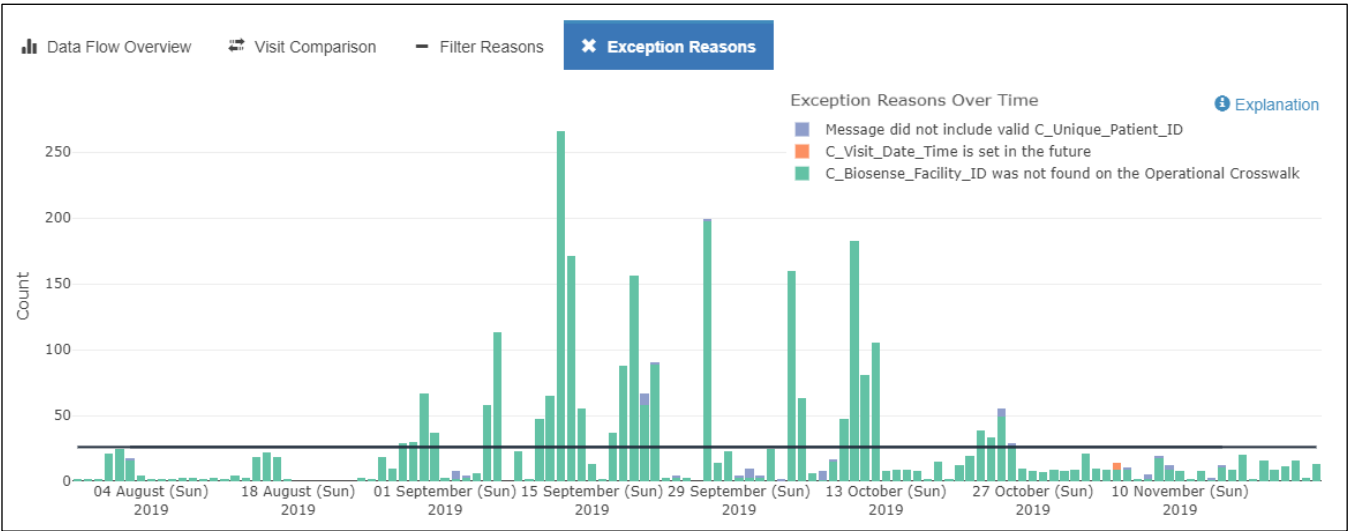


Figure 11. Exception Reasons Over Time Graph

The Exceptioned Message table displays your site’s 5,000 most recent (by arrived date) and Exceptioned message details including:

- Message ID
- Exception Reason Description
- Exception Reason Code
- Arrived Date
- Feed Name
- Sending Application
- Message Type
- Trigger Event
- Channel Name
- Exception Date
- Recorded Date Time
- Facility Name
- Receiving Facility
- Facility ID

You can select which columns to include by clicking **Column Visibility** in the upper-left corner of the table (Figure 12). You can also filter the table by searching for specific text, sorting the columns, or by downloading the table in Excel or as a CSV file.

Exceptioned Messages for 2019-07-27 to 2019-11-24

Column visibilityDownload

Search:

Message_ID	Exception_Description	Facility_Name	Exception_Date	Feed_Name	Arrived_Date_Time	Sending_Application	Message_Type
56575	C_Biosense_Facility_ID was not found on the Operational Crosswalk		2019-11-23T15:12:23Z		2019-11-23T15:11:00Z		
All	All	All	All	All	All	All	All

Showing 1 to 3,099 of 3,099 entries

For processing reasons, only the most recent 5000 messages will be displayed in this table. Change your date range to view older messages.

Figure 12. Exceptioned Messages Table

4. Timeliness

The Timeliness Page provides an overview of your site’s data timeliness. Timeliness is calculated as the difference in time between the first message received (adjusted for time zone) for a given visit and the c_visit_date of that message.

Data Refresh: Timeliness graphics are updated daily between 06:00:00 UTC and 10:00:00 UTC and are based on data that arrived on the platform by 00:00:00 UTC. Underlying calculation data include the previous 120 days of data (based on c_visit_date).

The Timeliness page includes the following visualizations:

4.1 Timeliness Banner

The Timeliness Banner (Figure 13) contains the percent of visits received within 24 hours, between 24 and 48 hours, and after 48 hours. The percent of visits is calculated as the number of visits where the first message was received within the specified timeframe—in less than 24 hours, between 24 and 48 hours, or after 48 hours—over the total visits.

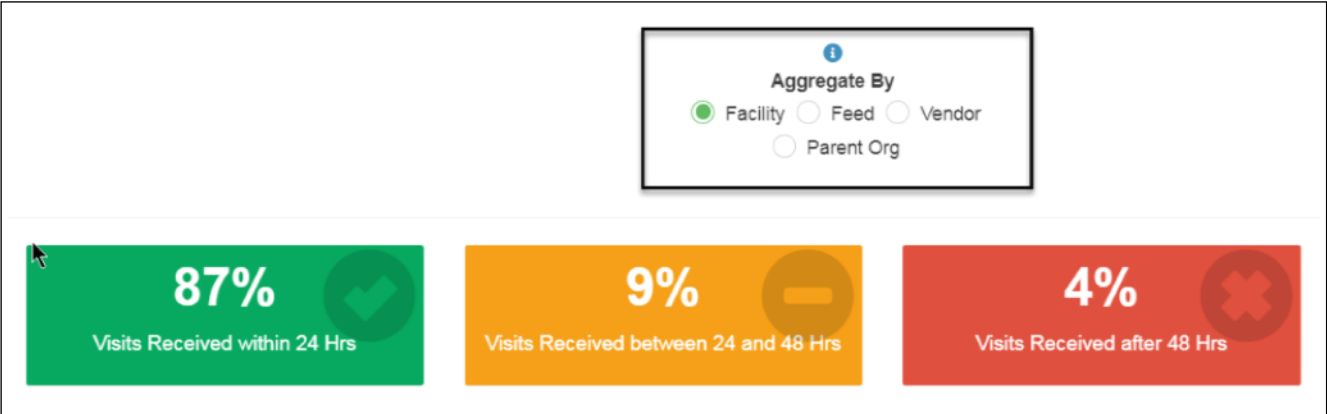


Figure 13. Timeliness Banner

4.2 Overall Timeliness

The Overall Timeliness graph (Figure 14) displays the percent of your site's visits received within the selected timeframe—less than 24 hours, between 24 and 48 hours, or after 48 hours—by facility, feed, vendor, or parent organization. You can click on **Less than 24**, **Between 24 and 48**, and **Greater than 48** in the graph legend to isolate the selected timeframe(s). You can also select to view the graph by facility, feed, vendor, or parent organization by selecting the corresponding **Aggregate By** radio button above the Timeliness Banner.

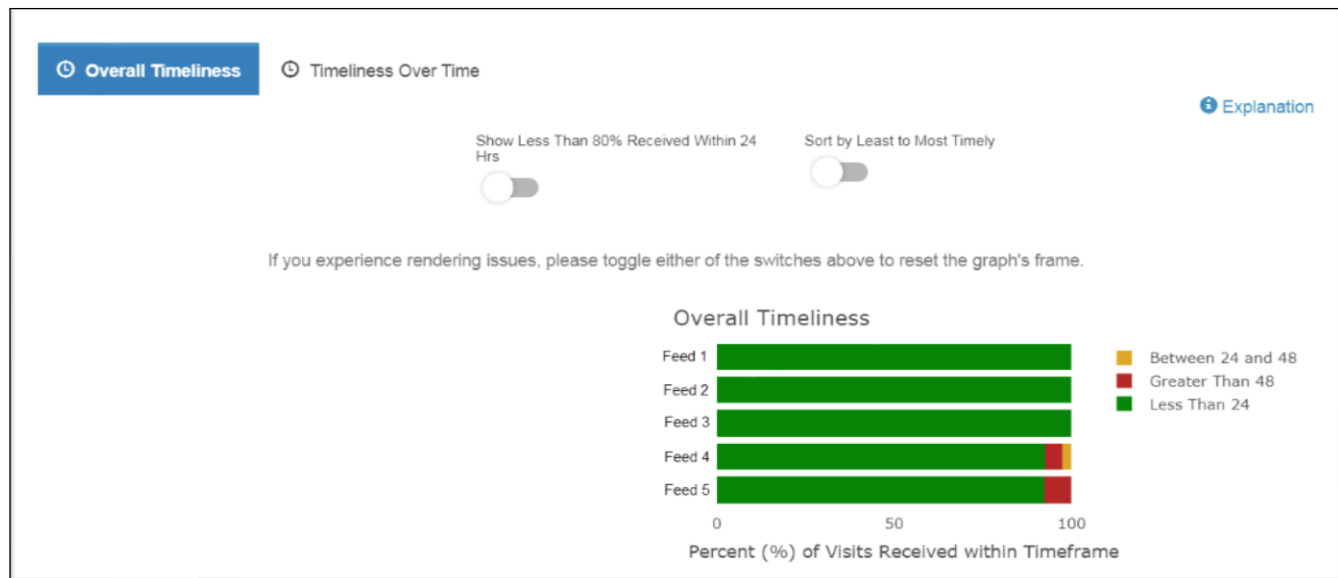


Figure 14. Overall Timeliness Graph

The x-axis shows the percent of the first message for a visit received, and the y-axis shows the selected entity (Facility, Feed, Vendor, or Parent Organization as selected by the **Aggregate By** radio buttons). To ensure timely and useful syndromic data, more than 80% of the first messages for a visit should be received within 24 hours. You can view entities that have less than 80% of the first messages for a visit received within 24 hours by selecting the **Show Less Than 80% Received Within 24 Hrs** toggle switch on the upper-left side of the graph.

4.3 Timeliness Over Time

The graph titled Time from Visit to First Message Received: Overall Average (Figure 15) displays the average hours delayed, or average time between visit and first message received, based on c_visit_date. The x-axis shows week ending in the specified c_visit_date, and the y-axis shows the average time between visit and first message received. The solid line represents the average across all selected entities (feeds, vendors, parent organizations, or facilities). You can view specific entities by filtering on Feed, Vendor, Parent Organization, or Facility.

Note: The Time from Visit to First Message Received: Overall Average graph is a *proportional average*. For example, if this graph is generated for all facilities but one facility contributes to 95% of the visits across all selected facilities, the overall plot will look most like the Timeliness Over Time plot of the individual facility contributing to 95% of the visits.

The dotted horizontal line indicates the NSSP timeliness threshold of 24 hours. To ensure timely and useful syndromic data, more than 80% of the first messages for visits should be received within 24 hours.

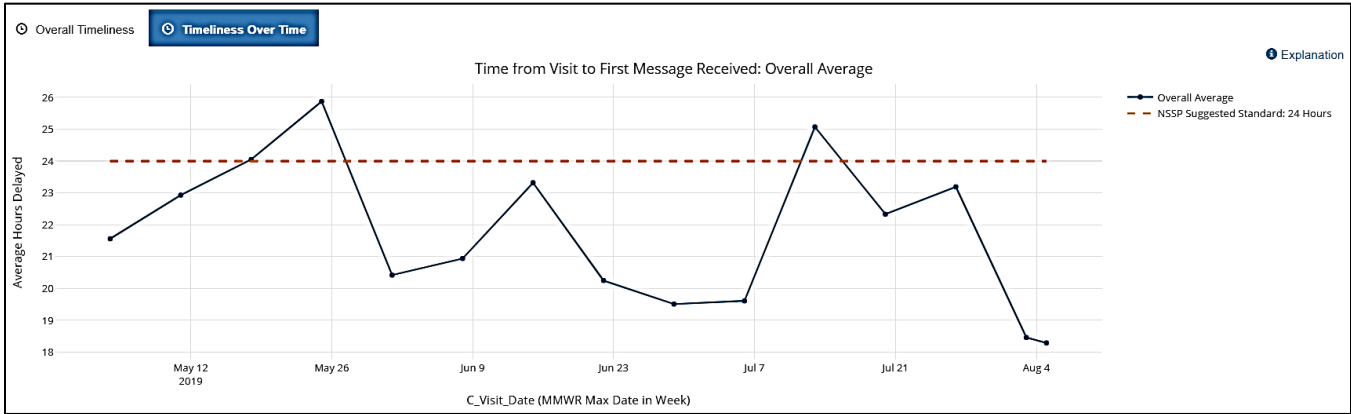


Figure 15. Time from Visit to First Message Received: Overall Average

The Time from Visit to First Message Received graph (Figure 16) displays the average hours between Visit and First Message Received by week (i.e., the average by week of times between visit and first message was received according to the specified c_visit_date for each week).

The x-axis shows the week ending for the specified set of c_visit_date, and the y-axis shows the average time between visit and first message received for that week.

Each line represents a selected entity (feed, vendor, parent organization, or facility). You can select to view the graph by facility, feed, vendor, or parent organization by selecting the corresponding **Aggregate By** radio button in the filter bar above. You can also select to view specific entities by filtering on Feed, Parent Organization, Vendor, Individual Elements, Facility Type, Facility, Date Range, or Group Elements in the filter bar above.

The dotted horizontal line indicates the NSSP timeliness threshold of 24 hours. To ensure timely and useful syndromic data, more than 80% of the first messages for visits should be received within 24 hours.

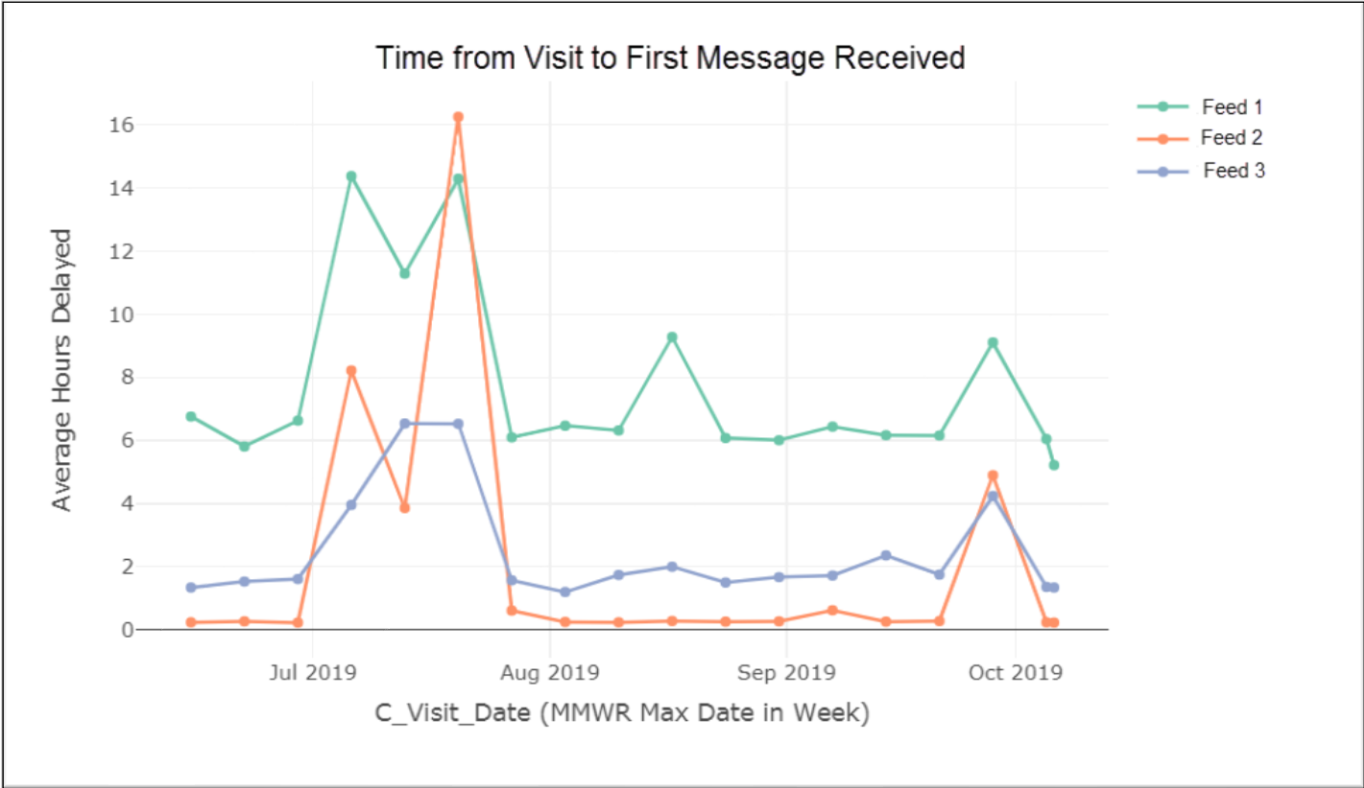


Figure 16. Time from Visit to First Message Received Graph

5. Completeness

The Completeness page provides an overview of your site's data completeness for specified data elements. For an element to be considered complete, it must contain at least one non-null message value per visit. The following data elements (Table 1) are assessed for completeness.

Table 1. List of Elements Assessed for Completeness		
Priority 1	Priority 2	
<ul style="list-style-type: none"> Facility_Type_Code Sending_Facility_ID Treating_Facility_ID Visit_ID C_FacType_Patient_Class Patient_Class_Code Admit_Date_Time Admit_Reason_Description C_Chief_Complaint Chief_Complaint_Text C_Patient_Age C_Patient_Age_Years Diagnosis_Code Diagnosis_Description Patient_Zip Processing_ID Trigger_Event 	<ul style="list-style-type: none"> Sending_Facility_ID_Source Discharge_Disposition Discharge_Date_Time First_Patient_ID Medical_Record_Number Recorded_Date_Time Admit_Reason_Code Chief_Complaint_Code Diagnosis_Type Administrative_Sex Age_Reported Age_Units_Reported C_Patient_Age_Units C_Patient_County Ethnicity_Code Ethnicity_Description Patient_City 	<ul style="list-style-type: none"> Patient_Country Patient_State Race_Code Race_Description Version_ID

You can select to view visualizations for specific groupings of data elements by using the Group Elements filter or for combinations of data elements by using the Individual Elements filter. By default, Priority 1 data elements will be displayed. The data element groups are of two types (see the Individual Elements dropdown list for which elements are included in each group):

- *Priority groups* (elements grouped by importance). These groups are Priority 1 and Priority 2.
- *Use groups* (elements grouped into related categories), for example Facility, Visit, Chief Complaint (CC) and Diagnosis, Patient Demographics, and Data Processing data elements. Individual elements may also be chosen.

Data Refresh: Completeness graphics are updated daily between 06:00:00 UTC and 10:00:00 UTC and are based on data that arrived on the platform by 00:00:00 UTC. Underlying calculation data include the previous 120 days of data (based on c_visit_date).

The Completeness page visualizations are shown below:

5.1 Completeness Banner

The Completeness banner (Figure 17) contains the percent completeness for emergency department (ED) facilities and for the site overall. Percent completeness is calculated as the average percent completeness of selected data elements. Percent completeness of each data element is calculated as the number of visits with at least one non-null message value for that data element, per visit, over total visits.

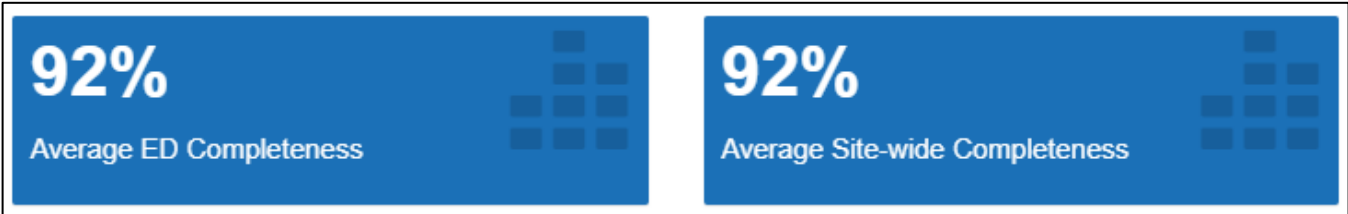


Figure 17. Completeness Banner

5.2 Completeness by Element

The Completeness by Element graph (Figure 18) displays the percent of visits with at least one non-null message value for selected data element(s). The x-axis shows the percent complete, and the y-axis shows the selected data element(s). The orange (top) bar displays completeness of selected data elements for your site (Site-Wide), and the blue (bottom) bar displays completeness of selected data elements for emergency care facilities (Emergency Only) within your site.

By default, the percent completeness of Priority 1 data elements will be displayed. You can select to view the completeness of different data elements by using the other Group Elements or Individual Element filters.

The grey vertical line indicates the 90% NSSP completeness threshold. To ensure complete and useful syndromic data, ideally, data elements have greater than 90% completeness (i.e., more than 90% of visits have at least one message with one non-NULL value for that data element). You can view data elements that have less than 90% completeness by selecting the toggle switch on the upper-left side of the graph.

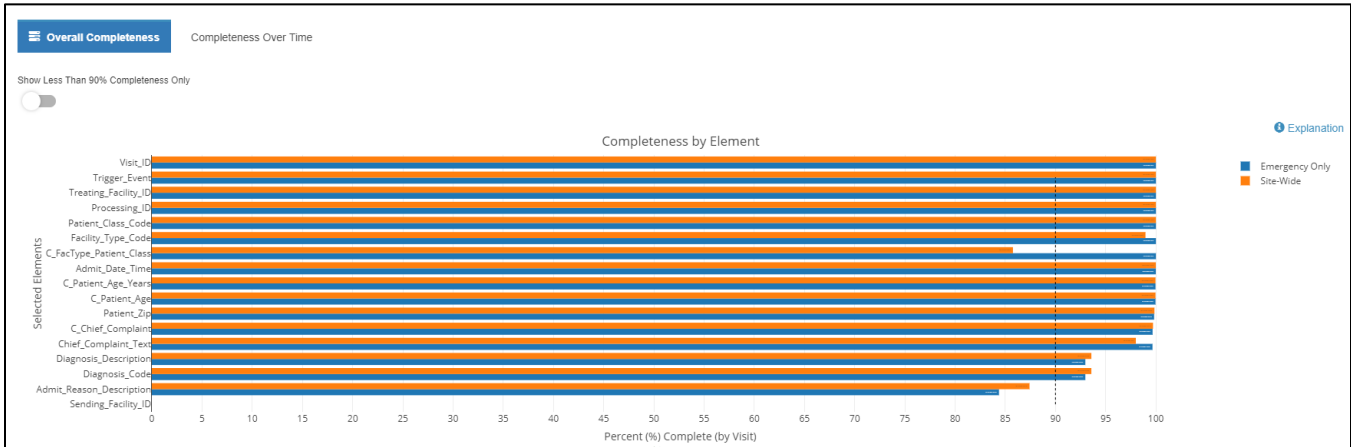


Figure 18. Completeness by Element Graph

5.3 Overall Completeness Over Time

The Overall Completeness Over Time graph (Figure 19) displays the average percent of visits with at least one non-NULL message for selected data elements. The x-axis shows the week ending in the specified c_visit_date, and the y-axis shows the average percent completeness.

By default, the average percent completeness of Priority 1 data elements will be displayed. You can select to view the average completeness of different data elements by using the data element filters.

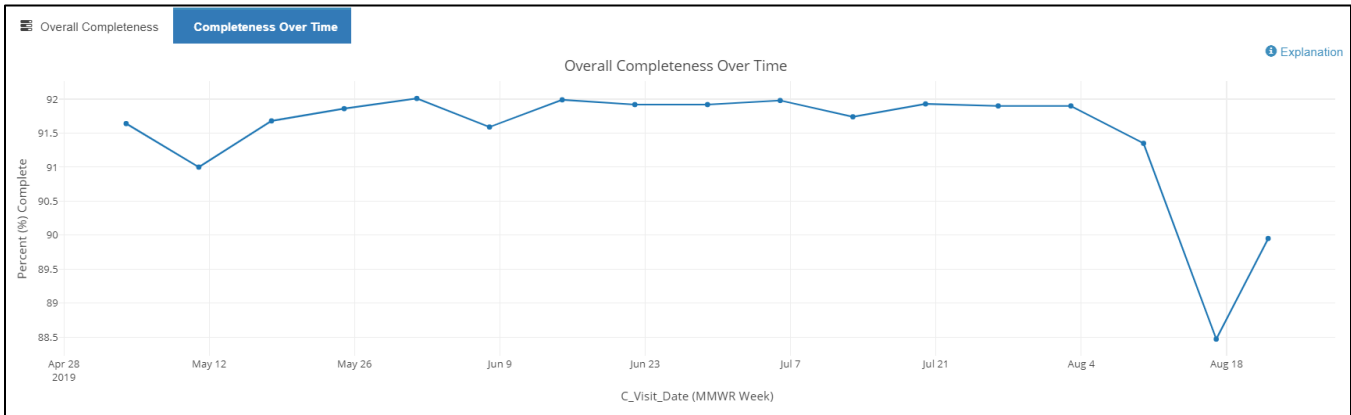


Figure 19. Overall Completeness Over Time Graph

5.4 Completeness by Element Over Time

The Completeness by Element Over Time graph (Figure 20) displays the percent of visits with at least one non-NULL message for selected data elements. The x-axis shows the week ending in the specified c_visit_date, and the y-axis shows the average percent completeness.

Each line represents a selected data element. By default, the percent completeness of Priority 1 data elements will be displayed. You can select to view the completeness of different data elements by using the data element filters.

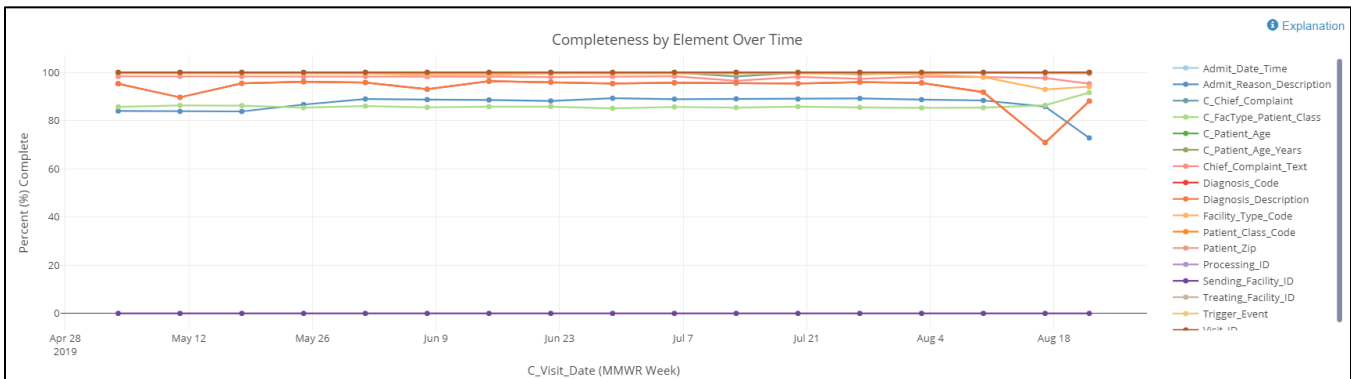


Figure 20. Completeness by Element Over Time Graph

6. Validity

Table 2 shows the specific data elements that are assessed for validity. A data element that (1) follows its rule listed under *Validity Definitions* AND (2) adheres to guidance referenced under *Location of Valid Values* is accepted as valid.

Table 2. List of Element Validity Definitions and Location of Valid Values		
Data Element	Validity Definitions	Location of Valid Values <i>Always refer to the current PHIN VADS* value list.</i>
Facility_Type_Code	Considered valid if value exists in PHIN VADS value list	PHVS FacilityVisitType SyndromicSurveillance
C_FacType_Patient_Class	Considered valid if value exists in PHIN VADS value list	Valid values include E, I, and O and are inferred using Facility_Type_Code
Patient_Class_Code	Considered valid if value exists in PHIN VADS value list	PHVS PatientClass SyndromicSurveillance
C_Chief_Complaint	Considered valid if value in its entirety is not found in the Non-informative Chief Complaint (NICC) terms list	Data Dictionary in NSSP Technical Resource Center
Diagnosis_Code	Considered valid if the value in its entirety is not found in the Non-Informative Chief Complaint (NICC) terms list	Data Dictionary in NSSP Technical Resource Center
C_Patient_Age	Considered valid if value is numeric and less than 120	
C_Patient_Age_Years	Considered valid if value is numeric and less than 120	Inferred using C_Patient_Age and C_Patient_Age_Source
Patient_Zip	Considered valid if formatted as ##### or #####-####	
Trigger_Event	Considered valid if value exists in PHIN VADS value list	PHVS EventType SyndromicSurveillance
Discharge_Disposition	Considered valid if value exists in PHIN VADS value list	PHVS DischargeDisposition HL7 2x
Diagnosis_Type	Considered valid if all values received within that message exist in PHIN VADS value list	PHVS DiagnosisType HL7 2x
Administrative_Sex	Considered valid if value exists in PHIN VADS value list	PHVS Gender SyndromicSurveillance
Age_Reported	Considered valid if value is numeric and less than 120	
Age_Units_Reported	Considered valid if value exists in PHIN VADS value list	PHVS AgeUnit SyndromicSurveillance
C_Patient_Age_Units	Considered valid if value exists in PHIN VADS value list	
Ethnicity_Code	Considered valid if all values received within that message exist in PHIN VADS value list	PHVS EthnicityGroup CDC
Race_Code	Considered valid if all values received within that message exist in PHIN VADS value list	PHVS RaceCategory CDC

* PHIN VADS: Public Health Information Network Vocabulary Access and Distribution System

You may view visualizations for specified groupings of data elements by selecting the Group Elements filter or for individual data elements by selecting the Individual Elements filter. By default, Priority 1 data elements will be displayed. The data element groupings include:

- *Priority groups* (elements grouped into groups of importance) including Priority 1 and Priority 2.
- *Use groups* (elements grouped into related categories) including Facility, Visit, Chief Complaint (CC) and Diagnosis, Patient Demographics, and Data Processing data elements. Individual elements may also be chosen.

Data Refresh: The Validity graphics are updated daily at 04:00:00 UTC, and the data in the underlying calculations are refreshed daily at 00:00:00 UTC. Underlying calculation data include the previous 120 days of data (based on c_visit_date).

The Validity page includes the following visualizations:

6.1 *Validity Banner*

The Validity banner (Figure 21) contains the percent validity for ED facilities and for the site overall. Percent validity is calculated as the average percent validity of selected data elements. Percent validity of each data element is calculated as the number of messages with a valid value for that data element, over total messages with **any** non-null value for that element.

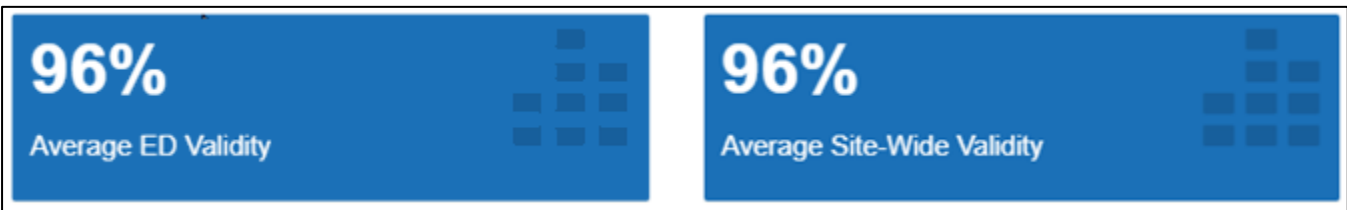


Figure 21. Validity Banner

6.2 *Overall Validity*

The Validity by Element graph (Figure 22) displays the percent of messages with a valid value for selected data elements. Percent validity of a data element is calculated as the number of messages with a valid value for that data element over messages with any non-null value for that element. The x-axis shows the percent of valid messages, and the y-axis shows selected data element(s). The orange (top) bar displays validity of selected data elements for your site (Site-Wide), and the blue (bottom) bar displays validity of selected data elements for ED facilities within your site.

By default, the percent validity of Priority 1 data elements will be displayed. You can select to view the validity of different data elements by using the data element filters.

The grey vertical line indicates the NSSP validity threshold of 80%. To ensure valid and useful syndromic data, ideally, data elements have greater than 80% validity. Or, put differently, more than 80% of messages have a valid value for that data element. You can view data elements that have less than 80% validity by selecting the toggle switch on the upper-left side of the graph.

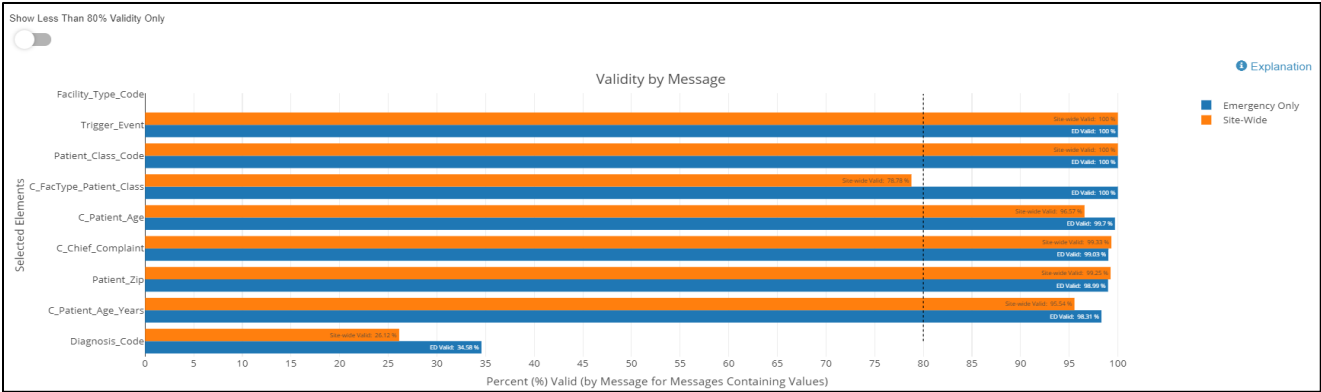


Figure 22. Validity by Message

6.3 Overall Invalidity

The Invalidity by Message graph (Figure 23) displays the percent of messages with an invalid non-null value for the selected data elements. The percent invalid for a data element is calculated as the number of messages with an invalid value for that data element over the total number of messages for that data element with non-null values. The x-axis shows the percent invalid and the y-axis shows the selected data element(s).

By default, the percent invalidity of Priority 1 data elements will be displayed. You can select to view the invalidity of different data elements by using the data element filters.

The grey vertical line indicates the NSSP invalidity threshold of 20%. For valid and useful syndromic data, ideally, data elements have less than 20% invalidity (i.e., less than 20% of messages have an invalid value for that data element). You can view data elements that have greater than 20% invalidity by selecting the toggle switch on the upper-left side of the graph.

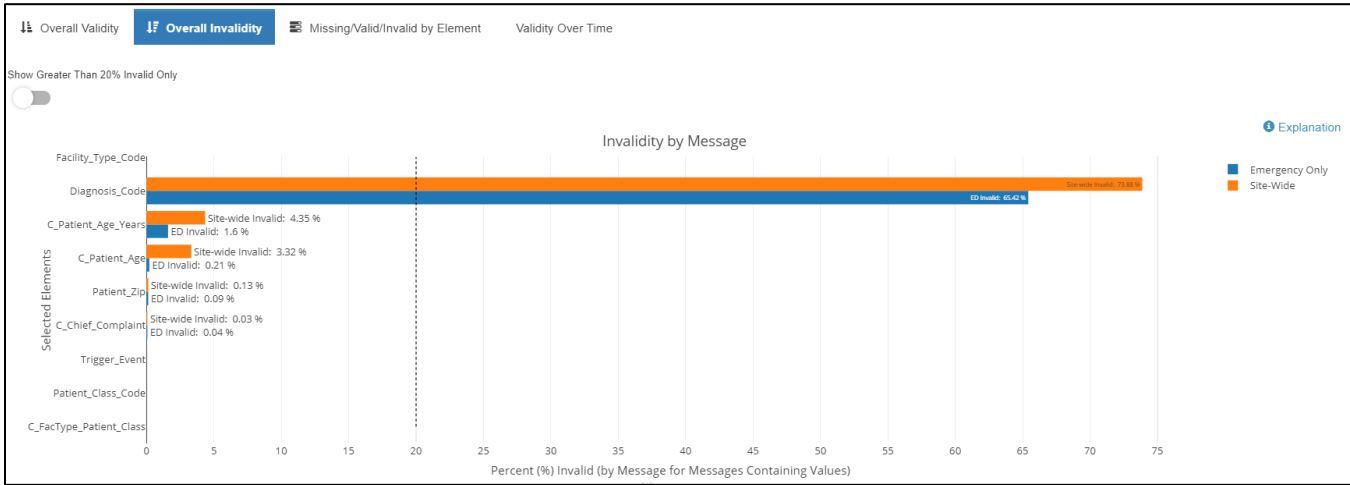


Figure 23. Invalidity by Message

6.4 Missing, Valid, Invalid by Element

The Missing, Valid, and Invalid by Element graph (Figure 24) displays the percent of messages with missing, valid, and invalid values for the selected data elements.

- Percent validity of a data element is calculated as the number of messages with a valid value for that data element over total messages—including *NULL* values.
- Percent invalidity of a data element is calculated as the number of messages with no valid value for that data element over total messages—including *NULL* values.
- Percent missing for each data element is calculated as the number of messages with a *NULL* value for that data element over total messages. The x-axis shows the percent, and the y-axis shows the selected data element(s).

By default, Priority 1 data elements will be displayed. You can select to view different data elements by using the data element filters. Please review the Calculation Information for details on how validity is assessed for each data element.

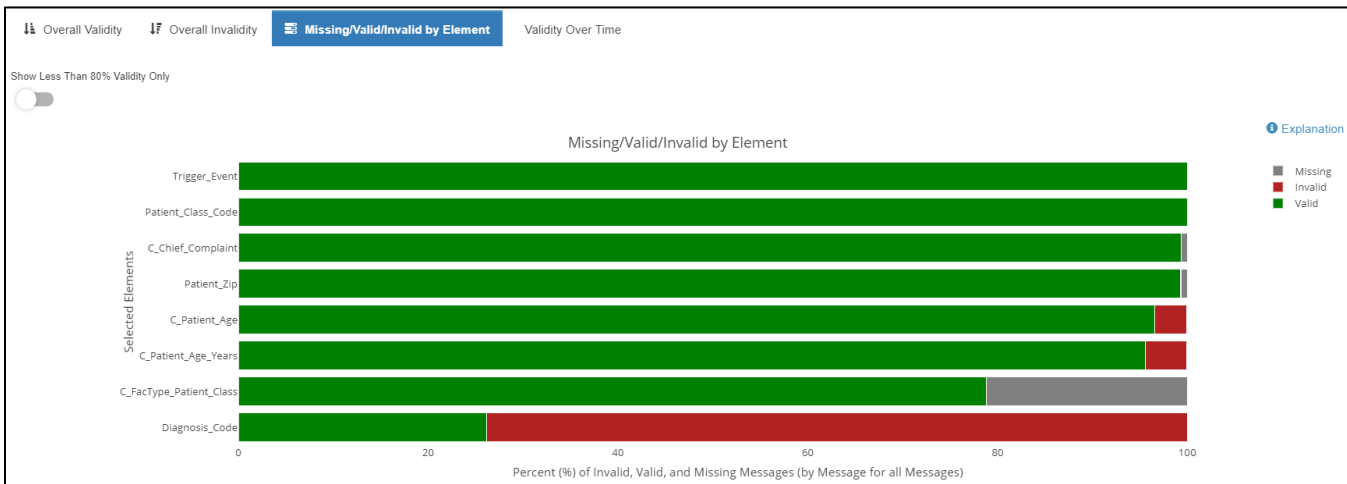


Figure 24. Missing/Valid/Invalid by Element

6.5 Overall Validity Over Time

The Overall Validity Over Time graph (Figure 25) displays the average percent of messages with a valid value for selected data elements. The x-axis shows the week ending in the specified c_visit_date, and the y-axis shows the average percent validity for that week for all selected data elements.

By default, the average percent validity of Priority 1 data elements will be displayed. You can select to view the average validity of different data elements by using the data element filters.

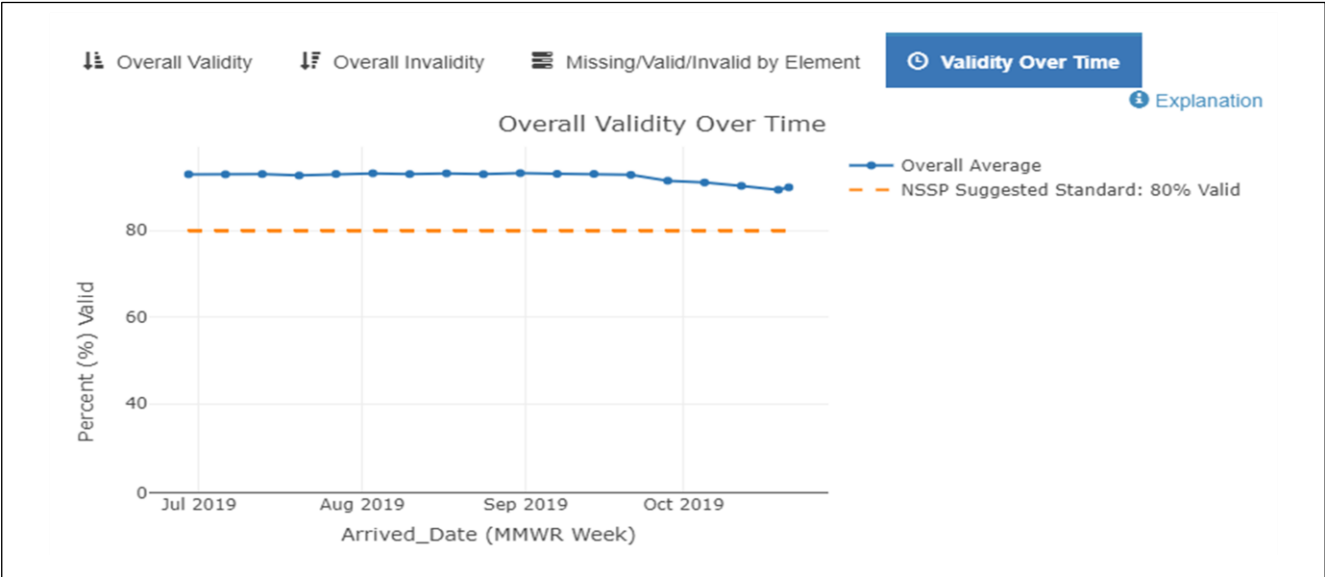


Figure 25. Overall Validity Over Time Graph

The Validity by Element Over Time graph (Figure 26) displays the percent of messages with a valid value for selected data elements. The x-axis shows the week ending in the specified c_visit_date, and the y-axis shows the average percent validity for the week for each selected data element.

Each line represents a selected data element. By default, the percent validity of Priority 1 data elements will be displayed. You can select to view the validity of different data elements by using the data element filters.

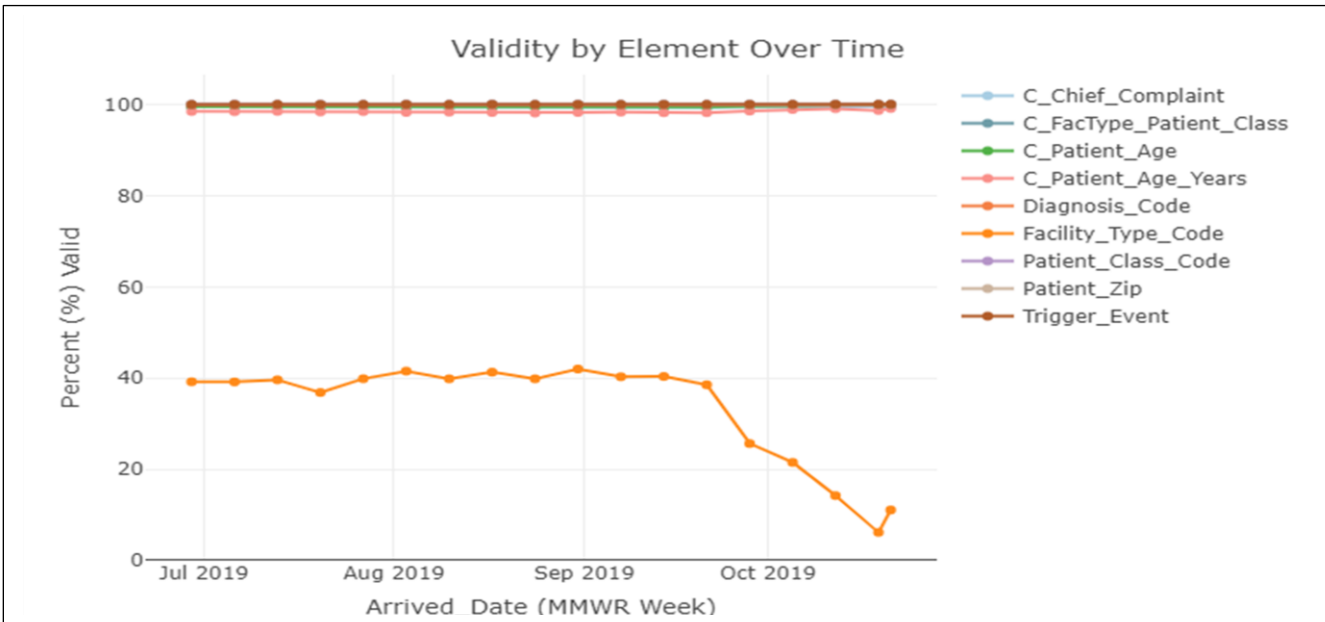


Figure 26. Validity by Element Over Time Graph

7. Plot Builder Overview

The Plot Builder page allows users to interactively create exportable chart visualizations by using the underlying dashboard data. Please refer to the **Information** buttons in the Data Flow, Timeliness, Completeness, and Validity pages to learn about how the terms (e.g., processed, filtered, exceptions, timeliness, completeness, and validity) are defined.

Here’s how to use the Plot Builder page:

1. Figure 27: Select your desired data quality metric: Operations Summary, Timeliness Summary, Completeness Summary, or Validity Summary.

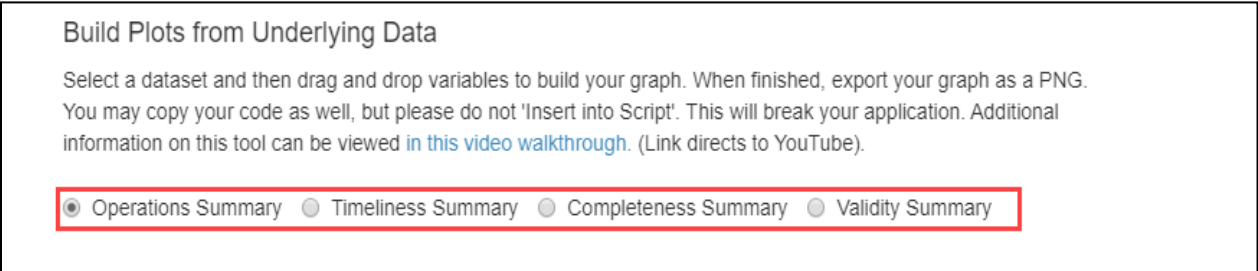


Figure 27. Data Quality Metric Options

2. Drag and drop (Figure 28) the available variables (based on the selected data quality metric) to the chart element fields (X-axis, Y-axis, Fill, Color, Size, Group, and Facet). Note: Orange variables are categorical, dark blue variables are numerical, and light blue variables are dates.

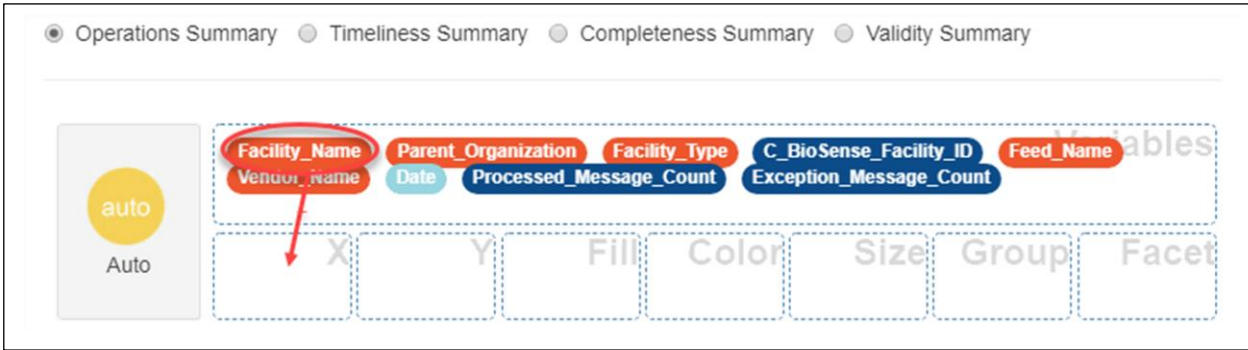


Figure 28. Chart Type Options

3. Add labels, plot options, and data filters as needed to annotate your plot (Figure 29).



Figure 29. Add Labels, Plot Options, and Data Filters

4. Export chart in .PNG or .PPTX format (Figure 30).

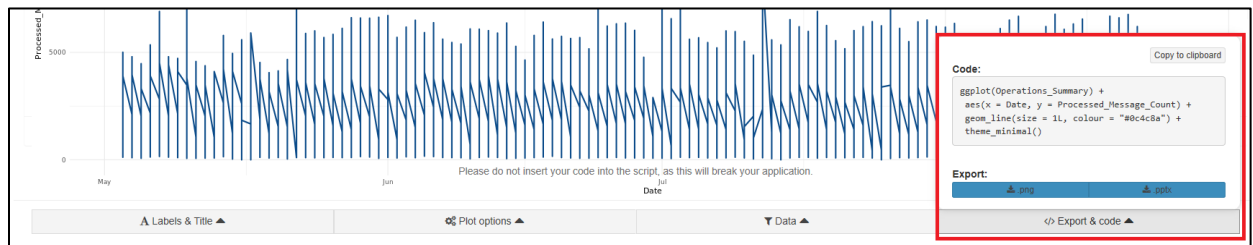


Figure 30. Copy Code or Export Chart in .PNG or .pptx Format

8. Plot Builder Reference Guide

The Plot Builder tab contains a third-party GUI (graphical user-interface) from the [dreamRS esquisse package](#) that allows you to drag and drop variables to construct your own plot. This utility is built from the R programming language, as is the rest of the Data Quality Dashboard.

There can be up to 7 steps in building a plot, but only the first 2 are required:

8.1 Step 1: Select the Dataset (Required)

The first step in building a plot is to select a dataset. Radio buttons here will show you the base reporting tables used to construct the metrics in the rest of the Data Quality Dashboard (Figure 31).

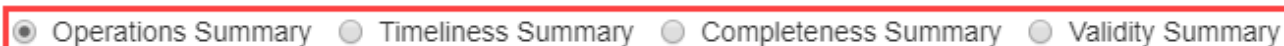


Figure 31. Base Reporting Table

There are multiple reporting tables used to construct the metrics on the Data Flow Tab and currently one, showing message count details, is offered to users: Operations Summary. An additional table, showing visit count details, is logged as an enhancement and will be made available in a future update.

Note for Operational Access (Super) users: if you have not yet selected a site from the Site Dropdown, a sample dataset will be shown until you select a site.

8.2 Step 2: Drag and Drop the Variables (Required)

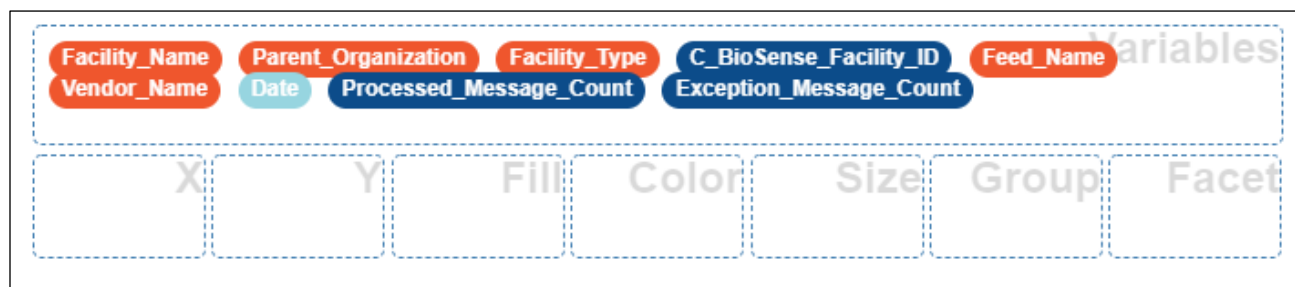


Figure 32. Column Names Populate Variables Boxes

After selecting a dataset from the radio buttons, the column names from that dataset will be populated in the Variables box, shown above in Figure 32. They are color coded into three groups: categorical variables (orange), numeric variables (dark blue), and date variables (light blue). You can drag them to any of the boxes below (X, Y, Fill, Color, Size, Group, Facet) to populate the graph. They can be dragged back to the variables box if you want to remove them. This may take a couple attempts if the variables box is crowded with options, but it is possible.

The **X** box holds the variable that will be on the X-axis (horizontal line). It is a good idea to put the Date variable here for time series. A helpful tip is to put Facility Name on the **Fill** or **Color** boxes because the axes labels will get crowded. This will populate the Facility Name in the legend instead of the axis. The **Y** box holds the variable that will be on the Y-axis (vertical line). It is a good idea to put a numeric metric here, such as Average Timeliness. The **Size** box will change the size of the marker (e.g., a dot) based on the variable you drag there. This works best with numerical variables. The **Group** box will add a legend showing the different groups within the variable. This works best with categorical variables. Finally, the **Facet** box will create multiple graphs where the common factor is the variable you drop in the Facet box.

8.3 Step 3: Select Graph Type (Optional)

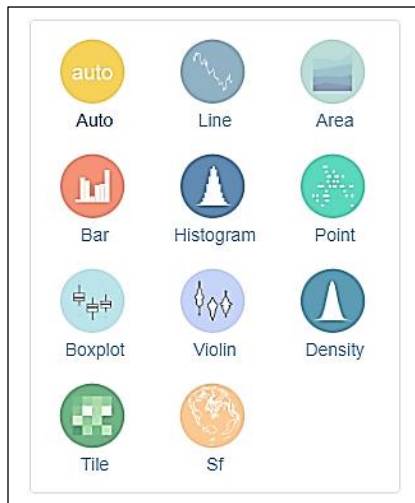


Figure 33. Graph Types

After dragging variables, you'll note that the graph will auto populate, typically with a bar or a scatter plot. If you've dragged a single variable, you will most likely see a histogram. You can alter the type of graph by clicking the circular **Auto** in the top left of the Plot Builder, pictured in Figure 33. Note that some of the options will be deactivated based on the input variables. Because we are dealing with relatively simple datasets, you will not encounter several of the graph types in any scenario. **Sf** for example, requires geographic data (latitude and longitude).

8.4 Step 4: Label the Graph (Optional)

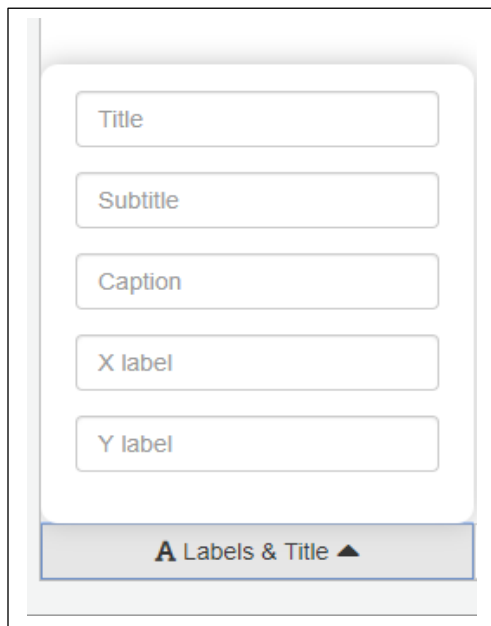


Figure 34. Free-text Fields

At the bottom left of the Plot Builder, you'll find a button to add labels to your graph (Figure 34). These are free-text fields that you can type into to populate your labels. The Title field will appear on the top of your graph, with the Subtitle entry appearing slightly below, in smaller font. The Caption field will be placed on the bottom right of your graph. The X and Y labels will appear near their respective axes.

8.5 Step 5: Change the Graph's Theme (Optional)

The image displays two side-by-side panels representing plot options. The left panel is for single-variable graphs and includes a 'Y-Axis transform' dropdown set to 'identity', a 'Choose a color' dropdown with a blue bar, a 'Theme' dropdown set to 'minimal', and a 'Legend position' section with five buttons: left arrow, up arrow, down arrow, right arrow, and 'X'. The right panel is for multivariable graphs and includes a 'Y-Axis transform' dropdown set to 'identity', a 'Position' section with three radio buttons (stack, dodge, fill) where 'dodge' is selected, a 'Flip coordinates' toggle switch, a 'Choose a palette' dropdown with a multi-colored bar labeled 'ggplot2', a 'Theme' dropdown set to 'minimal', and a 'Legend position' section with five buttons: left arrow, up arrow, down arrow, right arrow, and 'X'. Both panels have a 'Plot options' button at the bottom left.

Figure 35. Examples of Single-variable and Multivariable Plot Options

In **Plot Options**, you'll find a variety of options to change your graph's appearance depending on the dataset and variable(s) selected. Two examples of Plot Options are shown in Figure 35.

For single-variable graphs, you'll see a color option to select the color of the marker (bar/dot). For multivariable graphs, there may be a palette showing a range of color schemes.

The **Theme** option will change nuanced aspects of the graph such as the grid and background color.

Legend position allows you to choose where the legend is placed in relation to the main plot. Clicking **X** here will remove the legend, if desired.

8.6 Step 6: Subset Data (Optional)

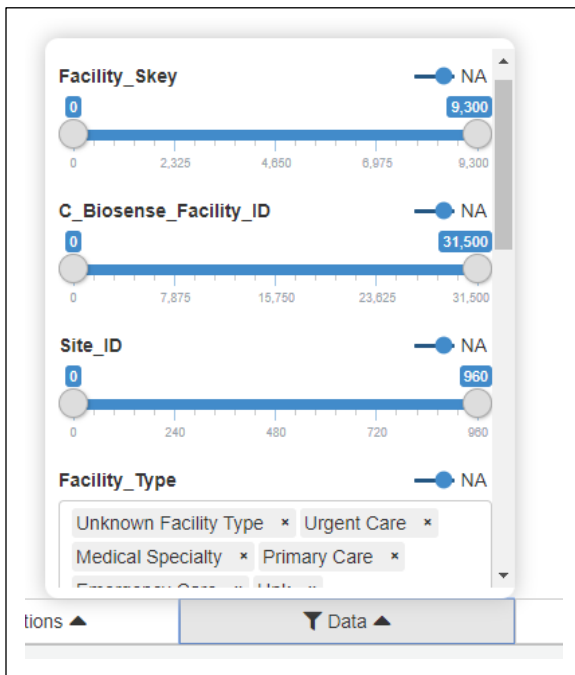


Figure 36. Click Data Button to View Details

If you would like to view a smaller selection of the overall dataset, click the **Data** button (Figure 36).

There are three types of interactions here:

1. For all data types, you can toggle the **NA** switch to remove NAs (e.g., empty data).
2. For numeric data, you will see a range slider, starting from 0 to the highest available value. You can slide this to select only values within that range.
3. For categorical variables such as Facility Type, you will be able to remove them by clicking the **X** on their respective boxes.

8.7 Step 7: Export Graph (Optional)

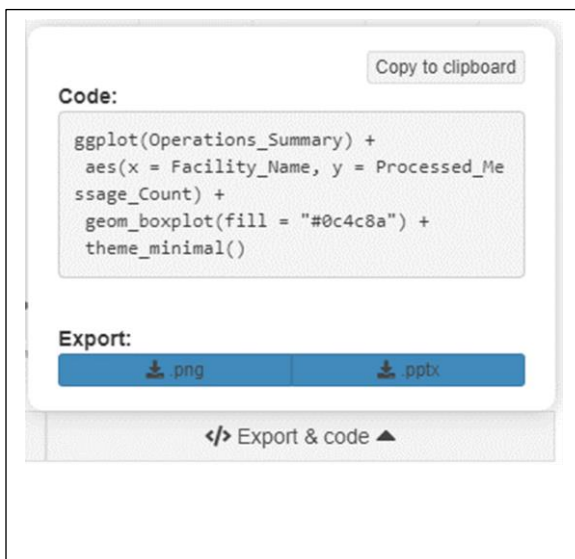


Figure 37. Export Image

Once you have completed your graph, you may export the image (Figure 37). If you have issues with how the plot renders, we recommend simply taking a screenshot of your graph.

To export, click **png** on the **Export** button.

If you press **Copy to clipboard**, you will be copying the **R** language code that can be used to reproduce the graph in the RStudio application.

8.8 Download Dashboard Data

After creating a plot, Figures 38 and 39 show how you can download the underlying dashboard data by clicking the **Download Data** button and selecting the desired data quality metric: Operations Summary, Timeliness Summary, Completeness Summary, or Validity Summary.

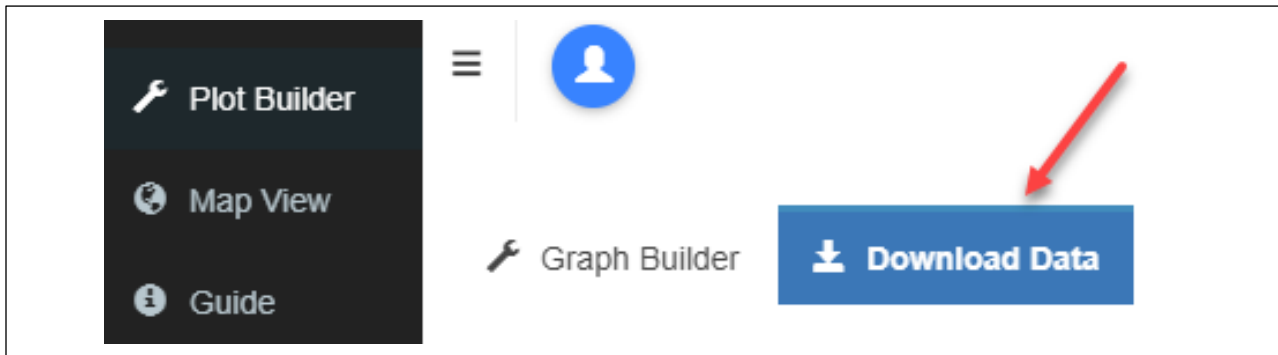


Figure 38. Select Download Data Button

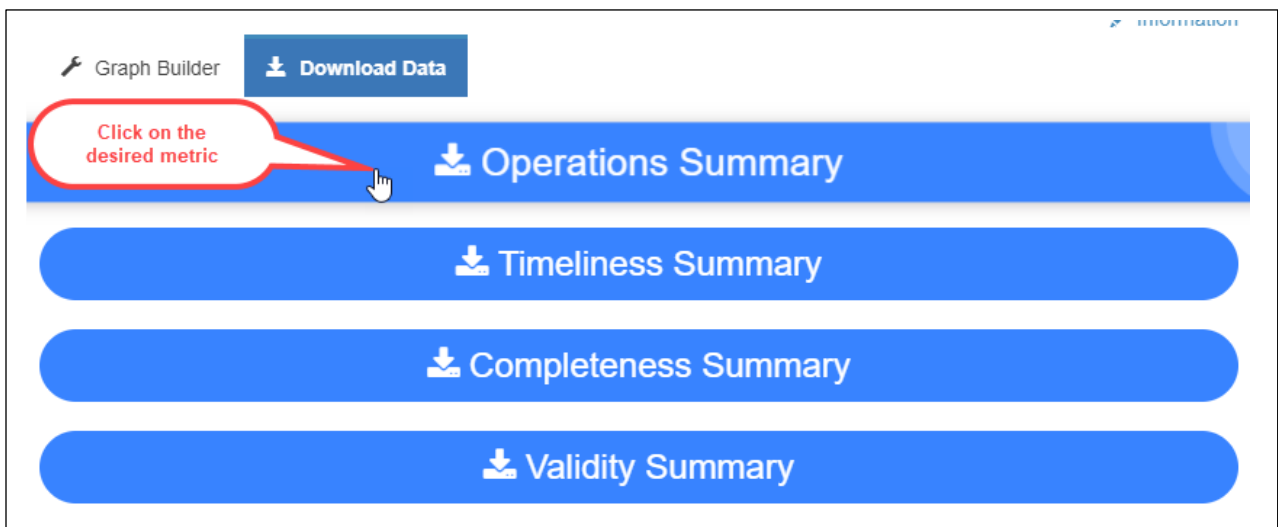


Figure 39. Download Options

After clicking on the desired metric bar, a CSV file will be automatically downloaded to your browser.

9. Map View

The Map View page displays a geographical representation of facilities registered in the Master Facility Table (MFT), their facility status values, and respective visit counts. The facility location in the map on the Map View page corresponds to the latitude and longitude values entered in the MFT.

Due to privacy issues, an example of a map is not displayed here, but highlights of a selected site’s Map View are described below:

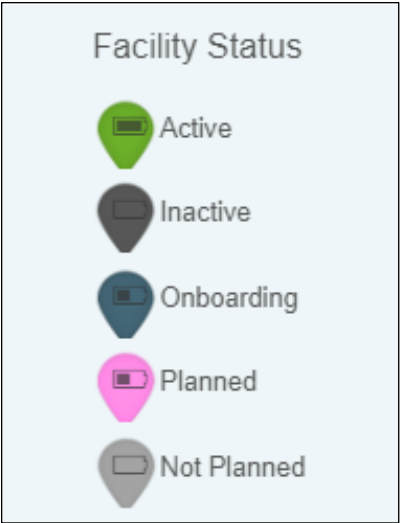



Figure 40. Facility Status Legend

An OpenStreetMap© map, similar to Google Maps, is shown for the geographic area being viewed.

Facility Status is designated by color-coded icons (Figure 40). These icons denote status of individual MFT facilities. In areas where there are multiple facilities, such as in an urban area, and depending on the map’s scale, a special icon containing the number of hidden facilities, for example, , may be displayed.

View individual facilities by zooming in or clicking on the **numbered icon**.

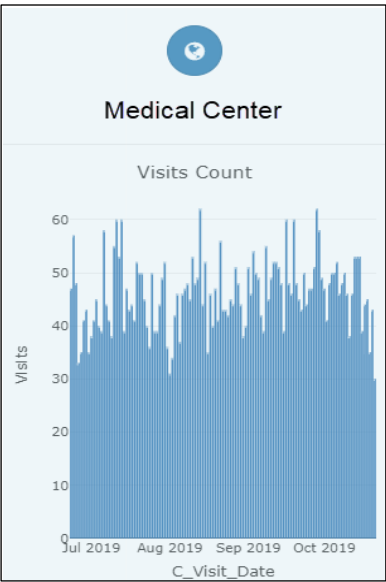


Figure 41. Number of Visits by Date

To access the Facility Information popup and Visit Count graph, first place your cursor on an icon (hover over it) to display the facility name.

If you click on the icon, a popup is displayed with additional facility information, such as Address, Status, Facility Type, and a small Visit Count graph is displayed in the upper-right corner of the page (Figure 41).