

Making the Best of a SyS Situation

Data Quality for Syndromic Surveillance

3:15PM EST | November 18th, 2020 | 2020 SyS Symposium

Moderated by Harold Gil, MSPH – Washington State Department of Health

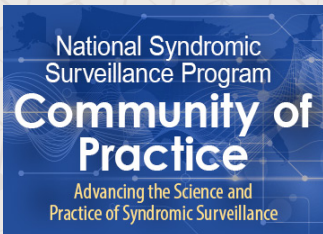
Speakers:

Dave Atrubin, MPH – Florida Department of Health

Viacony Herzi, MPH, MLS(ASCP)CM – Maricopa County Department of Public Health (Arizona) +

Sasha Harb, MPH – Pima County Health Department (Arizona)

PJ Gibson, BS – Indiana Department of Health



Council of State and Territorial Epidemiologists

Dave Atrubin, MPH
Florida Department of Health



Viacony Herzi– Maricopa County Department of Public Health (Arizona)
Sasha Harb, MPH – Pima County Health Department (Arizona)



Comparison of matched emergency department and inpatient hospital visits attributable to nonfatal opioid overdoses between hospital discharge and NSSP BioSense data ---Maricopa and Pima Counties, Arizona, 2018.

VIACONY HERZI, EPIDEMIOLOGIST
MARICOPA COUNTY DEPARTMENT OF PUBLIC HEALTH (MCDPH)

SASHA HARB, EPIDEMIOLOGIST
PIMA COUNTY HEALTH DEPARTMENT (PCHD)



ARIZONA DEPARTMENT
OF HEALTH SERVICES





1

Understand how Hospital Discharge and Syndromic Surveillance data compare to each other



2

Understand how health conditions such as nonfatal opioid overdoses are classified in each system



3

Provide a standardized protocol for how to match the data and do health condition comparisons.

Project Objectives

Background: Surveillance Systems for Opioid Overdose

Hospital Discharge Data (HDD)

- Bi-annual data distribution
- Administrative data
- ICD-10-CM diagnosis codes
- Valuable source about the patterns of care, public health, and the burden of chronic disease and injury morbidity

Syndromic Surveillance (NSSP ESSENCE)

- At minimum daily data
- Subset of hospitals who participate
- ICD 10 diagnosis codes & text
- Valuable data for an early alert for health events
- Syndrome categories that define health conditions (i.e. Chief Complaint and Discharge Diagnosis combinations)

Background: Maricopa County and Pima County (Arizona)

Large population and syndromic surveillance representation

- Maricopa County (MC) in 2018
 - Population: 4,294,460
 - 11 hospitals in ESSENCE/73 total hospitals
- Pima County (PC) in 2018
 - Population: 1,034,201
 - 5 hospitals in ESSENCE/17 total hospitals

METHODS: Matching

Match emergency department and inpatient hospital visit data from HDD and ESSENCE on unique patient-level variables

- MC: Facility Identifier and Patient Control Number (PCN)/Visit ID
- PC: Facility Identifier and Medical Records Number in combination with PCN, Discharge date and Admission date

METHODS: Classifying

Classify records according to definitions:

- HDD
 - CSTE Nonfatal Overdose Definition
- ESSENCE
 - CCDDcat: CDC Opioid Overdose v2, no changes
 - ICDonly: CDC Opioid Overdose v2, ICD codes only
 - CCDDext: CDC Opioid Overdose v2, applied to additional fields

METHODS: Classifying

Assess the extent of agreement between records identified in HDD and ESSENCE

- HDD as the “gold standard.”

	Case Definition Category 1	Case Definition Category 2	Case Definition Category 3
HDD	Nonfatal Opioid Overdoses	Nonfatal Opioid Overdoses	Nonfatal Opioid Overdoses
ESSENCE	CCDDcat	ICDonly	CCDDext

Examine visits where there was disagreement in classification

RESULTS:

Matching

County	Total Visits	Matched Visits	Percent of Visits Matched (2018)
MC	835,035	823,127	98.6%
PC	348,251	333,625	95.8%

RESULTS:

Classification

Among matched records, most ESSENCE queries classified more visits as a nonfatal opioid overdose than the HDD case definition in both counties.

	Case Definition Category 1 (HDD ~ CCDDcat)		Case Definition Category 2 (HDD ~ ICDonly)		Case Definition Category 3 (HDD ~ CCDDext)	
<i>County</i>	<i>MC</i>	<i>PC</i>	<i>MC</i>	<i>PC</i>	<i>MC</i>	<i>PC</i>
HDD	1,662	745	1,662	745	1,662	745
ESSENCE	2,972	1,462	1,196	609	2,994	1,462
% Agreement	99.8%	99.7%	99.9%	99.9%	99.8%	99.7%

RESULTS:

Classification by ESSENCE Query

Maricopa County		CCDDcat		ICDonly		CCDDext	
		Yes	No	Yes	No	Yes	No
HDD	Yes	1,326	336	1,127	535	1,340	322
	No	1,646	819,819	69	821,396	1,654	819,811

% Agreement: 99.8% 99.9% 99.8%

Pima County		CCDDcat		ICDonly		CCDDext	
		Yes	No	Yes	No	Yes	No
HDD	Yes	589	157	545	201	589	157
	No	875	332,004	66	332,813	875	332,004

% Agreement: 99.7% 99.9% 99.7%

RESULTS:

Why in ESSENCE but not HDD?

Referent diagnosis codes (ICD-10-CM)	Maricopa County (N = 1646)		Pima County (N = 875)	
	HDD	ESSENCE	HDD	ESSENCE
Opioid-related disorders (ORD) F-codes (F11.1-F11.9)	511*	34*	378*	28*
Opioid intoxication or overdose T-codes (T40.0X-T40.4x, T40.60, and T40.69)	56**	0	24**	0
Drug use, abuse, dependence F-codes (F12-F16, F18-F19)	245	0	84	0
Drug-poisoning T-codes (T36-T50)	142	0	87	0
No drug-related T-codes or F-codes	754	0	214	0

*F11 is not included in the CSTE case definition for acute opioid overdose identification

**Records with opioid overdose T-codes but not included in CSTE HDD case definition (ex T40605A)

- Records that do not have any drug-related T-codes and F-codes in any diagnosis field were likely identified in ESSENCE through terms like Narcan in the CCDD field.

RESULTS:

Why in HDD but not ESSENCE?

Referent diagnosis codes (ICD-10-CM)	Maricopa County (N = 336*)		Pima County (N = 157)	
	HDD	ESSENCE	HDD	ESSENCE
Opioid-related disorders (ORD) F-codes (F11.1-F11.9)	78	2	34	3
Opioid intoxication or overdose T-codes (T40.0X-T40.4x, T40.60, and T40.69)	336*	335	157*	157

*These records have opioid overdose T-codes that are included in the syndrome definitions such as T402X2A.

Conclusions

A high percentage (98.6%, MC; 95.8% PC) of the HDD and ESSENCE visits matched for Maricopa and Pima Counties.

This suggests that data collected in both systems have primarily the same records

However, there are variations of how nonfatal opioid overdoses are identified in each

These variations need to be considered before these two databases are combined.

Limitations

Opioid overdose definition changes

Hospital participation in ESSENCE may vary by jurisdiction

HDD is an administrative billing data set

Variability of hospital data in ESSENCE

Admission and discharge record differences between databases

Next Steps

1. Dissemination of standardized protocol for other AZ counties
 - Toolkit
 - Step by Step instructions
 - SAS program
 - Reference documents
 - Data dictionaries for both databases
 - Evaluation guide
2. Share with CSTE work groups such CSTE Injury Surveillance Workgroup to further study and may be improve case definition
3. Use the standardized protocol and toolkit to improve utilization of both HDD and ESSENCE for opioid overdose surveillance

Acknowledgements & Disclosure

Many thanks to the Center of Surveillance, Epidemiology, Laboratory Services, the National Center for Injury Prevention and Control, the Arizona Department of Health Services, the Maricopa County Department of Public Health, and the Pima County Health Department for their support of this project.

The project was supported by award # CDC-RFA-TP18-1802 Cooperative Agreement for Emergency Response: Public Health Crisis Response (Public Health Crisis Response NOFO).

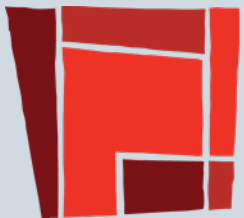
Disclosure: The contents of the presentation are solely the responsibility of the authors and do not reflect the official views of the US Department of Health and Human Services or the Centers for Disease Control and Prevention

Thank You!

VIACONY HERZI, EPIDEMIOLOGIST
CONTACT: VIACONY.HERZI@MARICOPA.GOV
MARICOPA COUNTY DEPARTMENT OF PUBLIC HEALTH (MCDPH)

SASHA HARB, EPIDEMIOLOGIST
CONTACT: SASHA.HARB@PIMA.GOV
PIMA COUNTY HEALTH DEPARTMENT (PCHD)

SYNDROMIC SURVEILLANCE PROGRAM
ARIZONA DEPARTMENT OF HEALTH SERVICES
SYNDROMICSURVEILLANCE@AZDHS.GOV



ARIZONA DEPARTMENT
OF HEALTH SERVICES



Maricopa County
Public Health
WeArePublicHealth.org

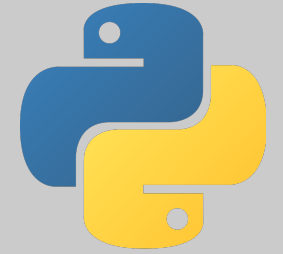


PIMA COUNTY
HEALTH DEPARTMENT



PJ Gibson, BS
Indiana Department of Health





ADTdq

- a python library for HL7 ADT data quality reporting

PJ Gibson

Syndromic Data Analyst



Indiana
Department
of
Health

The Problem

MSH|^~\&|EPICADT|DH|LABADT|DH|201301011226||ADT^A01|HL7MSG00001|P|2.3|

EVN|A01|201301011223||

PID|||MRN12345^5^M11||APPLESEED^JOHN^A^III||19710101|M||C|1 DATICA STREET^^MADISON^WI^53005-1020|GL|(414)379-1212|(414)271-3434|S||MRN12345001^2^M10|123456789|987654^NC|

NK1|1|APPLESEED^BARBARA^J|WIFE||||NK^NEXT OF KIN

PV1|1|I|2000^2012^01|||004777^GOOD^SIDNEY^J.||SUR|||ADM|A0|

*example HL7 ADT message, NOT real

- Difficult to read/parse
- High complexity -> hard to assess for DQ

A Solution

- Use NSSP Standards ([Data Dictionary](#))
 - Parse HL7 ADT messages for high-priority elements
 - Assess messages for element completion & validity

HL7 Archive Processed Name ▾	HL7 Archive Processing Description ▾	Processing Rationale ▾	Column Length ▾	Repeating field ▾	Priority ▾
Chief_Complaint_Text	OBX-5 segments where: * OBX-3 Observation Identifier is 8661-1 and/or 11292-0 * OBX-2 = "TX" or "CWE" or "CW" Select all non-null values and concatenate: IF OBX-2="TX" then Chief_Complaint_Text = OBX-5.1 IF OBX-2="CWE" or "CW" then Chief_Complaint_Text = concatenate(OBX-5.9, OBX-5.2, OBX-5.5)	The CWE Chief Complaint type is a legacy format of Chief Complaint. It is being included to capture legacy data feeds. The TX Chief Complaint is from PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015).	varchar(3000)	Yes	1

My Work

- 1 Created a function to parse HL7 messages for NSSP Priority Elements
- 2 Created supporting functions for
 - a) Data Visualization
 - b) Data Reporting
- 3 Packaged functions for others to use
(Catered to low-experience Python programmers)



1

NSSP_Element_Grabber

- Reads messages & parses them into neat data-frame
- ~550 messages parsed per minute

	Age_Reported	C_Patient_County	Age_Units_Reported	Race_Code	Ethnicity_Code	Ethnicity_Description	Patient_City
0	43	1793	YEAR	2131-1	2186-5	NON-HISPANIC	Naptown
1	5	1746	YEAR	2131-1	2186-5	NON-HISPANIC	Naptown
2	64	1826	NaN	2106-3	N	NaN	Naptown
3	74	1843	YEAR	2106-3	2186-5	Not Hispanic or Latino	Naptown
4	98	1791	YEAR	2106-3	2186-5	Not Hispanic or Latino	Naptown
...	...	...	...	...	...	...	...
9387	25	1753	NaN	2054-5	N	NaN	Naptown
9388	35	1853	YEAR	2131-1	2135-2	Hispanic or Latino	Naptown
9389	91	1734	YEAR	2054-5	2186-5	Not Hispanic or Latino	Naptown
9390	19	1788	YEAR	2054-5	2186-5	Not Hispanic or Latino	Naptown
9391	16	1548	YEAR	2131-1	2135-2	Hispanic or Latino	Naptown

*Records NOT real, randomized using examples

2a

Visualization_interactive

- Compares NSSP priority element completion across 2 time periods

- Heatmap expresses:

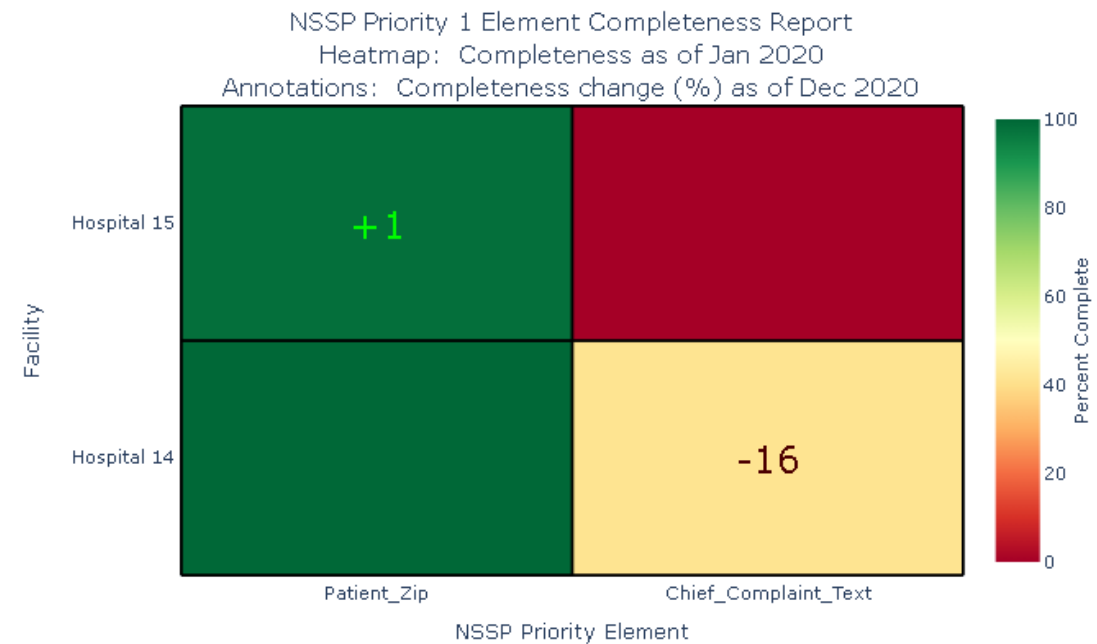
y-axis : facility names

x-axis : NSSP priority elements

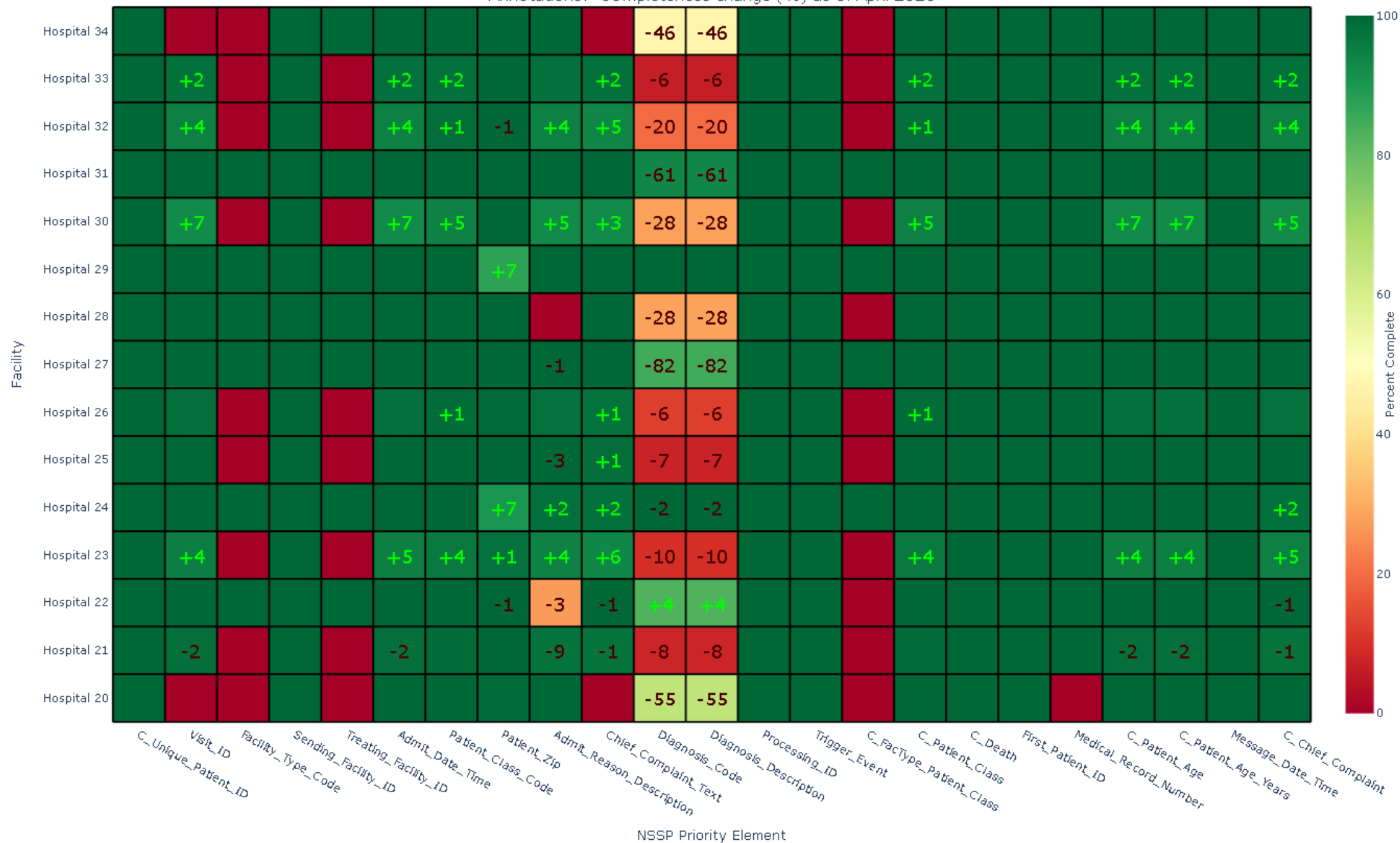
color : completion from time1 (%)

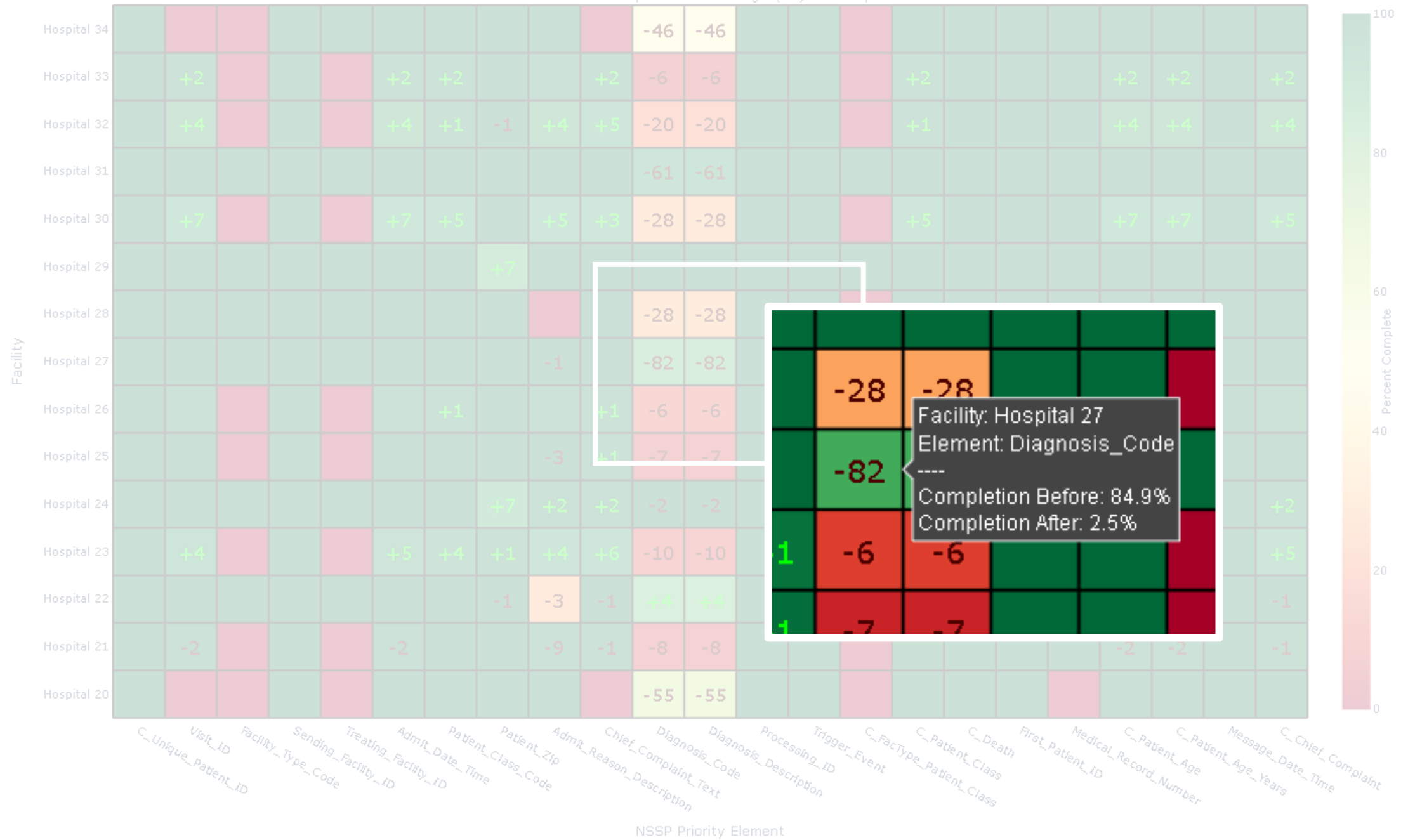
annotation: completion2 – completion1

- Can be saved as interactive HTML file



NSSP Priority 1 Element Completeness Report
 Heatmap: Completeness as of Jan 2020
 Annotations: Completeness change (%) as of April 2020







issues_in_messages

- Generates report on all errors in messages (missing or invalid)
- Can be exported neatly into Excel....may be large file

Grouper_ID	MESSAGE	Issue_Type	Element_Name	Priority	Description	Valid_Options	Message_Value	Suggestion	Comment
hospital MRN VisNum	some complicated message*blahblah 1	Missing or Null	Admit_Reason_Code	2	You must have a non-empty value for PV2-3.1 or PV2-3.4				
	some complicated message*blahblah 1	Invalid	Facility_Type_Code	1	You must have a OBX field where OBX-3 contains SS003 and where OBX-2='CWE'. When these criteria are met, the cooresponding OBX-5.1 or OBX-5.3 component must be non-empty	Value must be within the following list [261QE0002X, 1021-5,261QM2500X ,261QP2300X,261QU0200X]	225XN1300X	261QP2300X	You want a Fac_Type_Code representing Primary Care. Your current value set version is out of date. See the following link: https://phinvads.cdc.gov/vads/ViewValueSet.action?id=43E2BA38-DFDE-E411-8970-0017A477041A

3

ADTdq – a Python library

- Well documented on PyPi website
- Each function has lengthy descriptions with examples.
- Updates on horizon and as requested.

<https://pypi.org/project/ADTdq/>

The screenshot shows the PyPI project page for ADTdq 0.0.2. The page has a blue header with a search bar and navigation links (Help, Sponsor, Log in, Register). Below the header, the project name 'ADTdq 0.0.2' is displayed, along with a green 'Latest version' button and a release date of 'Released: Nov 4, 2020'. A code snippet 'pip install ADTdq' is shown. The main content area is divided into two columns. The left column contains a 'Navigation' menu with links to 'Project description', 'Release history', and 'Download files'. Below this is a 'Project links' section with a link to the 'Homepage'. The right column contains a 'Project description' section with the title 'ADTdq' and a link to the 'Github Project'. Below this is a 'Background' section with the title 'How it Started' and a paragraph of text. The bottom of the page features a 'Statistics' section with GitHub statistics (Stars: 0, Forks: 0, Open issues/PRs: 0) and a 'Functions' section.

ADTdq 0.0.2

pip install ADTdq

Released: Nov 4, 2020

A package designed to improve HL7 ADT Data Quality reporting in the field of public health informatics.

Navigation

- Project description
- Release history
- Download files

Project links

- Homepage

Statistics

GitHub statistics:

- Stars: 0
- Forks: 0
- Open issues/PRs: 0

View statistics for this project via [Libraries.io](#) or by using [our public](#)

Project description

ADTdq

[Github Project](#)

Background

How it Started

My name is PJ Gibson and I am a data analyst with the Indiana State Department of Health. My Informatics department arranged a grant with a group who could improve the quality of hospital reporting. We needed to track the progress of this hospital reporting, which required diving into HL7 Admission/Discharge/Transfer (ADT) messages and assessing for data completion and validity. Enter me.

The Goal

The main purpose of this package is to give data quality analysis functions to workers in public health informatics.

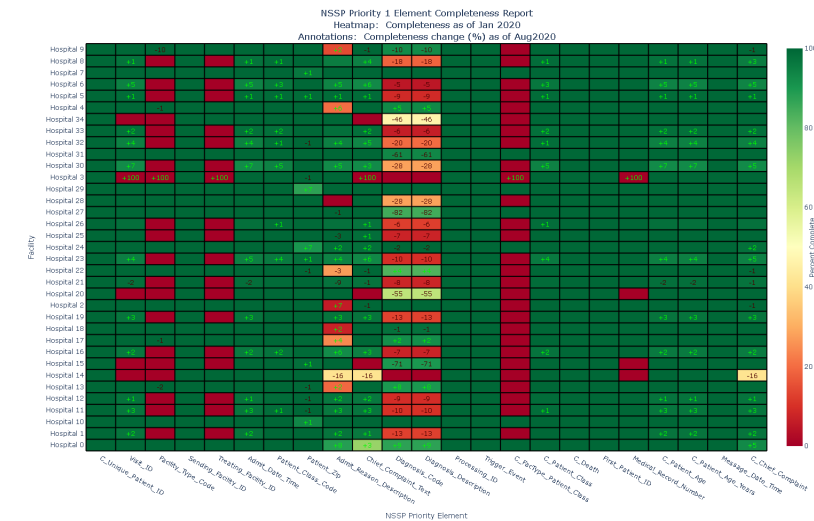
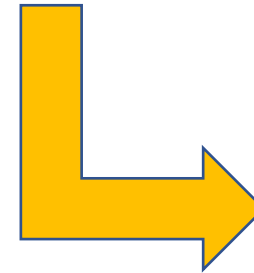
Functions

```

1  # Import my package.  Includes pandas
2  from ADTdq import *
3
4  # Read in raw-data (includes HL7 ADT MESSAGE field)
5  jan_data = pd.read_csv('January_pull.csv',usecols=['MESSAGE','FACILITY_NAME','PATIENT_MRN','PATIENT_VISIT_NUMBER'])
6  aug_data = pd.read_csv('August_pull.csv',usecols=['MESSAGE','FACILITY_NAME','PATIENT_MRN','PATIENT_VISIT_NUMBER'])
7
8  # put through NSSP parser function
9  before = NSSP_Element_Grabber(jan_data)
10 after = NSSP_Element_Grabber(aug_data)
11
12 # impliment visualization
13 Visualization_interactive(before,after,['Jan 2020','Aug2020'],priority='1',show_plot=True,outfile=True)

```

Python, it's not so bad!



Questions?

* For un-answered questions, please email me at **PGibson@isdh.IN.gov**

Downloading Python & installing ADTdq



Download Anaconda [here](#) –

- contains multiple environments for coding in Python (jupyter, spyder, ect.)



Command Prompt
App

Use Windows command prompt / terminal

- type in the following `pip install ADTdq`

Experiment yourself!

Request patches & updates?

National Syndromic Surveillance Program

Community of Practice

Advancing the Science and Practice of Syndromic Surveillance



Block Schedule

Training

Methods

Innovation

1-1:15PM EST	Daily Opening, Logistics, Overview		
1:15-2PM EST	Training: Call Up Your Data Using APIs	Breakout: Querying from Scratch - Cooking Up New Syndromes	Training: Data Sharing Step-by-Step Using the AMC
	...BREAK...		
2:15-3PM EST	Training: Telling Your Story - Insights and Tips to Promote Syndromic Surveillance	Breakout: The Art of Surveillance - Data Visualizations	Lightning: New Tactics for Old Problems
	...BREAK...		
3:15-4PM EST	Training: Dashboard Creation in NSSP ESSENCE	Breakout: Making the Best of a Sys Situation - Data Quality for Sys	Roundtable: COVID, Current Events, and More - Discussing Syndromic Surveillance Practices in 2020
	...BREAK...		
4:15-5PM EST	Virtual Happy Hour		