BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

Tweet about this Webinar! #BioSenseWebinars

BioSense v2.0 Webinar: Data Quality Checks and Assurance

Hosted by ISDS & the BioSense Redesign Team

May 22, 2014

Caleb Wiedeman, MPH
Epidemiologist
Tennessee Department of Health

Harold Gil, MSPH
Applied Public Health Informatics Fellow
County of San Diego Public Health Services

Learning Objectives

- **By the end of the Webinar, the audience will be able to:**

  - Describe processes that can be used to perform data quality checks in BioSense v2.0.
  - Identify specific data quality metrics in BioSense v2.0 (e.g., drops in visit counts).
  - Discuss the User Community Extension Project group's work on BioSense v2.0 data quality scripts.

Looking for CPH Credit for this Webinar?

- If you are seeking Certified in Public Health (CPH) recertification credit for this Webinar, please be sure to fill out the evaluation form at the end of the Webinar.

- One credit is available for attending this Webinar and completing the evaluation. The cost for non-members is $10. For ISDS members, the CPH credit is free.

- If you have any questions please contact us at syndromic@syndromic.org.

Upcoming ISDS Events

- ***TODAY: 2014 ISDS Conference Abstract Submission Site Opens***

- ***Meaningful Use Community Call***

Friday, June 6, 2014 – 1pm – 2pm EDT

- ***BioSense User Group (BUG) Meeting***

Tuesday, June 10, 2014 – 3pm – 4:30pm EDT

Visit www.syndromic.org for more information on upcoming events.

For more information about BioSense v2.0, please visit the BioSense Redesign Collaboration Web Site www.biosense2.org

New BioSense email address: BioSenseProgram@cdc.gov

BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

BioSense Data Quality: Tennessee's Experience

Caleb Wiedeman, MPH

Epidemiologist, Tennessee Department of Health

BioSense in TN

- **Two data feeds**
- **PHIN compliant, HL7 messages to TN (six Memphis area hospitals)**
  - Vendor to TN via SFTP
  - Routine message validation by TN (Informatics and IT)
  - TN to BioSense via SFTP
- **Non-standard CSV Files (three middle Tennessee hospitals)**
  - Hospital to regional health department via e-mail
  - Regional health department to TN via SFTP
  - Modified by TN and converted to CSV (epidemiologist)
  - TN to BioSense via SFTP

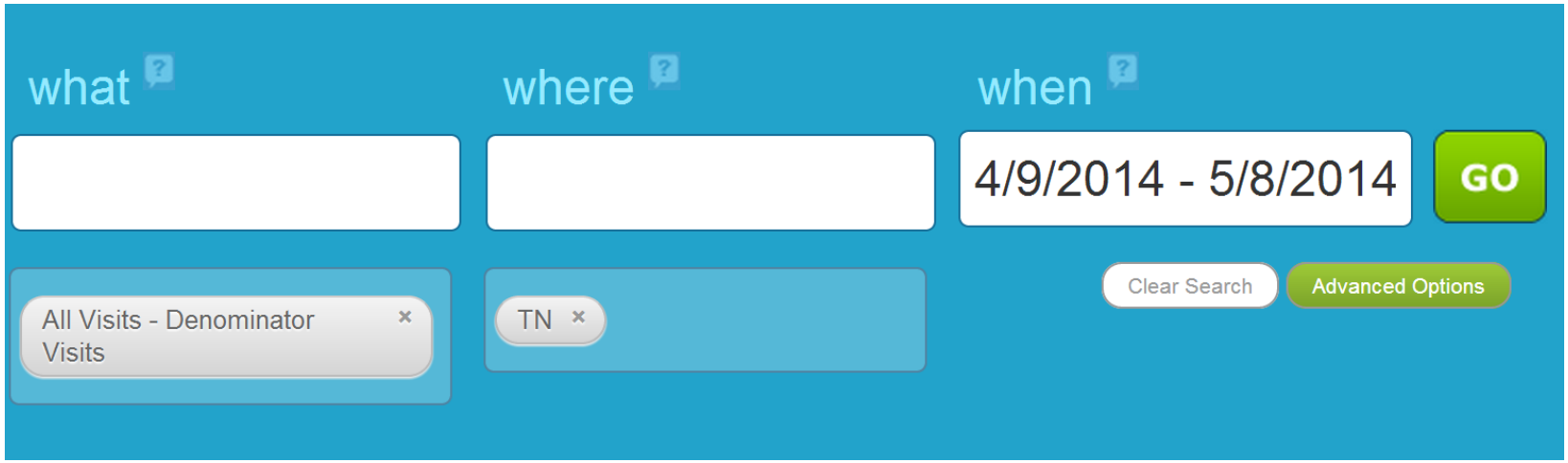
BioSense in TN

- **Regular checks of BioSense data**
  - Checking that data successfully transmitted
    - Visual check in frontend
    - More detailed queries and drill down in backend
  - Checking for syndrome aberrations
  - E-mail alerts set (but not functioning appropriately due to TN data tardiness)

BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

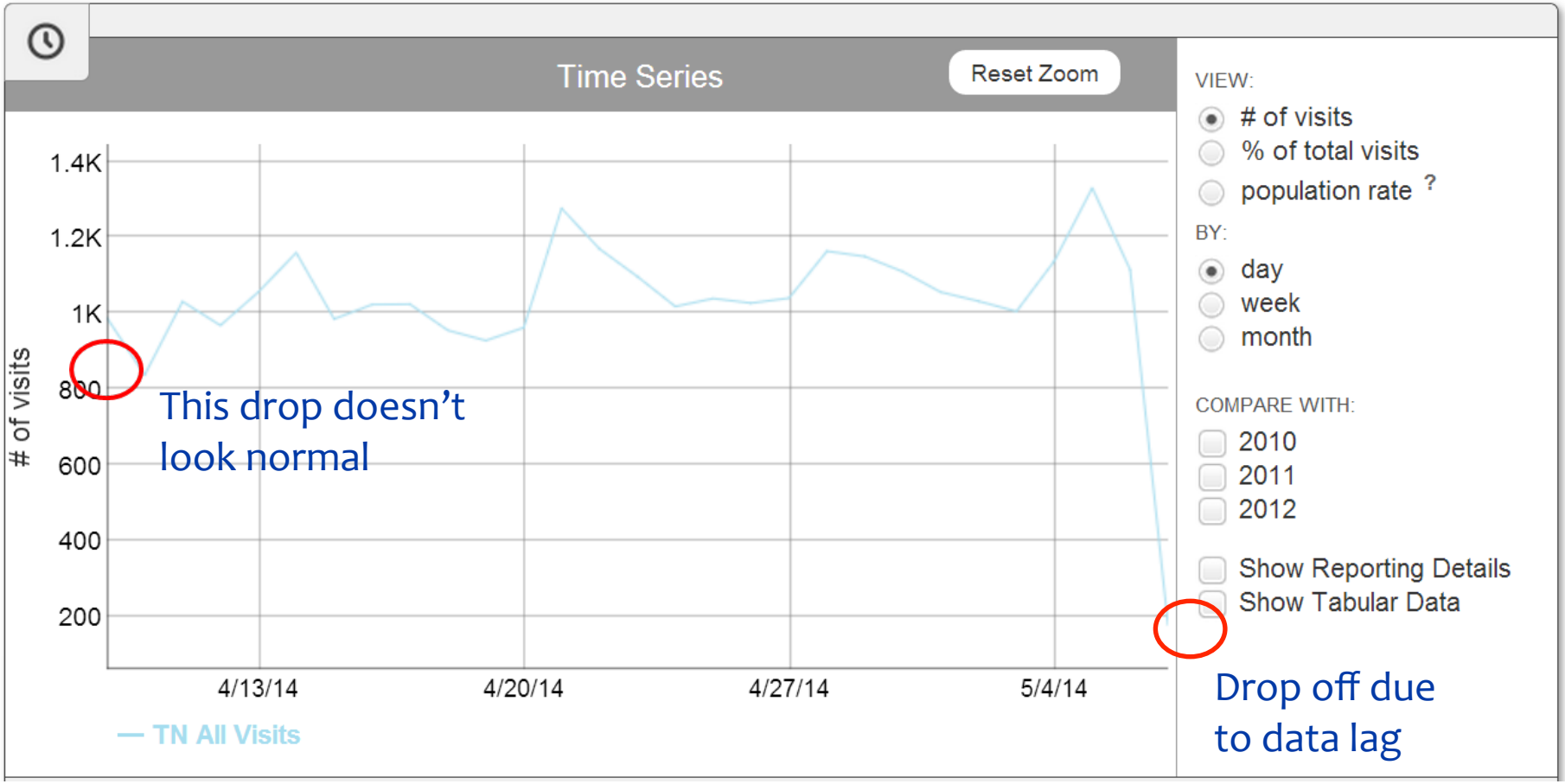
Viewing Data in BioSense

- **Checking the frontend to make sure previous day's data loaded**
  - All visits
  - TN only
  - Set source as TN only



Typical 1 Month of Data

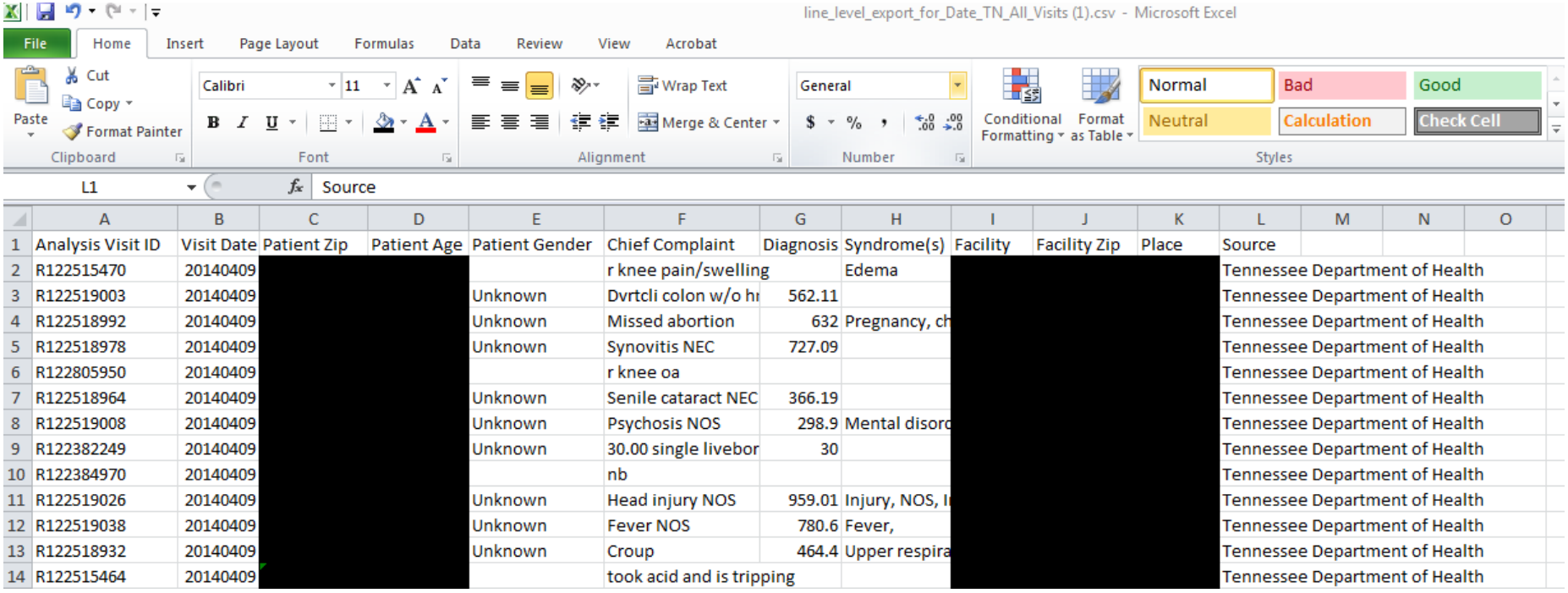
Viewing **29,834** visits for 1 syndrome in 1 location from **4/9/2014 - 5/7/2014**? from 1 source for **Male, Female, and Unknown**, ages **all ages**.



Export Frontend Data

- **On-the-fly tabling to get more info**
  - Frontend exports into CSV file
    - Data elements are “Analysis Visit ID” (which is linkable to backend binned data), “Visit Date”, “Patient Zip”, “Patient Age”, “Patient Gender”, “Chief Complaint”, “Diagnosis” (diagnosis code), “Facility”, “Facility Zip”, “Place” (county in TN’s case), and “Source”
  - CSV file can be quickly made into a pivot table in Excel

Example of Exported CSV File



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Analysis Visit ID	Visit Date	Patient Zip	Patient Age	Patient Gender	Chief Complaint	Diagnosis	Syndrome(s)	Facility	Facility Zip	Place	Source			
2	R122515470	20140409	[redacted]			r knee pain/swelling		Edema	[redacted]			Tennessee Department of Health			
3	R122519003	20140409			Unknown	Dvrtcli colon w/o hı	562.11					Tennessee Department of Health			
4	R122518992	20140409			Unknown	Missed abortion	632	Pregnancy, ch				Tennessee Department of Health			
5	R122518978	20140409			Unknown	Synovitis NEC	727.09					Tennessee Department of Health			
6	R122805950	20140409				r knee oa						Tennessee Department of Health			
7	R122518964	20140409			Unknown	Senile cataract NEC	366.19					Tennessee Department of Health			
8	R122519008	20140409			Unknown	Psychosis NOS	298.9	Mental disorc				Tennessee Department of Health			
9	R122382249	20140409			Unknown	30.00 single livebor	30					Tennessee Department of Health			
10	R122384970	20140409				nb						Tennessee Department of Health			
11	R122519026	20140409			Unknown	Head injury NOS	959.01	Injury, NOS, I				Tennessee Department of Health			
12	R122519038	20140409			Unknown	Fever NOS	780.6	Fever,				Tennessee Department of Health			
13	R122518932	20140409			Unknown	Croup	464.4	Upper respira				Tennessee Department of Health			
14	R122515464	20140409				took acid and is tripping						Tennessee Department of Health			

Pivot Table for Exported Data

Count of Visit Date	Column L									
Row Labels	Facility 1	Facility 2	Facility 3	Facility 4	Facility 5	Facility 6	Facility 7	Facility 8	Facility 9	Grand Total
20140409	105	108	9	25	229	120	127	159	106	988
20140410		49		16	229	132	152	148	110	836
20140411	89	141	9	17	222	125	140	158	129	1030
20140412	104	82	10	23	231	124	144	136	113	967
20140413	103	90	17	18	279	80	146	180	142	1055
20140414	134	162	18	20	250	139	162	166	108	1159
20140415	87	122	14	16	217	116	147	156	109	984
20140416	91	131	15	27	218	130	151	142	117	1022
20140417	98	138	15	22	221	106	158	155	110	1023
20140418	92	129	13	23	196	120	121	138	122	954
20140419	88	103	13	27	186	108	124	155	123	927
20140420	81	119	12	12	207	109	128	166	127	961
20140421	125	139	18	28	307	142	190	178	149	1276
20140422	94	153	16	16	261	160	170	191	108	1169
20140423	101	122	17	14	233	150	161	178	120	1096
20140424	110	127	20	13	201	135	142	152	117	1017
20140425	112	149	11	25	179	135	136	174	117	1038
20140426	103	79	16	16	233	130	133	187	129	1026
20140427	106	102	14	18	227	110	151	183	128	1039
20140428	105	131	14	16	208	143	187	226	133	1163
20140429	102	132	13	16	243	150	171	197	125	1149
20140430	93	146	22	19	193	160	167	185	124	1109
20140501	86	117	13	17	200	146	163	185	128	1055
20140502	69	126	11	22	204	149	139	187	124	1031
20140503	104	93	10	28	193	140	139	154	143	1004
20140504	99	97	15	23	259	149	177	172	146	1137
20140505	119	104	17	30	291	192	180	236	161	1330
20140506	90	92	8	23	259	146	196	171	129	1114
20140507	110		8	1	12	12	15	13	4	175
Grand Total	**2800**	**3283**	**388**	**571**	**6388**	**3758**	**4317**	**4828**	**3501**	**29834**

Drilling Down Further

- **Data appear to be missing from the frontend**
  - Could be missing for a number of reasons
    - Data were received by backend, but missed window of opportunity to be loaded to the frontend
    - Data were received, but never processed
    - Data were never received by BioSense; failed in transport
  - Are data present in the backend?

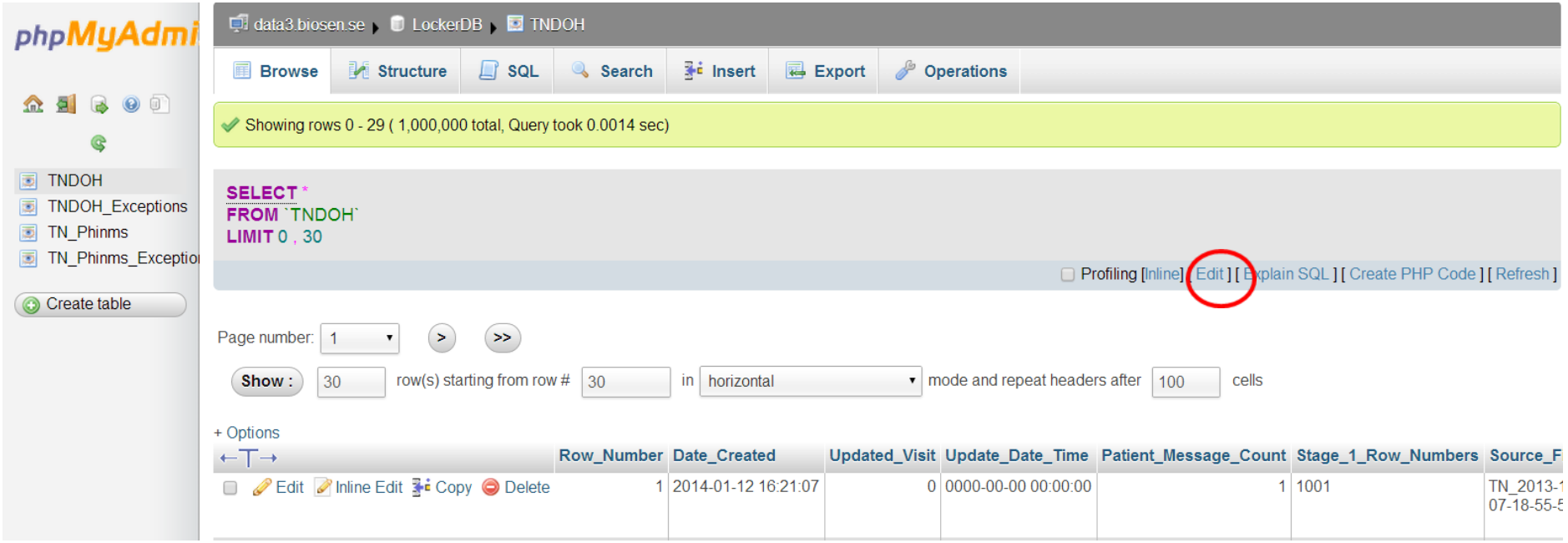
- **Need to look at backend data to answer questions**

phpMyAdmin

- Can quickly generate count tables to diagnose small data issues
- Can pull subsets of data and analyze using an external analysis package
- MySQL code used to generate tables and data subsets.

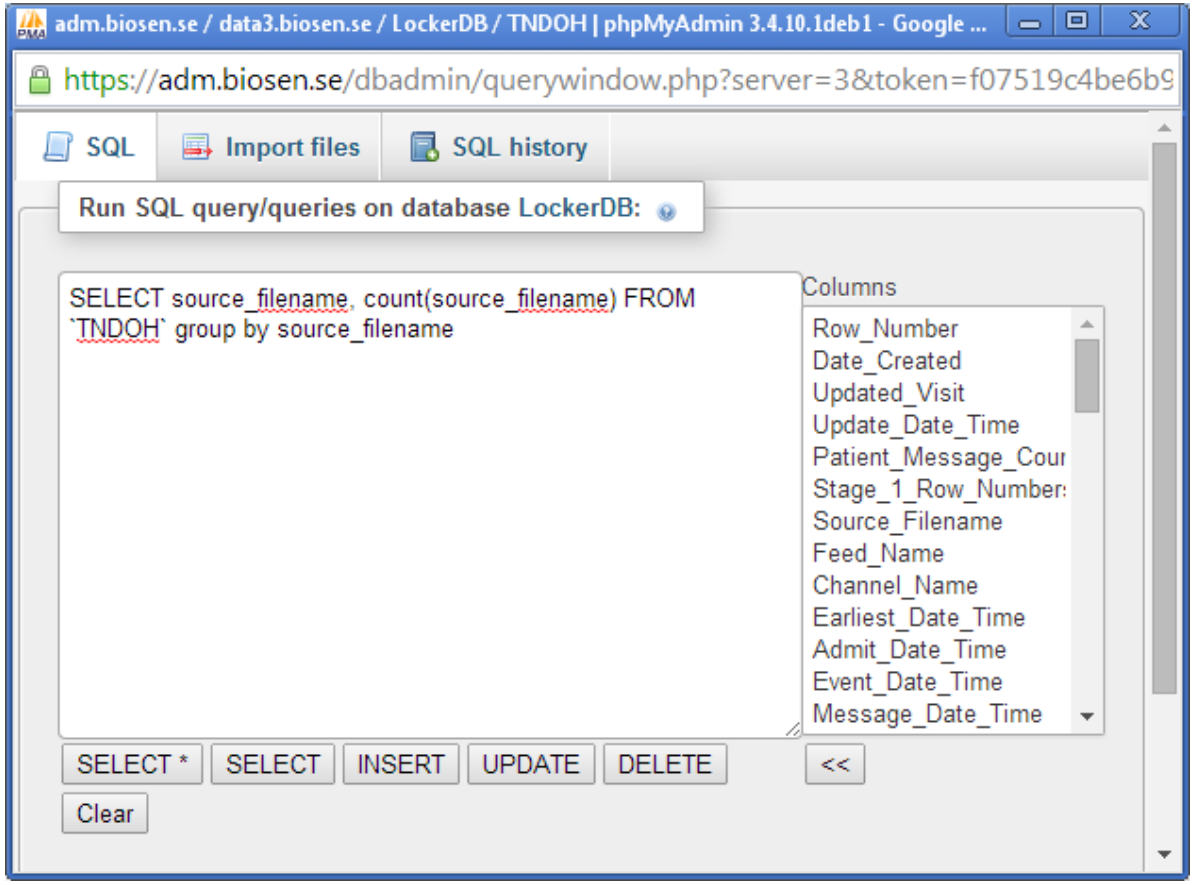
Submitting Queries

- Submit MySQL queries to data by clicking the "Edit" link



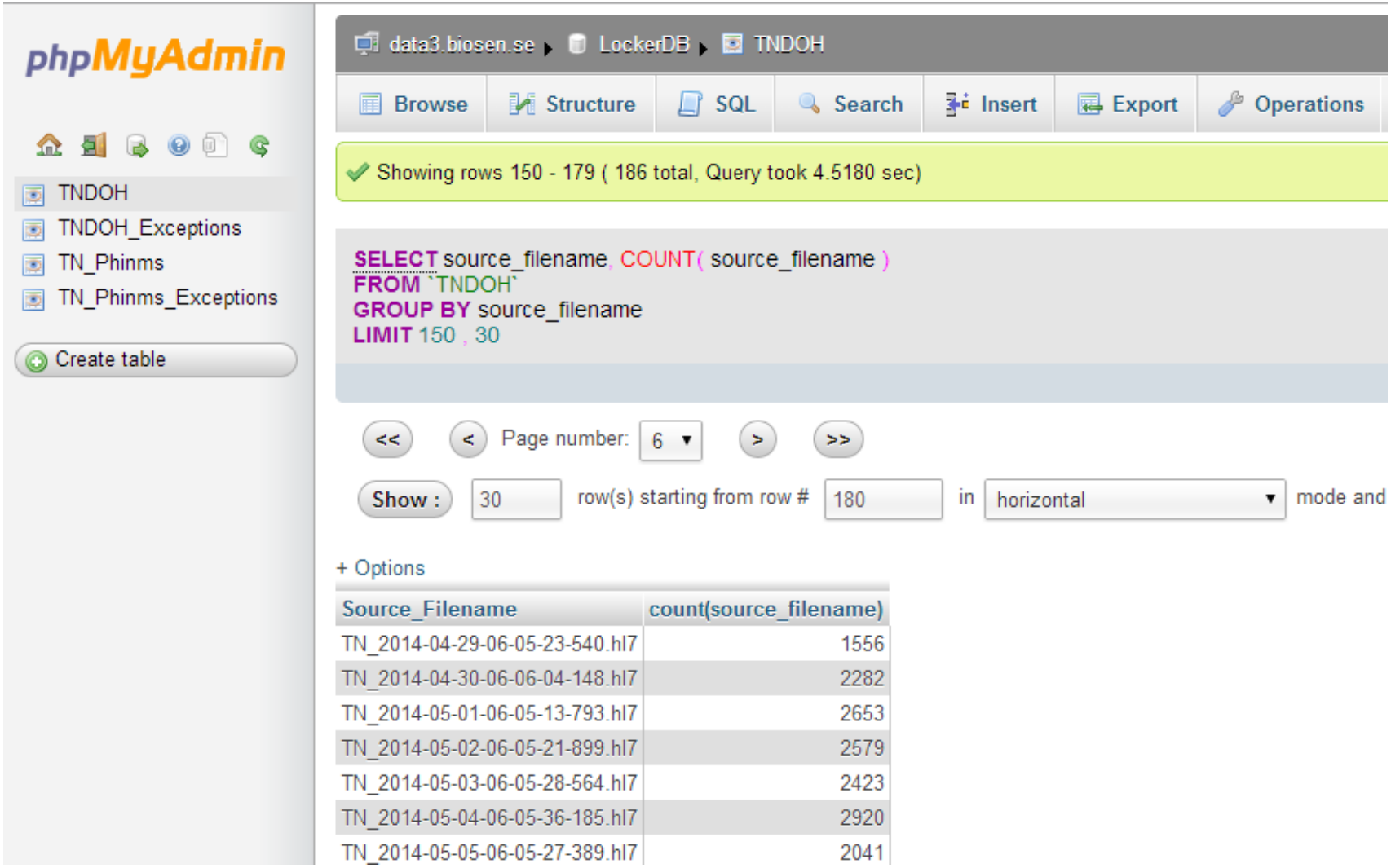
Daily Record Queries

- Check to see if BioSense received and processed files



BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

Query Result



Source_Filename	count(source_filename)
TN_2014-04-29-06-05-23-540.hl7	1556
TN_2014-04-30-06-06-04-148.hl7	2282
TN_2014-05-01-06-05-13-793.hl7	2653
TN_2014-05-02-06-05-21-899.hl7	2579
TN_2014-05-03-06-05-28-564.hl7	2423
TN_2014-05-04-06-05-36-185.hl7	2920
TN_2014-05-05-06-05-27-389.hl7	2041

Calendar Day Tables

- **Can approximate visits by using "distinct" call on unique_visiting_id**
  - Approximate visits can be broken out by day, facility, etc. using SQL code

- **Code for Tennessee's locker would be:**

```
SELECT date(earliest_date_time), facility_name,
count(distinct unique_visiting_id)
FROM `TNDOH`
GROUP BY 1, 2
```

Visit Dates, Facilities, and Approximate Visit Counts

```
SELECT DATE( earliest_date_time ) , facility_name, COUNT( DISTINCT unique_visiting_id )
FROM `TNDOH`
GROUP BY 1 , 2
LIMIT 3225 , 30
```

<< < Page number: 108 > >>

Show : 30 row(s) starting from row # 3255 in horizontal moc

+ Options

date(earliest_date_time)	Facility_Name	count(distinct unique_visiting_id)
2014-04-09		135
2014-04-09		138
2014-04-09		105
2014-04-09		9
2014-04-09		123
2014-04-10		264
2014-04-10		16
2014-04-10		123
2014-04-10		157
2014-04-10		159
2014-04-10		150
2014-04-10		90
2014-04-10		10
2014-04-10		150

Looking at Facilities by File by Date for a Date Range

- SELECT facility_name, source_filename, date(earliest_date_time), count(distinct unique_visiting_id)
  FROM `TNDOH`
  WHERE earliest_date_time BETWEEN "2014-04-10 00:00:00" and "2014-05-07 23:59:59" GROUP BY 1, 2, 3

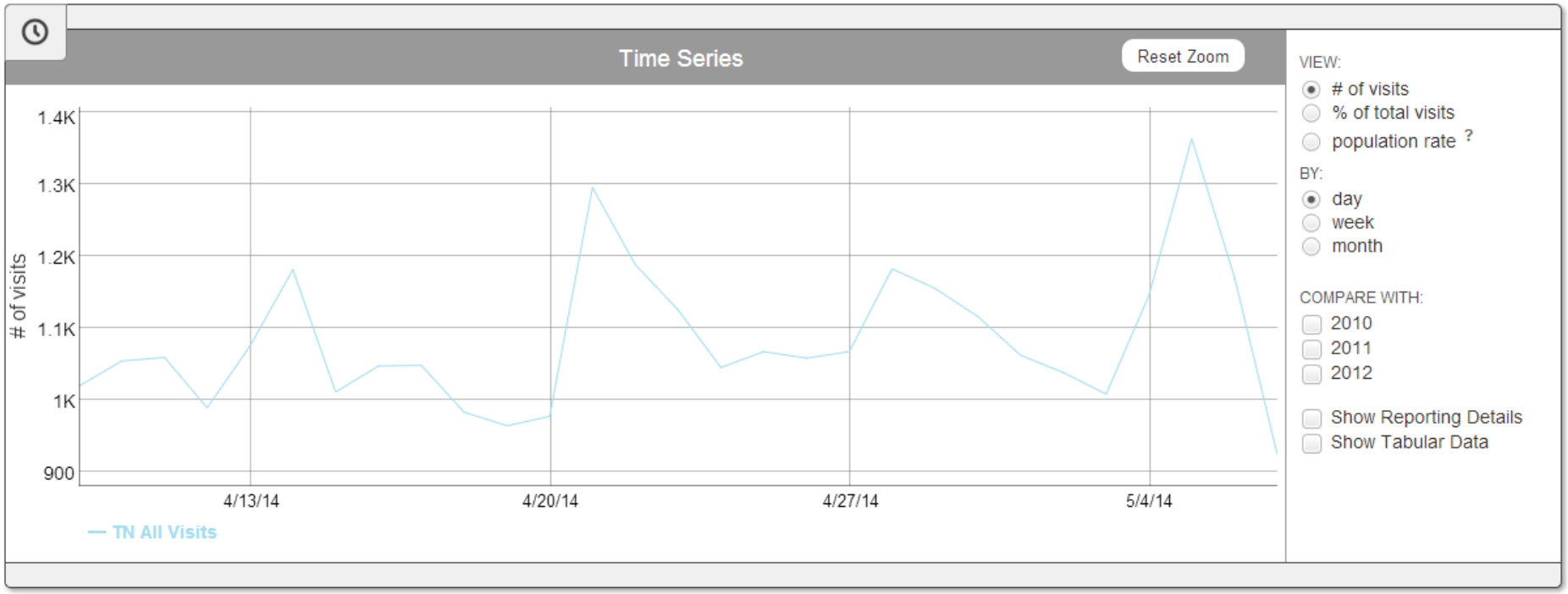
Facility_Name	Source_Filename	date(earliest_date_time)	count(distinct unique_visiting_id)
	TN_2014-05-08-06-05-50-307.hl7	2014-05-07	16
	TN_2014-05-09-06-05-30-457.hl7	2014-05-07	134
	TN_2014-05-09-06-05-30-535.hl7	2014-05-07	17
	TN_2014-05-13-06-05-29-401.hl7	2014-05-07	1
	TN_2014-05-14-06-05-18-538.hl7	2014-05-07	50
	TN_2014-05-08-06-05-50-307.hl7	2014-05-07	1
	TN_2014-05-09-06-05-30-457.hl7	2014-05-07	10
	TN_2014-05-09-06-05-30-535.hl7	2014-05-07	4
	TN_2014-05-14-06-05-18-538.hl7	2014-05-07	13
	TN_2014-05-08-06-05-50-307.hl7	2014-05-07	4
	TN_2014-05-09-06-05-30-457.hl7	2014-05-07	69
	TN_2014-05-09-06-05-30-535.hl7	2014-05-07	9
	TN_2014-05-13-06-05-29-401.hl7	2014-05-07	3
	TN_2014-05-14-06-05-18-538.hl7	2014-05-07	44

DQ Next Steps

- Records appear to be missing on frontend
- Backend queries show that records should exist
- Solution: Request frontend refresh for April 2014 to date from biosenseprogram@cdc.gov
- Check frontend data after successful refresh

BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

Viewing **31,415** visits for 1 syndrome in 1 location from **4/9/2014 - 5/7/2014** from 1 source for **Male, Female, and Unknown**, ages **all ages**.



Other Useful Queries

Looking at Visit Dates Contained in Transmitted Files

- SELECT source_filename, date(earliest_date_time), count(distinct unique_visiting_id)
  FROM `TNDOH`
  GROUP BY 1, 2

Looking at Visit Dates Contained in Transmitted Files

```
SELECT source_filename, DATE( earliest_date_time ) , COUNT( DISTINCT unique_visiting_id )
FROM `TNDOH`
GROUP BY 1 , 2
LIMIT 270 , 30
```

<< < Page number: 10 > >>

Show : 30 row(s) starting from row # 300 in horizontal mo

+ Options

Source_Filename	date(earliest_date_time)	count(distinct unique_visiting_id)
TN_2013-12-26-06-05-57-872.hl7	2013-12-23	3
TN_2013-12-26-06-05-57-872.hl7	2013-12-24	723
TN_2013-12-26-06-05-57-872.hl7	2013-12-25	65
TN_2013-12-27-06-06-01-409.hl7	2013-12-10	1
TN_2013-12-27-06-06-01-409.hl7	2013-12-11	2
TN_2013-12-27-06-06-01-409.hl7	2013-12-15	4
TN_2013-12-27-06-06-01-409.hl7	2013-12-16	28
TN_2013-12-27-06-06-01-409.hl7	2013-12-17	19
TN_2013-12-27-06-06-01-409.hl7	2013-12-18	2
TN_2013-12-27-06-06-01-409.hl7	2013-12-20	17
TN_2013-12-27-06-06-01-409.hl7	2013-12-21	28
TN_2013-12-27-06-06-01-409.hl7	2013-12-24	19
TN_2013-12-27-06-06-01-409.hl7	2013-12-25	680
TN_2013-12-27-06-06-01-409.hl7	2013-12-26	78

Linking frontend data to backend data

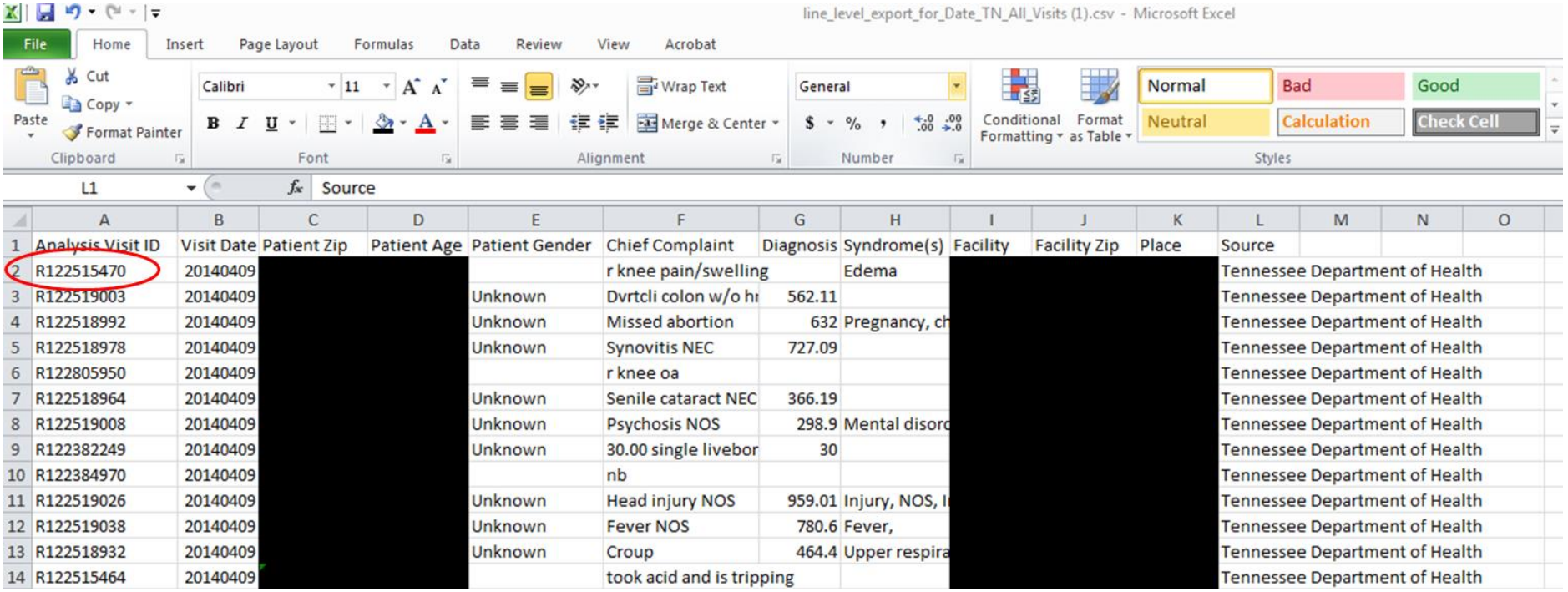
- **Data in the frontend are based on a pull of binned data**
  - Binned data are separated into two tables
  - Raw, binned, and frontend data can be linked using Visit ID

- **Frontend variable in line-level export is "Analysis Visit ID"**
  - The "R" at the beginning should be removed

- **Backend variable in the binned data (adm2.biosen.se) is "analysisvisitid"**
  - Can use "patientid" to link to the raw data (data3.biosen.se) variable "Unique_Patient_ID"

BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

Frontend Export



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Analysis Visit ID	Visit Date	Patient Zip	Patient Age	Patient Gender	Chief Complaint	Diagnosis	Syndrome(s)	Facility	Facility Zip	Place	Source			
2	R122515470	20140409	[redacted]			r knee pain/swelling		Edema	[redacted]			Tennessee Department of Health			
3	R122519003	20140409			Unknown	Dvrtcli colon w/o hr	562.11					Tennessee Department of Health			
4	R122518992	20140409			Unknown	Missed abortion	632	Pregnancy, ch				Tennessee Department of Health			
5	R122518978	20140409			Unknown	Synovitis NEC	727.09					Tennessee Department of Health			
6	R122805950	20140409				r knee oa						Tennessee Department of Health			
7	R122518964	20140409			Unknown	Senile cataract NEC	366.19					Tennessee Department of Health			
8	R122519008	20140409			Unknown	Psychosis NOS	298.9	Mental disorc				Tennessee Department of Health			
9	R122382249	20140409			Unknown	30.00 single livebor	30					Tennessee Department of Health			
10	R122384970	20140409				nb						Tennessee Department of Health			
11	R122519026	20140409			Unknown	Head injury NOS	959.01	Injury, NOS, I				Tennessee Department of Health			
12	R122519038	20140409			Unknown	Fever NOS	780.6	Fever,				Tennessee Department of Health			
13	R122518932	20140409			Unknown	Croup	464.4	Upper respira				Tennessee Department of Health			
14	R122515464	20140409				took acid and is tripping						Tennessee Department of Health			

Binned Data Query

- SELECT analysisvisitid, patientid, dateofvisit, patientzip, age, activitytext, activitycode, binvalue
  FROM `TNDOH_CC`
  WHERE `analysisvisitid` LIKE "122515470"

```
SELECT analysisvisitid, patientid, dateofvisit, patientzip, age, activitytext, activitycode, binvalue
FROM `TNDOH_CC`
WHERE `analysisvisitid` LIKE "122515470"
LIMIT 0 , 30
```

Profiling [Inline] [Edit

Show : 30 row(s) starting from row # 0 in horizontal mode and repeat headers after 100 cells

+ Options

	analysisvisitid	patientid	dateofvisit	patientzip	age	activitytext	activitycode	binvalue
Edit Inline Edit Copy Delete	122515470	V00032628893	2014-04-09	[redacted]		r knee pain/swelling	*NULL*	Edema
Edit Inline Edit Copy Delete	122515470	V00032628893	2014-04-09	[redacted]		r knee pain/swelling	*NULL*	Edema

- Note: Don't use `analysisvisitid` = 122515470
  It doesn't work.

Raw Data Query

- SELECT *
  FROM `TNDOH`
  WHERE unique_patient_id = "v00032628893“

```
SELECT *
FROM `TNDOH`
WHERE unique_patient_id = "v00032628893"
LIMIT 0 , 30
```

Show : 30 row(s) starting from row # 0 in horizontal mode and

+ Options

	Row_Number	Date_Created	Updated_Visit	Update_Da	Unique_Patient_ID	Chief_Complaint
Edit Inline Edit Copy Delete	418819	2014-04-16 11:49:10	0	0000-00-00	V00032628893	r knee pain/swelling

- One raw record returned; often more than one raw record will be returned

UCEP Data Quality Script

- **Simple functionality**
- **Provides standard data frame names for writing other R code**
  - Data frames already de-duplicated
  - Allows for easier sharing of code across jurisdictions
- **R provides a lot of power and flexibility**
  - On-the-fly keyword or code searching
  - Custom syndrome creation
  - Attractive graphics
  - More detailed data cleaning
  - Import and export external files

Questions?

Caleb Wiedeman
Caleb.Wiedeman@TN.gov

BioSense 2.0
Public Health Surveillance Through Collaboration
https://biosen.se

BioSense UCEP: Data Quality Tools

Harold Gil
Applied Public Health Informatics Fellow
County of San Diego, HHSA, Public Health Services

Contents

- UCEP Background
- Data Quality Metrics
- Binning Map
- Syndrome Definitions Comparison
- Future Plans

UCEP

- The User Community Extension Project (UCEP) is an informal collaboration of public health practitioners across jurisdictions.

- The group focuses on developing and sharing tools that complement BioSense functionality.

UCEP

- **Two current quality assurance projects:**
  - Data Quality Metrics
  - Binning Map

UCEP

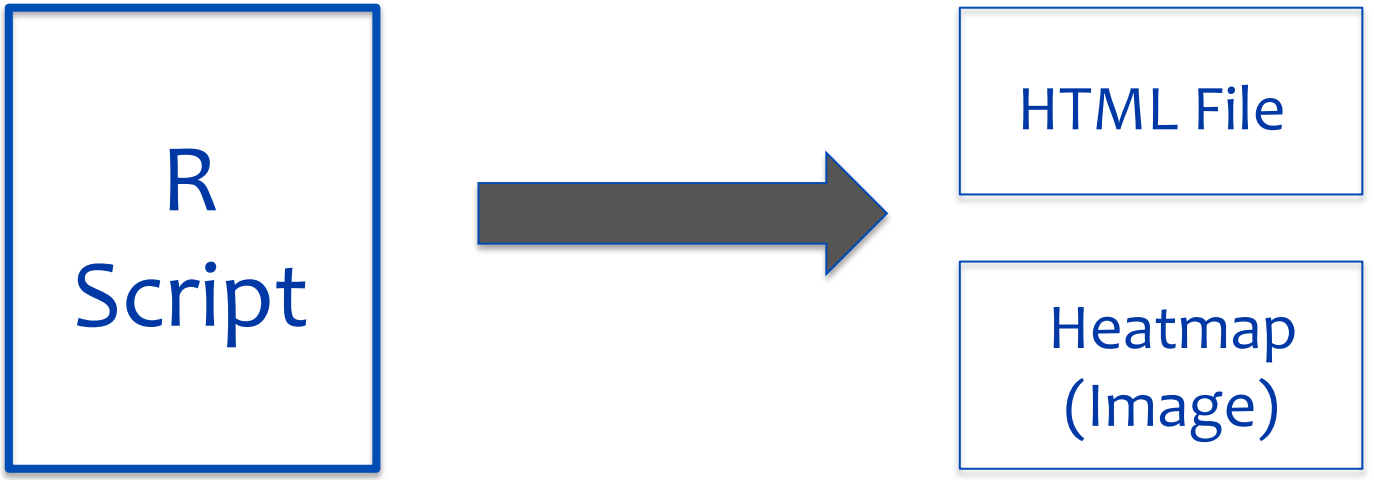
- **Data Quality Metrics**
  - Table: Aggregate level
  - Table: Hospital-stratified level
  - Heatmap: Hospital-stratified level

UCEP

- **Data Quality Metrics**
  - Table: Aggregate level
  - Table: Hospital-stratified level
  - Heatmap: Hospital-stratified level

- **Binning Map**
  - CSV file: four raw record fields used to create binned record ([CC, DC, DT, PC]---> syndrome)

Data Quality Metrics Script



Script Output

HTML File

Heatmap
(Image)

Script Output

BioSense - Data Quality Metrics

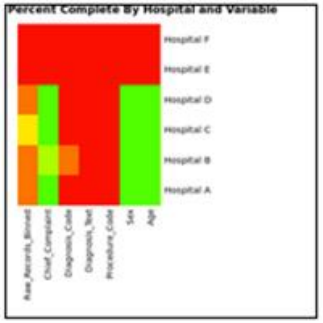
Time frame: '2014-04-07 00:00:00' - '2014-05-06 23:59:59'

Results

Aggregate summary metrics are shown in the table below.

	Total_Num_Visits	Total_Num_Binned	Total_Perc_Binned
1	22329	7219	32.33

Heatmap of stratified data quality metrics (at the facility level) is shown below.



Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Script Output

BioSense - Data Quality Metrics

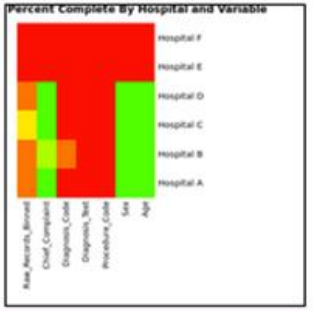
Time frame: '2014-04-07 00:00:00' - '2014-05-06 23:59:59'

Results

Aggregate summary metrics are shown in the table below.

	Total_Num_Visits	Total_Num_Binned	Total_Perc_Binned
1	22329	7219	32.33

Heatmap of stratified data quality metrics (at the facility level) is shown below.



Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Aggregate Level

Time frame: '2014-04-07 00:00:00' - '2014-05-06 23:59:59'

Results

Aggregate summary metrics are shown in the table below.

	Total_Num_Visits	Total_Num_Binned	Total_Perc_Binned
1	30727	11128	36.22

Table: Aggregate-level

Time frame: '2014-04-07 00:00:00' - '2014-05-06 23:59:59'

Results

Aggregate summary metrics are shown in the table below.

	Total_Num_Visits	Total_Num_Binned	Total_Perc_Binned
1	30727	11128	36.22

Table: Aggregate-level

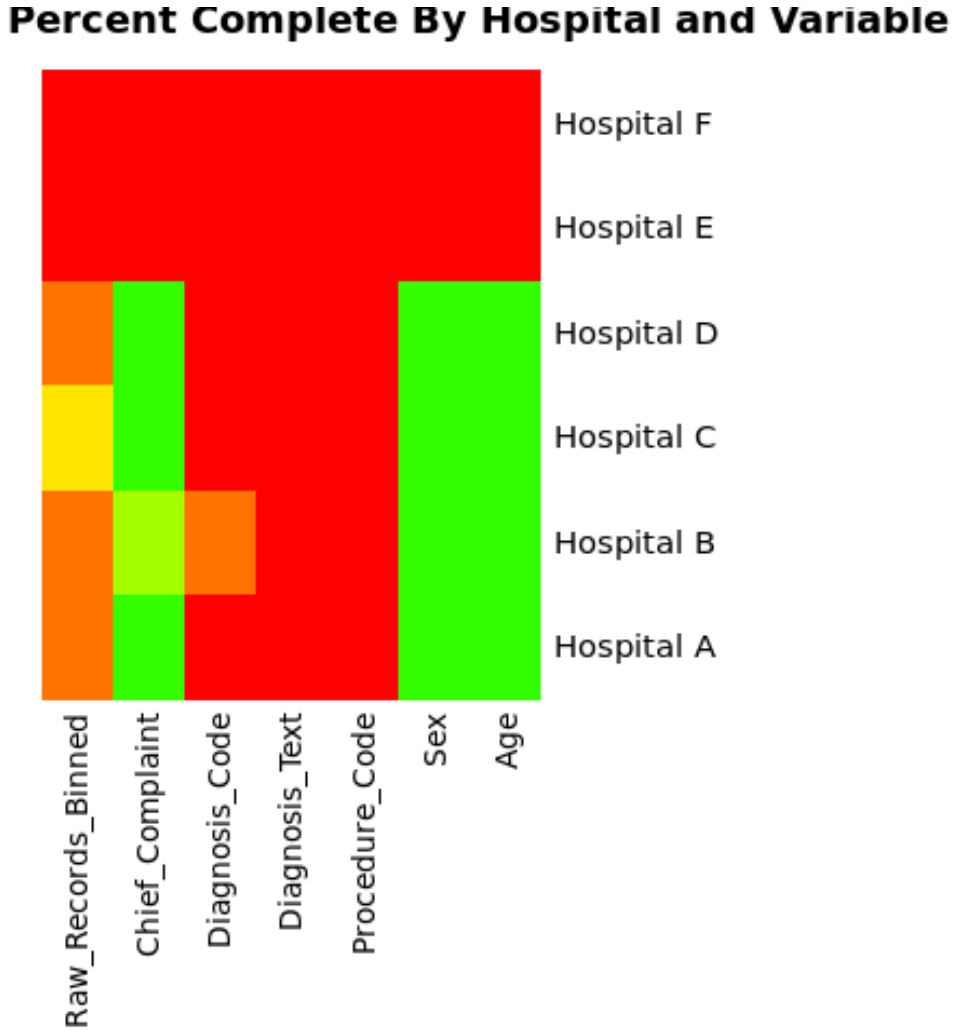
Time frame: '2014-04-07 00:00:00' - '2014-05-06 23:59:59'

Results

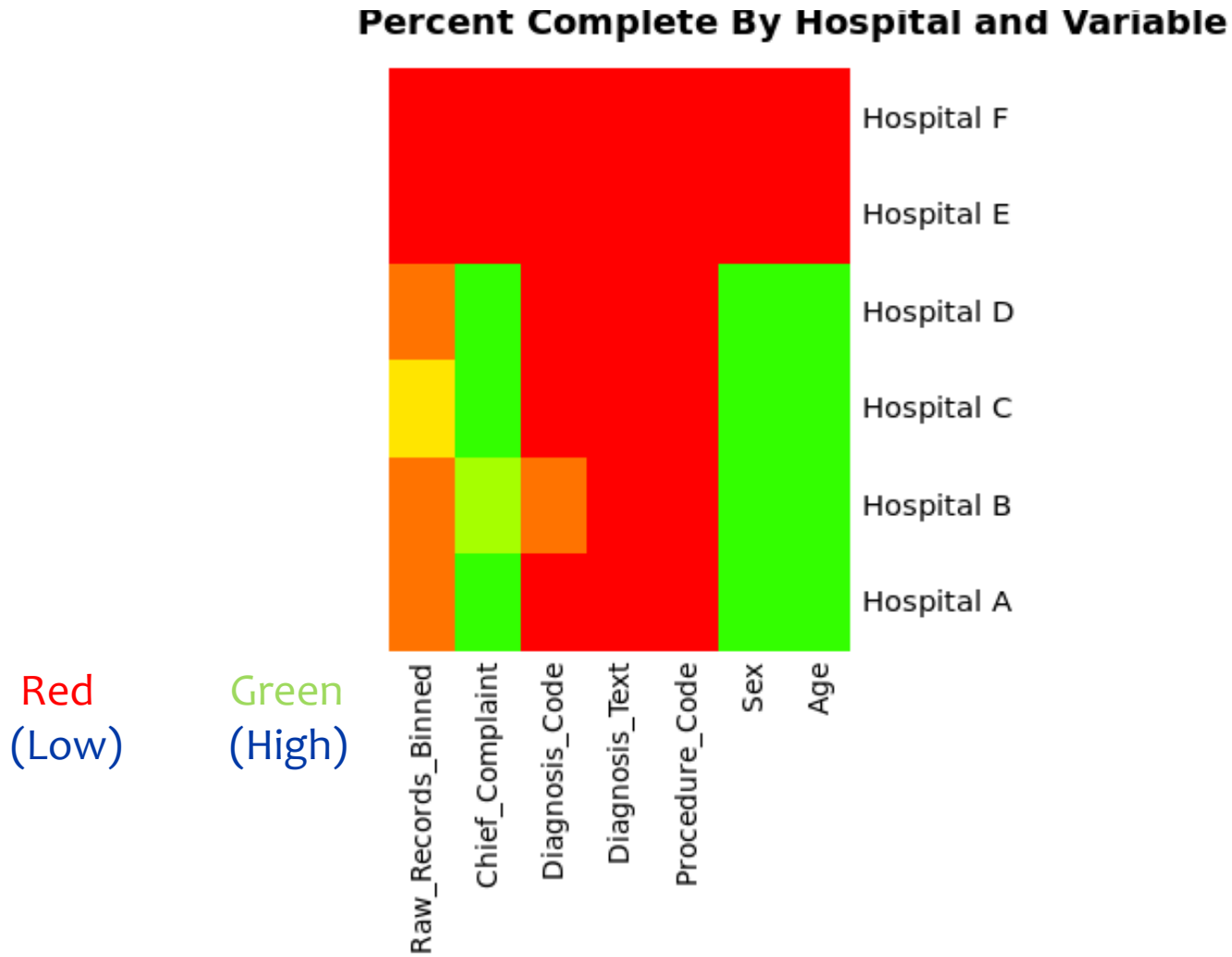
Aggregate summary metrics are shown in the table below.

	Total_Num_Visits	Total_Num_Binned	Total_Perc_Binned
1	30727	11128	36.22

Heatmap: Hospital-stratified level



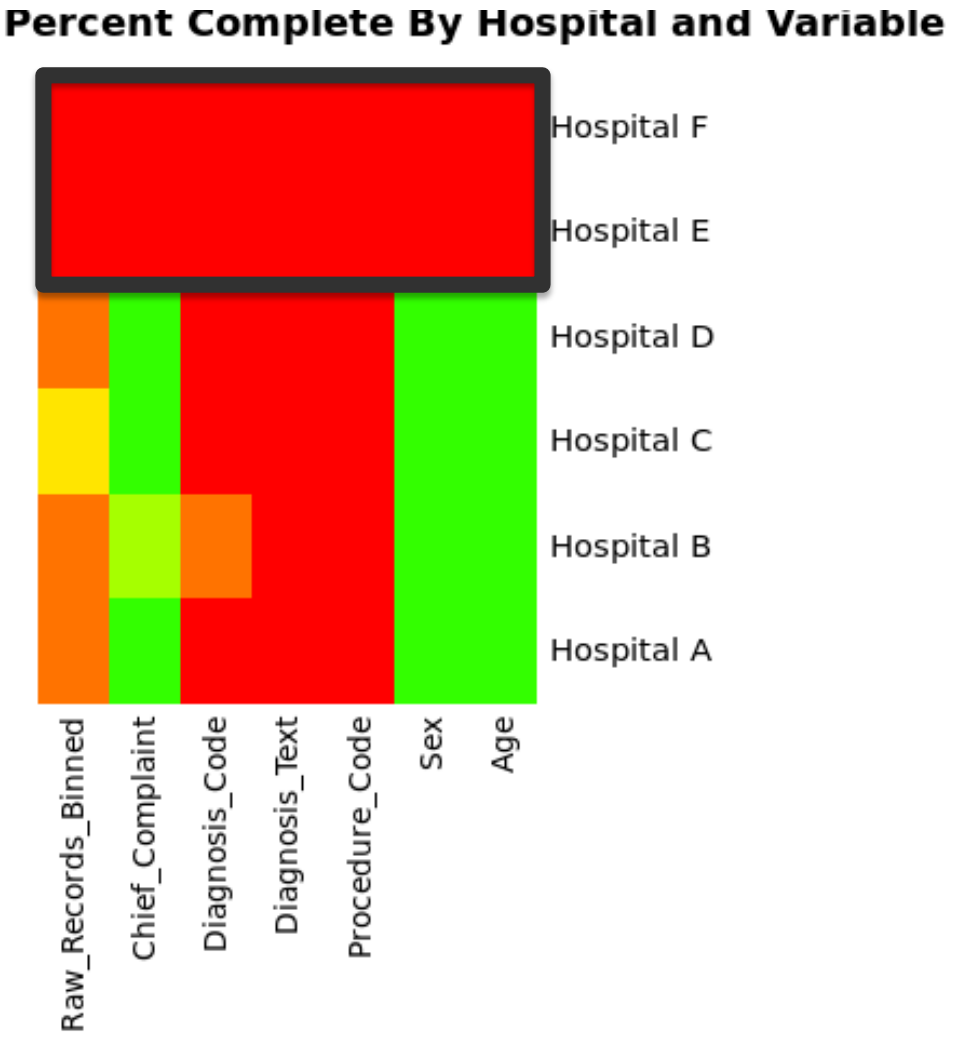
Heatmap: Hospital-stratified level



Red (Low)

Green (High)

Heatmap: Hospital-stratified level

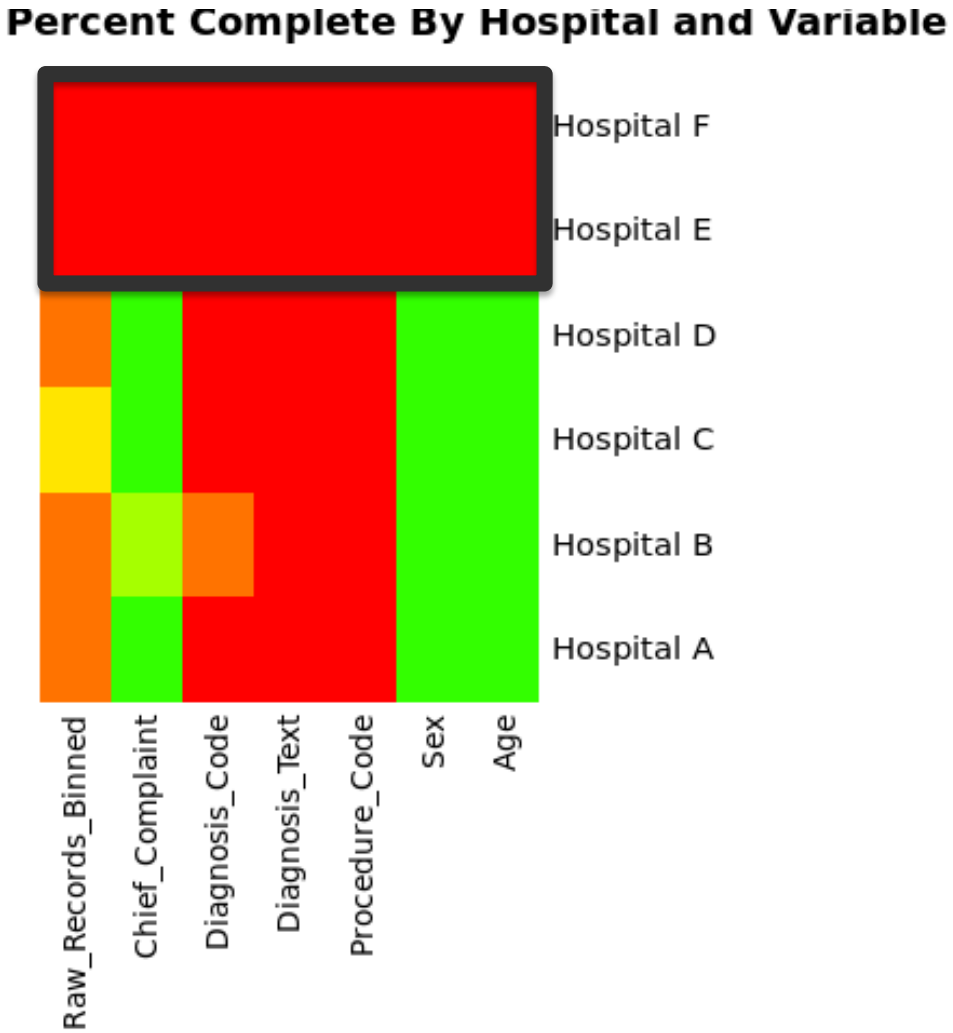


Red (Low)

Green (High)

Heatmap: Hospital-stratified level

These hospitals do NOT have raw records in the locker for specified time frame

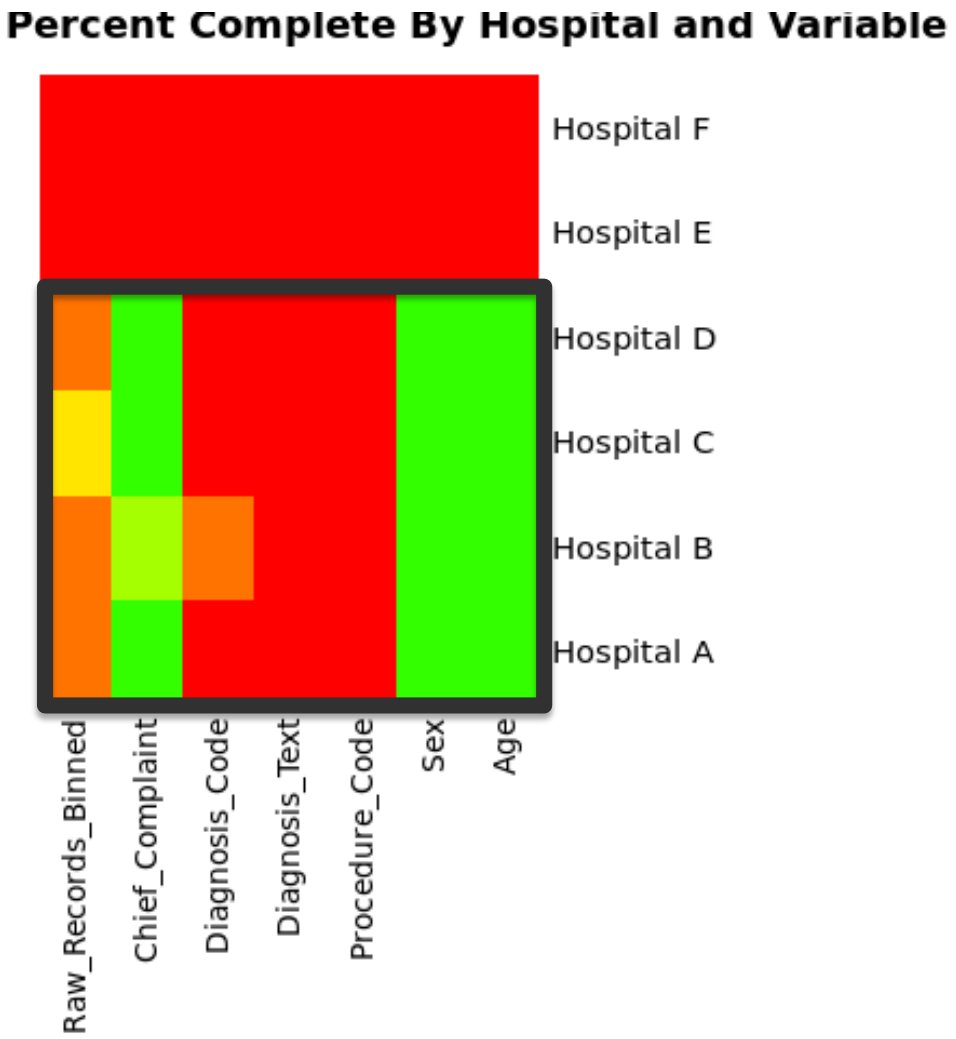


Red (Low)

Green (High)

Heatmap: Hospital-stratified level

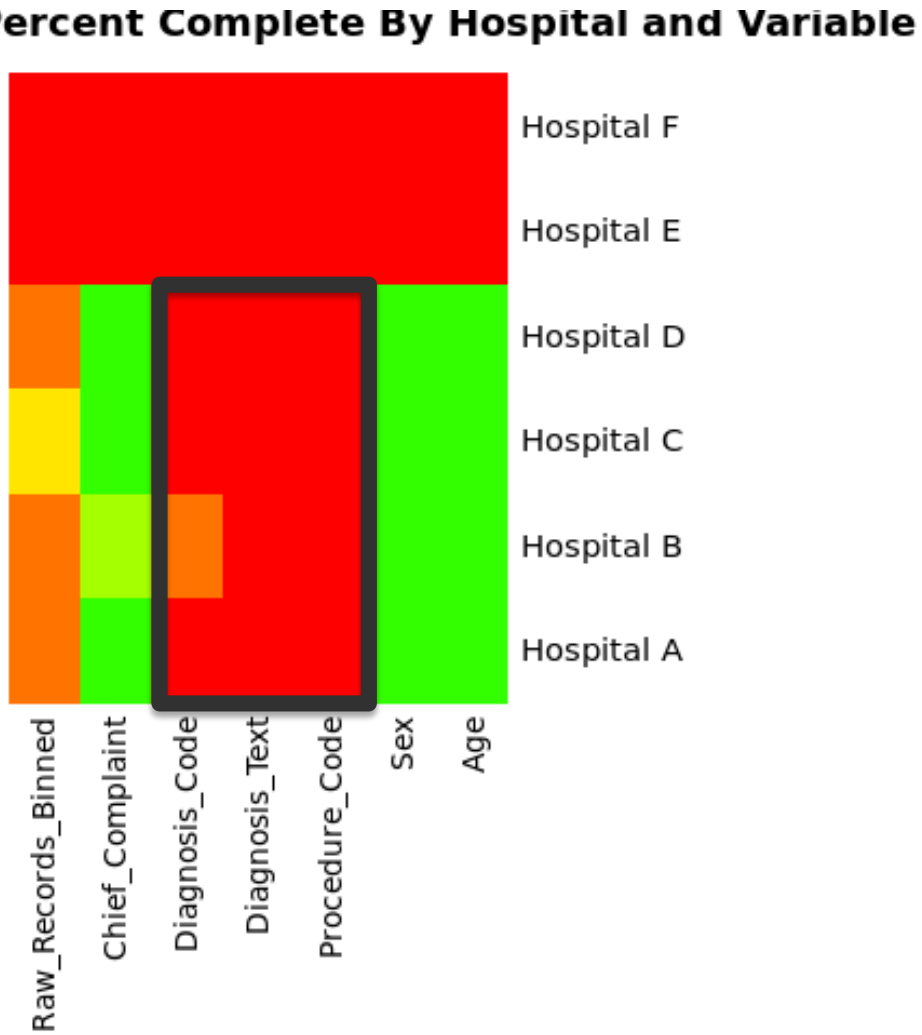
These hospitals have raw records in the locker for specified time frame



Red (Low)

Green (High)

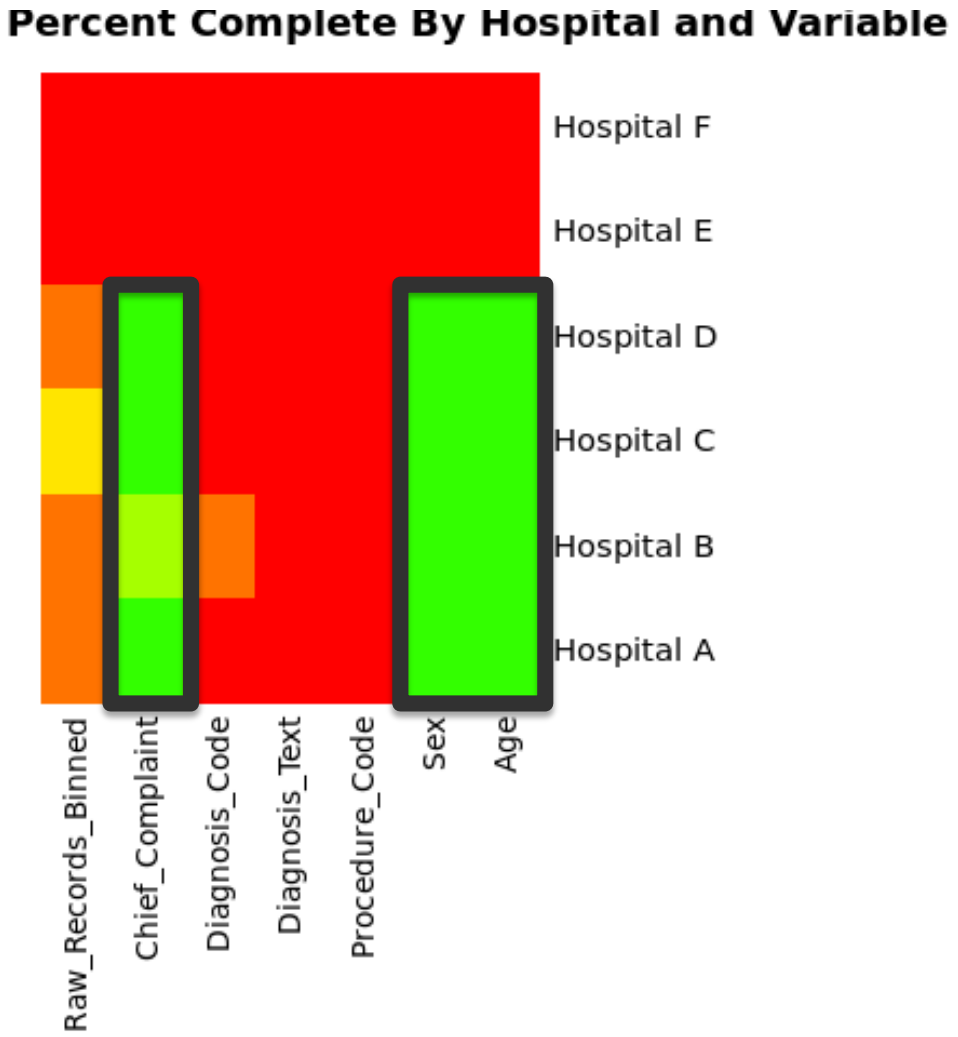
Heatmap: Hospital-stratified level



Red (Low)

Green (High)

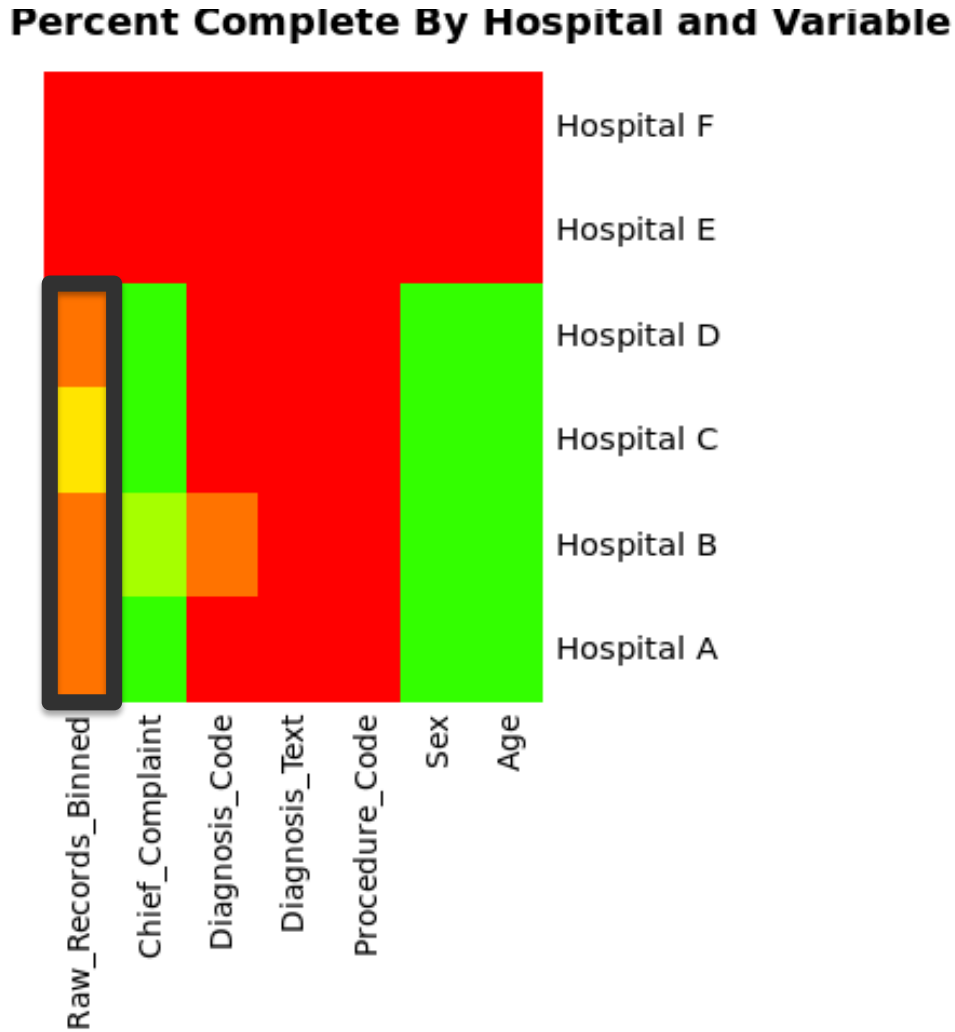
Heatmap: Hospital-stratified level



Red (Low)

Green (High)

Heatmap: Hospital-stratified level



Red (Low)

Green (High)

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Table: Hospital-stratified level

Stratified data quality metrics (at the facility level) are shown in the table below.

	Facility_Name	Num_Visits	Perc_Binned_Visits	Perc_Complete_CC	Perc_Complete_DC	Perc_Complete_DT	Perc_Complete_PC	Perc_Complete_Sex	Perc_Complete_Age	Last_Visit_Date
1	Hospital A	4782.00	38.58	90.13	6.21	0.00	0.00	100.00	99.98	2014-04-16
2	Hospital B	1494.00	22.69	71.49	20.95	0.00	0.00	100.00	100.00	2014-04-16
3	Hospital C	6663.00	40.61	85.98	17.38	0.00	0.00	100.00	100.00	2014-04-16
4	Hospital D	9390.00	24.80	87.93	16.99	0.00	0.00	99.99	99.99	2014-04-16
5	Hospital E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25
6	Hospital F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2013-10-25

Binning Algorithm

BioSense 2.0 Binning Algorithm

Introduction

This document is paired with a separate spreadsheet named BioSense Binning Instructions.xlsx. The flow of the BioSense 2.0 algorithm is described in this document. Details and specific instructions relating to parsing referred to in this document are found in the binning instructions spreadsheet.

Raw, HL7 encoded ADT records received by BioSense are Extracted, Translated and then Loaded (ETL) in the Jurisdictional Locker (or Meaningful Use Base [MUB] table). Periodically the binning algorithm looks at records that have not previously been binned and performs the binning of records into syndromes and sub-syndromes.

Binning

Binning is performed in distinct passes through the raw data that is housed in jurisdictional lockers.

First Pass

In the first pass data from the Diagnosis Code, Diagnosis text, Procedure Code, or Chief Complaint fields of the ADT record are viewed (see the "SEARCH TYPE" column in the binning instructions spreadsheet). The record is tested using, regular expressions, against all of the rules found in the binning instructions spreadsheet. All possible matches are recorded and temporarily stored.

- Chief Complaint
- Diagnosis Code
- Diagnosis Text
- Procedure Code

Binning Algorithm

BioSense 2.0 Binning Algorithm

Introduction

This document is paired with a separate spreadsheet named BioSense Binning Instructions.xlsx. The flow of the BioSense 2.0 algorithm is described in this document. Details and specific instructions relating to parsing referred to in this document are found in the binning instructions spreadsheet.

Raw, HL7 encoded ADT records received by BioSense are Extracted, Translated and then Loaded (ETL) in the Jurisdictional Locker (or Meaningful Use Base [MUB] table). Periodically the binning algorithm looks at records that have not previously been binned and performs the binning of records into syndromes and sub-syndromes.

Binning

Binning is performed in distinct passes through the raw data that is housed in jurisdictional lockers.

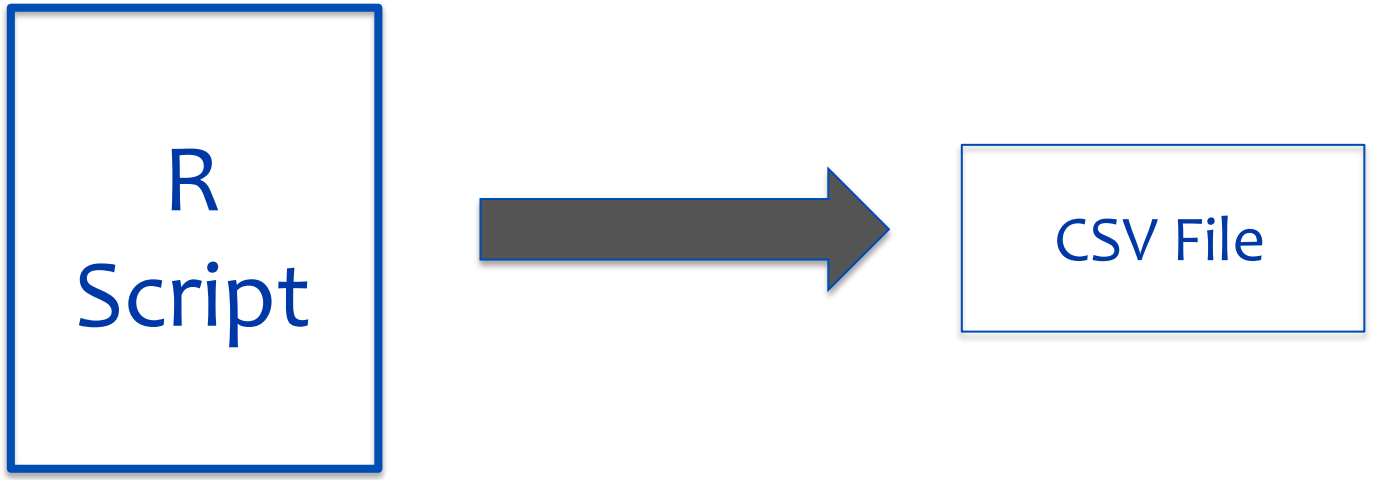
First Pass

In the first pass data from the Diagnosis Code, Diagnosis text, Procedure Code, or Chief Complaint fields of the ADT record are viewed (see the "SEARCH TYPE" column in the binning instructions spreadsheet). The record is tested using, regular expressions, against all of the rules found in the binning instructions spreadsheet. All possible matches are recorded and temporarily stored.

- Chief Complaint
- Diagnosis Code
- Diagnosis Text
- Procedure Code

Binvalue
(syndrome)

Binning Map Script



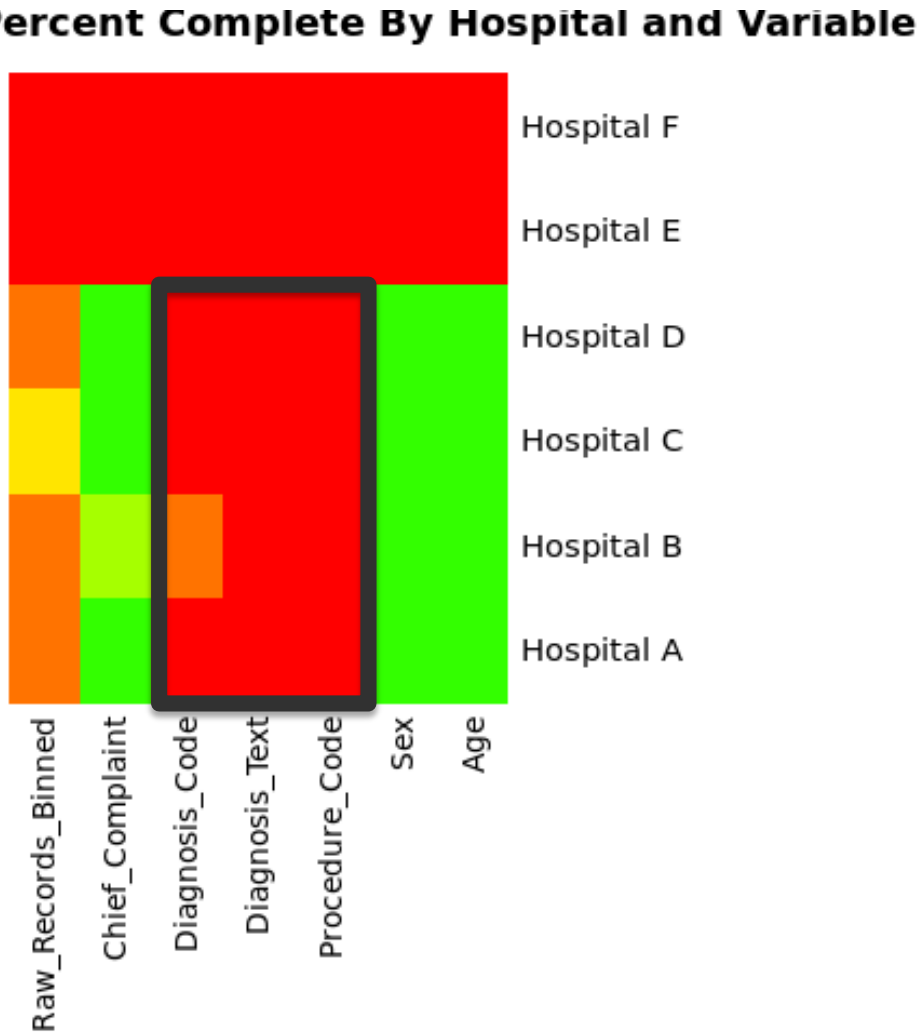
Binning Map

Chief_Complaint	Diagnosis_Code	Diagnosis_Text	Procedure_Code	binvalue
RLQ PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
PNEUMOINA	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
abd pain	NA	NA	NA	Abdominal pain
LLQ ABD PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
RIGHT LOWER QUADR/	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
VAG BLEEDING/ABD P/	NA	NA	NA	Abdominal pain
RIGHT FLANK PAIN,LOV	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ACUTE APPENDICITIS	NA	NA	NA	Abdominal pain
ABD/BACK PAIN	NA	NA	NA	Abdominal pain
SYMPTOMATIC CHOLEI	NA	NA	NA	Abdominal pain
ASCITES	NA	NA	NA	Abdominal pain
CHEST PAIN,RULE OUT	NA	NA	NA	Abdominal pain
DIABETIC KETOACIDOS	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain

Binning Map

Chief_Complaint	Diagnosis_Code	Diagnosis_Text	Procedure_Code	binvalue
RLQ PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
PNEUMOINA	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
abd pain	NA	NA	NA	Abdominal pain
LLQ ABD PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
RIGHT LOWER QUAD	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
VAG BLEEDING/ABD	NA	NA	NA	Abdominal pain
RIGHT FLANK PAIN,LC	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ACUTE APPENDICITIS	NA	NA	NA	Abdominal pain
ABD/BACK PAIN	NA	NA	NA	Abdominal pain
SYMPTOMATIC CHOL	NA	NA	NA	Abdominal pain
ASCITES	NA	NA	NA	Abdominal pain
CHEST PAIN,RULE OU	NA	NA	NA	Abdominal pain
DIABETIC KETOACIDO	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain

Heatmap: Hospital-stratified level



Red (Low)

Green (High)

Binning Map

Chief_Complaint	Diagnosis_Code	Diagnosis_Text	Procedure_Code	binvalue
RLQ PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
PNEUMOINA	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
abd pain	NA	NA	NA	Abdominal pain
LLQ ABD PAIN	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
RIGHT LOWER QUADR/	NA	NA	NA	Abdominal pain
ABD PAIN	NA	NA	NA	Abdominal pain
VAG BLEEDING/ABD P/	NA	NA	NA	Abdominal pain
RIGHT FLANK PAIN,LOV	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain
ACUTE APPENDICITIS	NA	NA	NA	Abdominal pain
ABD/BACK PAIN	NA	NA	NA	Abdominal pain
SYMPTOMATIC CHOLEI	NA	NA	NA	Abdominal pain
ASCITES	NA	NA	NA	Abdominal pain
CHEST PAIN,RULE OUT	NA	NA	NA	Abdominal pain
DIABETIC KETOACIDOS	NA	NA	NA	Abdominal pain
ABDOMINAL PAIN	NA	NA	NA	Abdominal pain

Syndrome Definitions Comparison

- **Purpose**
  - Rough performance assessment between BioSense v2.0 binning algorithm and County of San Diego (CoSD) SDADIC syndrome definitions
  - Question:
    - Do these different syndrome definitions result in similar syndrome classifications?

SDADIC: San Diego Aberration Detection and Incident Characterization; CoSD's custom syndromic surveillance system

Syndrome Definitions Comparison

- Number of records that mapped to a BioSense v2.0 bin:
  7,219 (32%)

- Number of records that did not map to a BioSense v2.0 bin:
  15,110 (68%)

- Number of unmapped records that would have mapped to a syndrome in CoSD's SDADIC:
  167 (1% of non-binned records)

Syndrome Definitions Comparison

- Number of BioSense v2.0 bins: 91
- Number of SDADIC syndromes: 22
- All SDADIC syndromes were found in BioSense v2.0 bins
- Many-to-one mapping from BioSense v2.0 bins to SDADIC syndromes

County of San Diego Syndromes vs. BioSense v2.0 Bins

SDADIC SYNDROMES	BIOSENSE 2.0 BIN NAMES
ABD PAIN	Abdominal pain
ALTERED NEURO	Paralysis, Speech Disturbance, Numbness
ASTHMA	Asthma, COPD
BLOODY DIARRHEA	Gastrointestinal Hemorrhage
CHEST PAIN	Chest Pain
COLD SYMPTOMS	Cough
DIARRHEA	Diarrhea
FEVER	Fever
GI	Gastrointestinal
HAZARDOUS-TOXIC	Poisoning by Medicine
HEADACHE	Headache
INFECTION	Skin infection, Septicemia/Bacteremia
INFLUENZA-LIKE-ILLNESS	Influenza-Like-Illness
NAUSEA	Nausea and Vomiting
NEUROTHER	Syncope and collapse
PSYCH	Mental Disorders
RASH	Rash
RESPIRATORY	Respiratory failure
RESPIRATORY W/BLOOD	Hemoptysis
SEIZURE	Convulsions
VOMITING	Nausea and Vomiting
WEAKNESS	Malaise and Fatigue

County of San Diego Syndromes vs. BioSense v2.0 Bins

SDADIC SYNDROMES	BIOSENSE 2.0 BIN NAMES
ABD PAIN	Abdominal pain
ALTERED NEURO	Paralysis, Speech Disturbance, Numbness
ASTHMA	Asthma, COPD
BLOODY DIARRHEA	Gastrointestinal Hemorrhage
CHEST PAIN	Chest Pain
COLD SYMPTOMS	Cough
DIARRHEA	Diarrhea
FEVER	Fever
GI	Gastrointestinal
HAZARDOUS-TOXIC	Poisoning by Medicine
HEADACHE	Headache
INFECTION	Skin infection, Septicemia/Bacteremia
INFLUENZA-LIKE-ILLNESS	Influenza-Like-Illness
NAUSEA	Nausea and Vomiting
NEUROTHER	Syncope and collapse
PSYCH	Mental Disorders
RASH	Rash
RESPIRATORY	Respiratory failure
RESPIRATORY W/BLOOD	Hemoptysis
SEIZURE	Convulsions
VOMITING	Nausea and Vomiting
WEAKNESS	Malaise and Fatigue

Syndrome Definitions Comparison

Conclusions

- High percentage of binned records binned accurately (using CoSD SDADIC syndrome definitions as a standard for comparison)

Syndrome Definitions Comparison

- **Steps for other jurisdictions**
  - Download Binning Map script [see later slide]
  - Run in RStudio Server (adm.biosen.se/rstudio)
  - Export CSV output file
  - Apply jurisdiction-specific syndrome definitions on CSV output file
  - Compare performance between jurisdiction-specific and BioSense v2.0 syndrome definitions

Future Plans

- **Finishing:**
  - Testing and validation of scripts
  - BioSense v2.0 Startup Guide
  - UCEP Web application

Download Scripts

- Data Quality Metrics:
  https://drive.google.com/file/d/0B9v3HiiWbFkpdU1tUjQ1UHdBazg/edit?usp=sharing

- Binning Map:
  https://drive.google.com/file/d/0B9v3HiiWbFkpc21PQVZXZjRzWG8/edit?usp=sharing

- Note: These scripts are still being tested and validated. They are not yet finalized nor guaranteed to be accurate.

Acknowledgements

- **Debugging and Validation**
  - Marcus Rennick
  - Caleb Wiedeman
- **Testing**
  - Robert Beum
  - Edward Castagna
  - Jenna Iberg Johnson
  - Danika Williams
- **Syndrome Definitions Comparison**
  - Brit Colanter

Acknowledgements

- This work was supported in part by an appointment to the Applied Public Health Informatics Fellowship Program administered by CSTE and funded by the Centers for Disease Control and Prevention (CDC) Cooperative Agreement 3U38HM000414-04W1.

Questions?

- Harold.Gil@sdcounty.ca.gov or hgil@med.miami.edu