



ScHARe

ScHARe Repository Introduction

November 20, 2024

Deborah Duran, PhD · NIMHD

Elif Dede Yildirim, PhD · NIMHD

Mark Aronson, PhD · NIMHD



ScHARe

Data Aggregation



BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

Downloading data to your own computer

Use the item's menu to download the original file

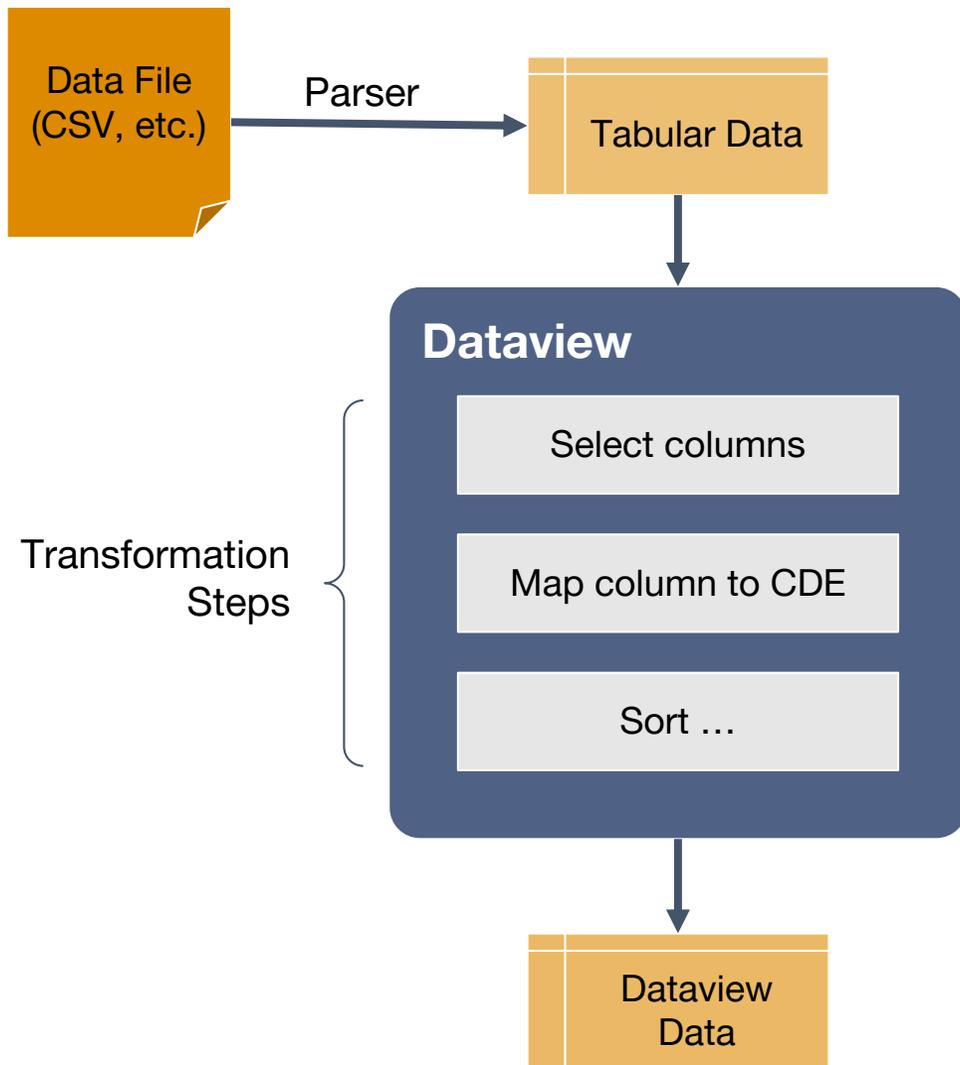
- Best for working directly with raw data

You can also export and download the tabular data in CSV, TSV, or Parquet format

- Best for working with mapped data, summary data, and other processed data

The screenshot shows a web interface for a data collection named "demo-files". The breadcrumb path is "scharedemo / demo-files / LIVE". The interface includes a header with navigation icons and an "Operations" dropdown. The main content area displays the "demo-files" abstract, which includes the text "November 20 Think-a-Thon Demonstration Files" and two columns: "Levels of Influence" (Individual) and "Domains of Influence" (Sociocultural Environment). Below the abstract are sections for "Links and Documents", "Data Items", and "Metadata and Statistics". The "Data Items" section contains a table with two rows: "example-data.csv" (25 KB, created an hour ago) and "example-dataview". A context menu is open over the "example-data.csv" row, listing actions such as "Rename...", "Copy To...", "Link To...", "Move To Folder...", "Configure Table", "Assign Data Dictionary...", "Create Dataview...", "Export Table Data..." (highlighted in blue), "Download", and "Delete". To the right of the main content is a sidebar with "Access Level" (Confidential), "Analysis Readiness" (Ready), "CDE Compliance - SCHARE" (9 / 19 CDEs assigned), and "Tags" (Cross-Sectional Study, State-Level Data). A lightbulb icon is visible in the bottom right corner of the interface.

Using Dataviews to Aggregate Data



Dataviews take data from one or more sources, apply a series of transformation steps to that data (*filtering, sorting, mapping, etc.*) resulting in a new table of data as output.

Uses of Dataviews:

- Creating subsets of data
- Hiding PHI/PII for publishing
- Summarizing individual-level data into subsets and estimates
- **Joining multiple datasets together**
- Mapping to CDEs
- ... many others!



Browsing and Searching for Data

Collections Create New Collection

My Collections

- demo-files** LIVE
November 20 Think-a-Thon Demonstration Files
18 hours ago

Shared with Me

You don't have any collections shared with you right now!

Public Collections

grouped by: **Domains of Influence**

Biological

- Example NMHSS Analysis** LIVE
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.
7 months ago

Health Care Systems and Clinical Care

- Example NMHSS Analysis** LIVE
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.
7 months ago
- Minority Health SVI** LIVE
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health...
8 months ago

Sociocultural Environment

- Minority Health SVI** LIVE
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health...
8 months ago

Repository About Docs Collections CDEs scharedemo

Search

Filters

- Minority Health SVI**
karl • 8 months ago
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.
- Example NMHSS Analysis**
karl • 7 months ago
This contains data from the 2018 National Mental Health Services Survey (N-MHSS) and links to Minority Health SVI data, also from 2018.

SciARc Repository - f1.2.1.d20241119 - bf.2.2.dev0-r-g4e4313bc.d20241119

Look for improvements to the data organization and search experience in 2025!



Linking and Aggregating Data

If you find public data that you would like to combine with your project data, you can link it into your own collection and use it just like a file that you had uploaded yourself.

▼ Data Items

STATUS	NAME	CREATED	SIZE
📄	MH SVI Overview_11.19.2021.pdf	7 months ago	193 KB
📄	MH SVI Fact Sheet_7.15.2021.pdf	7 months ago	276 KB
📄	mh_svi_county_2018.csv	7 months ago	7.5 MB
📄	MinorityHealthSVI_DataDictionary_2018.csv	7 months ago	25 KB
📄	mh_svi_county-ScHARe	7 months ago	4 KB
📄	mh-svi-by-state-2018		
📄	MinorityHealthSVI_DataDictionary_2018.pdd.js...		

Copy To...
Link To...
Export Table Data...
Download

> Metadata and Statistics

Link Item To...

My Collections

- ✓ ScHARe Example Data 2

Public Collections

- test unlisted data
- Example NMHSS Analysis
- Minority Health SVI

Link Cancel

ScHARe Example Data 2

Abstract
For purposes of demonstration, this project collected data according to ...

Levels of Influence
Individual
Community

Domains of Influence
Biological
Behavioral

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE
📄	test_data.xlsx	3 hours ago	14 KB
📄	mapped test data	an hour ago	1 KB
📄	mh_svi_county-ScHARe	a few seconds ago	4 KB

Page 1 of 1



Linking and Aggregating Data

If you find public data that you would like to combine with your project data, you can link it into your own collection and use it just like a file that you had uploaded yourself.

Minority Health SVI

Abstract
The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence
Community
Societal

Domains of Influence
Health Care Systems and Clinical Care
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	MH SVI Overview_11.19.2021.pdf	8 months ago	193 KB	⋮
	MH SVI Fact Sheet_7.15.2021.pdf	8 months ago	276 KB	⋮
	mh_svi_county_2018.csv	8 months ago	7.5 MB	⋮
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB	⋮
	mh_svi_county-SchARE	8 months ago	4 KB	⋮
	mh-svi-by-state-2018	7 months ago	2 KB	⋮
	MinorityHealthSVI_DataDictionary_2018.pdd.json	5 months ago	87 KB	⋮

<< Page 1 of 1 >>

> Metadata and Statistics

Access Level ⓘ
 Public

Analysis Readiness
 Ready >

CDE Compliance - SchARE
 1 / 19 CDEs assigned

Tags
Topics tagged in this collection



Linking Data

- My Collections >
- Starred >

Minority Health SVI

Abstract

The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence

Community
Societal

Domains of Influence

Health Care Systems and Clinical Care
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	MH SVI Overview_11.19.2021.pdf	8 months ago	193 KB	⋮
	MH SVI Fact Sheet_7.15.2021.pdf	8 months ago	276 KB	⋮
	mh_svi_county_2018.csv	8 months ago	7.5 MB	⋮
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB	⋮
	mh_svi_county-SchARE	8 months ago	4 KB	⋮
	mh-svi-by-state-2018			
	MinorityHealthSVI_DataDictionary_2018.pdd.json			

- Copy To...
- Link To...
- Export Table Data...
- Download

> Metadata and Statistics

Access Level ⓘ

Public

Analysis Readiness

Ready >

CDE Compliance - SchARE

1 / 19 CDEs assigned

Tags

Topics tagged in this collection

- Minority Health
- Household Income
- Household size
- Race
- Health Care Delivery



Linking Data

- Recent >
- My Collections >
- Starred >

Minority Health SVI

Abstract

The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (SVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic.

Levels of Influence
Community
Societal

Domains of Influence
Health Care
Sociocultural

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE
	MH SVI Overview_11.19.2021.pdf	months ago	193 KB
	MH SVI Fact Sheet_7.15.2021.pdf	months ago	276 KB
	mh_svi_county_2018.csv	8 months ago	7.5 MB
	MinorityHealthSVI_DataDictionary_2018.csv	8 months ago	25 KB
	mh_svi_county-SchARE	8 months ago	4 KB
	mh-svi-by-state-2018	7 months ago	2 KB
	MinorityHealthSVI_DataDictionary_2018.pdd.json	5 months ago	87 KB

> Metadata and Statistics

Link Item To...

- My Collections
 - ✓ demo-files
- Public Collections
 - ScHARe Think-a-Thon - Example Data
 - test unlisted data
 - Example NMHSS Analysis
 - Minority Health SVI

Link Cancel

Access Level ⓘ

Public

Analysis Readiness

✓ Ready >

CDE Compliance - ScHARe

1 / 19 CDEs assigned

Tags

Topics tagged in this collection

- Minority Health
- Household Income
- Household size
- Race
- Health Care Delivery



Linking Data

- Recent >
- My Collections >
- Starred >

scharedemo / demo-files / LIVE

Operations ▾

demo-files

Abstract
November 20 Think-a-Thon Demonstration Files

Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	example-data.csv	8 hours ago	25 KB	⋮
	example-dataview	7 hours ago	2 KB	⋮
	mh_svi_county-ScHARe	a few seconds ago	4 KB	⋮

<< Page 1 of 1 >

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ

Confidential

Analysis Readiness

Ready >

CDE Compliance - ScHARe ⓘ

9 / 19 CDEs assigned

Tags

Topics tagged in this collection

[Cross-Sectional Study](#) [State-Level Data](#)



Aggregating Data

scharedemo / demo-files / LIVE

demo-files

Abstract
November 20 Think-a-Thon Demonstration Files

Access Level ⓘ
Confidential

Recent >
My Collections >
Starred >

Operations ▾

Once linked into your collection, you can create a dataview that combines the public data with your own project data.

▼ Data Items

STATUS	NAME	CREATED	SIZE
EP	example-data.csv	8 hours ago	25 KB
EP	example-dataview	7 hours ago	2 KB
EP	mh_svi_county-SchARe		

Drag and Drop or [Browse Files](#) to Upload

- Rename...
- Copy To...
- Link To...
- Move To Folder...
- Configure Table
- Assign Data Dictionary...
- Create Dataview...
- Import as REDCap Data Dictionary

CDE Compliance - SchARe ⓘ
9 / 19 CDEs assigned

Tags
Topics tagged in this collection
Cross-Sectional Study State-Level Data

> Metadata and Statistics



Aggregating Data

My Collections >

Starred >

demo-files

Abstract

November 20 Think-a-Thon Demonstration Files

Levels of Influence
Individual

Domains of Influence
Sociocultural Environment

> Links and Documents

▼ Data Items

STATUS	NAME	CREATED	SIZE	
	example-data.csv	8 hours ago	25 KB	⋮
	example-dataview	7 hours ago	2 KB	⋮
	mh_svi_county-SchARe	minutes ago	4 KB	⋮

<< Page 1 of 1 >

Create Dataview

Drag and Drop or [Browse Files](#) to Upload

> Metadata and Statistics

Access Level ⓘ

Confidential

Analysis Readiness

Ready >

CDE Compliance - SchARe ⓘ

9 / 19 CDEs assigned

Tags

Topics tagged in this collection

[Cross-Sectional Study](#)

[State-Level Data](#)



Aggregating Data

- Recent >
- My Collections >
- Starred >

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status:

Item Operations

Source data from: [example-dataview](#)

Take

Number of rows Range of rows Limit output to Number of Rows

10

+ Add Step

Libraries:

Add Library

Add Data Elements

Clear Dataview

Save Dataview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Preview

age units	race ethnicity 1	race ethnicity 2	race ethnicity 3	race ethnicity 4	race ethnicity 5	race ethnicity 6	race ethnicity 7	zip code	sex at birth	gender	gender_select	
Years	no	no	no	yes	no	no	no	28752	Male	None of these describe me	Trans man / Trans woman / FTM	
Years	yes	no	no	no	no	no	no	10032	Male	None of these describe me	Trans man / Trans woman / FTM	
9784	18	Years	no	no	yes	no	no	yes	07501	Male	Trans	
5193	91	Years	no	no	no	no	yes	no	26726	None of these describe me	Tr	
8502	86	Years	yes	no	no	no	no	no	99737	Prefer not to answer	Prefer not to answer	

- Select Columns
- Filter Rows
- Sort
- Shuffle
- Join
- Take Rows
- Rename Column
- Map Column
- Aggregate Rows



Aggregating Data

☆ Starred >

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status: ⌵ Item Operations ▾

Source data from: [example-dataview](#) ↗

Take

Number of rows Range of rows Limit output to Number of Rows

Join Select

Join Table	Dataview Column	Matching Column from Join Table
<input type="text" value="Select table..."/>	<input type="text" value="Select column..."/>	<input type="text" value="Select column..."/>
<small>* Join table selection is required</small>	<small>* Column selection is required</small>	<small>* Column selection is required</small>

+ Add Step ▾ Libraries: ⓘ Add Library Add Data Elements Clear Dataview Save Dataview



Transformation Incomplete

There are one or more errors in the transformation above. Please correct the errors to preview the transformation result.

Developer Data >



Aggregating Data

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status:

Item Operations

Source data from: [example-dataview](#)

Take

Number of rows Range of rows Limit output to Number of Rows

10

Join Select

Join Table

Dataview Column

Matching Column from Join Table

mh_svi_county-SchARe

zip_code

zip_code

+ Add Step

Libraries:

Add Library

Add Data Elements

Clear Dataview

Save Dataview

Results

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

Transformation Preview

record_id	age	age_units	race_ethnicity_1	race_ethnicity_2	race_ethnicity_3	race_ethnicity_4	race_ethnicity_5	race_ethnicity_6	race_ethnicity_7	zip_code	sex_at_birth	gender
1943	92	Years	no	no	no	yes	no	no	no	28752	Male	N tr de me



Aggregating Data

- Recent >
- My Collections >
- Starred >

- Select Columns
- Filter Rows
- Sort
- Shuffle
- Join
- Take Rows
- Rename Column

scharedemo / demo-files / LIVE / aggregated_data

Advanced Explorer Table Dictionary Meta 575 bytes | a few seconds ago | text/prql | status:

Item Operations

Source data from: [example-dataview](#)

Take

Number of rows Range of rows Limit output to Number of Rows

10

Join Select

Join Table

mh_svi_county-SCHARe

Dataview Column

zip_code

Matching Column from Join Table

zip_code

+ Add Step

Libraries:

Add Library

Add Data Elements

Clear Dataview

Save Dataview

✓ Data available

✓ 0 parsing errors

✓ 0 validation errors

on Preview

age_units	race_ethnicity 1	race_ethnicity 2	race_ethnicity 3	race_ethnicity 4	race_ethnicity 5	race_ethnicity 6	race_ethnicity 7	zip_code	sex at birth	gender
Years	no	no	no	yes	no	no	no	28752	Male	N fr de me



Selecting Variables

mh_svi_county-SchARE zip_code zip_code

Select

Available Columns

- ST_ABBR
- FIPS
- M_TOTPOP
- E_HU
- M_HU
- E_HH

Selected Columns

- zip_code
- gender_mapped
- sex_at_birth
- STATE
- COUNTY
- LOCATION
- E_TOTPOP

Add All >> << Remove All

+ Add Step Libraries: Add Library Add Data Elements Clear Dataview Save Dataview

Results > ✓ Data available ✓ 0 parsing errors ✓ 0 validation errors

Transformation Preview

record id	age	zip_code	gender mapped	sex at birth	STATE	COUNTY	LOCATION	E_TOTPOP
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Avery	Avery County, North Carolina	
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Rutherford	Rutherford County, North Carolina	
1210	72	52223	Man	Female	IOWA	Delaware	Delaware County, Iowa	



Selecting Variables

Buttons: Add All >> << Remove All

Libraries: Add Library Add Data Elements

Clear Dataview Save Dataview

Results > ✓ Data available ✓ 0 parsing errors ✓ 0 validation errors

Transformation Preview

record_id	age	zip_code	gender_mapped	sex_at_birth	STATE	COUNTY	LOCATION	E_TOTPOP
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Avery	Avery County, North Carolina	17501
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	Rutherford	Rutherford County, North Carolina	66532
1210	72	52223	Man	Female	IOWA	Delaware	Delaware County, Iowa	17258
1943	92	28752	None of these describe me. I would like to consider additional options	Male	NORTH CAROLINA	McDowell	McDowell County, North Carolina	45109
8502	86	99737	Prefer not to answer	Prefer not to answer	ALASKA	Southeast Fairbanks	Southeast Fairbanks Census Area, Alaska	6876
9784	18	07501	Transgender	Male	NEW JERSEY	Passaic	Passaic County, New Jersey	504041
2618	80	35405	Prefer not to answer	Male	ALABAMA	Tuscaloosa	Tuscaloosa County, Alabama	206213
6809	58	11769	Transgender	None of these describe me	NEW YORK	Suffolk	Suffolk County, New York	1487901
6688	76	10032	None of these describe me. I would like to consider additional options	Male	NEW YORK	New York	New York County, New York	1632480
5193	91	26726	Transgender	None of these describe me	WEST VIRGINIA	Grant	Grant County, West Virginia	11641
3563	79	53186	Transgender	None of these describe me	WISCONSIN	Waukesha	Waukesha County, Wisconsin	398879
5193	91	26726	Transgender	None of these describe me	WEST VIRGINIA	Mineral	Mineral County, West Virginia	27278
1172	46	37208	Woman	Prefer not to answer	TENNESSEE	Davidson	Davidson County, Tennessee	6

rows per page: 25



ScHARe

Data Analysis

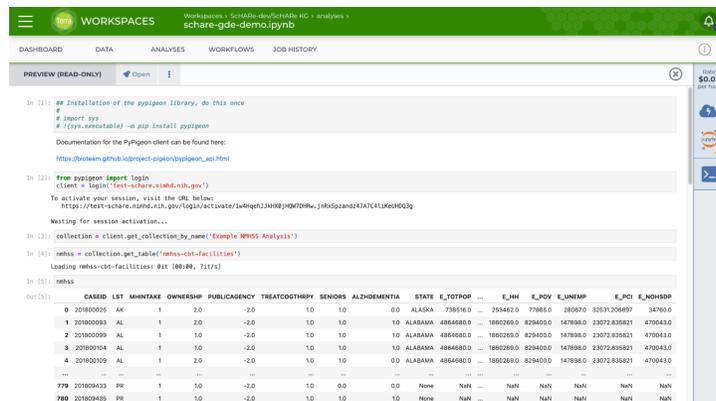


BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION

How do I analyze or aggregate data from the ScHARe Repository?

Three quick options:

Connect to a Terra Jupyter notebook



```
In [11]: # Installation of the pyPigeon library, do this once
import sys
! (sys.executable) -m pip install pyPigeon
Documentation for the PyPigeon client can be found here:
https://bioteam.github.io/project-pigeon/pyPigeon_api.html

In [12]: from pyPigeon import login
client = login('test-schare-nimbus.nih.gov')
To activate your session, visit the URL below:
https://test-schare-nimbus.nih.gov/login/activate/14f1qht3JH9WJHW70Rw_jshSpzand247AC1Lk6dR03g
Waiting for session activation...

In [13]: collection = client.get_collection_by_name('Example NHIS Analysis')

In [14]: nhis = collection.get_table('nhis-cst-facilities')
Loading nhis-cst-facilities: #11 (89188, 717s)

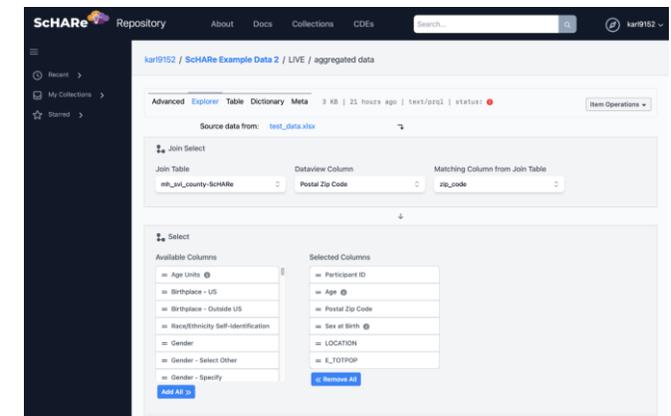
In [15]: nhis
Out[15]:
```

CASEID	LIST	MINITAKE	OWNERSHIP	PUBLICAGENCY	THREATCOSTHPY	SENIORS	ALZHEMIA	STATE	E_TOTPOP	E_HH	E_POV	E_UNEMP	E_PCI	E_HHSDP	
0	20180003	AL	1	2.0	-2.0	1.0	1.0	0.0	ALABAMA	788162.0	253482.0	77886.0	39050.0	3393106997	34765.0
1	20180009	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	1860269.0	829400.0	147890.0	23072835821	470043.0
2	20180010	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	1860269.0	829400.0	147890.0	23072835821	470043.0
3	201800104	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	1860269.0	829400.0	147890.0	23072835821	470043.0
4	201800109	AL	1	2.0	-2.0	1.0	1.0	0.0	ALABAMA	4864680.0	1860269.0	829400.0	147890.0	23072835821	470043.0
...
779	201809433	PR	1	1.0	-2.0	1.0	0.0	None	NAN	NAN	NAN	NAN	NAN	NAN	NAN
780	201809435	PR	1	1.0	-2.0	1.0	1.0	None	NAN	NAN	NAN	NAN	NAN	NAN	NAN

Download to your own computer



Use Dataviews on the Repository



Look for a future Think-a-Thon where we will share more about how to link your data to Terra and how to use the ScHARe Repository to create aggregated data sets!



Analyzing Data on Terra (Jupyter Notebooks)

The screenshot displays the Terra workspace interface for 'schare-gde-demo.ipynb'. The notebook contains several code cells:

- In [1]:** A code cell for installing the `pypigeon` library. The output includes documentation for the PyPigeon client and a link to the GitHub project page: https://bioteam.github.io/project-pigeon/pypigeon_api.html.
- In [2]:** A code cell for logging into the system. The output shows the activation URL: <https://test-schare.nimhd.nih.gov/login/activate/1w4HqehJJkHX0jHQW7DHRw.jnRx5pzandz47A7C4LiKeUHDQ3g> and the message 'Waiting for session activation...'. The right sidebar shows a rate of \$0.02 per hour.
- In [3]:** A code cell to retrieve a collection by name: `collection = client.get_collection_by_name('Example NMHSS Analysis')`.
- In [4]:** A code cell to retrieve a table: `nmhss = collection.get_table('nmhss-cbt-facilities')`. The output indicates 'Loading nmhss-cbt-facilities: 0it [00:00, ?it/s]'.
- In [5]:** A code cell to display the `nmhss` data.

Out [5]: A table with 17 columns: `CASEID`, `LST`, `MHINTAKE`, `OWNERSHP`, `PUBLICAGENCY`, `TREATCOGTHRPY`, `SENIORS`, `ALZHDementia`, `STATE`, `E_TOTPOP`, `E_HH`, `E_POV`, `E_UNEMP`, `E_PCI`, and `E_NOHSDP`. The table contains 780 rows of data, with the first few rows shown below:

	CASEID	LST	MHINTAKE	OWNERSHP	PUBLICAGENCY	TREATCOGTHRPY	SENIORS	ALZHDementia	STATE	E_TOTPOP	...	E_HH	E_POV	E_UNEMP	E_PCI	E_NOHSDP
0	201800025	AK	1	2.0	-2.0	1.0	1.0	0.0	ALASKA	738516.0	...	253462.0	77865.0	28067.0	32531.206897	34760.0
1	201800093	AL	1	2.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
2	201800099	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
3	201800104	AL	1	1.0	-2.0	1.0	1.0	1.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
4	201800109	AL	1	2.0	-2.0	1.0	1.0	0.0	ALABAMA	4864680.0	...	1860269.0	829400.0	147898.0	23072.835821	470043.0
...
779	201809433	PR	1	1.0	-2.0	1.0	0.0	0.0	None	NaN	...	NaN	NaN	NaN	NaN	NaN
780	201809435	PR	1	1.0	-2.0	1.0	1.0	1.0	None	NaN	...	NaN	NaN	NaN	NaN	NaN



Analyzing Data on Terra (Jupyter Notebooks)

Workspaces > ScHARe/ScHARe > Analyses

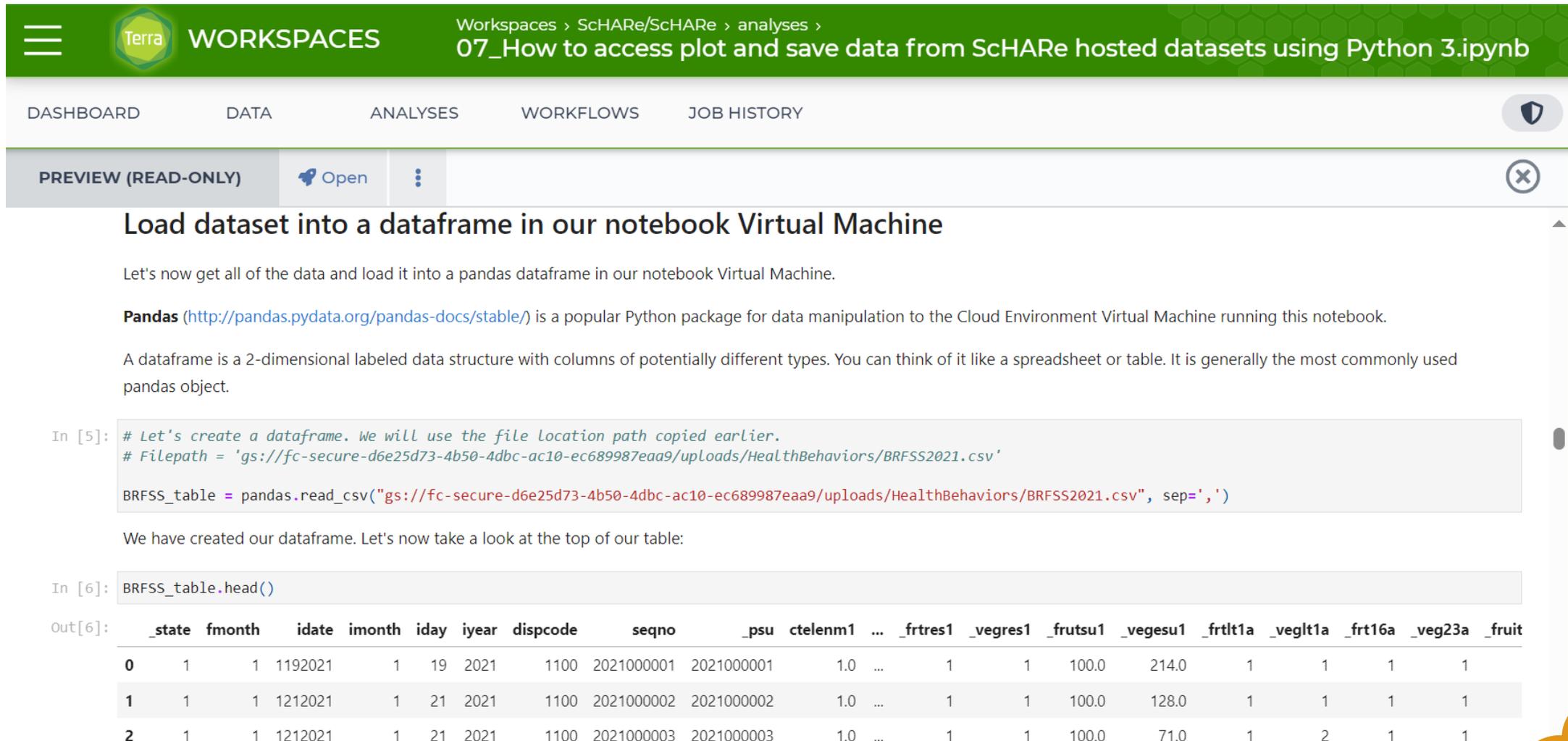
DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

Your Analyses [+ Start](#)

Application	Name ↓
Jupyter	00_List of Datasets Available on ScHARe.ipynb
Jupyter	01_Introduction to Terra Cloud Environment.ipynb
Jupyter	02_Introduction to Terra Jupyter Notebooks.ipynb
Jupyter	03_R Environment setup.ipynb



Analyzing Data on Terra (Jupyter Notebooks)



Terra WORKSPACES Workspaces > ScHARe/ScHARe > analyses > 07_How to access plot and save data from ScHARe hosted datasets using Python 3.ipynb

DASHBOARD DATA ANALYSES WORKFLOWS JOB HISTORY

PREVIEW (READ-ONLY) Open

Load dataset into a dataframe in our notebook Virtual Machine

Let's now get all of the data and load it into a pandas dataframe in our notebook Virtual Machine.

Pandas (<http://pandas.pydata.org/pandas-docs/stable/>) is a popular Python package for data manipulation to the Cloud Environment Virtual Machine running this notebook.

A dataframe is a 2-dimensional labeled data structure with columns of potentially different types. You can think of it like a spreadsheet or table. It is generally the most commonly used pandas object.

```
In [5]: # Let's create a dataframe. We will use the file location path copied earlier.
# Filepath = 'gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/HealthBehaviors/BRFSS2021.csv'

BRFSS_table = pandas.read_csv("gs://fc-secure-d6e25d73-4b50-4dbc-ac10-ec689987eaa9/uploads/HealthBehaviors/BRFSS2021.csv", sep=',')
```

We have created our dataframe. Let's now take a look at the top of our table:

```
In [6]: BRFSS_table.head()
```

```
Out[6]:
```

	_state	fmonth	idate	imonth	iday	iyear	dispcode	seqno	_psu	ctelenm1	...	_ftres1	_vegres1	_frutsu1	_vegesu1	_frlt1a	_veglt1a	_frt16a	_veg23a	_fruit
0	1	1	1192021	1	19	2021	1100	2021000001	2021000001	1.0	...	1	1	100.0	214.0	1	1	1	1	1
1	1	1	1212021	1	21	2021	1100	2021000002	2021000002	1.0	...	1	1	100.0	128.0	1	1	1	1	1
2	1	1	1212021	1	21	2021	1100	2021000003	2021000003	1.0	...	1	1	100.0	71.0	1	2	1	1	1



SCHARe

Conclusion

BE A PART OF THE FUTURE
OF KNOWLEDGE GENERATION



ScHARe Terra

In Summary:

- Register for the ScHARe/Terra Platform
- Access federated data – especially SDoH and population science
- Create your own workspace – can share with others
- Assess computational tools, such as python and R coding

ScHARe Repository

In Summary: (Comply with the NIH Data Management and Sharing Policy Requirement)

- When getting started, first create a collection, provide metadata, and upload documents.
- Upload your data and use the system to help you map to the ScHARe CDEs.
- View your data, see CDE compliance and analysis readiness.
- Data can be shared with your colleagues, and can be made publicly available after review.
- You can use dataviews to create subsets of your data, join datasets together and more.
- Data can also be analyzed on the ScHARe Terra workspace.



We are here to support you!

We want to hear your questions, issues and comments about the ScHARe Repository!

For any questions regarding how to use the Repository, please reach out to:

✉ schare@mail.nih.gov

Response time: within 24 hours

For Technical Support inquiries, contact:

✉ schare-repository-support@bioteam.net

🕒 Office Hours: Wednesdays, 4-5 PM US Eastern

Questions?



Slido Poll

What concerns or barriers do you have regarding sharing your data publicly?

Think-a-Thon poll

1. Rate how useful this session was:

- Very useful
- Useful
- Somewhat useful
- Not at all useful

Think-a-Thon poll

2. Rate the pace of the instruction for yourself:

- Too fast
- Adequate for me
- Too slow

Think-a-Thon poll

3. How likely will you participate in the next Think-a-Thon?

- Very interested, will definitely attend
- Interested, likely will attend
- Interested, but not available
- Not interested in attending any others

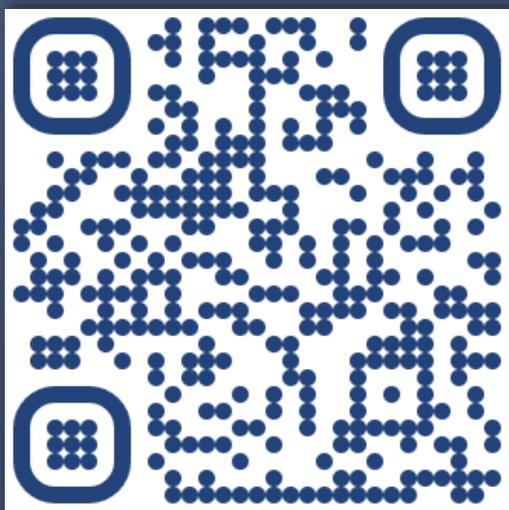
ScHARe

Thank you



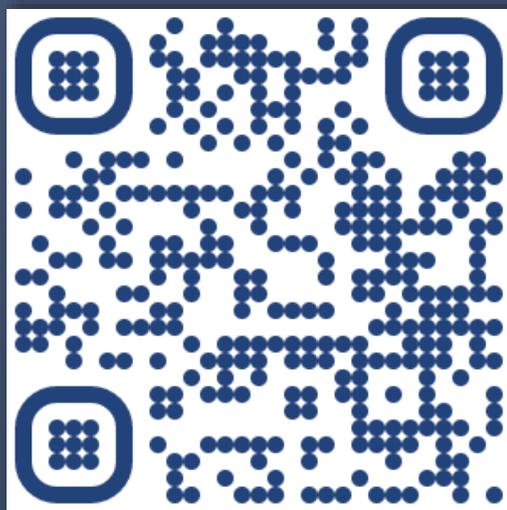
SchARE

Next Think-a-Thons:



bit.ly/think-a-thons

Register for SchARE:



bit.ly/join-schare



schare@mail.nih.gov

