

# allofus: An R package to facilitate use of the All of Us Research Workbench

Louisa Smith PhD, Robert Cavanaugh PhD

OHDSI Center, Roux Institute, Northeastern University, Portland, ME  
Bouvé College of Health Sciences, Northeastern University, Boston, MA

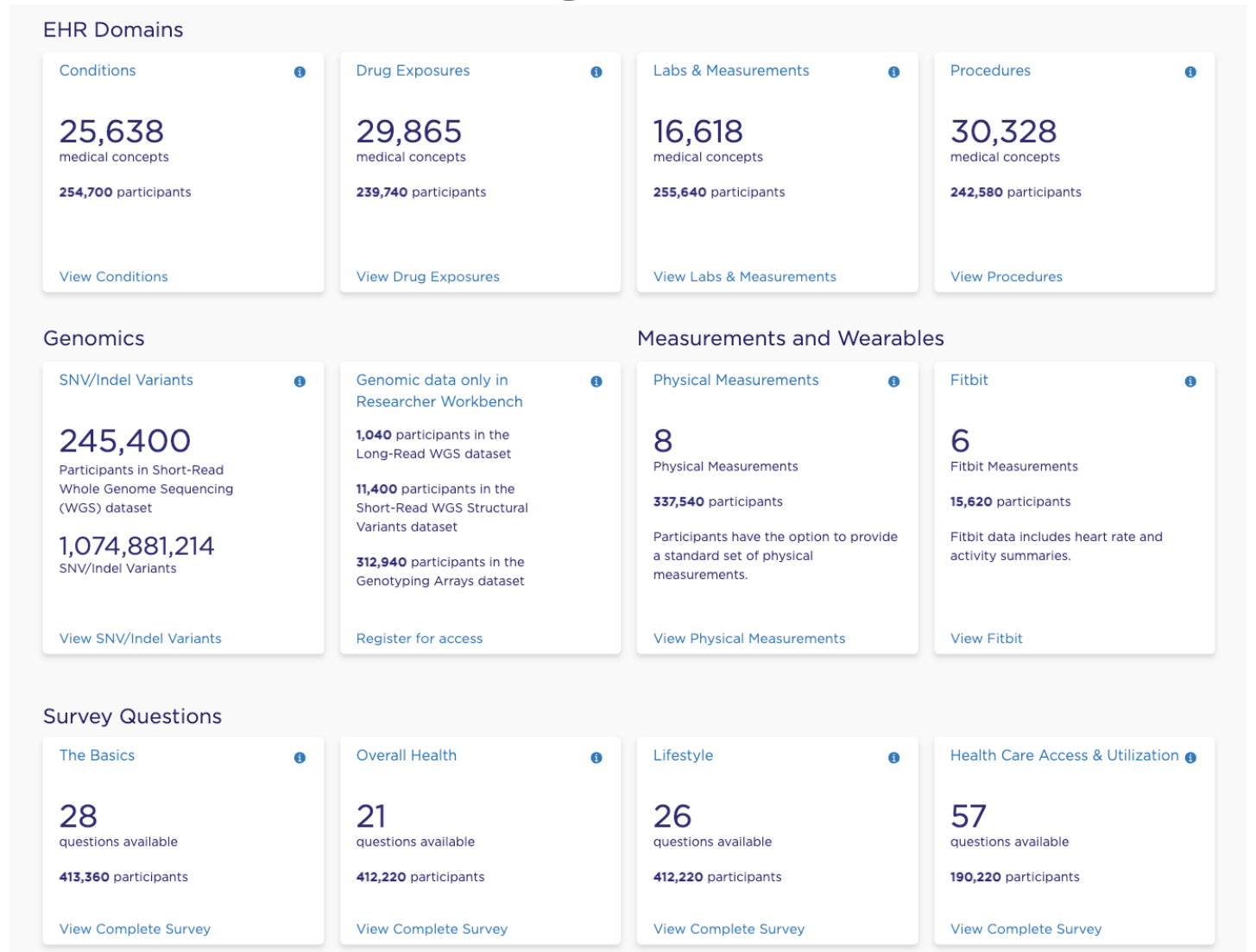
*The allofus R package was developed by Louisa Smith and Rob Cavanaugh at Northeastern University and is not affiliated with or endorsed by the All of Us Research Program*



**Northeastern University**  
**Observational Health Data**  
**Sciences and Informatics Center**



# Data in the *All of Us* research program



<https://www.nejm.org/doi/full/10.1056/NEJMSr1809937>

<https://databrowser.researchallofus.org/>

# Using the *All of Us* workbench

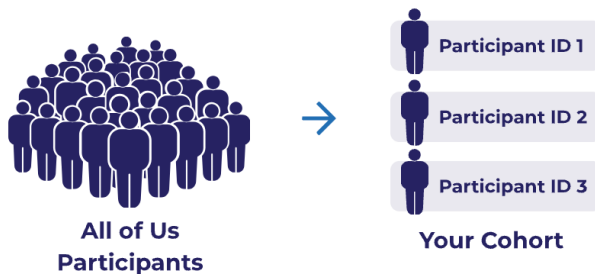
DATA

ANALYSIS

ABOUT

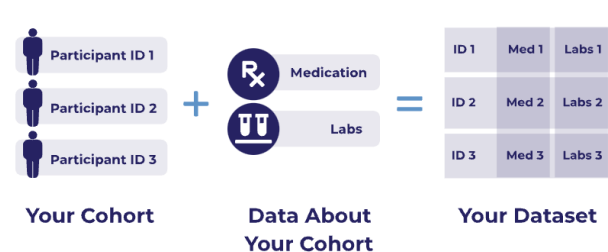
## Cohorts +

A cohort is a group of participants based on specific criteria.



## Datasets +

A dataset is a table containing data about a cohort that can be exported for analysis.



jupyter 01-testing (unsaved changes)

File Edit View Insert Cell Kernel Navigate Widgets Help Snippets Trusted Edit Mode R O

Code nbdiff

In [1]: library(allofus)

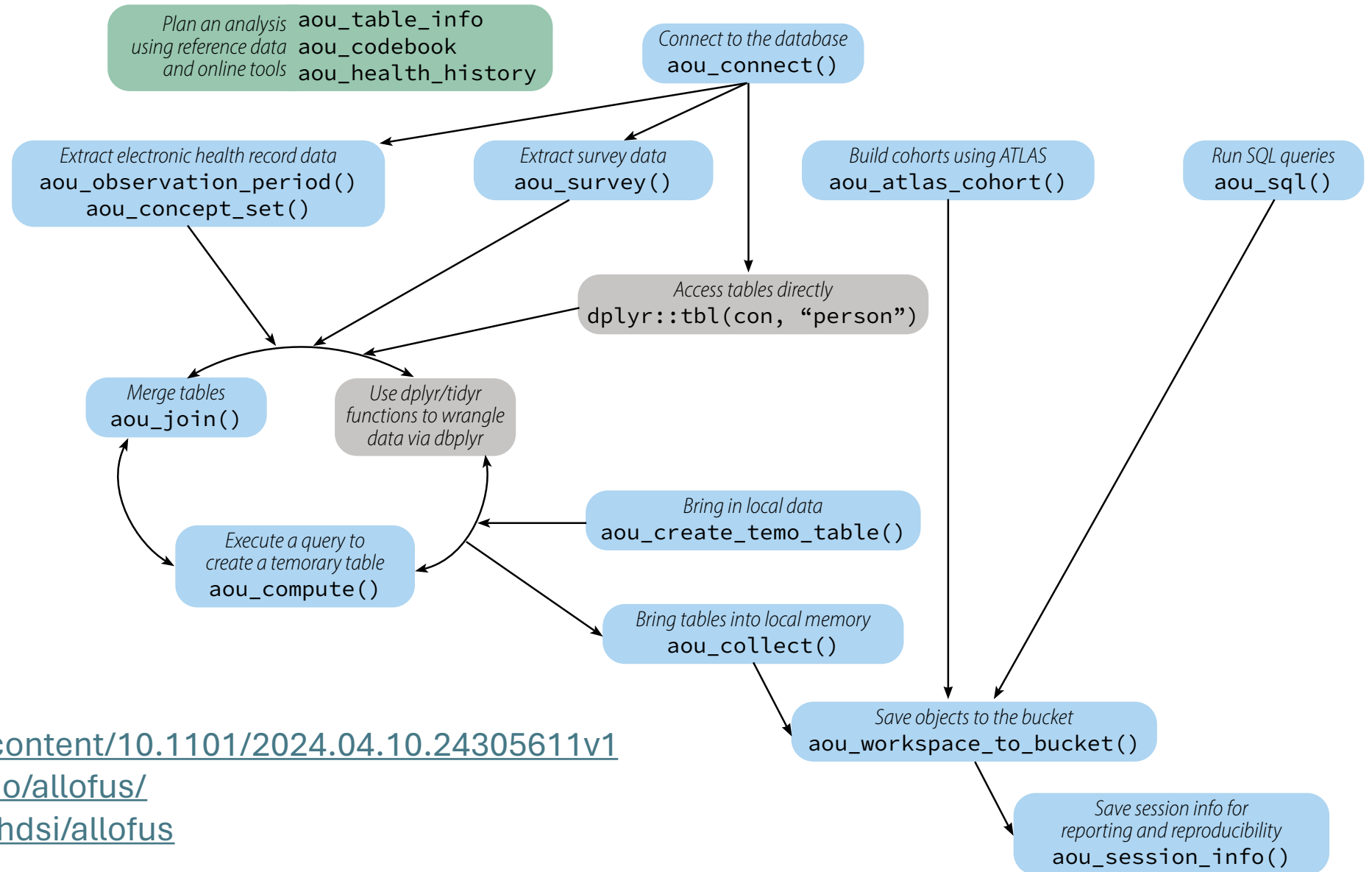
```
♥ Thank you for using the allofus R package! ♥
△ This package continues to be developed as All of Us grows and changes. Please report any issues to <https://github.com/roux-ohdsi/allofus/issues>.
i The allofus R package is not affiliated with or endorsed by the All of Us Research Program.
```

In [2]: con <- aou\_connect()
person\_table <- dplyr::tbl(con, "person")
dplyr::tally(person\_table)

```
✔ Connected successfully!
# Source:   SQL [1 x 1]
# Database: BigQueryConnection
           n
<int64>
1  410361
```



# {allofus} Package workflow



<https://www.medrxiv.org/content/10.1101/2024.04.10.24305611v1>

<https://roux-ohdsi.github.io/allofus/>

<https://github.com/roux-ohdsi/allofus>

# Integrating OHDSI Software & Standards

*Build cohorts using ATLAS*  
`aou_atlas_cohort()`

*Extract electronic health record data*  
`aou_observation_period()`  
`aou_concept_set()`

ATLAS

Home

Welcome to ATLAS.  
ATLAS is an open source application developed as a part of OHDSI intended to provide a unified interface to patient level data and analytics.

Documentation  
The ATLAS user guide can be found [here](#).

Getting Started

[Define a New Cohort](#) Begin performing research by defining the group of people you intend to study

[Search the Vocabulary](#) Search the different ontologies used to describe patient level data around the world

Release Notes

ATLAS Version 2.14.0 Release Notes  
WebAPI Version 2.14.0 DEV Release Notes

This latest release contains 24 feature enhancements and issue resolutions:

- Cannot pick up a concept from vocabulary for some cohort attributes
- Cannot find a concept by its id or code
- Admin cannot assign protected tag to entity
- JobServiceIT test fails
- Hydra v0.3 update
- Azure Synapse Analytics Dedicated dialect support
- Snowflake dialect support
- Refresh user names during scheduled user import
- PHOEBE 2.0 implementation for WebAPI
- Add description fields for all study asset types
- Support storage of map description fields

Apache 2.0  
open source software  
provided by  
OHDSI

OMOP Common Data Model

Background

Conventions

CDM Versions

CDM Additions

How to

Support

## Data Model Conventions

There are a number of implicit and explicit conventions that have been adopted in the CDM. Developers of methods that run against the CDM need to understand these conventions.

### General

The OMOP CDM is platform-independent. Data types are defined generically using ANSI SQL data types (VARCHAR, INTEGER, FLOAT, DATE, DATETIME, CLOB). Precision is provided only for VARCHAR. It reflects the minimal required string length and can be expanded within a CDM instantiation. The CDM does not prescribe the date and datetime format. Standard queries against CDM may vary for local instantiations and date/datetime configurations.

### Tables

# Using the survey data

concept_code	concept_id	concept_name	form_name	pdf
All	All	All	All	All
ahc_2	<a href="#">40192402</a>	Think abo...	Social Determinants of Health	<a href="#">Social Determinants of Health</a>
aian_ainspecific	<a href="#">1586150</a>	AIAN: AIA...	The Basics	<a href="#">The Basics</a>
aiannoneofthesedescribeme_aianfreetext	<a href="#">1585604</a>	AIAN None...	The Basics	<a href="#">The Basics</a>
alcohol_6ormoredrinksoccurence	<a href="#">1586213</a>	Alcohol: ...	Lifestyle	<a href="#">Lifestyle</a>
alcohol_alcoholparticipant	<a href="#">1586198</a>	Alcohol: ...	Lifestyle	<a href="#">Lifestyle</a>
alcohol_averagedailydrinkcount	<a href="#">1586207</a>	Alcohol: ...	Lifestyle	<a href="#">Lifestyle</a>
alcohol_drinkfrequencypastyear	<a href="#">1586201</a>	Alcohol: ...	Lifestyle	<a href="#">Lifestyle</a>
asian_asianspecific	<a href="#">1586151</a>	Asian: As...	The Basics	<a href="#">The Basics</a>
attemptquitsmoking_completelyquitage	<a href="#">1585870</a>	Attempt Q...	Lifestyle	<a href="#">Lifestyle</a>
audit_c_1	<a href="#">1332876</a>	In the pa...	COPE	<a href="#">COPE</a>

Extract survey data  
aou\_survey()

category	question
⊖ bone, joint, and muscle conditions	Including yourself, who in your family has had systemic lupus?
<b>relative</b>	<b>concept_code</b> <b>concept_id</b> <b>concept_id_rx_meds</b> <b>concept_id_on_txt</b>
self	skeletalmuscularcondition_systemiclupus_yes 1384389 43528881 43530396
daughter	daughterskeletalmuscularcondition_systemiclupus 43529888
son	sonskeletalmuscularcondition_systemiclupus 43529894
sibling	siblingskeletalmuscularcondition_systemiclupus 43529892
mother	motherskeletalmuscularcondition_systemiclupus 43529891
father	fatherskeletalmuscularcondition_systemiclupus 43529889
grandparent	grandparentskeletalmuscularcondition_systemiclupus 702784
⊕ bone, joint, and muscle conditions	Including yourself, who in your family has had pseudogout (CPPD)?
⊕ bone, joint, and muscle conditions	Including yourself, who in your family has had other arthritis?
⊕ bone, joint, and muscle conditions	Including yourself, who in your family has had other bone, joint, or muscle cor

# Functions for making life easier



*Connect to the database*  
`aou_connect()`

*Execute a query to  
create a temporary table*  
`aou_compute()`

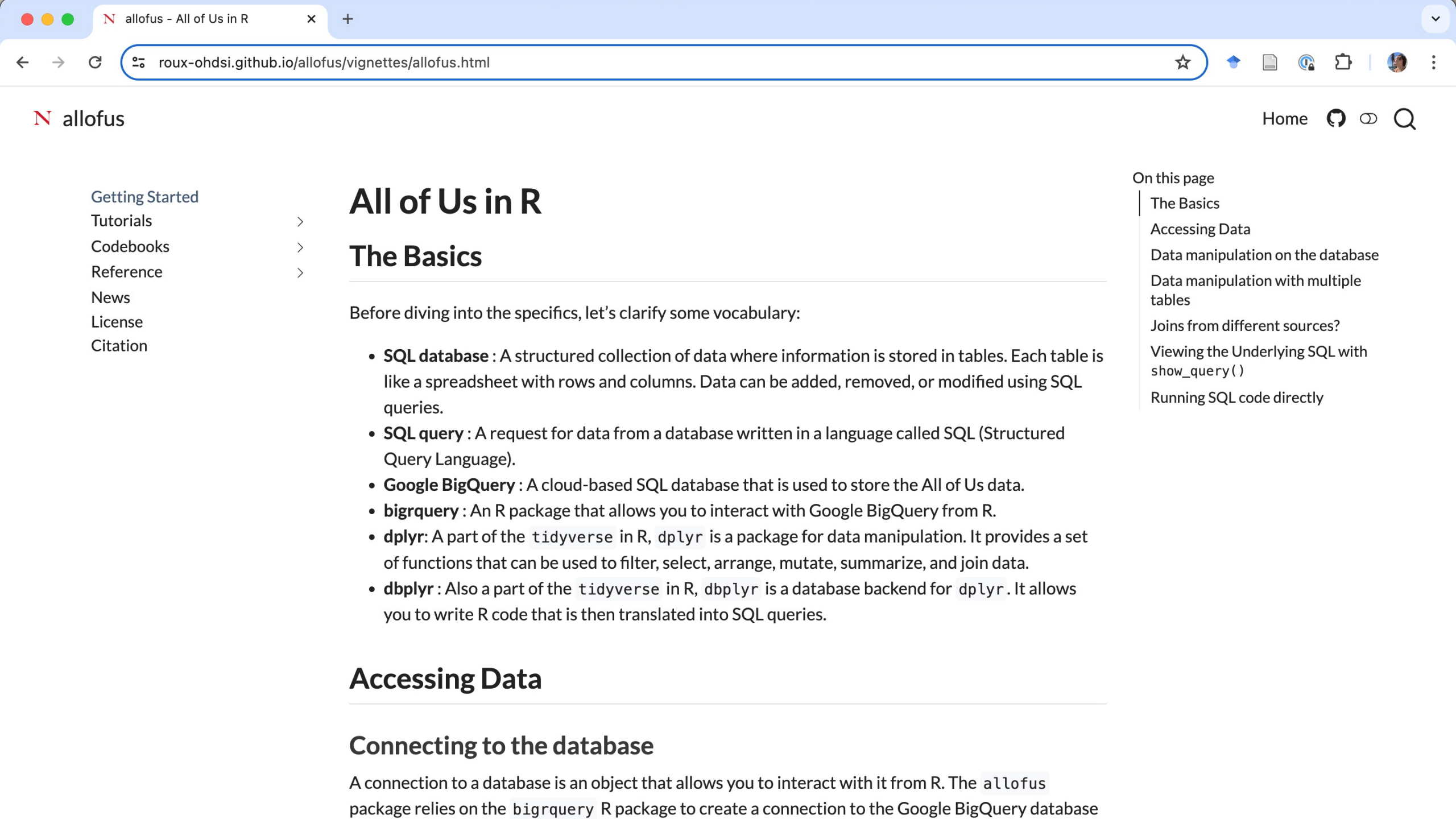
*Run SQL queries*  
`aou_sql()`

*Merge tables*  
`aou_join()`

*Bring in local data*  
`aou_create_temo_table()`

*Save objects to the bucket*  
`aou_workspace_to_bucket()`

*Save session info for  
reporting and reproducibility*  
`aou_session_info()`



Getting Started

Tutorials

Codebooks

Reference

News

License

Citation

# All of Us in R

## The Basics

Before diving into the specifics, let's clarify some vocabulary:

- **SQL database** : A structured collection of data where information is stored in tables. Each table is like a spreadsheet with rows and columns. Data can be added, removed, or modified using SQL queries.
- **SQL query** : A request for data from a database written in a language called SQL (Structured Query Language).
- **Google BigQuery** : A cloud-based SQL database that is used to store the All of Us data.
- **bigquery** : An R package that allows you to interact with Google BigQuery from R.
- **dplyr**: A part of the tidyverse in R, dplyr is a package for data manipulation. It provides a set of functions that can be used to filter, select, arrange, mutate, summarize, and join data.
- **dbplyr** : Also a part of the tidyverse in R, dbplyr is a database backend for dplyr . It allows you to write R code that is then translated into SQL queries.

## Accessing Data

### Connecting to the database

A connection to a database is an object that allows you to interact with it from R. The allofus package relies on the bigquery R package to create a connection to the Google BigQuery database

#### On this page

The Basics

Accessing Data

Data manipulation on the database

Data manipulation with multiple tables

Joins from different sources?

Viewing the Underlying SQL with show\_query()

Running SQL code directly



# Resources

- Preprint about the package:  
<https://www.medrxiv.org/content/10.1101/2024.04.10.24305611v1>
- R package site (read documentation and tutorials): <https://roux-ohdsi.github.io/allofus/>
- Source code (report issues) <https://github.com/roux-ohdsi/allofus>
- v1.1.0 on CRAN: <https://cran.r-project.org/web/packages/allofus/index.html>
- *All of Us* Data Browser (see what data is available)  
<https://databrowser.researchallofus.org/>
- Institutional DUA (if your institution does not already have one) <https://redcap.pmi-ops.org/surveys/?s=7N7TA9AHAA>
- Register for access (create an account once your institution has a DUA):  
<https://www.researchallofus.org/register/>

[l.smith@northeastern.edu](mailto:l.smith@northeastern.edu), [r.cavanaugh@northeastern.edu](mailto:r.cavanaugh@northeastern.edu)