allofus: An R package to facilitate use of the All of Us Researcher 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.
# Data in the *All of Us* research program

<table>
<thead>
<tr>
<th>EHR Domains</th>
<th>Genomics</th>
<th>Measurements and Wearables</th>
<th>Survey Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditions</td>
<td>Genomic data only in Researcher Workbench</td>
<td>Physical Measurements</td>
<td>The Basics</td>
</tr>
<tr>
<td>25,638 medical concepts</td>
<td>1,040 participants in the Long-Read WGS dataset</td>
<td>8</td>
<td>28 questions available</td>
</tr>
<tr>
<td>254,700 participants</td>
<td>11,400 participants in the Short-Read WGS Structural Variants dataset</td>
<td>Physical Measurements</td>
<td>21 questions available</td>
</tr>
<tr>
<td>Drug Exposures</td>
<td>SNV/Indel Variants</td>
<td>337,540 participants</td>
<td>413,360 participants</td>
</tr>
<tr>
<td>29,865 medical concepts</td>
<td>245,400 SNV/Indel Variants</td>
<td>Participants have the option to provide a standard set of physical measurements.</td>
<td>21 questions available</td>
</tr>
<tr>
<td>239,740 participants</td>
<td></td>
<td></td>
<td>412,220 participants</td>
</tr>
<tr>
<td>Labs &amp; Measurements</td>
<td>Genotyping Arrays dataset</td>
<td></td>
<td>26 questions available</td>
</tr>
<tr>
<td>16,618 medical concepts</td>
<td>Register for access</td>
<td></td>
<td>412,220 participants</td>
</tr>
<tr>
<td>255,640 participants</td>
<td>View Physical Measurements</td>
<td></td>
<td>View Complete Survey</td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
<td>Fitbit</td>
<td>Health Care Access &amp; Utilization</td>
</tr>
<tr>
<td>30,328 medical concepts</td>
<td></td>
<td>6 Fitbit Measurements</td>
<td>57 questions available</td>
</tr>
<tr>
<td>242,580 participants</td>
<td></td>
<td>18,620 participants</td>
<td>View Complete Survey</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fitbit data includes heart rate and activity summaries.</td>
<td></td>
</tr>
</tbody>
</table>
Using the All of Us workbench
{allofus} Package workflow

Plan an analysis using reference data and online tools
- `aou_table_info`
- `aou_codebook`
- `aou_health_history`

Connect to the database
- `aou_connect()`

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

Extract survey data
- `aou_survey()`

Build cohorts using ATLAS
- `aou_atlas_cohort()`

Access tables directly
- `dplyr::tbl(con, “person”)`

Merge tables
- `aou_join()`

Use `dplyr/tidyr` functions to wrangle data via `dplyr`

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

Bring in local data
- `aou_create_temo_table()`

Bring tables into local memory
- `aou_collect()`

Save objects to the bucket
- `aou_workspace_to_bucket()`

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

Run SQL queries
- `aou_sql()`

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()
```

**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

Extract survey data

aou_survey()
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_tempo_table()`

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

- Save session info for reporting and reproducibility
  - `aou_session_info()`
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.
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, r.cavanaugh@northeastern.edu