

Package: ds4owdanalytics (via r-universe)

October 9, 2026

Type Package

Title DS4OWD Analytics

Version 0.1.0

Description Data collected from the Data Science for Open WASH Data (DS4OWD) course and the associated website and data publishing endeavour.

License CC BY 4.0

Encoding UTF-8

LazyData true

Depends R (>= 3.5)

RoxygenNote 7.3.2

Language en-GB

Config/Needs/website rmarkdown

Date 2025-01-31

URL <https://github.com/openwashdata/ds4owdanalytics>

BugReports <https://github.com/openwashdata/ds4owdanalytics/issues>

Repository <https://openwashdata.r-universe.dev>

Date/Publication 2025-01-31 11:07:04 UTC

RemoteUrl <https://github.com/openwashdata/ds4owdanalytics>

RemoteRef HEAD

RemoteSha f0294f5a813a8b741310f1af9ac84643ac8d7baa

Contents

course_participation	2
ds4owd_browser	2
ds4owd_country	3
ds4owd_source	4
ds4owd_timeseries	4

locations	5
postsurvey	5
presurvey	6
publishing_metadata	7
Index	8

course_participation	<i>course_participation: Participation in the Course</i>
----------------------	--

Description

Sessions attended and participated in, for the course (anonymized)

Usage

course_participation

Format

- A tibble with 734 rows and 5 variables
- duration_minutes** Minutes spent in the session
- guest** Whether the user is a guest or not. The course instructor and their notetaker is not a guest
- date** Date of the session
- session** Session number
- anonymized_name** Anonymized hash of the name as an identifier

ds4owd_browser	<i>ds4owd_browser: Data on visits to the DS4OWD website by browser</i>
----------------	--

Description

This dataset contains information on visits to the DS4OWD website by browser used.

Usage

ds4owd_browser

Format

A tibble with 17 rows and 6 variables

browser Browser used to accesss the DS4OWD website

pageviews Number of pages viewed

visitors Total number of unique visitors

visit_duration Duration of visit (seconds)

visits Number of non-unique visits to the website

bounce_rate Percentage of users "bouncing" off website. A bounce is when a visitor spends less than 30 seconds on the website

ds4owd_country	<i>ds4owd_country: Data on visits to the DS4OWD website by country</i>
----------------	--

Description

Visits to the DS4OWD website by country

Usage

ds4owd_country

Format

A tibble with 88 rows and 6 variables

country Country of visit to the website

pageviews Number of pages viewed

visit_duration Duration of visit (seconds)

visitors Total number of unique visitors

visits Number of non-unique visits to the website

bounce_rate Percentage of users "bouncing" off website. A bounce is when a visitor spends less than 30 seconds on the website

ds4owd_source	<i>ds4owd_source: Data on source of website visits</i>
---------------	--

Description

Data on source of website visits

Usage

ds4owd_source

Format

A tibble with 27 rows and 6 variables

source Source of visit (Search engine, organic etc.)

pageviews Number of pages viewed

visit_duration Duration of visit (seconds)

visitors Total number of unique visitors

visits Number of non-unique visits to the website

bounce_rate Percentage of users "bouncing" off website. A bounce is when a visitor spends less than 30 seconds on the website

ds4owd_timeseries	<i>ds4owd_timeseries: Timeseries data on website traffic</i>
-------------------	--

Description

Timeseries data on website traffic to the DS4OWD website

Usage

ds4owd_timeseries

Format

A tibble with 468 rows and 6 variables

date Date of data collection

pageviews Number of pages viewed

visitors Total number of unique visitors

visit_duration Duration of visit (seconds)

visits Number of non-unique visits to the website

bounce_rate Percentage of users "bouncing" off website. A bounce is when a visitor spends less than 30 seconds on the website

locations	<i>locations: Location of the course participants</i>
-----------	---

Description

Location by country (and city) of the course participants. This is anonymized

Usage

locations

Format

A tibble with 212 rows and 3 variables

iso_country ISO Name for country of course participant

country Name of country

city Name of city

postsurvey	<i>postsurvey: Postsurvey responses</i>
------------	---

Description

Data from the postsurvey responses of the course participants in DS4OWD

Usage

postsurvey

Format

A tibble with 25 rows and 18 variables

course_completion Whether the course was completed by the respondent or not

rating_overall Overall course rating

expectations Expectations from the course

learning_objectives Whether learning objectives were met

rating_structure Rating for the structure of the course

rating_ins_knowledge Rating for the instructor's knowledge of the course material

rating_ins_clarity Rating for the instructor's clarity

rating_ins_comp Rating for the instructor's competency

rating_ins_app Rating for approachability of the instructor

feedback Feedback for the course
rating_self_r_comp Self rating for R competency
rating_self_vc_comp Self rating for version control competency
conf_skill_app Confidence in applying skills learned during the course
most_liked Most liked aspects of the course
least_liked Least liked aspects of the course
suggestion Suggestions for the course
other_comments Any other comments
reason_non_participation If course was not completed

presurvey

presurvey: Responses from the pre-course survey

Description

Data from the pre-course survey, including previous programming experience, ease of programming, data format, current level of LLM usage, organisation type, and barriers to learning.

Usage

presurvey

Format

A tibble with 212 rows and 15 variables

id ID of survey response

google_timestamp

zoom_timestamp

prog_exp Previous programming experience

prog_exp_r Previous programming experience in R

prog_ease_max_list Ease of programming

data_format Format of data

current_use_llm Current level of LLM usage

work_organisation Organisation type where participant works

barrier_time_availability If time availability is a barrier to learning

barrier_supervisor_interests If supervisor's interest is a barrier to learning

barrier_internet_connection If internet connection is a barrier to learning

barrier_electricity If electricity availability is a barrier to learning

barrier_computer If availability of a computer is a barrier to learning

interest_share_dataset If interested in sharing a dataset

publishing_metadata	<i>publishing_metadata: Publishing Metadata</i>
---------------------	---

Description

Metadata on packages published directly as a result of DS4OWD.

Usage

```
publishing_metadata
```

Format

A tibble with 34 rows and 14 variables

id Identifier of dataset

pkg_name Name of the package

maintainer Github ID of maintainer

source Source of data

difficulty Difficult of cleaning data

status Current status

description Description of data package

location Location of data in the package

date_published Date of publication

link_github Link to github repo

link_pkg_website Link to package website

doi Zenodo DOI of package

temporal_coverage Temporal coverage of data

keywords Keywords to identify data

Index

* **datasets**

- course_participation, [2](#)
- ds4owd_browser, [2](#)
- ds4owd_country, [3](#)
- ds4owd_source, [4](#)
- ds4owd_timeseries, [4](#)
- locations, [5](#)
- postsurvey, [5](#)
- presurvey, [6](#)
- publishing_metadata, [7](#)

course_participation, [2](#)

ds4owd_browser, [2](#)
ds4owd_country, [3](#)
ds4owd_source, [4](#)
ds4owd_timeseries, [4](#)

locations, [5](#)

postsurvey, [5](#)
presurvey, [6](#)
publishing_metadata, [7](#)