# **OPEN MNO MOBILITY DATA IN SPAIN:** A REPRODUCIBLE SHARING APPROACH VIA R PACKAGE **'SPANISHODDATA'**

**E. Kotov**, E. Vidal-Tortosa, O. Cantu-Ros, J. Burrieza-Galán,
R. Herranz, T. Gullón Muñoz-Repiso, R. Lovelace

MNO-MINDS, Paris, 24 Sep 2025

# NATIONAL STATISTICAL OFFICE EXPERIMENTING WITH COOL DATA INTERNALLY


Image: OpenAI DALL-E + Google NanoBanana

# NATIONAL STATISTICAL OFFICE EXPERIMENTING WITH COOL DATA INTERNALLY

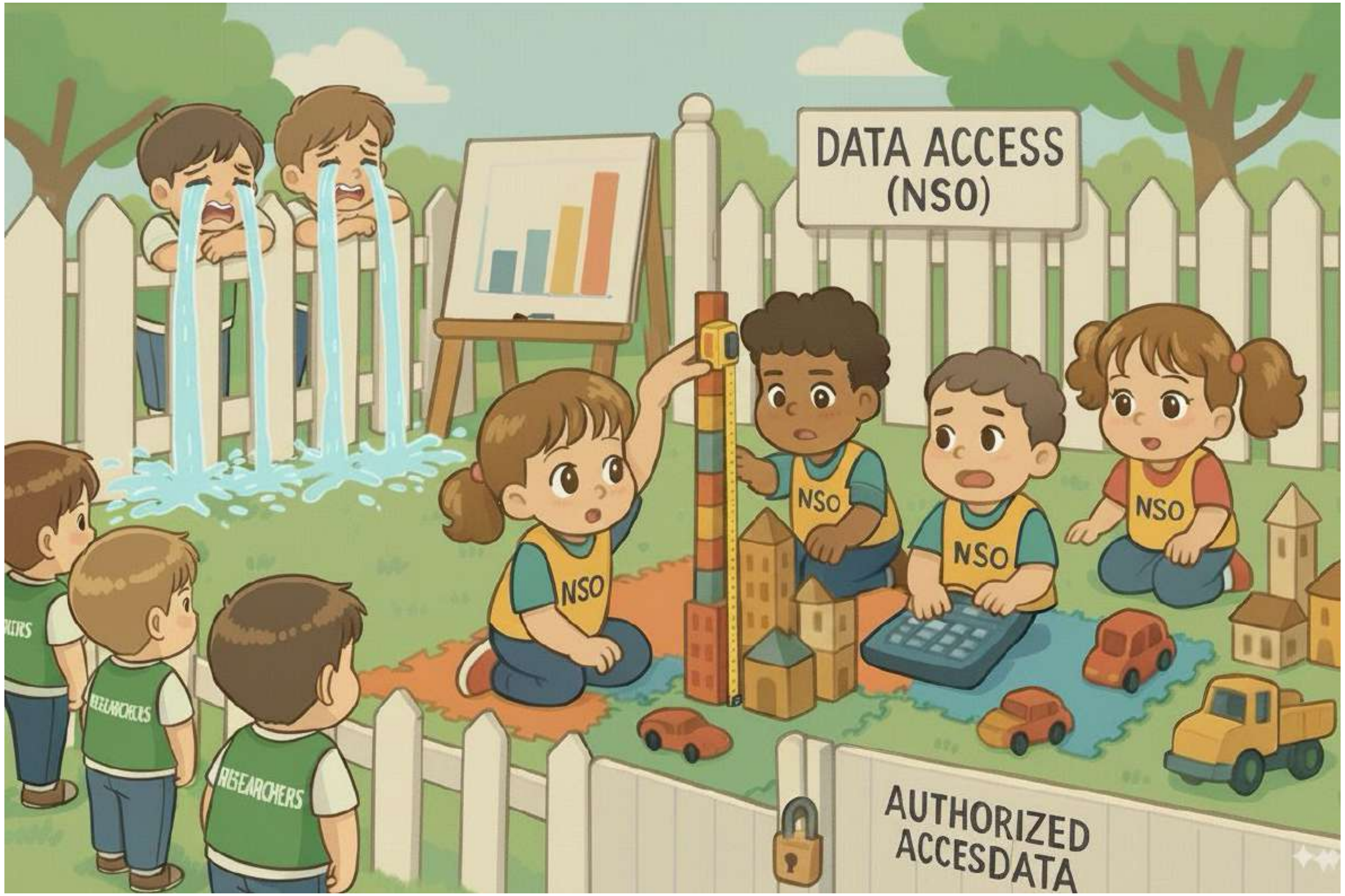



Image: OpenAI DALL-E + Google NanoBanana

# SPANISH OPEN MOBILITY BIG DATA

MINISTERIO
DE TRANSPORTES
Y MOVILIDAD SOSTENIBLE

Buscar

Ministerio Transporte Terrestre Carreteras Ferroviario Aéreo Marítimo Geo-información Movilidad Sostenible

Ministerio › Proyectos singulares › Estudio de la movilidad con Big Data

## Open Data Movilidad

En este espacio compartido están disponibles los datos de movilidad en España para el periodo desde enero de 2022 en adelante. El Ministerio de Transportes apuesta por los **datos abiertos** con el objetivo de fomentar la transparencia, la eficiencia, la participación ciudadana y el desarrollo económico, ya que los datos pueden consultarse, ser enriquecidos con nuevos datos, aplicaciones y servicios y generar nuevos negocios.

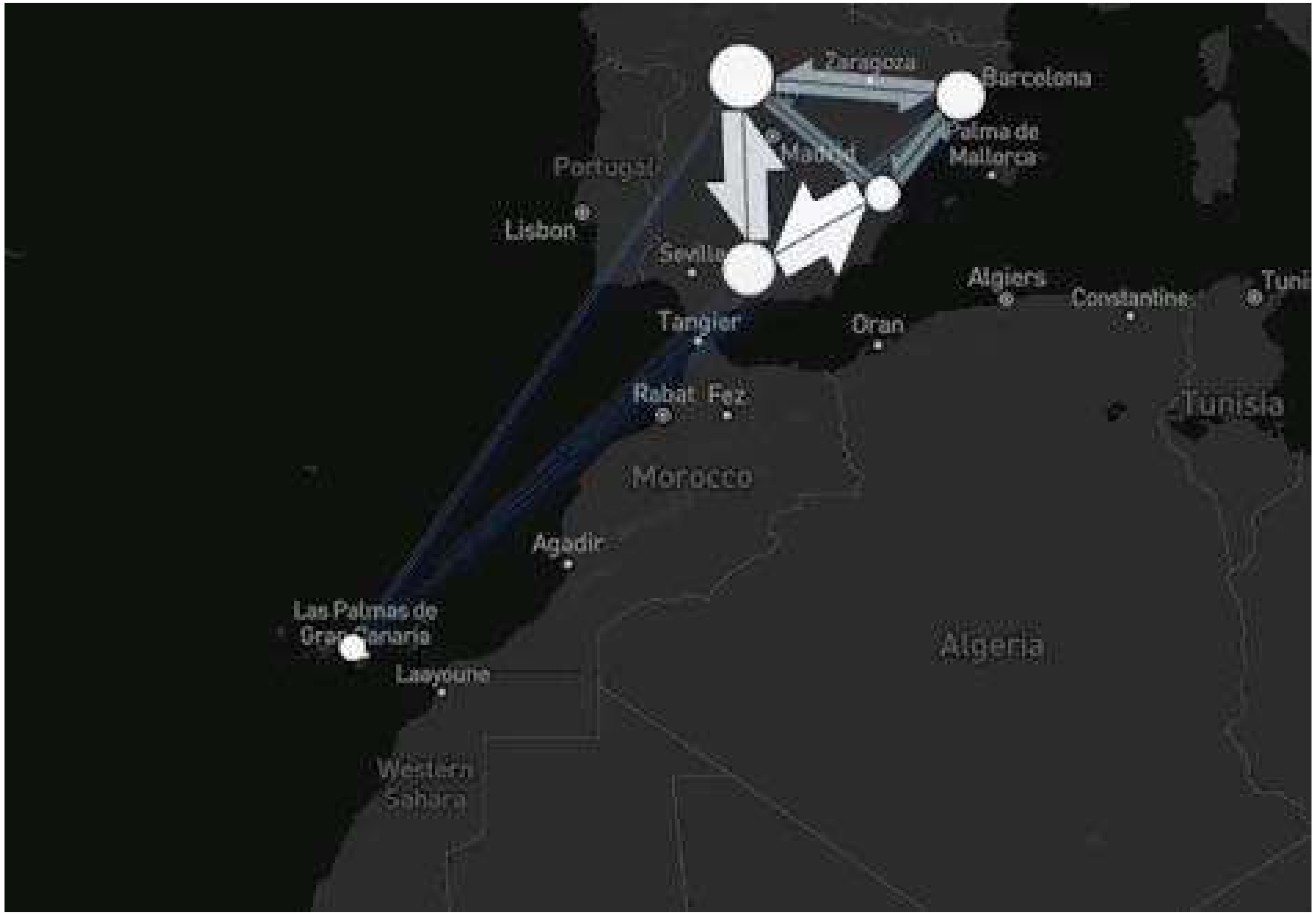
- Se ha utilizado como fuente principal de datos el posicionamiento de los **teléfonos móviles**, siendo una condición indispensable el cumplimiento de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- El contenido está estructurado en carpetas según el **tipo de estudio** y en subcarpetas por **zonificación** (municipal, distritos y GAUS). Para cada zonificación hay una carpeta por tipo de producto y otra para los metadatos referidos a la calidad del dato. Asimismo, se dispone de otra carpeta con la **zonificación** empleada y su información asociada.

El proyecto se divide en 3 **tipos de estudios**:

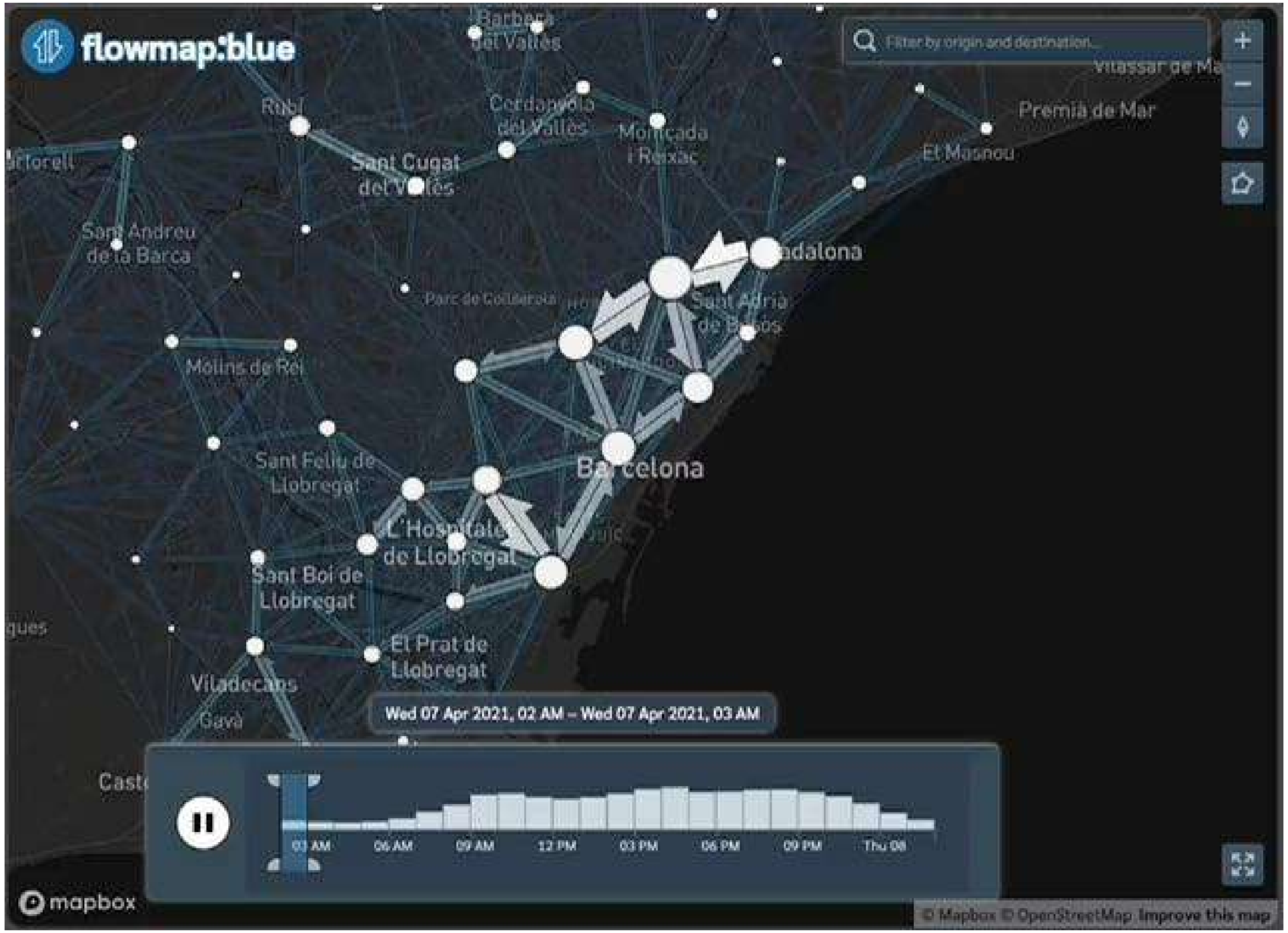
1. **Estudios básicos**, donde se analiza diariamente la movilidad de los residentes en España a nivel de distrito censal. Como resultados se dispone de las matrices de viajes, pernoctaciones y personas.

*https://www.transportes.gob.es/ministerio/proyectos-singulares/estudios-de-movilidad-con-big-data/opendata-movilidad*

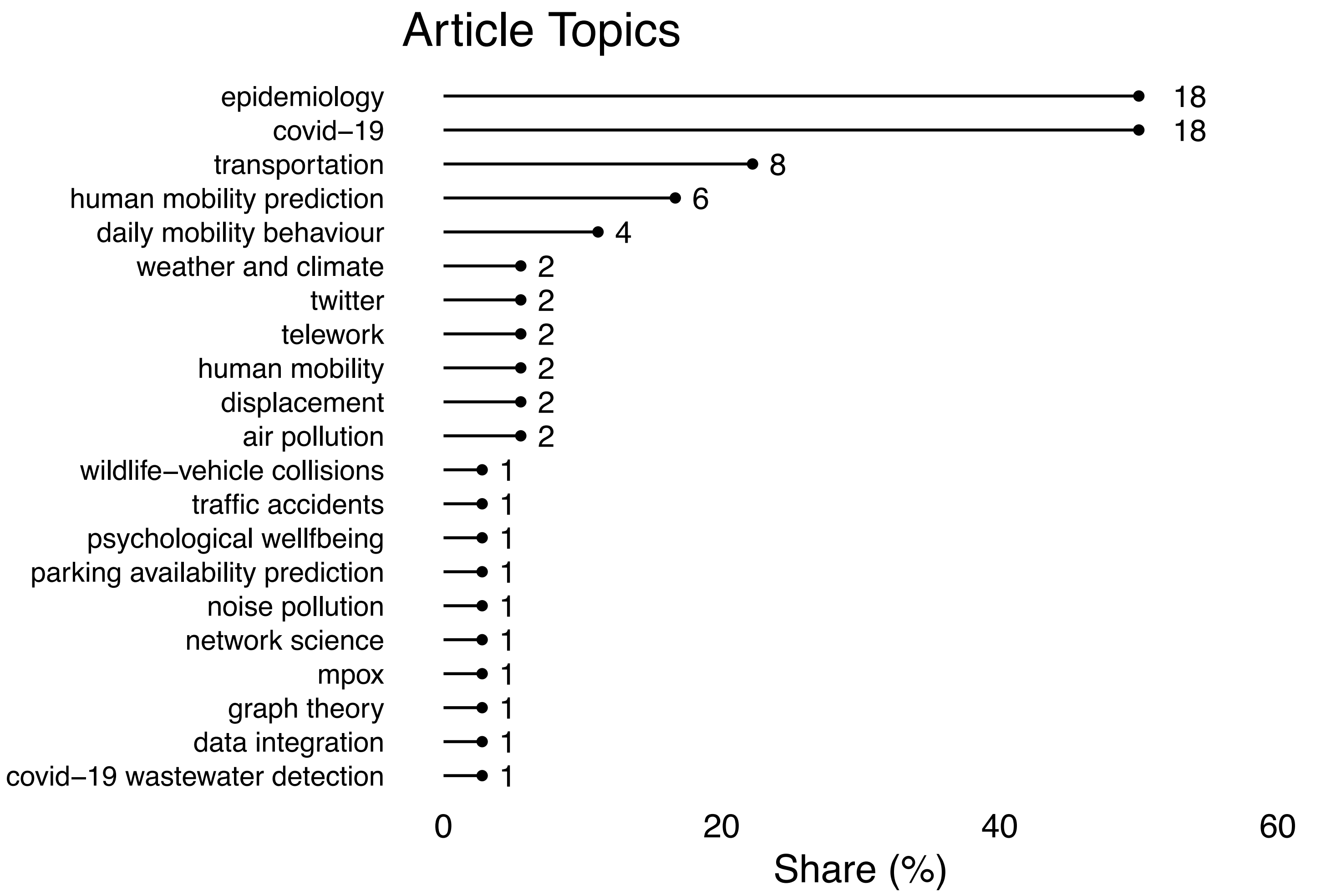
# SPANISH OPEN MOBILITY BIG DATA



# SPANISH OPEN MOBILITY BIG DATA



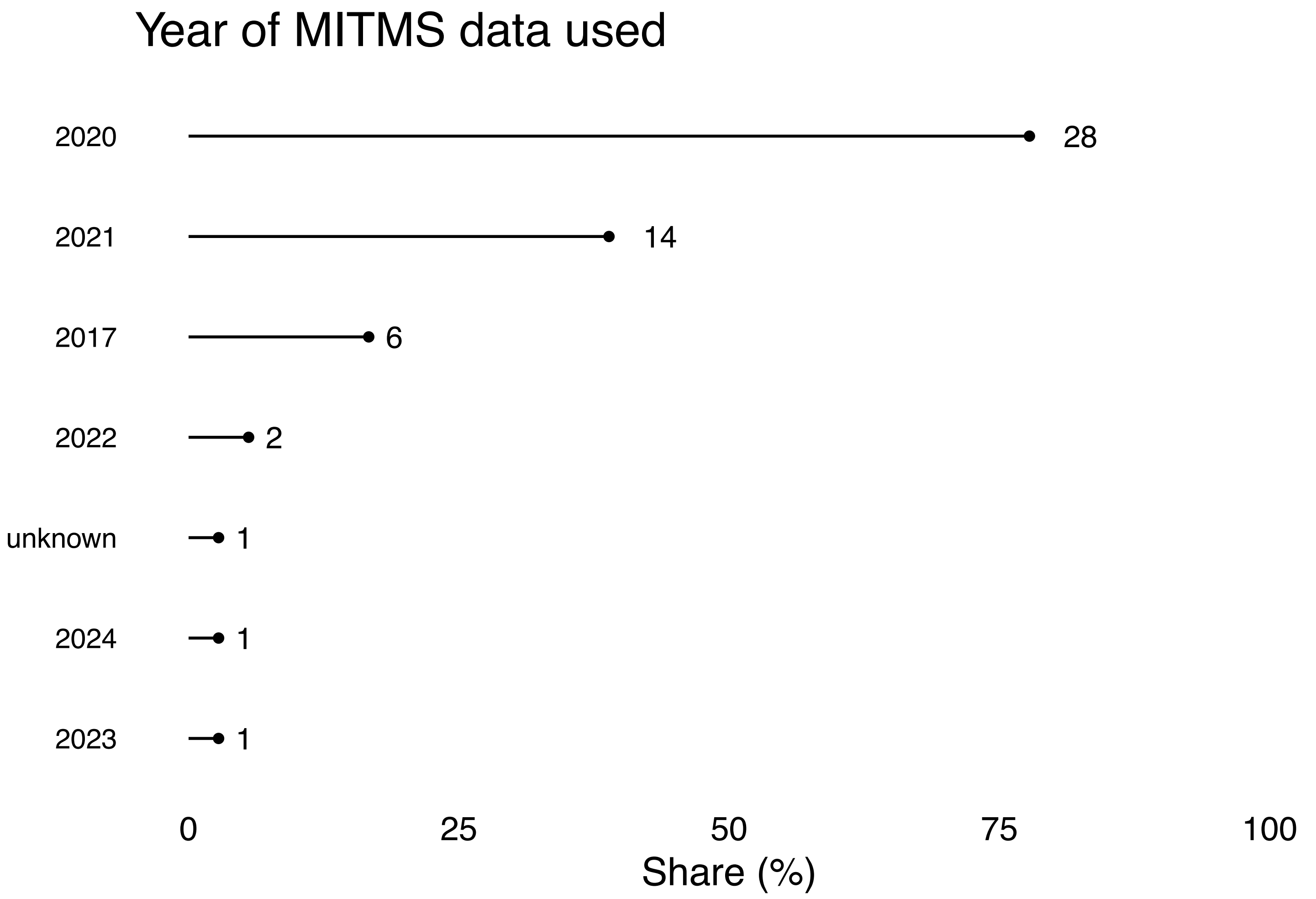
# TOPICS EXPLORED WITH MITMS DATA




May not sum to 100%
Share (on x axis) and Number (on plot) of Articles by Article Topic

https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

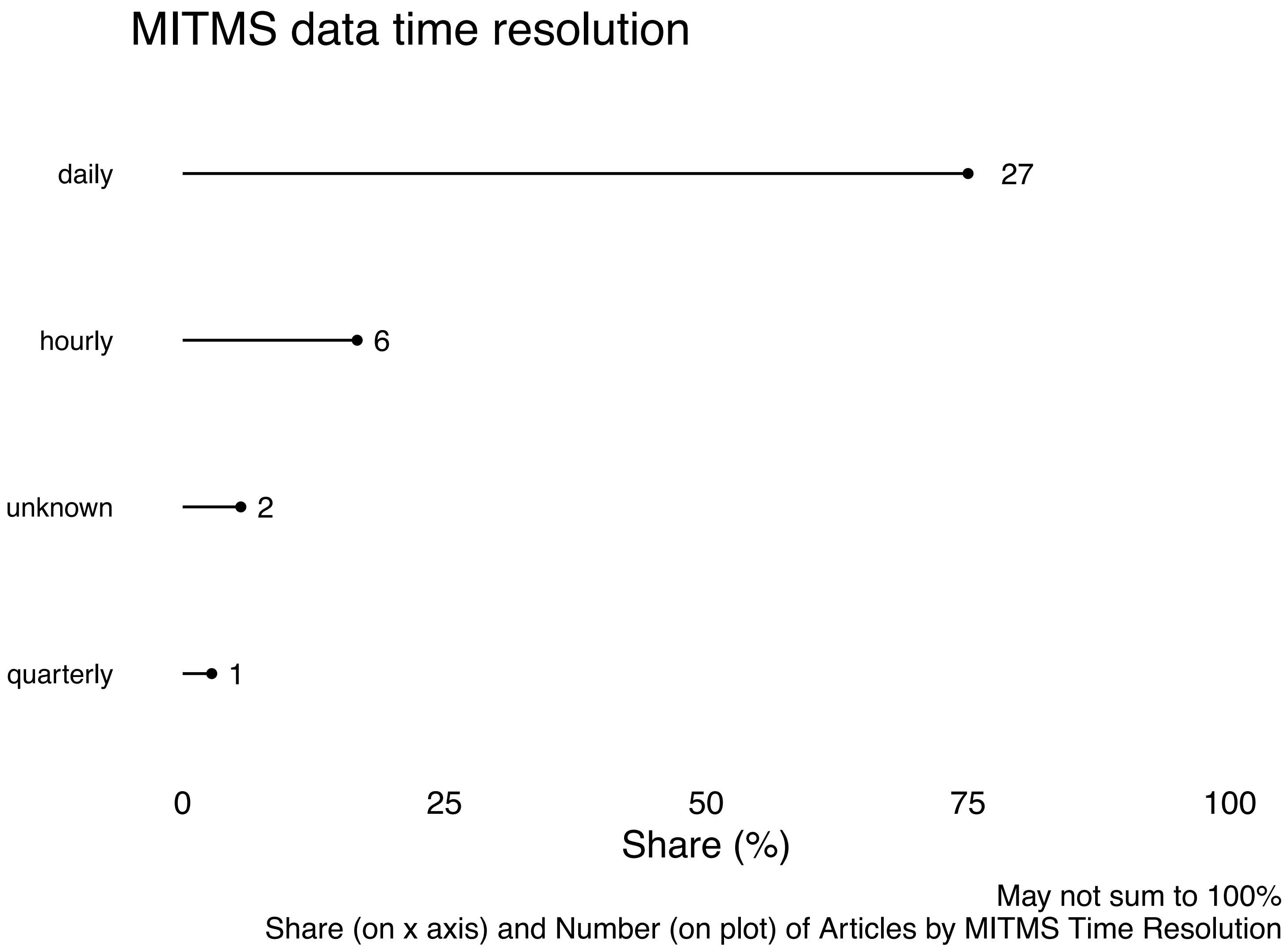
# POTENTIAL OF 2022+ DATA IS STILL UNTAPPED




May not sum to 100%
Share (on x axis) and Number (on plot) of Articles by Years of MITMS data

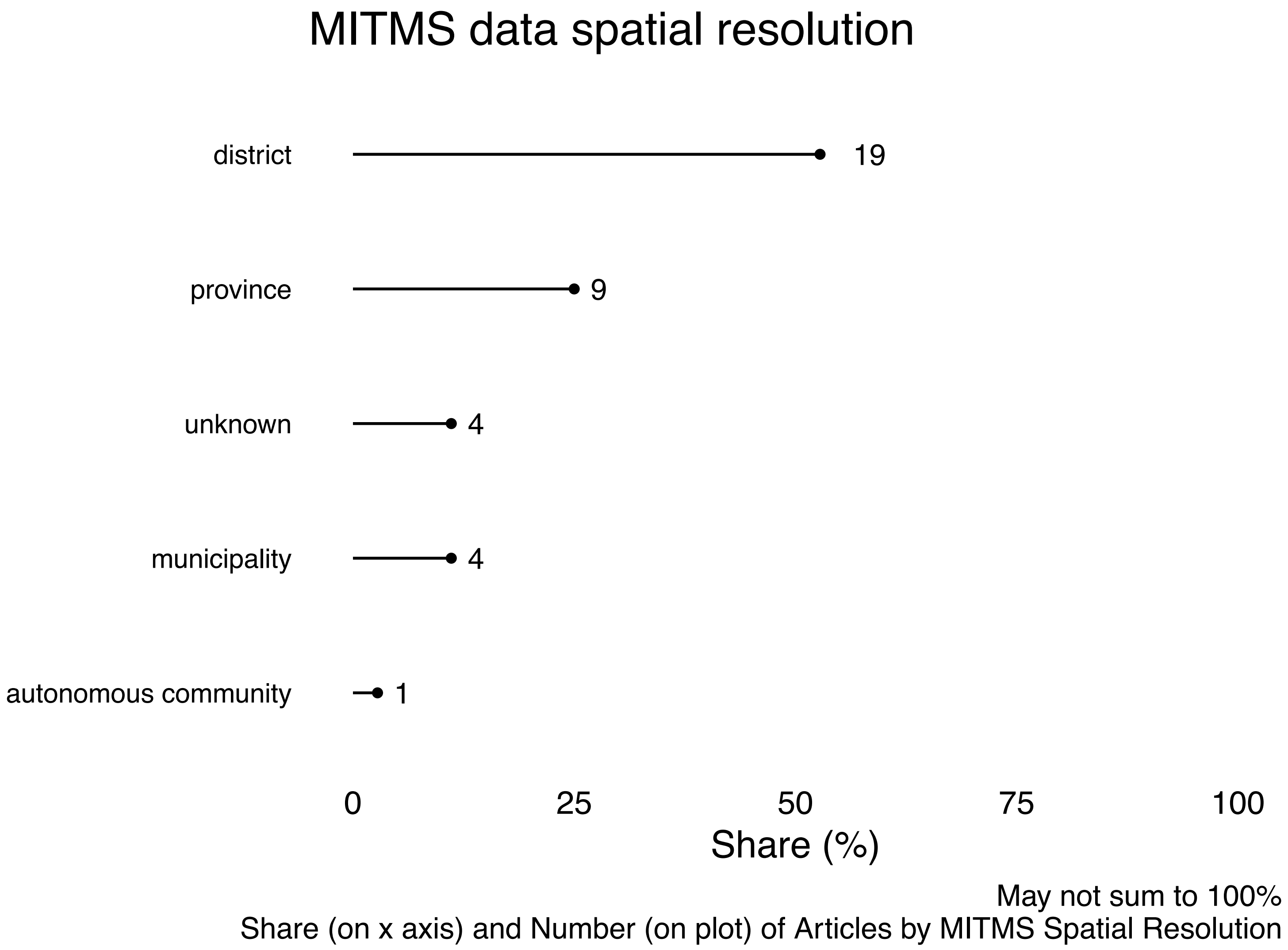
https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

# DAILY TIME RESOLUTION IS MOST POPULAR




https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

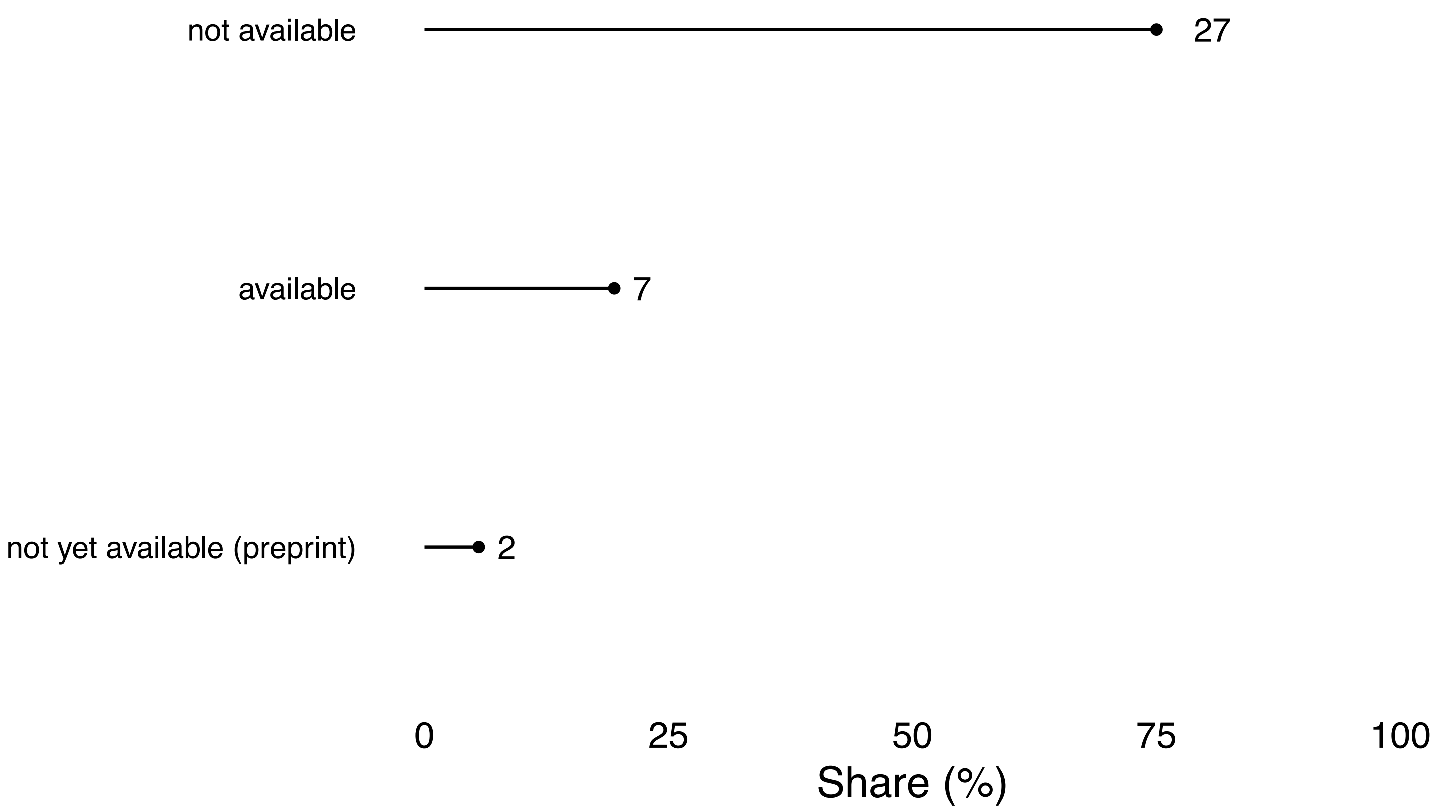
# FINEST DISTRICT SPATIAL RESOLUTION IS MOST POPULAR




https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

# PRE-PROCESSING CODE IS USUALLY NOT SHARED

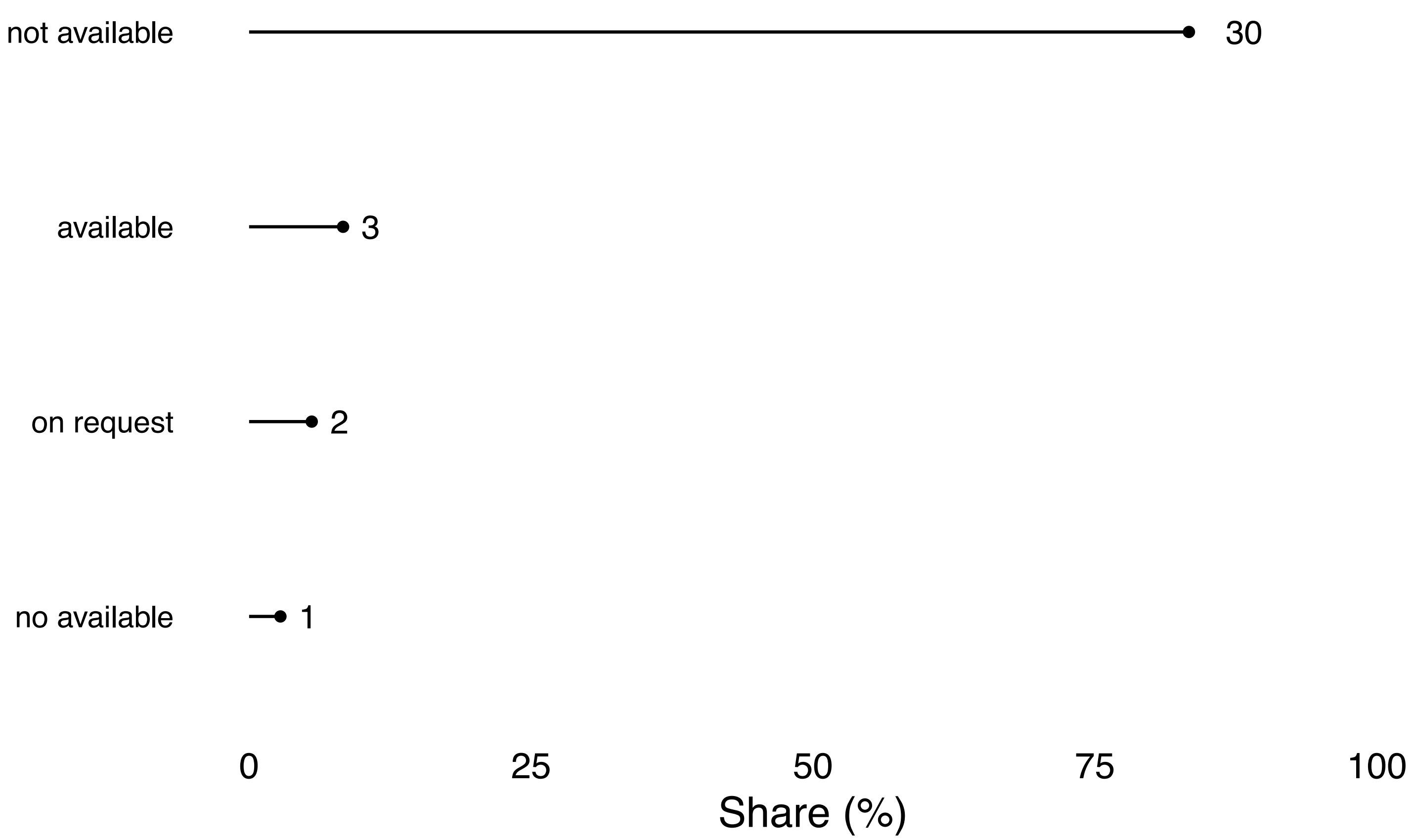
MITMS data pre–processing code




Share (on x axis) and Number (on plot) of Articles by Data Pre–Processing Reproducibility

https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

# PRE-PROCESSED DATA IS USUALLY NOT SHARED

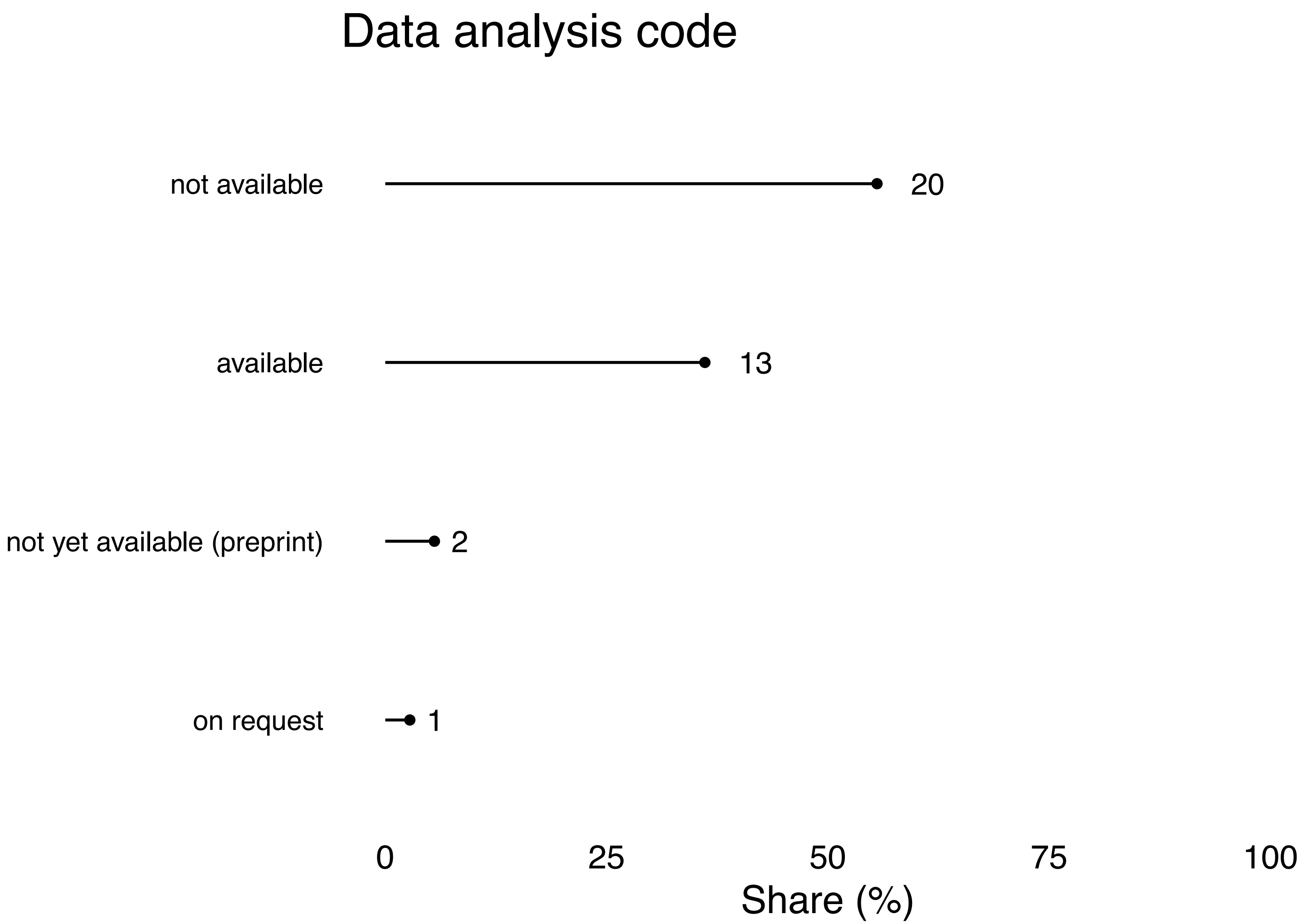



Share (on x axis) and Number (on plot) of Articles by Shared Pre–processed MITMS Data

*https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots*

**NOBODY** PROVIDES THE CODE HOW TO ACQUIRE THE DATA

# FEW PROVIDE ANALYSIS CODE




https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

# VERY LITTLE VALUE IN ANALYSIS CODE WITHOUT ACQUISITION AND PROCESSING

# VERY LITTLE VALUE IN ANALYSIS CODE WITHOUT ACQUISITION AND PROCESSING

How to get the data ???

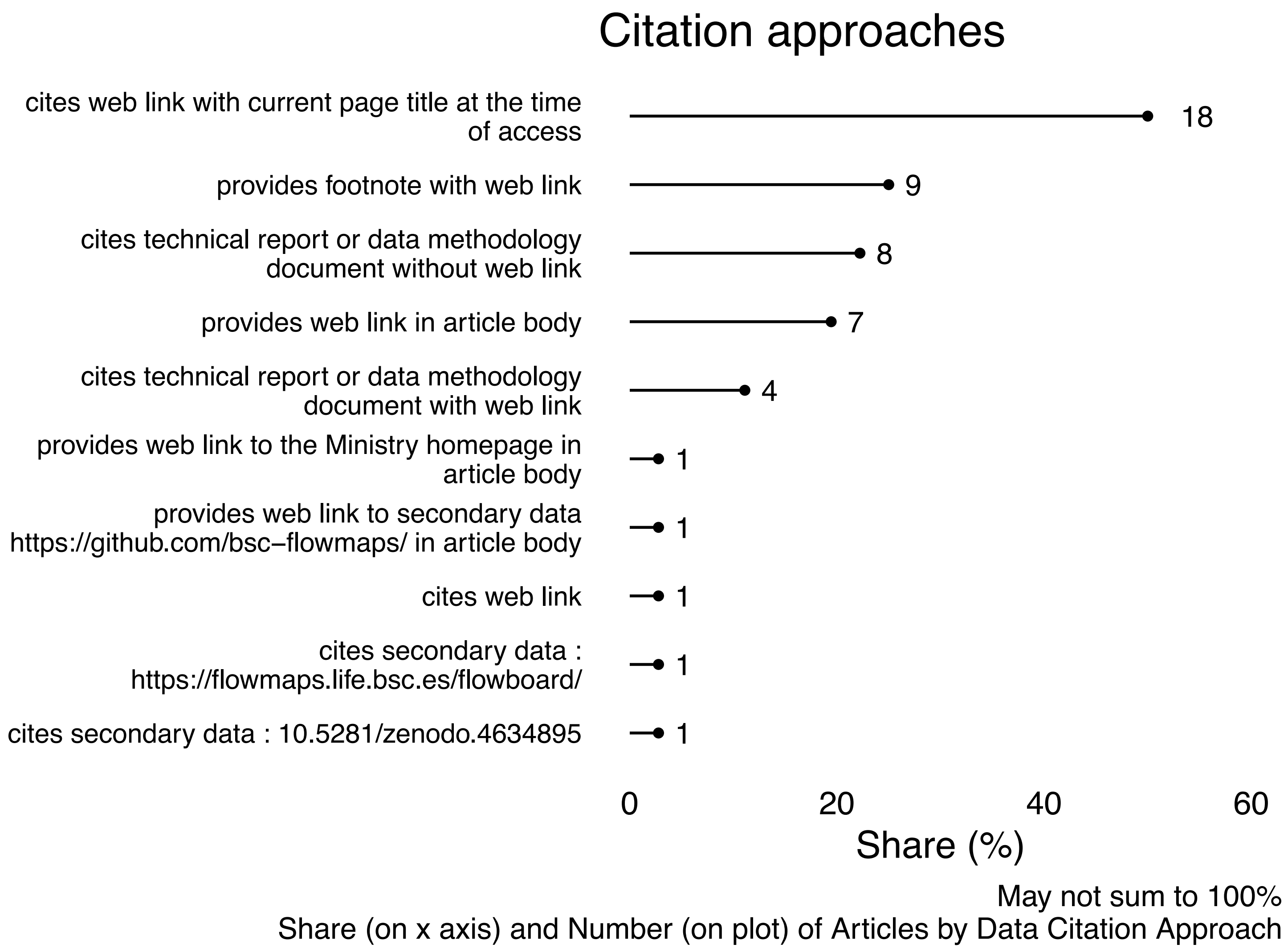
→

How to preprocess it?

→

~~analysis~~

# INCONSISTENT CITATION




https://github.com/e-kotov/spanishoddata-paper-supplement/tree/main/plots/supplement-plots

# DATA ACCESS

mitma-movilidad-v2 / estudios_basicos / por-municipios / viajes / ficheros-diarios / 2022-01

Mostrar 50 registros

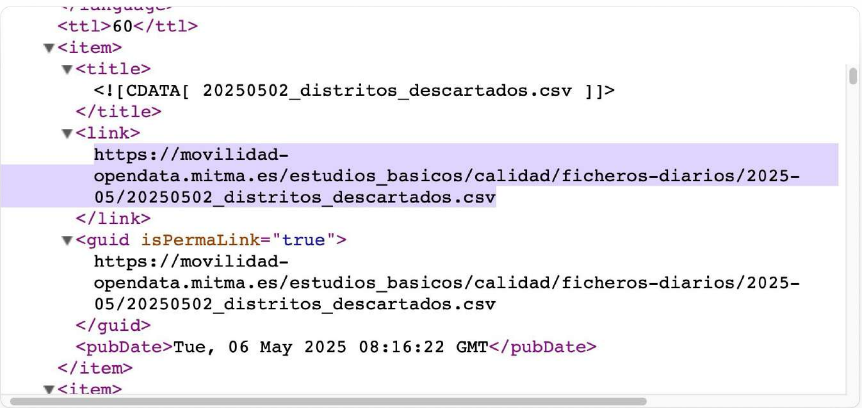
Buscar:

| Archivo | Fecha | Tamaño |
|---|---|---|
| [Volver a la carpeta padre] | | |
| 20220101_Viajes_municipios.csv.gz | 2023-06-29 16:19:58 | 94 MB |
| 20220102_Viajes_municipios.csv.gz | 2023-06-29 16:20:45 | 100 MB |
| 20220103_Viajes_municipios.csv.gz | 2023-06-29 16:21:35 | 115 MB |
| 20220104_Viajes_municipios.csv.gz | 2023-06-29 16:22:33 | 115 MB |
| 20220105_Viajes_municipios.csv.gz | 2023-06-29 16:23:33 | 114 MB |
| 20220106_Viajes_municipios.csv.gz | 2023-06-29 16:24:19 | 92 MB |

# DATA ACCESS

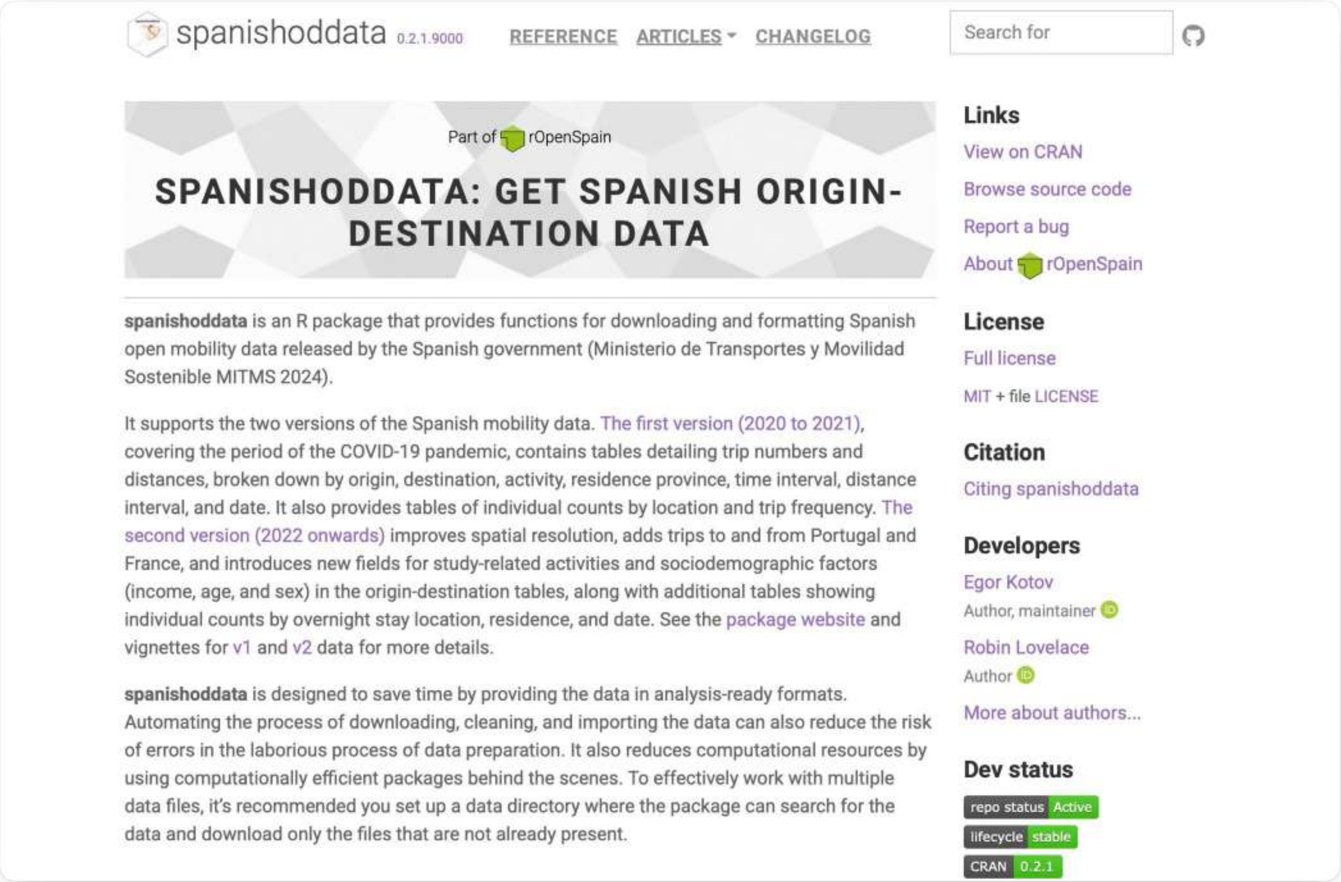

```
    <ttl>60</ttl>
  ▼<item>
    ▼<title>
        <![CDATA[ 20250502_distritos_descartados.csv ]]>
     </title>
    ▼<link>
        https://movilidad-
        opendata.mitma.es/estudios_basicos/calidad/ficheros-diarios/2025-
        05/20250502_distritos_descartados.csv
     </link>
    ▼<guid isPermaLink="true">
        https://movilidad-
        opendata.mitma.es/estudios_basicos/calidad/ficheros-diarios/2025-
        05/20250502_distritos_descartados.csv
     </guid>
     <pubDate>Tue, 06 May 2025 08:16:22 GMT</pubDate>
    </item>
  ▼<item>
```

# DATA ACCESS


- Custom code to download and import multiple days
- Variable names in Spanish
- No guarantee of consistent variable types
- Limited by available memory
- Slow data processing (raw csv data)

# SPANISHODDATA R PACKAGE



spanishoddata 0.2.1.9000 REFERENCE ARTICLES CHANGELOG

Search for

Part of rOpenSpain

## SPANISHODDATA: GET SPANISH ORIGIN-DESTINATION DATA

**spanishoddata** is an R package that provides functions for downloading and formatting Spanish open mobility data released by the Spanish government (Ministerio de Transportes y Movilidad Sostenible MITMS 2024).

It supports the two versions of the Spanish mobility data. The first version (2020 to 2021), covering the period of the COVID-19 pandemic, contains tables detailing trip numbers and distances, broken down by origin, destination, activity, residence province, time interval, distance interval, and date. It also provides tables of individual counts by location and trip frequency. The second version (2022 onwards) improves spatial resolution, adds trips to and from Portugal and France, and introduces new fields for study-related activities and sociodemographic factors (income, age, and sex) in the origin-destination tables, along with additional tables showing individual counts by overnight stay location, residence, and date. See the package website and vignettes for v1 and v2 data for more details.

**spanishoddata** is designed to save time by providing the data in analysis-ready formats. Automating the process of downloading, cleaning, and importing the data can also reduce the risk of errors in the laborious process of data preparation. It also reduces computational resources by using computationally efficient packages behind the scenes. To effectively work with multiple data files, it's recommended you set up a data directory where the package can search for the data and download only the files that are not already present.

**Links**

View on CRAN

Browse source code

Report a bug

About rOpenSpain

**License**

Full license

MIT + file LICENSE

**Citation**

Citing spanishoddata

**Developers**

Egor Kotov
Author, maintainer

Robin Lovelace
Author

More about authors...

**Dev status**

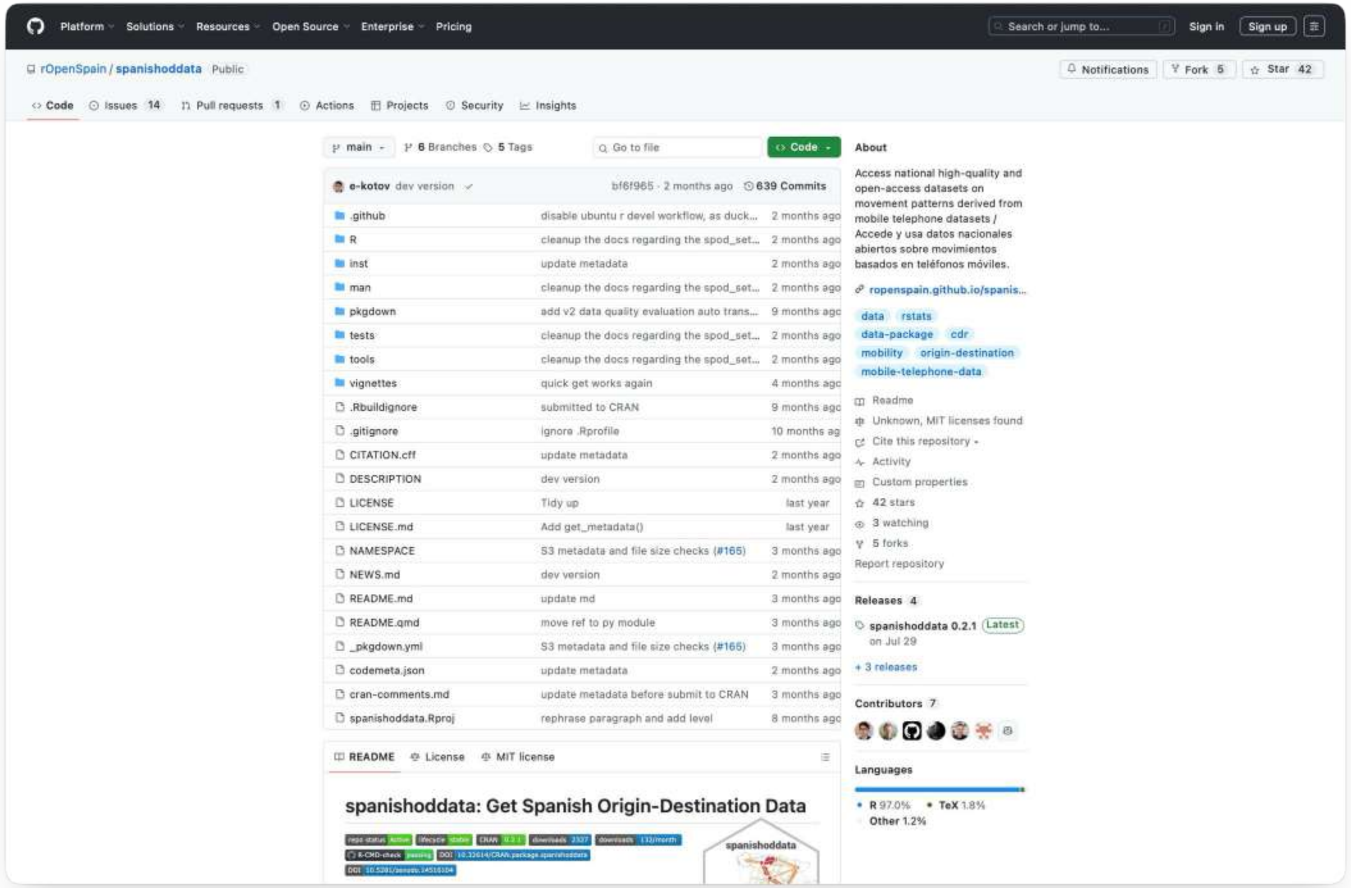
repo status Active

lifecycle stable

CRAN 0.2.1

*https://ropenspain.github.io/spanishoddata/ || Logos are the property of their respective owners.*

# SPANISHODDATA R PACKAGE




*https://ropenspain.github.io/spanishoddata/ || Logos are the property of their respective owners.*

# DATA ACCESS

mitma-movilidad-v2 / estudios_basicos / por-municipios / viajes / ficheros-diarios / 2022-01

Mostrar 50 registros    Buscar:

| Archivo | Fecha | Tamaño |
|---|---|---|
| [Volver a la carpeta padre] | | |
| 20220101_Viajes_municipios.csv.gz | 2023-06-29 16:19:58 | 94 MB |
| 20220102_Viajes_municipios.csv.gz | 2023-06-29 16:20:45 | 100 MB |
| 20220103_Viajes_municipios.csv.gz | 2023-06-29 16:21:35 | 115 MB |
| 20220104_Viajes_municipios.csv.gz | 2023-06-29 16:22:33 | 115 MB |
| 20220105_Viajes_municipios.csv.gz | 2023-06-29 16:23:33 | 114 MB |
| 20220106_Viajes_municipios.csv.gz | 2023-06-29 16:24:19 | 92 MB |


```
<ttl>60</ttl>
<item>
  <title>
    <![CDATA[ 20250502_distritos_descartados.csv ]]>
  </title>
  <link>
    https://movilidad-opendata.mitma.es/estudios_basicos/calidad/ficheros-diarios/2025-05/20250502_distritos_descartados.csv
  </link>
  <guid isPermaLink="true">
    https://movilidad-opendata.mitma.es/estudios_basicos/calidad/ficheros-diarios/2025-05/20250502_distritos_descartados.csv
  </guid>
  <pubDate>Tue, 06 May 2025 08:16:22 GMT</pubDate>
</item>
<item>
```


Use {spanishoddata} package

```r
library(spanishoddata)
spod_set_data_dir("data")

od_data <- spod_get(
  type = "origin-destination",
  zones = "districts",
  dates = c(
    start = "2022-03-01",
    end = "2022-03-07"
  )
)
```

# ACCESS TO TABULAR DATA

Use {spanishoddata} package

```r
library(spanishoddata)
spod_set_data_dir("data")

od_data <- spod_get(
  type = "origin-destination",
  zones = "districts",
  dates = c(
    start = "2022-03-01",
    end = "2022-03-07"
  )
)
```

```r
library(dplyr)
glimpse(od_data)

Rows: ??
Columns: 20
Database: DuckDB v1.2.1 [root@Darwin 24.4.0:R 4.5.0/
$ date                         <date> 2022-01-04, 202
$ hour                         <int> 0, 0, 0, 1, 1, 3
$ id_origin                    <fct> 01001, 01001, 01
$ id_destination               <fct> 01009_AM, 01009_
$ distance                     <fct> 2-10, 2-10, 2-10
$ activity_origin              <fct> home, frequent_a
$ activity_destination         <fct> frequent_activit
$ study_possible_origin        <lgl> FALSE, FALSE, FA
$ study_possible_destination   <lgl> FALSE, FALSE, FA
$ residence_province_ine_code  <fct> 01, 01, 01, 01,
$ residence_province_name      <fct> "Araba/Álava", "
$ income                       <fct> 10-15, >15, >15,
$ age                          <fct> NA, NA, NA, NA,
$ sex                          <fct> NA, NA, NA, NA,
$ n_trips                      <dbl> 4.894, 1.779, 1.
$ trips_total_length_km        <dbl> 27.966, 5.997, 4
$ year                         <int> 2022, 2022, 2022
$ month                        <int> 1, 1, 1, 1, 1, 1
$ day                          <int> 4, 4, 4, 4, 4, 4
```

# EASE OF USE – FAMILIAR TO R USERS

*Filter and summary*

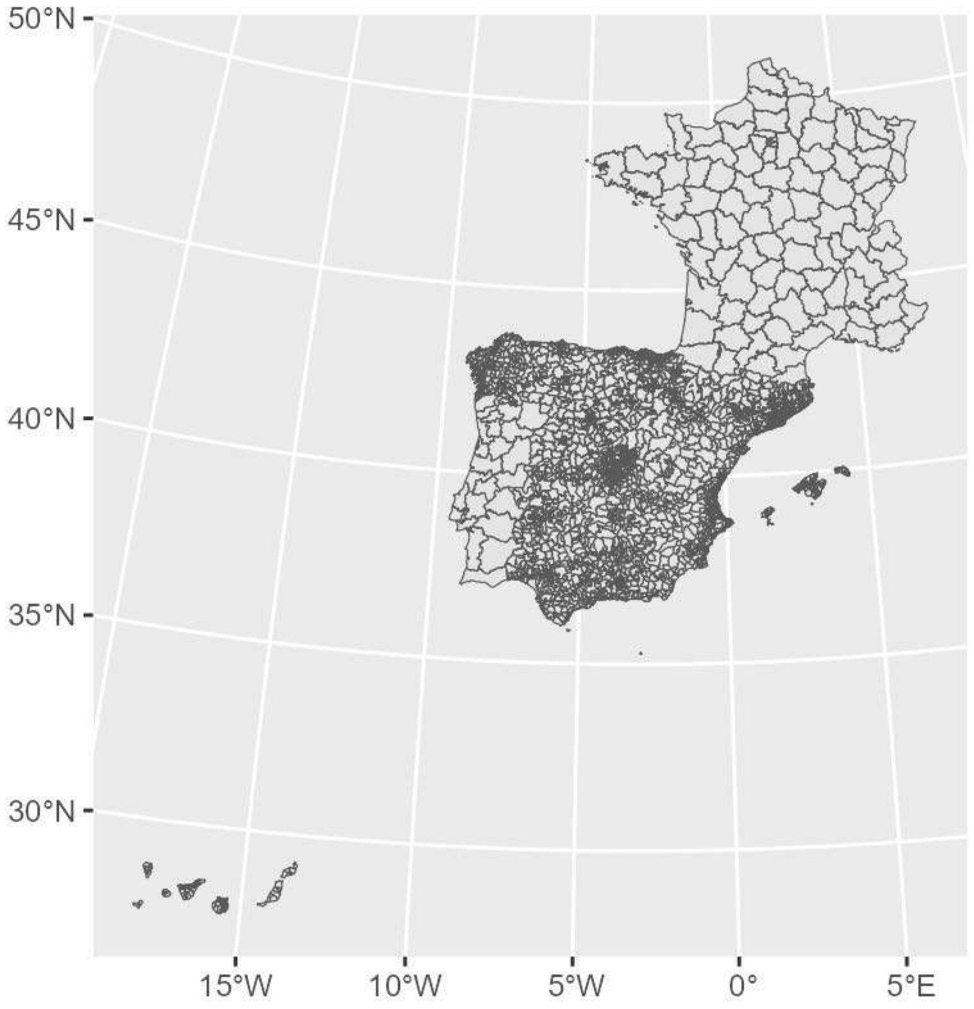
```r
library(dplyr)

od_data |>
  filter(
    year == 2022,
    month %in% c(2, 3, 4)
    ) |>
  summarise(
    n_trips = mean(n_trips)
  ) |>
  collect()
```

*Summary over multiple groups on full data set*

```r
library(dplyr)

od_data |>
  group_by(
    year,
    month,
    day,
    id_origin,
    id_destination
  )
  summarise(
    n_trips = mean(n_trips)
  ) |>
  collect()
```

# ACCESS TO SPATIAL DATA

```r
zones <- spod_get_zones(
  zones = "districts",
  ver = 2 # for 2022 data onwards
)

zones |>
  sf::st_geometry() |>
  ggplot2::ggplot() +
  ggplot2::geom_sf()
```

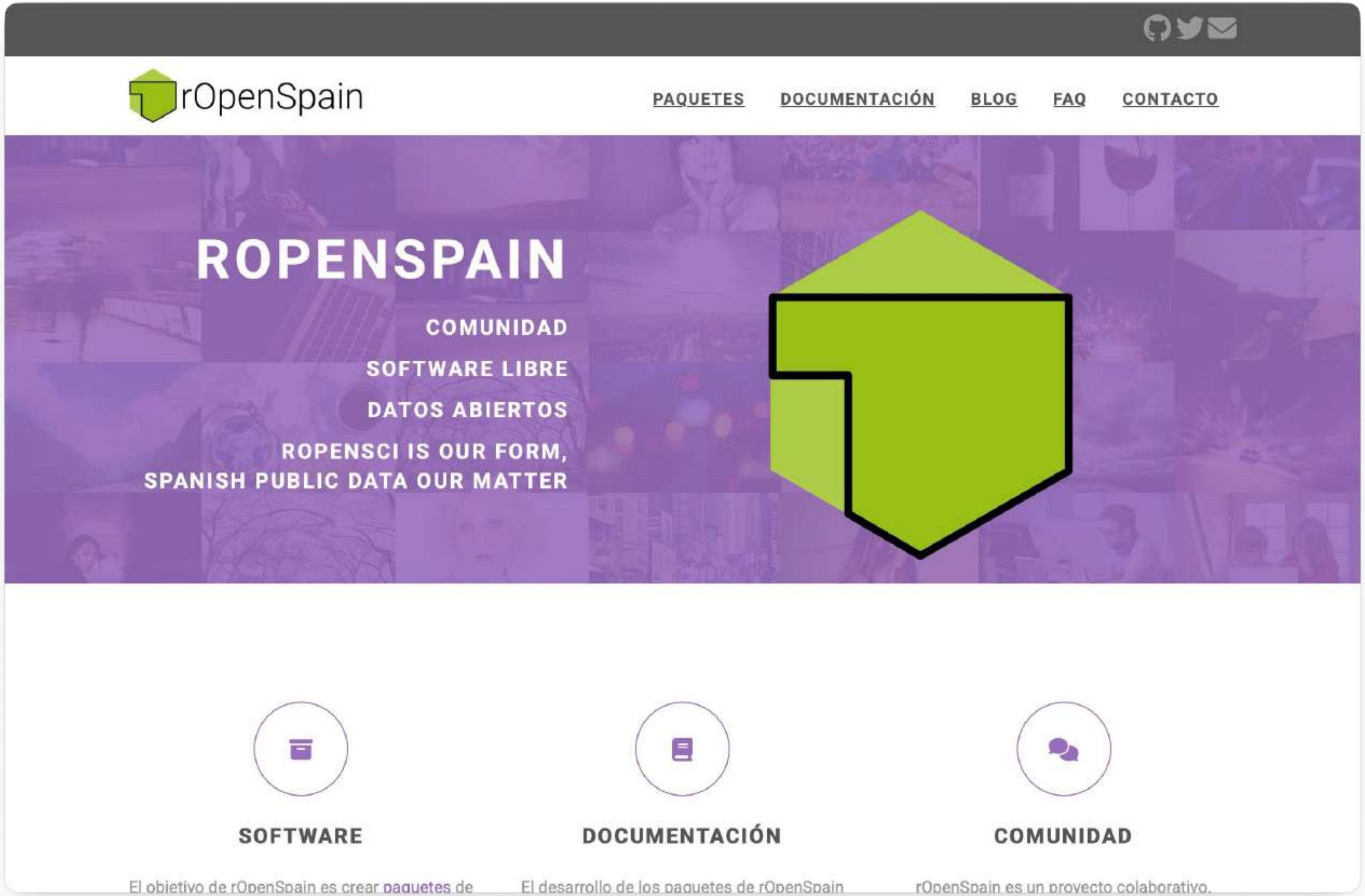


# DATA CITATION

```
spod_cite()
```

```
%% To cite the Ministry's mobility study website
@Misc{mitms_mobility_web,
  title = {Estudio de la movilidad con Big Data (Study of mobility with Big Data)},
  author = {{Ministerio de Transportes y Movilidad Sostenible (MITMS)}},
  year = {2024},
  url = {https://www.transportes.gob.es/ministerio/proyectos-singulares/estudio-de-movilidad-con-
big-data},
}

%% To cite the methodology for 2020-2021 data
@Manual{mitma_methodology_2020_v3,
  title = {Análisis de la movilidad en España con tecnología Big Data durante el estado de alarma
para la gestión de la crisis del COVID-19 (Analysis of mobility in Spain with Big Data technology
during the state of alarm for COVID-19 crisis management)},
  author = {{Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA)}},
  year = {2021},
  url = {https://cdn.mitma.gob.es/portal-web-drupal/covid-19/bigdata/mitma_-
_estudio_movilidad_covid-19_informe_metodologico_v3.pdf},
}
```

# SPANISHODDATA R PACKAGE




https://ropenspain.es

# SPANISHODDATA R PACKAGE – CODEBOOKS

## 2. Mobility data

All mobility data is referenced via `id_origin`, `id_destination`, or other location identifiers (mostly labelled as `id`) with the two sets of zones described above.

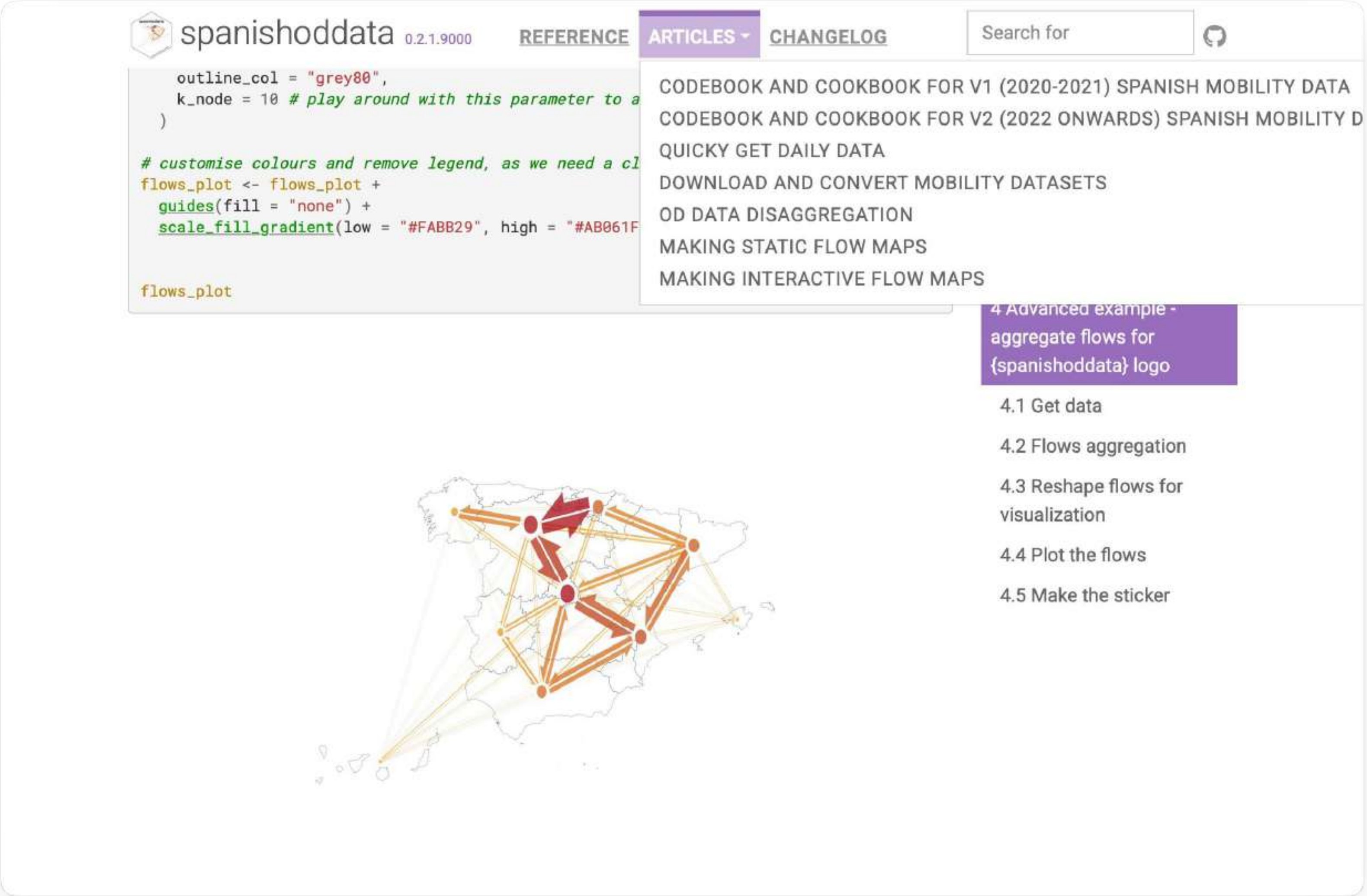
### 2.1. Origin-destination data

The origin-destination data contain the number of trips between `districts`, `municipalities`, or `large urban areas (LUAs)` in Spain for every hour of every day between 2022-02-01 and whichever currently available latest data (2024-06-30 at the time of writing). Each flow also has attributes such as the trip purpose (composed of the type of activity (`home` / `work_or_study` / `frequent_activity` / `infrequent_activity`) at both the origin and destination, but also age, sex, and income of each group of individuals traveling between the origin and destination), province of residence of individuals making this trip, distance covered while making the trip. See the detailed attributes below in a table.

Here are the variables you can find in the `district`, `municipality` and `large urban area` level data:

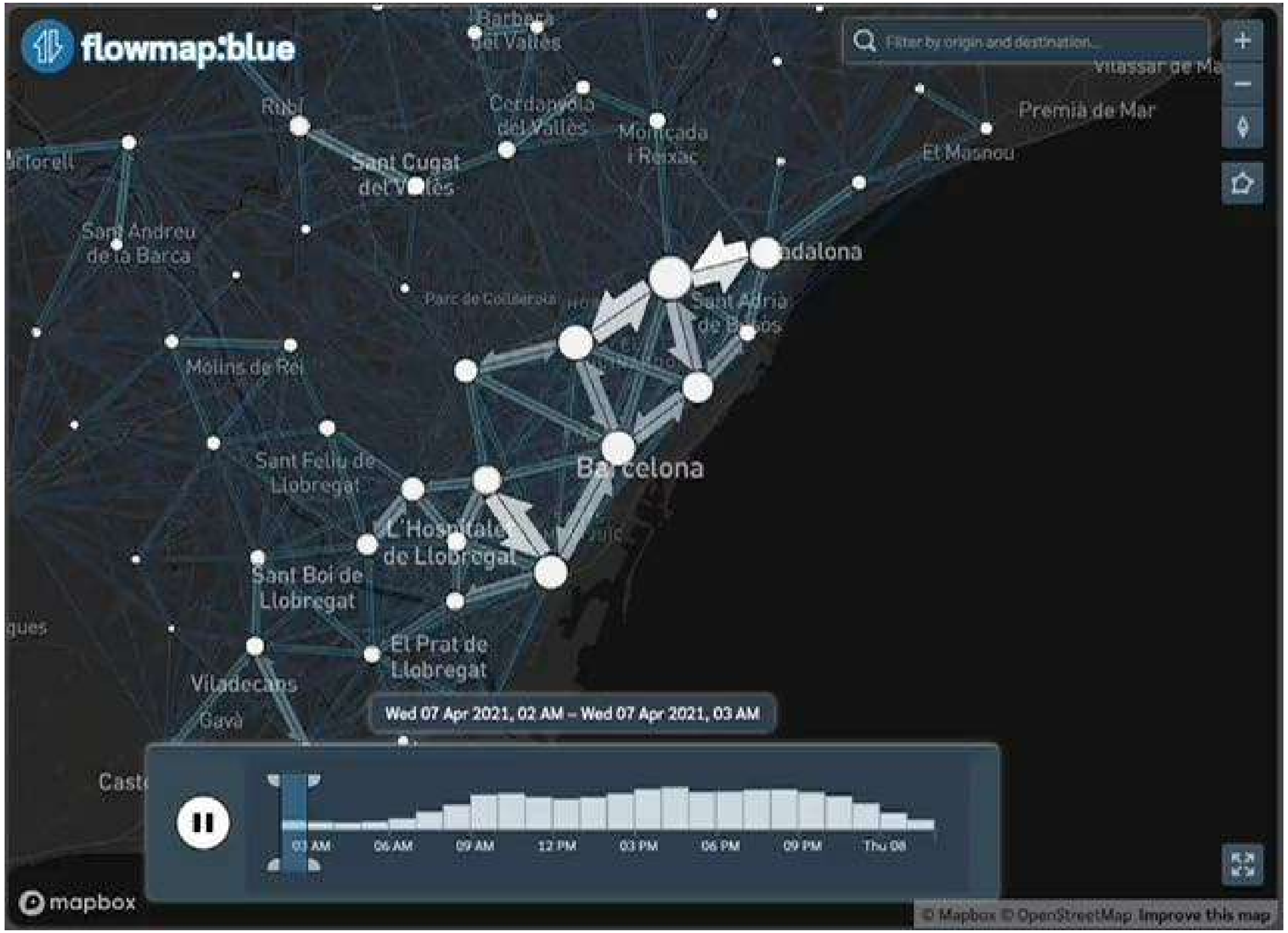
| English Variable Name | Original Variable Name | Type | Description |
|---|---|---|---|
| `date` | `fecha` | `Date` | The date of the recorded data, formatted as `YYYY-DD`. |
| `hour` | `periodo` | `integer` | The time slot durin which the trips occurred. *Note*: this column used to be called `time_slot`, still going to be available in the outp |

# SPANISHODDATA R PACKAGE – USER GUIDES




https://ropenspain.github.io/spanishoddata/ || Logos are the property of their respective owners.

# TUTORIALS TO CREATE INTERACTIVE FLOW MAPS

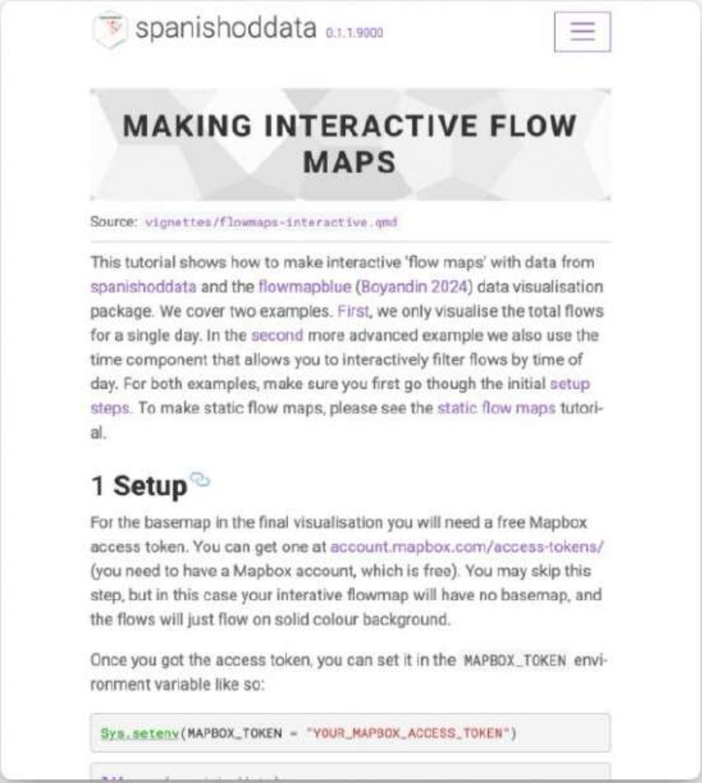


# ECOSYSTEM OF PACKAGES

## A package with many companions

DuckDB

## Detailed documentation



spanishoddata 0.1.1.9000

MAKING INTERACTIVE FLOW MAPS

Source: `vignettes/flowmaps-interactive.qmd`

This tutorial shows how to make interactive 'flow maps' with data from spanishoddata and the flowmapblue (Boyandin 2024) data visualisation package. We cover two examples. First, we only visualise the total flows for a single day. In the second more advanced example we also use the time component that allows you to interactively filter flows by time of day. For both examples, make sure you first go though the initial setup steps. To make static flow maps, please see the static flow maps tutorial.

### 1 Setup

For the basemap in the final visualisation you will need a free Mapbox access token. You can get one at account.mapbox.com/access-tokens/ (you need to have a Mapbox account, which is free). You may skip this step, but in this case your interative flowmap will have no basemap, and the flows will just flow on solid colour background.

Once you got the access token, you can set it in the `MAPBOX_TOKEN` environment variable like so:

```
Sys.setenv(MAPBOX_TOKEN = "YOUR_MAPBOX_ACCESS_TOKEN")
```

*Logos are the property of their respective owners.*

# SPANISHODDATA R PACKAGE – USER MANUAL

spanishoddata 0.2.1.9000 REFERENCE ARTICLES CHANGELOG Search for

## PACKAGE INDEX

### Analysing up to 1 week of data

Quickly download and analyse just a few days of mobility data

`spod_available_data()` stable
Get available data list

`spod_get_zones()` stable
Get zones

`spod_get()` stable
Get tabular mobility data

`spod_disconnect()` stable
Safely disconnect from data and free memory

### Analysing long time periods (months or even years)

Download data for longer periods, convert them into analysis ready format such as `DuckDB` or `Parquet` for out-of-memory analysis of this large data

`spod_available_data()` stable
Get available data list

`spod_get_zones()` stable
Get zones

`spod_download()` stable

**On this page**

- Analysing up to 1 week of data
- Analysing long time periods (months or even years)
- Analysing up to 1 day of trips with no extra variables
- Helper functions

*https://ropenspain.github.io/spanishoddata/ || Logos are the property of their respective owners.*

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

spanishoddata

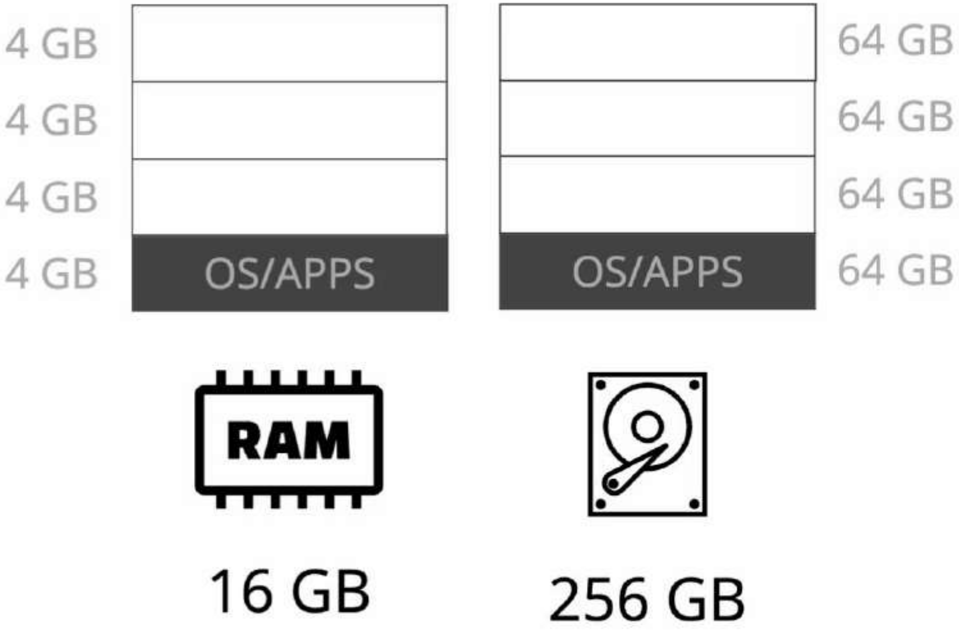
github.com/rOpenSpain/spanishoddata

DuckDB

*Image: DALLE-3; Logos are the property of their respective owners.*

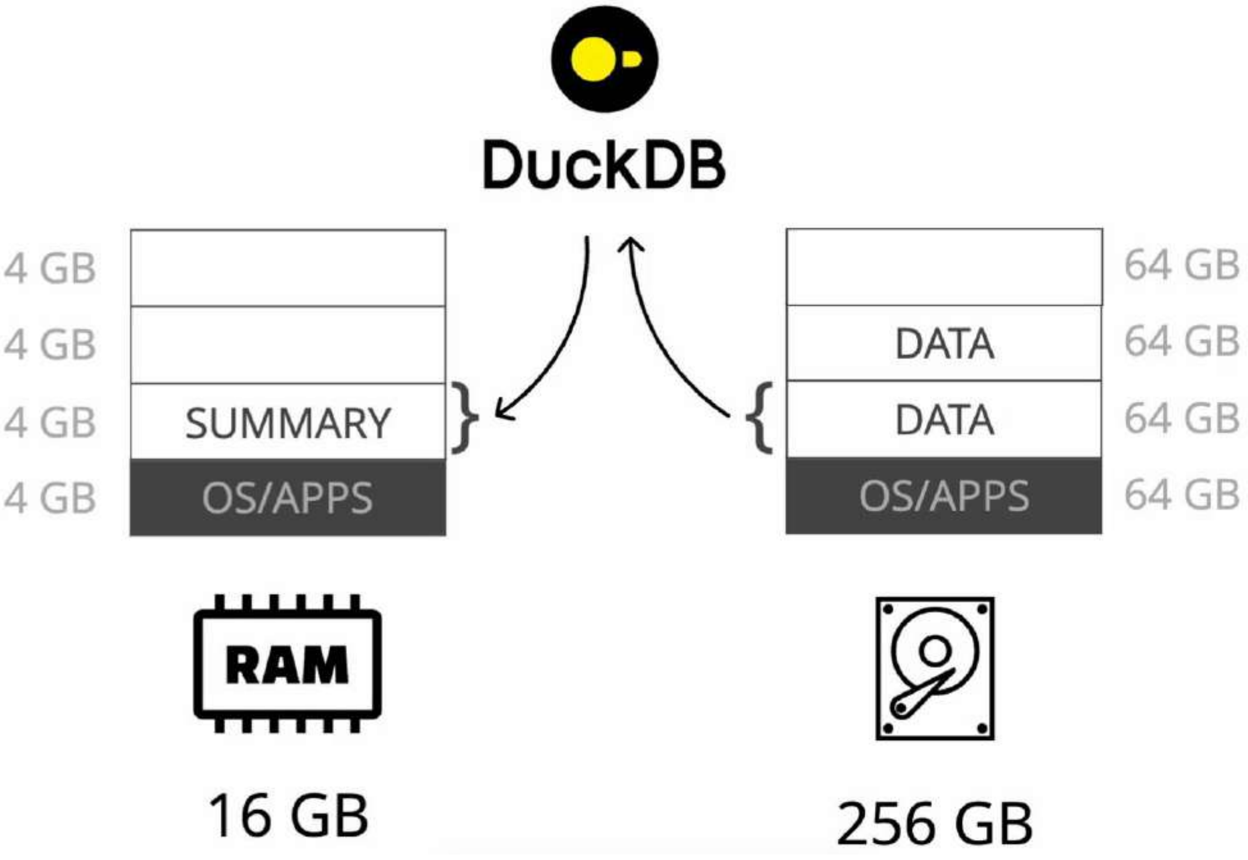
# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

DuckDB



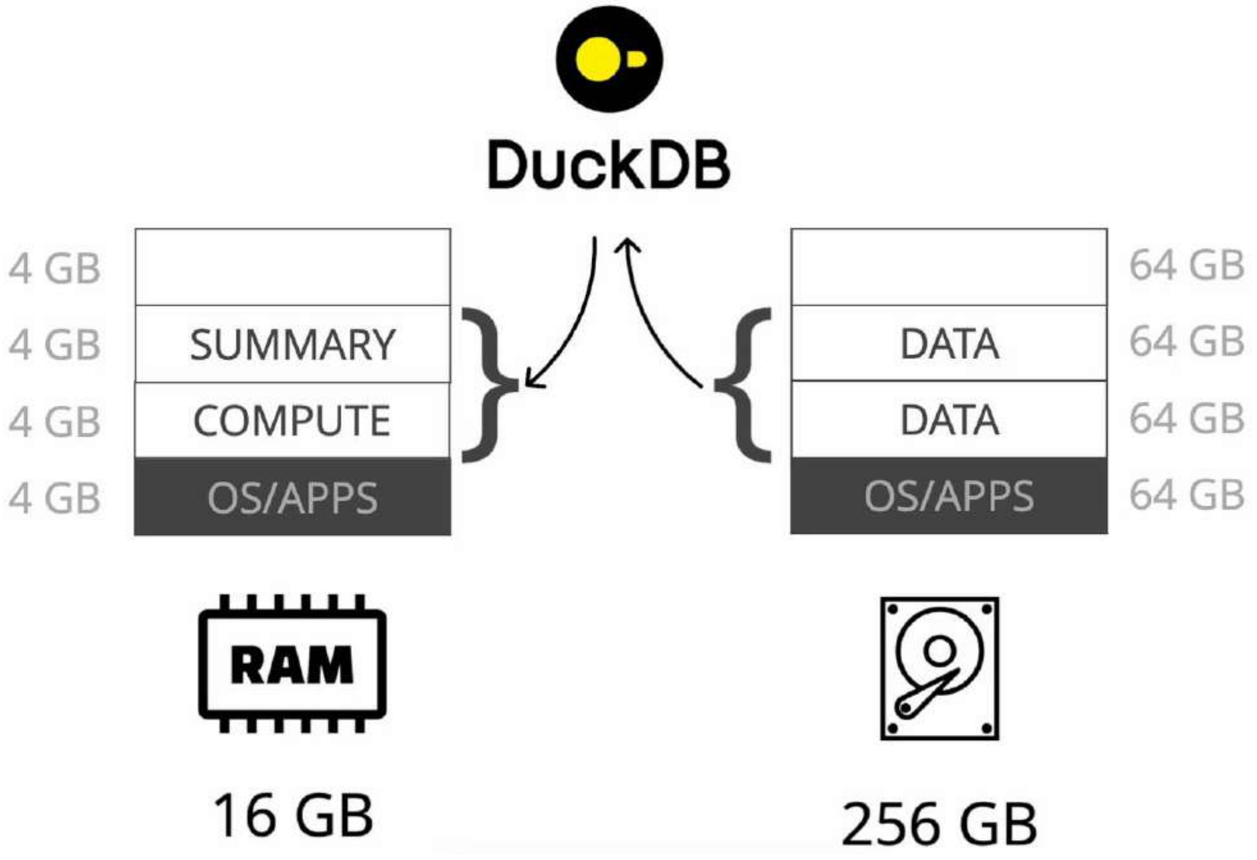

*Logos are the property of their respective owners.*

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB



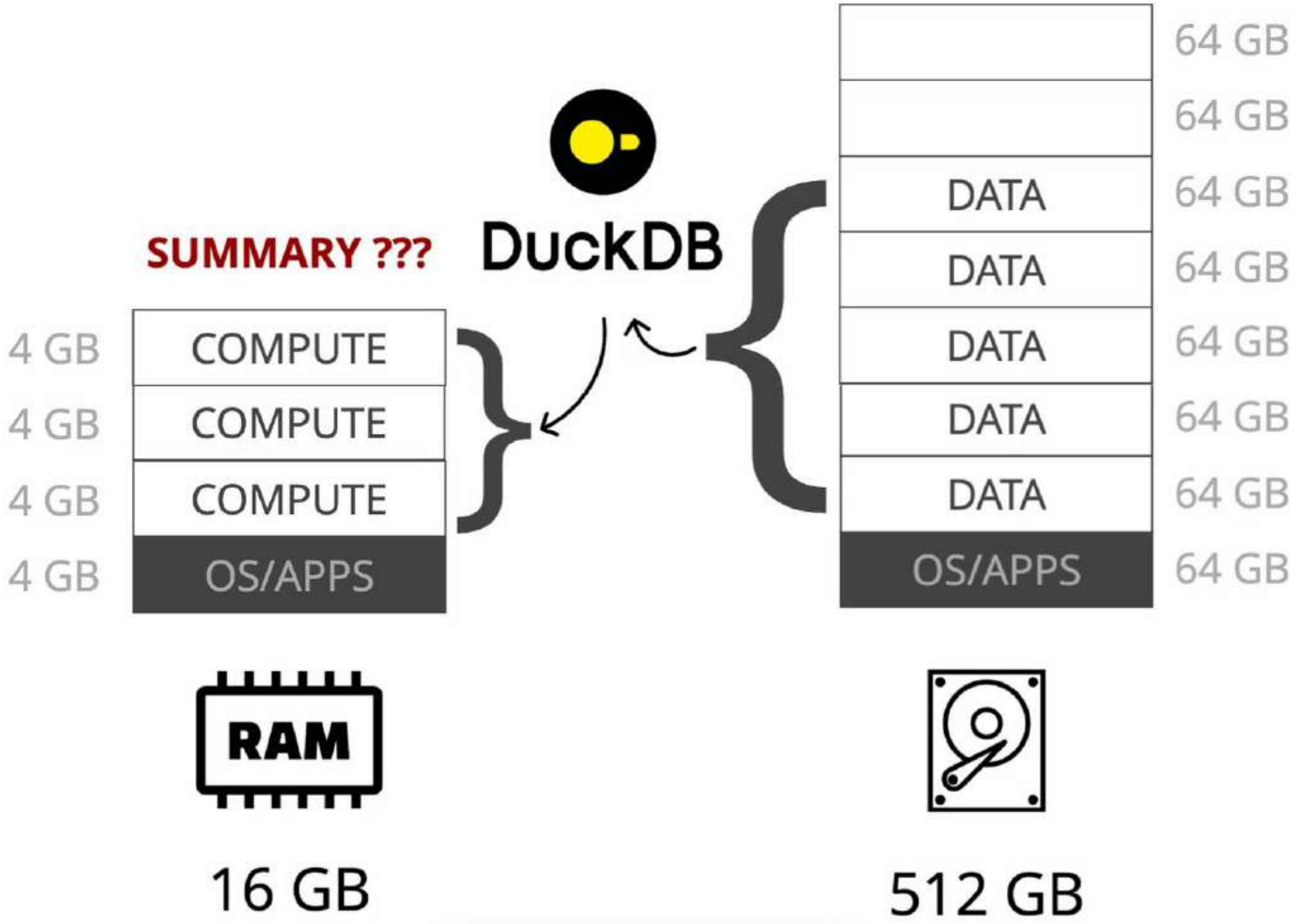

Logos are the property of their respective owners.

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB




*Logos are the property of their respective owners.*

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB




*Logos are the property of their respective owners.*

# EASE OF USE – FAMILIAR TO R USERS

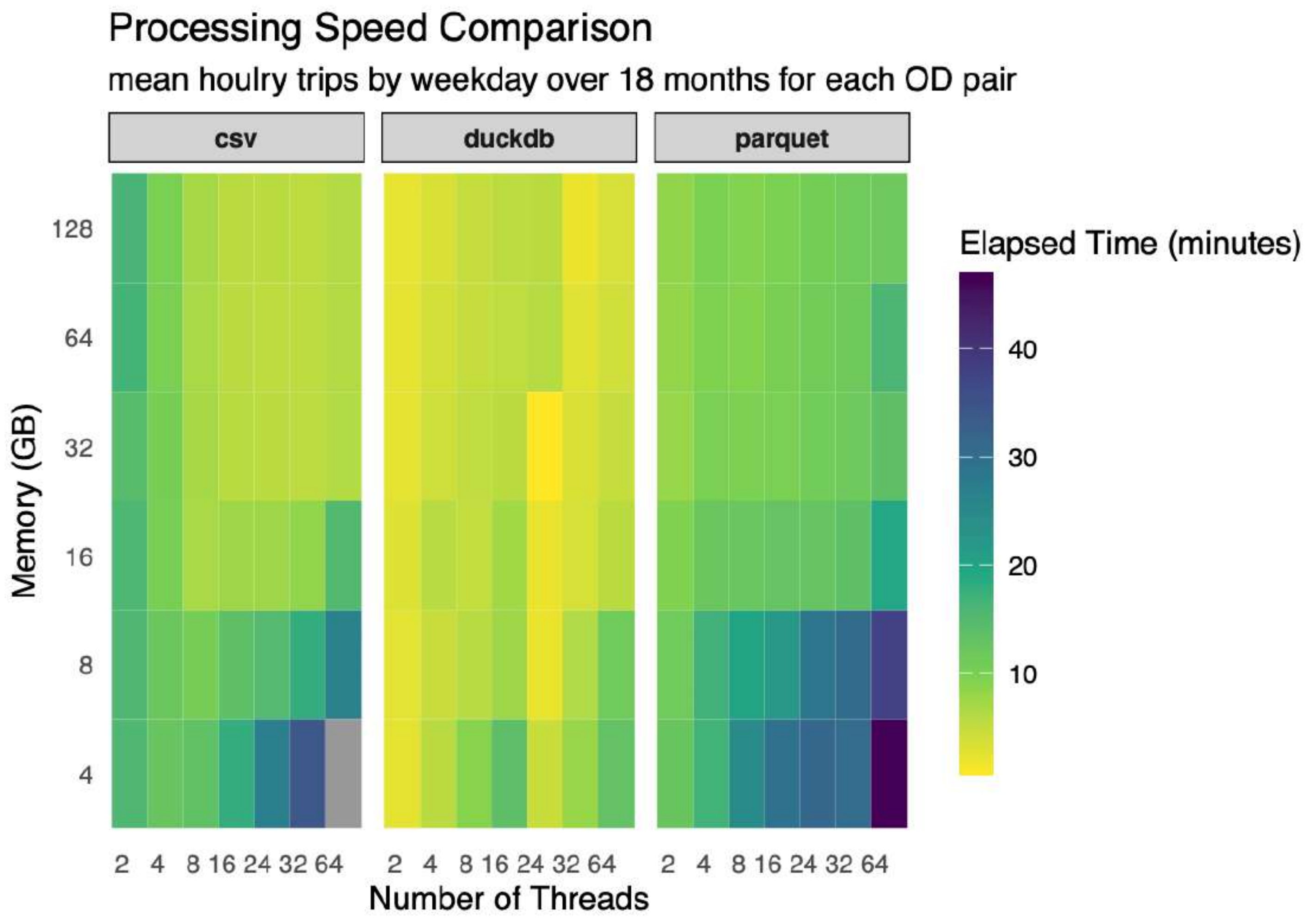
*Filter and summary*

```r
library(dplyr)

od_data |>
  filter(
    year == 2022,
    month %in% c(2, 3, 4)
    ) |>
  summarise(
    n_trips = mean(n_trips)
  ) |>
  collect()
```

*Summary over multiple groups on full data set*

```r
library(dplyr)

od_data |>
  group_by(
    year,
    month,
    day,
    id_origin,
    id_destination
  )
  summarise(
    n_trips = mean(n_trips)
  ) |>
  collect()
```

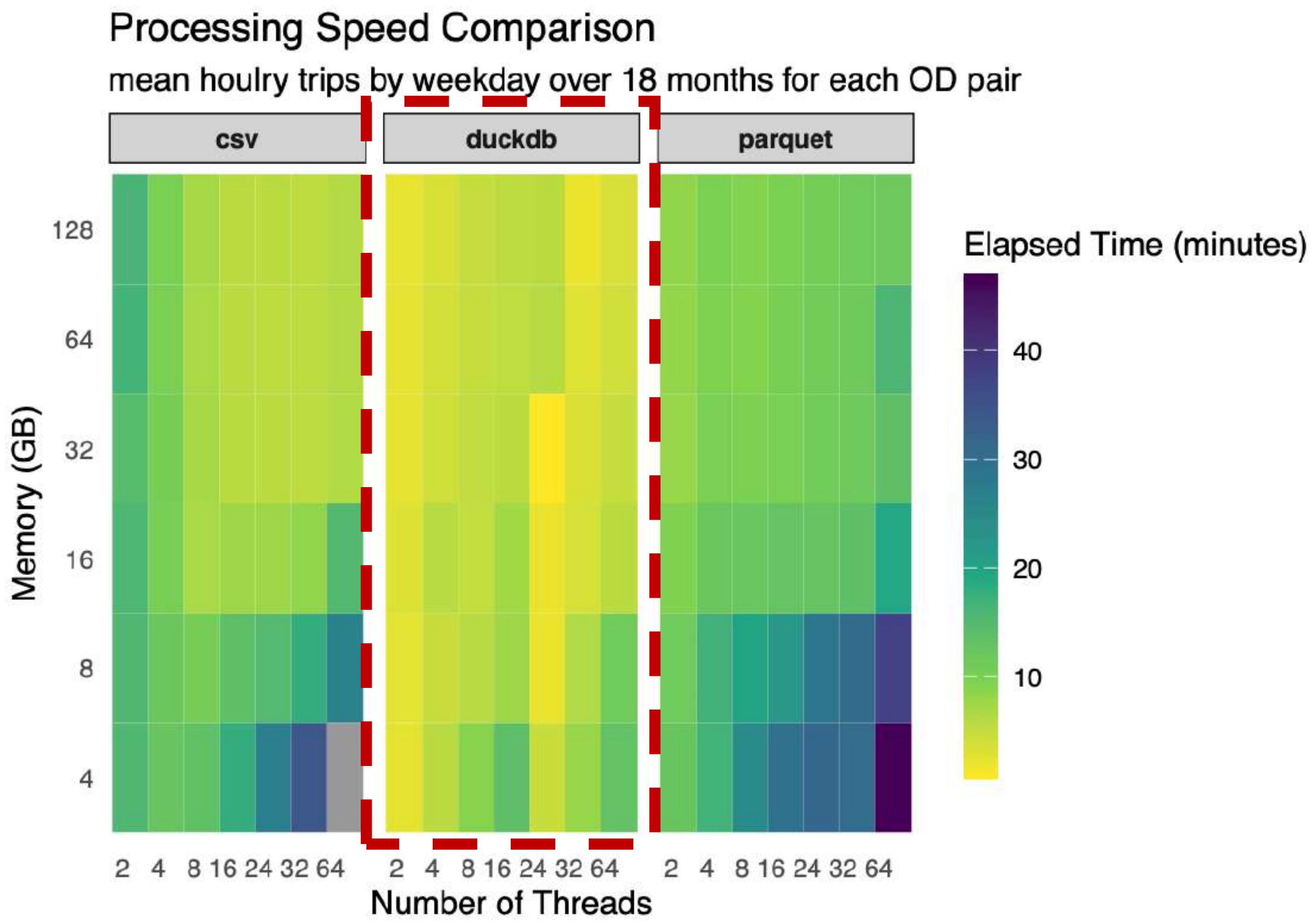
# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB




Data processing speed comparison: DuckDB engine running on CSV.gz files vs DuckDB database vs a folder of Parquet files

https://github.com/e-kotov/spanishoddata-paper-supplement/blob/main/plots/supplement-plots/12_analysis-speed.pdf

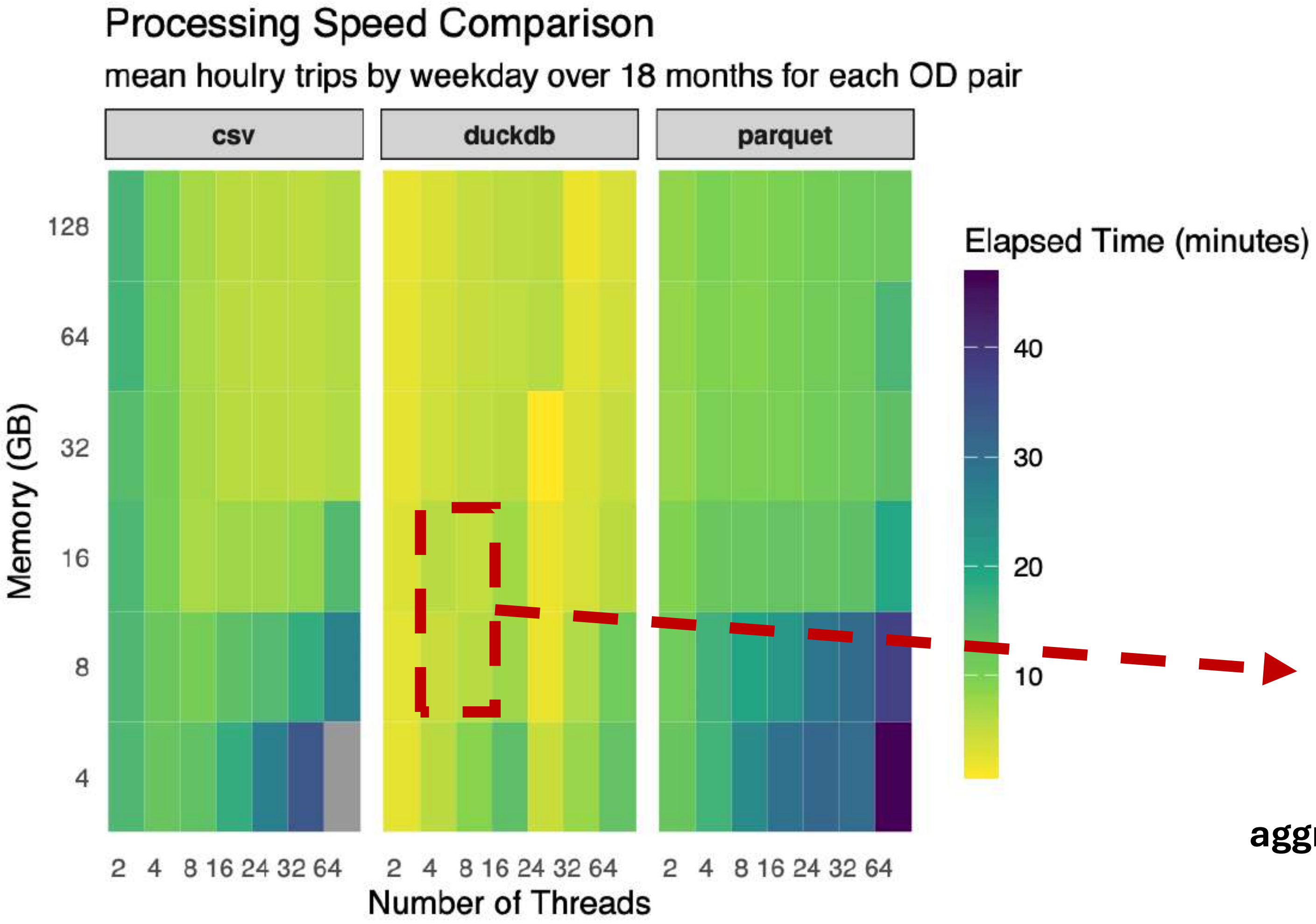
# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB




Data processing speed comparison: DuckDB engine running on CSV.gz files vs DuckDB database vs a folder of Parquet files

*https://github.com/e-kotov/spanishoddata-paper-supplement/blob/main/plots/supplement-plots/12_analysis-speed.pdf*

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

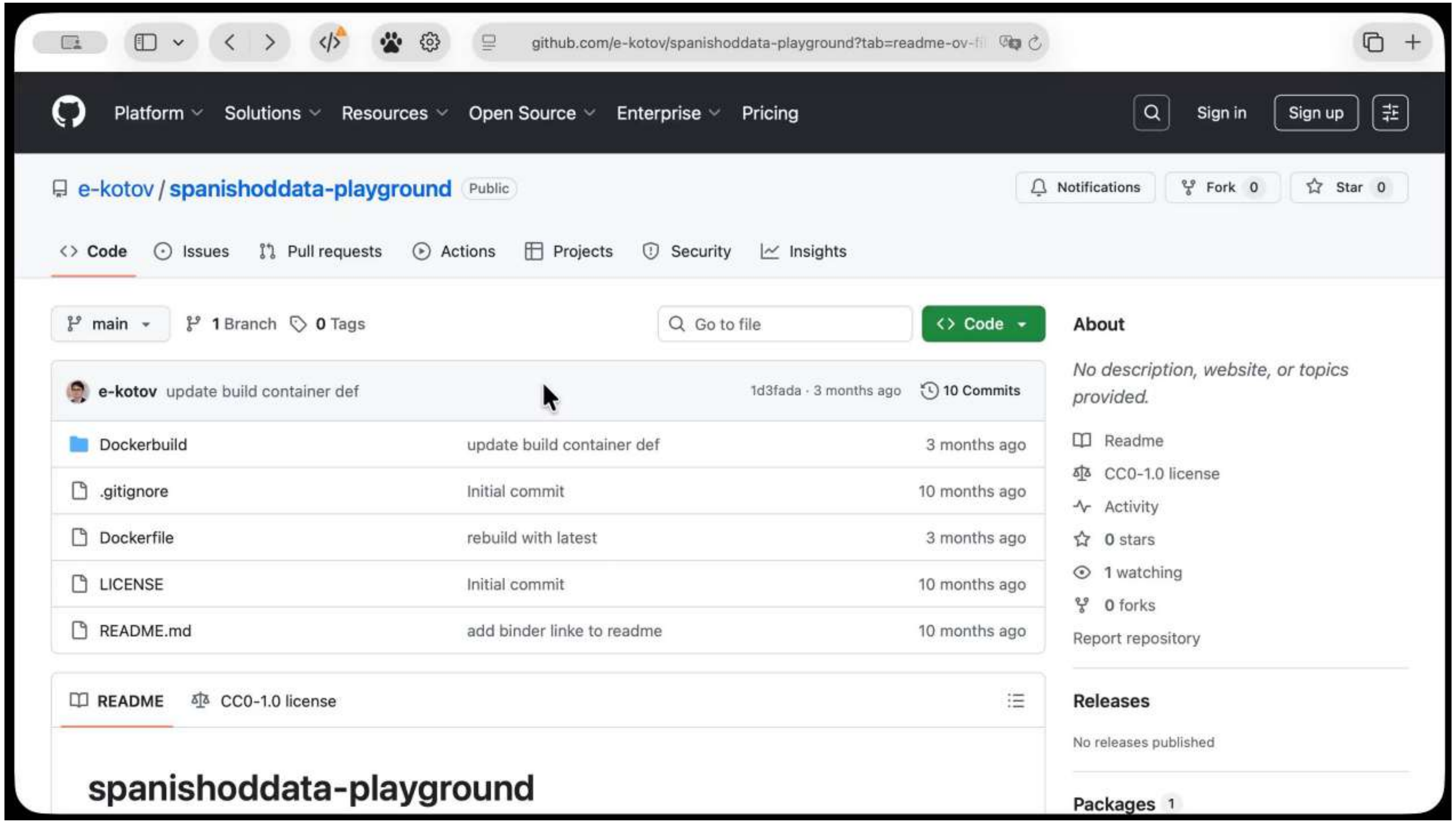



Data processing speed comparison: DuckDB engine running on CSV.gz files vs DuckDB database vs a folder of Parquet files

**3-4 year old laptop aggregates 18 months of data in ~ 10 minutes**

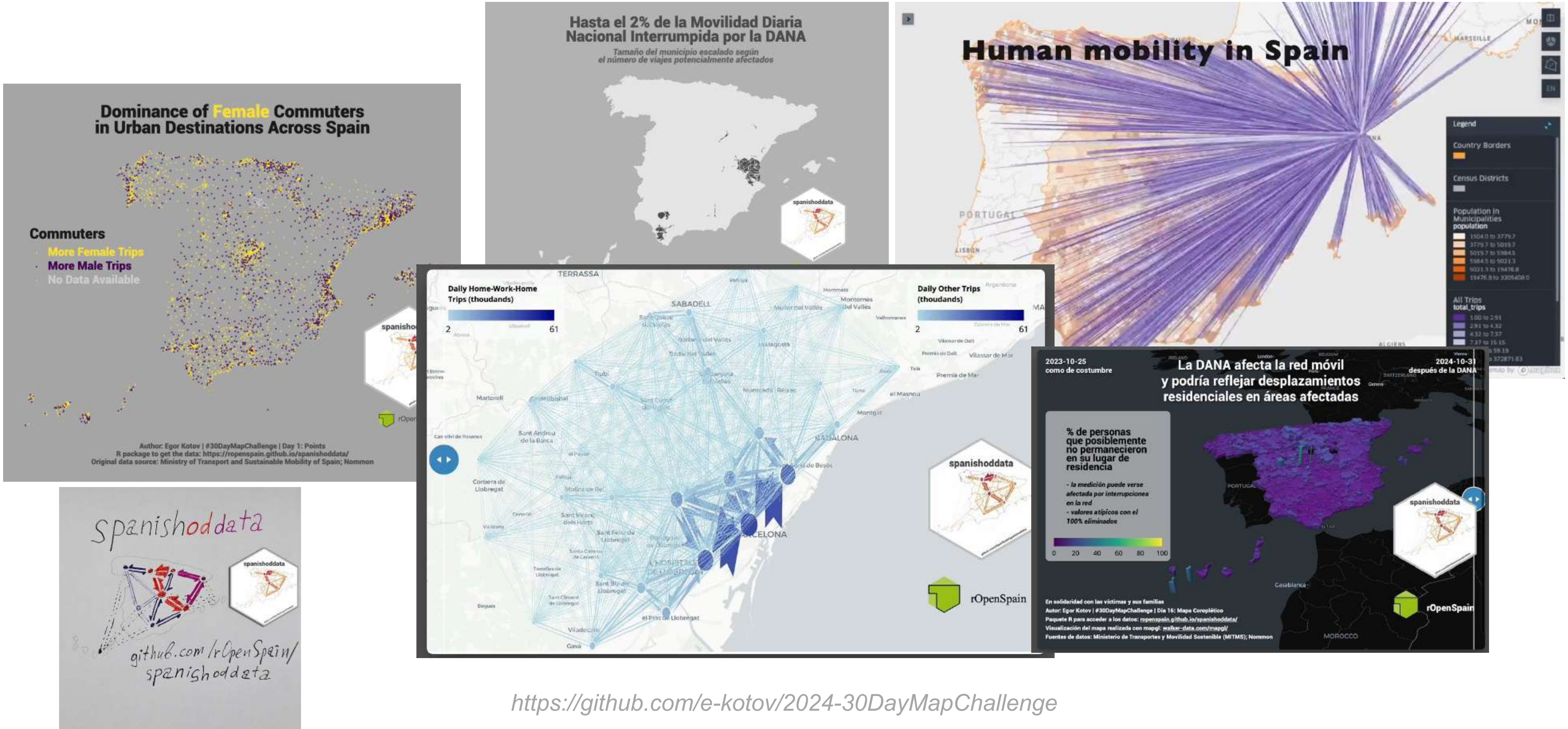
https://github.com/e-kotov/spanishoddata-paper-supplement/blob/main/plots/supplement-plots/12_analysis-speed.pdf

# DEMO




```
library(dplyr)
library(spanishoddata)
spod_set_data_dir("data")
x <- spod_get(
  "od",
  zones = "districts",
  dates = "2022-01-01")
x %>%
  group_by(age, sex) %>%
  summarize(
  trips = sum(n_trips,
  na.rm = TRUE))
```

# #30DayMapChallenge 2024



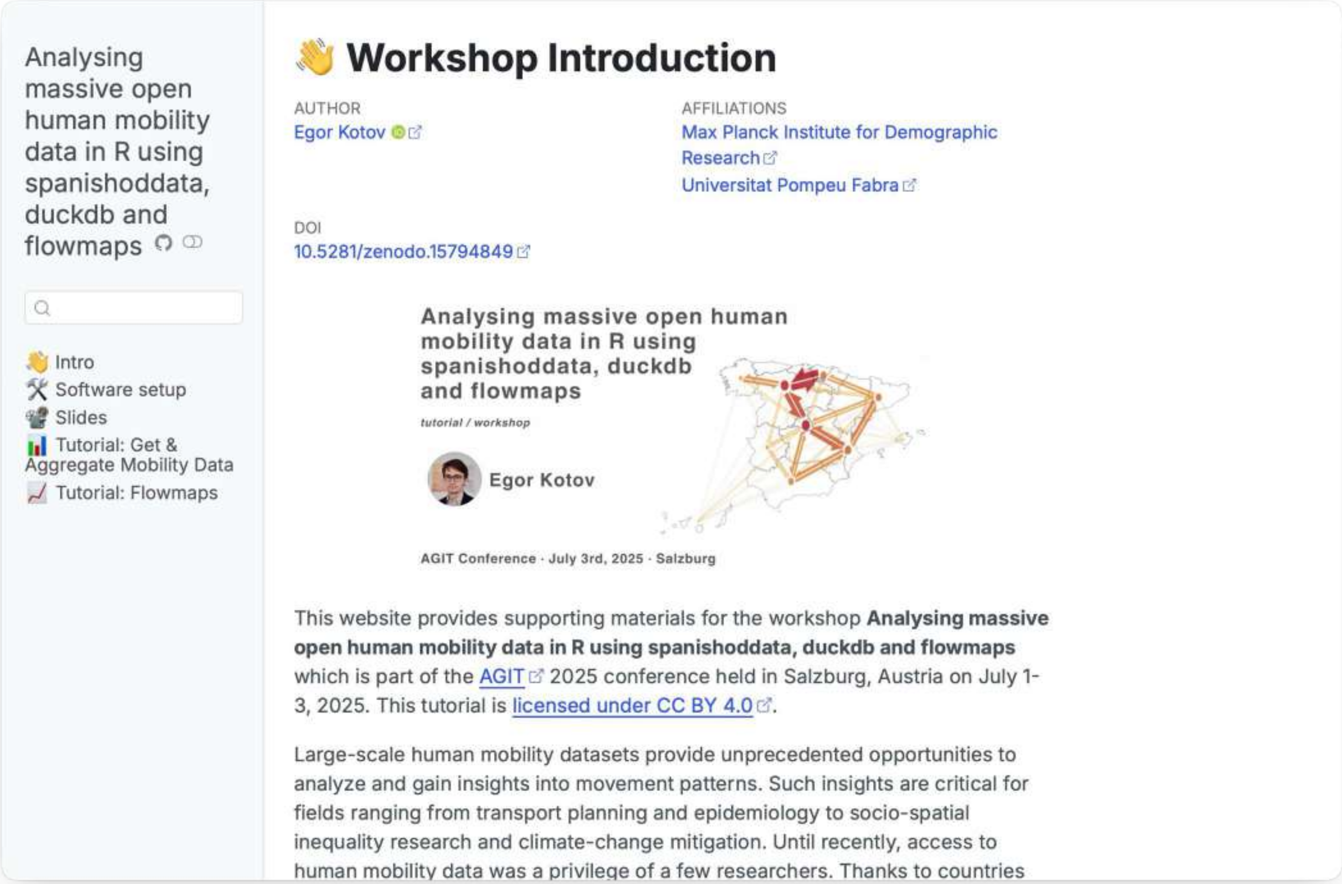

https://github.com/e-kotov/2024-30DayMapChallenge

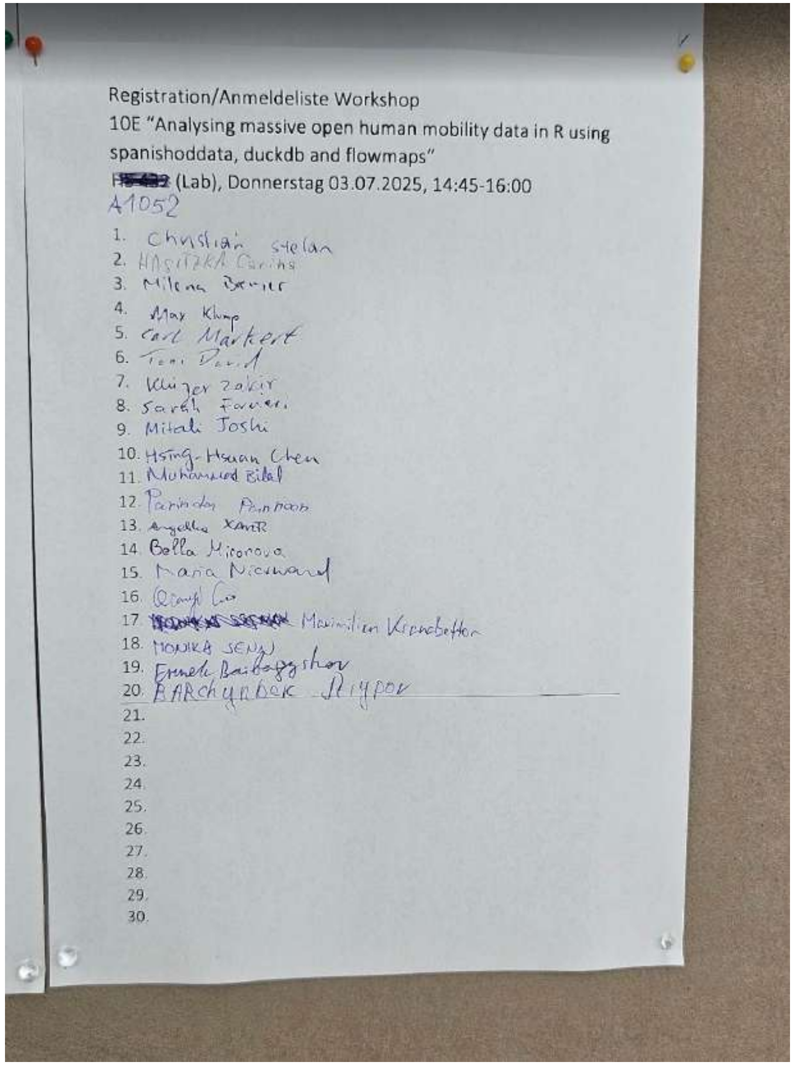
# SUMMER 2025 WORKSHOPS



https://www.ekotov.pro/agit-2025-spanishoddata/

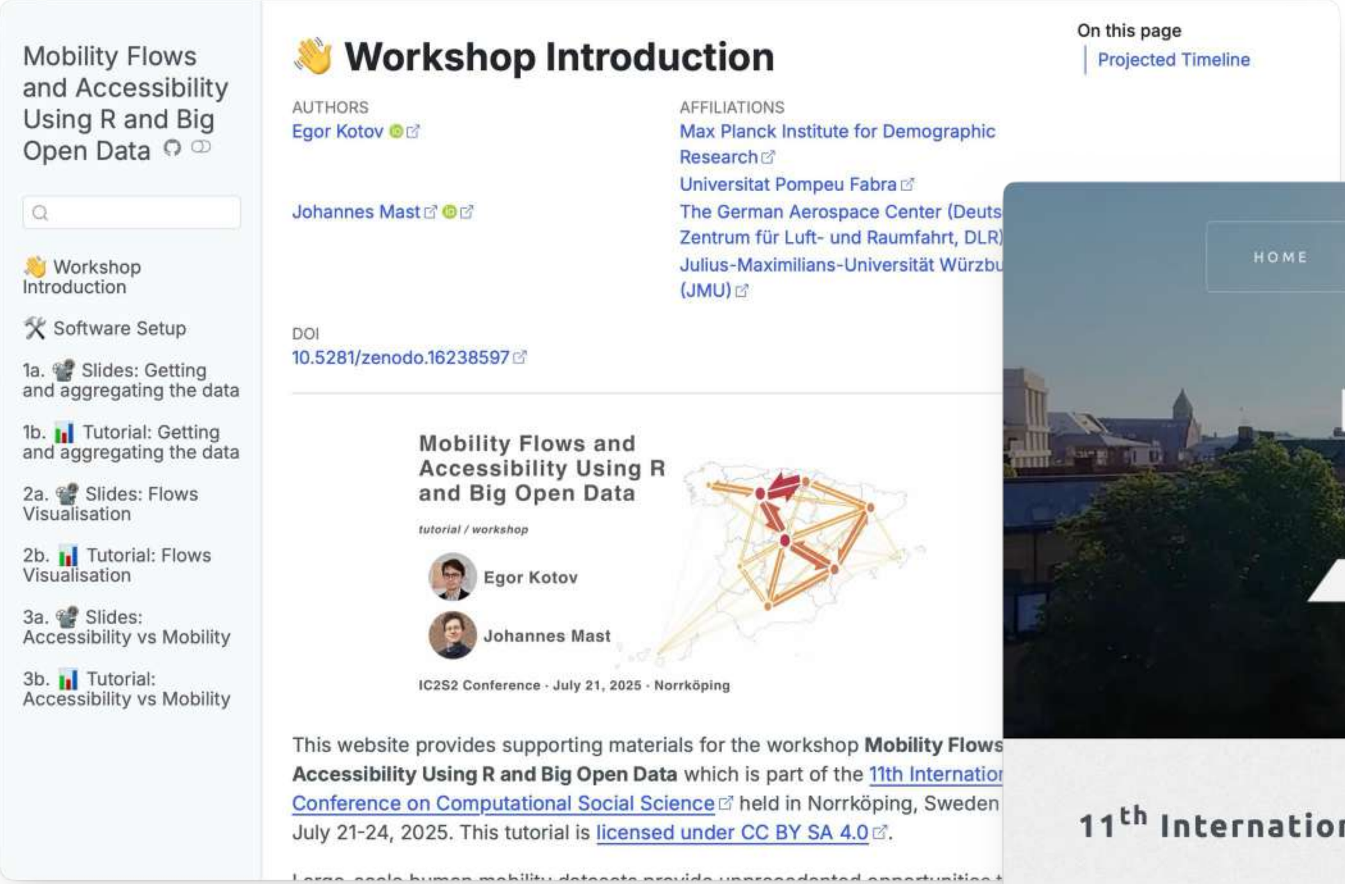
# SUMMER 2025 WORKSHOPS




https://www.ekotov.pro/agit-2025-spanishoddata/

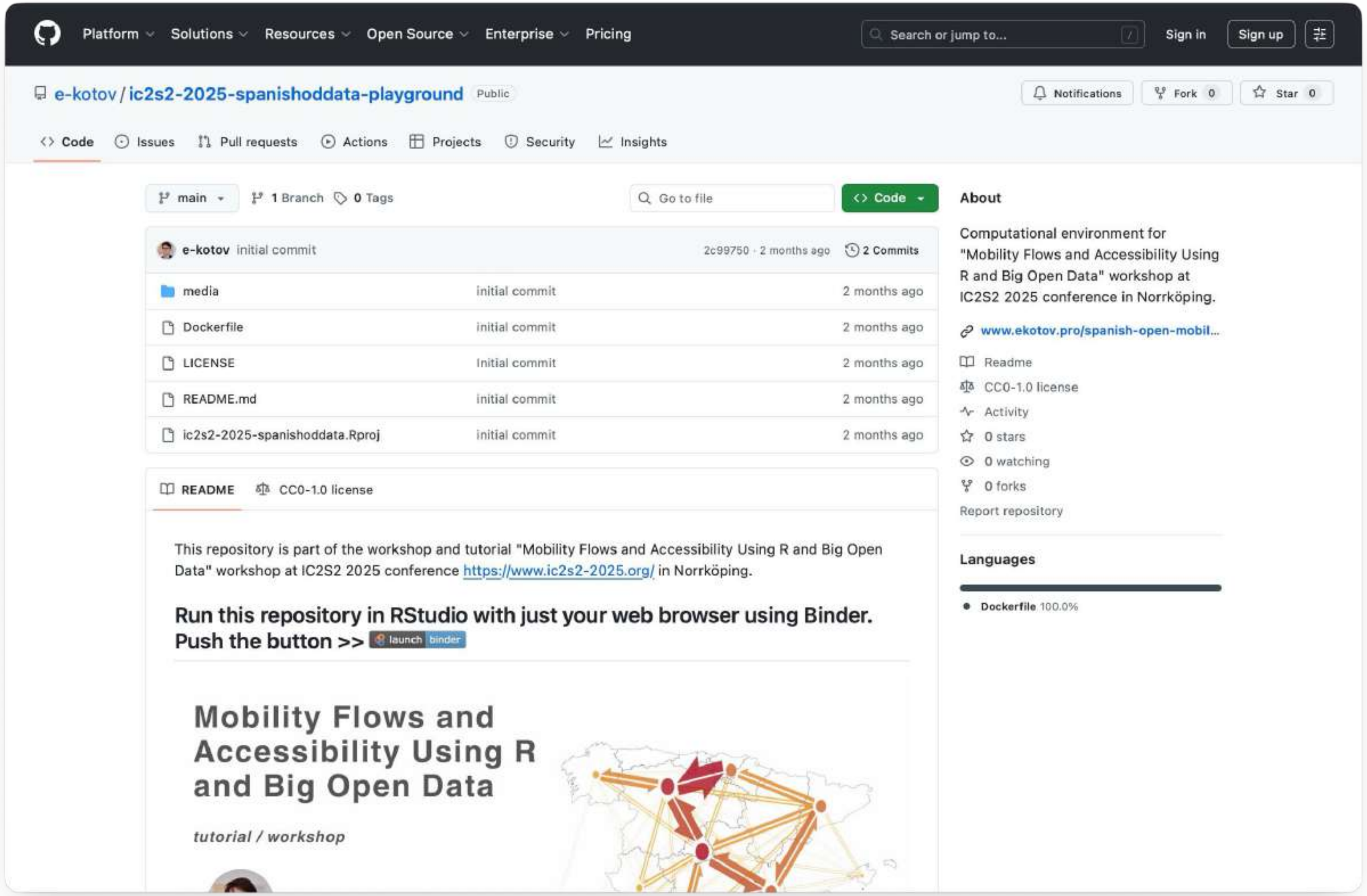
# SUMMER 2025 WORKSHOPS





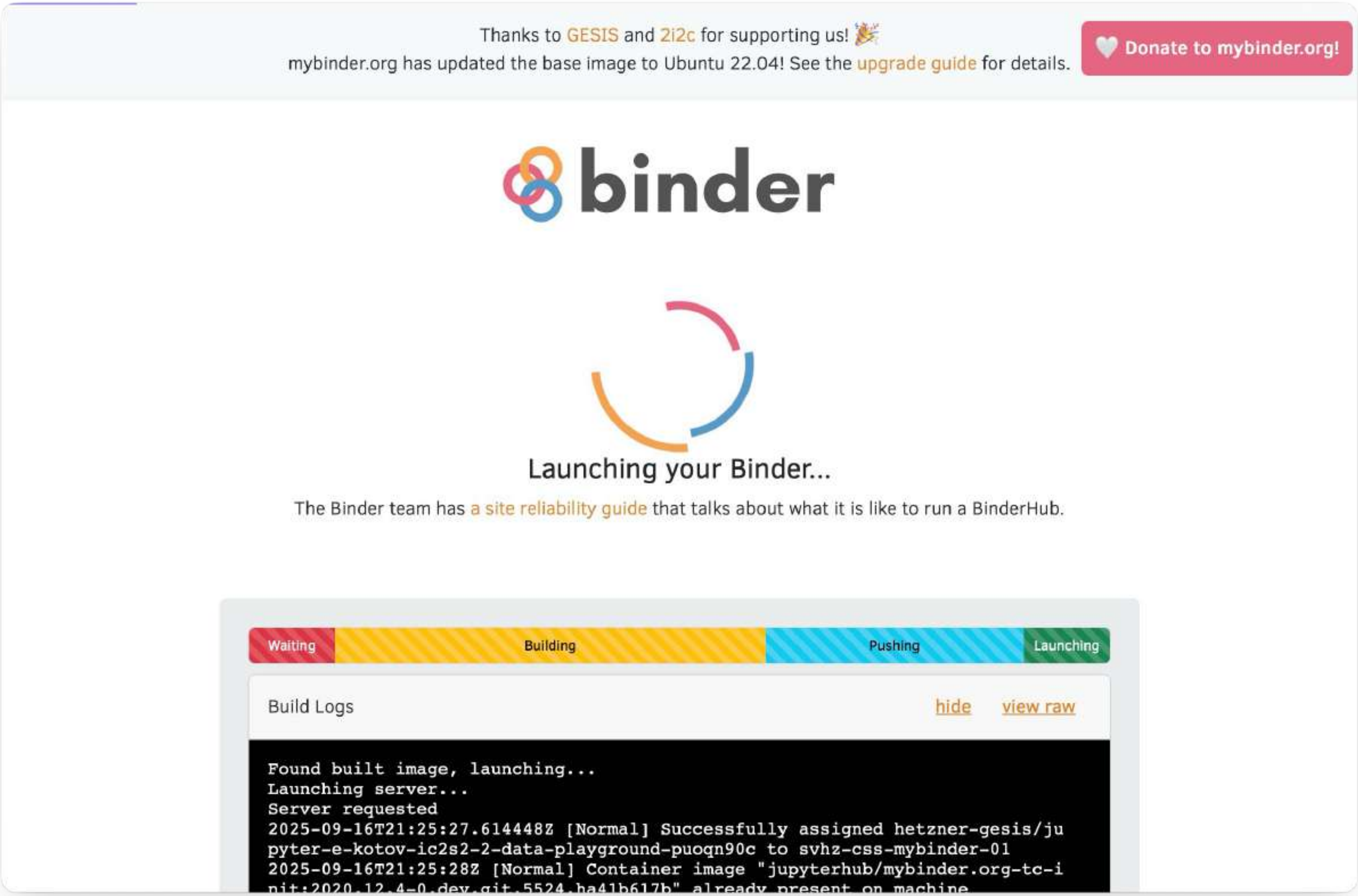
https://www.ekotov.pro/spanish-open-mobility-workshop-ic2s2-2025/

# INTERACTIVE PLAYGROUND IN WEB BROWSER



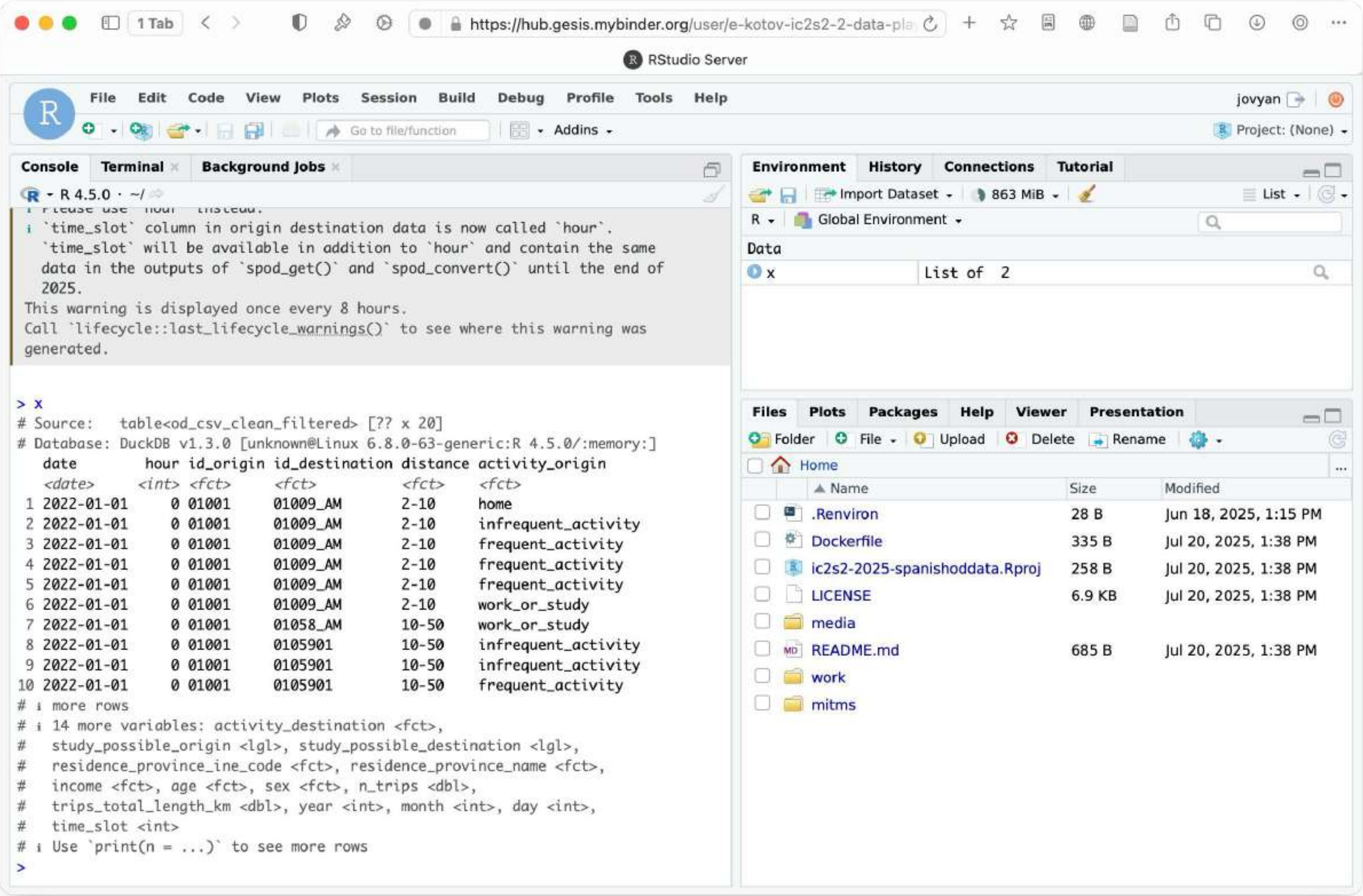

https://github.com/e-kotov/ic2s2-2025-spanishoddata-playground

# INTERACTIVE PLAYGROUND IN WEB BROWSER



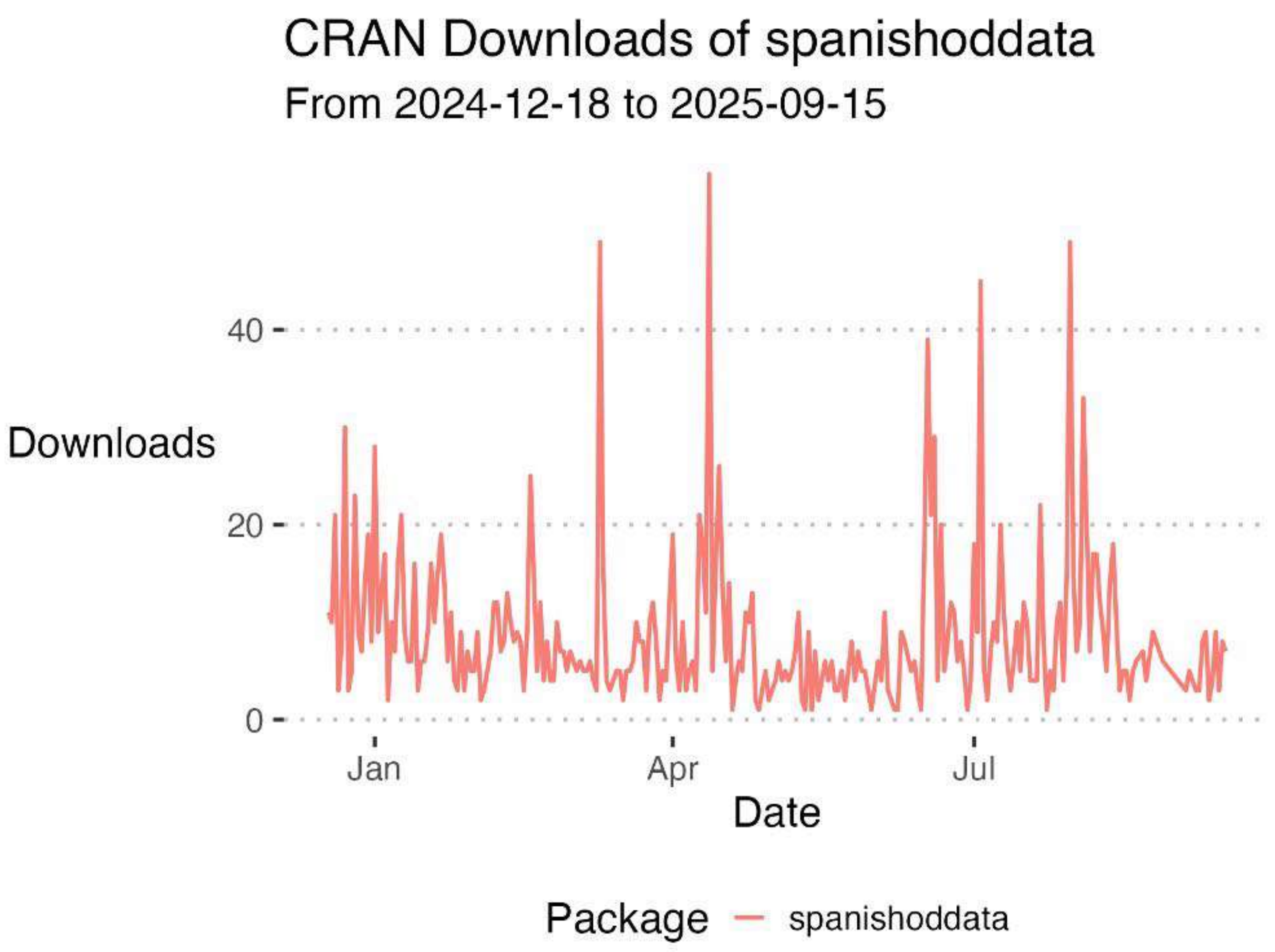

https://github.com/e-kotov/ic2s2-2025-spanishoddata-playground

# INTERACTIVE PLAYGROUND IN WEB BROWSER




*https://github.com/e-kotov/ic2s2-2025-spanishoddata-playground*

# 2,200+ DOWNLOADS IN THE PAST 9 MONTHS





CRAN 0.2.1

downloads 2282

downloads 107/month

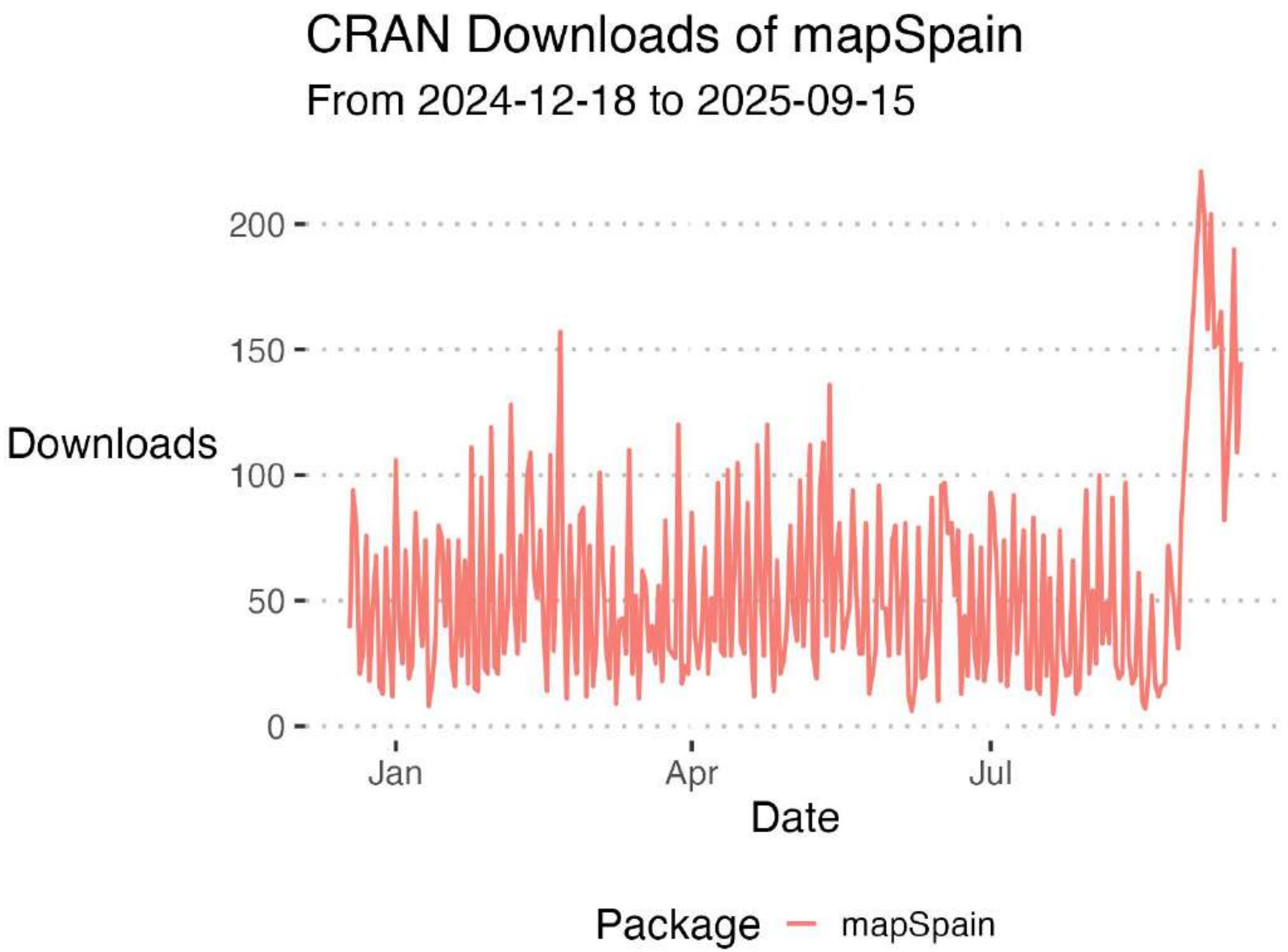
R-CMD-check passing


DOI 10.32614/CRAN.package.spanishoddata

DOI 10.5281/zenodo.14516104


*https://github.com/e-kotov/spanishoddata-paper-supplement/blob/main/plots/supplement-plots/12_analysis-speed.pdf*

# MAPSPAIN

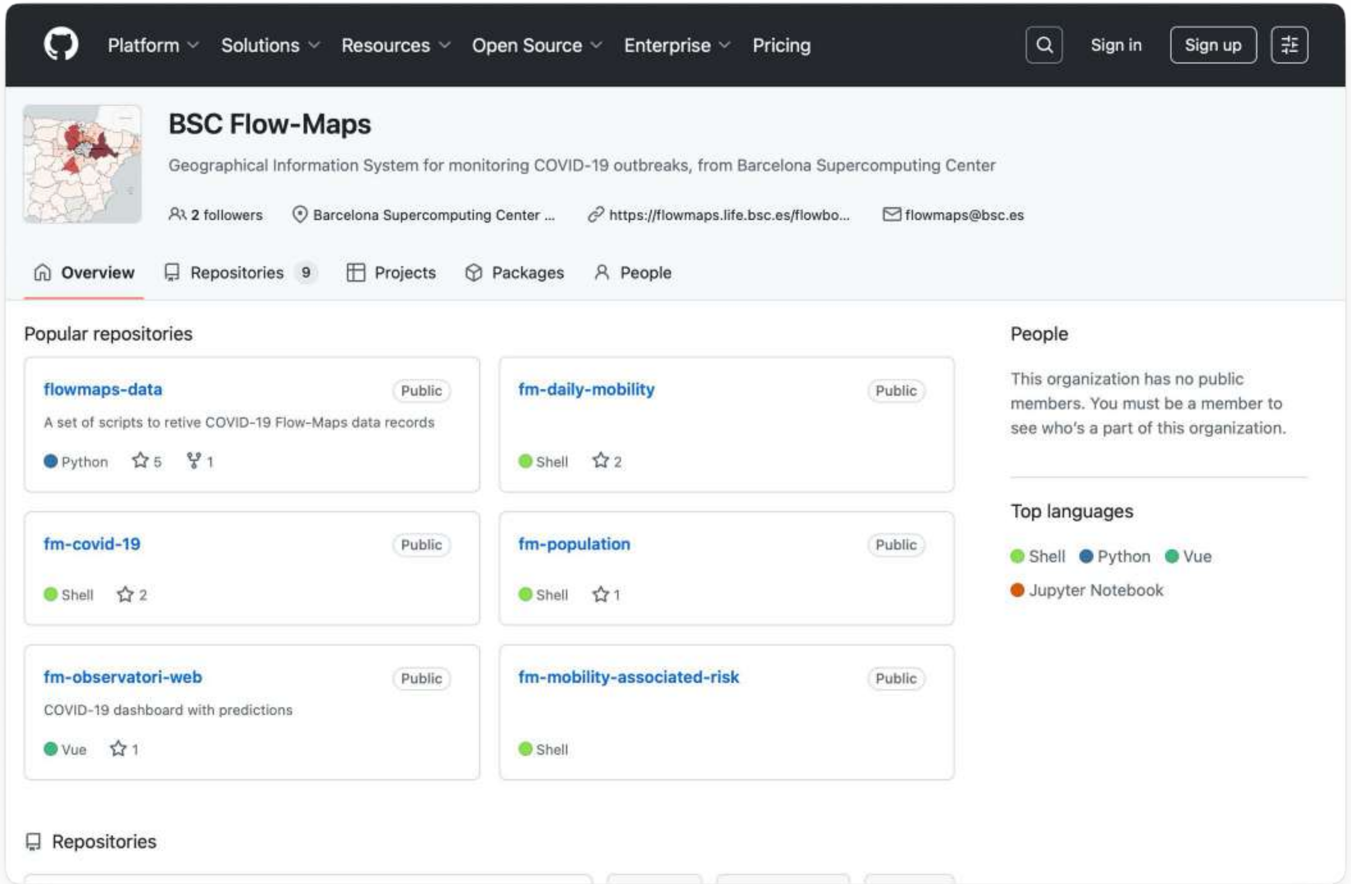




- **mapSpain** is an interface to Spanish spatial data such as province, municipality and census boundaries.
- With 14,321 downloads, is only 6 times more popular than much more specialized **spanishoddata** (2,200+ downloads) in the same period.

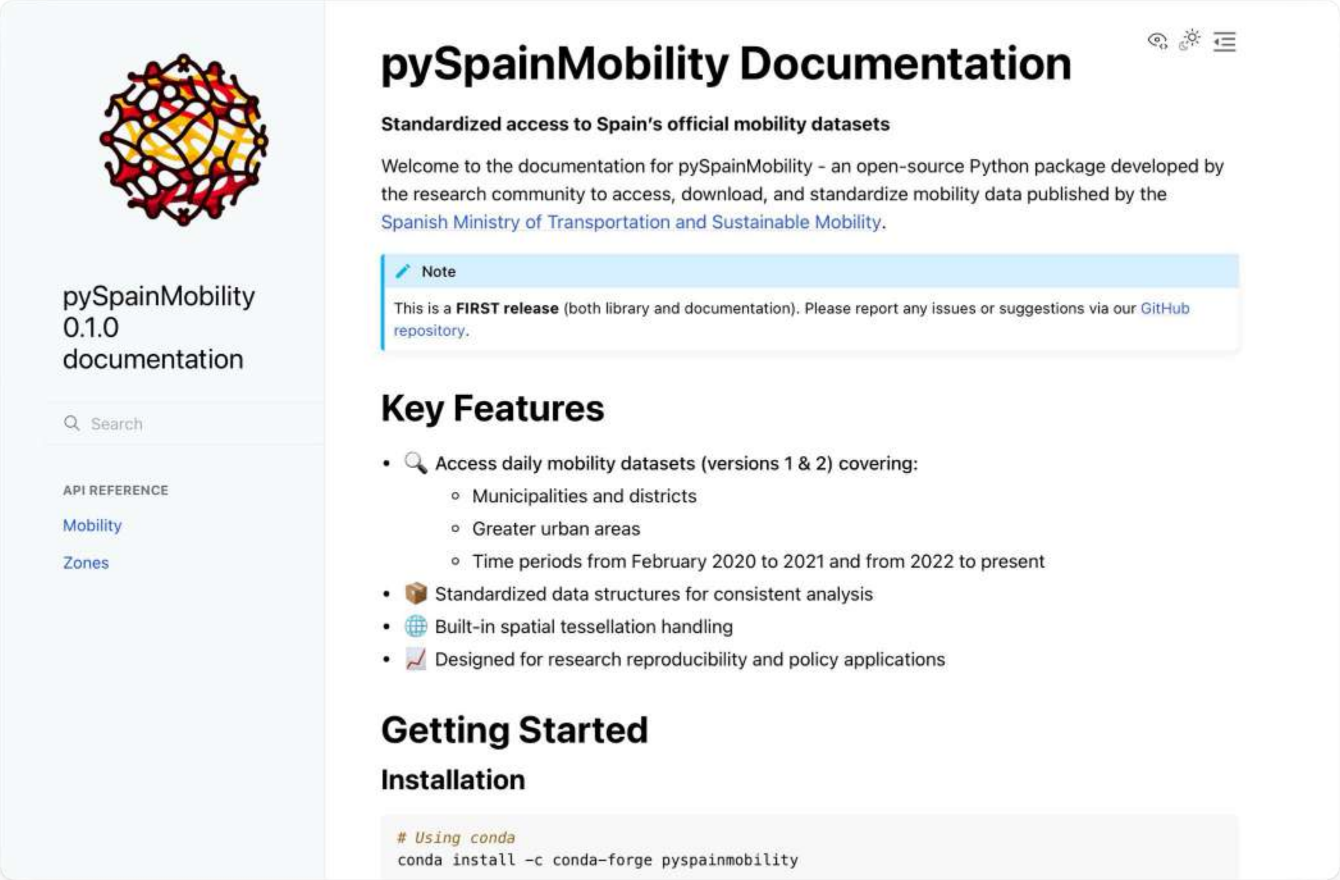
*https://github.com/e-kotov/spanishoddata-paper-supplement/blob/main/plots/supplement-plots/12_analysis-speed.pdf*

# BSC FLOW-MAPS




*https://github.com/bsc-flowmaps*

# PYSPAINMOBILITY



pySpainMobility 0.1.0 documentation

Search

API REFERENCE

Mobility

Zones

# pySpainMobility Documentation

**Standardized access to Spain's official mobility datasets**

Welcome to the documentation for pySpainMobility - an open-source Python package developed by the research community to access, download, and standardize mobility data published by the Spanish Ministry of Transportation and Sustainable Mobility.

Note

This is a **FIRST release** (both library and documentation). Please report any issues or suggestions via our GitHub repository.

## Key Features

- 🔍 Access daily mobility datasets (versions 1 & 2) covering:
  - Municipalities and districts
  - Greater urban areas
  - Time periods from February 2020 to 2021 and from 2022 to present
- 📦 Standardized data structures for consistent analysis
- 🌐 Built-in spatial tessellation handling
- 📈 Designed for research reproducibility and policy applications

## Getting Started

### Installation

```
# Using conda
conda install -c conda-forge pyspainmobility
```

*https://pyspainmobility.github.io/pySpainMobility/*

# SPANISHODDATA R PACKAGE

**https://ropenspain.github.io/spanishoddata/**

Use {spanishoddata} package

```r
library(spanishoddata)
spod_set_data_dir("data")

od_data <- spod_get(
  type = "origin-destination",
  zones = "districts",
  dates = c(
    start = "2022-03-01",
    end = "2022-03-07"
  )
)
```

CRAN 0.2.1

downloads 2282

downloads 107/month

R-CMD-check passing

DOI 10.32614/CRAN.package.spanishoddata

DOI 10.5281/zenodo.14516104

*Logos are the property of their respective owners.*

# Egor Kotov

- Human mobility researcher
- R package developer

spanishoddata

rJavaEnv

ducksf

rdocdump

gridmaker

r5rgui

and more...

ekotov.pro