



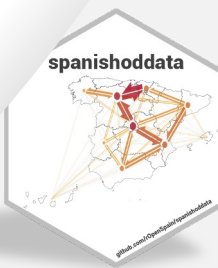
# SPANISHODDATA

## A PACKAGE FOR ACCESSING AND WORKING WITH SPANISH OPEN MOBILITY BIG DATA



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

AGILE, Tartu, 17 Jun 2026





 Search
  Access/Profile
  Cart

Environment and Planning B: Urban Analytics and City Science

Impact Factor: **3.1**  
 5-Year Impact Factor: **3.8**

 Open access
   Research article
 First published online January 17, 2026

**spanishoddata: A package for accessing and working with Spanish Open Mobility Big Data**

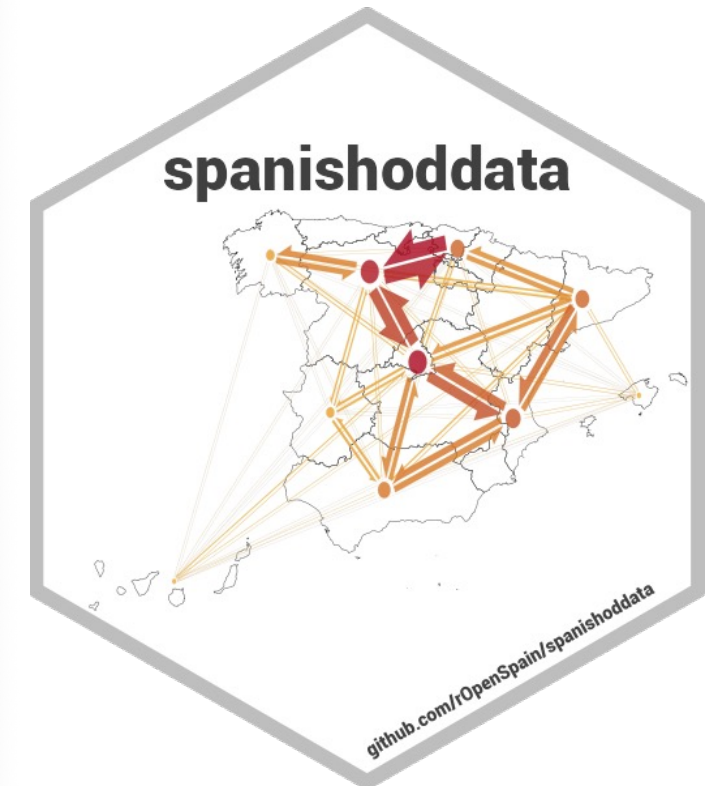
[Egor Kotov](#)  [Eugeni Vidal-Tortosa](#) , [...], and [Robin Lovelace](#)  [View all authors and affiliations](#)

[OnlineFirst](#) | <https://doi.org/10.1177/23998083251415040>

[Contents](#)
[PDF/EPUB](#)
[More](#)

### Abstract

We present **spanishoddata**, an R package that enables fast and efficient access to Spain's open, high-resolution origin-destination human mobility datasets, derived from anonymised mobile-phone records and released by the Ministry of Transport and Sustainable Mobility. The package directly addresses challenges of data accessibility, reproducibility, and efficient processing identified in prior studies. **spanishoddata** automates retrieval from the official source, performs file and schema validation, and converts the data to efficient, analysis-ready formats (DuckDB and Parquet) that enable multi-month and multi-year analysis on consumer-grade hardware. The interface handles complexities associated with these datasets, enabling a wide range of people – from data science beginners to experienced practitioners with domain expertise – to start using the data with just a few lines of code. We demonstrate the utility of the package with example applications in urban transport planning, such as assessing cycling potential or understanding mobility patterns by activity type. By simplifying data access and promoting reproducible workflows, **spanishoddata** lowers the barrier to entry for researchers, policymakers,



Kotov, E., Vidal-Tortosa, E., Cantú-Ros, O. G., Burrieza-Galán, J., Herranz, R., Gullón Muñoz-Repiso, T., & Lovelace, R. (2026). *spanishoddata: A package for accessing and working with Spanish Open Mobility Big Data. Environment and Planning B: Urban Analytics and City Science. DOI:10.1177/23998083251415040*

# NATIONAL STATISTICAL OFFICE EXPERIMENTING WITH COOL DATA INTERNALLY



Image: OpenAI DALL-E + Google NanoBanana



# NATIONAL STATISTICAL OFFICE EXPERIMENTING WITH COOL DATA INTERNALLY

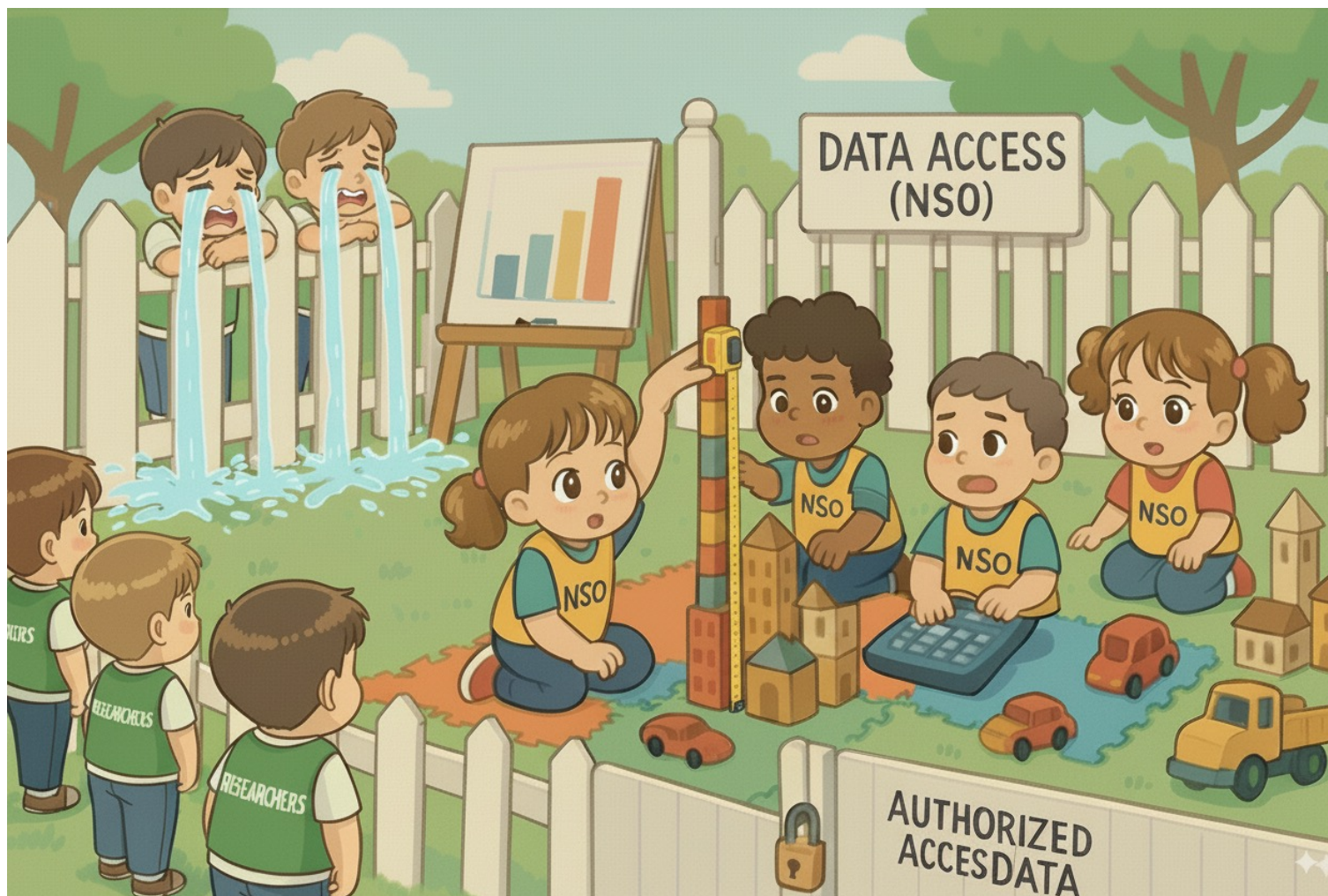


Image: OpenAI DALL-E + Google NanoBanana



# 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

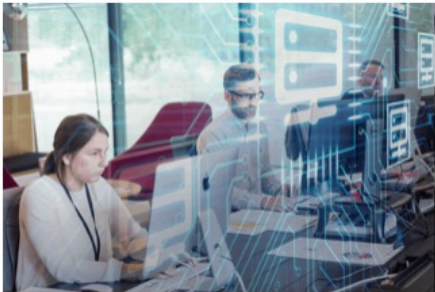
## Open Data Movilidad

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.

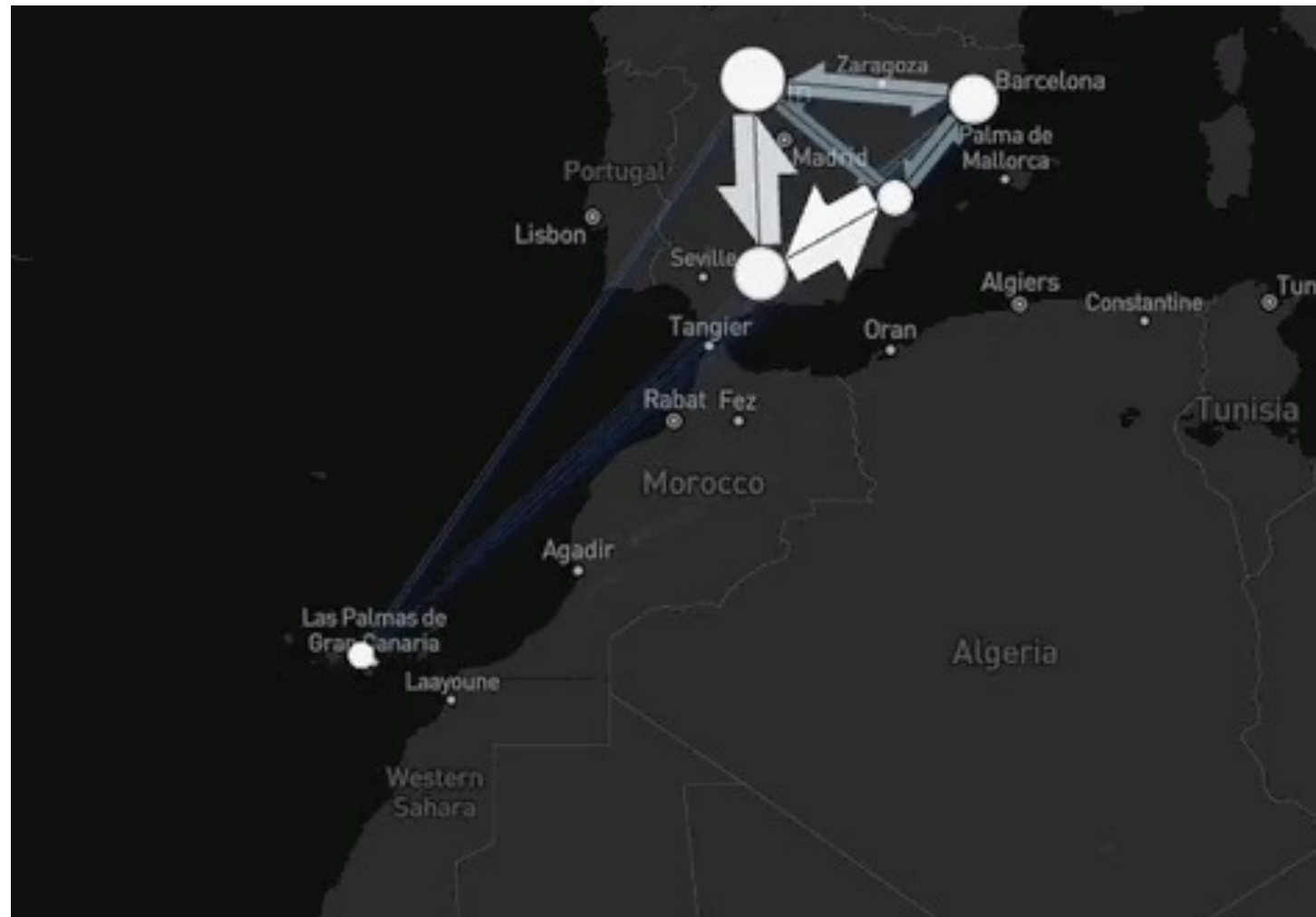
- Se ha utilizado como fuente principal de datos el posicionamiento de los **teléfonos móviles**, siendo una condición indispensable el cumplimiento de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- El contenido está estructurado en 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**:

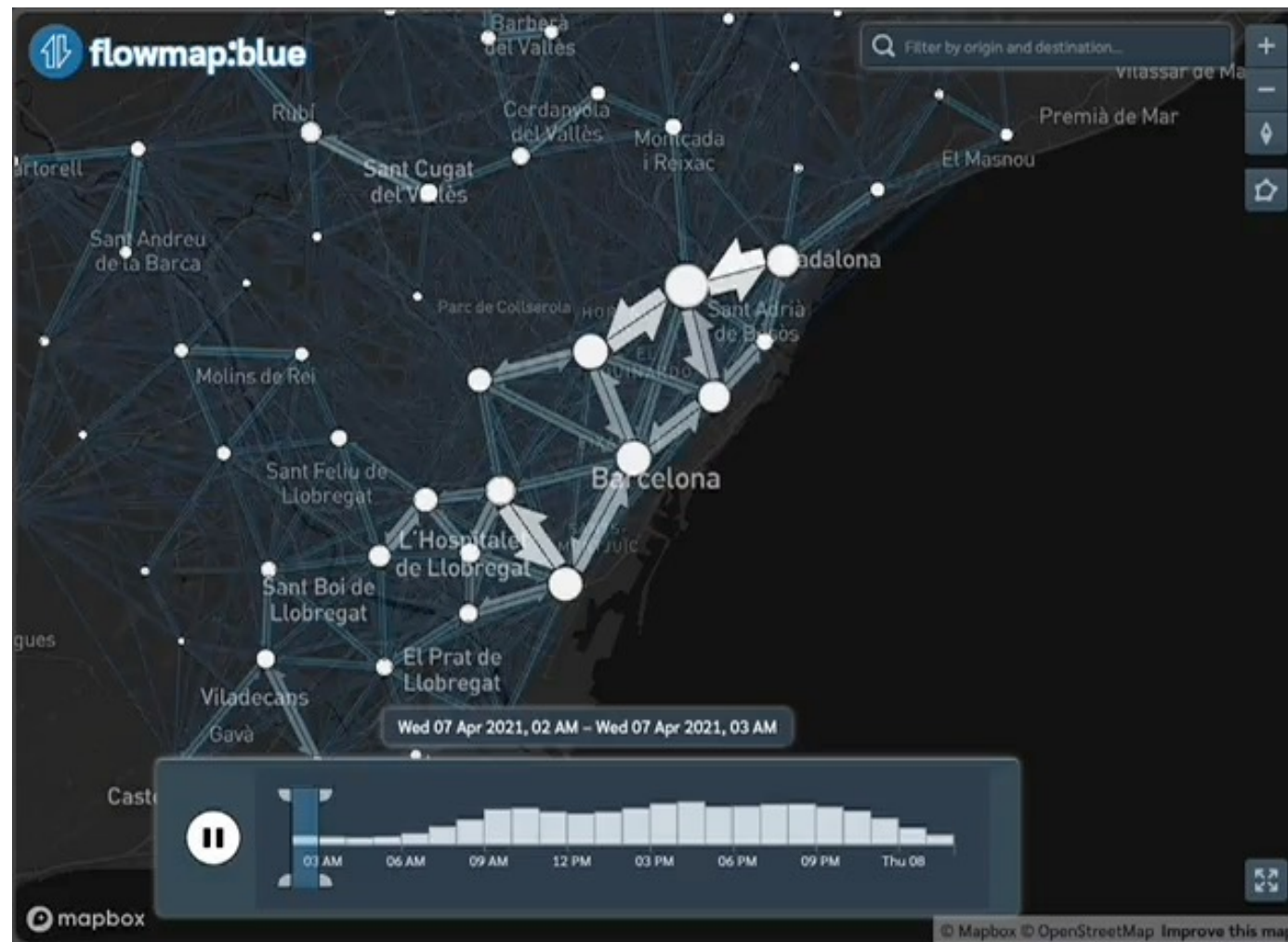
- Estudios básicos**, donde se analiza diariamente la movilidad de los residentes en España a nivel de distrito censal. Como resultados se dispone de las matrices de viajes, pernoctaciones y personas.



# SPANISH OPEN MOBILITY BIG DATA

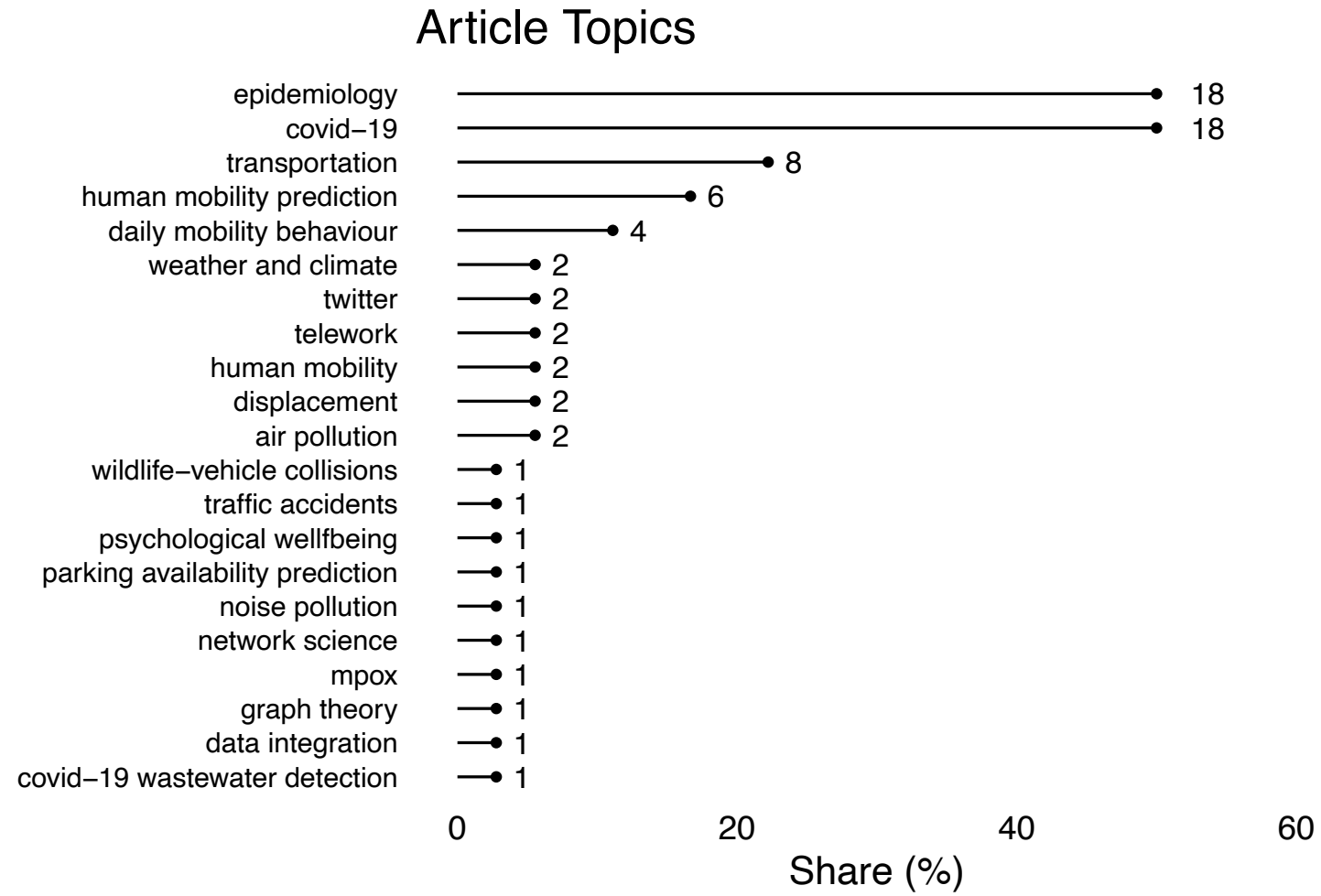


# SPANISH OPEN MOBILITY BIG DATA





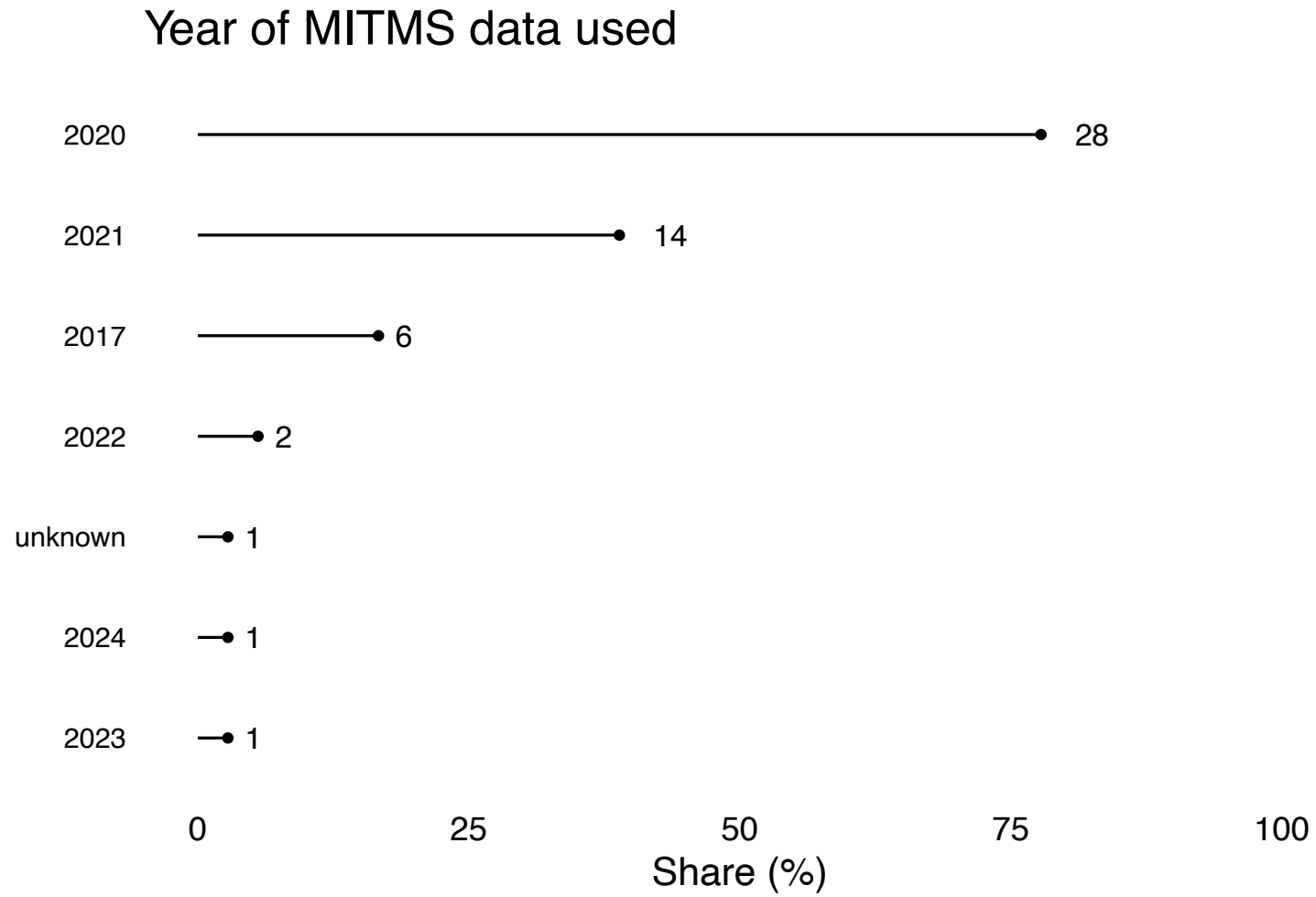
# TOPICS EXPLORED WITH MITMS DATA



May not sum to 100%

Share (on x axis) and Number (on plot) of Articles by Article Topic

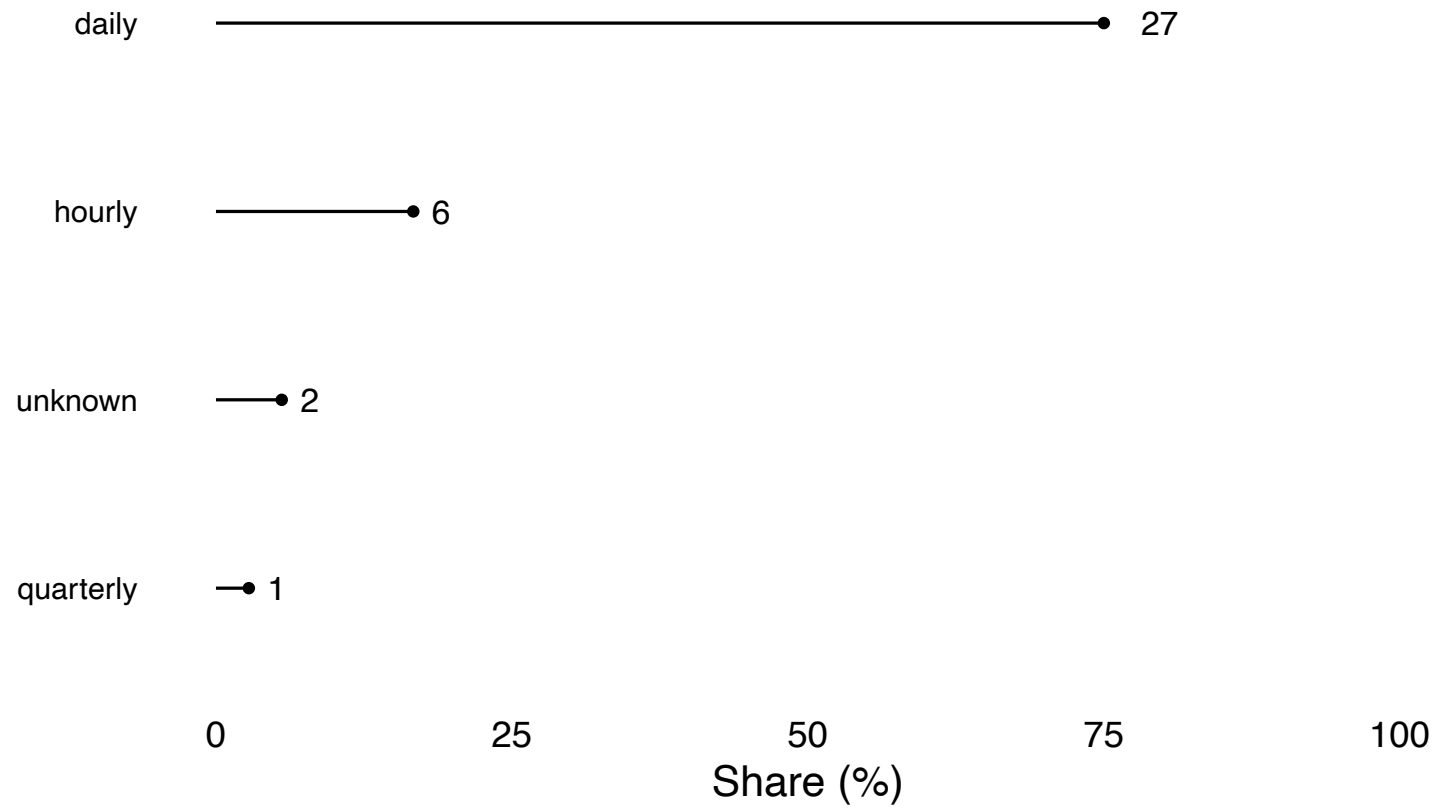
# POTENTIAL OF 2022+ DATA IS STILL UNTAPPED



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

# DAILY TIME RESOLUTION IS MOST POPULAR

## MITMS data time resolution



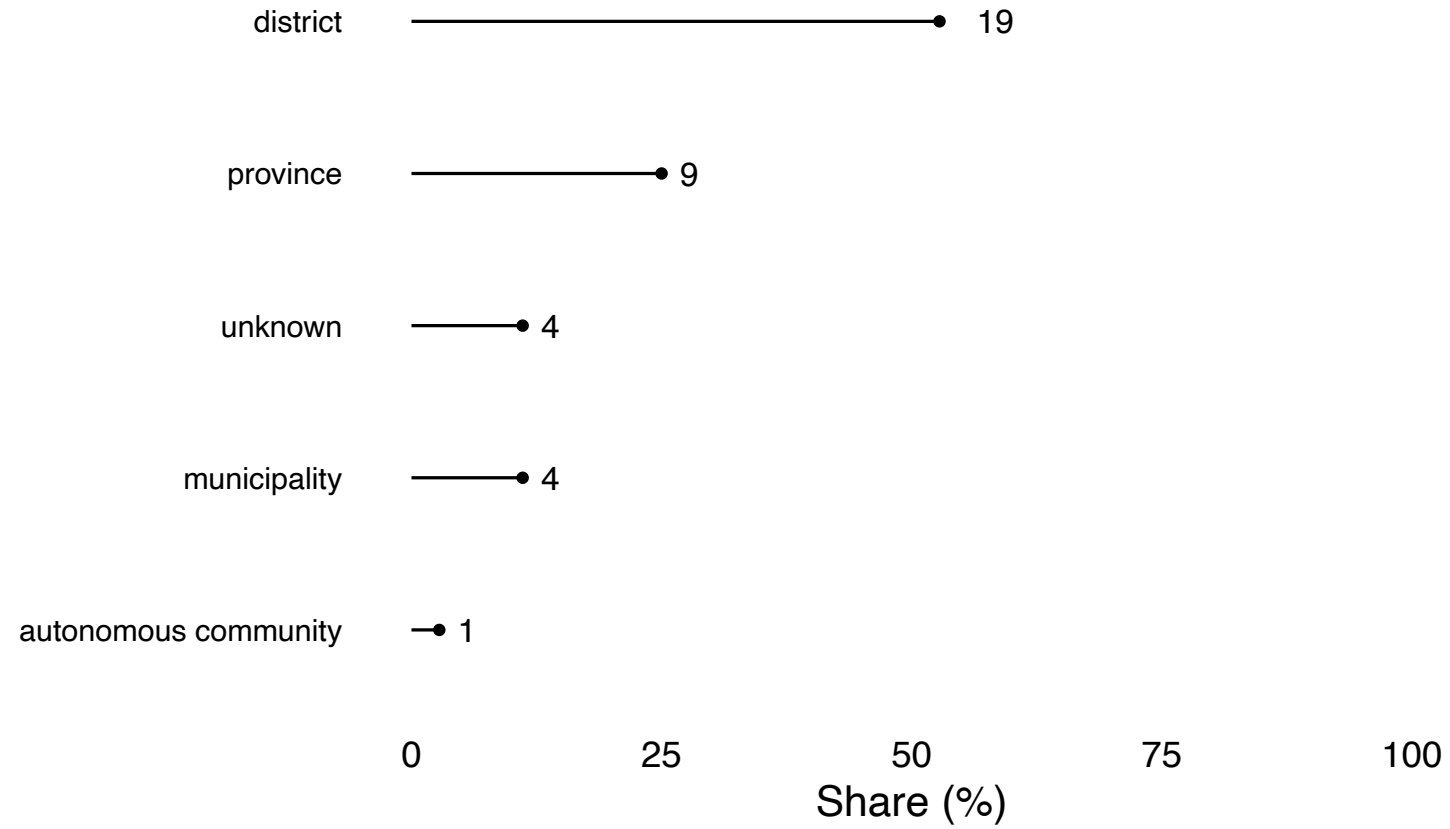
May not sum to 100%

Share (on x axis) and Number (on plot) of Articles by MITMS Time Resolution



# FINEST DISTRICT SPATIAL RESOLUTION IS MOST POPULAR

## MITMS data spatial resolution

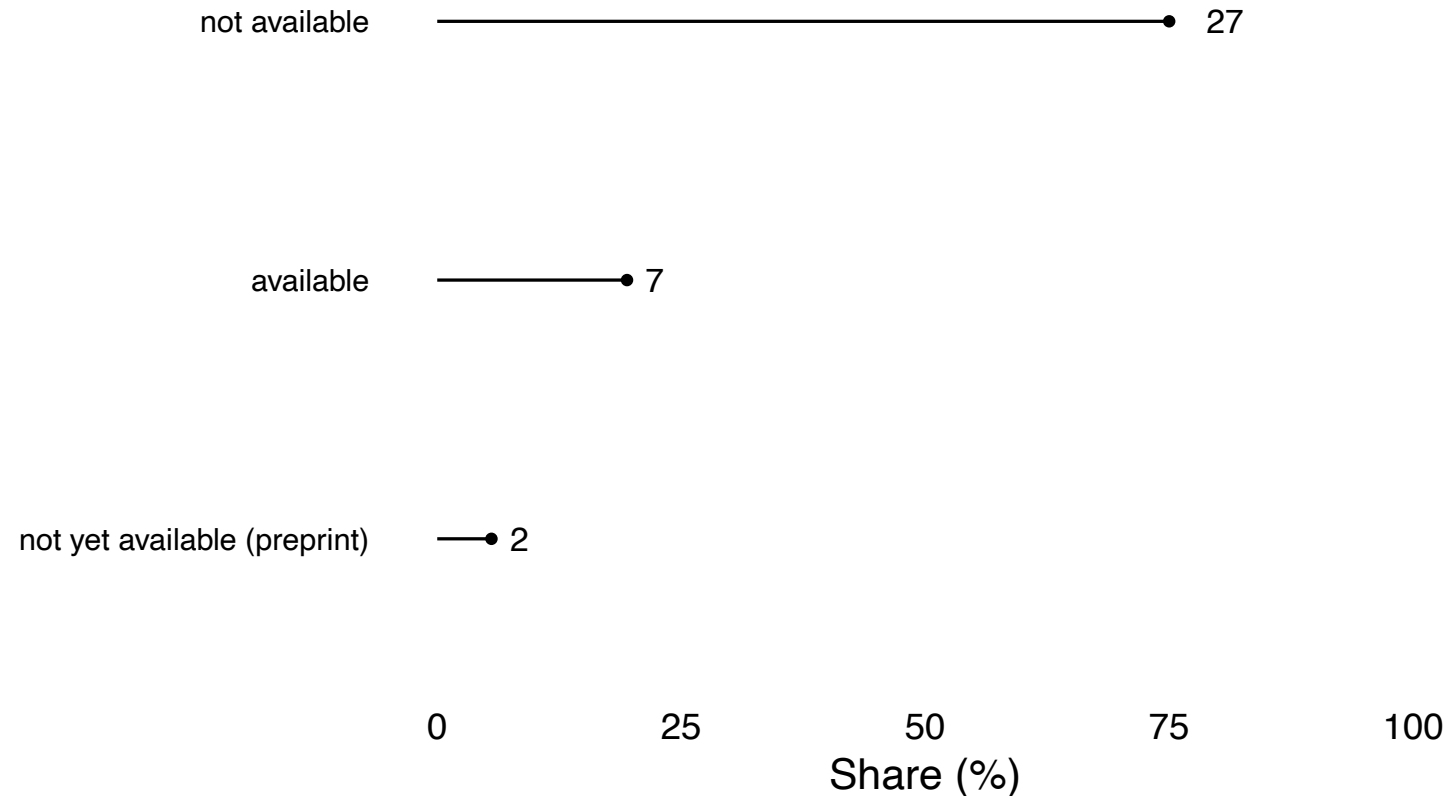


May not sum to 100%

Share (on x axis) and Number (on plot) of Articles by MITMS Spatial Resolution

# PRE-PROCESSING CODE IS USUALLY NOT SHARED

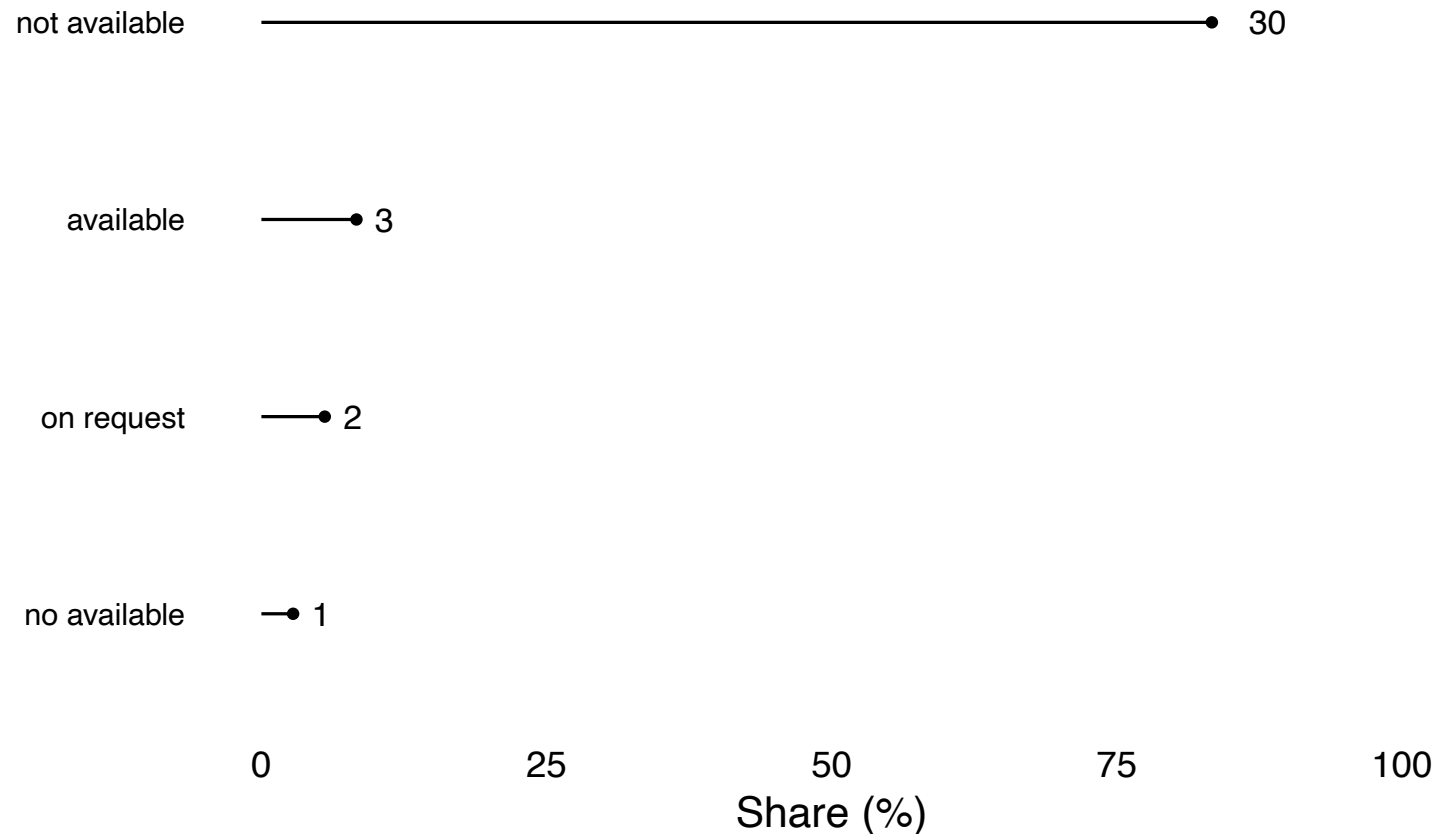
## MITMS data pre-processing code



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

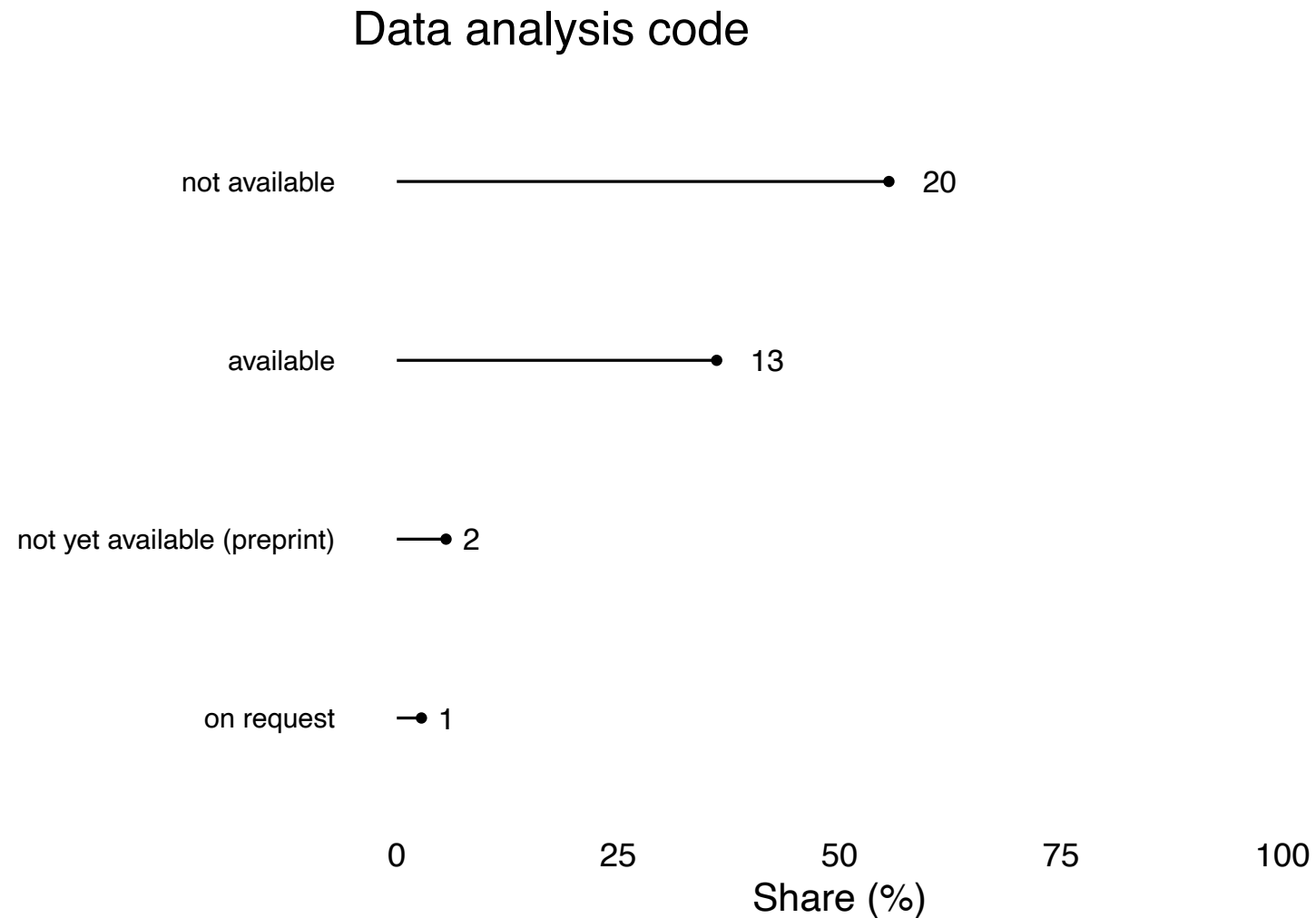


Share (on x axis) and Number (on plot) of Articles by Shared Pre-processed MITMS Data

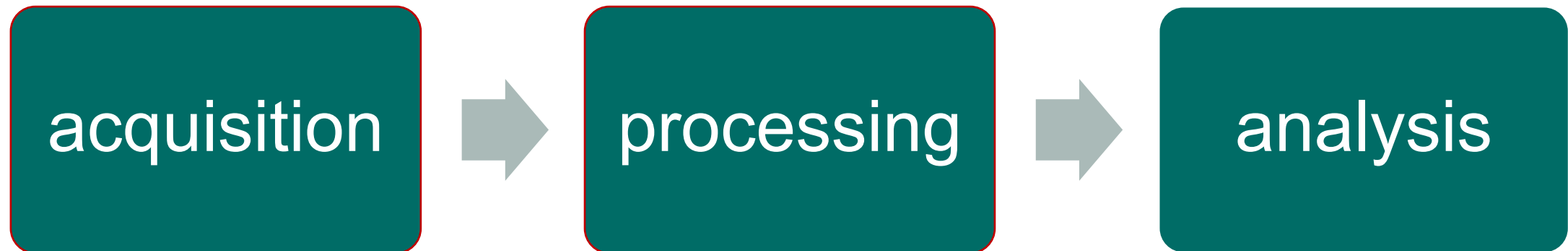
**NOBODY PROVIDES THE CODE HOW TO  
ACQUIRE THE DATA**



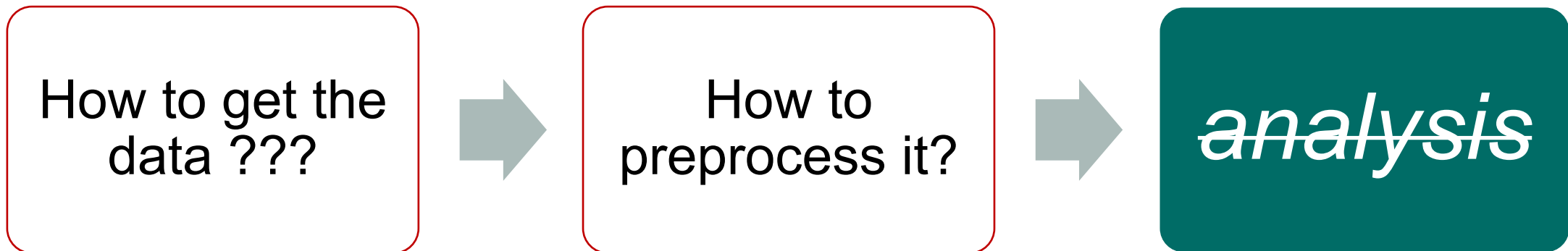
# FEW PROVIDE ANALYSIS CODE



## VERY LITTLE VALUE IN ANALYSIS CODE WITHOUT ACQUISITION AND PROCESSING

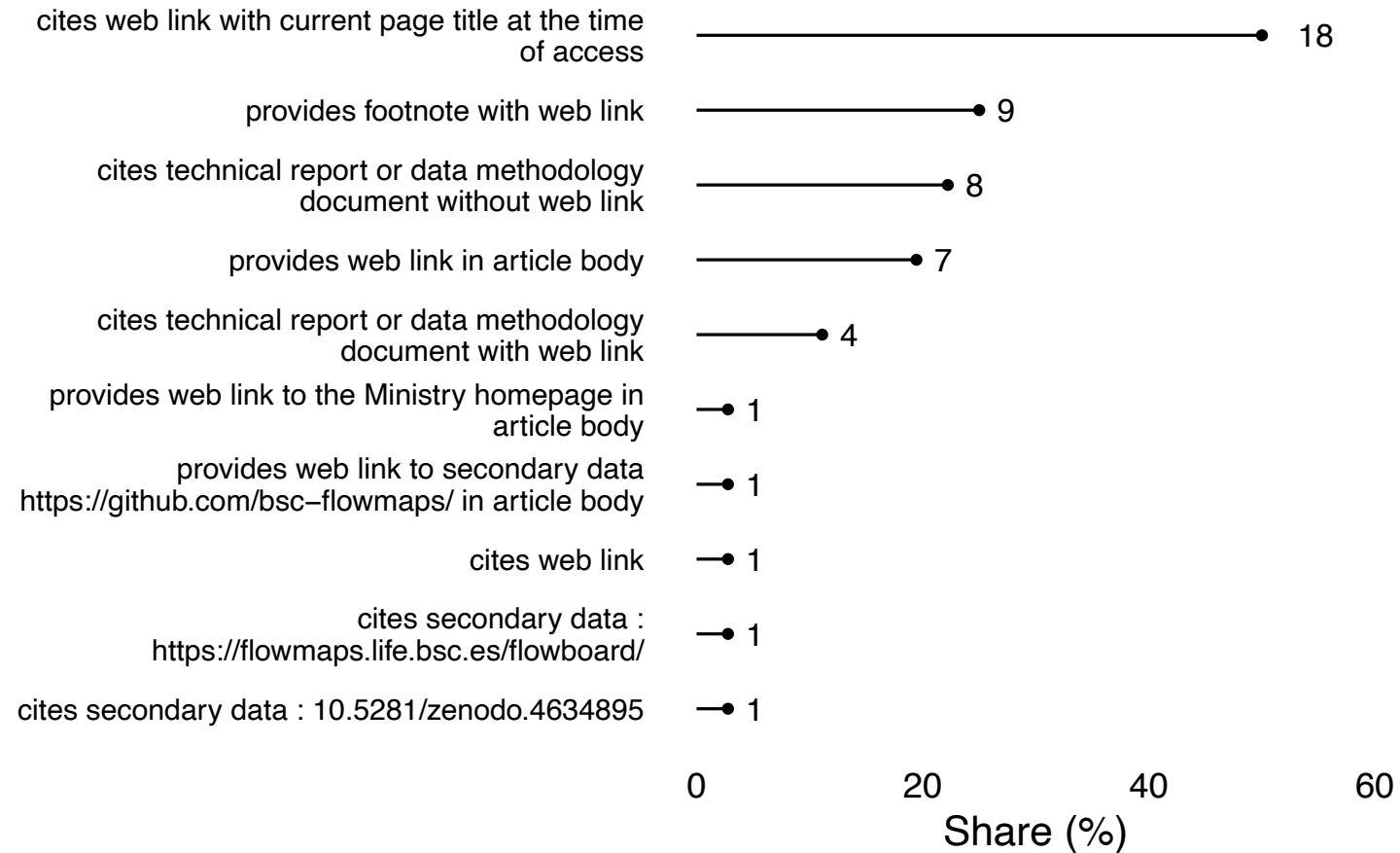


## VERY LITTLE VALUE IN ANALYSIS CODE WITHOUT ACQUISITION AND PROCESSING



# INCONSISTENT CITATION

## Citation approaches



May not sum to 100%  
Share (on x axis) and Number (on plot) of Articles by Data Citation Approach

# DATA ACCESS

| mitma-movilidad-v2 / estudios_basicos / por-municipios / viajes / ficheros-diarios / 2022-01                          |                     |           |                              |
|-----------------------------------------------------------------------------------------------------------------------|---------------------|-----------|------------------------------|
| Mostrar                                                                                                               | 50                  | registros | Buscar: <input type="text"/> |
| 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

```
</language>
<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



| 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   |



```
<?xml version='1.0' encoding='UTF-8'>
<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>
```

- 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

REFERENCEARTICLESCHANGELOG

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 !\[\]\(2fec083a2451e2c253b25d22c984fd4d\_img.jpg\) rOpenSpain](#)

**License**

- [Full license](#)
- MIT + file LICENSE

**Citation**

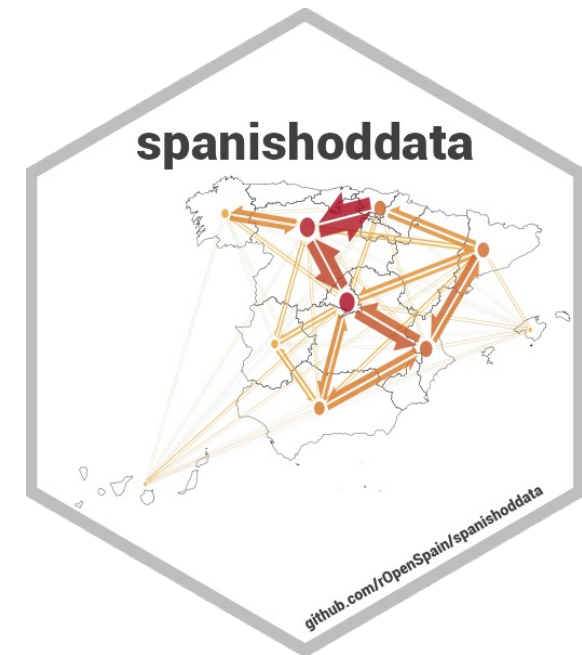
- [Citing spanishoddata](#)

**Developers**

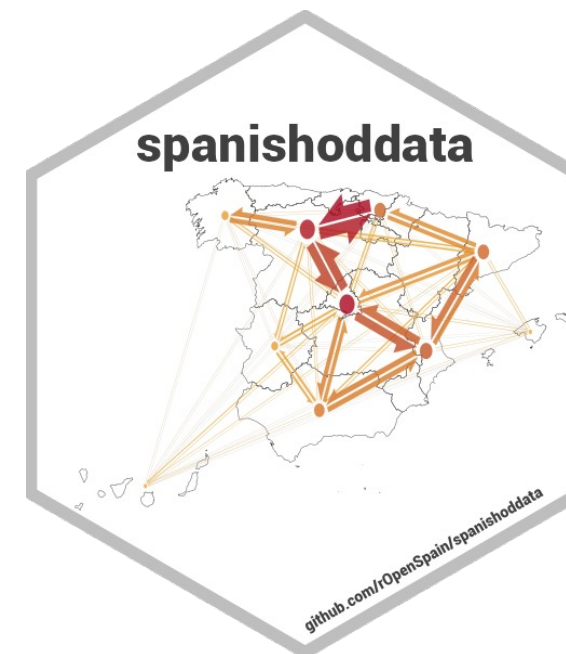
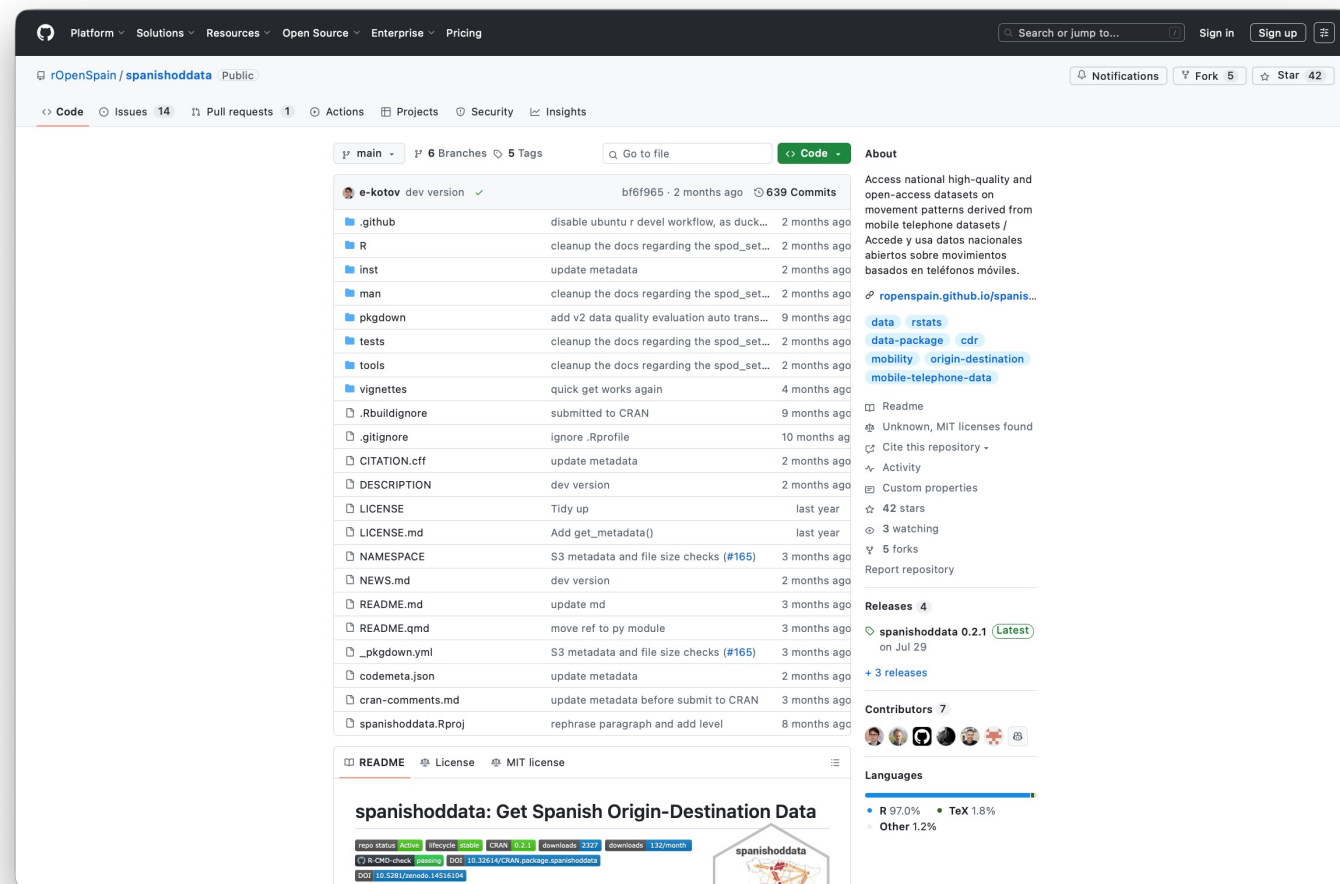
- [Egor Kotov](#)  
Author, maintainer 
- [Robin Lovelace](#)  
Author 
- [More about authors...](#)

**Dev status**

- repo status 
- lifecycle 
- CRAN 



# SPANISHODDATA R PACKAGE



# DATA ACCESS

mitma-movilidad-v2 / estudios\_basicos / por-municipios / viajes / ficheros-diarios / 2022-01

Mostrar 50 registros      Buscar:

| Archivo                                     | Fecha               | Tamaño |
|---------------------------------------------|---------------------|--------|
| <a href="#">[Volver a la carpeta padre]</a> |                     |        |
| 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  |

```
<?xml version="1.0" encoding="UTF-8" ?>
<language>es</language>
<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

```
1 library(spanishoddata)
2 spod_set_data_dir("data")
3
4 od_data <- spod_get(
5   type = "origin-destination",
6   zones = "districts",
7   dates = c(
8     start = "2022-03-01",
9     end = "2022-03-07"
10  )
11 )
```



# ACCESS TO TABULAR DATA

Use `{spanishoddata}`  
package

```
1 library(spanishoddata)
2 spod_set_data_dir("data")
3
4 od_data <- spod_get(
5   type = "origin-destination",
6   zones = "districts",
7   dates = c(
8     start = "2022-03-01",
9     end = "2022-03-07"
10  )
11 )
```

```
1 library(dplyr)
2 glimpse(od_data)
3
4 Rows: ??
5 Columns: 20
6 Database: DuckDB v1.2.1 [root@Darwin 24.4.0:R 4.5.0/
7 $ date <date> 2022-01-04, 202
8 $ hour <int> 0, 0, 0, 1, 1, 3
9 $ id_origin <fct> 01001, 01001, 01
10 $ id_destination <fct> 01009_AM, 01009_
11 $ distance <fct> 2-10, 2-10, 2-10
12 $ activity_origin <fct> home, frequent_a
13 $ activity_destination <fct> frequent_activit
14 $ study_possible_origin <lgl> FALSE, FALSE, FA
15 $ study_possible_destination <lgl> FALSE, FALSE, FA
16 $ residence_province_ine_code <fct> 01, 01, 01, 01,
17 $ residence_province_name <fct> "Araba/Álava", "
18 $ income <fct> 10-15, >15, >15,
19 $ age <fct> NA, NA, NA, NA,
20 $ sex <fct> NA, NA, NA, NA,
21 $ n_trips <dbl> 4.894, 1.779, 1.
22 $ trips_total_length_km <dbl> 27.966, 5.997, 4
23 $ year <int> 2022, 2022, 2022
24 $ month <int> 1, 1, 1, 1, 1, 1
25 $ day <int> 4, 4, 4, 4, 4, 4
```

## EASE OF USE – FAMILIAR TO R USERS

### *Filter and summary*

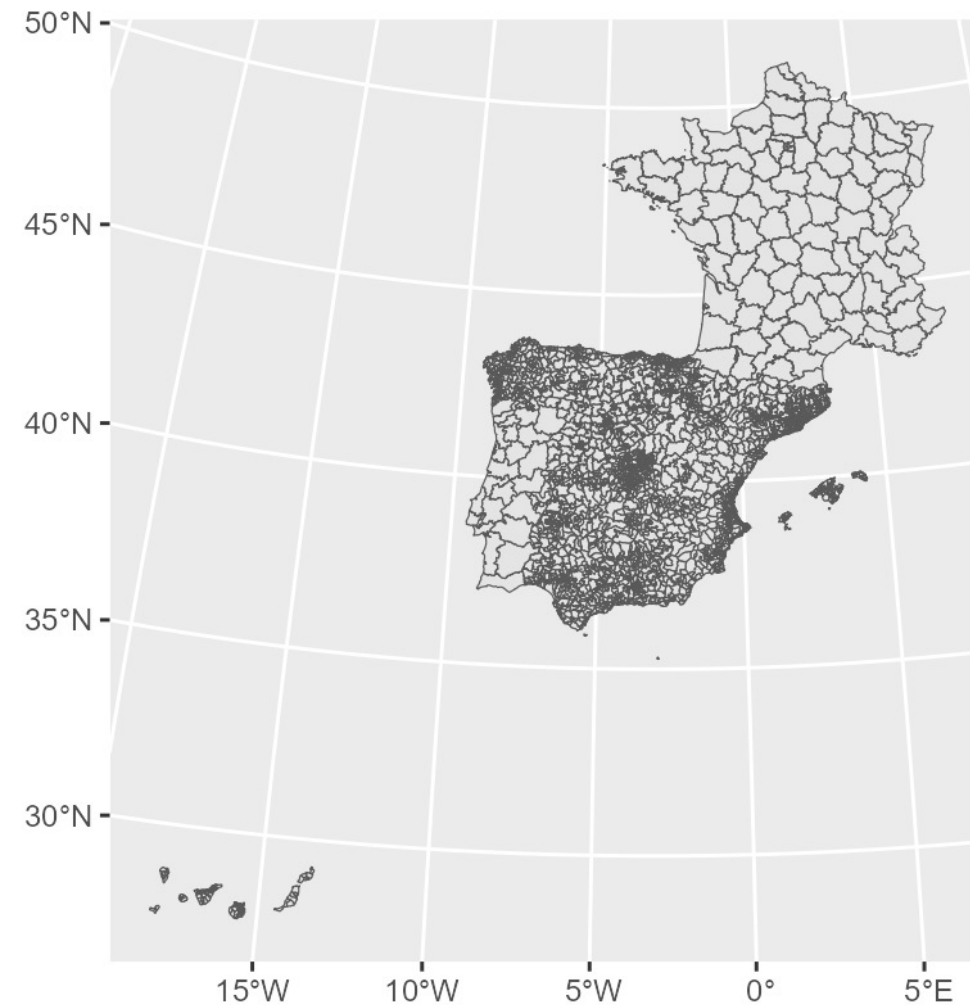
```
1 library(dplyr)
2
3 od_data |>
4   filter(
5     year == 2022,
6     month %in% c(2, 3, 4)
7   ) |>
8   summarise(
9     n_trips = mean(n_trips)
10  ) |>
11  collect()
```

### *Summary over multiple groups on full data set*

```
1 library(dplyr)
2
3 od_data |>
4   group_by(
5     year,
6     month,
7     day,
8     id_origin,
9     id_destination
10  )
11  summarise(
12    n_trips = mean(n_trips)
13  ) |>
14  collect()
```

# ACCESS TO SPATIAL DATA

```
1 zones <- spod_get_zones(  
2   zones = "districts",  
3   ver = 2 # for 2022 data onwards  
4 )  
5  
6 zones |>  
7   sf::st_geometry() |>  
8   ggplot2::ggplot() +  
9   ggplot2::geom_sf()
```





# DATA CITATION

```
1 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                                                                                                                                                |
|-----------------------|------------------------|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>date</code>     | <code>fecha</code>     | Date    | The date of the recorded data, formatted as YYYY-DD.                                                                                                       |
| <code>hour</code>     | <code>periodo</code>   | integer | The time slot during which the trips occurred. <i>Note:</i> this column used to be called <code>time_slot</code> , still going to be available in the outp |

### On this page

Install the package

Set the data directory

Overall approach to accessing the data

1. Spatial data with zoning boundaries

2. Mobility data

2.1. Origin-destination data

2.2. Population by trip count data

2.3. Population by overnight stay data

Advanced use

# 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**  
Download the data files of specified type, zones and dates

### 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

# SPANISHODDATA R PACKAGE – USER GUIDES



spanishoddata 0.2.1.9000

REFERENCE

ARTICLES

CHANGELOG

Search for

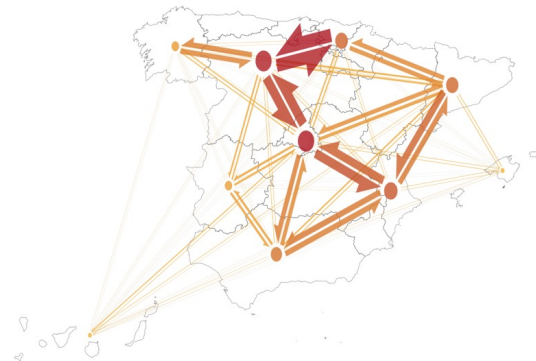


```
outline_col = "grey80",  
k_node = 10 # play around with this parameter to  
)  
  
# customise colours and remove legend, as we need a clean  
flows_plot <- flows_plot +  
  guides(fill = "none") +  
  scale_fill_gradient(low = "#FABB29", high = "#AB061F")  
  
flows_plot
```

CODEBOOK AND COOKBOOK FOR V1 (2020-2021) SPANISH MOBILITY DATA  
CODEBOOK AND COOKBOOK FOR V2 (2022 ONWARDS) SPANISH MOBILITY DATA  
QUICKY GET DAILY DATA  
DOWNLOAD AND CONVERT MOBILITY DATASETS  
OD DATA DISAGGREGATION  
MAKING STATIC FLOW MAPS  
MAKING INTERACTIVE FLOW MAPS

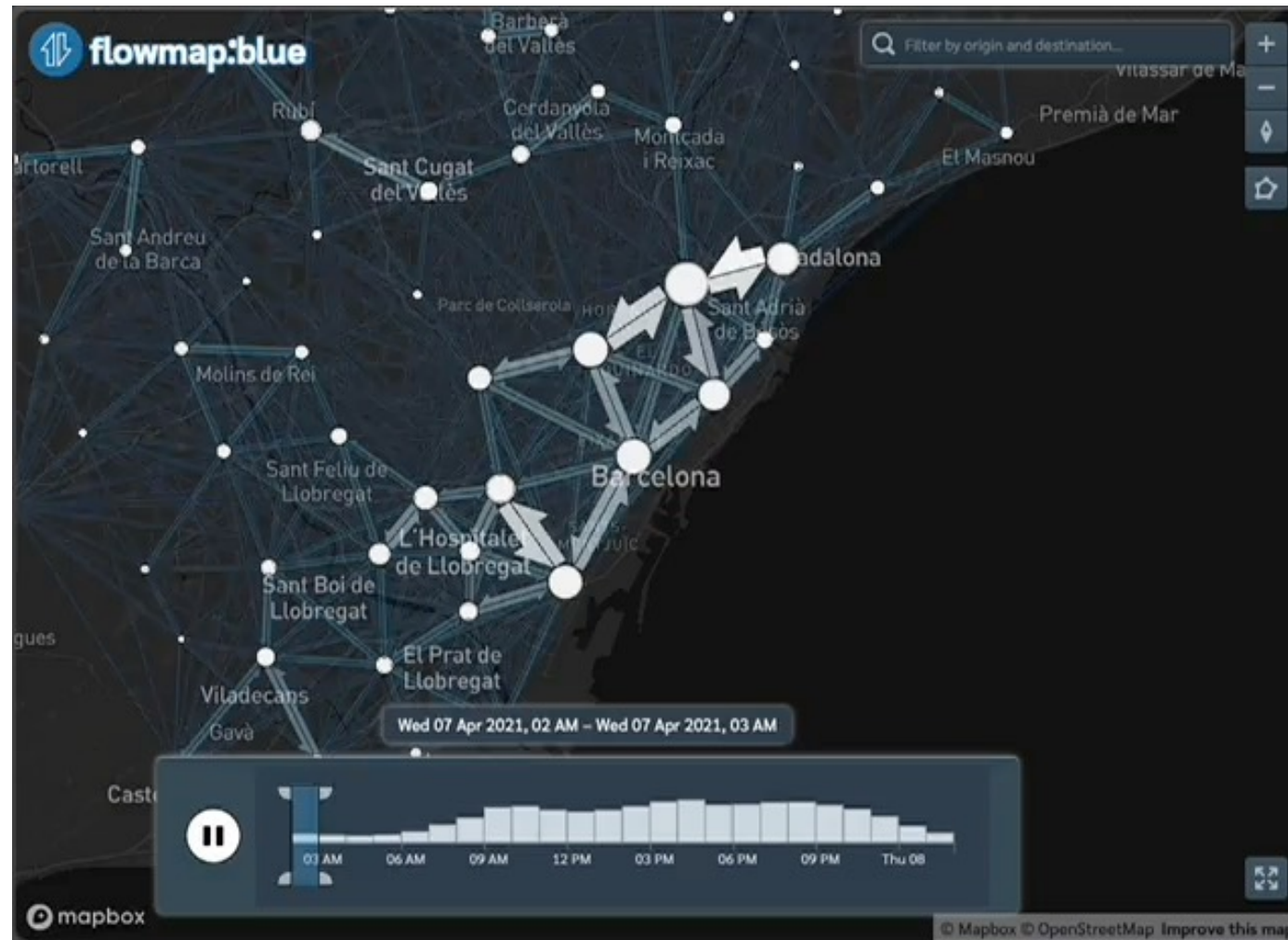
4 Advanced example -  
aggregate flows for  
{spanishoddata} logo

- 4.1 Get data
- 4.2 Flows aggregation
- 4.3 Reshape flows for visualization
- 4.4 Plot the flows
- 4.5 Make the sticker





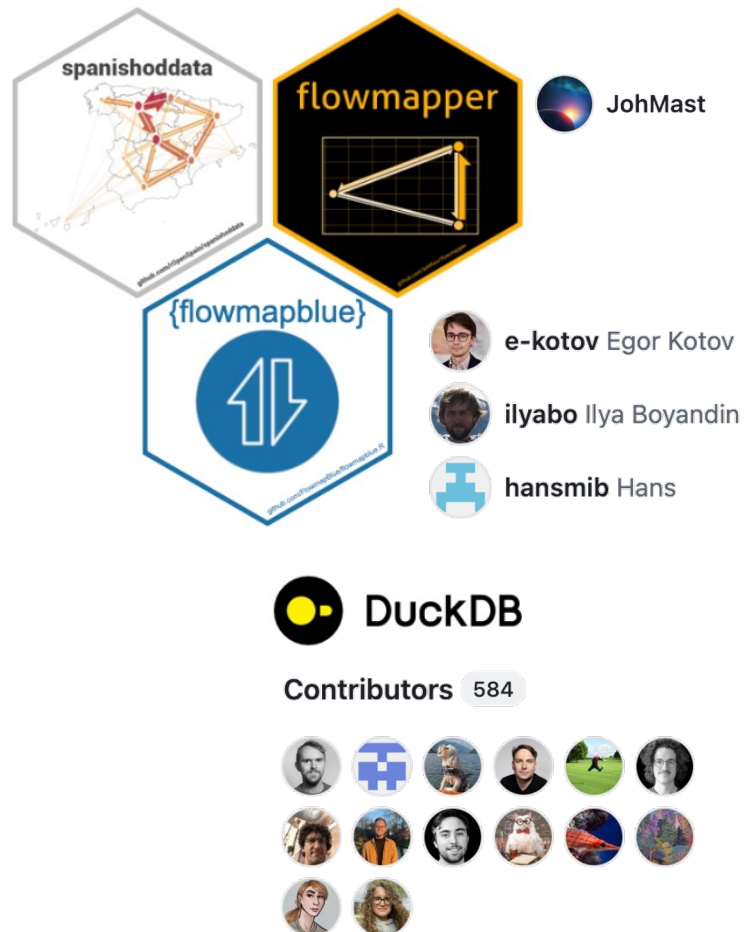
# TUTORIALS TO CREATE INTERACTIVE FLOW MAPS





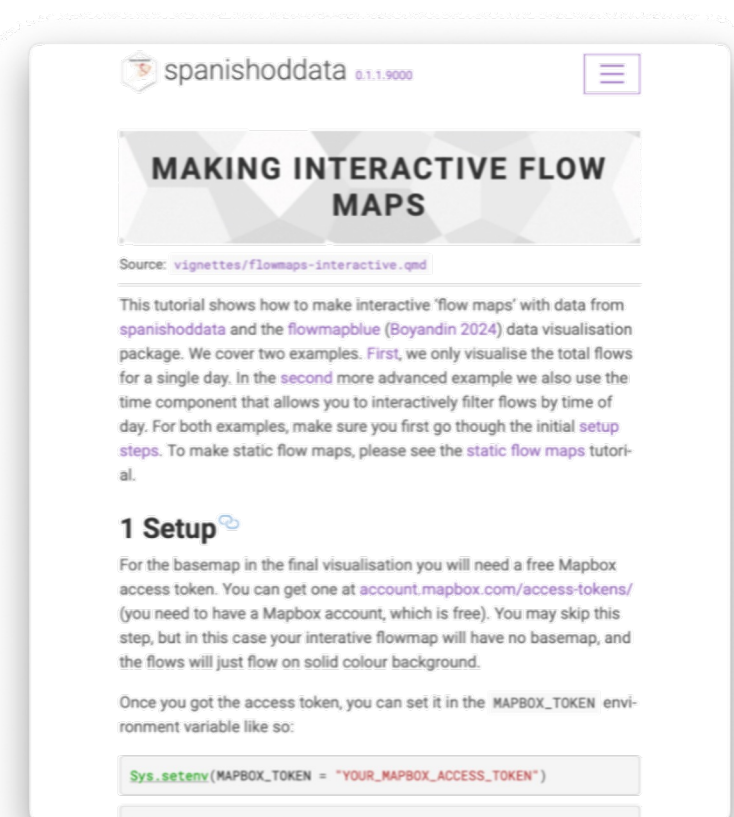
# ECOSYSTEM OF R PACKAGES

*A package with many companions*



[+ 570 contributors](#)

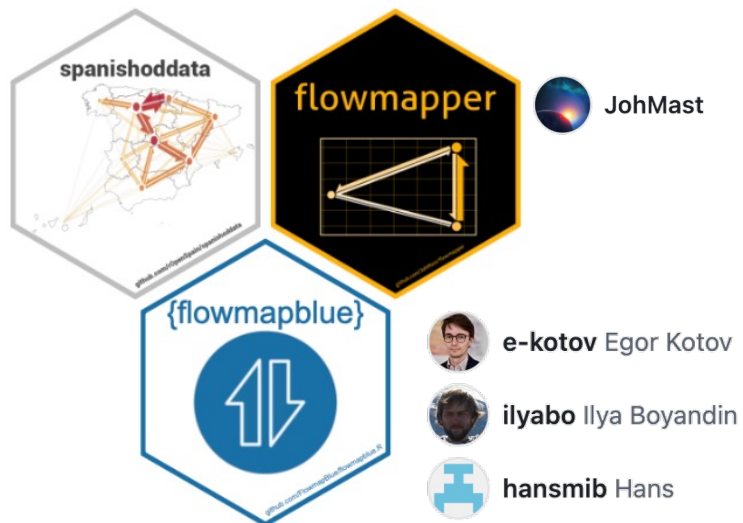
*Detailed documentation*



*Logos are the property of their respective owners.*

# ECOSYSTEM OF R PACKAGES

*A package with many companions*



 **JohMast**

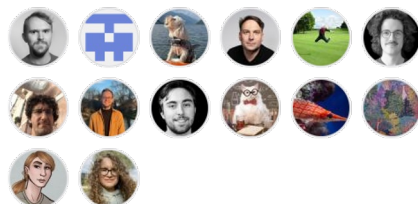
 **e-kotov** Egor Kotov

 **ilyabo** Ilya Boyandin

 **hansmib** Hans



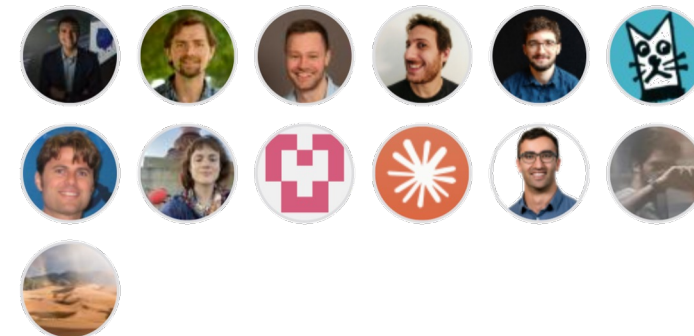
Contributors 584



[+ 570 contributors](#)



Kyle Walker



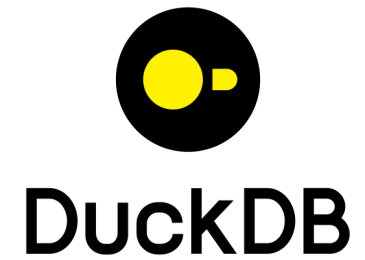
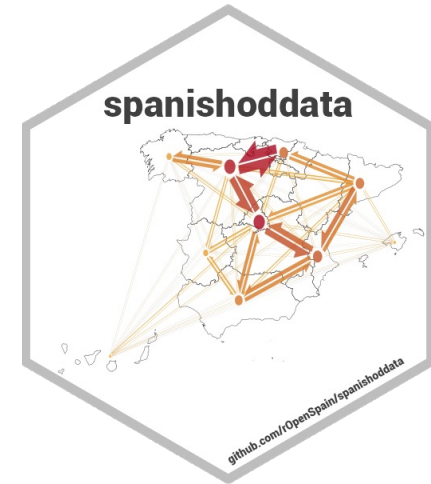
New version of  
**mapgl** flowmaps  
support coming  
soon!

 **ilyabo** Ilya Boyandin

 **igorDykhta** Igor Dykhta

 **AlexGodbehere** Alex Godbehere

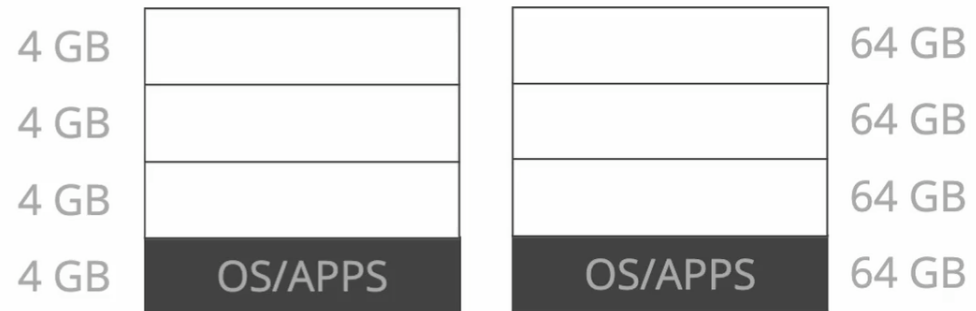
# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB



# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB



DuckDB

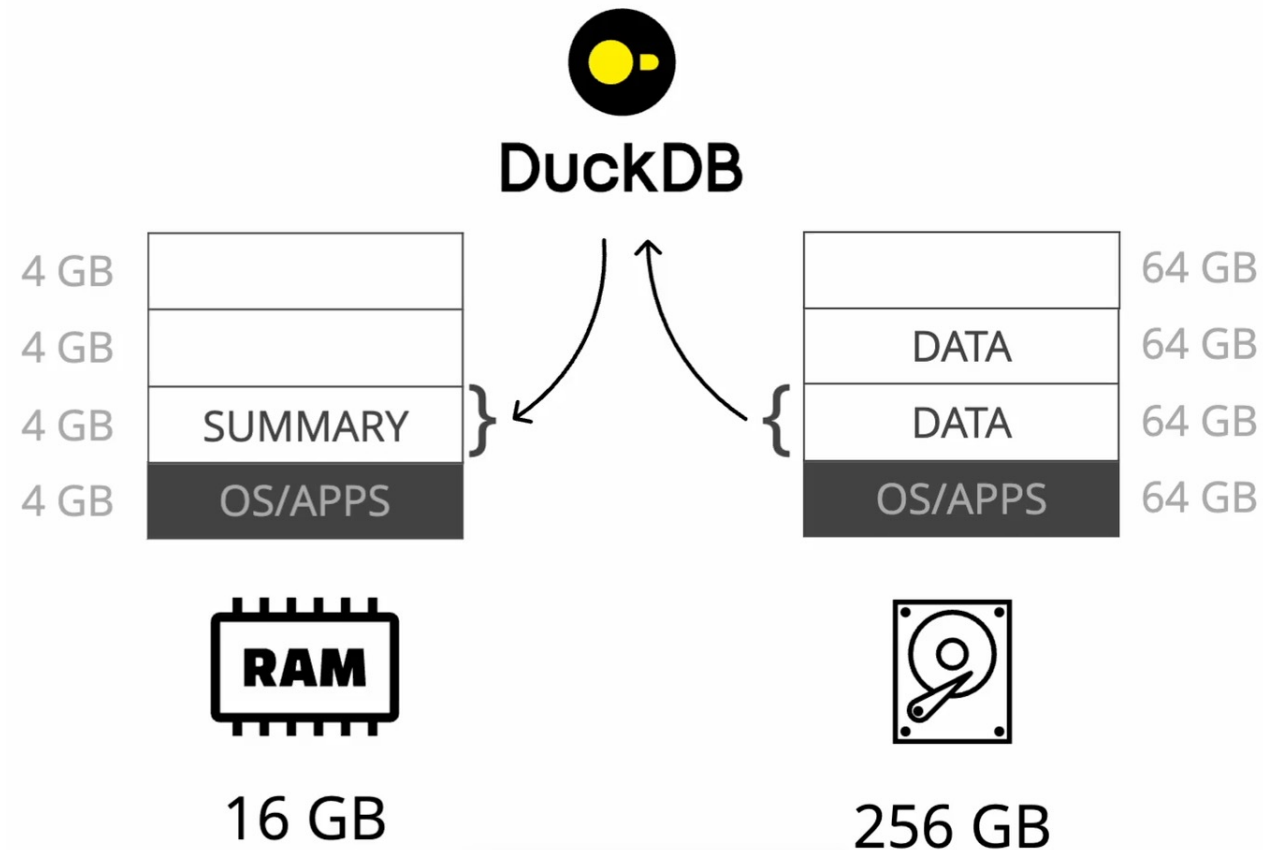


16 GB

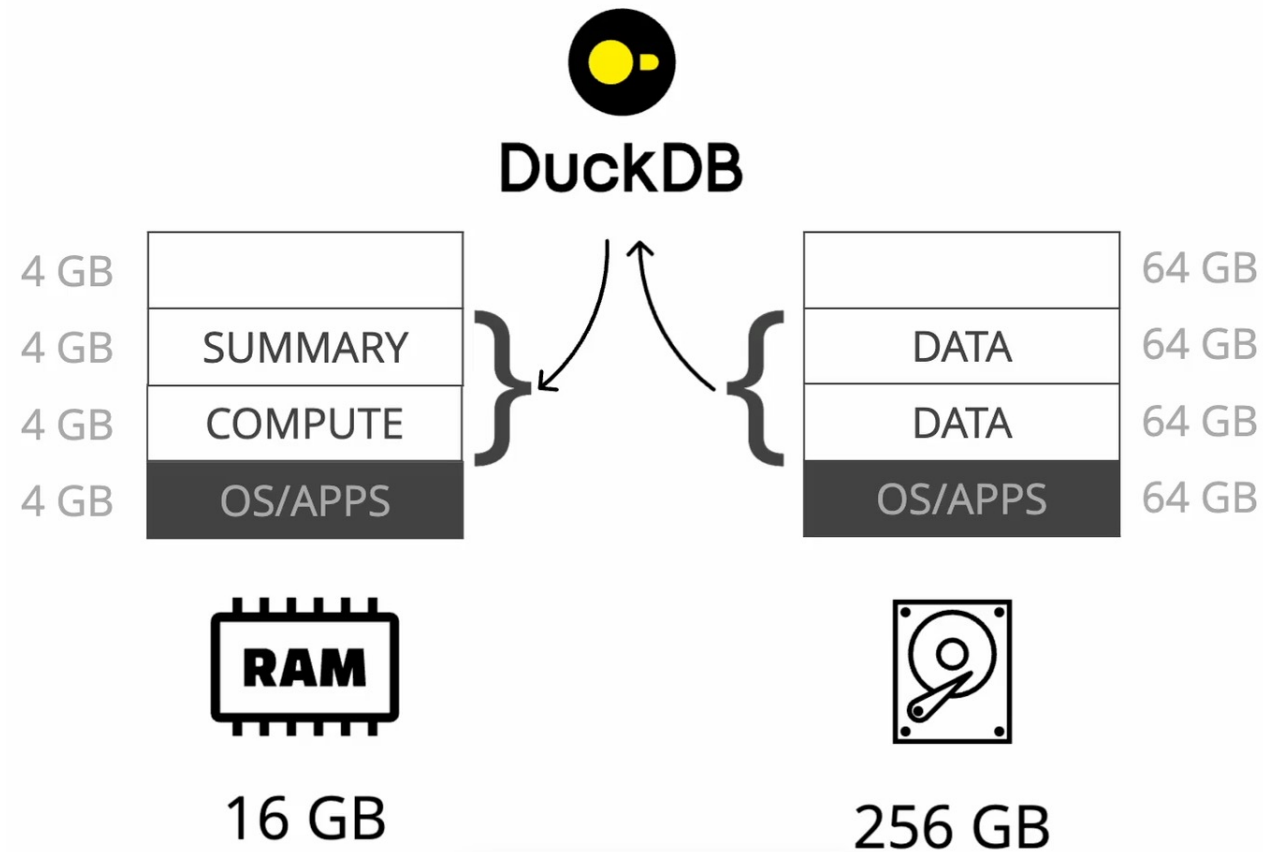


256 GB

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

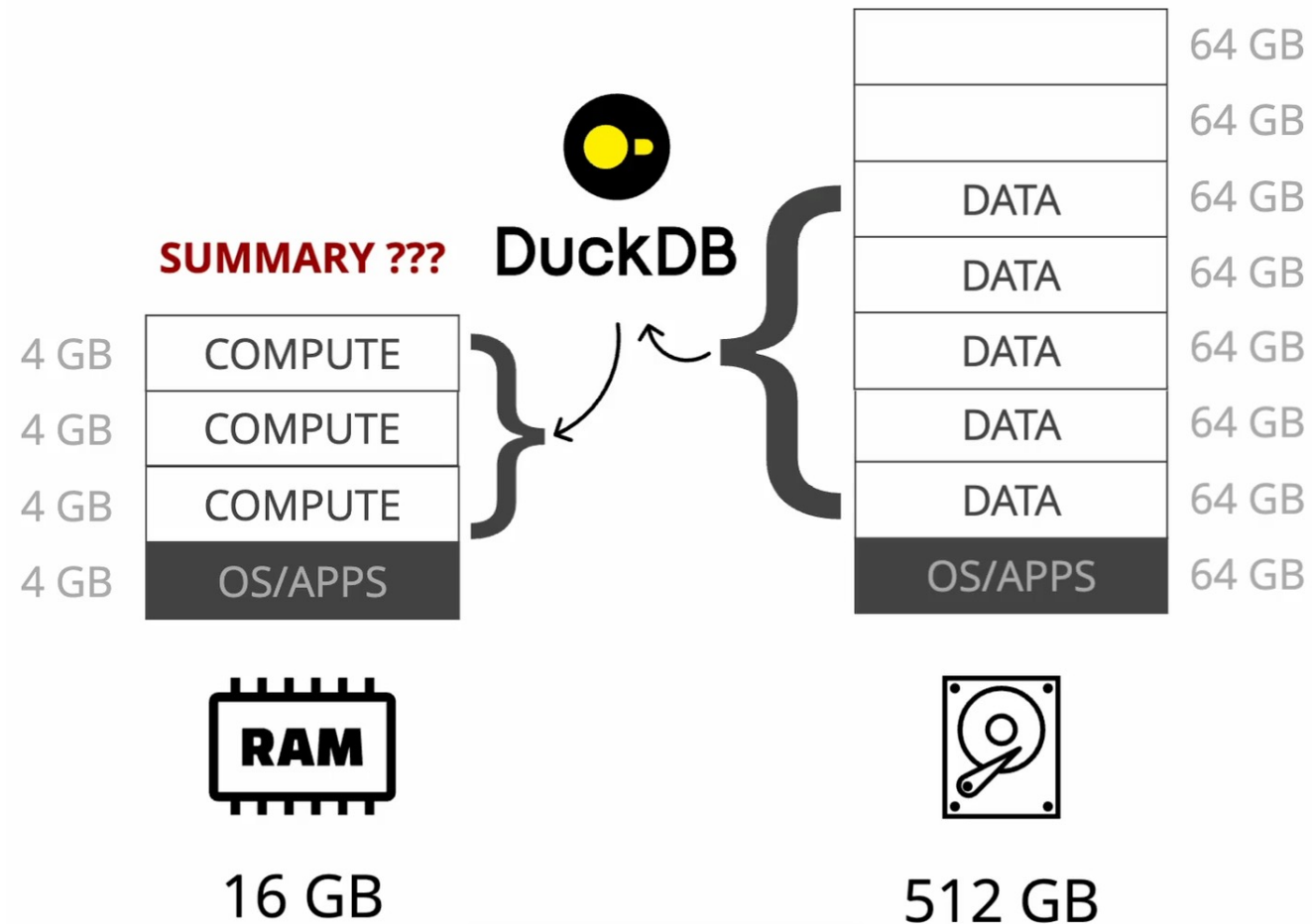


# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB





# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB



## EASE OF USE – FAMILIAR TO R USERS

### *Filter and summary*

```
1 library(dplyr)
2
3 od_data |>
4   filter(
5     year == 2022,
6     month %in% c(2, 3, 4)
7   ) |>
8   summarise(
9     n_trips = mean(n_trips)
10  ) |>
11  collect()
```

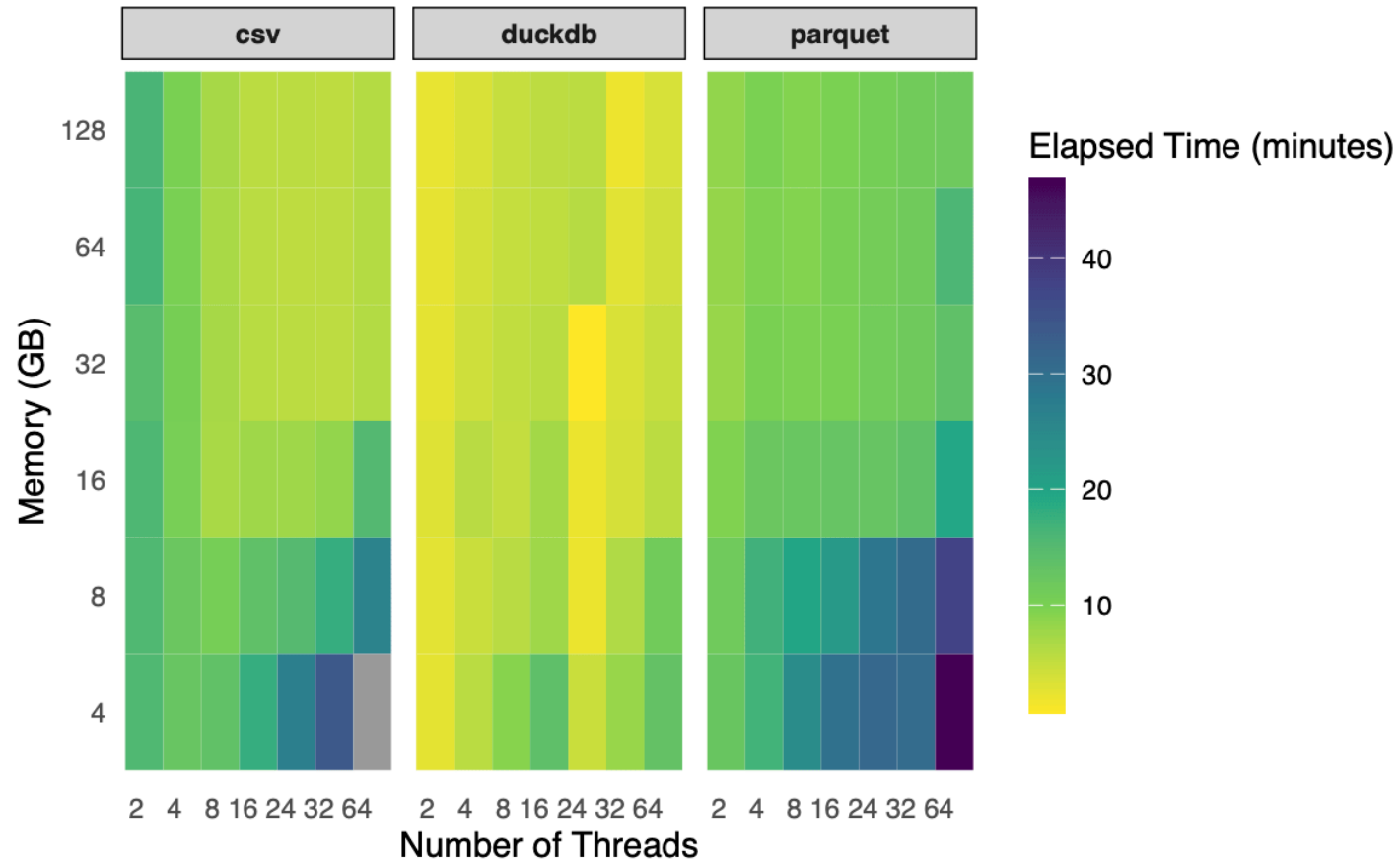
### *Summary over multiple groups on full data set*

```
1 library(dplyr)
2
3 od_data |>
4   group_by(
5     year,
6     month,
7     day,
8     id_origin,
9     id_destination
10  )
11  summarise(
12    n_trips = mean(n_trips)
13  ) |>
14  collect()
```

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

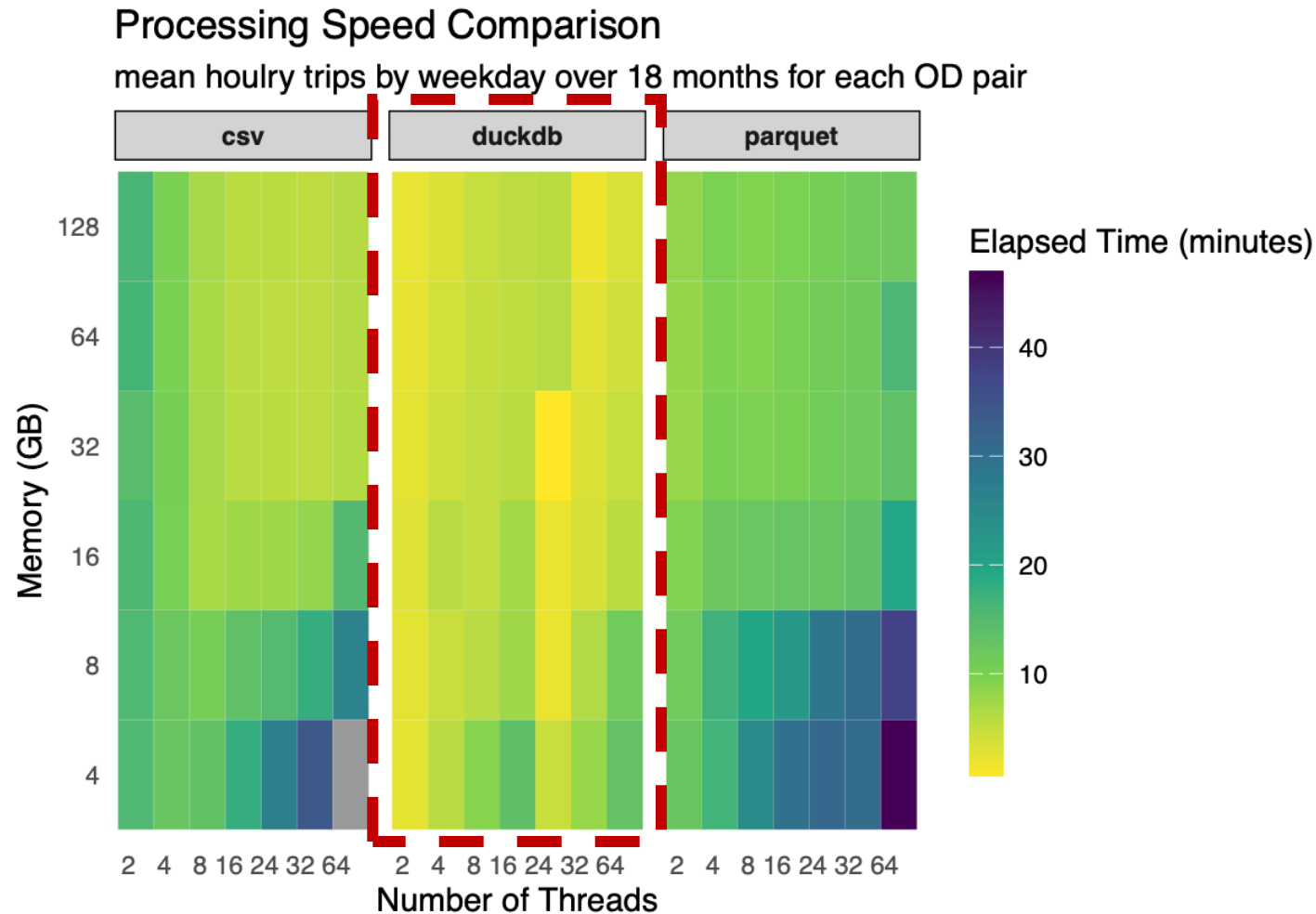
## Processing Speed Comparison

mean hourly trips by weekday over 18 months for each OD pair



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

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

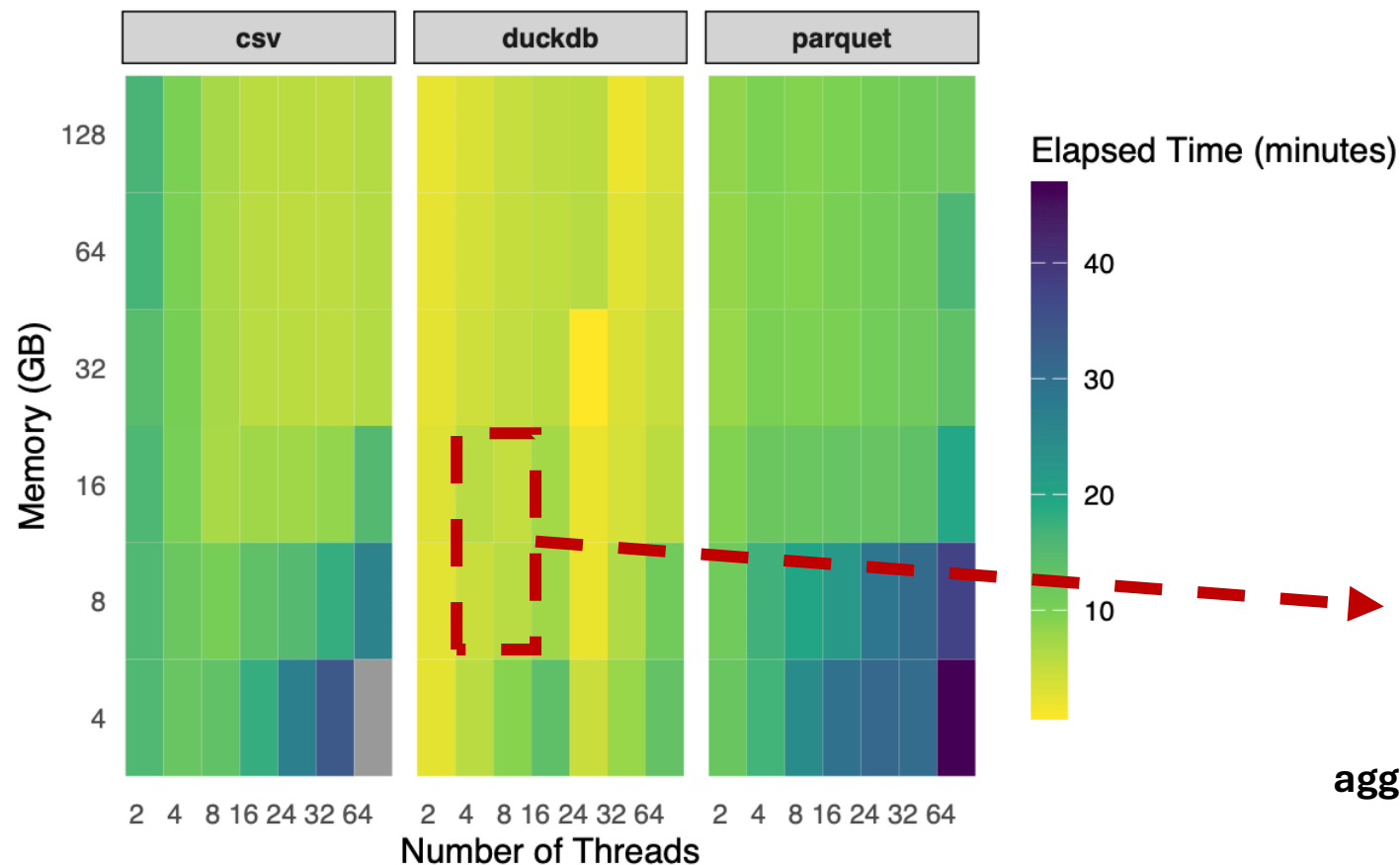


Data processing speed comparison: DuckDB engine running on CSV.gz files  
vs DuckDB database vs a folder of Parquet files

# BIG DATA ON CONSUMER-GRADE LAPTOP WITH DUCKDB

## Processing Speed Comparison

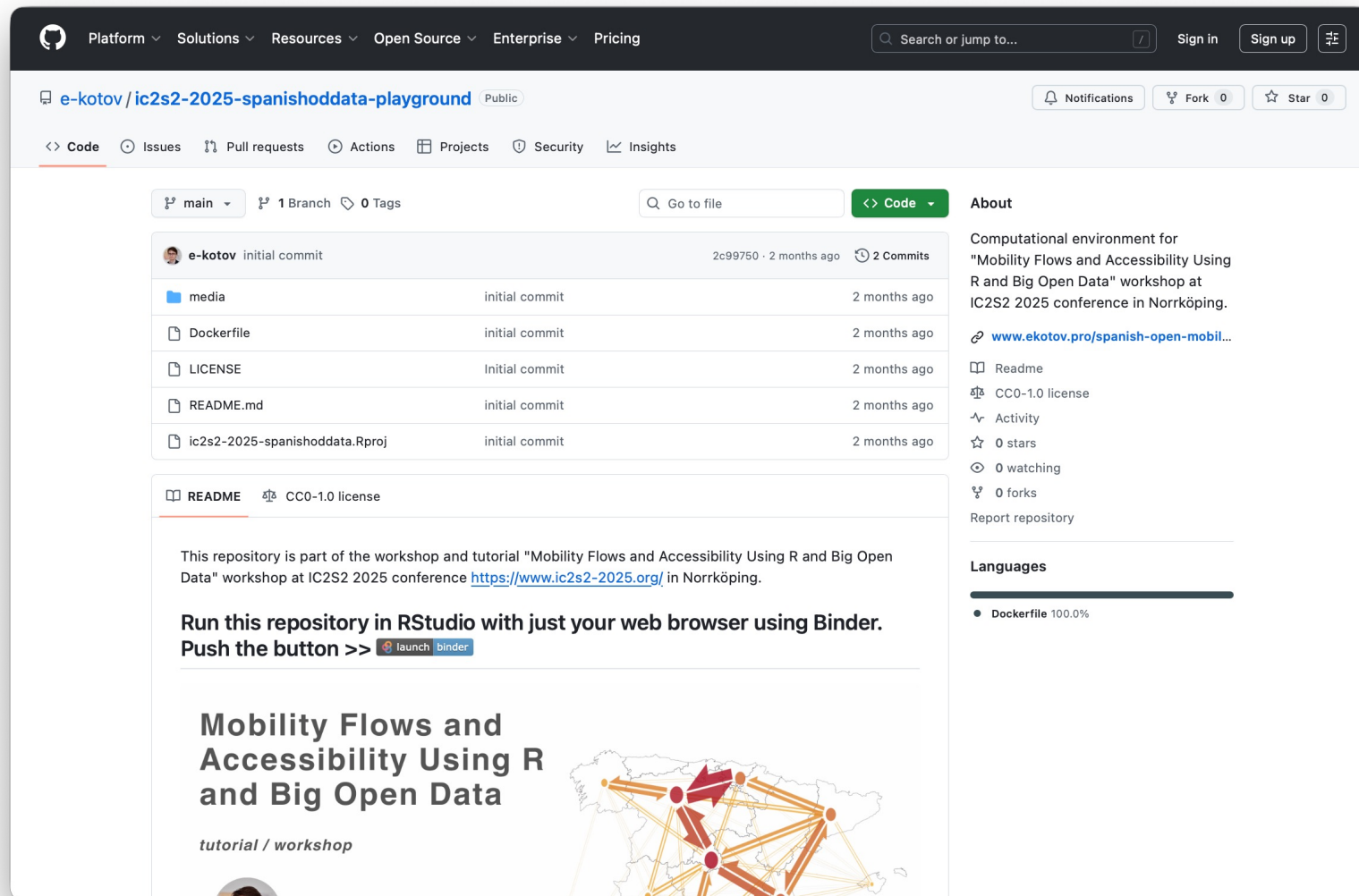
mean hourly trips by weekday over 18 months for each OD pair



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**

# INTERACTIVE PLAYGROUND IN WEB BROWSER



The screenshot shows a GitHub repository page for 'e-kotov/ic2s2-2025-spanishoddata-playground'. The repository is public and has 2 commits. The file list includes 'media', 'Dockerfile', 'LICENSE', 'README.md', and 'ic2s2-2025-spanishoddata.Rproj'. The README section describes the repository as part of a workshop and tutorial for 'Mobility Flows and Accessibility Using R and Big Open Data' at the IC2S2 2025 conference. It includes a link to the workshop website and a button to launch the repository in RStudio using Binder. The repository is licensed under CC0-1.0.

**Repository Details:**

- Repository: e-kotov / ic2s2-2025-spanishoddata-playground
- Public
- 2 Commits
- 1 Branch
- 0 Tags

**Files:**

| File                           | Commit         | Time         |
|--------------------------------|----------------|--------------|
| media                          | initial commit | 2 months ago |
| Dockerfile                     | initial commit | 2 months ago |
| LICENSE                        | Initial commit | 2 months ago |
| README.md                      | initial commit | 2 months ago |
| ic2s2-2025-spanishoddata.Rproj | initial commit | 2 months ago |

**About:**

Computational environment for "Mobility Flows and Accessibility Using R and Big Open Data" workshop at IC2S2 2025 conference in Norrköping.

[www.ekotov.pro/spanish-open-mobil...](http://www.ekotov.pro/spanish-open-mobil...)

**Languages:**

- Dockerfile 100.0%

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



# INTERACTIVE PLAYGROUND IN WEB BROWSER

Thanks to [GESIS](#) and [2i2c](#) for supporting us! 🎉  
mybinder.org has updated the base image to Ubuntu 22.04! See the [upgrade guide](#) for details.

♥ [Donate to mybinder.org!](#)



Launching your Binder...

The Binder team has [a site reliability guide](#) that talks about what it is like to run a BinderHub.

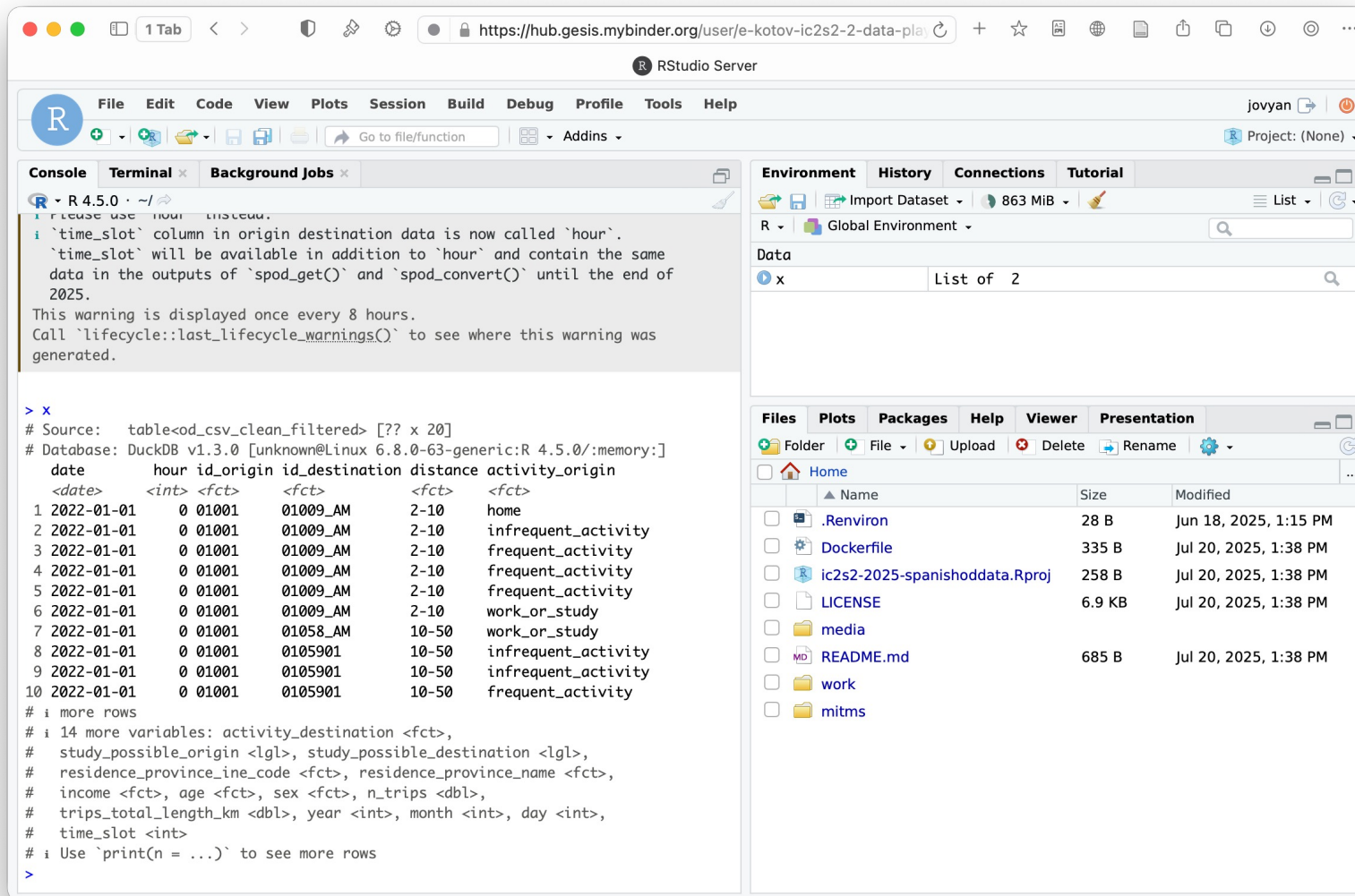
Waiting Building Pushing Launching

Build Logs [hide](#) [view raw](#)

```
Found built image, launching...
Launching server...
Server requested
2025-09-16T21:25:27.614448Z [Normal] Successfully assigned hetzner-geis/ju
pyter-e-kotov-ic2s2-2-data-playground-puoqn90c to svhz-css-mybinder-01
2025-09-16T21:25:28Z [Normal] Container image "jupyterhub/mybinder.org-tc-i
nit:2020.12.4-0-dev.git-5524-ba41b617b" already present on machine
```

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

# INTERACTIVE PLAYGROUND IN WEB BROWSER



RStudio Server

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Project: (None)

Console

```
R - R 4.5.0 - ~/
Please use 'hour' instead.
i 'time_slot' column in origin destination data is now called 'hour'.
'time_slot' will be available in addition to 'hour' and contain the same
data in the outputs of 'spod_get()' and 'spod_convert()' until the end of
2025.
This warning is displayed once every 8 hours.
Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
generated.

> x
# Source:   table<od_csv_clean_filtered> [?? x 20]
# Database: DuckDB v1.3.0 [unknown@Linux 6.8.0-63-generic:R 4.5.0/:memory:]
  date      hour id_origin id_destination distance activity_origin
  <date>    <int> <fct>    <fct>          <fct>    <fct>
1 2022-01-01 0 01001    01009_AM      2-10     home
2 2022-01-01 0 01001    01009_AM      2-10     infrequent_activity
3 2022-01-01 0 01001    01009_AM      2-10     frequent_activity
4 2022-01-01 0 01001    01009_AM      2-10     frequent_activity
5 2022-01-01 0 01001    01009_AM      2-10     frequent_activity
6 2022-01-01 0 01001    01009_AM      2-10     work_or_study
7 2022-01-01 0 01001    01058_AM      10-50     work_or_study
8 2022-01-01 0 01001    0105901       10-50     infrequent_activity
9 2022-01-01 0 01001    0105901       10-50     infrequent_activity
10 2022-01-01 0 01001    0105901       10-50     frequent_activity
# i more rows
# i 14 more variables: activity_destination <fct>,
#   study_possible_origin <lgl>, study_possible_destination <lgl>,
#   residence_province_ine_code <fct>, residence_province_name <fct>,
#   income <fct>, age <fct>, sex <fct>, n_trips <dbl>,
#   trips_total_length_km <dbl>, year <int>, month <int>, day <int>,
#   time_slot <int>
# i Use 'print(n = ...)' to see more rows
>
```

Environment

History

Connections

Tutorial

R - Global Environment

Data

x List of 2

Files

Plots

Packages

Help

Viewer

Presentation

Folder

File

Upload

Delete

Rename

Home

| Name                           | Size   | Modified              |
|--------------------------------|--------|-----------------------|
| .Renvi...                      | 28 B   | Jun 18, 2025, 1:15 PM |
| Dockerfile                     | 335 B  | Jul 20, 2025, 1:38 PM |
| ic2s2-2025-spanishoddata.Rproj | 258 B  | Jul 20, 2025, 1:38 PM |
| LICENSE                        | 6.9 KB | Jul 20, 2025, 1:38 PM |
| media                          |        |                       |
| README.md                      | 685 B  | Jul 20, 2025, 1:38 PM |
| work                           |        |                       |
| mitms                          |        |                       |

# DEMO



The screenshot shows the GitHub interface for the repository 'spanishoddata-playground' by user 'e-kotov'. The repository is public and has 0 forks and 0 stars. The main branch is 'main'. The repository contains several files: 'Dockerbuild', '.gitignore', 'Dockerfile', 'LICENSE', and 'README.md'. The 'README' file is selected, showing the title 'spanishoddata-playground' and the license 'CC0-1.0 license'. The repository description is 'No description, website, or topics provided.' The repository has 0 stars, 1 watching, and 0 forks. There are no releases published.

| File        | Commit Message             | Commit Date   |
|-------------|----------------------------|---------------|
| Dockerbuild | update build container def | 3 months ago  |
| .gitignore  | Initial commit             | 10 months ago |
| Dockerfile  | rebuild with latest        | 3 months ago  |
| LICENSE     | Initial commit             | 10 months ago |
| README.md   | add binder linke to readme | 10 months ago |

```
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

## Dominance of Female Commuters in Urban Destinations Across Spain

### Commuters

- More Female Trips
- More Male Trips
- No Data Available

Author: Egor Kotov | #30DayMapChallenge | Day 1: Points  
R package to get the data: <https://ropenspain.github.io/spanishoddata/>  
Original data source: Ministry of Transport and Sustainable Mobility of Spain; Nommon

## Hasta el 2% de la Movilidad Diaria Nacional Interrumpida por la DANA

Tamaño del municipio escalado según el número de viajes potencialmente afectados

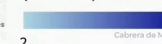
## Human mobility in Spain



### Daily Home-Work Trips (thoudands)



### Daily Other Trips (thoudands)



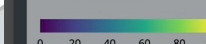
2023-10-25 como de costumbre

La DANA afecta la red móvil y podría reflejar desplazamientos residenciales en áreas afectadas

2024-10-31 después de la DANA

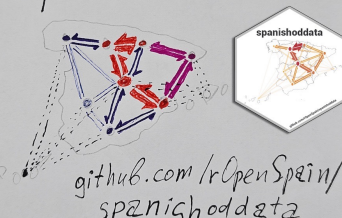
% de personas que posiblemente no permanecieron en su lugar de residencia

- la medición puede verse afectada por interrupciones en la red
- valores atípicos con el 100% eliminados



En solidaridad con las víctimas y sus familias  
Autor: Egor Kotov | #30DayMapChallenge | Día 16: Mapa Coroplético  
Paquete R para acceder a los datos: [ropenspain.github.io/spanishoddata/](https://ropenspain.github.io/spanishoddata/)  
Visualización del mapa realizada con mapgl: [walker-data.com/mapgl/](https://walker-data.com/mapgl/)  
Fuentes de datos: Ministerio de Transportes y Movilidad Sostenible (MITMS); Nommon

spanishoddata



[github.com/ropenspain/spanishoddata](https://github.com/ropenspain/spanishoddata)



# SUMMER 2025 WORKSHOPS

Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps  



-  Intro
-  Software setup
-  Slides
-  Tutorial: Get & Aggregate Mobility Data
-  Tutorial: Flowmaps

## Workshop Introduction

AUTHOR

Egor Kotov  

AFFILIATIONS

Max Planck Institute for Demographic

Research 

Universitat Pompeu Fabra 

DOI

[10.5281/zenodo.15794849](https://doi.org/10.5281/zenodo.15794849) 

### Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps

tutorial / workshop



Egor Kotov



AGIT Conference - July 3rd, 2025 - Salzburg

This website provides supporting materials for the workshop **Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps** which is part of the [AGIT](https://www.agit-conference.org/)  2025 conference held in Salzburg, Austria on 3, 2025. This tutorial is [licensed under CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) .

Large-scale human mobility datasets provide unprecedented opportunities to analyze and gain insights into movement patterns. Such insights are critical fields ranging from transport planning and epidemiology to socio-spatial inequality research and climate-change mitigation. Until recently, access to human mobility data was a privilege of a few researchers. Thanks to cour

agit2025

CONFERENCE 

DATES

REGISTRATION

SUBMISSIONS

CONTACT

ARCHIVE 



PARIS  
LODRON  
UNIVERSITY  
SALZBURG



**AGIT Conference for Geoinformatics** July 2–3, 2025 | Salzburg  
Shaping Geospatial Futures

## AGIT 2025

We are happy to announce that AGIT is back in 2025. It will take place from the 2nd to 3rd of July 2025 at [the Faculty of Natural and Life Sciences in Salzburg](https://www.unsalzburg.at/en/faculty-of-natural-and-life-sciences/). The importance of the event in combination with the Salzburg location is manifold. As an established user forum for Geoinformatics, it was a beacon, but also an

# SUMMER 2025 WORKSHOPS

Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps



- Intro
- Software setup
- Slides
- Tutorial: Get & Aggregate Mobility Data
- Tutorial: Flowmaps

## Workshop Introduction

AUTHOR

Egor Kotov

AFFILIATIONS

Max Planck Institute for Demographic

Research

Universitat Pompeu Fabra

DOI

10.5281/zenodo.15794849

Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps

tutorial / workshop



Egor Kotov



AGIT Conference - July 3rd, 2025 - Salzburg

This website provides supporting materials for the workshop **Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps** which is part of the [AGIT](#) 2025 conference held in Salzburg, Austria on July 1-3, 2025. This tutorial is [licensed under CC BY 4.0](#).

Large-scale human mobility datasets provide unprecedented opportunities to analyze and gain insights into movement patterns. Such insights are critical for fields ranging from transport planning and epidemiology to socio-spatial inequality research and climate-change mitigation. Until recently, access to human mobility data was a privilege of a few researchers. Thanks to countries

Registration/Anmeldeliste Workshop

10E "Analysing massive open human mobility data in R using spanishoddata, duckdb and flowmaps"

HS-102 (Lab), Donnerstag 03.07.2025, 14:45-16:00

A1052

1. Christian Stefan
2. HASITZKA Carina
3. Milena Bremer
4. Max Klump
5. Carl Markert
6. Tomi David
7. Kluger Zakir
8. Sarah Fournier
9. Mitah Joshi
10. Hsing-Hsuan Chen
11. Muhammad Bilal
12. Parinada Parinada
13. Angelika Kante
14. Bolla Hironova
15. Maria Nieuwand
16. Qiang Gu
17. ~~Maximilian Kranabetter~~ Maximilian Kranabetter
18. MONIKA SEN
19. Emel Barbagyhor
20. BARCHYBER Stijpke
- 21.
- 22.
- 23.
- 24.
- 25.
- 26.
- 27.
- 28.
- 29.
- 30.



# SUMMER 2025 WORKSHOPS

## Mobility Flows and Accessibility Using R and Big Open Data

🔍

🔧 Workshop Introduction

🔧 Software Setup

1a. 📄 Slides: Getting and aggregating the data

1b. 🇮🇹 Tutorial: Getting and aggregating the data

2a. 📄 Slides: Flows Visualisation

2b. 🇮🇹 Tutorial: Flows Visualisation

3a. 📄 Slides: Accessibility vs Mobility

3b. 🇮🇹 Tutorial: Accessibility vs Mobility

## 👋 Workshop Introduction

AUTHORS  
[Egor Kotov](#)  

[Johannes Mast](#)  

AFFILIATIONS  
[Max Planck Institute for Demographic Research](#)   
[Universitat Pompeu Fabra](#)   
[The German Aerospace Center \(Deutsches Zentrum für Luft- und Raumfahrt, DLR\)](#)   
[Julius-Maximilians-Universität Würzburg \(JMU\)](#) 

DOI  
[10.5281/zenodo.16238597](https://doi.org/10.5281/zenodo.16238597) 

### Mobility Flows and Accessibility Using R and Big Open Data

tutorial / workshop

 [Egor Kotov](#)

 [Johannes Mast](#)

IC2S2 Conference · July 21, 2025 · Norrköping

This website provides supporting materials for the workshop **Mobility Flows and Accessibility Using R and Big Open Data** which is part of the [11th International Conference on Computational Social Science](#)  held in Norrköping, Sweden July 21-24, 2025. This tutorial is [licensed under CC BY SA 4.0](#) .

Large-scale human mobility datasets provide unprecedented opportunities

On this page

[Projected Timeline](#)

[HOME](#) [SUBMIT](#) [ATTEND](#) [PROGRAM](#) [ORGANIZERS](#) [ABOUT](#)

# IC<sup>2</sup>S<sup>2</sup>

## NORRKÖPING 2025

ABOUT

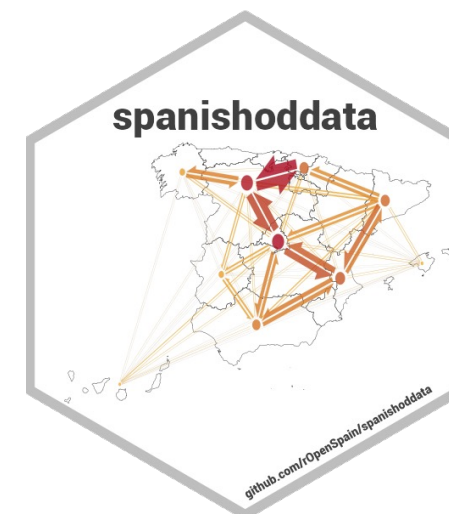
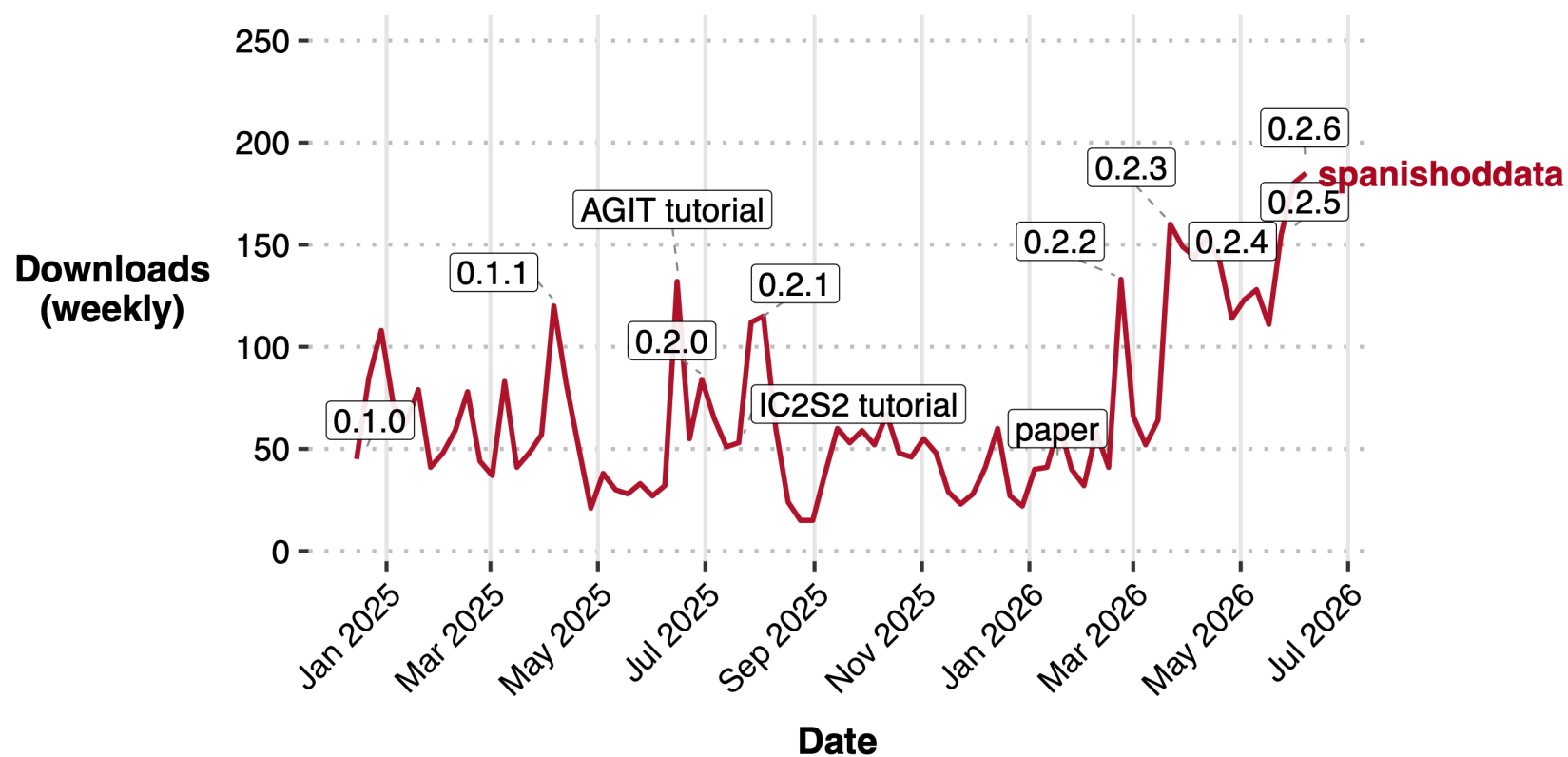
### 11<sup>th</sup> International Conference on Computational Social Science

Norrköping, Sweden | July 21-24, 2025

# 5k+ DOWNLOADS OVER PAST 18 MONTHS

## CRAN Downloads of spanishoddata

Weekly downloads from 2024-12-18 to 2026-06-07



repo status **Active**

lifecycle **stable**

CRAN **0.2.6**

downloads **5394**

downloads **680/month**

R-CMD-check **passing**

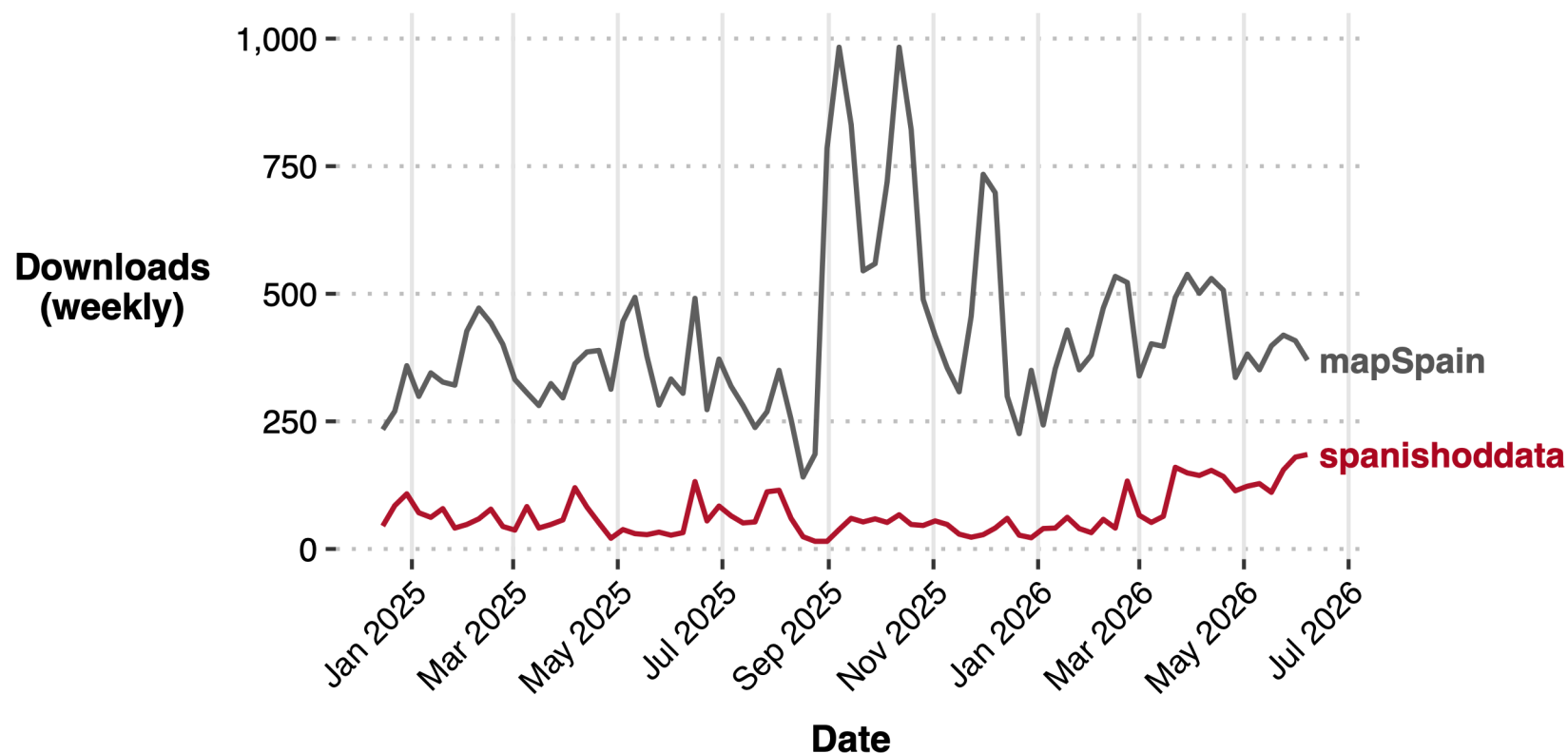
DOI **10.32614/CRAN.package.spanishoddata**

DOI **10.5281/zenodo.14516104**

# MAPSPAIN DOWNLOADS OVER LAST 18 MONTHS

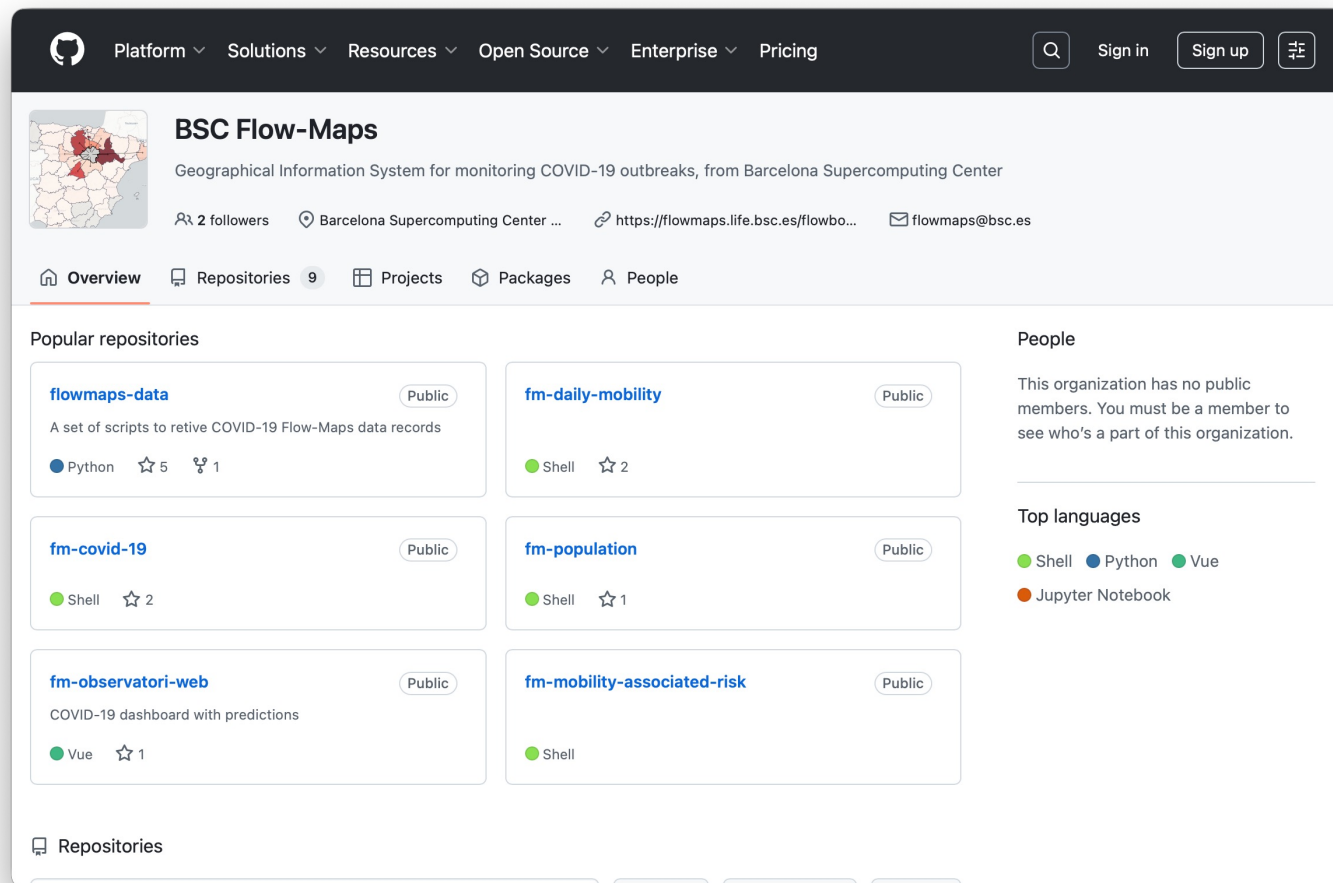
## CRAN Downloads: spanishoddata vs mapSpain

Weekly downloads from 2024-12-18 to 2026-06-07



- **mapSpain** is an interface to Spanish spatial data such as province, municipality and census boundaries.
- With 32k downloads, only 6 times more popular than much more specialized **spanishoddata** in the same period.

# BSC FLOW-MAPS



The screenshot shows the GitHub repository page for BSC Flow-Maps. The repository is described as a "Geographical Information System for monitoring COVID-19 outbreaks, from Barcelona Supercomputing Center". It has 2 followers and is located at the Barcelona Supercomputing Center. The repository is public and has a website link: <https://flowmaps.life.bsc.es/flowbo...> and an email: [flowmaps@bsc.es](mailto:flowmaps@bsc.es).

The repository is categorized under "Overview", "Repositories" (9), "Projects", "Packages", and "People".

**Popular repositories**

- flowmaps-data** (Public): A set of scripts to retrieve COVID-19 Flow-Maps data records. Languages: Python (5 stars, 1 fork).
- fm-daily-mobility** (Public): Languages: Shell (2 stars).
- fm-covid-19** (Public): Languages: Shell (2 stars).
- fm-population** (Public): Languages: Shell (1 star).
- fm-observatori-web** (Public): COVID-19 dashboard with predictions. Languages: Vue (1 star).
- fm-mobility-associated-risk** (Public): Languages: Shell.

**People**

This organization has no public members. You must be a member to see who's a part of this organization.

**Top languages**

- Shell
- Python
- Vue
- Jupyter Notebook

**Repositories**

# 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
```

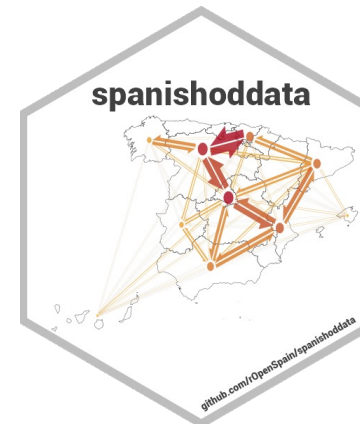


# SPANISHODDATA R PACKAGE

<https://ropenspain.github.io/spanishoddata/>

Use `{spanishoddata}`  
package

```
1 library(spanishoddata)
2 spod_set_data_dir("data")
3
4 od_data <- spod_get(
5   type = "origin-destination",
6   zones = "districts",
7   dates = c(
8     start = "2022-03-01",
9     end = "2022-03-07"
10  )
11 )
```



CRAN 0.2.1

downloads 2282

downloads 107/month

R-CMD-check passing

DOI 10.32614/CRAN.package.spanishoddata

DOI 10.5281/zenodo.14516104

# MAPGL WITH FLOWS SUPPORT COMING SOON

mapgl 0.4.9.9000

Reference

Articles

Changelog

Search for

## Getting started with mapgl

Source: [vignettes/getting-started.Rmd](#)

### Using Mapbox GL JS

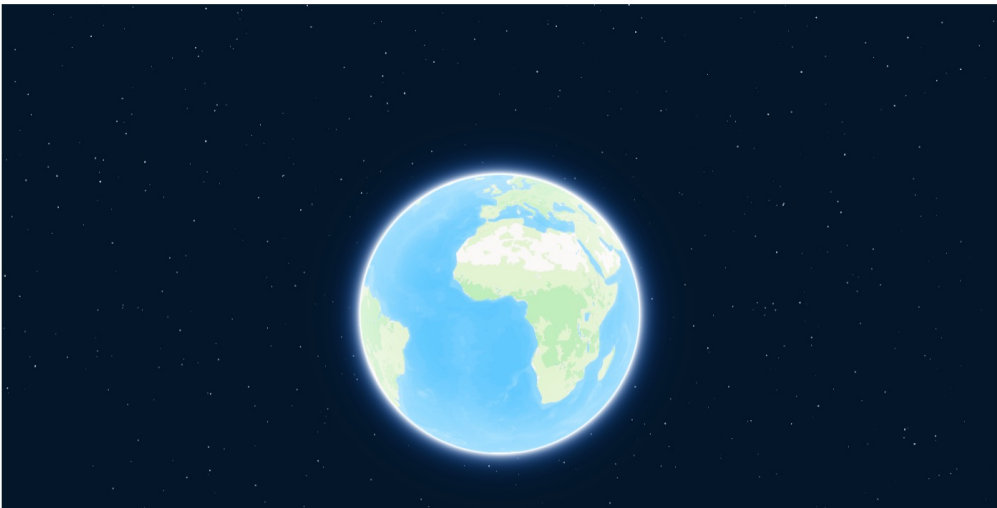
Your gateway to Mapbox GL JS v3 in R is the `mapboxgl()` function. Run the function with no arguments to get an interactive globe using Mapbox's Standard style:

```
library(mapgl)  
  
mapboxgl()
```



### On this page

- Using Mapbox GL JS
- Using Maplibre GL JS
- Quick data visualization with view functions
- Comparing map views



# MAPGL WITH FLOWS SUPPORT COMING SOON



# DUCKSPATIAL R PACKAGE – SAME ENGINE AS IN SPANISHODDATA

**duckspatial** 1.1.1 [Get Started](#) [Reference](#) [Articles](#) [News](#)

Search for

## Benchmark

2026-06-13

Source: [vignettes/benchmark.qmd](#)

This vignette benchmarks **{duckspatial}** against **{sf}** across several spatial operations, comparing computation time and memory usage as dataset size grows. We plan to extend it with additional operation types in future releases. Versions of packages used in this vignette:

- duckspatial** : v1.0.0
- duckdb** : v1.5.1
- sf** : v1.1.0

### TL;DR

**{duckspatial}** is substantially faster and allocates far less memory than **{sf}** in almost all cases, with the advantage becoming more pronounced on larger datasets. The one exception is pairwise distance calculation, where **{sf}** retains a memory advantage for Euclidean distances.

### Prepare data

We are going to test the speed of a bunch of functions using simulated data. We use the `make_points()` function to generate `n` random points within the globe, and 10,000 random rectangles.

► Set-up

### On this page

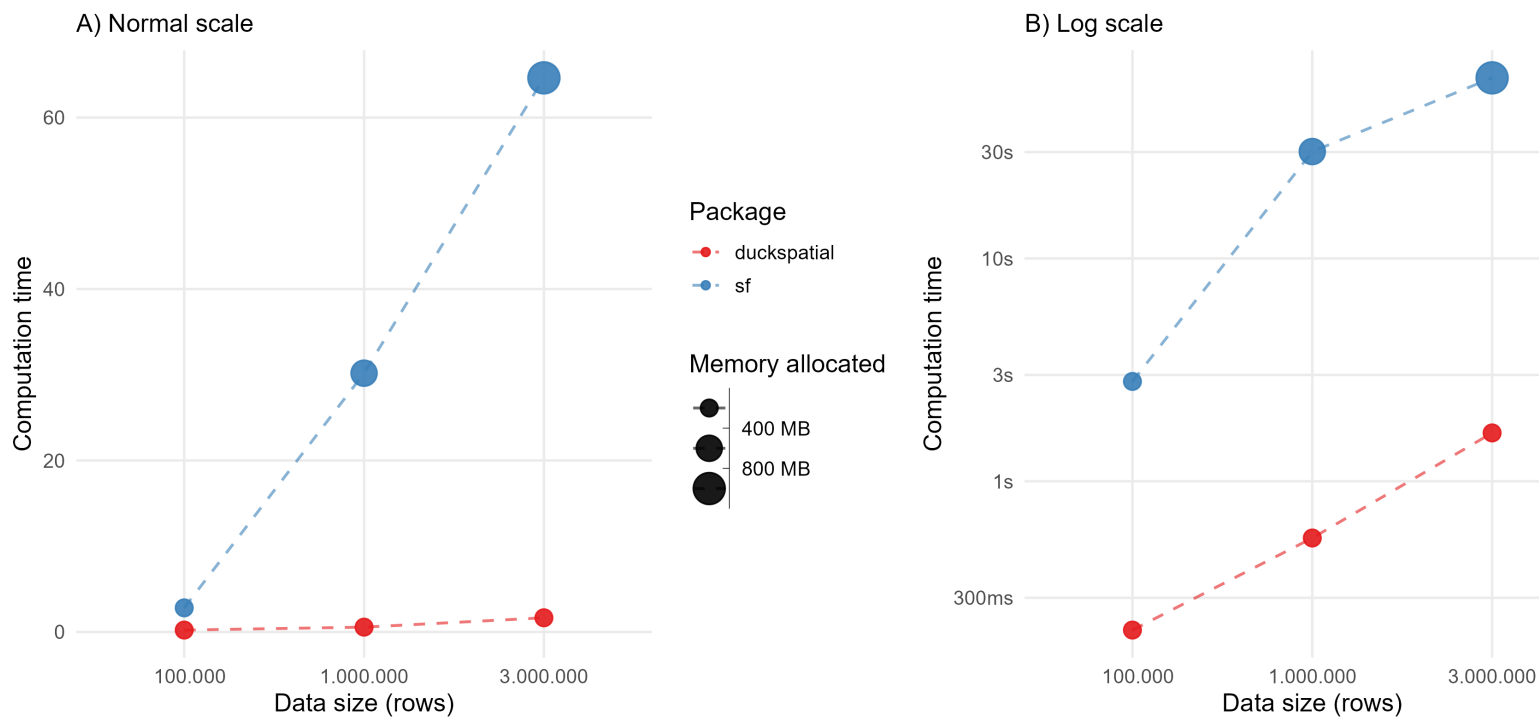
- TL;DR**
- Prepare data
- Spatial join
- Spatial filter
- Spatial distances
- Dissolving geometries
- Geometry predicates



# DUCKSPATIAL R PACKAGE – SAME ENGINE AS IN SPANISHODDATA

## Benchmark Comparison for ddbb\_join()

Execution time vs. data size







Egor Kotov

Human mobility researcher

R package developer



 [ekotov.pro](https://ekotov.pro)



and more...