Get your dataset ready!

Using R and GIS

Rosa Félix Gabriel Valença Rafael Pereira

2024-09-04

Table of contents

1 Introduction

This course aims to provide tools to deal with exploring and treating transportation datasets using R programming, an open-source and widely used tool for data analytics in urban mobility.

Additionally, this course provides guidance towards the use of reproducible methods to deal with large datasets that require manipulation and/or spatial analysis.

The course has a **hands-on** approach, where participants will learn the basics of **coding**, **data manipulation**, and **spatial analysis** for urban mobility and transportation.

1.1 Mobility data

There is an emerging increase in mobility data, through new forms of technology, which result in very large and diverse datasets.

Knowing how to get, treat and analyze complex datasets with the up-to-date technologies is extremely relevant for academia, policy makers and start-ups, since it allows them to:

1. acquire critical view on urban mobility based on data;
2. spatially identify locations in the city that require policy priorities;
3. and improve the efficiency of data analysis processes.

Why R and GIS

Most academic programs focus on teaching modelling and deep analysis of data. However, there is a need to learn how to explore and prepare a dataset for modelling. The use of **programming and GIS** techniques have enormous advantages, including their flexibility; reproducibility; and transparency and understanding the step-by-step process.

The use of GIS techniques in transportation is, traditionally, not considered in transportation learning programs, despite being of enormous relevance when doing accessibility analysis or reeling with georreferenced transportation data, such as bike sharing route trips' datasets, origin-destination flows datasets, home/work locations, GTFS public transit data, and so on. There is a need to learn how to locate these open datasets, how to explore them and

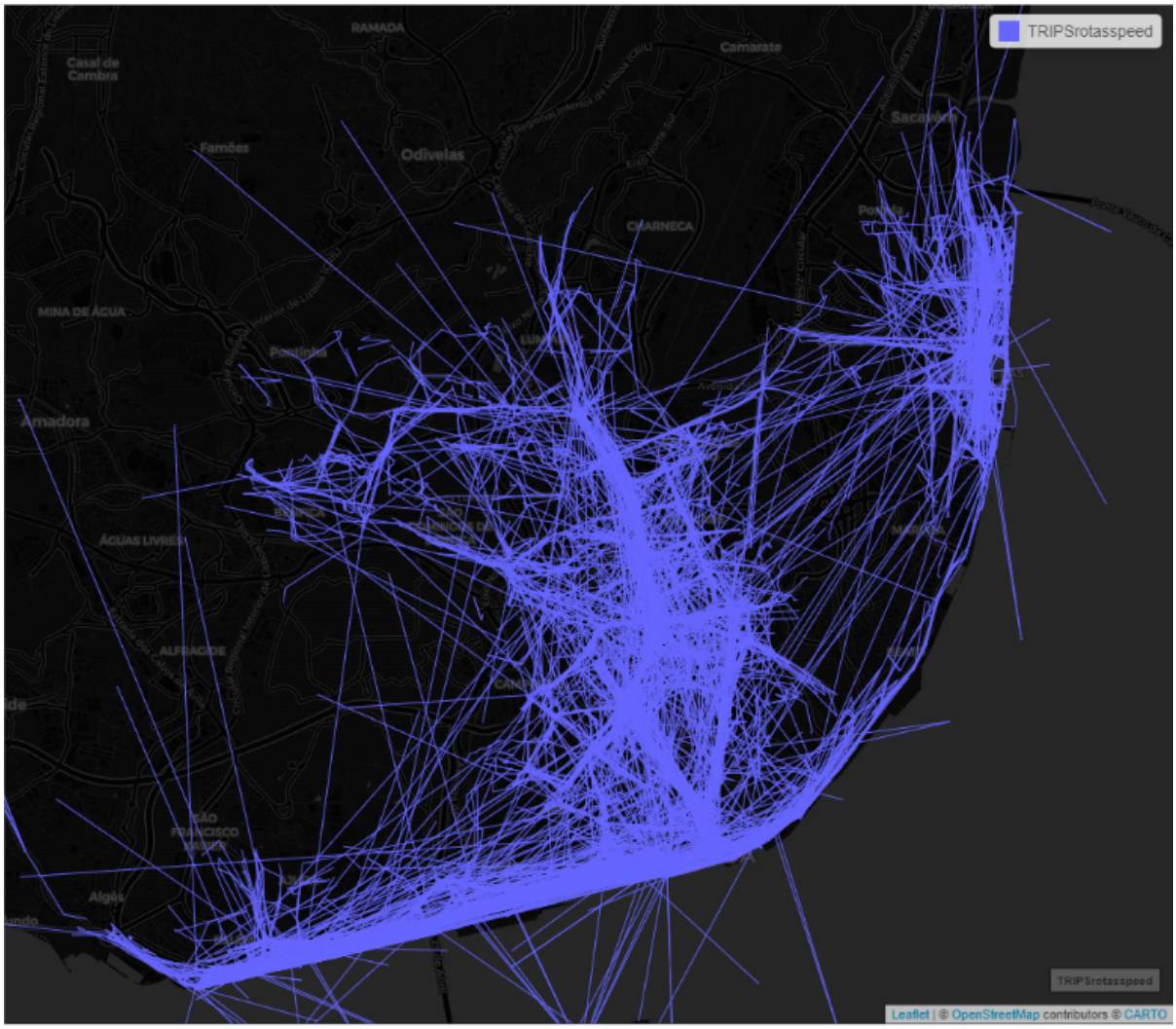


Figure 1.1: E-Scooter trip data in Lisbon. How to deal with it?

how to integrate them into transportation and urban analysis. Additionally, the use of open source software and datasets allows researchers to perform methods that are reproducible and transparent.

TLDR

- Open-source tools widely used in data analytics and spatial analysis
- Flexibility and reproducibility in data manipulation and visualization
- Critical for urban mobility and transportation research, with spatial relevance
- Large transportation datasets are becoming increasingly common

1.2 Course objectives

Introduce R Programming Basics

- Equip participants with foundational skills in R programming
- Emphasize reproducible research practices to ensure transparency and replicability in analyses

Teach Data Manipulation Techniques

- Use key R packages for data cleaning, manipulation, and summarization of datasets
- Enable participants to efficiently handle large and complex transportation datasets

Spatial Data Visualization

- Introduce methods for quick and effective spatial data visualization using R and GIS tools
- Provide hands-on experience with creating interactive maps and visualizations

Perform Basic Spatial Analysis

- Teach participants how to perform spatial analysis of transportation datasets using GIS techniques with R
- Cover practical applications such as georeferencing data, accessibility analysis, and routing ODs
- Utilize real-world transportation data for practical, hands-on learning

1.3 Target audience

- Ph.D. candidates from DTN and other researchers
- Policy makers and practitioners in urban mobility
- Beginners to intermediate R users, no prior experience needed

1.4 Recommended readings

- Engel, Claudia A. (2023) Introduction to R
- Lovelace, Robin, Nowosad, Jakub & Muenchow, Johannes. (2023) Geocomputation with R
- Pereira, Rafael H. M. & Herszenhut, Daniel. (2023) Introduction to urban accessibility: a practical guide with R. Ipea - Institute of Applied Economic Research

2 Course Structure

The course consists of an in-person 2-day course, taking place during the EIT DTN Annual Meeting on the **19th and 20th September 2024**.

The first day will focus on learning the basics of R programming and how to treat and explore datasets. The second day will focus on analyzing spatial datasets, and routing origins to destinations.

2.1 Day 1

Morning

- Introduction to **programming** techniques and **data structures**
- Introduction to R, and RStudio: **software installation** and main packages
- **R base and basics**: examples and exercises

Afternoon

- **Data manipulation**: using the dplyr package to select, filter, left-join, group and summarize
- Introduction to **GIS** and **spatial data**: import and visualize vector data
- R markdown and **interactive maps**

2.2 Day 2

Morning

- **Desire lines** from OD and transport zones
- **Georeference** coordinates: examples from surveys
- **Accessibility analysis**: from buffers to road networks

Afternoon

- **Open Transportation data**: where to find it
- **Routing with R**: multimodal and intermodal (*r5r demo* - Rafael Pereira)
- Group **exercise**

3 Detailed schedule (TBC)

Day 1	
9.30	Introductions and Presentation of the course contents
10.00	Introduction to programming techniques and data structures
10.30	Introduction to R and RStudio: hands-on to install software and main packages
11.00	*Coffee break*
11.15	(cont.)
11.30	R basics: examples and exercises
12.30	*Lunch break*
13.30	Data manipulation: examples and exercises (select, filter, left-join, subset, group and summarize, using dplyr package)
15.00	Introduction to GIS and spatial data: import and visualize vector data
15.30	*Coffee break*
15.45	(cont.)
16.15	View and export interactive maps
17.00	*End of day 1*

Day 2	
9.30	Desire-lines from OD pairs and transport zones: examples and exercises
10.30	Georeferenced coordinates from survey responses: example and exercises
11.00	*Coffee break*
11.15	(cont.)
11.30	Euclidean distance and buffers: example and exercises
12.30	*Lunch break*
13.30	Open Transportation data: where to find it (OSM and GTFS)

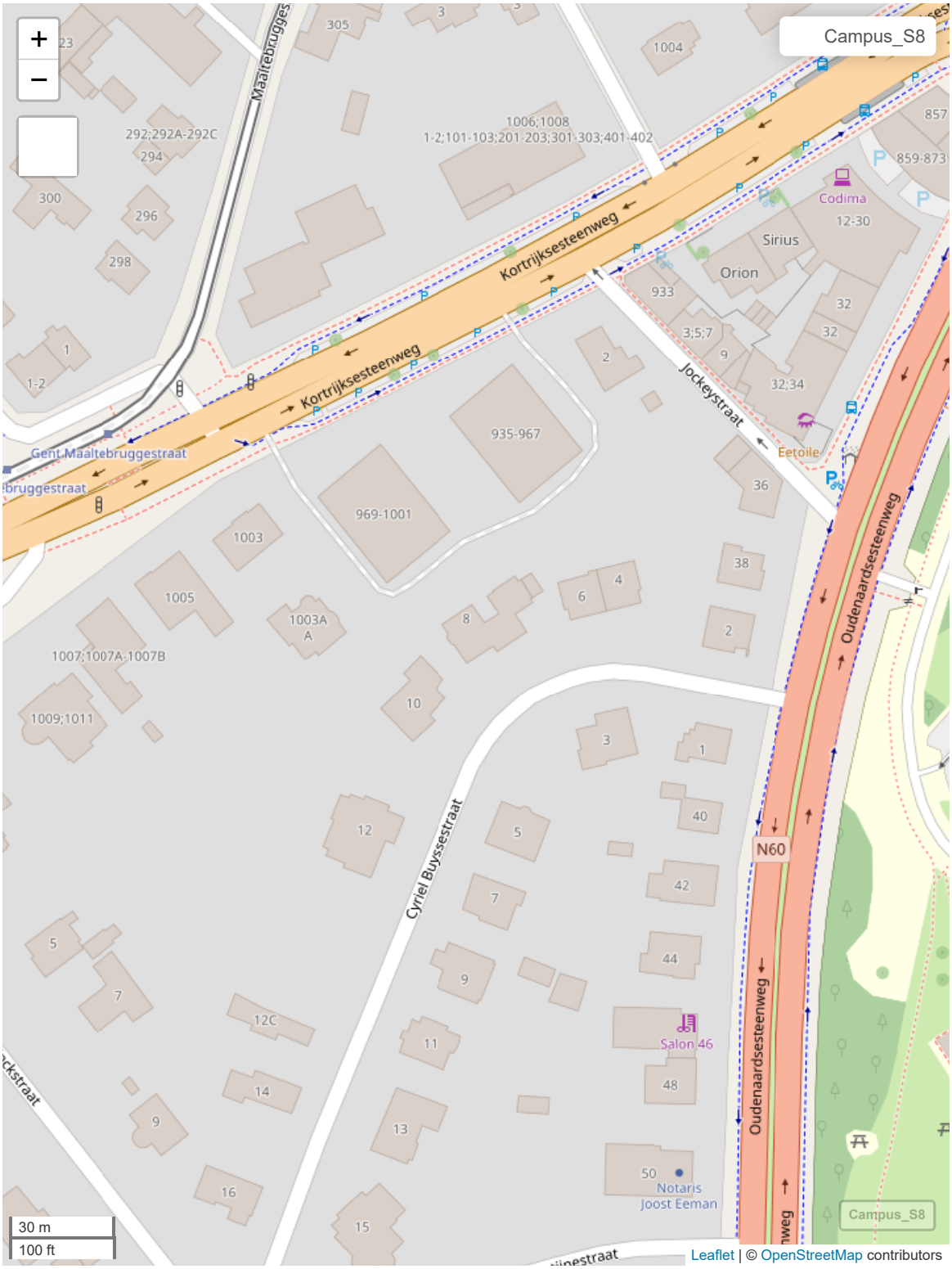
Day 2	
14.30	Uni-modal and Inter-modal Routing with r5r
15.30	Accessibility analysis with r5r
16.00	*Coffee break*
16.15	Using you data: manipulation and spatial analysis methods and further applications
16.45	Survey and feedback from participants
17.00	*End of day 2*

4 Location

The course will take place at Campus Sterre, Building S8, room 2.4.

```
Campus_S8_coord = c(3.7105372, 51.0241258)
Campus_S8 = sf::st_sfc(sf::st_point(Campus_S8_coord)) # create point
Campus_S8 = sf::st_as_sf(Campus_S8, crs = 4326) # assign crs

mapview::mapview(Campus_S8, map.types = "OpenStreetMap") # quick map view
```


Campus_S8
Kortrijksesteenweg
Kortrijksesteenweg
Jockeystraat
Oudenaardsesteenweg
Oudenaardsesteenweg
N60
Cyriel Buyssestraat
Gent Maaltebruggestraat
Codima
12-30
Sirius
Orion
Eetoile
Salon 46
Notaris
Joost Eeman
1006;1008
1-2;101-103;201-203;301-303;401-402
292;292A-292C
935-967
969-1001
1003
1005
1003A
A
1007;1007A-1007B
1009;1011
933
3;5;7
32;34
859-873
857
1004
305
300
296
298
294
1-2
30 m
100 ft
Leaflet | © OpenStreetMap contributors

5 Resources

- You laptop, with any OS
- Github repository with all the materials (data, code and guidelines)
- Survey datasets, school locations and public transport operator datasets

Part I

Day 1

6 Software

In this chapter we will guide you through the installation of R, RStudio and the packages you will need for this course.

R and **RStudio**[1] are separate downloads.

6.1 R

You will need **R** installed on your computer. **R stats** (how it is also known) is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing.

The download links live at The Comprehensive R Archive Network (aka CRAN). The most recent version is `4.4.1`, but you can use `>= 4.1.x` if you already have it installed.

6.1.1 Windows

Download R-4.4.1 for Windows and run the executable file.

You will also need to install Rtools, which is a collection of tools necessary to build R packages in Windows.

6.1.2 Mac

Download R-4.4.1 for MacOX. You will have to choose between the arm64 or the x86-64 version.

Download the `.pkg` file and install it as usual.

[1] We will use RStudio, although if you already use other studio such as VScode, that's also fine.

6.1.3 Ubuntu

> These are instructions for Ubuntu. If you use other linux distribution, please follow the instructions on The Comprehensive R Archive Network - CRAN.

You can look for R in the Ubuntu **Software Center** or install it via the terminal:

```
# sudo apt update && sudo apt upgrade -y
sudo apt install r-base
```

Or, if you prefer, you can install the latest version of R from CRAN:

```
# update indices
sudo apt update -qq
# install two helper packages we need
sudo apt install --no-install-recommends software-properties-common dirmngr
# add the signing key (by Michael Rutter) for these repos
wget -qO- https://cloud.r-project.org/bin/linux/ubuntu/marutter_pubkey.asc | sudo tee -a /
# add the R 4.0 repo from CRAN -- adjust 'focal' to 'groovy' or 'bionic' as needed
sudo add-apt-repository "deb https://cloud.r-project.org/bin/linux/ubuntu $(lsb_release -c
```

Then run:

```
sudo apt install r-base r-base-core r-recommended r-base-dev
```

[Optional] To keep up-to-date r version and packages, you can follow the instructions at r2u

After this installation, you don't need to open R base. Please proceed to install RStudio.

6.2 RStudio

RStudio Desktop is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

RStudio is available for free download from Posit RStudio.

6.2.1 Windows 10/11

Download RStudio 2024.04 and run the executable file.

6.2.2 MacOS

Download RStudio 2024.04 and install it as usual.

6.2.3 Ubuntu

> These are instructions for Ubuntu **22** / Debian 12. If you use other linux distribution, please follow the instructions on Posit RStudio.

Install it via the terminal:

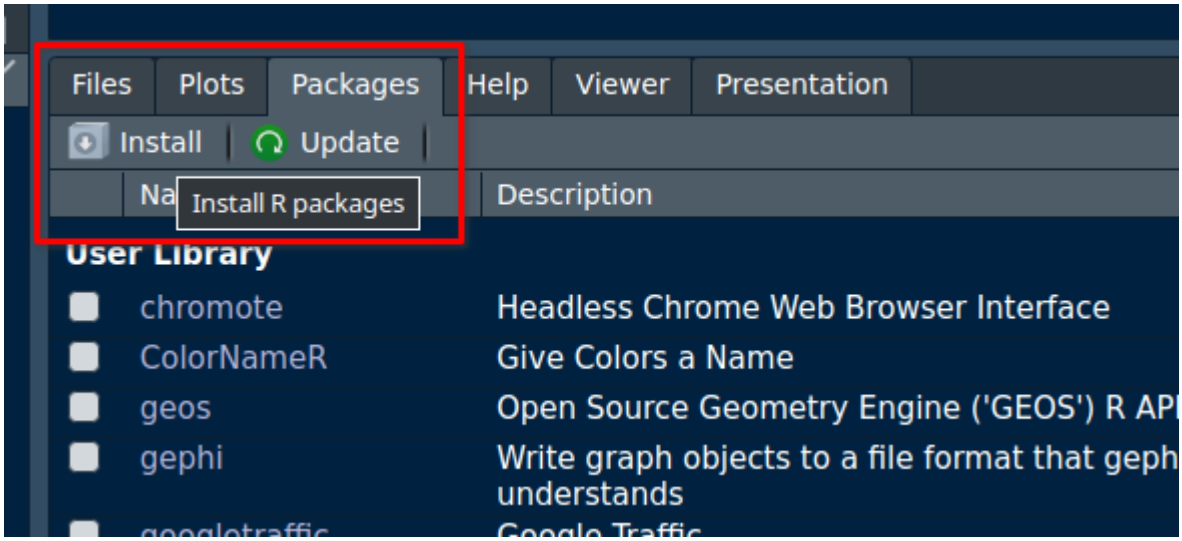
```
sudo apt install libssl-dev libclang-dev
wget https://download1.rstudio.org/electron/jammy/amd64/rstudio-2024.04.2-764-amd64.deb
sudo dpkg -i rstudio*
rm -v rstudio*
```

If you already use Ubuntu **24**, please check and replace the correct url from RStudio Dailies

6.3 R packages

You will need to install some packages to work with the data and scripts in this course.

You can install them in RStudio by searching for them in the **Packages** tab:



or by running the following code in the console:

```r
install.packages("tidyverse")
install.packages("readxl")

install.packages(c("remotes", "devtools", "usethis"))
install.packages("sf")
install.packages("mapview")
```

6.4 r5r

6.4.1 Java Development Kit

You will need this to work with the **r5r** package. It is also known as JDK 21.

6.4.2 Windows and MacOS

Go to Java Development Kit 21, download the latest 21 build corresponding to your operating system and run the executable file.

6.4.3 Ubuntu

Install it via the terminal:

```
sudo apt install -y openjdk-21-jdk openjdk-21-jre
java -version
```

Then, in R you will also need rJava package.

```r
install.packages("rJava")
```

6.4.4 r5r

Install the **r5r** package:

```r
install.packages("r5r")

# development version:
# devtools::install_github("ipeaGIT/r5r", subdir = "r-package")
```

6.4.5 Open Route Service

Sign up for an account and create a token. Copy your API.

In RStudio console, run:

```
# install.packages("openrouteservice")
openrouteservice::ors_api_key("YOUR-API-KEY")
```

This will store your key on your `.Renviron` file, meaning that every time you open RStudio, you won't need to run this command again.

This is useful also to write your `openrouteservice` scripts without sharing your key with others.

7 R basics

In this chapter we will introduce to the R basics and some exercises to get familiar to how R works.

7.1 Simple operations

7.1.1 Math operations

7.1.1.1 Sum

```
1+1
```

```
[1] 2
```

7.1.1.2 Subtraction

```
5-2
```

```
[1] 3
```

7.1.1.3 Multiplication

```
2*2
```

```
[1] 4
```

7.1.1.4 Division

```
8/2
```

```
[1] 4
```

7.1.1.5 Round the number

```
round(3.14)
```

```
[1] 3
```

```
round(3.14, 1) # The "1" indicates to round it up to 1 decimal digit.
```

```
[1] 3.1
```

You can use help `?round` in the console to see the description of the function, and the default arguments.

7.1.2 Basic shortpaths

7.1.2.1 Perform Combinations

```
c(1, 2, 3)
```

```
[1] 1 2 3
```

```
c(1:3) # The ":" indicates a range between the first and second numbers.
```

```
[1] 1 2 3
```

7.1.2.2 Create a comment with `ctrl + shift + m`

```
# Comments help you organize your code. The software will not run the comment.
```

7.1.2.3 Create a table

A simple table with the number of trips by car, PT, walking, and cycling in a hypothetical street segment at a certain period.

Define variables

```
modes <- c("car", "PT", "walking", "cycling") # you can use "=" or "<-"
Trips = c(200, 50, 300, 150) # uppercase letters modify
```

Join the variables to create a table

```
table_example = data.frame(modes, Trips)
```

Take a look at the table

Visualize the table by clicking on the "Data" in the "Environment" page or use :

```
View(table_example)
```

7.2 Practical exercise

Dataset: the number of trips between all municipalities in the Lisbon Metropolitan Area, Portugal (Instituto National de Estatística 2018).

7.2.0.1 Import dataset

You can click directly in the file under the "Files" pan, or:

```
data = readRDS("data/TRIPSmode_mun.Rds")
```

> Note that after you type " you can use `tab` to navigate between folders and files and `enter` to autocomplete.

7.2.0.2 Take a first look at the data

Summary statistics

```
summary(data)
```

```
 Origin_mun        Destination_mun        Total             Walk
Length:315         Length:315         Min.   :     7   Min.   :     0
Class :character   Class :character   1st Qu.:   330   1st Qu.:     0
Mode  :character   Mode  :character   Median :  1090   Median :     0
                                      Mean   : 16825   Mean   :  4033
                                      3rd Qu.:  5374   3rd Qu.:     0
                                      Max.   :875144   Max.   :306289
     Bike              Car             PTransit            Other
Min.   :   0.00   Min.   :     0   Min.   :     0.0   Min.   :    0.0
1st Qu.:   0.00   1st Qu.:   263   1st Qu.:     5.0   1st Qu.:    0.0
Median :   0.00   Median :   913   Median :   134.0   Median :    0.0
Mean   :  80.19   Mean   :  9956   Mean   :  2602.6   Mean   :  152.4
3rd Qu.:   0.00   3rd Qu.:  4408   3rd Qu.:   975.5   3rd Qu.:   62.5
Max.   :5362.00   Max.   :349815   Max.   :202428.0   Max.   :11647.0
```

Check the structure of the data

```
str(data)
```

```
'data.frame':   315 obs. of  8 variables:
 $ Origin_mun     : chr  "Alcochete" "Alcochete" "Alcochete" "Alcochete" ...
 $ Destination_mun: chr  "Alcochete" "Almada" "Amadora" "Barreiro" ...
 $ Total          : num  20478 567 188 867 114 ...
 $ Walk           : num  6833 0 0 0 0 ...
 $ Bike           : num  320 0 0 0 0 0 0 0 91 0 ...
 $ Car            : num  12484 353 107 861 114 ...
 $ PTransit       : num  833 0 81 5 0 ...
 $ Other          : num  7 214 0 0 0 0 0 0 0 0 ...
```

Check the first values of each variable

```
data # first 10 values
```

```
             Origin_mun     Destination_mun  Total   Walk Bike    Car PTransit
1             Alcochete           Alcochete  20478   6833  320  12484      833
2             Alcochete              Almada    567      0    0    353        0
3             Alcochete             Amadora    188      0    0    107       81
4             Alcochete            Barreiro    867      0    0    861        5
5             Alcochete             Cascais    114      0    0    114        0
6             Alcochete              Lisboa   2840     69    0   1994      775
7             Alcochete              Loures    634      0    0    634        0
```

8	Alcochete	Moita	261	0	0	256	5
9	Alcochete	Montijo	8714	130	91	7062	1431
10	Alcochete	Odivelas	129	0	0	129	0
11	Alcochete	Oeiras	130	0	0	130	0
12	Alcochete	Palmela	553	11	0	299	243
13	Alcochete	Seixal	118	0	0	58	0
14	Alcochete	Sesimbra	79	0	0	79	0
15	Alcochete	Setúbal	453	0	0	296	157
16	Alcochete	Sintra	573	0	0	524	25
17	Alcochete	Vila Franca de Xira	91	0	0	91	0
18	Almada	Alcochete	761	0	0	498	49
19	Almada	Almada	197248	60007	257	106852	27342
20	Almada	Amadora	1726	0	0	1331	393
21	Almada	Barreiro	1993	0	0	1972	21
22	Almada	Cascais	765	0	2	647	116
23	Almada	Lisboa	37180	370	218	23131	12322
24	Almada	Loures	1369	0	19	876	475
25	Almada	Mafra	15	0	0	15	0
26	Almada	Moita	464	0	0	225	239
27	Almada	Montijo	848	0	0	738	110
28	Almada	Odivelas	1133	0	0	1075	58
29	Almada	Oeiras	1854	0	0	1372	428
30	Almada	Palmela	538	0	0	511	21
31	Almada	Seixal	36018	3005	65	28500	3811
32	Almada	Sesimbra	2957	0	0	2611	216
33	Almada	Setúbal	2920	0	0	2528	283
34	Almada	Sintra	1278	0	0	536	731
35	Almada	Vila Franca de Xira	767	0	0	685	72
36	Amadora	Alcochete	190	0	0	108	81
37	Amadora	Almada	3207	0	0	2810	396
38	Amadora	Amadora	198980	108622	0	76436	13723
39	Amadora	Barreiro	212	0	0	141	71
40	Amadora	Cascais	3079	0	0	2957	113
41	Amadora	Lisboa	67532	2339	103	40353	24353
42	Amadora	Loures	6079	0	0	5334	489
43	Amadora	Mafra	517	0	0	488	28
44	Amadora	Moita	16	0	0	16	0
45	Amadora	Montijo	365	0	0	365	0
46	Amadora	Odivelas	11166	1120	0	8415	1630
47	Amadora	Oeiras	15385	768	516	11308	2757
48	Amadora	Palmela	131	0	0	131	0
49	Amadora	Seixal	2592	0	0	1950	642
50	Amadora	Sesimbra	192	0	0	104	88

51	Amadora	Setúbal	893	0	0	893	0
52	Amadora	Sintra	32926	3069	36	23667	5818
53	Amadora	Vila Franca de Xira	1090	270	0	728	27
54	Barreiro	Alcochete	855	0	0	849	5
55	Barreiro	Almada	1956	0	0	1639	317
56	Barreiro	Amadora	506	0	0	165	341
57	Barreiro	Barreiro	87630	32460	351	48667	5877
58	Barreiro	Cascais	35	0	0	0	35
59	Barreiro	Lisboa	10637	23	0	2674	7898
60	Barreiro	Loures	106	0	0	81	5
61	Barreiro	Moita	14990	2385	124	10847	1634
62	Barreiro	Montijo	1973	0	0	1937	37
63	Barreiro	Odivelas	81	0	0	81	0
64	Barreiro	Oeiras	601	0	0	467	134
65	Barreiro	Palmela	4658	0	0	3904	754
66	Barreiro	Seixal	2670	0	0	2483	186
67	Barreiro	Sesimbra	3800	62	0	3566	172
68	Barreiro	Setúbal	2118	0	0	1817	208
69	Barreiro	Sintra	923	0	0	923	0
70	Barreiro	Vila Franca de Xira	119	0	0	63	56
71	Cascais	Alcochete	114	0	0	114	0
72	Cascais	Almada	932	0	0	877	55
73	Cascais	Amadora	4021	0	0	3899	113
74	Cascais	Barreiro	35	0	0	0	35
75	Cascais	Cascais	268462	63849	5362	180633	17365
76	Cascais	Lisboa	35698	38	13	25985	9202
77	Cascais	Loures	2039	0	0	1997	42
78	Cascais	Mafra	976	0	0	976	0
79	Cascais	Moita	178	0	0	178	0
80	Cascais	Montijo	752	0	0	565	187
81	Cascais	Odivelas	3086	0	0	3086	0
82	Cascais	Oeiras	32698	970	401	28587	2642
83	Cascais	Palmela	299	0	0	299	0
84	Cascais	Seixal	1373	0	0	1006	354
85	Cascais	Sesimbra	83	0	0	83	0
86	Cascais	Setúbal	834	0	0	591	232
87	Cascais	Sintra	21105	48	16	18679	2269
88	Cascais	Vila Franca de Xira	894	0	0	894	0
89	Lisboa	Alcochete	2769	0	0	1953	816
90	Lisboa	Almada	33125	0	256	20519	11305
91	Lisboa	Amadora	68214	2185	73	41523	23754
92	Lisboa	Barreiro	8892	0	0	1875	7014
93	Lisboa	Cascais	35644	0	13	26516	8841

94	Lisboa	Lisboa	875144	306289	4954	349815	202428
95	Lisboa	Loures	81023	3376	195	51791	24911
96	Lisboa	Mafra	8815	0	0	8074	741
97	Lisboa	Moita	4550	0	0	1753	2798
98	Lisboa	Montijo	6179	0	0	4466	1712
99	Lisboa	Odivelas	48656	1638	32	26468	20107
100	Lisboa	Oeiras	62335	683	660	44846	15413
101	Lisboa	Palmela	4880	0	0	2686	2152
102	Lisboa	Seixal	20713	0	40	8813	11535
103	Lisboa	Sesimbra	4731	0	0	2975	1655
104	Lisboa	Setúbal	4408	0	31	2641	1654
105	Lisboa	Sintra	66397	0	9	40144	25768
106	Lisboa	Vila Franca de Xira	28636	0	0	18003	10189
107	Loures	Alcochete	312	0	0	312	0
108	Loures	Almada	1121	0	19	628	458
109	Loures	Amadora	6610	0	0	5239	994
110	Loures	Barreiro	86	0	0	81	5
111	Loures	Cascais	2665	0	0	2633	31
112	Loures	Lisboa	79430	3851	148	50809	23536
113	Loures	Loures	203061	69749	926	111399	18302
114	Loures	Mafra	7137	55	0	6944	138
115	Loures	Moita	81	0	0	45	17
116	Loures	Montijo	279	0	0	254	21
117	Loures	Odivelas	18891	562	19	16732	1476
118	Loures	Oeiras	3025	0	0	1977	1027
119	Loures	Palmela	456	0	0	433	23
120	Loures	Seixal	919	0	0	613	304
121	Loures	Sesimbra	224	0	0	212	12
122	Loures	Setúbal	437	0	0	437	0
123	Loures	Sintra	6747	0	0	6108	270
124	Loures	Vila Franca de Xira	14165	650	0	10860	1884
125	Mafra	Alcochete	7	0	0	7	0
126	Mafra	Almada	26	0	0	26	0
127	Mafra	Amadora	505	0	0	476	28
128	Mafra	Cascais	927	0	0	927	0
129	Mafra	Lisboa	9522	33	0	8550	940
130	Mafra	Loures	6893	12	0	6618	263
131	Mafra	Mafra	100468	16673	258	73570	6307
132	Mafra	Montijo	23	0	0	23	0
133	Mafra	Odivelas	1561	0	0	1507	54
134	Mafra	Oeiras	754	0	0	754	0
135	Mafra	Palmela	34	0	0	34	0
136	Mafra	Sesimbra	202	0	0	202	0

137	Mafra	Setúbal	47	0	0	47	0
138	Mafra	Sintra	4719	0	4	4488	193
139	Mafra	Vila Franca de Xira	344	0	0	333	11
140	Moita	Alcochete	235	0	0	235	0
141	Moita	Almada	312	0	0	73	239
142	Moita	Amadora	16	0	0	16	0
143	Moita	Barreiro	14746	2392	152	10668	1535
144	Moita	Cascais	178	0	0	178	0
145	Moita	Lisboa	5127	2	0	2725	2401
146	Moita	Loures	62	0	0	45	17
147	Moita	Moita	65792	34781	343	27998	2638
148	Moita	Montijo	1991	20	0	1885	0
149	Moita	Odivelas	390	0	0	298	92
150	Moita	Oeiras	55	0	0	23	32
151	Moita	Palmela	3854	91	0	2736	1026
152	Moita	Seixal	1448	0	0	1448	0
153	Moita	Sesimbra	821	0	0	821	0
154	Moita	Setúbal	2436	0	0	1044	687
155	Moita	Sintra	575	0	0	481	94
156	Moita	Vila Franca de Xira	474	0	0	474	0
157	Montijo	Alcochete	8376	18	91	6758	1508
158	Montijo	Almada	889	0	0	846	43
159	Montijo	Amadora	545	0	0	545	0
160	Montijo	Barreiro	1826	0	0	1720	0
161	Montijo	Cascais	873	0	0	686	187
162	Montijo	Lisboa	5276	0	24	3481	1761
163	Montijo	Loures	295	0	0	270	21
164	Montijo	Mafra	23	0	0	0	23
165	Montijo	Moita	2697	20	0	2591	0
166	Montijo	Montijo	68305	24638	930	40435	2224
167	Montijo	Odivelas	33	0	0	30	3
168	Montijo	Oeiras	315	0	0	315	0
169	Montijo	Palmela	5728	15	9	5384	284
170	Montijo	Seixal	1029	0	2	1021	0
171	Montijo	Sesimbra	120	0	0	120	0
172	Montijo	Setúbal	2692	0	42	2603	47
173	Montijo	Vila Franca de Xira	184	0	0	184	0
174	Odivelas	Alcochete	441	0	0	441	0
175	Odivelas	Almada	1455	0	0	1064	390
176	Odivelas	Amadora	10602	1192	0	8134	1275
177	Odivelas	Barreiro	25	0	0	25	0
178	Odivelas	Cascais	3240	0	0	3169	0
179	Odivelas	Lisboa	47756	1873	21	24736	20773

180	Odivelas	Loures	18702	572	0	16183	1812
181	Odivelas	Mafra	1329	0	0	1301	23
182	Odivelas	Moita	359	0	0	338	21
183	Odivelas	Montijo	30	0	0	30	0
184	Odivelas	Odivelas	148328	64016	210	73338	10624
185	Odivelas	Oeiras	4445	0	0	3670	685
186	Odivelas	Palmela	189	0	0	183	6
187	Odivelas	Seixal	1115	0	0	801	312
188	Odivelas	Sesimbra	90	0	0	77	13
189	Odivelas	Setúbal	124	0	0	106	16
190	Odivelas	Sintra	9291	305	0	7264	1646
191	Odivelas	Vila Franca de Xira	1656	0	0	1103	553
192	Oeiras	Alcochete	125	0	0	125	0
193	Oeiras	Almada	2222	0	0	1717	452
194	Oeiras	Amadora	15062	878	516	10676	2993
195	Oeiras	Barreiro	565	0	0	397	168
196	Oeiras	Cascais	34155	1425	52	29358	3304
197	Oeiras	Lisboa	61811	903	660	44697	15208
198	Oeiras	Loures	3206	0	0	2329	856
199	Oeiras	Mafra	827	0	0	827	0
200	Oeiras	Moita	90	0	0	58	32
201	Oeiras	Montijo	190	0	0	169	21
202	Oeiras	Odivelas	4404	0	0	3628	687
203	Oeiras	Oeiras	186797	64200	312	110224	11806
204	Oeiras	Palmela	651	0	0	616	25
205	Oeiras	Seixal	1974	0	0	1442	508
206	Oeiras	Sesimbra	468	0	0	468	0
207	Oeiras	Setúbal	776	0	0	740	37
208	Oeiras	Sintra	23933	602	0	19388	3912
209	Oeiras	Vila Franca de Xira	919	0	0	667	253
210	Palmela	Alcochete	439	11	0	193	234
211	Palmela	Almada	619	0	0	602	11
212	Palmela	Amadora	86	0	0	86	0
213	Palmela	Barreiro	4564	0	4	3567	993
214	Palmela	Cascais	337	0	0	337	0
215	Palmela	Lisboa	5675	0	0	3421	2218
216	Palmela	Loures	527	0	0	504	23
217	Palmela	Mafra	64	0	0	64	0
218	Palmela	Moita	5025	79	0	3334	1611
219	Palmela	Montijo	5471	0	9	5158	268
220	Palmela	Odivelas	69	0	0	59	10
221	Palmela	Oeiras	722	0	0	687	25
222	Palmela	Palmela	81747	22412	1422	52040	5661

223	Palmela	Seixal	3639	0	2	2113	1525
224	Palmela	Sesimbra	2518	0	0	2124	394
225	Palmela	Setúbal	20147	949	25	14221	4547
226	Palmela	Sintra	444	0	0	421	20
227	Palmela	Vila Franca de Xira	160	0	0	160	0
228	Seixal	Alcochete	231	0	0	172	0
229	Seixal	Almada	35845	2844	54	28261	4034
230	Seixal	Amadora	3234	0	0	2591	643
231	Seixal	Barreiro	3564	0	0	3491	73
232	Seixal	Cascais	1234	0	0	876	358
233	Seixal	Lisboa	19552	27	0	7287	11914
234	Seixal	Loures	928	0	0	616	312
235	Seixal	Moita	908	0	0	908	0
236	Seixal	Montijo	1028	0	0	1019	3
237	Seixal	Odivelas	968	0	0	657	308
238	Seixal	Oeiras	1798	0	0	1523	251
239	Seixal	Palmela	3487	0	2	2498	987
240	Seixal	Seixal	182892	67111	2103	103757	9125
241	Seixal	Sesimbra	6168	659	85	5201	226
242	Seixal	Setúbal	2428	4	2	2079	265
243	Seixal	Sintra	674	0	0	620	54
244	Seixal	Vila Franca de Xira	1340	0	0	1046	284
245	Sesimbra	Almada	2867	0	0	2561	176
246	Sesimbra	Amadora	289	0	0	201	88
247	Sesimbra	Barreiro	3612	62	0	3368	172
248	Sesimbra	Cascais	76	0	0	76	0
249	Sesimbra	Lisboa	4715	0	0	3225	1398
250	Sesimbra	Loures	203	0	0	203	0
251	Sesimbra	Mafra	227	0	0	227	0
252	Sesimbra	Moita	813	0	0	813	0
253	Sesimbra	Montijo	240	0	0	240	0
254	Sesimbra	Odivelas	109	0	0	90	19
255	Sesimbra	Oeiras	489	0	0	477	12
256	Sesimbra	Palmela	2252	0	0	2024	227
257	Sesimbra	Seixal	6540	671	85	5401	345
258	Sesimbra	Sesimbra	61503	20211	423	39661	840
259	Sesimbra	Setúbal	4699	112	43	4369	163
260	Sesimbra	Sintra	184	0	0	184	0
261	Sesimbra	Vila Franca de Xira	93	0	0	93	0
262	Setúbal	Alcochete	552	0	0	450	102
263	Setúbal	Almada	2625	0	0	2178	332
264	Setúbal	Amadora	918	0	0	913	5
265	Setúbal	Barreiro	2610	0	0	2003	475

266		Setúbal	Cascais	323	0	0	91	232
267		Setúbal	Lisboa	6323	0	0	4483	1794
268		Setúbal	Loures	519	0	0	507	12
269		Setúbal	Mafra	47	0	0	47	0
270		Setúbal	Moita	1896	0	0	501	692
271		Setúbal	Montijo	1871	0	44	1769	49
272		Setúbal	Odivelas	89	0	0	63	18
273		Setúbal	Oeiras	773	0	0	737	37
274		Setúbal	Palmela	20846	531	25	16783	3144
275		Setúbal	Seixal	2619	0	0	2352	186
276		Setúbal	Sesimbra	4824	202	43	4447	120
277		Setúbal	Setúbal	192242	60804	303	123742	6713
278		Setúbal	Sintra	80	0	0	48	32
279		Setúbal	Vila Franca de Xira	250	0	0	41	24
280		Sintra	Alcochete	349	0	0	276	49
281		Sintra	Almada	1069	0	0	673	396
282		Sintra	Amadora	31579	3061	22	21913	6217
283		Sintra	Barreiro	999	0	0	984	15
284		Sintra	Cascais	20909	32	12	18784	1957
285		Sintra	Lisboa	63751	312	32	40166	22774
286		Sintra	Loures	7569	0	0	6260	964
287		Sintra	Mafra	4693	0	4	4482	173
288		Sintra	Moita	596	0	0	481	115
289		Sintra	Montijo	12	0	0	0	12
290		Sintra	Odivelas	9883	496	0	7497	1802
291		Sintra	Oeiras	24142	570	0	19672	3875
292		Sintra	Palmela	411	0	0	388	20
293		Sintra	Seixal	675	0	0	651	25
294		Sintra	Sesimbra	191	0	0	191	0
295		Sintra	Setúbal	351	0	0	52	300
296		Sintra	Sintra	391270	127643	1341	238057	23595
297		Sintra	Vila Franca de Xira	1535	0	0	835	377
298	Vila Franca de Xira		Alcochete	96	0	0	96	0
299	Vila Franca de Xira		Almada	1058	0	0	981	67
300	Vila Franca de Xira		Amadora	1064	270	0	719	41
301	Vila Franca de Xira		Barreiro	124	0	0	38	86
302	Vila Franca de Xira		Cascais	949	0	0	949	0
303	Vila Franca de Xira		Lisboa	26739	0	11	16830	9485
304	Vila Franca de Xira		Loures	13739	690	0	11011	1270
305	Vila Franca de Xira		Mafra	322	0	0	311	11
306	Vila Franca de Xira		Moita	453	0	0	449	4
307	Vila Franca de Xira		Montijo	425	0	0	425	0
308	Vila Franca de Xira		Odivelas	1755	0	0	1250	505

```
309 Vila Franca de Xira              Oeiras    975      0    0    748      226
310 Vila Franca de Xira             Palmela    132      0    0    125        0
311 Vila Franca de Xira              Seixal   1459      0    0   1042      417
312 Vila Franca de Xira            Sesimbra     93      0    0     93        0
313 Vila Franca de Xira             Setúbal    490      0    0    129      176
314 Vila Franca de Xira              Sintra   1355      0    0    627      384
315 Vila Franca de Xira Vila Franca de Xira 161520  70582  269  82860     7719
    Other
1       7
2     214
3       0
4       0
5       0
6       0
7       0
8       0
9       0
10      0
11      0
12      0
13     60
14      0
15      0
16     24
17      0
18    214
19   2791
20      0
21      0
22      0
23   1145
24      0
25      0
26      0
27      0
28      0
29     53
30      6
31    639
32    130
33    110
34     12
35     10
```

36	0
37	0
38	200
39	0
40	9
41	381
42	258
43	0
44	0
45	0
46	0
47	35
48	0
49	0
50	0
51	0
52	332
53	65
54	0
55	0
56	0
57	276
58	0
59	40
60	20
61	0
62	0
63	0
64	0
65	0
66	0
67	0
68	91
69	0
70	0
71	0
72	0
73	9
74	0
75	1252
76	458
77	0
78	0

79	0
80	0
81	0
82	99
83	0
84	12
85	0
86	11
87	93
88	0
89	0
90	1042
91	672
92	2
93	275
94	11647
95	741
96	0
97	0
98	0
99	414
100	743
101	41
102	325
103	104
104	81
105	479
106	444
107	0
108	17
109	377
110	0
111	0
112	1089
113	2682
114	0
115	20
116	4
117	100
118	20
119	0
120	2
121	0

122	0
123	367
124	775
125	0
126	0
127	0
128	0
129	0
130	0
131	3666
132	0
133	0
134	0
135	0
136	0
137	0
138	35
139	0
140	0
141	0
142	0
143	0
144	0
145	0
146	0
147	32
148	86
149	0
150	0
151	0
152	0
153	0
154	703
155	0
156	0
157	0
158	0
159	0
160	106
161	0
162	7
163	4
164	0

165	86
166	79
167	0
168	0
169	37
170	6
171	0
172	0
173	0
174	0
175	0
176	0
177	0
178	71
179	350
180	136
181	5
182	0
183	0
184	142
185	90
186	0
187	4
188	0
189	2
190	77
191	0
192	0
193	53
194	0
195	0
196	15
197	343
198	20
199	0
200	0
201	0
202	90
203	254
204	10
205	24
206	0
207	0

208	34
209	0
210	0
211	6
212	0
213	0
214	0
215	37
216	0
217	0
218	0
219	37
220	0
221	10
222	212
223	0
224	0
225	407
226	3
227	0
228	60
229	651
230	0
231	0
232	0
233	327
234	0
235	0
236	6
237	4
238	24
239	0
240	797
241	0
242	76
243	0
244	10
245	130
246	0
247	10
248	0
249	93
250	0

251	0
252	0
253	0
254	0
255	0
256	0
257	39
258	370
259	12
260	0
261	0
262	0
263	115
264	0
265	132
266	0
267	44
268	0
269	0
270	703
271	7
272	7
273	0
274	365
275	81
276	12
277	678
278	0
279	185
280	24
281	0
282	362
283	0
284	126
285	469
286	345
287	35
288	0
289	0
290	85
291	28
292	3
293	0

```
294     0
295     0
296   630
297   322
298     0
299    10
300    34
301     0
302     0
303   411
304   770
305     0
306     0
307     0
308     0
309     0
310     7
311     0
312     0
313   185
314   344
315    88
```

```
head(data, 3) # first 3 values
```

```
  Origin_mun Destination_mun Total Walk Bike   Car PTransit Other
1  Alcochete       Alcochete 20478 6833  320 12484      833     7
2  Alcochete          Almada   567    0    0   353        0   214
3  Alcochete         Amadora   188    0    0   107       81     0
```

Check the number of rows (observations) and columns (variables)

```
nrow(data)
```

```
[1] 315
```

```
ncol(data)
```

```
[1] 8
```

Open the dataset

```
View(data)
```

7.2.0.3 Explore the data

Check the total number of trips

Use `$` to select a variable of the data

```
sum(data$Total)
```

```
[1] 5299853
```

Percentage of car trips related to the total

```
sum(data$Car)/sum(data$Total) * 100
```

```
[1] 59.17638
```

Percentage of active trips related to the total

```
(sum(data$Walk) + sum(data$Bike)) / sum(data$Total) * 100
```

```
[1] 24.44883
```

7.2.0.4 Modify original data

Create a column with the sum of the number of trips for active modes

```
data$Active = data$Walk + data$Bike
```

Filter by condition (create new tables)

Filter trips only with origin from Lisbon

```
data_Lisbon = data[data$Origin_mun == "Lisboa",]
```

Filter trips with origin **different** from Lisbon

```
data_out_Lisbon = data[data$Origin_mun != "Lisboa",]
```

Filter trips with origin **and** destination in Lisbon

```
data_in_Out_Lisbon = data[data$Origin_mun == "Lisboa" & data$Destination_mun == "Lisboa",]
```

Remove a column

Look at the first row

```
data[1,] #rows and columns start from 1
```

```
  Origin_mun Destination_mun Total Walk Bike   Car PTransit Other Active
1  Alcochete       Alcochete 20478 6833  320 12484      833     7   7153
```

Look at first row and column

```
data[1,1]
```

```
[1] "Alcochete"
```

Remove the first column

```
data = data[ ,-1] #first column
```

Create a table only with origin, destination and walking trips

There are many ways to do the same operation.

```
names(data)
```

```
[1] "Destination_mun" "Total"           "Walk"            "Bike"
[5] "Car"             "PTransit"        "Other"           "Active"
```

```
data_walk2 = data[ ,c(1,2,4)]
```

```
data_walk3 = data[ ,-c(3,5:9)]
```

7.2.0.5 Export data

Save data in **.csv** and **.Rds**

```
write.csv(data, 'data/dataset.csv', row.names = FALSE)
saveRDS(data, 'data/dataset.Rds') #Choose a different file.
```

7.2.0.6 Import data

```
csv_file = read.csv("data/dataset.csv")
rds_file = readRDS("data/dataset.Rds")
```

8 Data manipulation

In this chapter we will use some very useful `dplyr` functions to handle and manipulate data.

You can load the `dplyr` package directly, or load the entire tidy universe (`tidyverse`).

```
library(dplyr)
```

Using the same dataset as in R basics, we will do the same operations but in a simplified way.

```
data = readRDS("data/TRIPSmode_mun.Rds")
```

> Note that it is very important to understand the R basics, that's why we started from there, even if the following functions will provide the same results.

8.1 Select variables

Have a look at your dataset. You can open using `View()`, look at the information at the "Environment" panel, or even print the same information using `glimpse()`

```
glimpse(data)
```

We will create a new dataset with *Origin*, *Walk, Bike* and *Total*. This time we will use the `select()` function.

```
data_new = select(data, Origin_mun, Walk, Bike, Total) # the first argument is the dataset
```

The first argument, as usually in R, is the dataset, and the remainings are the columns to select.

With most of the `dplyr` functions you don't need to refer to `data$...` you can simply type the variable names (and even without the `"..."`!). This makes coding in R simpler :)

You can also remove columns that you don't need.

```
data_new = select(data_new, -Total) # dropping the Total column
```

8.1.1 Using pipes!

Now, let's introduce pipes. Pipes are a rule as: "**For this, do this.**"

This is usefull to skip the first argument of the functions (usually the dataset to apply the function).

Applying a pipe to the `select()` function, we can write as:

```
data_new = data |> select(Origin_mun, Walk, Bike, Total)
```

Two things to **note**:

1. The pipe symbol can be written as `|>` or `%>%`. [1] To write it you may also use the `ctrl+shift+m` shortcut.
2. After typing `select(` you can press `tab` and the list of available variables of that dataset will show up! `Enter` to select. With this you prevent typo errors.

8.2 Filter observations

8.3 Join data tables

8.4 `group_by` and summarize

8.4.0.1

[1] You can change this in RStudio > Tools > Global Options > Code.

9 Introduction to spatial data

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

10 Interactive maps

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

Part II

Day 2

11 OD pairs and desire lines

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

12 Georeferenced coordinates

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

13 Euclidean distances and buffers

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

14 Open transportation data

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

15 Routing and Accessibility with `r5r`

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
42 + 3.14
```

```
[1] 45.14
```

16 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
data = readRDS("data/TRIPSmode_mun.Rds")
data_walk = data[,c("Origin_mun", "Destination_mun", "Walk")]
head(data_walk)
```

```
  Origin_mun Destination_mun Walk
1  Alcochete       Alcochete 6833
2  Alcochete          Almada    0
3  Alcochete         Amadora    0
4  Alcochete        Barreiro    0
5  Alcochete         Cascais    0
6  Alcochete          Lisboa   69
```

References

Instituto National de Estatística. 2018. “Mobilidade e Funcionalidade Do Território Nas Áreas Metropolitanas Do Porto e de Lisboa: 2017.” Lisboa. https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=349495406&PUBLICACOESmodo=2&xlang=pt.

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.