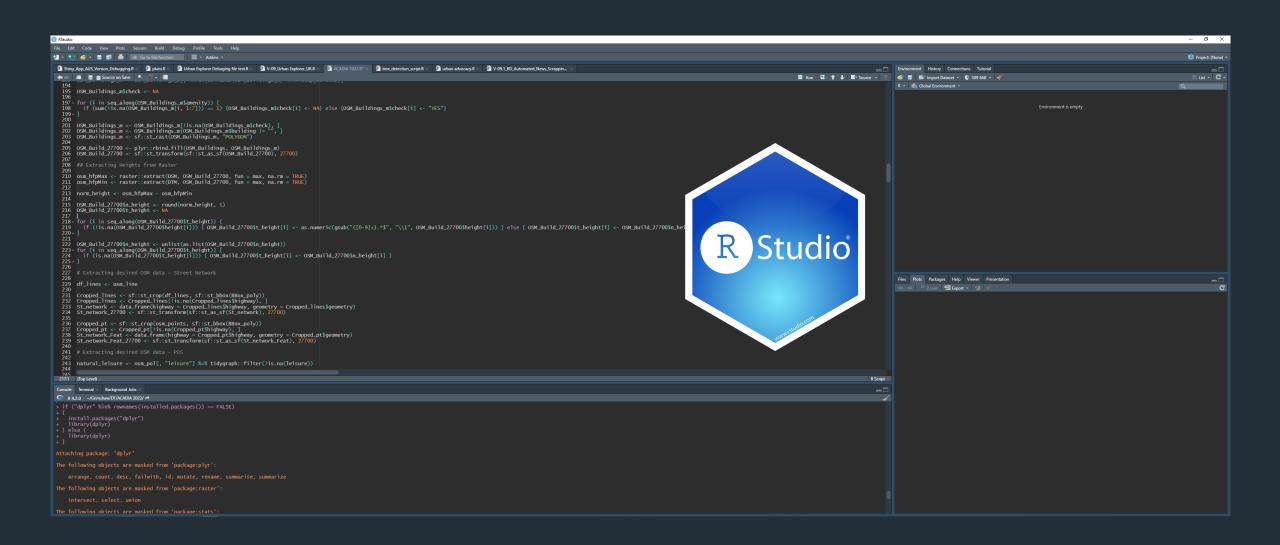
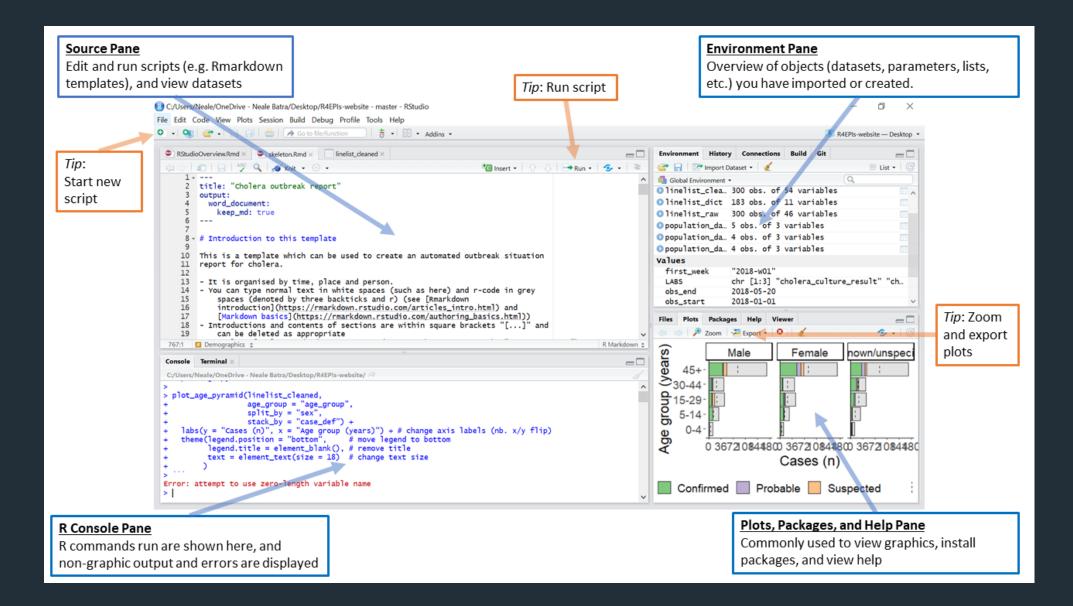
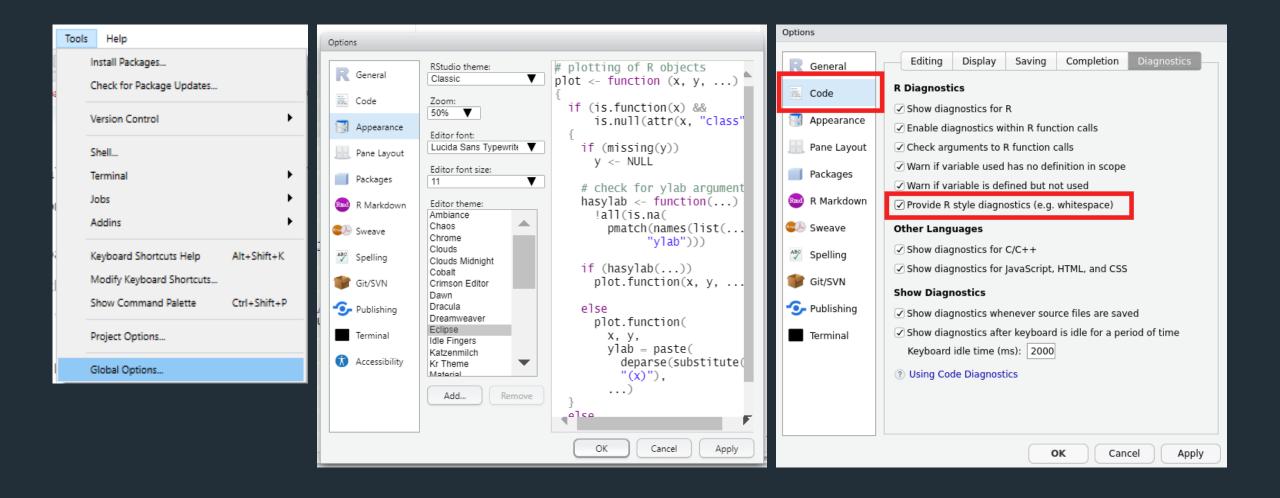
## ACADIA 2022

# A data-driven approach for urban design and master planning development

Jorge Sainz de Aja Curbelo – Esther Rubio Madronal







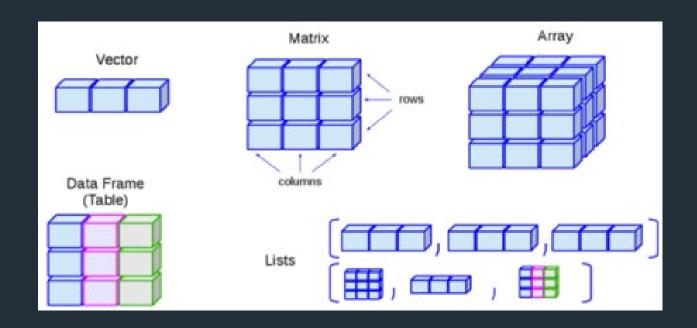
#### https://epirhandbook.com/en/r-basics.html



Bread <- dinner::eat(pastry::decorate(cooker::oven\_bake(...), ...)



Bread <- cooker::oven\_bake(...) %>% pastry::decorate(...) %>% dinner::eat(...)



id	Fruits	Dairies	Meat
1	apple	milk	NA
2	banana	cheese	NA
3	strawberry	cream	NA

Class	Explanation	Examples		
Character	These are text/words/sentences <b>"within quotation marks"</b> . Math cannot be done on these objects.	"Character objects are in quotation marks"		
Integer	Numbers that are <b>whole only</b> (no decimals)	-5, 14, or 2000		
Numeric	These are numbers and <b>can include decimals</b> . If within quotation marks they will be considered character class.	23.1 or 14		
Factor	These are vectors that have a <b>specified order</b> or hierarchy of values	An variable of economic status with ordered values		
Date	Once R is told that certain data are Dates, these data can be manipulated and displayed in special ways. See the page on Working with dates for more information.	2018-04-12 or 15/3/1954 or Wed 4 Jan 1980		
Logical	Values must be one of the two special values TRUE or FALSE (note these are <b>not</b> "TRUE" and "FALSE" in quotation marks)	TRUE or FALSE		
data.frame	A data frame is how R stores a <b>typical dataset</b> . It consists of vectors (columns) of data bound together, that all have the same number of observations (rows).	The example AJS dataset named linelist_raw contains 68 variables with 300 observations (rows) each.		
tibble	tibbles are a variation on data frame, the main operational difference being that they print more nicely to the console (display first 10 rows and only columns that fit on the screen)	Any data frame, list, or matrix can be converted to a tibble with as_tibble()		
list	A list is like vector, but holds other objects that can be other different classes	A list could hold a single number, and a dataframe, and a vector, and even another list within it!		

A NAN value in  $\underline{R}$  represents "NOT A NUMBER", It is basically any numeric calculations with an undefined result, such as '0/0'. This exists only in vectors with the numeric datatype.

A NA value in R represents "NOT AVAILABLE". This can exist in any sort of numeric or character vector. It is generally interpreted as missing values.

NULL is an object and is returned when an expression or function results in an undefined value.

Inf and <u>-Inf</u> stands for infinity (or negative infinity) and is a result of storing either a large number or a product that is a result of division by zero.

In numeric operations, you can avoid errors due to undefined values with the argument na.rm = TRUE

e.g:

#### 1. Loading all necessary packages

```
BOOL (TRUE/FALSE) +
                                                                  → FALSE
 package "raster" Is in
                                  The list of installed packages
if ("raster" %in% rownames(installed.packages()) == FALSE)
  install.packages("raster") < FALSE == FALSE: install and load the package
  library(raster)
} else {
  library(raster)
                                     < TRUE == FALSE: load the package
                                          FALSE
```

```
If (bool) { bool = True, do something } else { bool = False, do something else }

If (bool) { bool = True, do something } else if (if initial condition is False, test a second condition) {do...}
```

& = and, | = or > if (apple is green & apple is sweet), if (apple is green | apple is sweet)

#### 2. Setting up your working folder

setwd("C:/Users/user\_name/Documents/Folder/ACADIA 2022")

e.g: C:/Users/user\_name/Documents/Folder/ACADIA 2022/file.shp

Read("file.shp")

getwd() give back the folder on setwd – informative

R language does not recognize back slash as they are reserved for special characters, please change back slashes on your folder path by a single or double forward slash;

- The "forward slash" / is actually more common as it used by Unix, Linux, and macOS
- The "backward slash" \ is actually somewhat painful as it is also an escape character. So whenever you want one, you need to type two in the string: "C:\\TEMP".

<sup>\*</sup>REGEX (regular expression) Special characters

## 3. Setting up our area of analysis

Bounding box (bbox) method 1 Bounding box (bbox) method 2 Lat, long Distance\_1 Lat, long Distance\_2 Lat, long

bb\_poly: transform a bounding box (area) into a polygon which can be used to split/clip/crop other polygons etc.

#### 4. Loading all necessary data



Packages: OSMAR & OSMDATA

Query <- bbox %>% opq() %>% add\_osm\_feature("building")

buildings <- osmdata\_sf(Query)</pre>

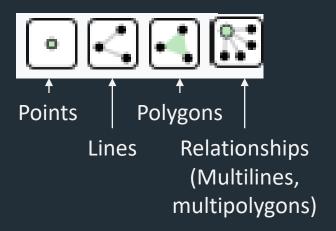
## Building This is used to identify individual buildings or groups of connected buildings. See the page Buildings for further details on the usage of this tag and man\_made=\* for tagging of various other structures. The building tags are intended for the physical description of a building: for functions in the building (e.g. police station, church, townhall, museum) you should add additional tags like amenity=\*, tourism=\*, shop=\* etc. For example mapping building=supermarket is not enough to mark place as having an active supermarket shop, it just marks that building has form typical for supermarket buildings. shop=supermarket must be mapped to indicate an active supermarket shop. On the other hand shop=\*/amenity=\* is not indicating building by itself, building must be mapped with building=\* tag.

active supern	active supermarket shop. On the other hand shop="/amenity=" is not indicating building by itself, building must be mapped with building=" tag.					
Key	Value	Comment	Photo			
Accommodation						
building	apartments	A building arranged into individual dwellings, often on separate floors. May also have retail outlets on the ground floor.				
building	barracks	Buildings built to house military personnel or laborers.				

https://wiki.openstreetmap.org/wiki/Map\_features

Query <- bbox %>% opq()

OSM\_sf <- osmdata\_sf(Query) # Returns whole osm features for the defined bounding box





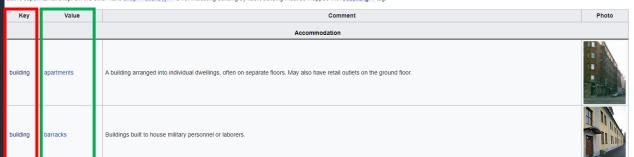
mypoints <- OSM\_sf\$points

## 4. Loading all necessary data

#### Building

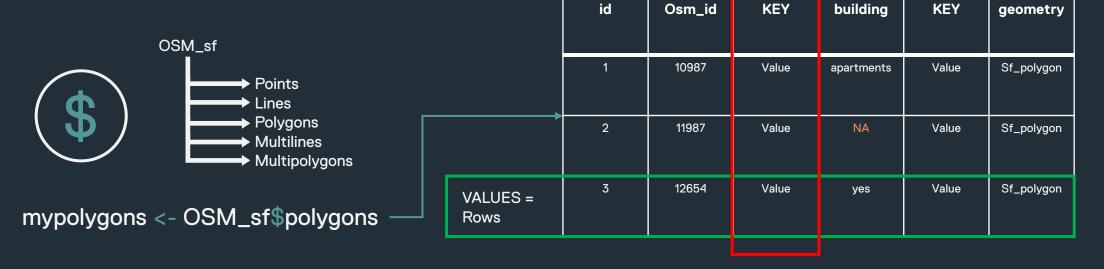
This is used to identify individual buildings or groups of connected buildings. See the page **Buildings** for further details on the usage of this tag and man\_made=\* for tagging of various other structures. The building tags are intended for the physical description of a building: for functions in the building (e.g. police station, church, townhall, museum) you should add additional tags like amenity=\*, tourism=\*, shop=\* etc.

For example mapping building=supermarket is not enough to mark place as having an active supermarket shop, it just marks that building has form typical for supermarket buildings, shop=supermarket must be mapped to indicate an active supermarket shop. On the other hand shop='/amenity=' is not indicating building by itself, building must be mapped with building=' tag.



mypolygons\$building[3] = "yes"

KEY = Columns



#### 4. Loading all necessary data

https://cran.r-project.org/web/packages/available\_packages\_by\_name.html

Available CRAN Packages By Name

#### ABCDEFGHIJKLMNOPQRSTUVWXYZ

A3 Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
AATtools Reliability and Scoring Routines for the Approach-Avoidance Task
ABACUS Apps Based Activities for Communicating and Understanding Statistics

abbreviate Readable String Abbreviation

abbyyR Access to Abbyy Optical Character Recognition (OCR) API
abc Tools for Approximate Bayesian Computation (ABC)

abc.data Data Only: Tools for Approximate Bayesian Computation (ABC)

ABC.RAP Array Based CpG Region Analysis Pipeline

abcADM Fit Accumulated Damage Models and Estimate Reliability using ABC

ABCanalysis Computed ABC Analysis

abclass Angle-Based Large-Margin Classifiers

ABCoptim Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2 Approximate Bayesian Computational Model for Estimating P2
abcrf Approximate Bayesian Computation via Random Forests

abcrlda Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis

 abctools
 Tools for ABC Analyses

 abd
 The Analysis of Biological Data

 abdiv
 Alpha and Beta Diversity Measures

 abe
 Augmented Backward Elimination

 abess
 Fast Best Subset Selection

 abglasso
 Adaptive Bayesian Graphical Lasso

 ABHgenotypeR
 Easy Visualization of ABH Genotypes

 abind
 Combine Multidimensional Arrays

<u>abjData</u> Databases Used Routinely by the Brazilian Jurimetrics Association

abjutils Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association

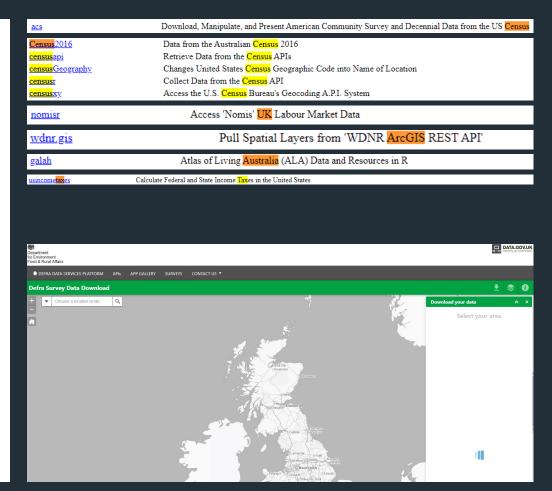
<u>abmR</u> Agent-Based Models in R

<u>bn</u> Modelling Multivariate Data with Additive Bayesian Networks

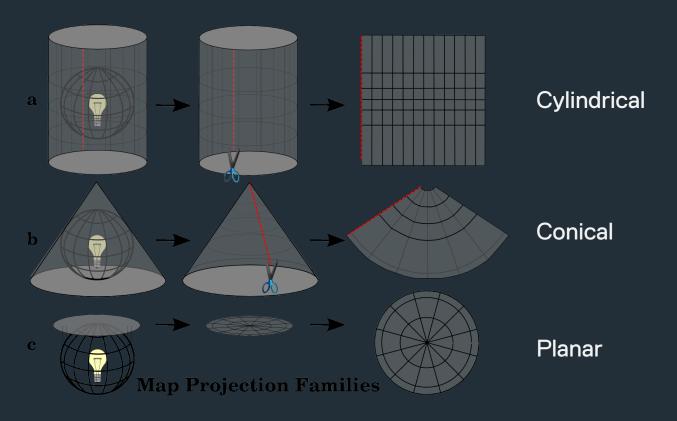
abnormality Measure a Subject's Abnormality with Respect to a Reference Population

abodOutlier Angle-Based Outlier Detection

ABPS The Abnormal Blood Profile Score to Detect Blood Doping



#### 4. Loading all necessary data



Latitude and Longitude coordinates do NOT have any projections, those are spherical coordinates

sf::sf\_use\_s2(FALSE), avoids using spherical calculations from the package "sp"

Lat & Lng can be directly translated to EPSG: 4326, CRS Mercator (e.g. google maps)

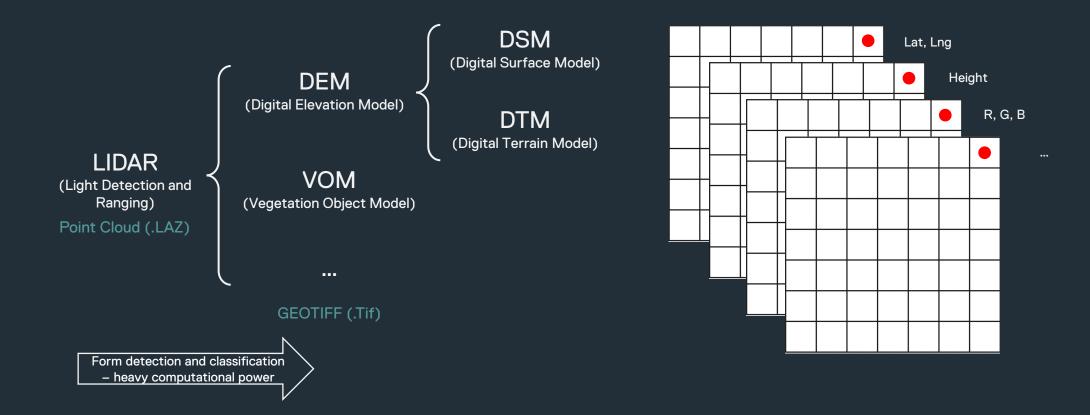
sf::st\_transform(sf\_geometry, 27700), allows you to reproject your geometry from one CRS to another

Measurements may change from one CRS to another

The full CRS string could be obtained if required by googling your EPSG code + "proj.4" or at EPSG.io



## 4. Loading all necessary data



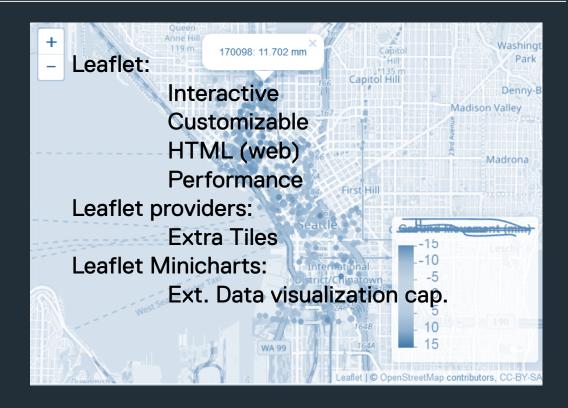
#### 5. Processing and visualizing loaded data

[Row, Column]

Buildings\_sf <- osm\_pol[!is.na(osm\_pol\$building), ]

! == "is not"
!is.na() == "is not NA"

Select all ROWS that are not NA at COLUMN building from osm\_pol, keeping all the columns





Empty map (initialize)

Add background map

Add building polygons, pop-up building column information when feature clicked (leaflet uses ~ rather than \$)

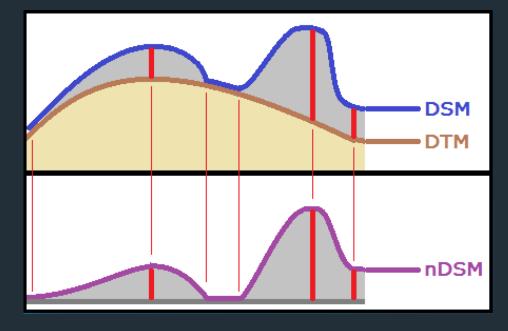
LEAFLET only plot geometry in lat and Ing, EPSG 4326 (Mercator CRS)

## 5. Processing and visualizing loaded data

osm\_hfpMax <- raster::extract(DSM, OSM\_Build\_27700, fun = max, na.rm = TRUE)

	10	10	10		10			
10					10			
	30	31	33					
10	30	31	33		10		10	
10	30	31	33		35		10	
30	30	31	33		35		10	
10	10	10	10		10	10		

fun = max == 35 fun = min == 30 fun = mean == 31.7



Norm\_DSM = DSM - DTM

DSM and DTM heights are based on sea levels we need to subtract DTM heights from DSM heights to obtain relative building heights

#### 5. Processing and visualizing loaded data

```
df_lines <- osm_line
```

Line features for column Highway represent classified street segments, such as primary, secondary...

```
Cropped_lines <- sf::st_crop(df_lines, sf::st_bbox(BBox_poly)) ←→ Crop all lines with bounding box polygon
Cropped_lines <- Cropped_lines[!is.na(Cropped_lines$highway), ] ←→ select only those which are NON NA values on Highway column
St_network <- data.frame(highway = Cropped_lines$highway, geometry = Cropped_lines$geometry) ←→ create a new data frame only with geometry and Highway column
St_network_27700 <- sf::st_transform(sf::st_as_sf(St_network), 27700) ←→ reproject to EPSG: 27700 (British)
```

Repeat the same operation with points features.

Points features for column Highway represent crosswalks, traffic lights, etc.

```
Cropped_pt <- sf::st_crop(osm_points, sf::st_bbox(BBox_poly))
Cropped_pt <- Cropped_pt[!is.na(Cropped_pt$highway), ]
St_network_Feat <- data.frame(highway = Cropped_pt$highway, geometry = Cropped_pt$geometry)
St_network_Feat_27700 <- sf::st_transform(sf::st_as_sf(St_network_Feat), 27700)
```

#### 5. Processing and visualizing loaded data

POI - Point of Interest

Based on OSM values, we should create a list of all amenities, this list might change based on different locations, project types and other considerations

```
amenity_list <- c("yes", "bar", "biergarten", "cafe", "fast_food", ...) \leftarrow \rightarrow list of amenities considered
```

amenity\_pol <- osm\_pol[, "amenity"] %>% tidygraph::filter(!is.na(amenity))  $\leftarrow \rightarrow$  selecting amenity column and filtering by non NA values amenity\_filt <- amenity\_pol[amenity\_pol\$amenity \%in\% amenity\_list, ]  $\leftarrow \rightarrow$  filtering again based on the list creates previously amenity\_pol\_filt <- sf::st\_centroid(amenity\_filt)  $\leftarrow \rightarrow$  calculating the centroid of each polygon so we have a list of points

```
amenity_pt <- osm_points[, c("amenity", "geometry")] %>% tidygraph::filter(!is.na(amenity))
amenity_pt_filt <- amenity_pt[amenity_pt$amenity %in% amenity_list, ] Same process as before but with the point data-frame
```

amenity\_pt\_total <- as.data.frame(rbind.fill(amenity\_pt\_filt, amenity\_pol\_filt))  $\leftarrow \rightarrow$  we now can join the result of the points data-frame and the polygon(centroid) data-frame

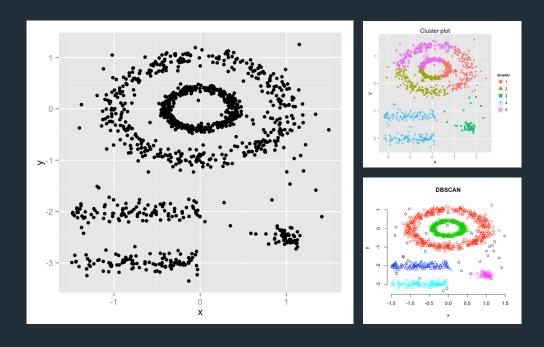
#### 5. Processing and visualizing loaded data

#### POI - Point of Interest

```
if (!is.null(amenity_pt_total)) {
 amenity_sustenance <- c(...)
 amenity_education <- c(...)
 amenity_finance <- c(...)
                                         amenity_list <- c()
 amenity_healthcare <- c(...)
 amenity_culture <- c(...)
 amenity_public <- c(...)
 amenity_others <- c(...)
for (i in seg_along(amenity_pt_total$amenity)) {
  if (amenity_pt_total$amenity[i] %in% amenity_sustenance) {
   amenity_pt_total$category[i] <- "sustenance"
   } else if (amenity_pt_total$amenity[i] %in% amenity_education) {
   amenity_pt_total$category[i] <- "education"}
  else if (amenity_pt_total$amenity[i] %in% amenity_finance) {
   amenity_pt_total$category[i] <- "finance"}
  else if (amenity_pt_total$amenity[i] %in% amenity_healthcare) {
   amenity_pt_total$category[i] <- "healthcare"}
  else if (amenity_pt_total$amenity[i] %in% amenity_public) {
   amenity_pt_total$category[i] <- "public"}
  else if (amenity_pt_total$amenity[i] %in% amenity_others) {
   amenity_pt_total$category[i] <- "others"}
```

If the value in column amenity is in the list amenity\_sustenance, then assign the value "sustenance" to the column Category. If not, then test another condition (else if)

#### 5. Processing and visualizing loaded data



data: data matrix, data frame or dissimilarity matrix (dist-object). Specify method = "dist" if the data should be interpreted as a dissimilarity matrix or object. Otherwise, Euclidean distances will be used.

eps: Reachability maximum distance

MinPts: Reachability minimum number of points

scale: If TRUE, the data will be scaled

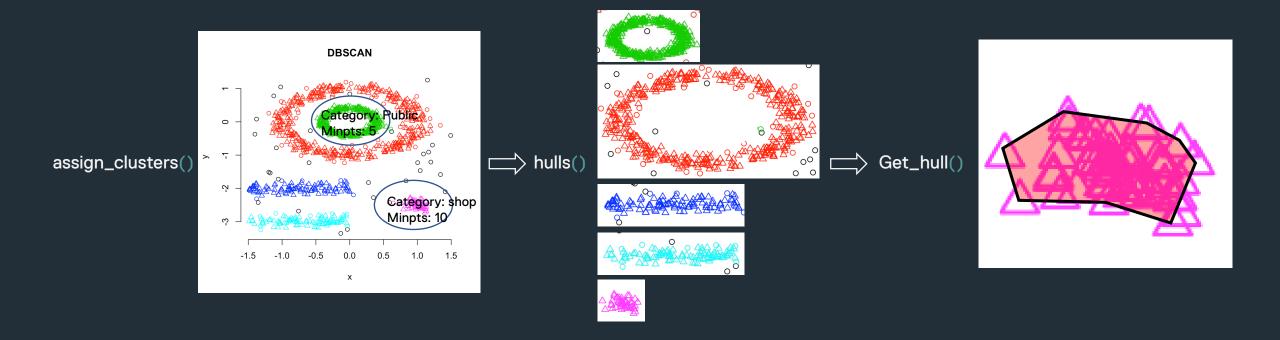
method: Possible values are:

- dist: Treats the data as a distance matrix
- raw: Treats the data as raw data
- hybrid: Expect also raw data, but calculates partial distance matrices

**DBSCAN** is a **density-based clustering algorithm**, introduced by Ester et al. 1996, which can be used to identify clusters of any shape in data set containing noise and outliers. DBSCAN stands for Density-Based Spatial Clustering and Application with Noise.

Clustering analysis is an unsupervised learning method that separates the data points into several specific bunches or groups, such that the data points in the same groups have similar properties and data points in different groups have different properties in some sense.

## 5. Processing and visualizing loaded data



#### 5. Processing and visualizing loaded data

#### ArcGIS RESTFUL API

Method 1 (package arcpullr):

arcpullr::get\_layer\_by\_poly("https://services5.arcgis.com/GfwWNkhOj9bNBqoJ/arcgis/rest/services/MAPPLUTO/FeatureServer/0", BBox\_sf, sp\_rel = "esriSpatialRelIntersects")

**ArcGIS Server** 

sf polygon

Spatial relationship: intersects, contains ...

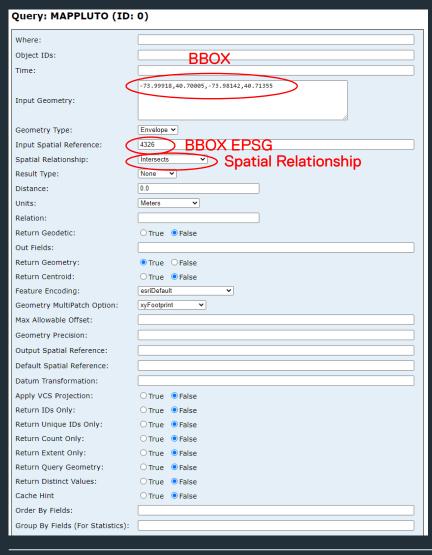
Method 2 (url):

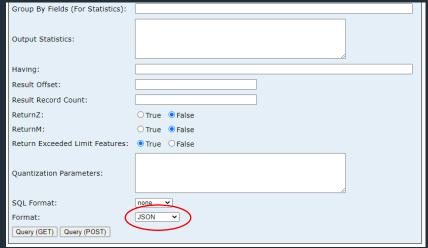
ESRI\_query <-

paste("https://onsinspire.esriuk.com/arcgis/rest/services/Census\_Boundaries/Lower\_Super\_Output\_Areas\_December\_2011\_Centroids/MapServer/0/query?where=1%3D1&outFields=\*& geometry=", as.character(bbox\_coord\$left), "%2C", as.character(bbox\_coord\$top), "%2C", as.character(bb

query\_result <- jsonify::from\_json(ESRI\_query)</pre>

#### 5. Processing and visualizing loaded data





https://services5.arcgis.com/GfwWNkhOj9bNBqoJ/arcgis/rest/services/MAPPLUTO/FeatureServer/0/query?where=&objectIds=&time=&geometry= 73.99918%2C40.70005%2C

73.98142%2C40.71355&geometryType=esriGeometryEnvelope&inSR=4326&spatialRel=esriSpatialRelIntersects&resultType=none&distance=0.0&units=esriSRUnit\_Meter&relationParam=&returnGeodetic=false&outFields=&returnGeometry=true&returnCentroid=false&featureEncoding=esriDefault&multipatchOption=xyFootprint&maxAllowableOffset=&geometryPrecision=&outSR=&defaultSR=&datumTransformation=&applyVCSProjection=false&returnIdsOnly=false&returnUniqueldsOnly=false&returnCountOnly=false&returnExtentOnly=false&returnQueryGeometry=false&returnDistinctValues=false&cacheHint=false&orderByFields=&groupByFieldsForStatistics=&outStatistics=&having=&resultOffset=&resultRecordCount=&returnZ=false&returnM=false&returnExceededLimitFeatures=true&quantizationParameters=&sqlFormat=none&f=pjson&token=

#### 5. Saving the Data

Save\_folder <- "C://Users//your\_name//Documents//ACADIA 2022//Results//"



