



osmlanduseR: An R package for the analysis of land use data contributed to OpenStreetMap

Andrés Esteban Duhour

2024-11-29

Laboratorio de Ecología, Departamento de Ciencias Básicas, Universidad Nacional de Luján - Instituto de Ecología y Desarrollo Sustentable (UNLu - CONICET);



Introduction

Introduction

Over the past 30 years, Argentina has been experiencing changes in its agricultural and urban development model that have drastically altered land use practices and patterns (Palmisano 2018; Pintos and Narodowski 2012).



Introduction

On a global scale, the expansion of capitalism through a process of accumulation by dispossession prevails. This is characterized by land privatization, the expulsion of farmers, the conversion or suppression of rights to the commons (Harvey 2004).



Introduction

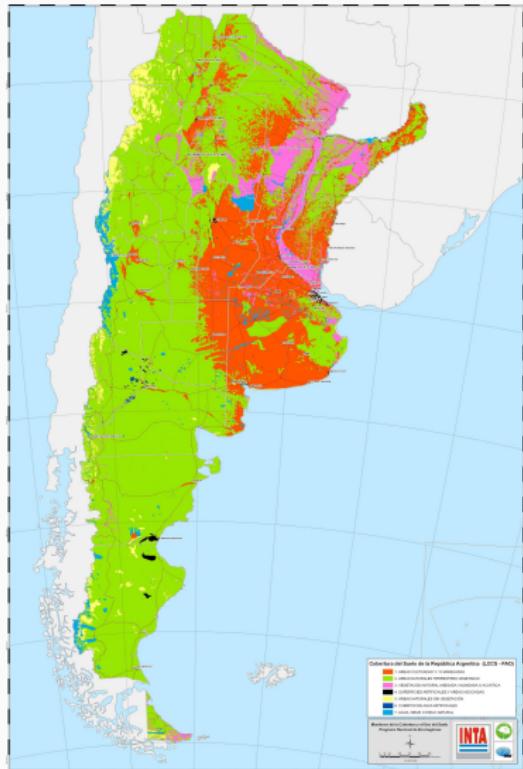


Figure 1: Land cover, INTA

According to the land cover inventory made by the National Institute of Agricultural Technology, (Volante 2009):

- Managed or cultivated areas covered 584353 km² (21.64 % of the total) and
- herbaceous crops accounted for 556720 km² (95.27 %).



FOSS4G
BELÉM, 2024

Introduction

LA NACION > Sociedad

Inundaciones: el agua entró en la Basílica de Luján

La municipalidad dispuso la mudanza de evacuados y el cierre de varias calles y del cementerio
11 de agosto de 2015 • 17:48

[lanacionar](#)



These changes have been associated with flooding in several regions (Pal et al. 2021; Pattison and Lane 2012) .

In particular, the Luján River basin (Buenos Aires, Argentina) has experienced extreme events historically.



Introduction

LA NACION > Sociedad

Inundaciones: el agua entró en la Basílica de Luján

La municipalidad dispuso la mudanza de evacuados y el cierre de varias calles y del cementerio

11 de agosto de 2015 • 17:48

lanacionar



The rising incidence and intensity of such events around the world has led to increased concern to study this relationship:

- China: Liu et al. (2023)
- Indonesia: Sugianto et al. (2023)
- Argentina: Pal et al. (2021)
- Brasil: da Silva et al. (2024)



FOSS4G
BELÉM, 2024

Introduction

The analysis of these changes in land use requires information typically obtained from remote sensing, which is validated and complemented with sampling programs.

The development of an updated land use map is a basic tool to study territorial phenomena such as flooding or land use change. Such an approach would enable the information to be compared to previously collected data on various scales and time periods.



Introduction

Although, geographic information, its format, or processing tools do not generally allow for its reuse or improvement, and they are not necessarily openly/freely available.

This type of data can be considered a *digital commons* and may be the susceptible of mercantilization processes.

Indeed, publicly produced information and processing tools should be made available in open-source formats and licenses to enforce knowledge construction (J. J. Arsanjani et al. 2015; Duféal and Noucher 2017).

Introduction



OpenStreetMap (OSM) is currently the main framework for volunteered geographic information, and because it constitutes a standardized database, it also allows it to be a preferred repository for contributions from universities and public sector research initiatives.

Recently, OSM has been registered as a *public good* by an agency affiliated to the United Nations: 

Introduction

The data contributed to OSM has already been incorporated into the creation and validation of land use and land cover maps, which are predominantly utilized in Europe:

- Vienna, Austria: A. J. Arsanjani et al. (2013);
- Paris, France; Milan, Italy: Fonte et al. (2016);
- Heidelberg, Germany: Schultz et al. (2017).

These authors have developed methods and algorithms for retrieving, analyzing, and classifying OSM data to produce land use maps.

Introduction

From these works also derived the osmlanduse.org  initiative, which delivers OSM land-use/land-cover information mapped to CORINE classes as a Web Map Service (WMS).



Introduction

In different countries, the local community of OSM users has contributed a substantial amount of geographic information that could be used for land use analysis, which might be expanded further, particularly in non-urban areas.

The general scheme for describing geographic elements provided by OSM can be applied to land use mapping by combining that information with country-specific classification systems.



Introduction

On this basis, there is a need for tools that can be applied at the local level to promote interaction with participatory mapping processes.

Such an approach would facilitate harmonization of OSM data with other classification systems, including those created by FAO or defined by users.



Introduction

- As a consequence, the objective of this study is to present the development of an  package for the analysis of land use data contributed to OSM.
- Overall, the aim is to increase the contribution of publicly generated information and its analysis tools in an open access format, such as those provided by OSM and  (R Core Team 2023).



Materials and Methods

Creation of a land use map in OpenStreetMap

Since 2016, land use data of the Lujan River's middle basin has been contributed to OSM as part of projects carried out by the National University of Luján.

Land use was visually assessed using satellite imagery, and representative polygons were added with appropriate tags. Geometries were created preferably as multipolygons.

Creation of a land use map in OpenStreetMap

The boundaries of each geometry were drawn to avoid sharing nodes with the road and rail networks.

The JOSM program was used for OSM edition, using imagery available within the program provided by Bing, Mapbox, Digital Globe, and the National Geographic Institute of Argentina. These images were captured between 2016 and 2022.



JOSM: Java editor for OSM



package implementation



The package is in its early stages of development



The code repository is accessible in
<https://github.com/aduhour/osmlanduseR>.



FOSS4G
BELÉM, 2024

The package comprises a number of features aiming at:

- 1) Downloading a set of land use related data from OSM using the overpass API: (**get_osmlanduse function**);
- 2) Removing overlaps, measuring polygon area and classifying polygons by mapping OSM tags to user-defined classes (**classify_osmlandse function**).
- 3) Providing example data sets for the translation of OSM land use related tags into other classification systems, such as the CORINE Land Cover and FAO Land Cover Classification System.
- 4) Creating a land use classification map (**map_osmlandse function**).

1. *get_osmlanduse* function

The *get_osmlanduse* function requires two arguments;

- 1) *area*: An objective region, defined as a polygon element in *sf* (simple features) format or as a character string containing the name of a region.

This input data is used to obtain the bounding box to set the overpass API query.

- 2) *crop_to*: A character string indicating if the final result should be cropped to the area polygon or to the bounding box.

1. get_osmlanduse function

The OSM retrieved features are those tagged as *natural* = *, *landuse* = *, *amenity* = *, *aeroway* = * and *leisure* = *.

At present, only the polygon objects are taken into account in the function implementation.

Finally, the invalid geometries are fixed and a list containing the objective area and the land use data is returned.

2. *classify_osmlanduse* function

The *classify_osmlanduse* function takes the output of the previously described function and:

- transforms the data to the selected coordinate reference system,
- detects and removes overlapping polygons and measures the area of elements.
- assigns a land use class to each polygon by matching existing OSM tag values to a user-defined list of class names.

2. classify_osmlanduse function

Overlaps are resolved by two optional methods:

- a) prioritizing the *smaller* polygons according to Schultz et al. (2017).
- b) prioritizing polygons in a *hierarchical* order of land use classes, as suggested by Fonte et al. (2016).

3. Data sets

Two data sets integrate the package to be used in classification:

- The data set *clc* is an example of the legend harmonization between OSM tags and Corine Land Cover classes proposed by Schultz et al. (2017).
- The dataset *intafao* is a translation of OSM tags to the FAO Land Cover Classification System adopted by the National Institute of Agricultural Technology (Argentina) (Volante 2009).



4. *map_osmlanduse* function

The *map_osmlanduse* function takes the classified set of data and returns a land use thematic map.

Dependencies

The implemented procedures use functions from the packages



(Padgham et al. 2017; Pebesma 2018; Tennekes 2018)



FOSS4G
BELÉM, 2024

Example application

Example application: Basic usage

- By calling the function `get_osmlanduse` with default arguments, the land use data of Luján in Buenos Aires, Argentina is downloaded:

```
lujan <- get_osmlanduse()
```

```
lujan$osmlanduse
```

```
> lujan$osmlanduse
Simple feature collection with 4111 features and 2 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -59.33842 ymin: -34.80437 xmax: -58.95423 ymax: -34.37462
Geodetic CRS:  WGS 84
First 10 features:
  value      key          geometry
1 water natural MULTIPOLYGON (((-59.18621 -...
```

value	key	geometry
1	water	natural MULTIPOLYGON (((-59.18621 -...
2	water	natural MULTIPOLYGON (((-59.11615 -...
3	water	natural MULTIPOLYGON (((-59.14118 -...
4	wood	natural MULTIPOLYGON (((-59.16937 -...
5	wood	natural MULTIPOLYGON (((-59.16727 -...
6	wood	natural MULTIPOLYGON (((-59.1238 -3...
7	wood	natural MULTIPOLYGON (((-59.16406 -...
8	wood	natural MULTIPOLYGON (((-59.12353 -...
9	wood	natural MULTIPOLYGON (((-59.06947 -...
10	wood	natural MULTIPOLYGON (((-59.06949 -...



Example application: Basic usage

- The result can then be classified into the classes of the CORINE Land Cover scheme

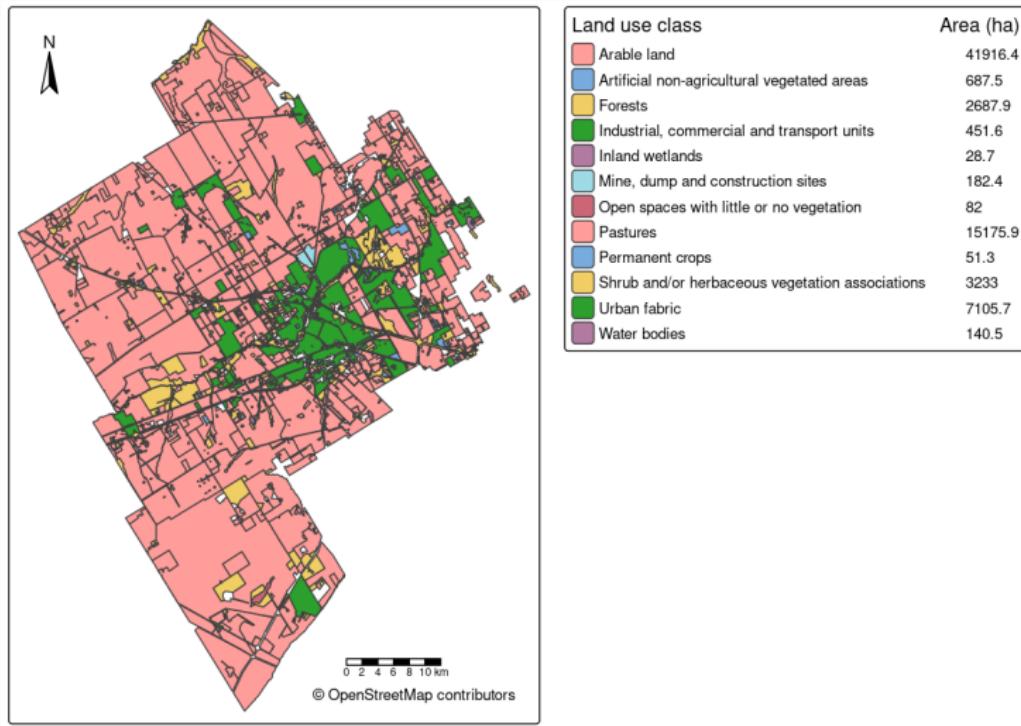
```
data(clc)

lujan.clc <- classify_osmlanduse(lujan$osmlanduse,
                                    clc$osm_tag,
                                    clc$class_name)
```

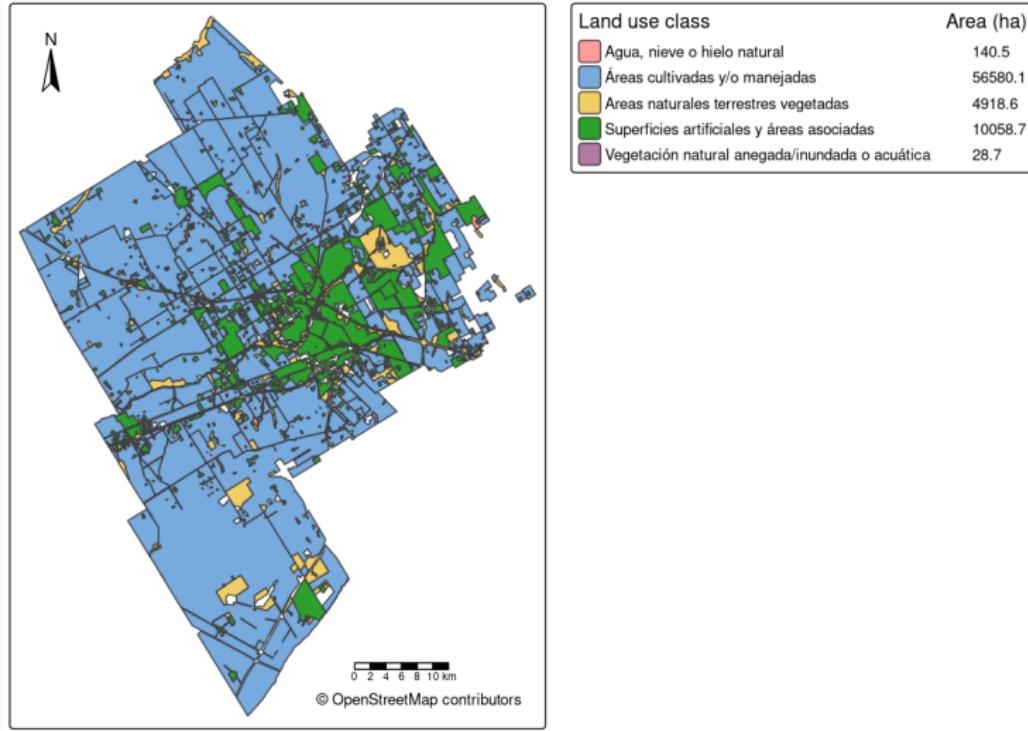
```
> subset(lujan.clc,subset = !is.na(lujan.clc$class_name))
Simple feature collection with 3575 features and 4 fields
Geometry type: GEOMETRY
Dimension: XY
Bounding box: xmin: 5560755 ymin: 6148759 xmax: 5595996 ymax: 6196458
Projected CRS: POSGAR 2007 / Argentina 5
First 10 features:
  value      key          class name      area      geometry
1129 industrial landuse Industrial, commercial and transport units 0.01632289 [ha] POLYGON ((5588312 6179957, ...
3878 park leisure Artificial non-agricultural vegetated areas 0.02528940 [ha] POLYGON ((5582932 6175043, ...
1128 industrial landuse Industrial, commercial and transport units 0.04771547 [ha] POLYGON ((5588306 6180175, ...
279 scrub natural Shrub and/or herbaceous vegetation associations 0.05574416 [ha] POLYGON ((5584017 6173754, ...
915 grass landuse Shrub and/or herbaceous vegetation associations 0.06931497 [ha] POLYGON ((5575766 6171394, ...
625 wood natural Forests 0.09567525 [ha] POLYGON ((5588331 6187822, ...
1064 commercial landuse Industrial, commercial and transport units 0.10574843 [ha] POLYGON ((5582367 6172797, ...
237 water natural Water bodies 0.11090597 [ha] POLYGON ((5583965 6179873, ...
3962 park leisure Artificial non-agricultural vegetated areas 0.13647602 [ha] POLYGON ((5579281 6175366, ...
277 scrub natural Shrub and/or herbaceous vegetation associations 0.14119399 [ha] POLYGON ((5583938 6173679, ...
```

Example application: Basic usage

```
map_osmlanduse(lujan.clc, "Partido de Luján")
```



Example application: Land use map of Luján



Example application: Land use map of Luján

The methods developed in the package *osmlanduseR* were used to download the elements associated with land use from OSM.

The geometric operations oriented to remove overlapping were successfully accomplished, and the matching of the present tags with a user-defined list of land use classes.

The land use polygons retrieved from OSM in the municipality of Luján were found to cover over 71742.8 ha, or 92.8% of the total area.

Example application: Flood risk and land use classes in Luján

The Buenos Aires province produced flood risk maps for the Luján river basin.

The corresponding area for the Luján city was selected a map of land use classes:

```
fr <- get_osmlanduse(floodrisk)
fr.intafao <- classify_osmlanduse(fr$osmlanduse,
                                    intafao$osm_tag,
                                    intafao$class_name)
map_osmlanduse(fr.intafao,
               "Land use and flood risk in Luján.")
```



Example application: Flood risk and land use classes in Luján

Land use and flood risk in Luján.



Land use class	Area (ha)
Agua, nieve o hielo natural	47
Áreas cultivadas y/o manejadas	1667.1
Áreas naturales terrestres vegetadas	625.4
Superficies artificiales y áreas asociadas	753.3

Future work and new features.

Future work and new features.

The results obtained in the development of *osmlanduseR* represent a preliminary model for the incorporation of additional functions and other features documented in previous research.

Future work and new features.

Linear elements

Linear elements, such as those associated with the road and rail network or waterways, could be included in the land use analysis, provided that appropriate assumptions are made about the width of each linear element, as suggested by Fonte et al. (2016).



Future work and new features.

Inference from other map elements

The current stage of the *osmlanduseR* package enables the incorporation of a tool that allows the inference of land use from elements such as POIs, buildings, road network, tags combination, etc.

This can be accomplished by using the tree decision approach proposed by A. J. Arsanjani et al. (2013).



Future work and new features.

Filling gaps

In their study, Schultz et al. (2017) described a method for using the information available in OSM as a training tool for supervised classification of remote sensing data.

There are also R packages that can be integrated with this intention. For example, the *sits* package (Simoes et al. 2021).



Future work and new features.

Map accuracy

An important component to be added in the future is a method for estimating the accuracy of a classification derived from other sources or constructed by other methods in a given area.

Future work and new features.

OpenStreetMap data improvements

One feature to be added in future work is the ability to integrate the package with workflows for enhancing OSM map data.

The package can implement detection of problems such as overlapping, tag consistency, invalid geometries, etc.



Conclusions

Conclusions

The work conducted since the initial development of *osmlanduseR* has made it possible to use data from OSM for land use analysis and to implement a classification system.

Additionally, the developed package provides a basis for integrating additional complementary methodologies referenced in previous research into a unified platform.



Conclusions

Regarding them, it is critical to continue improving *osmlanduseR* to allow the use of OSM tags as training to categorize unknown areas from remotely sensed imagery, and inferring land use classes from POIs, buildings, or linear features. It is also necessary to incorporate methods for calculating map accuracy.

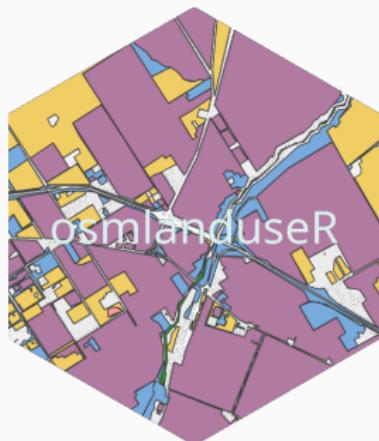


Conclusions

Finally, the approach of *osmlanuseR* enables a systematic analysis of the information available in OSM, resulting in a new tool that can be integrated into participatory mapping processes and used to enhance initiatives for open and public access to information.



Thank you! - ¡Muchas gracias!



Andrés Duhour: aduhour aduhour@unlu.edu.ar



References

References i

- Arsanjani, Amal Jokar, Marco Helbich, Mohamed Bakillah, Julian Hagenauer, and Alexander Zipf. 2013. "Toward Mapping Land-Use Patterns from Volunteered Geographic Information." *International Journal of Geographical Information Science*.
- Arsanjani, Jamal Jokar, Alexander Zipf, Peter Mooney, and Marco Helbich, eds. 2015. *OpenStreetMap in GIScience: Experiences, Research, and Applications*. Springer.
- <https://doi.org/10.1007/978-3-319-14280-7>.



References ii

- da Silva, Gean Carlos Gonzaga, Priscila Celebrini de Oliveira Campos, Marcelo de Miranda Reis, and Igor Paz. 2024. "Spatiotemporal Land Use and Land Cover Changes and Associated Runoff Impact in Itaperuna, Brazil." *Sustainability* 16 (1). <https://doi.org/10.3390/su16010325>.
- Duféal, Marina, and Matthieu Noucher. 2017. "Des TIC Au TOC. Contribuer à OpenStreetMap: Entre Commun Numérique Et Utopie Cartographique." *Communs Urbains Et Équipements Numériques* 31 (1/2).
<https://journals.openedition.org/netcom/2635>.

References iii

- Fonte, C., M. Minghini, V. Antoniou, L. See, J. Patriarca, M. Brovelli, and G. Milcinski. 2016. "An Automated Methodology for Converting OSM Data into a Land Use/Cover Map." In *6th International Conference on Cartography & GIS*.
<http://pure.iiasa.ac.at/id/eprint/13152/>.
- Harvey, David. 2004. "The 'New' Imperialism: Accumulation by Dispossession." *Socialist Register* 40: 63–87.
- Liu, Bin, Jie Yang, Jinxia Sha, Yun Luo, Xian Zhao, and Ruiting Liu. 2023. "Analysis of Runoff According to Land-Use Change in the Upper Hutuo River Basin." *Water* 15 (6).
<https://doi.org/10.3390/w15061138>.

- Padgham, Mark, Bob Rudis, Robin Lovelace, and Maëlle Salmon. 2017. “Osmdata.” *The Journal of Open Source Software* 2 (14). <https://doi.org/10.21105/joss.00305>.
- Pal, Sujan, Francina Dominguez, Pablo Bollatti, Steven P. Oncley, Yi Yang, Javier Alvarez, and Carlos M. Garcia. 2021. “Investigating the Effects of Land Use Change on Subsurface, Surface, and Atmospheric Branches of the Hydrologic Cycle in Central Argentina.” *Water Resources Research* 57 (11). <https://doi.org/10.1029/2021wr029704>.

- Palmisano, Tomás. 2018. *Tierras de Alguien: Una Propuesta de Análisis Sobre Las Transformaciones Económico-Productivas En La Actividad Agropecuaria Bonaerense Durante Los Albores Del Siglo XXI*. Teseo. <https://www.teseopress.com/tierrasdealguien>.
- Pattison, I., and S. N. Lane. 2012. “The Link Between Land-Use Management and Fluvial Flood Risk: A Chaotic Conception?” *Progress in Physical Geography* 36 (1): 72–92.
- Pebesma, E. 2018. “Simple Features for r: Standardized Support for Spatial Vector Data.” *The R Journal* 10 (1): 439–46. [https://doi.org/https://doi.org/10.32614/RJ-2018-009](https://doi.org/10.32614/RJ-2018-009).

- Pintos, Patricia, and Patricio Narodowski, eds. 2012. *La Privatopía Sacrílega. Efectos Del Urbanismo Privado En Humedales de La Cuenca Baja Del Río Luján.* 1era ed. Imago Mundi.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schultz, Michael, Janek Vossa, Michael Auera, Sarah Carterb, and Alexander Zipf. 2017. “Open Land Cover from OpenStreetMap and Remote Sensing.” Translated by mav. *International Journal of Applied Earth Observation and Geoinformation.*

References vii

- Simoes, Rolf, Gilberto Camara, Gilberto Queiroz, Felipe Souza, Pedro Andrade, Lorena Santos, Alexandre Carvalho, and Karine Ferreira. 2021. "Satellite Image Time Series Analysis for Big Earth Observation Data." *Remote Sensing* 13 (13): 2428. <https://doi.org/10.3390/rs13132428>.
- Sugianto, Sugianto, Anwar Deli, Edy Miswar, Muhammad Rusdi, and Muhammad Irham. 2023. "The Effect of Land Use and Land Cover Changes on Flood Occurrence in Teunom Watershed, Aceh Jaya." *Land* 11 (8). <https://doi.org/10.3390/land11081271>.

- Tennekes, Martijn. 2018. “tmap: Thematic Maps in R.” *Journal of Statistical Software* 84 (6): 1–39.
<https://doi.org/10.18637/jss.v084.i06>.
- Volante, José N. 2009. “Monitoreo de La Cobertura y El Uso Del Suelo a Partir de Sensores Remotos.” Instituto Nacional de Tecnología Agropecuaria.