# **populR:** an R package for transforming population data

Marios Batsaris, Dimitris Kavroudakis

**Marios Batsaris, PhD Candidate**
m.batsaris@aegean.gr
www.mbatsaris.gr

# Need

Importance

- ➢ Where is the population?

Applications

- ➢ Disaster management and planning
- ➢ Location analysis
- ➢ Decision making and urban planning

# Problem

- Population data → Census (ELSTAT)
- Optimal level of detail (LoD) - household
- Aggregated in respect of statistical confidentiality
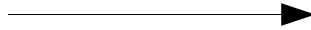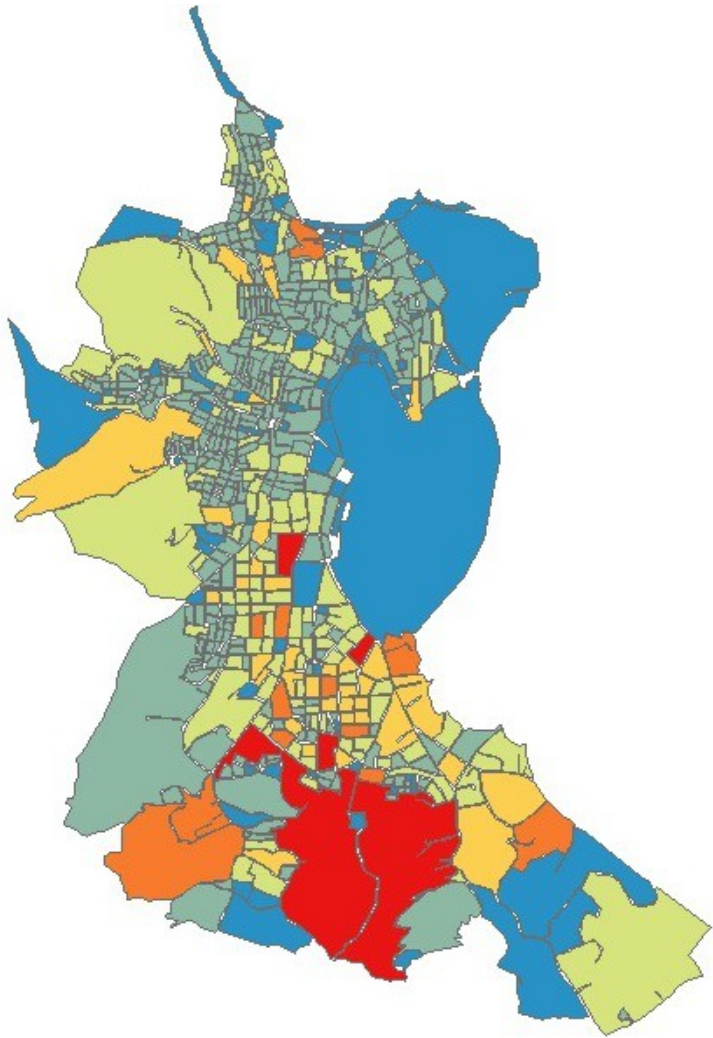- Residential population

## User defined LoD ≠ ELSTAT LoD

| Γεωγραφικός Κωδικός | Περιγραφή | Αριθμός οικοδομικού τετραγώνου | Μόνιμος πληθυσμός |
|---|---|---|---|
| 530101010 1 | Μυτιλήνη,η | 0000 | 190 |
| | | 0001 | 4 |
| | | 0007 | 1 |
| | | 0008 | 4 |



| FID | Shape * | ESYECODE |
|---|---|---|
| 0 | Polygon | 1 |
| 1 | Polygon | 2 |
| 2 | Polygon | 3 |
| 3 | Polygon | 9999 |
| 4 | Polygon | 4 |
| 5 | Polygon | 5 |
| 6 | Polygon | 6 |
| 7 | Polygon | 7 |
| 8 | Polygon | 8 |
| 9 | Polygon | 9 |
| 10 | Polygon | 10 |
| 11 | Polygon | 11 |
| 12 | Polygon | 12 |
| 13 | Polygon | 13 |
| 14 | Polygon | 14 |
| 15 | Polygon | 15 |
| 16 | Polygon | 16 |
| 17 | Polygon | 17 |
| 18 | Polygon | 18 |
| 19 | Polygon | 19 |
| 20 | Polygon | 20 |
| 21 | Polygon | 21 |
| 22 | Polygon | 22 |

# Problem

# Problem

➢ Despite the massive presence of down-scaling methods in the scientific literature there is a lack of implementation tools for due to increased complexity of existing methods

# Aim

Create and R extension to impel the use of population data transformation into finer scale

➢ Transform population data into finer scale

## By using:

➢ Down-scaling methods

➢ VGI (Volunteered Geographic Information)

# Tools

R programming language and Rstudio IDE with dependencies

- ➢ simple features - sf
- ➢ osmdata
- ➢ dplyr
- ➢ units

Package building

- ➢ roxygen
- ➢ devtools
- ➢ testthis
- ➢ rtools

# Availability

populR is available:

➢ Github → https://github.com/mbatsaris/populR

➢ CRAN → https://cran.r-project.org/web/packages/populR/

# populR

## Methods:

➤ area-based method → down-scaling using the area of intersection of target features

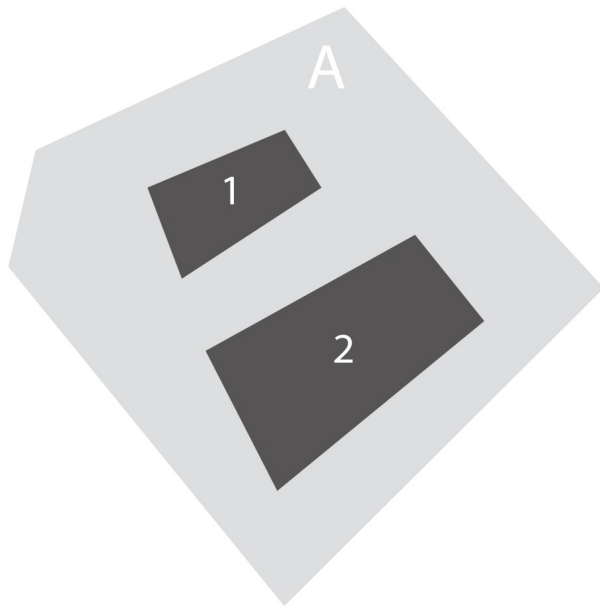| Advantages | Disadvantages |
|---|---|
| Easy and straightforward method | Even distribution throughout the source features |
| No need of ancillary information | |

# PopulR - contribution

New Methods:

➤ volume-based method → down-scaling using the area of intersection multiplied by the number of floors or height of target features

# PopulR – in depth

## Area-based method:



A - Population count: 20
1 - area: 50m$^2$
2 - area: 100m$^2$

## Population estimation
area -based method

intersected area
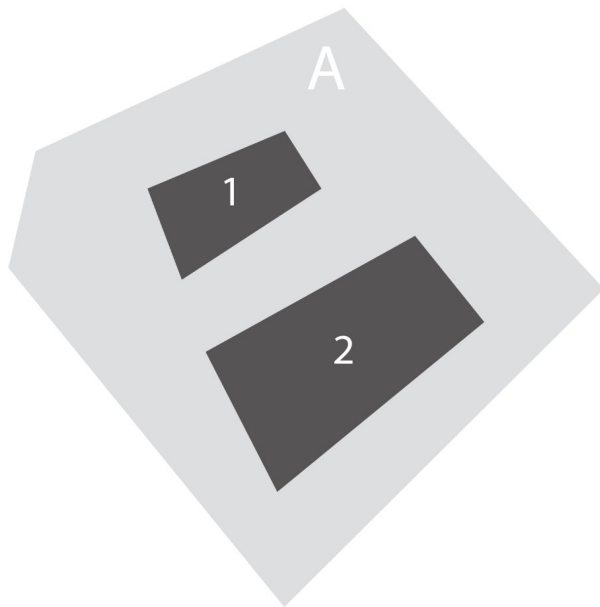50 + 100 = 150m$^2$

areal weights
50/150 = 0.333
100/150 = 0.667

population estimation
1 pop: 0.333 x 20
2 pop: 0.666 x 20

# PopulR – in depth

## Volume-based method:

A

1

2

A - Population count: 20
1 - area: 50m$^2$, floors: 4,
2 - area: 100m$^2$, floors: 1

## Population estimation
volume -based method

intersected area
4 x 50 + 1 x 100 = 300m$^2$

areal weights
200/300 = 0.667
100/300 = 0.333

population estimation
1 pop: 0.666 x 20
2 pop: 0.333 x 20

# Data

| Data | Format | Source | Year |
|---|---|---|---|
| Blocks - source | Spatial | Census - ELSTAT | 2011 |
| Buildings - target | Spatial | Census - ELSTAT | 2001 |

# Numerical Experiments
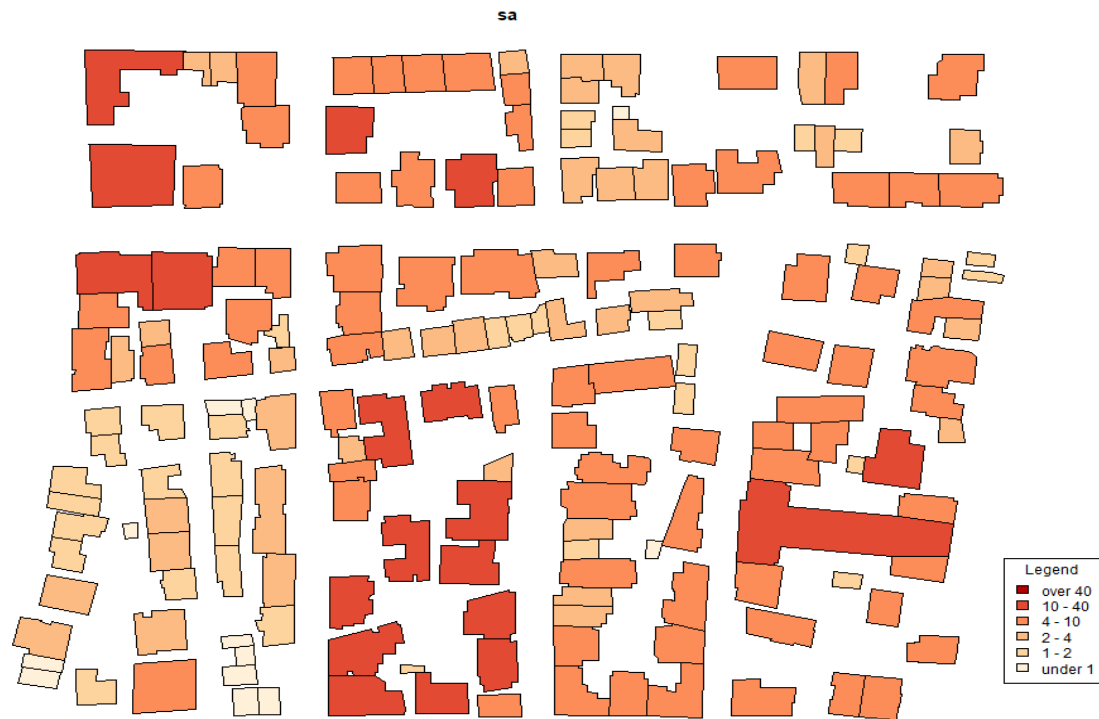


Subset area of Mytilini,
Greece

9 blocks (source zone)
179 buildings (target zone)
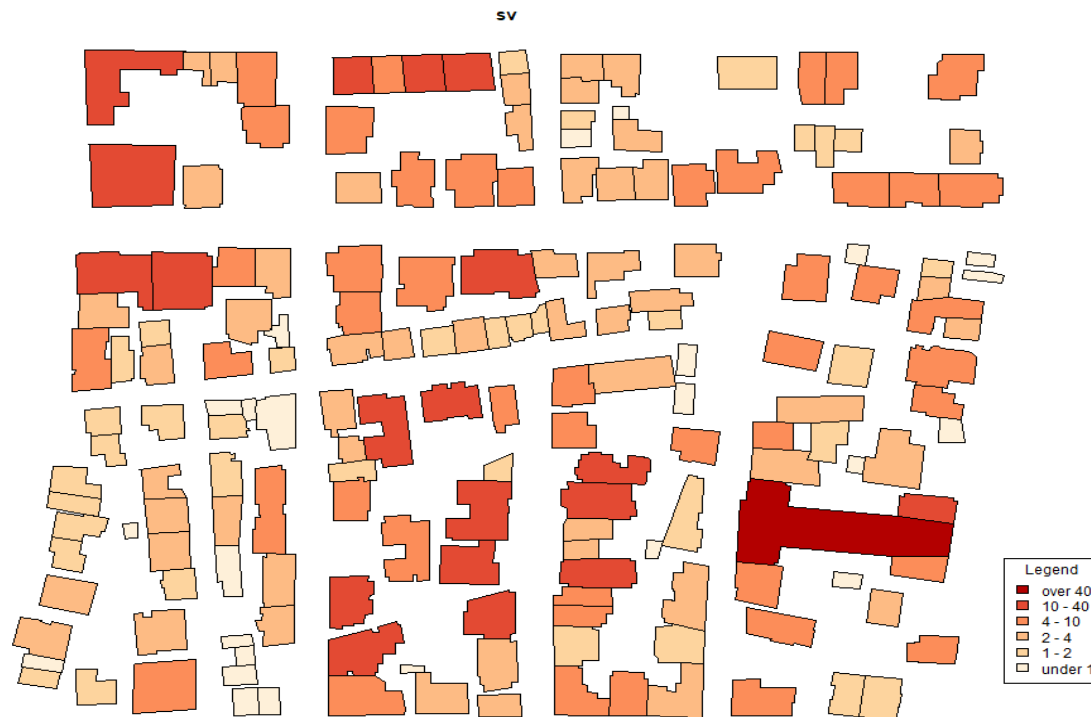Population: 911

# Results

Area-based method:

*pp_estimate(source = source, target = target, sourcecode = sid, sourcepop = pop)*
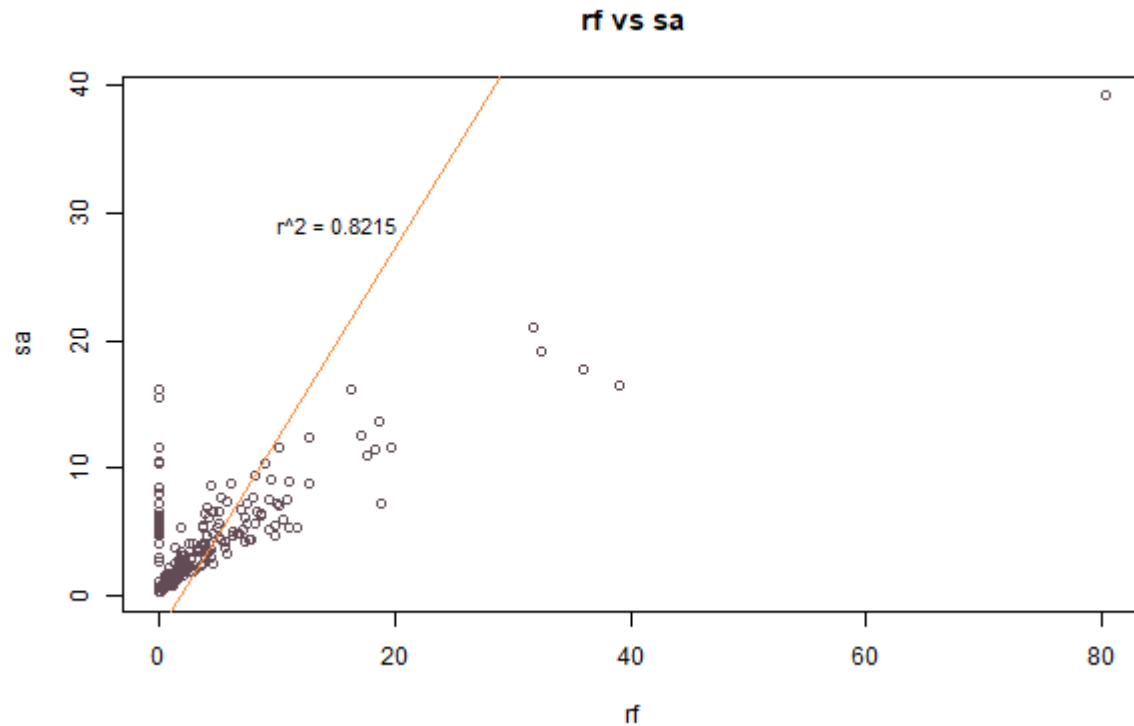
# Results

Area-based method:

*pp_estimate(source = source, target = target, sourcecode = sid, sourcepop = pop, volume = floors)*
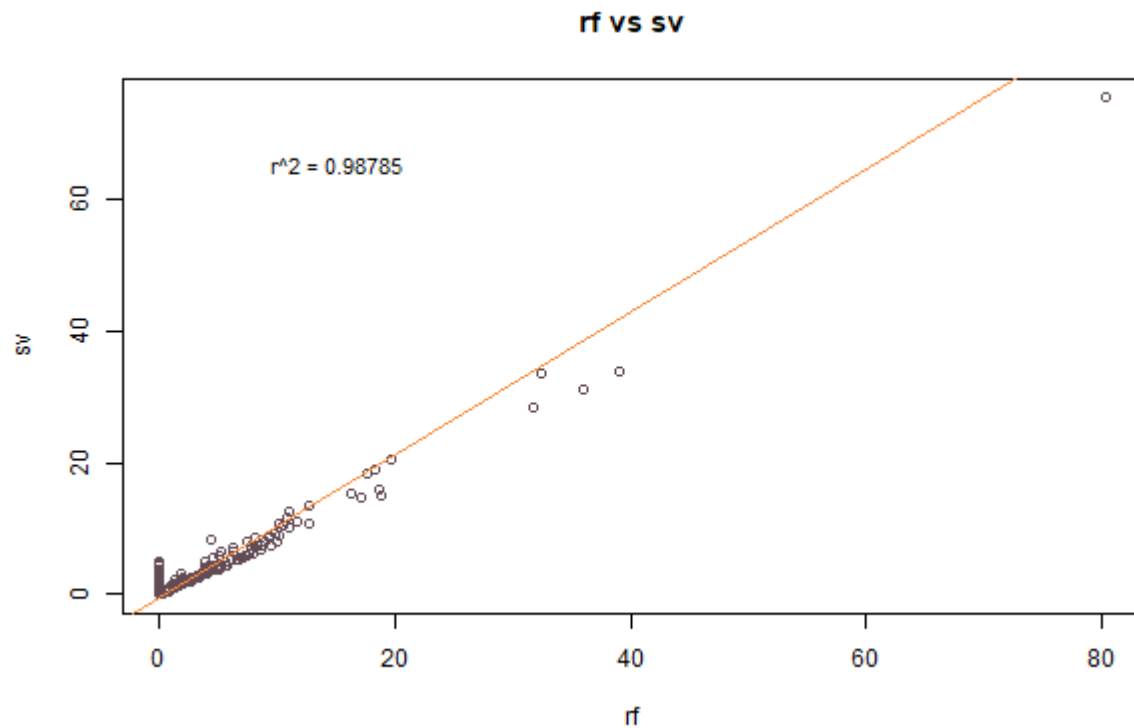
# Results
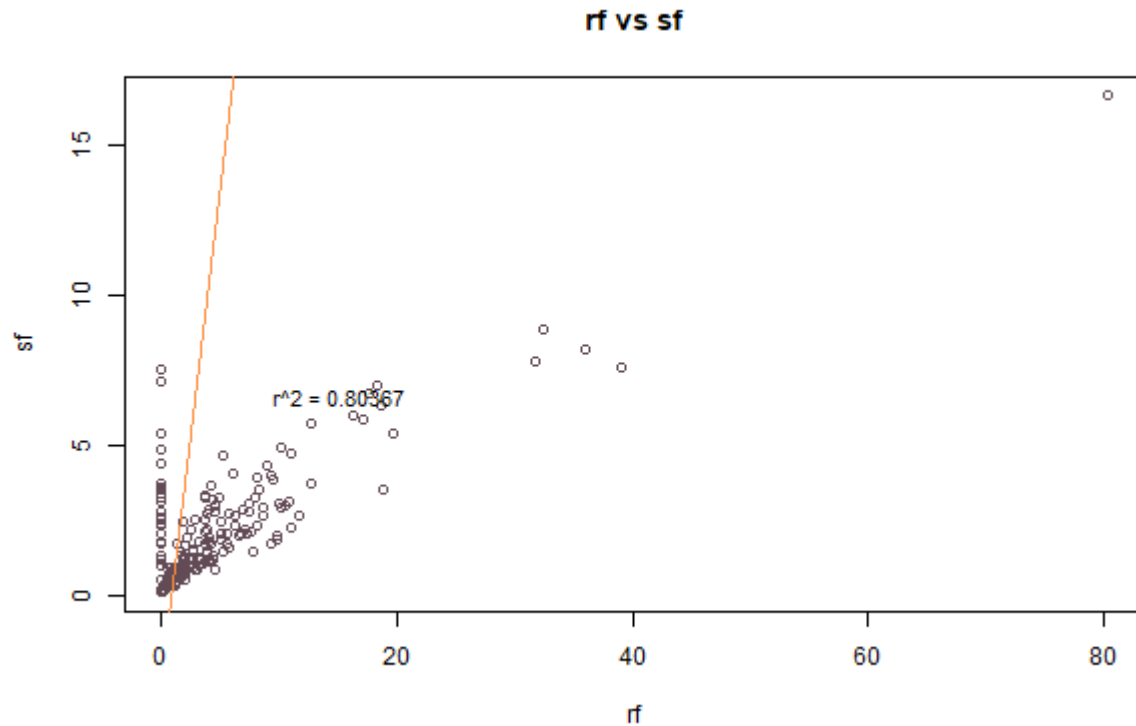
Comparison of area-based method with reference data:

# Results

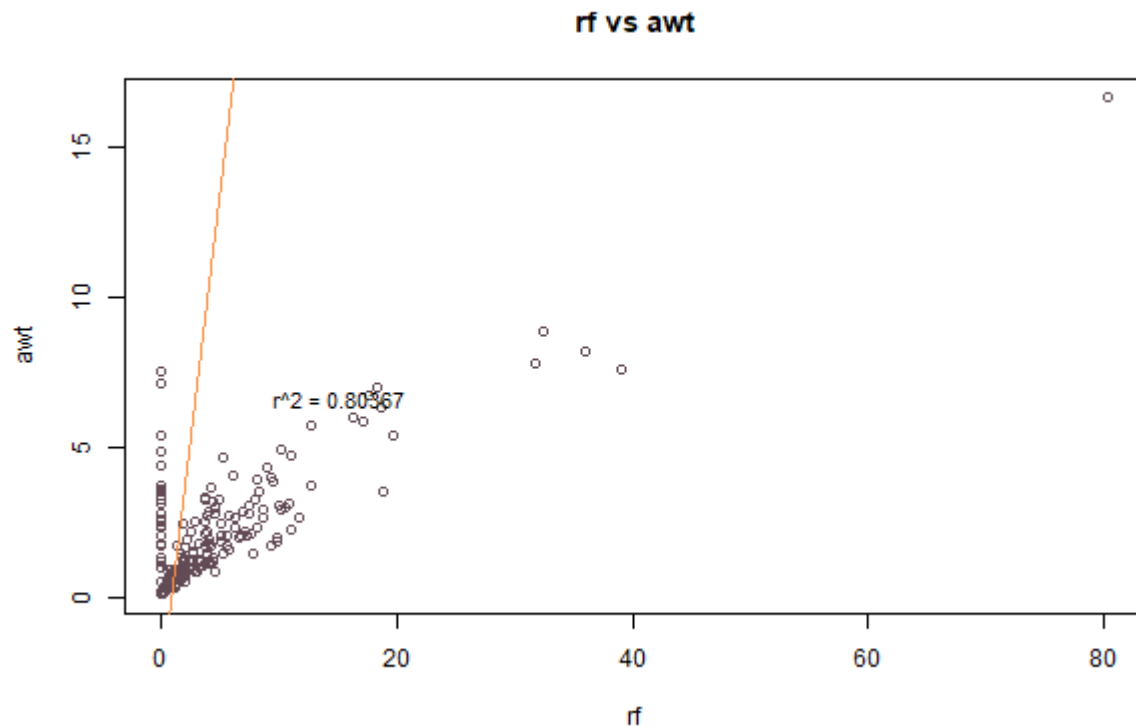Comparison of volume-based method with reference data:

# Results

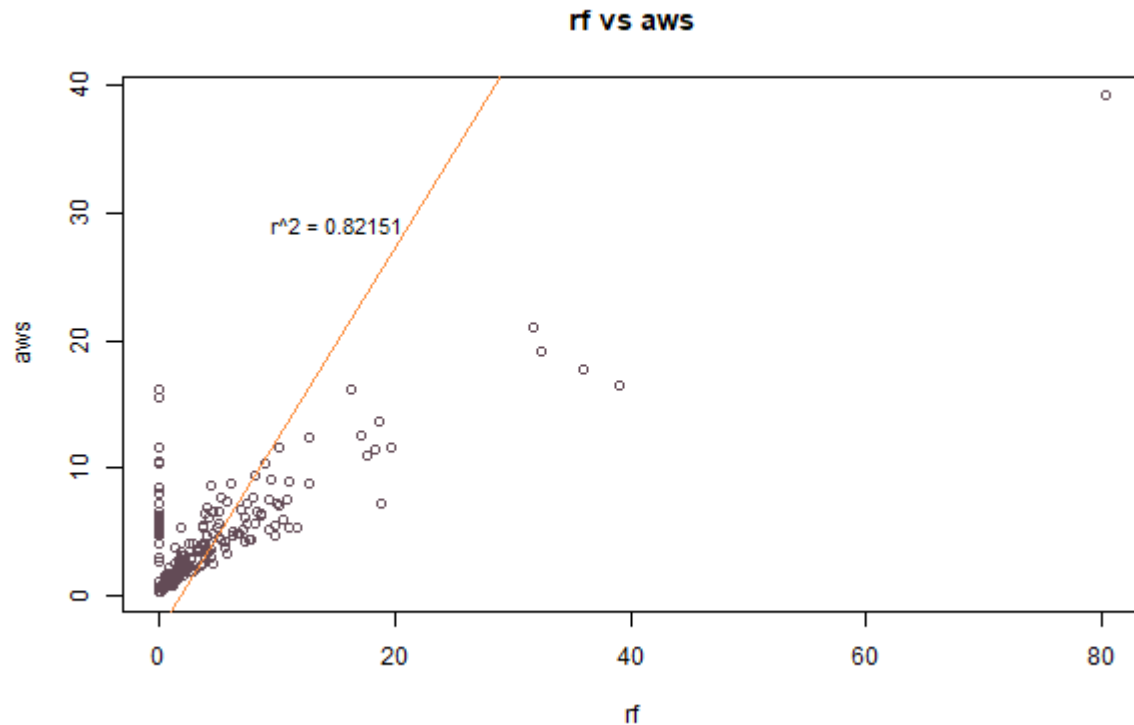Comparison to other alternatives – sf: st_interpolate_aw()



rf vs sf

# Results

Comparison to other alternatives – areal: aw_interpolate() - total weights

# Results

Comparison to other alternatives – areal: aw_interpolate() - sum weights

# Results

## RMSE and MAE values

| rf vs method | RMSE | MAE |
| --- | --- | --- |
| area-based | 5.33 | 2.75 |
| Volume based | 1.45 | 0.94 |
| sf | 7.41 | 3.66 |
| Areal: total weights | 7.41 | 3.66 |
| Areal: sum wights | 5.33 | 2.75 |

# Discussion

populR advantages:

➤ Simple and fast usage

➤ Offers the ability to add volume as extra parameter to calculate the weights

➤ Provides better results than existing alternatives (the volume-based method)

# Conclusions

Future improvements

➢ Use VGI as ancillary information to guide the down-scaling process

➢ Include other down-scaling methods

➢ Temporal estimations (daytime population)

# The end

Thank you for your time!