Analiza geo-prostornih podataka u R-u

Milutin Pejović & Petar Bursać

BelgradeR Meetup

2022-12-14, Startit, Beograd

doc. dr Milutin Pejović, dipl.geod.inž.

Petar Bursac, mast.inž.geod.

Građevinski fakultet Univerziteta u Beogradu
Odsek za Geodeziju i Geoinformatiku

- 2004 - Master akademske studije Geoinformatike.
- 2015 - Laboratorija za razvoj geoporostornih tehnologija otvorenog koda OSGL
- Kursevi i konferencije:
  - 2008 - GEOSTAT Belgrade workshop - lectures: Tomislav Hengl
  - 2011 - GEOSTAT 2011 workshop - lectures: Victor Olaya, Tomislav Hengl
  - 2014 - World Daily Meteo Conference - keynote speakers: Edzer Pebesma, Gerard Heuvelink, Tomislav Hengl
  - 2016 - GeoMLA (Geostatistics and Machine Learning) conference; keynote speakers: Mikhail Kanevski, Mirko Orlić, Wolfgang Wagner, Ole Einar Tveito, Tomislav Hengl.
- 2022 - Geoinformatika - novi studijski program na Građevinskom fakultetu

Geo-spatial data science

"Spatial data science treats location, distance, and spatial interaction as core aspects of the data and employs specialized methods and software to store, retrieve, explore, analyze, visualize and learn from such data. In this sense, spatial data science relates to data science as spatial statistics to statistics, spatial databases to databases, and geocomputation to computation."

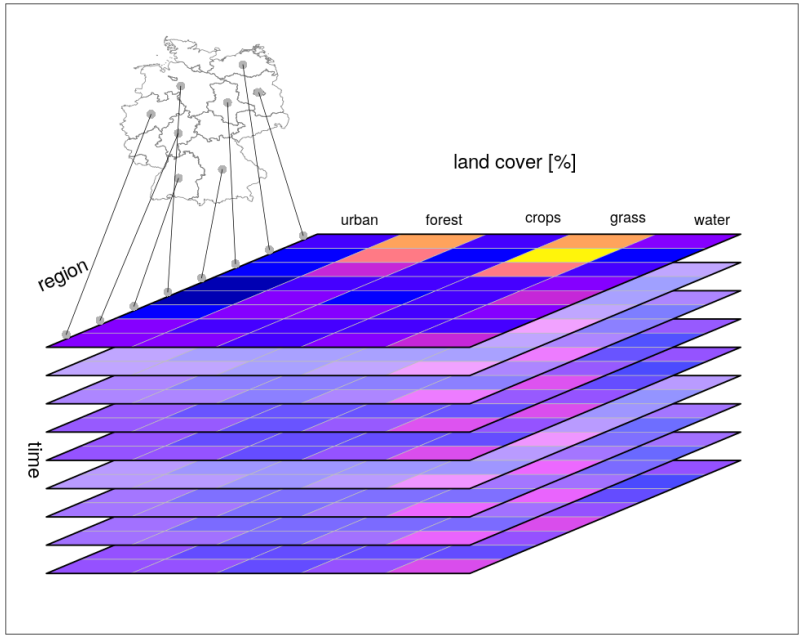
Luc Anselin, 2019

"Spatial Data Science" in The International Encyclopedia of Geography: People, the Earth, Environment, and Technology.

Šta su to geo-prostorni podaci?

Šta su to geo-prostorni podaci?

- Kombinuju informaciju sa geografskom lokacijom i vremenom geografskom lokacijom i vremenom na koje se odnose
- Geo-prostorni vs. prostorni podaci
- Digitalno predstavljaju prirodne i antropogene objekte, prostorne pojave i dogadjaje stavljajući ih u kontekst geografskog prostora
- Veliku ulogu je odigrao GPS!
- In 2008, more than 150 Earth observation satellites were in orbit, recording data with both passive and active sensors and acquiring more than 10 terabits of data daily. (Fifty Years of Earth-observation Satellites by A. Tatem, S. Goetz, S. Hay)
- Danas je preko 80% digitalnih podataka referisano na geografsku lokaciju (bez reference)

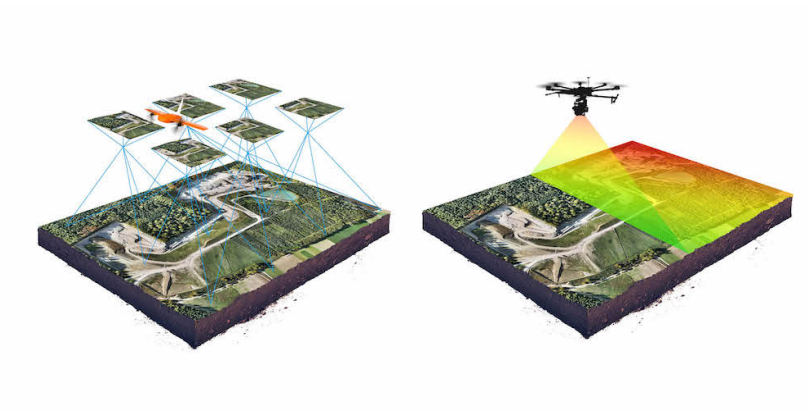


Izvori geo-prostornih podataka

- Javni podaci državnih službi (prostorni inventari, premer, popis i mape)
- Podaci permanentnih mernih stanica i senzora (senzori)
- Podaci daljinske detekcije (GNSS, LiDAR i satelitski snimci)

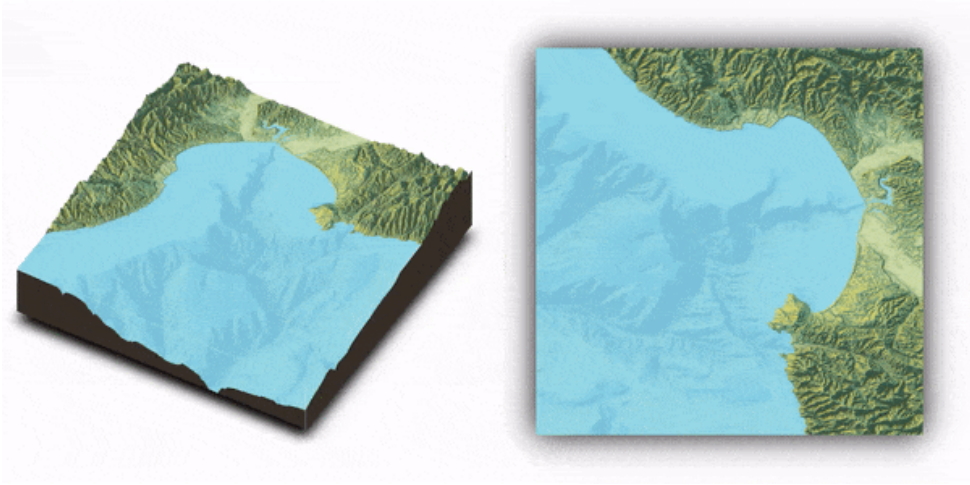
- Personalni podaci i podaci sa društvenih mreža

- Javno dostupni globalni podaci:
  - http://freegisdata.rtwilson.com/
  - https://data.nasa.gov/
  - https://land.copernicus.eu/
  - https://maps.elie.ucl.ac.be/CCI/viewer/
  - https://www.ecad .eu//dailydata/predefinedseries.php
  - http://www.worldclim.org/version2
  - https://www.geoportal.org/
  - ...



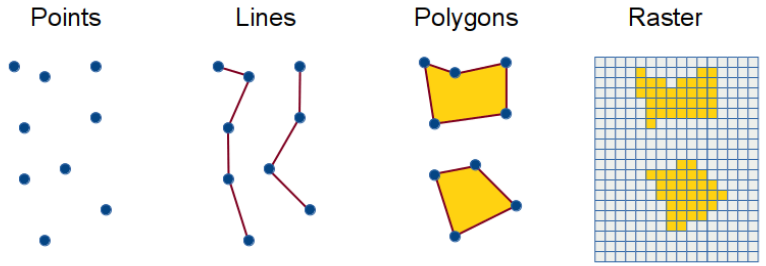
Primeri primene geo-prostornih podataka

- Ekologija
- Daljinska detekcija (satelitski snimci)
- Ekonomija
- Demografija
- Arheologija
- Transport
- Klimatologija
- Meteorologija
- Geo-morfometrija
- Hidrologija
- Poljoprivreda
- Rudarstvo
- Zemljište
- Turizam
- i mnogi drugi...

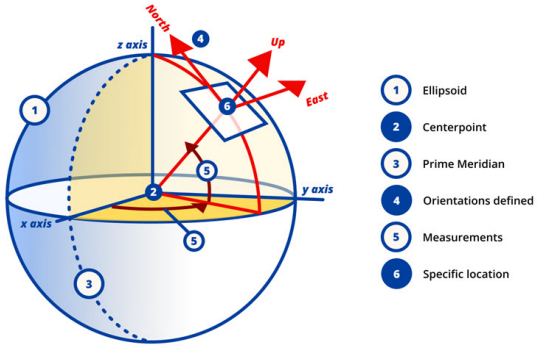


Komponente geo-prostornih podataka

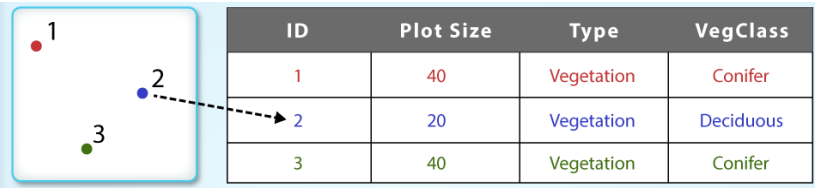
- Geometrija



- Koordinate i koordinatni sistem



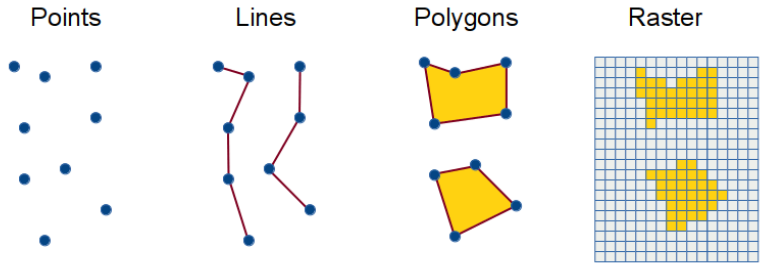
- Atributi i razmera

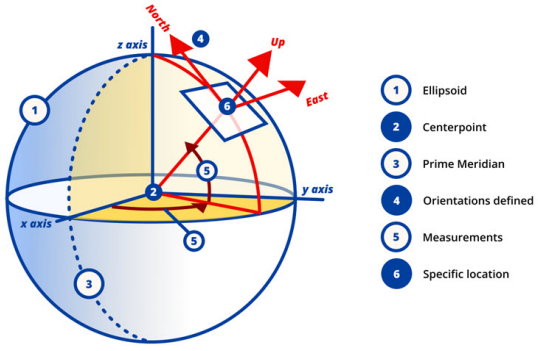


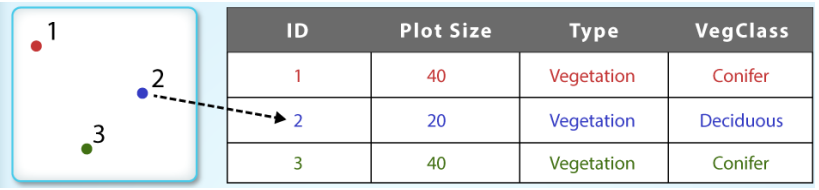
ID	Plot Size	Type	VegClass
1	40	Vegetation	Conifer
2	20	Vegetation	Deciduous
3	40	Vegetation	Conifer

Komponente geo-prostornih podataka

- Geometrija
- Koordinate i koordinatni sistem
- Atributi i razmera







ID	Plot Size	Type	VegClass
1	40	Vegetation	Conifer
2	20	Vegetation	Deciduous
3	40	Vegetation	Conifer

Koordinate i Koordinatni Referenti Sistem (CRS)

"Data are not just numbers, they are numbers with a context; "In data analysis, context provides meaning" (Cobb and Moore, 1997)

"Coordinates are not just numbers, they are numbers with a spatial context"

Koordinatni Referentni Sistem (CRS) definiše prostorni kontekst. On predstavlja set informacija kojim se opisuje geografski prostoru u matematičkom smislu (u odnosu na Zemlju) i sastoji se od:

- set matematičkih pravila kojim se definiše kako se koordinate dodeljuju tačkama u prostoru (koordinatni sistem)
- set parametara koji definišu poziciju početka koordiatnog sistema, razmera i orijentacija koordinatnog sistema (datum)

Možemo razlikovati dve glavne grupe koordinatnih referentih sistema:

- Geografski koordinatni sistemi
- Koordinatni sistemi u projekciji

Opis koordinatnog sistema

Do sada najzastupljeniji, ali u svom zalasku:

- `proj4string` - +proj=longlat +datum=WGS84 +no_defs

Dva preporučena pristupa:

- CRS codes, e.g., EPSG:4326
- `WKT2`

Više informacija:

- https://proj.org/usage/projections.html
- https://spatialreference.org/ref/epsg/
- https://www.youtube.com/watch?v=Va0STgco7-4
- https://www.gaia-gis.it/fossil/libspatialite/wiki?name=PROJ.6

```
## Coordinate Reference System:
##   User input: EPSG:4326
##   wkt:
## GEOGCRS["WGS 84",
##     DATUM["World Geodetic System 1984",
##         ELLIPSOID["WGS 84",6378137,298.257223563,
##             LENGTHUNIT["metre",1]]],
##     PRIMEM["Greenwich",0,
##         ANGLEUNIT["degree",0.0174532925199433]],
##     CS[ellipsoidal,2],
##         AXIS["geodetic latitude (Lat)",north,
##             ORDER[1],
##             ANGLEUNIT["degree",0.0174532925199433]],
##         AXIS["geodetic longitude (Lon)",east,
##             ORDER[2],
##             ANGLEUNIT["degree",0.0174532925199433]],
##     USAGE[
##         SCOPE["Horizontal component of 3D system."],
##         AREA["World."],
##         BBOX[-90,-180,90,180]],
##     ID["EPSG",4326]]
```

Geo-prostorni podaci u R-u

Istorija

- pre-2003. - razvoj drugih paketa namenjenih prostornoj analizi (MASS, `spatstat`, `geoR` itd.)
- 2003. - Vienna workshop, početak rada na `sp` paketu i definisanju klasa i metoda namenjenih geo-prostornim podacima
- 2003. - Publikovan paket `rgdal`
- 2005. - Publikovan paket `sp`
- 2005. - Applied Spatial Data Analysis with R, 1st ed.
- 2010. - Publikovan paket `raster`
- 2011. - Publikovan paket `rgeos`
- 2013. - Applied Spatial Data Analysis with R, 2nd ed.
- 2016. - Publikovan paket `sf`
- 2019. - Publikovan paket `stars`
- 2020. - Publikovan paket `terra`

Danas je dostupno preko 1000 paketa

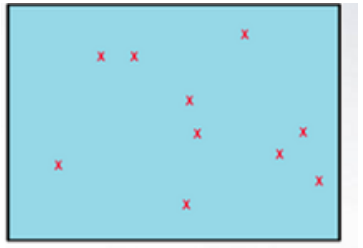
- **sf**, **sp**, **terra**, **raster**, **stars**, **gdalcubes** - klase i metode za skladištenje geo-prostornih podataka
- **rnaturalearth**, **osmdata**, **getlandsat** - download geo-prostornih podataka
- **rgrass7**, **qgisprocess**, **RSAGA**, **link2GI** - veza sa GIS softverima
- **gstat**, **mlr3**, **CAST**, **spatstat**, **spdep**, **spatialreg** - modeliranje geo-prosotrnih podataka
- **rasterVis**, **tmap**, **ggplot2**, **rayshader** - statična vizualizacija i kartiranje
- **leaflet**, **mapview**, **mapdeck** - interaktivna vizualizacija i kartiranje
- i mnogi drugi...

CRAN task view: https://cran.r-project.org/web/views/Spatial.html.

Vektorski model podataka

Vektorski podaci

- Tačke, linije, poligoni (kombinacije osnovnih geometrijskih entiteta)
- Predstavljaju diskretne objekte u prostoru
  - Uzorkovanja (merenja) na određenim lokacijama
  - Granice područja
  - Trase i trajektorije
- Poznati fajl formati: ESRI Shapefile (.shp, .shx, .dbf, .prj), GeoJSON, GPX (.gpx), geopackage (.gpk)


Point features


Line features


Polygon features

sf paket (revolucija u radu sa geo-prostornim podacima)

- sf je naslednik (2016) sp paketa koji omogućava klase i metode za vektorske geo-prostorne podatke.
- sf implmenetira OGC standard `ISO 19125-1:2004` Simple Features za reprezentovanje geo-prostornih podataka
- sf omogućava funkcionalnosti dobro poznatih i proverenih biblioteka: GDAL, PROJ and GEOS.
- sf klase su nadgradnja `data.frame` strukture!
- Komplementaran je sa tidyverse familijom paketa (`geom_sf` geometrija za ggplot)!

sf ekosistem

OSGeo

PROJ

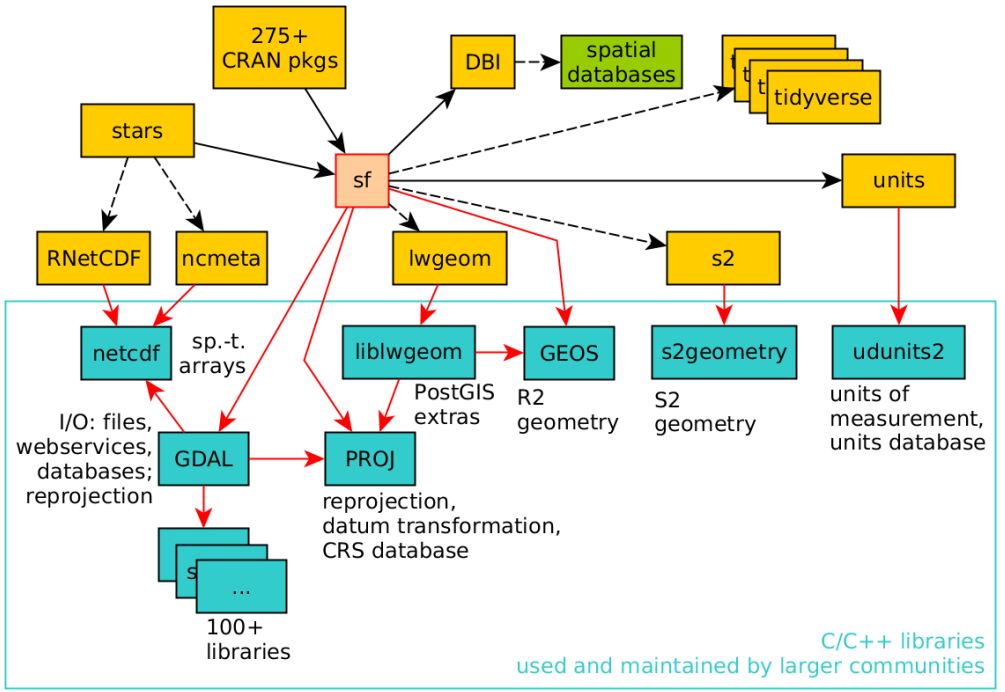
- "Konverzija koordinata"
- Prva verzija 1970s

GDAL

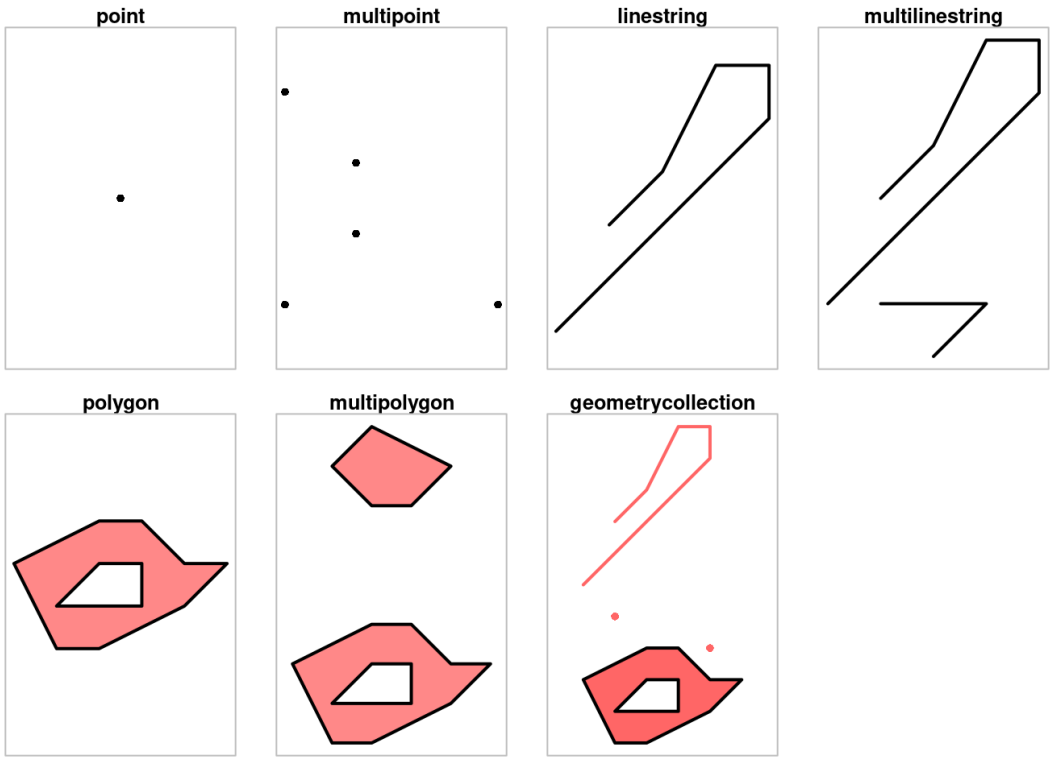
- Geospatial Data Abstraction Library, prva verzija 2000
- "200+ geo-prostornih fajl formata"

GEOS

- Geometry Engine – Open Source

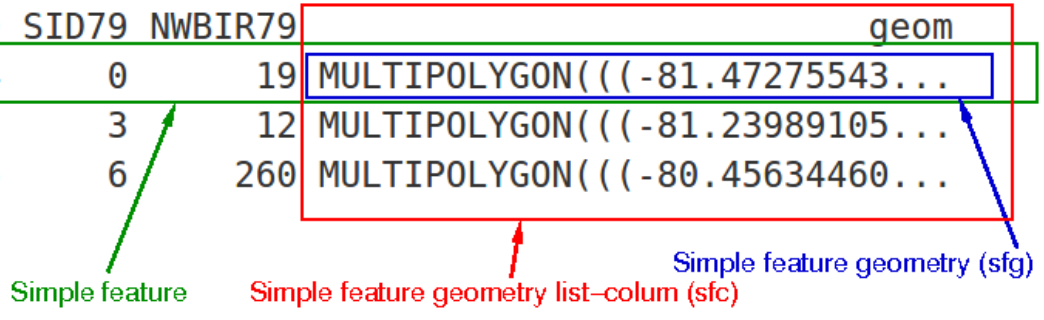


sf geometrije



sf klasa inside

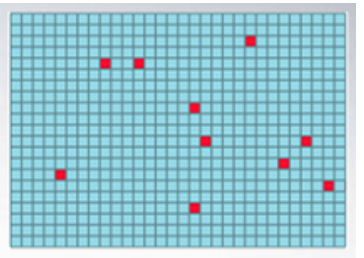
```
## Simple feature collection with 100 features and 6 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -84.32385 ymin: 33.88199 xmax: -75.45698 ymax: 36.58965
## epsg (SRID):    4267
## proj4string:    +proj=longlat +datum=NAD27 +no_defs
## precision:      double (default; no precision model)
## First 3 features:
##   BIR74 SID74 NWBIR74 BIR79 SID79 NWBIR79                              geom
## 1  1091     1      10  1364     0      19 MULTIPOLYGON(((-81.47275543...
## 2   487     0      10   542     3      12 MULTIPOLYGON(((-81.23989105...
## 3  3188     5     208  3616     6     260 MULTIPOLYGON(((-80.45634460...
```



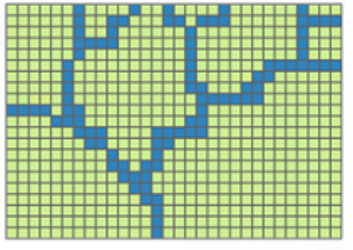
Rasterski model podataka

Rasterski podaci

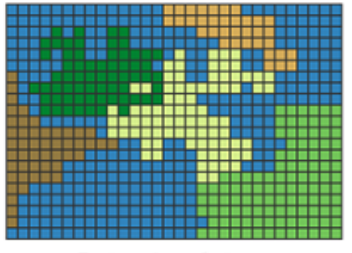
- Rasterski podaci predstavljaju mozaik manjih poligona (ćelija), najčešće kvadratnih i jednakih dimenzija, u kojem svaka ćelija predstavlja vrednost prostornog fenomena u odgovarajućoj prostornoj oblasti na koju se ćelija odnosi, u određenom trenutku
- Prirodni format satelitskih i aero-snimaka zemlje iz vazduha
- Odgovarajuči za predstavljanje kontinualnih objekata u prostoru (temperatura, nadmosrska visina, itd)
- Poznati fajl formati: GeoTIFF, Erdas Imagine Image (.img), HDF (.hdf), NetCDF (.nc)



Raster point features



Raster line features



Raster polygon features

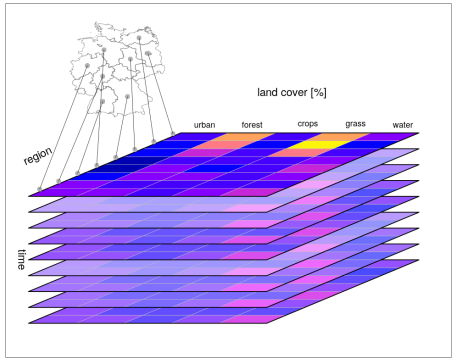
terra paket (naslednik raster paketa)

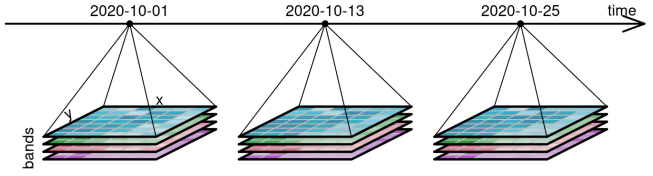
- **terra** omogućava klase (`SpatRaster`, `SpatVector`, `SpatExtent`) i metode za rasterske podatke
- Omogućava kreiranje, učitavanje, manipulaciju i pisanje jedno-dimenzionalnih i više-dimenzionalnih (single and multiband) rasterskih podataka
- Omogućava rad sa rasterskim podacima koji su skladišteni na disku (suviše veliki za učitavanje u memoriju)
- Omogućava funkcionalnosti rasterske algebre (lokalne, fokalne, zonalne, globalne)
- Dodatne funkcionalnosti (`crop`, `merge`, `aggregate`, `resample`) kao i dodatne funkcionalnosti za analizu DMT-a

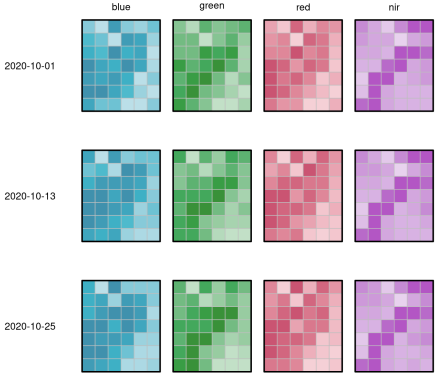
Data Cubes - Višedimenzionalne strukture geo-prostornih podataka

Data Cubes

- Uključivanje vremenske komponente u prostorne podatke
  - tro-dimenzionalni podaci (vektor)
  - četvoro-dimenzionalni podaci (raster)
  - više-dimenzionlani podaci
- Odgovarajući za predstavljanje podataka prostornog monitoringa







spacetime paket

- Publikovan 2010. godine
- Klase i metode za različite tipove prostorno-vremenskih vektorskih podataka
- Oslanja se na `sp` (prostor) i `xts` (vreme) klase.

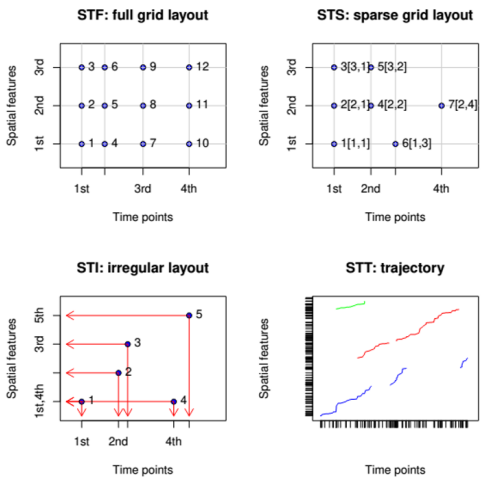
Journal of Statistical Software
November 2012, Volume 51, Issue 7. *doi: 10.18637/jss.v000.i00*

spacetime: Spatio-Temporal Data in R

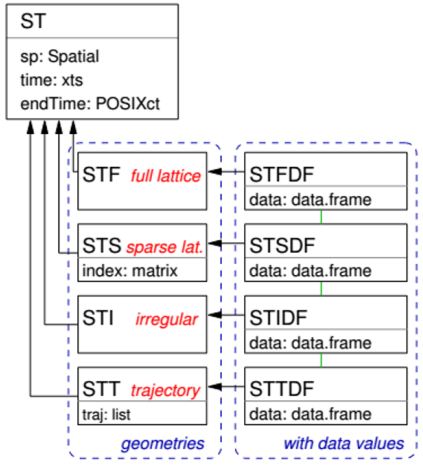
ifgi Institute for Geoinformatics University of Münster
52north exploring horizons

Edzer Pebesma

Slučajevi prostorno-vremenskih podataka



Klase u okviru „spacetime“ paketa



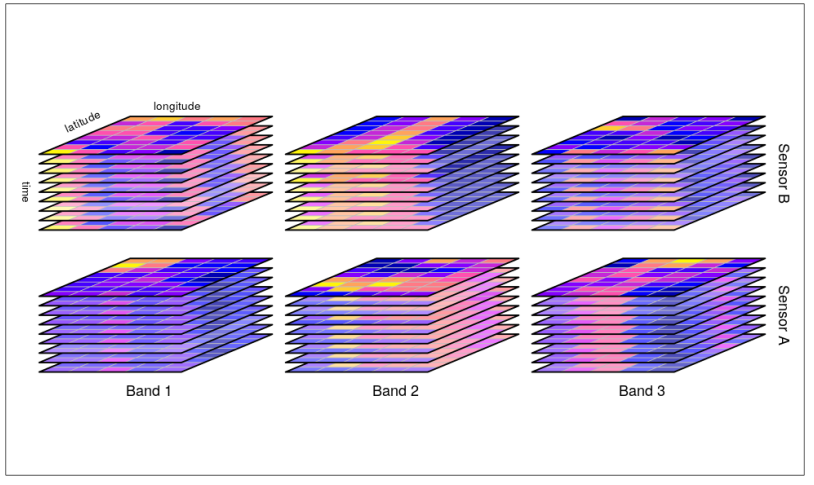
Abstract

This document describes classes and methods designed to deal with different types of spatio-temporal data in R implemented in the R package **spacetime**, and provides examples for analyzing them. It builds upon the classes and methods for spatial data from package **sp**, and for time series data from package **xts**. The goal is to cover a number of useful representations for spatio-temporal sensor data, and results from predicting (spatial and/or temporal interpolation or smoothing), aggregating, or subsetting them, and to represent trajectories. The goals of this paper are to explore how spatio-temporal data can be sensibly represented in classes, and to find out which analysis and visualisation methods are useful and feasible. We discuss the time series convention of representing time intervals by their starting time only. This vignette is the main reference for the R package `spacetime`; it has been published as Pebesma (2012), but is kept up-to-date with the software.

Keywords: Time series analysis, spatial data, spatio-temporal statistics, Geographic Information Systems.

stars paket

- **stars** obezbedjuje klase i metode za reprezentovanje prostorno-vremenskih (**rasterskih** i vektorskih) podataka
- Namenjen je radu sa vremenskim serijama prevashodno rasterskih podataka (satelitskim snimcima)
- Cilj razvijanja ovog paketa je da omogući:
    1. Rad sa rasterskim podacima koji je dobro integrisan sa `sf` i `tidyverse` paketima
    2. Rad sa rasterskim podacima kao više-dimenzionalnim nizovima (arrays)
    3. Rad sa velikim setovima podataka (out-of-memory) pomoću `raster_proxy` objekata
    4. Rad sa različitim vrstama rasterskih podataka (regular, rotated, sheared, rectilinear and curvilinear rasters)
- https://r-spatial.github.io/stars/



Harmonizacija vremenskih serija rasterskih podataka

`gdalcubes` je namenjen kreiranju harmonizovanih "data cubes" formi rasterskih podataka od različitih kolekcija (vremenskih serija) više-dimenzionalnih rasterskih podataka, u slučajevima kada:

- Prostorna rezolucija nije ista
- Prostorni obuhvat nije isti
- Koordinatni referentni sistemi nisu isti
- Vreme opažanja (uzorkovanja) nije isto

sftime paket

- `sftime` - predstavlja proširenje `sf` klasa za skladištenje i manipulaciju prostorno-vremenskim vektorskim podacima
- Prvi put publikovan u martu 2022. godine
- Kreiran je sa ciljem na nadomesti nedostatke `stars` paketa u pogledu geometrije i regularnosti vremenske serije podataka

Analiza geo-prostornih podataka

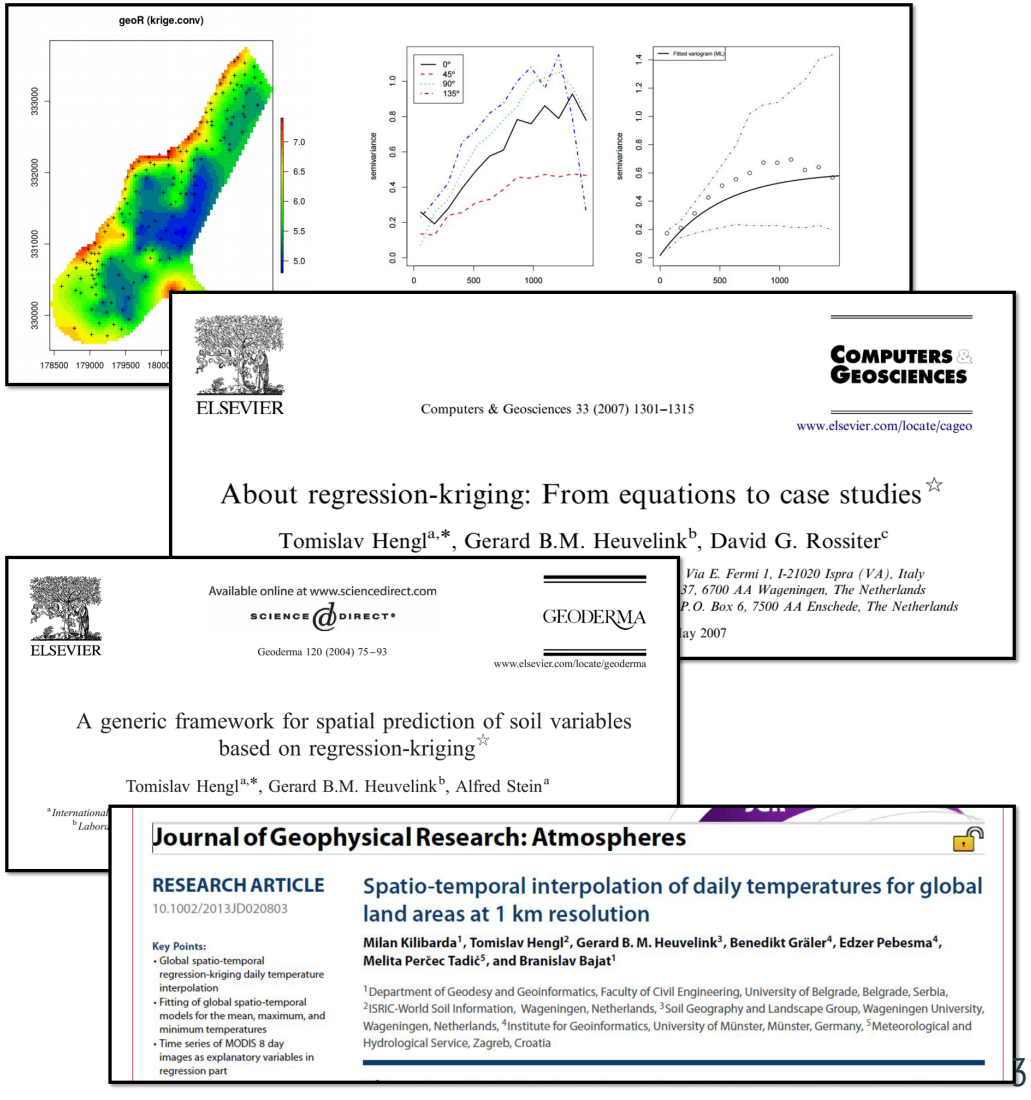
Šta obuhvata i zašto R?

1. Učitavanje i ispisivanje geo-prostornih podataka - gdal, osmdata, rnaturalearth, getlandsat
2. Skladištenje (klase i metode) podataka - sf (sp), terra (raster), stars, sftime
3. Pregled i manipulacija podacima - tidyverse
4. Deskriptivna i ekspanatorna analiza podataka - R
5. Konverzija i transformacija koordinata - proj
6. GIS analiza i kartiranje - ggplot, ggmap, maps, rasterVis tmap, mapview ...
7. Statistična i geostatistička analiza - gstat, spdep, spatstat ...
8. Modeliranje i prostorna predikcija - mlr, tidymodels
9. Komunikacija i publikacija - shiny, rmarkdown (quarto)

Tajna je u jedinstvenoj i potpuno zaokruženoj kombinaciji fantastičnih mogućnosti R-a i specijalizovanih paketa za geoprostorne podatke

Retrospektiva: do-2015.

- Geostatistika
  - Kvantifikacija prostorne zavisnosti (Variogram, Moran I, spatial cluster analysis ...)
  - Kriging (simple and ordinary)
  - Uključivanje pomoćnjih promenljvih - Regression Kriging
  - Prostorno-vremenska geostatistika (gstat-krigeST) - modeliranje prostorno vremenske zavisnosti
  - 2D/2D+T predikcija



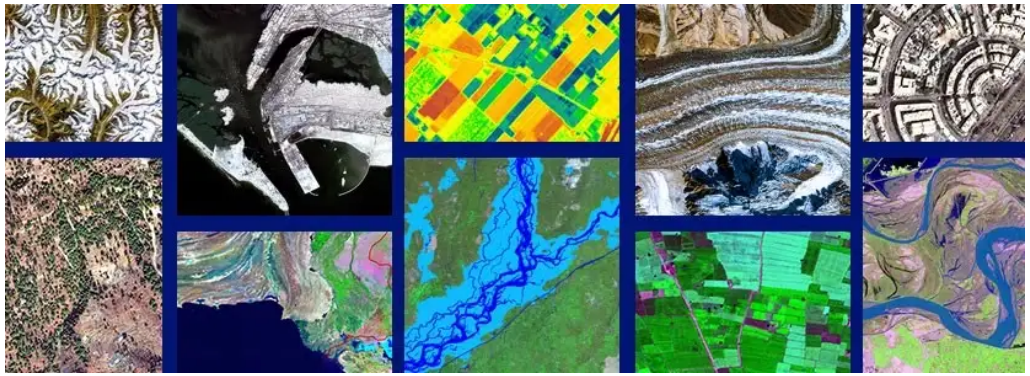
Retrospektiva: 2015-2020

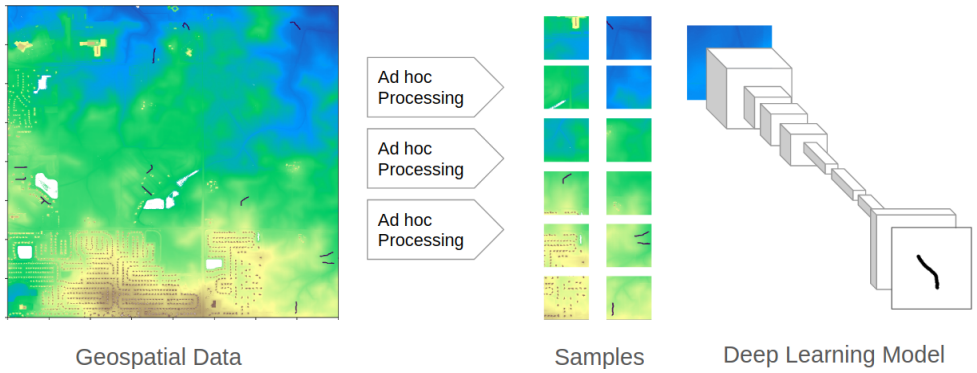
- Ogromna količina prostornih podataka postaje sve dostupnija!

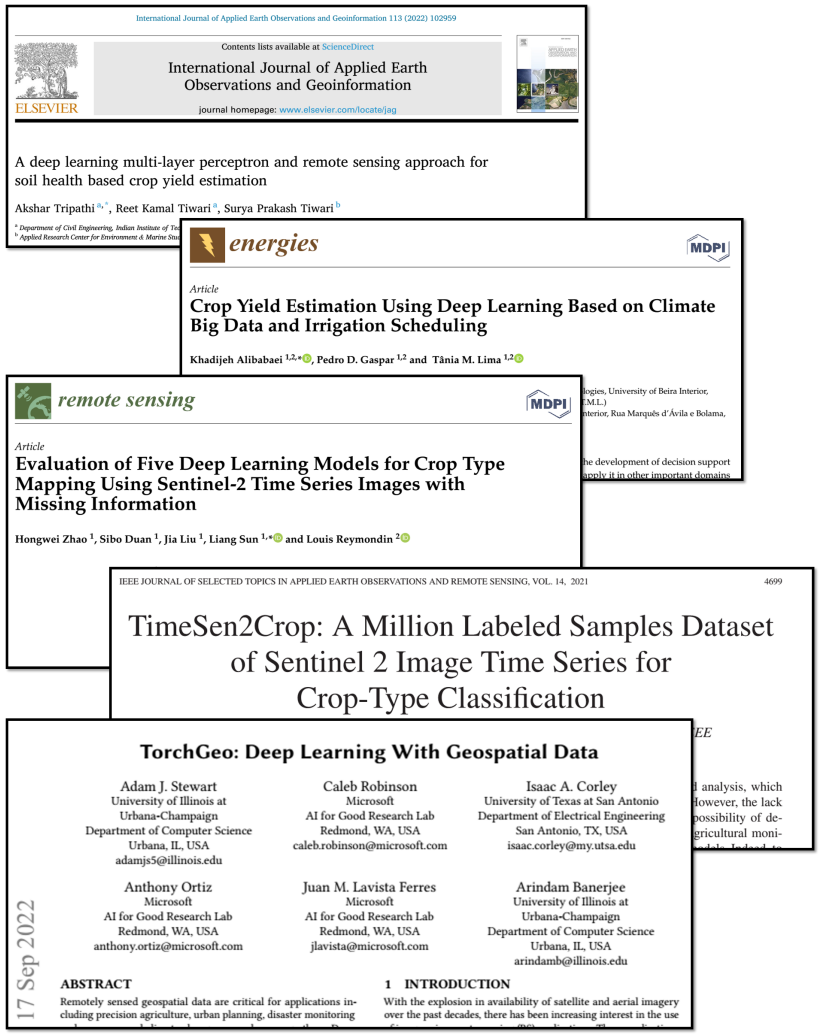
- Machine Learning with Spatial Data
  - Penalized linear models
  - Tree based models (Random Forest)
  - Ensemble ML models

Retrospektiva: od-2020

Deep Learning and Remote Sensing







Hvala na pažnji!