Зборник радова Факултета техничких наука, Нови Сад

УДК: 528
ДОИ: https://doi.org/10.24867/35KG03Mitrovic

Визуализација геопросторних и статистичких података коришћењем програмског језика *R*

Visualization of Geospatial and Statistical Data Using the R Programming Language

Михајло Митровић, *Факултет техничких наука, Нови Сад*

Студијски програм – ГЕОДЕЗИЈА И ГЕОИНФОРМАТИКА

Кратак садржај – *У оквиру мастер рада анализирана је употреба програмског језика у примени над геопросторним подацима у циљу визуализације. Анализом су обухваћени: програмски језик R, програмско окружење RStudio и геопросторни подаци који се односе на територију Републике Србије.*
Кључне речи: *визуализација, статистика, анализа, геопросторни подаци, R програмски језик*

Abstract – *The master thesis is about analysis of the programming language and its application to geospatial data for the purpose of visualization. The analzsis includes the R programming language, integrated development environment RStudio and geospatial data related to the territory of the Republic of Serbia.*

Keywords: *visualization, statistic, analysis, geospatial data, R programming language*

НАПОМЕНА: Овај рад проистекао је из мастер рада чији ментор је проф. др Душан Јовановић

1. УВОД

У савременом дигиталном добу подаци су извор од којих настају информације које се даље анализирају и на основу којих се доносе одлуке од локалних до глобалних размера. С обзиром на то да постоје различите класификације података, једној припадају и геопросторни подаци, који заузимају посебно место јер се на основу њих ствара и просторна димензија у анализи. На основу геопросторних података могу се анализирати кретања у простору, а ако геопросторним подацима додамо и временску димензију могуће је пратити одређене појаве како у простору, тако и током одређеног временског периода. Циљ рада је да се истражи и представи примена *R* програмског језика у визуализацији геопросторних података, односно података који имају геопросторни карактер, са фокусом на практичне аспекте обраде, анализе и приказа просторних информација путем различитих *R* библиотека.

Поред наведеног, рад има за циљ да покаже изазове са којима се корисници могу сусрети при раду са геопросторним подацима. У оквиру рада биће анализирани реални скупови података који укључују статистичке показатеље распоређене по територијалним јединицама, који се визуализују кроз разне типове картографских приказа. Овом приликом визуализација се односила на статички приказ, интерактивни приказ и приказ у виду анимације.

2. ТЕОРИЈСКЕ ОСНОВЕ

2.1. Визуализација

Визуализација представља процес трансформације података у визуални облик како би се олакшало њихово разумевање. Визуализација омогућава да се велики скупови података приказују на начин да буде лакши за препознавање образаца уколико они постоје. Визуализација просторних података, укључује информације о њиховој локацији, дистрибуцији и својствима (просторним и непросторним). Подаци који се обрађују за форму визуелне перцепције могу се односити на почетне и завршне резултате истраживачких анализа. Текстуални, статистички или просторни подаци могу се визуелизовати као једнодимензионални, дводимензионални, тродимензионални и вишедимензионални, укључујући анимацију, у односу на врсту и намену истраживања и коришћења. Визуално, класификовани подаци кроз табеле или представљени графиконима и дијаграмима не дају просторну дистрибуцију феномена који се проучава. То омогућава карта [1].

У смислу сврхе за које се визуализују подаци, визуализација може бити презентационог и истраживачког карактера.

Разлике између визуализација за презентацију и визуализација за истраживање леже и у форми и у пракси. Визуализација за потребе презентације је генерално статична и креира се да би се сумирале информације које треба представити. Са друге стране визуализација за потребе истраживања користи се за тражење резултата и треба да буде брза и информативна, а не спора и прецизна. Она није намењена за презентацију, тако да детаљне легенде и натписи нису потребни. Једна визуализација за потребе презентације биће намењена потенцијално хиљадама читалаца, док ће се креирати хиљаде визуализација за потребе истраживања података једног аналитичара [2].

2.2. Геопросторни подаци

Геопросторни подаци представљају темељ савремене картографије, географских информаци-оних система (ГИС) и бројних апликација које се ослањају на просторну анализу. Они описују објекте, појаве и процесе на Земљиној површини заједно са њиховим географским положајем. Другим речима, сваки геопросторни податак садржи две кључне компоненте: атрибутне информације (шта је објекат) и просторне информације (где се објекат налази).

Упркос богатству великих и разноврсних извора просторних података, просторни подаци могу бити и даље тешки за проналажење, тешки за коришћење или нелако доступни. Ове препреке ограничавају поновну употребу података и њихов потенцијални утицај. Ови изазови су препознати и довели су до развоја принципа за управљање подацима под називом „пронађиви, приступачни, интероперабилни и поново употребљиви“ (*FAIR*). Да би се максимизирала њихова вредност, подаци треба да буду *FAIR* (енгл. *Findable, Accessible, Interoperable and Reusable*) [3].

2.3. Координатни системи

Координатни системи су један од најважнијих чинилаца које треба узети у обзир при анализи геопросторних података. Просторни подаци су слични у односу на друге врсте података, али такође укључују нумеричке информације које омогућавају позиционирање на Земљи. Ови бројеви су део координатног система који обезбеђује референтни оквир за податке за лоцирање објеката на површини Земље, за поравнање података у односу на друге податке, за вршење просторно тачне анализе, за додавање и уређивање података и за креирање мапа.

Координатни референтни системи могу се идентификовати на различите начине. Тренутно најчешћи начини су: идентификатор *EPSG* (енгл. *European Petroleum Survey Group*) и потпунији *WKT2* (енгл. „*well-known text*“) дефинисан од стране *OGC* (енгл. *Open Geospatial Consortium*) и транспонован у *ISO* 19162 (енгл. *International Organization for Standardization 19162*) [4].

2.4. *R* програмски језик и *RStudio* интегрисано развојно окружење

R пружа свестран и моћан скуп алата за кориснике свих нивоа вештина, што га чини правим избором ако је потребно врхунско решење за анализе података. *R* има све што је потребно за унапређење анализе података, без обзира на ниво искуства [5].

RStudio је интегрисано развојно окружење (*IDE*) за *R* (слика 1) и *Python*. Садржи конзолу, едитор за истицање синтаксе који подржава директно извршавање кода и алате за цртање, историју, отклањање грешака и управљање радним простором. *RStudio* је доступан у издањима отвореног кода и комерцијалним издањима и ради на десктоп рачунарима (*Windows*, *Mac* и *Linux*) [6].



Слика 1: *R* и *RStudio*

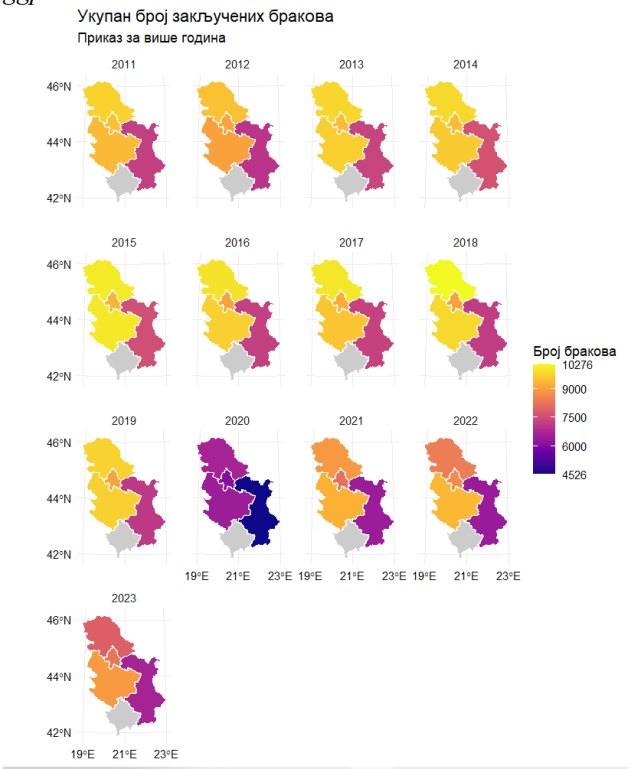
3. ПРАКТИЧНИ ПРИМЕР И ДИСКУСИЈА

Практични део се састојао из неколико делова:

- Претрага и преузимање *shp* фајлова који садрже границе административних јединица са портала Републичког геодетско завода;
- Претрага и преузимање извора података у *csv* формату са портала Републичког завода за статистику, који се односе на подручја из *shp* фајлова;
- Инсталација *R* програмског језика и *RStudio* програмског окружења;
- Писање и покретање кода у *RStudio* програмском окружењу.

3.1. Статички приказ

Слика испод (слика 2) је резултат визуализације укупног броја закључених бракова за сваку годину од 2011. године до 2023. године коришћењем пакета *ggplot2*.



Слика 2: Визуализација - статички приказ

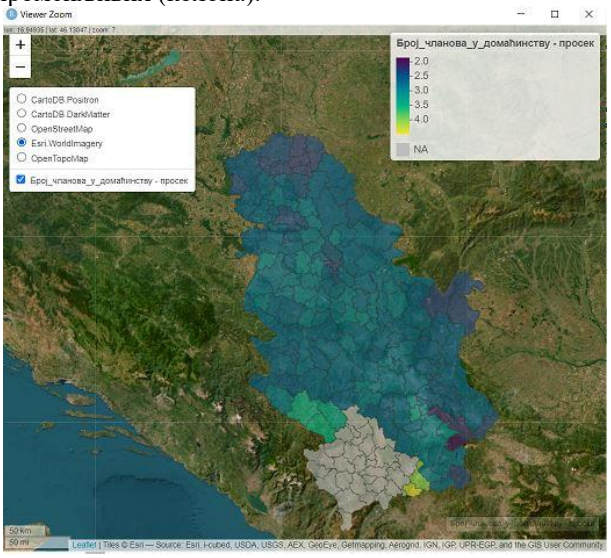
Изворни *csv* фајл садржи 1248 обсервација и 14 варијабли. У овом случају се ради о статичком приказу, што значи да нема опција које пружају анимиране или интерактивне мапе. Визуализација садржи назив, поднаслов, географску мрежу, легенду и ближе описује податке на основу године и територијалне јединице на коју се односи приказ. Садржај ове визуализације мора се описати кодом, јер се не подразумева, као у следећем подпоглављу.

Пошто су подаци приказани имајући у виду временску и просторну димензију они се могу анализирати с аспекта временске одреднице и просторне јединице. На основу ове визуализације лако се види да је више закључених бракова било у регионима Војводине, Београда и Западне Србије и Шумадије у односу на регион Јужне и Источне Србије, а с обзиром на то да

портал Републичког Завода за Статистику као коришћен извор података не садржи податке о броју бракова за регион Косова и Метохије, регион је приказан сивом бојом.
Оно што се још види јесте пораст закључених бракова од 2015. године до 2018. године, у односу на период од 2011. године до 2014. године, а од 2019. године до 2023. године је броја закључених бракова у опадању. Оно што се посебно примећује јесте знатно нижи број закључених бракова у 2020. години. „Зашто је број закључених бракова у 2020. години знатно нижи у односу на било коју годину?“ јесте питање које се поставља на основу овакве визуализације. Сада можемо кренути у потрагу за одговорима и тражити их у различитим друштвеним, политичким, економским, здравственим и другим сферама живота. С обзиром на то да је у 2020. години избила пандемија корона вирусом и да је за собом донела одређене мере које су се примењивале у циљу заштите становништва, у великој мери се може рећи да је пандемија узрок мањег броја закључених бракова, који након те године није достигао бројеве из година пре пандемије. Још једна ствар на коју указује ова визуализација, јесте мањи број склопљених бракова у региону Јужне и Источне Србије за сваку годину у односу на друге регионе, што може да укаже на мањи број становника у том региону. Коришћени пакети су: *sf, readr, dplyr, ggplot2.*

3.2. Интерактивни приказ

Визуализација на слици испод (слика 3) је резултат *mapview* пакета, који се користи за ефикасан и једноставан интерактивни приказ геопросторних података. У овом случају подаци у изворном *csv* фајлу су организовани у 5298 обсервација (редова) и 14 променљивих (колона).



Слика 3: Визуализација - интерактивни приказ

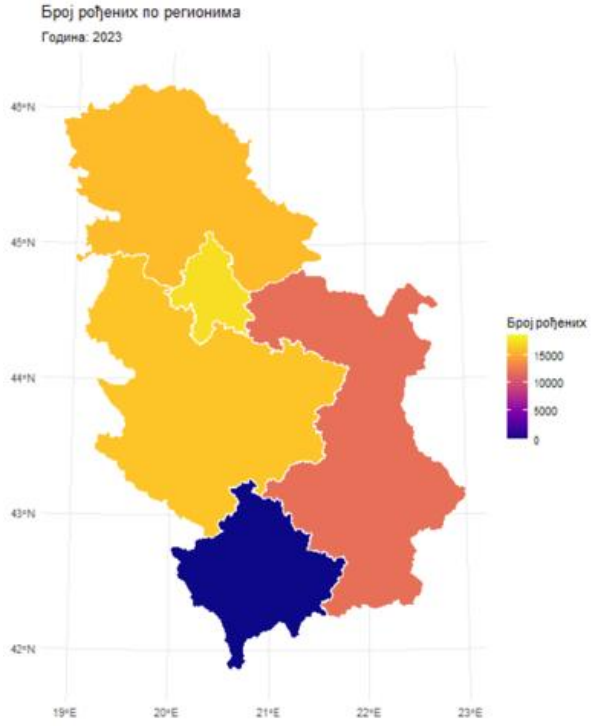
Предност визуализације помоћу овог пакета је што нема потребе за неким додатним подешавањима. Приказ овакве мапе не захтева пуно времена у смислу времена које је потребно да се мапа прикаже од тренутка покретања кода. На слици изнад се види палета у виду избора различитих слојева на којима се приказује мапа, што говори о једном делу интерактивности коју нуди. Поред наведеног постоје опције за увеличавање и умањење приказа, додавање легенде, размерника, мапу је могуће померати и увећати на жељено место, као и очитавати географске ширине и дужине преласком мишем на одређену локацију, аутоматски избор палете боја и интерактивност на начин да преласком мишем преко територију општине даје неку вредност, у овом случају је то просечан број чланова у домаћинству за 2022 годину. У табели испод (табела 1) дати су називи колона и вредности које се односе на те колоне за општину са називом Град Нови Сад. Табела (табела 1) је идентична табели која се добија у виду искачућег прозора при клику мишем на то подручје, што је још један облик интерактивности. Једном речју овакав начин приказа не тражи пуно, а даје много. Коришћени пакети су: *sf, readr, dplyr, mapview* и *stringi*.

Табела 1: Искачући прозор

Матични број општине	89010
Општина	Град Нови Сад
Година	2022
Индикатор	Просечан број чланова домаћинства
просек	2.336198186
fillColor	#1D4D85
mvFeaturedId	108

3.3. Приказ у виду анимације

На слици испод (слика 4) приказан је резултат једног дела визуализације у виду анимације, то јест приказ који се односи на укупан број рођених по регионима 2023. године.



Слика 4: Визуализација у виду анимације

Изворни *csv* фајл се садржи 312 обсервација и 14 варијабли. Анимација се састоји од 13 слика које се редом приказују тако што се приказ укупног број рођених претходне године замењује приказом укупног броја рођених наредне године и тако све у круг од 2011.

године до 2023. године, без могућности интеракције у смислу заустављања анимације у неком тренутку. Анимација је сачувана у *.GIF* формату и као таква је погодна за дељење путем апликација за размену порука. Садржај визуализације је пропраћен географском мрежом са назначеним вредностима географских координата (географска ширина и географска дужина), насловом (Број рођених по регионима), поднасловом (Година: 2011-2023), размерником у боји који остаје исти приликом промене приказа у анимацији и називом који ближе упућује на шта се размерник односи и на основу ког реалног параметра се мења приказ у анимацији (Број рођених). Коришћени пакети су: *sf, readr, dplyr, ggplot2, gganimate* и *viridis.*

4. ЗАКЉУЧАК

На основу приказаног, јасно се уочавају резултати у виду могућности да један или више великих скупова података учинимо „малим“ , односно да извршимо скалирање података, а да се при томе не изгуби њихова вредност и суштина онога што представљају. Очигледно је да је то могуће постићи трансформацијом сирових или сређених табеларних података у слике, у овом случају у мапе. Са тако креираних визуализација изведена су и шира запажања. На тај начин подацима није умањена вредност, већ су визуализацијом подаци добили на вредности.

Из претходно наведеног следи једно тумачење визуализације као трансформације података у мапе. Обрадом конкретних тема, односно тема из реалног окружења и доношењем закључака на основу обраде и упоређивања у смислу временске и просторне димензије у оквиру једне теме или идентификовање очигледних узрочно-последичних веза између две или више тема, представља још један пример како су геопросторни подаци добили на вредности применом *R* програмског језика у обради и визуализацији.

У неким случајевима код се извршава брже, а у неким случајевима спорије. Када су као извор коришћени геометријски слојеви директно са сервера, у случају интерактивних приказа, извршавање кода је било брже у односу на извршавање кода када су геопросторни подаци (подаци и гометрије) били складиштени на рачунару. Брзина извршавања кода зависила је и од тога о каквом типу визуализације је реч. Тако се у случају визуализације статичке мапе код извршавао брже него у случају визуализације у виду анимације, за коју је било потребно око 15 минута да се изврши након покретања кода. Дакле, најбрже се извршавао код у случају интерактивних мапа када је коришћена библиотека *mapview*, затим када је визуализација била статичког типа и најспорије се код извршавао када је визуализација била у виду анимације.

С аспекта могућих праваца даљег рада над подацима или примене у другим сферама живота, могу се разматрати други алати, друге библиотеке у оквиру *R* програмског језика, тродимензионалне визуализације, примена математичких теорија из области статистике и вероватноће, а касније и визуализација тако добијених резултата, а посебно примена вештачке интелигенције.

5. ЛИТЕРАТУРА

[1] J. M. Јовановић, T. M. Јанковић, M. Марковић, M. Стојановић и M. Симић, „Визуелизација просторних података од значаја за развој туризма у заштићеним подручјима,“ у *Дванаести научно - стручни скуп са међународним учешћем Планска и нормативна заштита простора и животне средине, Београд, октобар 2023.*, Београд, 2023.

[2] C.-h. Chen, W. Hardle и A. Unwin, Handbook of Data Visualization, Берлин, Хајделберг: Springer-Verlag, 2008.

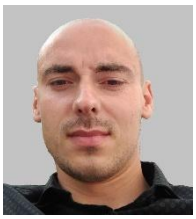
[3] P. Moraga и L. Baker, „rspatialdata: a collection of data sources and tutorials on downloading and visualising spatial data using R,“ *F1000 Research,* т. 11, 13 Новембар 2024.

[4] P. Maesen и E. Salingros, „Introduction to Reproducible Geospatial Analysis and Figures in R: A Tutorial Article,“ *Data,* т. 9, p. 58, 20 Април 2024.

[5] J. Roa, „Install R and Rstudio: why R rocks,“ 02 Март 2023, https://www.hertiecodingclub.com/learn/rstudio/Install-r-studio/ [Последњи приступ 13 Август 2025].

[6] „The most trusted IDE for open source data science,“ https://posit.co/products/open-source/rstudio/?sid=1 [Последњи приступ 13 Август 2025].

Кратка биографија:



Михајло Митровић рођен је у Нишу 1996. год. Мастер рад на Факултету техничких наука из области Геодезија и геоинформатика одбранио је 2026.год.

Контакт: mitrovic.mihajlo711@gmail.com