

Mapiranje zaraze koronavirusom u Hrvatskoj

Vesna Gotovac Đogaš, David Živaljić

Sažetak

Cilj ovog rada je bio otkrivanje županija u Hrvatskoj visokog rizika od zaraze i smrtnosti koronavirusom u prvoj polovini 2021. godine uz pomoć metoda prostorne statistike. U svrhu procjene rizika od zaraze koristimo standardizirani omjer incidencije. Za procjenu rizika smrtnost nam je poslužio standardizirani omjer smrtnosti i stopa smrtnosti.

Ključni pojmovi: koronavirus, statistički modeli, mapiranje, distribucije, procjenitelj

1. Uvod

Nit vodilja prostorne analize je takozvani Toblerov zakon koji glasi ovako: „Sve je povezano sa svime, ali bliske stvari su povezanije jedna s drugom.”

Prostorna statistika je samostalna grana statistike koja proučava metode prikupljanja i analize podataka koristeći njihova topološka (geografska) svojstva. Široko je primjenjiva i započela je razvoj iz različitih disciplina. Dovoljno je nabrojiti samo tri sljedeća problema u kojoj se primjenjuje kako bi dobili dojam široke rasprostranjenosti: analiza položaja dijelova čipova, analiza i predviđanje položaja galaksija u svemiru, epidemiologija.

U ovom radu razmatrat ćemo primjenu prostorne statistike u epidemiologiji. Preciznije govoreći, razmatrat ćemo mapiranje zaraze koronavirusom po hrvatskim županijama u prvoj polovini 2021. godine.

Općenito, cilj mapiranja zaraze je prikazati prostornu distribuciju rizika koji se može manifestirati na dva načina: mortalitet i morbiditet (broj ljudi koji su oboljeli). Nakon prikazane distribucije mogu se uočiti regije koje imaju povećani rizik u odnosu na ostale. Sukladno tome možemo poduzeti određene mjere zaštite ili pak pokušati otkriti razlog zašto dolazi do povećanog rizika.

Cilj ovog rada nije pružiti detaljan i opsežan opis svih metoda koje se trenutno koriste u prostornoj epidemiologiji, već pokazati one koje se koriste u širokom obujmu.

Glavni zadatak ovog rada je primjena svih iznesenih metoda uporabom računala, programskog jezika **R** i prostornih podataka korištenjem takozvanog **GIS-a** [2], računalni sustav koji analizira i prikazuje geografski referencirane informacije.

U prvom poglavlju kratko se navode osnove mapiranja zaraze uz prikaz učitanih podataka te dobnu raspodjelu za Hrvatsku po županijama. Pokazuje se podjela podataka na slojeve čiji cilj može biti uklanjanje zbunjujućih varijabli.

U drugom poglavlju se iznose razne metode procjene rizika od morbiditeta i mortaliteta te komentiraju njihove prednosti i nedostatke.

2. Osnove pri mapiranju zaraze

Cilj mapiranja zaraze je pružiti prikaz prostorne distribucije bolesti gdje pretpostavljamo da su podaci podjeljeni na disjunktne regije, u našem slučaju županije. Bolest se može reflektirati na dva načina koje nazivamo **morbiditet** i **mortalitet**. Morbiditet označava broj oboljenja dok mortalitet označava broj umrlih od određene bolesti.

U ovom radu promatramo obje pojave u slučaju bolesti izazvane koronavirusom (COVID-19) te pokušavamo zaključiti postoji li županija koja ima povećan rizik bilo mortaliteta ili morbiditeta. Kako takvi zaključci mogu biti varljivi, dobro je dane podatke razmatrati podijeljene u više slojeva kako bi izbjegli efekt zbunjujućih faktora¹ (engl. Confounding factor)

Slojevi mogu biti: spol, godine (gdje možemo na više načina definirati dobne skupine), županije u kojima ljudi žive, razni obrasci ponašanja ili karakteristike osoba za koje smatramo da mogu utjecati na morbiditet ili mortalitet (npr. podjela na pušače i nepušače ili osobe koje su pretile i one koje to nisu)...

¹ Čimbenik koji zbunjuje, može prikriti stvarnu povezanost ili lažno pokazati prividnu povezanost između varijabli u istraživanju ako ne postoji stvarna povezanost između njih.

Krećemo od oznaka P_{ij} , O_{ij} i U_{ij} koji označavaju populaciju, broj promatranih slučajeva zaraze, te broj umrlih u regiji i i sloju j . Sumiranjem po svim slojevima j možemo dobiti ukupnu populaciju P_i , slučajeve O_i i U_i broj umrlih u nekoj regiji i . Sada na isti način sumiranjem svih P_i dobijemo P_+ , analognogno dobijemo O_+ i U_+ .

Dakle, P_+ je ukupan broj populacije koju promatramo, O_+ je ukupan broj opaženih slučajeva zaraze, a U_+ ukupan broj opaženih smrtnih slučajeva. Kako nam brojevi opaženih slučajeva ne daje puno informacija o samoj distribuciji zaraze potrebno ih je usporediti sa prosječnim brojem slučajeva kojeg ćemo računati na sljedeći način: $E_i = P_i r_+$ gdje je $r_+ = \frac{O_+}{P_+}$. Slično u slučaju smrtnosti definiramo $E'_i = P_i q_+$, gdje je $q_+ = \frac{U_+}{P_+}$. Ovdje se radi o najjednostavnijem slučaju standardizacije kojeg nazivamo *indirektna standardizacija*.

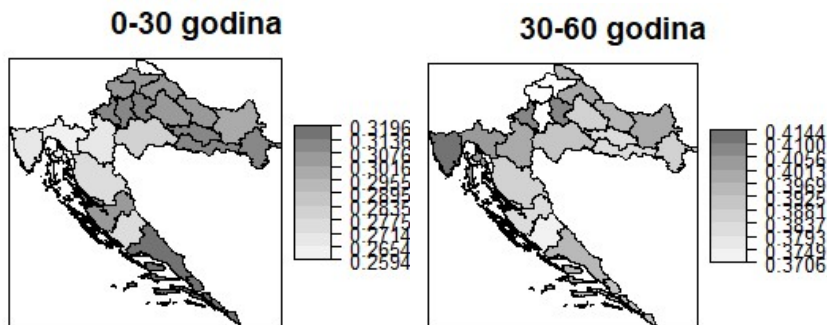
Kada su nam podaci grupirani u slojeve umjesto računanja globalnog omjera $\frac{O_+}{P_+}$, odnosno $\frac{U_+}{P_+}$ kojeg koristimo za sve regije, za svaki pojedinačni sloj možemo izračunati $r_j = \frac{\sum_i O_{ij}}{\sum_i P_{ij}}$ te $q_j = \frac{\sum_i U_{ij}}{\sum_i P_{ij}}$. Ova standardizacije se naziva *interna standardizacija*.

Pogledajmo sada podatke o zarazi koronavirusom Hrvatskoj i njihovu raspodjelu po županijama. Podaci su o slučajevima za prvu polovicu 2021. godine.

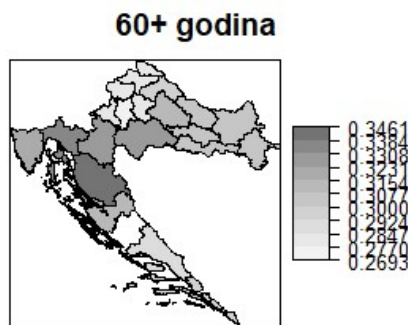


Slika 1. Broj zaraženih osoba COVID-19 virusom u period od 1.1.2021 do 1.7.2021. po svakoj županiji u Republici Hrvatskoj.

Na slici 1 vidimo broj zaraženih po svakoj županiji, ali očigledno je da gledajući samo te podatke ne možemo ništa zaključiti o povećanom riziku zaraze, odnosno smrtnosti u nekoj županiji. U tu svhu smo napravili prikaz na slikama 2 i 3.



(a) Udio zaraženih u dobnoj skupini od 0 do 30 godina. (b) Udio zaraženih u dobnoj skupini od 30 do 60 godina.

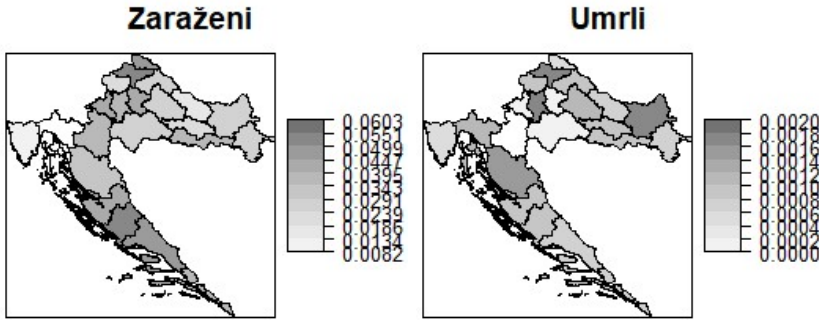


(c) Udio zaraženih u dobnoj skupini preko 60 godina.

Slika 2.

Iz ovih prikaza stvari su nešto jasnije nego kada imamo same brojeve slučajeva po županijama.

U sljedećem prikazu ćemo tamnijim bojama prikazivati županije sa većim brojem slučajeva.



(a) Ukupan broj zaraženih.

(b) Ukupan broj umrlih.

Slika 3.

3. Statistički modeli

Uobičajena statistička pretpostavka za modeliranje promatranih slučajeva zaraze ili smrtnosti u regiji i i sloju j je ta da je broj opaženih slučajeva izveden iz Poissonove distribucije s parametrom $\theta_i E_{ij}$, gdje θ_i označava relativan rizik incidencije u regiji i , a $E_{ij} = P_{ij} r_j$ za broj zaraženih, odnosno s parametrom $\Theta_i E'_{ij}$, gdje je Θ_i relativan rizik smrtnosti u regiji i , a $E'_{ij} = P_{ij} q_j$ za broj umrlih.

Relativni rizici θ_i i Θ_i će biti presudni parametri kod zaključivanja o povećanom riziku od incidencije, odnosno smrtnosti u nekoj od županija.

Kada je relativni rizik 1 to znači da je rizik prosječan, a od interesa će nam biti rizici koji su značajnije veći od 1, što ukazuje povećan rizik.

Napomenimo da ćemo pretpostavljati da nema interakcije između rizika i populacijskog sloja, tj. θ_i i Θ_i ovise samo o regiji i .

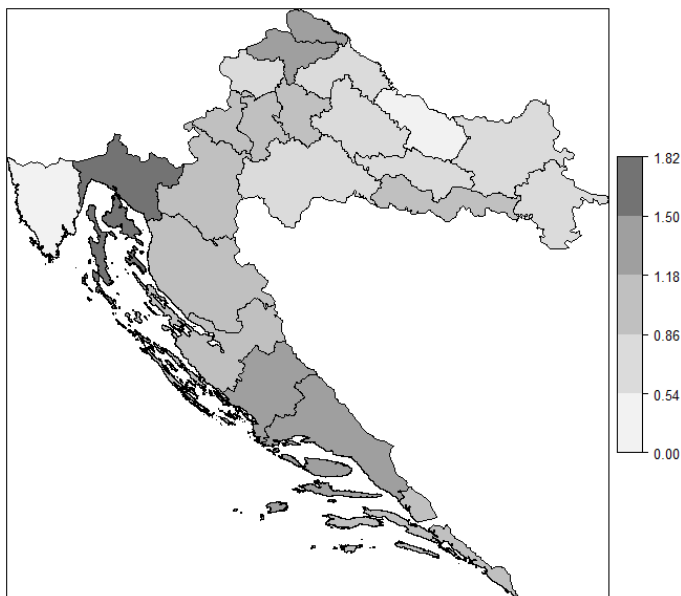
U nastavku rada se koncentriramo na načine kako možemo procijeniti relativne rizike θ_i i Θ_i iz dostupnih podataka o broju zaraženih i umrlih.

3.1. Standardizirani omjer incidencije (engl. Standardized Incidence Ratio SIR)

Jedan način procjene relativnog rizika θ_i incidencije kojeg ćemo koristiti u ovom radu definira se na sljedeći način:

$$SIR_i = \frac{O_i}{E_i}$$

kojeg nazivamo *Standardizirani omjer incidencije* (engl. *Standardized Incidence Ratio*) kojeg ćemo kroz tekst kraće označavati sa *SIR*.



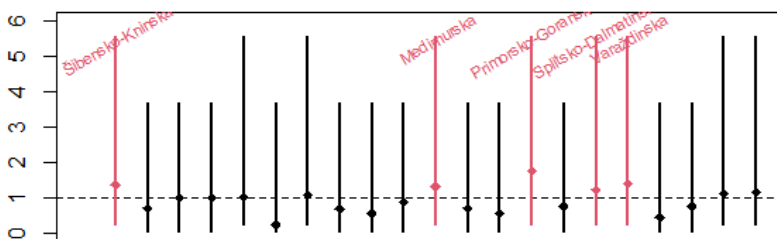
Slika 4. Grafički prikaz SIRa u prvoj polovini 2021. godine po županijama u Republici Hrvatskoj

Uz pomoć SIR -a i njegovog grafičkog prikaza na slici 4 dobivamo dosta više informacija i bolji dojam o varijabilnosti rizika nego na slici 1 gdje koristimo same brojeve slučajeva. Vidljivo je da prikazom SIR -a dobivamo dosta benefita, ali on ima neke nedostatke.

SIR u nekim slučajevima i nije baš najbolji prikaz distribucije bolesti ili pak nije dobar za određivanje područja većeg relativnog rizika. Naime, u područjima gdje je populacija dosta manja nazivnik u izrazu SIR_i je mali pa mala varijacija u O_i može dramatično promjeniti cijeli razlomak što nas može dovesti do krivog zaključka. Crvenom bojom na slici 5 ćemo označiti samo one županije koje imaju veći relativan rizik od 1.2. Sada koristeći činjenicu da je O_i iz Poissonove distribucije možemo izračunati interval pouzdanosti za svaki SIR_i . U tu svrhu u R-u koristimo paket **epitools** i funkciju `pois.exact`.

Povećan SIR_i , u našem slučaju veći od 1, sugerira da je zabilježen broj slučajeva veći nego što bismo prethodno očekivali.

Najveću vrijednost za relativni rizik ima Primorsko-goranska županija i to 1.769, slijede Varaždinska i Međimurska s vrijednostima 1.37 odnosno 1.31. Uočimo, kako je u našem slučaju $SIR_i = \frac{O_i}{E_i} = \frac{O_i}{P_i} \frac{\sum_i P_i}{\sum_i O_i}$ to povlači da u Primorsko-goranskoj županiji ima oko 77% više zaraženih nego u



Slika 5. 95% intervali pouzdranosti za SIR

općoj populaciji jer za taj i vrijedi $\frac{O_i}{P_i} = 1.769 \frac{\sum_i O_i}{\sum_i P_i}$.

Na slici 5 su prikazani intervali 95%-tne pouzdranosti za SIR pomoću funkcije **pois.exact** iz paketa **epitools** u R-u. Vidljivo je da su ti intervali dosta „široki” pa ne možemo sa sigurnošću zaključiti nešto o regijama povećanog rizika.

Napomena 1. Način procjene relativnog rizika Θ_i smrtnosti analogan SIR u definira se na sljedeći način:

$$SMR_i = \frac{U_i}{E'_i}$$

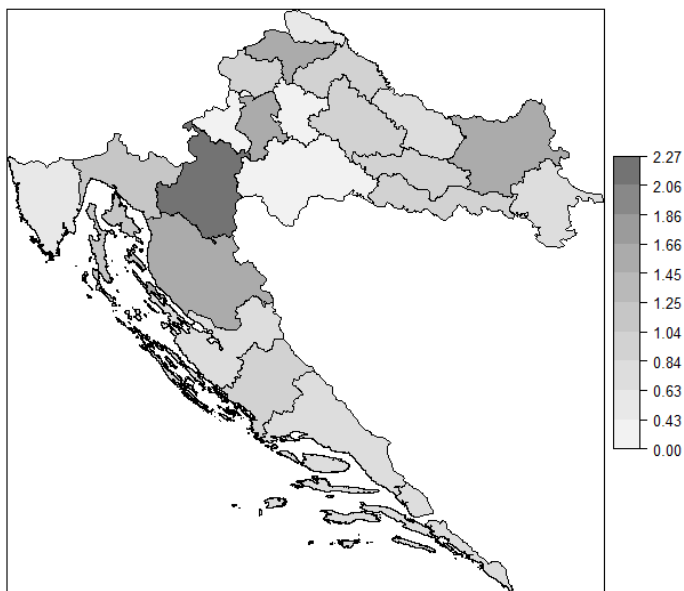
te se naziva Standardizirani omjer smrtnosri (engl. *Standardized Mortality Ratio*). No, treba imati na umu da ovaj procjenitelj posjeduje iste nedostatak kao i SIR .

Na slici 6 prikazane su procjene relativnog rizika od smrti po županijama u Republici Hrvatskoj korištenjem SMR a.

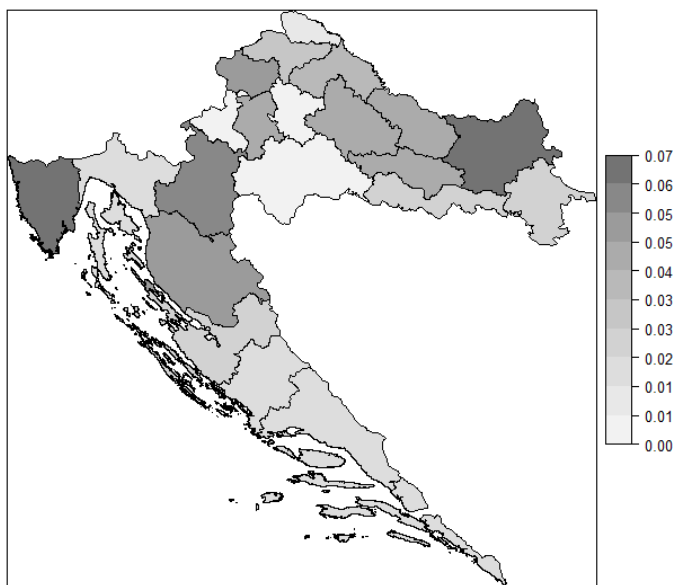
3.2. Stopa smrtnosti (engl. Case Fatality Rate CFR)

Stopa smrtnosti (u daljnjem tekstu ćemo koristiti kraticu CFR) mjeri težinu određene bolesti te predstavlja omjer ukupnog broja smrtnih slučajeva i broja zaraženih u određenom vremenskom intervalu. Možemo ga definirati na sljedeći način: $CFR_i = \frac{U_i}{O_i}$, gdje je U_i broj umrlih u regiji i , a O_i broj potvrđenih slučajeva zarazom u regiji i . Nakon dijela programa u R-u na slici 7 prikazujemo grafički prikaz od stope smrtnosti od koronavirusa u prvoj polovini 2021. godine. Najveće vrijednosti CFR -a imaju županije: Istarska, Osječko-baranjska, Karlovačka i Ličko-senjska.

Ovakav prikaz je informativan te može ukazivati na razne probleme kao što je lošiji zdravstveni tretman od prosjeka u državi. No razlozi



Slika 6. SMR po županijama u Republici Hrvatskoj

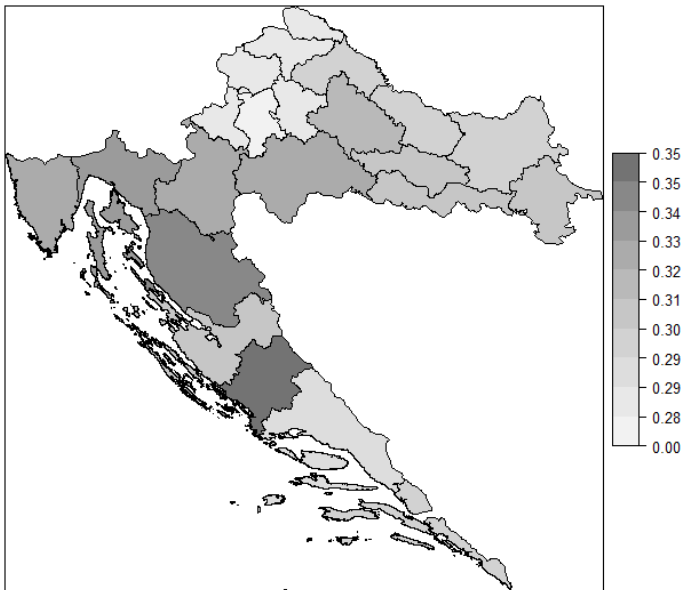
Slika 7. *CFR* bolesti izazvane koronavirusom za prvu polovinu 2021. godine po županijama u Republici Hrvatskoj.

mogu biti svakakvi, primjerice možda je baš u ovim županijama povećan broj ljudi starijih od 60 godina za koje je poznato da se teže odupiru koronavirusom. Ako ove rezultate usporedimo sa udjelom starijih od 60 godina iz slike 2 možemo vidjeti da je upravo u tim županijama (osim Osječko-baranjske) povećaj udio starijih osoba.

Nedostaci ovog omjera slični su kao oni u *SIR*-u. Ako iz slike 3 uočimo da je broj zaraženih u Istarskoj županiji relativno mali, tada moramo biti oprezni u donošenju zaključaka za Istarsku županiju jer mala promjena u brojniku može znatno utjecati na konačni CFR_i .

Globalni CFR za cijelu Hrvatsku u prvoj polovini 2021. godine je 2.8% dok je na svjetskoj razini 2.1%.

Učitali smo podatke za prvu polovinu 2022. godine te napravili sličan postupak. Dobivamo sljedeće rezultate:



Slika 8. CFR bolesti izazvane koronavirusom za prvu polovinu 2022. godine po županijama u Republici Hrvatskoj.

Uspredimo li ove rezultate sa udjelom populacije starije od 60 godina dobijamo još bolje poklapanje. Šibensko-kninska i Ličko-senjska sada imaju najveći CFR_i , a ujedno su to županije s najvećim udjelom starijih od 60.

Globalni CFR za Hrvatsku u prvoj polovici 2022. godine iznosi 0.082%. Mnogo faktora je moglo utjecati na ovoliku razliku u stopi smrtnosti. Mnoge pristranosti su se mogle dogoditi u odnosu na 2021. Na primjer, veći broj testiranja (i točniji testovi) gdje otkrivamo sve vrste zaraze

za razliku od prije. No, ovo je možda samo indikacija da postizemo kolektivni imunitet. Još jedan razlog za ovo može biti procijepljnost stanovništva koja je dosta veća u 2022. nego u 2021. godini. No, ovdje ne možemo donijeti siguran zaključak jer je moguće da je bolest lakša kod novih sojeva virusa.

4. Zaključak

Izvorni cilj ovog rada je bio upoznavanje osnovnih metoda prostorne statistike te otkrivanje županija u Hrvatskoj visokog rizika od zaraze koronavirusom. Koristili smo dvije različite metode i iznijeli sve prikaze koje smo stvarali uz pomoć programskog jezika R.

Vodeći se tim prikazima zaključili smo sljedeće: što se tiče morbiditeta, najveći rizik ima Primorsko-goranska županija, a slijede je Međimurska i Varaždinska. Što se tiče mortaliteta, najveći rizik ima Karlovačka županija, a slijede Varaždinska, Grad Zagreb, Ličko-senjeska i Osječko-baranjska.

Cijeli kod koji je korišten pri izradi ovog rada zajedno sa svim potrebnim podacima za učitati te slikama može se pronaći na github profilu na linku [8].

Također tu možete pronaći i kod u Pythonu koji je korišten kako bi se lakše preradili strojno čitljivi podaci sa stranice koronavirus.hr.

Literatura

- [1] R. Bivand, E. Pebesma, V. Gómez-Rubio *Applied Spatial Data Analysis with R*, Drugo izdanje, (2013) New York, Springer
- [2] Geographic information system. Dostupno na https://en.wikipedia.org/wiki/Geographic_information_system[20. kolovoz 2022.]
- [3] Programski jezik R. Dostupno na <https://www.r-project.org/>
- [4] O. Shabir, *What is Case Fatality Rate (CFR)?* [https://www.news-medical.net/health/What-is-Case-Fatality-Rate-\(CFR\).aspx](https://www.news-medical.net/health/What-is-Case-Fatality-Rate-(CFR).aspx)[2. rujna 2022.]
- [5] Shapefile Hrvatske: <https://www.diva-gis.org/datadown>
- [6] K. Weinberger, *Estimating probabilities from data*. Dostupno na <https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote04.html> [28. kolovoza 2022.]

- [7] D. Živaljić, *Mapiranje zaraze koronavirusom u Hrvatskoj*, Diplomski rad, (2013) Split: Sveučilište u Splitu, Prirodoslovno-matematički fakultet.
- [8] D. Živaljić, github profil, https://github.com/dzivaljic/spatial_statistic[5. rujna 2022.]

Vesna Gotovac Đogaš
Sveučilište u Splitu, Prirodoslovno matematički fakultet, Ruđera Boškovića
33, 21000 Split
E-mail adresa: vgotovac@pmfst.hr

David Živaljić
Positive Equity d.o.o.
E-mail adresa: david.zivaljic@gmail.com