

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Seminar

**Grafički prikaz razlike u dnevnom broju
Covid-19 slučajeva u R programskom
okruženju**

Rijeka, siječanj 2023.

Ivor Cvetkovski,
David Katalinić
Entoni Korlević,

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski studij računarstva

Seminar

**Grafički prikaz razlike u dnevnom broju
Covid-19 slučajeva u R programskom
okruženju**

Mentor: dr.sc. Renato Filjar

Rijeka, siječanj 2023.

Ivor Cvetkovski,
David Katalinić
Entoni Korlević,

Sadržaj

Popis slika	iv
1 Uvod	1
2 Metodologija rada	2
2.1 Korišteni podaci	2
3 Razvijena programska podrška u R-u	3
3.1 Preuzimanje dnevnog broja Covid-19 novozaraženih osoba	4
3.2 Kreiranje korisničkog sučelja uz pomoć Shiny paketa	5
3.3 Poslužiteljska strana programskog rješenja	6
4 Rezultati	9
5 Zaključak	12
Bibliografija	13

Popis slika

4.1	Slika zemlje i prikaz država pri inicijalizaciji programa	10
4.2	Prikaz novozaraženih osoba Covid-19 virusom na datum 21.11.2020.	11

Poglavlje 1

Uvod

U današnjem svijetu svakodnevno se na vijestima te internetskim portalima mogu vidjeti informacije o broju Covid-19 novozaraženih osoba. Covid-19 je virusna bolest uzrokovana novim koronavirusom SARS-CoV-2 zbog koje je 11.3.2020 proglašena epidemija. Cilj ovog projekta je razvoj programske podrške u programskom okruženju za statističko računarstvo R koja će:

- preuzeti dnevni broj Covid-19 novozaraženih osoba s Our World in Data GitHub repozitorija
- izdvojiti opažanja za izabrani dan u razdoblju od 2020. - 2022. godine putem grafičkog korisničkog sučelja
- za svaku državu odrediti razliku u dnevnom broju Covid-19 novozaraženih osoba u odnosu na prethodni dan
- grafički prikazati digitalnu kartu svijeta s konturama granica pojedinih država te s odgovarajućim bojama (crveno - porast, plavo - pad)

Prije govora o samom programskom rješenju potrebno je definirati metodologiju rada koja je obrađena u sljedećem poglavlju. Nakon toga u trećem poglavlju dan je opis programskog rješenja u programskom okruženju za statističko računarstvo R. U sljedećem poglavlju predstavljeni su te interpretirani rezultati dobiveni izvođenjem programskog rješenja. Na samom kraju u zaključku kratko je naznačen problem te je predstavljen sažetak vlastitih rezultata.

Poglavlje 2

Metodologija rada

Postupak za provedbu predloženog rješenja je jednostavan. Za provedbu rješenja potrebno je preuzeti 4.2.1 ili noviju verziju R-a s poveznice Cloud R Project te RStudio 2022.07.0 Build 548 ili noviji s poveznice R Studio Desktop. Nakon toga potrebno je preuzeti programski kod. Naposljetku za pokrenuti programski kod potrebno je samo pritisnuti *Run App* gumb u R studiju nakon čega će se pojaviti skočni prozor s aplikacijom. Koristeći dobiveno grafičko korisničko sučelje potrebno je unijeti datume za koje je potrebno izračunati razliku u broju Covid-19 slučajeva.

2.1 Korišteni podaci

Kao izvor informacija o dnevnom broju novozaraženih Covid-19 osoba korištena je csv datoteka preuzeta sa GitHub repozitorija Our World in Data. Our World in Data je internetska arhiva izvještaja o stanju Covid-19 pandemije u državama svijeta. Konkretno korišten je stupac `new_cases` koji sadrži nove potvrđene Covid-19 slučajeve. Stupac može sadržavati i vjerojatne slučajeve, ako su prijavljeni. Za prikaz kontura granica država korišteni su podaci dobiveni funkcijom `map_data("world")`.

Poglavlje 3

Razvijena programska podrška u R-u

U razvoju programske podrške u programskom okruženju za statističko računarstvo R korištene su četiri knjižnice: shiny, shinythemes, dplyr i ggplot2. Shiny je R paket koji omogućuje jednostavan razvoj interaktivnih web aplikacija. Aplikacije mogu biti samostalne web stranice ili mogu biti unutar R markdown dokumenta. Aplikacije kreirane pomoću Shiny paketa mogu izvoditi bilo koji R izračun i analizu jer izvršavaju R kod na poslužiteljskoj strani. U sklopu programskog rješenja Shiny paket se koristio za razvoj grafičkog korisničkog sučelja putem kojeg korisnik unosi potrebne datume. Shinythemes paket omogućuje dodatno uređivanje Shiny aplikacije pomoću predefiniranih tema. Dplyr je paket koji omogućuje manipulaciju podacima odnosno omogućuje rad s podatkovnim okvirima unutar i van memorije. Ggplot2 je sistem za deklarativno kreiranje grafike. Omogućuje kreiranje velikog broja različitih grafova te grafičkih prikaza digitalnih karti. U programskom rješenju Ggplot2 korišten je za grafički prikaz digitalne karte svijeta s konturama granica pojedinih država te s odgovarajućom bojom ovisno da li je riječ o porast ili padu broja novozaraženih osoba.

3.1 Preuzimanje dnevnog broja Covid-19 novozaraženih osoba

Prvi korak u razvoju programskog rješenja je preuzimanje potrebnih podataka za rad. Na kodu prikazanom ispod (Ispis 3.1) u varijablu *url* sprema se poveznica s koje će se preuzeti potrebni podaci. U ovom slučaju podaci o broju Covid-19 novozaraženih preuzimaju se s GitHub repozitorija Our World in Data. U varijablu *destfile* navodi se lokacija za spremanje preuzete csv datoteke. Uz pomoć funkcije *download.file()* preuzimaju se podaci na temelju vrijednosti zapisanih u *url* i *destfile* varijablama. U varijablu *data* u obliku podatkovnog okvira sprema se preuzeta csv datoteka uz pomoć funkcije *read.csv()*.

```
1
2 url <- 'https://covid.ourworldindata.org/data/owid-covid-data.
   csv'
3 destfile <- 'C:\\Users\\User\\Documents\\Fakultet\\Usluge
   zasnovane na lokaciji\\owid-covid-data.csv'
4 download.file(url,destfile)
5 data <-
6   as.data.frame(
7     read.csv(
8       'C:\\Users\\User\\Documents\\Fakultet\\Usluge zasnovane
       na lokaciji\\owid-covid-data.csv',
9       header = TRUE,
10      sep = ',',
11    )
12  )
```

Ispis 3.1 Preuzimanje dnevnog broja Covid-19 novozaraženih osoba

3.2 Kreiranje korisničkog sučelja uz pomoć Shiny paketa

Za kreiranje korisničkog sučelja poziva se funkcija *fluidPage()* uz pomoć koje se definira raspored panela u sučelju. Uz pomoć paketa *shinythemes* odabire se tema *lumen* za korisničko sučelje. U *sidebarLayout()* funkciji širina panela bočne trake (eng. *Sidebar panel*) postavlja se na 3. Širina panela označava koliko stupaca od mogućih 12 u rešetki zauzima jedna panela. Nadalje u bočnoj traci definira se *dateRangeInput()* uz pomoć kojeg korisnik unosi datume. Korisnik unosi dva datuma u razdoblju od 1.1.2020. do 31.12.2022. Na samom kraju definira se *mainPanel()* unutar kojeg će se prikazati dobiveni grafički prikaz karte svijeta. Za *mainPanel()* širina se postavlja na 9.

```
1 ui <- fluidPage(  
2   theme = shinytheme("lumen"),  
3   titlePanel("Covid-19 data"),  
4   sidebarLayout(  
5     sidebarPanel(width = 3,  
6       # Select date range to be plotted  
7       dateRangeInput(  
8         "date",  
9         strong("Date range"),  
10        start = "2020-01-01",  
11        end = "2022-12-31",  
12        min = "2020-01-01",  
13        max = "2022-12-31"  
14      ),  
15    ),  
16    # Output: Description, lineplot, and reference  
17    mainPanel( width = 9,  
18      plotOutput(outputId = "lineplot", width = "90%"),  
19    ) ) )
```

Ispis 3.2 Definiranje korisničkog sučelja

3.3 Poslužiteljska strana programskog rješenja

Na poslužiteljskoj strani (Ispis 3.3) u *output lineplot* varijabli zove se funkcija *renderPlot()* uz pomoć koje će se grafički prikazati karta svijeta. Na početku prvo se provjerava da li nedostaje korisnikov unos za početni ili krajnji datum. Nakon toga vrši se provjera da li je početni datum manji od krajnjeg datuma. Provjera se vrši pomoću funkcija *validate()* i *need()*. Ako jedno od ova dva uvjeta nije ispunjen ispisat će se poruka s opisom greške. Zatim se filtriraju podatci prema datumu, stvarajući novi podatkovne okvire *chosenDay* i *dayBefore*. Određene države imaju različite nazive u podacima preuzetim s Our World in Data repozitorija te u podacima dobivenim s funkcijom *map_data("world")* (Ispis 3.4). Potrebno je uskladiti imena tih država kako bi se osigurao ispravan grafički prikaz digitalne karte. Za određene države ne postoje dostupna opažanja o broju novih Covid-19 slučajeva te su te države navedene u komentaru programskog koda (Ispis 3.3).

```
1 server <- function(input, output) {  
2  
3   output$lineplot <- renderPlot({  
4     validate(need(!is.na(input$date[1]) & !is.na(input$date[2]))  
5       , "Error: Please provide both a start and an end date.")  
6     validate(need(input$date[1] < input$date[2], "Error: Start  
7       date should be earlier than end date.")  
8  
9     chosenDay <- filter(data, date == input$date[2])  
10    dayBefore <- filter(data, date == input$date[1])  
11  
12    chosenDay$location[chosenDay$location == "United States"] =  
13      "USA"  
14    chosenDay$location[chosenDay$location == "United Kingdom"]  
15      = "UK"  
16    chosenDay$location[chosenDay$location == "Cote d'Ivoire"] =  
17      "Ivory Coast"  
18    chosenDay$location[chosenDay$location == "Democratic  
19      Republic of Congo"] = "Democratic Republic of the Congo"  
20    chosenDay$location[chosenDay$location == "Congo"] = "  
21      Republic of Congo"
```

Poglavlje 3. Razvijena programska podrška u R-u

```
14
15     # Zemlje koje nisu pokrivene COVID datasetom
16     # Turkmenistan, Luxembourg, Yemen, Zambia, Zimbabwe, French
    Guiana, Antarctica
17
18     dayBefore$location[dayBefore$location == "United States"] =
    "USA"
19     dayBefore$location[dayBefore$location == "United Kingdom"]
    = "UK"
20     dayBefore$location[dayBefore$location == "Cote d'Ivoire"] =
    "Ivory Coast"
21     dayBefore$location[dayBefore$location == "Democratic
    Republic of Congo"] = "Democratic Republic of the Congo"
22     dayBefore$location[dayBefore$location == "Congo"] = "
    Republic of Congo"
```

Ispis 3.3 Preuzimanje i validacija podataka

Za grafički prikaz digitalne karte svijeta potrebno je prvo izračunati razliku u dnevnom broju Covid-19 novozaraženih (Ispis 3.4). Razlika se izražava pomoću postotaka. Od svakog postotka oduzme se 100 jer se gleda razlika u odnosu na prethodni pa ako je broj novozaraženih jednak u oba dana razlika iznosi 0%. Pomoću funkcije `map_data("world")` u varijablu `world` spremaju se podatci potrebni za prikaz grafičke karte svijeta s konturama granica pojedinih država. U sljedećem koraku podatkovnom okviru `world` dodaje se nova varijabla koja se zove `variable` te se ispuni podacima varijable `casesDifference`. Funkcija `match()` koristi se za povezivanje varijable `region` u `world` podatkovnom okviru s varijablom `location` u okviru podataka `chosenDay`. Rezultat korištenja `match()` funkcije je vektor indeksa koji odgovara zemljama u okviru podataka `chosenDay`. Na ovaj način podatci o postotnoj razlici novozaraženih se povezuju s odgovarajućim državama u podatkovnom okviru `world`. U sljedećem koraku stvara se grafički prikaz digitalne karte. Prvo se kreira `ggplot` objekt pod imenom `world_map` iscrtavanjem podataka u podatkovnom okviru `world`. Nakon toga u `geom_polygon()` sloju uz pomoć argumenta `aes()` određuje se koja će se varijabla koristiti za bojanje država. Nadalje u istom sloju postavlja se boja granica između država u bijelu te na širinu 0.1. Na samom kraju na objekt `world_map` do-

Poglavlje 3. Razvijena programska podrška u R-u

daje se `scale_fill_gradientn()` funkcija. U `scale_fill_gradientn()` funkciji u varijabli `limits` definira se raspon vrijednosti za preslikavanje ljestvice boja. Varijabla `labels` za svaki definirani prekid (eng. breaks) određuje jednu oznaku. Oznake i prekidi definirani su u varijabli `b`. Na samom kraju određen je vektor od četiri boje te je u argumentu `values` određen raspon za svaku boju.

```
1      casesDifference = ((chosenDay$new_cases /
2      dayBefore$new_cases) * 100) - 100
3
4      world <- map_data("world")
5
6      world$variable <- casesDifference[match(world$region,
7      chosenDay$location)]
8
9      b <- c(-10000, 0, 10000)
10
11     world_map <- ggplot(world, aes(x = long, y = lat)) +
12       geom_polygon(aes(group = group, fill = variable),
13         col = "white",
14         lwd = 0.1)
15
16     world_map + scale_fill_gradientn(
17       limits = c(-10000,10000),
18       labels = b,
19       breaks = b,
20       colors = c("#0093ff", "#c0e4ff", "#ffa6b6", "#ff0000"),
21       values = scales::rescale(c(-10000, -1, 1, 10000))
22     )
23   })
24 }
25
26 shinyApp(ui = ui, server = server)
```

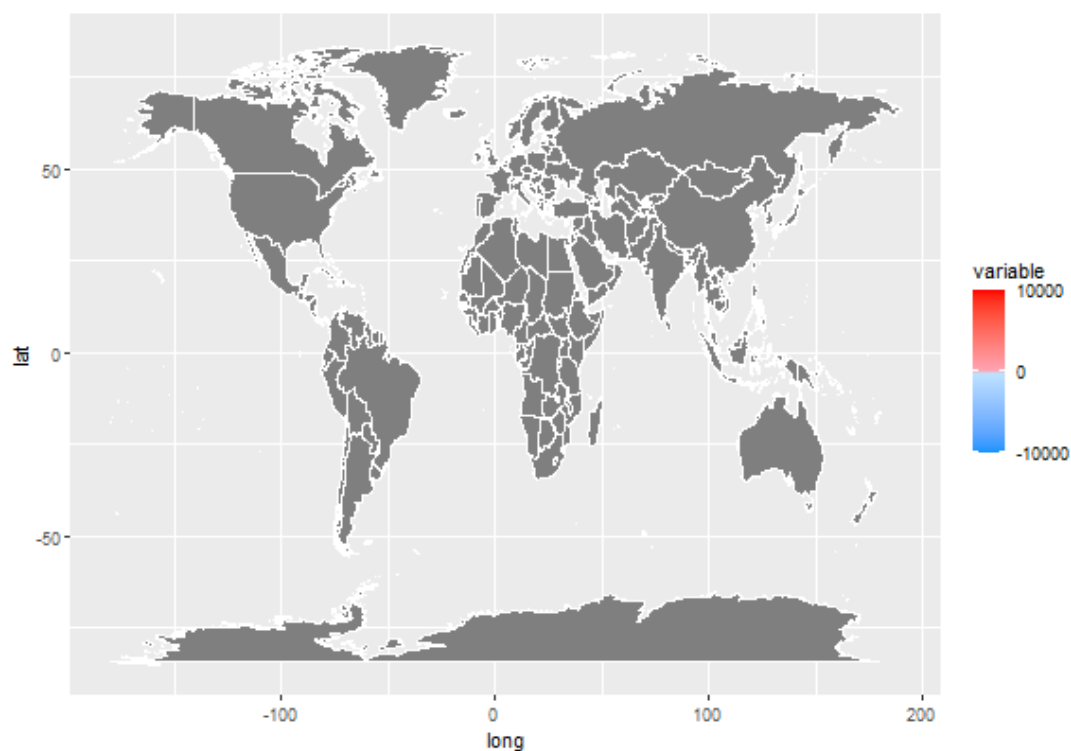
Ispis 3.4 Kreiranje mape svijeta

Poglavlje 4

Rezultati

Pokretanjem prethodno opisanog koda, pojavljuje se prozor grafičkog korisničkog sučelja. Na njemu se prikazuje karta svijeta sa jasno označenim granicama država. Kraj same karte se nalazi izbornik pomoću kojeg definiramo točan datum za koji nas zanima postotak porasta/pada dnevnog broja Covid-19 slučajeva. Sve države koje su predale te iste podatke za određeni datum te dan prije će biti označene specificiranom bojom, dok će ostale države poprimiti neutralnu sivu boju. Prateći upute zadatka, u slučaju porasta broja slučajeva zaraze u odnosu na prethodni dan implementirane su funkcije koje će prostor te određene države obojati u crvenu boju. Pojavom pada broja novozaraženih slučajeva u odnosu na prethodni dan, funkcije u programu takvim državama dodijeljuju plavu boju. Intenzitet dodijeljenih boja ovisi o postotku povećanja vrijednosti u odnosu na prethodni dan koji je izračunat posebnom varijablom. Izračunati intenzitet varira od 1% za najmanji mogući pad/porast do 10000% za najdrastičnije promjene u padu/porastu te se pomoću tog izračuna određuje boja prostora svake određene države na karti.

Poglavlje 4. Rezultati



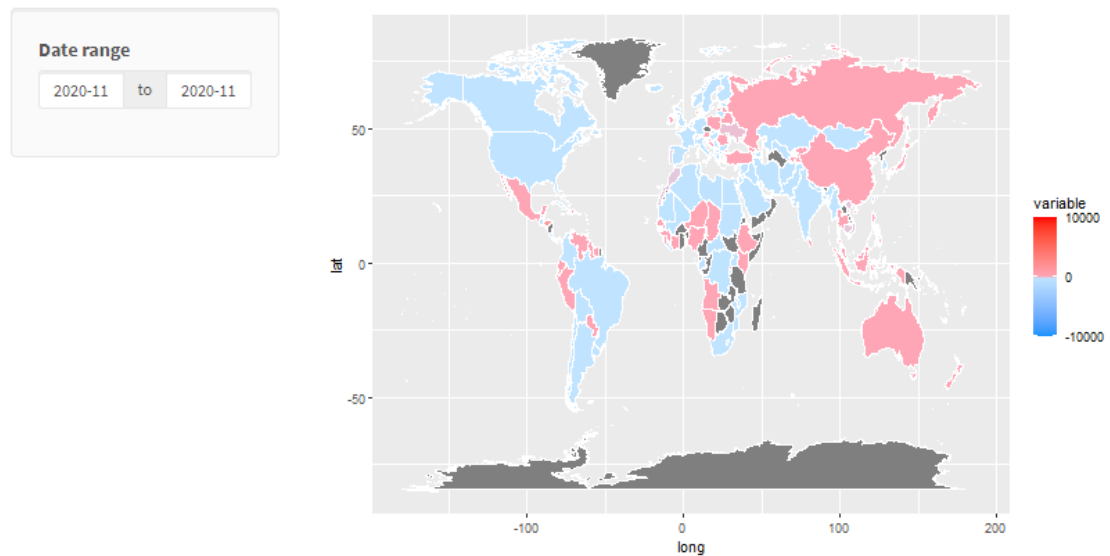
Slika 4.1 Slika zemlje i prikaz država pri inicijalizaciji programa

Nakon korištenja grafičkog sučelja za odabir datuma program u stvarnom vremenu računa podatke te prikazuje obojanu kartu svijeta sa pripadajućom legendom s desne strane. U obzir treba uzeti da određene države i njihovi podatci o Covid-19 zaraženim osobama nisu pokriveni u zadanom datasetu te će se te države neovisno o datumu uvijek prikazivati sivom bojom. Nekonzistencijom podataka u datasetu neke države imaju određene dane za koje nema navedenih podataka o novozaraženima. To dovodi do slučaja u kojima će na karti države koje su inače obojane crvenom ili plavom bojom biti sive za određeno razdoblje u kojem ili nema podataka o novozaraženima za specificirani datum ili za dan prije toga.

Poglavlje 4. Rezultati

U nastavku je prikazan primjer aplikacije u kojemu je odabran određeni datum te na kojem se vide sve navedene funkcionalnosti i nedostaci zadanog dataseta.

Covid 19 data



Slika 4.2 Prikaz novozaraženih osoba Covid-19 virusom na datum 21.11.2020.

Poglavlje 5

Zaključak

Obradom zadanih informacija i izgradnjom predstavljenog programa dobijamo uvid u versatilnost R programskog jezika te mogućnosti manipulacije i prikaza raznih statističkih i lokacijskih podataka oblikovanih u opširnim data setovima. Uvidom i istraživanjem o svim mogućnostima prikaza podataka saznajemo da smo otkrili samo najmanji dio funkcionalnosti i upotrebe raznih funkcija i paketa koji se koriste pomoću R programskog jezika. U ovom su radu obrađeni i statistički prikazani podatci o pandemiji Covid-19 virusa te njegovom utjecaju na razne države iz dana u dana u proteklih par godina sa namjerom proširivanja znanja o uslugama i podacima koji su bazirani na lokaciji.

Bibliografija

- [1] "Our World in Data", s Interneta, <https://github.com/owid/covid-19-data/tree/master/public/data>, 20.1.2023
- [2] " Spatial Data Science with R", s Interneta, <https://rspatial.org/>, 12.1.2023
- [3] "R documentation", s Interneta, <https://www.rdocumentation.org/>, 12.1.2023
- [4] "Cran R project", s Interneta, <https://cran.r-project.org/>, 12.1.2023