

Statističko i računalno modeliranje širenja pandemije COVID-19

Gabrovec, Marko

Undergraduate thesis / Završni rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Pula / Sveučilište Jurja Dobrile u Puli**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:137:781096>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-07-17**



Repository / Repozitorij:

[Digital Repository Juraj Dobrila University of Pula](#)



SVEUČILIŠTE JURJA DOBRILE U PULI
PREDDIPLOMSKI SVEUČILIŠNI STUDIJ INFORMATIKA

**STATISTIČKO I RAČUNALNO MODELIRANJE
ŠIRENJA PANDEMIJE COVID-19**

Završni rad

Pula, 2020

SVEUČILIŠTE JURJA DOBRILE U PULI
PREDDIPLOMSKI SVEUČILIŠNI STUDIJ INFORMATIKA

**STATISTIČKO I RAČUNALNO MODELIRANJE
ŠIRENJA PANDEMIJE COVID-19**

Završni rad

Student: Marko Gabrovec, redovni student

Smjer: Informatika

Kolegij: Statistika

Mentor: doc. dr. sc. Siniša Miličić

Pula, 2020



IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisan Marko Gabrovec, kandidat za prvostupnika Informatike ovime izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega necitiranog rada, te da ikoji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

 Student

U Puli, 19. rujna 2020.



IZJAVA O KORIŠTENJU AUTORSKOG DJELA

Ja, dolje potpisan Marko Gabrovec, dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj Završni rad pod nazivom *Statističko i računalno modeliranje širenja pandemije COVID-19* koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama. Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

Student


U Puli, 19. rujna 2020.

Contents

1. Uvod	1
2. Epidemiološki pojmovi	3
2.1. SIR model.....	3
2.2. Indeks reprodukcije	5
2.3. Veza SIR modela i indeksa reprodukcije.....	6
3. Statistički pojmovi	7
3.1. Klasteriranje k-sredina.....	7
3.2. Linearna regresija	9
3.3. log-log graf i potencijnska ovisnost	11
3.4. Korijen srednje kvadratne pogreške.....	12
3.5. 68 – 95 – 99.7 pravilo	13
4. Modeliranje COVID-19.....	15
4.1. Izvor podataka	15
4.2. Opis dobivenog grafa.....	16
4.3. RMSD vrijednosti	18
4.3.1. Norveška.....	20
4.3.2. Nizozemska.....	21
4.3.3. Srbija	22
5. Zaključak.....	23
6. Literatura	24
7. Popis slika	25
8. Dodatak 1 – kod.....	26
9. Dodatak 2 – bulk grafovi RMSD.....	32

1. Uvod

Razne bolesti oduvijek su se kretale među ljudima. Često u formi endemija jedne te iste bolesti kružile su istom populacijom vezanom jednim područjem. Tek ponekad, neke su bile dovoljno infektivne da zarade naziv epidemije pa čak i pandemije. Dok epidemija nije striktni termin, generalno se definira kao nekontroljivo širenje bolesti na nekom mjestu u nekom vremenu. Pandemija je eskalacija epidemije u trenutku kad nekontroljivim širenjem, bolest zahvati cijelu državu odnosno cijeli svijet.

Krajem 2019. godine, svijet je pogodila, početno epidemija u Kini, a ubrzo i svjetska pandemija COVID-19 virusa. COVID-19, virus je sličan SARS-u, MERS-u i običnoj prehladi. Na našu sreću, ne radi se o novoj kugi. Ipak, COVID-19 je smrtonosan za one najslabije, a nalazi se i sve više dokaza o dugoročnoj štetnosti bolesti čak i na najzdravija tijela. COVID-19 nije izuzetna bolest, ali ju pronalazimo u izuzetnoj etapi razvoja čovječanstva. Globalizacija, kombinirana sa mlakom i neodgovornom reakcijom svjetskih lidera, dovela je do neviđene brzine širenja bolesti. Ono što samu pandemiju čini posebnom, sveopća je dostupnost velikog broja podataka koja po prvi puta zahtijeva znanstvenike za podatke i statističare kako bi se mogli izraditi kvalitetni modeli predikcije širenja pandemije.

Fokus ovog rada je na modeliranju i zornijem prikazu pandemije COVID-19. U tu svrhu koristiti će se već predefinirani epidemiološki modeli (pojmovi: SIR, R_0 ...), ali i određena količina modela iz mašinskog učenja (regresija, algoritam k-sredina...). Mi se nećemo uopće doticati bioloških mehanizama širenja bolesti budući da niti je to fokus rada, niti smo dovoljno upoznati sa tom materijom.

Klasični modeli, tako i najpopularniji SIR model, pretpostavljaju konstantan bazičan indeks reprodukcije bolesti. Posljednji izračuni, i iz njih izvedene predikcije, pokazali su se neadekvatni u predviđanju širenja COVID-19 bolesti. U ranijim fazama su neki to pripisali izuzetnosti COVID-19 bolesti. Daleko vjerojatniji razlog jesu brojne mjere poduzete od strane ljudi ka spriječavanju širenja bolesti te poboljšanje same brige o bolesnicima, koje trenutni klasični modeli ne uzimaju u obzir.

Za potrebe ovog rada modelirati ćemo potencijski odnos novooboljelih i aktivnih bolesnika te ćemo utvrditi prosjek pogreške osnovnog SIR modela za promatrane

skupine. Od modeliranih država, izdvojiti ćemo ekstreme vrijednosti pogreške te ćemo prokomentirati potencijalne razloge velike greške modela tih slučajeva.

Mi ne stvaramo model koji je spreman za korištenje sa točnijim prediktivnim mogućnostima. Mi isključivo testiramo model i potencijalno navodimo propuste modela čijim popravljanjem bi on postao efektivniji u ispravnom modeliranju stvarnog toka širenja bolesti.

2. Epidemiološki pojmovi

Pri izradi modela širenja bolesti nužno je proći kroz ustaljeni vokabular. Termini su brojni, ali mi ćemo se fokusirati isključivo na bazičnu terminologiju koja je osnova trenutačnih modela širenja bolesti.

2.1. SIR model

SIR (*Susceptible – Infected – Removed*) prvi je epidemiološki model predložen 1927. godine od strane Kermack-a i McKendrick-a. Spada pod kategorizirane epidemiološke modele. Kategorizirani epidemiološki modeli su oni koji promatraju širenje bolesti kroz populaciju podijeljenu na kategorije. Jedna od kategorija mogu biti zdrave osobe, druga bolesne, a treća ozdravljene. Same kategorije definiraju se ovisno o bolesti. Neke bolesti imaju isključivo bolesne i zdrave kategorije jer je bolest nepreboljiva i nesmrtonosna. U slučaju COVID-19 bolesti, gdje nakon zaraze nastupa smrt ili ozdravljenje sa statistički zanemarivom mogućnosti reinfekcije koristi se regularni SIR model. (Martcheva, 2015., str. 9. – 10.)

Svako slovo skraćenice modela opisuje uključene kategorije populacije u modelu. *Susceptible* (S) jeste kategorija zdravih i neimuniziranih pojedinaca. *Infected* (I) jeste kategorija infektivnih odnosno zaraznih pojedinaca. *Removed* (R) jeste kategorija maknutih pojedinaca, odnosno kategorija oporavljenih, umrlih ili cijepljenih pojedinaca. Sam redoslijed znakova u skraćenici predstavlja i tok pojedinaca kroz kategorije. (S -> I -> R). Ovisno o tipu zaraze pa tako i prigodnom modelu, znakovi se mijenjaju odnosno nadodaju. Tako model za bolest sa mogućnošću ponovne zaraze pišemo kao SIRS (*Susceptible -> Infectious -> Recovered -> Susceptible*). (Jones, 2007., str. 1. – 2.)

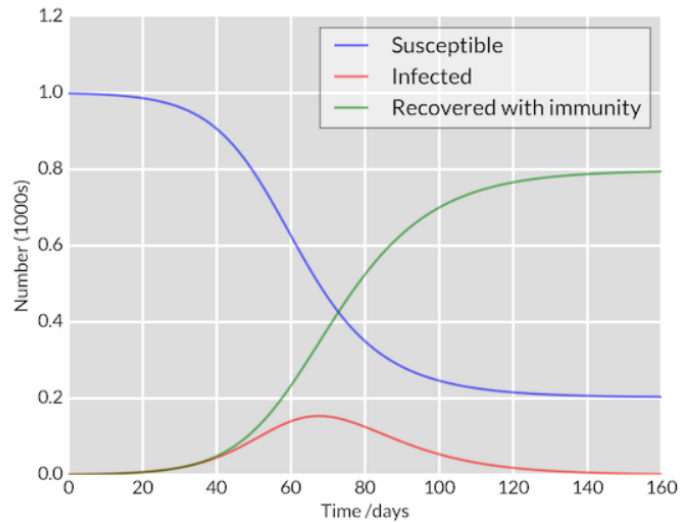
Matematički, SIR model je opisan sustavom diferencijalnih jednadžbi:

$$S'(t) = -\beta IS,$$

$$I'(t) = \beta IS - \alpha I,$$

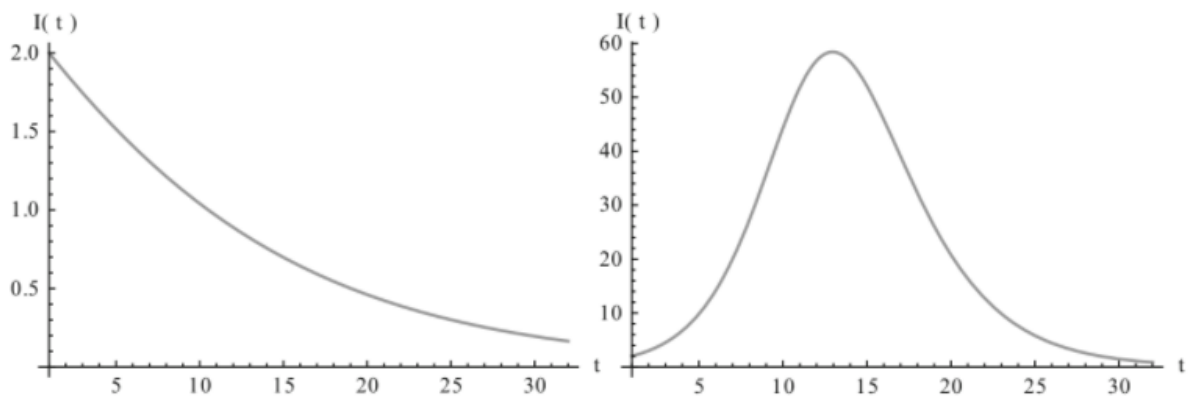
$$R'(t) = \alpha I,$$

gdje su, S, I i R oznake kategorija, β vjerojatnost da infektivna osoba zarazi zdravu, neimuniziranu osobu, a α vjerojatnost da infektivna osoba postane dio kategorije maknutih pojedinaca. (Martcheva, 2015., str. 10. – 12.)



Slika 1. Graf SIR modela (Izvor: <https://scipython.com/book/chapter-8-scipy/additional-examples/the-sir-epidemic-model/>)

Slika 1. predstavlja osnovni graf SIR modela. Plava linija prikazuje zdravu neimuniziranu populaciju, crvena infektivnu, a zelena onu maknutu (mrtvu ili imuniziranu).



Slika 2. Grafovi infektivne populacije SIR modela (Izvor: Martcheva, 2015., str. 12.)

Na slici 2 se jedan do drugoga nalaze grafovi varijante moguće promjene u kategoriji infektivnih. Lijevo broj zaraženih monotono pada kako stalno raste broj zdravih. Desno broj zaraženih nekim događajem naglo raste do svog maksimuma da bi opet počeo monotono padati.

Ciljevi za korištenje modela su razni. On se koristi za predviđanje: kako se bolest širi, ukupnog broja zaraženih, trajanja epidemije, ali i epidemioloških parametara kao što je indeks reprodukcije. Modeli daju i uvid u efektivnost raznih javnih zdravstvenih odluka te mogu pomoći pri daljnjem donošenju odluka o zaoštavanju ili ublaživanju pojedinih odluka pa čak i kod odlučivanja o procijepljivanju populacije kod limitiranih količina cjepiva. (Brauer, Chavez, Feng, 2019., str. 3. – 4.)

2.2. Indeks reprodukcije

Indeks reprodukcije, inače R_0 , može se zamisliti kao očekivani broj sekundarnih slučajeva direktno uzrokovanih od jednog zaraženog člana inače na tu bolest neimunizirane populacije. Indeks reprodukcije ne predstavlja realnu situaciju već je aproksimacija sustava bez ljudskog djelovanja na širenje bolesti i bez imuniziranih odnosno umrlih pojedinaca. Ipak, predstavlja glavnu mjeru širenja bolesti. Vrijednost ne predstavlja stopu rasta nego sveukupni broj sekundarno oboljelih pojedinaca bez vremenske dimenzije, odnosno u cijelom predviđenom trajanju infektivnosti. Indeks reprodukcije u osnovi računamo kroz slijedeću formulu:

$$R_0 = \tau \times \bar{c} \times d ,$$

gdje, τ predstavlja zaraznost bolesti odnosno vjerojatnost zaraze sa danim kontaktima zdrave, neimunizirane osobe i one zarazne. \bar{c} predstavlja prosjek kontakata između te dvije skupine populacije. d predstavlja trajanje infektivnosti pojedinca. (Jones, 2007., str. 1.)

Računanje indeksa reprodukcije u bilo kojem datom trenutku lagan je zadatak, međutim ispravan izračun je nešto potupuno drugačije. Ovisno o načinu računanja i korištenoj matematici, razni istraživači tako pronalaze razne vrijednosti (COVID-19, startnih 1.4 sve do 2.8). Problem je uspostavljanje ispravnosti indeksa na stvarno

širenje bolesti. U tu svrhu, validnost indeksa reprodukcije provjerava se kroz razne epidemiološke modele. (<https://www.healthline.com/health/r-nought-reproduction-number#covid-19-r-0>, rujan 2020.)

Nužnost konstante provjere ispravnosti R_0 je strahovita važnost tog indeksa. Gotovo sve odluke bitne po javno zdravstvo za vrijeme epidemije koriste upravo taj indeks. Velika pogreška indeksa reprodukcije potencijalno dovodi do velike pogreške u ljudskoj reakciji na njeno širenje.

2.3. Veza SIR modela i indeksa reprodukcije

SIR model, strogo je određen vrijednosti indeksa reprodukcije. Ako je korišten, izračunat R_0 manji od stvarnoga, ekstrem brojnosti infektivne populacije raste, gornji limes brojnosti preboljele populacije se diže, a donji limes brojnosti zdrave populacije pada. Suprotno, ako je korišten R_0 veći od stvarnoga, gornji i donji limesi se približavaju dok se ekstrem brojnosti infektivne populacije obuzdaje. Prevedeno, pogreške u korištenom indeksu reprodukcije drastično mijenjaju broj populacije koji će se u danom trenutku nalaziti u jednoj od SIR kategorija. (Jones, 2007., str. 2. – 6.)

Matematički, indeks reprodukcije se ponaša kao prag dinamike modela. Ako vrijedi $R_0 < 1$, broj infektivnih opada sve dok bolest ne biva iskorijenjena iz populacije. Ako vrijedi $R_0 > 1$, broj infektivnih se kad tad stabilizira i bolest postaje endemična populaciji. Vrijednost i njena definicija ne razlikuju se mnogo od βIS dijela sustava diferencijalnih jednadžbi. Ona samo u izračun startno uključuje cjelokupnu populaciju zdrave neimunizirane populacije te vrijeme infektivnosti bolesti. (Martcheva, 2015., str. 21.)

Formula R_0 izvedena iz veze sa SIR modelom tako glasi:

$$R_0 = \frac{\beta N}{\alpha} .$$

3. Statistički pojmovi

U svrhu izrade statističkog modela širenja bolesti, moramo posjedovati set alata za manipulaciju podacima te stvaranje znanja iz istih. Pri izradi modela, koristili smo Python programski jezik. Korištene knjižnice, između ostalih, koje su nam dale te alate jesu scikit -learn, knjižnica fokusirana na mašinsko učenje, i bokeh, knjižnica-alat za izradu grafova i obavljanje jednostavnijih grafičkih statističkih izračuna. Kako bi se kasnije moglo opisati stvaranje i rad modela, proći ćemo kroz unutarnji rad i teoriju korištenih dijelova tih knjižnica.

3.1. Klasteriranje k-sredina

Klasteriranje k-sredina (k-means clustering), model je strojnog učenja bez nadzora. Prvenstveno se koristi pri klasifikaciji i regresiji danih podataka. Ideja je određivanje predstavnika svake skupine te priključivanja svake točke skupinama na način da zbroj kvadrata udaljenosti točaka od predstavnika skupine kojoj pripadaju bude minimalan. (Wagstaff, Cardie, Rogers, Schroedl, 2001., str. 2. – 3.)

Klasteriranje se može implementirati kroz nekoliko algoritama. Ipak, Lloyd algoritam, je standard. Činjenica da je daleko najčešće korišten, smatra ga se pravim predstavnikom algoritama klasteriranja k-sredina te on još nosi i naziv algoritam k-sredina ili standardni algoritam. Po svojoj prirodi naivan, temelji se na iteraciji jednostavnih, ali skupih funkcija sve do jedne iteracije u kojoj se ne mijenja stanje klastera. To dovodi do sporijeg vremena izvođenja, ali sa razumnim vrijednostima u potpunosti je sposoban brzo dati kvalitetne rezultate.

Za Lloyd algoritam vrijedi; ako je zadan skup točaka,

$$X = \{x_1, x_2, \dots, x_m\},$$

u n dimenzionalnom prostoru R_n cilj algoritma je pronaći, za unaprijed zadani $k \geq 2$, optimalnu particiju skupa X ,

$$\pi = \{C_1, C_2, \dots, C_k\},$$

odnosno razmjestiti m točaka skupa X u k skupina (klastera). Svakoj skupini pridružena je točka,

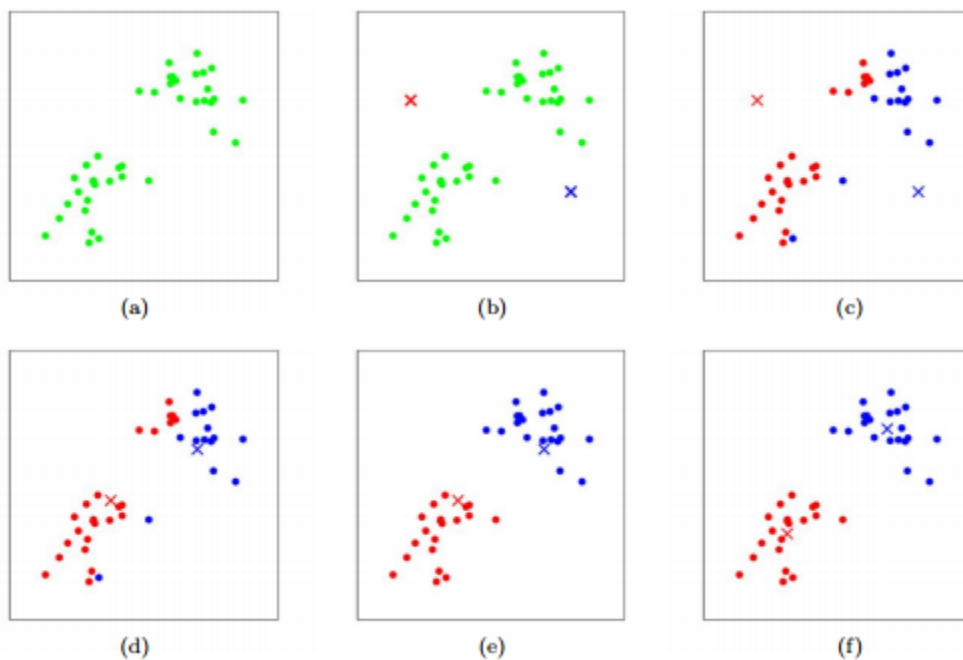
$$c_i = \frac{1}{|C_i|} \times \sum_{x \in C_i} x,$$

koju nazivamo predstavnikom skupine (ona ne mora biti član skupa X). Temeljni algoritam jednostavan je činjenica da optimalni izbor predstavnika skupine je samo središte skupine. Kako bi uopće započeo rad algoritma potrebno je postaviti početne vrijednosti predstavnika skupina, najčešće slučajnim odabirom k točaka X skupa.

Algoritam k-sredina je iterativni algoritam izvršavan u dva osnovna koraka:

- Pridruživanje – Za svaku točku x vršimo pridruživanje najbližem predstavniku skupine, na kraju koraka svaka točka dio je jednog C skupa.
- Prepravljanje – Iz dobivenih skupova, računa se novi predstavnik skupa tako postavljen da varijanca bude najmanja moguća.

Ti se koraci ponavljaju sve dok raspored skupova i lokacija predstavnika ne krenu konvergirati kroz sekventne iteracije. (Mirošević, 2016., str. 91. – 93.)



Slika 3. Lloyd algoritam kroz korake na neoznačenim grafovima (Izvor: <https://stanford.edu/~cpiech/cs221/img/kmeansViz.png>)

Na grafu (a), vidimo skup nekih neoznačenih podataka. Na grafu (b), slučajnim odabirom za $k = 2$ odrede se dva centra budućih klastera. Graf (c) označava korak pridruživanja. Svaka točka nadodaje se svom najbližem skupu. Kod grafa (d) dešava se prepravljjanje te se centar svakog klastera pomiče. Graf (e), ponovljeno pridruživanje za nove centre, i završno (f), ponovljeno prepravljjanje po novim klasterima. Kad bi se dodao novi graf (g), naše točke bi prilikom novog pridruživanja opet bile dio istog klastera kojemu su pripadale prije tog pridruživanja. Tada, naš algoritam prestaje sa svojim radom i mi imamo dva klastera koji cjelokupni skup podataka dijele sa najmanjom mogućom udaljenosti od centara klastera.

3.2. Linearna regresija

Regresijska analiza, skup je statističkih procesa za estimaciju veze jedne ili više nezavisnih varijabli X i jedne ili više zavisnih varijabli Y . Linearna regresija najstariji je i najpopularniji oblik regresijske analize. Fokusira se na razdiobu uvjetne vjerojatnosti od Y za zadani X pronalaženjem pravca koji je najbliže prilagođen danim podacima. Prevedeno, iz nekog skupa rezultata populacije, estimiraju se osobine populacije. Razlog njezine popularnosti leži u jednostavnosti korištenja i čitkosti dobivenih rezultata u odnosu na modele koji nelinearno ovise o parametrima.

Za linearnu regresiju vrijedi; ako imamo skup podataka,

$$\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n,$$

od n elemenata, linearna regresija pretpostavlja linearni odnos zavisne varijable y i novonastalog multivektora regresora x . Linearnost se objašnjava kroz postojanje varijable pogreške ε – slučajne varijable koja dodaje šum u linearni odnos zavisne varijable i regresora. Model tako glasi,

$$y = X \times \beta + \varepsilon,$$

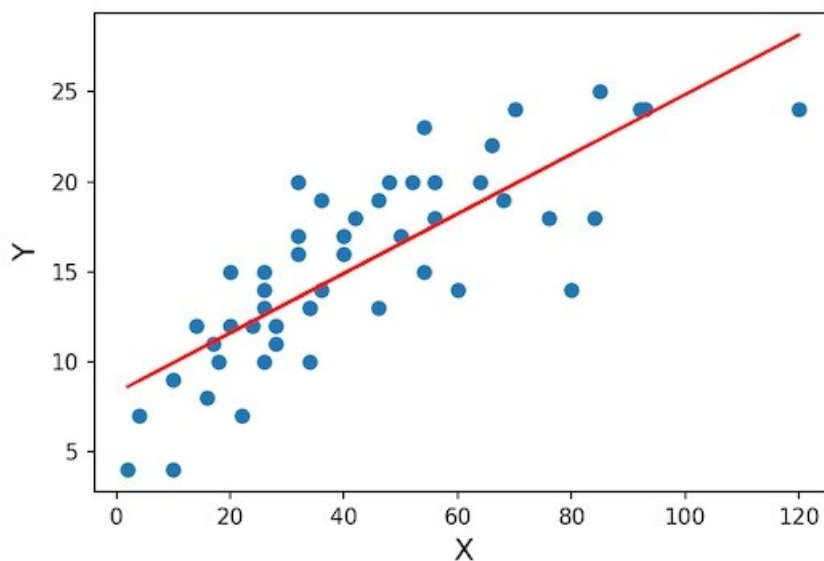
gdje,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

sa značenjem; y je vektor zavisnih odnosno predviđenih varijabli, X je vektor regresora, nezavisnih varijabli odnosno prediktornih varijabli, β je parametarski vektor regresijskih koeficijenata koji prikazuje promjenu zavisne varijable za jedinku u odnosu na ostale nezavisne varijable, a ε je vektor vrijednosti šuma regresije koji za svaku jedinku definira odstupanje od pretpostavljenog linearnog odnosa. (<https://online.stat.psu.edu/stat462/node/132/>, rujan 2020.)



Slika 4. Graf sa linearnom regresijom (crveno) skupa podataka (plavo) (Izvor:

Na grafu možemo vidjeti primjer linearne regresije. Za jedan skup podataka označen plavim točkama, računa se crveni pravac linearne regresije. Kad bi se povukle linije od plavih točaka do crvenog pravca paralelno na y -os te se te linije zbrojile, za ispravno izračunatu linearnu regresiju, dobili bi najmanji mogući zbroj u odnosu na bilo koji mogući pravac.

3.3. log-log graf i potencijnska ovisnost

Log-log graf koristi varijable x i y , ali u njihovoj logaritmiranoj verziji $\log(x)$ i $\log(y)$. Za razliku od normalnog, jednostrukog log grafa primjenjenog u epidemijologiji koji najčešće služi boljem prikazu naglo rastućih vrijednosti u vremenskom intervalu, log-log graf ima manje vizualnu, ali svakako praktičnu primjenu. Prilikom unosa točaka na log-log graf, ako postoji potencijnska ovisnost nazirati će se linearnost aproksimacije vrijednosti unesenih točaka. Svođenjem točaka na jedan pravac linearnom regresijom steći će se potencija međuodnosa kvantiteta x i y . (<https://statisticsbyjim.com/regression/log-log-plots/>, rujan 2020.)

Formula (i njene izvedenice) koja u linearnosti log-log grafa za logaritmirani (po bilo kojoj bazi) monom,

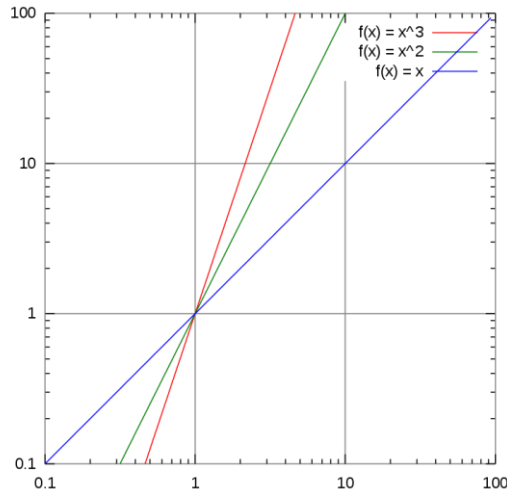
$$y = A \times x^n,$$

$$\log(y) = \log(A) + n \times \log(x),$$

opisuje odnos kvantiteta $X = \log(x)$ i $Y = \log(y)$, što odgovara log-log grafu, glasi:

$$Y = n \times X + b$$

Formula je zapravo jednostavan račun pravca. n poput k predstavlja nagib pravca dok b predstavlja sjecište y -osi, ovdje $\log(y)$ -osi. (<https://datascientia.blog/2017/08/14/dss-llrm/>, rujan 2020.)



Slika 5. Log-log graf pravaca različite potencijske ovisnosti (Izvor: https://upload.wikimedia.org/wikipedia/commons/9/92/LogLog_exponentials.svg)

Potencijska ovisnost, pak, odnos je dviju nekih količina gdje relativna promjena kvantitete jedne količine uzrokuje proporcionalnu relativnu promjenu druge količine, nevezano za početne kvantitete tih količina. Prevedeno, jedna kvantiteta varira kao potencija druge kvantitete.

Značaj dokaza potencijske ovisnosti je bitan. Trenutni epidemiološki modeli (SIR i izvedenice) za monom koji opisuje ovisnost startno zaključuju potencijnsku ovisnost 1. To bi značilo da će se po SIR modelu za trenutni broj aktivnih slučajeva, isključivo po indeksu reprodukcije, desiti točno izračunljiv broj novih slučajeva (npr. za $R_0 = 3$ i 1000 slučajeva, uvijek 50 novih). I dok se ta vrijednosti ne odmiče uvijek od 1, u većini slučajeva, odmak je statistički značajan te se stalno mijenja. (dokaz mjenjanja R_0) (<https://www.statisticshowto.com/power-law/>, rujan 2020.)

3.4. Korijen srednje kvadratne pogreške

Korijen srednje kvadratne pogreške, skraćeno RMSD (root-mean-square deviation) ili RMSE (root-mean-square error), često je korištena mjera razlike vrijednosti nekog modela i stvarno izračunatih vrijednosti. RMSD pokazuje preciznost modela u odnosu na prave vrijednosti te tako definira vjerojatnost nastanka i veličine pogreške u predikcijama modela.

RMSD je uvijek pozitivan. Vrijednost 0 označava potpunu podudarnost modela podacima (nešto što skoro nikada nije slučaj). Kako RMSD raste, smanjuje se preciznost modela. Prevedeno, što je broj bliži 0, to smo više potvrdili ispravnost SIR modela za primjer.

Formula za izračun vrijednosti RMSD glasi:

$$RMSD = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}$$

gdje, \hat{y}_t predstavlja pretpostavljenu funkcijsku vrijednost člana populacije u jedinici vremena prema postavljenom linearnom modelu, y_t stvarnu vrijednost člana populacije u jedinici vremena i T predstavlja broj jedinica vremena u kojima je računat y . (<https://statweb.stanford.edu/~susan/courses/s60/split/node60.html>, rujan 2020.)

3.5. 68 – 95 – 99.7 pravilo

68 – 95 – 99.7 pravilo, ili empirijsko pravilo, skup je postotaka vrijednosti koje se nalaze uz prosječnu vrijednost unutar udaljenosti od jedne, dvije te na poslijetku tri standardne devijacije.

Matematički prikaz pravila glasi:

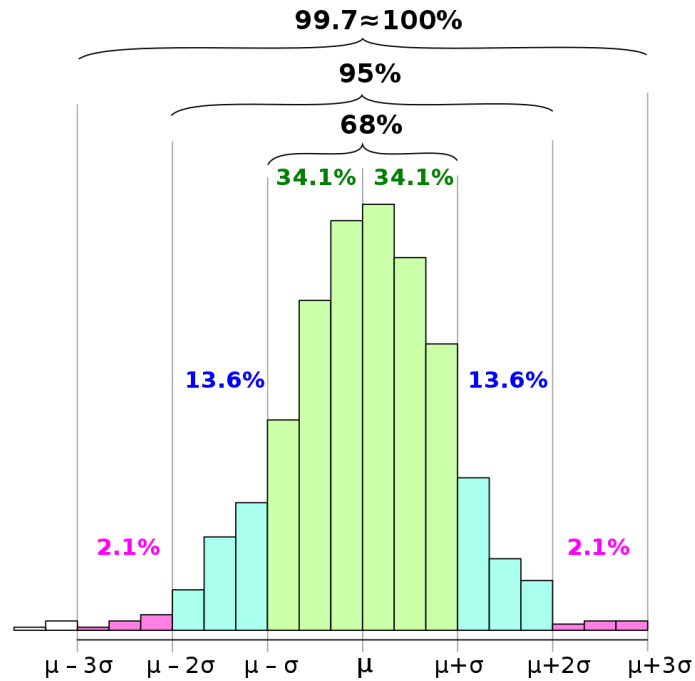
$$Pr (\mu - 1\sigma \leq X \leq \mu + 1\sigma) \approx 0.6827,$$

$$Pr (\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.9545,$$

$$Pr (\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0.9973,$$

gdje se standardna devijacija računa kao,

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}.$$



Slika 6. 68 – 95 – 99.7 pravilo prikazano grafički (Izvor: https://upload.wikimedia.org/wikipedia/commons/2/22/Empirical_rule_histogram.svg)

Pravilo, time i ispituje ekstremnih vrijednosti, jedan je od najmoćnijih alata u statistici. Ispitivanjem količine tih ekstrema, možemo saznati vrijedi li uopće naš postavljeni model za skup podataka. Promatranjem direktnih slučajeva ekstremnih vrijednosti možemo doći do novog uvida u generalni rad modela. (Utts, Heckard, 2007., str. 47. – 54.)

4. Modeliranje COVID-19

U ovom poglavlju proći ćemo kroz korištene podatke, opisati proceduru izrade modela, te na kraju i prikazati i opisati neke dobivene grafove i pokušati objasniti postojanje ekstremnih vrijednosti korijena srednje kvadratne pogreške.

4.1. Izvor podataka

Sreća u nesreći, COVID-19 pandemija je prva pandemija sa donekle centraliziranim, globalnim prikupljanjem podataka o toku bolesti. John Hopkins University uzeo si je zadatak prikupljanja i organiziranja podataka o epidemiji te dijeljenja istih svakome tko te podatke zatreba. Podaci se unose u nekoliko različitih baza i preko git-a.

Za izradu modela koristili smo raw git podatke. Iz mape, *csse_covid_19_data* pa unutar nje *csse_covid_19_time_series*, pri pokretanju aplikacije za izradu grafova modela povlače se .csv datoteke *time_series_covid19_confirmed_global*, *time_series_covid19_deaths_global* i *time_series_covid19_recovered_global*. Sve .csv datoteke sastoje se od dijela koji je opisnik države (možda i država unutar država federacija ili pojedinih regija velikih država) i niza datuma počevši od 22.01.2020. sve do današnjeg dana (zbog vremena unosa potencijalno jučerašnjeg) sa nadodanim vrijednostima oboljelih, mrtvih odnosno oporavljenih ljudi (korelacija sa SIR). (https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series)

Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	1/31/20
	Afghanistan	33.93911	67.709953	0	0	0	0	0	0	0	0	0	0
	Albania	41.1533	20.1683	0	0	0	0	0	0	0	0	0	0
	Algeria	28.0339	1.6596	0	0	0	0	0	0	0	0	0	0
	Andorra	42.5063	1.5218	0	0	0	0	0	0	0	0	0	0
	Angola	-11.2027	17.8739	0	0	0	0	0	0	0	0	0	0
	Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0	0	0	0	0
	Argentina	-38.4161	-63.6167	0	0	0	0	0	0	0	0	0	0
	Armenia	40.0691	45.0382	0	0	0	0	0	0	0	0	0	0
Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0	0	0	0	0	0	0	0
New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4	4	4	4	4

Slika 7. Tablični prikaz djeliča korištenih podataka (Izvor:

[https://github.com/CSSEGISandData/COVID-](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv)

[19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv](https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv))

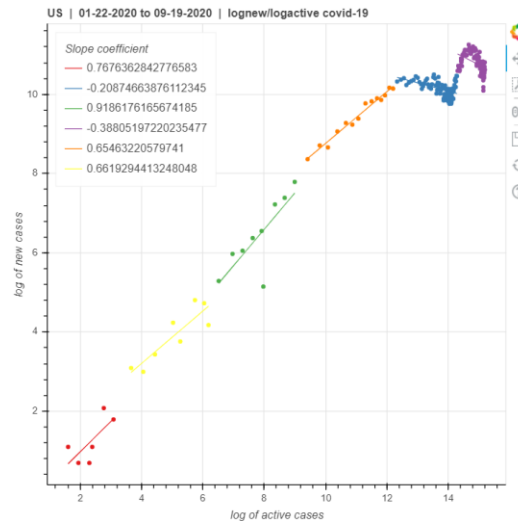
Same podatke nismo mogli samo skinuti i već krenuti sa izradom modela. Te podatke smo prvo morali očistiti od nepotrebnih podataka te im za potrebe ovog rada smanjiti granularnost. Tako su neki stupci opisnici zbog redundancije obrisani, a razni retci grupirani kako bi granularnost sveli isključivo na države. Tu obradu vršili smo pomoću pandas knjižnice za Python. Cjelokupni unos i prvotnu obradu podataka u kodu naše aplikacije obavlja funkcija `import_data()`.

Značajno, `set_filter_sample()` i `set_date_sample()`, miču određene retke odnosno stupce iz izrade modela. `set_filter_sample()` za primljeni argument definira broj slučajeva potrebnih za uključenje u izradu grafa modela. `set_date_sample()` bez argumenta dopušta definiranje početnog i završnog datuma uzorka koji se koristi za izradu grafa modela, odnosno, sa argumerentom definira raspon dana uzorka prije posljednjeg koji se koristi za izradu grafa modela.

4.2. Opis dobivenog grafa

Sa odabranim datumom te selektiranom državom (ako je neselektirana cijelim svijetom) u aplikaciji se kreiraju grafovi po našem modelu. Sama pretvorba podataka u prikaz modela događa se `kmeans_regression()` funkcijom. Funkcija prima argument države koju promatramo (pušteno prazno, "World") i količine klastera na koje želimo podijeliti skup podataka (pušteno prazno, 1). Produkt funkcije je log-log graf, obični graf slučajeva te vrijednost korijena srednje kvadratne pogreške o kojoj će više govora biti kasnije.

Za potrebe opisivanja grafova, koristiti ćemo najrazvikaniji primjer širenja bolesti, Sjedinjene Američke Države od datuma 22.01.2019. sve do 19.09.2020. sa 6 pojedinačna klastera.



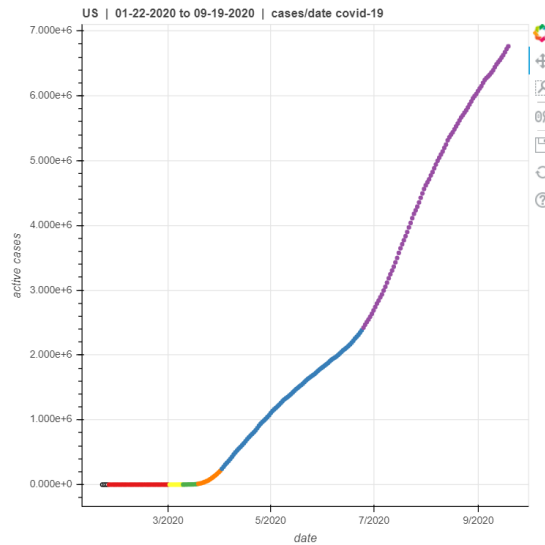
Slika 8. Log-log graf po modelu za SAD, od 22.01.2019. do 19.09.2020. podjeljen kroz 6 klastera (Izvor: izrada autora)

Na slici 6. nalazi se log-log graf modela. Na x -osi nalazi se logaritamska vrijednost aktivnih slučajeva. Aktivni slučajevi su zbroj svih slučajeva na dani datum umanjeno za sve umrle i oporavljene osobe. Broj u svakom trenutku definira infektivni skup populacije, iz kojega bi širenjem trebali nastati novi slučajevi. Na y -osi nalazi se logaritamska vrijednost novih slučajeva. Novi slučajevi su zbroj svih slučajeva na dani datum umanjeno za zbroj svih slučajeva na jučerašnji datum. Broj definira novonastale slučajeve u tom trenutku.

Po samom značenju osi x i y da se zaključiti da bi veza između te dvije vrijednosti po samom modelu trebala biti velika. Dok u nekim fazama, potencijalna ovisnost novonastalih slučajeva pokazuje snažnu vezu sa brojem aktivnih slučajeva, u drugim fazama vidljiva je daleko slabija međuovisnost tih vrijednosti.

Problematici, za graf modela, su periodi naglih pozitivnih promjena u broju bolesnih. Kako se radi o grafu bez vremenske crte, promjene koje reduciraju populacije x ili y osi mogu stvoriti statistički problematična područja grafa. Taj problem ne javlja

se često te je više endemičan za primjere država koje ozbiljno drže pandemiju pod kontrolom ili država koje neozbiljno vode evidenciju oboljelih i ozdravljenih.



Slika 9. Obični graf za SAD, od 22.01.2019. do 19.09.2020. sa odgovarajućim točkama obojanim u boju klastera kojemu pripadaju na log-log grafu (Izvor: izrada autora)

Na slici 7. nalazi se običan graf toka pandemije. x -os je vremenska os sa početnim datumom 22.01.2019. te završnim 19.09.2020. y -os je skup slučajeva za dani datum. Prikaz je isti najčešćem grafu primjenjenom u epidemiološke svrhe, nelogaritamskom prikazu broja slučajeva po danu. Ono što ovu iteraciju grafa čini posebnom je bojanje pojedinačnih podatkovnih čvorova u boje klastera kojima pripadaju u log-log grafu. To znači da svaku točku opisuje potencijalna ovisnost klastera kojoj pripada. Tako, možemo promatrati ponašanje krivulje u odnosu na izračunatu potencijalnu ovisnost klastera.

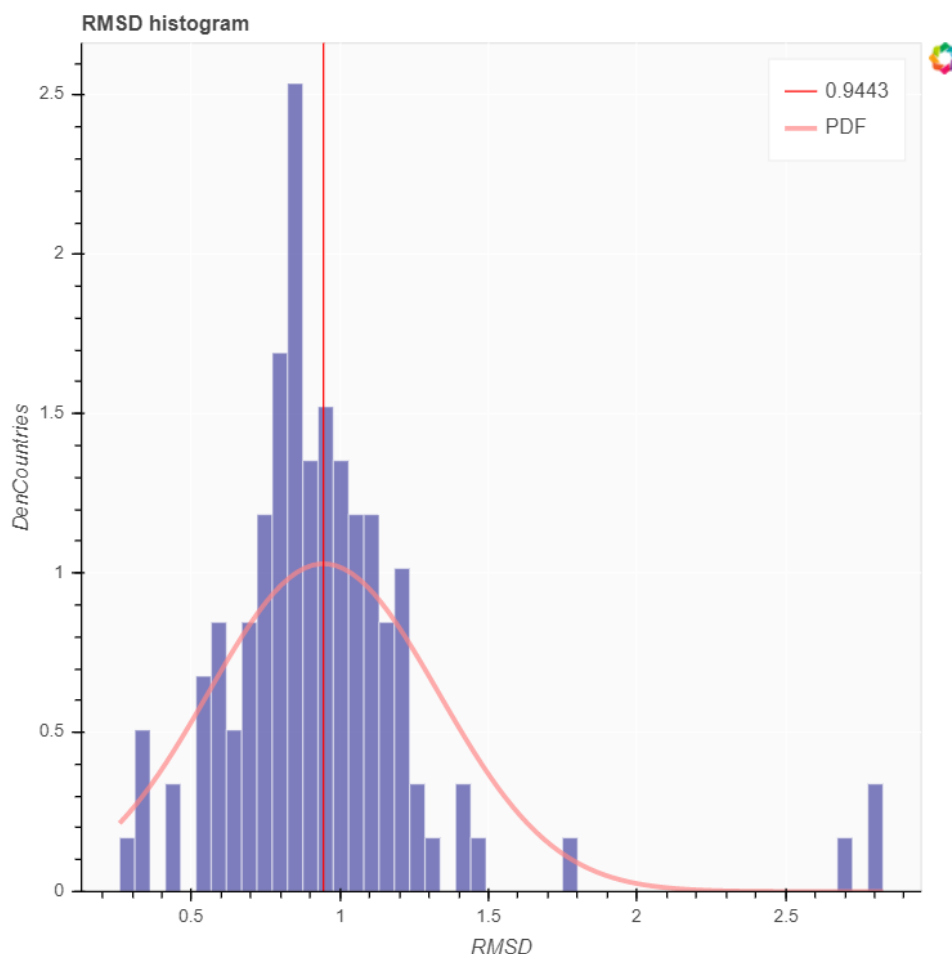
Značaj grafa nije jednak ovom prethodnom, ali za razliku od log-log grafa, puno je smisleniji i čitljiviji. Na kraju, on pomaže u usidrenju znanja stečenog prethodnim log-log grafom, te je koristan i tokom analize ekstrema premo histogramu.

4.3. RMSD vrijednosti

RMSD vrijednosti svakako se puno razlikuju od države do države. Kako bi se samoj vrijednosti pridodao smisao, nužno je izračunati prosjek vrijednosti. Za svih

115 država sa preko 5000 slučajeva (statistička relevantnost, izbacivanje država sa nekonstantnim prikupljanjem koje dovodi do pogrešaka izračuna) računaju se pojedinačni RMSD-ovi te dolazimo do generalnog globalnog stupnja pogreške. Pri našem izračunu za 6 klastera prosjek RMSD-a iznosi 0.9443.

Vrijednost dobivenog RMSD-a pokazuje postojeću pogrešku u skoro svim državnim modelima. Ipak, pogreške se mogu pojaviti zbog raznih stvari. Rubne vrijednosti, pogreške u prikupljanju podataka, proglašavanja velikog broja bolesnih ili zdravih svi mogu pridonijeti vrijednosti RMSD-a. Iz toga proizlazi kako je barem minimalna pogreška neizbježna. 0, ustvari, definira savršen model. RMSD zato više treba koristi kao vrijednost u odnosu na referentni prosječni RMSD. Tako vrijednosti bliske prosjeku RMSD-a jesu unutar prosjeka pogreške, one udaljenije pak ili manje pogrešne, ako su bliže 0, ili više, ako se udaljuju od prosjeka i 0.

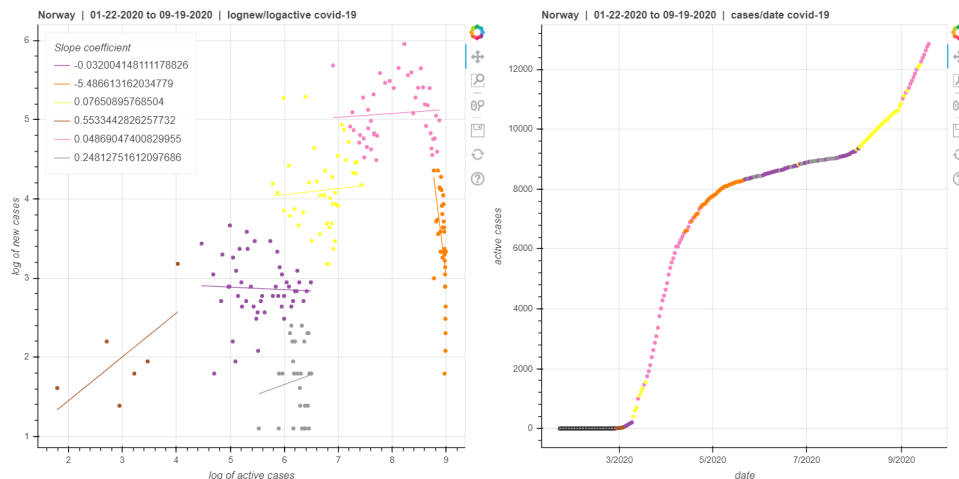


Slika 10. Histogram sa objedinjenim vrijednostima RMSD-a za statistički relevantne države (Izvor: izrada autora)

Na slici 8 nalazi se histogram. Na y-osi nalazi se gustoća frekvencije. Jedna uniformna visina predstavlja jedan RMSD izračunat za pojedinu državu. x-os predstavlja izračunati RMSD za pripadajući skup. Crvena linija odgovara prosjeku izračunatih RMSD-ova. PDF, *probability density function* ili funkcija gustoće vjerojatnosti, krivulja je čija udaljenost od x-osi označava vjerojatnost pojave te vrijednosti. Rubne vrijednosti su malo vjerojatne dok su one centralne vrijednosti česte.

Za prethodni primjer SAD-a za razdoblje od 22.01.2019. do 19.09.2020. sa 6 pojedinačnih klastera, izračun RMSD-a iznosi 1.1875. Ta vrijednost nalazi se unutar prve standardne devijacije prosjeka RMSD-a te se tako može smatrati uistinu prosječnom vrijednosti normalne distribucije nad tom populacijom. U uvodu spomenuto, mi ćemo ukratko promotriti države sa ekstremnim RMSD vrijednostima. Kako SAD ne odgovara svojim RMSD-om mi ga napuštamo te promatramo države sa vrijednostima RMSD-a izvan 3 standardne devijacije: Norveška, Nizozemska i Srbija.

4.3.1. Norveška



Slika 11. Grafovi za Norvešku (Izvor: Izrada autora)

Norveška se postavlja kao najveći pozitivni ekstrem populacije. Strahovito bogata sa informiranom i savjesnom populacijom, Norveška je jedna od država sa najboljom reakcijom na pandemiju. Norvežani su tako, ranije od gotovo svih drugih,

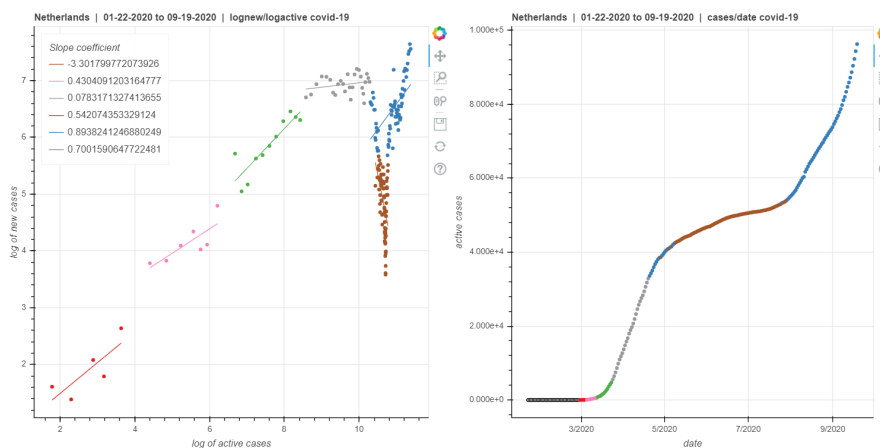
uveli stroge mjere kontrole kretanja populacije, a isto tako su bili i među najsporijima za ukidanje istih. Norveška je i dalje najtestiranija država u Europi, a visok standard nudi bolje i sigurnije liječenje. (<https://www.lifeinnorway.net/coronavirus-in-norway/>, rujan 2020.)

Neizostavan dio uspjeha Norveške u borbi sa pandemijom je i izuzetno mala gustoća naseljenosti. Pri dnu ljestvica gustoće naseljenosti i bez sindroma velikog grada okruženog pustoši, u Norveškoj je pandemija osim zdravstvenog sustava i opreznosti ljudi trebala preći i čistu geografsku barijeru velike, slabo napućene zemlje.

	Belarus	70	207,000	9,700,000
	Montenegro	45	13,810	622,387
	Lithuania	43	65,300	2,795,674
	Faroe Islands (Denmark)	37	1,399	51,095
	Latvia	30	64,589	1,921,800
	Estonia	29	45,227	1,319,133
	Russia (European part)	29	3,960,000	113,112,406
	Sweden	23	450,295	10,215,250
	Finland	16	338,424	5,521,533
	Norway	14	385,203	5,323,933
	Iceland	3.5	102,775	355,620
	Total	73	10,160,000	740,000,000

Slika 12. Dno tablice gustoća populacije u Europi (Izvor: https://en.wikipedia.org/wiki/Area_and_population_of_European_countries)

4.3.2. Nizozemska

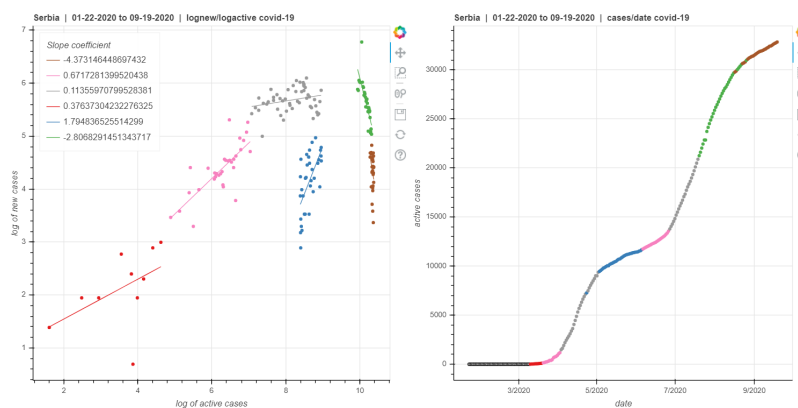


Slika 13. Grafovi za Nizozemsku (Izvor: Izrada autora)

Nizozemska je puno gore prošla od Norveške u početnim fazama pandemije. Za razliku od Norveške radi se ipak o puno gušće naseljenoj državi sa daleko većim globalnim integracijama. Unatoč tome, Nizozemska dijeli puno osobina Norveške,

dobro organiziranu, bogatu državu sa dobro informiranom savjesnom populacijom. To joj, slično kao i Norveškoj, nudi mogućnost bržeg testiranja većih dijelova populacije te pružanje bolje skrbi bolesnima. Tako je unatoč velikom broju prvotno oboljelih, prema kojemu bi se dala predvidjeti situacija sličnija Španjolskoj, Nizozemska uspjela izuzetno dobro staviti pandemiju pod kontrolu. (<https://dutchreview.com/news/coronavirus-netherlands/>, rujan 2020.)

4.3.3. Srbija



Slika 14. Grafovi za Srbiju (Izvor: Izrada autora)

Srbija je veoma različita u odnosu na Norvešku i Nizozemsku. Poput svojih balkanskih susjeda, ona nije niti blizu njima po stupnju organiziranosti države. K tomu, izlaze i brojne informacije o prikrivanju stvarnih brojki oboljelih i umrlih radi nadolazećih izbora. Ipak, Srbija posjeduje izuzetno kvalitetan epidemiološki kadar naslijeđen iz bivše Jugoslavije koji je unatoč raznim državnim manipulacijama spriječio pretvaranje Srbije u novu Španjolsku. Vrlo vjerojatno, puni podaci bi pokvarili stanje koje statistika prikazuje, ali Srbija bi vjerojatno i dalje imala dobre brojke slične svojim organiziranijim Balkanskim susjedima, Mađarska, Hrvatska, Grčka... (<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/comparisonsofallcausemortalitybetweeneuropeancountriesandregions/januarytojune2020#age-standardised-mortality-rates>, rujan 2020.)

5. Zaključak

COVID-19, prvi je veliki, javno univerzalno vidljiv, zalogaj moderne statistike i mašinskog učenja. Već u samim počecima pandemije neki od stečenih znanja pokazali su se nezamjenjivi te su zasigurno spasili tisuće života. Posao ne prestaje te kako bolest bude jenjavala fokus će se sa ispravnog predviđanja ove pandemije prebaciti na kvalitetnije modeliranja budućih epidemija i pandemija.

Ono što se iz rada da zaključiti je da su nominalno ispravni modeli manjkavi te da nisu u mogućnosti ispravno prikazati stvarnost. To ne umanjuje njihovu važnost sve dok su oni naš jedini alat. Ipak, postavlja se pitanje treba li dogmatično prihvaćati istinitost modela, ili treba pokušati ispraviti te modele kako bi u budućnosti mogli donositi bolje odluke, odnosno, kako se te ispravne odluke ne bi mogle propitivati od strane poluupučenih individua.

6. Literatura

Knjige i stručne publikacije

1. Maia Martcheva, *An Introduction to Mathematical Epidemiology*, Springer Science+Business Media, New York 2015.
2. James Holland Jones, *Notes On R_0* , Stanford University, Stanford 2007. (dostupno na: <https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf>)
3. Fred Brauer, Carlos Castillo-Chavez, Zhilan Feng, *Mathematical Models in Epidemiology*, Springer Science+Business Media, New York 2019.
4. Kiri Wagstaf, Claire Cardie, Seth Rogers, Stefan Schroedl, *Constrained K-means Clustering with Background Knowledge*, Cornell University, Ithaca 2001. (dostupno na: <https://web.cse.msu.edu/~cse802/notes/ConstrainedKmeans.pdf>)
5. Ivančica Mirošević, *Algoritam k-sredina*, 2016. (dostupno na: <https://hrcak.srce.hr/file/256969>)
6. Jessica Utts, Robert Heckard, *Mind on Statistics*, Thomson Learning, Duxbury 2007.

Internetski izvori

1. <https://www.healthline.com/health/r-nought-reproduction-number#coronavirus> (rujan 2020.)
2. <https://online.stat.psu.edu/stat462/node/132/> (rujan 2020.)
3. <https://statisticsbyjim.com/regression/log-log-plots/> (rujan 2020.)
4. <https://datascientia.blog/2017/08/14/dss-llrm/> (rujan 2020.)
5. <https://www.statisticshowto.com/power-law/> (rujan 2020.)
6. <https://statweb.stanford.edu/~susan/courses/s60/split/node60.html> (rujan 2020.)
7. <https://www.lifeinnorway.net/coronavirus-in-norway/> (rujan 2020.)
8. <https://dutchreview.com/news/coronavirus-netherlands/> (rujan 2020.)

7. Popis slika

Slika 1. Graf SIR modela

Slika 2. Grafovi infektivne populacije SIR modela

Slika 3. Lloyd algoritam kroz korake na neoznačenim grafovima

Slika 4. Graf sa linearnom regresijom (crveno) skupa podataka (plavo)

Slika 5. Log-log graf pravaca različite potencijske ovisnosti

Slika 6. 68 – 95 – 99.7 pravilo prikazano grafički

Slika 7. Tablični prikaz djeliča korištenih podataka

Slika 8. Log-log graf po modelu za SAD, od 22.01.2019. do 19.09.2020. podjeljen kroz 6 klastera

Slika 9. Obični graf za SAD, od 22.01.2019. do 19.09.2020. sa odgovarajućim točkama obojanim u boju klastera kojemu pripadaju na log-log grafu

Slika 10. Histogram sa objedinjenim vrijednostima RMSD-a za statistički relevantne države

Slika 11. Grafovi za Norvešku

Slika 12. Dno tablice gustoća populacije u Europi

Slika 13. Grafovi za Nizozemsku

Slika 14. Grafovi za Srbiju

8. Dodatak 1 – kod

```
import math
import itertools
import pandas as pd
import numpy as np
from datetime import date, timedelta, datetime
from bokeh.plotting import figure, output_file, show
from bokeh.layouts import row
from bokeh.palettes import Set1_9 as palette
from bokeh.models.formatters import DatetimeTickFormatter
from bokeh.models import Span, Slope
from sklearn.cluster import KMeans
import scipy.special
from scipy.stats import norm

colors = itertools.cycle(palette)

confirmed_global = pd.DataFrame()
deaths_global = pd.DataFrame()
recovered_global = pd.DataFrame()
first_date = ""
last_date = ""
affected_countries = set()

first_date_sample = ""
last_date_sample = ""
affected_countries_sample = set()

bottom_limit_sample = 500

def make_plot(title, hist, edges, avg, x, mu, sigma):
    p = figure(title=title, tools='', background_fill_color="#fafafa",
              y_axis_type="linear")
    p.quad(top=hist, bottom=0, left=edges[:-1], right=edges[1:],
           fill_color="navy", line_color="white", alpha=0.5)
    vline = Span(location=avg, dimension='height', line_color='red', line_width=1)
    p.line([], [], legend_label=str(round(avg,4)), line_color='red',
           line_width=1)
    pdf = norm.pdf(x, mu, sigma)
    p.line(x, pdf, line_color="#ff8888", line_width=3, alpha=0.7,
           legend_label="PDF")

    p.y_range.start = 0
    p.xaxis.axis_label = 'RMSD'
    p.yaxis.axis_label = 'DenCountries'
    p.grid.grid_line_color="white"
    p.renderers.extend([vline])

    return p
```



```

def bulk_print(clusters=1):
    set_filter_sample(5000)
    rmsd_avg = 0
    rmsd_bank = []
    for i in affected_countries_sample:
        print(i)
        temp = kmeans_regression(i, clusters)
        rmsd_bank.append(temp)
        rmsd_avg = rmsd_avg + temp
    rmsd_avg = rmsd_avg / len(affected_countries_sample)
    print("RMSD average: " + str(rmsd_avg))

    hist, edges = np.histogram(rmsd_bank, density=True, bins=50)

    # Normal Distribution
    mu = np.mean(rmsd_bank)
    sigma = np.std(rmsd_bank)
    print(sigma)
    x = np.linspace(min(rmsd_bank), max(rmsd_bank), 100)

    p1 = make_plot("RMSD histogram", hist, edges, rmsd_avg, x, mu, sigma)
    show(p1)

def kmeans_regression(country="World", clusters=1):
    title = country + " | " + first_date_sample.strftime("%m-%d-%Y") +
        " to " + last_date_sample.strftime("%m-%d-%Y")
    delta1 = (first_date_sample - first_date).days + 1
    delta2 = (last_date_sample - first_date).days + 2
    if (country == "World"):
        confirmed_temp = confirmed_global.sum()
        deaths_temp = deaths_global.sum()
        recovered_temp = recovered_global.sum()
        del confirmed_temp['Country/Region']
        del deaths_temp['Country/Region']
        del recovered_temp['Country/Region']
        confirmed = confirmed_temp.values.tolist()[delta1:delta2]
        deaths = deaths_temp.values.tolist()[delta1:delta2]
        recovered = recovered_temp.values.tolist()[delta1:delta2]
    else:
        confirmed = confirmed_global.iloc[confirmed_global.index
            [confirmed_global['Country/Region'] == country],
            delta1:delta2].values.tolist()[0]
        deaths = deaths_global.iloc[deaths_global.index
            [deaths_global['Country/Region'] == country],
            delta1:delta2].values.tolist()[0]
        recovered = recovered_global.iloc[recovered_global.index
            [recovered_global['Country/Region'] == country],
            delta1:delta2].values.tolist()[0]

    i = 0
    last_confirmed = 0
    log_sum_new = []
    log_sum_active = []

```

```

all_sum_active = []
new_scatter = []
active_scatter = []
color_list = []
while i < len(confirmed):
    sum_new = confirmed[i] - last_confirmed
    sum_active = confirmed[i] - deaths[i] - recovered[i]
    all_sum_active.append(sum_active)
    if (sum_new > 1 and sum_active > 1):
        log_sum_new.append(math.log(sum_new))
        log_sum_active.append(math.log(sum_active))
        new_scatter.append(math.log(sum_new))
        active_scatter.append(math.log(sum_active))
    else:
        new_scatter.append(0)
        active_scatter.append(0)
    color_list.append(0)
    last_confirmed = confirmed[i]
    i += 1
data = np.vstack([log_sum_active, log_sum_new]).T
model = KMeans(n_clusters=clusters, random_state=0).fit(data)
y = model.labels_
output_file("plot.html")
p = figure(title=title+" | lognew/logactive covid-19",
           x_axis_label="log of active cases",
           y_axis_label="log of new cases")
p1 = figure(title=title+" | cases/date covid-19",
            x_axis_type='datetime', x_axis_label="date",
            y_axis_label="active cases")
rmsd_table = []
for i in range(clusters):
    cluster_i = data[y == i, :]
    cluster_zip = list(zip(*cluster_i))
    this_color = next(colors)
    i = 0
    while i < len(cluster_zip[0]):
        j = 0
        while j < len(new_scatter):
            if(active_scatter[j] == cluster_zip[0][i] and
               new_scatter[j] == cluster_zip[1][i]):
                color_list[j] = this_color
            j += 1
        i += 1

    if(len(cluster_zip[0]) > 1):
        coef = np.polyfit(cluster_zip[0], cluster_zip[1], 1)
        slope = coef[0]
        rmsd_table.append([len(cluster_zip[0]), slope])
        yp = np.polyval(coef, cluster_zip[0])

```

```

        p.line(cluster_zip[0], yp, color=this_color,
              legend_label=str(slope))
        p.scatter(cluster_zip[0], cluster_zip[1], color=this_color)
    else:
        slope = 0
        p.scatter(cluster_zip[0], cluster_zip[1], color=this_color,
              legend_label=str(slope))

p.legend.location = 'top_left'
p.legend.title = 'Slope coefficient'

rmsd = 0
rmsd_time = 0
for i in rmsd_table:
    rmsd = rmsd + i[0] * pow((1 - i[1]), 2)
    rmsd_time = rmsd_time + i[0]
rmsd = pow((rmsd / rmsd_time), 0.5)
print("RMSD (of power law): " + str(rmsd))

datelist = pd.date_range(start=first_date_sample,
                        end=last_date_sample).tolist()
p1.xaxis.formatter = DatetimeTickFormatter(days="%d-%m-%Y")
p1.scatter(datelist, confirmed, color=color_list)

show(row(p, p1))
return rmsd

def import_data():
    url_confirmed_global =
'https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_
confirmed_global.csv'
    url_deaths_global =
'https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_
deaths_global.csv'
    url_recovered_global =
'https://raw.githubusercontent.com/CSSEGISandData/COVID-
19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_
recovered_global.csv'
    global confirmed_global, deaths_global, recovered_global
    confirmed_global = pd.read_csv(url_confirmed_global)
    deaths_global = pd.read_csv(url_deaths_global)
    recovered_global = pd.read_csv(url_recovered_global)
    del confirmed_global['Province/State'], confirmed_global['Lat'],
        confirmed_global['Long']
    del deaths_global['Province/State'], deaths_global['Lat'],
        deaths_global['Long']
    del recovered_global['Province/State'], recovered_global['Lat'],
        recovered_global['Long']

```

```

global first_date, last_date
first_date = datetime.strptime(confirmed_global.columns[1],
                              "%m/%d/%y").date()
last_date = datetime.strptime(confirmed_global.columns[-1],
                              "%m/%d/%y").date()
confirmed_global = confirmed_global
                    .groupby(['Country/Region'], as_index=False).sum()
deaths_global = deaths_global
                .groupby(['Country/Region'], as_index=False).sum()
recovered_global = recovered_global
                  .groupby(['Country/Region'], as_index=False).sum()
global affected_countries
affected_countries = set(confirmed_global['Country/Region'].unique())
global first_date_sample, last_date_sample, affected_countries_sample
first_date_sample = first_date
last_date_sample = last_date
affected_countries_sample = affected_countries

def info_data():
    print("Imported all COVID-19 cases from " +
          first_date.strftime("%m-%d-%Y") + " to " +
          last_date.strftime("%m-%d-%Y"))

def set_filter_sample(bottom=0):
    global bottom_limit_sample, affected_countries_sample
    if (bottom > 0):
        bottom_limit_sample = bottom
        delta1 = (first_date_sample - first_date).days + 1
        delta2 = (last_date_sample - first_date).days + 1
        confirmed_global_sample = pd.concat([confirmed_global.iloc[:, 0],
                                             confirmed_global.iloc
                                            [:, delta1:delta2]], axis=1)

        remove_countries = set()
        for i in affected_countries:
            sum_country = confirmed_global_sample
                          .loc[confirmed_global_sample['Country/Region'] == i,
                               confirmed_global_sample.columns[-1]].sum()
            if (sum_country <= bottom_limit_sample):
                remove_countries.add(i)
        affected_countries_sample = affected_countries - remove_countries

def reset_filter_sample():
    global bottom_limit_sample
    bottom_limit_sample = 500
    set_filter_sample()

def set_date_span(days=0):
    global first_date_sample, last_date_sample

```

```

if (days == 0):
    span_start_input = input("Start date of span(mm-dd-yyyy): ")
    span_end_input = input("End date of span(m m-dd-yyyy): ")
    span_start = datetime.strptime(span_start_input, "%m-%d-%Y").date()
    span_end = datetime.strptime(span_end_input, "%m-%d-%Y").date()
else:
    span_start = last_date - timedelta(days=days)
    span_end = last_date
if (span_start < first_date):
    span_start = first_date
if (span_end > last_date):
    span_end = last_date
if (span_start >= span_end):
    return
first_date_sample = span_start
last_date_sample = span_end
set_filter_sample()

def reset_date_span():
    global first_date_sample, last_date_sample
    first_date_sample = first_date
    last_date_sample = last_date
    set_filter_sample()

def info_sample():
    print("Sample includes all affected countries in the span of " +
          first_date_sample.strftime("%m-%d-%Y") + " to " +
          last_date_sample.strftime("%m-%d-%Y"))
    print("NOTE: countries with less than " + str(bottom_limit_sample) +
          "total cases in dataset excluded")
    print(sorted(affected_countries_sample))

```

9. Dodatak 2 – bulk grafovi RMSD

Skup grafova i vrijednosti RMSD za sve statistički relevantne države u rasponu od 22.01.2019. do 19.09.2020. sa podacima podjeljenima na 6 klastera i minimalnim brojem slučajeva od 5000.

Izrađeno podacima sa poveznice:

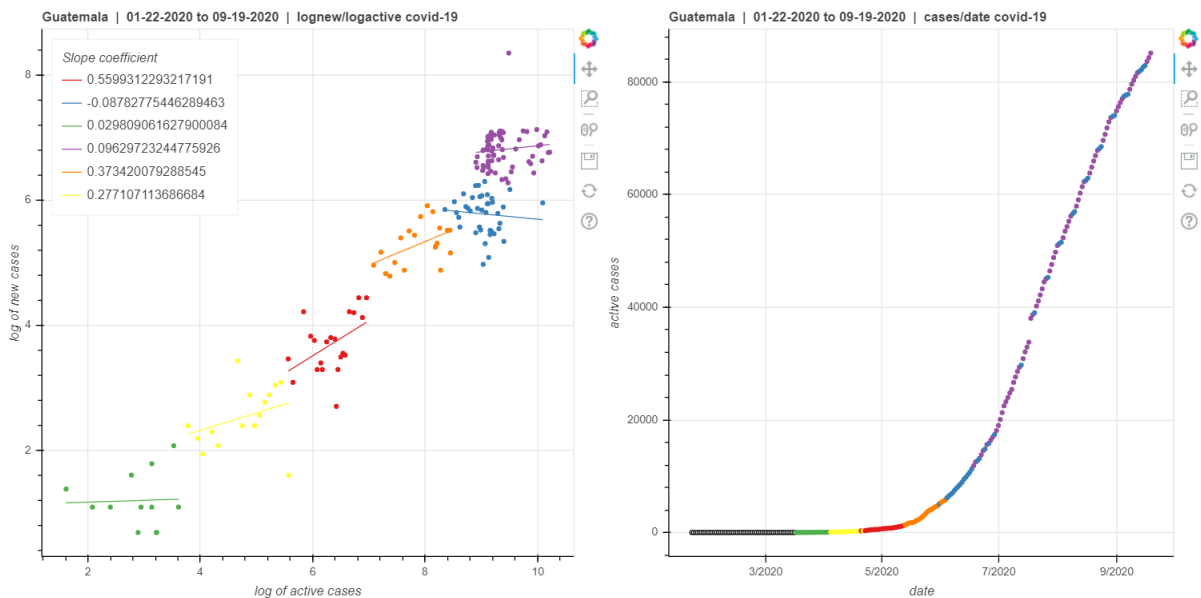
https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

Korištene datoteke sa poveznice:

[time_series_covid19_confirmed_global](#)

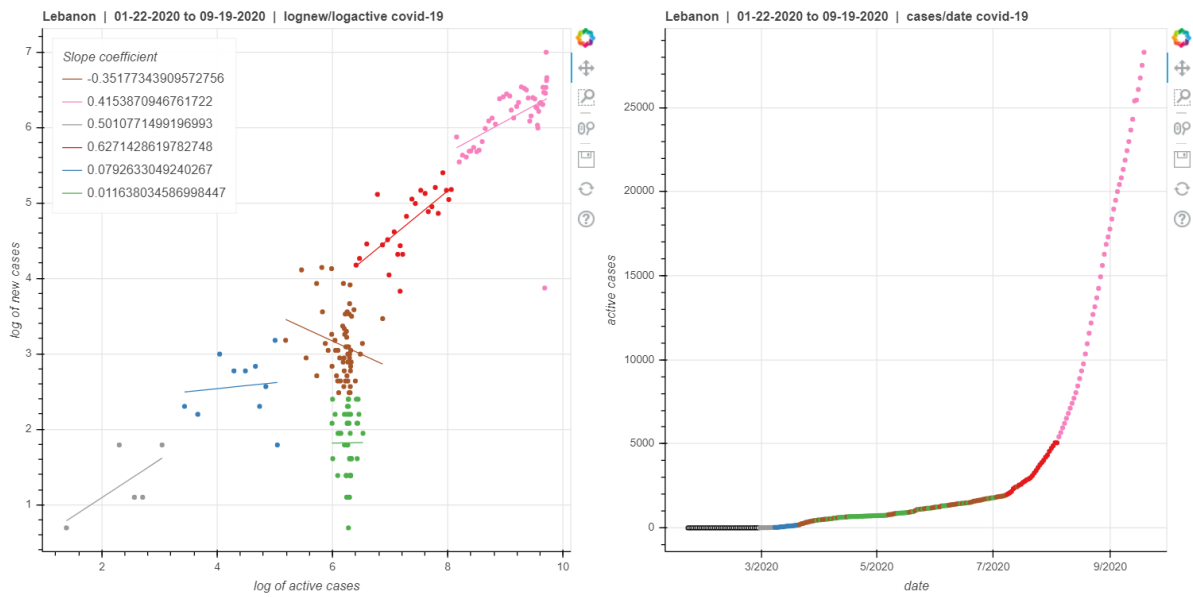
[time_series_covid19_deaths_global](#)

[time_series_covid19_recovered_global](#)



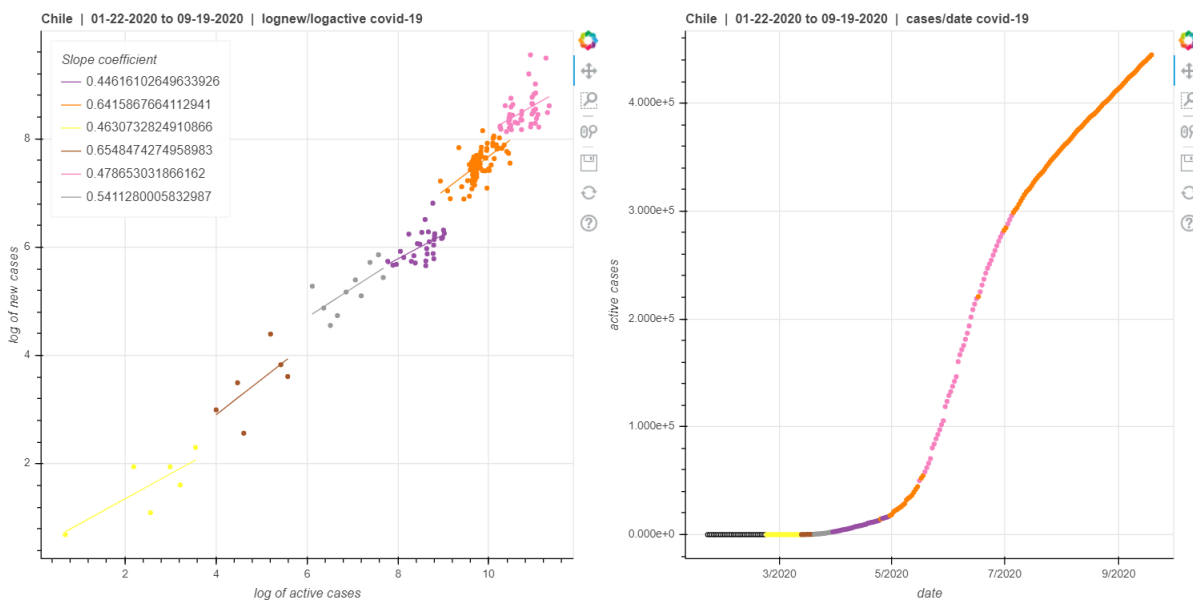
Slika 1. Gvatemala (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Gvatemala): 0.8724871098721656



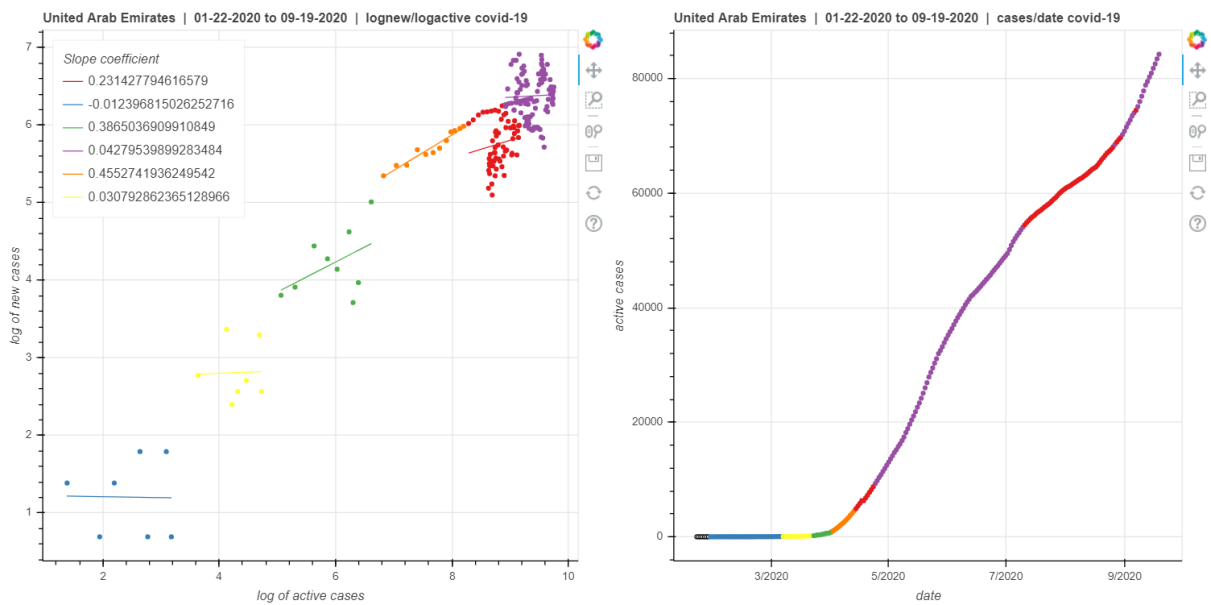
Slika 2. Libanon (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Libanon): 0.977705079293903



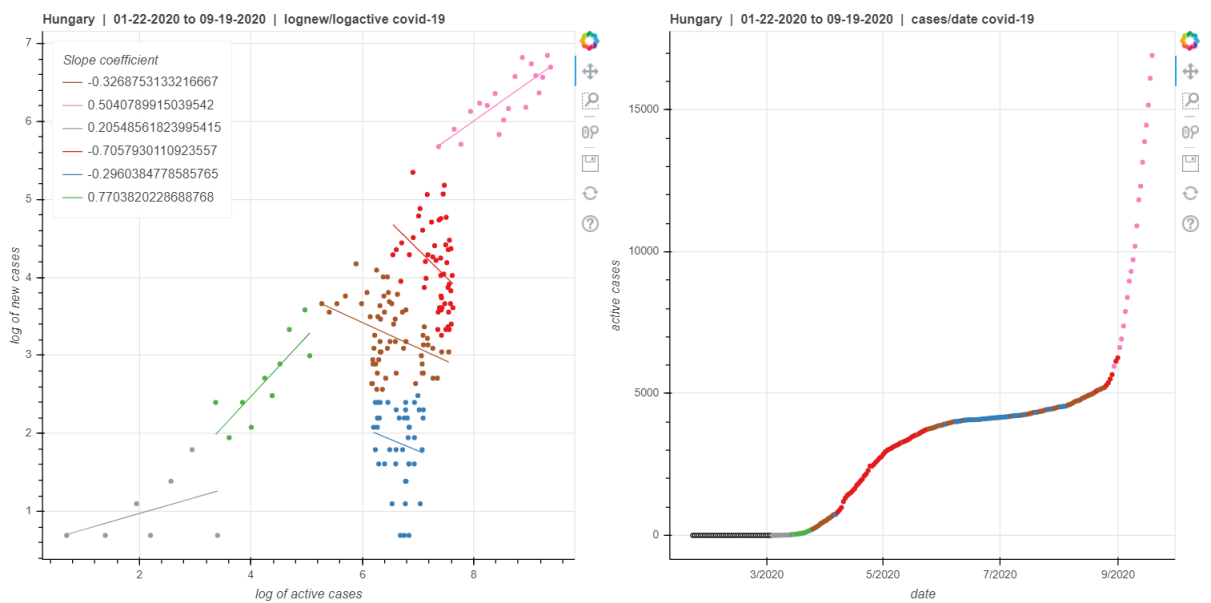
Slika 3. Čile (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Čile): 0.44517993317559834



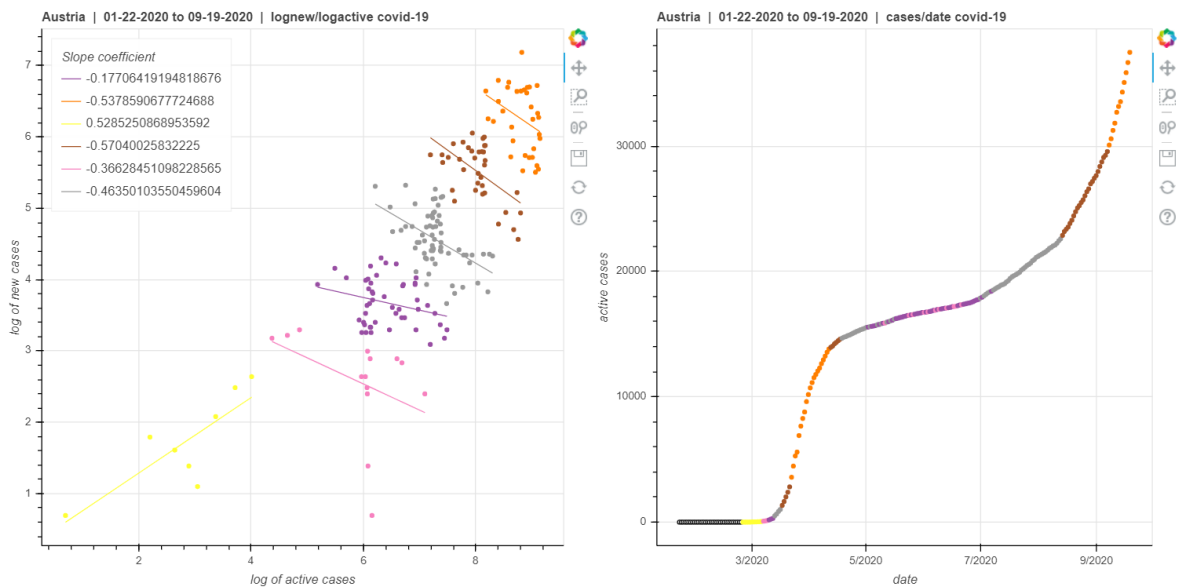
Slika 4. Ujedinjeni Arapski Emirati (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Ujedinjeni Arapski Emirati): 0.8695542901616287



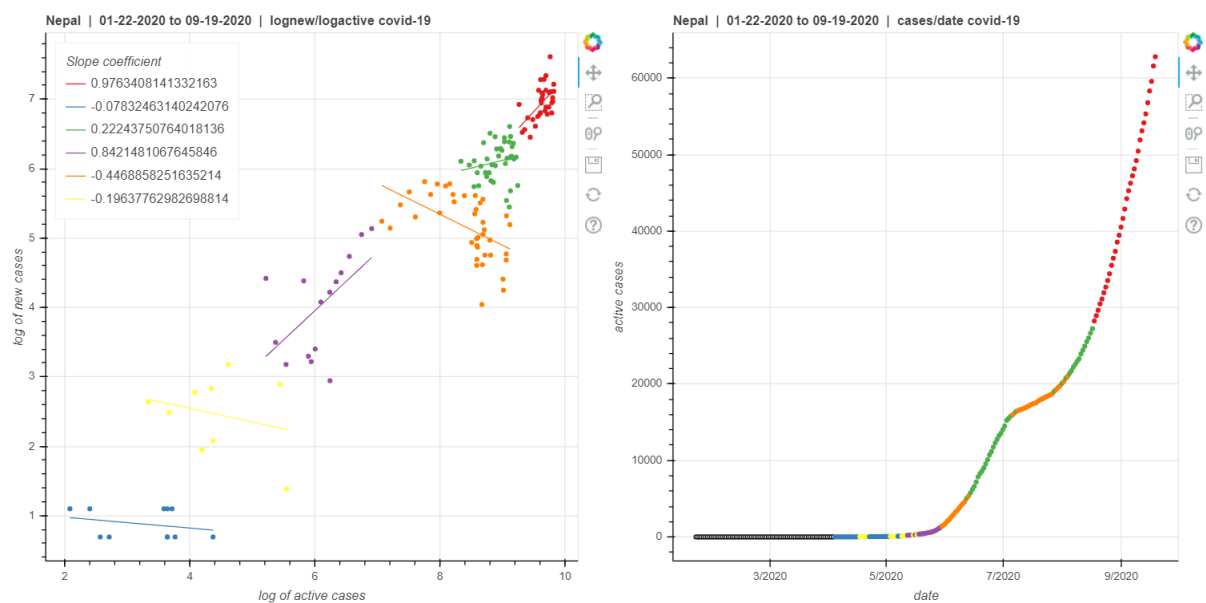
Slika 5. Mađarska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Mađarska): 1.3316424671926463



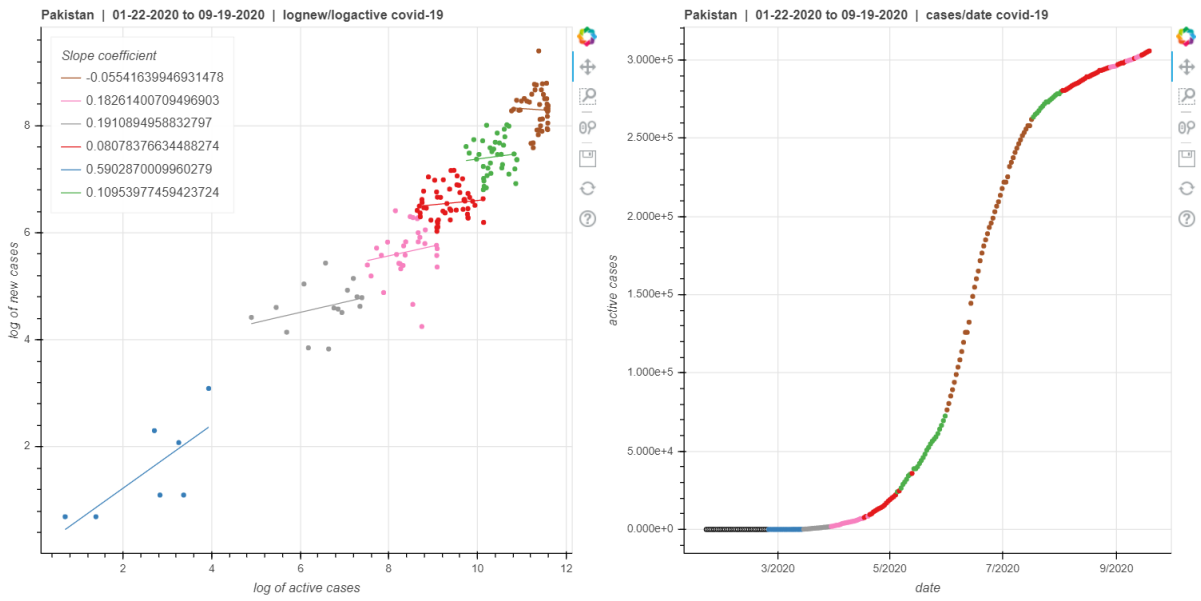
Slika 6. Austrija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Austrija): 1.4013995576708227



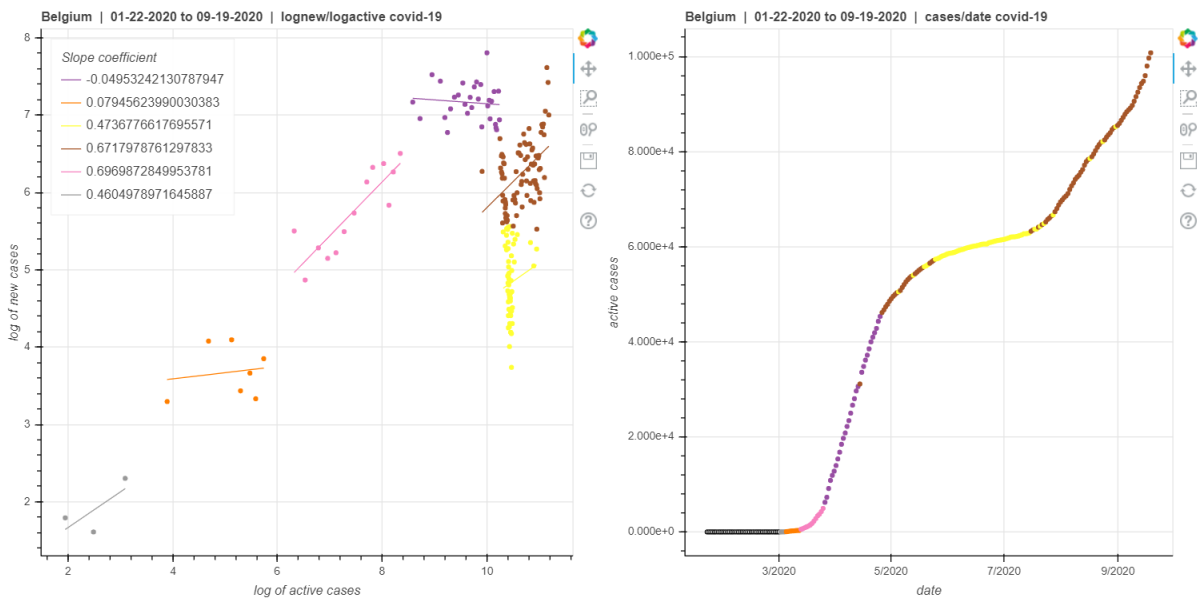
Slika 7. Nepal (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Nepal): 0.9514853612860603



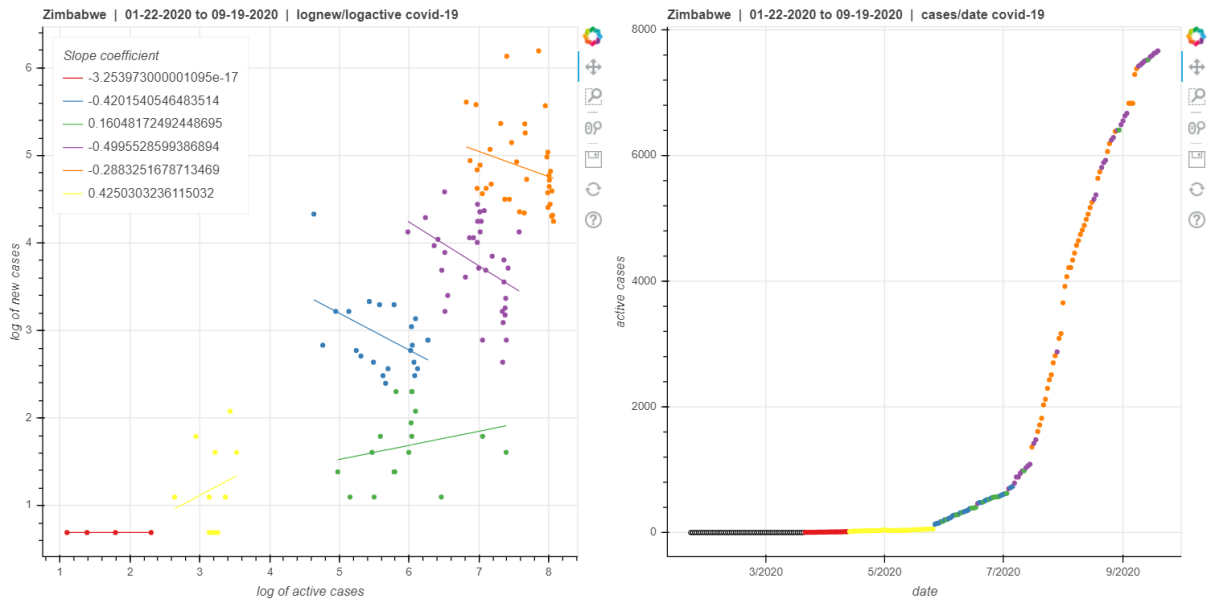
Slika 8. Pakistan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Pakistan): 0.9102588463184484



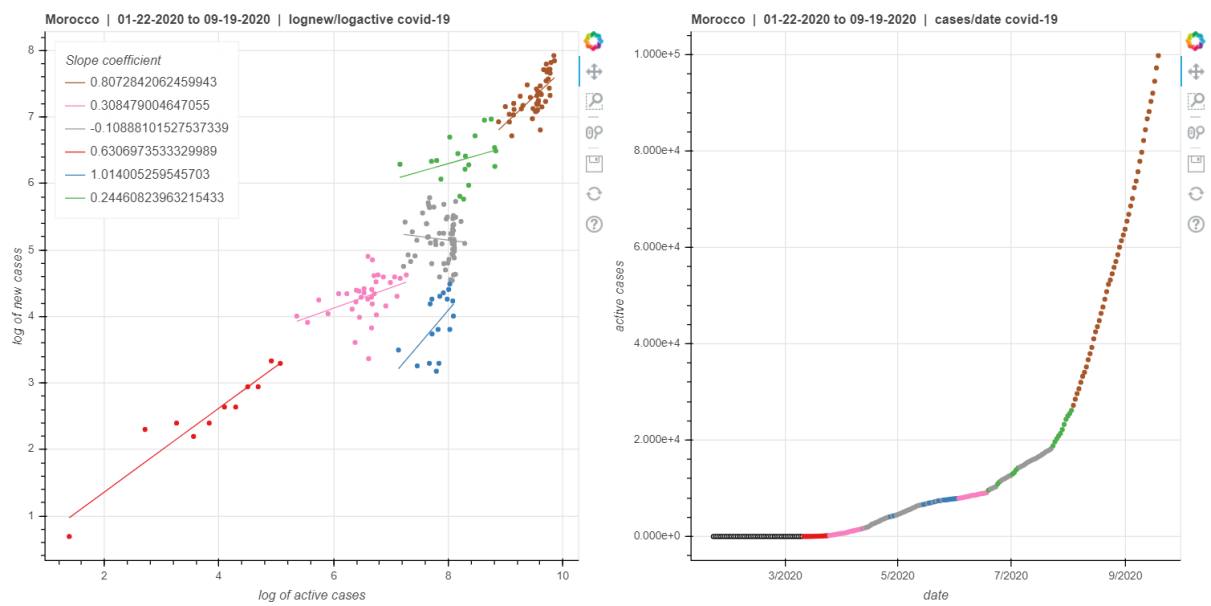
Slika 9. Belgija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Belgija): 0.5829444629246975



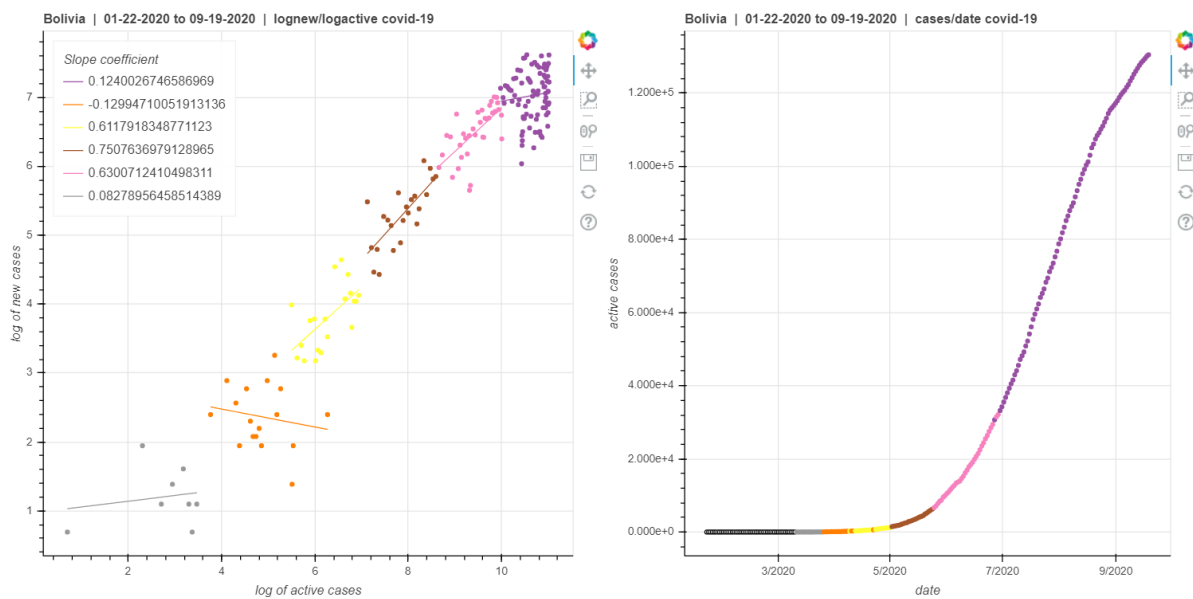
Slika 10. Zimbabve (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Zimbabve): 1.2706199408918064



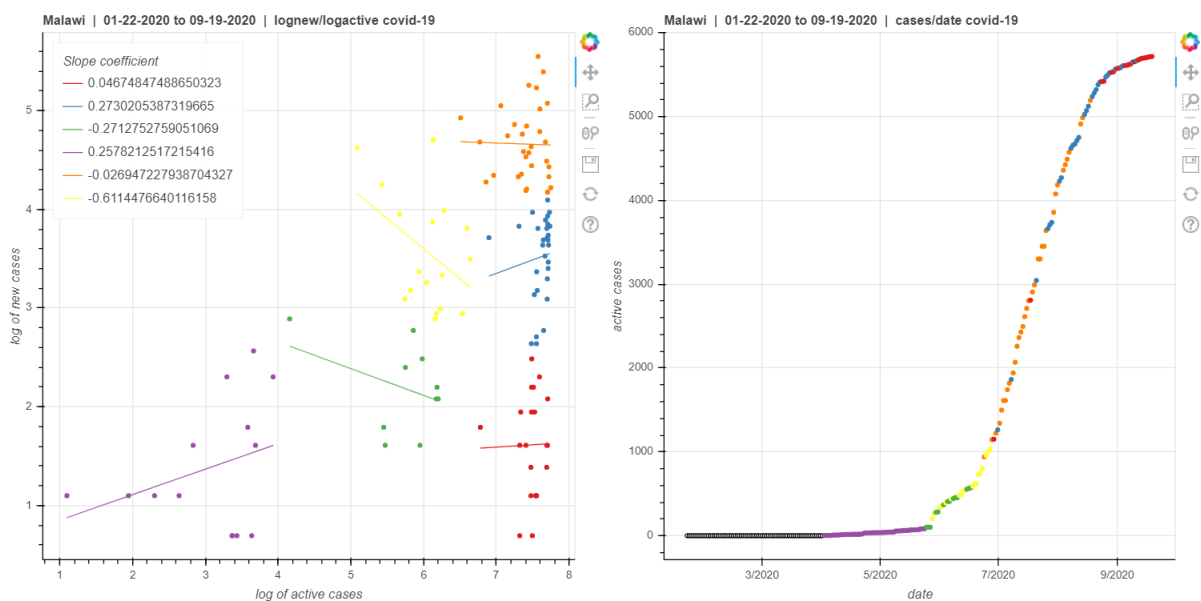
Slika 11. Maroko (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Maroko): 0.7486602790565945



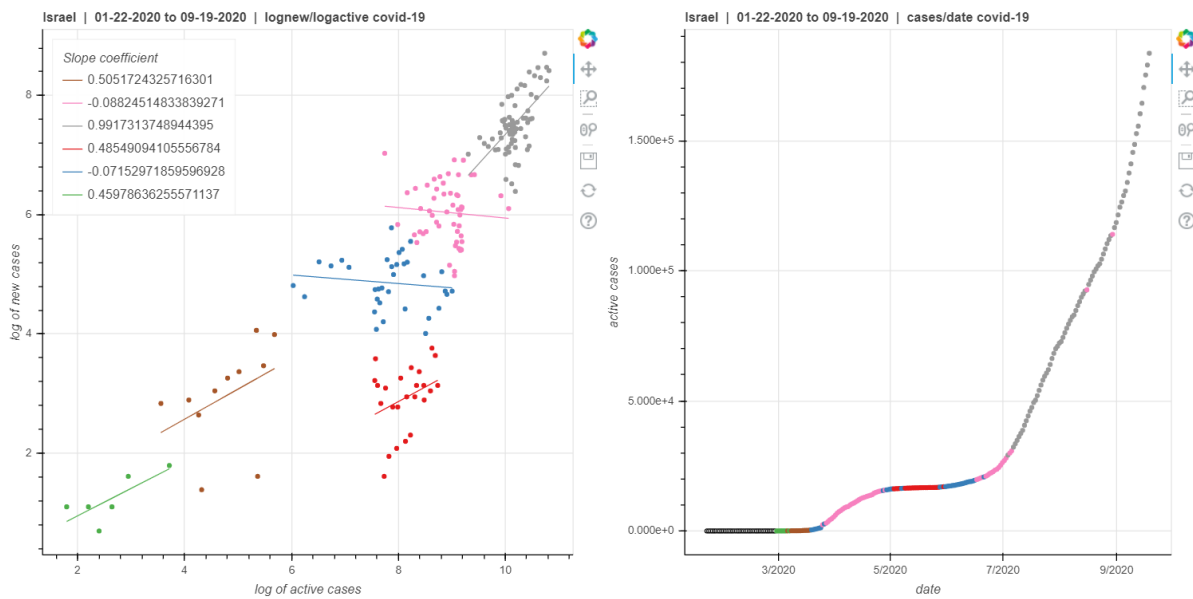
Slika 12. Bolivija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bolivija): 0.7382039526686109



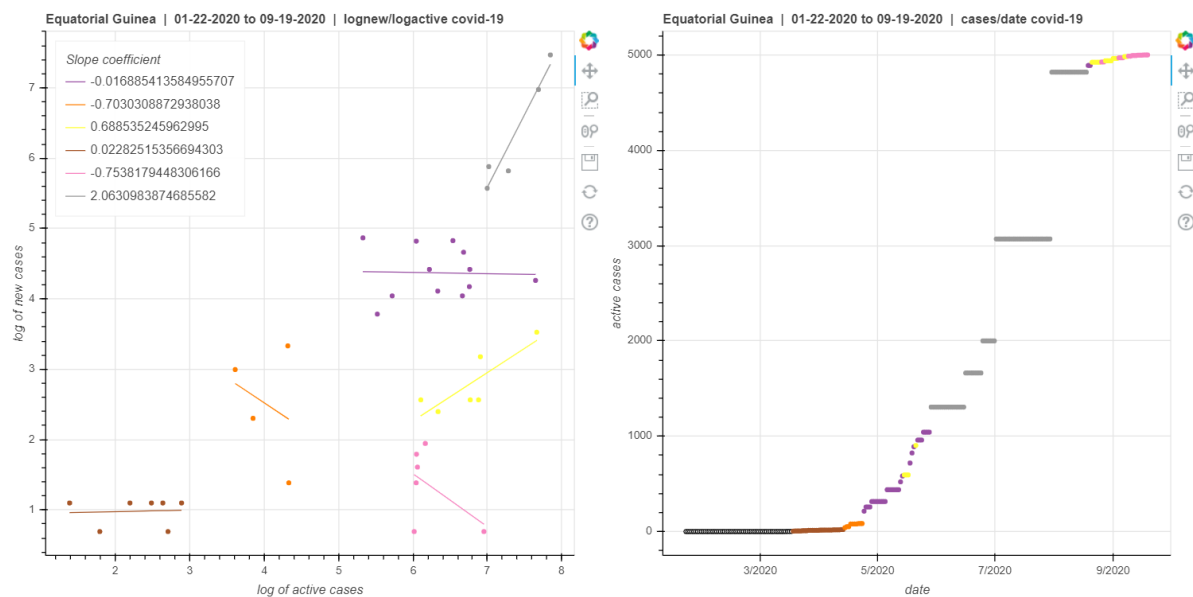
Slika 13. Malavi (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Malavi): 1.0521939583020223



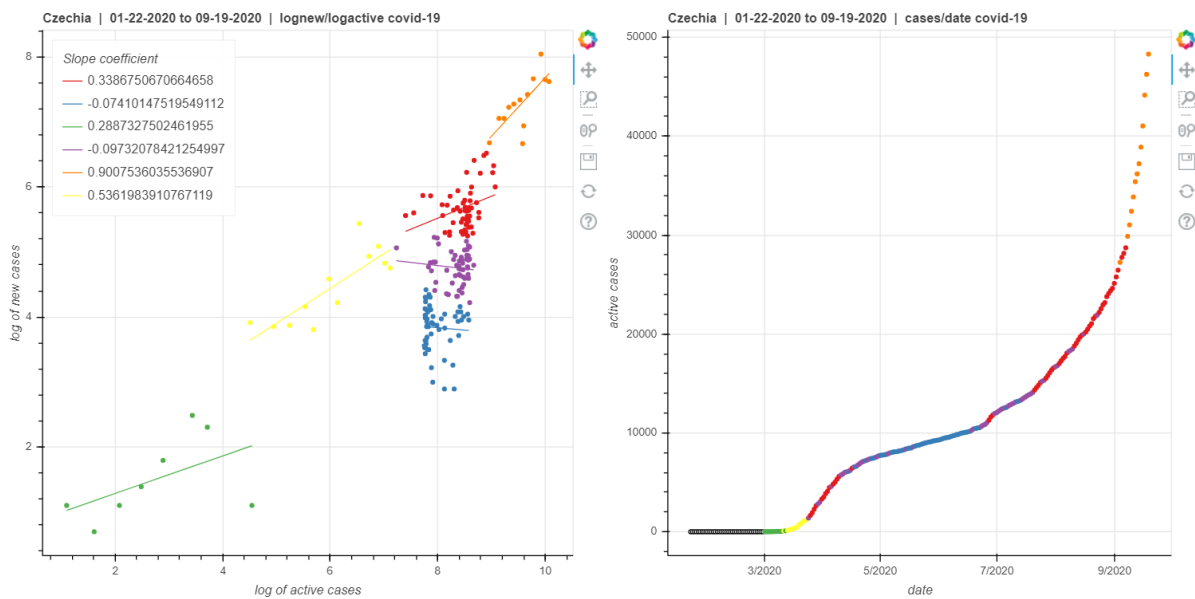
Slika 14. Gvatemala (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Israel): 0.7466340337568887



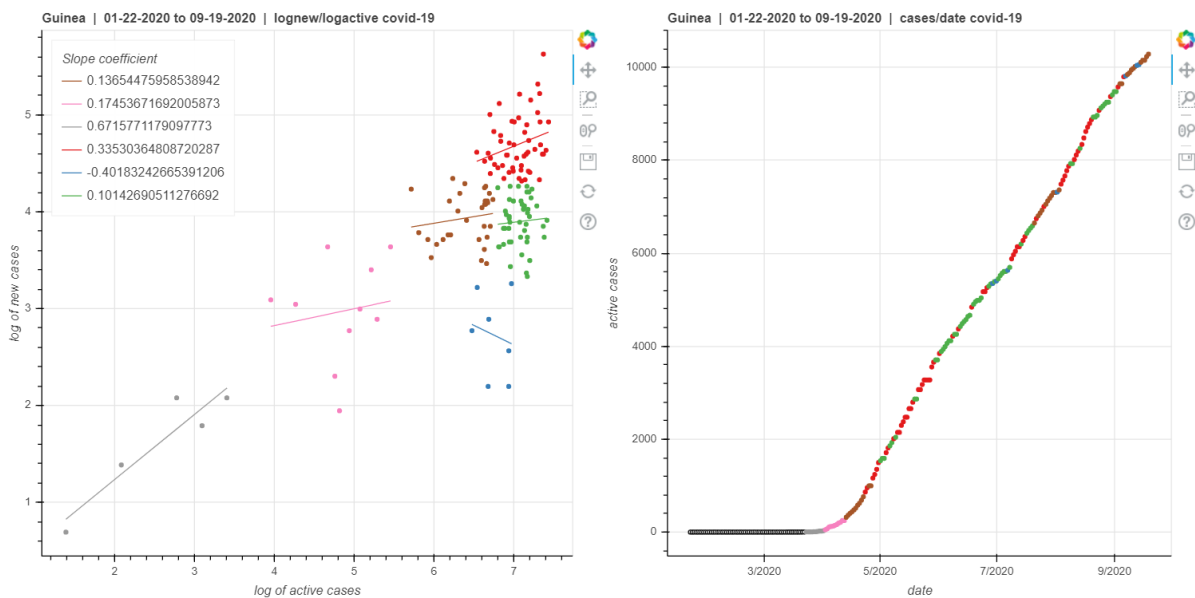
Slika 15. Ekvatorijalna Gvineja (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Ekvatorijalna Gvineja): 1.1766715756218245



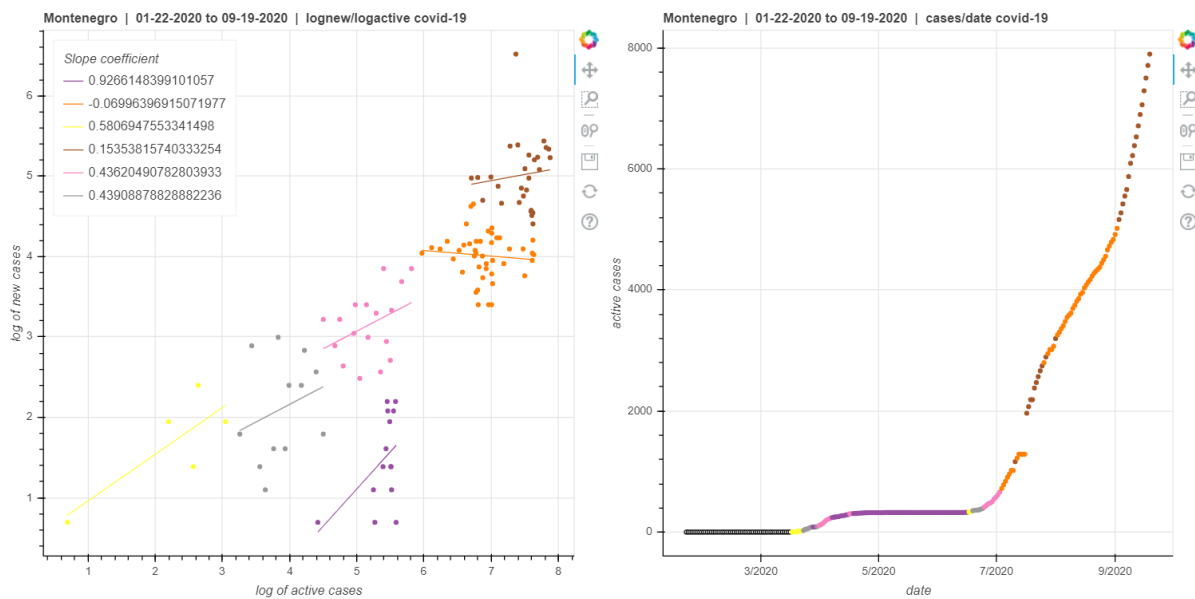
Slika 16. Češka (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Češka): 0.9002258324181088



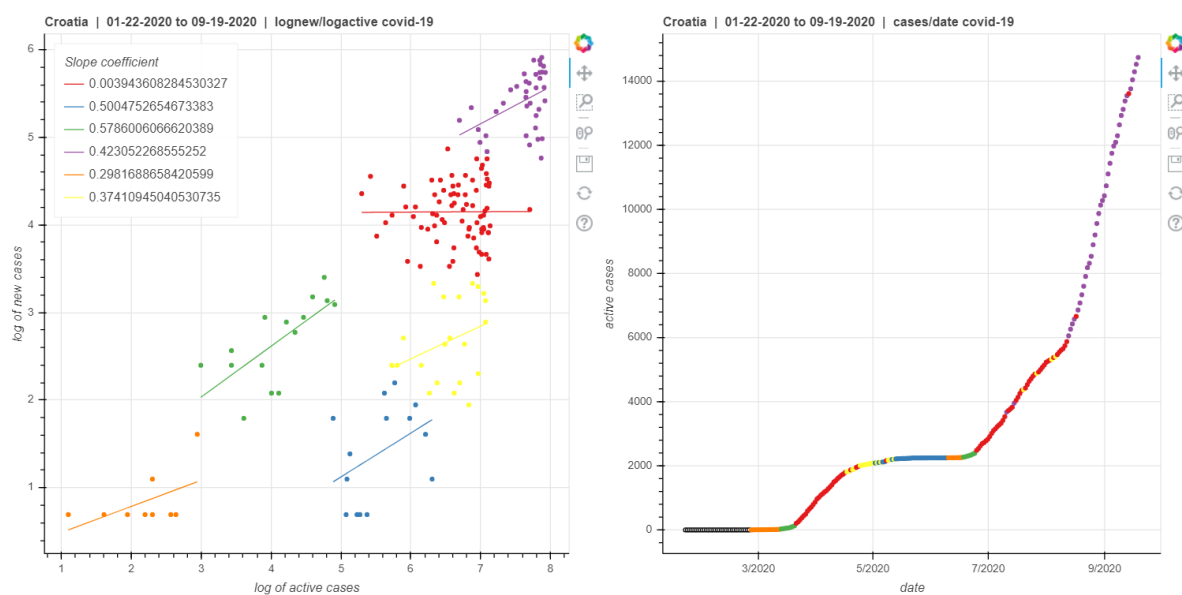
Slika 17. Gvineja (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Gvineja): 0.8312521364122792



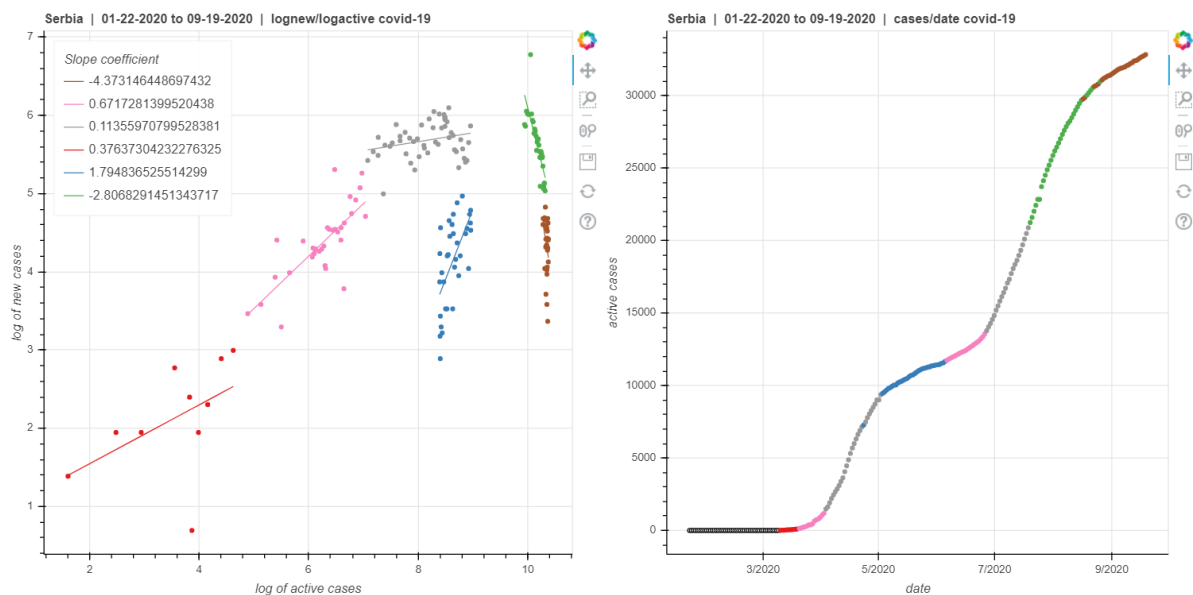
Slika 18. Crna Gora (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Crna Gora): 0.8193889057047675



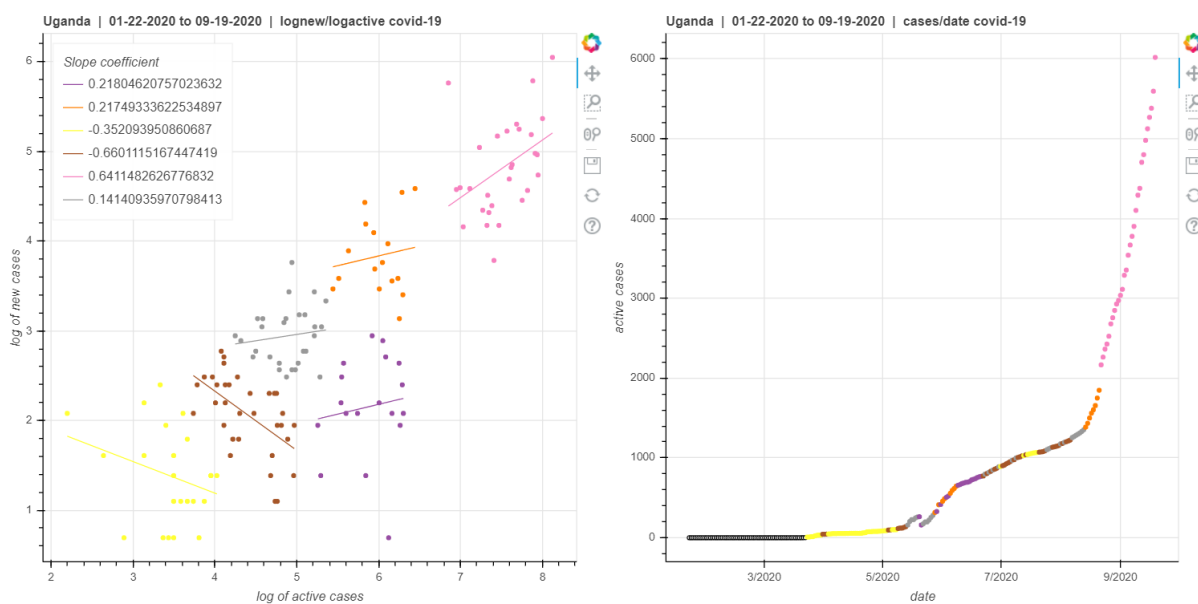
Slika 19. Hrvatska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Hrvatska): 0.7859584029981042



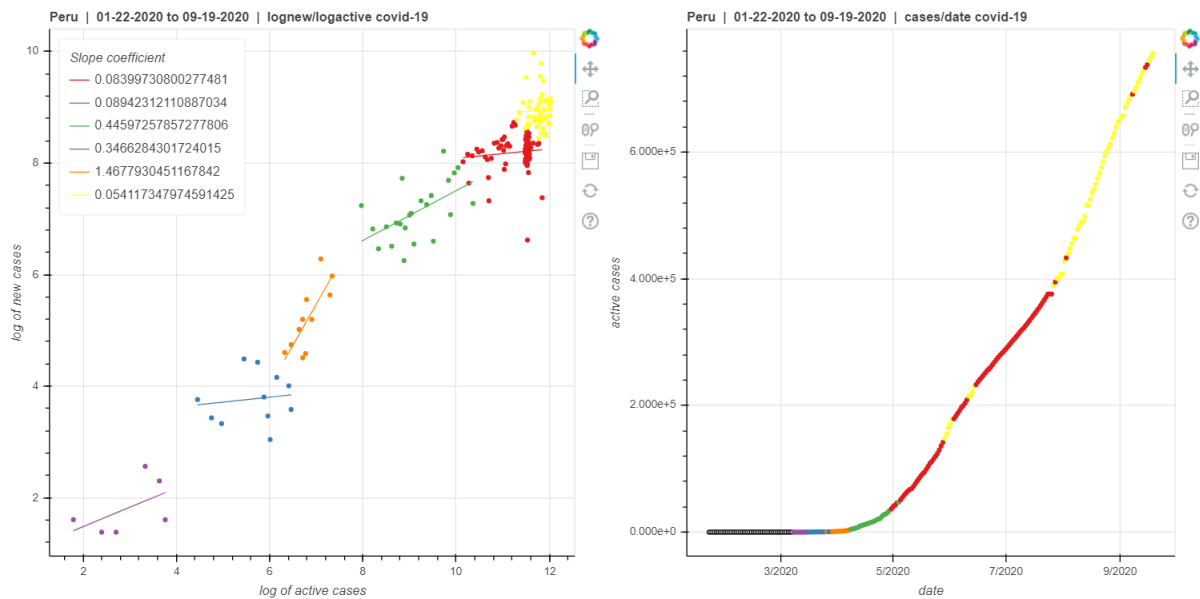
Slika 20. Srbija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Srbija): 2.685340670184227



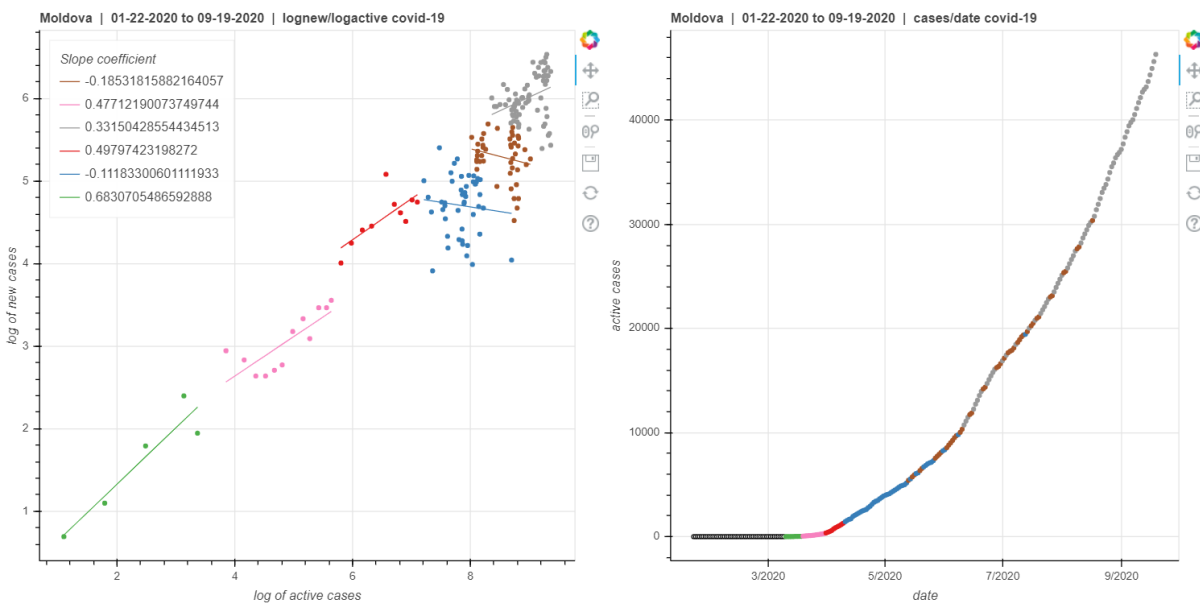
Slika 21. Uganda (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Uganda): 1.1009374323703291



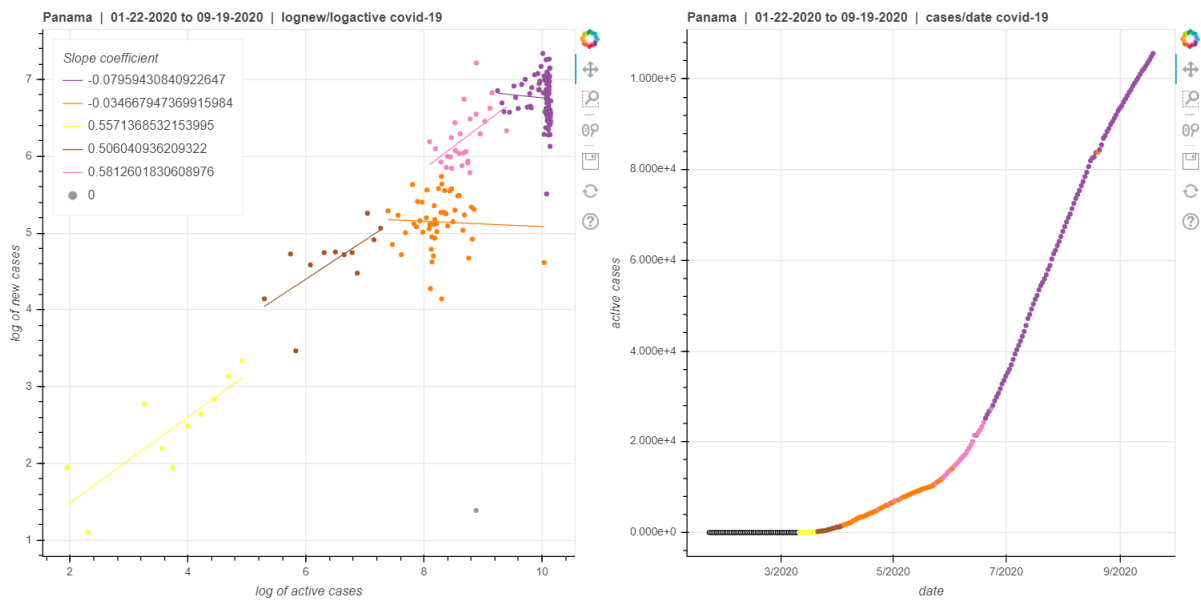
Slika 22. Peru (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Peru): 0.8585786495670994



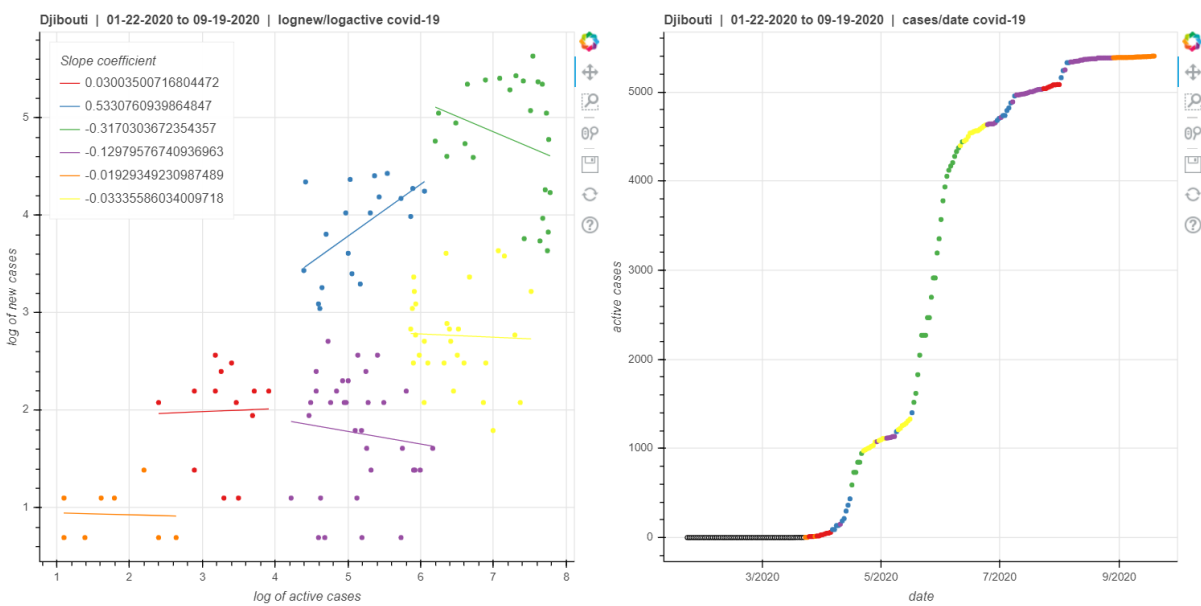
Slika 23. Moldavija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Moldavija): 0.9135447730920664



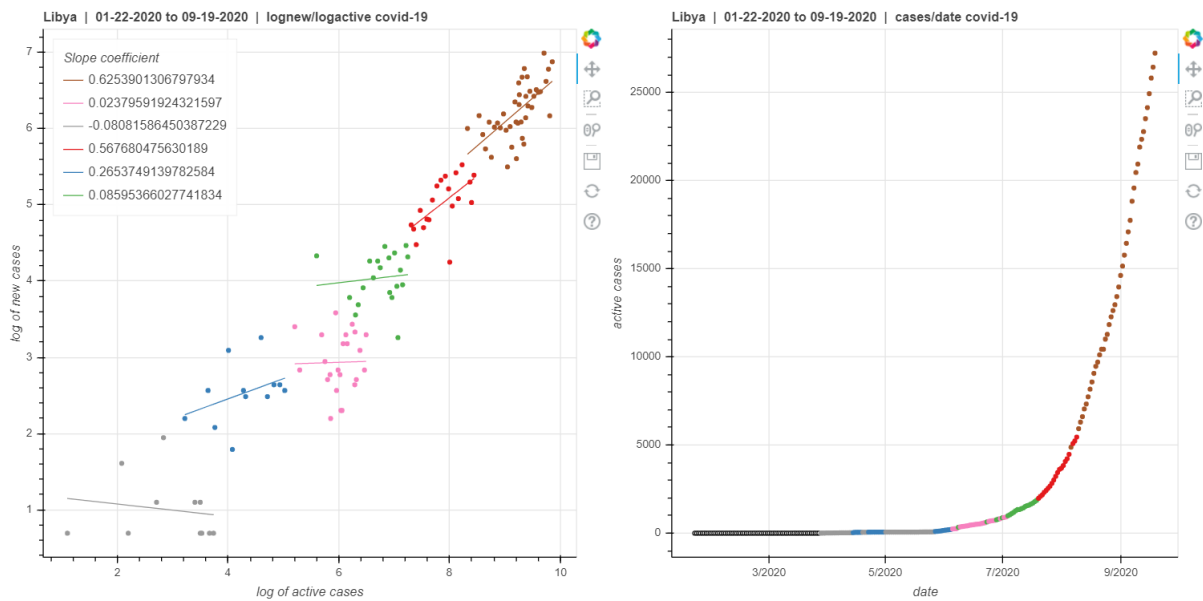
Slika 24. Panama (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Panama): 0.9424306085012564



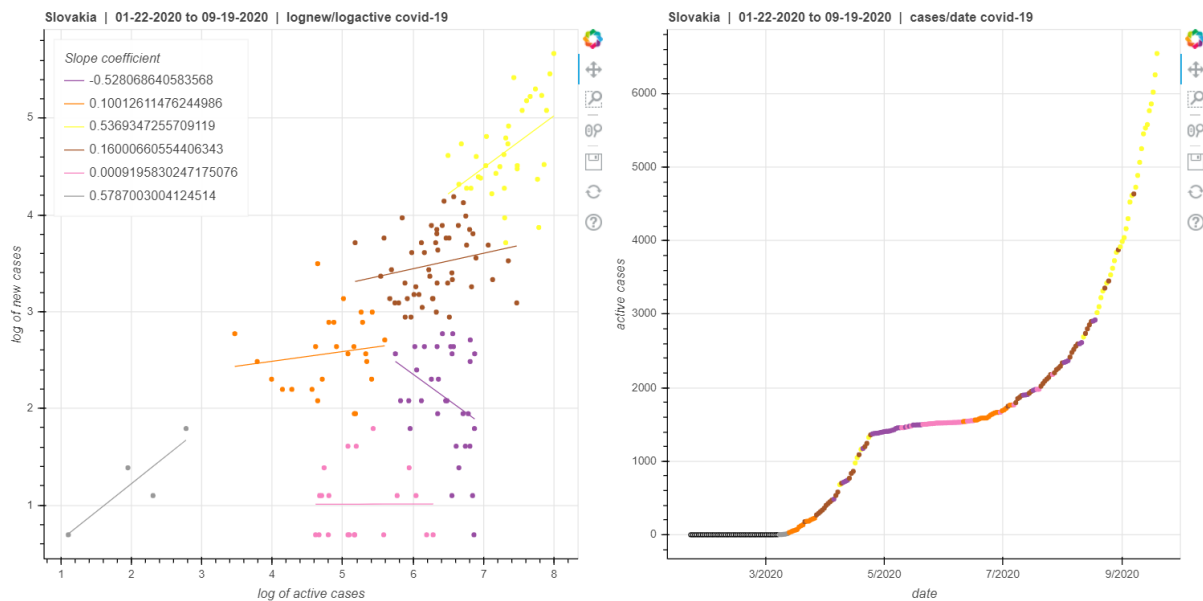
Slika 25. Džibuti (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Džibuti): 1.0541945904030214



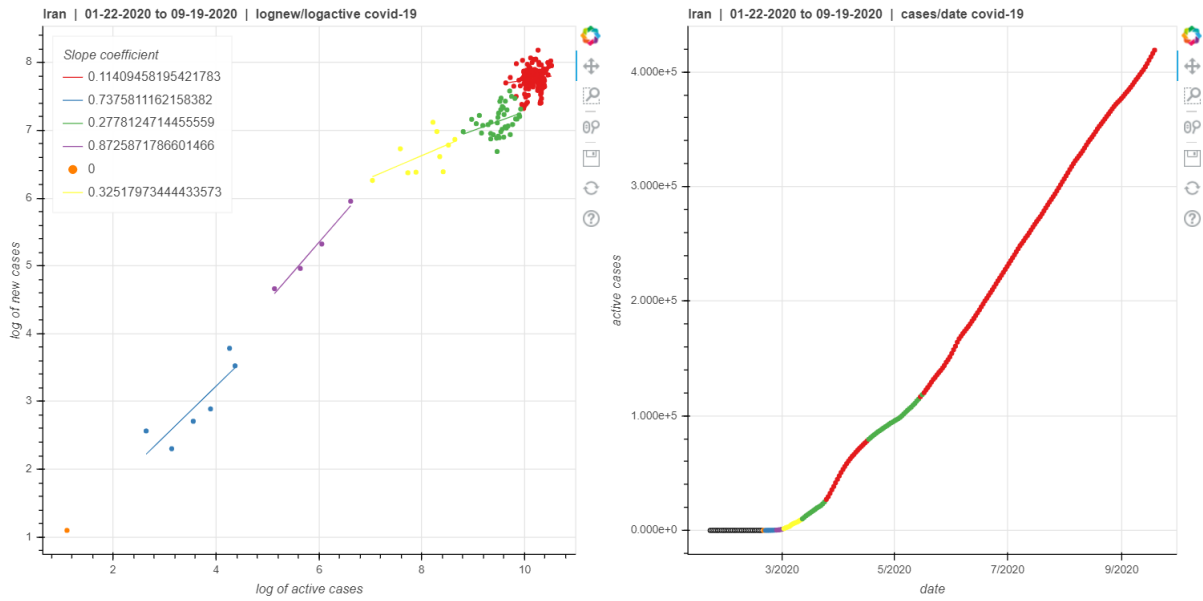
Slika 26. Libija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Libija): 0.7287436292582825



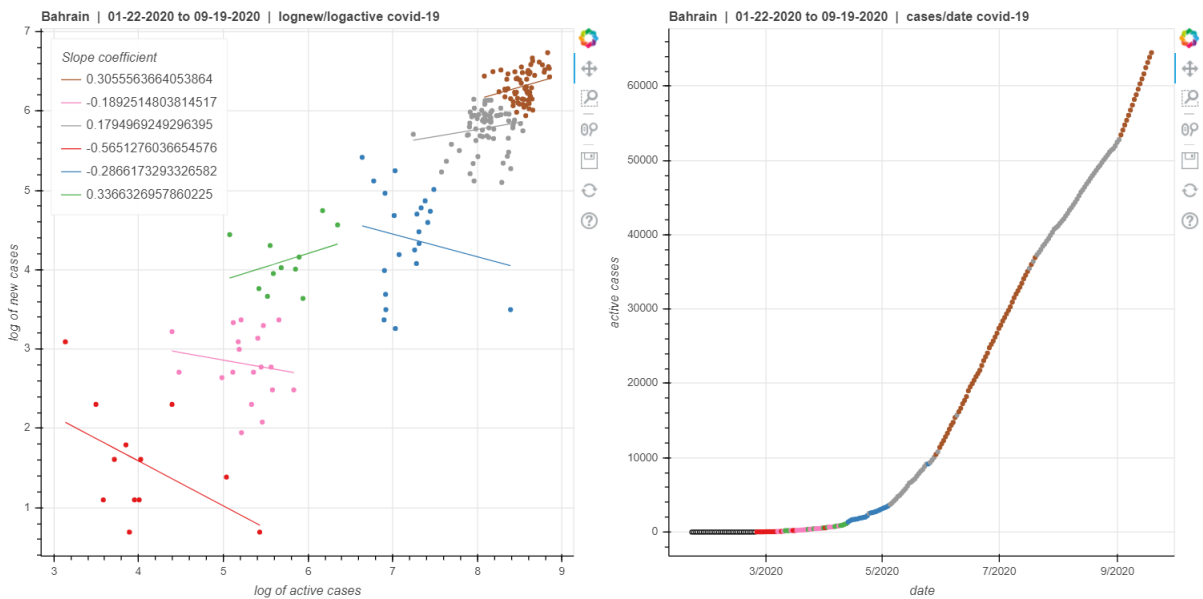
Slika 27. Slovačka (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Slovačka): 0.9784206862087258



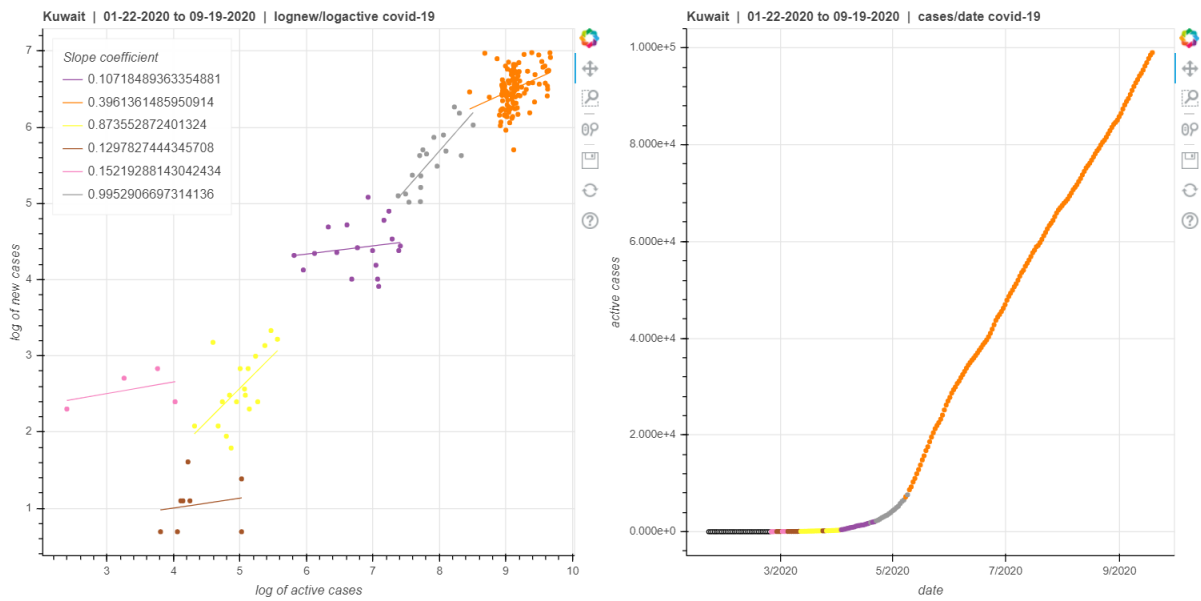
Slika 28. Iran (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Iran): 0.826054232671932



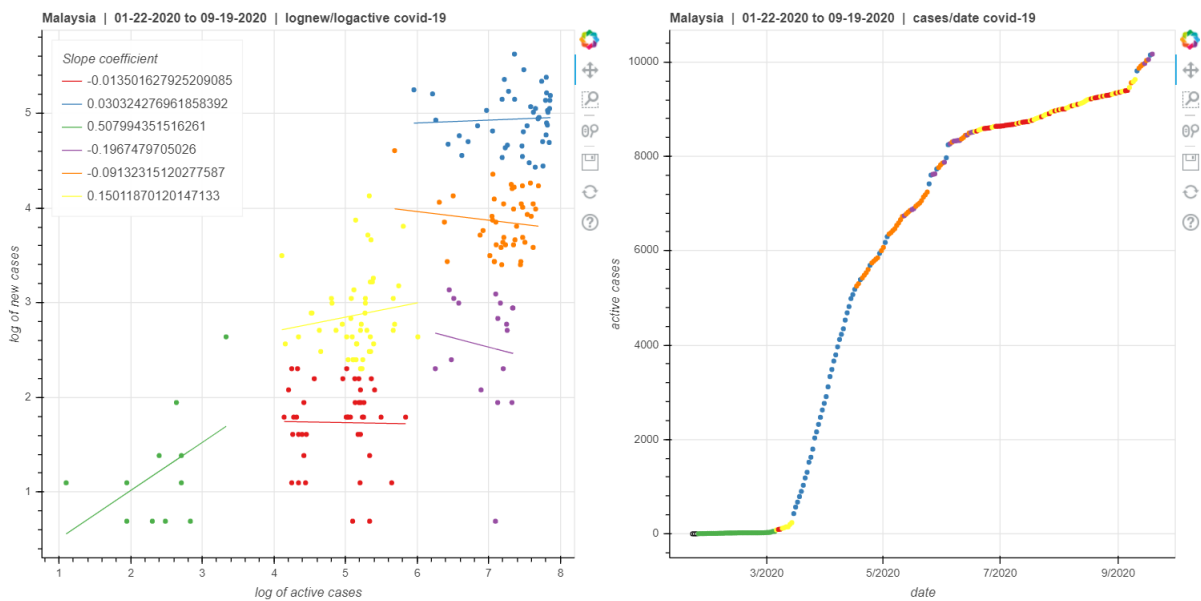
Slika 29. Bahrein (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bahrein): 0.938261177749893



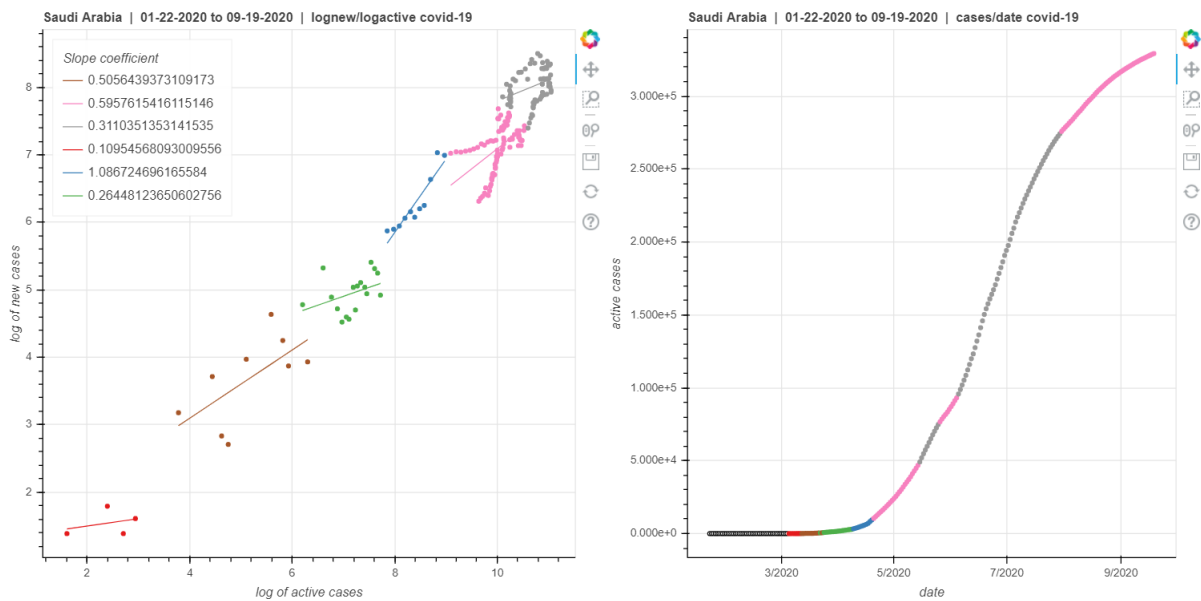
Slika 30. Kuvajt (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kuvajt): 0.6018013695531942



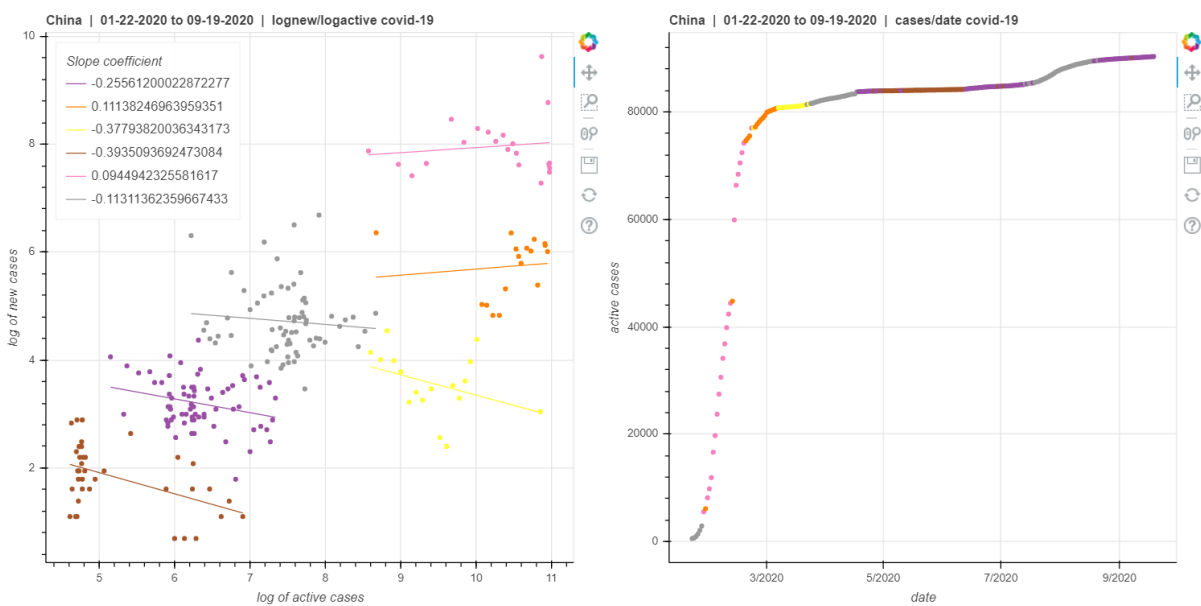
Slika 31. Malazija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Malazija): 0.9812488889397668



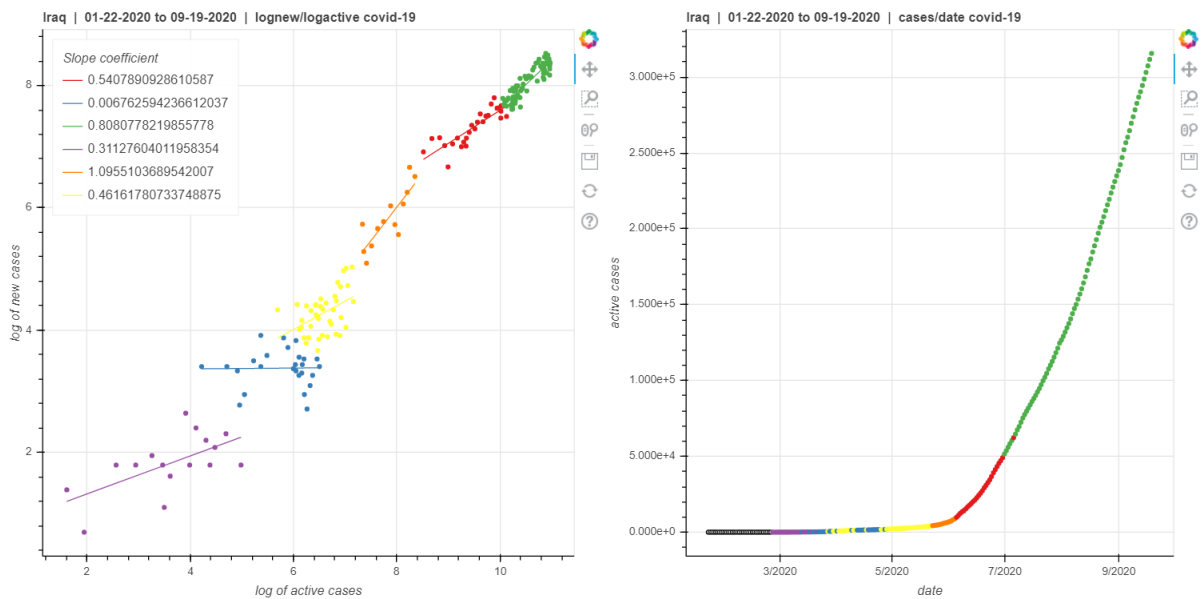
Slika 32. Saudijska Arabija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Saudijska Arabija): 0.5584720838212048



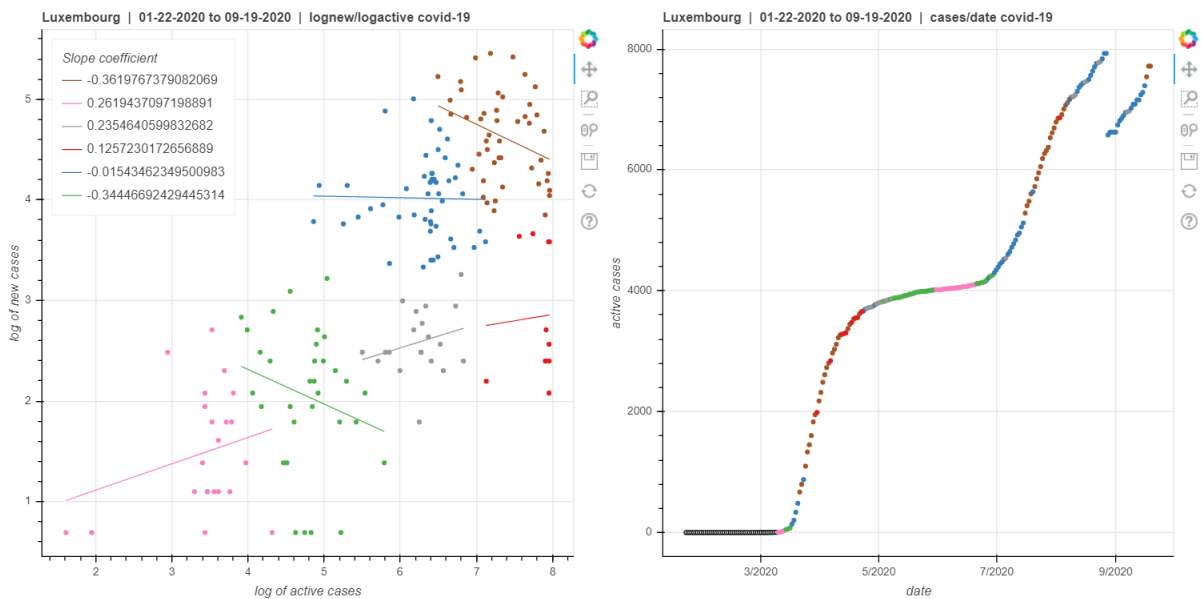
Slika 33. Kina (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kina): 1.1979793830640355



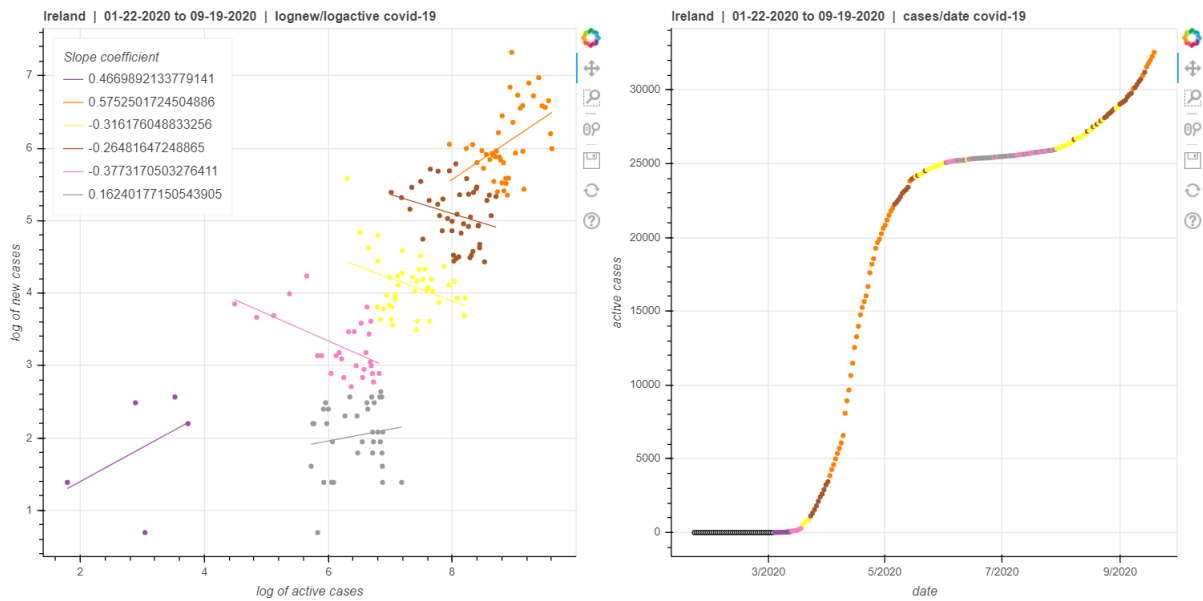
Slika 34. Irak (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Irak): 0.5150801878695836



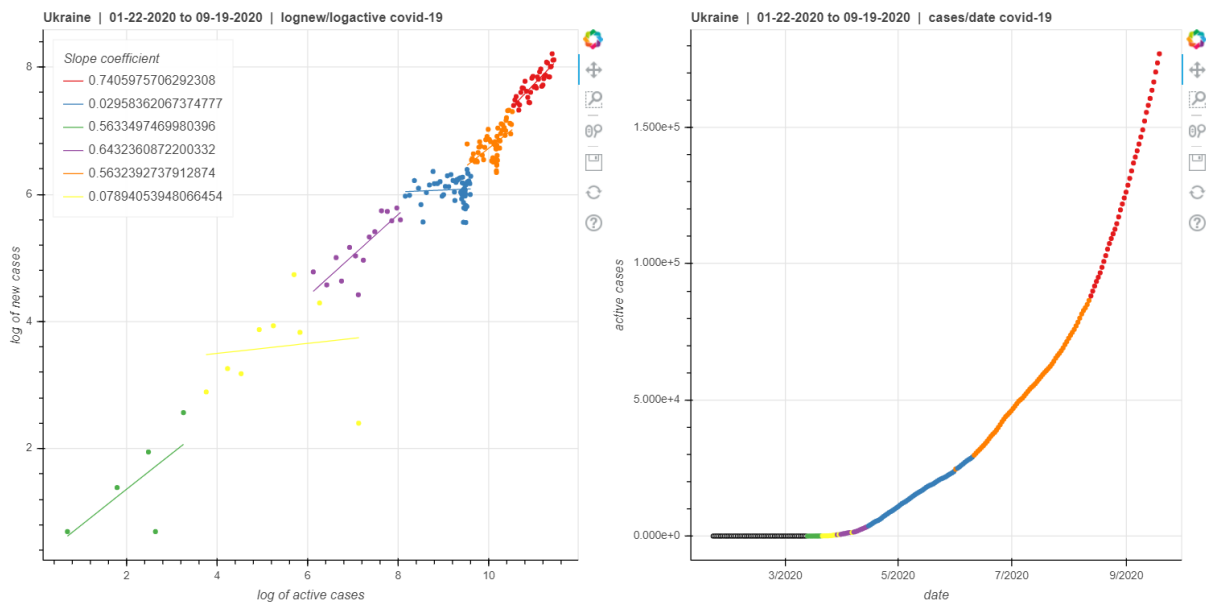
Slika 35. Luksemburg (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Luksemburg): 1.119966662961902



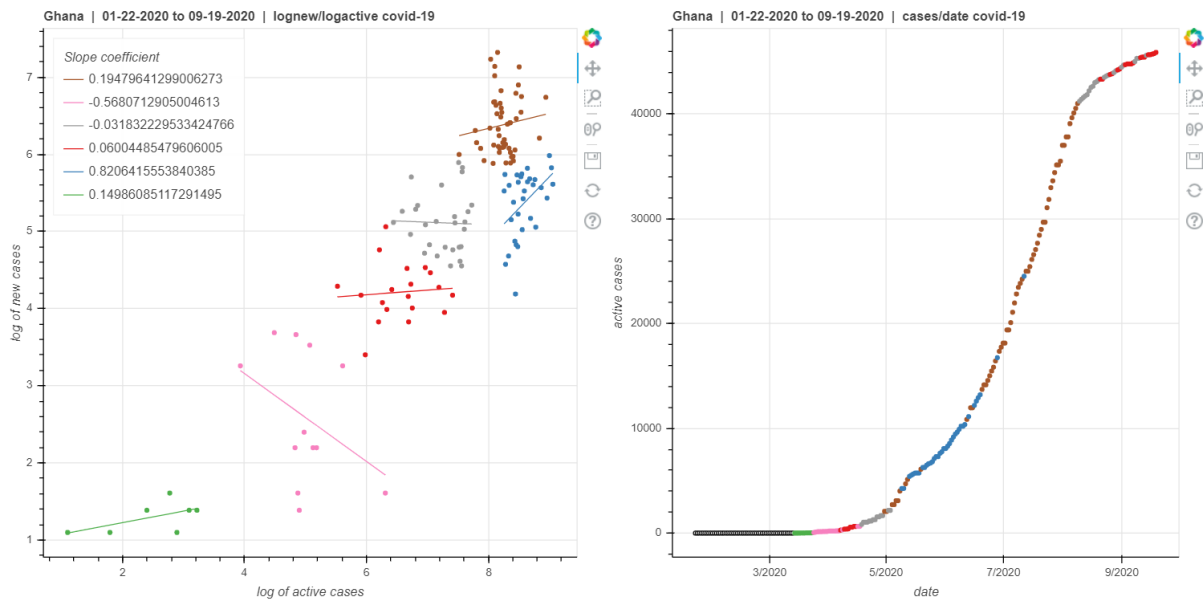
Slika 36. Irska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Irska): 1.0836349585624792



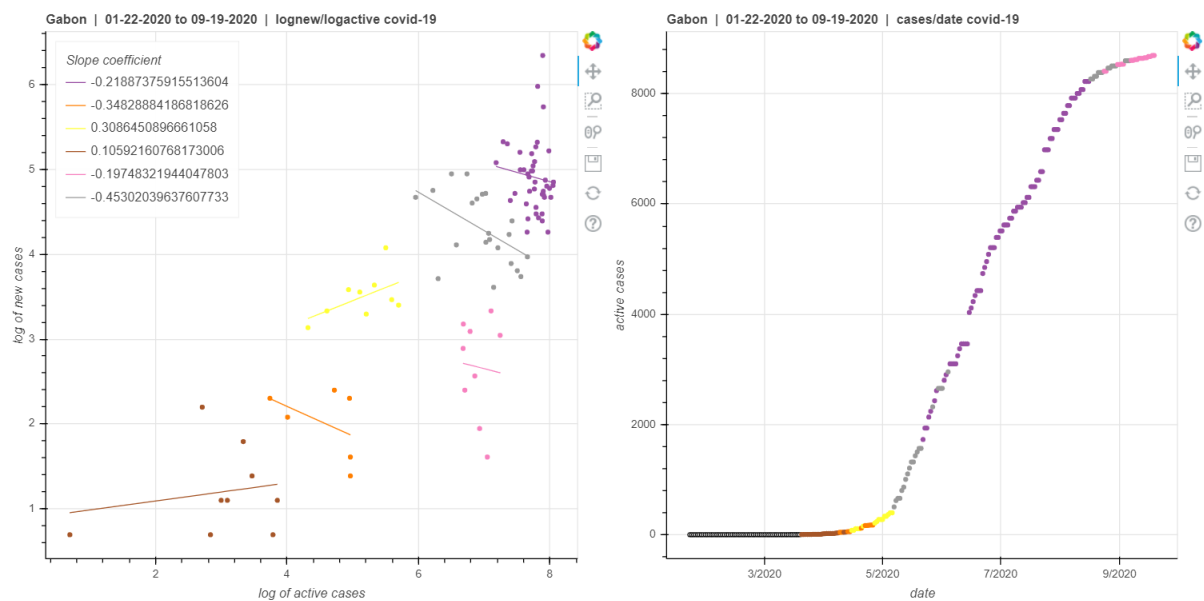
Slika 37. Ukrajina (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Ukrajina): 0.6458380357940374



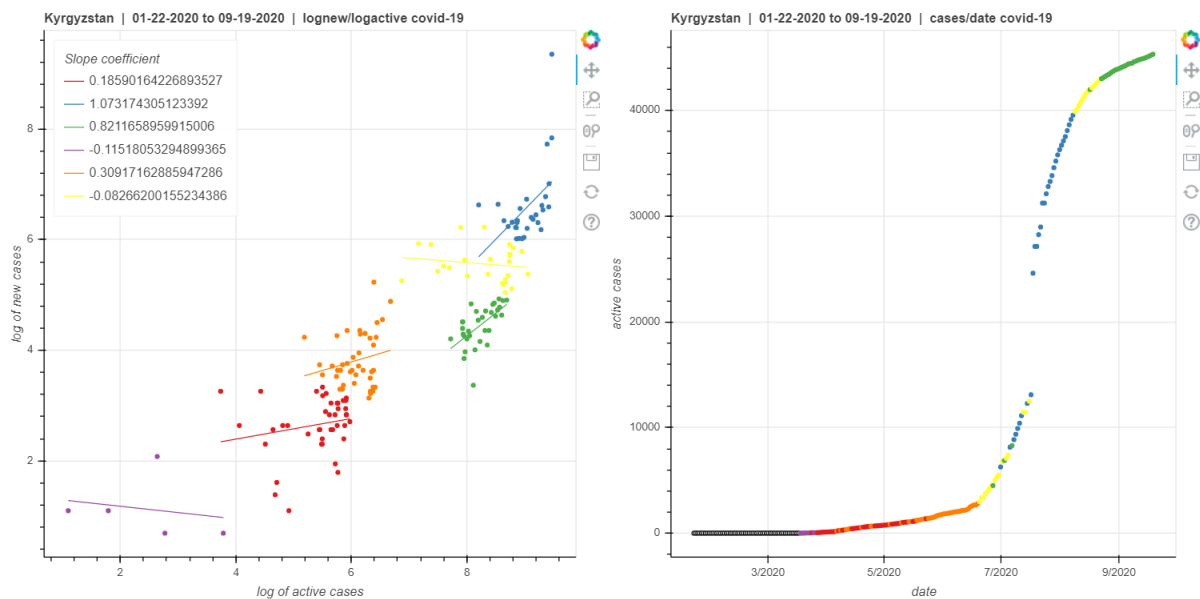
Slika 38. Gana (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Gana): 0.8828602077381392



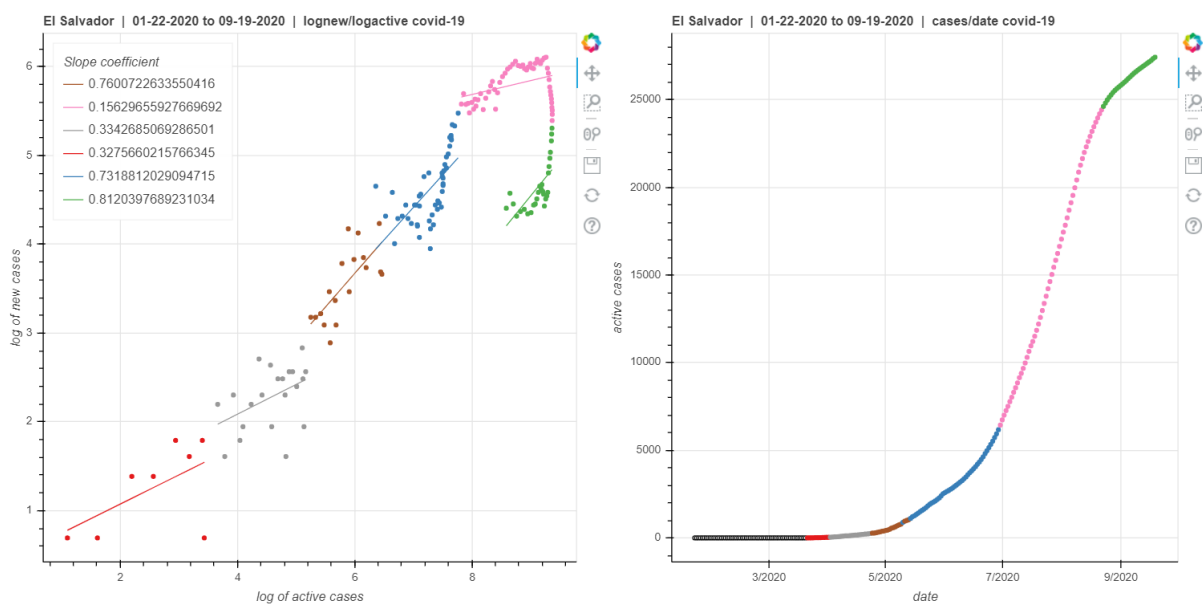
Slika 39. Gabon (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Gabon): 1.2164204902284643



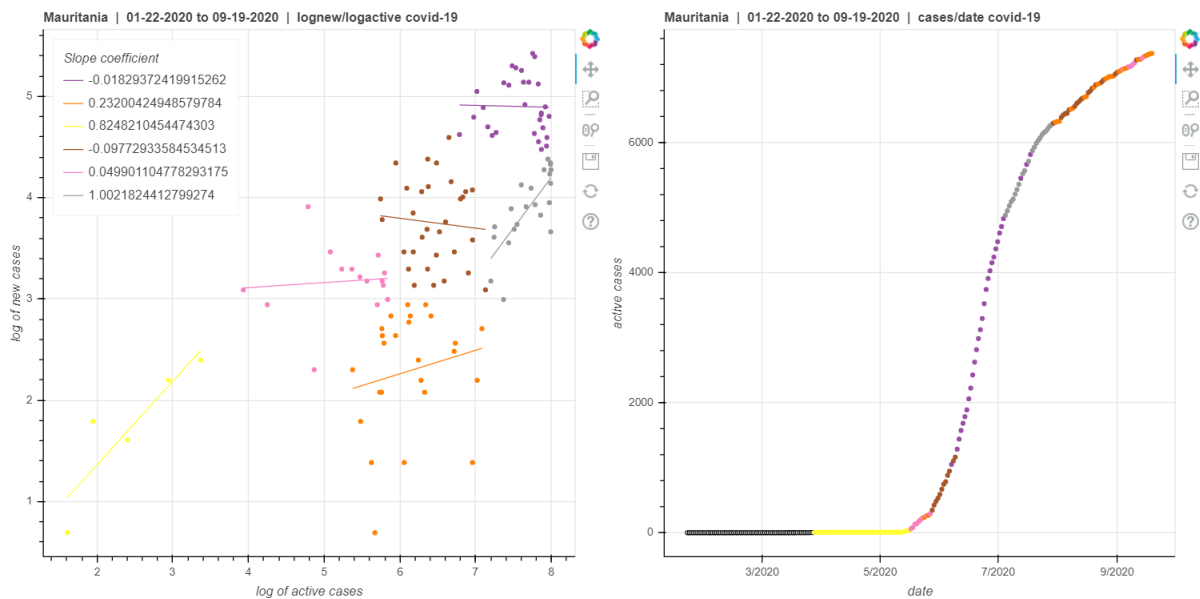
Slika 40. Kirgistan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kirgistan): 0.7001431224364252



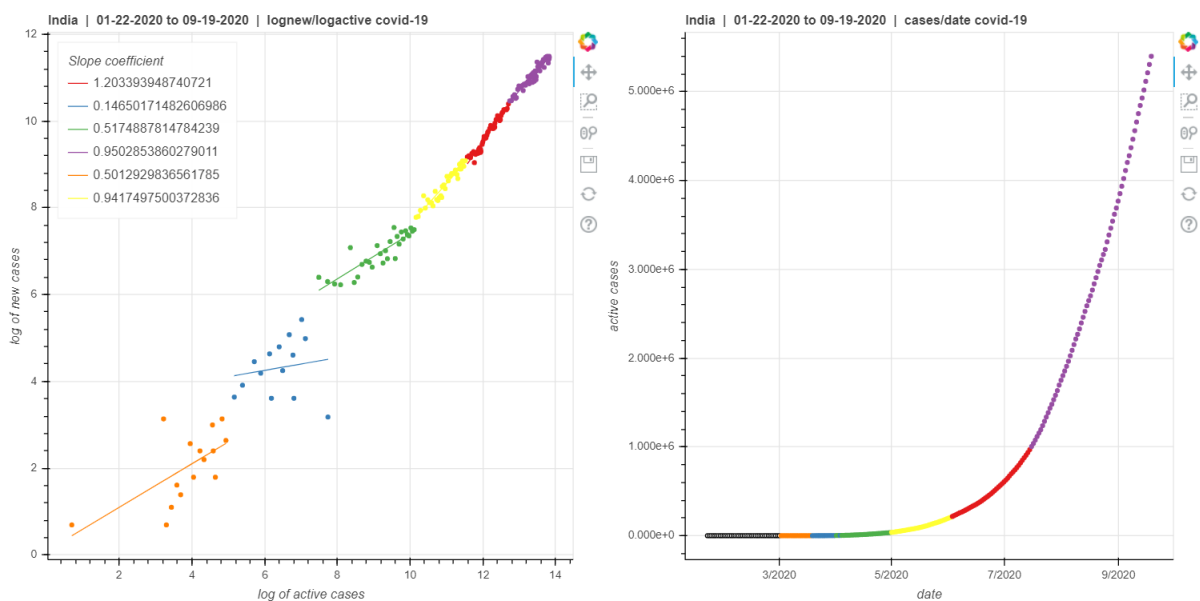
Slika 41. El Salvador (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (El Salvador): 0.566681396170367



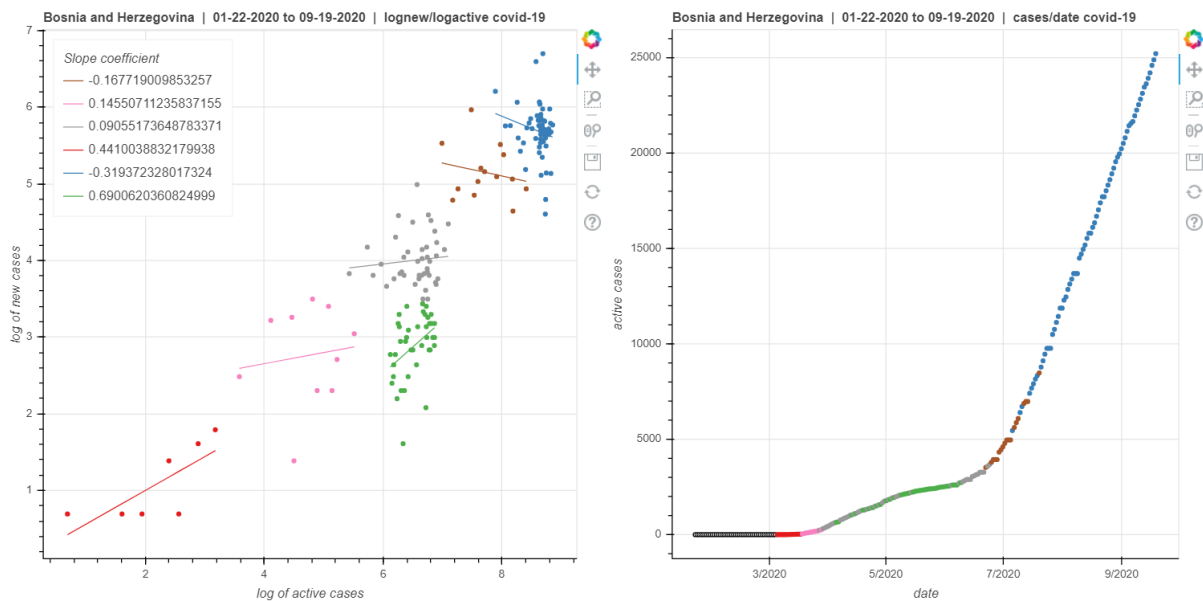
Slika 42. Mauritanija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Mauritanija): 0.8680759874699819



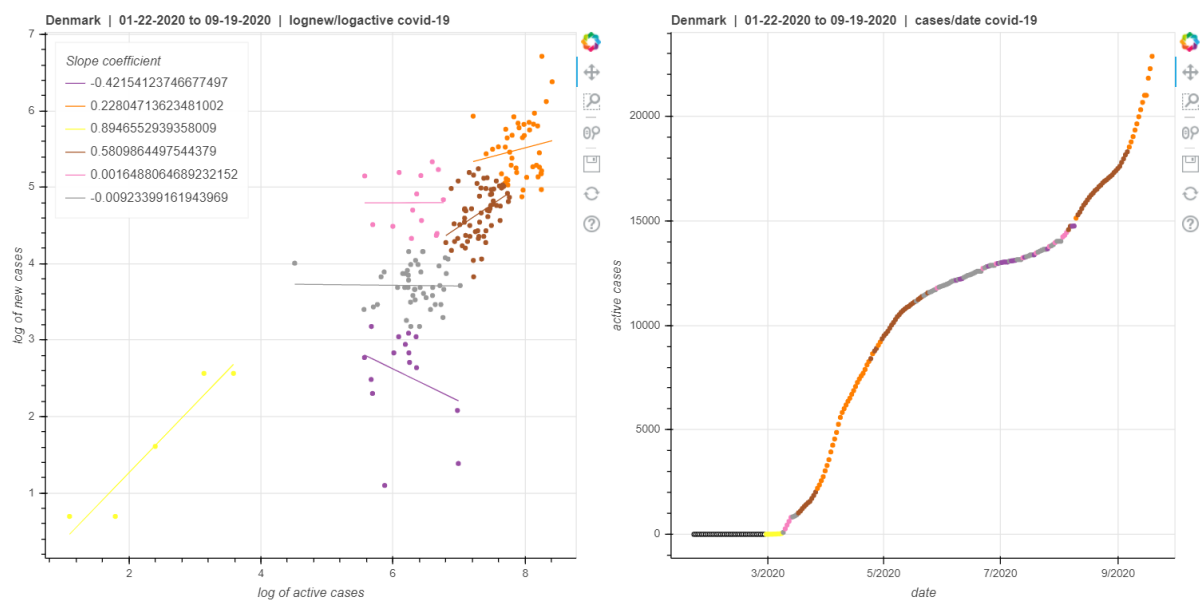
Slika 43. Indija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Indija): 0.3371551655170861



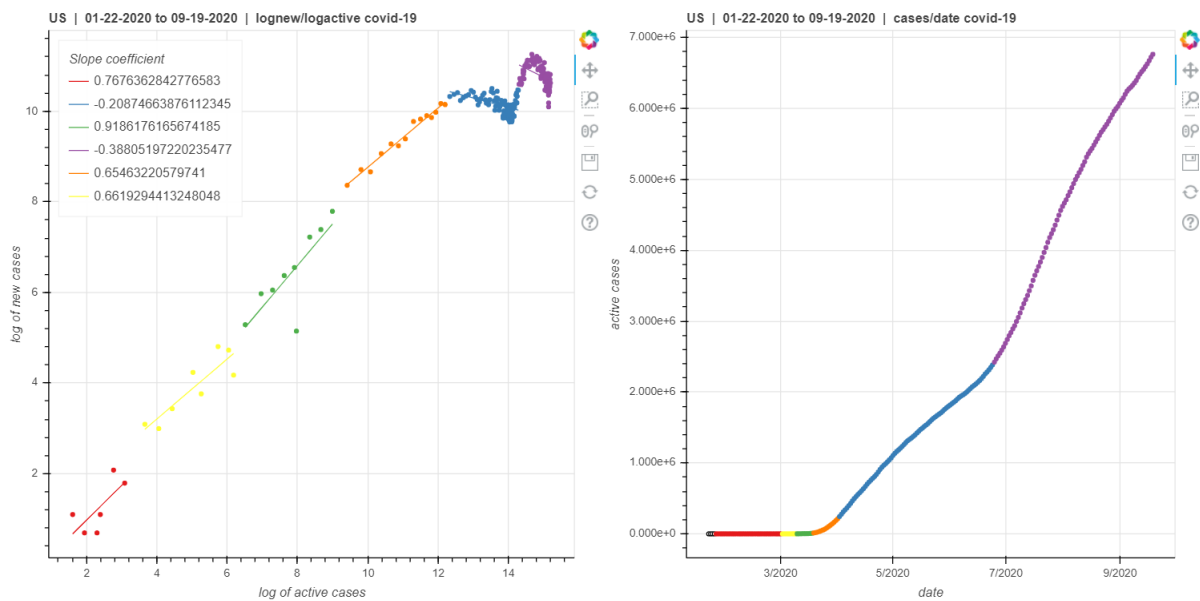
Slika 44. Bosna i Hercegovina (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bosna i Hercegovina): 1.0002441186242559



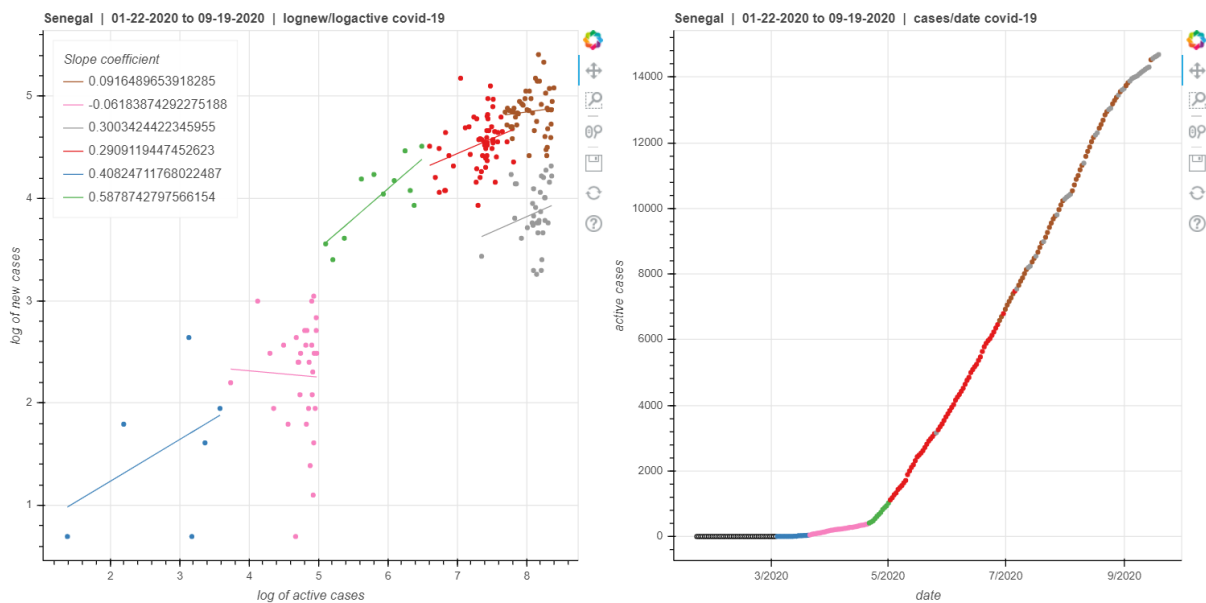
Slika 45. Danska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Danska): 0.8260875903415326



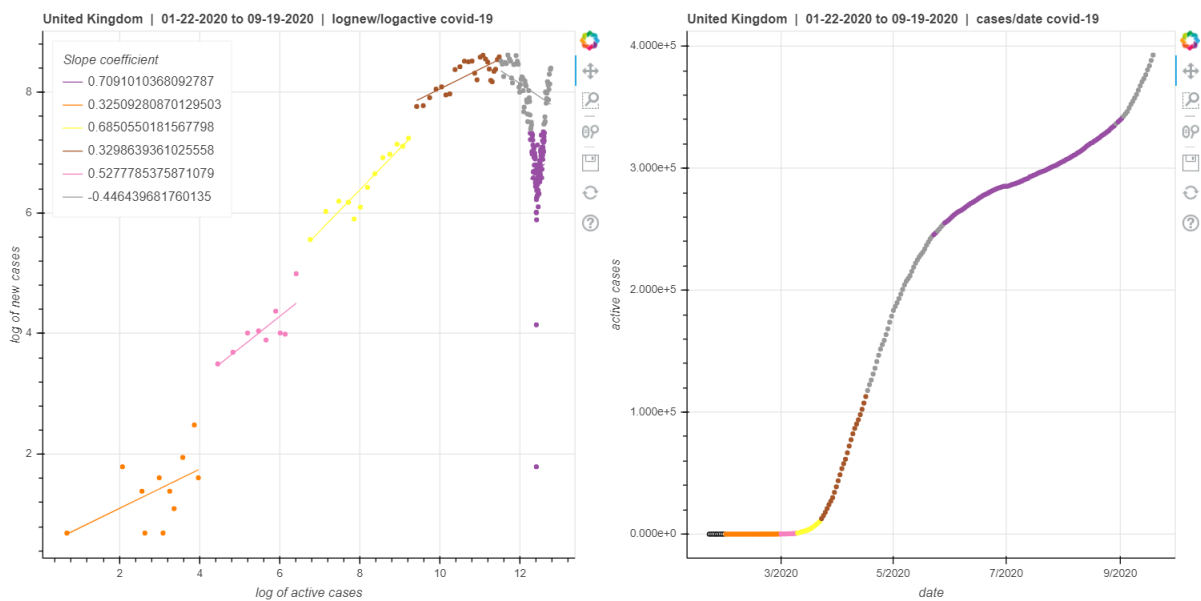
Slika 46. SAD (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (SAD): 1.1875140332085048



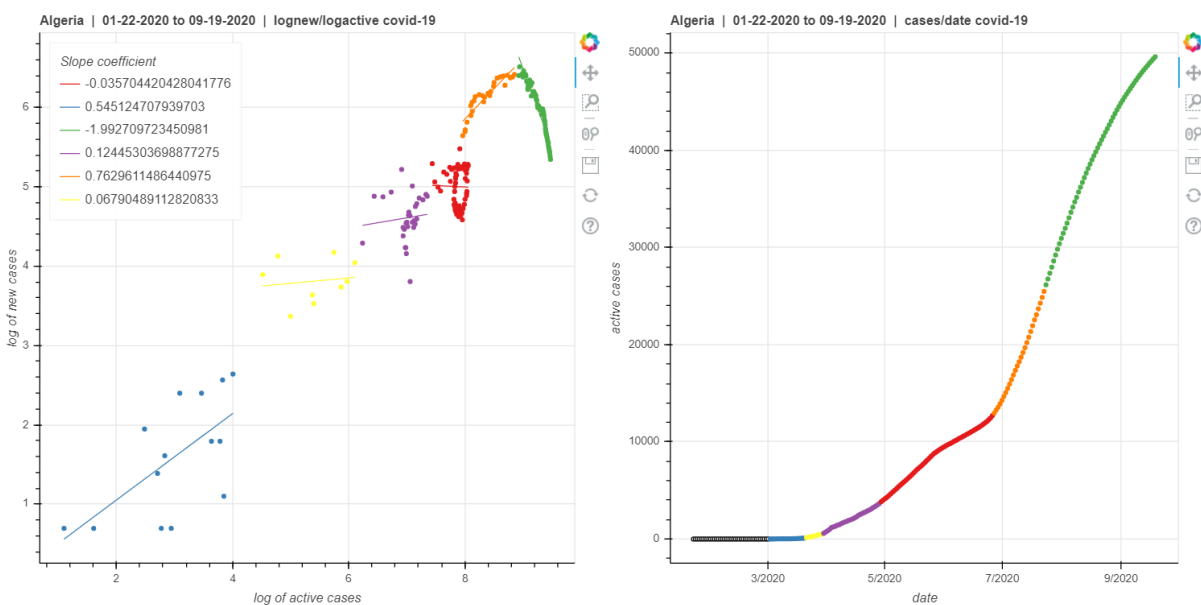
Slika 47. Senegal (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Senegal): 0.8137649201131414



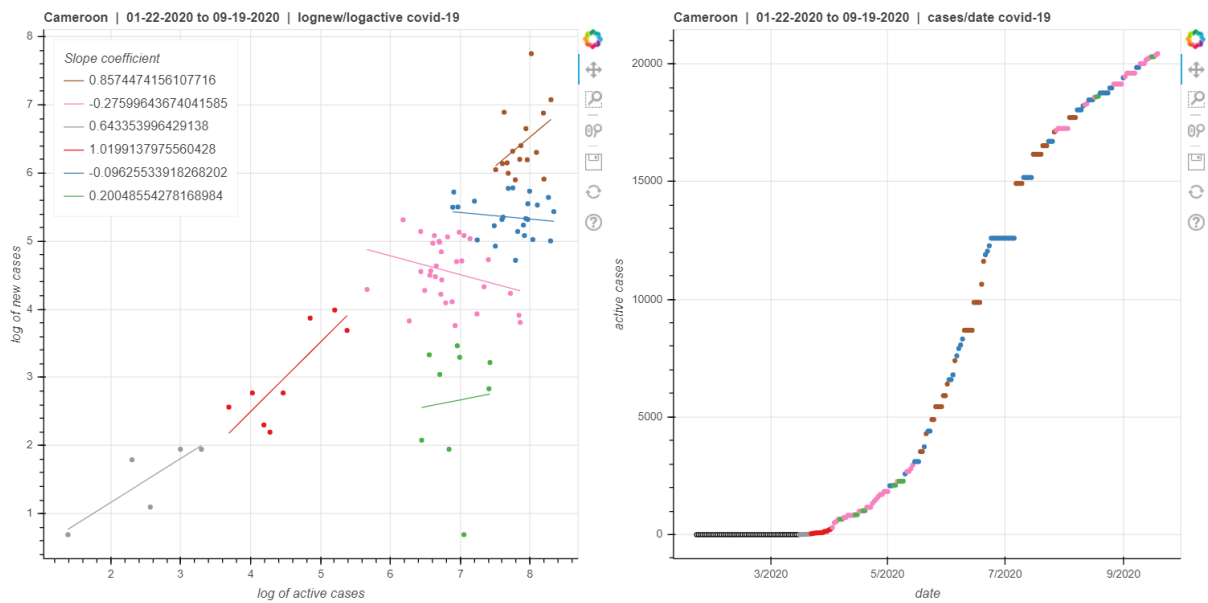
Slika 48. Ujedinjeno Kraljevstvo (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Ujedinjeno Kraljevstvo): 0.8349095699336785



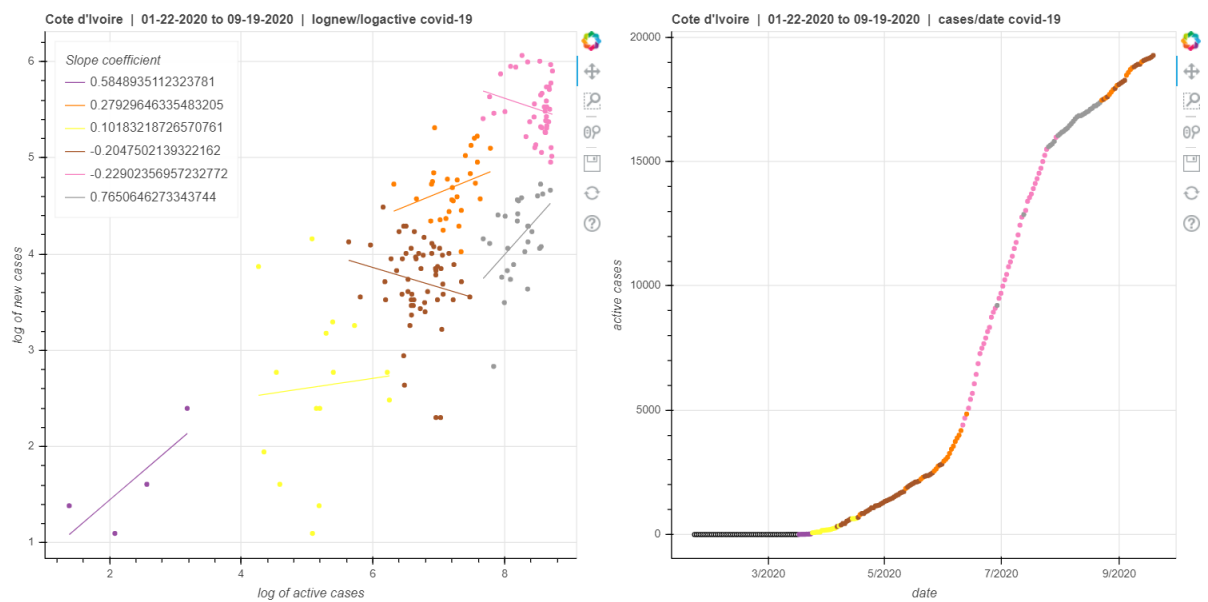
Slika 49. Algerija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Algerija): 1.77115195150171



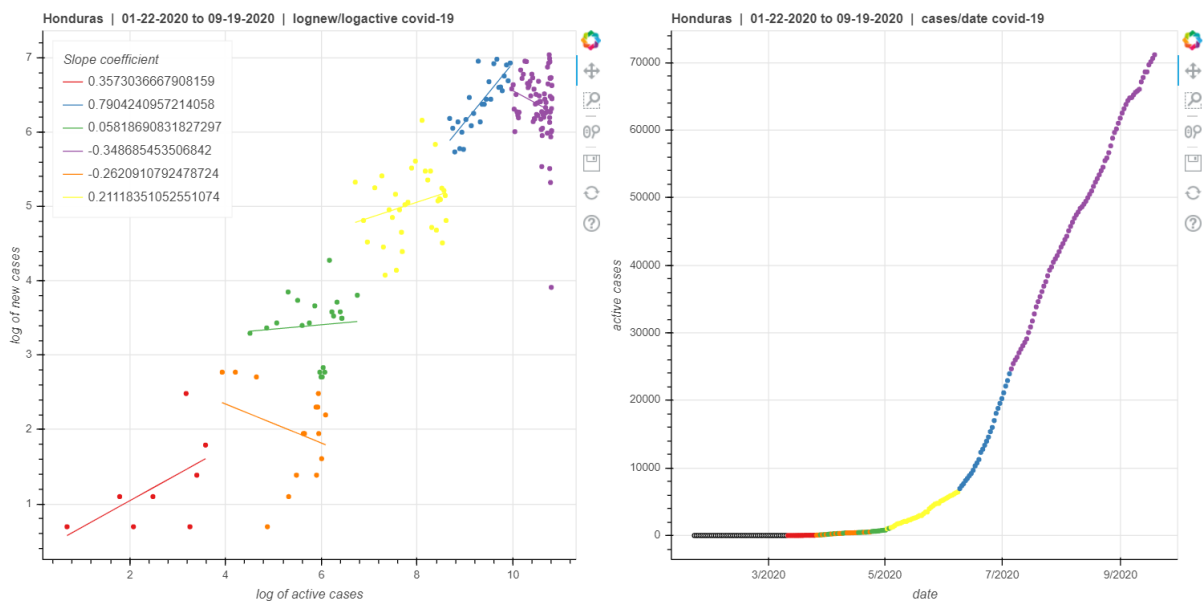
Slika 50. Kamerun (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kamerun): 0.9656948987789319



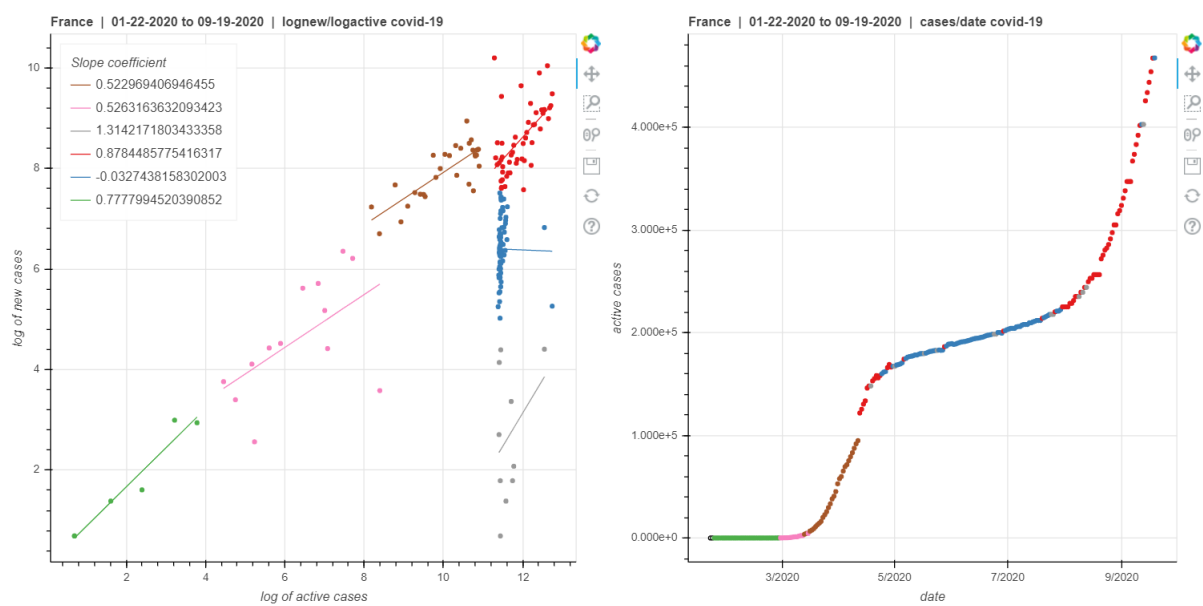
Slika 51. Obala Bjelokosti (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Obala Bjelokosti): 1.0043191224762105



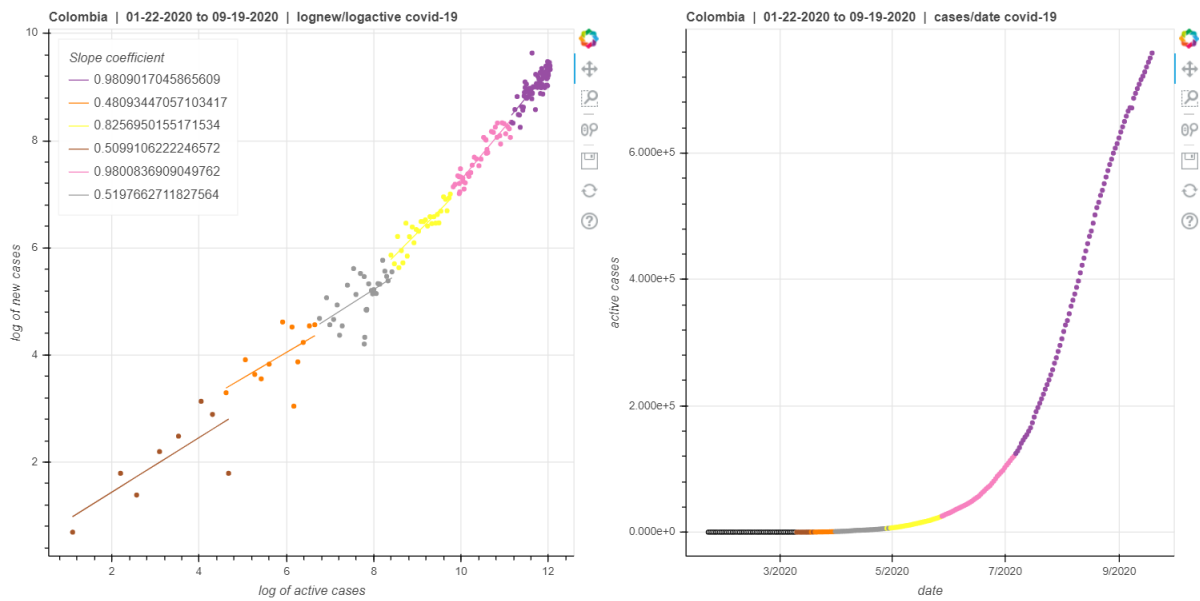
Slika 52. Honduras (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Honduras): 1.067461339752071



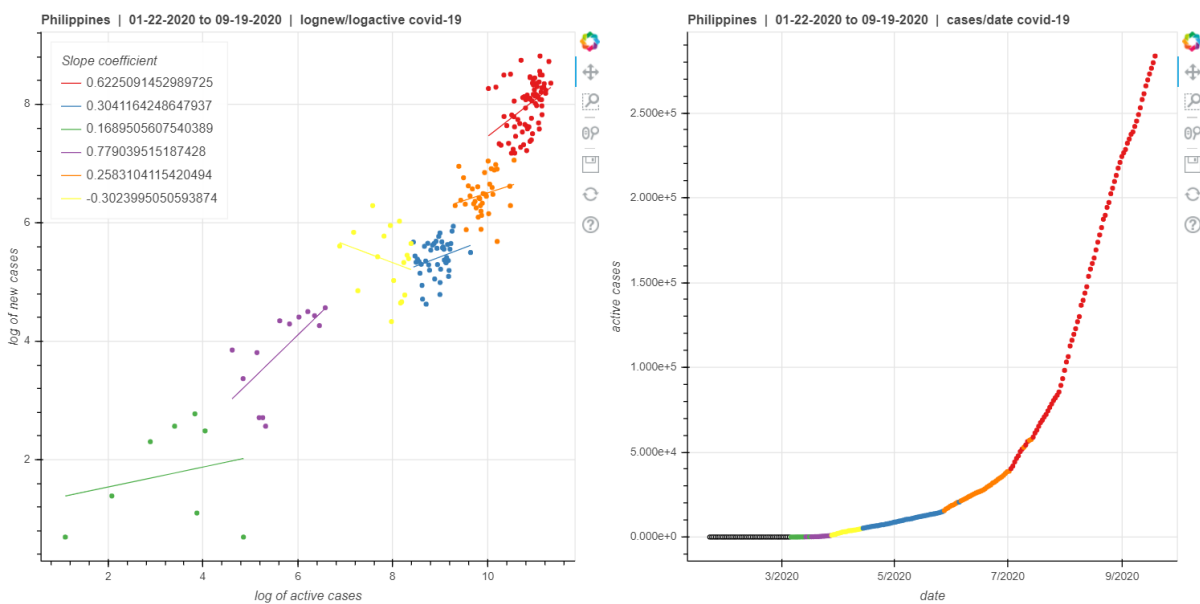
Slika 53. Francuska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Francuska): 0.6958850293108145



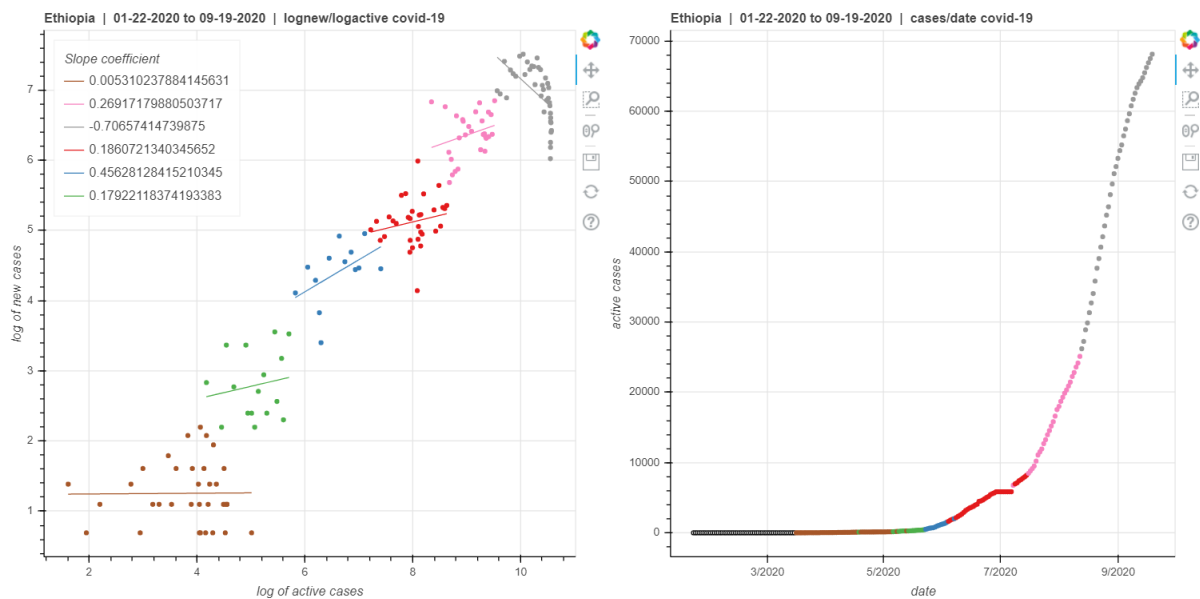
Slika 54. Kolumbija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kolumbija): 0.2578628083596066



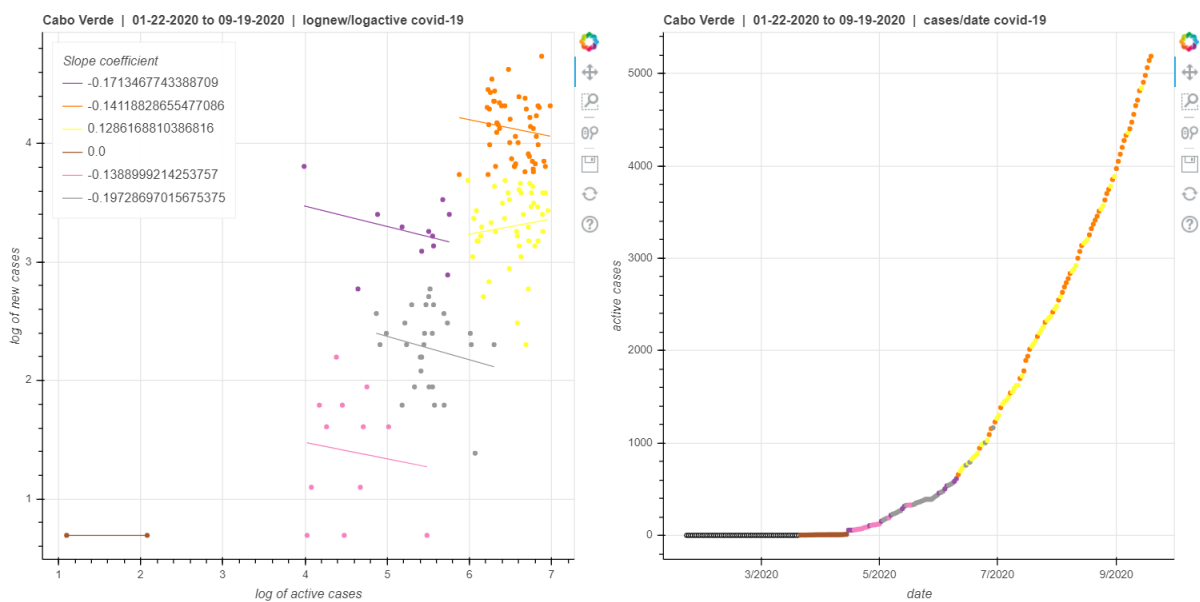
Slika 55. Filipini (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Filipini): 0.6703333373096008



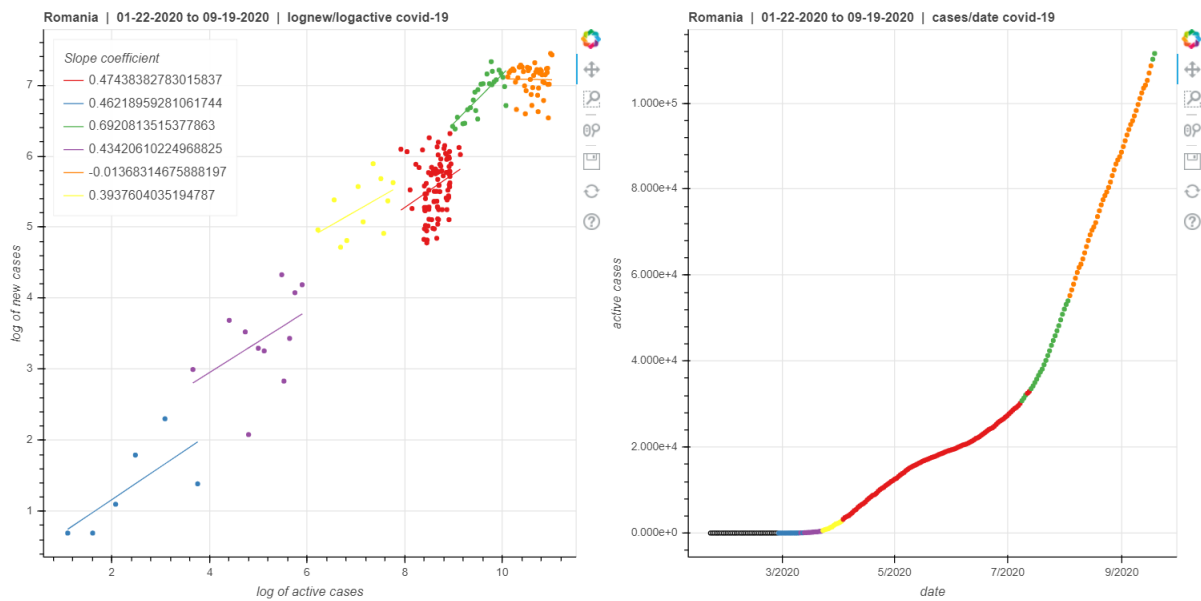
Slika 56. Etiopija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Etiopija): 1.0988810845410786



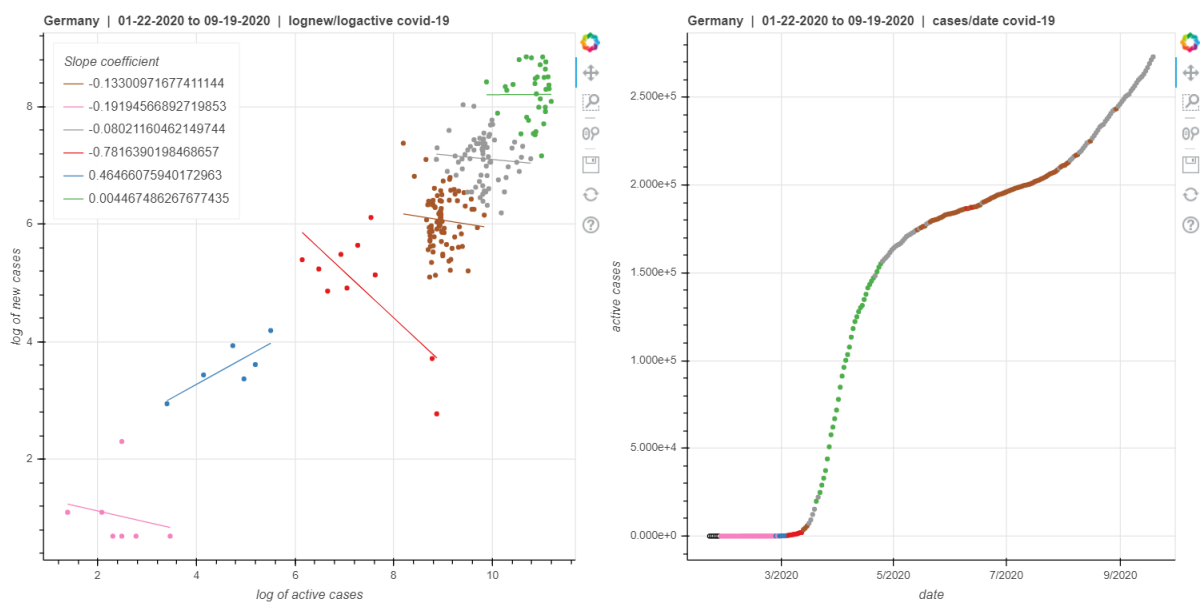
Slika 57. Zelenortski Otoci (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Zelenortski Otoci): 1.0787213505238131



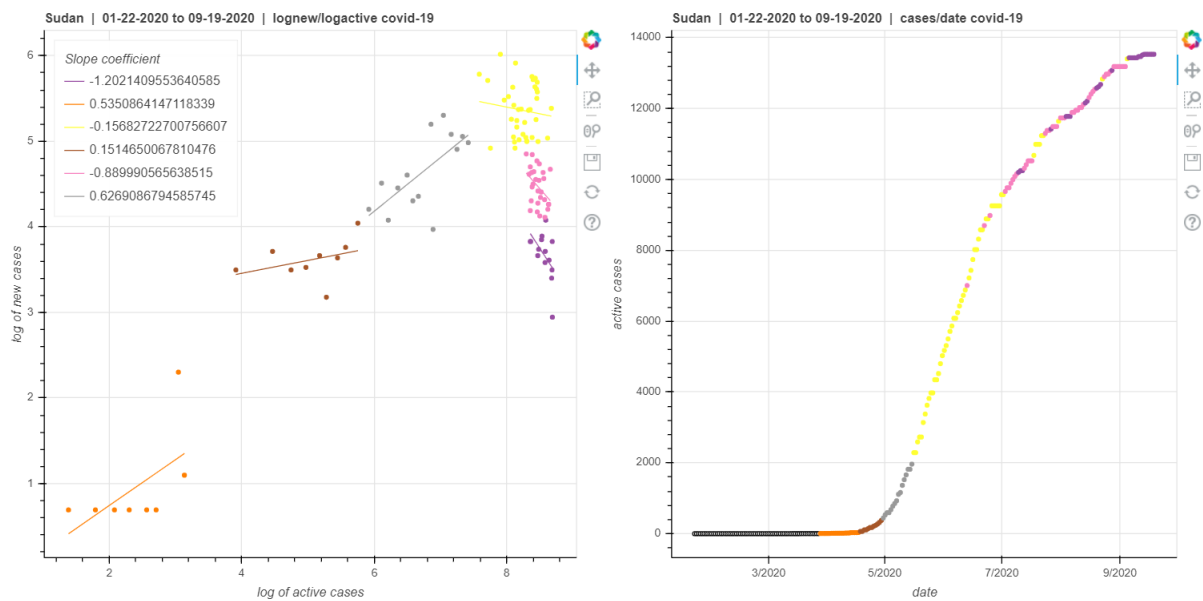
Slika 58. Rumunjska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Rumunjska): 0.6565736537701222



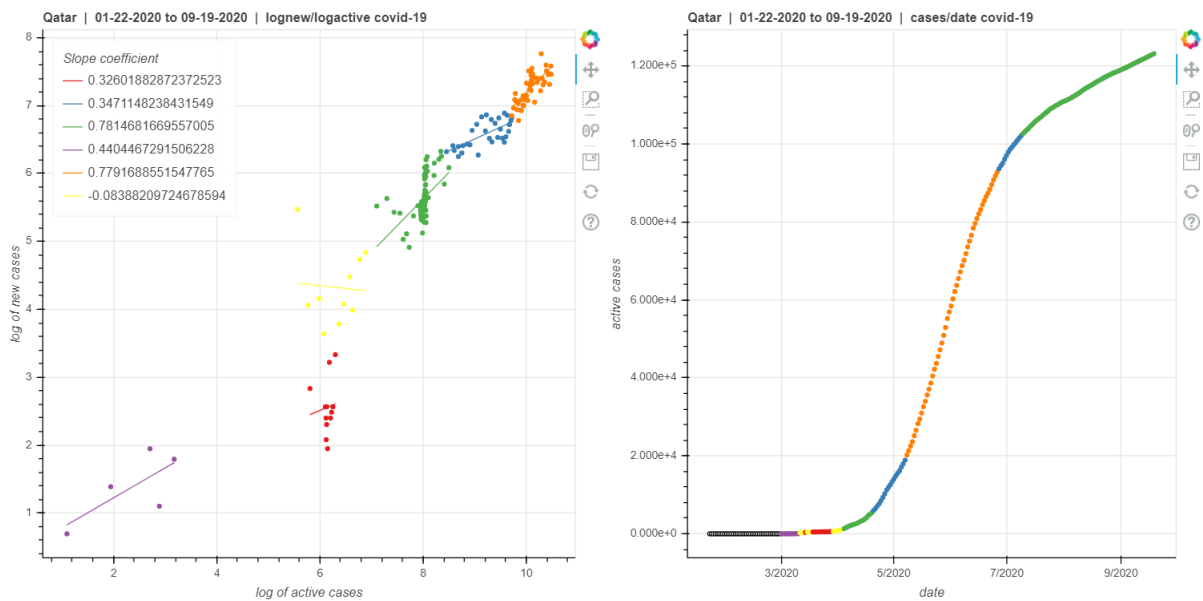
Slika 59. Njemačka (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Njemačka): 1.1236431271062424



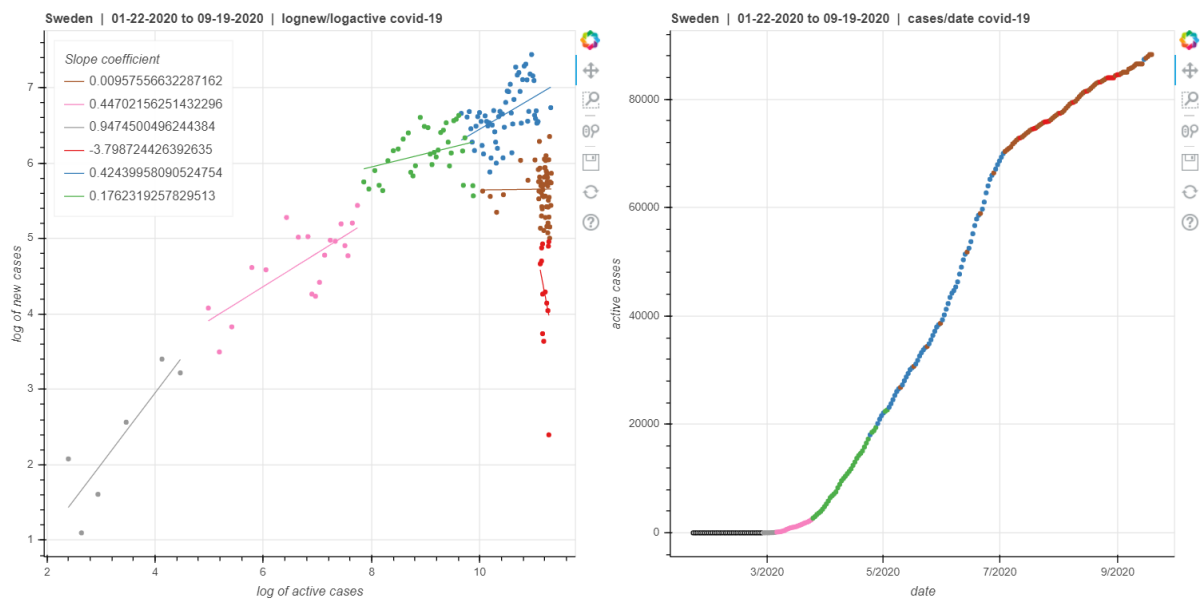
Slika 60. Sudan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Sudan): 1.4360896640258547



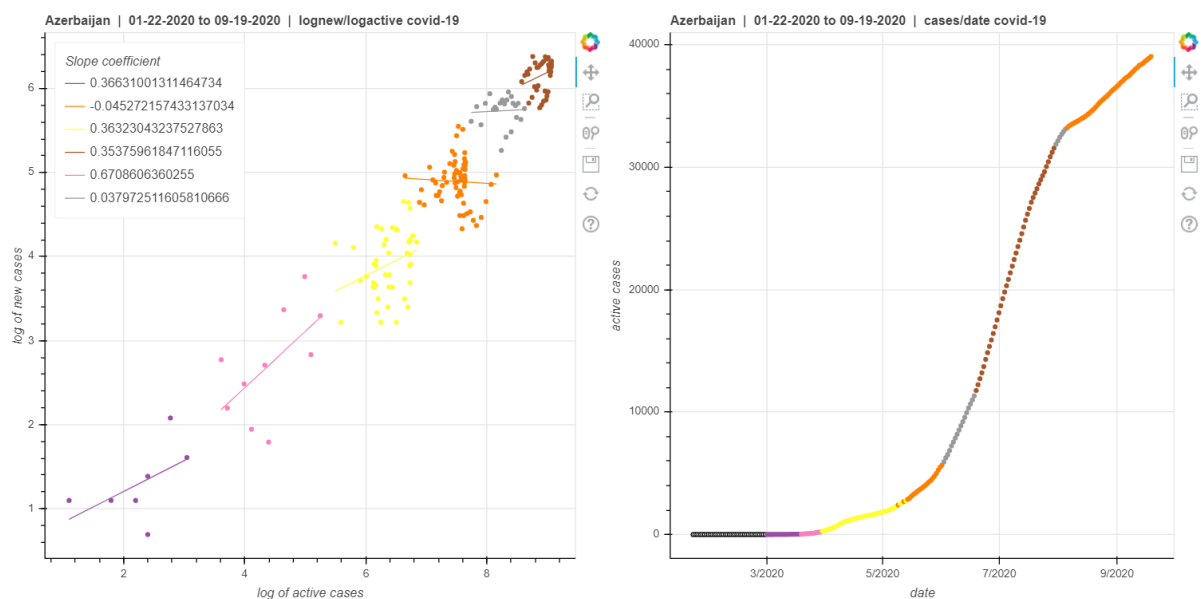
Slika 61. Katar (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Katar): 0.4438893695525539



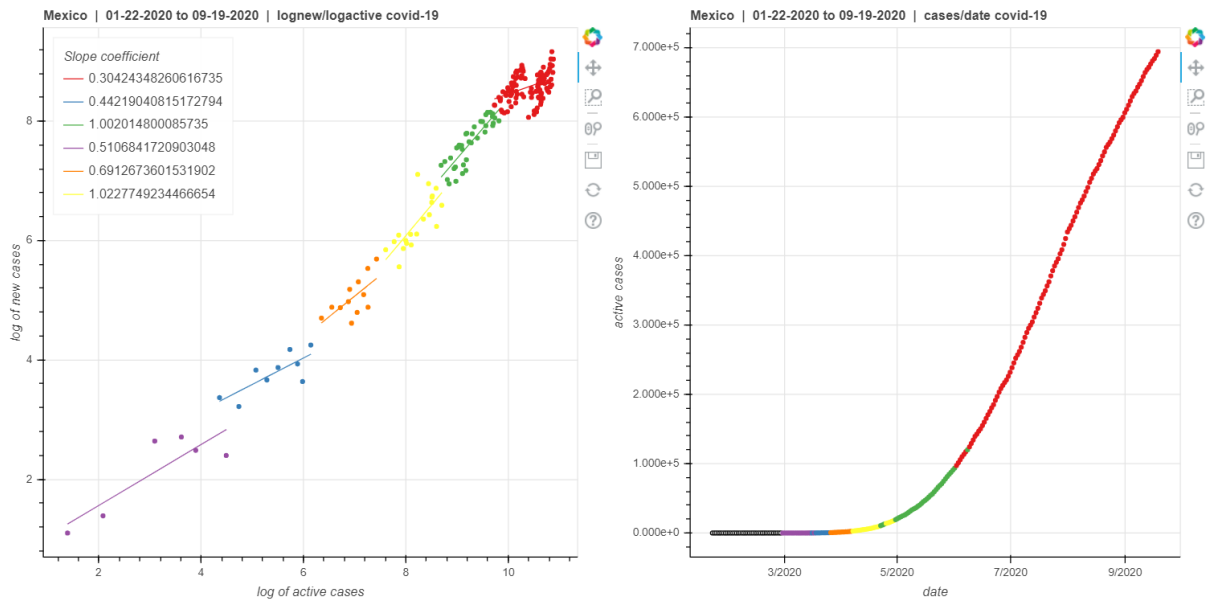
Slika 62. Švedska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Švedska): 1.4493771713503247



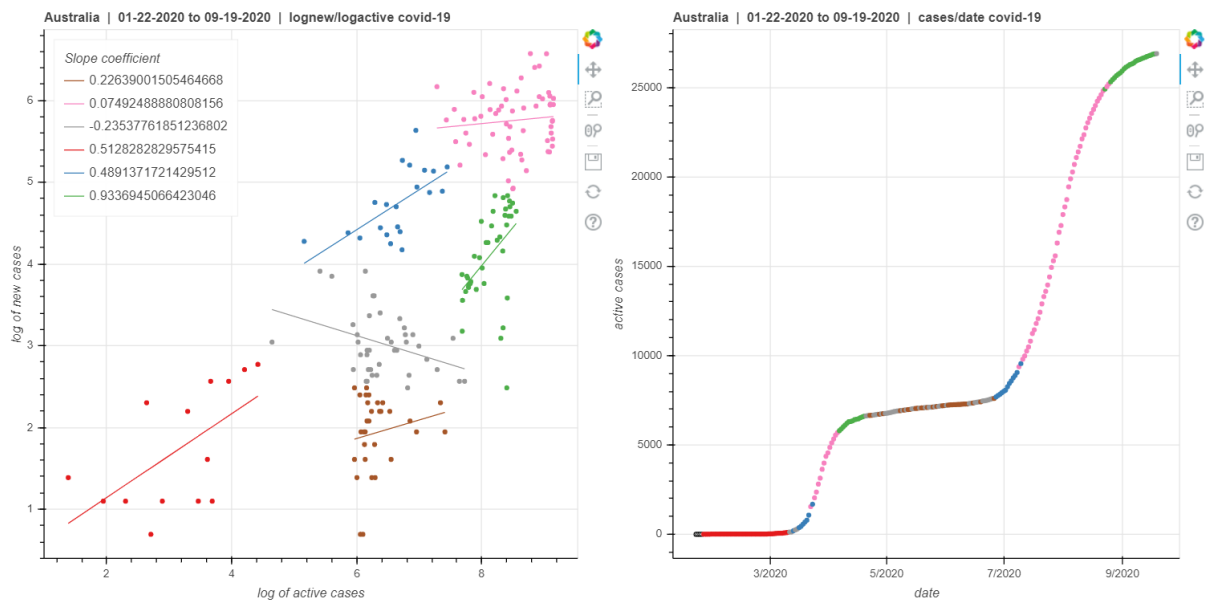
Slika 63. Azerbejdžan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Azerbejdžan): 0.8321663432161339



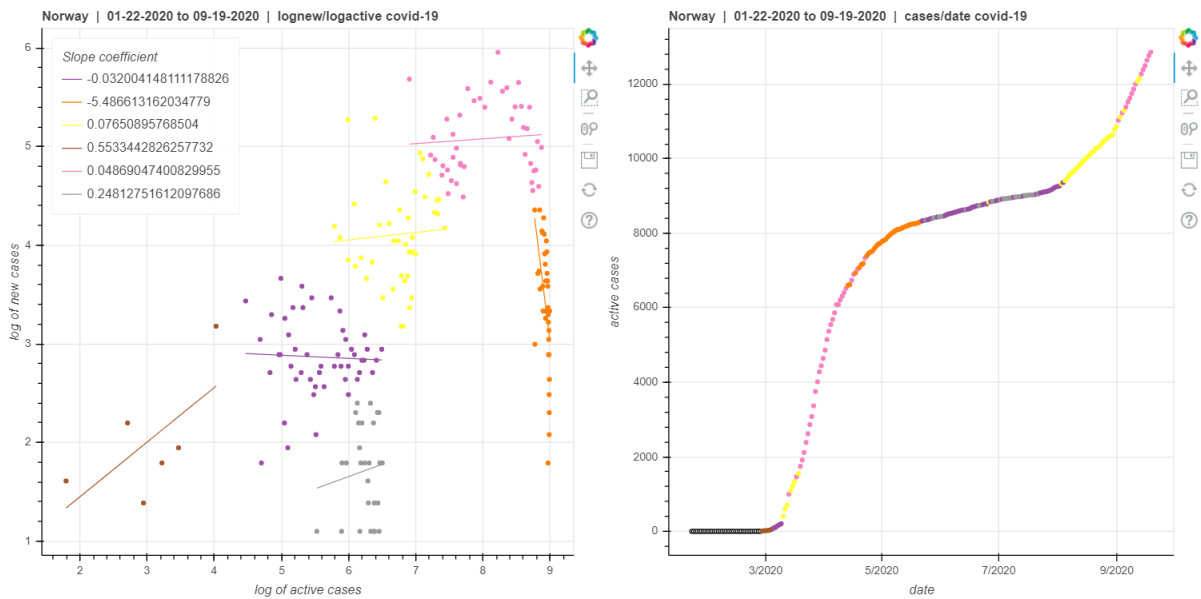
Slika 64. Meksiko (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Meksiko): 0.5489304972790374



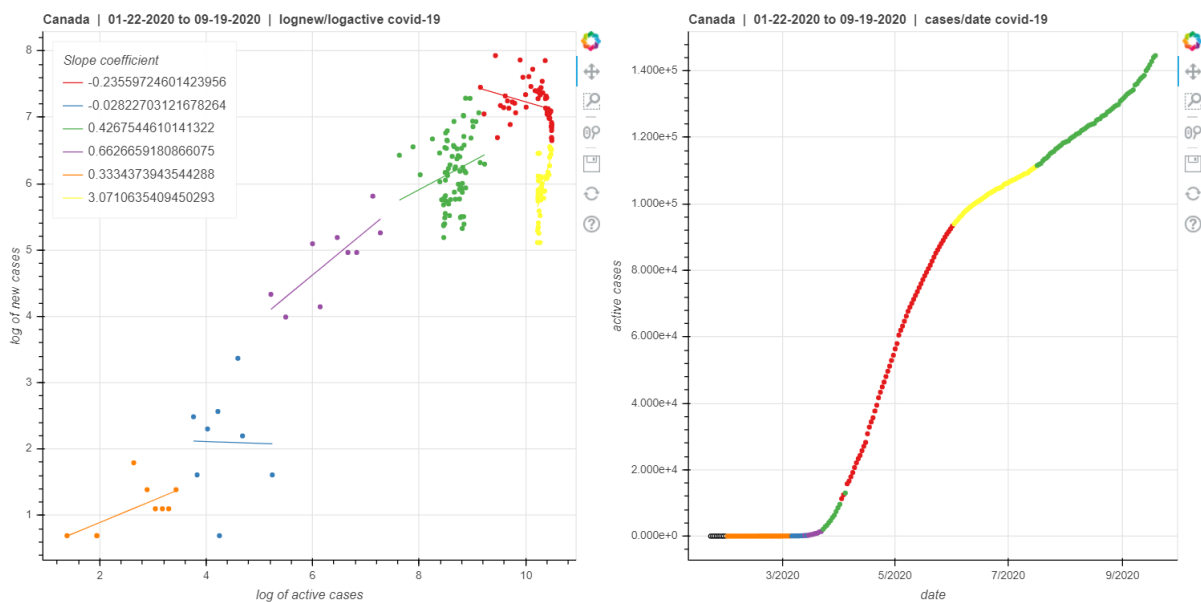
Slika 65. Australija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Australija): 0.826156432476934



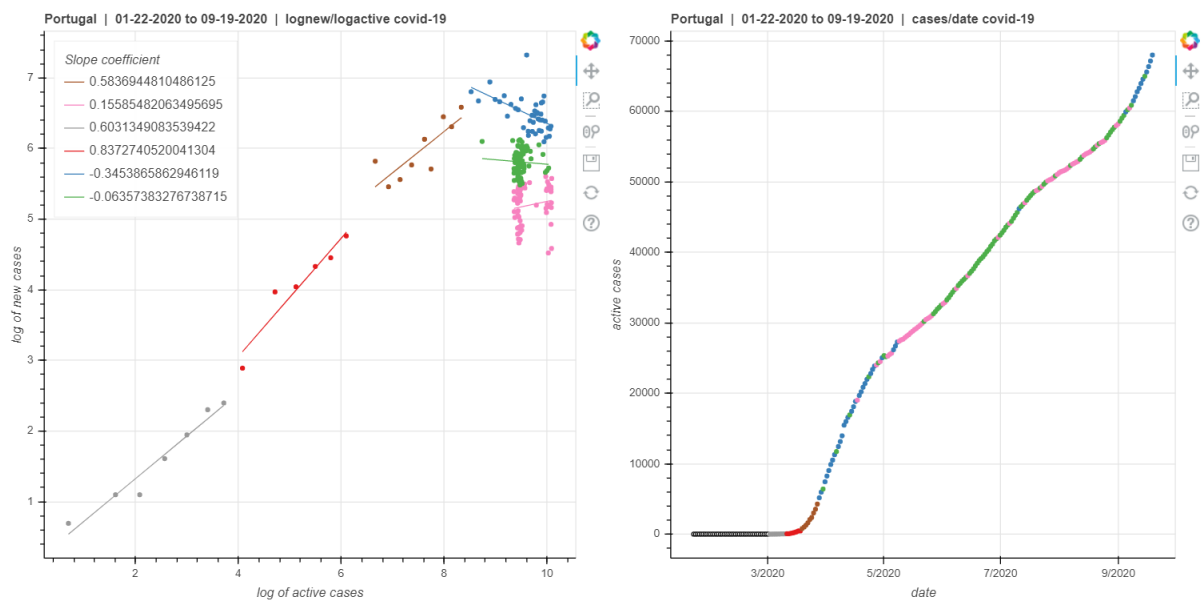
Slika 66. Norveška (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Norveška): 2.82851298309199



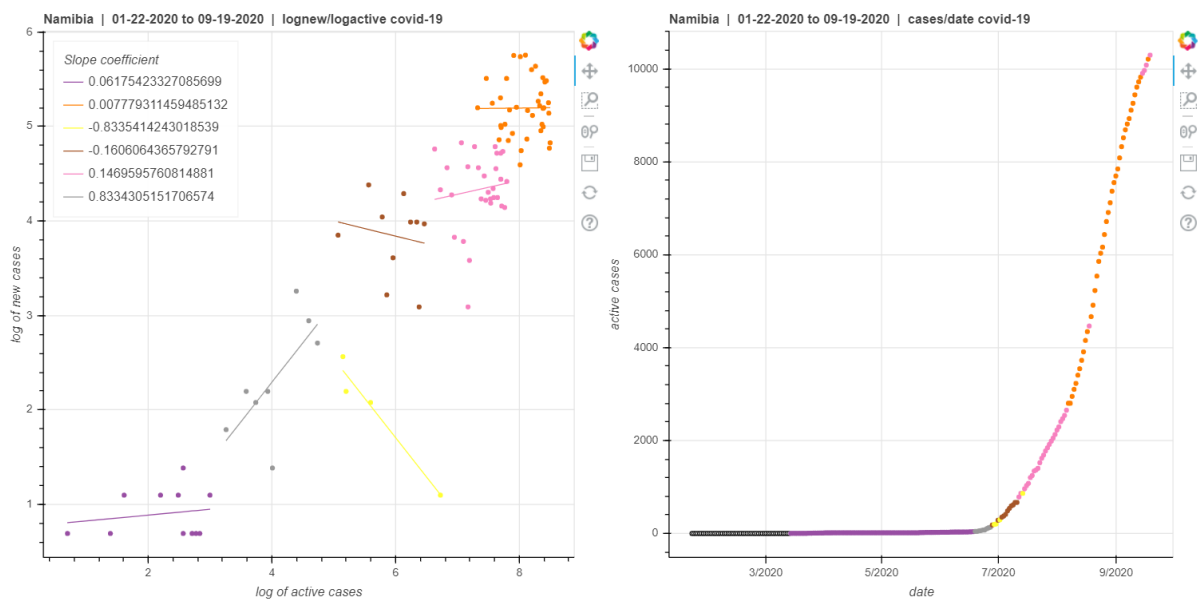
Slika 67. Kanada (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kanada): 1.2493012973206128



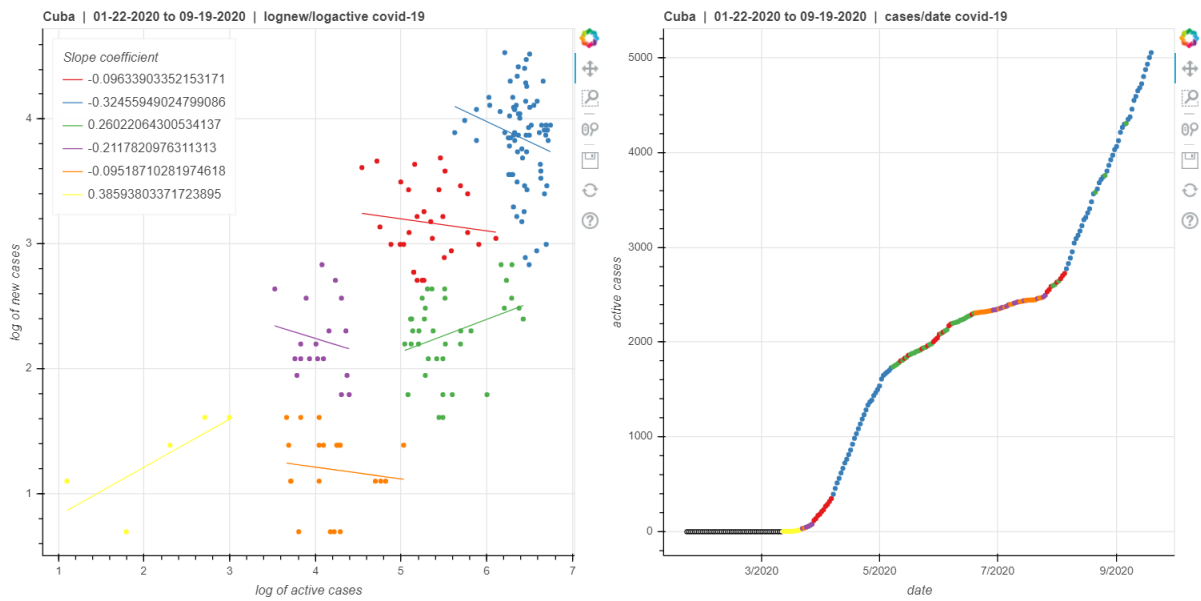
Slika 68. Portugal (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Portugal): 1.0163554003671622



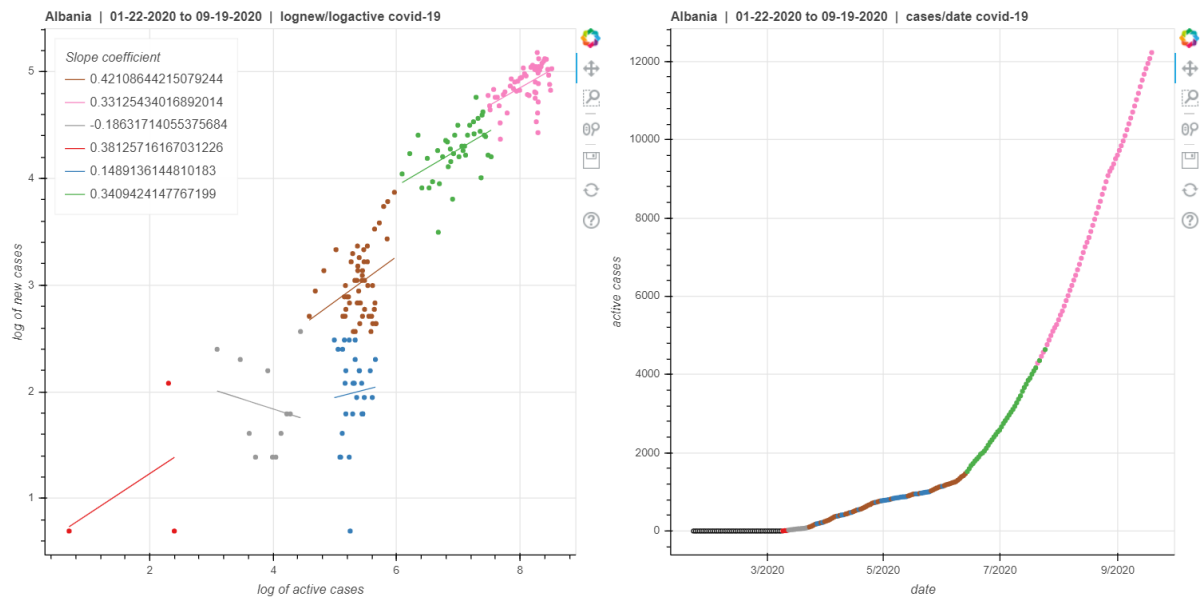
Slika 69. Namibija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Namibija): 0.9752598030299736



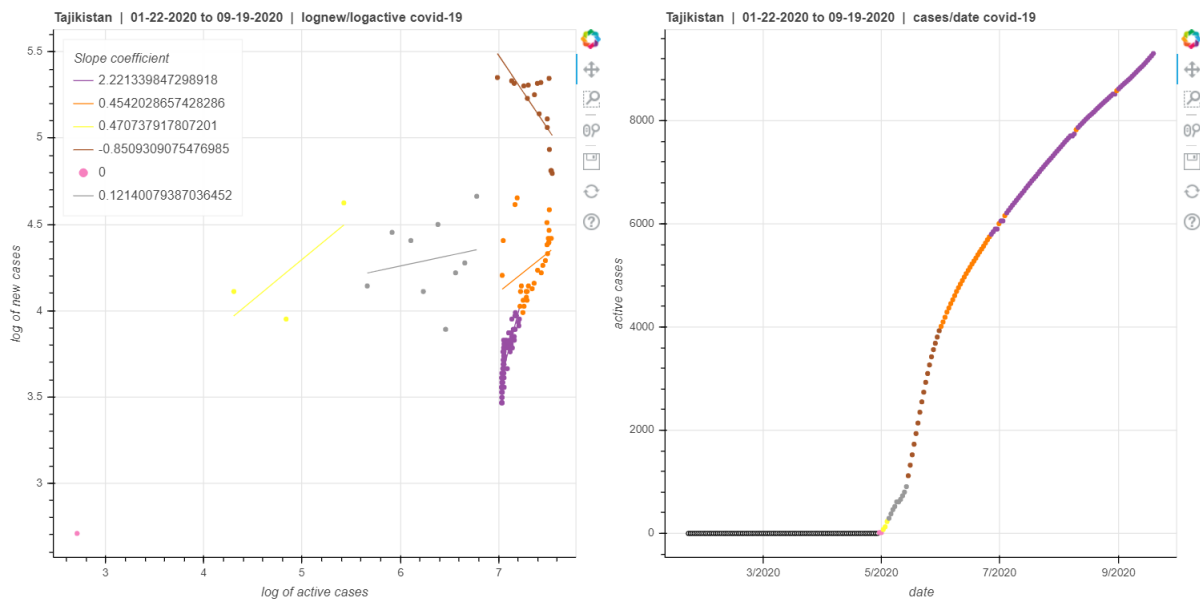
Slika 70. Kuba (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kuba): 1.139070346731958



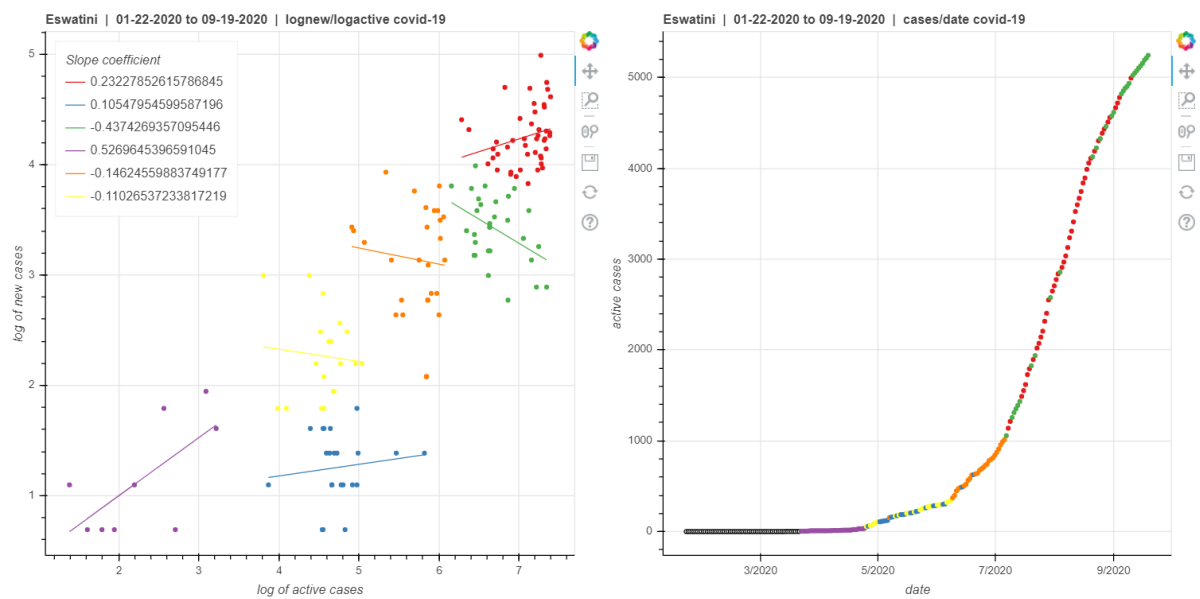
Slika 71. Albanija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Albanija): 0.7133623917683996



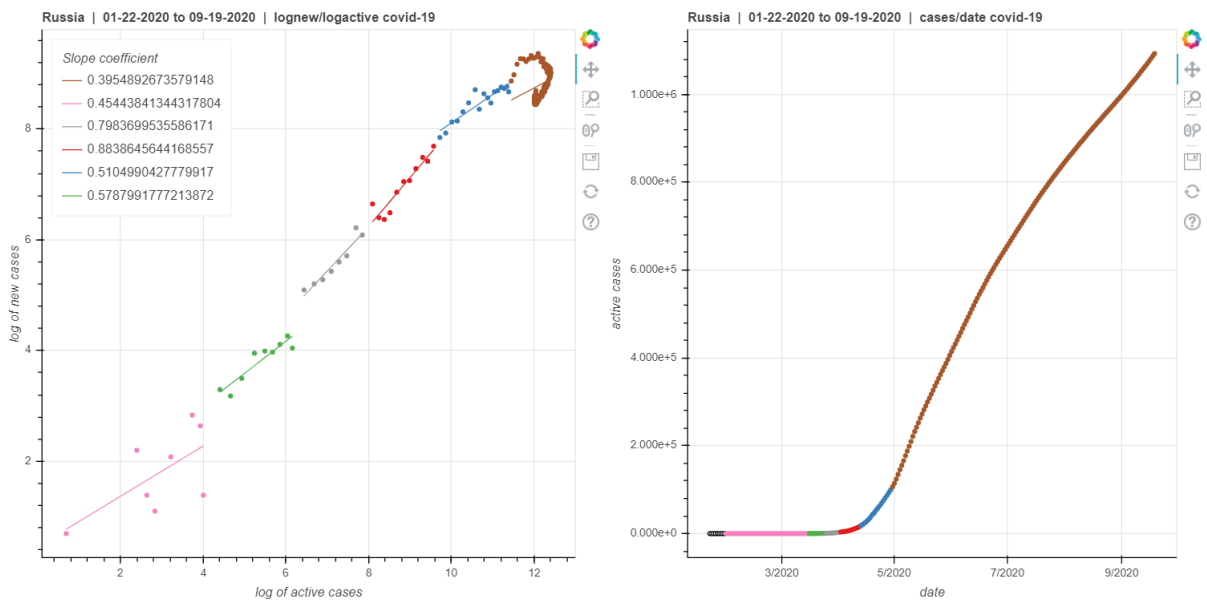
Slika 72. Tadžikistan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Tadžikistan): 1.1814259736646209



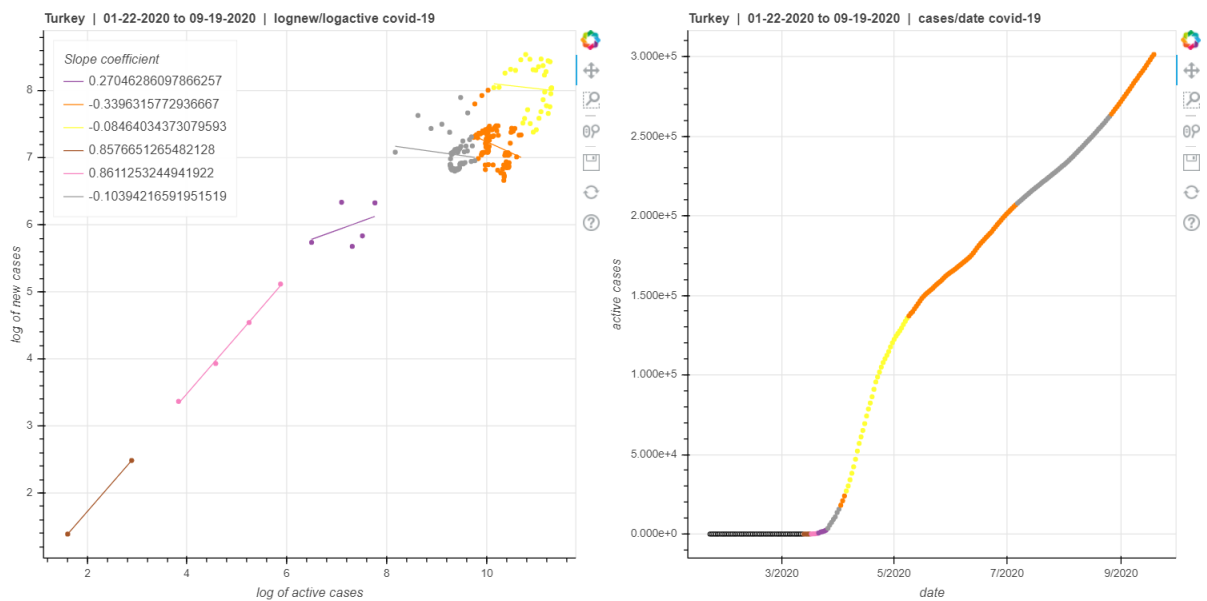
Slika 73. Svazi (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Svazi): 1.0482506671420821



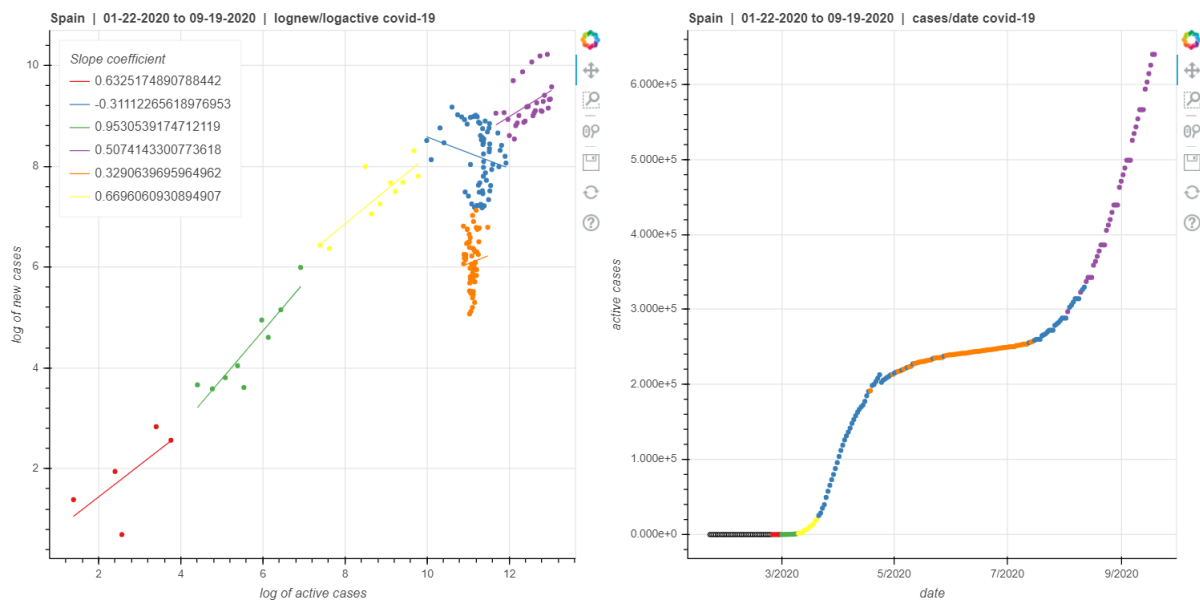
Slika 74. Rusija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Rusija): 0.55687599594374



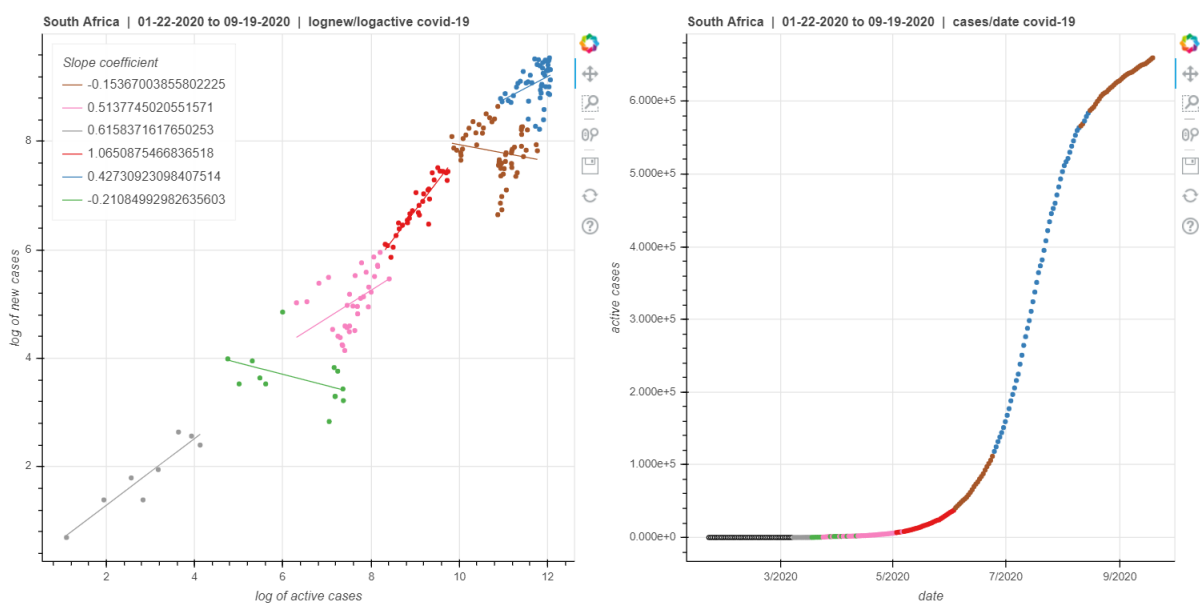
Slika 75. Turska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Turska): 1.1903459183705427



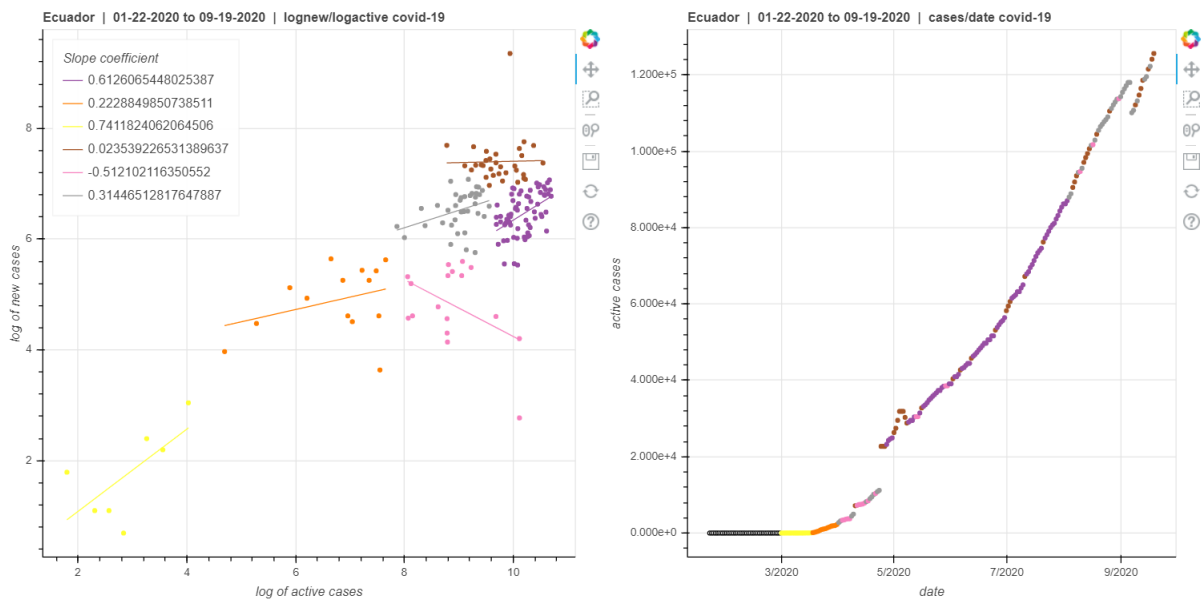
Slika 76. Španjolska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Španjolska): 0.9043159934742746



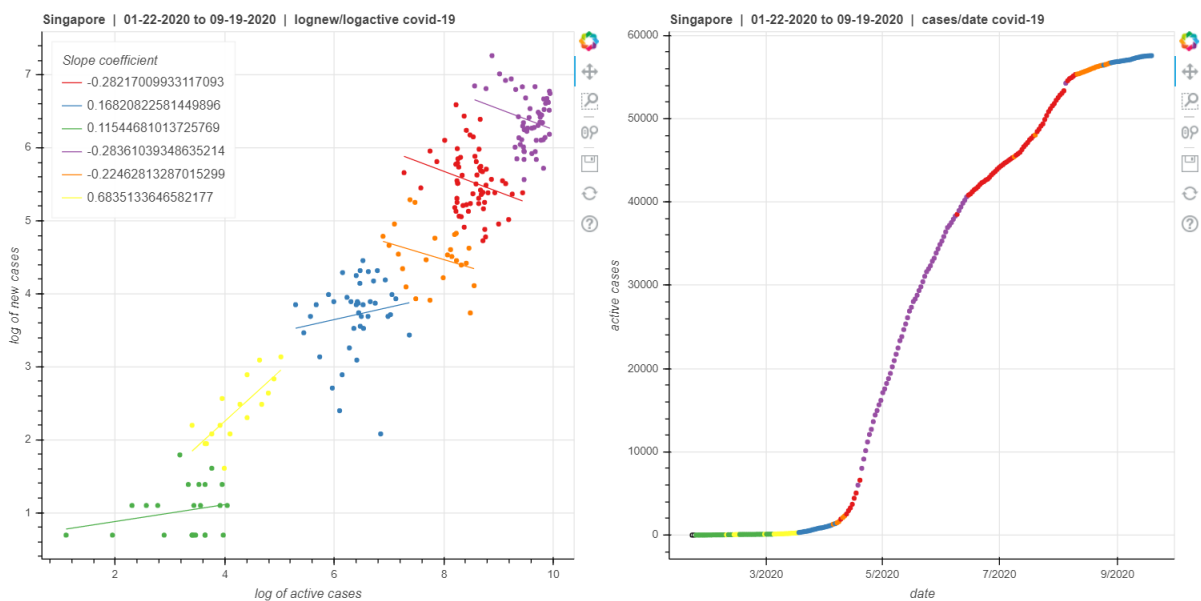
Slika 77. Južna Afrika (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Južna Afrika): 0.79302449589472



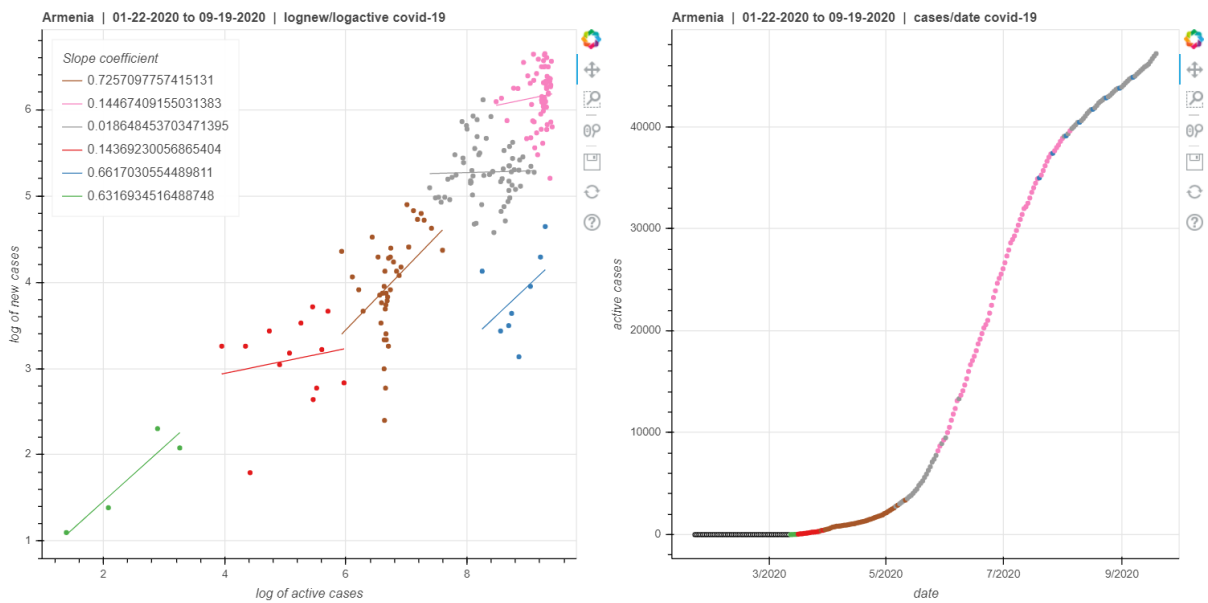
Slika 78. Ekvador (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Ekvador): 0.7800917388976972



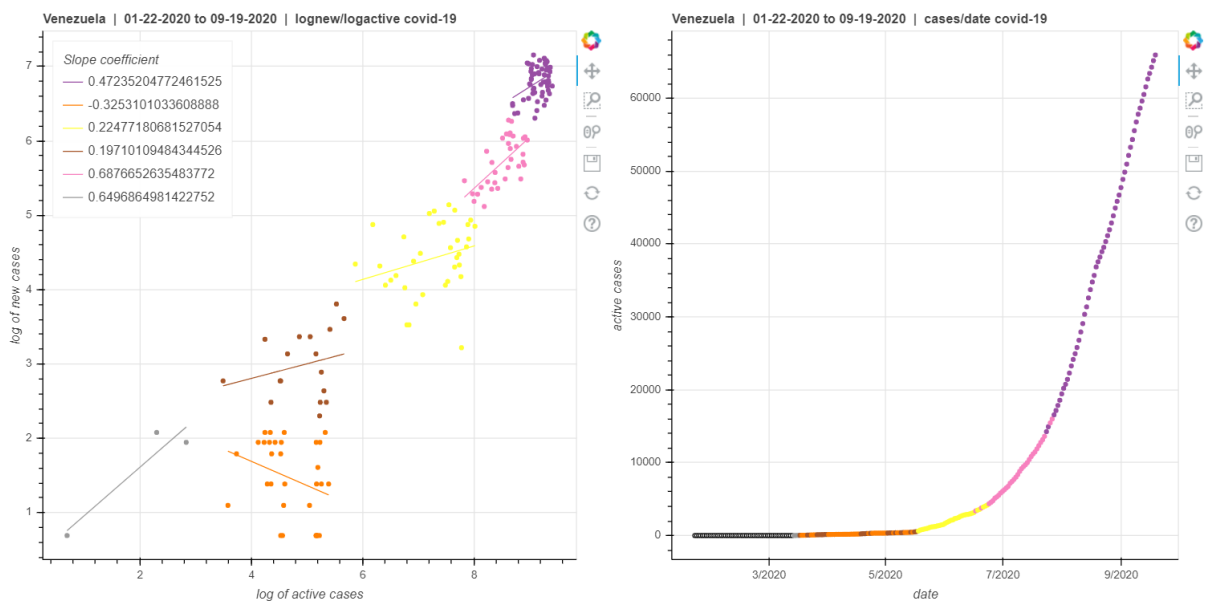
Slika 79. Singapur (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Singapur): 1.1197945592943972



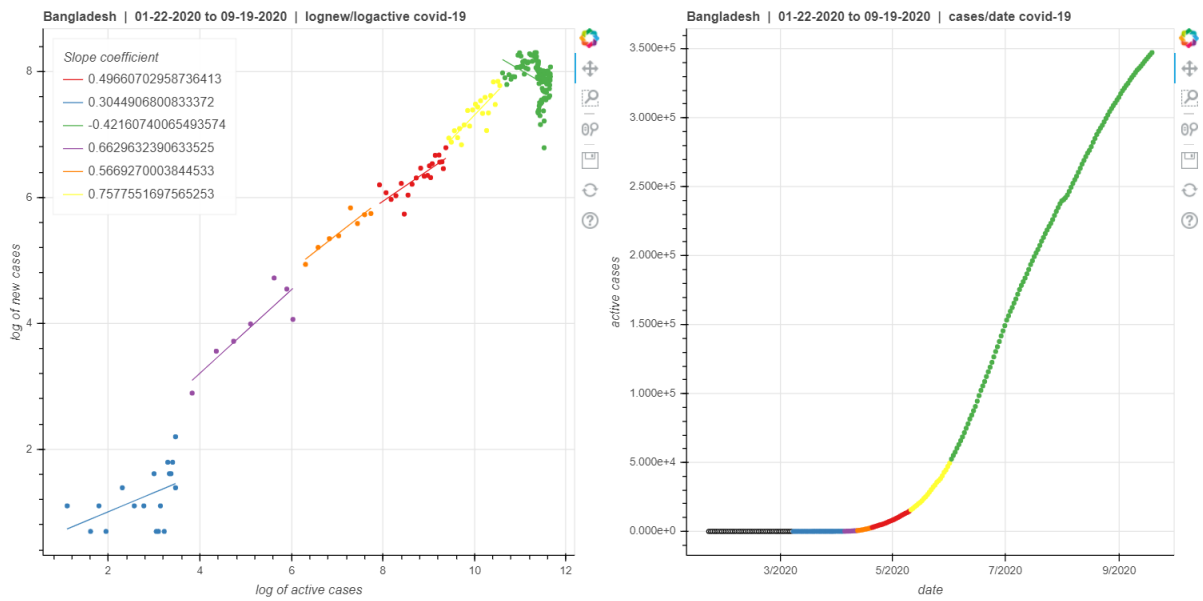
Slika 80. Armenija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Armenija): 0.7969878026093468



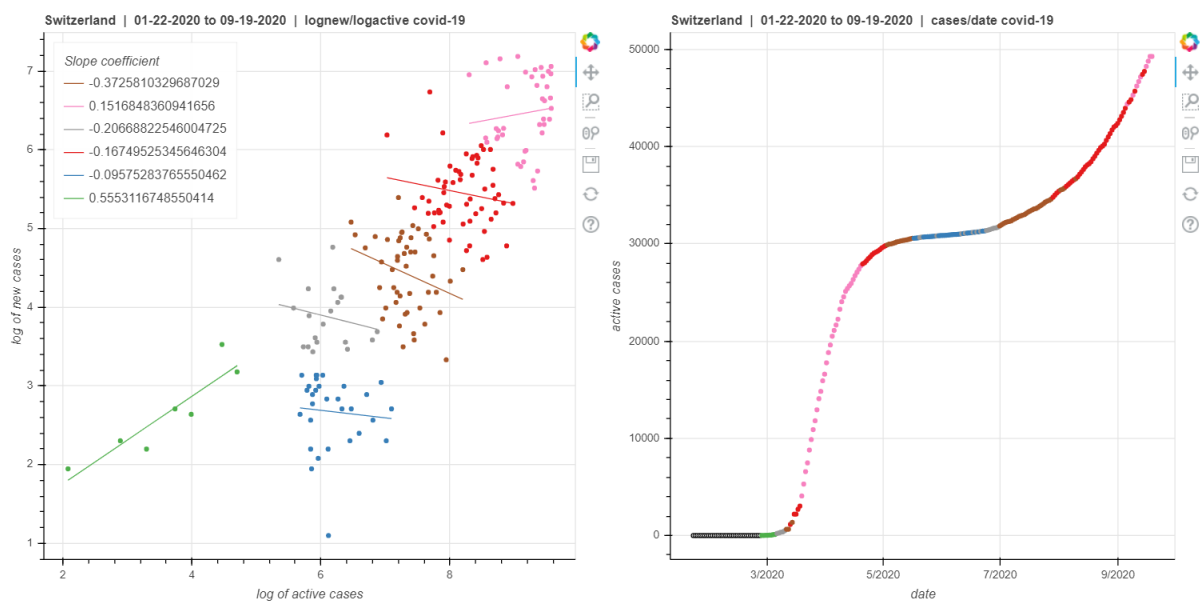
Slika 81. Venezuela (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Venezuela): 0.7750335580363885



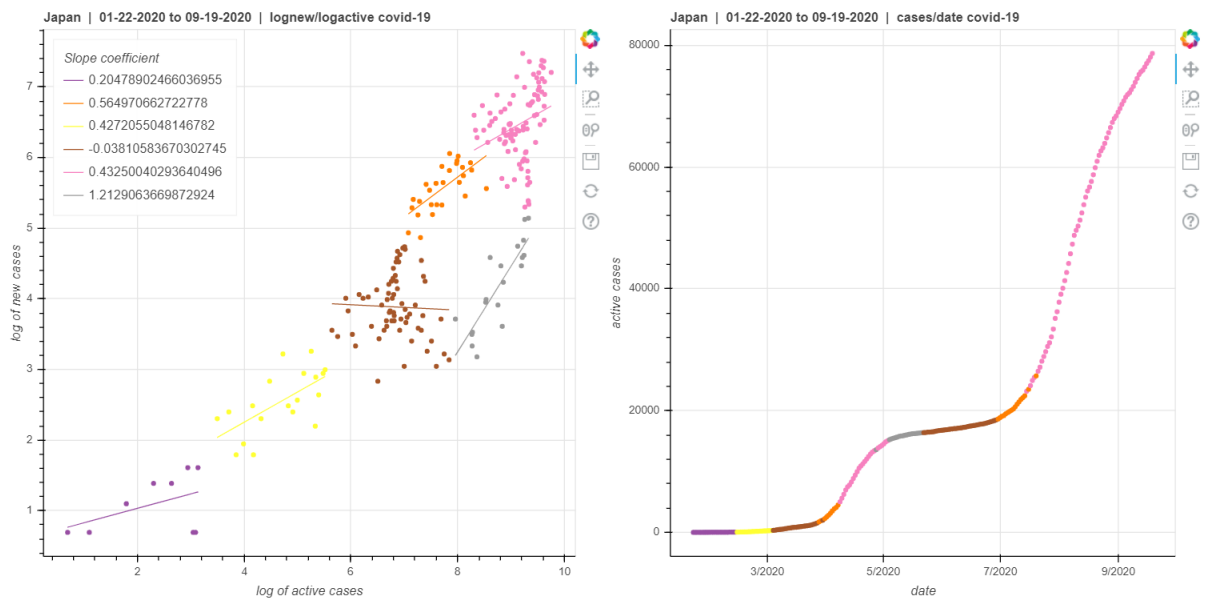
Slika 82. Bangladeš (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bangladeš): 1.135738400014731



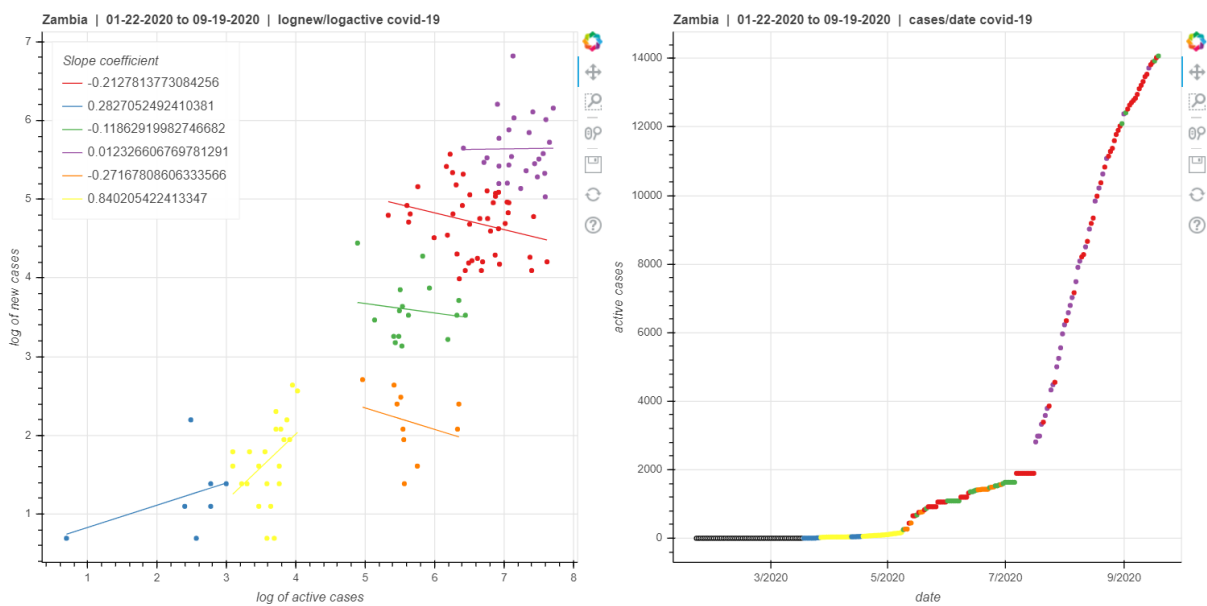
Slika 83. Švicarska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Švicarska): 1.1419899435008711



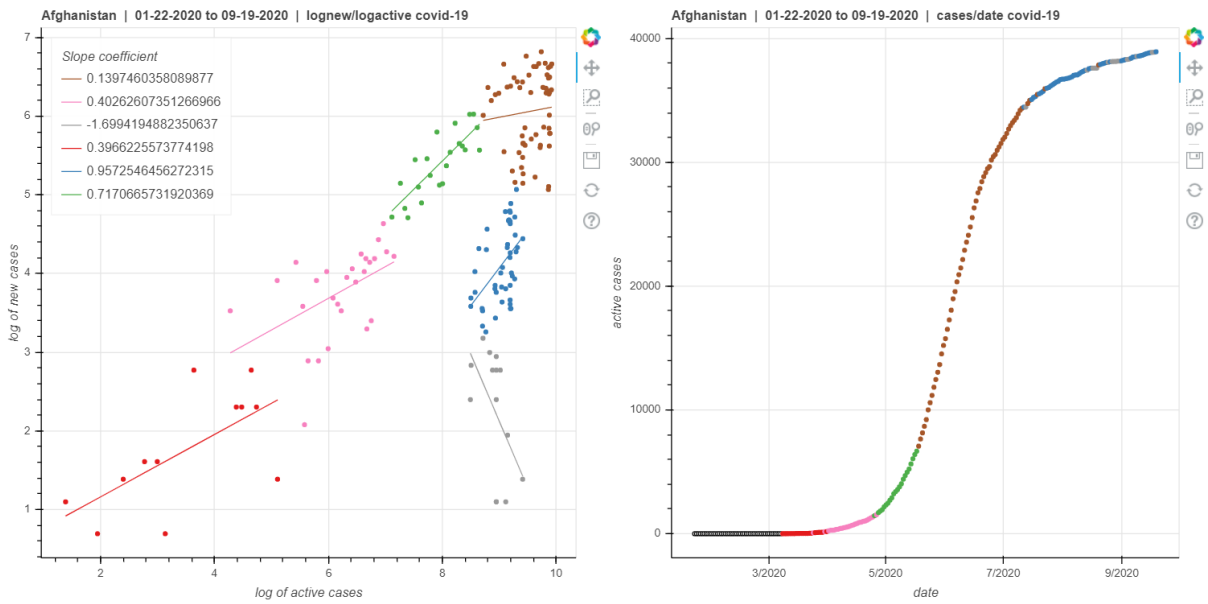
Slika 84. Japan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Japan): 0.7128259149938536



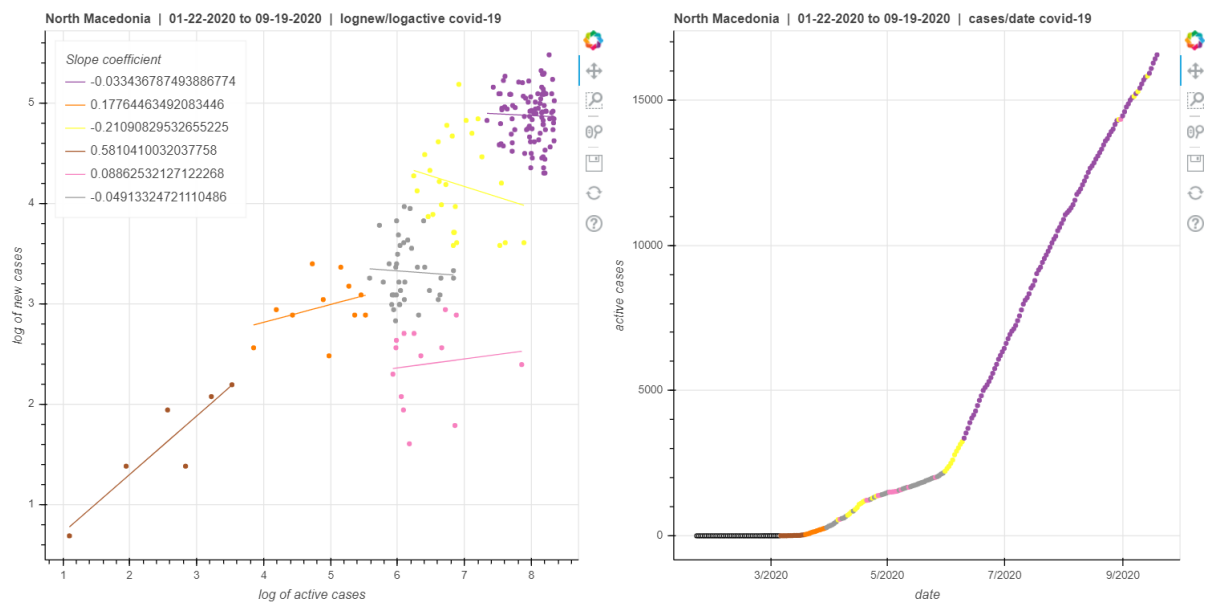
Slika 85. Zambija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Zambija): 1.0198249335896967



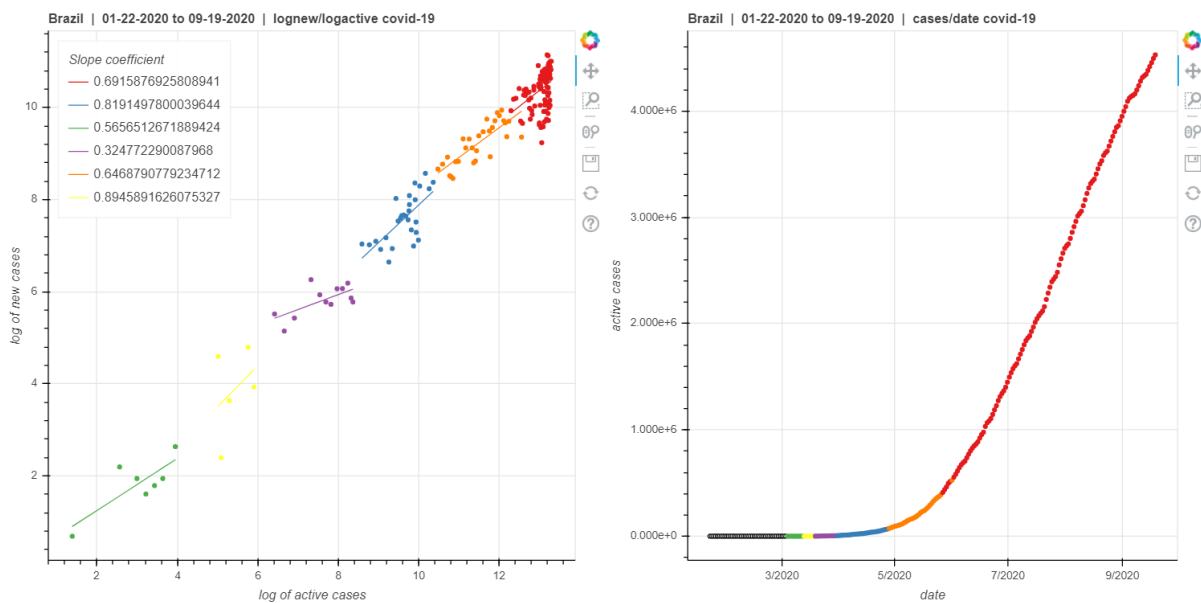
Slika 86. Afganistan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Afganistan): 0.944239941968269



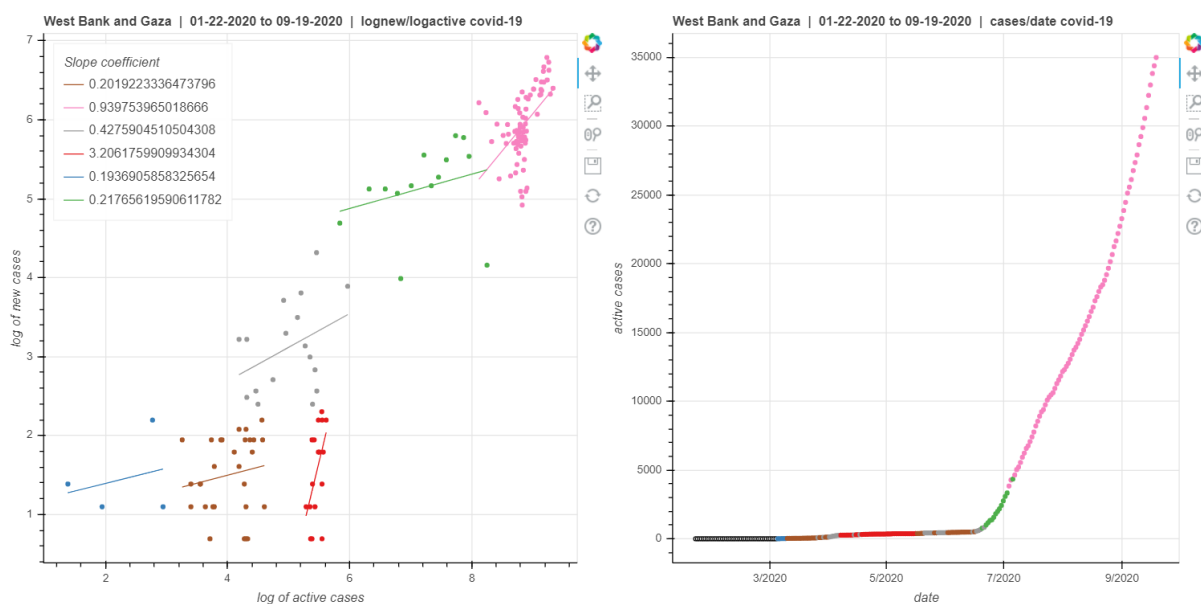
Slika 87. Sjeverna Makedonija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Sjeverna Makedonija): 1.0298291985100003



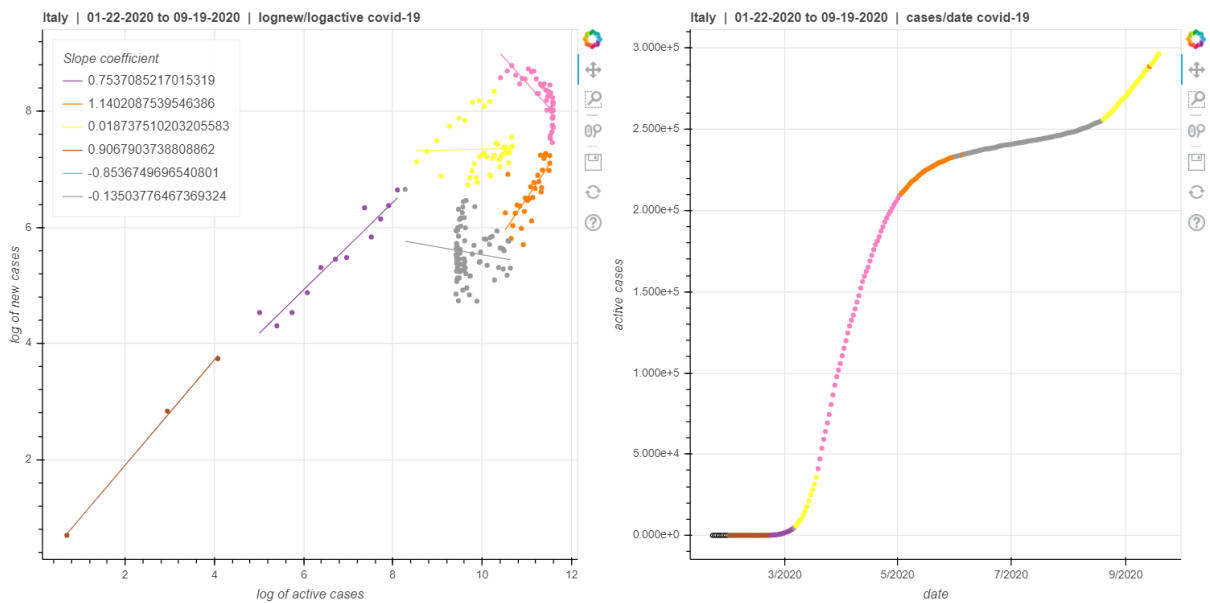
Slika 88. Brazil (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Brazil): 0.33752283571105735



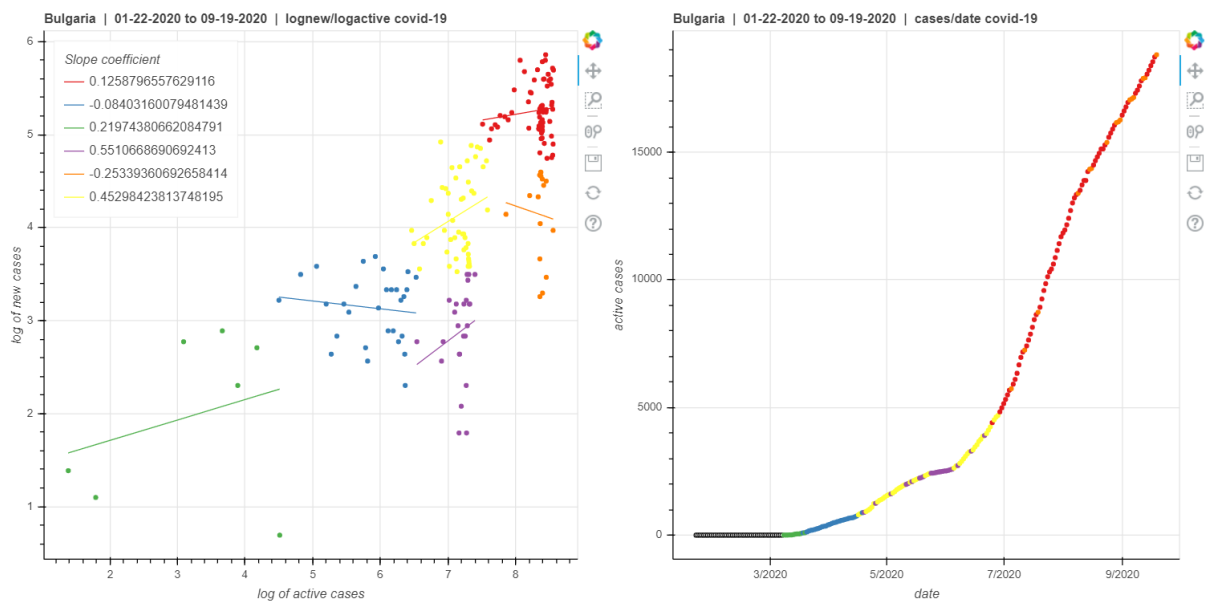
Slika 89. Pojas Gaze i Zapadna obala (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Pojas Gaze i Zapadna obala): 0.8650473840019192



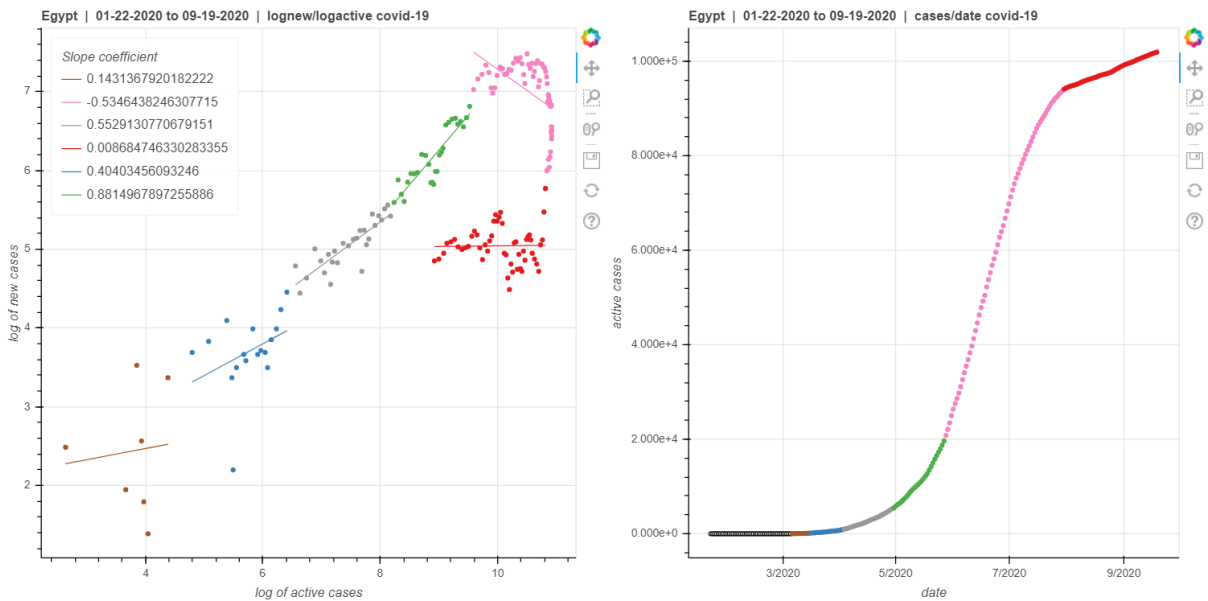
Slika 90. Italija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Italija): 1.1860146601177



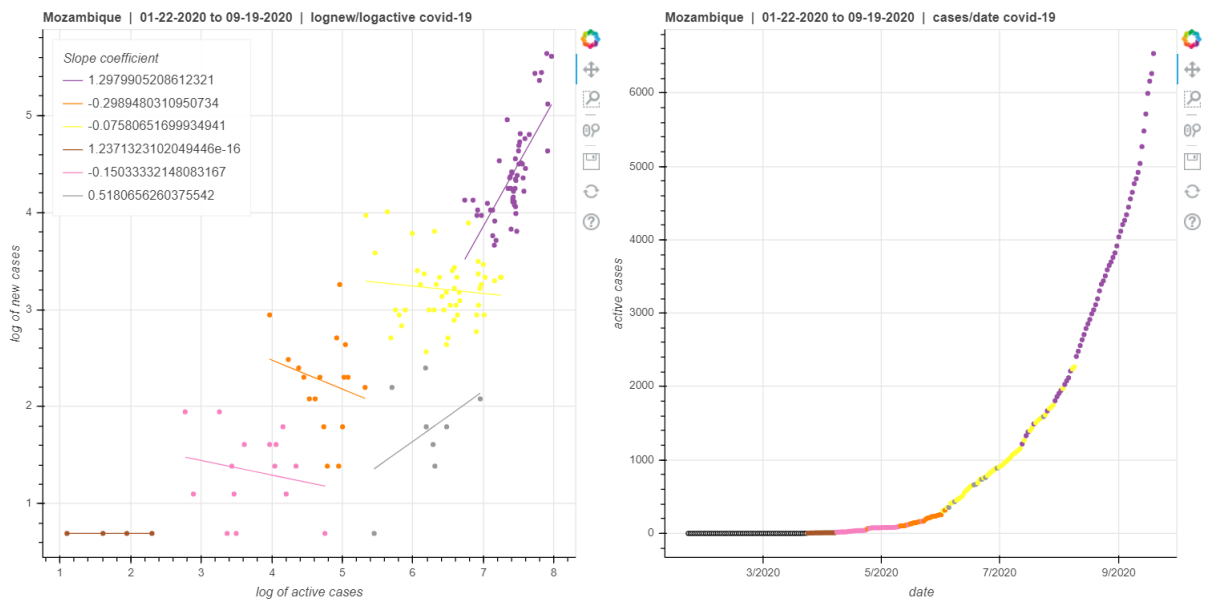
Slika 91. Bugarska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bugarska): 0.8369923628564906



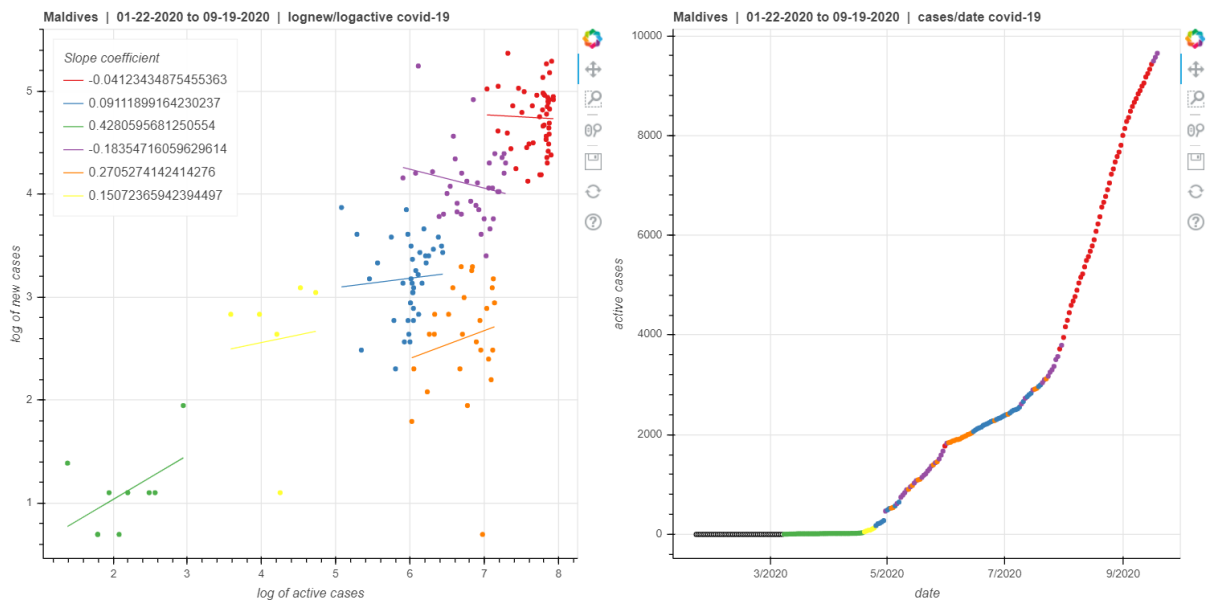
Slika 92. Egipat (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Egipat): 1.0596003150233906



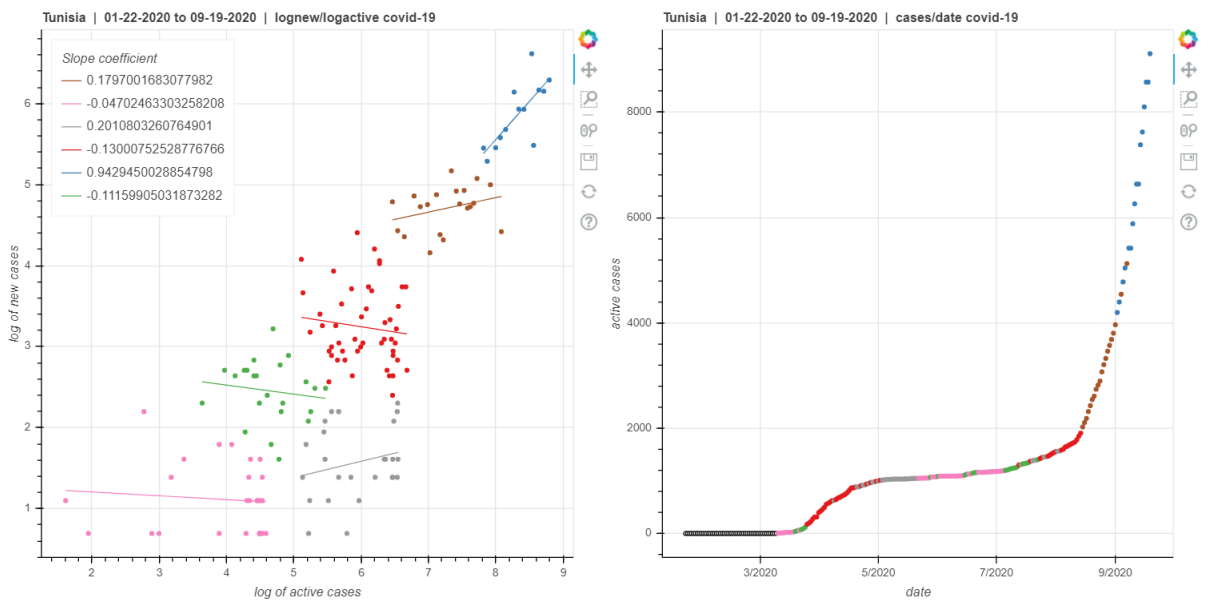
Slika 93. Mozambik (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Mozambik): 0.8850545434758721



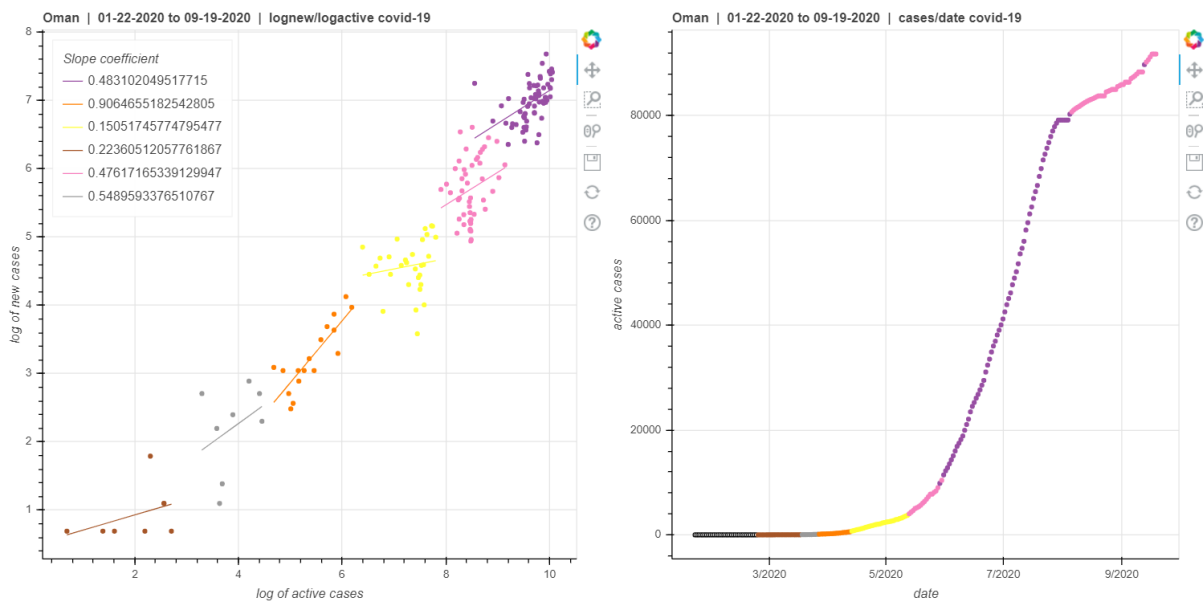
Slika 94. Maldivi (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Maldivi): 0.9740473545490694



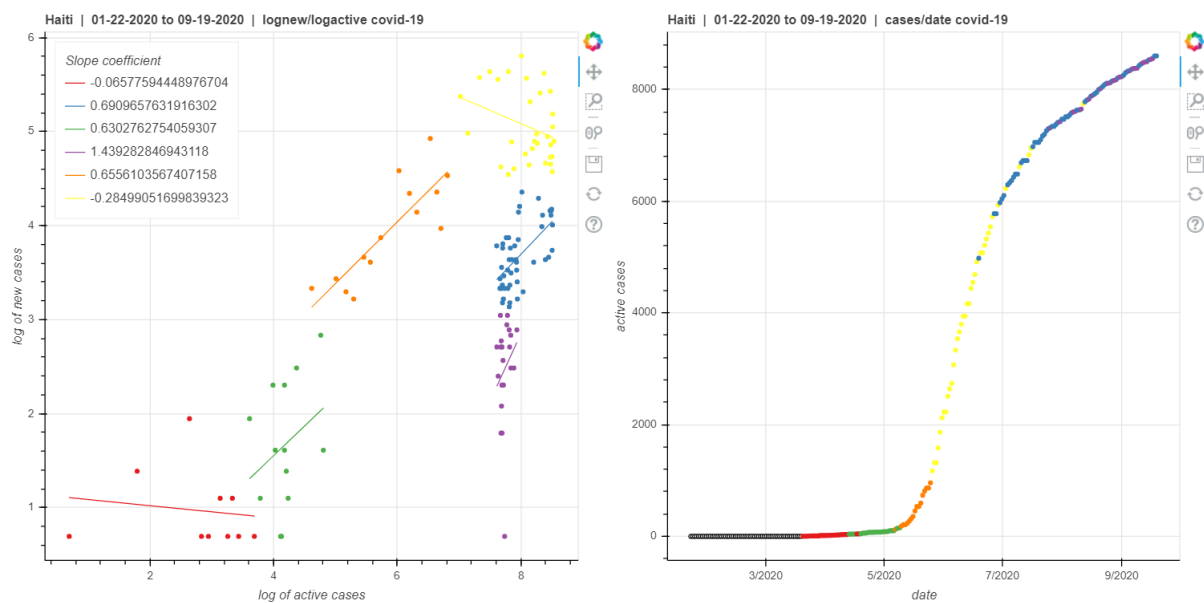
Slika 95. Tunis (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Tunis): 0.9765455512804325



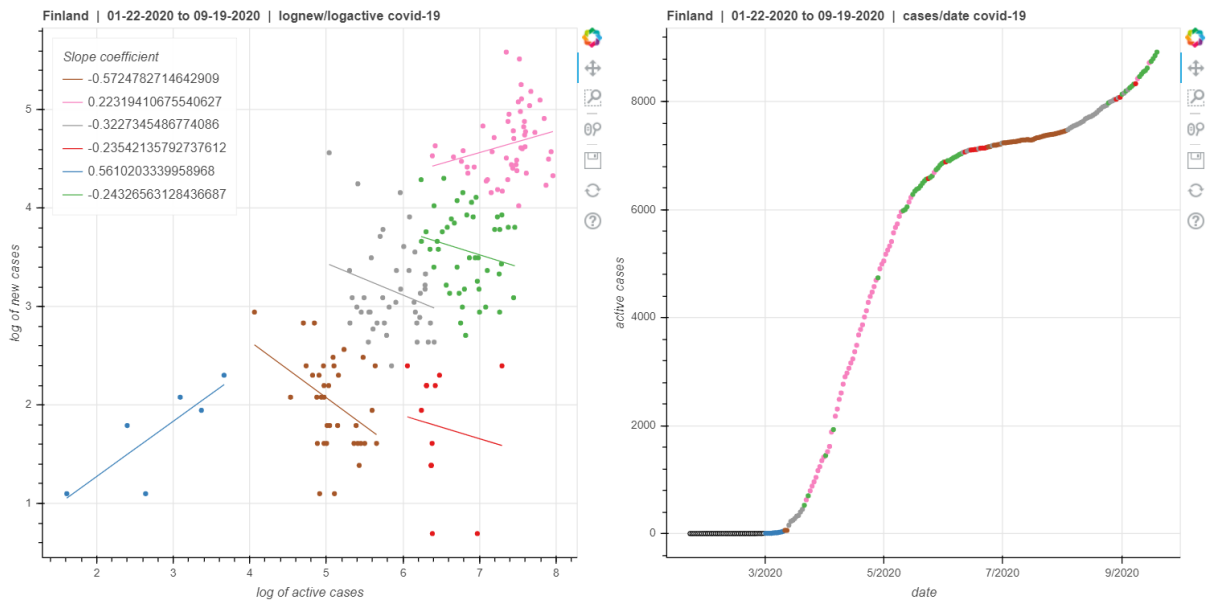
Slika 96. Oman (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Oman): 0.5782630003639373



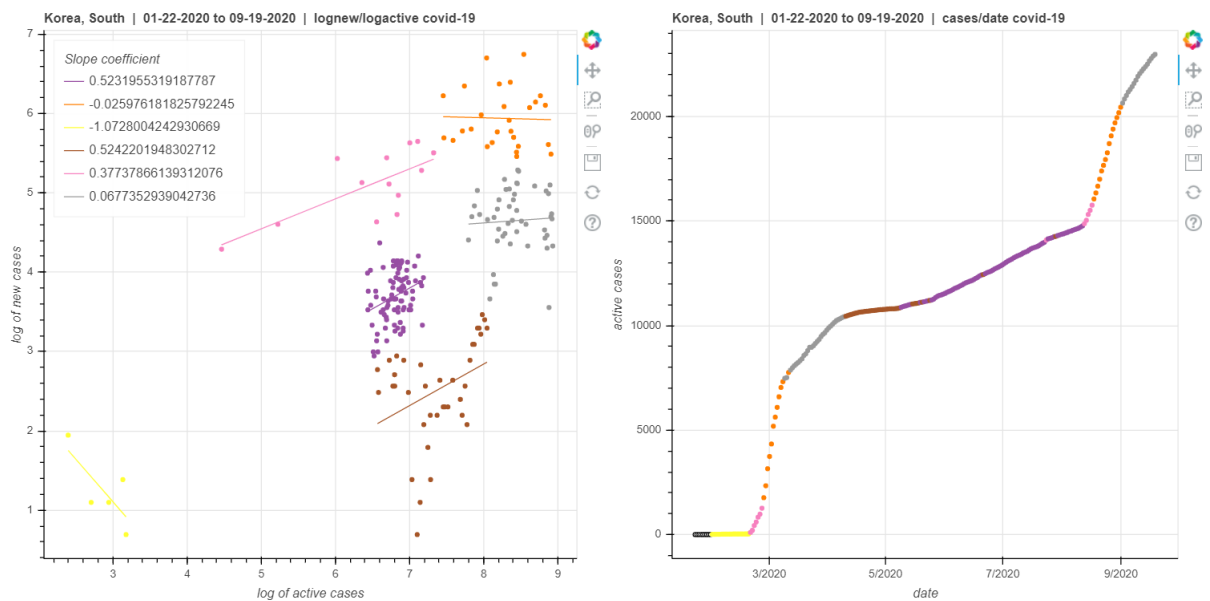
Slika 97. Haiti (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Haiti): 0.7461489094306812



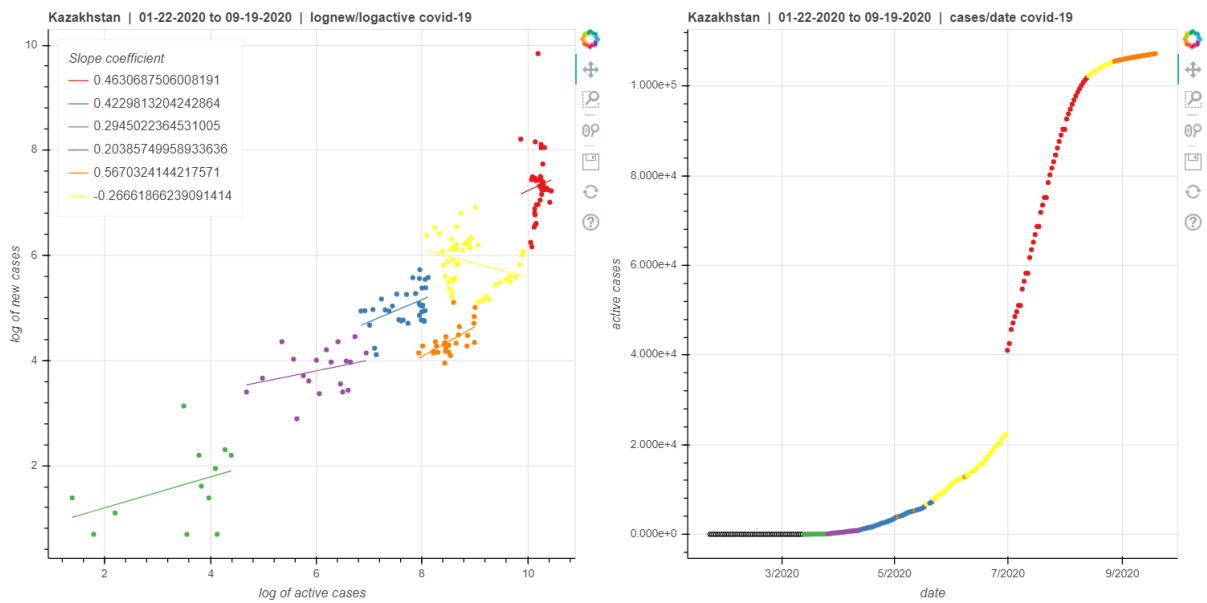
Slika 98. Finska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Finska): 1.2097888087313757



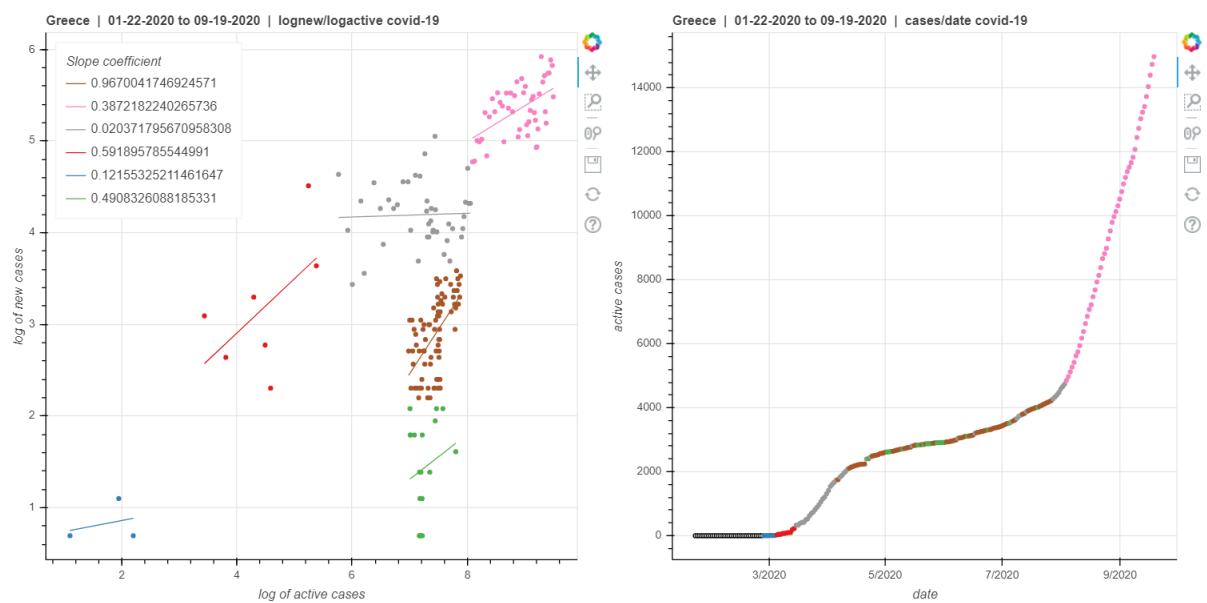
Slika 99. Južna Koreja (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Južna Koreja): 0.7584090188764212



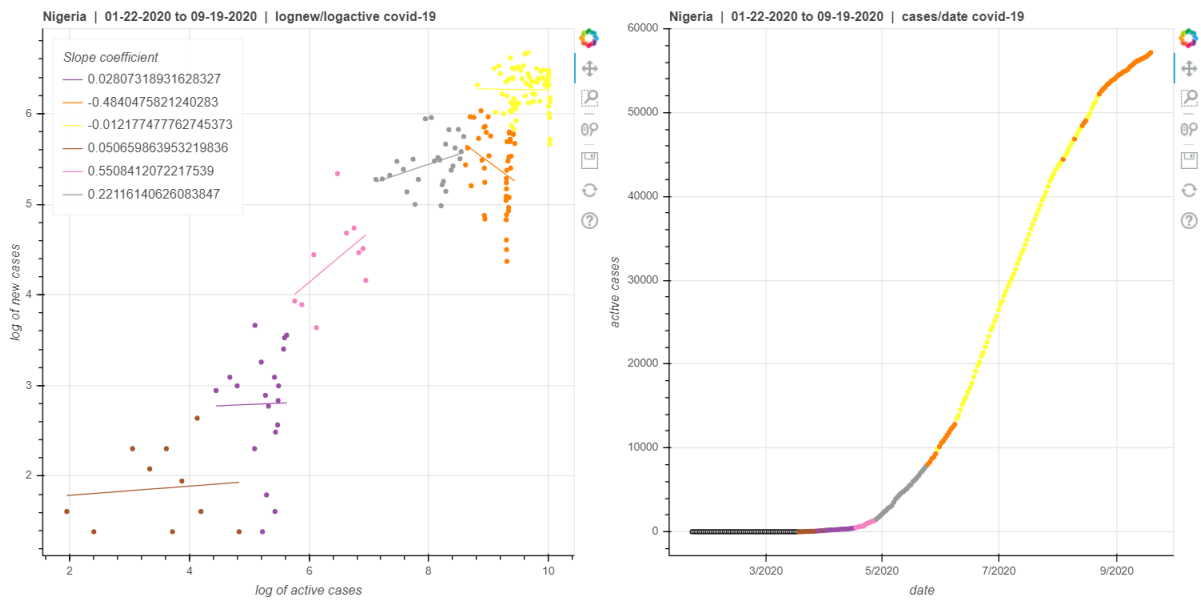
Slika 100. Kazahstan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kazahstan): 0.8294476658191999



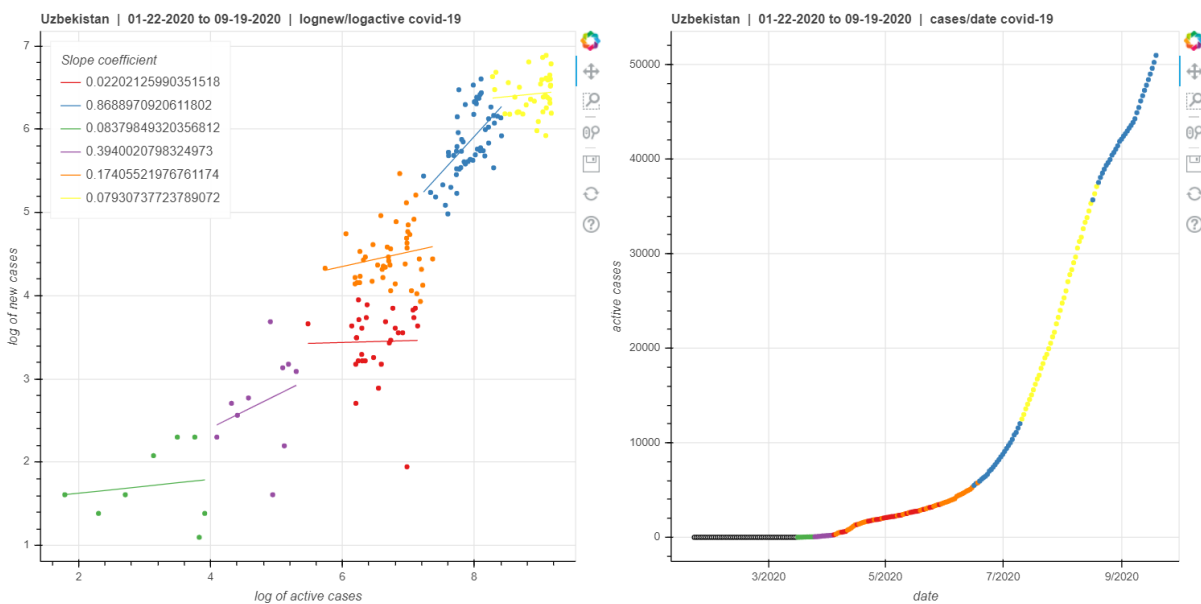
Slika 101. Grčka (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Grčka): 0.5852312082035794



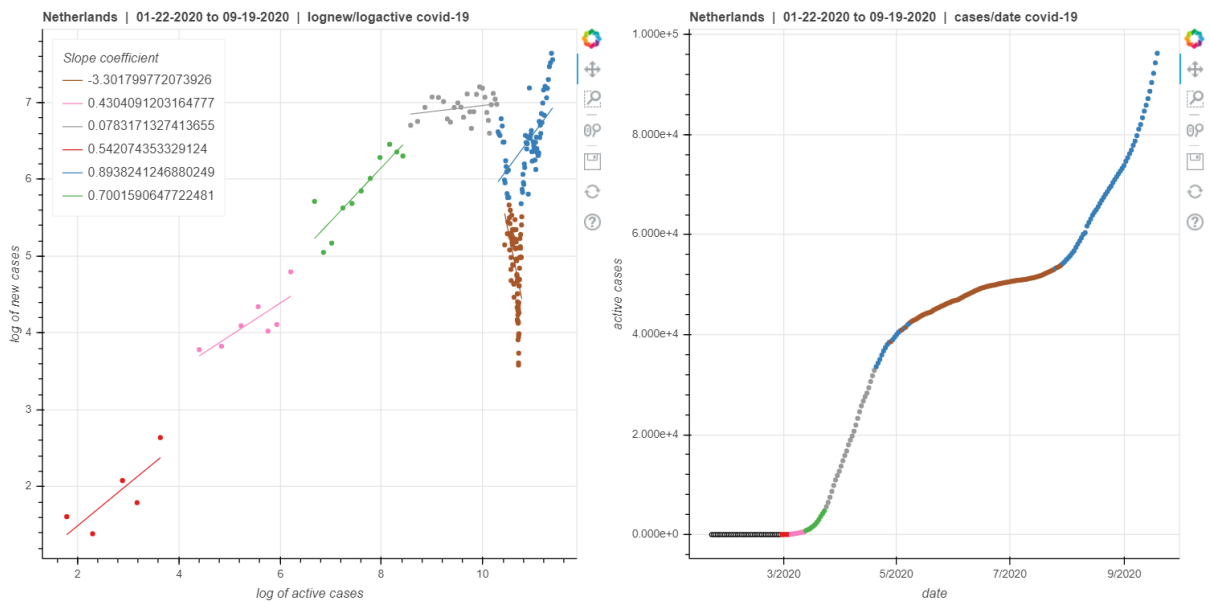
Slika 102. Nigerija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Nigerija): 1.0974486569212627



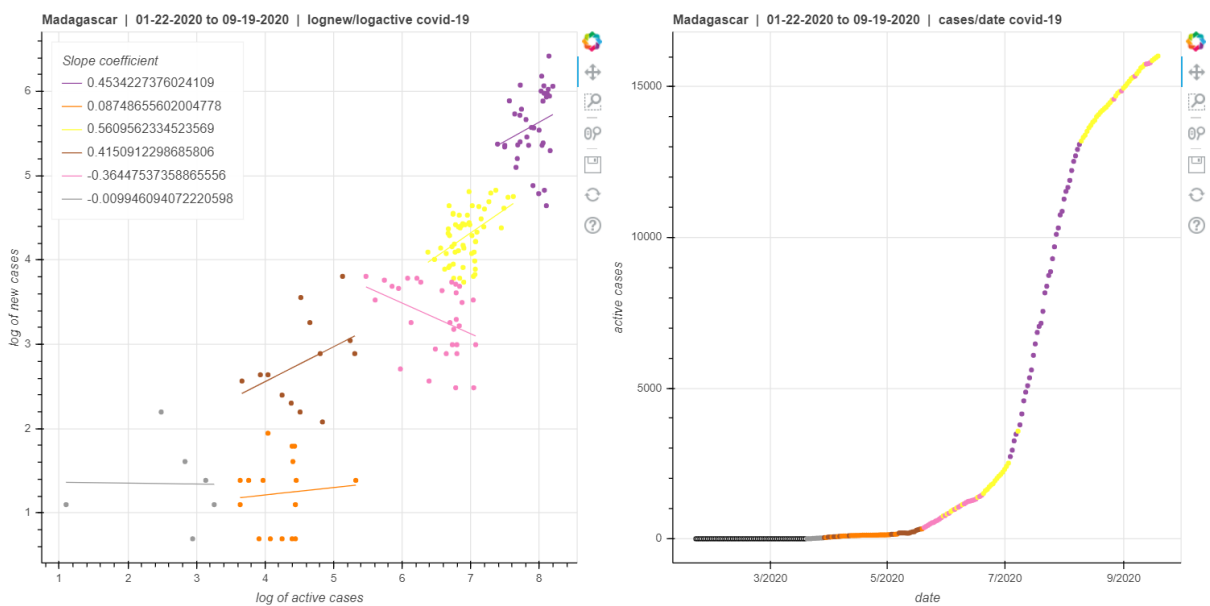
Slika 103. Uzbekistan (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Uzbekistan): 0.7425890068191426



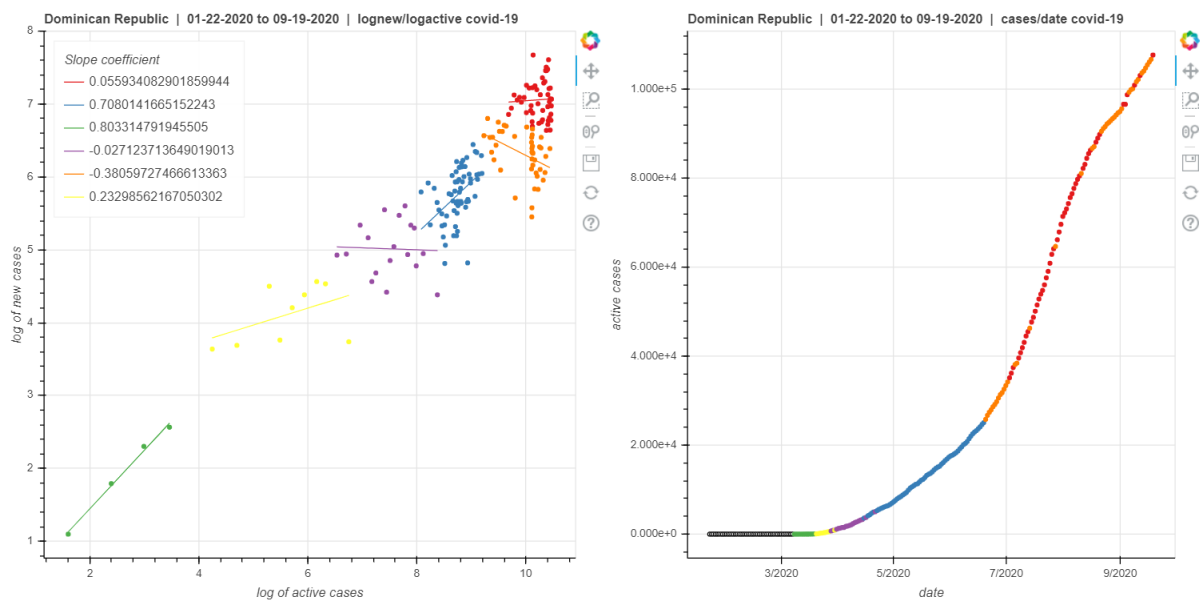
Slika 104. Nizozemska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Nizozemska): 2.792039382051389



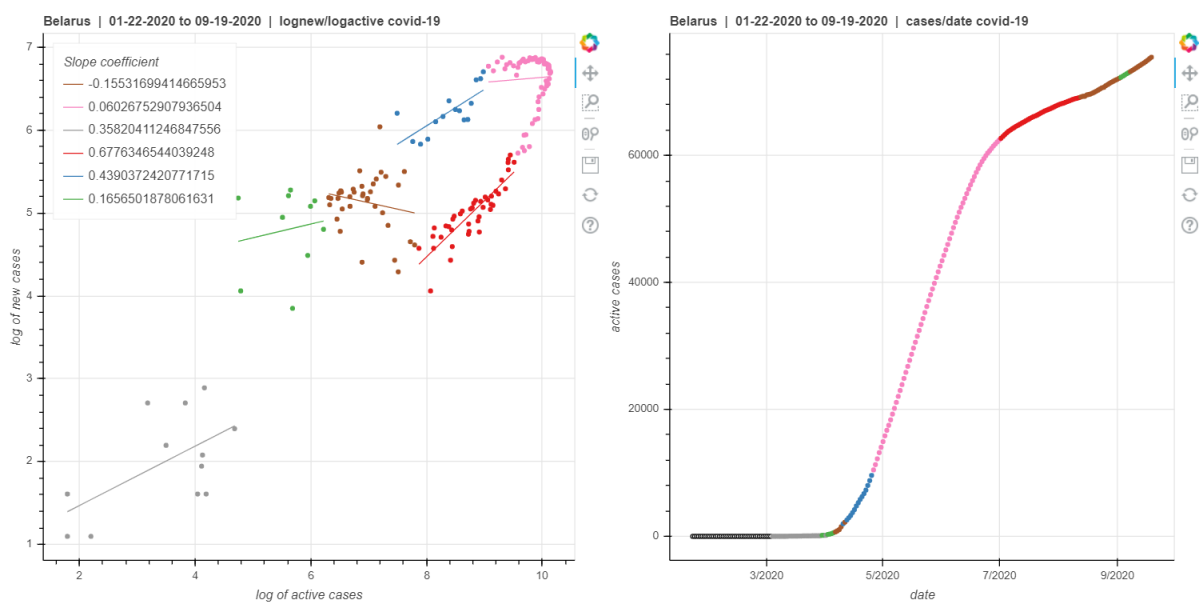
Slika 105. Madagaskar (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Madagaskar): 0.8014478073376488



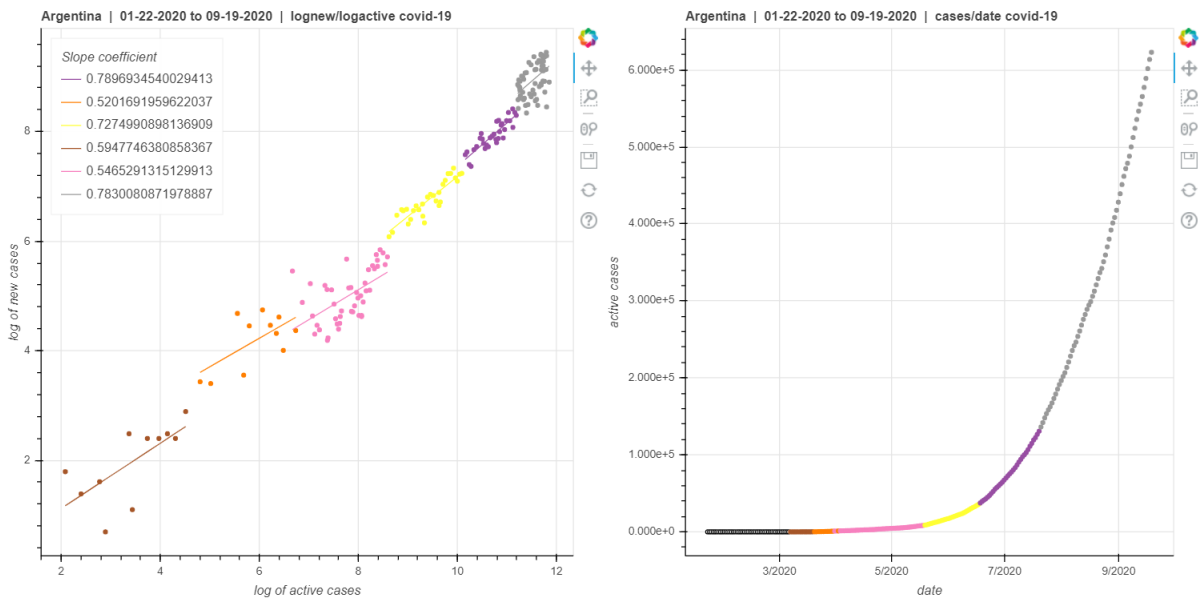
Slika 106. Dominikanska Republika (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Dominikanska Republika): 0.9117990143317909



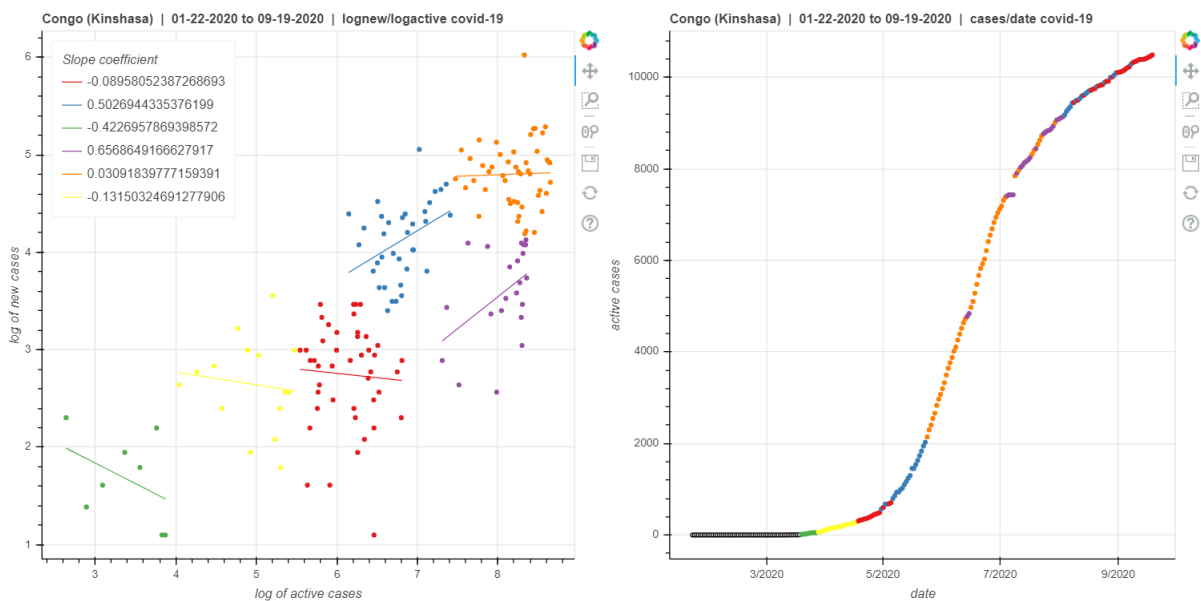
Slika 107. Bjelorusija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Bjelorusija): 0.8413577121963672



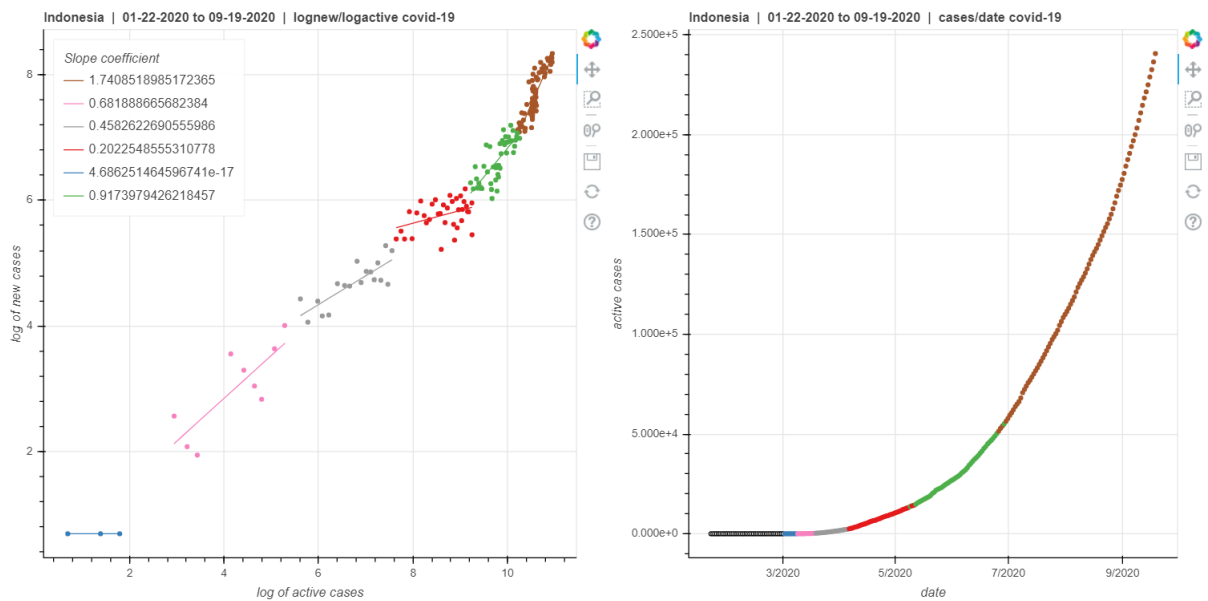
Slika 108. Argentina (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Argentina): 0.3248195514997901



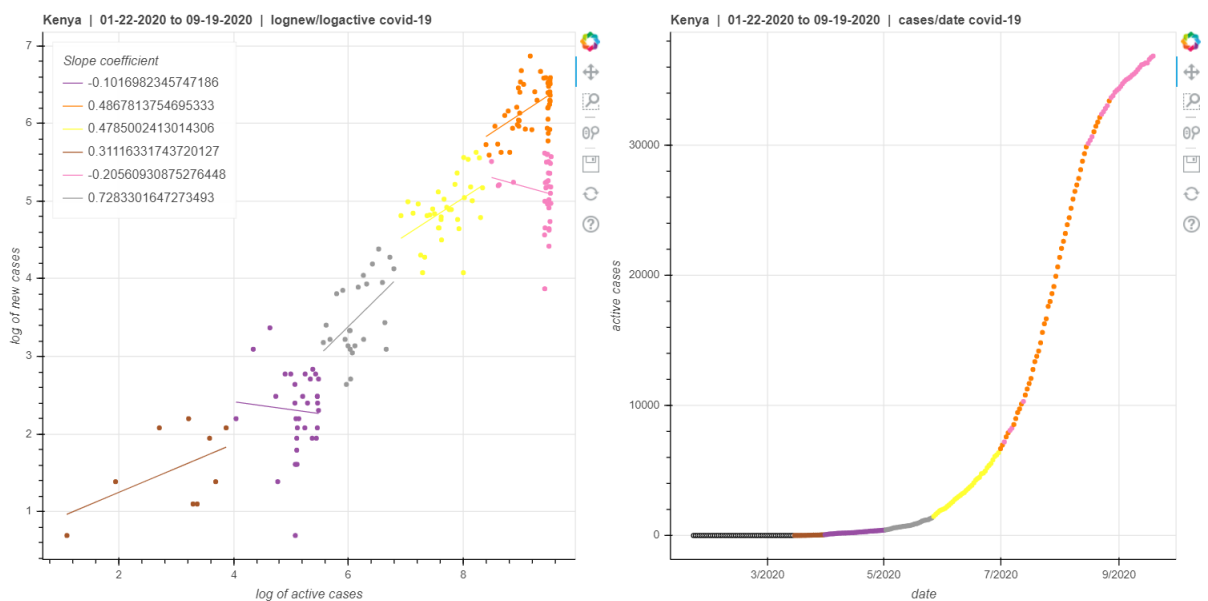
Slika 109. Demokratska Republika Kongo (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Demokratska Republika Kongo): 0.9148552982094338



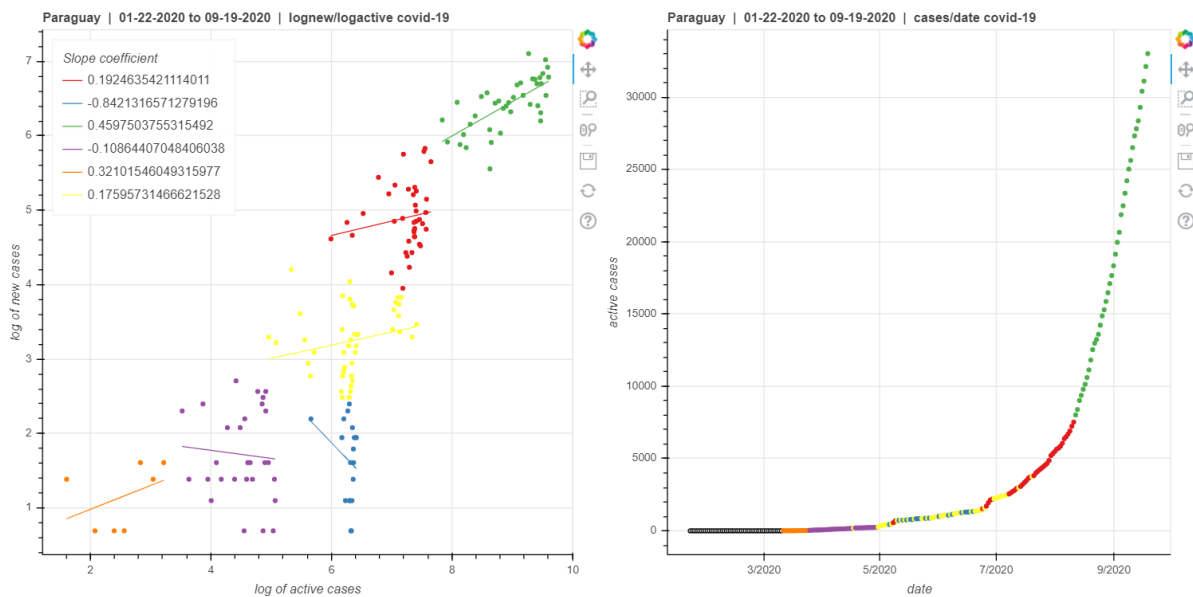
Slika 110. Indonezija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Indonezija): 0.6310140557371566



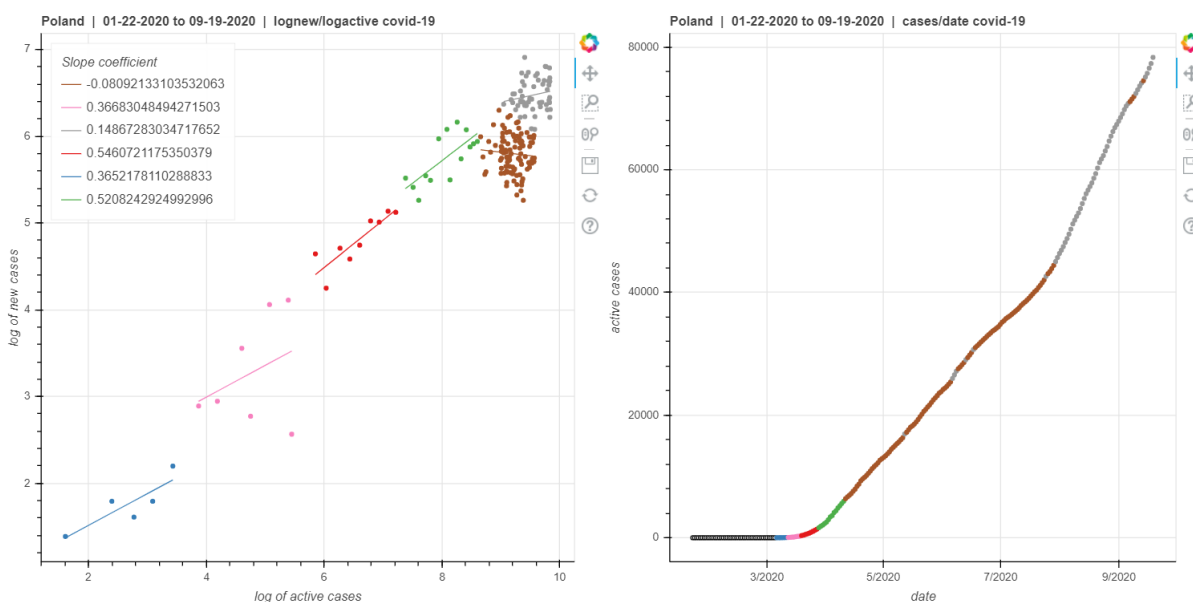
Slika 111. Kenija (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kenija): 0.7964655262336845



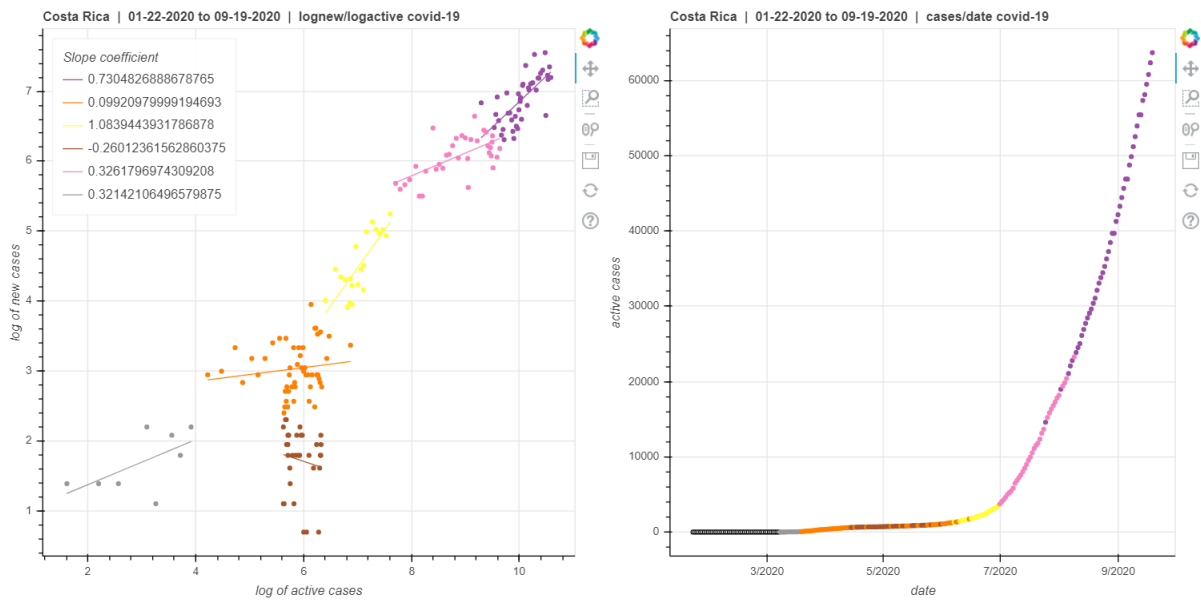
Slika 112. Paragvaj (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Paragvaj): 0.977677384716555



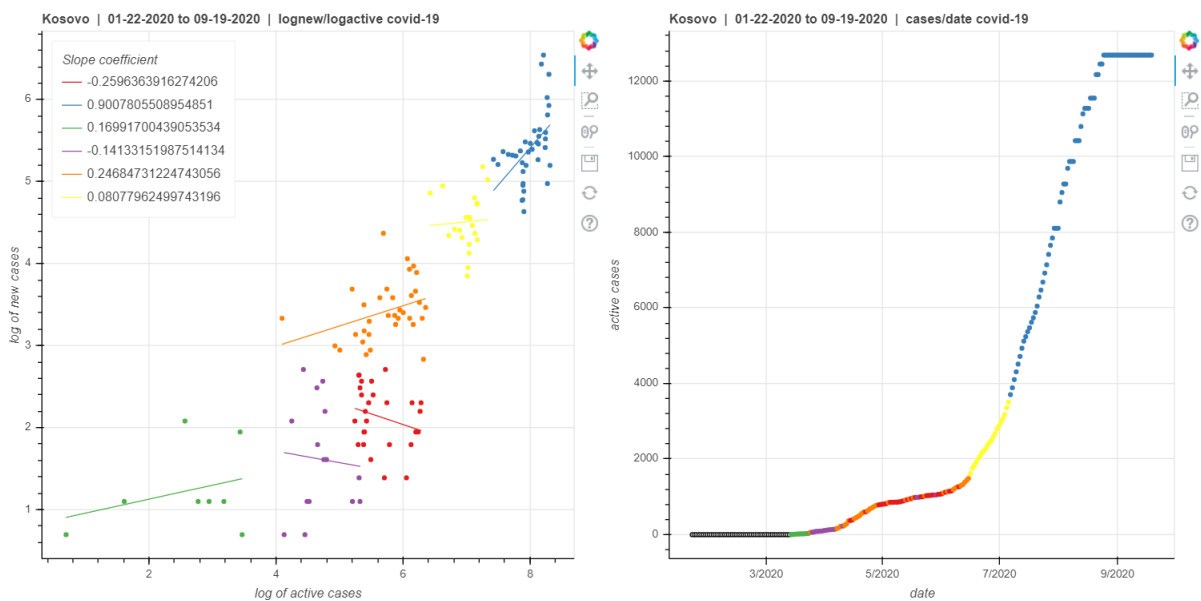
Slika 113. Poljska (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Poljska): 0.9418891519771821



Slika 114. Kostarika (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kostarika): 0.7774287354530822



Slika 115. Kosovo (a) log/log novo-zaraženih/zaraženih, (b) klasteri raspoređeni po ukupnim slučajevima

RMSD (Kosovo): 0.8610793668875262