

Potek širjenja do viška Covid-19 v prvem valu v Evropski uniji

Matej Kalc

28. julij 2020

1 Uvod

1.1 Motivacija

“Koronavirus je hujši kot vojna, kjer je sovražnik še vedno človek, s katerim se še vedno lahko ukvarjamo, medtem ko je kakršenkoli dogovor s smrtonosnim virusom, ki ogroža naše preživetje, nemogoč. (...)”.

Tako je izjavil G. Zuccarini. Lahko bi izjavili, da je Koronavirus tretja svetovna vojna, kjer se neviden sovražnik skriva med ljudmi. Ogroža ljudem življenje, nekaterim pa ga tudi odvzame. Ljudje lahko premagamo nevidnega sovražnika, le če primerno in pravočasno ukrepamo s pravim orožjem, kot so samozavest in ukrep človeka. V taki bitki tudi študiji in analize podatkov so dobro orožje proti virusu, saj izvemo nekaj novega o našem sovražniku. Mogoče eden izmed teh nam bom dal možnost odkritja cepiva zdravilo proti virusu, toda dokler tega ne ugotovimo ostaja edina možnost uporaba mask, razkužil in podobno. Zanima me kako so ljudje odzvali na epidemijo. Zanima me katere države so bile najboljše in katere najslabše organizirane za preprečevanje okužbe. Ker je epidemija še v teku, bom kot vzorec izbral države Evropske Unije, ker se je v teh epidemija sprožila približno sočasno.

1.2 Cilji

Trdimo lahko, da so vse države v Evropski uniji preživele prvi val Koronavirus pred 19. julijem 2020. V seminarski bom analiziral kako se je virus širil po državah Evropske unije. Predvsem bom analiziral interval od začetka širjenja do vrhunca prvega vala virusa v vsaki evropski državi, ker je ta interval najzanimivejši, saj so v tem intervalu evropske države prvič pod vplivom virusa. Zanima nas tudi kako je vsaka država Evropske unije začela testirati, tako da bo zmanjšala bodoče okužbe.

Cilj študija je analiza:

- Vpliva števila dni do viška dnevnih okuženih primerov na delež okuženih do viška,
- Deleža okuženih primerov v prvem delu prvega vala,
- Vpliva mediane starosti EU države na delež okuženih,
- Vpliva media starosti populacije EU države na fatalnost virusa in
- Vpliva števila dni do viška dnevnih okuženih primerov na delež opravljenih testov do viška.

Testiral bom korelacijo med spremenljivkami in izračunal bom intervale zaupanja, saj podatki niso realni, ker v teh niso vsebovani asimptomatiki.

1.3 Raziskave o virusu

Veliko je spletnih strani, ki analizirajo in grafično prikazujejo podatke Covid-19. Našel bom eno, ki me je motivirala za seminarsko.

Inštitut za zdravstvene meritve in vrednotenje IHME nudi spletno stran o Koronavirus, kjer so grafično prikazani podatki o okuženih, mrtvih, testih, socialna distanca ipd, ampak najzanimivejše so projekcije v času, ki stran nudi. IHME-ove projekcije COVID-19 so bile razvite kot odziv na zahteve Medicinske šole Univerze v Washingtonu in drugih ameriških bolnišničnih sistemov in vladnih držav, ki si prizadevajo za določitev, kdaj bo COVID-19 premagal njihovo sposobnost oskrbe bolnikov. Napovedi kažejo povpraševanje po bolnišničnih storitvah, dnevne in kumulativne smrti zaradi COVID-19, stopnje okužbe in testiranja ter vpliv socialne distanciranja, ki ga organizira država in država (za izbrane lokacije).

1.4 Poglavlja

1. Uvod
2. Opis virusa in njegovo širjenje
3. Podatki
4. Izračuni in rezultati
5. Zaključki
6. Literatura

2 Opis virusa in njegovo širjenje

COVID-19 je nalezljiva bolezen, ki jo povzroča virus SARS-CoV-2. Dihalni virus se širi preko kašlja in kihanja. Prvi okuženec Covid-19 je bil zaznan na Kitajskem novembra 2019. Najprej se je dihalni virus širil na Kitajskem, Hubei in Wuhan. Na začetku leta 2020 se je začelo širjenje virusa po celem svetu. 11. marca 2020 je Svetovna zdravstvena organizacija WHO proglasila pandemijo. Iz statističnih podatkov je razvidno, da do vključno 19. julija 2020 je bilo okuženih več kot 14.2 milijonov ljudi v 188 državah, od katerih 600 tisoč je mrtvih in 8.02 milijonov je ozdravelih. Trdimo lahko, da je ta virus leta 2020 močno vplival na države po celem svetu.

3 Podatki

Podatki so bili izbrani iz spleta. Podatke, ki bom rabil za statistični študij, so:

1. Seznam Evropskih držav
2. Mediana starosti populacije vsake Evropske države
3. Število prebivalcev vsake Evropske države

4. Dnevno število mrtvih v vsaki Evropski državi
5. Dnevno število okuženih v vsaki Evropski državi
6. Datum vrhunca števila okuženih prvega vala v vsaki Evropski državi
7. Datum vrhunca števila mrtvih prvega vala v vsaki Evropski državi
8. Datum prve zaznane okužbe virusa v vsaki Evropski državi
9. Datum prve zaznane smrti zaradi virusa v vsaki Evropski državi
10. Dnevno število opravljenih testov v vsaki Evropski državi
11. Število mrtvih do vrhunca prvega vala v vsaki Evropski državi
12. Število okuženih do vrhunca prvega vala v vsaki Evropski državi

Število mrtvih do vrhunca prvega vala v vsaki Evropski državi (9) je seštevek dnevno število mrtvih v vsaki Evropski državi(4) od datuma prvega mrtvega zaradi virusa do datum vrhunca števila mrtvih prvega vala v vsaki Evropski državi(7). Število okuženih do vrhunca prvega vala v vsaki Evropski državi (10) je seštevek dnevno število okuženih v vsaki Evropski državi(5) od datuma prvega mrtvega zaradi virusa do datum vrhunca števila okuženih prvega vala v vsaki Evropski državi(6). Nabrani podatki so vidni v spodnji tabeli.

DR	KOD	MED	DPO	DPS	DVO	DVS	OV	MV	PREB	OTO	OTS
Austria	AT	44.0	2020-02-25	2020-03-12	2020-03-27	2020-04-23	7029	494	9025715	38809	201454
Belgium	BE	41.4	2020-02-04	2020-03-10	2020-04-11	2020-04-12	32778	4616	11602522	103714	109427
Bulgaria	BG	42.7	2020-03-08	2020-03-12	2020-06-12	2020-06-06	3086	160	6943915	91083	81084
Croatia	HR	43.0	2020-02-25	2020-03-25	2020-04-02	2020-04-20	963	47	4101782	8110	25566
Cyprus	CY	36.8	2020-03-09	2020-03-25	2020-04-02	2020-03-25	320	3	1190007	8468	3849
Czechia	CZ	42.1	2020-03-01	2020-03-23	2020-03-27	2020-04-15	2062	161	10715154	36089	148586
Denmark	DK	42.2	2020-02-27	2020-03-15	2020-04-08	2020-04-05	5071	161	5793679	62063	50097
Estonia	EE	42.7	2020-02-27	2020-03-26	2020-03-27	2020-04-03	538	11	1328655	9010	18172
Finland	FI	42.5	2020-01-29	2020-03-21	2020-04-05	2020-04-22	1882	141	5542713	34486	76173
France	FR	41.4	2020-01-24	2020-02-15	2020-04-01	2020-04-04	51477	6493	65283211	233494	282205
Germany	DE	47.1	2020-01-28	2020-03-09	2020-03-20	2020-04-16	18323	3569	83951077	595836	2019592
Greece	GR	44.5	2020-02-26	2020-03-12	2020-04-22	2020-04-05	2401	68	10420046	58847	26200
Hungary	HU	42.3	2020-03-04	2020-03-15	2020-04-10	2020-04-24	1190	250	9659639	29041	57641
Ireland	IE	36.8	2020-03-01	2020-03-11	2020-04-10	2020-04-26	7393	1063	4953657	68922	142512
Italy	IT	45.5	2020-01-29	2020-02-22	2020-03-21	2020-03-28	53578	9136	60465251	239558	428323
Latvia	LV	43.6	2020-03-02	2020-04-04	2020-03-24	2020-04-22	180	9	1883138	8281	40057
Lithuania	LT	43.7	2020-02-28	2020-03-20	2020-04-04	2020-04-12	771	23	2714541	21467	40951
Luxembourg	LU	39.3	2020-03-01	2020-03-13	2020-03-24	2020-04-12	875	62	628614	11189	29881
Malta	MT	41.8	2020-03-07	2020-04-09	2020-04-08	2020-06-02	293	9	441612	14119	73236
Netherlands	NL	42.6	2020-02-27	2020-03-06	2020-03-24	2020-04-08	4749	2101	17138553	45825	109414
Poland	PL	40.7	2020-03-05	2020-03-12	2020-06-05	2020-04-25	25048	494	37850596	1006819	271678
Portugal	PT	42.2	2020-03-02	2020-03-17	2020-04-11	2020-04-04	15472	246	10193282	179542	112892
Romania	RO	41.1	2020-02-26	2020-03-23	2020-04-12	2020-05-01	5990	717	19210031	64385	175728
Slovakia	SK	40.5	2020-03-06	2020-04-07	2020-04-17	2020-04-16	977	6	5461415	42768	40048
Slovenia	SI	44.5	2020-03-04	2020-03-17	2020-03-13	2020-04-06	141	28	2079553	4228	28453
Spain	ES	42.7	2020-01-31	2020-03-04	2020-04-01	2020-06-20	94417	28315	46785134	466271	3627852
Sweden	SE	41.2	2020-01-31	2020-03-15	2020-06-23	2020-04-22	58932	1765	10108080	467798	105806

Legenda:

- DR - Ime države
- KOD - Koda države
- MED - Mediana starosti populacije

- DPO - Datum prvega zaznanega okuženca
- DPS - Datum prve zaznane smrti
- DVO - Datum vrhunca okuženih v prvem valu
- DVS - Datum vrhunca smrti v prvem valu
- OV - Število okuženih od prve zaznane okužbe do vrhunca okuženih v prvem valu
- MV - Število mrtvih od prve zaznane smrti do vrhunca smrti v prvem valu
- PREB - Število prebivalcev
- OTO - Število opravljenih testov do vrhunca okužb v prvem valu
- OTS - Število opravljenih testov do vrhunca smrti v prvem valu

4 Izračuni in rezultati

Podatki iz prvega vala nam povejo koliko je bila država pripravljena na tako epidemijo.

4.1 Vpliv števila dni do viška dnevnih okuženih primerov na delež okuženih do viška

Zanima me kako število dni do viška dnevnih okuženih primerov vpliva na delež okuženih do viška. Naj bosta

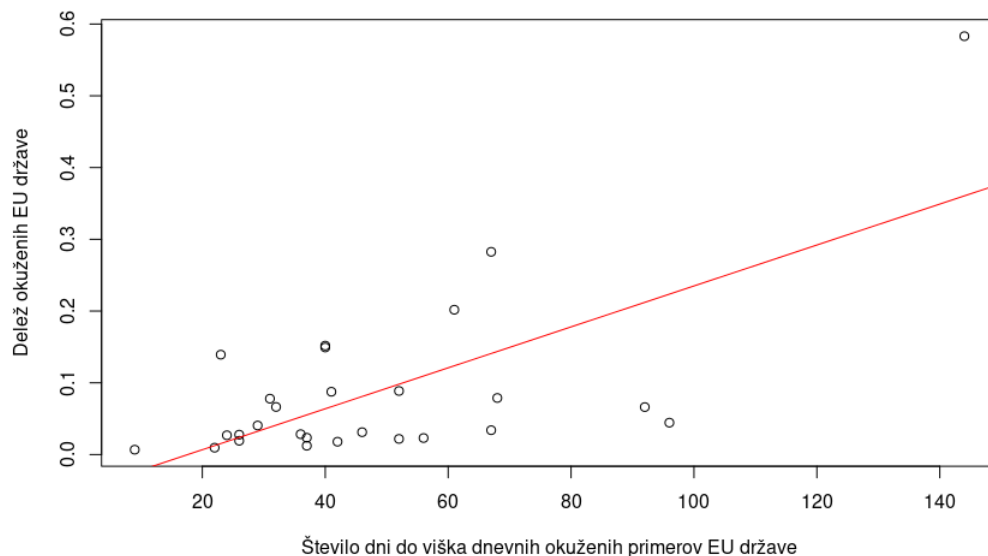
število dni do viška = datum vrhunca okuženih – datum prvega zaznanega okuženca

in

$$\text{delež okuženih} = \frac{100 \times \text{število okuženih}}{\text{število prebivalcev}}.$$

Naj bo spremenljivka N število dni do viška prvega vala, ki je razlika v dnevih datuma vrhunca okuženih v prvem valu (stolpec DVO v bazi) in datuma prvega zaznanega okuženca (stolpec DPO v bazi). Naj bo spremenljivka D delež okuženih, ki je razmerje med številom okuženih (stolpec OV v bazi) in številom prebivalcev (stolpec PREB v bazi). Spodnji graf prikazuje poidatke o spremenljivkah.

Vpliv števila dni do viška dnevnih okuženih primerov na delež okuženih v prvi polovici prvega



Iz grafa je razvidno da obstaja nek trend. Izračunajmo povezanost spremenljivk s Pearsonovim koeficientom korelacije:

$$r = \frac{Cov(N, D)}{\sigma_N \sigma_D} = 0.676$$

kjer je σ_N standardni odklon spremenljivke N in σ_D standardni odklon spremenljivke D. Koeficient ni dovolj velik, da bi lahko smatrali, da obstaja močna povezanost med spremenljivkima.

4.2 Delež okuženih primerov v prvem delu prvega vala

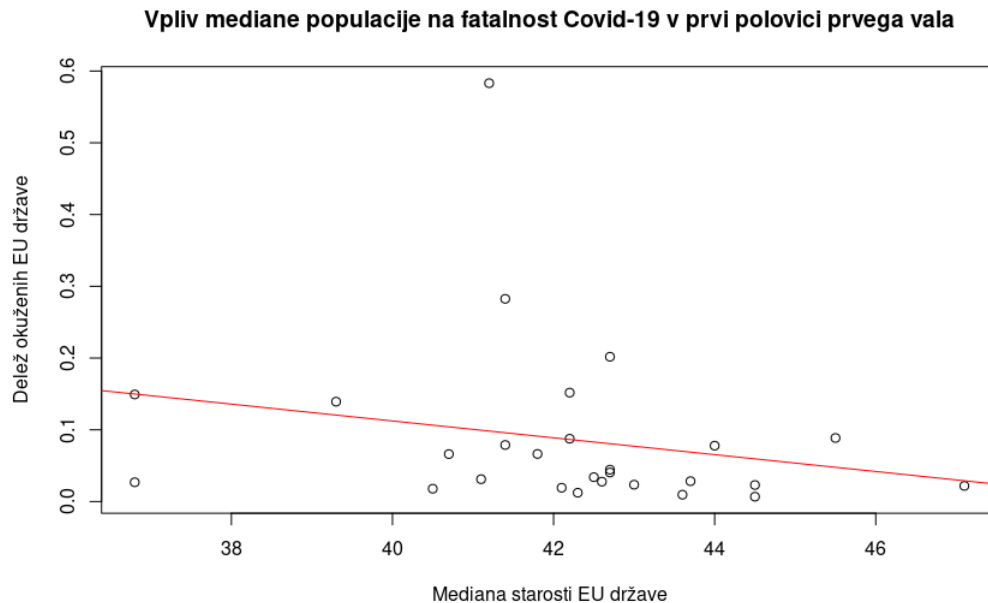
Iz analize opazamo, da obstaja neka povezanost med številom dni do viška okuženih primerov in deležom okuženih. Ker je veliko asimptomatikov, je pravo število okuženih neznano.

4.3 Vpliv mediane starosti EU države na delež okuženih

Zanima me kako se je širjenje razlikovalo med Evropskimi državami. Najprej definiramo delež okuženih.

$$\text{delež okuženih} = \frac{100 \times \text{število okuženih}}{\text{število prebivalcev}}$$

Naj bo spremenljivka S mediana starosti populacije (stolpec MED v bazi). Naj bo spremenljivka D delež okuženih, ki je razmerje med številom okuženih (stolpec OV v bazi) in številom prebivalcev (stolpec PREB v bazi). Izbrani podatki so podatki iz intervala od datuma prvega zaznanega okuženca do datuma vrhunca okuženih v prvem valu (Stolpca DPO in DVO v bazi v poglavju 3). Podatke iz baze prikažemo v grafu.



Na prvi pogled trend ne obstaja, kljub temu da smo vzeli primeren vzorec, vse države evropske unije. Graf nam, da slutiti, da sta spremenljivki S in D neodvisni. Z računi lahko ugotovimo več o povezanosti zgornjih spremenljivk. Izračunajmo Pearsonov koeficient korelacije:

$$r = \frac{Cov(S, D)}{\sigma_S \sigma_D} = -0.22$$

kjer je σ_S standardni odklon spremenljivke S in σ_D standardni odklon spremenljivke D. Koeficient je zelo blizu ničli, kar pomeni, da njihova linearna povezanost je zelo majhna. Mogoče pa sta si neliarno povezani. To lahko preverimo s Spearmanovim koeficientom korelacije. Izračunajmoga takole:

$$\rho = 1 - \frac{6 \sum_i d_i^2}{N(N^2 - 1)} = -0.3254892$$

kjer je d_i razlika med rangoma za i-to enoto in N pa število vseh enot (parov rangov). Kot Pearsonov koeficient tudi Spearmanov koeficient je zelo blizu ničli. Trdimo lahko, da spremenljivki S in F si nista močno povezani. To seveda še ne pomeni, da sta spremenljivki neodvisni. Izvedemo lahko f-test. Najprej nastavimo ničelno hipotezo H_0 :

H_0 : Spremenljivki S (mediana starosti populacije) in D (delež okuženih) sta si neodvisni

Računamo R^2 (Coefficient of determination) in F:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} = 0.04852426$$

kjer so y_i podatki, ki jih imamo na razpolago, \hat{y}_i pa vrednosti, ki bi jih imela odvisna spremenljivka D (delež okuženih) v primeru, da bi njene vrednosti ležale natanko na regresijski premici, ki smo jo dobili z metodo najmanjših kvadratov.

$$F = \frac{R^2}{(p-1)(1-R^2)(n-p)} = 0.00204$$

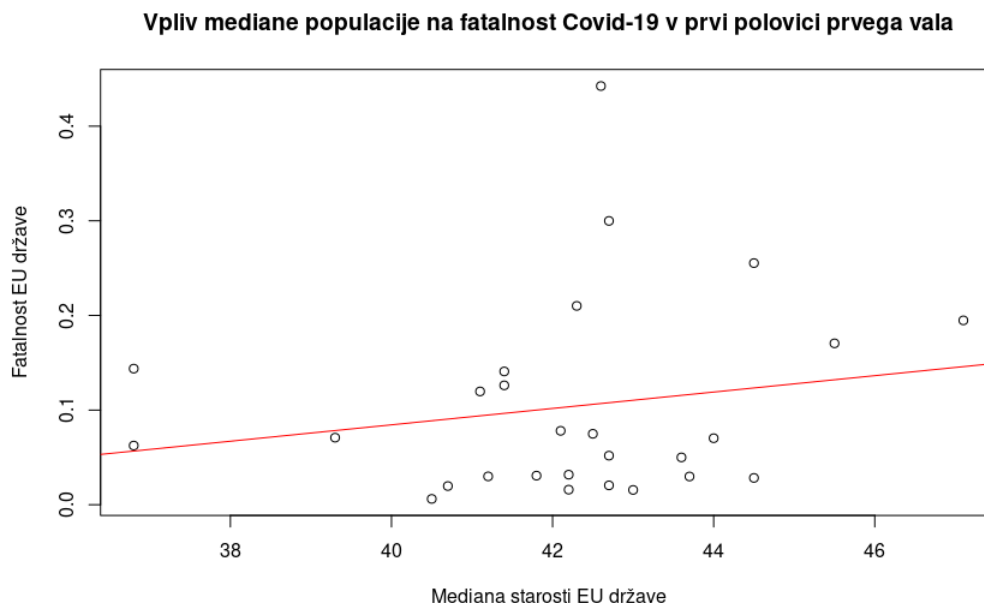
kjer p je število spremenljivk, R^2 je Koeficient odločnosti in n je število držav evropske unije. Ničelno hipotezo H_0 lahko zavrremo, če $F \geq F_t$. F_t je vrednost, ki preberemo iz porazdelitvene tabele distribucije F . Izberemo α (significance level) = 0.05. Iz tabele lahko razberemo, da je $F_t = 4.2417$ (Gledamo stolpec $p - 1 = 1$ in vrstico $n - 2 = 25$). Velja da, $F < F_t$, torej ne moremo zavreči ničelne hipoteze H_0 . Ker je vrednost F veliko manjša od vrednosti F_t , si lahko predstavljamo, da sta si spremenljivki S in F neodvisni.

4.4 Vpliv mediane starosti EU države na fatalnost virusa Covid-19

Začetna analiza je privleka do zanimivih rezultatov. Sprašujem se, če enako velja za fatalnost virusa. Opravi bom podobno analizo kot v poglavju 4.1. Je fatalnost virusa pod vplivom mediane starosti populacije? Najprej definiramo fatalnost.

$$\text{fatalnost} = \frac{100 \times \text{število mrtvih}}{\text{število okuženih}}$$

Naj bo fatalnost prve polovice prvega vala razmerje med številom mrtvih prve polovice prvega vala (stolpec MV v bazi) in številom okuženih prve polovice prvega vala (stolpec OV bazi). Za analizo vzamemo spremenljivki F in S , kjer F predstavlja fatalnost in S mediano starosti države. Spremenljivki F in S vsake države evropske unije sta prikazani v spodnjem grafu.



Gledano navidezno tudi v spodnjem grafu ni nobenega trenda. Na prvi vtis sta si spremenljivki neodvisni. To lahko potrdim s sledečimi računi. Izračunamo Pearsonov koeficient korelacije, ki nam pove linearno povezanost spremenljivk S in F :

$$r = \frac{Cov(S, F)}{\sigma_S \sigma_F} = 0.22$$

kjer je σ_S standardni odklon spremenljivke S in σ_F standardni odklon spremenljivke F. Koeficient je zelo majhen, kar pomeni, da njihova linearna povezanost je zelo majhna. Kot v prejšnji analizi, bomo izračunali neliarno povezanost. To lahko preverimo s Spearmanovim koeficientom korelacije. Izračunajmo:

$$\rho = 1 - \frac{6 \sum_i d_i^2}{N(N^2 - 1)} = 0.27$$

kjer je d_i razlika med rangoma za i -to enoto in N pa število vseh enot (parov rangov). Spearmanov koeficient je zelo majhen. Trdimo lahko, da spremenljivki S in F si nista močno povezani. To seveda še ne pomeni, da sta spremenljivki neodvisni. Kot v prejšnji analizi, bomo izvedli f-test. Najprej nastavimo ničelno hipotezo H_0 :

H_0 : Spremenljivki S (mediana starosti populacije) in F (fatalnost virusa) sta si neodvisni

Računamo R^2 (Coefficient of determination) in F:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} = 0.049$$

kjer so y_i podatki, ki jih imamo na razpolago, \hat{y}_i pa vrednosti, ki bi jih imela odvisna spremenljivka F (fatalnost virusa) v primeru, da bi njene vrednosti ležale natanko na regresijski premici.

$$F = \frac{R^2}{(p - 1)(1 - R^2)(n - p)} = 0.002055029$$

kjer p je število spremenljivk, R^2 je Koeficient odločnosti in n je število držav evropske unije. Velja podobno kot v analizi 4.1. Ničelno hipotezo H_0 lahko zavrnilo, če $F \geq F_t$. F_t je vrednost, ki preberemo iz porazdelitvene tabele distribucije F. Izberemo α (significance level) = 0.05. Iz tabele lahko razberemo, da je $F_t = 4.2417$ (Gledamo stolpec $p - 1 = 1$ in vrstico $n - 2 = 25$). Velja da, $F < F_t$, torej ne moremo zavreči ničelne hipoteze H_0 . Ker je vrednost F veliko manjša od vrednosti F_t , si lahko predstavljamo, da sta si spremenljivki S in F neodvisni.

4.5 Vpliv velikosti populacije na število opravljenih testov

5 Zaključki

6 Literatura

https://en.wikipedia.org/wiki/List_of_countries_by_median_age
<https://en.wikipedia.org/wiki/Europe>

https://en.wikipedia.org/wiki/List_of_European_countries_by_population

<https://covid19.healthdata.org/>

<https://covid19.who.int/info>

[https://github.com/KalcMatej99/Seminarska – VS – Covid – 19](https://github.com/KalcMatej99/Seminarska-VS-Covid-19)

Martina Lizza – Pandemijski koronavirus : biološki, mikrobiološki i kemijski izsledki te okužbe [https://www.cebm.net/covid – 19/covid – 19 – what – proportion – are – asymptomatic/](https://www.cebm.net/covid-19/covid-19-what-proportion-are-asymptomatic/)