

SVEUČILIŠTE U RIJECI
ODJEL ZA INFORMATIKU

Diplomski studij informatike
Informacijsko – komunikacijski sustavi

Priprema i vizualizacija podataka

Seminarski rad iz kolegija Sustavi za podršku odlučivanju

Nositelj kolegija: izv. prof. dr. sc. Sanda Martinčić - Ipšić

Autorica : Mia Puriš

Rijeka, lipanj 2020

1. Izvor podataka

Ideja seminara je prikazati koliko iznosi ukupna stopa fertiliteta razvijenih i nerazvijenih zemalja te kako drugi čimbenici utječu na to.

Podaci koji su bili potrebni za izradu seminara preuzeti su s nekoliko stranica: [stranica A](#). Na stranici nalazi se puno informacija, skupova podataka te tema (demografija, politika, bolesti, stanje u svijetu...) koje mogu zadovoljiti potrebe korisnika. S stranice uzeti su [podaci](#) vezani za ukupnu stopu plodnosti odnosno za prosječan broj djece koja bi se rodila po ženi koja bi živjela do kraja svog plodnog razdoblja od 15. do 49. godine života. [Drugi skup podataka](#) vezan je za kontracepciju odnosno prikazuje postotak žena reproduktivne dobi (od 15-49) koje su u braku ili zajednici i koje koriste ili čiji seksualni partneri koriste neku od metoda kontracepcije. Te informacije pokazatelj su zdravstvenih usluga, razvoja, osnaživanje žena i korisne su za razumijevanje prošlih, sadašnjih i budućih trendova razvoja plodnosti odnosno koliko korištenje kontracepcije utječe na to. [Treći skup podataka](#) prikazuje prosječnu dob majki pri rođenju prvog djeteta. Te informacije koristan su pokazatelj programa planiranja obitelji koje imaju za cilj smanjiti smrtnost majki i povećati upotrebu kontracepcijskih sredstava (posebno među adolescentnim parovima u braku i neoženjenima) odnosno kako bi se odgodilo prerano stupanje u brak i na taj način poboljšalo buduće zdravlje novorođenčadi. [Četvrti skup podataka](#) prikazuje troškove obrazovanja kao postotak BDP¹-a. Obrazovanje svakako ima utjecaj na razinu plodnosti na način da ako se mladi ljudi zaduže i dignu kredit za školovanje koji moraju vratiti nakon završetka fakulteta odmah traže posao. Ali tu je prisutan još i trend dužeg obrazovanja i težnja ka situiranosti prije stvaranja obitelji što dovodi do pomicanja granica rađanja te samim time utječe na stopu fertiliteta.

[Stranica B](#) također sadrži puno informacija, raznovrstan sadržaj i skupove podataka koji su vezani za svijet. [Prvi skup podataka](#) prikazuje razvijene i nerazvijene zemlje koje se određuju prema kriteriju HDI (eng. *Human Development Index*)². Temelji se na razvoju zemlje kao što je obrazovanje, zdravlje, životni vijek i dohodak. HDI je u rasponu od 0-1 te zemlje koje imaju HDI veći od 0.8 smatraju se razvijenima, u suprotnome nerazvijenima. [Drugi skup podataka](#) sadrži podatke o stopi pismenosti po zemljama. Iako je globalni postotak pismenosti prilično visok ipak postoje još neke zemlje u južnoj i zapadnoj Aziji i u subsaharskoj Africi gdje je postotak pismenosti još uvijek mali. Kako se povećava stopa pismenosti fertilitet se smanjuje iz razloga da žene koje znaju čitati i pisati postaju svjesnije o planiranju obitelji vjerojatnije je

¹ Bruto domaći proizvod

² Ljudski razvojni indeks

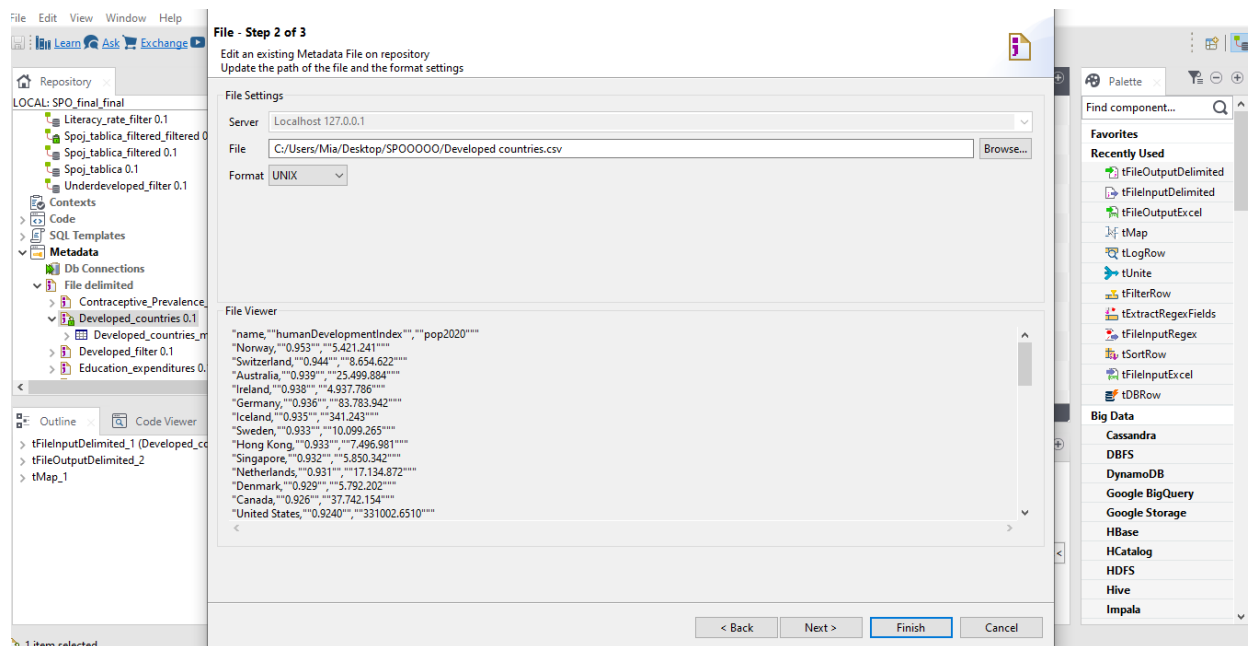
da će koristiti neku od metoda kontracepcije. Udavaju se kasnije, imaju manje djece te su bolje informirane u vezi zdravlja, higijene, žive u boljim uvjetima i većina njihove djece preživi.

[Stranica C](#) sadrži podatke koji su vezani za svjetsko stanovništvo. [Skup podataka](#) sadrži podatke koliko će djece majka roditi prema svojoj dobi (u tisućama) u plodnom razdoblju od 15. do 49. godine svog života.

2. Čišćenje podataka u alatu Talend Data Integration

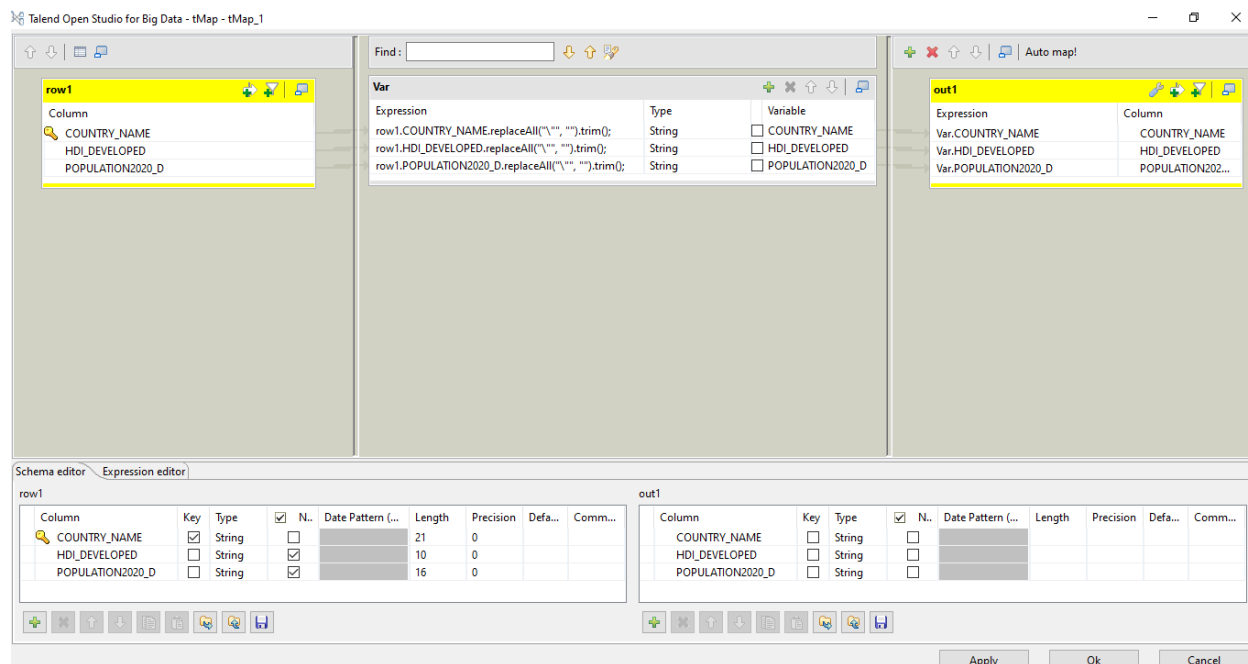
Nakon izrade projekta u Talend-u potrebo je izraditi Job unutar kojega se vrši sva potrebna obrada podataka.

Prvo se u program unose početne tablice klikom na opciju File delimited unutar koje se uređuje postojeća tablica.



Slika 1 Prikaz uređenja tablice Developed Countries

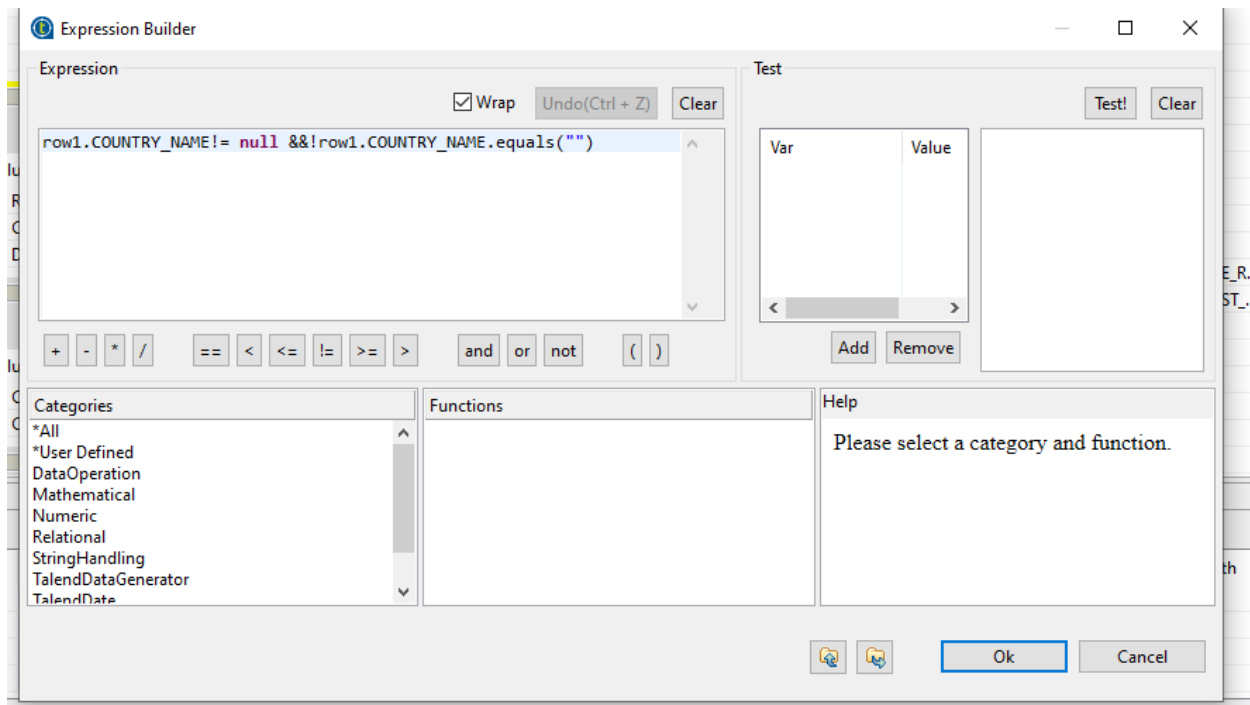
Prva na redu je tablica Developed Countries. Pošto sadrži nepotrebne znakove (u ovom slučaju navodnike tj. meta podatke u imenu stupaca i podataka u redcima) potrebno ih je je odstraniti jer onemogućavaju spajanje tablice. Korištena je multifunkcionalna komponenta tMap unutar koje se pomoću funkcije **replaceAll('"', '')** otklanjaju navedeni znakovi te se s funkcijom trim() otklanjaju nepotrebne praznine ukoliko postoje. Nakon tMap-a tablice se generiraju u jednu tablicu upotrebom funkcije *tFileOutputDelimited*.



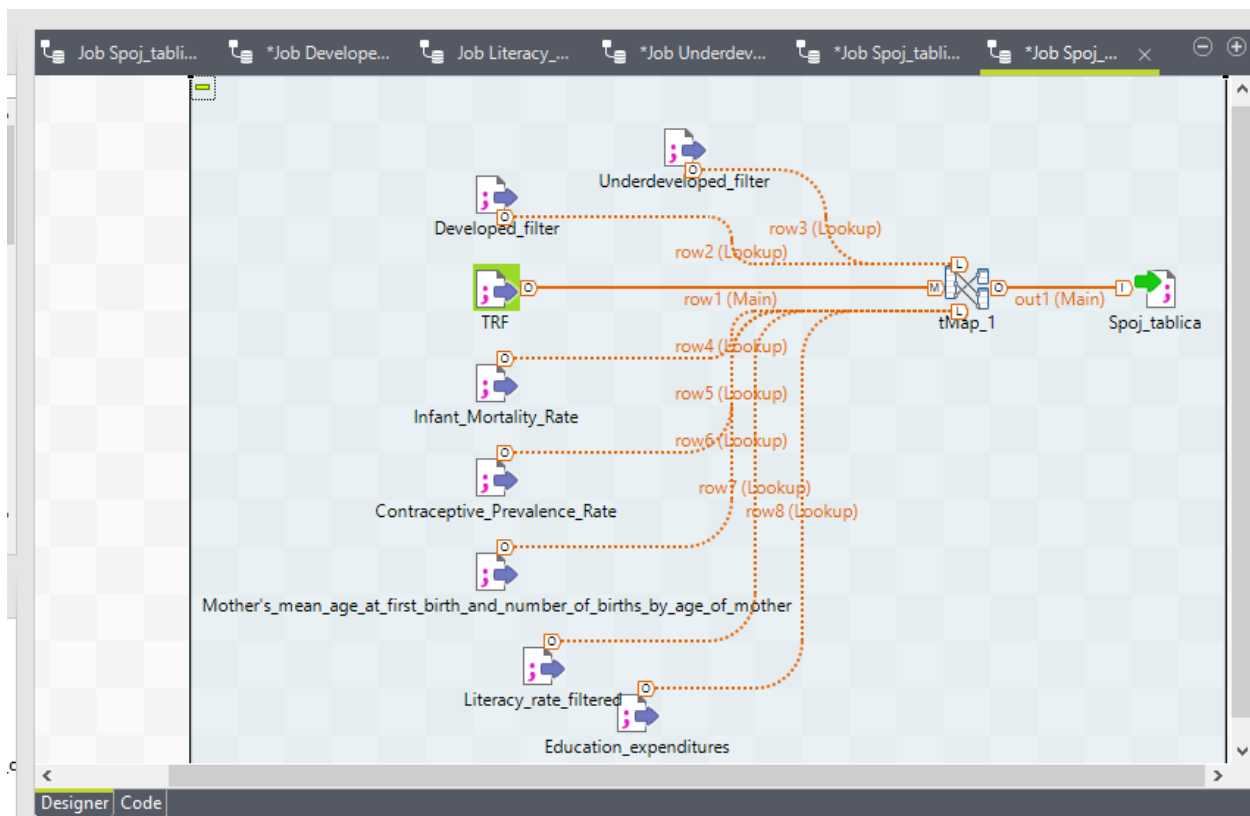
Slika 2 Prikaz funkcije `replaceAll`

Isti postupak ponavlja se i za tablice `Underdeveloped_countries` i `Literacy_rate_2020` koje isto sadrže meta podatke koje treba maknuti.

Slijedeće na redu je spajanje preostalih tablica koje će se do kraja očistiti kroz 2 Job-a radi očuvanja preglednosti. Tablice su međusobno povezane primarnim ključem `COUNTRY_NAME`. Za spajanje opet će se upotrijebiti komponenta `tMap` kojom će se potrebni podaci spojiti u jednu tablicu, odnosno atributi izlazne tablice koje se nalaze na desnoj strani koji su odabrani iz svih ulaznih tablica s lijeve strane. Zajednička datoteka nosi naziv `Spoj_tablica`. Treba naglasiti da je odmah postavljen uvjet u „Expresion“ filteru unutar kojega se filtriraju ulazni zapisi koji nisu jednaki vrijednosti `null` i koji su prazni za atribut `COUNTRY_NAME`.

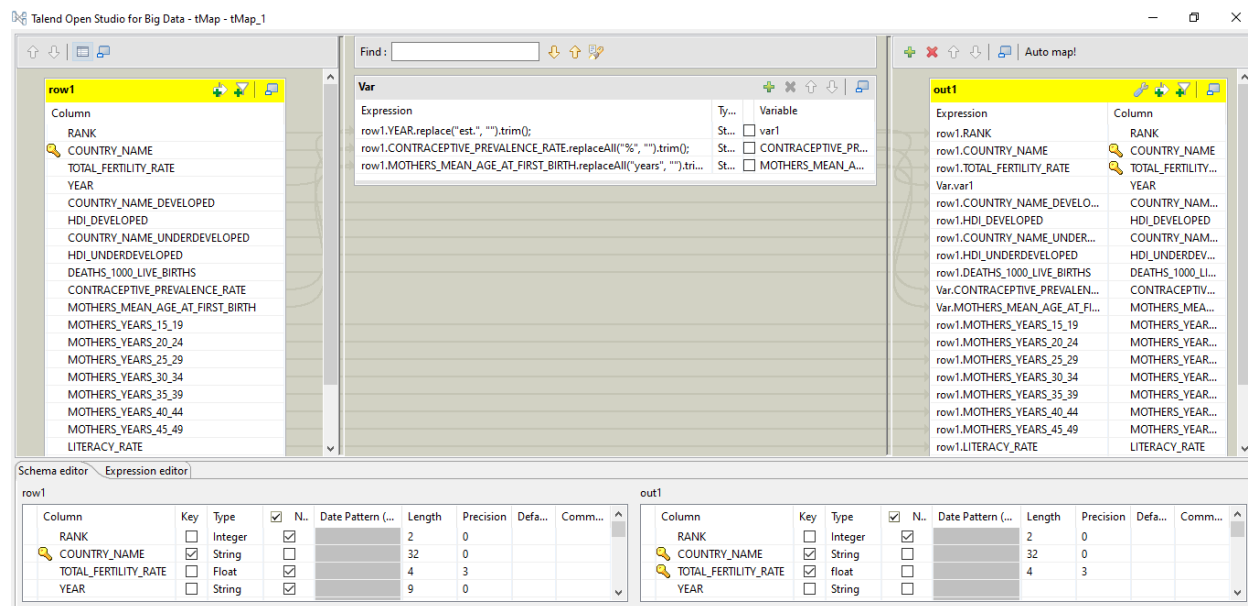


Slika 3 Prikaz „Expression filtera“



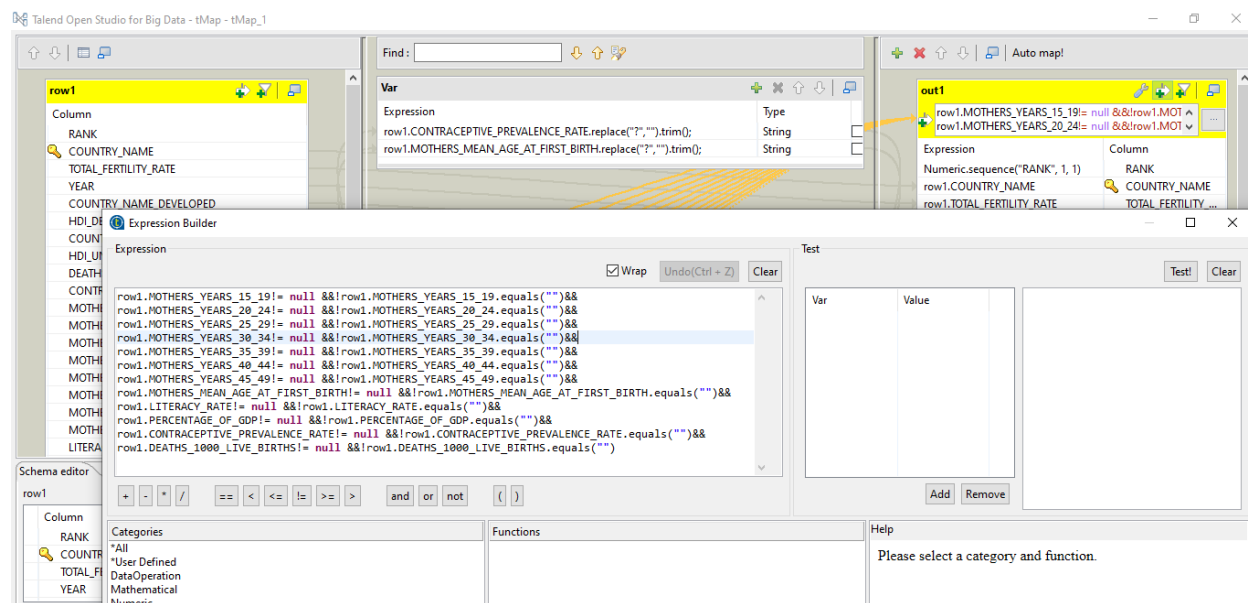
Slika 4 Prikaz početnih datoteka spojenih pomoću komponente tMap

Nakon spajanja tablica na redu je drugo filtriranje. Opet se filtriraju nepotrebni znakovi.



Slika 5 Prikaz filtriranja

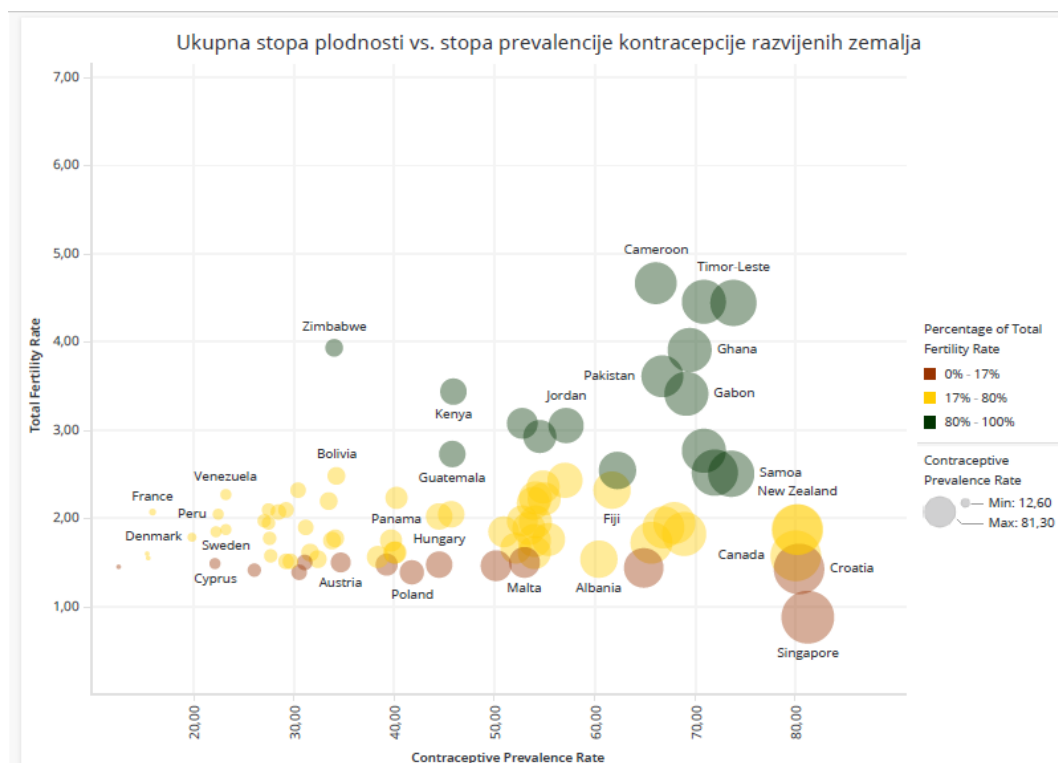
I u zadnjem filtriranju opet se miču nepotrebni znakovi te se u „Expression filteru“ dodaje izraz prema kojemu se filtriraju svi podaci osim Developed i Underdeveloped Countries te atributi koji su vezani za njih iz razloga što te dvije tablice ne sadrže iste vrijednosti te se na taj način omogućuje njihovo spajanje. I još na samom kraju za izlaz se postavi .xls datoteka radi boljeg učitavanja podataka u preostalim alatima.



Slika 6 Prikaz filtriranja u tMap-u

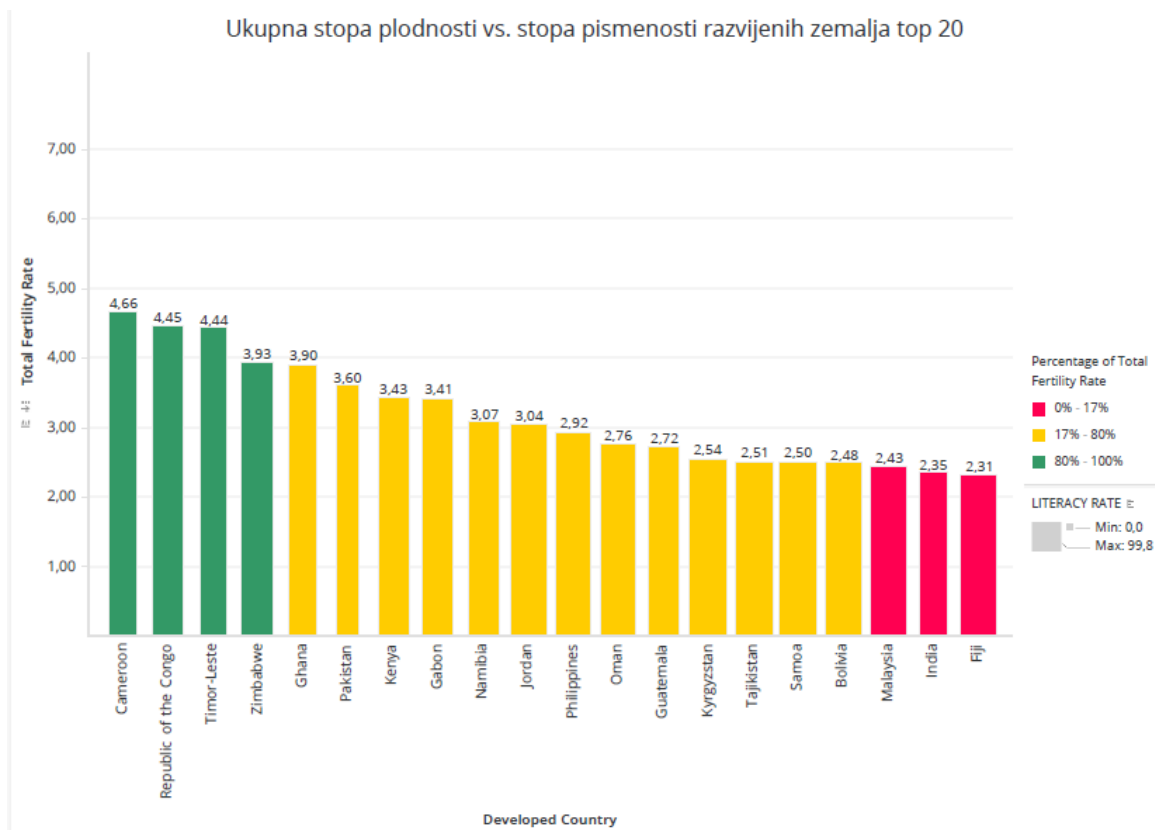
3. Vizualizacije

3.1 Microstrategy



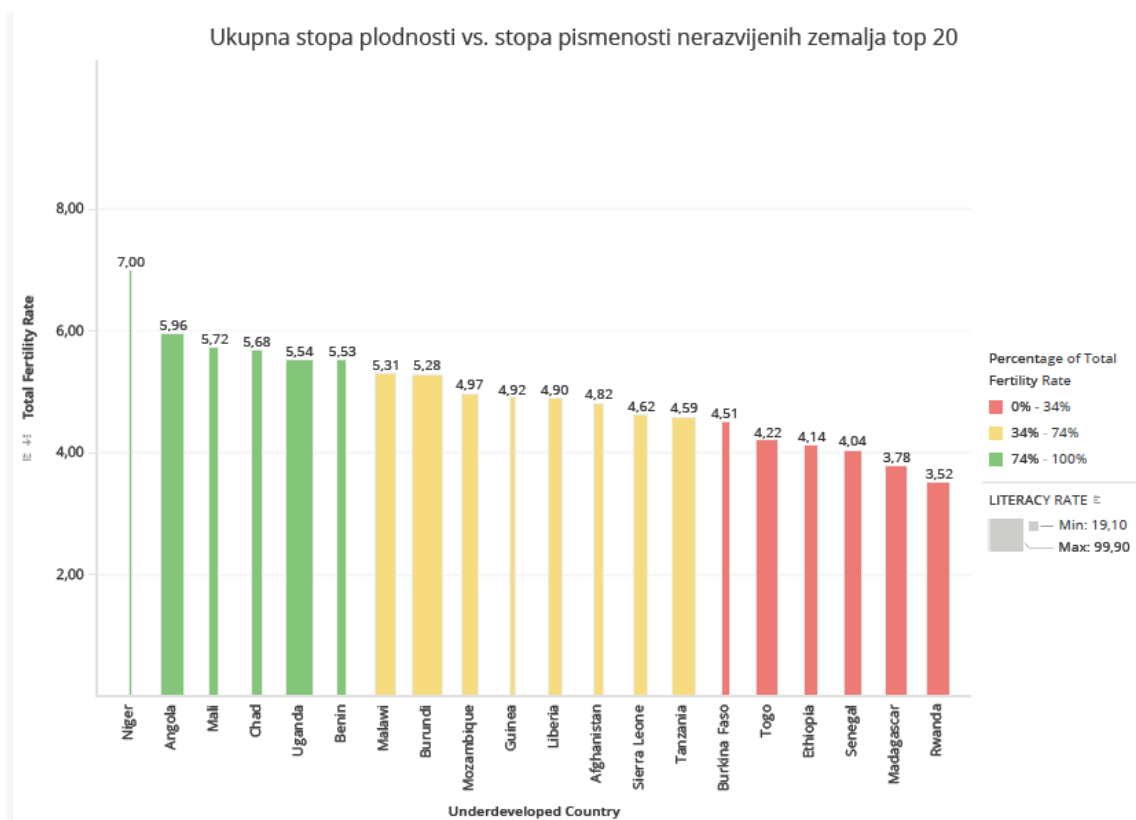
Slika 7 Prikaz ukupne stope plodnosti i stope prevalencije kontracepcije

Slika 7 prikazuje stopu prevalencije kontracepcije kako utječe na ukupnu stopu plodnosti kod razvijenih zemalja. Države su obojane prema stopi plodnosti (crveno - mala stopa plodnosti, žuto - srednja stopa plodnosti, zeleno – visoka stopa plodnosti), veličina kruga ovisi o postotku kontracepcije svake države.



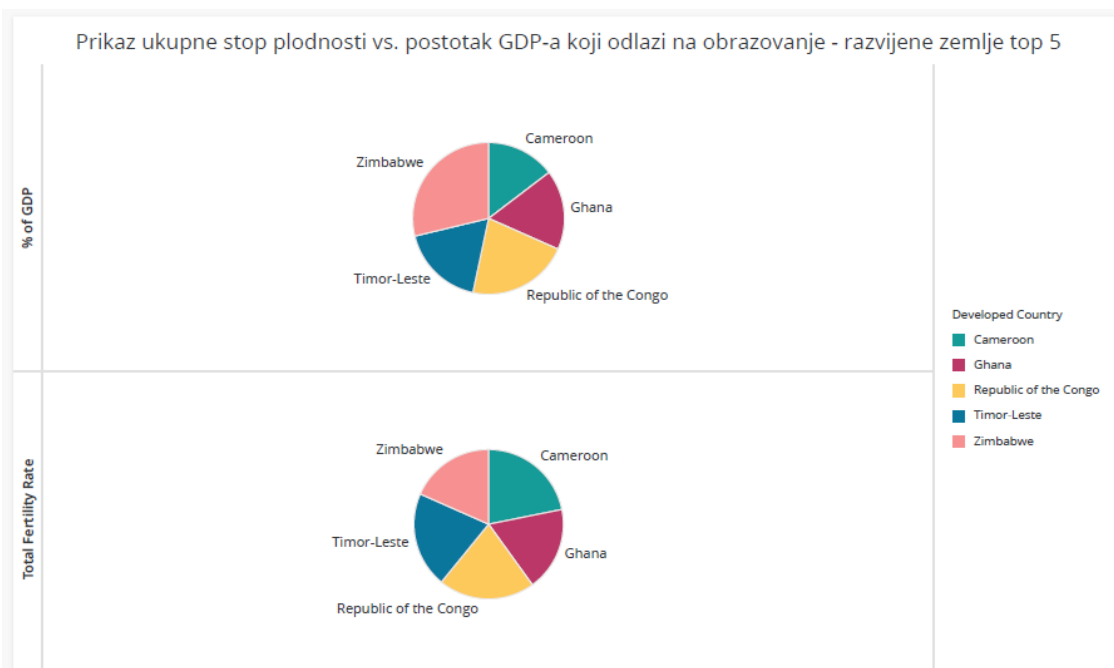
Slika 8 Prikaz ukupne stope plodnosti i stope pismenosti razvijenih zemalja

Slika 8 prikazuje kako na ukupnu stopu plodnosti utječe stopa pismenosti razvijenih zemalja. Graf je obojan prema stopi plodnosti (zeleno – visoka stopa plodnosti, žuto – srednja stopa plodnosti, crveno – niska stopa plodnosti), a veličina stupaca određena je stopom pismenosti. Kao što se odmah da zaključiti kod razvijenih zemalja stopa pismenosti je izrazito velika pa razlika između država nije pretežito naglašena. Skup podataka filtriran je za prvih 20 zemalja te je odabran stupčasti graf zbog preglednosti.



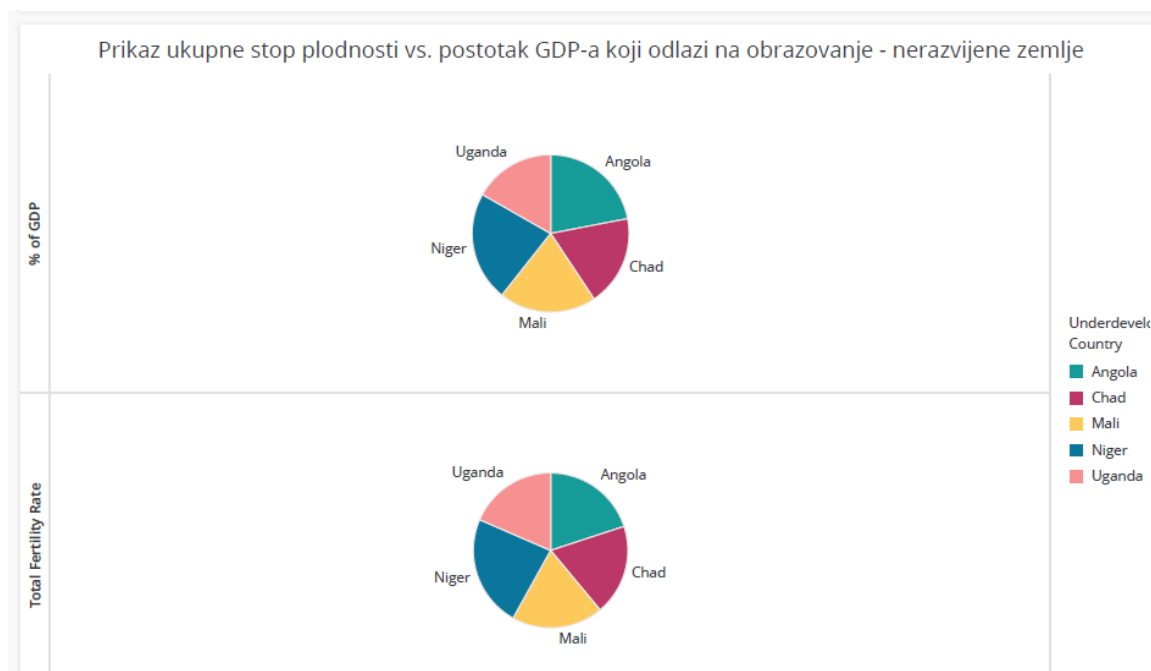
Slika 9 Prikaz ukupne stope plodnosti i stope pismenosti nerazvijenih zemalja

Slika 9 prikazuje kako na ukupnu stopu plodnosti utječe stopa pismenosti nerazvijenih zemalja. Graf je obojan prema stopi plodnosti (zeleno – visoka stopa plodnosti, žuto – srednja stopa plodnosti, crveno – niska stopa plodnosti), a veličina stupaca određena je stopom pismenosti. Kod nerazvijenih zemalja razlika je očita u odnosu na razvijene zemlje. Niger je prvi prema stopi plodnosti ali i također je zadnji prema stopi pismenosti. To je jedan od razloga zašto je stopa plodnosti tako visoka u Nigeru. Skup podataka filtriran je za prvih 20 zemalja.



Slika 10 Prikaz ukupne stope plodnosti i troškova obrazovanja

Slika 10 prikazuje razmjer između ukupne stope plodnosti i troškova u obrazovanju. Podaci su filtrirani na 5 država koje imaju veću stopu plodnosti. Kod nekih država na stopu plodnosti utječu troškovi obrazovanja.

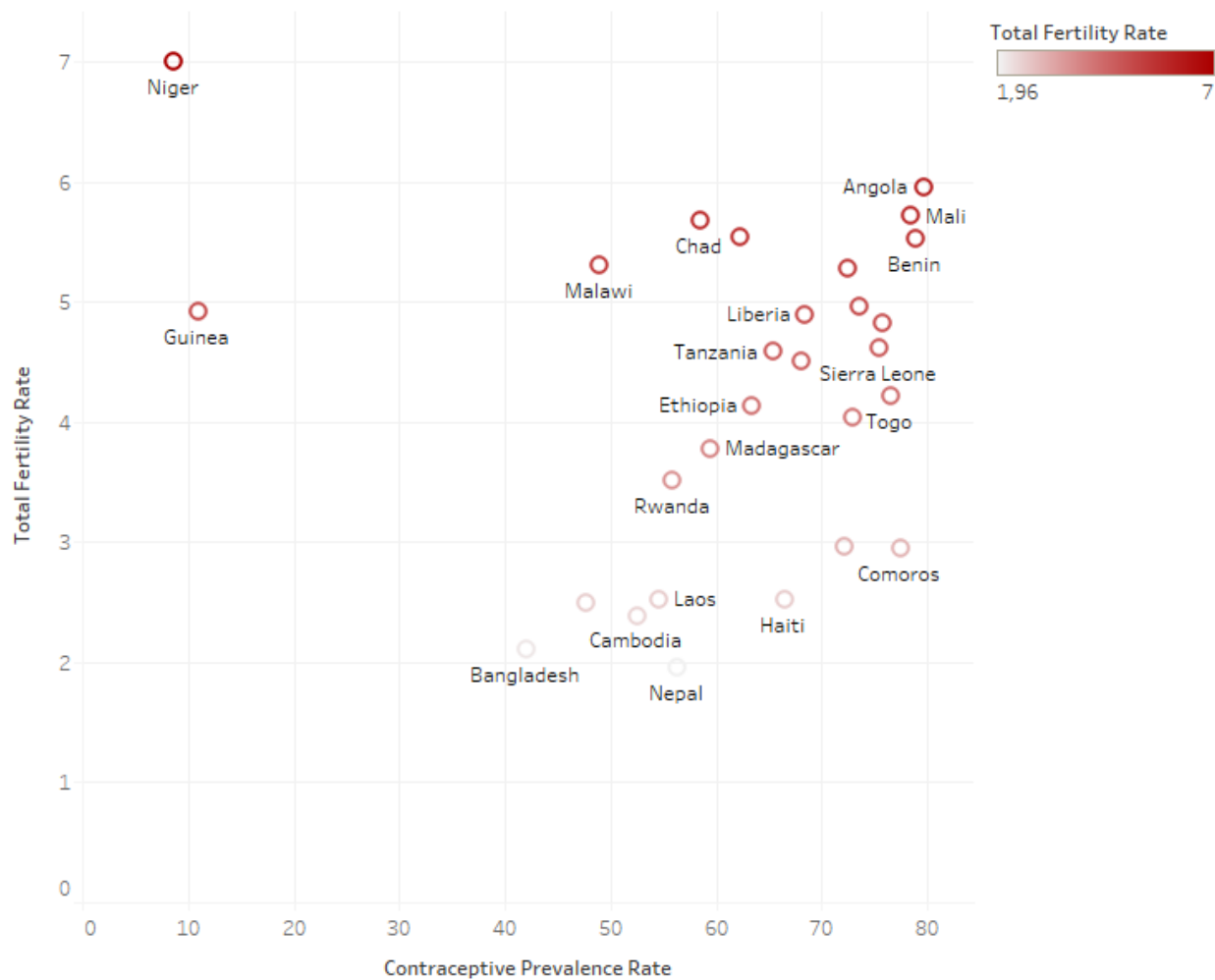


Slika 11 Prikaz ukupne stope plodnosti i troškova obrazovanja

Slika 11 prikazuje razmjer između ukupne stope plodnosti i troškova u obrazovanju. Podaci su filtrirani na 5 država koje imaju veću stopu plodnosti. Kod nekih država na stopu plodnosti utječu troškovi obrazovanja. Kod nerazvijenih zemalja kako se i može pretpostaviti što je trošak manji to je stopa plodnosti veća.

3.2 Tableau

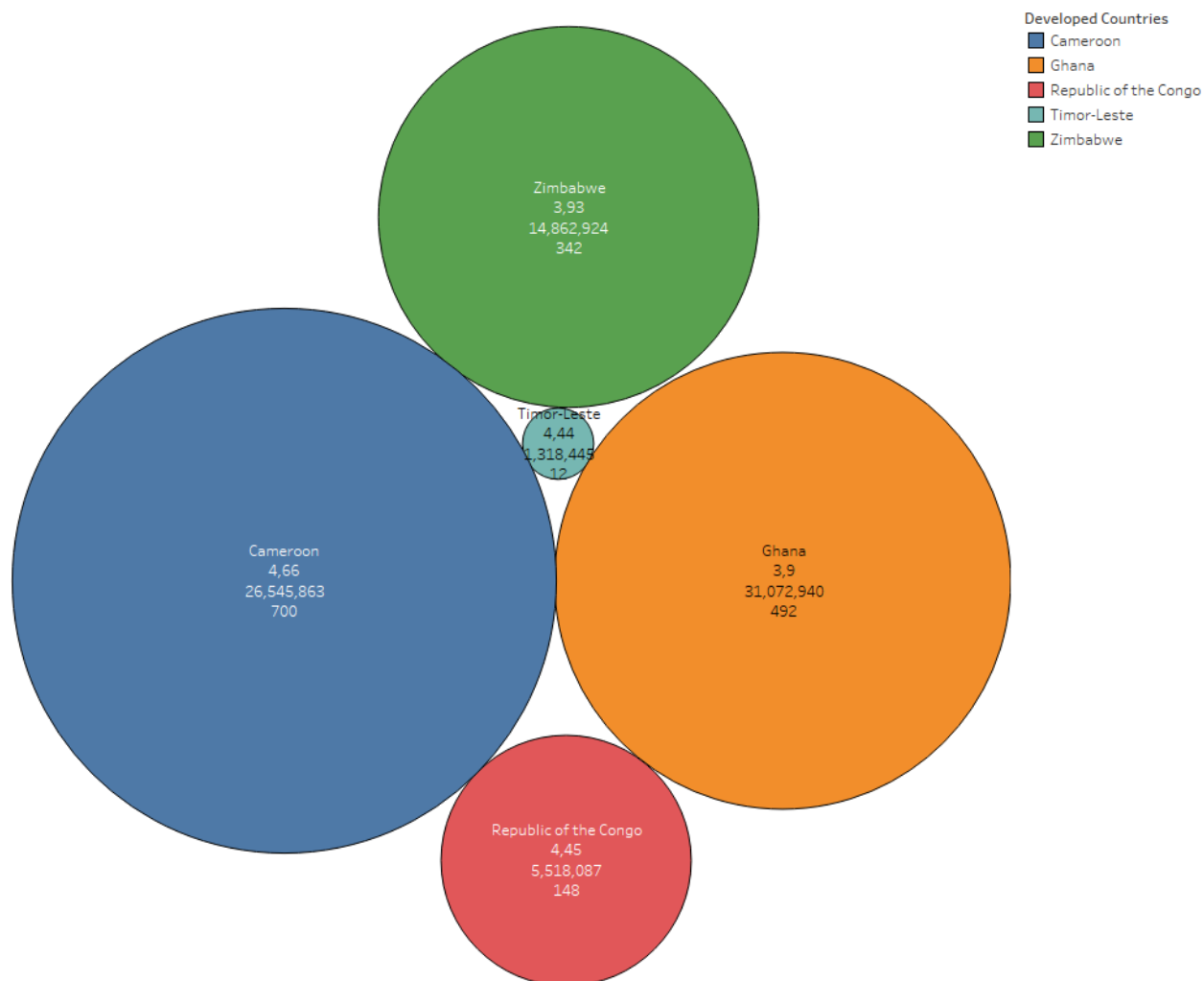
Ukupna stopa plodnosti vs. stopa prevalencije kontracepcije nerazvijenih zemalja



Slika 12 Prikaz ukupne stope plodnosti i stope prevalencije kontracepcije

Slika 12 prikazuje stopu prevalencije kontracepcije kako utječe na ukupnu stopu plodnosti kod nerazvijenih zemalja. Države koje imaju slabiju stopu plodnosti slabije su obojane crvenom bojom, a one koje imaju višu sve su više crvenije.

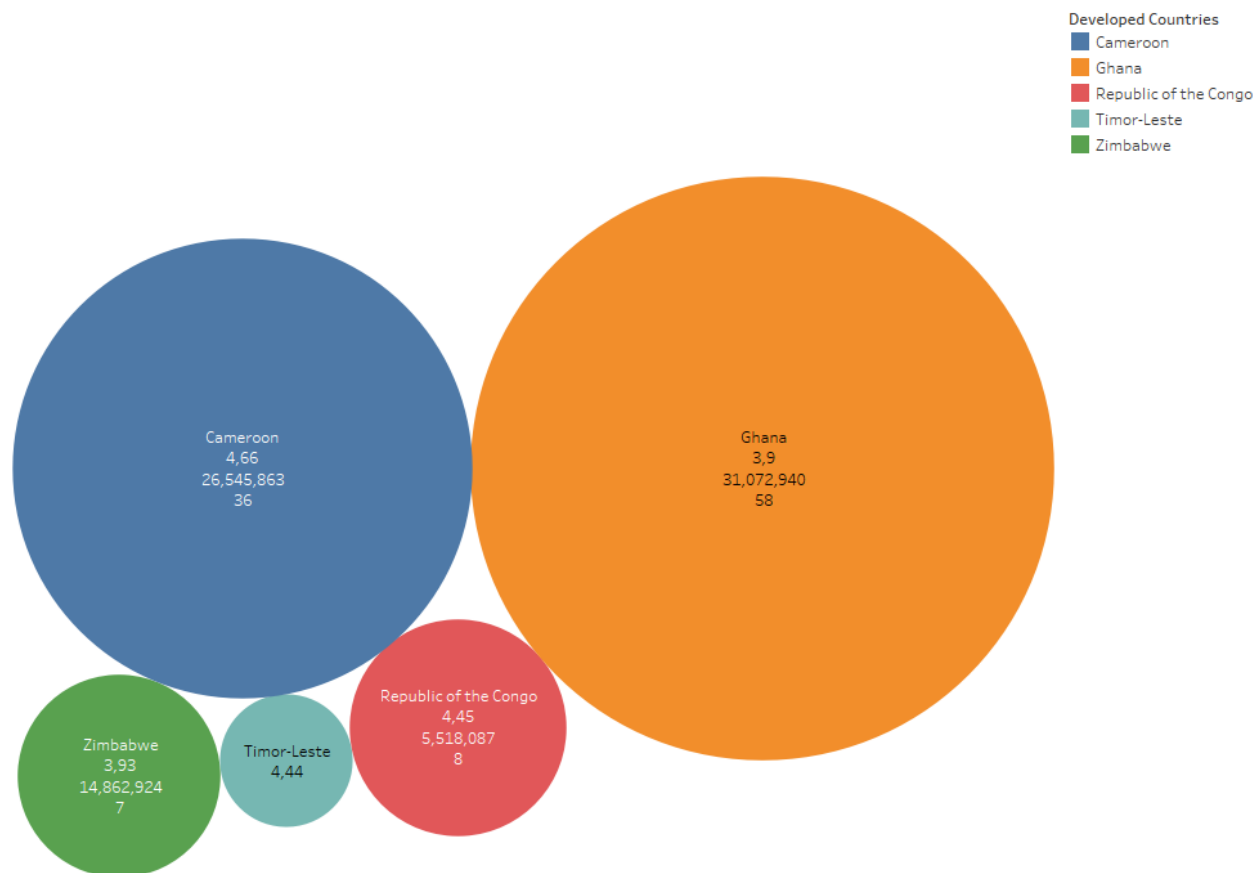
Pregled ukupne stope plodnosti i broj djece koje žena rodi u periodu od 15-19. godine
- razvijene zemlje



Slika 13 Prikaz ukupne stope plodnosti i broja djece koje žena rodi u periodu od 15.-19. godine života

Slika 13 prikazuje ukupnu stopu plodnosti i broj djece koje će žena roditi na početku svog plodnog razdoblja u razvijenim zemljama. Podaci su filtrirani na 5 država radi preglednosti. Veličina krugova ovisi o broju djece koje je žena rodila. Cameroon je zemlja s najvećom stopom plodnosti i s najvećim brojem rođene djece. Ovaj graf najbolje prikazuje odnos između veličina podataka.

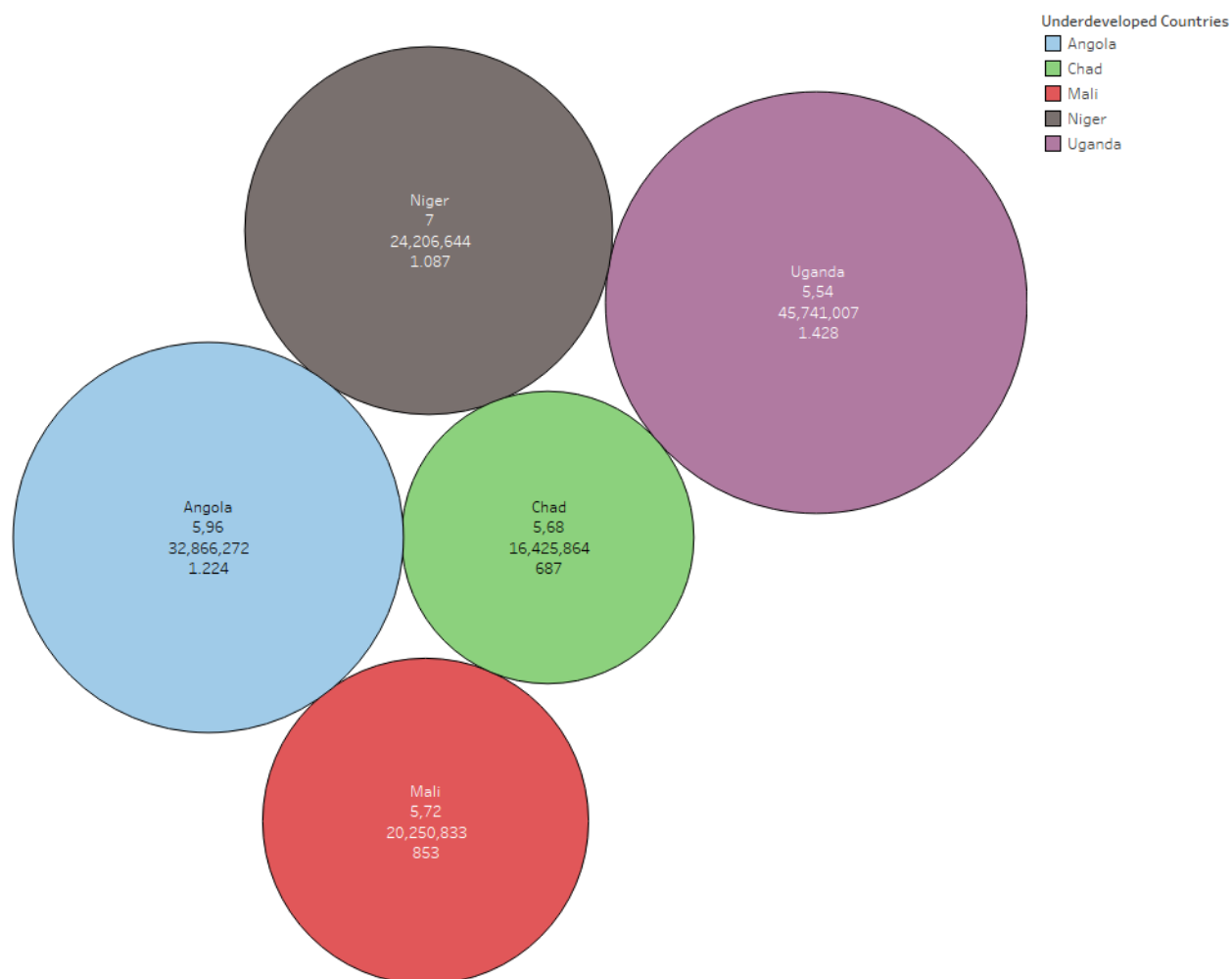
Pregled ukupne stope plodnosti i broj djece koje žena rodi u periodu od 45-49. godine
- razvijene zemlje



Slika 14 Prikaz ukupne stope plodnosti i broja djece koje žena rodi u periodu od 45.-49. godine života

Slika 14 prikazuje ukupnu stopu plodnosti i broj djece koje će žena roditi na kraju svog plodnog razdoblja u razvijenim zemljama. Podaci su filtrirani na 5 država radi preglednosti. Veličina krugova ovisi o broju djece koje je žena rodila.

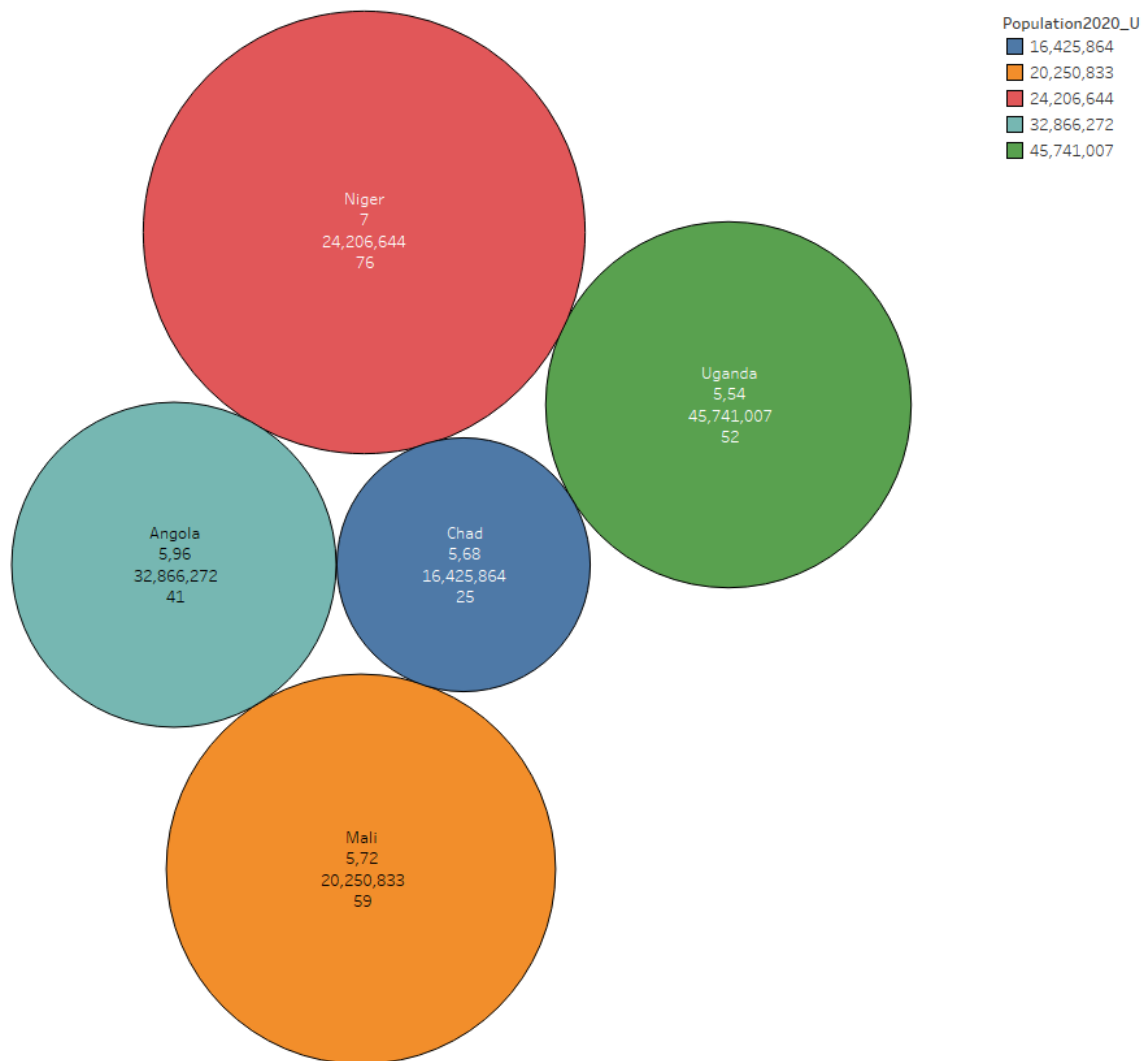
Pregled ukupne stope plodnosti i broj djece koje žena rodi u periodu od 15-19. godine
- nerazvijene zemlje



Slika 15 Prikaz ukupne stope plodnosti i broja djece koje žena rodi u periodu od 15.-19. godine života

Slika 15 prikazuje ukupnu stopu plodnosti i broj djece koje će žena roditi na početku svog plodnog razdoblja u nerazvijenim zemljama. Podaci su filtrirani na 5 država radi preglednosti. Veličina krugova ovisi o broju djece koje je žena rodila.

Pregled ukupne stope plodnosti i broj djece koje žena rodi u periodu od 45-49. godine
- nerazvijene zemlje



Slika 16 Prikaz ukupne stope plodnosti i broja djece koje žena rodi u periodu od 45.-49. godine života

Slika 16 prikazuje ukupnu stopu plodnosti i broj djece koje će žena roditi na kraju svog plodnog razdoblja u nerazvijenim zemljama. Podaci su filtrirani na 5 država radi preglednosti. Veličina krugova ovisi o broju djece koje je žena rodila.

4. Zaključak

Iz podataka vidljivo je da na ukupnu stopu plodnosti utječu različiti čimbenici i da rezultati nisu isti kod razvijenih i nerazvijenih zemalja. Pošto živimo u 21. stoljeću neki podaci su šokantni za nerazvijene zemlje s obzirom na mogućnosti koje se danas nude. Nadajmo se da će se stopa plodnosti s godinama sve više smanjivati radi dobrobiti Zemlje ali i samih ljudi jer djeca trebaju dolaziti planski , a ne slučajno.