

# Obrada skupa podataka: UFO Sightings

Projektni rad iz predmeta: Osnove statističkog programiranja, 2023./2024.

Autori: Luka Blagaić, Valentina Ivanić, Dominik Ožvald, Nikola Vlahović

## 1 Uvodna obrada i prilagodba podataka

Podaci se nalaze u datoteci nazvanoj `complete.csv`, pri čemu su podaci odvojeni zarezima. Za učitavanje ovih podataka koristimo funkciju `read_csv()`. Nakon inicijalnog učitavanja i pregleda dokumentacije povezane s ovim skupom podataka, primijetili smo da se u stupcu "comments" nalaze nizovi znakova poput "&#44", koji označavaju točke, zareze i druge simbole koji mogu izazvati probleme pri spremanju u formatu csv datoteke. Zamijenili smo ove nizove znakova razmakom koristeći regularne izraze.

Mnogi reci podataka nisu imali navedenu državu, stoga smo poduzeli sljedeće korake kako bismo odredili pripadnost državi za svaki podatak:

- Kreirali smo popis država u SAD-u i svaki redak koji ima vrijednost u stupcu "state" koja se nalazi unutar ovog popisa smo dodijelili SAD-u.

- Svaki redak koji u atributu "city" ima navedenu državu u zagradi smo pridružili odgovarajućoj državi (na primjer, "chester (uk/england)" smo dodijelili Velikoj Britaniji).

- Za sve preostale podatke na temelju atributa "city", odredili smo pripadajući grad i pridružili ga državi u kojoj se taj grad nalazi.

- Podatke koji nisu imali niti jedan od atributa "city", "state" ili "country" smo označili kao pripadajuće državi "Nepoznato".

Početno, atribut "country" je sadržavao kratice država ("us", "ca", "gb"...). Ove kratica smo zamijenili punim imenima država na hrvatskom jeziku. Sve navedene promjene smo pohranili u posebnu csv datoteku koju smo koristili za daljnju analizu podataka.

Daljnja obrada podataka, uključujući izlučivanje specifičnih podataka i izradu grafova, zahtijevala je dodatne korake, koji će biti opisani uz svaki pojedini graf.

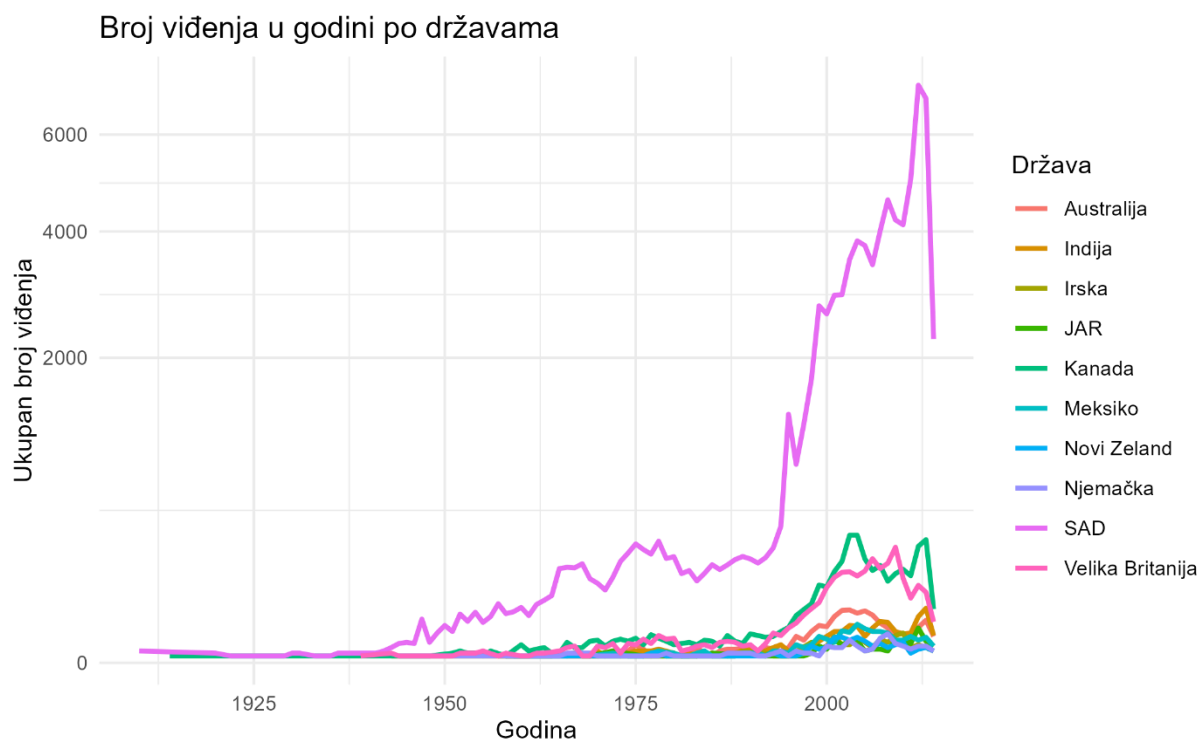
## 2 Analiza podataka

### 2.1 UFO – viđanja kroz godine

Na početku naše analize, usmjerenje je bilo na proučavanje podataka o UFO događajima kroz protekle godine. Cilj je bio istražiti dinamiku njihovog rasta, padanja te uočiti eventualne veze s važnim događajima poput slijetanja na Mjesec ili razvoja dronova.

#### 2.1.1 Broj UFO – viđanja u godini po državama

Prva pretpostavka koje smo željeli provjeriti bila je da će broj prijavljenih UFO viđenja naglo rasti nakon što su dronovi postali lako dostupni. Dronovi se u lošim uvjetima lako mogu zamijeniti za nepoznatu letjelicu. Da bi smo to provjerili stvorili smo graf(slika 1) koji prikazuje broj UFO viđenja svake godine za nekoliko različitih država. Na tom grafu očekujemo nagli rast oko 2000-te godine.

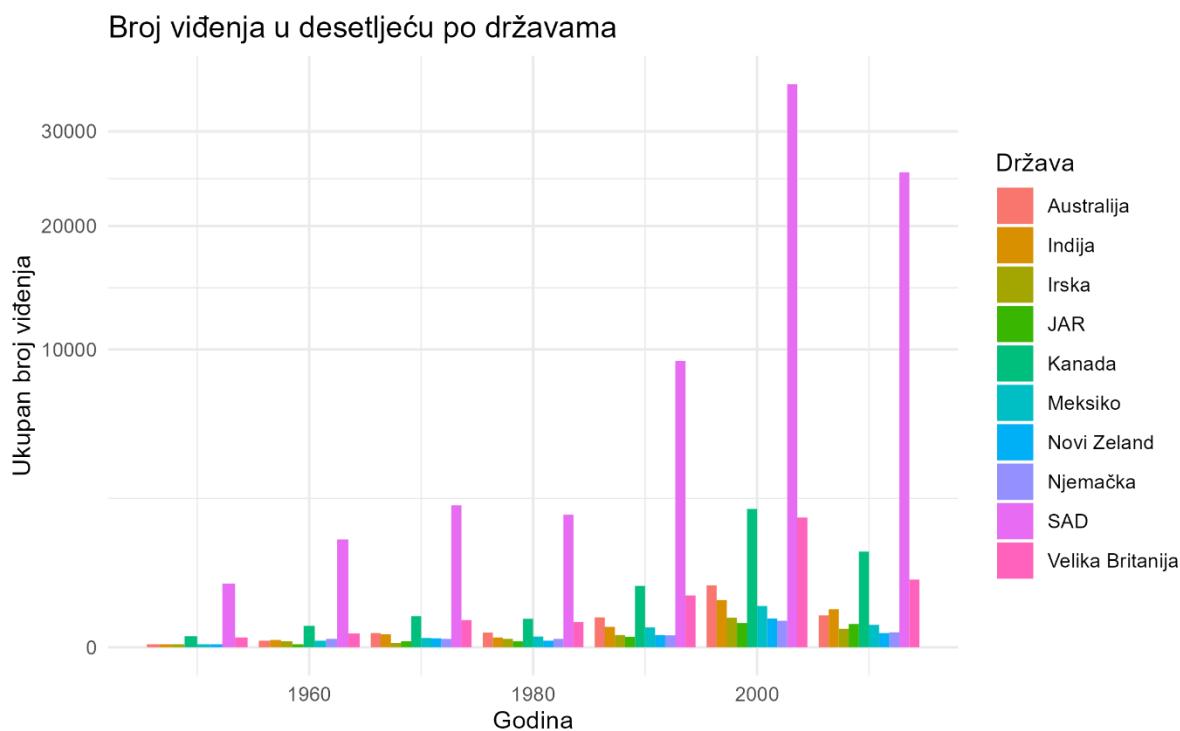


Slika 1 Graf: Broj viđenja u godini po državama

Prikazano na graf(slika 1) možemo vidjeti da nije došlo do naglog rasta oko 2000-te godine nego je taj rast krenuo već u devedesetim godinama što nam je opovrgnulo pretpostavku.

### 2.1.2 Broj UFO – viđanja u desetljećima po državama

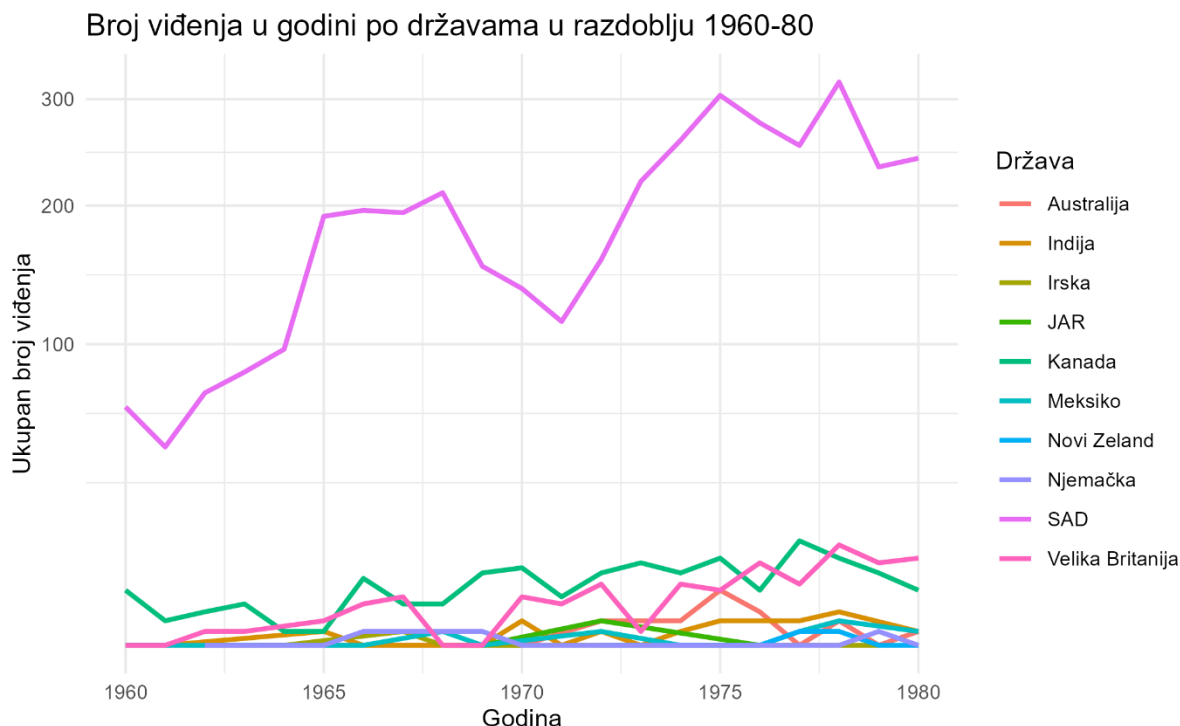
Nadalje smo željeli provjeriti kolika je razlika u broju viđenja u prošlih nekoliko desetljeća. Možemo vidjeti na sljedećem grafu(slika 2) kako od šezdesetih do osamdesetih godina broj viđanja ostaje malo promijenjen dok se u devedesetima naglo udvostručio, te nakon toga u dvije tisućitima utrostručio.



Slika 2 Graf: Broj viđenja u desetljeću

### 2.1.3 Broj UFO – viđanja oko 1969. godine (slijetanje na Mjesec)

Naša sljedeća pretpostavka bila je da će se broj viđanja povećati oko 1969. godine kada je čovjek prvi put sletio na Mjesec. Ova pretpostavka temelji se na činjenici da je u to vrijeme, posebno u SAD-u, postojalo izraženo zanimanje za svemirsku tehnologiju, a slijetanje na Mjesec je bilo veliko postignuće koje je potaknulo interes za vanzemaljskim životom i svemirom u cjelini. To bi moglo povećati vjerojatnost da se neki obični događaj interpretira kao nešto nepoznato iz svemira. Na grafu (slika 3) smo prikazali razdoblje od 1960. do 1980. godine kako bi provjerili ovu pretpostavku.



Slika 3 Graf: Broj viđenja u razdoblju od 1960. do 1980.

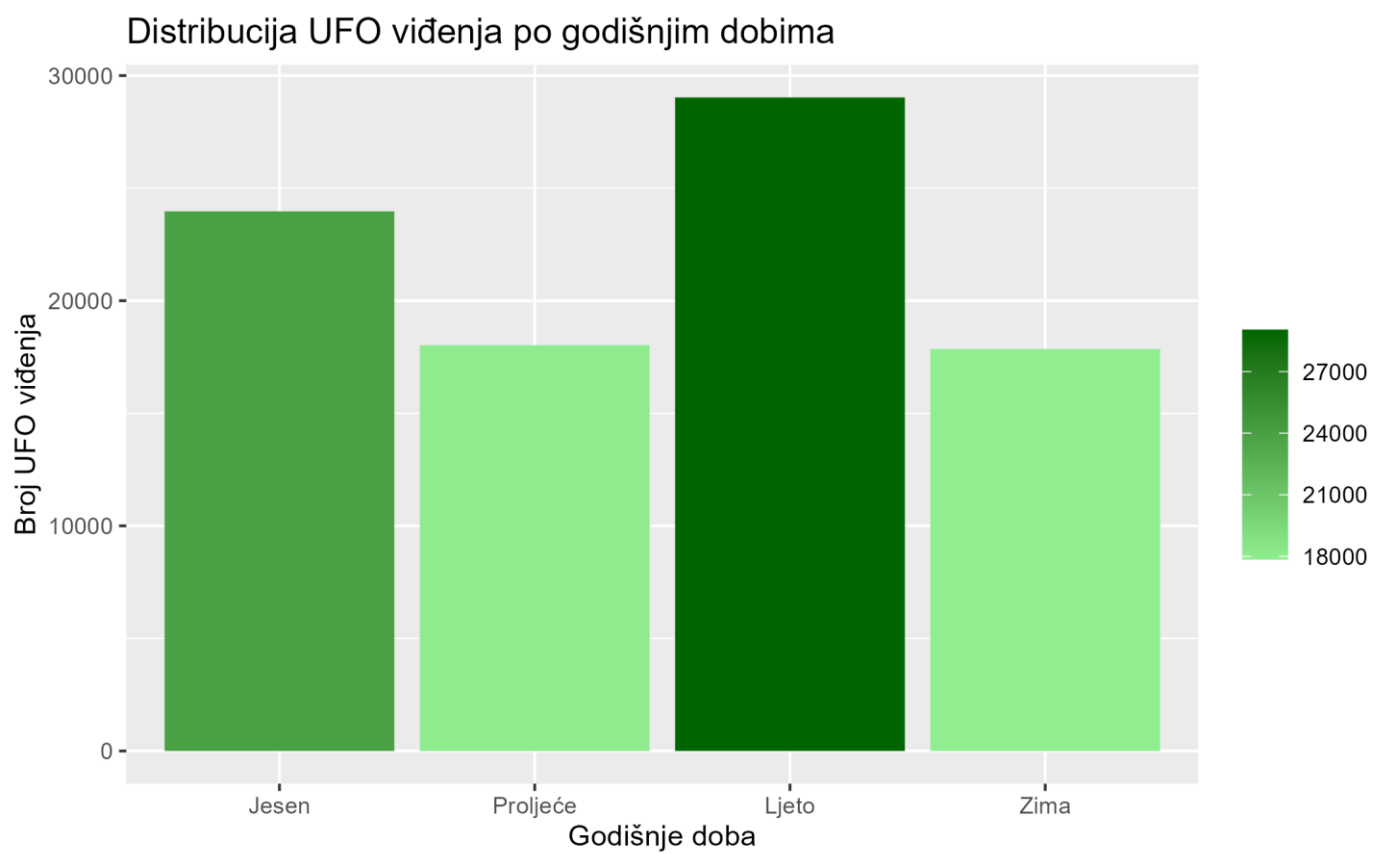
Pomalo začuđujuće, ali broj viđenja UFO-a 1969. godine u SAD-u pada, dok u ostalim državama dolazi do manjih promjena ili stabilnosti. Lokalni minimum u SAD-u je oko 1972. godine, ali možemo vidjeti da je broj slučajeva počeo padati već prije 1969. godine što je suprotno našoj pretpostavci. Suprotno našim očekivanjima, moguće je da slijetanje na Mjesec i povećano zanimanje za svemirsku tehnologiju nisu imali očekivani utjecaj na broj prijave viđenja UFO-a.

## 2.2 Analiza UFO – viđenja prema godišnjim dobima, distribucije tijekom tjedna i dana

U drugoj fazi naše analize, usredotočili smo se na istraživanje varijacija u broju prijave viđenja UFO-a ovisno o godišnjim dobima, kao i njihovu distribuciju tijekom tjedna i dana. Cilj je bio otkriti razlike u učestalosti ovih događaja ovisno o vremenskom razdoblju, što bi moglo pružiti uvid u eventualne obrasce ili vanjske čimbenike koji utječu na pojavu viđenja UFO-a.

### 2.2.1 Broj UFO – viđenja po godišnjim dobima

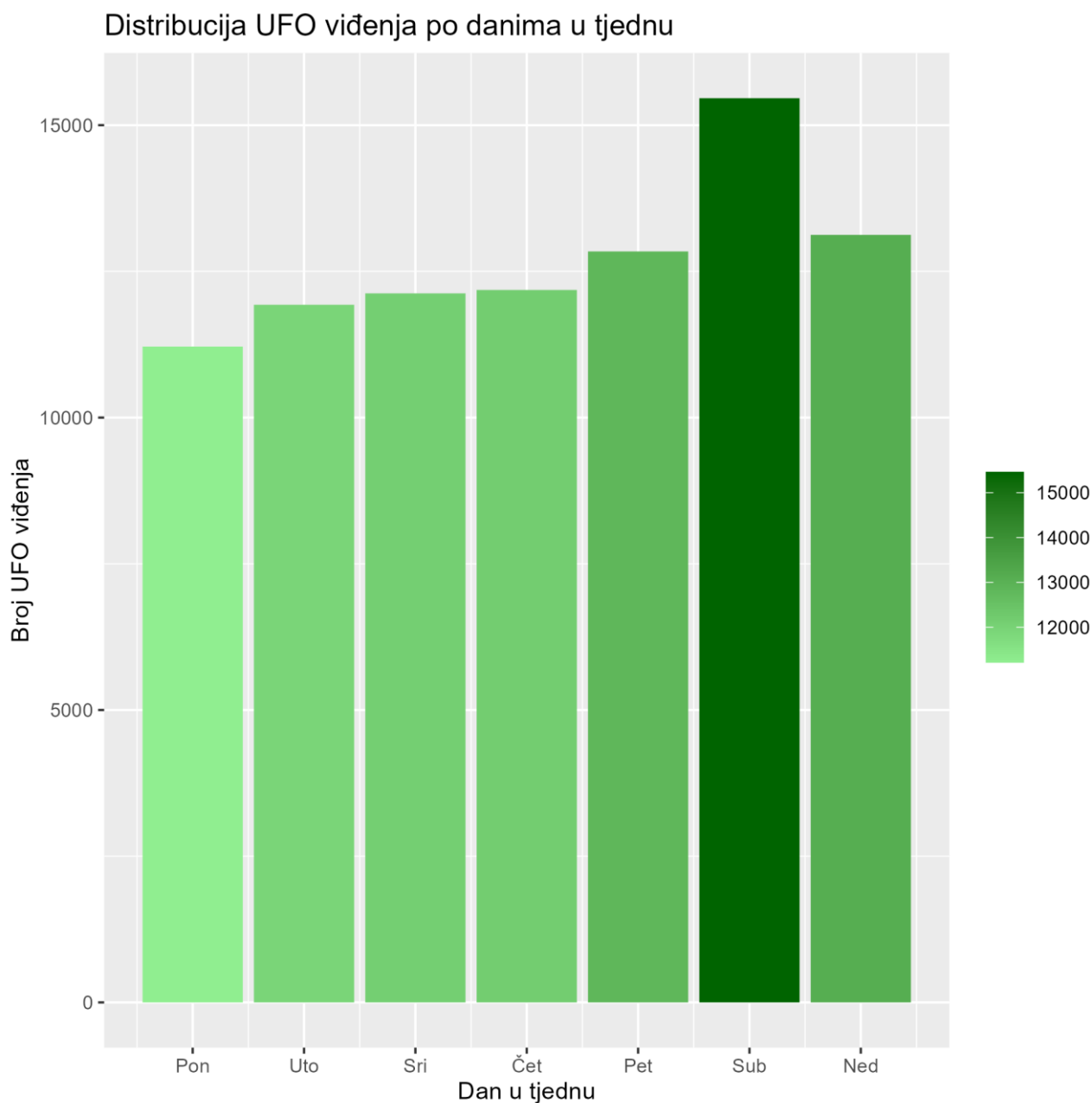
Graf (slika 4) prikazuje distribuciju UFO pojava prema godišnjim dobima. Primjećujemo da se većina pojava događa tijekom ljetnih mjeseci, dok su pojave tijekom zimskih i proljetnih razdoblja manje zastupljene. Ovaj trend može biti rezultat većeg broja aktivnosti ljudi na otvorenom tijekom ljeta, što povećava vjerojatnost opažanja neobičnih fenomena na nebu. Također, povoljni vremenski uvjeti tijekom ljeta olakšavaju promatranje neba u potrazi za UFO pojavama.



Slika 4 Graf: Distribucija UFO viđenja po godišnjim dobima

### 2.2.2 Broj UFO – viđenja tijekom tjedna

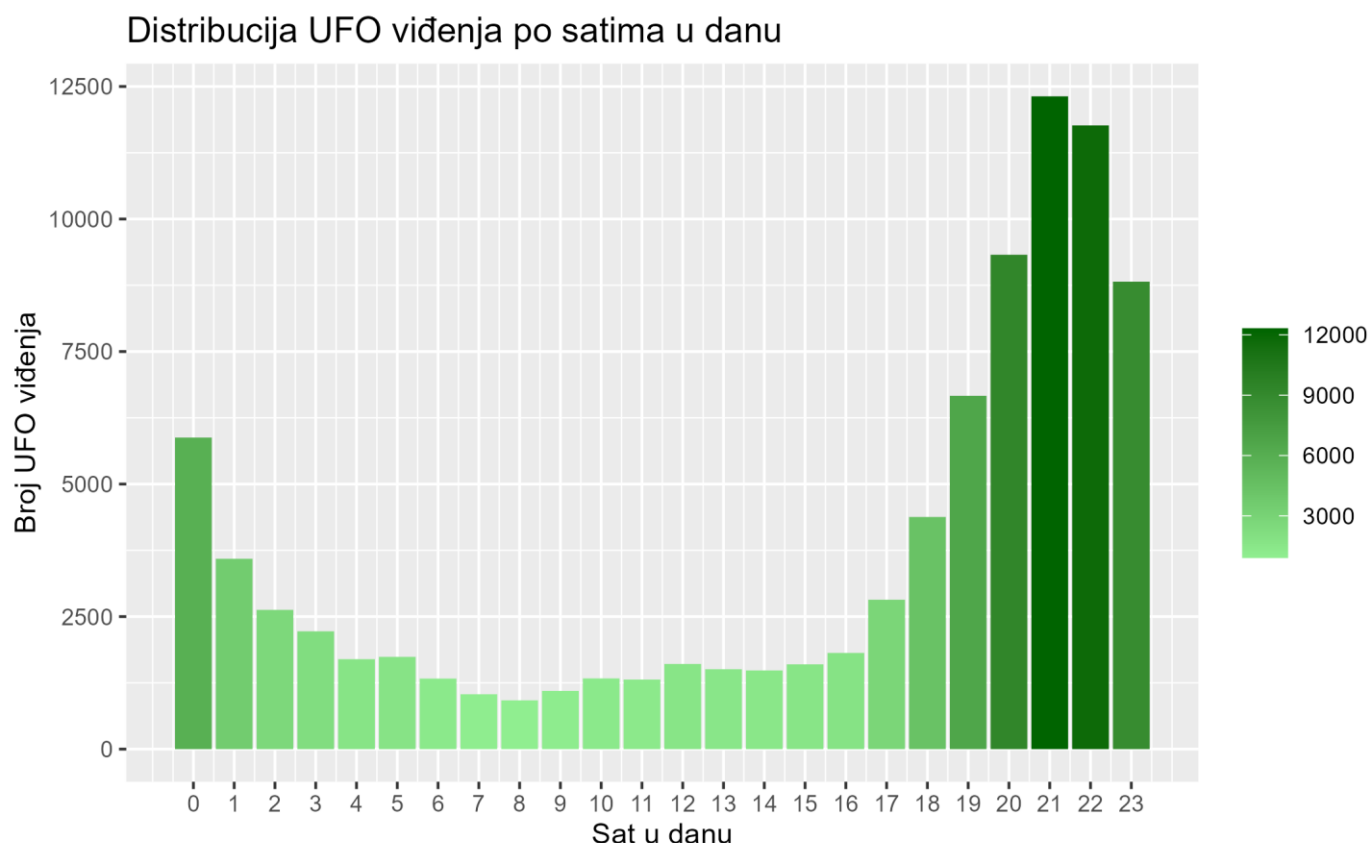
Sljedeći graf prikazuje distribuciju UFO pojava tijekom različitih dana u tjednu. Primjećujemo da je subota dan s najviše zabilježenih NLO pojava, dok je srijeda dan s najmanje takvih događaja. Ovo bi moglo biti povezano s činjenicom da su subote obično dan kada ljudi imaju više slobodnog vremena i veća je vjerojatnost da će provoditi vrijeme na otvorenom ili promatrati nebo, dok su srijede obično radni dani s manje prilika za takve aktivnosti.



Slika 5 Graf: Distribucija UFO - viđenja tijekom tjedna

### 2.2.3 Broj UFO – viđenja tokom dana

Graf sa slike 6 prikazuje distribuciju UFO pojava tijekom različitih sati u danu. Iz grafa se jasno vidi da se većina pojava događa tijekom večernjih i noćnih sati, dosežući vrhunac između 20:00 i 22:00 sata. Ovo bi moglo biti povezano s činjenicom da su ti sati često period kada je nebo tamnije i kada je lakše primijetiti svjetleće objekte na nebu. Također, večernji sati su vrijeme kada ljudi često idu u izlaske, odnosno provode vrijeme na otvorenom.



Slika 6 Graf: distribucija UFO - viđenja tokom dana

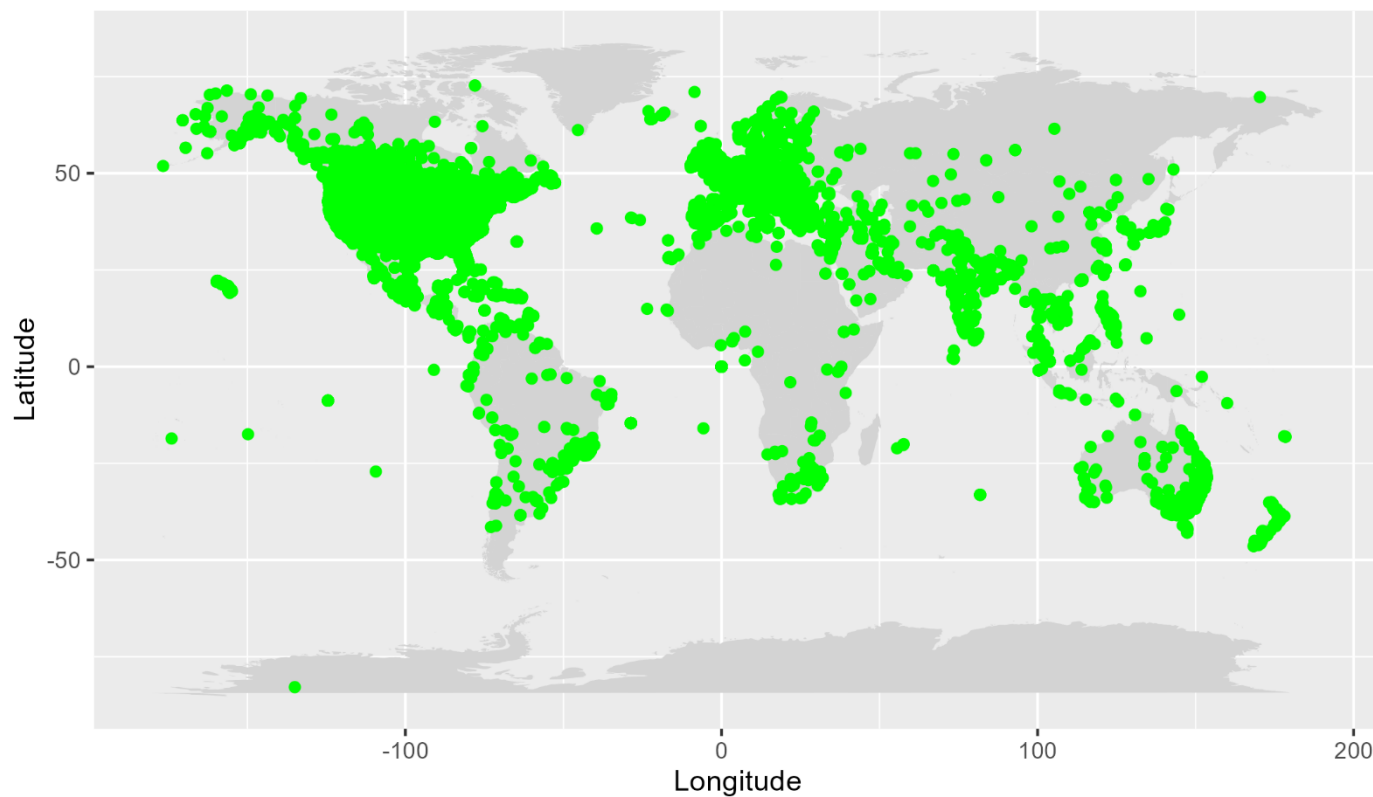
### 2.3 Analiza UFO – viđenja prema geografskom položaju

U trećoj fazi naše analize, istraživali smo viđenja UFO-a s obzirom na njihov geografski položaj. Izradili smo kartografski prikaz prijave UFO-a po svijetu. Ova analiza omogućila je uvid u geografsku distribuciju viđenja UFO-a te identifikaciju potencijalnih obrazaca ili područja s povećanom incidencijom ovih događaja. Također, izradili smo posebne grafove koji su prikazivali distribuciju viđenja po državama, pružajući detaljniji uvid u incidenciju viđenja UFO-a na regionalnoj razini te odgovor na našu pretpostavku da je SAD država sa najviše UFO viđenja.

### 2.3.1 UFO – viđanja na karti svijeta

Ovaj graf prikazuje da se diljem svijeta najviše UFO pojava bilježi u SAD-u i Europi. Ovo zapažanje može biti rezultat većeg tehnološkog razvoja u tim regijama. Primjerice, sve veća upotreba dronova može izazvati vizualna opažanja koja se pogrešno tumače kao UFO pojave, što doprinosi povećanju broja zabilježenih slučajeva u tim područjima.

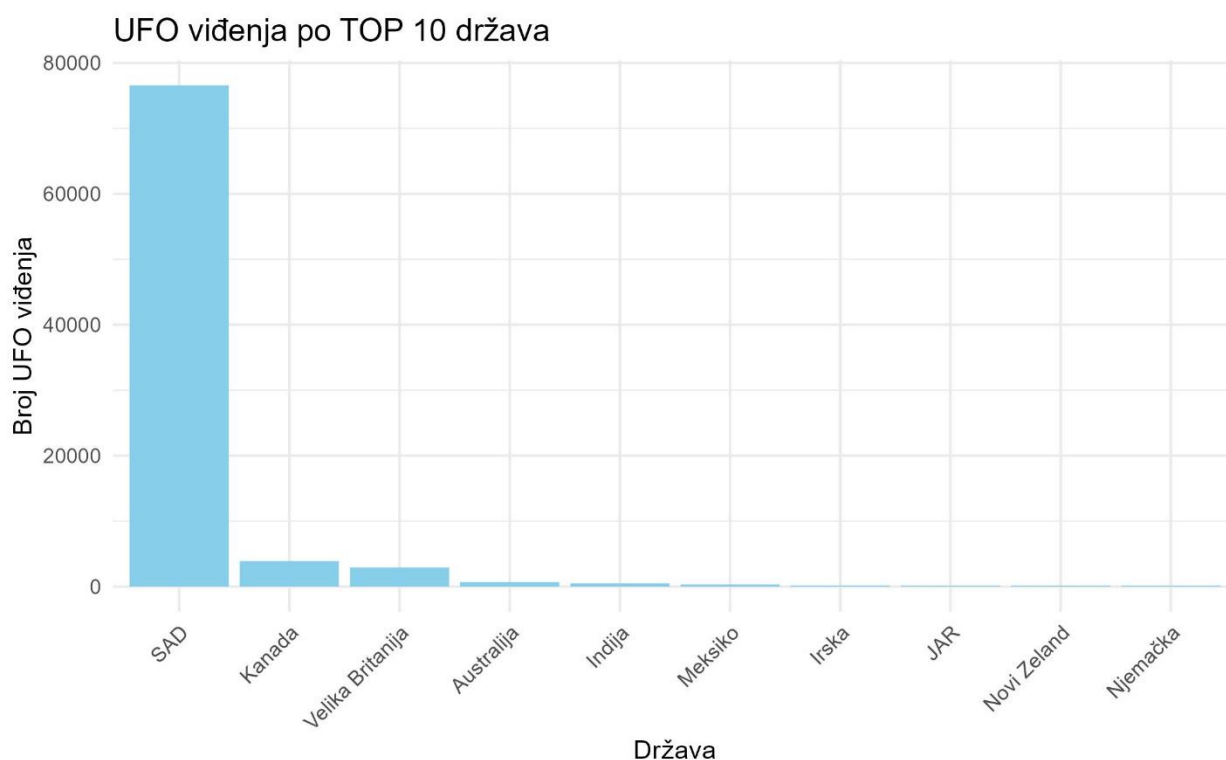
Distribucija UFO viđenja širom svijeta



Slika 7 Graf: Distribucija UFO - viđenja na karti svijeta

### 2.3.2 Broj UFO – viđenja po državama

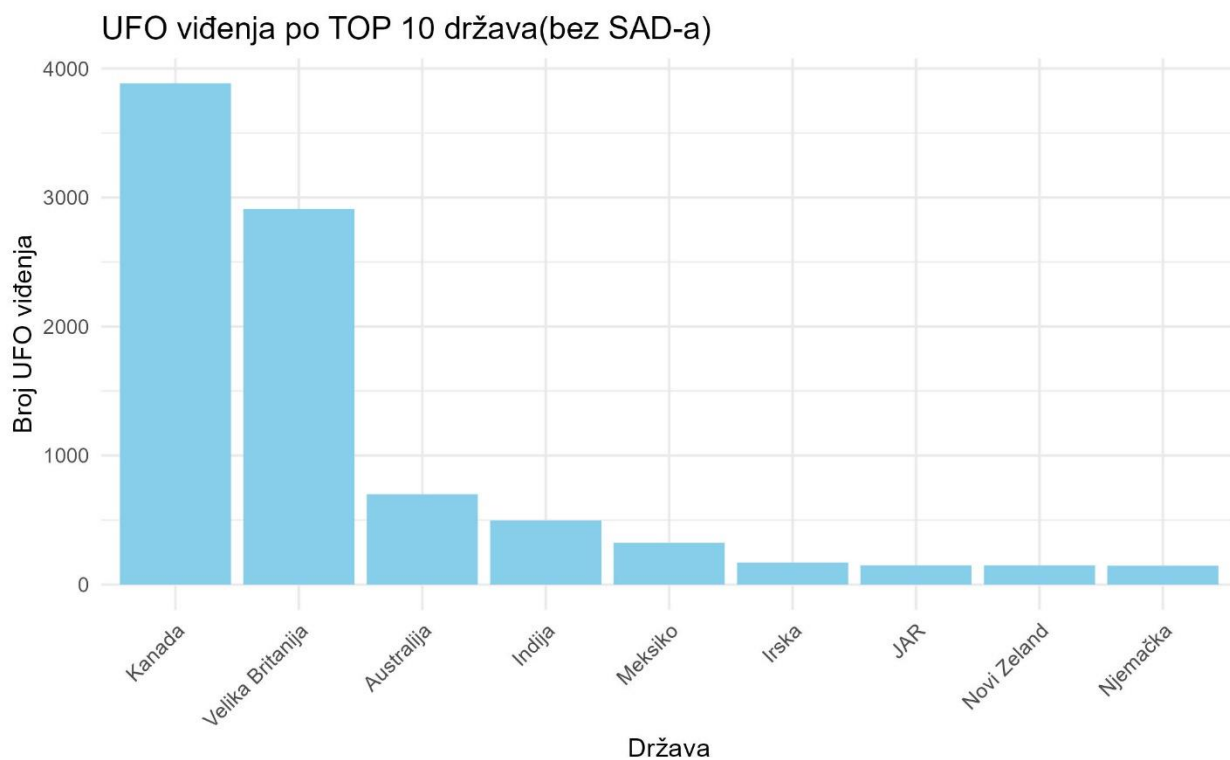
Sljedeći graf predstavlja broj UFO viđenja za 10 država s najvećim brojem zabilježenih slučajeva pruža dublji uvid u stvarni opseg događanja po pojedinim državama. S obzirom na prijašnji graf koji je prikazao geografsku distribuciju viđenja na svjetskoj karti, ovaj graf pruža konkretan uvid u razmjere viđenja po državama. Primarno, uočavamo da Sjedinjene Američke Države značajno prednjače po broju zabilježenih viđenja UFO-a što potvrđuje našu pretpostavku. Postoji nekoliko mogućih razloga za takvu dominaciju. Opći napredak tehnologije, razvoj avionske i svemirske industrije koji može značajno utjecati na povećanje broja prijavljenih slučajeva viđenja neidentificiranih letećih objekata (UFO-a). Ovo može proizaći iz činjenice da su novi tipovi letjelica ili eksperimentalni zrakoplovi, osobito u nejasnim atmosferskim uvjetima, podložni tumačenju kao neidentificirani objekti. Nadalje, postoji općenita zainteresiranost američke populacije za ovakve fenomene. Također, vlada SAD-a ozbiljnije pristupa ovim pitanjima i ima bolje razvijene sustave za bilježenje i analizu ovakvih događaja.



Slika 8 Graf: UFO viđenja u top 10 država

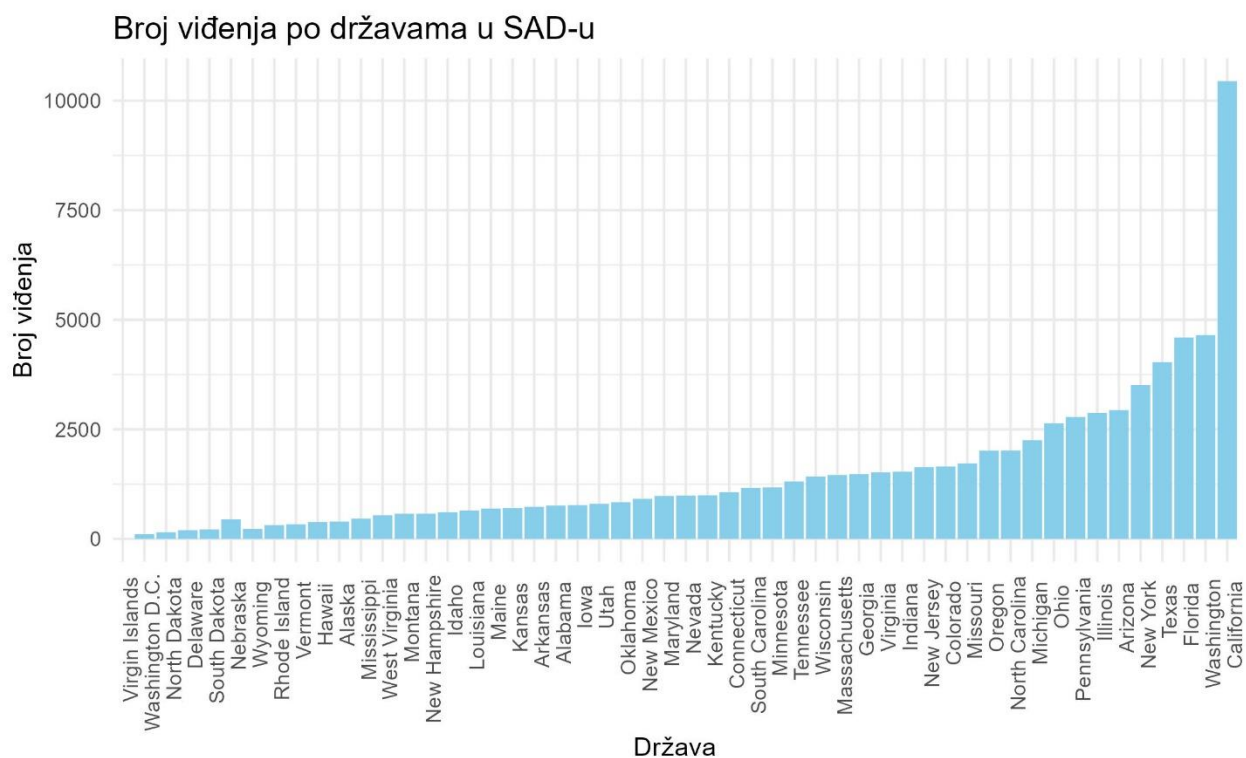
Kako bismo jasnije vidjeli razdiobu viđenja po ostalim državama donji graf prikazuje UFO viđenja za 9 država s najvećim brojem zabilježenih slučajeva isključujući Sjedinjene Američke Države. Možemo uočiti da mnogo država pripada anglofonskom svijetu (države u kojima je engleski službeni jezik). Ta činjenica može biti povezana s nekoliko čimbenika. Anglosaksonske države često dijele slične kulturne, tehnološke i medijske trendove. Stoga trendovi i fenomeni popularni u SAD-u lako se putem društvenih mreža i ostalih medija mogu prenijeti u druge anglofonske države. Isti razlog moguć je i za Meksiko zbog svoje geografske blizine SAD-u. Također razvoj avio industrije u SAD-u može utjecati na veći broj UFO viđenja u njoj geografski bliskim državama Meksiku i Kanadi. Većem broju zapisa UFO viđenja u državama poput Kanade, Velike Britanije, Australije, Irske, Novog Zelanda i Njemačke može doprinijeti i veća svijest o istome. Indiji kao sveukupno petoj državi po UFO viđenjima može doprinositi velika populacija.





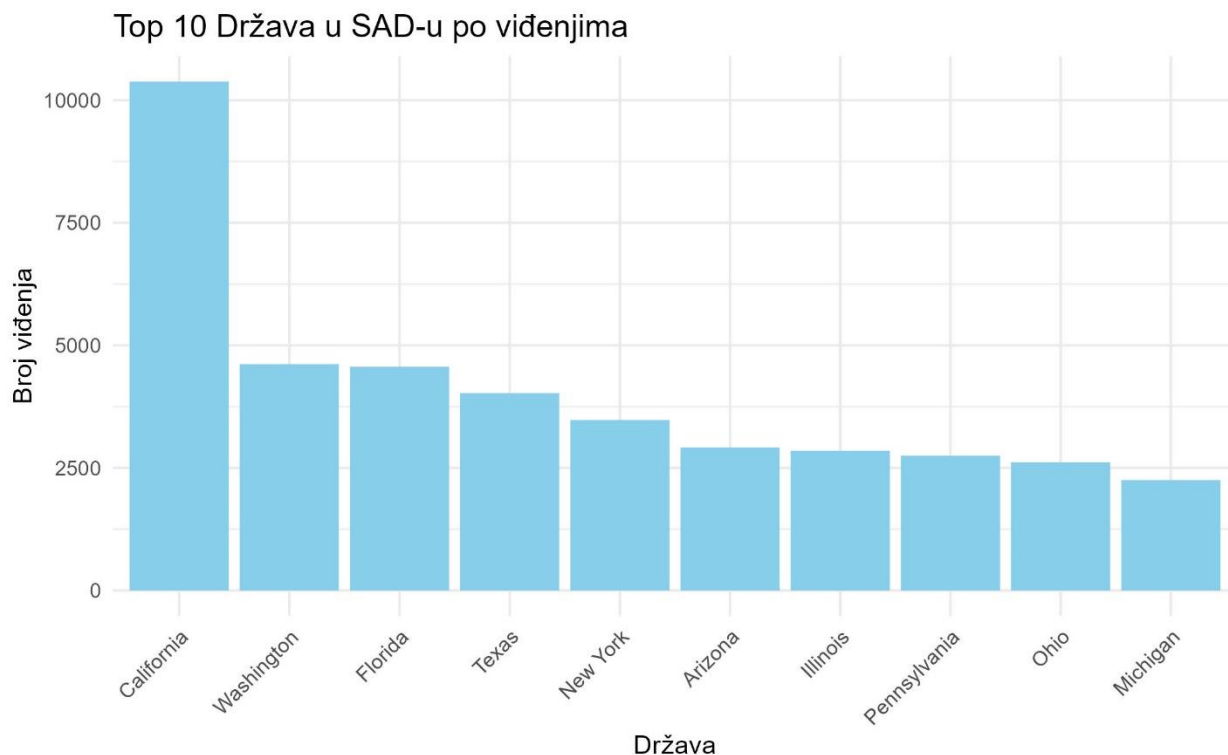
Slika 9 Graf: UFO viđanja u top 10 država bez SAD-a

### 2.3.3 Broj UFO – viđanja po državi (state) u SAD-u



Slika 10 Graf: Broj viđenja po državama u SAD-u

Budući da SAD toliko prednjači po broju zabilježenih UFO događaja napravili smo poseban graf za države(states) unutar SAD-a koju prikazuje donji graf. Države s intenzivnom avionskom i svemirskom industrijom mogu imati veći broj neidentificiranih letjelica ili eksperimentalnih letova koji bi mogli biti percipirani kao UFO-i. Na primjer, Kalifornija ima značajnu prisutnost zračnih i svemirskih tvrtki poput SpaceX-a i NASA-e, što može povećati broj zabilježenih slučajeva i objasniti dominaciju Kalifornije. Također veće države poput Kalifornije, Texasa, Floride i New Yorka često imaju veću populaciju i veću gustoću stanovništva, a što je veća populacija, to je veća vjerojatnost da će se dogoditi više prijava o neobičnim pojavama poput viđenja UFO-a. U nastavku radi preglednosti prilažemo graf koji prikazuje prvih 10 država u SAD-u po zabilježenim UFO viđenjima.



Slika 11 Graf: Države s najvećim brojem UFO viđenja unutar SAD-a

## 2.4 Analiza UFO – viđenja po trajanju

U četvrtoj fazi naše analize, proučavali smo viđenja UFO-a u kontekstu trajanja tih događaja. Analizirali smo distribuciju trajanja viđenja kako bismo identificirali raspodjelu događaja prema njihovoj duljini. Ova analiza omogućila nam je bolje razumijevanje raznolikosti trajanja viđenja UFO-a te identifikaciju eventualnih uzoraka ili neobičnih pojava u vezi s trajanjem tih događaja.

### 2.4.1 Trajanje događaja za TOP 10 država



Slika 12 Graf: Prosječno trajanje događaja za prvih 10 država s najvećim brojem viđenja

Gornji graf prikazuje prosječno vrijeme trajanja (u satima) događaja za prvih 10 država s najviše zabilježenih UFO događaja. Vidimo da je prosjek za većinu ovih država barem 1 sat što se za nešto poput UFO viđenja čini poprilično dugo. Razlog tome su pojedini zapisi koji su imali zabilježeno trajanje i po nekoliko godina. Takve podatke smatrali smo nevaljanima i da zapravo „kvare“ prosjek jer se nešto sigurno za više od godine može o tom događaju više otkriti, pogledati što je to, doći tome bliže, a i male su vjerojatnosti da je nešto stvarno 3 godine bilo konstantno na istome mjestu. Stoga smo takve događaje smatrali krivo zapisanima te u sljedećem koraku napravili novi graf gdje smo izdvojili samo podatke gdje je trajanje događaja kraće od 24h smatrajući da su to vjerojatno najstvarniji događaji.

## 2.4.2 Trajanje događaja za TOP 10 država filtrirano



Slika 13 Graf: Prosječno trajanje događaja za prvih 10 država s najvećim brojem viđenja (max 24h)

Gornji graf prikazuje prosječno trajanje(u minutama) događaja za prvih 10 država s najviše zabilježenih UFO događaja koji su trajali manje ili jednako 24h. Za razliku od prijašnjeg grafa gdje je Velika Britanija prednjačila, sada se ona nalazi tek na 9.-tom mjestu. Slično možemo uočiti i za Kanadu. Zanimljivo je da Irska na oba grafa stoji na posljednjem mjestu sa najkraćim prosječnim vremenom od ovih 10 država. Slično uočavamo i za Njemačku koja se zadržava pri vrhu u oba grafa.

Ovakva zapažanja moguće da proizlaze iz različitih metoda i obrazaca prijavljivanja i istraživanja takvih događaja u tim zemljama.

## 2.4.3 Trajanje događaja u zemljama s najvećim i najmanjim prosjekom trajanja

U nastavku slijedi graf koji opisuje 10 država s najvećim prosječnim vremenom događaja(u minutama) te graf koji prikazuje 10 država s najmanjim prosječnim vremenom događaja(u sekundama). Oba grafa obuhvaćaju samo događaje kraće od 24h. Nismo uspjeli uvidjeti neki razlog za rezultate koje prikazuju osim što većina država s tih dvaju grafa imaju mali broj zapisa te stoga lako odstupaju od top 10 država koje imaju veći broj zapisa.



Slika 14 Graf: Prvih 10 država s najvećim prosječnim vremenom trajanja događaja



Slika 15 Graf: Prvih 10 država s najmanjim prosječnim vremenom trajanja događaja

## 2.5 Analiza komentara o UFO – viđanjima

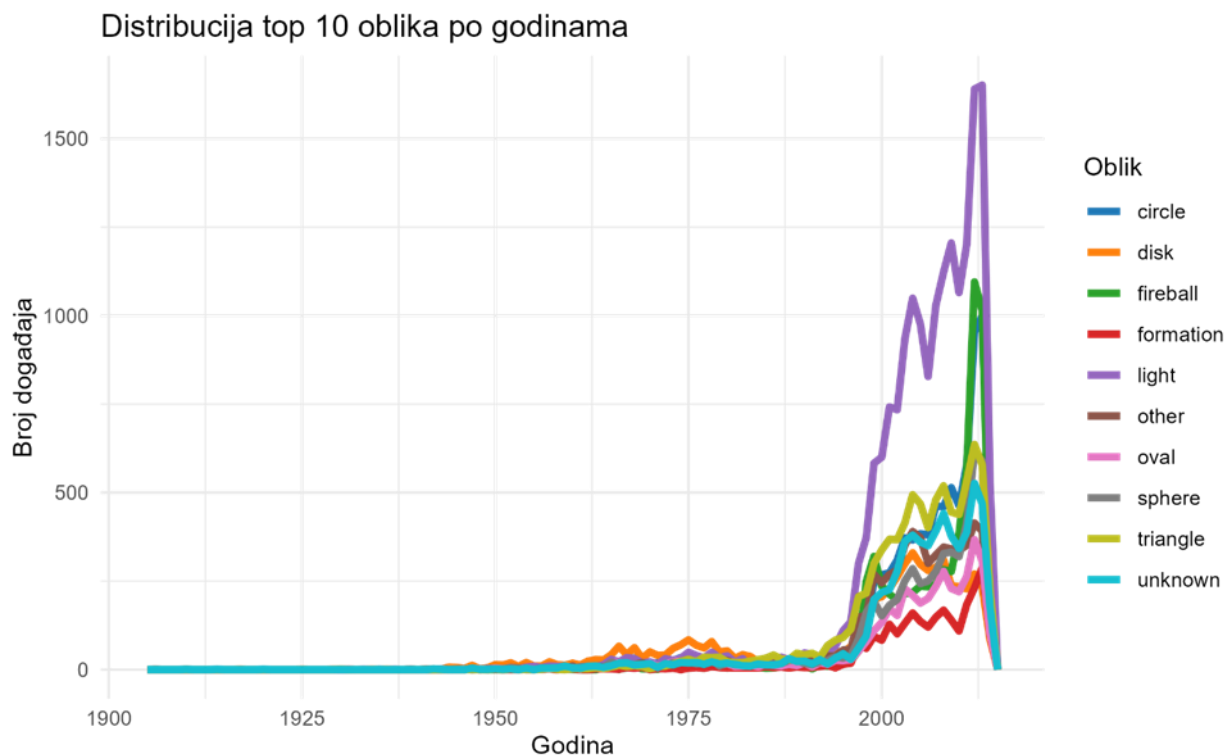
Graf „Oblak riječi“ pruža uvid u najčešće korištene pojmove koji se pojavljuju u opisima doživljaja UFO pojava.



**Slika 16 Graf: „Oblak riječi“ s najčešće korištenim pojmovima unutar opisa doživljaja UFO pojava**

## 2.6 Analiza UFO – viđanja po oblicima

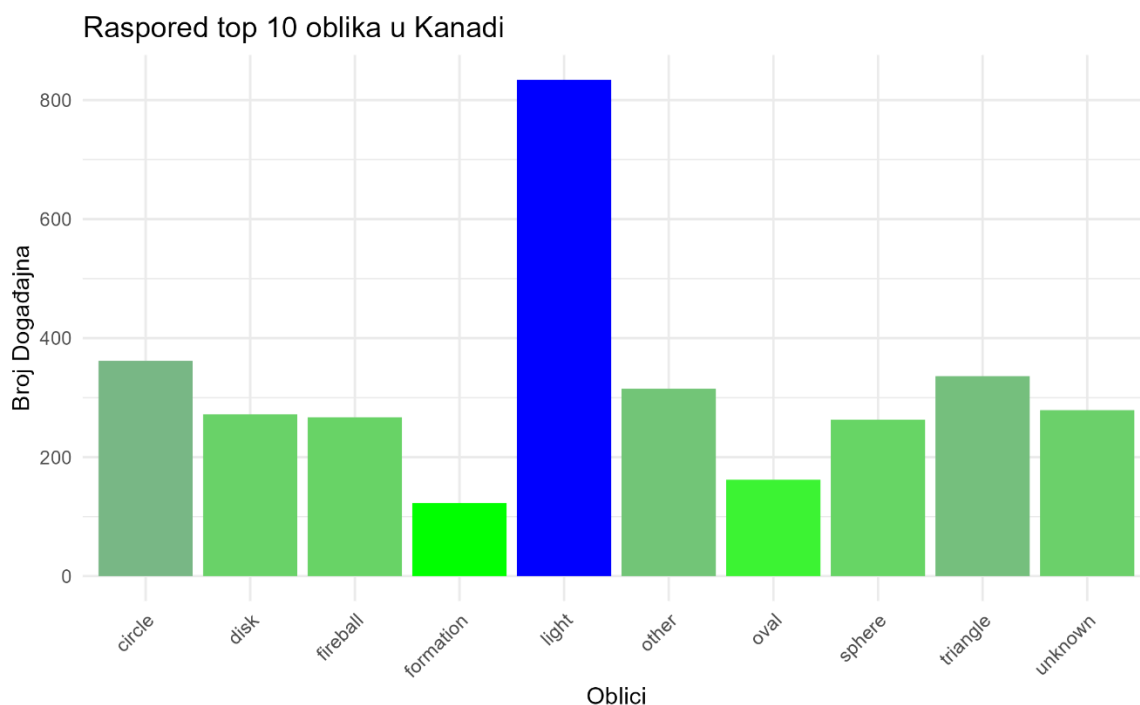
Zadnji graf prikazuje distribuciju top 10 vrsti oblika UFO viđenja po godinama te otkriva zanimljive uzorke. Najviše spominjana pojava je svjetlo, što je očekivano. Interesantan je rast oblika "diska" oko 1975. godine. Taj porast može biti povezan s kulturnim i medijskim utjecajima tog vremena. Diskoidni oblik je često povezan s popularnim predstavama vanzemaljskih letjelica i znanstvenom fikcijom tog razdoblja, što može potaknuti više prijava takvih oblika. Općenito, primjećujemo rast broja prijava svih oblika kako se približavamo sadašnjosti. To može biti rezultat veće svijesti o fenomenu UFO-a, većeg broja medija i platformi za prijavljivanje događaja, kao i veće zainteresiranosti javnosti za ovu temu. Vrijednosti padaju nakon 2015. godine zbog nedostupnosti novijih podataka.



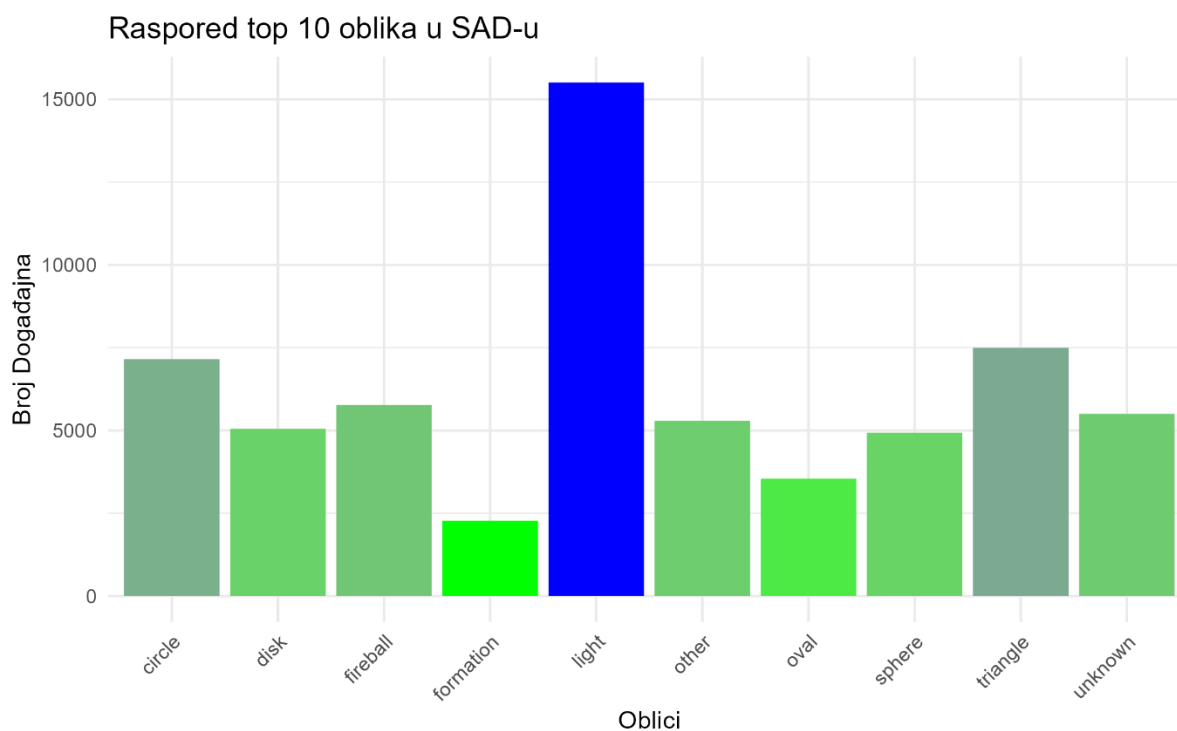
Slika 17 Graf: Distribucija top 10 oblika po godinama

## 2.7 Ostali grafovi

Unutar naše analize, napravili smo nekoliko grafikona koji nisu bili predmet detaljne obrade, ali pružaju uvide u fenomen viđenja neidentificiranih letećih objekata. Ti grafikoni obuhvaćaju različite aspekte, uključujući distribuciju oblika viđenja UFO-a po državama te promjenu broja viđenja kroz godine. Dok možda nisu bili temeljito istraženi, ti grafikoni nude potencijalno zanimljive informacije čitatelju koje bi mogle potaknuti daljnju analizu i razmišljanje o prirodi ovog fenomena.

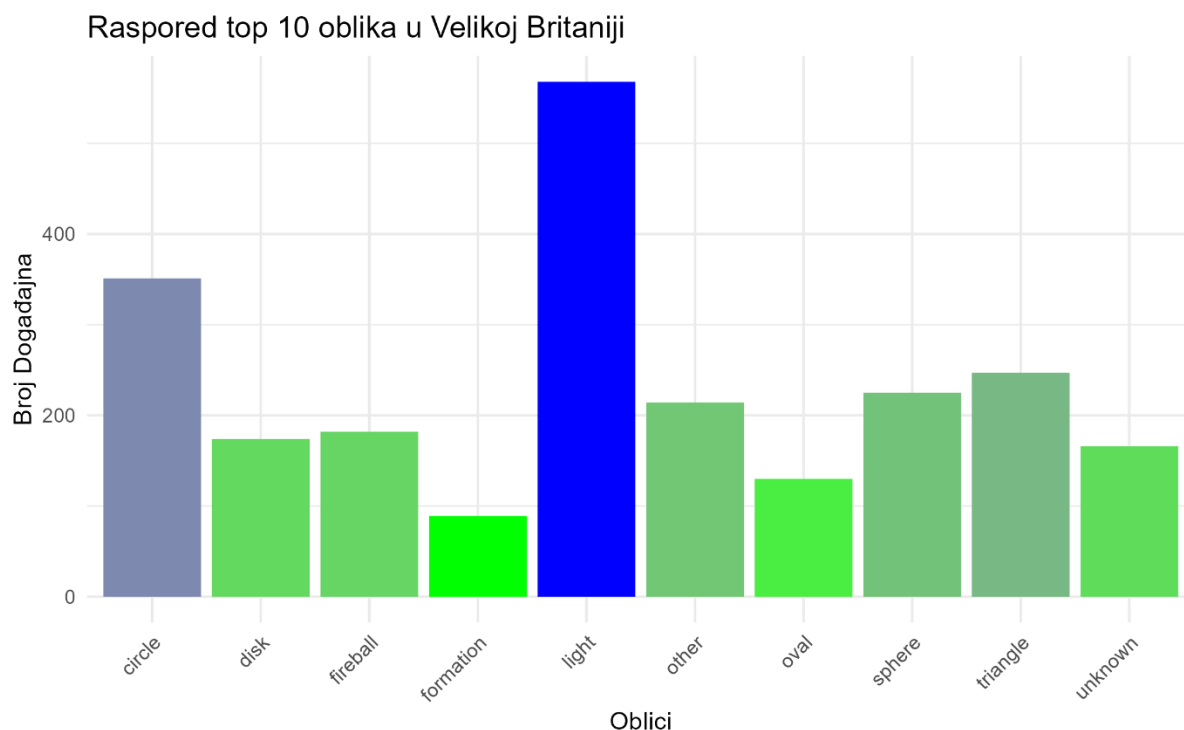


Slika 18 Graf: raspored top 10 oblika u Kanadi

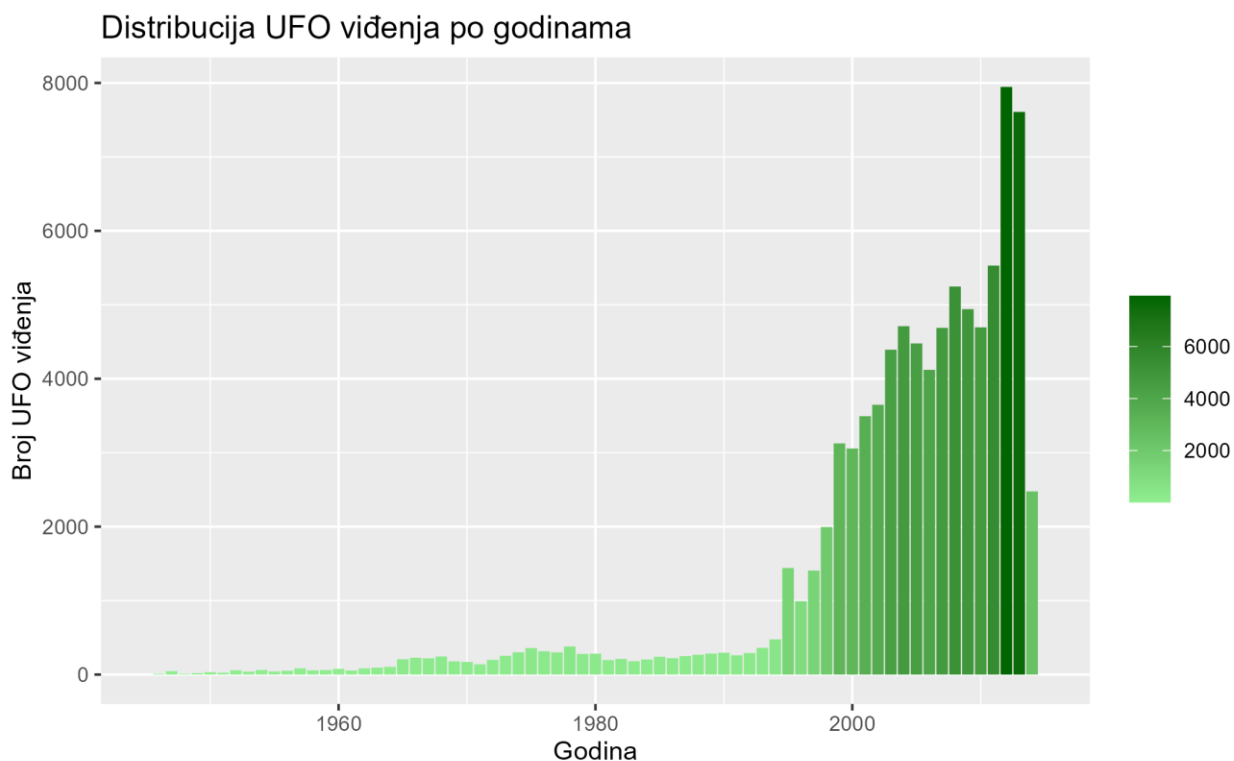


Slika 19 Graf: raspored top 10 oblika u SAD-u





Slika 20 Graf: raspored top 10 oblika u Velikoj Britaniji



Slika 21 Graf: UFO viđenja po godinama

### 3 Zaključak

Uzimajući u obzir sve segmente naše analize viđenja UFO-a, možemo istaknuti nekoliko ključnih spoznaja. Počevši od prve faze istraživanja, gdje smo promatrali trendove viđenja kroz godine, do analize geografske distribucije, trajanja događaja i oblika viđenja, naša analiza pruža dublji uvid u ovaj fenomen.

Prvo, unatoč očekivanjima da će razvoj tehnologije, poput dronova, povećati broj prijava viđenja UFO-a, naša analiza ne pokazuje nagli porast u predviđenim razdobljima. Također čak ni razvoj svemirske tehnologije i postignuća poput prvog slijetanja na mjesec nisu uzrokovala porast viđanja UFO-a. Ovo sugerira da drugi faktori, poput promjena u svijesti i medijskom pokrivanju, mogu imati veći utjecaj na prijavljivanje ovakvih događaja.

Drugo, geografska analiza ukazuje na to da Sjedinjene Američke Države dominiraju u broju prijava, što može biti rezultat veće svijesti i prijavljivanja ovih događaja u zemlji. Također, uočavamo zanimljive obrasce u distribuciji viđenja po državama, što može ukazivati na razlike u kulturi, tehnološkom razvoju i dostupnosti informacija.

Treće, analiza trajanja događaja otkriva varijabilnost u prijavama, s pojedinim događajima koji traju nekoliko sekundi pa do nekih koji traju i nekoliko godina. No, filtriranjem podataka, možemo bolje razumjeti stvarnu distribuciju trajanja događaja.

Jedna od stvari koja(barem za nas) nije bila iznenađujuća je frekvencija riječi u opisu događaja te viđen oblik.

Ova analiza pokazala nam je da neki događaji koje bi očekivali ipak nisu povezani sugerirajući da je ovo područje kompleksnije od očekivanog. Daljnja analiza i istraživanja mogli bi produbiti naše razumijevanje ovog fenomena i otkriti nove aspekte koji stoje iza prijava UFO-a.