

Izbor alata i dobre prakse kod izrade vizualizacija podataka

Gabelić, Ira

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, Faculty of Science / Sveučilište u Splitu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:166:082073>

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

Download date / Datum preuzimanja: **2024-12-01**

Repository / Repozitorij:

[Repository of Faculty of Science](#)



SVEUČILIŠTE U SPLITU
PRIRODOSLOVNO MATEMATIČKI FAKULTET

DIPLOMSKI RAD

**Izbor alata i dobre prakse kod izrade
vizualizacija podataka**

Ira Gabelić

Split, Rujan 2024.

Temeljna dokumentacijska kartica

Diplomski rad

Sveučilište u Splitu
Prirodoslovno-matematički fakultet
Odjel za Informatiku
Ruđera Boškovića 33, 21000 Split, Hrvatska

IZBOR ALATA I DOBRE PRAKSE KOD IZRADE VIZUALIZACIJA PODATAKA

Ira Gabelić

SAŽETAK

Zbog velikog broja podataka koji se generiraju suočeni smo s potrebom za većim brojem vizualizacija koja često rezultira s lošim vizualizacijama. Područje je specifično po tome što se s vizualizacijama susreću svi, a ne samo stručne osobe koje mogu primijetiti da ona ne odgovara podacima ili da je zbunjujuća pa je stoga važno precizno prikazati podatke. U radu je napravljeno istraživanje kojim su testirane prakse koje su u znanstvenim člancima opisane kao dobre, a rezultati ankete će ih potvrditi ili demantirati.

Ključne riječi: Vizualizacija podataka, dobre prakse, alati, Python, Power BI, Tableau

Rad je pohranjen u knjižnici Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu

Rad sadrži: [48] stranica, [44] grafičkih prikaza, [4] tablice i [40] literaturnih navoda. Izvornik je na hrvatskom jeziku.

Mentor: **Doc. dr. sc. Goran Zaharija**, *docent Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu*

Ocjenjivači: **Doc. dr. sc. Goran Zaharija**, *docent Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu*

Dr. sc. Divna Krpan, *docent Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu*

Dr. sc. Monika Mladenović, *docent Prirodoslovno-matematičkog fakulteta,
Sveučilišta u Splitu*

Rad prihvaćen: Rujan 2024.

Basic documentation card

Graduate Thesis

University of Split
Faculty of Science
Department of Computer science
Ruđera Boškovića 33, 21000 Split, Croatia

SELECTING TOOLS AND GOOD PRACTICES FOR DATA VISUALIZATIONS

Ira Gabelić

ABSTRACT

Abstract

With the increasing amount of data generated daily, the need for effective visualization has become necessary. However, this often leads to poor or misleading visualizations, which is a significant concern as data visualization is not confined to scientific fields—it's something everyone encounters, often unknowingly. This paper explores the principles deemed effective in scientific literature, testing their validity through empirical research. The results of this survey aim to either support or challenge these established best practices.

Key words: Data visualization, good practices, tools, Python, Power BI, Tableau

Thesis deposited in library of Faculty of science, University of Split

Thesis consists of: [48] pages, [44] figures, [4] tables and [40] references

Original language: Croatian

Mentor: **Goran Zaharija, Ph.D.** *Assistant Professor of Faculty of Science, University of Split*

Reviewers: **Goran Zaharija, Ph.D.** *Assistant Professor of Faculty of Science, University of Split*

Divna Krpan, Ph.D. *Assistant Professor of Faculty of Science, University of Split*

Monika Mladenović, Ph.D. *Assistant Professor of Faculty of Science, University of Split*

Thesis accepted: September 2024.

IZJAVA

kojom izjavljujem s punom materijalnom i moralnom odgovornošću da sam diplomski rad s naslovom IZBOR ALATA I DOBRE PRAKSE KOD IZRADE VIZUALIZACIJA PODATAKA izradila samostalno pod voditeljstvom Doc. Dr. sc. Gorana Zaharije. U radu sam primijenila metodologiju znanstvenoistraživačkog rada i koristila literaturu koja je navedena na kraju diplomskog rada. Tuđe spoznaje, stavove, zaključke, teorije i zakonitosti koje sam izravno ili parafrazirajući navela u diplomskom radu na uobičajen, standardan način citirala sam i povezala s fusnotama s korištenim bibliografskim jedinicama. Rad je pisan u duhu hrvatskog jezika.

Studentica

Ira Gabelić

Hvala Ivi i Martini koje su sa mnom prošle svaku prepreku i slavile svaki uspjeh kao da je njihov, na podršci, zajedničkom učenju, neprospavanim noćima, ali i neizmjereno puno smijeha i lijepih trenutaka.

Hvala mom Mati koji je kroz cijelo vrijeme bio bezuvjetna podrška i motivator.

I na kraju hvala mojim roditeljima i baki Mirjani, da nisu bili uporni ne bih nikad završila faks.

Sadržaj

UVOD.....	1
1 MOTIVACIJA.....	2
2 VIZUALIZACIJA PODATAKA OPĆENITO	3
2.1 Što?.....	3
2.2 Zašto?.....	3
2.2.1 Analiza.....	4
2.2.2 Pretraga.....	4
2.2.3 Upiti.....	5
2.2.4 Ciljevi	5
2.3 Kako?.....	6
3 PREGLED LITERATURE	8
4 METODOLOGIJA ISTRAŽIVANJA	13
4.1 Metrike.....	14
4.2 Odabir grafikona	15
4.3 Odabir alata.....	17
4.3.1 Power BI.....	17
4.3.2 Tableau	18
4.3.3 Python biblioteke	18
4.3.4 Matplotlib	18
4.3.5 Seaborn.....	19
4.4 Odabir skupa podataka	19

5	PRAKTIČNI DIO.....	21
5.1	Korišteni alati.....	21
5.1.1	Stupčasti grafikon.....	21
5.1.2	Mali višekratnici – linijski grafikon.....	27
5.1.3	Kružni grafikon.....	32
5.1.4	Raspršeni ili linijski grafikon?.....	33
5.1.5	Složeni stupčasti dijagram.....	39
5.2	Upitnik.....	42
6	REZULTATI ISTRAŽIVANJA.....	44
6.1	Rezultati usporedbe biblioteka i alata.....	44
6.2	Rezultati upitnika.....	44
7	ZAKLJUČAK.....	48
8	TABLICA SLIKA.....	49
9	LITERATURA.....	51

UVOD

Poznata izreka kaže da slika vrijedi tisuću riječi, a upravo to je jedan od razloga za korištenje vizualizacije podataka. Gotovo svakim klikom koji se napravi na internetu generira se neki podatak. S obzirom na milijune korisnika interneta diljem svijeta, očekivano je rapidno porastao i broj podataka koje je potrebno interpretirati. S tim pomaže vizualizacija koja podacima daje svrhu te je postala neophodna kako bi ih ljudi mogli razumjeti i prikazati. Podaci se mogu prikazati koristeći grafikone, ali pažljivo treba odabrati koji će najbolje i najjasnije prikazati podatke na način na koji i laici mogu razumjeti. Svrha vizualizacije je da se brojčane vrijednosti prikažu vizualno jer je ponekad jedino tako moguće ispravno interpretirati podatke. Zapravo, za vizualizaciju bi bilo bolje reći kako otkriva podatke umjesto da ih prikazuje. Primjer takve situacije bi bio Anscombeov kvartet, odnosno poznati vizualizacijski primjer kojim se prikazuje važnost eksplorativne analize podataka (engl. Exploratory data analysis) i vizualizacije [1]. Naime, u spomenutom kvartetu se nalaze četiri skupa podataka (engl. Dataset) koji imaju ista statistička svojstva kao što su standardna devijacija, varijanca, prosjek i ostalo, ali kad se vizualno prikažu na grafikonu, izgledaju skroz drugačije. To govori o važnosti vizualizacije podataka prije primjene raznih algoritama za izradu modela iz njih, što sugerira da se značajke podataka moraju iscertati kako bi se vidjela distribucija uzoraka koji mogu pomoći da se identificiraju različite anomalije prisutne u podacima kao što su outlieri, raznolikost podataka, linearna odvojivost podataka itd. [2]

U ovom radu će se obraditi definicija vizualizacije podataka, objasniti za što se ona koristi, na koji način odabrati grafikone te kroz nekoliko alata na primjerima usporediti njihove funkcionalnosti te prednosti i mane. Također će se prikazati primjeri dobre, ali i loše prakse.

1 MOTIVACIJA

Tijekom prošle godine proizvedeno je oko 120 zetabajta podataka, odnosno u mjeri bližoj laiku 120 milijuna gigabajta. U skladu s tim se razvija sve veća potreba za prikazom tih podataka pa se tako svugdje može susresti s nekom vrstom vizualizacije podataka bilo to u trgovini gdje je prikazana karta države s označenim gradovima i mjestima gdje se nalaze poslovnice te trgovine, bolnice ili ambulante koje infogramima prikazuju koje krvne grupe nedostaju ili na dnevniku gdje se takvim prikazima mogu pratiti porasti cijena i režija, inflacija ili recimo kretanje oblaka na vremenskoj prognozi. Upravo iz razloga što su takvi prikazi postali „in“ i nezaobilazni, sve više se viđaju vizualizacije koje nisu napravljene u skladu s načelima dobre prakse te krajnje korisnike, a pogotovo laike, zbunjuju. Iz tog razloga je važno da se prate načela i eventualne novosti, a cilj ovog rada je prikazati dobre prakse te ispraviti neke loše koje su već završile na internetu kroz skup podataka koji je zanimljiv i blizak široj populaciji te opširan zbog čega otvara mogućnost za stvaranje dobrih i korisnih vizualizacija te analiza trenutno popularnih alata i biblioteka.

2 VIZUALIZACIJA PODATAKA OPĆENITO

Vizualizacija podataka je grafički prikaz informacija i podataka. Korištenjem vizualnih elemenata kao što su grafikoni i mape, alati za vizualizaciju podataka pružaju pristupačan način da se vide i razumiju trendovi, ekstremi i uzorci u podacima. Odličan je način za prikaz podataka i rezultata osobama koje nisu tehničke struke. [3]

Često osoba koja koristi vizualizaciju podataka nema određeni cilj ili pitanje. Umjesto toga, osoba možda jednostavno ispituje podatke kako bi saznala više o njima s ciljem da dođe do novih otkrića ili da stekne uvid u podatke. [4]

Osim pojednostavljivanja prikaza i interpretacije, cilj vizualizacije podataka je prepoznavanje trendova u podacima, traženje korelacije, ušteda vremena te davanje konteksta.

Kod vizualizacije podataka razlikujemo nekoliko ključnih koncepata koje je Tamara Munzner pojasnila kroz tri pitanja [5]:

- Što?
- Zašto?
- Kako?

2.1 Što?

Glavno pitanje ovog poglavlja je što su podaci koje korisnik vidi. Ovdje je bitno spomenuti tipove podataka te prije svega analizirati koji podaci su na raspolaganju. Osim tipa podatka, važno je znati i njegovu semantiku, odnosno kontekst. Cilj ovog koraka je definiranje strukture u vidu skupa podataka (engl. Dataset) i tipa podataka svakog od atributa.

2.2 Zašto?

Ovo se pitanje odnosi na razlog korištenja vizualizacije. U ovom će se poglavlju obraditi akcije i ciljevi. Akcije opisuju krajnji cilj vizualizacije i one su neovisne jedna od drugoj, a one su analiza, pretraga i upit. S obzirom da su neovisne, potrebno je opisati akcije na sve tri razine prilikom planiranja vizualizacije.

Razlog korištenja vizualizacije možemo prikazati kao kombinaciju akcije i cilja.

2.2.1 Analiza

U slučajevima kada koristimo vizualizaciju kao alat za analizu podataka uglavnom postoje dva glavna cilja. Najčešći slučaj je konzumacija (engl. consume) informacija koje su već prethodno generirane i spremljene u obliku prikladnom za prikaz ili obradu.

Osim konzumacije, cilj može biti i stvaranje novih materijala. Tu je cilj označavanje, zabilježavanje i izvođenje.

Cilj označavanja je dodati nove grafičke ili tekstualne anotacije koje su povezane sa jednom ili više već postojećih vizualizacija.

Zabilježavanje mora spremiti ili zapisati elemente vizualizacije kao trajne podatke. To može biti u obliku slike ekrana, popisa zabilježenih elemenata ili lokacija, vrijednosti parametara, zapisa interakcija ili čak anotacija. Ključna razlika u odnosu na označavanje je u trajnom spremanju.


Cilj izvođenja ili deriviranja je proizvesti nove podatkovne elemente na temelju postojećih. Novi atributi se mogu izvući iz postojećih ili mogu biti pretvoreni iz jednog oblika u drugi.

2.2.2 Pretraga

Kod pretrage je najvažnije odgovoriti na pitanje zna li se što pretražujemo ili ne i gdje se to nalazi.

Ako je unaprijed poznato što se traži i gdje se nalazi onda je akcija pretrage jednostavno očitavanje (engl. lookup). Tražnje poznatog podatka na nepoznatoj lokaciji je akcija pronalaska ili lociranje (engl. locate). Ukoliko se ne zna identitet podatka, ali zna se gdje tražiti (uglavnom po atributu) onda je to razgledavanje (engl. browse). Ako se ne zna niti lokaciju podatka radi se o akciji istraživanja (engl. explore). Opisane akcije mogu se vidjeti i na Slika 1 Pretraga (prema), gdje su jednostavno prikazane tablicom.

➔ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

Slika 1 Pretraga (prema [5])

2.2.3 Upiti

Kad je ciljani podatak pronađen, moguće je postavljati upite nad tim podatkom. Ovisno o rasponu ciljeva, upiti se mogu podijeliti na:

1. identificiranje (engl. identify) – ima samo jedan cilj, a to je prikaz karakteristika pronađenog podatka
2. usporedba (engl. compare) – ima više ciljeva, zahtjevniji zadatak (složeniji idiomi)
3. sažimanje (engl. summarize) - puni skup mogućih ciljeva, služi za detaljni pregled svih podataka

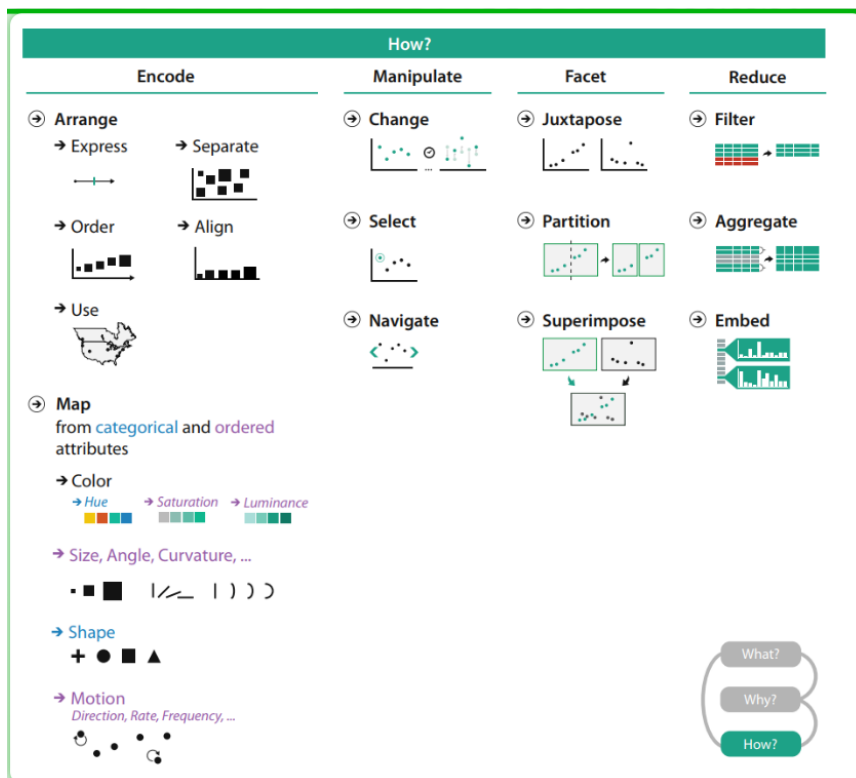
2.2.4 Ciljevi

Sve prethodno navedene akcije se odnose na neki cilj tj. neki dio podataka koji je od interesa korisniku. Ciljevi se mogu podijeliti u četiri kategorije: potpuni podaci, atributi, mrežni podaci, prostorni podaci.

Metode izrade vizualizacije se mogu podijeliti na četiri glavne kategorije:

1. enkodiranje (engl. encode)
2. manipulacija (engl. manipulate)
3. izdvajanje/naglašavanje (engl. facet)
4. reduciranje (engl. reduce)

Na slici se mogu vidjeti primjeri svake od spomenutih kategorija.



Slika 2 Metode izrade vizualizacije (Prema [5])

2.3 Kako?

Pitanje „kako?“ se odnosi na vizualnu reprezentaciju, odnosno kako podatke pretvoriti u vizualni oblik. Ovaj korak se odnosi na odabir ispravnih grafikona te oznaka i kanala. Oznake (engl. Marks) i kanali (engl. Channels) čine osnovu vizualnog enkodiranja. Oznake su osnovni geometrijski likovi koji prikazuju elemente. Primjer oznake su točke, linije... Kanali su zaslužni za izgled oznaka. Primjer su boja (nijansa, saturacija, svjetlina), veličina (dužina, površina, volumen), prostorna pozicija (horizontalna, vertikalna)... [6]

Tako su točke 0-dimenzionalne oznake, pravci i krivulje 1-dimenzionalne, površine 2-dimenzionalne, a 3-dimenzionalne oznake predstavljaju volumen.

Vizualni kanal je način upravljanja izgledom oznaka neovisno o njihovoj dimenzionalnosti.

Ljudska percepcija ima dva različita senzorska modaliteta i prema tome postoje dva tipa kanala:

- identiteta (engl. identity) - pružaju nam informaciju što ili gdje je nešto.
- veličine (engl. magnitude) - govori nam koliko nečega ima.

Također postoje dva načela korištenja kanala, a to su:

- Izražajnost (engl. Expressiveness)
- Djelotvornost (engl. Effectiveness)

Ključni elementi izražajnosti kažu kako se vizualna reprezentacija mora poklapati s tipom podataka, mora izbjegavati elemente koji korisnika mogu navesti na pogrešan zaključak te mora izraziti sve informacije.

Djelotvornost kaže kako bi se ključni atributi trebali prikazati kanalom koji će najprije biti primijećen. Istraživanja su pokazala kako su neki prikazi bolji od drugih pa su redom pozicija, duljina, nagib, površina i jačina boje najbolji prikazi.

3 PREGLED LITERATURE

„Vizualizacija je isključivo ljudska kognitivna aktivnost i nema nikakve veze s računalima“ – riječi su iz knjige Information Visualization [7], dok s druge strane Craig Starbuck kaže kako je to više umjetnost nego znanost [8]. Ove izjave svakako postavljaju čvrst temelj za istraživanje o ljudskoj percepciji kod vizualizacije. Kod odabira grafikona, kao i kod odabira vizualizacijskih alata može nastati problem zbog prevelikog broja opcija i mogućnosti. S obzirom da je cilj vizualizacije pojednostavljivanje i lakše razumijevanje podataka, kao i ubrzavanje procesa, važno je odabrati dobre alate i grafikone koji će nedvosmisleno prikazati ono što trebaju. Izbor grafikona se uglavnom svodi na pitanja koliko varijabli se pokušava pokazati te definiranje njihovog odnosa. Osim toga, postoji još nekoliko stvari koje se moraju uzeti u obzir kao što su boje, orijentacija, oznake na grafikonu i slično i također dodavanje samo bitnih elemenata. U radu iz 2001. [9] je uveden pojam „data-ink ratio“ koji govori koliko je tinte utrošeno na prikazivanje podataka u odnosu na ukupnu količinu potrošene tinte. Cilj je da taj odnos bude što manji, odnosno da se većina tinte koristi za prikaz bitnoga. Na primjer, stupčasti grafikon koji prikazuje odnos dviju varijabla te na osima ima iskazane oznake je dovoljan za prikaz podataka i nije potrebno dodavati legendu sa strane. Na taj način bi se količina tinte povećala pa bi „data-ink ratio“ bio nepovoljniji, isti podatak bi bio dvaput prikazan te bi se potencijalno krajnjeg korisnika navelo na krivi zaključak.

Stephen Midway je u radu [10] iz 2020. godine opisao deset načela kojima bi se trebali voditi pri izradi vizualizacija. Ta načela će biti primijenjena i u ovom radu, a ona su:

- Prije izrade vizualizacije treba se prioritizirati informacija koju želimo prenijeti
- Odabir dobrog alata
- Odabir geometrije (grafikona)
- Boje imaju značenje
- Uključivanje podataka u koje nismo sigurni
- Za prikazivanje promjene je dobro koristiti male višekratnike (engl. Small multiples)
- Nije dobro stavljati rezultate modela u grafikon ako korisnik ne može doći do detalja
- Jednostavne vizualizacije, kompleksno objašnjenje
- Ponekad je dobro koristiti infografike
- Dobro je pitati za savjet

Zaključio je da što dulje netko gleda vizualizaciju i ne razumije poruku, to je vjerojatnije da neće ništa dobiti od te vizualizacije i možda će čak izgubiti nešto od razumijevanja rada. Slijeđenje smjernica i preporuka iz ovog rada može pomoći u izradi dobrih vizualnih prikaza koji izbjegavaju mnoge uobičajene zamke neučinkovitih vizualizacija.

Iako ga znanstvenici ne preferiraju, Bačić, Krbanjević i Jukić su proveli eksperiment čiji su rezultati pokazali kako unatoč većem kognitivnom naporu, kružni grafikon pokazuje bolje rezultate od, recimo, stupčastog grafikona. Njihovo istraživanje je obuhvatilo više vrsta grafikona, a to su kružni, vertikalni i horizontalni stupčasti, lizalica (engl. Lollipop), mjehurićasti (engl. Bubble), složeni stupčasti, mapiranje stabla (engl. Treemap) i prstenasti (engl. Donut) grafikon. U eksperimentu su se mjerile korisničke performanse u vidu točnosti (engl. Accuracy), brzine (engl. Speed) i napora (engl. Effort). Osim navedenog, mjerili su se i pokreti očiju (engl. Eye-tracking). [11]

Dr. Sadhna Bagchi je poglavlje svoje knjige „Words in Motion The Power of Media and Arts“ posvetila bojama. Smatra kako boja igra važnu ulogu u tome da mozak prima informacije, procesira ih i donosi zaključke. Svrha članka je istaknuti vizualizaciju podataka i teoriju boja kako bi prezentacije podataka bile dojmljivije i učinkovitije. Navodi da ako želite ostaviti dojam da su podaci uzastopni ili da variraju u stupnju, a ne vrsti, monokromatska paleta bi dobro funkcionirala [12]. Sve ovo ovisi o upotrebi nijansi (engl. Hue) i zasićenosti (engl. Saturation) u vizualizaciji za prikaz različitih nijansi boja. Također tvrdi, kao i Wan i Weng u članku „Study of the visualization and interaction of data“ [13], kako je temelj teorije boja HSB model, odnosno nijansa (engl. Hue), zasićenje (engl. Saturation) i svjetlina (engl. Brightness) (u nekim izvorima i Lightness).



Slika 3 HSB model (prema [12])

U ljudskoj je prirodi da boje u danom kontekstu asociraju na neku od kategorija. Primjerice želi li se prikazati odnos šuma i vodenih površina, intuitivno je šume prikazati zelenom, a vode plavom bojom. Također crvena boja u pravilu asocira na zabranu, a zelena na dozvolu. Ipak, Starbuck navodi kako boje imaju značenje te da bi crvenu i zelenu boju trebalo izbjegavati iz dva razloga. Prvi je zbog daltonista koji čine oko 4,5% ukupne populacije, a u pravilu su

osjetljivi upravo na te dvije boje, a drugi je taj što u istočnim kulturama crvena boja ne simbolizira nešto negativno ili zabranu, već uspjeh i sreću [8]. Iako možda gledajući s objektivne strane ne bi trebale postojati muške i ženske boje, kod vizualizacije takva podjela može biti praktična. Istraživanje provedeno 2008. godine govori kako djeca zapravo preferiraju primarne boje u usporedbi sa sekundarnima [14], dok drugo iz 2006. godine objašnjava kako ženska djeca uglavnom biraju roze, a muška plave predmete iz razloga što ih roditelji u ranoj dobi okružuju predmetima i odjećom te boje pa ih asociraju na nešto poznato i prirodno. [15]

Osim boje, bitnu ulogu igra i sama orijentacija stupčastog grafikona. U svom istraživanju iz 2005. godine Fischer, Dewulf i Hill su postavili pitanje „je li orijentacija stupčastog grafikona bitna?“ te su mjerenjem zaključili kako je bitna, a njihovi rezultati su pokazali kako je ljudima lakše i brže čitati x os grafikona, odnosno slijeva nadesno nego y os od gore prema dole jer je to prirodnije. [16]

Stupčasti grafikon je u članku „Data visualizations: A literature review and opportunities for technical and professional communication“ opisan kao grafikon koji privlači pažnju te daje informacije koje zahtijevaju minimalan kognitivni napor za razumijevanje. [17]

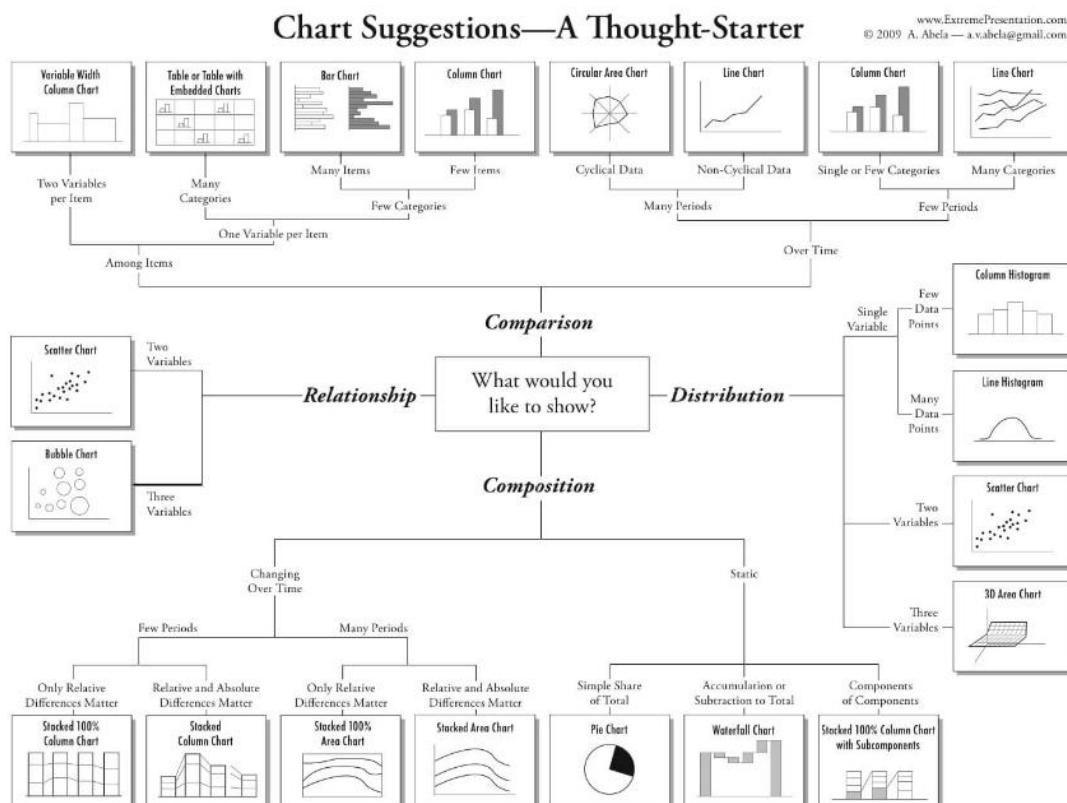
Starbuck navodi još jedan element na kojeg se često ne obraća pažnja, a to je da vrijednosti na y osi kreću od nule (engl. Zero-baseline). Kod takvog prikaza nastaje problem u slučaju kada stupčastim grafikonom prikazujemo slične, istovremeno visoke vrijednosti pa je vizualna razlika ponekad neprimjetna, a to ne bi smjela biti osim ako vrijednosti nisu identične. [8]

Istraživanje koje su proveli Lauer i O'Brien tvrdi ovakav pristup može biti obmanjujuć. Ljudi nisu skloni čitanju osi te ne obraćanja pažnje na detalje jer se smatra da os y uvijek počinje od nule i to može dovesti do krivih zaključaka, a posebno u slučajevima kada osobe koje stoje iza vizualizacije to žele. Kao primjer navode američku predsjedničku kampanju iz 2016. godine u kojoj je jedan od kandidata plasirao stupčaste grafikone bez početne vrijednosti (engl. Baseline) ili s povećanom početnom vrijednosti te svoju prednost od 2% pretvorio u prednost koja vizualno izgleda kao 11% [18]. Sličan primjer prikazuju i Lo, Cao, Yang u radu „Why Change My Design: Explaining Poorly Constructed Visualization Designs with Explorable Explanations“ koji također navodi pomicanje minimalne vrijednosti y osi kao lošu praksu [19].

Friendly i Denis raspršeni grafikon nazivaju najsvestranijim i najkorisnijim izumom u povijesti statističke grafike [20], a Tufte je u radu iz 1983. godine procijenio kako je između 70 i 80 posto grafikona korištenih u znanstvenim radovima upravo raspršeni grafikon [21], što samo za sebe govori o važnosti ovog grafikona. Wang, Han, Zhu, Deussen i Chen u radu „Line Graph

or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series“ ističu kako se linijski grafikon uglavnom uzima za najbolji, iako je on u suštini raspršeni grafikon čiji su podaci (engl. Data points) povezani linijama. Upravo zbog njihove velike sličnosti su odlučili napraviti algoritam koji automatizmom odlučuje hoće li se za dane podatke koristiti linijski ili raspršeni grafikon. [22]

Osim primjene spomenutih načela, važno je odabrati i ispravan način prikaza. Postoje četiri skupine grafikona, ovisno o tome što je potrebno prikazati. To su odnos (engl. Relationship), usporedba (engl. Comparison), distribucija (engl. Distribution) i kompozicija (engl. Composition). Svaka od ovih podjela ima još nekoliko svojih podjela te se uz pomoć slike lako može odrediti idealan grafikon za dani skup podataka. [23]



Slika 4 Tipovi vizualizacija [24]

Kao što je vidljivo na Slika 4 Tipovi vizualizacija , postoji jako puno različitih grafikona kojima se mogu prikazati podaci. Za ispravan odabir grafikona prije svega je potrebno odrediti što se

točno želi prikazati, odnosno na koji se način žele usporediti podaci. Također je potrebno odrediti odnos između podataka koje se prikazuje kako bi se znalo prikazati na ispravan i jasan način.

4 METODOLOGIJA ISTRAŽIVANJA

Metodologija korištena za pregled područja u kontekstu percepcije vizualizacija podataka je u skladu s PRISMA smjernicama. Pretraživanjem baza Web of Science i IEEE Xplore, a pretraga se izvršila uz ključne riječi „data visualization“, „charts“ i „good practices“. Na Web of Scienceu se nalazi 33 članaka koji odgovaraju upitu, a na IEEE Xplore 46. Uz te ključne riječi dodani su sljedeći kriteriji:

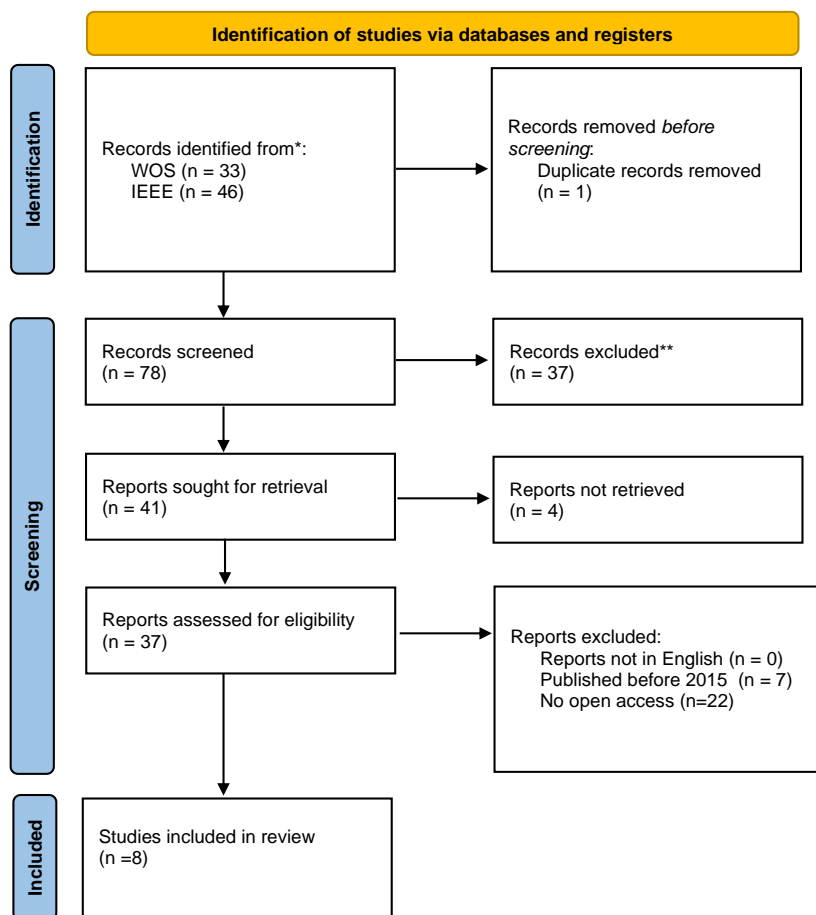
Uključujući kriteriji:

- Članci objavljeni unazad 9 godina (od početka 2015.)
- Članci na engleskom jeziku

Isključujući kriterij:

- Članci koji nisu javno dostupni

Nakon pregleda naslova i sažetka, preostalo je 41 rad i još dodatna 4 su eliminirana jer nisu vezana za područje. Od 37 radova, 29 ih je eliminirano zbog gornjih kriterija tako da je pribrajanjem članaka pronađenih drugim metodama ovim radom obuhvaćen ukupno 20 članaka. Razlog primjene kriterija da su radovi objavljeni od 2015. godine je taj što je te godine izdana prva verzija Power BI-ja koji je Microsoftov alat što je vizualizaciju podataka dodatno približilo običnim korisnicima, dodatno i zbog toga što je za razliku od Tableaua ovaj alat besplatan.



Slika 5 PRISMA dijagram

4.1 Metrike

Prije početka usporedbe, moraju se odrediti metrike po kojima će se određivati je li neki alat dobar ili ne. Metrike su podaci mjere kvantitativne procjene koje se obično koriste za procjenu, usporedbu i praćenje učinka ili proizvodnje. [26]

Ove metrike osiguravaju da su vizualizacije jasne, točne i olakšavaju razumijevanje.

U radu iz 2013. godine Min Chen, Luciano Floridi i Rita Borgo navode kako se ušteda vremena može gledati kao temeljna metrika te da bi trebala biti u fokusu kod izrade vizualizacije [27]. S druge strane u članku „How to evaluate data visualizations across different levels of understanding“ ne osporavaju tu tvrdnju, ali sugeriraju kako svrha vizualizacije nije samo lakše razumijevanje informacija, već da primljene informacije mogu dalje primijeniti [28].

U radu „Measuring effective data visualization“ su definirane kvantitativne i kvalitativne mjere. U ovom će se radu koristiti kvalitativne mjere, konkretno usporedba stručnih i nestručnih osoba te anketa koja je kombinacija opservacije i intervjua. [29]

	<i>Quantitative measurements</i>	<i>Qualitative measurements</i>
Accuracy	<ul style="list-style-type: none"> • Measure the number of interpretation errors 	<ul style="list-style-type: none"> • Interview • Observation • Expert-novice comparison
Utility	<ul style="list-style-type: none"> • Measure the number of achieved benchmark goals • Record the number of times a visualization design is selected by users to conduct a task 	<ul style="list-style-type: none"> • Interview • Observation • Expert-novice comparison
Efficiency	<ul style="list-style-type: none"> • Record task completion time • Record eye movements • Measure the learning curve 	<ul style="list-style-type: none"> • Visualization complexity analysis • Interview • Observation • Expert-novice comparison

Slika 6 Mjere vizualizacije (prema [29])

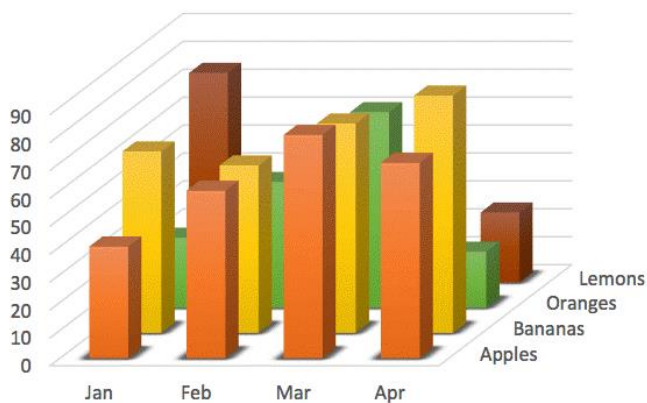
Dobre prakse primijenjene u ovom radu su izbor boja u skladu s kategorijom koja se prikazuje, prikazivanje podataka monokromatskom paletom kad se podaci razlikuju samo u stupnju, vertikalna orijentacija stupčastog grafikona, prikaz grafikona u 2d umjesto 3d, korištenje zero-baselinea, uredan prikaz linijskog grafikona korištenjem malih višekratnika, pretvaranje običnog stupčastog grafikona u složeni kad je to moguće, izrada algoritma koji na temelju učitanih podataka samostalno određuje je li bolje koristiti linijski ili raspršeni grafikon te korištenje jednostavnih vizualizacija uz objašnjenja.

U ovom radu će se u prvi plan staviti dobre prakse kod izrade vizualizacije, a ušteda vremena će se mjeriti i usporediti u različitim alatima. Uz to, vizualizacija podataka je sama po sebi ušteda vremena (uz brojne ostale prednosti) jer je pojednostavljena slika podataka na čiju bi analizu izgubili jako puno vremena. Također će se relevantnim smatrati rezultati istraživanje u obliku ankete koje će se provesti u sklopu ovog rada.

4.2 Odabir grafikona

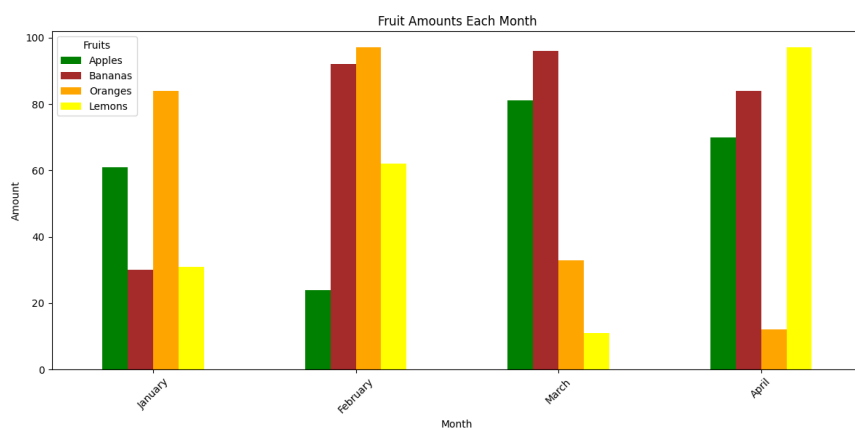
Sve gore spomenuto vrijedi samo u slučaju da je ispravno odabran prikladan način prikaza naših podataka. U protivnom će vizualizacije biti nejasne, besmislene te mogu navesti na pogreške zaključke. Prije svega je potrebno odrediti koliko je varijabli potrebno prikazati te u koju od tih kategorija prikaza spadaju podaci.

Loš primjer vizualizacije će zbuniti korisnika te se iz njega neće moći ništa iščitati. Jedan takav primjer je vidljiv na slici gdje se određeni stupci uopće ne vide od drugih. [30]



Slika 7 Primjer loše vizualizacije 3d stupčastog grafikona [25]

Rješenje ovog problema bi bio obični stupčasti grafikon (engl. bar chart) kojem su stupci poredani jedan do drugog, a ne jedan iza drugog, odnosno dvodimenzionalno umjesto trodimenzionalno. Na taj način bi se jasno vidjela količinu voća po mjesecima i lako uspoređivali njihovu količinu, svako voće bi bilo prezentirano jednom bojom, a vizualizacija bi također sadržavala legendu koja bi povezivala boju i voće. Jedan od principa dobre vizualizacije je da nema potrebe činiti trodimenzionalnima one stvari koje to nisu. U konkretnom slučaju dolazi do okluzije, odnosno stapanja stupaca. Stupci koji nisu u prvom redu su djelomično ili čak potpuno prekriveni stupcima ispred njih. Osim toga, dodijeljene boje nisu intuitivne jer bi prirodno bilo zaključiti kako je u ovom kontekstu jabukama dodijeljena zelena, a narančama narančasta boja, no u ovom slučaju to nije tako. Na slici niže se nalazi prijedlog poboljšane gornje vizualizacije s popraavljenim elementima koji su detektirani kao loša praksa.



Slika 8 Popravljeni primjer stupčastog grafikona

Grafikoni obrađeni u ovom radu su odabrani u svrhu što boljeg opisivanja principa dobre vizualizacije, konkretno po (barem) jedan primjer iz svake kategorije. Primarno je određeno što će se točno prikazivati, a potom su odabrani prikladni grafikoni uz obzir kako će ih u sklopu upitnika analizirati i laici.

Vizualizacije koje će se koristiti u ovom radu su:

- stupčasti grafikon
- kružni grafikon
- linijski grafikon
- malim višekratnici (linijski grafikon)
- raspršeni grafikon
- složeni stupčasti grafikon

4.3 Odabir alata

Kao i u bilo kojem drugom području, vizualizacijski alati su u zadnjih desetak godina iznimno napredovali te danas postoji široka paleta izbora alata koji je najprikladniji za potrebe danog skupa podataka. U ovom će radu biti predstavljeno nekoliko popularnih alata za vizualizaciju podataka.

Alati koji će se u ovom radu koristiti su Tableau i Power BI te biblioteke Pythona Matplotlib i Seaborn uz neke sporedne biblioteke koje nisu namijenjene samoj vizualizaciji već pripremi i oblikovanju podataka te predviđanju. Ovi alati su odabrani zbog velikog broja mogućnosti te činjenice da su jedni od najkorištenijih alata za vizualizaciju podataka. Također je važno spomenuti kako Tableau i Power BI ne zahtijevaju pisanje koda već se uglavnom baziraju na „drag and drop“ sistemu, dok je za preostale biblioteke potrebno kodiranje i poznavanje jezika kojima pripadaju. Osim svoje jednostavnosti, alati imaju veliku zajednicu (engl. Community) zbog čega se može rješenje nekog problema potražiti na internetu ili postaviti pitanje na koje će drugi korisnik odgovoriti. Priprema i učitavanje podataka je također jednostavno.

4.3.1 Power BI

Power BI je skup softverskih usluga, aplikacija i konektora koji rade zajedno kako bi izvore podataka pretvorili u koherentne, vizualno sveobuhvatne i interaktivne izvještaje. Razvijen je 2011. godine. Pruža mogućnosti skladištenja podataka uključujući njihovu pripremu, rudarenje

i interaktivne nadzorne ploče. Podaci mogu biti Excel proračunska tablica ili zbirka hibridnih skladišta podataka temeljenih na oblaku (engl. Cloud) i na lokaciji. Power BI omogućuje jednostavno povezivanje s izvorima podataka, vizualizaciju i dijeljenje. [31] Postoje tri verzije Power BI-ja, desktop, service i mobilna aplikacija. Ako se planiraju objaviti vizualizacije, to se može učiniti preko Power BI servicea.

4.3.2 Tableau

Tableau je osnovan 2003. kao rezultat projekta koji je imao za cilj poboljšati tijek analize i učiniti podatke dostupnijima ljudima putem vizualizacije. Njihova misija je pomoći ljudima da razumiju podatke. [32] Tableau proizvodi postavljaju upite relacijskim bazama podataka, bazama podataka u oblaku i proračunskim tablicama za generiranje vizualizacija podataka u obliku grafikona. Softver također može izdvojiti, pohraniti i dohvatiti podatke iz podatkovnog mehanizma u memoriji. [33]

4.3.3 Python biblioteke

Matplotlib je Pythonova biblioteka koja služi za prikaz običnih grafova, dok se za estetski zahtjevnije vizualizacije (npr. Karta svijeta) koristi Seaborn.

4.3.4 Matplotlib

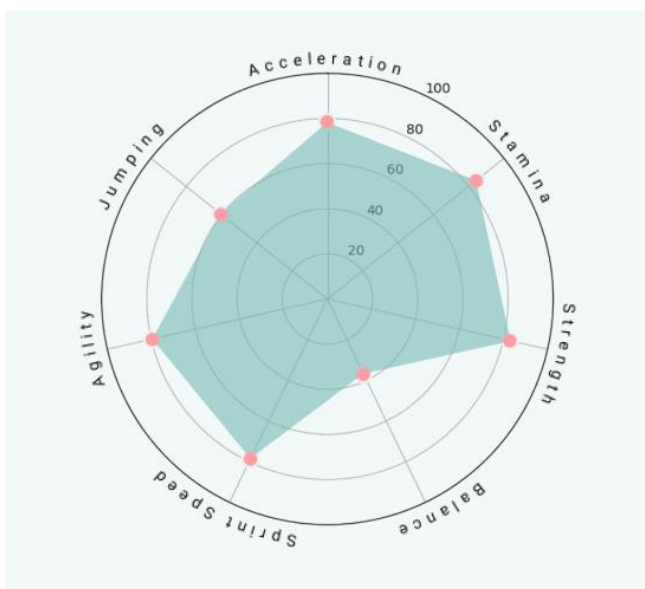
Matplotlib je biblioteka za vizualizaciju podataka te grafičko crtanje koja je izgrađena na NumPy nizovima. U Matplotlibu se *figure* smatra spremištem (engl. container) osi, grafikona, teksta i labela. Pozivanjem funkcije *axes* bez parametara dobivamo prazni okvir u kojem su vidljive samo osi *x* i *y*. `matplotlib.pyplot` je skup funkcija zbog kojih matplotlib radi poput MATLAB-a. Svaka funkcija `pyplot`-a vrši neke promjene na slici: npr. Stvara lik, stvara područje crtanja na slici, iscrtava crte u području crtanja, ukrašava ploču oznakama i slično. [34]

4.3.5 Seaborn

Seaborn je Pythonova biblioteka za vizualiziranje podataka temeljena na Matplotlibu, a pomaže korisniku razumjeti podatke. Glavna ideja Seaborna je da pruža naredbe visoke razine za stvaranje različitih vrsta grafikona korisnih za istraživanje statističkih podataka pa čak i za određeno usklađivanje statističkog model [35]. Seaborn i Matplotlib imaju slične funkcije, ali Seaborn ima dosta modernije i preglednije sučelje namijenjeno za crtanje grafikona. Seaborn pomaže istražiti podatke, te ih prikazati na razumljiviji način [36].

4.4 Odabir skupa podataka

Vizualizacija podataka u sportu se često koristi u razne svrhe. Neki od primjera su: prikaz statistike nakon utakmice ili meča, prikaz mjesta na koje neki igrač puca penal u nogometu, kružni površinski grafikon (engl. Circular area chart) koji prikazuje performanse nekog igrača. Ovakav se grafikon mogao vidjeti i u nekim video igrama što govori kako je dosta jednostavan za razumjeti čak i djeci i mladima.



Slika 9 Primjer kružnog površinskog grafikona [37]

Skup podataka (engl. Dataset) koji će se koristiti u ovom radu je vezan za povijest Olimpijskih igara, točnije za sportaše, njihove tjelesne karakteristike, sport kojim se bave, godinu te odličje. S obzirom na veliki broj natjecatelja i sportova na Olimpijskim igrama, ali i činjenicom da statistike vezano za sport više nego bilo koje drugo područje prate laici, dobra vizualizacija

ovog skupa podataka će simplificirati veliku količinu podataka te je učiniti njima razumljivima i jasnijima. Ovaj skup podataka s obzirom na svoja polja omogućava da se na zanimljiv način prikažu svi principi i sve vrste grafikona obuhvaćeni ovim radom.

Skup podataka je preuzet s platforme Kaggle koja sebe naziva najvećom zajednicom podatkovne znanosti, a na njoj su dostupni brojni skupovi podataka, bilježnice, modeli, tečajevi te prostor za komunikaciju korisnika.

5 PRAKTIČNI DIO

Praktični dio ovog rada će se sastojati od dvije cjeline. U prvoj će se testirati odabrani alati i biblioteke, a u drugom će biti provedeno istraživanje kojim će se testirati dobre prakse u vizualizaciji.

5.1 Korišteni alati

U ovom poglavlju će se uz pomoć spomenutih grafikona testirati alati Tableau i Power BI te Pythonove biblioteke za vizualizaciju podataka Matplotlib i Seaborn.

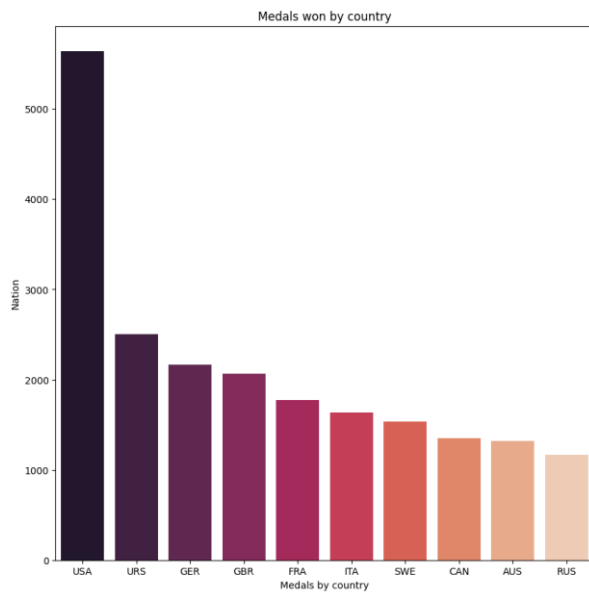
5.1.1 Stupčasti grafikon

Za prezentaciju ovog grafikona odabrano je prikazati broj osvojenih medalja po državi, a u grafikon je uključeno najboljih deset.

Osim što će se grafikon prikazati monokromatski, zbog istraživanja o orijentaciji stupčastog grafikona će primjer dobre prakse prikazan u ovom radu biti vertikalalan.

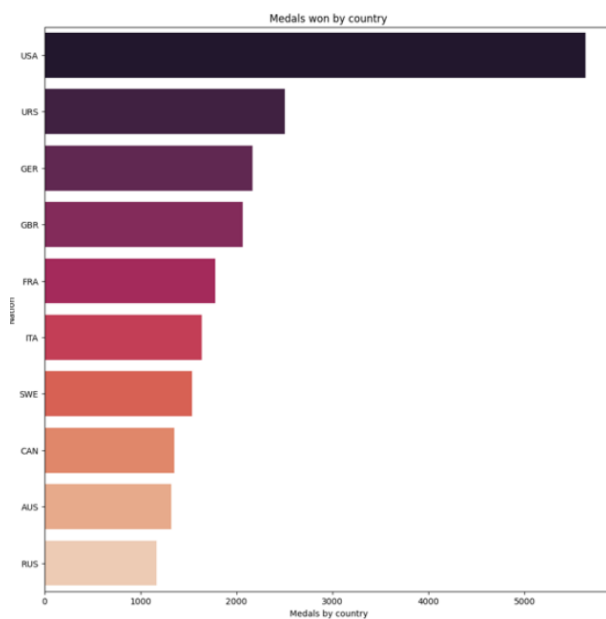
Python

Za obradu i pripremu podataka u Pythonu se koristila biblioteka Pandas, dok su za prikaz korištene biblioteke Seaborn i Matplotlib. Iako ima velik izbor paleta kojima se mogu bojati elementi grafikona, Python nema mogućnost prikaza palete te biranja konkretne boje iz te palete već se boje biraju „na slijepo“ ili pretraživanjem interneta.



Slika 10 Stupčasti grafikon sa zero-baselineom

Zamjenom podataka x i y osi dobiva se horizontalni grafikon.

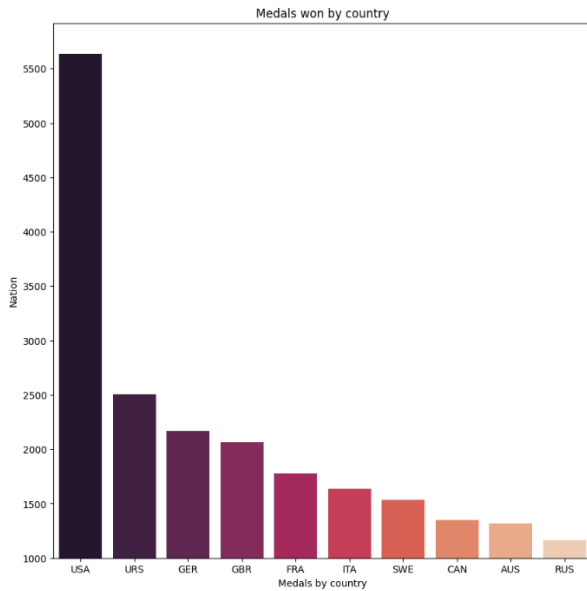


Slika 11 Horizontalni stupčasti grafikon

Baseline

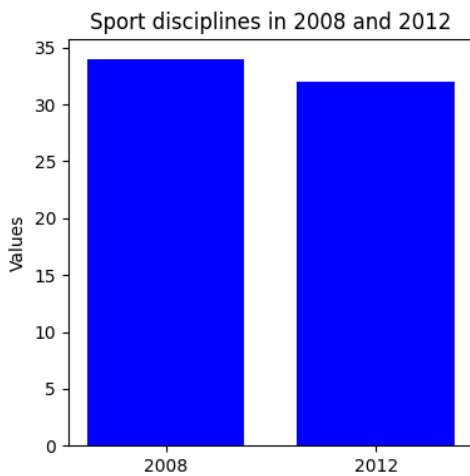
Kad na grafikonu treba prikazati slične vrijednosti, može se postaviti da vrijednost ne počinje od nule, već od neke veće vrijednosti kako bi se razlika bolje vidjela. To se može dobiti na način da je početna vrijednost na y-osi u ovom slučaju 1000, umjesto 0 kao u gornjem primjeru.

Vrijednost u kojoj počinju stupci se zove osnovna linija (engl. baseline) te je ona u pravilu nula (engl. Zero-baseline), ali je moguće prilagoditi na bilo koju drugu vrijednost sljedećim kodom: `plt.ylim(1000)`. Na ovaj način je vrijednost s defaultne 0 postavljena na 1000.

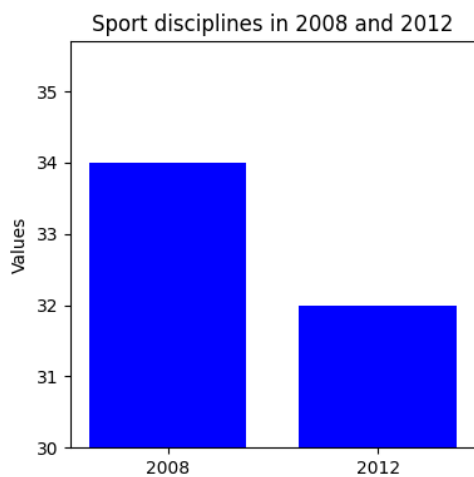


Slika 12 Stupčasti grafikon – baseline = 1000

Primjer će se prikazati na broju disciplina po godini, u ovom slučaju će se uspoređivati 2008. i 2012. godina. 2008. godine su se sportaši natjecali u 34 discipline, a 2012. u dvije manje, odnosno 32. Prvi grafikon prikazuje dva dosta slična stupca, što je i prikladno s obzirom da se radi o maloj razlici. No, postavljanjem minimuma y osi na 30, umjesto defaultnih 0, situacija se mijenja i prvi stupac je duplo veći od drugog, što ne prikazuje realno stanje za ovaj primjer.



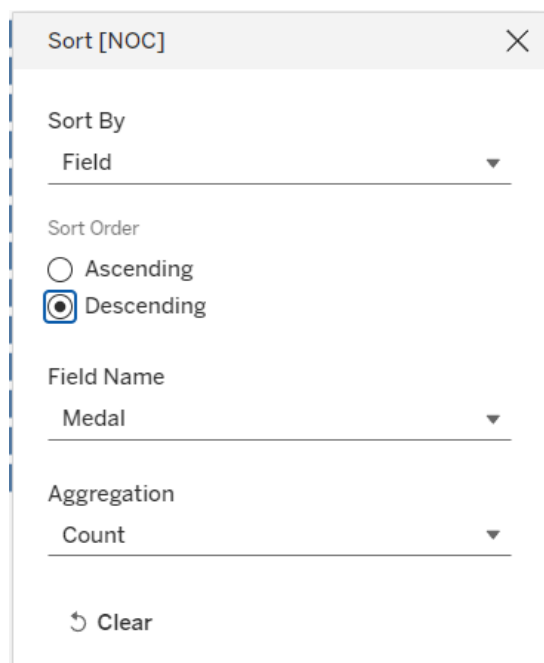
Slika 13 Broj disciplina, baseline = 0



Slika 14 Broj disciplina, baseline = 30

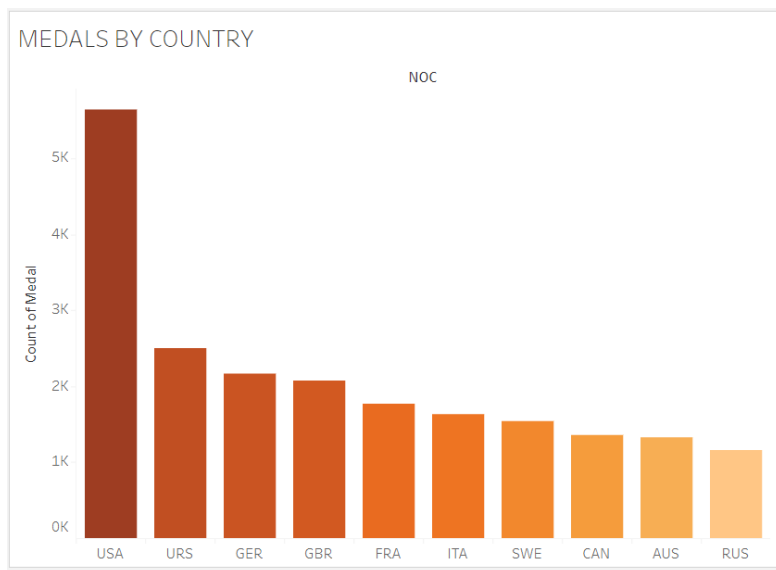
Tableau

Stupčasti grafikon je u Tableau jedan od najjednostavnijih prikaza. Ključni detalj pri izradi je sortiranje grafikona silazno (engl. Descending) te filtriranje da se pokaže samo prvih deset vrijednosti.



Slika 15 Sortiranje vrijednosti u Tableuu

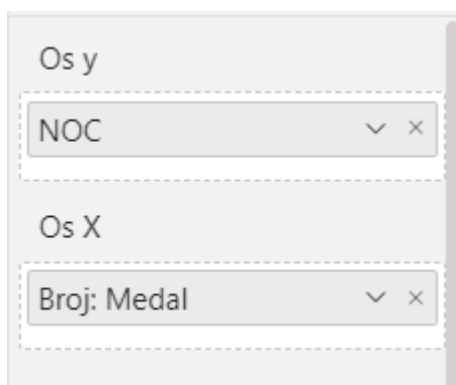
Mana Tableaui, barem za ovaj primjer, je ta što ne postoji ugrađena naredba za gradijent, već svakom od stupaca treba boju dodjeljivati ručno.



Slika 16 Stupčasti grafikon u Tableuu

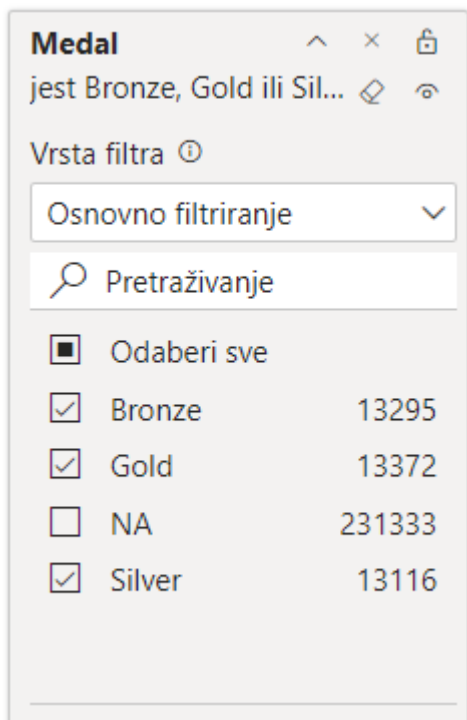
Power BI

Nakon učitavanja datasea te stvaranja nove stranice, na os X je dodan stupac 'Medal', a na os Y stupac NOC, odnosno troslovna kratica imena države.



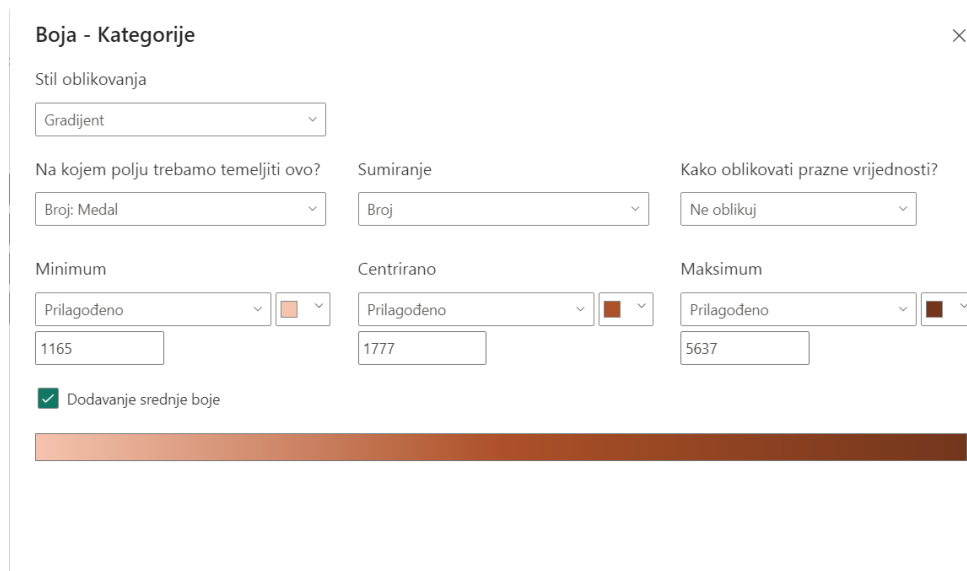
Slika 17 Postavljanje x i y osi

Korištenjem filtera je na jednostavan način odabrano da se prikazuju sve osvojene medalje. Slično bi se, micanjem oznake sa srebra i bronce brzo mogao dobiti grafikon koji prikazuje države s najviše osvojenih zlata.



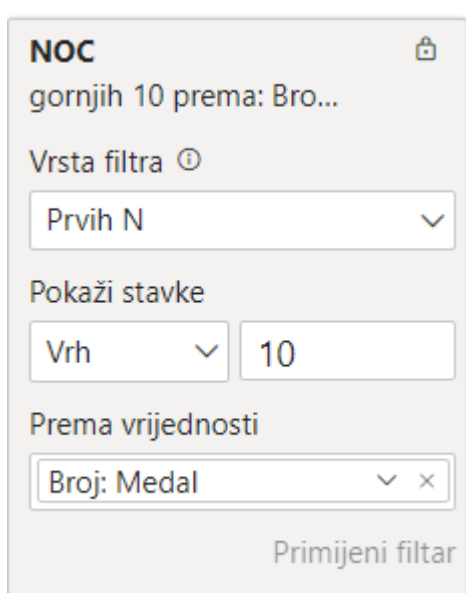
Slika 18 Odabir kategorija

U postavkama se mogu odabrati načini bojanja stupaca pa je tako ovdje odabran gradijent.

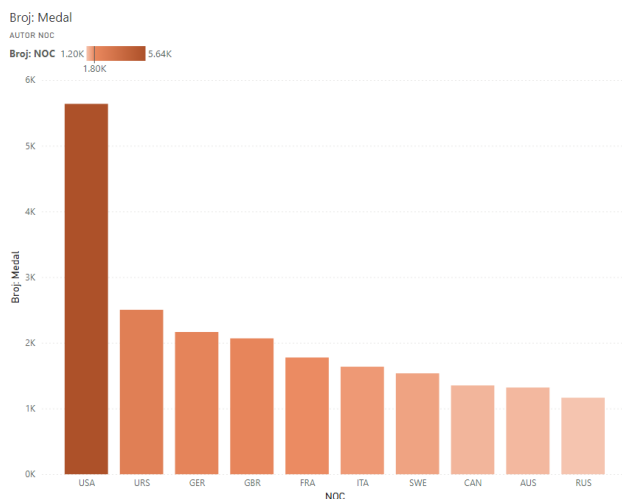


Slika 19 Postavljanje boje

Filtriranje prvih 10 država po broju medalja



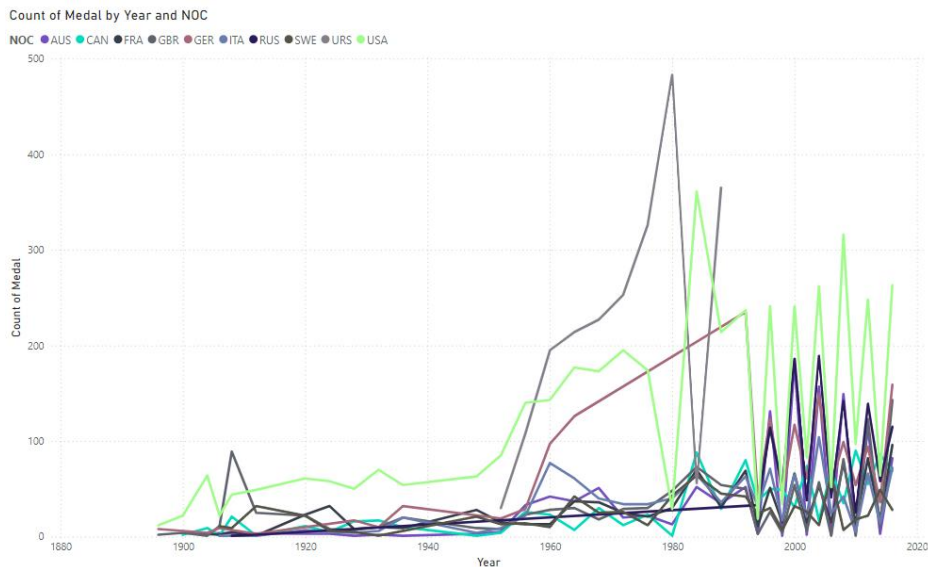
Slika 20 Filtriranje u Power BI-ju



Slika 21 Konačan prikaz stupčastog grafikona

5.1.2 Mali višekratnici – linijski grafikon

Mali višekratnici (engl. Small multiples) je način prikazivanja grafikona koji koriste iste osi kroz više sličica zbog urednijeg i jasnijeg prikaza te jednostavne usporedbe. Ovakav prikaz posebno važnu ulogu igra kod uspoređivanja puno vrijednosti jer prikazivanjem tolikog broja linija na jednom grafikonu vizualizacija postala nepregledna i samim time beskorisna kao na primjeru niže koji prikazuje linijski grafikon s deset linija, u Power BI alatu s bojama koje je alat sam odredio za prikaz.

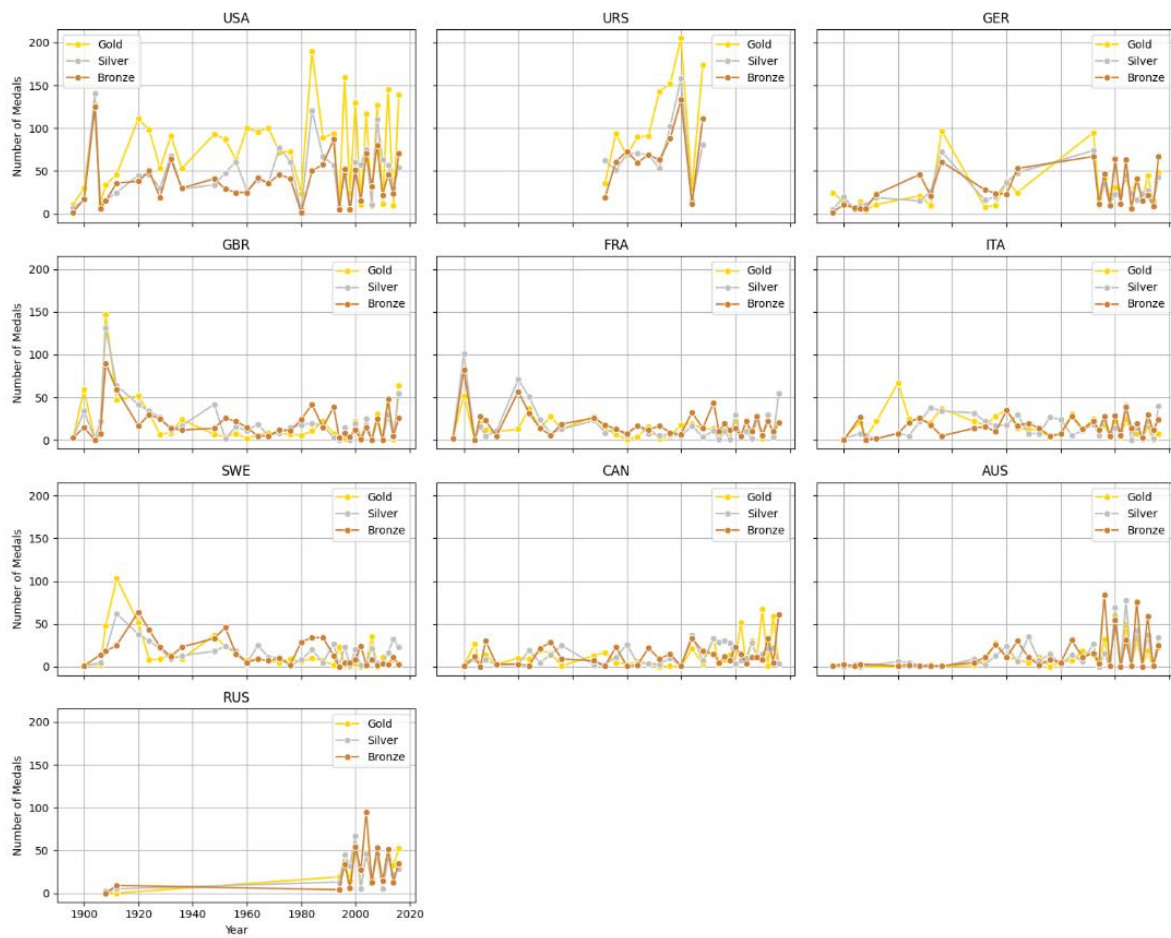


Slika 22 Linijski grafikon sa svih 10 prikaza

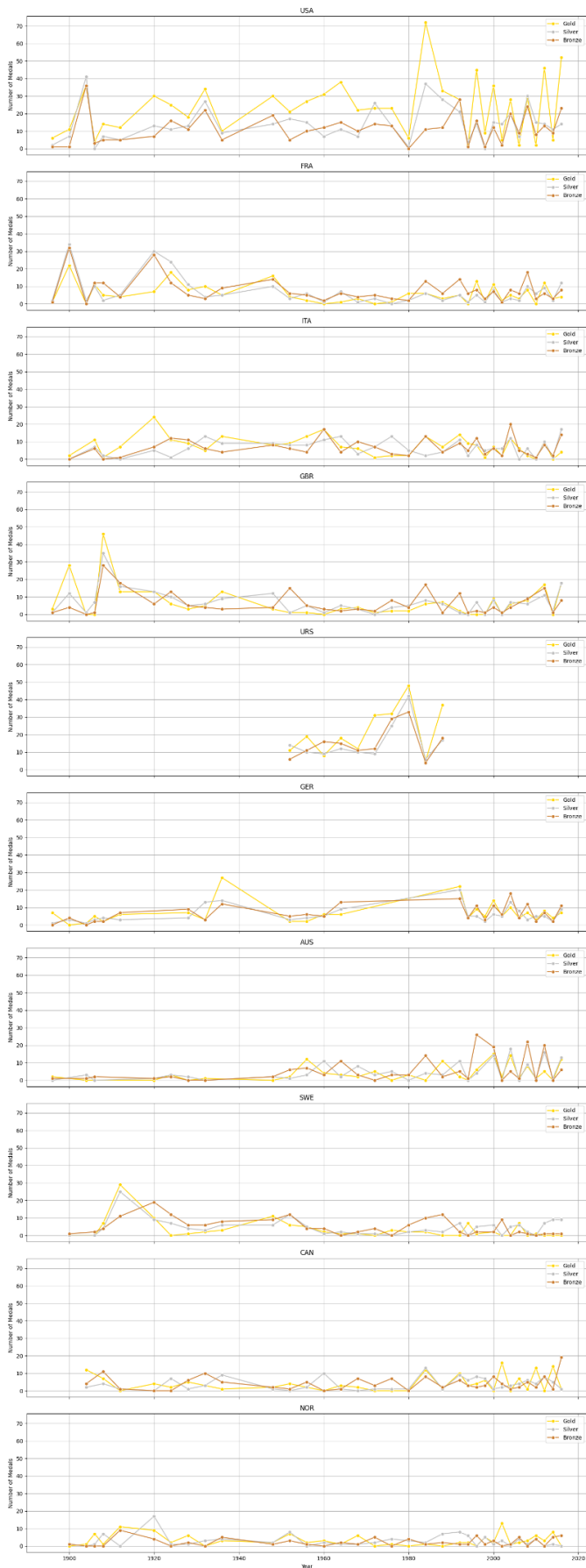
Ovakav primjer bi se mogao popraviti izradom malih višekratnika s deset sličica u kojem je na svakoj sličici bojom izdvojena po jedna država, a sivom su blago ocrtane sve ostale linije.

Python

U Pythonu postoje parametri kojima se može prikazati sličice u koliko god redova i stupaca je potrebno. To su parametri `nrows` kojim se određuje broj redaka, `ncols` kojim se određuje broj stupaca te `figsize` koji određuje veličinu okvira u kojem su grafikoni prikazani. Kod gornjeg prikaza je broj redaka četiri, a broj stupaca tri, dok je kod donjeg prikaza broj redaka deset, a stupaca jedan. Ako je bitno preko grafikona izvući što precizniju vrijednost, u ovom slučaju broj medalja, pristup s jednim stupcem i deset redaka se čini primjerenijim jer je širi te ga je lakše čitati, dok je za sticanje dojma kako su se otprilike podaci kretali kroz vrijeme sažeti prikaz, u ovom slučaju od tri stupca i četiri retka prikladniji jer su sličice veće i dovoljne da daju neke općenite informacije.



Slika 23 Mali višekratnici u Pythonu, prikaz 3 x 4

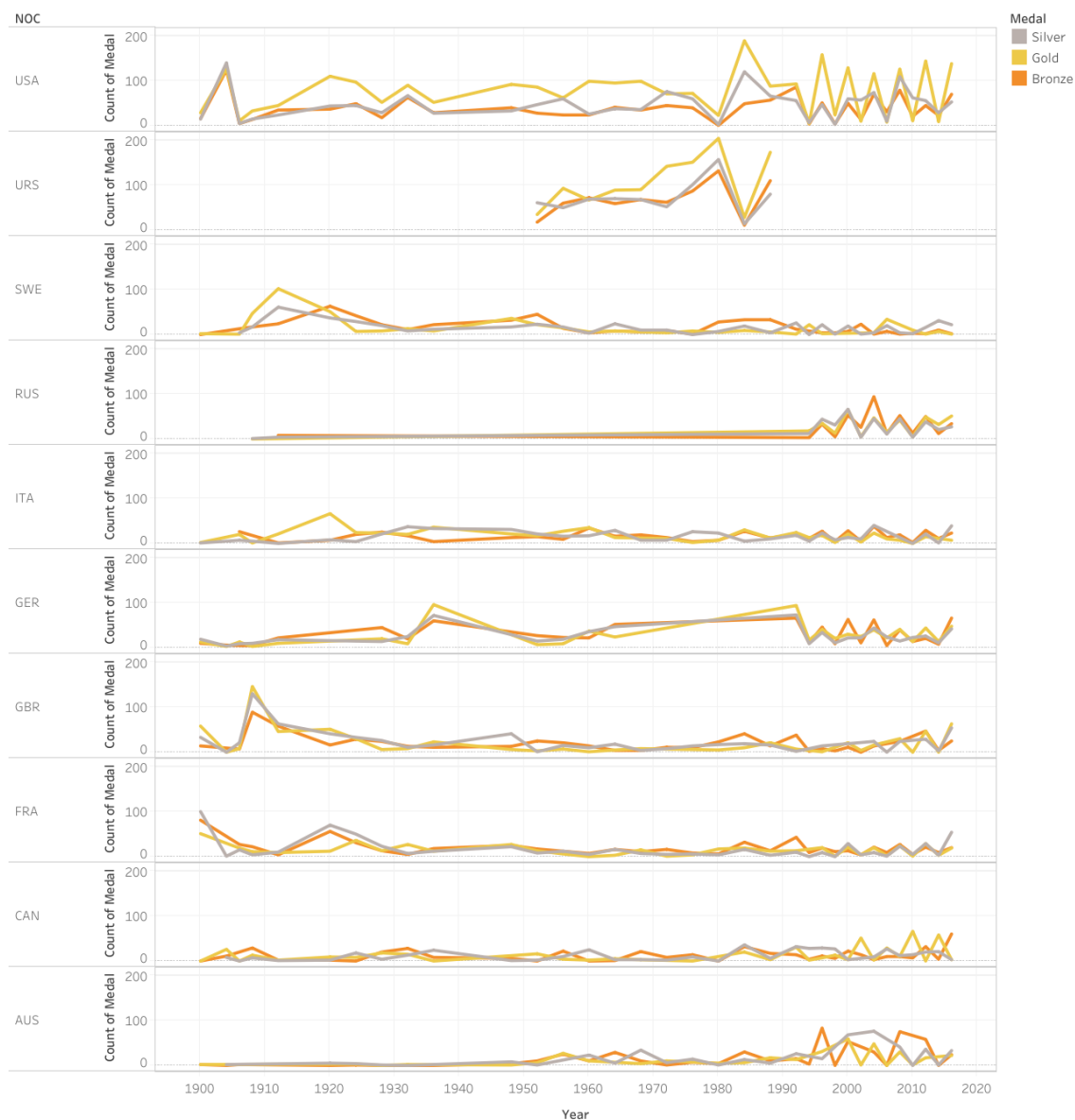


Slika 24 Mali višekratnici u Pythonu, prikaz 1 x 10

Tableau

Tableau po defaultu ima prikaz s deset redaka i jednim stupcem, ali je moguće postaviti prikaz onako kako to korisniku odgovara.

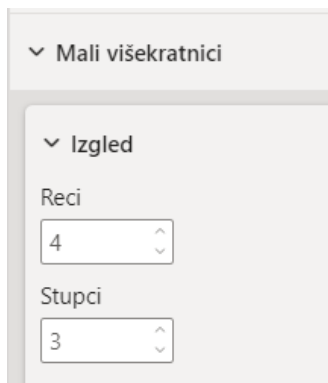
Sheet 2



Slika 25 Mali višekratnici u Tableuu, prikaz 1 x 10

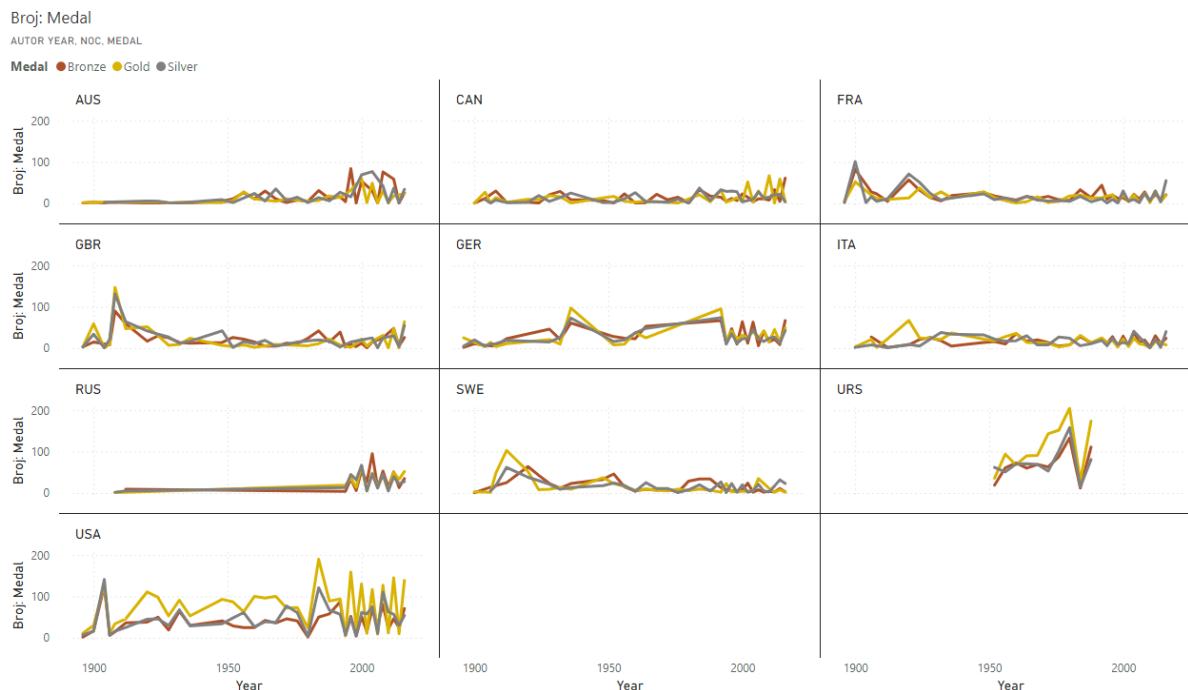
Power BI

U Power BI izgled malih višekratnika biramo preko sučelja sa slike. Maksimalan broj grafikona koji se u ovom alatu može prikazati je 36 (6x6).



Slika 26 Izbornik za postavljanje dimenzija

Grafikoni su interaktivni pa tako povlačenjem miša (engl. Hover) preko bilo kojeg dijela linijskog grafikona možemo dobiti točan podatak o broju medalja i godini.



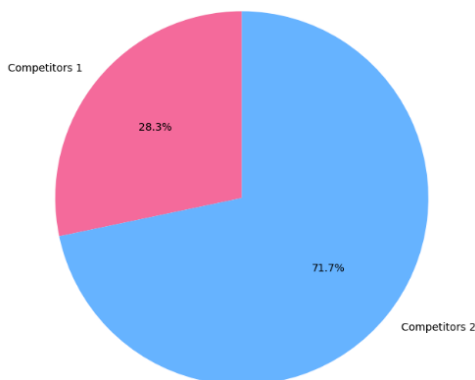
Slika 27 Mali višekratnici u Power BI, prikaz 3 x 4

5.1.3 Kružni grafikon

Kružni grafikon (engl. Pie chart) je vrsta grafikona oblika kruga s isječcima koji predstavljaju dio jedne cjeline. Korištenje boja može pomoći gledatelju da se lakše kreće kroz podatke. [12]

Na primjeru niže je prikazan kružni grafikon (engl. Pie chart) koji prikazuje udio muškaraca i žena u ukupnom broju natjecatelja na Olimpijskim igrama. Napomena kako prikazivanje grafikona bez opisa i oslanjanje na boje ili bilo koje drugo svojstvo nije dobra praksa, ali je

uvršteno u ovo istraživanje kako bi se prikazala važnost dobrog izbora boja te testiralo ispitanike u svrhu potvrde važnosti boja.



Slika 28 Kružni grafikon u Pythonu

5.1.4 Raspršeni ili linijski grafikon?

Raspršeni grafikon (engl. Scatter plot) je grafikon kojim se svaka vrijednost prikazuje kao jedna točka. Ovaj grafikon se koristi za prikazivanje odnosa između numeričkih vrijednosti. S obzirom da se radi o jako malim oznakama prikaza, odnosno točkama, važno je voditi računa da podaci ne prekrivaju jedan drugog. U ovom primjeru je prikazan broj sportskih disciplina po natjecanju kroz godine. Iz ove vizualizacije se može zaključiti kako se u pravilu sportaši na zimskim olimpijskim igrama natječu u manje disciplina od onih koji se natječu na ljetnima.

Linijski grafikon povezuje niz podatkovnih točaka pomoću linije. Predstavlja sekvencijalne vrijednosti koje pomažu u prepoznavanju trendova. Grafikon je najčešće korišten za prikazivanje podataka tijekom vremena. [38]

Iako postoje jasne razlike između raspršenog i linijskog grafikona, često se događa kako se ta dva grafikona miješaju te pogrešno koriste.

Python

U ovom poglavlju će se po uzoru na istraživanje [22] primijeniti sličan pristup uz pomoć funkcija LOWESS (locally weighted scatterplot smoothing) te KDE (Kernel density

estimation), konkretno u Pythonu, a rezultati će biti uspoređeni s rezultatima ankete. Cilj ovog istraživanja je pronaći grafikon koji bolje prikazuje skup podataka koji je vezan za ovaj rad, ne općenito.

Lowess

Lowess je jedna od regresijskih tehnika, odnosno funkcija zaglađivanja koja pokušava shvatiti uzorke među podacima te smanjiti šum (engl. Noise) i čini minimalne pretpostavke o odnosima među varijablama. S obzirom da je LOWESS funkcija kojom se određuje trend podataka, kvaliteta automatske selekcije grafikona ovisi o njoj. Faktor zaglađivanja (engl. Smoothing factor) je jedan od opcionalnih parametara funkcije te će on u ovom radu biti 0.4, kao što je sugerirano u radu po kojem se istraživanje radi. On je u Pythonu izražen varijablom `frac` te označava stupanj zaglađivanja. Ako je on manji, krivulja više prati podatke, a ako je veći, cilj joj je da bude što glađa. Ova je funkcija također neparametarska (engl. Non-parametric), što znači da se ne pretpostavlja da funkcija prati neku određenu formu, kao na primjer da je u linearnoj regresiji odnos među varijablama linearan ili da su kod Gaussovog modela varijable normalno distribuirane, već se oblik veze između podataka kroji isključivo prema samim podacima.

Također bitan parametar je r^2 (engl. R-squared) koji je koeficijent određenosti (engl. Coefficient of determination) koji u ovom kontekstu govori koliko dobro zaglađena krivulja odgovara originalnim podacima, odnosno koliko dobro podaci odgovaraju regresijskoj liniji. Raspon ove varijable je od 0 do 1 jer predstavlja odnos između dviju varijabli. U slučaju da je koeficijent određenosti 1, podaci bi savršeno bili poslagani na liniju, a u slučaju da je 0, ne bi bilo podataka na liniji [39]. Granica (engl. Threshold) je postavljena na 0.5, što znači da će za manje vrijednosti sugerirati raspršeni dijagram, a za veće linijski. Posljednji bitan parametar je `regular` kojim se određuje jesu li vrijednosti varijable x ravnomjerno raspršene.

KDE

Uz LOWESS, za određivanje grafikona će se koristiti i KDE funkcija. KDE je također neparametarska funkcija koja nam prikazuje procijenjenu distribuciju. Ona stvara glatku krivulju koja prikazuje gustoću podataka na temelju koje se određuje koji će se grafikon koristiti. Bitan parametar kod KDE je propusnost (engl. Bandwidth) koji regulira glatkoću krivulje gustoće podataka. Odnosi se na širinu funkcije o čemu ovisi koliko je raspršen svaki

individualni podatak, a računa se uz pomoć Silvermanovog pravila (engl. Silverman's rule of thumb) koje je prikazano na slici.

$$h_y = 1.06 \times \min\left\{\hat{\sigma}_y, \frac{R}{1.34}\right\} n^{-1/5},$$

Slika 29 Silvermanovo pravilo

σ_y predstavlja standardnu devijaciju, R je interkvartil, a n broj podataka (engl. Data points) koji se prikazuju.

`kde_has_peaks`, kao što joj samo ime govori je varijabla tipa `bool` koja govori ima li KDE krivulja vrh (engl. Peak) u kojem je x veći od 0.1. Ako ima, to sugerira da gustoća podataka varira.

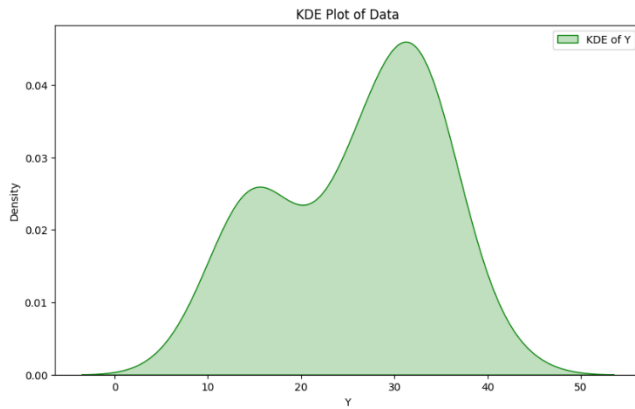
Odabir

Uzevši u obzir ranije navedene parametre, algoritam određuje je li bolji prikaz koristeći raspršeni ili linijski grafikon s obzirom na dane podatke. Uvjeti koje, dakle, treba ispuniti da bi odluka pala na linijski grafikon su da `regular` istina (engl. True), da je r^2 veći od granice (engl. Threshold) te da je `kde_has_peak` laž (engl. False). Ako je bilo koja od ovih varijabli suprotna od naznačenog, algoritam će odabrati raspršeni grafikon.

```
if regular and r_squared > r2_threshold and not kde_has_peaks:
    plt.plot(x_smooth, y_smooth, label=f'LOESS Line Plot (R²={r_squared:.2f})')
    print("Algorithm chose a LOESS smoothed line chart.")
else:
    plt.scatter(x, y, label=f'Scatter Plot (R²={r_squared:.2f})', color='red')
    print("Algorithm chose a scatter plot.")
```

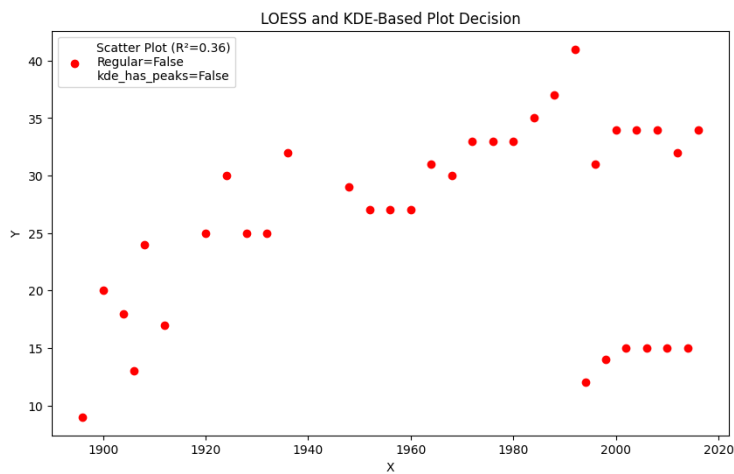
Slika 30 Kod za odabir grafikona

Varijabla `regular` je False, r^2 iznosi 0.36, a `kde_has_peaks` je laž jer nijedna vrijednost ne premašuje 0.1, Zbog prva dva kriterija je algoritam odlučio da je za ovaj primjer primjerenije koristiti raspršeni grafikon.



Slika 31 KDE krivulja

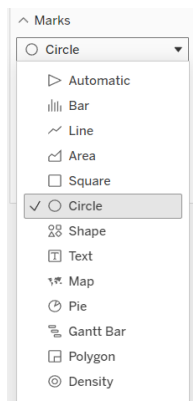
Krajnji rezultat je prikazani grafikon zajedno s varijablama koje su bile ključne za donošenje odluke.



Slika 32 Raspršeni dijagram

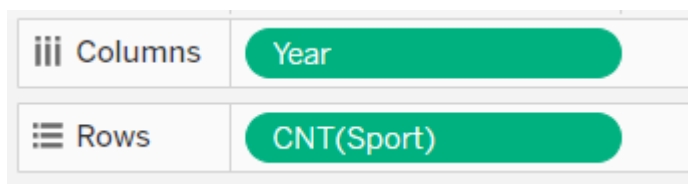
Tableau

Na slici se vidi izbornik oblika kojima se mogu prikazati podaci koji se mogu koristiti u Tableau, a kombiniranjem polja može se dobiti razne grafikone. Za raspršeni grafikon je odabrana oznaka (engl. Mark) „circle“.



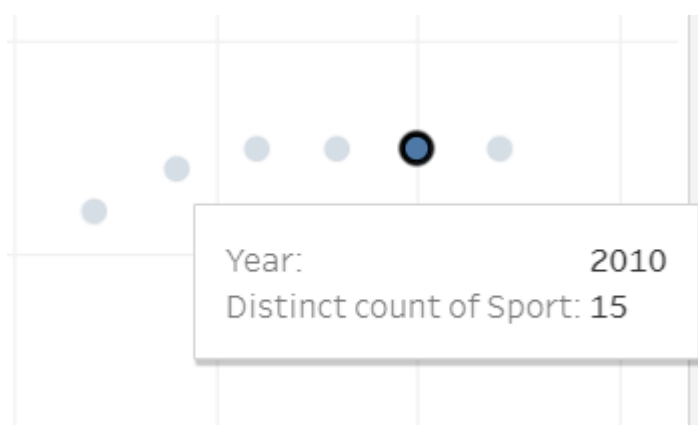
Slika 33 Odabir oznake

Što se podataka tiče, na x osi se nalaze godine što je uobičajeno kod ovakvih grafikona, dok je na y osi prikazan broj sportskih disciplina. Svaki kružić reprezentira godinu u kojoj su se održale Olimpijske igre te broj disciplina koje su se te godine održale.



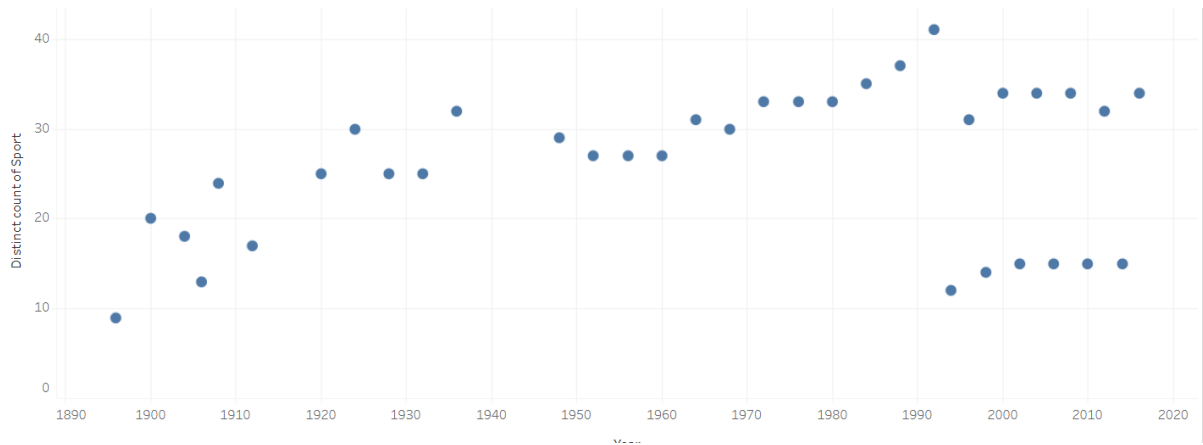
Slika 34 Postavljanje osi x i y

Pređe li se mišem (engl. Hover) preko kružića, vidjet će se točan broj disciplina za određenu godinu.



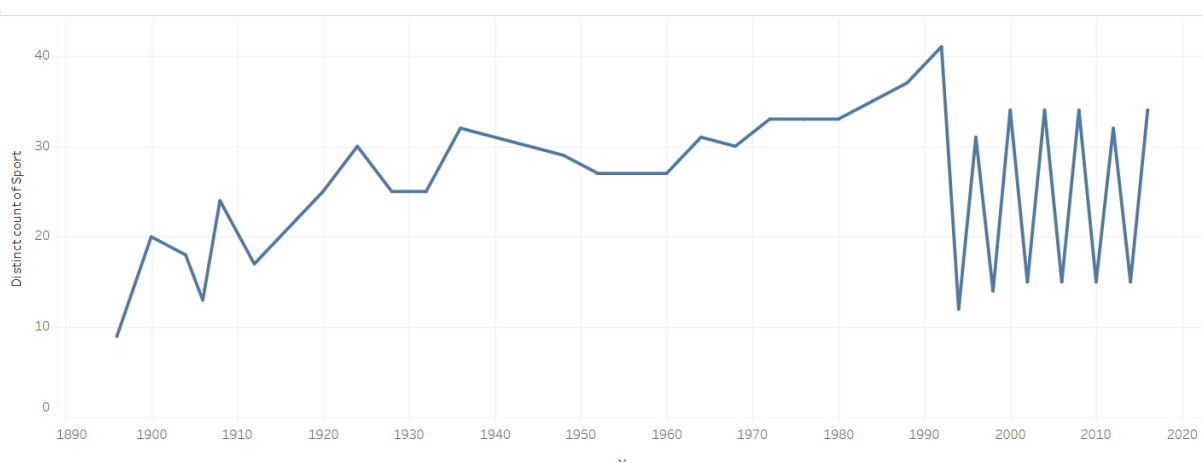
Slika 35 Prikaz individualnog podataka u raspršenom grafikonu

Važno je napomenuti kako je bilo potrebno uključiti razlikovno brojanje (engl. Distinct) jer bi u suprotnom bili uračunati svi sportaši koji su se te godine natjecali umjesto samih disciplina.



Slika 36 Raspršeni grafikon u Tableauu

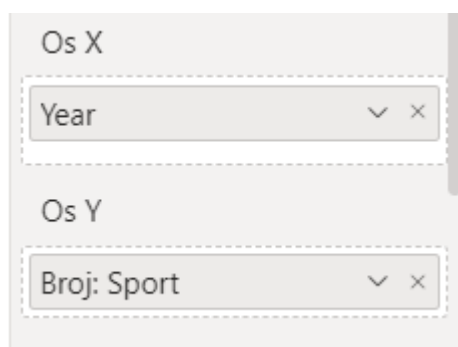
Velika prednost alata koji su obrađeni u ovom radu je ta da je moguće jednim klikom promijeniti oblik kojim prikazujemo grafikone, konkretno u ovom primjeru iz točki u linije te vrlo brzo dobiti linijski grafikon.



Slika 37 Linijski grafikon u Tableauu

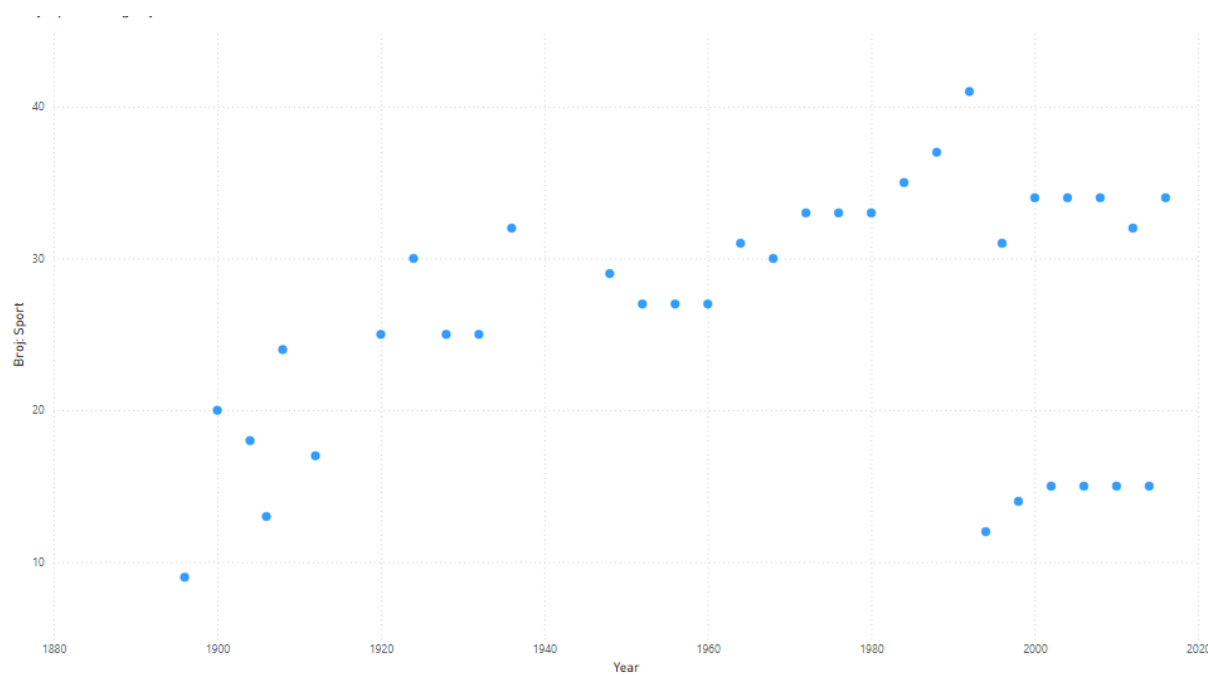
Power BI

Izrada scatter plota u Power BI je također dosta jednostavna, po principu „drag and drop“.



Slika 38 Postavljanje osi x i y za izradu raspršenog dijagrama

Kao i Tableau ima po defaultu opciju kojom se prelaskom preko oznake na grafikonu dobiju dodatne informacije o godini i broju disciplina za razliku od Pythona.



Slika 39 Raspršeni grafikon u Power BI

5.1.5 Složeni stupčasti grafikon

Složeni stupčasti grafikon (engl. Stacked bar chart) je u suštini stupčasti grafikon (eng. bar chart) koji daje jednu dodatnu informaciju. U kontekstu skupa podataka koji je korišten u ovom radu prikazuje se deset država s najvećim brojem osvojenih medalja (ukupno) kroz godine, a ono što čini razliku između običnog i složenog dijagrama je ta što se na ovaj način dodatno može prikazati udio svake vrste osvojene medalje (zlato, srebro, bronca) i to koristeći prikladne boje.

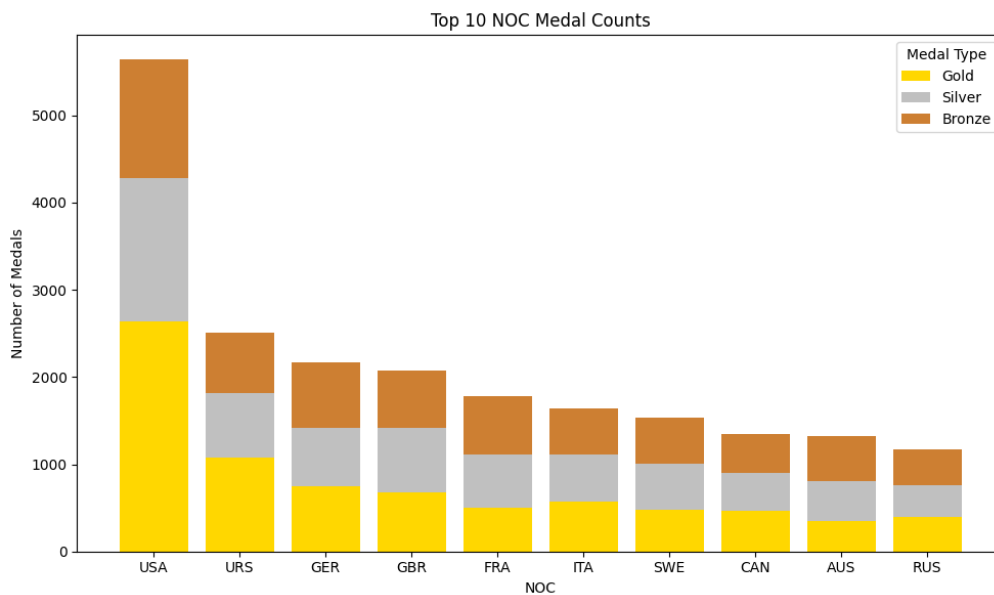
Python

Za dobivanje složenog stupčastog grafikona korišten je sljedeći kod:

```
top_10 = df.groupby('NOC')['Medal'].count().nlargest(10).reset_index(name='Medals')
medal_counts = df.groupby(['NOC', 'Medal']).size().unstack(fill_value=0)
top_10_nocs = medal_counts.sum(axis=1).nlargest(10).index
top_10_medals = medal_counts.loc[top_10_nocs]
fig, ax = plt.subplots(figsize=(10, 6))
ax.bar(top_10_medals.index, top_10_medals['Gold'], label='Gold', color='gold')
ax.bar(top_10_medals.index, top_10_medals['Silver'], bottom=top_10_medals['Gold'], label='Silver', color='silver')
ax.bar(top_10_medals.index, top_10_medals['Bronze'], bottom=top_10_medals['Gold'] + top_10_medals['Silver'], label='Bronze', color='#cd7f32')
ax.set_xlabel('NOC')
ax.set_ylabel('Number of Medals')
ax.set_title('Top 10 NOC Medal Counts')
ax.legend(title='Medal Type')
plt.show()
```

Slika 40 Kod za izradu složenog stupčastog grafikona u Pythonu

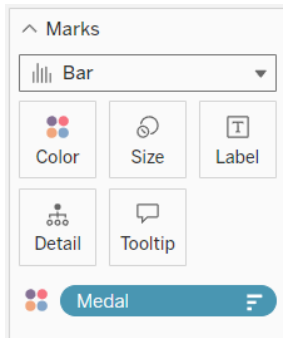
U varijablu `top_10` spremljen je okvir podataka (engl. Dataframe) u kojem se nalazi deset država s najvećim brojem osvojenih medalja te točan broj medalja. `Medal_counts` je po sličnom principu okvir u kojem su medalje dodatno podijeljene na zlatne, srebrene i bronzane, a `top_10_medals` je iste strukture, ali su filtrirane države pa tako sadrži samo deset najbolje rangiranih. Sljedeći dio koda je dio Matplotlib biblioteke i odnosi se na crtanje grafikona. Ideja je napraviti tri grafikona (svaki za svoje odličje) te ih složiti jedan na drugi i svakome dodijeliti pripadajuću boju. Krajnji rezultat je prikazan uz pomoć iste biblioteke uz pripadajuću legendu te oznake (engl. Labels) na osima.



Slika 41 Složeni stupčasti grafikon u Pythonu

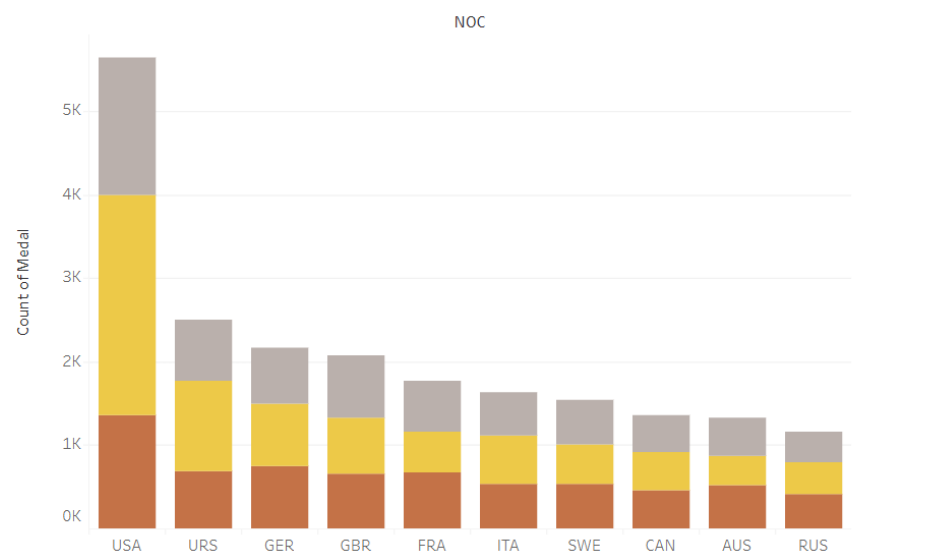
Tableau

Pretvaranje običnog stupčastog grafikona u složeni u Tableau nije baš intuitivno. Logika je ista kao i u Power BI, ali ovdje ne postoji legenda, već se dodatno polje 'medal' dodaje u oznake (engl. Marks).



Slika 42 Dodavanje parametra kako bi stupčasti grafikon postao složen u Tableauu

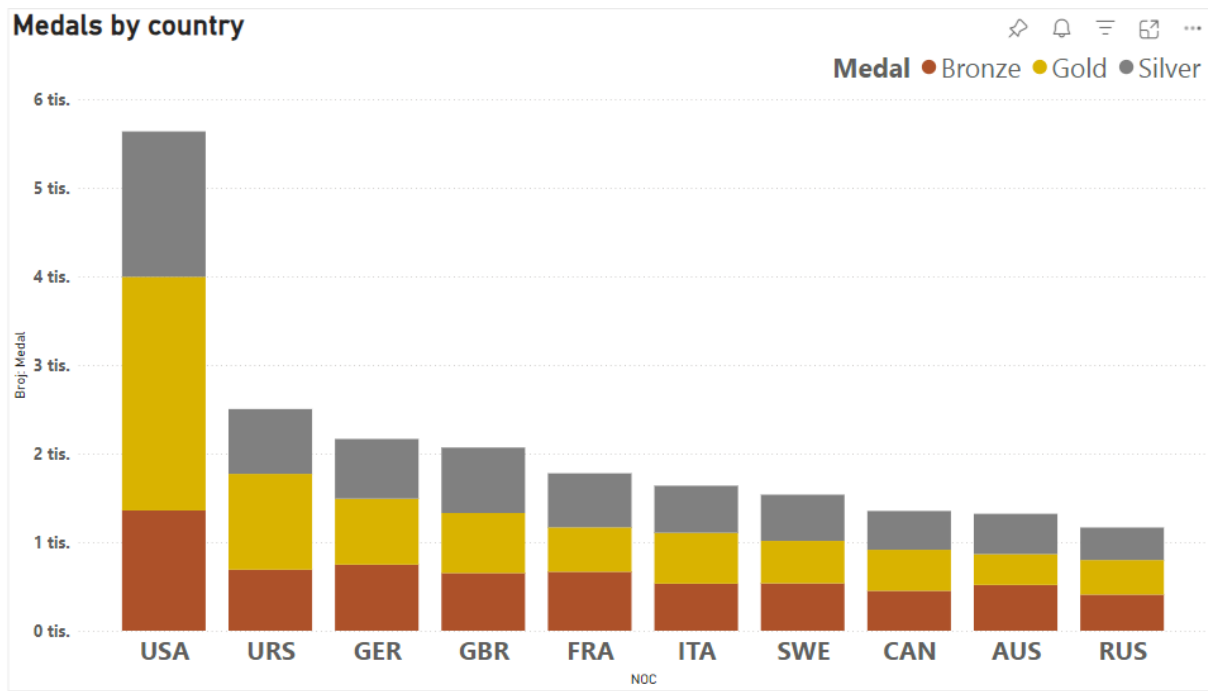
Tableau ima velik broj paleta te se može odabrati točno ona boja koja je potrebna za savršeno prikazati podatke.



Slika 43 Složeni stupčasti grafikon u Tableauu

Power BI

Power BI također nema ugrađenu opciju složenog stupčastog dijagrama, ali dodavanjem polja 'Medal' u legendu te odabiranjem vrijednosti 'gold', 'silver' i 'bronze', odnosno izbacivanjem vrijednosti 'None' dobije se grafikon sa slike.



Slika 44 Složeni stupčasti grafikon u Power BI

Zadatak konverzije

U sklopu ovog rada postavljen je zadatak da se u svim korištenim alatima i bibliotekama stupac sa kilogramima konvertira u funte (engl. Pounds). Dodatan izazov je što je po defaultu tip podataka tekstualni (engl. String), a ne decimalni broj (engl. float) zbog čega je prvo bilo potrebno prilagoditi tip podataka zadatku pa ga tek onda izvršiti.

5.2 Upitnik

Istraživanje među ispitanicima je provedeno uz pomoć upitnika koji se sastoji od ukupno 15 pitanja općenitog tipa, poznavanja alata te rada na računalu općenito te pitanja koja testiraju dobre prakse korištene u ovom radu.

Pitanje 8 testira istraživanje [16] koje se tiče orijentacije stupčastog grafikona. Ispitanicima su ponuđena dva grafikona (Slika 10 Stupčasti grafikon sa zero-baselineom, Slika 11 Horizontalni stupčasti grafikon) koja prikazuju iste podatke te se od njih traži da odaberu onaj za kojeg smatraju da preglednije i jasnije prikazuje države po broju osvojenih medalja.

Pitanje 9 koristi isti grafikone te iste podatke, ali je cilj prikazati kako iako na prvi pogled izgleda jasnije i bolje koristiti neku veću vrijednost y osi, to može dovesti do pogrešne interpretacije. Ispitanici će u ovom pitanju birati između dva grafikona (Slika 10 Stupčasti grafikon sa zero-baselineom, Slika 12 Stupčasti grafikon – baseline = 1000) od kojih je jednom vrijednost y 0, a drugom 1000 te im je zadatak procijeniti koji od grafikona jasnije prikazuje vrijednosti. [8]

Pitanje 10 po istom principu prikazuje druge podatke. Sadrži stupčasti grafikon (Slika 14 Broj disciplina, baseline = 30) koji prikazuje broj sportskih disciplina u 2008. i 2012. godini. Svrha ovog pitanja je potvrditi istraživanje [18] na način da su ispitanicima vrijednosti grafikona vizualno prikazane kao 50% veće, a u stvarnosti, čitanjem y osi koja počinje od 30 je jasno da je razlika otprilike 6%.

Cilj pitanja 11 je prikazati težinu čitanja linijskog grafikona koji ima previše linija (Slika 22 Linijski grafikon sa svih 10 prikaza). Od ispitanika se traži da iščitaju točnu vrijednost za jedan primjer, preciznije za Veliku Britaniju 1980. godine. To je primjer loše prakse koji bi se trebao prikazati kroz neprecizne rezultate istraživanja.

12. pitanje ispitanicima prikazuje prijedlog bolje prakse u vidu prikaza grafikona preko malih višekratnika koji u jednom vizualnom prikazu kreiraju 10 malih grafikona za koje ispitanici moraju odabrati smatraju li ih boljim prikazom od onoga iz pitanja 11 (Slika 23 Mali višekratnici u Pythonu, prikaz 3 x 4).

U 13. se pitanju od ispitanika traži da zaključe o kakvoj se podjeli natjecatelja radi isključivo na temelju boja isječaka u kružnom dijagramu (Slika 28 Kružni grafikon u Pythonu). Cilj je potvrditi istraživanja [15] [14] [12] koja govore kako se boje dijele po spolovima i kako bi se kod izrade vizualizacije trebalo obratiti pažnju na izbor boja.

Slika u 14. pitanju prikazuje usporedbu linijskog i raspršenog grafikona kroz podatke o broju sportskih disciplina kroz godine (Slika 36 Raspršeni grafikon u Tableauu, Slika 37 Linijski grafikon u Tableauu). Svrha je provjeriti hoće li se rezultati ispitivanja poklopiti sa rezultatima koji su dobiveni izradom algoritma u Pythonu na temelju istraživanja [22].

Posljednje 15. pitanje prikazuje dva stupčasta grafikona od kojih jedan prikazuje broj osvojenih medalja po državama, a drugi je složeni pa uz spomenuto prikazuje još i raspodjelu osvojenih medalja (zlato, srebro, bronca) (Slika 41 Složeni stupčasti grafikon u Pythonu, Slika 21 Konačan prikaz stupčastog grafikona). Cilj je provjeriti hoće li ispitanici prepoznati dodatne informacije koje čine razliku između običnog i složenog stupčastog grafikona.

6 REZULTATI ISTRAŽIVANJA

U ovom poglavlju će biti opisani rezultati rada koji su temeljeni na osobnom dojmu korištenja i izrade vizualizacije u alatima Tableau i Power BI te Pythonovim bibliotekama kao i istraživanje provedeno u obliku upitnika.

6.1 Rezultati usporedbe biblioteka i alata

Unatoč iskustvu u Pythonu, izrada vizualizacija je bila puno jednostavnija u Power BI i Tableau alatima koji su bili potpuno nepoznati i koji nikad nisu bili korišteni. Kod Pythona su priprema i oblikovanje podataka najteži dio, dok je sama izrada vizualizacije nešto jednostavnija te je sintaksa lako dostupna i jasno objašnjena na internetu. Nakon učitavanja i proučavanja skupa podataka (engl. Dataset) te osnovne pripreme, za izradu svake vizualizacije je bilo potrebno do sat vremena, dok je vrijeme izrade vizualizacija u ostala dva alata oko pola sata što je uključivalo i vrijeme upoznavanja s alatom, učitavanje podataka te pronalaženje svega potrebnog za izradu vizualizacije. Power BI i Tableau su otprilike podjednako dobri alati čije vizualizacije mogu stati uz bok Pythonovima uz veliku razliku što za korištenje njih nije potrebno znanje programiranja već malo naprednije znanje korištenja računala uz uvjet da je osoba upoznata s grafikonima u teoriji te osnovama vizualnog prikaza. Ipak, mala prednost ide Tableuu zbog jednostavnijeg i intuitivnijeg korisničkog sučelja iz razloga što su takvi alati namijenjeni osobama ograničenog znanja koji će se potencijalno teško snalaziti u dosta nabijenom sučelju i izbornicima koji su dio Power BI. Općenito se bilo lakše snaći i pronaći što je potrebno u Tableuu nego u Power BI te se smatra kako je bolji alat za početnike upravo iz tog razloga. Još jedan razlog zbog kojeg Tableau ima prednost nad Power BI je zadatak konverzije koji je opisan ranije i proveden kao dio istraživanja u sklopu ovog rada. Dok je izrada u Pythonu i Tableau bila trivijalna i trajala je ispod 10 minuta, dok u Power BI alatu rješavanje ovog zadatka nije ni najmanje intuitivno te zahtjeva dodatna znanja van snalaženja u samom alatu.

6.2 Rezultati upitnika

Upitnik je popunilo 50 ljudi, od čega je 50% muškaraca i 50% žena. Većina ih je u dobnoj skupini od 23 do 33 godine, točnije njih 84%. Svi su ocijenili svoje sposobnosti korištenja

računala ocjenom 3 i više, dok ih je 50% sadašnjih ili bivših studenata smjerova koji sadrže programerske kolegije, a 50% ne.

Spol	Broj ispitanika	Postotak
Muški	25	50%
Ženski	25	50%

Tablica 1 Raspodjela po spolu

Dobna skupina	Broj ispitanika	Postotak
18-23	3	6%
23-28	29	58%
28-33	13	26%
34+	5	10%

Tablica 2 Raspodjela po dobi

Ocjena poznavanja rada na računalu	Broj ispitanika	Postotak
1	0	0%
2	0	0%
3	7	14%
4	19	38%
5	24	48%

Tablica 3 Poznavanje rada na računalu

Ispitanik je bivši student smjera koji se bavi programiranjem.	Broj ispitanika	Postotak
Da	25	50%
Ne	25	50%

Tablica 4 Pohadanje programerskih kolegija

Što se poznavanja alata i biblioteka tiče, za Pythonove biblioteke je čulo čak 94% ispitanika, a 46% ih koristi ili ih je koristilo prije. Za Power BI 42% ispitanika nikad nije čulo, a isto toliko ih je čilo za njega, ali ga nikad nisu koristili. 12% ga je koristilo nekoliko puta, a tek 4% često. Slične rezultate ima i Tableau za kojeg 40% ispitanika nije čulo, 42% je čulo, ali nije koristilo, a samo 16% ga je u nekom trenutku koristilo.

U praktičnom dijelu ankete 82% ispitanika je odgovorilo kako preferiraju vertikalni nad horizontalnim stupčastim grafikonom što potvrđuje kako je korisnicima lakše čitati oznake napisane jednu uz drugu nego jednu ispod druge. 76% ispitanika je u pitanju 9 uspješno uočilo razliku među grafikonima koji imaju drugačiju početnu vrijednost y osi. Svi ispitanici koji su svoje računalne sposobnosti ocijenili ocjenom '3' su odabrali vertikalni grafikon nad horizontalnim te su uspješno uočili razliku između grafikona iz pitanja 9. S druge strane 35 ispitanika, odnosno 70% je uspješno riješilo zadatak iz pitanja 10, od čega ih je 20 sadašnjih ili bivših studenata smjerova vezanih za računala. S obzirom da se radi o trivijalnom izračunu, može se zaključiti kako 30% ispitanika nije primijetilo da vrijednost ne počinje od nule, već su samo zaključili da je vrijednost upola manja jer je i grafikon upola manji. U ovim su pitanjima ispitanici većinom gledali samo ono što se od njih traži; u 9. pitanju se tražila razlika na grafikonu, dok je u 10. bilo potrebno računati pa su ispitanici uz pomoć vrijednosti sa y osi izračunali o kolikoj se razlici radi.

U 11. pitanju nitko od ispitanika nije točno odgovorio na pitanje. 33 ispitanika (66%) je odgovorilo u rasponu od 400 do 500, što implicira da su pomiješali linije Velike Britanije i Sovjetskog Saveza jer su na grafikonu slične boje. Ovakvi rezultati potvrđuju kako na ovoliko velikom broju država nije pametno koristiti jedan linijski grafikon. Točan odgovor na ovo pitanje je bio 48. U 12. pitanju se 86% ispitanika složilo s tvrdnjom kako je individualni prikaz država korištenjem malih višekratnika bolja opcija od primjera iz prethodnog pitanja.

U 13. pitanju je 56% ispitanika prepoznalo da roza i plava boja predstavljaju podjelu po spolu, dok ih 34% smatra kako je jednaka vjerojatnost da se radi o bilo kojoj od kategorija. Iako potpuno oslanjanje na boje nije poželjno te je potrebno grafikon opisati, svakako je dobra praksa da korisniku prikaz bude intuitivan kroz boje. Za prikaz broja sportskih disciplina kroz godine putem linijskog grafikona se odlučilo 66% ispitanika, dok je 34% odabralo raspršeni. Odabrani algoritam je s blagom prednošću odabrao raspršeni grafikon. U posljednjem 15. pitanju je 96% ispitanika, a svi studenti smjera vezanog za računala, prepoznalo kako složeni stupčasti grafikon daje više informacija od običnog.

7 ZAKLJUČAK

U ovom istraživanju testirani su alati i biblioteke za vizualizaciju podataka te su kroz upitnik testirane dobre prakse pri izradi vizualizacije na skupu podataka vezanim za Olimpijske igre.

Zbog ograničenja u ovaj rad nisu uključene kvantitativne mjere kao što je praćenje pokreta očiju, mjerenja vremena potrebnog za rješavanje zadataka te mjerenja koliko je puta neka vizualizacija odabrana. Budući radovi bi trebali uključivati ove mjere jer bi upotpunile upitnik te pružile dodatna saznanja.

Korištenje svih alata i biblioteka obrađenih u ovom radu ima prednosti i mane te je najprimjereniji najbolje odabrati prema vlastitim sposobnostima i zahtjevima. Alati sadrže moderna sučelja te su idealni za početnike, dok je rad u Pythonu primjereniji za malo iskusnije korisnike. Dodatna prednost Pythona je integrirana umjetna inteligencija Gemini koja na temelju imena varijable ili prethodnog koda predlaže kod te znatno ubrzava proces pisanja koda te smanjuje mogućnost grešaka u pisanju. Osim toga, umjetna inteligencija pomaže i kod interpretacije grešaka u kodu te predlaže rješenja. To znatno pomaže kod uštede vremena te omogućuje da se korisnik fokusira na druge dijelove rada. Iako je za stvaranje vizualizacije i općenito rad u Pythonu potrebno znanje programiranja, integracija umjetne inteligencije svakako smanjuje potrebnu razinu znanja te izrade ovakvih vizualizacija čini malo dostupnijima i onima manje iskusnima. Primjena dobrih praksa je cijeli nevidljivi proces koji se nalazi iza jedne vizualizacije koja je predstavljena krajnjem korisniku. Iznimno je bitna te je temelj izrade kvalitetnih i korisnih vizualizacija koje su postale neophodne za interpretaciju podataka. Usporedbom rezultata uočeno je kako su ispitanici prepoznali dobre prakse te birali grafikone koji su napravljeni u skladu s njima. Prepoznali su ih čak i oni koji svoje poznavanje računala ocjenjuju kao loše te nisu programerske struke, što je pokazatelj da vizualizacija nije rezervirana samo za računalni svijet, već bi svi trebali biti sposobni razumjeti barem osnove.

Iz tog je razloga bitno razvijati područje te tražiti nove načine kako bi sve vizualizacije bile u duhu dobre prakse.

8 TABLICA SLIKA

Slika 1 Pretraga (prema [5]).....	5
Slika 2 Metode izrade vizualizacije (Prema [5]).....	6
Slika 3 HSB model (prema [12]).....	9
Slika 4 Tipovi vizualizacija [24]	11
Slika 5 PRISMA dijagram	14
Slika 6 Mjere vizualizacije (prema [29])	15
Slika 7 Primjer loše vizualizacije 3d stupčastog grafikona [25]	16
Slika 8 Popravljeni primjer stupčastog grafikona	16
Slika 9 Primjer kružnog površinskog grafikona [37]	19
Slika 10 Stupčasti grafikon sa zero-baselineom.....	22
Slika 11 Horizontalni stupčasti grafikon	22
Slika 12 Stupčasti grafikon – baseline = 1000.....	23
Slika 13 Broj disciplina, baseline = 0	23
Slika 14 Broj disciplina, baseline = 30	24
Slika 15 Sortiranje vrijednosti u Tableauu	24
Slika 16 Stupčasti grafikon u Tableuu.....	25
Slika 17 Postavljanje x i y osi	25
Slika 18 Odabir kategorija	26
Slika 19 Postavljanje boje.....	26
Slika 20 Filtriranje u Power BI-ju	27
Slika 21 Konačan prikaz stupčastog grafikona	27
Slika 22 Linijski grafikon sa svih 10 prikaza.....	28
Slika 23 Mali višekratnici u Pythonu, prikaz 3 x 4	29
Slika 24 Mali višekratnici u Pythonu, prikaz 1 x 10	30
Slika 25 Mali višekratnici u Tableauu, prikaz 1 x 10.....	31

Slika 26 Izbornik za postavljanje dimenzija	32
Slika 27 Mali višekratnici u Power BI, prikaz 3 x 4	32
Slika 28 Kružni grafikon u Pythonu	33
Slika 29 Silvermanovo pravilo	35
Slika 30 Kod za odabir grafikona.....	35
Slika 31 KDE krivulja	36
Slika 32 Raspršeni dijagram	36
Slika 33 Odabir oznake.....	37
Slika 34 Postavljanje osi x i y	37
Slika 35 Prikaz individualnog podataka u raspršenom grafikonu.....	37
Slika 36 Raspršeni grafikon u Tableauu.....	38
Slika 37 Linijski grafikon u Tableauu	38
Slika 38 Postavljanje osi x i y za izradu raspršenog dijagrama	39
Slika 39 Raspršeni grafikon u Power BI.....	39
Slika 40 Kod za izradu složenog stupčastog grafikona u Pythonu	40
Slika 41 Složeni stupčasti grafikon u Pythonu.....	40
Slika 42 Dodavanje parametra kako bi stupčasti grafikon postao složen u Tableauu	41
Slika 43 Složeni stupčasti grafikon u Tableauu	41
Slika 44 Složeni stupčasti grafikon u Power BI.....	42

9 LITERATURA

- [1] M. Cornell, M. Li, G. Kindlmann i C. Scheidegger, »Looks Good To Me: Visualizations As Sanity Checks,« 2018.
- [2] »Towards data science,« 17 Svibanj 2024. [Mrežno]. Available: <https://towardsdatascience.com/importance-of-data-visualization-anscombes-quartet-way-a325148b9fd2>.
- [3] »Tableau,« 17 Svibanj 2024. [Mrežno]. Available: <https://www.tableau.com/learn/articles/data-visualization>.
- [4] J.-D. Fekete, J. J. van Wijk, J. T. Stasko i C. North, »The Value of Information Visualization,« 2008.
- [5] T. Munzner, *Visualisation Analysis & Design*, 2014..
- [6] M. Fabien, »maelfabien,« [Mrežno]. Available: <https://maelfabien.github.io/machinelearning/Dataviz/#marks-and-channels>. [Pokušaj pristupa 18 September 2024].
- [7] R. Spence, »Information visualization,« u *Information visualization*, 2014.
- [8] C. Starbuck, »Data Visualization,« u *The Fundamentals of People Analytics*, 2023.
- [9] E. Tufte, »The Visual Display of Quantitative Information,« 2001.
- [10] S. Midway, »Principles of effective data visualizations,« 2020..
- [11] D. Bačić, A. Krbanjevic i N. Jukić, »Exploring Pie Charts and Part-To-Whole,« 2023..
- [12] S. Bagchi, »The Role of Colour Theory in Data Visualisation,« u *Words in Motion: The Power of Media and Arts*, 2024.

- [13] J. Wen i X. Wang, »Study of the visualization and Interaction of data : Take the Historical Data of the Winter Olympics as an Example,« 2020..
- [14] A. Franaklin , N. Pitchford, L. Hart, I. R. Davies, S. Clause i S. Jennings, »Salience of primary and secondary colours in infancy,« 2008..
- [15] S. W. Chiu, S. Gervan, C. Fairbrother, L. L. Johnson, A. Owen-Anderson, S. J. Bradley i K. J. Zuccker, »Sex-Dimorphic Color Preference in Children with Gender Identity Disorder: A Comparison to Clinical and Community Controls,« 2006..
- [16] M. H. Fischer, N. Dewulf i R. L. Hill, »Designing bar graphs: Orientation matters,« 2005..
- [17] L. Meloncon i E. Warner, »Data visualizations: A literature review and opportunities for technical and professional communication,« 2017.
- [18] C. Lauer i S. O'Brien, »How People Are Influenced by Deceptive Tactics in,« 2020..
- [19] L. Y.-H. Lo, Y. Cao, L. Yang i H. Qu, »Why Change My Design: Explaining Poorly Constructed,« 2024.
- [20] M. Friendly i D. Denis, »The Early Origins and Development of the Scatterplot,« 2005..
- [21] E. Tufte, The Visual Display of Quantitative Information, 1983.
- [22] Y. Wang, F. Han, L. Zhu, O. Deussen i B. Chen, »Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series,« 2018..
- [23] »eazybi,« [Mrežno]. Available: <https://eazybi.com/blog/data-visualization-and-chart-types>. [Pokušaj pristupa Svibanj 2024.].
- [24] F. Bouhafer, A. El Haddadi i M. Heyouni, »XEW 2.0 : Big Data Analytics tool based on swarm intelligence,« 2020.
- [25] »Future Learn,« [Mrežno]. Available: <https://www.futurelearn.com/info/courses/charting-and-data-visualization-in-excel-sc/0/steps/231657>. [Pokušaj pristupa 18 September 2024].
- [26] »investopedia,« [Mrežno]. Available: <https://www.investopedia.com/terms/m/metrics.asp>. [Pokušaj pristupa Srpanj 2024.].

- [27] M. Chen, L. Floridi i R. Borgo, »What Is Visualization Really For?,« 2013..
- [28] A. Burns, C. Xiong, S. Franconeri, A. Cairo i N. Mahyar, »How to evaluate data visualizations across different levels of understanding,« 2020..
- [29] Y. Zhu, »Measuring effective data visualization,« 2007.
- [30] »futurelearn,« [Mrežno]. Available: <https://www.futurelearn.com/info/courses/charting-and-data-visualization-in-excel/0/steps/231657>. [Pokušaj pristupa Srpanj 2024.].
- [31] »Power BI,« [Mrežno]. Available: <https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>. [Pokušaj pristupa Svibanj 2024.].
- [32] »Tableau,« [Mrežno]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>. [Pokušaj pristupa Svibanj 2024.].
- [33] »Wikipedia,« [Mrežno]. Available: https://en.wikipedia.org/wiki/Tableau_Software. [Pokušaj pristupa Svibanj 2024.].
- [34] »Matplotlib,« [Mrežno]. Available: <https://matplotlib.org/stable/tutorials/introductory/pyplot.html>. [Pokušaj pristupa Svibanj 2024.].
- [35] »jakevdp,« [Mrežno]. Available: <https://jakevdp.github.io/PythonDataScienceHandbook/02.03-computation-on-arrays-ufuncs.html>. [Pokušaj pristupa Svibanj 2024.].
- [36] »Seaborn,« [Mrežno]. Available: <https://seaborn.pydata.org/introduction.html>. [Pokušaj pristupa Svibanj 2024.].
- [37] »Cj-Mayes,« [Mrežno]. Available: <https://cj-mayes.com/2021/08/01/tackling-radar-charts/>. [Pokušaj pristupa September 2024].
- [38] »Tableau,« [Mrežno]. Available: <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/line-charts>. [Pokušaj pristupa Rujan 2024.].
- [39] »Stat psu,« [Mrežno]. Available: <https://online.stat.psu.edu/stat501/lesson/1/1.5>. [Pokušaj pristupa Rujan 2024.].

[40] J. VanderPlas, »Python Data Science Handbook,« [Mrežno]. Available: <https://jakevdp.github.io/PythonDataScienceHandbook/02.03-computation-on-arrays-ufuncs.html> .