

UDK: 004:65

Stručni rad

## VIZUALIZACIJA PODATAKA

### DATA VISUALIZATION

**Petar Subić<sup>1</sup>***prof.u penziji Visoka tehnička škola, Zrenjanin*<sup>1</sup>psubic@mts.rs

**Rezime:** Vizualizacija podataka je veoma značajna pomoć za neku analizu podataka i za identifikovanje značajnih odnosa i zakonitosti. Efektivno projektovanje tabela i dijagrama je takođe potrebno radi saopštavanja analize podataka drugima. Neki od najopštijih oblika vizualizacije podataka su tabele, dijagrami, unapređeni dijagrami, komandna tabla.

**Ključne reči:** Vizualzacija podataka, analiza poslovanja.

**Abstract:** Data visualization is very important for helping someone analyze data and identify important relations and patterns. The effective design of tables and charts is also necessary to communicate data analysis to others. Some of the most common forms of visualizing data are tables, charts, advanced charts, data dashboards.

**Key words:** Data visualization, Business analytics.

#### 1. UVOD

Brojne studije istraživanja u psihologiji i u drugim oblastima pokazuju da način mišljenja ljudi može obraditi mnogo brže i uspešnije vizualne slike kao što je dijagram ili da interpretira niz brojeva. Stoga prvi korak u naporu tumačenja podataka je neki oblik vizualizacije podataka. Vizualizacija podataka može biti stvaranje zbirne tabele ili može zahtevati kreiranje dijagrama kao pomoć u tumačenju, analizi ili zaključivanju na osnovu podataka. Vizualizacija podataka je veoma korisna i za identifikovanje pogrešnih podataka i za smanjenje našeg skupa pri otkrivanju značajnih odnosa i trenda.

Vizualizacija podataka je takođe značajna za prenošenje naše analize drugima. Iako, poslovna analitika se obavlja zbog donošenja boljih odluka, u mnogim slučajevima, krajnji donosioci odluka nisu lica koja analiziraju podatke. Stoga, lica koja analiziraju podatke obavljaju analizu za druge i trude se da ih ona razumeju. Prikladna tehnika vizualizacije podataka poboljšava sposobnost donosilaca odluka da lakše prihvate analizu. Ipak, neki način vizualizacije podataka je bolji od drugih.

U ovom radu se razmatraju neki opšti koncepti za vizualizaciju podataka kao pomoć analizi podataka i za prenošenje analize drugima, kao i postupci kako da projektujemo

tabele i dijagrame. Takođe, obuhvaćen je koncept dijagrama geografskog informacionog sistema.

Za vizualizaciju podataka najčešće se koristi softverski alat Microsoft Excel. U novije vreme pojavili su se takođe i mnogi drugi unapređeni softverski alati koji omogućuju specifičnu vizualizaciju podataka. [1] U ovom radu ukazano je samo uopšteno na mogućnosti njihovog korišćenja bez ukazivanja na konkretne primere.

## 2. TABELA

Prva odluka u vezi vizualizacije podataka je da li će tabela ili dijagram biti efektivnije. Uopšteno, dijagram može često preneti informaciju brže i lakše čitaocu, ali u nekim slučajevima tabela je mnogo podesnija. [2] Tabela treba biti korišćena kada:

1. Čitaoca treba uputiti u specifične numeričke vrednosti.
2. Čitaoc treba da uradi preciznije upoređenje između različitih vrednosti a ne samo upoređenje relativnih vrednosti.
3. Prikazane vrednosti imaju različite jedinice ili su sa veoma različitim značajem.

Statističke serije, koje predstavljaju rezultat deskriptivne analize, da bi bile preglednije i sistematičnije daju se u vidu tabele. Svaka tabela sastoji se od niza horizontalnih redova i vertikalnih kolona. Prvi horizontalni red naziva se zaglavlje a prva vertikalna kolona pretkolona. U zaglavlju i koloni nalazi se oznaka obeležja i njihovih modaliteta, vremenski intervali ili geografski pojmovi. Dobru tabelu karakteriše jasnoća, preglednost, jednostavnost i uočljivost osnovnih karakteristika obeležja pojave. Svaka tabela ima naziv koji se stavlja iznad nje i koji odmah treba da ukaže na to šta se podacima prikazuje, na koga se odnose, za koji vremenski period i sl. Iznad tabele u desnom uglu stavlja se oznaka za jedinicu mere (dinari, kilogrami, tone i sl.). Ukoliko ona nije ista za celu tabelu onda se oni naznačuju u odgovarajućoj koloni. Ispod tabele se obično navode izvor podataka i razna obaveštenja koja imaju značaj za pravilno tumačenje podataka.

Tabele zavisno od sadržine dele se na proste, složene i kombinovane.

Prosta tabela prikazuje podatke samo jedne serije bilo da je to vremenska serija ili serija strukture.

Složena tabela dobija se ako se dve ili više prostih tabela spoje u jednu. U takvim slučajevima pretkolona mora biti zajednička.

Putem kombinovane tabele prikazuju se podaci serija koji su dobiveni grupisanjem prema dva ili više obeležja. Svakako da ovakva tabela može da ima veći analitički značaj ali samo pod uslovom da se vodi računa da ne bude nepregledna, pretrpana podacima i komplikovana za razumevanje. Smatra se da ovakve tabele ne bi trebalo da imaju više od tri obeležja.

Sva polja u svim vrstama tabela moraju biti ispunjena tj ne bi trebala biti prazna. To znači da u svakom polju treba da stoji brojčani podatak ili neki znak koji označava zašto podatak nije unet. Ti znaci u našoj praksi su:

- nema pojave
- ... ne raspolaže se podatkom
- o podatak je manji od 0,5 od date jedinice mere
- () nepotpun odnosno nedovoljno proveran podatak

### 3.GRAFIKON

Serije statističkih podataka dobivenih u deskriptivnoj analitici mogu se radi razumljivijeg i pristupačnijeg istraživanja prikazati i pomoću grafikona. Grafički prikazi u većini slučajeva očiglednije i jasnije ukazuju na osnovne karakteristike posmatrane pojave. Zbog toga grafički način prikazivanja ima značajno mesto i često se koristi. Međutim, grafičko prikazivanje nije uvek podesno za prikazivanje dužih serija i nije uvek dovoljno precizno kao iskazivanje podataka u tabelama. [3]

Grafički prikazi serija podataka mogu biti u vidu geometrijskih oblika i u vidu oznaka na geografskim kartama. Grafikoni koje smo dobili korišćenjem geometrijskih oblika nazivaju se grafikoni, a grafikoni sa oznakama na geografskim kartama kartogrami.

Postojim više vrsta dijagrama koji se međusobno razlikuju u zavisnosti od toga da li se za prikazivanje podataka koriste tačke, linije (duži), površine ili prostorno prikazivanje u tri i više dimenzija. Otuda i njihovi različiti nazivi. Grafikoni u vidu tačaka nazivaju se stigmogramima, ako koristimo linije onda je to linijski dijagram, ako koristimo površine dobijamo površinski dijagram a ako se služimo prostornim prikazivanjem imamo stereogram. Ovo poslednji ređe se upotrebljavaju zbog njihovog komplikovanog prikazivanja.

Linijskim dijagramom, s obzirom da imaju samo jednu dimenziju, mogu se prikazati podaci samo jednog obeležja dok su prostorni dijagrami koji imaju tri dimenzije pogodni za prikazivanje više obeležja. Oznakama na geografskim kartama možemo se koristiti samo kod geografskih serija.

Može se izvršiti još jedna podela prikazivanja i to zavisno od toga da li se grafičko prikazivanje vrši u koordinatnom sistemu ili izvan njega.

U koordinatnom sistemu, pravouglom i polulogatitamskom ili logaroritamskom, prikazuju se podaci rasporeda frekvencija, vremskih serija, ili odnosi (veze) dve pojave. Grafičko prikazivanje van koordinatnog sistema vrši se obično površinskim dijagramima u obliku pravougaonika, krugova i sl.

Za prikazivanje statističkih serija često se koriste i razne figure u obliku simbola ili slika koje su srazmerne veličini pojave. Mada grafičko prikazivanje na ovaj način nije najpreciznije ono na veoma pristupačan način ukazuje na obim, strukturu i promene posmatrane pojave u određenom vremenskom periodu.

Potrebno je istaći da se bez obzira kojom vrstom grafikona se služimo, međusobni odnosi između simbola, dužina linija i veličina površina moraju biti u srazmeri sa veličinom vrednosti pojave na koju se odnose.

U ovom radu se ukazuje samo na najopštije korišćenje dijagrama za prikaz i analizu podataka. Za kreiranje dijagrama najčešće korišćen softverski paket je Excel.

### ***Grafičko prikazivanje atributivnih serija***

Za grafičko prikazivanje atributivnih serija najčešće se koriste površinski dijagram tj. histogram u obliku pravougoinika, krugova i drugih površina. Histogram u obliku pravougaonika može biti postavljen tako da pravougaone površine budu jedna iznad druge. U tom slučaju visina te površine je jednaka jedinici a dužina označava veličinu pojave.

Pravougaone površine mogu biti postavljene i uspravno i tada se naziju „stubićima“. U tom slučaju njihova osnova je jednaka jedinici, a visina treba da je srazmerna podacima pojave na koje se podnosi. Jednostruke stubiće koristimo kada prikazujemo odnose podataka samo jedne serije. Ako želimo da uporedimo dve ili više serija za grafičko prikazivanje koriste se dvostruki ili višestruki stubići, ali oni tada moraju biti različito obojeni ili šrafirani i naznačena legenda.

Za grafičko prikazivanje strukture skupa koristi se pravougaona ili kružna površina. Pravougaona površina „stubić“ ili kružna površina predstavljaju celinu pojave a delovi pojave, iskazani bilo u relativnim bilo u apsolutnim iznosi iskazuju se srazmernim delovima tih površina. Svi ti delovi moraju se različito obojiti ili šrafirati kako bi grafičko predstavljanje bilo preglednije. Međutim, ova preglednost se smanjuje kada se prikazuje veći broj podataka. Zbog toga se ovaj način prikazivanja najčešće koristi kada je obeležje manje izdiferencirano tj kada je grupisano u manje modaliteta. Kada strukturu skupa prikazujemo kružnom površinom odgovarajući delovi skupa moraju se preračunati u stepene.

Za grafičko prikazivanje prostornog razmeštaja skupa mogu se koristiti geografske karte na kojima se odgovarajuće teritorije različito šrafiraju ili boje. Svako grafičko obeležavanje uz pomoć bojenja, senčenja, tačkica itd. odgovara određenom intezitetu ili učestalosti pojave koje razmatramo. Ovakvo grafičko prikazivanje naziva se kartogramom. Kartogrami su našli veću primenu u regionalnim analizama ali je njihovo korišćenje ipak ograničeno na grafičko prikazivanje geografskih serija.

### ***Grafičko prikazivanje serija distribucije frekvencija***

Za grafičko prikazivanje serije distribucije frekvencija koristi se pravougli koordinatni sistem. Pri tome se na apcionu osu raspoređuju vrednosti obeležja (x) a na ordinatnu osu njihove odgovarajuće frekvencije (y). Na ordinatnoj osi merilo bi trebalo uvek da počne od nule.

Svaki raspored frekvencija moguće je prikazati u tri oblika i to kao histogram frekvencija, poligon frekvencija i kriva frekvencija.

*Histogram frekvencija* je dijagram koji je ucrtan u vidu stubića na pravouglom koordinatnom sistemu. Površina jednog stubića izražava veličinu dela skupa koji ima dato obeležje a površina svih stubića pokazuje veličinu celokupne posmatrane mase. U slučajevima kada želimo da seriju distribucije frekvencija sa neprekidnim vrednostima obeležja prikažemo uz pomoć histograma, tada će se na apcisonoj osi naznačiti granice grupnih intervala tj. širina stubića biće određena širinom grupnih intervala, a visina stubića frekvencijama pojedinih grupa. Međutim, u slučajevima kada je širina grupnih intervala različita tada se prilikom grafičkog prikazivanja moraju korigovati veličine frekvencija tih grupa kako bi se obezbedilo da površina stubića bude srazmerna veličini frekvencije

Histogram frekvencija može se crtati u dve ili više dimenzija. Ako se želi da se prikaže dve serije distribucije frekvencija kod kojih je grupisanje izvršeno po istom numeričkom obeležju koristi se dvostruki histogram frekvencijama sa horizontalnim stubićima. Ovakav vid grafičkog prikazivanja često se koristi u demografskoj statistici kada se prikazuje struktura stanovništva prema starosti i polu u obliku tzv. starosne piramide.

*Poligon frekvencija* se izrađuje takođe u pravougaonom koordinatnom sistemu. Iznad sredine grupnih intervala određuje se visina koja odgovara veličini frekvencija te grupe i obeležava tačkom. Spajanjem tako označenih vrhova ordinata dobija se izlomljena linija koja se spaja sa apcisonom u tačkama koje odgovaraju zamišljenim sredinama grupnih intervala pre stvarnog prvog intervala i posle stvarnog poslednjeg intervala. U slučajevima kada su prvi i zadnji grupni intervali otvoreni, izlomljena linija se ne spaja sa apcisonom.

*Kriva frekvencija* se naziva ona grafički predstavljena kontinuirana linija koja je najbolje prilagođena histogramu frekvencija. Kriva frekvencija, za razliku od histograma i poligona frekvencija, predstavlja teorijsku krivu. Prilikom njenog ucrtavanja prepostavlja se da su grupni intervali beskrajno mali a broj obuhvaćenih jedinica beskrajno velik. Pomoću nje je moguće, pod uslovom da su vrednosti obeležja neprekidne, utvrditi teoretsku frekvenciju za svaku vrednost obeležja i za one koje nisu date u empiriskoj distribuciji frekvencija,

U praksi, radi jasnijeg isticanja nekih karakteristika serije distribucije frekvencija često se koristi i grafičko prikazivanje kumuliranih vrednosti u pravougaonom koordinatnom sistemu. Dobiven linijski dijagram naziva se kumulanta ili ogiva. Pre nego što se pristupi ucrtavanje kumulante nužno je kumulirati frekvencije. To kumuliranje može otpočeti od najmanje vrednosti obeležja na taj način da se frekvencije postepeno dodaju zbiru prehodnih frekvencija. Kumulativ najveće vrednosti obeležja ujedno predstavlja i zbir svih frekvencija. Na taj način dobija se kumulacija „ispod“ (rastuća) za razliku od kumulacije „iznad“ (opadajuća) kada se sabiranje započinje od najveće vrednosti obeležja.

### **Grafičko prikazivanje vremenskih serija**

Za grafičko prikazivanje vremenskih serija veoma je podesan pravougli koordinatni sistem. Najčešće se koristi histogram, linijski dijagram, polarni krug itd. Na koji način ćemo grafički prikazati podatke zavisi od cilja istraživanja i od prirode podataka. Pri tome se mora voditi računa da se na apcisu ucrtavaju oznake za vreme a na ordinatu merilo za određivanje veličine posmatrane pojave. Merilo na ordinatnoj osi po pravilu počinju od nule. Međutim, u slučajevima kada je početna veličina velika, da bi se izbegao veći prazan prostor koji ne bi bio iskorišćen, označava se presek ordinatne ose sa dve talasaste linije i odmah iza toga stavlja se merilo raspoređeno u jednakim razmacima.

Linijski dijagram je pogodniji za grafičko prikazivanje ovih podataka jer na adekvatniji način ističe promene pojave u vremenu i njihov pravac kretanja. Oni se crtaju tako što se iznad vremenskih oznaka stavlja tačka na onoj visini koja odgovara veličini pojave. Zatim se tačke spoje pravim linijama tako da dobijemo izlomljenu liniju koju ne zivamo linijski dijagram.

U zavisnosti od toga da li je na ordinatnoj osi odstojanja brojeva u aritmetičkoj progresiji ili u logaritamskoj, razlikujemo aritmetički dijagram ili polulogaritamski dijagram. Aritmetički dijagram pokazuje apsolutne varijacije posmatrane pojave tj. apsolutne promene tokom vremena. Na aritmetičkom dijagramu mogu se prikazivati i analizirati dve ili više pojava pod uslovom da su njihove veličine izražene u istim jedinicame mere i da nisu velike razlike u njihovim veličinama.

U statističkim istraživanjima često se koristi i grafičko prikazivanje više serija koje su sastavni deo jedne celine. Na taj način moguće je izrazitije uočiti promene u strukturi posmatrane pojave. Takav dijagram naziva se kumulativni dijagram dinamike.

Polulogaritamski dijagram vremenskih serija konstruiše se kada želimo da na jednom grafikonu analiziramo kretanje više serija a između njih su veće apsolutne razlike u nivou podataka, tako da ih je teško upoređivati na aritmetičkom dijagramu. Kod ovog dijagrama na ordinati koja služi za određivanje veličine pojave, aritmetička skala se zamenjuje logaritamskom. Na apcisu ucrtava se merilo koje označava vremenski period po aritmetičkoj progresiji.

Za razliku od aritmetičkog dijagrama koji pokazuje apsolutne varijacije posmatrane pojave, ovaj dijagram ističe proporcionalne, odnosno relativne (procentualne) varijacije. Kolika je ta proporcionalna varijacija zaključuje se na osnovu nagiba linije prema apcisi. Manji nagib ukazuje da je proporcionalna promena manja ili obrnuto. Upravo zato što pokazuje relativne promene on može da se iskoristi za upoređivanje serija u kojima su znatne razlike u visini apsolutnih podataka.

Za grafičko prikazivanje vremenskih serija, naročito ako raspolažemo mesečnim podacima, često se koristi polarni koordinacioni sistem. Ovaj sistem sastoji se od pola, kako se naziva tačka iz koje polazi duž koja se naziva polarnom osom, zatim, jednog broja potega i polarni ugao. Na polarnoj osi ucrtava se merilo kojim se odmeravaju veličinje posmatrane pojave i nanose na potege. Te tačke na potezima spajamo pravim

linijama i formiramo tzv. polarni dijagram. Jedan polarni dijagram može se iskoristiti i za analiziranje kretanja pojave po mesecima i za više godina, pa ukoliko pojava raste ili opada linija na dijagramu dobija oblik spirale.

### ***Unapredena vizualizacija podataka***

Napred u ovom radu je ukazano samo na neke osnovne ideje za efektivnije korišćenje vizualizacije kako za analizu podataka tako i za informisanje drugih o analizi podataka. Navedeni dijagrami se najčešće koriste i verovatno će zadovoljiti većinu potreba za vizualizaciju podataka. Ipak, mnogi unapredeni koncepti, dijagrami i alati mogu biti korišćeni za poboljšanje tehnika vizualizacije podataka. Dalje se ukazuje samo na neke od njih.

*Unapređen grafički prikaz.* Napred navedeni linijski dijagrami dosta se koriste za aplikacije vizualizacije podataka, ipak, postoje i drugi unapredeni grafički prikazi koji mogu biti veoma korisni u određenim situacijama. Jedan od takvih prikaza za ispitivanje podataka sa više od dve varijable je *parallel-coordinates plots* (poralelno-koordinatni nacrt). To je tehnika koja se koristi za grafičko predstavljanje višedimenzionalnih podataka. Da bi se prikazao skup podataka u n-dimenzionalnom prostoru na dvodimenzionalnoj površini (ekran računara) formira se grafički prikaz koji se sastoji od n-vertikalnih osa, jednako raspoređenih paralelnih linija. To znači da on obuhvata različite ose za svaku varijablu. Svako opaženje u skupu podataka je predstavljeno sa ucrtanom linijom na paralelnom-koordinatnom dijagramu spajanjem svake vertikalne ose. Najviša tačka linije na svakoj vertikalnoj osi predstavlja vrednost dobivena posmatranjem koja je predstavljena na vertikalnoj osi.

*Prikaz podataka geografskog informacionog sistema* spaja geografsku kartu i vizualni prikaz statistički podataka za određenu geografsku oblast radi davanja specifičnih informacija za različite geografske oblasti. Informacije predstavljaju podatke o veštačkim i prirodnim objektima na nekom prostoru. To su, npr. infrastrukturni objekti, stambeni, sportski i drugi objekti, hidroakumulacije, reke, namena zemljišta, podaci o terenu, geologija terena i dr. Informacije o geometriji objekata mogu biti u obliku rastera i vektora. Raster se sastoji od redova i kolona ćelija, koje se nazivaju pikseli, pri čemu svaka od tih ćelija ima jednu, određenu, brojnu vrednost. U slučaju slike, ta brojčana vrednost, predstavlja broj boje (boje su kodirane brojevima). U slučaju prikaza neke druge informacije vrednost piksela ne predstavlja boju već predstavlja prostorni podatak. Pošto jedan piksel ima jednu vrednost, grubo rečeno, raster ima onoliko informacija koliko ima piksela. Rasteri se mogu prikazivati po kanalima, RGB kanalima, odnosno u crvenom, zelenom i plavom delu spektra vidljive svetlosti. Preklapanjem tako pripremljenih rastera, može se dobiti znatno veći broj informacija o nekom području.

Prikaz podataka geografskog informacionog sistema ima uticaj na geografiju kao disciplinu, jer može da uključi fizičke, biološke, demografske, prostorne i ekonomske podatke, pa predstavlja dragoceni alat u prirodnim, medicinskim, tehničkim i društvenim naukama, kao i u analizi i planiranju poslovanja preduzeća.

*Dashboard (komandna, kontrolna tabla)* je alat vizualizacije podataka koji ilustruje višestruki metrički sistem i koji se automatski ažurira sa novim podacima koji su na raspolaganju. Dashboard obično prikazuje seriju grafika, grafikona i drugih vizualnih pokazatelja koji se mogu lako pratiti i tumačiti. Ovaj alat je sličan automobilske instrument tabli koja obezbeđuje informacije o brzini vižnje, nivou goriva i temperaturi motora tako da vozač može preduzeti efikasne aktivnosti. Tako i ovaj alat obezbeđuje važan metrički sistem za koji donosilac odluke ima potrebu da bi brže preduzeo odgovarajuće aktivnosti za korekciju tekućih poslovnih aktivnosti. Te informacije obično se nazivaju osnovni pokazatelji poslovanja. Ova tabla treba da obezbedi blagovremeno sve informacije koje su značajne za donosica odluke i to na takav način koji je pogodan da informiše pre nego što nastane odstupanje od planiranih aktivnosti. Idealno je, da se svi indikatori poslovanja budu prikazani na jednom ekranu računara na način da korisnik može brzo da ih pregleda i da uoči sve promene tekućeg poslovanja i neuobičajene vrednosti koje zahtevaju pažnju. Kolor treba da bude korišćen da upozori na specifične vrednosti u odnosu na propisane vrednosti, ali korišćenje kolora trebalo da bude uzdržano.

#### **4.ZAKLJUČAK**

Vizualizacija podataka je veoma značajna pomoć za neku analizu podataka i za identifikovanje značajnih odnosa i zakonitosti. Efektivno projektovanje tabela i dijagrama je takođe potrebno radi saopštavanje analize podataka drugima. Neki od najopštijih oblika vizualizacije podataka su tabele, dijagrami, unapređeni dijagrami, komandna tabla.

#### **LITERATURA**

- [1] Schnidejans, M., Schniderjans, D., Starkej, C., (2015), *Business Analytics Principles, Concept and Applications with SAS*, Pearson Education, Inc.
- [2] Sharda, R., Delen, D., Turban, E., (2014), *Business Intelligence and Analytics*, Pearson Education Limited.
- [3] Jeffrey D. Camm, James J. Cochran, Michael J. Fry, Jeffrey W. Ohlmann, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, (2017) *Essentials of Business Analytics*, Cengage Learning.