

# DEO II

## Priprema datoteke s podacima

Priprema datoteke s podacima za analizu obuhvata više koraka. Prvo se napravi prazna datoteka i u nju unesu podaci dobijeni u istraživanju, u obliku definisanom u šifarniku (objašnjeno u poglavlju 2). Zatim u datoteci s podacima treba pronaći eventualne greške i ukloniti ih. U drugom delu knjige razmatraju se upravo ta dva koraka. U poglavlju 4 obrađeni su postupci za pravljenje datoteke i unošenje podataka u nju. U poglavlju 5 objašnjen je postupak uklanjanja grešaka iz datoteke s podacima.





# Pravljenje datoteke za podatke i unošenje podataka

Postupak izrade datoteke s podacima i njihove analize sastoji se od više koraka. Dijagram toka prikazan na sledećoj stranici sadrži osnovne potrebne korake. U ovom poglavlju provešću vas kroz postupak pravljenja datoteke i unošenja podataka u nju.

Objasnićemo tri ključna koraka u pripremi datoteke s podacima:

- *Korak 1.* Prvi korak je pregled i, po potrebi, izmena opcija koje IBM SPSS koristi za prikazivanje podataka i dobijenih rezultata.
- *Korak 2.* Sledeći korak je izgradnja strukture datoteke s podacima definisanjem promenljivih.
- *Korak 3.* Poslednji korak je unošenje podataka, tj. vrednosti koju je svakoj promenljivoj dodelio svaki učesnik ili ispitanik.

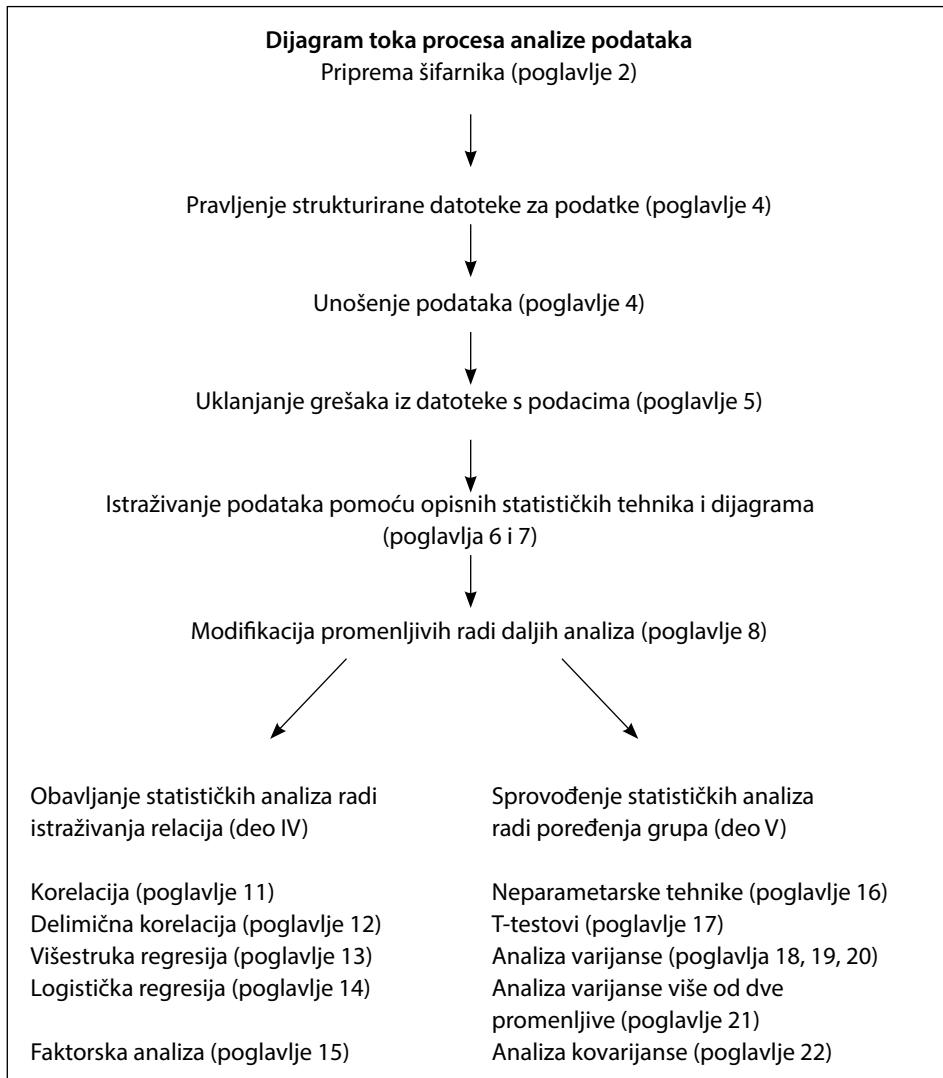
Navedene postupke ćemo ilustrovati na primeru datoteke **survey.sav**, opisanoj u dodatku knjige. U dodatku je i šifarnik pomoću kojeg su dobijeni numerički podaci za analizu rezultata.

Datoteke s podacima mogu se uvesti i iz drugih programa za tabelarne proračune (npr. iz Excela). To može umnogome da olakša postupak unošenja podataka, naročito studentima koji na svojim računarima nemaju IBM SPSS. U Excelu korisnik može da napravi praznu datoteku i da u nju unese podatke na svom kućnom računaru. Kada to završi, datoteku uveze u IBM SPSS i obradu i analizu podataka nastavi u njemu. Pri kraju poglavlja biće navedena uputstva za unošenje podataka u Excelu.

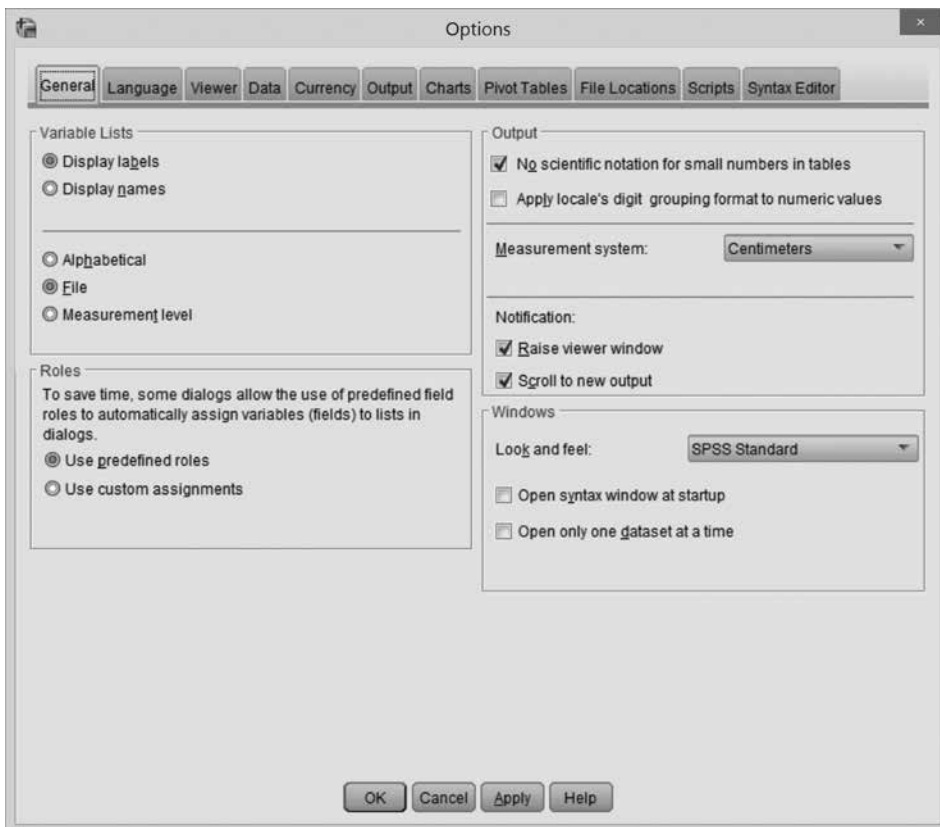
## Izmena opcija u IBM SPSS-u

Pre nego što napravite datoteku s podacima, bilo bi dobro da proučite opcije (postavke) IBM SPSS-a od kojih zavisi način prikazivanja podataka i rezultata. To su opcije za definisanje načina prikazivanja promenljivih, vrste tabela u kojima će rezultati biti prikazani i mnogih drugih aspekata programa. Možda će vam početka nešto od ovoga biti nerazumljivo, ali čim počnete da unosite i analizirate podatke u IBM SPSS-u, možete se vratiti na ovaj odeljak.

Ukoliko računar na kome je instaliran IBM SPSS delite s drugim korisnicima (npr. u računarskoj laboratoriji), morate znati ove postavke zato što ih drugi studenti mogu promeniti, a to će uveliko izmeniti izgled programa. Trebalo bi da znate kako da vratite opcije u staro, poželjno stanje.



Da biste otvorili prozor **Options**, pritisnite **Edit** u glavnom meniju na vrhu ekrana, pa izaberite stavku **Options**. Otvoriće se prozor prikazan na slici 4.1. Tu je navedeno mnogo opcija; većinu nećete morati da menjate. U nastavku ću opisati samo ključne opcije, podeljene po karticama na kojima se nalaze. (S kartice na karticu prelazite pritiskom na jezičak one na koju hoćete da pređete.) Nemojte pritiskati **OK** dok na svim karticama ne izmenite sve opcije koje nameravate da izmenite.



Slika 4.1 Prozor Options.

### Kartica General

Kada dođe vreme za obavljanje analiza, promenljive mogu biti poredane abecedno ili redom kojim se pojavljuju u datoteci. Ja uvek koristim redosled pojavljivanja u datoteci, zato što je on jednak redosledu stavki u upitniku i u šifarniku. Da bi se promenljive prikazivale po redosledu pojavljivanja u datoteci, samo se uverite da je uključena opcija **File** u odeljku **Variable Lists** (gore levo).

U odeljku **Output** (gore desno), treba da potvrdite polje **No scientific notation for small numbers in tables**. Tako ćete sprečiti pojavljivanje nekih veoma čudnih brojeva među rezultatima vaših statističkih analiza.

### Kartica Data

Pritisnite karticu **Data** da biste podesili način prikazivanja datoteke s podacima. Ako promenljive s kojima radite nisu decimalne vrednosti, mogli biste promeniti format prikazivanja svih promenljivih. U odeljku **Display Format for New Numeric Variables**, za broj decimalnih mesta zadajte 0. To znači da se decimalna mesta neće prikazivati ni za jednu novu promenljivu. Tako smanjujete datoteku s podacima i pojednostavljujete njen izgled.

## Kartica Output

Opcije na ovoj kartici služe za prilagođavanje načina prikazivanja običnih imena i dužih, opisnih imena promenljivih u rezultatima. Na samom dnu, u odeljku **Variable values in labels are shown as**, u padajućoj listi izaberite **Values and Labels**. Tako ćete u tabelama generisanim u prozoru **Viewer** moći da vidite i numeričke vrednosti i duža, opisna imena promenljivih.

U odeljku **Output Display** izaberite **Pivot tables and charts**. To je posebno važno ako nameravate da koristite postupke za neparametarske testove u SPSS-u.

## Kartica Pivot Tables

Većinu rezultata statističkih analiza IBM SPSS prikazuje u izvedenim tabelama. Na kartici **Pivot Tables** možete sa opsežnih lista izabrati format tih tabela. Stil koji vam najviše odgovara naći ćete eksperimentisanjem. Kada tek počnem s nekom analizom, upotrebljavam stil **CompactBoxed**, čime štedim prostor (i papir prilikom štampanja).

Možda će vam biti korisna i opcija koja se nalazi na dnu kartice **Pivot Tables** – označena sa **Copying wide tables to the clipboard in rich text form**. Otvorite padajuću listu i izaberite **Shrink width to fit**. To je korisno kada komandom **Paste** prenosite rezultate iz IBM SPSS-a u Microsoft Word a oni su preširoki za datu stranicu (čest problem u nekim statističkim postupcima predstavljenim kasnije u ovoj knjizi).

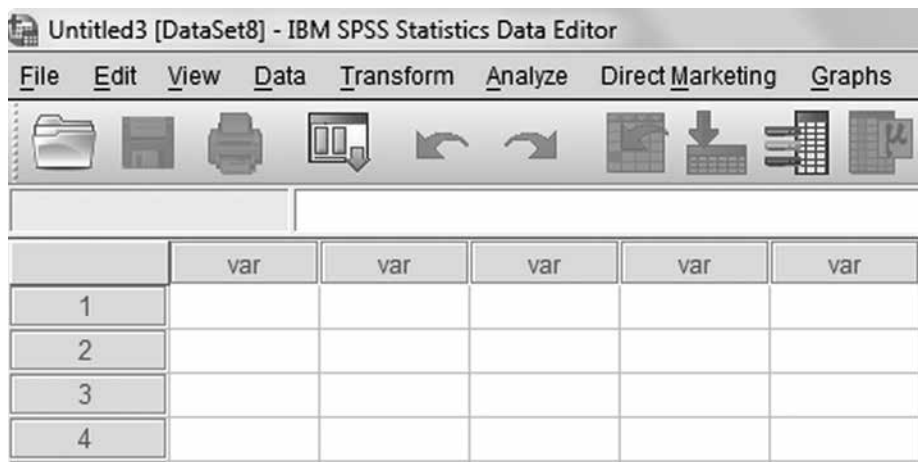
Stilove tabela možete menjati koliko god puta hoćete, samo ne zaboravite da stil treba promeniti *pre* analize. Kada se tabele već pojave u rezultatima, više im se ne može menjati stil, ali se mnoga njihova svojstva (recimo, veličina fontova, širina kolona) ipak mogu menjati u prozoru **Pivot Table Editor**. Otvorićete ga kada dvaput pritisnete tabelu koju treba izmeniti.

Nakon što izmenite sve što želite na raznim karticama iz okvira za dijalog **Options**, pritisnite **OK**. Zatim možete preći na definisanje promenljivih i unošenje podataka.

## Definisanje promenljivih

Pre nego što počnete da unosite podatke, IBM SPSS-u morate saopštiti kako vam se promenljive zovu i dati mu uputstva za šifrovanje (kodiranje), tj. pretvaranje vrednosti promenljivih u brojeve. To se naziva definisanje promenljivih, a obavlja se u prozoru **Data Editor** (slika 4.2). Prozor **Data Editor** se sastoji od dva različita prikaza: **Data View** i **Variable View**. Iz jednog u drugi prelazite tako što pritisnete jezičke njihovih kartica na dnu ekrana, s leve strane.

Svaka kolona u prozoru **Data View** odgovara jednoj promenljivoj (engl. *variable*) i nosi ime **var** (slika 4.2). Ta imena će biti zamenjena imenima promenljivih koje ste naveli u svom šifarniku. Duž leve strane su brojevi **1**, **2**, **3** itd. To su redni brojevi analiziranih slučajeva, koje IBM SPSS dodeljuje svakom redu s podacima. To *nisu* vaši brojevi za identifikaciju slučajeva; osim toga, brojevi redova se menjaju kada sortirate datoteku ili je podelite da biste podatke analizirali u podskupovima.



Slika 4.2 Prozor Data Editor.

### Postupak definisanja promenljivih

Da biste definisali sve promenljive od kojih se vaši podaci sastoje, najpre pritisnite jezičak kartice **Variable View** na dnu ekrana, levo. U tom prikazu (slika 4.3), duž leve strane su navedeni redni brojevi promenljivih, a duž vrha njihovi atributi (ime, tip, širina, broj decimala, duže opisno ime itd.).

The screenshot shows the Variable View window for a file named '\*survey5ED.sav [DataSet1]'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The main area is a table with the following data:

	Name	Type	Width	Decimals	Label	Values	Missing
1	id	Numeric	3	0		None	None
2	sex	Numeric	3	0	sex	{1, MALES}...	None
3	age	Numeric	3	0		None	None
4	marital	Numeric	8	0	marital status	{1, SINGLE}...	None
5	child	Numeric	5	0	child	{1, YES}...	None
6	educ	Numeric	5	0	highest educ c...	{1, PRIMAR...	None

Slika 4.3 Variable View.

Vaš zadatak je da sve promenljive definišete tako što ćete u šifarniku navesti potrebne informacije za svaku od njih. Deo tih informacija morate sami da obezbedite (npr. ime); ostale automatski dodeljuje IBM SPSS, koristeći podrazumevane vrednosti. Naravno, te vrednosti se mogu menjati po potrebi. U nastavku ćemo navesti osnovne neophodne informacije. Upotrebila sam zaglavlja koja odgovaraju zaglavljinama kolona prikazanim u prozoru **Variable View**. Sledi opis jednostavnih koraka od kojih se postupak sastoji; međutim, postoje i brojne prečice koje ćete moći da koristite kada upoznate postupak. Njih navodim kasnije, u odeljku „Opcione prečice“. Najpre treba upoznati osnovne tehnike.

**Name**

U ovu kolonu upišite kratko ime svake promenljive navedene u šifarniku. Pomoću imena će svaka promenljiva u datoteci s podacima biti identifikovana. Imena treba da su što kraća i da ne premašuju 64 znaka. Uz to, moraju zadovoljiti IBM SPSS-ovu konvenciju o imenovanju (objašnjenu u poglavlju 2). Imena moraju biti jedinstvena, početi slovom, i ne smeju sadržati razmake (beline) niti specijalne znakove. Primere dobrih imena za promenljive pročitajte u šifarniku u dodatku. Tamo su navedena imena promenljivih koje su korišćene u datotekama s podacima napravljenim za ovu knjigu. (Pojednosti o tim datotekama videti na str. ix.)

**Type**

Kada upišete prvo ime promenljive, u koloni **Type** će se automatski pojaviti podrazumevana vrednost tipa promenljive – **Numeric**. To će vam odgovarati za numeričke promenljive, ali za tekstualne informacije (recimo, prezime) tip treba promeniti u **String**. Ukoliko među vašim podacima ima i datuma, tu je opcija **Date**. Tip promenljive menjate tako što pritisnete unutar ćelije; trebalo bi da se pojavi okvir s tri tačke i da vam se prikažu dostupne opcije. U istom prozoru može se podešavati širina promenljive i broj njenih decimalnih mesta.

**Width**

Podrazumevana vrednost širine je 8 (osim ako je promenjena korišćenjem instrukcija iz prozora **Options** predstavljenim ranije u ovom poglavlju). To je obično dovoljno za većinu podataka. Ukoliko vaša promenljiva može imati veoma velike vrednosti ili je u pitanju tekstualna promenljiva tipa *string*, promenite podrazumevanu vrednost. U protivnom, ostavite je na 8.

**Decimals**

Podrazumevana vrednost za broj decimalnih mesta obično je 2 (to se može izmeniti u prozoru **Options**, opisanom ranije u ovom poglavlju). Ako vaša promenljiva ima decimale, promenite ovu postavku kako god vam odgovara.

**Label**

Kolona **Label** služi da promenljivoj date duže ime od onog u koloni **Name**. Duža, opisna imena promenljivih navode se u rezultatima svih analiza koje obavljate u IBM SPSS-u. Na primer, promenljivoj kratkog imena UOPTIM mogli biste dodeliti duže opisno ime Ukupan optimizam.

**Values**

U koloni **Values** definišete značenje brojeva kojima ste šifrovali tekstualni sadržaj promenljivih, tj. zamenili ga. Pokazaću taj postupak za promenljivu pol (sex) s prethodne slike.



1. Pritisnite unutar ćelije pod naslovom **Values** za promenljivu koju želite da zadate. Trebalo bi da se na desnoj strani ćelije pojavi okvir s tri tačke. Otvoriće se okvir za dijalog **Value Labels**.
2. Pritisnite polje **Value** i u njega upišite 1.
3. Pritisnite polje **Label** i u njega upišite „muški“.
4. Pritisnite dugme **Add**. U zbirnom polju pisaće 1=muški.
5. Ponovite za ženski pol: za **Value** upišite 2, za **Label** upišite „ženski“. Pritisnite dugme **Add**.
6. Kada definišete sve brojeve navedene u šifarniku, pritisnite dugme **OK**.

### **Missing**

Ponekad istraživači koriste specifične vrednosti kako bi pokazali da podatak nedostaje, tj. ne postoji. To ne morate da radite, pošto IBM SPSS svaku praznu ćeliju tumači kao nepostojanje podatka. Dakle, ukoliko nameravate da ostavite prazno mesto kada neki delić informacija nije dostupan, u kolonu **Missing** prikaza **Variable View** ne treba ništa da upisujete.

S druge strane, ako nepostojanje vrednosti nameravate da kodirate određenim brojem (npr. 99=nema vrednost), *morate* ga upisati u odeljak **Missing**, inače će ga IBM SPSS u svim statističkim analizama smatrati za legitimnu vrednost (s potencijalno katastrofalnim posledicama!). Pritisnite unutar ćelije a zatim zasenčen okvir s tri tačke koji će se pojaviti. Izaberite stavku **Discrete missing values** i upišite taj broj (npr. 99) u predviđeno polje. Mogu se zadati do tri vrednosti. Pritisnite **OK**.

Ako koristite te posebne kodove, preporučljivo je da se vratite i dodelite duža, opisna imena odgovarajućim vrednostima u koloni **Values**.

### **Columns**

Podrazumevana širina kolona obično je 8, što je dovoljno za većinu namena. Izmenite je samo ukoliko očekujete ogromne vrednosti ili dugačka imena promenljivih.

### **Align**

Vrednosti u kolonama obično su poravnate kako treba. Nema potrebe da to menjate.

### **Measure**

Zaglavlje kolone **Measure** odnosi se na način merenja svake promenljive pojedinačno. Podrazumevana vrednost je **Scale**, što znači da se koristi intervalna merna skala, pogodna za neprekidne (kontinualne) podatke. Kada se promenljiva sastoji od kategorija (npr. pol, bračno stanje...), pritisnite tu ćeliju i zatim dugme sa strelicom koje će se pojaviti. Za kategorijske podatke izaberite nominalnu skalu (**Nominal**), a za one za koje je važan poredak mogućih vrednosti (npr. nivo obrazovanja), izaberite ordinalnu skalu (**Ordinal**). Važno je da ispravno zadate merne nivoe promenljivih jer će inače SPSS sprečiti da koristite neke postupke (npr. izradu dijagrama).

## **Role**

Nema potrebe da išta menjate u ovom odeljku; samo ostavite podrazumevanu vrednost, **Input**.

## **Opcione prečice**

Kada datoteka s podacima ima mnogo promenljivih, opisani postupak ume da bude zametan. Ima više prečica pomoću kojih ga možete ubrzati. Ukoliko vise promenljivih ima iste atribute (npr. tip, širina kolone, broj decimala), definišite prvu promenljivu a zatim njene atribute kopirajte na ostale promenljive.

### ***Kopiranje atributa iz definicije jedne promenljive na drugu promenljivu***

---

1. U prikazu **Variable View** pritisnite ćeliju sa atributom koji treba kopirati (npr. širinu, Width).
  2. Otvorite meni **Edit** i u njemu izaberite **Copy**.
  3. Pritisnite ćeliju istog atributa koja pripada drugoj promenljivoj (onoj na koju treba primeniti taj atribut).
  4. Otvorite meni **Edit** i u njemu izaberite **Paste**.
- 

### ***Kopiranje atributa iz definicije jedne promenljive na više drugih promenljivih***

---

1. U prikazu **Variable View** pritisnite ćeliju sa atributom koji treba kopirati (npr. širinu, Width).
  2. Otvorite meni **Edit** i u njemu izaberite **Copy**.
  3. Pritisnite ćeliju istog atributa koja pripada prvoj promenljivoj na koju ga treba primeniti, zadržite levi taster miša pritisnut i prevucite kursor niz kolonu birajući tako sve promenljive na koje taj atribut treba primeniti.
  4. Otvorite meni **Edit** i u njemu izaberite **Paste**.
- 

## **Zadavanje istih atributa nizu novih promenljivih**

Kada podaci obuhvataju skale koje se sastoje od više pojedinačnih stavki, odjednom napravite sve te nove promenljive i definišite njihove atribute. Kao primer, sledi opis tog postupka za šest stavki (od optim1 do optim6) skale optimizma. Ako želite da uradite ovo kao vežbu, otvorite novu datoteku za podatke (**File, New, Data**).

1. U prikazu **Variable View** definišite atribute prve promenljive (optim1) prema prethodnim uputstvima. Trebalo bi da zadate i duža opisna imena koja odgovaraju mogućim vrednostima te promenljive: 1=nimalo se ne slažem, 2=ne slažem se, 3=neutralno, 4=slažem se, 5=potpuno se slažem.
2. Dok je izabran prikaz **Variable View**, pritisnite broj reda u kome se promenljiva nalazi (trebalo bi da se time izabere ceo taj red).