

Laboratorio di Statistica e Analisi Dati: Lezione 6

Tommaso C. & Marco G.

30 Novembre - 2 Dicembre 2016

Esercizio I

1. Importare il dataset `autos.csv`
2. Visualizzare la struttura e le prime 6 righe
3. Modificare i dati nei seguenti modi:([Aiuto modifica fattori](#))
 - *nella colonna seller:*
 - `privat` → `privato`
 - `gewerblich` → `negozio`
 - *nella colonna offerType:*
 - `angebot` → `offerta`
 - `gesuch` → `richiesta`
 - *nella colonna notRepairedDamage:*
 - `ja` → `si`

- nein → no
- *nella colonna vehicleType:*
 - limousine → limo
 - andere → altro
 - kombi → s.w.
 - kleinwagen → util.
- *eliminare tutte le righe con un anno di registrazione non nell'intervallo [1901, 2016] (2 modi possibili)*
- *eliminare tutte le righe con prezzo superiore ai 100000 euro*

Esercizio I (cont)

4. Contare il numero di brand di auto disponibili
5. Come sono distribuiti i tipi di veicolo?
6. Tracciare un diagramma a barre delle frequenze assolute del tipo di veicolo
7. Fare un grafico a torta delle frequenze assolute del tipo di veicolo
8. Tracciare il grafico a torta dei chilometri (**Attenzione a cosa disegnate**)
9. Visualizzare le marche delle auto con powerPS uguale a 30
10. Calcolare media, mediana, max, min, primo e terzo quartile degli anni di registrazione
11. Calcolare lo scarto interquartile degli anni di registrazione (2 modi)

possibili)

12. Tracciare un istogramma e un boxplot degli anni di registrazione; i conti tornano con i dati precedenti?
13. Visualizzare il grafico a barre dei brand con frequenza maggiore di 5000 e calcolate l'indice di Gini di questo sottoinsieme (per l'indice di Gini non usate il for)
14. Calcolare la moda dei brand
15. Calcolare la funzione cumulativa empirica degli anni di registrazione e plottarla (2 modi possibili)
16. Fare un boxplot della potenza delle auto di marca trabant; quale valore di potenza vi sembra elevato? (potete rispondere a spanne oppure usare un filtro per identificare le offerte relative)
17. Contare il numero di suv dell'Alfa Romeo; perchè non si può usare

length?

18. Trovare il modello di auto più comune

Modifica Fattori

Modificare una colonna di fattori in R non è facile come modificare un dato numerico. Per modificare un dato numerico di un dataframe basta sovrascriverlo.

```
v <- 1:10
pippo <- data.frame(mieiDati = v)
pippo$mieiDati
## [1] 1 2 3 4 5 6 7 8 9 10
#prendiamo tutte le righe della colonna "mieiDati" e li moltiplichiamo per 3
pippo[,"mieiDati" ] <- pippo[, "mieiDati"] * 3
pippo$mieiDati
## [1] 3 6 9 12 15 18 21 24 27 30
```

Ci si aspetterebbe una cosa simile anche per i fattori

Modifica Fattori (cont.)

I fattori sono liste di variabili qualitative e se si prova ad inserire un valore *non* compreso nei *levels* viene aggiunto un NA

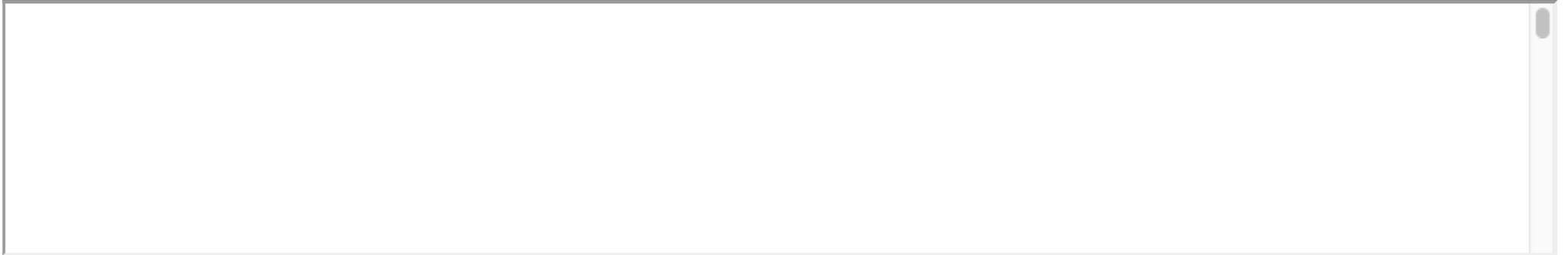
```
v <- c("pippo", "pluto", "pluto", "paperino", "paperino", "pluto", "paperino", "paperino")
personaggi <- data.frame(mieiDati = as.factor(v))
personaggi$mieiDati
## [1] pippo    pluto    pluto    paperino paperino pluto    paperino paperino
## Levels: paperino pippo pluto
# ora prendiamo tutte le righe della colonna "mieiDati" che contengono pippo e li sostituiamo con topolino
personaggi[personaggi$mieiDati == "pippo", "mieiDati"] <- "topolino"
## Warning in `[<-.factor`(`*tmp*`, iseq, value = "topolino"): invalid factor
## level, NA generated
personaggi$mieiDati
## [1] <NA>    pluto    pluto    paperino paperino pluto    paperino paperino
## Levels: paperino pippo pluto
```

Niente da fare???

Modifica Fattori (cont.)

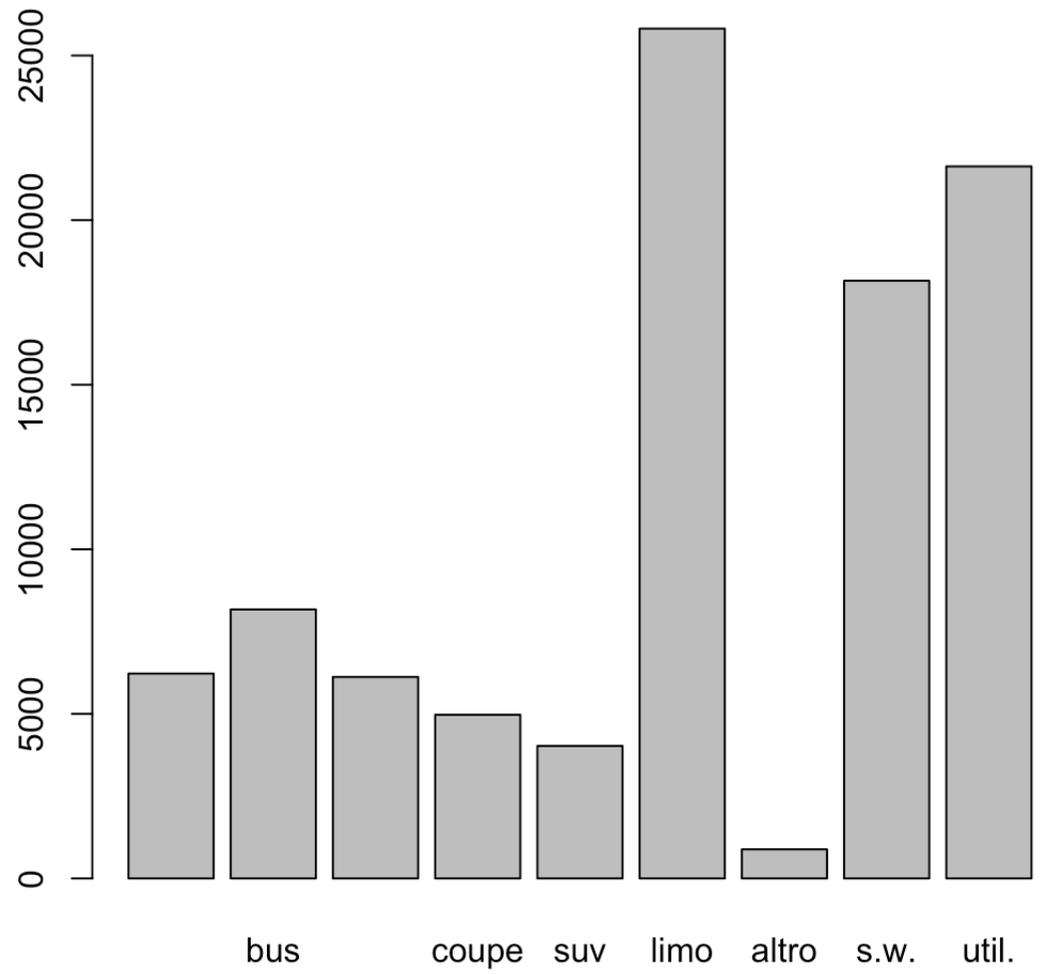
```
v <- c("pippo", "pluto", "pluto", "paperino", "paperino", "pluto", "paperino", "paperino")
personaggi <- data.frame(mieiDati = as.factor(v))
personaggi$mieiDati
## [1] pippo    pluto    pluto    paperino paperino pluto    paperino paperino
## Levels: paperino pippo pluto
# prima aggiungiamo ai levels, i nuovi possibili valori
levels(personaggi$mieiDati) <- c(levels(personaggi$mieiDati), "topolino")
# ora prendiamo tutte le righe della colonna "mieiDati" che contengono pippo e li sostituiamo con topolino
personaggi[personaggi$mieiDati == "pippo", "mieiDati"] <- "topolino"
levels(personaggi$mieiDati)
## [1] "paperino" "pippo"    "pluto"    "topolino"
unique(personaggi$mieiDati)
## [1] topolino pluto    paperino
## Levels: paperino pippo pluto topolino
# levels e unique non coincidono
# usando il comando droplevels eliminiamo dal fattore i valori che non compaiono neanche una volta
levels(personaggi$mieiDati) <- droplevels(personaggi$mieiDati)
personaggi$mieiDati
## [1] paperino pluto    pluto    topolino topolino pluto    topolino topolino
## Levels: topolino pluto paperino
```

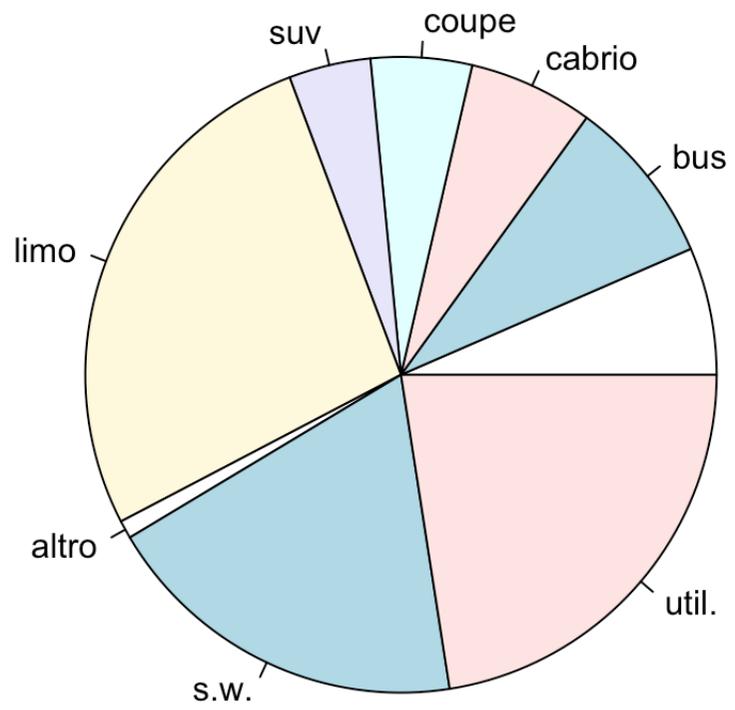
Questionario

A large, empty rectangular box with a thin black border, intended for a questionnaire. The box is positioned horizontally across the upper portion of the page. On the right side of the box, there is a small, vertical grey scrollbar, indicating that the box is designed to accommodate a scrollable list of questions or responses.

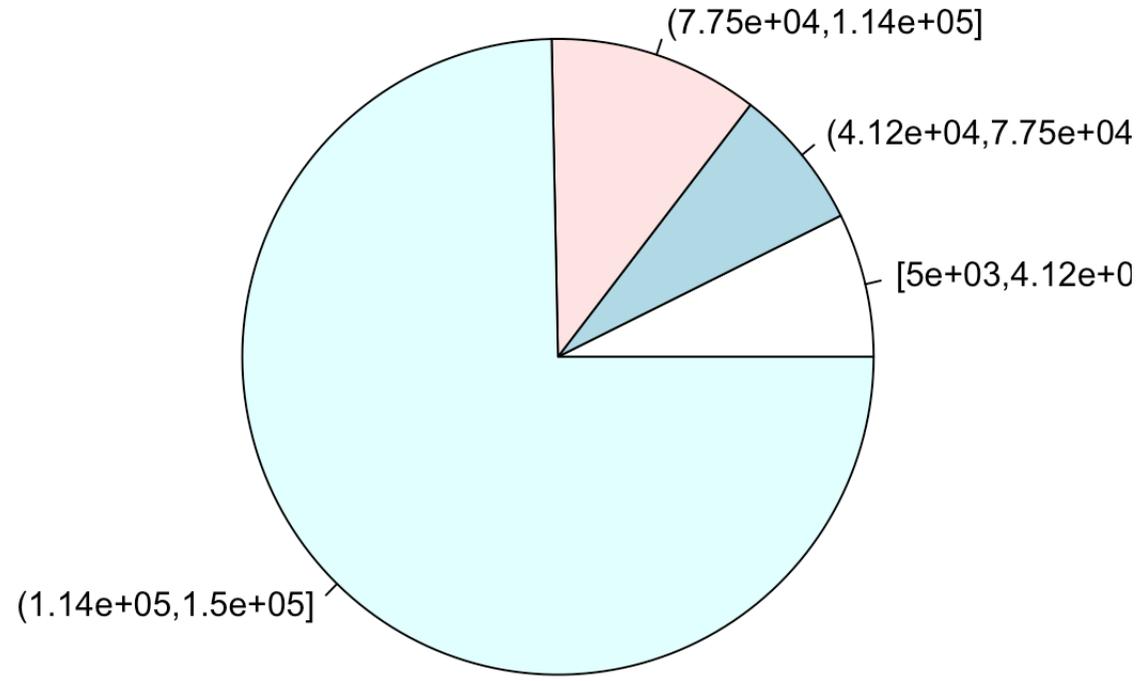
Risultati esercizio I

Usateli solo per controllare quando siete sicuri di quello che avete scritto





chilometri



Anni registrazione

