



UNIVERSITÀ
DEGLI STUDI DELLA
TUSCIA

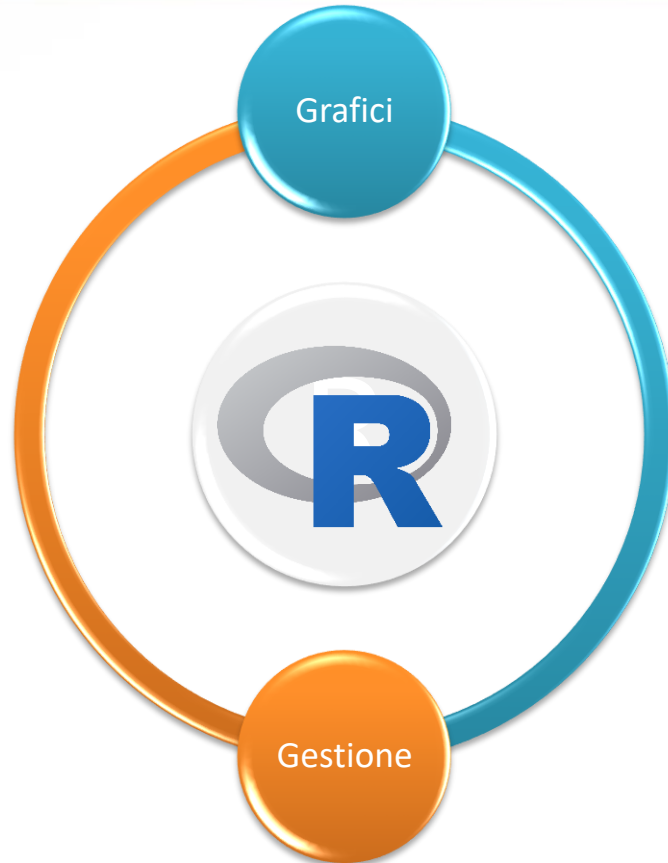
INFORMATICA

Linguaggio R
Grafici

Dott. Franco Liberati
franco.liberati@unitus.it

LINGUAGGIO R

Argomenti del corso





LINGUAGGIO R

GRAFICI

- ❑ La rappresentazione grafica scientifica è una strategia necessaria per illustrare (mediante figure, linee o segmenti, superfici o aree, solidi, simboli convenzionali) delle informazioni probabilistiche, statistiche e matematiche
- ❑ Oltre a questo consente di visualizzare immediatamente l'andamento di un fenomeno nonché avere una efficace e una globale descrizione dei dati
- ❑ In più offre la possibilità, in poco spazio, di confrontare una moltitudine di misurazioni aventi caratteristiche comuni o relazioni logiche
- ❑ Infine mette in risalto i casi 'anomali' (particolari picchi grafici), imputabili a errori nei dati o a casi da approfondire



LINGUAGGIO R

GRAFICI

- ❑ La visualizzazione dei dati non è limitata a semplici grafici a barre o a torta; nella realtà, bisogna correlare la visualizzazione più adeguata al giusto insieme di informazioni.
- ❑ Sono molti i metodi di visualizzazione disponibili per presentare i dati in modo efficace e interessante: Grafico ad area, Grafico a barre, Diagramma a scatola e baffi, Diagramma a bolle, Grafico a barre orizzontali, Cartogramma, Diagramma a torta, Mappa della distribuzione di punti, Diagramma di Gantt, Grafico heat-map, Highlight table, Istogramma, Matrice, Diagramma a rete, Area polare, Albero radiale, Grafico a dispersione (2D o 3D), Streamgraph, Tabelle di testo, Linea temporale, Diagramma ad albero, Grafico wedge stack, Nuvola di parole, Imbuto, Cascata, Azioni, Combinato



LINGUAGGIO R

GRAFICI in R

- ❑ Il Linguaggio R ha numerosi pacchetti che consentono di derivare una rappresentazione grafica dai dati analizzati (medie, varianze, correlazioni, distribuzioni e tabelle di contingenza)



Grafico a Barre



LINGUAGGIO R

GRAFICI A BARRE

- ❑ I grafici a barre, o ortogrammi, sono caratterizzati dall'avere un solo asse in scala graduata, dipendente dall'unità di misura scelta per rappresentare i dati, cioè i campioni rilevati. Sull'altro asse figurano i valori dei campioni (le modalità qualitative), equidistanti, che hanno un ordine arbitrario, se la serie è sconnessa, oppure, naturale, se la serie è rettilinea
- ❑ Il grafico a barre a colonne è una successione di colonne (segmenti verticali, rettangoli) equidistanti ed in numero pari ai campioni rilevati. La base delle colonne è spesso uguale (più raramente arbitraria). L'altezza delle colonne è proporzionale alla frequenza (assoluta o relativa) o all'intensità della modalità di rappresentazione



LINGUAGGIO R

GRAFICI A BARRE

- ❑ I grafici a barre, o ortogrammi, sono caratterizzati dall'avere un solo asse in scala graduata, dipendente dall'unità di misura scelta per rappresentare i dati, cioè i campioni rilevati. Sull'altro asse figurano i valori dei campioni (le modalità qualitative), equidistanti, che hanno un ordine arbitrario, se la serie è sconnessa, oppure, naturale, se la serie è rettilinea
- ❑ Il grafico a barre a colonne è una successione di colonne (segmenti verticali, rettangoli) equidistanti ed in numero pari ai campioni rilevati. La base delle colonne è spesso uguale (più raramente arbitraria). L'altezza delle colonne è proporzionale alla frequenza (assoluta o relativa) o all'intensità della modalità di rappresentazione



LINGUAGGIO R

GRAFICI A BARRE: disegno

❑ La funzione utilizzata nel Linguaggio R per generare un ortogramma a colonne è:

barplot(height, width = 1, space = NULL, names.arg = NULL, legend.text = NULL, beside = FALSE, horiz = FALSE, density = NULL, angle = 45, col = NULL, border = par("fg"), main = NULL, sub = NULL, xlab = NULL, ylab = NULL, xlim = NULL, ylim = NULL, xpd = TRUE, log = "", axes = TRUE, axisnames = TRUE, cex.axis = par("cex.axis"), cex.names = par("cex.axis"), ...)

dove i parametri più interessanti e usati sono **height**, che corrisponde al vettore dei campioni da rappresentare; **xlab/ylab**, ovvero l'etichetta dell'asse x/y; **main** è il titolo del grafico; **names.arg**, un vettore che riporta le etichette da apporre sotto ogni barra; e, infine, **col**, indica il colore da assegnare alle barre del grafico



LINGUAGGIO R

GRAFICI A BARRE: esempio

```
temperatura <- c(29,29,30,29,28,30,30,27,24,26,26,25,23,24) #Temperatura Roma Agosto 2022
```

```
#impostazione delle componenti di supporto del grafico
```

```
mymain<-"Temperature del mese di Agosto 2022 a Roma (primi 14 giorni)"
```

```
myxlab<-"Giorni"
```

```
myylab<-"Gradi centigradi °C"
```

```
mynamx<-as.character(1:14)
```

```
mycolr<-"royalblue1"
```

```
#Creazione di un file PNG
```

```
png(file = "C://Users/flibe/Documents/Esercizi/LIBRO/ Imm_X_Barchart_Temp_Roma_Ago_2022.png")
```

```
#Creazione del grafico
```

```
barplot(height=temperatura, col=mycolr, main=mymain, xlab=myxlab,ylab=myylab,names.arg=mynamx)
```

```
#Funzione di salvataggio del file nel formato PNG
```

```
dev.off()
```

LINGUAGGIO R

GRAFICI A BARRE: esempio- risultato

Temperature del mese di Agosto 2022 a Roma (primi 14 giorni)

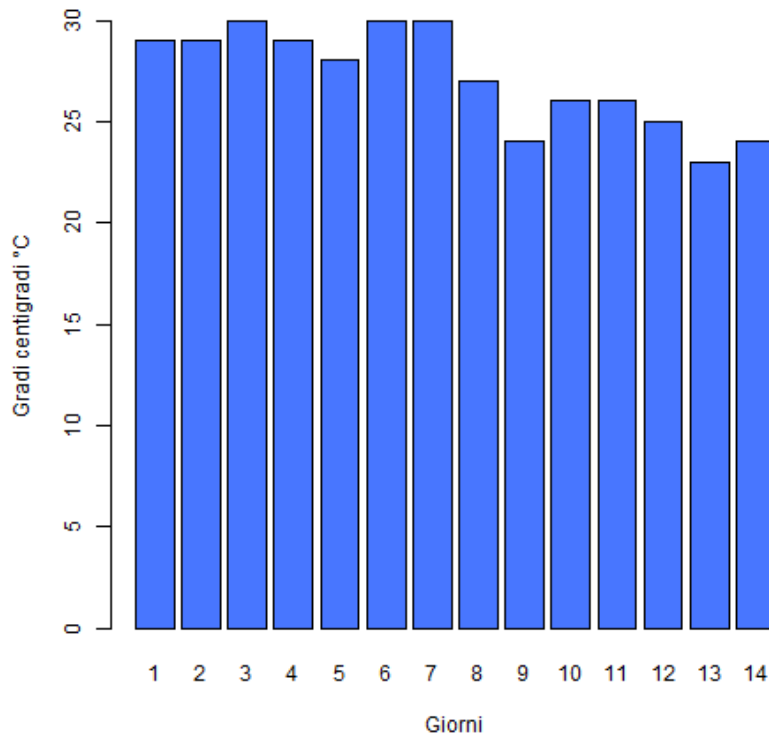




Grafico a Torta



LINGUAGGIO R

GRAFICO A TORTA

I diagrammi circolari (o aerogrammi), comunemente noti come diagrammi a torta, sono usati per rappresentare serie sconnesse o rettilinee in modo da mettere in evidenza l'importanza relativa delle singole modalità rispetto al totale.

Esistono diversi tipi: a spicchi (o settori) variabili, con angoli al centro corrispondenti alle frequenze assolute delle singole modalità e raggio fisso, e a spicchi fissi, con angoli al centro uguali e raggio variabile corrispondente alle frequenze assolute (o relative) delle singole modalità



LINGUAGGIO R

GRAFICO A TORTA: disegno

Nel Linguaggio R si usa la funzione:

```
pie(x, labels = names(x), edges = 200, radius = 0.8, clockwise = FALSE, init.angle =  
if(clockwise) 90 else 0, density = NULL, angle = 45, col = NULL, border = NULL, lty =  
NULL, main = NULL, ...)
```

dove i campi più interessanti sono x, cioè il vettore dei campioni (valori non negativi); labels, ovvero un vettore di nomi che riporta le etichette dei settori; la “torta” è disegnata all’interno di un “ipotetico” quadrato di lato 2, il radius (da -1 a 1), che va ridotto se le etichette hanno identificatori lunghi; e clockwise che imposta l’ordine di rappresentazione dei campioni (FALSE, antiorario; TRUE, orario)



LINGUAGGIO R

GRAFICO A TORTA: esempio

#Creazione di un file PNG

```
png(file = "C://Users/flibe/Documents/Esercizi/LIBRO/Imm_X_Pie_Rip_Geo_Pop_2019.png")
```

#impostazione del grafico

#Ripartizione della popolazione italiana nel 2019 per fasce geografiche (milioni di persone)

```
popolazione <- c(11,27,19,20,23)
```

#Etichette delle percentuali

```
mylabels <- c("Isole", "Nord-Ovest", "Nord-Est", "Centro", "Sud")
```

#Definizione delle componenti di supporto

```
mymain<-"Residenti in Italia (2019) per fascia geografica"
```

```
mycolr<-rainbow(length(popolazione)) #La funzione rainbow(n) genera un vettore di n colori differenti
```

#Creazione del grafico

```
pie(popolazione, labels=mylabels, main = mymain, col=mycolr)
```

#Funzione di salvataggio del file

```
dev.off()
```

LINGUAGGIO R

GRAFICO A TORTA: esempio - risultato

Residenti in Italia (2019) per fascia geografica

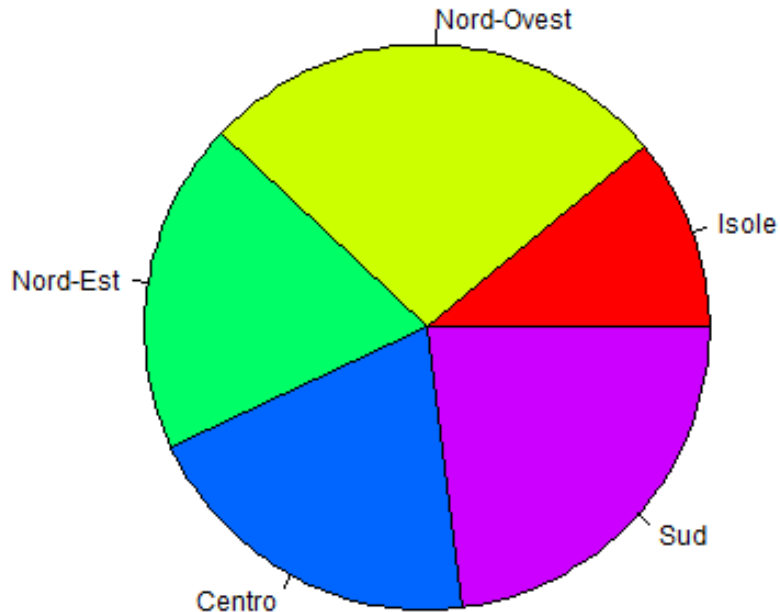




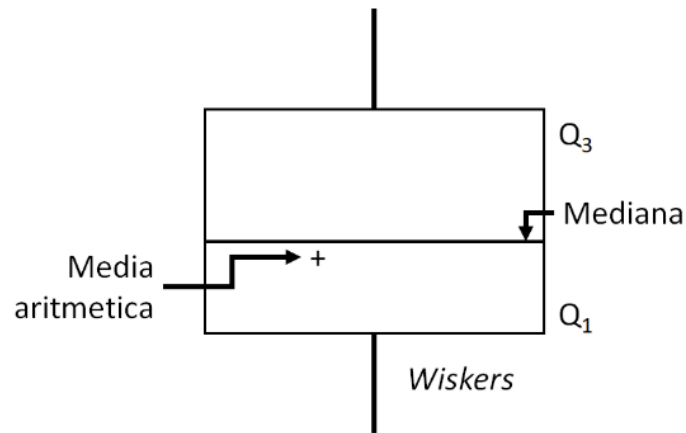
Diagramma a scatola



LINGUAGGIO R

DIAGRAMMA A SCATOLA

- ❑ Tra le rappresentazioni di tipo statistico ha un particolare rilievo la scatola (boxplot), la quale consente di visualizzare in maniera chiara e coerente le principali caratteristiche di una distribuzione statistica semplice
- ❑ Si tratta di un rettangolo in cui il lato inferiore e quello superiore indicano rispettivamente il 1° quartile (Q_1) e il 3° quartile (Q_3) della distribuzione considerata; la linea all'interno indica la posizione della mediana (2° quartile, Q_2); il simbolo “+” rappresenta la media aritmetica; la linea verticale tratteggiata (*whisker*) va da Q_3 fino al massimo e da Q_1 fino al minimo





LINGUAGGIO R

DIAGRAMMA A SCATOLA: disegno

□ Un diagramma a scatola nel Linguaggio R è ottenuto con la funzione:

boxplot(formula, data=NULL, notch=TRUE, varwidth=FALSE, ..., subset, na.action=NULL, xlab = mklab(y_var = horizontal), ylab = mklab(y_var !=horizontal), add = FALSE, ann = !add, horizontal = FALSE, drop = FALSE, sep = ".", lex.order = FALSE)

dove formula indica una funzione matematica (di solito si mettono in relazione le variabili dipendenti con quelle indipendenti; in questo caso si usa l'operatore ~); data è un segmento di dati; impostando notch a TRUE si evidenzia la tacca e varwidth se impostato a TRUE traccia una scatola proporzionale alla grandezza



LINGUAGGIO R

DIAGRAMMA A SCATOLA: esempio (dominio)

- ❑ Si applica la rappresentazione a scatola ad un segmento di dati formato da una lista di persone a cui sono associati i redditi mensili, l'aspettativa di vita e la fascia territoriale di residenza.
- ❑ Si prendono in considerazione i guadagni degli utenti che hanno la stessa aspettativa di vita. valutandone la media, la mediana e i quartili.
- ❑ Per svolgere l'analisi si usa l'operatore **tilde** $\sim$, utile per definire la relazione tra variabili dipendenti (il reddito degli utenti) e variabili indipendenti (l'aspettativa di vita)



LINGUAGGIO R

DIAGRAMMA A SCATOLA: esempio

#Persone Reddito e Aspettativa di vita e Area geografica

```
nome <- c("Marco", "Mario", "Jonathan", "Filippo", "Ugo", "Antonio", "Francesco", "Antonello", "Carlo")
```

```
reddito <- c(1260,2456,1120,1987,2750,1990,2200,1640,2340)
```

```
life_exp<-c(75,83,75,75,88,76,83,76,88)
```

```
area_geo<-c("Sud", "Nord", "Centro", "Centro", "Nord", "Nord", "Sud", "Isole", "Isolae")
```

#Creazione data frame

```
df1 <- data.frame(nome, reddito, life_exp, area_geo)
```

#Creazione dati per il grafico

```
myylab<-"Guadagno mensile"
```

```
myxlab<-"Aspettativa di vita (anni)"
```

```
mymain<-"Rapporto tra reddito e aspettativa di vita"
```

```
mycolor<-c("green", "yellow", "purple", "blue")
```

#Definizione di un file

```
png(file = "C://Users/flibe/Documents/Esercizi/LIBRO/Imm_X_Boxplot_Utenti.png")
```

#Produzione del grafico

```
boxplot(reddito ~ life_exp, data=df1, xlab=myxlab, ylab=myylab, main=mymain, col=mycolor)
```

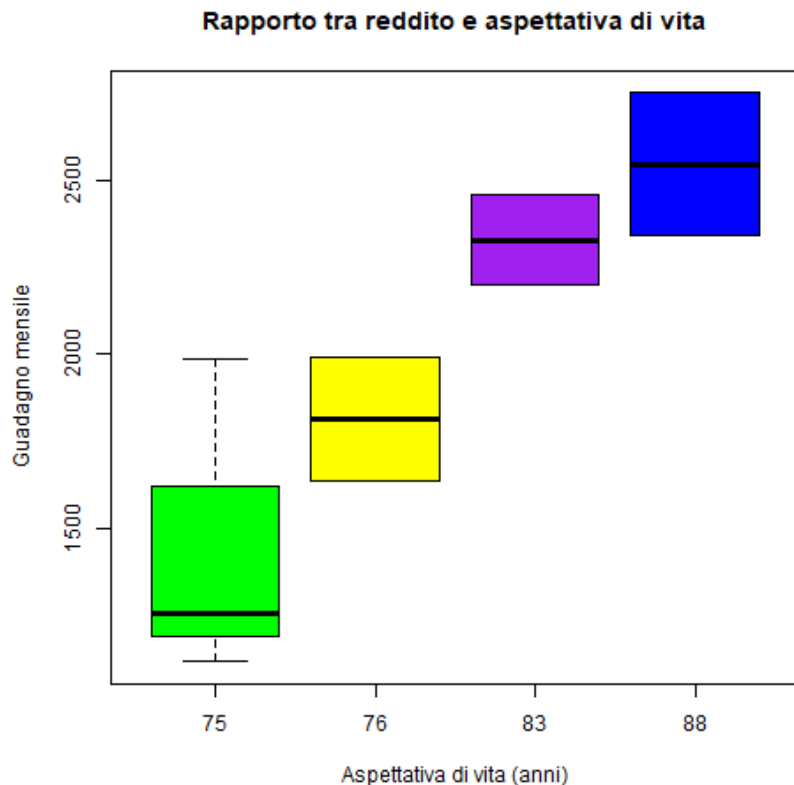
#Salvataggio del file

```
dev.off()
```

LINGUAGGIO R

DIAGRAMMA A SCATOLA: esempio

nome	reddito	life_exp	area_geo
1 Marco	1260	75	Sud
2 Mario	2456	83	Nord
3 Jonathan	1120	75	Centro
4 Filippo	1987	75	Centro
5 Ugo	2750	88	Nord
6 Antonio	1990	76	Nord
7 Francesco	2200	83	Sud
8 Antonello	1640	76	Isole
9 Carlo	2340	88	Isole



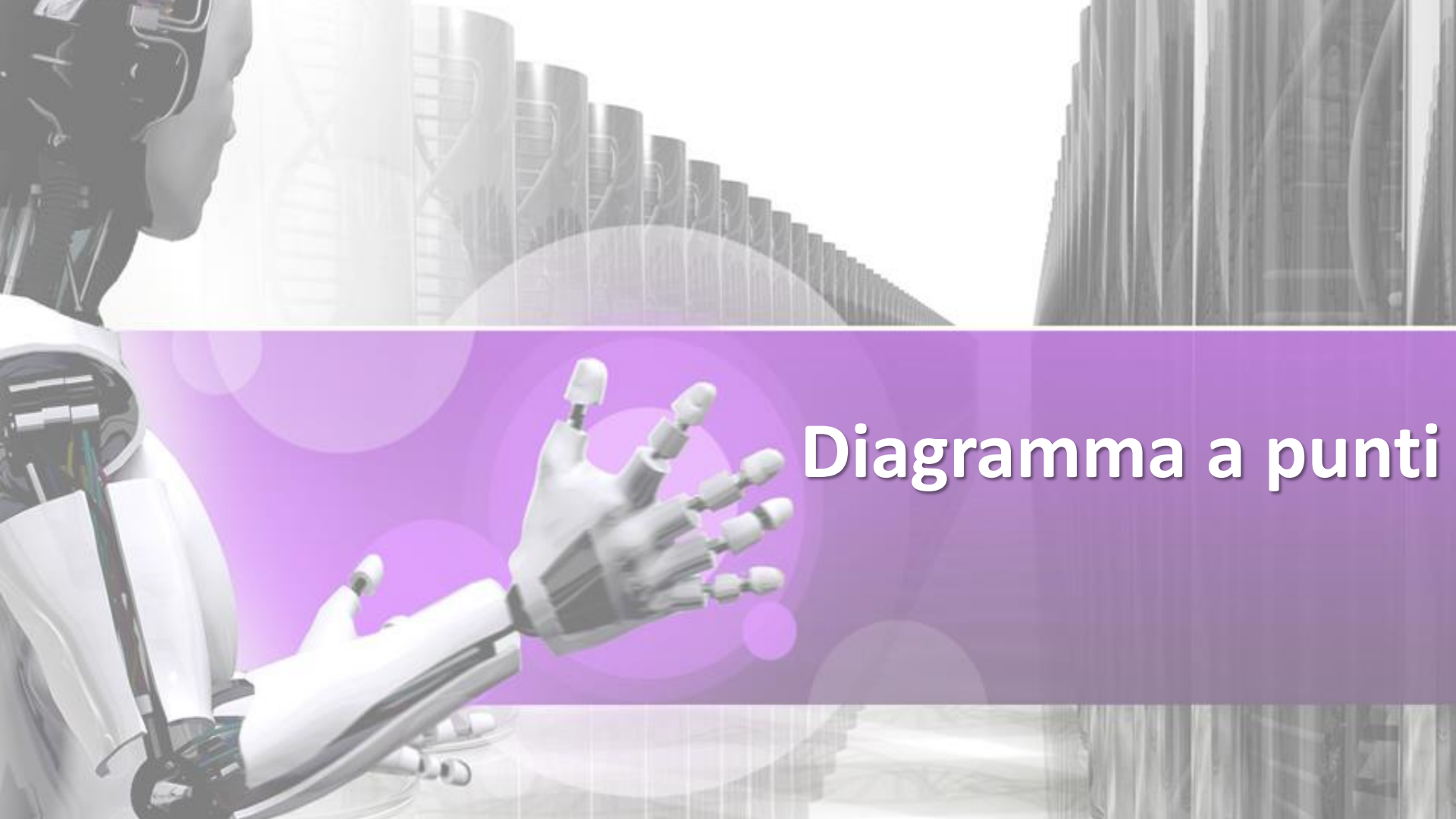


Diagramma a punti



LINGUAGGIO R

DIAGRAMMA A PUNTI

- ❑ La nuvola di punti (o scatter plot) è una rappresentazione grafica molto utilizzata nel caso di una distribuzione statistica doppia, in cui entrambi i dati rilevati sono quantitativi, perché consente di visualizzare la correlazione esistente fra le variabili
- ❑ La nuvola dei punti consente di vedere la dispersione tra le unità statistiche ossia la loro vicinanza o distanza, indicanti rispettivamente la loro somiglianza o dissomiglianza. La rappresentazione grafica, consente di analizzare la forma assunta dalla nuvola dei punti, e derivare da essa indicazioni sul tipo di relazione esistente tra le due variabili messe in relazione.
- ❑ Ad esempio, se i punti tendono a disporsi lungo una retta, le due variabili hanno una relazione di tipo lineare; mentre se la nuvola di punti assume una forma sferica, tra le due variabili non vi è un legame di tipo lineare.



LINGUAGGIO R

DIAGRAMMA A PUNTI: disegno

Nel Linguaggio R si sfrutta la funzione:

plot(x, y, type, main, sub, xlab, ylab, asp,...)

impostando type="p".



LINGUAGGIO R

DIAGRAMMA A PUNTI: esempio

Si applica la rappresentazione a nuvola di punti ad un segmento di dati formato da una lista di persone, in cui sono campionate alcune caratteristiche fisiche

Si mette in relazione l'altezza con il peso per dimostrare che c'è una correlazione di tipo lineare (è logico supporre che più lo studente è alto, minore sarà il suo peso e viceversa).



LINGUAGGIO R

DIAGRAMMA A PUNTI: esempio

```
#Universitari
```

```
nome <- c("Marco", "Mario", "Jonathan", "Filippo", "Ugo", "Antonio", "Francesco", "Antonello")
```

```
altezza <- c(1.85, 1.65, 1.76, 1.56, 1.77, 1.76, 1.65, 1.55)
```

```
eta <- c(23, 20, 21, 22, 22, 23, 25, 18)
```

```
peso <- c(84, 63, 77, 55, 78, 73, 85, 54)
```

```
area_geo <- c("Sud", "Nord", "Centro", "Centro", "Nord", "Nord", "Sud", "Isole")
```

```
media_voti <- c(23.5, 24.8, 27.9, 29.9, 30, 27.5, 24.3, 25.5)
```

```
universitari <- data.frame(nome, eta, altezza, peso, area_geo, media_voti)
```

```
#Etichette del grafico
```

```
myylab <- "Guadagno mensile"
```

```
myxlab <- "Peso"
```

```
mymain <- "Rapporto tra altezza e peso di studenti universitari"
```

```
#Scrittura dei dati
```

```
plot(x=universitari$peso, y=universitari$altezza, type="p", xlab=myxlab, ylab=myylab, main=mymain)
```

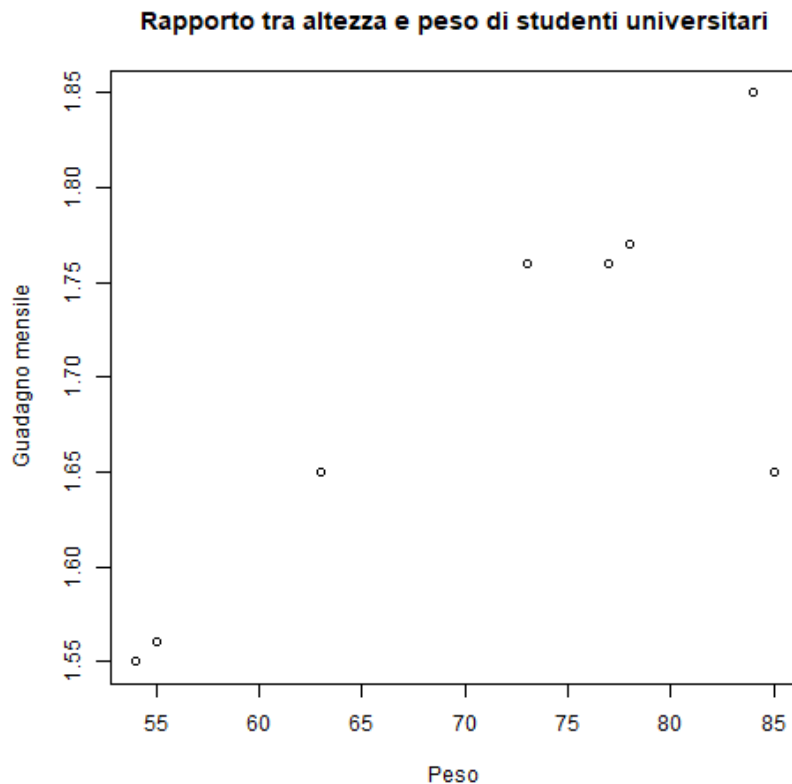
```
#Equivalente a #plot(altezza ~ peso, data=universitari, type="p", xlab=myxlab, ylab=myylab, main=mymain)
```

LINGUAGGIO R

DIAGRAMMA A PUNTI: esempio risultato

nome eta altezza peso area_geo media_voti

1	Marco	23	1.85	84	Sud	23.5
2	Mario	20	1.65	63	Nord	24.8
3	Jonathan	21	1.76	77	Centro	27.9
4	Filippo	22	1.56	55	Centro	29.9
5	Ugo	22	1.77	78	Nord	30.0
6	Antonio	23	1.76	73	Nord	27.5
7	Francesco	25	1.65	85	Sud	24.3
8	Antonello	18	1.55	54	Isole	25.5





Fine