

# Corso di Statistica Sociale

---

CORSO DI LAUREA: SCIENZE DELL'EDUCAZIONE

DOCENTE: FRANCESCO SANTELLI

# Esempi di rappresentazioni grafiche: la *Barcolana* sul Piccolo (1)



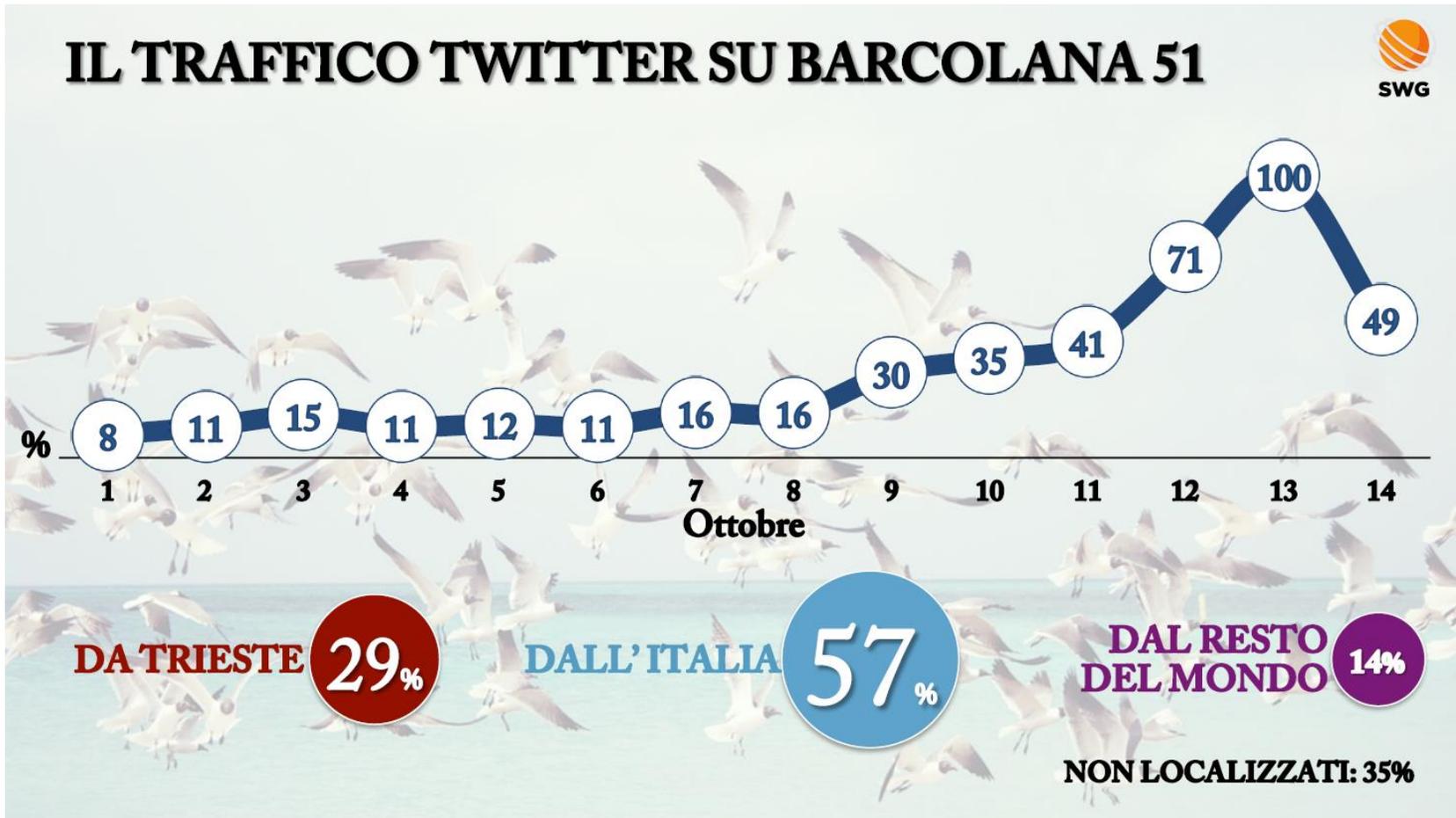
Dati di sondaggio

Ottica: campione rappresentativo

Ogni dato è percentuale che risponde «sì, conosco la Barcolana»

Effetto temporale

# Esempi di rappresentazioni grafiche: la *Barcolana* sul Piccolo (2)

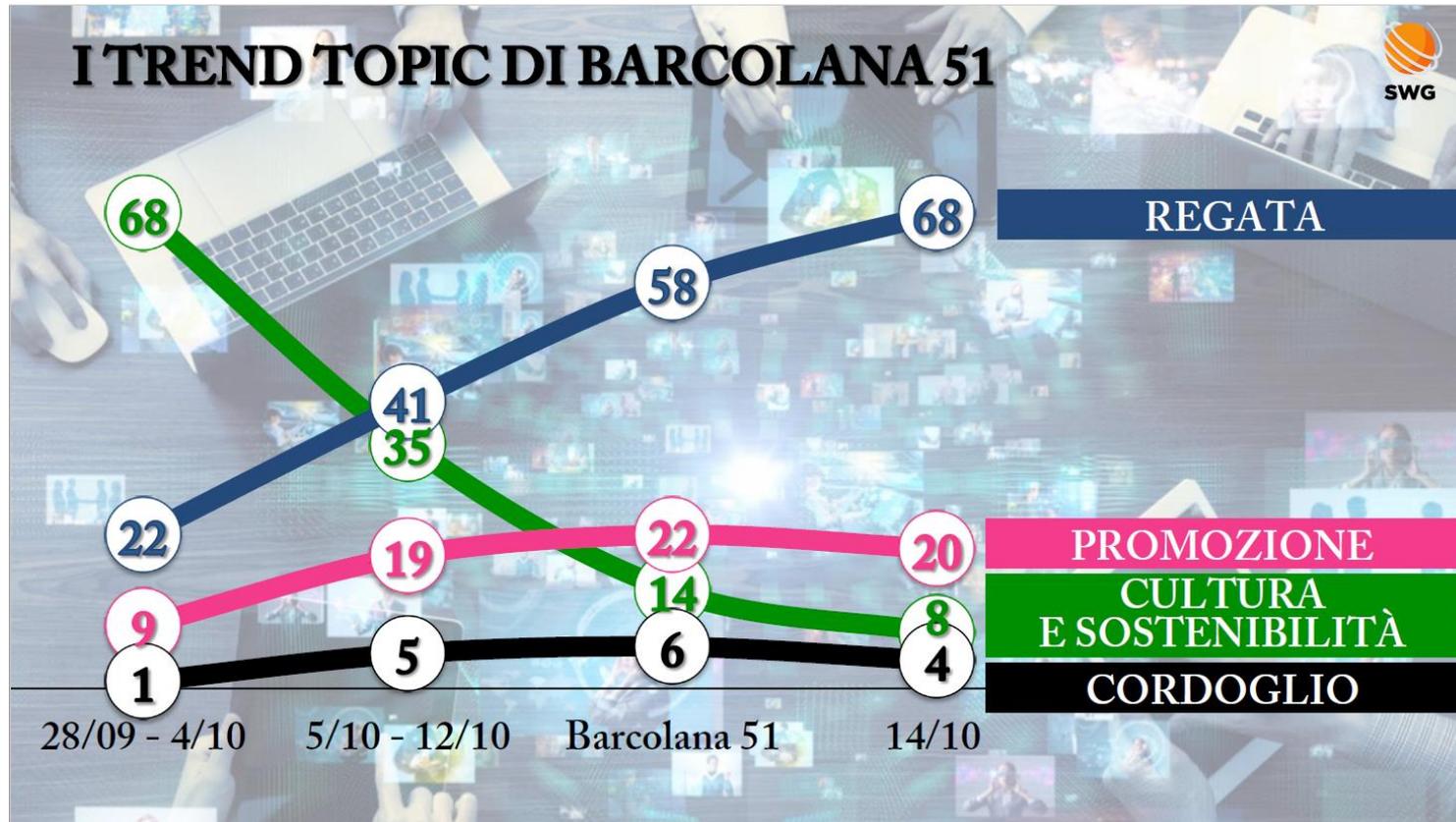


Dati provenienti da Big Data

Ottica: dati twitter, dati di piattaforma Online, non rapporto campione-popolazione

Come leggere le percentuali?

# Esempi di rappresentazioni grafiche: la *Barcolana* sul Piccolo (3)



**Classificazione:** come leggere  
Percentuali e dati temporali?

Si noti che si deve trovare una sintesi tra «precisione» statistica e obiettivi di un giornale che deve raggiungere un pubblico molto ampio

# Esempi di rappresentazioni grafiche: la *Barcolana* sul Piccolo (4)

## I TOP HASHTAG DI BARCOLANA 51



trieste	326
regata	140
vela	112
fvg	66
fvglive	52
flegophotography	51
sailing	50
triestesocial	49
barcolanachallenge	42
regatabarcolana	42
regatta	33
igersfvg	32
friuliveneziagiulia	22
igerstrieste	21
discovertrieste	20

**Wordcloud:** hashtag più frequenti  
Rappresentati con dimensione maggiore

Quale hashtag importante sembra mancare?

Questo è uno strumento utile per descrivere da un punto di vista statistico (frequenze) dei **dati testuali**

# Esempi di rappresentazioni grafiche: la *Barcolana* sul Piccolo (5)

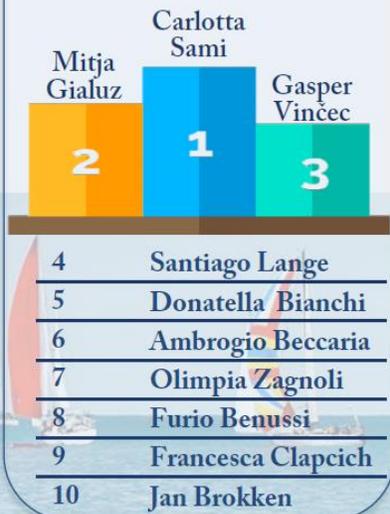
## I PROTAGONISTI DI BARCOLANA 51 SU TWITTER



### TOP TEAM



### TOP PERSONAGGI



### TOP BRAND



### TOP INFLUENCER



Classifiche di team velistici,  
Personaggi legati alla regata,  
Di brand e di influencer.

Manca la parte realmente statistica:  
Qui abbiamo solo output utile a fini  
Giornalistici, non è «importante»  
Sapere il procedimento statistico che  
Ha portato a tali classifiche

# Dati numerici in classi? Istogramma!

Abbiamo N=100 osservazioni di 100 studenti che hanno tentato l'esame di statistica ottenendo

La seguente distribuzione di punteggi:

<u><b>Voti</b></u>	<u><b>F</b></u>
0-10	18
11-16	25
17-19	12
20-22	15
23-26	15
27-30	15

Per le  $X_i$  (modalità) non abbiamo un solo valore ma un range di valore  
Per questo scrivere i dati così vuol dire avere **dati in classi**.

Abbiamo *perso informazione originaria*:  
non sappiamo i 18 nella prima classe se al compito hanno preso 0 o 1 o 2..o 10.  
Sappiamo solo che sono in quel **range**!

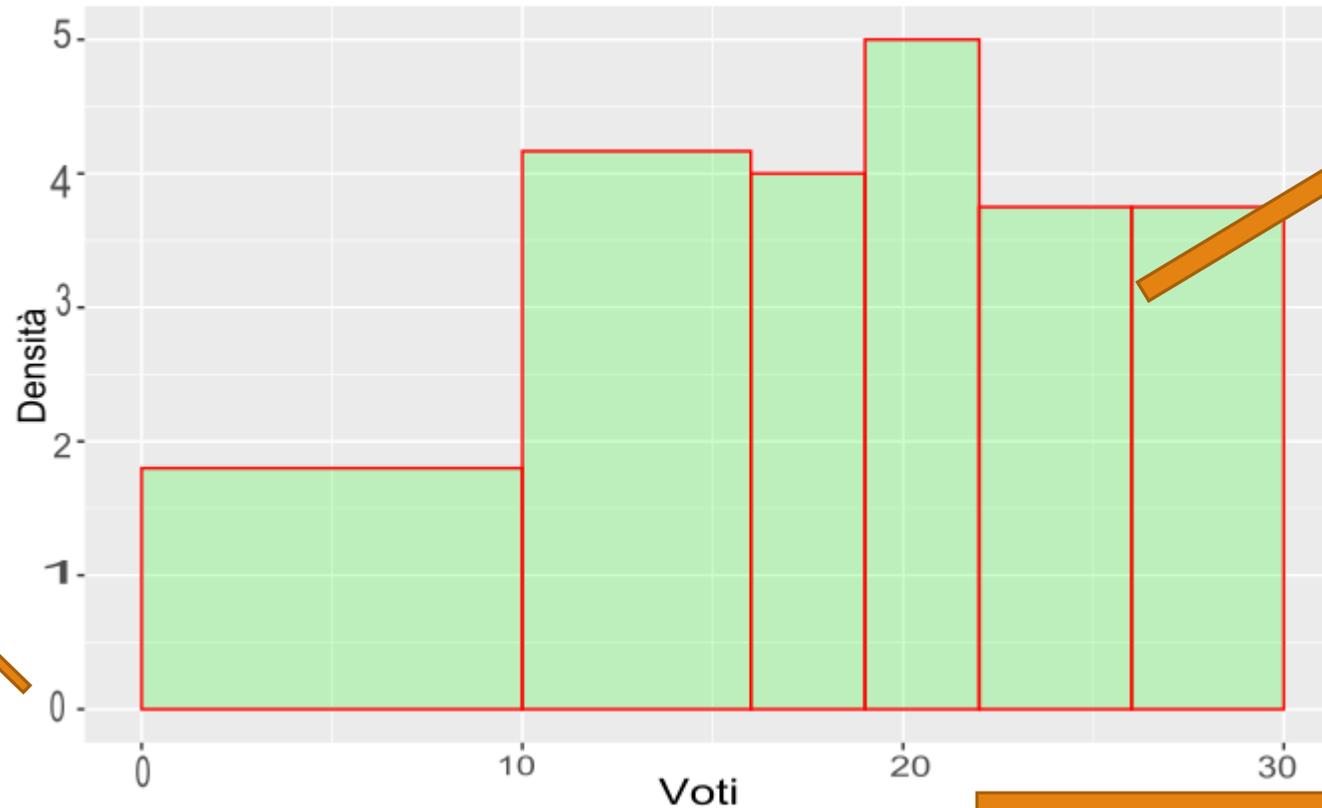
1) Quanti studenti Sono stati promossi?

2) Quanti studenti hanno Preso almeno 23?

3) Quanti studenti hanno Preso tra 20 e 26?

# Costruire un istogramma

Istogramma per voti in statistica



Asse y, nuova quantità:  
Densità di frequenza

Area di ogni rettangolino:  
La frequenza, cioè quanti  
Studenti ci sono dentro

Asse X: i voti, le classi

# Densità di frequenza

Come ogni rettangolo, ogni rettangolino dell'istogramma ha una sua area che è pari a:

$$\underline{\text{Area} = \text{Base} \times \text{Altezza}}$$

La base è data dalle classi: 0-10, 11-16 ecc.

L'altezza è data dalla densità di frequenza: 1,8, 4,2 ecc.

Si trova dividendo la frequenza F per ampiezza classe

L'area è uguale alla frequenza

$$\underline{\text{Frequenza} = \text{ampiezza classe} \times \text{densità frequenza}}$$

<u>Voti</u>	<u>F</u>	Ampiezza	Densità
0-10	18	10-0=10	18/10=1,8
11-16	25		
17-19	12		
20-22	15		
23-26	15		
27-30	15		

