



Università
di Genova

DIMA DIPARTIMENTO
DI MATEMATICA



DIPARTIMENTO
DI ECCELLENZA

MUR

Trasmettere informazioni in modo chiaro (e bello)

...per interpretare il mondo che ci circonda



Chi siamo

Sara Sommariva

- Liceo Scientifico
- Laurea magistrale e dottorato in Matematica e Applicazioni (Genova)
- Ricercatrice (Genova-Helsinki)
- Consulente Data Scientist presso Roche S.p.A e fondatrice di una Start-up



Francesco Porro

- Liceo Scientifico
- Laurea in matematica (Pavia)
- Consulente aziendale
- Dottorato in Statistica Metodologica ed Applicata (Milano)
- Ricercatore (Sassari – Genova)



Visualizzare i dati: perchè?

	week	category	weekly_rank	show_title	season_title	weekly_hours_viewed	runtime	weekly_views	cumulative_weeks_in_top_10
0	2024-01-07	Films (English)	1	The Equalizer 3	NaN	26800000	1.8167	14800000.0	1
1	2024-01-07	Films (English)	2	Rebel Moon ? Part One: A Child of Fire	NaN	25100000	2.2667	11100000.0	3
2	2024-01-07	Films (English)	3	Leave the World Behind	NaN	18700000	2.3667	7900000.0	5
3	2024-01-07	Films (English)	4	Exodus: Gods and Kings	NaN	18600000	2.5000	7400000.0	1
4	2024-01-07	Films (English)	5	Aquaman	NaN	16800000	2.3833	7000000.0	1
...
5275	2021-07-04	TV (Non-English)	6	Elite	Elite: Season 1	10530000	NaN	NaN	1
5276	2021-07-04	TV (Non-English)	7	Elite	Elite: Season 3	10200000	NaN	NaN	1
5277	2021-07-04	TV (Non-English)	8	Elite	Elite: Season 2	10140000	NaN	NaN	1
5278	2021-07-04	TV (Non-English)	9	Katla	Katla: Season 1	9190000	NaN	NaN	1
5279	2021-07-04	TV (Non-English)	10	Record of Ragnarok	Record of Ragnarok: Season 1	9140000	NaN	NaN	1

Dati settimanali riguardo alle visualizzazioni su netflix.

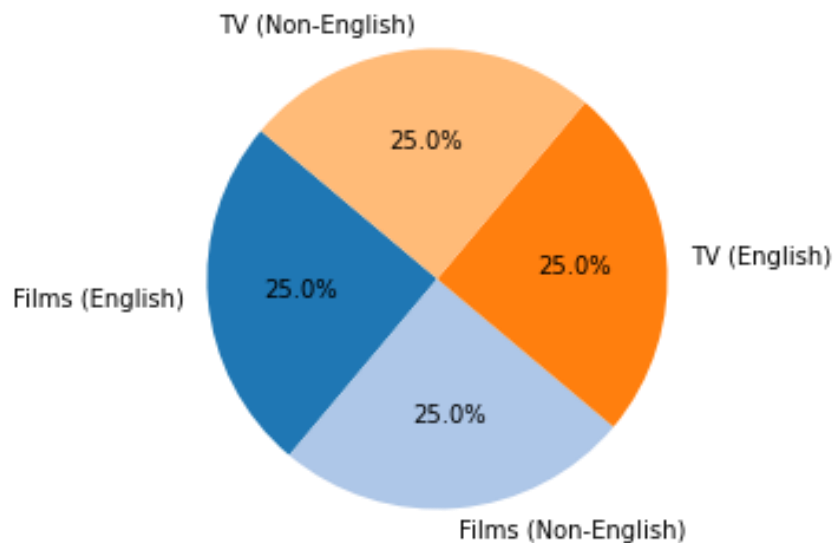
Fonte: kaggle.

Quali metodi di visualizzazione conoscete?

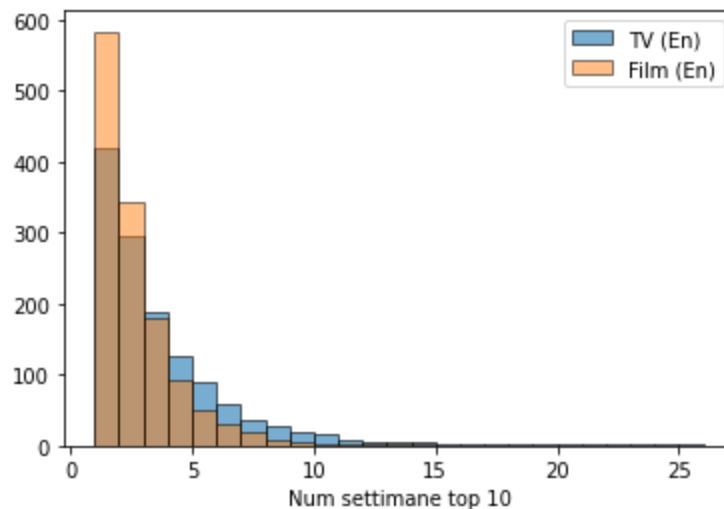
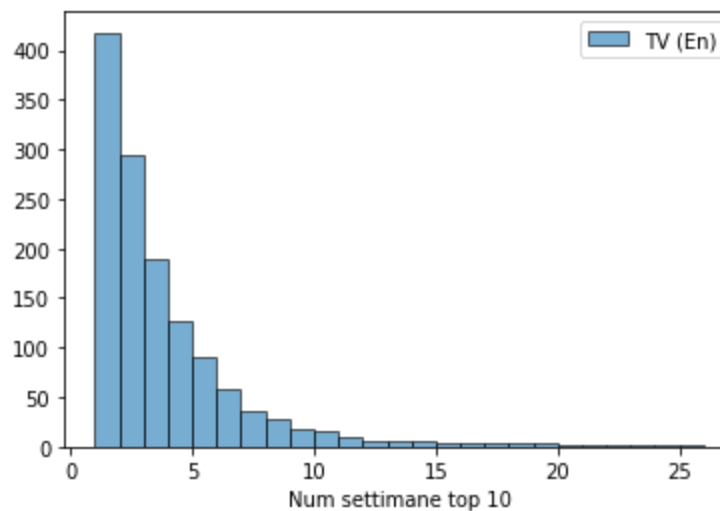


Visualizzare i dati: come?

Areogrammi (o torte)

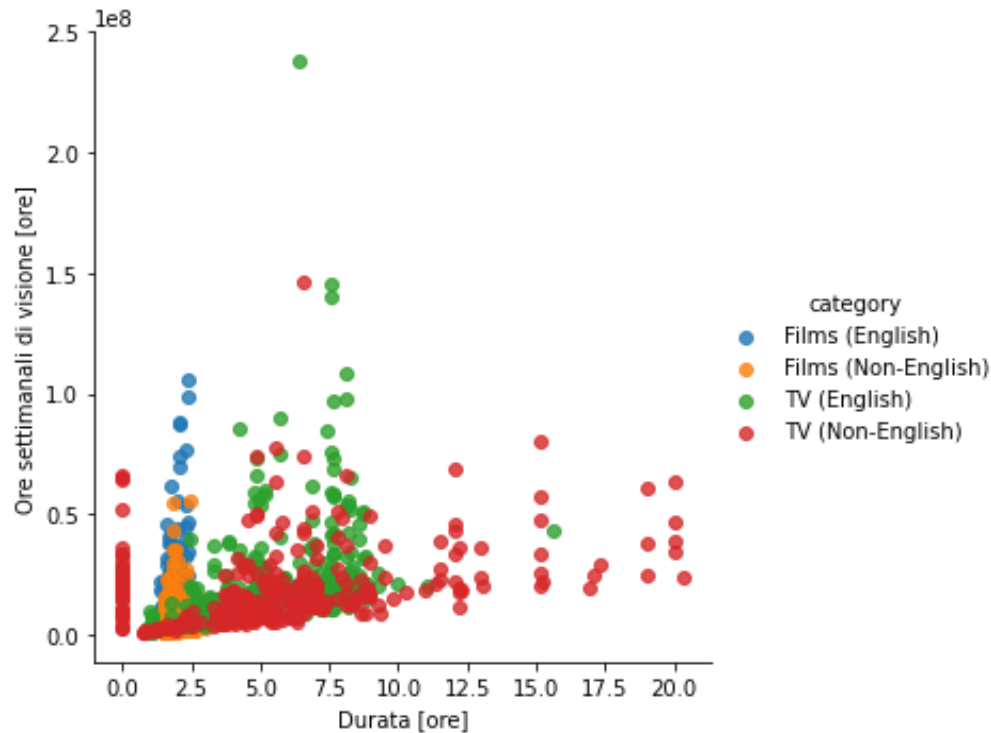


Istogrammi



Visualizzare i dati: come?

Scatterplot



Grafici fuorvianti

Tutto è relativo

	LAUREA 5	LAUREA 3	DIPL_SUP	MEDIA_IN	ELEMENT SENZA TIT.	Totale
Piemonte	29667	25771	99670	55332	4922	215362
ValleAosta	901	830	2887	1717	136	6471
Liguria	10156	8554	31695	15117	1707	67229
Lombardia	74943	66007	227809	123538	13380	505677
TrentinoAA	7837	7389	30264	11231	967	57688
Veneto	33917	31820	125964	52799	5840	250340
FriuliVGiulia	8554	7944	29827	11206	1069	58600
EmiliaRom	31723	675	99259	49541	5764	214962
Toscana	22118	23569	87228	43180	4426	180521
Umbria	6663	7888	24521	8712	972	48756
Marche	11851	13269	39701	16571	1668	83060
Lazio	43629	39188	153357	56287	5259	297720
Abruzzo	10938	11045	39362	13849	1262	76456
Molise	3020	5716	9065	3426	264	21491
Campania	42703	35330	177610	103864	13053	372560
Puglia	29419	25107	116477	68195	6071	245269
Basilicata	4774	6522	18746	6961	407	37410
Calabria	16962	16403	63977	27485	3947	128774
Sicilia	28244	30862	148748	90511	11801	310166
Sardegna	8952	9644	45709	29575	437	94317
Totale	426971	401533	1571876	789097	83352	3272829

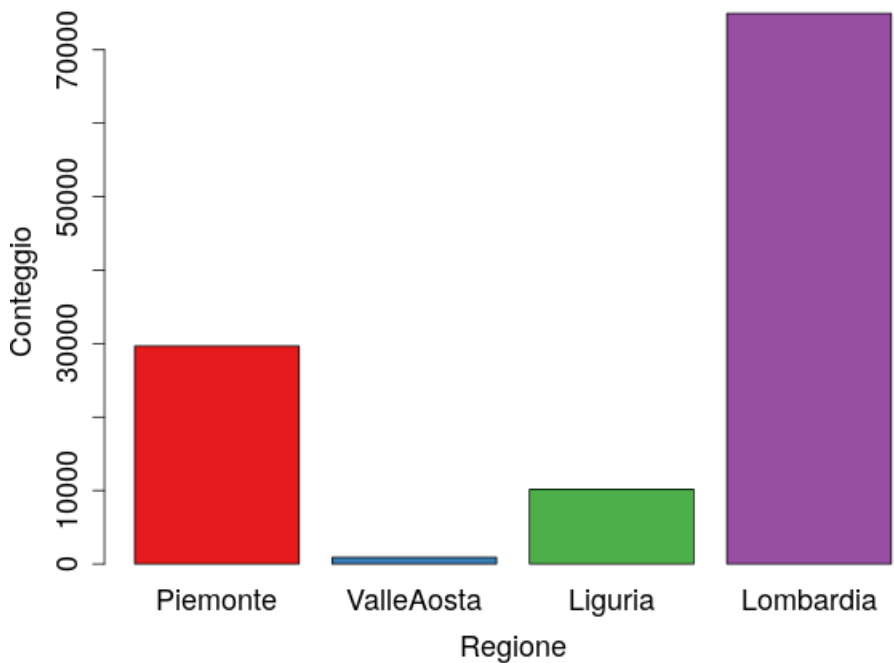
Ci sono più laureati in:

- A) Piemonte
- B) Valle d'Aosta
- C) Liguria
- D) Lombardia?

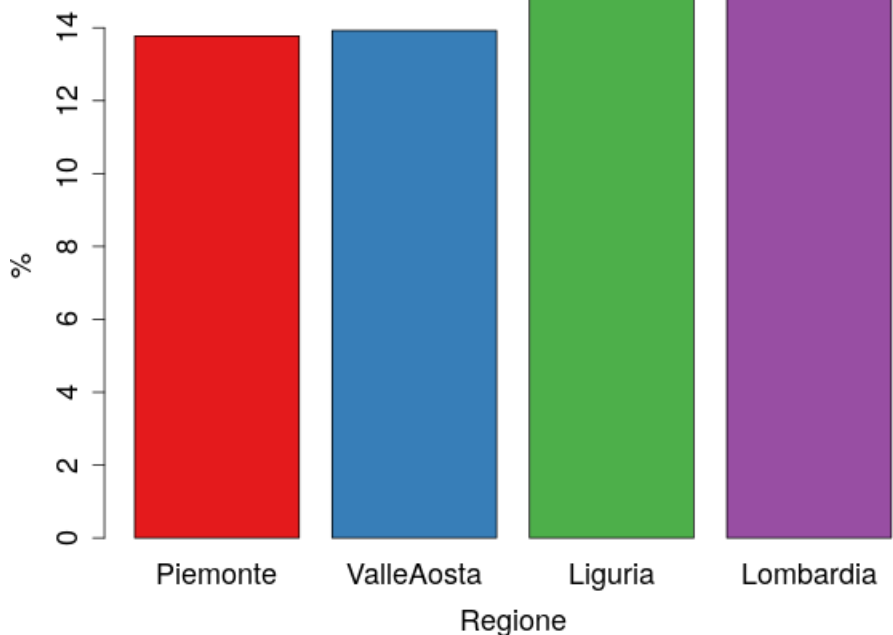


dati da moltiplicare per 10

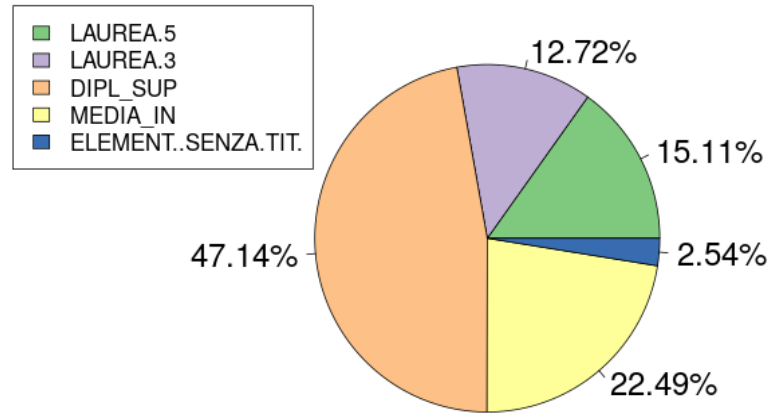
Laurea magistrale



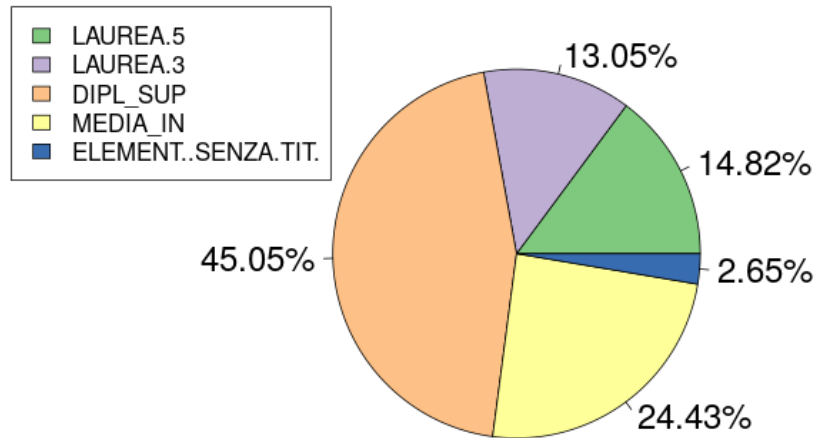
Laurea Magistrale



Liguria

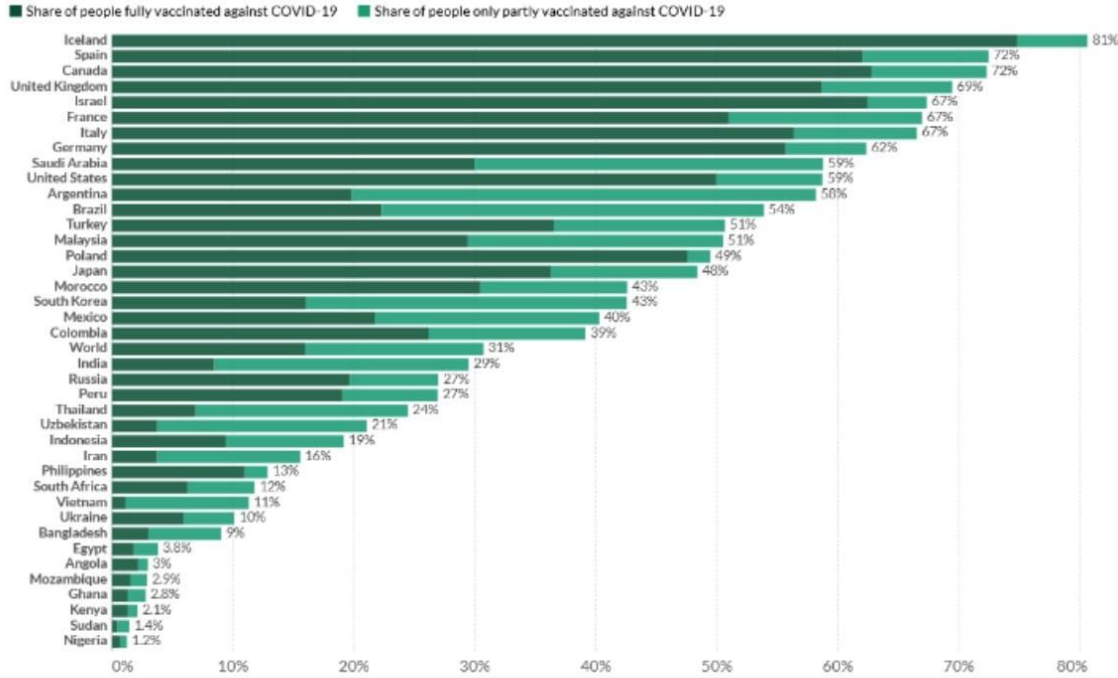


Lombardia



Share of people vaccinated against COVID-19, Aug 11, 2021

Our World in Data



Islanda: 81% vaccinati
Nigeria: 1.2% vaccinati

Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World in Data

LINEAR LOG



Islanda: 323.60 casi
Nigeria: 2.70 casi

L'islanda ha 120 volte più casi!!!!



Bello, ma non troppo

Provenienza delle persone di 6 anni o più che nell'ultimo anno (2018) hanno visitato almeno una volta siti archeologici e monumenti.



Ci sono più turisti provenienti dal Centro o dal Nord-est?

■ Nord-ovest ■ Nord-est ■ Centro
■ Sud ■ Isole

Fonte: Istat

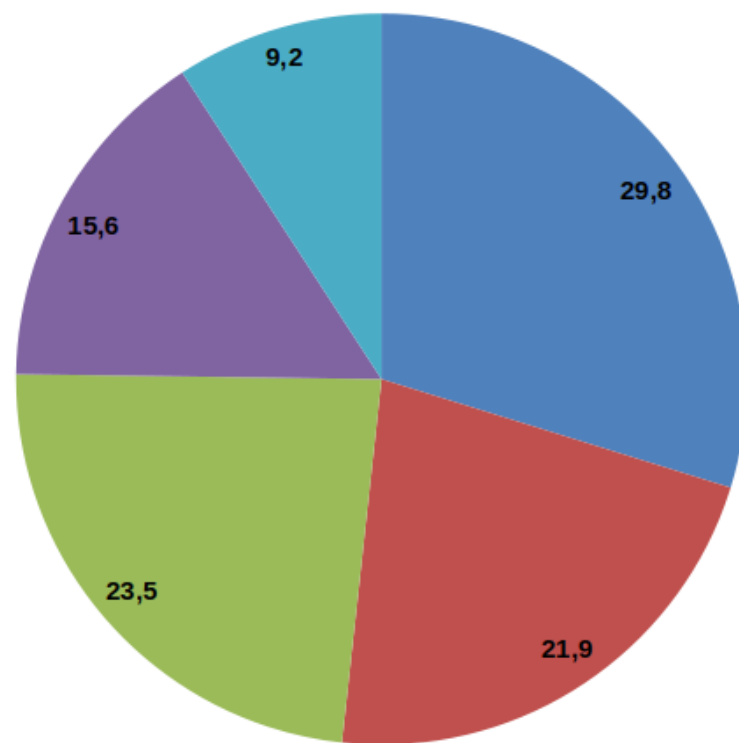


Bello, ma non troppo

Provenienza delle persone di 6 anni o più che nell'ultimo anno (2018) hanno visitato almeno una volta siti archeologici e monumenti.



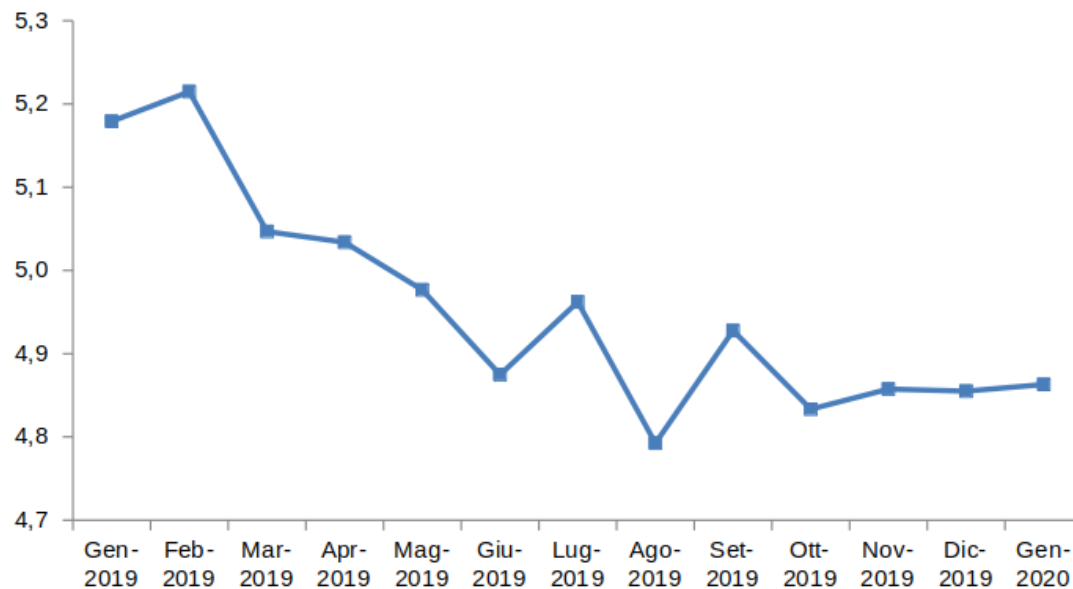
■ Nord-ovest ■ Nord-est ■ Centro
■ Sud ■ Isole



■ Nord-ovest ■ Nord-est ■ Centro ■ Sud ■ Isole

Gestione della disoccupazione

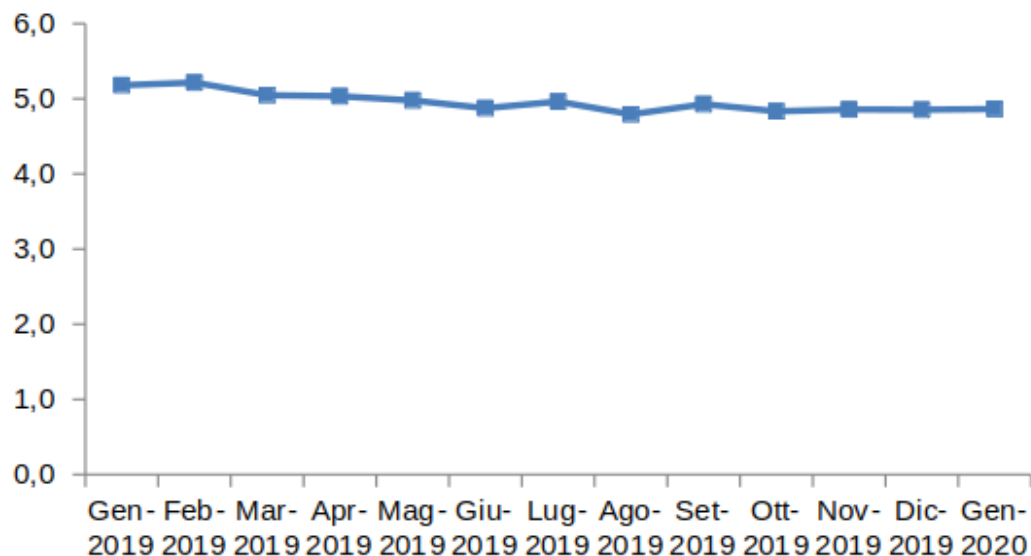
Valore percentuale del tasso di disoccupazione tra Gennaio 2019 e Gennaio 2020.
Dati destagionalizzati



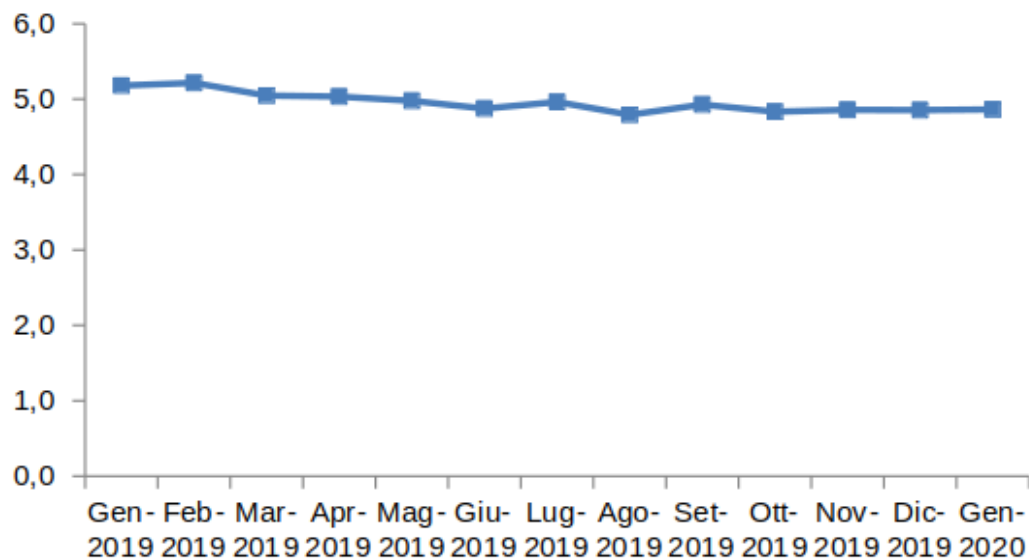
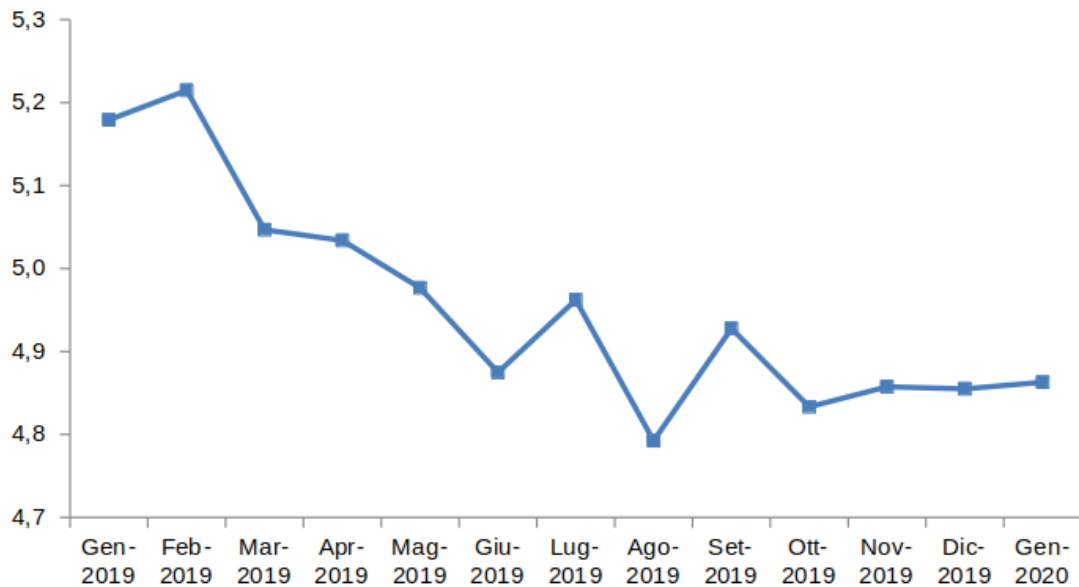
Come è variato durante il 2019
il tasso di disoccupazione?

Gestione della disoccupazione

Valore percentuale del tasso di disoccupazione tra Gennaio 2019 e Gennaio 2020.
Dati destagionalizzati



Siete sicuri?



Fonte: Istat



Ma quindi?

Una buona rappresentazione grafica è:

Sintetica. Permette in poco spazio di sintetizzare i dati e confrontare più variabili

Semplice. E' facilmente comprensibile anche a non esperti

Completa. Una buona rappresentazione grafica deve contenere:

- Un titolo o un caption
- Una breve descrizione delle variabili rappresentate
- L'unità di misura usata per graduare gli assi
- Il tipo di valore rappresentato (assoluto/percentuale)
- ...

**Una buona rappresentazione grafica
deve trasmettere un messaggio.**



Referenze e materiale

- Pacchetti didattici istat
(<https://www.istat.it/it/informazioni-e-servizi/per-studenti-e-docenti/pacchetti-didattici>)
- Gapminder (<https://www.gapminder.org/>)

Dalla raccolta dati alla visualizzazione

Fase 1: Raccolta dati

Fase 2: Scegliamo il tipo di grafico

Fase 3: Proviamo





Università
di Genova

DIMA DIPARTIMENTO
DI MATEMATICA



DIPARTIMENTO
DI ECCELLENZA

MUR

**Trasmettere informazioni in
modo chiaro (e bello)**

**...per interpretare il mondo
che ci circonda**



La soglia dei gruppi

Due gruppi:

G1: 6 sampdoriansi

G2: 5 genoani

“I genoani sono tutti più furbi dei sampdoriansi”

Disegnati in ordine di furbizia:



La soglia dei gruppi

Due gruppi:

G1: 6 sampdoriansi

G2: 5 genoani

“I genoani sono tutti più furbi dei sampdoriansi”

Disegnati in ordine di furbizia:



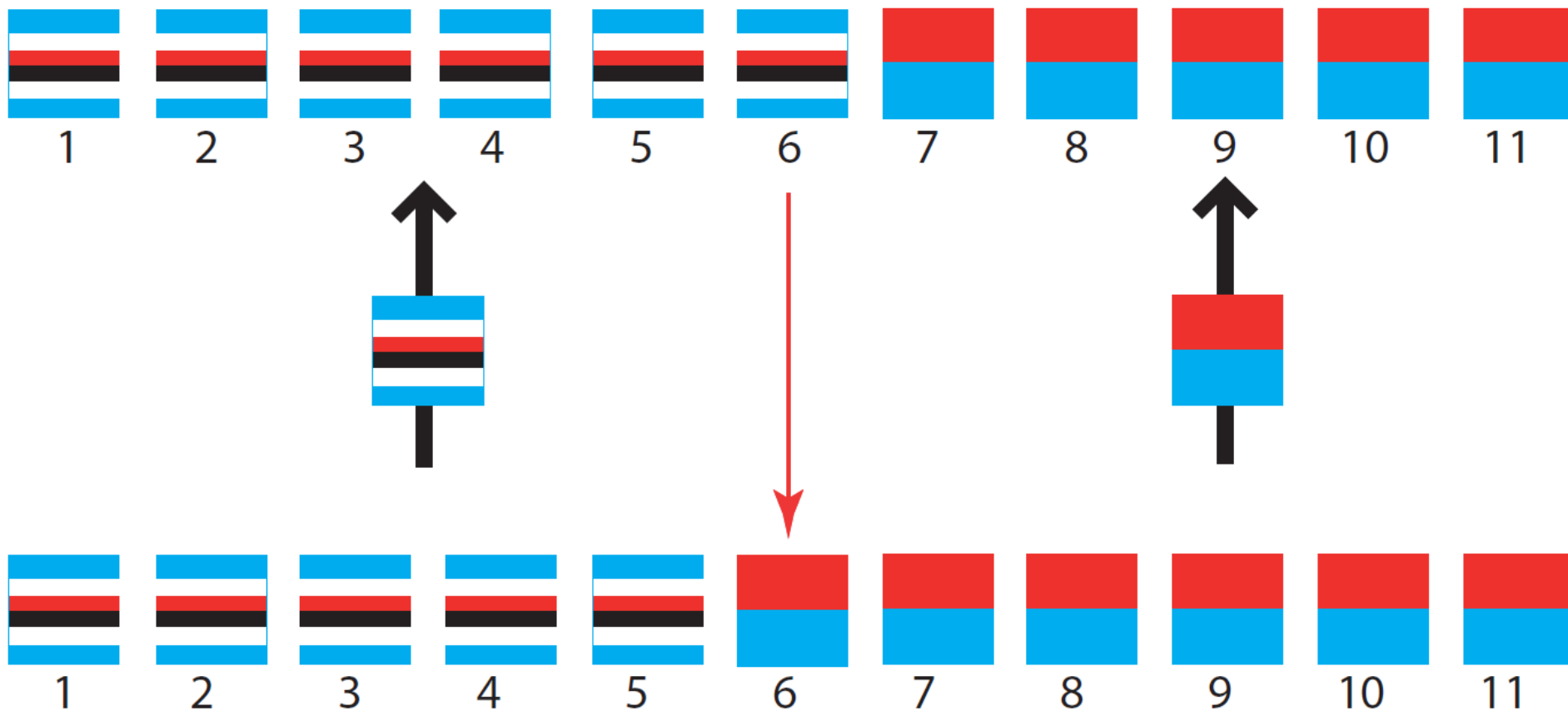
Media dei sampdoriansi



Media dei genoani

La soglia dei gruppi

Se il più furbo dei sampdoriansi diventa genoano:



Cosa succede alle medie della furbizia dei due gruppi?

La soglia dei gruppi

Se il più furbo dei sampdoriansi diventa genoano,

Cosa succede alle medie della furbizia dei due gruppi?



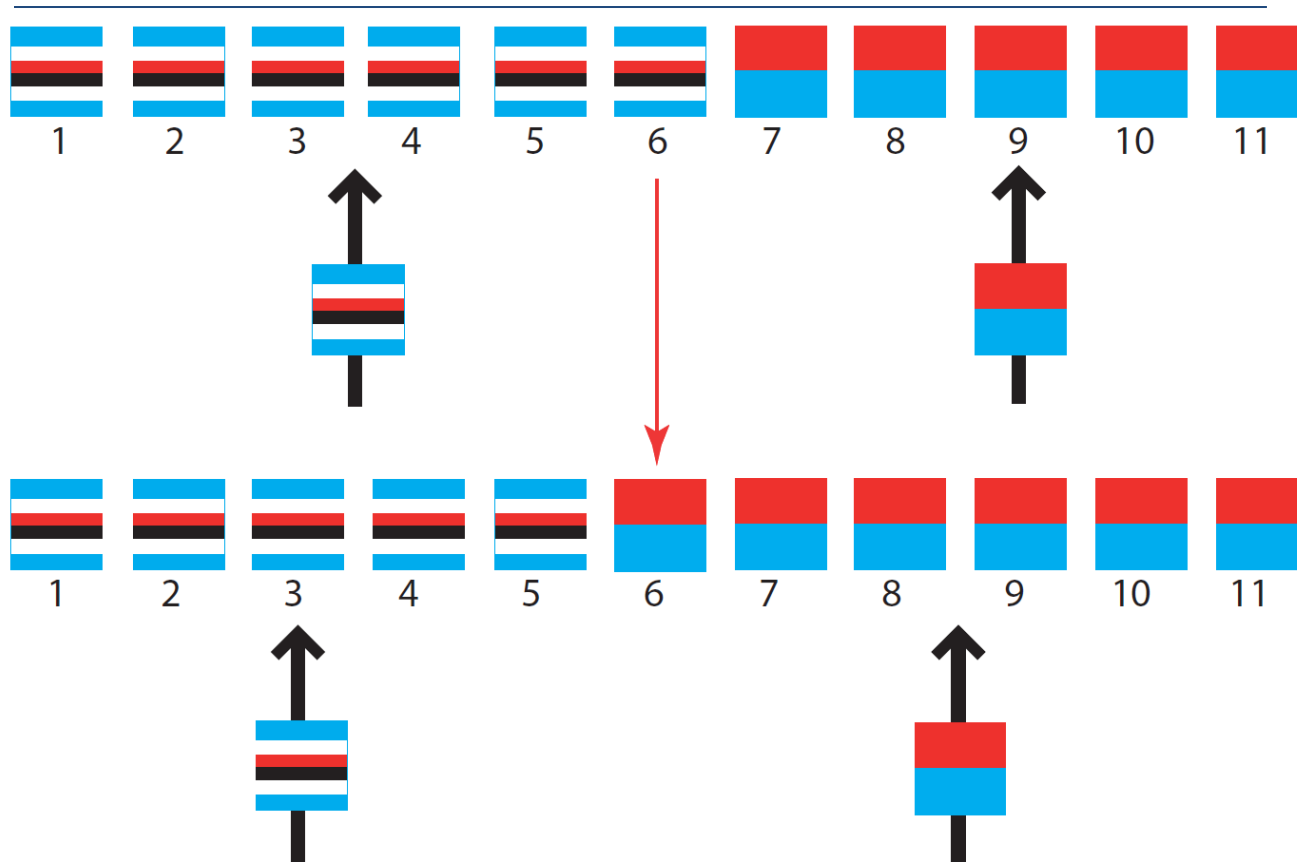
La soglia dei gruppi

Se il più furbo dei sampdoriansi diventa genoano,

Cosa succede alle medie della furbizia dei due gruppi?



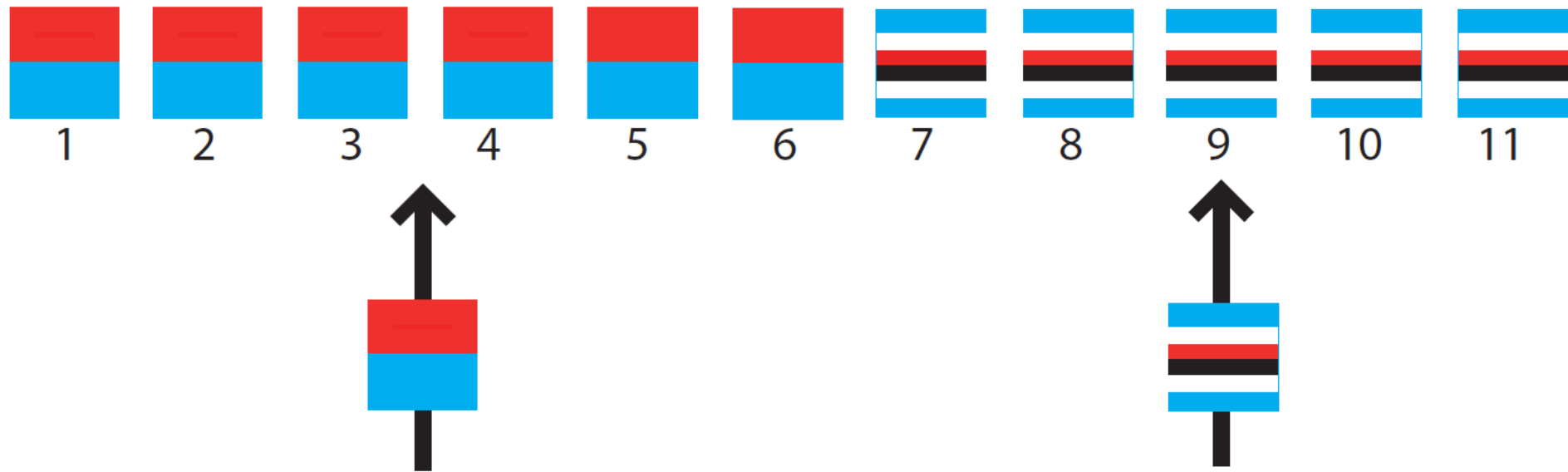
La soglia dei gruppi



...abbassa la media di entrambi i gruppi!!!

La soglia dei gruppi

... forse preferite questo?



Paradosso di Simpson



Il paradosso di Simpson indica una situazione in cui una relazione tra due fenomeni appare modificata, o perfino invertita, dai dati in possesso a causa di altri fenomeni non presi in considerazione nell'analisi (variabili nascoste).

Paradosso di Simpson



Il paradosso di Simpson indica una situazione in cui una relazione tra due fenomeni appare modificata, o perfino invertita, dai dati in possesso a causa di altri fenomeni non presi in considerazione nell'analisi (variabili nascoste).

Ma... cosa significa?

Smid

Paradosso di Simpson



200
✓ 104

Sì: 52%

200
✓ 80

Sì: 40%



Smid

Paradosso di Simpson



200
✓ 104

Sì: 52%



200
✓ 80

Sì: 40%

Paradosso di Simpson



Discriminazione di genere?



200
✓ 104









Sì: 52%












200
✓ 80

Sì: 40%










Paradosso di Simpson

	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%










Paradosso di Simpson

	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30% 
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%









Paradosso di Simpson

	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30% 
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%









Paradosso di Simpson

	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80% 
Fisica	80	20	48	16	60%	80%

Paradosso di Simpson









	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%

Paradosso di Simpson

	 Domande #		 Accett. 			 % 
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%

Il tasso di accettazione delle ragazze è in tutti e 4 i corsi **maggiore** di quello dei ragazzi!









Paradosso di Simpson

	 Domande #		 Accett. 			 % 
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%

Il tasso di accettazione delle ragazze è in tutti e 4 i corsi **maggiore** di quello dei ragazzi!

Discriminazione di genere?

Paradosso di Simpson

	 Domande #		 Accett. 		 % 	
Scienze Formazione	20	80	4	24	20%	30%
Infermieristica	20	80	4	24	20%	30%
Ingegneria	80	20	48	16	60%	80%
Fisica	80	20	48	16	60%	80%

Il tasso di accettazione delle ragazze è in tutti e 4 i corsi **maggiore** di quello dei ragazzi!

Discriminazione di genere? NO



Paradosso di Simpson

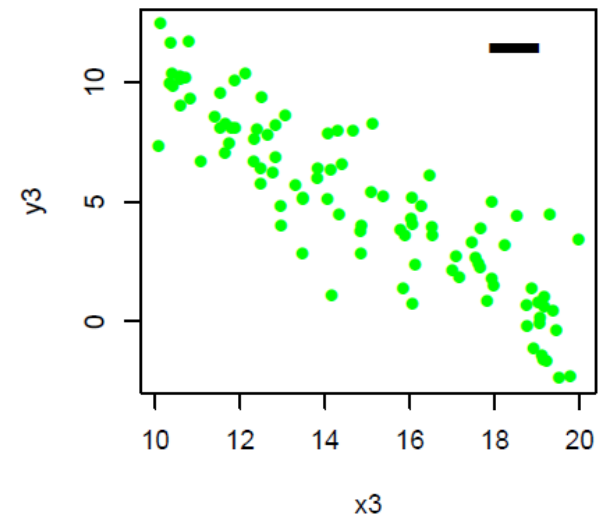
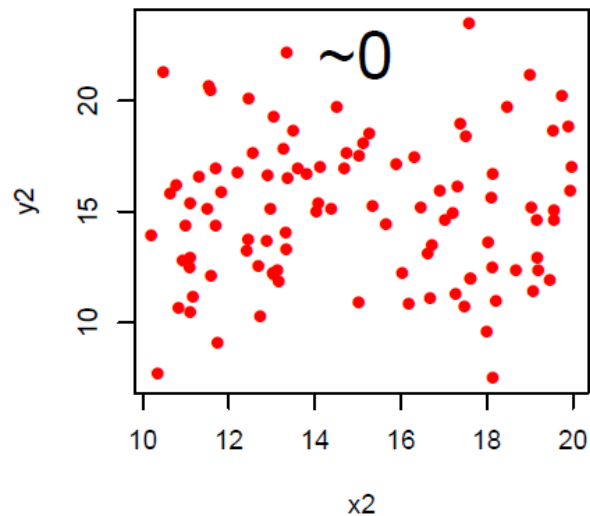
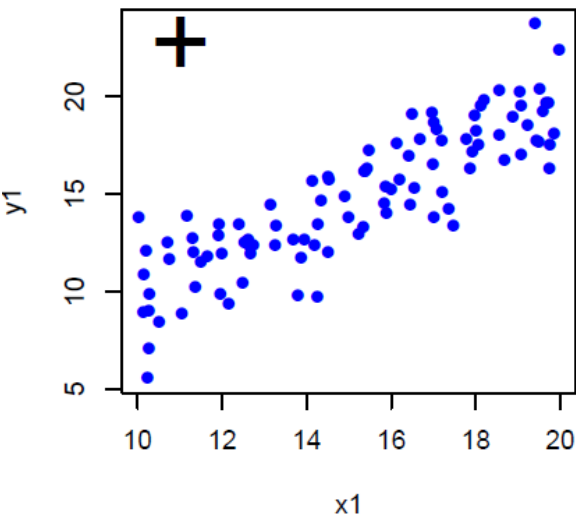


Il paradosso di Simpson indica una situazione in cui una relazione tra due fenomeni (**genere e accettazione alla scuola**) appare modificata, o perfino invertita, dai dati in possesso a causa di altri fenomeni non presi in considerazione nell'analisi (**i singoli corsi di studio**).

Popolazioni e sottopopolazioni: cosa c'è dentro

La covarianza e la correlazione (lineare) sono:

- positive** se - in media - a alti valori di X corrispondono alti valori di Y (e a bassi valori di X, bassi valori di Y)
- negative** se - in media - a alti valori di X corrispondono bassi valori di Y (e a bassi valori di X, alti valori di Y)
- circa 0** altrimenti

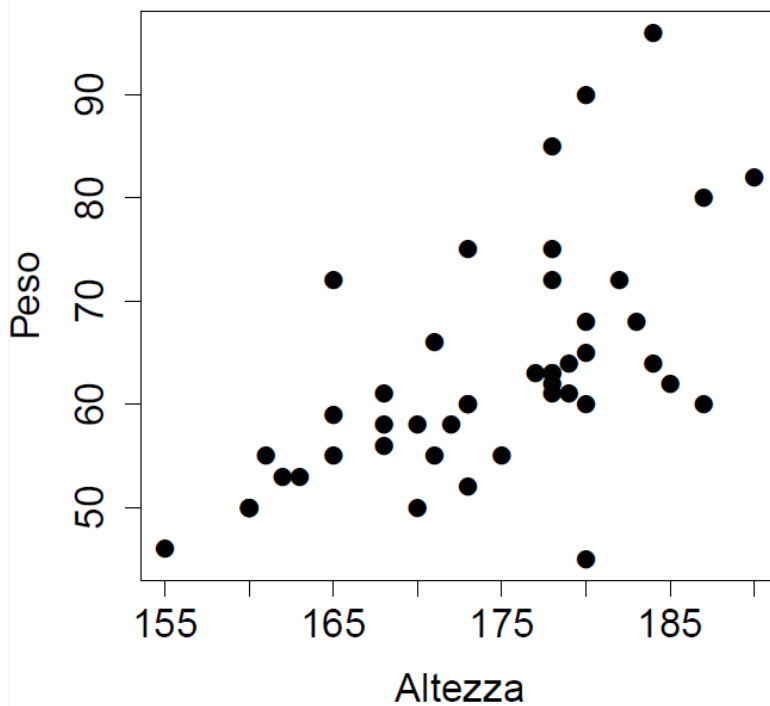


La correlazione (lineare) è compreso fra -1 e 1

Popolazioni e sottopopolazioni: cosa c'è dentro

Altezza e peso di alcuni studenti del primo anno di Università

$$\rho_{TOT} = 0,61$$



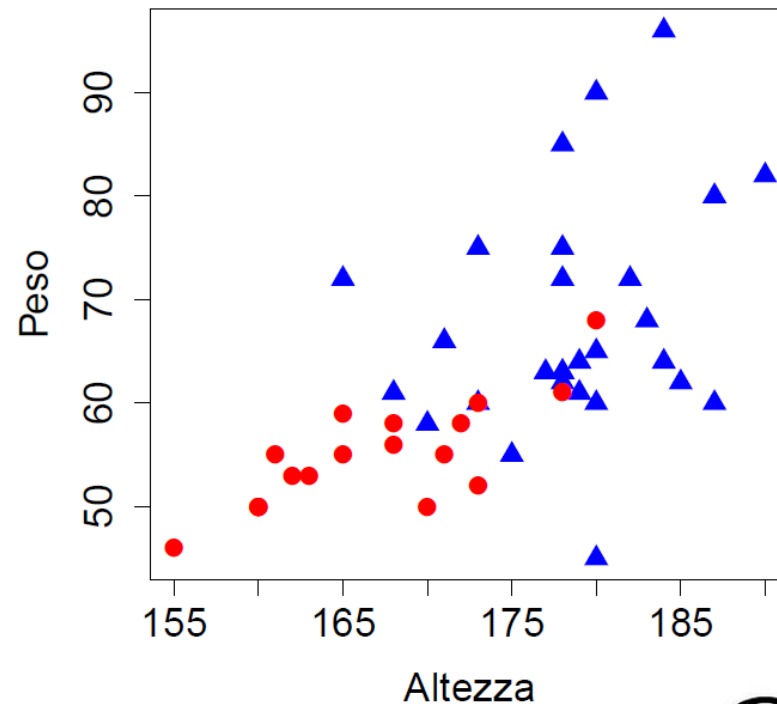
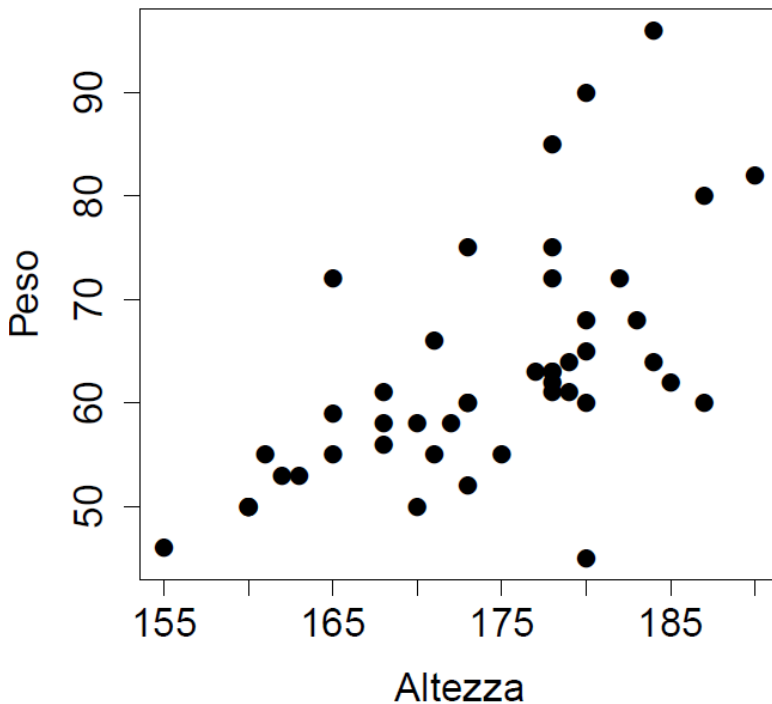
Popolazioni e sottopopolazioni: cosa c'è dentro

Altezza e peso di alcuni studenti del primo anno di Università

$$\rho_{TOT} = 0,61$$

$$\rho_M = 0,26$$

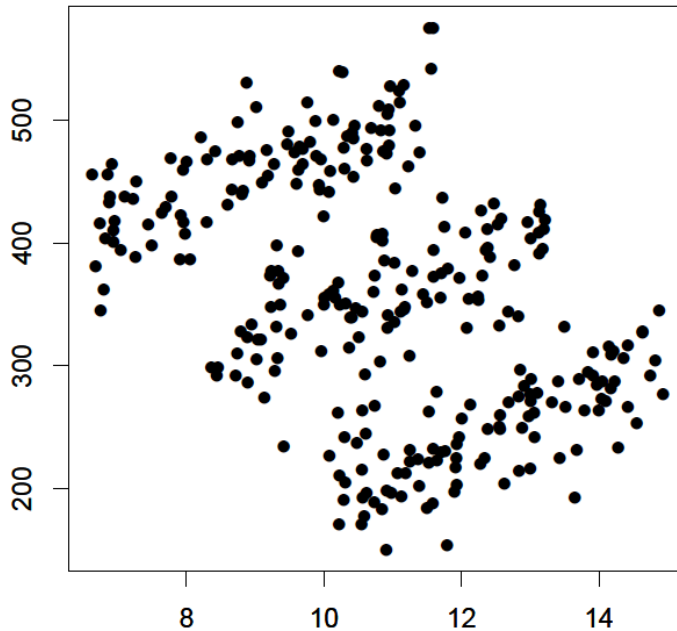
$$\rho_F = 0,78$$



Popolazioni e sottopopolazioni: cosa c'è dentro

Paradosso 1:

$$\rho_{TOT} < 0$$



Popolazioni e sottopopolazioni: cosa c'è dentro

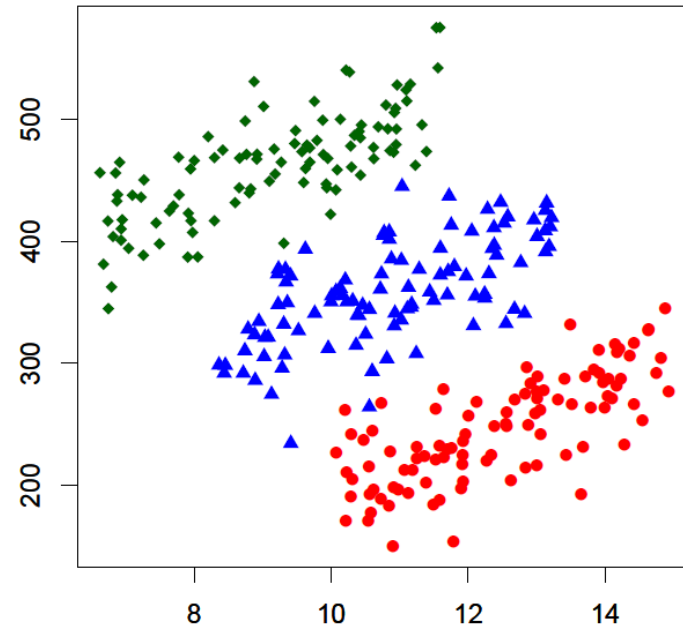
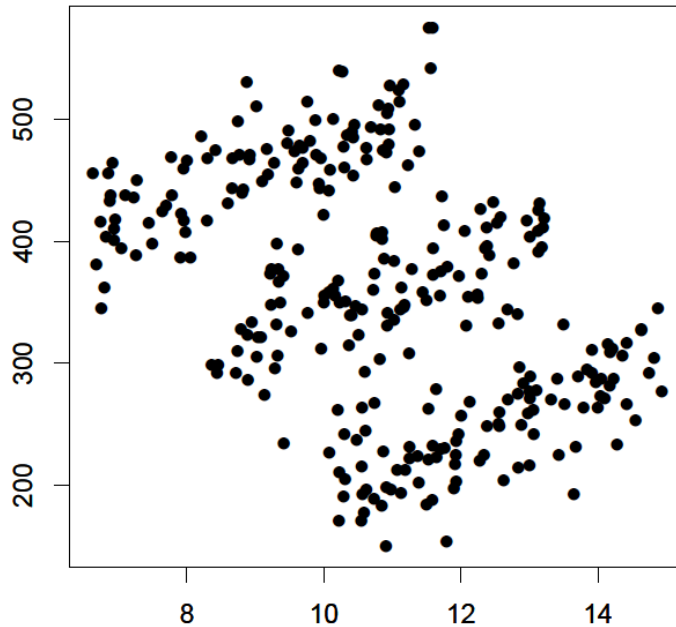
Paradosso 1:

$$\rho_{TOT} < 0$$

$$\rho_1 > 0$$

$$\rho_2 > 0$$

$$\rho_3 > 0$$



Popolazioni e sottopopolazioni: cosa c'è dentro

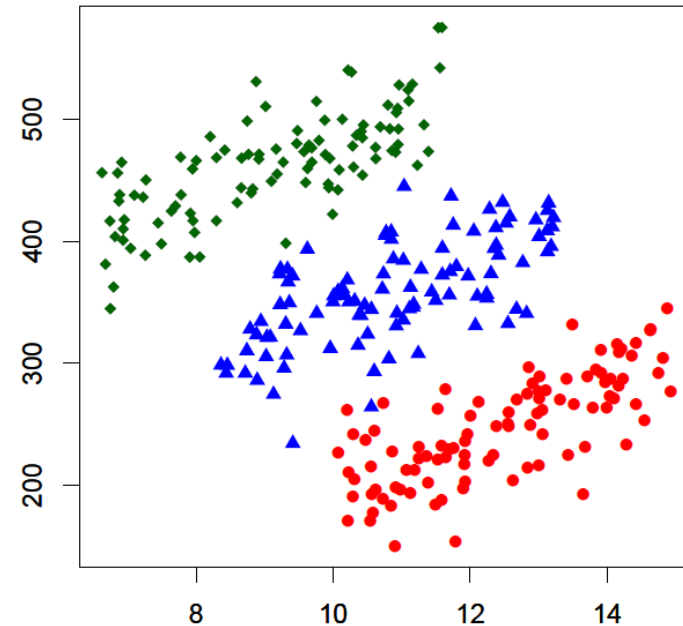
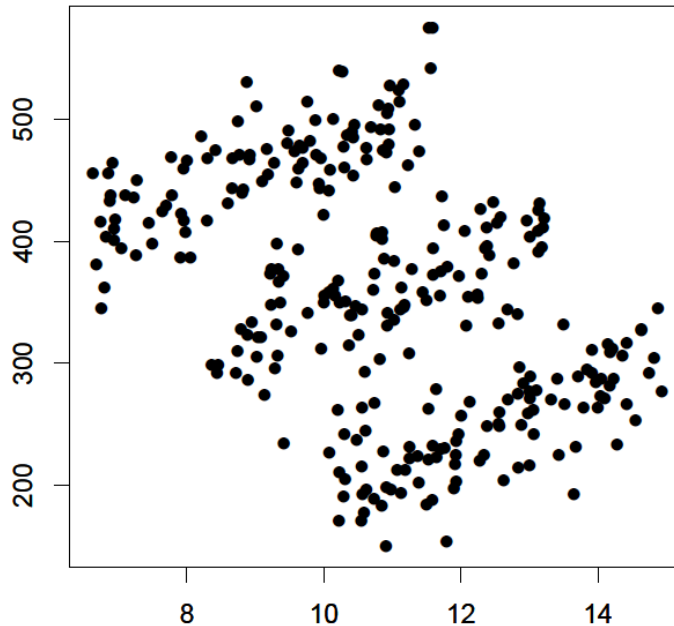
Paradosso 1:

$$\rho_{TOT} < 0$$

$$\rho_1 > 0$$

$$\rho_2 > 0$$

$$\rho_3 > 0$$



... vi ricorda qualcosa?

Popolazioni e sottopopolazioni: cosa c'è dentro

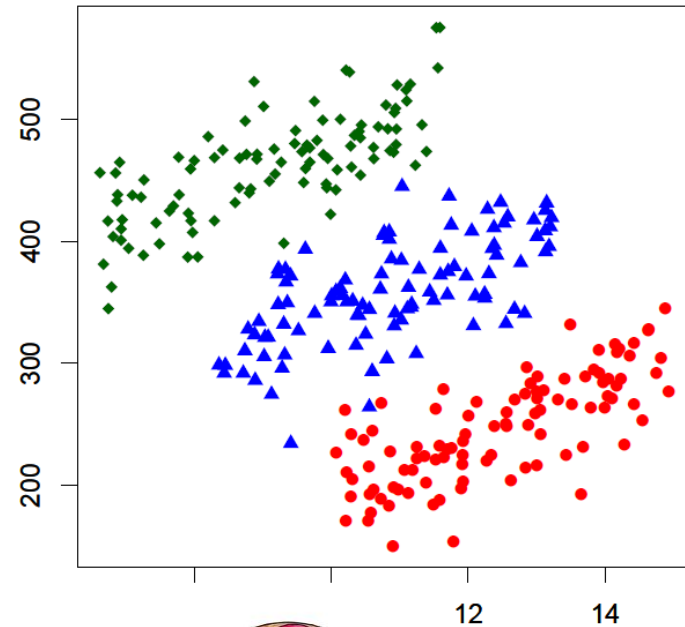
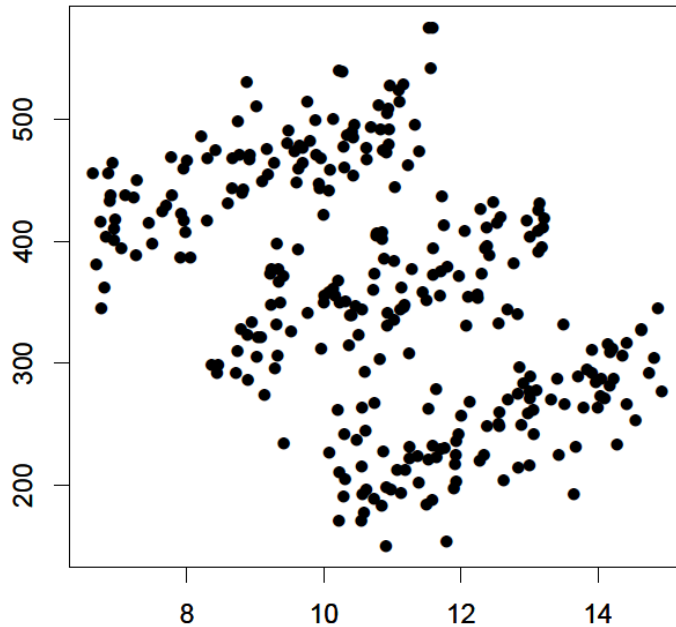
Paradosso 1:

$$\rho_{TOT} < 0$$

$$\rho_1 > 0$$

$$\rho_2 > 0$$

$$\rho_3 > 0$$



... vi ricorda qualcosa?

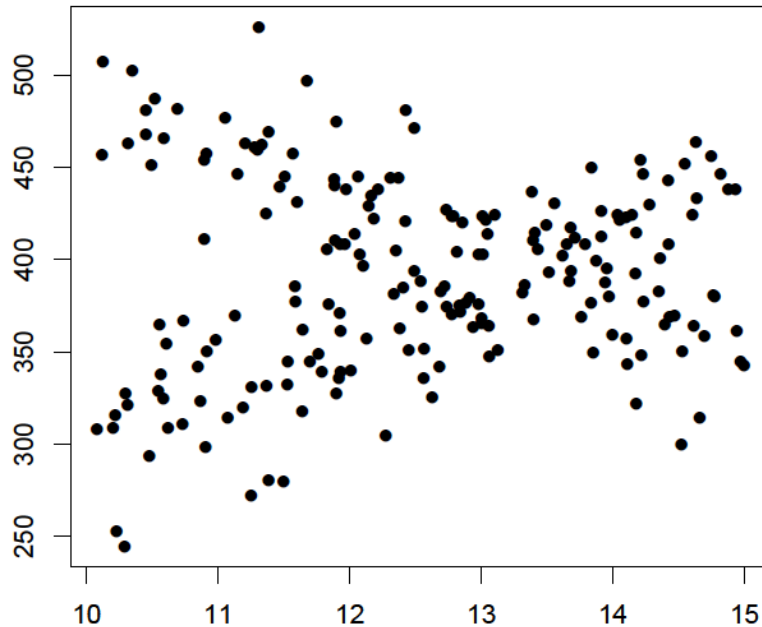


Smid

Popolazioni e sottopopolazioni: cosa c'è dentro

Paradosso 2:

$$\rho_{TOT} \approx 0$$



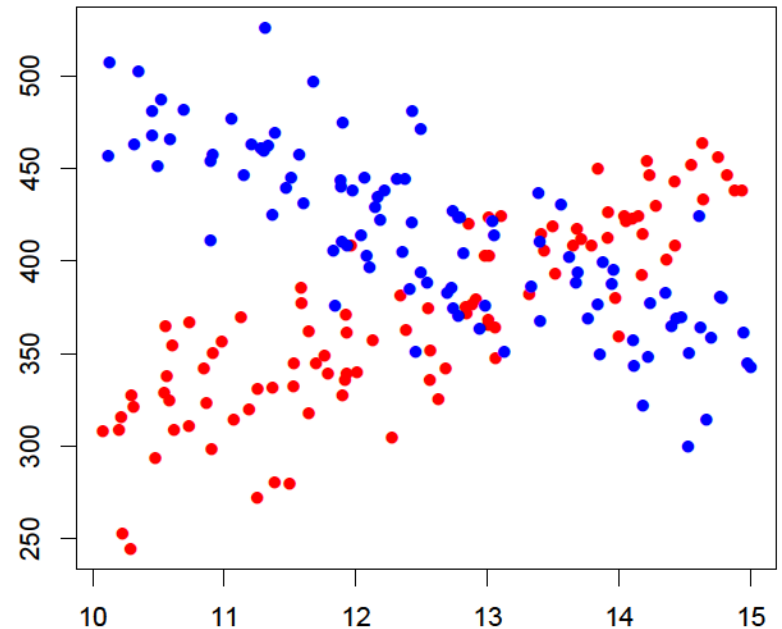
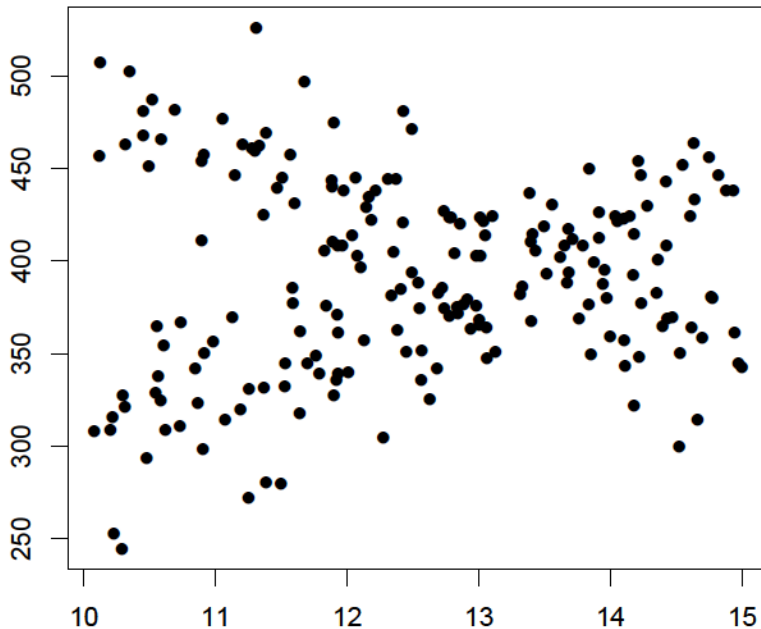
Popolazioni e sottopopolazioni: cosa c'è dentro

Paradosso 2:

$$\rho_{TOT} \approx 0$$

$$\rho_1 < 0$$

$$\rho_2 > 0$$



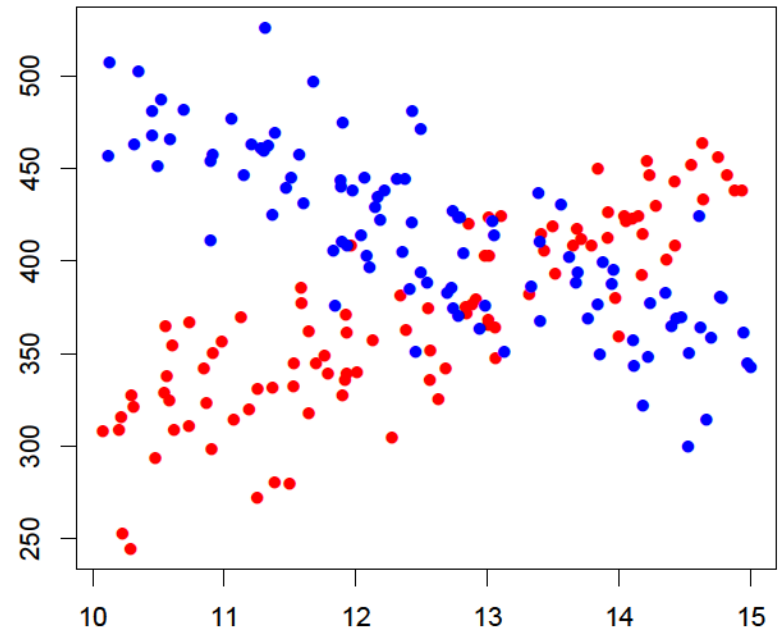
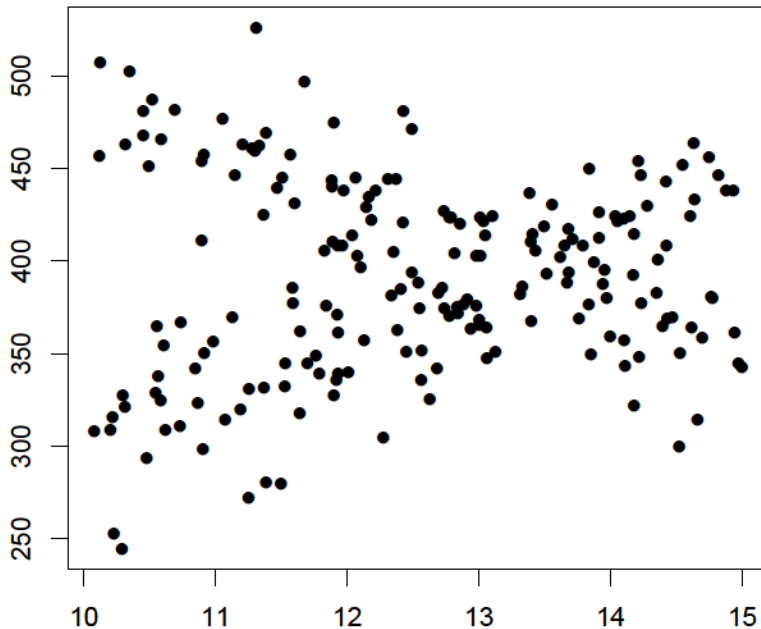
Popolazioni e sottopopolazioni: cosa c'è dentro

Paradosso 2:

$$\rho_{TOT} \approx 0$$

$$\rho_1 < 0$$

$$\rho_2 > 0$$



... vi ricorda qualcosa?

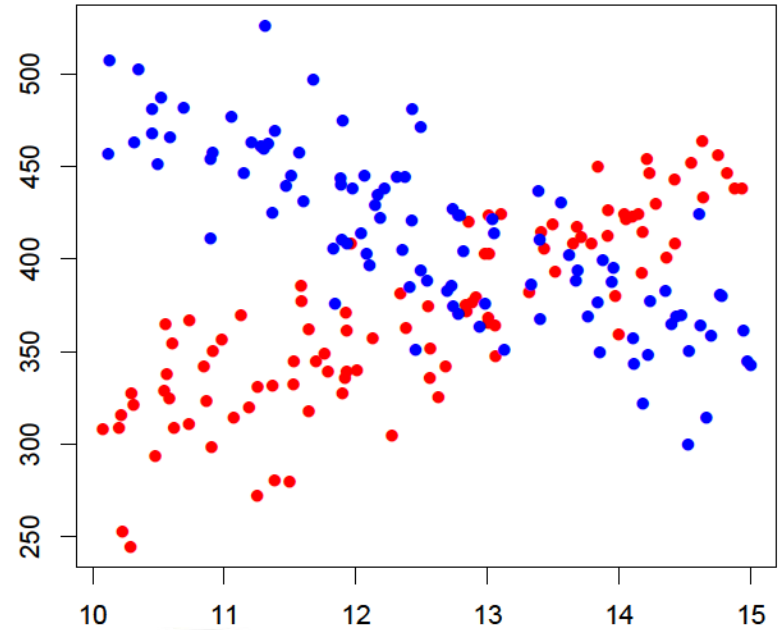
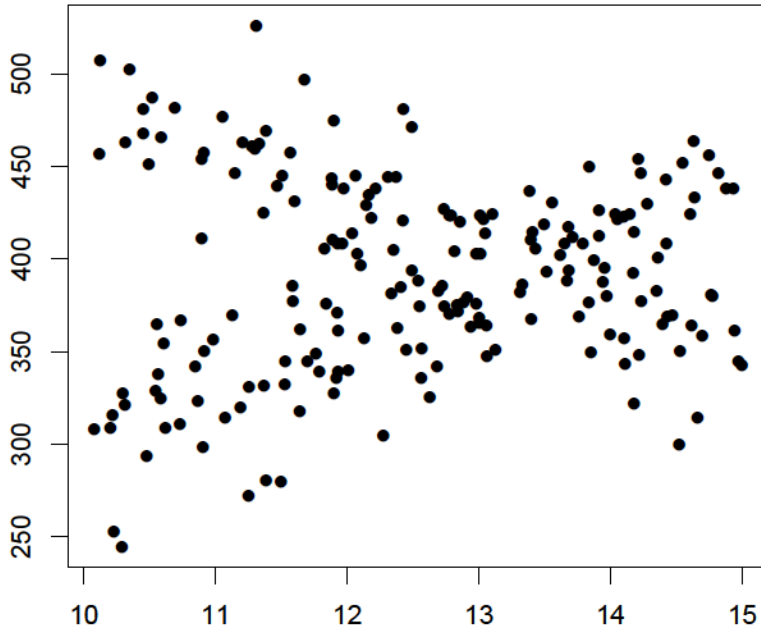
Popolazioni e sottopopolazioni: cosa c'è dentro

Paradosso 2:

$$\rho_{TOT} \approx 0$$

$$\rho_1 < 0$$

$$\rho_2 > 0$$



... vi ricorda qualcosa?



Correlazione \neq Causalità

Se due variabili statistiche sono correlate

NON SIGNIFICA

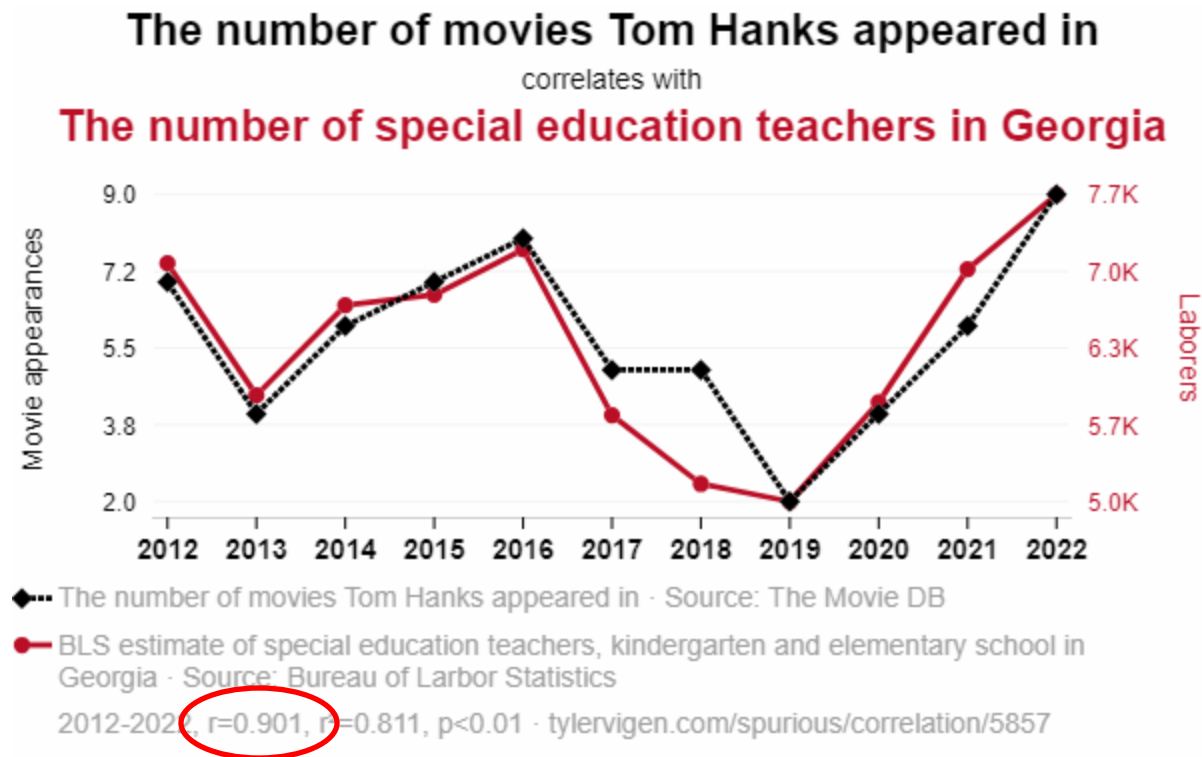
che esiste un legame di causalità tra le due

Si possono trovare numerosi esempi di variabili (anche molto) **correlate** tra di loro, che però non hanno **nessun legame causale** tra loro!

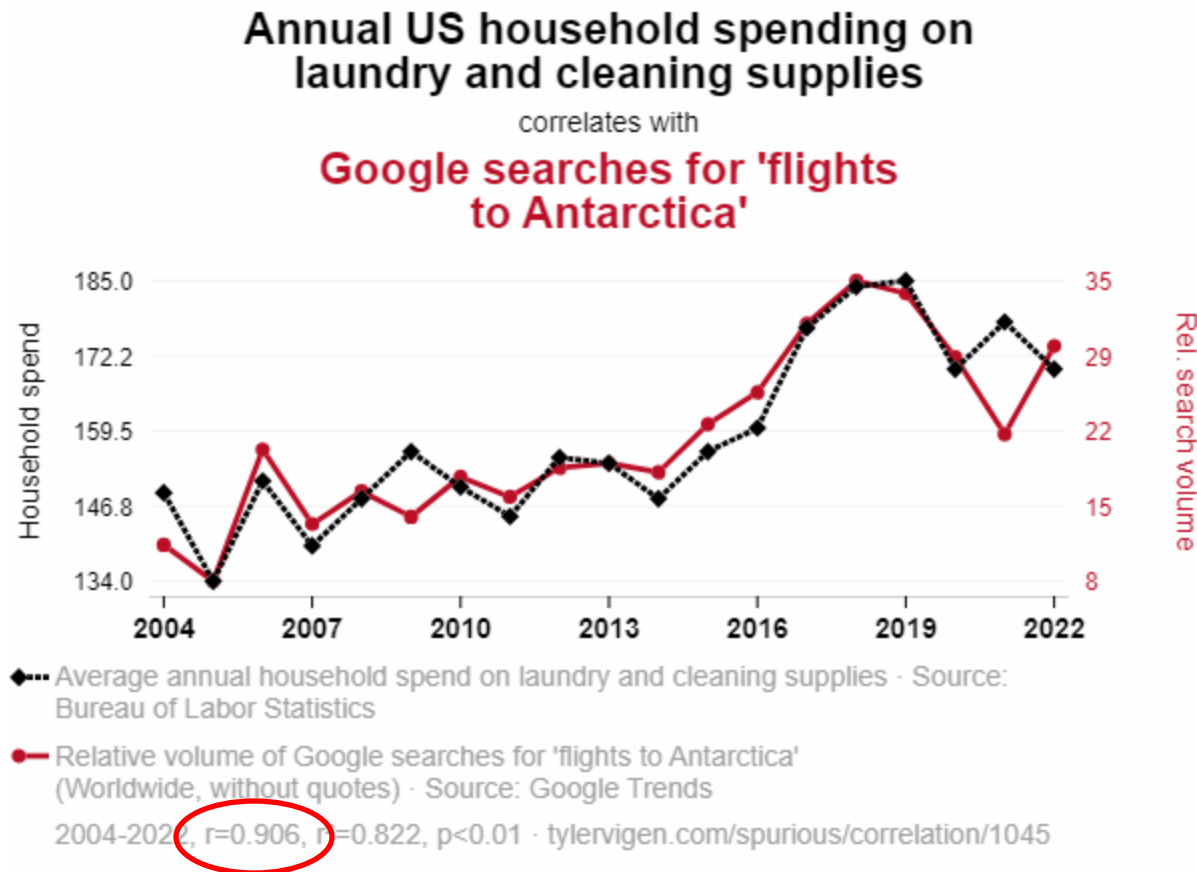
Ecco qualche esempio...



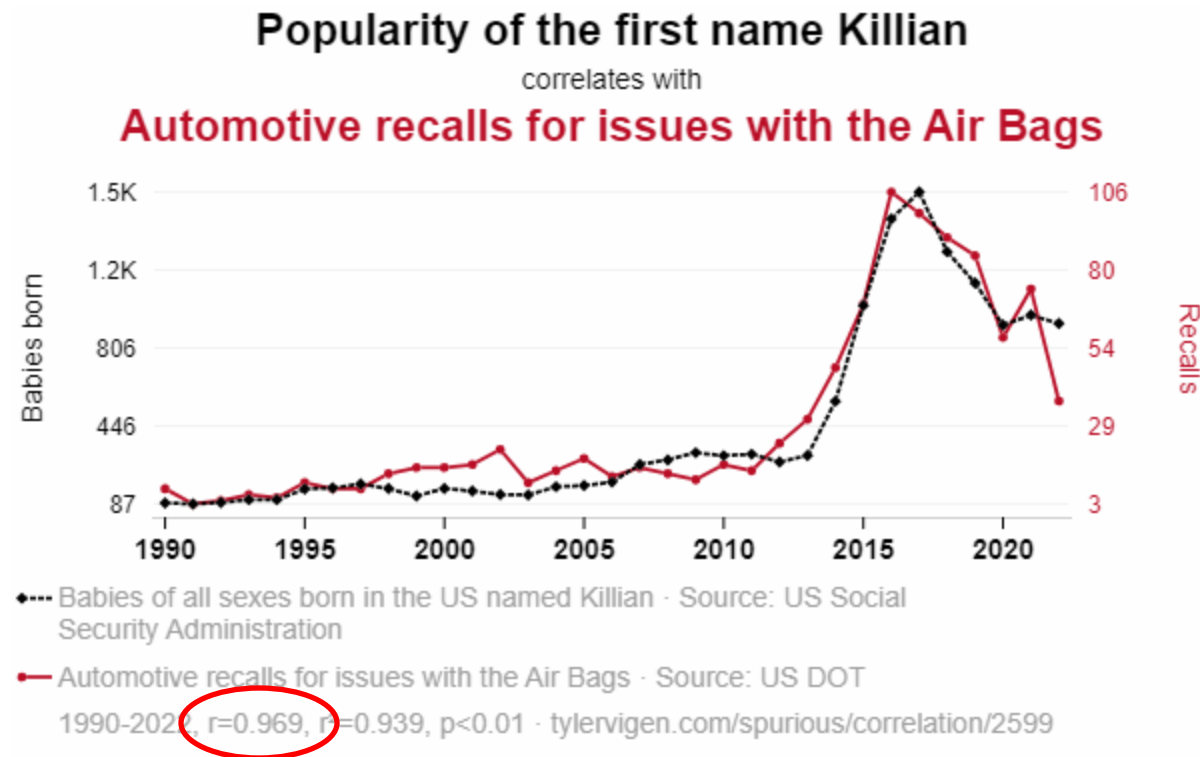
Correlazione \neq Causalità



Correlazione \neq Causalità



Correlazione \neq Causalità



..e quindi? Aveva ragione Trilussa?

Trilussa (1851-1950) poeta dialettale romano diceva:

«Da li conti che se fanno - seconno le statistiche d'adesso - risurta che te tocca un pollo all'anno: - e, se nun entra nelle spese tue, - t'entra ne la statistica lo stesso - perch'è c'è un antro che ne magna due.»

... non ci si può fidare di niente?

..e quindi? Aveva ragione Trilussa?

Trilussa (1851-1950) poeta dialettale romano diceva:

«Da li conti che se fanno - seconno le statistiche d'adesso - risurta che te tocca un pollo all'anno: - e, se nun entra nelle spese tue, - t'entra ne la statistica lo stesso - perch'è c'è un antro che ne magna due.»

... non ci si può fidare di niente?

NO, i dati - come i fatti del mondo - devono essere letti e **interpretati** con attenzione. Inoltre la conoscenza approfondita delle metodologie statistiche rende **consapevoli** anche della loro **applicabilità** alle situazioni concrete.



Smid

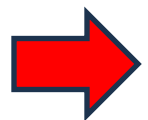
..e quindi? Aveva ragione Trilussa?

Trilussa (1851-1950) poeta dialettale romano diceva:

«Da li conti che se fanno - seconno le statistiche d'adesso - risurta che te tocca un pollo all'anno: - e, se nun entra nelle spese tue, - t'entra ne la statistica lo stesso - perch'è c'è un antro che ne magna due.»

... non ci si può fidare di niente?

NO, i dati - come i fatti del mondo - devono essere letti e **interpretati** con attenzione. Inoltre la conoscenza approfondita delle metodologie statistiche rende **consapevoli** anche della loro **applicabilità** alle situazioni concrete.



Ottima ragione per studiare Statistica!

Smid

GRAZIE per il tempo
che ci avete dedicato!!

Smid

GRAZIE per il tempo
che ci avete dedicato!!

**In bocca al lupo per i
vostri elaborati!!**

Smid

That's all, folks!

