L'analisi, il confronto e la valutazione critica della credibilità e dell'affidabilità delle fonti di dati, informazioni e contenuti digitali

Antonio D'Ambrosio

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II

Outline

Statistica e statistiche

I like to think of the constant presence in any sound Republic of two guardian angels: Statistician and the Historian of Science. The former keeps his finger on the pulse of Humanity, and gives the necessary warning when things are not as they should be. If Statistician is like a physician, the Historian is like a priest, – the guardian of man's most precious heritage, of the one treasure which, whatever may happen, can never be taken away from him – for the past is irrevocable

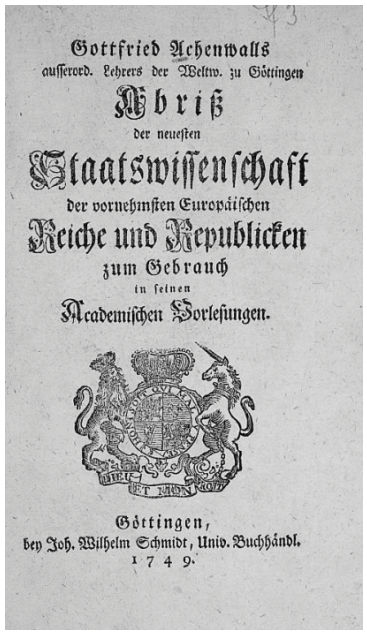
(Sarton G., 1935. "'Preface to Volume XXIII of Isis (Quetelet)," Isis, vol. 23, pp. 6–24.

Statistica, da dove deriva il termine?

- *Status* (stato politico): raccolta di informazioni organizzate e gestite dallo Stato;
- *Ratio status* (Ragion di stato):
  Stato è un dominio fermo sopra popoli; e Ragione di Stato è notizia di mezi atti a fondare, conservare, e ampliare un Dominio così fatto. Egli è vero che, se bene, assolutamente parlando, ella si stende alle tre parti sudette, nondimeno pare, che più strettamente abbracci la conservatione, che l'altre; e dell'altre più l'ampliatione, che la fondatione. (Botero G. (1589). "Della ragione di stato".
- *Staatswissenschaf* (Scienza dello stato): determinazione della forza politica di un Paese su questioni di stato;
- ...

Dispute sulla derivazione del nome

Prima volta in cui il termine "'statistik" è stato pubblicato: Achenwall (1749, Abriß der neuesten Staatswissenschaft der heutigen vornehmsten europäischen Reiche und Republiken zum Gebrauch in seinem Academischen Vorlesungen).


Gottfried Achenwalls
ausserord. Lehrers der Weltw. zu Göttingen
Abriß
der neuesten
Staatswissenschaft
der vornehmsten Europäischen
Reiche und Republicken
zum Gebrauch
in seinen
Academischen Vorlesungen.
Göttingen,
bey Joh. Wilhelm Schmidt, Univ. Buchhändl.
1749.

...la statistica non è un argomento che può essere compreso subito da un patè vuoto. Appartiene a una filosofia ben digerita, esige una conoscenza approfondita dello stato europeo e della storia naturale, insieme a una moltitudine di concetti e principi e una capacità di comprendere sufficientemente bene articoli molto diversi delle Costituzioni di Regni contemporanei.

Dispute sulla derivazione del nome

Giuseppe Ferrario (1839, Ragionamenti sull'utilità e necessità della statistica patologica, terapeutica e clinica e pensamenti sull'istituzione pubblica di una statistica clinica, nazionale e magistrale consentanea alla filosofia medica del secolo XIX).

RAGIONAMENTI

SULL' UTILITÀ E NECESSITÀ

DELLA

STATISTICA PATOLOGICA, TERAPEUTICA E CLINICA

E

PENSAMENTI SULL'ISTITUZIONE PUBBLICA

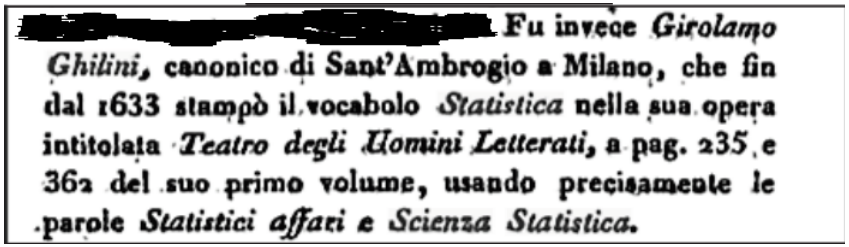
DI UNA

STATISTICA CLINICA, NAZIONALE E MAGISTRALE

CONSENTANEA ALLA FILOSOFIA MEDICA
DEL SECOLO XIX

DI

GIUSEPPE FERRARIO



[redacted] Fu invece *Girolamo Ghilini*, canonico di Sant'Ambrogio a Milano, che fin dal 1633 stampò il vocabolo *Statistica* nella sua opera intitolata *Teatro degli Uomini Letterati*, a pag. 235 e 362 del suo primo volume, usando precisamente le parole *Statistici affari* e *Scienza Statistica*.

Dispute sulla derivazione del nome

Ancora Giuseppe Ferrario (1838)

STATISTICA MEDICA

DI MILANO

DAL SECOLO XV FINO AI NOSTRI GIORNI

ESCLUSO IL MILITARE

DI

GIUSEPPE FERRARIO

DOTTORE DI MEDICINA, CHIRURGIA ED OSTETRICIA
MEDICO-CHIRURGO DELL'ACCADEMIA DE' FILODRAMMATICI DI MILANO,
PREMIATO PIÙ VOLTE DALL'I. R. GOVERNO

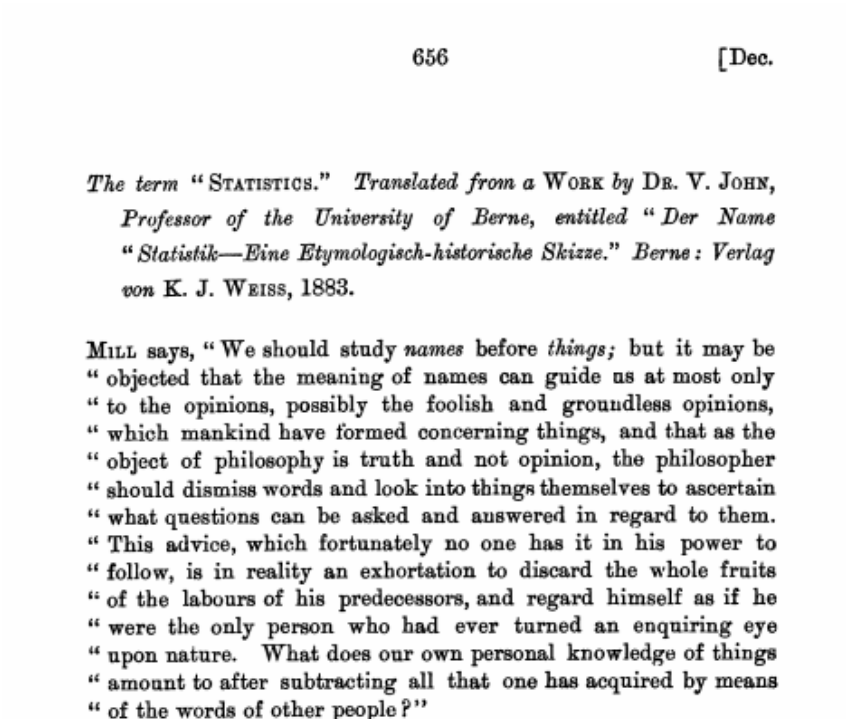
AUTORE DELLA STATISTICA DELLE MORTI INSORTE DALL'ANNO 1750 AL 1834
PUBBLICATA DALL'I. R. ISTITUTO DI SCIENZE, LETTERE ED ARTI
DEL REGNO LOMBARDO-VENETO.

MILANO

Abbiamo infatti già provato in principio dell'attuale opera al capo I *Etimologia*, ec., che il nostro Ghilini, canonico della basilica di sant'Ambrogio in Milano, usò i vocaboli *Statistica* e *Statistici* fin dall'anno 1633 a pag. 35, 36a del suo volume primo *Teatro degli Uomini Letterati*, in maniera tale da lasciar travedere che i detti vocaboli fossero già comuni nel secolo XVI; ed il Segneri parimenti a que' tempi adoperò nella sua predica 33, 3, la parola *Statisti* nel senso d'uomini di governo o di Stato. Dunque non alla Germania, ma all'Italia è dovuto il più antico uso del vocabolo *Statistica*, ed alla medesima Italia spetta anche la primitiva applicazione pratica della statistica scienza.

Dispute sulla derivazione del nome

John, V. (1838, Der Name Statistik – Eine Etymologisch-historische Skizze, Berne - published as English translation in "Journal of Royal Statistical Society" 1883, vol. 46, pp. 656–657.)



656 [Dec.

The term "STATISTICS." *Translated from a* WORK *by* DR. V. JOHN, *Professor of the University of Berne, entitled* "*Der Name* "*Statistik—Eine Etymologisch-historische Skizze.*" *Berne: Verlag von* K. J. WEISS, 1883.

MILL says, "We should study *names* before *things;* but it may be "objected that the meaning of names can guide us at most only "to the opinions, possibly the foolish and groundless opinions, "which mankind have formed concerning things, and that as the "object of philosophy is truth and not opinion, the philosopher "should dismiss words and look into things themselves to ascertain "what questions can be asked and answered in regard to them. "This advice, which fortunately no one has it in his power to "follow, is in reality an exhortation to discard the whole fruits "of the labours of his predecessors, and regard himself as if he "were the only person who had ever turned an enquiring eye "upon nature. What does our own personal knowledge of things "amount to after subtracting all that one has acquired by means "of the words of other people?"

The conflicting opinions which were expressed as to the true definition of the word led at last to the determination to combine the two general meanings, and thus "statistik" came to denote the *science of the condition of a State.* This rendering was accepted by Butte, who says, on p. 158 of his "Statistik als Wissenschaft" (1808): "According to its derivation "it is a branch of learning (*disciplin*) which treats of the *condition* "*of the State.*"

Kniess, in his work entitled "Die Statistik als Selbstständige "Wissenschaft," forcibly protested against this acceptation, and says: "Granted that the word is of Latin origin, and derived from "*status*, signifying *state* or *condition*, either the former or the latter "meaning must be taken, as in no case could it mean both at the "same time."

...e così 'statistik' finì per denotare la scienza della condizione di uno Stato.

... Secondo la sua derivazione 'è una branca del sapere (disciplinare) che tratta della condizione" dello Stato'.

Kniess (...) protesta fortemente contro questa accettazione e dice: 'Ammesso che la parola sia di origine latina e derivata da 'status', che significa stato o condizione, o il primo o il secondo significato deve essere accolto, poiché in nessun caso potrebbe significare entrambe le cose allo stesso tempo.

Di cosa si occupa la statistica?

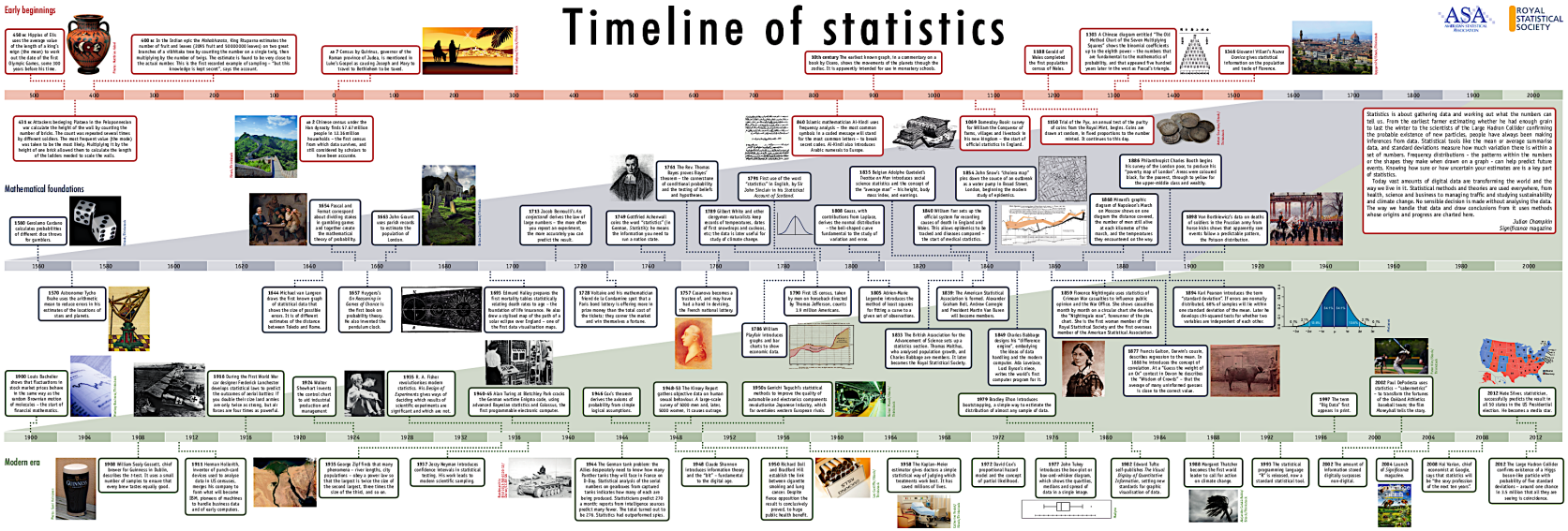
Sir John Sinclair (1791–1799, Statistical Account of Scotland, 20 volumi!!)

In Germania, (intendo la statistica come) un'inchiesta allo scopo di accertare la forza politica di un paese o affari riguardanti questioni di Stato; mentre, l'idea che associo al termine, è un'indagine sullo stato di un paese allo scopo di accertare il grado di felicità di cui godono i suoi abitanti e i mezzi del suo futuro miglioramento.

Pearson K. (*The history of statistics in the 17th and 18th centuries against the changing background of intellectual, scientific, and religious thought : lectures by Karl Pearson given at University College, London, during the academic sessions, 1921-1933 / edited by E. S. Pearson.*)

Uno scozzese ha rubato le parole 'Statistics' e 'Statistik' e le ha associate a metodi di 'Aritmetica Politica'.... E' come se qualcuno rubasse la nostra parola 'biometria' e la applicasse in un senso completamente diverso da quello dei suoi creatori!. In ogni caso, dobbiamo benedire Sinclair, visto che per un po' l' 'aritmetica' (o meglio l' 'algebra') è servita a marcare il carattere essenzialmente matematico della nostra scienza.

Timeline della Statistica



Distinguiamo la statistica intesa come ‘statistiche’ (raccolta di dati, tabelle, grafici) dalla statistica intesa come ‘scienza statistica’. Le statistiche esistono da quando esiste l’essere umano: censimenti, ‘stime’ dell’altezza delle fortificazioni per la costruzione di catapulte e di macchine per gli assedi, determinazione del gettito delle imposte, ecc.

Da attività di conteggio, nel XVII secolo si passa all’osservazione delle proprietà di un *insieme* di dati, del quale si cerca di studiarne aspetti come la variabilità, la sintesi, la dipendenza o l’indipendenza.

Numeri.....

- La Statistica moderna, nota come scienza delle decisioni in condizioni di incertezza, può aiutare attraverso opportuni scenari sperimentali ad avere un'idea migliore delle conseguenze delle decisioni intraprese.
- I numeri da sempre sono un potente strumento di comunicazione: sono importanti, sono meravigliosi...
- ...ma sono ingannevoli!

Numeri.....

I dati

- Sono raccolti correttamente (censimento, indagine campionaria, sondaggio, *merging* di data base, ...)
  - la fonte?
  - repository?
  - report riproducibile?
  - data cleaning?
  - dal report si evince la strategia di *missing data imputation*?
  - casi anomali o influenti (es. *outliers*) sono stati omessi? Perchè?
  - richiamo ad appendice tecnica?
- Non sono corretti (modificati, creati ad hoc, ...)
  - frode!!

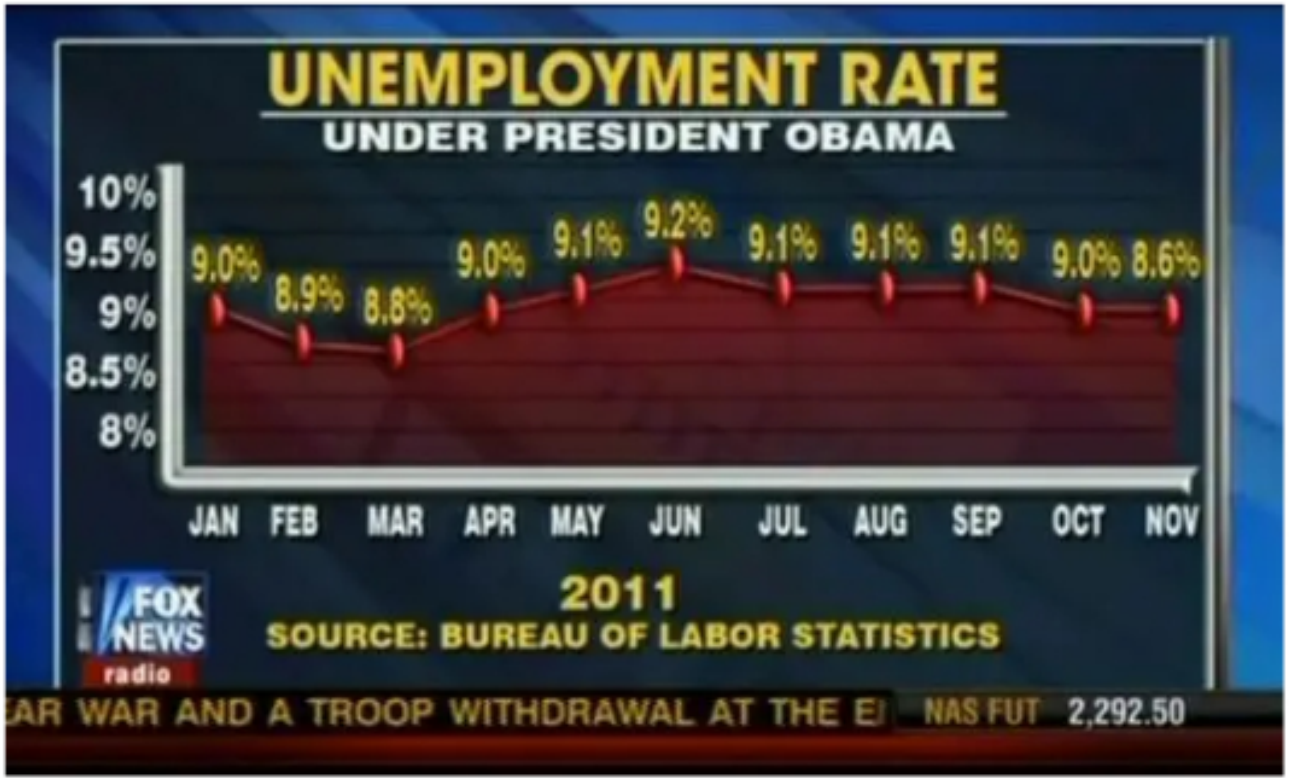

Governo Italiano
Ministero dell'Istruzione
Ministero dell'Università e della Ricerca
STAMPA E COMUNICAZIONE
Notizie
Comunicati
Rassegna Stampa
Dichiarazioni
Interviste
Discorsi
FOIA
Video
Social network
Campagne
Ufficio Stampa
FILTRI DI RICERCA
DATA DI PUBBLICAZIONE
2020
2019
2018
Home » Stampa e Comunicazione » Comunicati »
Covid-19, Ministero: dati su scuole forniti a ISS per puntuale analisi. Dal sistema scolastico grande impegno per la raccolta di informazioni
Giovedì, 29 ottobre 2020
Con riferimento ai dati relativi ai casi di infezione da SARS-CoV-2 rilevati in ambito scolastico, il Ministero dell'Istruzione comunica di aver condiviso, fin dall'inizio, i risultati del monitoraggio condotto attraverso i dirigenti scolastici con l'Istituto Superiore di Sanità, per contribuire alle analisi epidemiologiche.
Il sistema scolastico si è impegnato, fin dalla prima settimana di lockdown, per dare il proprio sostegno alla raccolta di informazioni, contribuendo a completare il quadro epidemiologico generale e a disegnare il trend di quanto osservato nelle scuole italiane in relazione al Covid-19.
nei confronti della sorveglianza epidemiologica, con lo scopo di contribuire a ridurre la diffusione dell'epidemia.
Tutti i dati sono in possesso delle autorità sanitarie, a cui sono trasferiti settimanalmente per la loro analisi nell'ambito del quadro epidemiologico generale.


epicentro
L'epidemiologia per la sanità pubblica
Istituto Superiore di Sanità
Coronavirus
SARS-CoV-2
Consulta i dati
Dati epidemiologici
Analisi sui decessi
Indagine nelle RSA
Novità

Numeri.....

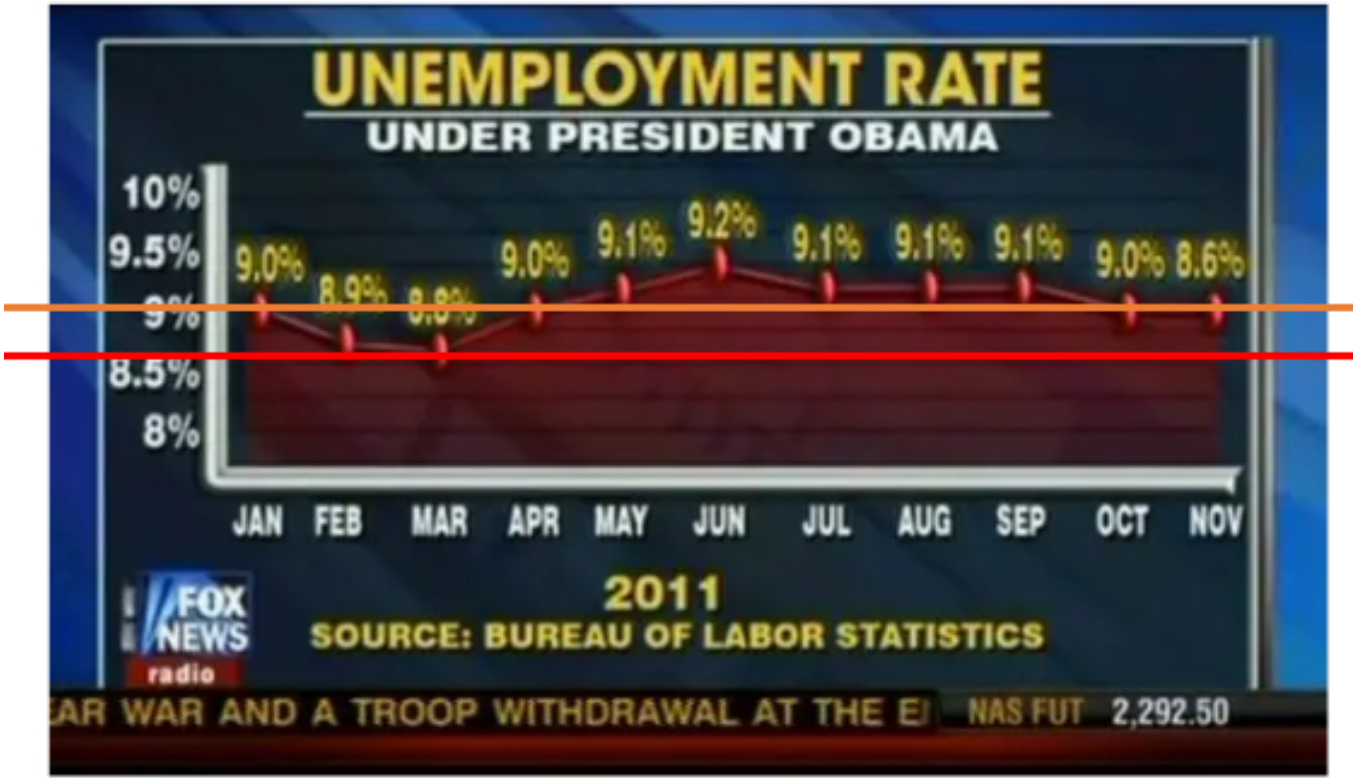
I dati sono fonte di informazione. L'informazione va comunicata bene. Soprattutto, l'informazione non deve essere fuorviante

Fox news 1



Si guarda il telegiornale in TV, si sta cenando, si presta poca attenzione ai commenti ma l'occhio cade sul grafico. Il grafico ci informa immediatamente su ciò che i commentatori stanno dicendo, perché 'concentrarsi' ad ascoltare? Eppure...

Orientare le opinioni: distorsione del risultato



Si guarda il telegiornale in TV, si sta cenando, si presta poca attenzione ai commenti ma l'occhio cade sul grafico. Il grafico ci informa immediatamente su ciò che i commentatori stanno dicendo, perché 'concentrarsi' ad ascoltare? Eppure il tasso di disoccupazione non è stagnante

Fox news 1

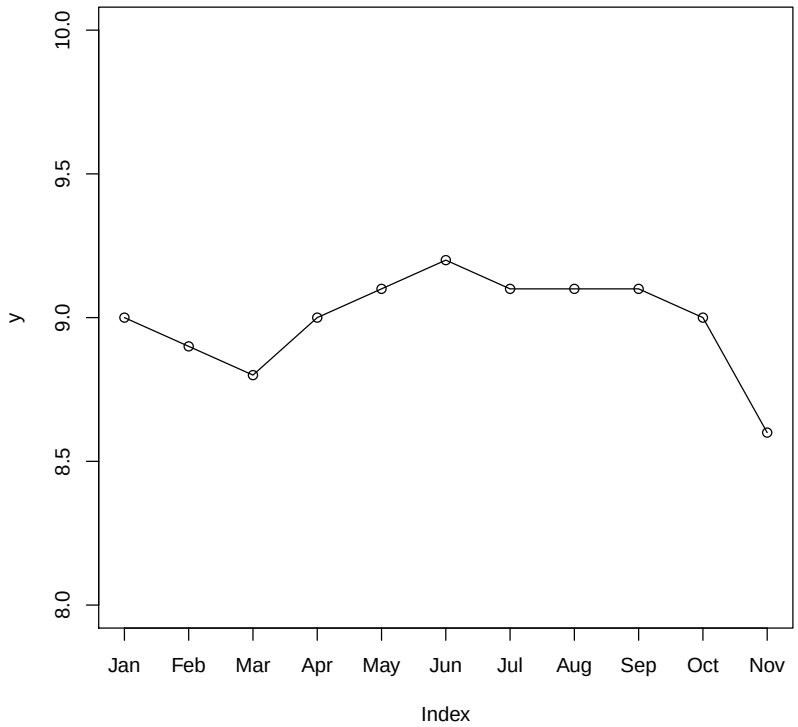
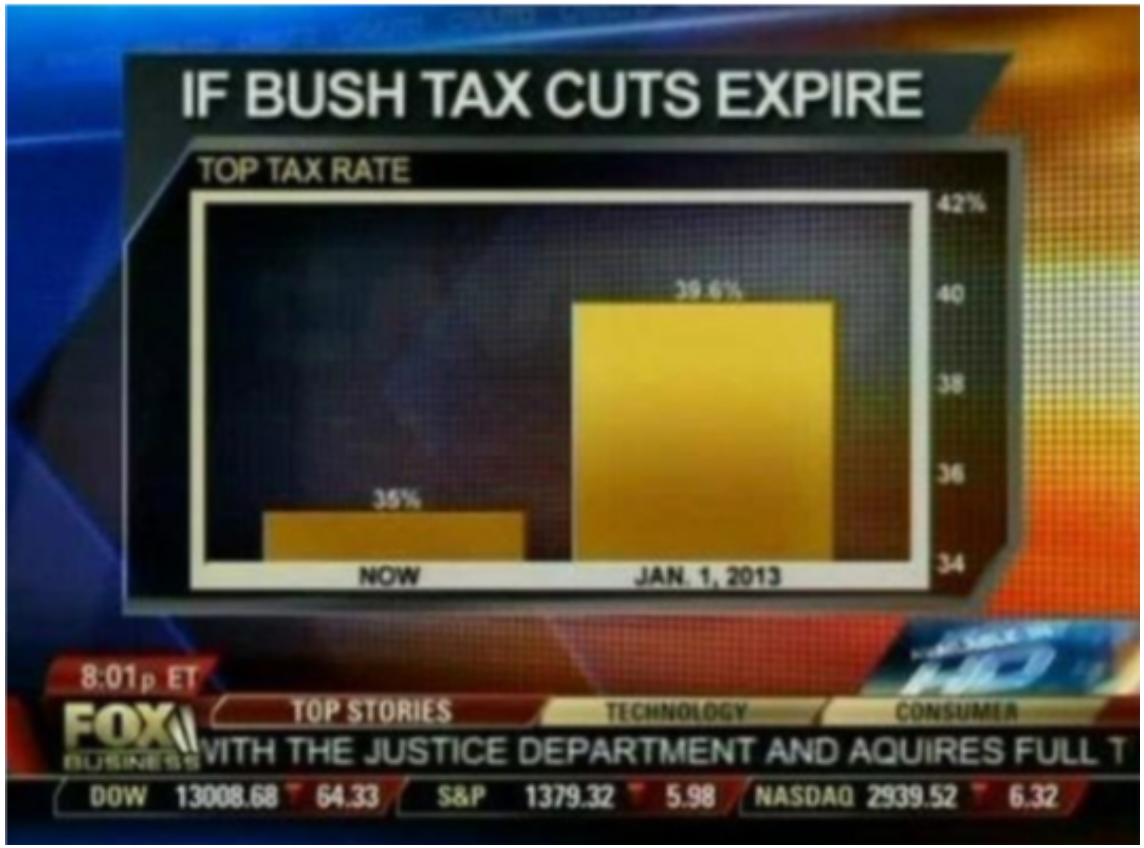


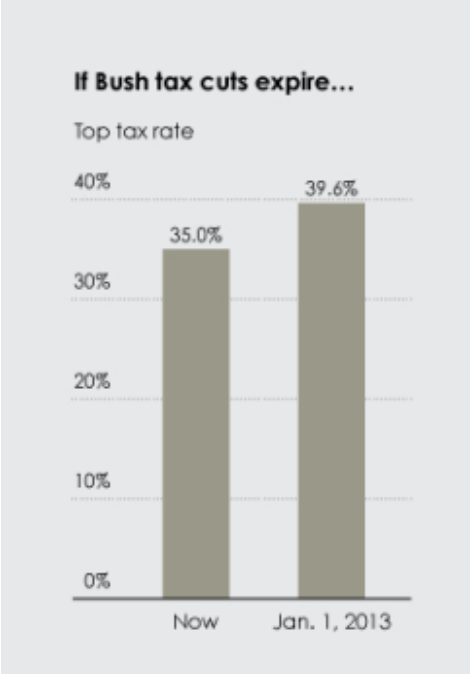
Grafico rifatto utilizzando i dati stampati sul grafico originale di Fox news

Orientare le opinioni: ancora Fox news



Si guarda il telegiornale in TV, si sta cenando, si presta poca attenzione ai commenti ma l'occhio cade sul grafico. Il grafico ci informa immediatamente su ciò che i commentatori stanno dicendo, perché 'concentrarsi' ad ascoltare? Il grafico mostra l'aumento delle tasse (delle top-tax) se il taglio delle tasse voluto da Bush dovesse decadere.

Fox news 2



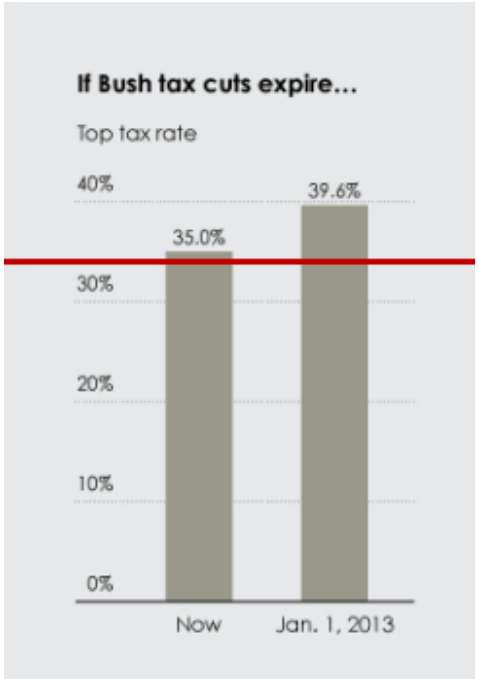


Grafico costruito utilizzando i dati stampati sul grafico originale di Fox news, e facendo partire i valori sull'asse Y da 0 e non dal 34%. L'incremento c'è, ovviamente, ma l'effetto (emotivo) non è lo stesso

La differenza tra i due grafici è evidente. Anche se l'informazione è la stessa (la differenza di tasse è e resta del 4.6%), l'incremento delle tasse sembra che sia 5 volte superiore al livello attuale.

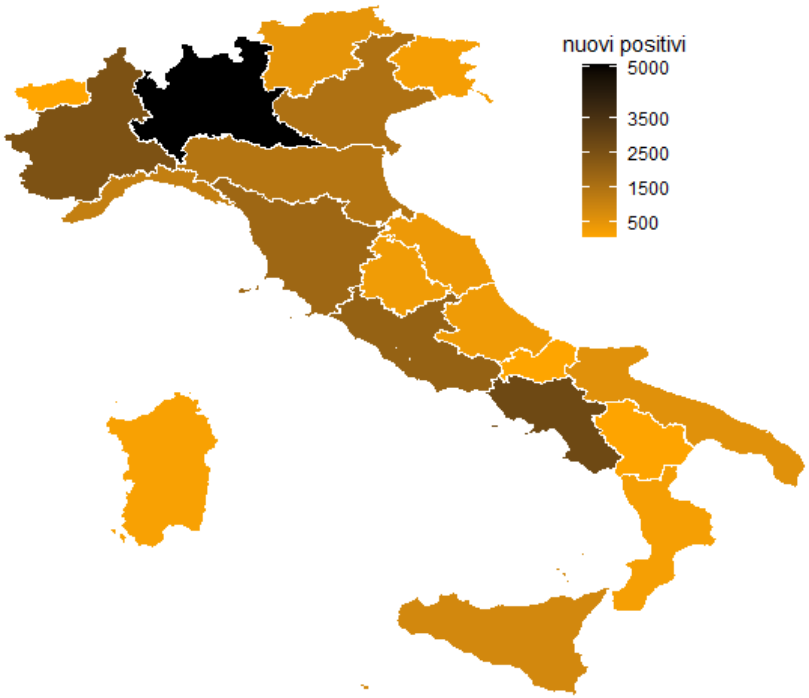
Informazione incompleta

Si guarda il telegiornale in TV, si sta cenando, si presta poca attenzione ai commenti ma l'occhio cade sul grafico. Il grafico ci informa immediatamente su ciò che i commentatori stanno dicendo, perché 'concentrarsi' ad ascoltare? Grafico mostrato a 'di martedì' del 27 Ottobre

P.s. I commenti in studio erano corretti.

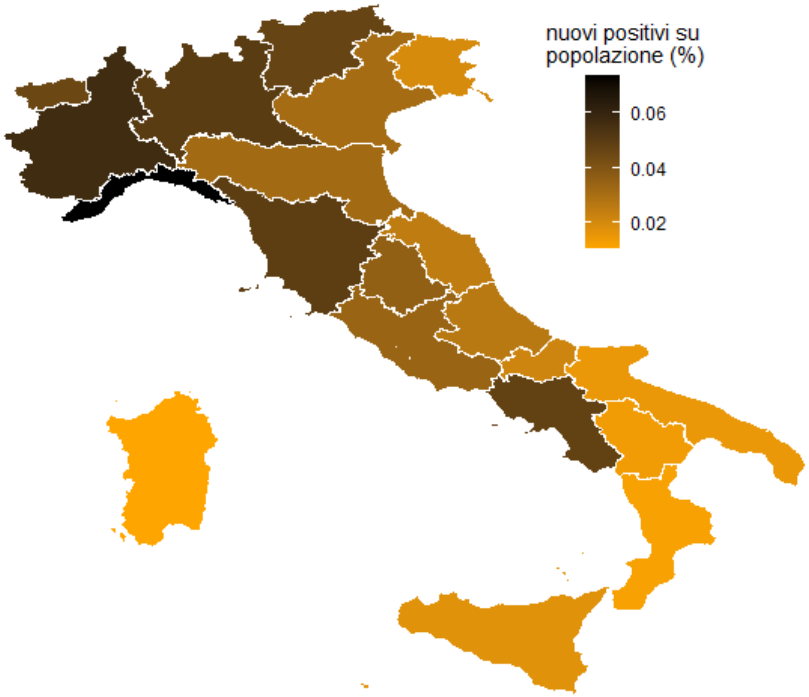
Informazione incompleta...

fonte `https://github.com/pcm-dpc/COVID-19`



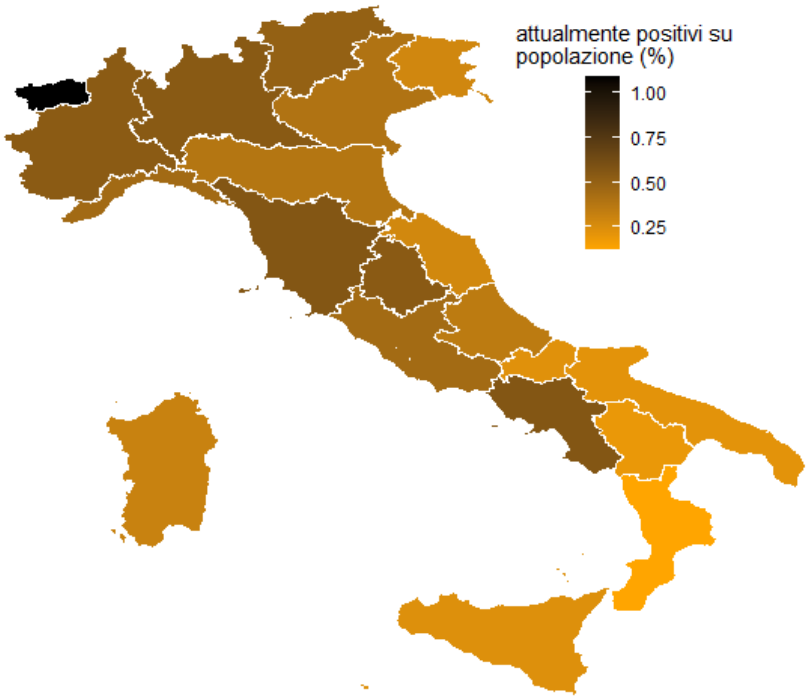
Stesso grafico visto in TV. Scala diversa, base dati differente, informazione sostanzialmente uguale. Analisi condotta sui dati del 27 Settembre 2020

Informazione incompleta...



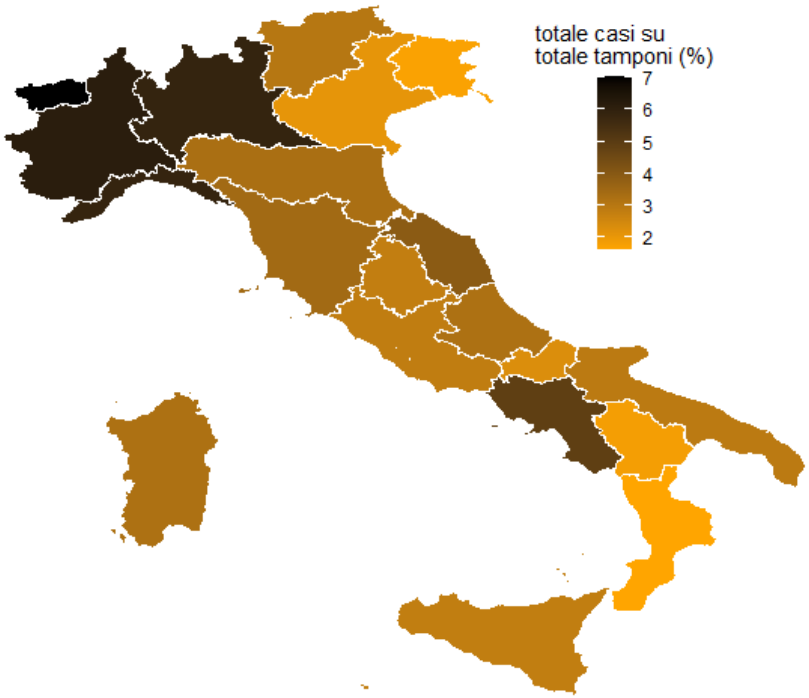
Normalizziamo i nuovi positivi rispetto alla popolazione di ogni regione

Informazione incompleta...



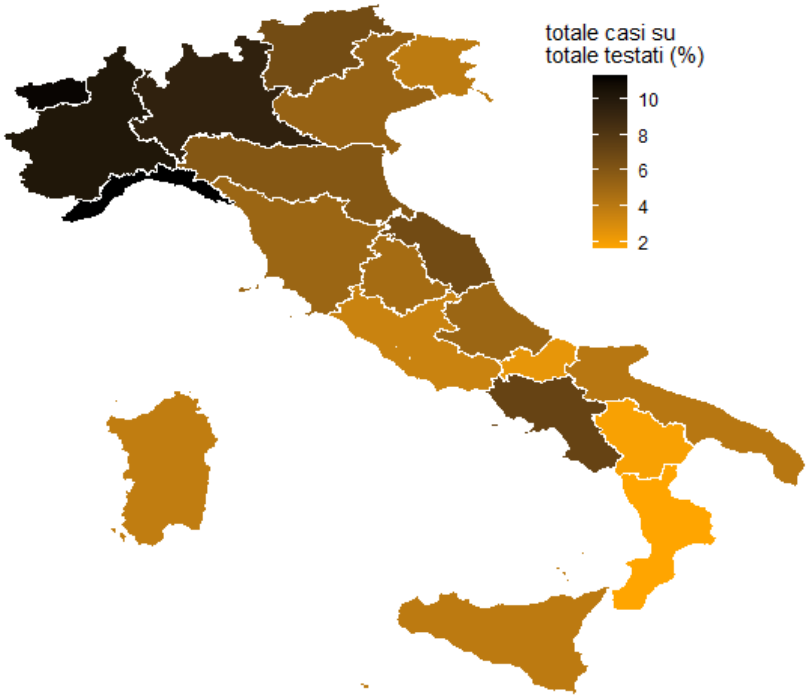
Normalizziamo gli *attualmente positivi* rispetto alla popolazione di ogni regione

Informazione incompleta...



Diamo un'occhiata ai casi totali sul totale tamponi effettuati in ogni Regione

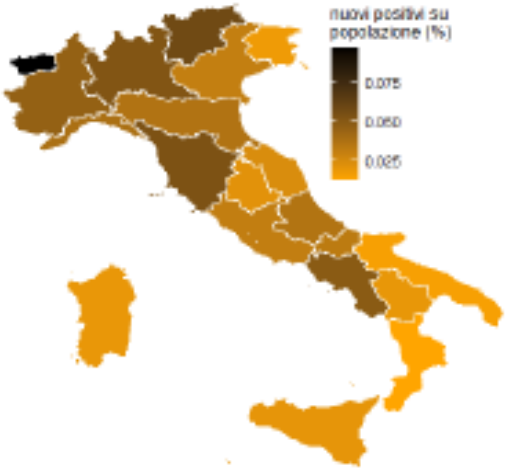
Informazione incompleta...

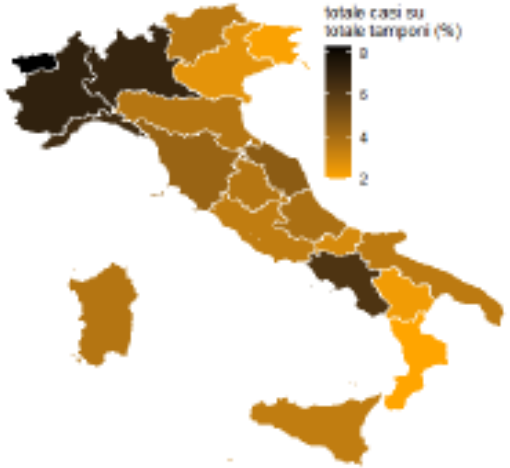


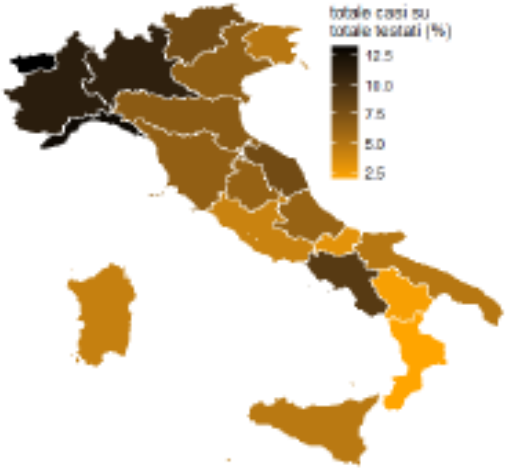
Diamo un'occhiata ai casi totali sul totale casi testati in ogni Regione

Informazione incompleta...

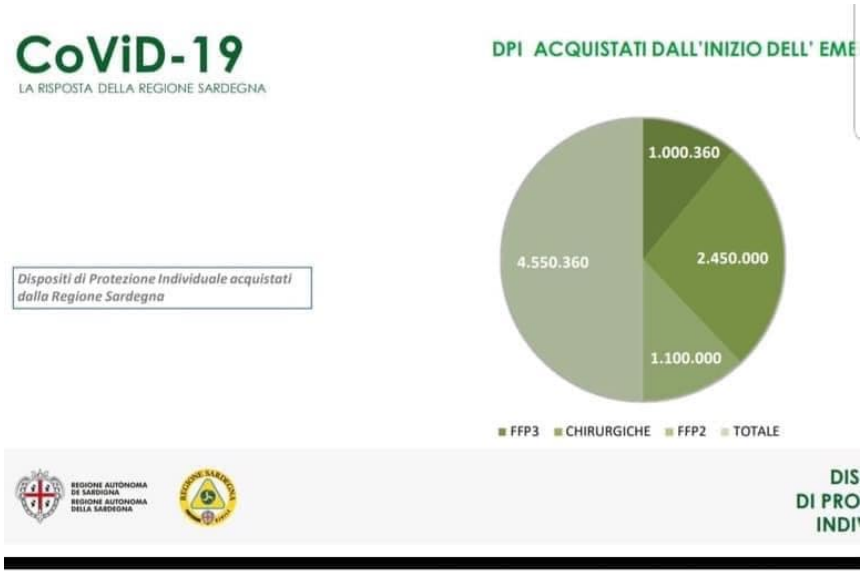
Vediamoli tutti insieme





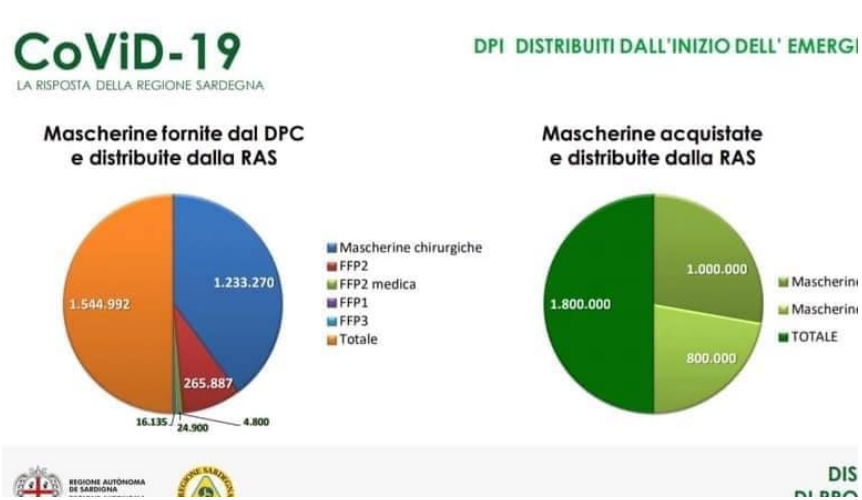


Torte...



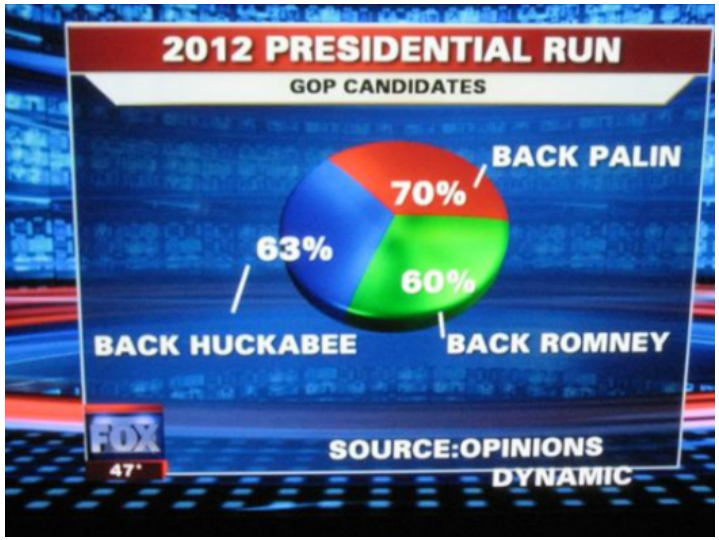
Facciamo due conti:

1.000.360+2.450.000+1.100.000=4.550.360. Quindi dal primo *pie chart* l'utente dovrebbe immediatamente capire dal rapporto tra gli spicchi che il 21, 98419% delle mascherine acquistate sono di tipo FFP3, il 53, 84189% sono chirurgiche e il restante 24, 17391% sono di tipo FFP2.



Nelle torte si vedono immediatamente due metà, perché una è per costruzione la somma delle componenti delle altre metà!

Torte...

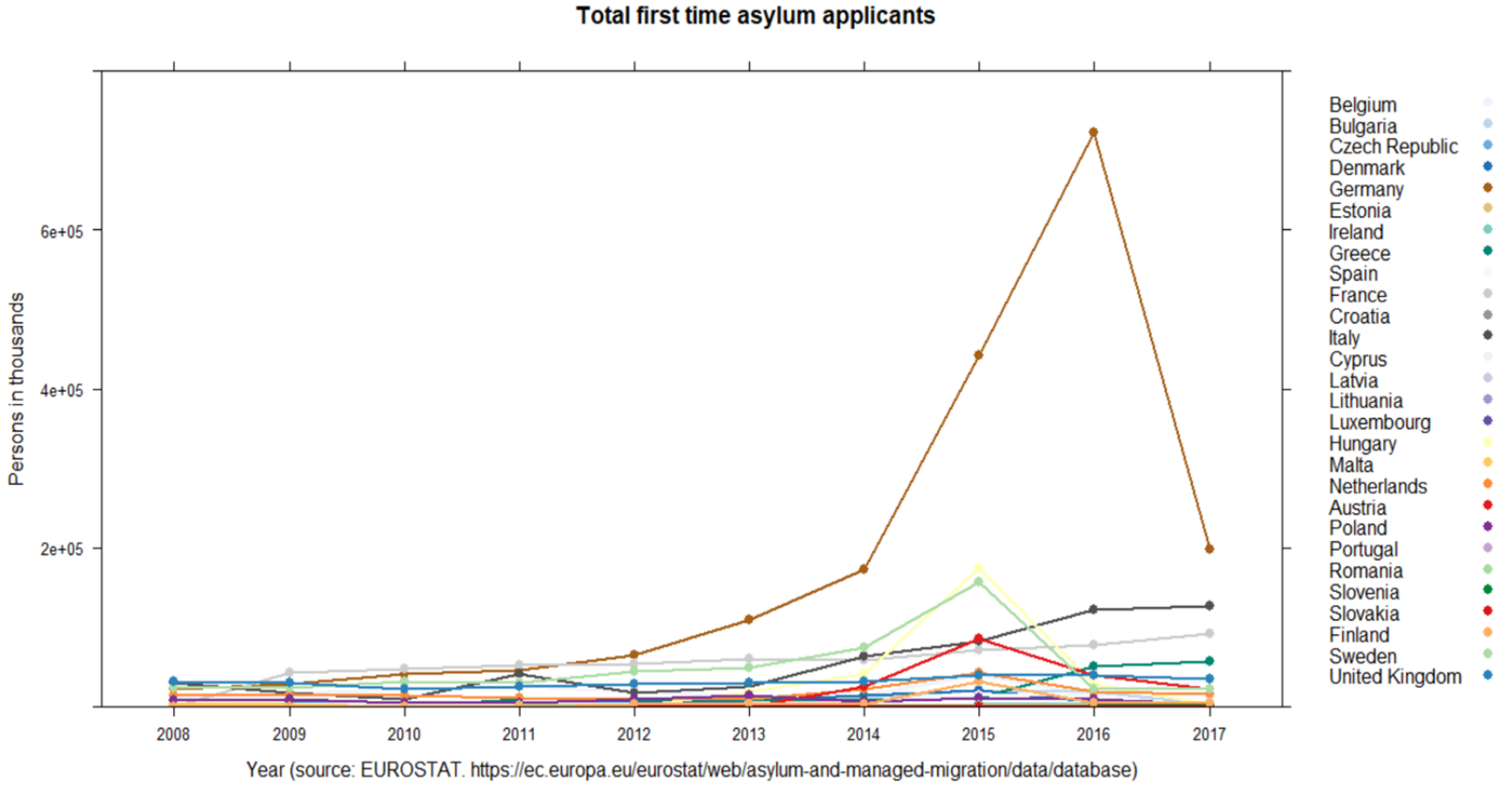


Facciamo due conti:

$63\% + 70\% + 60\% = 193\%$.

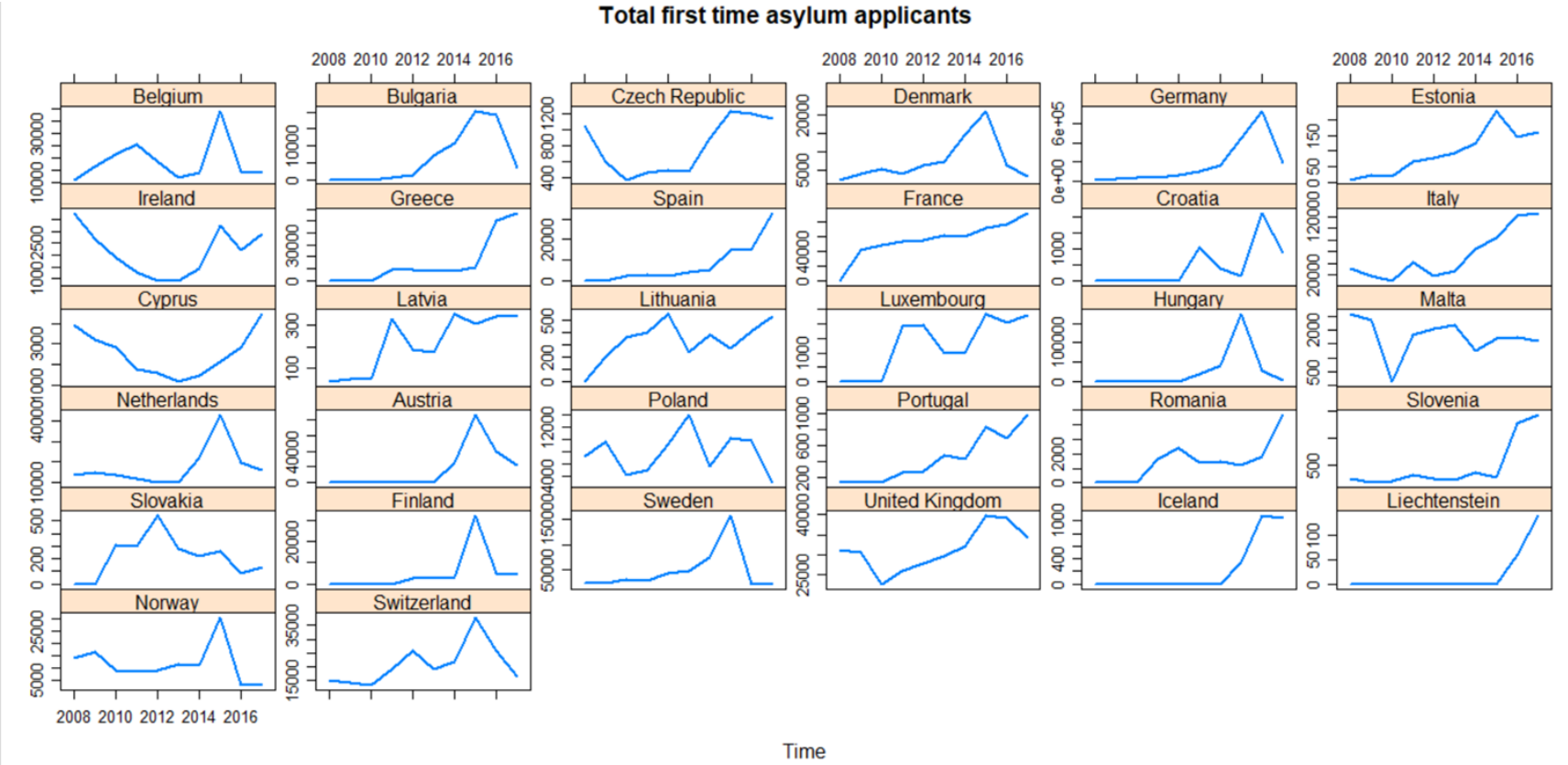
Il grafico a torta è la scelta sbagliata quando ai rispondenti è data la possibilità di poter fornire risposte multiple (categorie non mutualmente escludentesi).

Dati come fonte di discussione: un esempio sul fenomeno migranti



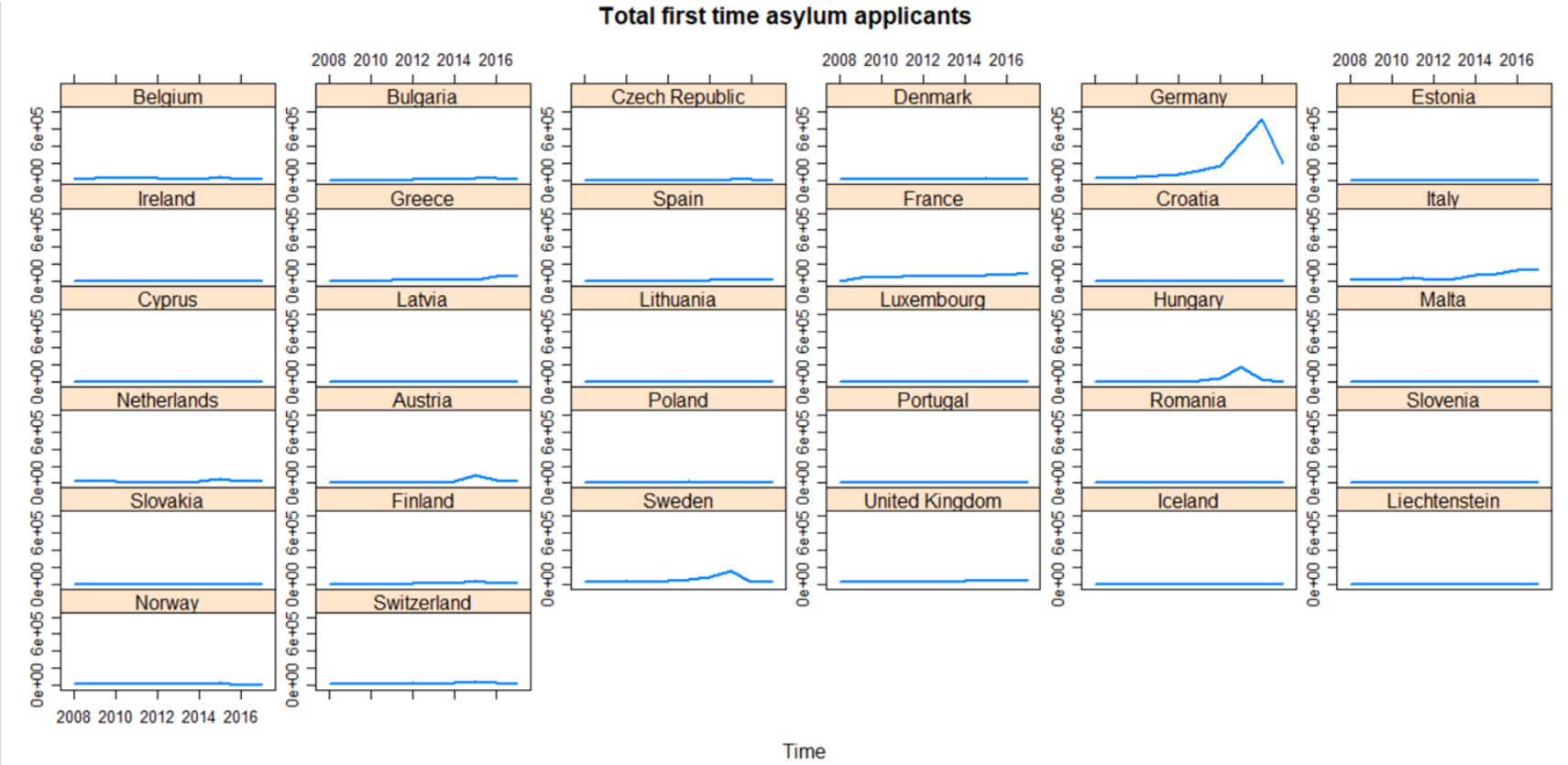
Andamento richiedenti asilo nei paesi UE+ Islanda, Liechtenstein, Norvegia e Svizzera.

Dati come fonte di discussione: un esempio sul fenomeno migranti



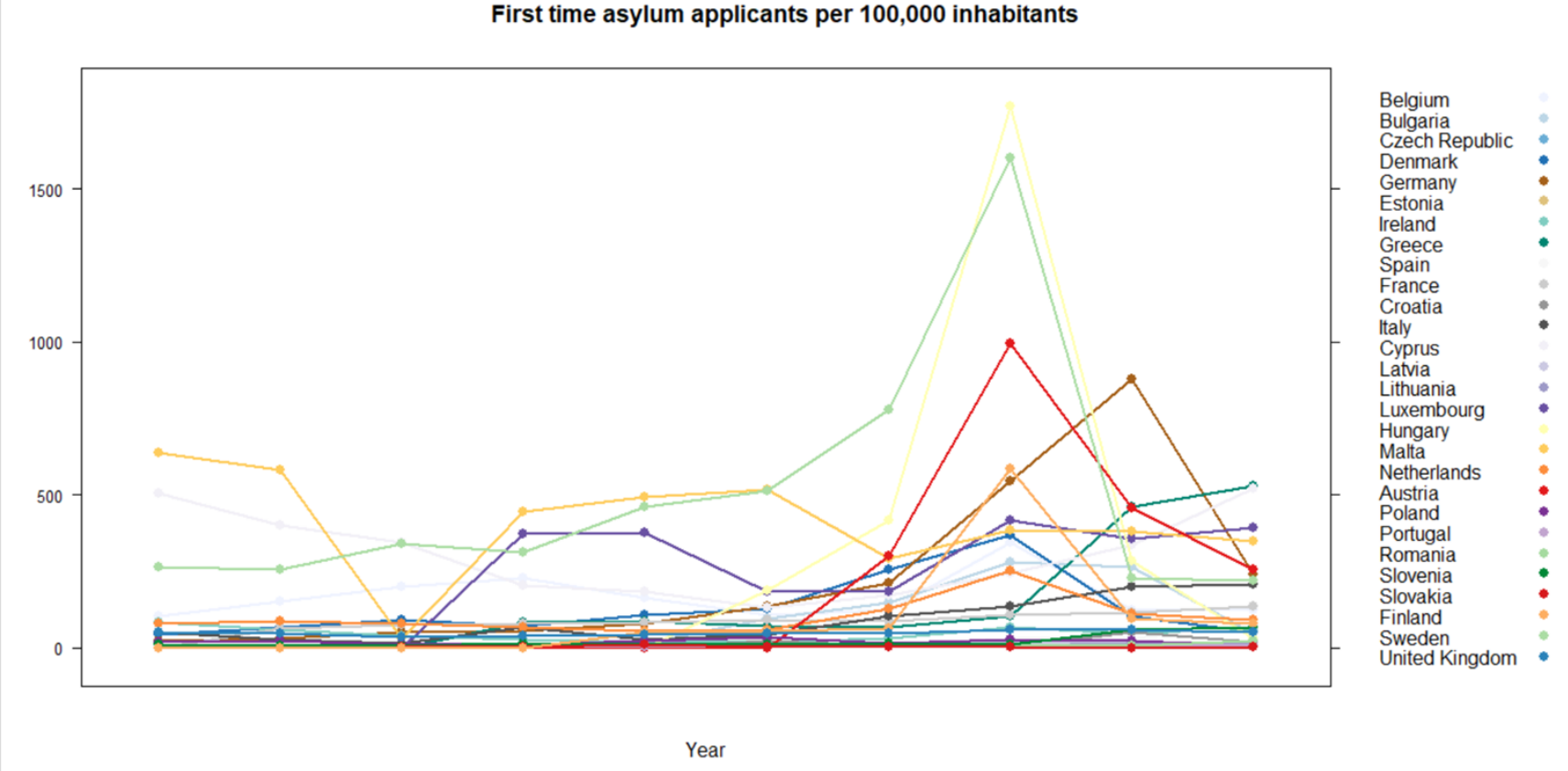
Stesso grafico di prima, con un pannello diverso per ognuno dei 32 Paesi. Immagine chiara?

Dati come fonte di discussione: un esempio sul fenomeno migranti



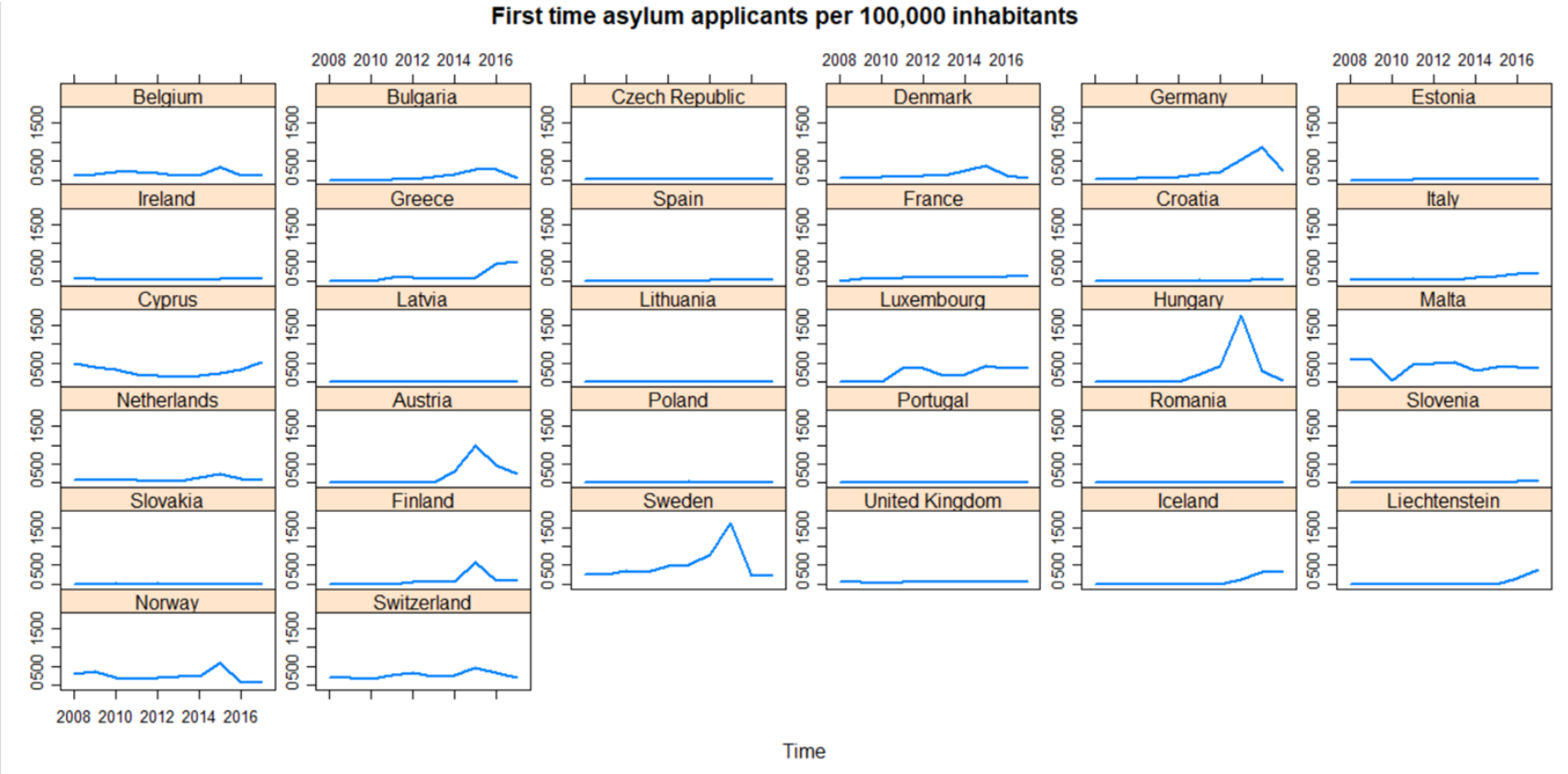
Stesso grafico di prima, ma più 'onesto': l'asse verticale ha lo stesso range per tutti

Dati come fonte di discussione: un esempio sul fenomeno migranti



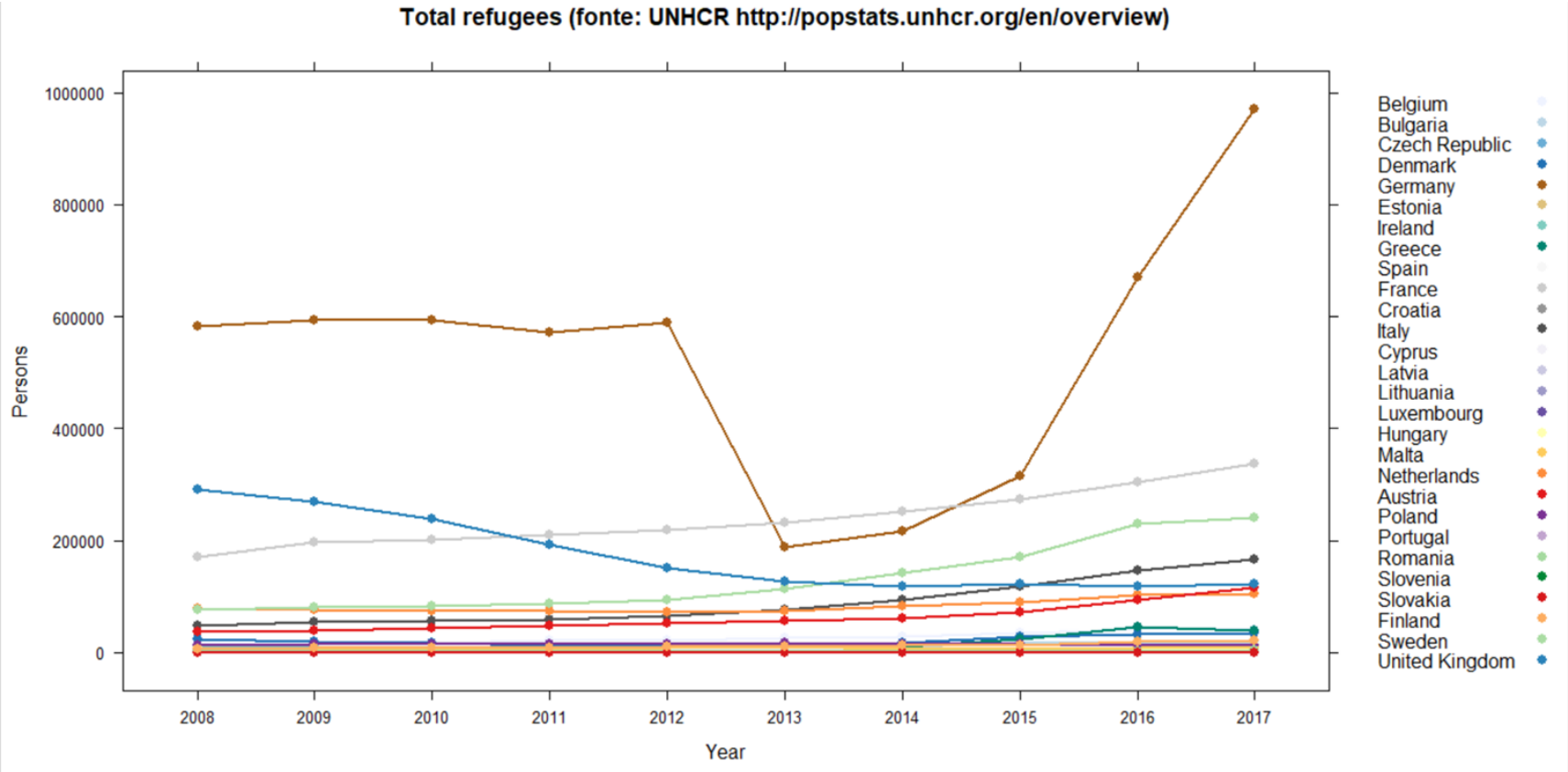
Ma qual è l'informazione che interessa? Proviamo a normalizzare il tutto per la popolazione di ogni Stato, rendendo i valori per 100.000 abitanti

Dati come fonte di discussione: un esempio sul fenomeno migranti



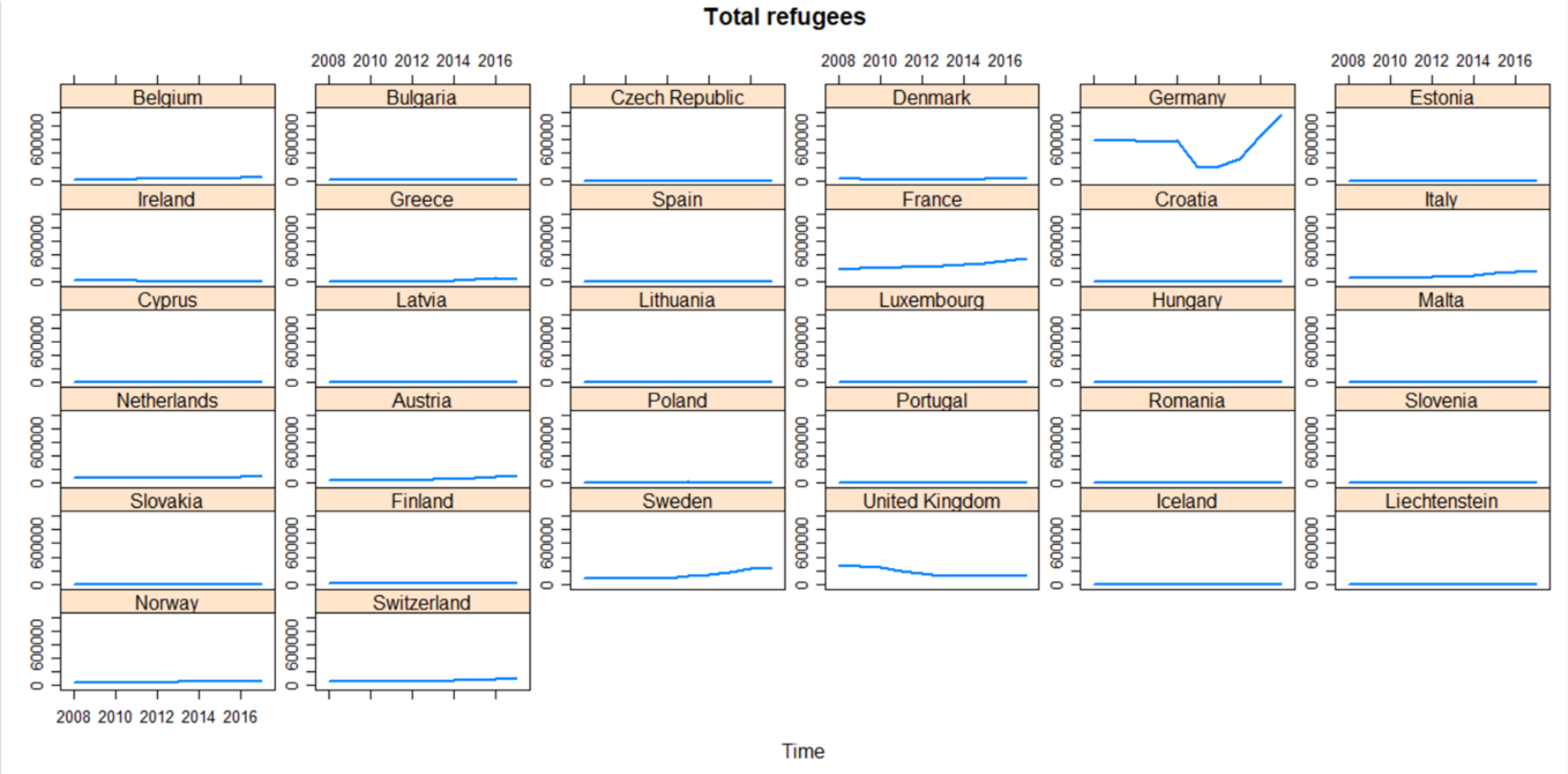
Richiedenti asilo per 100.000 abitanti, grafico particolareggiato

Dati come fonte di discussione: un esempio sul fenomeno migranti



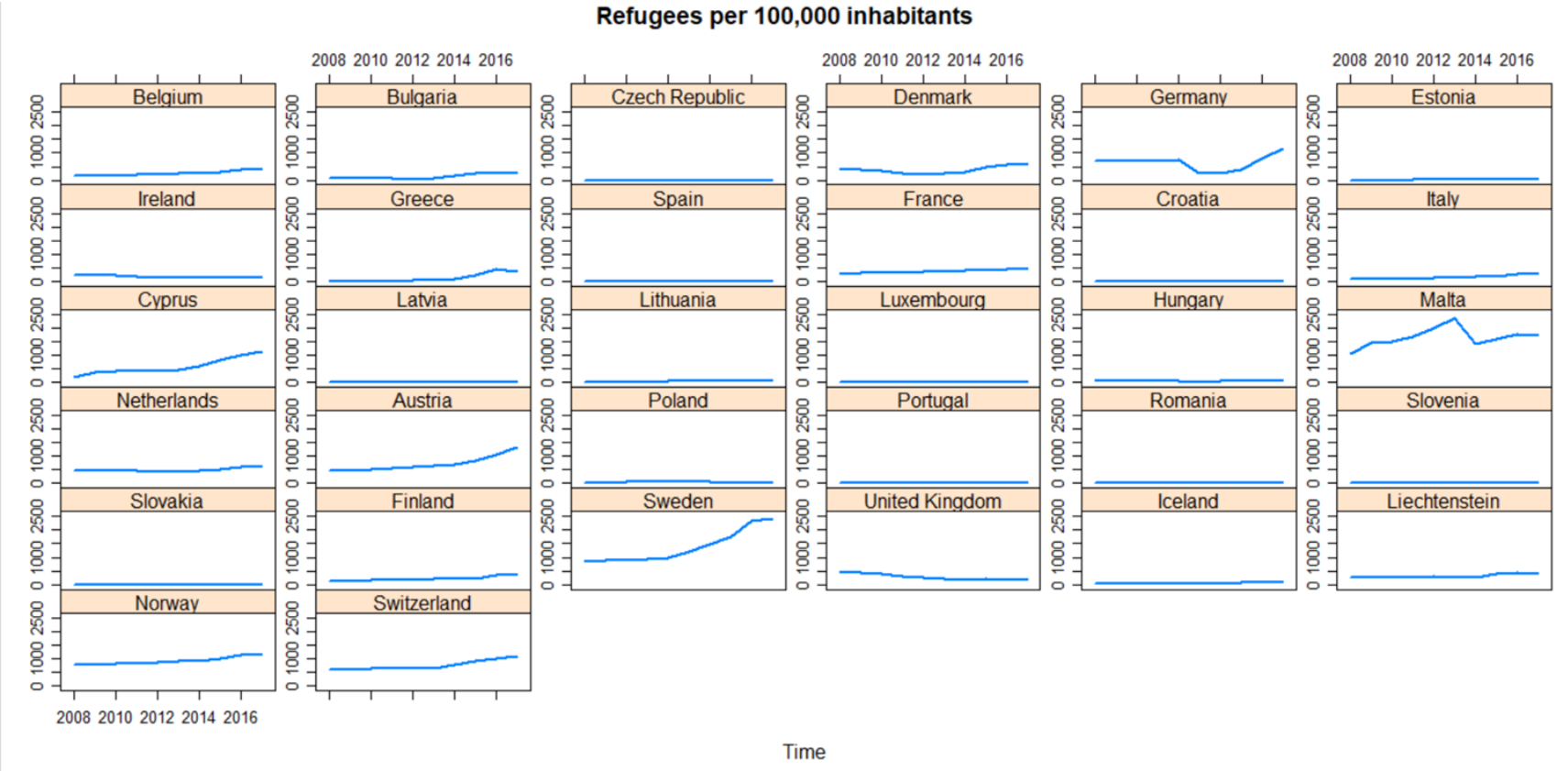
Andamento rifugiati nei paesi UE+ Islanda, Liechtenstein, Norvegia e Svizzera.

Dati come fonte di discussione: un esempio sul fenomeno migranti



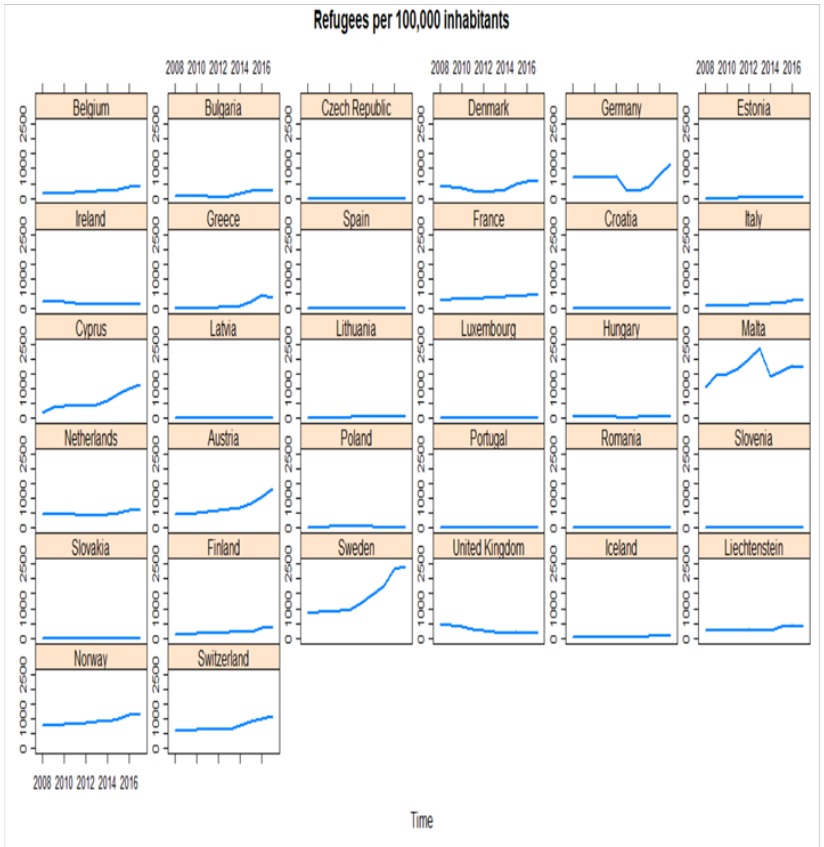
Flusso rifugiati, asse Y uguale per tutti i grafici.

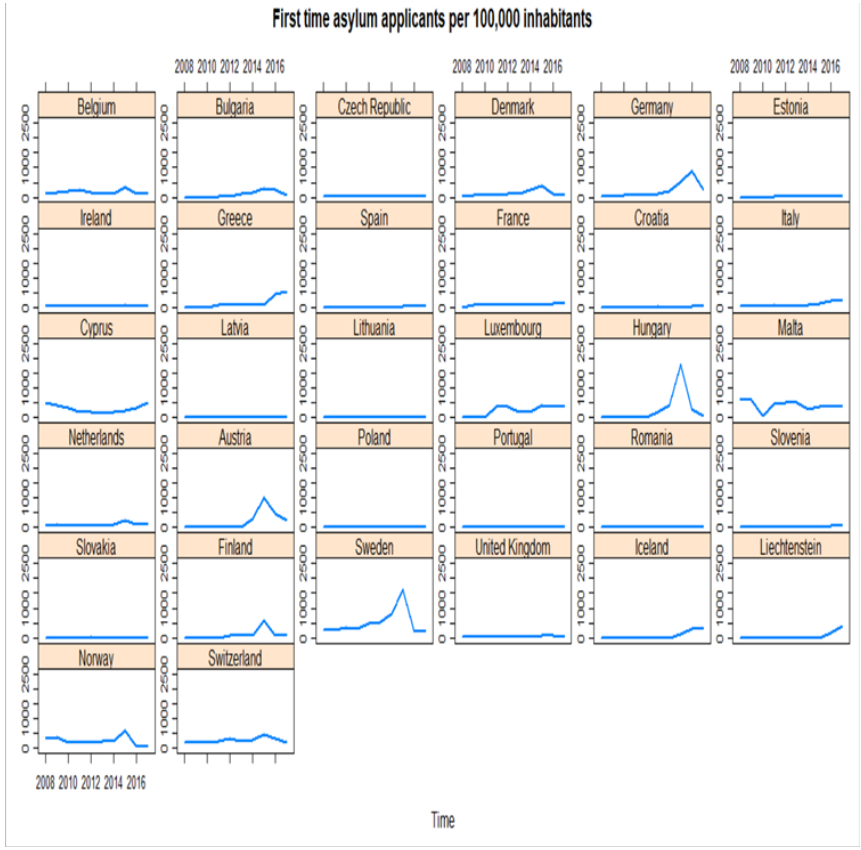
Dati come fonte di discussione: un esempio sul fenomeno migranti



Flusso rifugiati per 100.000 abitanti.

Dati come fonte di discussione: un esempio sul fenomeno migranti





Confronto 'onesto' tra flusso rifugiati e flusso richiedenti asilo

Confusione tra correlazione e relazione causa-effetto: una storiella simpatica

- Subito dopo la fine della seconda guerra mondiale fu notato un incremento significativo della nascita dei bimbi a Londra e, allo stesso tempo, della nidificazione di cicogne
- Dalla raccolta di molti dati al riguardo, da una formale analisi scaturì l'alta concomitanza dell'evento A =*almeno una nascita si è verificata in un caseggiato* e B =*almeno un nido di cicogna è presente sullo stesso caseggiato*

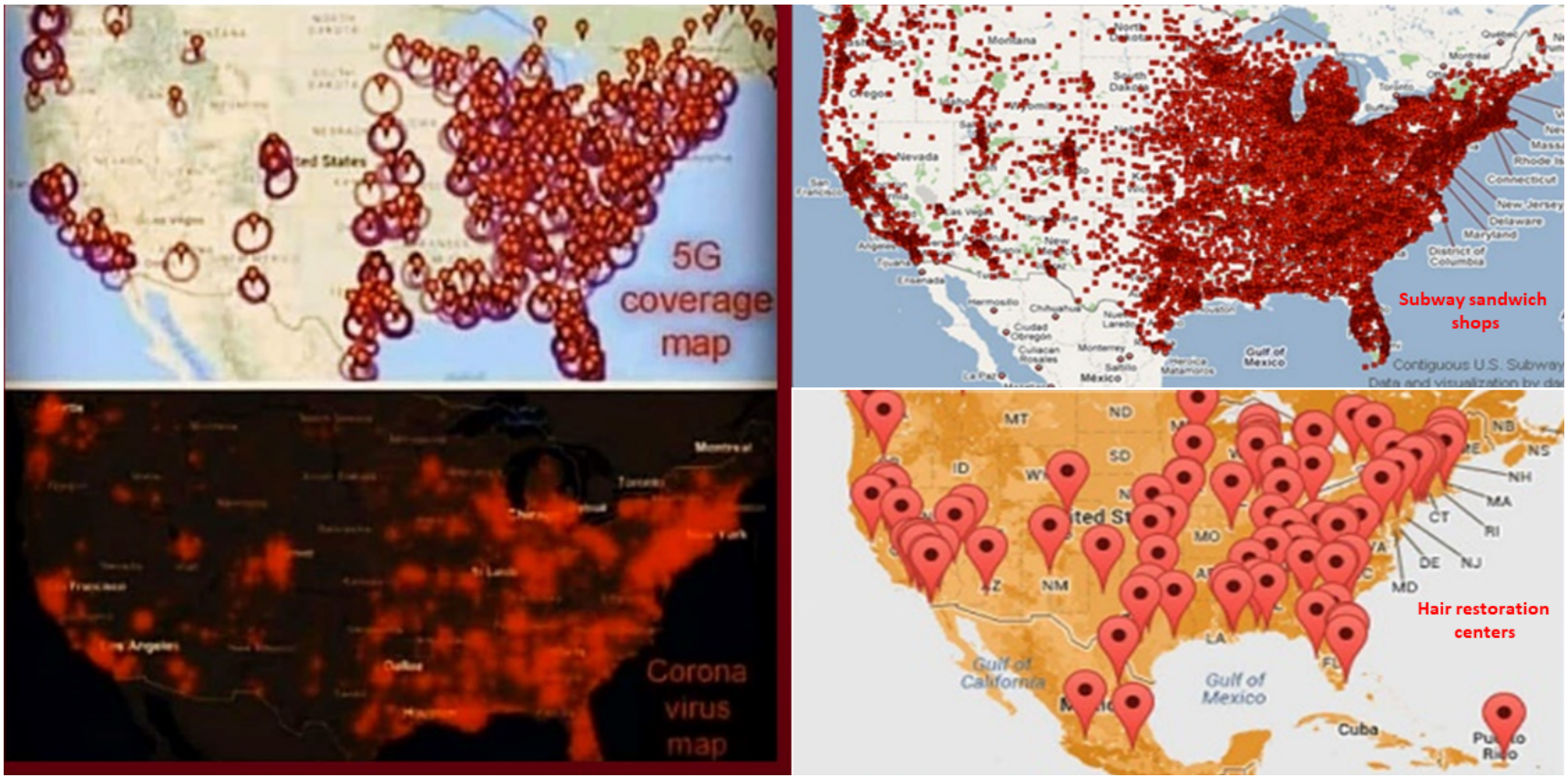
Conclusione: i bimbi vengono effettivamente portati dalle cicogne!!!

Confusione tra correlazione e relazione causa-effetto: una storiella meno simpatica

E' circolata molto su internet questa immagine, che ha convinto moltissime persone che esista una relazione evidente tra diffusione del SARS COVID-2 (Coronavirus) e diffusione della tecnologia 5G

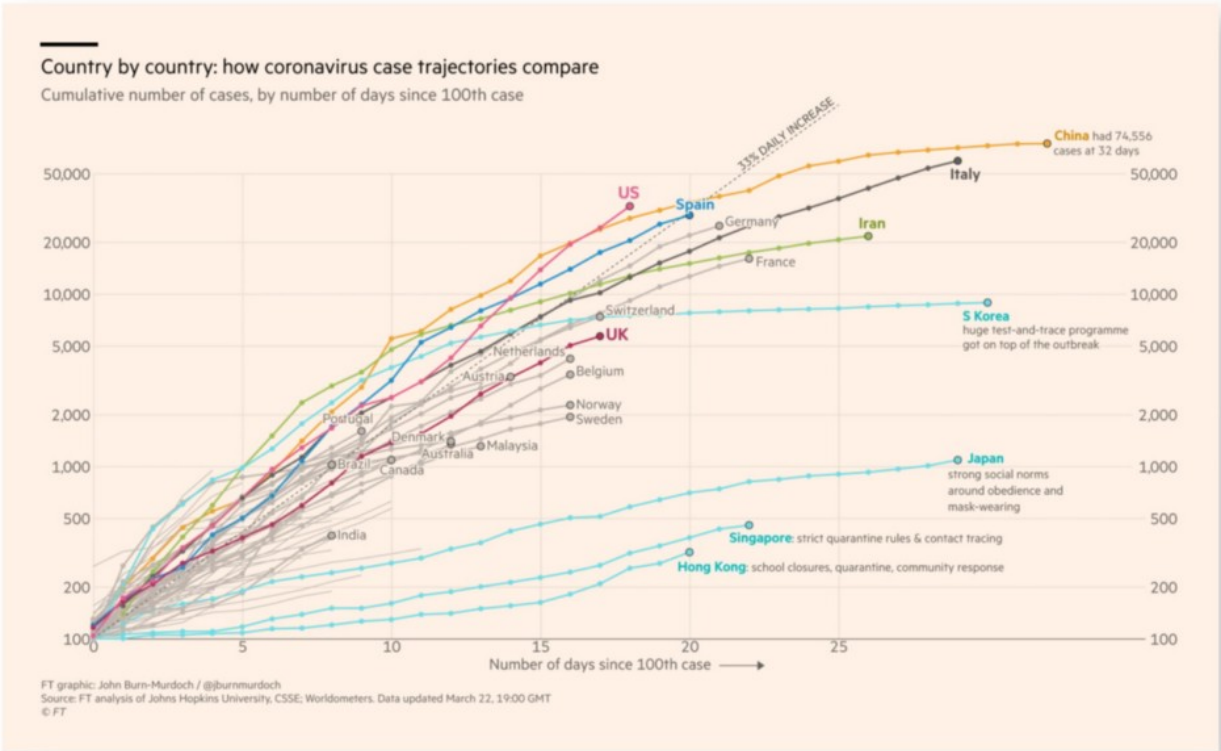
Confusione tra correlazione e relazione causa-effetto: una storiella meno simpatica resa più simpatica

Il 5G causa molte cose negli States....



Confusione tra correlazione e relazione causa-effetto

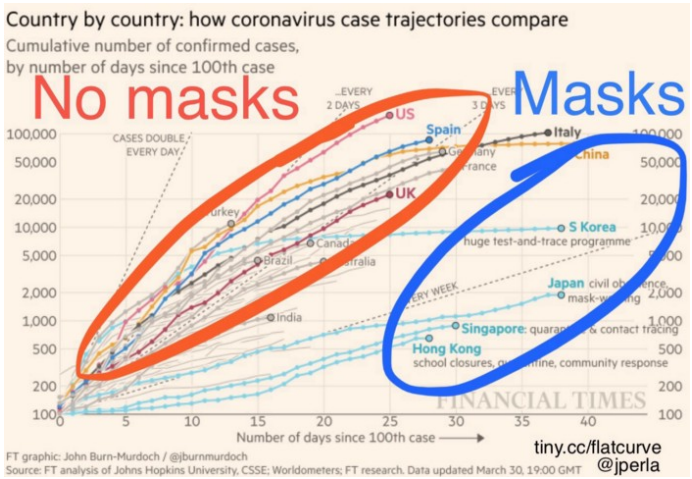
La visualizzazione in basso è stata creata da John Murdoch per il Financial Times. (`https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938`)



L'asse X rappresenta il numero di giorni dal centesimo caso in quel paese, l'asse Y rappresenta il numero cumulativo di casi (in scala logaritmica, quindi le linee rette rappresentano effettivamente una crescita esponenziale) e ogni curva rappresenta un paese diverso

Confusione tra correlazione e relazione causa-effetto

Il problema è che si confonde correlazione con relazione causa-effetto: è possibile giungere a una conclusione corretta da premesse errate ed è possibile accettare che una conclusione sia vera anche se non si accetta ogni argomento per essa.



Un altro commentatore, Joseph Perla, ha aggiunto un paio di cerchi disegnati a mano. L'argomentazione implicita è che le mascherine aiutano ad "appiattire la curva" (cioè abbassare il tasso di crescita del dato cumulativo dei casi), come evidenziato dal fatto che i paesi con utilizzo di mascherine hanno tassi di crescita inferiori rispetto ai paesi senza utilizzo di maschere.

A prescindere dalla legittima opinione di chiunque riguardo a questo tema particolare, e a prescindere dalla mia convinzione personale che l'uso delle mascherine sia un fattore importante per rallentare l'andamento della pandemia, la correlazione non implica causalità.

La variabile x è correlata alla variabile y, non importa quanto sia forte la correlazione, non significa che la variabile x causi la variabile y. E' possibile che la variabile y causi la variabile x. E' possibile che entrambe le variabili siano causate da una terza variabile confondente z

.

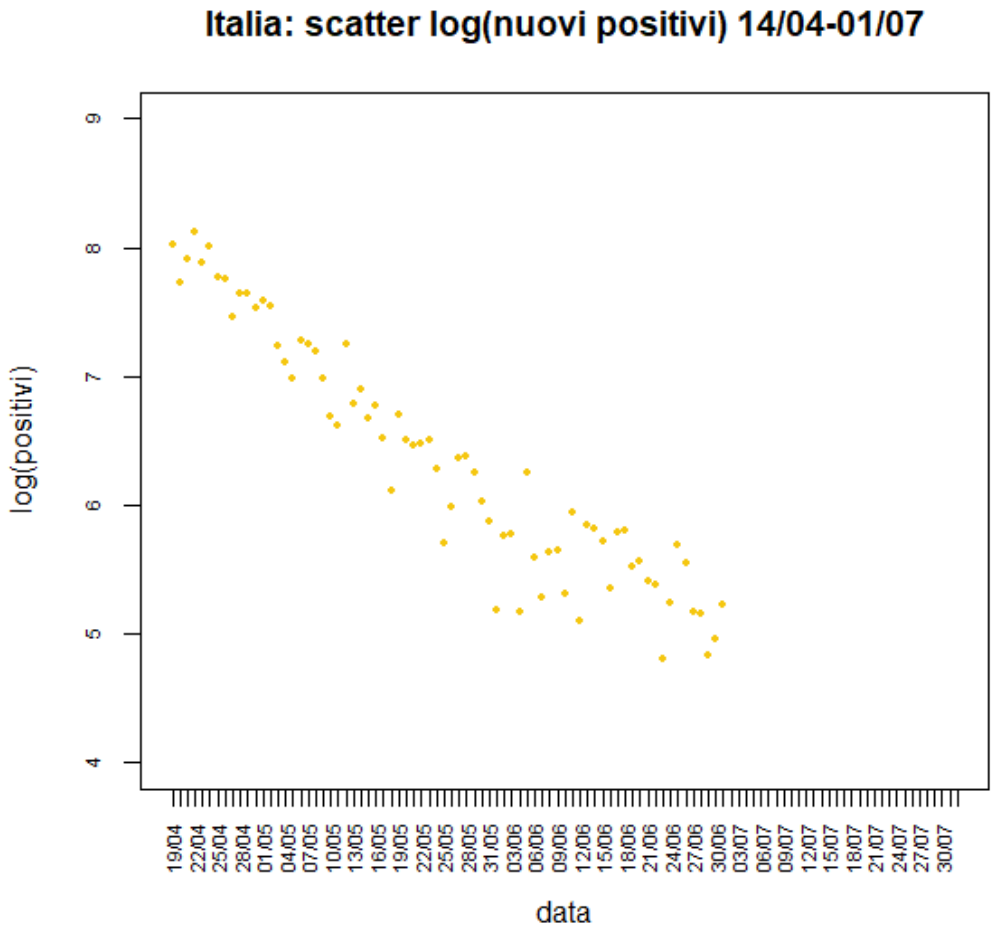
Scienza delle decisioni in condizioni di incertezza

- La scienza statistica è stata definita, tra l'altro, come scienza delle decisioni in condizioni di incertezza.
- I dati, la loro reperibilità, la loro bontà, il loro trattamento, sono la risorsa più importante.
- *Garbage in, garbage out*
- Gli esempi che seguono sono stati elaborati con software *R* utilizzando un P-spline smoother utilizzando un GLM con per dati di conteggio -distribuzione di Poisson-*.
- Data repository: `https://github.com/pcm-dpc/COVID-19`
- Dal dato arriva l'informazione, l'informazione genera e alimenta la conoscenza. Le decisioni spettano al decisore

*
Eilers P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science*. 11, 89-121.;
Currie, I. D., M. Durban, and P. H. C. Eilers (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*. 4, 279-298;
Currie, I. D., M. Durban, and P. H. C. Eilers (2006). Generalized linear array models with applications to multidimentional smoothing. *Journal of the Royal Statistical Society*. Series B. 68, 259-280.

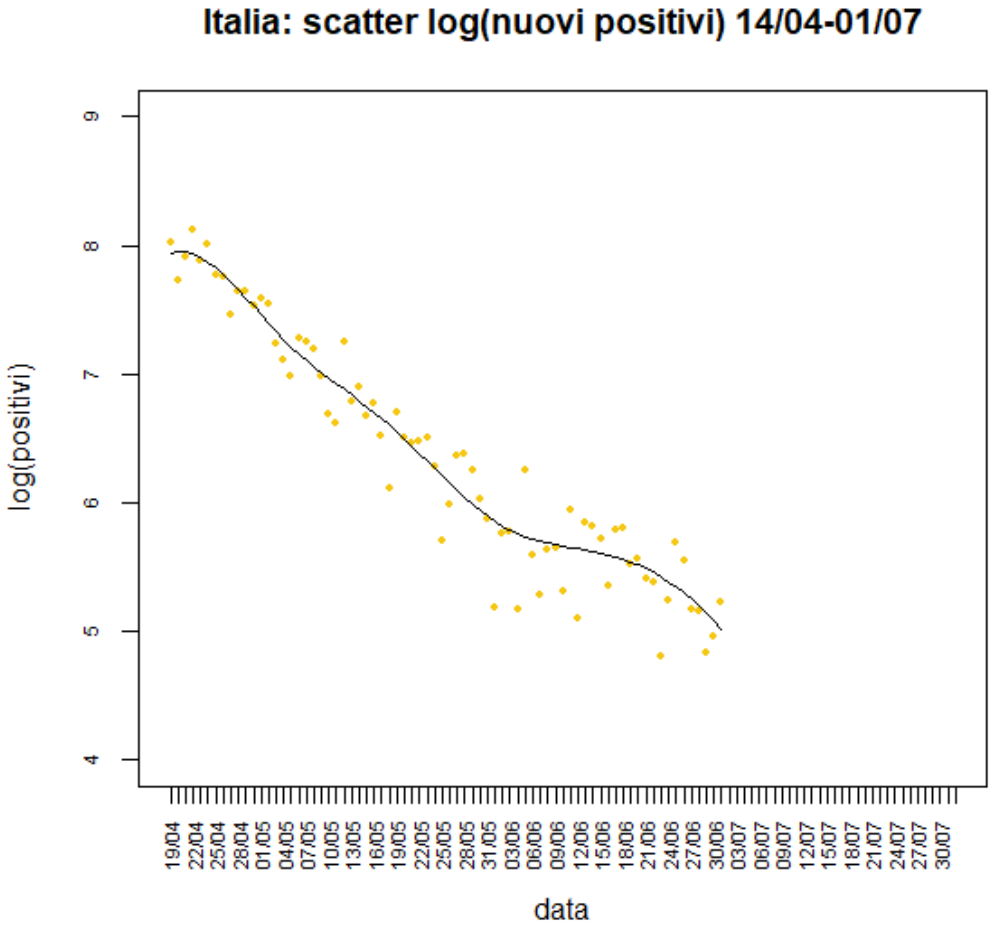
Scienza delle decisioni in condizioni di incertezza

Immaginiamo di trovarci al giorno 1 Luglio. Si osserva lo *scatter-plot* del logaritmo dei nuovi positivi giornalieri dal 19 Aprile.



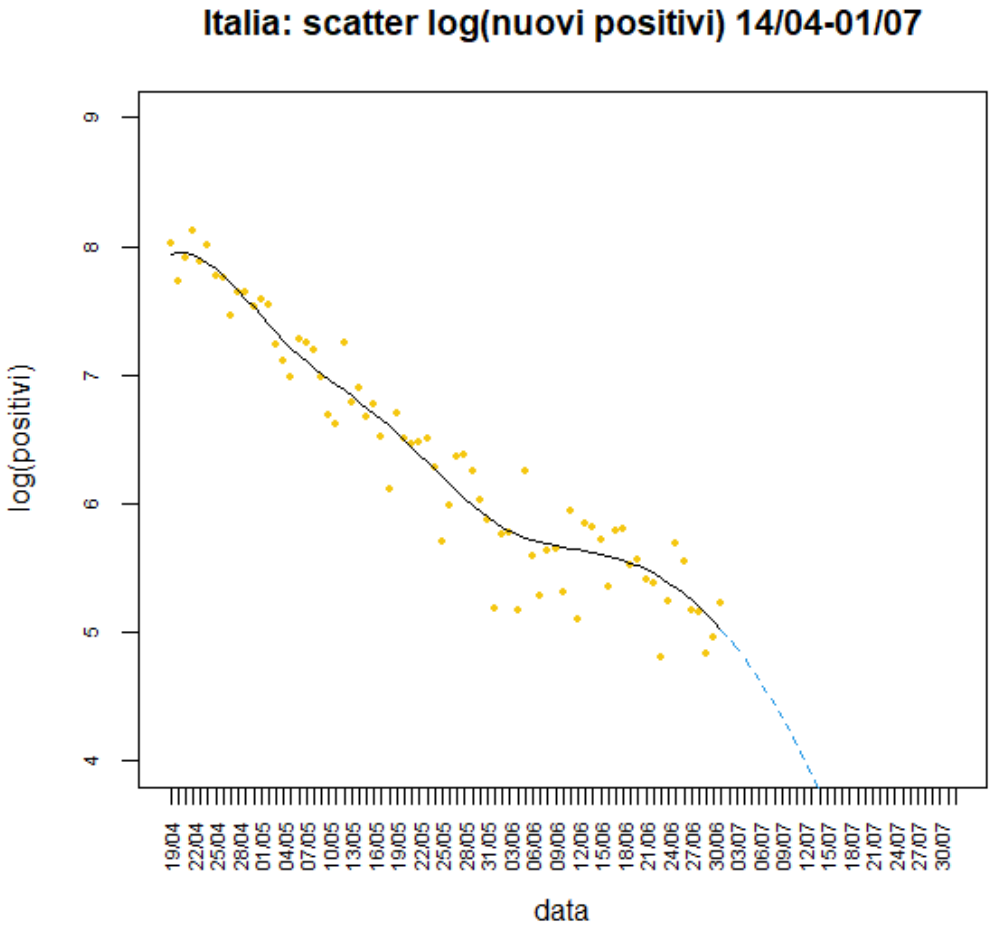
Scienza delle decisioni in condizioni di incertezza

Si stima il modello, $(\mathbb{E}[y|x])$, lo si riporta sul grafico, e valutiamo il *fit* a livello grafico.



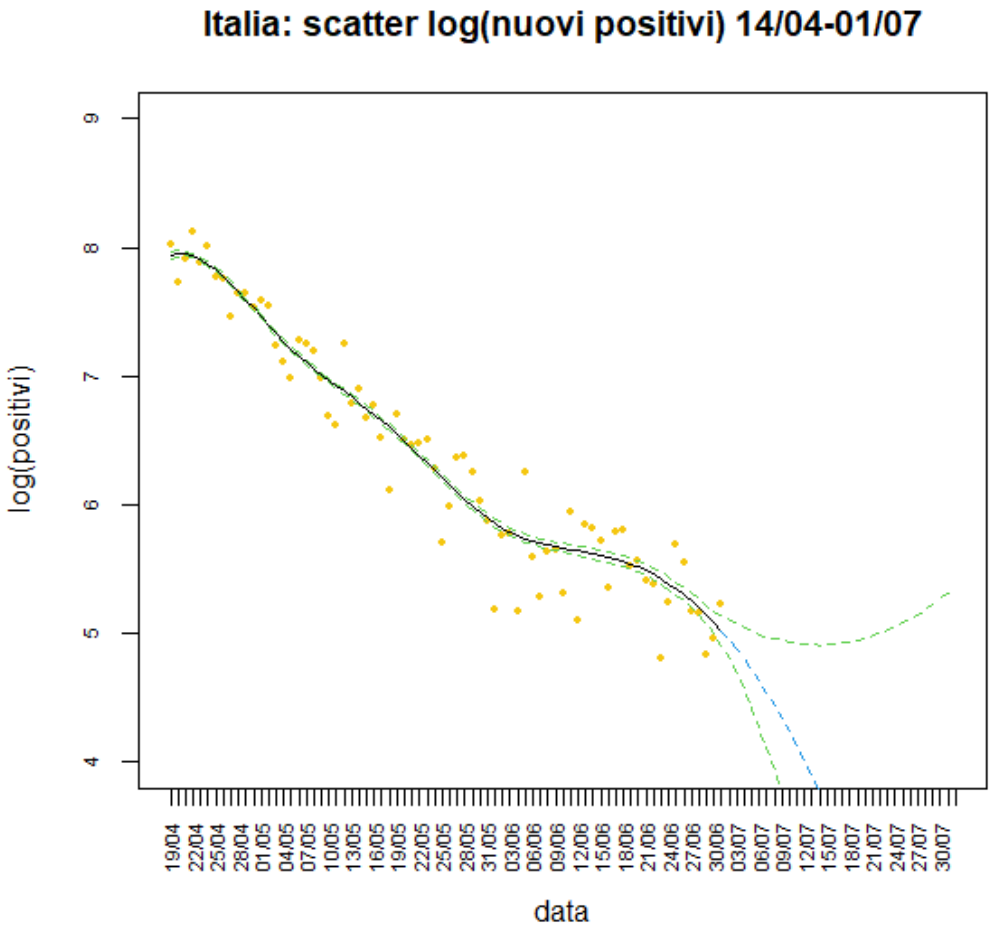
Scienza delle decisioni in condizioni di incertezza

Si procede al *forecasting*, scegliendo l'orizzonte temporale di previsione. Nel nostro caso, poco più che scolastico, l'orizzonte temporale è 30 giorni. Siamo contenti del risultato



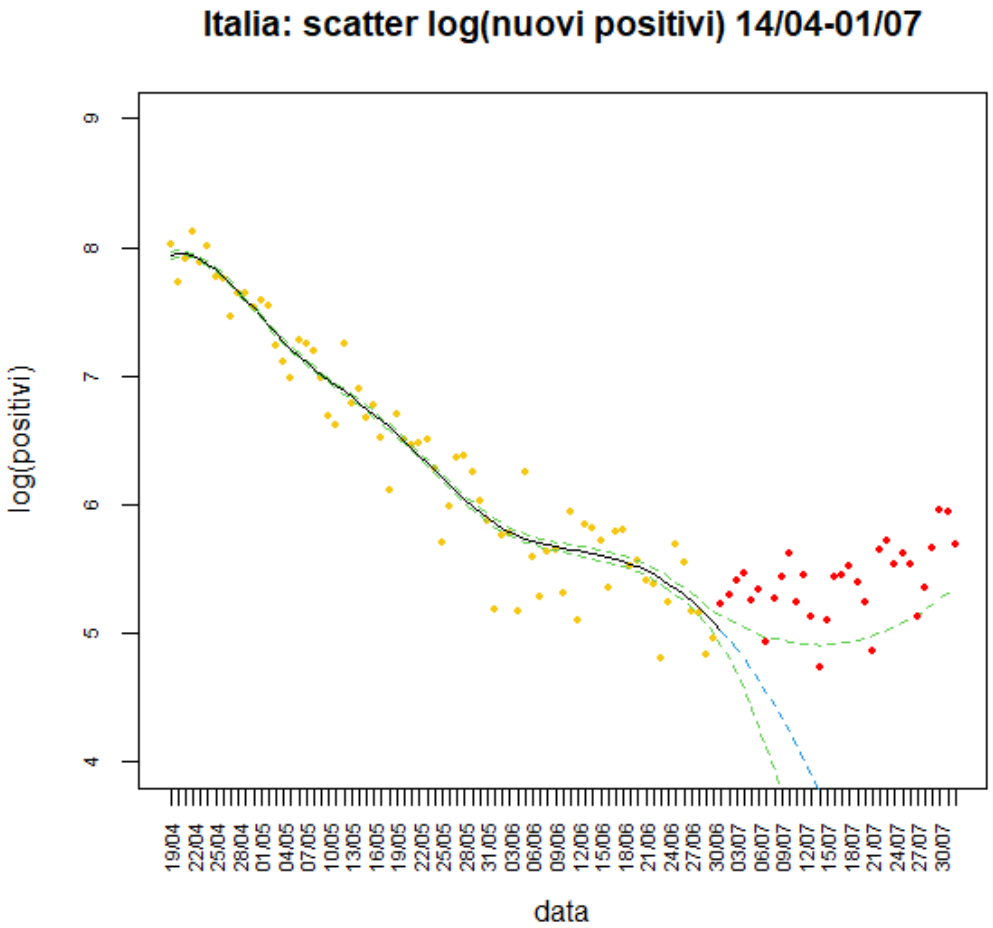
Scienza delle decisioni in condizioni di incertezza

Sappiamo che la stima può essere puntuale o per intervalli. Calcoliamo e riportiamo sul grafico le *bande di confidenza*. Abbiamo scelto un livello di confidenza del 99%. Osserviamo le bande di confidenza, valutiamo il loro andamento. La banda superiore non ci piace

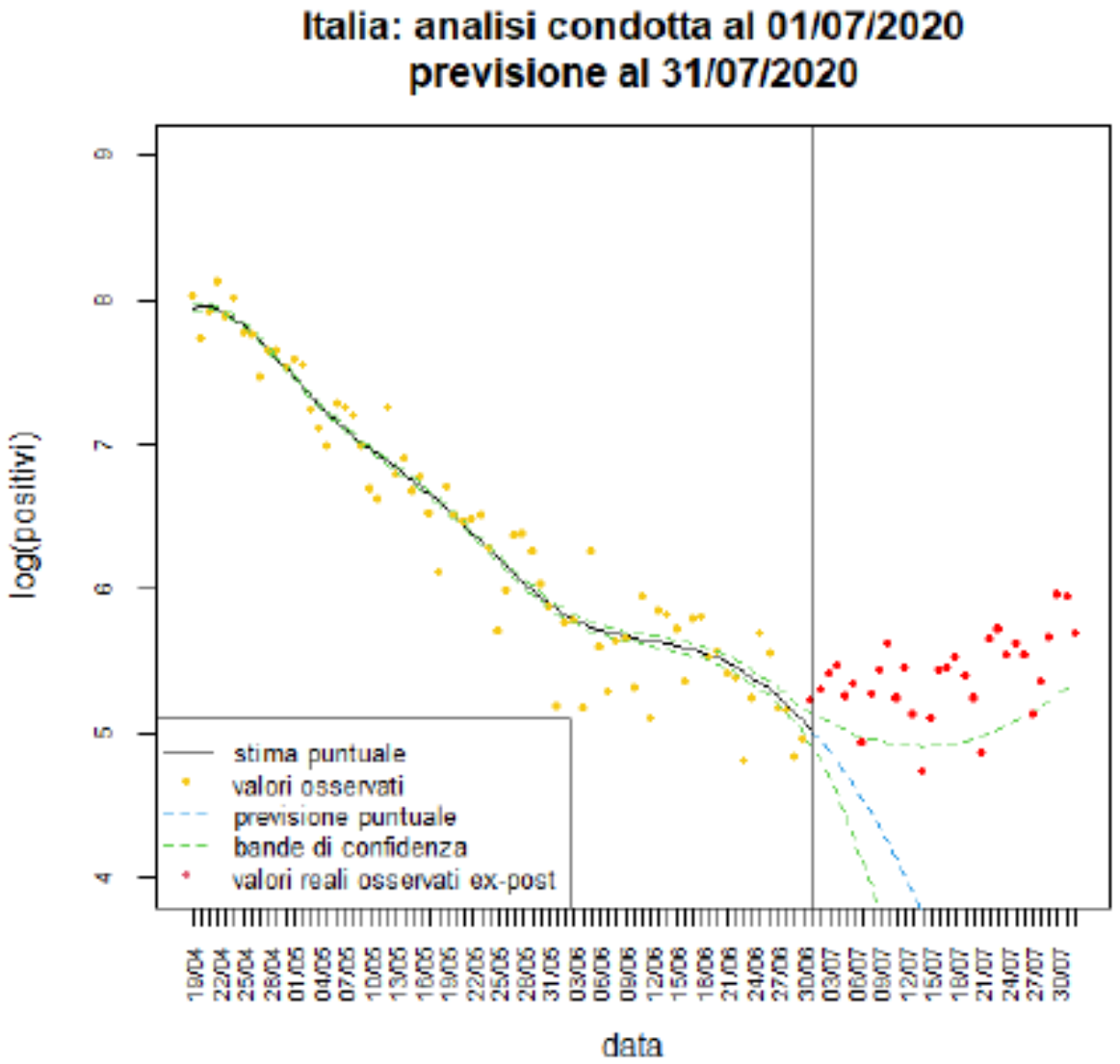


Scienza delle decisioni in condizioni di incertezza

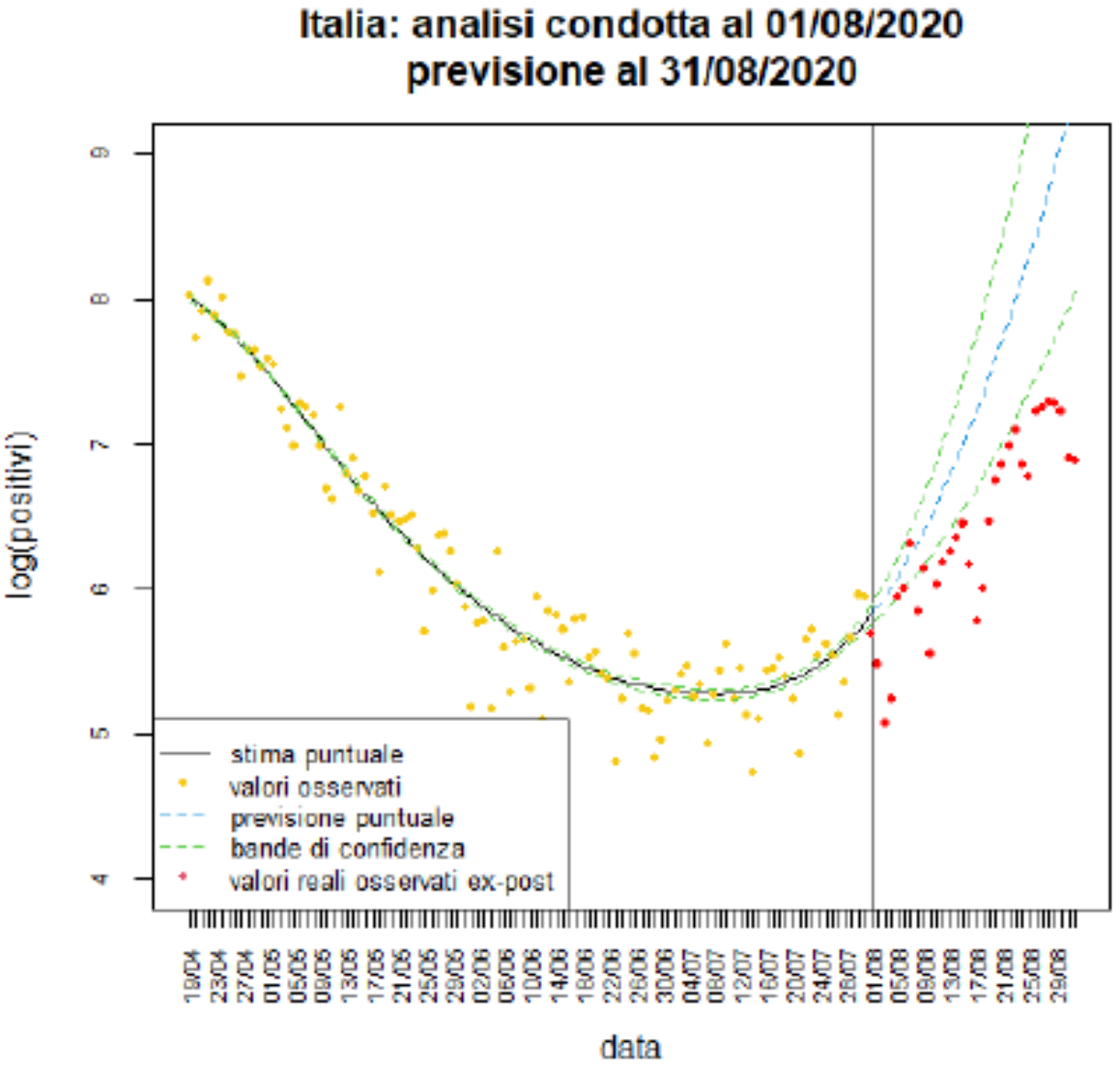
Siamo arrivati al 31 Luglio. Facciamo una valutazione ex-post del risultato. Inseriamo sullo stesso grafico i valori *realmente* osservati. Facciamo le nostre valutazioni. Cosa avevamo pensato il primo Luglio?



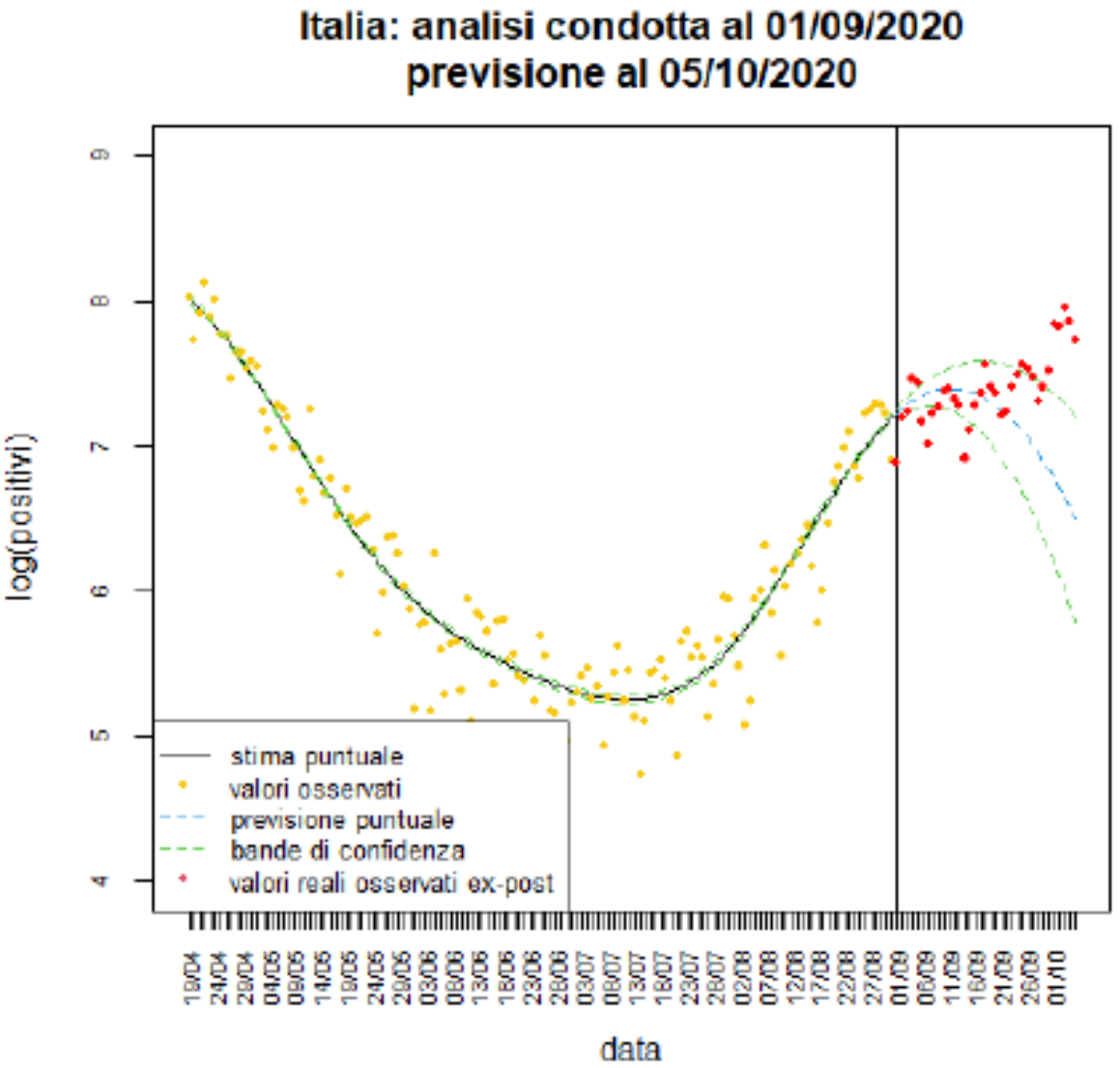
Scienza delle decisioni in condizioni di incertezza



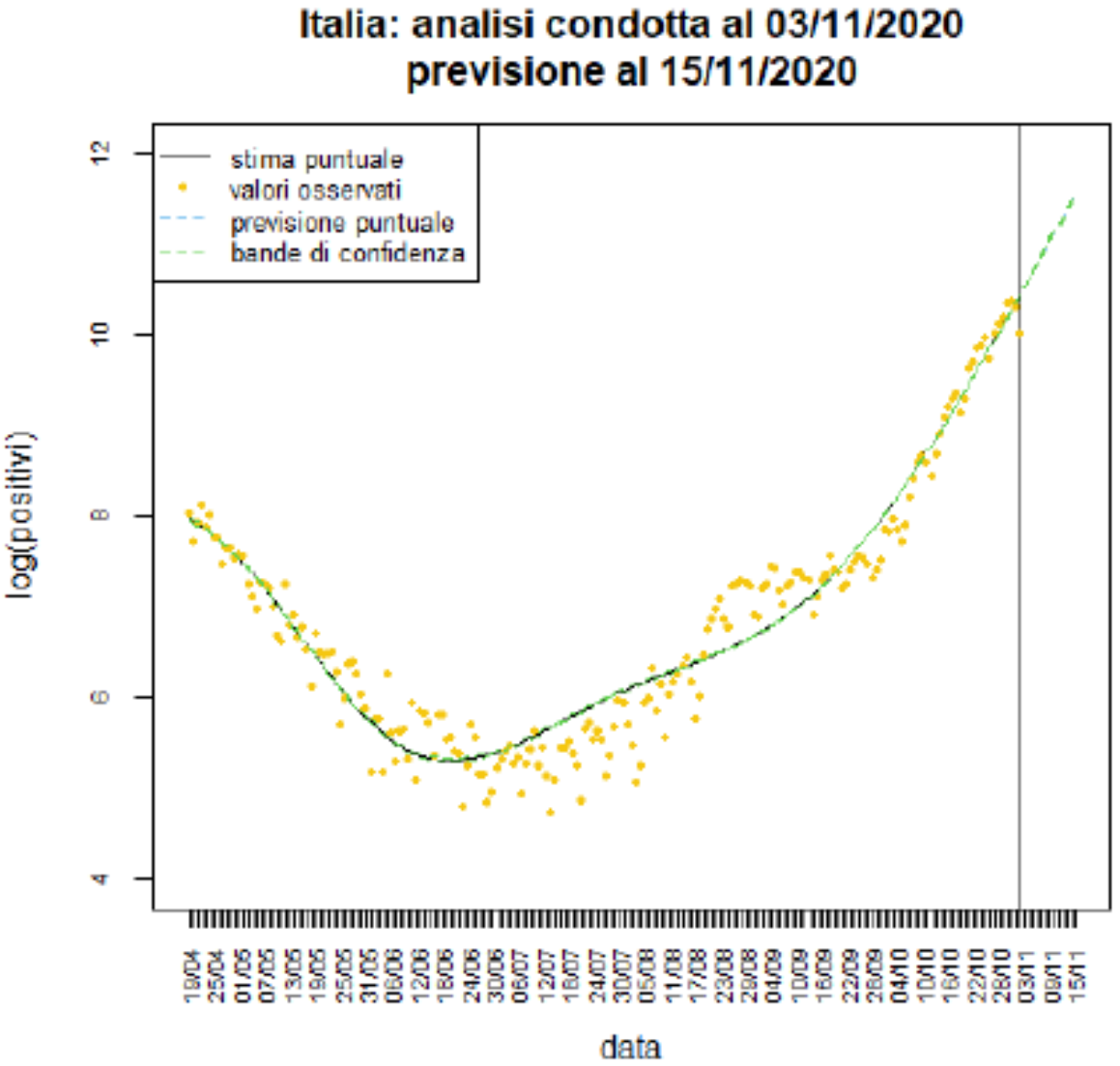
Scienza delle decisioni in condizioni di incertezza



Scienza delle decisioni in condizioni di incertezza



Scienza delle decisioni in condizioni di incertezza



Credibilità e affidabilità fonti di dati

- Le fonti di dati più credibili e affidabili sono (dovrebbero essere) quelle ufficiali.
- Problema 1: non sempre si trovano i dati grezzi, ma risultati aggregati in forma tabellare in formati digitali 'scomodi'
- Problema 2: i *data warehouse*, laddove presenti, non sempre consentono di arrivare a microdati
- Problema 3: sono tutti diventati esperti di *data analysis*, ma nessun (o quasi nessun) *data analyst* formato come tale è preposto alla raccolta e diffusione di dati, soprattutto negli uffici pubblici
- *Garbage in, garbage out*: do you remember?

Credibilità e affidabilità fonti di dati

- Diffidare sempre di informazioni date sulla base di dati senza che sia indicata la fonte degli stessi
- Anche se la fonte dei dati fosse indicata, dovrebbe essere possibile accedere ai dati
- Se non fosse possibile accedere ai dati (copyright), non dovrebbero mai mancare informazioni sul campione intervistato (numerosità e popolazione di riferimento), sul piano di campionamento adottato (probabilistico?, ragionato?, stratificato?, ponderato?, ...)
- Finalmente la statistica di base si insegna nelle scuole a partire dalla prima media. Educare sempre di più a non farsi influenzare dagli 'effetti speciali' degli strumenti di visualizzazione.
- *Garbage in, garbage out*: do you remember?

Credibilità e affidabilità fonti di dati

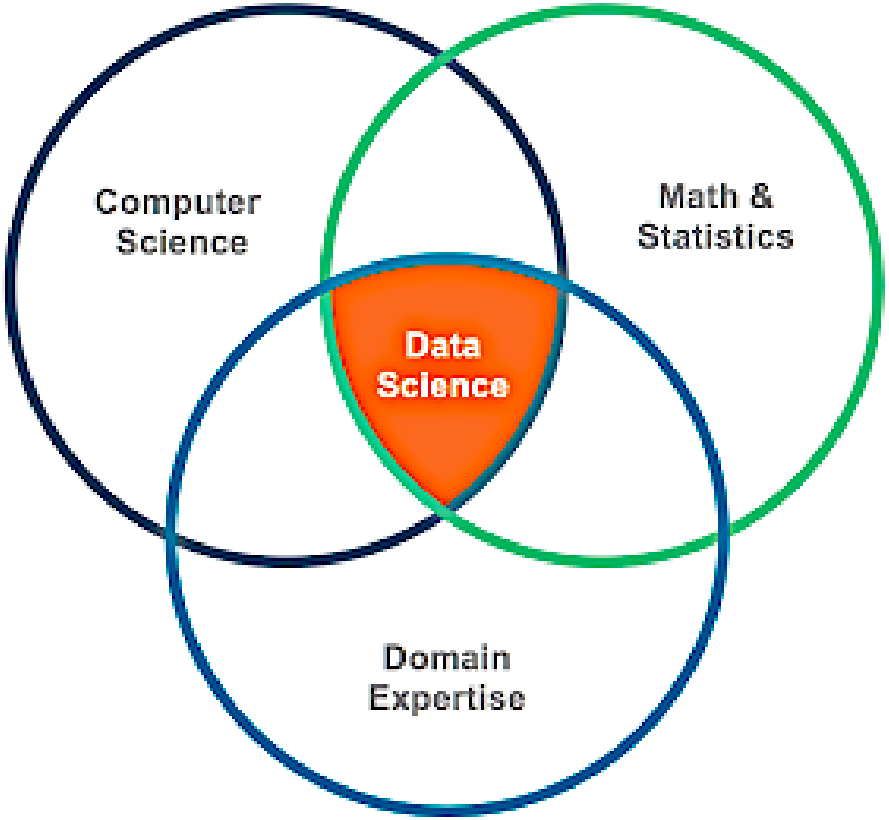
- Siamo sommersi da *fake news*. Cosa è una fake news? Qualcuno ha cercato di darne formale definizione *Fake news is a news article that is intentionally and verifiably false.*(Shu K., Sliva A., Wang S., Tang J., and Liu H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* vol. 19(1), 22–36.)
- Come identificare una fake news?
- Statistica e Data Science

Statistica e Data Science

When the Lord created the world and people to live in it - an enterprise which, according to modern science, took a very long time - I could well imagine that He reasoned with Himself as follows: 'If I make everything predictable, these human beings, whom I have endowed with pretty good brains, will undoubtedly learn to predict everything, and they will thereupon have no motive to do anything at all, because they will recognise that the future is totally determined and cannot be influenced by any human action. On the other hand, if I make everything unpredictable: they will gradually discover that there is no rational basis for any decision whatsoever and, as in the first case, they will thereupon have no motive to do anything at all. Neither scheme would make sense. I must therefore create a mixture of the two. Let some things be predictable and let others be unpredictable, They will then, amongst many other things, have the very important task of finding out which is which.

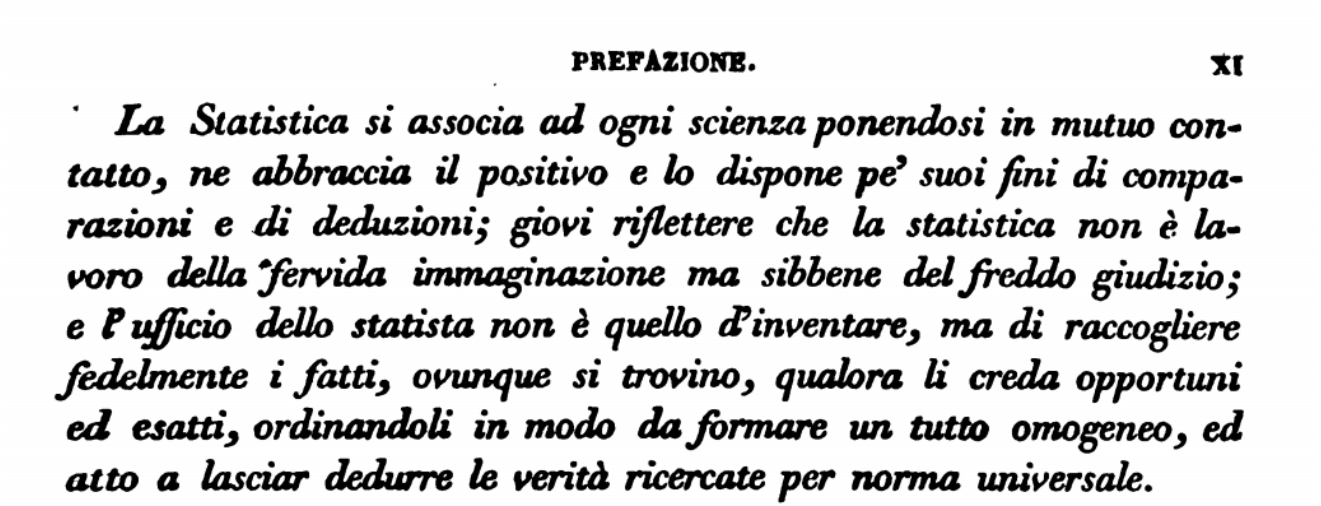
E. F. Schumaker. *Small is beautiful*

Statistica e Data Science



Statistica e Data Science

Mi piace concludere riportando quanto scritto da Ferrario (A.D. 1838)



PREFAZIONE. XI

La Statistica si associa ad ogni scienza ponendosi in mutuo contatto, ne abbraccia il positivo e lo dispone pe' suoi fini di comparazioni e di deduzioni; giovi riflettere che la statistica non è lavoro della fervida immaginazione ma sibbene del freddo giudizio; e l'ufficio dello statista non è quello d'inventare, ma di raccogliere fedelmente i fatti, ovunque si trovino, qualora li creda opportuni ed esatti, ordinandoli in modo da formare un tutto omogeneo, ed atto a lasciar dedurre le verità ricercate per norma universale.