

# DATI, FONTI, DECISIONI

Author Giovambattista Vieri

<https://www.linkedin.com/in/giovambattista-vieri-26799/>

(c) 2020 All rights reserved

License: GNU FDL

# DATI E FONTI

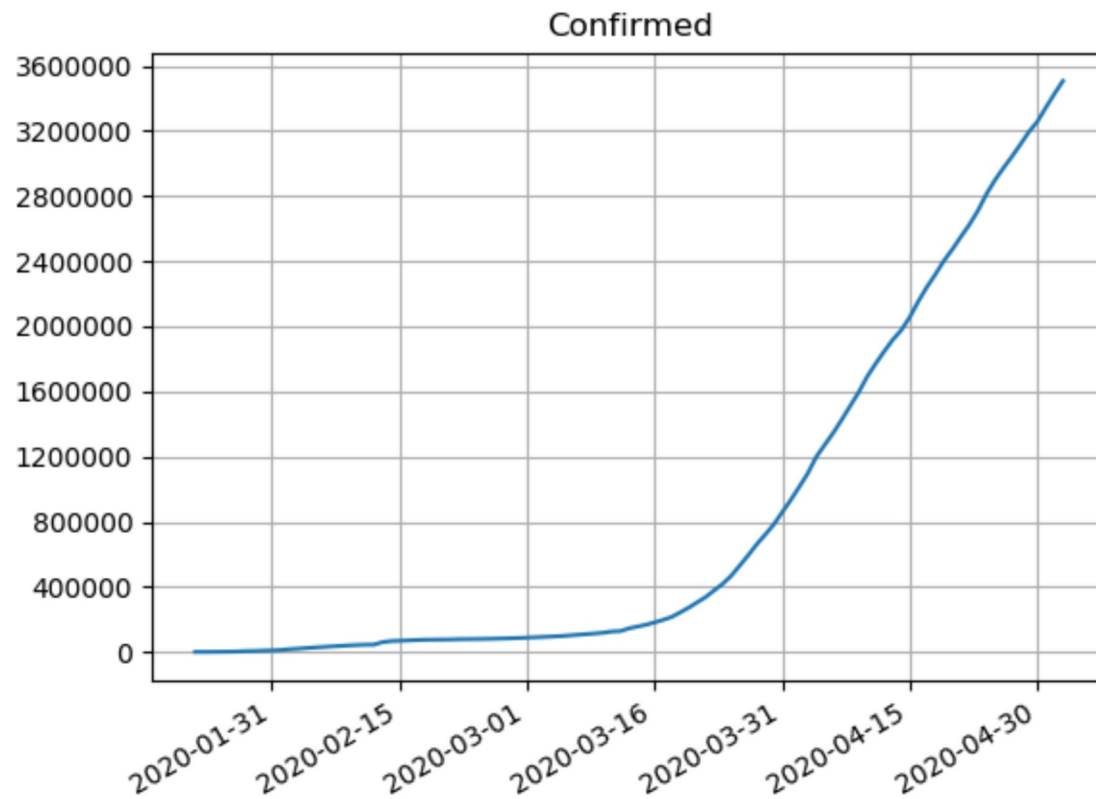
- Attualmente viviamo in un oceano di dati.
- Per gli altri diventiamo anche noi dei dati da collezionare. Almeno, lo diventano le nostre azioni, soprattutto quelle di interesse economico.
- Questi dati per essere usati vengono spesso aggregati e poi presentati a dei decisori che possono essere umani o meno.

# Presentazione dei dati:

- Tipicamente nei convegni o nei documenti tecnici dei grafici sono presenti.
- Da anni si usano infografiche geo-referenziate per mostrare le aggregazioni piu' disparate.

# Un esempio di grafico

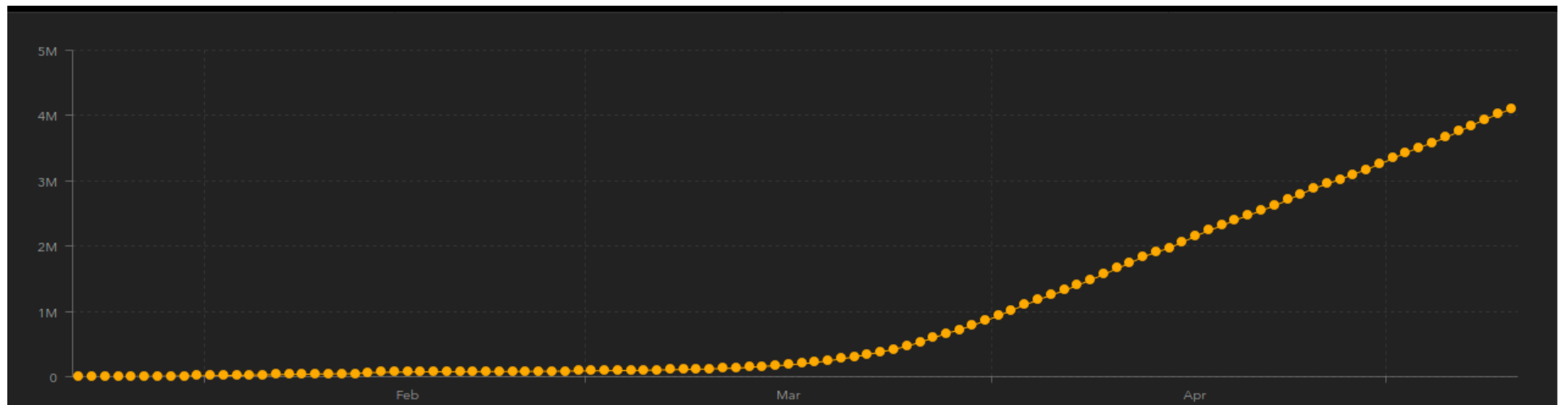
- Che reincontreremo dopo:



# Cosa mostra?

- Sicuramente un fenomeno
- Che inizia
- Cresce e, a un certo punto aumenta la sua velocità'.

# Altro fenomeno



# Quale fenomeno

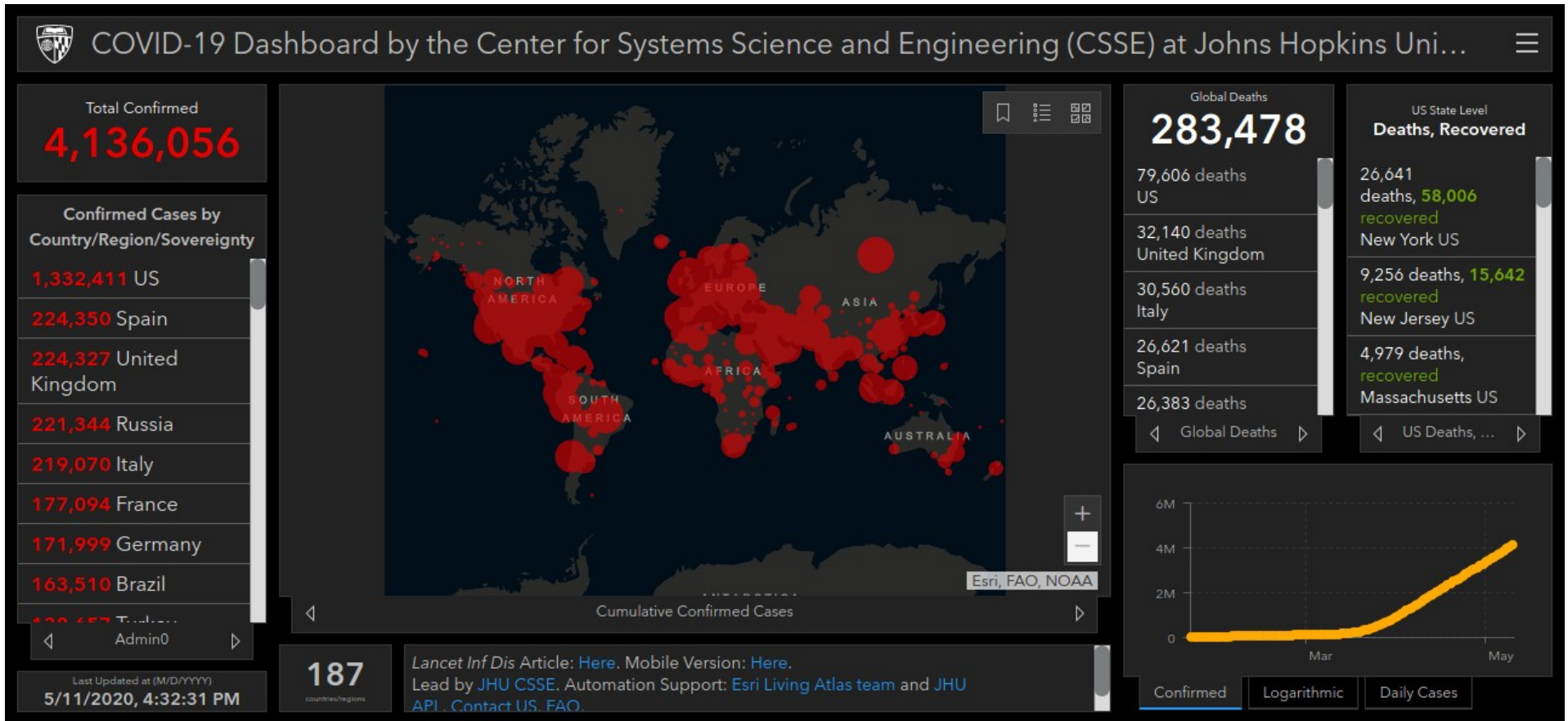
- Cresce piu' rapidamente ?

# Covid19

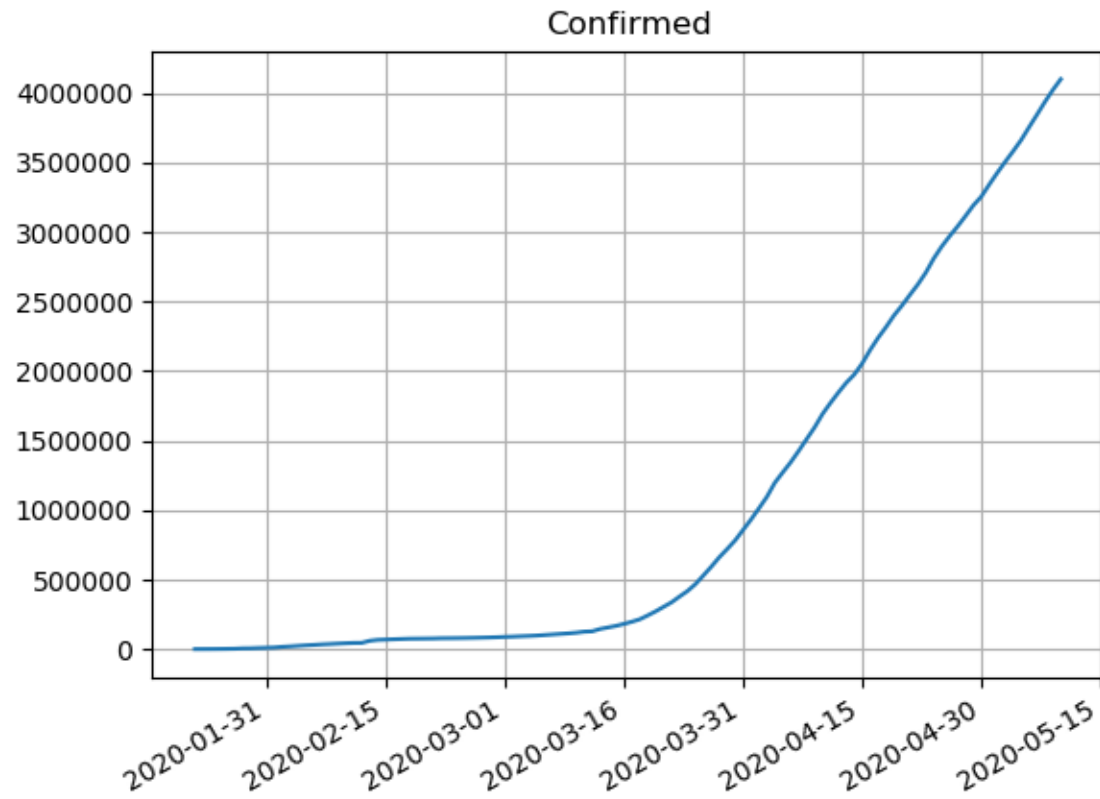
- Come non parlarne?
- Tante fonti. Tanti dati. Tanto Rumore.
- Per fortuna gli stati e le organizzazioni internazionali sono intervenute insieme a enti/organizzazioni note e degne di fiducia.
- CDC, ministeri nazionali, WHO/OMS, universita' , Facebook, Etc.
- Personalmente ho scelto come fonte privilegiata:
- JOHNS HOPKINS UNIVERSITY.



# dashboard



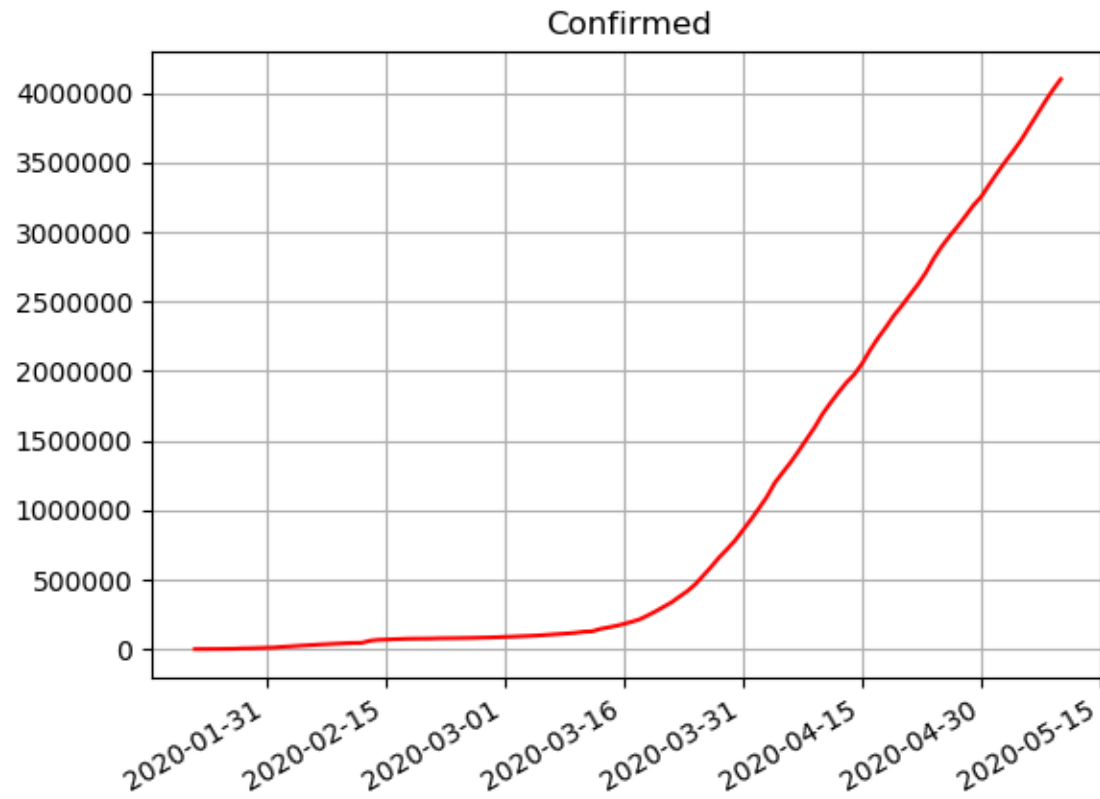
# Grafico:



# Domande e considerazioni

- Fenomeni diversi?
- Quale sembra crescere più rapidamente ?
- Quale sembra più soggetto a impressionare il lettore ?

# Forse questo ?



# Ora

- Se ritenete che la grandezza che sale di più sia l'ultima:
  - Le proporzioni contano
  - I colori contano
  - La legenda e' illegibile o non la guarda nessuno.
- Almeno nei grafici.

# Autopromozione ?

- <https://github.com/gvieri/coviutil/>
- Un semplice repository con utility in python che usano i dati della nota universita' americana per ricavare grafici e tabelle.
- Se volete risparmiare tempo:
- <https://colab.research.google.com/github/gvieri/coviutil/blob/master/covid19glodata.ipynb>

# Colab:

- E' un servizio di google che vi dà la possibilità di far girare script in python
- Collegati eventualmente al vostro 'Drive'
- Usando python GPU TPU on demand
- E senza il sistemista che deve installare tutto questa roba sul vs computer e gestire i conflitti :-)

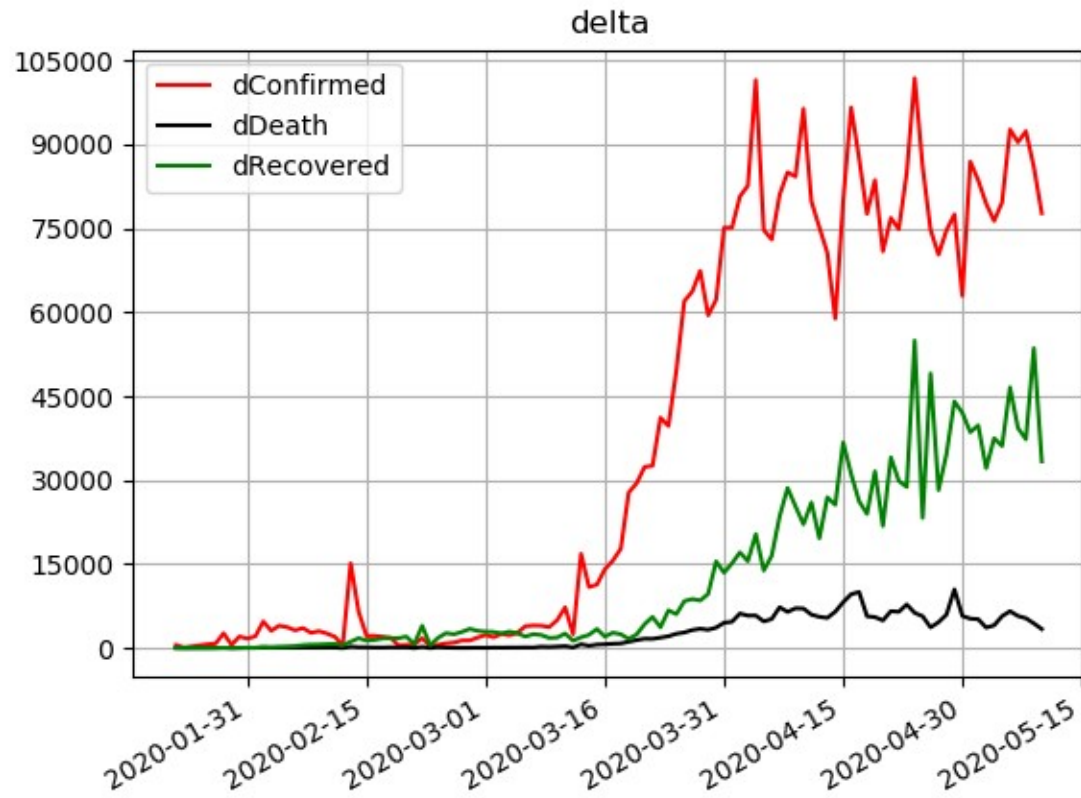
# Non era autopromozione

- Avete visto I grafici ?
- Andate a vedere I tabulati.
- Guardate verso il 3 marzo.. riduzione degli incrementi ...
- Ma anche il 12 febbraio ...



02-07-2020	34392	719	2011	3574	85	524	0.021	0.058
02-08-2020	37121	806	2616	2729	87	605	0.022	0.07
02-09-2020	40151	906	3244	3030	100	628	0.023	0.081
02-10-2020	42763	1013	3946	2612	107	702	0.024	0.092
02-11-2020	44803	1113	4683	2040	100	737	0.025	0.105
02-12-2020	45222	1118	5150	419	5	467	0.025	0.114
02-13-2020	60370	1371	6295	15148	253	1145	0.023	0.104
02-14-2020	66887	1523	8058	6517	152	1763	0.023	0.12
02-15-2020	69032	1666	9395	2145	143	1337	0.024	0.136
02-16-2020	71226	1770	10865	2194	104	1470	0.025	0.153
02-17-2020	73260	1868	12583	2034	98	1718	0.025	0.172
02-18-2020	75138	2007	14352	1878	139	1769	0.027	0.191
02-19-2020	75641	2122	16121	503	115	1769	0.028	0.213

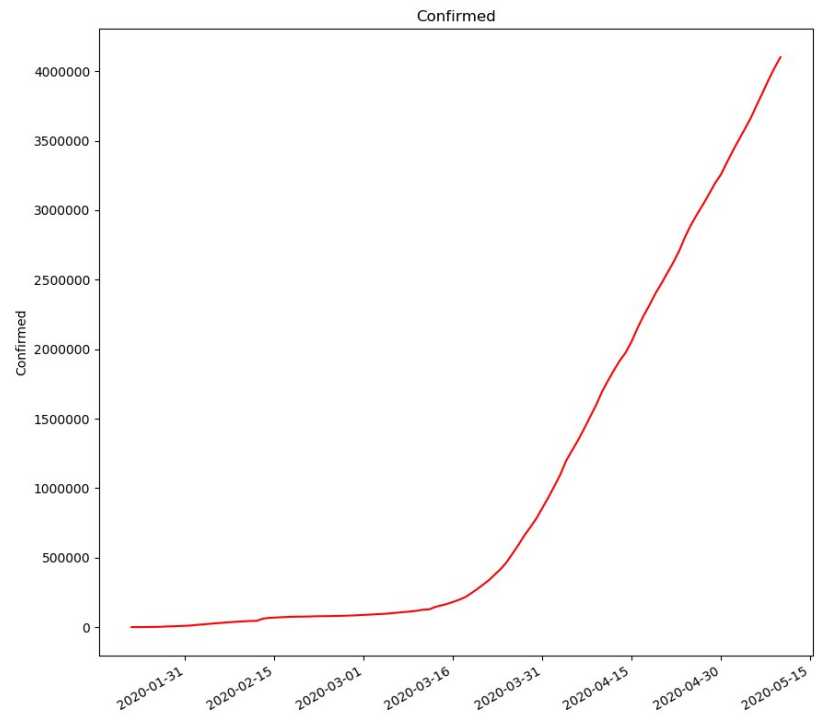
# Visto che parliamo di incrementi:



# Vi lascio

- Il compito di immaginare la versione 'con medie mobili'
- Ovvero senza 'le punte'

# Torniamo al grafico



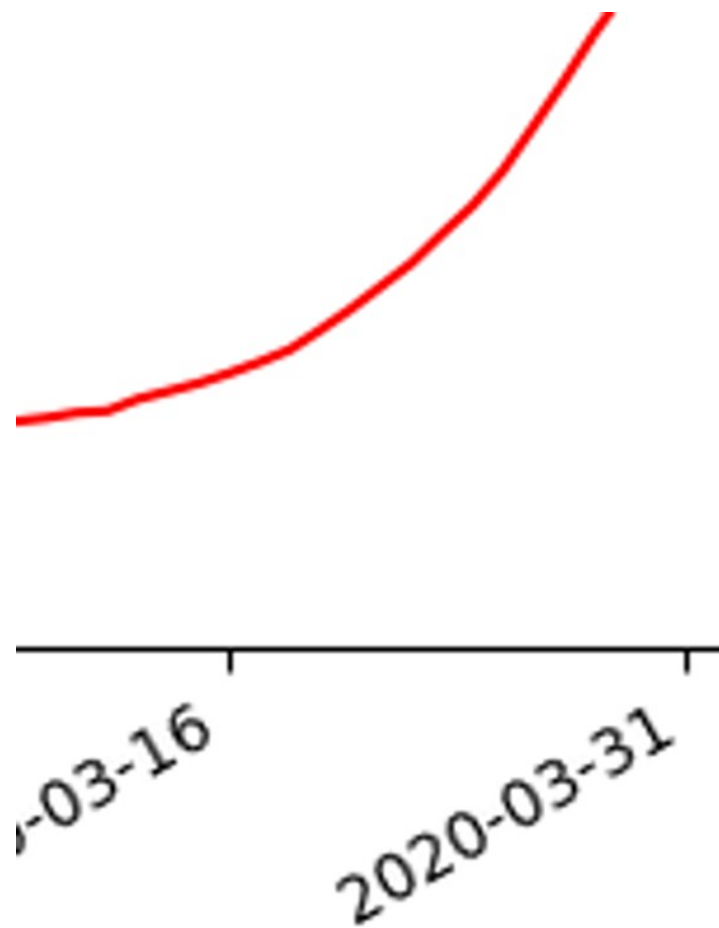
Consideriamone una arbitraria parte



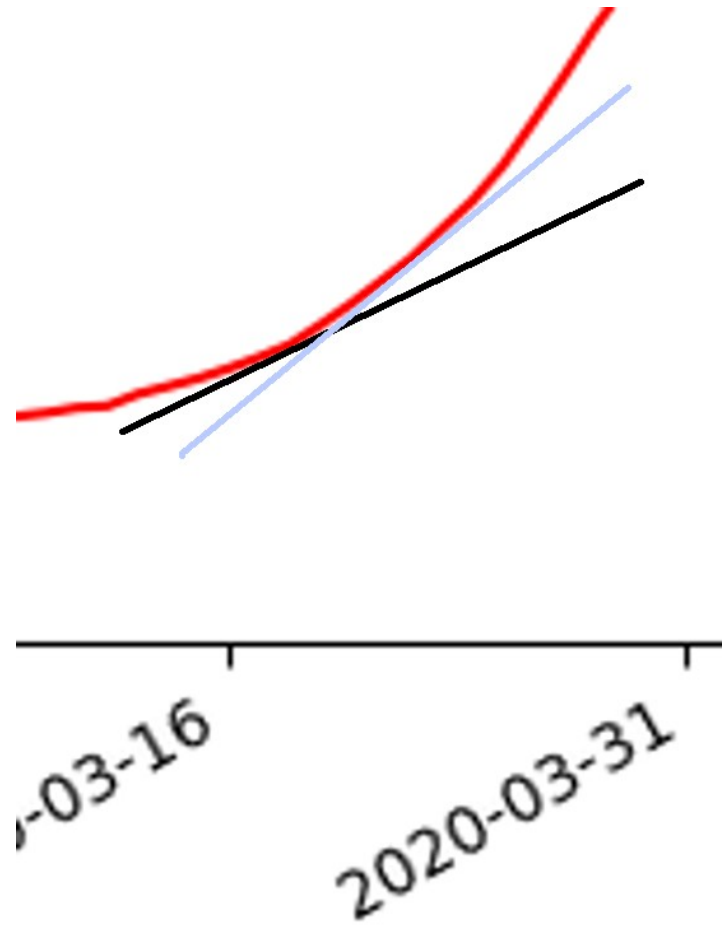
Torturiamo.  
Interrompiamolo. Aggiungiamo una  
retta per il trend:



'massaggiamo'



e...





# tecnicità

- Ho usato solo rette, ovvero equazioni semplici tipo  $y=ax+b$ .
- Usare modelli più complessi era possibile (forse doveroso) ma senza nessun guadagno ai fini della presentazione.
- Sklearn, scikit-learn, pytorch etc. possono dare molte soddisfazioni.

# Conclusioni.

- Le fonti sono tutto. TUTTO.
- La fonte originale come vedete è fondamentale.
- Poi occorre controllare I dati. Se osservate Il codice noterete diversi modi di processare le date. Sono cambiate in corso d'opera. Quindi anche le fonti migliori vanno controllate.

# I dati

- I dati sono fondamentali ma, non vanno creduti ciecamente. Sono l'anello debole della catena.
- Vanno presentati adeguatamente. Visto come cambiare scala e colore impatta sulle vostre 'sensazioni'?
- Ove possibile vanno usati colori culturalmente standard: rosso per pericolo, nero per irreparabile e verde per lieto fine.

# Le decisioni

- Ponetevi nei panni di un decision maker:
- Da quel grafico deve decidere cosa fare per la sua attività industriale o peggio.
- Quindi e' opportuno controllare come i dati vengono presentati. Non solo la fonte o integrità.

# previsioni

- E' molto piu' facile decidere da un bouquet di previsioni automaticamente calcolate.
- Questo però richiede ancora maggiore attenzione su dove cominciare e finire a scrivere il modello.

# Ricordate:

- IF YOU TORTURE THE DATA LONG ENOUGH, IT WILL CONFESS TO ANYTHING
- 
- Se massaggiare i vostri dati abbastanza a lungo, vi diranno tutto quel che volete.

Fonti e integrità

**NON SONO PIÙ ABBASTANZA**

# Dal vivo e domande

- Disclaimer: ho usato il tema covid perchè credo tutti abbiamo visto abbastanza grafici, e, modelli per avere una idea delle tecniche usate.
- <https://colab.research.google.com/github/gvieri/coviutil/blob/master/covid19glodata.ipynb>

<https://www.linkedin.com/in/giovambattista-vieri-26799/>