

COVID-19 NEGLI USA

DATA SCIENCE LAB IN BIOSCIENCES – A.A 2020/2021



MATTEO CAMPIRONI

matricola 801850

SERENA DI MAGGIO

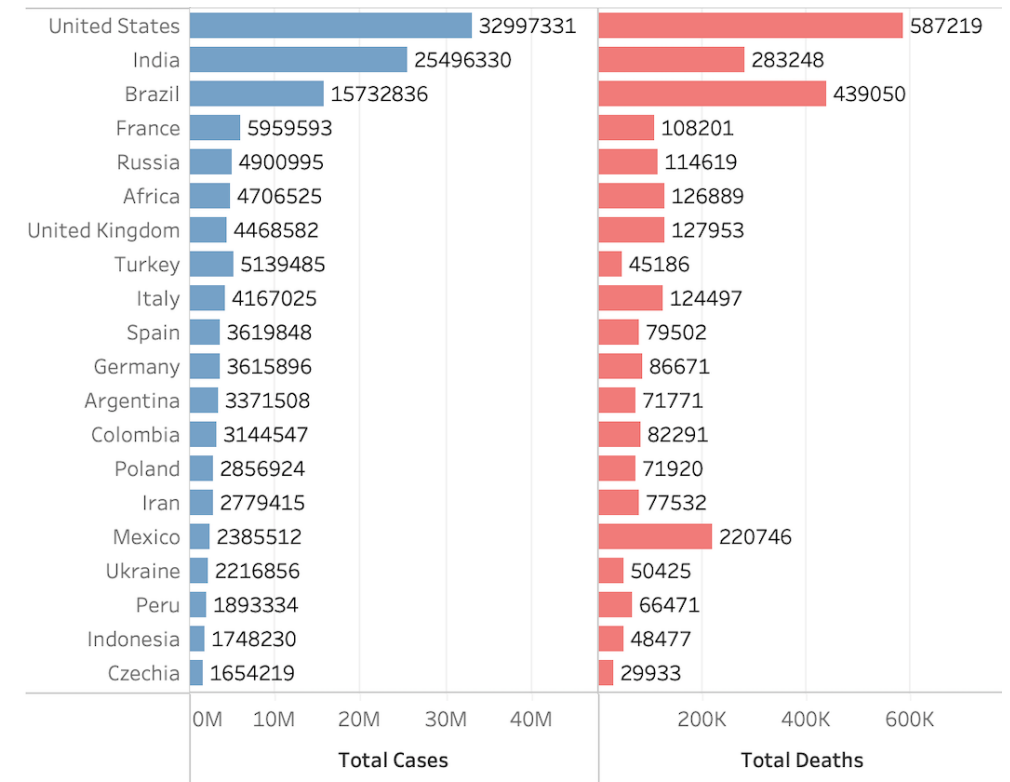
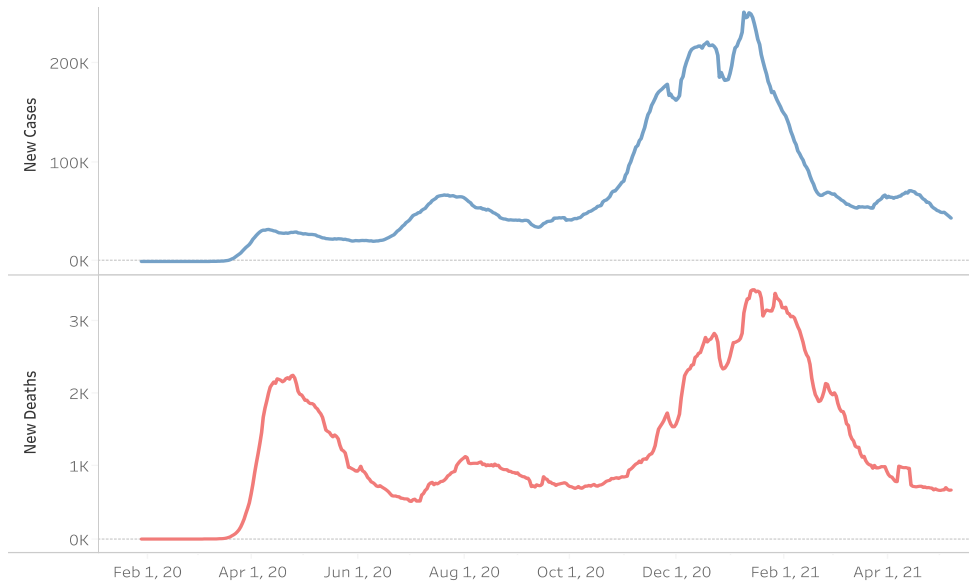
matricola 821063



USA vs The World

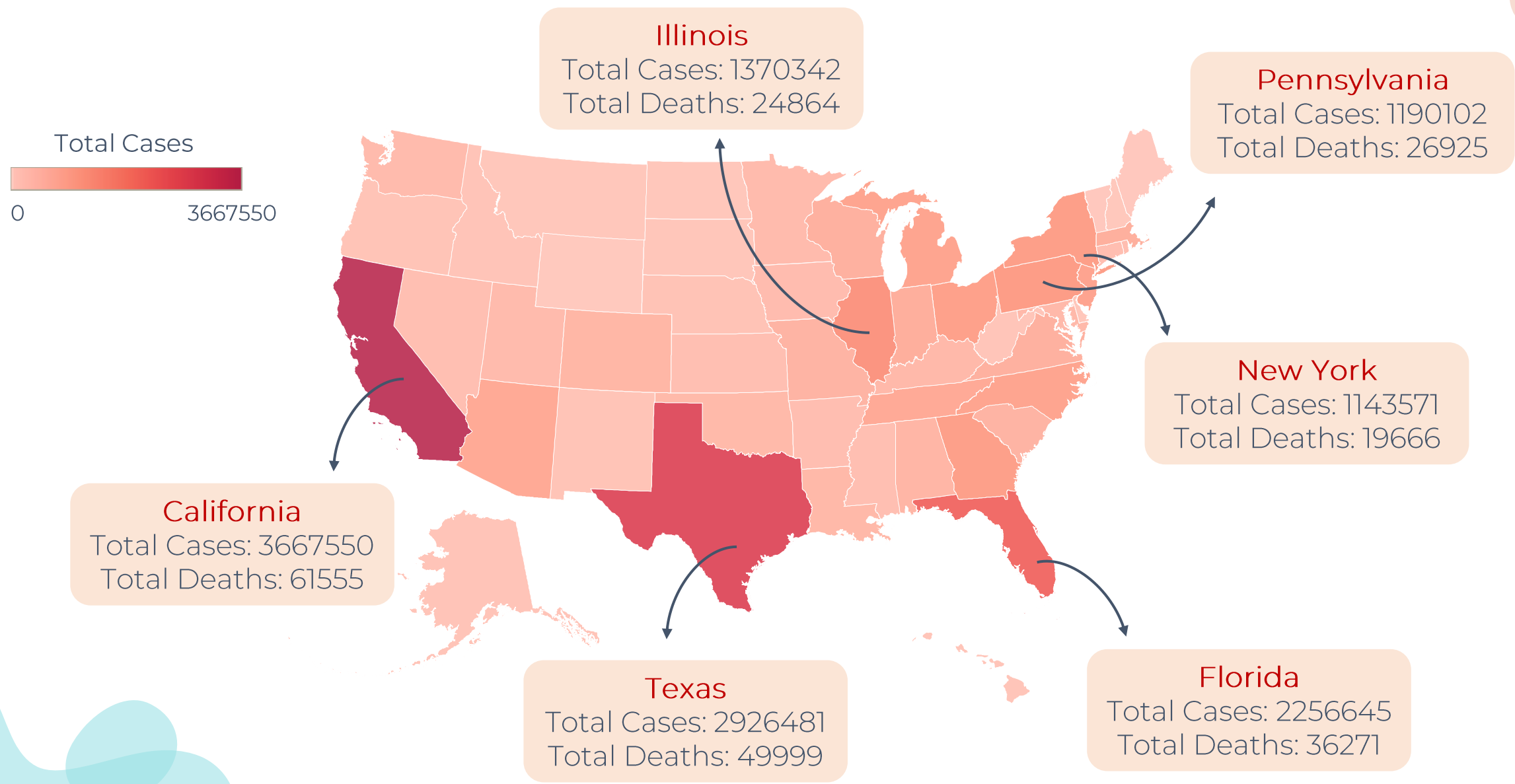
Il primo caso di COVID-19 negli Stati Uniti è stato registrato il 13 gennaio 2020. Esattamente due mesi dopo, il presidente ha dichiarato lo stato di emergenza nazionale in tutto il Paese per contrastare il diffondersi dei contagi da coronavirus.

Nel corso del tempo, è stato effettuato un numero sempre maggiore di test diagnostici e analisi delle sequenze genomiche del virus SARS-CoV-2, ancora poco conosciuto e studiato.



Gli Stati Uniti sono tra i paesi più colpiti al mondo, con più di 32 milioni di casi confermati e più di mezzo milione di decessi.

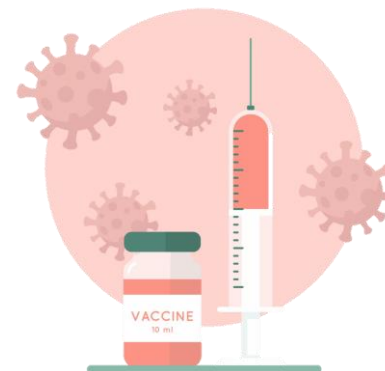
Stati degli USA più colpiti



Vaccinazioni

Il 15 dicembre 2020 è stato somministrato il primo vaccino a New York, dando il via alla campagna di vaccinazione contro il COVID-19 negli USA. Attualmente, negli Stati Uniti sono state somministrate 283941223 dosi, di cui 153075845 (53,9%) della **Pfizer-BionTech** e 120493252 di **Moderna** (42,4%).

Il **38,9%** della popolazione totale è stata completamente vaccinata, mentre il **61,1%** della popolazione con età maggiore o uguale 18 ha ricevuto almeno la prima dose di vaccino.



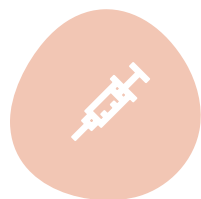
Gli Stati Uniti procedono con la distribuzione dei vaccini ad un buon ritmo, risultando **tra i primi 5 Paesi nel mondo** con la percentuale della popolazione completamente vaccinata più elevata, dopo Israele, Regno Unito, Ungheria, Bahrain e Cile.

I primi 4 Stati per numero di vaccinazioni complete negli USA sono **California, Texas, Stato di New York e Florida**. Maggiori dettagli sulla distribuzione tra i vari stati sono reperibili al seguente link:
inserire dashboard vaccinazioni

Vaccinazioni

Il 15 dicembre 2020 è stato somministrato il primo vaccino a New York, dando il via alla campagna di vaccinazione contro il COVID-19 negli USA.

Gli Stati Uniti procedono con la distribuzione dei vaccini ad un buon ritmo, risultando tra i primi 5 Paesi nel mondo con la percentuale della popolazione completamente vaccinata più elevata, dopo Israele, Regno Unito, Ungheria, Bahrain e Cile.



283.941.223

**Dosi di vaccino
somministrate**

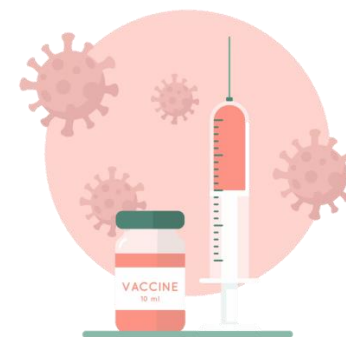


38,9%

**Popolazione
completamente vaccinata
(61,1% almeno prima dose).**

I primi 4 Stati per numero di vaccinazioni complete negli USA sono **California, Texas, Stato di New York e Florida**. Maggiori dettagli sulla distribuzione tra i vari stati sono reperibili al seguente link:

<https://public.tableau.com/profile/matteo.campironi#!/vizhome/DSLlab/DEMS>



Obiettivi principali



- Quanto e con che frequenza hanno pubblicato gli Stati Uniti?

L'idea è rispondere utilizzando i metadati delle pubblicazioni scaricate da PubMed

- Su quali argomenti si concentrano in particolare le pubblicazioni?

L'idea è rispondere utilizzando tecniche di text mining per creare dei cluster ed estrarne il topic principale

- Perché questo determinato argomento è tanto studiato?

L'idea è capire il motivo del perché gli USA si sono concentrati particolarmente su un determinato argomento uscito dall'analisi precedente



- Qual è lo stato della diffusione delle varianti ad Aprile 2021?

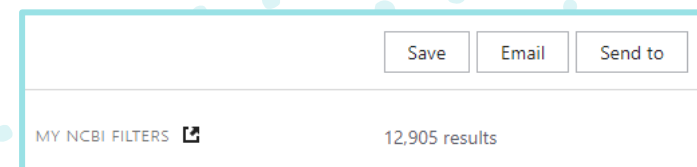
L'idea è di utilizzare le sequenze presenti su GISAID per stimare la diffusione delle varianti negli USA

Criticità pubblicazioni PubMed

Discrepanza numero pubblicazioni

Il numero di pubblicazioni restituito dalla query effettuata sul sito di PubMed (12905) non corrispondeva con quella effettuata tramite Entrez (circa 45000).

Dal momento che nelle 45000 pubblicazioni comparivano anche paper datati prima del 2019 e che il numero era molto distante da quello presente su LitCovid, si è deciso di scaricare gli ID utilizzando il sito e di proseguire l'analisi con Entrez.



Save Email Send to

MY NCBI FILTERS 12,905 results

Literature	
Bookshelf	818
MeSH	0
NLM Catalog	333
PubMed	45,394

Dati sulle pubblicazioni



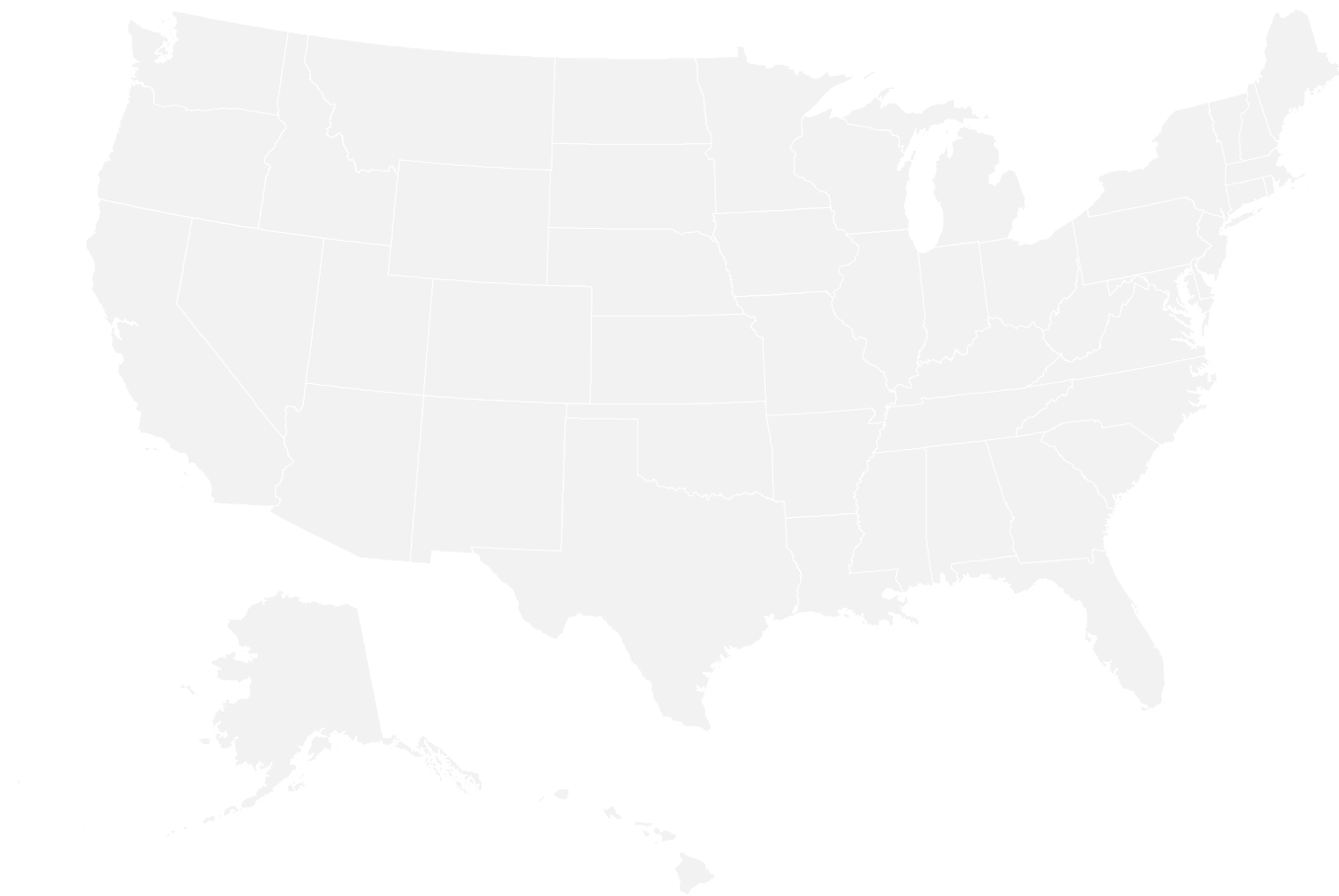
12.905

Numero di pubblicazioni
inviolate dagli USA



7.504

Numero di pubblicazioni
con il campo *abstract*
non vuoto



Su quali argomenti si concentrano le pubblicazioni?

Per rispondere a questa domanda è stato utilizzato il **Text Mining** sugli abstract delle pubblicazioni.

Dopo un primo lavoro di preprocessing è stato utilizzato un modello preaddestrato per la generazione degli **embedding** delle pubblicazioni.

Successivamente è stato utilizzato un algoritmo per la **riduzione della dimensionalità**, in modo tale che ad ogni pubblicazione corrispondesse un punto sul piano cartesiano.

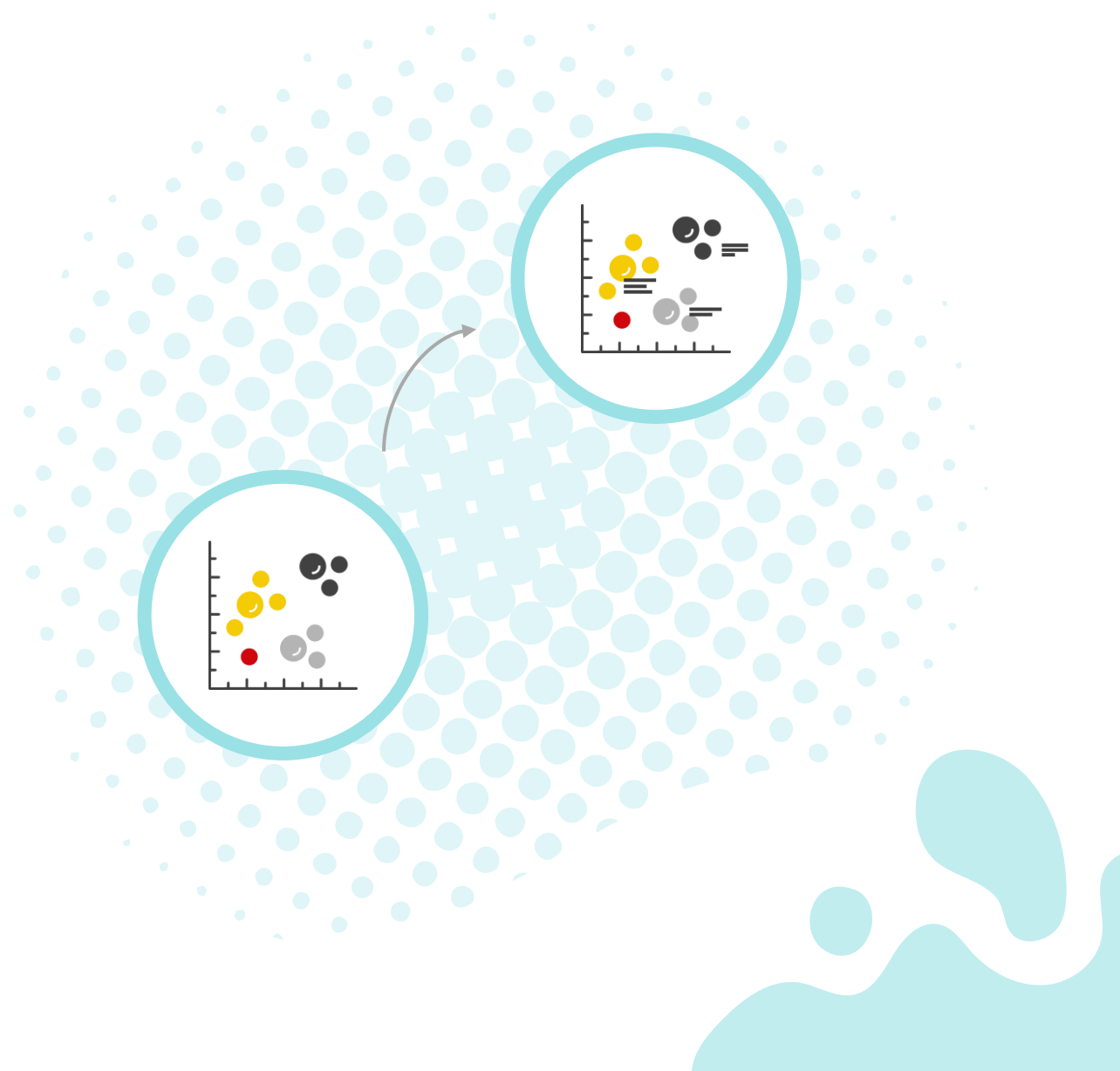


Su quali argomenti si concentrano le pubblicazioni?

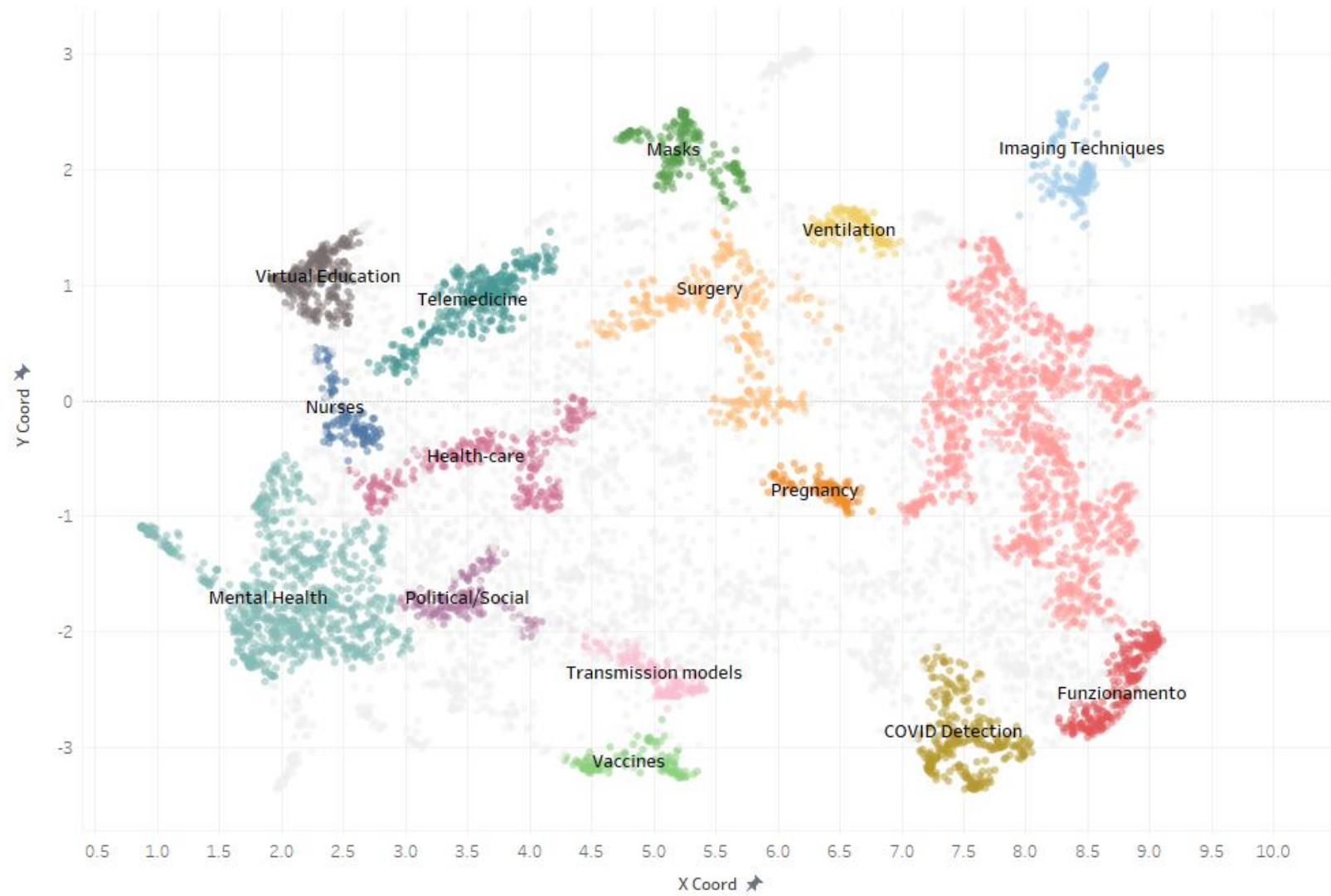
Infine è stato utilizzato DBSCAN come algoritmo di clustering basato su densità.

Per ogni cluster trovato è stato effettuato un lavoro di estrazione delle keywords, eliminando le parole più frequenti come *covid*, *coronavirus*, ecc...

In questo modo sono stati assegnati manualmente dei nomi ai cluster in base ai temi trattati



Sheet 1



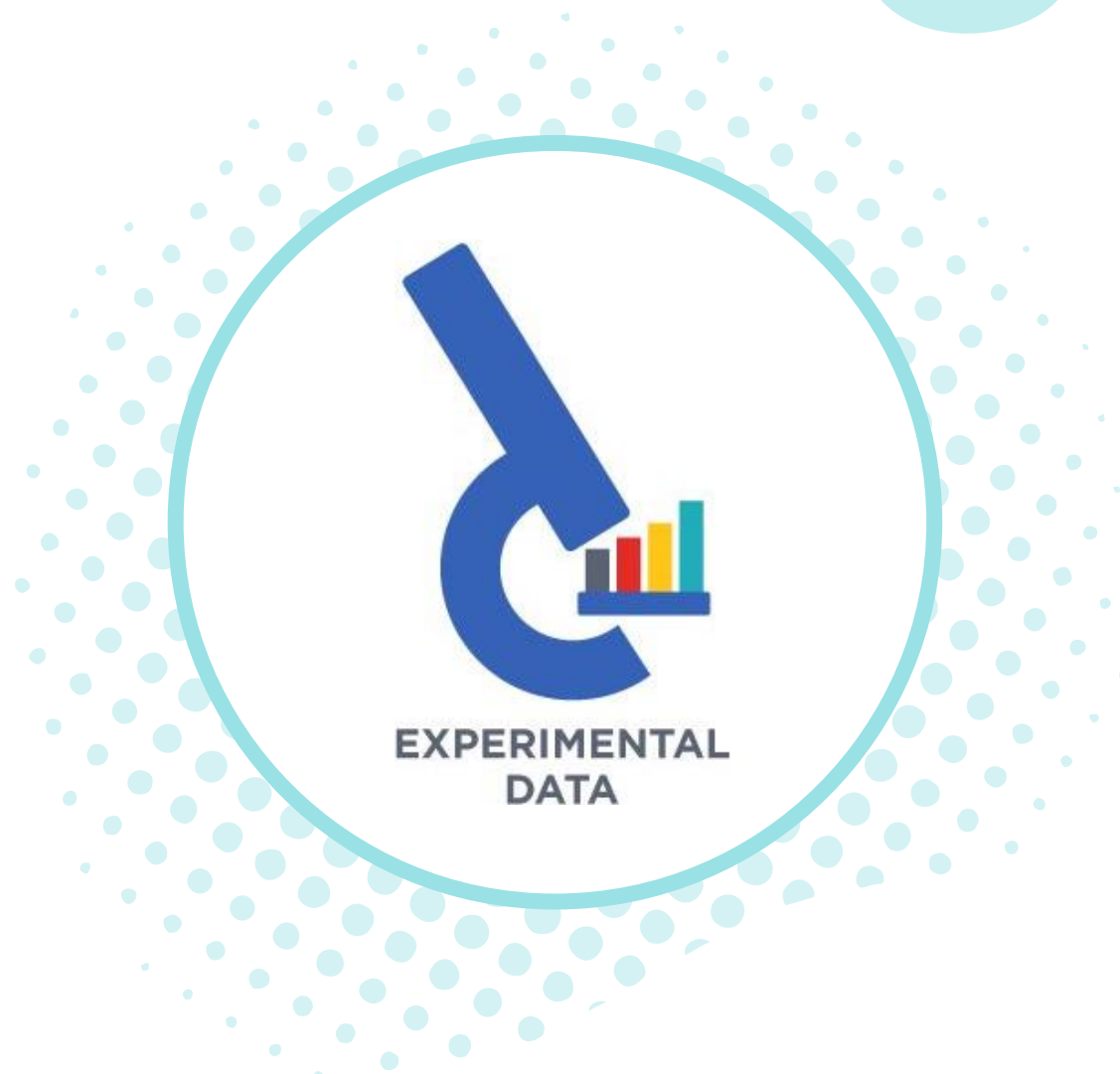
Perché la salute mentale è tanto studiata?

Household Pulse Survey

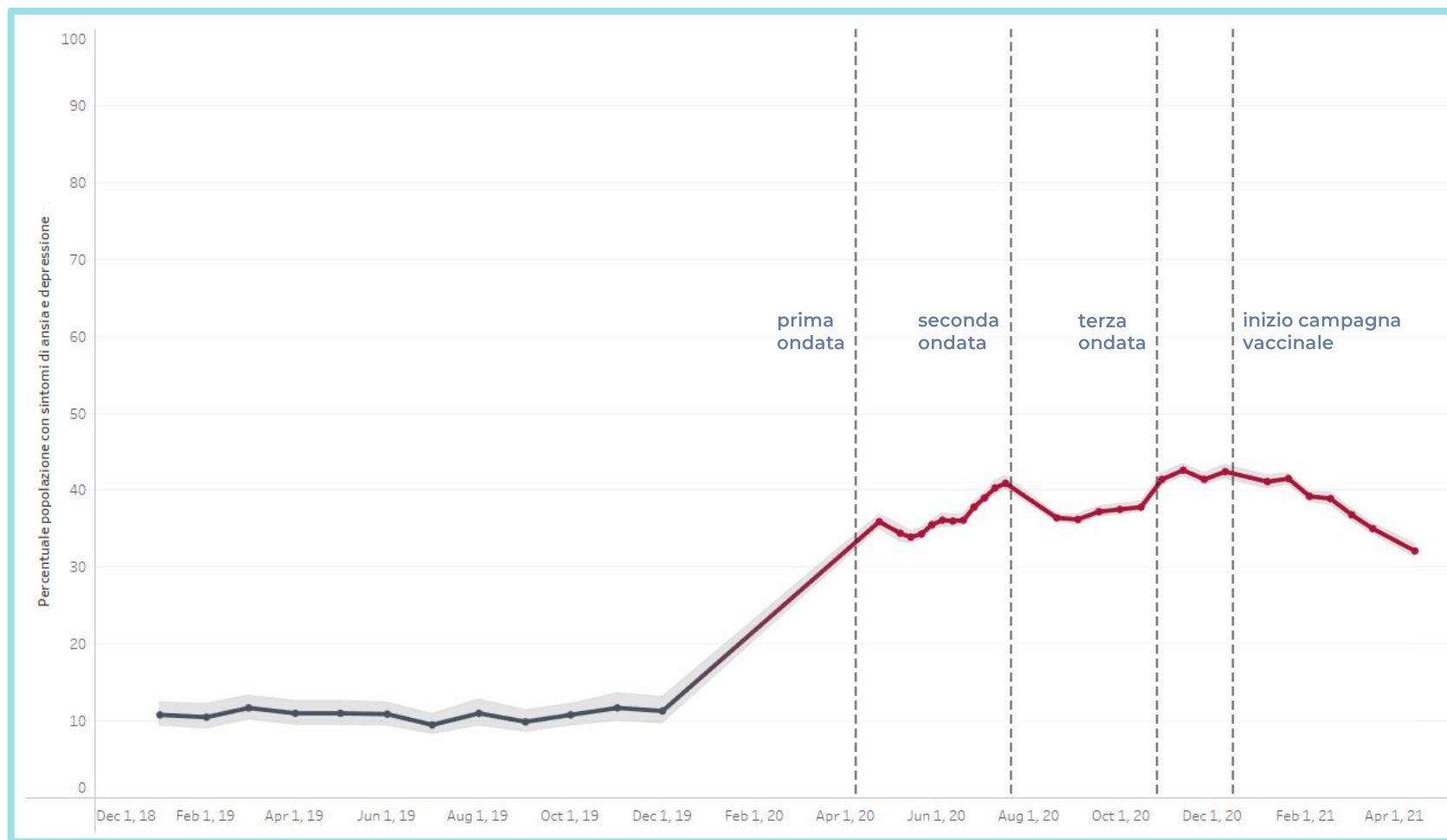
Da una collaborazione del National Center for Health Statistics (NCHS) e del Census Bureau, nasce una raccolta di dati sperimentale chiamata Household Pulse Survey.

È un questionario online dalla durata di 20 minuti, al fine di monitorare i cambiamenti nella salute mentale degli statunitensi a causa dell'impatto della pandemia.

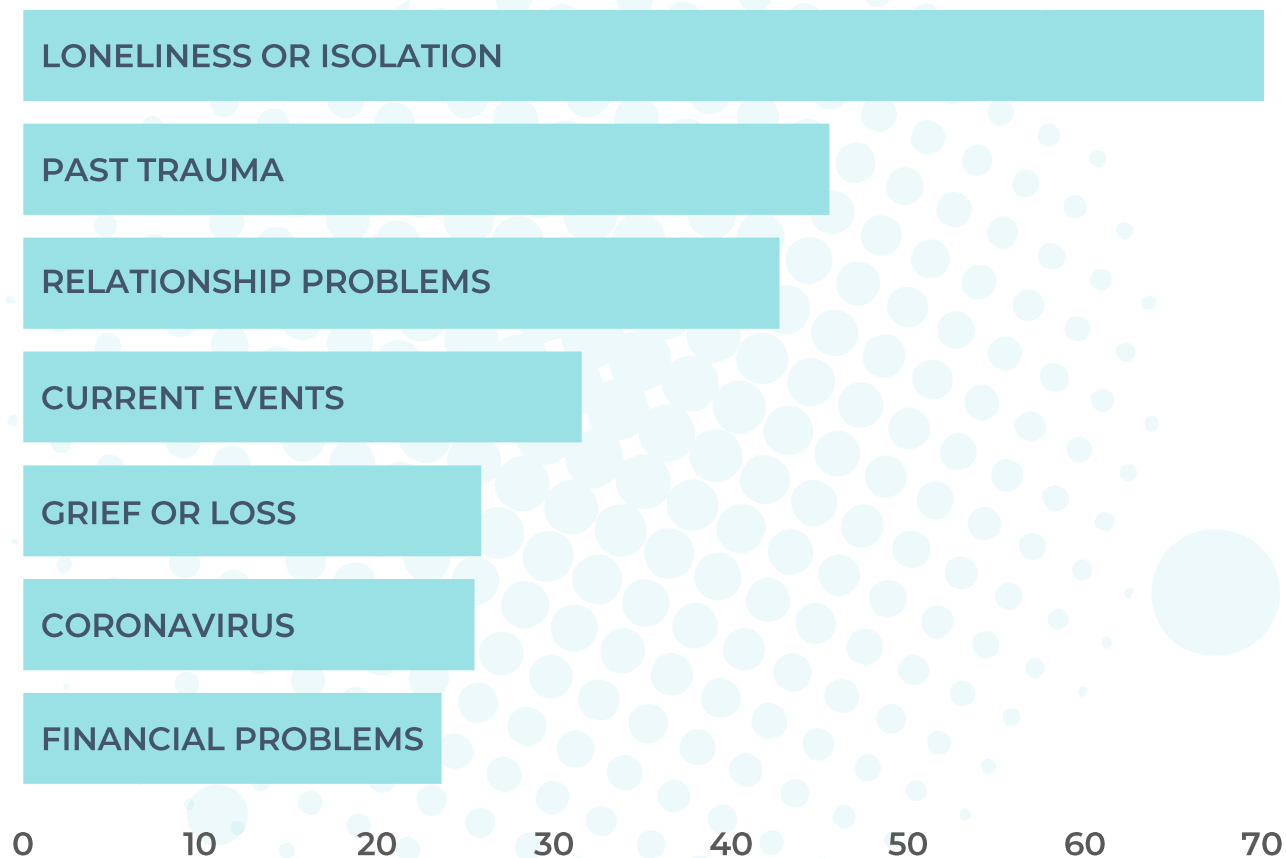
Il NCHS ha incluso delle domande per ottenere informazioni sulla frequenza dei sintomi di ansia e depressione.



Salute mentale negli USA dal 2019 ad oggi



È possibile osservare come nel 2019 in media circa l'11% della popolazione soffriva di disturbi di salute mentale, contro il 38% del 2020. Il trend sembra in discesa dopo l'inizio della campagna vaccinale negli USA.



Cause dei problemi mentali nel 2020

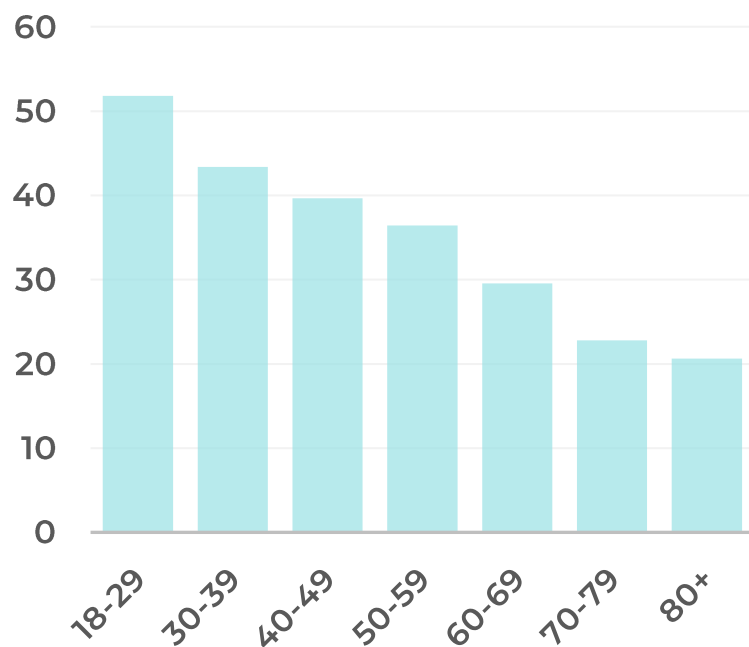
Ad ogni partecipante è stata posta la seguente domanda:

Think about your mental health test. What are the main things contributing to your mental health problems right now? Choose up to 3

Salute mentale nel dettaglio

Non tutte le fasce d'età sono state colpite allo stesso modo:

I giovani hanno subito maggiormente l'impatto della pandemia, con un tasso maggiore di sintomi di ansia e depressione.

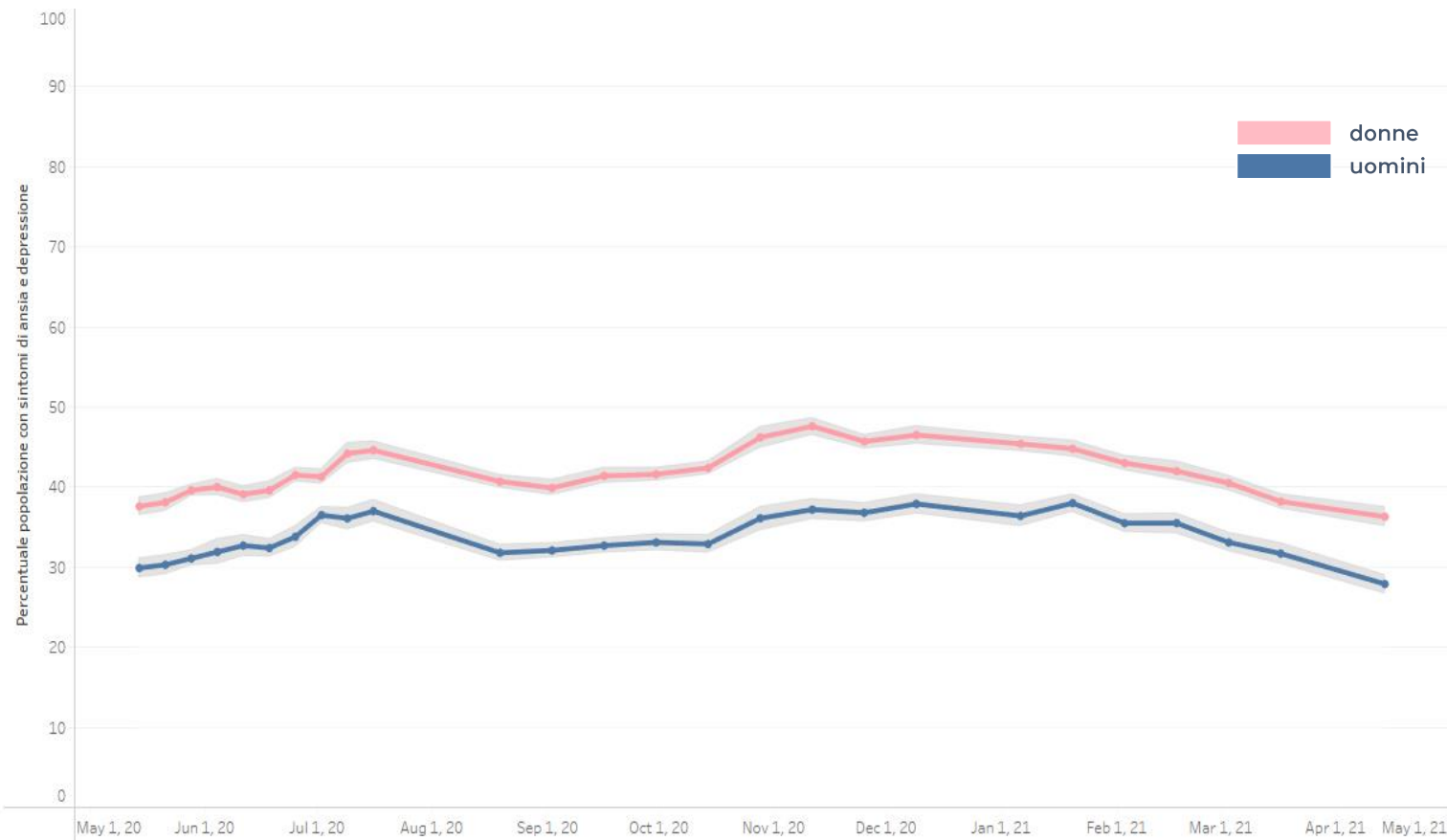


Less than an high school diploma	39.19
Some college/Associate's degree	35.73
High school diploma or GED	32.62
Bachelor's degree or higher	28.49

Anche il livello d'istruzione gioca un ruolo importante:

I laureati sembrano aver subito meno l'impatto della pandemia, seguite dalle persone diplomate e infine da quelle con un titolo inferiore al diploma.

Salute mentale per genere



Il genere contribuisce alla risposta:

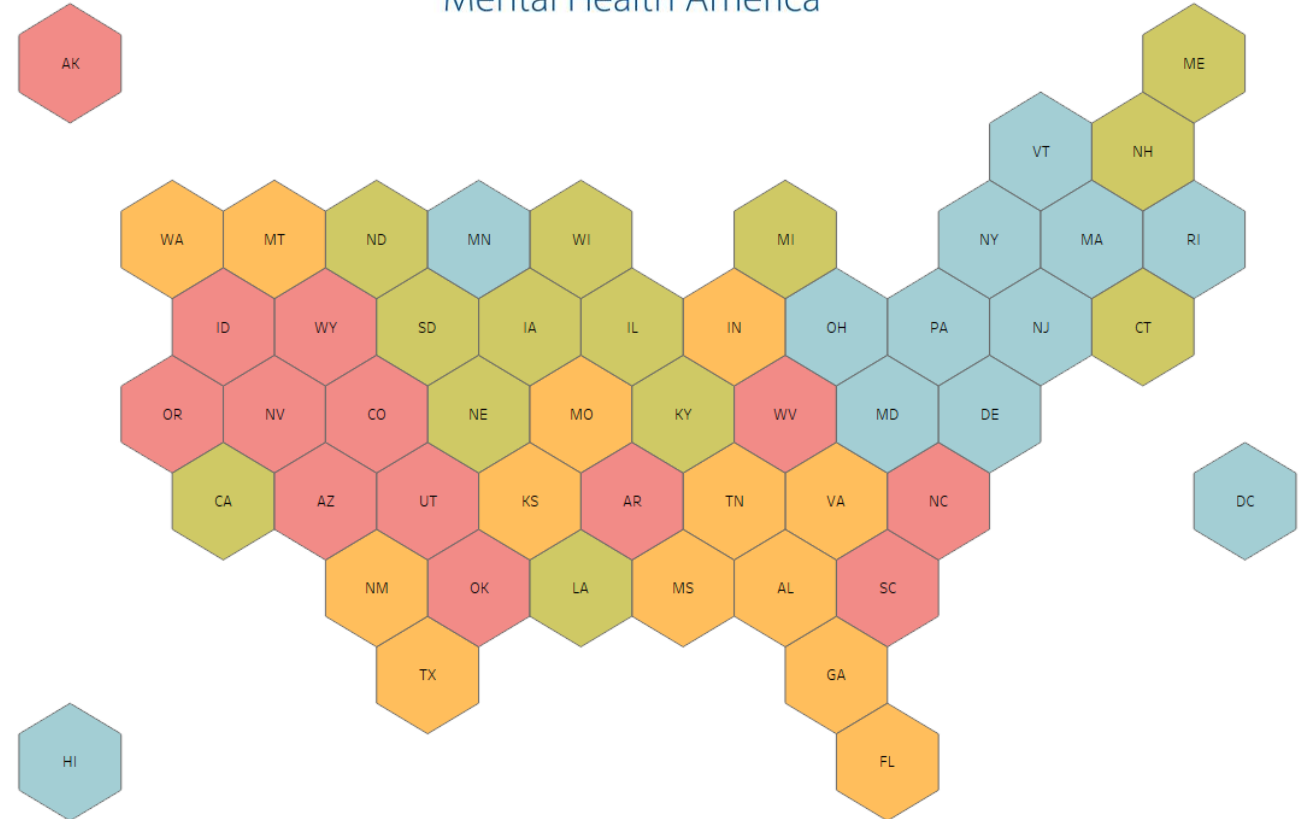
Durante tutto il periodo in cui è stato effettuato il sondaggio le **donne** risultano aver subito maggiormente l'impatto della pandemia. La media nel 2020 per il genere **femminile** risulta del 42%, contro il 34% per gli **uomini**.

Mental Health America

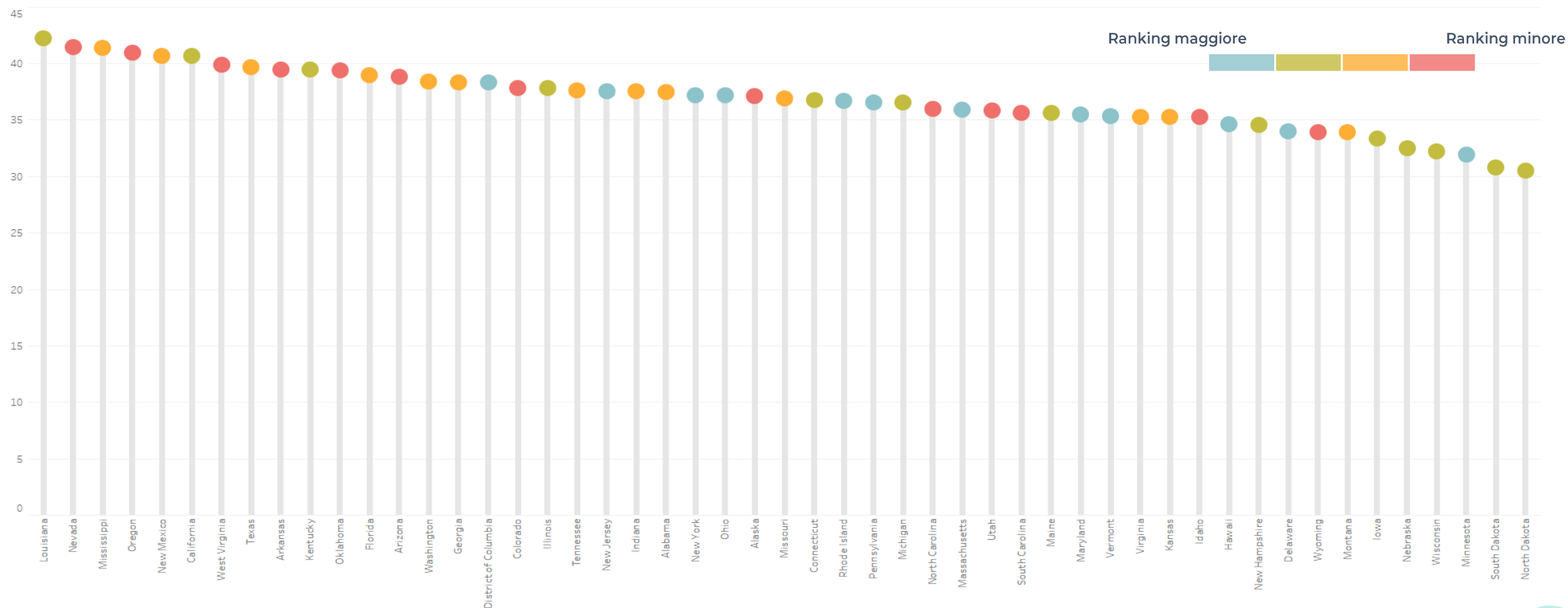
2020 Ranking The States

Mental Health America è un organizzazione non profit che si occupa dei problemi di salute mentale negli USA.

Ogni anno viene stilata una classifica degli stati in base alla prevalenza di disturbi mentali nella popolazione e alla capacità di poter accedere a strutture per curarsi.



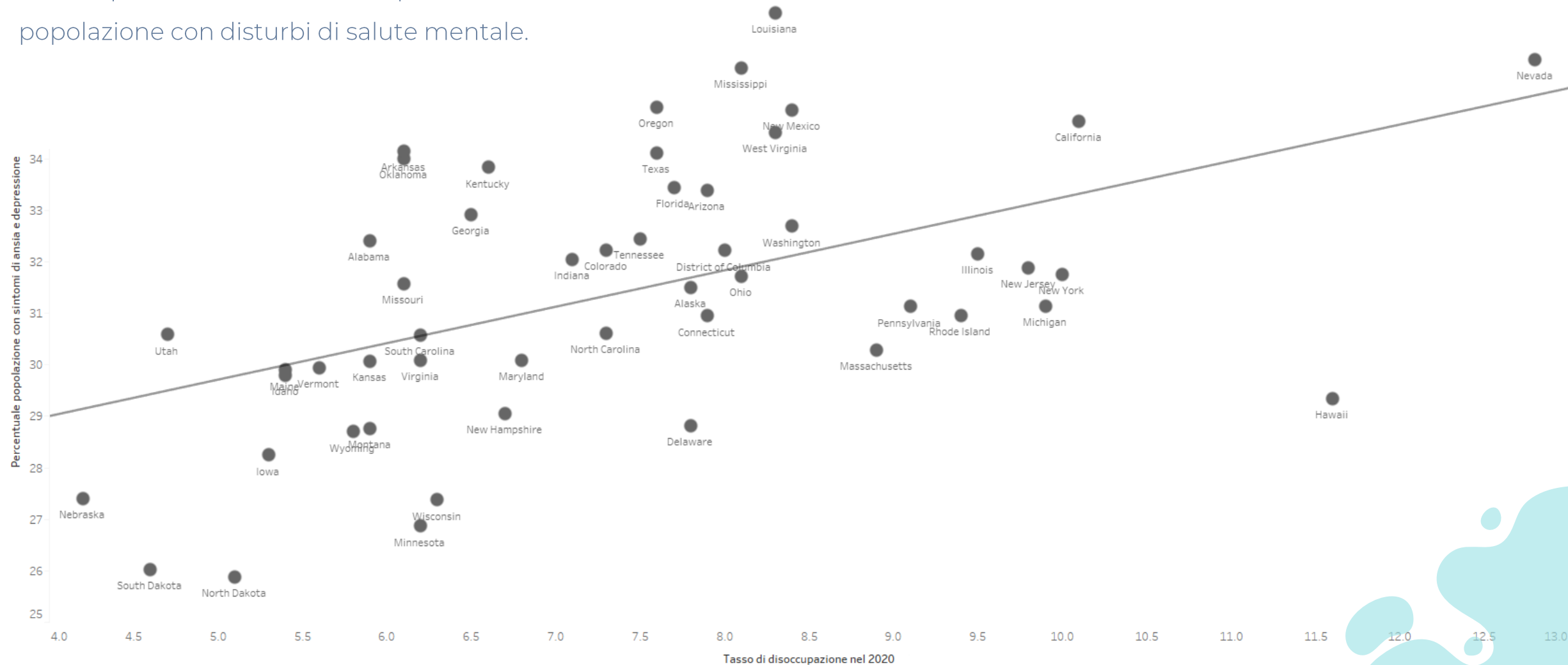
Salute mentale per stato degli USA



Lo stato con la percentuale minore di casi è il **North Dakota**, mentre quello che ha subito di più è la **Louisiana**. Non sempre però gli stati che offrono un supporto migliore dal punto di vista della sanità mentale sono quelli che ne escono meglio.

Disoccupazione e salute mentale

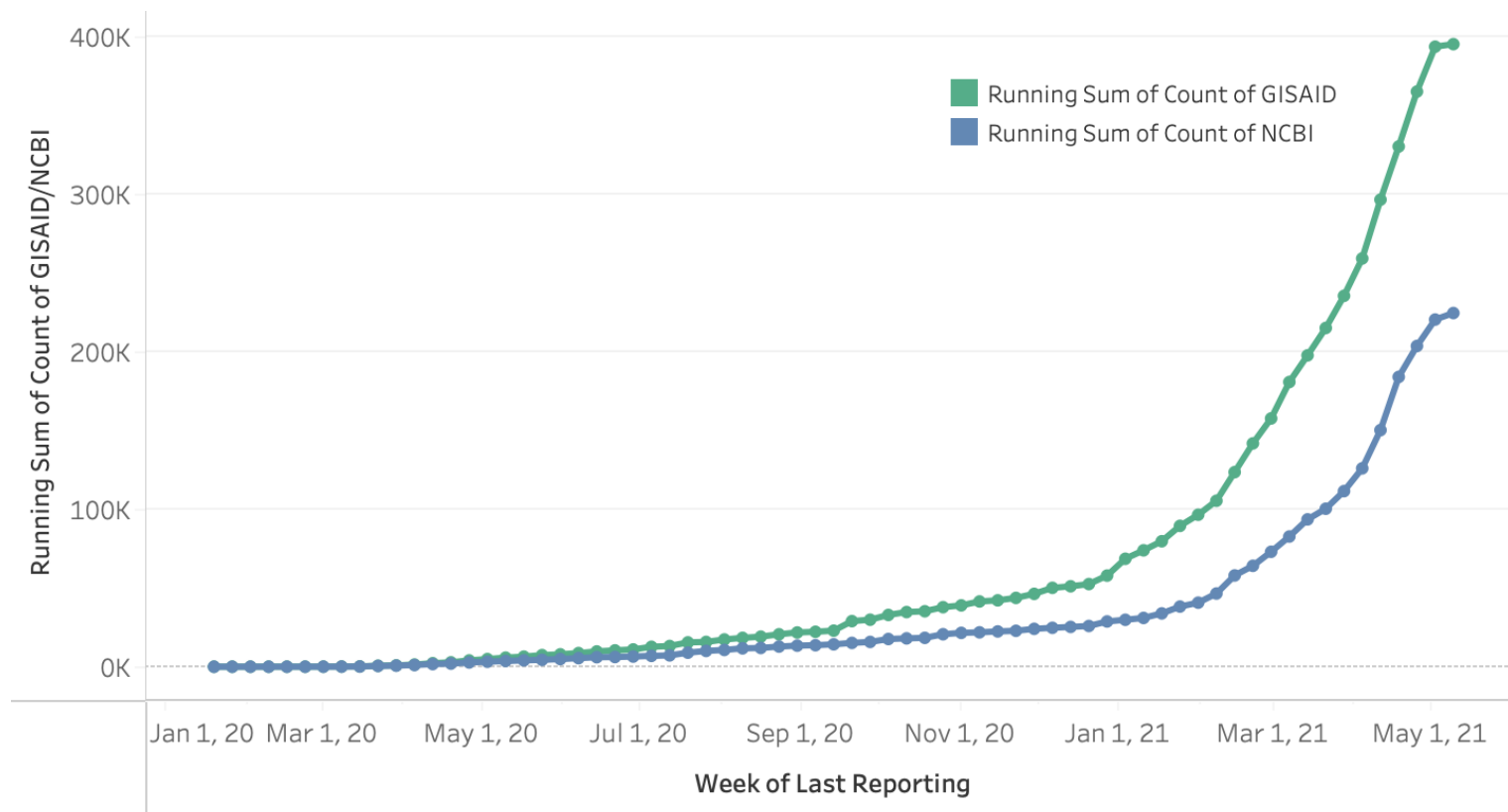
È possibile osservare come all'aumentare del tasso di disoccupazione cresca anche la percentuale di popolazione con disturbi di salute mentale.



Sequenze disponibili: NCBI e GISAID

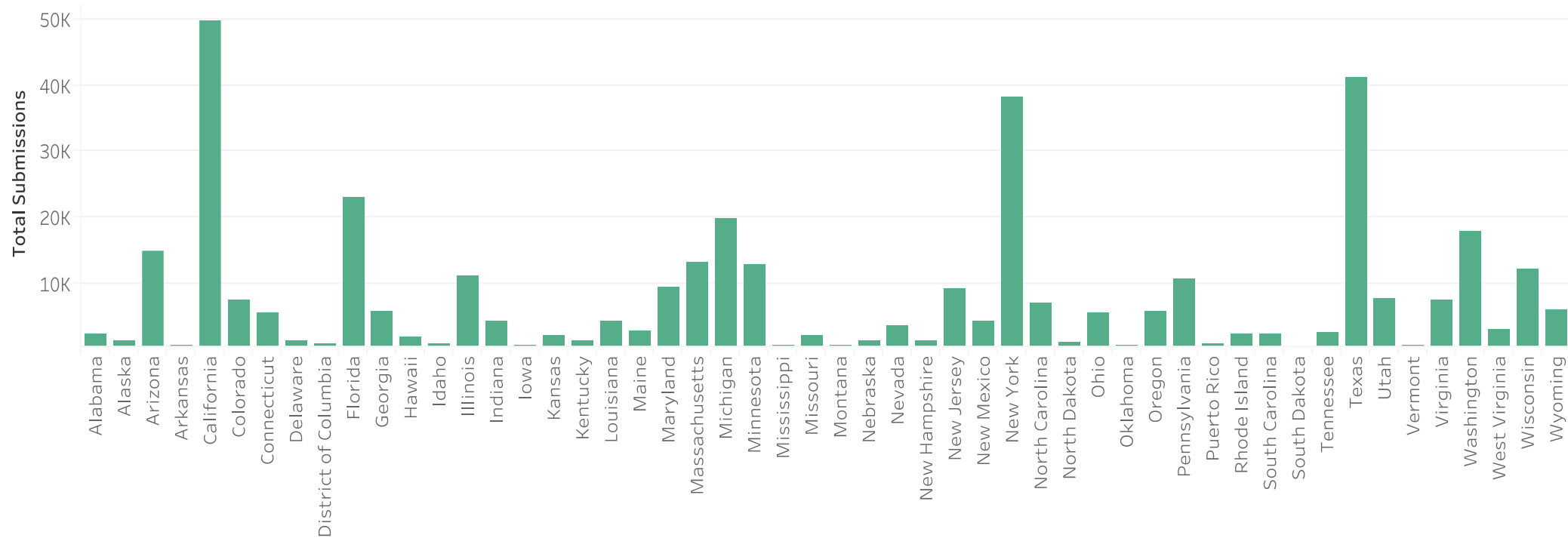
L'analisi ed il sequenziamento del virus sono indispensabili per lo studio delle varianti SARS-CoV-2 negli Stati Uniti e negli altri paesi.

Due tra i principali archivi pubblici che rendono disponibili sequenze genomiche del virus sono: **NCBI**, National Center for Biotechnology Information, e **GISAID**.



Invio delle sequenze per stato degli USA

GISAID mette a disposizione i dati relativi al numero di inserimenti di sequenze relative al SARS-CoV-2 dei vari stati degli USA effettuate a partire da gennaio 2020 fino al presente.



Numero sequenze USA gennaio 2020 – maggio 2021



È possibile osservare come il numero di inserimenti di sequenze su GISAID da parte degli Stati Uniti sia notevolmente aumentato nel corso del tempo, essendo aumentate le risorse a disposizione e rendendo evidente quanto questo sia necessario per una maggiore conoscenza di questa virus.

Classificazione delle varianti

Sequenziamento e analisi hanno consentito di monitorare la diffusione di varianti emergenti nel mondo ed in particolare negli Stati Uniti.

Il CDC, Centers of Disease Control and Prevention, riporta lo schema di classificazione sviluppato dagli Stati Uniti per definire le varianti SARS-CoV-2.



Variants of Concern	First detected
B.1.1.7	United Kingdom
B.1.351	South Africa
B.1.427	California (USA)
B.1.429	California (USA)
P.1	Brazil

Variants of Interest	First detected
B.1.525	UK/Nigeria
B.1.526	New York (USA)
B.1.526.1	New York (USA)
B.1.617	India
B.1.617.1	India
B.1.617.2	India
P.2	Brazil

Qual è lo stato della diffusione delle varianti ad aprile 2021?

Basandosi sulle varianti di interesse scelte dal CDC, si è deciso di concentrarsi sulla diffusione di quest'ultime nei vari stati degli USA nel mese di aprile 2021.

