

WEBINAR

Harnessing the Power of Visual Analytics & Data Science in Health Trends

Practical Guidance and Insights from TIBCO Spotfire® and PerkinElmer

LIVE PRESENTATION WITH:



Michael O'Connell,
Chief Analytics Officer,
TIBCO Software® Inc.



Dan Weaver,
Senior Product Manager and Solutions
Architect, Research Informatics

TIBCO Spotfire®

April 21, 11 AM ET

REGISTER NOW


PerkinElmer®
For the Better



Agenda

COVID-19 Basics

- Reproduction Numbers
- Case Fatality Rates
- Interventions : NPIs

Genomics

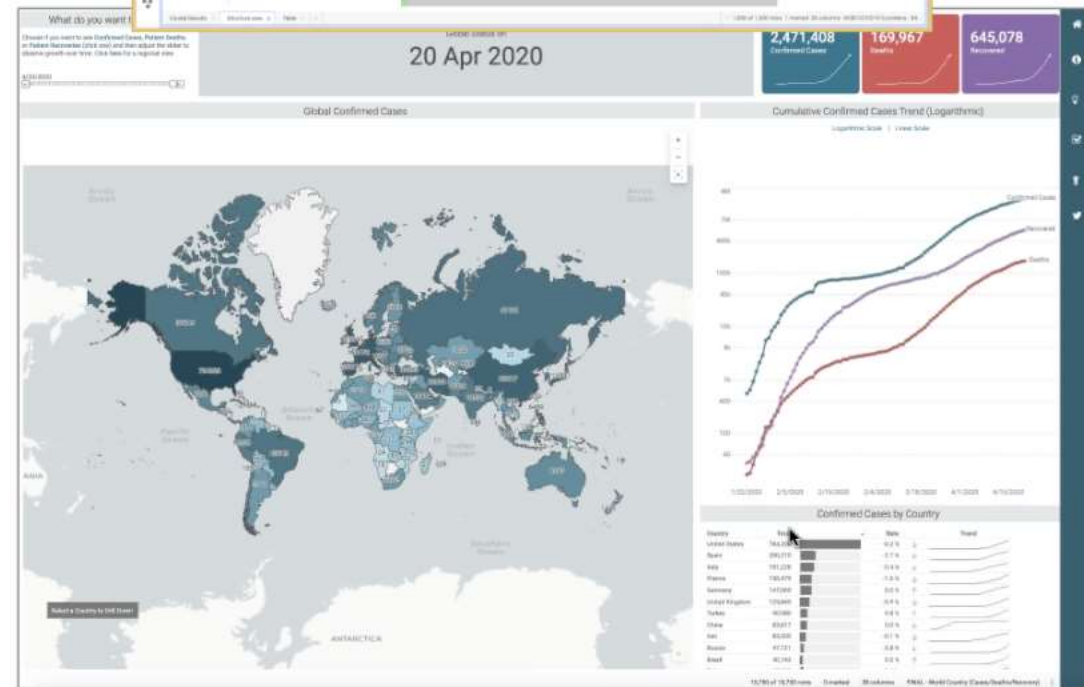
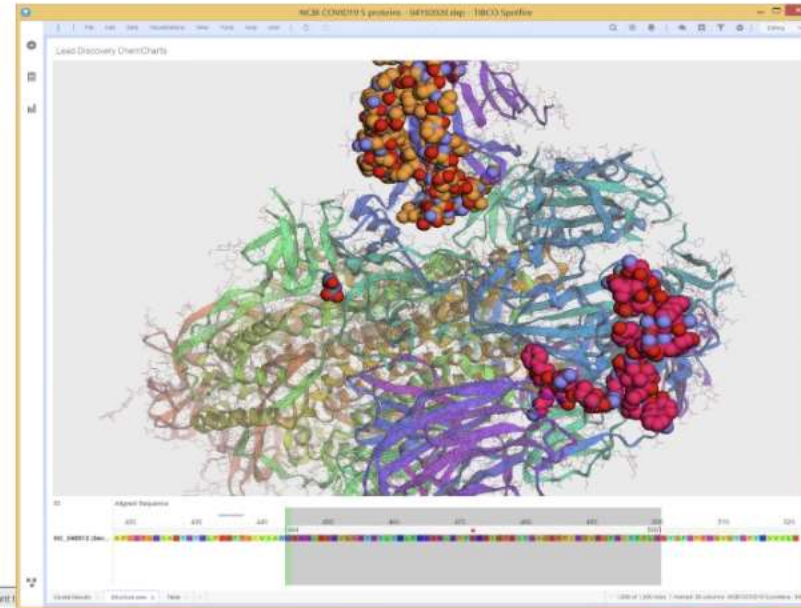
- COVID19 Spike protein genomics

Demonstrations

- Data Science
- Genomics

Rejoinder

- Current controversies / analyses
- Genomics directions
- Back to work directions



The COVID-19 Webinar Panel

TIBCO & PerkinElmer



Michael O'Connell - chief analytics officer at TIBCO.



Neil Kanungo - Spotfire visual analytics ninja.



Dan Weaver – Perkin Elmer Bioinformatics.



Gerard Conway – Perkin Elmer Spotfire guru.



TIBCO Data Science and COVID-19

Provide data-driven review of issues & analytics work in the space

- Epidemiology, Data Science, Data Journalism

Contribute data and analytics to the community

- Data, Spotfire apps, R and Python code and global intervention data timelines

Bring a visual analytics and data science perspective

- Recognize the many sources of variability; potential for wide array of outcomes

Show TIBCO Predict & Unify products at work

- **Visual Data Science**
- **GeoSpatial Data Science**
- **Unified Data** and **Data Wrangling**
- **Real-time data updates**

Virulence of Diseases

Compare diseases by and

Show me...

Compare Disease Characteristics

Quick Insights

COVID-19

Compared to Deadliest

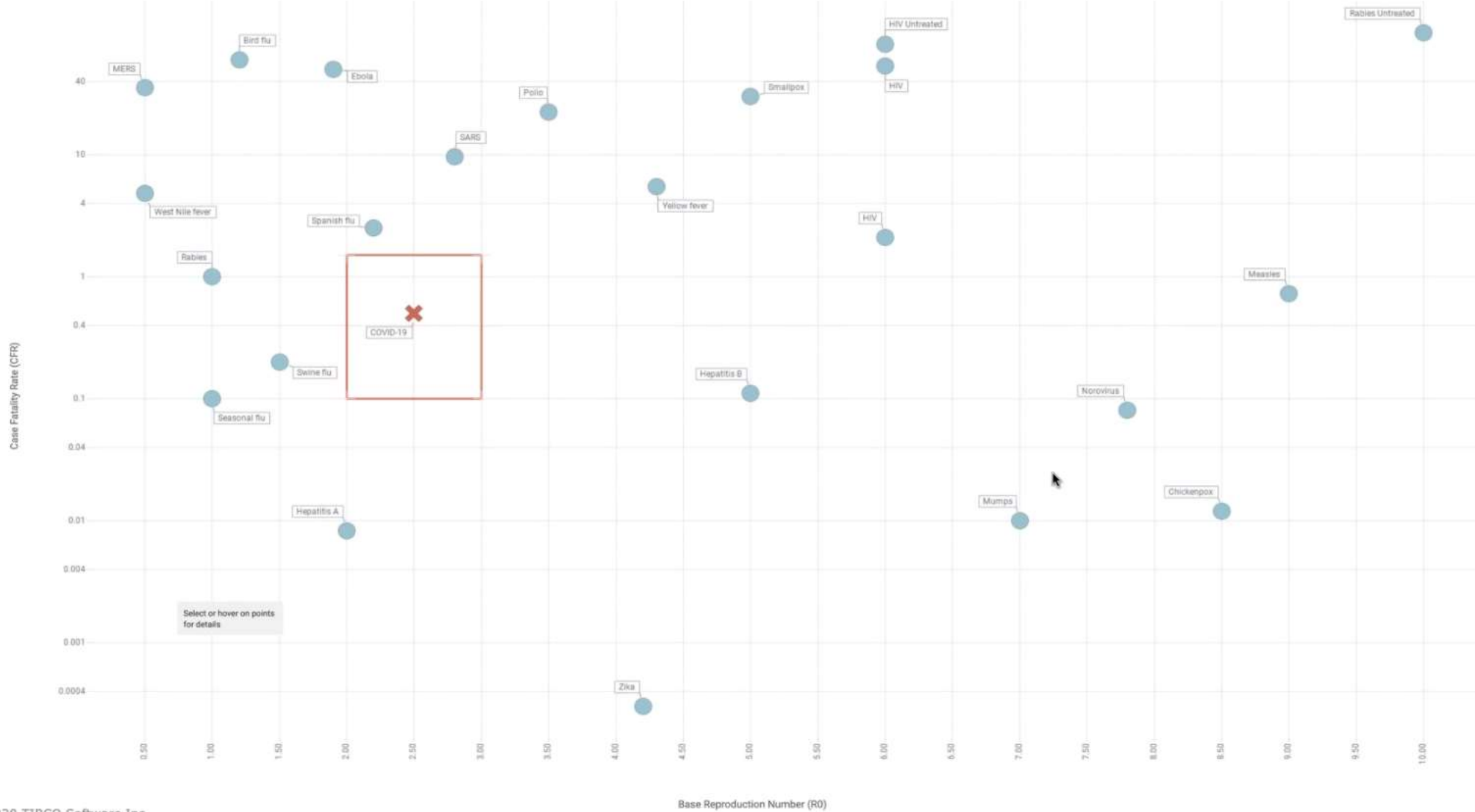
Pathogen Type

- Bacterium
- Parasite
- Prion
- Virus

Diseases

Type to search in list

- (All) 32 values
- Bird flu
- Chickenpox
- Chikungunya
- Common cold
- COVID Uncertainty
- COVID-19
- Dengue fever
- Dengue fever Untreated
- Ebola
- Hand, foot & mouth
- Hantavirus



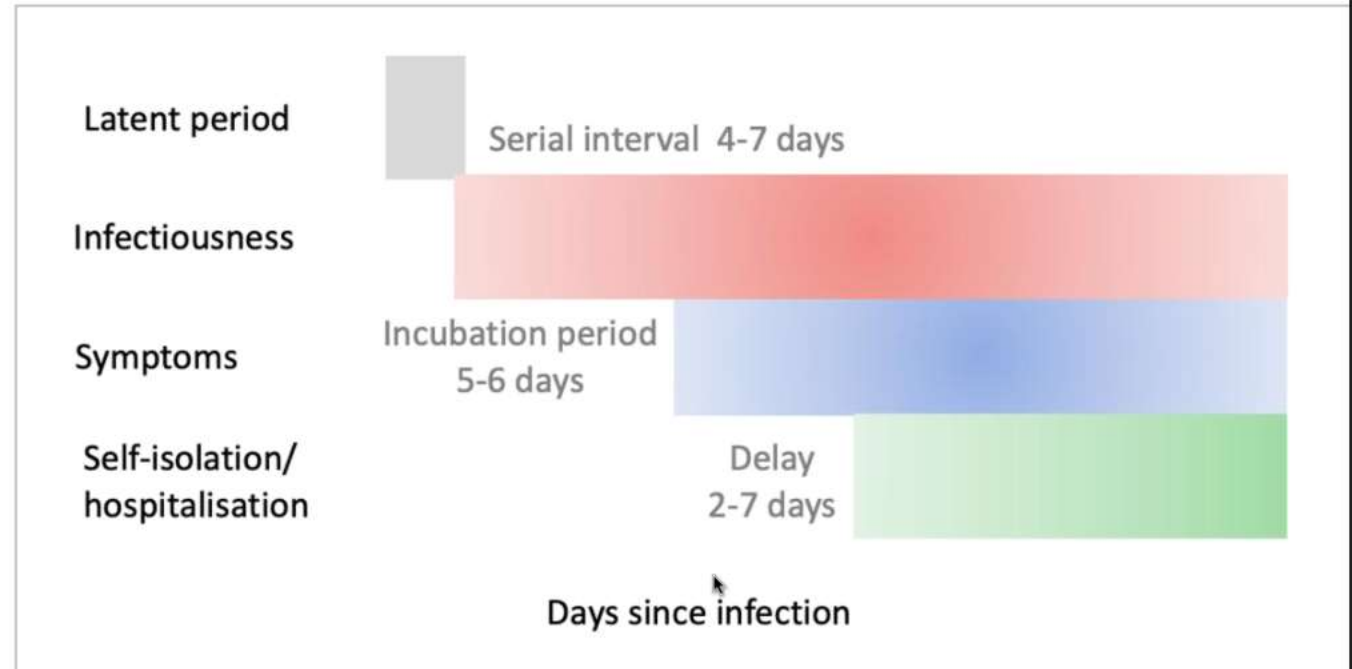
Select or hover on points for details

R0: the reproduction number; and Re, Rt

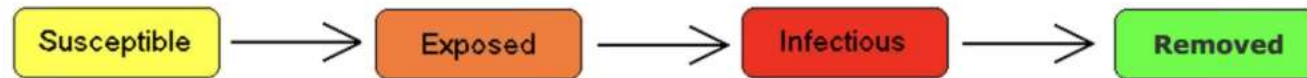
Latent Period = time between occurrence of infection and onset of infectiousness (when the infected individual becomes infectious).

Serial Interval = duration of time between the onset of symptoms in primary case and onset of symptoms in secondary case infected by primary case.

Incubation Period = time period between the occurrence of infection (or transmission) and the onset of disease symptoms



Common epidemiology model: SEIR



Removed = recovered or died

R0: the reproduction number; and Re, Rt

R0 = Basic Reproduction Number

The number of people infected from an infected person, in usual behavior (no interventions e.g. distancing)

Re = Effective Reproduction Number

Shows the effects of interventions

Re < 1 and virus stops spreading

Re is a leading indicator of case curve bending and epidemic curve flattening

COVID-19 Reproduction Numbers

- **R0** is in the range 2-3

Non-Pharmaceutical Interventions (NPIs) can drop the Re dramatically

- China: 2.4 to 1.1 with severe travel / lockdown restrictions

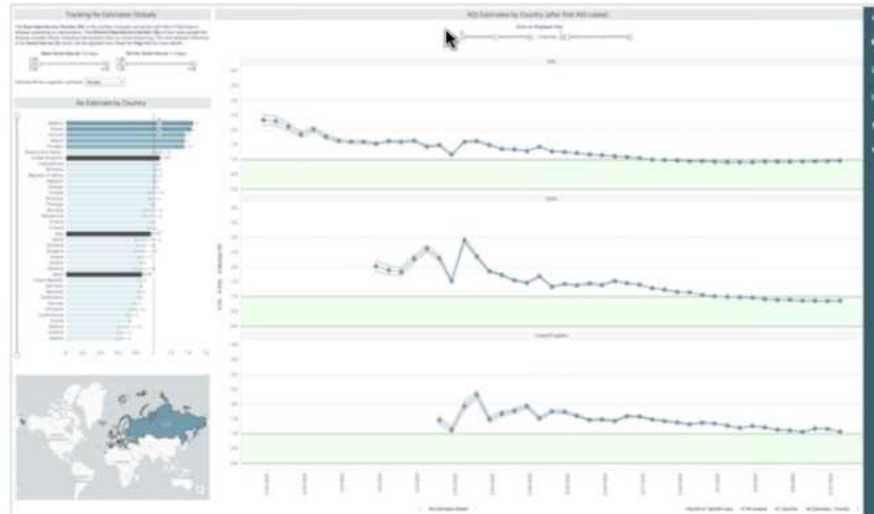
Time Lags & Reproduction Numbers

Time lags

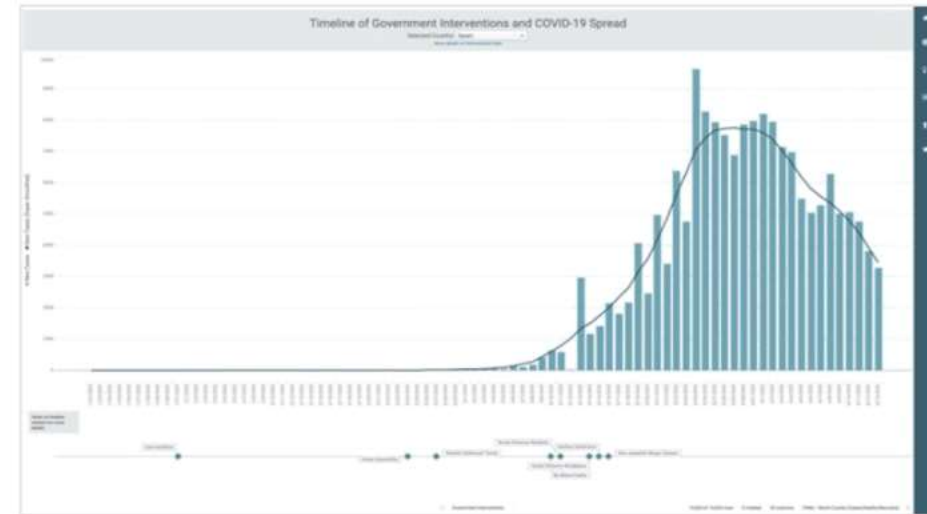
Cases & fatalities that we see now, were initiated a few weeks ago

The world we are in today, doesn't show its COVID status for some weeks in the future

slow the transmission (social distancing)



... and the case counts drop



CFR: the case fatality rate

CFR - risk that someone who develops symptoms will eventually die from infection

Best way to calculate CFR would be to track a large group of people from the point when they develop symptoms until they later die or recover

- Then calculate the proportion of all these cases who had died.
- This is not possible in the real world.

It is incorrect to just divide the total number of deaths by total number of cases as this does not account for *unreported cases or the delay from illness to death.*

- *Many unreported cases* eg due to unavailable test kits. In the US, Bedford estimates ~10-20X under-reporting of cases as of March 13.
- *Time delay*, consider 20 new people admitted to a hospital with confirmed COVID-19 infection on a given day -- that doesn't mean the CFR is zero!. We need to wait to see what happens to them.
- Conversely any deaths that occur now, are people who showed symptoms some weeks before.

Early estimates of CFR in epidemics is typically high as focus is on the sickest of the sick.

- [Early CDC estimates](#) suggested a wide range of 0.25%-3.0% back in January.

CFR: the case fatality rate

see TIBCO blogs:
community.tibco.blog

Data

- [Wu et al.](#) : CFR of COVID-19 in Wuhan of **1.4% (0.9–2.1%)**
 - Lower than naïve confirmed case fatality risk of $2,169/48,557 = 4.5\%$
 - Lower than the approximator of deaths/(deaths + recoveries): $2,169/(2,169 + 17,572) = 11\%$
 - Risk of symptomatic infection increased with age
 - Those above 59 years were 5.1X (4.2–6.1) more likely to die after developing symptoms, compared to those aged 30–59.
- Other recent estimates support 0.5 – 1.5% CFR (Wuhan, International)

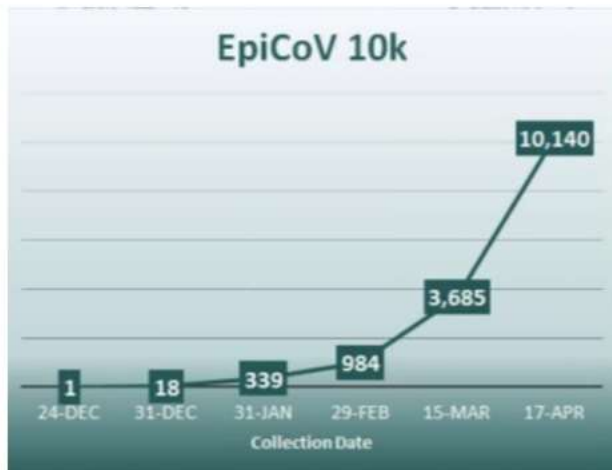
Experts

- **Kucharski:** COVID-19 is **~1% ie ~10X+** more deadly than Flu
- **Fauci et al.** "if one assumes that the number of asymptomatic or minimally symptomatic cases is several times as high as the number of reported cases, the case fatality rate **may be considerably less than 1%**. This suggests that the overall clinical consequences of Covid-19 may ultimately be more akin to those of a severe seasonal influenza (which has a case fatality rate of approximately 0.1%) or a pandemic influenza (similar to those in 1957 and 1968) rather than a disease similar to SARS or MERS, which have had case fatality rates of 9 to 10% and 36%, respectively."
- Some others posit that CFR is lower than this eg **Gupta** (U Oxford), **Bendavid and Bhattacharya** (Stanford); they suggest are larger proportion of the population is already infected.

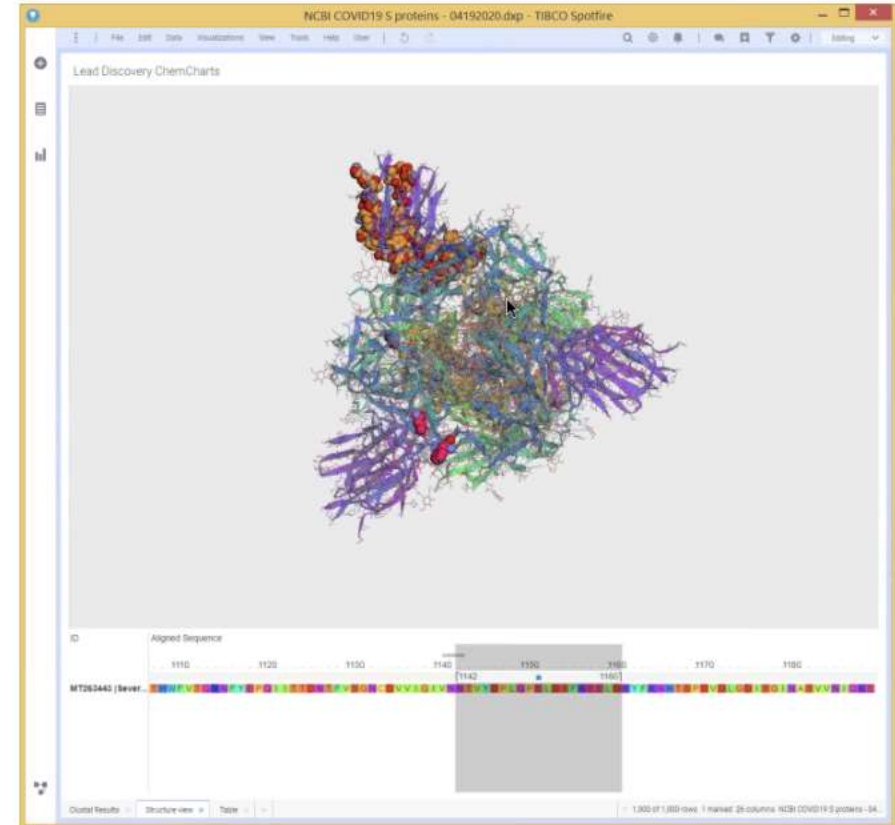
Genomics

The world research community has move at remarkable speed:

- >10,000 SARS-CoV-2 genomes sequence¹
- Crystal structure solved²
- Spike protein mutations that correlate with severity being identified



We gratefully acknowledge the Authors from Originating and Submitting laboratories of sequence data on which the analysis is based.



1. <https://www.epicov.org/epi3/frontend#>

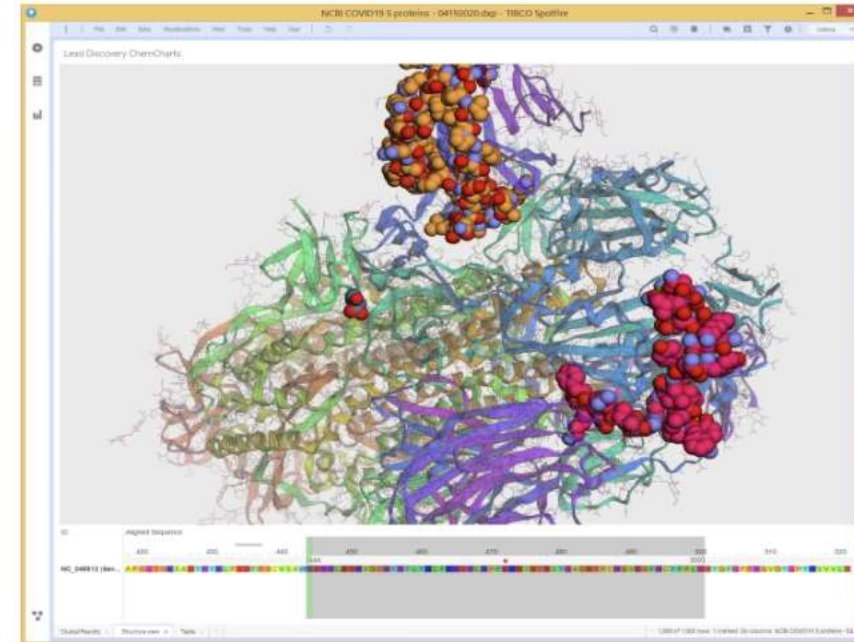
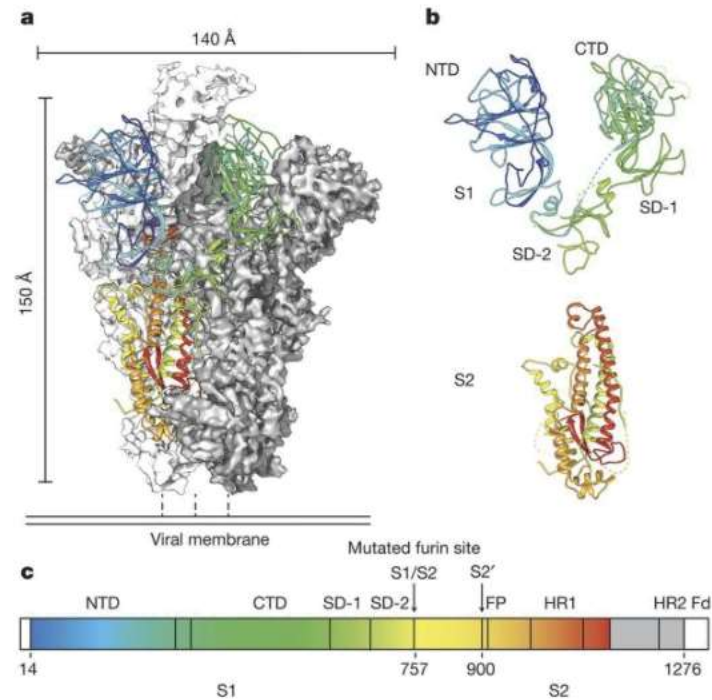
2. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. [Cell](https://doi.org/10.1016/j.cell.2020.02.058). 2020 Apr 16;181(2):281-292.e6. doi: 10.1016/j.cell.2020.02.058. Epub 2020 Mar 9.

The structure of the Spike

- The top of the Spike contains the Receptor Binding Domain (RBD) in 2 parts NTD and CTD.
- The bottom of the Spike mediate membrane fusion.
- The top has to be cleaved off for the bottom to work.

Structural overlay of mutations found:

- The red region is where most of the known mutations suspected to impact severity have been found.
- The orange region is where there seems to be mutations of unknown impact.
- The blue amino acid is the D614G mutation



1. Kirchdoerfer, R., Cottrell, C., Wang, N. et al. Pre-fusion structure of a human coronavirus spike protein. Nature 531, 118–121 (2016).

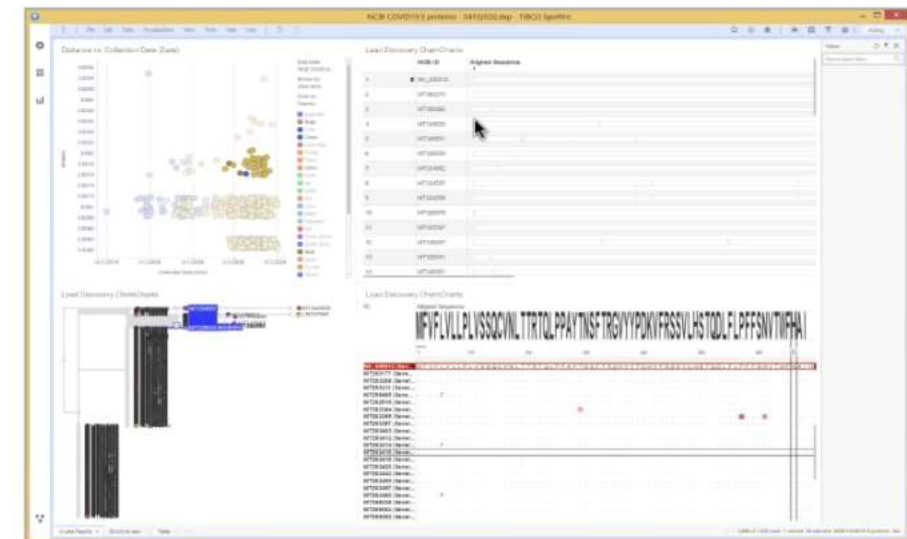
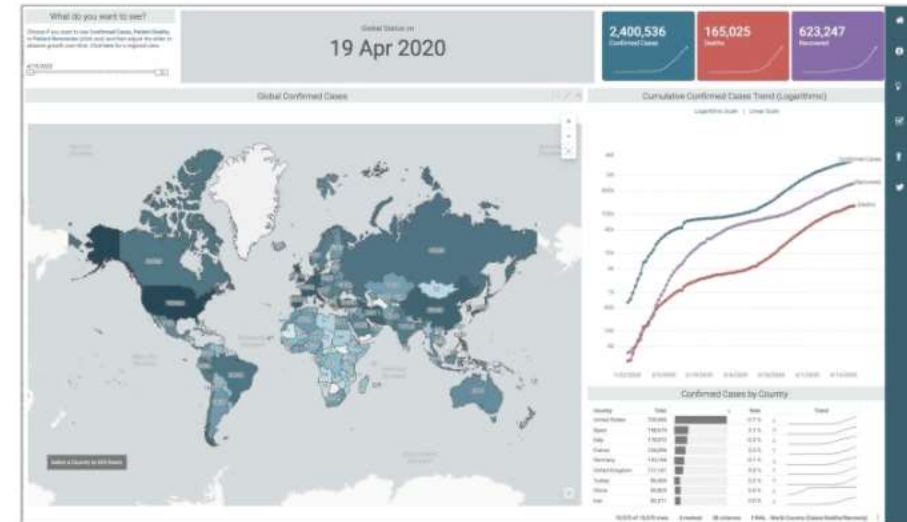
Demonstrations

Data Science & Epidemiology

- Reproduction Number Predictions
- Cases and Fatalities
- Interventions
- Healthcare locators

Genomics

- COVID19 Genome



Trouble viewing?
Use Cmd +/- to zoom

Scroll down to view this
entire page

COVID-19 Live Report

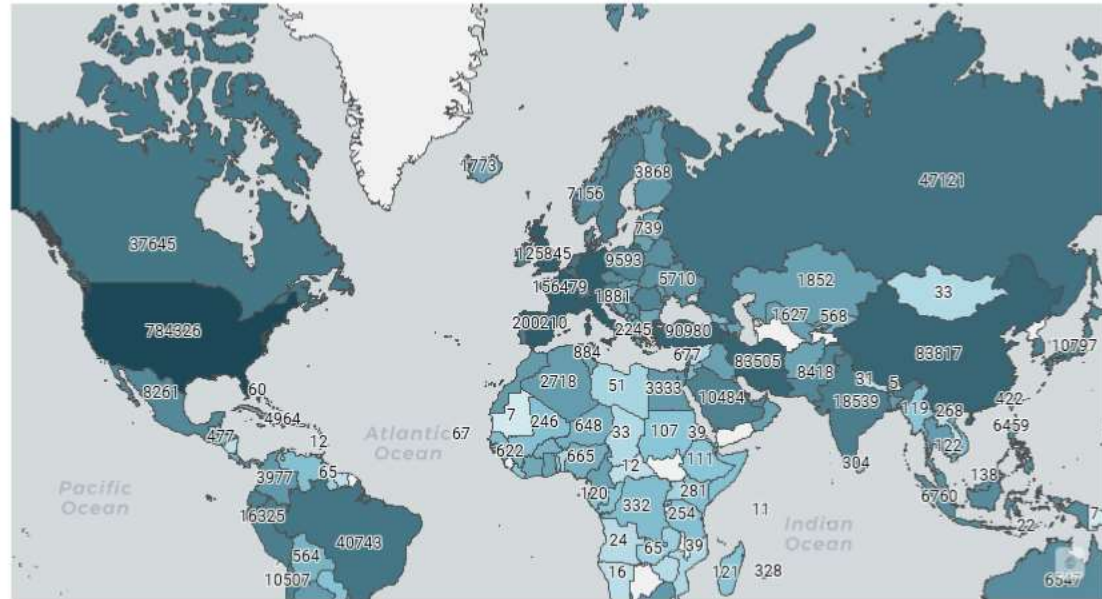
TIBCO Software, Inc. | Last Updated: 4/20/2020 7:20 PM | See Site Map

Global Status Check

As of April 20, there are at least 2,471,408 cases confirmed worldwide.

Dive deeper...

- Global Overview
- Regional Overview
- US Counties

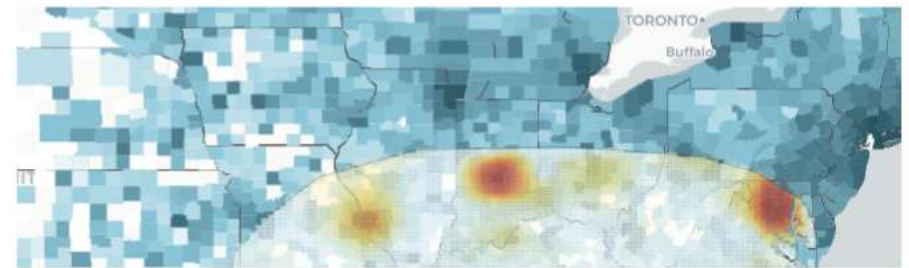


Top 10 Countries				
Country	Cases	Rate	Deaths	Rate
United States	784,326	-0.2%	42,094	-1.3%
Spain	200,210	-2.7%	20,852	-0.1%
Italy	181,228	-0.4%	24,114	0.1%
France	156,479	-1.6%	20,292	0.7%
Germany	147,065	0.0%	4,862	3.1%
United Kingdom	125,845	-0.9%	16,550	-0.9%
Turkey	90,980	0.8%	2,140	-0.2%
China	83,817	0.0%	4,636	0.0%
Iran	83,505	-0.1%	5,209	0.1%
Russia	47,121	-3.8%	405	-1.0%

Analyze Outbreak Hotspots – United States

Using local county level data, lasso select different regions of the United States to find the biggest outbreaks in a region. The calculations are relative to your selection so highlighting different regions will emphasize outbreaks in that selection.

Find Hotspots

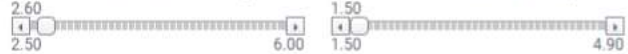


Tracking Re Estimates Globally

The Base Reproduction Number (R_0) is the number of people one person will infect if they have a disease, assuming no interventions. The Effective Reproduction Number (R_e) is how many people the disease *actually* infects, including interventions such as social distancing. The time between infections is the Serial Interval (SI) which can be adjusted here. Read the Page Info for more details.

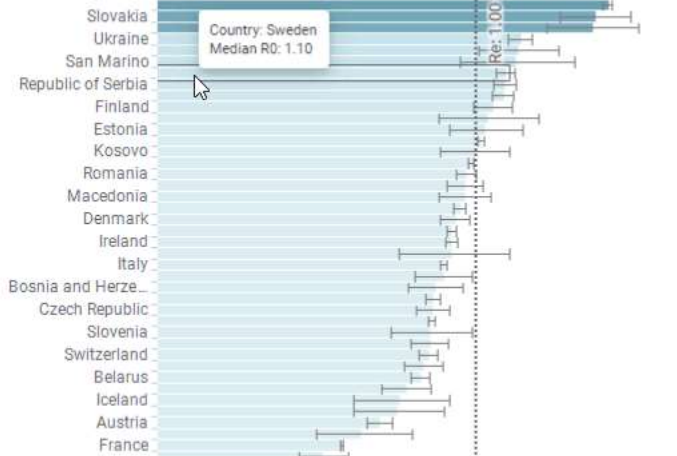
Mean Serial Interval: 2.6 days

Std Dev Serial Interval: 1.5 days

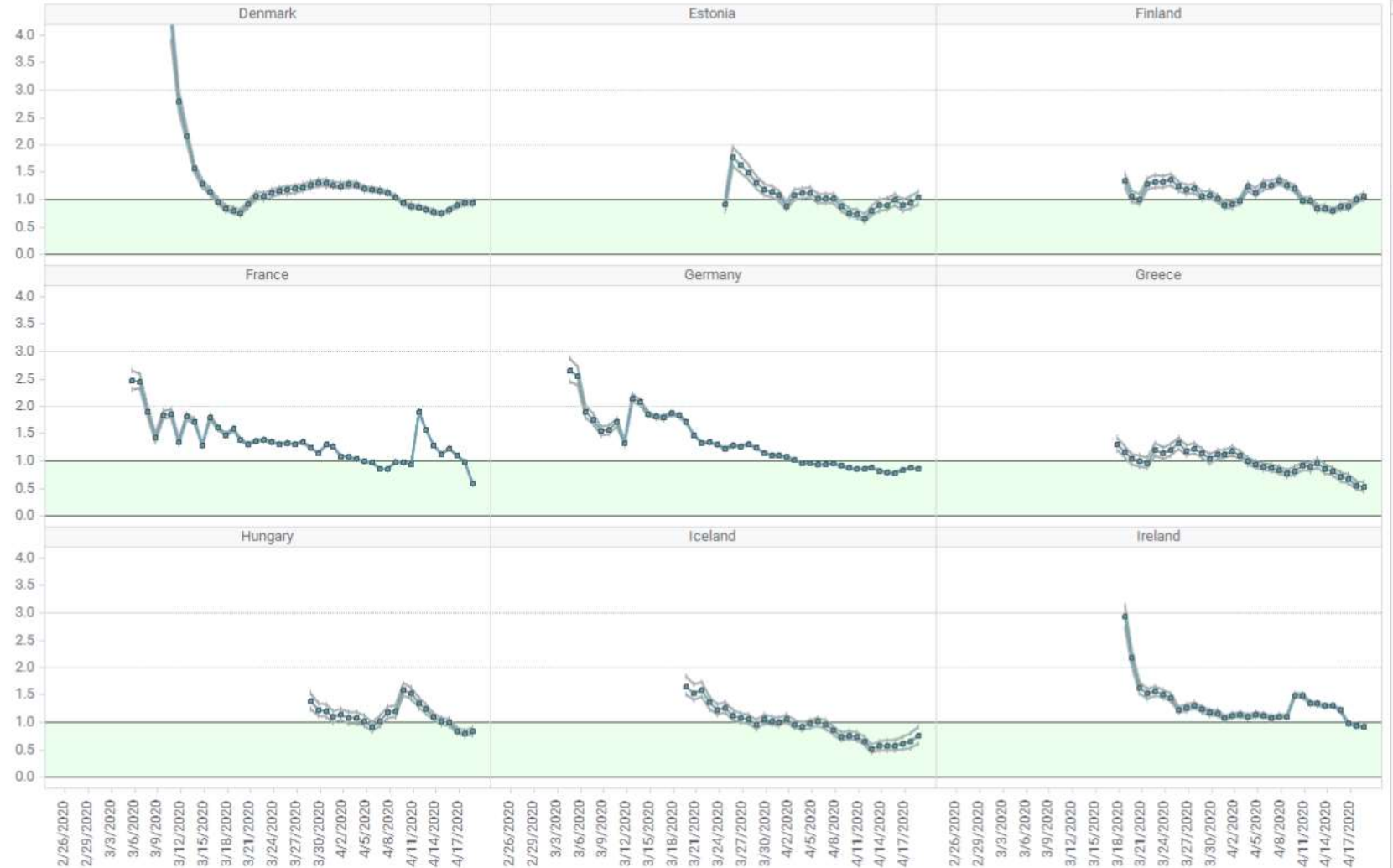


Estimate Re for a specific continent: Europe

Re Estimate by Country



R(t) Estimates by Country (after first 400 cases)



Compare Growth Rates

Confirmed Cases are growing exponentially around the world. Here, each country is shown time normalized after it's 100th case where reporting is assumed to be more consistent.

Type to search in list

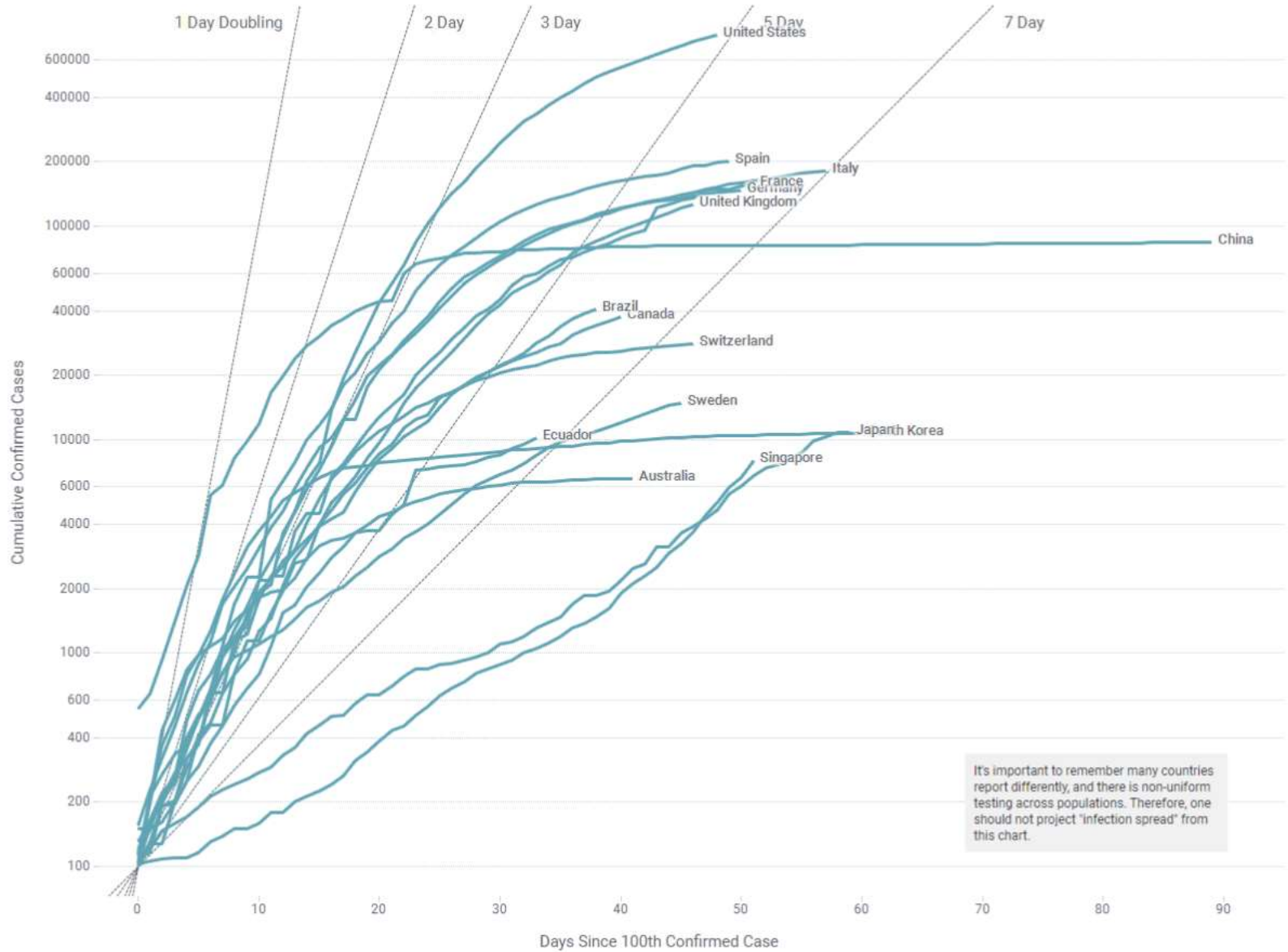
- (All) 175 values
- Afghanistan
- Albania
- Algeria
- Andorra

Choose countries to compare from the checklist.

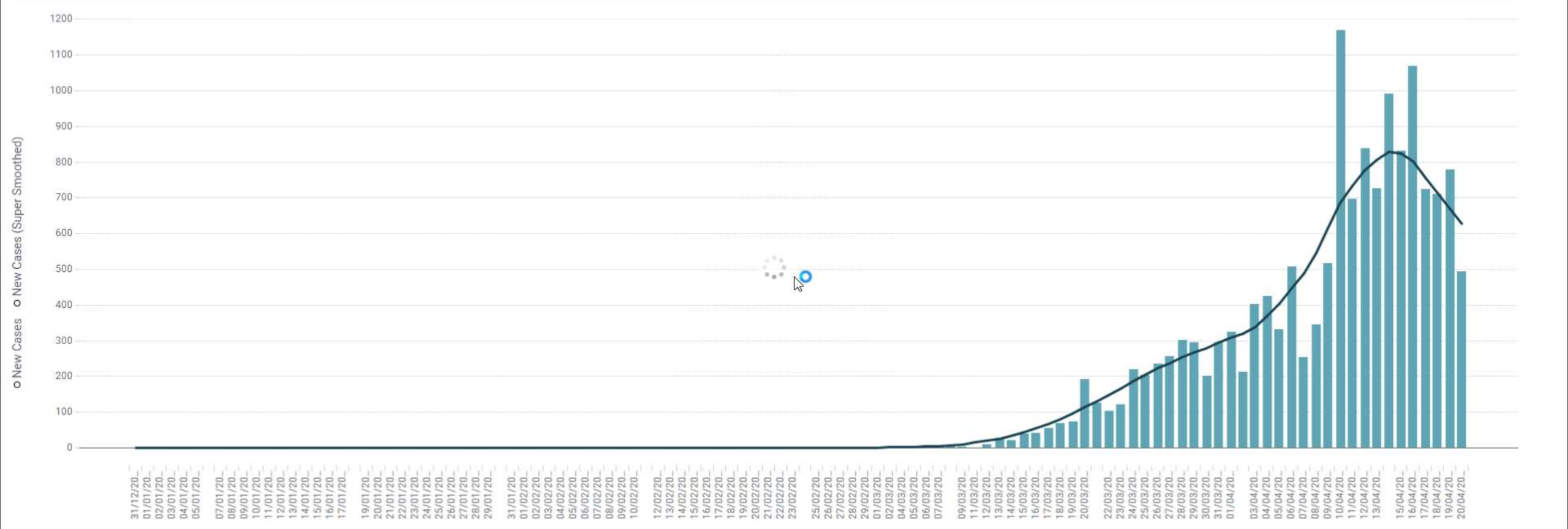
Cumulative Cases over Time



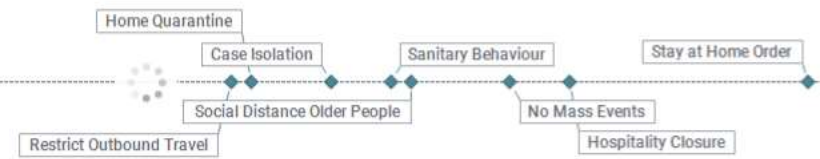
COVID-19 Case Trajectory by Country



It's important to remember many countries report differently, and there is non-uniform testing across populations. Therefore, one should not project "infection spread" from this chart.



Hover on timeline markers for more details



Where do you want to search?

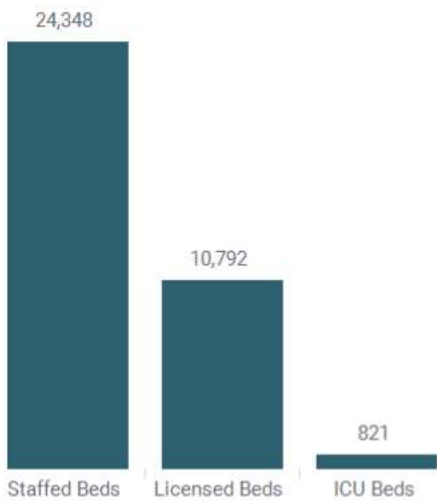
Enter a starting point below and the miles willing to drive, then click "Find Nearby Resources" to get hospital info within your desired drive radius.

Street Address (optional) Zip Code (required)

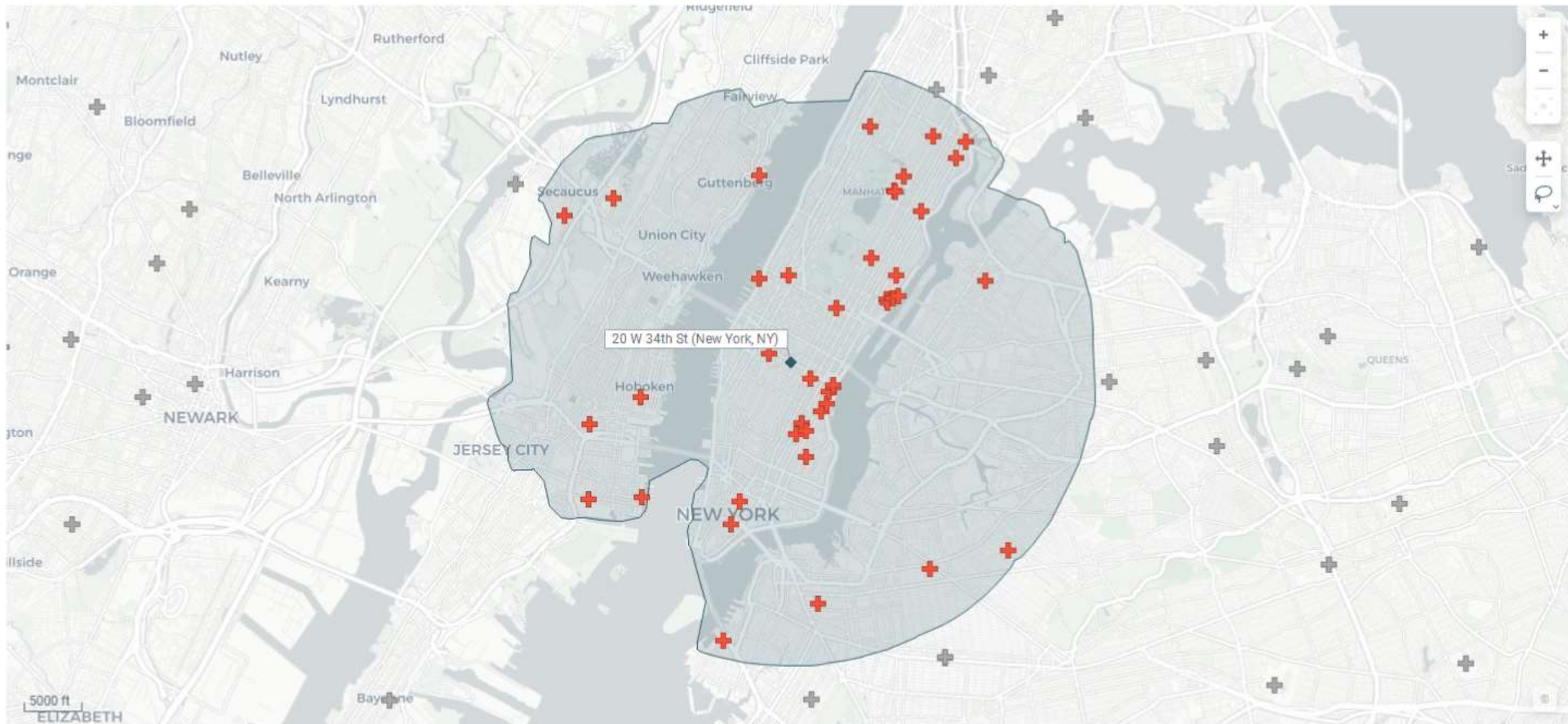
Driving Miles: 5.0
 (Scale: 1.00 to 20.00)

Find Nearby Resources

Known Beds within 5.0 Miles Drive



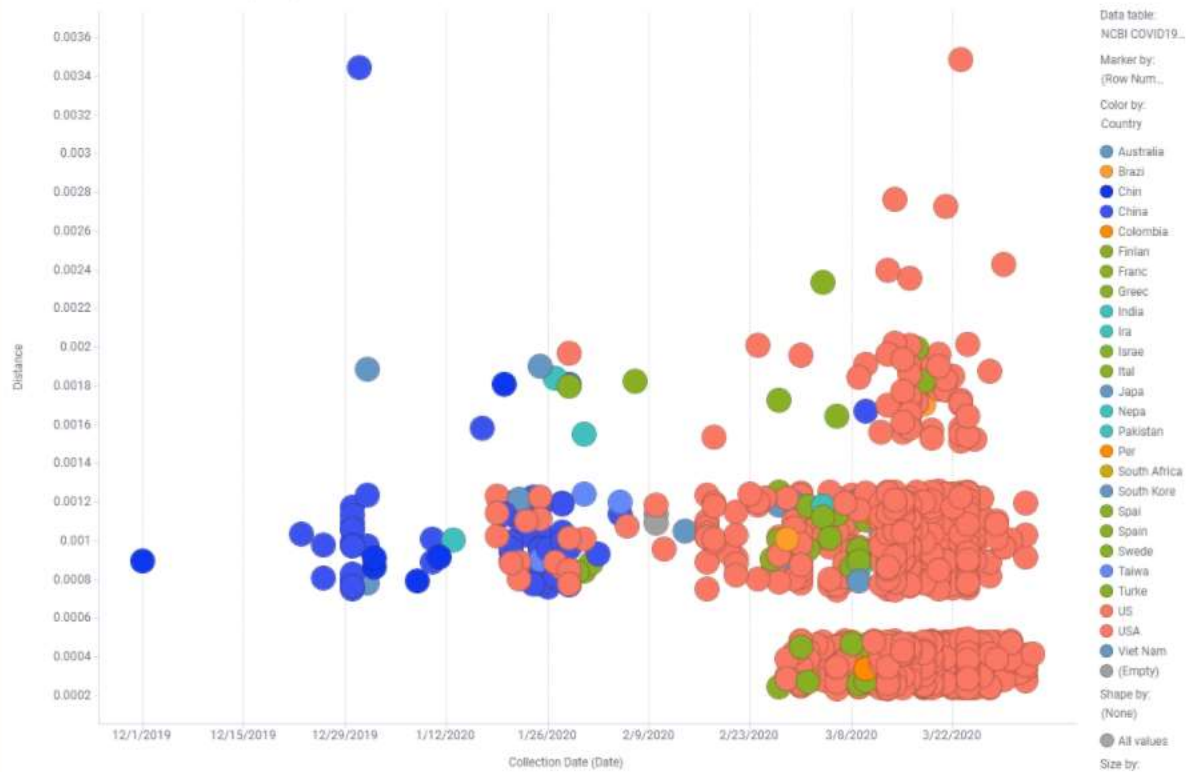
Hospitals within 5.0 miles of 20 W 34th St (New York, NY)



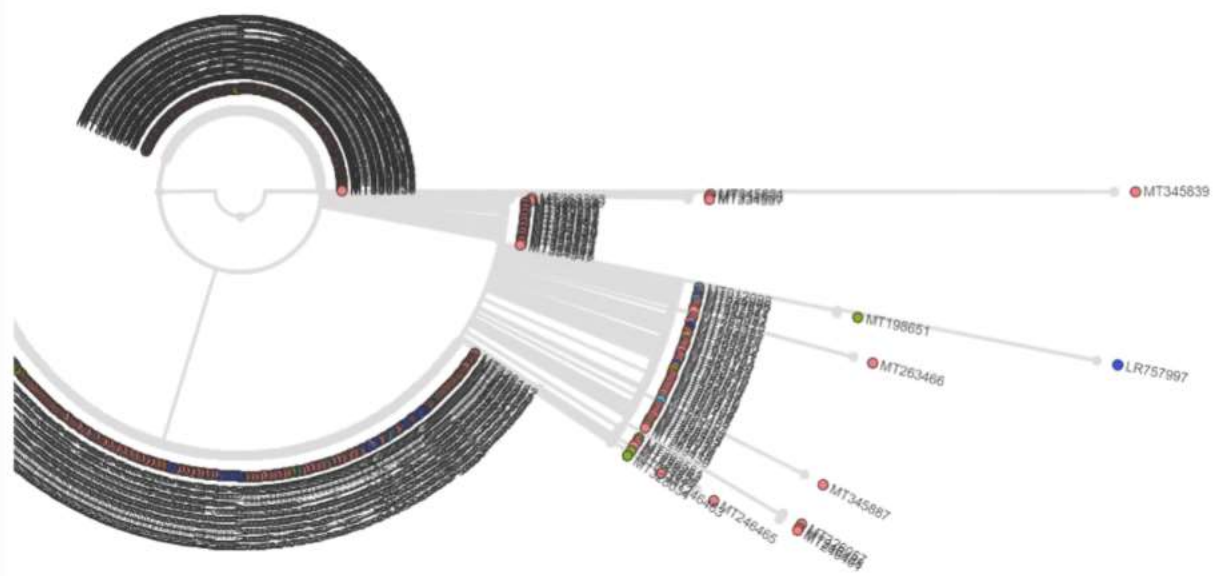
Hospital ID	Hospital Name	Hospital Type	Travel Miles	Phone	Address	City	State	Staffed Beds	Licensed Beds	ICU Beds	Bed Utilization
1012227	Javits Center Field Hospital (Temporari...	Short Term Acute Care Hos...	0.39		429 11th Ave	New York	NY				
582544	ArchCare Health System	Health System	0.41	855-951-2273	205 Lexington Ave	New York	NY	306			95.5%
2843	Tisch Hospital	Short Term Acute Care Hos...	0.79	212-263-6906	550 1st Ave	New York	NY	725	844	212	79.0%
579801	NYU Langone Health (FKA NYU Langon...	Health System	0.79	646-929-7870	550 1st Ave	New York	NY	2313			83.3%
850870	NYU Langone Hassenfeld Childrens Ho...	Childrens Hospital	0.80	212-263-7300	424 E 34th St	New York	NY	68			
981238	Helen L and Martin S Kimmel Pavillon	Short Term Acute Care Hos...	0.81	212-263-7300	424 E 34th St	New York	NY	306			
2844	NYC Health and Hospitals - Bellevue (F...	Short Term Acute Care Hos...	0.91	212-562-4141	462 1st Ave	New York	NY	527	912	45	69.5%
2841	VA NY Harbor Health Care System - Ma...	VA Hospital	0.96	212-686-7500	423 E 23rd St	New York	NY				
579802	Rusk Rehabilitation at NYU Langone Ort...	Long Term Acute Care Hosp...	1.03	212-263-6034	301 E 17th St	New York	NY	190	225		

DEFINITIVE HEALTHCARE
 Note: This dataset is intended to be used as a baseline for understanding the typical bed capacity and average yearly bed utilization of hospitals reporting such information. The date of last update received from each hospital may be varied. While the dataset is not updated in real-time, this information is critical for understanding the impact of a high utilization event, like COVID-19.
 Hospital Bed Data provided by

Distance vs. Collection Date (Date)



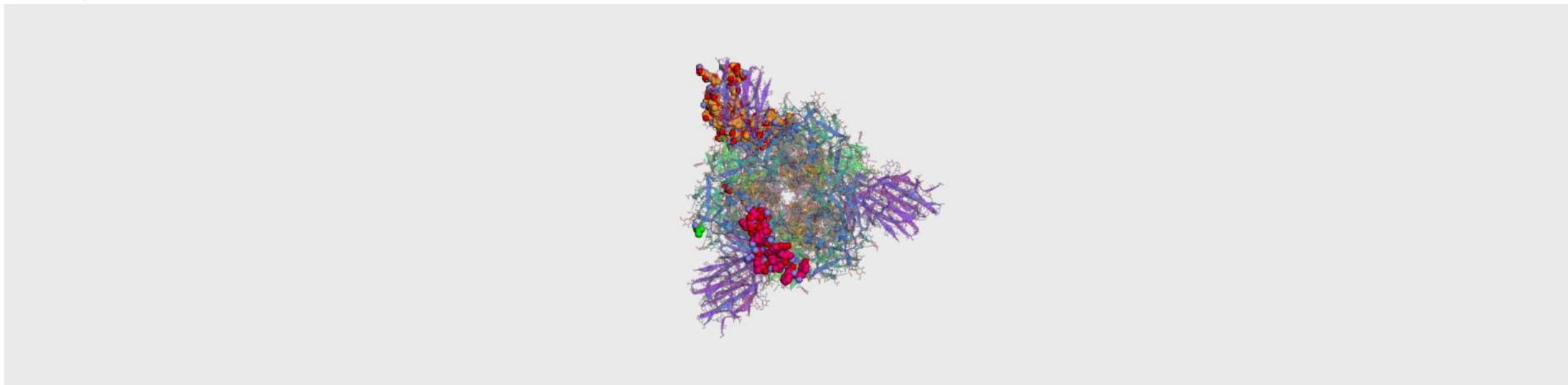
Lead Discovery ChemCharts



Lead Discovery ChemCharts

	NCBI ID	Country	Aligned Sequence	Dendr...	Distan...	Length	Authors	Geo_Location	GenBank_Title	Collection Dat...
1	NC_045512	China530.540.550.560.570.580.590.600.610. .	row_0	0.00	29903	Baranov,P.V.,H...	China	Severe acute re...	12/1/2019
2	MT350236	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_1	0.00	29865	Fink,L	USA: VA	Severe acute re...	3/1/2020
3	MT350237	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_2	0.00	29866	Fink,L	USA: VA	Severe acute re...	3/1/2020
4	MT350238	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_3	0.00	29826	Fink,L	USA: VA	Severe acute re...	3/1/2020
5	MT350239	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_4	0.00	29897	Fink,L	USA: VA	Severe acute re...	3/1/2020
6	MT350240	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_5	0.00	29899	Fink,L	USA: VA	Severe acute re...	3/1/2020
7	MT350241	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_6	0.00	29857	Fink,L	USA: VA	Severe acute re...	3/1/2020
8	MT350242	USA	HAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTS	row_7	0.00	29865	Fink,L	USA: VA	Severe acute re...	3/1/2020

Lead Discovery ChemCharts



ID Aligned Sequence

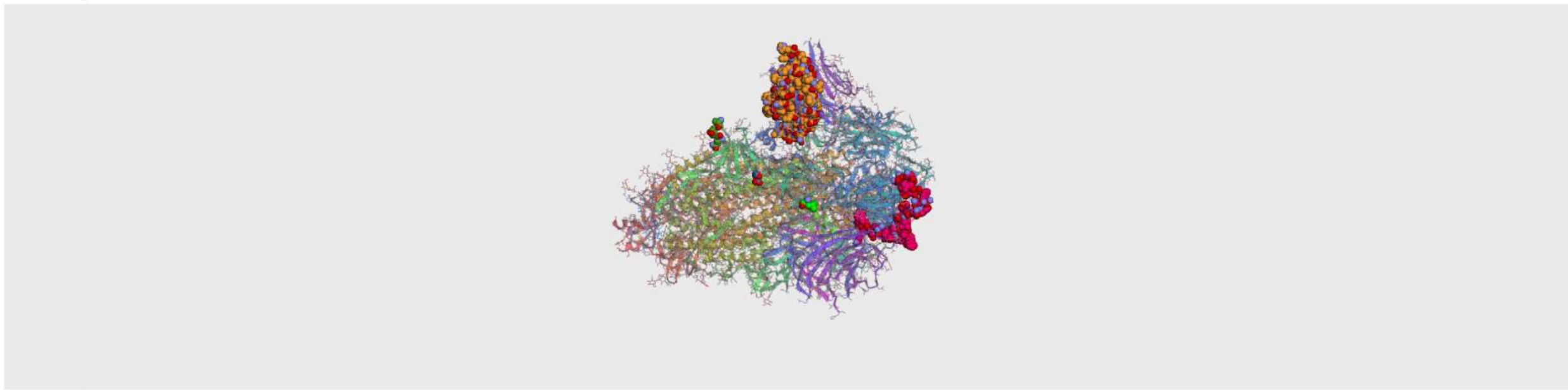
1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170

[2

NC_045612 | Sev... MEVFLVLLPLVSSQCVLITITQLUPPAYTNSFTIGVYYPDVVFSSVLSSTQGLPLPFFSNVITWFAITVVGTINGTIRFNPVLPFPIIGVYFASTERSHIIGWIFGITLDSITQSLIVNVAITVVIIVCFEFCVNPFLGVVYHNNHSWMESEFVYSSANNCITFELY



Lead Discovery ChemCharts

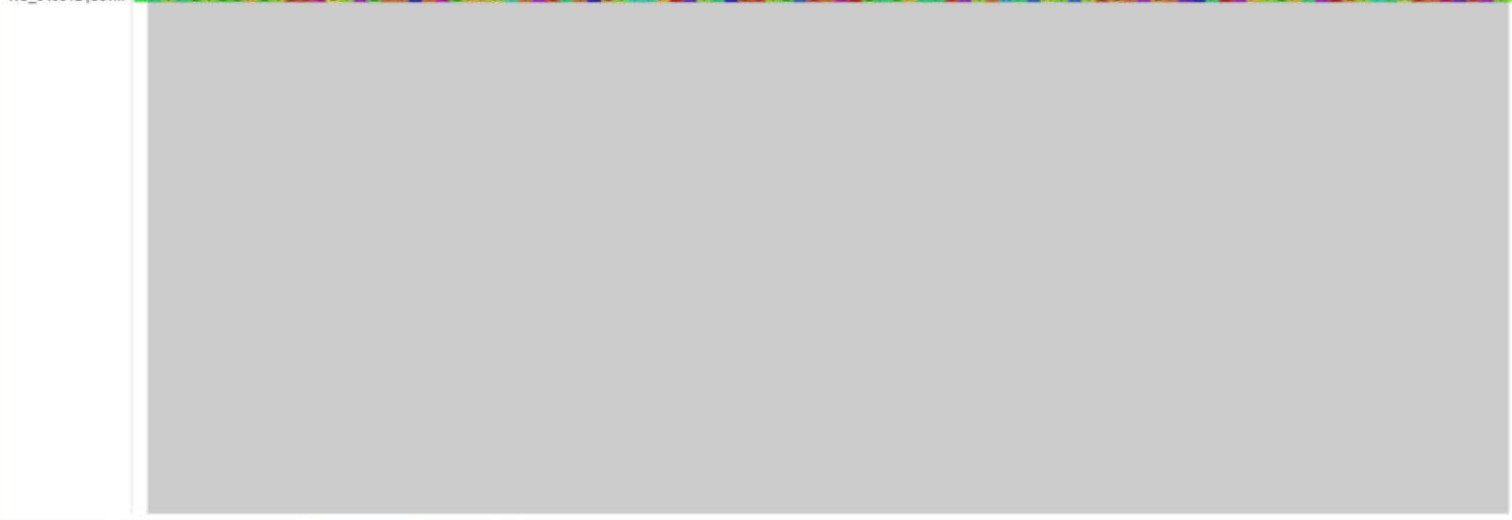


ID Aligned Sequence

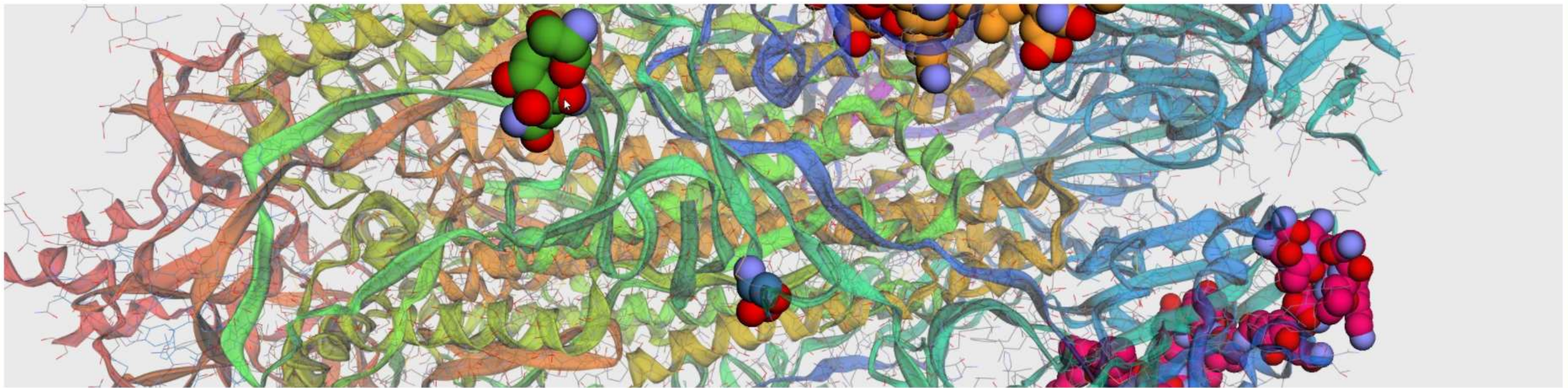
1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170

[2

NC_045512 | Sev... MEVFLVLLPLVSSQCVNLTITQLUPPAYTNSFTIGVYYPQVVFSSVLRSTQGLPLPFFSNVITWFAITVSGTNGTIRFNPVLPFPIIGVYFASTERSHIIGWIFGTLDSITQSLIVNVAITVVIIVCFEFCVNPFLGVVYHNNHSWMESEFVYSSANNCITFELY



Lead Discovery ChemCharts



ID

Aligned Sequence

	.540	.550	.560	.570	.580	.590	.600	.610	.620	.630	.640	.650	.660	.670	.680	.690	.700																																																																																																																																	
NC_045512 Sev...	S	T	N	L	V	N	C	V	N	F	N	F	G	L	T	G	T	G	V	L	T	E	S	N	F	E	P	F	Q	D	F	G	S	I	A	L	T	T	A	V	D	P	T	S	L	L	I	T	R	C	S	F	G	G	V	E	V	I	T	P	G	I	N	T	S	N	Q	V	A	V	L	Y	C	V	K	C	T	E	V	P	V	A	I	H	A	Q	L	T	P	T	W	V	Y	S	T	G	S	N	V	F	O	T	A	G	C	L	I	G	A	S	H	V	N	S	Y	C	L	I	P	I	G	A	G	I	C	A	S	Y	V	T	O	T	N	S	P	A	S	V	A	S	I	T	A	Y	T	M

[682 - 687]

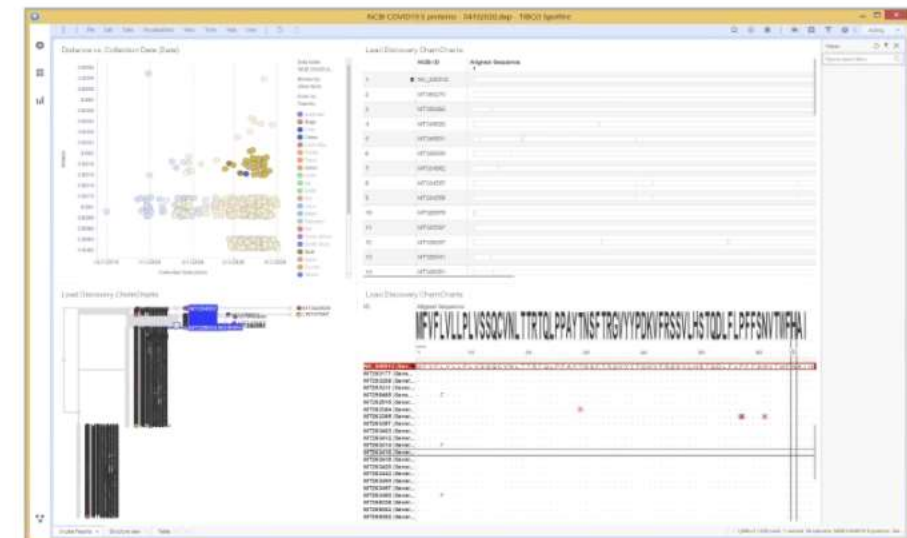
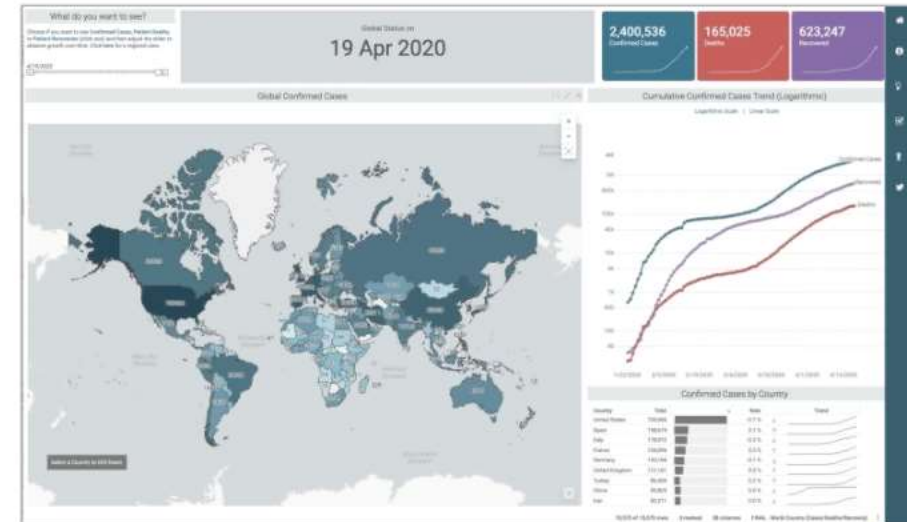
Demonstrations

Data Science & Epidemiology

- Reproduction Number Predictions
- Cases and Fatalities
- Interventions
- Healthcare locators

Genomics

- COVID19 Genome



CFR: the case fatality rate

Issue: How to assess the total number of COVID-19 infections?

Gupta (Oxford)

Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic

Jose Lourenco, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, Sunetra Gupta

doi: <https://doi.org/10.1101/2020.03.24.20042291>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Gupta described a number of scenarios. **The media picked up on the most extreme, which implied the majority of the UK had already been infected.** The Gupta analysis analysis focused solely on deaths.

Kucharski: Based on multiple studies it seems about 20–80% of people infected with Covid-19 could show symptoms. If this range turns out to be correct – and we combine it with our estimate that one in 15 people with symptoms are being reported – it would mean that hundreds of thousands of people in the UK have probably been infected with Covid-19 already, but not tens of millions.

Bendavid and Bhattacharya (Stanford)

WSJ | OPINION Michael O'Connell ▾
Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ Magazine Q

OPINION | COMMENTARY

Is the Coronavirus as Deadly as They Say?

Current estimates about the Covid-19 fatality rate may be too high by orders of magnitude.

[Bendavid and Bhattacharya](#) suggest CFR could be more like 0.01 to 0.1% ie more in line with seasonal Flu.

Argument based on under-reporting, lack of diagnostics
Existing evidence suggests that the virus is highly transmissible and that the number of infections doubles roughly every three days. An epidemic seed on Jan. 1 implies that by March 9 about six million people in the U.S. would have been infected.

We don't know the true infection rate in the U.S.
Antibody testing of representative samples to measure disease prevalence (including the recovered) is crucial.

Bendavid and Bhattacharya (Stanford)

COVID-19 Antibody Seroprevalence in Santa Clara County, California

Eran Bendavid, Bianca Mulaney, Neeraj Sood, Soleil Shah, Emilia Ling, Rebecca Bromley-Dulfano, Cara Lai, Zoe Weissberg, Rodrigo Saavedra, James Tedrow, Dona Tversky, Andrew Bogan, Thomas Kupiec, Daniel Eichner, Ribhav Gupta, John Ioannidis, Jay Bhattacharya

doi: <https://doi.org/10.1101/2020.04.14.20062463>

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Posted April 17, 2020.

 **Download PDF**

 Supplementary Material

 Data/Code

 Email

 Share

 Citation Tools

 Tweet

 Like 109K

Results

The unadjusted prevalence of antibodies to SARS-CoV-2 in Santa Clara County was 1.5% (exact binomial 95CI 1.11-1.97%), and the population-weighted prevalence was 2.81% (95CI 2.24-3.37%). Under the three scenarios for test performance characteristics, the population **prevalence of COVID-19 in Santa Clara ranged from 2.49% (95CI 1.80-3.17%) to 4.16% (2.58-5.70%). These prevalence estimates represent a range between 48,000 and 81,000 people infected in Santa Clara County by early April, 50-85-fold more than the number of confirmed cases.** Conclusions The population prevalence of SARS-CoV-2 antibodies in Santa Clara County implies that the infection is much more widespread than indicated by the number of confirmed cases. Population prevalence estimates can now be used to calibrate epidemic and mortality projections.

Assay Sensitivity and Specificity

A combination of both data sources provides us with a combined sensitivity of 80.3% (95 CI 72.1-87.0%) and a specificity of 99.5% (95 CI 98.3-99.9%).

The total number of positive cases by either IgG or IgM in our unadjusted sample was 50, a crude prevalence rate of 1.50% (exact binomial 95% CI 1.11-1.97%). **After weighting our sample to match Santa Clara County by zip, race, and sex, the prevalence was 2.81% (95% CI 2.24-3.37 without clustering the standard errors for members of the same household, and 1.45-4.16 with clustering).** We further **improved our estimation using the available data on test kit sensitivity and specificity**, using the three scenarios noted above. The estimated prevalence was **2.49%** (95CI 1.80%-3.17%) under the S1 scenario, **4.16%** (95CI 2.58%-5.70%) under the S2 scenario, and 2.75% (95CI 2.01%-3.49%) under the S3 scenario.

Bendavid and Bhattacharya - Issues

- B&B publish a confidence interval on the specificity of the test of 98.3% - 99.9%.
- But only 1.5% of all the tests are positive !
- Thus, if true specificity is close to 98.3%, then ALL of the positive results may be accounted as false positives.
- B&B then report a 95% confidence interval for the prevalence of COVID-19 in Santa Clara County of 2.01% to 3.49% (using delta method, adjusting for patient sample demographics and test specificity).

Concerns with the B&B paper

- unstable population weights, adjustment, error propagation
- can specificity be this high?
- consent bias
- consistency with other serosurvey data?

Statistical Appendix

In this appendix, we describe our statistical methods in more detail. Section 1 describes our approach to calculating population weights. Section 2 describes our approach to adjusting our population prevalence estimate for the sensitivity and specificity properties of the LFA test kit we are using. Finally, Section 3 describes our approach to incorporating three separate sources of uncertainty in our prevalence estimates: sampling variability, error in the sensitivity estimate, and error in the specificity estimate.

We solve for π as a function of the sample prevalence, sensitivity, and specificity:

$$\pi = \frac{q + s - 1}{r + s - 1}$$

There is one important caveat to this formula: it only holds as long as (one minus) the specificity of the test is higher than the sample prevalence. If it is lower, all the observed positives in the sample could be due to false-positive test results, and we cannot exclude zero prevalence as a possibility. As long as the specificity is high relative to the sample prevalence, this expression allows us to recover population prevalence from sample prevalence, despite using a noisy test.

See: John Cherian (@jjcherian) for alternate bootstrap analysis; avoiding asymptotic assumptions of delta method
https://github.com/jjcherian/medrxiv_experiment



Sensitivity and Specificity

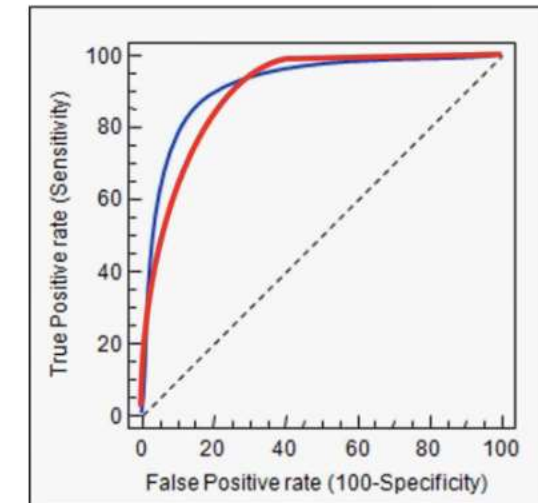
- **SENSITIVITY measures how well we can detect patients with disease.** Imagine a nasal swab from someone with SARS-CoV-2 infection. Sensitivity is the probability this sample will test positive.
- **With RT-PCR**, the specimen is declared positive if viral RNA is detected. Patients with COVID19 often have high viral loads in their throats. Thus it is relatively easy to detect virus, and the **test is typically highly sensitive**.
- **When sensitivity is low, we get FALSE NEGATIVES.** These are swabs from which no virus is detected even though the person is infected. This could occur if the specimen was degraded, virus didn't amplify well, etc.

		VIRUS	
		+ve	-ve
TEST	+ve	a	b
	-ve	c	d

$$\text{sensitivity} = a / (a+c)$$

$$\text{specificity} = b / (b+d)$$

ROC Curve



See: Natalie Dean (@nataliexdean)

Sensitivity and Specificity

- **SENSITIVITY measures how well we can detect patients with disease.** Imagine a nasal swab from someone with SARS-CoV-2 infection. Sensitivity is the probability this sample will test positive.
- **With RT-PCR**, the specimen is declared positive if viral RNA is detected. Patients with COVID19 often have high viral loads in their throats. Thus it is relatively easy to detect virus, and the **test is typically highly sensitive.**
- **When sensitivity is low, we get FALSE NEGATIVES.** These are swabs from which no virus is detected even though the person is infected. This could occur if the specimen was degraded, virus didn't amplify well, etc.
- **SPECIFICITY measures how well we rule out infection for people who are tested but aren't infected.** Imagine a nasal swab from someone without SARS-CoV-2 infection. This should come back negative on RT-PCR. Specificity measures how often these come back negative.
- **Because RT-PCR measures viral RNA, it is rare for someone to test positive if they aren't infected.** It can happen, though e.g. if there is random contamination in the lab. In general, **RT-PCR is usually highly specific and there are few FALSE POSITIVES.**

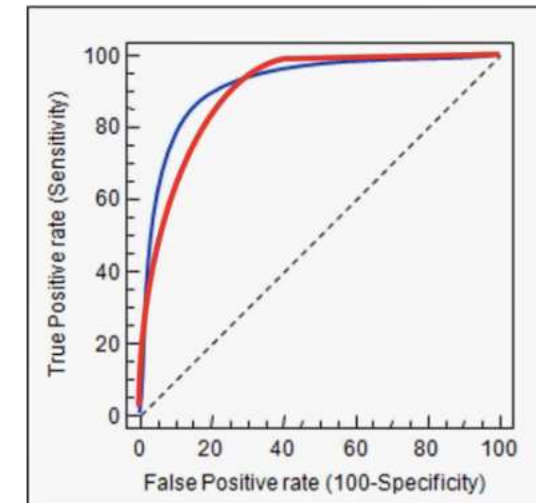
See: Natalie Dean (@nataliexdean)

		VIRUS	
		+ve	-ve
TEST	+ve	a	b
	-ve	c	d

$$\text{sensitivity} = a / (a+c)$$

$$\text{specificity} = b / (b+d)$$

ROC Curve



So what is %population infected?

What is the under-reporting of cases?

Table 1: Posterior model estimates of percentage of total population infected as of 28th March 2020.

Country	% of total population infected (mean [95% credible interval])
Austria	1.1% [0.36%-3.1%]
Belgium	3.7% [1.3%-9.7%]
Denmark	1.1% [0.40%-3.1%]
France	3.0% [1.1%-7.4%]
Germany	0.72% [0.28%-1.8%]
Italy	9.8% [3.2%-26%]
Norway	0.41% [0.09%-1.2%]
Spain	15% [3.7%-41%]
Sweden	3.1% [0.85%-8.4%]
Switzerland	3.2% [1.3%-7.6%]
United Kingdom	2.7% [1.2%-5.4%]

Source: [report by @MRC Outbreak](#), March 30

Antistoffen tegen SARS-CoV-2 bij ~3% Nederlandse bloeddonors

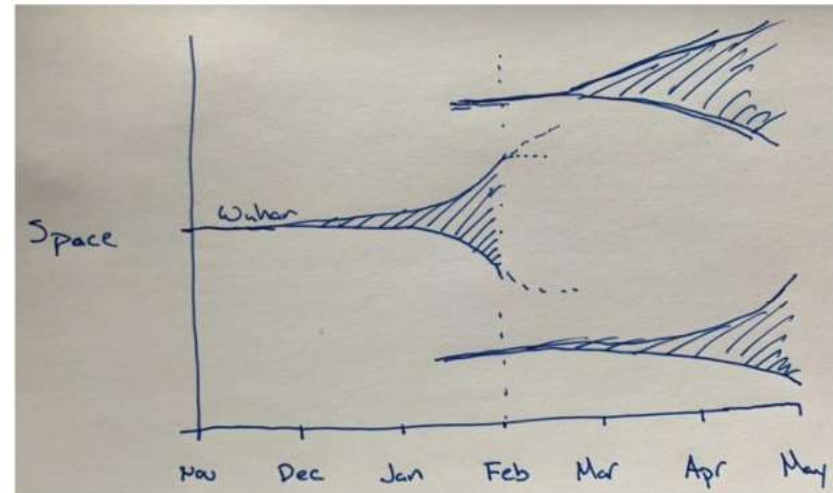
18-30 jaar	25 / 688	3,6%
31-40 jaar	17 / 494	3,4%
41-50 jaar	26 / 752	3,5%
51-60 jaar	38 / 1234	3,1%
61-70 jaar	29 / 1030	2,8%
71-80 jaar	0 / 10	(0%)

Source: [@rivm](#) (Dutch CDC) presentation for members of parliament.

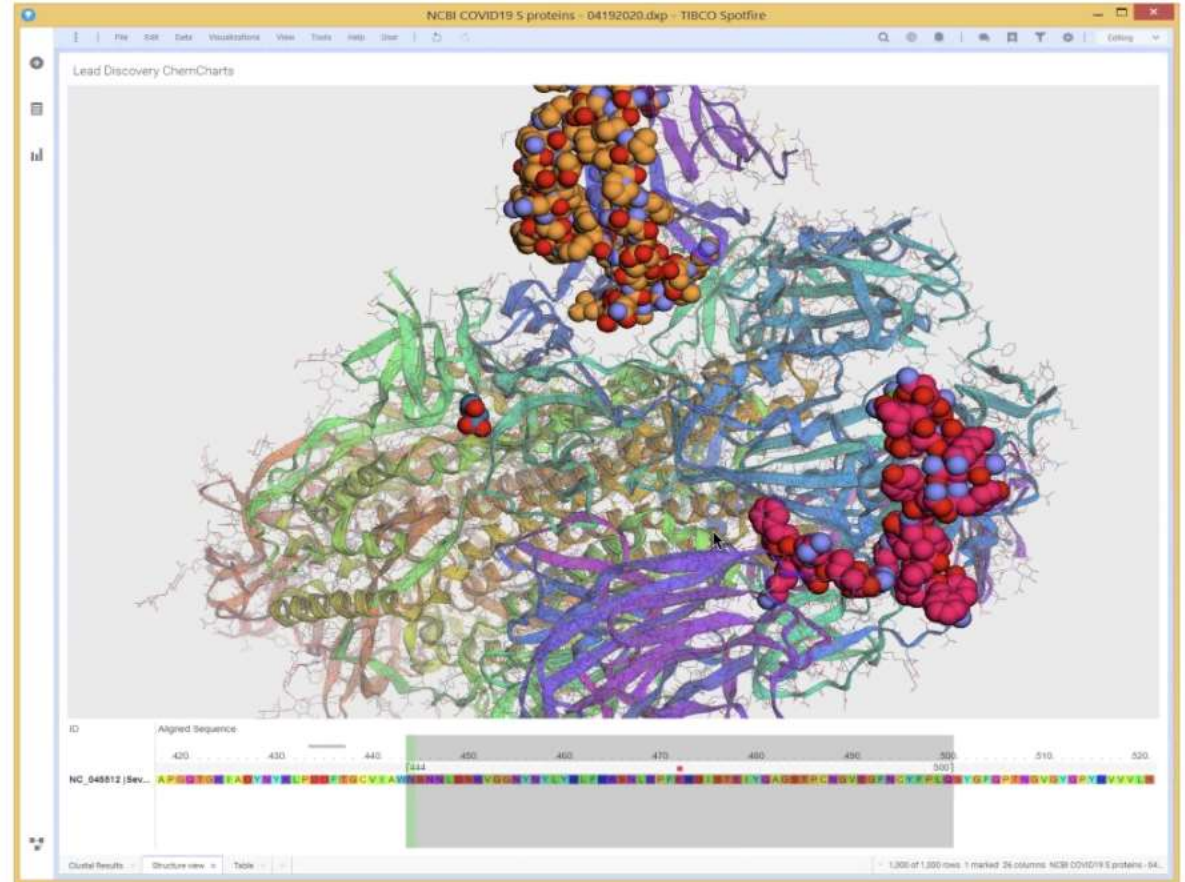
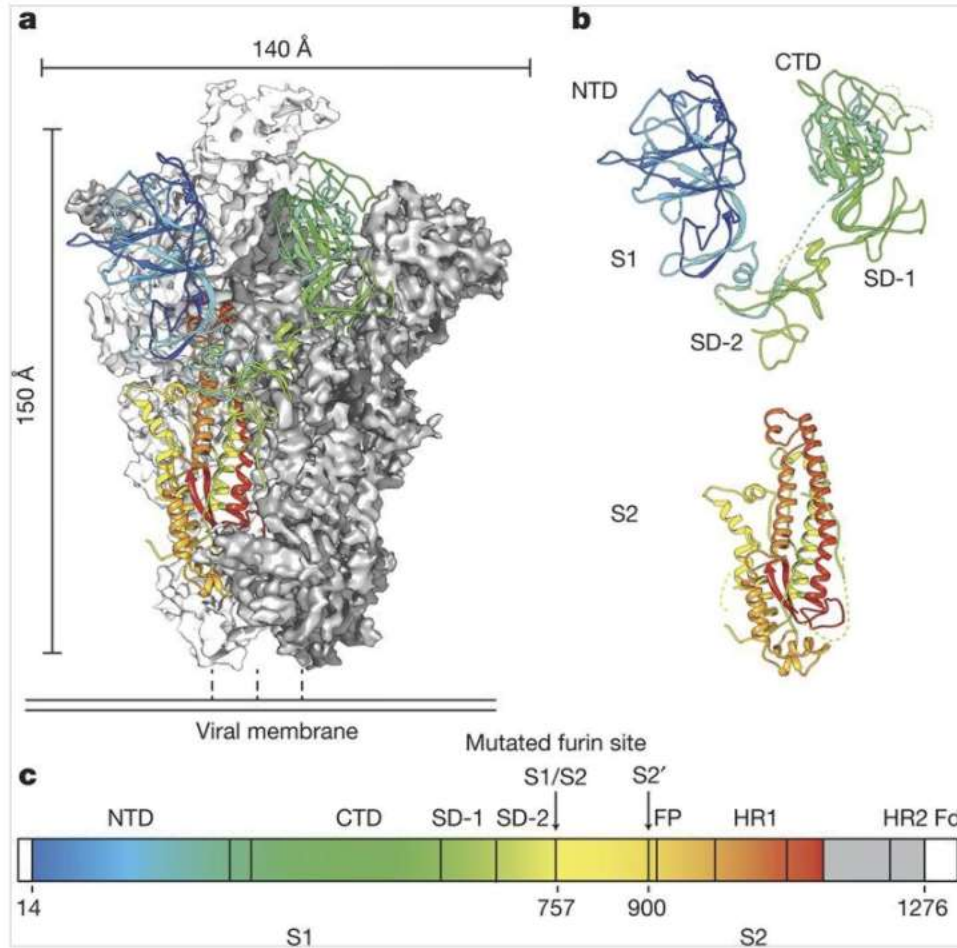
Trevor Bedford (@trvr) on the US undercount :- predicts 10-20X more people in the US are infected c.f. reported cases; based on ~20 sparks that caught between Jan 15 and Feb 15 and have grown into outbreaks.

So, we have ~760k cases as of Apr 20 in US. With 10-20% under-reporting, this means we may have 7.5-15M infected people in the US at this time, or around 2-4% of the US population.

This is consistent with these other WW results.

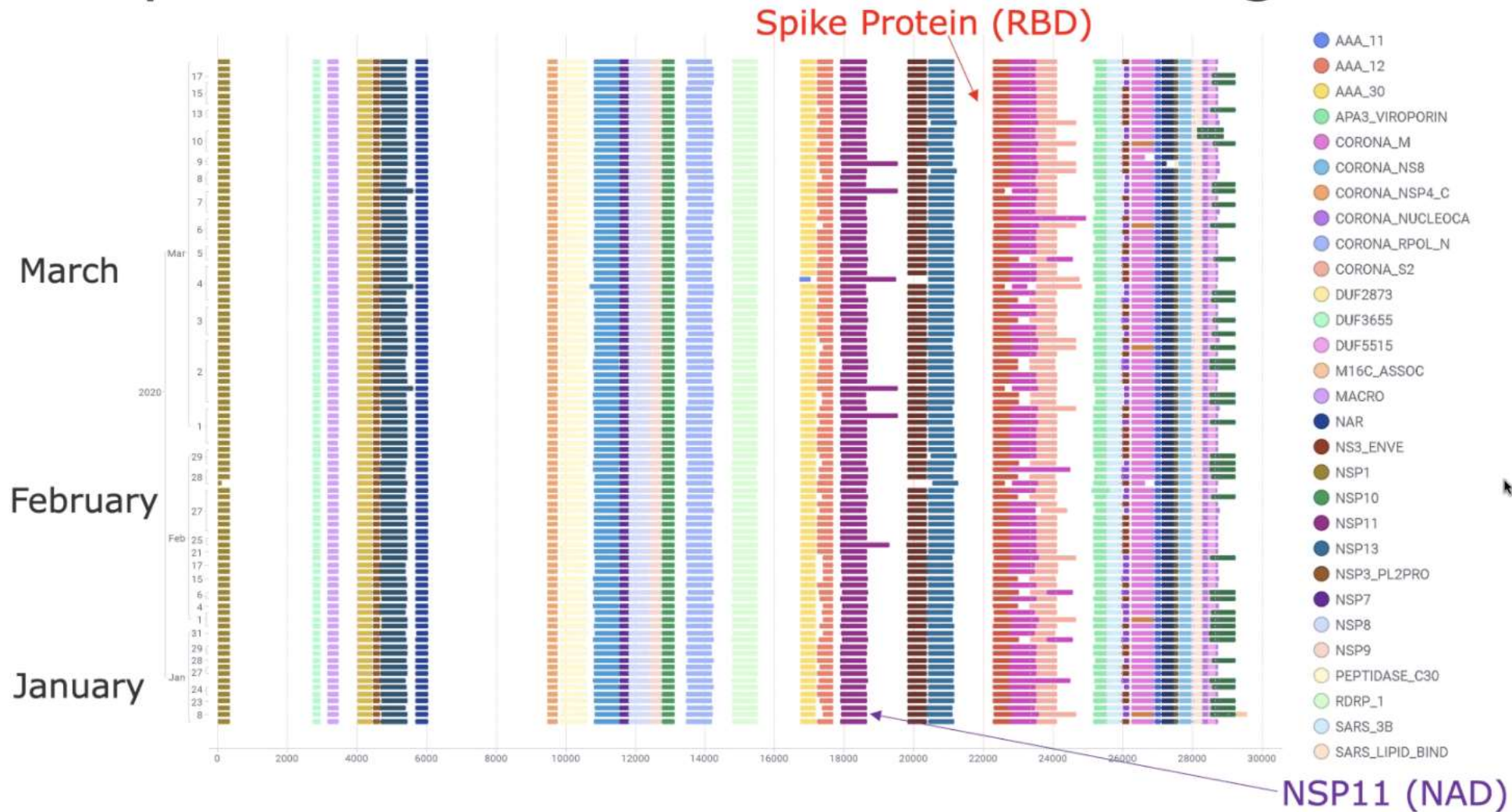


The structure of the Spike



Kirchdoerfer, R., Cottrell, C., Wang, N. et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* 531, 118–121 (2016).

Spike in context of domain of full genome

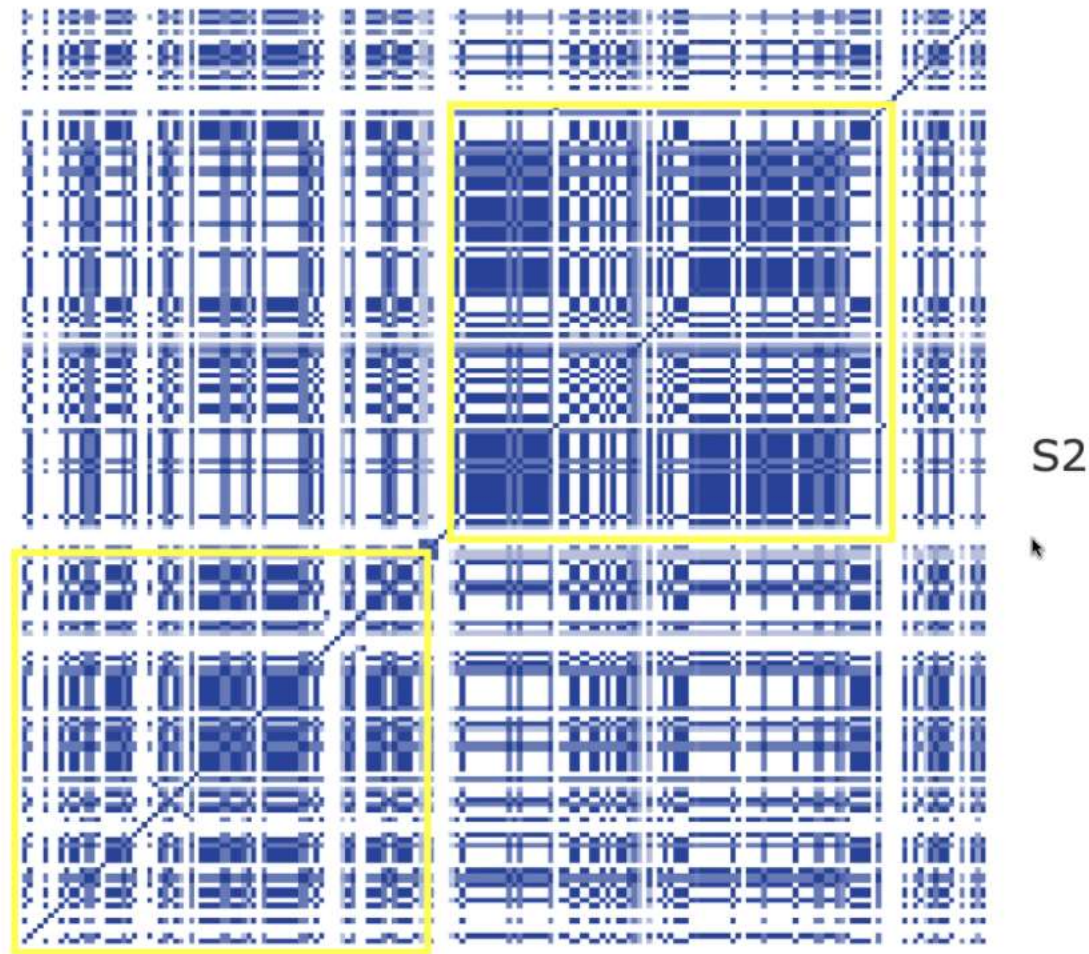


We gratefully acknowledge these Data provided by William Buscher and Jack Bramley of Washington University

Preliminary Coupling Analysis near Spike RBD Domain

- Researchers are beginning to look for more complex genetic trends.
- “Coupling Analysis” shows statistical signals of sequences that co-vary and thus may be structurally important.

Spike
RBD



We gratefully acknowledge these Data provided by William Buscher and Jack Bramley of Washington University

Wrap-Up

Common Misconceptions

Back to Work Planning

Don't divide case curves by population

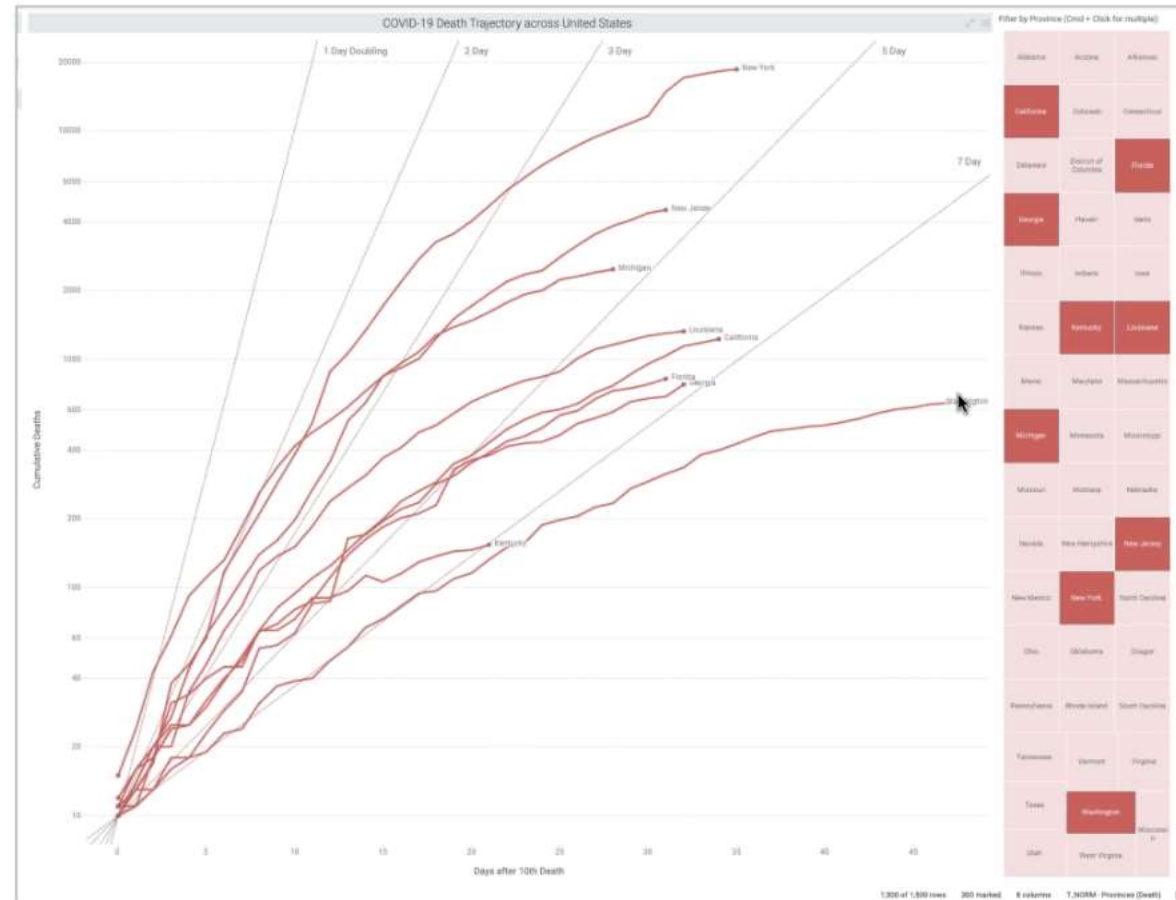
without %pop-normalizing

Logarithmic scale on the vertical axis with start at same number of cases is best

- x-axis is days since first <n> cases
- slopes reflect growth rates
- heights reflect prevalence
- relative timing is maintained

If you divide by population

- Need to start at same % of population infected
- x-axis is days since first <%> population infected



Don't divide case curves by population

without %pop-normalizing

Logarithmic scale on the vertical axis with start at same number of cases is best

- x-axis is days since first <n> cases
- slopes reflect growth rates
- heights reflect prevalence
- relative timing is maintained

If you divide by population

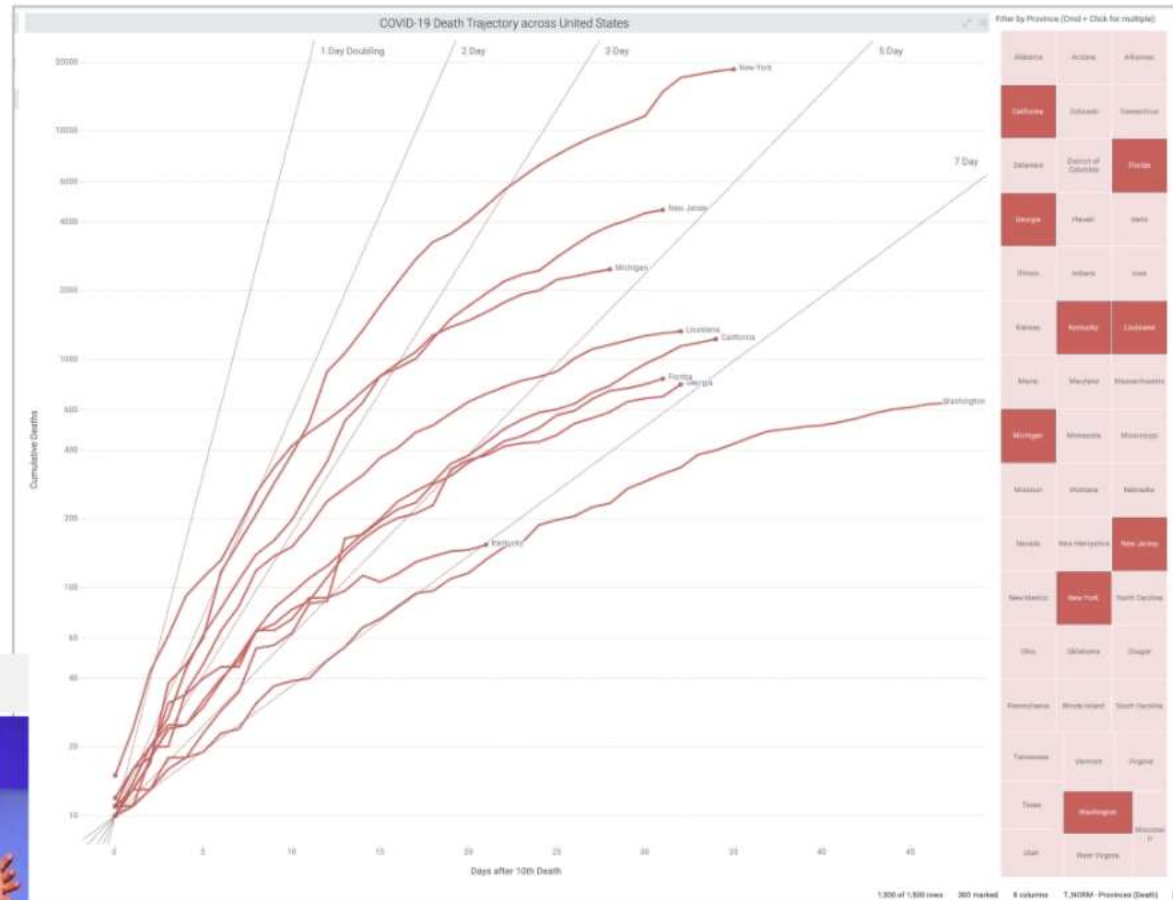
- Need to start at same % of population infected
- x-axis is days since first <%> population infected

Exponential Growth is alarming

1 and 2 weeks growth

- $2^7 = 128$
- $2^{14} = 16,384$

Ray Kurzweil



Don't divide case curves by population

without %pop-normalizing

Logarithmic scale on the vertical axis with start at same number of cases is best

- x-axis is days since first <n> cases
- slopes reflect growth rates
- heights reflect prevalence
- relative timing is maintained

If you divide by population

- Need to start at same % of population infected
- x-axis is days since first <%> population infected

Pay close attention to axis scales

- Especially in the media
- the logarithmic scale !! (Yann LeCun, Twitter)

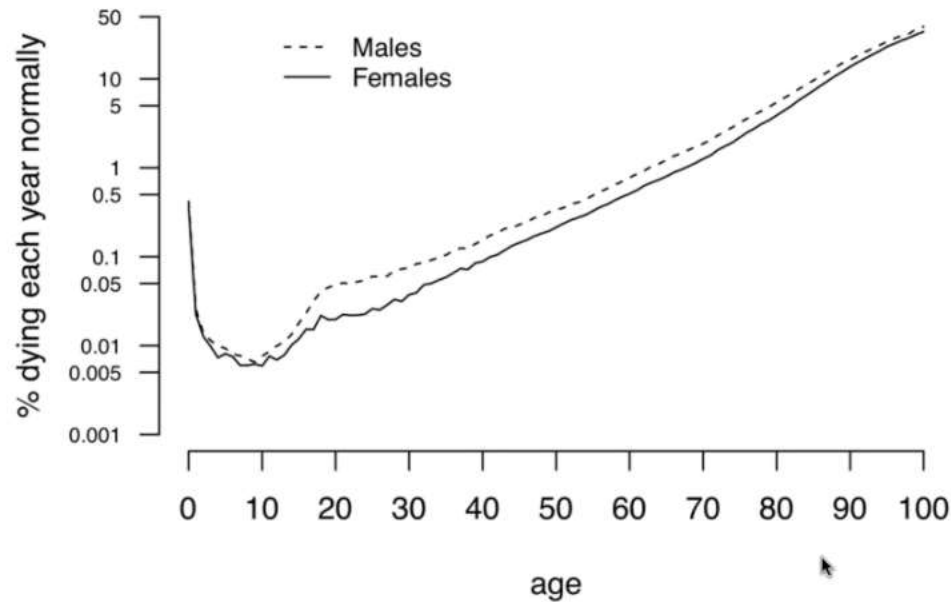


Effects of Age on COVID Fatalities

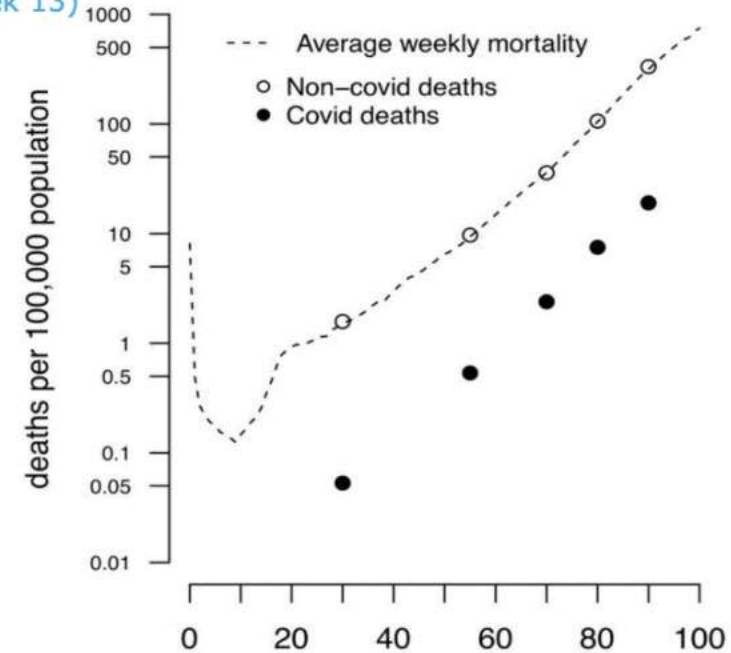
- Risk(death) increases exponentially with age in COVID and non-COVID people
- COVID packs ~1 years risk into 3 weeks
- COVID multiplies baseline risk similarly across age bands
- Males : 5.8% of deaths are COVID, Females : 3.8% are COVID

Risk of Dying vs Age for non-COVID infected people. UK ONS

- proportion increases ~9% per year
- average risk of death doubles every ~8 years

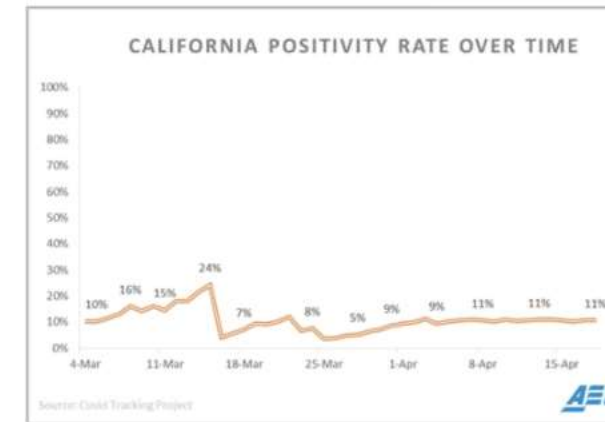
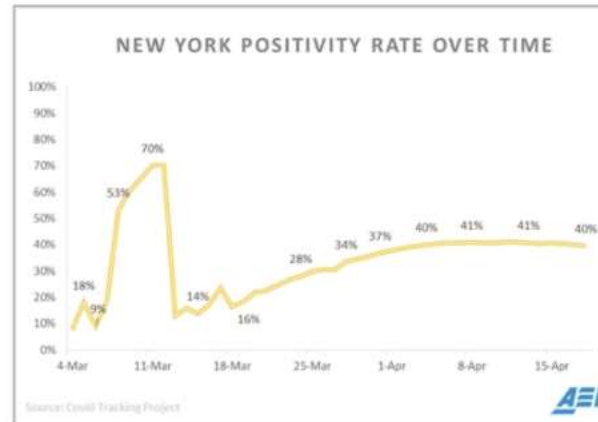
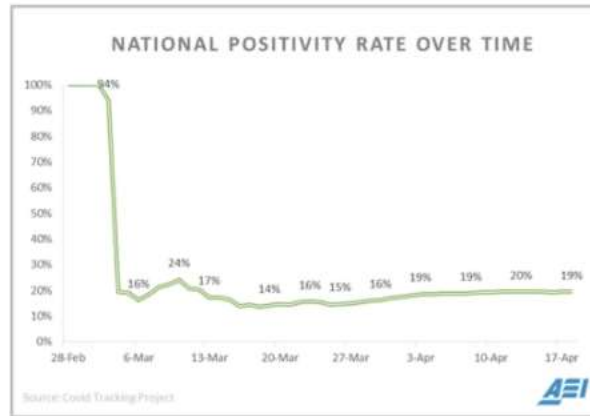


UK REGISTERED deaths: COVID and non-COVID. March 21 - 27 (week 13)

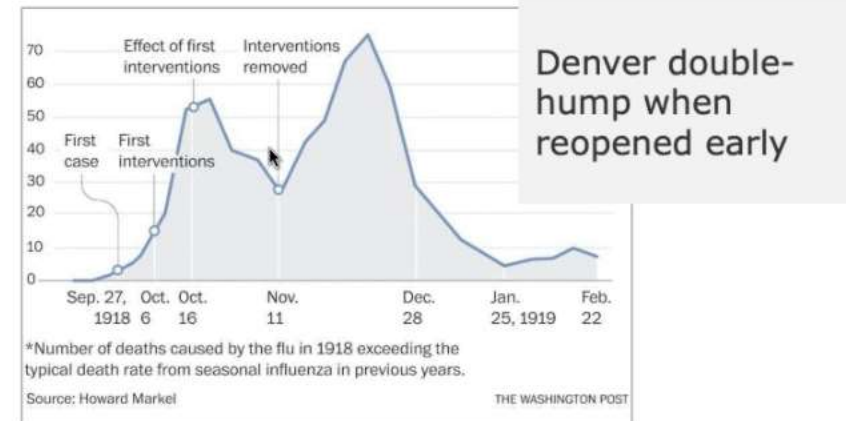
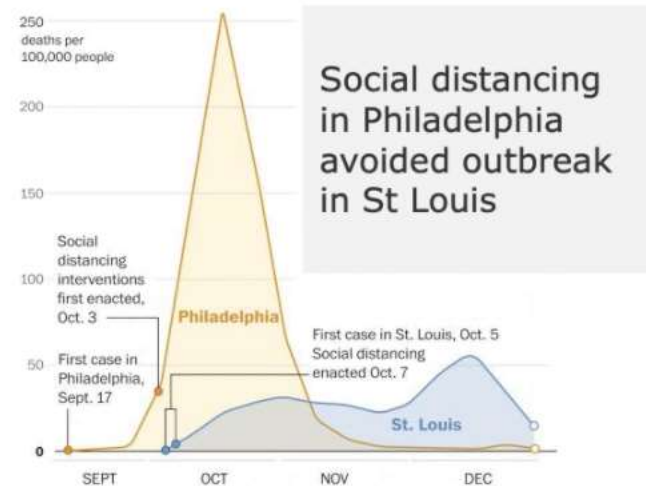


Back to Work Planning

Positive Test Rates remain high – epidemic still underway



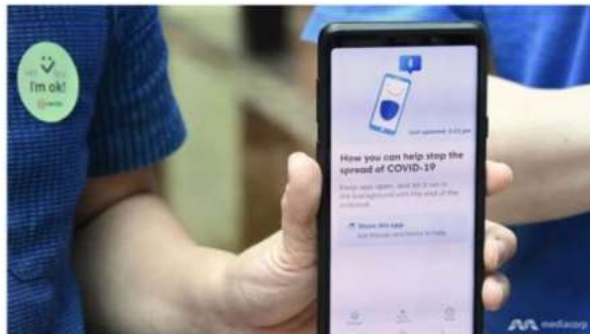
Learnings from 1918 Spanish Flu



Back to Work Planning

Singapore

Singapore launches TraceTogether mobile app to boost COVID-19 contact tracing efforts



The TraceTogether app was launched on Mar 20, 2020 to support ongoing COVID-19 contact tracing efforts. (Photo: Hanidah Amin)

SINGAPORE: A new mobile application called TraceTogether was launched on Friday (Mar 20) to support ongoing contact tracing efforts amid the COVID-19 outbreak in Singapore.

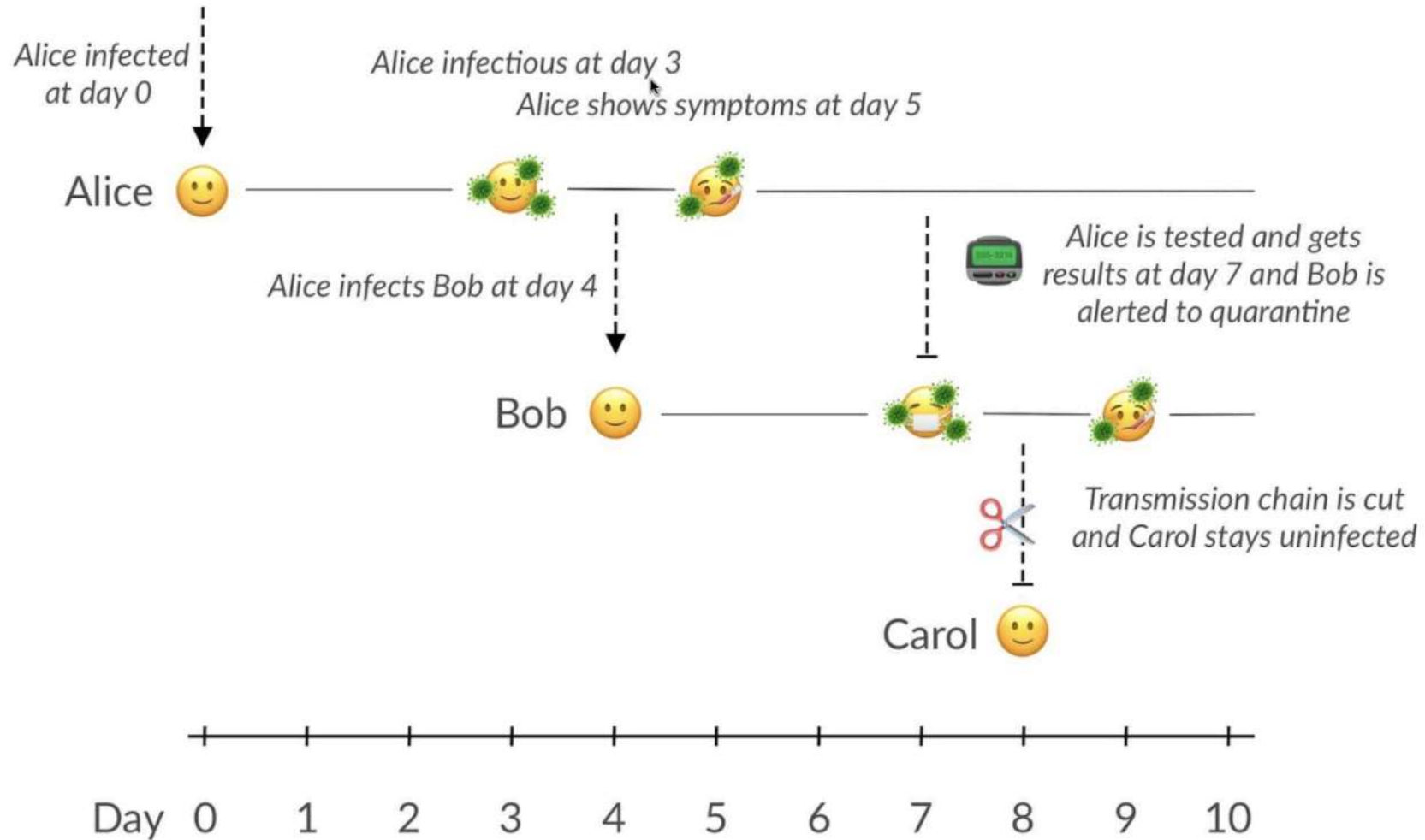
Category/Study Name	1° Data source(s)	Institution(s)	Country
Digital Surveillance			
DETECT Study and COVID-19 Wearables	Smartwatch, smart ring, EHR	Scripps, Stanford	US
Kinsa	Smart thermometer	NA	US
COVID-19 Citizen Science	Smartphone +	UCSF	US
Corona-Datenspende-App	Smartwatch, body temp	Robert Koch Institute	Germany
Symptom checkers	Multiple smartphone apps with daily inputs	NA	Israel, UK , Canada, Switzerland, Germany
Aggregated Mobility/ Geolocation	Cellphone	NA	Germany, China and other Asia, EU countries
Contact Tracing			
Google/Apple	Smartphone	NA	Global
PACT (Private Automatic Contact Tracing)	Smartphone	MIT, CSAIL	US
Trace Together	Smartphone	NA	Singapore

Apple, Google Bring Covid-19 Contact-Tracing to 3 Billion People

By [Mark Gurman](#)

April 10, 2020, 10:00 AM PDT Updated on April 10, 2020, 4:53 PM PDT

Test Trace Isolate



Resources

1) TIBCO COVID-19 Visual Analysis Hub
tibco.com/covid19

2) Community Headquarters
community.tibco.com/covid19

Community Tech Blogs
community.tibco.com/blog

3) Live Spotfire App
[Live Spotfire App](#)

4) Social
LinkedIn
linkedin.com/in/michaelo15/

Twitter : Daily update
[@MichOConnell](https://twitter.com/MichOConnell)

April 29 : Hands-on building session

If you need a Spotfire license we will provide

Go to TIBCO Community HQ to get

- Spotfire
- Spotfire starter app
- Data
- R and Python scripts

TIBCO Community



community.tibco.com/covid19



3



1.08k
views

TIBCO Community COVID-19 Visual Data Science Headquarters

[TIBCO Spotfire*](#) [TIBCO* Streaming](#) [TIBCO* Data Science - Workbench](#) [TIBCO* Data Science](#)
[TIBCO* Enterprise Runtime for R \(TERR™\)](#)

By: Mike Alperin

Last updated: 2:49pm Apr 17, 2020

7



Resources

1) TIBCO COVID-19 Visual Analysis Hub
tibco.com/covid19

2) Community Headquarters
community.tibco.com/covid19

Community Tech Blogs
community.tibco.com/blog

3) Live Spotfire App
[Live Spotfire App](#)

4) Social
LinkedIn
linkedin.com/in/michaelo15/

Twitter : Daily update
[@MichOConnell](https://twitter.com/MichOConnell)

April 29 : Hands-on building session

If you need a Spotfire license we will provide
Go to TIBCO Community HQ to get

- Spotfire
- Spotfire starter app
- Data
- R and Python scripts

TIBCO Community



community.tibco.com/covid19



3



1.08k
views

7



TIBCO Community COVID-19 Visual Data Science Headquarters

[TIBCO Spotfire*](#) [TIBCO* Streaming](#) [TIBCO* Data Science - Workbench](#) [TIBCO* Data Science](#)
[TIBCO* Enterprise Runtime for R \(TERR™\)](#)

By: Mike Alperin

Last updated: 2:49pm Apr 17, 2020