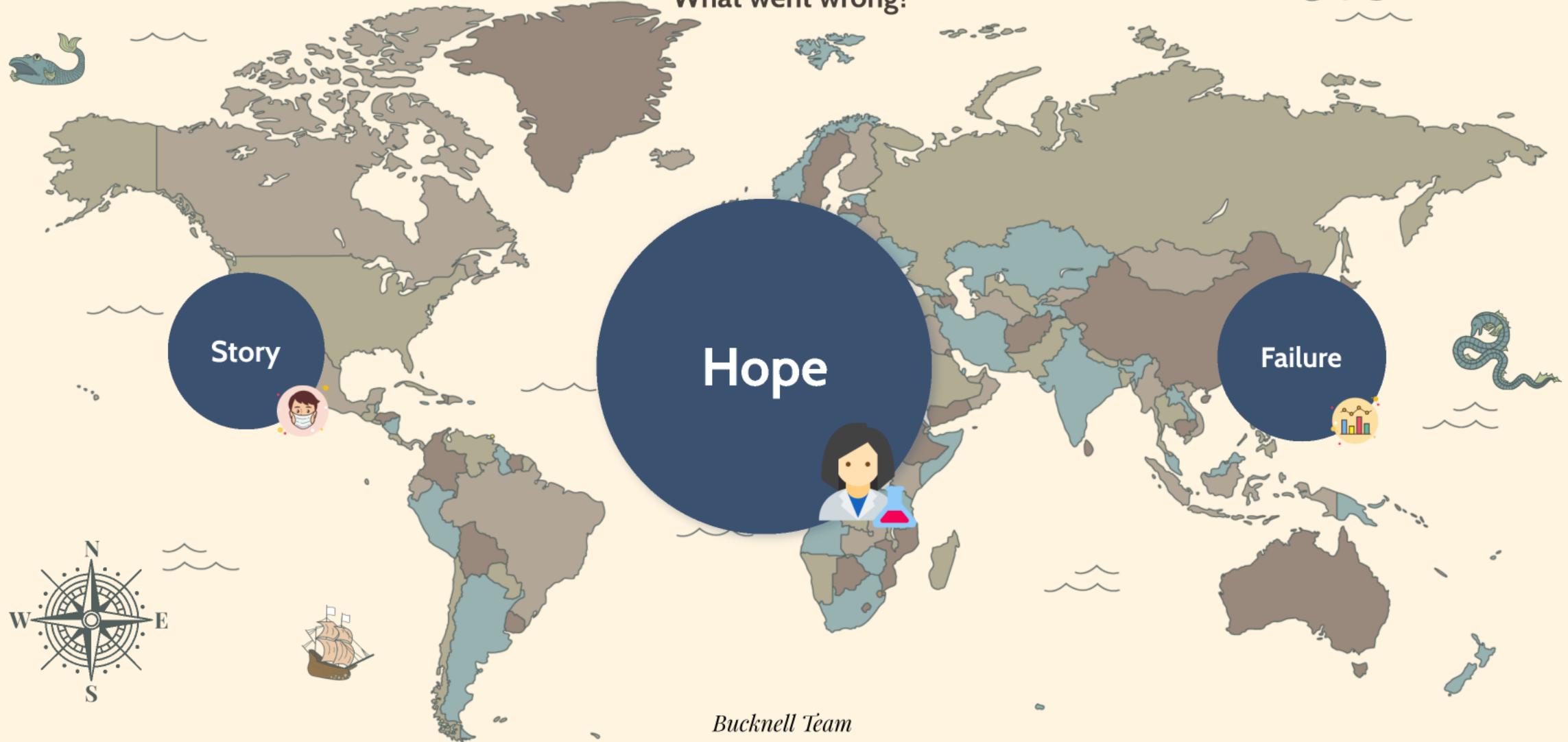




# Data, Society, and Influence

What went wrong?



# What went wrong?



# Where are we?

---



December 31 2019



March 11 2020

...



6.19 Million

Not sure? We have data!

Understanding  
Data Set

Missing  
Values

2(Independent)+1(Bloomberg ranking)



RANK	COUNTRY	BLOOMBERG ECONOMIC VALUE	INDUSTRIAL CAPACITY	INDUSTRIAL TRADE INDEX
1	China	74.2	244.0	-11.3%
2	U.S.	73.5	287.5	-27.7%
3	U.K.	73.5	188.0	-30.4%
4	Ireland	73.5	80	-36.7%
5	Canada	73.5	174.1	-43.4%
6	Australia	73.5	174.1	-43.4%
7	Turkey	73.5	181.2	-17.2%
8	Spain	73.5	377.0	-17.8%
9	Belgium	73.5	180.0	-34.2%
10	UK	68.2	188.0	-35.7%
11	Ireland	67.5	100.0	-27.2%
12	U.S.	67.5	244.0	-11.3%

53 countries (Selected based on economic values)  
Focus on social and economic disruption instead of health impact



225 countries and regions (with overlap)

But not all countries appear on the same date

c6\_stay\_at\_home\_requirements  
c2\_flag  
v1\_vaccine\_prioritisation  
v2a\_vaccine\_availability  
c7\_restrict\_on\_internal\_move  
governmentresponseindex  
c7\_flag  
containmenthealthindexfordisplay  
stringencyindexfordisplay  
c4\_restrictions\_on\_gatherings  
contract\_relief  
stringencyindex  
c5\_close\_public\_transport  
stringencyindex  
goveresponseindexdisplay  
e3\_fiscal\_measures  
h6\_facial\_coverings  
v3\_vaccine\_financial\_support  
c4\_flag  
c3\_cancel\_public\_events  
stringencylegacyindexfordisplay  
e1\_flag  
confirmeddeaths  
c8\_international\_travel\_controls  
c5\_flag  
e1\_income\_support  
e4\_international\_support  
h8\_flag  
h8\_protection\_of\_elderly\_people  
h5\_investment\_in\_vaccines  
containmenthealthindex  
v2g\_frontline\_workers\_healthcare  
c3\_flag  
h6\_flag  
h4\_emergency\_invest\_in\_hlthcre  
h1\_public\_information\_campaigns  
c6\_flag  
c1\_school\_closing  
h7\_vaccination\_policy  
e2\_debt  
h7\_flag  
h1\_flag  
economicsupportindexfordisplay  
c1\_flag  
v2f\_frontline\_workers\_nonhealth  
m1\_wildcard  
v2e\_education

186 countries and regions

All countries appear on 2020/01/01, but not all countries have data

RANK	CHANGE	ECONOMY	BLOOMBERG RESILIENCE SCORE	VACCINE DOSES PER 100	LOCKDOWN SEVERITY	FLIGHT CAPACITY	VACCINATED TRAVEL ROUTES
1	▲1	Chile	74.1	224.9	41	-17.3%	278
2	▲2	Ireland	72.9	181.2	50	-21.8%	397
3	▼2	U.A.E.	72.9	207.6	52	-27.7%	406
4	▼1	Finland	71.6	166.9	35	-39.4%	402
5	▲2	Canada	70.7	173.7	63	-36.7%	405.5
6	▲8	Colombia	70.4	124.1	47	0.3%	401
7	▼1	Turkey	70.2	150.2	55	-17.2%	397
8	▼3	Spain	70.1	177.5	44	-17.6%	399
9	▼1	Sweden	69.8	165.1	19	-35.4%	240
10	▲2	U.K.	68.3	190.9	49	-35.7%	402.5
11	▼1	Denmark	67.4	185	39	-27.2%	315.5
12	▲1	U.S.	67.1	150.7	48	-11.7%	401

53 countries (Selected based on economic values)

Focus on social and economic disruption instead of health impact

# Data reliability

56.81% of the rows in Covid Variables data set have missing value(s)



## Missing important indicators

total_deaths		
total_deaths	Frequency	Percent
•	20127	12.99
Non-missing	134784	87.01

hosp_patients_per_million		
hosp_patients_per_million	Frequency	Percent
•	133335	86.07
Non-missing	21576	13.93

new_tests		
new_tests	Frequency	Percent
•	93804	60.55
Non-missing	61107	39.45

people_fully_vaccinated		
people_fully_vaccinated	Frequency	Percent
•	119408	77.08
Non-missing	35503	22.92

new_deaths		
new_deaths	Frequency	Percent
•	19960	12.88
Non-missing	134951	87.12

weekly_icu_admissions		
weekly_icu_admissions	Frequency	Percent
•	150611	97.22
Non-missing	4300	2.78

total_tests		
total_tests	Frequency	Percent
•	92165	59.50
Non-missing	62746	40.50

total_boosters		
total_boosters	Frequency	Percent
•	141803	91.54
Non-missing	13108	8.46

Non-missing	27	0.0174
Non-missing	40	0.0258
Non-missing	624	0.4028
Non-missing	17	0.0110
Non-missing	669	0.4319

total_deaths		
total_deaths	Frequency	Percent
.	20127	12.99
Non-missing	134784	87.01

new_deaths		
new_deaths	Frequency	Percent
.	19960	12.88
Non-missing	134951	87.12

hosp_patie

weekly_i

## **hosp\_patients\_per\_million**

<b>hosp_patients_per_million</b>	<b>Frequency</b>	<b>Percent</b>
	133335	86.07
<b>Non-missing</b>	21576	13.93

## **weekly\_icu\_admissions**

<b>weekly_icu_admissions</b>	<b>Frequency</b>	<b>Percent</b>
	150611	97.22
<b>Non-missing</b>	4300	2.78

ercent

86.07

13.93

rcent

97.22

2.78

### new\_tests

new_tests	Frequency	Percent
	93804	60.55
<b>Non-missing</b>	61107	39.45

### total\_tests

total_tests	Frequency	Percent
	92165	59.50
<b>Non-missing</b>	62746	40.50

people

to

## people\_fully\_vaccinated

people_fully_vaccinated	Frequency	Percent
	119408	77.08
Non-missing	35503	22.92

## total\_boosters

total_boosters	Frequency	Percent
	141803	91.54
Non-missing	13108	8.46

# Data reliability

56.81% of the rows in Covid Variables data set have missing value(s)



## Missing important indicators

total_deaths		
total_deaths	Frequency	Percent
•	20127	12.99
Non-missing	134784	87.01

hosp_patients_per_million		
hosp_patients_per_million	Frequency	Percent
•	133335	86.07
Non-missing	21576	13.93

new_tests		
new_tests	Frequency	Percent
•	93804	60.55
Non-missing	61107	39.45

people_fully_vaccinated		
people_fully_vaccinated	Frequency	Percent
•	119408	77.08
Non-missing	35503	22.92

new_deaths		
new_deaths	Frequency	Percent
•	19960	12.88
Non-missing	134951	87.12

weekly_icu_admissions		
weekly_icu_admissions	Frequency	Percent
•	150611	97.22
Non-missing	4300	2.78

total_tests		
total_tests	Frequency	Percent
•	92165	59.50
Non-missing	62746	40.50

total_boosters		
total_boosters	Frequency	Percent
•	141803	91.54
Non-missing	13108	8.46

# Data reliability

56.81% of the rows in Covid Variables data set have missing value(s)

new_deaths	Frequency	Percent
NaN	20127	12.99
Non-missing	134784	87.01

hosp_patients_per_million	Frequency	Percent
NaN	133335	86.07
Non-missing	21576	13.93

new_tests	Frequency	Percent
NaN	93804	60.55
Non-missing	61107	39.45

people_fully_vaccinated	Frequency	Percent
NaN	119408	77.08
Non-missing	35503	22.92



## Is our data set Functional?

total_deaths		
total_deaths	Frequency	Percent
NaN	20127	12.99
Non-missing	134784	87.01

hosp_patients_per_million		
hosp_patients_per_million	Frequency	Percent
NaN	133335	86.07
Non-missing	21576	13.93

new_tests		
new_tests	Frequency	Percent
NaN	93804	60.55
Non-missing	61107	39.45

people_fully_vaccinated		
people_fully_vaccinated	Frequency	Percent
NaN	119408	77.08
Non-missing	35503	22.92

new_deaths		
new_deaths	Frequency	Percent
NaN	19960	12.88
Non-missing	134951	87.12

weekly_icu_admissions		
weekly_icu_admissions	Frequency	Percent
NaN	150611	97.22
Non-missing	4300	2.78

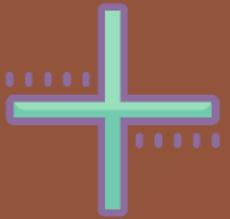
total_tests		
total_tests	Frequency	Percent
NaN	92165	59.50
Non-missing	62746	40.50

total_boosters		
total_boosters	Frequency	Percent
NaN	141803	91.54
Non-missing	13108	8.46

# Decision on missing values



80%



20%



We are removing **small economies**

# Decision on missing values

## Why rank only 53 places?

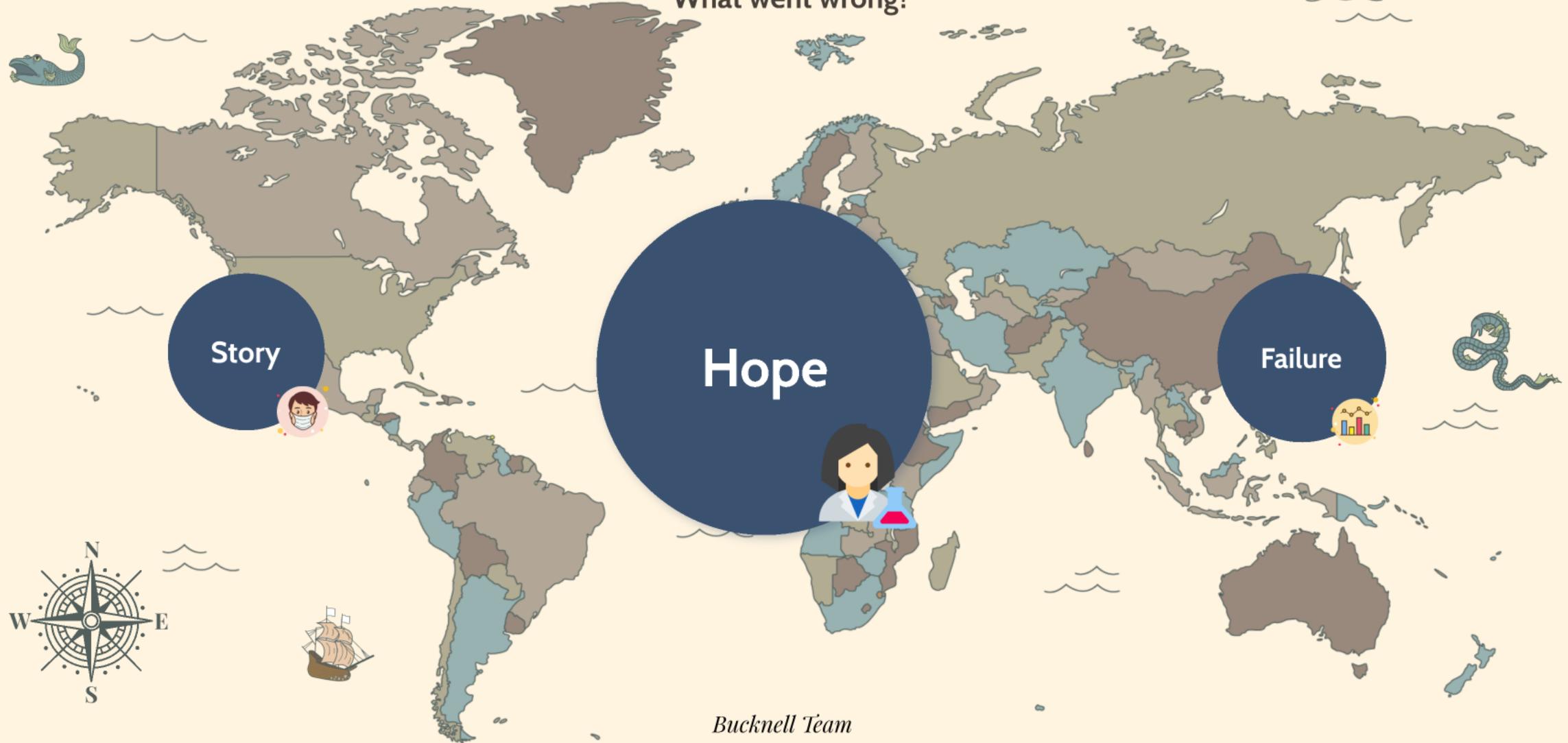
We decided for brevity and relevance to limit the Ranking to economies valued at more than \$200 billion prior to the pandemic.

We are removing **small economies**



# Data, Society, and Influence

What went wrong?



# Go back to our goals

---

Health crisis

Actionable insights

Critical decision making

...

Does our data set **provide** insights?

A faint background map of the world is visible, showing landmasses in various shades of brown, tan, and light blue.

Model  
Input

Relationship

# **Statistical significance vs Practical significance**

# T-test by Nearest Neighbors

Null hypothesis: There is **no difference** in new death per million, between countries with high vs low policy severity

Alternative hypothesis: There is **a difference** in new death per million...

P value and statistical significance:

The two-tailed P value is less than 0.0001

Reject null hypothesis, there is sufficient evidence to support the alternative hypothesis at 0.05 significance level

## Test results

More school closure correlates to **higher** death

Group	Kuwait (High)	Tajikistan (Low)
Mean	0.87688	0.02062
SD	0.85507	0.07079
SEM	0.03349	0.00284
N	652	623

More public event cancellation correlates to **higher** death

Group	Myanmar (High)	Taiwan (Low)
Mean	0.53680	0.05344
SD	1.19856	0.18918
SEM	0.04680	0.00733
N	656	667

More gathering restrictions correlates to **higher** death

Group	Kazakhstan (High)	Taiwan (Low)
Mean	1.45916	0.05344
SD	7.08252	0.18918
SEM	0.27569	0.00733
N	660	667

More Stay at Home policies correlates to **higher** death

Group	Kenya (High)	Tanzania (Low)
Mean	0.15100	0.01834
SD	0.15270	0.42511
SEM	0.00594	0.01660
N	661	656

# More school closure correlates to higher death

Group	Kuwait (High)	Tajikistan (Low)
Mean	0.87688	0.02062
SD	0.85507	0.07079
SEM	0.03349	0.00284
N	652	623

## More public event cancellation correlates to higher death

Group	Myanmar (High)	Taiwan (Low)
Mean	0.53680	0.05344
SD	1.19856	0.18918
SEM	0.04680	0.00733
N	656	667

# More gathering restrictions correlates to higher death

Group	Kazakhstan (High)	Taiwan (Low)
Mean	1.45916	0.05344
SD	7.08252	0.18918
SEM	0.27569	0.00733
N	660	667

# More Stay at Home policies correlates to higher death

Group	Kenya (High)	Tanzania (Low)
Mean	0.15100	0.01834
SD	0.15270	0.42511
SEM	0.00594	0.01660
N	661	656

# T-test by Nearest Neighbors

Null hypothesis: There is **no difference** in new death per million, between countries with high vs low policy severity

Alternative hypothesis: There is **a difference** in new death per million...

P value and statistical significance:

The two-tailed P value is less than 0.0001

Reject null hypothesis, there is sufficient evidence to support the alternative hypothesis at 0.05 significance level

## Test results

More school closure correlates to **higher** death

Group	Kuwait (High)	Tajikistan (Low)
Mean	0.87688	0.02062
SD	0.85507	0.07079
SEM	0.03349	0.00284
N	652	623

More public event cancellation correlates to **higher** death

Group	Myanmar (High)	Taiwan (Low)
Mean	0.53680	0.05344
SD	1.19856	0.18918
SEM	0.04680	0.00733
N	656	667

More gathering restrictions correlates to **higher** death

Group	Kazakhstan (High)	Taiwan (Low)
Mean	1.45916	0.05344
SD	7.08252	0.18918
SEM	0.27569	0.00733
N	660	667

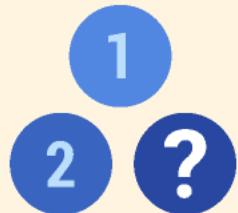
More Stay at Home policies correlates to **higher** death

Group	Kenya (High)	Tanzania (Low)
Mean	0.15100	0.01834
SD	0.15270	0.42511
SEM	0.00594	0.01660
N	661	656

# T-test by Nearest Neighbors

Null hypothesis: There is **no difference** in new death per million, between countries with high vs low policy severity

Alternative hypothesis: There is **a difference** in new death per million...



## Why? Confounding

More school closure correlates to **higher** death

Group	Kuwait (High)	Tajikistan (Low)
Mean	0.87688	0.02062
SD	0.85507	0.07079
SEM	0.03349	0.00284
N	652	623

More public event cancellation correlates to **higher** death

Group	Myanmar (High)	Taiwan (Low)
Mean	0.53680	0.05344
SD	1.19856	0.18918
SEM	0.04680	0.00733
N	656	667

More gathering restrictions correlates to **higher** death

Group	Kazakhstan (High)	Taiwan (Low)
Mean	1.45916	0.05344
SD	7.08252	0.18918
SEM	0.27569	0.00733
N	660	667

More Stay at Home policies correlates to **higher** death

Group	Kenya (High)	Tanzania (Low)
Mean	0.15100	0.01834
SD	0.15270	0.42511
SEM	0.00594	0.01660
N	661	656

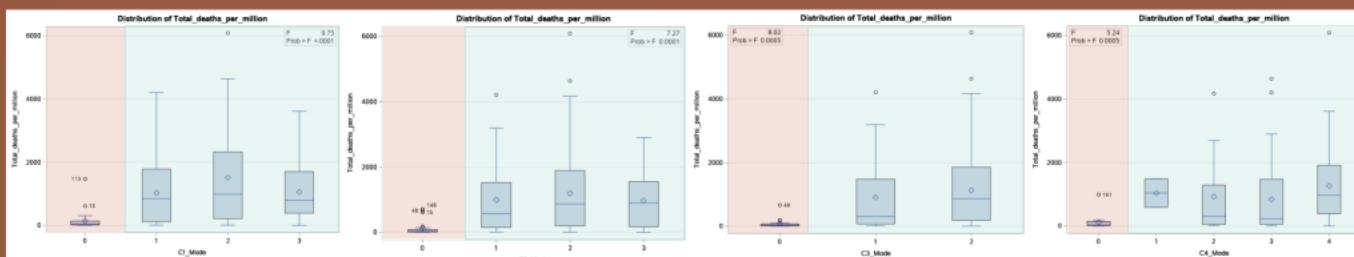
# ANOVA (Total death per million)

Null hypothesis: There is **no difference** in the mean of total death per million, between countries with different policy severity

Alternative hypothesis: There is **at least one** mean of total death per million that is **different**

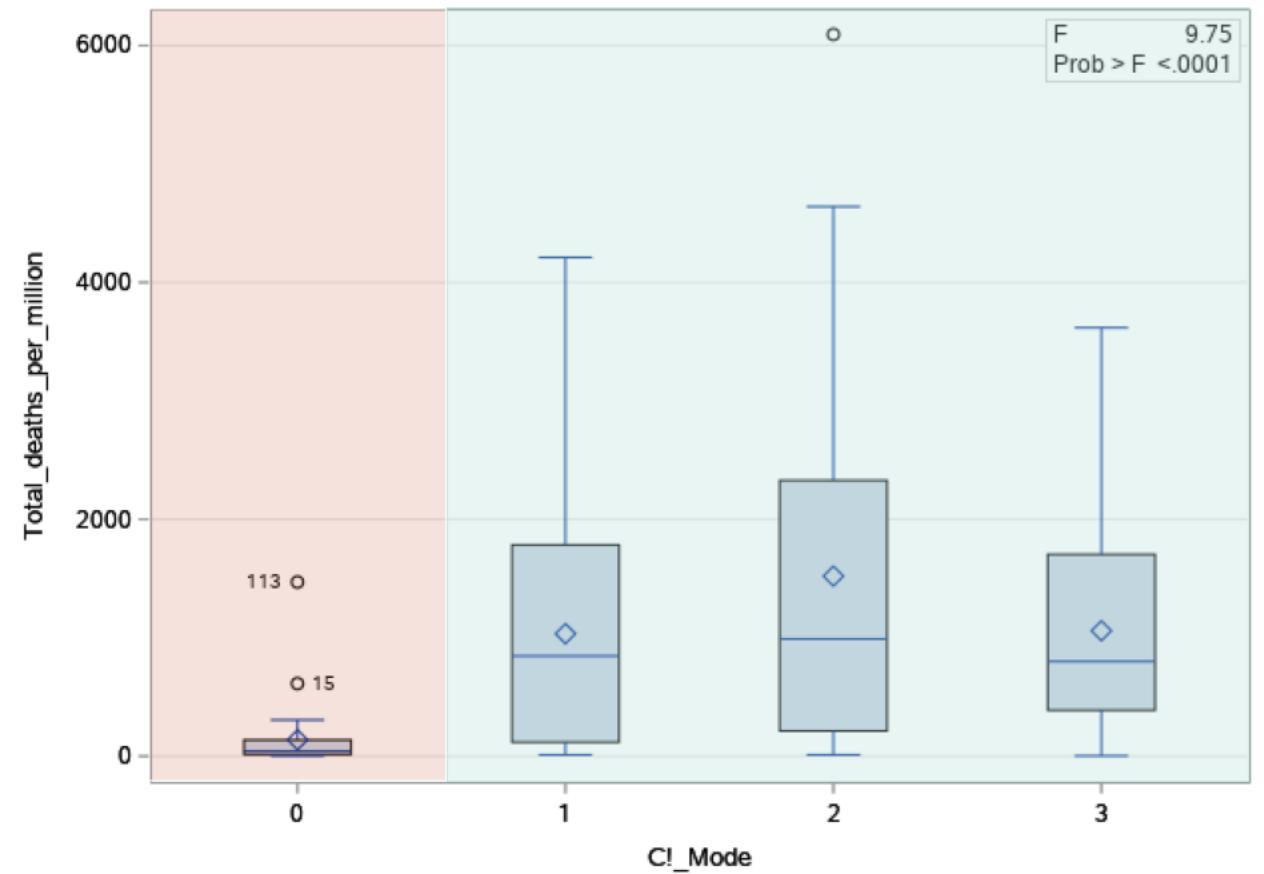
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	29523319.9	9841106.6	9.75	<.0001
Error	170	171525091.5	1008971.1		
Corrected Total	173	201048411.4			

## Test results

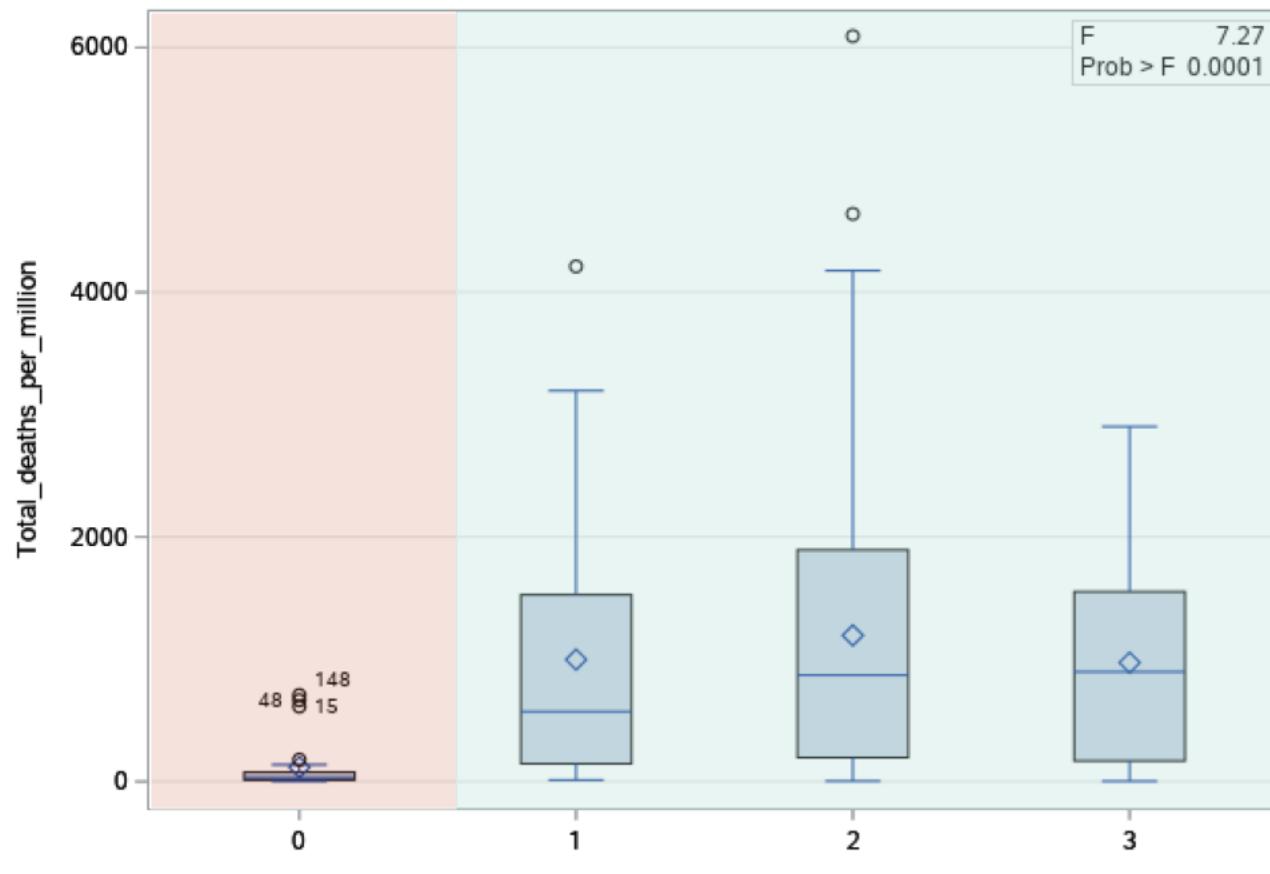


Based on test results, closures and lock downs are **ineffective** policies.

Distribution of Total\_deaths\_per\_million

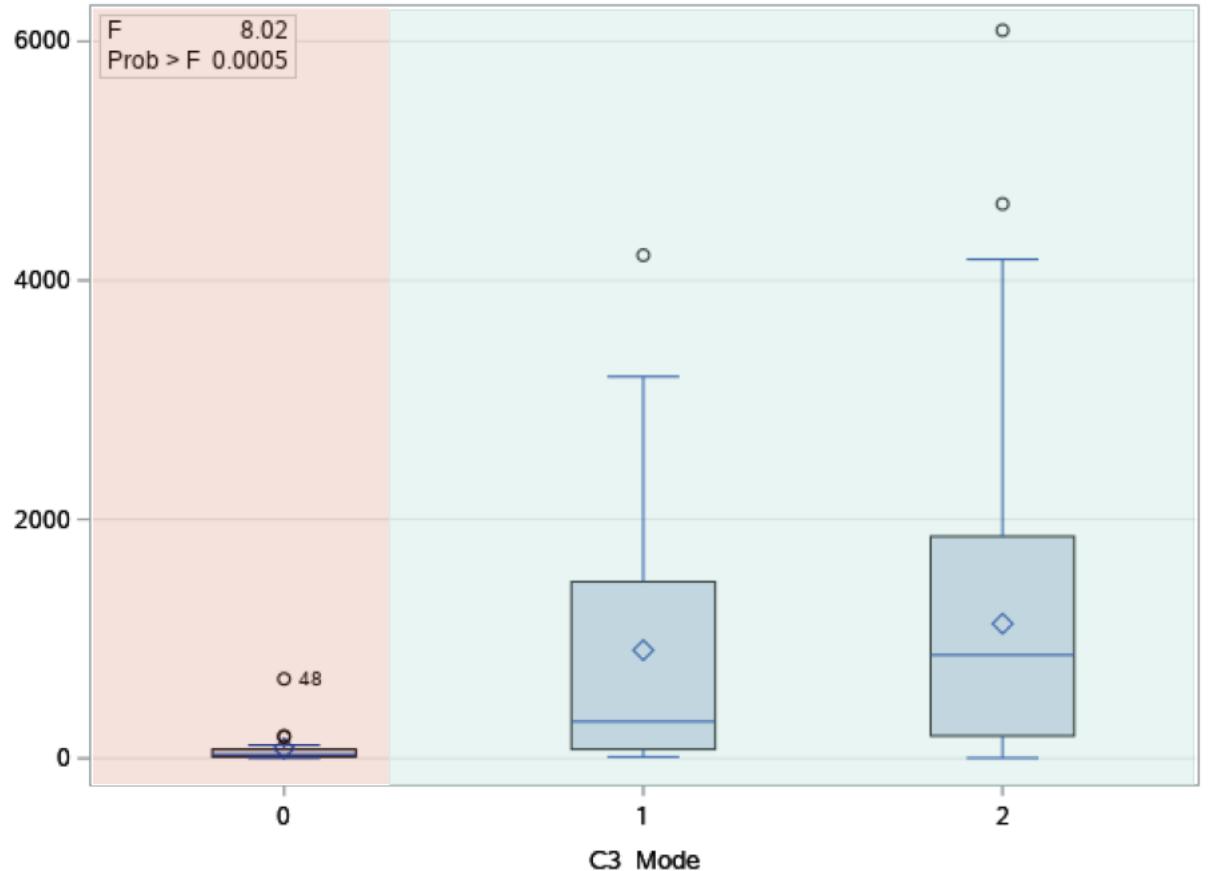


Distribution of Total\_deaths\_per\_million

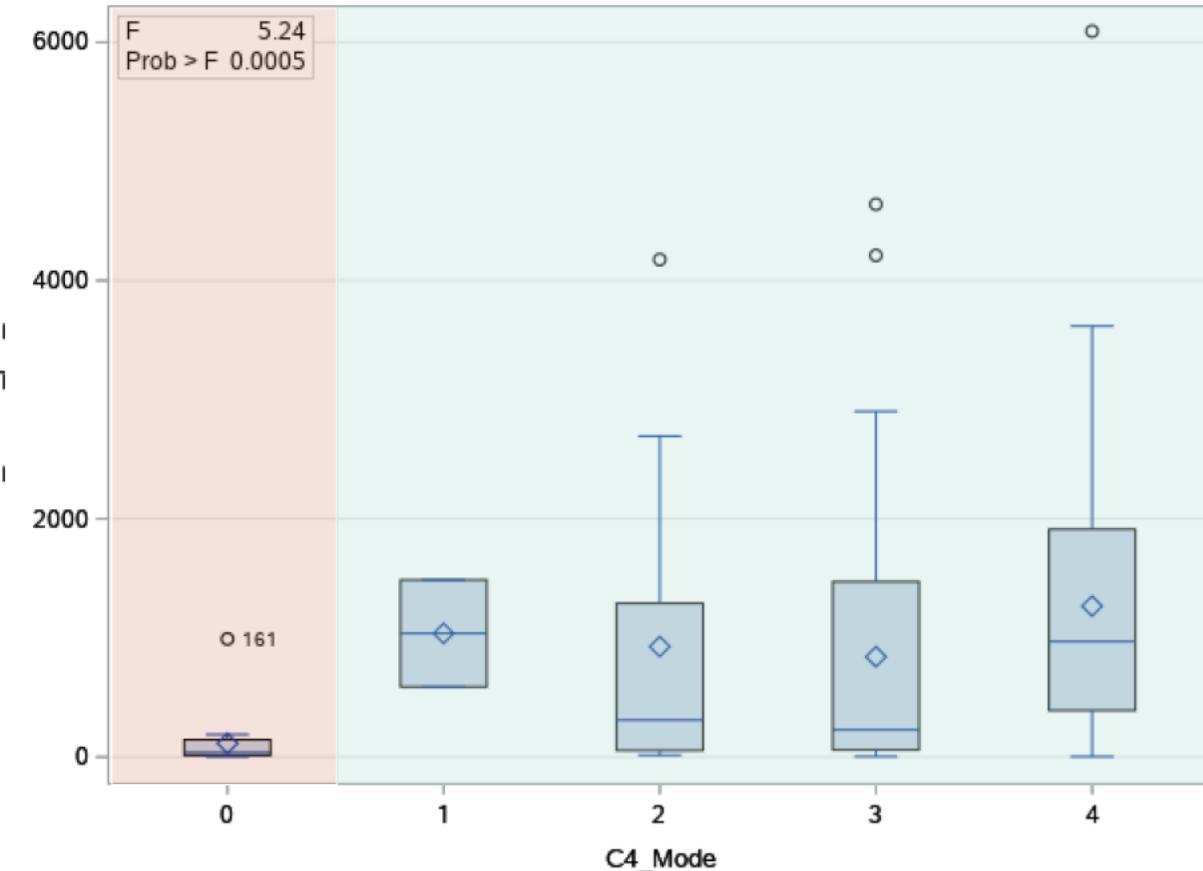


Based on test results, closures and lock downs are **ineffective** policies.

Distribution of Total\_deaths\_per\_million



Distribution of Total\_deaths\_per\_million



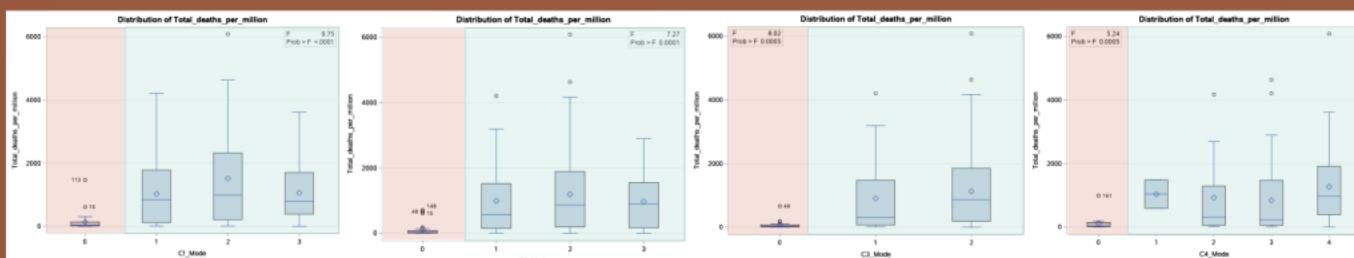
# ANOVA (Total death per million)

Null hypothesis: There is **no difference** in the mean of total death per million, between countries with different policy severity

Alternative hypothesis: There is **at least one** mean of total death per million that is **different**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	29523319.9	9841106.6	9.75	<.0001
Error	170	171525091.5	1008971.1		
Corrected Total	173	201048411.4			

## Test results



Based on test results, closures and lock downs are **ineffective** policies.

# ANOVA (Total death per million)

Null hypothesis: There is **no difference** in the mean of total death per million, between countries with different policy severity

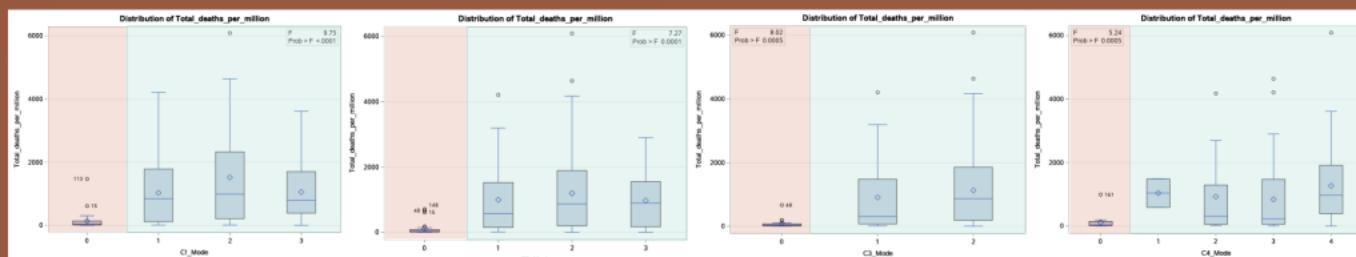
Alternative hypothesis: There is **at least one** mean of total death per million that is **different**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
--------	----	----------------	-------------	---------	--------



## Unrelated Data

### Test results



Based on test results, closures and lock downs are **ineffective** policies.

# Finding an Indicator: New case (per million)

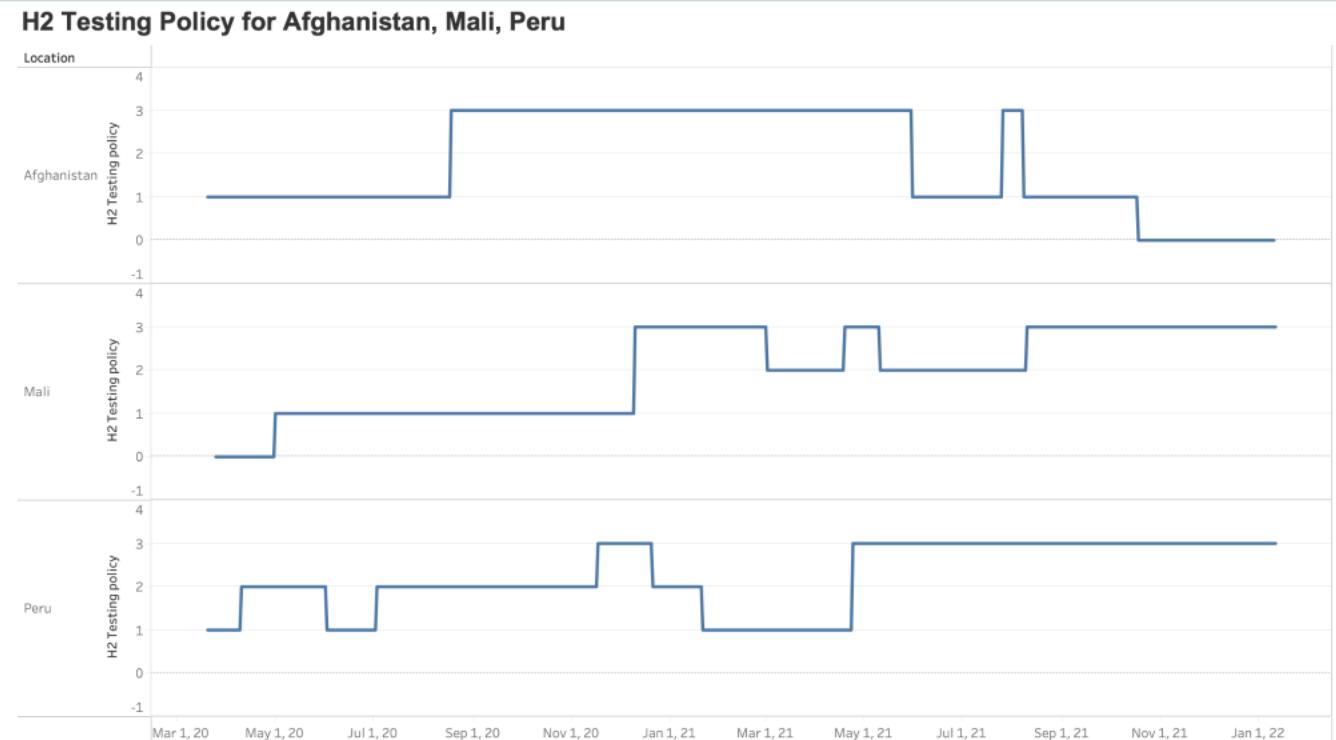
## Testing Policy

0 - no testing policy

1 - only those who both (a) have symptoms AND  
(b) meet specific criteria (eg key workers, admitted  
to hospital, came into contact with a known case,  
returned from overseas)

2 - testing of anyone showing Covid-19 symptoms

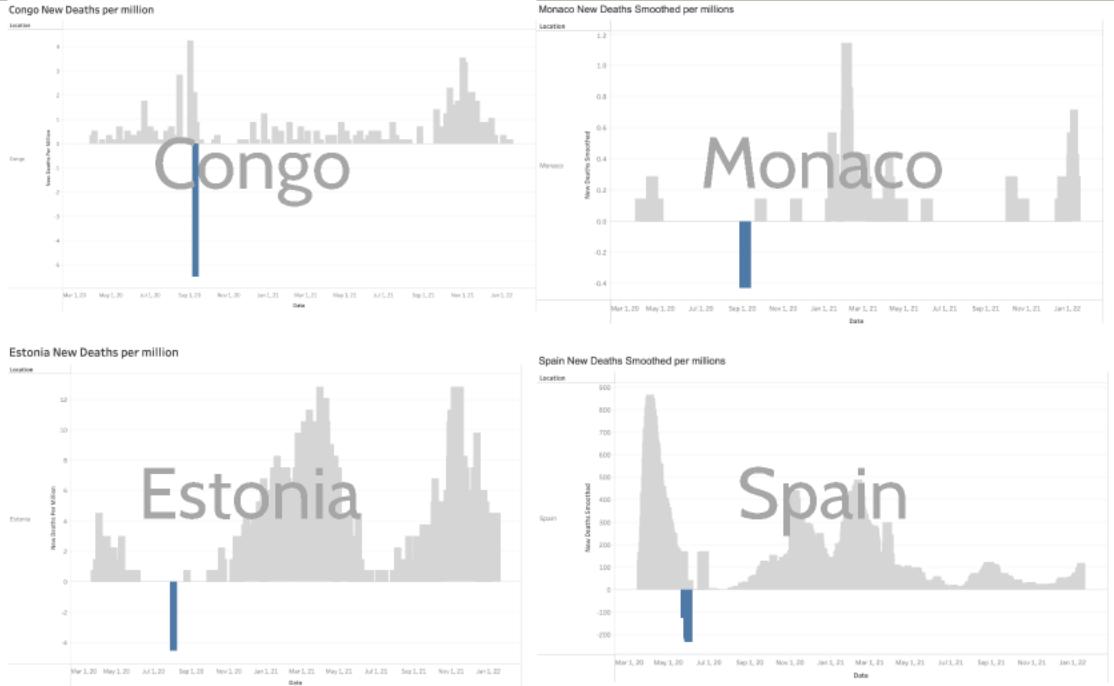
3 - open public testing (eg "drive through" testing  
available to asymptomatic people)



The trend of sum of H2 Testing policy for Date broken down by Location. Color shows details about Location. The view is filtered on Location and Date. The Location filter keeps Afghanistan, Mali and Peru. The Date filter ranges from 3/20/2020 to 1/15/2022.

"New cases" is an **inconsistent** indicator

# Finding an Indicator: New death (per million) (smoothed)

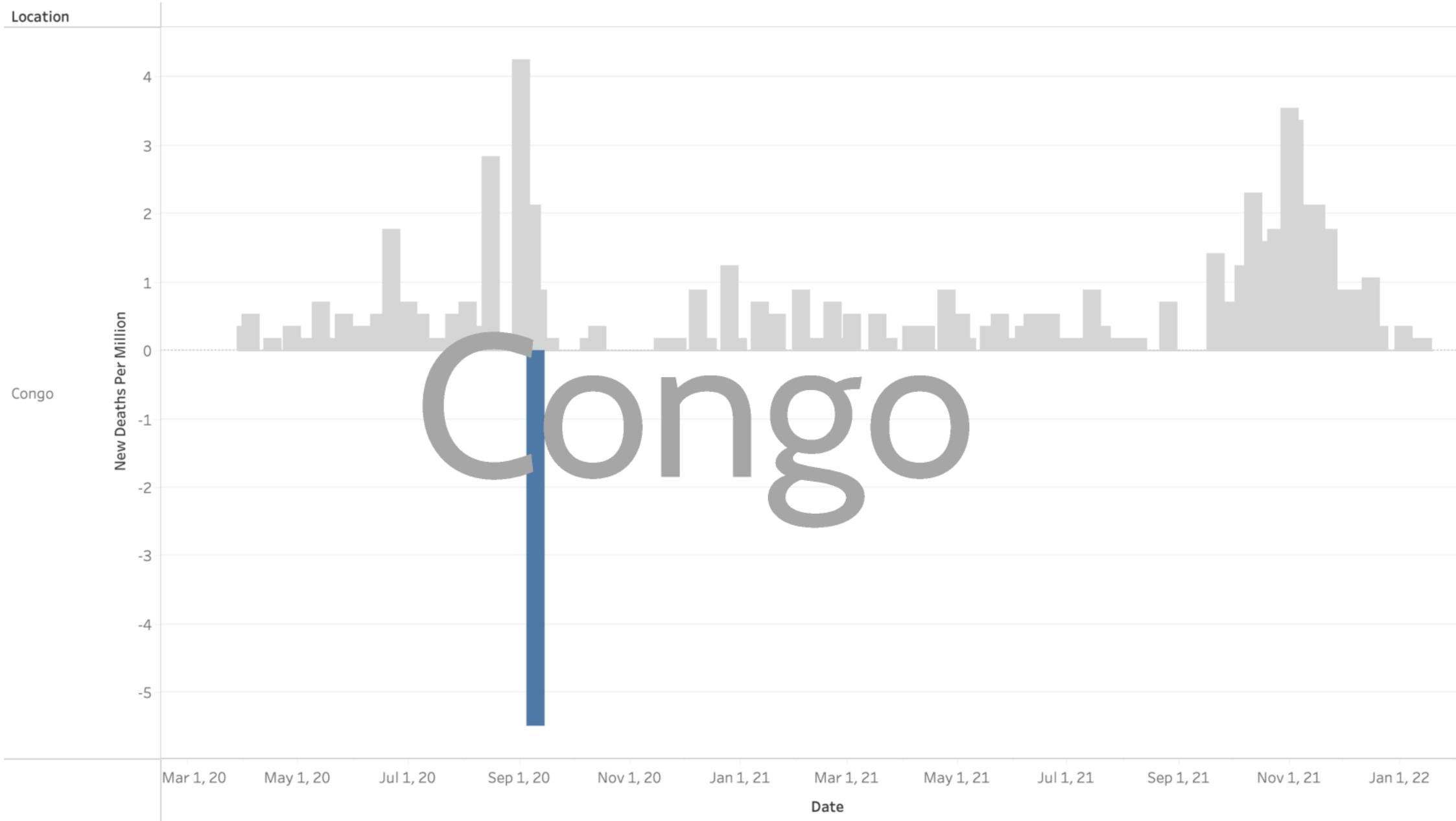


location	date	new_deaths_per_million	new_deaths_smoothed_per_million
Monaco	9/2/2020	-75.911	-10.844
Kyrgyzstan	8/21/2020	-66.834	-9.397
Anguilla	11/18/2021	-66.116	0
Spain	5/25/2020	-41.031	-2.674
Bonaire Sint Eustatius and Saba	10/11/2021	-37.814	0
Bonaire Sint Eustatius and Saba	10/18/2021	-37.814	0
Sweden	4/4/2020	-22.834	1.884
Peru	7/10/2021	-20.354	2.325
Sao Tome and Principe	6/3/2021	-17.908	0
Kazakhstan	8/31/2021	-14.793	0
Cuba	1/11/2022	-14.579	0.076
Chile	10/8/2021	-11.087	0.491
Belgium	8/26/2020	-10.058	-1.105
Switzerland	2/8/2021	-9.982	1.164
Aruba	10/14/2020	-9.329	1.333
Belize	12/8/2021	-7.409	2.47
Congo	9/10/2020	-5.48	-0.48
Estonia	8/2/2020	-4.528	-0.647

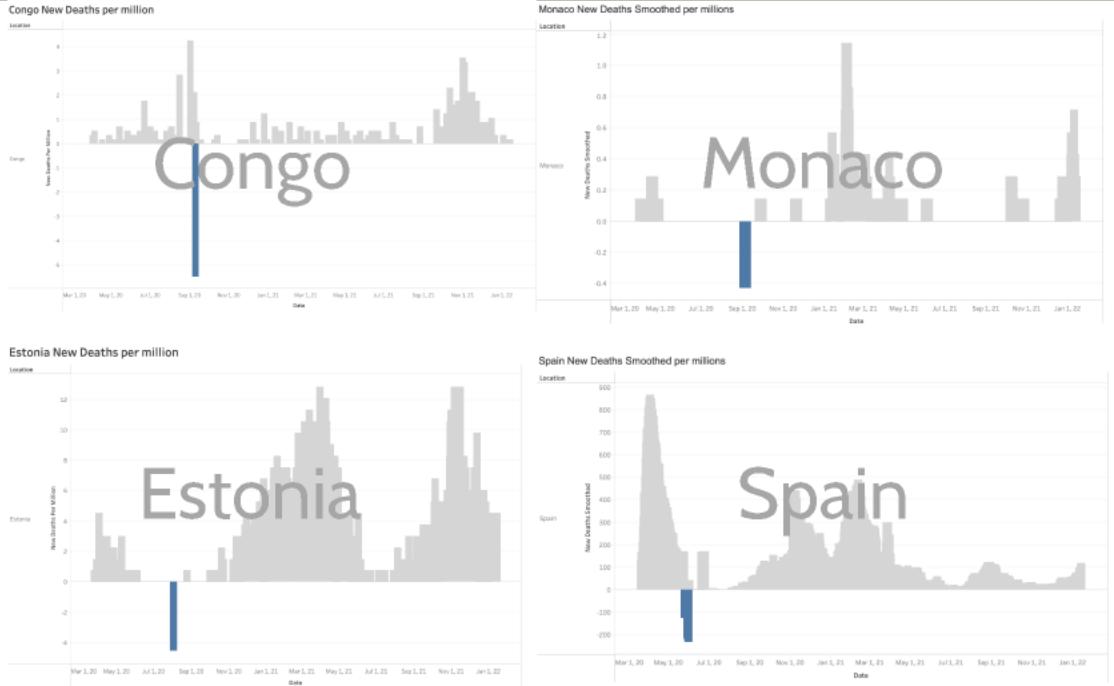
Hypothesis: Abnormal peaks represent a **correction** from previous days

Analytical judgement: 60 out of 183 countries had made +/- corrections

# Congo New Deaths per million



# Finding an Indicator: New death (per million) (smoothed)



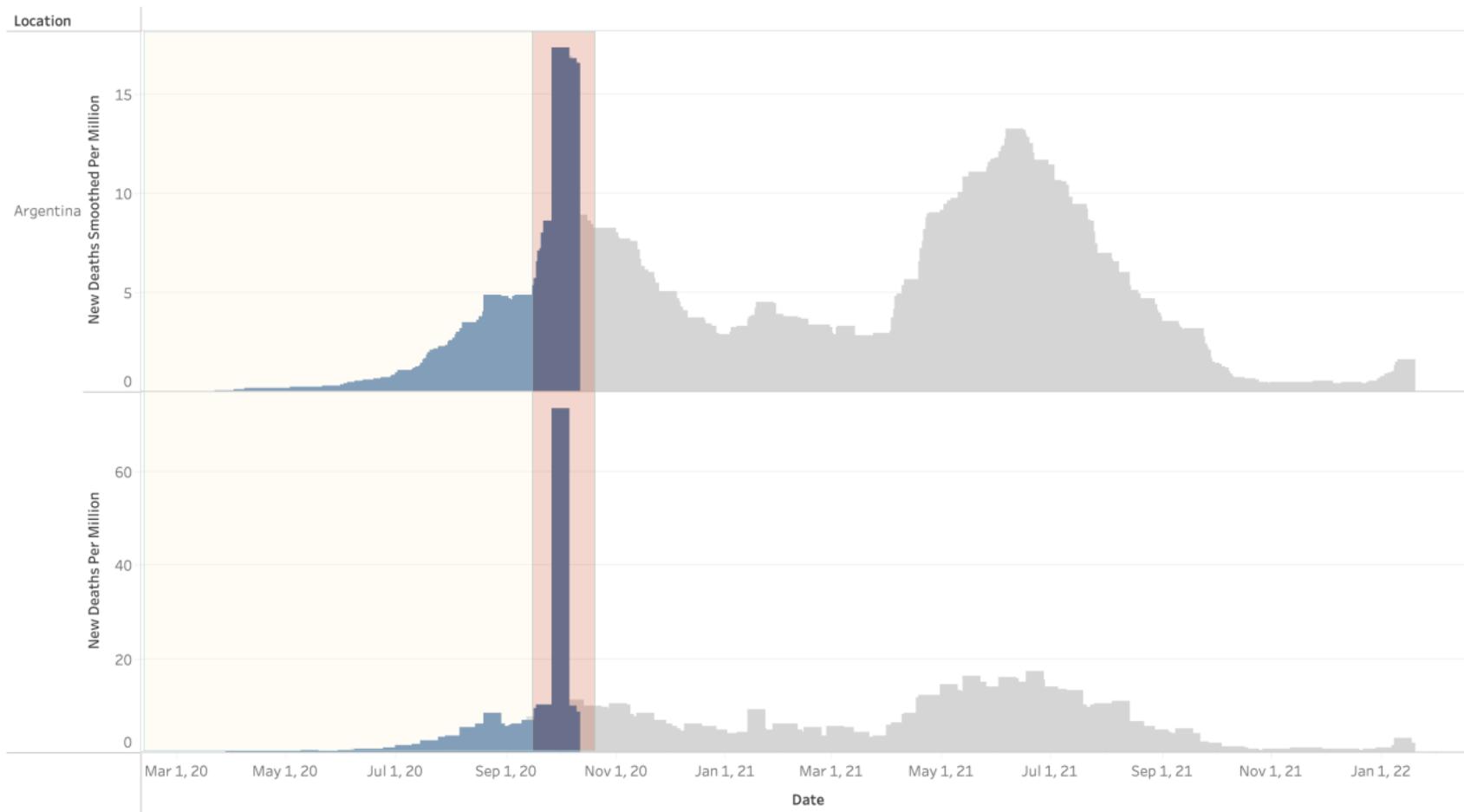
location	date	new_deaths_per_million	new_deaths_smoothed_per_million
Monaco	9/2/2020	-75.911	-10.844
Kyrgyzstan	8/21/2020	-66.834	-9.397
Anguilla	11/18/2021	-66.116	0
Spain	5/25/2020	-41.031	-2.674
Bonaire Sint Eustatius and Saba	10/11/2021	-37.814	0
Bonaire Sint Eustatius and Saba	10/18/2021	-37.814	0
Sweden	4/4/2020	-22.834	1.884
Peru	7/10/2021	-20.354	2.325
Sao Tome and Principe	6/3/2021	-17.908	0
Kazakhstan	8/31/2021	-14.793	0
Cuba	1/11/2022	-14.579	0.076
Chile	10/8/2021	-11.087	0.491
Belgium	8/26/2020	-10.058	-1.105
Switzerland	2/8/2021	-9.982	1.164
Aruba	10/14/2020	-9.329	1.333
Belize	12/8/2021	-7.409	2.47
Congo	9/10/2020	-5.48	-0.48
Estonia	8/2/2020	-4.528	-0.647

Hypothesis: Abnormal peaks represent a **correction** from previous days

Analytical judgement: 60 out of 183 countries had made +/- corrections

# Finding an Indicator: New death (per million) (smoothed)

Argentina New Deaths per million and New Deaths Smoothed per million

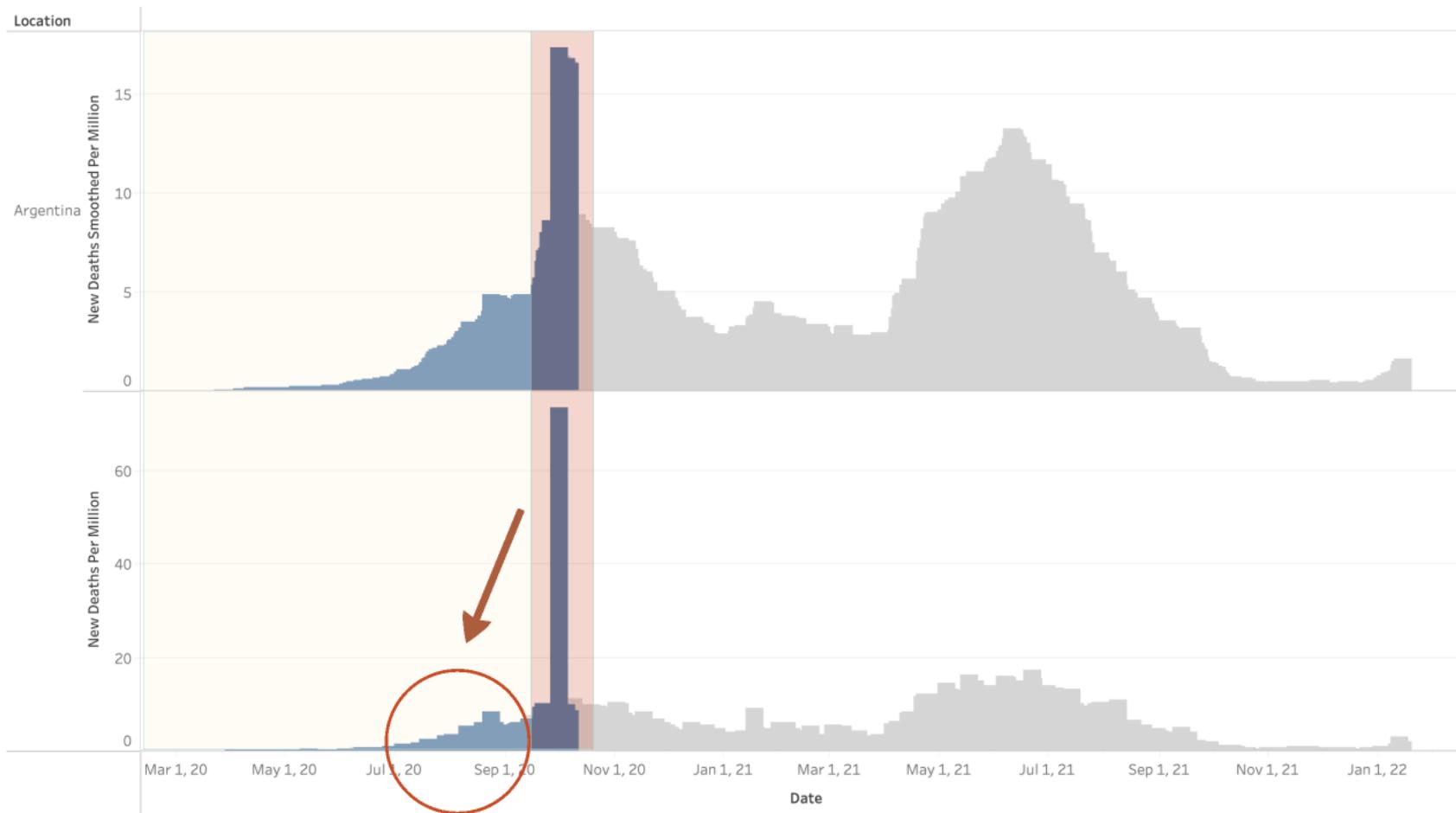


The plots of sum of New Deaths Smoothed Per Million and sum of New Deaths Per Million for Date broken down by Location. Color shows details about Date & Location (group). The view is filtered on Date and Location. The Date filter ranges from 3/20/2020 to 1/15/2022. The Location filter keeps Argentina.

The correction misleads analysis prior to the correct date

# Finding an Indicator: New death (per million) (smoothed)

Argentina New Deaths per million and New Deaths Smoothed per million

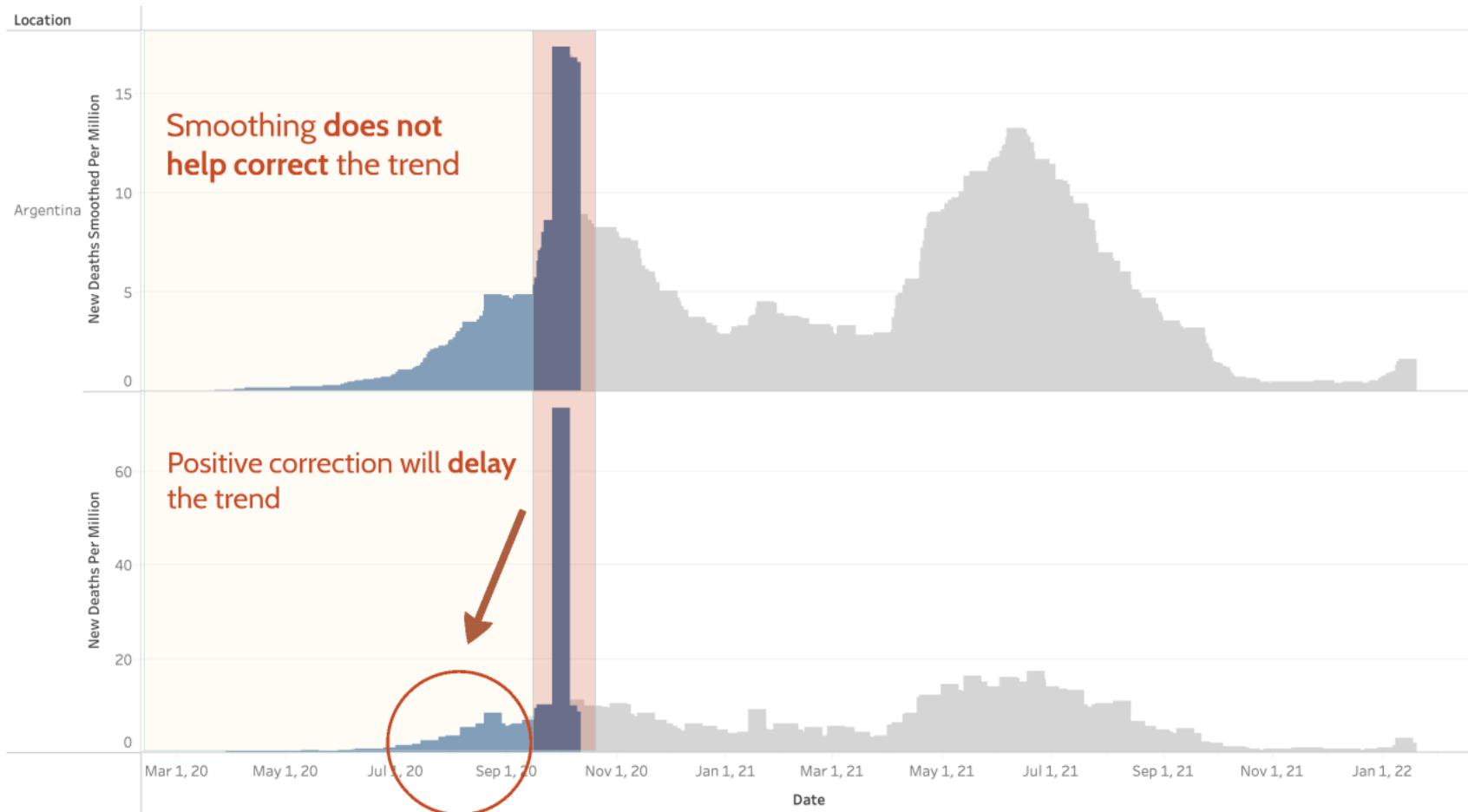


The plots of sum of New Deaths Smoothed Per Million and sum of New Deaths Per Million for Date broken down by Location. Color shows details about Date & Location (group). The view is filtered on Date and Location. The Date filter ranges from 3/20/2020 to 1/15/2022. The Location filter keeps Argentina.

The correction misleads analysis prior to the correct date

# Finding an Indicator: New death (per million) (smoothed)

Argentina New Deaths per million and New Deaths Smoothed per million



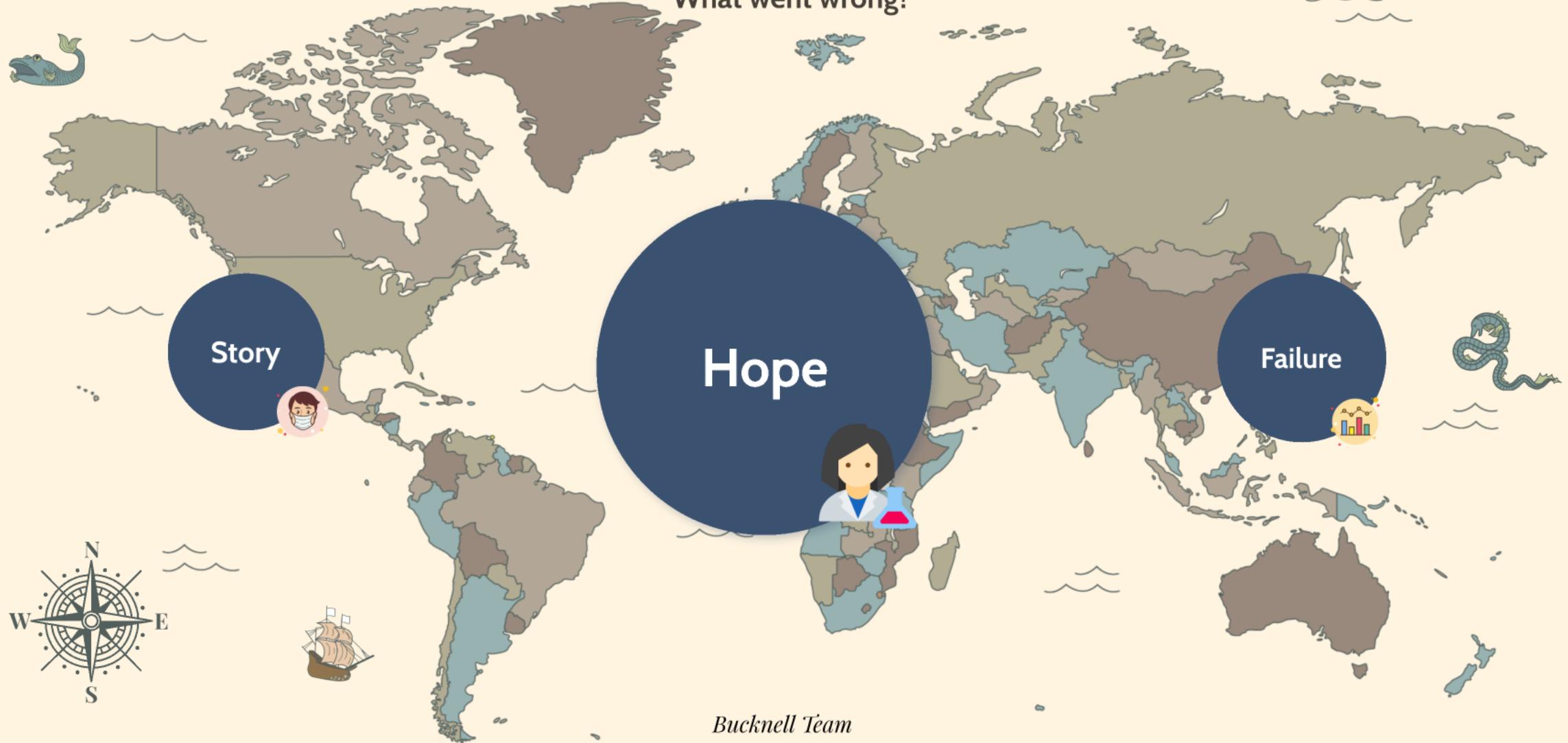
The plots of sum of New Deaths Smoothed Per Million and sum of New Deaths Per Million for Date broken down by Location. Color shows details about Date & Location (group). The view is filtered on Date and Location. The Date filter ranges from 3/20/2020 to 1/15/2022. The Location filter keeps Argentina.

The correction misleads analysis prior to the correct date



# Data, Society, and Influence

What went wrong?



# Recommendations

## Our Data

- Missing values
- Confounding
- Inconsistent data record
- Overgeneralized indicators



## Clinical Trials

- Experimental design
- Avoid Con-founder

## Model Assumption

- Complete data
- Causal analysis
- Continuity and accuracy
- Standardized measurement



## Data Collection Body

- Observational studies
- Minimize Con-founder's effects

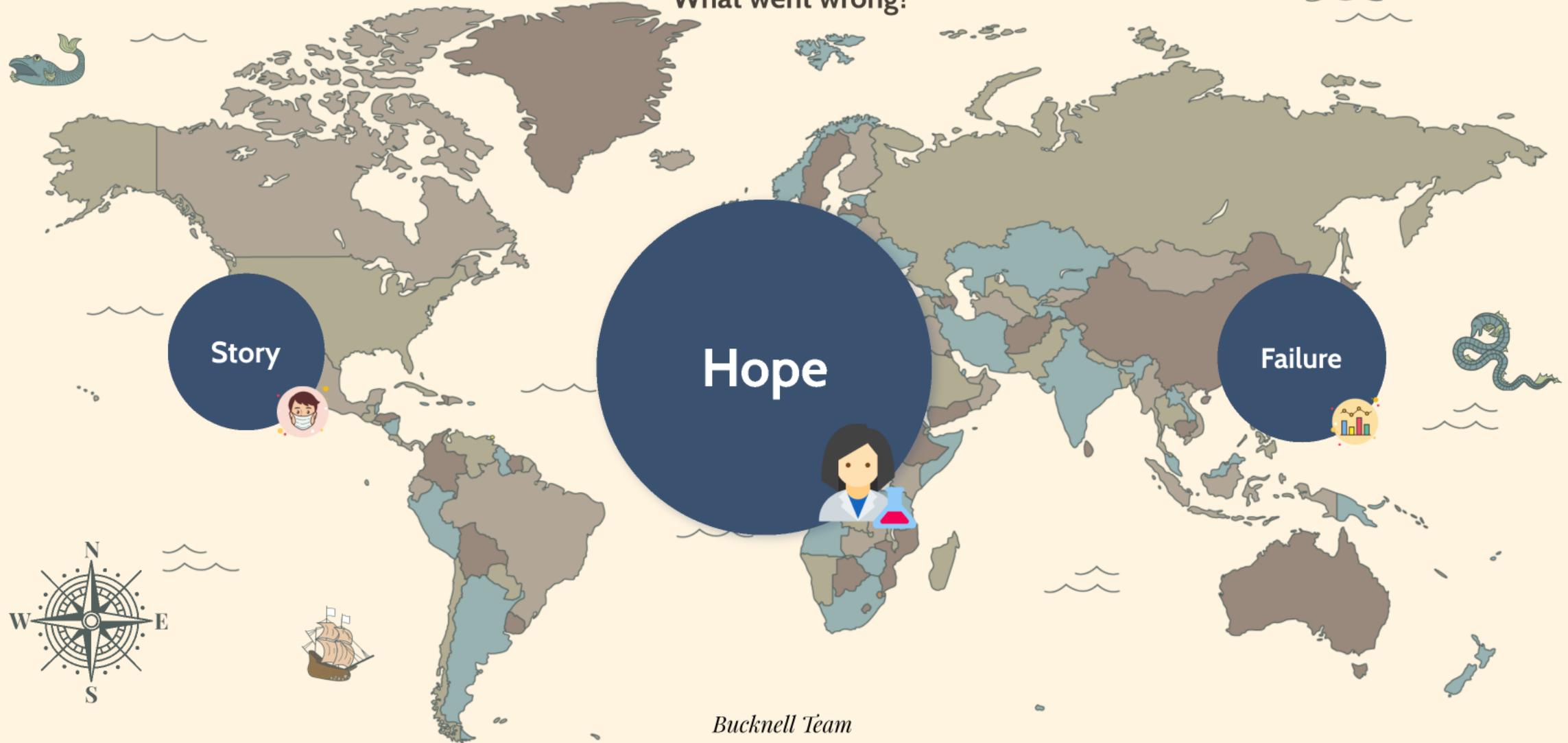
*As the future generation of data  
scientists, we prioritize the usefulness  
and validity of our analysis.*

- The Bucknell Team -



# Data, Society, and Influence

What went wrong?





# Data, Society, and Influence

What went wrong?

