The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences

Alessandro Romano, Chiara Sotis*, Goran Dominioni and Sebastián Guidi

Data Visualization Workshop- Banco de Portugal
18th December 2020

Overview

INTRODUCTION
AND MOTIVATION

RELATED
LITERATURE

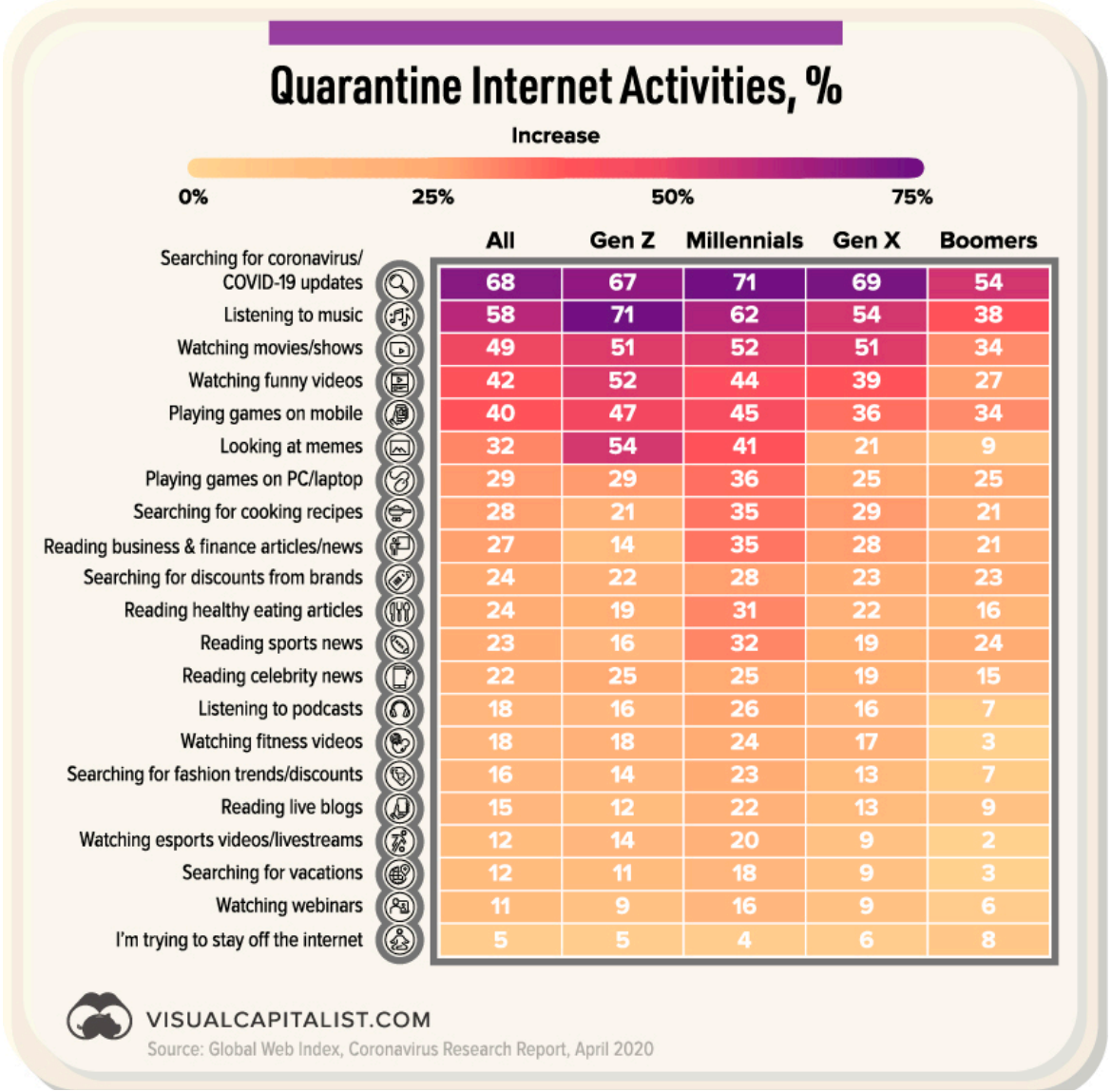
DATA &
METHODOLOGY

RESULTS

LOOKING AHEAD

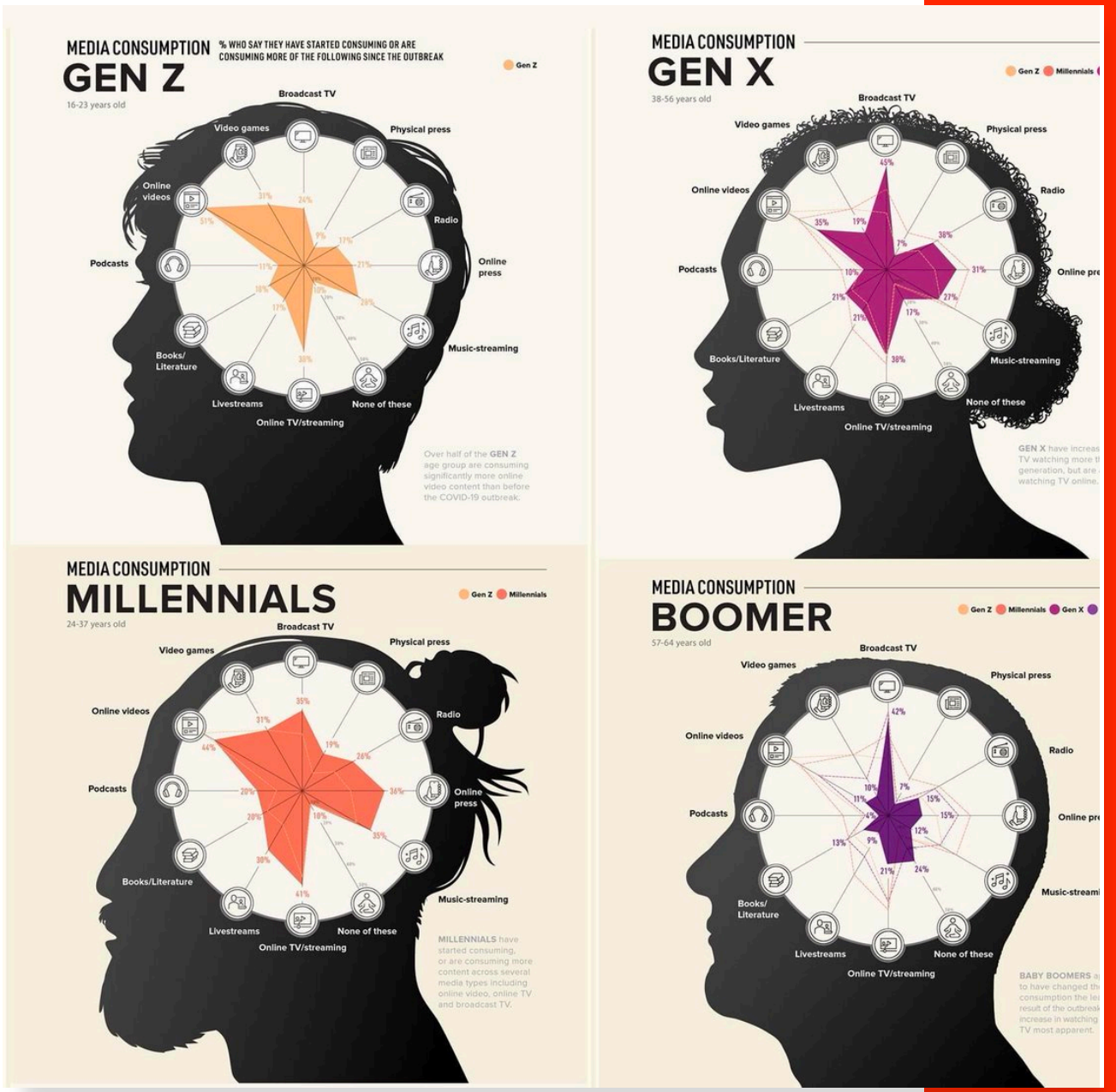
Introduction and Motivation

- The pandemic has fostered people's interest in data and news
- New information about the virus, new government guidance, reports from other countries and the number of cases and deaths on a given day were routinely searched by people in all age categories
- For the first time in a long time even printed newspapers saw an increase in demand
- What matters isn't just how much news people consume, but how much they understand from what they read
- This is likely to shape their opinions and influence their behaviors



Quarantine Internet Activities: a general rise in internet activities

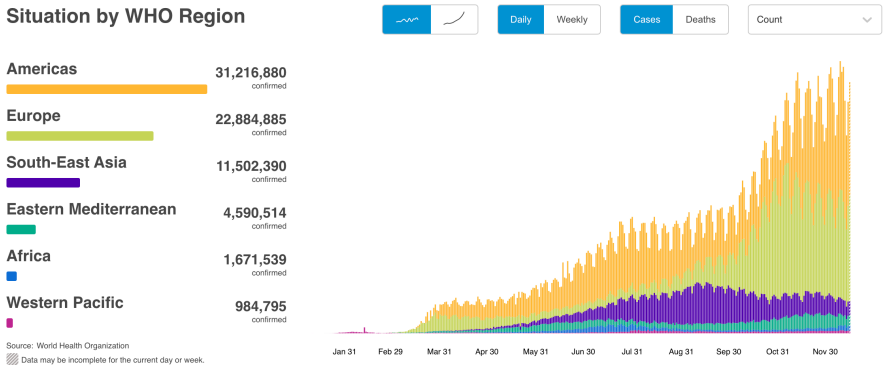
The increase in news consumption is not well predicted by age



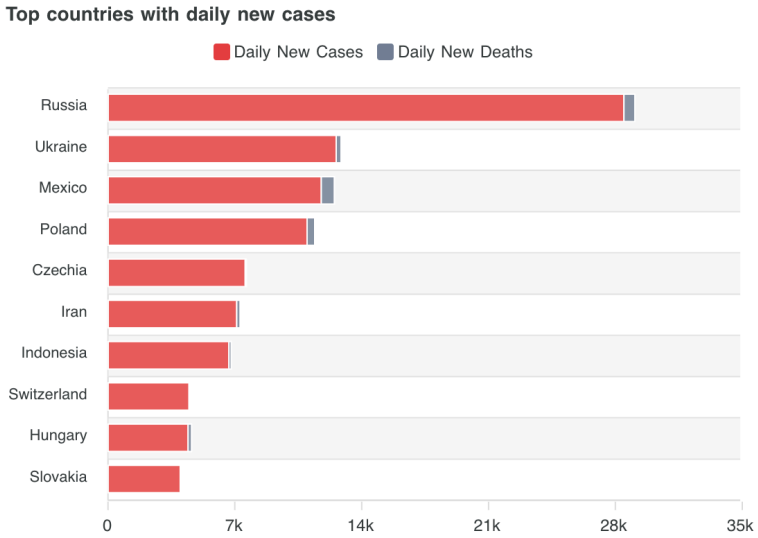
So before we start a couple of polls

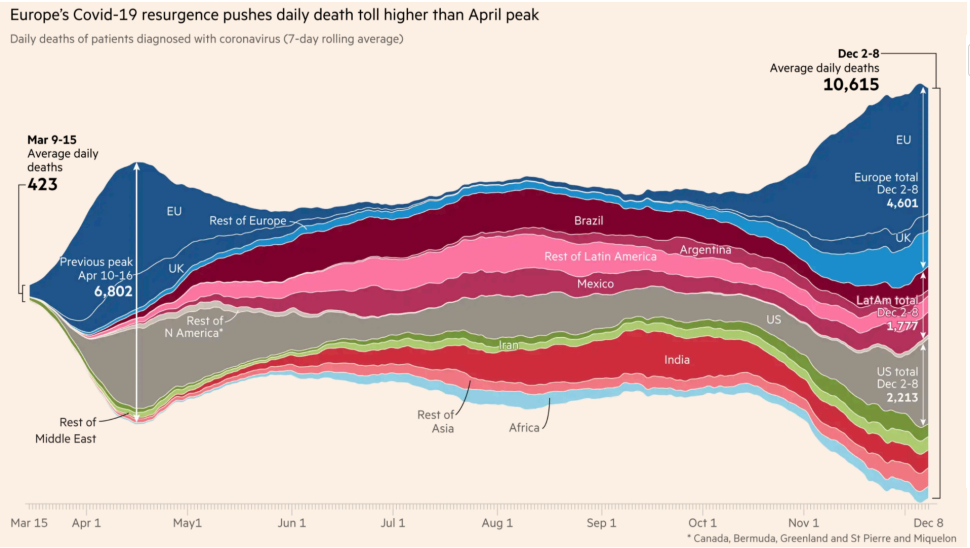
Which of the following plots do you find most informative about the Covid-19 trends?

A)

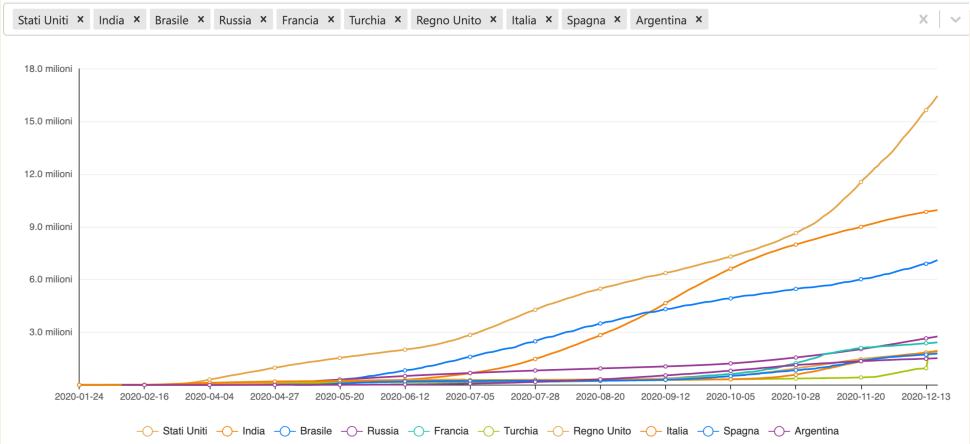


B)





C)

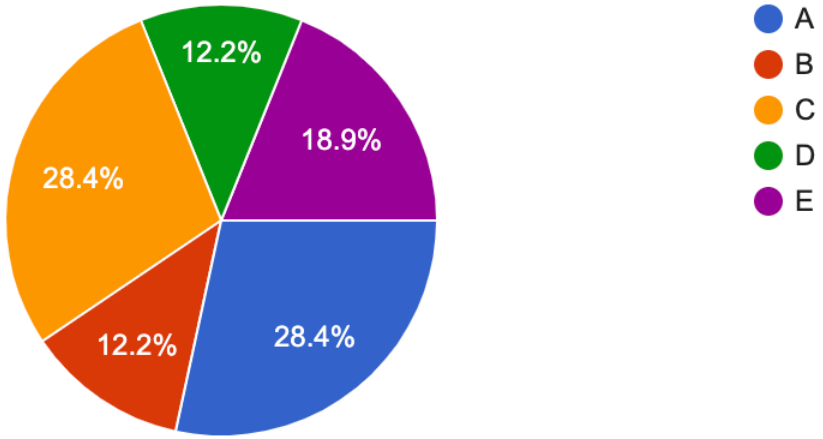


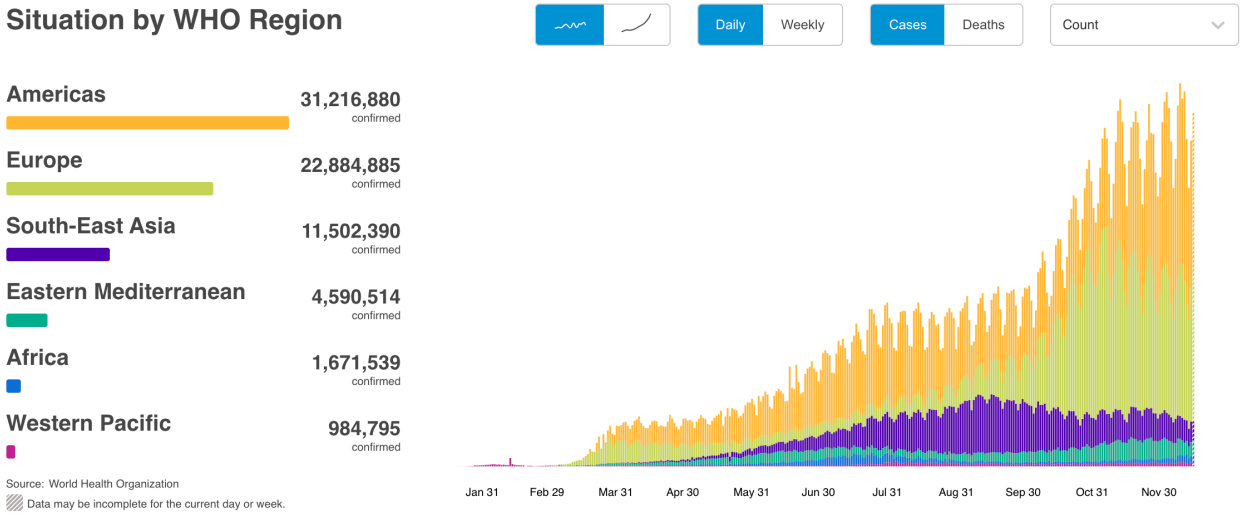
D)

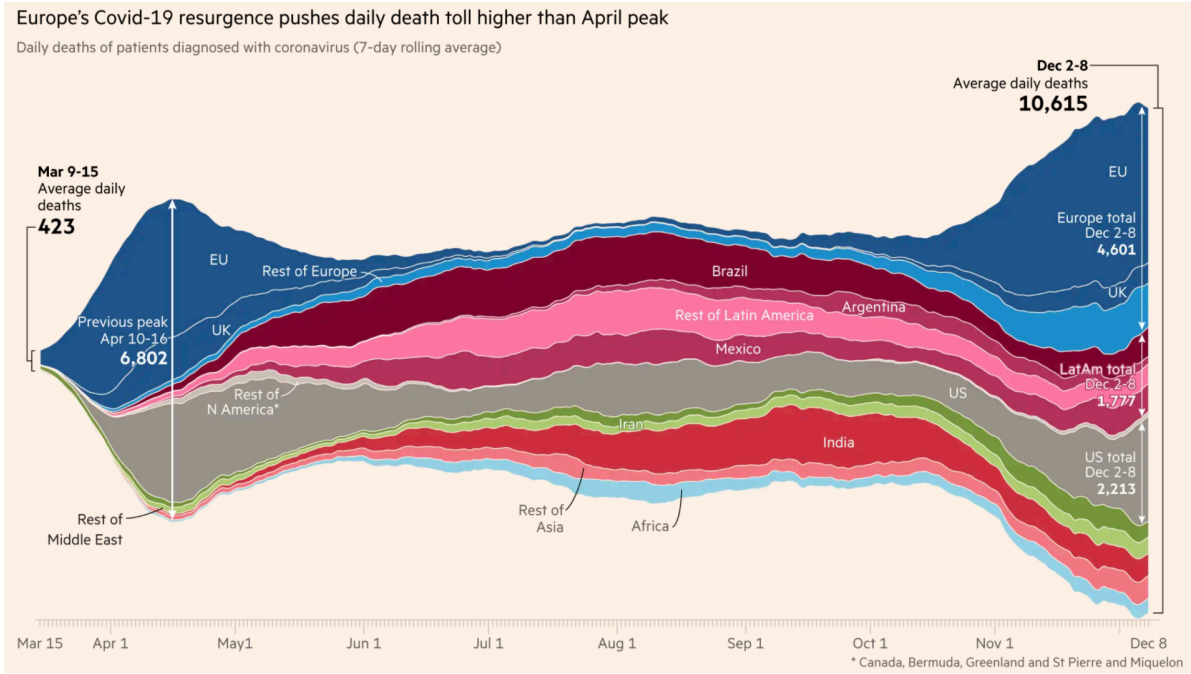
E)

Most informative

74 responses



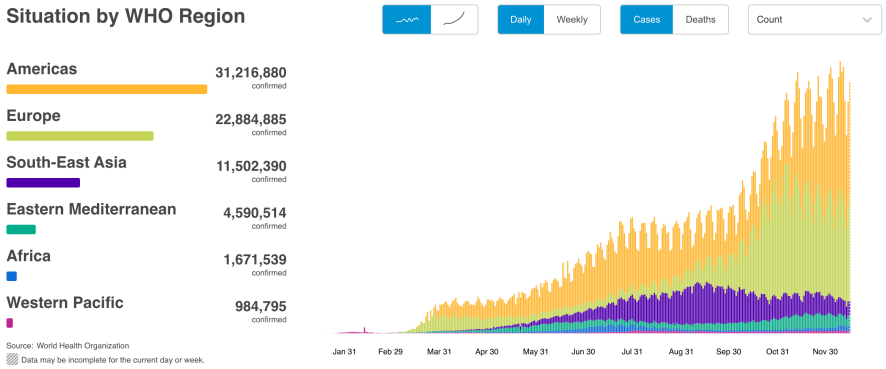




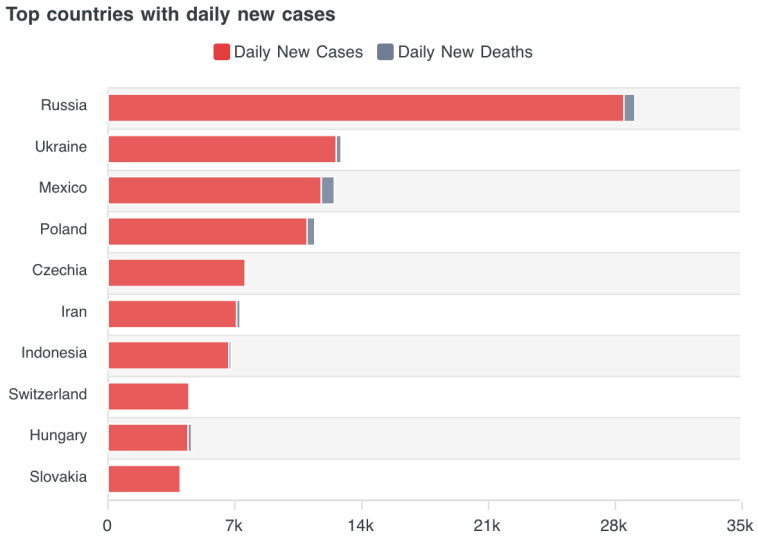
So before we start a couple of polls

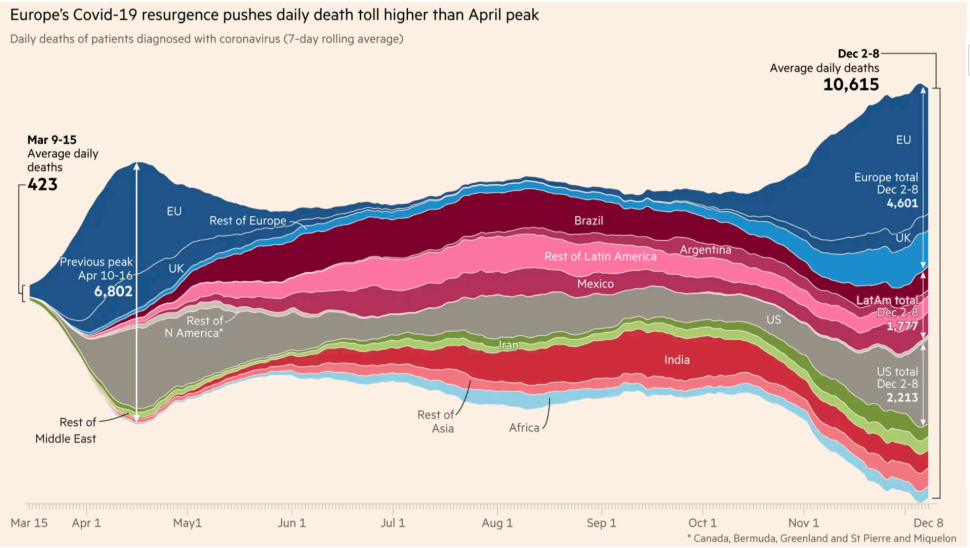
Which of the following plots would you suggest for a newspaper aimed at a general audience?

A)

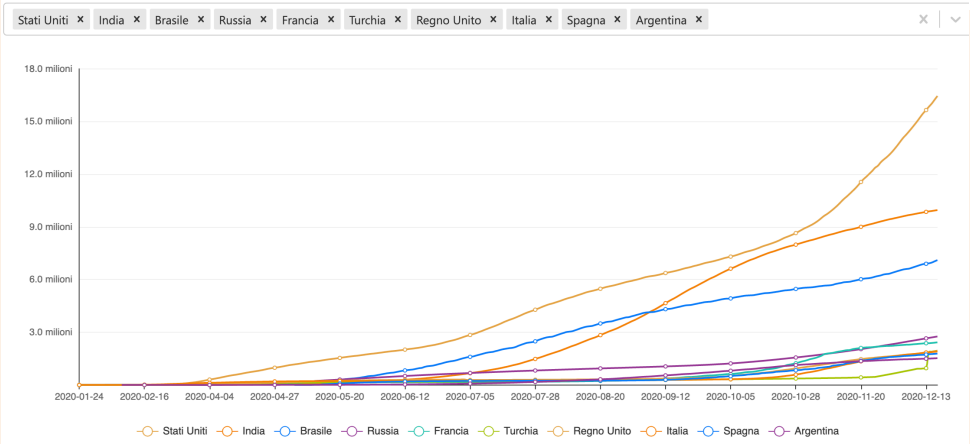


B)

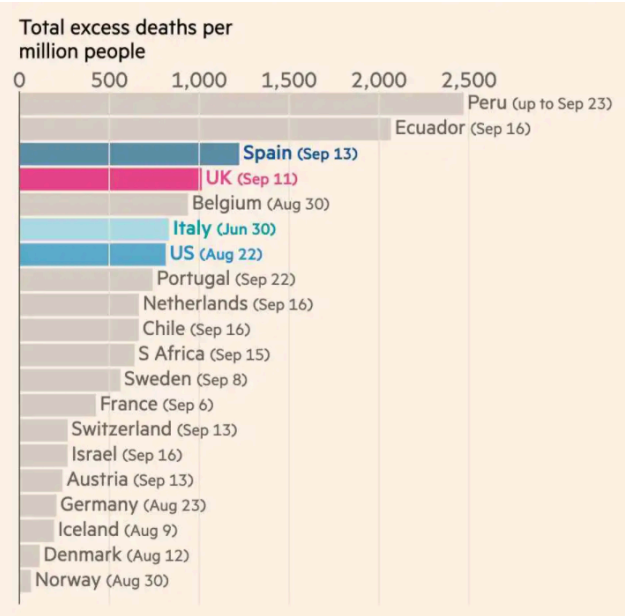




C)



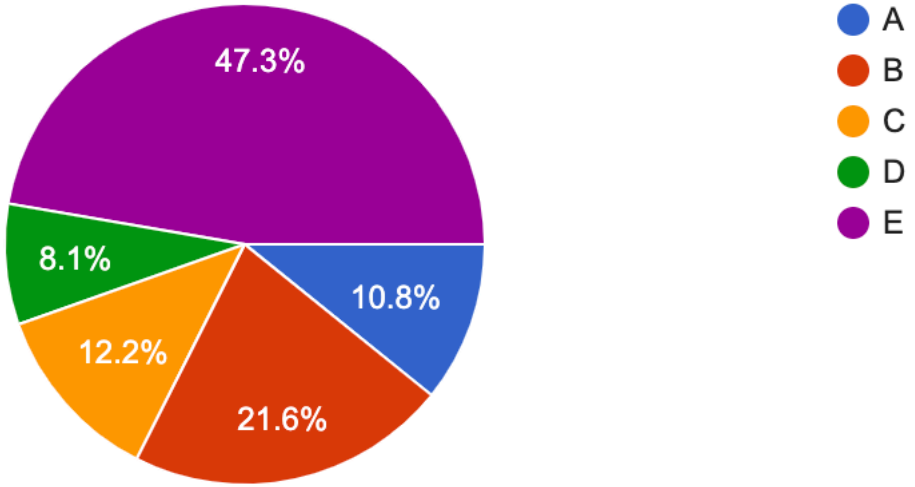
D)

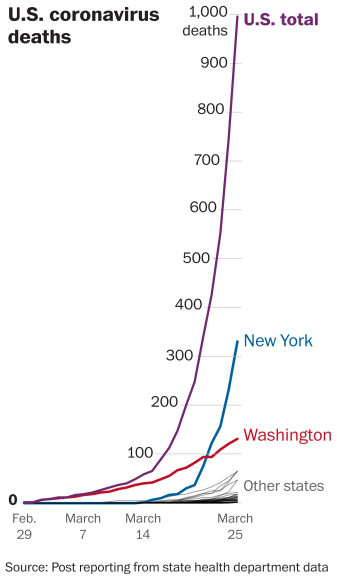


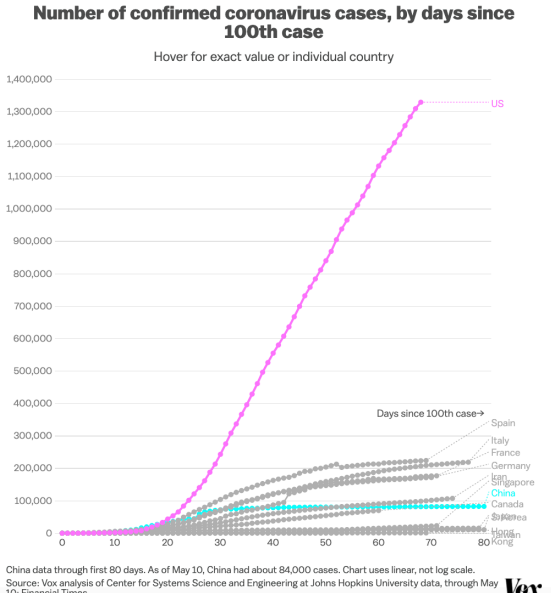
E)

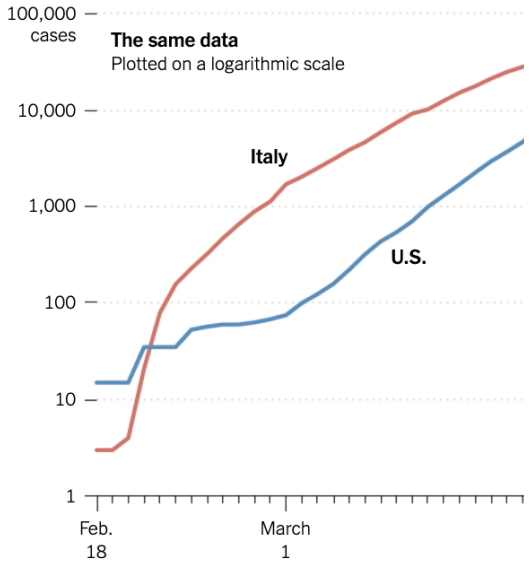
For a general audience

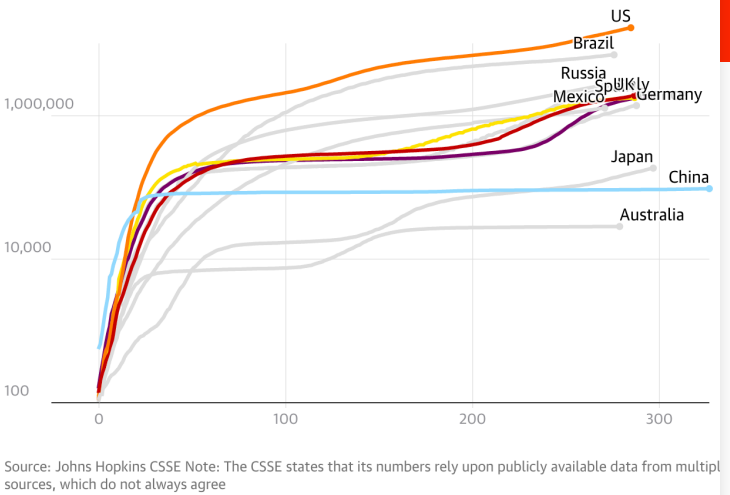
74 responses











Information was delivered in different ways

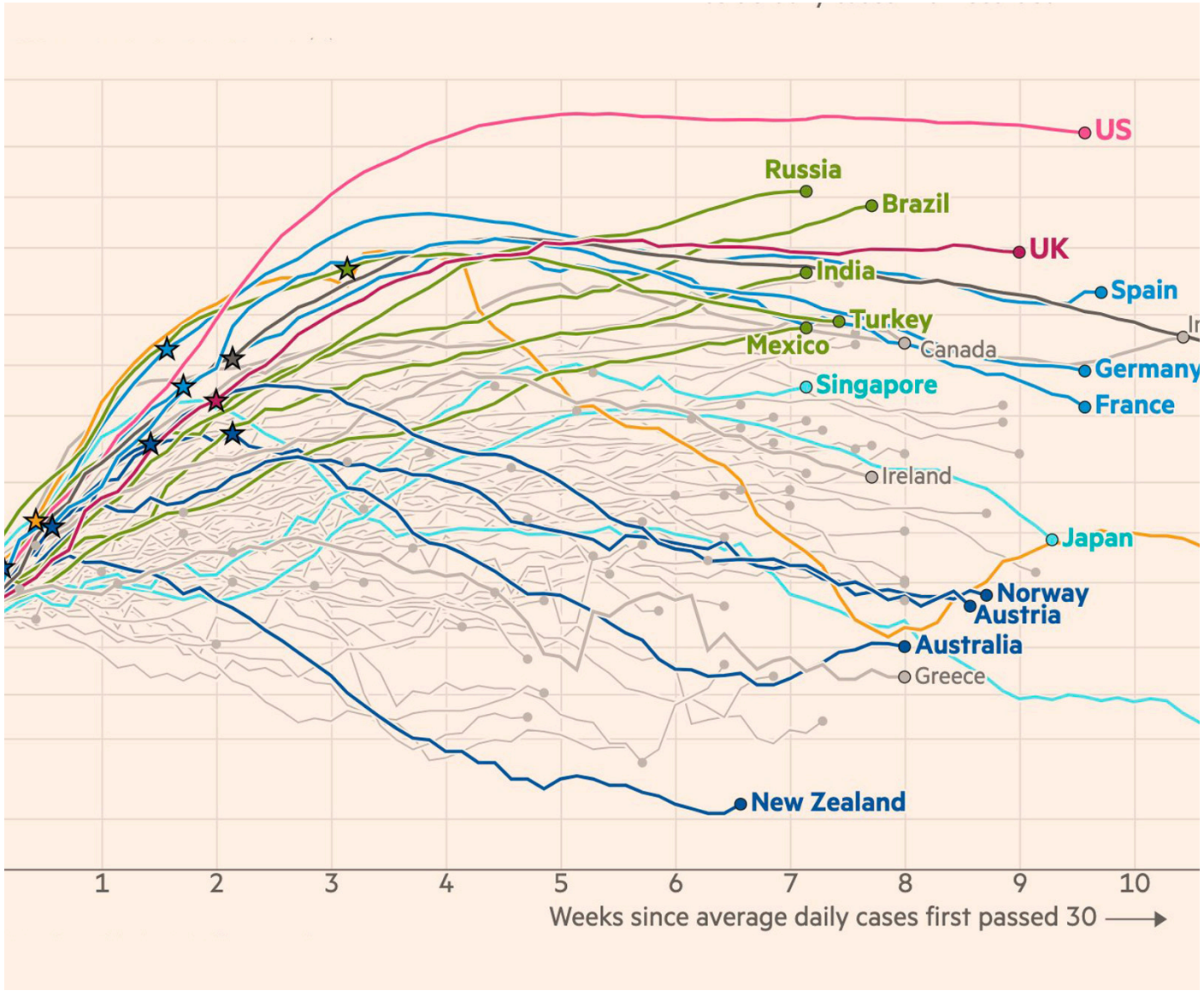
- Washington Post and Vox: linear scale
- New York Times and Guardian UK: logarithmic scale
- Does this information look the same to a reader?

Guardian UK: logarithmic scale

Introduction and Motivation

- In May 2020, reflecting on his famous Covid-19 data tracker Burn-Murdoch said

  "Without data […] it would be very easy for governments or the media or other people with vested interests to just tell their own story about how things were going, and there'd be very little anyone could do to push back on that."



Many outlets chose deliberately the scale to adopt

- For instance, The New York Times has explained that the logarithmic scale helps better visualize exponential growth (New York Times, 2020): “Logarithmic scales can emphasize the rate of change in a way that linear scales do not […]. Unconstrained, the coronavirus spreads exponentially, the caseload doubling at a steady rate. That curve, plotted linearly, is a skyrocketing curve. Plotted logarithmically, however, it transforms into a straight line — which means that deviations from the exponential spread of the virus become much easier to discern.”
- Some epidemiology journals (Gladen, 1983; Levine, Ahmad, & Asa, 2010) and data visualization handbooks (Kosslyn, 2006) agree that the logarithmic scale is preferable during pandemics.

But do people understand the data we provide?

- Garfin, Silver, and Holman (2020) note that during a public health crisis, the general public relies on the media to convey **accurate and understandable information**, so that it can take informed decisions regarding health protective behaviors. Absent information of this kind, people cannot form informed preferences or take informed decisions.
- If the public is provided with data they don't understand the decisions to comply with lockdowns or support government policies is not an informed one
- People may end up regretting their decisions!

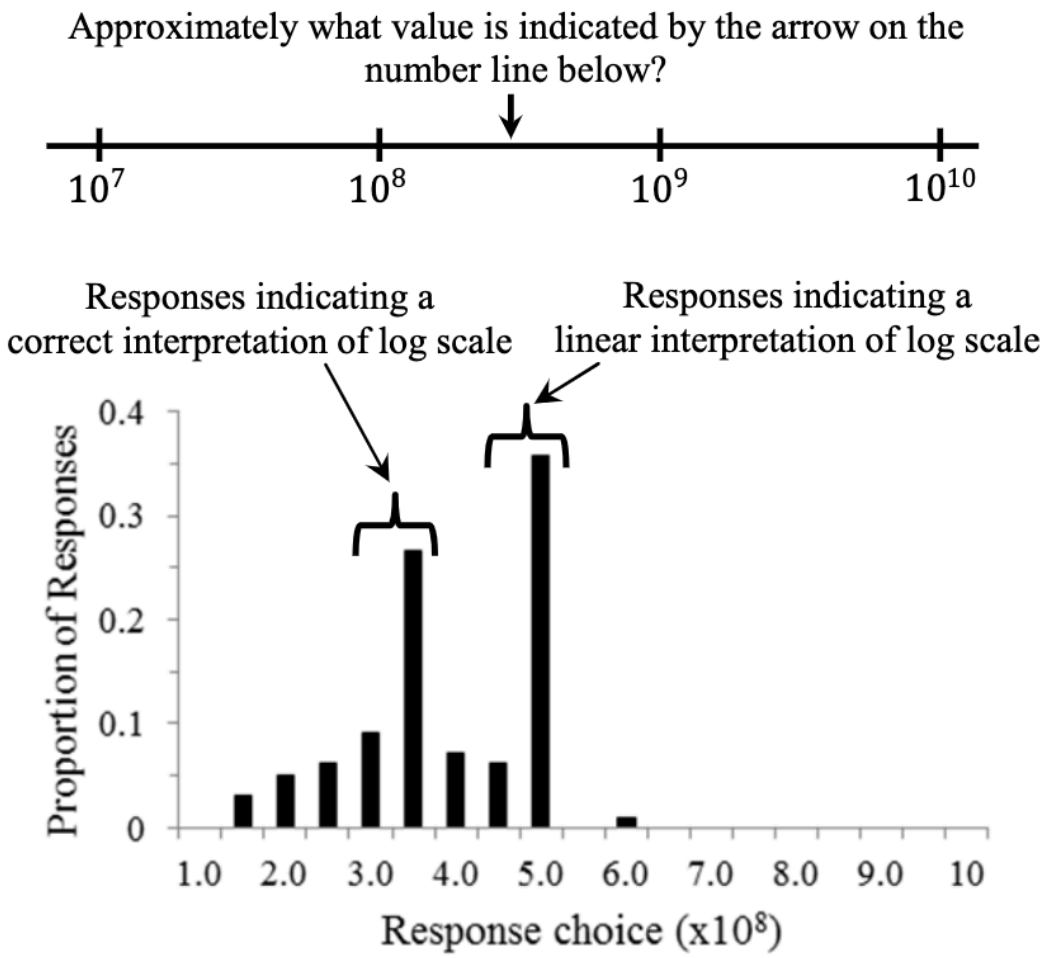


Fig 2. Determine-the-value question with minor tick marks absent. Almost half of student responses indicated a linear interpretation of log scale.

Previous literature shows log scales are hard for people to understand

- Heckler, et al (2013) find that the majority of university sophomore, junior, and senior engineering students in a standard introductory materials science engineering course have a variety of difficulties reading correct values from simple logarithmic graphs
- "Students often unknowingly interpreted the log scale as linear"

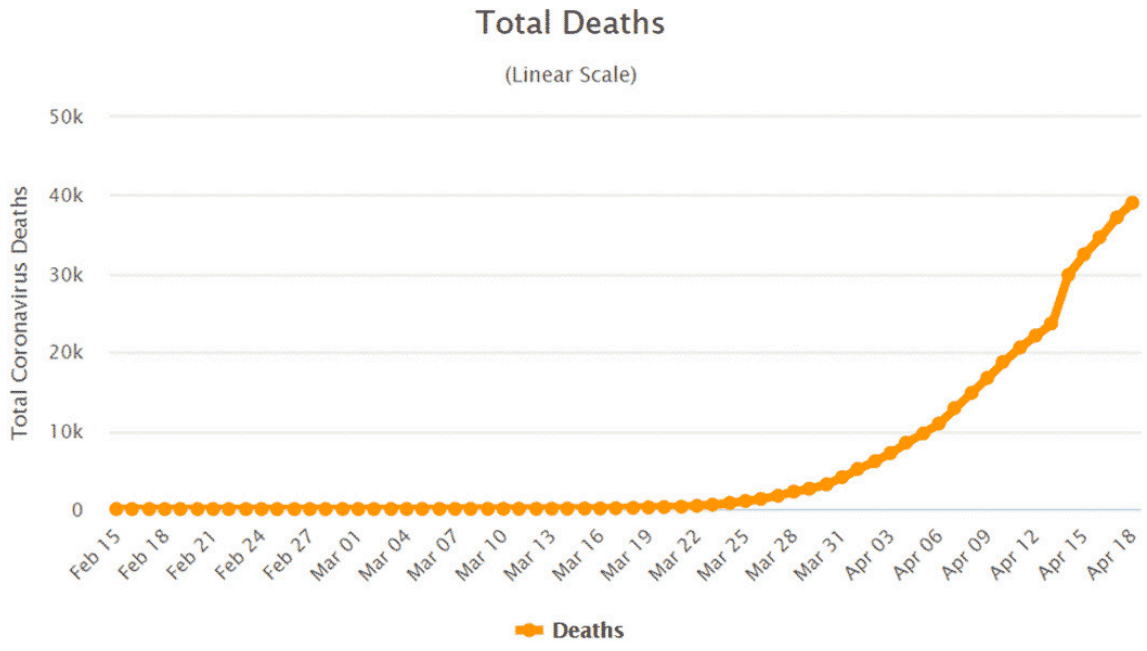
Previous literature shows log scales are hard

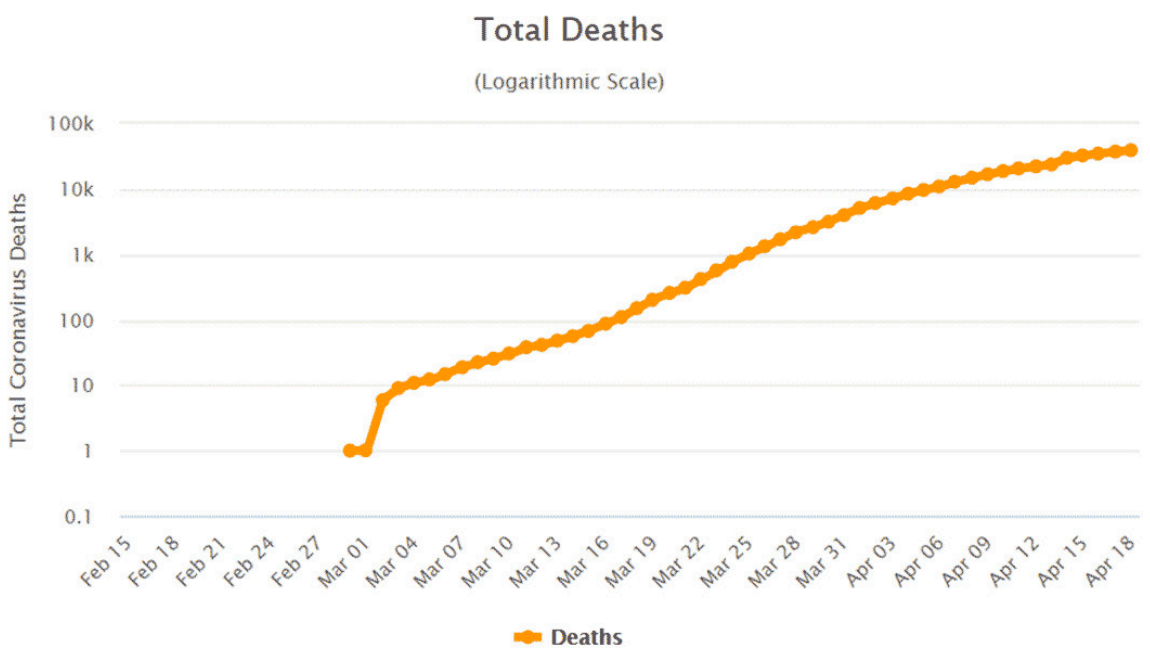
- Menge et al (2018) survey members of the Ecological Society of America and find similar results: Many more respondents interpreted graphs correctly when the graphs had linear–linear axes than when they had at least one log-scale axis even though 69% of the respondents had PhDs
- “Given the frequency of incorrect responses and the regularity with which ecologists are exposed to logarithms, we suspect that confusion about log-scaled data is common among many scientists, not just ecologists.”
- Log scale plots might be even harder to understand for laymen.

So where does this leave us? Method

- Double-blind online experiment on Cloud Research with $n \sim 2000$ after exclusion criteria (above 18 years old, giving consent, US residents)
- **Half** of the participants were randomly assigned to the "Linear Group", in which they were shown the evolution of COVID-19 deaths in the U.S. on a **linear scale** and **half** to the "Log group" in which they were shown the evolution of COVID-19 deaths in the U.S. on a **logarithmic scale**
- Graphs are taken from www.worldometers.info

Graphs shown: linear and log group





The figure above shows the total number of COVID 19 deaths in the US from February 15th to April 18th. The numbers are being measured on what is called a linear scale: A straight line means linear growth, and the steeper or more convex a line, the faster the total number of deaths due to COVID 19 is doubling .

The figure above shows the total number of deaths due to COVID 19 in the US from February 15 to April 18th. The total number of deaths on the 18th of April is 39014. The numbers are being measured on what is called a logarithmic scale: The steeper a line, the faster the total number of deaths due to COVID 19 is doubling

	Graph shown								
	Log Scale			Linear Scale			Total		
	No.	Col %	Cum %	No.	Col %	Cum %	No.	Col %	Cum %
Age									
18-25 years old	126	11.6	11.6	122	12.4	12.4	248	12.0	12.0
26-35 years old	351	32.3	43.9	309	31.3	43.7	660	31.8	43.8
36-45 years old	234	21.5	65.4	237	24.0	67.7	471	22.7	66.5
46-55 years old	182	16.7	82.2	150	15.2	82.9	332	16.0	82.5
56-65 years old	129	11.9	94.0	107	10.8	93.7	236	11.4	93.9
66-75 years old	57	5.2	99.3	52	5.3	99.0	109	5.3	99.1
>75 years old	8	0.7	100.0	10	1.0	100.0	18	0.9	100.0
Education									
Less than high school degree	4	0.4	0.4	5	0.5	0.5	9	0.4	0.4
High school graduate (diploma or equivalent)	88	8.1	8.5	83	8.4	8.9	171	8.3	8.7
Some college but no degree	210	19.3	27.8	168	17.0	26.0	378	18.2	26.9
Associate degree in college (2-year)	97	8.9	36.7	101	10.2	36.2	198	9.6	36.5
Bachelor's degree in college	478	44.0	80.8	402	40.8	77.0	880	42.5	79.0
Master's degree or Professional Degree (JD, MD, etc)	190	17.5	98.3	203	20.6	97.6	393	19.0	97.9
Doctoral degree	19	1.7	100.0	24	2.4	100.0	43	2.1	100.0
Income									
Less than $10,000	48	4.4	4.4	36	3.7	3.7	84	4.1	4.1
$10,000 to $19,999	64	5.9	10.3	56	5.7	9.3	120	5.8	9.9

Sample

Sample

	Graph shown								
	Log Scale			Linear Scale			Total		
	No.	Col %	Cum %	No.	Col %	Cum %	No.	Col %	Cum %
Political orientation									
Other	352	32.4	32.4	292	29.6	29.6	644	31.1	31.1
Democrat	441	40.6	73.0	426	43.2	72.7	867	41.8	72.9
Republican	294	27.0	100.0	269	27.3	100.0	563	27.1	100.0
Total	1087	100.0		987	100.0		2074	100.0	
Gender									
Other/Prefer not to declare	8	0.7	0.7	14	1.4	1.4	22	1.1	1.1
Female	571	52.5	53.3	524	53.1	54.5	1095	52.8	53.9
Male	508	46.7	100.0	449	45.5	100.0	957	46.1	100.0
Live in city with <50K People									
No	680	62.6	62.6	601	60.9	60.9	1281	61.8	61.8
Yes	407	37.4	100.0	386	39.1	100.0	793	38.2	100.0
Total	1087	100.0		987	100.0		2074	100.0	
Live in city with >500K People									
No	851	78.3	78.3	769	77.9	77.9	1620	78.1	78.1
Yes	236	21.7	100.0	218	22.1	100.0	454	21.9	100.0

Survey structure and variables of interest

- Participants are shown the graph before they start answering questions
- The questions asked can be grouped into:

    1) determinants of worry
    2) policy preferences
    3) understanding

- Participants see the graph again between 2) and 3)
- Participants are asked standard demographics questions at the end

Block 1: Determinants of worry

- Participants see a description of the scale used in the graph and the graph to which they are assigned –> participants are asked about how worried they are about the **health** crisis and the **economic** crisis caused by COVID-19)
- The scale shown impacts people's level of worry for the health crisis (but not for the economic crisis)
- People seeing the linear scale plots are **more worried** about the health crises caused by COVID-19 *regardless* of their understanding of the graphs and demographics

	(1) Worry About Health Crisis	(2) Worry About Health Crisis	(3) Worry About Health Crisis
Worry About Health Crisis			
In Linear Group	0.141*	0.258*	0.327**
	(0.081)	(0.091)	(0.038)
COVID-19 News Checking		0.500***	0.434***
		(<0.001)	(<0.001)
Male		-0.806***	-0.654***
		(<0.001)	(<0.001)
Democrat			0.732***
			(<0.001)
Republican			-0.282**
			(0.017)
Worry About Economic Crisis			0.707***
			(<0.001)
Live in city with <50K People			0.0156
			(0.880)
Live in city with >500K People			-0.132
			(0.280)
Education			-0.0258
			(0.473)
Age			-0.00132
			(0.694)

Block 2: Policy preferences

- We ask participants:

1. their **support for closing non-essential businesses** (essential businesses are defined to be supermarkets, pharmacies, etc.)
2. **until when they would keep these businesses closed** (date)
3. **how often they would use a mask if the government sent a supply**
4. their **support for a tax that finances the distribution of masks** for everyone in their State

Policy preferences

- People in the Linear Group prefer that nonessential businesses remain closed for longer.
- However, they support less strongly the idea of closing nonessential business in the first place

	(1) Likelihood to Wear Masks	(2) Likelihood to Wear Masks	(3) Likelihood to Wear Masks	(4) Support for Mask-Buying Tax	(5) Support for Mask-Buying Tax	(6) Support for Mask-Buying Tax
In Linear Group	0.00311	-0.314**	-0.350**	-0.0218	0.307**	0.305**
	(0.970)	(0.045)	(0.029)	(0.780)	(0.042)	(0.046)
Worry About Health Crisis		0.907***	0.908***		0.481***	0.471***
		(<0.001)	(<0.001)		(<0.001)	(<0.001)
COVID-19 News Checking		0.138***	0.129***		0.0403	0.0682
		(0.003)	(0.006)		(0.341)	(0.116)
Male		-0.255***	-0.270***		0.0372	0.0455
		(0.007)	(0.005)		(0.673)	(0.612)
Democrat			0.161			0.378***
			(0.154)			(<0.001)
Republican			-0.384***			-0.261**
			(0.001)			(0.024)
Worry About Economic Crisis			-0.132**			-0.0979*
			(0.021)			(0.069)
Live in city with <50K People			0.0832			0.115
			(0.424)			(0.240)
Live in city with >500K People			0.588***			0.0488
			(<0.001)			(0.681)
Education			-0.0767**			-0.0209
			(0.040)			(0.543)
Age			0.00713**			-0.00942***
			(0.041)			(0.004)

Policy preferences

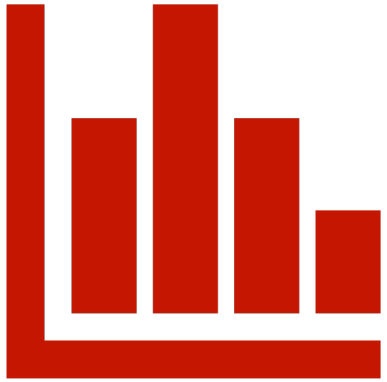
- People in the Linear Group would wear government-supplied masks less often
- However, they support more strongly government provision

	(1) Likelihood to Wear Masks	(2) Likelihood to Wear Masks	(3) Likelihood to Wear Masks	(4) Support for Mask-Buying Tax	(5) Support for Mask-Buying Tax	(6) Support for Mask-Buying Tax
In Linear Group	0.00311	-0.314**	-0.350**	-0.0218	0.307**	0.305**
	(0.970)	(0.045)	(0.029)	(0.780)	(0.042)	(0.046)
Worry About Health Crisis		0.907***	0.908***		0.481***	0.471***
		(<0.001)	(<0.001)		(<0.001)	(<0.001)
COVID-19 News Checking		0.138***	0.129***		0.0403	0.0682
		(0.003)	(0.006)		(0.341)	(0.116)
Male		-0.255***	-0.270***		0.0372	0.0455
		(0.007)	(0.005)		(0.673)	(0.612)
Democrat			0.161			0.378***
			(0.154)			(<0.001)
Republican			-0.384***			-0.261**
			(0.001)			(0.024)
Worry About Economic Crisis			-0.132**			-0.0979*
			(0.021)			(0.069)
Live in city with <50K People			0.0832			0.115
			(0.424)			(0.240)
Live in city with >500K People			0.588***			0.0488
			(<0.001)			(0.681)
Education			-0.0767**			-0.0209
			(0.040)			(0.543)
Age			0.00713**			-0.00942***
			(0.041)			(0.004)

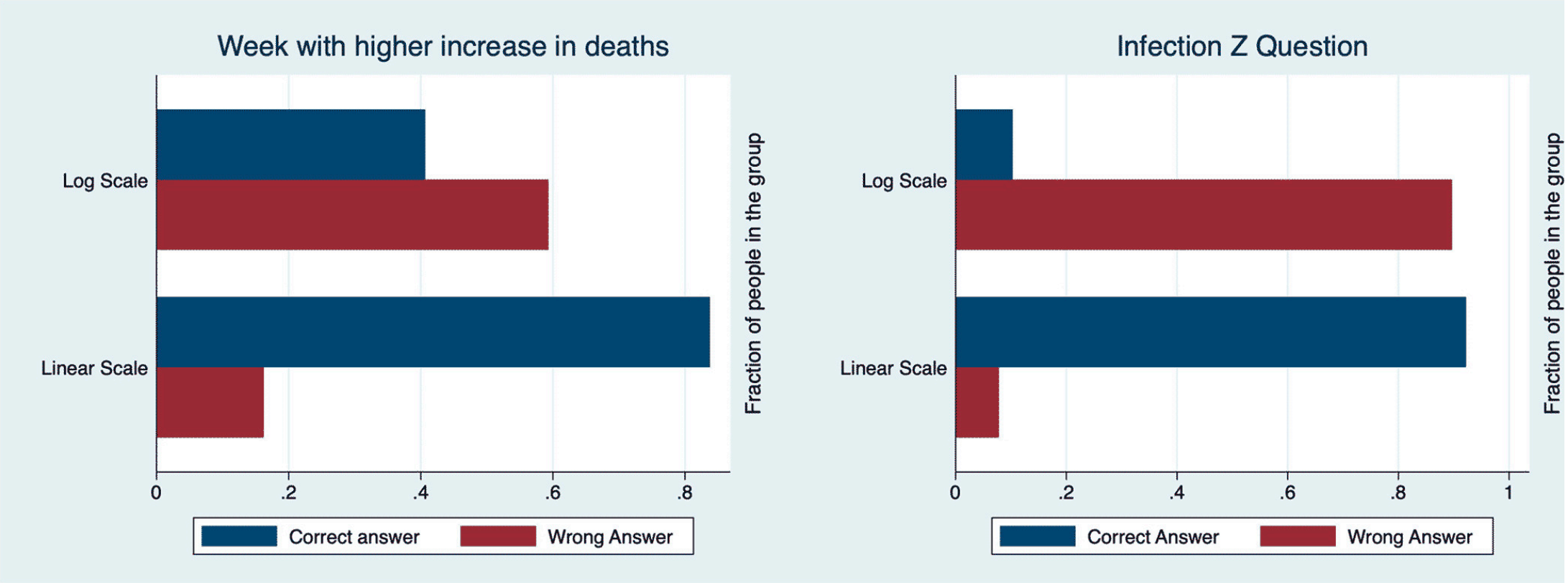
Block 3: Understanding questions

1. Participants are asked whether the number of deaths increased more between March 31st and April 6th or between April 6th and April 12th.
2. Second, we show them a graph describing non-COVID-19 related data on the number of deaths from a hypothetical infection Z (taken from Okan, Galesic, & Garcia-Retamero, **2016**) and asked them a similar question.
3. We test whether respondents can make predictions based on the curve. In particular, by asking them to make a prediction on the total number of deaths on April 25th, one week after we launched the experiment.

Forecasting: what's the benchmark?

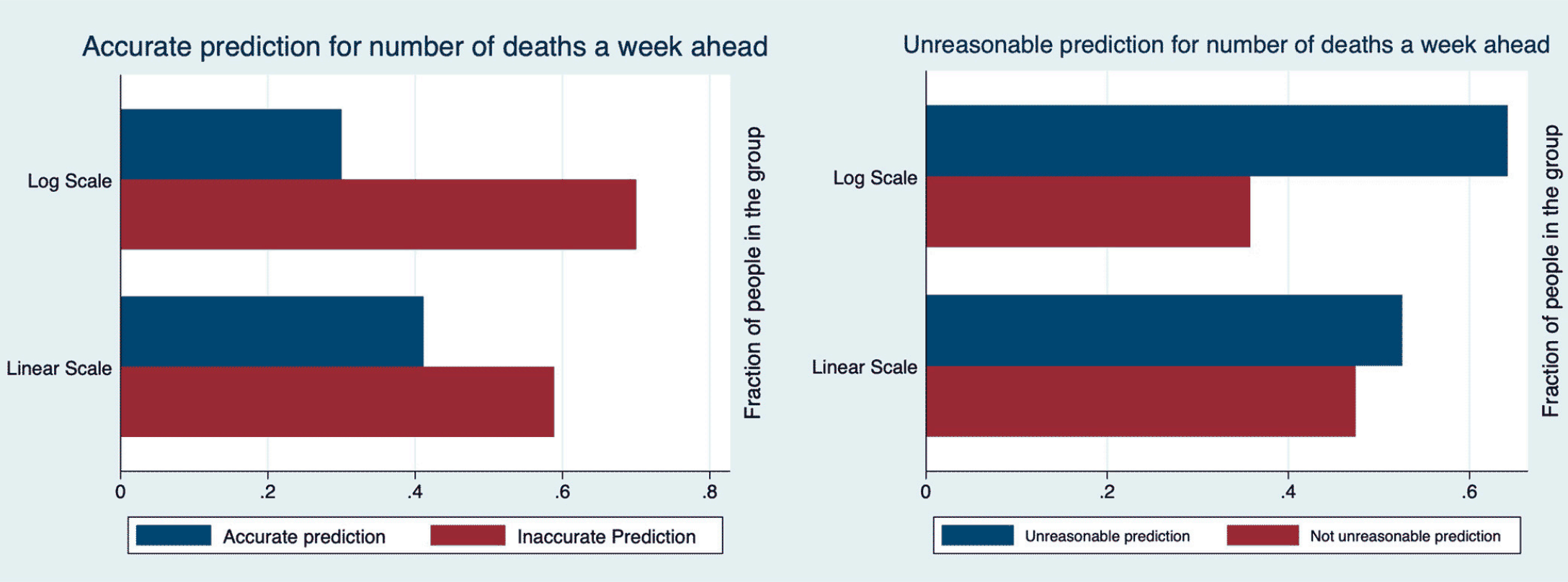


- Predicting the number of COVID-19 related deaths in a week is very difficult, but some predictions are more reasonable than others.
- We forecast the number of total deaths on April 25th using an ARIMA model (model as in Benvenuto, Giovanetti, Vassallo, Angeletti, & Ciccozzi, 2020) -> ARIMA(0,2,1)
- Focus on two measures:
  95% $\boldsymbol{CI = [49,203.15\ to\ 62,559.27]}$, **and**
  99% $\boldsymbol{CI = [46,895.47, 64,685.95]}$
  (the actual number of deaths on the 25th of April was 54,256, while our ARIMA predicted 55,791)
- If the participant's prediction is within the 95% CI we deem it "accurate", if outside the 95% CI but within the 99% we deem it "unlikely", if outside the 99% CI we deem it "unreasonable"



Understanding questions: increase in deaths

- The left panel reports the percentage of correct and incorrect answers provided by the members of the two groups to the understanding question related to COVID-19 real world data. The right panel reports the percentage of correct and incorrect answers provided by the members of the two groups to the understanding question related to Infection Z hypothetical data



Understanding questions: forecasts

- The left panel reports the percentage of accurate and inaccurate (i.e., not accurate) predictions provided by the members of the two groups. The right panel reports the unreasonable and reasonable (i.e., not unreasonable) predictions provided by the members of the two groups

	(1) Worry About Health Crisis	(2) Likelihood to Wear Masks	(3) Support for Mask-Buying Tax	(4) Support for Closing Businesses	(5) Understanding Q.1: Real Data	(6) Understanding Q.2: Hypothetical	(7) Accurate Prediction	(8) Unreasonable Prediction
In Linear Group	1.387*	0.705*	1.356*	0.654*	7.800***	123.9***	1.619***	0.619***
	(0.218)	(0.113)	(0.207)	(0.110)	(0.902)	(23.13)	(0.159)	(0.0594)
COVID-19 News Checking	1.543***	1.138**	1.071	1.078	1.081	1.090	1.162**	0.840***
	(0.0718)	(0.0537)	(0.0464)	(0.0514)	(0.0578)	(0.0886)	(0.0563)	(0.0398)
Male	0.520***	0.763**	1.047	0.915	0.864	1.379	0.968	0.985
	(0.0486)	(0.0735)	(0.0937)	(0.0900)	(0.0972)	(0.241)	(0.0988)	(0.0980)
Understanding Q.1: Real Data	1.006	1.014	1.184	1.141				
	(0.107)	(0.112)	(0.120)	(0.127)				
Confidence in Understanding Q.1	0.998	1.005	1.006	1.008*	1.009***			
	(0.00361)	(0.00379)	(0.00355)	(0.00375)	(0.00253)			
Understanding Q.2: Hypothetical	0.799	1.267	0.636**	1.416*				
	(0.131)	(0.212)	(0.101)	(0.247)				
Confidence in Understanding Q.2	0.996	1.003	0.989**	1.000		1.031***		
	(0.00368)	(0.00385)	(0.00360)	(0.00379)		(0.00424)		
Accurate Prediction	1.244	1.539*	1.152	1.569*				
	(0.238)	(0.290)	(0.213)	(0.302)				
Unreasonable Prediction	1.384	1.638**	1.159	1.084				
	(0.260)	(0.301)	(0.209)	(0.202)				
Confidence in Prediction	1.006**	1.003	1.007***	0.996			0.998	1.002
	(0.00225)	(0.00231)	(0.00221)	(0.00236)			(0.00234)	(0.00229)
Democrat	2.080***	1.175	1.459***	1.725***	1.004	1.091	1.096	0.900
	(0.225)	(0.133)	(0.152)	(0.200)	(0.131)	(0.216)	(0.130)	(0.104)
Republican	0.754*	0.681**	0.770*	0.612***	0.981	0.833	0.834	1.247
	(0.0893)	(0.0822)	(0.0891)	(0.0735)	(0.141)	(0.186)	(0.111)	(0.161)

The effects of seeing data on different scales

What's driving the results?

An hypothesis

- Merely changing the scale can alter public policy preferences and the level of worry, despite the endless flow of COVID-19 related information to which everyone is exposed.
- We cannot be sure about the mechanism driving the result for each of the participants
- But the results are robust to a number of controls and specifications (OLogit, OProbit, OLS)
- The shape of the curves could explain these findings: flat logarithmic curve (plateau) vs constantly growing line (the end isn't near)

Conclusions

- Regardless of the reasons behind our findings, it is noteworthy that changing the scale can alter policy preferences, intentions to adopt precautionary measures, and level of worry for the health consequences of the pandemic.
- Unclear information conveyed by the media can undermine how effective government policies are and affect people's trust in science, another key predictor of compliance with COVID-19 guidelines (Brzezinski, Kecht, Van Dijcke, & Wright Austin, 2020; Phlol & Musil, 2020). It also prevents people from making informed decisions
- Using linear scale graphs can help people understand the data better and should be preferred for a general audience

Thank you!

Questions/comments/feedback? Feel free to raise your point during the discussion and/or email me at c.sotis@lse.ac.uk

On Twitter as @chiarawouldsay

On LinkedIn as Chiara Sotis