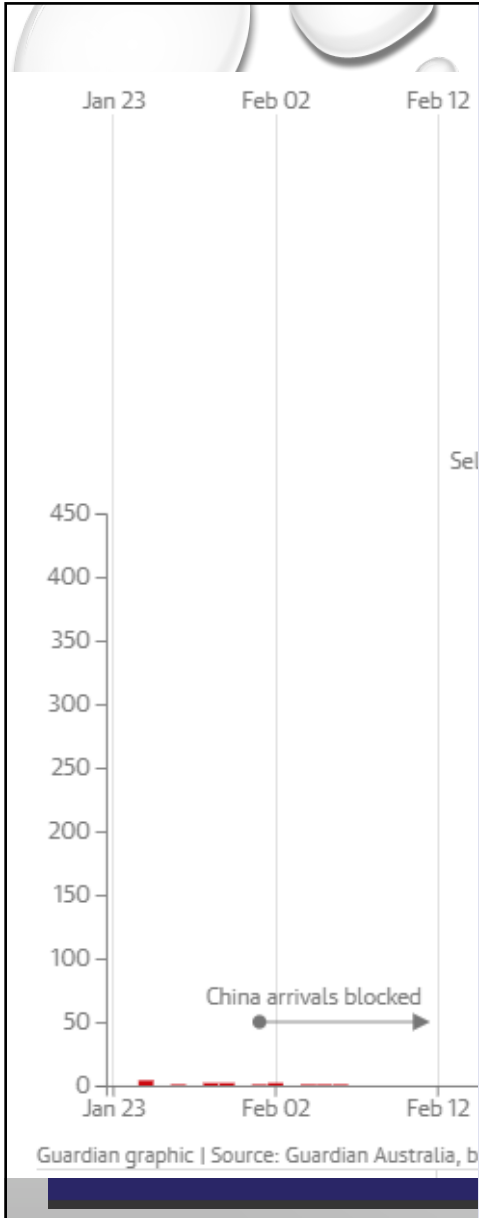


The background of the slide is a light gray gradient with several realistic water droplets of various sizes scattered across it. The droplets have highlights and shadows, giving them a three-dimensional appearance. The text is centered on the slide.

THINKING ABOUT DATA

BEING A DISCERNING CONSUMER OF DATA



THERE ARE EVEN STUDIES TO TELL US HOW INFORMATION IS SPREADING

The COVID-19 Social Media Infodemic

Matteo Cinelli¹, Walter Quattrociocchi^{*2,1,3}, Alessandro Galeazzi⁴,
Carlo Michele Valensise⁵, Emanuele Brugnoti¹, Ana Lucia
Schmidt², Paola Zola⁶, Fabiana Zollo^{2,1}, and Antonio Scala^{1,3}

¹CNR-ISC, Roma

²Universit Ca Foscari di Venezia

³Big Data in Health Society, Roma

⁴Universit di Brescia

⁵Politecnico di Milano

⁶CNR-IIT, Pisa

Abstract

We address the diffusion of information about the COVID-19 with a massive data analysis on Twitter, Instagram, YouTube, Reddit and Gab. We analyze engagement and interest in the COVID-19 topic and provide a differential assessment on the evolution of the discourse on a global scale for each platform and their users. We fit information spreading with epidemic models characterizing the basic reproduction numbers R_0 for each social media platform. Moreover, we characterize information spreading from questionable sources, finding different volumes of misinformation in each platform. However, information from both reliable and questionable sources do not present different spreading patterns. Finally, we provide platform-dependent numerical estimates of rumors' amplification.

⁴<https://mediabiasfactcheck.com> classifies news sources that are considered reliable and news sources that are considered unreliable

<https://arxiv.org/pdf/2003.05004.pdf?fbclid=IwAR08xVUpifbp5QKhwP3FPqu7J6oAswtPoLJliG6plgumFfQfnVRml5Z8Sho>

DOING RESEARCH

- There are lots of ways of collecting data about humans, including:
 - Biomedical testing
 - Surveys and other quantitative methods
 - Observation
 - Qualitative information (Interviews, focus groups)
- All of them have their place, depending on what questions you are trying to answer
- All of them have pros and cons, and may be subject to bias, misinterpretation and mishandling

DOING RESEARCH

- Researchers have to make choices and put limits on their research
- Once researchers have their data, they analyse and interpret.
- Our real example

Thinking about data

Thank you!

My name is Dr Jenni Greig, and I am an academic at Charles Sturt University, located in Bathurst, Australia.

I have designed this short survey to use as a teaching tool - looking at how the questions you ask, and the answer options you give in a survey shapes the kind of data you can collect. There are questions about Covid-19 because it is topical, not because I am conducting any research into the pandemic.

It is completely up to you if you want to participate. I'm hoping the survey will be spread widely, so it isn't just my friends participating. Participation in the survey is completely anonymous, and the data collected will only be used as a teaching tool. It should only take you 3 to 5 minutes.

WHO WERE THE PARTICIPANTS?

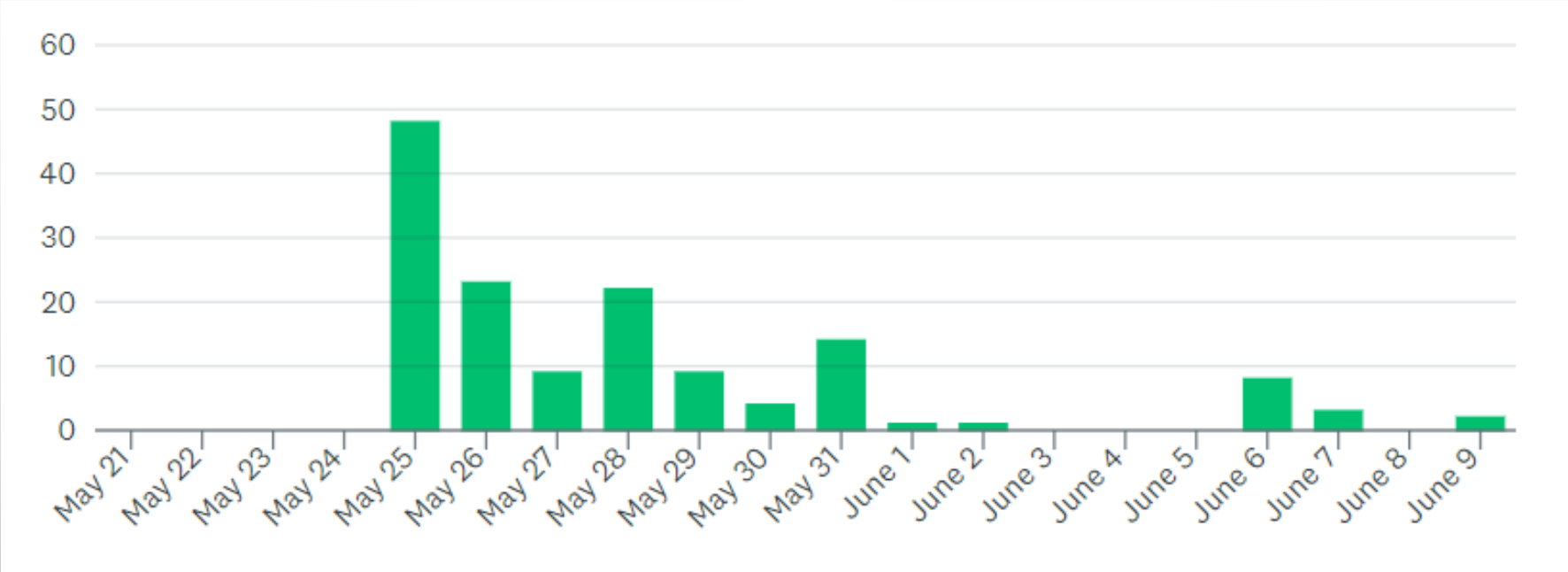
- Who we get the data from matters
 - Who participated? Why did they participate?
 - How were they recruited?

Survey Results

100% of respondents believe pandas are cuter than koalas (in a survey of pandas).

Permalink: <http://thedoghousediaries.com/5314>

RECRUITING PARTICIPATION



WHO WERE THE PARTICIPANTS?

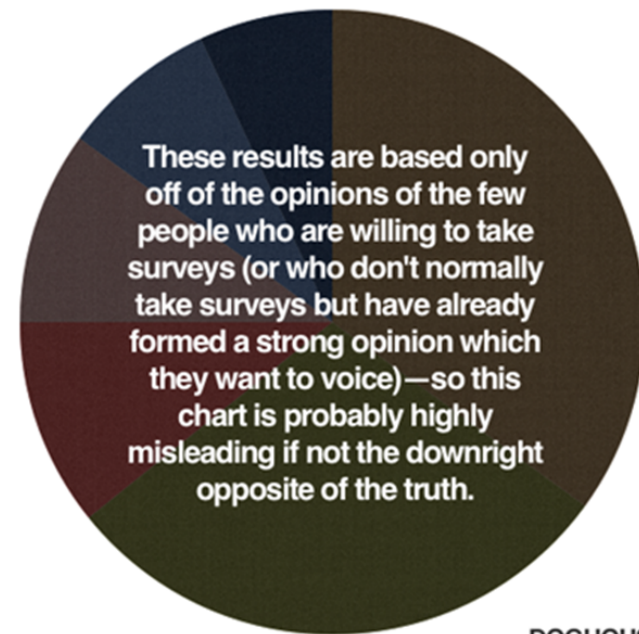
- Who we get the data from matters
 - Who participated? Why did they participate?
 - How were they recruited?
 - Who didn't participate? Why?

Survey Results

100% of respondents believe pandas are cuter than koalas (in a survey of pa

Permalink: <http://thedoghousediaris.com/5314>

What I See When I Look At Any Survey Results



DOGHOUSEDIARIES

TO WHAT GROUP/S DOES THE DATA APPLY?



- What was the population of interest?
- How was the sample recruited?
- Do the findings apply outside of the sample?
- Does it tell me anything about people not in the population of interest?



WHAT TYPE OF RESEARCH IS IT?

1. **EXPERIMENTAL RESEARCH**

- Concerned with establishing the *causes* of something
 - Such as the spread of a virus, causes of a specific behaviour, the effect of one factor on another factor.
 - Requires controlling factors so we can isolate the cause-effect relationship

2. **DESCRIPTIVE RESEARCH**

- Concerned with *describing* something
 - This may also be about the way a virus spreads, observations of factors statistically associated with a specific behaviour, or the strength of an association between two factors
 - May be able to control for some factors, but can't isolate cause-effect relationship

The way we gather data defines:

- What kind of information we get
- What kind of conclusions we can draw

* 4. Thinking about the current restrictions imposed wherever you are at the time you are doing this survey, do you think they are effective in keeping your community safe:

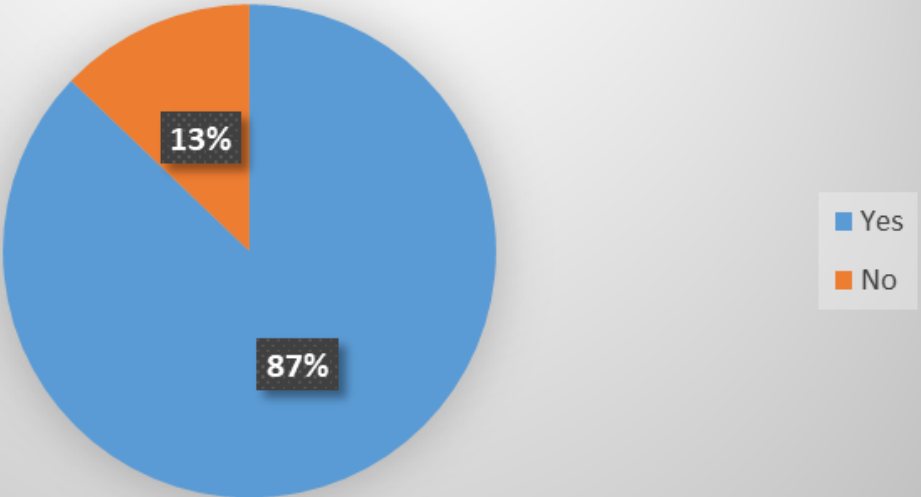
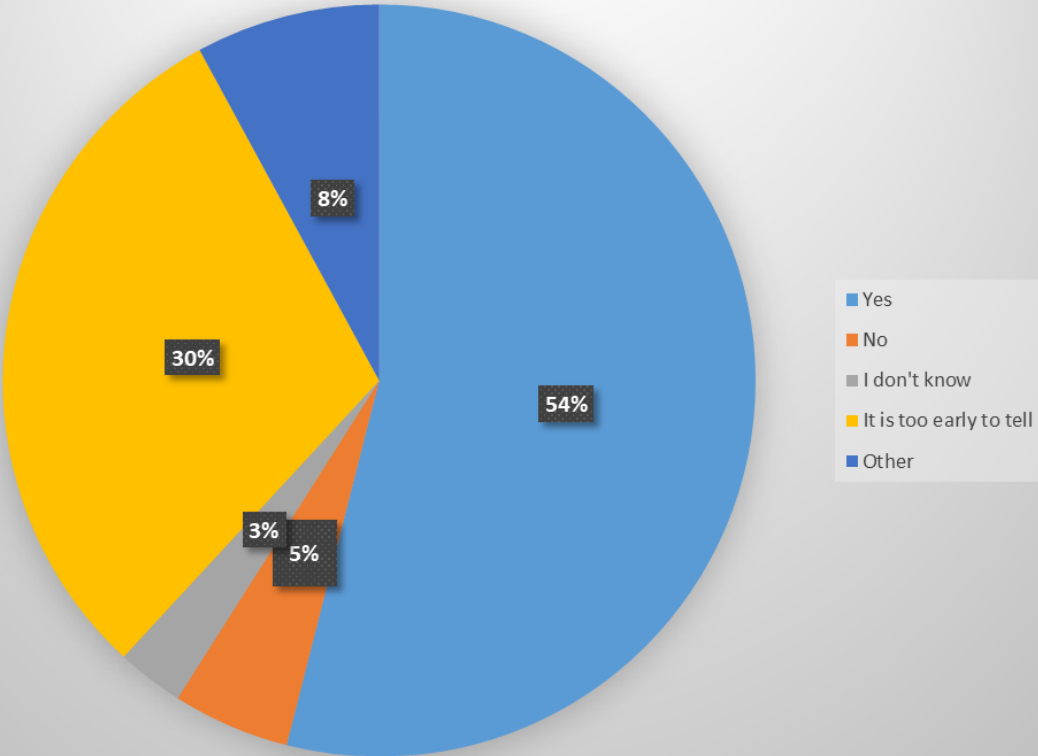
- Yes
- No

* 5. Now, I'd like to ask you the same question, but give you some different options for answering. Thinking about the current restrictions imposed wherever you are at the time you are doing this survey, do you think they are effective in keeping your community safe:

- Yes
- No
- I don't know
- It is too early to tell
- Other (please specify)

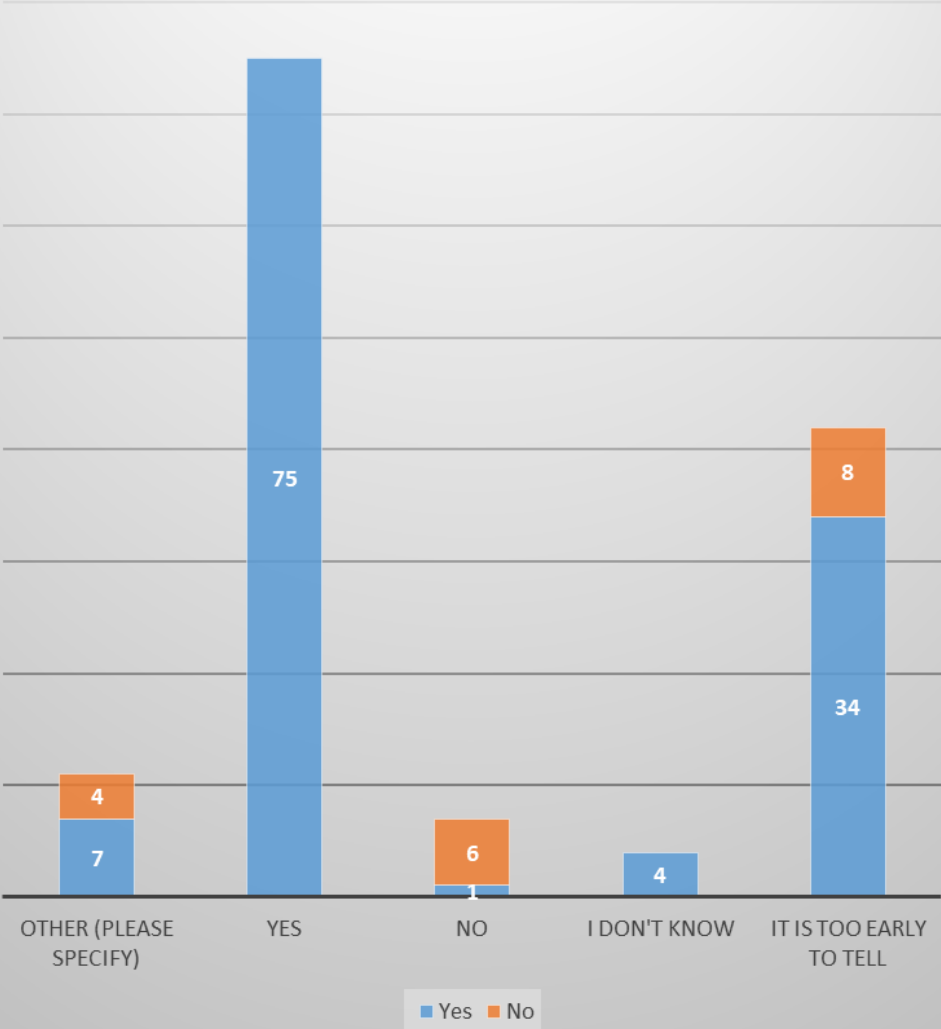
Do you think restrictions are effective in keeping your community safe?

Do you think restrictions are effective in keeping your community safe?

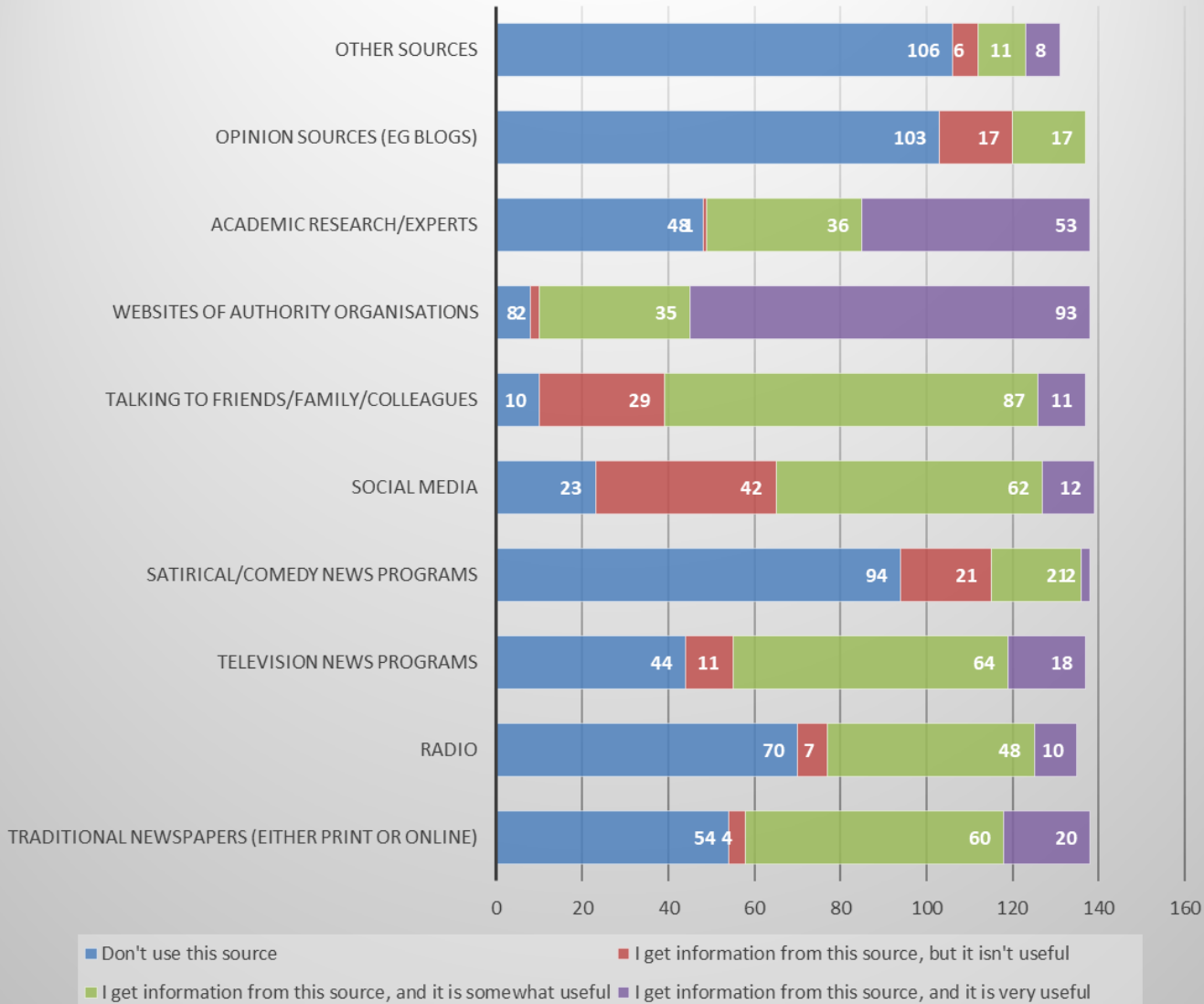


■ Yes
■ No

Do you think restrictions are effective in keeping your community safe?



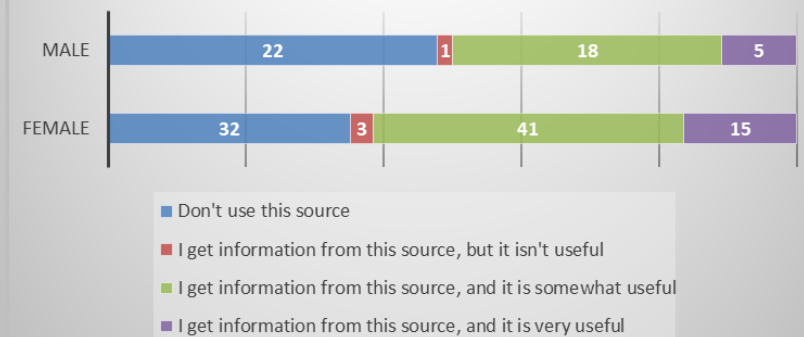
How useful are different sources of information?



WHAT MAKES DATA IMPORTANT TO A RESEARCHER?

- The question you start with
- Other interesting patterns

How useful men and women find traditional newspapers



A VERY QUICK LESSON IN STATISTICAL SIGNIFICANCE

Count		What is your gender?		Total
		Female	Male	
Social media	Don't use this source	10	13	23
	I get information from this source, but it isn't useful	27	14	41
	I get information from this source, and it is somewhat useful	45	17	62
	I get information from this source, and it is very useful	10	2	12
Total		92	46	138

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	8.053 ^a	3	.045
Likelihood Ratio	7.892	3	.048
Linear-by-Linear Association	7.243	1	.007
N of Valid Cases	138		

We look for “statistical significance” to tell us if there is a pattern between different variables.

- Reflects *probability* of something meaningful (rather than just occurred by *chance*)
- Reported as *p value* (probability value)
- Psychologists accept results when *probability of less than 5% ($p < .05$)*
- Smaller *p value* = more certain of results

In this example, I ran a Chi Square Test of association.

THE WAY WE REPORT DATA ALSO CHANGES THE IMPRESSION WE GIVE

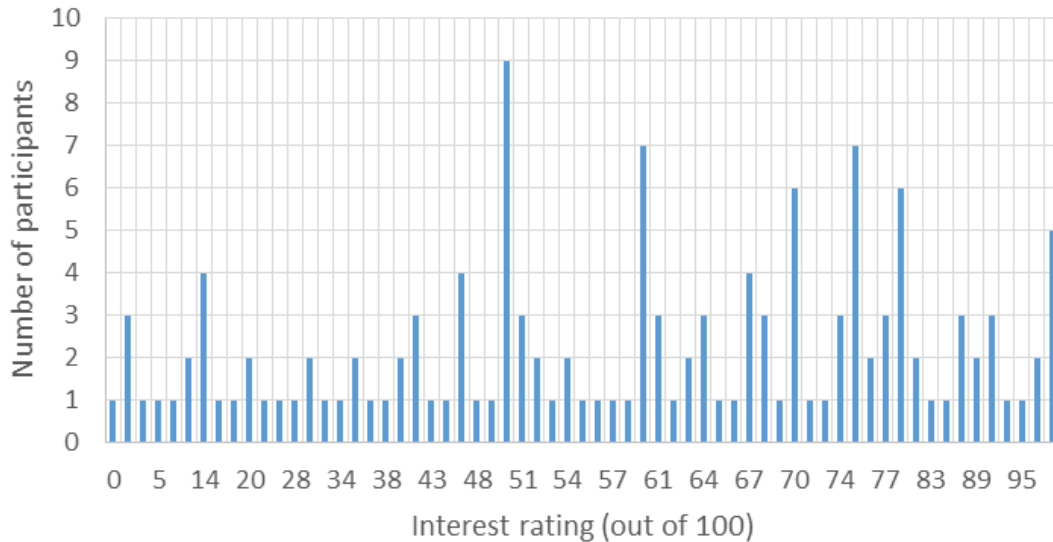
10. On the following scale, please indicate how interesting you found participating in this survey

Very boring Very interesting



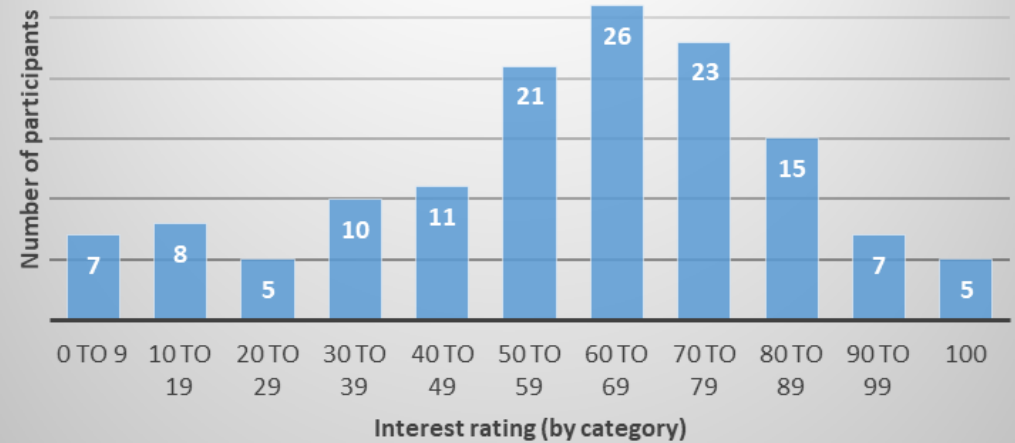
The image shows a horizontal slider scale. On the left side, there is a white circle with a grey border, representing the current selection. The text 'Very boring' is positioned to the left of this circle. On the right side, there is a grey square, representing the maximum value. The text 'Very interesting' is positioned to the left of this square. A horizontal grey line connects the circle to the square, with a white dot at the circle's position.

How interesting was the survey?



THE WAY WE REPORT DATA
ALSO CHANGES THE
IMPRESSION WE GIVE

How interesting was the survey?



Descriptive Statistics

	N	Min	Max	Mean	Std. Deviation
How interesting you found participating in this survey	138	0	100	57.50	24.925
Valid N (listwise)	138				

WHAT TO DO WITH FINDINGS?

- REPORTING RESULTS IN AN ACADEMIC CONTEXT
- MEDIA – HEADLINES AND DETAILS

Dear News Media,

When reporting poll results, please keep in mind the following suggestions:

1. If two poll numbers differ by less than the margin of error, it's not a news story.
2. Scientific facts are not determined by public opinion polls.
3. A poll taken of your viewers/internet users is not a scientific poll.
4. What if all polls included the option "Don't care"?

Signed,
-Someone who took a basic statistics course.

JORGE CHAM © 2010
WWW.PHDCOMICS.COM

WHERE HAS THE DATA COME FROM?


The Sydney Morning Herald

Such wet markets are notorious breeding grounds for disease as stressed animals from all corners of the world are caged close together, trading unfamiliar diseases, and then handled and butchered by humans. It happened that way with SARS, the first deadly coronavirus outbreak of the modern era, and then again with the second, MERS, this time along camel trade routes and slaughterhouses in the Middle East.

So why does the bioweapon rumour persist? Some of the trouble started with a draft paper, which was not peer reviewed, claiming to find traces of HIV artificially inserted into the new virus's genetic code. The work was quickly [withdrawn by its authors](#) after it was [debunked for using incorrect data but](#) not before it was seized upon by blogs and right-wing pundits as "proof" of manipulation, eventually finding its way into a US senator's talking points and Fox News bulletins.

President Donald Trump has said he is confident the Wuhan lab was connected to the outbreak, using public White House briefings on the

By the time we get data, it has usually come via various channels, all of which have interpreted it for their own purposes.

Participant  researcher(s)  peer reviewed article

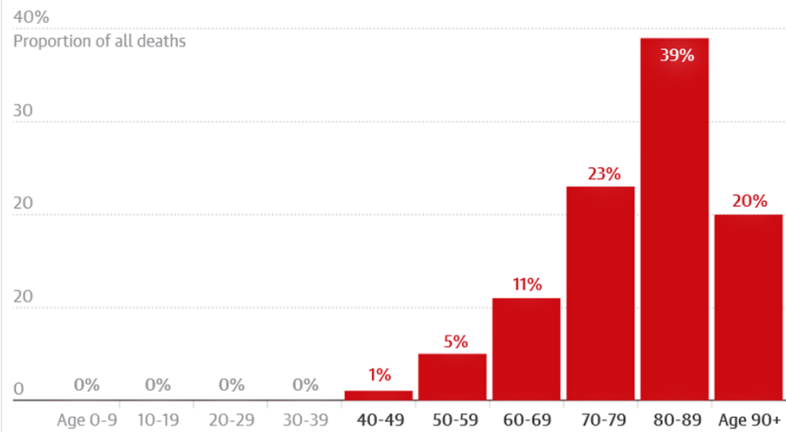
 media office  government/funding body  social media  traditional media  you

READ THE FINE PRINT

Pensioners 34 times more likely to die of Covid-19 than working age Brits, data shows

Death rate among working population is relatively low but risk varies by gender, ethnicity, class and 'exposure to people'

Older people have a greater risk of dying from Covid-19, ONS figures show

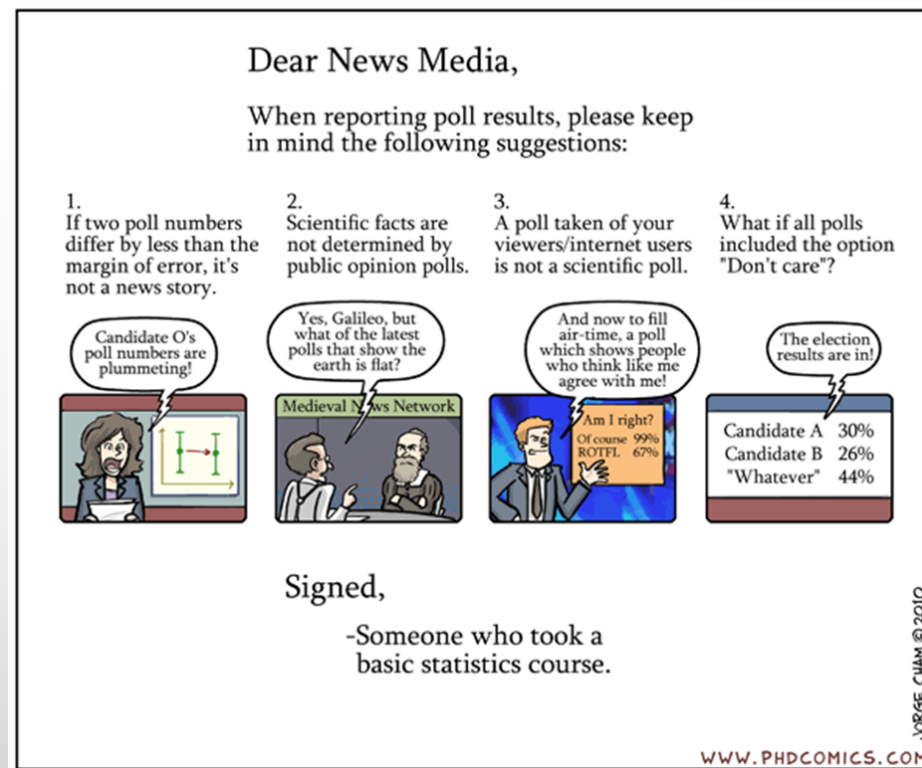


Guardian graphic. Source: ONS

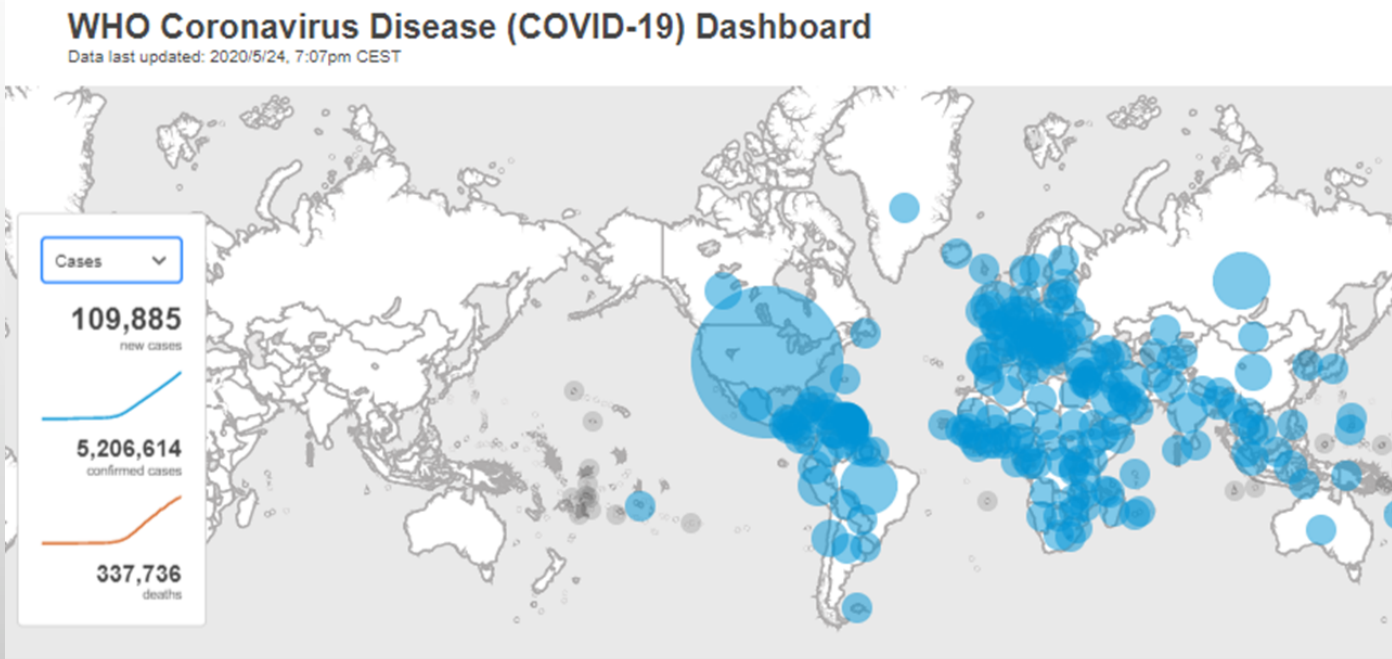
- The source – The Guardian
- What is the context?
- What is the big idea?
- Are there any hidden nuggets of information?
- Is this verifiable?

WHAT TO DO WITH FINDINGS?

- REPORTING RESULTS IN AN ACADEMIC CONTEXT
- MEDIA – HEADLINES AND DETAILS
- GRAPHS



GRAPHS AND FIGURES CONDENSE LOTS OF INFORMATION

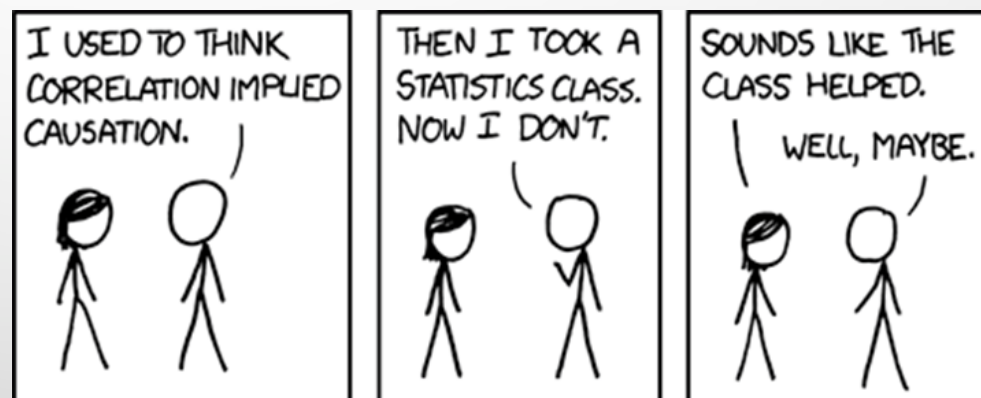


Globally, as of 7:07pm CEST, 24 May 2020, there have been 5,206,614 confirmed cases including 337,736 deaths, reported to WHO.

Date reported	Country	Country	WHO_re	New_ca	Cumulat	New_de	Cumulati
2020-01-25T00:00:00	AU	Australia	WPRO	3	3	0	0
2020-01-26T00:00:00	AU	Australia	WPRO	1	4	0	0
2020-01-27T00:00:00	AU	Australia	WPRO	1	5	0	0
2020-01-28T00:00:00	AU	Australia	WPRO	0	5	0	0
2020-01-29T00:00:00	AU	Australia	WPRO	2	7	0	0
2020-01-30T00:00:00	AU	Australia	WPRO	2	9	0	0
2020-01-31T00:00:00	AU	Australia	WPRO	0	9	0	0
2020-02-01T00:00:00	AU	Australia	WPRO	3	12	0	0
2020-02-02T00:00:00	AU	Australia	WPRO	0	12	0	0
2020-02-03T00:00:00	AU	Australia	WPRO	0	12	0	0
2020-02-04T00:00:00	AU	Australia	WPRO	1	13	0	0
2020-02-05T00:00:00	AU	Australia	WPRO	1	14	0	0
2020-02-06T00:00:00	AU	Australia	WPRO	1	15	0	0
2020-02-07T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-08T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-09T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-10T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-11T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-12T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-13T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-14T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-15T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-16T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-17T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-18T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-19T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-20T00:00:00	AU	Australia	WPRO	0	15	0	0
2020-02-21T00:00:00	AU	Australia	WPRO	2	17	0	0
2020-02-22T00:00:00	AU	Australia	WPRO	4	21	0	0
2020-02-23T00:00:00	AU	Australia	WPRO	1	22	0	0
2020-02-24T00:00:00	AU	Australia	WPRO	0	22	0	0
2020-02-25T00:00:00	AU	Australia	WPRO	1	23	0	0
2020-02-26T00:00:00	AU	Australia	WPRO	0	23	0	0
2020-02-27T00:00:00	AU	Australia	WPRO	0	23	0	0
2020-02-28T00:00:00	AU	Australia	WPRO	0	23	0	0
2020-02-29T00:00:00	AU	Australia	WPRO	0	23	0	0

BEWARE OF CORRELATIONS

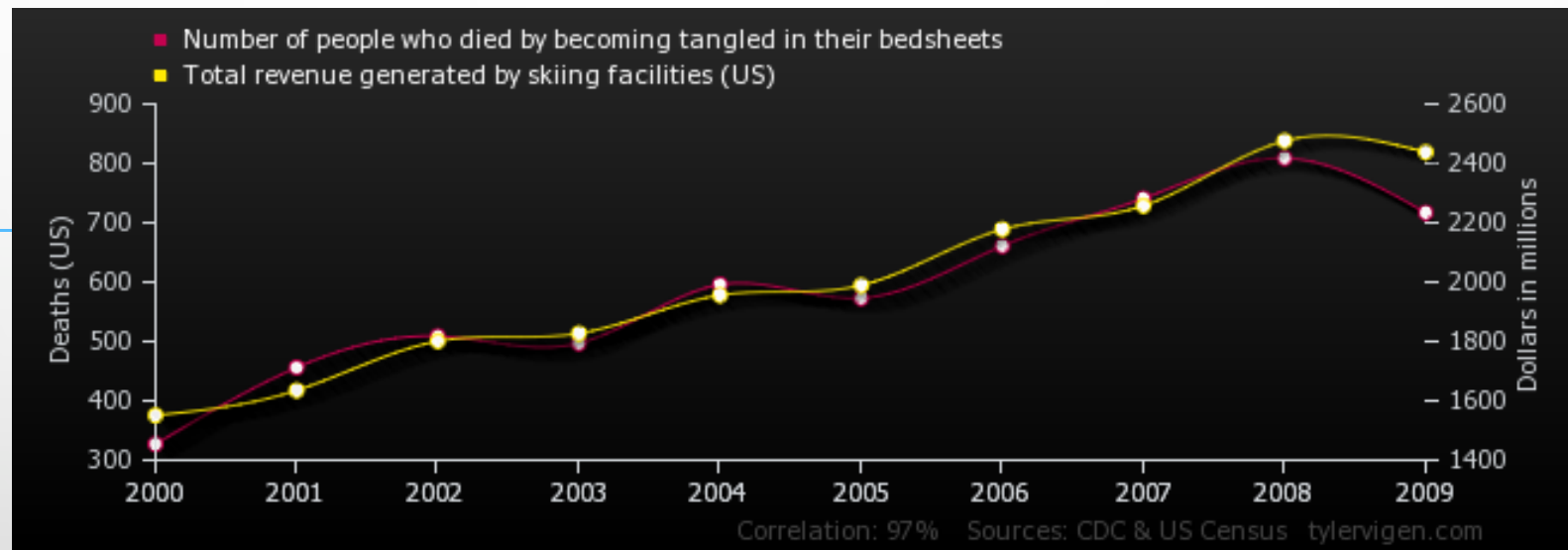
Correlation \neq Causation



Number of people who died by becoming tangled in their bedsheets

correlates with

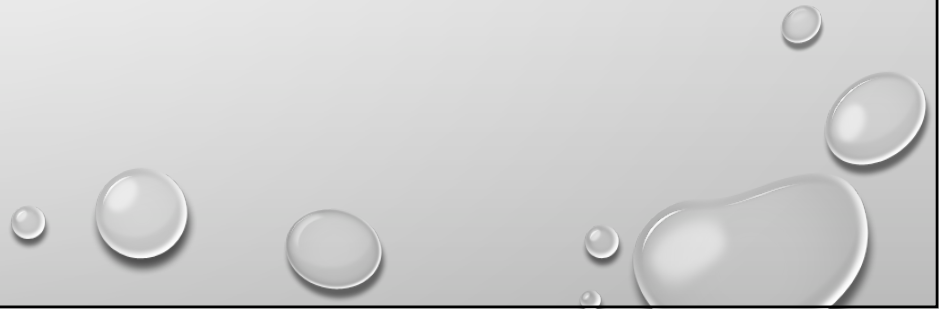
Total revenue generated by skiing facilities (US)



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
<i>Number of people who died by becoming tangled in their bedsheets</i> Deaths (US) (CDC)	327	456	509	497	596	573	661	741	809	717
<i>Total revenue generated by skiing facilities (US)</i> Dollars in millions (US Census)	1,551	1,635	1,801	1,827	1,956	1,989	2,178	2,257	2,476	2,438
Correlation: 0.969724										



QUESTIONS TO ASK WHEN YOU READ ABOUT DATA

- How many stages, or reinterpretations, has the data been through?
 - Is it traceable?
 - If I wanted to, could I find out the how the data was gathered, and how findings were arrived at?
 - Where can I go to verify this information?
 - What is the goal of the source I'm looking at?
 - Commercial, public information, getting votes, winning an argument ...
 - Is there something here I can, or should, act on?
- 

EXAMPLE 1



Advocating for freedom
from violence

MEDIA RELEASE

3 April 2020 – For Immediate Release

New Domestic Violence Survey in NSW Shows Impact of COVID-19 on the Rise

Women's Safety NSW is surveying its members who are frontline women's domestic violence workers across NSW on a weekly basis to map the impact COVID-19 is having on themselves as workers, their services and the women, children and families they are supporting.

This week's results have just come in and reveal the impact of COVID-19 is on the rise with a higher percentage of frontline domestic violence workers reporting:

- An increase in client numbers overall (50% up from 41% last week)
- Increasing complexity of client needs (75% up from 70.8% last week)
- Escalating or worsening violence (50% up from 44.9% last week)
- Violence specifically relating to COVID-19 (50% up from 36.2% last week)
- A sharp increase in violence being reported for the first time (47.5% up from 15.9% last week).

https://www.womenssafetynsw.org.au/wp-content/uploads/2020/04/03.04.20_New-Domestic-Violence-Survey-in-NSW-Shows-Impact-of-COVID-19-on-the-Rise.pdf

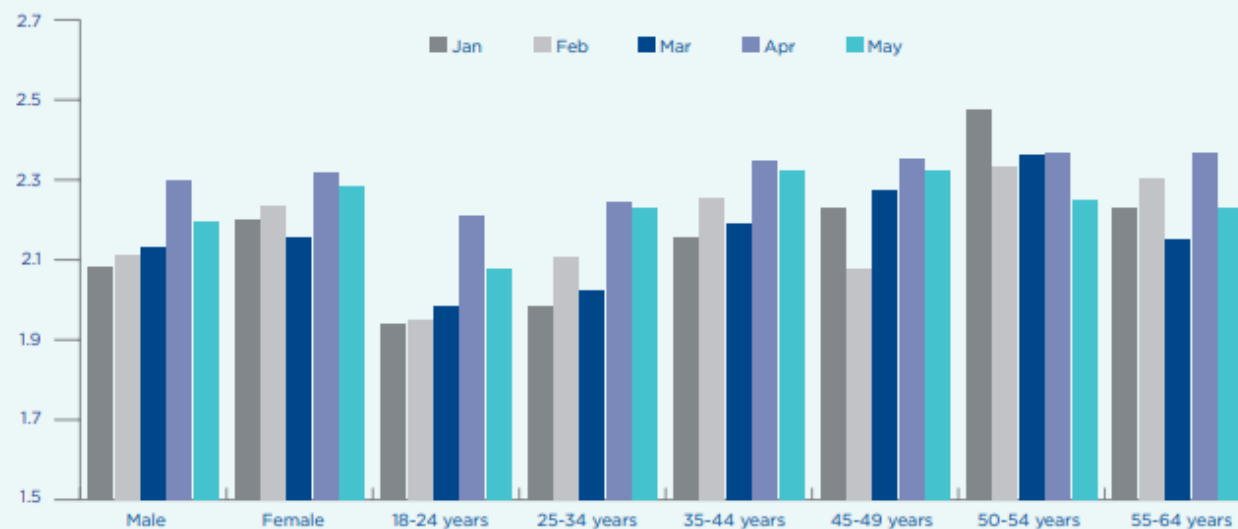
Did JobSeeker and JobKeeper achieve their aims?

Supporting the economic recovery from the COVID-19 pandemic by protecting employment and income through targeted government policy.

https://melbourneinstitute.unimelb.edu.au/data/assets/pdf_file/0004/3396415/ri2020n12.pdf

EXAMPLE 2

Figure 5: Average score of comparison of current financial situation to 12 months ago, by gender and age



Note: A higher score indicates fewer families are better off and more families are worse off.

Source: Authors' calculations based on 6,000 observations from CASIE Survey data for the months January to May 2020. The sample is weighted to be representative of the Australian population on gender, age and location.

QUESTIONS

