Introduction to Research Design

Francisco Villamil

Research Design for Social Sciences

MA Computational Social Science, UC3M

Fall 2025

Introduction

- What is this course about?
- Expectations?

Introduction

- What is this course about?
- Expectations?

- **What is research? And why do we need to design it?**

A real example

- Racing team deciding whether to race or not in the last session

A real example

- Racing team deciding whether to race or not in the last session
- Car engine has blown out in 7 out of 24 past races

A real example

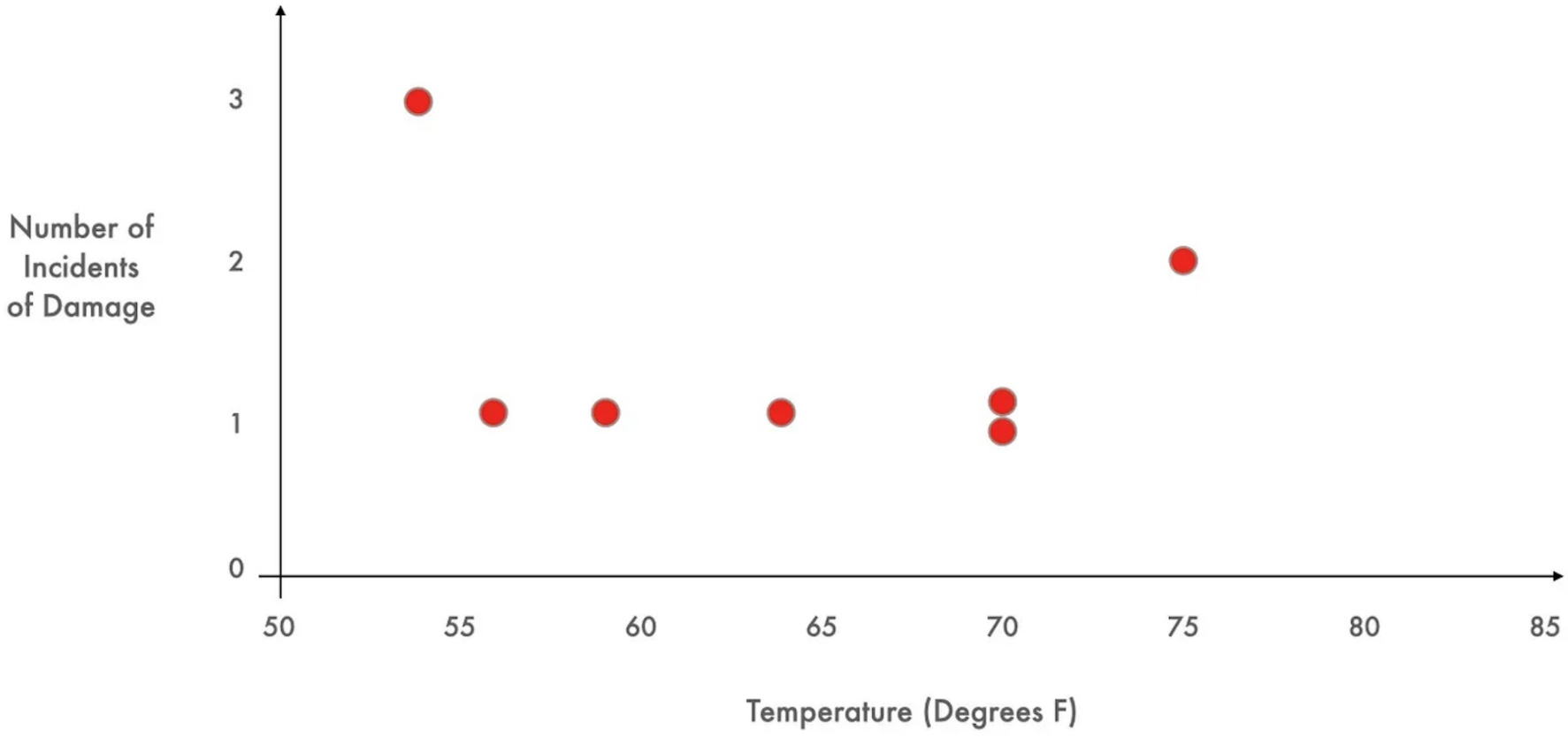
- Racing team deciding whether to race or not in the last session
- Car engine has blown out in 7 out of 24 past races
  - $\rightarrow$ Should we risk an explosion and go bankrupt or race?

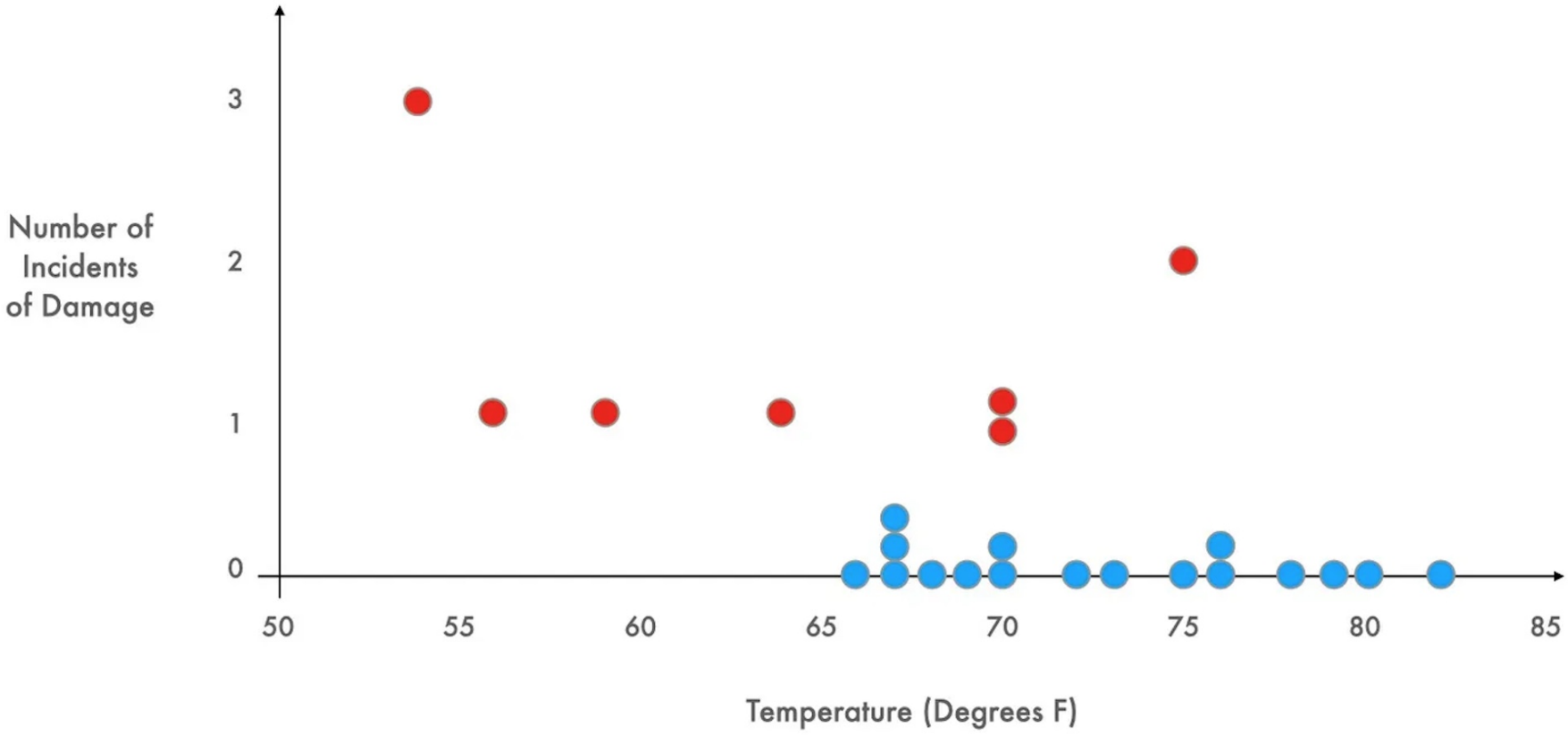
A real example

- Racing team deciding whether to race or not in the last session
- Car engine has blown out in 7 out of 24 past races
  - $\rightarrow$ Should we risk an explosion and go bankrupt or race?
- A mechanic has a last-minute gut feeling that it might have to do with temperature (forecast for race: very cold day)

A real example

- Racing team deciding whether to race or not in the last session
- Car engine has blown out in 7 out of 24 past races
  - $\rightarrow$ Should we risk an explosion and go bankrupt or race?
- A mechanic has a last-minute gut feeling that it might have to do with temperature (forecast for race: very cold day)
- Someone says: "show me the data of the past failures!"


Number of Incidents of Damage
3
2
1
0
50
55
60
65
70
75
80
85
Temperature (Degrees F)


Number of Incidents of Damage
3
2
1
0
50
55
60
65
70
75
80
85
Temperature (Degrees F)

Space Shuttle *Challenger* disaster

From Wikipedia, the free encyclopedia

Coordinates: 28°38′24″N 80°16′48″W

On January 28, 1986, the Space Shuttle *Challenger* broke apart 73 seconds into its flight, killing all seven crew members aboard. The spacecraft disintegrated 46,000 feet (14 km) above the Atlantic Ocean, off the coast of Cape Canaveral, Florida, at 11:39 a.m. EST (16:39 UTC). It was the first fatal accident involving an American spacecraft while in flight.

The mission, designated STS-51-L, was the tenth flight for the orbiter and the twenty-fifth flight of the Space Shuttle fleet. The crew was scheduled to deploy a communications satellite and study Halley's Comet while they were in orbit, in addition to taking school teacher Christa McAuliffe into space. The latter resulted in a higher than usual media interest and coverage of the mission; the launch and subsequent disaster were seen live in many schools across the United States.

The cause of the disaster was the failure of the primary and secondary redundant O-ring seals in a joint in the shuttle's right solid rocket booster (SRB). The record-low temperatures on the morning of the launch had stiffened the rubber O-rings, reducing their ability to seal the joints. Shortly after liftoff, the seals were breached, and hot pressurized gas from within the SRB leaked through the joint and burned through the aft attachment strut connecting it to the external propellant tank (ET), then into the tank itself. The collapse of the ET's internal structures and the rotation of the SRB that followed threw the shuttle

Space Shuttle *Challenger* disaster

Challenger's solid rocket boosters fly uncontrollably after the breakup of the external tank separated them from the shuttle stack. The remains of the orbiter and tank leave thin white contrails as they fall toward the Atlantic Ocean.

Challenger example

- Explanatory research
- What is the role of design here?

Challenger example

- Explanatory research
- What is the role of design here?
- Which observations to use
- What *variation* we need to answer the question?

Types of research

- *Theoretical* and *empirical* research

Types of research

- *Theoretical* and *empirical* research
- *Qualitative* and *quantitative* empirical research

Types of research

- *Theoretical* and *empirical* research
- *Qualitative* and *quantitative* empirical research
  - → Descriptive vs explanatory

Empirical Research

- Goal: answer a question using empirical evidence
  - → Usual problems: unanswerable questions, wrong data to question, concepts do not correspond to measurements...

Empirical Research

- Goal: answer a question using empirical evidence
  - → Usual problems: unanswerable questions, wrong data to question, concepts do not correspond to measurements...
- How? exploiting **empirical variation**
- Research design is essentially knowing where to look at, how to measure it, how to analyze it, how to interpret it, etc –

  it's about inference

Empirical Research

- Goal: answer a question using empirical evidence
  - → Usual problems: unanswerable questions, wrong data to question, concepts do not correspond to measurements...
- How? exploiting **empirical variation**
- Research design is essentially knowing where to look at, how to measure it, how to analyze it, how to interpret it, etc –

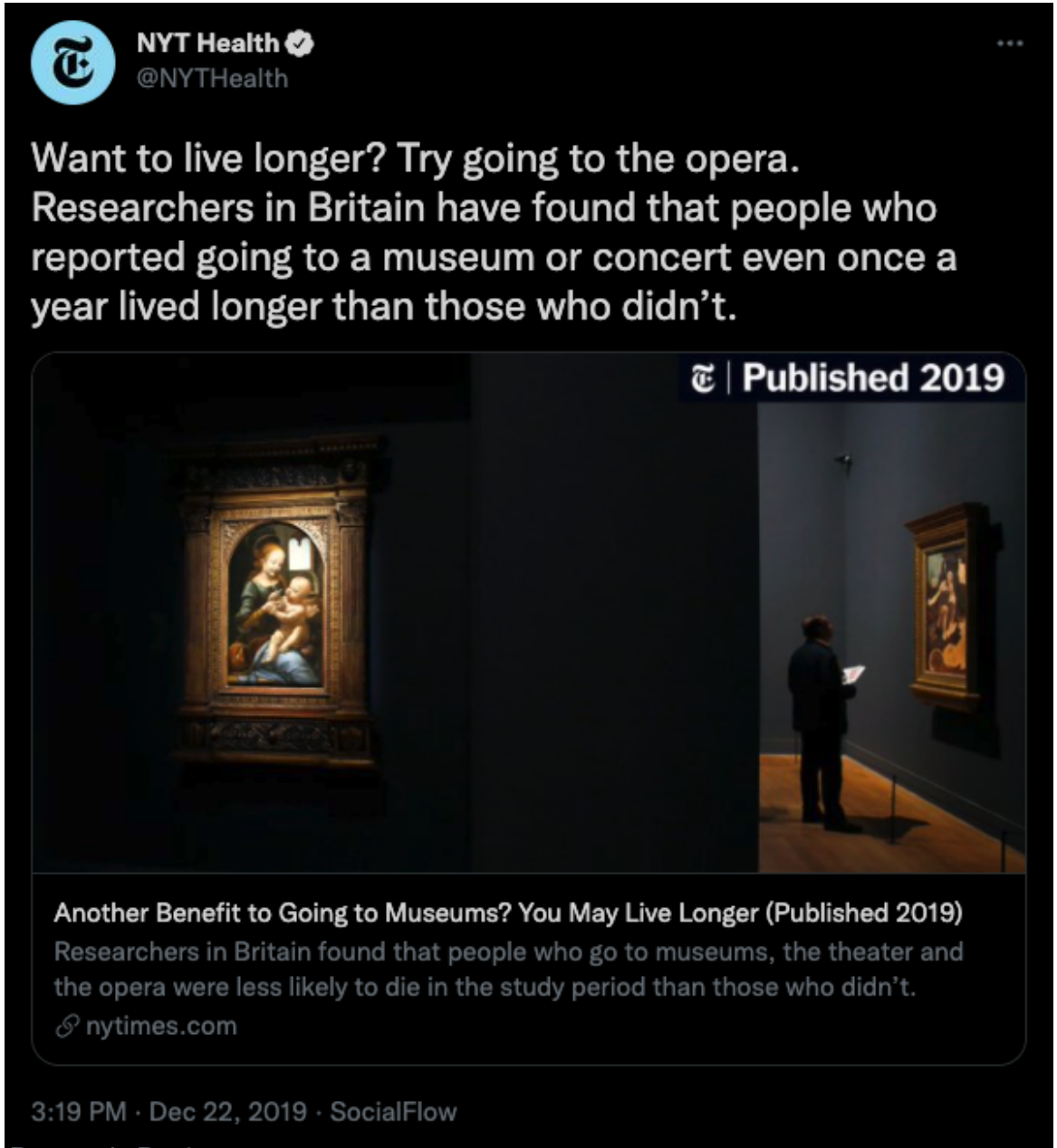
  it's about inference
- Example:
  - → Hotel interested in increasing bookings
  - → Data used: online bookings and actual visits

Empirical research

- Usually includes many things we vaguely associate with "data"
  - → Statistical analysis
  - → Experiments
  - → Simulations, cases without variation?
  - → etc
- Idea is learning about **making claims**, or how can we learn from the observable world the right way
- Two things we should learn:

  how to answer questions & and how to evaluate answers

Empirical evidence and claims



Empirical evidence and claims

- Let's reverse-engineer this:
  - → What did they probably do?
- How would you answer the question better?

Empirical evidence and claims

- Let's reverse-engineer this:
  - → What did they probably do?
- How would you answer the question better?
- Causal inference and experimental methods
  - → Observation vs manipulation

Empirical evidence and claims

- Let's reverse-engineer this:
  - → What did they probably do?
- How would you answer the question better?
- Causal inference and experimental methods
  - → Observation vs manipulation
- Gold standard?

REPORT

Analytic Thinking Promotes Religious Disbelief

WILL M. GERVAIS AND ARA NORENZAYAN Authors Info & Affiliations

SCIENCE · 27 Apr 2012 · Vol 336, Issue 6080 · pp. 493-496 · DOI: 10.1126/science.1215647

Overthinking Religion?

Many theories of human cognition make a distinction between System I, which tends to be rapid and to rely on heuristics or rules of thumb, and System II, which tends to be more deliberative and analytic. This dual-process framework, within which both processes may operate simultaneously and competitively, has been used to explain a variety of situational influences upon decision-making. **Gervais and Norenzayan** (p. 493) studied the application of a dual-process framework to religious disbelief and found that triggering analytic thinking processes through a variety of experimental manipulations resulted in a tendency for subjects to report lower levels of religious belief.

We adopted three complementary strategies to test for robustness and generality. First, study 1 tested whether individual differences in the tendency to engage analytic thinking are associated with reduced religious belief. Second, studies 2 to 5 established causation by testing whether various experimental manipulations of analytic processing, induced subtly and implicitly, encourage religious disbelief. These manipulations of analytic processing included visual priming, implicit priming, and cognitive disfluency (*18*, *19*). Third, across studies, we assessed religious belief using diverse measures that focused primarily on belief in and commitment to religiously endorsed supernatural agents. Samples consisted of participants from diverse cultural and religious backgrounds (*20*).

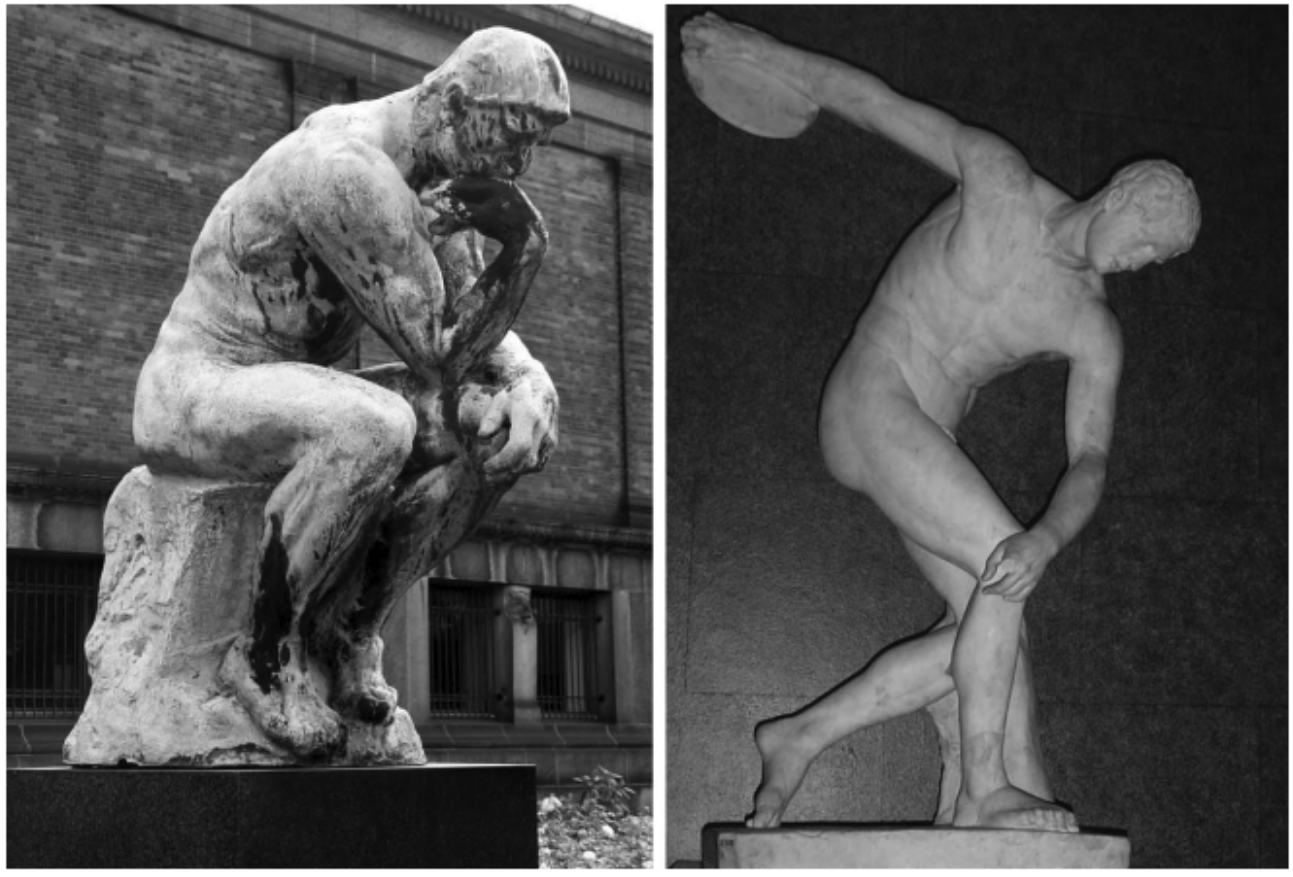


Fig. 1. Sample images of *The Thinker* **(left)** and *Discobolus* **(right)** used in study 2. The images shown here are similar to, but not the exact same ones used in the study. [Source: *Wikimedia*]

ipants only very rarely detected a connection between manipulations and religious belief measures (*20*).

Study 2 used a visual priming paradigm in which a sample of Canadian undergraduates rated their belief in God (from 0 to 100) after being randomly assigned to view four images (samples provided in Fig. 1) of either artwork depicting a reflective thinking pose (Rodin's *The Thinker*; $N = 26$) or control artwork matched for surface characteristics like color and posture (*Discobolus* of Myron; $N = 31$). A pilot test with different participants ($N = 40$) revealed that this novel priming procedure significantly improved performance on a syllogistic reasoning task that measures analytic tendencies (*20*). In the present study, as hypothesized, viewing *The Thinker* significantly promoted religious

Opinions about claims?

Table 2. Summary of experimental methods and findings (studies 2 to 5). *d* reflects effect size estimates (Cohen's *d*).

Study Belief measure (possible range)	Condition: sample stimuli	*N*	*M*	SD	*t*	*P*	*d*
2: Art	Control: *Discobolus*	31	61.55	35.68	2.24	0.03	0.59
Belief in God (0–100)	Analytic: *The Thinker*	26	41.42	31.47			
3: Implicit	Control: hammer, shoes, jump, retrace, brown	43	12.65	5.29	2.11	0.04	0.44
Supernatural agents (3–21)	Analytic: think, reason, analyze, ponder, rational	50	10.12	6.13			
4: Implicit	Control: hammer, shoes, jump, retrace, brown	75	40.16	16.73	2.20	0.03	0.36
Intrinsic religiosity (10–70)	Analytic: think, reason, analyze, ponder, rational	70	34.39	14.77			
5: Disfluency	Control: sample font	88	12.16	5.99	2.06	0.04	0.31
Supernatural agents (3–21)	Analytic: *sample font*	91	10.40	5.44			

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer[1,16], Anna Dreber[2,16], Felix Holzmeister[3,16], Teck-Hua Ho[4,16], Jürgen Huber[3,16], Magnus Johannesson[2,16], Michael Kirchler[3,5,16], Gideon Nave[6,16], Brian A. Nosek[7,8,16*], Thomas Pfeiffer[9,16], Adam Altmejd[2], Nick Buttrick[7,8], Taizan Chan[10], Yiling Chen[11], Eskil Forsell[12], Anup Gampa[7,8], Emma Heikensten[2], Lily Hummer[8], Taisuke Imai[13], Siri Isaksson[2], Dylan Manfredi[6], Julia Rose[3], Eric-Jan Wagenmakers[14] and Hang Wu[15]

Being able to replicate scientific findings is crucial for scientific progress[1–15]. We replicate 21 systematically selected experimental studies in the social sciences published in *Nature* and *Science* between 2010 and 2015[16–36]. The replications follow analysis plans reviewed by the original authors and pre-registered prior to the replications. The replications are high powered, with sample sizes on average about five times higher than in the original studies. We find a significant effect in the same direction as the original study for 13 (62%) studies, and the effect size of the replications is on average about 50% of the original effect size. Replicability varies between 12 (57%) and 14 (67%) studies for complementary replicability indicators. Consistent with these results, the estimated true-

a significant effect in the same direction as the original studies for 61% of replications[13]. Both the RPP and the EERP had high statistical power to detect the effect sizes observed in the original studies. However, the effect sizes of published studies may be inflated even for true-positive findings owing to publication or reporting biases[40–42]. As a consequence, if replications were well powered to detect effect sizes smaller than those observed in the original studies, replication rates might be higher than those estimated in the RPP and the EERP.

We provide evidence about the replicability of experimental studies in the social sciences published in the two most prestigious general science journals, *Nature* and *Science* (the Social Sciences Replication Project (SSRP)). Articles published in these

‘Your view of human nature will change profoundly as you read this brilliant book.’

DANIEL KAHNEMAN
Author of *Thinking, Fast and Slow*

THE MARSHMALLOW TEST

Understanding self-control and how to master it

WALTER MISCHEL

Why Rich Kids Are So Good at the Marshmallow Test

Affluence—not willpower—seems to be what's behind some kids' capacity to delay gratification.

By Jessica McCrory Calarco

Watts and his colleagues were skeptical of that finding. The original results were based on studies that included fewer than 90 children—all enrolled in a preschool on Stanford's campus. In restaging the experiment, Watts and his colleagues thus adjusted the experimental design in important ways: The researchers used a sample that was much larger—more than 900 children—and also more representative of the general population in terms of race, ethnicity, and parents' education. The researchers also, when analyzing their test's results, controlled for certain factors—such as the income of a child's household—that might explain children's ability to delay gratification and their long-term success.

Ultimately, the new study finds limited support for the idea that being able to delay gratification leads to better outcomes. Instead, it suggests that the capacity to hold out for a second marshmallow is shaped in large part by a child's social and economic background—and, in turn, that that background, not the ability to delay gratification, is what's behind kids' long-term success.

Other types of inference

- We are used to cases where we compare many observations with different values on key variables
- What if we don't have enough observations to answer a question?
- Any idea about examples?

Europe

Violent storm kills six on Corsica as island raises new alert

By **Marc Angrand** and **Benoit Van Overstraeten**

August 18, 2022 8:34 PM GMT+2 · Updated a year ago

PARIS, Aug 18 (Reuters) - A violent and unexpected storm battered the French Mediterranean island of Corsica on Thursday, killing at least six people including a teenage girl, and meteorologists predicted more bad weather to come.

Hail, heavy rain and winds peaking at 224 km per hour (140 mph) swept the island early in the day. Two of the victims were killed when trees fell in campsites.

"Storms formed at sea will affect large parts of the western Corsica coast throughout the night from Thursday to Friday," Meteo France forecaster said.

Opinions? How do you think it's done?

Article Type: Research Article

Open access

View license

Anthropogenic Warming Had a Crucial Role in Triggering the Historic and Destructive Mediterranean Derecho in Summer 2022

Juan Jesús González-Alemán, Damián Insua-Costa, Eric Bazile, Sergi González-Herrero, Mario Marcello Miglietta, Pieter Groenemeijer, and Markus G. Donat

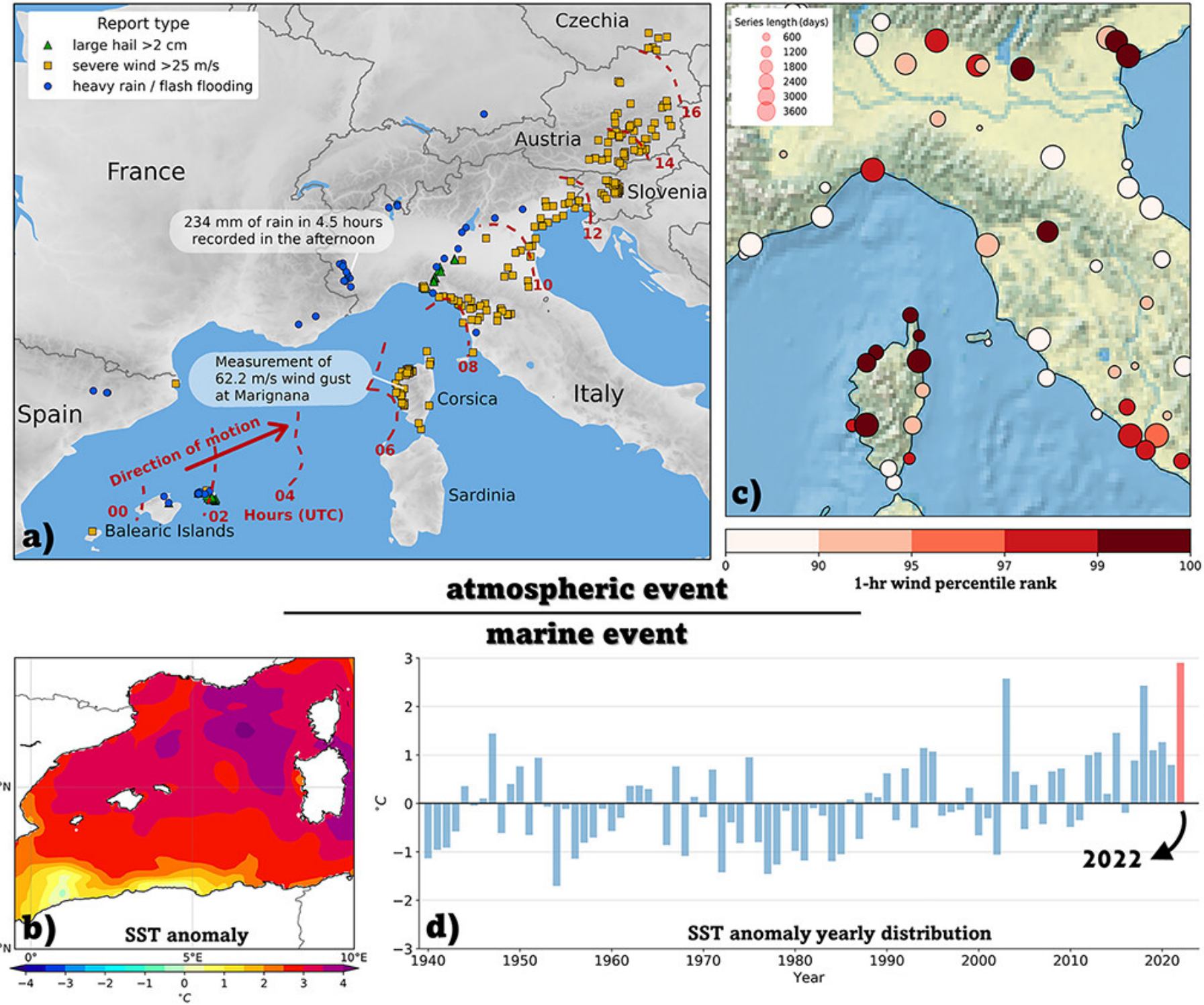
Online Publication: 31 Aug 2023

Print Publication: 01 Aug 2023

DOI: https://doi.org/10.1175/BAMS-D-23-0119.1

Page(s): E1526–E1532

Article History | Download PDF | © Get Permissions


Report type
large hail >2 cm
severe wind >25 m/s
heavy rain / flash flooding
Czechia
Austria
France
Slovenia
234 mm of rain in 4.5 hours recorded in the afternoon
Measurement of 62.2 m/s wind gust at Marignana
Spain
Corsica
Italy
Sardinia
Direction of motion
00
02
04
Hours (UTC)
06
08
10
12
14
16
Balearic Islands
a)
Series length (days)
600
1200
1800
2400
3000
3600
c)
0
90
95
97
99
100
1-hr wind percentile rank
atmospheric event
marine event
40°N
35°N
0°
5°E
10°E
−4
−3
−2
−1
0
1
2
3
4
°C
b)
SST anomaly
°C
3
2
1
0
−1
−2
−3
2022
d)
SST anomaly yearly distribution
1940
1950
1960
1970
1980
1990
2000
2010
2020
Year

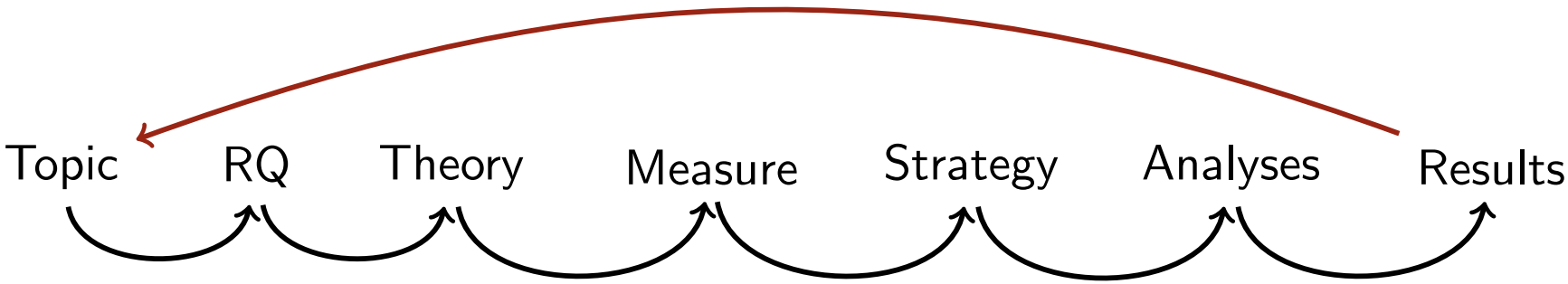
Not only about data or empirical evidence

- The "non-data" part is also important:

  thinking about an answerable question, a coherent argument, and how to connect it to data

- Also problematic because we all probably think we already know how it's done, but e.g. many (most?) final theses fail on this

Empirical Research process



Key ingredientes

In every case, when we talk about *quantitative empirical research*, we will usually work with **variation** across **observations** at a given **unit of analyses/observation**, from which we infer stuff

1. Observation

Key ingredientes

In every case, when we talk about *quantitative empirical research*, we will usually work with **variation** across **observations** at a given **unit of analyses/observation**, from which we infer stuff

1. Observation
2. Unit of analysis or unit of observation

Key ingredientes

In every case, when we talk about *quantitative empirical research*, we will usually work with **variation** across **observations** at a given **unit of analyses/observation**, from which we infer stuff

1. Observation
2. Unit of analysis or unit of observation
3. Variables

Key ingredientes

In every case, when we talk about *quantitative empirical research*, we will usually work with **variation** across **observations** at a given **unit of analyses/observation**, from which we infer stuff

1. Observation
2. Unit of analysis or unit of observation
3. Variables
4. Variation

A few questions on variation

- Why do we talk so much about *variation*?

Halley's Eclipse: a coup for Newtonian prediction and the selling of science

300 years ago, on the 3rd of May 1715, a rare solar eclipse occurred over England. It was an opportunity too good to miss for those promoting new astronomical theories – and their own careers.

40

A few questions on variation

- Why do we talk so much about *variation*?
- Critique of having "few observations"?

A few questions on variation

- Why do we talk so much about *variation*?
- Critique of having "few observations"?
- Why do we use statistics at all?

A few questions on variation

- Why do we talk so much about *variation*?
- Critique of having "few observations"?
- Why do we use statistics at all?
- What about experiments?

A few questions on variation

- Why do we talk so much about *variation*?
- Critique of having "few observations"?
- Why do we use statistics at all?
- What about experiments?
- Where's the variation in attribution science?

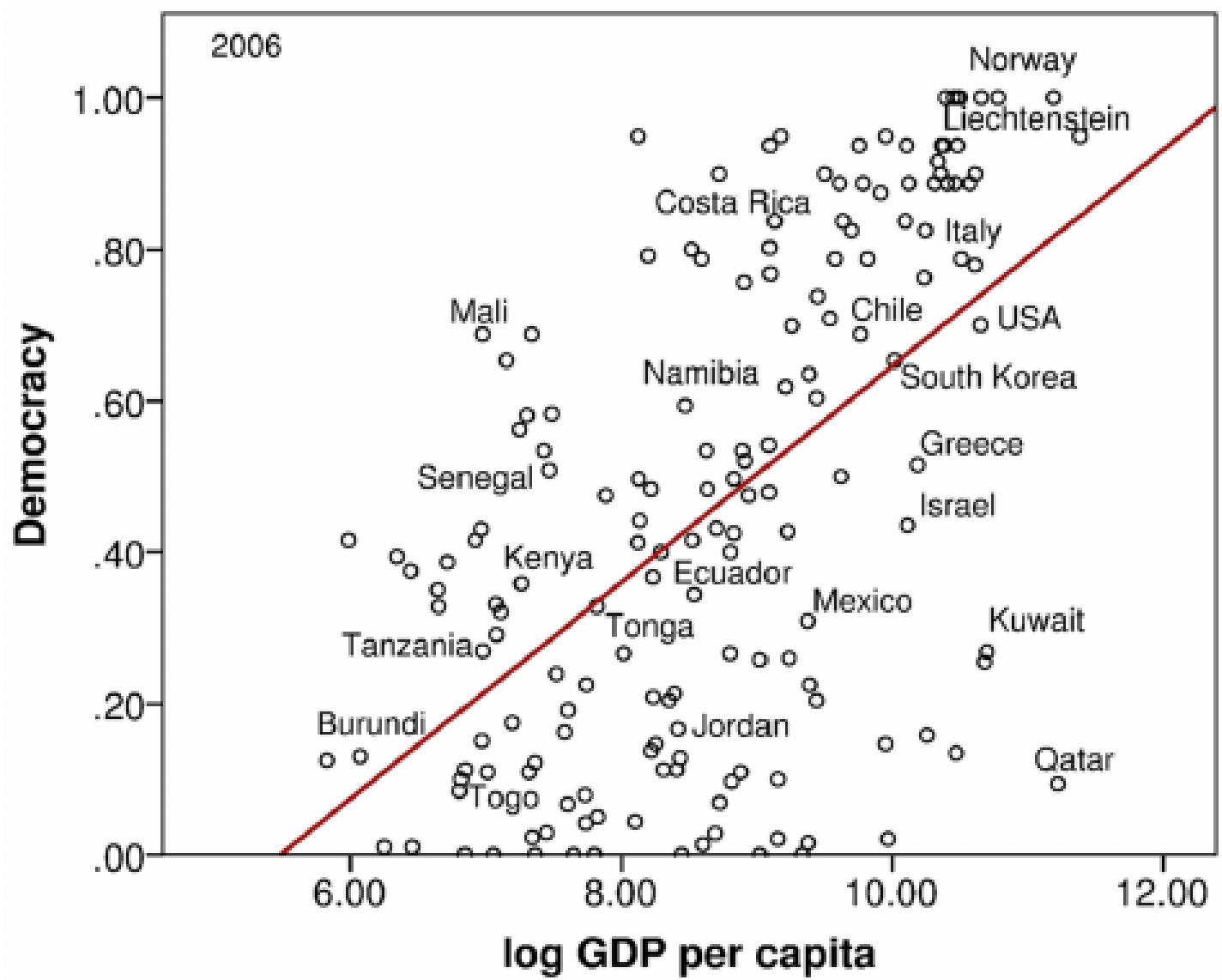
On the unit of analysis

- Why is this important?

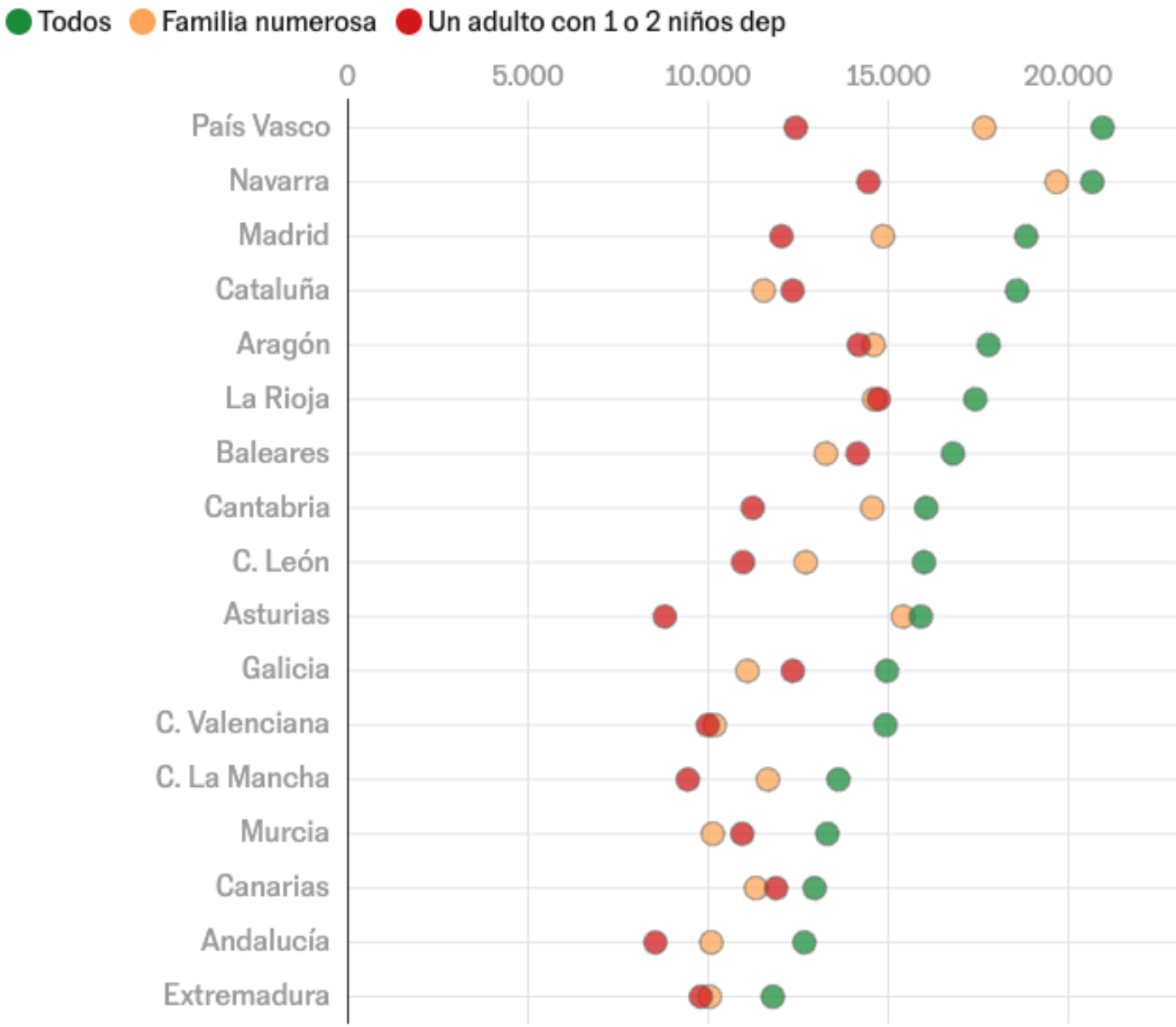
On the unit of analysis

- Why is this important?
- Able to quickly identify it?

What's the unit of analyses in the data behind this graph?



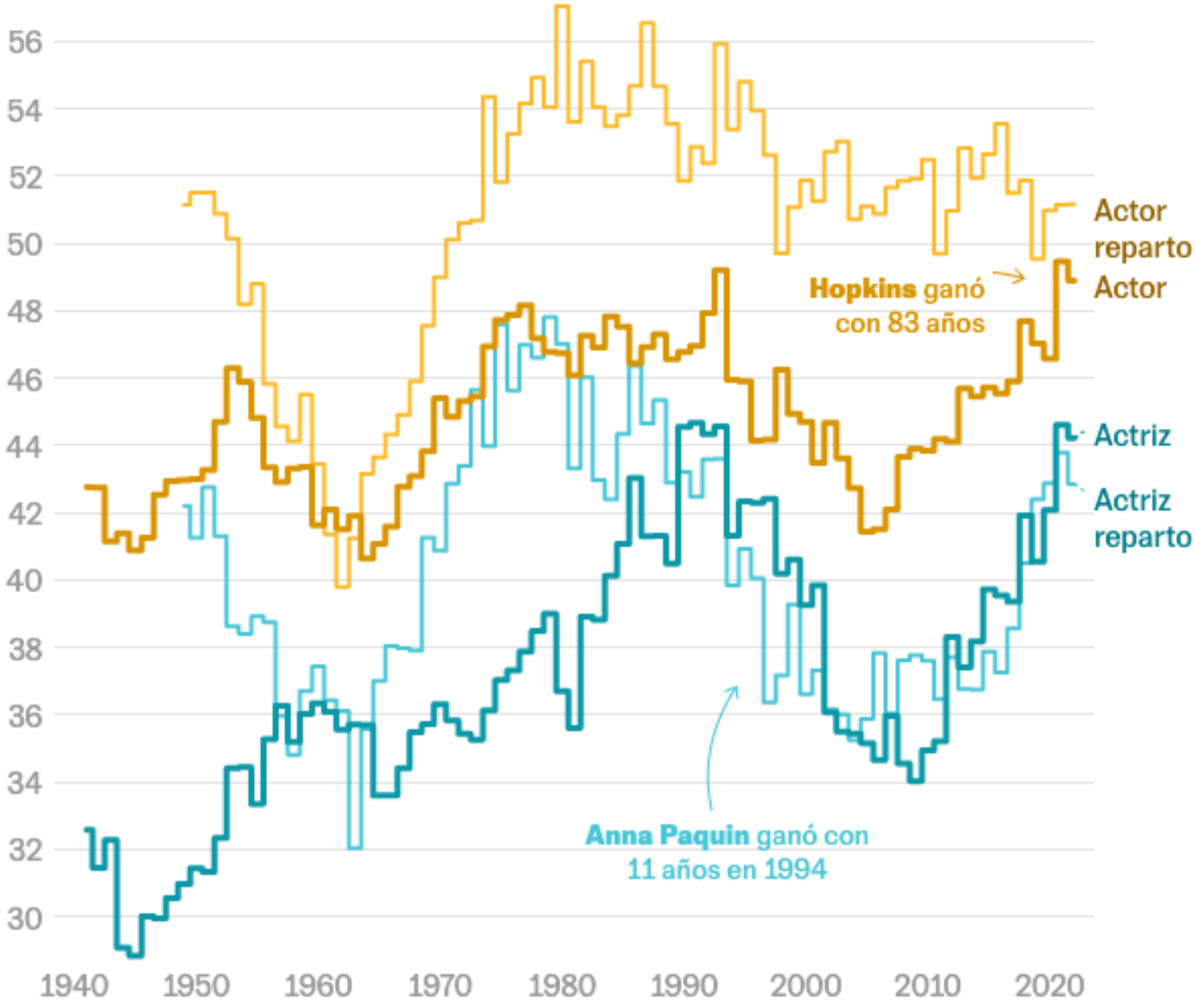
Renta mediana de por unidad de convivencia de cada tipo de familia en cada comunidad autónoma



Los datos de Extremadura, Aragón, Canarias, Cantabria, La Rioja y Asturias cuentan solo con entre 50 y 100 entrevistas

ECV 2019, 2020 y 2021 (INE) / EL PAÍS

Edad de los actores premiados (media móvil de 12 años)



Crowdflower y elaboración propia / EL PAÍS

On the unit of analysis

- Most empirical designs can be adapted to different units of analyses
- The unit you use is related to the question and theory
- Example: think again about the hotel that wants to increase bookings
  - → How many different units could you use and how are they related to different questions/theories?

On the unit of analysis

- Most empirical designs can be adapted to different units of analyses
- The unit you use is related to the question and theory
- Example: think again about the hotel that wants to increase bookings
  - → How many different units could you use and how are they related to different questions/theories?
- Another example: children educational attainment and contextual factors (peers, school, family, etc)

Reading

BOOKS

WHAT DATA CAN'T DO

When it comes to people—and policy—numbers are both powerful and perilous.

By Hannah Fry

March 22, 2021

Research process in detail (1)

- From **topic** (or problem) to the **research question**

Research process in detail (1)

> From **topic** (or problem) to the **research question**

- Difference between motivation and question

Research process in detail (1)

> From **topic** (or problem) to the **research question**

- Difference between motivation and question
- What are good & bad research questions?
  - → Answerable
  - → Relevant: importance and connection theory/empirics (*)

Research process in detail (1)

> From **topic** (or problem) to the **research question**

- Difference between motivation and question
- What are good & bad research questions?
  - → Answerable
  - → Relevant: importance and connection theory/empirics (*)

Examples, good and bad

- → What is the best Netflix show?
- → What can we do to help poor countries develop?
- → What are the shopping patterns of the Spanish population?
- → Do individuals from minorities support the use of violence?

Research process in detail (2)

- From **RQ** to **theory**

Research process in detail (2)

> From **RQ** to **theory**

- What is a theory? Importance

Research process in detail (2)

> From **RQ** to **theory**

- What is a theory? Importance
- Arguments and mechanisms
- Micro-level, macro-level, and both

Research process in detail (2)

> From **RQ** to **theory**

- What is a theory? Importance
- Arguments and mechanisms
- Micro-level, macro-level, and both
- Good theories or arguments
  - → Testable
  - → Credible

Research process in detail (3)

- From **theory** to **measurement**

Research process in detail (3)

> From **theory** to **measurement**

- Concepts

Research process in detail (3)

> From **theory** to **measurement**

- Concepts
- Operationalization, unit of analysis

Research process in detail (3)

> From **theory** to **measurement**

- Concepts
- Operationalization, unit of analysis
- Measurement

Research process in detail (4)

> From **measurement** to **strategy**

- Once you have your variables measured, what variation are you going to look at?
- Descriptive and causal inference
- (Experiments) (*)

Research process in detail (5)

> From **strategy** to **analyses**

- We're **not** covering this in this course
- This basically means choosing how you are going to analyze the variation you have
- Comparisons, models, statistics
- **Question:** what is the role of statistics in research design?

Research process in detail (6)

> From **analyses** to **results** and back to **topic**

- Interpretation
- Relevance
- **External validity**

Simplyfing

1. Formulate one (or more) RQ from a topic/problem
2. Develop a theoretical argument related to that RQ
3. Building on theory, think about what and how to measure
4. What variation are you going to look at? (and what data analysis tools do you need?)
5. What can you really learn from the results?

Let's go through an example

You are hired by the city government as a quantitative analyst to tackle the problem of **urban traffic in Madrid** and offer suggestions to improve it

Course logistics

- Tuesdays 18:00–21:00 (not always)
- Seven sessions between today and October 20th

- My email: `francisco.villamil@uc3m.es`
- Office hours

- https://franvillamil.github.io/res_design/

Course logistics: evaluation

- Participation (15%)
- Research papers reviews (15%)
- Workshop group presentation (20%)
- Workshop feedback (10%)
- Final essay (40%)

Workshop and final essay

- In the workshop we will have around 12-15 slots (10 minute presentation & 10 feedback) so we need to do **group presentations**
  - → We can adjust time once we know the final number of groups
- I do not mind if you do the final essay individually or in groups (max 3/4)
- Two good options:
  - → Do a group presentation covering different strategies and individual essay with each of them
  - → Do a group final essay

Final essay

- Get a topic or problem and go through all the steps in the research process
- No need to do data analyses or statistics (but you could show something if it's relevant)
- **Most important thing**: show me you understand how you can learn something about the topic from quantitative data, what different strategies you could use and what are the general and specific limitations of them
- Word limit: **5000 words** (we can talk in case of group essays)
  - → Appendix obviously does not count
- Deadline: **October 28th, 23.59h** (exam week)

Course logistics: calendar

- 5 lectures (including today)
  - → In sessions 2–4, we'll discuss a paper in the second half
  - → You have to submit some comments/critique (5% each), **before** class (email or paper)
  - → Final lecture on advanced topics, overview, and questions
- Last day: Workshop session (15h-21h)

Course logistics: calendar

- Sept 16: Introduction
- Sept 23: Elements of quantitative data
- Sept 30: Causality and experimental evidence
- Oct 6 (Monday): Causal inference with observational data
- Oct 14: Advanced topics and overview
- Oct 20 (Monday, 15h-21h): Workshop

Textbooks and resources

- **Nick Huntington-Klein, The Effect: An Introduction to Research Design and Causality (Chapman and Hall/CRC, 2021).**
- Kosuke Imai, *Quantiative Social Science: An Introduction* (Princeton UP, 2017).
- Dimiter Toshkov, *Research Design in Political Science* (Palgrave, 2016)
- Scott Cunningham, *Causal Inference: The Mixtape* (Yale University Press, 2021).

Papers to read

1. Carl Müller-Crepon, Philipp Hunziker, and Lars-Erik Cederman (2021) Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa. *Journal of Conflict Resolution* 65(2–3): 563–590.
2. Andrew M. Guess *et al.* (2023) How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381(6656): 398–404.
3. Francisco Villamil and Laia Balcells (2021) Do TJ policies cause backlash? Evidence from street name changes in Spain. *Research & Politics* 8(4).

Next week's reading

Journal of Conflict Resolution
2021, Vol. 65(2-3) 563-590

DOI: 10.1177/0022002720963674
journals.sagepub.com/home/jcr

Roads to Rule, Roads to Rebel: Relational State Capacity and Conflict in Africa

Carl Müller-Crepon[1], Philipp Hunziker[2], and Lars-Erik Cederman[3]

Abstract

Weak state capacity is one of the most important explanations of civil conflict. Yet, current conceptualizations of state capacity typically focus only on the state while ignoring the relational nature of armed conflict. We argue that opportunities for conflict arise where relational state capacity is low, that is, where the state has less control over its subjects than its potential challengers. This occurs in ethnic groups that are poorly accessible from the state capital, but are internally highly interconnected. To test this argument, we digitize detailed African road maps and convert them into a road atlas akin to Google Maps. We measure the accessibility and internal connectedness of groups via travel times obtained from this atlas and simulate road networks for an instrumental variable design. Our findings suggest that low relational state capacity increases the risk of armed conflict in Africa.

Questions?