

# Complementar los datos convencionales de encuestas web con nuevas oportunidades de medición para obtener mejores o nuevos conocimientos

*Madrid, 23 de Junio 2023*

**Melanie Revilla | IBEI**

**Agradecimientos:** Este proyecto ha recibido financiación del Consejo Europeo de Investigación (ERC, por sus siglas en inglés) en el marco del programa de investigación e innovación Horizonte 2020 de la Unión Europea (acuerdo de subvención n.º 849165); del Ministerio de Ciencia e Innovación de España en el marco del programa "Proyectos de I+D+i" (número de subvención PID2019-106867RB-I00/AEI/10.13039/501100011033 (2020-2024)); y de la Fundación BBVA en el marco de su programa de subvenciones a equipos de investigación científica en economía y sociedad digital, 2019.

Quiero agradecer a Oriol Bosch y Patricia Iglesias por sus comentarios y Carlos Ochoa por su ayuda en preparar esta presentación.

# ¿Qué nuevas oportunidades?

## Creciente uso de Internet (especialmente en móviles)

Cada vez más aspectos de la vida de las personas ocurren **online**

Tiempo diario promedio <sup>1</sup> que pasa cada usuario de Internet online

|         |             |                               |
|---------|-------------|-------------------------------|
| en 2021 | <b>6h58</b> | ) <b>+30 min</b><br>en 5 años |
| en 2016 | <b>6h29</b> |                               |

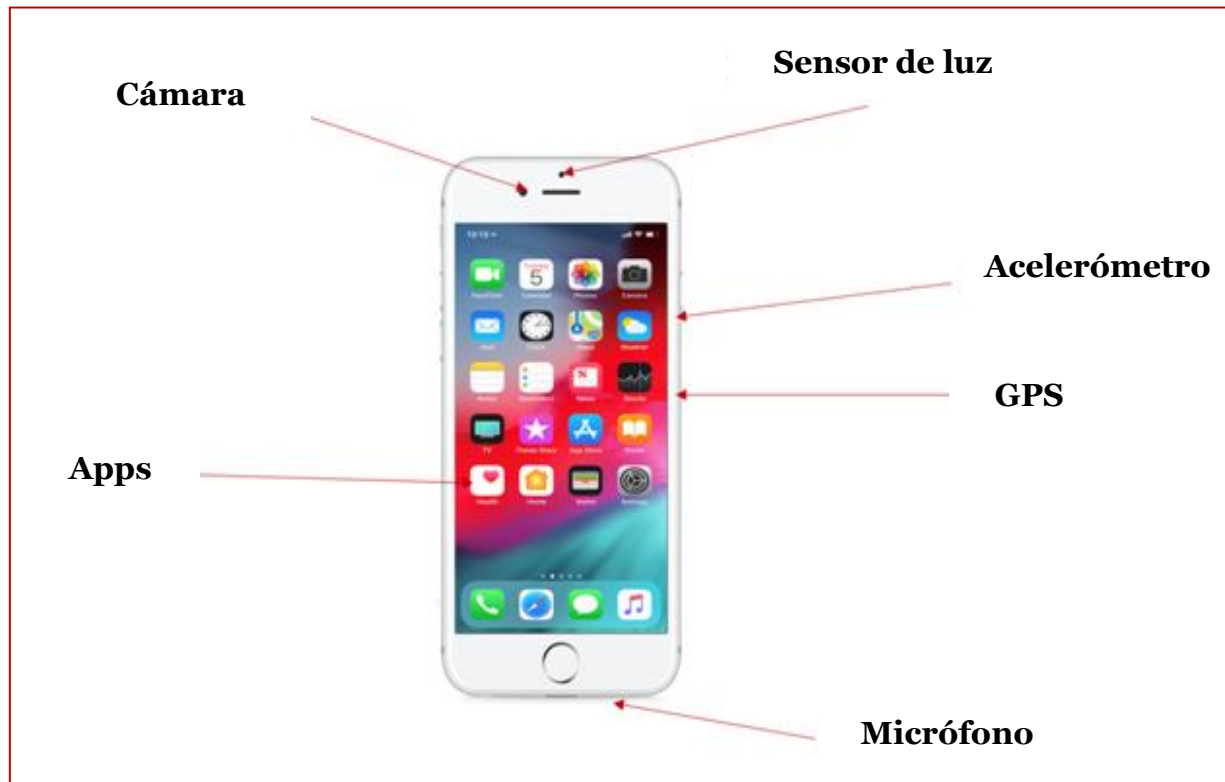
Cada vez una mayor parte de las actividades online se realiza a través de **smartphones**

**83%** de la población mundial tiene un smartphone<sup>2</sup>

**92%** de los usuarios Internet en el mundo acceden a Internet a través de smartphones<sup>1</sup>

¿QUÉ NUEVAS OPORTUNIDADES?

# Posibilidad de recolectar muchos tipos de datos diferentes



Muchos tipos de  
datos



Cada uno tiene sus  
propios beneficios y  
limitaciones

## Nuevos tipos de datos estudiados

### DATOS VISUALES



Capturas de pantalla  
Fotos/videos capturados durante la encuesta  
Fotos/videos almacenados en el dispositivo

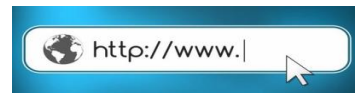
### DATOS de VOZ



Dictado  
Grabación de voz

La mayoría de esos datos también se pueden obtener desde PCs

### DATOS de NAVEGACION



Obtenidos a través de una aplicación de rastreo digital (“meter”) instalada por los participantes en sus dispositivos para registrar (al menos) las URLs de las páginas web visitadas.  
Por lo general, se emplea en paneles.

### DATOS de GEOLOCACION



Obtenidos a través de una aplicación de rastreo instalada en los dispositivos de los participantes para registrar sus coordenadas GPS.

ENCUESTAS “EN EL MOMENTO” desencadenadas por esos datos

## *Ejemplo 1*

Estudiar las **condiciones** en las que **teletrabajan** las personas mediante el uso de datos visuales

# El teletrabajo en el centro del debate

Economía

## Patronal y sindicatos pactan salvar el teletrabajo y se olvidan de la jornada de cuatro días

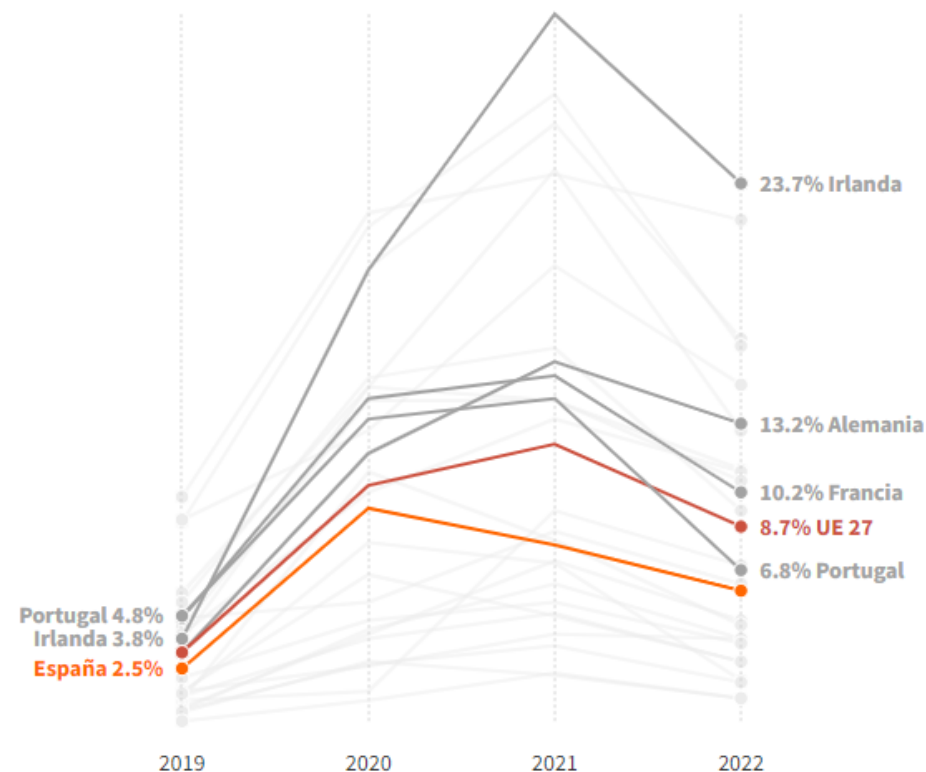
Javier Esteban

6:00 - 12/05/2023

El V Acuerdo Estatal de Negociación Colectiva que acaban de firmar patronal y sindicatos dedica un capítulo completo a propuestas para impulsar la implantación del teletrabajo en los convenios colectivos, una de las claves que, según los interlocutores sociales **explican el escaso peso de esta modalidad en nuestro país tras la pandemia**. Pero aunque el texto hace una decidida apuesta por la flexibilidad del tiempo de trabajo obvia el debate sobre la reducción del tiempo de trabajo, que tiene su máximo exponente **en la jornada de cuatro días**.

## Evolución del teletrabajo entre 2019 y 2022


En porcentaje sobre el total de asalariados.



Fuente: Eurostat.

elEconomista.es

# Necesidad de estudiar las condiciones de teletrabajo

Malas condiciones de trabajo  Baja productividad  
Problemas de salud física (vista, cuello, espalda...) y mental





# Algunos estudios ya existentes

## Teletrabajo: el 97% de los hogares españoles no están preparados

por Germán Cañavate | 23 Sep 2020 | Productividad, Riesgos psicosociales, Teletrabajo | 2 Comentarios

### Resultados de la evaluación de teletrabajo de 1056 hogares españoles

Durante el tiempo del confinamiento por la pandemia de Covid19, de marzo a junio de 2020, se recabaron 1.056 encuestas de autoevaluación de las condiciones laborales sobre el teletrabajo.

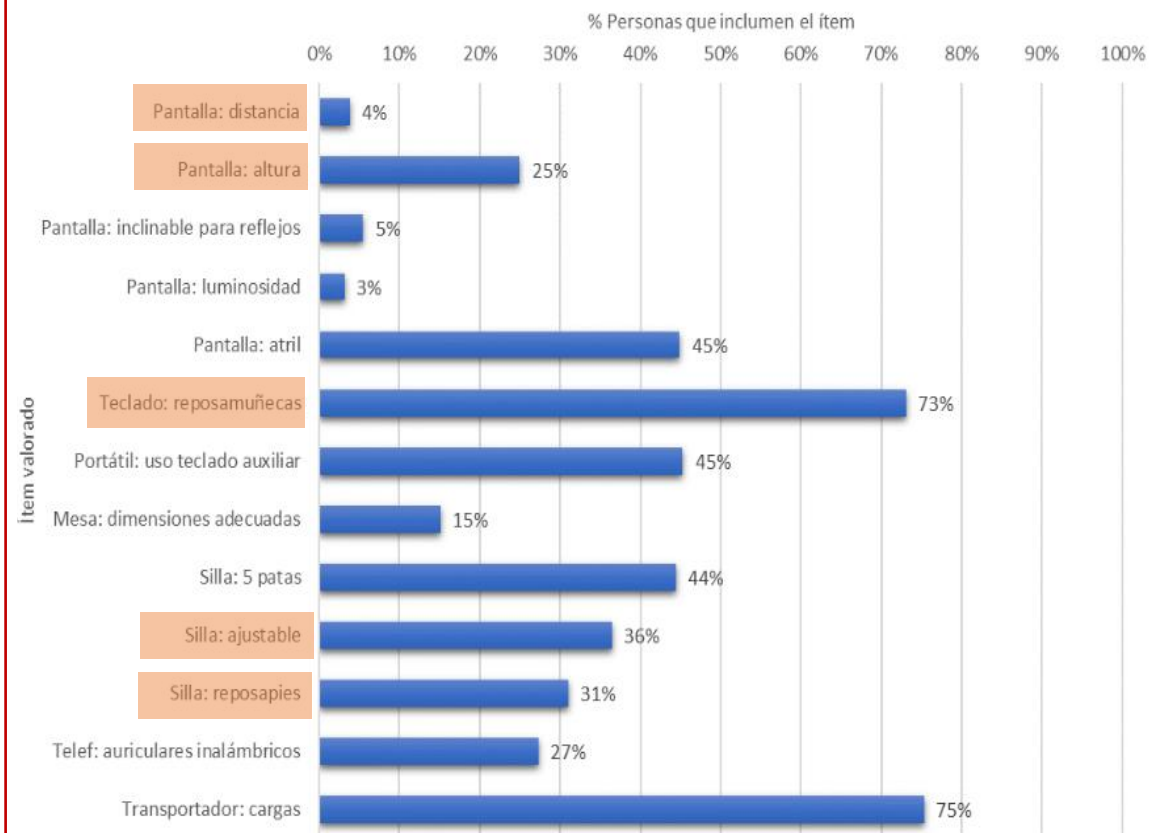
El martes 22/09/2020 se publicó por parte del Gobierno la noticia de la aprobación de la **nueva Ley de Teletrabajo**. Y siendo objetivos, hay mucho trabajo que abordar en los hogares españoles para un cumplimiento eficaz de esta nueva norma.

Es importante recordar que **unas condiciones de teletrabajo mal gestionadas pueden ser causa y motivo de riesgos psicosociales**, lo que provocará más absentismo laboral, pérdida de productividad, fuga de talentos y bajas por motivos laborales. Cosa que ninguna organización se puede permitir.

Según el primer análisis de los resultados, solamente el 3% de todas las situaciones evaluadas obtuvieron la puntuación máxima: es decir, hogares que tuvieron los 43 ítems evaluados perfectamente controlados.

**El 97% de los hogares tenían alguna deficiencia en materia de condiciones de teletrabajo.**

### ERGONOMÍA DEL PUESTO DE TELETRABAJO



# Pedir fotos del lugar donde teletrabajan podría ayudar

## Principales beneficios esperados

### Investigadores

- Reducción de **errores de medición**, especialmente los generados por **deseabilidad social**
- Información más **detallada**
- **Nueva información**: posibilidad de medir conceptos no medidos hasta ahora

*“Una imagen  
vale más que  
1.000  
palabras”*

### Participantes

- Reducción del **tiempo** necesario para proporcionar la información
- Reducción del **esfuerzo** cognitivo
- Más **ameno**

# Pero no es tan fácil...

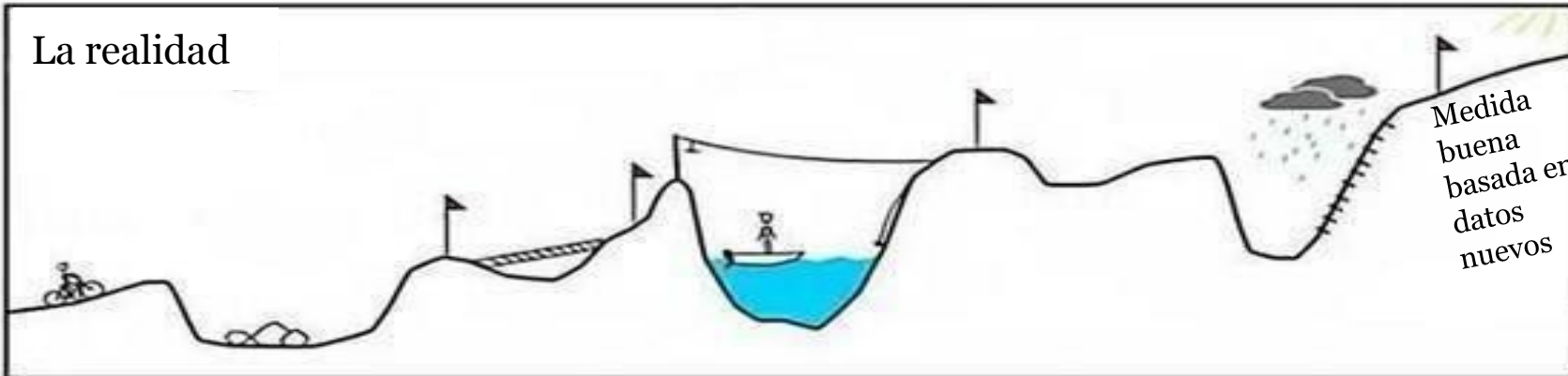
Lo que suele pensar la gente "normal" sobre usar nuevos tipos de datos



Medida  
buena  
basada en  
datos nuevos



La realidad



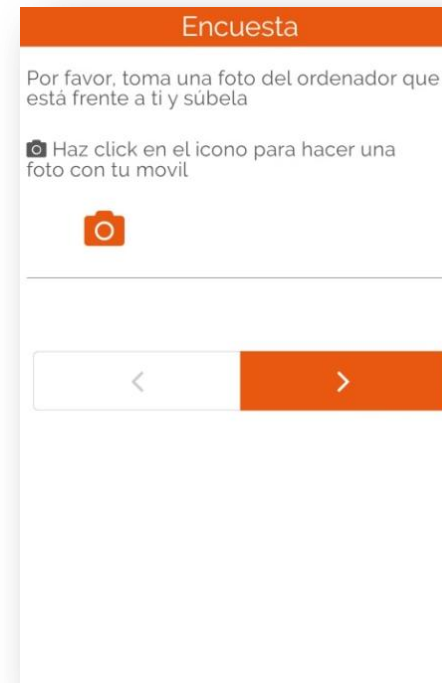
Medida  
buena  
basada en  
datos  
nuevos

# Problema 1: se necesita una herramienta especial

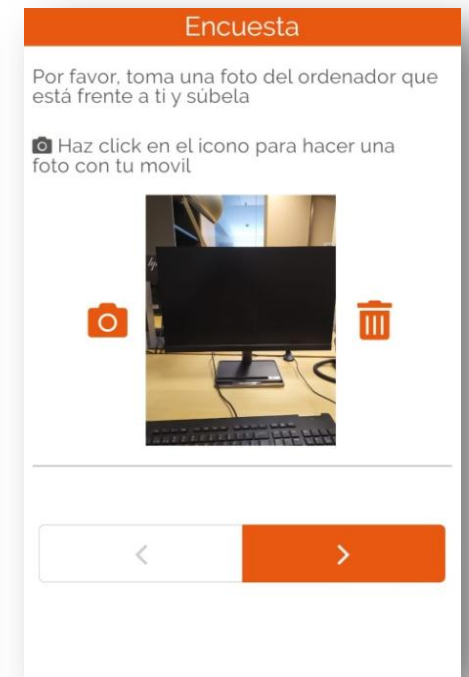
Hemos desarrollado **WebdataVisual** (Revilla et al., 2022)




Recolecta datos visuales ya guardados en el dispositivo



O capturados durante la encuesta



## Problema 2: los participantes tienen que enviar imágenes

Investigación previa:  de los respondientes **comparten imágenes** cuando se lo pedimos en encuestas web (Bosch et al., 2018)



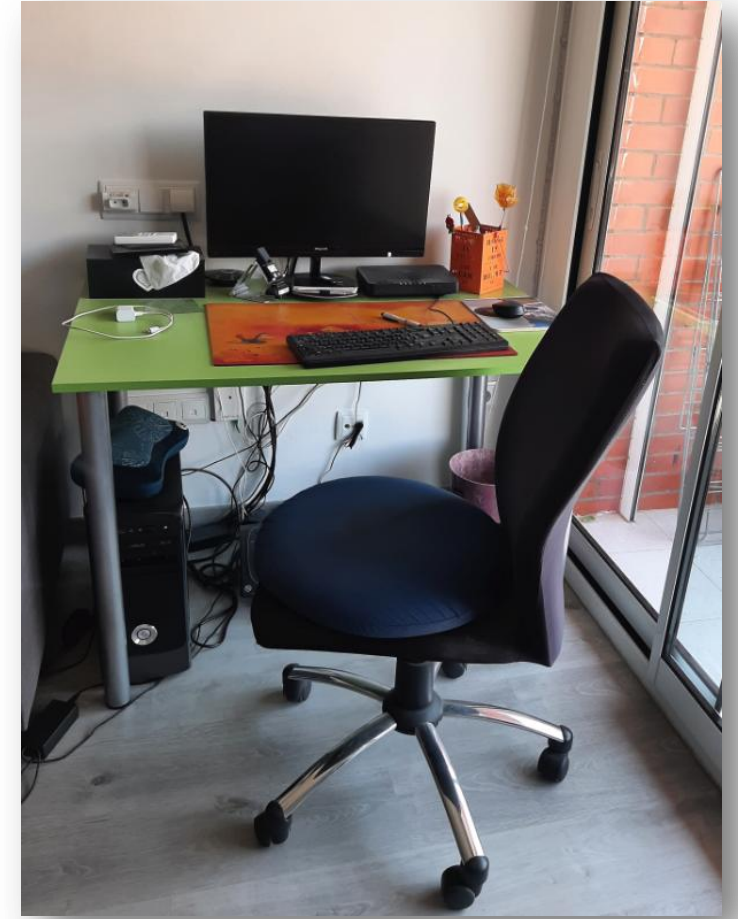
Usando datos de un opt-in online panel en España, encontramos que la **disponibilidad** parece ser el factor más limitante para la participación (Iglesias & Revilla, 2023)

## Problema 3: como extraer la información relevante

Calidad de los datos depende de cómo se extrae la información = “**clasificación**”



¿Como definir  
correctamente qué  
extraer?



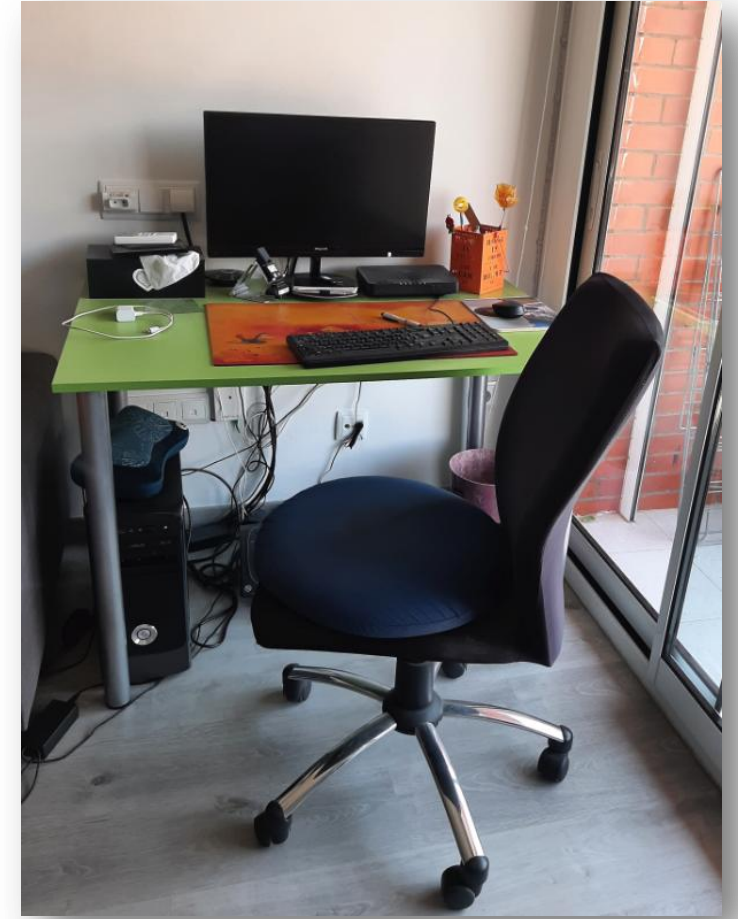


## Problema 3: como extraer la información relevante

Calidad de los datos depende de cómo se extrae la información = “**clasificación**”



¿Qué método de clasificación es el más adecuado?



# Problema 3: como extraer la información relevante

Método de clasificación: manual o automático (“machine learning”)

Ejemplo: resultados de la API Google Vision: “etiquetas”

Objects Labels Text Properties Safe Search

|                   |     |
|-------------------|-----|
| Computer          | 97% |
| Table             | 97% |
| Personal Computer | 95% |
| Furniture         | 95% |
| Plant             | 95% |
| Computer Desk     | 95% |
| Property          | 94% |
| Desk              | 93% |
| Peripheral        | 91% |
| Computer Keyboard | 90% |
| Computer Monitor  | 89% |
| Writing Desk      | 89% |

Objects Labels Text Properties Safe Search

|                   |            |
|-------------------|------------|
| Armrest           | 84%        |
| Building          | 83%        |
| Floor             | 82%        |
| Desktop Computer  | 80%        |
| Flooring          | 80%        |
| Computer Keyboard | 80%        |
| Automotive Design | 77%        |
| Office Equipment  | 77%        |
| <b>Comfort</b>    | <b>67%</b> |
| Room              | 67%        |
| Curtain           | 67%        |
| Auto Part         | 64%        |

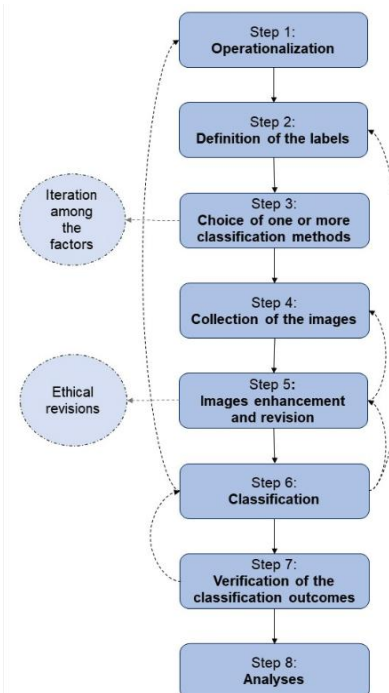


## Problema 3: como extraer la información relevante

Método de clasificación: manual o automático (“machine learning”)

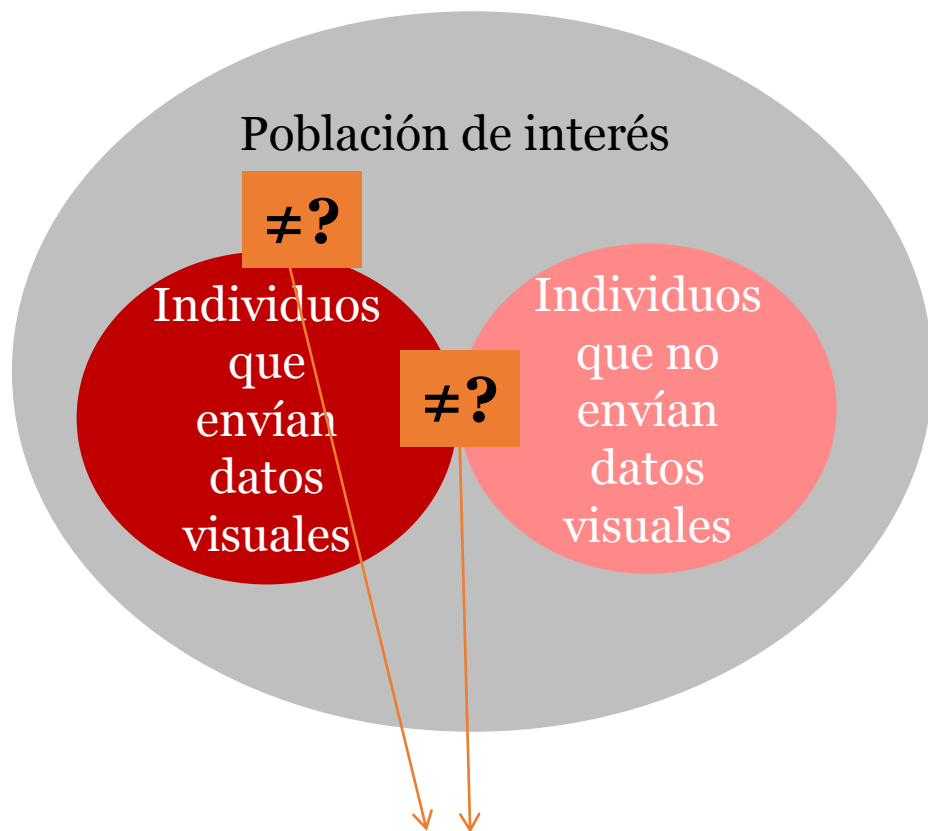


Guía práctica para ayudar a los investigadores interesados en usar imágenes con estos problemas (Iglesias et al., 2022)



# Más problemas

## ¿Sesgo de **selección**?



Depende población objetivo + incentivo + etc.

## Protección de datos y consideraciones **éticas**



¿Cómo asegurarse de que el **consentimiento** sea realmente **informado**?

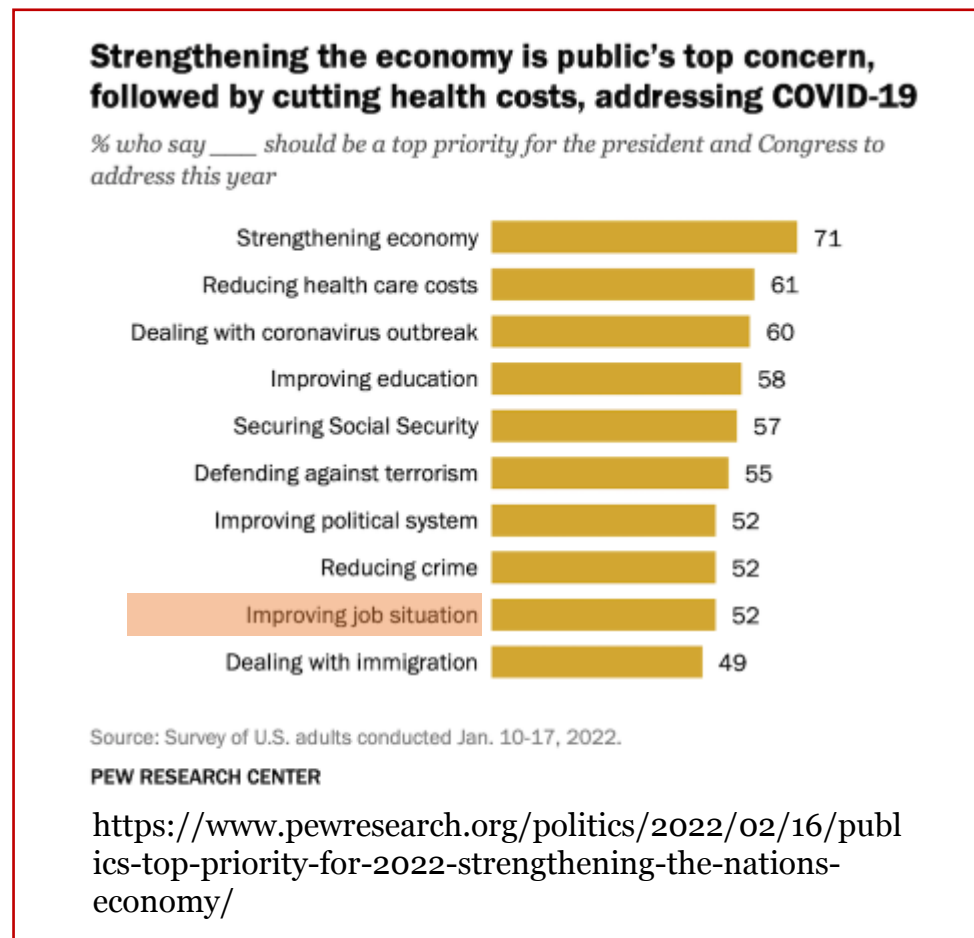


¿Cómo tratar imágenes que contienen **datos personales** (propios o de terceros)?

## *Ejemplo 2*

Estudiar cómo la gente **postula a trabajos online** mediante el uso de datos de navegación y encuestas “en el momento”

# El empleo: una preocupación central



➔ Entender cómo la gente **busca y postula a trabajos** (especialmente online) es clave

# Algunos estudios ya existentes (principalmente en EEUU)

## How do jobseekers search for jobs? New data on applications, interviews, and job offers

*By Michael R. Dalton and Jeffrey A. Groen*

This **Beyond the Numbers** article explores aspects of job search using data from a supplement to the **Current Population Survey (CPS)** in May and September 2018 which obtained information about the job search of those who were not employed and asked whether people applied for and received unemployment insurance (UI) benefits. The CPS is a monthly survey of about 60,000 households that provides data on employment and unemployment in the **United States**, including the national unemployment rate. In the supplement, questions about job search were asked of those who were without a job and had looked for work recently. In this analysis, we restrict the sample to those who were unemployed at the time of the survey and had looked for work in the past 4 weeks.

*Fuente:* <https://www.bls.gov/opub/btn/volume-9/how-do-jobseekers-search-for-jobs.htm>

- **65% of currently employed individuals** are looking for new jobs as of 2022.
- The average time to find a job in 2023 is **five months**.
- **80%** of all job searches are done online as of 2022.

*Fuente:* <https://www.zippia.com/advice/job-search-statistics/>

# Usar datos de navegación podría ayudar

## Principales beneficios esperados

### Investigadores

- Reducción de **errores de medición**, especialmente los generados por **problemas de memoria**
- Información más **detallada**
- En **tiempo real** / continuo
- **Nueva** información

Ej: ¿A cuántas ofertas ha postulado un candidato?

Ej: lista de todas las URLs que ha visitado un candidato (y tiempo dedicado en cada una) hasta postular a una oferta de trabajo

### Participantes

- **Esfuerzo** casi nulo (solo instalar la app)
- Reciben **incentivo** sin tener que participar activamente

# Pero no es tan fácil...

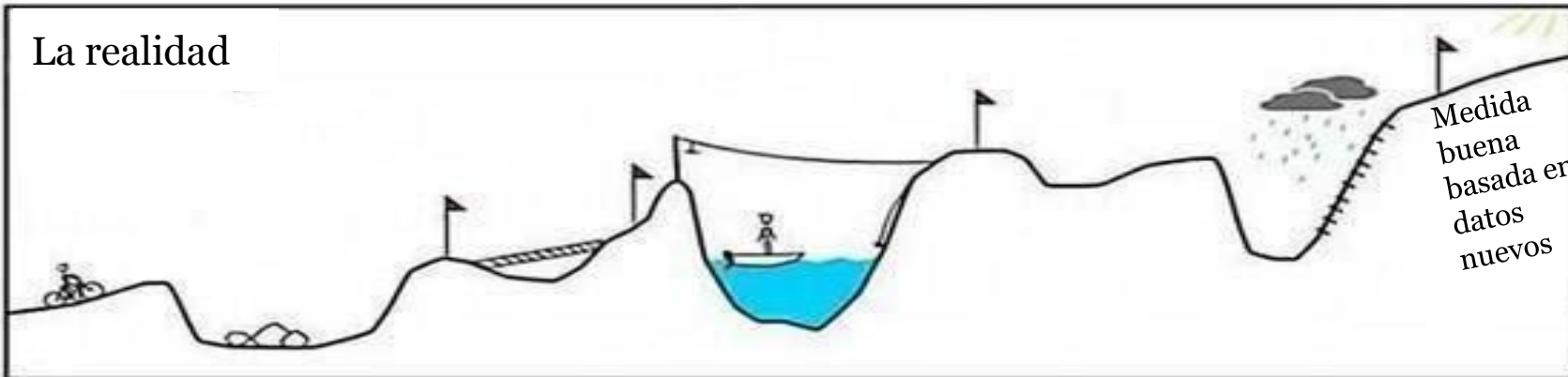
Lo que suele pensar la gente "normal" sobre usar nuevos tipos de datos



Medida  
buena  
basada en  
datos nuevos



La realidad



Medida  
buena  
basada en  
datos nuevos

## Problema 1: los datos de navegación también tienen errores

Es crucial:

1. **Identificar** los potenciales errores
2. Estimar su **tamaño**
3. Encontrar maneras para **minimizarlos**
4. Y/o **corregir** esos errores



**TEM = “Total error framework for metered data”** (Bosch & Revilla, 2022)



Adaptación del TSE



Visión general de todos los posibles errores y sus causas



# Problema 1: los datos de navegación también tienen errores

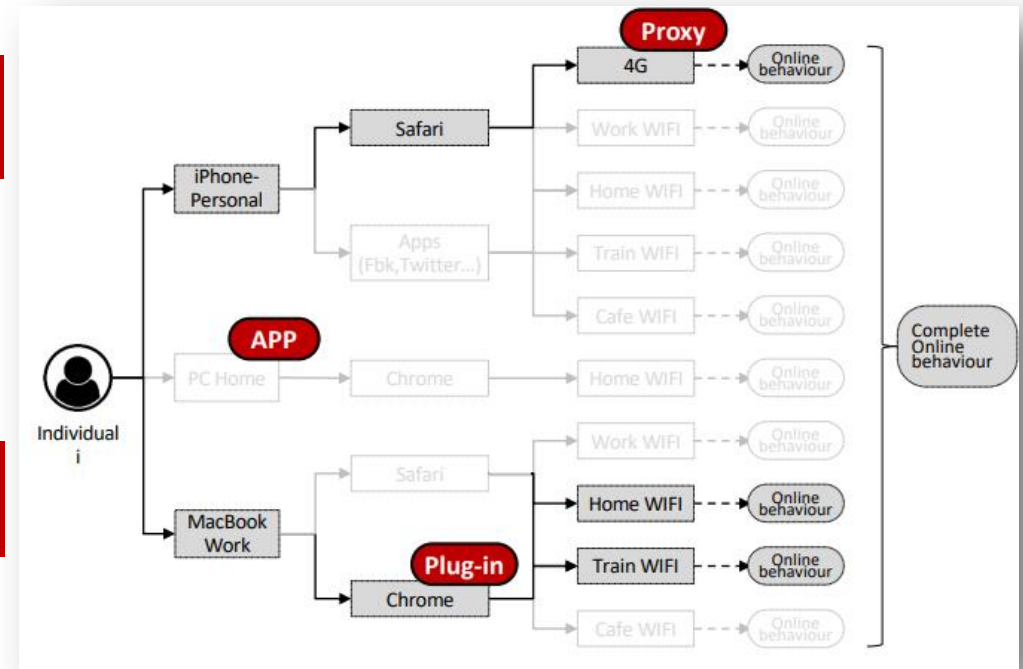
| Error components    | Specific error causes   |
|---------------------|---|
| Specification error | <ul style="list-style-type: none"> <li>- Measuring concepts from which not enough data is available</li> <li>- Inferring attitudes</li> <li>- Defining valid information</li> </ul>   |
| Measurement error   | <ul style="list-style-type: none"> <li>- Non-trackable target</li> <li>- Meter not installed</li> <li>- Uninstalling the meter</li> <li>- New non-tracked device</li> <li>- Technology limitations</li> <li>- Technology errors</li> <li>- Hidden behaviours</li> <li>- Shared device</li> <li>- Social desirability</li> <li>- Extraction error</li> </ul>                   |
| Processing error    | <ul style="list-style-type: none"> <li>- Coding error</li> <li>- Aggregation at the domain level</li> <li>- Data anonymization</li> </ul>   |
| Coverage error      | <ul style="list-style-type: none"> <li>- Non-trackable individuals</li> </ul>   |
| Sampling error      | <ul style="list-style-type: none"> <li>- Same error causes than for surveys</li> </ul>  |
| Missing data error  | <ul style="list-style-type: none"> <li>- Noncontact</li> <li>- Non-consent</li> <li>- Non-trackable target</li> <li>- Meter not installed</li> <li>- Uninstalling the meter</li> <li>- New non-tracked device</li> <li>- Technology limitations</li> <li>- Technology error</li> <li>- Hidden behaviour</li> <li>- Social desirability</li> <li>- Extraction error</li> </ul> |

**Dispositivos compartidos**

**Meter no instalado**

**Limitaciones tecnológicas**

“Tracking undercoverage” afecta al 80-85 % de los participantes en los datos de TRI-POL<sup>1</sup> (Bosch et al., 2022)



<sup>1</sup><https://www.upf.edu/web/tri-pol>

## Problema 2: poca información subjetiva

Los datos de navegación aportan mucha información sobre los **comportamientos online**

... pero no nos dicen el **porqué**

¿Por qué una persona postula a una oferta de trabajo y no a otra?

¿Por qué una persona busca trabajo en una web y no en otra?

... ni nos dan información sobre **conocimientos, sentimientos**, etc.

A esa persona, ¿le ha gustado cómo se presentaba la información de la oferta de trabajo?

¿Ha entendido el contenido correctamente?

De todas las ofertas a las que postuló, ¿tiene una preferencia fuerte por alguna? ¿Por cuál?

# Usar encuestas “en el momento” desencadenadas por datos de navegación podría ayudar



➔ Experimento implementado en el panel Netquest en España para estudiarlo (Ochoa, 2023)

# Pero no es tan fácil...

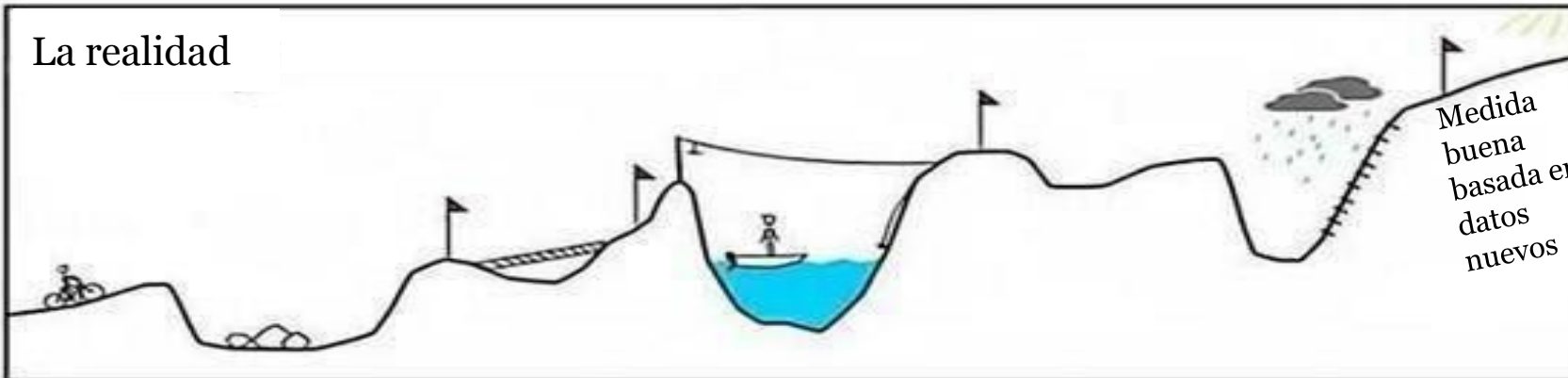
Lo que suele pensar la gente "normal" sobre usar nuevos tipos de datos



Medida  
buena  
basada en  
datos nuevos

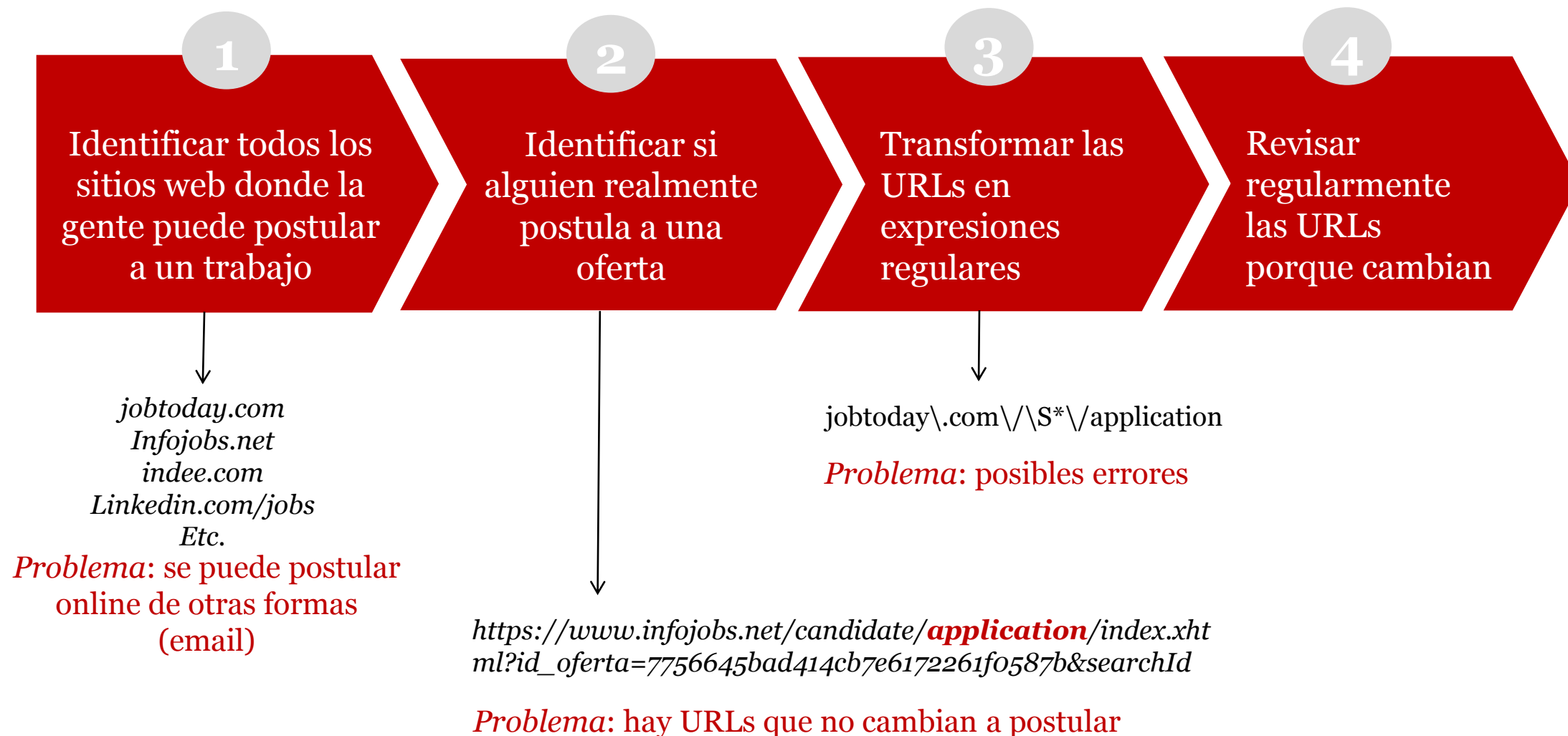


La realidad



Medida  
buena  
basada en  
datos nuevos

# Problema 1: Identificar los eventos de interés



## Problema 1: Identificar los eventos de interés

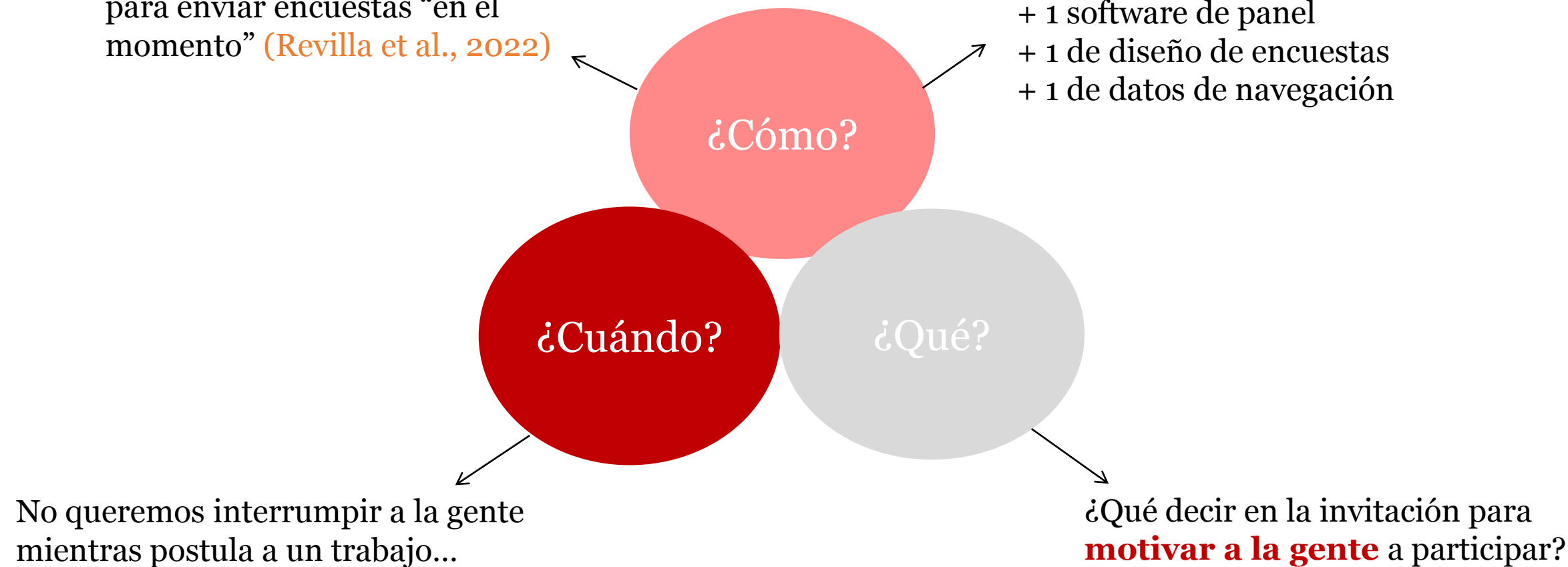


- ➔ No podemos detectar (ni invitar) a todos los que postulan online
- ➔ Podemos invitar a personas que no postularon (dispositivo compartido)

## Problema 2: Invitar a los participantes

**WebdataNow:** una herramienta para enviar encuestas “en el momento” (Revilla et al., 2022)

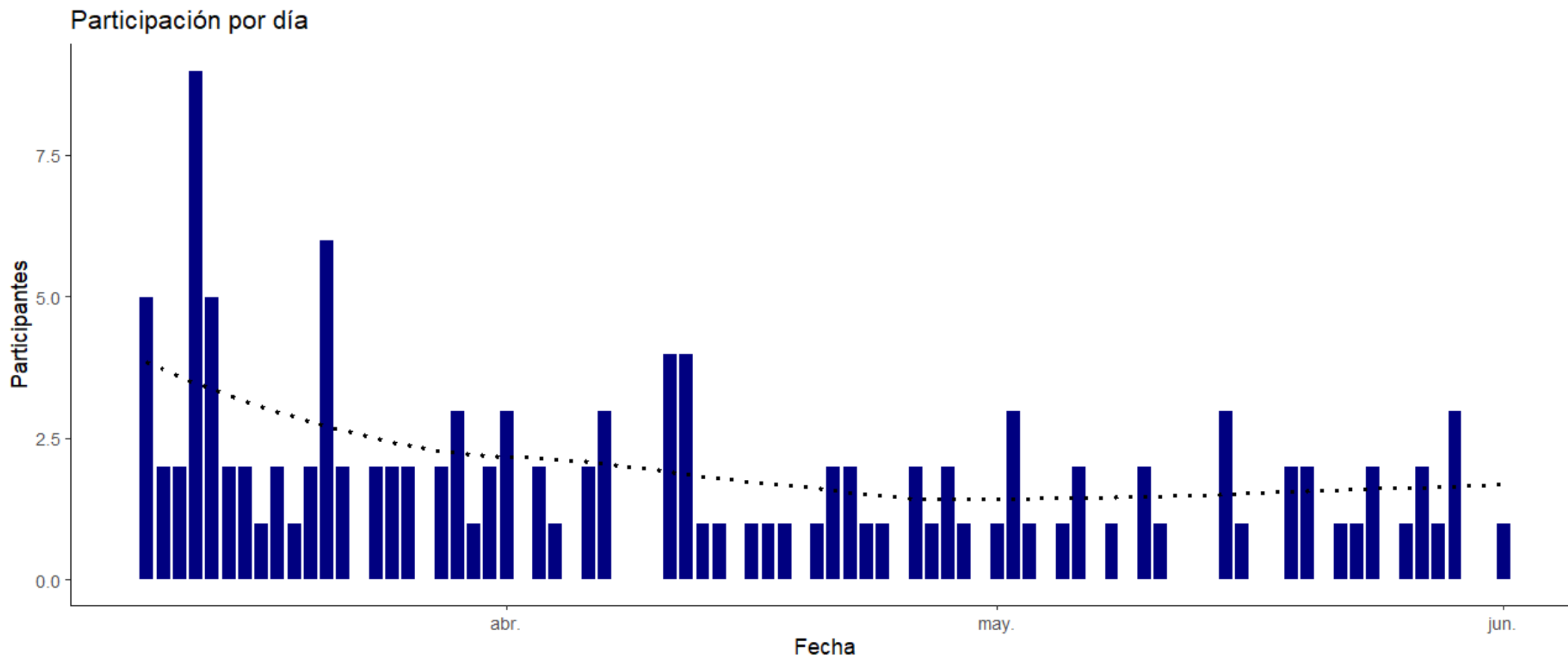
Se necesitan también:  
+ 1 software de panel  
+ 1 de diseño de encuestas  
+ 1 de datos de navegación



➔ Estas decisiones afectan al tamaño + a la representatividad de la muestra + a la calidad de los resultados

# Aun así, primeros resultados alentadores

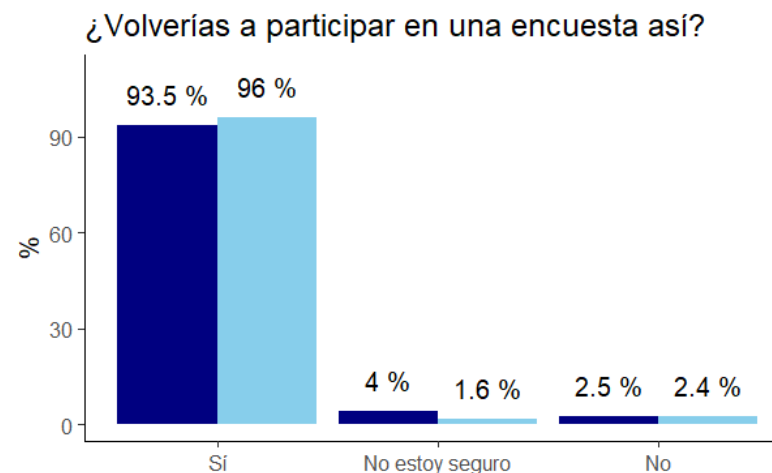
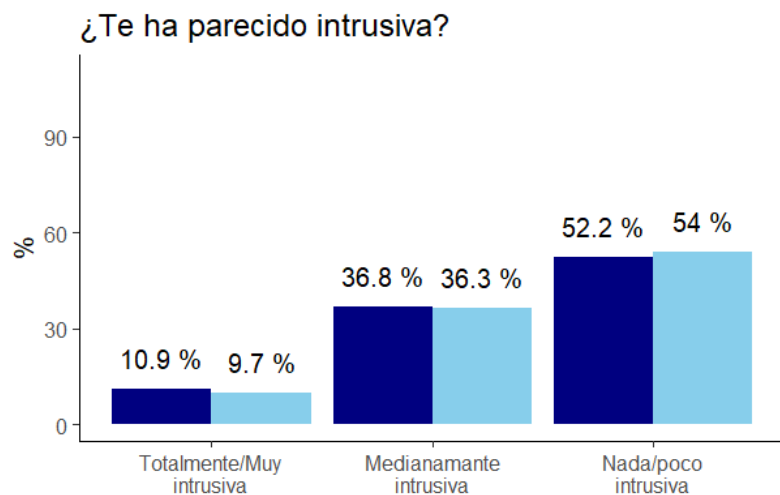
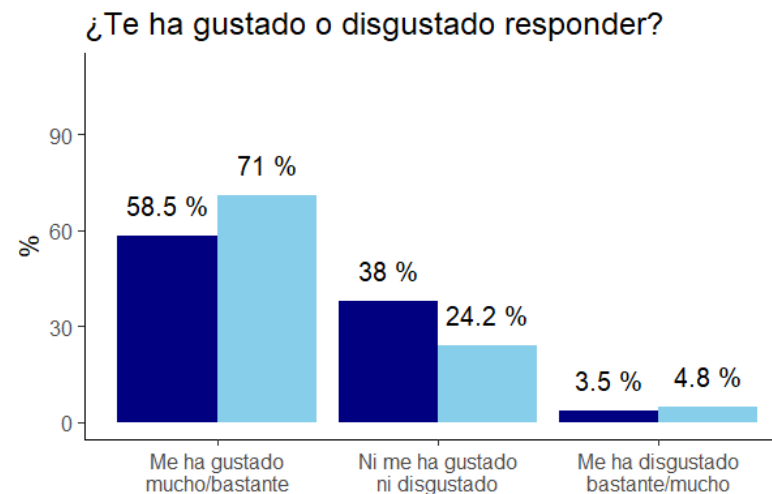
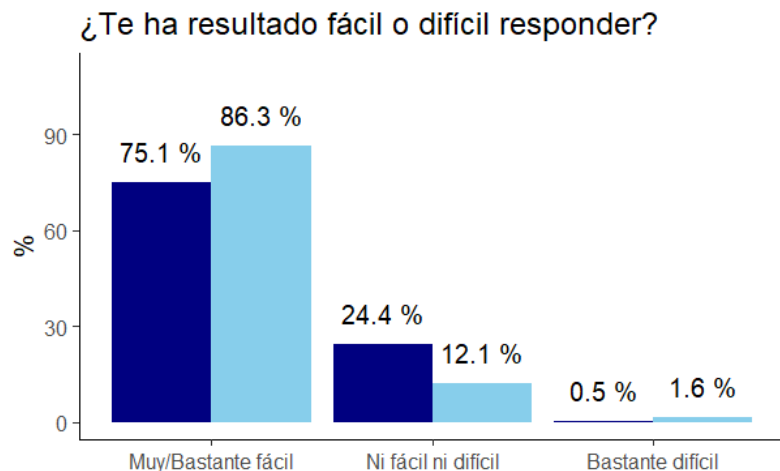
Trabajo de campo lento pero más o menos constante





# Aun así, primeros resultados alentadores

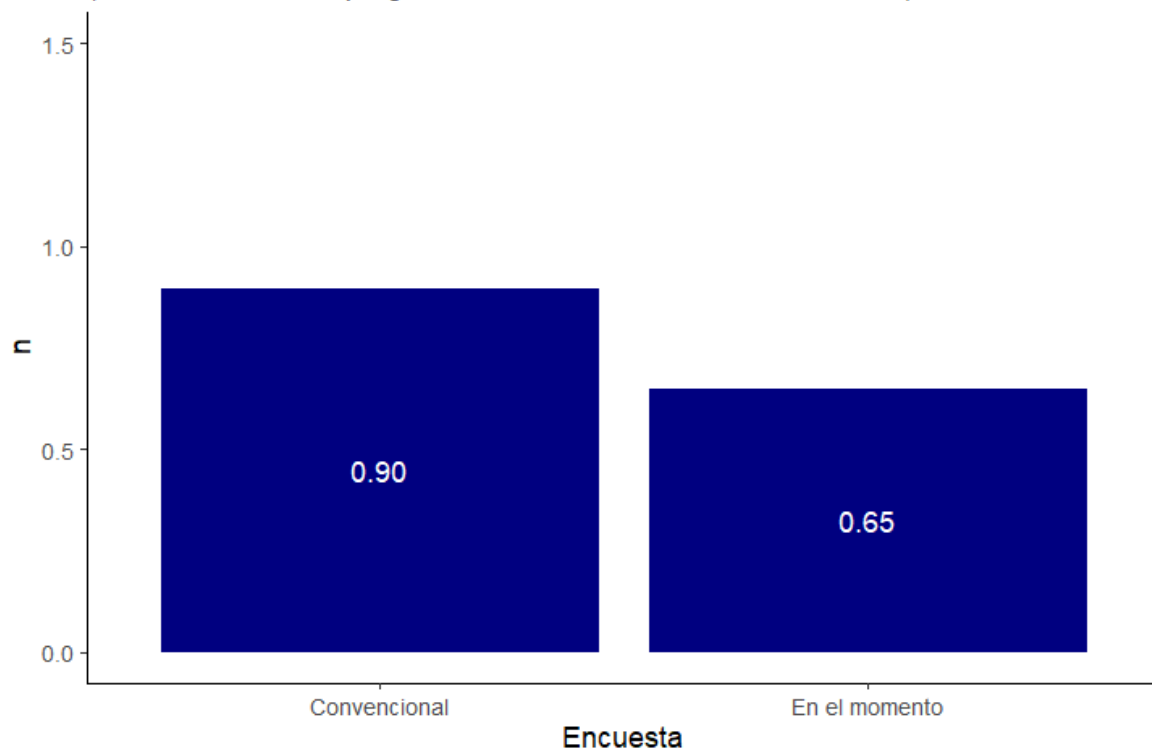
## Participantes contentos de su experiencia



# Aun así, primeros resultados alentadores

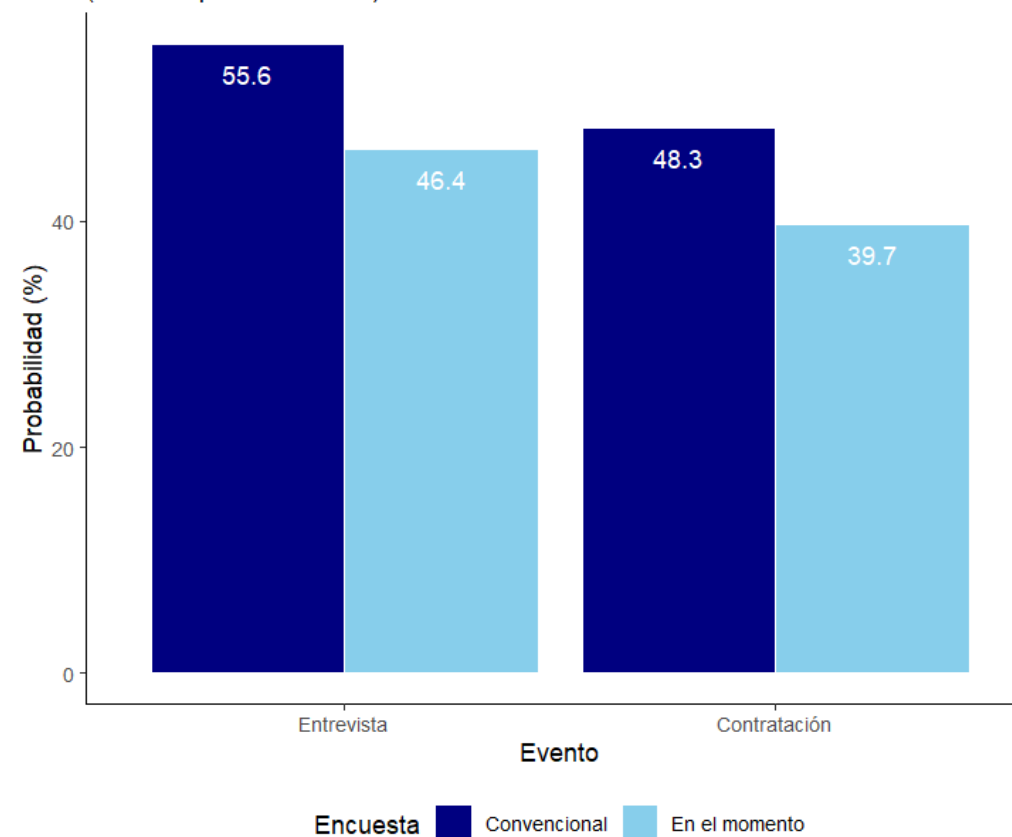
## Menos “no lo recuerdo”

Número medio de respuestas 'no lo recuerdo'  
(sobre un total de 4 preguntas sobre características de la oferta)



## Resultados diferentes

Probabilidad que el aplicante creía tener de ser entrevistado/contratado  
(cuando aplicó a la oferta)



# Conclusiones

# Los nuevos tipos de datos tienen un gran potencial

Posible estudiar cosas que los respondientes no saben/recuerdan

Menos esfuerzo



Más ameno

Gran cantidad de información

Menos problemas de deseabilidad social

## Más investigación necesaria

1 Aprender más sobre los **errores** de esos datos en diferentes contextos

Tipos de errores

Tamaño

Efecto sobre los resultados

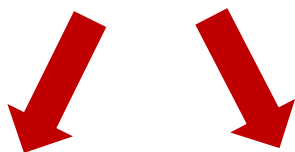
2 Entender mejor **cuándo** usar estos datos

Necesario  
entender mejor  
los mecanismos

Necesario identificar cuándo  
beneficios > desventajas  
(investigadores y participantes)

## Más investigación necesaria

3 Entender mejor **cómo** usar estos datos



¿**Reemplazando** preguntas?



Siempre habrán errores



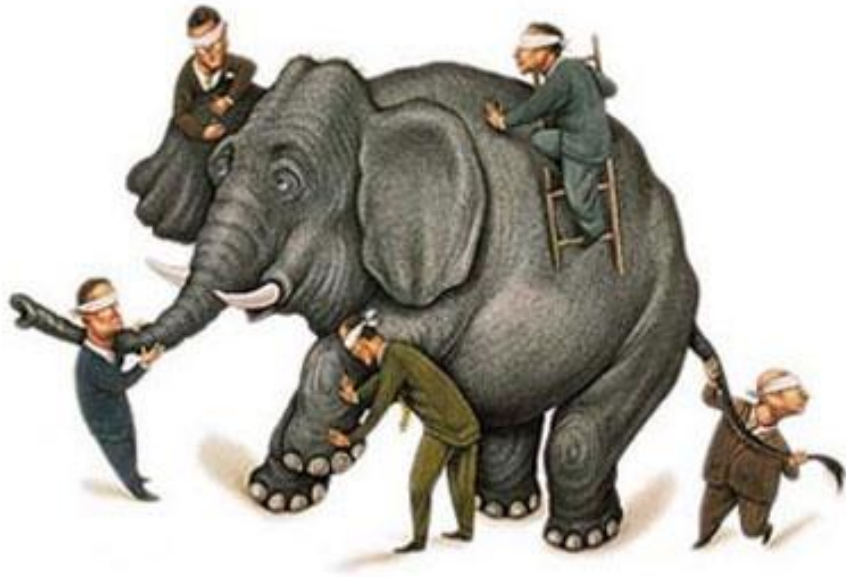
Crucial **identificarlos** y pensar en sus **consecuencias**

¿**Combinando** con preguntas?



Proporciona información **diferente** pero **complementaria**

# Mirar desde diferentes perspectivas



THE BLIND MEN AND THE ELEPHANT

“And so these men of research  
Disputed loud and long,  
Each in his own opinion  
Exceeding stiff and strong,  
Though each was partly in the right,  
And all were in the wrong!”

John Godfrey Saxe (1816-1887)

# Gracias!

## *¿Preguntas?*

Melanie Revilla | IBEI



[mrevilla@ibei.org](mailto:mrevilla@ibei.org)



<https://www.upf.edu/web/webdataopp>



INSTITUT  
BARCELONA  
ESTUDIS  
INTERNACIONALS





# Referencias

- Bosch, O.J., Revilla, M., & E. Paura (2018). Answering mobile surveys with images: an exploration using a computer vision API. *Social Science Computer Review*, 37(5): 669-683. <https://doi.org/10.1177/0894439318791515>
- Bosch, O.J., & Revilla, M. (2022). When survey science met web tracking: presenting an error framework for metered data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. <https://doi.org/10.1111/rssa.12956>
- Bosch, O.J., Sturgis, P., & Kuha, J. (2022). Track me but not really: Tracking undercoverage in metered data collection. AAPOR Annual Conference, 11th-13th May 2022.
- Iglesias, P., & Revilla, M. (2023). Skills, availability, willingness, expected participation and burden of sharing visual data within the frame of web surveys. *Quality and Quantity*.
- Iglesias, P., Ochoa, C., and Revilla, M. (2022). What do I do with these images? A practical guide to the classification of images sent by survey participants, RECSM Webinar, 31st March 2022. Available at: <https://www.youtube.com/watch?v=IQoKbO4XsZI>
- Ochoa, C. (2023 ). Researching the moment of truth: An experiment comparing in-the-moment and conventional surveys to investigate online job applications. UPF PhD workshop (March 20-21).
- Revilla, M., Iglesias, P., Ochoa, C., & Antón, D. (2022). WebdataVisual: a tool to gather visual data within the frame of web surveys. OSF. <http://doi.org/10.17605/OSF.IO/R7CAX>
- Revilla, M., Ochoa, C., Iglesias, P., Antón, D. (2022). WebdataNow: a tool to send in-the-moment surveys triggered by passive data. OSF. <http://doi.org/10.17605/OSF.IO/G3MSC>.