


**Dados de Laplace**






| Javier Álvarez Liébana

Follow @DatosdeLaplace

29.4K followers

28 Jul, 24 tweets, 12 min read

Maybe Scrolly?

Bookmark

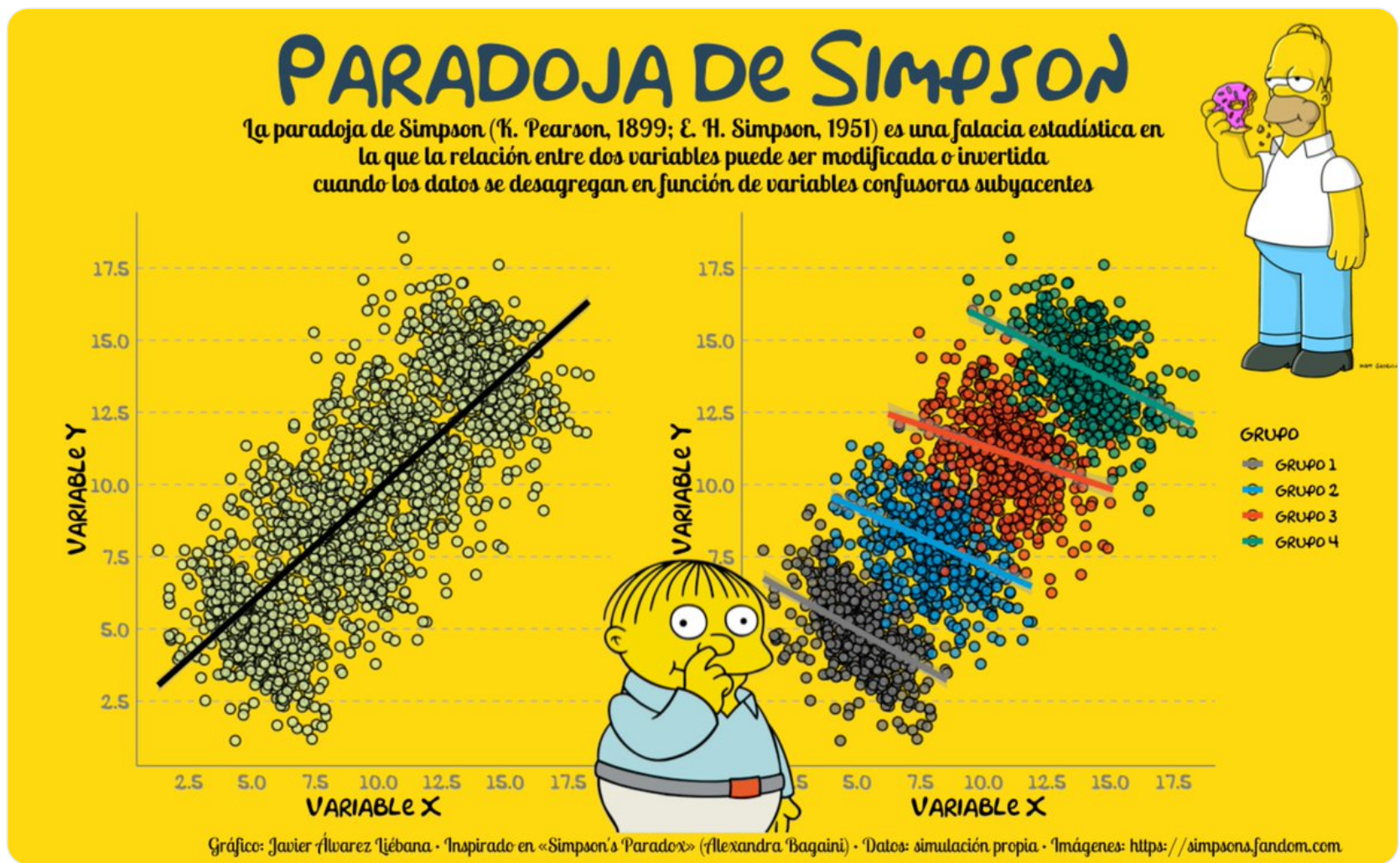
Save as PDF

My Authors

La paradoja de Simpson es una de las falacias estadísticas más comunes en medicina y sociología: la tendencia de datos agregados puede ser contraria a la de los datos desagregados


¿Por qué es importante conocerla para combatir el negacionismo de las vacunas?

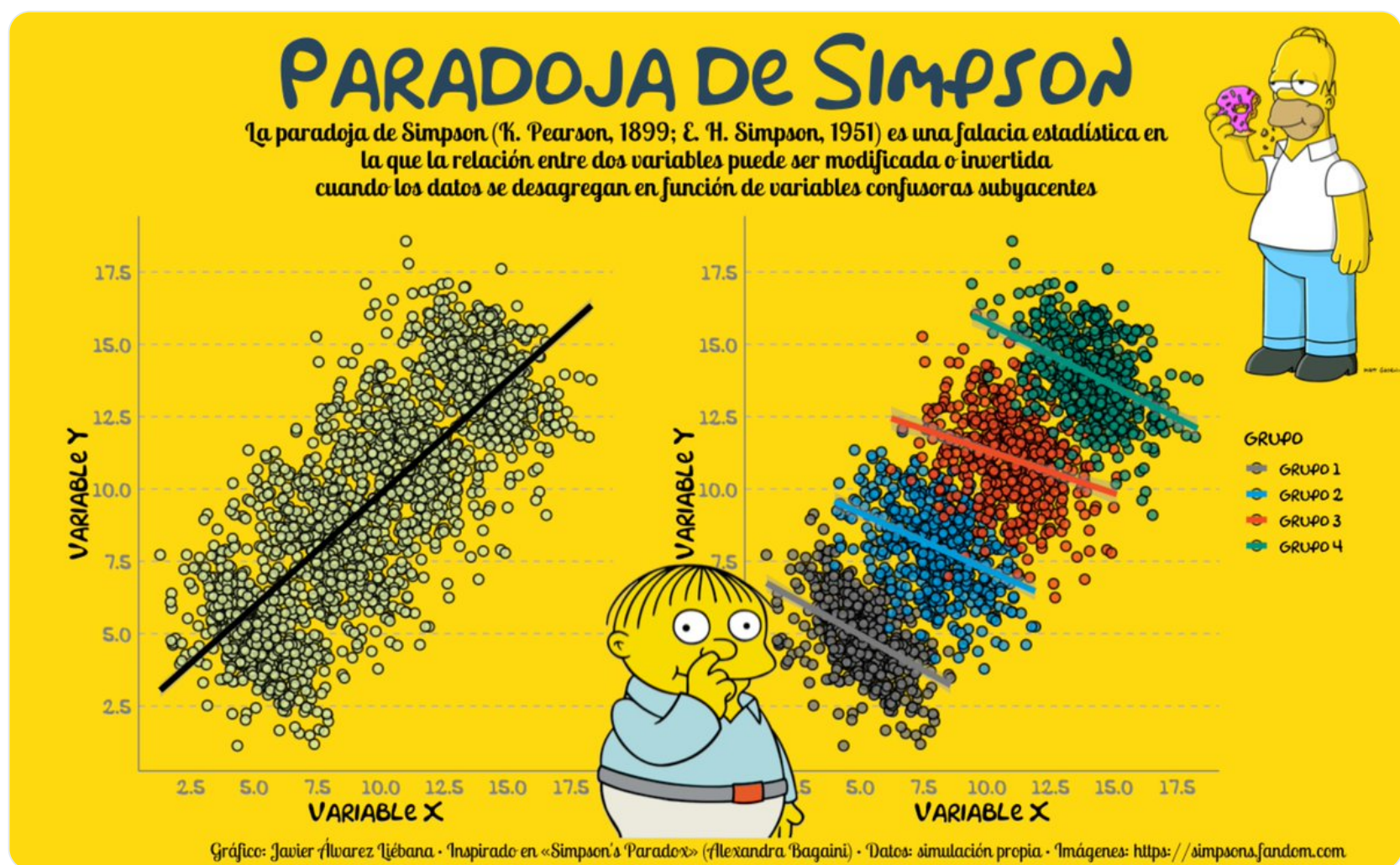
Breve hilo🧵👉



1/ Aunque no lleve su nombre, la paradoja ya fue observada por los estadísticos Karl Pearson (1899) y Udny Yule (1903)

[royalsocietypublishing.org/doi/pdf/10.1098/rsta.1903.0001](https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1903.0001)

Su nombre actual se lo puso Blyth en el 1972, paradoja de Simpson, por el famoso artículo de E. Simpson   [jstor.org/stable/2984065](https://www.jstor.org/stable/2984065)



2/ En dicho artículo Simpson se centra en algo muy simple: la media de la suma no es la suma de las medias 🤖

💡 Imagina las cuentas de un restaurante: una mesa de 2 personas de 40€ y otra de 3 personas de 30€

Si hacemos la media total nos sale  $70/5 = 14€$  por cabeza

Pero...

🤔 ¿Qué sucede si hacemos la media de cada mesa por separado (desagregamos) y luego realizamos la media de esas medias?

Resulta que entonces tenemos  $40/2 = 20€$  y  $30/3 = 10€$ , de forma que si miramos las estadísticas desagregadas, la media nos sale de  $(20 + 10) / 2 = 15€$

3/ La paradoja de Simpson es una falacia y error muy común en estudios médicos y sociológicos

Veamos el ejemplo con la gráfica: he generado datos aparentemente aleatorios, de forma que cuando los vemos todos juntos parecen tener un patrón

Según crece la variable X ==> crece Y

❏ ¿El problema?

Que dichos datos no son totalmente aleatorios: he generado 4 grupos de normales, que he pegado lo suficiente para que parezcan del mismo conjunto de datos pero en realidad tenemos 4 grupos distintos

La paradoja de Simpson sucede cuando al analizar los datos desagregados por grupos resulta que las conclusiones estadísticas obtenidas entre la variable X y la Y son OPUESTAS

En lugar de tener una relación creciente tenemos en todos casos relaciones decrecientes muy débiles

4/ ¿Por qué es tan importante en medicina? 🧐 🏥 🧑

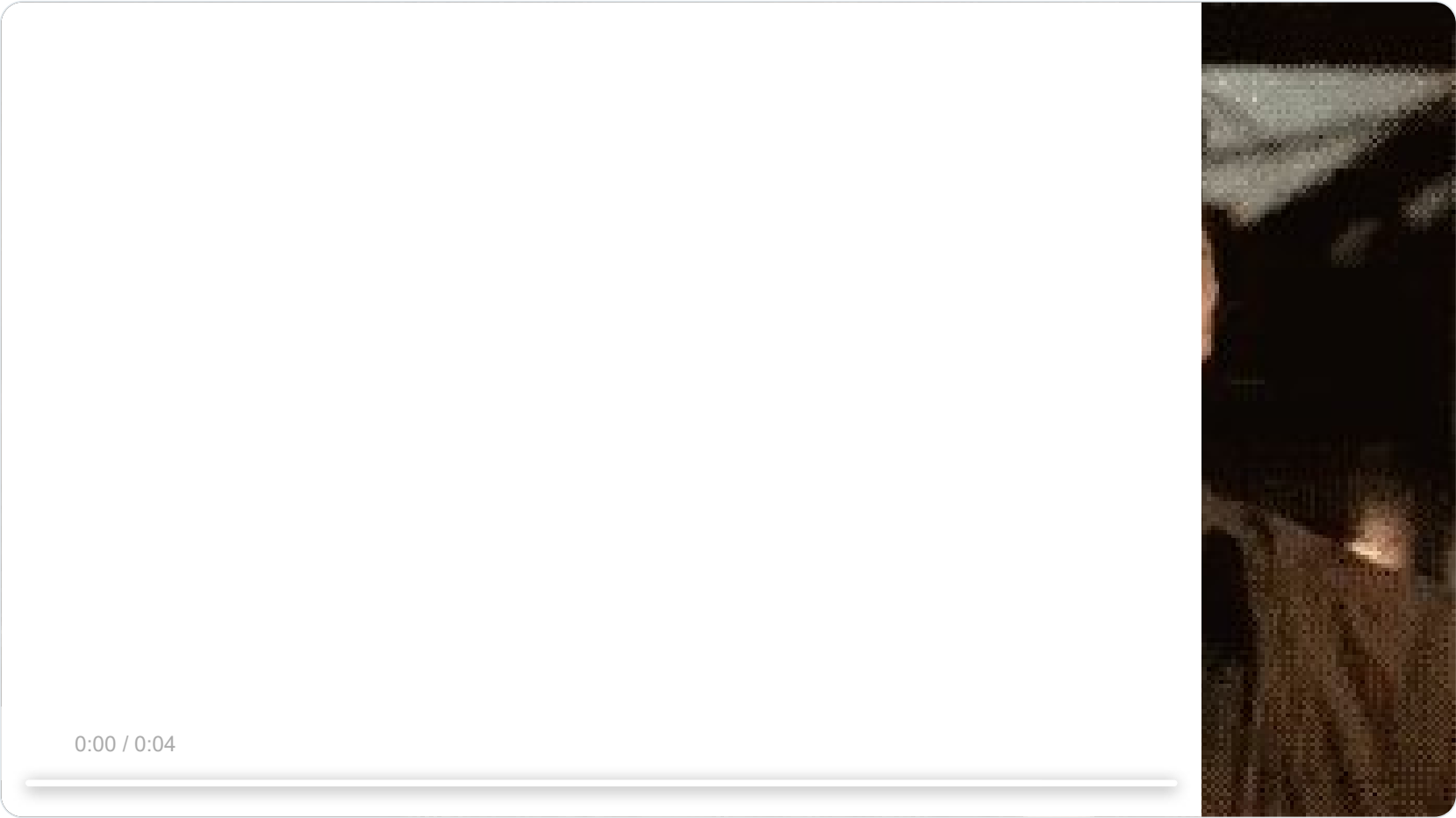


Imagina que comparamos dos tratamientos para eliminar cálculos renales: **A** incluye cirugía invasiva y el **B** no invasiva

- A** 78% (273/350) de éxito
- B** 83% (289/350) de éxito

Parece obvio que el mejor es el **B**...

¿No?

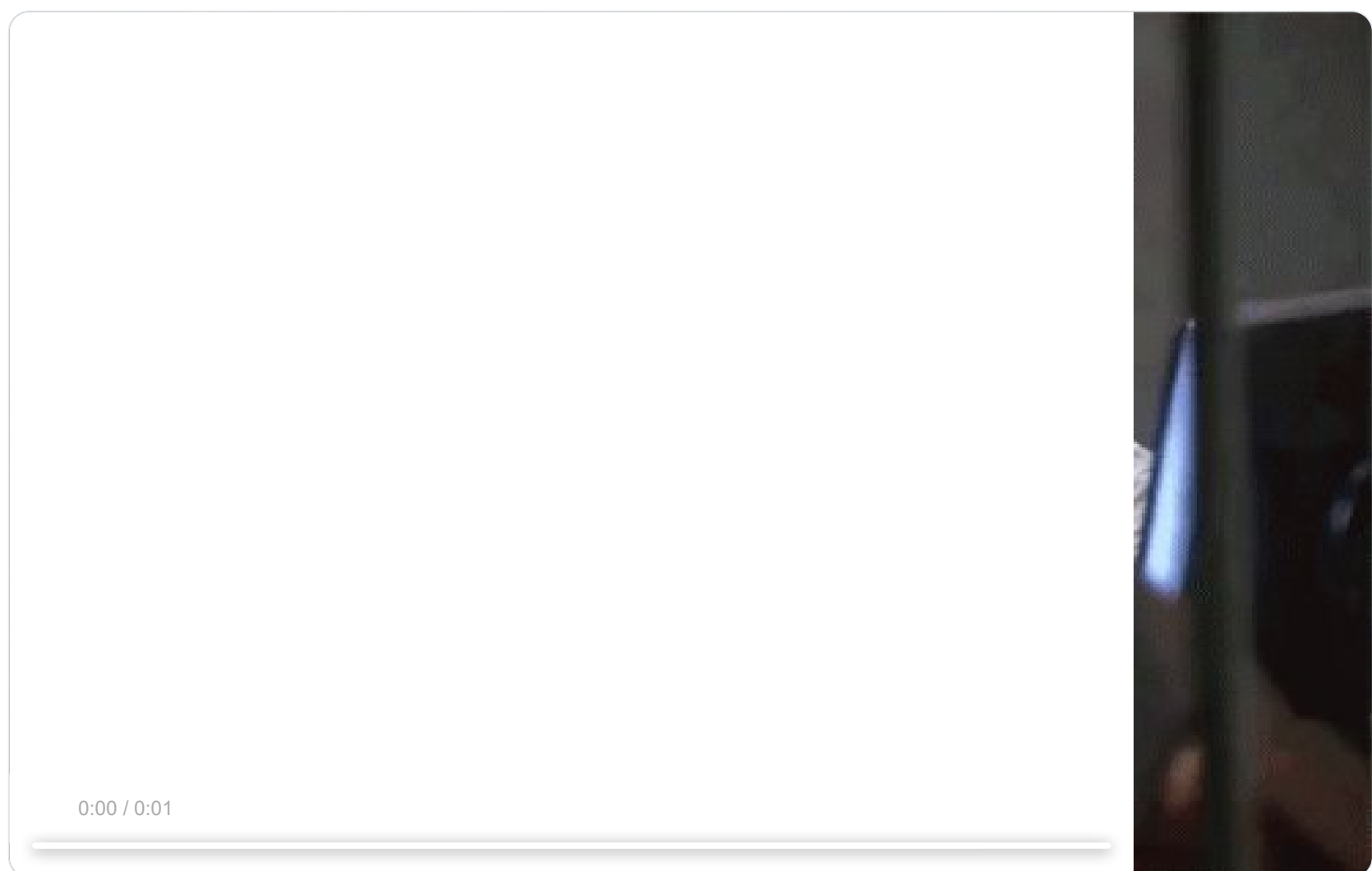


¿Pero qué sucede cuando consideramos el tamaño de los cálculos como nueva variable: grande **↑** vs pequeño**↓**?

- A** 93% (81/87) para **↑**/ 73% (192/263) para **↓**
- B** 87% (234/270) para **↑**/ 69% (55/80) para **↓**

Cuando desagregamos por tamaño

¡El tratamiento **A** es mejor en cada subgrupo!



La paradoja es que cuando agregamos los datos, sin tener en cuenta el efecto del tamaño del cálculo (variable confusora), obtenemos conclusiones opuestas sobre la efectividad de cada uno

Conclusiones que sin desagregar de forma adecuada nos llevarían a resultados erróneos


Esto sucede por dos motivos:

- 1) El efecto del tamaño eclipsa al tratamiento (influye más la causalidad de la gravedad que  $\boxed{A}$   $\boxed{B}$ )
- 2) Los grupos no tienen el mismo número: los grupos ( $\boxed{A}$ ,  $\downarrow$ ) y ( $\boxed{B}$ ,  $\uparrow$ ) tienen mucho más peso que ( $\boxed{A}$ ,  $\uparrow$ ) y ( $\boxed{B}$ ,  $\downarrow$ ), peso distorsionado al agregar

5/ ¿Por qué conocer esta paradoja nos puede ayudar contra el negacionismo vacunal? 🖋️

Hace dos semanas me preguntó [@elena\\_turri](#) para [@Newtral](#) si tenía sentido una afirmación que rulaba

«Hay un 885% más probabilidad de morir por COVID-19 entre vacunados»



Un informe británico no demuestra que todos los vacunados tienen “885% más probabilidad de morir” por COVID... Las vacunas sí están funcionando contra la COVID-19, al contrario de lo que se afirma en un contenido viral, en el que se sacan de contexto unas cifras de un informe del del organismo Public Health En...

<https://www.newtral.es/vacunados-probabilidad-morir-por-covid-informe-inglaterra/20210714/>

Dicho % ha sido estas semanas propagado por los negacionistas usando como fuente el informe del 25 de junio del PHE en Inglaterra GB 🖋️🔗

En dicho informe se incluye esta tabla de contingencia de casos y fallecidos, de vacunados y no vacunados

En dicha tabla se observa como

7235 infectados por delta estaban --> 50 de ellos fallecidos (0.691% letalidad)  
53 822 infectados por delta NO lo estaban --> 44 de ellos fallecidos (0.081% letalidad)

¡Mayor letalidad entre vacunados!

¿Dónde está el truco?

¿Qué sucede si en lugar de calcular con datos agregados los desagregamos por EDAD?

<50a CON : 3689 casos, 0 (0%)  
<50a SIN : 52846 casos, 6 (0.011%)

>=50a CON : 3546 casos, 50 (1.410%)  
>=50a SIN : 976 casos, 38 (3.893%)

Chorprecha

El problema de analizar los datos vacunales sin desagregar por tramos etarios es que se obvia la paradoja de Simpson, obviando una variable confusora «oculta» con un efecto causal clave en los fallecidos: la variable edad

Las vacunas reducen drásticamente la letalidad

6/ La paradoja de Simpson es un error muy habitual en los estudios sociológicos y en los análisis de rendimiento deportivo

El ejemplo más conocido es el del promedio de bateo de los jugadores de béisbol Derek Jeter y David Justice

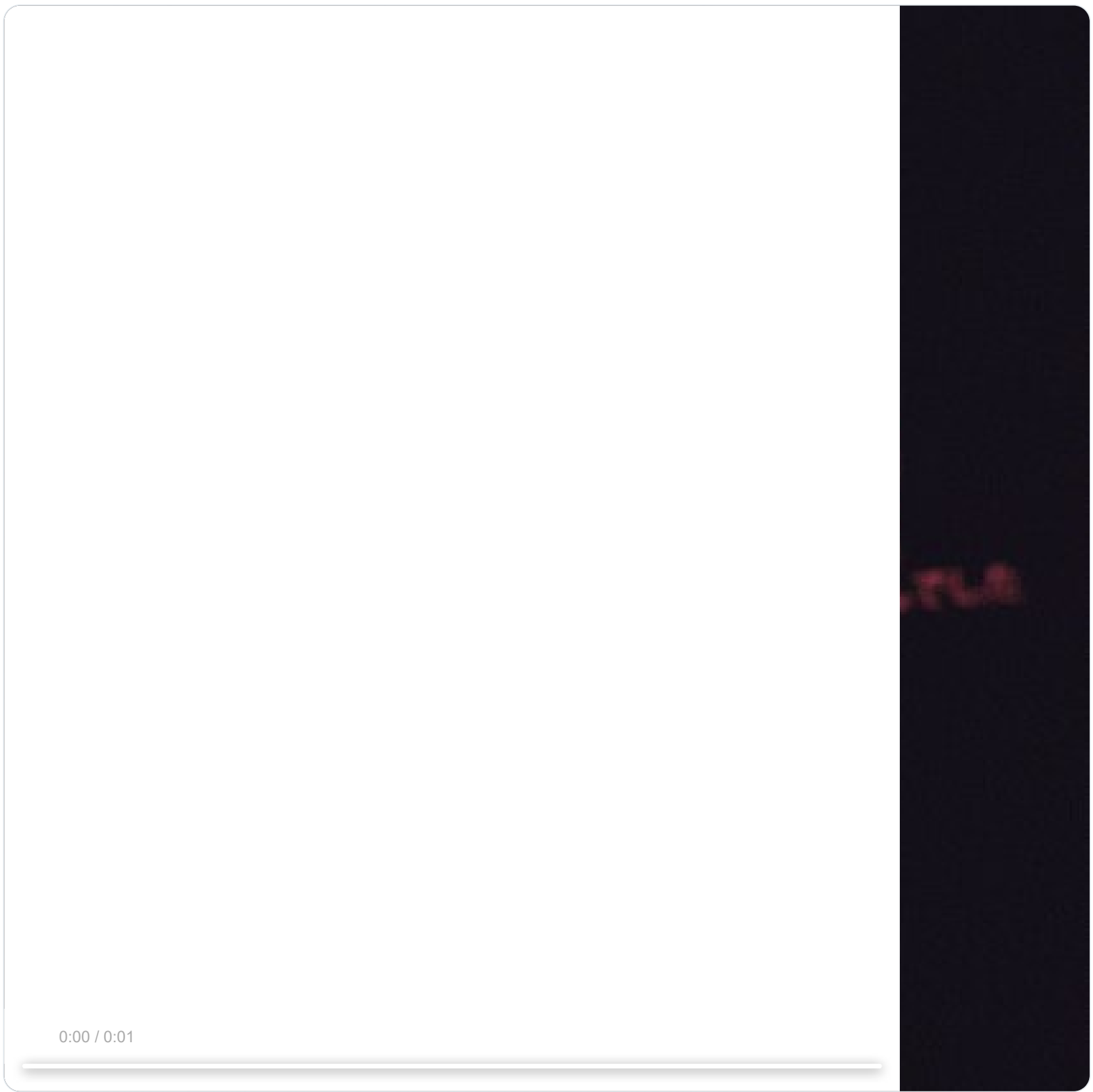
El matemático Ken Rosse estudió su promedio de bateo durante los años 1995 y 1996

Jeter: 12/48 (25%) en el 95, 183/582(31.4%) en el 96  
Justice: 104/411 (25.3%) en el 95, 45/140 (32.1%) en el 96

Sin embargo en el total  
Jeter: 27% de acierto  
Justice: 31%

Tanto en 1995 como en 1996, en ambos años POR SEPARADO, Justice tuvo un promedio de bateo más alto que Jeter

Sin embargo, al analizar los datos agregados (con frecuencias de bateo muy distintas en cada año), da la sensación errónea de que Jeter tuvo mayor que acierto que Justice



Hasta aquí la importancia de conocer esta paradoja/falacia estadística

Si te ha gustado, un RT al inicio del hilo se agradece ♡ 😊

📊 Todos los gráficos están realizados con R, sin ningún tipo de edición posterior [#dataviz](#) [#rstats](#)

💻 Código R: [github.com/dadosdelaplace...](https://github.com/dadosdelaplace/hilostwitter)

dadosdelaplace/  
**hilostwitter**

los hilos de Twitter (R codes of Twitter activity)  
https://twitter.com/dadosdelaplace

0 Issues 7 Stars 2 Forks

GitHub – dadosdelaplace/hilostwitter: Códigos R de los hilos de Twitter (R codes of Twitter activity) https://twitter.com/dadosdelaplace – GitHub – dadosdelaplace/hilostwitter: Códigos R de los hilos de Twitter (R codes of Twitter activity) https://twitter.com/dadosdelaplace

<https://github.com/dadosdelaplace/hilostwitter>

PD: ha sido un año intensito, mil gracias por la cantidad de gente que somos en este proyecto de divulgación estadística y en Telegram ([t.me/dadosdelaplace](https://t.me/dadosdelaplace))

Me tomaré vacaciones en agosto así que quién quiera saber de mis andanzas, en IG me tenéis

🐦 Follow Us on Twitter!

🐦 Tweet

📌 Share

[instagram.com/javieralvarezliebana](https://www.instagram.com/javieralvarezliebana)

Dados de Laplace

Canal de Telegram de la comunidad Laplaciana. Más divulgación: Instagram (instagram.com/javieralvarezliebana) y Twitter (https://twitter.com/i/events/1345095058882785281). Dropbox con material: https://t.me/dadosdelaplace

PD2: os dejo hilos que he realizado sobre otras falacias estadísticas clásicas

Gerrymandering (precisamente importante por la paradoja de Simpson):

Unroll available on Thread Reader

Dados de Laplace Δ( ) | Javier Álvarez Liébana

@DadosdeLaplace

Se acercan elecciones EE.UU.

Con permiso de @Nanisimo #Destino3N, el hilo hoy va de estadística y elecciones: el GERRYMANDERING

Y sí, también de arte, con permiso @elbarroquista y #TwitterCultural, del retratista de G. Washington

[instagram.com/javieralvarezliebana](https://www.instagram.com/javieralvarezliebana)  
#MatesEnDomingo

4:34 AM · Sep 20, 2020

211

9

Share this Tweet

Sesgo del superviviente:

Unroll available on Thread Reader

Dados de Laplace Δ( ) | Javier Álvarez Liébana

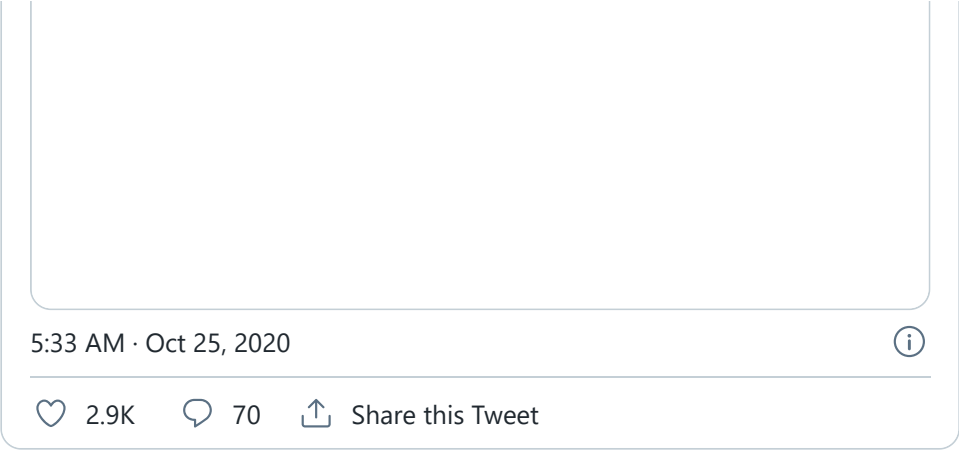
@DadosdeLaplace


SEGUNDA GUERRA MUNDIAL

La Marina Real era fuertemente golpeada por el ejército nazi, y buscaban mejorar sus aviones analizando los impactos balísticos

¿Qué partes del avión reforzarías tú?

Hilo con la historia del SESGO DEL SUPERVIVIENTE  
#MatesEnDomingo



 Regresión a la media:

Unroll available on Thread Reader



\* Ni un hilo sin su errata ☐

Los % de éxito de bateo agregados están intercambiados: Sin embargo en el total

Jeter: 31% de acierto

Justice: 27%

Gracias [@Vdot\\_Spain](#) por la corrección :)

Añado para obligado visionado

El vídeo de [@AnaBayes](#) donde cuenta en [@AlFerPodcast](#) el caso del proceso de admisión de la Universidad de Berkeley, y un caso muy actual: la comparación de mortalidad entre hospitales



LO QUE LOS DATOS ESCONDEN: ANABEL NOS CUENTA LA PARADOJA DE SIMPSON - ALFERPODCAST #2 ...


...


Missing some Tweet in this thread? You can try to [force a refresh](#)

 Tweet


 Share


 Email


 Keep Current with [Datos de Laplace !\[\]\(f5126919fb264baa65afc980ba29ad65\_img.jpg\) | Javi...](#)




Stay in touch and get notified when new unrolls are available from this author!


 Add to "My Authors"

 Read all threads


 This Thread may be Removed Anytime!




Twitter may remove this content at anytime! Save it as PDF for later use!

 Save this thread as PDF

Try unrolling a thread yourself!





Tweet your reply





Andrew Chen @andrewchen · Dec 27


12/ One of the best things out of @ericries's Lean Startup is the concept of "vanity metrics" - unimportant, inflated metrics that don't have much to do with business value. But the effects of using them are different when you're a startup versus a big co.

 1

 8

 52





Andrew Chen @andrewchen · Dec 27

13/ For big, established companies, focusing on vanity metrics is silly but not

1. Follow [@ThreadReaderApp](#) to mention us!


2. From a Twitter thread mention us with a keyword "unroll"

[@threadreaderapp](#) unroll

[Practice here](#) first or read more on our [help page](#)!

More from @DatosdeLaplace





**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)

26 Jul

Madresita, más de 24k por aquí y 1500 por Telegram, infinitas gracias 😊❤️ Os dejo un top 5 de hilos de estadística y matemáticas, de resumen de «temporada» 😊

Read 5 tweets





**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)

16 Jul

Es recurrente leer «es imposible ya que tiene probabilidad 0»  
Un evento con probabilidad 0 puede suceder La probabilidad de que midas exactamente 170.000 cm, si pudiésemos disponer de precisión infinita, es 0 Y la de 165.428 cm Y la de

Read 9 tweets



**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)

1 Jul

Hace una semana el INE publicó la encuesta anual de estructura salarial, donde se incluye la distribución salarial  
● Moda: 18 489 € brutos anuales ● Mediana: 20 351 €  
● Media: 24 395 € ¿Por qué es importante diferenciarlas para datos asimétricos como los salarios?👉

1/ Esta gráfica es algo así como un diagrama de barras pero con los puntos unidos con una línea continua (función de densidad) Es asimétrica porque el área debajo de la curva es

Read 12 tweets




**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)


30 Jun

Me han gustado muchos los gráficos que ha hecho @CedScherer sobre las anomalías térmicas que vive estos días Portland (Canadá) Así que me he animado y he hecho rápidamente en plan «guarriplei» algunas gráficas similares sobre algunas ciudades en España es 📊👉 #dataviz

1) Evolución anual de las temperaturas diarias máximas coloreadas por décadas ● 1960-1970 ● 1970-1980 ● 1980-1990 ● 1990-2000 ● 2000-2010 ● 2010-2020 ●

Read 10 tweets



**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)

27 Jun

📅 ¿Hasta qué punto influye tu día de nacimiento? 😊 Hace semanas @kikollan analizaba su influencia en deportistas pero... 📅 ¿Y si pudiera incluso mandarte a la guerra? Hablemos del «Draft Lottery» del 69, el sorteo que mandó a Vietnamus a miles de personas 📊👉 #MatesEnDomingo

📖 Algo de contexto En 1955 los EE.UU. us se embarcaron en la guerra de Vietnam para impedir su reunificación

Read 35 tweets



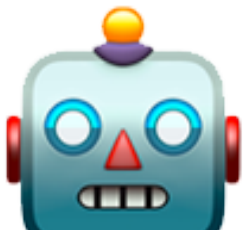
**Datos de Laplace**  | **Javier Álvarez Liéba...**  
[@DatosdeLaplace](#)

17 Jun

Siempre estamos con la turra @elartedeldato @dr\_xeo o servidor de la importancia de la visualización de datos📊 en cualquier análisis estadístico 📊 ¿No basta con hacer cuentas y medias? 😊 Dejadme que os hable del cuarteto más famoso de la estadística: el cuarteto de Anscombe 📊👉

«Toxicidad fuera, mala correlación fuera, me llamas outlier, te doy la mano» (necesitaba hacer el chiste, perdón) El

Read 26 tweets




Did Thread Reader help you today?

Support us! We are indie developers!

This site is made by just two indie developers on a laptop doing marketing, support and

Become a Premium Member (\$3/month or \$30/year) and get exclusive features!

 [Become Premium](#)


Too expensive? Make a small donation by buying us coffee (\$5) or help with server cost (\$10)

 [Donate via Paypal](#)    [Become our Patreon](#)

♥♥ Thank you for your support! ♥♥

[Help](#) | [About](#) | [TOS](#) | [Privacy](#) | [Covid19 Threads](#)

 [Follow Us on Twitter!](#)

 [Tweet](#)  [Share](#)