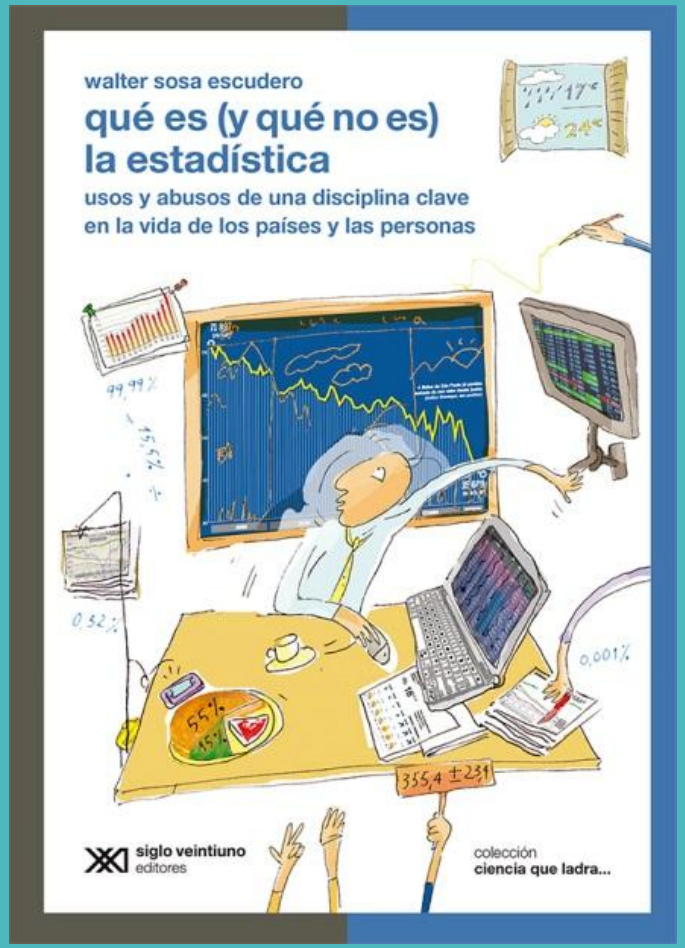
Estadísticamente

El desafío de razonar en términos estadísticos en tiempos de twitter y big data.

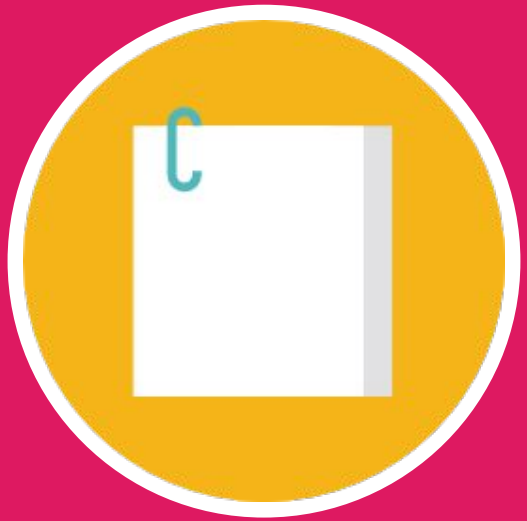
Walter Sosa Escudero | Universidad de San Andrés & CONICET.


walter sosa escudero
qué es (y qué no es) la estadística
usos y abusos de una disciplina clave en la vida de los países y las personas
99,99%
0,32%
0,001%
355,4 ± 23,1
siglo veintiuno editores
colección
ciencia que ladra...

Información nutrimental
Porción 1 manzana grande
(242g / 8 oz.)
Cantidad por porción
Calorías 130
Grasas totales 0g
Grasas saturadas 0g
Grasas trans 0g
Colesterol 0mg
Sodio 0mg
Potasio 260mg
Carbohidratos totale
Sexo: MUJER
19 julio 2012
Metabolismo Lipídico
Resultado
Unidad
221 mg/dl
63 mg/dl
108 mg/dl
100 mg/dl
Proteínas
632
625
611
U. de Chile
NÚMERO DE PARADOS Y TASA DE PARO
Número de parados (escala izquierda)
Porcentaje de población activa en paro (escala derecha
5.000
4.000
3.000
Cambio metodológico
facebook
Facts & Figures
Mission
Timeline
2004
Otro, 10,1
Nu/Nc, 1
Luis Patti, 2,0
Martín Sabatella, 1,7

Razonar Estadísticamente

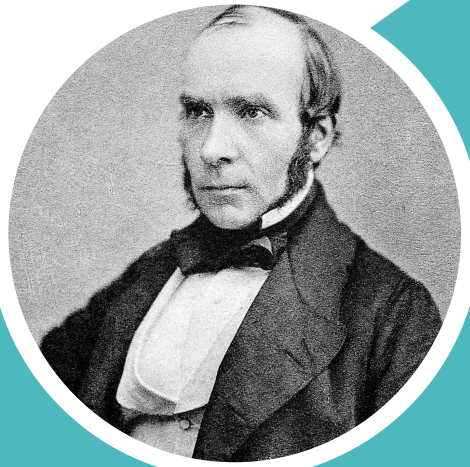
- Técnicas, algoritmos, métodos.
- Estadística como razonamiento.
- Elegir por sobre usar.



Dos casos

La estadística en tiempos de cólera.

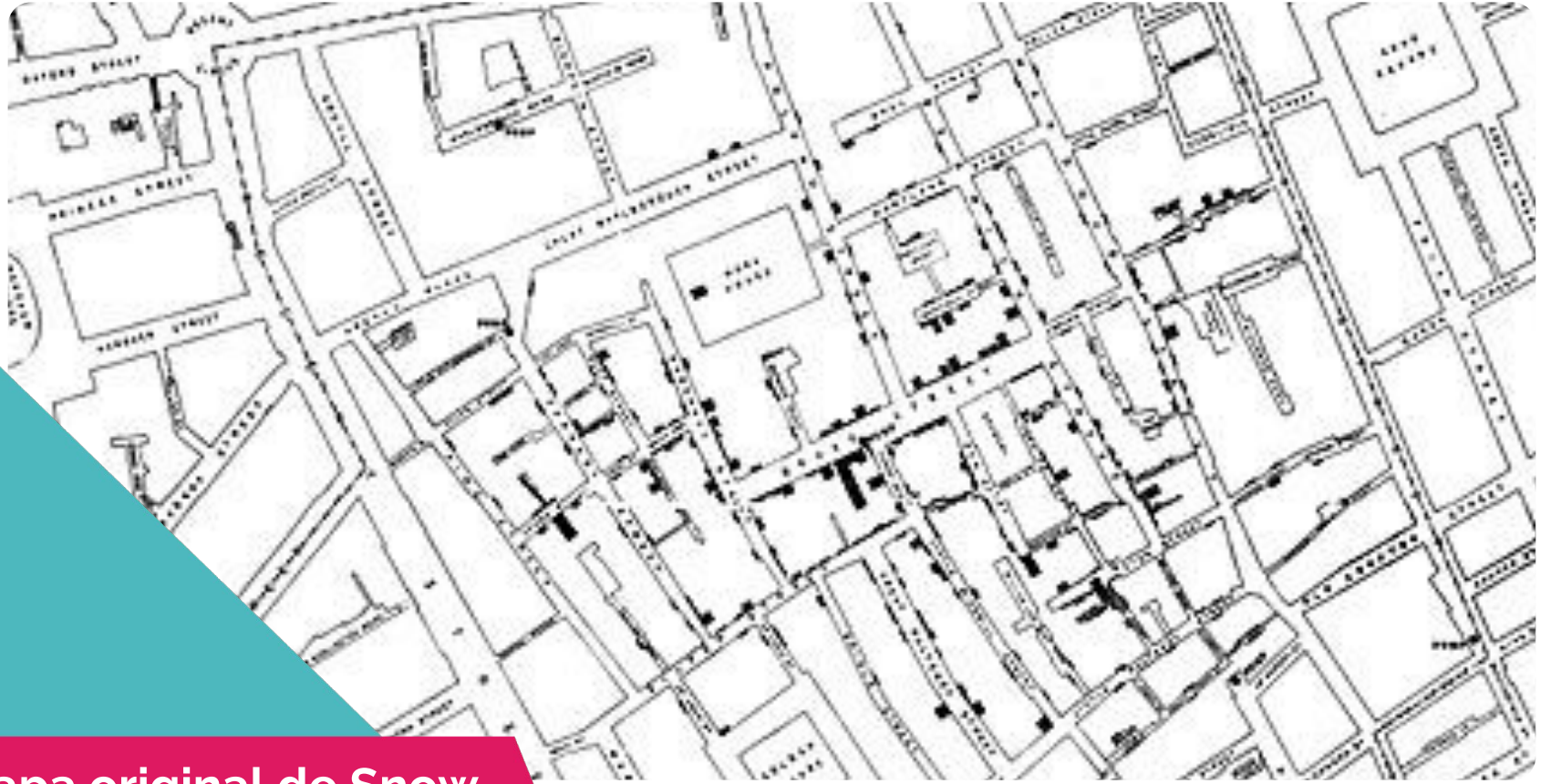
La estadística en tiempos de cólera



John Snow, 1854.
Soho, Londres.

- 31/8: John Snow reporta un brote de cólera.
- 3/9: 127 muertos.
- 6/9: 500 muertos.
- 7/9: **milagro**.

Gran logro de la epidemiología, la salud pública y **la estadística.**



Mapa original de Snow

El Google Map de Snow

Razonamientos

Edward Tufte.
"Joya de la Comunicación Visual"

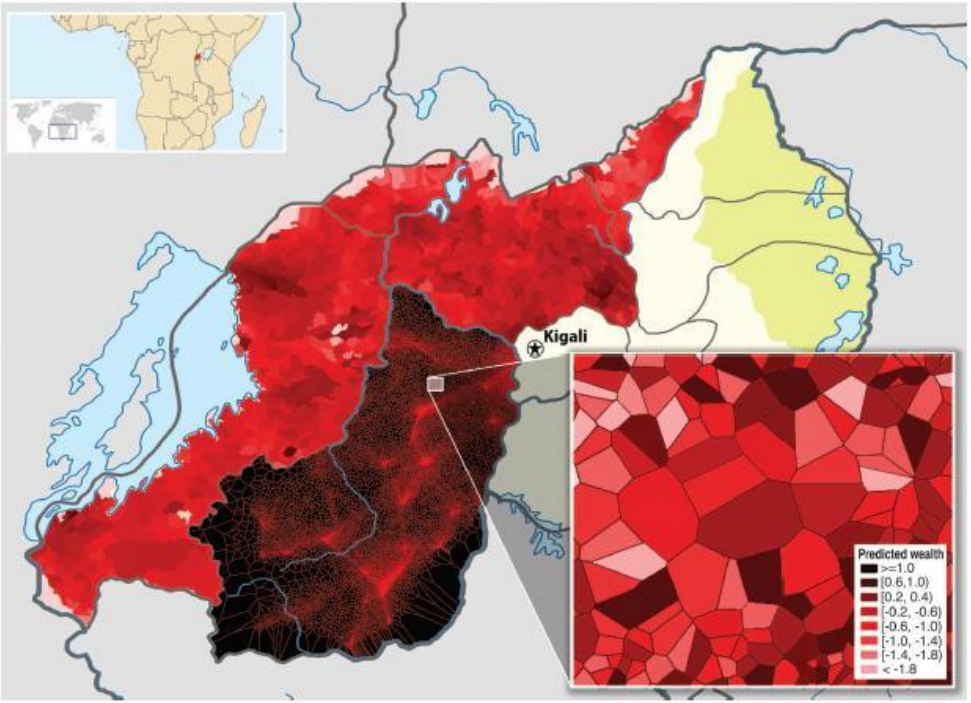
- Argumento estadístico vs. causal.
- Argumento visual.
- Argumento comunicacional.

ECONOMICS

Predicting poverty and wealth from mobile phone metadat

Joshua Blumenstock,[1*] Gabriel Cadamuro,[2] Robert C

Accurate and timely estimates of population characte
and economic research and policy. In industrialized e
enabling new approaches to demographic profiling, b
sources of big data exist. We show that an individual's
be used to infer his or her socioeconomic status. Furt
dicted attributes of millions of individuals can, in t
bution of wealth of an entire nation or to infer th
ed of just a few households. In resource-const
hold surveys are rare, this approach create
formation at a fraction of the cost of tra



Predicción de la Pobreza

La estadistica de lo oculto

Pregunta

En el último año.

¿Ha evadido impuestos?

Sí [ ] **No** [ ]

- Pregunta delicada.
- No respuestas o mentiras.
- Drogas, racismo, evasión, delitos.

¡La estadística al rescate!

Instrucciones

A cada persona le damos un dado y una moneda.

- Tire el dado
- **¿1 o 2?** Tire la moneda. Cara = si. Ceca = no.
- **¿3, 4, 5 o 6?** Responda la verdad.

El método

- W = proporción que responde “sí” a nuestro experimento (mienta o no).

- P = proporción de evasores en el grupo

$$P \approx W * 1.5 - 0.25$$

Ejemplo: 30% responden que si, P ≈ 0.3*1.5-0.25 = 0.2 (20% de consumidores).

Ver para creer

- Encuesta = 2/6 Moneda + 4/6 Verdad
- Despejamos "Verdad":
- Verdad = 6/4 Encuesta – 2/4 Moneda
- Moneda = 1 / 2
- Verdad = 1.5 Encuesta – 0.5 x 0.5
- Verdad = 1.5 Encuesta – 0.25

Computar para creer

```
Experimento sexo, drogas, estadistica

#valor verdadero = 1/3
x<-c(rep(1,30), rep(0,60))
dado<-sample(c(1:6), 90, rep=T)
coin<- sample(c(1,0), 90, rep=T)
rp<-x
rp[dado<3]<-coin[dado<3]
ag<-mean(rp)
1.5*ag-0.25

#Experimento de Montecarlo
expe<-rep(0,1000)
for(i in 1:1000){
 dado<-sample(c(1:6), 90, rep=T)
 coin<- sample(c(1,0), 90, rep=T)
 rp<-x
 rp[dado<3]<-coin[dado<3]
 ag<-mean(rp)
 expe[i]<-1.5*ag-0.25
 }
mean(expe)
```

Razonamientos

- Cambiar el problema.
- Cuestiones complejas, estadísticas complejas.
- Estadísticas, matemática, experimentos.

II.2. LA ECONOMÍA OCULTA EN LA ARGENTINA: EVIDENCIA BASADA EN ENCUESTAS DE GASTO[1]

Walter Sosa Escudero[2]
Verónica Alaimo

1. INTRODUCCIÓN

Mas allá de su significado preciso, el término "economía oculta"[3] sugiere que cualquier intento de medir las actividades que éste designa debe sortear un difícil problema empírico. Ya sea porque hay un esfuerzo explícito de los agentes económicos en que sus actividades no aparezcan

La economía oculta

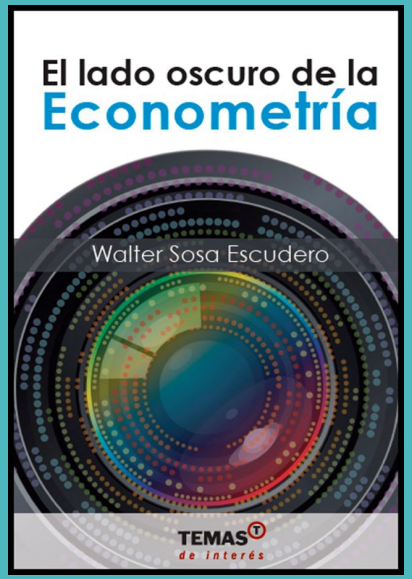
El desafío de
pensar en términos estadísticos

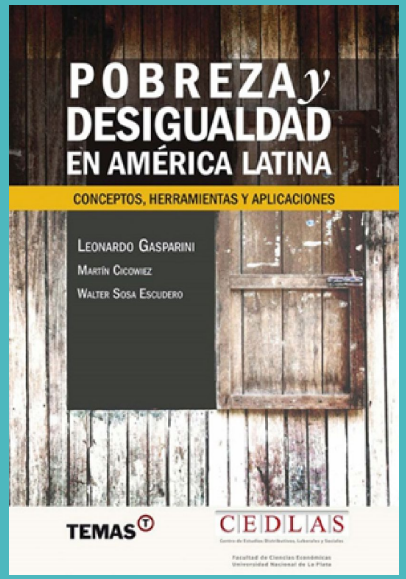
- Box: “Todos los modelos están mal, pero algunos son útiles”.
- Lo bueno, lo malo, lo útil y lo inútil.
- Propósito: determina el método.

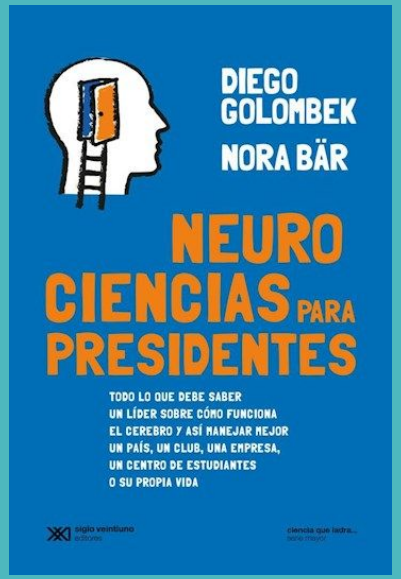
El desafío

“En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el Mapa de una sola Provincia ocupaba toda una Ciudad, y el Mapa del Imperio, toda una Provincia. Con el tiempo, estos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el Tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era **Inútil** y no sin Impiedad lo entregaron a las Inclemencias del Sol y los Inviernos. En los Desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas.”

Jorge Luis Borges, Del Rigor en la Ciencia







walter sosa escudero

databorges

un paseo borgeano por el mundo de la información y las estadísticas en tiempos de big data

enero de 2017

versión preliminar para comentarios. No citar sin autorización previa.
wsosa@udesa.edu.ar

Sugerencias

- **Esta charla:** waltersosa.weebly.com
- **Un libro:** Freakonomics (Dunber y Levitt).
- **Otro libro:** Casualmente (Martin Tetaz)
- **Una película:** Una mente brillante.
- **Otra película:** The Big Short.
- **Una charla TED:** Social experiments to fight poverty (Esther Duflo).
- **Otra charla TED:** Bienvenidos al mundo de la estadistica (WSE).

Walter Sosa Escudero

Profesor Plenario, Director

Departamento de Economía

Universidad de San Andrés

Investigador Principal

CONICET

wsosa@udesa.edu.ar

@wsosaescudero

waltersosa.weebly.com

PhD en Economía (University of Illinois), Lic. en Economía (UBA). Especialista en econometría y estadística aplicada a cuestiones sociales. Director del Departamento de Economía de la Universidad de San Andrés. Investigador principal del CONICET. Autor de numerosas publicaciones académicas, incluyendo artículos en Journal of Econometrics o Econometric Theory, entre otros. Orador en TEDXRioDeLaPlata y autor de Que es (y que no es) la Estadística. Escribe frecuentemente para los principales diarios argentinos. Ha recibido los premios Konex, Fulvio Pagani/ARCOR, Bernardo Houssay, EGSO Award (como el mejor profesor del PhD en Economía de la Universidad de Illinois) y elegido por la Fundación BGH como uno de los 100 argentinos innovadores. Ex presidente de la Asociación Argentina de Economía Política. Miembro de número de la Academia Nacional de Ciencias Económicas.