

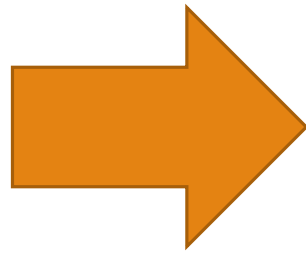
# Data Science para Economía y Negocios

---

CLASE 2

JAVIER FERNÁNDEZ Y ESTEBAN LÓPEZ

# Libro “Guía” BDWS



# big data

walter sosa escudero

¡tercera edición!



breve manual para conocer la ciencia de datos  
que ya invadió nuestras vidas

 **siglo veintiuno**  
editores

**ciencia que ladra...**  
serie mayor

# BDWS-Cap1: Google Flu Trends (GFT)



# ¿ Cómo hacerse cargo de una pandemia?

---

- En el año 2009 surge la gripe A (H1N1) , una pandemia que cobró alrededor de 18.337 victimas en el mundo
- Esto generó un problema de Salud Pública Mundial
- Hasta ese año el monitoreo en EEUU estaba a cargo del “Centro para el control de enfermedades y Prevención de enfermedades) (CDC)
- El CDC generaba un reporte a través de la agregación de la información a nivel nacional
- Este proceso tomaba alrededor de **10 días**

# Solución de Google


---

- Predecir la intensidad de la gripe en tiempo real y con precisión geográfica
- ¿Cómo?
  - Generar un modelo alimentado con:
    - Datos históricos de prevalencia: Semanales entre 2003 -2007 para 9 unidades hospitalarias ->2340 datos
    - Datos de la intensidad de búsqueda de términos y frases relacionados con la gripe

$$\widehat{intensidadgripe} = f(intensidad\_busqueda, otros)$$

Letter | Published: 19 November 2008

# Detecting influenza epidemics using search engine query data

Jeremy Ginsberg, Matthew H. Mohebbi , Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant

GFT

¿Elvis de DS?



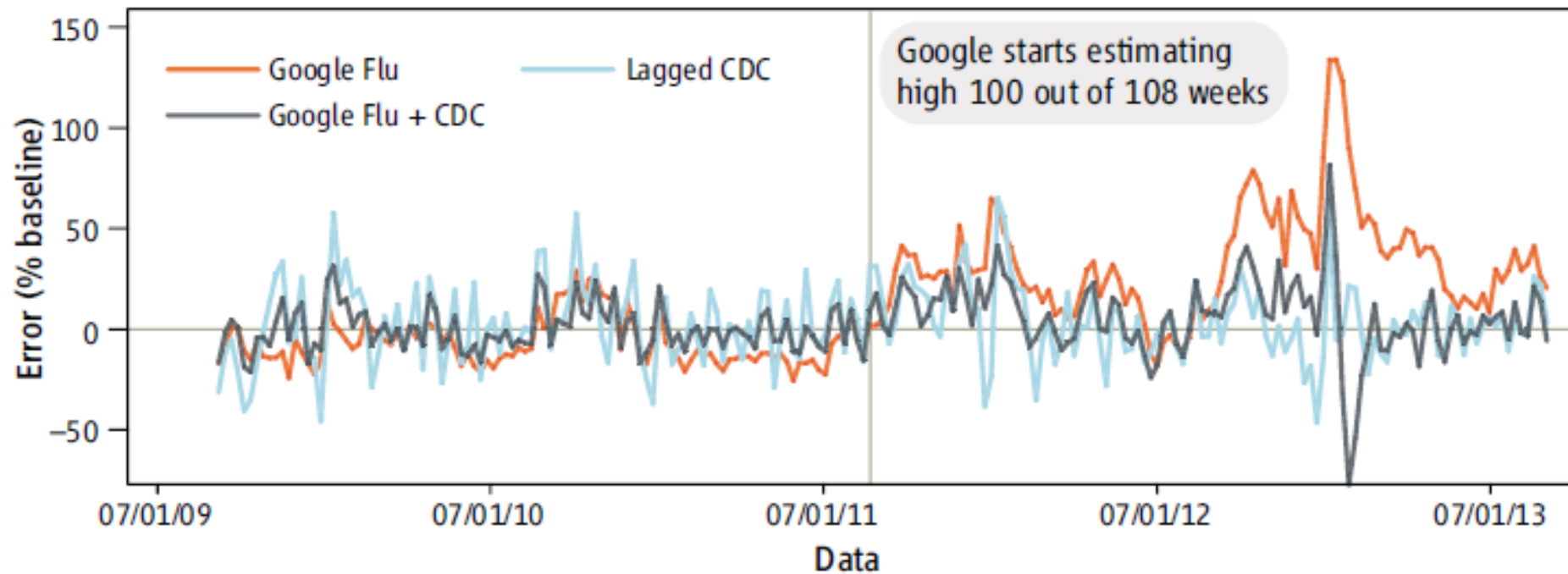
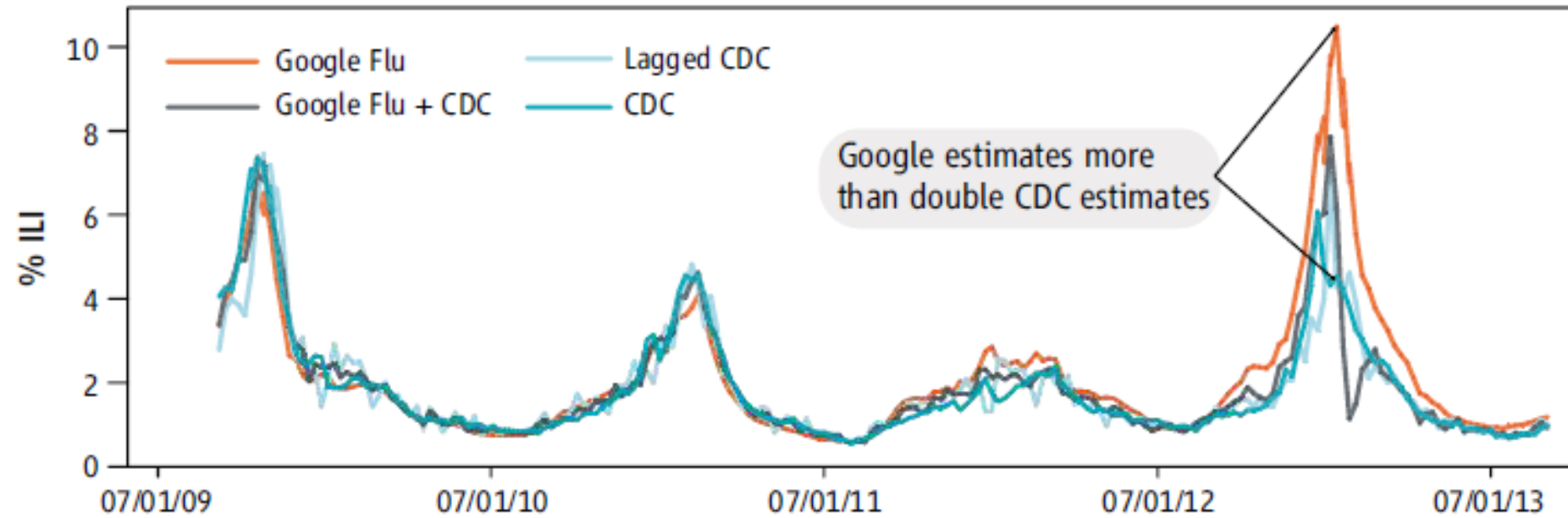
BIG DATA

# The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,<sup>1,2\*</sup> Ryan Kennedy,<sup>1,3,4</sup> Gary King,<sup>3</sup> Alessandro Vespignani<sup>3,5,6</sup>

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.





# Aprendizajes

---

- Ambiciosas estrategias y modelamiento de big data no siempre solucionarán todo.
  - Más no siempre es preferido a menos
  - Es importante encontrar un correlato de los datos con la realidad
  - Tener cuidado con inducir “correlaciones espureas”
- Pero, GFT sí resolvió un problema en un comienzo
  - Redujo mucho el tiempo en que la PP podía tomar acción sobre un posible crecimiento de la epidemia
  - El día de hoy coopera con CDC y podemos intuir que ha ayudado a mejorar los pronósticos de este organismo