



Aprendiendo a ser científic@s de datos

Ana Freire

Científica de Datos

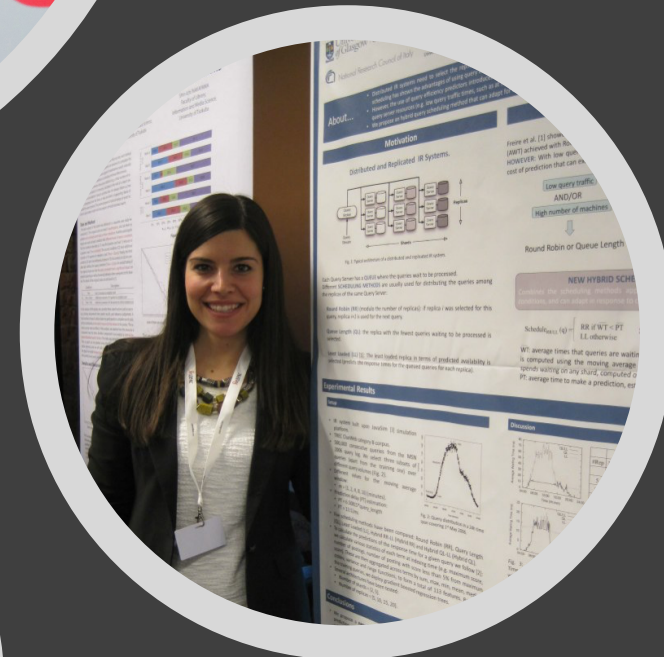


Universitat  
Pompeu Fabra  
*Barcelona*



# Ana Freire

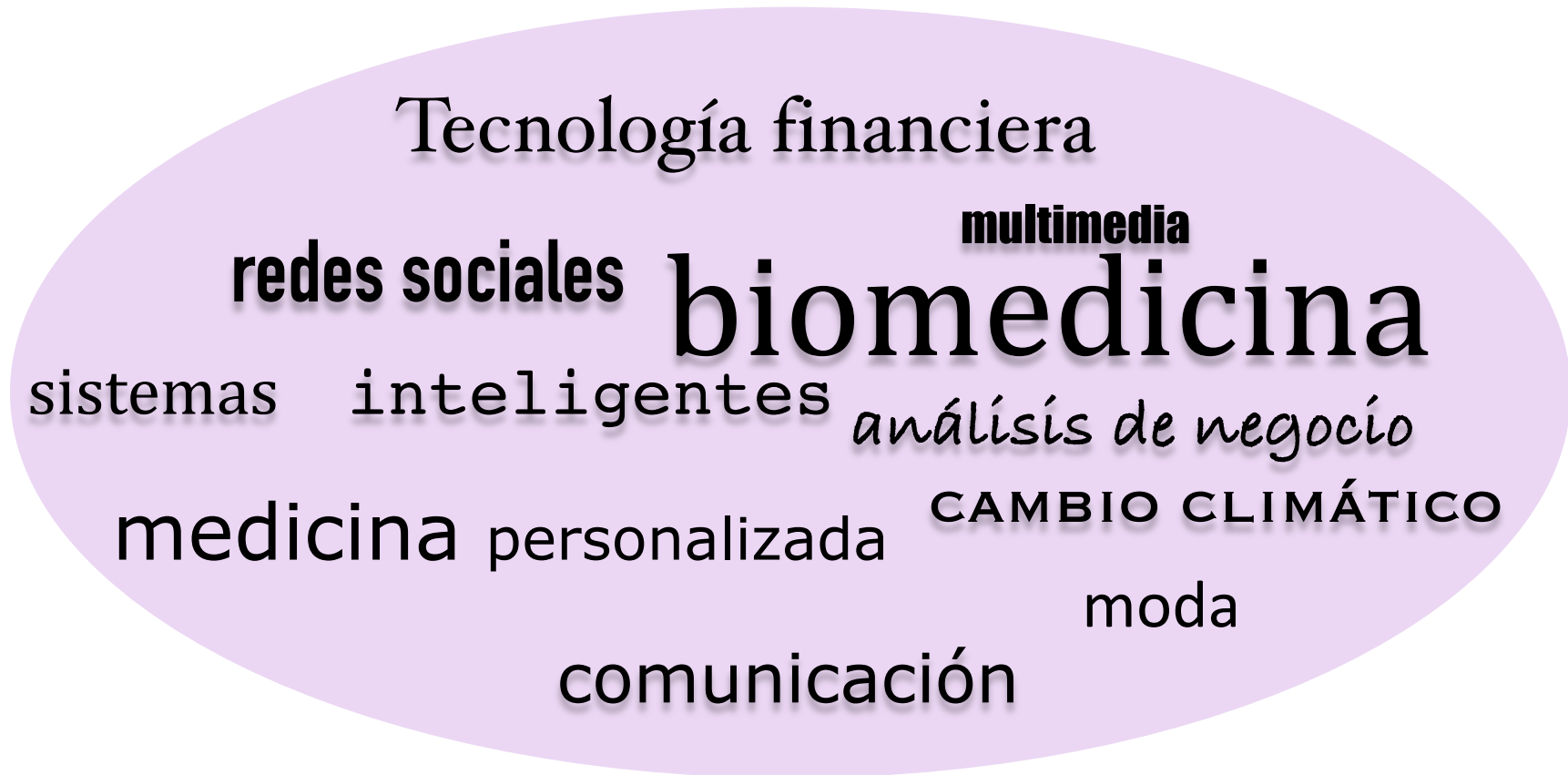
## Doctora en Informática





# Ciencia de Datos

Necesidad de interpretar grandes volúmenes de datos generados por nuestra sociedad



Datos

Procesamiento

Inteligencia



**Matemática + Estadística + Computación + Sociedad**





¿Qué datos se generan  
en la sociedad?

Ciudades inteligentes



¿Qué datos se generan  
en la sociedad?

Hospitales



A silver MacBook Pro is open on a wooden table. The screen shows a blue background with the text "Online Banking" at the top. Below it are two light blue input fields: the first is labeled "Username" with a person icon, and the second is labeled "Password" with a lock icon. A green "Login" button is centered below the fields. A black smartphone is lying on the table to the left of the laptop.

# Online Banking

¿Qué datos se generan  
en la sociedad?

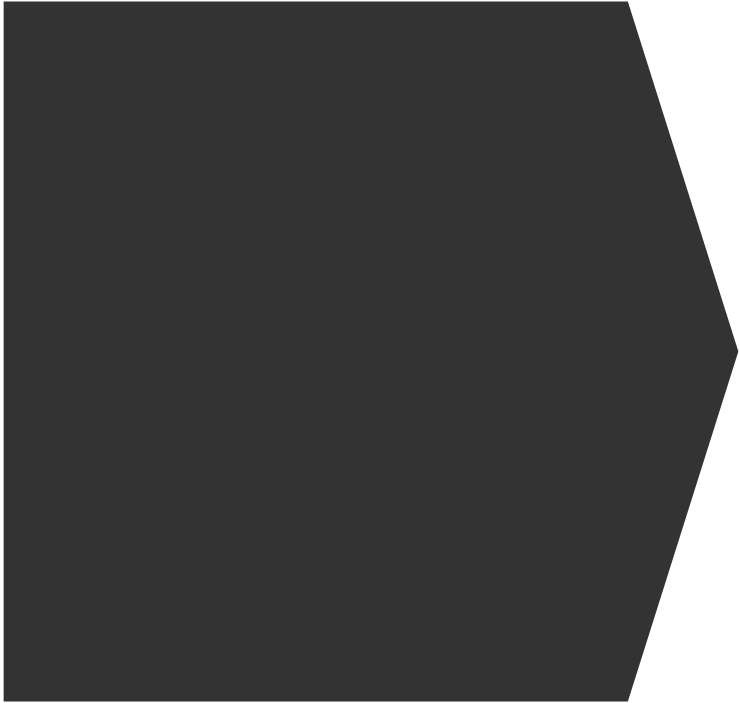
Finanzas



¿Qué datos se generan en la sociedad?

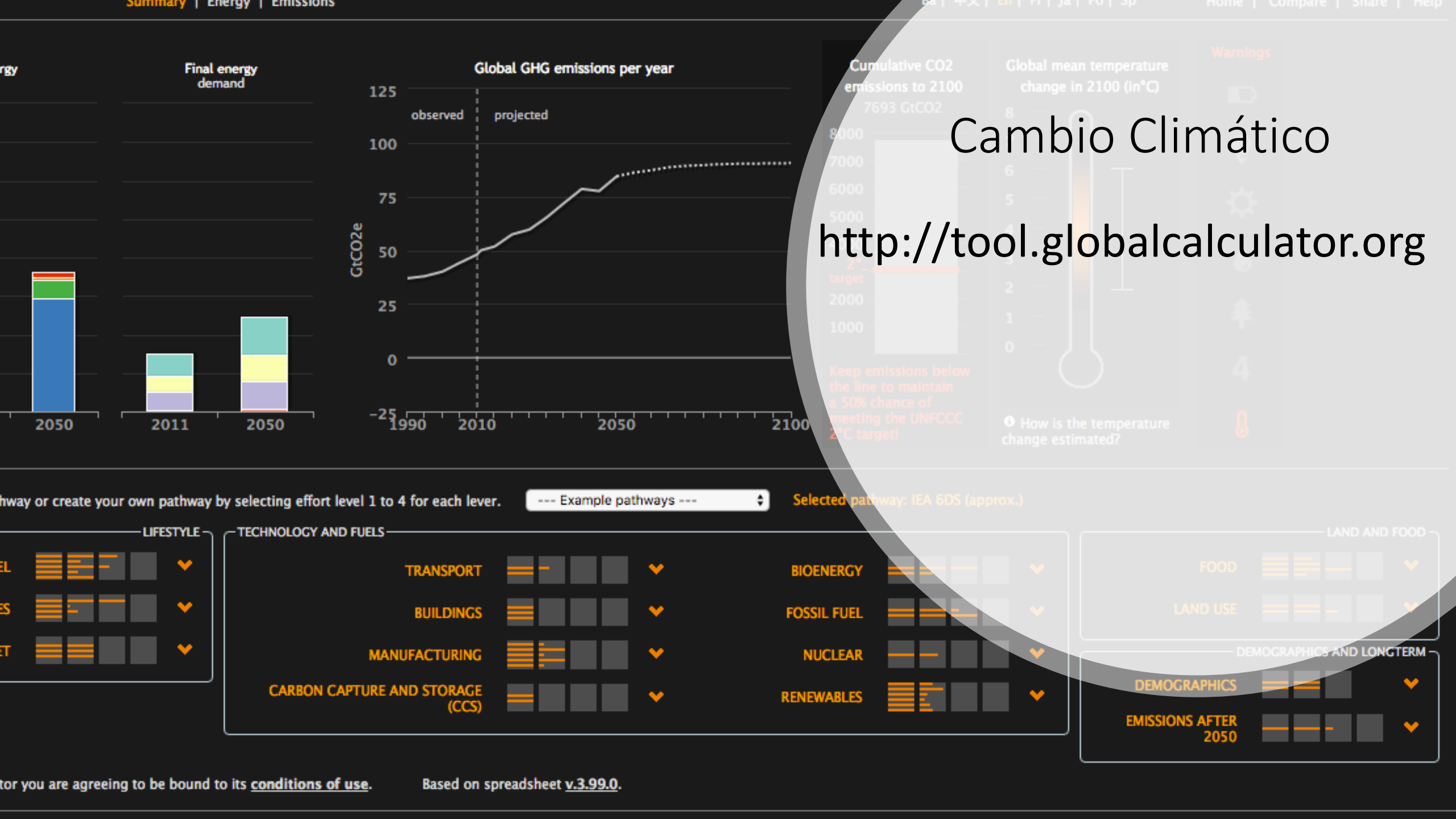
Comportamiento humano





# Cambio Climático

¿Qué podemos hacer para ralentizarlo?



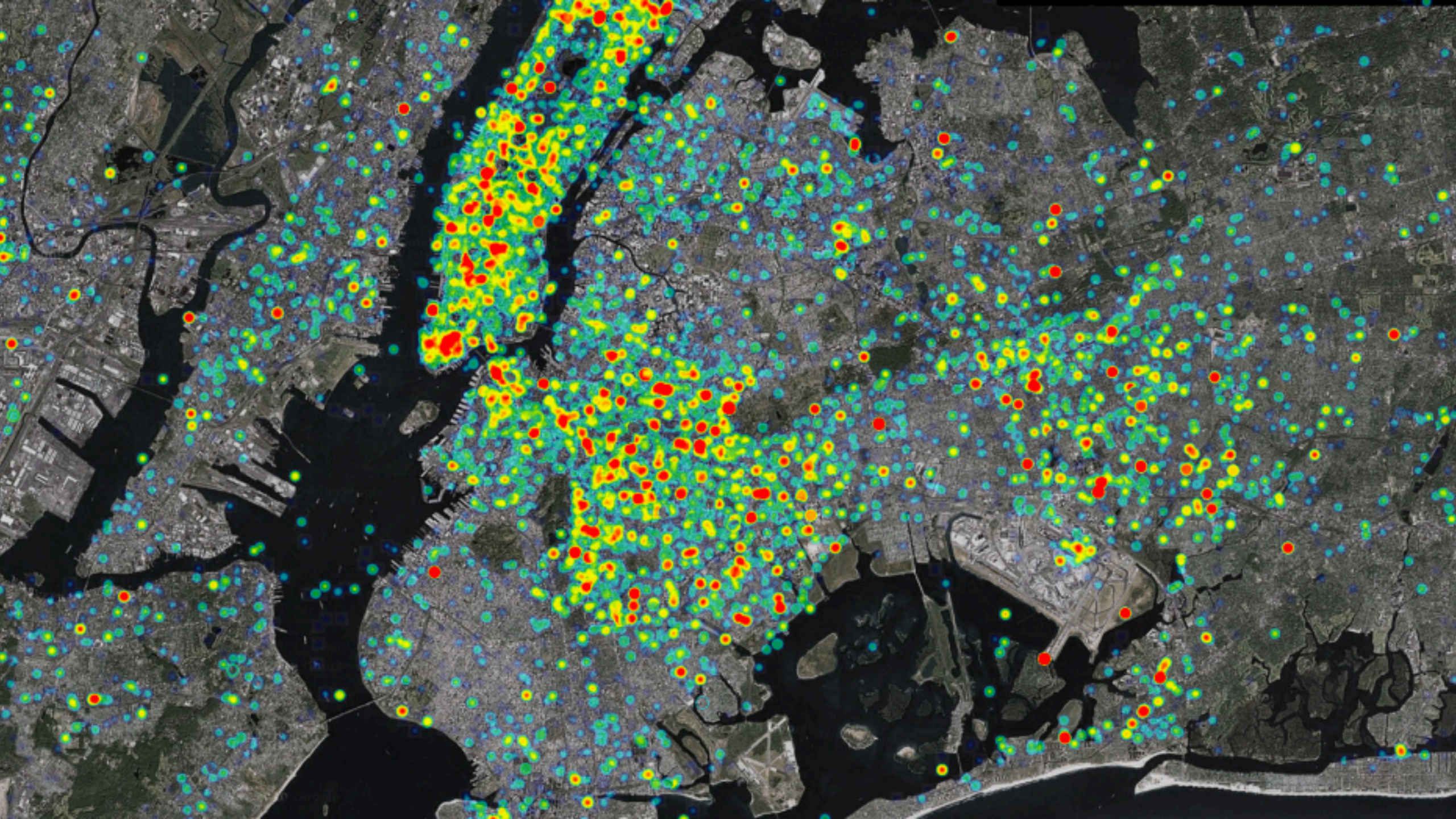




¿Qué datos se generan en la sociedad?

Comportamiento humano

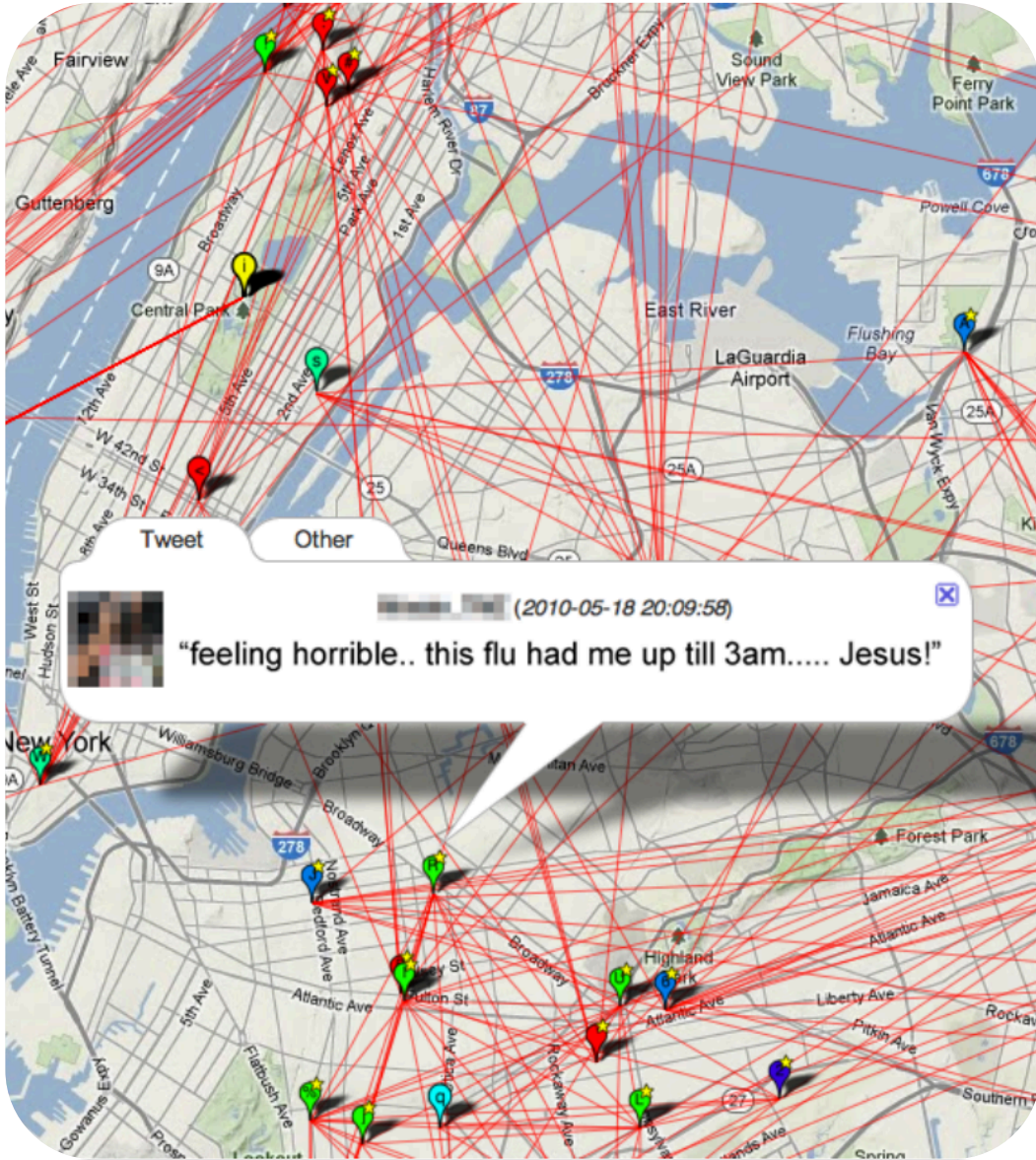




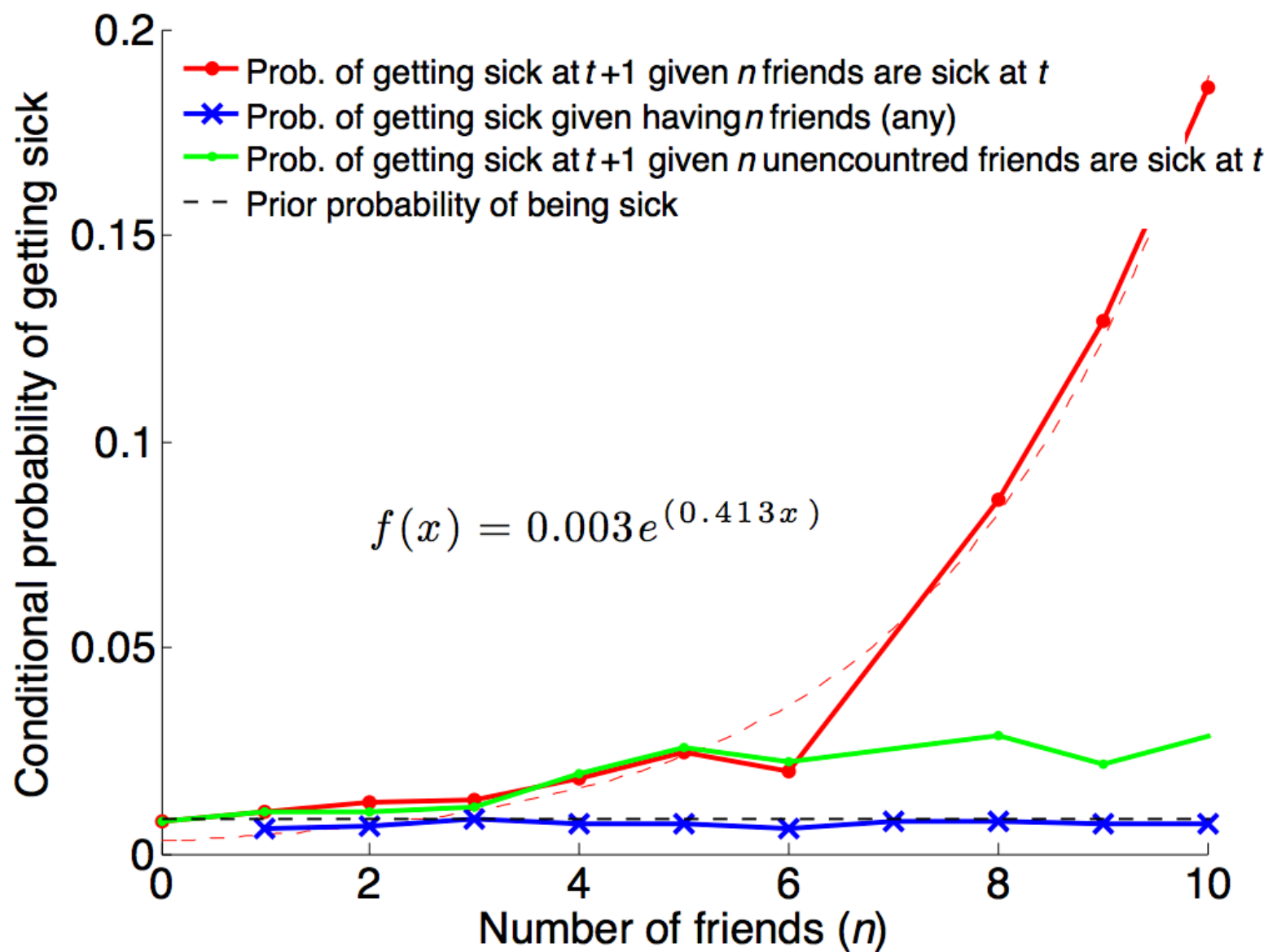


# Gripe

Un algoritmo de inteligencia artificial creado por científicos de la Universidad de Rochester y probado con éxito en la ciudad de Nueva York acierta en 9 de cada 10 pronósticos.







(b) Effects of friendship and co-location

# En 1 segundo...

¿Cuántos **Tweets** se publican?

¿Y fotos en **instagram**?

¿Y búsquedas en **Google**?



<https://www.internetlivestats.com>



Search Google or type URL



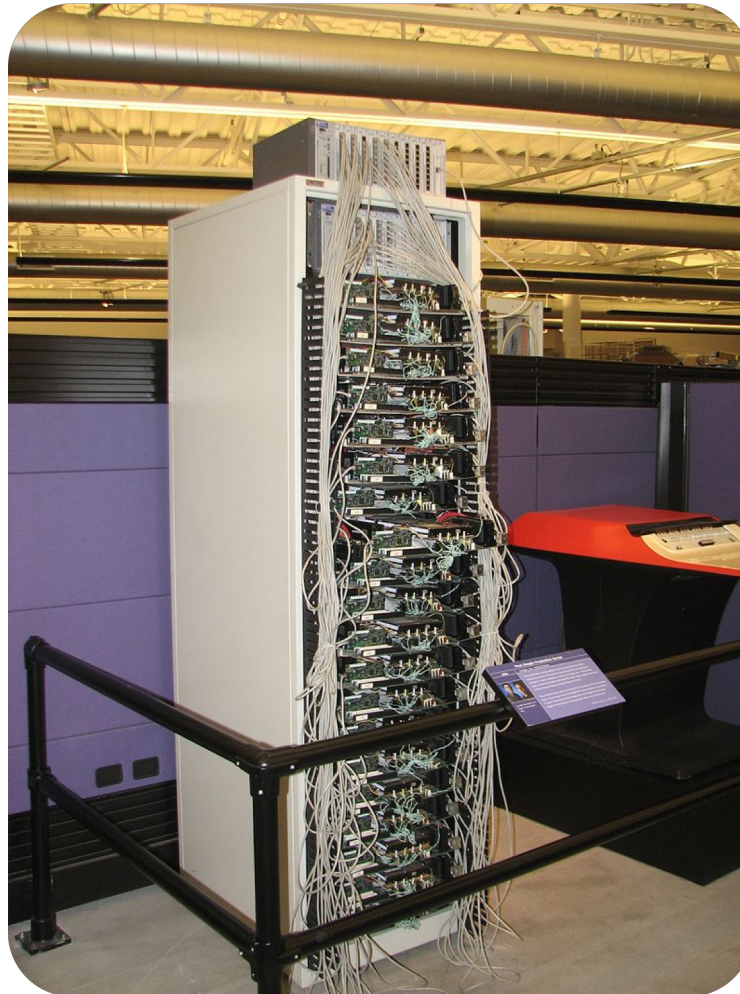
# Google: un motor de búsqueda



|                       | 1999       | 2009        | 2018         |
|-----------------------|------------|-------------|--------------|
| Documents             | 70 million | 1 trillion  | 130 trillion |
| Queries (per day)     | 100,000    | 8.5 million | 3.5 billion  |
| Average Latency (sec) | 1          | 0.5         | 0.5          |



# Primer Servidor de Google (1998)







Centro de datos de Google en Council Bluffs, Iowa.



# ¿Big data?

Grandes y complejas colecciones de datos, difíciles de almacenar y procesar utilizando las herramientas tradicionales de almacenamiento y procesamiento.

El reto de big data incluye capturar, limpiar, almacenar, buscar, compartir, transferir, analizar y visualizar estos datos.

# Explore what the world is searching

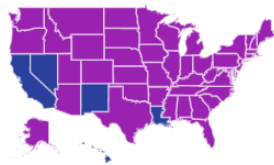
Enter a search term or a topic



Or start with an example

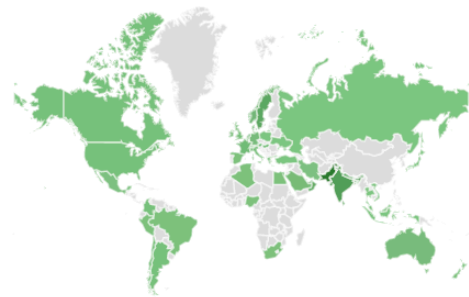
HIDE

● Taylor Swift ● Kim Kardashian



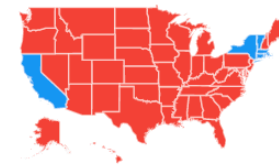
Interest by sub-region, Past 7 days, United States

● World Cup



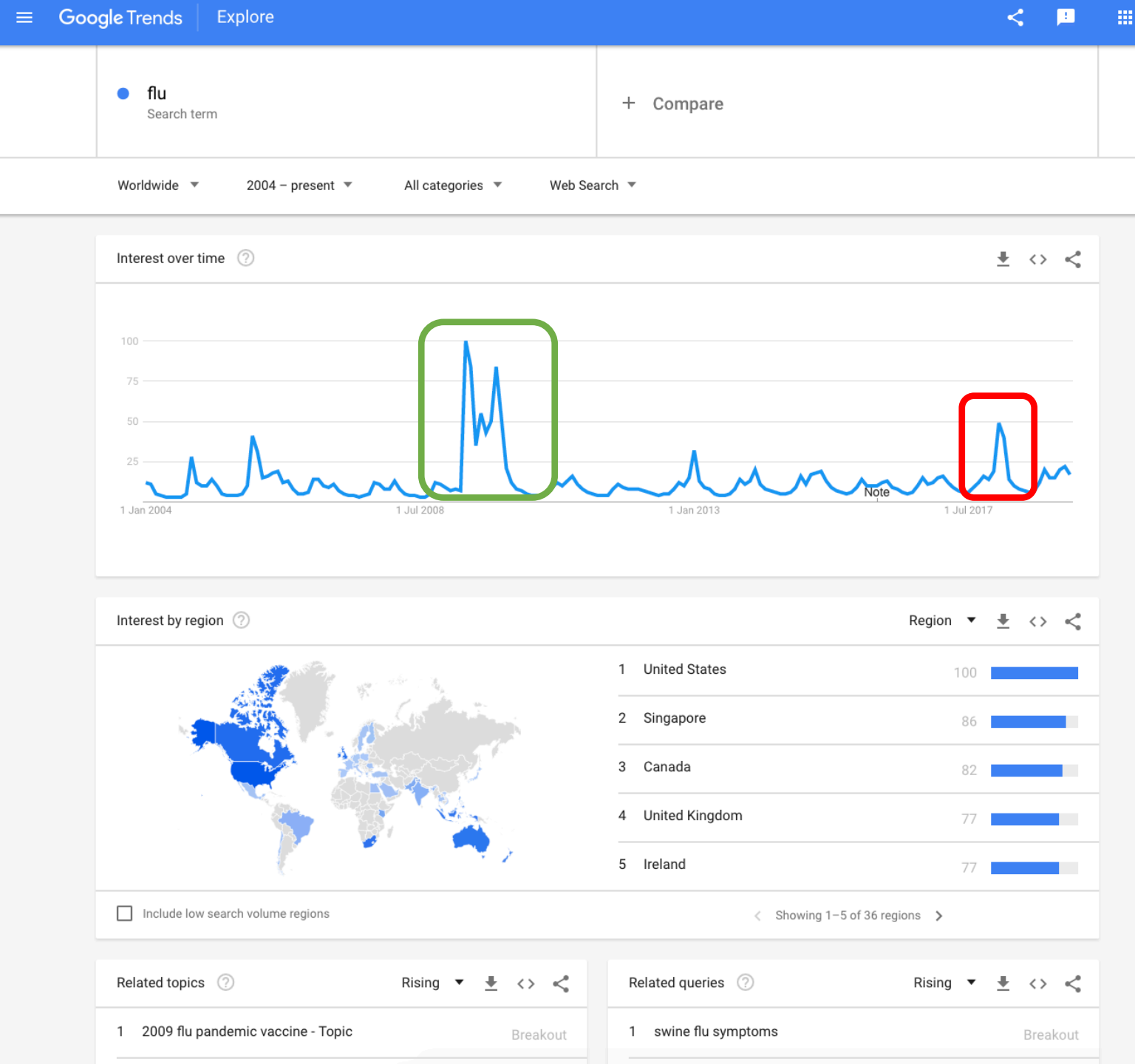
Interest by region, Past 7 days, Worldwide

● Football ● American football



Interest by sub-region, 2004 – present, United States





Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

## 2009 flu pandemic

From Wikipedia, the free encyclopedia

The **2009 flu pandemic** or *swine flu* was an **influenza pandemic** that lasted from early 2009 to late 2010, and the second of the two pandemics involving **H1N1 influenza virus** (the first of them being the **1918 flu pandemic**), albeit in a new version. First described in April 2009, the virus appeared to be a new strain of H1N1 which resulted when a previous triple **reassortment** of bird, swine and human flu viruses further combined with a Eurasian pig flu virus,<sup>[2]</sup> leading to the term "**swine flu**".<sup>[3]</sup>

Unlike most strains of influenza, H1N1 does not disproportionately infect adults older than 60 years; this was an unusual and characteristic feature of the H1N1 **pandemic**.<sup>[4]</sup> Even in the case of previously very healthy people, a small percentage will develop **pneumonia** or **acute respiratory distress syndrome (ARDS)**. This manifests itself as increased breathing

**Pandemic H1N1/09 Influenza**

Electron microscope image of the reassorted H1N1 influenza virus. The viruses are ~100 nanometres in diameter.<sup>[1]</sup>

**Specialty** Infectious disease, pulmonology

The New York Times

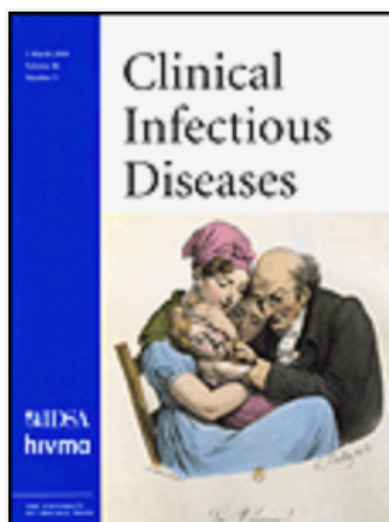
## This Flu Season Is the Worst in Nearly a Decade

An emergency room nurse treating a flu patient in Vista, Calif., this month. Mike Blake/Reuters

By Donald G. McNeil Jr.

Jan. 26, 2018

f t e p



#### JOURNAL ARTICLE

## Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks

Herman Anthony Carneiro and Eleftherios Mylonakis  
*Clinical Infectious Diseases*

Vol. 49, No. 10 (15 November 2009), pp. 1557-1564

Published by: [Oxford University Press](http://www.oxforduniversitypress.com)

Stable URL: <https://www.jstor.org/stable/27799390>

Page Count: 8

**Topics:** [Search engines](#), [Surveillance](#), [Infectious diseases](#), [Disease outbreaks](#), [West Nile virus](#), [Search terms](#), [Preventive medicine](#), [Influenza A virus](#), [Surfing](#)

# NETFLIX

## Top Picks for ana



139 millones de suscriptores a nivel mundial

# Sistemas Recomendadores



The Amazon logo is centered on a dark gray background. It features the word "amazon" in a white, lowercase, sans-serif font. A bright orange curved arrow, representing a smile, is positioned below the text, starting under the 'a' and ending under the 'n', with the arrowhead pointing towards the right. A small registered trademark symbol (®) is located to the right of the word.

Compare with similar items

**Renewed (13)** from \$115.95 Details

**Add an Accessory:**

- ☐ Microsoft Office 365 Home **\$99.00**
- ☐ G Suite Business | Unlimited Storage | Monthly Sub... **\$10.00**

7 - 8 inch Tablet Sleeve,  
Evecase 7 ~ 8 inch Tablet  
Ad feedback

Total price: **\$141.42**

[Add all three to Cart](#)

[Add all three to List](#)

- ✓ **This item:** Samsung Chromebook (Wi-Fi, 11.6-Inch) - Silver (Certified Refurbished) **\$116.94**
- ✓ AmazonBasics 11.6-Inch Laptop Sleeve **\$8.49**
- ✓ mCover iPearl mCover Hard Shell Case (mCover-Samsung-Chromebook-11.6-PINK) **\$15.99**

[Ad feedback](#)

*“If I have 3 million customers on the Web, I should have 3 million stores on the Web”*

Jeff Bezos, CEO of Amazon.com

\$135.6 billion estimated, ranked no. 1 in the Forbes list of world's billionaires







*Hipermercados*

Market basket analysis

---

Data mining



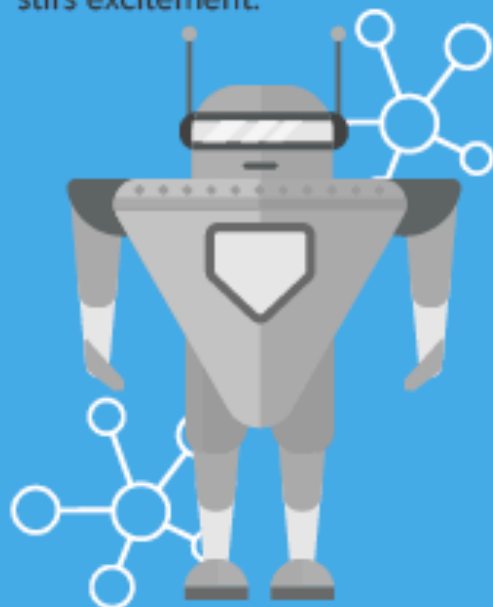
Inteligencia

Artificial

Aprendizaje Automático

# ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



# MACHINE LEARNING

Machine learning begins to flourish.

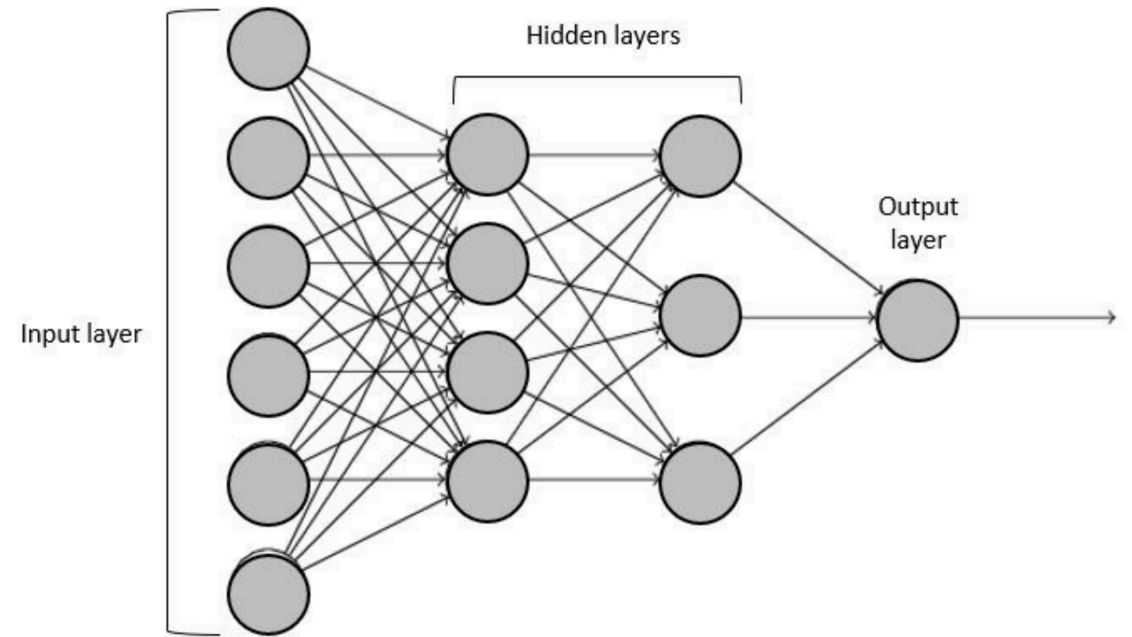
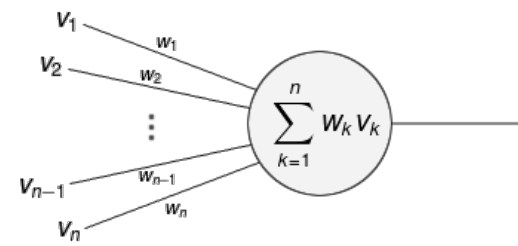
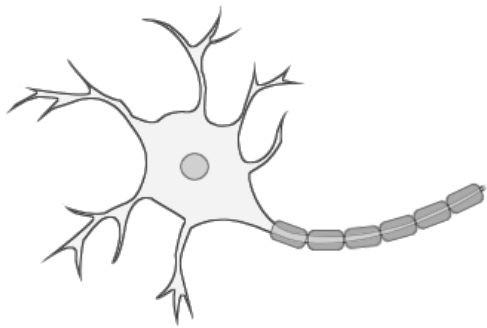


# DEEP LEARNING

Deep learning breakthroughs drive AI boom.



Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.





# ARTIFICIAL INTELLIGENCE

## Early artificial intelligence stirs excitement.



# MACHINE LEARNING

Machine learning begins to flourish.



# DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

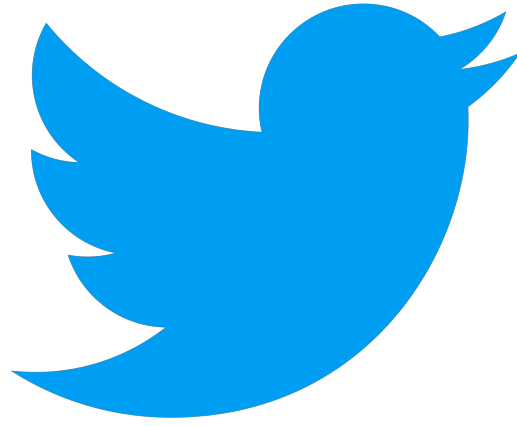
1980's

1990's

2000's

2010's

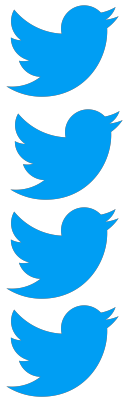
Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.



# Detección de Enfermedades en Redes Sociales



## Recopilación de datos y etiquetado



“Me estoy medicando contra la depresión”

“Perdió el Barsa, qué depresión”

“La depresión afecta a un alto % de la población”

“La depresión afecta a un alto % de la población”

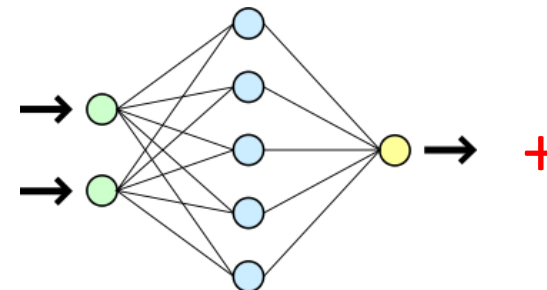
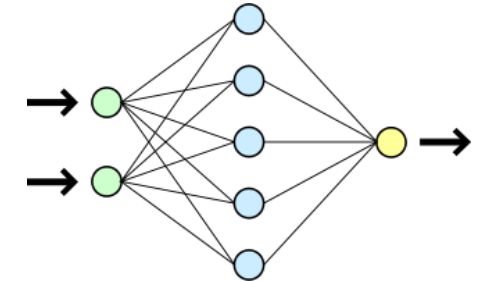


|      |   |     |      |      |   |
|------|---|-----|------|------|---|
| 0,43 | 1 | ... | 0,98 | 0,36 | ? |
|------|---|-----|------|------|---|

## Extracción de características

|   |      |   |     |      |      |   |
|---|------|---|-----|------|------|---|
| + | 0,89 | 1 | ... | 0,50 | 0,23 | 1 |
| - | 0,45 | 0 | ... | 0,54 | 0,53 | 0 |
| - | 0,99 | 0 | ... | 0,89 | 0,45 | 0 |
| - | 0,43 | 1 | ... | 0,98 | 0,36 | 0 |

## Entrenamiento



Ejemplos









# Sesgo en Big Data





# Sistema de etiquetado automático de imágenes de Google, 2015



Google

All **Images** Videos News Maps More Settings Tools

clip art uniform thank you day registered h >

Foreign-Educated Nurse  
nurse.org

Nursing Jobs  
nursingworld.org

documenting properly on patient charts  
nurse.com

Registered Nurse  
nursinglicensemap.com

Nurse - Nursing - Youth.gov.hk Career ...  
youth.gov.hk

Nurse interview questions and answers ...  
snagajob.com

Nurses | TaleMed  
talemed.com

How Nurses Stay Healthy through ...  
stevenson.edu

Nurse Staffing  
nursingworld.org

Google

All **Images** Maps Videos News More Settings Tools

female cartoon clip art patient stethoscope >

Choosing the Best Primary Care Doctor ...  
rd.com

How to Spot a Bad Doctor | MD Ma...  
mdmag.com

Well Connection | MyBlue  
myblue.bluecrossma.com

How to find an English-speaking doctor ...  
expatincroatia.com

Lehigh County Physician Listing | LCM...  
lcmadsoc.org

Tips for Choosing a Doctor - Scripps Health  
scripps.org

Bulk Billing Doctors In Chapel...

Kids' Checkups: What to Expect, Why...

Find a Doctor, Doctor Reviews & Rating...



in

arquitecto

Home

My Network

Jobs

Messaging

People

Connections

Locations

Current companies

All Filters

Showing 1,366,681 results

**Ruben Fernandez Retuerto** • 1st  
Arquitecto de Sistemas Unix, Windows & Middleware - Responsable técnico equi...  
A Coruña Area, Spain  
Current: Arquitecto de sistemas - Dept. Middleware - Responsable técnico equipo Tecnomat Indra  
Jorge Prieto Trabada, Emma Iglesias Alejandro, and 65 other shared connections

Message

**Cristian Felipez Ongay** • 2nd  
Arquitecto  
Barcelona Area, Spain  
Adrián Pérez Diéguez, CONSTRUIRE S.A.S., and 2 other shared connections

Connect

**Iago Nóvoa Pérez** • 1st  
Graphic Designer, Architect, and Founding Partner at 988  
Ourense Area, Spain  
Past: Arquitecto at South Eastern Education and Library Board  
Blanca Antolín, Raquel Sabin Pedre, and 20 other shared connections

Message

**Javier Novo Rodriguez** • 1st  
Software Architect, Ph.D. in Computer Science, PMP®  
A Coruña Area, Spain  
Past: Arquitecto de Software at Grupo Cándido Hermida  
Iria García López, Raquel Sabin Pedre, and 126 other shared connections

Message

**Miguel Borrell Noguera** • 2nd  
Arquitecto + Arquitecto técnico  
Barcelona Area, Spain  
Juan Luis Hermo, Diego Barreto de Paz, and 21 other shared connections

Connect

**Roberto Zaffari** • 2nd  
Architect  
Barcelona Area, Spain  
Past: Arquitecto y BIM Manager at b720 Arquitectos  
Samuel Fernández Alonso, Diego Guillemette Rodríguez, and 2 other shared connections

Connect

**Lucía Sineiro Cascallar** • 2nd  
Arquitecto  
Barcelona Area, Spain  
Current: Arquitecto superior at INEL, ARQUITECTURA, INGENIERÍA Y MANAGEMENT  
Diego Rubio Pita, GALICIA VERDE FOUNDATION, and 7 other shared connections

Connect

in

arquitectura

Home

My Network

Jobs

Messaging

People

Connections

Locations

Current companies

All Filters

JIRA Simplified - Using JIRA Has Never Been Easier With WalkMe's Intuitive OnScreen

Showing 1,237,767 results

**Paula Daniela Garcia** • 2nd  
Arquitecta  
Barcelona Area, Spain  
GALICIA VERDE FOUNDATION, Francisco José Cruz Oria, and 4 other shared connections

Connect

**Laura Coladas Bilbao** • 2nd  
Arquitecta  
Barcelona Area, Spain  
Summary: Arquitecta por la Universidad de A Coruña en 2008. Desde entonces he colaborado en distintos...  
Iago Nóvoa Pérez is a shared connection

Connect

**Patricia Rodríguez Luque** • 2nd  
Arquitecta tecnica  
Barcelona Area, Spain  
Miguel Carreras Dutra, Juan Luis Hermo, and 4 other shared connections

Connect

**Anna Perramon Dominguez** • 2nd  
Arquitecta  
Barcelona Area, Spain  
Jaume Hernández Condal and Amador Gómez Hidalgo are shared connections

Connect

**Marta gonzalez guzman** • 2nd  
Arquitecta  
Barcelona Area, Spain  
Past: arquitecta at Pere Casajoana Arquitectes S.L.  
Marti Umbert is a shared connection

Connect

**olga felip ordis** • 2nd  
Arquitecta  
Barcelona Area, Spain  
Luz Rello and Amador Gómez Hidalgo are shared connections

Connect

**Eugenia Santiago** • 2nd  
Arquitecta - Facility & Project Manager  
Barcelona Area, Spain  
Past: ARQUITECTA - OFICINA TECNICA DE OBRA - EDIFICIO DE VIVIENDAS at ESTUDIO PISANO  
Carlos Castillo (ChaTo), Alexander Steblin, and 10 other shared connections

Connect

in

architect

Home

My Network

Jobs

Messaging

People

Connections

Locations

Current companies

All Filters

Showing 3,386,753 results

**Sergi Mayordomo** • 2nd  
Enterprise Solutions Architect  
Barcelona Area, Spain  
ETIC UPF, Ricardo Baeza-Yates, and 3 other shared connections

Connect

**Paula Martínez Louro** • 2nd  
Solutions Architect  
Barcelona Area, Spain  
Past: Solutions Architect at Claranet Spain  
Pablo Casaliau, Raúl Varela Barros, and 18 other shared connections

Connect

**Jose M Chenlo, Ph.D.** • 1st  
Lead Data Scientist & Solution Architect  
Vigo Area, Spain  
Current: Lead Data Scientist & Solution Architect at DXC Technology  
Daniel Valcarce, Jorge Novo Buján, and 123 other shared connections

Message

**Sergio Pérez Rodríguez** • 2nd  
DevOps Architect  
Barcelona Area, Spain  
Jean-Paul Kirschstein, Ignacio Candau Pérez, and 2 other shared connections

Connect

**Miguel Toscano** • 2nd  
Solutions Architect  
Barcelona Area, Spain  
Current: Principal Consultant at Oracle  
Ignacio Candau Pérez, Albert Giménez, and 2 other shared connections

Connect

**René Balay Lorenzo** • 2nd  
Software Architect at Inditex  
Spain  
Nazaret Hermida Fernández, Pablo Casaliau, and 34 other shared connections

Connect

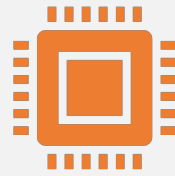
**Daniel Gómez Ferro** • 2nd  
Chief Architect at Splice Machine  
Barcelona Area, Spain  
Past: Software Architect at Splice Machine  
Jose M Chenlo, Ph.D., Pablo Suárez Gómez, and 36 other shared connections

Connect

**Tomás Casquero Sierra** • 2nd  
Javascript Architect  
Barcelona Area, Spain

Connect

Empresas con  
equipos más  
diversos  
tienen un  
19% más de  
ingresos



Intentar contrarrestar el sesgo  
de los algoritmos



Dotar a las empresas de mayor  
diversidad

# #Cambiemoslosdatos de la presencia femenina en el sector TIC

25%

mujeres  
estudiando  
carreras  
técnicas

15%

programadoras  
en Twitter,  
Google,  
Facebook y  
Apple

17%

mujeres en  
startups



1078

estudiantes  
internacionales

400

referentes  
femeninos en TIC

22

centros  
premiados



# Wisibilíзалas

4<sup>a</sup> edición

<https://www.upf.edu/web/wisibilizalas>  
@wisibilizalas





Aprendiendo a ser científic@s de datos

Ana Freire

Científica de Datos



Universitat  
Pompeu Fabra  
*Barcelona*