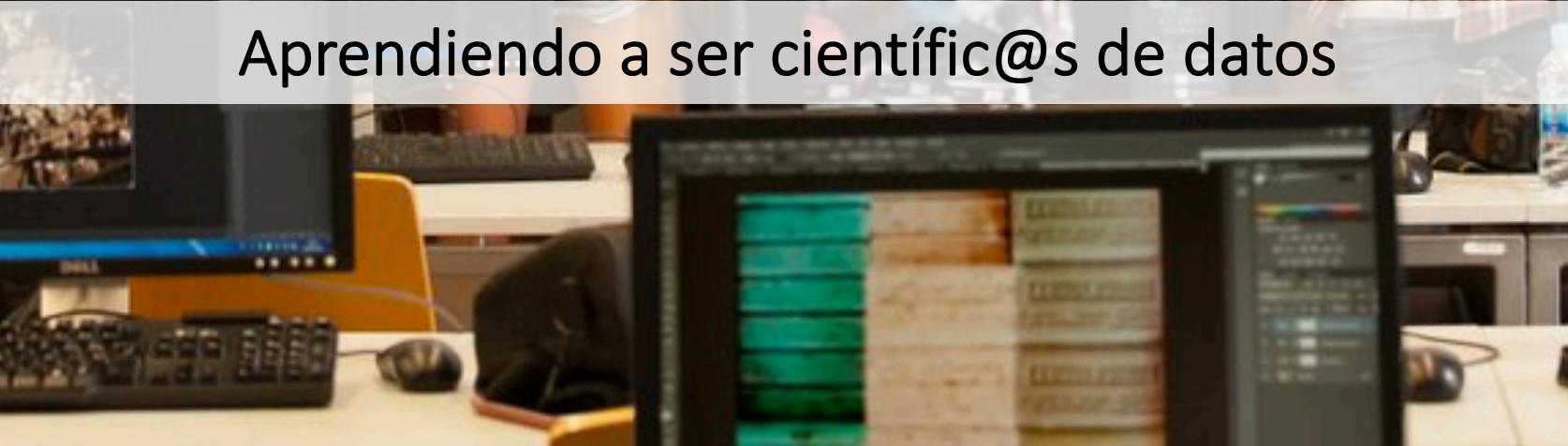




Aprendiendo a ser científic@s de datos



Ana Freire

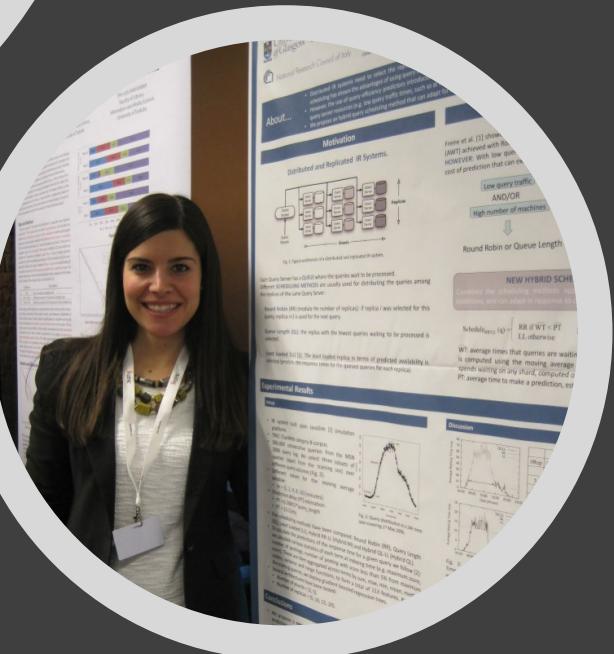
Científica de Datos



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



¿Qué datos se generan  
en la sociedad?

Finanzas



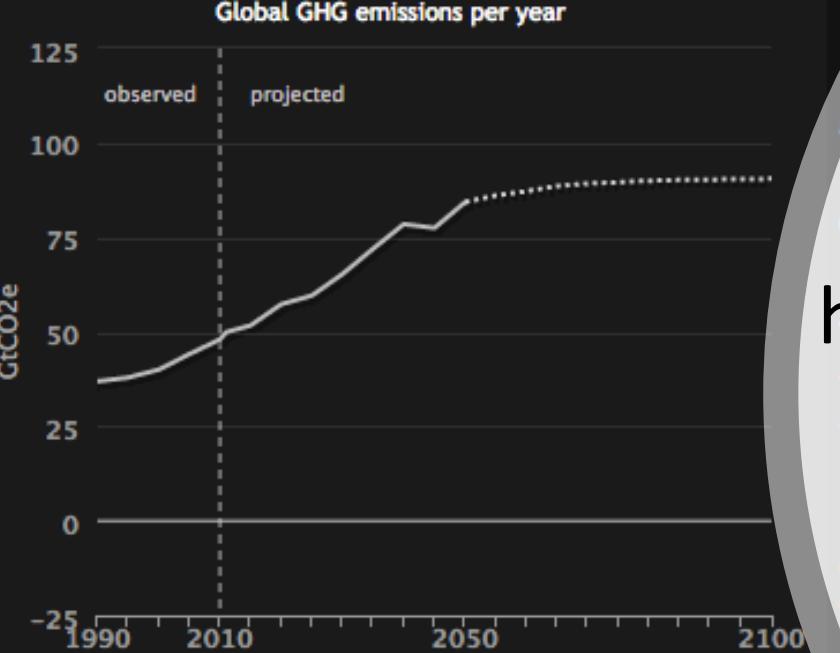
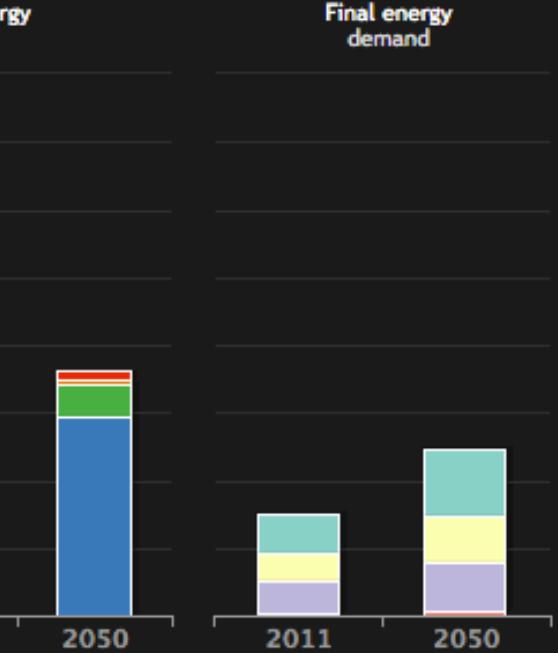
¿Qué datos se generan en la sociedad?

Comportamiento humano



# Cambio Climático

¿Qué podemos hacer para ralentizarlo?



# Cambio Climático

<http://tool.globalcalculator.org>

Keep emissions below the line to maintain a 50% chance of meeting the UNFCCC 2°C target!

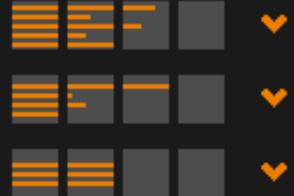
How is the temperature change estimated?

Choose a pathway or create your own pathway by selecting effort level 1 to 4 for each lever.

--- Example pathways ---

Selected pathway: IEA 6DS (approx.)

#### LIFESTYLE



#### TECHNOLOGY AND FUELS



#### BIOENERGY

#### FOSSIL FUEL

#### NUCLEAR

#### RENEWABLES

#### LAND AND FOOD



#### DEMOGRAPHICS AND LONGTERM

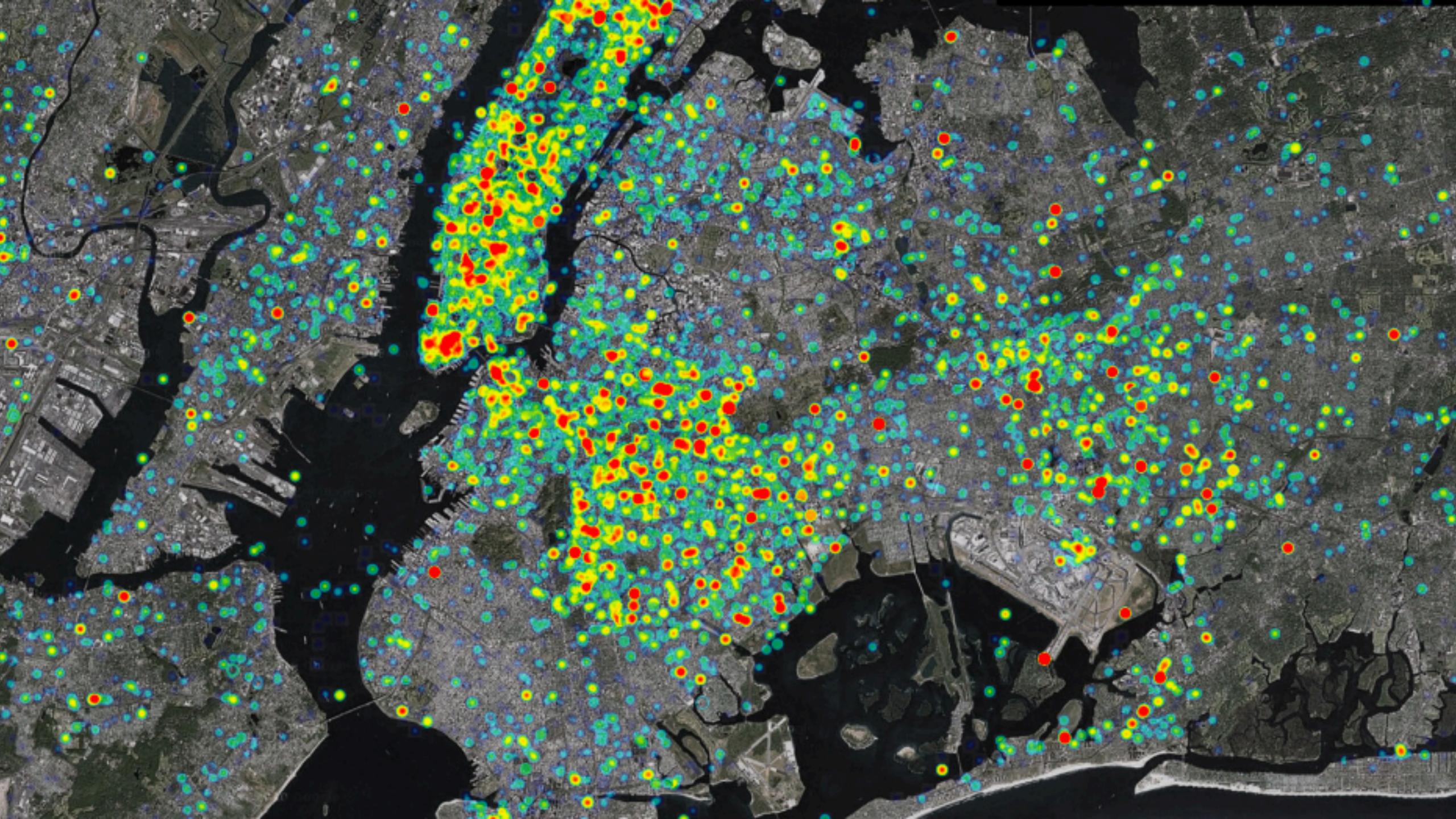
#### DEMOGRAPHICS

#### EMISSIONS AFTER 2050



¿Qué datos se generan en la sociedad?

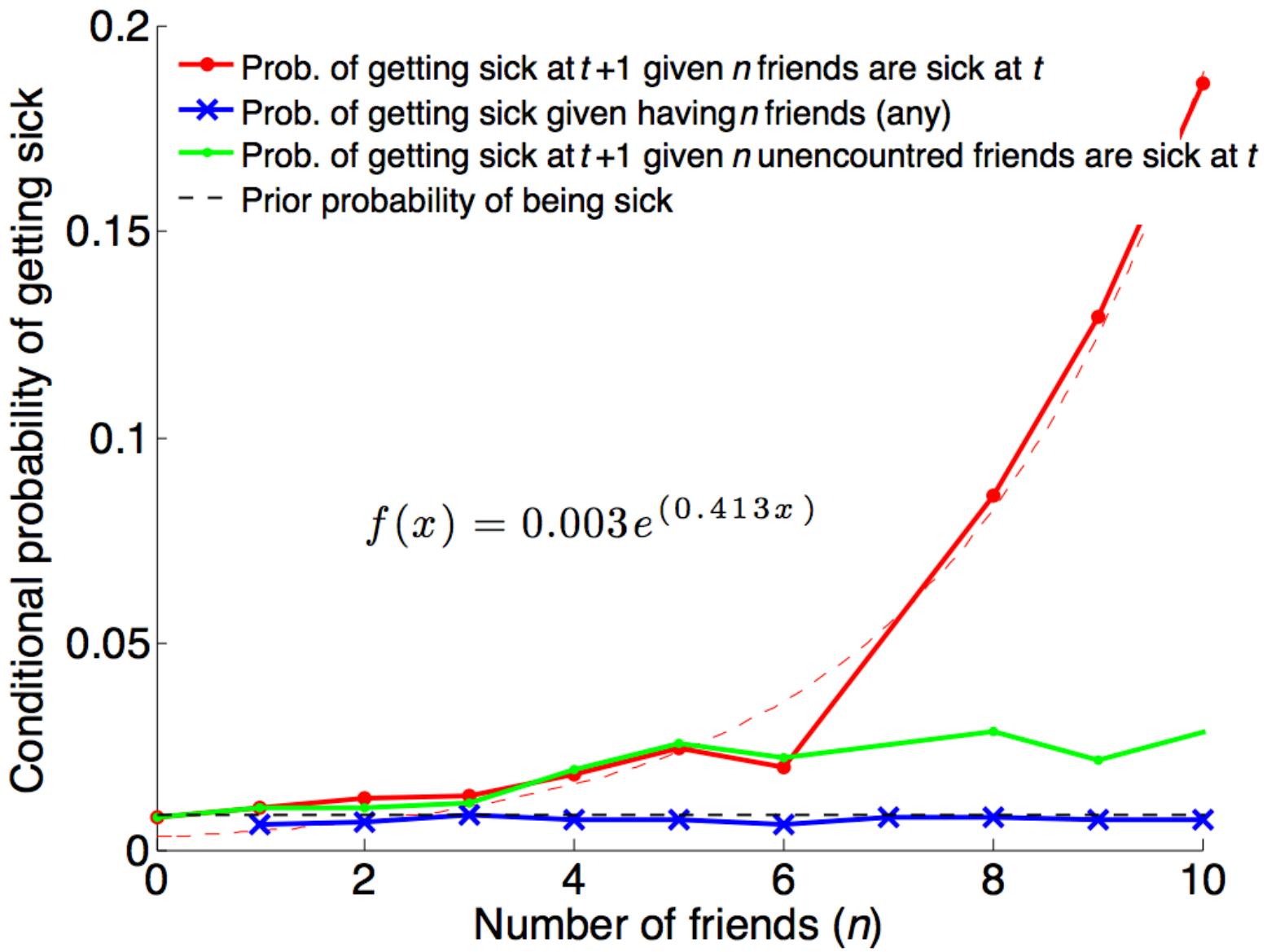
Comportamiento humano



# 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>



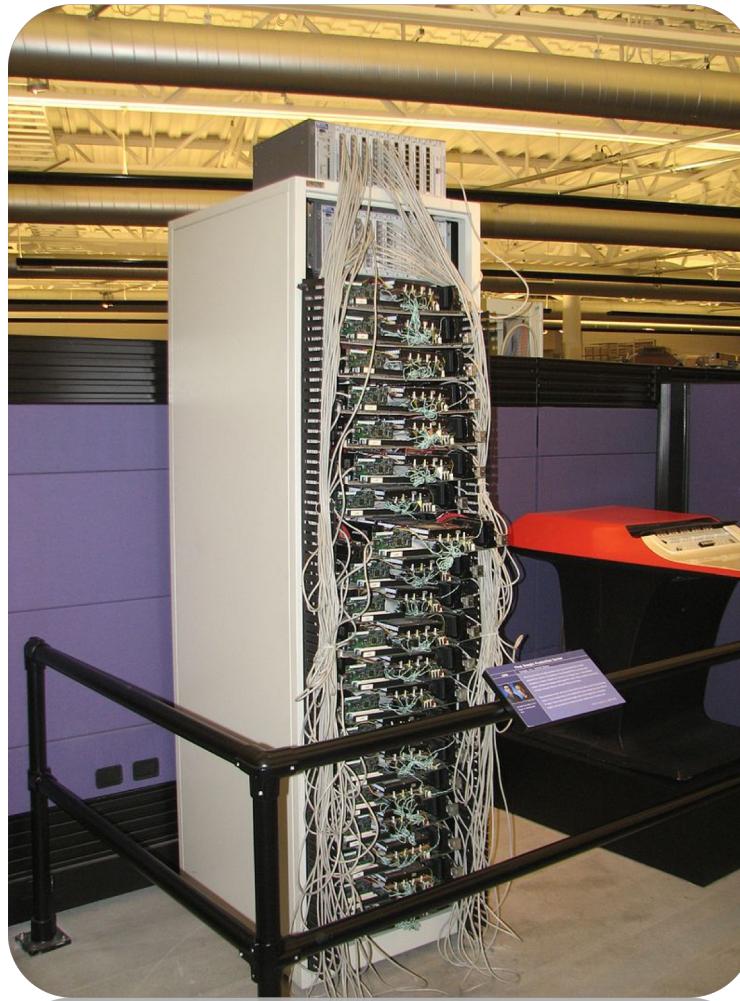
Search Google or type URL



# 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



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



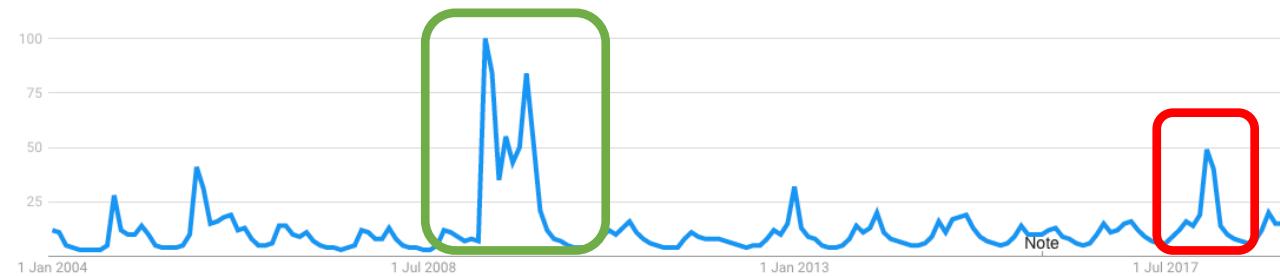
Interest by sub-region, 2004 – present, United States

flu  
Search term

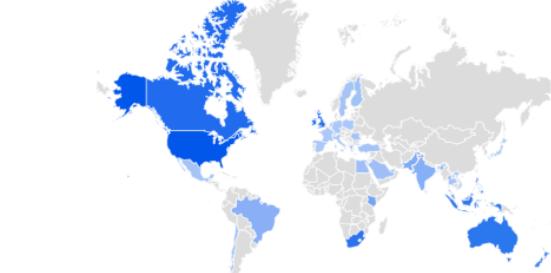
+ Compare

Worldwide ▾ 2004 – present ▾ All categories ▾ Web Search ▾

Interest over time



Interest by region



Region

1	United States	100	
2	Singapore	86	
3	Canada	82	
4	United Kingdom	77	
5	Ireland	77	

< Showing 1–5 of 36 regions >

Related topics

Rising

1 2009 flu pandemic vaccine - Topic

Breakout

Related queries

Rising

1 swine flu symptoms

Breakout



WIKIPEDIA  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
[Donate to Wikipedia](#)  
[Wikipedia store](#)  
  
Interaction  
Help  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact page](#)  
  
Tools  
[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Permanent link](#)

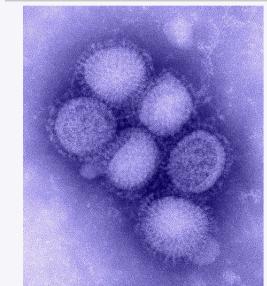
## 2009 flu pandemic

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

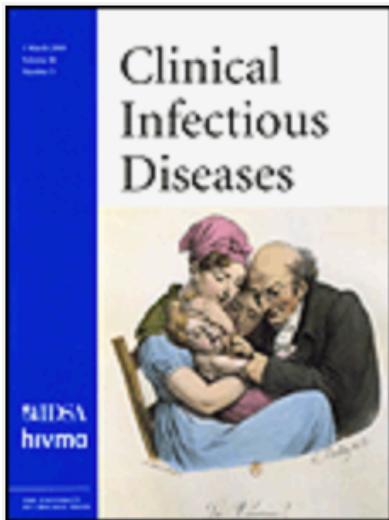
## This Flu Season Is the Worst in Nearly a Decade



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

By Donald G. McNeil Jr.

Jan. 26, 2018



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](#)

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



Shop Today's Deals

EN Hello, Sign In Account & Lists Orders Cart

Electronics - Departments - Your Amazon.com Today's Deals Gift Cards Registry Sell Help

Electronics > Computers & Accessories > Computers & Tablets > Laptops > Traditional Laptops

Samsung Chromebook (Wi-Fi, 11.6-Inch) - Silver (Certified Refurbished) by Samsung ★★★☆☆ 572 customer reviews | 288 answered questions

Price: \$116.94 Color: Silver

\$116.94 \$129.00

In Stock. Qty: 1 Add to Cart Buy Now

This item ships to Spain. Want it Thursday, March 28? Choose AmazonGlobal Priority Shipping at checkout. Learn more

Product works and looks like new. Comes with a 90-day warranty. This Amazon Renewed product is professionally inspected and tested by an Amazon qualified supplier. Box and accessories may be generic. Learn more

- Chrome OS, Samsung Exynos 5250 Dual Core Processor
- Display: 11.6" LED HD 1366 x 768 16:9; Ports: HDMI, Headphone/MIC combo, 1 x USB 3.0 + 1 x USB 2.0, 3- in-1 (SD/SDHC/SDXC)
- Memory: 2 GB DDR3L RAM, 16GB Solid State Drive
- Built-in dual band Wi-Fi 802.11 a/b/g/n, Webcam, 3W Stereo Speaker

See more product details

Compare with similar items

Renewed (13) from \$115.95 Details

Roll over image to zoom in

Frequently bought together

Total price: \$141.42 Add all three to Cart Add all three to List

These items are shipped from and sold by different sellers. Show details

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

Sponsored products related to this item

Page 1 of 69 Renewed (13) from \$115.95 Details

Acer Chromebook 11, Celeron N3060, 11.6" HD, 4GB DDR3L, 16GB Storage, CB3-132-C4VV \$199.00 prime 309 ★★★☆☆ 309

ASUS Chromebook C202SA-VS02 11.6" Ruggedized and Water Resistant Design with... \$199.00 prime 1507 ★★★☆☆ 1507

Samsung Chromebook Pro Convertible Touch Screen Laptop, 12.3 (XE510C24-K01US) \$509.00 prime 399 ★★★☆☆ 399

2018 HP 17.3" HD+ Notebook Laptop, Intel Core i5-7200U up to 3.10GHz,... \$611.00 prime 151 ★★★☆☆ 151

ASUS Chromebook Flip C302CA-DH54 12.5-inch Touchscreen Convertible Chromebook Intel... \$633.98 prime 30 ★★★☆☆ 30

Apple MacBook Pro 13.3" 2.5GHz i5, 16GB RAM, 500GB SSD, Space Gray, 2018 Model \$569.99 prime 569.99 prime

Apple MacBook Pro 13.3" 2.5GHz i5, 16GB RAM, 500GB SSD, Space Gray, 2018 Model \$569.99 prime 569.99 prime

Ad feedback

Customers also shopped for

Page 1 of 2

Apple iPad with Retina Display, MD511LL/A (32GB, Wi-Fi, Black) 4th Generation (Refurbished) \$202.95 498 ★★★☆☆ 498

Samsung Galaxy Tab E Lite 7.0" 8GB Wi-Fi (White) (Certified Refurbished) \$94.95 33 ★★★☆☆ 33

Samsung Galaxy Tab A 10.1" 16GB (Wi-Fi), White (Certified Refurbished) \$189.99 12 ★★★☆☆ 12

Samsung XE303C12-A01US Samsung Exynos 5250 X2 1.7GHz 2GB 16GB SSD... \$139.95 2 offers from 139.95 \$349.99

iView 1760AIO All in One Computer/Tablet, 17.3" IPS 1920 x 1080 Touch Screen, Intel Apollo Lake... ★★★☆☆ 28

Dell Chromebook 11 11.6-inch ★★★☆☆ 64

Ad feedback

Other Sellers on Amazon

\$116.95 Add to Cart & FREE Shipping on eligible orders. Details Sold by: Kanga Supply

\$117.98 Add to Cart & FREE Shipping on eligible orders. Details Sold by: SupremeDealsCR

Share Email Facebook Twitter LinkedIn 2K+ Deliver to Spain Add to List

Have one to sell? Sell on Amazon

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

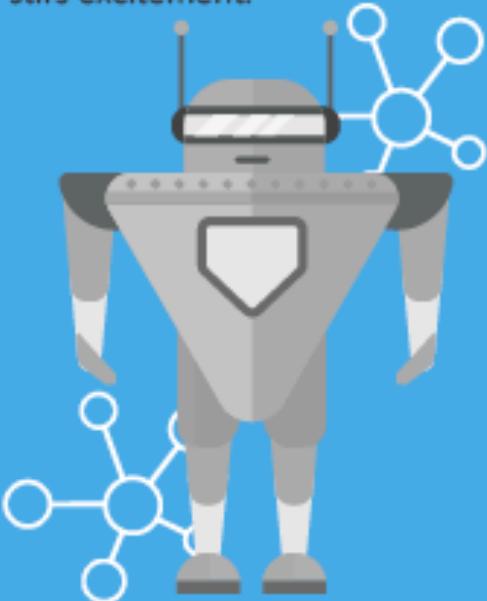
# Inteligencia

# Artificial

## Aprendizaje Automático

## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



1950's

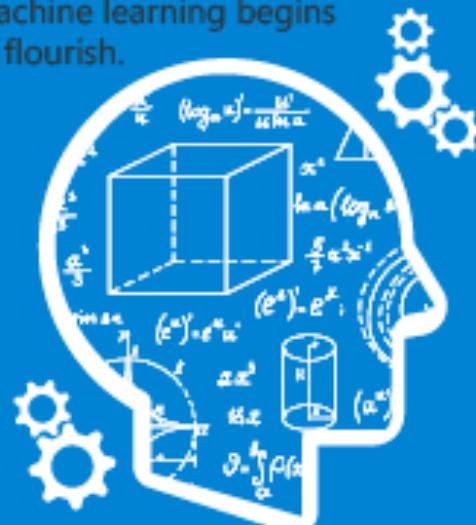
1960's

1970's

1980's

## MACHINE LEARNING

Machine learning begins to flourish.



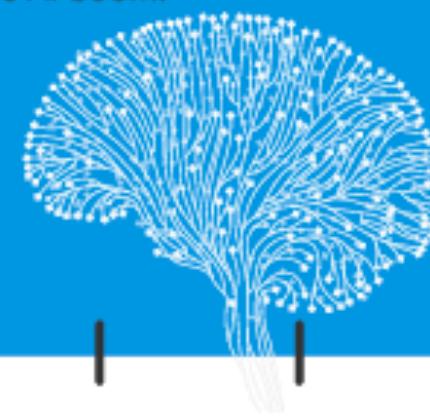
1990's

2000's

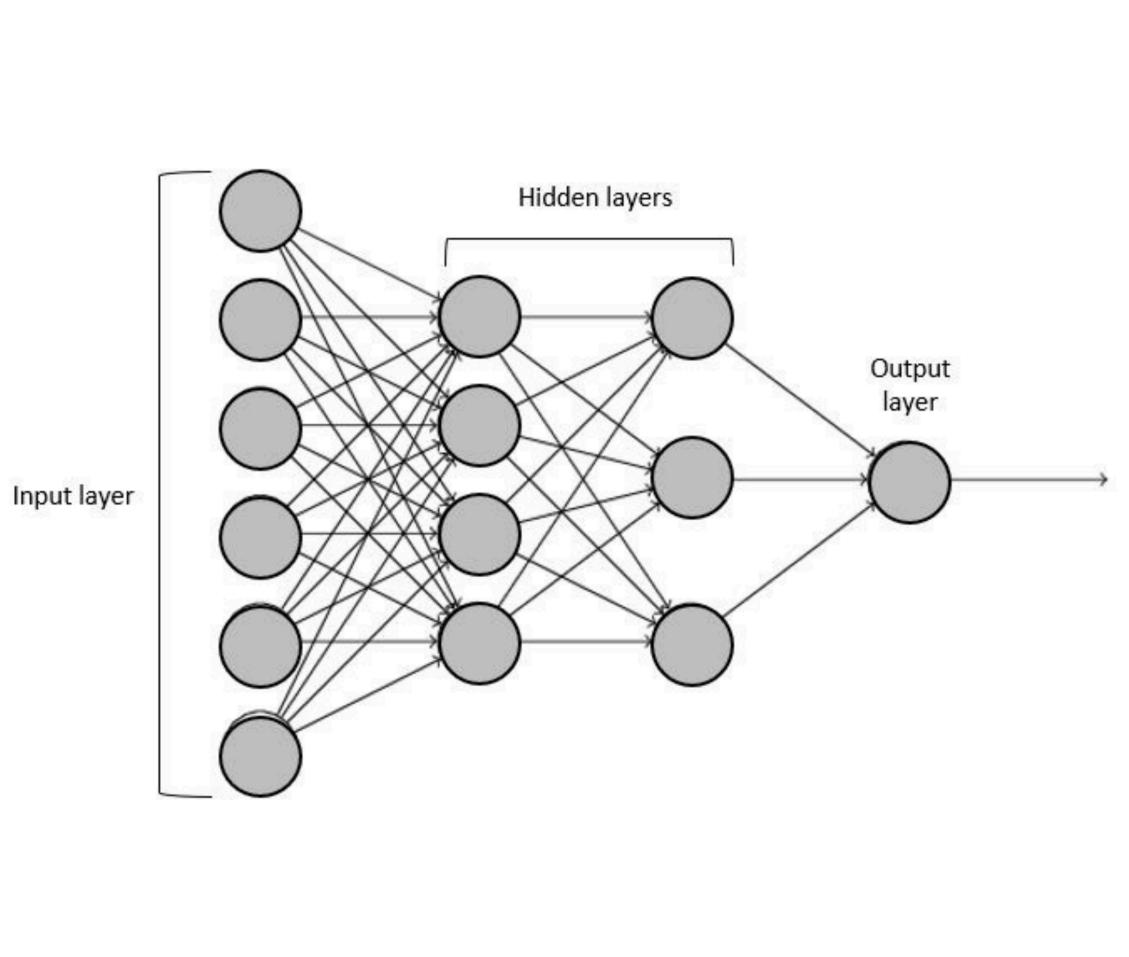
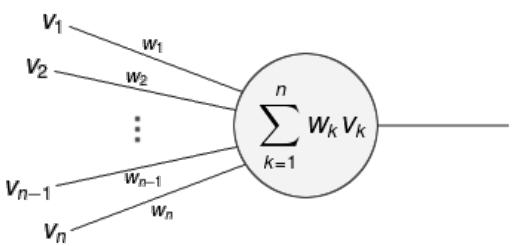
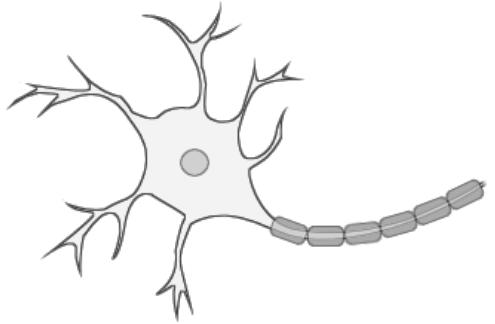
2010's

## DEEP LEARNING

Deep learning breakthroughs drive AI boom.

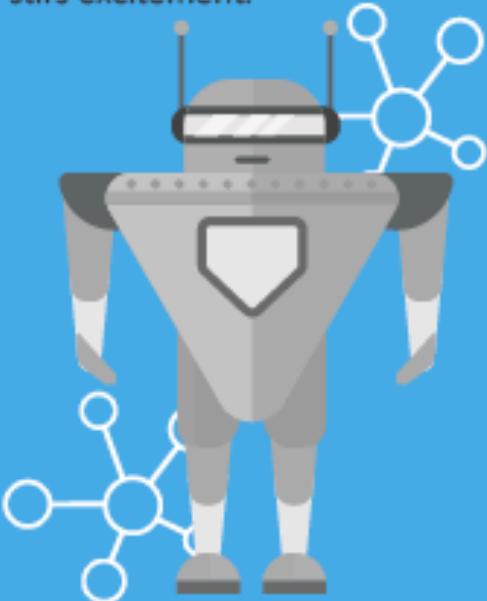


Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.



## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



1950's

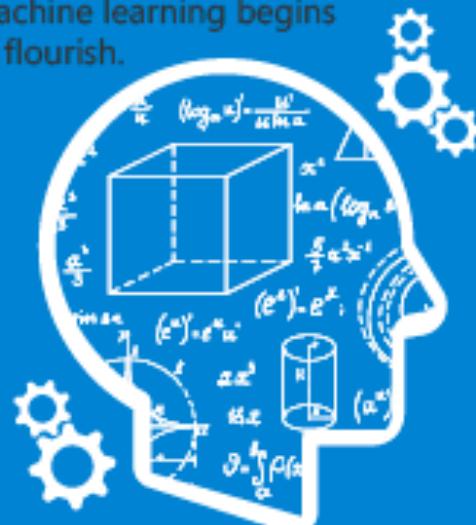
1960's

1970's

1980's

## MACHINE LEARNING

Machine learning begins to flourish.



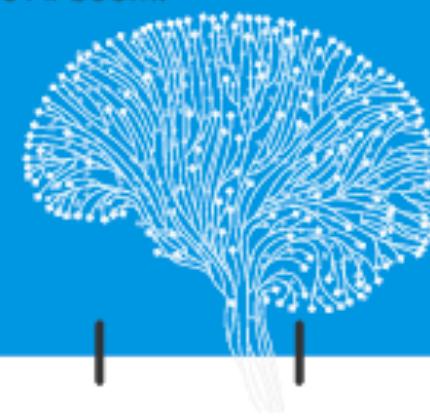
1990's

2000's

2010's

## 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.

# 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”

## Extracción de características

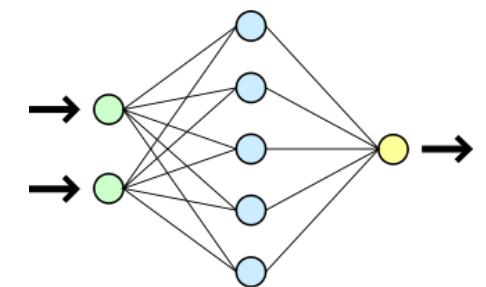
+	0,89	1	...	0,50	0,23	1
---	------	---	-----	------	------	---

-	0,45	0	...	0,54	0,53	0
---	------	---	-----	------	------	---

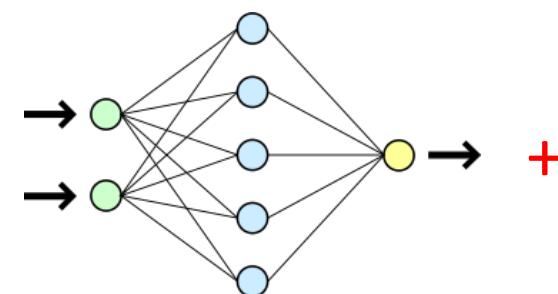
-	0,99	0	...	0,89	0,45	0
---	------	---	-----	------	------	---

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

## Entrenamiento



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



# Ejemplos





# Sesgo en Big Data

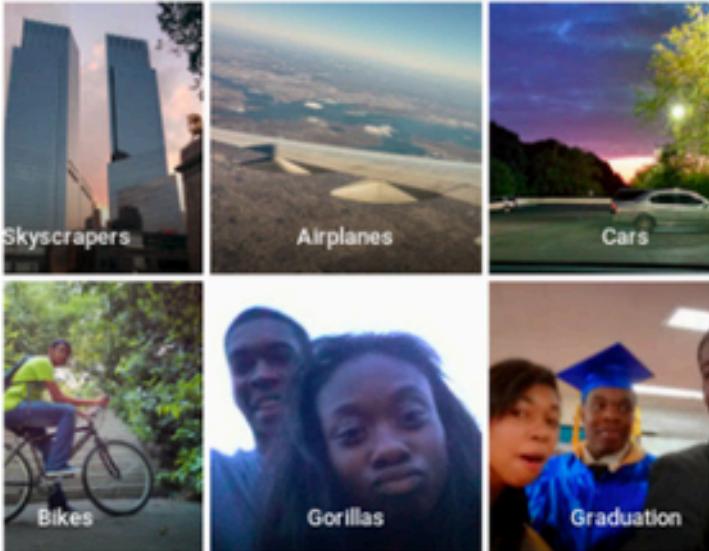


# Sistema de etiquetado automático de imágenes de Google, 2015

 **Jacky Alciné**  
@jackyalcine

Follow

Google Photos, y'all [REDACTED] up. My friend's not a gorilla.



 Skyscrapers	 Airplanes	 Cars
 Bikes	 Gorillas	 Graduation

RETWEETS 3,356 FAVORITES 1,930

8:22 PM - 28 Jun 2015

Reply Retweet Like Share Save More

# Google Images - nurse

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 Images - doctor

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...  
lcmedsoc.org

Tips for Choosing a Doctor - Scripps Health  
scripps.org

Bulk Billing Doctors In Cheltenham  
kidcheckups.whattoexpect.when...

Find a Doctor, Doctor Reviews & Ratings  
findadoctor.com

Showing 1,366,681 results

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

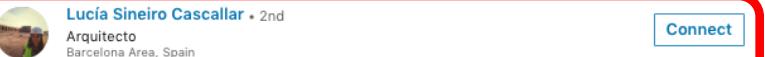
**Cristian Felipez Ongay** • 2nd  
Arquitecto  
Barcelona Area, Spain  
  Adrián Pérez Diéguez, CONSTRUIRE S.A.S., and 2 other shared connections

**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 Sabín Pedre, and 20 other shared connections

**Javier Novo Rodríguez** • 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 Sabín Pedre, and 126 other shared connections

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

**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

**Lucía Sineiro Cascalier** • 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

Showing 1,237,767 results

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

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

**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

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

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

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

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

**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

Showing 3,386,753 results

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

**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

**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

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

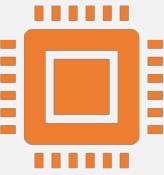
**Miguel Toscano** • 2nd  
Solutions Architect  
Barcelona Area, Spain  
Current: Principal Consultant at Oracle  
  Ignacio Candau Pérez, Albert Giménez, and 2 other shared connections

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

**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

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

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



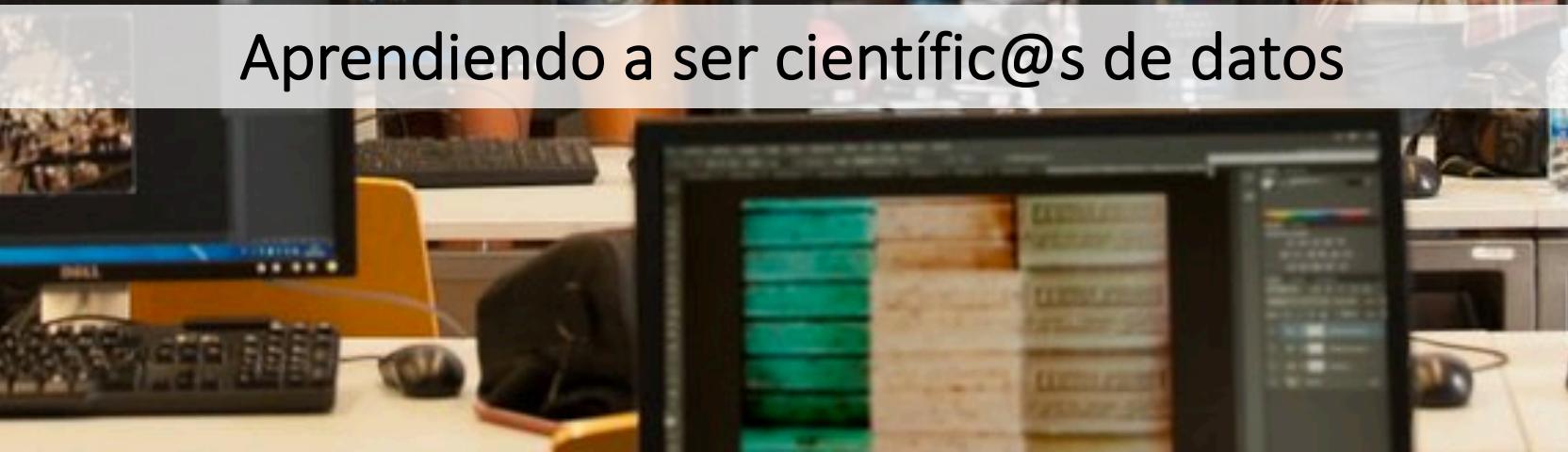
# Wisibilízalas

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