



Big Data y Ciencias Sociales

Desafíos, oportunidades y primeros
pasos

De qué vamos a hablar

- Qué es la ciencia de datos y el Big Data
- Especialidades, usos y costumbres
- El rol de las ciencias sociales
- El trabajo del científico de datos: luz y sombra
- Lenguajes, comunidades, formación y puestos de trabajo en ciencia de datos

¿Qué es la ciencia de datos?

La disciplina que aplica técnicas de programación para analizar datos, también conocido como “El trabajo más sexy del siglo XXI” según Thomas Davenport and DJ Patil



COMUNICACIÓN Y MONITOREO

¿Ciencia de datos == Big Data?

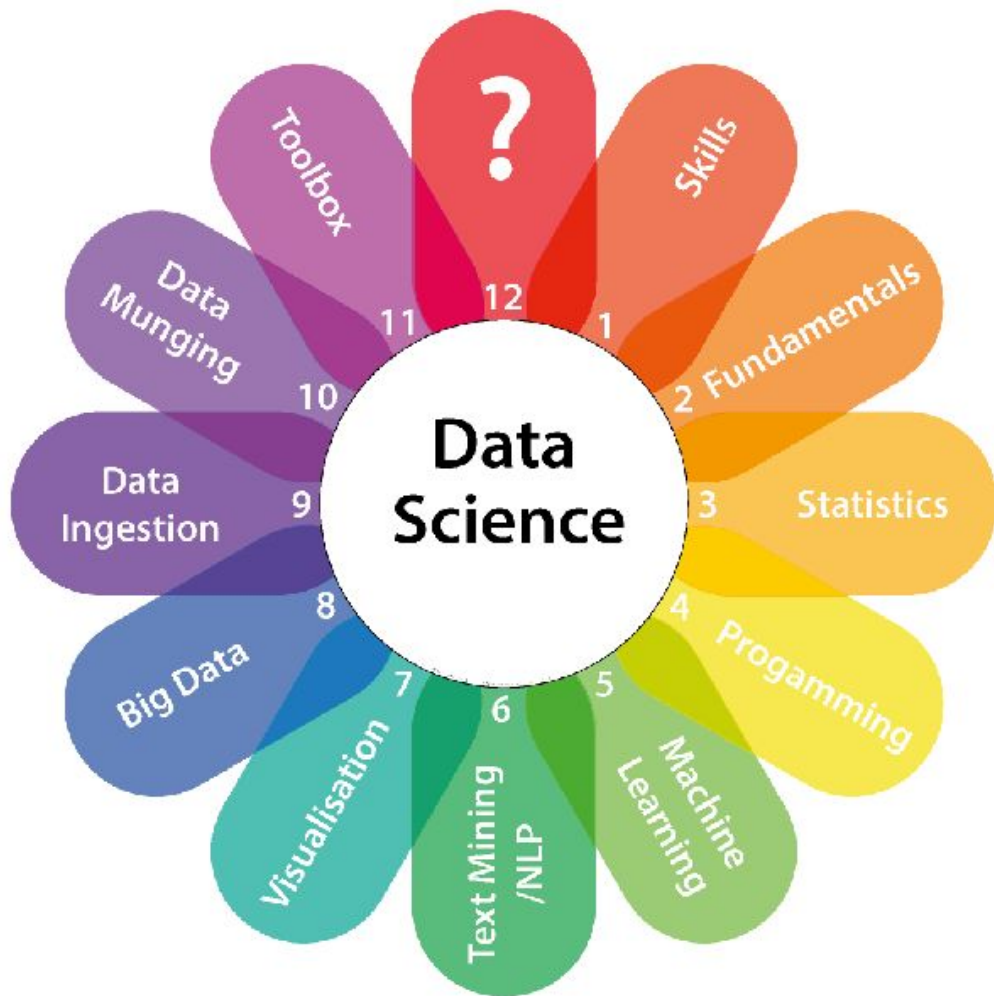
Big Data es el conjunto de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles. Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde **30-50 Terabytes a varios Petabytes**.

Las 3 V's que ahora son 7



Big Data

Podemos pensar el Big Data como una parte de la Ciencia de Datos que requiere una arquitectura y un análisis específico



¿Por qué el auge?



¿Para qué sirve?

- Entender y segmentar clientes (campañas personalizadas)
- Optimizar los procesos de negocio (mejorar rutas)
- Cuantificar y optimizar el rendimiento personal (IOT)
- Mejora de la Salud Pública (análisis genéticos y medicación)
- Mejora del rendimiento deportivo. Esto se puede ver en el tenis, que utiliza el análisis predictivo para determinar patrones.
- Ciencia e investigación (imágenes)
- Mejora de la seguridad y cumplimiento de la ley (identificar anomalías)
- Mejora de las ciudades. Se puede agilizar el flujo del tráfico.
- Trading financiero (intentar predecir comportamiento de los mercados)

¿Qué áreas están involucradas?

Figura 1—Técnicas de la Ciencia de Datos



ALGUNAS INDUSTRIAS BENEFICIADAS POR EL USO DE BIG DATA





**¿Quién puede
hacer ciencia de
datos?: todos/as**

¿Qué competencias se requieren?

- Conocimientos matemáticos y de análisis.
- **Entendimiento del dominio**
- Control de bases de datos como PL/SQL o SQL.
- Habilidades de diferentes lenguajes de programación.
- **Intuición y pensamiento crítico en el análisis de datos.**
- **Capacidad de trabajo en equipo e interdisciplinario**
- **Interpretación de estadísticas y parámetros de mercado.**
- **Capacidad de comunicación**
- **Creatividad para abordar la información**

Ciencias Sociales en Big Data

Cientistas sociales en Big Data

- Capacidad para hacer preguntas interesantes
- Entrenamiento en traducir el interés en objetivos claros
- Ejercicio de armar revisión bibliográfica
- Foco en la relevancia de los determinantes sociales
- Análisis crítico de resultados
- Puesta en contexto de la información
- Capacidad de comunicación o habilidades blandas (muy difícil de reemplazar)
- Evaluación del impacto de los modelos y resultados a nivel social
- “Guiar a los algoritmos hacia un lugar más ético y humanitario” (Boston Review)

Data Harm: ejemplos de daños

Los modelos, si no cuentan con supervisión y análisis crítico de personas, pueden perpetuar o agravar las inequidades y discriminación. Algunos ejemplos donde los algoritmos fueron dañinos:

- Cambridge Analytica
- Recomendador del sistema judicial en EEUU (sesgo racial / de NSE)
- Modelos en reclutamiento de RRHH (sesgo por género)
- Aumento de primas por búsquedas en Internet (errores de interpretación)
- Reducción de crédito por comportamiento de usuarios con compras similares (falacias ecológicas)

**“En la era de los algoritmos, los seres humanos nunca fueron tan importantes”
(Hannah Fry)**

**“Es más fácil desintegrar un átomo que un prejuicio”
(Einstein)**

SOCIEDAD

The tech bias: why Silicon Valley needs social theory

🕒 2 julio, 2018 ✍️ AEON



13,623 views | Aug 6, 2019, 10:32pm

Why Computer Science Needs The Humanities



Kalev Leetaru Contributor

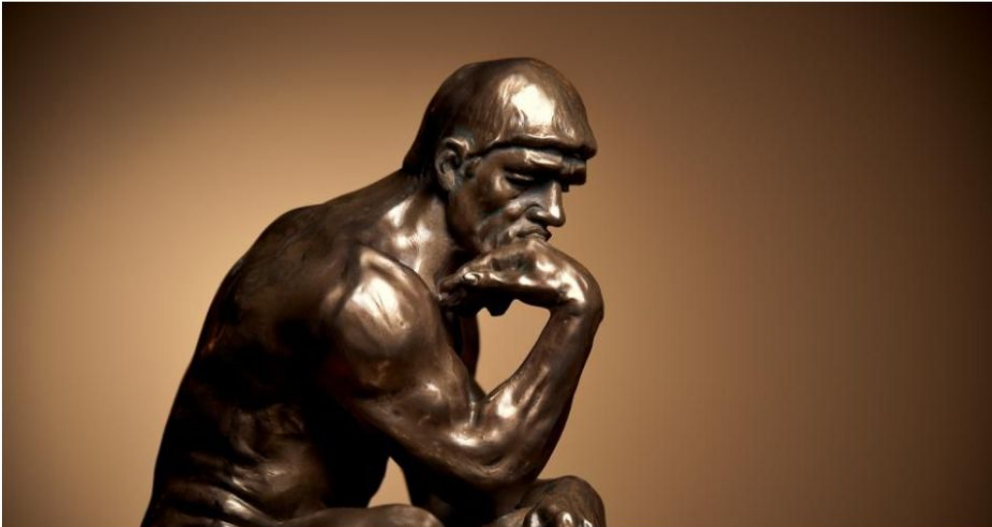
[AI & Big Data](#)

I write about the broad intersection of data and society.

f

t

in



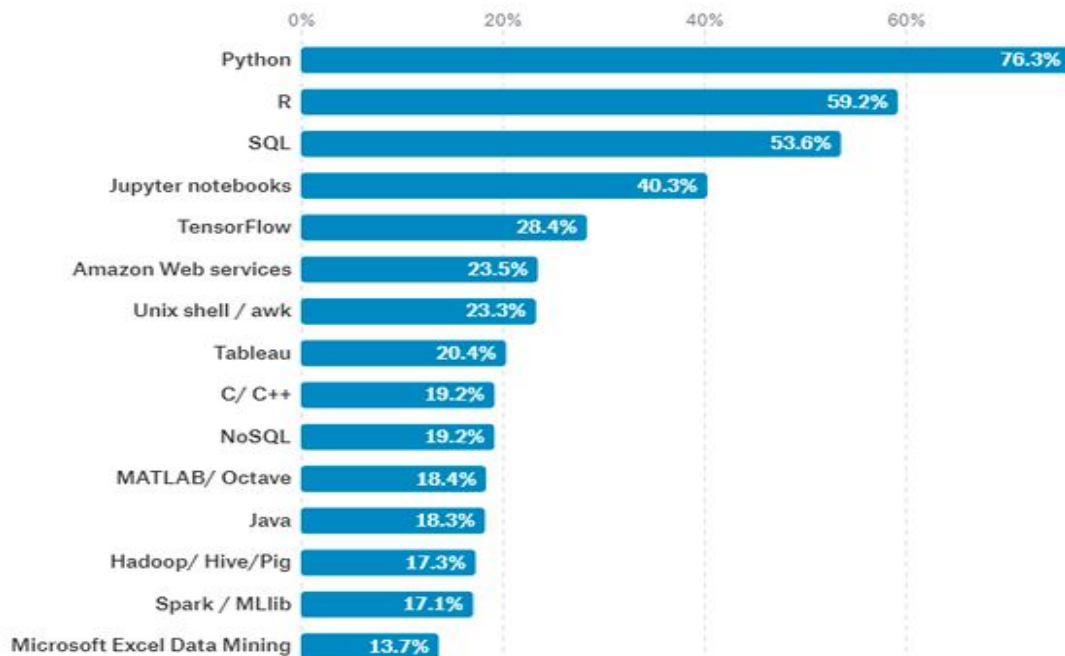
Ad closed by Google

Stop seeing this ad

Why this ad? ⓘ

**¿Y el resto de lo
que piden?**

¿Qué lenguajes son los más populares?



7,955 responses

Only displaying the top 15 answers. There are 38 answers not shown.

¿R o Python?



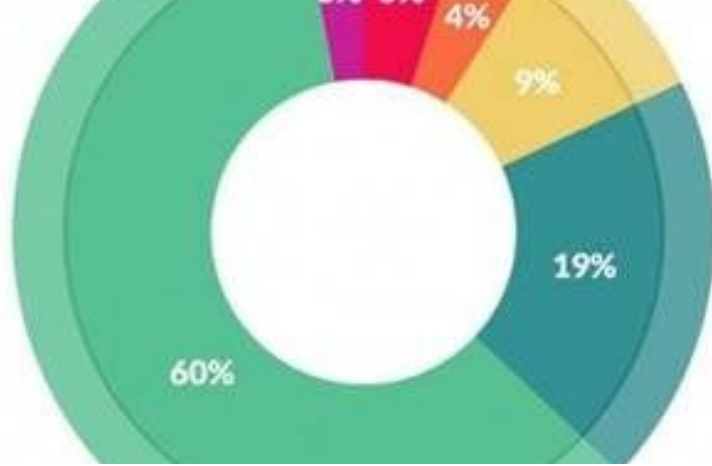
¿Para qué elegir?

```
13
14 ▾ ```{python}
15 import pandas
16 flights = pandas.read_csv("flights.csv")
17 flights = flights[flights['dest'] == "ORD"]
18 flights = flights[['carrier', 'dep_delay', 'arr_delay']]
19 flights = flights.dropna()
20 ```
21
22 ▾ ```{r, fig.width=7, fig.height=3}
23 library(ggplot2)
24 ggplot(py$flights, aes(carrier, arr_delay)) + geom_point() + geom_jitter()
25 ```
26
```

¿Por qué estudiar ciencia de datos?

- Dinamismo de la disciplina
- Desafíos constantes
- Gimnasia analítica
- Alta demanda laboral (por ahora)
- Salarios competitivos
- HYPE





- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

¿Por qué no estudiar ciencia de datos?

- Distribución del tiempo:
 - 80% obtener una base limpia
 - 20% modelado y comunicación de resultados
- Cada vez más competencia
- Cada vez más automatización de recursos
- Es posible que no exista una posibilidad de buena paga en el país si el área de interés es muy específica

¿Qué comunidades hay disponibles?

- Foros (ej. stack overflow)
- R-Ladies
- Pydata
- Data Science Argentina en Facebook
- Meetups
- Eventbrite

Data Science Argentina

Public group

- About
- Discussion
- Mentorship
- Announcements
- Members
- Events
- Videos
- Photos

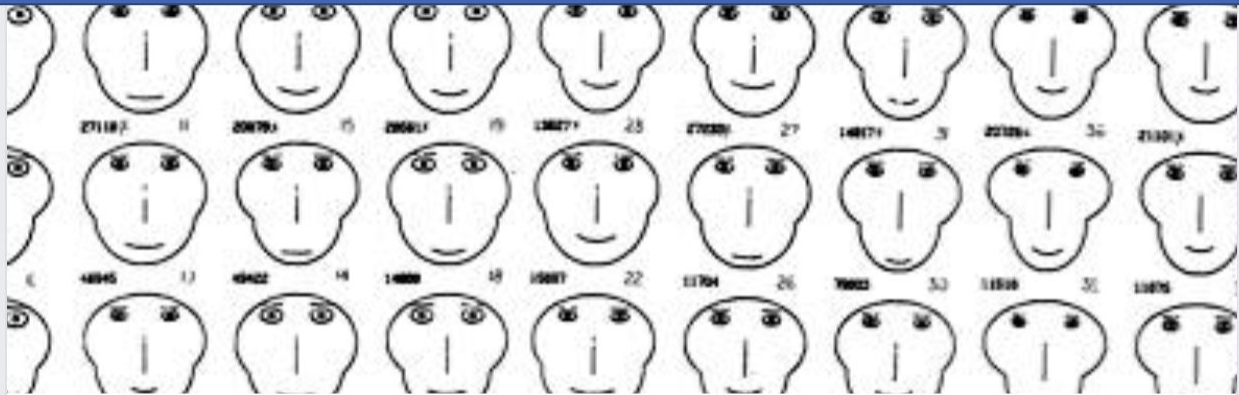
Files

Recommendations

Search this group

Shortcuts

Data Science Argentina



Joined Notifications Share More

Files

Create Doc Upload File

Name	Type	Modified	
 Cómo profundizar en Data Science	Document	Victoria O'Donnell 27 seconds ago	
 Concurso_Beca_Posdoctoral_PICT2016-	PDF	August 12 at 3:10 PM	

Chat (10)



Crear un nuevo grupo | Iniciar sesión | Registrarse



R en Córdoba

📍 Córdoba, Argentina
👤 323 miembros · Grupo público
👤 Organizado por **Patricio Del B.** y otras 2 personas

Compartir:

[Sobre nosotros](#) | [Eventos](#) | [Miembros](#) | [Fotos](#) | [Conversaciones](#)

Unirse a este grupo



Crear un nuevo grupo | Iniciar sesión | Registrarse

```
library(dplyr)
rladies_global %>%
  filter(city == 'Córdoba')
```



Parte de **R-Ladies** – 165 grupos

R-Ladies Córdoba

- Córdoba, Argentina
- 102 miembros · Grupo público
- Organizado por R-Ladies G. y alguien más



Compartir:

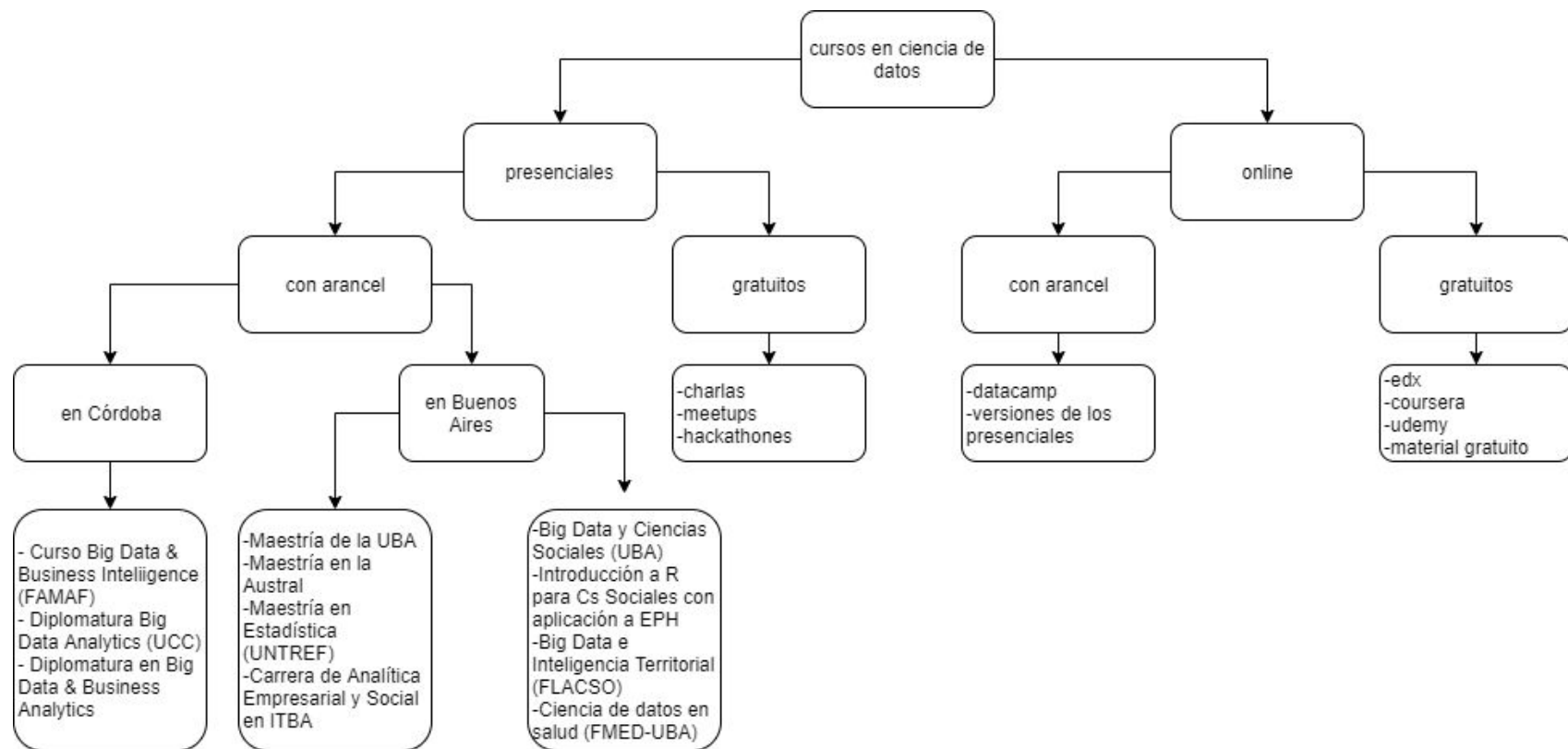
[Sobre nosotros](#) | [Eventos](#) | [Miembros](#) | [Fotos](#) | [Conversaciones](#)

[Solicitar unirse](#)

Lo que hacemos

Organizadores

¿Cuál es la oferta de formación?



¿Dónde puedo conseguir datos?

- INDEC
- Datos Abiertos Córdoba
- Kaggle Datasets
- Apis de redes sociales (como Twitter por ejemplo) y descargar
- Bases específicas por tema en Google

Datos Abiertos

¿Qué dato buscás?

Economía

Sociedad

Territorio

Censos

Todos los DataSet

Datos Abiertos Destacados

Índice del Costo de la Construcción

¿Qué preguntarse?

- ¿Quiero hacer ciencia de datos? ¿Por qué y para qué?
- ¿En qué dominio o campo de aplicación me interesa aplicarla?
- ¿Qué tarea de la ciencia de datos me interesa más?
- ¿Qué herramientas o técnicas son relevantes para lo que me interesa?
- ¿Qué oferta de formación hay disponible en lo que me interesa?
- ¿Cuánto tiempo y plata tengo disponible para mi formación?
- ¿Quiénes son los referentes del área y cuál es la mejor forma de acercarme a esa red?

¿Cuál sería el paso a paso recomendado?

- Definir área de interés (tanto el perfil como el campo de aplicación)
- Investigar la oferta que más se ajusta a lo que me interesa considerando los recursos que tengo disponibles
- Identificar y contactar referentes en el campo que me interesa
- Probar con cursos cortos o gratuitos; si me sigue interesando avanzar a otros
- Dentro de lo posible y si no sé: aprender inglés
- Unirme a las comunidades relevantes: en esas espacios se comunican mucha oferta de formación, eventos y oportunidades de trabajo
- Participar de alguna competencia o desafío como Kaggle, Tidy Tuesday
- Abrir una cuenta de github, blog de medium, cuenta de Twitter profesional, etc
- Actualizarse constantemente
- Asistir a conferencias, meetups, etc

¿Cómo buscar trabajo?

- LinkedIn (pedir validaciones)
- Comunidades (slack por ejemplo)
- Twitter (seguir cuentas de referentes)
- Grupos de facebook (de Data Science o Lenguajes)
- Recomendaciones de conocidos
- Networking en eventos de tipo Meetup
- Dar a conocer las páginas de los proyectos que realicen

¡Gracias!

vickyodonnell@gmail.com

Referencias

- <https://www.argentina.gob.ar/modernizacion/setic/grupo-de-trabajo/big-data>
- <https://www.powerdata.es/big-data>
- <https://www.master-bigdata.com/10-usos-reales-big-data/>
- <https://qph.fs.quoracdn.net/main-qimg-ef7afb3dea1e5e124168629caa435d3a-c>
- <https://www.questionpro.com/es/que-es-big-data.html>
- https://cran.rstudio.com/web/packages/reticulate/vignettes/r_markdown.html
- <https://datosestadistica.cba.gov.ar/>
- <https://www.meetup.com/es/rladies-cordoba/>
- <https://www.meetup.com/es/R-en-Cordoba/>
- <https://fjs.ucc.edu.ar/curso.php?id=12645>
- <https://fjs.ucc.edu.ar/curso.php?id=12645>
- https://www.facebook.com/groups/DataScienceArgentina/?ref=group_header
- <https://21.edu.ar/content/diplomatura-en-big-data-business-analytics>
- <https://datajusticelab.org/data-harm-record/>
- <http://bostonreview.net/science-nature/tom-slee-silicon-valley-liberal-arts-majors-we-want-you>
- <https://collateralbits.net/the-tech-bias-why-silicon-valley-needs-social-theory/>

