VISUALIZACIÓN DE DATOS

MSc Carlos Córdova BSc Carlos Ramírez





Introducción

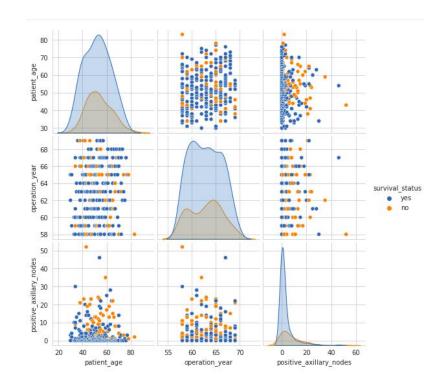
¿Por qué es importante la visualización de datos?

• • • • • •

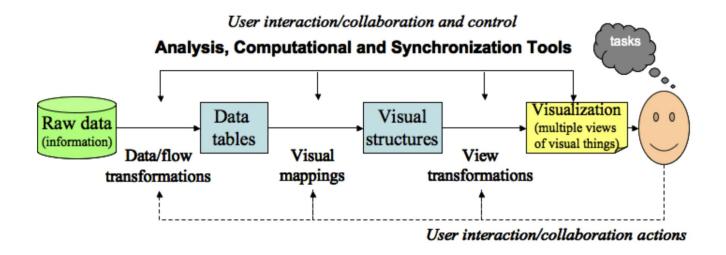
"Una imagen vale más que mil palabras"

La visualización de datos es importante porque:

- Permite comprender r\u00e4pidamente grandes vol\u00e4menes.
- Crucial como paso previo al entrenamiento de algoritmos de ML.
- Facilita la detección de patrones, tendencias y anomalías.
- Ayuda a comunicar los hallazgos de manera clara y efectiva.



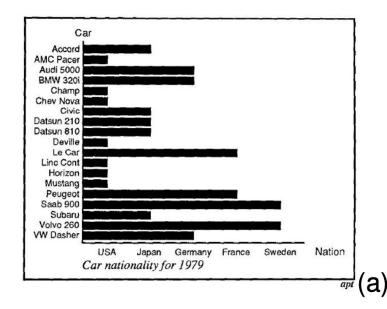
Pipeline de visualización

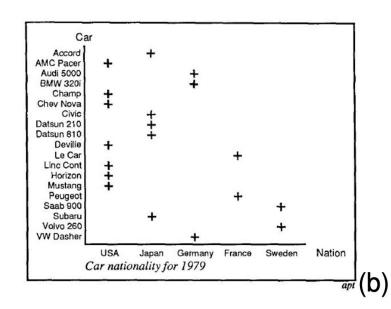


- Primer paso: preprocesamiento y transformación de datos.
- Segundo paso: mapear los datos en representaciones gráficas.
 - o Propiedades gráficas (e.g., posición, forma, color, textura, etc)

¿Qué visualización será la correcta?

• •

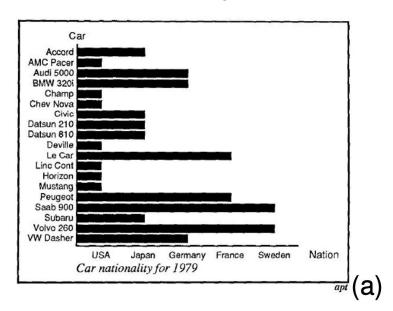


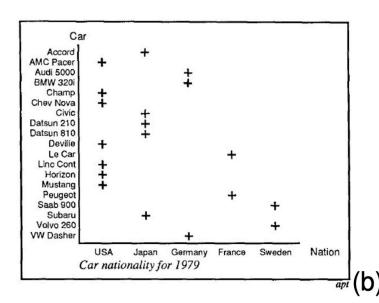


¿Qué visualización será la correcta?

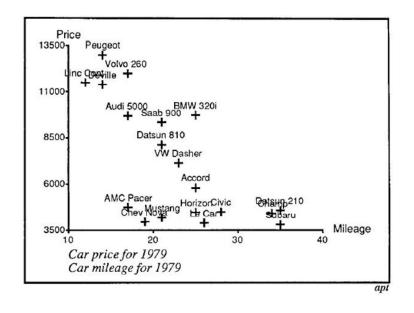
- Mal uso de un gráfico de barras.
- Buen uso de un gráfico de dispersión.

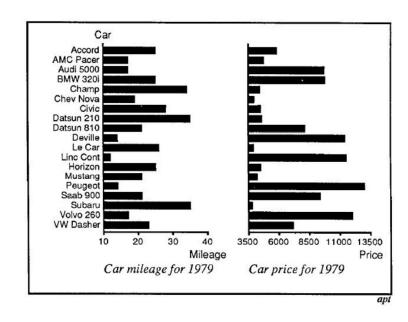
La visualización debe escogerse correctamente.





¿Qué carro por encima de 11,000\$ presenta mayor kilometraje?

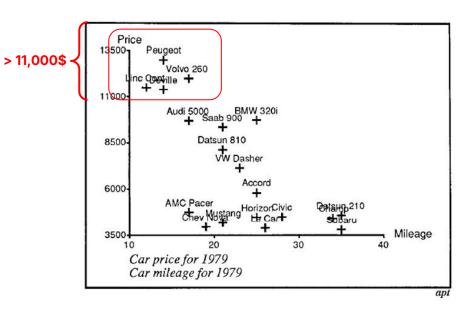


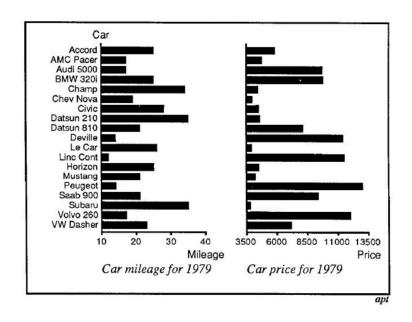


• •

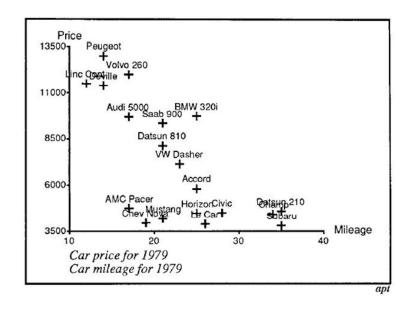
¿Qué carro por encima de 11,000\$ presenta mayor kilometraje?

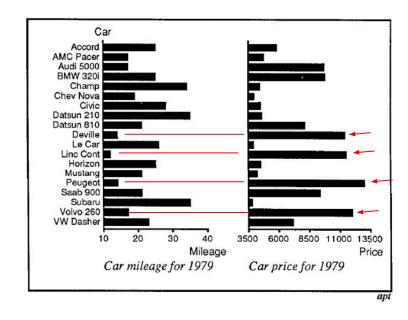
• •





¿Qué carro por encima de 11,000\$ presenta mayor kilometraje?





La **calidad** de una visualización se mide por:

Expresividad

o Transmitir toda la información deseada.

Efectividad

o Ser interpretada correcta y rápidamente.

Análisis de un gráfico

Percepción Inicial:

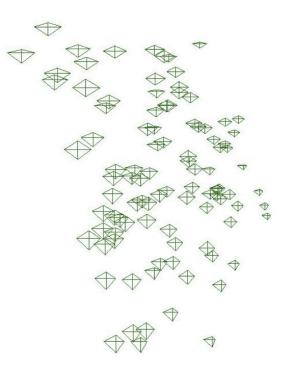
 Percibimos grupos de objetos de manera intuitiva.

Categorización Cognitiva:

 Intentamos categorizar estos grupos y darles un sentido.

• Análisis de Casos Especiales:

 Analizamos elementos sin relación dentro de los grupos o elementos sin grupos.







LAS 8 VARIABLES VISUALES

¿Cómo representar los datos visualmente?

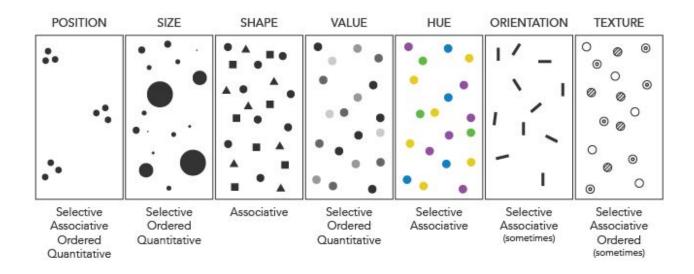
.

Las 8 variables visuales

• Los datos se representan mediante el mapeo de diferentes ítems y atributos en variantes visuales.

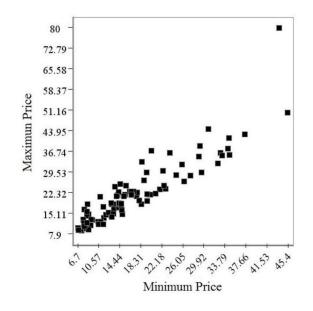
• Codificación Visual:

- Hasta 8 variables visuales pueden ser codificadas.
- Organizadas por prioridad para transmitir información de manera efectiva.



Posición

- 1D, 2D, 3D.
- La lectura inicia por la distribución espacial.

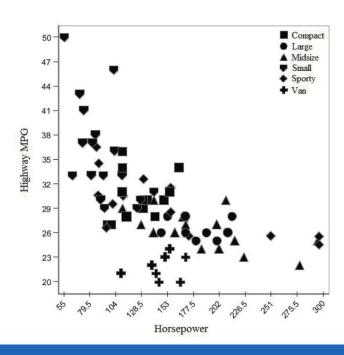




Marcador

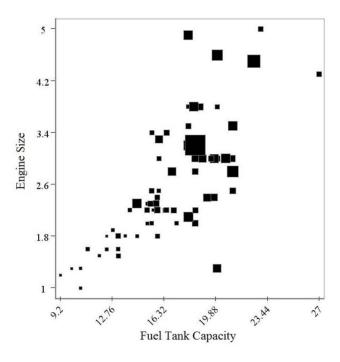
- Puede ser cualquier elemento gráfico (e.g., símbolos, letras)
- Lo más diferentes entre sí.
 - Fácilmente distinguibles.
 - Evitar predominancias.





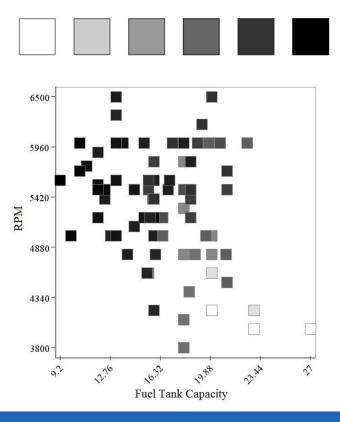
Tamaño

- Permite mapear atributos continuos y categóricos.
- Esta variable y las siguientes solo afectan la apariencia de las anteriores.



Brillo

Usar un conjunto reducido de variantes de brillo.

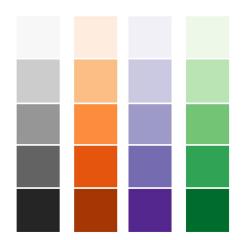


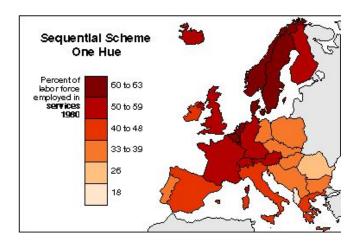
Color

• Los colormaps (mapas de colores) determinan como son mapeados los valores en colores.

Colormaps secuenciales:

- o Para datos ranqueados.
- o Gradientes de un mismo color que varían en intensidad.





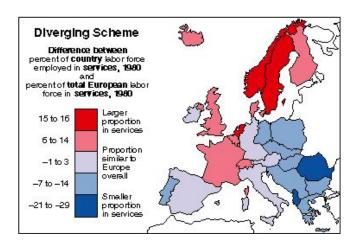
Color

- Los colormaps (mapas de colores) determinan como son mapeados los valores en colores.
 - Generadores de paletas <u>online</u>.

• Colormaps divergentes:

- Para datos con un punto medio reconocible (e.g., desviaciones respecto a una media, positivos y negativos respecto del punto 0)
- o Gradientes de un mismo color que varían en intensidad.



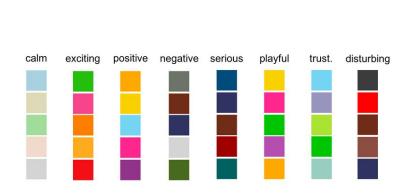


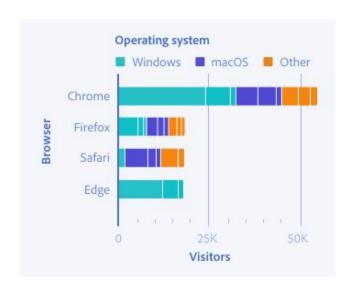
Color

• Los colormaps (mapas de colores) determinan como son mapeados los valores en colores.

• Colormaps categóricos:

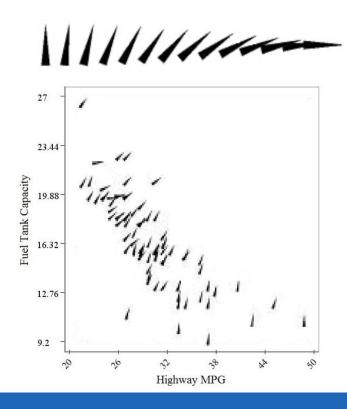
- Para datos categóricos.
- Distintos colores sin orden específico.





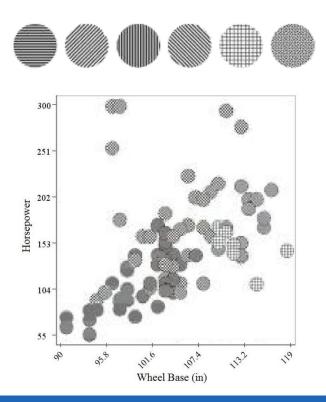
Orientación

- Se procesa de manera intuitiva.
 - o Debe aplicarse solo a marcadores que tengan una orientación natural.



Textura

• Combina diversas variables gráficos (e.g., forma, color, orientación)



Movimiento

- Se puede asociar a cualquiera de las otras variables visuales.
 - Puede indicar variación temporal.



VISUALIZACIONES

Algunos de los gráficos más usados

• • • • • •

• •

• •

• •

•

Visualizaciones Comparativas

Gráfico de barras

- Fácil de entender.
 - Muestra diferencias claras entre categorías.
- No es recomendado para datasets muy grandes.

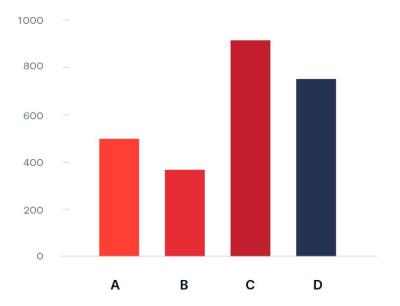


Gráfico de barras

- Variantes:
 - o Agrupadas
 - Comparan una variable bajo una misma escala.

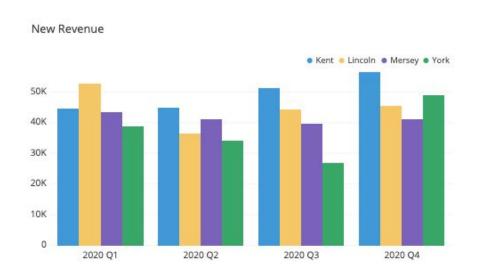
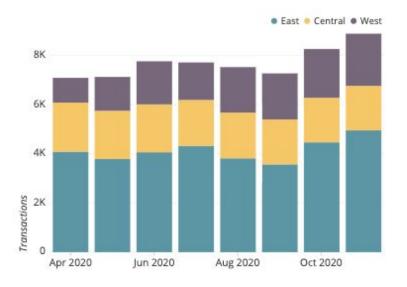


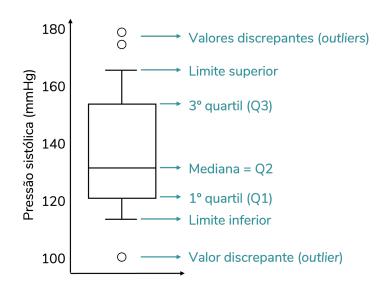
Gráfico de barras

- Variantes:
 - Apiladas
 - Describen la composición de una variable y permiten la comparación.



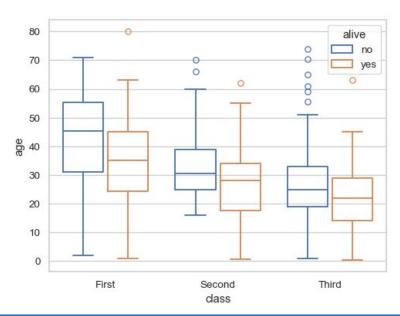
Boxplot (Diagrama de caja)

- Permite comparar la distribución de los datos entre grupos.
- Identifica:
 - Medianas.
 - Rangos intercuartiles.
 - o Outliers.
- Su interpretación no es tan simple.



Boxplot (Diagrama de caja)

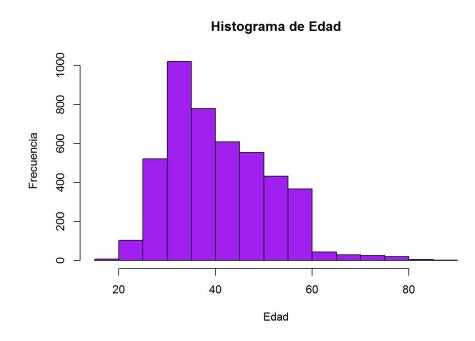
- Permite comparar la distribución de los datos entre grupos.
- Identifica:
 - Medianas.
 - Rangos intercuartiles.
 - Outliers.
- Su interpretación no es tan simple.



Visualizaciones de distribución

Histogramas

- Muestra la distribución de datos continuos.
- Permite identificar la asimetría y dispersión de datos.
- El número de bins es muy determinante.



Visualizaciones temporales

Gráfico de líneas

- Muestra tendencias a lo largo del tiempo
- Fácil de entender.
- La comparación puede degradarse por el número de líneas.

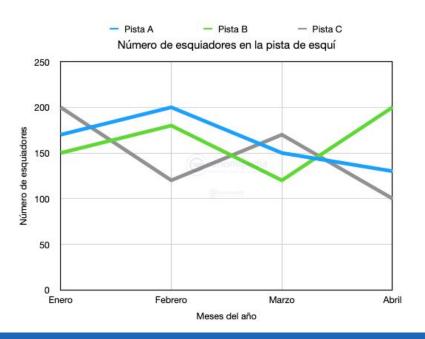
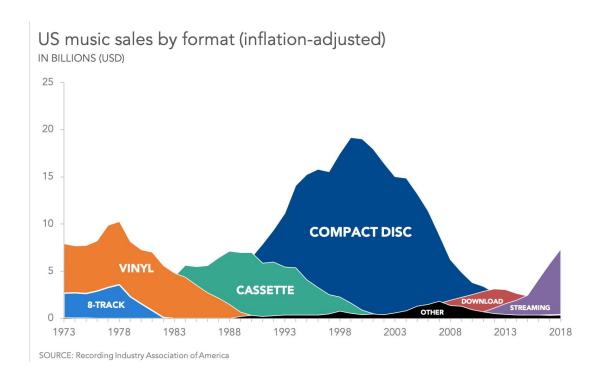


Gráfico de área

- Destaca más el volumen.
- Permite comparar las composiciones en el tiempo.



Visualizaciones de relaciones

Gráfico de dispersión (Scatter plot)

- Permite identificar tendencias.
- Útil para detectar outliers.

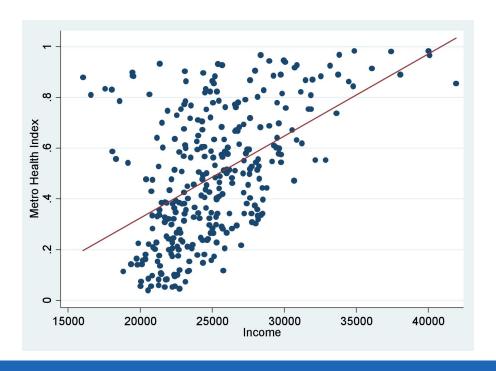
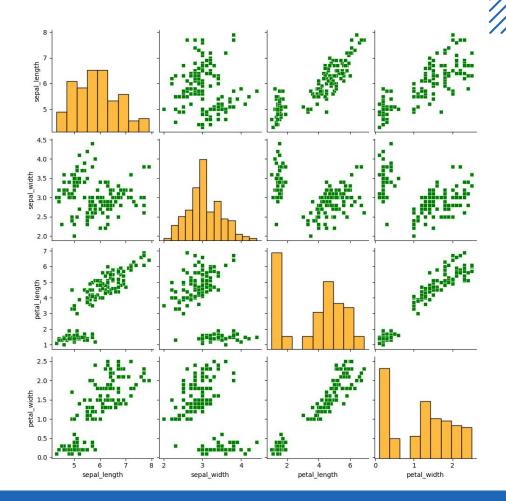


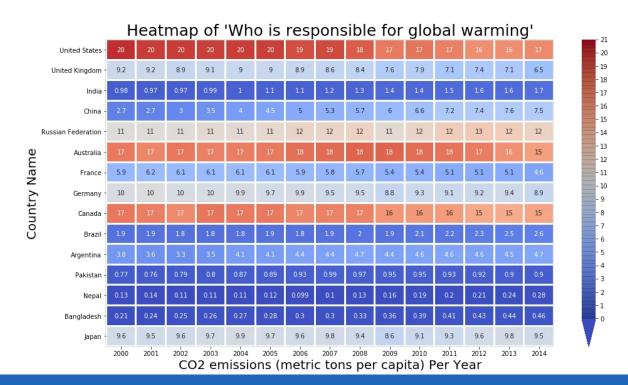
Diagrama de pares (Pair plot)

 Permite visualizar relaciones entre varios pares de variables



Heatmap (Mapa de calor)

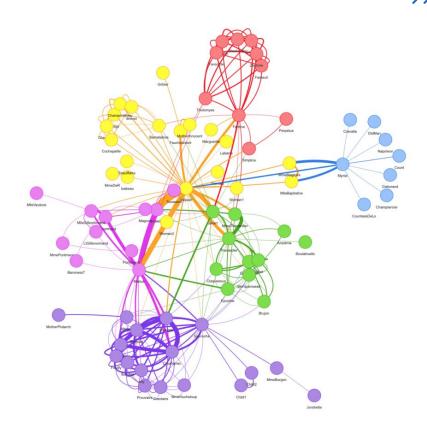
- Los valores se representan con colores.
 - Paleta divergente.
- La elección de los colores es importante porque puede generar confusión.



Otras visualizaciones

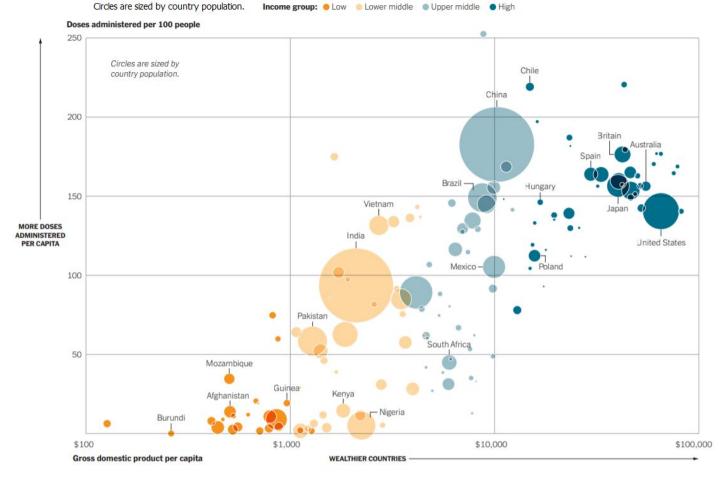
Redes

- Permite ver las relaciones, encontrar comunidades.
- Permite mapear la intensidad entre las relaciones.
- Puede estar sobrecargado cuando hay muchos nodos y aristas, es difícil identificar nodos específicos.
- Herramienta como <u>Gephi</u> pueden generar estos gráficos.





Vacunación por país



Sources: Vaccination data from local governments via Our World in Data; income classifications and gross domestic product data from the World Bank. | Note: Data is as of Dec. 8.

Glossary: G.D.P. per capita is the Gross Domestic Product, or wealth of a country divided by its population size.

• Ejemplos (<u>NY Times</u>)

Diversidad en los deportes

. .

Racial disparity in major sports leagues

■ Players of color ■ Head coaches of color

National Football League	Women's National Basketball Association	National Basketball Association	Major League Baseball	Major League Soccer
2020	2019	2020	2020	2020
74%	83%	83%	40%	60%
13%	25%	30%	20%	41%

Note: Managers are shown instead of head coaches for M.L.B. The most recent available data is used for each league. The data is self-reported and may not include all league members.

Source: Institute for Diversity and Ethics in Sport

• Ejemplos (NY Times)

Los trabajos post pandemia

• •

How Might Future Jobs Be Affected by the Pandemic

The 10 industries in which the "strong-impact scenario" differs most from the original projection, both positively and negatively.

Epidemiologists			+25.3%	
Medical scientists, except epidemiologists			+23.2%	
Web developers and digital interface designers			+10.5%	
Biochemists and biophysicists	+10.0%			
Network and computer systems administrators			+9.8%	
Computer network architects			+9.7%	
Information security analysts			+9.0%	
Microbiologists			,	
Biological technicians			+5.6%	
Database administrators and architects			+5.4%	
Hosts and hostesses, restaurant, lounge and coffee shop				
Bartenders	-18.6%			
Reservation and transportation ticket agents and travel clerks	-16.7%			
Hotel, motel and resort desk clerks -16.2%				
Vaiters and waitresses -16.0%				
Receptionists and information clerks -13.5%				
Cashiers -13.5				
Flight attendants -11.79				
Subway and streetcar operators				
Bus drivers, transit and intercity	-10.9%			



Source: B.L.S. • By The New York Times

¿Preguntas?