



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN  
IIC2026 — VISUALIZACIÓN DE INFORMACIÓN

## Ayudantía 4

### Repaso Tarea 1

31 de Agosto

### Primer caso: Elecciones presidenciales EUA 2020

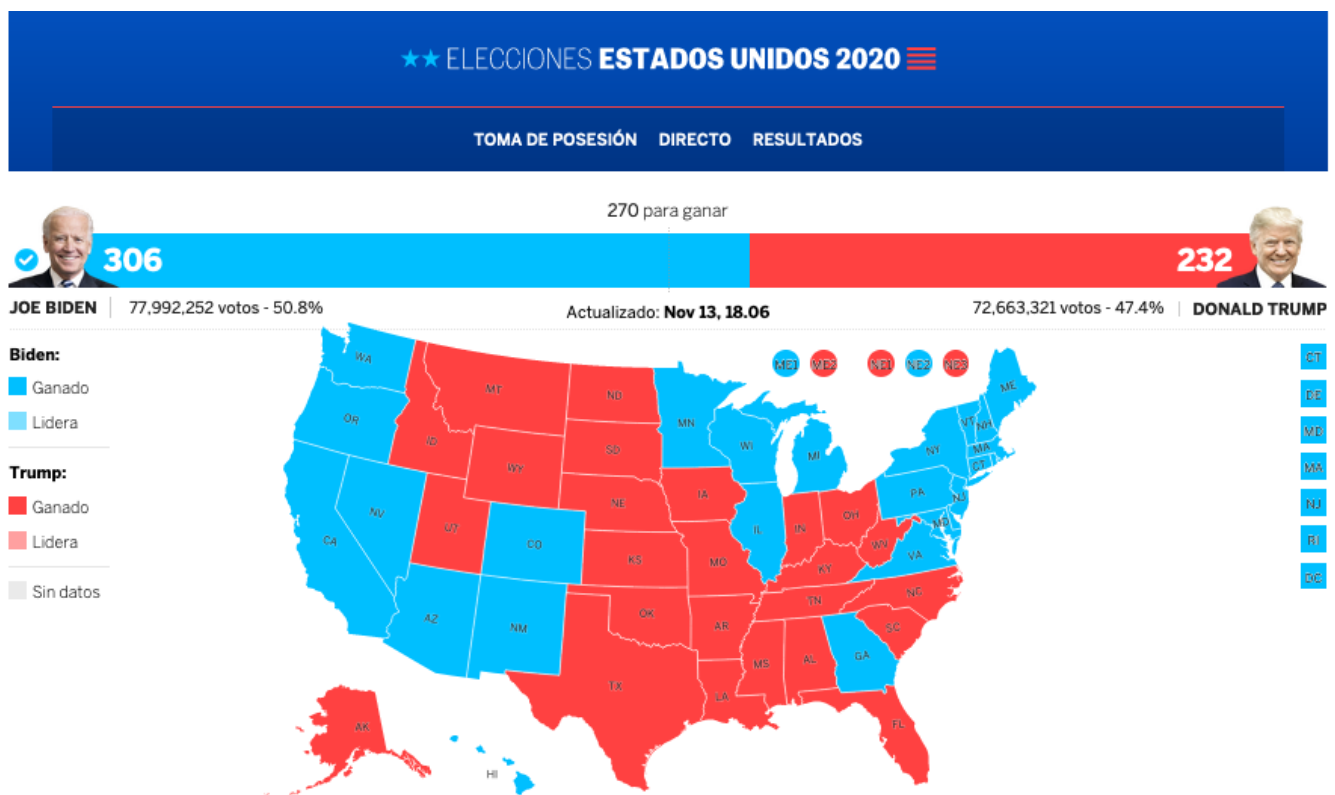


Figura 1: Captura de pantalla de visualización interactiva sobre las elecciones presidenciales de EE.UU. 2020. La visualización original la pueden encontrar [aquí](#).

### Visualización

El sistema de elección de presidentes de EE.UU. es uno interesante, donde las personas del país votan indirectamente por quien se volverá presidente. Indirectamente, porque no se considera la mayoría de todos los votos a lo largo del país para realizar la decisión, sino que a nivel de estado la mayoría determina a quienes irían cierta cantidad de votos del estado. Eso se conoce como El Colegio Electoral, y produce que ciertos estados tengan mayor influencia en el resultado, ya que tienen mayor cantidad de votos, o electores. Quien obtenga la mayoría de votos de electores de estados, es quien gana la elección.

Por eso, es usual que visualizaciones de los resultados, como en el caso de estudio, se muestren los

resultados a nivel de estados. En este caso se usa un mapa con los estados, y los colores muestran el ganador en cada estado, o quien lleva la delantera (considerando que la visualización se actualizaba en vivo mientras se realizaban los conteos de votos). Eso a su vez comunica luego a qué candidato irían los votos electores del estado. Para el caso de estados muy pequeños, geográficamente al este del país, se agregan pequeños cuadrados que les representa, para hacerlos más visibles. También, debido a dos estados excepción, Maine y Nebraska, donde la decisión de votos se toma a nivel de distritos en vez del estado completo, se agregan círculos pequeños que muestran cada distrito por separado.

Sobre el mapa se muestra un resumen del resultado final y principal, que es mostrar quien gana o ganaría la elección. Las barras muestran el total de votos de electores que fueron a dos de los candidatos principales (306 para Biden y 232 para Trump). Bajo ella se muestra también el total de votos individuales que obtuvo cada candidato.

### 1.1 Descomposición y clasificación

La **marca** predominante es de áreas (items<sup>1</sup>). Cada área representa un estado y por lo mismo tienen la forma de estos, el color de estas se ve representado por ganador en el estado (azul-demócrata, rojo-republicano). También se presentan algunas excepciones, ya que algunos estados no logran ser visibles, por lo que se incluyen a la derecha en unos cuadrados. Otra marca presente de tipo contención es la forma a EUA, conteniendo los estados. Por último, las barras en la parte superior son marcas unidimensionales que representan la cantidad de puntos que tiene cada candidato. Los **canales** presentes en estas marcas son

- Color, se observa relación clara entre azul-demócrata, rojo-republicano y gris-sin datos, permitiendo de esta forma diferenciar los estados y los puntajes de cada candidato. Usándose, el color como un canal de identidad.
- Largo de las barras superiores tienen una relación con la cantidad de puntos de cada representante, usando este canal como uno de magnitud.
- La forma de EUA ayuda a entender de qué se hablan datos correspondientes a ese país. Aplicándose se como un canal de identidad.

La visualización está compuesta por datos del tipo ítem y atributo. Los ítems considerados son los estados y los candidatos (J. Biden y D. Trump). Los atributos correspondientes al ítem de estados son cantidad de puntos que este entrega y cantidad de votos para cada candidato (solo visibles cuando se coloca el *mouse* encima), además los nombres de estados (siglas) también son consideradas como atributos. Por otro lado, el ítem de candidato tiene como atributo cantidad de votos totales que obtiene (y su porcentaje), la cantidad de puntos totales y nombre del candidato. El tipo de *dataset* original era dinámico (actualmente no porque esos datos no cambian), ya que se iba actualizando constantemente con la nueva información de las elecciones.

Las clasificaciones de los atributos serían las siguientes:

- Cantidad de puntos que entrega cada estado, es un dato ordenado, cuantitativo, secuencial, no cíclico y de valor.
- Cantidad de votos que tiene cada estado, atributo ordenado, cuantitativo, secuencial, no cíclico y de valor.

---

<sup>1</sup>Como dice Tamara Munzner en su libro *"...on table datasets, where a mark always represents an item"*. De esta forma, un dataset tabular siempre tiene marcas de tipo ítem.

- Siglas de cada estado, atributo categórico, de llave.
- Cantidad de puntos y cantidad de votos que tienen cada candidato son atributos ordenados, cuantitativos, secuenciales, no cíclicos y de valor.
- Nombres de los candidatos son atributos categóricos de valor.

## 1.2 Críticas y propuestas

De la propuesta se observa que cumple el **principio de expresividad**. Se observa que la cantidad de puntos que obtiene cada candidato se representa en el largo de la barra superior, de esta forma un atributo ordenado se representa correctamente en un canal de magnitud. Por otro lado, el color se representa como un canal de identidad, es decir, no busca orden si no qué categoriza. Finalmente, la contención de los estados da forma a EUA, permitiendo identificar que se trata de EUA la visualización, por lo que es utilizado de manera correcta el canal de identidad.

Se observa que la información más importante que trata de entregar esta visualización es la cantidad de votos entre los dos candidatos a tiempo real. Con esto, podemos mencionar que la información principal que se trata de entregar logra resaltar al utilizar longitud como canal adecuado para cantidad de votos para cada candidato. Luego, la información secundaria, como el estado de votaciones de un estado, quedan representados por canales menos efectivos. Con esta información, se puede decir que la propuesta cumple con el **principio de efectividad**. No obstante, no se cumple en su totalidad porque disponer las barras una vs. la otra hace que la estimación de qué candidato tiene más votos sea difícil cuando la diferencia de votos es mínima, por lo tanto, se vuelve más difícil determinar un orden de los candidatos en dicho caso, pero cuando la diferencia de votos se vuelve mayor, este canal es bien efectivo.

Respecto a la percepción de la visualización, se logra el criterio de agrupación, ya que se perceptivamente agrupamos los elementos de colores similares, dando rápidamente una idea general respecto a como se distribuyen los votos para cada candidato en torno al territorio norteamericano. Los ítems de candidatos se distinguen inmediatamente por su color, por lo que su valor de canal (canal de identidad) logra resaltar los atributos de estos. Cumpliendo así, con el criterio *popout*.

Siguiendo con el color, se observa que la propuesta cumple con el principio de diseño gráfico de la **consistencia**. Se utilizan los colores de los partidos políticos que apoyan a los candidatos como el color en la visualización. Además, presenta una muy buena **autocontención**, ya que es muy autoexplicativa, presenta una leyenda adecuada y textos capaces de contextualizar cierta información.

Finalmente, los colores utilizados siguen cumpliendo su función luego de modificar la visualización con un efecto en blanco y negro. Se sigue observando un contraste notable entre los dos candidatos. Por esto se concluye que cumple con el principio de **lograrlo en blanco y negro**.

Además, el **data ink ratio** en la propuesta se acerca bastante a 1. Las barras superiores ayudan a entregar una mejor percepción sobre la cantidad de votos. Mientras que los estados (áreas) logran resaltar, de manera útil, como se distribuyen los puntos en los estados que componen a EUA. Además, la leyenda y estados (y distritos) que son representados afuera de la contención de EUA, permite aclarar ciertos puntos de la visualización que de otra forma no lograrían ser visibles. Cabe destacar, que las fotos de cada candidato no aportan a la visualización.

Para esta visualización se recomiendan las siguientes propuestas

- Eliminar las fotos de los candidatos para justificar la mayoría de los píxeles utilizados, ya que estas no aportan a una mejor comprensión de los datos. Así, se mejoraría el *data ink ratio*.

- Crear una columna en la parte derecha de la visualización que guarde los Estados que han sido seleccionados y sus respectivos votos. De esta manera se puede comparar mejor los votos que han recibido los candidatos según las zonas que se seleccionen. Esta propuesta haría que la visualización cumpliera con el principio de IHC “los ojos le ganan a la memoria”, ya que se podría ver los datos mientras se compara, al contrario de la visualización actual, donde si alguien quiere comparar los datos entre dos Estados, tiene que memorizar alguno de los dos.

## Segundo caso: Energías limpias y seguras

# What Are the Safest and Cleanest Energy Sources?

Data: Our World in Data, 2020. Square root scales for x and y axes.

Graph by: @rubenbmathisen (Twitter)

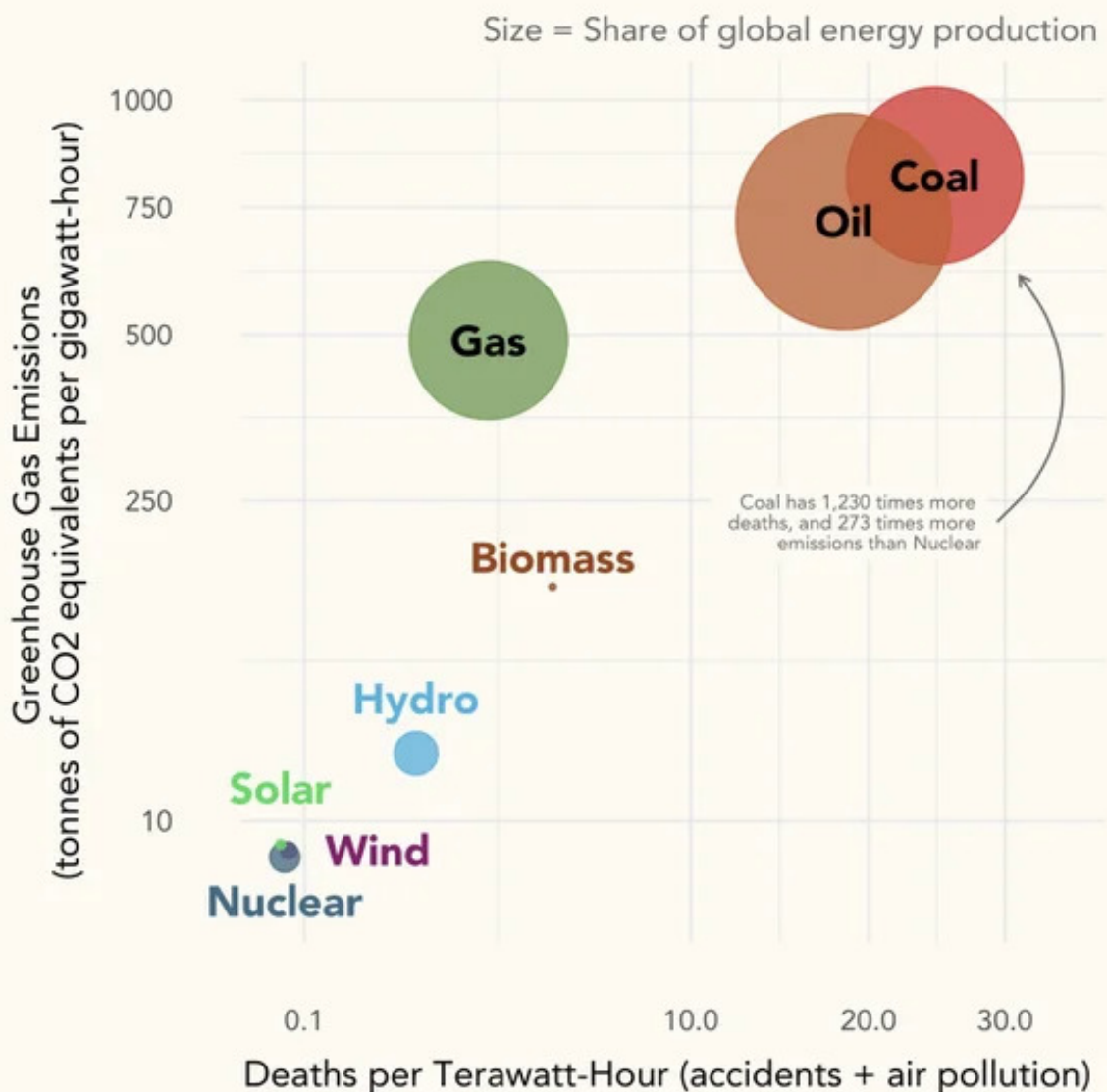


Figura 2: Gráfico sobre las energías más seguras y menos contaminantes. Pueden encontrar la visualización original [aca](#).

La siguiente visualización muestra gráficamente las diferentes fuentes de energía representadas por círculos en el gráfico. Cada círculo (tipo de fuente de energía) tiene un tamaño que corresponde a cuanto de la energía global viene de esta fuente (mientras más grande, más energía de esa fuente). Además, cada fuente energética está representada por un color distinto. El **eje x** representa la cantidad de muertes por TeraWatt-Hora en una escala de raíz cuadrada, mientras que el **eje y** representa la emisión de gases de efecto invernadero (en toneladas de  $CO_2$  por GigaWatt-Hora) de cada fuente energética en una escala de raíz cuadrada.

## 2.1 Descomposición y clasificación

Se observa que la **marca** principal de este gráfico es de ítem, específicamente de áreas. Cada área representa una fuente de energía, y tienen un tamaño específico según cuanta energía global es producida por estas. Los **canales** principales son

- Color para poder diferenciar entre cada tipo de fuente de energía, usando este canal como uno de identidad.
- Tamaño de las circunferencias indica la cantidad de energía que estos suministran a la red de energía global, de esta forma actúa como un canal de magnitud.
- La posición horizontal en una escala común de los círculos indica cantidad de muertes para cada fuente energética, por lo que se usa como un canal de magnitud.
- La posición vertical en una escala común indica cantidad de contaminación generada por cada fuente energética, por lo que se usa como un canal de magnitud en un eje.

Respecto a los datos presentes en la visualización, se observan datos de tipo ítem y atributos. Los ítems presentes son las distintas fuentes de energías y que se diferencian a través del color. Estos ítems presentan cuatro atributos claves, el primero es el nombre que recibe cada fuente energética. El segundo atributo, cantidad de energía global que cada fuente energética produce. El tercer atributo es la cantidad de muertes por TeraWatt-Hora. Y el último atributo es la cantidad de emisiones de gases de efecto invernadero por GigaWatt-Hora.

La clasificación de los atributos sería la siguiente:

- El nombre que recibe cada fuente energética es un atributo categórico (las fuentes de energía no tienen un orden intrínseco). También se puede considerar como llave, ya que nombre es único y sirve para identificar el tipo de fuente energética.
- La cantidad de energía que genera cada fuente energética es un atributo ordenado, cuantitativo, secuencial, no cíclico y de valor.
- La cantidad de muertes por TeraWatt-Hora es un atributo ordenado, cuantitativo, secuencial, no cíclico y de valor.
- La cantidad de emisiones de gases de efecto invernadero por GigaWatt-Hora es un atributo ordenado, cuantitativo, secuencial, no cíclico y de valor.

## 2.2 Críticas y propuestas

De la propuesta visual se logra observar que el **principio de expresividad** se cumple y se respeta totalmente, ya que se expresa la totalidad de la información del dataset y cada atributo se muestra en el canal

adecuado. Los colores de las fuentes de energía indican el tipo y estos no tienen ningún orden establecido, por lo que está bien que un atributo categórico sea expresado a través de un canal de identidad. Respecto a la posición horizontal y vertical, se observa que estas se basan en escalas comunes para cada eje, de esta forma se expresan atributos cuantitativos a través de un canal de magnitud. Por último, el tamaño representa la cantidad de energía que cada fuente de energía produce a nivel global, de esta forma, se usa un canal de magnitud con un atributo ordenado.

Respecto al **principio de efectividad**, este se cumple parcialmente. El nombre de la visualización sugiere que lo más importante son las fuentes de energía más limpias y seguras, de esta forma se espera que esta información esté codificada con los canales más efectivos posibles. Como se ve en la visualización, todos comparten canales de magnitud como posición horizontal y vertical en una escala común, el canal de posición es el más importante y codifica la información clave. Se indicó que se cumple **parcialmente** el principio de efectividad porque la escala de raíz cuadrada ocupada para posicionar los datos no es tan efectiva para la estimación de datos en comparación a una escala lineal, pero se entiende que esta decisión se tuvo que tomar para asegurar que todos los datos se vieran en pantalla. Por otro lado, la visualización utiliza el canal de tamaño, el cual es un canal de menor importancia, pero se usa para un atributo secundario.

La percepción de la visualización no es la ideal, ya que se ven ciertos tipos de fuentes energéticas agrupadas sin que estas tengan una relación directa, lo que nos lleva a ver ciertos grupos como una unidad. De esta forma se rompe con el criterio de **agrupación**. Además, se usan dos canales (posiciones y tamaños) de una forma en la que uno interfiere en el otro, ya que el tamaño de los círculos no permite determinar el verdadero valor en el eje correspondiente, ya que el tamaño hace que se vuelva un poco más difícil identificar el centro del círculo para luego encontrar la posición exacta. Con esto se pasaría a llevar el criterio de **separabilidad**.

Por otro lado, podemos analizar el **lie factor** y observar que su valor se aleja levemente del 1. Esto se debe a que la magnitud de los datos no se traduce de manera correcta con la codificación de ciertos atributos. Por ejemplo, las posiciones se basan en una escala de raíz cuadrada, esto hace que estar “el doble de lejos” no implica que el dato duplicó su valor. También se puede mencionar que él se utiliza el tamaño del círculo, pero no se especifica si se ocupa el área o diámetro del círculo. En ambos casos no se logra estimar correctamente la magnitud, y ante la ausencia de una leyenda, tampoco se puede apreciar si el cambio visual del tamaño refleja realmente el cambio de los datos. En conclusión, el uso de escala de raíz cuadrada ya nos aleja el **lie factor** del valor ideal, el 1, y la ausencia de la leyenda sobre el tamaño del círculo y especificar si se ocupa el área o diámetro, dificultan el proceso para calcular correctamente dicho valor **lie factor**.

Por último, la propuesta no cumple totalmente con el principio de **consistencia**, los colores de cada tipo de energía no coinciden con lo que uno está acostumbrado, por ejemplo, verde para nuclear, amarillo/naranja para el sol, coal (carbón) negro, etc. Por otro lado, la consistencia interna falla parcialmente al tener los nombres de cada tipo de fuente energética afuera y adentro de los círculos, en vez de hacer uno de los dos. Sin embargo, podemos decir que cumple parcialmente el principio de **autocontención**, ya que el contenido es autoexplicativo, hay leyendas que permiten entender ciertos atributos, títulos y ejes bien explicados, solo faltaría explicar si el tamaño de los círculos se mide en área o diámetro y agregar el rango de valores del tamaño.

Esta visualización presenta varias características positivas, sin embargo, se proponen las siguientes mejoras

- Ser consistente con la forma de nombrar cada energía. Las más pequeñas tienen el texto del mismo color que el punto, pero los más grandes ponen el texto adentro y en negro, buscaría un mismo patrón, en este caso, todos los textos afuera. Otra opción, sería agregar una leyenda que indique

relación de color con tipo de fuente energética.

- Especificar si *size* es área o diámetro de cada círculo para poder entender de mejor manera la magnitud que este canal representa, y poder establecer de mejor manera el ***lie factor***. También se recomienda incluir una leyenda para determinar rango de valores.