

Buenas prácticas sobre la visualización de datos

Dr. José Luis López
Aguirre



Introducción

*(La visualización consiste en **transformar información en imágenes que facilitan la extracción de significado**)*

Diseñar una visualización implica decidir qué codificaciones visuales se utilizarán para representar los datos.

La visualización de datos es la presentación de los resultados del trabajo de planeación, investigación, análisis y realización.

Según Victoria Pascual Cid, existen 6 reglas para desarrollar buenas visualizaciones.

Además, para Alberto Cairo existen 5 cualidades o prácticas para las buenas (y grandes) visualizaciones de datos.

Las 6 reglas para desarrollar buenas visualizaciones

Una propiedad visual preatentiva o preaplicada es aquella que se procesa en la memoria sin nuestra acción consciente; es decir, es **muy notoria a la vista y fácil de recordar.**

Cuando se desarrolla una visualización de datos es importante **decidir qué codificaciones se utilizarán para representar cada atributo** (categoría, variable, indicador) del conjunto de datos. A esto se le conoce como **mapeo visual**.

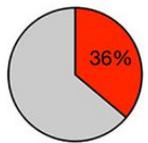
Las 6 reglas para desarrollar buenas visualizaciones

Las principales codificaciones visuales que llaman la atención en una visualización de datos son: **posición, forma, color y movimiento.**

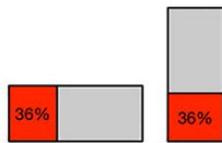
Colin Ware menciona que **la elección de estas codificaciones es el paso más importante para una visualización efectiva.**

Para saber cuáles escoger, es necesario seguir las seis reglas.

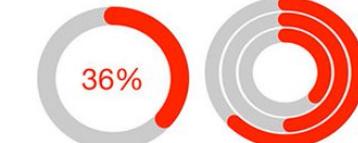
Visualizing Percentages & Parts of a Whole



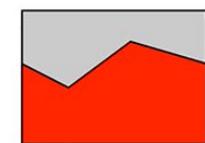
Pie Chart



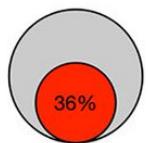
Stacked Bar/Column



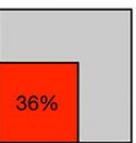
Donut/Circular Bar



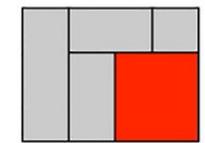
Stacked Area



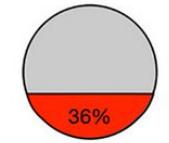
Proportional Area
(Nested Shapes)



Grid of Icons



Tree Map



Filled Shapes



Line of Icons



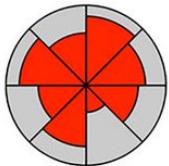
Half Circle



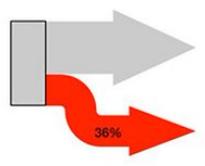
Arcs



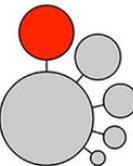
Gauges



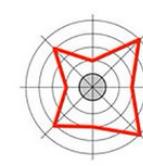
Rose Diagram
Polar Area Chart



Sankey



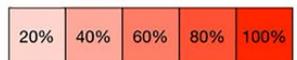
Bubbles



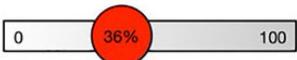
Circular Column



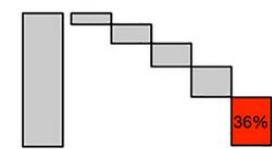
Radar



Color Scale/Heatmap



Sliding Scale



Waterfall

FORMA

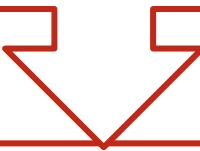
La **longitud** es el atributo que nuestro cerebro procesa más fácilmente. Por eso los gráficos de barras son tan populares.

1. Empezar con preguntas

En función del **objetivo informacional**, es decir, de lo que se quiere saber y lo que se quiere mostrar al público, **uno debe de escoger cuál puede ser la mejor forma de presentar los datos, de manera que sean comprensibles y fáciles de recordar.**

Es importante **preguntarse por la audiencia y su relación con el tema** a visualizar. **Cómo se busca ayudar al usuario**, el fin de la presentación, el tiempo disponible...

Es importante que tanto la visualización de datos como la investigación respondan las seis preguntas básicas: ¿Qué? ¿Quién? ¿Dónde?
¿Cuándo? **¿Por qué?** ¿Cómo?



En la mayoría de los casos, estas preguntas pueden ser las mismas que en la metodología de la investigación.

2. Utilizar gráficos autoexplicativos

Al diseñar nuestras visualizaciones es importante buscar **que el usuario sea capaz de entenderla**.

Para ello, se dispone de un **recurso elemental** que mejora la legibilidad de las visualizaciones: **el texto** dentro de los mismos gráficos.

Gracias al texto es más fácil entender la historia que se esconde tras los datos.

Moonville: active users over time

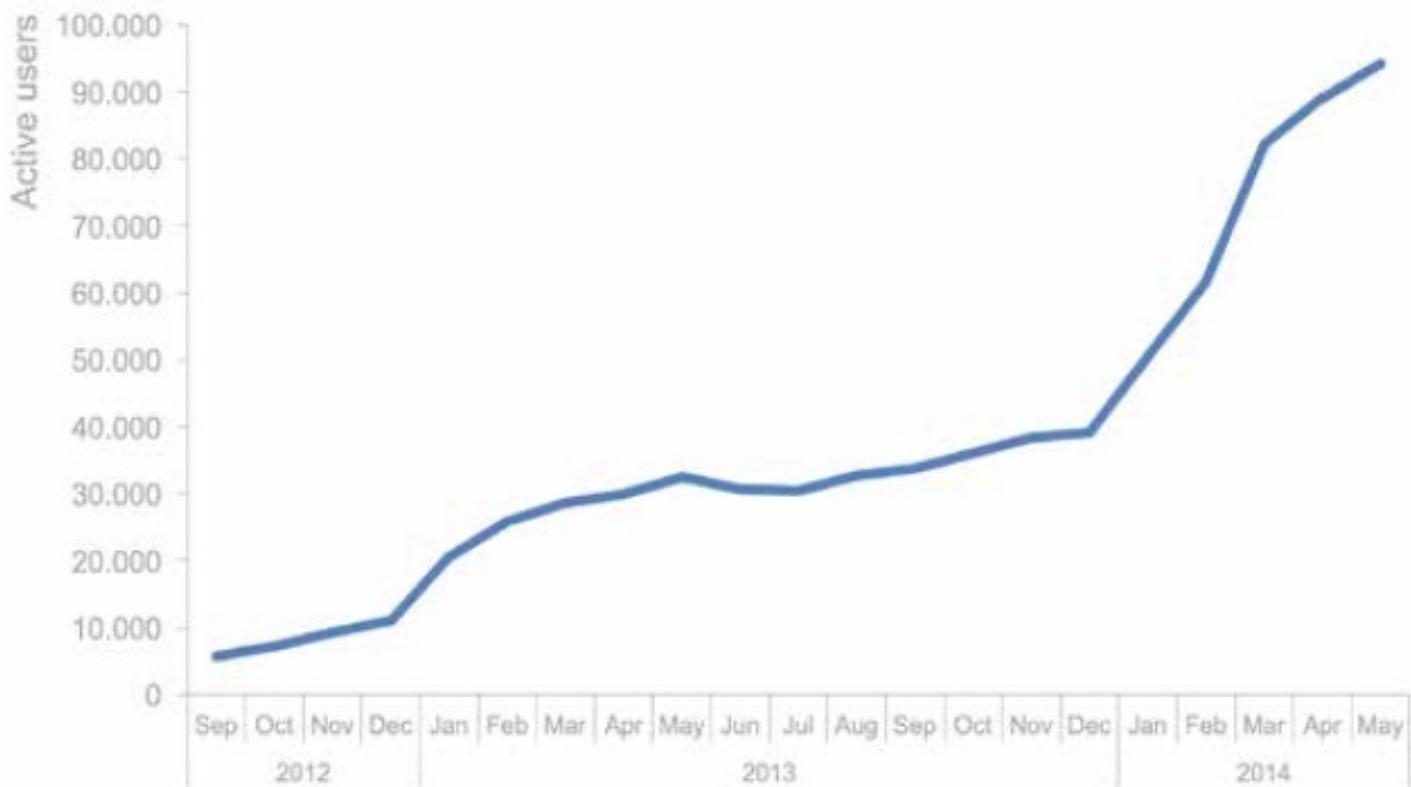


Figura 2.5. Visualización utilizada para mostrar la evolución de los usuarios activos de una aplicación móvil. Imagen cedida por Cole Nussbaumer.

Moonville: active users over time



Figura 2.8. Visualización final que resume todo el discurso pronunciado previamente. Imagen cedida por Cole Nussbaumer.



PUBLICACIONES TOTALES

VS.



PUBLICACIONES RELEVANTES

R. ANAYA



438

◦ 77

J.A. MEADE



277

◦ 109

AMLO



111

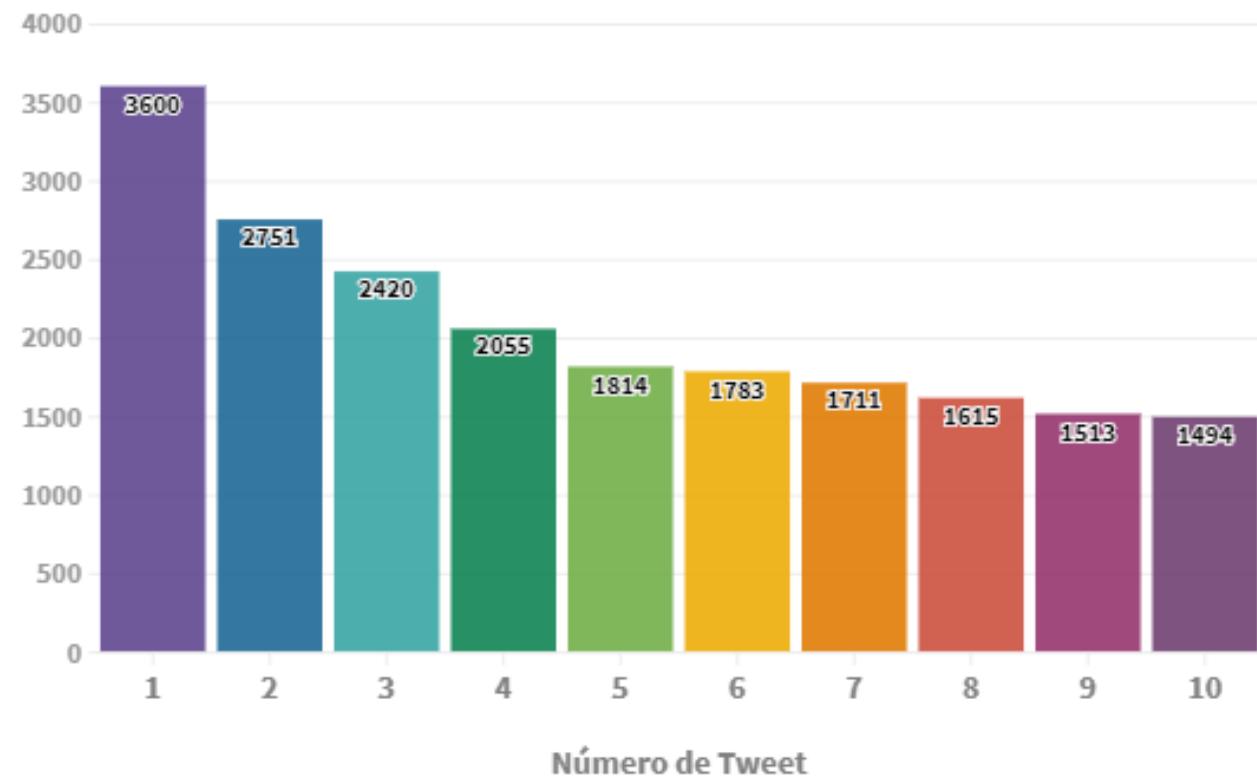
◦ 100

Infografía acerca de las **publicaciones totales** en comparación con las **publicaciones relevantes** realizadas por los candidatos presidenciales durante la campaña de 2018 en México.

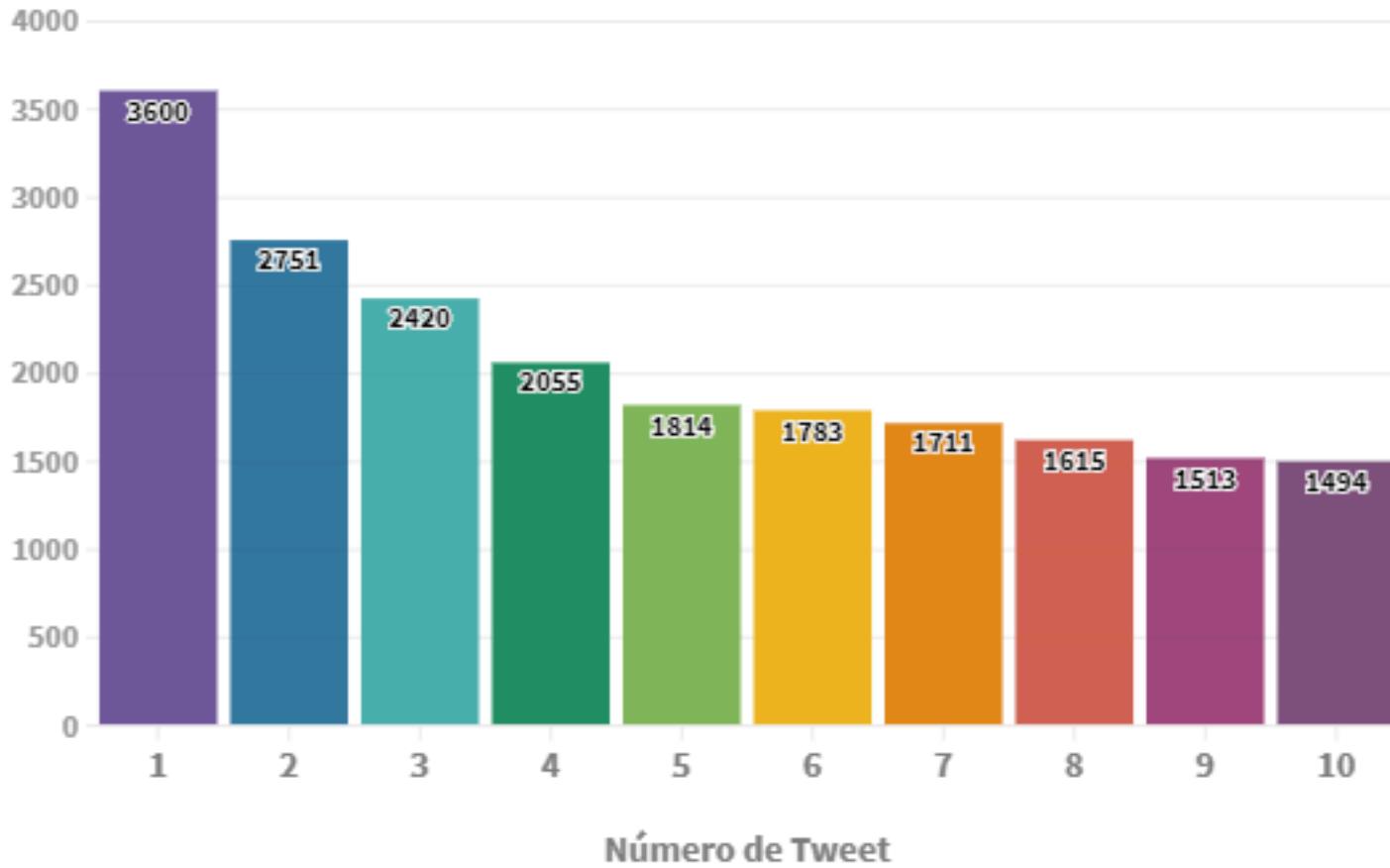


Fuente: Elaboración propia.

Fórmula de Interacción



Fórmula de Interacción



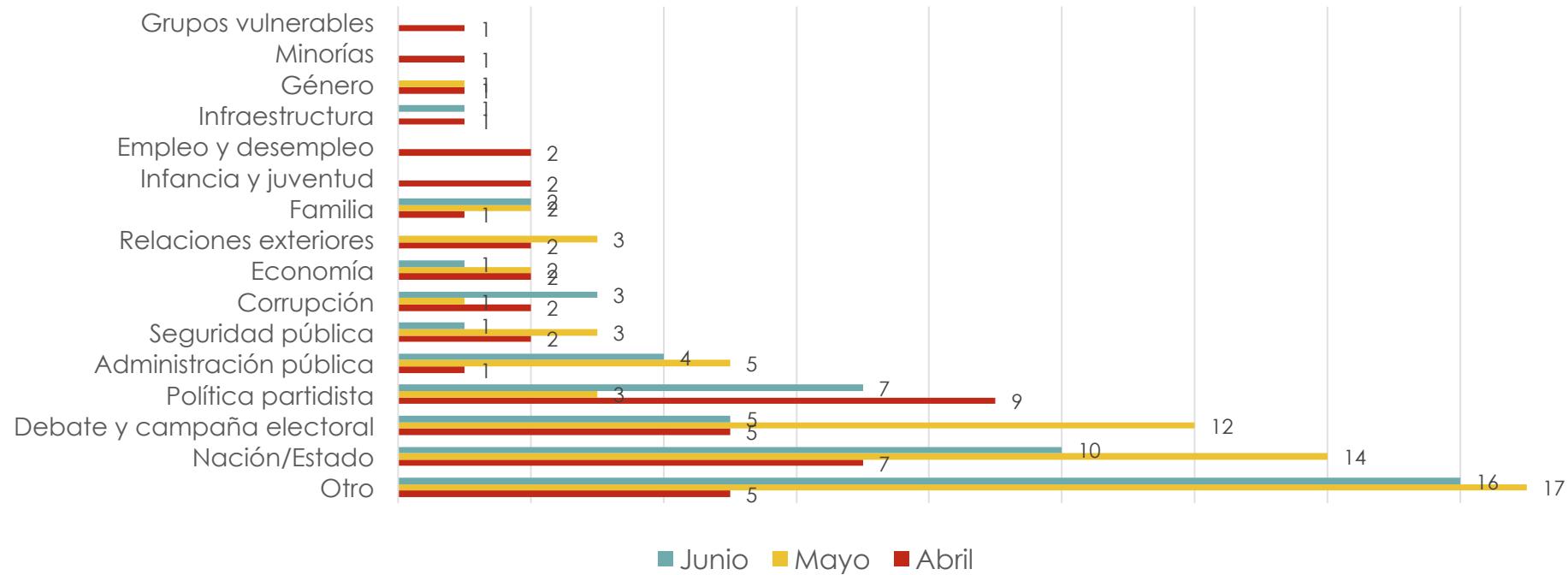
Fuente: Elaboración propia



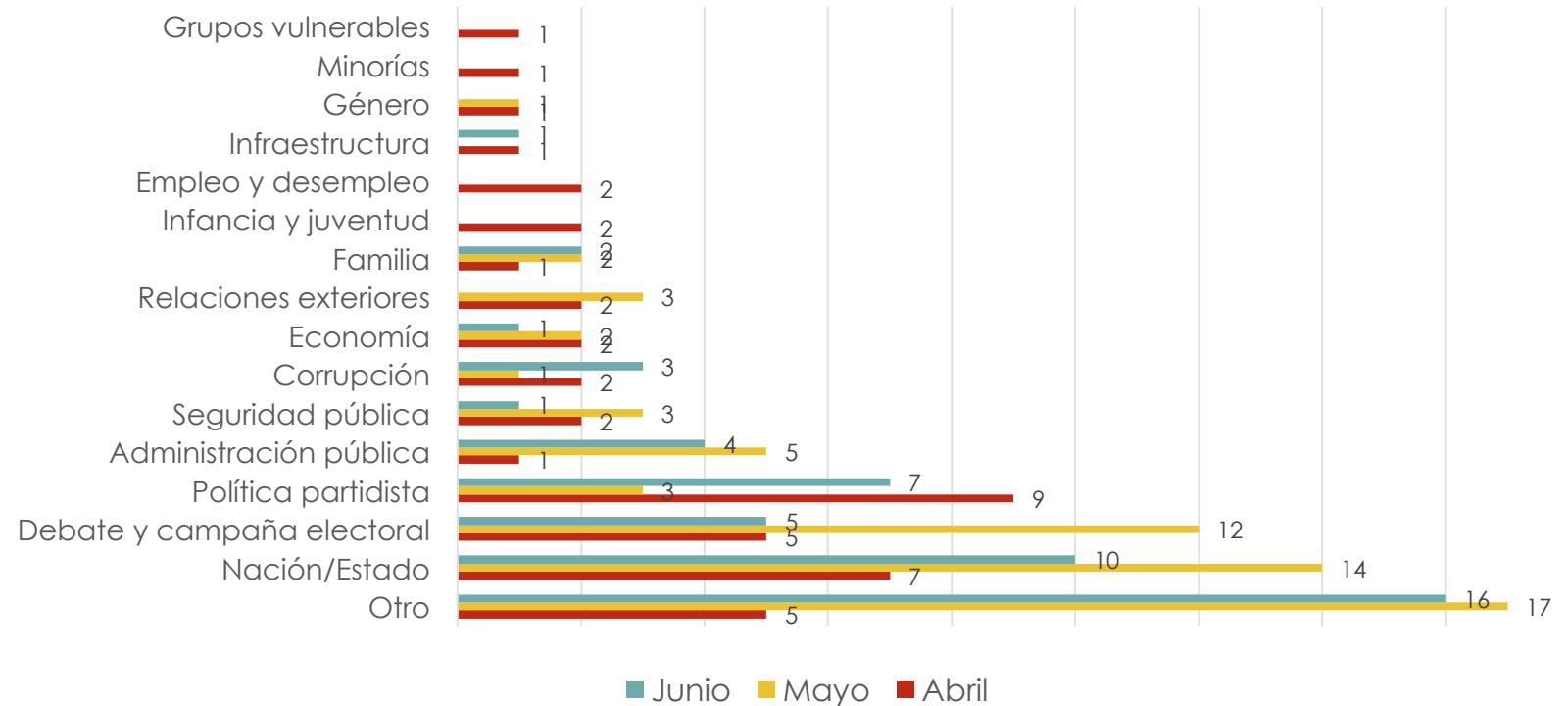
El Club América registró el grado más alto de interacción durante el periodo futbolístico del Clausura 2020 entre las escuadras analizadas.

“Se generan demasiados datos y abundan las malas interpretaciones de las gráficas”

Un gráfico simple puede convertirse en un **relato personalizado y emotivo**. Los datos son exactamente los mismos, pero **gracias al texto los situamos en contexto, entendemos qué hay detrás y qué significado tienen**. Como vemos, en una visualización, el texto puede ser incluso más importante que los propios datos numéricos.



Gráfica 1. Frecuencia de los tópicos en las publicaciones de @RicardoAnayaC que obtuvieron mayor grado de interacción en Twitter durante la campaña electoral. **Fuente:** Elaboración propia.



Gráfica 1. Frecuencia de los tópicos en las publicaciones de @RicardoAnayaC que obtuvieron mayor grado de interacción en Twitter durante la campaña electoral. **Fuente:** Elaboración propia.

- ▶ Los **tópicos más frecuentes** en los tweets **relevantes** fueron **Otro, Nación/Estado y Debate y campaña electoral**.
- ▶ La categoría Política partidista fue impulsada por las **publicaciones en las que se criticó a otros candidatos**, particularmente a Andrés Manuel López Obrador.

3. La funcionalidad sobre la estética

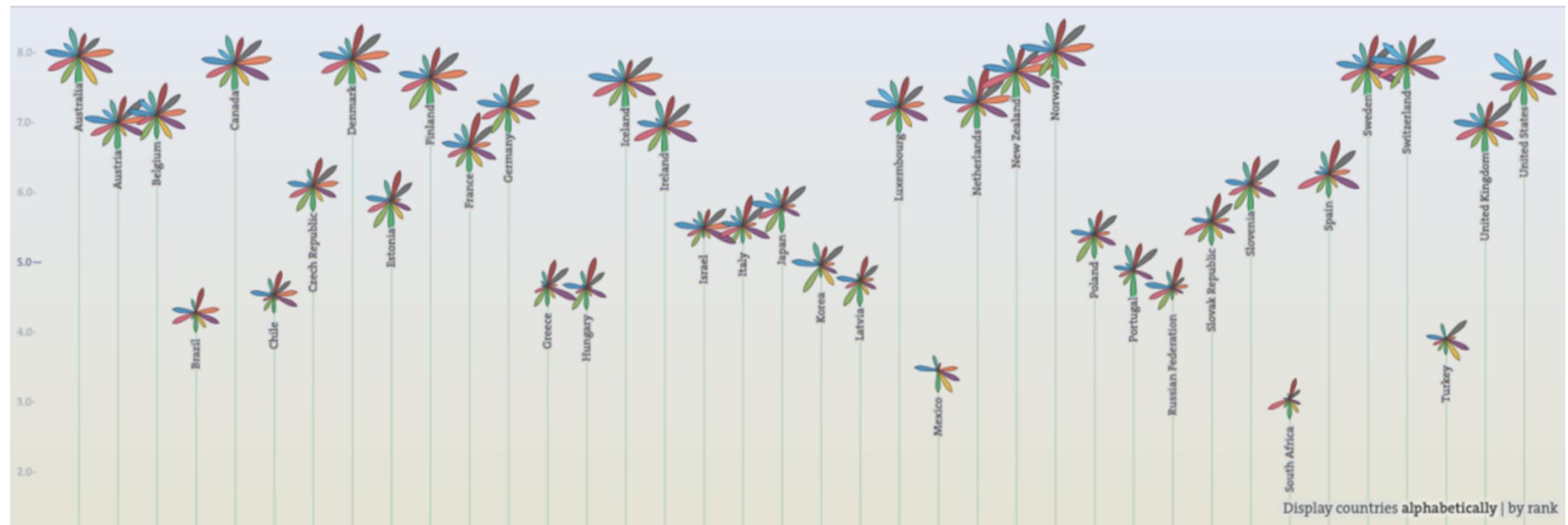
Es importante señalar que por más bonita que se vea la visualización, su objetivo principal es **generar conocimiento sobre los datos con los que se trabaja.**

No se debe caer en un punto en que la visualización se vea muy atractiva, pero que sea confuso interpretarla.

El uso de color de fondo y el tipo de visualización (estática o interactiva) también son aspectos que deben tenerse en cuenta.

Ejemplo

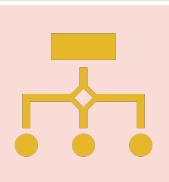
Figura 11. Flores representando barras en el «Better Life Index» de la OECD.



4. Uso de la interacción



Es importante saber qué tipo de gráfico se utilizará y qué uso se le dará, para saber cuál escoger **en función del objetivo de la visualización y el perfil de la audiencia (competencias del visualizador)**.



Los **gráficos estáticos** comunican un conjunto concreto de datos de una única manera, mientras que **los interactivos pueden permitir al usuario que decida en qué fijarse (personalización)**.

5. La forma sigue la necesidad

Siguiendo la regla anterior, es necesario escoger un tipo de visualización que muestre mejor los datos. Alberto Cairo menciona que mientras los buenos gráficos nos hacen más inteligentes, otros nos toman desprevenidos y mal equipados para interpretarlos.

“La forma escogida dependerá siempre del **objetivo informacional**”.

Comparaciones

Para comparar diferentes variables o categorías entre si.



Mapas

Para encontrar patrones geográficos.



Tendencias

Para entender la evolución temporal de variables cuantitativas.



Partes de un total

Para entender la contribución de diferentes categorías en un total.



Distribuciones

Para entender la forma y propiedades de una variable.



Correlaciones

Para entender la relación entre diferentes variables.



Conexiones, relaciones y redes

Para entender la relación entre los elementos de un conjunto de datos.



Ejemplo

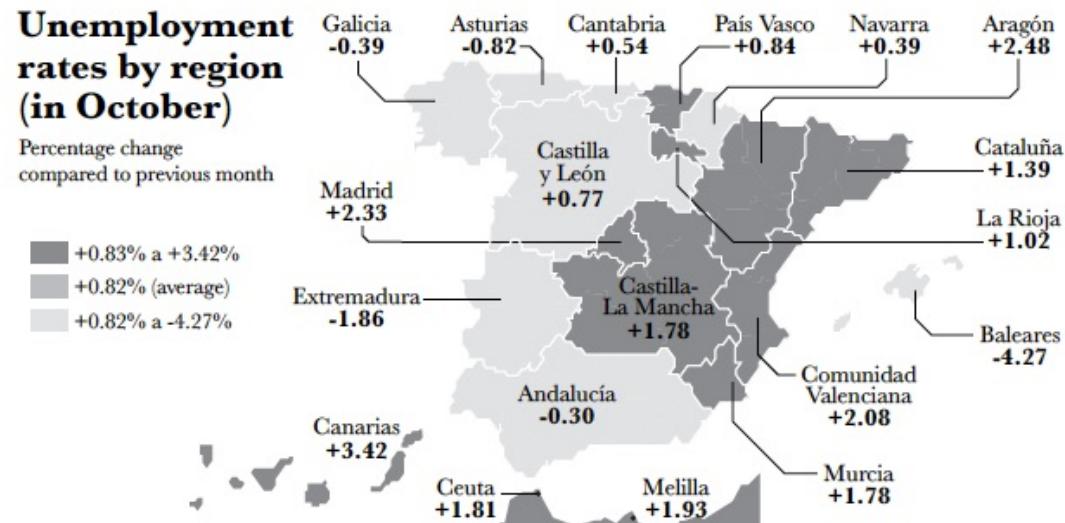


Figura 5.10. Representación de la tasa de paro por comunidad autónoma. Imagen cedida por Alberto Cairo, de su libro *El arte funcional*¹⁰.

5. La forma sigue la necesidad

Ejemplo

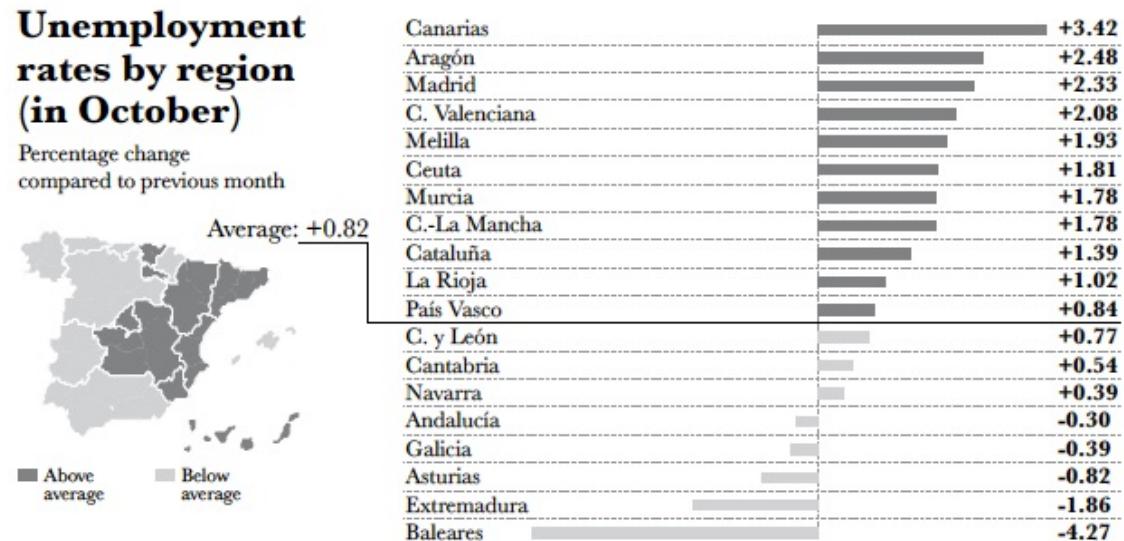


Figura 5.11. Rediseño planteado por Alberto Cairo. Imagen cedida por el mismo autor, de su libro *El arte funcional*.

5. La forma sigue la necesidad

6. Preparar bien los datos

No está de más recordar que antes de representar los datos se debe trabajar con ellos. Un ejemplo es el **proceso de normalización de bases de datos**, que consiste en migrar el almacén de datos a estructuras más pequeñas, estables y comprensibles, de manera que se puedan hacer cálculos con ellos.

Las 5 cualidades de las grandes visualizaciones de datos

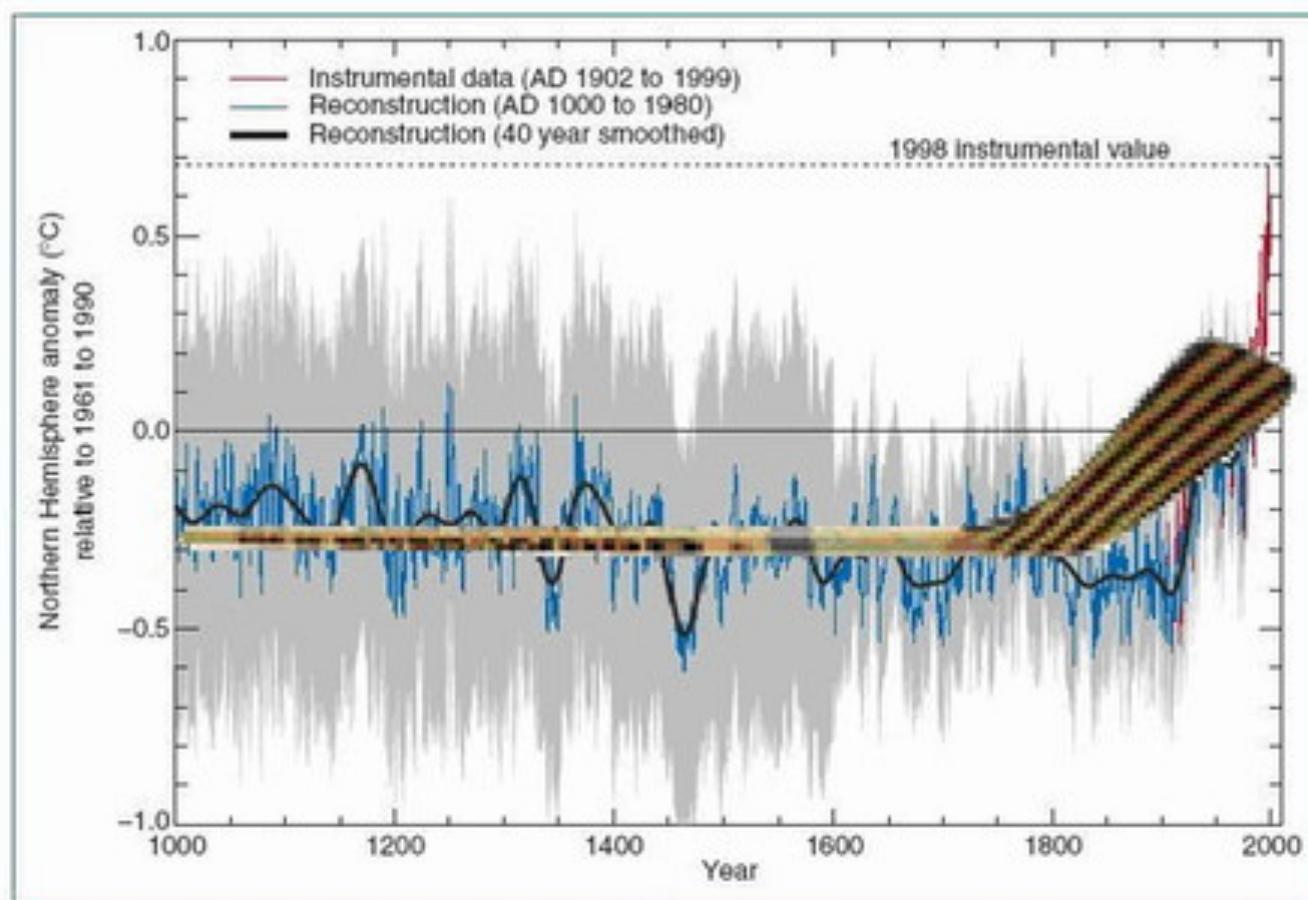
”
“

Alberto Cairo menciona en su libro The Truthful Art los principios de una buena visualización de datos.

Describe cinco cualidades que deberían ser la base de trabajo de una exitosa visualización de datos.



Explica que **una visualización se vuelve exitosa cuando cualquier persona le encuentra algo útil para su vida diaria o trabajo.**



Las cinco cualidades:
**veraz, funcional,
hermosa, perspicaz y
esclarecedora.**
Estas trabajan
relacionadas entre sí
e interactuando
constantemente.

1. Ser veraz

Existe la posibilidad de una **comunicación visual clara, precisa y efectiva**. Para Cairo, la verdad y la falsedad no son absolutas. Él recomienda:

- ▶ Evitar el autoengaño
- ▶ Ser honesto con la audiencia
- ▶ Prestar atención a las funciones y propósitos del gráfico

1. Ser veraz

Cairo **exhorta a los usuarios a solicitar los datos en los cuales se basó la visualización** y a luchar por el derecho de saber si la información es veraz y comprobable.

Los especialistas en marketing son conscientes de que la mayoría de las personas no prestan mucha atención a lo que ven y que los cuadros y mapas de datos pueden ser muy persuasivos, ya que se ven muy científicos. “Los datos los apantanallan, pero los engañan”.

2. Ser funcional

Un gráfico no se debe escoger por cuestiones de gusto, sino por pensamientos racionales, de acuerdo con las tareas que se desea habilitar. Es importante **ayudar al público a interpretar la información correctamente y que comprenda el propósito de la visualización.**

Para Cairo, el propósito de los gráficos es guiar la decisión de cómo dar forma a la información.

3. Ser hermosa

Una visualización hermosa es aquella que no se puede dejar de notar, que **obliga a uno a detenerse y contemplar su composición**, que puede ser considerada elegante y **apreciarse sin saber nada sobre lo que pretende representar**.

Para Cairo, el sentimiento de belleza está en **la forma en que percibimos la apariencia de un objeto y su profunda conexión con su propósito**; y consiste en una mezcla equilibrada de placer sensual e intelectual.

3. Ser hermosa

La belleza es una medida de la experiencia emocional de **asombro**, placer o mera sorpresa. Es importante porque las cosas atractivas y agradables funcionan mejor.

La belleza es subjetiva, por lo cual lo importante es que **las visualizaciones sean experimentadas (entendidas) por la mayor cantidad de personas posible**.

Hay que diferenciar entre diseñadores de infografías o visualización de datos y **decoradores de datos**. Estos últimos no **interpretan los datos antes de crear una visualización**.

4. Ser perspicaz

Las buenas visualizaciones **despejan el camino para hacer descubrimientos valiosos que serían inaccesibles si la información se presentara de una manera diferente**. Muestra datos relevantes en un formato visual que revela tendencias o relaciones.

5. Ser esclarecedora



Uno de los objetivos de una visualización de datos debe ser dar a las personas acceso a la información que necesitan para aumentar su bienestar. Es importante elegir temas de manera ética y prudente, así como arrojar luz sobre cuestiones relevantes.



Una visualización esclarecedora se da cuando se unen todas las características mencionadas anteriormente, además de la **iluminación**. Esta se refiere a que **cada componente importante se ha combinado con éxito para producir una nueva y valiosa revelación**.

Primeiro Plano

Diagrama

A NOTÍCIA EM PERSPECTIVA

Quem consegue um desenvolvimento limpo

Ao enriquecer, vários países emergentes aumentam sua poluição. Isso não é inevitável

Alberto Cairo, Juliana Elias e Marco Vergotti

DADOS RECENTES do Banco Mundial mostram que a China ultrapassou os Estados Unidos e é hoje o país mais poluidor. A China é líder em emissões de gás carbônico, principal responsável pelas mudanças climáticas. O fato é usado como argumento para que os grandes países emergentes, como o Brasil, a Índia e a Indonésia, assumam outras opções de redução nas emissões. A justificativa de gigantes como a China vem se baseando no tamanho de sua população. O argumento é que, com 1,3 bilhão de pessoas para sustentar, seria natural (e até justo) que a China emitisse mais gás carbônico. Palavras-chave: China, poluição. Só que os limites de emissão devem ser considerados por pessoa. Isto justifica, mas nem tanto. Os chineses já têm uma emissão per capita maior que a de algumas nações ricas, como a Suécia, o que levanta outra questão: os emergentes precisam reproduzir o caminho poluidor de nações como os EUA ou devem criar um novo caminho, mais limpo, como o dos países emergentes?

OS MAIS POLUIDORES EM 2009

A China é o país que mais aumenta as emissões



O aumento na renda nem sempre leva a mais poluição

Comparação entre evolução do PIB e das emissões per capita

1 EUA

Os americanos são grandes consumidores. Mas, nos últimos 30 anos, produziram menos riqueza sem aumentar suas emissões. E hoje em dia menos do que em 1980.

2 Rússia

As emissões russas descreveram com a crise da União Soviética, quando o colapso de algumas indústrias. Mas, com a recuperação, o país voltou a emitir mais para gerar riqueza nas últimas décadas.

3 Alemanha

O país gerou riqueza e também diminuiu suas emissões, um dos maiores usuários da energia elétrica. Tem 220 mil turbinas hidrelétricas em cata-ventos, ou quase duas Itaipus.

4 Suécia

Com política intensa de redução das emissões, a Suécia é uma das maiores índices da Europa. Na capital, Estocolmo, os ônibus elétricos são os principais de carro e ônibus.

5 China

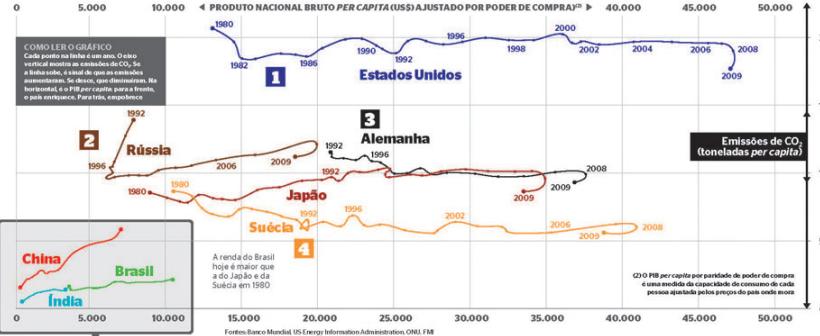
Embora seja o país que mais investe em energia elétrica, a China ainda fura 700 mil toneladas de carbono mineral, a fonte mais importante da poluição.

6 Índia

A Índia é o terceiro país que mais investiu em energia elétrica entre China e os EUA em 2010. Mas tem a quinta maior renda per capita, que pode extrair-lhe a custos altos. A fórmula é metade de energia do país.

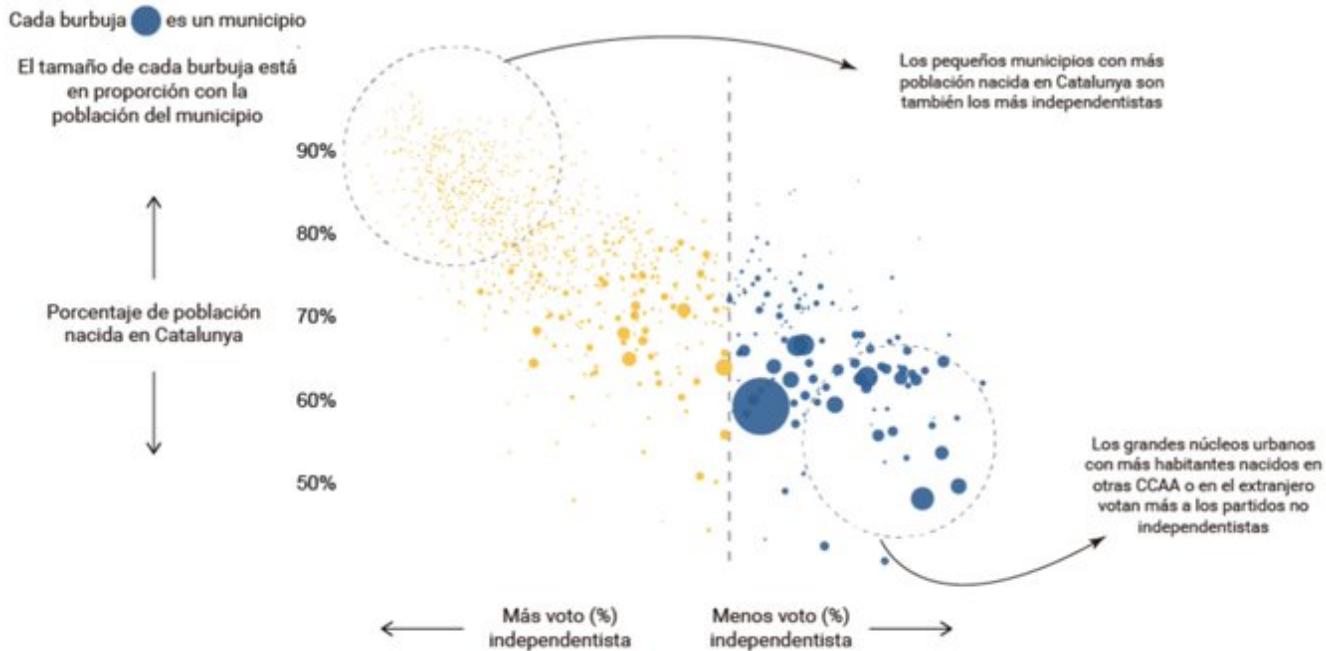
7 Brasil

O Brasil vem se desenvolvendo com uma energia hidrelétrica limpa e substituindo gasolina por álcool, que é uma menor contribuição para o carbono da atmosfera. Os dados sobre desmatamento e florestamento, responsável por 70% de nossas emissões – e por pouco progresso econômico



Cómo leer este gráfico

Comparativa del porcentaje de población nacida en Catalunya de cada municipio con el voto a cada bloque el 21D



Corolario

- ▶ Las visualizaciones de datos tienen como **prioridad ayudar a generar conocimiento en torno a los datos** y no solamente a la generación de imágenes atractivas.
- ▶ Es **nuestra responsabilidad poner la información correcta frente a la audiencia correcta**, de manera que se adapte a su conjunto de habilidades, expectativas y propósitos.
- ▶ Un analista o diseñador de datos **debe escoger aquella visualización que mejor ayude a cumplir el objetivo informacional**.
- ▶ La **ética** debe permear todo el proceso de realización de nuestras visualizaciones de datos.

Corolario

- ▶ Dos competencias básicas: la **capacidad de síntesis** y la **capacidad estética**; y ambas deben combinarse correctamente.
- ▶ “La visualización ideal no solo debe **comunicar con claridad**, sino que también debe **estimular la participación y la atención del espectador**”. Fernanda Viégas y Martin Wattenberg, líderes del Big Picture Data Visualization Group de Google.
- ▶ Un visualizador de datos debe **comprender, interpretar, aclarar, sintetizar, explicar, transmitir contenidos y ponerlos en contexto para que sean relevantes**.
- ▶ Para tener éxito, **un gráfico debe ser muy simple** (que no exista impedimento entre el usuario y la comprensión del dato) **y tener** un poco de movimiento, **de chispa**, que llame la atención.