

Tendencias en la disciplina de la Visualización de Información

Trends in the discipline of Information Visualization

Mario Pérez-Montoro

Pérez-Montoro, Mario (2022). "Tendencias en la disciplina de la Visualización de Información". *Anuario ThinkEPI*, v. 16, e16a02.

<https://doi.org/10.3145/thinkepi.2022.e16a02>

Publicado en *IweTel* el 3 de febrero de 2022

Mario Pérez-Montoro

<https://orcid.org/0000-0003-2426-8119>

Universitat de Barcelona

Facultat d'Informació i Mitjans Audiovisuais

Melcior de Palau, 140. 08014 Barcelona

perez-montoro@ub.edu



Resumen: Panorámica sobre la evolución de la disciplina de la visualización de la información durante 2021. Por un lado, se presentan propuestas que han destacado por incorporar innovaciones respecto a los conceptos y recursos visuales. Por otro, se recogen algunos ejemplos de visualizaciones que han llamado la atención, no tanto por sus aportaciones visuales, sino por el conjunto de datos y contenidos que han representado. Por último, para completar esa visión de la evolución anual de la disciplina, se analizan algunos avances relacionados con el desarrollo tecnológico que se ha producido en el campo de la visualización de información.

Palabras clave: Visualización de la información; 2022; Comunicación visual; Tendencias; Visualización de datos; Visualización interactiva; infografía; Visualizaciones multimedia; Tecnología de análisis de datos; Recursos visuales.

Abstract: Evolution overview of the discipline of information visualization during 2021. We select, on the one hand, various proposals that incorporate innovations regarding visual concepts and resources, and on the other, some examples of visualizations that have attracted attention, not so much for their visual contributions but for the set of data and content they have represented. Finally, to complete this vision of the evolution of this discipline over one year, we address some advances related to the technological development in the field of information visualization.

Keywords: Information visualization; 2022; Visual communication; Trends; Data visualization; Interactive visualization; Infographics; Multimedia visualization; Data analysis technology; Visual resources.

1. Introducción

En nota para la edición del año pasado del *Anuario ThinkEPI* concentramos nuestros esfuerzos en intentar recoger algunas de las propuestas de comunicación visual relacionadas con la pandemia provocada por el SARS-CoV-2.

Este año, aunque continúe el azote de la pandemia, para intentar abrir el foco en el análisis hemos querido recuperar nuestro esquema original llevado a cabo en otras ediciones y abordar, ya sin restricciones temáticas, algunas de las propuestas visuales desarrolladas durante el pasado año y que destacaron por sus planteamientos innovadores y su contribución a la evolución de la propia disciplina de la visualización de la información.

Por tanto, el objetivo de este trabajo es revisar algunas de esas visualizaciones innovadoras realizadas a lo largo de todo 2021. Para abordarlas, recuperaremos la clasificación defendida en trabajos anteriores. Así, por un lado, presentaremos una selección de aquellas que han destacado por incorporar novedades conceptuales y nuevas propuestas de visualización. A continuación, mostraremos otras que han llamado la atención por la importancia del conjunto de datos que han visualizado. Por último, acabaremos este trabajo dando cuenta de algunos de los avances y propuestas tecnológicas que se han llevado a cabo dentro de la disciplina a lo largo de este año.

2. Conceptos

Comencemos presentando nuestra selección con una serie de propuestas que han destacado a lo largo de ese año por incorporar innovaciones respecto a los conceptos y recursos visuales y que, al hacerlo, han permitido representar nuevos tipos de relaciones y contenidos semánticos. Concretamente, la selección recoge propuestas de visualización que incluyen el uso gráfico de las tres dimensiones, alternativas para la representación de la producción y la colaboración científica, la utilización de minigráficas y una propuesta para la representación de las relaciones topológicas entre conjuntos.

La primera de las propuestas visuales que queremos destacar tiene que ver con el uso gráfico de las 3D (tres dimensiones). Tradicionalmente, y en términos generales, desde las propuestas conceptuales de la disciplina de la Visualización de Información se ha desaconsejado el uso de las 3D para representar contenidos semánticos. Las razones principales que justifican su no utilización son el peligro del solapamiento visual de los elementos gráficos y las dificultades de comprensión que pueden generar en sus usuarios.

Sin embargo, en algunas propuestas visuales, su uso no sólo queda justificado, sino que permite comunicar la información de una forma más eficiente. Y es justo esto lo que nos plantean Raúl Sánchez y Analía Plaza en su trabajo “España vive en pisos: por qué hemos construido nuestras ciudades en vertical”, publicado por el periódico español *elDiario.es*. En esta pieza periodística (basada en navegación con *parallax scrolling* (navegación mediante el uso del cursor)), los autores representan los datos sobre la vivienda y la densidad de ocupación no sólo en un mapa en dos dimensiones, sino que introducen en el mapa una tercera dimensión: la altura del edificio (figura 1). De esta manera podemos ver cómo de poblado se encuentra un territorio, pero también si esa población está dispuesta de forma horizontal (con poca densidad, en edificios de poca altura) o de forma vertical (con mucha densidad, en comunidades dentro de edificios de muchas plantas).



Figura 1. Mapa de densidad de población representado en 3D
<https://especiales.eldiario.es/espana-vive-en-pisos>

La segunda de las propuestas visuales que queremos analizar está centrada en el tema de la producción científica; concretamente sobre los trabajos indizados en *PubMed* (base de datos sobre literatura biomédica) que tratan como tema principal la Covid-19. En esta línea, el diseñador de la información Jeff MacInnes publicó la pieza “Following the science en the pudding” con la intención de mostrar que el desarrollo de una vacuna en un tiempo récord no es tarea de una sola persona, sino que es el fruto de la articulación de la producción y hallazgos de muchas instituciones de investigación y de los científicos que las conforman.

En esta pieza interactiva (también basada en navegación con *parallax scrolling*) se incorporan una serie de visualizaciones interesantes, creadas utilizando *React* con una combinación de *three.js*, *d3.js* y *p5.js* a partir de los datos tratados con bibliotecas de *Python* y cuadernos de *Jupyter*.

Así, por ejemplo, muestra con una figura dinámica cómo ha cambiado drásticamente la densidad de trabajos publicados sobre Covid-19 en 2019 (casi nula) frente a 2020 (mucho más alta). En esa figura, a la izquierda, circulan de abajo a arriba, los artículos científicos producidos a lo largo de 2019; y, a la derecha, los publicados en 2020 (figura 2). Y mediante la introducción del coloreado, destacan los que han versado sobre el tema del coronavirus, evidenciando cómo ha variado la distribución.

Pero también nos muestra la colaboración entre científicos. Mediante un mapa dinámico que evoluciona con el tiempo y uniendo mediante arcos los científicos que colaboran en sus investigaciones, se muestra la geografía de esas alianzas y los clústeres (Estados Unidos y Europa, principalmente) que concentran una parte importante de esos científicos (figura 3).



Figura 2. Densidad de trabajos publicados sobre en Covid-19 en 2019 frente a 2020
<https://pudding.cool/2021/03/covid-science>



Figura 3. Mapa dinámico de la colaboración científica
<https://pudding.cool/2021/03/covid-science>

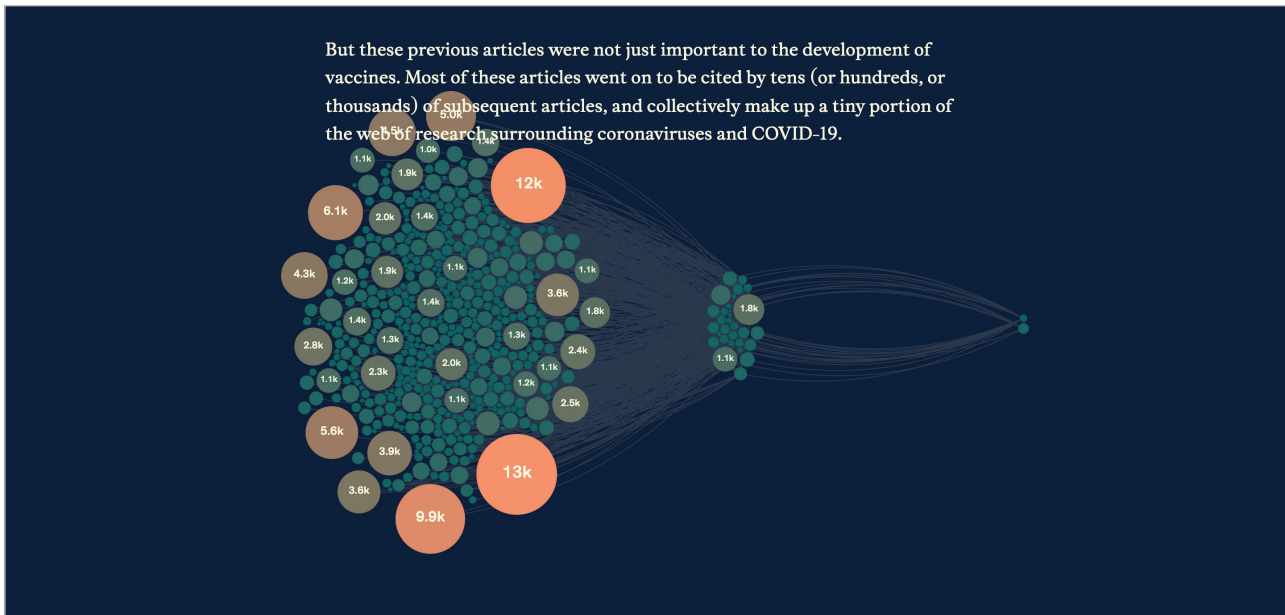


Figura 4. Cadena de citas tras los artículos donde se presentan las vacunas
<https://pudding.cool/2021/03/covid-science>

Y la última propuesta visual que queremos destacar de esta pieza es la que representa la cadena de citas que desemboca en los dos artículos (**Lindsey et al.**, 2021; **Polack et al.**, 2020) en los que se presentan respectivamente los resultados y el éxito de la tercera fase del estudio para el desarrollo de las vacunas de *Pfizer* y *Moderna* (figura 4). En el extremo de la derecha se sitúan los dos artículos referidos. Hacia la izquierda, la cadena de citas que incluyen esas dos piezas. Cada uno de los puntos muestra el trabajo científico publicado que representa y nos ofrece sus datos al pasar el ratón por encima de la figura circular.

La tercera propuesta visual que queremos destacar se centra en el uso de los que de forma técnica se han denominado “minigráficas” (*small multiple, trellis chart, lattice chart, grid chart o panel chart*, en inglés). Popularizadas por **Tufte** (1990), una minigráfica es una propuesta visual formada por una serie de pequeños gráficos similares que utilizan la misma escala y los mismos ejes, lo que permite compararlos fácilmente. Se utiliza para mostrar diferentes patrones dentro de un (o varios) conjunto de datos.

Normalmente, estos gráficos se integran en alguna de las celdas de una tabla o se ubican en una composición de parrilla (con filas y columnas) para facilitar su análisis visual. Pero existen otros usos más creativos de esta representación de información.

Por ejemplo, Nathaniel Rakich y Jasmine Mithani publicaron el trabajo “What absentee voting looked like in all 50 states” en el medio digital *FiveThirtyEight*. En él intenta analizar el comportamiento de los votantes en las elecciones presidenciales de Estados Unidos de 2020. Y para llevar a cabo ese análisis y mostrar la expansión masiva de la votación por correo, utilizan una propuesta visual creativa que incorpora unas minigráficas integradas en un mapa de símbolos graduados (figura 5).

Habitualmente, un mapa de símbolos graduados (*graduated symbol map*, en inglés) es un tipo especial de mapa en el que, para representar los valores cuantitativos asociados a las regiones físicas codificadas que lo conforman, se insertan pequeños símbolos (círculos, columnas, barras o gráficos circulares, por ejemplo) dotados de ciertos atributos de codificación como el tamaño del símbolo, su forma o su color, entre otros. Pero la novedad en este caso es que los autores no incorporan esas representaciones más clásicas, sino que nos presentan unas pequeñas gráficas de líneas (unas gráficas de pendiente, en un sentido más técnico) ubicadas geográficamente para mostrar el cambio porcentual del voto electrónico en cada estado entre las dos últimas elecciones presidenciales. Los ejes de todas las gráficas son idénticos. La inclinación de la línea indica si ha habido aumento o disminución en el porcentaje. Los colores de los puntos y las líneas nos indican si en ese estado se requirió algún criterio

Se recogen propuestas de visualización que han representado datos sobre el tema de la mentira y su comunicación, sobre el cambio climático, sobre la vertiente más luctuosa de la pandemia del coronavirus y sobre el mal diseño de las gráficas para fundamentar una campaña publicitaria engañosa

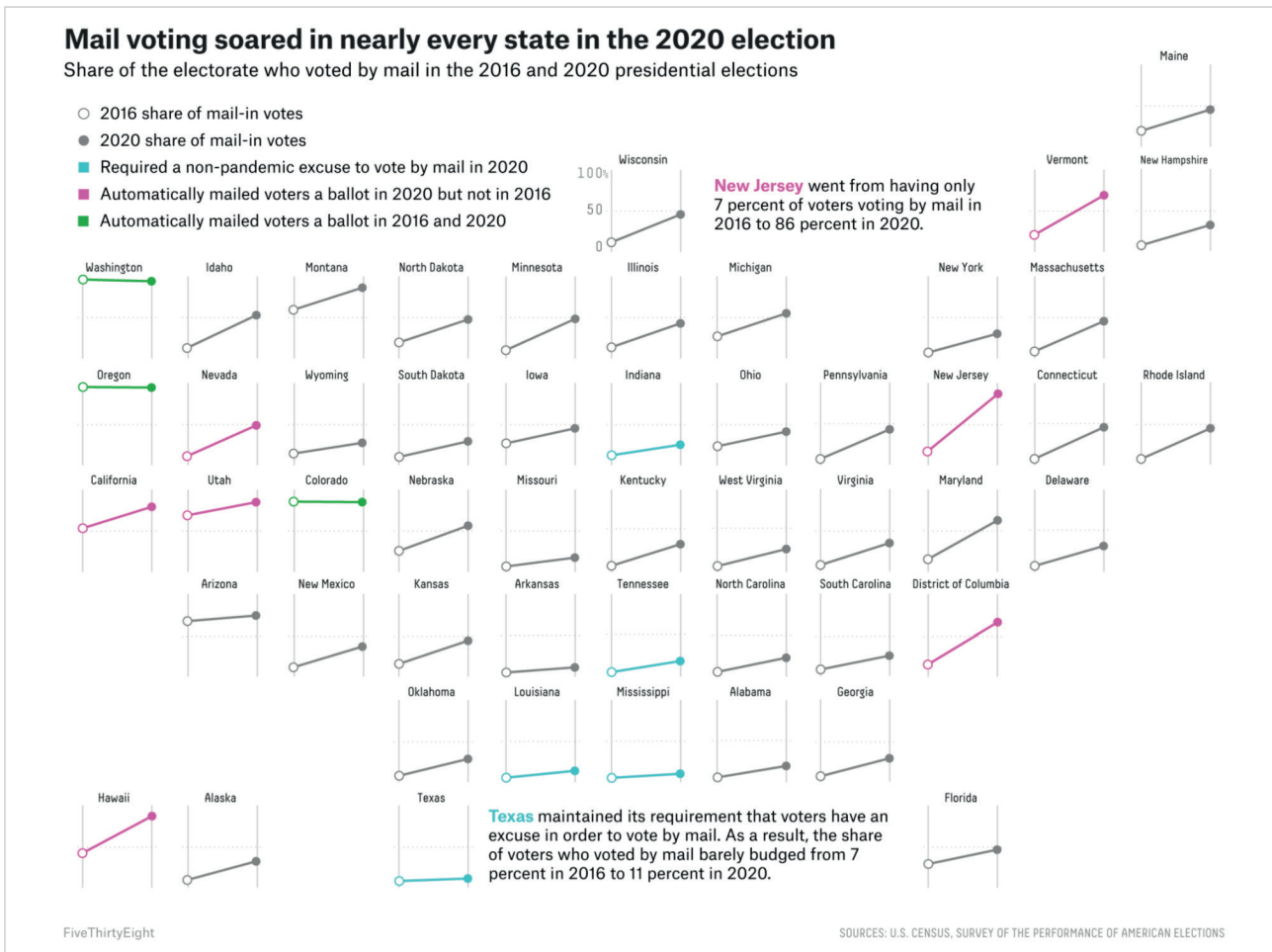


Figura 5. Aumento del voto por correo en las elecciones norteamericanas de 2020 frente a las de 2016 <https://fivethirtyeight.com/features/what-absentee-voting-looked-like-in-all-50-states>

no relacionado con la pandemia para poder solicitar el voto, si se produjo el envío automático de las papeletas a los votantes en 2020 pero no en 2016 o si el envío se realizó en 2016 pero no en 2020.

La última de las propuestas de visualización que queremos presentar en este apartado intenta codificar gráficamente, dentro de la teoría de conjuntos, el espacio topológico (unión, intersección o disyunción) que se puede producir cuando se analizan simultáneamente una serie de conjuntos.

Tradicionalmente, para analizar visualmente ese espacio lógico o topológico, se han utilizado los diagramas de Venn. Esos diagramas funcionan relativamente bien cuando se comparan visualmente sólo dos o tres conjuntos. En esos casos es relativamente cómodo identificar las zonas de unión, intersección y disyunción entre esos conjuntos. Pero cuando aumentamos el número de conjuntos puestos en relación, esa propuesta visual no funciona como estrategia de representación.

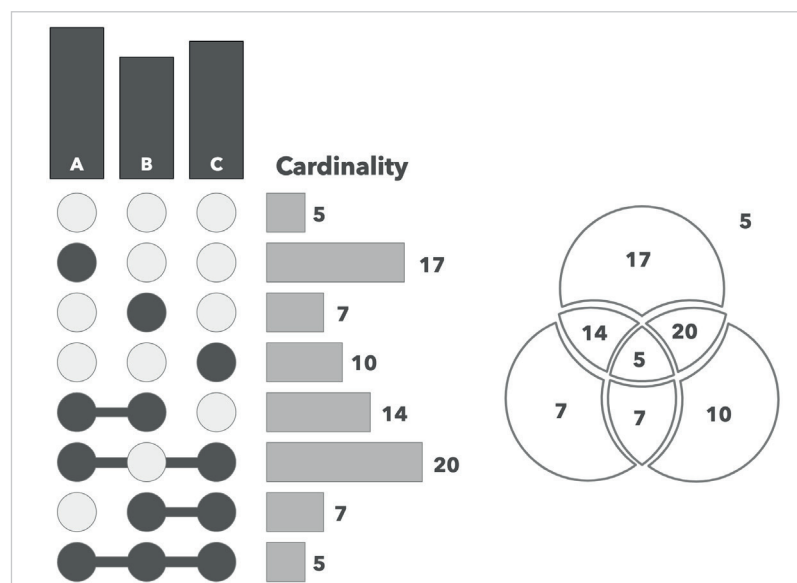


Figura 6. Explicación de la representación gráfica del espacio topológico en el análisis simultáneo de una serie de conjuntos <https://upset.app/#upset-vs-venn-diagrams>

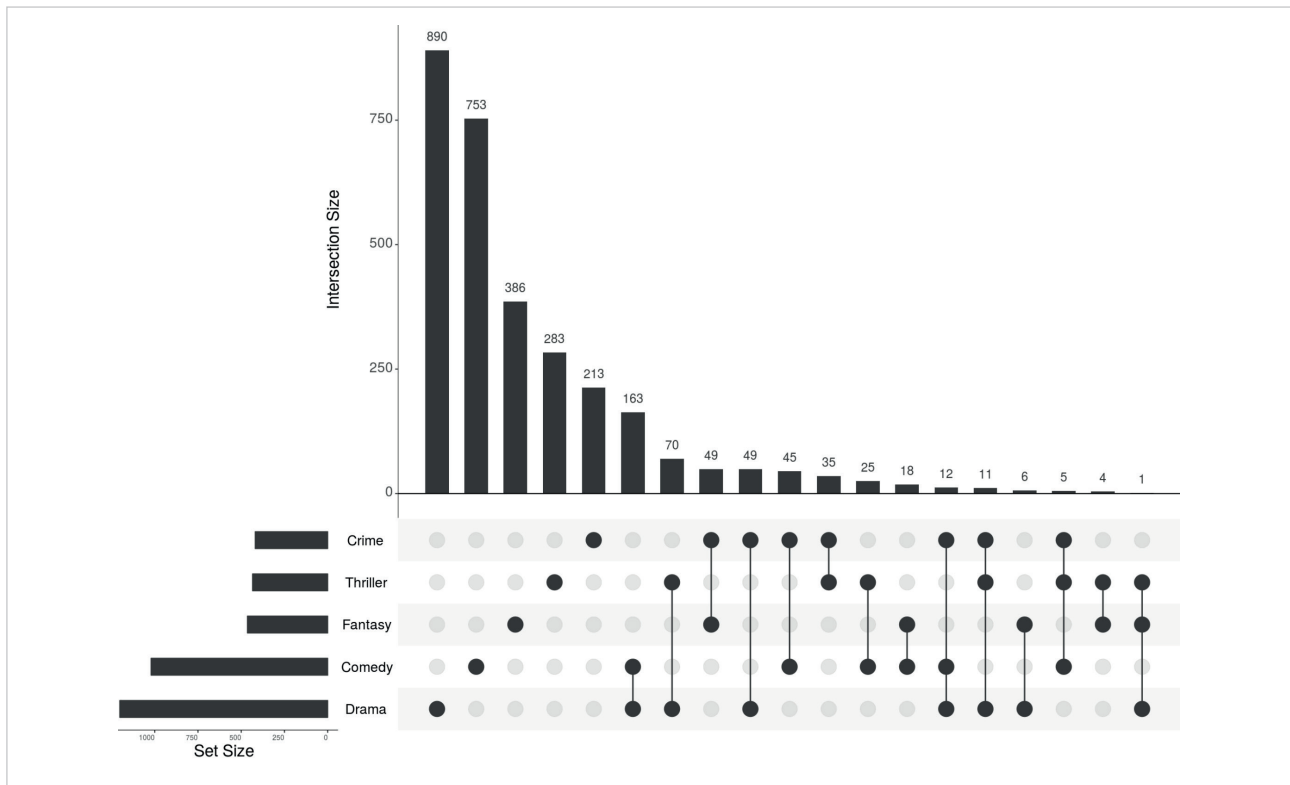


Figura 7. Representación gráfica del espacio lógico algebraico entre los conjuntos de películas clasificadas por género <https://upset.app/#upset-vs-venn-diagrams>

Para solventar ese problema, nace *UpSet*. Surgido originalmente en Harvard del trabajo colaborativo de los investigadores Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot y Hanspeter Pfister, esta aplicación se consolida ahora como una propuesta innovadora para poder representar, comprender y analizar las relaciones e interacciones que se producen entre más (sin límite numérico) de tres o cuatro conjuntos.

La propuesta, como muestra la figura 6 y frente a un diagrama de Venn convencional, representa los tamaños de los conjuntos analizados (cosa que no permiten los diagramas de Venn clásicos), las posibles relaciones topológicas (uniones, intersecciones y disyunciones) booleanas que se producen entre esos conjuntos y el tamaño de éstas (cosa que tampoco codifican los diagramas de Venn), independientemente del número de conjuntos analizados. En la figura 7 se aplica al caso de una serie de películas clasificadas por género cinematográfico en cinco conjuntos no disjuntos.

3. Contenidos

En este segundo grupo de productos seleccionados vamos a recoger algunos ejemplos de visualizaciones que han llamado la atención, no tanto por sus innovaciones conceptuales y de codificación, sino por el conjunto de datos y contenidos que han representado. Concretamente, nuestra selección recoge propuestas de visualización que han representado datos sobre el tema de la mentira y su comunicación, sobre el cambio climático, sobre la vertiente más luctuosa de la pandemia del coronavirus y sobre el mal diseño de las gráficas para fundamentar una campaña publicitaria engañosa.

Comencemos destacando un trabajo relacionado con el tema de la mentira y de su comunicación intencionada para conseguir unos fines determinados. A principios de 2021, los periodistas Glenn Kessler, Meg Kelly, Salvador Rizzo, Leslie Shapiro y Leo Dominguez publicaron en el periódico *The Washington post* el trabajo "A term of untruths. The longer Trump was president, the more frequently he made false or misleading claims", en el que analizaron las mentiras (30.573) pronunciadas por el presidente estadounidense Trump durante su período en la Casa Blanca (desde enero de 2017 a enero 2021). Concretamente, aderezada también con *parallax scrolling*, proponen una línea del tiempo donde recogen mediante una gráfica de barras las mentiras (o afirmaciones no ciertas) proferidas diariamente, recontándolas, clasificándolas temáticamente según su contenido (salud, Rusia, economía, comercio, inmigración, Ucrania o coronavirus, entre otros temas) y destacándolas dentro de esa clasificación mediante el uso

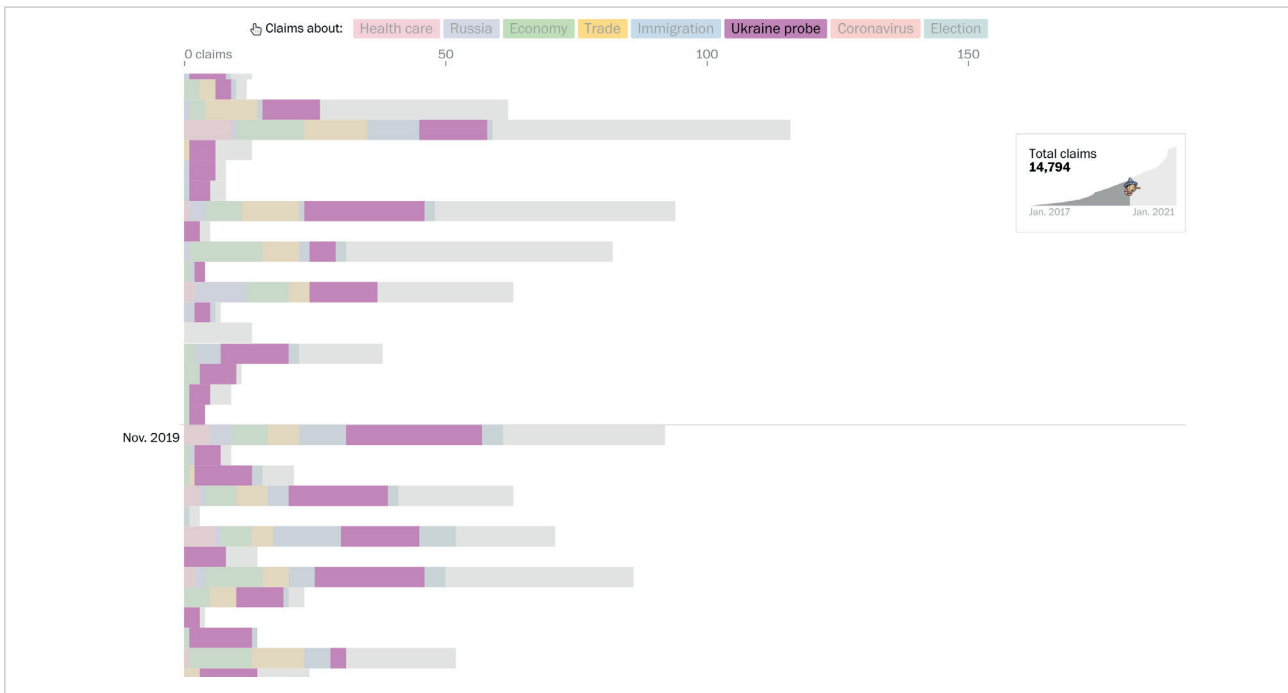


Figura 8. Línea del tiempo donde recogen mediante una gráfica de barras las mentiras proferidas diariamente por Donald Trump

<https://www.washingtonpost.com/politics/interactive/2021/timeline-trump-claims-as-president/>

de un código de color (figura 8). Es curioso observar cómo a medida que avanza su mandato, el presidente profiere un número mayor de mentiras al cabo del día, superando las cien en muchas de esas unidades temporales y llegando a profiere más de 500 en la jornada del 2 de noviembre de 2020. La base de datos de todas esas afirmaciones no ciertas se puede obtener en la propia web del periódico (*The Washington post*, 2021).

La segunda pieza que queremos destacar en este apartado versa sobre la comunicación de la emergencia climática. Concretamente la que nos propone *The Intergovernmental Panel on Climate Change (IPCC)*, organismo dependiente de *Naciones Unidas* y que persigue como objetivo la evaluación de la ciencia relacionada con el cambio climático.

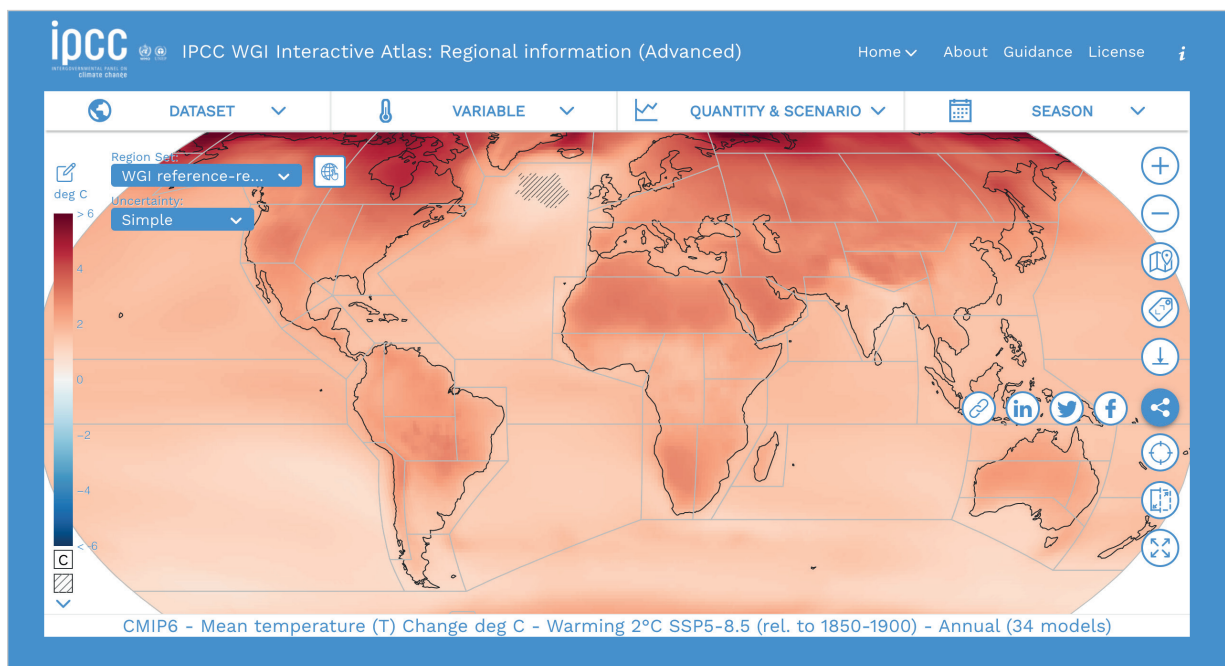


Figura 9. Atlas interactivo sobre el cambio climático propuesto por *The Intergovernmental Panel on Climate Change (IPCC)*
<https://interactive-atlas.ipcc.ch/regional-information>

Este organismo ofrece un atlas interactivo dotado de muchas funciones y que permite proyectar cómo sería el clima sobre la faz de nuestro planeta si cambiaran ciertas variables implicadas (figura 9). Entre esas funciones, podemos seleccionar los diferentes conjuntos de datos incluidos y sus diferentes visualizaciones y estadísticas de resumen. También permite descargar y compartir los datos mostrados y elegir las variables atmosféricas, oceánicas y socioeconómicas que queramos visualizar. Habilita la posibilidad de seleccionar el período de tiempo implicado (pasado reciente para la línea de base histórica y período de tiempo futuro o el nivel de calentamiento para la información de proyección) y mostrar los valores absolutos o los cambios relativos a esa línea histórica.

Existe la opción de poder seleccionar la época del año que queremos mostrar. Y, además de mostrar la media anual (la predeterminada), se puede elegir una temporada preestablecida o un período específico de meses. También se pueden elegir diferentes conjuntos de regiones geográficas sobre las que agregar datos y activar la opción de mostrar información sobre la importancia de las tendencias observadas o información sobre la incertidumbre de las proyecciones. Por último, el atlas ofrece una barra lateral de navegación para explorar otras posibilidades, entre las que se incluye acercar y alejar en el mapa, cambiar su proyección o explorar los metadatos para ver cómo se procesaron los datos.

La siguiente propuesta visual que queremos destacar está relacionada con la pandemia, con su vertiente más luctuosa, con los fallecimientos que ha provocado. Como ya vimos en nuestra contribución del año pasado (Pérez-Montoro, 2021), *The New York Times* publicó en su portada del 24 de mayo de 2020 la selección de los nombres de un millar de personas como representación de todos los fallecidos por coronavirus en el país norteamericano. El resultado y el efecto fue impactante. El mismo rotativo, en su portada impresa del 21 de febrero de 2021, volvió a proponer una nueva codificación visual de esas muertes. En este caso, se codificó mediante un punto en esa portada cada uno de los casos registrados entre el casi medio millón de fallecimientos por coronavirus en Estados Unidos hasta esa fecha. En esa propuesta se puede observar cómo va aumentando la densidad de muertos a lo largo de la pandemia, los picos de muertes en las olas y cómo de elevado es el número de fallecidos a medida que nos acercamos a la fecha de la portada.

La última de las piezas que queremos incluir en este apartado intenta reivindicar la importancia



Figura 10. Portada de *The New York Times* del 21 de febrero de 2021 <https://static01.nyt.com/images/2021/02/21/nytfrontpage/scan.pdf>

de la alfabetización gráfica y numérica en nuestra sociedad para ser poder llegar a ser ciudadanos libres y no manipulados. No es suficiente con saber leer y escribir.

Concretamente, me gustaría recuperar el sonado caso la multinacional farmacéutica norteamericana *Purdue Pharma* (y el tratamiento audiovisual que se hace del mismo en la serie *Dopesick* (*Disney+*) estrenada el pasado año). Para situar el tema, remarcar que la multinacional farmacéutica utilizó un gráfico con una escala logarítmica y no lineal para basar una campaña publicitaria engañosa y hacer creer así a la comunidad médica que su producto *OxyContin* (un opiáceo) no creaba adicción como el resto de los opiáceos. Y, por tanto, que podía recetarse para cualquier dolor leve. Esto creó unos ingresos enormes para la farmacéutica, pero también una oleada de adictos y violencia asociada a su consumo.

En el gráfico 11, con la escala lineal en el eje Y, se ve el pico de su molécula en sangre después de la ingesta y cómo luego esa concentración va bajando. Eso ocurre con todos los opiáceos. Pero, dejando al margen el tema de los posibles sobornos, cuando hicieron el informe para que la *FDA* (*Food and Drug Administration*) terminara aprobando el medicamento y aceptara que no creaba adicción cambiaron la escala por una logarítmica en el eje Y (figura 12). Esto puede dar a entender erróneamente que no hay pico (euforia) ni bajada drástica cuando se consume la droga y que por tanto no crea adicción. El pico del gráfico lineal se transforma en un valle poco elevado en la logarítmica.

Eso lo exageraron todavía más cuando hicieron la publicidad entre la comunidad médica responsable de la prescripción de la sustancia (figura 13). En el gráfico superior (con escala lineal en el eje Y) se representa la presencia en sangre de los opiáceos después de su ingesta: pico con euforia y bajada de la concentración hasta la nueva ingesta (se genera adicción). En el gráfico inferior (con escala logarítmica) se representa la presencia en sangre de su molécula *OxyContin* después de la ingesta: sin pico, por tanto, sin euforia ni bajada acusada (no provoca

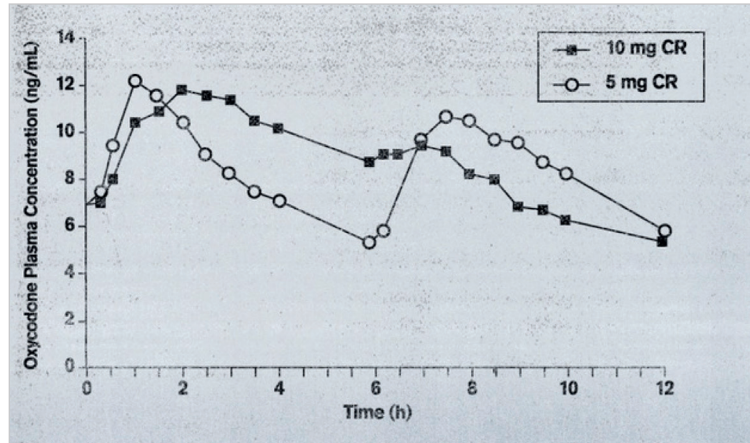


Figura 11. Evolución de la concentración en sangre del opiáceo *OxyContin* (gráfico con escala lineal en el eje Y) https://twitter.com/benjamin_rubin/status/1456752559155302412

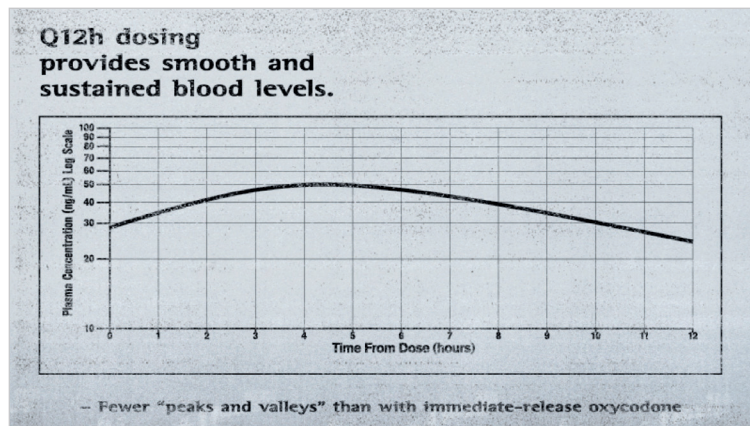


Figura 12. Evolución de la concentración en sangre del opiáceo *OxyContin* (gráfico con escala logarítmica en el eje Y) https://twitter.com/benjamin_rubin/status/1456752559155302412

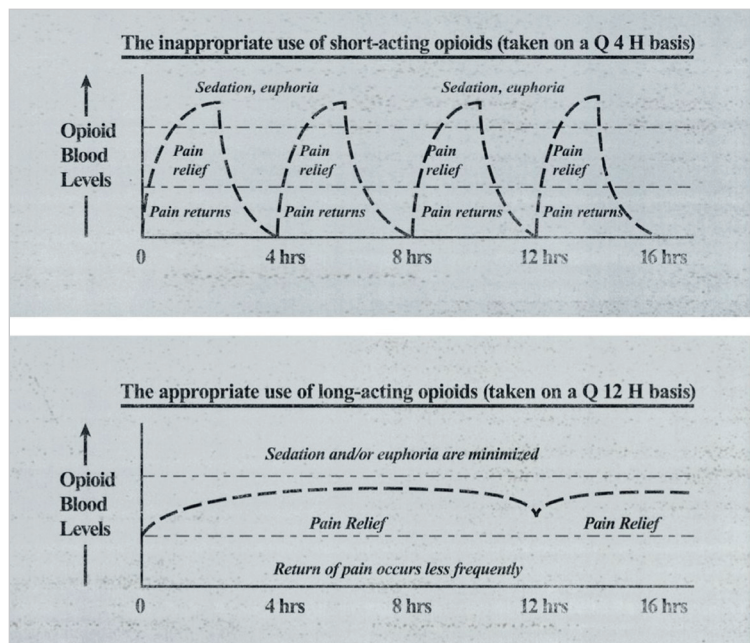


Figura 13. Evolución de la concentración en sangre de otros opiáceos (gráfico superior, con escala lineal) y *OxyContin* (gráfico inferior con escala logarítmica en el eje Y) https://twitter.com/benjamin_rubin/status/1456752559155302412

adicción). Además, para rematar y completar el engaño, hacen desaparecer los valores numéricos en las marcas de los ejes Y en las dos propuestas visuales para que nadie descubra o pueda darse cuenta de que hay una variación de escalas entre esas dos gráficas.

4. Tecnología

Completemos nuestra visión personal del desarrollo anual de la disciplina abordando algunos avances relacionados con el progreso tecnológico que se ha producido en el campo de la visualización de la información a lo largo de este año. Concretamente, vamos a remarcar algunas aportaciones encaminadas a facilitar el uso de las tecnologías por parte de un público no especialista o con un nivel limitado de conocimientos y competencias relacionados con éstas.

La primera aportación que queremos destacar nos la ofrece la empresa de software de visualización *Flourish*. Preocupada porque los periodistas deportivos puedan utilizar buenos recursos visuales en sus crónicas que publican en los cibermedios, en el post “How to visualize the Olympics” del blog de la compañía se ofrecen una serie de plantillas visualmente muy impactantes para representar datos sobre competiciones deportivas y, de forma especial, sobre los juegos olímpicos.

Entre estas plantillas ofrece unos recursos para representar una carrera de atletismo (figura 14), una competición de natación, una carrera ciclista dentro de un circuito cerrado, una cuenta atrás para codificar el tiempo que queda para comenzar el evento deportivo, un desglose histórico de atletas olímpicos por género, un mapa marcador ampliable e interactivo de los lugares donde se desarrollan los eventos, una tabla interactiva con el reparto de medallas o trofeos o para visualizar el análisis histórico

de una competición de atletismo (figura 14), una competición de natación, una carrera ciclista dentro de un circuito cerrado, una cuenta atrás para codificar el tiempo que queda para comenzar el evento deportivo, un desglose histórico de atletas olímpicos por género, un mapa marcador ampliable e interactivo de los lugares donde se desarrollan los eventos, una tabla interactiva con el reparto de medallas o trofeos o para visualizar el análisis histórico

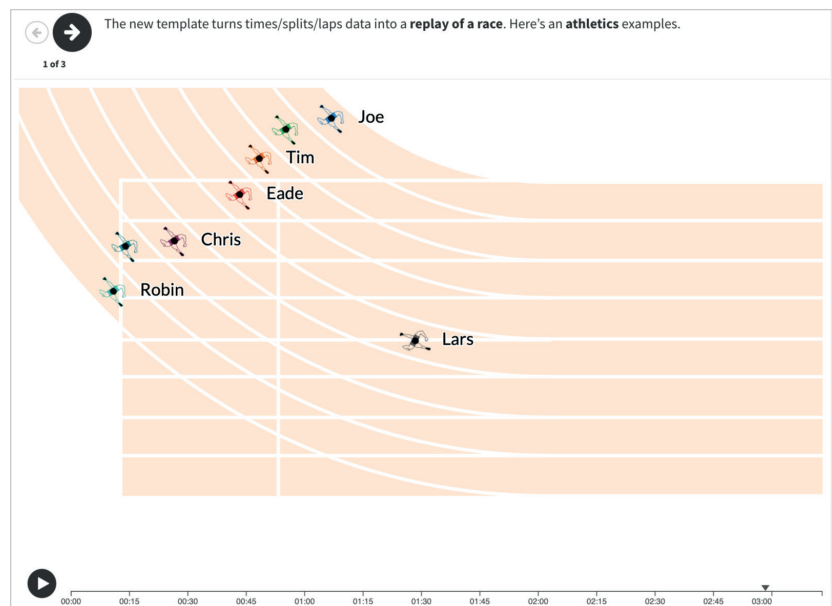


Figura 14. Plantilla para representar una carrera de atletismo
<https://flourish.studio/blog/olympics-tokyo-2021>

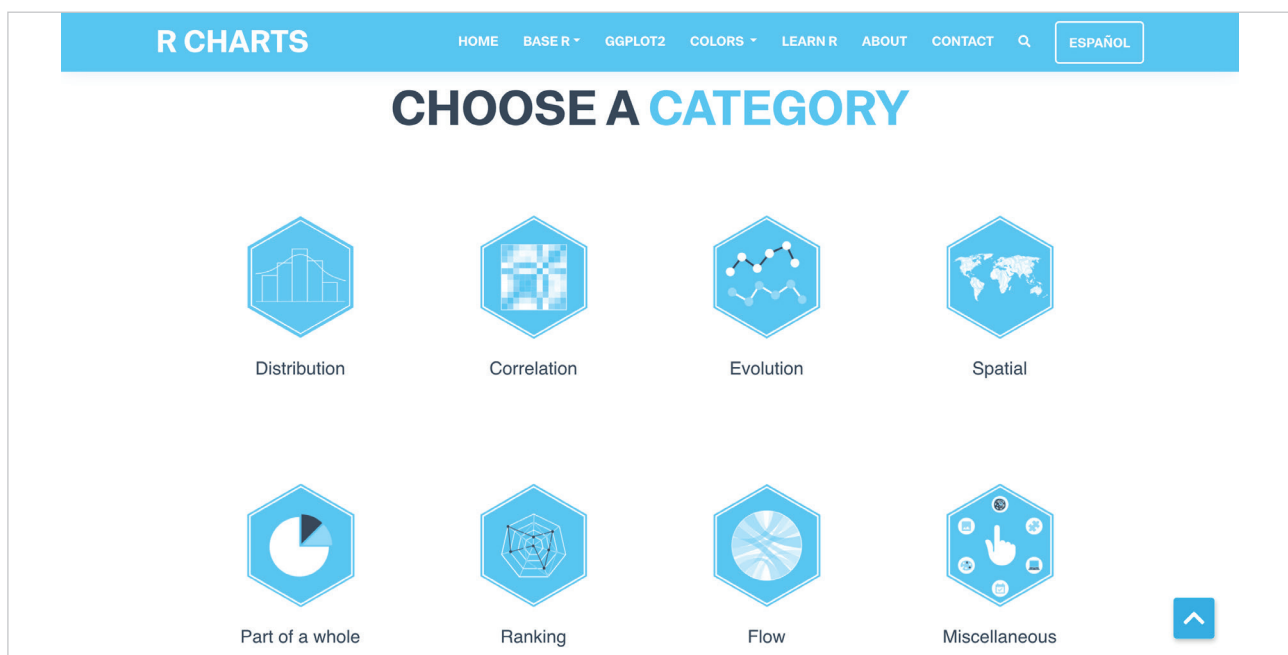
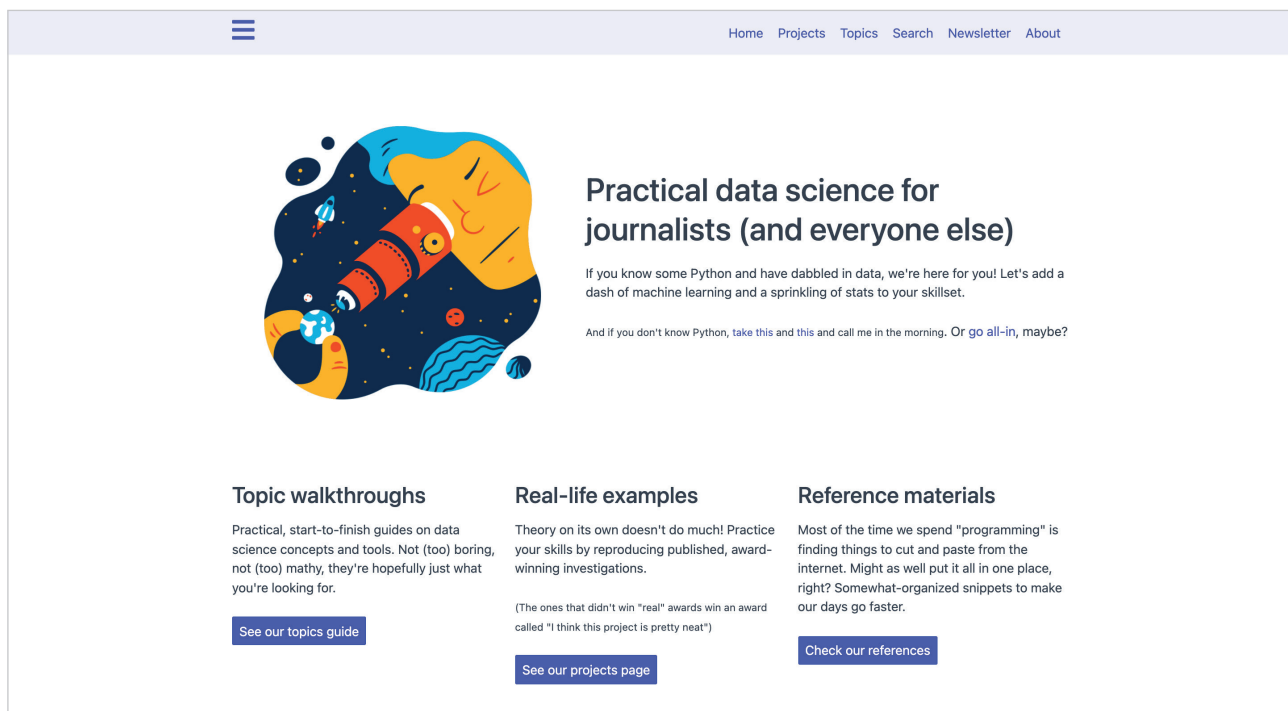


Figura 15. RCharts
<https://r-charts.com>

de resultados mediante el uso de un mapa de calor interactivo. Sólo se tienen que cambiar los datos por aquellos que queramos ilustrar y el recurso muestra una propuesta visual dinámica y preparada para incorporar a nuestra pieza digital.

En la misma línea, queremos remarcar la propuesta *RCharts*, creada por José Carlos Soage. Se trata de una web (figura 15) donde podemos encontrar recursos interesantes para usuarios de la herramienta *R*, independientemente de su conocimiento y competencia respecto a la misma. La web recoge ejemplos de más de mil gráficos con código reproducible que están organizados en 8 categorías y más de 50 tipos de gráficos. Los ejemplos se encuentran organizados bajo la clasificación *Financial Times visual vocabulary Financial* (*Financial Times visual vocabulary*, 2018) y utilizan *ggplot2* y gráficos base *R*, entre otras librerías.

Por último, queremos completar este trabajo destacando la página web *Practical data science for journalists (and everyone else)* (*Investigate.ai*, 2021) (figura 16). De forma parecida al destacado anterior, en esta página se recogen una serie de recursos relacionados con la Ciencia de Datos y orientados, principalmente, hacia la praxis periodística. Concretamente, y de forma especial utilizando *Python*, ofrece recursos para, entre otros, el análisis textual (contaje de palabras, análisis avanzado de palabras, análisis de los sentimientos, exportación de documentos a texto y tratamiento de conceptos, personas y lugares), para la clasificación automatizada [técnicas (evaluación de clasificadores, tratamiento de categóricas y clasificación con texto) y proyectos (*BuzzFeed*, *WaPo chat*, *The New York times* y *LA Times*)] y para el análisis de correlaciones y regresiones. La página web se completa ofreciéndonos una serie de trabajos periodísticos reales (implementados en medios como *The New York times*, *LA times*, *USA today*, *FiveThirtyEight*, *Dallas morning news*, *Tampa bay times*, *ProPublica* o el *Boston globe*, entre otros) en los que se han aplicado estas tecnologías, mostrándonos el tratamiento estadístico y el código desarrollado para extraer, limpiar, depurar, gestionar y analizar los datos implicados en esas piezas generadas.



Home Projects Topics Search Newsletter About

Practical data science for journalists (and everyone else)

If you know some Python and have dabbled in data, we're here for you! Let's add a dash of machine learning and a sprinkling of stats to your skillset.

And if you don't know Python, take this and this and call me in the morning. Or go all-in, maybe?

Topic walkthroughs

Practical, start-to-finish guides on data science concepts and tools. Not (too) boring, not (too) mathy, they're hopefully just what you're looking for.

[See our topics guide](#)

Real-life examples

Theory on its own doesn't do much! Practice your skills by reproducing published, award-winning investigations.

(The ones that didn't win "real" awards win an award called "I think this project is pretty neat")

[See our projects page](#)

Reference materials

Most of the time we spend "programming" is finding things to cut and paste from the internet. Might as well put it all in one place, right? Somewhat-organized snippets to make our days go faster.

[Check our references](#)

Figura 16. Practical data science for journalists (and everyone else) (<https://investigate.ai>)

5. Referencias

Financial Times visual vocabulary (2018).

<https://raw.githubusercontent.com/ft-interactive/chart-doctor/master/visual-vocabulary/poster.png>

Flourish (2021). "How to visualize the Olympics". *Flourish*, 23 July.

<https://flourish.studio/blog/olympics-tokyo-2021>

Investigate.ai (2021). "Practical data science for journalists (and everyone else)". *Investigate.ai*.

<https://investigate.ai>

Kessler, Glenn; Kelly, Meg; Rizzo, Salvador; Shapiro, Leslie; Domínguez, Leo (2021). "A term of untruths. The longer Trump was president, the more frequently he made false or misleading claims". *The Washington post*, 23 Jan. <https://www.washingtonpost.com/politics/interactive/2021/timeline-trump-claims-as-president/>

Baden, Lindsey, R.; El-Sahly, Hana M.; Essink, Brandon; Kotloff, Karen; Frey, Sharon; Novak, Rick; Diemert, David; Spector, Stephen; Rouphael, Nadine; Creech, Buddy; McGettigan, John; Khetan, Shishir; Segall, Nathan; Solis, Joel; Brosz, Adam; Fierro, Carlos; Schwartz, Howard; Neuzil, Kathleen; Corey, Lawrence; Gilbert, Peter; Janes, Holy; Follmann, Dean; Marovich, Mary; Mascola, John; Polakowski, Laura; Ledgerwood, Julie; Graham, Barney S.; Bennett, Hamilton; Pajon, Rolando; Knightly, Conor; Leav, Brett; Deng, Weiping; Zhou, Honghong; Han, Shu; Ivarsson, Melanie; Miller, Jacqueline; Zaks, Tal (2021). "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine". *The New England journal of medicine*, v. 384, p. 403-416. <https://doi.org/10.1056/NEJMoa2035389>

MacInnes, Jeff (2021). "Following the science". *The pudding*. <https://pudding.cool/2021/03/covid-science>

Pérez-Montoro, Mario (2021). "Comunicación visual de una emergencia sanitaria mundial: el caso del Covid-19". *Anuario ThinkEPI*, v. 15, e15d01. <https://doi.org/10.3145/thinkepi.2021.e15d01>

Polack, Fernando P.; Thomas, Stephen J.; Kitchin, Nicholas; Absalon, Judith; Gurtman, Alejandra; Lockhart, Stephen; Perez, John L.; Pérez-Marc, Gonzalo; Moreira, Edson D.; Zerbini, Cristiano; Bailey, Ruth; Swanson, Kena A.; Roychoudhury, Satrajit; Koury, Kenneth; Li, Ping; Kalina, Warren V.; Cooper, David; Frenck, Robert W.; Hammitt, Laura L.; Türeci, Özlem; Nell, Haylene; Schaefer, Axel; Ünal, Serhat; Tresnan, Dina B.; Mather, Susan; Dormitzer, Philip R.; Şahin, Uğur; Jansen, Kathrin U.; Gruber, William C. (2020). "Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine". *The New England journal of medicine*, v. 383, p. 2603-2615. <https://doi.org/10.1056/NEJMoa2034577>

Rakich, Nathaniel; Mithani, Jasmine (2021). "What absentee voting looked like in all 50 states". *FiveThirtyEight*, Feb 9. <https://fivethirtyeight.com/features/what-absentee-voting-looked-like-in-all-50-states>

RCharts (2021). <https://r-charts.com>

Rubin, Benjamin (2021). *Twitter*. https://twitter.com/benjamin_rubin

Sánchez, Raúl; Plaza, Analía (2021). "España vive en pisos: por qué hemos construido nuestras ciudades en vertical". *Eldiario.es*. <https://especiales.eldiario.es/espana-vive-en-pisos>

IPCC (2021). *The Intergovernmental Panel on Climate Change*. <https://interactive-atlas.ipcc.ch/regional-information>

The New York times (2021). Portada del 21 de febrero de 2021. <https://static01.nyt.com/images/2021/02/21/nytfrontpage/scan.pdf>

The Washington post (2021). "In four years, President Trump made 30,573 false or misleading claims". *The Washington post*, Jan 20. https://www.washingtonpost.com/graphics/politics/trump-claims-database/?itid=ik_interstitial_manual_9

Tufte, Edward R. (1990). *Envisioning information*. Cheshire: Graphics Press.

UpSet (2020). <https://upset.app/#upset-vs-venn-diagrams>