

INTRODUCCIÓN AL PERIODISMO DE DATOS
PROF. WALTER CALDERON RAMÍREZ

Correo: wcalderon@unan.edu.ni

I. DATOS GENERALES DEL CURSO

Horas Teóricas	Créditos HT	Horas Prácticas	Créditos HP	Total de Horas	Total de Créditos
8.4	0.56	12.6	0.42	21	0.98

II. INTRODUCCIÓN AL CURSO

El curso del Diplomado en Periodismo Científico en Salud con énfasis en Covid-19, contempla la identificación de grandes volúmenes de datos, la visualización interactiva de dichos datos y el desarrollo de tareas que incluyen extracción de datos, depuración del set de datos y el uso de dichos datos para producir información verificada accesible a diversas audiencias de los medios de comunicación. En consecuencia, solo se puede realizar periodismo de datos, si y sólo si, se cuenta con datos para la construcción de una narrativa que represente estos datos interpretados.

El periodismo de datos es la aplicación del método científico de la investigación social y comportamental a la práctica del periodismo de investigación. Las actitudes, conocimientos y habilidades de investigación desarrollados en el presente curso facilitarán el abordaje, en el curso subsiguiente, redacción periodística en el producirán un reportaje de divulgación científica sobre la Covid-19.

III. OBJETIVOS DEL CURSO

- a) Comprender los principales conceptos del periodismo de datos y su relación con el periodismo de investigación y el periodismo de precisión.
- b) Aplicar el método científico a la investigación periodística basada en datos para la cobertura de la COVID-19.
- c) Crear una base de datos sobre el tema de investigación periodística de la COVID-19.
- d) Valorar el cumplimiento de normas y altos estándares éticos que regulan la profesión periodística.

IV. PLAN TEMÁTICO

		Horas teóricas		Horas prácticas		Total de horas
Unidad de aprendizaje	Contenidos	Presenciales	Independiente	Presenciales	Independiente	
Construcción de bases de datos	1. ¿Qué es el periodismo de datos?	2	4	2	4	12
	2.Procedimientos del periodismo de datos	1	2	1	8	12
	3. Técnicas de periodismo de datos	1	2	1	8	12
	4.Scraping y crowdsourcing	1	2	1	8	12
Contar historias con datos	1. Presentación de datos	1	2	1	8	12
	2. Aplicaciones con datos	1	2	1	8	12
	3.Visualizaciones de datos	1	2	1	8	12
TOTALES		8	16	8	16	

V. METODOLOGIA DEL CURSO

La fase presencial de este curso estará basada en el desarrollo de una Conferencia Magistral con enfoque activo y participativo, iniciando la sesión con una evaluación diagnóstica basada en la estrategia didáctica de la ubicación contextual, sobre la cual la (UNAN-Managua, 2011, págs. 38-39) nos explica:

El estudiante aprende en un entorno contextualizado, cuyo punto de partida es el contexto personal. En este sentido él llega al momento de aprendizaje con interés causado por la incertidumbre, la necesidad y la novedad. Por eso, es obligatorio que el docente contextualice la información con el objeto de que el discente la incorpore significativamente a su estructura mental. Por ello, el aprendizaje debe ser situado en un contexto que revele su origen: social, político, económico, científico, cultural, a fin de que se comprenda el porqué y el para qué se estudia tal o cual saber. Entonces y solo entonces, el sujeto será consciente de lo significativo del aprendizaje.

Se realizará además el contrato didáctico, en el que se establecerán de común acuerdo entre docentes y estudiantes, las reglas del curso en cuanto a: puntualidad en la realización de actividades, formas de participación presencial y en línea, justificación de ausencias o entregas tardías de trabajos, entre otros aspectos propios del contexto educativo.

La conferencia se apoyará en diversos medios audiovisuales y consistirá en bloques de una hora, después de cada uno de los cuales se irá discutiendo el contenido desarrollado, con base en la estrategia didáctica conocida como Guía de Cuestionamiento de lo que se aprende, definida en (UNAN-Managua, 2011, pág. 39) de la siguiente manera:

Se pueden elaborar guías de preguntas en los dominios del saber (conocimiento teórico sobre conceptos, datos, hechos, acontecimientos, lugares). Estas desarrollan esquemas de acción que permiten reconocer o recordar información, explicar un mismo contenido en diferentes formas, interpretar el significado de alguna información, comparar o relacionar eventos y ejemplificar ...

En la fase *on line* se orientará y asesorará, con el soporte de la Plataforma Teams, un proceso de investigación documental orientado al reforzamiento de los temas abordados, asimismo se generarán foros de discusión con la moderación del docente. Se trabajará la redacción de casos basados en el contexto laboral de los participantes, actividad en la que se implementará la Técnica del Coach como una forma de generar retroalimentación entre los mismos. Se interactuará con un sistema de información en línea basado en los principios de la Creación y Gestión del Conocimiento. También se elaborará sobre un caso de toma de decisión grupal a fin de analizar el papel que juegan en estos procesos, los datos, la información y el conocimiento.

VI. EVALUACIÓN DEL CURSO

Se deberán realizar actividades que cumplan el doble propósito de la consolidación de los conocimientos y de la evaluación de los mismos. La base de interacción para la asesoría, retroalimentación y entrega de tareas lo constituirá la plataforma virtual seleccionada, de la cual deben aprovecharse los mecanismos de comunicación sincrónica y asincrónica para potenciar el trabajo en equipo y las metodologías participativas. Los participantes deberán demostrar que se han apropiado de los conocimientos propuestos y que son capaces de contextualizar y entender la aplicación de los mismos en su quehacer laboral.

Se sugieren las siguientes actividades con sus correspondientes pesos de evaluación.

- Actividad 1: Elaboración de base de datos sobre la COVID-19, a partir de temas de investigación (20 %).
- Actividad 2: Curación de la base de datos (20 %).
- Actividad 3: Extracción de datos y depuración del set de datos (30 %).
- Actividad 4: Visualización interactiva de datos (30 %).

VI. DESARROLLO DEL CURSO

MINUTOS	ACTIVIDADES	METODOLOGIA	EVALUACIÓN	RECURSOS	PRODUCTO
15 minutos	Presentación metodología de trabajo	El facilitador presenta la metodología de trabajo en grupos y la agenda del día.	Formativa	Audio y Micrófono	Explicada la metodología de trabajo del día
30 minutos	Conferencia interactiva	El facilitador presenta contenido temático e invita a la discusión a través de preguntas generadoras -Periodismo científico -Periodismo de precisión -Periodismo de datos	Formativa	Audio y Micrófono	Apropiación del conocimiento
15 minutos	Trabajo de grupo. Actividad 1. Instrucciones y desarrollo	El facilitador explica las instrucciones de la Actividad 1: Elaboración de la base de datos	Evaluativa	Computadora, proyector, micrófono, dossier	Instrucciones del trabajo de grupo clarificadas
10 minutos	PAUSA ACTIVA				
30 minutos	Visionado de documental: Cosmos, <i>"El descubrimiento que nos cambió la vida"</i>	El facilitador explica el trabajo en grupo: Guía de análisis de periodismo científico	Formativa	Audio y Micrófono	Apropiación del conocimiento
20 minutos	Trabajo de grupo: Foro interactivo (Guía de identificación de criterios de periodismo científico (Viera, 2004).	Plenaria y presentación del trabajo en grupo	Formativa	Audio y Micrófono	Apropiación del conocimiento
10 minutos	Conclusiones de Plenaria	El facilitador destaca los principales aprendizajes del trabajo en grupo	Formativa	Audio y Micrófono	Apropiación del conocimiento
10 minutos	PAUSA ACTIVA				
30 minutos	Conferencia interactiva:	El facilitador presenta el	Formativa	Audio y Micrófono	Apropiación del

	Curación de contenidos de la base de datos	tema e invita a participantes a cuestionar los siguientes conceptos de ciencia: -Certeza/Incertidumbre -Búsqueda, extracción y limpieza de datos			conocimiento.
30 minutos	Trabajo en grupo	Actividad 2. Aplicación de procedimientos para la curación de contenidos de la base de datos.	Evaluativa	Computadora, proyector, micrófono, dossier	Apropiación del conocimiento
15 minutos	Evaluación de la primera sesión de aprendizaje	Evaluación conforme a criterios de evaluación de aprendizajes	Formativa	Google Forms	Apropiación del conocimiento
SEGUNDA SESIÓN DE APRENDIZAJE					
10 minutos	Presentación de la agenda y expectativas de la sesión	El facilitador presenta la metodología de trabajo en grupos y la agenda del día.	Formativa	Audio y Micrófono	Explicada la metodología de trabajo del día
30 minutos	Conferencia interactiva: análisis de datos	El facilitador explicará las principales técnicas y herramientas para el procesamiento y análisis de dato: -Scraping y crowdsourcing - Extracción de datos y depuración del set de datos	Formativa	Audio y Micrófono	Apropiación del conocimiento
20 minutos	Trabajo en grupos	Actividad 3. Extracción de datos y depuración del set de datos	Evaluativa	Computadora, proyector, micrófono, dossier	Apropiación del conocimiento
10 minutos	PAUSA ACTIVA				
30 minutos	Conferencia interactiva: visualización de datos	El facilitador explicará las principales técnicas y herramientas para la visualización de datos -Creación de gráficos -Herramientas para la visualización	Formativa	Audio y Micrófono	Apropiación del conocimiento
50 minutos	Trabajo en grupos	Actividad 4. Elaboración de gráficos (Tablas, infografías, mapas interactivos etc.)	Evaluativa	Audio, micrófono y pantalla	Apropiación del conocimiento
10 minutos	Cierre de la sesión: -Indicaciones para el subproducto integrador - Evaluación 360º	El Facilitador indicará las bases orientadoras de la acción para la entrega del subproducto integrador (Elaboración de	Formativa	Audio, micrófono y Google Forms	Apropiación del conocimiento

VII. BIBLIOGRAFIA DEL CURSO

Lecturas obligatorias (Dossier del curso)

- Crucianelli, S. (2013). *¿Qué es el periodismo de datos?* <https://www.cuadernosdeperiodistas.com/media/2013/12/106-1241.pdf>
- Dader, J., & Fernández, P. (1993). "Periodismo de precisión": una nueva metodología para transformar el periodismo. *Anàlisi: quaderns de comunicació i cultura*, (15), 99-116.
- Renó, L., & Saad, E. (2017). Reportaje soportado por el computador: procedimientos y tecnología para el periodismo de datos contemporáneo. *Razón y Palabra*, 21(97), 128-141.
- Sagan, C. (2017). No hay preguntas tontas. En: Sagan, C. *El mundo y sus demonios*. Editorial Planeta. pp. 345-364.
- Tola, E. (2020). Cómo interpretar las estadísticas científicas, En: Knight Science Journalism Program at MIT, *Manual de Edición Científica por KSJ*. pp. 42-66.
- Viera, C. (2004). *Breve manual de comunicación de la ciencia*. http://www.piel.com.ve/ediciones/94/breve_manual_comunicacion_ciencia.pdf

LECTURAS COMPLEMENTARIAS

- Berganza, M. & Ruiz, J. (2005). *Investigar en Comunicación: Guía práctica de métodos y técnicas de investigación social en comunicación*. Editorial McGraw Hill.
- Bilbeny, N. (2012). *Ética del periodismo*. Publicacions i Edicions de la Universitat de Barcelona.
- Blanco, P. (2001). *Una aproximación al periodismo de precisión*. Publicaciones UCAB.
- Callegari, N. (2019). *Géneros periodísticos de hoy: Este es el manual*. Ecoe Ediciones.
- Cytrynblum, A. (2009). *Periodismo social: Una nueva disciplina*. La Crujía.
- De Lara, A. & Gómez, A. (2021). *Ciencia y Periodismo. Una es de Marte y otra es de Venus*. Universidad Miguel Hernández de Elche.
- Elías, C. & Teira, D. (2021). *Manual de periodismo y verificación de noticias en la era de las fake news*. Universidad Nacional de Educación a Distancia.
- García, J. (2015). *Comunicar en la Sociedad Red: Teorías, modelos y prácticas*. Editorial UOC.
- Hunter, M. (2013). *La investigación a partir de historias: manual para periodistas de investigación*. UNESCO.
- Irigaray, F. (2015). *Reflexiones móviles. El periodismo en la era de la movilidad*. Editorial de la Universidad del Rosario.
- Jarvis, J. (2015). *El fin de los medios de comunicación de masas*. Gestión 2000.
- Orozco, G. (2019). *TVmorfosis 8. Periodismo en la era de la posverdad*. Productora de contenidos culturales Sahagón Repoll.
- Oyarce, J. Belda, F., Velarde, H., Gutiérrez, N. (2021). *Infodemia: la otra cara de la información*. RIA Editorial.
- Ramonet, I. (2014). *La explosión del periodismo: De los medios de masas a la masa de medios*. Editorial Universitaria. Universidad de Guadalajara.
- Renó, L. (2018). *Manual de Periodismo de Datos*. RIA Editorial.
- Yaguana, H., Sánchez, A., Pisco, G. (Coords.). (2022). *Nuevas miradas de la comunicación y el periodismo*. RIA Editorial.

VIII. AUTORES

Nombre y Firma del Autor 1

Nombre y Firma del Autor 2

Nombre y Firma del Autor 3

Nombre y Firma del
Coordinador Académico del
Programa de Estudios de
Postgrado

Nombre y Firma del
Responsable de Postgrado de
la Facultad, Instituto o Centro
de Investigación

BASES ORIENTADORAS DE LA ACCION

Elaboración de un proyecto de periodismo visual

Nombre del componente curricular: Introducción al Periodismo de Datos

Actividad integradora: Elaboración de proyecto de periodismo visual

Trabajo a entregar: (10 de septiembre de 2022, sin prórroga)

II. OBJETIVOS DE APRENDIZAJE:

- a) Aplicar el método científico a la investigación periodística basada en datos para la cobertura de la COVID-19.
- b) Elaborar una base de datos sobre COVID-19
- c) Aplicar procedimientos para la curación de contenidos de la base de datos.
- d) Realizar extracción de datos y depuración del set de datos
- e) Elaborar gráficos (Tablas, infografías, mapas interactivos etc.)
- f) Valorar el cumplimiento de normas y altos estándares éticos que regulan la profesión periodística.

III. COMPETENCIAS:

- Habilidad de comprender la realidad sociodemográfica de nuestro país frente a la pandemia por COVID-19
- Ser capaz de comunicarse mediante el uso de gráficos que visualicen los datos.
- Destreza para utilizar las TIC como apoyo para mejorar el aprendizaje de en diferentes ámbitos de actuación y campos de acción profesional.

IV. ACCIONES

Estimado estudiante para la realización de un proyecto de periodismo visual realizo las siguientes acciones:

ELABORO BASE DE DATOS

1. Me organizo en un grupo de trabajo para seguir trabajando con mi tema de investigación, para lo cual elaboraré una base de datos. La identificación y selección de los datos los obtendré buscando ya sea en los repositorios o bases de datos del Consejo Nacional de Universidades/UNAN-Managua o las instituciones públicas del Estado de Nicaragua: Ministerio de Salud, INIDE, Ministerio de Hacienda y Crédito Público, Banco Central de Nicaragua, Ministerio de Educación, Asamblea Nacional de la República de Nicaragua etc. Los datos los voy a reunir en una tabla de Excel, los cuales deben ser guardados en el formato CSV.

REALIZO CURACIÓN DE LA BASE DE DATOS (EXTRACCION Y DEPURACIÓN DEL SET DE DATOS)

La curación de los datos, no es más que el procedimiento mediante el cual, mediante el cual me aseguro que voy a utilizar los datos que sean significativos y me permitan evitar sesgos. Por ello, debo deshacerme de todos los datos que no cumplen con los criterios de calidad para su procesamiento y análisis.

ELABORO ANÁLISIS DEL SET DE DATOS

Después del proceso de búsqueda, extracción y limpieza de las informaciones deseadas, ahora es hora de hacer el análisis de los datos. En realidad, ahí puede estar la noticia, lo que interesa al periodismo de datos. La herramienta utilizada para eso procedimiento es Google Sheets, que es una herramienta completa para esta etapa. Es posible trabajar con diversos procesos para analizar los datos, como, por ejemplo, clasificación, fórmulas y funciones, filtros, tablas dinámicas y gráficos.

DISEÑO LA VISUALIZACIÓN DE MIS DATOS

Visualización de datos es la representación visual de las informaciones para mostrar evidencias, permitir análisis e informar puntos fundamentales para el reportaje. En los procesos de visualización se puede presentar verdades ocultas en datos complejos. La utilización de gráficos permite la comparación entre los contenidos y construye de manera sencilla la transmisión de informaciones.

V. RESULTADOS ESPERADOS

Desarrollado el nivel de habilidades y competencias propuestas en el componente curricular para el análisis y visualización de datos.

VI. EVALUACIÓN

Proceso de Autoevaluación - Técnicas, desde el primer día se deberá recoger información referente (base de datos, extracción y depuración del set de datos, procesamiento y análisis y visualización de datos.

PROYECTO DE PERIODISMO VISUAL
HOJA DE COTEJO

Elaboración de un proyecto de periodismo visual			Calificación
Elaboración de base de datos	<p>Cumplido La base de datos se ha construido en Excel y los datos se han guardado en formato CSV</p>	<p>Incumplido La base de datos en Excel, está incompleta y el formato no permite realizar el análisis</p>	20
Curación de la base de datos	Los datos se han extraído y procesado adecuadamente evitando errores y sesgos.	Los datos no han sido extraídos, ni procesados. No es posible realizar ningún procedimiento por la falta de calidad en los datos.	20
Análisis del set de datos	El análisis estadístico de los datos nos permite obtener algunas medidas de tendencia central y asegurar la validez y confiabilidad de los datos	Los datos no pueden ser analizados porque la muestra no cumple con los criterios de validez y confiabilidad	30
Visualización de datos	La representación visual de los datos, le permite a la audiencia comprender de forma fácil y bajo criterios de periodismo científico sobre la pandemia de la COVID-19	La representación de los datos es confusa y no es posible comprender el fenómeno descrito.	30
		TOTAL	100

BREVE MANUAL SOBRE COMUNICACIÓN DE LA CIENCIA -RESUMEN

Cassio Viera

Julio 2004

Introducción

En 1999 publiqué un folleto de 50 páginas llamado *Breve manual sobre comunicación de la ciencia – Consejos prácticos para científicos y comunicadores de la ciencia*. Esta empresa surgió de mi trabajo en *Ciencia Hoje*, una revista de comunicación de la ciencia publicada por la Asociación Brasileña para el Avance de la Ciencia.

Ciencia Hoje presenta ciertas similitudes con *Scientific American*. La revista se compone principalmente de artículos redactados por científicos y su objetivo ha sido siempre fomentar un acceso asequible a la información de la ciencia. Mi trabajo como editor de ciencias puras en la revista incluye la “traducción” de textos científicos en un lenguaje apropiado para lectores no especializados. Sin embargo, los materiales que recibía contenían con frecuencia un lenguaje técnico poco apropiado y difícil, incluyendo algunas veces fórmulas complejas y jerga propia del sector. Para empeorar la situación la revista no tenía ninguna guía para redactores. Por ello decidí preparar una yo mismo. Como resultado nació el Pequeño manual: un conjunto de normas de redacción científica basadas en el sentido común para lectores no especializados.

El resumen mostrado a continuación es una adaptación del manual, incluyendo normas seguidas de una breve explicación. Tal vez la vuelta a la simplicidad de estas normas básicas refleje la situación que se vive en Latinoamérica, donde la comunicación escrita con el público no especializado no está extendida dentro de nuestra comunidad científica como lo está entre los comunicadores europeos y norteamericanos. A pesar de ello espero que sean útiles también para científicos, redactores y comunicadores de ciencia de otros países.

LENGUAJE

Sea consciente de la existencia de “lenguajes” diferentes

El lenguaje propio de los artículos de comunicación de la ciencia debe ser diferente del empleado en los artículos científicos. Aunque parezca increíble, algunos miembros de la comunidad científica todavía no son conscientes de este hecho.

Cautive al lector

En la comunicación de la ciencia, el primer párrafo o la introducción es esencial para llamar la atención del lector, motivándole a leer el resto del texto. Las novelas y las historias cortas generalmente dejan lo mejor para el final. Sin embargo los artículos de comunicación de la ciencia deben comenzar por mensajes de impacto con cuestiones o hechos que llamen la atención del lector.

Evite asustar al lector

Si el primer párrafo o la introducción contienen conceptos y fórmulas demasiado científicas, conseguiremos que el lector no pase de las primeras líneas.

Use las analogías con prudencia

Las analogías son un elemento esencial dentro de la comunicación de la ciencia. Los mejores ejemplos de analogías describen conceptos con el lenguaje cotidiano de los lectores. Sin embargo, a veces puede ser necesario establecer cuáles son los límites de las analogías ante el lector para evitar extrapolaciones que puedan inducir a error. Veamos el siguiente ejemplo: “De acuerdo a la fórmula más famosa de la historia de la ciencia, $E = mc^2$ (E de energía, m de masa y c de velocidad de la luz), un kilogramo de materia proporciona 25 billones de kWh de energía, lo suficiente para cubrir la demanda actual de energía eléctrica de Brasil durante 8 meses. Sin embargo, los físicos todavía no saben cómo

transformar la materia en energía con 100% de eficacia. En términos prácticos el porcentaje es muy bajo. Incluso en el caso de explosiones atómicas no pasa del 1%”.

Sea riguroso

Cualquier información, ya sea científica o de otro tipo, debe ser siempre rigurosa. En términos de comunicación de la ciencia es vital distinguir la especulación de los resultados probados.

Concéntrese en el objetivo

Tal vez esta sea la norma más importante de las presentadas en este manual: no olvide quién es público destino. Incluso Einstein llegó a ser consciente de eso (lea el prefacio de *Evolution of Physics*). Esta norma es válida para cualquier tipo de lector, ya sean niños, público no especializado o especialistas.

Evite las fórmulas

Evite en general la utilización de fórmulas. Si cree necesario incluir alguna explique el significado de los elementos. Incluso fórmulas tan conocidas como $E=mc^2$ deberían ser explicadas. Por este mismo motivo deben evitarse las fórmulas químicas.

¡Recuerde que los doctores también se olvidan!

Otra cuestión a considerar es que incluso los doctores se olvidan de conceptos básicos de campos en los que no son especialistas, por lo que se hace necesario explicarlos. Después de todo y en términos de comunicación de la ciencia, nadie quiere tener que usar un diccionario para leer un artículo.

Aproveche su sentido del humor

El humor puede hacer que los artículos sean más amenos para el lector, asegurándole que deseen leerlos hasta el final. Por otro lado es importante no ofender a los lectores, por lo que hay que tener tacto al valerse de este recurso.

Evite estilo rococó

Utilice un lenguaje informal y sencillo. Recuerde: lenguaje simple no es incompatible con contenidos ricos.

Sea sucinto

Compare "Está expresamente prohibido fumar en esta sala" con "Prohibido fumar". Es fácil observar cuál es el estilo deseable. No olvide que el espacio es oro en los periódicos y revistas (consulte la sección "Ajústese al tamaño solicitado" mostrada más adelante).

Evite el uso de jerga

Evite el uso de jerga en términos generales, ya que hace que los artículos sean más "pesados". Si cree que es absolutamente necesario hacerlo, explique el término entre paréntesis o en un glosario (consulte la siguiente sección).

Explique siempre

Hemos comentado que no se debe utilizar jerga. Sin embargo es casi imposible evitar el uso de conceptos científicos, por lo que es conveniente explicarlos de la manera más simple posible. Ejemplos: cloruro de sodio (sal); hidróxido de sodio (soda cáustica); mitocondria (fábricas productoras de energía) etc. Evite el uso de términos científicos para explicar otros, tales como: fermios (partículas que obedecen a las estadísticas de Fermi-Dirac). Si ofrecer una explicación concreta le parece imposible, inténtelo de nuevo. Válgase de analogías, o al menos explique parcialmente el concepto. Siempre será mejor que presentarlo de manera ininteligible en nombre de la pureza conceptual.

Marque los conceptos complejos en cuadros

¿Necesita explicar conceptos más complicados o de carácter técnico? Destáquelos en cuadros o en bloques de texto adicionales y no olvide simplificarlos lo máximo posible.

Indique quién, qué, donde...

Utilice la expresión "el físico danés Niels Bohr (1885-1962)" en vez de simplemente "Bohr". Utilice la expresión "en 1998 el físico británico Joe Olmi del Departamento de Inteligencia Artificial de la Universidad de Gran Bretaña publicó un artículo sobre nanorobots en el *Journal of Robotics* (vol 20, n 456, p 457)..." en vez de "Según Olmi (J. of Rob...) 1998)...".

Escriba acrónimos sólo una vez

Nadie tiene la obligación de conocer todos los acrónimos. Por ello, use por ejemplo "Asociación Brasileña para el Avance de la Ciencia (ABAC)" en vez de sólo "ABAC".

No incluya notas a pié de página ni menciones

No hay espacio para este tipo de texto en los diarios y revistas. Evite también incluir citas bibliográficas tales como (*Science* 43 (6543):53, 1992), ya que normalmente no añaden información útil para el lector no especializado. Si necesita referirse a un artículo use la siguiente fórmula: "[...], publicado en la revista científica americana *Science* (vol 43, no 6.543, 1992, p 53)."

No aliente falsas esperanzas

En artículos médicos, deje claro cuándo los tratamientos o medicinas en cuestión están en periodo de prueba y no están a disposición del público en general. No olvide que el lector (o algún miembro de su familia) puede sufrir la enfermedad a la que se está refiriendo.

FORMA

Siga las reglas "espacio-tiempo"

Todos os medios de comunicación sufren la llamada dictadura "espacio-tiempo". En los diarios y revistas, por ejemplo, el número de palabras escritas debe ajustarse al área reservada para su artículo. Sea, por tanto, conciso y piense que incluso la Encyclopaedia Britannica no contiene toda la información referente a cada asunto.

Ajústese al tamaño solicitado

Redacte el número de palabras solicitadas por el editor. Caso contrario, los artículos que superen el límite tendrán que ser reducidos y los más cortos tendrán que incorporar más palabras. Algunos periodistas afirman que este último caso es más difícil que reducir el número de palabras. La mejor manera de evitar sorpresas es preguntarle al editor cuál es el número de palabras esperado.

Sugiera títulos

El título es la primera parte a ser leída. Decidir cuál es el más apropiado es un arte que normalmente esta bajo la responsabilidad de editores con experiencia. En los periódicos se siguen unas normas estrictas al respecto del número de palabras. Lo que sí es cierto es que los editores siempre agradecen sugerencias.

Facilite la vida del lector

Se recomienda el uso de párrafos cortos en vez de los largos. Las publicaciones diarias normalmente prefieren los cortos. Incluso cuando se trata de revistas, el sentido común nos dice que es bueno facilitar la vida de los lectores para que puedan asimilar mentalmente el contenido del texto.

Incluya imágenes cuando sea posible

Tanto si escribe para un diario como para una revista, intente incluir buenas imágenes o, al menos, indique al editor donde puede encontrarlas. Es preferible incluir imágenes de alta resolución (300 dpi o

superior). Evite mandar gráficos complicados, esquemas o tablas. Muchas personas tienen cierta dificultad para interpretarlas.

Ponga leyenda en las imágenes

En cierta ocasión, mientras editaba un artículo, observé la foto incluida y llamé a la criatura "oruga". Resultó ser un pez. Este tipo de error es factible cuando las imágenes no incluyen leyendas. No olvide pues incluir la leyenda.

No olvide el copyright

Nunca jamás olvide incluir el copyright de la fotografía. Si se trata de ilustraciones, tablas, diagramas, esquemas y en casos parecidos cite siempre la fuente. En caso de duda consulte al editor. Normalmente él o ella sabrá si la imagen puede ser reproducida.

Incluya otras perspectivas

Si fuera posible, incluya diferentes perspectivas sobre el tema tratado en el artículo. Caso contrario podemos transmitir al lector la idea errónea de que poseemos la verdad absoluta sobre el asunto.



No existen las preguntas tontas

Carl Sagan

Siempre es interesante conocer los puntos de vista de los grandes maestros de la divulgación científica. El siguiente fragmento, tomado del libro El mundo y sus demonios, muestra algunas de las opiniones del famoso Carl Sagan, quien inspiró a muchos de los que hoy nos dedicamos a compartir nuestro gusto por la ciencia.

¿Por qué es difícil para los científicos comunicar la ciencia? Algunos científicos, incluyendo algunos muy buenos, me dicen que les encantaría divulgar, pero sienten que carecen de talento en esa área. Saber y explicar, dicen, no son la misma cosa. ¿Cuál es el secreto?

Sólo hay uno, creo yo: no le hables al público general como le hablarías a tus colegas científicos. Hay términos que comunican lo que quieres decir en forma inmediata y precisa a tus colegas expertos. En tu trabajo profesional puedes hablar sólo tres frases al día. Pero ante una audiencia de no especialistas, eso no sirve más que para causar confusión. Usa el lenguaje más sencillo posible. Sobre todo, recuerda cómo era antes de que tú mismo comprendieras lo que sea que estás explicando. Recuerda los malentendidos en los que casi caíste, y menciónalos explícitamente. Mantén muy en mente que hubo un tiempo en que tú tampoco entendías nada de esto. Recapitula los primeros pasos que te llevaron de la ignorancia al conocimiento. Nunca olvides que la inteligencia está ampliamente

distribuida en nuestra especie. De hecho, es el secreto de nuestro éxito.

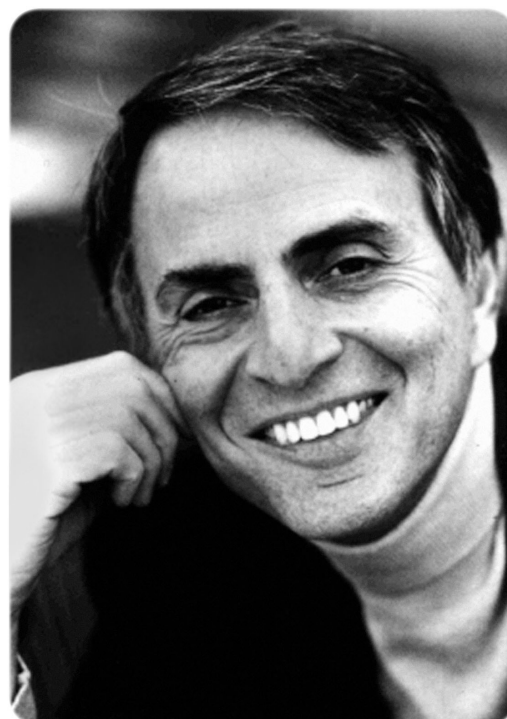
El esfuerzo que esto requiere es pequeño; los beneficios, grandes. Entre los escollos potenciales están la sobresimplificación, la necesidad de ser cauto con las cualificaciones (y las cuantificaciones), no dar el crédito adecuado a los muchos científicos involucrados, y no hacer las distinciones suficientes entre una analogía útil y la realidad. Sin duda, habrá que hacer concesiones.

Conforme vayas haciendo más presentaciones de este tipo, estará más claro qué funciona y qué no. Habrá una selección natural de metáforas, imágenes, analogías, anécdotas. Después de un tiempo encontrarás que puedes llegar casi a donde quieras, caminando sobre piedras probadas y garantizadas. Entonces podrás afinar tus presentaciones según las necesidades de cada audiencia particular.

Al igual que algunos editores y productores de televisión, ciertos científicos creen que el público es demasiado ignorante o estúpido para entender la ciencia, que la empresa de la divulgación es fundamentalmente una causa perdida, o incluso que equivale a fraternizar, si no es que abiertamente cohabitar, con el enemigo. Entre las muchas crí-

ticas que se podrían hacer de este tipo de juicios además de su insufrible arrogancia y su desdén por los numerosos y muy exitosos ejemplos de divulgación de la ciencia, es que se auto-confirman. Y también, para los científicos involucrados, se auto-derrotan:

El apoyo amplio del gobierno a la ciencia es bastante nuevo; data sólo desde la segunda guerra mundial, aunque el patrocinio de unos pocos científicos por los ricos y poderosos es mucho más viejo.



Con el final de la guerra fría, el pretexto de la defensa nacional, que había servido para apoyar todo tipo de ciencia fundamental se volvió virtualmente inaceptable. Es sólo por esta razón, creo, que hoy muchos científicos aceptan la idea de divulgar la ciencia. (Como casi todo el apoyo para la ciencia viene de las arcas públicas, si los científicos se opusieran a la divulgación competente estarían coqueteando con el suicidio.) Es más probable que el público apoye lo que entiende y aprecia. No me refiero a escribir artículos para *Scientific American*, digamos, que son leídos por entusiastas de la ciencia y científicos de otros campos. No estoy hablando sólo de impartir cursos introductorios para estudiantes de licenciatura. Hablo de esfuerzos para comunicar la sustancia y el enfoque de la ciencia en periódicos, revistas, en radio y televisión, en conferencias para el público general, y en libros de texto elementales, de enseñanza media y de bachillerato.

Desde luego que hay que establecer criterios en la divulgación. Es importante no confundir ni ser condescendiente. A veces, al intentar estimular el interés del público, los científicos han ido demasiado lejos por ejemplo, al sacar conclusiones religiosas injustificadas.

[...]Las publicaciones periódicas y la televisión pueden iniciar una reacción cuando nos permiten echar un vistazo a la ciencia, y esto es muy importante. Pero aparte de talleres, clases y seminarios bien estructurados, la mejor manera de popularizar la ciencia es a través de libros de texto, libros populares, *cd-roms* y discos láser. Uno puede repasar las cosas una y otra vez, ir a su propio paso, revisar las partes difíciles, comparar textos, excavar profundo. Tiene que hacerse bien, sin embargo, y en las escuelas normalmente no se hace así. En ellas, como comenta el filósofo John Passmore, la ciencia muchas veces se presenta «como una cuestión de aprender principios y aplicarlos mediante procedimientos rutinarios. Se aprende


en libros de texto, no al leer los trabajos de grandes científicos o incluso las contribuciones de cada día a la literatura científica. El científico principiante, a diferencia del humanista principiante, no tiene un contacto inmediato con el genio. De hecho[...] los cursos escolares pueden atraer a la ciencia precisamente al tipo equivocado de persona: chicos y chicas poco imaginativos a los que les gusta la rutina.»

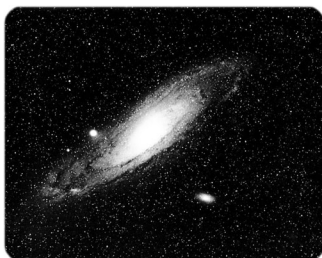
Sostengo que la divulgación de la ciencia es exitosa si, en principio, no hace más que encender el sentido de lo maravilloso. Para hacerlo, basta con proporcionar un vistazo de los hallazgos de la ciencia sin explicar detalladamente cómo fueron logrados. Es más fácil mostrar el destino que el viaje. Pero, cuando sea posible, los divulgadores deberían tratar de relatar algunos de los errores, comienzos falsos, callejones sin salida, y la confusión aparentemente

irremediable a lo largo del camino. Al menos de vez en cuando deberíamos proporcionar la evidencia y dejar que el lector sacara sus propias conclusiones. Esto convierte la asimilación obediente de nuevos conocimientos en un descubrimiento personal. Cuando uno logra el hallazgo por sí mismo, aun si es la última persona en la tierra en darse cuenta de las cosas, nunca lo olvida.

Cuando yo era joven, me inspiraba en los libros y artículos de divulgación científica de George Gamow, James Jeans, Arthur Eddington, J. B. S. Haldane, Julian Huxley, Rachel Carson y Arthur C. Clarke, todos ellos con formación científica, y muchos de ellos practicantes destacados de la ciencia. La popularidad de los libros de ciencia bien escritos, bien explicados, profundamente imaginativos, que tocan nuestros corazones así como nuestras mentes, parece mayor en los últimos veinte años que nunca antes, y el número y la diversidad de disciplinas de las que provienen los científicos que los escriben tampoco tiene precedentes. Entre los mejores divulgadores contemporáneos de la ciencia considero a Stephen Jay Gould, E. O. Wilson, Lewis Thomas y

Richard Dawkins, en biología; Stephen Weinberg, Alan Lightman y Kip Thorne en física; Roald Hoffman en química; y los primeros trabajos de Fred Hoyle en astronomía. Isaac Asimov escribió talentosamente sobre todos los temas. (Y, aunque su lectura requiere conocimientos de cálculo, me parece que el trabajo de divulgación más consistentemente emocionante, provocativo e inspirador de las últimas décadas es el volumen I de las *Lecciones de física* de Richard Feynman.) Sin embargo, los esfuerzos actuales están muy lejos de estar en proporción a las necesidades públicas. Y, desde luego, si no sabemos leer, no podemos beneficiarnos de estas obras, no importa qué tan inspiradoras resulten.

[...]Sostengo que la ciencia es una herramienta absolutamente esencial para cualquier sociedad que quiera tener alguna esperanza de sobrevivir en el nuevo siglo con sus valores fundamentales intactos. No sólo la ciencia como la practican sus profesionales, sino la ciencia entendida y adoptada por la comunidad humana entera. Y si los científicos no lo llevan a cabo, ¿quién lo hará? 



Tomado del libro *The demon-haunted world, science as a candle in the dark*, Nueva York, Random House, 1996, capítulo 19. Traducción de Martín Bonfil Olivera (publicado en español como *El mundo y sus demonios*, Planeta, 1997).

04 Cómo interpretar las estadísticas científicas



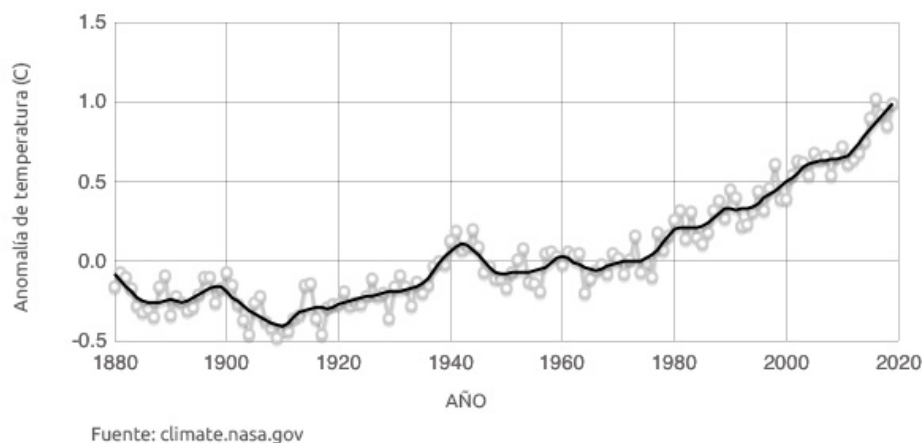
Por Elisabetta Tola

Introducción

Para la mayoría de las personas, los números transmiten una sensación de precisión, exactitud y objetividad. Sin embargo, los números no son más que una descripción de los hechos, una interpretación de los fenómenos. Son un medio para comparar cosas, ver patrones y evaluar riesgos. Pero no son la "verdad" absoluta.

Considera lo siguiente: los defensores del cambio climático señalan la temperatura media mundial registrada en 1998 de $14,61^{\circ}$ Celsius y la comparan con la de años más recientes, señalando un descenso de las temperaturas medias durante más de una década. Al fin y al cabo, en 2012, la temperatura media fue más baja, de $14,56^{\circ}$ Celsius. Según ellos, esto es una prueba de que el calentamiento global no es real.

Las cifras son ciertas, pero las conclusiones son profundamente erróneas. El año 1998 fue inusualmente caluroso debido al inesperado e intenso fenómeno de El Niño. No obstante, cuando los datos se consideran como una tendencia, la realidad del cambio climático es indiscutible.



En efecto, 1998 fue más caluroso que algunos de los años siguientes, pero la temperatura en los últimos años se ha incrementado mucho más.

Esta selección de cifras es solo un ejemplo de cómo se pueden malinterpretar los datos científicos. Otras formas incluyen:

- Cuando las cifras se proporcionan sin contexto.
- Cuando se omiten referencias en los márgenes de error, las metodologías, el tamaño de las muestras o las composiciones.
- Cuando se confunde correlación y causalidad.
- Cuando se sacan conclusiones a partir de datos insuficientes.

¿Cómo puedes evitar caer en estas trampas? Es importante que, como redactor, te acerques a ellas con el mismo escepticismo y cuidado que tienes con las palabras. ¿Qué dicen y qué no dicen las estadísticas? ¿Qué cifras son relevantes para tu historia? ¿Cuáles tienen un significado para los lectores?

La clave, por supuesto, es entender lo que significa una determinada cifra o estadística. Tomemos el concepto estadístico de "riesgo". En Estados Unidos, se producen unos 95 accidentes automovilísticos de carácter mortal al día. Dicho de otro modo, a efectos prácticos hay un 100% de probabilidades de que se produzcan docenas de accidentes automovilísticos mortales en Estados Unidos hoy en día. Pero, ¿qué riesgo corres cuando estás al volante? ¿Qué tan nervioso deberías estar? No mucho. Tu riesgo individual de morir en un accidente automovilístico es de menos de 1 entre 4000 en un año determinado.

A ti, como editor, te corresponde asegurarte de que las cifras y las estadísticas estén debidamente enmarcadas y contextualizadas para representar una realidad honesta, y que tus lectores las entiendan correctamente.

Cuestionar los datos y tener en cuenta la incertidumbre

Los números son afirmaciones. Cuando te encuentres con números y cifras de documentos científicos, instituciones o informes oficiales, debes interrogarlos como lo harías con cualquier otra afirmación. No los utilices tal cual, aunque tengas prisa. Asegúrate de comprenderlas, de dónde proceden y cómo se han obtenido. ¿Han examinado los datos otras fuentes? ¿Qué dicen esas fuentes? No puedes asumir que una cifra, solo porque esta se encuentra en un documento científico, es la *verdad*.

El escritor científico del *New York Times*, Carl Zimmer, [proporciona](#)¹ un ejemplo de esto a partir de la pandemia de coronavirus. En abril de 2020, un equipo de investigadores publicó un artículo preimpreso en el que afirmaban que la tasa de mortalidad del COVID-19 era mucho más baja de lo que algunos expertos habían estimado. Interesante, ¿verdad?

Sin embargo, tras el examen de otros estadísticos, quedó claro que las "cifras de interés periodístico" eran esencialmente el producto de un error estadístico. Al mes siguiente, se publicó en *The Lancet* un importante estudio sobre la hidroxiclороquina. El estudio afirmaba que la hidroxiclороquina era ineficaz y posiblemente peligrosa como tratamiento del COVID-19, lo que hizo que el informe recibiera una rápida atención de los medios de comunicación internacionales y que la Organización Mundial de la Salud suspendiera los ensayos clínicos del fármaco. Un centenar de científicos respondieron publicando una carta abierta, en la que expresaban su preocupación por la calidad de los datos del estudio. Cuando *The Guardian* trató de obtener información específica sobre los datos utilizados por los autores del estudio y señaló varias incoherencias que aparecían en los resultados, el autor principal no pudo responder por la exactitud de los datos y pidió a *The Lancet* una retractación. "Se trata de un ejemplo escandaloso de falta de ética en la investigación, en medio de una emergencia sanitaria mundial", declaró el director de *The Lancet* a [The Guardian](#)². (En un [podcast de FiveThirtyEight.com](#)³ se ofrece un relato completo de esta historia).

No importa que las investigaciones posteriores acabaran estableciendo la ineficacia de la hidroxiclороquina en el tratamiento del COVID-19; la equivocación puso en duda el proceso científico y afectó la confianza del público.

En contra de la creencia popular, la ciencia está lejos de ser una fuente de certeza. El método científico se basa en la duda y en el método de ensayo y error. Los datos y los hechos se definen siempre con un cierto nivel limitado de confianza. E incluso cuando muchas pruebas apuntan en una dirección, siempre debe haber espacio para el escepticismo constructivo. Lo que la ciencia aporta a nuestro pensamiento crítico no son números, datos y hechos escritos en piedra. Por el contrario, su contribución es la capacidad de investigar continuamente el mundo con una mente abierta.

Una de las lecciones más importantes que hay que aprender sobre el uso de los datos en una historia, es aceptar la incertidumbre. "Lo que muchas personas pasan por alto", dice Heather Krauss, fundadora de [Datassist](#)⁴, una empresa que apoya a los periodistas y a las organizaciones sin ánimo de lucro en el trabajo con datos, "es que, para contar una historia con datos precisos, hay que proporcionar contexto. Y eso incluye la incertidumbre".

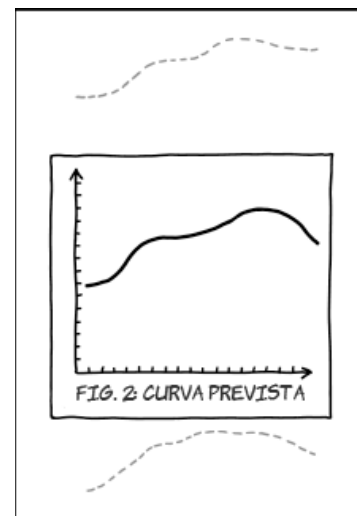
Medir la confianza

Los científicos utilizan varias herramientas⁵ para medir y tener en cuenta la incertidumbre: valores p, niveles de confianza, intervalos de confianza y márgenes de error.

El **valor p** es un concepto que incluso los expertos pueden encontrar difícil de expresar en un lenguaje sencillo. De hecho, en 2015, Christie Aschwanden, principal escritora de ciencia de FiveThirtyEight, pidió a docenas de científicos que lo explicaran de la forma más sencilla posible. El **resultado**⁶ a la vez humorístico y esclarecedor; si bien existe una clara definición técnica, explicarla en términos profanos es un reto importante. No obstante, vale la pena intentarlo, ya que el valor p es una medida de significancia comúnmente aceptada.

Esta es nuestra explicación: El valor p es un número que nos permite conocer el grado de confianza y fiabilidad — la importancia, en términos estadísticos — de un resultado científico. Para ser más específicos: si tu hipótesis no es cierta, ¿qué probabilidad hay de que obtengas un resultado que sea consistente con tu hipótesis? Cuanto menos probable sea, mejor, ya que demuestra que el resultado es estadísticamente relevante.

Como ejemplo, veamos las pruebas de medicamentos para encontrar un tratamiento para el COVID-19. Los investigadores parten de una hipótesis inicial (en este caso, que el medicamento que se está probando es eficaz), pero para ponerlo a prueba tienen que ser capaces de rechazar lo que se conoce como *hipótesis nula*. En este caso, la *hipótesis nula* es que no hay diferencia entre tomar o no el medicamento.



CONSEJO CIENTÍFICO: SI TU MODELO ES LO BASTANTE MALO, LOS INTERVALOS DE CONFIANZA CAERÁN FUERA DEL ÁREA IMPRIMIBLE.

Tomado de: xkcd.com

Mediante un ensayo de control aleatorio en dos grupos de pacientes, los investigadores comparan los resultados de los pacientes que tomaron el medicamento con los que no lo hicieron. A partir de estos datos, se calcula el valor p.

Si el valor p es alto, significa que hay una alta probabilidad de obtener un resultado que demuestre que el medicamento es eficaz, aunque no lo sea. Por el contrario, si el valor p es bajo, es poco probable que se obtenga un buen resultado con el medicamento si este no es efectivo. El valor p es siempre una cifra entre 0 y 1 y, convencionalmente, los científicos consideran que si el valor p es inferior a 0,05, los resultados son válidos; en caso contrario, no lo son.

Sin embargo, [hay quienes sostienen⁷](#) que los valores p no se consideran una medida de fiabilidad absoluta; al fin y al cabo, el valor p es solo un dato que puede utilizarse para determinar la fiabilidad. Lo más importante, [como indica un estudio de Stanford⁸](#), es que un valor p no garantiza que algo es verdad.

Además del valor p, los artículos científicos suelen ofrecer otros indicadores de incertidumbre: **márgenes de error e intervalos de confianza**. Los resultados suelen escribirse con un margen de error dentro de un nivel de confianza definido (normalmente, el 95 % o el 99 %). Los márgenes de error y los intervalos de confianza, que también se utilizan en las encuestas políticas, son una forma de transmitir el grado de confianza de un investigador en la exactitud en un rango de valores.

Por ejemplo, [un documento de investigación de 2016 de los Centros para el Control y la Prevención de Enfermedades de EE. UU.⁹](#) sobre la salud neonatal estima que, entre 2010 y 2017, el 83,8 % de los bebés nacidos en EE. UU. fueron amamantados alguna vez. La cifra tenía un margen de error de +/-1,2, con un intervalo de confianza del 95 %. Los investigadores decían que si el estudio se reprodujera 100 veces, al menos 95 de los estudios reportarían una cifra entre el 82,6 % y el 85 %.

Lo que es fundamental entender — y es el tema de la siguiente sección — es que los márgenes de error y los intervalos de confianza se basan en gran medida en el tamaño del grupo estudiado. Es importante tener esto en cuenta, ya que hay muchos casos en los que hay escasez de datos — la violencia doméstica, los delitos de odio y la injusticia racial, por ejemplo — o cuando se trata de sucesos poco frecuentes, como enfermedades genéticas raras. En esos momentos, los modelos estadísticos pueden quedarse cortos. Es entonces cuando resulta aún más importante añadir el respectivo contexto y perspectiva, para así llenar las lagunas con la ayuda de expertos.

La incertidumbre también entra en juego de otras maneras. En 2019, el periodista y diseñador visual Alberto Cairo analizó cómo se comunica la incertidumbre en el ámbito de la predicción de huracanes. En colaboración con el Centro Nacional de Huracanes, investigó cómo la mayoría de las personas interpretan los mapas de predicción de huracanes, que indican la probabilidad de tocar tierra. El resultado: no los interpretan muy bien.

Tradicionalmente, los meteorólogos utilizan un mapa de "cono de trayectoria" para señalar dónde podría tocar tierra un huracán. Pero, **como descubrió Cairo¹⁰**, las personas no entienden realmente los mapas y pueden adoptar medidas inadecuadas debido a ese malentendido. La razón, según ha averiguado Cairo, es que los lectores malinterpretan los mapas de diversas maneras, incluyendo la forma en que se representa la incertidumbre.

El cono es un rango de posibles trayectorias con una "línea central" que indica el centro probable de la tormenta a futuro. En realidad, solo el 60-70% de las veces el centro del huracán termina dentro del cono proyectado. Además, el alcance de la tormenta y su impacto devastador, es mayor de lo que indica el cono o la línea central. La ignorancia de una persona promedio sobre cómo interpretar el riesgo personal de un mapa de cono de trayectoria, conlleva a un potencial de resultados fatales. La previsión de la llegada a tierra de un huracán es un excelente ejemplo de por qué siempre hay que preguntarse, al escribir o mostrar gráficamente lo que es posible o probable: ¿Estamos transmitiendo con precisión la incertidumbre?

Cairo sugiere que se utilicen mapas alternativos, como un "mapa de calor" del impacto, que demuestran mejor el verdadero riesgo. En otros casos, los redactores deben describir mejor la importancia de la incertidumbre y su aspecto real. Un buen ejemplo de este concepto se encuentra en **un artículo de *Scientific American* sobre la importancia de la estadística¹¹**, relacionado con el bosón de Higgs, una partícula elemental de la física. Aunque está dedicado a la idea de la incertidumbre, el artículo es capaz de expresar en términos sencillos algunas matemáticas complicadas:

Como ejemplo específico, hay algo menos del uno por ciento de posibilidades de lanzar una moneda normal 100 veces y obtener al menos 66 caras. Pero si mil personas lanzan 100 veces monedas idénticas, es probable que unas pocas obtengan al menos 66 caras cada una. Uno de estos eventos, por sí solo, no debería interpretarse como una prueba de que las monedas de alguna manera estaban manipuladas.



Varios días antes de que el huracán Katrina tocara tierra, este pronóstico mostraba el cono de predicción. Aunque Nueva Orleans está en el cono, pocos espectadores de este gráfico pensarían que estaba en la zona prevista una vez llegara a tierra. Fuente: NOAA

Confiar únicamente en los números para que te hagan el trabajo, les fallará tanto a ti y como a tus lectores. La clave es explicar los conceptos subyacentes.

Tamaño de la muestra

Uno de los factores que impulsan la confianza en la estadística es el tamaño del grupo de estudio, también conocido como tamaño de la muestra. Rara vez disponemos de datos perfectos sobre una población en su totalidad. El censo de los EE. UU. trata de hacer eso precisamente: contar cada década el número de habitantes en el país.

En términos de datos, se puede pensar en ello de la siguiente manera: imagina que hay 100 personas en una isla, y te gustaría determinar su altura media. Si mides a las 100 personas, no tendrías ningún margen de error y tendrías un nivel de confianza del 100%. Por lo tanto, tu cálculo sería perfecto.

Pero estas situaciones son poco frecuentes. Son más comunes los grupos de muestra que sustituyen a toda una población. En nuestro ejemplo de la isla, podríamos tomar una muestra de las alturas de 20 personas elegidas al azar para estimar la altura media de la población. Probablemente nos acercáramos, pero no coincidiríamos, con la altura media real. Cuanto mayor sea el tamaño de la muestra, mejor será la estimación. Cuanto más pequeña sea la muestra, mayor será el margen de error y menor el intervalo de confianza.

Los experimentos científicos suelen realizarse mediante un *muestreo aleatorio*. En términos estadísticos, una muestra aleatoria está diseñada para representar a toda la población de estudio. Puede que este concepto te resulte familiar en las encuestas electorales. Un enfoque similar se utiliza en las pruebas de drogas o en la descripción de las características biológicas de un subgrupo de individuos.

Observaciones vs. experimentos

Los estudios observacionales son estudios en los que las variables y los datos no están controlados y no son recabados por los científicos que realizan la prueba. Normalmente, en estos estudios se aplican modelos estadísticos a datos recopilados por otras entidades públicas o privadas. Su valor radica en que pueden dar una visión del mundo real, al no ocurrir en un entorno controlado.

Los estudios experimentales proporcionan datos recopilados directamente por los científicos que los analizan. Pueden ser aleatorios o no, con o sin un grupo de control. Por ejemplo, un ensayo clínico habitual suele ser aleatorio y controlado. Los estudios experimentales facilitan a los investigadores el control de las variables y reducen la incertidumbre estadística, conocida como "ruido".

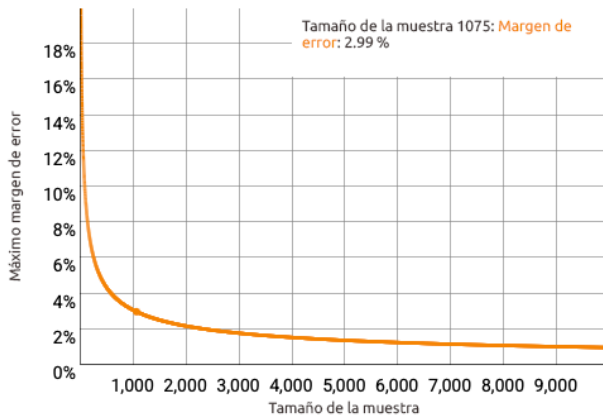
Por el contrario, las muestras no aleatorias, como los grupos de personas voluntarias, no dicen nada sobre la población en su totalidad. Entonces, los estudios que constan de tales muestras deben considerarse con cierto escepticismo.

Pero a veces incluso las muestras mejor diseñadas pueden resultar sesgadas. Esto ocurre cuando muchas personas no responden o un subgrupo informa sus datos de manera inexacta. Por ejemplo, un grupo de personas que **informan sobre su consumo de alimentos**¹² podría ser una muestra aleatoria y su autoinforme ser defectuoso hasta el punto de hacer que los datos sean irrelevantes.



Tomado de: xkcd.com

El tamaño también importa. Las muestras pequeñas son más fáciles de sesgar por los valores atípicos y es más probable que les afecten los errores aleatorios.



Observa en [este gráfico interactivo](#) cómo el margen de error disminuye con el tamaño de la muestra.

Además, como escribe John Allen Paulos en [A Mathematician Reads the Newspaper](#)¹³, "lo fundamental de una muestra aleatoria es su tamaño absoluto, no el porcentaje de la población". Puede parecer contradictorio, pero una muestra aleatoria de 500 personas tomada de la población total de EE. UU. suele ser mucho más predictiva — es decir, tiene un menor margen de error — que una muestra aleatoria de 50, tomada de una población de 2500. Diferentes variables, incluyendo el tamaño de la población, determinan la fiabilidad de una muestra. Por ejemplo, como regla general, el Instituto Nacional de Estadística de Italia utiliza muestras de entre 1500 y 2000 personas para realizar encuestas periódicas de poblaciones, independientemente de la población total.

Pero el tamaño de la muestra tiene una advertencia: como [señala](#)¹⁴ el médico y escritor científico Ben Goldacre en *The Guardian*, los efectos pequeños (pero significativos) son difíciles de medir en poblaciones con muestras pequeñas.

Evaluar el riesgo

Muchos artículos científicos, especialmente los de salud, intentan cuantificar el riesgo. A menudo, se hace en forma de relación o porcentaje, como el riesgo estimado por la Asociación Americana del Corazón de que un adulto en Estados Unidos contraiga una enfermedad cardiovascular a lo largo de su vida es: 48 por ciento. Un porcentaje enorme, ¿verdad?

En primer lugar, hay que entender qué son las enfermedades cardiovasculares. No solo se incluyen las enfermedades coronarias, la insuficiencia cardíaca y los

accidentes cerebrovasculares, sino también la hipertensión arterial. Si se elimina la última categoría, el riesgo se reduce hasta el 9 por ciento.

Riesgo absoluto

Esa cifra del 9 por ciento es una estimación del *riesgo absoluto*, o el porcentaje de personas de toda la población estadounidense que tienen probabilidades de desarrollar la enfermedad. No tiene en cuenta los factores de riesgo individuales, como la genética, el tabaquismo o el ejercicio regular, a menos que la propia población esté limitada por estos factores. (En otras palabras, si toda la población estudiada está compuesta por fumadores, el riesgo absoluto es para las personas que tienen ese comportamiento).

Por supuesto, la realidad es que cada persona tiene un riesgo diferente, dependiendo de varios factores. Algunos son independientes del comportamiento del individuo. La edad, los antecedentes familiares y el sexo están fuera del control de una persona. Otros factores que probablemente estén bajo el control de una persona son la nutrición, el ejercicio, el tabaquismo y el consumo de alcohol. También hay que tener en cuenta las probabilidades condicionales: el riesgo de sufrir un infarto puede ser mayor si ya se ha sufrido un infarto anteriormente.

Riesgo relativo

Aquí es donde entran en juego el *riesgo relativo* y la *razón de probabilidades*. Ambas cifras son proporciones, pero son dos medidas diferentes que se aplican a distintos tipos de estudios. Los investigadores calculan el riesgo relativo cuando realizan un estudio prospectivo; es decir, un "estudio de cohortes", en el que se observa a las personas y se recopilan datos que las siguen en tiempo real. La razón de probabilidades se calcula en un estudio retrospectivo, también llamado "estudio de casos y controles", en el que los investigadores comparan dos muestras seleccionadas de personas, por ejemplo, un grupo de personas con VIH y otro sin la enfermedad.

El *riesgo relativo* (a veces abreviado como "RR" y también llamado razón de riesgo) se calcula dividiendo el riesgo de la población expuesta a un determinado factor (tabaquismo; dieta poco saludable; estilo de vida sedentario) por el riesgo de la población no expuesta. Para cada cohorte, los investigadores miden la incidencia de la enfermedad calculando la proporción de personas afectadas con respecto al número total de personas de esa cohorte. Por último, dividen la incidencia en la población expuesta (por ejemplo, los fumadores) con la de la población no expuesta (los no fumadores).

Como ejemplo, hagamos algunos cálculos utilizando los datos de [un artículo](#)¹⁵ de la revista médica británica *Heart*. Según los datos, la tasa de mortalidad por cáncer de pulmón para los fumadores es de 209 por cada 100.000 personas; para los no fumadores, es de 14 por cada 100.000. Dividiendo 209 entre 14 se obtiene 15 (14,93, para ser exactos), que es el aumento del riesgo relativo de morir de cáncer de pulmón si fumaras.

La cifra "15" puede expresarse de varias maneras: 15 veces, por un factor de 15, o un aumento de 15 veces. Todas significan lo mismo. También se puede convertir en un aumento porcentual. Para ello, usa la fórmula "NOO", que se describe más adelante en este capítulo. Si lo haces, sabrás que fumar aumenta el riesgo de morir de cáncer de pulmón en un 1400%. Sin embargo, estos porcentajes son difíciles de entender para la mayoría de las personas, por lo que recomendamos mantener la expresión "15 veces", para mayor claridad.

Un riesgo relativo inferior a "1" significa que existe una relación inversa entre el factor y el resultado. Por ejemplo, si el grupo de fumadores tuviera *menos* muertes por cáncer de pulmón, se obtendría un riesgo relativo inferior a 1, lo que sugeriría que fumar es una medida profiláctica contra la muerte por cáncer de pulmón. Por lo tanto, se esperaría ver cifras por debajo de 1 en las investigaciones de medicamentos eficaces.

Cuando se reporta el riesgo relativo, es importante mencionar siempre el riesgo absoluto original. Si la incidencia original de una enfermedad es muy baja, informar solo sobre un riesgo relativo elevado conducirá a una exageración de los efectos de ese factor de riesgo. Por ejemplo, si un estudio informa de que en un grupo de 100.000 personas solamente muere una persona de cáncer de pulmón, mientras que en el grupo de comparación mueren dos personas de cada 100.000, eso es una duplicación del riesgo relativo, no obstante, el riesgo absoluto sigue siendo bastante pequeño.

Por otro lado, cuando la enfermedad tiene una incidencia muy alta, incluso una pequeña diferencia en el riesgo relativo, equivale a un elevado número de personas que podrían ser afectadas.

Razón de probabilidades

A veces, un estudio presenta una razón de probabilidades en lugar de un riesgo relativo. Esto es común en un "estudio de casos y controles", en el que los investigadores identifican a las personas con una enfermedad o problema de interés y comparan ese grupo con otro grupo que se le parece, pero sin la enfermedad o problema. En un estudio de este tipo, los controles no permiten determinar el riesgo relativo. En su lugar, los investigadores solo pueden describir las probabilidades.

Por ejemplo, imagina que los investigadores analizan dos grupos de exsocorristas. Un grupo contrajo cáncer de piel, mientras que el otro grupo no lo hizo. Al observar el uso del protector solar de óxido de zinc por parte de los socorristas, los investigadores pueden calcular sus probabilidades de desarrollar cáncer de piel si utilizaron o no el óxido de zinc. Sin embargo, los investigadores no podrían calcular una reducción generalizada del riesgo que el uso de óxido de zinc proporcionaría en una población aleatoria.

Para profundizar en este ejemplo de razón de probabilidades, supongamos que el grupo de "caso de estudio" de socorristas que desarrollaron cáncer de piel, está formado por 100 personas, de las cuales 10 utilizaron óxido de zinc. El grupo de "control" está formado por 400 socorristas, de los cuales 300 utilizaron óxido de zinc. Para calcular la razón de probabilidades que te indicará si la utilización de óxido de zinc previene el cáncer de piel, primero tendrías que mirar el número de socorristas que usaron óxido de zinc, frente a los que no lo hicieron en el grupo de socorristas del caso estudio que desarrollaron cáncer. A continuación, se divide el resultado por la proporción entre los que usaron óxido de zinc y los que no lo hicieron en el grupo de control.

Entonces, según el cálculo, las personas (10) que utilizaron de óxido de zinc, divididos por los usuarios que no utilizan óxido de zinc en el grupo de cáncer de caso de estudio (90), nos da un resultado de 0,11 (o 1 de cada 9). A continuación, divide los consumidores de óxido de zinc entre los no consumidores de óxido de zinc del grupo sin cáncer (300/100), en donde obtienes 3. Por último, divide la primera cifra (0,11) entre la segunda (3) para obtener 0,037, que es tu razón de probabilidades. Podrías decir: "Los socorristas que utilizan óxido de zinc tienen 0,037 veces más probabilidades de padecer cáncer de piel que los socorristas que no utilizan óxido de zinc", pero eso es difícil de entender. En su lugar, divide la proporción entre 1 y dale la vuelta a la frase para decir de forma más sencilla: "Los socorristas que no utilizan óxido de zinc tienen 27 veces más probabilidades de contraer cáncer de piel que los socorristas que sí utilizan óxido de zinc".

Esto es diferente a una evaluación de riesgos, ya que no conocemos el riesgo general de contraer cáncer de piel, ni hubo controles para otros factores. Todo lo que sabemos es que, basándonos en nuestro caso de estudio hipotético, las probabilidades de desarrollar cáncer de piel son 27 veces mayores para los socorristas que no se aplican óxido de zinc.

En resumen, cuando calculamos el riesgo, nos referimos al número de casos dentro de una población. Pero cuando calculamos las probabilidades, estamos comparando los casos con los que no lo son.

Media, mediana y moda

A menudo, cuando intentamos transmitir el significado de los datos, utilizamos características descriptivas representativas como **la media, la mediana y la moda**. Cualquiera de las tres se denomina a veces media, aunque la mayoría de las veces "media" se asocia con "promedio".

Imagina que tienes un avión lleno de pasajeros y mides la altura de cada individuo. Si representas esa información en un gráfico, probablemente verás unos cuantos puntos en el extremo inferior, otros tantos en el extremo superior y un montón de puntos agrupados en el centro. Esa es tu curva de campana estándar, también llamada "distribución normal". En este caso, la media, la mediana y la moda tienden a ser similares, si no es que llegan a coincidir, en el punto medio del gráfico.

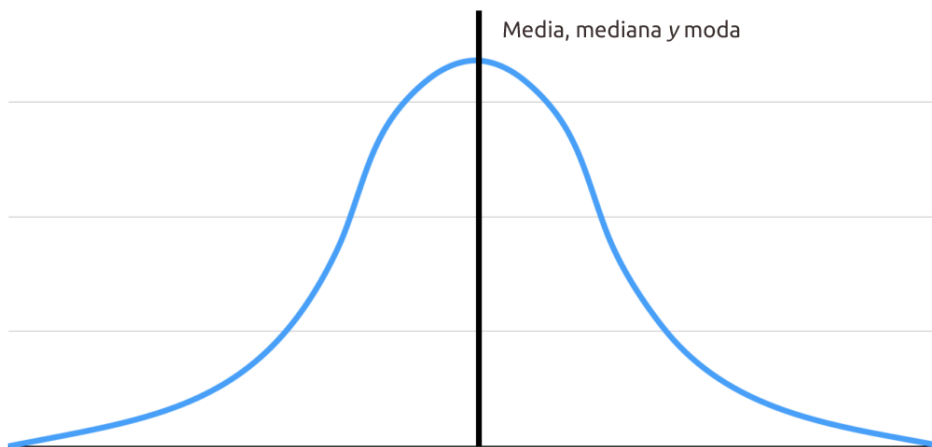
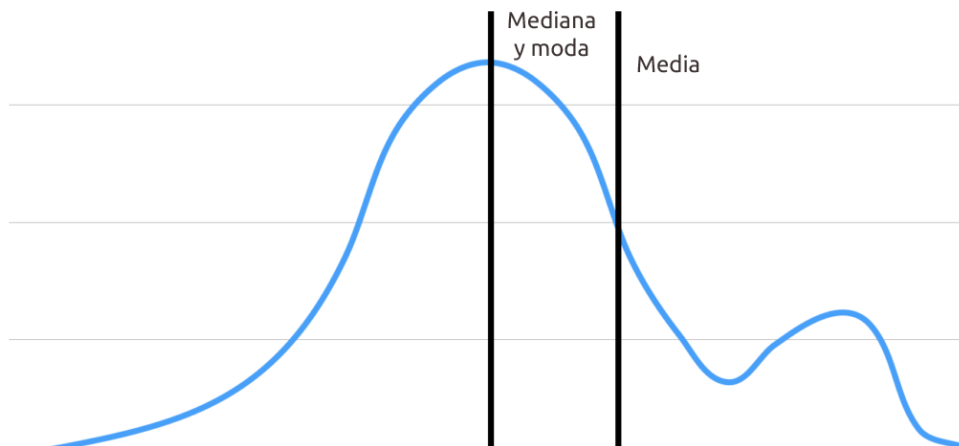


Gráfico característico de una distribución "normal", también conocida como "curva de campana".

Las características biológicas, como la altura y el peso, suelen comportarse así. Para calcular la media, basta con sumar todos los valores del conjunto de datos y dividirlos por el número de elementos. Si la muestra es lo suficientemente grande, es probable que los valores se distribuyan de forma simétrica alrededor de la media.

Asociada a la media está la "desviación estándar". Este término se refiere a lo lejos que están los puntos de información, en promedio, de la media. En otras palabras, ¿es la campana angosta (los datos están agrupados juntos)? ¿O es voluminosa (los datos están dispersos)? Las desviaciones estándar pequeñas sugieren que los datos son más uniformes.

Sin embargo, en una muestra pequeña, incluso las características biológicas pueden distribirse de forma poco normal. Por ejemplo, ¿qué pasaría si un equipo de baloncesto profesional estuviera a bordo de nuestro vuelo imaginario? La presencia de esos valores atípicos altos afectaría a la distribución de los valores. Ahora nuestra curva de campana tiene un relieve a la derecha, y la media se desplaza hacia la derecha.



La distribución "normal" cambia cuando aparece un grupo de valores atípicos en el conjunto de datos.

En estos casos, puede ser mejor utilizar la **mediana**, que es el número medio de todos nuestros valores, si los ordenamos de menor a mayor.

A menudo se utiliza la palabra "mediana" para hablar de los ingresos, especialmente en poblaciones con gran desigualdad. La presencia de unos pocos multimillonarios desplazaría la media muy por encima del nivel de ingresos que experimentan la mayoría de los ciudadanos. Por ello, la mediana indica mejor el punto medio de una distribución: la mitad de las personas ganan más que la mediana y la otra mitad menos.

Otra forma de describir los datos es a través de la moda, que proporciona el valor más frecuente o más común. Como ejemplo, salgamos de nuestro plano imaginario y dirijámonos en cambio a las carreteras.

En 2018, según la Administración Nacional de Seguridad del Tráfico en Carreteras, esta fue la distribución de los accidentes mortales por hora en un día:

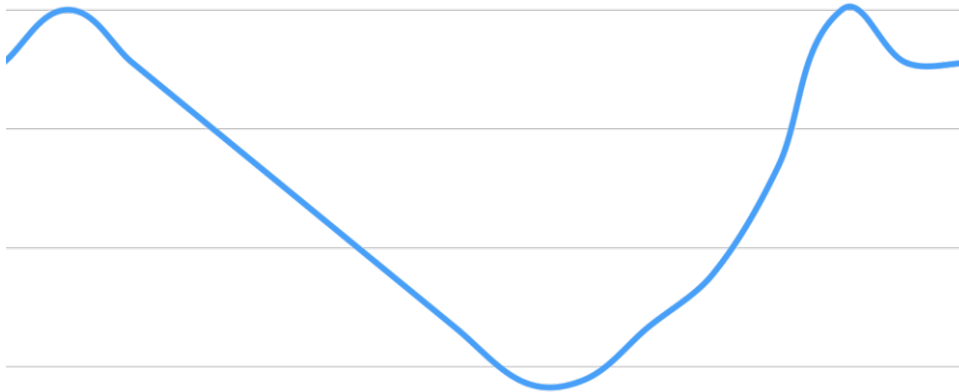
Medianoche - 3:59 am	4574
4:00 - 7:59 am	4388
8:00 - 11:59 am	4154
Mediodía - 3:59 pm	5943
4:00 - 7:59 pm	7329
8:00 - 11:59 pm	7022

Accidentes mortales de tráfico en Estados Unidos en 2018, por hora en un día. Fuente: Administración Nacional de Seguridad del Tráfico en las Carreteras

Si tuviéramos la hora exacta de cada accidente, se podría calcular una hora media de los accidentes: sería en algún momento de las primeras horas de la tarde. Asimismo, se podría determinar la hora mediana, que también sería en las primeras horas de la tarde. Sin embargo, ninguna de las dos opciones nos diría mucho sobre las horas del día a las que los agentes de seguridad pública deberían acudir para reducir los accidentes mortales. La hora más habitual, o la **moda**, de los accidentes mortales es en la tarde.

Por lo tanto, desde el punto de vista de la seguridad pública, los agentes deberían centrarse en mejorar la seguridad por la noche, tal vez a través de la aplicación de la ley, la iluminación, la señalización o las marcas viales, con el fin de tener el mayor impacto en los accidentes mortales.

Algunos grupos de información no tienen un solo patrón. Pueden existir dos puntos que representen los datos más frecuentes. Esto se conoce como distribución bimodal. (Más de dos patrones se llamaría "distribución multimodal").



En una distribución bimodal o multimodal, el punto más alto se repite. Este gráfico muestra el índice de accidentes de tráfico según la edad del conductor, siendo los conductores más jóvenes y los de más edad los que tienen más accidentes que el resto de grupos.

Un ejemplo de esto lo encontramos en las edades más comunes de los conductores implicados en accidentes mortales. Los conductores nuevos, jóvenes y mayores, sobrepasan a los conductores de entre 24 y 65 años en términos de accidentes. Por lo tanto, las políticas destinadas a reducir los accidentes mortales podrían centrarse en estos dos grupos de edad.

Correlaciones, causas y datos a lo largo del tiempo

No es nada raro ver historias que dicen algo como: "Las prescripciones de antidepresivos han aumentado en los últimos 20 años" o "El número de personas que no tienen acceso al agua potable ha disminuido en los últimos 10 años". Sin embargo, a lo largo del tiempo la comparación de datos exige un estudio cuidadoso. El significado de unas cifras determinadas puede variar en función del momento en que se midan. El dinero es un clásico ejemplo: las comparaciones financieras deben tener en cuenta la inflación; de lo contrario, es imposible sacar conclusiones significativas.

Para calcular el impacto de la inflación, utiliza la [calculadora de inflación del Índice de Precios al Consumidor de la Oficina de Estadísticas Laborales de los Estados Unidos](#)¹⁶.

El problema, por supuesto, va mucho más allá del ámbito financiero. Hay muchos otros factores que pueden determinar la calidad de los datos como se observa a través del tiempo. Otro ejemplo: Las capacidades de diagnóstico han mejorado a lo largo de los años en una amplia variedad de problemas de salud. Informar sobre el aumento o la disminución de la prevalencia de una enfermedad en comparación con una época en la que los datos no estaban disponibles o se medían con estándares diferentes no tiene mucho sentido.

Cuando vemos que las cifras cambian con el tiempo, la pregunta natural que nos hacemos es: "¿Por qué?".

Para responder a esa pregunta, los científicos suelen utilizar una serie de pruebas estadísticas denominadas análisis de regresión para ver si puede establecer una relación entre dos variables. La forma más habitual de expresar esa correlación es utilizar el índice r , que va de -1 a $+1$. Hay correlaciones negativas, en las que una variable crece mientras la otra disminuye, y positivas, en las que ambas variables se mueven en la misma dirección.

Por ejemplo, existe una correlación negativa entre el aumento de las temperaturas y el uso de combustible para calefacción. Asimismo, existe una correlación positiva entre el aumento de las temperaturas y el uso de la electricidad (para el aire acondicionado). Los valores de las correlaciones débiles se acercan a cero; las correlaciones fuertes se acercan a los extremos de -1 o $+1$.

Sin embargo, el hecho de que dos cosas estén correlacionadas no significa que tengan que ver entre sí. Tyler Vigen presenta una serie de lo que él llama **correlaciones espurias**¹⁷ en su sitio web y en su libro del mismo nombre. Muestra datos reales que crean correlaciones absurdamente falsas, como la estrecha correlación de la tasa de divorcios en Maine y el consumo de margarina por habitante.

Las gráficas de Vigen demuestran dos errores comunes en las correlaciones. El primero es la tendencia a sacar conclusiones sobre los individuos a partir de un conjunto de datos que describen una población, algo conocido como falacia ecológica, que la estadística Heather Krause explica en un [video de YouTube](#)¹⁸.



Tomado de: xkcd.com

Como ejemplo, es generalmente cierto que existe una correlación entre el nivel de ingresos y la esperanza de vida en la población general. Sin embargo, no es cierto que las personas ricas vivirán más tiempo que las pobres, ni que una persona muy vieja y sana también deba ser rica.

El segundo error más común es interpretar erróneamente una correlación como causalidad. Puede haber muchas explicaciones detrás de la razón por la que las variables están relacionadas. Por ejemplo, existe una relación de causalidad entre el aumento de las temperaturas y el incremento del uso de combustible para calefacción. Pero *no* hay causalidad entre la disminución del uso del combustible para calefacción y el aumento del uso de las piscinas. En este último caso, una tercera variable — el aumento de las temperaturas — explica la aparente relación.

Para establecer una relación causal, los científicos se adentran en la estadística y realizan complejas pruebas con sus datos. Para confirmar un efecto causal y descartar todas las demás explicaciones posibles, deben elaborar estudios experimentales con diseños aleatorios y grupos de control. Esto es especialmente importante en campos como la epistemología ambiental, en la que los investigadores quieren saber si un contaminante concreto puede ser la causa de una enfermedad. Es complicado encontrar una relación única entre la presencia de una sustancia y su impacto en la salud de la población. Vivimos en un entorno complejo y hay muchos factores en juego: estilo de vida, estado nutricional, afecciones previas y predisposición genética, entre otros.

Ayudar a los lectores a entender las cifras

Las cifras por sí solas pueden ser impactantes, pero a menudo carecen del contexto necesario. ¿Qué significa, por ejemplo, que **cada año mueran en Estados Unidos unas 650.000 personas por enfermedades del corazón**¹⁹? Es una cifra importante, y son muchas personas. De hecho, es la principal causa de muerte en EE. UU., representando aproximadamente una cuarta parte de todas las muertes. Pero, ¿debería ser alarmante para nosotros? Para contextualizar, es útil hacer una comparación.

Quizás debas fijarte en el Reino Unido, que tiene una dieta y un nivel de vida similares a los de EE. UU. En el Reino Unido, unas **170.000 personas mueren cada año por enfermedades cardiovasculares**²⁰. Estados Unidos parece estar bastante mal en esta comparación. Pero el porcentaje de todas las muertes en el Reino Unido atribuibles a enfermedades del corazón, alrededor del 27 %, es un poco más alto que el de Estados Unidos.

Por supuesto, esto está en función de las diferencias en el tamaño total de la población. Estados Unidos tiene unos 330 millones de habitantes, y el Reino Unido apenas llega a los 68 millones. Así que, en términos de porcentaje de la población total, alrededor de 2 personas de cada mil mueren de enfermedades cardiovasculares en los EE. UU., en comparación con 2,5 en el Reino Unido. Resulta que, en realidad, Estados Unidos está un poco mejor en ese aspecto si se tiene en cuenta la población de cada país.

Como editor, la pregunta "¿Entonces qué?" es fundamental a la hora de analizar los datos. ¿Qué significa esta estadística? ¿Qué dice? ¿Es significativa, sorprendente, esclarecedora? No caigas en la trampa de pensar que los números son significativos en sí mismos. Es necesario explicarlos, describirlos y contextualizarlos.

Porcentajes, tasas y ratios

Un porcentaje no es más que una relación específica: una comparativa con 100. Pero eso no significa que sea el único que hay que utilizar. A veces, pueden ser útiles otros denominadores, como 10 o, como en el ejemplo de las tasas de mortalidad anterior, 1000. O incluso 1, como en las comparaciones per cápita.

Por ejemplo, si quieres calcular y comparar el consumo de energía de Nueva York y Texas, se puede ver el consumo total de cada estado. Tanto Nueva York como Texas son grandes y están poblados. Pero lo que realmente importa es el número de personas que utilizan esa energía. La mejor manera de comparar es dividir el consumo total de energía de cada población, para obtener una cifra *per cápita*. Así se obtendría un resultado comparable y más fácil de relacionar con los lectores.

Texas es uno de los mayores consumidores de energía, con **470,2 BTU por persona²¹**, mientras que Nueva York tiene la tasa per cápita más baja, **188,6 BTU por persona²²**, menos de la mitad que Texas.

Asegúrate de que su comparación es una cifra que tiene sentido intuitivo. Decir que la tasa de mortalidad cardiovascular en EE. UU. es de 0.0028 per cápita puede ser cierto, pero es poco probable que la mayoría lo entienda.

Un error común se produce cuando la gente trata de comparar porcentajes. Si comparas la expansión de la tierra agrícola brasileña de 1965 a 2015, verás que creció del 20% al 28%. Una forma de describir esa diferencia es como un aumento de 8 *puntos porcentuales*. Ten en cuenta, sin embargo, que esto no es lo mismo que un "cambio porcentual". Si quisieras expresar el cambio de esa manera, tendrías que seguir la regla "NOO" (lo nuevo menos lo anterior, con el

resultado dividido por lo anterior). En ese caso, calcularías $28 - 20 = 8$. Luego calcula $8 \div 20 = 0,4$. Eso produce el porcentaje de aumento: $0,4 * 100 = 40\%$.

Traducir números grandes y pequeños

Las analogías y las metáforas son otras herramientas que hacen que los números sean relacionables. Por ejemplo, nuestro ADN es un mensaje compuesto por cuatro ácidos, que representamos con las letras A, C, G y T. En total, un hilo de ADN es una cadena de 6500 millones de pares de bases de ácidos. Así que, escrito, serían 6500 millones de caracteres A, C, G y T. Es un número bastante grande, difícil de entender para la mayoría de las personas. Para contextualizar, se podría comparar una cadena de ADN con la novela *La guerra y la paz*. La novela contiene aproximadamente 2,5 millones de caracteres, lo que significa que cada una de nuestras células contiene un mensaje que mide lo mismo que 2600 ejemplares de la obra maestra de Tolstoi.

Para reflexionar

Millón, mil millones y billón suenan como si no fueran tan diferentes entre sí. Pero son *significativamente* diferentes. Por ejemplo:

Un **millón** de segundos equivale a unos 12 días.

Mil millones de segundos equivalen a unos 32 años.

Un **billón** de segundos duraría casi 31.710 años.

Los números pequeños también pueden ser difíciles de comprender. Por ejemplo, las cifras que se utilizan cuando se habla del cambio climático, como la cantidad de dióxido de carbono en la atmósfera (400 partes por millón) o el objetivo de mantener el cambio de temperatura a 2 grados centígrados. En ambos casos, el contexto histórico puede aportar un mayor significado que las pequeñas cifras, que por sí mismas no pueden. Este fragmento de un [artículo de la revista Time²³](#) lo explica mejor:

Vivimos rodeados en un invernadero atmosférico de gases y vapor de agua que ha mantenido las condiciones de la vida durante cientos de millones de años; el CO2 forma parte de esa mezcla. Pero en los últimos tres millones de años nuestro sistema de invernadero ha sido muy inestable. El registro de CO2 atrapado en el hielo polar revela

que en los últimos 800.000 años, durante los dramáticos cambios entre la época glacial y los períodos cálidos, el CO2 ha oscilado entre 180 y 280 ppm. En el último calentamiento acelerado del último período glacial, el CO2 pasó a 260 ppm, y luego osciló en torno a 275 ppm. Desde entonces, durante unos 9000 años, nuestro clima se ha mantenido relativamente estable. Surgió la agricultura, las civilizaciones y los estados, y la **población mundial**²⁴ creció de varios millones al final de la última Era de Hielo a unos 1200 millones en 1850.

Enmarcar las cifras

Repite conmigo: *Los números no cuentan historias por sí solos*. La elección de palabras y cifras que elijas destacar sirven para dar una perspectiva particular a la información. Esto se conoce como efecto marco, descrito minuciosamente por el premio Nobel Daniel Kahneman en su libro *Pensar rápido, pensar despacio*, y tiene un impacto considerable en los lectores y oyentes; nuestra toma de decisiones depende enormemente de la forma en que se presentan los hechos. Por ejemplo, decir que un tipo de cirugía tiene un 70% de probabilidades de éxito, suena muy diferente a decir que es probable que tres personas de cada 10 mueran durante dicha operación. Unos ligeros cambios en la redacción y el marco, pueden hacer que las personas interpreten la información de forma muy diferente.

Los datos no solamente son objetivos intrínsecos, sino que pueden ser francamente subjetivos en determinadas circunstancias, como cuando se omiten estadísticas o contextos cruciales. Es esencial describir la fuente de los datos y los conflictos de intereses que puedan estar escondidos detrás de las hojas de cálculo. ¿Quién pagó la investigación? ¿Cómo se recogieron los datos? ¿Qué datos se excluyeron? El poder de las cifras para influir en la opinión pública es muy notorio. Basta con ver a las empresas tabacaleras que facilitaron investigaciones para minimizar los riesgos — y promocionar los supuestos beneficios — del tabaco para la salud de las personas. Ahora sabemos que esos estudios eran erróneos, pero llenaron las noticias y los anuncios durante años. Y por supuesto, establecieron el modelo a seguir por otras empresas e industrias.

¿De dónde proviene la información y los estudios que llegan a tu mesa?

- **Las preimpresiones**, son documentos con información y resultados que aún no han sido sometidos al proceso de revisión por pares; se publican en "servidores de preimpresiones" y no en revistas. Plantean un reto especial cuando se utilizan como fuente de información pública. Entonces, ¿por

qué existen? Una de las razones es que ofrecen la posibilidad de que otros científicos e investigadores proporcionen comentarios tempranos que puedan utilizarse para mejorar un estudio o documento.

- **Los estudios primarios**, son los datos y resultados de una serie de experimentos originales, derivados de una hipótesis y generados mediante una metodología muy definida.
- **El metaanálisis y las revisiones sistemáticas**, son documentos que **recogen, analizan, revisan y discuten otros artículos científicos**²⁵ sobre un tema determinado, en lugar de reflejar una nueva investigación. Estos documentos son útiles para proporcionar una visión general de la investigación y los conocimientos en un campo específico.
- **Los sistemas de vigilancia y seguimiento**, son recopilaciones de datos realizadas por organismos públicos de investigación de manera continua para hacer un seguimiento de los avances en un área determinada. Pueden ser la vigilancia epidemiológica, los rastreadores de emisiones de CO2 o las redes de seguimiento de la actividad sísmica o volcánica. A menudo, estos datos se hacen públicos a través de plataformas abiertas.
- **Los informes periódicos**, son reportes e informes anuales, mensuales o de otro tipo, publicados por organizaciones internacionales. A menudo hacen referencia y sintetizan otros trabajos de investigación, estudios e informes.

Consejos para escribir sobre números

A nuestra profesión debería avergonzarle que muchos periodistas se enorgullecen de no entender nada de números o matemáticas. Sin embargo, como señala Sarah Cohen en el imprescindible libro de referencia **Numbers in the Newsroom**²⁶, "El miedo a los números convierte a un reportero en alguien que solo puede hacer una parte del trabajo". Estos son algunos consejos para que los números y las matemáticas te sirvan como periodista:

- **Conoce tu universo.** Sarah Cohen y Jennifer La Fleur, dos premiadas periodistas de datos, dicen que el paso más importante es conocer algunas de las cifras básicas de tu especialidad. Por ejemplo, si estás editando historias sobre el cambio climático, debes conocer la temperatura promedio de tu región, la concentración actual de dióxido de carbono en la atmósfera (416 partes por millón) y el año más cálido registrado (2019, al momento de escribir esto, a mediados de 2020), entre otras estadísticas

básicas. Los que editan la cobertura de COVID-19 deberían haber interiorizado información básica como la tasa de mortalidad (estimada en un 1 %) y su comparación con otras enfermedades, como la gripe (0,1 %), el SARS (11 %) y el MERS (35%), así como las tasas de transmisión (o R_0), que para el actual coronavirus se estima entre 2 y 2,5. Por supuesto, el R_0 depende en gran medida de factores ambientales, como el hecho de que la gente lleve o no mascarillas y practique el distanciamiento social. La cuestión es que conocer las estadísticas y los datos básicos te permitirá detectar errores y poner un punto más preciso a las historias.

- **Utiliza las cifras grandes y pequeñas con cuidado.** La gente es pésima a la hora de entender números muy grandes o muy pequeños. Es difícil poner en contexto un "billón" o entender lo diminuto que es un virus (menos de 100 nanómetros). Así que es mejor relacionar los números con algo cercano a las experiencias de la gente. Afirmar que 800 millones de personas no comen lo suficiente cada día, puede entenderse mejor de esta manera: "más del doble de la población de Estados Unidos se acuesta con hambre cada día". Así mismo, encuentra formas creativas de visualizar lo más pequeño. Por ejemplo, al describir el tamaño de las bacterias, puedes decir "unas 150.000 bacterias caben en la punta de un cabello".
- **Cuidado con la falsa precisión.** Imagina que un científico dice que el esqueleto de un dinosaurio tiene 100 millones de años. Cinco años después, ¿dirías que ese esqueleto tiene 100.000.005 años? Por supuesto que no. Las cifras de "cien millones", como muchas medidas de la naturaleza, es una estimación con cierta incertidumbre inherente, no una cifra precisa. Utilizar cifras "exactas" con decimales hasta las centésimas, no suele ser la mejor manera de presentar los números. A menudo, es mejor utilizar una buena aproximación dentro del orden de magnitud apropiado. En lugar de "23,7%", di "un poco menos que uno de cada cuatro". Del mismo modo, el 44% se convierte en "más de 4 de cada 10".
- **No abusos de los números en un párrafo.** Lo ideal es escribir frases concisas y párrafos cortos, describiendo el sentido de los datos sin mostrar necesariamente los números, a menos que sean necesarios. Sarah Cohen sugiere reducir el número de dígitos para mejorar el empleo de aquellos que sí utilices. Como regla general, dice que vale la pena reducir el número de dígitos en un párrafo a ocho. Por ejemplo, decir "Más de 200.000 personas murieron en Estados Unidos por el coronavirus en 2020" ya tiene 10 dígitos. Pero es una regla general, no una ley. Solo hay que ser prudente y utilizar los dígitos cuando sea necesario explicar algunas ideas.

Lecturas y recursos adicionales

Libros

- Darrell Huff, *How to Lie With Statistics* (1957), Penguin Books. La mejor y más inmediata aproximación al mundo de los números en las noticias.
- John Allen Paulos, *A Mathematician Reads the Newspaper* (1995), Anchor Books. Informativo, lleno de ejemplos y fácil de leer.
- Hans Rosling, *Factfulness* (2018), Flatiron Books.
- Alberto Cairo, *Cómo mienten los gráficos* (2019), W.W. Norton and Company. Un útil enfoque sobre cómo deben dibujarse y utilizarse los gráficos en las noticias.
- Sarah Cohen. *Numbers in the Newsroom* (2014) Investigative Reporters & Editors. Las matemáticas de forma sencilla para los periodistas, con ejemplos de cómo calcular las fórmulas más esenciales, necesarias para entender las estadísticas.

Sitios web

- [GapMinder](#) – Enseña a fomentar una visión del mundo basada en hechos, explorando datos y hechos ampliamente disponibles sobre salud, riqueza, demografía y medio ambiente, y cómo combinar los datos en narrativas convincentes.
- [Sección de Ciencia de FiveThirtyEight](#) – Buenos ejemplos del uso de datos científicos de forma accesible y precisa.
- [Science Feedback](#) – Un proyecto sin ánimo de lucro que reúne a científicos con experiencia relevante de diferentes instituciones y disciplinas (actualmente limitado a temas de clima y salud) para revisar y comentar artículos de noticias que contienen información y datos científicos, señalando inexactitudes y razonamientos defectuosos.

Acerca de la autora

Elisabetta Tola es una científica y periodista de datos que tiene un doctorado en microbiología. Fue becaria Tow-Knight en 2019 en la Escuela de Periodismo Craig Newmark de la City University of New York. Como fundadora de **Facta** y directora ejecutiva de la agencia de comunicación científica formicablu, Tola aplica la metodología científica al periodismo. Tola es autora de las investigaciones internacionales de datos y multimedia **Seedversity.org**, **SEEDcontrol.eu** y **Hearing Voices**. Es presentadora de radio en **RAI Radio3scienza** y ha trabajado como independiente para publicaciones como ESMH, IIBOLive, AGI y WiredIT. También es instructora de medios de comunicación y profesora de periodismo científico de datos y autora de un **manual para periodistas sobre ciencia medioambiental para el Instituto Forestal Europeo**.

Referencias

- 1 <https://www.nytimes.com/article/how-to-read-a-science-study-coronavirus.html>
- 2 <https://www.theguardian.com/world/2020/jun/04/COVID-19-lancet-retracts-paper-that-halted-hydroxychloroquine-trials>
- 3 <https://fivethirtyeight.com/features/what-happened-with-that-bogus-hydroxychloroquine-study/>
- 4 <https://idatassist.com/>
- 5 https://www.researchgate.net/publication/45891209_P-Values_Confidence_Intervals_or_Confidence_Levels_for_Hypotheses
- 6 <https://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/>
- 7 <http://www.nature.com/nmeth/journal/v12/n3/abs/nmeth.3288.html>
- 8 <https://med.stanford.edu/news/all-news/2016/03/misleading-p-values-showing-up-more-often-in-journals.html>
- 9 https://www.cdc.gov/breastfeeding/data/nis_data/results.html
- 10 <https://www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html>
- 11 <https://blogs.scientificamerican.com/observations/five-sigmawhats-that/>
- 12 <http://www.uniteforsight.org/global-health-university/nutrition-study>
- 13 <https://www.ams.org/notices/199603/comm-kolata.pdf>
- 14 <https://www.theguardian.com/commentisfree/2011/aug/12/bad-science-exaggerated-study-results>
- 15 <https://heart.bmj.com/content/85/5/491>
- 16 https://www.bls.gov/data/inflation_calculator.htm
- 17 <https://tylervigen.com/spurious-correlations>
- 18 https://www.youtube.com/watch?time_continue=332&v=RmYlPnzbQtM&feature=emb_logo
- 19 https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_06-508.pdf
- 20 <https://www.bhf.org.uk/-/media/files/research/heart-statistics/bhf-cvd-statistics-uk-factsheet.pdf?la=en>
- 21 <https://www.forbes.com/pictures/598363d84bbe6f5997eede26/6-texas>
- 22 <https://www.forbes.com/pictures/598363d84bbe6f5997eede26/51-new-york/?sh=370e9b36b7ce>
- 23 <https://time.com/5680432/climate-change-history-carbon/>
- 24 <https://ourworldindata.org/world-population-growth>
- 25 <https://www.bmj.com/content/315/7109/672.full>
- 26 <https://www.ire.org/product/numbers-in-the-newsroom-using-math-and-statistics-in-news-second-edition-e-version>

¿Qué es el periodismo de datos?

Esta disciplina periodística se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella, se trabaja **con grandes volúmenes de datos**, se aprovecha al máximo la **visualización interactiva** y se incorpora al **programador** al equipo periodístico.

SANDRA CRUCIANELLI

Muchos colegas me preguntan qué es el periodismo de datos (PdD). No encuentran lógica en la denominación, ya que, desde siempre, los periodistas nos valemos de datos para producir contenidos periodísticos. Corresponde, en realidad, llamarlo “periodismo de base de datos” (PBD); pero, por uso y costumbre, predomina la denominación más corta.

- ¿Es periodismo de investigación (PI)?

Ciertamente, se usan técnicas propias del periodismo investigativo.

- ¿Es periodismo en profundidad (PP)?

También, dependiendo del tema que se esté analizando.

- ¿Es periodismo de precisión (PPr)?

En ocasiones, efectivamente, hay que recurrir a los métodos de investigación social definidos en la década de los 60 por

Philip Meyer para su aplicación en el campo del periodismo.

- ¿Es periodismo asistido por computadora (PAC)?

Casi siempre, porque tenemos que analizar datos y nos valemos de hojas de cálculo, así como de los procedimientos propios de esta disciplina. De hecho, creo que el PAC solo queda resumido a cuando manejamos volúmenes de datos pequeños y, para casos más grandes en cuanto a volumen de información, ha dejado de ser PAC para convertirse en periodismo de datos.

- ¿Es periodismo analítico (PA)?

También, ya que en general usamos métodos analíticos para llevar los datos a sistemas de información geográfica que nos permitan visualizar mejor el comportamiento de las variables que tenemos.

En realidad, lo que hoy se conoce como

Sandra Crucianelli es periodista especializada en periodismo de investigación y periodismo de precisión, fundadora de Sololocal.info, miembro del Consejo Asesor del Centro de Periodismo Digital de la Universidad de Guadalajara (México), docente del Programa de Educación Virtual del Centro Knight de la Universidad de Texas y del Centro de Medios de la Universidad de Florida (ambas en EE. UU.).

periodismo de datos –que, como ya dije, es verdaderamente periodismo de base de datos– involucra todo lo anteriormente citado, a lo que debemos agregar:

- 1.- Grandes volúmenes de datos, que muchas veces ni siquiera podían alojarse en una hoja de cálculo de Excel
- 2.- Visualización interactiva
- 3.- Incorporación del programador al equipo periodístico para desarrollar tareas que van desde la extracción de datos y la depuración de los sets de datos hasta el diseño de aplicaciones de noticias (*news apps*)

Entonces, si pudiéramos escribirlo a modo de fórmula, podemos decir que tenemos una suma de métodos conocidos a los que se añaden tres frutos de la innovación tecnológica:

PI + PP + PPr + PA + PAC + volumen de datos + visualización interactiva + programación = PdD o PBD
--

Para que comencemos desde cero y quede todo claro, las definiciones son las siguientes:

- **Periodismo de investigación:** se requieren tres condiciones necesarias para que un reportaje se precie de ser investigativo: que sea un tema de relevancia social, a lo que se añade que alguien pretenda mantenerlo en secreto y, finalmente, se requiere un trabajo adicional del periodista, ajeno a las prácticas cotidianas de cobertura informativa. Son los casos en los que se investigan asuntos relacionados con irregularidades en la administración del dinero público, como los casos de corrupción, cuando se

revelan situaciones de ineficiencia o abuso por parte de compañías privatizadas de servicios públicos y todo aquello que cause una afectación a la sociedad (por ejemplo, un caso de delito ecológico).

- **Periodismo de profundidad:** es lo mismo que el PI, solo que en este caso nadie intenta mantener en secreto lo que el periodista investiga. Así, por ejemplo, un estudio sobre las variables de pobreza con datos de los censos nacionales.

- **Periodismo de precisión:** formulado como disciplina –como ya apunté– por el profesor Philip Meyer, involucra el uso de métodos de la investigación social aplicados al periodismo tales como la investigación cualitativa (análisis de casos, grupos focales, análisis del discurso, etc.) y la cuantitativa. Esta última incluye la descriptiva (cuando se consideran todos los casos: por ejemplo, un censo entre legisladores) y lo que se denomina investigación inferencial, es decir, cuando se trabaja sobre un universo y, dada la dificultad para abordar a cada miembro de ese universo, se diseña una muestra representativa, a la cual se procede a poner bajo estudio. Es el caso de las encuestas de opinión o las electorales con fines predictivos.

- **Periodismo analítico:** analiza una realidad compleja, con el fin de lograr la comprensión del público. Combina algunos aspectos del periodismo de investigación y la información explicativa. Su misión es ofrecer pruebas basadas en interpre-

taciones de la realidad, explicándolas de un modo comprensible. Contextualiza un tema de fondo mediante la descripción de detalles históricos y datos estadísticos. El resultado es una explicación completa, destinada a dar forma a la percepción del público del fenómeno. Aspira a recoger datos dispares y hacer conexiones que no son inmediatamente evidentes, y es eficaz a menudo en el análisis entre los hechos y no en los hechos mismos. Utiliza como pilar de la visualización los llamados sistemas de información geográfica (SIG).

- **Periodismo asistido por computadora:** conocido como CAR (*Computer Assisted Reporting*) por sus siglas en inglés, se define como “cualquier proceso que utiliza la asistencia de computadoras durante la recolección y procesamiento de datos”. Ha basado su práctica en el uso de hojas de cálculo de Excel para el análisis de información contenida en tablas.

Los productos del periodismo de datos

Hay cuatro productos diferentes al menos, generalmente complementarios, que surgen de esta práctica:

- 1.- Artículos basados en datos
- 2.- Visualizaciones interactivas
- 3.- Conjuntos de datos abiertos
- 4.- Aplicaciones de noticias

- **Artículos basados en datos:** generalmente, son artículos cortos construidos a partir de grandes volúmenes de datos como los que podrían estar contenidos en bases de datos o conjuntos de datos diseñados

a partir de una o varias bases de datos (<http://www.lanacion.com.ar/1484852-los-millones-de-la-ape-como-se-repartio-en-2011-la-caja-que-era-de-moyano>).

- **Visualizaciones interactivas:** en general, son complementos de los artículos basados en datos, pero pueden ser contenidos en sí mismos sin necesidad de que exista un artículo referencial: en algunos casos, es suficiente un título, un subtítulo y una explicación a la audiencia sobre cómo leer los datos desde la visualización (<http://www.lanacion.com.ar/1546303-los-bienes-de-los-funcionarios-en-la-primera-news-application-de-la-nacion>).

- **Conjunto de datos abiertos** (conocidos como set de datos): los medios de comunicación no necesariamente se abocan a realizar trabajos periodísticos basados en datos. También pueden, como el caso de *La Nación* de Argentina, dedicar una gran parte de su esfuerzo a la apertura de datos públicos. En países en los que no existen leyes de acceso a la información pública, estas no funcionan como deberían, en los sitios webs del Gobierno hay algunos datos pero otros no, están ocultos bajo distintas capas de información o los formatos para mostrar los datos en la web no son accesibles (ver un dato en la pantalla no es lo mismo que acceder al mismo). Para que un dato sea accesible se debe mostrar en un formato reutilizable. Por ejemplo, un documento oficial subido a Flickr. En estas circunstancias, medios y organizaciones pueden dedi-

carse a la búsqueda, extracción y apertura de datos públicos, tal como *La Nación* hace a través de su catálogo de datos. Esto es, ya que hay datos a los que los Gobiernos o las organizaciones no permiten acceder –sea por la razón que fuere–, los medios de comunicación pueden incursionar en ese terreno, y así seguramente harán grandes aportes. Los datos pueden gestionarse o administrarse desde plataformas sencillas y gratuitas como Google Drive (<http://data.lanacion.com.ar/dashboards/5068/inflacion-y-precios/>).

- **Aplicaciones de noticias:** conocidas en inglés como *news apps*, son el fruto de la labor conjunta de periodistas y programadores informáticos. Muchas veces, el volumen de datos es tan grande que resulta imposible encontrar una noticia si no se diseña una aplicación que nos permita agrupar y analizar variables, por ejemplo, por localización geográfica, por fecha, por nombre de compañía, etcétera (<http://www.gastopublicobahiense.org/> - http://www.nacion.com/CustomerFiles/nacioncom/Generales/Subsitios/Investigacion/2012/propiedadessubvaloradas/props_diputados/index.html - <http://www.texastribune.org/library/data/government-employee-salaries/> - <http://projects.propublica.org/docdollars/>).

Sin cultura de datos abiertos

No existe una cultura de datos abiertos en la mayoría de los países de habla hispana. Sin embargo, lo que no hagan los

Gobiernos en materia de datos abiertos lo hará la comunidad, las organizaciones de la sociedad civil o los medios de comunicación.

El pronóstico no es nuevo. Un ejemplo vio la luz en 2010, cuando el diario *Texas Tribune* solicitó al Gobierno estatal el detalle de los salarios de los 674.000 empleados públicos, con la intención de diseñar una aplicación que permitiera a los ciudadanos consultar los haberes que se pagan con fondos públicos, incluidos maestros de escuelas y profesores de universidades públicas (<http://www.texastribune.org/library/data/government-employee-salaries/>).

Para hacer periodismo de datos se puede prescindir de muchas cosas, ya que hay herramientas digitales gratuitas que todos los periodistas podemos usar para sortear casi cualquier obstáculo, pero de lo que no podemos prescindir es de los datos. Son una condición necesaria sin la cual no podemos siquiera comenzar a diseñar hipótesis periodística alguna.

En muchos casos, el formato en que se visualizan los datos en las webs no es reutilizable. Verdaderamente, los periodistas podemos resolver estos escollos aprendiendo técnicas específicas. Pero ¿qué ocurre con los ciudadanos? ¿Acaso los contribuyentes no tienen derecho a un mejor acceso a los datos directamente relacionados con el dinero público?

La información con atraso, desordenada, aún desactualizada o en formatos inadecuados se puede leer en pantalla

ciertamente, pero visualizar datos no es lo mismo que acceder a ellos. Es el clásico ejemplo de una imagen convertida en un PDF en el que se lee el detalle de una tabla, como la ejecución de un presupuesto. Habría que recurrir a una técnica de reconocimiento óptico de caracteres (OCR, por sus siglas en inglés) para extraer los datos; aunque estos programas no siempre resultan eficientes y no queda otro recurso que la carga manual de datos, lo cual complica enormemente la tarea de todo reportero, en especial si el contenido de la tabla que queremos procesar es extenso.

En el filme *Ojos bien cerrados*, dirigido en 1999 por Stanley Kubrick poco antes de su muerte, el eje argumental versa sobre dos sentimientos concretos: la incertidumbre y la insatisfacción. Por poner un ejemplo, en Argentina, sin los componentes del *thriller* que rodean aquella oscura historia, 103 años después, ciudadanos y periodistas nos encontramos con *Datos bien cerrados*. Ante ellos, experimentamos incertidumbre y nadie puede sentirse satisfecho con un escenario en el que hay que tener habilidades específicas, cuando no pagar por recursos adicionales, para acceder a datos legítimamente públicos.

Cómo armar un set de datos

Los datos ya han sido recuperados de la web o descargados en el PC. Tal vez, el periodista haya tenido que recurrir a la entrada de datos manual, ya que esto es frecuente cuando los documen-

tos, mediante petición, son entregados en formato papel y no resulta posible la extracción de la información y su posterior reestructuración en tablas, considerando que no siempre funcionan con eficiencia los programas de reconocimiento óptico de caracteres.

En periodismo de datos, consideramos que un set de datos es una colección de elementos de datos agrupados, que permiten su fácil registro. En ellos, los datos se encuentran organizados en variables, por lo general, medidas a lo largo del tiempo y en soportes descargables.

Visualizar datos no es lo mismo que acceder a ellos

La forma más sencilla de gestionar sets de datos es usando Google Drive, la herramienta que reemplazó a Google Docs por sus mayores y mejores funcionalidades. Es necesaria su libre descarga a partir de una cuenta de Gmail y, una vez instalada, permite alojar todo tipo de documentos, en el caso que nos ocupa, desde hojas de cálculo en Excel hasta tablas, gracias a su novedosa función Google Tablas Dinámicas (<https://support.google.com/fusiontables/answer/2571232?hl=es>).

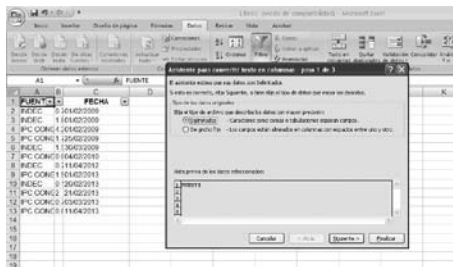
Las formas más comunes de gestionar y alojar sets de datos se llevan a cabo mediante planillas de Excel o mediante un formato denominado CSV, igualmente abierto y reutilizable, en el que las columnas están separadas por comas.

Cuando se descarga un fichero de datos en CSV, la imagen que se obtiene es del siguiente tipo:

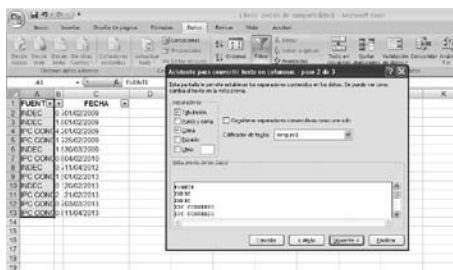
FUENTE	FECHA
2. INDEC	01/01/2009
3. INDEC	11/01/2009
4. IPC CONGRESO	01/02/2009
5. IPC CONGRESO	15/02/2009
6. INDEC	11/03/2009
7. IPC CONGRESO	04/03/2010
8. INDEC	01/10/2012
9. IPC CONGRESO	15/01/2013
10. INDEC	01/30/2013
11. IPC CONGRESO	21/02/2013
12. IPC CONGRESO	03/03/2013
13. IPC CONGRESO	11/04/2013

En estos casos, hay que volver a estructurar las columnas. El procedimiento es sencillo, sin mover el cursor de la Fila A:

- 1.- Ir a la pestaña “Datos”
- 2.- Hacer clic en “Texto en columnas”
- 3.- Se abrirá un cuadro de diálogo, marcar “Delimitados” + “Siguiete”



- 4.- Marcar “Tabulación” + “Comas”



5.- Nótese que, durante el procedimiento, la columna A permanece marcada o “pin-tada”. Luego, hacer clic en “Siguiete” y lo que se obtendrá será este resultado:

FUENTE	VARIACION	FECHA
2. INDEC	0,7	01/01/2009
3. INDEC	1,87	01/01/2009
4. IPC CONGRESO	4,2	01/02/2009
5. IPC CONGRESO	1,7	21/02/2009
6. INDEC	1,5	03/03/2009
7. IPC CONGRESO	0,8	04/03/2010
8. INDEC	0,7	11/04/2010
9. IPC CONGRESO	1,52	01/02/2013
10. INDEC	0,11	03/30/2013
11. IPC CONGRESO	2	21/02/2013
12. IPC CONGRESO	0,7	03/03/2013
13. IPC CONGRESO	0,8	11/04/2013

6.- Para lograr el resultado que ofrece la vista de arriba se centraron los valores de las celdas y solo resta colocar las variables que no aparecen en las columnas B y C, que originalmente eran “Variación” y “Fecha”.

FUENTE	VARIACION	FECHA
INDEC	0,7	1/31/2011
INDEC	1,87	1/31/2011
IPC CONGRESO	4,2	2/28/2011
IPC CONGRESO	1,7	2/28/2011
INDEC	1,5	3/31/2011
IPC CONGRESO	0,8	3/31/2011
INDEC	0,7	4/30/2011
IPC CONGRESO	1,52	4/30/2011
INDEC	0,11	5/31/2011
IPC CONGRESO	2	7/30/2011
IPC CONGRESO	0,7	6/30/2011
IPC CONGRESO	0,8	6/30/2011
INDEC	1,87	7/30/2011
IPC CONGRESO	0,8	8/31/2011
INDEC	1,89	8/31/2011

Con este último paso, el procedimiento se completa. El lector puede probar el procedimiento descargando este set de datos en formato CSV en: <http://data.lanacion.com.ar/datastreams/69218/inflacion-indec-y-estimaciones-privadas-variacion-mensual/>. En este caso, se trata de un set

perteneciente al catálogo de datos abiertos del diario *La Nación* de Argentina.

Estos sets están alojados dentro de una plataforma denominada Junar. Con una cuenta gratuita, este recurso permite gestionar hasta cinco sets de datos. La plataforma posibilita a las empresas, Gobiernos y otras organizaciones liberar sus datos y exponerlos como datos abiertos, impulsando la colaboración y la transparencia.

Crónicas cortas, con conclusiones contundentes y una gran visualización

La Plataforma Junar Open Data permite recoger, refinar y publicar al mismo tiempo. Varios Gobiernos, medios de comunicación como *La Nación* de Argentina y ONG la usan para compartir sus colecciones de datos con la audiencia. La pestaña inferior izquierda de cada set de datos, denominada “Acciones”, permite la exportación de información tanto en Excel como en CSV, así como obtener un código para incrustar en un blog o sitio web, visualizar como Google Hojas de Cálculo y socializar a través de Twitter, Facebook, LinkedIn y Google Plus.

Hay otras plataformas desde donde se pueden alojar, gestionar y descargar sets de datos, como Socrata, de la que se valen igualmente tanto medios de comunicación como varias ONG. Un ejemplo es la colección de set de datos denominada Finances, del Banco Mundial. Todos

los sets de datos financieros del Banco Mundial están aquí: <https://finances.worldbank.org/page/datasets>. El ejemplo más conocido dentro del campo periodístico es el que se visualiza en este enlace: <https://finances.worldbank.org/Procurement/Major-Contract-Awards-FY2007-FY2013-Beta-version/kdwi-wcs3>, que contiene información completa de todos los préstamos otorgados desde 2007 hasta la actualidad por países.

A la derecha de cada columna, hay una flecha que abre un cuadro desde donde se pueden seleccionar opciones como ocultar columnas o seleccionar un solo país.

Desde el momento en que se activa el proceso de selección de datos, la plataforma permite guardar esa selección como propia, descargarla en el PC, así como visualizar los datos de la serie, entre otras funciones.

Para ver más recursos en Socrata, se puede consultar en: <https://opendata.socrata.com/>.

Visualización para principiantes

El periodismo de datos maneja un volumen grande de datos. Sin una adecuada visualización que acompañe el análisis de esos datos, sería imposible visualizar sencillamente las variables que se ponen en estudio. Las crónicas deberían ser extensas y podrían pecar de aburridas si están plagadas de números que la gente no entienda con facilidad al no ser expuestos adecuadamente. Por ello, se sugieren crónicas cortas, con conclusiones contundentes y una visualización

importante, dominante en el artículo, a la hora de contar una historia.

Los expertos en visualización interactiva son capaces de diseñar, a partir de una gran cantidad de datos, un gráfico que permite al lector decidir qué desea visualizar y cómo. La visualización de datos explica de manera comprensible las relaciones entre gran cantidad de información que se genera por distintas vías.

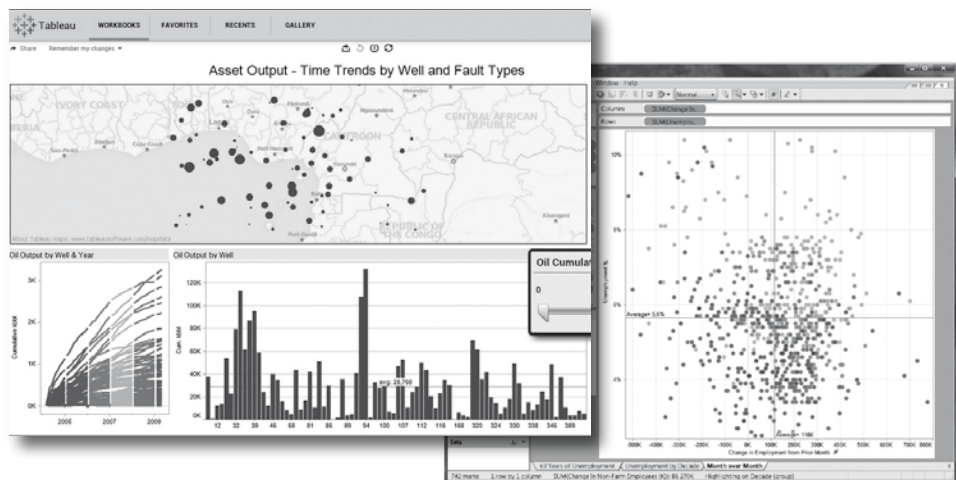
Para principiantes en visualización: ChartGo y, después, Many Eyes

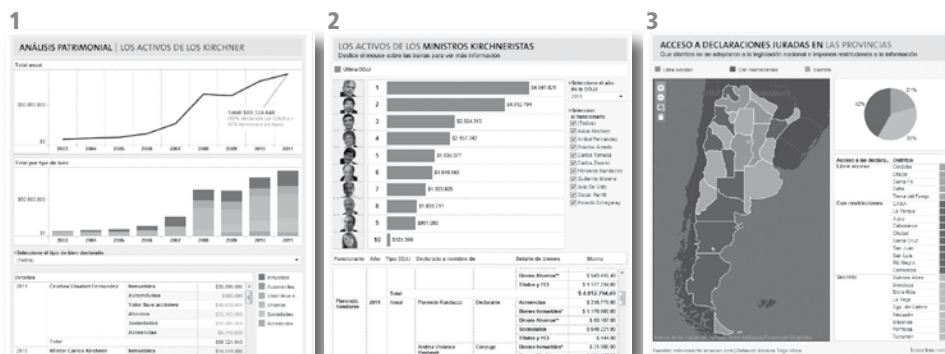
El clásico ejemplo para el reportero es el presupuesto público. El Ministerio de Economía entrega un pesado reporte lleno de tablas y números que no producen ninguna información interesante a simple vista ni, mucho menos, una noticia

que no sea la anunciada en el parte oficial de prensa. Aunque un presupuesto contiene variables, es decir, elementos que son medidos a lo largo del tiempo. Y si se dispone de un presupuesto del año en curso, probablemente tenga a mano el del año anterior, que tiene... las mismas variables. Sin embargo, muchas veces, hasta la más sencilla de las tablas puede resultar muy aburrida.

La comparación es quizá una de las prácticas más usadas por el periodismo, pero, cuando hay que procesar gran cantidad de información numérica, el asunto se complica. Por ello, se recurre a distintas formas de visualizar el conjunto de datos; de ese modo, podemos analizar mejor ciertos procesos y cambios no detectados de un vistazo.

Uno de los muchos ejemplos que podemos mencionar es la herramienta Tableau Public:





Hay una larga lista de ejemplos que el diario *La Nación* de Argentina puede mostrar en el uso de este recurso. Por ejemplo, las siguientes:

Declaraciones juradas patrimoniales¹

- Tipo de representación visual: fiebre + barras + tabla / barras + tabla
- Interactividad: menú desplegable + *tooltip* con datos al *roll over*
- Programa: ambos en Tableau Public

Declaraciones juradas patrimoniales²

- Tipo de representación visual: barras + imágenes fotográficas + tabla / círculos apilados + tabla
- Interactividad: menú desplegable, selector + *tooltip* con datos al *roll over*
- Programa: ambos en Tableau Public
- Extras: enlaces a Documents Cloud en el *tooltip*

Declaraciones juradas patrimoniales³

- Tipo de representación visual: geolocalización con polígonos + torta + tabla
- Interactividad: *tooltip* con datos y filtro al *roll over*
- Programa: Tableau Public

Desempleo en Argentina⁴

- Tipo de representación visual: fiebres + tabla
- Interactividad: menú desplegable + selector + *tooltip* con datos al *roll over*
- Programa: Tableau Public

Subsidios a transporte público de pasajeros⁵

- Tipo de representación visual: semicírculos y barras apiladas
- Interactividad: filtro al *roll over* y aparición de barras al clic
- Programa: Flash

- 1.- <http://www.lanacion.com.ar/1525583-el-crecimiento-de-los-bienes-de-los-kirchner-de-7-a-89-millones-de-pesos>.
- 2.- <http://www.lanacion.com.ar/1535554-la-evolucion-de-la-fortuna-personal-de-los-ministros-kirchneristas>.
- 3.- <http://www.lanacion.com.ar/1546537-que-provincias-esconden-bajo-llave-el-patrimonio-de-sus-gobernadores>.
Enlace directo: <http://public.tableausoftware.com/shared/Z47YDQC9M>.
- 4.- <http://www.lanacion.com.ar/1552132-el-desempleo-en-personas-como-vo>.
Enlace directo: <http://public.tableausoftware.com/shared/TWDPMN4P>.
- 5.- <http://www.lanacion.com.ar/1380725-colectivos-insaciables-un-cheque-diario-de-10-millones-en-subsidios>.

Asimismo, hay hoy otros recursos que compiten por la delantera, como Google Fusion Tables, que permite diseñar desde cero una tabla y trasladar esa información hacia un mapa. Hay una galería de imágenes que se puede consultar en: <https://sites.google.com/site/fusiontablestalks/stories>. Por supuesto, se puede comenzar desde cero, aunque eso requiere entrenamiento. Se puede ver un tutorial en: <http://support.google.com/fusiontables/answer/184641?hl=en>. Y más en: <http://pushmatrix.github.com/fusiontable-slides/#/2>.

A modo ilustrativo, el proyecto Censo del *New York Times* ha utilizado este recurso para mostrar los datos del censo 2010: <http://projects.nytimes.com/census/2010/map>. Pero no es el único, ya que otros medios han usado esta herramienta para el mapeo interactivo del censo, como el *Texas Tribune* (<http://www.texastribune.org/library/data/census-2010/>) y el *Washington Post* (<http://www.washingtonpost.com/wp-srv/special/nation/census/2010/>).

Sin embargo, estos recursos no son sencillos de usar para un principiante. Requieren entrenamiento y mucha práctica, aunque esto no debe ser visto como un impedimento, sino como un desafío. Un ejemplo es el del bloguero Andy Tow, quien usa la herramienta con frecuencia para su blog de datos, tal como se visualiza aquí: <http://towsa.com/wordpress/>. Entonces, la conclusión es: si Andy puede hacerlo, otros pueden seguir su ejemplo.

No obstante, hay recursos sencillos que posibilitan graficar de manera interactiva

sin necesidad de mayores conocimientos. Quizá el más usado y versátil es Many Eyes: permite diseñar varios tipos de visualizaciones diferentes, creando una cuenta gratuita; no es necesario descargar ningún programa (lo cual es ventajoso para muchos periodistas que trabajan en redacciones en las que, por cuestiones de seguridad informática, sus terminales no les permiten la bajada de programas).

Sugiero a los principiantes que comiencen con ChartGo (<http://www.chartgo.com/>), el cual usé experimentalmente y sin experiencia previa alguna en 2010, cuando comenzaba a incursionar en periodismo de datos (ver nota en: <http://www.sololocal.info/noticias/1-de-bahia/1968-pauta-oficial-el-reparto-de-la-torta-.html>).

Luego, un segundo paso sería aprender el uso de Many Eyes, al que considero muy conveniente por las razones antes apuntadas.

Esto va dirigido especialmente al reportero que no dispone de un profesional especialista en la materia dentro de su equipo: esta es la realidad de miles de reporteros, blogueros y periodistas de portales de noticias en línea, que deben cumplir varias funciones al mismo tiempo por carencia de recursos.

Afirmar que para hacer periodismo de datos se requiere de un equipo visualizador especializado como condición sine qua non sería establecer una brecha demasiado grande y peligrosa entre unos y otros.

Por eso, los recursos más complicados de usar quedan en manos de pro-

fesionales o en las de aquellos que se lanzan a descubrir como autodidactas la maravillosa tarea de llevar a una imagen algo tan aburrido como una tabla. Y para quienes no hemos recibido instrucción formal en estas prácticas nos quedan los recursos más sencillos o los que yo llamo “*ready to use*”, es decir, listos para usar.

Gráficos interactivos inteligentes

La visualización de datos está teniendo en nuestros días un desarrollo espectacular, que impacta directamente en el periodismo; no solamente en medios digitales, también la prensa gráfica y la televisión se están valiendo de estas herramientas para presentar visualmente sus historias de una forma más interesante.

Hace algunos años, para ser periodista, con buscar datos, abordar fuentes y saber escribir era suficiente. En nuestro tiempo, eso ha cambiado, porque los periodistas comenzamos a percibir que, si no manejamos adecuadamente estos recursos, podemos incluso perder la oportunidad de encontrar primicias. Esto requiere que el reportero desarrolle habilidades especiales.

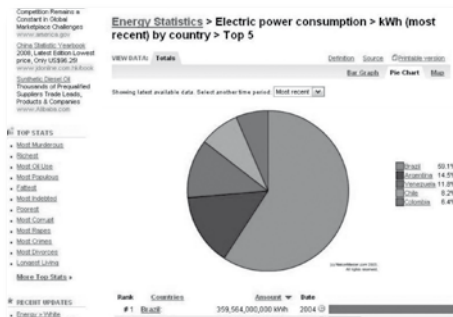
Buena parte de la información se encuentra en bases de datos u oculta en las webs. A finales de los 70, cuando comencé mi carrera como reportera, el escenario era de pocos datos a la vista, por lo que los periodistas de mi generación debimos adaptarnos a trabajar en un terreno más dominado por las fuentes orales que por las documentales.

Eso ha cambiado en el siglo XXI, en el que la sobreabundancia de información puede incluso sumirnos en un caos. Antes, el reportero escribía y el equipo de infografía hacía su labor sobre la base de los datos que el periodista le proporcionaba. Esa relación ha cambiado drásticamente en los últimos años.

Ahora, para encontrar noticias, el periodista necesita muchas veces apoyarse en herramientas visuales como Google Maps y Google Earth u otras similares como la de generación de gráficos.

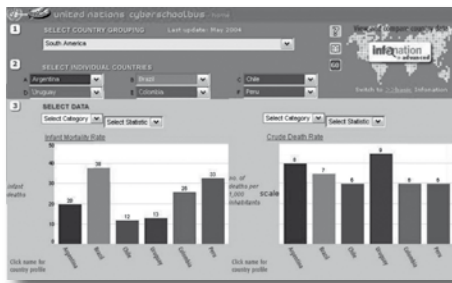
Por ejemplo, varias crónicas periodísticas de los años 90 fueron ilustradas o surgieron sobre la base de recursos tales como NationMaster (www.nationmaster.com), que permite obtener gráficos e información comparada entre varios países del mundo.

Se ingresa mediante la pestaña “Countries A-Z”, desde donde se seleccionarán las variables. La imagen inferior (<http://www.nationmaster.com/countries>) corresponde a una captura de pantalla de un gráfico de torta realizado a través de NationMaster para comparar el consumo de energía en cinco países de América del Sur.



Otro similar, muy usado en el pasado, era InfoNacion, un sitio web educativo de las Naciones Unidas de uso frecuente en liceos y universidades, al que se accede mediante el siguiente enlace: http://cyberschoolbus.un.org/infonation/s_infonation.htm.

La imagen de abajo (<http://cyberschoolbus.un.org/infonation3/ie&nn4.html>) corresponde a una captura de pantalla del recurso para la comparación de distintas variables correspondientes a seis países de América del Sur. Generar las gráficas demandó apenas 40 segundos.



Estos recursos, sin embargo, poseen la desventaja de que los datos no están tan actualizados como quisiéramos. Por ejemplo, varios indicadores corresponden al año 2004. Ya nadie los usa, salvo como referenciales.

Los administradores de medios independientes o blogueros han debido aprender a utilizar otros de más reciente aparición o que ofrecen mayores ventajas.

Tablas en línea

Insertar una hoja de cálculo en una página web puede resultar un problema si no se tiene mucha experiencia con Excel

o si lo que se desea es obtener el código HTML que traduzca esa información de un modo sencillo y amigable.

COUNTRY	POPULATION	RANK
China	1,34,73,50,000	1
India	1,21,01,93,422	2
USA	31,48,90,000	3

Tableizer (<http://tableizer.journalistopia.com/>) convierte cualquier hoja de cálculo en una tabla HTML. Lo único que hay que hacer es copiar el contenido de la hoja de cálculo y pegar esa información en el cuadro que tenemos a nuestra disposición. Posteriormente, se pueden configurar preferencias en cuanto al formato y, automáticamente, un botón genera el código. El resultado es una vista muy sencilla pero muy útil, ya que en menos de unos minutos se puede obtener el resultado. En la imagen de arriba, en la parte izquierda, se puede observar la introducción de datos en Tableizer; a la derecha, lo que resulta de pegar el código HTML (fuente: <http://img.labnol.org/di/tableizer2.gif>).

Sistemas de información geográfica

Se los define como sistemas de integración de datos, capaces de mostrar información geográficamente localizada. Los datos censales, por ejemplo, suelen ser aburridos; aunque, administrados mediante programas de SIG, pueden ofrecer una visión clara de las variables en estudio.

Los usuarios de estos sistemas pueden crear mapas interactivos, poner bajo la lupa información espacial y editar datos y mapas, de tal modo que se ofrezcan mejores representaciones de la realidad.

En palabras sencillas, un SIG es un *software* que permite una mejor representación visual de datos numéricos, conforme estos se distribuyan geográfica y espacialmente.

El Instituto de Periodismo Analítico (<http://www.analyticjournalism.com/>) viene trabajando desde hace años en la incorporación de estas herramientas al campo de las comunicaciones. El sitio contiene una gran cantidad de valiosas herramientas que serán de mucha ayuda para docentes de periodismo y reporteros interesados en el tema.

Un SIG permite una mejor representación visual de datos numéricos

Para comprobar cómo los sistemas de información geográfica impactan en el periodismo, se puede ver la ponencia de Tom Johnson, director del Instituto de Periodismo Analítico, en el siguiente enlace: <http://www.slideshare.net/jtjohnson>.

Las universidades y empresas privadas en América Latina vienen interesándose por esta tecnología desde hace algunos años e, igualmente, los Gobiernos incorporan los SIG a distintas áreas de gestión.

También en el área de transporte, el recurso <http://mtcgeo.mtc.gob.pe/website/telecom/viewer.htm> es de libre consulta, como el relacionado con vivienda, por lo que es fácil imaginar que estas herramientas dan soporte esencial para la visualización de datos en estas materias.

En Argentina, por ejemplo, el sitio <http://www.mapaeducativo.edu.ar/> ofrece una introducción a estos sistemas, tanto para docentes como para alumnos; pero también hay herramientas para periodistas, como la localización de escuelas y su distribución espacial, que bien podrían formar parte de una crónica.

Por su parte, el sitio <http://www.sig.gov.ar/> reúne información de diferentes oficinas estatales, generando recursos de libre uso para los usuarios registrados.

Los sistemas de información geográfica disponibles en una larga lista de plataformas de E-Gov mejoran el acceso a la información pública por parte de los ciudadanos, por lo que los periodistas no deberían quedar al margen de este proceso. En el futuro, consultar estos sistemas será tan frecuente como hoy en día lo es consultar el pronóstico del tiempo en cualquier portal sobre meteorología.

Así como un reportero puede consultar estos sistemas, también puede manejarlos, cosa que requiere un cierto entrenamiento y práctica. La mayoría de los programas de SIG son de pago, como el *software* ArcView (<http://www.esri.com/>), que sigue siendo

uno de los SIG estándar más utilizados en el mundo. Si bien la mayoría de los medios está usando ArcView, otros utilizan Maptitude, ya que tiene una interfaz en español, como se ve en el enlace <http://www.caliper.com/maptitude/MappingSoftware.htm>, así como la versión profesional de MapInfo, que se obtiene en el sitio <http://www.pbinsight.com/>, la cual dispone asimismo de una versión básica gratis, aunque en inglés.

Gráficos listos para usar

- **Sistema Regional de Visualización y Monitoreo (Servir):** una serie de recursos interactivos se encuentra disponible en <http://www.servir.net/america-latina-caribe>, que integra observaciones de la tierra con modelos de pronóstico y datos del escenario geográfico. En Centroamérica y el Caribe, el equipo de Servir ha dado respuesta a más de 20 desastres naturales y diez amenazas ambientales. El portal geoespacial permite el acceso a los datos y metadatos de la región.

- **Portal de datos:** <http://maps.cathalac.org/Portal/>.

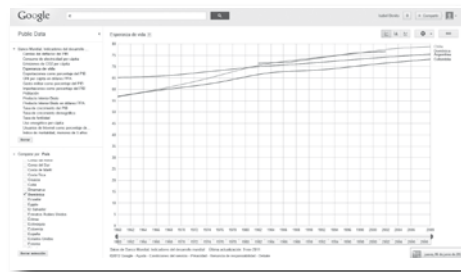
- **Visualizador de mapas:** <http://share1.cathalac.org/servirmaps/index1.html>.

- **Pronóstico para Mesoamérica:**
http://www.servir.net/MM5_Mesoam
%C3%A9rica.

- **Análisis de desastres:** http://www.servir.net/An%C3%A1lisis_de_Desastres/Desastres/.

- **GapMinder:** muchos de nosotros conocemos <http://www.gapminder.org/>, un clásico en este tema. Explorar esta herramienta puede resultar un excelente ejercicio, a fin de comprender sus potencialidades. Para obtener gráficas o mapas en GapMinder, el acceso (a la izquierda) es GapMinder World, en el que en la barra inferior se puede seleccionar la variable y, en el cuadro de la derecha, marcar el país o los países que se desean incluir. La página posee un tutorial en vídeo y en PDF.

- **Google PublicData:** http://www.google.com/publicdata?ds=wb-wdi&met=sp_dyn_le00_in&idim=country. Una herramienta interesante que permite analizar la expectativa de vida entre países de un modo muy sencillo: simplemente, se señala los países que se desea comparar y se obtiene un resultado como el que se muestra para Chile, Argentina, Colombia y República Dominicana en la captura de pantalla:



El servicio es muy sencillo de usar, recomendado especialmente para quienes no están muy familiarizados con el manejo de gráficos. Posee otras herramientas como Data Visualizer (<http://devdata.worldbank.org/DataVisualizer/>), que permite realizar gráficas sobre cualquiera de los 49 indicadores para 209 países. Las variables pertenecen a los campos de la economía, finanzas, información general, tecnología y medio ambiente.

Los *mashups* permiten combinar datos de diferentes fuentes

- **Google Finance:** <http://www.google.com/finance>. Para quienes cubren la sección de economía y finanzas, Google Finance es otra herramienta eficaz. En especial, si se trata de obtener gráficos sobre compañías públicas y privadas, ya que organiza la información relacionada con el mundo financiero: cotizaciones en Bolsa y datos financieros de las empresas. La búsqueda por nombre de empresa, como la que se ve en la captura de pantalla para Dow Chemical, una compañía que opera en la ciudad de Bahía Blanca (Argentina), proporciona datos en tiempo real, lo cual constituye un notable aporte para quienes trabajan en medios que manejan información de último momento.

Esta herramienta tiene un antecesor, Yahoo Finance: <http://espanol.finance>.



yahoo.com/. Esta se ha convertido en la favorita de muchos periodistas, ya que ofrece las cotizaciones de la Bolsa, tasas de cambio de valores, comunicados de prensa de empresas, además de algunas herramientas de gestión para finanzas personales. Pero habrá que probar uno y otro servicio para escoger el que se adapte mejor a las preferencias de cada usuario.

- **Google Maps:** <http://maps.google.com/>. Divide el mapa completo en pequeñas imágenes con un tamaño de 256 x 256 píxeles. Permite encontrar y mostrar rutas y direcciones para llegar a un lugar determinado. Una interesante característica es la de brindar distintas vistas de los mapas: la satelital y la que corresponde al relieve. El cursor se arrastra permitiendo los movimientos hacia cualquier dirección, mientras que la herramienta de zoom y alejamiento nos da una vista puntual o más general del espacio en estudio.

Las imágenes de Google Maps son tomadas mediante un satélite de teledetección, que viene funcionando desde 2001, ubicado a menos de 500 kilómetros sobre la superficie terrestre, en el espacio.

Aunque aún hay zonas geográficas del globo terráqueo que no tienen acce-

so a las vistas, actualmente constituye una herramienta de uso cotidiana para los reporteros.

Cualquier usuario con una cuenta de Google puede ingresar al sistema, buscar sus propios mapas, subir fotos y, de esta manera, ubicarse geográficamente en el mundo o ubicar su comercio, oficina o empresa, siendo útil para el mercadeo de actividades comerciales u organizacionales.

Las aplicaciones se multiplican: si se instala Google Gears (<http://gears.google.com/>) en nuestro navegador, va a aparecer un botón en la interfaz de Google Maps sobre la herramienta de zoom y bajo la herramienta de vista panorámica que, al pulsarlo, nos mostrará la posición en la que nos encontramos. Esta ubicación se obtiene a través de las redes wifi disponibles o mediante nuestra IP.

Otro recurso interesante que aparece en el menú es el llamado ¿Qué hay aquí?, con el que se puede encontrar el resultado que mejor represente a ese lugar, ya sea una dirección específica, parte del paisaje o el nombre de un sitio. Esta función tiene en cuenta el nivel de zoom que se está mirando y ofrece el dato geográfico más apropiado desde ese punto de vista.

- **Google Earth:** <http://earth.google.es/>. Como el propio nombre del recurso dice, Google Earth te permite volar a cualquier lugar de la Tierra y ver imágenes de satélite, mapas, relieve y hasta edificios en 3D. Ahora también puedes ver galaxias

del espacio exterior y hasta sumergirte en el océano. Se puede explorar un amplio contenido geográfico, guardar los lugares que se visitan y compartirlos con otras personas. El programa es de libre descarga y se accede en el siguiente enlace: <http://earth.google.es/download-earth.html>. Las imágenes que muestra Google Earth son obtenidas mediante un satélite llamado QuickBird, lanzado al espacio en octubre de 2001.

El recurso es muy útil cuando se investigan compañías o casos de propiedades adquiridas por funcionarios. O, por ejemplo, el impacto ecológico de la polución.

Puede citarse como caso práctico la crónica “Discurso oficial: a contramano de Google Earth” (<http://tinyurl.com/acntramanodegoogleearth>), que representa la manera más sencilla de presentar esta utilidad en el contexto de una noticia. Hay, por supuesto, otras mucho más sofisticadas.

Un dato que se debe tener en cuenta es que la búsqueda avanzada de Google permite encontrar lugares, personas, empresas o grupos seleccionando el formato que devuelve imágenes registradas por Google Earth. Esto resulta de mucha utilidad; en especial, cuando hay que monitorear movimientos portuarios, tal como se observa en la imagen disponible en <http://www.portalwhite.com.ar/archivos/IngenieroWhite-BuenosAires-Argentina.kmz>, obtenida tras una búsqueda en la que se introdujo en el cuadro para “la frase exacta” el término “Ingeniero Whi-

te” –que corresponde al puerto que funciona en Bahía Blanca– y se seleccionó el formato KMZ, que es el que devuelve vistas de Google Earth. El otro formato apto para selección es el KML.

Los formatos KMZ y KML sirven para representar datos geográficos. Un fichero KML especifica un lugar, una imagen o un polígono. Puede contener un título, una descripción básica del lugar, sus coordenadas (latitud y longitud), así como otros datos. En cuanto al formato KMZ, no es más que un fichero KML comprimido. En concreto, KMZ usa el formato de compresión ZIP.

Aplicaciones híbridas: ‘mashups’

Existe una variada gama de aplicaciones de Google Maps y de Google Earth. Por ejemplo, Eosnap (<http://www.eosnap.com/>) es un portal dedicado a la observación de la Tierra desde el espacio. Cada día, el sitio genera imágenes satelitales que resultan de indudable interés en el campo de las comunicaciones.

Muchos de estos recursos son híbridos (*mashup* o *remezcla*), es decir, sitios que usan contenidos de otras aplicaciones para crear nuevas funciones. Los *mashups* permiten la combinación de datos que existen en diferentes fuentes, como YouTube, Flickr, Google Maps y Google Earth, entre otras muchas.

El escenario en el que se desarrollan los *mashups* es increíblemente versátil y todos los días aparecen nuevos recursos. A estas alturas, cualquier reportero podría sentir algo de fastidio por el hecho

de que la mayoría de estas herramientas esté disponible en inglés. Ciertamente, el reportero que no domina ese idioma –para su lectura, al menos– tendrá dificultades para ponerse al día en el uso de estas tecnologías. No obstante, se están desarrollando aplicaciones híbridas en español, que tienen un enorme potencial. Entre esos *mashups* destacan:

- **Tagzania** (<http://tagzania.com/>): permite etiquetar lugares, por lo cual los usuarios forman una base de datos colaborativa de enorme potencial. Cualquier periodista podría sorprenderse cuando introduce el nombre de la ciudad donde vive, ya que tal vez encuentre más resultados de los que espera, como ocurrió en mi caso. Un recurso similar es Flof (<http://flof.com.ar/bin/home/>).

- **WikiLoc** (<http://es.wikiloc.com/wikiloc/home.do>): para marcar rutas, senderos, trayectorias de caminatas o ciclismo. Combina el uso del GPS, cuando se crean puntos de interés.

- **Woices** (<http://woices.com/>): encontrar o grabar audios geolocalizados.

- **Proyecto Simile** (<http://simile-widgets.org/>): es una cita obligada dentro del campo de la visualización de datos. Semantic Interoperability of Metadata and Information in Unlike Environments (Simile) es un proyecto de investigación que apunta a maximizar la interacción entre diferentes colecciones digitales. Buena

parte de la información contenida en bibliotecas digitales resulta invisible para los motores de búsqueda en línea. Por eso, este recurso agrupa una interesante variedad de herramientas que tratan de resolver ese inconveniente. Simile es un recurso desarrollado por The World Wide Web Consortium (www.w3c.org), con el apoyo del Instituto Tecnológico de Massachusetts (EE. UU.), entre otras organizaciones que participan en la iniciativa.

Las herramientas desarrolladas en Simile utilizan las tecnologías de web semántica para mejorar el intercambio de los recursos web, lo que ayuda a desbloquear el contenido de las bibliotecas digitales de todo el mundo.

Su recurso más conocido es Piggy Bank (http://simile.mit.edu/wiki/Piggy_Bank), una extensión de Firefox que permite introducir al navegador en la web semántica, facilitando el procesamiento de páginas webs, de tal modo que se logre recuperar información no visible desde la web. Una vez instalado, aparecerán dos iconos, que deberán añadirse a la barra de herramientas.

Pero no es el único. Exhibit (<http://www.simile-widgets.org/exhibit/>), por ejemplo, posibilita la creación de presentaciones de elementos mediante un carrusel, que puede ser personalizado gracias a la selección de varios parámetros.

Otra maravilla de este proyecto es Time Plot (<http://www.simile-widgets.org/timeplot/>), sistema que permite graficar estadísticas de manera original, dando la opción de insertar una línea

temporal mediante la incorporación de eventos concretos.

Por su parte, Citeline (<http://citeline.mit.edu/>) facilita la edición de bibliografías interactivas.

Condiciones necesarias

La respuesta a cuáles son las condiciones mínimas necesarias para procesar datos desde una sala de redacción y llegar a conclusiones útiles a los ciudadanos es compleja. Vivimos en un mundo en el que casi todo se expresa con números. Por lo tanto, el manejo de grandes masas de datos requiere ciertas habilidades por parte de quienes pretenden encontrar noticias relevantes detrás de ellos.

La misión no es sencilla: se tiene que encontrar un significado tangible a partir de lo abstracto, hay que contarle a la gente cómo esa gran maraña de números afecta su vida. Solo de ese modo se logra el impacto necesario para vivir en sociedades mejor informadas.

El periodismo de datos requiere entrenamiento permanente y conocimiento de materias generalmente alejadas de las preferencias del periodista convencional. Las fuentes aceitadas no son suficientes: hay que conocer el funcionamiento de la Administración Pública, saber interpretar leyes, decretos y ordenanzas. Si un periodista no comprende cómo funcionan las instituciones de su país, difícilmente podrá emprender el camino con éxito.

- Dominio del idioma inglés: la mayoría de las herramientas de uso frecuente en periodismo de datos no tiene corre-

lato al español. Actualmente, se está formando una brecha importante entre quienes acceden a estos recursos por su condición bilingüe y los que no.

- Habilidades desarrolladas en materia de búsquedas en la web: con usar Google de manera tradicional no alcanza. Hay que aprender a sacar mayor provecho de las búsquedas avanzadas, por formato, por dominio y por fecha; pero también, ante la carencia de una iniciativa nacional de datos abiertos, controlar los posibles sitios oficiales en los que hay datos.

El periodismo de datos requiere entrenamiento permanente

- Conocer los principios básicos del programa Excel o similar, capaz de procesar información numérica en hojas de cálculo.

- Tener conocimientos de matemática básica y de estadística, tanto descriptiva como inferencial: un periodista “anumérico” no tendrá la capacidad suficiente para enfrentar el proceso conocido como “la entrevista a los números”. Esto nos interesa especialmente cuando se trata de análisis de presupuestos públicos, impuestos y otros temas como criminalidad y contaminación ambiental.

El periodista se encuentra frente a un escenario desconocido. Hemos sido formados para un mundo en el que prevalece la carencia de datos, no para un mundo en el que hay sobreabundancia de

información. Los datos nos llegan desde varios frentes y en grandes cantidades.

No hay todavía instrucción formal sobre periodismo de datos en las universidades, salvo excepciones. Por lo tanto, los futuros periodistas de datos deberán invertir –básicamente, en tiempo– para afrontar su propia capacitación en esta disciplina.

El periodismo de datos implica un trabajo duro y, muchas veces, costoso. Hay que acceder a decenas, cuando no a cientos o a miles de planillas, textos, gráficos. Hay que recurrir, en ocasiones, a programadores para que diseñen aplicaciones capaces de capturar datos contenidos en páginas webs. El esfuerzo siempre será un componente con fuerte presencia en esta materia. A veces, la rutina es decepcionante, porque no siempre encontrará lo que busca. Tendrá que leer pesados expedientes, montañas de papeles, hacer llamadas que nadie le devolverá y llamar a puertas que probablemente no se abrirán. El periodista de datos siempre deberá considerar la posibilidad de que, en ocasiones, no hay noticias detrás de los datos. Aunque claro, eso no es lo más habitual.

Ahora que sabe lo peor del periodismo de datos, si desea seguir adelante, entonces está preparado para enfrentar el desafío; y, seguramente, lo logrará, porque en esta bendita profesión no brillan las personas con suerte, con buenas conexiones, ni siquiera con mentes brillantes: en esta tarea solo brillan los perseverantes. Ahí está la diferencia. ■



Razón y Palabra

ISSN: 1605-4806

octavio.islas@uhemisferios.edu.ec

Universidad de los Hemisferios

Ecuador

Renó, Luciana; Saad Corrêa, Elizabeth
Reportaje soportado por el computador: procedimientos y tecnología para el periodismo
de datos contemporáneo
Razón y Palabra, vol. 21, núm. 97, abril-junio, 2017, pp. 128-141
Universidad de los Hemisferios
Quito, Ecuador

Disponible en: <http://www.redalyc.org/articulo.oa?id=199552192008>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

**Reportaje soportado por el computador:
procedimientos y tecnología para el periodismo de
datos contemporáneo**

**Feature Supported by the Computer: Procedures and
Technology for Contemporary Data Journalism**

**Reportagem suportada pelo computador:
procedimentos e tecnologia para o jornalismo de dados
contemporâneo**

Luciana Renó

Universidad de São Paulo (Brasil)

luciana.lorenzi@gmail.com

Elizabeth Saad Corrêa

Universidad de São Paulo (Brasil)

bethsaad@gmail.com

Fecha de recepción: 1 de marzo de 2017

Fecha de recepción evaluador: 15 de marzo de 2017

Fecha de recepción corrección: 1 de mayo de 2017

Resumo

O texto aqui apresentado resulta da tese de doutorado desenvolvida pela primeira autora sobre o tema, que foi realizada a partir da metodologia quase-experimental. Os resultados oferecidos no artigo direcionam à prática do jornalismo de dados contemporâneo, que avança sobre os conceitos do tradicional modelo de reportagem assistida por computador, sendo agora suportado por ele. Espera-se, como texto, colaborar com o crescimento do

jornalismo de dados nas redações pois, ainda que seja uma tendência, ainda não está consolidado pelo desconhecimento sobre suas ferramentas e procedimentos.

Palavras-chave: Jornalismo; Jornalismo de dados; Ecologia dos Meios; Mídia e Tecnologia

Abstract

This text to present some results of a study developed for the first author, which was based on the quasi-experimental methodology. The results offers information for the practice of contemporary data journalism, which advances on the concepts of traditional model of computer-assisted reporting, and is now supported by it. It is hoped, as a text, to collaborate with the growth of data journalism in newsrooms, although it is a trend, not yet consolidated by the lack of knowledge about its tools and procedures.

Key Word: Journalism; Data Journalism; Media Ecology; Media and Technolgy.

Resumen

El texto aquí presentado resulta de la tesis doctoral de la primera autora sobre el tema, que ha sido desarrollada a partir de la metodología casi-experimental. Los resultados ofrecidos en el artículo direccionan a la práctica del periodismo de datos contemporáneo, que avanza en los conceptos del tradicional modelo de reportaje asistido por el computador, siendo ahora soportado por él. Si espera, con el texto, colaborar con el crecimiento del periodismo de datos en las redacciones pues, aunque sea una tendencia, todavía no está consolidado por desconocimiento sobre sus herramientas y procedimientos.

Palabras-clave: Periodismo; Periodismo de datos; Ecología de los Medios; Medios y Tecnología.

Introducción

El periodismo es, por si, una actividad que tiene como base del proceso la investigación e interpretación de datos para la construcción de una narrativa que represente estos datos interpretados. No existe periodismo sin investigación, y tampoco sin datos. Pero no es redundante decir periodismo de datos, pues entre todos los métodos periodísticos y sus respectivos géneros, hay uno que recibe ese nombre por dedicarse al estudio de los datos a partir de procedimientos muy peculiares que reúnen en un único espacio, con una técnica, actividades estadísticas procedentes de las ciencias exactas, con los procedimientos tradicionales y sociales del periodismo, además de adoptar cada vez más en sus procesos estructuras tecnológicas, como direcciona Meyer.

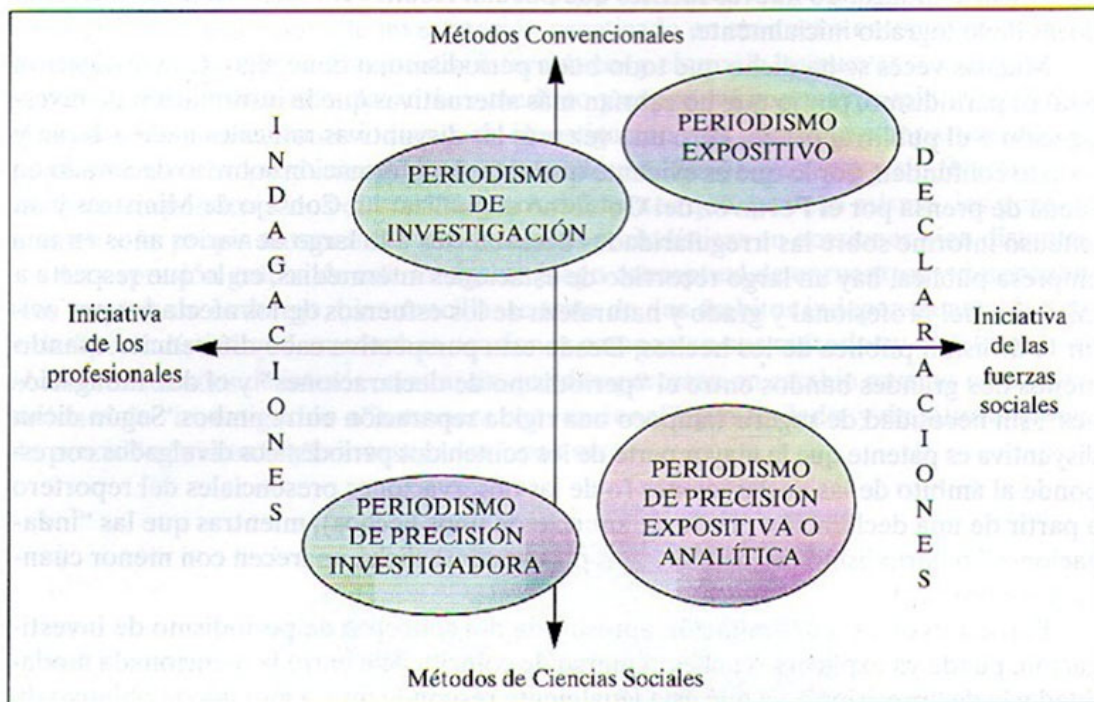
Meyer la define como “la aplicación de métodos científicos de investigación social y comportamental a la práctica del periodismo” (1989, p.196) y ya en 1973 había señalado que tales métodos eran, básicamente, el sondeo o encuesta de opinión, el experimento sociológico y el análisis de contenido. A ellos habría que agregar, según lo esbozado página atrás, las estrategias de búsqueda documental y el rastreo informático general de todo tipo de bases de datos, textuales o de listado alfanuméricos (Dader, 1997, p.20).

Dader defiende que el periodismo de datos es la evolución del periodismo de investigación. Según el autor, periodismo de datos o de precisión es muchas veces considerado exactamente como periodismo de investigación lo que, para Dader, es un equívoco, pues el periodismo de precisión es la evolución del periodismo de investigación. Es más allá.

(...) cada vez que comenzaba y comienzo una intervención sobre “periodismo de precisión” ante a un auditorio nuevo, tanto los asistentes con sus preguntas, como los propios promotores del acto en su presentación, no cesan de referirse a todo lo expuesto como cuestiones de “periodismo de investigación”. En las primeras ocasiones me resultaba un tanto desalentador que, tras llevar quizá una hora hablando sobre el periodismo de precisión y sus potencialidades, cuantos me interpelaban siguieran refiriéndose a todo lo supuestamente descrito por mí como asuntos de “periodismo de investigación”. (Dader, 1997, p.25)

Esa justificación de Dader es apoyada en la idea de que todo periodismo es de investigación, aunque algunos dedican su tiempo con procedimientos más detallados del estudio, además de adoptar técnicas y herramientas diferentes. Lo que ocurre es que, además de esas herramientas y procedimientos específicos, el periodismo de precisión tiene otras más, que garantizan o por lo menos disminuyen las posibilidades de errores provocadas por las características naturales de la psicología humana, que Dader (1997, p.25) denomina “esquemas y encuadres cognitivos”.

Gráfico 1 – Métodos periodísticos



Fuente: Dader, 1997, p.28

En el gráfico anterior, Dader presenta una visualización que corresponde exactamente con la relación entre periodismo de investigación y periodismo de precisión o de datos. En realidad, percibimos que el periodismo de datos ofrece resultados analíticos y no solamente de inspiración del periodista, de impresiones personales, aunque apoyado en técnicas y procedimientos profesionales.

El periodismo de datos también es practicado a partir de la adopción de diversos procedimientos y métodos. Uno de ellos, actualmente lo más eficaz y preferido por los periodistas de datos, es el RAC – Reportaje Asistido por Computador, que tiene como técnica fundamental el cruce de datos y estadísticas a partir de comandos desarrollados en el software Excel, pero también en conjunto con otros programas de computador desarrollados específicamente para esa actividad.

La difusión del RAC es intensa en los Estados Unidos, donde solo una entidad – el Instituto Nacional para el Reportaje Asistido por Computador (NICAR) – había preparado, hasta el inicio de 1999, 12 mil reporteros en técnicas de investigación computadorizada. (...) Utilizando el RAC, por ejemplo, reporteros probaran (e ilustraran con tablas y gráficos) que una tercera parte de la población del Estado de Nueva Jersey vive en áreas en que la contaminación excede en hasta 20 veces el límite recomendable. (Lage, 2003, p.162)

Lo que pasa es que, aunque los datos sean procesados en ambientes digitales, su recolección sigue por los caminos tradicionales, es decir, sin adoptar una de las posibilidades de reunión de datos en ambientes digitales. Por esa razón, es importante definir ambientes donde existan algoritmos dedicados a dar soporte a esos procedimientos laborales.

Cualquier observador puede decir, sin esfuerzo, que la introducción de los computadores ha modificado mucho la práctica del periodismo. Alguno que estudie bien el asunto, mientras tanto, concluirá que esa modificación es más profunda que se parece al primer vistazo y que el proceso de cambios está lejos de terminar; en verdad, promete tornarse permanente. (Lage, 2003, p.153)

Mientras tanto, es visible y seguro que por más que se construyan espacios digitales para apoyar la labor del periodismo de datos, jamás estos aparatos tecnológicos van a ocupar el espacio que actualmente está el periodista. Los procesos y ambientes digitales sirven para apoyar, complementar el trabajo periodístico, y no para ejercer su trabajo de manera autónoma, pues la máquina no tiene algo que el periodista debe tener, siempre: sensibilidad.

Los medios sociales y la noticia: nuevo escenario

Cuando pensamos en medios sociales y noticia, la primera cosa que viene a la cabeza es la circulación de la información. En realidad, los medios sociales son una eficaz herramienta para promocionar la circulación de información. Pero no está limitada a eso, especialmente cuando el tema Big Data es considerado.

Podemos considerar que el *Big Data* es una estrategia de investigación de datos en la web a partir de algoritmos apropiados para el cruce de datos sucios y limpios. La labor de cruzar los datos sigue un parámetro que contiene, además de los códigos de programación y la definición de tareas y secuencias, la definición de palabras clave o términos que son buscados. Eso ocurre, por ejemplo, al empezar una búsqueda en Google, pues definimos las palabras y el programa contesta con los resultados sobre los temas definidos. La diferencia es que Google ofrece, normalmente, datos limpios y disponibles, oficiales o no, mientras el *Big Data* propone que se siga más allá de lo limpio.

En este sentido, Manovich (2013) presenta la idea de que hoy en día el software sigue al mando, es decir, necesitamos del software para entender y desarrollar tareas en la web, no solamente para descubrir informaciones, pero también para realizar procedimientos personales o profesionales en la nube. El autor propone que:

Convertir todo en datos y usar algoritmos para analizarlo, cambia lo que significa conocer algo. Crea nuevas estrategias que, unidas, crean una epistemología del software. La epistemología es una rama de la filosofía que se interroga sobre lo que es el conocimiento, como se adquiere y en qué medida un sujeto puede ser conocido. El código digital, la visualización de datos, los GIS, la búsqueda de información, las técnicas de

aprendizaje automático, el constante incremento de la velocidad de procesadores y el decremento de sus costos, las tecnologías de análisis de *big data*, los medios sociales y otras partes del moderno universo tecno-social introducen nuevas formas de adquisición de conocimientos y, en el proceso, redefinen lo que es el conocimiento (Manovich, 2013, pp.337–338)

Manovich propone el desarrollo de algoritmos para obtener resultados concretos a partir del concepto de Big Data. El trabajo de búsqueda en la nube es posibilitado por los algoritmos desarrollados específicamente para la tarea.

Por ejemplo, siempre es posible inventar nuevos algoritmos (o nuevas formas de escalar algoritmos existentes para analizar más rápido el *big data*) que puedan analizar los datos existentes de hoy en formas que los algoritmos previos no podían. Como resultado, podemos extraer patrones adicionales y generar nuevas informaciones de los datos viejos ya analizados. (Manovich, 2013, p.338)

El autor también hace una discusión sobre la relación entre los usuarios y los medios sociales. Para Manovich, la utilización de estos espacios es casi libre y los usuarios publican lo que deciden publicar. Eso es un espacio donde informaciones pueden surgir de manera natural. Según el autor:

Los desarrollos de los años 90 se han diseminado a cientos de millones de personas que están escribiendo blogs, subiendo fotos y videos a los sitios sociales, y usando de forma libre (o casi) herramientas de software de producción y de edición que hace algunos años costaban decenas de miles de dólares. (Manovich, 2013, p.1)

De hecho, los medios sociales son responsables por el descubrimiento de informaciones diversas por parte de los periodistas contemporáneos. Actuar como periodista sin tener una cuenta en los diversos medios sociales es como trabajar sin máquina de escribir en los años 1940. Las noticias fluyen en los medios sociales, y los ciudadanos tienen interés en hacer circular, como defiende Gillmor (2005), para quien los ciudadanos contemporáneos son los “seres medio”.

Adoptar estrategias de construcción de la noticia a partir de la recolección de informaciones desde los medios sociales es común en el periodismo internacional, especialmente en situaciones extremas. En el año 2012, los periodistas estadounidenses cubrieron la trayectoria (y la destrucción) del huracán Isaac a partir de informaciones que circulaban en los medios sociales, especialmente el Twitter y el YouTube. A partir de eso, algunos periodistas pasaron a definir esa recolección de informaciones como periodismo híbrido, por una mezcla de actividades y procedimientos en la búsqueda de la información, como propone la periodista Gina Masullo Chen en un texto publicado en Nieman Lab. Según Chen (2012), “La mezcla de medios me ofreció – a quien acababa de mudarse de un país de huracanes tres semanas antes de la tormenta – una experiencia multimedia, que yo llamaría, incomparable con las viejas épocas de medios impresos, televisión y radio”. Eso ocurrió debido a la gran cantidad de informaciones en los medios sociales, muchos de ellos con la documentación audiovisual, lo que fue posible la

comparación entre diversas noticias sobre el tema y la publicación segura por parte de los medios tradicionales. Sin embargo, fue una superación posible para algunos periodistas, especialmente los que estaban acostumbrados con esa nueva realidad mediática y social, donde la noticia circula entre las personas y los periodistas asumen el papel de descubrir y comprenderla.

En otro momento de 2012, también en los Estados Unidos, el periódico *The New York Times* creó una página web distinta para el cubrimiento de la destrucción provocada por otro huracán – el Sandy –, que ha dejado Nueva York y Nueva Jersey sin accesibilidad, además de la falta de energía y de comunicación convencional. Para superar la falta de noticia por problemas incluso de accesibilidad, el periódico construyó un mapa interactivo con los contenidos publicados por los ciudadanos a partir de sus dispositivos móviles y de la tecnología *Foursquare* para descubrir se la localización y el registro eran realmente donde y cuando decían los usuarios. De esa manera, fue producido un mapa de Manhattan y región con los puntos clave y las imágenes ofrecidas por los ciudadanos. Fue otro paso adelante para la construcción de la noticia a partir de medios sociales y de la participación ciudadana.

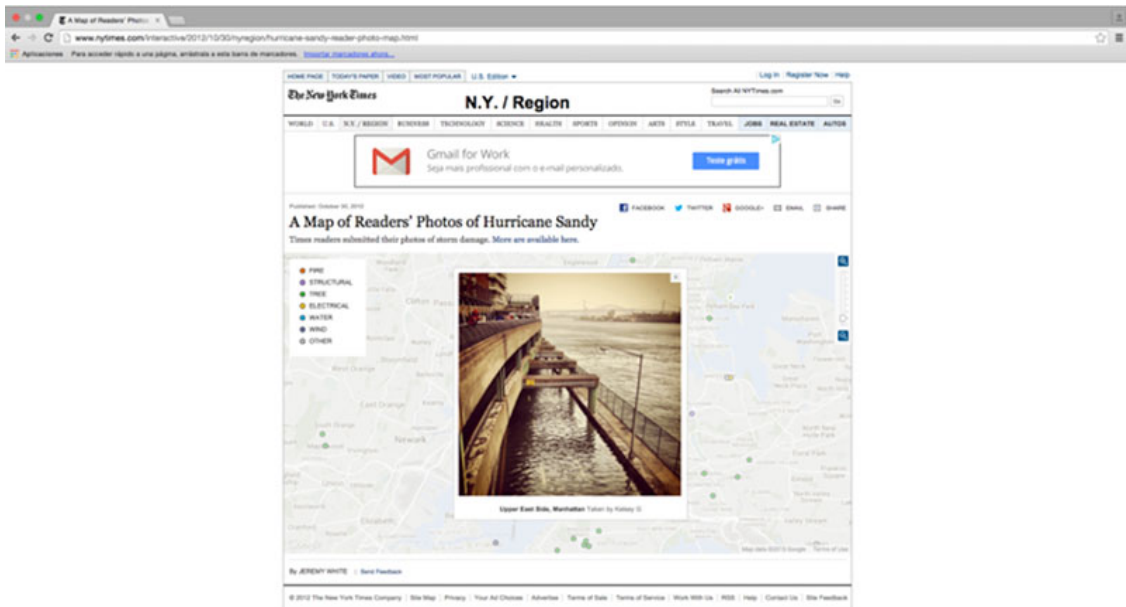
La página, bautizada como *A Map of Readers' Photos of Hurricane Sandy*¹, fue publicada como parte de otras informaciones complementarias, como mapas interactivos sobre la condición climática de la región, la trayectoria del huracán Sandy por los Estados Unidos, etc. Sin embargo, la publicación con mejor audiencia fue la de fotos, que ofreció un registro visual por parte de los ciudadanos, ampliando la circulación del contenido de manera viral.

Gráfico 2 – Manhattan al fondo parcialmente sin luz



Fuente: Reeve Jolliffe²

Gráfico 3 – El agua sobre la calle, en Manhattan



Fuente: Kelsey G.³

Sin embargo, esos cambios fueron posibles gracias a una redacción mixta, compuesta por profesionales del periodismo y de la tecnología de la información. Sólo así las ideas surgieron y ganaron cuerpo, tornándose reales y posibles. Eso es una prueba de que es fundamental compartir espacios, especialmente hoy, cuando la ecología mediática gana nuevo formato y pasa a ser un complejo y amplio mundo diverso.

Mapas interactivos

Uno de los diferenciales del periodismo de datos es la visualización de los mismos. Por tanto, se presenta aquí un estudio originalmente desarrollado en 2013 por la autora y que ofrece, para la tesis, una actualización de las informaciones. Se trata de una reinterpretación necesaria para se comprender e imaginar los mapas interactivos para la labor periodística de datos.

Los mapas interactivos son históricamente usados para representación de territorios de forma simplificada, o sea, más facilidad de comprensión, pero con una función informativa a partir de caminos hipermediáticos y/o hipertextuales, dependiendo de la capacidad creativa del autor del mismo. Pueden existir diversas informaciones en mapas interactivos, como, por ejemplo, datos geográficos, físicos, políticos, sociales o las informaciones apoyadas en contenidos audiovisuales, pues es posible compartir fotos, videos y audio en estos íconos interactivos. La información es construida a partir del cruce de datos y normalmente son adoptados en conjunto con otras plataformas de comunicación de manera que la información queda comprensiva. Por tanto, es necesario obtener los datos para, enseguida, tratarlos y construir las plataformas.

Según el video presentado por el español Carlos Martínez de la Serna durante taller sobre mapas interactivos con elementos visuales (realizado por *Knight Center Foundation*, en 2012), la iconografía y el uso de una tipografía apropiada y la diferenciación de colores deben facilitar la identificación y la jerarquía de los datos. Es fundamental garantizar al lector un mapa claro y preciso.

Para fortalecer el debate, Edward Tufte (1990) dijo, ironizando a Paul Klee, que la relación de color con la información es tan básica y sencilla como la técnica de color en el arte, o sea, la pintura es simplemente la colocación de color en el lugar correcto. Y más, que poner un buen color en el punto correcto es una cuestión compleja, pero que valora la información. En realidad, esa colocación de color es tan difícil y sutil cómo evitar una catástrofe comunicacional, pues puede ocultar o destacar una información importante. Pero es importante aclarar que la principal función del mapeamiento en la comunicación periodística es informar sobre algo a través de un mapa para que el lector pueda encontrar un asunto, o sea, identificar un evento, en especial cuando se trata de lugares desconocidos y ubicados lejos de las grandes ciudades.

Existe una necesidad intrínseca en los mapas de transferir la información de forma eficaz para que el lector pueda interpretar los datos correctamente. La gran cantidad de información geográfica en formato digital le proporciona al lector una mejor profundidad en los datos referentes al lugar presentado, o al tema. El mapa interactivo también es usado frecuentemente como apoyo referencial geográfico para la construcción de noticias, permitiendo al usuario la navegabilidad en la información. Pero estos recursos todavía son poco explorados por los medios de comunicación.

La interactividad en los mapas interactivos también está relacionada con la navegabilidad dentro de ellos, interactuando con las informaciones relatadas. Haciendo un clic sobre estos puntos informativos, es obtenida la información de manera destacada. Otra manera de interactividad con los mapas es la posibilidad de inserción de datos, que pueden ser ofrecidos para los usuarios con o sin filtros y mediaciones. De esa manera, el usuario se transforma en un coautor de la información presentada. Mapas interactivos contemporáneos ofrecen una manera de participar en la lectura de la noticia, aproximando el proceso de un momento recreativo que desempeña un papel social e informativo con un proceso de absorción más eficaz, aunque discreto. De lo contrario, no se puede ponerlos en la ecología mediática contemporánea, o considerarlos como parte de una narrativa transmedia.

Actualmente, algunos contenidos ofrecen navegación a partir de una visualización geográfica. La importancia de los mapas interactivos no puede ser ignorada por el periodismo, especialmente su lenguaje apoyado en la gamificación, que explora de manera expresiva la interactividad y su visualización de contenido. Ese lenguaje bien definido permite la utilización de mapas de manera analítica por parte del público, una vez que esas plataformas ofrezcan datos que ayuden en la interpretación de los mismos y en la construcción de la opinión pública, lo que es el verdadero papel del periodismo.

Procedimientos de periodismo de datos

Para la realización del periodismo de datos, es necesario conocer en la práctica las más importantes herramientas para su desarrollo, así como sus procedimientos. El procedimiento usual para recopilación de datos es:

Búsqueda → Extracción/Limpieza → Análisis → Visualización

Búsqueda

La manera más eficiente en trabajar con búsqueda es utilizando la herramienta del Google⁴. Esta nos permite hacer trabajos simples como búsquedas directas en la caja de texto que aparece en el sitio como por ejemplos cálculos aritméticos, cconversor de monedas y medidas, reloj mundial, listado informaciones de personas públicas en un cuadro a la derecha de la pantalla, entre otras cosas.

También es posible hacer búsquedas avanzadas que contiene mayores recursos. A través de ella se puede usar varios filtros como por ejemplos por palabras, palabras o frase exactas, eliminar palabras, formatos de los documentos (PDF, CSV, etc.), últimas 24hs, en tiempo real, sitio o dominio, palabra-clave, idioma, región, etc.

Otra utilización de la herramienta de búsqueda del Google es la búsqueda por imágenes. Utilizando una imagen deseada o una *url* es posible buscar informaciones sobre ella. También es posible hacer un filtro por tipo de imagen (cara, fotografía, animada, etc.), colores, formato (JPG, BMP, GIF, etc.), tamaño, derechos de uso, entre otras posibilidades.

Extracción

A través de algunas herramientas de extracción de datos es posible obtener un contenido deseado que está en un formato no adecuado pasando para una otra extensión deseada. Por ejemplo, extraer texto de imágenes, pasar una tabla que está en PDF para XLS u otra extensión deseada.

Herramientas de extracción de imágenes:

1. Zamzar⁵
2. Free OCR⁶
3. ABBYY FineReader Online⁷
4. Tabula⁸

Herramientas de extracción de tablas y/o listas:

1. Google Sheets⁹
2. IFTTT¹⁰
3. Web scraper¹¹

Limpieza

1. Google Sheets¹²
2. Data Wrangler¹³
3. Open Refine¹⁴

Análisis

Después del proceso de búsqueda, extracción y limpieza de las informaciones deseadas, ahora es hora de hacer el análisis de los datos. En realidad, ahí puede estar la

noticia, lo que interesa al periodismo de datos. La herramienta utilizada para eso procedimiento es Google Sheets, que es una herramienta completa para esta etapa. Es posible trabajar con diversos procesos para analizar los datos, como por ejemplo, clasificación, fórmulas y funciones, filtros, tablas dinámica y gráficos

Visualización

Visualización de datos es la representación visual de las informaciones para mostrar evidencias, permitir análisis e informar puntos fundamentales para el reportaje. En los procesos de visualización se puede presentar verdades ocultas en datos complejos. La utilización de gráficos permite la comparación entre los contenidos y construye de manera sencilla la transmisión de informaciones. Las destacadas herramientas utilizadas son:

1. Infogr.am¹⁵
2. Timeline.js¹⁶
3. Silk¹⁷
4. Tableau Public¹⁸
5. Carto DB¹⁹
6. Odyssey.js²⁰
7. Storymap²¹

A partir de estas posibilidades, es posible construir contenidos complejos con calidad profesional. En realidad, reportajes profesionales trabajan con estas herramientas y alcanzan resultados sólidos y confiables. Por esa razón, si espera colaborar con el desarrollo del periodismo de datos en un escenario contemporáneo.

Conclusiones

El uso de las tecnologías en el periodismo ha sido, en toda su historia, una mezcla de facilitación y dificultad. La facilidad en el desarrollo de la profesión en sus distintas maneras fue positivo en todos sus momentos. Por otra parte, la dificultad ha sido relacionado a la superación y el aprendizaje de los periodistas que tenían que descubrir nuevos métodos y técnicas para ajustarse a estas tecnologías nuevas. En el campo del periodismo en base de datos, conocido por periodismo de datos, el desarrollo de las tecnologías ha sido fundamental.

Aunque la sociedad no esté totalmente digitalizada, es posible determinar una tendencia de digitalización creciente de contenidos y de la relación entre medios y ciudadanos, lo que actualmente implica un cambio en la brecha digital.

La sociedad contemporánea está en constante transformación y la conformación de nuevas estructuras basadas en lo digital (estructuras líquidas) son las que sostienen los procesos sociales. En este sentido, el espacio digital usado para desarrollo del periodismo de datos, apoyado por la participación social, es una realidad.

Entre los desarrollos tecnológicos, pasamos por el método RAC, que ha facilitado la búsqueda por informaciones periodísticas, y en el periodismo de datos eso se sobresa, específicamente por su mezcla con otros saberes. El periodismo de datos contemporáneo es vuelve a los orígenes formativos como profesión interdisciplinaria en términos de filosofías y rutinas. Esto constituye, además, una tendencia creciente del ecosistema mediático vigente en donde el conocimiento sobre estadística, algoritmos y lógica deben componer el currículo actual en la formación del periodista de nuevo perfil. Esta disyuntiva ha ocurrido en otros momentos históricos de la comunicación, como la radio, la televisión e internet, pero ahora esta incorporación e integración de conocimientos ocurre en el proceso de producción de las noticias y no sólo con las posibilidades tecnológicas ofrecidas.

Esperamos, con el texto, poder colaborar con el crecimiento de la práctica del periodismo de datos en las redacciones pues, aunque sea una tendencia, todavía no está consolidado por desconocimiento sobre sus herramientas y procedimientos.

Bibliografía

- Chen, G. M. (2012). Hurricane Isaac coverage shows the promise of hybrid model for news. *Nieman Lab*. Disponible en: <http://www.niemanlab.org/2012/08/hurricane-isaac-coverage-shows-the-promise-of-a-hybrid-model-for-news>. Consultado el 10/01/2015.
- Dader, J. L. (1997). *Periodismo de precisión – vía socioinformática de descubrir noticias*. Madrid: Síntesis.
- Gillmor, D. (2005). *Nós, os media*. Lisboa: Editora Presença.
- Lage, N. (2003). *A reportagem*. Rio de Janeiro: Record.
- Manovich, L. (2013). *Software takes command*. New York/London: Bloomsbury.
- Tufte, E. (1990). *Envisioning Information*. Cheshire: Graphics Press.

Notas

¹ La página fue producida en conjunto con tecnología *Foursquare*. Disponible en <http://www.nytimes.com/interactive/2012/10/30/nyregion/hurricane-sandy-reader-photo-map.html>. Consultado el 19/02/2015.

² Disponible en <http://www.nytimes.com/interactive/2012/10/30/nyregion/hurricane-sandy-reader-photo-map.html>. Consultado el 13/10/2015.

³ Disponible en <http://www.nytimes.com/interactive/2012/10/30/nyregion/hurricane-sandy-reader-photo-map.html>. Consultado el 13/10/2015.

⁴ www.google.es

⁵ www.zamzar.com

⁶ Reconocimiento Óptico de Caracteres. www.free-ocr.com

⁷ <http://finereaderonline.com/en-us>

⁸ <http://tabula.technology/>

⁹ <https://www.google.com/sheets/about/>

¹⁰ <https://ifttt.com/>

¹¹ <https://chrome.google.com/webstore/detail/web-scraper/jnhgnonknehpejjehhllklipmbmhn>

¹² <https://www.google.com/sheets/about/>

¹³ <http://vis.stanford.edu/wrangler/app/>

¹⁴ <http://openrefine.org/>

¹⁵ <https://infogr.am/>

¹⁶ <http://timeline.knightlab.com/>

¹⁷ <https://www.silk.co/>

¹⁸ <http://public.tableau.com>

¹⁹ <https://cartodb.com/>

²⁰ <http://cartodb.github.io/odyssey.js/>

²¹ <https://storymap.knightlab.com/>