

Reconstrucción y análisis de la base de datos de jurisprudencia de cuantificación de daños de la Cámara Nacional de Apelaciones en lo Civil

Gonzalo R. Ramírez Muñoz de Toro*

*«For the rational study of the law
the black-letter man may be
the man of the present,
but the man of the future
is the man of statistics
and the master of economics».*

Oliver W. Holmes Jr.,
jurista y exjefe de la Corte Suprema de los
Estados Unidos de América, 1897.

RESUMEN

El análisis jurisprudencial se ha basado tradicionalmente en el estudio de decisiones judiciales de modo concreto e inconexo, mientras que el derecho ha sido renuente en adoptar estrategias de empleo de datos. Para este trabajo, he desarrollado software de raspado web y análisis estadístico a efectos de recolectar y normalizar datos relativos a más de veinte mil sentencias de cuantificación de daños de la Cámara Nacional de Apelaciones en lo Civil, para el análisis exploratorio que aquí se emprende. Se ha trazado un perfil general de siniestros y víctimas. Se otorga libre acceso a las bases de datos resultantes para su uso en investigaciones científicas o académicas sin fines de lucro.

Palabras clave: raspado web, datos abiertos, jurisprudencia, daños, estadística.

ABSTRACT

Jurisprudential analysis has been traditionally based on the study of judicial decisions in a concrete and unrelated fashion, while Law has been reluctant to adopt data-driven strategies. For this paper, I have developed web scraping and statistical analysis software to harvest and normalize data of more than twenty thousand rulings where damages have been awarded by Argentina's National Civil Court of Appeals, for the exploratory analysis that is undertaken herein. A general torts and victims profile has been made. Free access is granted on the resulting databases for its non-profit usage on scientific or academic research.

Key words: web scraping, open data, jurisprudence, torts, statistics.

* Abogacía, Departamento de Derecho, Universidad Nacional del Sur (Bahía Blanca, Provincia de Buenos Aires, República Argentina).

Agradezco al Dr. Hugo A. Acciarri por haber dirigido este trabajo, así como por sus valiosos consejos. Va de suyo que todos los desatinos con que pudiera importunar al clemente lector han de ser achacados exclusivamente a mi persona.

Asimismo, se agradece el envío de comentarios y sugerencias por vía de correo electrónico a grrmdt@gmail.com.

§ 1. Introducción

El Poder Judicial de la Nación publica en su sitio web fichas de datos de sentencias en que se cuantifican daños emitidas por las diversas salas que componen la Cámara Nacional de Apelaciones en lo Civil¹. Los datos son recolectados y publicados por el Centro de Datos Informatizados de la Excelentísima Cámara Nacional de Apelaciones en lo Civil (Sistema Judicial para Cuantificar los Daños a la Persona) y abarcan precedentes judiciales desde el año 1993 hasta la actualidad. Los usuarios pueden efectuar consultas a través de diversos formularios de búsqueda provistos por el mencionado sitio, las que se canalizan a cinco bases de datos, a saber: Fallecimiento (valor vida), Lesiones, Daño extrapatrimonial, Responsabilidad de los profesionales de la salud y Daños derivados de violencia hacia la mujer. Todas ellas se encuentran divididas en dos grandes áreas temporales de búsqueda², según se trate de sentencias de primera instancia anteriores o posteriores a agosto de 2015. Los resultados se presentan en el lenguaje de marcado HTML (siglas en inglés para *HyperText Markup Language*, “Lenguaje de Marcado de HiperTexto”), no pueden ser descargados por el usuario ni se hallan rígidamente estructurados, documentados o mínimamente ambiguos; al mismo tiempo, se presentan sin metadatos que permitan su adecuada categorización, procesamiento, comparación, análisis estadístico y representación gráfica³.

A fin de superar estos obstáculos, he desarrollado para este trabajo dos programas informáticos. Uno de ellos, de tipo *web scraper* (“raspador web”), fue utilizado para descargar masivamente y de forma automática las mencionadas fichas de sentencias de cuantificación de daños; el otro, de análisis estadístico, fue usado para la corrección y normalización de la información recolectada, así como para su posterior análisis y visualización mediante gráficos.

En el presente trabajo propongo explicar someramente la metodología utilizada y analizar la información obtenida. Se trazará un perfil estadístico de las bases de datos reconstruidas, así como de las víctimas involucradas para, finalmente, analizar con algo más de detalle la base de datos Lesiones, por ser la más nutrida de las aquí recopiladas. La totalidad de la información recabada será publicada para facilitar su uso en investigaciones y proyectos científicos y académicos sin fines de lucro.

§ 2. Datos: estructura, análisis y *web scraping*

Según el Instituto Nacional de Estándares y Tecnología⁴ del Departamento de Comercio de los Estados Unidos de América, se define como “datos estructurados” (*structured data*) a “una *organización de información*, usualmente en memoria, para una mayor eficiencia de algoritmos, [...] o una *unidad conceptual*, tal como el nombre y dirección

¹ Las mismas pueden consultarse en <http://consultas.pjn.gov.ar/cuantificacion/civil/>.

² Si bien el criterio utilizado para dividir las series temporales no se encuentra explicitado, cabe presumir que el mismo se basa en la fecha a partir de la cual ha entrado en vigor el Código Civil y Comercial de la Nación. Si bien originariamente se dispuso que ese cuerpo legal regiría tan solo a partir del 1 de enero de 2017 (art. 7, ley 26.994), su vigencia fue anticipada al 15 de agosto de 2015 (art. 7, ley 27.077).

³ Ello implica una violación a la Carta Internacional de Datos Abiertos (<https://opendatacharter.net/principles-es/>), que establece diversos principios y buenas prácticas para la liberación de datos abiertos gubernamentales.

⁴ *National Institute of Standards and Technology, U. S. Department of Commerce* (<https://www.nist.gov>).

de una persona...”⁵ (el énfasis pertenece al autor de este trabajo). Los datos estructurados, por su propia naturaleza, son inmediatamente aptos para la realización de búsquedas, análisis y procesamiento con el auxilio de recursos informáticos apropiados. Así acontece, por ejemplo, con la información que podemos encontrar en bases de datos relacionales, normalmente administradas mediante el uso del lenguaje SQL, o bien en hojas de cálculo (Microsoft Excel, Google Spreadsheets, Apple Numbers, LibreOffice Calc, etc.), las que suelen exhibir sus registros como conjuntos de datos *organizados* en filas y columnas en torno a una *unidad conceptual*.

A pesar del diluvio de datos⁶ que experimentamos en nuestra denominada “sociedad de la información”, la mayor parte de aquellos que podemos encontrar en la web no se encuentra en formatos rígidamente estructurados, bien documentados ni mínimamente ambiguos, lo que dificulta su procesamiento automático mediante el uso de herramientas de software e impone al investigador un campo de acción reducido. Si se pretende acceder a algún tipo de abstracción (estadística) y, así, a un nuevo nivel de análisis “macrojurídico”, se debe, como paso previo, hallar alguna estrategia que permita una adecuada estructuración de la información.

Debe descartarse, por inviable, la estrategia manual tradicional, en tanto ella exigiría contar con una nutrida dotación de personal destinada a la interpretación y adecuada categorización de los datos. El tiempo que tal tarea demandaría, combinado con el tedio y la proclividad a errores que llevaría aneja, nos obligan a buscar otra alternativa. Es por ello que, necesariamente, debemos acudir a procedimientos automatizados de estructuración de datos.

En las disciplinas informáticas se entiende por *data scraping* (“raspado de datos”) a:

*... una técnica en la que un programa de computadora extrae datos de una respuesta legible por seres humanos proveniente de otro programa.*⁷

⁵ La entrada completa de la expresión entiende por “datos estructurados” a “una organización de información, usualmente en memoria, para una mayor eficiencia de algoritmos tales como cola, pila, lista enlazada, montículo, diccionario y árbol, o una unidad conceptual, tal como el nombre y dirección de una persona. Puede llegar a incluir información redundante, como el tamaño de una lista o el número de nodos en un subárbol” (“*An organization of information, usually in memory, for better algorithm efficiency, such as queue, stack, linked list, heap, dictionary, and tree, or conceptual unity, such as the name and address of a person. It may include redundant information, such as length of the list or number of nodes in a subtree*”), entrada “*Data structure*” en BLACK, Paul E. y PIETERSE, Vreda (editores), *Dictionary of Algorithms and Data Structures [online]*, disponible en el sitio web oficial del *National Institute of Standards and Technology* en: <https://xlinux.nist.gov/dads/HTML/dataStructure.html>.

⁶ La capacidad tecnológica mundial para almacenar información alcanzó en 2014 los 4,6 zettabytes óptimamente comprimidos que, de ser almacenados en CD-ROM de 730 megabytes y 1,2 milímetros de espesor, obligarían a construir veinte pilas de discos con una altura tal que cada una de ellas llegaría al sol. Al respecto, véase: HILBERT, Martin, “Quantifying the Data Deluge and the Data Drought”, *Social Sciences Research Network* (SSRN), disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2984851> (documento digital, p. 3 y ss.).

⁷ “*Data scraping is a technique in which a computer program extracts data from human-readable output coming from another program*”, entrada “*Data scraping*” en AA. VV., *Wikipedia*, disponible en el sitio web oficial de la enciclopedia *Wikipedia*: https://en.wikipedia.org/wiki/Data_scraping.

Asimismo, se ha dicho que:

Normalmente, la transferencia de datos entre programas es lograda mediante el uso de estructuras de datos aptas para el procesamiento automático por computadoras, no personas. Tales formatos y protocolos de intercambio se encuentran típicamente estructurados de forma rígida, bien documentada, fácilmente procesable y mantienen la ambigüedad al mínimo. Muy frecuentemente, estas transmisiones no son en absoluto legibles por un ser humano. [...] En consecuencia, el elemento clave que distingue el raspado de datos del procesamiento regular es que la respuesta a raspar se hallaba destinada a ser presentada a un usuario en lugar de servir como entrada para otro programa, y es por ello que usualmente no se encuentra bien documentada ni estructurada para un adecuado procesamiento...⁸.

A su vez, y como especie del género *data scraping*, se ha afirmado que *web scraping* (“raspado web”, también conocido como “cosecha web” o “extracción de datos web”) es:

... raspado de datos usado para la extracción de datos provenientes de páginas web. El software de raspado web puede acceder a la World Wide Web directamente usando el Protocolo de Transferencia de Hipertexto, o por medio de un navegador web. Mientras el raspado web puede ser realizado manualmente por un usuario de software, el término habitualmente refiere a los procesos automatizados implementados mediante el uso de un bot o web crawler. Es una forma de copiar en la que los datos específicos son recolectados y copiados desde una web, frecuentemente en una base de datos u hoja de cálculo, para su recuperación o análisis posterior...⁹.

Los programas de raspado web usualmente recorren e interpretan el código fuente de páginas web objetivo a fin de recolectar y almacenar los datos allí cautivos, de forma tal que se vuelvan, estructuración mediante, inmediatamente aptos para su procesamiento automático con el auxilio de recursos computacionales.

⁸ “Normally, data transfer between programs is accomplished using data structures suited for automated processing by computers, not people. Such interchange formats and protocols are typically rigidly structured, well-documented, easily parsed, and keep ambiguity to a minimum. Very often, these transmissions are not human-readable at all. [...] Thus, the key element that distinguishes data scraping from regular parsing is that the output being scraped was intended for display to an end-user, rather than as input to another program, and is therefore usually neither documented nor structured for convenient parsing [...]”, AA. VV., *ibid*.

⁹ “Web scraping (web harvesting or web data extraction) is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis”, entrada “Web scraping” en AA. VV., *Wikipedia*, disponible en el sitio web oficial de la enciclopedia *Wikipedia* en: https://en.wikipedia.org/wiki/Web_scraping.

3. *Web scraping* en otras disciplinas

Si bien el uso de la metodología del *web scraping* es reciente debido a la disponibilidad técnica de internet, la web y software adecuado, podemos encontrar diversos antecedentes en los que se analizan datos recolectados gracias a la misma.

En el área de los estudios políticos se la ha utilizado para estudiar casos en que el gobierno estadounidense discutió sobre la democratización en otros estados, o bien consideró o impuso medidas de política exterior relativas al liberalismo en el extranjero¹⁰. En otro ámbito, se la ha aplicado también para determinar probabilidades de victoria y estrategias de apuestas durante el desarrollo de un juego de cricket¹¹.

Sin duda alguna, el campo que ha sido más fecundo en resultados ha sido el de las disciplinas económicas. Robert M. Groves señala que la recolección de datos por medio del *web scraping* goza de un costo marginal cercano a cero en algunos casos y de inmediatez respecto de los hechos a los que tales datos se refieren¹². Por su parte, Ainslie Restieaux aconseja explorar su uso para mejorar la metodología tradicional utilizada para la realización del Índice de Precios al Consumidor (IPC) que lleva adelante la *Office for National Statistics* del Reino Unido¹³, tarea que acometieran luego Robert Breton y otros¹⁴. Finalmente, Ben Powell y otros lidian con los inconvenientes derivados de datos faltantes en los relevamientos realizados mediante *web scraping* a través de diversos criterios de imputación, a fin de alcanzar la conformación de series de precios completas

¹⁰ ELSHEHAWY, Ashrakat, MARINOV, Nikolay y NANNI, Federico, “Quantifying Attention to Foreign Elections with Text Analysis of U.S. Congress and the Presidency”, *Social Sciences Research Network* (SSRN), disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2981486> or <http://dx.doi.org/10.2139/ssrn.2981486> (documento digital, p. 8 y ss.).

¹¹ NORTON, Hugh, GRAY, Stephen y FAFF, Robert W., “Yes, One-Day International Cricket ‘In-Play’ Trading Strategies Can Be Profitable!”, *Social Sciences Research Network* (SSRN), disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2465536> o <http://dx.doi.org/10.2139/ssrn.2465536> (documento digital, pp. 8-9).

¹² GROVES, Robert M., “Three Eras in Survey Research”, *Public Opinion Quarterly*, vol. 75, nro. 5, disponible en el sitio web oficial de la *Public Opinion Quarterly* en: <https://academic.oup.com/poq/article/75/5/861/1831518/Three-Eras-of-Survey-Research> (documento digital, p. 868).

¹³ RESTIEAUX, Ainslie, “Work Programme for Consumer Price Statistics”, *Office for National Statistics*, disponible en el sitio web oficial de la *Office for National Statistics* en: <https://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/work-programme-for-consumer-price-statistics.pdf> (documento digital, p. 15).

¹⁴ Véase: BRETON, Robert *et al.*, “Research indices using web scraped data”, *Office for National Statistics*, disponible la web oficial de la *Office for National Statistics* en: <http://webarchive.nationalarchives.gov.uk/20160106041631/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/research-indicies-using-web-scraped-data.pdf> (documento digital, p. 2 y ss.) y BRETON, Robert *et al.*, “Research indices using web scraped data: May 2016 update”, *Office for National Statistics*, disponible en la web oficial de la *Office for National Statistics* en: <https://www.ons.gov.uk/file?uri=/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/may2016update/researchindicesusingwebscrapeddatamay2016.pdf> (documento digital, p. 4 y ss.).

con el objeto de predecir la evolución de precios mediante el uso de estadística bayesiana¹⁵.

Pueden encontrarse, a su vez, diversos proyectos recientes que se han valido de la metodología que aquí se trata. Así, por ejemplo, se la ha usado para el relevamiento de precios de productos electrónicos de consumo y pasajes de avión, en el marco de diversas iniciativas realizadas por el *Istituto Nazionale di Statistica* (Istat) italiano para modernizar sus mecanismos de recolección de datos¹⁶. También se la ha utilizado para la reconstrucción de la base de datos del Banco Mundial mediante la realización de más de 23 millones de peticiones a su sitio web, con el objeto de producir diversos indicadores económicos¹⁷. Asimismo, se la ha aplicado para el análisis del mercado inmobiliario mediante el relevamiento de sitios populares de avisos en línea¹⁸.

El uso del *web scraping* ha tenido particular protagonismo en nuestro país para la producción de indicadores de coyuntura económica, en un contexto signado por serias acusaciones de manipulación de las estadísticas oficiales publicadas por el Instituto Nacional de Estadísticas y Censos (INDEC), intervenido desde el año 2007 por el gobierno nacional presidido por Néstor Carlos Kirchner, obra que prosiguiera durante el mandato de su sucesora, Cristina Elisabet Fernández¹⁹. Estos indicadores han permitido que la población contara con guarismos económicos confiables, relevados con una

¹⁵ POWELL, Ben *et al.*, “Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting”, *Journal of the Royal Statistical Society: Series A (Statistics in society)*, vol. 181, nro. 3, disponible en la web oficial de la *Royal Statistical Society* en: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12314> o <https://doi.org/10.1111/rssa.12314> (documento digital, pp. 738-741).

¹⁶ POLIDORO, Federico *et al.*, “Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation”, *Statistical Journal of the IAOS*, vol. 31, nro. 2, DOI: 10.3233/SJI-150901, disponible en la web oficial de *IOS Press* en: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji901> (documento digital, p. 165 y ss.).

¹⁷ DYKSTRA, Sarah, DYKSTRA, Benjamin y SANDEFUR, Justin, “We Just Ran Twenty-Three Million Queries of the World Bank's Website”, *Center for Global Development*, documento de trabajo n° 362, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2458086> o <http://dx.doi.org/10.2139/ssrn.2458086> (documento digital, p. 6).

¹⁸ Véase: BOEING, Geoff y WADDELL, Paul, “New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings”, *Journal of Planning Education and Research*, DOI: 10.1177/0739456X16664789, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2781297> (documento digital, pp. 3 y 4); FENG, Jay, “A Week of Mining Seattle's Craigslist Apartment Pricing”, *Data and Stuff by Jay*, 23/12/2014, disponible en el sitio web personal del autor en: <http://www.racketracer.com/2014/12/23/a-week-of-seattles-craigslist-apartment-pricing/> y FENG, Jay, “San Francisco Rent Analysis: Part 1 – Cheapest month to find housing in SF?”, *Data and Stuff by Jay*, 19/09/2018, disponible en el sitio web personal del autor en: <http://www.racketracer.com/2018/09/19/san-francisco-rent-analysis-part-1-cheapest-month-to-find-housing-in-sf/>.

¹⁹ Véase: AA. VV., “El Indec, la máquina de la mentira. Crónica de la destrucción de las estadísticas públicas durante el kirchnerismo”, *La Nación*, 15/04/2015, disponible en la web oficial del periódico de difusión nacional *La Nación* en: <http://casos.lanacion.com.ar/indec-la-maquina-de-la-mentira> y CABOT, Diego, “Acusan al Gobierno de manipular datos del Indec”, *La Nación*, 10/02/2007, disponible en la web oficial del periódico de difusión nacional *La Nación* en: <https://www.lanacion.com.ar/882393-acusan-al-gobierno-de-manipular-datos-del-indec>.

periodicidad mayor que la acostumbrada por la metodología de recolección manual (lo que les ha valido el nombre de “indicadores de alta frecuencia”).

Entre ellos se destaca *The Billion Prices Project*²⁰, desarrollado en la *Massachusetts Institute of Technology* (MIT) *Management Sloan School* por Alberto F. Cavallo²¹ y Roberto Rigobon. El mismo realiza un seguimiento diario de la evolución de precios publicados en páginas de *e-commerce* de diversos países, entre los cuales se encuentra la Argentina.

Al mismo tiempo, puede mencionarse el proyecto IPCOnline²², en el que participé en carácter de fundador junto al Ing. Juan I. Uriarte y el Dr. Juan M. C. Larrosa. El mismo

²⁰ Puede hallarse el sitio web oficial del proyecto en: <http://www.thebillionpricesproject.com/>.

²¹ Es posible hallar una nutrida bibliografía perteneciente al autor en la que profundiza diversas temáticas de naturaleza económica, a partir de datos recabados mediante *web scraping*. Al respecto, véase: CAVALLLO, Alberto, “The Billion Prices Project: using online prices for measurement and research”, *Journal of Economic Perspectives*, vol. 30 (2), 151-178; *National Bureau of Economic Research*, documento de trabajo n° 22.111, DOI: 10.3386/w22111, disponible en la web oficial del *National Bureau of Economic Research* en: <https://www.nber.org/papers/w22111> (documento digital, pp. 2-6); CAVALLLO, Alberto, “Are online and offline prices similar? Evidence from large multi-channel retailers”, *American Economic Review*, vol. 107 (1), 283-303; *National Bureau of Economic Research*, documento de trabajo n° 22.142, DOI: 10.3386/w22142, disponible en la web oficial del *National Bureau of Economic Research* en: <https://www.nber.org/papers/w22142> (documento digital, pp. 2-4 y 10-11); CAVALLLO, Alberto y RIGOBÓN, Roberto, “The distribution of the size of price changes”, *National Bureau of Economic Research* (NBER), documento de trabajo n° 16.760, DOI: 10.3386/w16760, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w16760> y en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=1754912> (documento digital, pp. 6-7); CAVALLLO, Alberto, “Online and official price indexes: Measuring Argentina’s inflation”, *Journal of Monetary Economics*, vol. 60, nro. 2, 152-165, disponible en la web oficial de *Science Direct* en: <https://www.sciencedirect.com/science/article/abs/pii/S0304393212000967> o <http://dx.doi.org/10.1016/j.jmoneco.2012.10.002> (documento digital, p. 2); CAVALLLO, Alberto, CAVALLLO, Eduardo y RIGOBON, Roberto, “Prices and supply disruptions during natural disasters”, *Review of Income and Wealth*, vol. 60, S449-S471; *National Bureau of Economic Research* (NBER), documento de trabajo n° 19.474, DOI: 10.3386/w19474, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w19474> (documento digital, pp. 6-8); CAVALLLO, Alberto, “Scraped data and sticky prices”, *The Review of Economics and Statistics*, vol. 100 (1), 105-119; *National Bureau of Economic Research* (NBER), documento de trabajo n° 21.490, DOI: 10.3386/w21490, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w21490> (documento digital, pp. 5-7); y CAVALLLO, Alberto, “Scraped data and sticky prices”, *MIT Sloan*, documento de investigación n° 4.976-12, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=1711999> o <http://dx.doi.org/10.2139/ssrn.1711999> (documento digital, pp. 6-8).

²² Véase: URIARTE, Juan I., RAMÍREZ MUÑOZ de TORO, Gonzalo R. y LARROSA, Juan M. C., “Web Scraping Based Online Consumer Price Index: The Case of the IPC Online” (título tentativo, bajo revisión de comité editorial); LARROSA, Juan M. C. (director), *Índice de Precios al Consumidor Online e Índice de Costo de la Construcción Online: Medición de la inflación en Bahía Blanca, 2014-2017* (título tentativo, consultado por gentileza del autor) y RINALDI, Juan F., “En Bahía, la inflación también puede medirse a través de Internet”, *La Nueva Provincia*, 01/12/2014, disponible en la web oficial del periódico de difusión local de la ciudad de Bahía Blanca *La Nueva Provincia* en: <http://www.lanueva.com/nota/2014-12-1-0-26-0-en-bahia-la-inflacion-tambien-puede-medirse-a-traves-de-internet>.

fue desarrollado en el ámbito del Departamento de Economía de la Universidad Nacional del Sur²³, en la ciudad de Bahía Blanca, y se basó en un sistema informático que desarrollé junto al Ing. Juan I. Uriarte para la recopilación y procesamiento de precios en línea, con el cual hemos logrado construir una base de datos que, al tiempo de realizar este trabajo, cuenta con más de quince millones de precios de superficies comerciales de diverso tamaño que operan en la mentada ciudad. El software referido permitió la confección de distintos indicadores económicos de publicación periódica y totalmente automatizada, entre los que se destacan un índice de precios al consumidor de alta frecuencia²⁴, un índice de costo de la construcción, y canastas básicas alimentaria y total para la determinación de líneas de ingreso de indigencia y pobreza²⁵.

Como producto derivado del anterior, hemos efectuado un análisis de modelo hedónico de precios para el mercado nacional de automotores usados utilizando información recolectada de sitios populares de avisos en línea²⁶. En el mismo hemos establecido estadísticamente de qué modo diversos elementos (combustible, color, acondicionador de aire, ubicación geográfica, etc.) contribuyen a la determinación del precio de reventa de las unidades.

§ 4. Algunas precisiones sobre la metodología seguida en este trabajo

Para los fines aquí propuestos, desarrollé en el lenguaje informático Ruby un sistema de software tipo *web scraper* para simular la navegación que realizaría un ser humano en las páginas web pertinentes del sitio oficial de la Cámara Nacional de Apelaciones en lo Civil. Ante los formularios de búsqueda, el programa simula la selección de los distintos

²³ Actualmente, en el referido departamento se encuentra en funcionamiento un Proyecto de Grupo de Investigación (PGI) denominado “Base de precios minoristas, inflación e índice de precios al consumidor de Bahía Blanca”, bajo la dirección del Dr. Juan M. C. Larrosa e integrado por el autor de este trabajo, donde se están desarrollando diversas líneas de investigación relativas a la metodología analizada.

²⁴ Para este indicador hemos desarrollado software adicional encargado de la redacción automática, sin intervención humana, del texto de cada informe. Logramos así la automatización total, de punta a punta, del proceso de producción de la información, que se extiende desde el relevamiento de precios hasta la publicación del informe respectivo en formato de texto legible por seres humanos, pasando por el procesamiento de los datos relevados y el cálculo económico pertinente. Damos así por cumplido nuestro objetivo inicial, que fuera construir un sistema de software que pudiese calcular de forma integral y autónoma el mencionado indicador, como procedimiento extensible, sobre todo, a institutos oficiales de estadística, a fin de aventar los peligros derivados de la intervención potencialmente manipuladora de autoridades y otros agentes que pudieran abrigar algún interés en la falsificación de guarismos de inflación. La adopción de mecanismos similares por tales organismos ha de fungir como una suerte de anticuerpo institucional frente al avasallamiento del poder político, y evitará la reproducción de aciagos sucesos como los que afectaron al Instituto Nacional de Estadísticas y Censos (INDEC) durante el periodo 2007-2015.

²⁵ La totalidad del material publicado puede encontrarse en el sitio web oficial del proyecto en: <http://ipconlinebb.wordpress.com>. También en las redes sociales Twitter (@ipconlinebb) y Facebook (<https://www.facebook.com/ipconlinebb>).

²⁶ RAMÍREZ MUÑOZ de TORO, Gonzalo R., URIARTE, Juan I., DELBIANCO, Fernando y LARROSA, Juan M. C., “Un modelo hedónico de precios en línea de automóviles usados en Argentina”, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, vol. 24 (2017), disponible en el sitio web oficial de la *Revista de Métodos Cuantitativos para la Economía y la Empresa* en <https://www.upo.es/revistas/index.php/RevMetCuant/article/view/2879> (documento digital, pp. 25-53).

criterios para cada menú y los clics para el envío de las consultas pertinentes, de modo combinado y sucesivo para permitir el acopio de la totalidad de la información remota disponible. A efectos de minimizar el impacto en el rendimiento de los servidores, se dispuso una espera en intervalos de sesenta segundos entre cada petición y tan solo se realizó un cúmulo mínimo de consultas por día a lo largo de varias jornadas. Los resultados de dichas búsquedas fueron así raspados, mínimamente procesados y almacenados en diversas tablas de una base de datos relacional SQLite.

Una vez finalizada la recolección, las bases de datos fueron exportadas a efectos de su corrección, normalización, análisis y representación mediante gráficos, para lo cual desarrollé software adicional en el lenguaje de computación estadística y gráficos denominado R, con RStudio como entorno de desarrollo.

§ 5. Datos recolectados

§ 5. a) Análisis general

Se recabaron inicialmente datos para un total de 20.642 sentencias en las que se cuantifican daños de la Cámara Nacional de Apelaciones en lo Civil. Durante la labor de corrección y normalización, se eliminaron 18 fallos correspondientes a la base de datos Lesiones²⁷ por contener anomalías de origen, en su mayoría relativas a las fechas asignadas al acaecimiento del hecho dañoso o a la emisión de la sentencia. De esta manera, la base de datos utilizada para el presente trabajo quedó conformada con un total de 20.624 fallos de segunda instancia para el lapso temporal que va desde el 2 de febrero de 1993 al 21 de abril de 2016, correspondientes a sentencias de primera instancia dictadas entre el 15 de abril de 1988 y el 13 de octubre de 2015. Ellos se distribuyen de la siguiente manera:

²⁷ Fallos “Medina, Leandro c/Ferrocarriles Argentinos” (caso n° 6.696); “Ledesma, Yolando Hilario c/Eaton Icsa” (caso n° 5.093); “Espinoza, Carlos A. c/Codum S. A.” (caso n° 2.072); “Segovia, Teresita c/M. C. B. A.” (caso n° 1.849); “Muñoz, Ramón c/Ferrocarriles Metropolitanos” (caso n° 6.458); “Rodríguez, Maximiliano Daniel c/Asociación de Fútbol Argentina” (caso n° 11.918); “Paredes de Rodríguez, Mirna c/Mihocka, Bárbara y otro” (caso n° 14.266); “Costa, Jorge Rubén y otro c/Castillo Núñez, Celestino Ramón (2/2)” (caso n° 15.176); “Mansilla, Mirta Mabel y otro c/Ruiz, Ricardo Fabián y otros” (caso n° 15.173); “Domínguez, Mabel Luján c/Araujo, Julio César y otros” (caso n° 11.846, expresa incapacidad psíquica en valores negativos, a contramano del resto de las bases de datos); “Romero, Karina c/Encina, Juan” (caso n° 7.060); “Mariño, Juan Ramón y otro c/Ávalos, Jorgelina Fernanda y ot. (2/2)” (caso n° 17.252); “Romano, Cristian Adrián y otro c/Carpineta, Rosana Mabel y otros” (caso n° 15.465); “Waseman, Raquel c/Los Constituyentes S. A.” (caso n° 8.837); “Ragone, Carmen c/Espósito, Raúl A.” (caso n° 7.670); “Osorio, María Isabel c/Transportes M. O. Este S. A.” (caso n° 6.473); “Tapia, María Isabel c/Rubo, Miguel A. y otros” (caso n° 3.663) y “Deza, Amelia c/Cabrera, Carlos y otro” (caso n° 1425).

Tabla 1. Composición de las bases de datos reconstruidas en este trabajo

	Base de datos	Recuento	Porcentaje
1	Lesiones	16.236	78,7%
2	Fallecimientos	2.273	11,0%
3	Extrapatrimonial	1.357	6,6%
4	Responsabilidad de profesionales de la salud ²⁸	728	3,5%
5	Violencia contra la mujer	30	0,1%
	Total	20.624	100%

Como puede observarse, existe un amplio predominio de fallos correspondientes a lesiones, seguidos a lo lejos por sentencias relativas a fallecimientos.

Se registró un número sorprendentemente bajo de sentencias relativas a violencia de género, lo que podría deberse a una publicación meramente parcial de la jurisprudencia del tribunal o, lo que sería alarmante, a obstáculos culturales, institucionales, económicos, entre otros, con que deben enfrentarse las víctimas de la mencionada violencia para el acceso a la justicia. No menos llamativo resulta el hecho de que el primer fallo relevado en la materia (“P., M. c/Cía. de Servicios Hoteleros”, caso n° 15) es de fecha en alguna medida reciente, datado para resolución de primera instancia al 12 de junio de 2000 y para segunda instancia al 5 de junio de 2001.

§ 5. b) Víctimas

§ 5. b) 1. Sexo

Se detectaron presuntas anomalías en la publicación de las sentencias relativas a lesiones generadoras de responsabilidad de los profesionales de la salud, en razón de que en la totalidad de las 526 sentencias relevadas se consignó que el lesionado pertenecía al sexo femenino. Es por ello que las mismas no serán representadas en los guarismos de la presente sección. A su vez, se excluyeron dos víctimas cuyo sexo fue consignado como “sin datos”²⁹.

De esta forma, se computó un total de 20.429 víctimas, con la siguiente distribución: 12.067 (59,0%) correspondientes al sexo masculino y 8.362 (40,9%) al sexo femenino. La distribución por sexo en las distintas bases de datos se consigna en la Tabla 2.

²⁸ En esta base de datos se incluyeron 202 casos por fallecimientos y 526 por lesiones derivadas de responsabilidad de profesionales de la salud.

²⁹ Fallos “C., P. J. c/ Editorial Arte Gráfico Argentino” (caso n° 89) y “Rubino, Laura c/ Htal. de agudos Dr. Ignacio Pirovano” (caso n° 177).

Tabla 2. Sexo de víctimas y porcentajes por base de datos

Base de datos	Femenino		Masculino		Totales	
	Cantidad	Porcentaje	Cantidad	Porcentaje	Cantidad	Porcentaje
Lesiones	6.906	42,5%	9.330	57,5%	16.236	100%
Fallecimiento	623	27,4%	1.650	72,6%	2.273	100%
Extrapatrimonial	699	41,4%	989	58,6%	1.688	100%
Profesionales de la salud	104	51,5%	98	48,5%	202	100%
Violencia de género	30	100%	0	0%	30	100%

§ 5. b) 2. Edad

Para este subinciso solo se tuvieron en cuenta edades de víctimas expresadas como números enteros, excluyéndose variables categóricas tales como que la víctima es “mayor” o “nonato”³⁰. Por las mismas razones expuestas en el subinciso anterior, se descartaron fallos por lesiones generadoras de responsabilidad de los profesionales de la salud.

En total, se computaron edades para 18.999 víctimas, correspondiendo 7.763 (40,9%) al sexo femenino y 11.236 (59,1%) al masculino.

Tabla 3. Sumario estadístico de edades de víctimas

	Femenino	Masculino	General
Mínima	0	0	0
1° cuartil	26	23	24
Mediana	42	33	37
Media	42,2	36,2	38,6
3° cuartil	57	48	53
Máxima	92	93	93

De lo consignado resulta que la edad mínima de las víctimas es, previsiblemente, 0 (cero), en tanto que la máxima llegó a 93 años. Se destaca, a su vez, una mediana de 37 años y una media 38,55 años; a este respecto, el sexo femenino se ubicó algo por encima (mediana de 42 y media de 42,2) y el sexo masculino hizo lo propio por debajo (mediana de 33 y media de 36,2).

³⁰ Se excluyeron 8 víctimas cuya edad fue consignada como “mayor”, 38 como “nonato” y 1.393 que no registraron valor alguno.

Figura 1. Histograma general de edades de víctimas

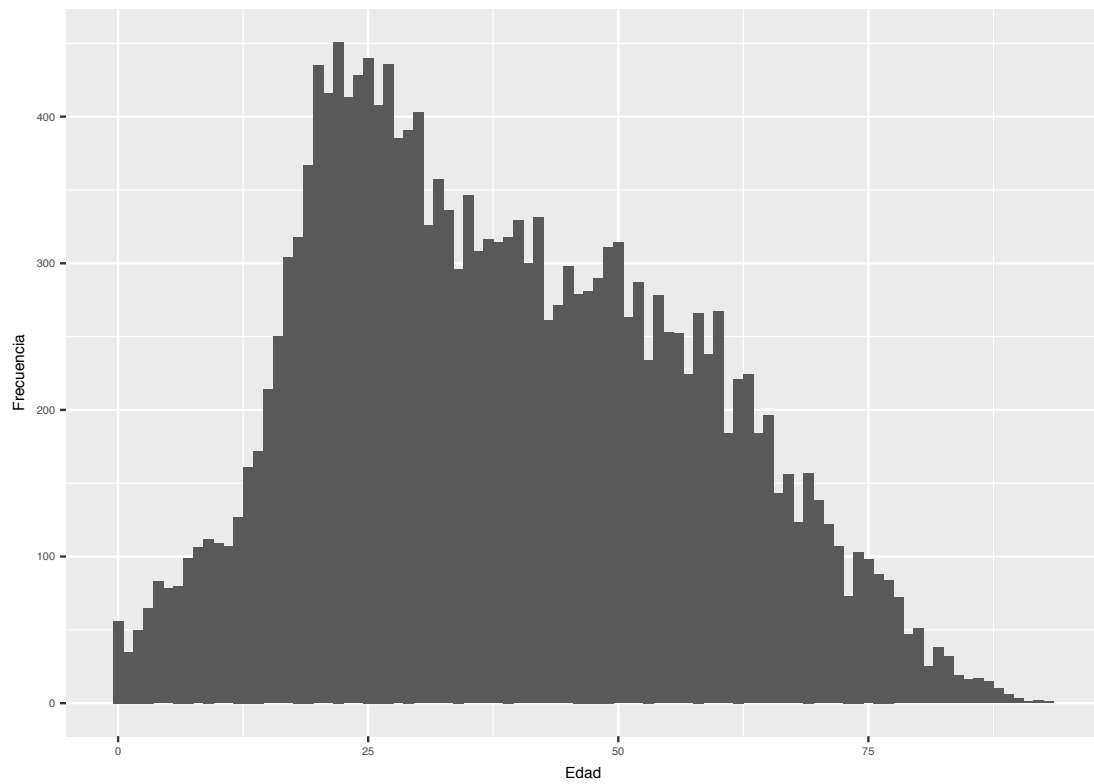


Figura 2. Histograma de edades de víctimas por sexo

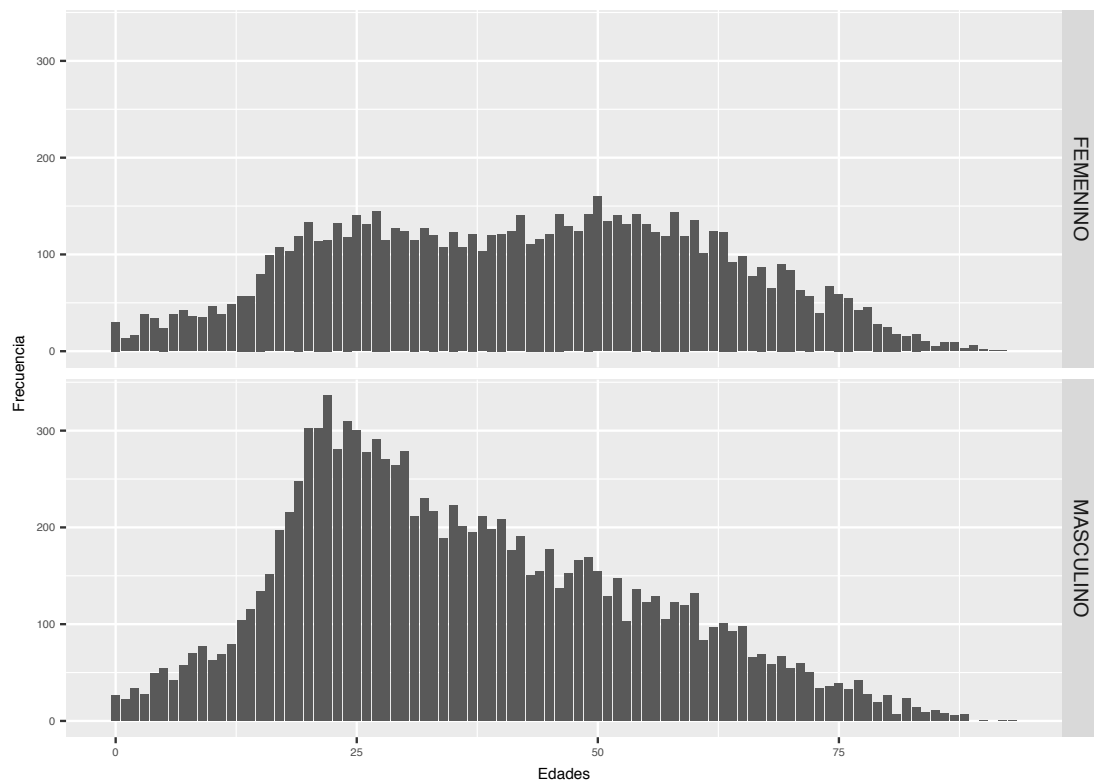


Figura 3. Diagrama de caja general de edades de víctimas

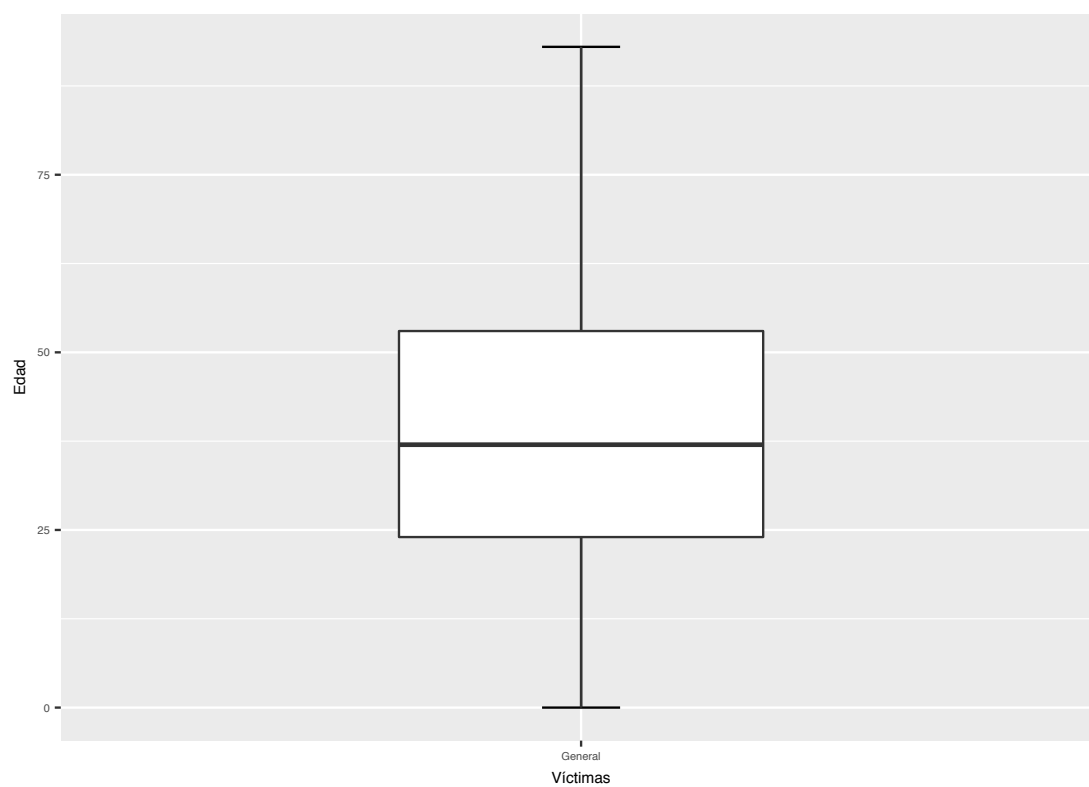
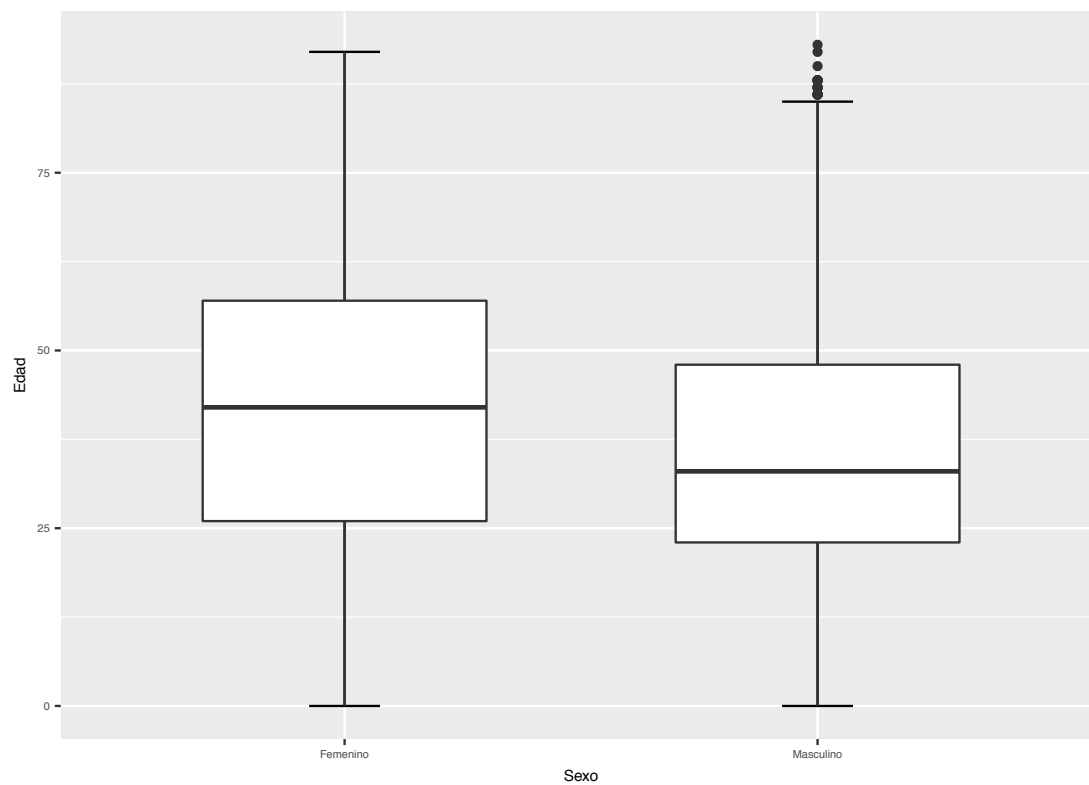


Figura 4. Diagrama de caja de edades de víctima por sexo



§ 5. c) Lesiones

En este inciso se emprende un somero análisis de la base de datos Lesiones, exclusivamente, por ser la más nutrida de las relevadas. El estudio de las restantes extendería en demasía el presente trabajo y quedará, por ello, encomendado a futuras indagaciones.

En la Figura 5 se representa la dispersión de los siniestros por sexo y edad de 16.236 lesionados. Se observa claramente el protagonismo de los siniestros viales, que conforman las áreas más nutridas del gráfico y aportan las seis primeras frecuencias a la comparación: “Pasajero” (4.199 registros, lo que representa 25,9% del total), “Peatón” (3.305, 18,7%), “Conductor” (2.551, 15,7%), “Motociclista” (1.891, 11,6%), “Acompañante” (1.690, 10,4%) y, finalmente, “Ciclista” (725, 4,5%). En particular, se destaca la diversa composición por sexo evidenciada. Se percibe que algunos registran una marcada presencia femenina, tales como:

- ❖ “Acompañante”: para un total de 1.690 lesionados, 1.060 (62,7%) corresponden al sexo femenino y 630 (37,3%) al masculino.
- ❖ “Pasajero”: para un total de 4.199 lesionados, 2.657 (63,3%) pertenecen al sexo femenino y 1.542 (36,7%) al masculino.

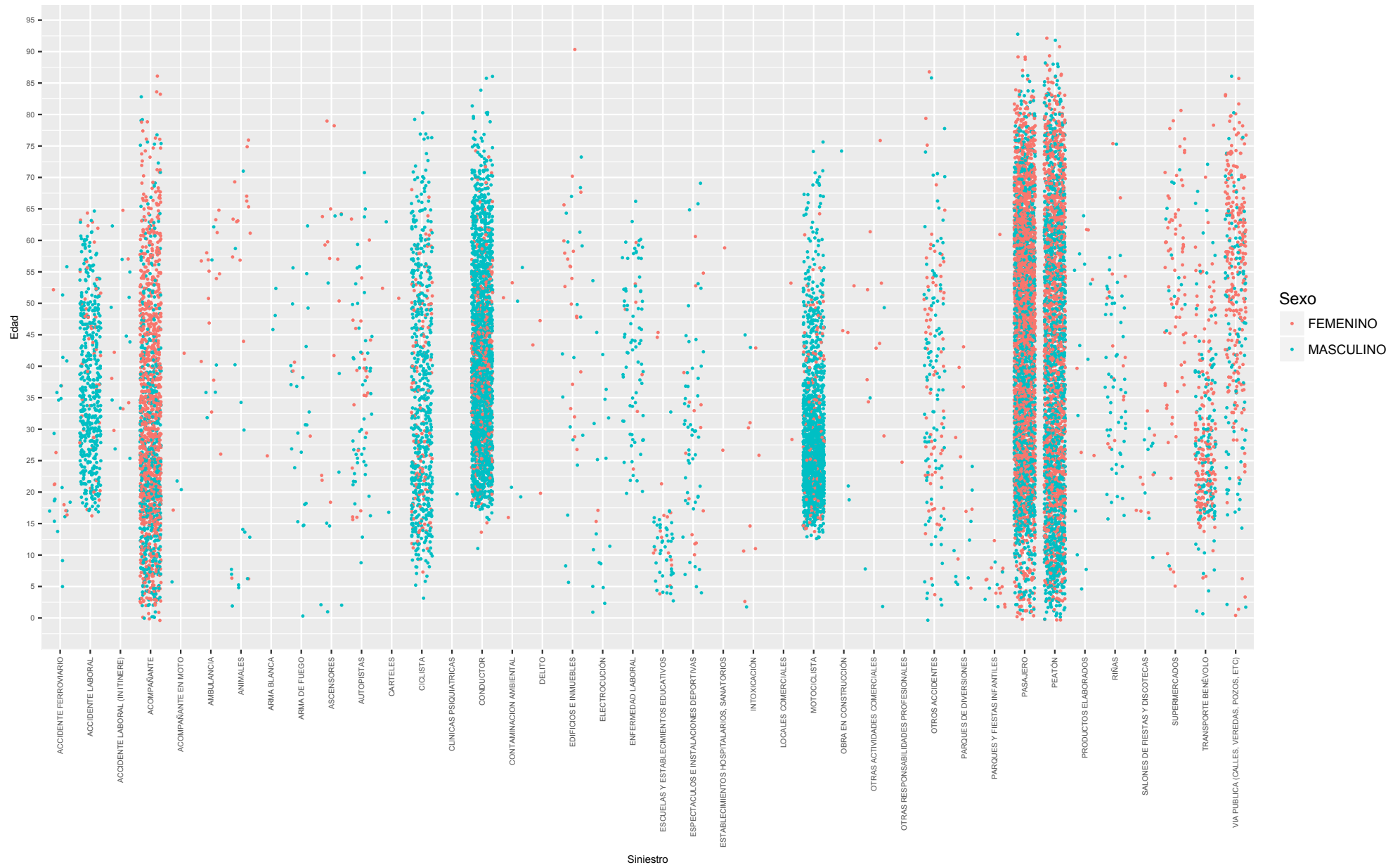
Por el contrario, otros registran una amplia mayoría de masculinos, a saber:

- ❖ “Conductor”: en un total de 2.551 víctimas, 2.022 (79,3%) son de sexo masculino, mientras que tan solo 529 (20,7%) son de sexo femenino.
- ❖ “Ciclista”: para un total de 725 víctimas, 585 (80,7%) corresponden al sexo masculino y 140 (19,3%) al femenino.
- ❖ “Motociclista”: para un total de 1.891 víctimas, 1.737 (92,9%) pertenecen al sexo masculino y 154 (8,1%) al femenino.

De este modo, se aprecia una desigual distribución por género según se trate de siniestros sufridos por sujetos *activos* del transporte, donde predominan los masculinos, o sujetos *pasivos* del mismo, donde predominan los femeninos. Ello podría sugerir que, dados ciertos supuestos, existe una propensión a la conducción riesgosa por parte de sujetos pertenecientes al sexo masculino.

Por su parte, el protagonismo casi excluyente de las sentencias derivadas de siniestros viales torna aconsejable asignar competencia exclusiva en materia de daños por causa de tránsito a una o más salas de la Cámara Nacional de Apelaciones en lo Civil. La especialización técnica redundará, presumiblemente, en una mayor celeridad de los procesos, lo que congruaría una muestra deseable de sana administración judicial.

Figura 5. Dispersión de siniestros por sexo y edad de víctimas para base de datos de Lesiones



§ 6. ¿Por qué el análisis de datos debe ocupar un lugar protagónico en el derecho?

En el año 2002, la humanidad almacenó por primera vez más información en formato digital que analógico, lo que dio inicio a la denominada “era digital”³¹. Recientemente, la revista inglesa *The Economist* tituló: “El recurso más valioso del mundo ya no es el petróleo, sino los datos”³². En la sociedad de la información, el dato se convierte en la medida del conocimiento y transforma a su paso saberes de toda índole.

Sin embargo, y a diferencia de lo que acontece en otras disciplinas sociales como la economía, el derecho ha sido refractario a adoptar estrategias integrales de análisis de datos como herramienta central para la producción de conocimiento³³. Este déficit podría colocarlo, en un futuro próximo, ante una verdadera crisis de legitimidad³⁴, dada su manifiesta incapacidad para respaldar sus especulaciones en hechos comprobables. Ha afirmado Augusto M. Morello que “los jueces no pueden ser fugitivos de la realidad”; no pueden, por ello, desentenderse de los efectos de sus sentencias, y es a todas luces razonable suponer que la realidad fáctica en que sus fallos han de regir puede ser aprehendida de mejor manera lejos de la anécdota y en proximidad a un enfoque sistemático de tipo estadístico. A su vez, el análisis de datos ofrecerá una base empírica sobre la cual erigir demandas y sentencias, permitirá captar las grandes tendencias decisorias de los tribunales y contribuirá a un control más efectivo de ciertas zonas umbrías de la administración de justicia, hoy ocultas al tradicional enfoque “microjurídico”.

Por su parte, las diversas ramas de la inteligencia artificial, que suelen hacer uso intensivo de *datasets* gigantescos para el entrenamiento de algoritmos, se avizoran como terrenos fértiles para eventuales revoluciones jurídicas, las que deberán esperar en tanto no se destinen ingentes esfuerzos a una radical transformación de la enseñanza del derecho, que habrá de ofrecer a los estudiantes los rudimentos indispensables para el desarrollo de software y el análisis estadístico. Otro tanto tendrá que dirigirse a la estructuración no solo de los datos jurídicos, sino de los mismísimos procesos judiciales, sobre la base de contemplar a la administración de justicia como una verdadera factoría de datos. Todo ello, claro, si es que el derecho pretende dar ese salto hacia delante desde adentro.

Vale aquí, finalmente, recordar las palabras de Oliver W. Holmes Jr. con las que abriera este trabajo: “Para el análisis racional de la ley, el hombre de las letras góticas puede ser el hombre

³¹ HILBERT, *loc. cit.*

³² AA. VV., “The world’s most valuable resource is no longer oil, but data”, London, *The Economist*, 06/05/2017, disponible en la web oficial de la revista de difusión internacional *The Economist* en: <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>.

³³ Quizás ello explique por qué el saber popular tiende a emparentar al derecho con el campo de las denominadas “ciencias blandas”, mientras que afilia a la economía —que hace uso intensivo del análisis de datos mediante procedimientos estadísticos— al conjunto pretendidamente más desarrollado de “ciencias duras”, a pesar de ser esta tan *blanda* como aquel. Al respecto, véase: RAMÍREZ MUÑOZ de TORO, Gonzalo R., “El derecho, la economía y sus desafíos en la era de la información”, disertación breve presentada en el *Congreso Internacional de Gobernanza Inteligente e Innovación Inclusiva*, organizado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y otros, del 22 al 24 de noviembre de 2017, Ciudad Autónoma de Buenos Aires, versión por escrito aprobada por el comité organizador disponible en: <http://bit.ly/2JFF5i4> (documento digital, pp. 3-4).

³⁴ *Ibidem*.

del presente, pero el hombre del futuro es el hombre de la estadística y el maestro de la economía”³⁵. Y también que fueron escritas en 1897.

§ 7. Divulgación de este trabajo

Con el fin de divulgar la información acumulada en el desarrollo de la presente investigación, he desarrollado un sistema de software para la publicación automática de datos relevantes de cada fallo en la red social Twitter (@unfalloporthora). Se trata de lo que en la jergonza informática se conoce como *bot*, esto es, un programa de computadora que ejecuta determinadas tareas en forma autónoma, normalmente como consecuencia del cumplimiento de ciertas condiciones o en intervalos de tiempo preestablecidos.

El mencionado *bot* ha sido desarrollado íntegramente en lenguaje Python. A la fecha de publicación del presente trabajo se encontraba difundiendo datos relativos a la base de datos Lesiones, habiendo ya emitido más de cuarenta y siete mil tuits desde el año 2017.

Asimismo, las bases de datos reconstruidas han sido publicadas para su uso en la realización de trabajos de índole académica o científica que no persigan fines de lucro. Las mismas se encuentran disponibles en formato .csv y .xlsx en: <https://github.com/GonzaloRMDT/danos-cniv>.

§ 8. Conclusiones

- ❖ Se ha logrado reconstruir, corregir y normalizar exitosamente las bases de datos de sentencias de cuantificación de daños publicadas en el sitio web de la Cámara Nacional de Apelaciones en lo Civil de la República Argentina, las que comprenden resoluciones de segunda instancia para el lapso temporal que va desde el 2 de febrero de 1993 al 21 de abril de 2016.
- ❖ Se han explorado datos relativos a más de veinte mil fallos, y trazado un somero perfil estadístico de víctimas (damnificados directos) y siniestros generadores de lesiones.
- ❖ Se ha argumentado sobre el papel protagónico que es de esperar asuma el análisis de datos en la formación, investigación y práctica jurídicas.
- ❖ Se ha divulgado en forma novedosa la información recolectada, con el auxilio de un *bot* publicador para la red social Twitter.
- ❖ Se ha realizado un aporte al movimiento de datos abiertos, así como a la investigación científica y académica, mediante la publicación de las bases de datos aquí reconstruidas.

§ 9. Anexo de datos

Integran el presente trabajo las bases de datos mencionadas en § 7, al que se remite para conocer el modo en que puede accederse a ellas.

³⁵ “For the rational study of the law the black-letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics”, HOLMES, Oliver W., “The Path of the Law”, *Harvard Law Review*, vol. 10, no. 8, DOI: 10.2307/1322028, disponible en la web oficial de JSTOR en: www.jstor.org/stable/1322028 (documento digital, p. 469).

§ 10. Referencias

1. AA. VV., “El Indec, la máquina de la mentira. Crónica de la destrucción de las estadísticas públicas durante el kirchnerismo”, *La Nación*, 15/04/2015, disponible en la web oficial del periódico de difusión nacional *La Nación* en: <http://casos.lanacion.com.ar/indec-la-maquina-de-la-mentira>.
2. AA. VV., “The world’s most valuable resource is no longer oil, but data”, London, *The Economist*, 06/05/2017, disponible en la web oficial de la revista de difusión internacional *The Economist* en: <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>.
3. AA. VV., *Wikipedia*, disponible en el sitio web oficial de la enciclopedia *Wikipedia* en: <https://en.wikipedia.org>.
4. BLACK, Paul E. y PIETERSE, Vreda (editores), *Dictionary of Algorithms and Data Structures [online]*, disponible en el sitio web oficial del *National Institute of Standards and Technology* en: <https://xlinux.nist.gov/dads/HTML/dataStructure.html>.
5. BOEING, Geoff y WADDELL, Paul, “New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings”, *Journal of Planning Education and Research*, DOI: 10.1177/0739456X16664789, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2781297>.
6. BRETON, Robert, CLEWS, Gareth, METCALFE, Liz, MILLIKEN, Natasha, PAYNE, Christopher, WINTON, Joe y WOODS, Ainslie, “Research indices using web scraped data”, *Office for National Statistics*, disponible en la web oficial de la *Office for National Statistics* en: <http://webarchive.nationalarchives.gov.uk/20160106041631/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/research-indices-using-web-scraped-data>.
7. BRETON, Robert, FLOWER, Tanya, MAYHEW, Matthew, METCALFE, Elizabeth, MILLIKEN, Natasha, PAYNE, Christopher, SMITH, Thomas, WINTON, Joe y WOODS, Ainslie, “Research indices using web scraped data: May 2016 update”, *Office for National Statistics*, disponible en la web oficial de la *Office for National Statistics* en: <https://www.ons.gov.uk/file?uri=/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/may2016update/researchindicesusingwebscrapeddatamay2016.pdf>.
8. CABOT, Diego, “Acusan al Gobierno de manipular datos del Indec”, *La Nación*, 10/02/2007, disponible en la web oficial del periódico de difusión nacional *La Nación* en: <https://www.lanacion.com.ar/882393-acusan-al-gobierno-de-manipular-datos-del-indec>.
9. CAVALLO, Alberto, “Are online and offline prices similar? Evidence from large multi-channel retailers”, *American Economic Review*, vol. 107 (1), 283-303; *National Bureau of Economic Research*, documento de trabajo n° 22.142, DOI:

- 10.3386/w22142, disponible en la web oficial del *National Bureau of Economic Research* en: <https://www.nber.org/papers/w22142>.
10. CAVALLO, Alberto, "Online and official price indexes: Measuring Argentina's inflation", *Journal of Monetary Economics*, vol. 60, nro. 2, 152-165, disponible en la web oficial de *Science Direct* en: <https://www.sciencedirect.com/science/article/abs/pii/S0304393212000967> o <http://dx.doi.org/10.1016/j.jmoneco.2012.10.002>.
 11. CAVALLO, Alberto, "Scraped data and sticky prices", *The Review of Economics and Statistics*, vol. 100 (1), 105-119; *National Bureau of Economic Research* (NBER), documento de trabajo n° 21.490, DOI: 10.3386/w21490, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w21490>.
 12. CAVALLO, Alberto, "Scraped data and sticky prices", *MIT Sloan*, documento de investigación n° 4976-12, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=1711999> o <http://dx.doi.org/10.2139/ssrn.1711999>.
 13. CAVALLO, Alberto, "The Billion Prices Project: using online prices for measurement and research", *Journal of Economic Perspectives*, vol. 30 (2), 151-178; *National Bureau of Economic Research*, documento de trabajo n° 22.111, DOI: 10.3386/w22111, disponible en la web oficial del *National Bureau of Economic Research* en: <https://www.nber.org/papers/w22111>.
 14. CAVALLO, Alberto, CAVALLO, Eduardo y RIGOBON, Roberto, "Prices and supply disruptions during natural disasters", *Review of Income and Wealth*, vol. 60, S449-S471; *National Bureau of Economic Research* (NBER), documento de trabajo n° 19.474, DOI: 10.3386/w19474, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w19474>.
 15. CAVALLO, Alberto y RIGOBÓN, Roberto, "The distribution of the size of price changes", *National Bureau of Economic Research* (NBER), documento de trabajo n° 16.760, DOI: 10.3386/w16760, disponible en la web oficial del *National Bureau of Economic Research* (NBER) en: <http://www.nber.org/papers/w16760> y en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=1754912>.
 16. DYKSTRA, Sarah, DYKSTRA, Benjamin y SANDEFUR, Justin, "We Just Ran Twenty-Three Million Queries of the World Bank's Website", *Center for Global Development*, documento de trabajo n° 362, disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2458086> o <http://dx.doi.org/10.2139/ssrn.2458086>.
 17. ELSHEHAWY, Ashrakat, MARINOV, Nikolay y NANNI, Federico, "Quantifying Attention to Foreign Elections with Text Analysis of U.S. Congress and the Presidency", *Social Sciences Research Network* (SSRN), disponible en la web oficial de la *Social Sciences Research Network* (SSRN) en: <https://ssrn.com/abstract=2981486> or <http://dx.doi.org/10.2139/ssrn.2981486>.
 18. FENG, Jay, "A Week of Mining Seattle's Craigslist Apartment Pricing", *Data and Stuff by Jay*, 23/12/2014, disponible en el sitio web personal del autor en:

<http://www.racketracer.com/2014/12/23/a-week-of-seattles-craigslist-apartment-pricing/>.

19. FENG, Jay, “San Francisco Rent Analysis: Part 1 – Cheapest month to find housing in SF?”, *Data and Stuff by Jay*, 19/09/2018, disponible en el sitio web personal del autor en: <http://www.racketracer.com/2018/09/19/san-francisco-rent-analysis-part-1-cheapest-month-to-find-housing-in-sf/>.
20. GROVES, Robert M., “Three Eras in Survey Research”, *Public Opinion Quarterly*, vol. 75, nro. 5, 01/12/2011, disponible en el sitio web oficial de la *Public Opinion Quarterly* en: <https://academic.oup.com/poq/article/75/5/861/1831518/Three-Eras-of-Survey-Research>.
21. HILBERT, Martin, “Quantifying the Data Deluge and the Data Drought”, *Social Sciences Research Network (SSRN)*, disponible en la web oficial de la *Social Sciences Research Network (SSRN)* en: <https://ssrn.com/abstract=2984851>.
22. HOLMES, Oliver W., “The Path of the Law”, *Harvard Law Review*, vol. 10, no. 8, DOI: 10.2307/1322028, disponible en la web oficial de *JSTOR* en: www.jstor.org/stable/1322028.
23. LARROSA, Juan M. C. (director), *Índice de Precios al Consumidor Online e Índice de Costo de la Construcción Online: Medición de la inflación en Bahía Blanca, 2014-2017* (título tentativo, consultado por gentileza del autor).
24. NORTON, Hugh, GRAY, Stephen y FAFF, Robert W., “Yes, One-Day International Cricket ‘In-Play’ Trading Strategies Can Be Profitable!”, *Social Sciences Research Network (SSRN)*, disponible en la web oficial de la *Social Sciences Research Network (SSRN)* en: <https://ssrn.com/abstract=2465536> o <http://dx.doi.org/10.2139/ssrn.2465536>.
25. POLIDORO, Federico *et al.*, “Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation”, *Statistical Journal of the IAOS*, vol. 31, nro. 2, DOI: 10.3233/SJI-150901, disponible en la web oficial de *IOS Press* en: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji901>.
26. POWELL, Ben, NASON, Guy, ELLIOTT, Duncan, MAYHEW, Matthew, DAVIES, Jennifer y WINTON, Joe, “Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting”, *Journal of the Royal Statistical Society: Series A (Statistics in society)*, vol. 181, nro. 3, disponible en la web oficial de la *Royal Statistical Society* en: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12314> o <https://doi.org/10.1111/rssa.12314>.
27. RAMÍREZ MUÑOZ de TORO, Gonzalo R., “El derecho, la economía y sus desafíos en la era de la información”, disertación breve presentada en el *Congreso Internacional de Gobernanza Inteligente e Innovación Inclusiva*, organizado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y otros, del 22 al 24 de noviembre de 2017, Ciudad Autónoma de Buenos Aires, versión por escrito aprobada por el comité organizador disponible en: <http://bit.ly/2JFF5i4>.

28. RAMÍREZ MUÑOZ de TORO, Gonzalo R., URIARTE, Juan I., DELBIANCO, Fernando y LARROSA, Juan M. C., “Un modelo hedónico de precios en línea de automóviles usados en Argentina”, *Revista de Métodos Cuantitativos para la Economía y la Empresa*, vol. 24 (2017), disponible en el sitio web oficial de la *Revista de Métodos Cuantitativos para la Economía y la Empresa* en <https://www.upo.es/revistas/index.php/RevMetCuant/article/view/2879>.
29. RESTIEAUX, Ainslie, “Work Programme for Consumer Price Statistics”, *Office for National Statistics*, disponible en el sitio web oficial de la *Office for National Statistics* en: <https://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/work-programme-for-consumer-price-statistics.pdf>.
30. RINALDI, Juan F., “En Bahía, la inflación también puede medirse a través de Internet”, *La Nueva Provincia*, 01/12/2014, disponible en la web oficial del periódico de difusión local de la ciudad de Bahía Blanca *La Nueva Provincia* en: <http://www.lanueva.com/nota/2014-12-1-0-26-0-en-bahia-la-inflacion-tambien-puede-medirse-a-traves-de-internet>.
31. URIARTE, Juan I., RAMÍREZ MUÑOZ de TORO, Gonzalo R. y LARROSA, Juan M. C., “Web Scraping Based Online Consumer Price Index: The Case of the IPC Online” (título tentativo, bajo revisión de comité editorial).