

El internet: fuente de datos para estudios académicos.

Alphonse L. MacDonald

Ex Profesor Metodología, Estadística y Desarrollo Económico, Departamento
CCSS, PUC, Lima, Perú

Asesor Principal, Oficina General de Estadística de Suriname (ABS)

Primer Vice-Presidente, Instituto Interamericano de Estadística (IASI)

Ex alto funcionario del Fondo de Población de las Naciones Unidas, FNUAP,

Antecedentes históricos

- Contar y registrar: hay evidencia que los humanos sabían contar antes de escribir.
- La evidencia más temprana del conteo por humanos involucra huesos prehistóricos con muescas.
- El Hueso de Lebombo es un peroné de babuino con 29 muescas fue encontrado en la Cueva de la Frontera en Sudáfrica, que data aproximadamente de 44,000 aC.



- Se desconoce lo que es, y cual era su función.
- Se asemeja a los palos calendarios todavía en uso por los bosquimanos en Namibia.

El internet

- El internet, un sistema electrónico descentralizado de redes de comunicación interconectadas de alcance mundial. Originalmente creado para el intercambio de información científica.
- Actualmente la gran mayoría de la información en el internet no es de origen científico.
- Usado por agencias internacionales gubernamentales y no-gubernamentales, gobiernos nacionales y sub-nacionales, el sector privado, e individuos produce información de características, utilidad e accesibilidad muy diversas.

La Revolución digital



Temas de la presentación

- En esta presentación trataremos los datos estadísticos
 - disponible en el dominio público, es decir, de acceso gratuito, o de acceso con un mínimo de condiciones; y
 - que tienen como unidad de observación el país (estado-nación o territorio) y individuos.
- En el análisis se pueden estudiar las relaciones entre los países o entre las personas por separado.
- Es posible de relacionar datos del nivel de país o sus subdivisiones con características de personas.
- También es posible de usar datos agregados del nivel personal para caracterizar el país.

Datos estadísticos

- ¡No todas las cifras o números son estadísticas!
- Estadísticas son valores numéricos asociados a características u atributos de unidades de observación específicas,
 - que derivan su significado de un marco de referencia (marco teórico),
 - que son identificados en el espacio y el tiempo,
 - que tienen una distribución de probabilidades y
 - son obtenidos por un procedimiento específico y verificable.
- Las estadísticas deben ser verificables y replicables.

Disponibilidad de estadísticas de calidad

- La disponibilidad de datos estadísticas de alta calidad se deben a cuatro procesos de desarrollo en los últimos 40 años.
 - La creación de los sistemas nacionales de estadista,
 - La estandarización de las defunciones, conceptos y procedimientos,
 - El numero crecientes de organizaciones no gubernamentales que producen datos estadísticos, y
 - La aceptación generalizada que las estadísticas son bienes públicos, necesarios para una sociedad democrática. .

Estrategia de búsqueda de datos

- Independiente del tema de la investigación es recomendable consultar: el web de la oficina nacional de estadística, y verificar los rubros “datos” y “publicaciones” y “enlaces”.
- Para obtener datos regionales es preferible usar los sitios de las comisiones regionales de las agencias de las Naciones Unidas, y otras agencias con carácter sub-regional regional, el Grupo Andino y el Mercosur.
- Para obtener datos internacionales (de dos o mas regiones) es preferible usar los sitios de las agencias de las Naciones Unidas, incluso el Banco Mundial y el Fondo Monetario Internacional. Algunas centros universitarios mantienen repositorios especiales que incluyen los datos de las agencias de las NNUU.
- El idioma del internet es el Ingles, aunque hay sitios que tienen información disponibles el Español. Si el sitio no es de un país de habla Español es preferible consultar la versión inglesa, puesto que esta tiende ser más completo

Datos a nivel de país

- En el Perú: el INEI <https://www.inei.gob.pe/> y los ministerios sectoriales (agricultura, salud, educación, etc.). Puede permitir análisis sub-nacional.
- Datos de los países de la región sobre población, económica y ambiental del la CEPAL: CEPALSTAT
<http://estadisticas.cepal.org/cepalstat/portada.html?idioma=spanish>
- Datos de población a nivel mundial y los países: División de población de las NNUU. <https://www.un.org/development/desa/es/about/desa-divisions/population.html>
- Datos económicos globales y de los países: Banco Mundial
<https://datos.bancomundial.org/>
- Datos financieros globales y de los países: el Fondo Monetario Internacional,
<http://www.imf.org/external/spanish/pubs/ft/ar/2015/html/spotlight-free-data.htm>
- Datos ambientales globales y de los países: Programa del Medio Ambiente de las NNUU, UNEP, <http://geodata.grid.unep.ch/>

Banco Mundial Datos de libre acceso

- La fuente de datos sobre el desarrollo a nivel de país más completo,
- Este sitio <https://datos.bancomundial.org/> ha sido diseñado para facilitar la búsqueda, la descarga y el uso de los datos del Banco Mundial. Todos estos datos se pueden utilizar de manera gratuita y bajo mínimas restricciones.
- Tiene información para todo los países y territorios sobre 20 temas:
 - Agricultura y desarrollo rural
 - Ciencia y tecnología
 - Desarrollo social
 - Deuda externa
 - Educación
 - Energía y minería
 - Infraestructura
 - Pobreza
 - Salud
 - Sector privado
 - Cambio climático
 - Comercio
 - Desarrollo urbano
 - Economía y crecimiento
 - Eficacia de la ayuda
 - Género
 - Medio ambiente
 - Protección social y trabajo
 - Sector financiero
 - Sector público

Datos estadísticas no convencionales

- Algunas agencias internacionales pero sobre todo organizaciones no-gubernamentales presenten información estadística sobre temas que no forma parte del sistema de estadísticas públicas pero pueden ser de interés para la investigación científica.
- Algunos ejemplos son:
 - El Índice del Desarrollo Humano (IDH) del Programa de las Naciones Unidas para el Desarrollo PNUD, <http://hdr.undp.org/en/2016-report> como alternativa del PIB como medida de desarrollo.
 - El proyecto DOING BUSINESS del Banco Mundial proporciona una medición objetiva de las regulaciones para hacer negocios en 190 economías y en algunas ciudades seleccionadas <http://espanol.doingbusiness.org/> ;
 - La clasificación de la Capacidad Estadística de los países, territorios y regiones del Banco Mundial. Un índice anual basado en 3 dimensiones (metodología (10, 2), fuentes de datos (5, 3) y periodicidad (10, 5) <http://datatopics.worldbank.org/statisticalcapacity/>
- Hay centenares organizaciones no-gubernamentales o unidades comerciales privadas con objetivos específicos de interés general o partidario que publican índices sobre características de países o grupos de países.

Organizaciones no-gubernamentales

- Heritage Foundation (EEUU) ;
<https://www.heritage.org/about-heritage/mission> - Index of Economic Freedom 2018: <https://www.heritage.org/index/>
- Reporteros sin frontera (España) , <https://rsf.org/es/> Libertad de prensa. “La Clasificación Mundial de la Libertad de Prensa”: <https://rsf.org/es/la-clasificacion-mundial-de-la-libertad-de-prensa>
- The Economist Intelligence Unit (GB): “Democracy index 2017”: http://pages.eiu.com/rs/753-riq-438/images/democracy_index_2017.pdf
- Transparency International (Alemania): anti corrupción
<https://www.transparency.org/>
 - Transparency International España
 - <http://www.transparencia.org.es/>
 - Índice de Percepción de la Corrupción 2017;
<https://transparencia.org.es/ipc-2017/>

Evaluación de las fuentes

- Antes de usar los datos de organizaciones no-gubernamentales hay que verificar:
- la naturaleza de la organización;
 - ¿que clase de organización es?;
 - ¿quienes son los responsables política y técnicamente?;
 - ¿cuales son sus intereses?;
 - ¿cuales son sus fuentes de financiamiento?;
- La metodología usada de colección de datos: ¿sigue normas y practicas establecidas?; ¿son verificables los datos?;
- La calidad de los datos y sus dimensiones; y
- La metodología del análisis y la presentación de los datos
- Combinando estadísticas publicas con datos no convencionales puede enriquecer el análisis. Hay que crear un marco conceptual común y establecer las relaciones entre los variables.

Percepción de corrupción en América Latina y el Caribe

País	Democracia 2017	Percepción Corrupción 2017	Capacidad estadística 2017	PIB p C \$ 2016
Argentina	6.96	39	85.56	10,254.12
Bolivia	5.49	33	67.78	2,454.05
Brasil	6.86	37	74.44	10,771.01
Chile	7.84	67	93.33	14,692.14
Colombia	6.67	37	88.89	7,526.31
Costa Rica	7.88	59	87.78	9,585.47
Ecuador	6.02	32	68.89	5,210.28
El Salvador	6.43	33	92.22	3,815.17
Guatemala	5.86	28	78.89	3,077.89
Guyana	6.46	38	55.56	3,771.86
Haití	4.03	22	36.67	735.77
Honduras	5.72	29	71.11	2,380.84
Jamaica	7.29	44	71.11	4,938.79
México	6.41	29	93.33	9,972.72
Nicaragua	4.66	26	68.89	1,945.62
Panamá	7.08	37	68.89	11,098.97
Paraguay	6.31	29	67.78	3,930.49
Perú	6.49	37	90.00	6,094.22
República Dominicana	6.66	29	83.33	6,881.14
Suriname	6.76	41	72.22	8,268.81
Trinidad y Tabaco	7.04	41	46.67	15,722.12
Uruguay	8.12	70	85.56	14,015.14
Venezuela	3.87	18	68.89	14,300.00

Micro datos

- Micro datos es información a nivel de la unidad de observación del estudio (personas, entidades, hechos). Sus fuentes son registros administrativos, censos y encuestas.
- Tratando de micro datos es importante que se garanticen la anonimidad y confidencialidad del informante y que las variables son comparables sobre el espacio y el tiempo.
- Cuando se requiere información de un solo país es mejor consultar la Oficina Nacional del país.
- En Perú el INEI tiene una rica colección de información reciente accesible por el público en la Base de Datos (<https://www.inei.gob.pe/bases-de-datos/>). Están disponibles:
 - 15 censos nacionales recientes;
 - 264 encuestas de hogares,
 - 40 encuestas de empresas y establecimientos, y
 - 19 archivos basados en registros administrativos.
- Cuando se quiere comparar dos o más países es preferible utilizar un repositorio internacional, de preferencia de una de las agencias internacionales (ONU) o de una universidad.

IHSN:

International Household Survey Network

- El IHSN (<http://www.ihsn.org/>) es un sistema de coordinar y fomentar la cooperación entre las organizaciones internacionales y nacionales que promuevan o que brindan apoyo financiero y/o técnico para la implementación de programas de encuestas nacionales y/o internacionales en países en desarrollo.
- La misión de IHSN es mejorar la disponibilidad, accesibilidad y calidad de los datos de las encuestas y promover el análisis y el uso de los datos por los responsables de la toma de decisiones nacionales e internacionales, la comunidad investigadora y otras partes interesadas.
- El IHSN no solo es un repositorio de datos sino también promueve la metodología y el uso de datos a través de:
 - proyectos,
 - manuales,
 - paquetes de procesamiento de datos (software) y
 - un catálogo de encuestas.

IHSN Catalogo de encuestas

- El Catálogo Central de Encuestas de IHSN es un inventario de encuestas y censos de países de ingresos bajos y medios. El catálogo se mantiene en colaboración con el Banco Mundial. El catálogo proporciona metadatos que incluyen, cuando disponibles, los cuestionarios de la encuestas, los manuales y los informes, y una lista de citas relevantes. No proporciona acceso directo a micro datos, pero proporciona un enlace a catálogos externos donde se pueden obtener los datos.
- El IHSN también mantiene una lista de otros catálogos de encuestas donde los investigadores pueden encontrar datos y metadatos.
- Al 13 de agosto de 2018 el catalogo contenía 6,261 archivos; de 3,517 encuestas hay datos disponibles de repositorios externos.
- Para el Perú hay 61 archivos de 18 estudios distintos.

World Bank Microdata Library

- <http://microdata.worldbank.org/index.php/home> La Biblioteca de micro datos facilita el acceso a los datos recopilados mediante encuestas de hogares, establecimientos comerciales u otras entidades.
- La Biblioteca también contiene datos y documentación de los censos de población, vivienda o agricultura, procesos administrativos y encuestas apoyados por el Banco, y otras organizaciones internacionales y nacionales. Al 13 de agosto contenía 2,577 estudios. Hay 29 archivos de 12 estudios distintos para el Perú.
- Los datos son accesible por diversas modalidades. En casos muy raros los datos no son disponibles, y existen solamente tabulaciones o informes.

Global Health Data Exchange (GHDx)

- <http://ghdx.healthdata.org/>
- GHDx, el catálogo más completo de encuestas, censos, estadísticas vitales, informes y otros datos relacionados con la salud. Para el Peru hay 516 entradas.
- Organizado y mantenido por Institute for Health Metrics and Evaluation (IHME) de la University of Washington, Seattle.
- Tipos de accesibilidad a los datos o a la información:
 - Microdata access: Download (descarga libre)
 - Microdata access: Register (registrar)
 - Microdata access: Request (registrar y justificación)
 - Tabulations only (Solo tabulaciones o el informe)
- Para acceso a los datos se refiere a la fuente de los datos o el archivo donde están localizados.

Encuestas DHS

- Desde 1984, el Programa de Encuestas de Demografía y Salud (DHS) ha proporcionado asistencia técnica a más de 300 encuestas en más de 90 países, avanzando la comprensión global de las tendencias de salud y población en los países en desarrollo. Hay 13 encuestas DHS para el Perú.
- Son encuestas de hogares a nivel nacional sobre la fecundidad, planificación familiar, salud materna e infantil, el género, el VIH / SIDA, la malaria y la nutrición.
- El Programa DHS está financiado por la Agencia de Estados Unidos para el Desarrollo Internacional (USAID), con contribuciones de otros donantes y fondos de los países participantes.
- Los datos de todas las encuestas son disponibles para descargar así como un sistema exhaustivo de apoyo técnica en línea.

UNICEF: Encuestas MICS

- El programa de las encuestas Multiple Indicator Cluster Surveys (MICS) de UNICEF <http://mics.unicef.org/> se inició en 1995 y hasta la actualidad realizaron 304 encuestas en 112 países en 5 rondas, la sexta esta en ejecución actualmente.
- Son encuestas nacionales de hogares sobre mujeres y niños tratando Discapacidad infantil, Mortalidad infantil, Nutrición infantil, Desarrollo infantil temprano, Protección infantil, Pobreza y privación infantil, Abuso y negligencia infantil, Educación, Salud materna, VIH / SIDA, y Agua y saneamiento. Sus contenidos y metodologías son estandarizadas. Tiene un sistema de apoyo técnico para los equipos nacionales.
- Los datos de las encuestas son libremente accesibles a los investigadores, pero se deben registrarse y proporcionar una breve descripción del tema de su estudio; reciben los datos y metadatos de las encuestas que desean utilizar. Si publican los resultados deben mencionar MICS como a fuente de información.
- El Perú no participó en MICS.

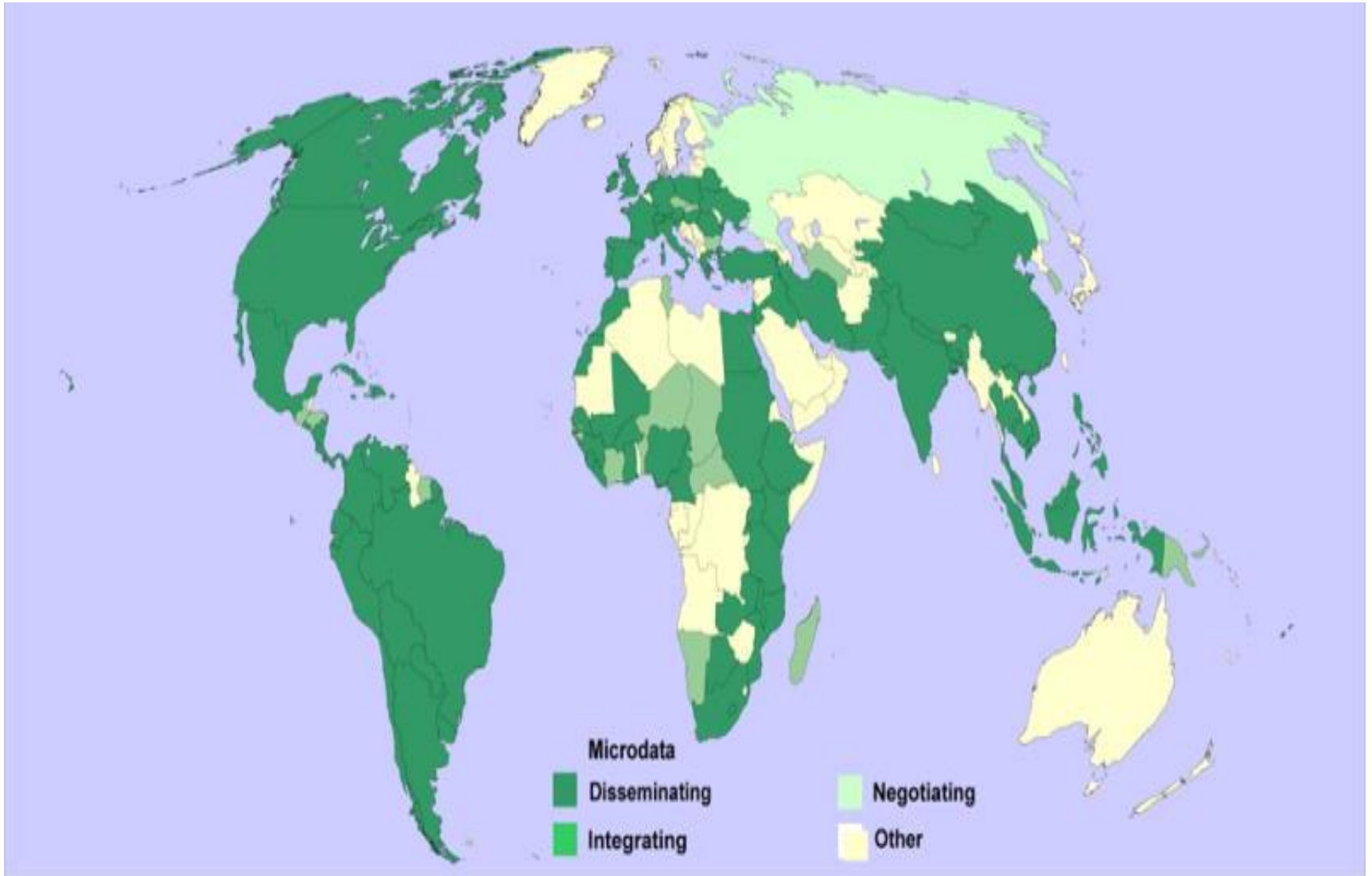
La colección IPUMS

- El Integrated Public Use Microdata Series (IPUMS) (<https://www.ipums.org/>) del Minnesota Population Center (MPC) la Universidad de Minnesota es una colección de nueve programas de proveer datos internacionales y de los EEUU derivados de censos y encuestas de hogar cubriendo una serie de temas: población, educación, salud, medio ambiente a veces relacionados con información geográfico (GOS, GPS).
- La colección se inició con el proyecto IPUMS US liderado por el Prof. Steven Ruggles en los años 90, poniendo a disposición de la comunidad científica muestras de los censos de los EEUU comenzando con aquel de 1850, primer año en el cual registraron individuos en el censo.
- A finales de los años 90 el Prof. Robert McCaa sugirió y propulsó la creación de una colección de censos internacionales, IPUMS International

IPUMS International

- IPUMS International proporciona, de forma gratuita, datos censales internacionales anonimizados e armonizados para la investigación científica social, económica, de la salud y el medio ambiente. Actualmente dispone de datos de 988 millones individuos censados de 345 censos de 89 países.
(<https://international.ipums.org/international/>)
- Salvo Guatemala y Honduras, todo los países de América Latina participan en el IPUMS a partir de la ronda de censos de 1960 con datos de 75 censos y dos encuestas intercensal (México).

Países que proporcionan datos censales



Metodología de IPUMS

- Los archivos de datos disponibles de IPUMS son muestras (10 %) de hogares de los censos confiados a IPUMS en forma voluntaria.
- Los datos de los hogares y sus miembros son verificados y si necesario editados, creando un archivo jerárquica.
- Datos identificadores del hogar y los miembros son modificado para garantizar la anonimidad. Aun así es posible de identificar áreas sub-nacionales.
- IPUMS también crea variables nuevas basándose en las variables originales, por ejemplo relaciones entre miembros del hogar.
- Las variables originales y derivadas son armonizadas de tal forma que el mismo código armonizado tiene el misma significado para todos los censosa lo largo del tiempo.
- IPUMS proporciona una documentación completa del material disponible de cada censo y incluso materiales originales.
- Además hay un servicio de consultas técnicas en línea y gran cantidad de publicaciones basados en los datos IPUMS.

Armonización de Estado civil

Armonizado		Datos de los censos		
Code	Descripción	Bangladesh 2011	México 1970	Kenya 1999
100	Soltero	1 = No casado	8 = Soltero	1 = Jamás casado
200	Casado o en unión	2 = Casado		
210	Casado formal		1 = Casado, Civil y religioso	
211	Civil		2 = Casado, civil	
212	Religioso		3 = Casado, religioso	
213	Civil y religioso			
214	Monógamo			2 = Monógamo
215	Polígamo			3 = Polígamo
220	Unión consensual		4 = Unión consensual	
300	Divorciado/Separado	4 = Divorciado/Separado		
310	Separado		7 = Separado	6 = Separado
320	Divorciado		6 = Divorciado	5 = Divorciado
400	Viudo	3 = Viudo	5 = Viudo	4 = Viudo

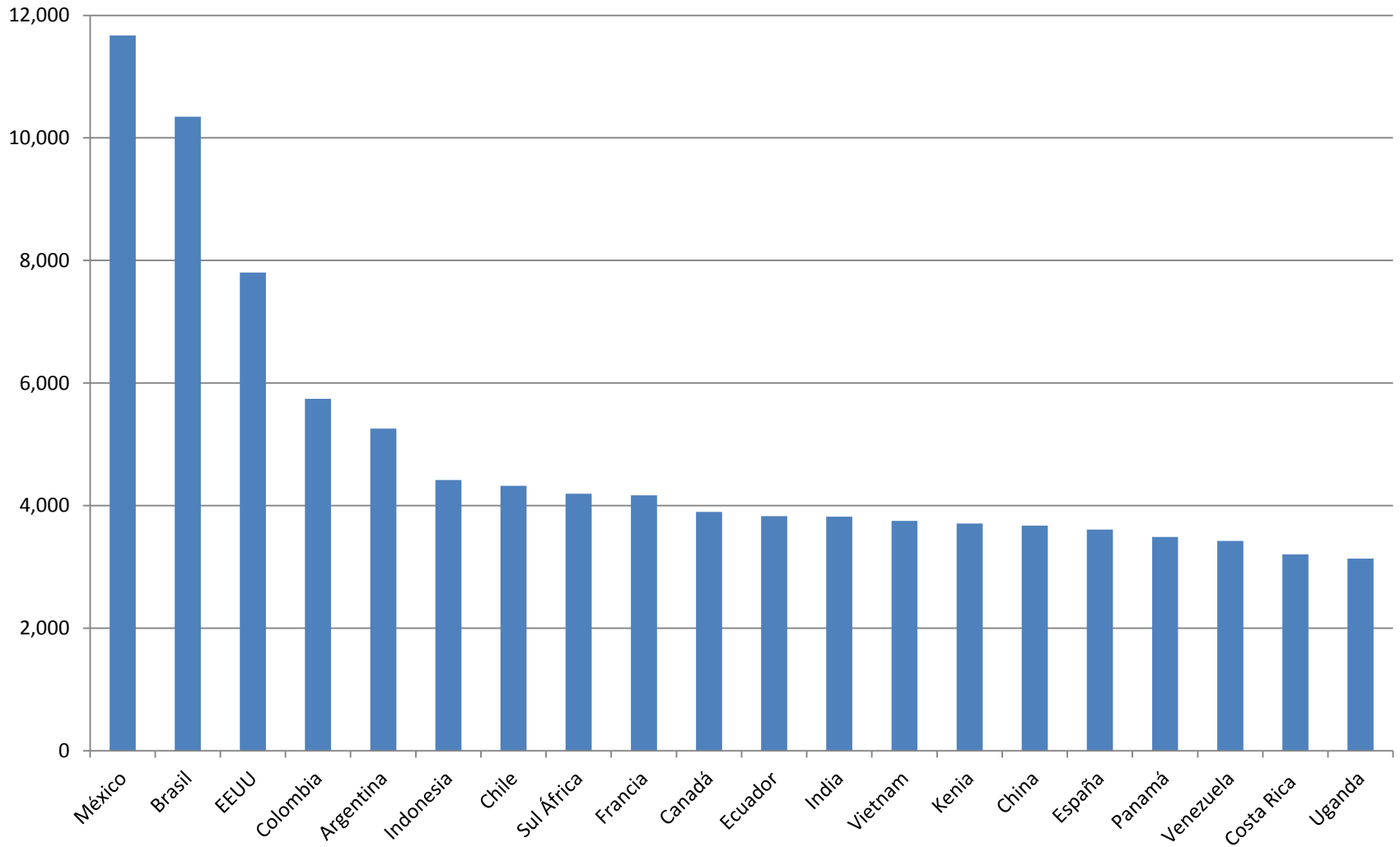
Crterios para acceso

- Hay dos opciones para usar los datos de IPUMS:
 - Como investigador individual:
<https://international.ipums.org/international-action/menu>
 - Como estudiantes en una clase con profesor registrado.
 - https://international.ipums.org/international/classroom_accounts.shtml
- Un investigador individual debe registrarse a través de un proceso en línea, presentar su perfil profesional, una descripción de su investigación y que tipo de datos de que países requiere.
- IPUMS permite a profesores de crear una cuenta de clase.
 - El profesor crea la cuenta y especifica los temas y datos;
 - Los estudiantes de la clase pueden registrarse individualmente en forma simplificado; y
 - El profesor puede verificar quien esta registrado y que datos tiene.
- IPUMS pone a disponibilidad un juego de variables de los países identificados que el investigador/estudiante puede usar en línea o descargarlo para uso posterior.

Los usuarios

- Como institución académica MPC debe asegurar la alta calidad de las investigaciones que apruebe. En los años recientes la tasa de aprobación es aproximadamente 60 %, con 2,000 nuevos usuarios anuales.
- La gran mayoría de usuarios tiene relaciones académicos: 66 % son estudiantes, 11 % docentes y 14 % investigadores académicos.
- Entre las personas sin afiliación académica hay funcionarios de las Naciones Unidas, el Banco Mundial, la Organización Mundial de la Salud y el Banco Inter-americano de Desarrollo.
- Por disciplinas académicas:
 - Un poco más del 50% de economía,
 - 16% de demografía,
 - 8% de sociología y
 - 5% cada uno de políticas públicas y de estadísticas
- Los 10 países con el mayor número de usuarios en forma descendientes son los Estados Unidos, el Reino Unido, Alemania, Canadá, Francia, España, Brasil, China, Argentina, y Japón.

Los 20 países más analizados



Datos y usuarios de América Latina

- De los 20 censos nacionales más analizados 11 son de países de América y 9 son de América Latina.
- De los 89 países contribuyentes al IPUMS 18 son de América Latina.
- De los 345 censos disponibles para el análisis 77 o sea el 22 % son de América Latina.
- De los 2021 usuarios nuevos en 2015, 216 o sea 11 % eran de América Latina.
- En 2015 los países con 5 y más usuarios en orden ascendente eran México, Brasil, Chile, Colombia, Uruguay y Ecuador. Países con dos usuarios eran Costa Rica, Haití, Perú y Venezuela, mientras hubo un usuario en Nicaragua y República Dominicana.
- Países sin usuarios eran Bolivia, Cuba, El Salvador, Panamá y Paraguay.

Liderazgo de IPUMS en armonización

- Proyecto IPUMS-DHS: <https://www.idhsdata.org/idhs/> En este proyecto, manejado por IPUMS, variables de 18 encuestas de DHS de 24 países Africanos y 4 países Asiáticos han sido armonizadas de tal forma que son comprable por el espacio (entre países) y el tiempo.
- La armonización de las variables permite crear poblaciones sintéticas uniendo dos o más encuestas.
- UNICEF reconociendo el liderazgo de IPUMS en armonizar de variables y archivar datos estadísticas esta colaborando con IPUMS para el desarrollo del prototipo del MICS Tabulator, una plataforma de análisis de datos basada en el internet que permitirá una investigación rápida, sencilla, exhaustiva y comparativa de distintos países en el tiempo.
- Ver: http://mics.unicef.org/news_entries/104/NEW-PARTNERSHIP-WITH-IPUMS

Conclusiones y opciones futuras

- En el internet hay ahora datos de alta calidad a nivel de países y micro datos que permite realizar estudios científicos sobre temas diversos.
- Agencias no gubernamentales tratan temas no convencionales relacionadas a la democracia, transparencia y responsabilidad (accountability) que pueden ser analizados en conjunto con estadísticas publicas.
- Muchos de los datos estadísticos en el internet son series históricos. Este permite de estudiar una serie de problemas con dimensiones históricos en población, desarrollo y el medio ambiente, por separado o en combinación..
- Esto requiere que se construyen modelos dinámicos basándose en ,los datos disponibles a nivel de país, pero también se puede incluir en el modelo datos derivados de micro datos si no existen a nivel país.
- Los modelos también pueden ser de niveles múltiples, relacionando datos del nivel global, regional, nacional y su-nacional con características individuales.

Más conclusiones y opciones futuras

- Los micro datos armonizados permite combinar datos de dos o mas encuestas en una sintética, respectando las reglas metodológicas (definiciones) y estadísticas (tasas de selección y ponderaciones).
- Los archivos de IPUMS International en este sentido son de especial interés para el mundo académico puesto que ya han creado datos limpios de alta calidad. **¡Lo único que el investigador debe hacer es el trabajo intelectual!**
- Los países de América Latina participan muy activamente en IPUMS International como proveedor de datos.
- Los investigadores latino americanos no han aprovechado esta fuente de datos de alta calidad.
- **¡Recomiendo que se utilicen esta fuente de datos en la enseñanza académica por su alta calidad y las ventajas practicas que ofrece!**



¡GRACIAS POR SU ATENCIÓN!

Referencias

- Robert McCaa, Thanks to 70 years of Inter American Statistical cooperation, the world's largest integrated census microdata dissemination site www.ipums.org/international, *Estadastica*. 2013 June ; 65(184): 31–45.
<https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4445070&blobtype=pdf>
- Alphonse L. MacDonald, IPUMS International: A review and future prospects of a unique global statistical cooperation programme; *Stat J IAOS*. 2016; 32(4): 715– 727.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5565170/>
- Kristen Jeffers, Miriam King, Lara Cleveland and Patricia Kelly Hall, Data Resource Profile: IPUMS-International, *International Journal of Epidemiology*, 2017, 390–391, doi: 10.1093/ije/dyw321, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5837405/>