



**Representación mediante grafos de conocimiento de
trayectorias académicas y producción científica de becarios/as
de doctorado BECAS CHILE y CONICYT (2014–2016) en
disciplinas STEM y Ciencias Sociales (Versión Beta para
completar índice preliminarmente)**

Rodrigo Molina Ávila

Magíster en Ciencia de Datos
Universidad de Chile

Profesor guía: Sebastián Ferrada Aliaga
Profesora co-guía: Roxana Chiappa Baros

Santiago, Chile
16 de julio de 2025

Índice

1	Introducción	3
1.1	Justificación del estudio	3
1.2	Contexto general: educación superior, desigualdades y STEM	3
1.3	Perspectiva analítica y propuesta metodológica	3
2	Antecedentes	4
2.1	Desigualdades estructurales en educación superior chilena	4
2.2	Modelo androcéntrico en disciplinas STEM	4
2.3	Necesidad de un enfoque interseccional	4
2.4	Limitaciones de la evidencia nacional	4
2.5	Evidencia de la intersección clase-género	5
2.6	Relevancia de los grafos de conocimiento	5
2.7	Marco institucional	5
3	Marco Teórico	6
3.1	Teoría feminista y crítica de la ciencia	6
3.2	Interseccionalidad como enfoque analítico	6
3.3	Campo científico y capital académico	6
3.4	Grafos de conocimiento y modelamiento relacional	6
4	Pregunta de investigación y objetivos	7
4.1	Pregunta de investigación	7
4.2	Objetivo general	7
4.3	Objetivos específicos	7
5	Metodología	8
5.1	Enfoque metodológico	8
5.2	Modelado conceptual y esquema entidad-relación	8
5.3	Fuentes de datos utilizadas	8
5.4	Transformaciones aplicadas a los datos	9
5.5	Construcción del grafo de conocimiento	9
5.6	Ánalysis estructural y detección de patrones	10
5.7	Justificación del enfoque grafo	10

5.8 Resultados esperados	10
Bibliografía	12
Anexos: EDA en Power BI	14

1 Introducción

1.1 Justificación del estudio

La producción científica y académica se encuentra atravesada por múltiples ejes de desigualdad que no sólo afectan la cantidad de resultados, sino también su contenido, orientación temática, legitimación epistémica y circulación. En este marco, comprender cómo el género y la clase social configuran las trayectorias formativas y epistémicas de los/as investigadores/as se vuelve un imperativo analítico y político, particularmente en contextos de alta estratificación como el chileno.

1.2 Contexto general: educación superior, desigualdades y STEM

La educación superior en Chile ha estado históricamente marcada por profundas desigualdades sociales que se expresan en el acceso, permanencia y éxito académico. En el caso específico de las disciplinas STEM, estas desigualdades adoptan formas estructurales donde la hegemonía masculina y la exclusión de sectores populares se encuentran fuertemente institucionalizadas.

1.3 Perspectiva analítica y propuesta metodológica

Este estudio adopta una perspectiva interseccional y relacional, operativizada mediante el uso de grafos de conocimiento. Esta propuesta metodológica permite integrar dimensiones múltiples de análisis, habilitando formas de visualización y exploración estructurada de trayectorias académicas complejas.

2 Antecedentes

2.1 Desigualdades estructurales en educación superior chilena

Estudios recientes han documentado desigualdades estructurales en la educación superior chilena mediadas por clase y género (Chiappa & Pérez Mejías, 2019; Rodríguez & Archer, 2022). Sin embargo, persiste una carencia de análisis empíricos que integren múltiples dimensiones de estratificación social en la explicación de trayectorias académicas.

2.2 Modelo androcéntrico en disciplinas STEM

La educación superior en STEM ha reproducido históricamente una organización androcéntrica, donde el género actúa como principio estructurante del poder institucional (Acker, 1991). Sólo el 19,7 % de las matrículas de primer año en STEM en Chile en 2023 correspondieron a mujeres (MINEDUC, 2023).

2.3 Necesidad de un enfoque interseccional

El análisis exclusivo del género resulta insuficiente. La perspectiva interseccional (Crenshaw, 1989; Collins, 2019) permite identificar dinámicas de poder surgidas de la interacción simultánea de género, clase y otras categorías. Ignorar estas complejidades conduce a interpretaciones sesgadas e ineficaces.

2.4 Limitaciones de la evidencia nacional

En Chile, la investigación sobre trayectorias académicas ha privilegiado el análisis del género como eje único, descuidando el efecto conjunto de la clase social (Rodríguez & Archer, 2022; Sepúlveda & Lizama-Loyola, 2022). Esto debilita la formulación de políticas públicas eficaces orientadas a la equidad estructural.

2.5 Evidencia de la intersección clase-género

El tipo de establecimiento escolar de origen actúa como proxy de clase social, y junto al género, media el acceso, permanencia y legitimación en la academia STEM (Chiappa & Pérez Mejías, 2019; Canales, 2016). Investigadores de clases sociales bajas presentan menor acceso a redes académicas de élite y menor legitimación epistémica.

2.6 Relevancia de los grafos de conocimiento

La complejidad estructural de las trayectorias académicas interseccionales, donde múltiples atributos género, clase social, prestigio institucional, campos disciplinares se articulan de manera no lineal, exige métodos capaces de representar y analizar relaciones múltiples de alta dimensionalidad.

Los grafos de conocimiento, definidos formalmente como estructuras de datos orientadas a representar entidades y relaciones semánticamente Enriquecidas (Hogan et al., 2021), permiten modelar trayectorias académicas como redes dinámicas, en las cuales los nodos representan agentes, instituciones y objetos científicos, y las aristas codifican vínculos de formación, producción o colaboración.

Esta metodología supera las limitaciones de los enfoques tabulares tradicionales, al capturar explícitamente patrones de asociación compleja, jerarquías relacionales y configuraciones de capital social y simbólico (Bordes et al., 2013).

2.7 Marco institucional

Este proyecto se enmarca en el Fondecyt Regular N°11241304, Re-conociendo las desigualdades de la Academia Chilena: Un análisis interseccional de género y clase en las trayectorias de personas con grado de doctorado. Utiliza un corpus que integra datos administrativos, información escolar y universitaria, rankings internacionales, producción científica y proyectos FONDECYT adjudicados.

3 Marco Teórico

3.1 Teoría feminista y crítica de la ciencia

La crítica feminista ha develado cómo el conocimiento científico y académico ha sido históricamente producido desde marcos epistémicos androcéntricos que excluyen, subordinan o marginan otras formas de producción de conocimiento (Harding, 1991; Haraway, 1988). La ciencia no es neutral ni objetiva, sino situada, permeada por relaciones de poder y estructurada por jerarquías sociales.

3.2 Interseccionalidad como enfoque analítico

Desde Crenshaw (1989) hasta Collins (2019), la interseccionalidad se ha consolidado como un enfoque que permite analizar cómo múltiples sistemas de dominación género, clase, raza interactúan de forma simultánea y co-constitutiva en la configuración de desigualdades. No se trata de sumar variables, sino de comprender estructuras relacionales complejas que operan de manera contextual.

3.3 Campo científico y capital académico

Desde la sociología de la ciencia, autores como Bourdieu (2008) han aportado herramientas clave para pensar la estructuración del campo académico como un espacio de luchas simbólicas, donde distintos agentes compiten por legitimidad, prestigio y recursos. En este campo, el capital científico cultural, simbólico, social se acumula diferencialmente según el origen social, género, institución y red de contactos.

3.4 Grafos de conocimiento y modelamiento relacional

El uso de grafos de conocimiento en ciencias sociales permite representar datos complejos como redes de relaciones entre entidades heterogéneas (personas, universidades, disciplinas, publicaciones). Este enfoque relacional y semántico permite identificar patrones estructurales, jerarquías y trayectorias académicas con mayor precisión analítica

que los enfoques tradicionales (Hogan et al., 2021; Bordes et al., 2013).

4 Pregunta de investigación y objetivos

4.1 Pregunta de investigación

¿Cómo incide el posicionamiento interseccional de **clase social y género** en la **configuración temática, estructural e institucional** de la producción científica de los/as becarios/as doctorales en Chile (2014–2016), según representaciones mediante **grafos de conocimiento**?

4.2 Objetivo general

Comprender cómo el posicionamiento interseccional de **clase social y género** incide en la **configuración temática** de la producción científica de los/as becarios/as doctorales chilenos/as de disciplinas **STEM y Ciencias Sociales**, a partir de representaciones estructuradas mediante **grafos de conocimiento**.

4.3 Objetivos específicos

1. Identificar las **entidades, atributos y relaciones** relevantes que caracterizan las trayectorias académicas de los/as becarios/as, considerando variables **socio-demográficas, institucionales y de producción científica**.
2. Conocer las formas de asociación entre atributos sociales (**clase social, género**) y áreas temáticas de investigación, diferenciando las configuraciones disciplinares entre **STEM y Ciencias Sociales**.
3. Explorar las estructuras generadas por los **grafos de conocimiento**, focalizando en **patrones de movilidad, segmentación temática, jerarquización institucional y desigualdad epistémica interseccional**.

5 Metodología

5.1 Enfoque metodológico

El presente estudio adopta una estrategia cuantitativa de tipo estructural, orientada a representar trayectorias académicas y científicas mediante grafos de conocimiento. Este enfoque permite la integración de datos multifuente provenientes de registros administrativos (ANID), bases educacionales (secundaria, pregrado, posgrado), sistemas de información científica (Scopus) y registros de financiamiento público (FONDECYT).

5.2 Modelado conceptual y esquema entidad-relación

Siguiendo los lineamientos clásicos de modelado entidad-relación (Chen, 1976), se construyó un modelo conceptual que identifica:

- **Entidades:** personas, instituciones, programas académicos, publicaciones, proyectos FONDECYT, disciplinas OCDE y países.
- **Atributos observables:** género, clase social de origen (proxy: tipo de secundaria), país de formación, tipo de institución, área temática, indicadores bibliométricos.
- **Relaciones:** formación, afiliación, publicación, financiamiento, jerarquización institucional.

La validación estructural del modelo fue realizada mediante la herramienta ERDoc, asegurando la coherencia semántica y la consistencia lógica entre entidades y relaciones.

5.3 Fuentes de datos utilizadas

[label=)]**Becas y formación:** Base de datos administrativa ANID del programa BECAS CHILE.

1. ▪ *maestra_latest*: tabla central con características del becario/a (FOLIO, género, etnia, área OCDE, año de inicio).

- *secundaria*: tipo de establecimiento, región, GSE.
 - *pregrado*: nombre de universidad, acreditación, región.
 - *posgrado*: universidad y su jerarquía (QS, Shanghai).
2. **Producción científica:** API Scopus.
- *author_id, scopus_author_search, abstract_search*.
 - Información detallada sobre productividad, afiliaciones y posicionamiento temático.
3. **Ranking institucional:** QS World University Rankings y Shanghai Ranking.
4. **Proyectos Fondecyt:** base de datos de adjudicación (nombre, monto, área, duración, macrozona, etc.).
5. **Emparejamiento de autoría:** vinculación entre FOLIO y author_id Scopus.

5.4 Transformaciones aplicadas a los datos

- **Homologación institucional:** diccionario normalizado para QS y Shanghai.
- **Unión de entidades por claves:** FOLIO, author_id, scopus_id, ID.
- **Modelo estrella en Power BI:** preparación estructural para el modelado grafo.

5.5 Construcción del grafo de conocimiento

Sobre un universo de 3.143 becarios/as (cohorte 20142016), se construyó un grafo orientado con los siguientes elementos:

- **Nodos:** persona, secundaria, pregrado, posgrado, publicación, proyecto Fondecyt, disciplina, país.
- **Relaciones:** estudió en, obtuvo posgrado en, publicó en, fue financiado por, pertenece a disciplina.
- **Atributos:** género, clase social, indicadores bibliométricos, posicionamiento institucional.

La base de datos utilizada es Neo4j, compatible con consultas en lenguaje Cypher y análisis de redes complejas. Se contempla el uso de MillenniumDB para etapas posteriores.

5.6 Análisis estructural y detección de patrones

Se aplicarán algoritmos clásicos de análisis de redes sociales, incluyendo:

- **Centralidad de cercanía (closeness):** identificación de nodos centrales en la red.
- **Intermediación (betweenness):** nodos puente en trayectorias académicas.
- **Detección de comunidades (Leiden):** particiones temáticas y disciplinares.
- **Similitud topológica (Jaccard):** evaluación de trayectorias análogas.

5.7 Justificación del enfoque grafo

Según Bro y Mendoza (2021), los grafos permiten modelar fenómenos de estructuración social mediante relaciones semánticas y jerárquicas no visibles en enfoques tabulares. La adopción de esta técnica metodológica constituye una innovación frente a la agregación tradicional, al capturar simultáneamente redes de afinidad, exclusión y posicionamiento epistémico.

5.8 Resultados esperados

- Construcción de un grafo de conocimiento para la cohorte 20142016.
- Identificación de patrones de movilidad, segmentación temática y vínculos epístémicos.
- Métricas de estructura: centralidad, modularidad, grado, etc.
- Agrupamientos disciplinares según género y clase social.
- Análisis de desigualdades en redes de prestigio académico.

Este enfoque busca fortalecer la capacidad explicativa de los análisis de trayectorias académicas en Chile mediante un modelo formal replicable, visualizable y analíticamente potente.

Bibliografía

- Acker, J. (1990). Hierarchies, jobs, bodies: A theory of gendered organizations. *Gender & Society*, 4(2), 139158. <https://doi.org/10.1177/089124390004002002>
- Bro, N., & Mendoza, M. (2021). Surname affinity in Santiago, Chile: A network-based approach that uncovers urban segregation. *PLOS ONE*, 16(1), e0244372. <https://doi.org/10.1371/journal.pone.0244372>
- Canales, A. (2016). *La reproducción de la élite en Chile: Educación y meritocracia*. Ediciones Universidad Diego Portales.
- Chen, P. P.-S. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems*, 1(1), 936. <https://doi.org/10.1145/320434.320440>
- Chiappa, R., & Pérez Mejías, P. (2019). Inequalities in higher education in Chile: Discourses, practices and policies. In A. Furlong, F. Kintrea, & T. Biggart (Eds.), *Higher education and social inequalities: University admissions, experiences and outcomes* (pp. 225245). Routledge.
- Christensen, A. D., & Jensen, S. Q. (2012). Doing intersectional analysis: Methodological implications for qualitative research. *NORA - Nordic Journal of Feminist and Gender Research*, 20(2), 109125. <https://doi.org/10.1080/08038740.2012.673505>
- Collins, P. H. (2019). *Intersectionality as critical social theory*. Duke University Press.
- Connell, R. (2007). *Southern theory: The global dynamics of knowledge in social science*. Polity Press.
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), 139167. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8/>
- Guzmán-Valenzuela, C., Ortega, L., Montero, M., & Pérez Mejías, P. (2022). Situated knowledge, cognitive injustice and epistemic resistance in Chilean higher education: Toward a decolonial science. *Frontiers in Sociology*, 7, 938407. <https://doi.org/10.3389/fsoc.2022.938407>

- Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutiérrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), 137. <https://doi.org/10.1145/3447772>
- McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society*, 30(3), 17711800. <https://doi.org/10.1086/426800>
- Rodríguez, J. M., & Archer, L. (2022). Posicionamientos sociales y trayectorias científicas en Chile: una mirada interseccional. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 20(1), 118. <https://doi.org/10.11600/1692715x.2019200108>
- UNESCO. (n.d.). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. <https://unesdoc.unesco.org/ark:/48223/pf0000253479>

Anexos: Eda En Power BI

Tabla:

Becarios

Fuente:

.Solicitada por la Investigadora principal del proyecto Fondecyt a través de portal de transparencia.

N Total

3143



Consideraciones:

- Se construyó la variable Sexo a través del paquete genderizR de R más supervisión humana individual.
- Se corrigió 2 casos de becarios con dos folios diferentes.

Becarios Cohorte 2014-2016

Proporción Sexo



1328

Femenino

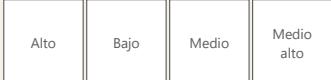
1815

Masculino

Sexo



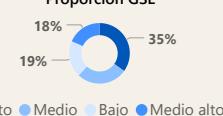
Grupo Socio-económico



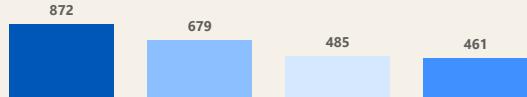
3143

N

Proporción GSE



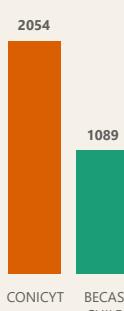
Total GSE



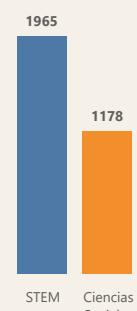
País Posgrado



Total Tipo Beca



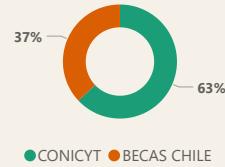
Total Área OCDE



País Posgrado Ciencias Sociales STEM Total

País Posgrado	Ciencias Sociales	STEM	Total
Alemania	38	64	102
ARGENTINA	3	1	4
AUSTRALIA	20	41	61
AUSTRIA	1	3	4
BÉLGICA	7	3	10
BRASIL	8	13	21
Canadá	16	15	31
Total	1178	1965	3143

Proporción Tipo Beca



Proporción Área OCDE



Tabla:

Pregrado

Fuente:

- Portal del Investigador.
- Exploración de personal técnico del proyecto Fondecyt.
- Comisión Nacional de Investigación (CNA)

N Muestra	N Total
3084	3143



Consideraciones:

Solo existen 3084 valores con información del paso por la institución de pregrado del total de los becarios.
Luego de la exploración y búsqueda se acepta este número.

Pregrado

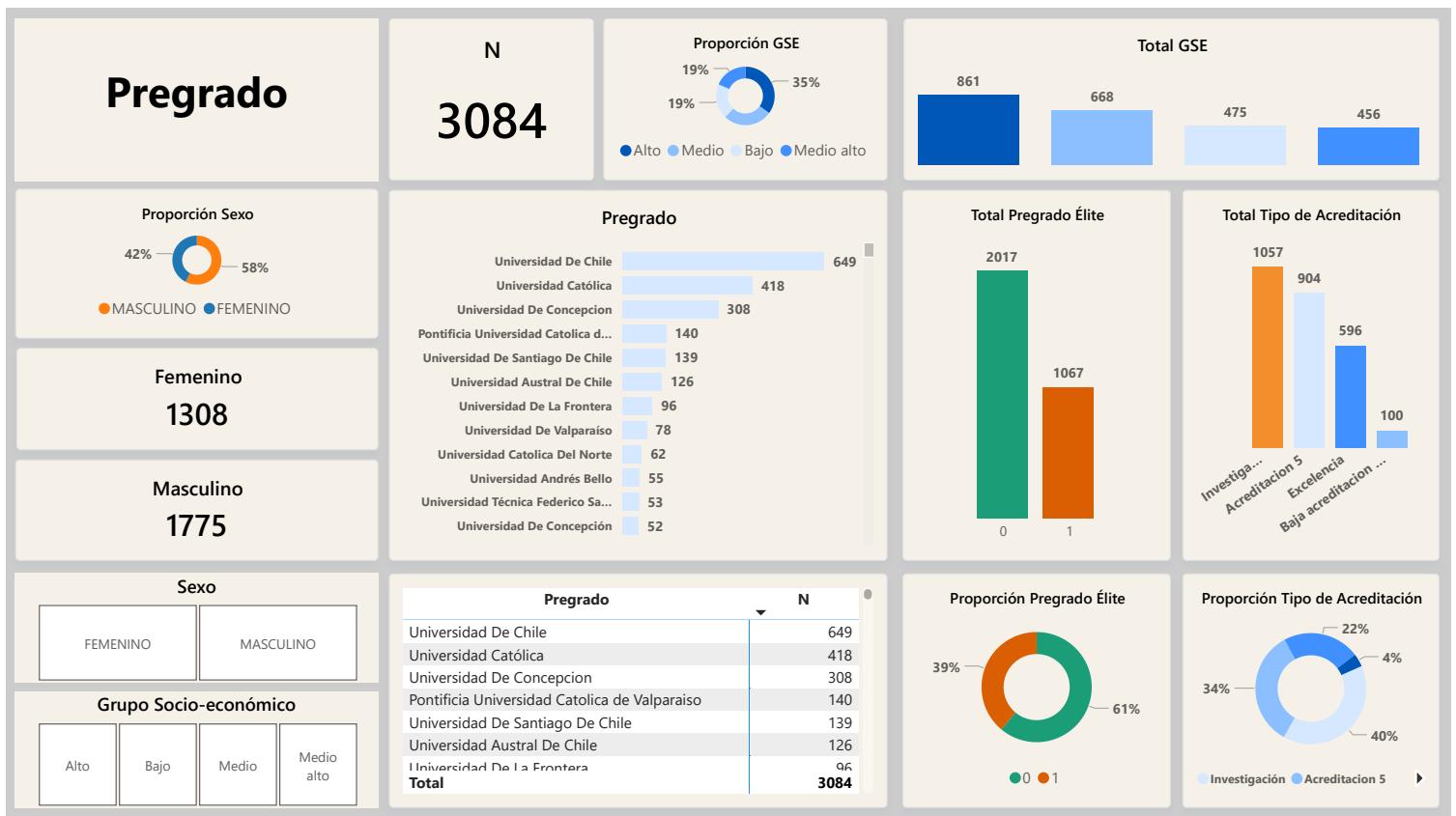


Tabla:

Secundaria

Fuente:

- Ministerio de Educación.
- Extracción por Webscraping: Python + Selenium

N Muestra	N Total
2619	3143



Consideraciones:

- Solo existen 2619 valores con información valida.
- La información se cruza con el valor RBD de los becarios.
- El valor RBD de los becarios proviene de lo otorgado desde el portal de investigador.
- Existe parte de la muestra que estudió en el extranjero su enseñanza secundaria.

Secundaria

N
2619



Total GSE



Proporción Sexo



Femenino

1074

Masculino

1498

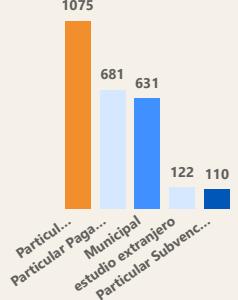
Secundaria

estudio extranjero	122
LICEO INSTITUTO NACIONAL	94
LICEO CARMELA CARVAJAL DE PRAT	39
LICEO CAMILO HENRIQUEZ DE TEMUCO	35
LICEO JOSE VICTORINO LASTARRIA	28
LICEO DE NIÑAS	25
LICEO ABATE MOLINA	24
COLEGIO SAN IGNACIO	23
LICEO DE APLICACION RECTOR J...	23
COLEGIO SAINT GEORGE S COLL...	21
LICEO JAVIERA CARRERA	18
LICEO MANUEL BARROS BORG...	18

Total Emblemático



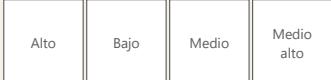
Total Tipo de Establecimiento



Sexo



Grupo Socio-económico

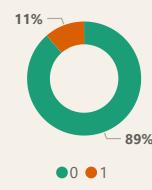


Nombre

Recuento de ID

estudio extranjero	122
LICEO INSTITUTO NACIONAL	94
LICEO CARMELA CARVAJAL DE PRAT	39
LICEO CAMILO HENRIQUEZ DE TEMUCO	35
LICEO JOSE VICTORINO LASTARRIA	28
LICEO DE NIÑAS	25
LICEO DRATE MOJINA	24
Total	2619

Proporción Emblemático



Proporción Tipo de Establecimiento



Tabla:

Calidad Posgrado: Ranking Shanghái

Fuente:

- Página web Ranking Shanghai
- Extracción por Webscraping: Consola de Explorador a través de código en lenguaje Javascript



Consideraciones:

- Solo existen 2112 valores de becarios con universidades de posgrado que sean parte de este ranking.
- Se entiende que no todas las universidades a las cuales se opta en las Becas Chile y Conicyt pertenecen a un programa de una universidad perteneciente a un ranking.
- Los valores son del ranking 2024.
- Construcción de variable si pertenece al top 100 de este ranking.

Calidad Posgrado: Ranking Shanghái

N
2112



Total GSE



Proporción Sexo



Femenino

865

Masculino

1247

Sexo



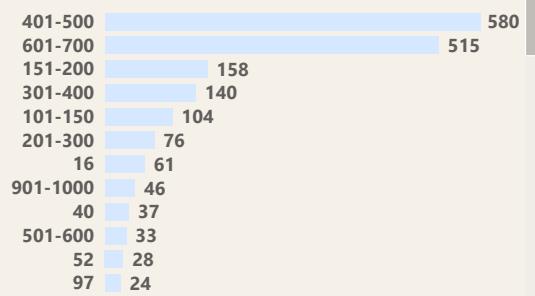
Grupo Socio-económico



País Ranking Shanghái



Total Rank



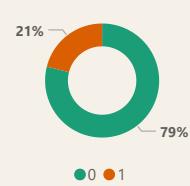
Nombre Universidad

Rank

N

University of Chile	401-500	552
Pontifical Catholic University of Chile	601-700	507
University College London	16	61
Autonomous University of Barcelona	301-400	58
University of Barcelona	151-200	52
Universidad Andrés Bello	901-1000	40
Total	1	2112

Proporción Top 100



Total Top 100

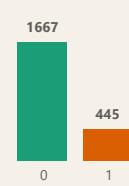


Tabla:

Calidad Posgrado: Ranking QS

Fuente:

- Página web Ranking Shanghai
- Descarga base de datos de la página.

N Muestra	N Total
3077	3143



Consideraciones:

- Existen 3077 valores de becarios con universidades de posgrado que sean parte de este ranking.
- Se entiende que no todas las universidades a las cuales se opta en las Becas Chile y Conicyt pertenecen a un programa de una universidad perteneciente a un ranking.
- Los valores son del ranking 2025.
- Construcción de variable si pertenece al top 100 de este ranking.

Calidad Posgrado: Ranking QS



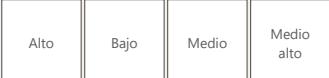
Femenino
1298

Masculino
1779

Sexo



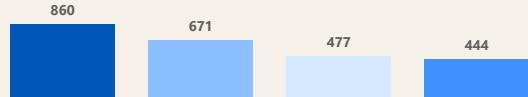
Grupo Socio-económico



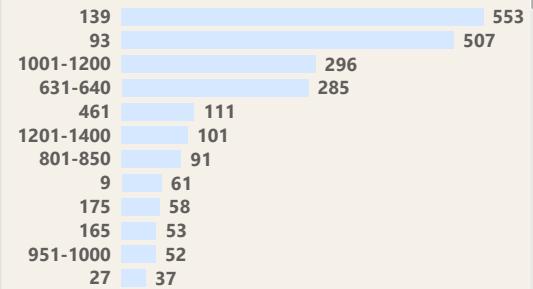
N
3077



Total GSE

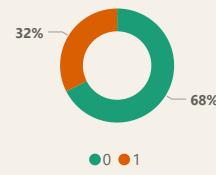


Total Rank



Nombre Universidad	Rank	N
University of Chile	139	552
Pontifical Catholic University of Chile	93	507
Universidad De Concepción	631-640	283
Universidad De Santiago De Chile	461	111
Universidad Austral De Chile	1001-1200	104
Universidad De La Frontera	1001-1200	91
Pontificia Universidad Católica De	801-850	87
Total	1	3077

Proporción Top 100



Total Top 100

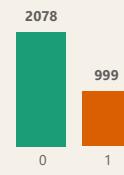


Tabla:

Scopus Author API

Fuente:

- Extracción desde la API de programación de ElSevier.
- Extracción posibilitada a través de una API_KEY y una INST_TOKEN que permiten una quota de extracción de los servidores a través de requests por HTML.
 - Endpont: author_retrieval



Consideraciones:

- La identificación y validación de los becarios se realizó por el personal técnico del proyecto.
- Existen solo 2712 valores de becarios con identificadores como autor dentro de la página web y bases de datos de Scopus/ElSevier.
- Se entiende que no todos los becarios optan por carreras académicas.
- Se entiende que no todos los académicos publican exclusivamente en revistas indexadas.

Scopus Author API

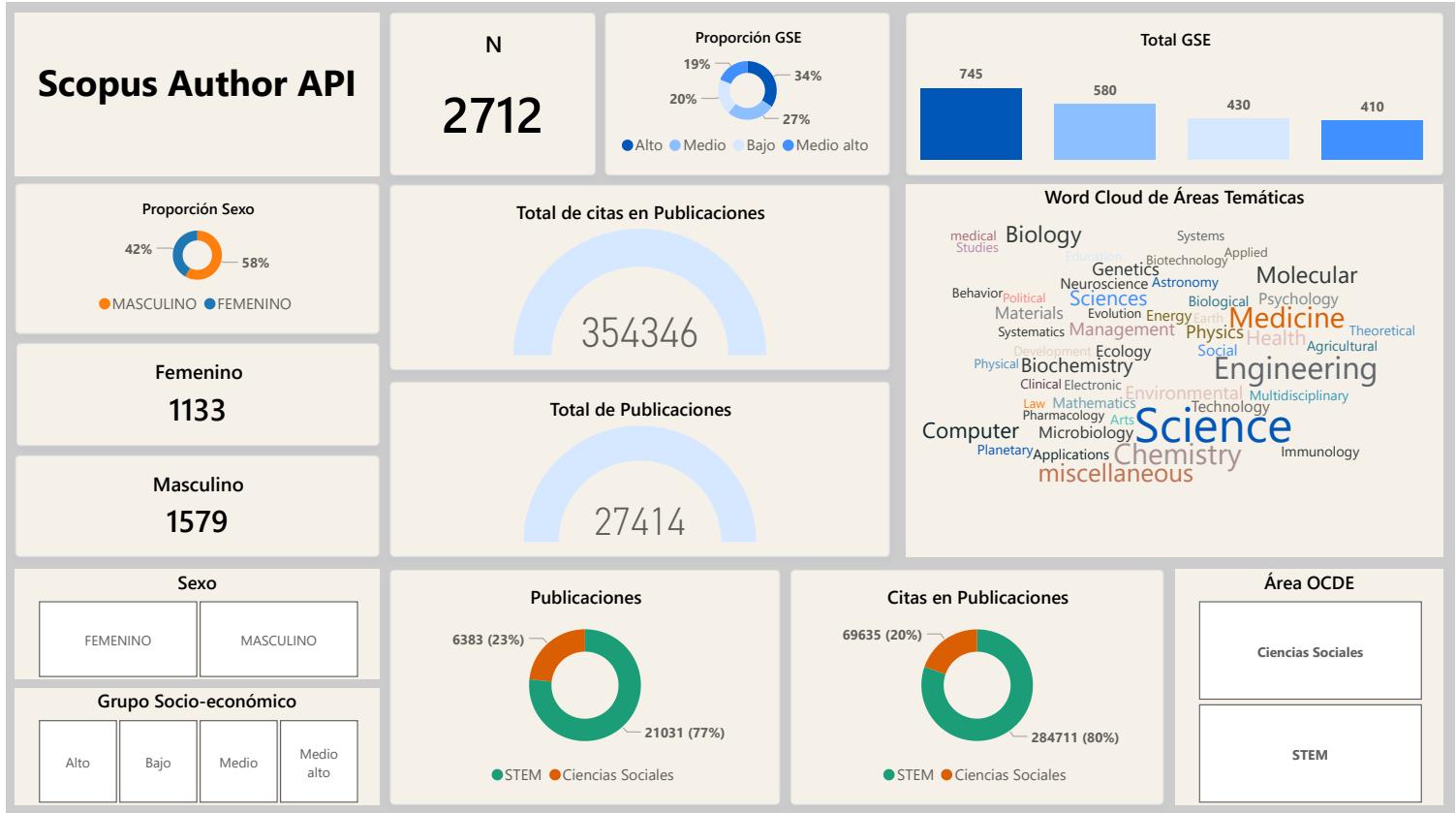


Tabla:

Scopus Abstract API

Fuente:

- Extracción desde la API de programación de ElSevier.
- Extracción posibilitada a través de una API_KEY y una INST_TOKEN que permiten una quota de extracción de los servidores a través de requests por HTML.
- Endpont: abstract_retrieval



Consideraciones:

- La información de los ID de las publicaciones se consiguió a través de Author API.
- Se pudieron identificar y obtener datos de 16769 publicaciones en Scopus/ElSevier de un total de 27414 que corresponde a la suma de document_count de cada becario individualmente indicado en Author API.

Scopus Abstract API

