

# **Desigualdades interseccionales en la ciencia**

Diego Kozlowski<sup>1</sup>, Vincent Larivière<sup>2</sup>, Cassidy R. Sugimoto<sup>3</sup> and Thema Monroe-White<sup>4</sup>

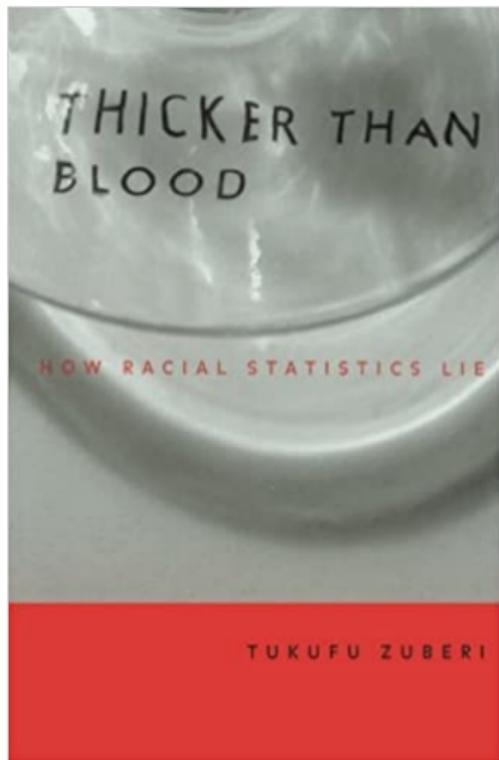
<sup>1</sup> University of Luxembourg, Luxembourg

<sup>2</sup> Université de Montréal, Montréal, QC, Canada

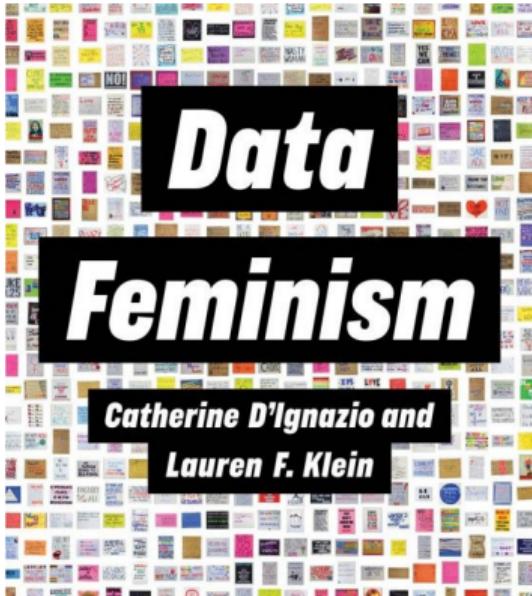
<sup>3</sup> Georgia Institute of Technology, Atlanta, GA, USA

<sup>2</sup> Berry College, Mount Berry, GA, USA

(Des)programando la (des)igualdad



*“La racialización de los datos es un artefacto tanto de las luchas por preservar como por destruir la estratificación racial.” (Zuberi 2002)*



*“Contar y medir no siempre tienen que ser herramientas de opresión. También podemos usarlos para hacer que el poder rinda cuentas.”* (D'Ignazio and Klein 2020)

## FIRST CENSUS OF THE UNITED STATES.

*of the United States as returned at the First Census, by states: 1790.*

DISTRICT.	Free white males of 16 years and upward, including heads of families.	Free white males under 16 years.	Free white females, including heads of families.	All other free persons.	Slaves.	Total.
	22,435	22,328	40,505	255	1 16	2 85,559
	36,086	34,851	70,160	630	158	141,885
	24,384	24,748	46,870	538	None.	96,540
	95,453	87,289	190,582	5,463	None.	378,787
	16,019	15,799	32,652	3,407	948	68,825
	60,523	54,403	117,448	2,808	2,764	237,946

- ▶ El censo de EE. UU. es uno de esos casos donde el racismo estructural se codifica. Por ejemplo, el primer censo de EE. UU. (1790) distingue "Esclavos", "Blancos libres" y "Otra persona libre",
- ▶ sin embargo, es una herramienta que puede servirnos para a comprender (y superar) las desigualdades.

## Bases de datos bibliometricas

- ▶ Nuestro objeto de estudio es la discriminación en la academia (En EE. UU.)
- ▶ Hay mucha evidencia cualitativa de discriminación estructural en la academia,
- ▶ pero el análisis bibliométrico a gran escala tiene el poder de poner un número a aquellas experiencias individuales.
- ▶ Las bases de datos bibliométricas no tienen información sobre la raza<sup>1</sup> autopercibida del autor,
- ▶ Si pudiéramos conocer la raza de los autores, seríamos capaces de desentrañar múltiples dimensiones del racismo estructural en la academia.

## Bases de datos bibliometricas

- ▶ Nuestro objeto de estudio es la discriminación en la academia (En EE. UU.)
- ▶ **Hay mucha evidencia cualitativa de discriminación estructural en la academia,**
- ▶ pero el análisis bibliométrico a gran escala tiene el poder de poner un número a aquellas experiencias individuales.
- ▶ Las bases de datos bibliométricas no tienen información sobre la raza<sup>1</sup> autopercibida del autor,
- ▶ Si pudiéramos conocer la raza de los autores, seríamos capaces de desentrañar múltiples dimensiones del racismo estructural en la academia.

## Bases de datos bibliometricas

- ▶ Nuestro objeto de estudio es la discriminación en la academia (En EE. UU.)
- ▶ Hay mucha evidencia cualitativa de discriminación estructural en la academia,
- ▶ pero el análisis bibliométrico a gran escala tiene el poder de poner un número a aquellas experiencias individuales.
- ▶ Las bases de datos bibliométricas no tienen información sobre la raza<sup>1</sup> autopercibida del autor,
- ▶ Si pudiéramos conocer la raza de los autores, seríamos capaces de desentrañar múltiples dimensiones del racismo estructural en la academia.

## Bases de datos bibliometricas

- ▶ Nuestro objeto de estudio es la discriminación en la academia (En EE. UU.)
- ▶ Hay mucha evidencia cualitativa de discriminación estructural en la academia,
- ▶ pero el análisis bibliométrico a gran escala tiene el poder de poner un número a aquellas experiencias individuales.
- ▶ **Las bases de datos bibliométricas no tienen información sobre la raza<sup>1</sup> autopercebida del autor,**
- ▶ Si pudiéramos conocer la raza de los autores, seríamos capaces de desentrañar múltiples dimensiones del racismo estructural en la academia.

---

<sup>1</sup>Siguiendo la teoría crítica de la raza, utilizamos el concepto de 'raza' como constructo social.

## Bases de datos bibliometricas

- ▶ Nuestro objeto de estudio es la discriminación en la academia (En EE. UU.)
- ▶ Hay mucha evidencia cualitativa de discriminación estructural en la academia,
- ▶ pero el análisis bibliométrico a gran escala tiene el poder de poner un número a aquellas experiencias individuales.
- ▶ Las bases de datos bibliométricas no tienen información sobre la raza<sup>1</sup> autopercebida del autor,
- ▶ Si pudiéramos conocer la raza de los autores, seríamos capaces de desentrañar múltiples dimensiones del racismo estructural en la academia.

---

<sup>1</sup>Siguiendo la teoría crítica de la raza, utilizamos el concepto de 'raza' como constructo social.

# Fuentes de datos

## US Census

- ▶ Datos del censo de EE. UU. de 2010 sobre apellidos y grupos raciales.
- ▶ Todos los nombres con más de 100 apariciones,
- ▶ 162 253 nombres (90% de la población).

## Web of Science

- ▶ primeros autores, basados en EE. UU. de WOS,
- ▶ entre 2008-2019,
- ▶ 5.4M de artículos, 1.6M de primeros autores.

Nos enfocamos en los EE. UU. porque las categorías raciales son una construcción social que solo puede tener significado en un contexto específico.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

### limitaciones

- ▶ Dado que el método se basa en la asociación estadística, las poblaciones pequeñas, o no consideradas en el censo, no pueden ser tenidas en cuenta
- ▶ En nuestro caso, debemos excluir a personas de pueblos originarios y de más de una raza ( 0.69% y 1.76% de los autores),
- ▶ A su vez, sólo se puede inferir el género de forma binaria.
- ▶ La única manera de resolver estas limitaciones es con encuestas.
- ▶ Para ello, es necesario que instituciones, como universidades y editoriales, tomen la tarea.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

### limitaciones

- ▶ Dado que el método se basa en la asociación estadística, las poblaciones pequeñas, o no consideradas en el censo, no pueden ser tenidas en cuenta
- ▶ En nuestro caso, debemos excluir a personas de pueblos originarios y de más de una raza ( 0.69% y 1.76% de los autores),
- ▶ A su vez, sólo se puede inferir el género de forma binaria.
- ▶ La única manera de resolver estas limitaciones es con encuestas.
- ▶ Para ello, es necesario que instituciones, como universidades y editoriales, tomen la tarea.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

### limitaciones

- ▶ Dado que el método se basa en la asociación estadística, las poblaciones pequeñas, o no consideradas en el censo, no pueden ser tenidas en cuenta
- ▶ En nuestro caso, debemos excluir a personas de pueblos originarios y de más de una raza ( 0.69% y 1.76% de los autores),
- ▶ A su vez, sólo se puede inferir el género de forma binaria.
- ▶ La única manera de resolver estas limitaciones es con encuestas.
- ▶ Para ello, es necesario que instituciones, como universidades y editoriales, tomen la tarea.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

### limitaciones

- ▶ Dado que el método se basa en la asociación estadística, las poblaciones pequeñas, o no consideradas en el censo, no pueden ser tenidas en cuenta
- ▶ En nuestro caso, debemos excluir a personas de pueblos originarios y de más de una raza ( 0.69% y 1.76% de los autores),
- ▶ A su vez, sólo se puede inferir el género de forma binaria.
- ▶ La única manera de resolver estas limitaciones es con encuestas.
- ▶ Para ello, es necesario que instituciones, como universidades y editoriales, tomen la tarea.

## Datos & Métodos

- ▶ Usamos los apellidos de los autores para inferir su probable raza, y su nombre de pila para el género.
- ▶ Esto abre la posibilidad a análisis de gran escala, utilizando bases bibliométricas.
- ▶ Pero también es limitado, porque no puede dar cuenta de poblaciones pequeñas, reforzando su invisibilización.

### limitaciones

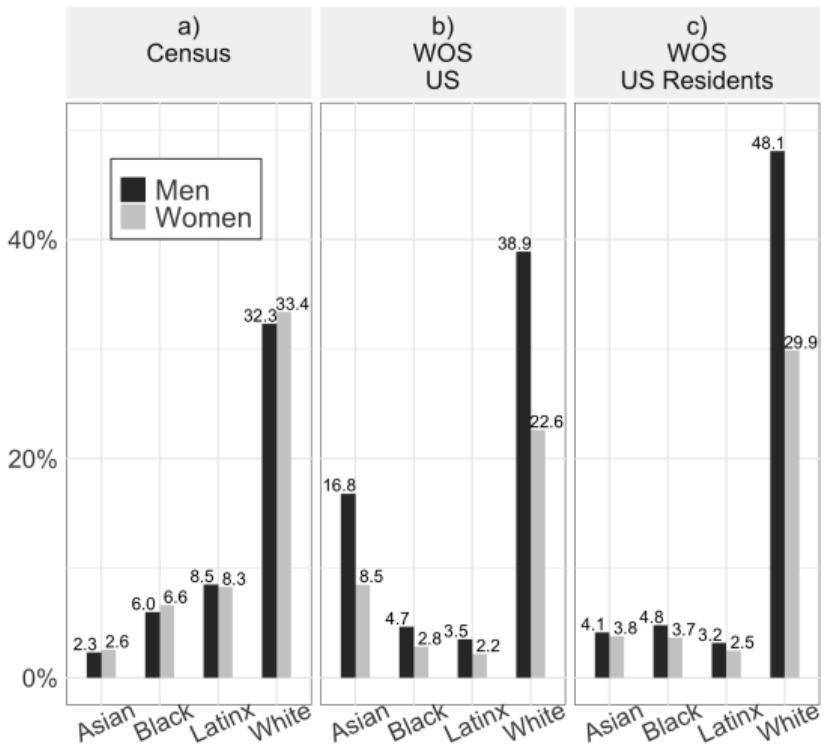
- ▶ Dado que el método se basa en la asociación estadística, las poblaciones pequeñas, o no consideradas en el censo, no pueden ser tenidas en cuenta
- ▶ En nuestro caso, debemos excluir a personas de pueblos originarios y de más de una raza ( 0.69% y 1.76% de los autores),
- ▶ A su vez, sólo se puede inferir el género de forma binaria.
- ▶ La única manera de resolver estas limitaciones es con encuestas.
- ▶ Para ello, es necesario que instituciones, como universidades y editoriales, tomen la tarea.

# Resultados



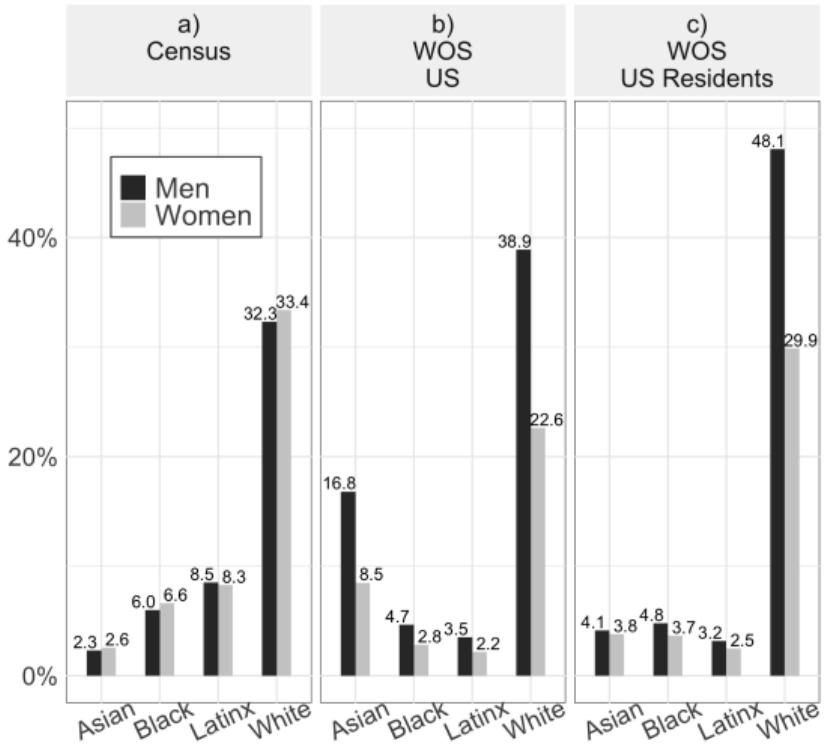
más resultados en [sciencebias.uni.lu/app](http://sciencebias.uni.lu/app)

# distribución general



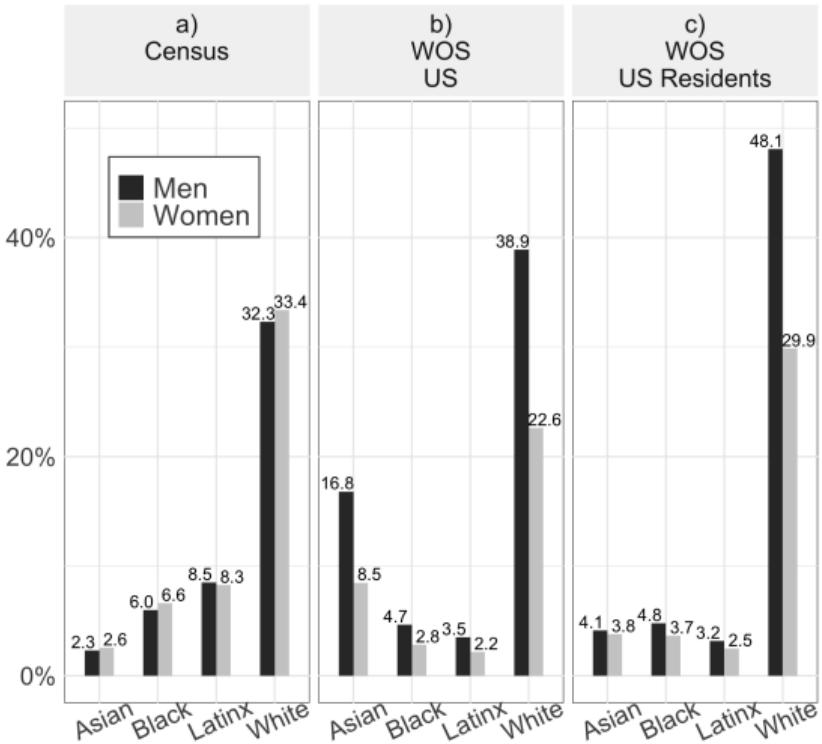
- ▶ En general, hay una representación excesiva de hombres blancos y asiáticos,
- ▶ una gran proporción de autores asiáticos son no residentes de EE. UU.,
- ▶ es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- ▶ **Las mujeres** están subrepresentadas,
- ▶ **Negres y Latinxs** están subrepresentadas,
- ▶ **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# distribución general



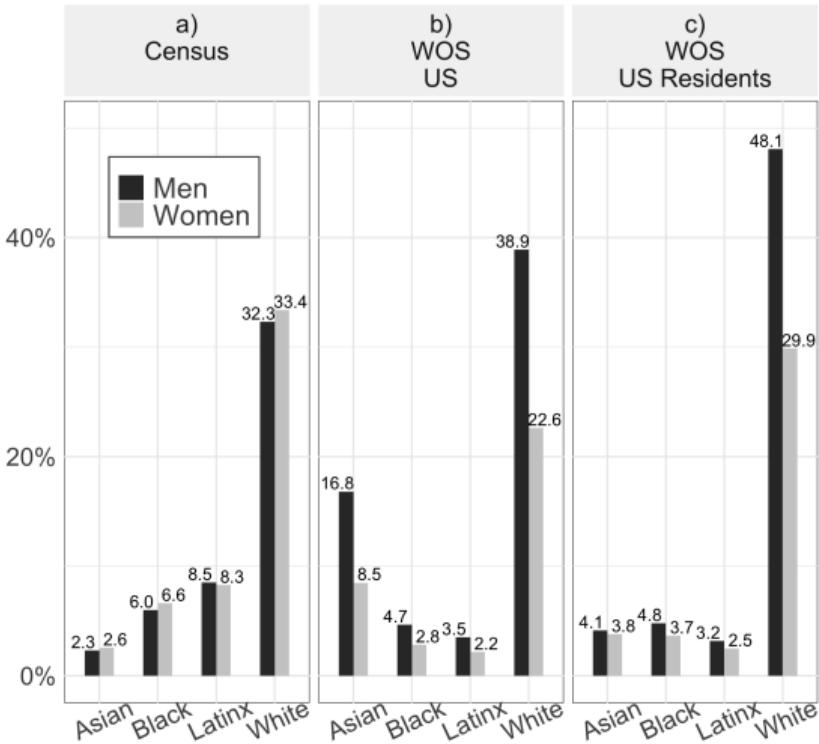
- ▶ En general, hay una representación excesiva de hombres blancos y asiáticos,
- ▶ una gran proporción de autores asiáticos son no residentes de EE. UU.,
- ▶ es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- ▶ **Las mujeres** están subrepresentadas,
- ▶ **Negres y Latinxs** están subrepresentadas,
- ▶ **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# distribución general



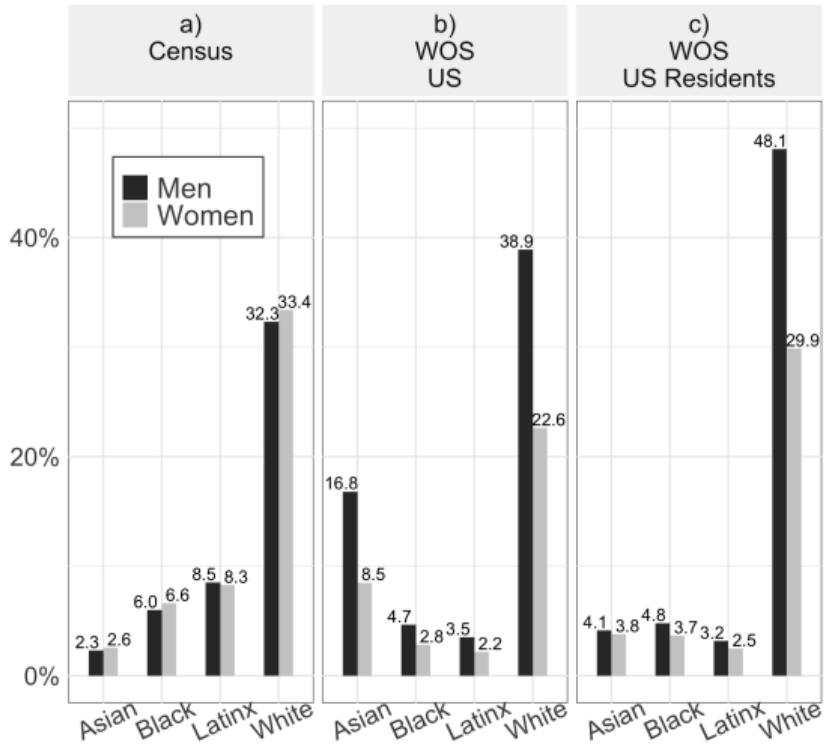
- ▶ En general, hay una representación excesiva de hombres blancos y asiáticos,
- ▶ una gran proporción de autores asiáticos son no residentes de EE. UU.,
- ▶ es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- ▶ **Las mujeres** están subrepresentadas,
- ▶ **Negres y Latinxs** están subrepresentadas,
- ▶ **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# distribución general



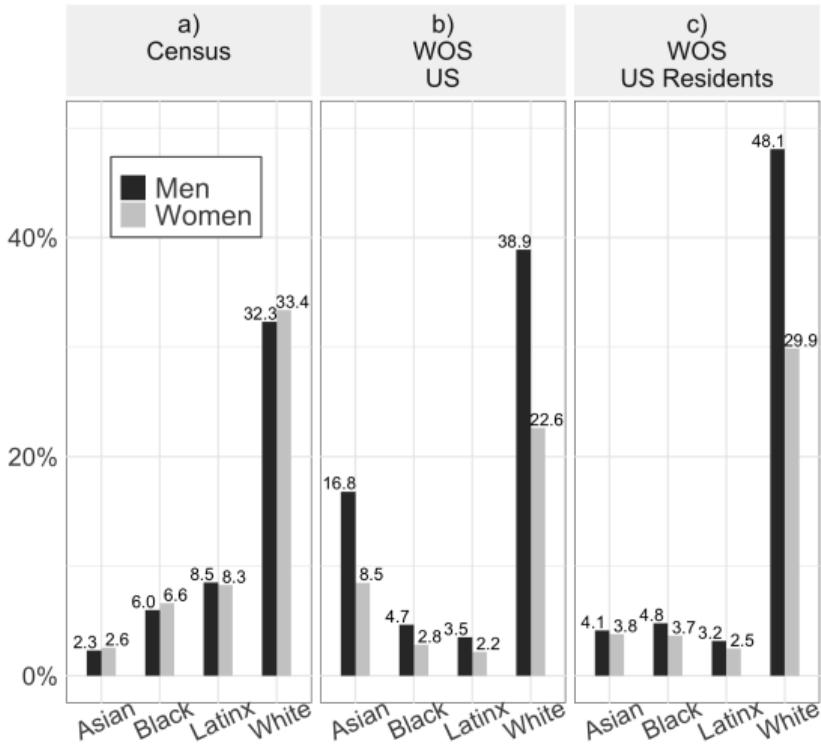
- En general, hay una representación excesiva de hombres blancos y asiáticos,
- una gran proporción de autores asiáticos son no residentes de EE. UU.,
- es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- **Las mujeres** están subrepresentadas,
- **Negres y Latinxs** están subrepresentadas,
- **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# distribución general



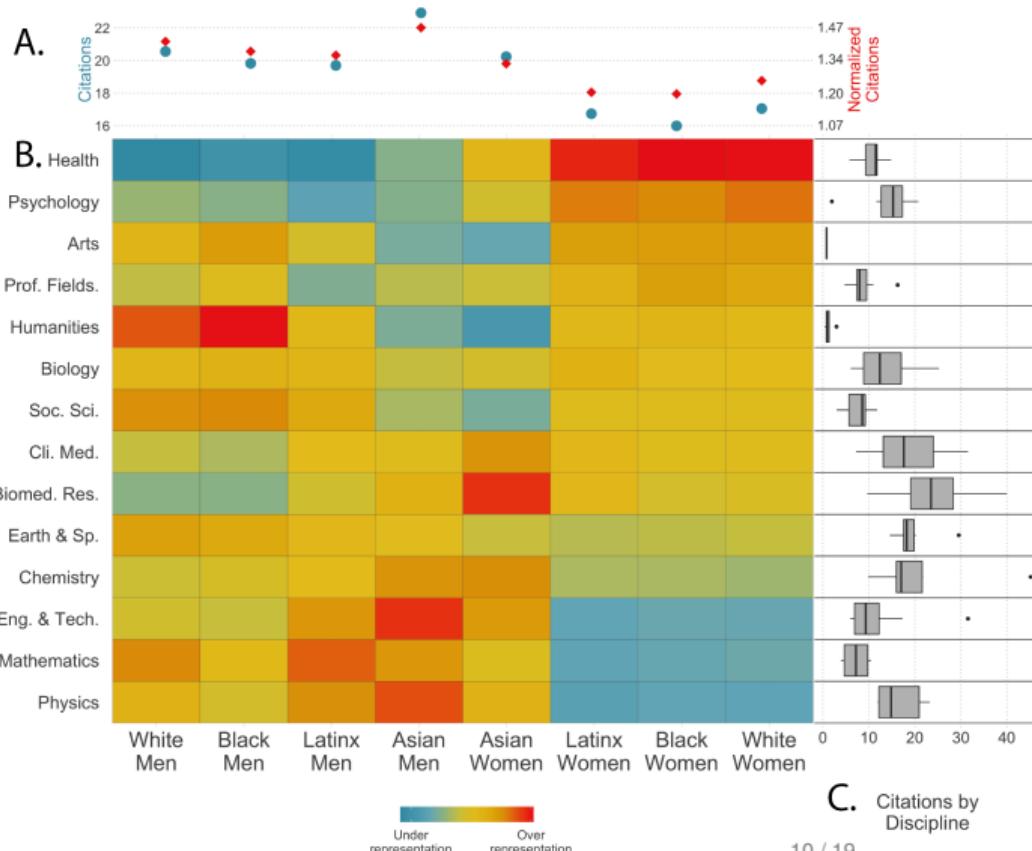
- ▶ En general, hay una representación excesiva de hombres blancos y asiáticos,
- ▶ una gran proporción de autores asiáticos son no residentes de EE. UU.,
- ▶ es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- ▶ **Las mujeres** están subrepresentadas,
- ▶ **Negres y Latinxs** están subrepresentadas,
- ▶ **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# distribución general



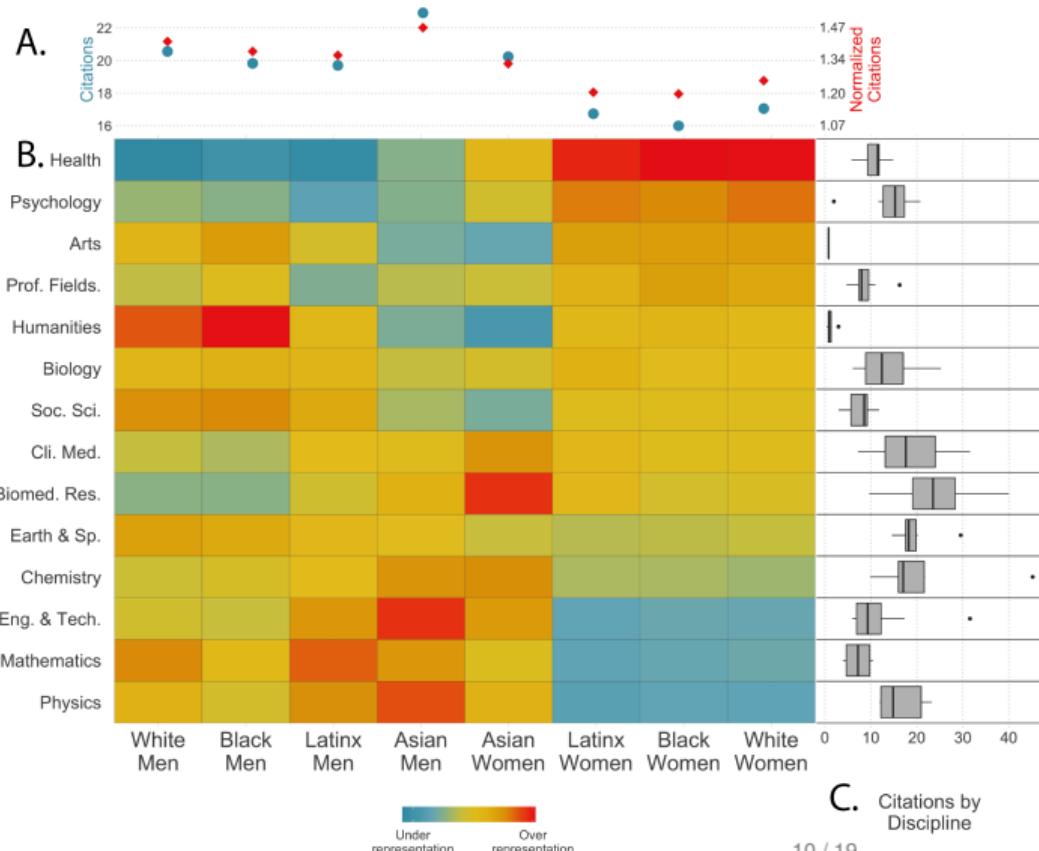
- ▶ En general, hay una representación excesiva de hombres blancos y asiáticos,
- ▶ una gran proporción de autores asiáticos son no residentes de EE. UU.,
- ▶ es decir, el Censo no es una buena referencia para definir *sobrerepresentación* de este grupo.
- ▶ **Las mujeres** están subrepresentadas,
- ▶ **Negres y Latinxs** están subrepresentadas,
- ▶ **Mujeres Negras y Latinas** son los grupos más subrepresentados.

# Heterogeneidad de las disciplinas



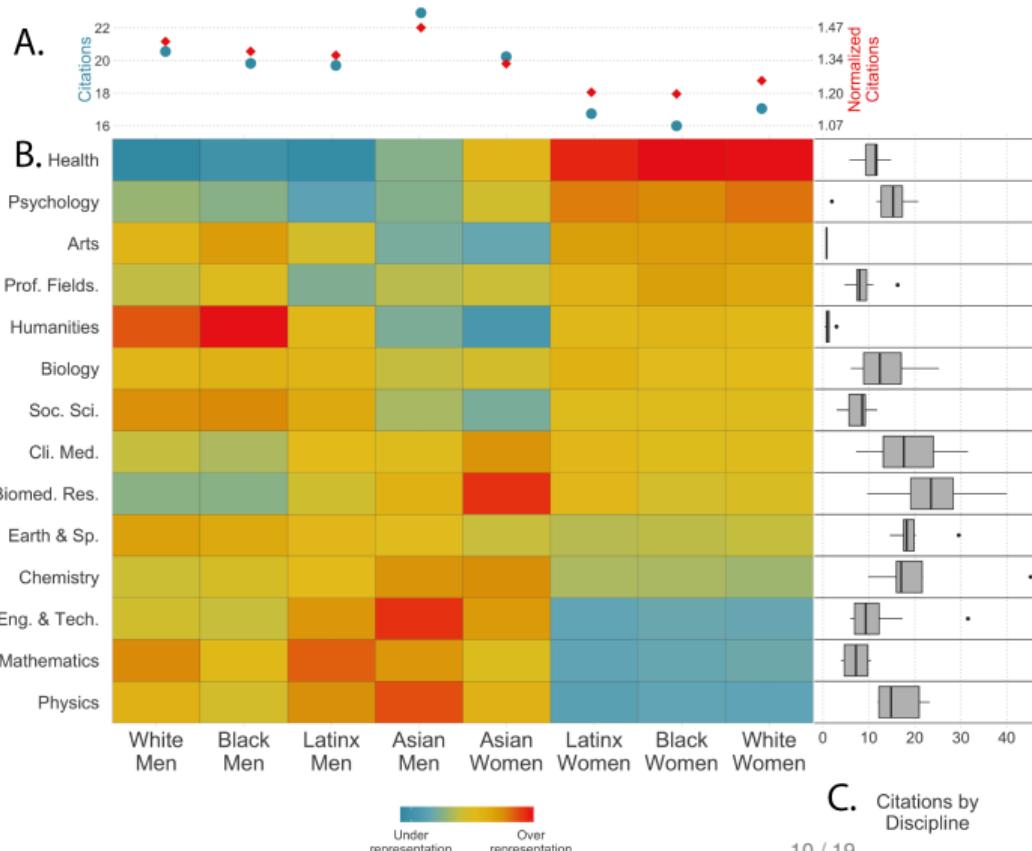
- ▶ La distribución por disciplinas sigue un patrón claro por género, excepto en los autores asiáticos.
- ▶ Las mujeres tienen en promedio menos citas,
- ▶ la normalización por disciplina en las citas reduce la brecha, pero no completamente.

# Heterogeneidad de las disciplinas



- ▶ La distribución por disciplinas sigue un patrón claro por género, excepto en los autores asiáticos.
- ▶ Las mujeres tienen en promedio menos citas,
- ▶ la normalización por disciplina en las citas reduce la brecha, pero no completamente.

# Heterogeneidad de las disciplinas



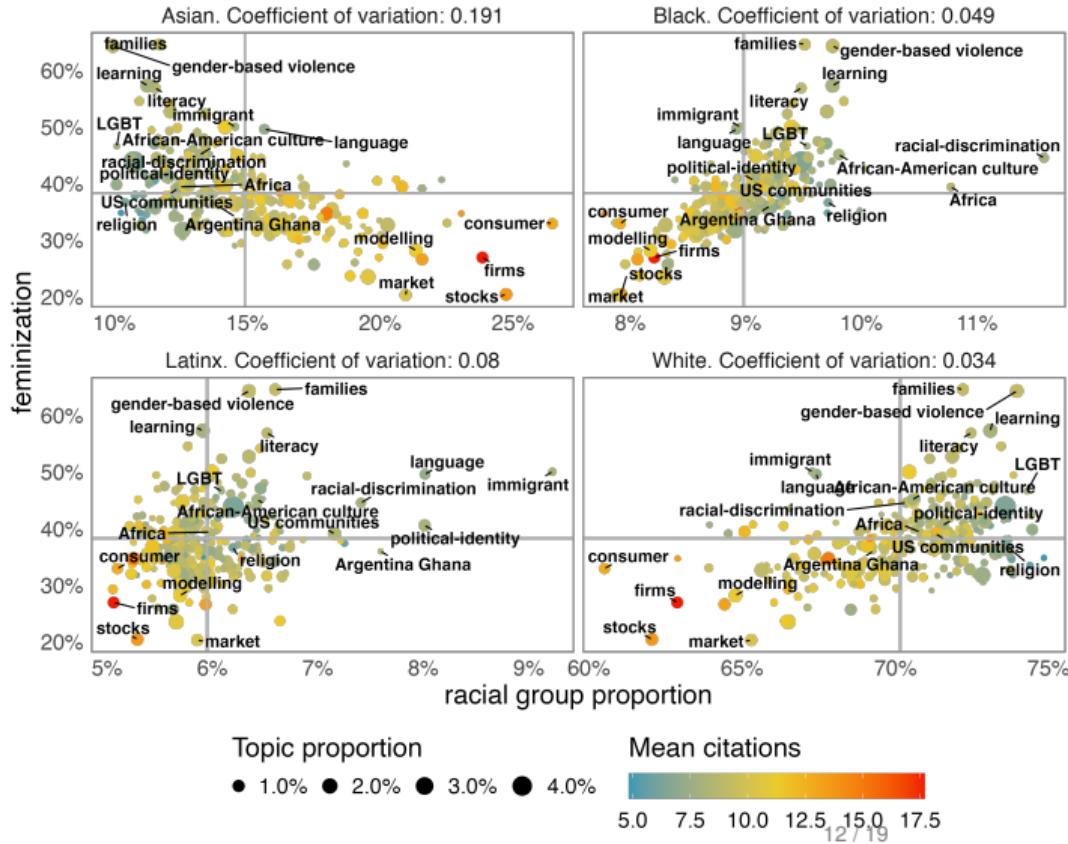
- ▶ La distribución por disciplinas sigue un patrón claro por género, excepto en los autores asiáticos.
- ▶ Las mujeres tienen en promedio menos citas,
- ▶ la normalización por disciplina en las citas reduce la brecha, pero no completamente.

- ▶ Queremos profundizar en la relación de los temas de investigación con la raza y el género,
- ▶ por eso nos enfocamos en Ciencias Sociales, y definimos temas de investigación usando LDA (Blei, Ng, and Jordan 2003).
- ▶ Para cada raza y género definimos la participación promedio en cada tema de investigación.

- ▶ Queremos profundizar en la relación de los temas de investigación con la raza y el género,
- ▶ por eso nos enfocamos en Ciencias Sociales, y definimos temas de investigación usando LDA (Blei, Ng, and Jordan 2003).
- ▶ Para cada raza y género definimos la participación promedio en cada tema de investigación.

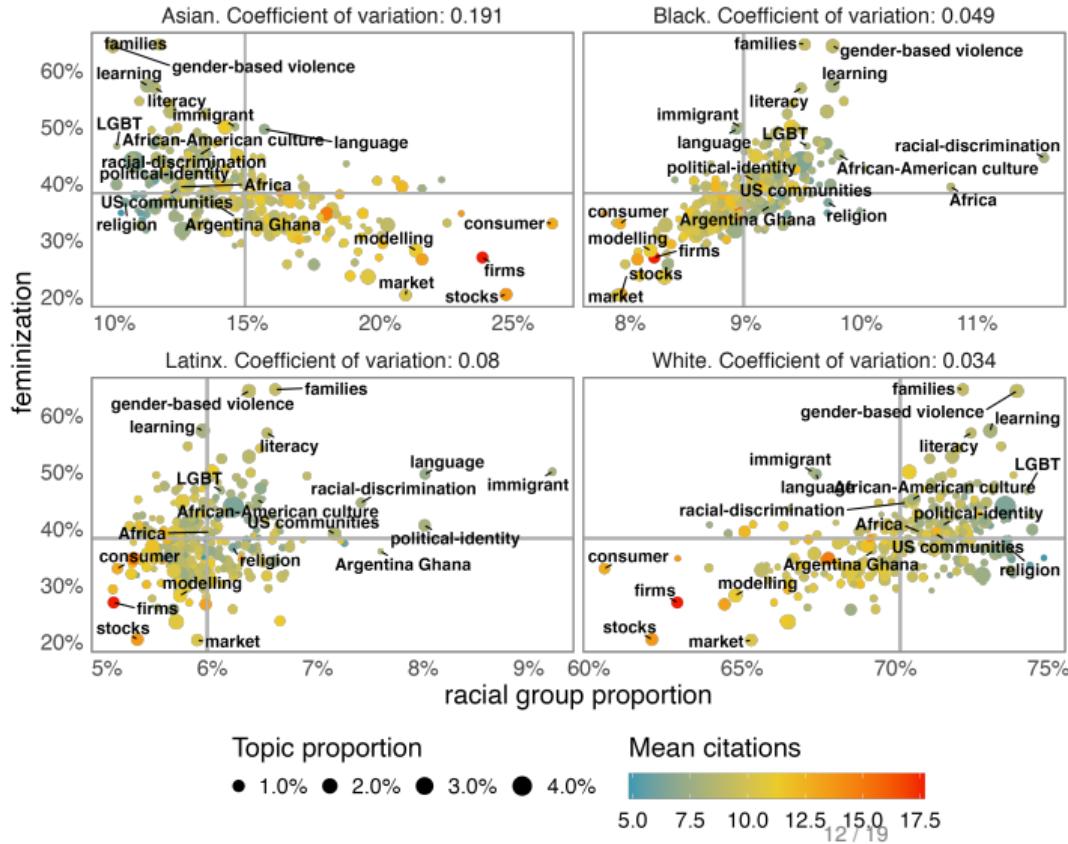
- ▶ Queremos profundizar en la relación de los temas de investigación con la raza y el género,
- ▶ por eso nos enfocamos en Ciencias Sociales, y definimos temas de investigación usando LDA (Blei, Ng, and Jordan 2003).
- ▶ Para cada raza y género definimos la participación promedio en cada tema de investigación.

# Líneas de investigación



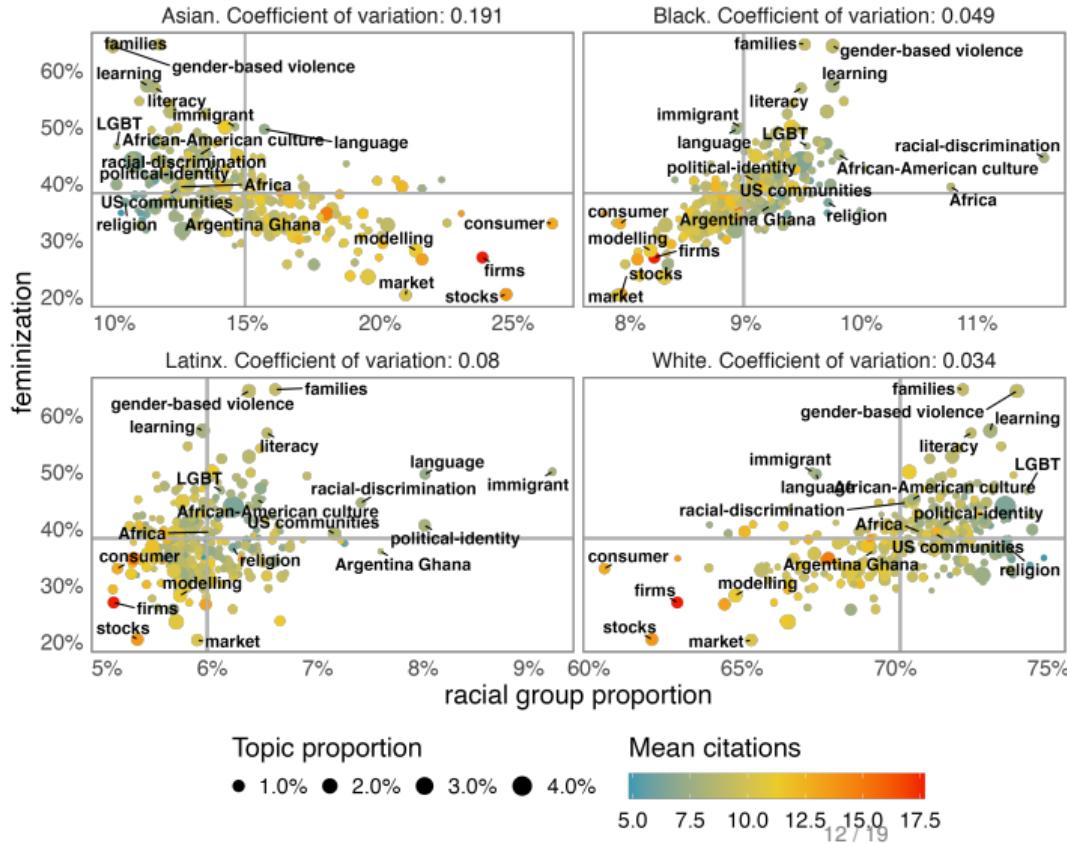
- ▶ Las mujeres publican más sobre *familias, violencia de género y aprendizaje*,
- ▶ Les autores negros se enfocan en *discriminación racial, cultura afroamericana y África*,
- ▶ Les autores latinx se enfocan en *migración, lenguaje e identidad política*,
- ▶ Los autores asiáticos se enfocan en *economía*, mientras que los autores blancos se distribuyen más uniformemente en todos los temas.

# Líneas de investigación



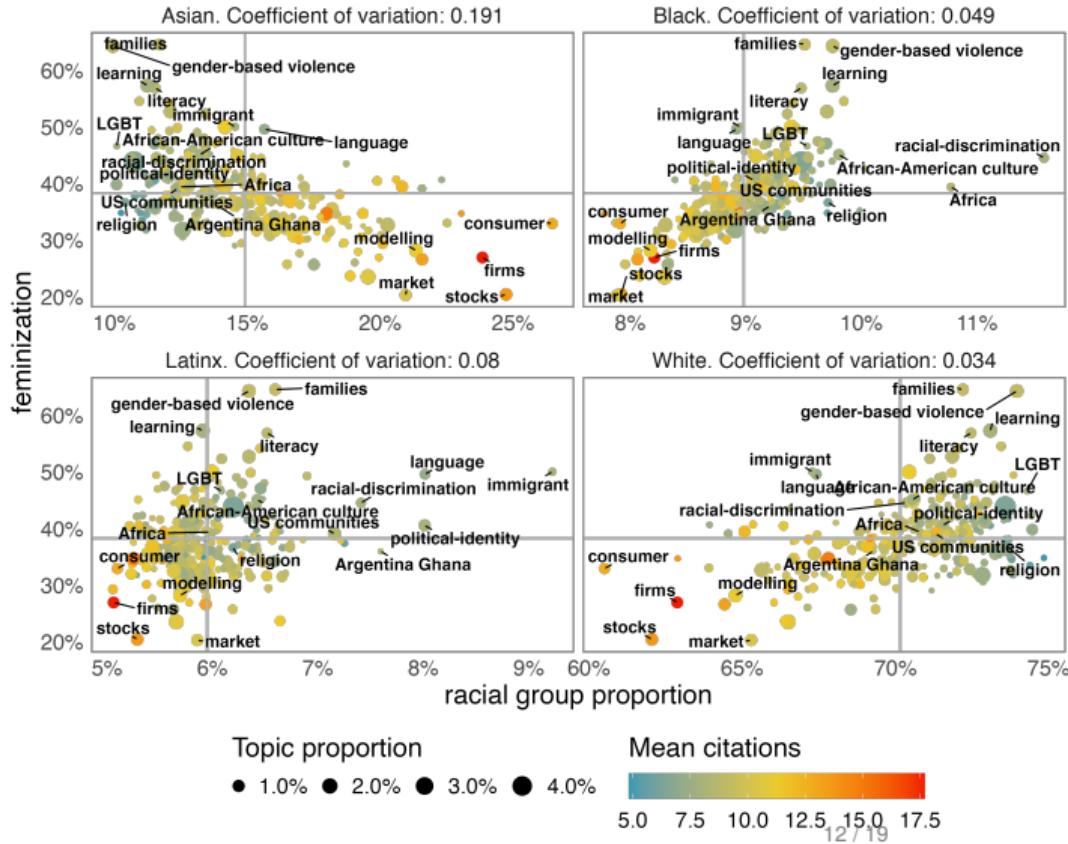
- ▶ Las mujeres publican más sobre *familias, violencia de género y aprendizaje*,
- ▶ Les autores negros se enfocan en *discriminación racial, cultura afroamericana y África*,
- ▶ Les autores latinx se enfocan en *migración, lenguaje e identidad política*,
- ▶ Les autores asiáticos se enfocan en *economía*, mientras que los autores blancos se distribuyen más uniformemente en todos los temas.

# Líneas de investigación



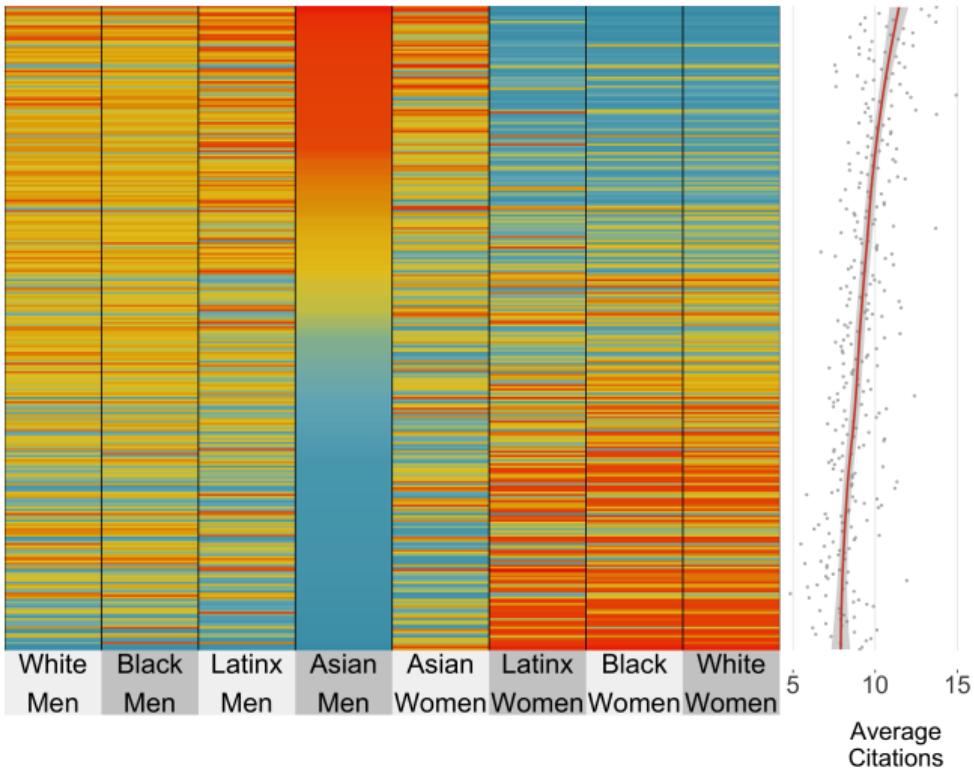
- ▶ Las mujeres publican más sobre *familias, violencia de género y aprendizaje*,
- ▶ Les autores negros se enfocan en *discriminación racial, cultura afroamericana y África*,
- ▶ Les autores latinx se enfocan en *migración, lenguaje e identidad política*,
- ▶ Les autores asiáticos se enfocan en *economía*, mientras que los autores blancos se distribuyen más uniformemente en todos los temas.

# Líneas de investigación



- ▶ Las mujeres publican más sobre *familias, violencia de género y aprendizaje*,
- ▶ Les autores negros se enfocan en *discriminación racial, cultura afroamericana y África*,
- ▶ Les autores latinx se enfocan en *migración, lenguaje e identidad política*,
- ▶ Los autores asiáticos se enfocan en *economía*, mientras que los autores blancos se distribuyen más uniformemente en todos los temas.

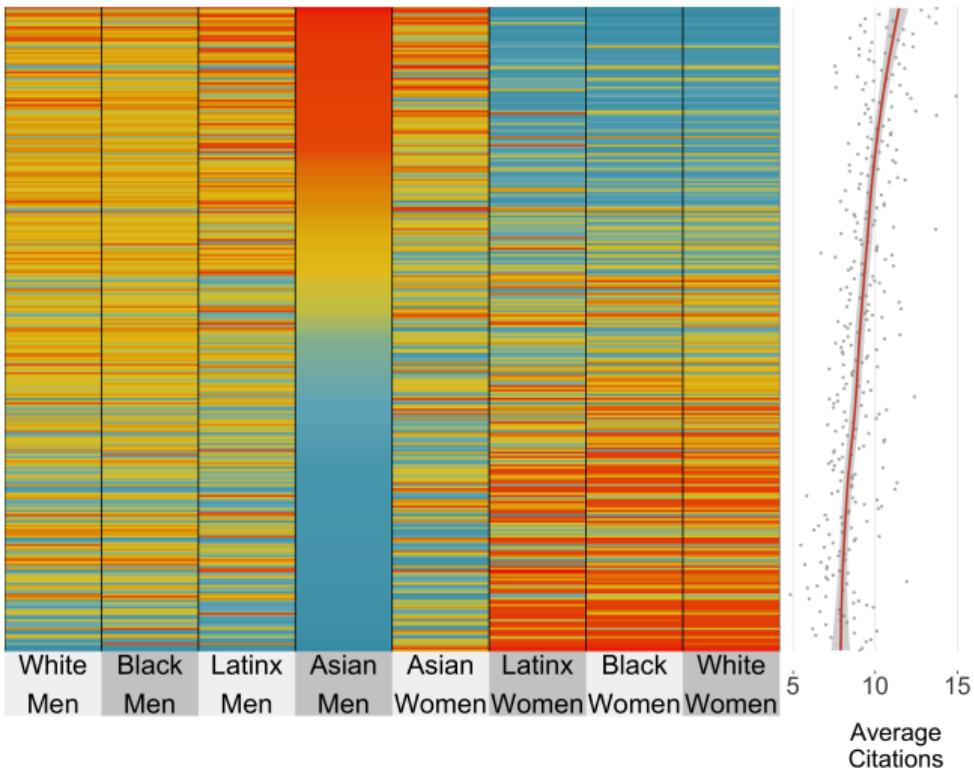
# Tópicos y citas



¿Cómo afectan los tópicos de investigación a las brechas de citas?

- ▶ Si ordenamos los tópicos por la participación de hombres asiáticos, correlaciona positivamente con el número promedio de citas por tópico.
- ▶ Esto significa que los hombres asiáticos tienden a investigar sobre temas más citados.
- ▶ También podemos ver patrones de género en los tópicos de investigación.

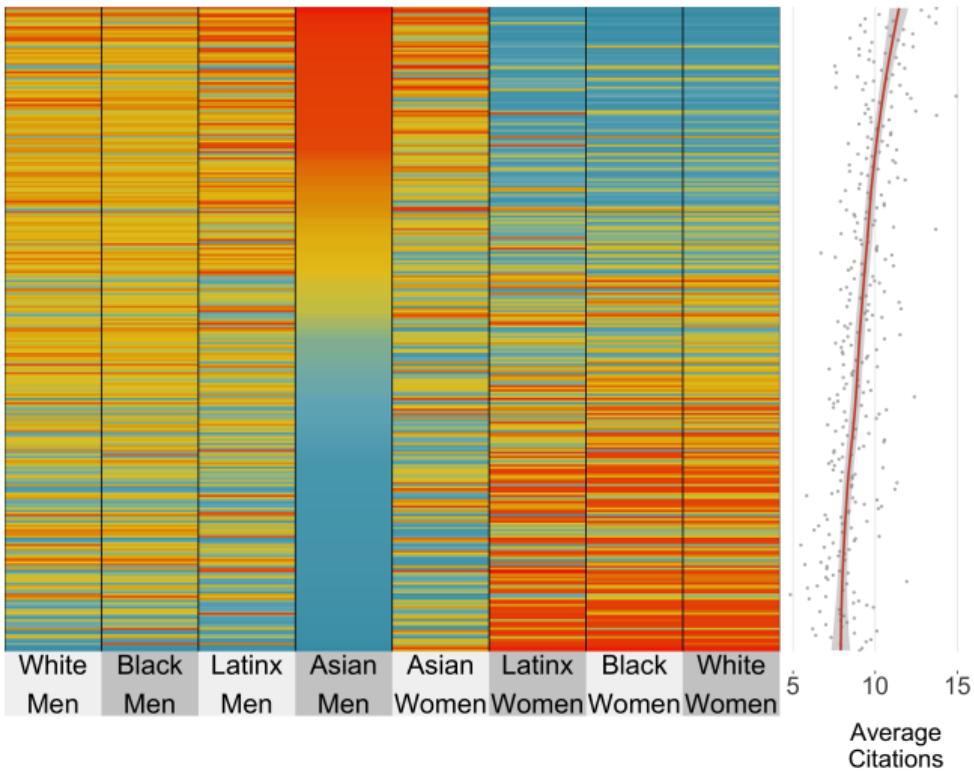
# Tópicos y citas



¿Cómo afectan los tópicos de investigación a las brechas de citas?

- ▶ Si ordenamos los tópicos por la participación de hombres asiáticos, correlaciona positivamente con el número promedio de citas por tópico.
- ▶ Esto significa que los hombres asiáticos tienden a investigar sobre temas más citados.
- ▶ También podemos ver patrones de género en los tópicos de investigación.

# Tópicos y citas

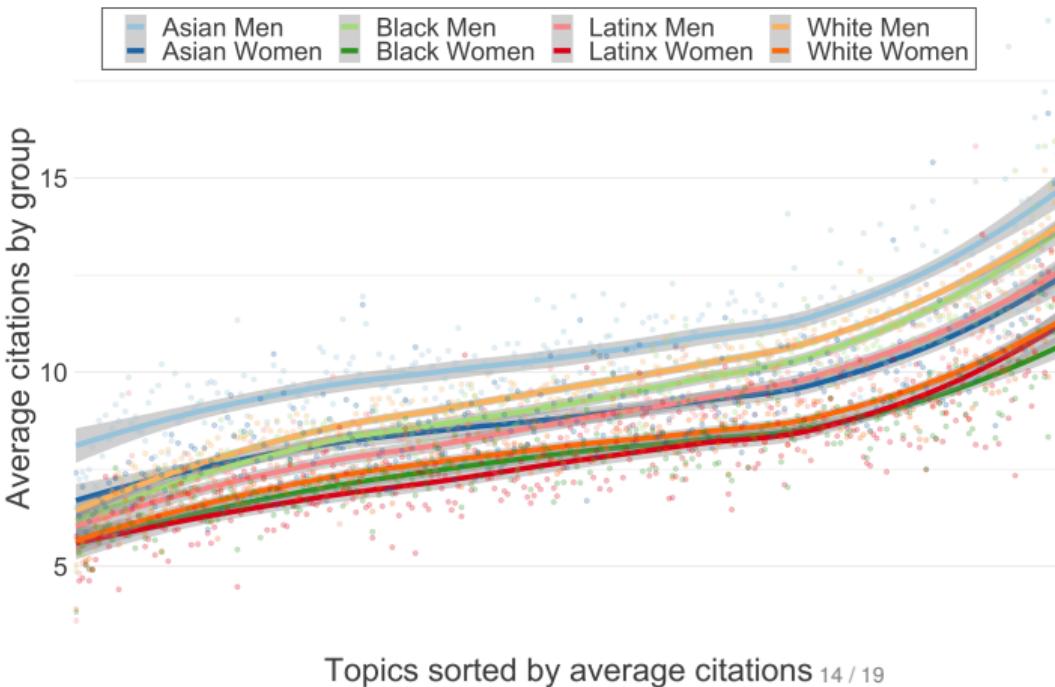


¿Cómo afectan los tópicos de investigación a las brechas de citas?

- ▶ Si ordenamos los tópicos por la participación de hombres asiáticos, correlaciona positivamente con el número promedio de citas por tópico.
- ▶ Esto significa que los hombres asiáticos tienden a investigar sobre temas más citados.
- ▶ También podemos ver patrones de género en los tópicos de investigación.

# Tópicos y citas

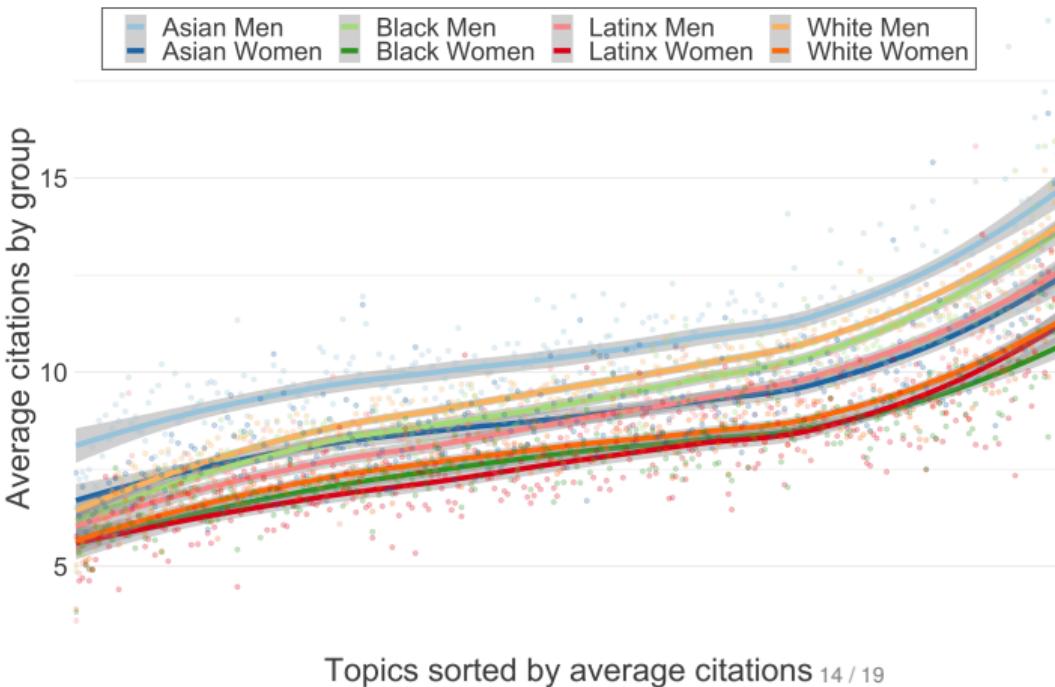
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

# Tópicos y citas

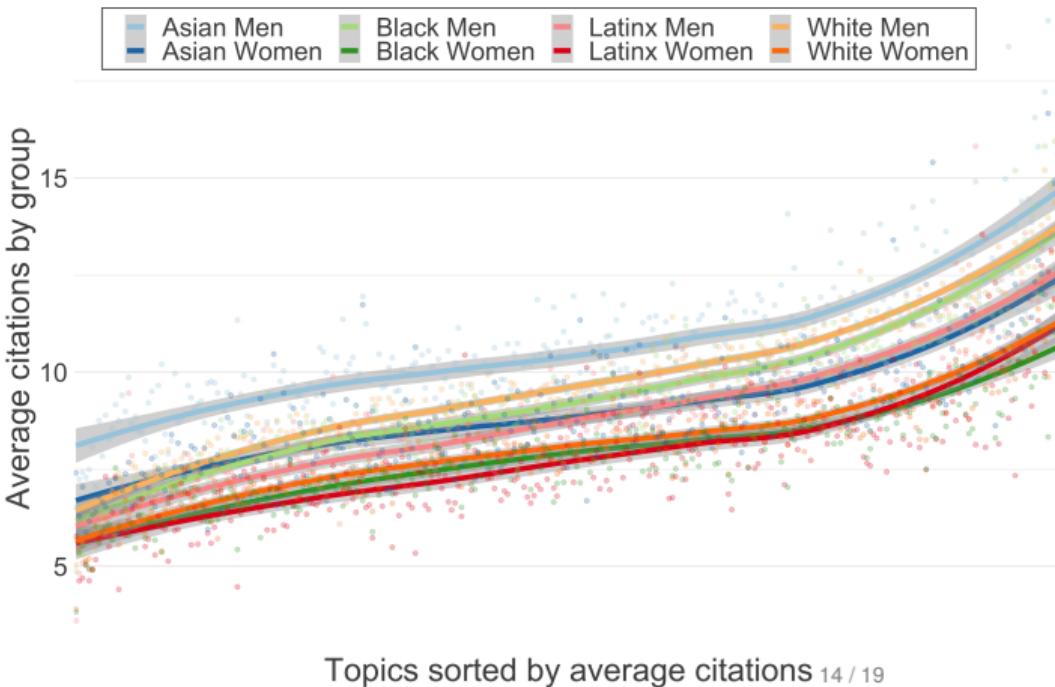
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

# Tópicos y citas

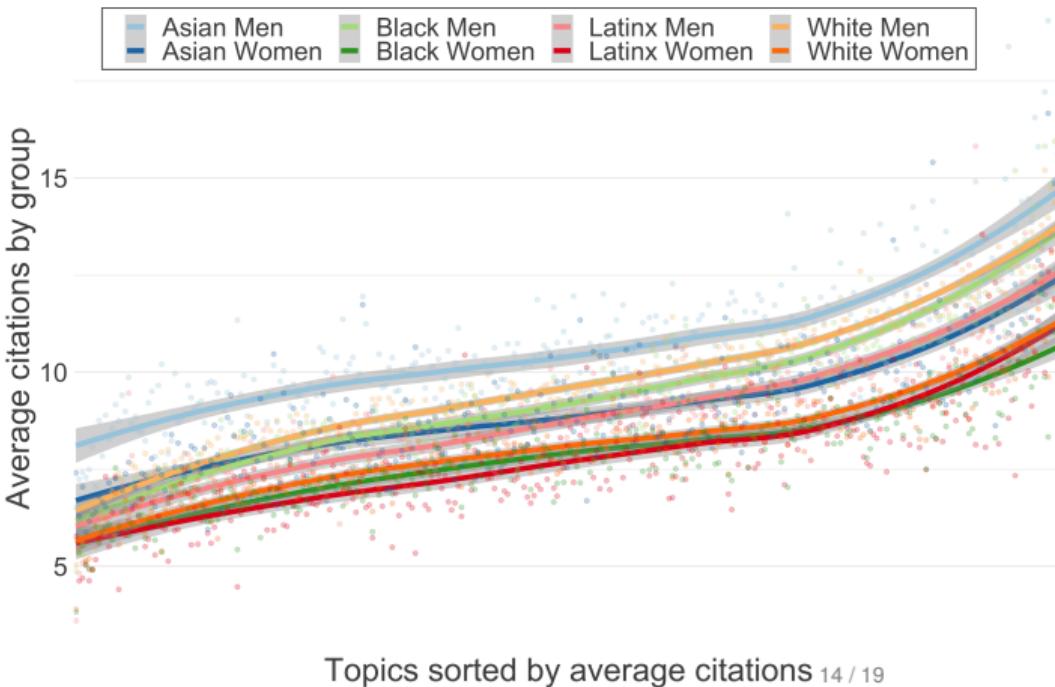
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

# Tópicos y citas

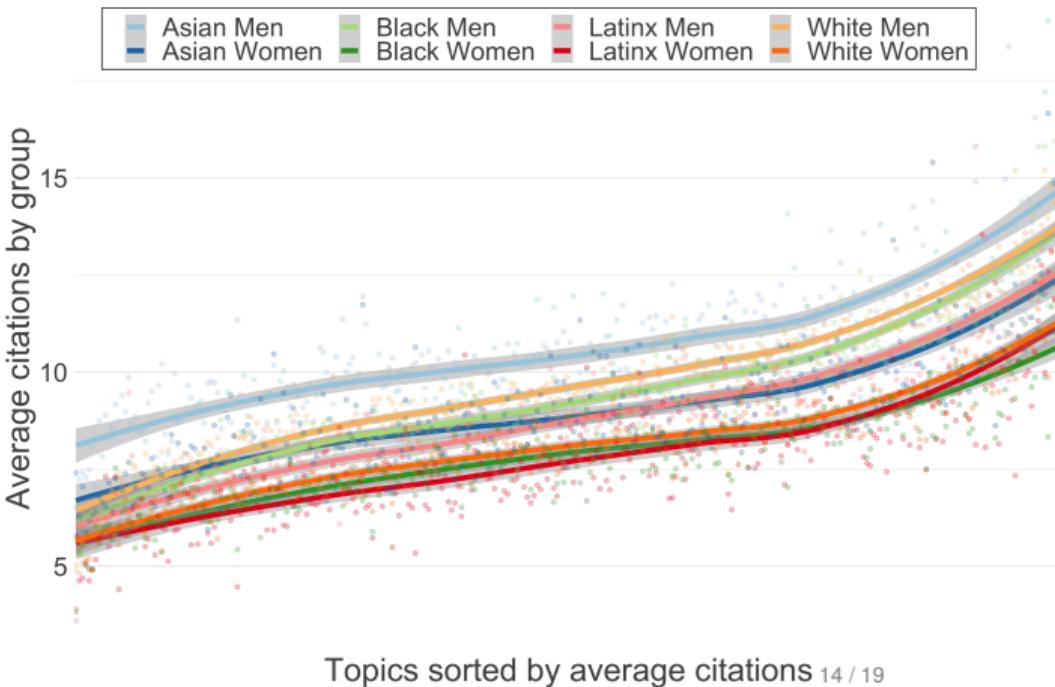
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

# Tópicos y citas

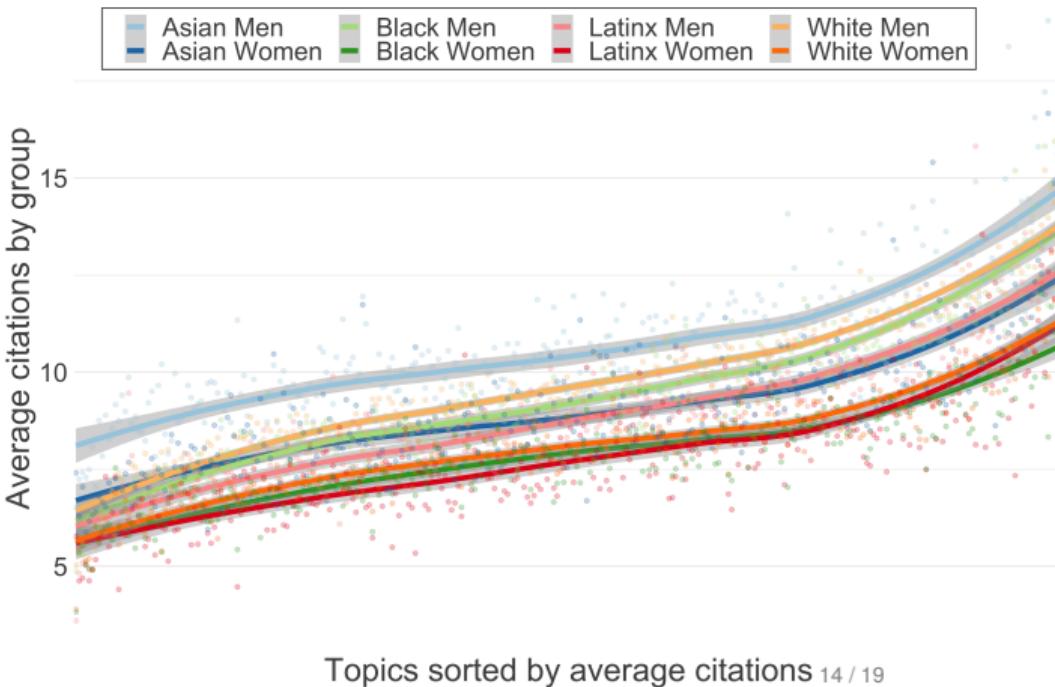
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

# Tópicos y citas

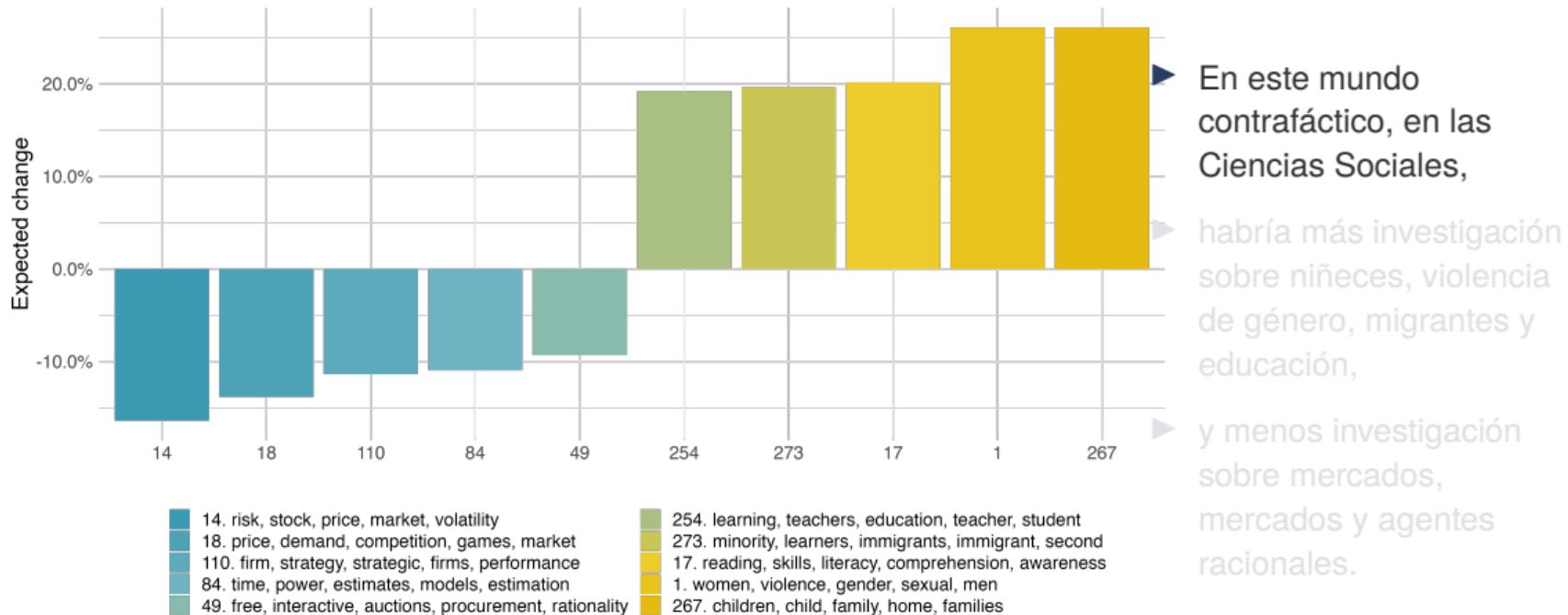
¿Las brechas de citas se explican por la distribución en los temas de investigación?



- ▶ Si ordenamos los temas por sus citas promedio y modelamos la distribución de citas por grupos, podemos ver que
- ▶ Los hombres asiáticos tienen más citas en todos los temas,
- ▶ luego los hombres blancos y negros
- ▶ seguidos por hombres Latinx y mujeres asiáticas,
- ▶ Las mujeres blancas, negras y latinas tienen menos citas en todos los temas.
- ▶ Hay un sesgo inter e intra

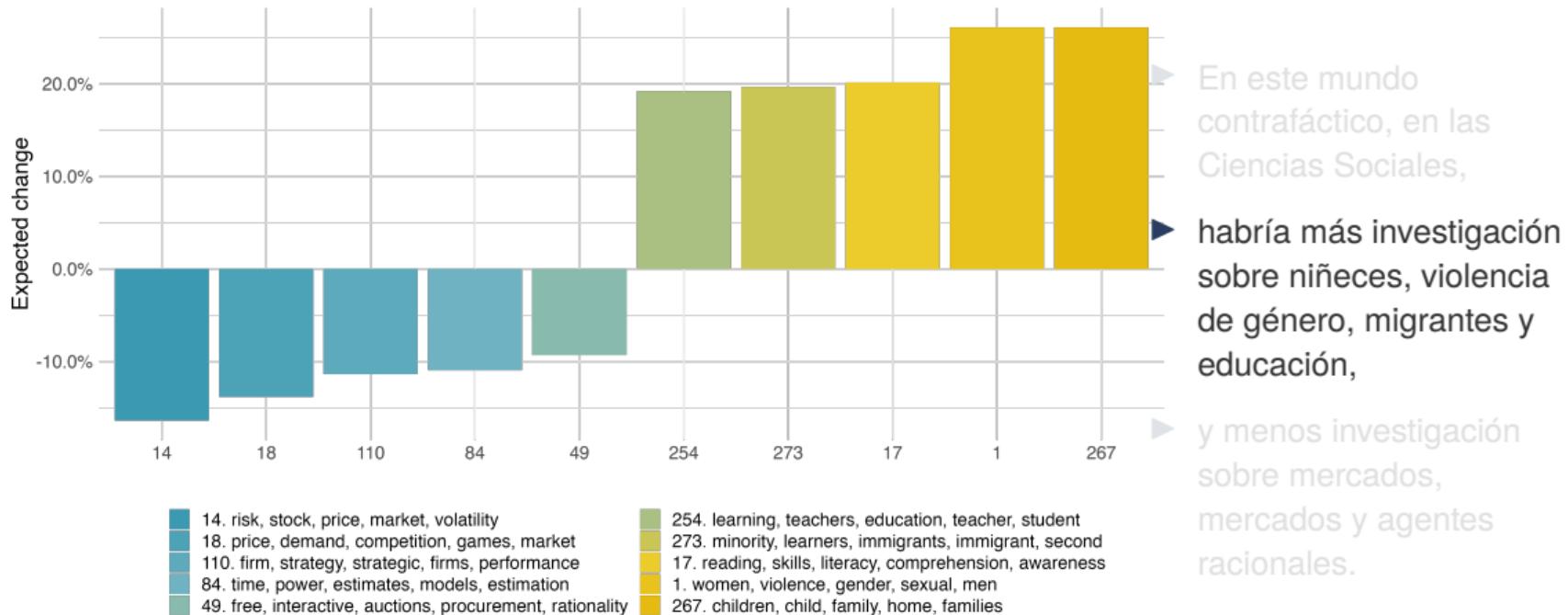
# Análisis contrafáctico

¿Qué pasaría si la proporción de autores por raza y género fuera equivalente al Censo de EE. UU. de 2010?



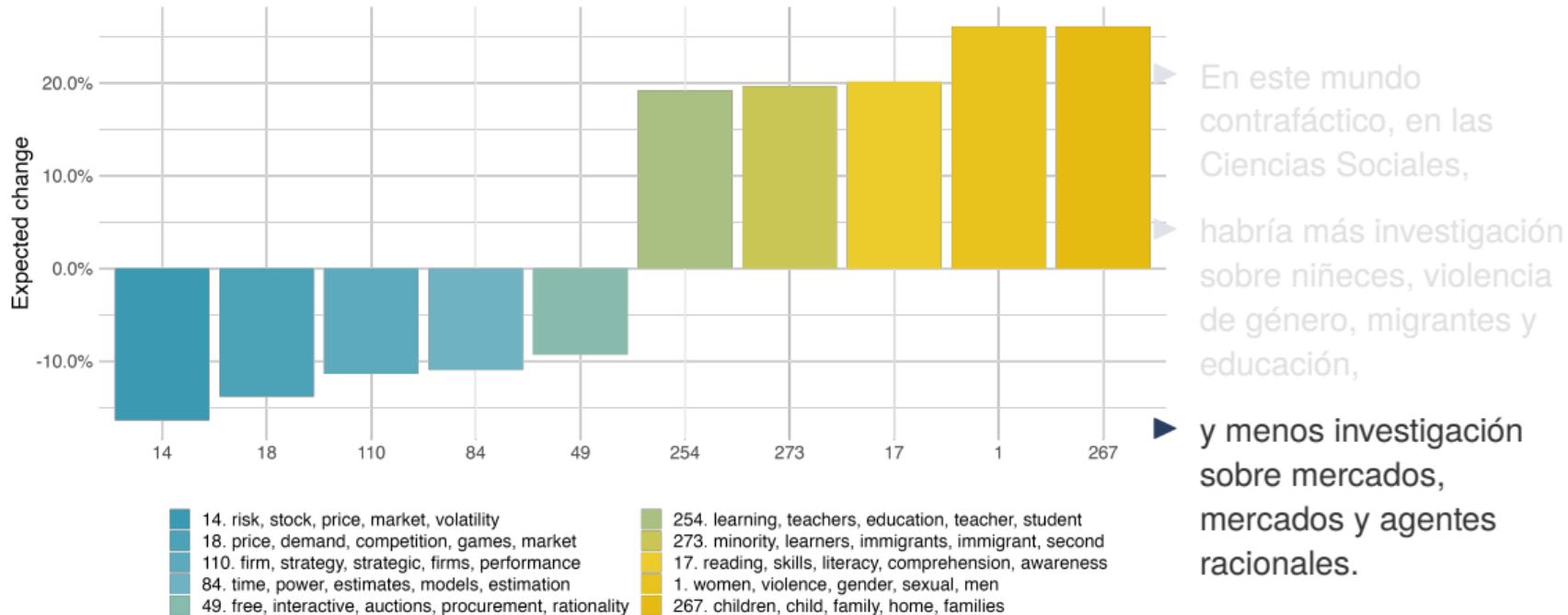
# Análisis contrafáctico

¿Qué pasaría si la proporción de autores por raza y género fuera equivalente al Censo de EE. UU. de 2010?



# Análisis contrafáctico

¿Qué pasaría si la proporción de autores por raza y género fuera equivalente al Censo de EE. UU. de 2010?



# Conclusiones

- ▶ Hay una **subrepresentación** de grupos marginalizados (en la intersección de raza y género),
- ▶ estos grupos tienen intereses de investigación específicos, lo que crea una **falta de investigación de temas poco estudiados**.
- ▶ También hay una **brecha de citas** debido a la distribución de temas de investigación, y al sesgo dentro de cada tema.

# Conclusiones

- ▶ Hay una **subrepresentación** de grupos marginalizados (en la intersección de raza y género),
- ▶ estos grupos tienen intereses de investigación específicos, lo que crea una **falta de investigación de temas poco estudiados**.
- ▶ También hay una **brecha de citas** debido a la distribución de temas de investigación, y al sesgo dentro de cada tema.

# Conclusiones

- ▶ Hay una **subrepresentación** de grupos marginalizados (en la intersección de raza y género),
- ▶ estos grupos tienen intereses de investigación específicos, lo que crea una **falta de investigación de temas poco estudiados**.
- ▶ También hay una **brecha de citas** debido a la distribución de temas de investigación, y al sesgo dentro de cada tema.

# Gracias!

 @Diego\_Koz

 diego.kozlowski@uni.lu

 sciencebias.uni.lu/app

## Acknowledgement

The Doctoral Training Unit **Data-driven computational modelling and applications** (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

<https://driven.uni.lu>



# Bibliography

- [1] Tukufu Zuberi. "Thicker Than Blood: How Racial Statistics Lie". In: 31 (2002), p. 529. issn: 0094-3061. doi: 10.2307/3090025.
- [2] Catherine D'Ignazio and Lauren Klein. "4. "What Gets Counted Counts"". In: *Data Feminism*. <https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp>. Mar. 16, 2020. URL: <https://data-feminism.mitpress.mit.edu/pub/h1w0nbqp>.
- [3] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichlet allocation". In: *the Journal of machine Learning research* 3 (2003), pp. 993–1022.