

# What Data Science Can Learn from Sociology

Stephanie Kirmer  
satRdays Chicago  
April 27, 2019



@data\_stephanie | [www.stephaniekirmer.com](http://www.stephaniekirmer.com)

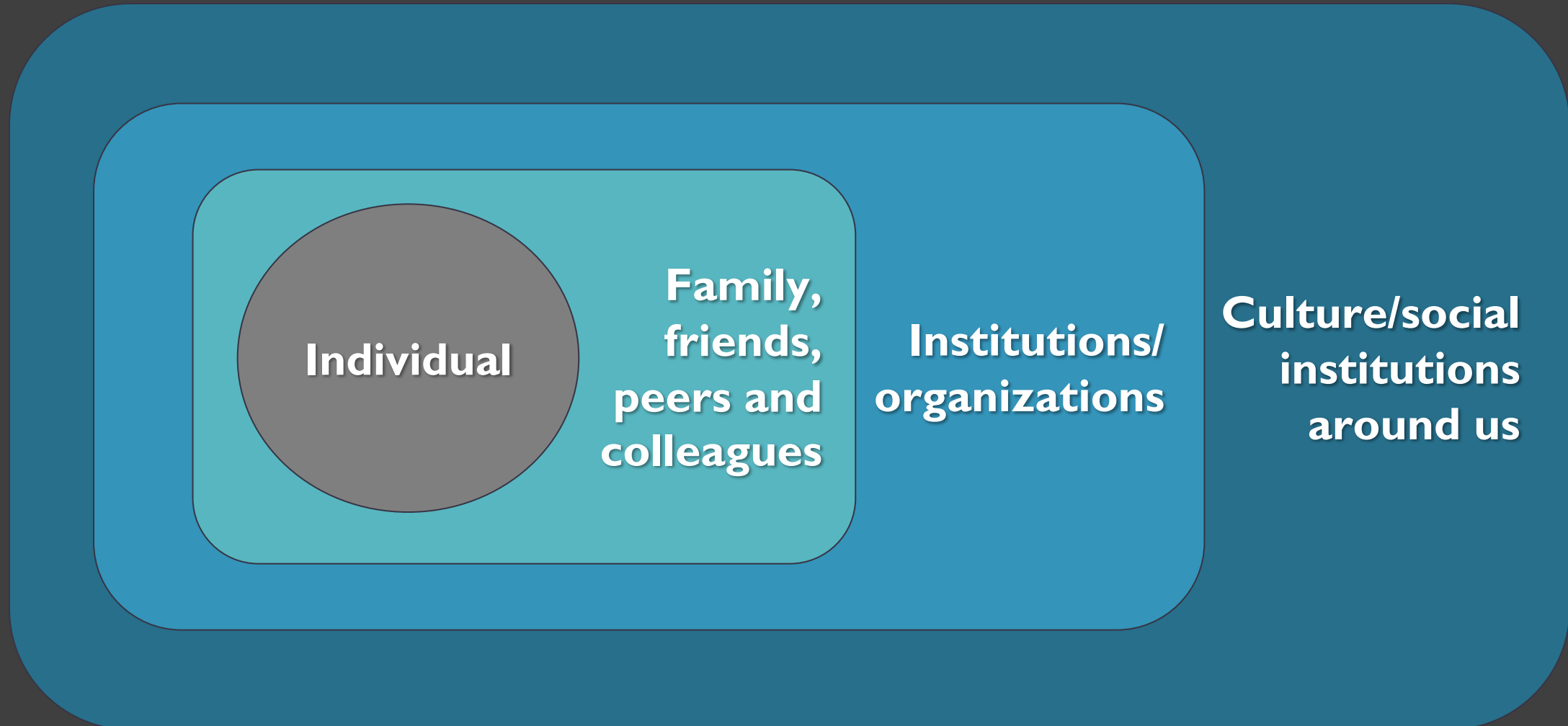
# Sociology? What?

We're at a data conference, right?

# What Sociology Studies

- Race, class, gender, and other forms of social stratification
- Why people and institutions behave the way they do
- Real life experiences of different kinds of people in society
- Power dynamics between people

# Structures



# Sociology of Work/Professions

- Stratification of job types
- Perceived prestige of work
- Explicit or implicit rules or expectations
- Social norms within workplaces
- How workplaces and jobs intersect with other parts of our lives

# Key Terms

**Privilege:** **unearned benefits** available to particular groups based on their location within a social hierarchy.

**Power:** An entity or individual's ability to **control, influence, or direct others**

**Cultural capital:** the combined **social assets of a person** (education, style of speech, etc.) that promote social mobility in a stratified society.

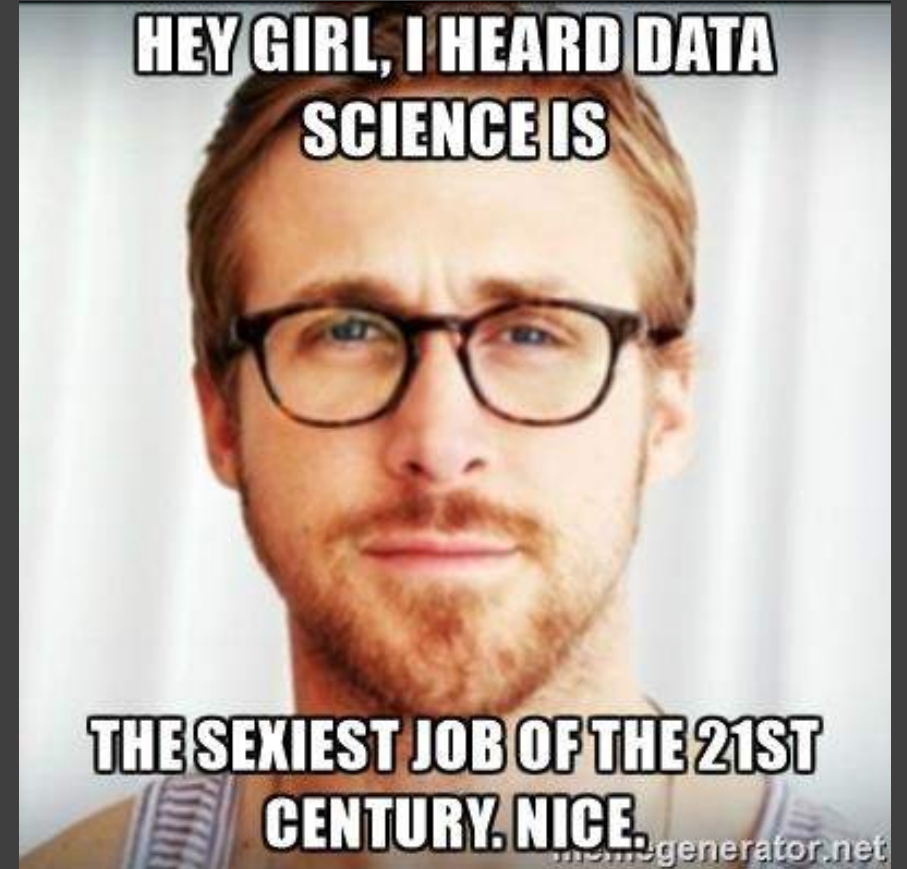


# So, About Data Science...

Getting back to today's event!

# Data Science

- Relatively new field
- Prestigious, well compensated, “sexy”
- Visible inequalities in all kinds of areas





# Areas Where Sociology Can Help

- Who you hire
- Who you boost
- What projects you do
- What analytic questions you ask
- What data you use (or don't use)
- How you structure your organization
- Priorities you set
- ... among others

# LESSON 1.

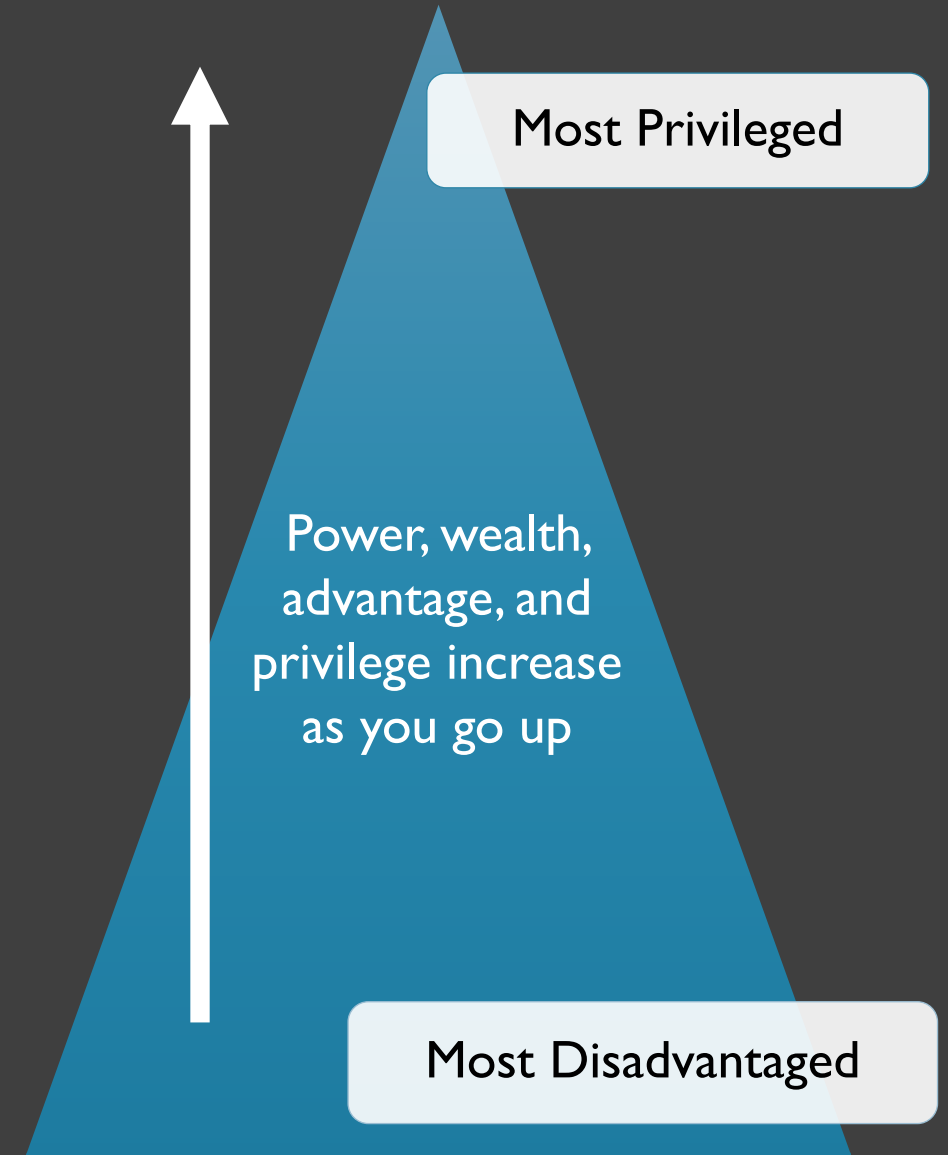
Power and inequality are  
everywhere.

# Stratification

Classification frameworks:

- Are everywhere
- Are conscious and subconscious
- Come with social status (for good or bad)

No one gets to opt out of social stratification.



# Data Science Example

## Professional Privilege

- Part of social status comes from our job/profession
- Whiter, more male professions (like ours) = higher social status
- Gaining prestige status affects the profession, for better or worse

“In every job category, higher median base salaries continue to be seen among data scientists when compared to other predictive analytics professionals.”  
- Burtch Works 2018

“[B]ecause of the prestige of the data scientist job title, companies like Lyft will hire for data science job titles, but with data analyst skillsets.”  
– Vicki Boykis, 2019

LESSON 1.

Power and inequality are  
everywhere.

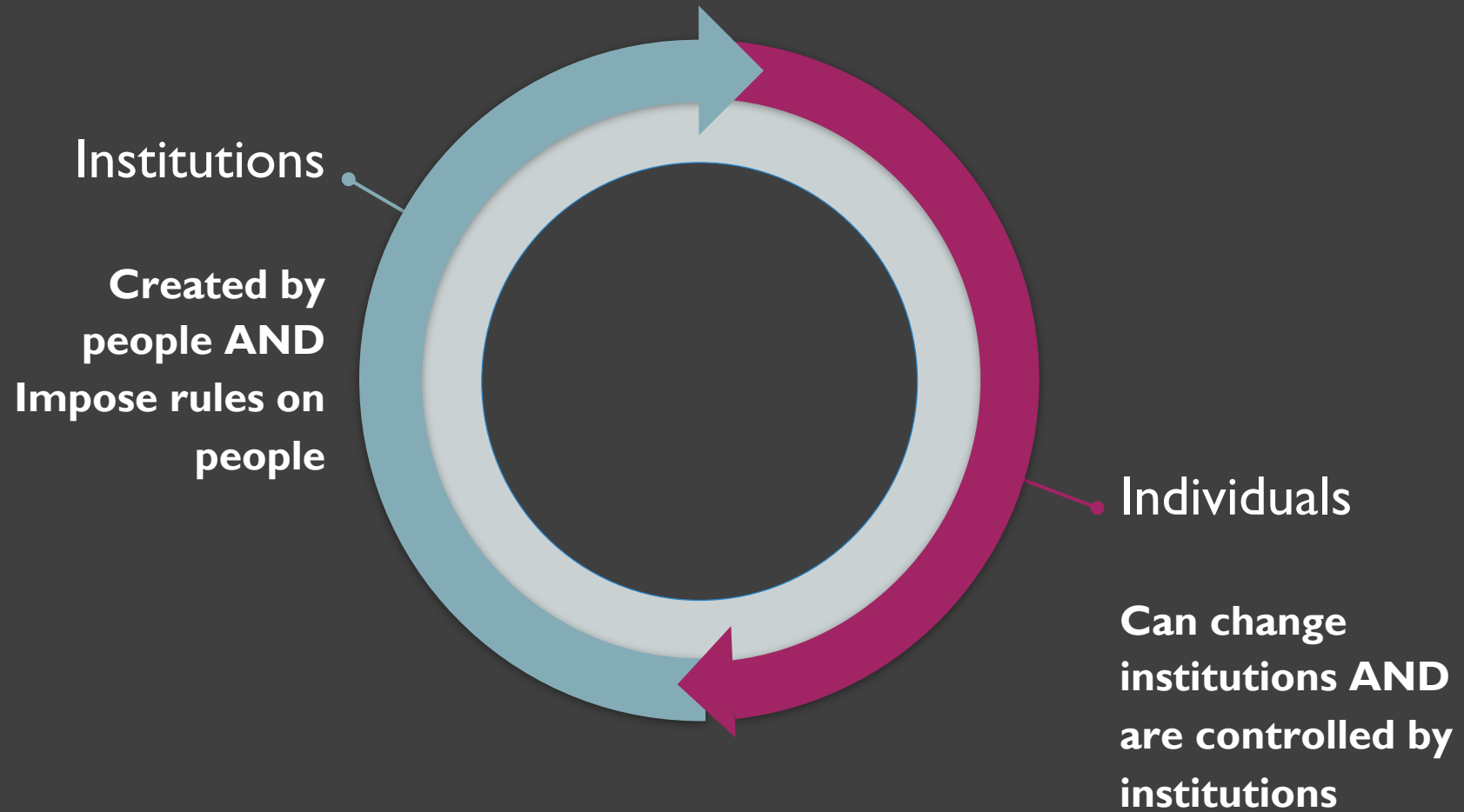
LESSON 2.

Your organization is an  
institution.

# Institutions

- Establish and enforce social order
- Have culture, norms, and explicit and implicit rules
- Made up of individuals, but bigger than individuals
- To individuals, institutions can seem indelible; actually they're products of human choices
- They evolve and change

# Paradox



# Data Science Example

“Casual” Workplace - what is this really?

- Dress code
- Social behavior norms
- Fewer formal structures or explicit processes

Just because an institution's norms aren't formally written out, that doesn't mean they don't exist.





# Clothing as Cultural Capital

- Social status intersects with clothing
- Clothing means different things depending on who wears it
- Clothing = an important signifier for belonging, power, and status

Casual dress does not make clothing less important!

# Workplaces as Institutions

- Unwritten/informal rules of behavior = hard to enforce
- “Casual” used as excuse for insufficient worker protections
- Strong HR infrastructure can protect less powerful workers

Members of an institution can change how the institution operates and protect each other.

LESSON 1.

Power and inequality are everywhere.

LESSON 2.

Your organization is an institution.

LESSON 3.

External forces affect institutions.

# Social Context

Even if your organization is working really hard to be conscious, aware, and socially responsible, **the other institutions you interact with may not be.**

You have to deal with that.



# Data Science Example

## Hiring for Diversity

- Diversity is a known issue inside data science... and also outside
- Discrimination can be a problem...
  - In business
  - In academia
  - Pretty much everywhere
- Our choices in hiring could lend credibility to that discrimination

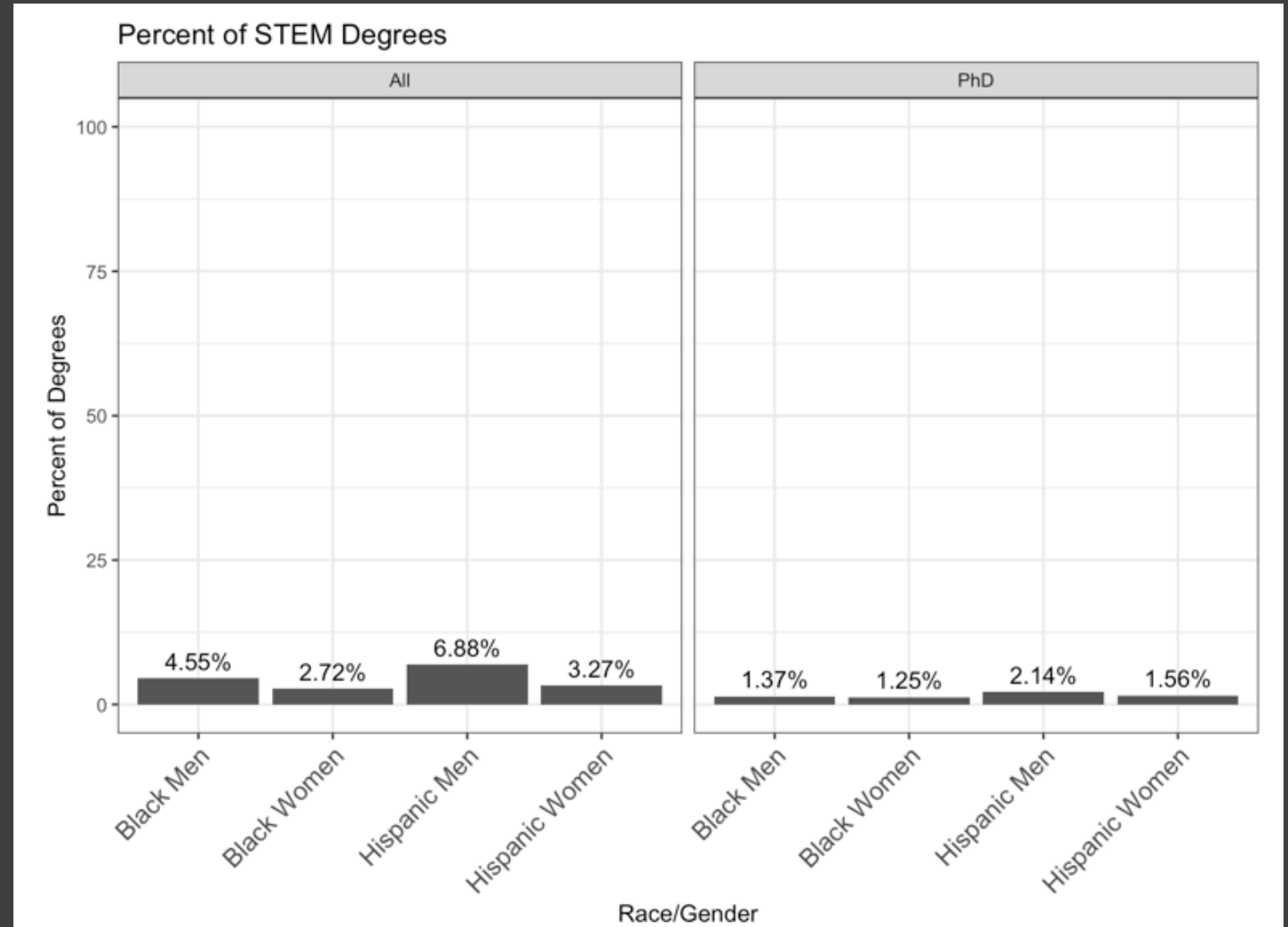
# Visualize the Problem

< 8% of STEM degrees are awarded to black people.  
For PhDs, < 3%.

< 11% are awarded to Latinx people.  
For PhDs, < 4%.

Your job requirements matter.

\*NCES 2017



# Threading the Needle

- Diploma != qualification/ability
- Skill can be acquired in many ways
- Privilege influences decisionmaking

Data science is an institution – it imposes norms, sets rules, and holds power over us.

But, at the same time, we make it what it is.

# Wrapping Up



# Hard Questions You Need To Ask

- Where do you hold power?
  - Who holds power in your organization? How is that power used?
- Are the systems in your organization producing the results you want?
  - What would you like to change?
- Where is cultural capital in your life?
  - Is cultural capital rewarded in your organization, and if so, how?
- Are you considering the impact of your everyday choices in a sociological way?
  - Are you working to overcompensate for external forces and institutions perpetuating bias in your sphere?

# Recap of Lessons

1. Power and inequality are everywhere.
2. Your organization is an institution.
3. External forces affect institutions.

Use this knowledge and awareness to make your workplace or organization a better place!

# Thank you.

 @data\_stephanie  
www.stephaniekirmer.com

## Citations

Boykis, V. 2019, Feb 13. <http://veekaybee.github.io/2019/02/13/data-science-is-different/> “Data Science is Different Now”

Cao, Longbing. 2017, Jun. Data Science: A Comprehensive Overview. ACM Computing Surveys, Vol. 50, No. 3, Article 43.

Giang, V. 2015, Mar 13. Quartz. “The casual workplace is making sexual harassment harder to identify – and stop.”  
<https://qz.com/361740/the-casual-workplace-is-making-sexual-harassment-harder-to-identify-and-stop/>

Harris, A. 2019, April 19. The Atlantic. “The Disciplines Where No Black People Earn Ph.D.s”.  
<https://www.theatlantic.com/education/archive/2019/04/lack-of-black-doctoral-students/587413/>

Livni, E. 2018, Aug 21. Quartz. “All career advice for women is a form of gaslighting.” <https://qz.com/work/1363399/all-career-advice-for-women-is-a-form-of-gaslighting>

Miller, C. C. 2016, Mar 18. NY Times. “As women take over a male dominated field, the pay drops.”  
<https://www.nytimes.com/2016/03/20/upshot/as-women-take-over-a-male-dominated-field-the-pay-drops.html>

NCES Digest of Education Statistics, 2017. Jan 2018. <https://nces.ed.gov/programs/digest/d17/> U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics.

The Burtch Works Study: Salaries of Data Scientists, May 2018. [www.burtchworks.com](http://www.burtchworks.com)

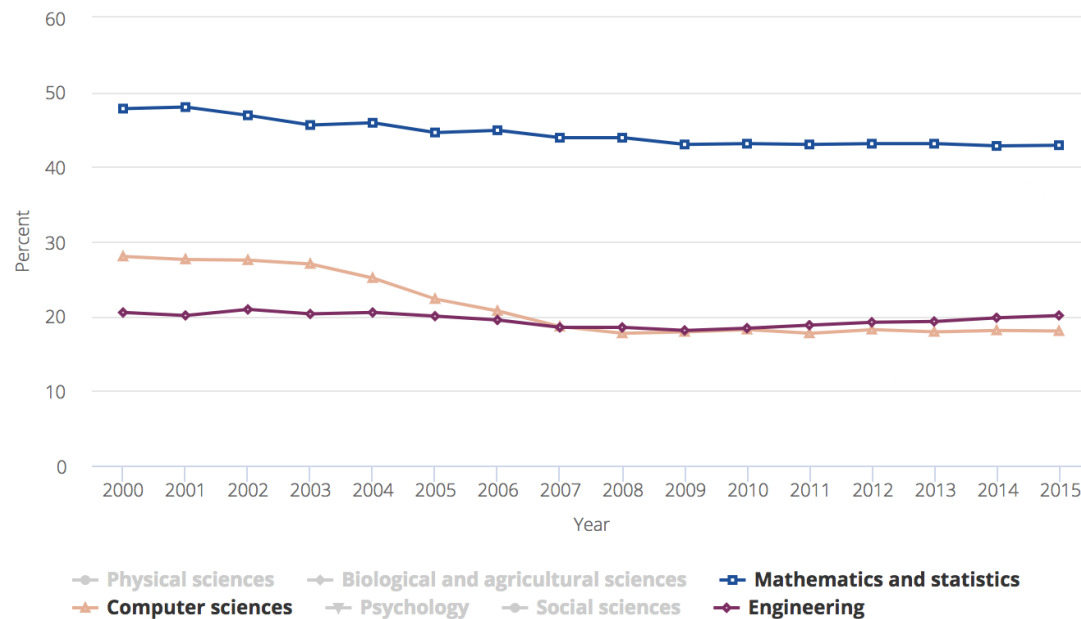
# Appendix

# Science/Engineering Bachelors' by Recipient Gender

Figure 2-11



Women's share of S&E bachelor's degrees, by field: 2000–15



## Note(s)

Physical sciences include earth, atmospheric, and ocean sciences.

## Source(s)

National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Completions Survey; National Science Foundation, National Center for Science and Engineering Statistics, WebCASPAR database, <https://ncesdata.nsf.gov/webcaspar/>.

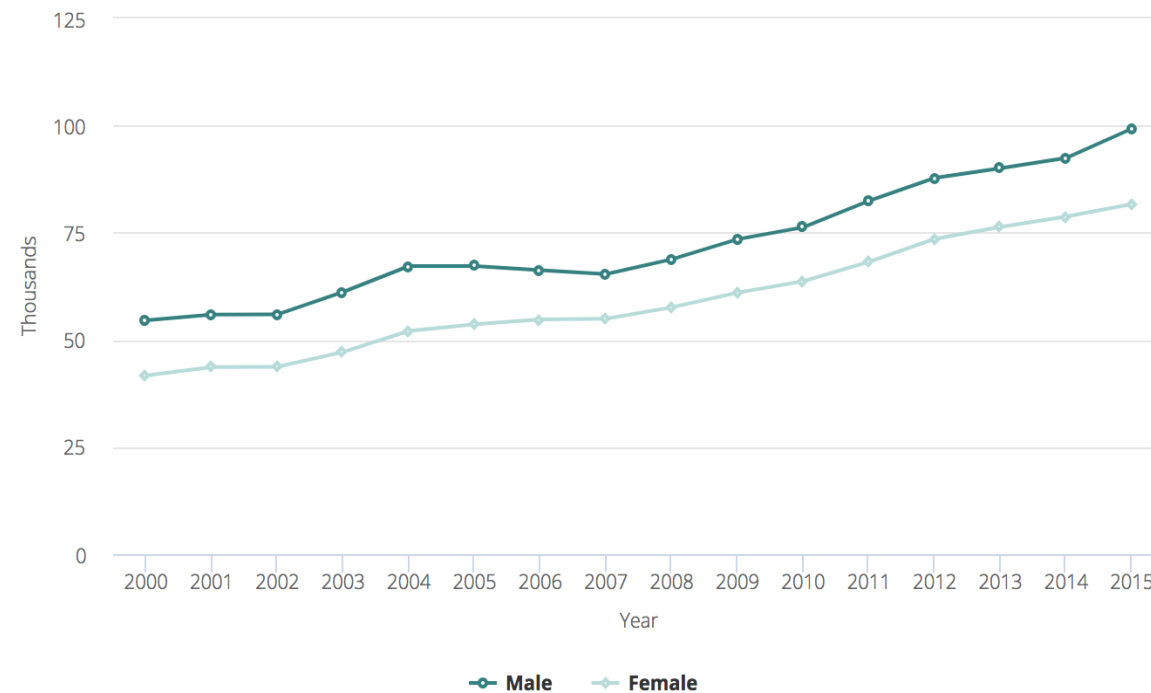
*Science and Engineering Indicators 2018*

# Science/Engineering Masters' by Recipient Gender

Figure 2-14



S&E master's degrees, by sex of recipient: 2000–15



**Source(s)**

National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Completions Survey; National Science Foundation, National Center for Science and Engineering Statistics, WebCASPAR database, <https://ncesdata.nsf.gov/webcaspar/>.

*Science and Engineering Indicators 2018*

# Data Science Terminology

Table I. Some Key Terms in Data Science

Key terms	Description
Advanced analytics	Refers to theories, technologies, tools, and processes that enable an in-depth understanding and discovery of actionable insights in big data, which cannot be achieved by traditional data analysis and processing theories, technologies, tools, and processes.
Big data	Refers to data that are too large and/or complex to be effectively and/or efficiently handled by traditional data-related theories, technologies, and tools.
Data analysis	Refers to the processing of data by traditional (e.g., classic statistical, mathematical, or logical) theories, technologies, and tools for obtaining useful information and for practical purposes.
Data analytics	Refers to the theories, technologies, tools, and processes that enable an in-depth understanding and discovery of actionable insight into data. Data analytics consists of descriptive analytics, predictive analytics, and prescriptive analytics.
Data science	Is the science of data.
Data scientist	Refers to those people whose roles very much center on data.
Descriptive analytics	Refers to the type of data analytics that typically uses statistics to describe the data used to gain information, or for other useful purposes.
Predictive analytics	Refers to the type of data analytics that makes predictions about unknown future events and discloses the reasons behind them, typically by advanced analytics.
Prescriptive analytics	Refers to the type of data analytics that optimizes indications and recommends actions for smart decision-making.
Explicit analytics	Focuses on descriptive analytics typically by reporting, descriptive analysis, alerting, and forecasting.
Implicit analytics	Focuses on deep analytics, typically by predictive modeling, optimization, prescriptive analytics, and actionable knowledge delivery.
Deep analytics	Refers to data analytics that can acquire an in-depth understanding of why and how things have happened, are happening, or will happen, which cannot be addressed by descriptive analytics.