

While we wait to begin, please remember to fill out the pre-workshop knowledge survey if you haven't already!

Reproducible Data Science (mini) Workshop

Instructor: Jessi Rick

✉ rick.jessica@cornell.edu

🐦 [@jessi_rick](https://twitter.com/jessi_rick)



Note that this workshop will be recorded, and this recording and my slides will be available to participants afterward.

What is **data science**?

Data science is the field of study that combines **domain expertise, programming skills**, and knowledge of **mathematics and statistics** to extract meaningful insights from data

"What is data science?"

<https://www.datarobot.com/wiki/data-science/>

What is **reproducible research**?

Research is reproducible when others can reproduce scientific results given only **the original data, code, and documentation**

Essawy et al. (2020)

<https://doi.org/10.1016/j.envsoft.2020.104753>

What is **open science**?

...the practice of science in such a way that **others can collaborate and contribute**, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods

Foster Open Science, "What is Open Science?"
<https://www.fosteropenscience.eu/learning/what-is-open-science/>

...the idea that scientific knowledge of all kinds should be **openly shared as early as is practical** in the discovery process... the overall goal of open science is to **accelerate scientific progress and discoveries** and to turn these discoveries into *benefits for all*

Masuzzo & Martens (2017), "Do you speak open science? Resources and tips to learn the language"
<https://doi.org/10.7287/peerj.preprints.2689v1>

Principles of open science

Principles of open science



Why is open and reproducible science beneficial?

contributes to knowledge
equity and justice

accelerates scientific progress and discoveries and to turn these discoveries into benefits for all

reduces research misconduct (both intentional and unintentional)

increases the transparency and accessibility of scientific research

makes it **easier to remember** what you did, explain to others, or to re-do analyses on the same or future projects

enables **follow-up research** and meta-analyses

Increases paper **citation rates**

Knowledge equity & knowledge justice

Knowledge equity & knowledge **justice**

facts, information, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject

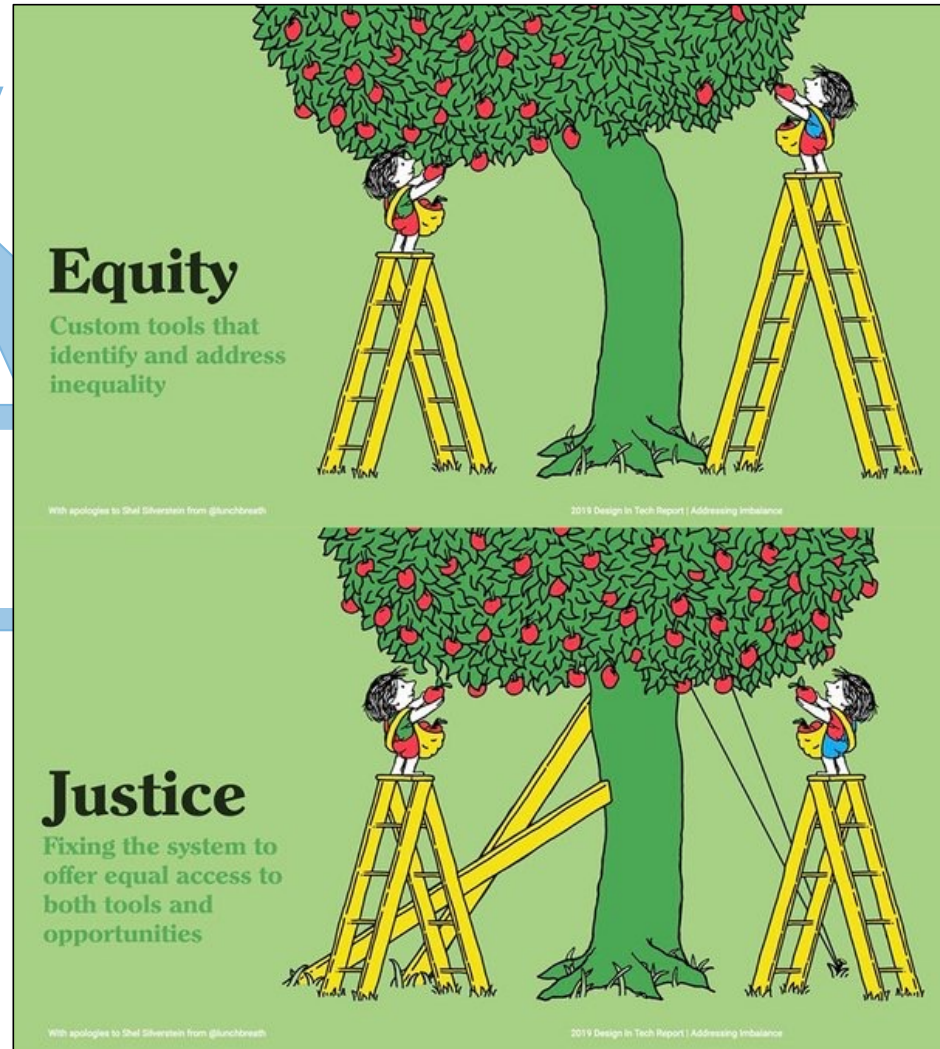
the quality of being fair and impartial

the quality of being just, impartial, or fair

Oxford English Dictionary

Knowledge equity

facts, information, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject



Tony Ruth's Equity Series (2019)

April 2023

nary

Knowledge equity & knowledge **justice**

facts, information, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject

the quality of being fair and impartial

the quality of being just, impartial, or fair

"Justice can take equity one step further by fixing the systems in a way that leads to long-term, sustainable, equitable access for generations to come."

Erdmann (2021), "Defining: Equity, Equality and Justice"

<https://achievebrowncounty.org/2021/05/defining-equity-equality-and-justice/>

Knowledge equity & knowledge justice

Both can be advanced by facing and acknowledging biases in **who generates** knowledge, **who has access** to information, and **who controls** the distribution of information

*The movement toward open science is a **step toward** restructuring the system to make it more equitable & just*

Principles of open science

How does each of these address the issues of knowledge (in)equities or knowledge (in)justices?

Open
data

Open
code

Open access
papers

Open
peer review

The reality of open and reproducible research

What are two to three tools/practices in your field that facilitate reproducibility or make it **easier** for researchers to make their work reproducible?

What is one area where you think that there is room for **improvement**?

- 1 Brainstorm by yourself
- 2 Share in small groups
- 3 Share common themes with the larger group

Okay, *but how?*

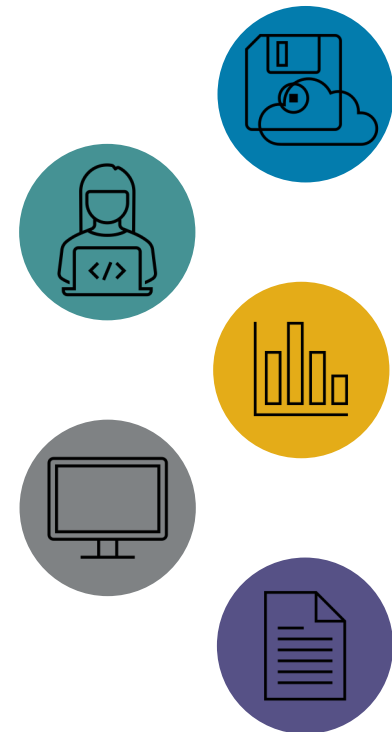
Sound **data management** practices

Thorough data and analysis **documentation**

Data-driven table and figure generation

Public archiving of data and analyses

Manuscript **pre-printing** (as possible)



Okay, *but how?*

Sound **data management** practices

Thorough data and analysis **documentation**

Data-driven table and figure generation

Public archiving of data and analyses

Manuscript **pre-printing** (as possible)



RESEARCH DATA MANAGEMENT SERVICE GROUP
Comprehensive Data Management Planning & Services

The Research Data Management Service Group (RDMSG) is a collaborative, campus-wide organization that assists with creating and implementing data management plans, applying best practices for managing data, and finding data management services at any stage of the research process.

<https://data.research.cornell.edu/>