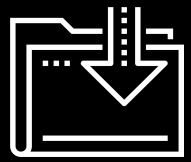
Legal and Ethical Use of Data

Data Boot Camp

Module 10 Class 1



Class Objectives

By the end of today's class you will be able to:

Explain how copyright and contract law can protect datasets.

List ethical concerns to think through when using data about people.

Describe why it's important to be able to reproduce and replicate other people's work.

Store data in a Mongo database.


WELCOME

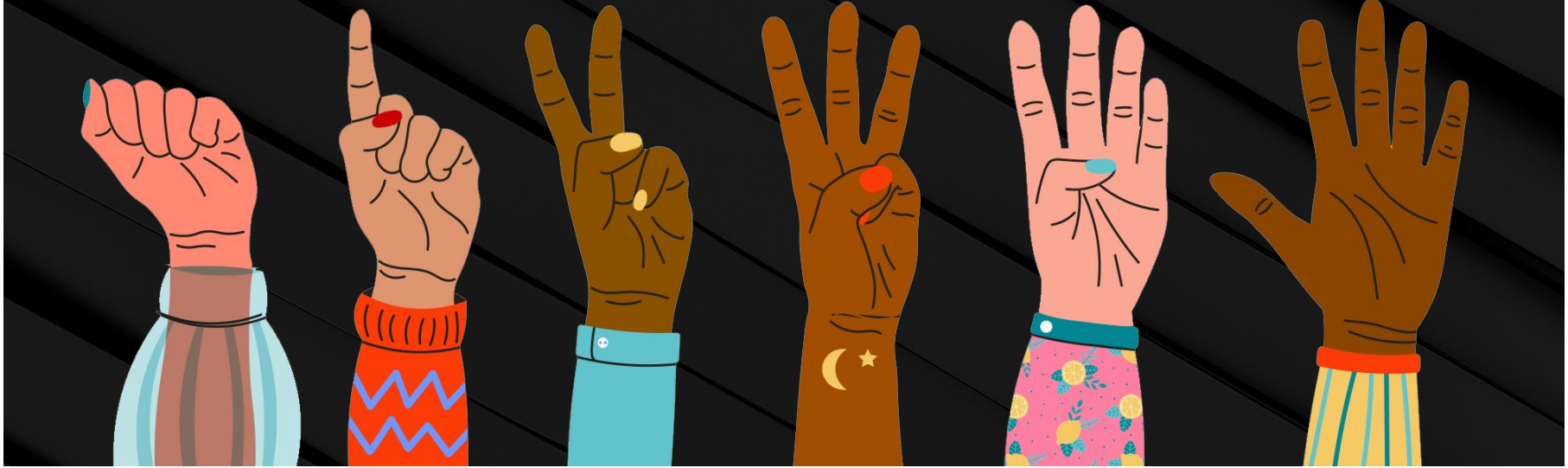
How are you feeling about your progress so far?

You’re starting to build your skill set.

It’s okay to feel overwhelmed, but don’t give up.

FIST TO FIVE:

How comfortable do you feel with this topic?



Using Data Legally and Ethically

The legal and ethical issues of using datasets are important to people who work with data.

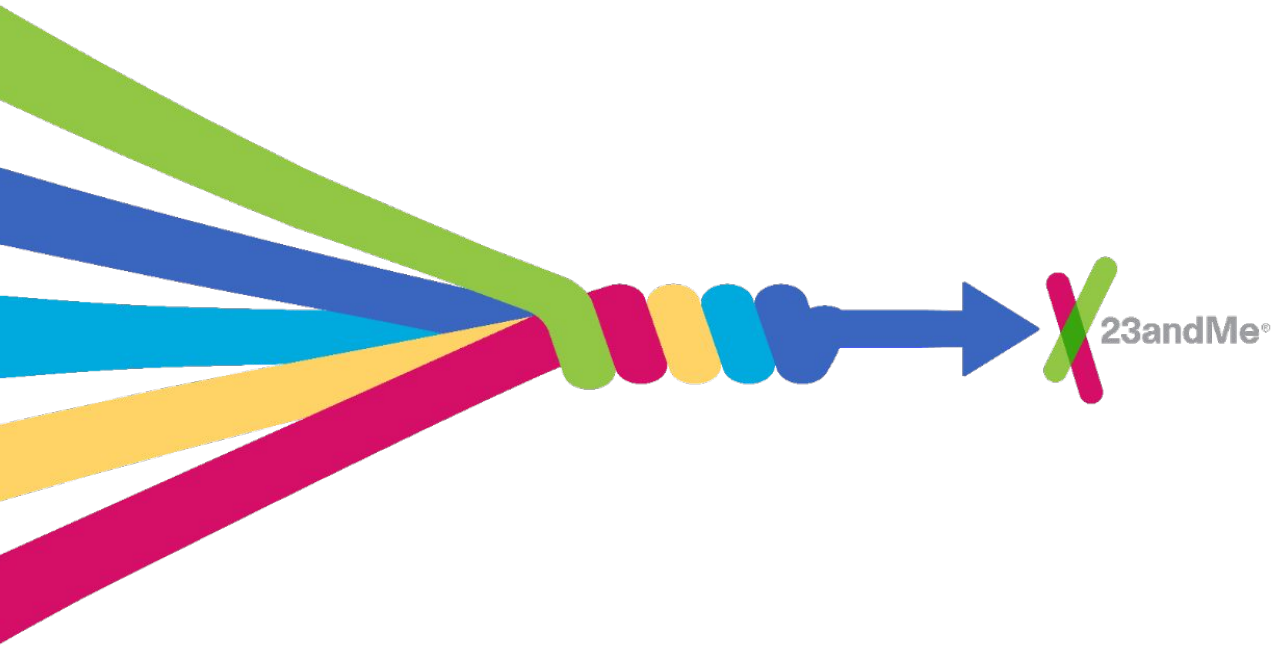
Case Study: 23andME

The genetic testing company, 23andMe, has access to the genetic information of millions of consumers.

Because family members share genetic information, this impacts not only the consumers who opted in to the company services but also their family members.

Case Study: 23andME

When companies get sold or merged, the data that they own or have access to might be part of the sale or merger. Transferring data also creates cybersecurity risks.



Data protection

Fears over DNA privacy as 23andMe plans to go public in deal with Richard Branson

Genetic testing company with 10 million customers' data has 'huge cybersecurity implications'

Richard Branson's company is joining forces with 23andMe through what is known as a 'blank-check company'. Photograph: Daniel Leal-Olivas/AFP/Getty Images

Kari Paul *in San Francisco*

Tue 9 Feb 2021 16.52 EST

The genetic testing company 23andMe will go public through a partnership with a firm backed by the billionaire Richard Branson, in a deal that has raised fresh privacy questions about the information of millions of customers.

https://www.theguardian.com

Case Study: Supreme Court

The Supreme Court ruled that a police officer did not break the law when he searched a license-plate database for an unauthorized purpose.

https://www.cnn.com

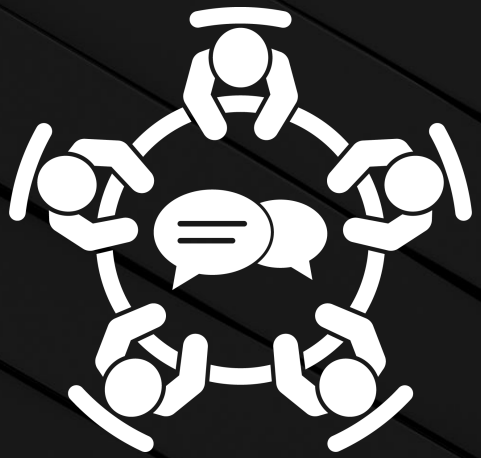
What’s legal and what’s ethical aren’t always clear. In any situation, you have to use your best judgment. And, erring on the side of caution is usually best.

Activity: Clearview AI

In this activity, you will learn about Clearview AI, a facial recognition company that was fined £17 million (about $23 million) by the Information Commissioner’s Office of the United Kingdom for gathering data about UK citizens without their knowledge.

Suggested Time:

10 Minutes



Group Discussion

Suggested Time:

5 Minutes

Questions?

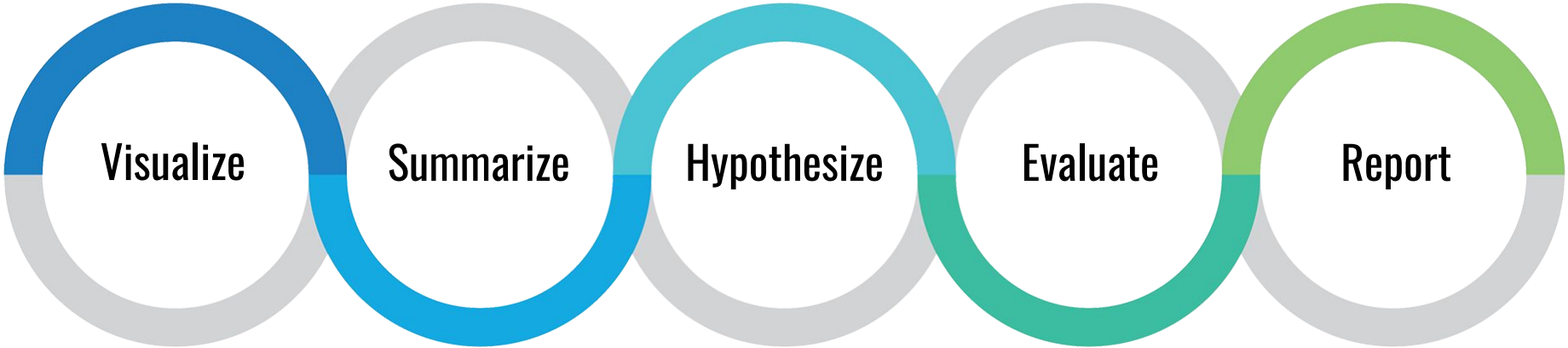
Reproducibility, Replicability, and Mongo

Replicability refers to replicating the results of someone else's analysis by using the same dataset and methods.

Reproducibility refers to reproducing their work by using a different dataset and methods.

Reproducibility and Replicability

Reproducibility and replicability are part of the **scientific method**, which is how we try to establish what's true about the world.



Both help ensure that people produce accurate work with valid results.

Any analytical or experimental result might be a chance occurrence.

That's a statistical truth: a certain percentage of the time, a result doesn't reflect the truth of the world.

We’re all human, so we make mistakes.

Unfortunately, cases of fraudulent work also exist.

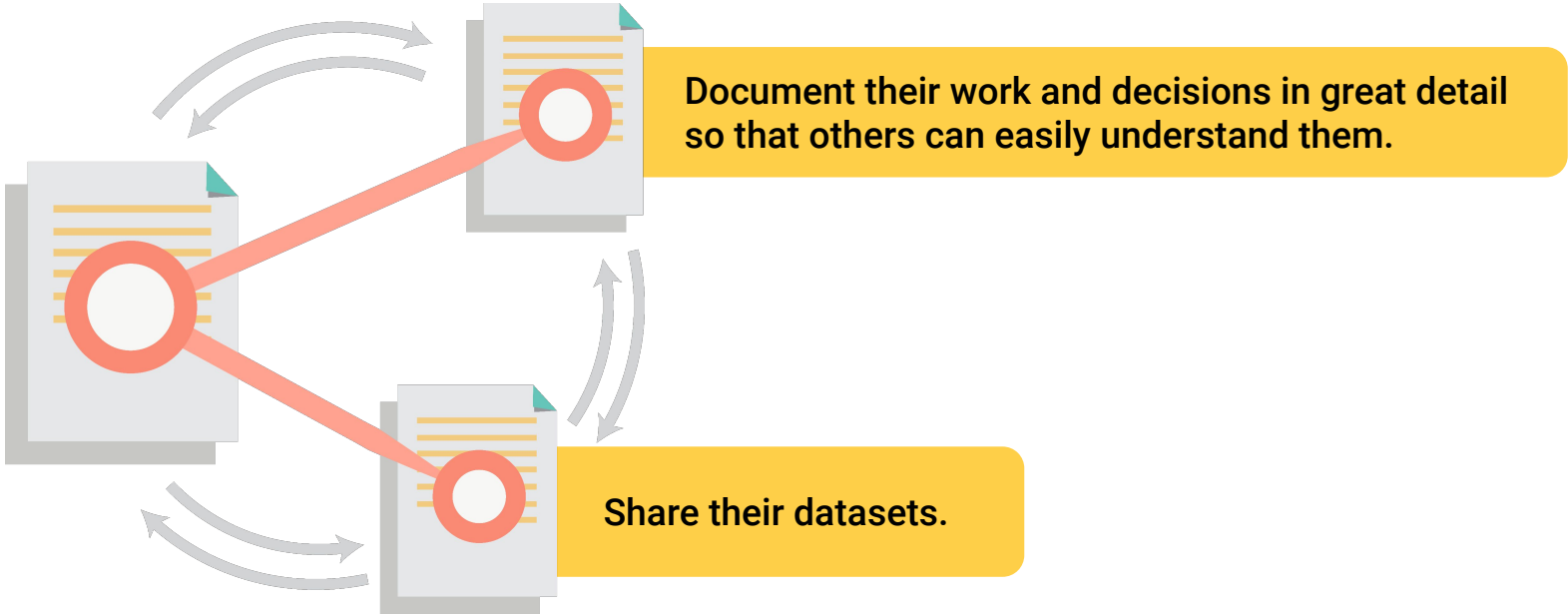
We have many examples of researchers behaving unethically by falsifying data and results.

Replication studies can reveal these ethical breaches.

Reproducibility and Replicability

A growing consensus states that researchers and analysts should set up their work to make replication and reproduction as easy as possible.

This means that they should take two main actions:

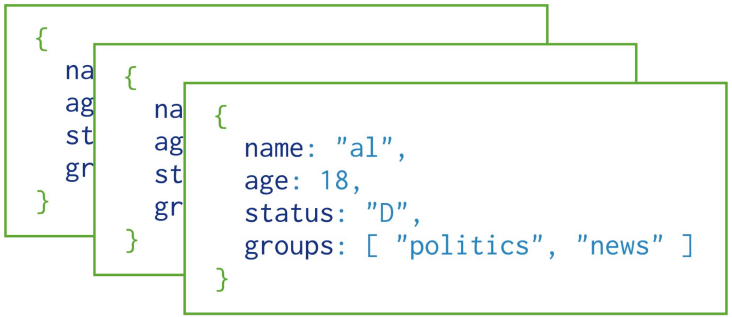


Mongo

Mongo offers an excellent way to organize and share data. The reason is twofold.

01

You can group multiple data files in collections.



02

It doesn't put many constraints on data structures.

Mongo

Whether you have a table or a series of API calls, saving that data to a database in Mongo is straightforward. **That saves you lots of work!**

- You can use Mongo to build an API of your own allowing others to query your data.
- We'll keep our use of Mongo fairly simple. But, know that it can handle more-complex data structures.
- We'll import existing data files into Mongo and then use PyMongo to explore them.

Questions?

Instructor Demonstration

Mongo

Questions?

Activity: Create a Mongo Database

In this activity, you will create your own Mongo database by following the steps from the previous demo. But this time, you'll use a version of the customer database that's in a JavaScript Object Notation (JSON) file.

Suggested Time:

10 Minutes

Time’s Up! Let’s Review.

Questions?


Recap

Recap

Let’s review the skills that we covered today:

Lesson 10.1	Explaining common legal issues that relate to using datasets.
Lesson 10.2	Explaining common ethical issues that relate to using datasets.
Lesson 10.3	• Reproducibility and replicability • Storing data with MongoDB

Questions?

Challenge Q and A

Questions?


The End