My personal experience

- You've got data

- I need data and a question

- 80% of my time will be to help you figure out the question

- 80% of the remaining time will be spent getting your data into a usable state for your question

Need to <u>find</u> the right Big Data for your question

Need to <u>see</u> if your question is answerable given available data

My personal experience

- You've got data

- I need data <u>and a question</u>

- 80% of my time will be to help you figure out the question

- 80% of the remaining time will be spent getting your data into a usable state for your question

Need to <u>find</u> the right Big Data for your question

↕

Need to <u>see</u> if your question is answerable given available data

Biases and ethics

* Our learning is only as good as the training data that is used

* Training data is often biased

    * Funding biases

    * Historical biases

    * Incompatibilities