The Data Science Cycle

Ask an interesting question.

Get the data.

Explore the data.

Model the data.

Communicate and visualize the results.

What is the scientific goal? What would you do if you had all the data? What do you want to predict or estimate?

How were the data sampled? Which data are relevant? Are there privacy issues?

Plot the data. Are there anomalies? Are there patterns?

Build a model. Fit the model. Validate the model.

What did we learn? Do the results make sense? Can we tell a story?

Image is work of Joe Blitzstein.