Data Science Process

1. Ask an interesting question.

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

2. Get the data

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

3. Explore the data

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

4. Model the data

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

5. Communicate and visualize the results

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