EDA

Exploratory Data Analysis

Problem Motivation

- In order to answer a question using data, you must first understand your data.
- Exploratory data analysis (EDA) helps you understand your data.

Key EDA Questions

"Soft" EDA Questions

- Where did this data come from? How was it produced?
- Is it trustworthy?

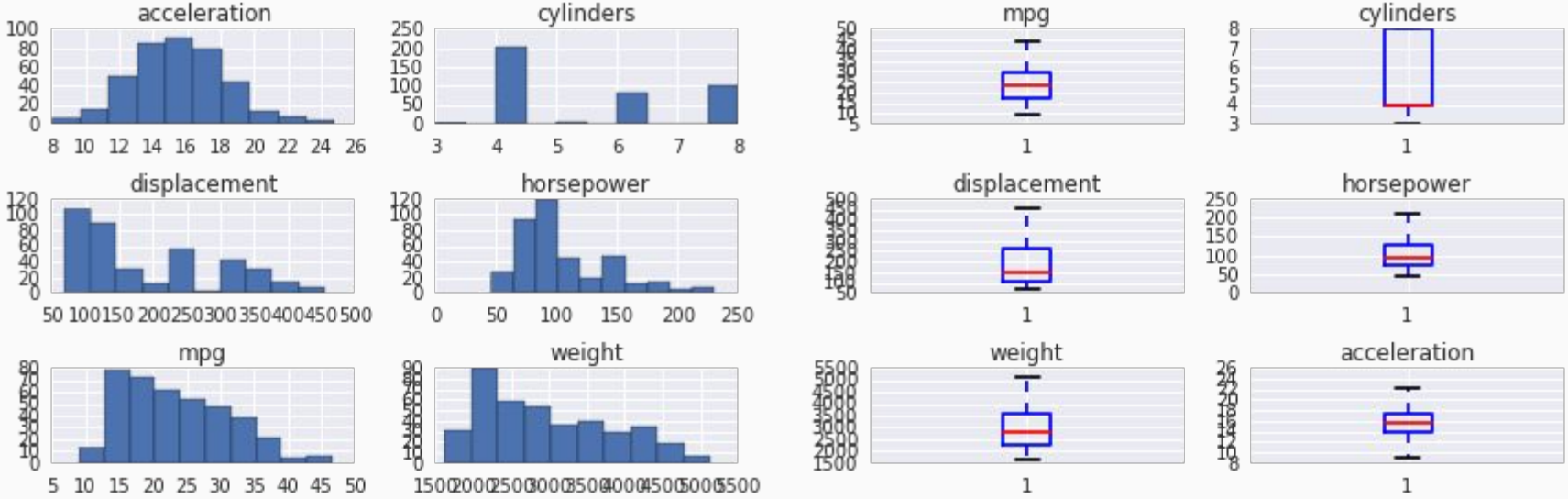
Key EDA Questions

Technical EDA Questions

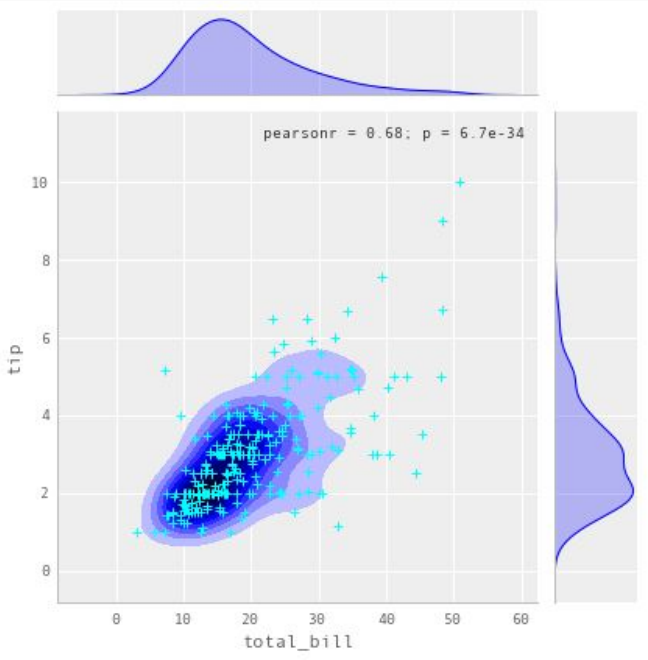
- What are the feature names and types?
- Which features are continuous and which are categorical?
- Is any data missing?
- What is the distribution of the features?
- What is the distribution of the target?
- How do the features relate to the target?
- How do the variables relate to each other?

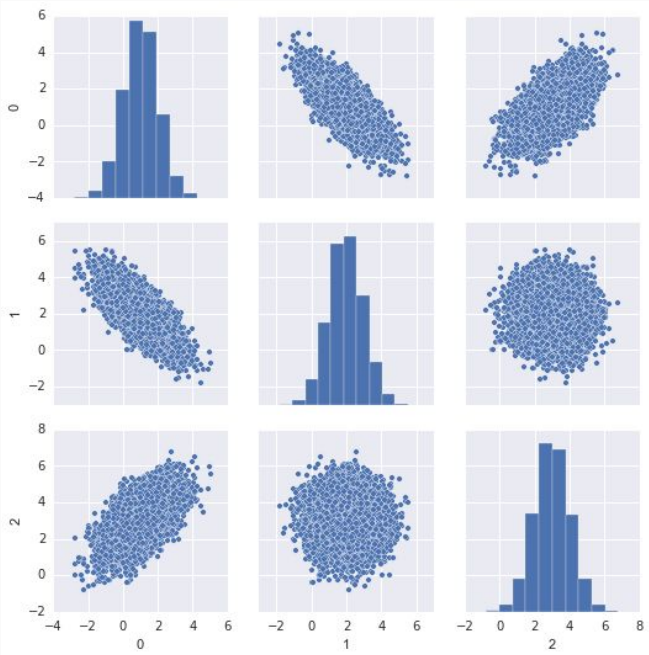
Exploring Features Individually

Look at distribution



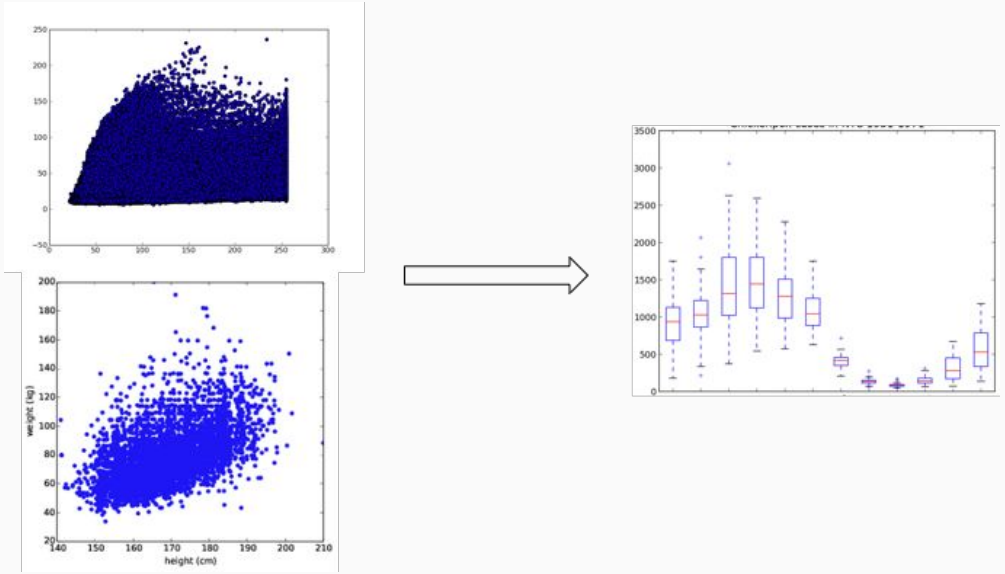
Relating Features



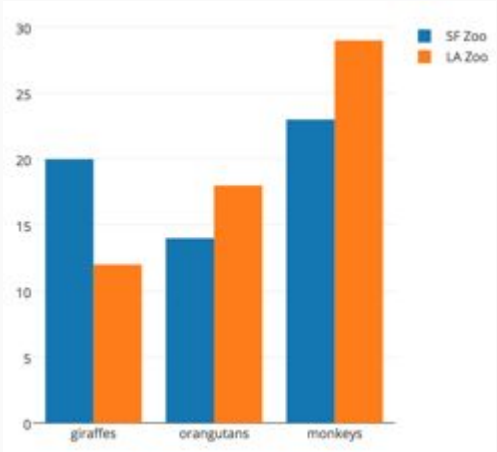


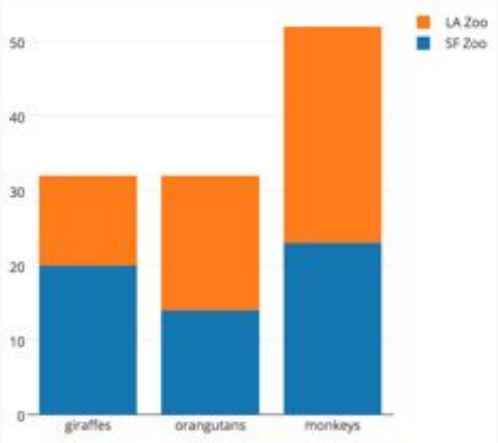
Relating Features

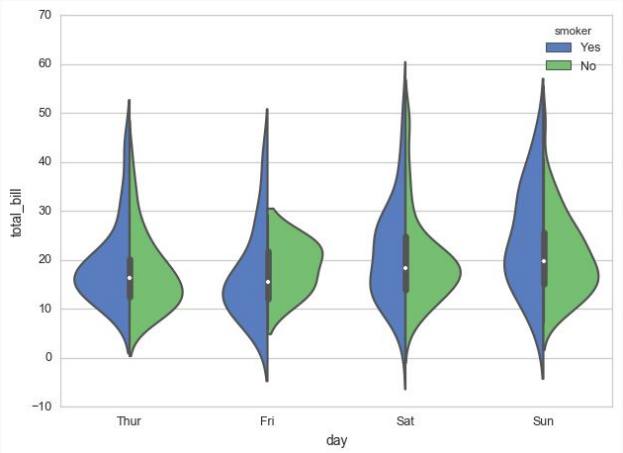
Sometimes useful to bin one of the quantitative variables



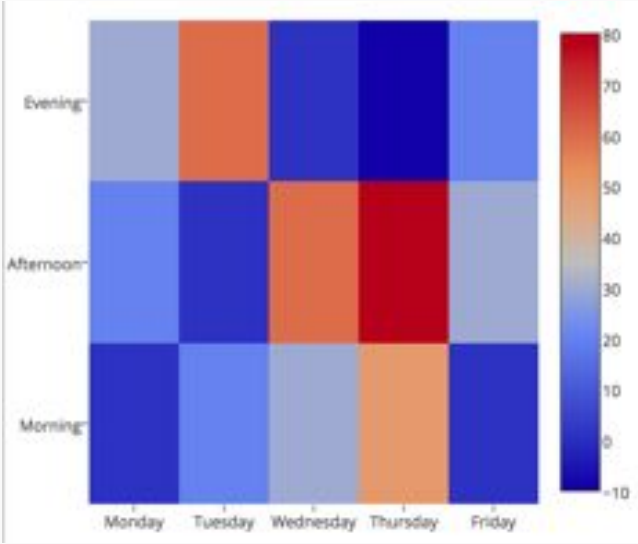
Relating Features

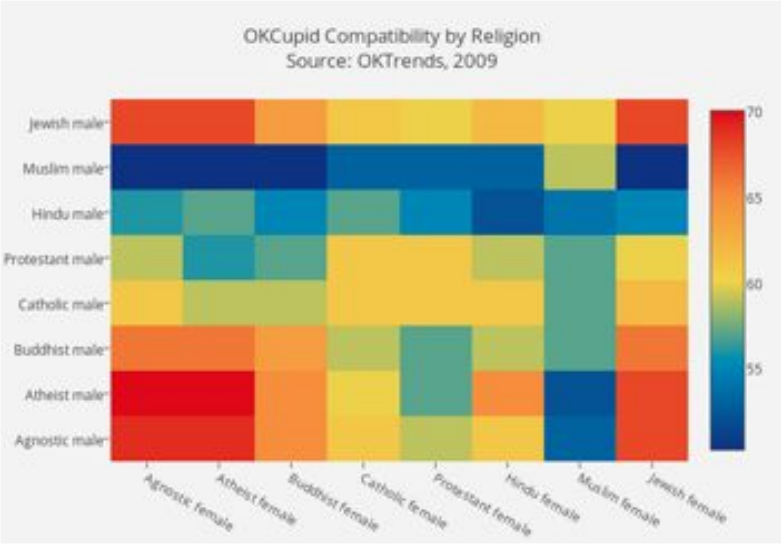






Relating Features





Relating Features