PHYS 453 - Homework 2

Before we can do fancy machine learning we need to know how to visualize a dataset and find basic relationships in the data. This process is called **Exploratory Data Analysis** (EDA), and it is the necessary first step to machine learning. You will get a mysterious dataset with little context to analyze. Remember: people only learn by doing, so the only way to learn *how* to do this is to give it a try.

My final advice is that it that it is *not* enough to just "look" at the data.
You need to have specific questions that you are trying to answer!

The *Pasty Python* is a new restaurant that has hired you to help. The staff have been collecting data on customers for you to study. You meet with the owners and they talked about lots of things:
- how much money does each person spend at lunch and at dinner?
- do people spend more on weekends?
- what is the average tip percentage?
- do men or women tip more?
- do smokers spend more than non-smokers?
- do smokers tip more?
- we need to hire the right number of kitchen and wait staff for each meal, so how many customers came for each day of the week?
- what is the difference in number of customers between lunch and dinner?

and a few other questions you forgot to write down.

Homework Instructions:
- Pick any 3 of the questions above to answer (or write your own). Make sure you state them clearly!
- Your python notebook is the report that you are giving to the owners. Make it look nice, easy to read, and have clear questions and answers.
- Some questions are best answered using pandas commands, while some need a seaborn plot. Some need both. Your final report needs **at least two** seaborn plots.
- I will post a rubric on canvas for grading this assignment.

Getting Started:
This code loads the dataset

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

tips = sns.load_dataset('tips')
```