Exercises:

1. Read Chapter 1: What is Data Science? from The Data Science Design Manual by Skiena, then respond to the questions listed below. Be sure to fully explain or justify your answer. You may also be interested to watch the corresponding video lecture here `http://www3.cs.stonybrook.edu/~skiena/data-manual/lectures/`.

    (a) What is the relationship between data science and the discplines of computer science, mathematics, and statistics?

    (b) What are some substantive application domains where data science is relevant? Explain your answer.

    (c) Provide an argument for or against the following statement: "Data science is an indepedent discipline and not simply an application or a rehashing of traditional ideas from established fields such as statistics."

    (d) Give a distinction between hypothesis-driven science and data-driven science.

    (e) Provide several examples of the types of general questions a data scientist may ask.

    (f) Find a source of data different from but in the same vein as those described in sections 1.2.1 - 1.2.4 from the reading. Describe this data source and list examples of questions that you could use your data source to answer. Use sections 1.2.1 - 1.2.4 as templates.

    (g) List the "taxonomy of data." Explain the following terms in your own words and provide an example to illustrate the concept:
    - quantitative data
    - categorical data

    (h) Provide an example of a classification problem different from the examples given in the reading.

    (i) Provide an example of a regression problem different from the examples given in the reading.

    (j) What is Kaggle and what are Kaggle challenges?

2. Visit `http://data.gov`, and identify five data sets that sound interesting to you. For each write a brief description and propose three interesting things you might do with them.

3. You would like to conduct an experiment to establish whether your friends prefer the taste of Coke or Dr. Pepper. Briefly outline a design for such a study.