

**DSCI-549: Introduction to  
Computational Thinking and  
Data Science**

**USC Viterbi School  
of Engineering**

## Homework 5

### Goals

The purpose of this homework is to learn to analyze data in real data science scenarios and to describe data collection, pre-processing, and analysis steps using technical data science terms you learned in class. This homework is the Part 2 of the class project that you started in homework 1, consisting of a total of 3 parts done in 3 separate homework assignments.

### Assignment

Describe the project that you proposed in homework 1 in expanded data science terms (100 points). Complete this assignment in a word/PDF document (max size 12 font, single spacing, 2-5 pages, not including the appendix), **with sections that correspond to the requirements below**. The homework should be submitted on D2L.

1. Include a short description of the project that you proposed in homework 1 in expanded data science terms you learned in the class. And underline the terms.
2. Describe in technical terms (underline each of them) the characteristics of the input datasets.
3. List the data pre-processing steps that you believe are involved.
4. State the analysis you intend to do, on what variable/dataset, and list the main analysis steps.
5. Sketch high-level workflows that show both the data pre-processing and analysis steps, illustrating the dataflow. Include figures and **descriptions** of the workflow.
6. Include as an appendix the original description that you wrote for homework 1, which you can update it based on comments or new understanding. **Highlight your changes in red font**. Inform the instructor if the homework need to exceed the 5 page limit.

---

### IMPORTANT NOTES

*Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in the Student Handbook <https://policy.usc.edu/studenthandbook/>. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <https://policy.usc.edu/research-and-scholarship-misconduct/>.*

*A number of USC’s schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the American Language Institute <http://dornsife.usc.edu/ali>, which sponsors courses and workshops specifically for international graduate students.*

*For more information, see the class syllabus and the USC web site.*