

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

**USC Viterbi School
of Engineering**

Homework 4

Communications

Please contact TA **Tae Woo Kang** (taewooka@usc.edu) for this homework and include “DSCI549” in the subject line of your email.

Assignment

The purpose of this homework is to use data analysis workflows for image processing, network analysis, and data exploration. You must complete this assignment using the Jupyter Notebook. Answer the questions in separate word documents, along with relevant screenshots. Submit it on D2L.

For this assignment, you will be using three notebooks. You can access them here:

https://github.com/doctorningwangusc/DSCI549/tree/main/Homework%20Assignments/Assignment4_ImageProcessing_SocialNetworkAnalysis

1. Image processing: [Use Notebook1_Image_Processing] (35 points)
 - a. Provide relevant screenshot(s) to compare the results of segmentation with and without smoothing first. How are they different? Why do you think they are different?
 - b. Provide relevant screenshot(s) to compare the results of Canny edge detection with and without segmenting. How are they different? Why do you think they are different?
 - c. What is the effect of sigma on the Canny Edge Detection?
 - d. Repeat questions a-c for the long jump image. How are the results different between the two images? Provide relevant screenshot(s) to support your statement.
 - e. Based on the differences between the rocks image and the long jump image, write some hypotheses about how workflows are likely to work with different types of images.
 - f. Find 3 different kinds of images that will let you test your hypotheses, and show them here. Based on the hypotheses you just created, describe how you expect the workflow results to be.
 - g. Upload the images to Google Colab (or add links in the first cell) and run the code. Show the input image and the results. Discuss whether your hypotheses were confirmed or disproved. [Note: it is fine if your hypotheses were disproved, since they were based on your observations from only two examples.]
2. Network visualization: [Use Notebook2_AnalyzingNetworks] (35 points)
 - a. Run the notebook using the dataset “Revolution.txt”, which contains network data about people and organizations during the American Revolution. Include both the visualizations.
 - b. In the people visualization, choose the names of 3 different individuals that exhibit completely different patterns (e.g., isolated, belong to many groups, etc.), and discuss those patterns in network terms.
 - c. In the organization visualization, are there any two organizations that do not have any common members? Is there a network property that captures this?

- d. Create your own network, a simple one, following carefully the format of the file. You can see an example in “HighSchoolClubs.txt”. Show the people and organization visualizations. Are the networks as you expected?
3. Time series - Data Exploration: [Use Notebook3_Collision_Data_Exploration] (30 points)
- a. What trend do you observe in the number of collisions in LA county between 2010 and 2019?
 - b. Where do they most occur?
 - c. Look at the histogram of collisions per age group. How do you explain the increase in the 99 age group? How do you explain the spikes at 25, 30, 35, 40, 45... years old?
 - d. When are collisions most frequent? Can you form a hypothesis as to why that is? What other data (if any) would you need to confirm your hypothesis?

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.