15-110 Hw6 - General Guide

For the last three assignments of the course, instead of working on a set of written and programming problems, you will code a **programming project** from start to finish. This project will guide you through the process of building a CS tool for a specific domain. It will also allow you to make an in-depth study of either data analysis, simulation, or machine learning.

You may choose one of five different projects to work on for this assignment. Each project covers a different topic and a different CS-as-a-tool focus, but all have been standardized to be about the same level of difficulty. The projects are described on the next page; for more information, check each project's official writeup.

Hw6 has two checks and a full assignment. These are broken down as follows:

Check6-1: complete by **Friday 4/23 at noon EST**
   ● [45] complete a short written assignment on week11-12 content
   ● [55] complete Step 1 of your project, which will focus on **data organization**

Check6-2: complete by **Friday 4/30 at noon EST**
   ● [45] complete a short written assignment on week12-13 content
   ● [55] complete Step 2 of your project, which will focus on **data processing**

Hw6: complete by **Friday 5/07 at noon EST**
   ● [20] complete Step 1, as mentioned above
   ● [20] complete Step 2, as mentioned above
   ● [60] complete Step 3 of your project, which will focus on **data presentation**

Note that revision deadlines work differently for Hw6. Check6-1 revisions are due by Wed 4/28 at noon EST; Check6-2 revisions are due by Wed 5/05 at noon EST; and Hw6 does **not** have a revision deadline (on-time submissions only). We've changed the policy because A) Check6-2 relies on Check6-1 to work and Hw6 relies on both checks to work; you cannot reasonably complete all three at the same time, and B) Hw6 is due on the last day of the semester and cannot be pushed back any further.

Once you have chosen a project, you will not be allowed to switch to a different project (except in extreme circumstances, with permission from the course instructors). You must select a project by **Monday 4/19 at noon EST**. Notify the instructors about your project selection by filling out this form: https://forms.gle/JBH9wsh28nGQbSad7

Project: Battleship

CS Focus: Simulation

Description: use the class simulation framework to program an interactive game of Battleship. If you've never played Battleship before, try an example game here: https://www.mathsisfun.com/games/battleship.html

Requirements: 2D lists, tkinter, and event-based simulation

Project: Circuit Simulator

CS Focus: Simulation

Description: build a simple circuit simulator that can take a boolean expression in text format, generate an interactive circuit for the expression using simulation, and create a truth table for the expression automatically. This will allow you to explore what different boolean expressions output.

Requirements: recursion, trees, tkinter, and event-based simulation

---

Project: Language Modeling

CS Focus: Data Analysis/Machine Learning

Description: build and test your own language model based on the works of two famous folklore and fairytale authors, Andersen and the Grimm Brothers. You'll use the language model you create to generate text that is similar to the texts written by these two authors, and compare how often they used different words and phrases.

Requirements: 2D lists, dictionaries, File I/O, and matplotlib

---

Project: Protein Sequencing

CS Focus: Data Analysis

Description: use data analysis to process and analyze DNA sequences for the gene p53, which is used to suppress cancer in organisms. Specifically, you will compare the p53 genes in humans and elephants, to identify what they have in common and how they are different.

Requirements: 2D lists, dictionaries, File I/O, and matplotlib

---

Project: Social Media Analytics

CS Focus: Data Analysis

Description: analyze a dataset that maps tweets and facebook messages made by politicians to information about the text's bias, partisanship, and message intent. You will use sentiment analysis to add additional data on message sentiment, and will compare the data across several vectors.

Requirements: 2D lists, dictionaries, File I/O, matplotlib, pandas, and nltk