

Portfolio Component 1: Data Exploration

Objectives:

- In class, we covered how to do data exploration with statistical functions in R
- In this assignment, you recreate that functionality in C++ code
- This will prepare us to write algorithms in C++ in future assignments

Turn in:

- Upload your C++ code and document to your portfolio, and create a link to it on your index page
- Upload your C++ code and document to eLearning

Instructions:

1. In the C++ IDE of your choice:
 - a. Read the csv file (now reduced to 2 columns) into 2 vectors of the appropriate type. See the reading in cpp picture in Piazza.
 - b. Write the following functions:
 1. a function to find the sum of a numeric vector
 2. a function to find the mean of a numeric vector
 3. a function to find the median of a numeric vector
 4. a function to find the range of a numeric vector
 5. a function to compute covariance between rm and medv (see formula on p. 74 of pdf)
 6. a function to compute correlation between rm and medv (see formula on p. 74 of pdf); Hint: sigma of a vector can be calculated as the square root of $\text{variance}(v, v)$
 - c. Call the functions described in 1-4 for rm and separately for medv. Call the covariance and correlation functions for rm and medv together. Print results for each function.
2. Write a short document:
 - a. copy/paste runs of your code showing the output
 - b. describing your experience using built-in functions in R versus coding your own functions in C++
 - c. describe the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning
 - d. describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?
3. Create a link to this document and your code on your index page.

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.

Grading Rubric:

Element	Points
Step 1 C++ code	70
Step 2 Overview document	20
Step 3 Create links to the document and code on the index page	10
Total	100

Grading Rubric:

- 90 and above for exceptional work
- 80-89 for good work
- 70-79 for average work
- below 70 for low quality work

Caution: All course work is run through plagiarism detection software comparing students' work as well as work from previous semesters and other sources.