William Alec Akin

IN300 – Programming for Data Analysis (Java, R, Python)

Unit 7 Assignment

Dr. Robert Kayl

June 26th, 2022

# #1 – R Source Code (Mean, Median, Mode, Average Deviation, Standard Deviation):

# W. Alec Akin

# IN300 Programmging for Data Analysis (R, Python, Java)

# Dr. Robert Kayl

# Unit 7 Assignment - R Source Code

# Description: Write an R program that reads the text file into a data frame.

# Using that data frame, print the mean, median, mode, average deviation,

# and standard deviation. Ensure that each output is labeled.

# Read in the IN300\_Dataset2.txt file

data <- scan("/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt")

# Print the mean to stdout

print(paste("The mean of the data is: ", mean(data)))

# Print the median of the data

print(paste("The media of the data is: ", median(data)))

# Create a function to get the mode of the data

getmode <- function(v) {

  uniqv <- unique(v)

  uniqv[which.max(tabulate(match(v, uniqv)))]

}

# Print the mode to stdout

mode <- getmode(data)

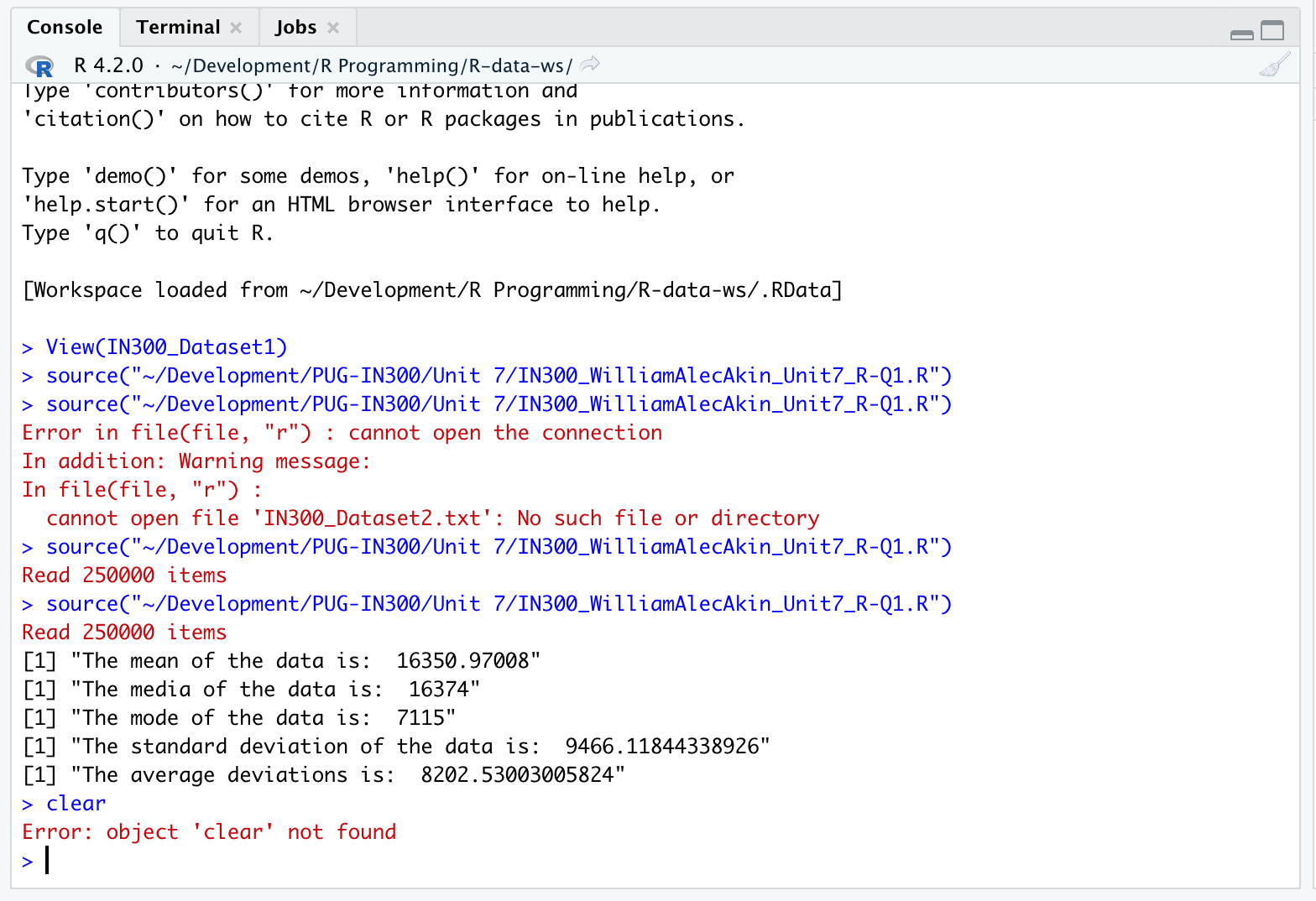
print(paste("The mode of the data is: ", mode))

# Print the standard deviatino to stdout

print(paste("The standard deviation of the data is: ", sd(data)))

# Print the average deviation of the data to stdout

print(paste("The average deviations is: ", sum(abs(data - mean(data))) / length(data)))



# #2 – Java Source Code (Mean, Median, Mode, Standard Deviation)

/\* William Alec Akin

Purdue University Global

IN300 - Programming for Data Analysis (Python, R, Java)

Dr. Robert Kayl

Unit 7 Assignment - Java Source Code (Mean, median, mode, standard deviation)

Description:

Write a Java program that reads the text file. Using meaningful print statements, output the mean, median, mode,

and standard deviation.

File name: IN300\_Dataset2.txt

\*/

//Import necessary libraries

import java.util.\*;

import java.io.File;

public class IN300\_WilliamAlecAkin\_Unit7\_JavaQ2 {

    public static void main(String[] args) {

        try {

            //Read the IN300\_Dataset2.txt file into an array

            //Specify the full file path if there are issues with loading it

            Scanner scanner = new Scanner(new File("/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt"));

            int [] result = new int [250000];

            int i = 0;

            while(scanner.hasNextInt()){

                result[i++] = scanner.nextInt();

            }

            //Calculate the mean and print it to standard out

            OptionalDouble mean = Arrays.stream(result).average();

            if(mean.isPresent()) {

                System.out.println("The mean (average) is : " + mean.getAsDouble());

            }

            else {

                System.out.println("There are no values to average. Please try again.");

            }

            //Calculate the median and print it to standard out

            if(result.length%2 == 0) { //This is for an even number of elements

                double mid1 = result[(result.length/2)-1];

                double mid2 = result[result.length/2];

                double median = (mid1 + mid2)/2;

                System.out.println("The median is : " + median);}

            else{

                double median = result[(result.length/2)+1];

                System.out.println("The median is : " + median);

            }

            //Calculate the mode and print it to standard out

            int modeCount = 0;

            double mode = 0;

            int tempCnt = 0;

            for (double intValue : result) {

                tempCnt = 0;

                for (double value : result) {

                    if (intValue == value) {

                        tempCnt++;

                    }

                }

                if (tempCnt > modeCount) {

                    modeCount = tempCnt;

                    mode = intValue;

                    }

                }

                //Wait for the result to process

                System.out.println("Mode " + mode + " appears " + modeCount + " times.");

                //Calculate the standard deviation and print it to standard out

                int sdSum = 0;

                for (double dValue : result) {

                    sdSum += Math.pow((dValue - mean.getAsDouble()), 2);

                }

                System.out.println("The standard deviation is : " + Math.sqrt( sdSum / ( result.length ) ));

                scanner.close();

            }

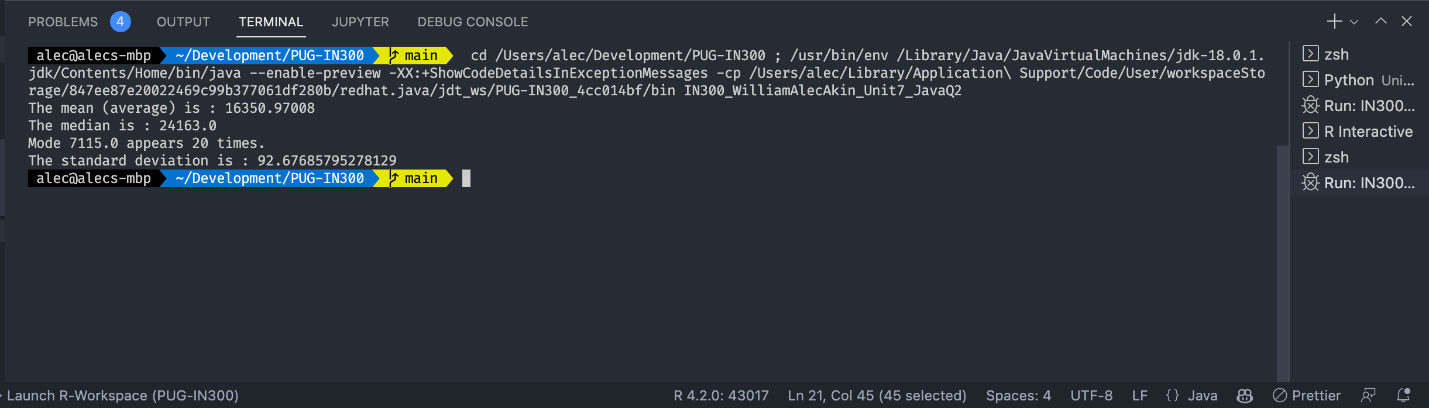
            catch (Exception ex) {

                System.out.println("Error: " + ex.getMessage());

            }

            }

    }



# #3 – Python Source Code (NumPy, SciPy, Mean, Media, Mode, Standard Deviation)

'''William Alec Akin

Purdue University Global

IN300 - Programming for Data Analysis (Python, R, Java)

Dr. Robert Kayl

Unit 7 Assignment - Python Source Code

June 19th, 2022

Description: Write a Python program which completes the following:

1. Read in IN300\_Dataset2.txt  using NumPy

2. Using meaningful statments and NumPy and SciPy, calculate and print the following:

    a. The mean of the data

    b. The median of the data

    c. the mode of the data

    d. The standard deviation of the data

'''

# Import required modules

# Note for Dr. Kayl: 'stats' is a submodule of the 'scipy' module now (it looks like) so I had to use 'scipy.stats' to get this to work

# Not sure if other students had the same issue, but wanted to let you know!

import numpy as np

from scipy import stats

# Read in the file IN300\_Dataset2.txt with numpy

# Specify as full path to avoid errors

data = np.loadtxt('/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt', delimiter = ' ', dtype = int)

# Print the mean to stdout with a message

print("\nThe mean of the data is: ")

print(np.mean(data))

# Print the median to stdout with a message

print("\nThe median of the data is: ")

print(np.median(data))

# Calculate the mode of the data using SciPy

mode = int(stats.mode(data)[0])

# Print the mode to stdout with a message

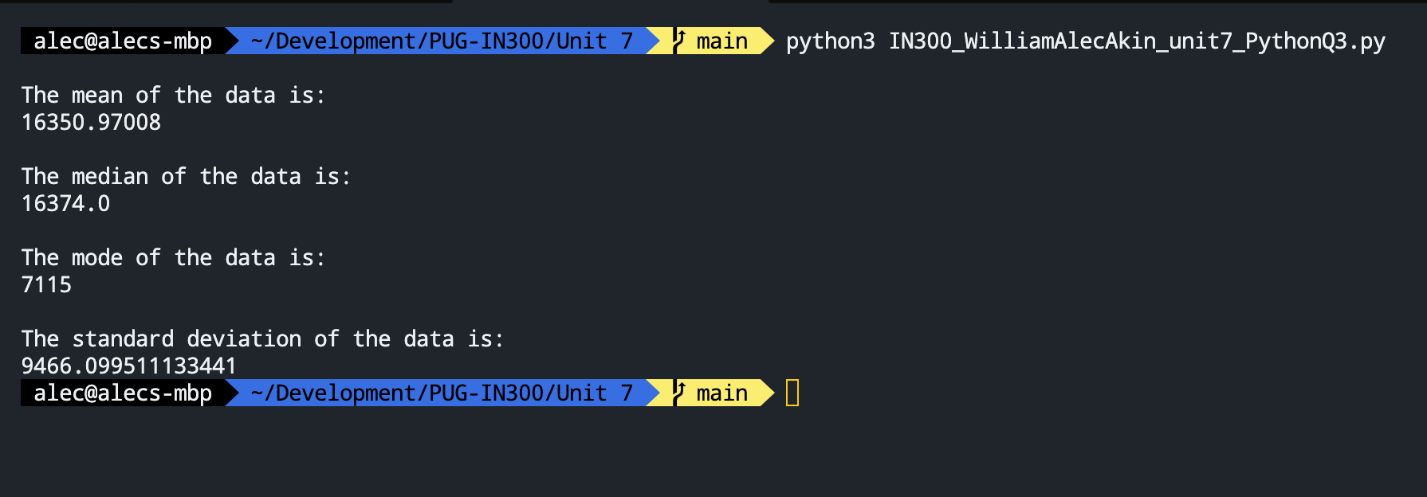
print("\nThe mode of the data is: ")

print(mode)

# Print the standard deviation to stdout with a message

print("\nThe standard deviation of the data is: ")

print(np.std(data))



# #4 – R Source Code (Read txt file, 2d array, 3d array, slice)

# W. Alec Akin

# IN300 Programmging for Data Analysis (R, Python, Java)

# Dr. Robert Kayl

# Unit 7 Assignment - R Source Code

# Description: Write an R program that does the following:

#   a. Reads in the IN300\_Dataset2.txt file

#   b. Using that data, create a two-dimensional array that is 2,500 rows by 100 columns

#   c. Also create a three-dimensional array using a start column index of 2, end of 5

#   d. Print the results of the arrays and slice

# Read in the IN300\_Dataset2.txt file

data <- scan("/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt")

# Create a two-dimensional matrix of 2500 rows by 100 columns

my2dMatrix <- matrix(data, nrow = 2500, ncol = 100)

# Print the two dimensional matrix

print(my2dMatrix)

# Create a three-dimensional array that is 2500 by 10 by 10

my3dArray <- array(data, dim = c(2500,10,10))

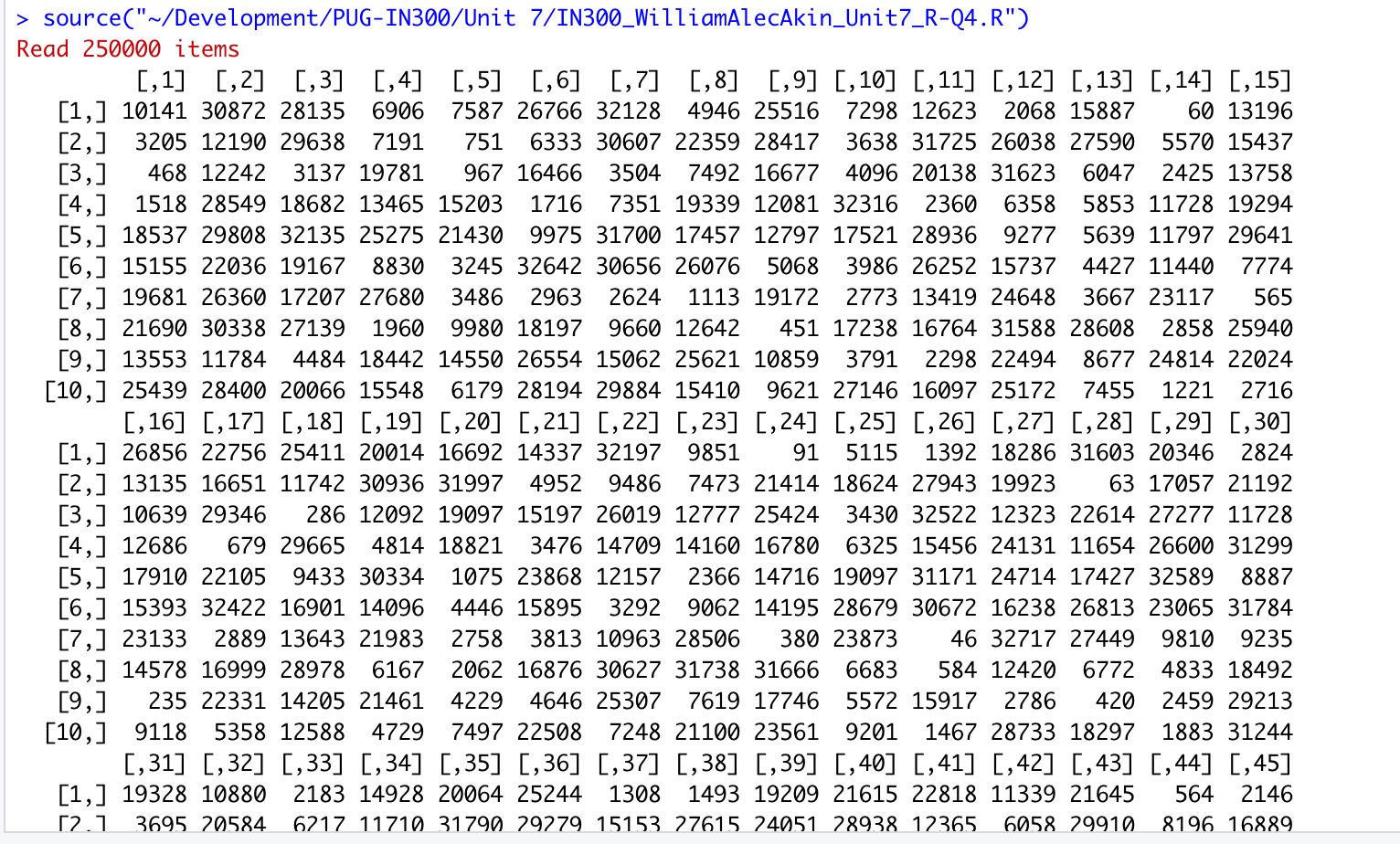
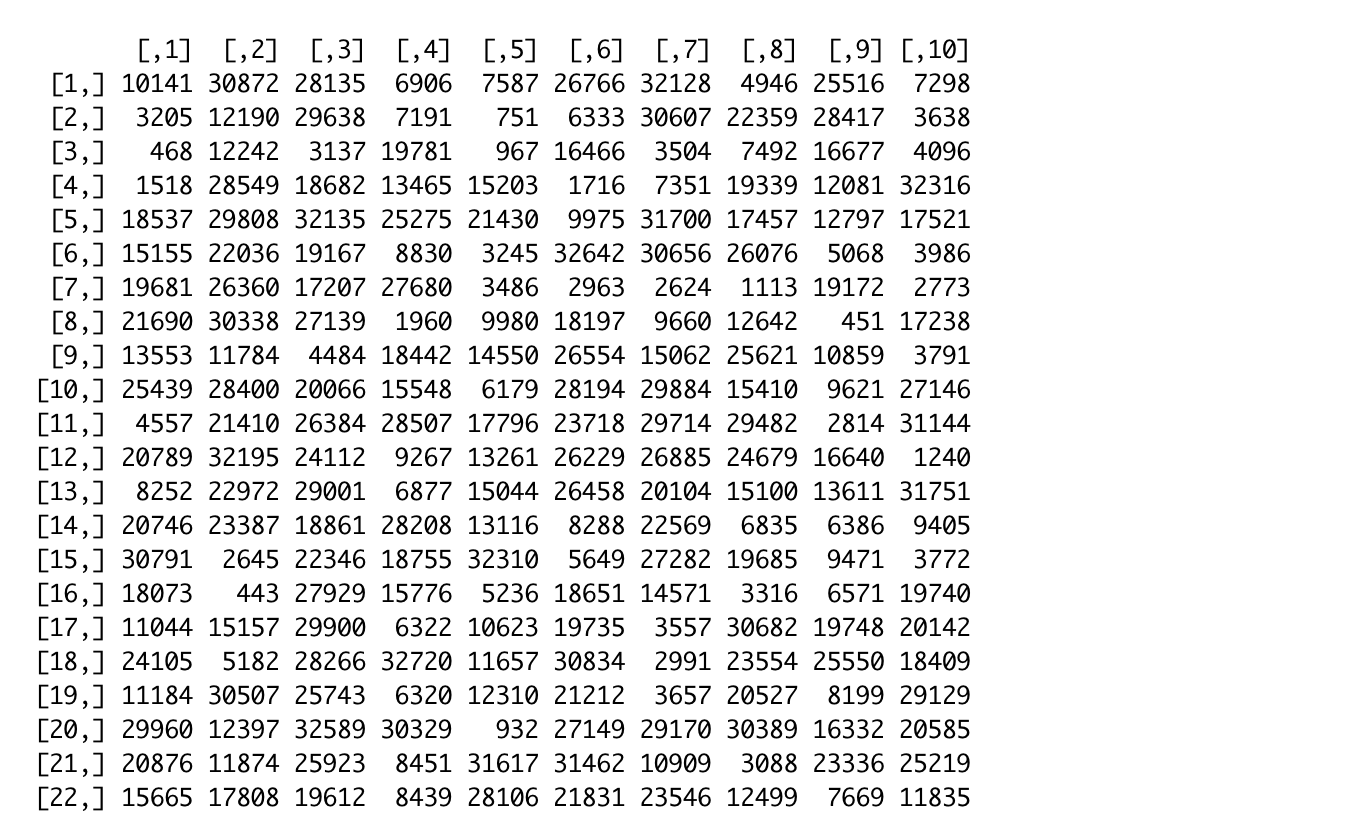
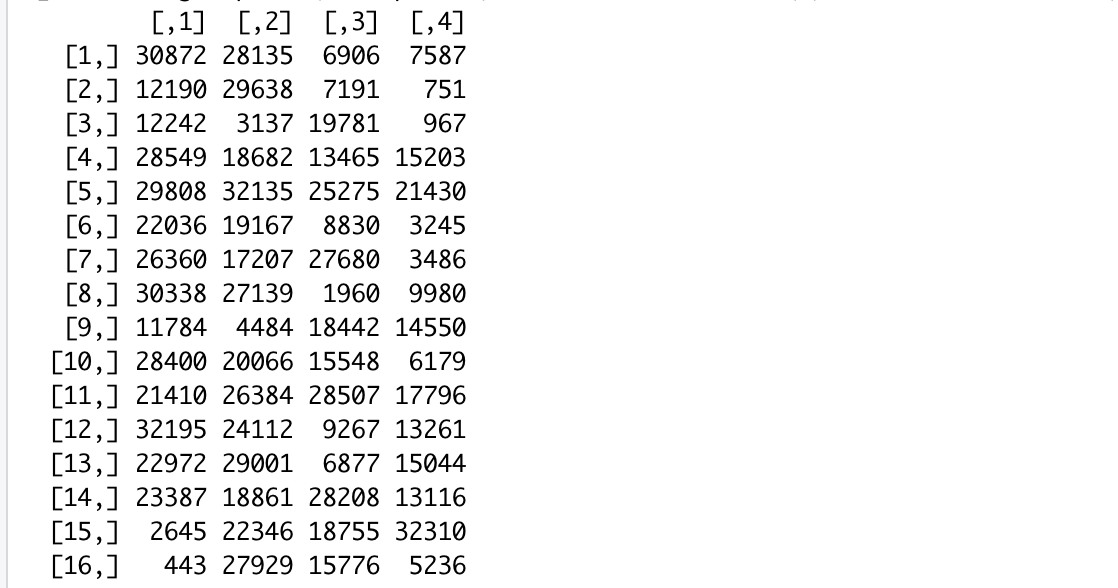
# Print the three-dimensional array

print(my3dArray)

# Slice the 2d array starting at index 2 and ending at index 5

# Then, print the results

print(my2dMatrix[,2:5])



# #5 – Java Source Code (Read TXT File, 2d array, Slice)

/\* William Alec Akin

Purdue University Global

IN300 - Programming for Data Analysis (Python, R, Java)

Dr. Robert Kayl

Unit 7 Assignment - Java Source Code (Mean, median, mode, standard deviation)

Description:

Write a Java program that does the following:

1. Reads in the text file IN300\_Dataset2.txt

2. Using that data, creates a two-dimensional array that is 2,500 rows by 100 columns

3. Slices the two-dimensional array using a starting column index of 2 and an ending column index of 5

4. Prints the restuls of the arrays and slice.

\*/

//Import necessary libraries

import java.util.\*;

import java.io.File;

public class IN300\_WilliamAlecAkin\_Unit7\_JavaQ5 {

    public static void main(String[] args) {

        try {

            //Create a two-dimensional array that is 2,500 rows by 100 columns

            int cols = 2500;

            int rows = 100;

            int[][] my2dArray = new int[cols][rows];

            //Read in IN300\_Dataset2.txt - may need full file path if there are issues with loading it

            Scanner scanner = new Scanner(new File("/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt"));

            //Fill the two-dimensional array with the values from the text file

            while(scanner.hasNextInt()){

                for(int i = 0; i < cols; i++) {

                    for(int j = 0; j < rows; j++) {

                        my2dArray[i][j] = scanner.nextInt();

                    }

                }

            }

            //Copy a slice of the array

            int[][] newArray = Arrays.copyOfRange(my2dArray, 2,5);

            for(int[] row : newArray) {

                System.out.println(Arrays.toString(row));

            }

            //This is the placeholder for the slice

            int[][] myNewArray = new int[3][];

            //Print the results of the slice

            for(int i = 2, j=0; i < 5; i++, j++) {

                myNewArray[j] = Arrays.copyOfRange(my2dArray[i], 2,5);

                System.out.println(Arrays.toString(newArray[j]));

            }

            //Close the scanner

            scanner.close();

        }

        //Catch any expections

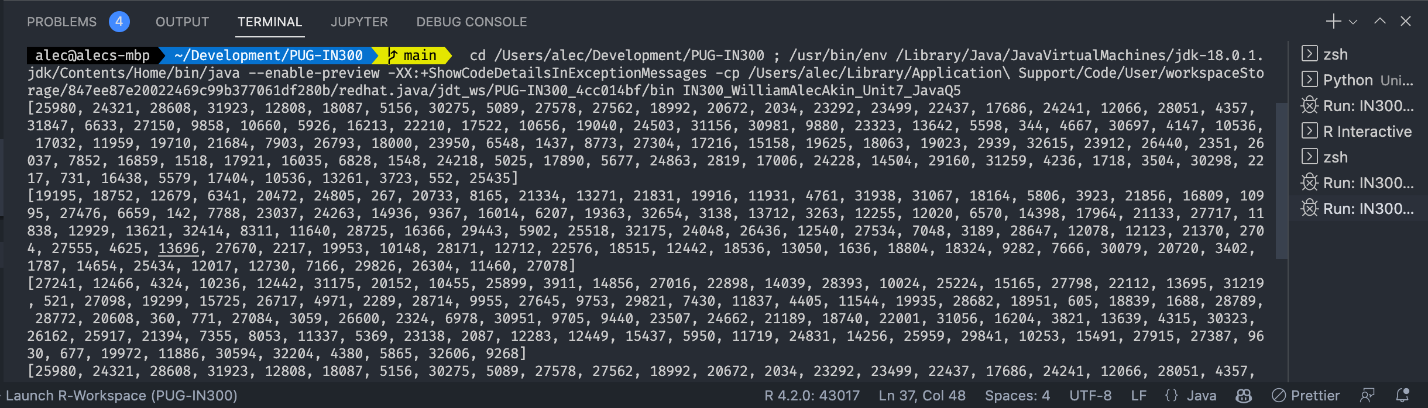
        catch(Exception e) {

            System.out.println("There was an error. Please try again." + e.getMessage());

        }

    }

}



# #6 – Python Source Code (Read TXT w/ NumPy, 2d array, 3d array, Slice)

#!/usr/bin/python3

'''William Alec Akin

Purdue University Global

IN300 - Programming for Data Analysis (Python, R, Java)

Dr. Robert Kayl

Unit 7 Assignment - Python Source Code

June 19th, 2022

Description: Write a Python program which completes the following:

1. Read in IN300\_Dataset2.txt  using NumPy

2. Using the data read in from the text file, create a two-dimensional array that is 2,500 rows by 100 columns

3. Also create a three-dimensional array that is 2,500 by 10 by 10

4. Slice the two-dimensional array using a starting index of 2, an ending index of 5, and steps by 10

5. Print the results of the slice

'''

# Import the necessary libraries

import numpy as np

# Read in the file IN300\_Dataset2.txt with numpy

# Specify absolute path to file to avoid errors

data = np.loadtxt('/Users/alec/Development/PUG-IN300/Unit 7/IN300\_Dataset2.txt', delimiter = ' ', dtype = int)

# Create the two-dimensional array/matrix by reshaping the data frame

matrix\_2d = data.reshape(2500,100)

# Print the results of the maxtrix to stdout with a message

print("The two-dimensional array/matrix is: \n")

print(matrix\_2d)

# Reshape the data to a three-dimensional array/matrix

matrix\_3d = data.reshape(2500,10,10)

# Print the results of the three-dimensional array/matrix to stdout with a message

print("The three-dimensional array/matrix is: \n")

print(matrix\_3d)

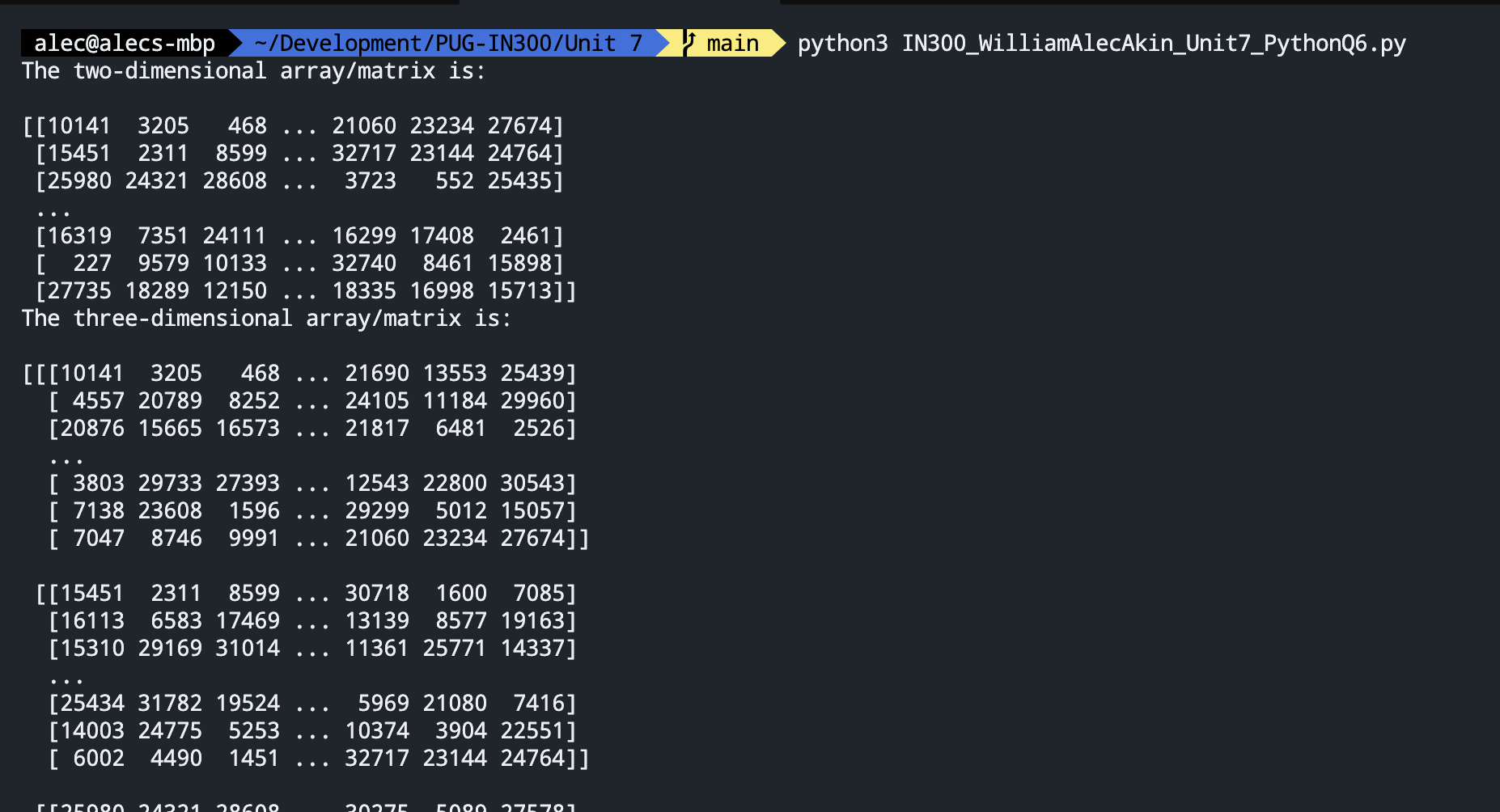
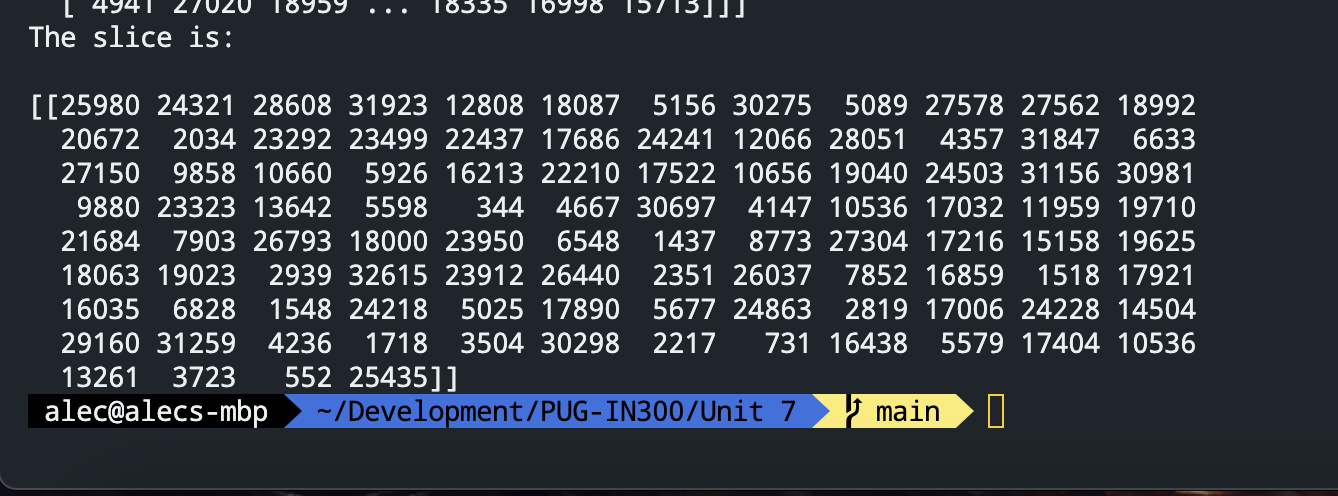
# Slice by the 2nd and 5th index and steps of 10 of the two-dimensional array/matrix

slice = matrix\_2d[2:5:10, :]

# Print the results of the slice to stdout with a message

print("The slice is: \n")

print(slice)



# #7 – Compare and Contrast Languages

In my opinion, in this assignment, R came out ready to compete and showed how simple, practical, and easy it can be to calculate ordinary statistical equations and solve problems. I still think that in my mind, Python was just a little bit superior at this task – even though R is specifically designed for this type of work. The reason that I say this is that even with Python needing to import two separate libraries to complete this task, the structure, methods, and use of both were still more straightforward for me to use than they were for R. Plus, R’s syntax and operators, specifically things like “<-“ instead of “=” make it more tedious to program in. Also, it needs to use print(paste(“print something”)) is confusing and seems unnecessary. Python remains consistent at its core and still uses nearly the same number of lines of code, so I think it wins. Java felt so far out of left field compared to the other two. It required exponentially more work and complexity, and it was far slower at computation than the other two. Given a choice, I would not use Java for this type of task in the future.