Parallel Programming

R Programming – Program1\_Stats

R Studio contains many datasets to experiment with and use in your code.

One of them is called the MASS package. To include in your script you simply add the first line:

library(MASS)

In this package, we are going to use the painter’s dataset (a compilation of technical information of a few eighteenth century classical painters.

To see the data – just type:

painters

in the console or add print(painters) to your script.

To see individual columns you can use:

print(painters$School)

The painters table is a built-in data frame so it can be accessed with the data frame functions.

The Frequency Distribution of a data variable is a summary of the data occurrence in a collection of non-overlapping categories.

So in the data set painters, the frequency distribution of the School variable is a summary of the number of painters in each school.

So if asked to find the frequency distribution of the painter schools in the data set painters you would:

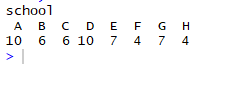
library(MASS)

school = painters$School

fq = school.freq = table(school)

print(x)

And get:



Now we can use the cbind ( ) to print the result in column format like:

library(MASS)

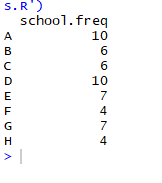
school = painters$School

fq = school.freq = table(school)

fq\_col = cbind(school.freq)

print(fq\_col)

and get:



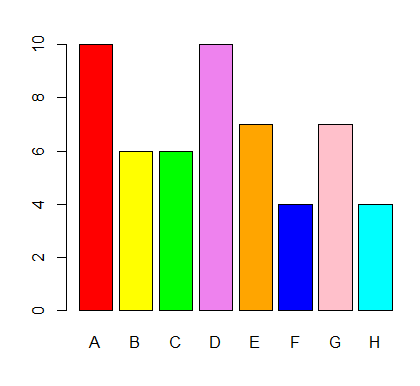
To create a bar graph of our data:

#Add coloring:

colors = c("red", "yellow", "green", "violet","orange", "blue", "pink", "cyan")

barplot(school.freq, col=colors)

We get:



Part I of the Assignment is to:

1. Write a script that finds the frequency distribution of the Expression column for our painters.
2. Then use the above code to find and print out the highest frequency of our Expression column and the Expression it represents.
3. Represent and print the Expression frequency distribution in a column format.
4. Then plot a bar graph (in color) representing the frequency distribution of the Expressions.

Submit your code and a screen shot of your code / bar graph working in R Studio here:

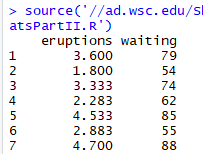
Part II of the assignment:

In this assignment, we are going to use a data frame named faithful from the same MASS library.

This data is quantitative data so it consists of numeric data that support arithmetic operations. We can now practice the R tools for quantitative data analysis techniques discussed in our stats lectures.

Let’s use the faithful data set to find the frequency distribution and more info on the waiting column of faithful’s eruptions.

Snip-it of data set:



eruptions = numeric Eruption time in mins

waiting = numeric Waiting time to next eruption (in mins)

You will use this data set to find:

* the range
* interval of time
* display frequency distribution as a table
* calculate the mean and median of the data set
* create a histogram and a box and whiskers chart (screen shots added below (comment out unused code to continue)
* add code to determine the interval that has the most waiting times of the data set.

Submit your code and a screen shot of your code / histogram / box and whiskers graph working in R Studio here:

Part III – Finally created a plot – using the plot function that shows the values in a graph with the Eruption time (min) on the x axis and the Waiting time between eruptions (min) on the y axis.

Submit copy of code and screen shot of it working here: