Day 3 – Exercises Name:

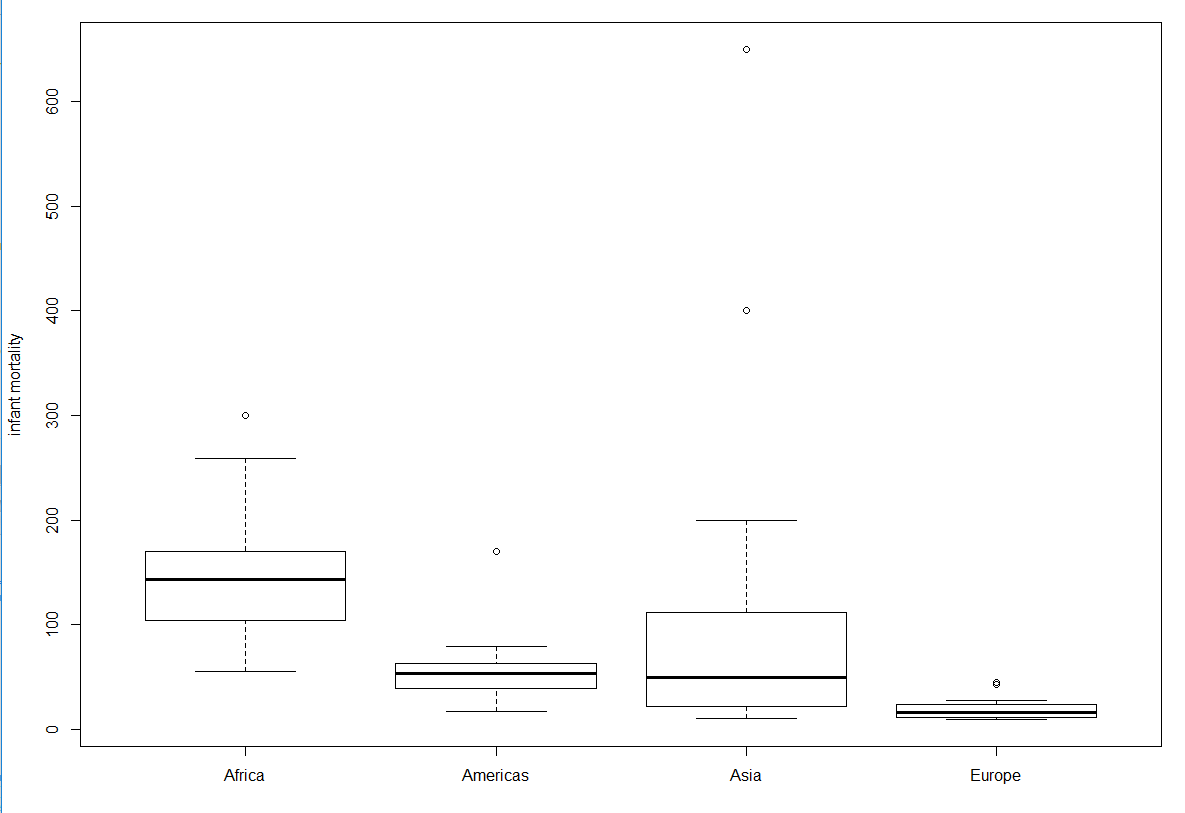
1. Import the dataset “speciesarea.csv”. This dataset contains data on land area and number of mammal species for 14 islands in Southeast Asia. Please do the following:

1. Calculate the correlation between island area (A) and the number of mammal species (M). Is the correlation significant at p=0.05 level?
2. Transform both A and S by their natural logarithms (lnA and lnM, respectively), and then calculate the correlation between lnA and lnM. Is the correlation here bigger or smaller than the one in (a)?
3. Now you also want to study if the number of mammal species is correlated with species number of other animals. So you ask for data from a Singapore colleague, who sends you a dataset ‘birds.csv’. The dataset ‘bird.csv’ contains two columns, one is island name, and one is the number of bird species (B). What you need to do is to merge the two datasets into one by the column of island name, and to calculate the correlation between M and B.

2. The dataset “infant mortality.csv” contains infant mortality data from ~100 countries around the world in 1970s. Columns include: income (per-capita income in U. S. dollars), infant (infant mortality rate per 1000 live births), region (large geographic regions each country belongs to), and oil (whether or not the country is an oil-exporting country.) For Questions 2(a) and 2(b), please also print the boxplots comparing variances of different groups.

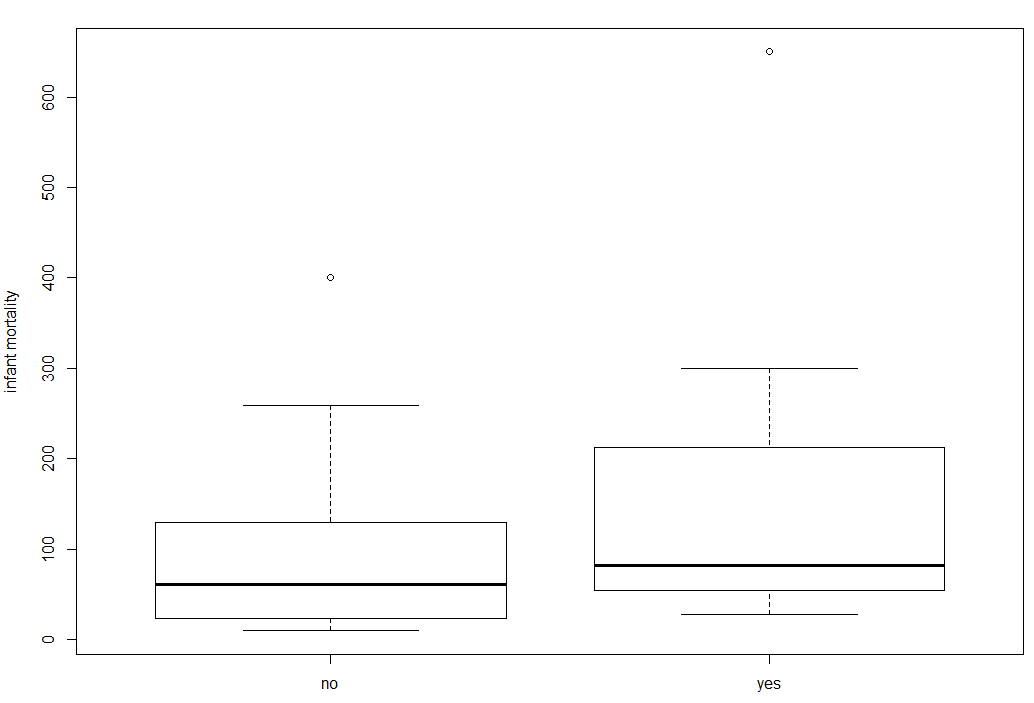
1. Is there significant difference among different regions in infant mortality rate?

Yes

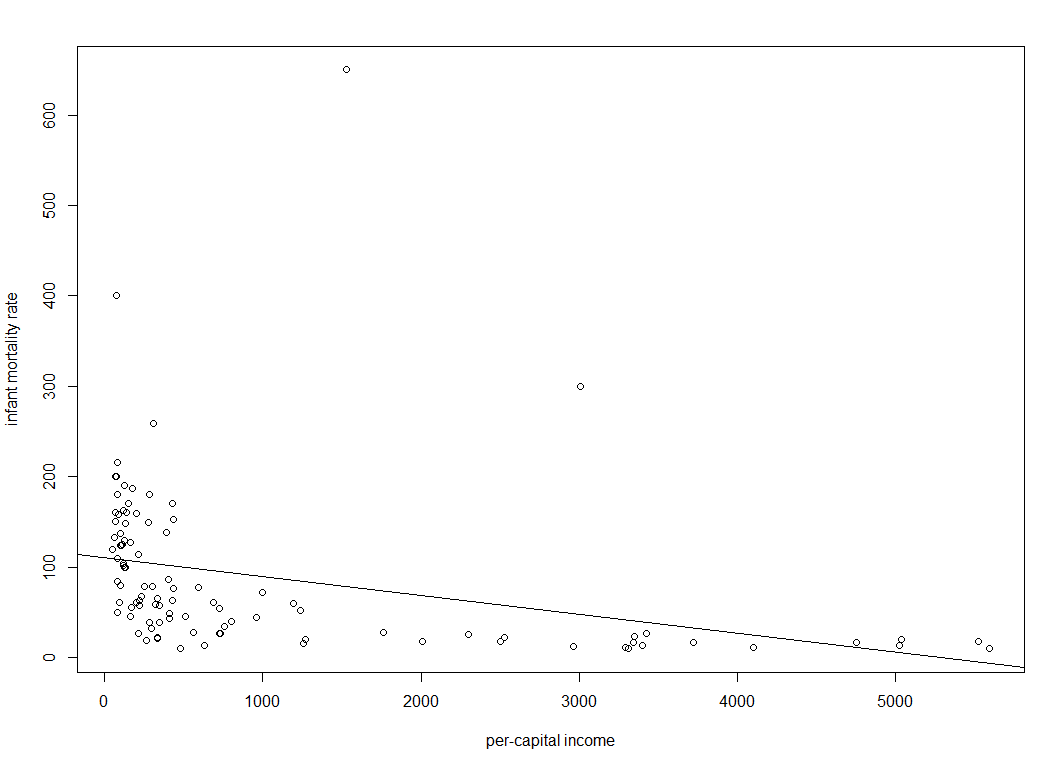


1. Is there significant difference between oil-exporting and non-oil-exporting countries in infant mortality rate?

Yes



1. What is the linear regression relationship between per-capita income (x) and infant mortality rate (y)?

y=-0.020907x+110.421086

1. The power-function regression between per-capita income and infant mortality is y~x^a. Which model better describe the regression relationship between per-capita income and infant mortality, linear or power function?

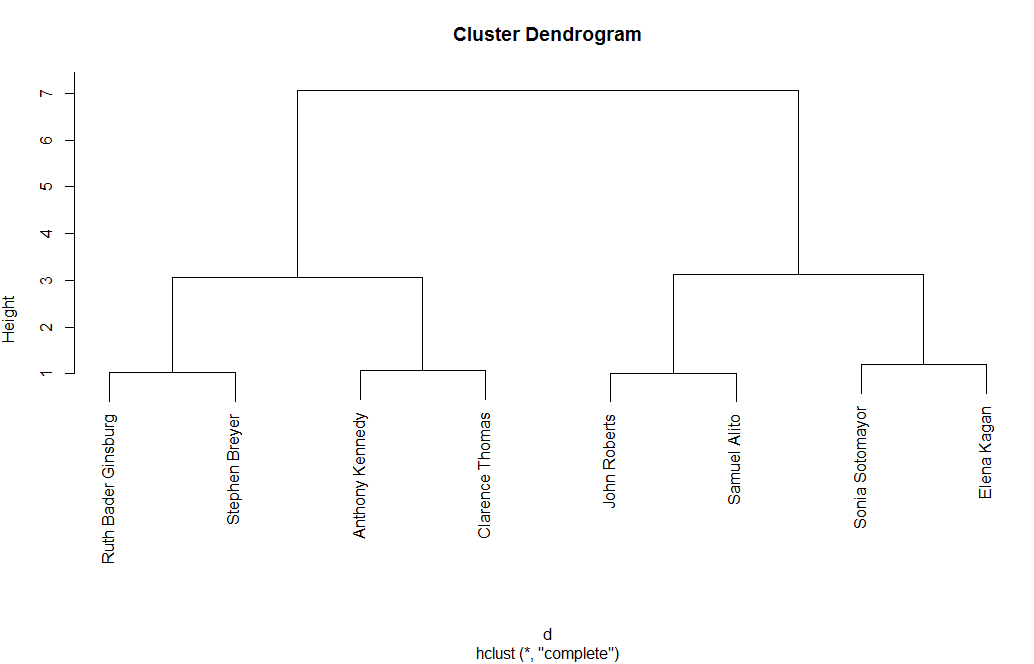
The power function regression better describes.

3. The dataset ‘justice.csv’ contains information about the leaning score on a number of issues by the current 8 justices in the US Supreme Court (not including newly-appointed Justice Gorsuch).

1. Does a justice’s leaning on First Amendment affect his/her leaning on Economic?

Yes

1. Perform a hierarchical clustering for the justices based on their scores on all the 7 issues.



4. Write a for() loop that prints all the letters in y <- c("q", "w", "e", "r", "z", "c").

y <- c("q", "w", "e", "r", "z", "c")

for (x in y) {

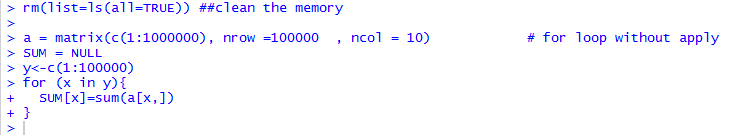
print(x) ##task statement

}

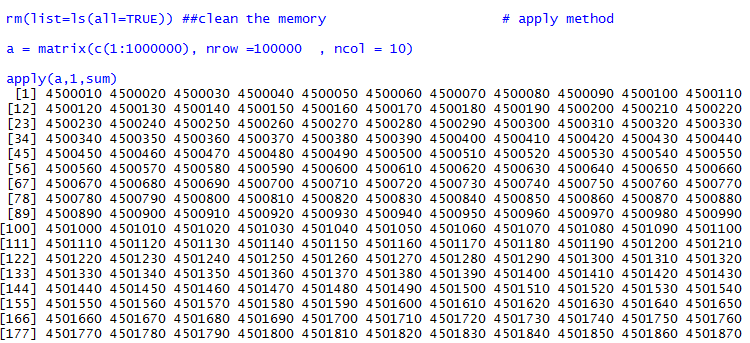
5. Create a 100,000 rows by 10 columns matrix with the numbers 1:1,000,000. Make a for-loop that calculates the sum of column values for each row (i.e. row sums) of the matrix. Verify that your results are consistent with what you obtain with the appropriate apply function.

Hint: To obtain a specific row from a matrix 🡪 matrix.name[x,]

To obtain sum of elements use sum()







**Name your files: LastName\_Exercise# (e.g. Knott\_Exercise1.docx)**

Answer each question and paste screenshots of any relevant output into this document. Submit your annotated code and answers to Blackboard by the start of class on Friday.