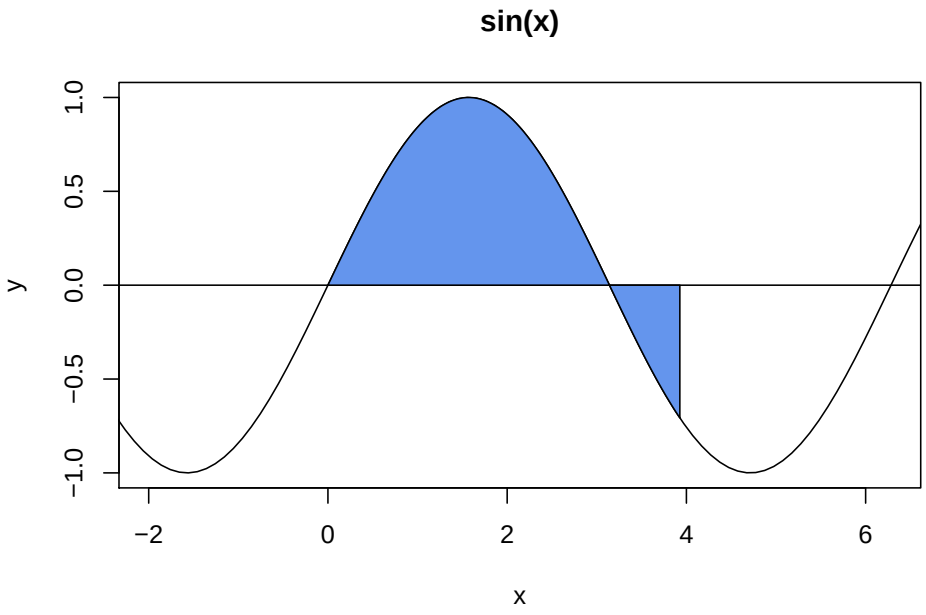
Homework 1 for R language and data analysis

Qiang Shen

Oct.1,2016

Use the knowledge learned from the course to solve the following questions and submit it to Shenwei Huang(黄玾炜,11420008@zju.edu.cn) before the midnight of Oct. 12.

1. Use R to Estimate what is the total area covered in blue of sin(x) function (from 0 to 5/4*pi). (1% deviation from the true value is acceptable. hint: in R π is "pi".)



2. If we list all the natural numbers below 10 that are multiples of 3 or 5, we get 3, 5, 6 and 9. The sum of these multiples is 23.Find the sum

of all the multiples of 3 or 5 below 1000.(hint: summation function in R is *sum*.For example, **sum(1:4) equals 10**)

3. Install package *ggplot2* and then export the build-in dataset available in the package named *presidential* to files "presidential.csv"" and "presidential.txt". Once exported, then import files "presidential.csv" back to R accordingly.(Note for each variable in the imported data file,it should have the same type as the orginal dataset *presidential* and you can use the command *identical* to check whether the two datasets are the same or not, if it's identical, then *TRUE* will be returned. Hints: use **as.Date** to convert corresponding character into date).

4. In data file named "Restaurant.dta" attached:

   1>. Find the row name/index of the observations with the missing value in variable named *mealprice*.

   2>. to generate a data file without the observations with missing value in *mealprice* and exported it as *Restaurant_new.csv*.

5. In data file named "Grades.dta" attached: Generate a new variable called *group* and classify those who have *gpa* equal or large than 3.5 as *excellent*, equal or large than 2 but smaller than 3.5 as *pass* and the rest as *fail* and then exported the a new data file called *Grades_new.csv*.