9Міністерство освіти і науки України  
ІФНТУНГ

Кафедра ІТТС

Лабораторна робота №3

НАПИСАННЯ ФУНКЦІЙ В R ДЛЯ ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ

Виконав:  
ст. гр. СІм-11  
Вірстюк В.  
Перевірила:  
Штаєр Л. О.

Івано-Франківськ  
2016

**ЛАБОРАТОРНА РОБОТА 1.3**

НАПИСАННЯ ФУНКЦІЙ В R ДЛЯ ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ

**Мета роботи**: одержати практичні навики у написанні функцій для обробки статистичних даних в R.

**Обладнання:**

- ПК ІВМ РС x86 CPU з встановленою операційною системою;

- встановлене програмне забезпечення R з оболонкою RStudio;

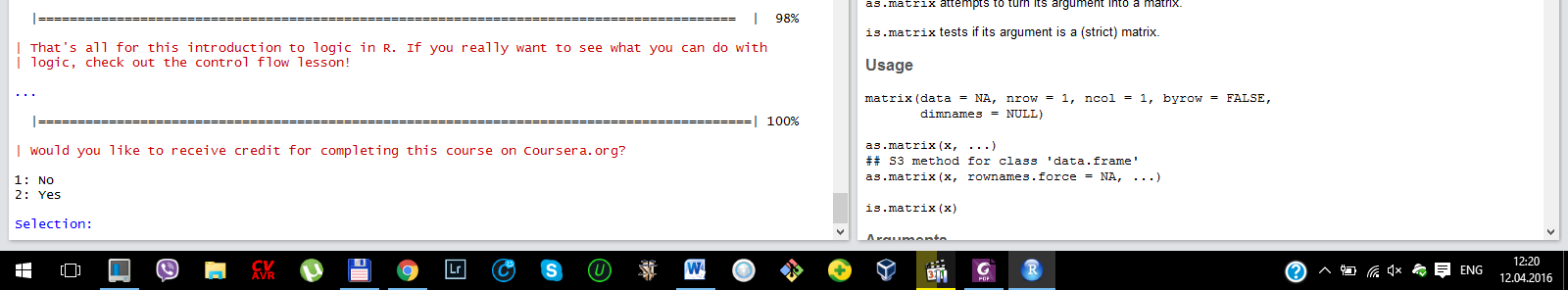
- встановлений в R пакет swirl;

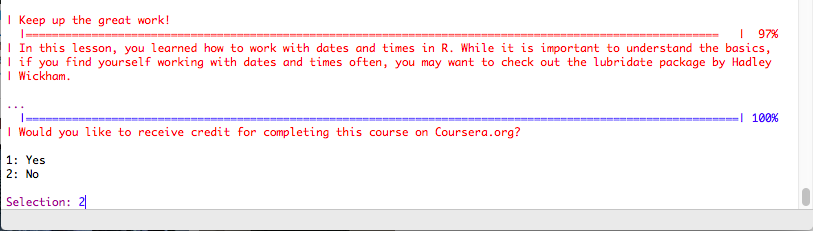
* - доступ до мережі інтернет.
* **1.3.1 Теоретичні відомості**
* Для одержання практичних навичок в написанні функцій в R виконайте наступні уроки в навчальному середовищі swirl():
* 8. Logic
* 9. Functions
* 14. Dates and Times

**Порядок виконання роботи:**

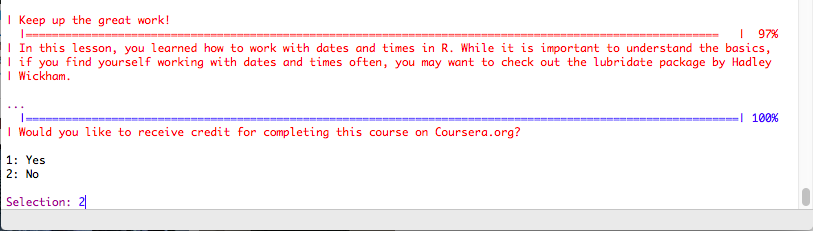
1 Виконайте уроки 8, 9 і 14 в навчальному середовищі swirl.

8

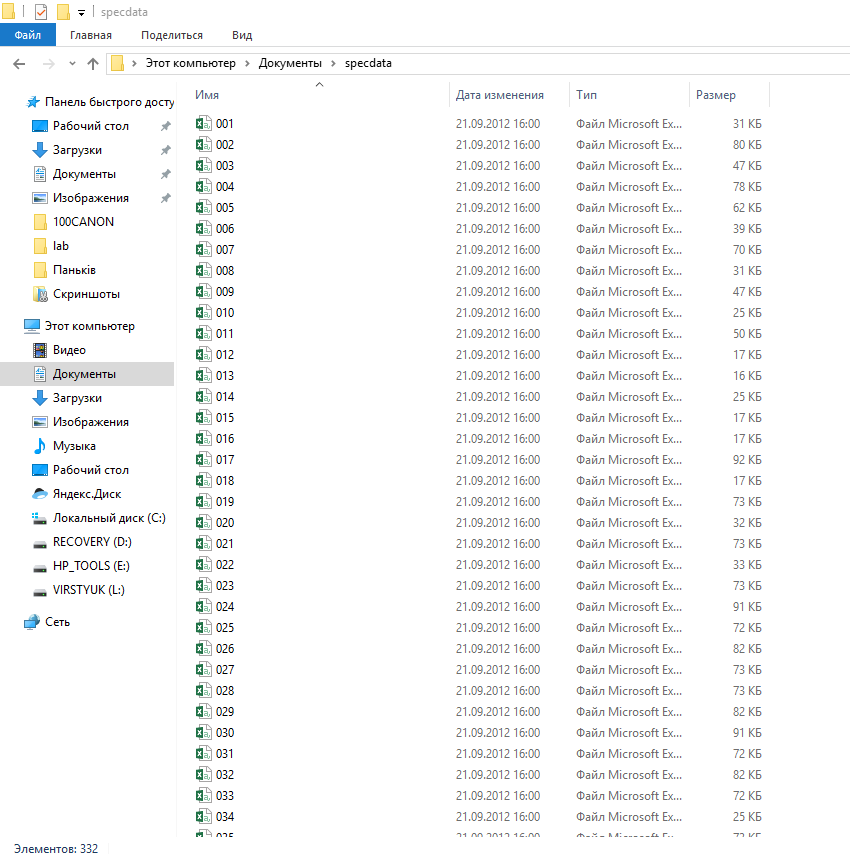


9. 

14.



2 Завантажити файл specdata.zip



3. Напишіть функцію з ім'ям 'pollutantmean', яка обчислює середнє значення для забруднюючої речовини (сульфат або нітрат) для певного переліку точок моніторингу. Функція 'pollutantmean' приймає три аргументи: “каталог”, “забруднювач” і “ідентифікатор”. Враховуючи заданий вектор точок моніторингу (їх ID), функція 'pollutantmean' читає дані про відповідний тип забруднення з каталогу, який відповідає точці забруднення і повертає середнє значення забруднення по всіх точках моніторингу, ігноруючи пропущені значення (NA).

**Функція pollutantmean.R :**

pollutantmean <- function(directory, pollutant, id = 1:332) {

setwd(file.path(getwd(), directory))

total = 0

stupidValues = 0

for (i in id)

{

if (i <10) {

data <- read.csv(paste("0","0", as.character(i), ".csv", sep=""), ## 'id' =7 -> 007.csv

header = T,

na.strings=c("NA","NaN", " "))

}

else if (i>=10 & i<100) {

data <- read.csv(paste("0", as.character(i), ".csv", sep=""), ## 'id' = 17 -> 017.csv

header = T,

na.strings=c("NA","NaN", " ")

)

}

else {

data <- read.csv(paste(as.character(i), ".csv", sep=""),

header = T,

na.strings=c("NA","NaN", " ")

)

}

data = na.omit(data) ## Handle Na

stupidValues = stupidValues + nrow(data)

if (pollutant == "sulfate") {

total = total + sum(data$sulfate) ## Commands starting with a '$' sign need to be executed from a Unix or Linux shell

}

else {

total = total + sum(data$nitrate)

}

}

setwd("..") ## reset

return (total/stupidValues) #return mean()

}

**Результат**:

> source("pollutantmean.R")

> pollutantmean("specdata", "sulfate", 1:10)

[1] 4.064128

> pollutantmean("specdata", "sulfate", 1:10)

[1] 4.064128

> pollutantmean("specdata", "nitrate", 23)

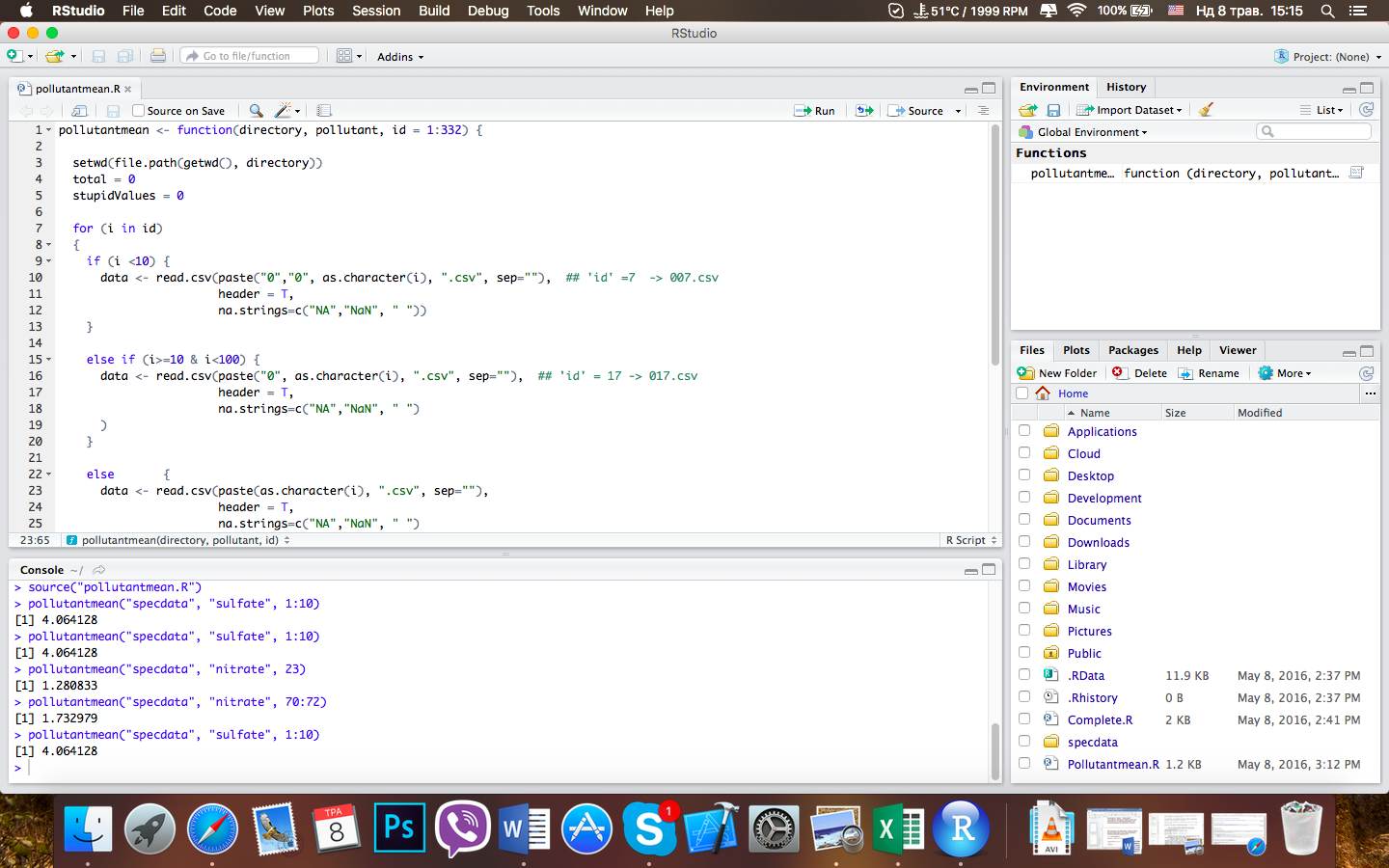
[1] 1.280833

> pollutantmean("specdata", "nitrate", 70:72)

[1] 1.732979

> pollutantmean("specdata", "sulfate", 1:10)

[1] 4.064128



4 Написати функцію, яка зчитує каталог з файлами і повідомляє про кількість повністю спостережуваних випадків в кожному файлі даних. Функція повинна повертати фрейм даних, де перший стовпець - це ім'я файлу, а другий стовпець - число повних випадків.

Функція complete.R :

complete <- function(directory, id = 1:332) {

dataframe = NULL

setwd(file.path(getwd(), directory))

for (i in id)

{

if (i <10) {

data <- read.csv(paste("0","0", as.character(i), ".csv", sep=""),

header = T,

na.strings=c("NA","NaN", " "))

}

else if (i>=10 & i<100) {

data <- read.csv(paste("0", as.character(i), ".csv", sep=""),

header = T,

na.strings=c("NA","NaN", " ")

)

}

else {

data <- read.csv(paste(as.character(i), ".csv", sep=""),

header = T,

na.strings=c("NA","NaN", " ")

)

}

data = na.omit(data)

data = as.matrix(data) ##make matrix for each successive row of our dataframe

dataframe = rbind(dataframe, c(i,nrow(data))) # fill each successive row of our dataframe

}

setwd("..")

dataframe = data.frame(dataframe) # from matix to data frame

names(dataframe) = c('id', 'nobs') # output

return (dataframe)

}

**Результат:**

> source("complete.R")

> complete("specdata", 1)

id nobs

1 1 117

> complete("specdata", c(2, 4, 8, 10, 12))

id nobs

1 2 1041

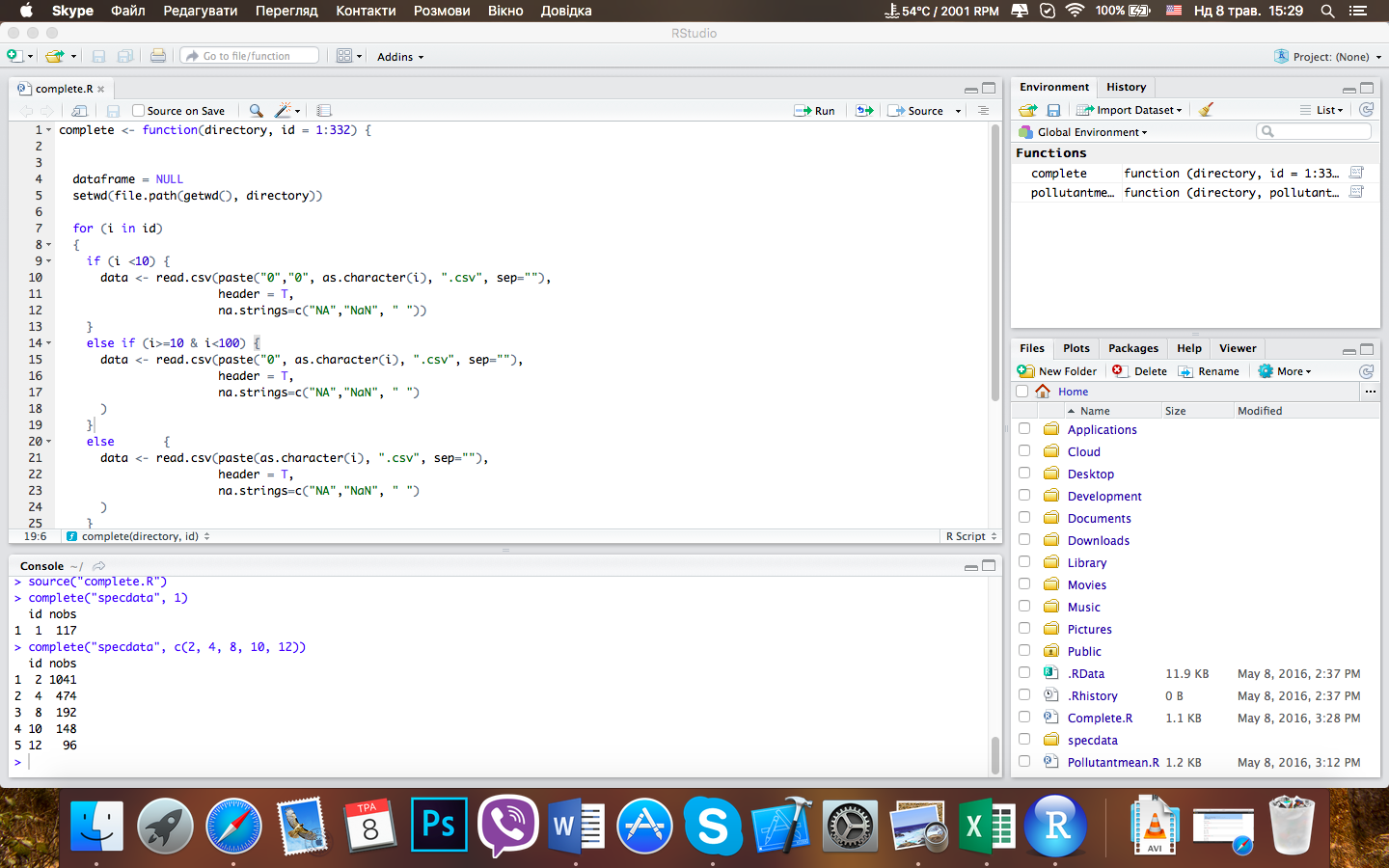
2 4 474

3 8 192

4 10 148

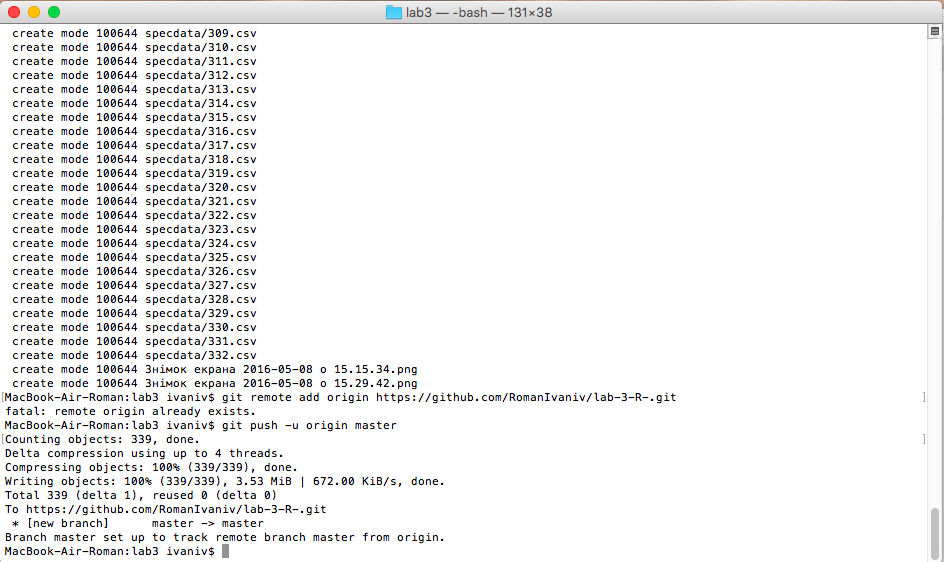
5 12 96

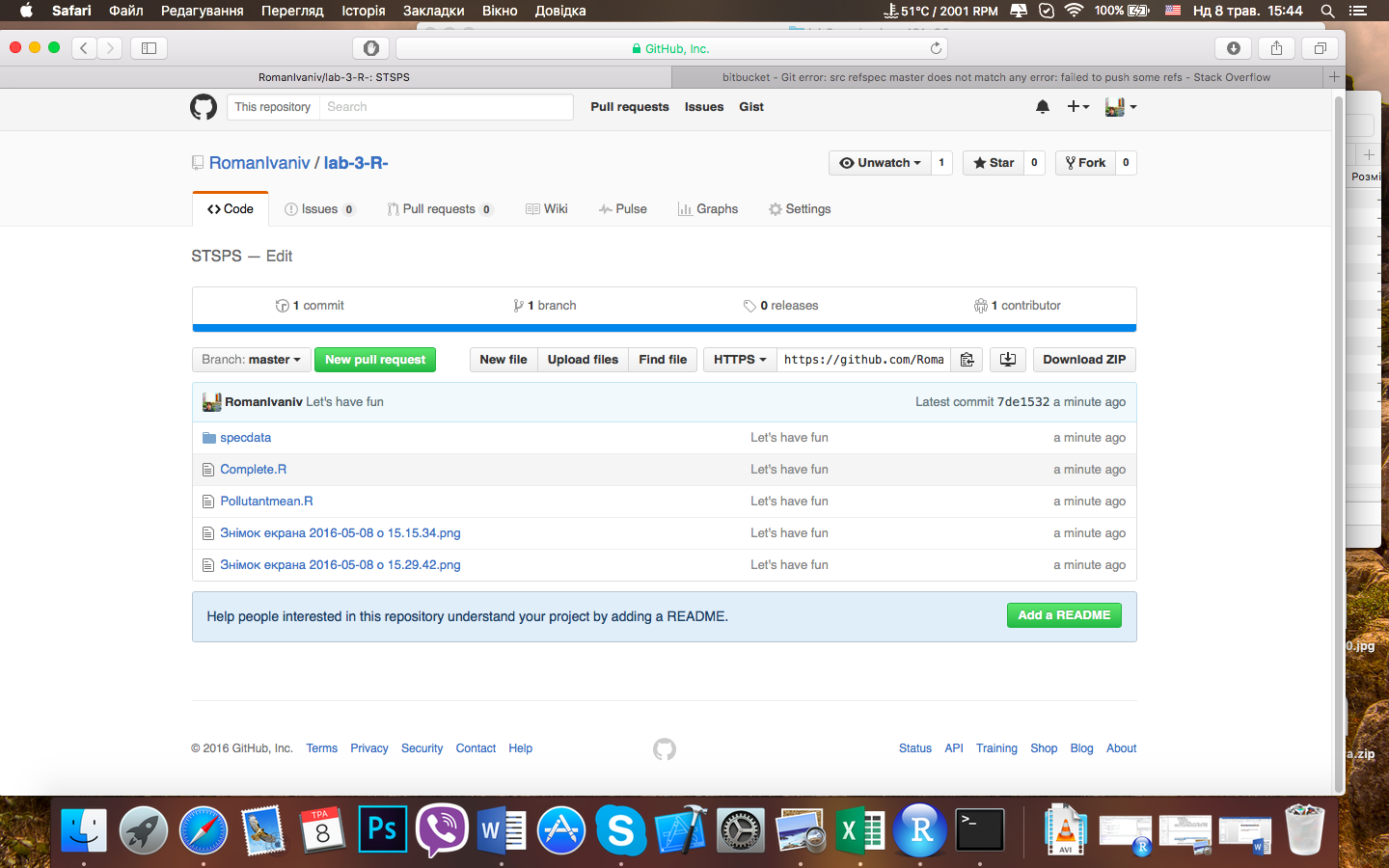
>



5**. Копії екранів з відображенням вмісту створеного репозиторію на GitHub.**

https://github.com/RomanIvaniv/lab-3-R-





**Висновок**: під час виконання даної роботи я засвоїв знання по роботі з функціями в R.