소프트웨어 융합 최신기술 레포트

소프트웨어학과

20103308 김성곤

1. 웹크롤링(R 코드)

library(rvest)

library(dplyr)

url1 <- "http://www.kma.go.kr/weather/climate/past\_table.jsp?stn=108&yy=2011&obs=12&x=26&y=7"

url2 <- "http://www.kma.go.kr/weather/climate/past\_table.jsp?stn=108&yy=2012&obs=12&x=26&y=7"

url3 <- "http://www.kma.go.kr/weather/climate/past\_table.jsp?stn=108&yy=2013&obs=12&x=26&y=7"

url4 <- "http://www.kma.go.kr/weather/climate/past\_table.jsp?stn=108&yy=2014&obs=12&x=26&y=7"

url5 <- "http://www.kma.go.kr/weather/climate/past\_table.jsp?stn=108&yy=2015&obs=12&x=26&y=7"

data1 <- read\_html(url1,encoding="EUC-KR")

data2 <- read\_html(url2,encoding="EUC-KR")

data3 <- read\_html(url3,encoding="EUC-KR")

data4 <- read\_html(url4,encoding="EUC-KR")

data5 <- read\_html(url5,encoding="EUC-KR")

dataTable1 <- html\_table(html\_nodes(data1,"table")[[2]], fill=TRUE)

dataTable2 <- html\_table(html\_nodes(data2,"table")[[2]], fill=TRUE)

dataTable3 <- html\_table(html\_nodes(data3,"table")[[2]], fill=TRUE)

dataTable4 <- html\_table(html\_nodes(data4,"table")[[2]], fill=TRUE)

dataTable5 <- html\_table(html\_nodes(data5,"table")[[2]], fill=TRUE)

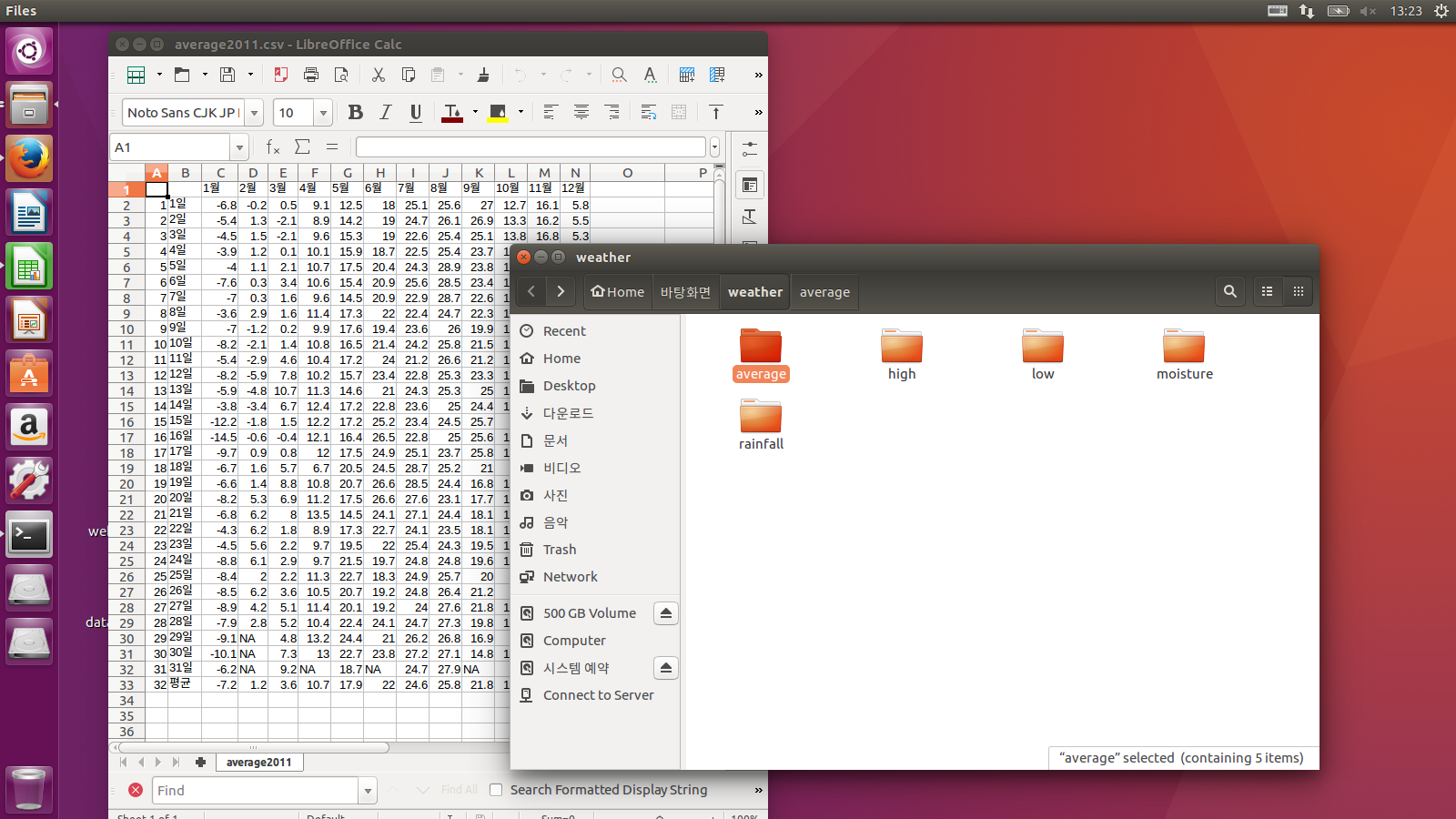
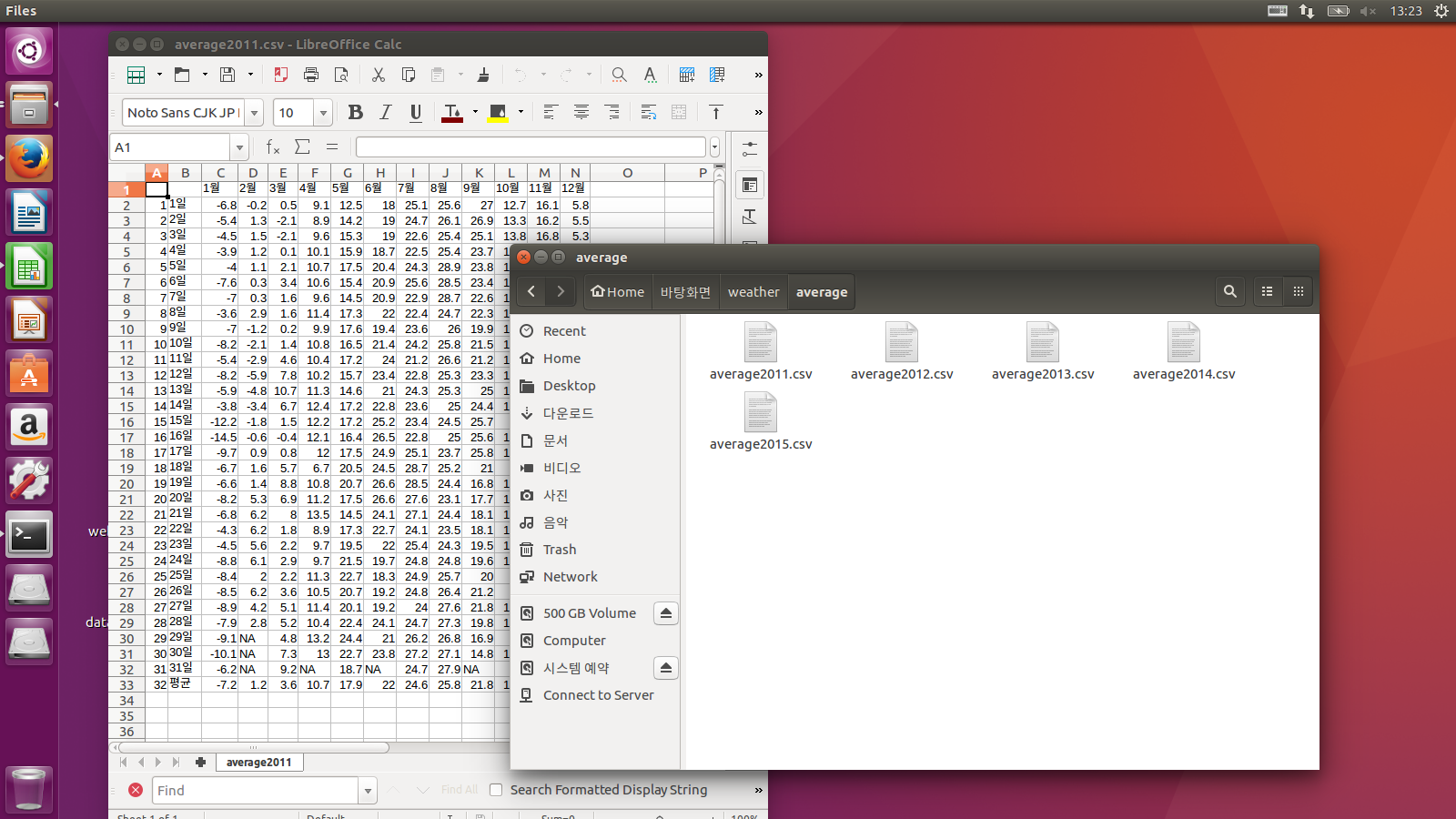
write.csv(dataTable1, "/home/kimsunggon/바탕화면/weather/moisture2011.csv")

write.csv(dataTable2, "/home/kimsunggon/바탕화면/weather/moisture2012.csv")

write.csv(dataTable3, "/home/kimsunggon/바탕화면/weather/moisture2013.csv")

write.csv(dataTable4, "/home/kimsunggon/바탕화면/weather/moisture2014.csv")

write.csv(dataTable5, "/home/kimsunggon/바탕화면/weather/moisture2015.csv")



1. 데이터 가공(데이터처리)(1)
2. 결측값처리 : 대부분 엑셀에서 직접 처리(노가다).
3. 잡음완화 : 처음에는 1000이상을 엑셀 함수로 처리할까 하다가 단순히 1000이상이 넘는 것을 다 잡음처리 하면 문제가 생길것으로 예상하여 해당일의 하루 후나 하루 전의 모기수와 1000이상 차이가 나고 해당일의 하루 후나 하루 전의 모기수가 1000미만일 때에는 1000으로 처리하였고 나머지 경우에는 그대로 데이터 사용함.
4. 이상점분석 : 평균값에 올림을 하여 계산, 엑셀로 if문을 사용하여 처리.
5. 데이터 가공(데이터 결합)(2)

R을 이용하여 데이터 통합 (R 코드)

library(rvest)

library(dplyr)

data1 <- read.csv("/media/kimsunggon/4480EF3B80EF325A/Users/KimSunggon/Desktop/수업/sw융합최신기술/mosq1.csv", stringsAsFactors=FALSE)

high2011 <- read.csv("/home/kimsunggon/바탕화면/weather/high/high2011.csv", header=F, stringsAsFactors=FALSE)

high2012 <- read.csv("/home/kimsunggon/바탕화면/weather/high/high2012.csv", header=F, stringsAsFactors=FALSE)

high2013 <- read.csv("/home/kimsunggon/바탕화면/weather/high/high2013.csv", header=F, stringsAsFactors=FALSE)

high2014 <- read.csv("/home/kimsunggon/바탕화면/weather/high/high2014.csv", header=F, stringsAsFactors=FALSE)

high2015 <- read.csv("/home/kimsunggon/바탕화면/weather/high/high2015.csv", header=F, stringsAsFactors=FALSE)

low2011 <- read.csv("/home/kimsunggon/바탕화면/weather/low/low2011.csv", header=F, stringsAsFactors=FALSE)

low2012 <- read.csv("/home/kimsunggon/바탕화면/weather/low/low2012.csv", header=F, stringsAsFactors=FALSE)

low2013 <- read.csv("/home/kimsunggon/바탕화면/weather/low/low2013.csv", header=F, stringsAsFactors=FALSE)

low2014 <- read.csv("/home/kimsunggon/바탕화면/weather/low/low2014.csv", header=F, stringsAsFactors=FALSE)

low2015 <- read.csv("/home/kimsunggon/바탕화면/weather/low/low2015.csv", header=F, stringsAsFactors=FALSE)

average2011 <- read.csv("/home/kimsunggon/바탕화면/weather/average/average2011.csv", header=F, stringsAsFactors=FALSE)

average2012 <- read.csv("/home/kimsunggon/바탕화면/weather/average/average2012.csv", header=F, stringsAsFactors=FALSE)

average2013 <- read.csv("/home/kimsunggon/바탕화면/weather/average/average2013.csv", header=F, stringsAsFactors=FALSE)

average2014 <- read.csv("/home/kimsunggon/바탕화면/weather/average/average2014.csv", header=F, stringsAsFactors=FALSE)

average2015 <- read.csv("/home/kimsunggon/바탕화면/weather/average/average2015.csv", header=F, stringsAsFactors=FALSE)

moisture2011 <- read.csv("/home/kimsunggon/바탕화면/weather/moisture/moisture2011.csv", header=F, stringsAsFactors=FALSE)

moisture2012 <- read.csv("/home/kimsunggon/바탕화면/weather/moisture/moisture2012.csv", header=F, stringsAsFactors=FALSE)

moisture2013 <- read.csv("/home/kimsunggon/바탕화면/weather/moisture/moisture2013.csv", header=F, stringsAsFactors=FALSE)

moisture2014 <- read.csv("/home/kimsunggon/바탕화면/weather/moisture/moisture2014.csv", header=F, stringsAsFactors=FALSE)

moisture2015 <- read.csv("/home/kimsunggon/바탕화면/weather/moisture/moisture2015.csv", header=F, stringsAsFactors=FALSE)

rainfall2011 <- read.csv("/home/kimsunggon/바탕화면/weather/rainfall/rainfall2011.csv", header=F, stringsAsFactors=FALSE)

rainfall2012 <- read.csv("/home/kimsunggon/바탕화면/weather/rainfall/rainfall2012.csv", header=F, stringsAsFactors=FALSE)

rainfall2013 <- read.csv("/home/kimsunggon/바탕화면/weather/rainfall/rainfall2013.csv", header=F, stringsAsFactors=FALSE)

rainfall2014 <- read.csv("/home/kimsunggon/바탕화면/weather/rainfall/rainfall2014.csv", header=F, stringsAsFactors=FALSE)

rainfall2015 <- read.csv("/home/kimsunggon/바탕화면/weather/rainfall/rainfall2015.csv", header=F, stringsAsFactors=FALSE)

for(i in 1:9838){

tempString <- data1[i,1]

as.POSIXlt(tempString) -> tempDate

tYear <- tempDate$year + 1900

tMonth <- tempDate$mon + 3

tDay <- tempDate$mday + 1

high <- switch(tYear - 2010, high2011, high2012, high2013, high2014, high2015)

data1[i,5] <- high[tDay, tMonth]

low <- switch(tYear - 2010, low2011, low2012, low2013, low2014, low2015)

data1[i,6] <- low[tDay, tMonth]

average <- switch(tYear - 2010, average2011, average2012, average2013, average2014, average2015)

data1[i,7] <- average[tDay, tMonth]

moisture <- switch(tYear - 2010, moisture2011, moisture2012, moisture2013, moisture2014, moisture2015)

data1[i,8] <- moisture[tDay, tMonth]

rainfall <- switch(tYear - 2010, rainfall2011, rainfall2012, rainfall2013, rainfall2014, rainfall2015)

data1[i,9] <- rainfall[tDay, tMonth]

}

write.csv(data1, "/media/kimsunggon/4480EF3B80EF325A/Users/KimSunggon/Desktop/수업/sw융합최신기술/mosq1\_new.csv")