Lab Assignment: Chapter 14 - Data Reshaping

Michael Beebe

library(stringr)  
library(plyr)  
library(reshape2)  
library(ggplot2)  
library(scales)

## 1. cbind and rbind

Create 3 vectors: start, finish, and numeric

* The start vector has the first 10 capital letters of the alphabet
* The finish vector has the last 10 capital letters of the alphabet
* The numeric vector has the numbers 1-10

start <- c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J")  
finish <- c("Q", "R", "S", "T", "U", "V", "W", "X", "Y", "Z")  
numeric <- c(1:10)

Use cbind to bind these vectors into table1 by column

table1 <- cbind(start, finish, numeric)  
table1

## start finish numeric  
## [1,] "A" "Q" "1"   
## [2,] "B" "R" "2"   
## [3,] "C" "S" "3"   
## [4,] "D" "T" "4"   
## [5,] "E" "U" "5"   
## [6,] "F" "V" "6"   
## [7,] "G" "W" "7"   
## [8,] "H" "X" "8"   
## [9,] "I" "Y" "9"   
## [10,] "J" "Z" "10"

Use rbind to bind these vectors into table2 by row

table2 <- rbind(start, finish, numeric)  
table2

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]  
## start "A" "B" "C" "D" "E" "F" "G" "H" "I" "J"   
## finish "Q" "R" "S" "T" "U" "V" "W" "X" "Y" "Z"   
## numeric "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"

Create 2 CSV files - data1.csv and data2.csv

* data1.csvshould have the column titles “Course” and “Title”  
  – List 7 courses and titles in the Data Analytics Degree
* data2.csv should have the column titles “Credits” and “Semester”  
  – List the credit hours and the semester you should take the courses in data1.csv

# DFs THAT WE WILL USE TO MAKE THE CSVs  
data1 <- data.frame(  
 Course = c("ITCS 1010", "ITDB 1300", "ITDB 1400", "ITDB 1405", "ITCS 1880", "ITON 1205", "ITCS 1870"),  
 Title = c("Programming Logic", "Intro to Database Theory", "Intro to SQL", "Oracle PL/SQL Programming", "R Programming", "Network+ and Networking Essentials", "Python Programming")  
)  
  
data2 <- data.frame(  
 Credits = c(3, 3, 2, 2, 3, 3, 3),  
 Semester = c(2, 2, 2, 2, 3, 3, 4)  
)  
  
  
# MAKING CSVs FROM DF  
data1.csv <- write.csv(  
 x = data1,  
 file = "./Part1CSVs/data1.csv"  
)  
  
data2.csv <- write.csv(  
 x = data2,  
 file = "./Part1CSVs/data2.csv"  
)

Import and use cbind to bind these 2 datasets together.

data3.csv <- write.csv(  
 x = cbind(data1, data2),  
 file = "./Part1CSVs/data3.csv"  
)

## 2. Joins

Unzip the US\_Foreign\_Aid.zip file into a directory on your computer

foreign\_aid <- unzip("../../Data/US\_Foreign\_Aid.zip", exdir = "./US\_Foreign\_Aid")  
foreign\_aid

## [1] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_40s.csv"  
## [2] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_50s.csv"  
## [3] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_60s.csv"  
## [4] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_70s.csv"  
## [5] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_80s.csv"  
## [6] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_90s.csv"  
## [7] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_00s.csv"  
## [8] "./US\_Foreign\_Aid/US\_Foreign\_Aid\_10s.csv"

aid\_data <- dir("US\_Foreign\_Aid/", pattern = "\\.csv")  
  
for (i in aid\_data) {  
 us\_foreign\_aid <- str\_sub(string = i, start=12, end=18)  
 temp <- read.table(file = file.path("US\_Foreign\_Aid/", i),  
 header = TRUE, sep = ",", stringsAsFactors = FALSE)  
 assign(x=us\_foreign\_aid, value = temp)  
}

Merge the data from the 80s and 90s and display to the console

aid80s90s <- merge(  
 x = Aid\_80s,  
 y = Aid\_90s,  
 by.x = c("Country.Name", "Program.Name"),  
 by.y = c("Country.Name", "Program.Name")  
)  
  
head(aid80s90s)

## Country.Name Program.Name FY1980 FY1981  
## 1 Afghanistan Child Survival and Health NA NA  
## 2 Afghanistan Department of Defense Security Assistance NA NA  
## 3 Afghanistan Development Assistance NA NA  
## 4 Afghanistan Economic Support Fund/Security Support Assistance NA NA  
## 5 Afghanistan Food For Education NA NA  
## 6 Afghanistan Global Health and Child Survival NA NA  
## FY1982 FY1983 FY1984 FY1985 FY1986 FY1987 FY1988 FY1989 FY1990 FY1991 FY1992  
## 1 NA NA NA NA NA NA NA NA NA NA NA  
## 2 NA NA NA NA NA NA NA NA NA NA NA  
## 3 NA NA NA NA NA NA NA NA NA NA NA  
## 4 NA NA NA NA NA NA NA NA NA NA NA  
## 5 NA NA NA NA NA NA NA NA NA NA NA  
## 6 NA NA NA NA NA NA NA NA NA NA NA  
## FY1993 FY1994 FY1995 FY1996 FY1997 FY1998 FY1999  
## 1 NA NA NA NA NA NA NA  
## 2 NA NA NA NA NA NA NA  
## 3 NA NA NA NA NA NA NA  
## 4 14178135 2769948 NA NA NA NA NA  
## 5 NA NA NA NA NA NA NA  
## 6 NA NA NA NA NA NA NA

Merge the data from the 90s and 00s and display to the console

aid90s00s <- merge(  
 x = Aid\_90s,  
 y = Aid\_00s,  
 by.x = c("Country.Name", "Program.Name"),  
 by.y = c("Country.Name", "Program.Name")  
)  
  
head(aid90s00s)

## Country.Name Program.Name FY1990 FY1991  
## 1 Afghanistan Child Survival and Health NA NA  
## 2 Afghanistan Department of Defense Security Assistance NA NA  
## 3 Afghanistan Development Assistance NA NA  
## 4 Afghanistan Economic Support Fund/Security Support Assistance NA NA  
## 5 Afghanistan Food For Education NA NA  
## 6 Afghanistan Global Health and Child Survival NA NA  
## FY1992 FY1993 FY1994 FY1995 FY1996 FY1997 FY1998 FY1999 FY2000 FY2001  
## 1 NA NA NA NA NA NA NA NA NA NA  
## 2 NA NA NA NA NA NA NA NA NA NA  
## 3 NA NA NA NA NA NA NA NA NA 4110478  
## 4 NA 14178135 2769948 NA NA NA NA NA NA 61144  
## 5 NA NA NA NA NA NA NA NA NA NA  
## 6 NA NA NA NA NA NA NA NA NA NA  
## FY2002 FY2003 FY2004 FY2005 FY2006 FY2007 FY2008  
## 1 2586555 56501189 40215304 39817970 40856382 72527069 28397435  
## 2 2964313 NA 45635526 151334908 230501318 214505892 495539084  
## 3 8762080 54538965 180539337 193598227 212648440 173134034 150529862  
## 4 31827014 341306822 1025522037 1157530168 1357750249 1266653993 1400237791  
## 5 NA 3957312 2610006 3254408 386891 NA NA  
## 6 NA NA NA NA NA NA 63064912  
## FY2009  
## 1 NA  
## 2 552524990  
## 3 3675202  
## 4 1418688520  
## 5 NA  
## 6 1764252

Use the plyr package and join function to do the same for both sets of data

aid80s90s\_join <- join(  
 x = Aid\_80s,  
 y = Aid\_90s,  
 by = c("Country.Name", "Program.Name")  
)  
  
aid90s00s\_join <- join(  
 x = Aid\_90s,  
 y = Aid\_00s,  
 by = c("Country.Name", "Program.Name")  
)

Combine the 8 foreign aid data frames into 1 data frame by putting them in a list and joining toghether

* Test your list with different types of subscripts

df\_names <-str\_sub(string = aid\_data, start = 12, end = 18)  
df\_list <- vector("list", length(df\_names))  
names(df\_list) <- df\_names  
for (i in df\_names) {  
 df\_list[[i]] <- eval(parse(text = i))  
}  
  
all\_aid <- Reduce(function(...) {  
 join(..., by = c("Country.Name", "Program.Name"))  
}, df\_list)  
  
dim(all\_aid)

## [1] 2453 67

## 3. reshape2

Use melt so that each row represents a Country-Program-Year entry with Dollars stored in a column

aid\_melt <- melt(  
 data = all\_aid,  
 id.vars = c("Country.Name", "Program.Name"),  
 variable.name = "Year",  
 value.name = "Dollars"  
)  
  
aid\_melt$Year <- as.numeric(str\_sub(aid\_melt$Year, start = 3, end = 6))  
  
tail(aid\_melt)

## Country.Name Program.Name Year Dollars  
## 159440 Zimbabwe Other Food Aid Programs 1999 1645691  
## 159441 Zimbabwe Other State Assistance 1999 191624  
## 159442 Zimbabwe Other USAID Assistance 1999 155788  
## 159443 Zimbabwe Peace Corps 1999 2075542  
## 159444 Zimbabwe Title I 1999 12764713  
## 159445 Zimbabwe Title II 1999 NA

Aggregate and plot the data

melt\_agg <- aggregate(  
 Dollars ~ Program.Name + Year,  
 data = aid\_melt,  
 FUN = sum,  
 na.rm = TRUE  
)  
  
melt\_agg$Program.Name <- str\_sub(melt\_agg$Program.Name, start = 1, end = 10)  
  
ggplot(melt\_agg, aes(x=Year, y=Dollars)) +  
 geom\_line(aes(group=Program.Name))

