cuny_607_week2

Valerie Briot

February 7, 2016

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

This week assignment is to illustrate data manipulation.

1. retrieve a data set from UCI Machine Learning repository, Data Set = Flags

```
flags <- read.csv("http://archive.ics.uci.edu/ml/machine-learning-databases/flags/flag.data", header = F
```

2. Add Column Names The data set has no header, hence we will append column headers, column headers will be assigned based on the metadata information in https://archive.ics.uci.edu/ml/machine-learning-databases/flags/flag.names.txt

```
names(flags) <- c("name", "landmass", "zone", "area", "population", "language", "religion", "bars", "str
                  "red", "green", "blue", "gold", "black", "white", "orange", "mainhue", "circles", "cr
                 "quarter", "sunstars", "crescent", "triangle", "icon", "animate", "text", "topleft",
```

3. Subset Rows and Columns

we will now show the first few rows of the data set

```
head(flags)
```

```
##             name landmass zone area population language religion bars
## 1    Afghanistan        5    1  648         16       10        2    0
## 2        Albania        3    1   29          3        6        6    0
## 3        Algeria        4    1 2388         20        8        2    2
## 4 American-Samoa        6    3    0          0        1        1    0
## 5        Andorra        3    1    0          0        6        0    3
## 6         Angola        4    2 1247          7       10        5    0
##   stripes colours red green blue gold black white orange mainhue circles
## 1       3       5   1     1    0    1     1     1      0   green       0
## 2       0       3   1     0    0    1     0     1      0     red       0
## 3       0       3   1     1    0    0     1     0      0   green       0
## 4       0       5   1     0    1    1     1     0      1    blue       0
## 5       0       3   1     0    1    1     0     0      0    gold       0
## 6       2       3   1     0    0    1     0     1      0     red       0
##   crosses saltires quarter sunstars crescent triangle icon animate text
## 1       0        0       0        1        0        0    1       0    0
## 2       0        0       0        1        0        0    0       1    0
## 3       0        0       0        1        1        0    0       0    0
## 4       0        0       0        0        0        1    1       1    0
```

```
## 5       0        0       0        0        0        0    0       0    0
## 6       0        0       0        1        0        0    1       0    0
##   topleft botright
## 1   black    green
## 2     red      red
## 3   green    white
## 4    blue      red
## 5    blue      red
## 6     red    black
```

we will only consider rows where landmass = N. America (1) or Europe (3)

```
myflags <- subset(flags, landmass == "1" | landmass == "3")
```

We will only consider the columns name, landmass, and bars through text.

```
myflags_1 <- subset(myflags, select = name:landmass)
myflags_2 <- subset(myflags, select = bars:text)
myflags_new <- cbind(myflags_1, myflags_2)
```

4. Map value of landmass as follows 1 : N_America and 3 : Europe

```
require(plyr)
```

```
## Loading required package: plyr
```

```
myflags_new$landmass <- mapvalues(myflags_new$landmass,
                                  from=c("1","3"),
                                  to=c("N_America","Europe"))
```

5. Display transformed data set

```
myflags_new
```

```
##                     name  landmass bars stripes colours red green blue
## 2                Albania    Europe    0       0       3   1     0    0
## 5                Andorra    Europe    3       0       3   1     0    1
## 7               Anguilla N_America    0       1       3   0     0    1
## 8       Antigua-Barbuda N_America    0       1       5   1     0    1
## 12               Austria    Europe    0       3       2   1     0    0
## 13               Bahamas N_America    0       3       3   0     0    1
## 16              Barbados N_America    3       0       3   0     0    1
## 17               Belgium    Europe    3       0       3   1     0    0
## 18                Belize N_America    0       2       8   1     1    1
## 20               Bermuda N_America    0       0       6   1     1    1
## 25  British-Virgin-Isles N_America    0       0       6   1     1    1
## 27              Bulgaria    Europe    0       3       5   1     1    1
## 32                Canada N_America    2       0       2   1     0    0
## 34       Cayman-Islands N_America    0       0       6   1     1    1
## 43           Costa-Rica N_America    0       5       3   1     0    1
```

```
## 44                  Cuba N_America    0       5       3   1     0    1
## 45                Cyprus    Europe    0       0       3   0     1    0
## 46          Czechoslovakia    Europe    0       0       3   1     0    1
## 47                Denmark    Europe    0       0       2   1     0    0
## 49               Dominica N_America    0       0       6   1     1    1
## 50    Dominican-Republic N_America    0       0       3   1     0    1
## 53            El-Salvador N_America    0       3       2   0     0    1
## 56                Faeroes    Europe    0       0       3   1     0    1
## 59                Finland    Europe    0       0       2   0     0    1
## 60                 France    Europe    3       0       3   1     0    1
## 65            Germany-DDR    Europe    0       3       3   1     0    0
## 66            Germany-FRG    Europe    0       3       3   1     0    0
## 68              Gibraltar    Europe    0       1       3   1     0    0
## 69                 Greece    Europe    0       9       2   0     0    1
## 70              Greenland N_America    0       0       2   1     0    0
## 71                Grenada N_America    0       0       3   1     1    0
## 73              Guatemala N_America    3       0       2   0     0    1
## 77                  Haiti N_America    2       0       2   1     0    0
## 78               Honduras N_America    0       3       2   0     0    1
## 80                Hungary    Europe    0       3       3   1     1    0
## 81                Iceland    Europe    0       0       3   1     0    1
## 86                Ireland    Europe    3       0       3   0     1    0
## 88                  Italy    Europe    3       0       3   1     1    0
## 90                Jamaica N_America    0       0       3   0     1    0
## 102         Liechtenstein    Europe    0       2       3   1     0    1
## 103            Luxembourg    Europe    0       3       3   1     0    1
## 109                 Malta    Europe    2       0       3   1     0    0
## 113                Mexico N_America    3       0       4   1     1    0
## 115                Monaco    Europe    0       2       2   1     0    0
## 117            Montserrat N_America    0       0       7   1     1    1
## 122           Netherlands    Europe    0       3       3   1     0    1
## 123 Netherlands-Antilles N_America    0       1       3   1     0    1
## 125             Nicaragua N_America    0       3       2   0     0    1
## 131                Norway    Europe    0       0       3   1     0    1
## 139                Poland    Europe    0       2       2   1     0    0
## 140              Portugal    Europe    0       0       5   1     1    1
## 141           Puerto-Rico N_America    0       5       3   1     0    1
## 143               Romania    Europe    3       0       7   1     1    1
## 145            San-Marino    Europe    0       2       2   0     0    1
## 157                 Spain    Europe    0       3       2   1     0    0
## 160        St-Kitts-Nevis N_America    0       0       5   1     1    0
## 161             St-Lucia N_America    0       0       4   0     0    1
## 162            St-Vincent N_America    5       0       4   0     1    1
## 166                Sweden    Europe    0       0       2   0     0    1
## 167           Switzerland    Europe    0       0       2   1     0    0
## 177  Turks-Cocos-Islands N_America    0       0       6   1     1    1
## 181                   UK    Europe    0       0       3   1     0    1
## 183      US-Virgin-Isles N_America    0       0       6   1     1    1
## 184                  USA N_America    0      13       3   1     0    1
## 187          Vatican-City    Europe    2       0       4   1     0    0
## 191            Yugoslavia    Europe    0       3       4   1     0    1
##     gold black white orange mainhue circles crosses saltires quarter
## 2      1     0     1      0     red       0       0        0       0
## 5      1     0     0      0    gold       0       0        0       0
```

```
## 7      0     1     0     1   white       0       0       0       0
## 8      1     1     1     0     red       0       0       0       0
## 12     0     1     0     0     red       0       0       0       0
## 13     1     0     1     0    blue       0       0       0       0
## 16     1     0     1     0    blue       0       0       0       0
## 17     1     0     1     0    gold       0       0       0       0
## 18     1     1     1     1    blue       1       0       0       0
## 20     1     1     1     0     red       1       1       1       1
## 25     1     1     0     1    blue       0       1       1       1
## 27     1     1     0     0     red       0       0       0       0
## 32     0     1     0     0     red       0       0       0       0
## 34     1     1     0     1    blue       1       1       1       1
## 43     0     1     0     0    blue       0       0       0       0
## 44     0     1     0     0    blue       0       0       0       0
## 45     1     1     0     0   white       0       0       0       0
## 46     0     1     0     0   white       0       0       0       0
## 47     0     1     0     0     red       0       1       0       0
## 49     1     1     1     0   green       1       0       0       0
## 50     0     1     0     0    blue       0       1       0       0
## 53     0     1     0     0    blue       0       0       0       0
## 56     0     1     0     0   white       0       1       0       0
## 59     0     1     0     0   white       0       1       0       0
## 60     0     1     0     0   white       0       0       0       0
## 65     1     0     1     0    gold       0       0       0       0
## 66     1     0     1     0   black       0       0       0       0
## 68     1     1     0     0   white       0       0       0       0
## 69     0     1     0     0    blue       0       1       0       1
## 70     0     1     0     0   white       1       0       0       0
## 71     1     0     0     0    gold       1       0       0       0
## 73     0     1     0     0    blue       0       0       0       0
## 77     0     0     1     0   black       0       0       0       0
## 78     0     1     0     0    blue       0       0       0       0
## 80     0     1     0     0     red       0       0       0       0
## 81     0     1     0     0    blue       0       1       0       0
## 86     0     1     0     1   white       0       0       0       0
## 88     0     1     0     0   white       0       0       0       0
## 90     1     0     1     0   green       0       0       1       0
## 102    1     0     0     0     red       0       0       0       0
## 103    0     1     0     0     red       0       0       0       0
## 109    0     1     1     0     red       0       1       0       0
## 113    0     1     0     1   green       0       0       0       0
## 115    0     1     0     0     red       0       0       0       0
## 117    1     1     1     0    blue       0       2       1       1
## 122    0     1     0     0     red       0       0       0       0
## 123    0     1     0     0   white       0       0       0       0
## 125    0     1     0     0    blue       0       0       0       0
## 131    0     1     0     0     red       0       1       0       0
## 139    0     1     0     0   white       0       0       0       0
## 140    1     1     0     0     red       1       0       0       0
## 141    0     1     0     0     red       0       0       0       0
## 143    1     1     0     1     red       0       0       0       0
## 145    0     1     0     0   white       0       0       0       0
## 157    1     0     0     0     red       0       0       0       0
## 160    1     1     1     0   green       0       0       0       0
```

```
## 161    1     1     1      0    blue       0       0        0       0
## 162    1     1     0      0   green       0       0        0       0
## 166    1     0     0      0    blue       0       1        0       0
## 167    0     1     0      0     red       0       1        0       0
## 177    1     1     0      1    blue       0       1        1       1
## 181    0     1     0      0     red       0       1        1       0
## 183    1     1     0      0   white       0       0        0       0
## 184    0     1     0      0   white       0       0        0       1
## 187    1     1     1      0    gold       0       0        0       0
## 191    1     1     0      0     red       0       0        0       0
##     sunstars crescent triangle icon animate text
## 2          1        0        0    0       1    0
## 5          0        0        0    0       0    0
## 7          0        0        0    0       1    0
## 8          1        0        1    0       0    0
## 12         0        0        0    0       0    0
## 13         0        0        1    0       0    0
## 16         0        0        0    1       0    0
## 17         0        0        0    0       0    0
## 18         0        0        0    1       1    1
## 20         0        0        0    1       1    0
## 25         0        0        0    1       1    1
## 27         1        0        0    1       1    0
## 32         0        0        0    0       1    0
## 34         4        0        0    1       1    1
## 43         0        0        0    0       0    0
## 44         1        0        1    0       0    0
## 45         0        0        0    1       1    0
## 46         0        0        1    0       0    0
## 47         0        0        0    0       0    0
## 49        10        0        0    0       1    0
## 50         0        0        0    0       0    0
## 53         0        0        0    0       0    0
## 56         0        0        0    0       0    0
## 59         0        0        0    0       0    0
## 60         0        0        0    0       0    0
## 65         0        0        0    1       0    0
## 66         0        0        0    0       0    0
## 68         0        0        0    1       0    0
## 69         0        0        0    0       0    0
## 70         0        0        0    0       0    0
## 71         7        0        1    0       1    0
## 73         0        0        0    0       0    0
## 77         0        0        0    0       0    0
## 78         5        0        0    0       0    0
## 80         0        0        0    0       0    0
## 81         0        0        0    0       0    0
## 86         0        0        0    0       0    0
## 88         0        0        0    0       0    0
## 90         0        0        1    0       0    0
## 102        0        0        0    1       0    0
## 103        0        0        0    0       0    0
## 109        0        0        0    1       0    0
## 113        0        0        0    0       1    0
```

```
## 115        0        0        0    0        0    0
## 117        0        0        0    1        1    0
## 122        0        0        0    0        0    0
## 123        6        0        0    0        0    0
## 125        0        0        0    0        0    0
## 131        0        0        0    0        0    0
## 139        0        0        0    0        0    0
## 140        0        0        0    1        0    0
## 141        1        0        1    0        0    0
## 143        2        0        0    1        1    1
## 145        0        0        0    0        0    0
## 157        0        0        0    0        0    0
## 160        2        0        1    0        0    0
## 161        0        0        1    0        0    0
## 162        0        0        0    1        1    1
## 166        0        0        0    0        0    0
## 167        0        0        0    0        0    0
## 177        0        0        0    1        1    0
## 181        0        0        0    0        0    0
## 183        0        0        0    1        1    1
## 184       50        0        0    0        0    0
## 187        0        0        0    1        0    0
## 191        1        0        0    0        0    0
```