Week 2 Assignment

Dan Brooks

February 5, 2016

```
# Task:
#choose one dataset, then study the data and its associated description of the data
#(i.e. "data dictionary"). You should take the data, and #create an R data frame with
#a subset of the columns in the dataset. Your deliverable is the R code to perform
#these #transformation tasks.
```

```
#Grab the data from a web source. The data is in CSV (Comma separated format)
theURL <- "https://archive.ics.uci.edu/ml/machine-learning-databases/bridges/bridges.data.version2"
GetBridgeData <- read.csv(file = theURL, header = FALSE, sep = ",", as.is = TRUE)

#Set the names for the columns of the data.
names(GetBridgeData) <- c("Identifier", "River", "Location", "Erected", "Purpose", "Length",
                          "Lanes", "ClearG","TorD", "Material", "Span", "RelL", "Type")

#Show the headers and the top few rows of the data
head(GetBridgeData)
```

```
##   Identifier River Location Erected  Purpose Length Lanes ClearG    TorD
## 1         E1     M        3  CRAFTS  HIGHWAY      ?     2      N THROUGH
## 2         E2     A       25  CRAFTS  HIGHWAY MEDIUM     2      N THROUGH
## 3         E3     A       39  CRAFTS AQUEDUCT      ?     1      N THROUGH
## 4         E5     A       29  CRAFTS  HIGHWAY MEDIUM     2      N THROUGH
## 5         E6     M       23  CRAFTS  HIGHWAY      ?     2      N THROUGH
## 6         E7     A       27  CRAFTS  HIGHWAY  SHORT     2      N THROUGH
##   Material   Span RelL Type
## 1     WOOD  SHORT    S WOOD
## 2     WOOD  SHORT    S WOOD
## 3     WOOD      ?    S WOOD
## 4     WOOD  SHORT    S WOOD
## 5     WOOD      ?    S WOOD
## 6     WOOD MEDIUM    S WOOD
```

```
#Parse out the data frame into individual vectors for easier manipulation
GetBridgeData_Identify <- as.vector(as.character(GetBridgeData$`Identifier`))
GetBridgeData_River <- as.vector(as.character(GetBridgeData$`River`))
GetBridgeData_Location <- as.vector(as.character(GetBridgeData$`Location`))
GetBridgeData_Erected <- as.vector(as.character(GetBridgeData$`Erected`))
GetBridgeData_Purpose <- as.vector(as.character(GetBridgeData$`Purpose`))
GetBridgeData_Length <- as.vector(as.character(GetBridgeData$`Length`))
GetBridgeData_Lanes <- as.vector(as.character(GetBridgeData$`Lanes`))
GetBridgeData_ClearG <- as.vector(as.character(GetBridgeData$`ClearG`))
GetBridgeData_TorD <- as.vector(as.character(GetBridgeData$`TorD`))
GetBridgeData_Material <- as.vector(as.character(GetBridgeData$`Material`))
GetBridgeData_Span <- as.vector(as.character(GetBridgeData$`Span`))
GetBridgeData_RelL <- as.vector(as.character(GetBridgeData$`RelL`))
GetBridgeData_Type <- as.vector(as.character(GetBridgeData$`Type`))
```

```r
#Go through all of the individual vectors and change the '?' to N/a
GetBridgeData_Length <- replace(GetBridgeData_Length, GetBridgeData_Length == '?', 'N/a')
GetBridgeData_Span <- replace(GetBridgeData_Span, GetBridgeData_Span == '?', 'N/a')
GetBridgeData_Identify <- replace(GetBridgeData_Identify, GetBridgeData_Identify == '?', 'N/a')
GetBridgeData_River <- replace(GetBridgeData_River, GetBridgeData_River == '?', 'N/a')
GetBridgeData_Location <- replace(GetBridgeData_Location, GetBridgeData_Location == '?', 'N/a')
GetBridgeData_Erected <- replace(GetBridgeData_Erected, GetBridgeData_Erected == '?', 'N/a')
GetBridgeData_Purpose <- replace(GetBridgeData_Purpose, GetBridgeData_Purpose == '?', 'N/a')
GetBridgeData_Lanes <- replace(GetBridgeData_Lanes, GetBridgeData_Lanes == '?', 'N/a')
GetBridgeData_ClearG <- replace(GetBridgeData_ClearG, GetBridgeData_ClearG == '?', 'N/a')
GetBridgeData_TorD <- replace(GetBridgeData_TorD, GetBridgeData_TorD == '?', 'N/a')
GetBridgeData_Material <- replace(GetBridgeData_Material, GetBridgeData_Material == '?', 'N/a')
GetBridgeData_RelL <- replace(GetBridgeData_RelL, GetBridgeData_RelL == '?', 'N/a')
GetBridgeData_Type <- replace(GetBridgeData_Type, GetBridgeData_Type == '?', 'N/a')
```

```r
#Binds the individual vecotrs back into one whole factor. Then displays the top and bottom
#rows of the data.
SubSet_df <- cbind.data.frame(GetBridgeData_Identify, GetBridgeData_Material, GetBridgeData_Lanes)
head(SubSet_df)
```

```
##   GetBridgeData_Identify GetBridgeData_Material GetBridgeData_Lanes
## 1                     E1                  WOOD                   2
## 2                     E2                  WOOD                   2
## 3                     E3                  WOOD                   1
## 4                     E5                  WOOD                   2
## 5                     E6                  WOOD                   2
## 6                     E7                  WOOD                   2
```

```r
tail(SubSet_df)
```

```
##     GetBridgeData_Identify GetBridgeData_Material GetBridgeData_Lanes
## 103                   E85                  STEEL                   4
## 104                   E84                  STEEL                   6
## 105                   E91                  STEEL                   6
## 106                   E90                  STEEL                   6
## 107                  E100                    N/a                 N/a
## 108                  E109                    N/a                 N/a
```