Lab Exercise 1

Laurence Aguas

2024-03-07

```
#1a.
# A. Load the built-in warpbreaks dataset.
data("warpbreaks")

# 1. Find out, in a single command, which columns of warpbreaks are either numeric or integer. What are
column_types <- sapply(warpbreaks, class)
column_types
```

```
##    breaks      wool   tension
## "numeric"  "factor"  "factor"
```

```
# 2. How many observations does it have?
num_observations <- nrow(warpbreaks)
num_observations
```

```
## [1] 54
```

```
# 3. Is numeric a natural data type for the columns which are stored as such? Convert to integer when n
warpbreaks$breaks <- as.integer(warpbreaks$breaks)

# 4. Error messages in R sometimes report the underlying type of an object rather than the user-level c
observe <- nrow(warpbreaks)
```

#B. Load the exampleFile.txt

```
#2.
FileLines <- readLines("exampleFile.txt")
```

```
## Warning in readLines("exampleFile.txt"): incomplete final line found on
## 'exampleFile.txt'
```

```
comments <- FileLines[grepl("//", FileLines)]
data <- FileLines[!grepl("//", FileLines)]

comments
```

```
## [1] "// Survey data. Created : 21 May 2013"
## [2] "// Field 1: Gender"
## [3] "// Field 2: Age (in years)"
## [4] "// Field 3: Weight (in kg)"
```

```
data
```

```
## [1] "M;28;81.3"      "male;45;"       "Female;17;57,2" "fem.;64;62.8"
```

```
#3.c
CreationDate <- gsub(".*Created : ", "", comments[1])
cat("It was created", CreationDate, "data.\n")
```

```
## It was created 21 May 2013 data.
```

```
#4.c
  #4.a
    Split <- strsplit(data,comments,split = ";")
    Split
```

```
## [[1]]
## [1] "M"    "28"   "81.3"
##
## [[2]]
## [1] "male" "45"
##
## [[3]]
## [1] "Female" "17"     "57,2"
##
## [[4]]
## [1] "fem." "64"   "62.8"
```

```
  #4.b
    MaxField <- max(sapply(Split, length))
    MaxField
```

```
## [1] 3
```

```
    Split <- lapply(Split, function(x) c(x, rep(NA, MaxField - length(x))))

  #4.c
    DataMatrix <- matrix(unlist(Split), nrow = length(Split), byrow = TRUE)
     DataMatrixcol <- matrix(unlist(Split), ncol = length(Split),)
    DataMatrix
```

```
##      [,1]     [,2] [,3]
## [1,] "M"      "28" "81.3"
## [2,] "male"   "45" NA
## [3,] "Female" "17" "57,2"
## [4,] "fem."   "64" "62.8"
```

```
  #4.d
    col <- colnames(DataMatrixcol)
    col
```

```
## NULL
```

```
    colnames(DataMatrix)
```

```
## NULL
```

```
    colnames(DataMatrix) <- gsub("^Field \\d+: ", "", comments[2:4])

    DataMatrix
```

```
##      // Field 1: Gender // Field 2: Age (in years) // Field 3: Weight (in kg)
## [1,] "M"                "28"                       "81.3"
## [2,] "male"             "45"                       NA
## [3,] "Female"           "17"                       "57,2"
## [4,] "fem."             "64"                       "62.8"
```