# CS 102 (Data Preparation)

## Cedric

## 2024-03-06

```r
data("warpbreaks")

#View(warpbreaks)
```

How many observations does it have? 54

```r
nrow(warpbreaks)
```

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

```r
str(warpbreaks)
```

```
## 'data.frame':    54 obs. of  3 variables:
##  $ breaks : num  26 30 54 25 70 52 51 26 67 18 ...
##  $ wool   : Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
##  $ tension: Factor w/ 3 levels "L","M","H": 1 1 1 1 1 1 1 1 1 2 ...
```

```r
#warpbreaks$wool <- as.integer(warpbreaks)
```

```r
sapply(warpbreaks, function(x) {paste(class(x), collapse=",")})
```

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

B.

1.Read the complete file using readLines.

```r
sample <- readLines("exampleFile.txt")
```

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

```r
sample
```

```
## [1] "// Survey data. Created : 21 May 2013"
## [2] "// Field 1: Gender"
## [3] "// Field 2: Age (in years)"
## [4] "// Field 3: Weight (in kg)"
## [5] "M;28;81.3"
## [6] "male;45;"
## [7] "Female;17;57,2"
## [8] "fem.;64;62.8"
```

2.Separate the vector of lines into a vector containing comments and a vector containing the data. Hint: use grepl.

```r
file_content <- readLines("exampleFile.txt")
```

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

```r
comment <- file_content[grepl("^//", file_content)]
data_lines <- file_content[grepl("^//", file_content)]
```

3.Extract the date from the first comment line and display on the screen "It was created data."

```r
date_lines <- comment[1]
dat <- gsub("^// Created on ","", date_lines)
#cat("created on", date, "\n")
```

4.

```r
#spli_data way na time na
```