Exercise 2

Your Name

09 Oct 2015

Create a data frame called my.patients using the instructions in the slides. Change the data if you like.

```r
age     <- c(50, 21, 35, 45, 28, 31, 42, 33, 57, 62)
weight <- c(70.8, 67.9, 75.3, 61.9, 72.4, 69.9, 63.5,
71.5, 73.2, 64.8)
firstName <- c("Adam", "Eve", "John", "Mary", "Peter",
"Paul", "Joanna", "Matthew", "David", "Sally")
secondName <- c("Jones", "Parker", "Evans", "Davis",
"Baker","Daniels", "Edwards", "Smith", "Roberts", "Wilson")

consent <- c(TRUE, TRUE, FALSE, TRUE, FALSE, FALSE,
FALSE, TRUE, FALSE, TRUE)

sex <- c("Male", "Female", "Male", "Female", "Male", "Male",
"Female", "Male", "Male", "Female")

patients <- data.frame(First_Name = firstName,
                       Second_Name = secondName,
                       Full_Name = paste(firstName, secondName),
                       Sex = factor(sex),
                       Age = age,
                       Weight = weight,
                       Consent = consent,
                       stringsAsFactors = FALSE)
```

Remake your data frame with three new variables: country, continent, and height

- Make up the data
- Make country a character vector but continent a factor

```r
library(gapminder)
data("gapminder")
randind <- sample(1:nrow(gapminder),10)
Country <- as.character(gapminder$country[randind])
Continent <- gapminder$continent[randind]
Height <- rnorm(n=10, mean = 64,sd = 3)
my.patients <- data.frame(First_Name = firstName,
                       Second_Name = secondName,
                       Full_Name = paste(firstName, secondName),
                       Sex = factor(sex),
                       Age = age,
                       Weight = weight,
                       Consent = consent,
```

```r
                     Country,
                     Continent = factor(Continent),
                     Height,
                     stringsAsFactors = FALSE)
summary(my.patients)
```

```
##    First_Name         Second_Name          Full_Name              Sex
##  Length:10           Length:10           Length:10             Female:4
##  Class :character    Class :character    Class :character      Male  :6
##  Mode  :character    Mode  :character    Mode  :character
##
##
##
##       Age              Weight           Consent            Country
##  Min.   :21.00    Min.   :61.90    Mode :logical    Length:10
##  1st Qu.:31.50    1st Qu.:65.58    FALSE:5          Class :character
##  Median :38.50    Median :70.35    TRUE :5          Mode  :character
##  Mean   :40.40    Mean   :69.12    NA's :0
##  3rd Qu.:48.75    3rd Qu.:72.17
##  Max.   :62.00    Max.   :75.30
##    Continent      Height
##  Africa:3    Min.   :57.82
##  Asia  :6    1st Qu.:60.07
##  Europe:1    Median :62.80
##              Mean   :63.25
##              3rd Qu.:67.26
##              Max.   :67.92
```

```r
my.patients <- data.frame(patients,
                     Country,
                     Continent = factor(Continent),
                     Height)
```

```r
my.patients <- cbind(patients,
                     Country,
                     Continent = factor(Continent),
                     Height)
```

Try the summary function on your data frame. What does it do? How does it treat vectors (numeric, character, logical) and factors? (What does it do for matrices?)

Use logical indexing to select the following patients:

- Patients under 40

```r
patients[patients$Age < 40,]
```

```
##    First_Name Second_Name    Full_Name    Sex Age Weight Consent
## 2         Eve      Parker    Eve Parker Female  21   67.9    TRUE
```

```
## 3       John       Evans    John Evans   Male  35    75.3    FALSE
## 5      Peter       Baker   Peter Baker   Male  28    72.4    FALSE
## 6       Paul     Daniels  Paul Daniels   Male  31    69.9    FALSE
## 8    Matthew       Smith Matthew Smith   Male  33    71.5     TRUE
```

- Patients who give consent to share their data

```
patients[patients$Consent == TRUE,]
```

```
##     First_Name Second_Name     Full_Name    Sex Age Weight Consent
## 1         Adam       Jones    Adam Jones   Male  50   70.8    TRUE
## 2          Eve      Parker   Eve Parker Female  21   67.9    TRUE
## 4         Mary       Davis    Mary Davis Female  45   61.9    TRUE
## 8      Matthew       Smith Matthew Smith   Male  33   71.5    TRUE
## 10       Sally      Wilson  Sally Wilson Female  62   64.8    TRUE
```

```
patients[patients$Consent,]
```

```
##     First_Name Second_Name     Full_Name    Sex Age Weight Consent
## 1         Adam       Jones    Adam Jones   Male  50   70.8    TRUE
## 2          Eve      Parker   Eve Parker Female  21   67.9    TRUE
## 4         Mary       Davis    Mary Davis Female  45   61.9    TRUE
## 8      Matthew       Smith Matthew Smith   Male  33   71.5    TRUE
## 10       Sally      Wilson  Sally Wilson Female  62   64.8    TRUE
```

- Men who weigh as much or more than the average European male (70.8 kg)

```
patients[patients$Sex == "Male" & patients$Weight >= 70.8,]
```

```
##    First_Name Second_Name     Full_Name  Sex Age Weight Consent
## 1        Adam       Jones    Adam Jones Male  50   70.8    TRUE
## 3        John       Evans    John Evans Male  35   75.3   FALSE
## 5       Peter       Baker   Peter Baker Male  28   72.4   FALSE
## 8     Matthew       Smith Matthew Smith Male  33   71.5    TRUE
## 9       David     Roberts David Roberts Male  57   73.2   FALSE
```