

# Lab 7

Yi Chen (yc3356)

November 27, 2016

## Instructions

Before you leave lab today make sure that you upload a .pdf file to the canvas page (this should have a .pdf extension). This should be the PDF output after you have knitted the file, we don't need the .Rmd file (don't upload the one with the .Rmd extension). The file you upload to the Canvas page should be updated with commands you provide to answer each of the questions below. You can edit this file directly to produce your final solutions. Note, however, in the file you upload you should the above header to have the date, your name, and your UNI. Similarly, when you save the file you should replace **UNI** with your actualy UNI.

## Introduction

In today's lab we will use data on the 2829 fastest men's and women's 100m sprint times saved as dataframes `sprint.m.csv` and `sprint.f.csv`.

1. Load the dataframe `sprint.m.csv` and save it as `sprint.m`. Append a column to the dataframe called `CityDate` that is defined by concatenating the string entries in the `City` and `Date` columns. For example, entries "Berlin" and "16.08.2009" in the `City` and `Date` columns, respectively, produce an entry of "Berlin 16.08.2009" in the `CityDate` column. Note that there will be some weird characters in the city names.

```
setwd("C:/Users/cheny/Desktop/study/statistical computing and intro to data science/lab/lab 7")
```

```
sprint.m <- read.csv('sprint.m.csv',header = TRUE)
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.4.2
```

```
sprint.m <- sprint.m %>%  
  unite(CityDate,c(City, Date),sep = " ",remove = FALSE)  
head(sprint.m,3)
```

```
##   Rank Time Wind      Name Country Birthdate      CityDate      City  
## 1    1  9.58  0.9 Usain Bolt      JAM   21.08.86 Berlin 16.08.2009 Berlin  
## 2    2  9.63  1.5 Usain Bolt      JAM   21.08.86 London 05.08.2012 London  
## 3    3  9.69  0.0 Usain Bolt      JAM   21.08.86 Beijing 16.08.2008 Beijing  
##           Date  
## 1 16.08.2009  
## 2 05.08.2012  
## 3 16.08.2008
```

2. We assume that every unique combination of city and date in the `CityDate` column corresponds to a unique track meet. How many unique track meets occur? How many other sprint times were recorded in the same track meet as Usain Bolt's legendary time of 9.58 seconds?

```
length(unique(sprint.m$CityDate))
```

```
## [1] 1181
```

```
nrow(sprint.m[sprint.m$CityDate==sprint.m[sprint.m$Time==9.58,]$CityDate,])
```

```
## [1] 16
```

3. Compute a reduced version of `sprint.m` that only keeps the fastest time from each track meet. For example, of all rows that correspond to sprint times recorded at the “Berlin 16.08.2009” track meet, we will only keep Usain Bolt’s row since his time of 9.58 was fastest. Hint: There are many ways to do this, `tapply()` or `split()` might be helpful. You can do this without using a loop. Call the result `sprint.m.fastest` and check that the number of rows is the same as the number of unique men’s track meets. Display the first five rows. Note that if there are ties for first place, choose only one runner to represent the race in the reduced data frame. How you chose doesn’t matter.

```
sprint.m.fastest <- tapply(sprint.m$Time, INDEX = sprint.m$CityDate, FUN = min)
```

```
length(sprint.m.fastest)==length(unique(sprint.m$CityDate))
```

```
## [1] TRUE
```

```
g <- sprint.m$CityDate
l <- split(sprint.m,g)

fastest <- function(df){
  min(df[2])
}
l <- lapply(l,fastest)
```

4. Load the women’s dataframe `sprint.w.csv` and repeat steps (1) and (2) on this dataset so that what remains is `sprint.w.fastest`. Display the first five rows.

```
setwd("C:/Users/cheny/Desktop/study/statistical computing and intro to data science/lab/lab 7")
```

```
sprint.w <- read.csv('sprint.w.csv',header = TRUE)
```

```
library(tidyr)
sprint.w <- sprint.w %>%
  unite(CityDate,c(City, Date),sep = " ",remove = FALSE)
head(sprint.w,3)
```

```
##   Rank  Time Wind      Name Country Birthdate
## 1    1  10.49  0.0 Florence Griffith-Joyner    USA  21.12.59
## 2    2  10.61 +1.2 Florence Griffith-Joyner    USA  21.12.59
## 3    3  10.62 +1.0 Florence Griffith-Joyner    USA  21.12.59
##                CityDate      City      Date
## 1 Indianapolis 16.07.1988 Indianapolis 16.07.1988
## 2 Indianapolis 17.07.1988 Indianapolis 17.07.1988
## 3      Seoul 24.09.1988      Seoul 24.09.1988
```

```
sprint.w.fastest <- tapply(sprint.w$Time, INDEX = sprint.w$CityDate, FUN = min)
```