Practical 4

Jumping Rivers

Data Manipulation

This section will hopefully help you get more comfortable with some of the **dplyr** functionality for "wrangling" your data. We will do some data wrangling and the use that to create some plots. Make sure you load in the **dplyr** package and the `movies` data set.

```
library("dplyr")
data("movies", package = "jrIntroduction")
```

1. We want to look at how film budgets for films in English have changed over time for both the Comedy and non Comedy films. To start with, we should filter the data set such that it only contains films spoken in English. Try using the `%>%` notation, i.e. `movies %>% filter(...)`
2. We want to look at comedy and non comedy films in each year, this is some sort of grouping structure which suggests use of the `group_by()` function. Create this grouping structure on the filtered movies data
3. Use the `summarise()` function to calculate the average budget in each year for both comedy and non comedy films.

Question 2

Run the folowing R code:

```
data(USnames, package = "jrIntroduction")
```

The tibble `USnames` is a collection of names given to babies born in the US between 2011 and 2014.

1. Make sure you are comfortable with what the data looks like using `head()` and `colnames()`.
2. How many children were born in 2012? Hint: use `filter()` then `summarise()`
3. Were more male or females children born during the four years? Hint: use `group_by()`
4. Tricky: How many names in 2011 were used fewer than ten times?

Solutions

Solutions to the practical questions are contained within the package

```
vignette("solutions4", package = "jrIntroduction")
```