Stringr Practical

Jumping Rivers

Some of the questions in this practical might not exactly be things you would do in the real world, they are just intended to get you comfortable using some of the **stringr** functions you have seen so far.

Question 1

We'll start by loading the necessary packages and data sets

```
library("tidyverse")
data(names, package = "jrTidyverse2")
```

Here we have a data set containing 800 people with the names: "Abigail", "Alexander", "Ava", "Benjamin", "Charlotte", "Emily", "Emma", "Ethan", "Harper", "Isabella", "Jacob", "James", "Liam", "Mason", "Mia", "Michael", "Noah", "Olivia", "Sophia" and "William".

Using various functions from **stringr** and `count()` from **dplyr**, work out the frequency of each name. Which name occcurs the most?

```
## Warning: Detecting old grouped_df format, replacing `vars` attribute by
## `groups`
```

Question 2: Movies

We'll start off by loading the data

```
data(movies, package = "jrTidyverse")
```

1. How many movie titles contain the word "The" ?
2. Do any titles contain your name?
3. Mutate a new column that is the count of how many characters each movie title contains, call this new column `title_length`
4. How many characters does the longest title contain? Hint: Use `summarise()` and `max()`
5. Which film has the longest title?
6. Use **ggplot2** to produce a histogram of the title lengths
7. Produce a scatter plot of title length against rating, do you think longer title effect the rating?