Exercise: **String Functions**

Day 1, Part D

Use the following code to create a vector of all the files in the answer key folder:

```
> main_dir <- "C:/Users/ngraetz/Documents/repos/r_training_penn/"  # CHANGE TO YOUR COPY OF THE TRAI
> files <- list.files(paste0(main_dir, "exercises/answer_keys/"))
```

1. From the `files` vector, create a new vector (`pdfs`) of just the PDF files...
    a. Using `grepl()`.
    b. Using `grep()` with `value = F`.
    c. Using `grep()` with `value = T`.
2. From the `pdfs` vector...
    a. Create a new vector (`day1`) of just exercises from day 1.
    b. Remove "exercise_" and "_answers.pdf" from the elements of `day1`. (hint: use `gsub()`)
    c. Replace underscores with spaces in `day1`.
    d. Remove the lesson number/letter (e.g., "1a") from `day1`. (hint: use `substr()`)

Bonus

3. Load the Nigeria health metrics data set by running the following:

```
> data <- read.csv(paste0(main_dir, "data/nigeria_healthmap.csv"), stringsAsFactors = F)
```

    a. Keep only the rows where the indicator name includes "immunization coverage".
    b. Replace "immunization coverage" with "Coverage" in the `indicator` variable.
    c. Remove "(%)" from the `indicator` variable. (hint: see the `fixed` argument for gsub)
    d. The `indicator_type` variable has an extra space at the end–remove this space. (hint: use `gsub()` with regular expressions, or try the `trimws()` function)
    e. Using the `indicator_type`, `indicator`, `location_name`, and `year` variables, create a new variable (`full_title`) that has values similar to: "Childhood immunizations: BCG Coverage, Nigeria, 2000 (Percent)"