

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 TRAINING MATERIALS
> 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()`.

```
> pdfs <- files[grepl("pdf", files)]
```

b. Using `grep()` with `value = F`.

```
> pdfs <- files[grep("pdf", files)]
```

c. Using `grep()` with `value = T`.

```
> pdfs <- grep("pdf", files, value = T)
```

2. From the `pdfs` vector...

a. Create a new vector (`day1`) of just exercises from day 1.

```
> day1 <- grep("exercise_1", pdfs, value = T)
```

b. Remove “exercise_” and “_answers.pdf” from the elements of `day1`. (hint: use `gsub()`)

```
> day1 <- gsub("exercise_", "", day1)
> day1 <- gsub("_answers.pdf", "", day1)
```

c. Replace underscores with spaces in `day1`.

```
> day1 <- gsub("_", " ", day1)
```

d. Remove the lesson number/letter (e.g., “1a”) from `day1`. (hint: use `substr()`)

```
> day1 <- substr(day1, 4, 100)
```

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”.

```
> data <- data[grepl("immunization coverage", data$indicator), ]
```

b. Replace “immunization coverage” with “Coverage” in the `indicator` variable.

```
> data$indicator <- gsub("immunization coverage", "Coverage", data$indicator)
```

c. Remove “(%)” from the `indicator` variable. (hint: see the `fixed` argument for `gsub`)

```
> data$indicator <- gsub(" (%)", "", data$indicator, fixed = T)
```

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)

```
> data$indicator_type <- gsub(" $", "", data$indicator_type)
> # or
> data$indicator_type <- trimws(data$indicator_type)
```

- 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)”

```
> data$full_title <- paste0(data$indicator_type, ": ", data$indicator, ", ",
+                           data$location_name, ", ", data$year, " (Percent)")
```