Data science with R: tidyverse

II Data Import (readr & tibble)

Assignment

Create *R* script called *assignment_2.R*. From course sources download zip file called **data_import.zip**, extract its content in your **data** folder inside your *R*'s project folder.

Exercise 1

In this exercise you will create a **tibble** called **continents**, using data from the table shown on Figure 1. After you have created a tibble use it to calculate given table summaries:

- total area
- total population
- sum of percentage - total landmass
- sum of percentage - total population

Figure 1: Table - Continents

	Date (data published)	Continent	Area (km2)	Percent total landmass	Population	Percent total pop.
1	2017-11-10	Africa	30370000	20.4	1287920000	16.9
2	2017-11-10	Antarctica	14000000	9.2	4490	0.1
3	2017-11-10	Asia	44579000	29.5	4545133000	59.5
4	2017-11-10	Europe	10180000	6.8	742648000	9.7
5	2017-11-10	North America	24709000	16.5	587615000	7.7
6	2017-11-10	South America	17840000	12.0	428240000	5.6
7	2017-11-10	Australia	8600000	5.9	41264000	0.5

Exercise 2

In this exercise you will import **.csv** file called **flights_02.csv**, which is located in zip file. When you are importing the file, try to consider the following:

- assign imported object to *R* object named **df2**
- for importing use function from library **readr**
- inside function for importing, define column parsing

- maybe you should inspect the **.csv** file before actual import
- after the import, check structure of **df2** object with **str()**

Exercise 3

Now you will import **.csv** file called **flights_03.csv**. When you are importing the file, you should consider the following:

- assign imported object to *R* object named **df3**
- for importing use function from library **readr**
- inspect the **.csv** file before actual import
- function for import should include some additional import strategies (compared to previous example!)
- **HINT:** at the point of import: maybe you should parse all columns as characters (***col_types = cols(.default = "c")***)
- **HINT:** then inside *R* you can convert column types and add column names, after the import operation is executed
- after the import, check structure of **df3** object with **str()**

Exercise 4

In the last exercise you will import **.csv** file called **flights_04.csv**, which is a larger flat file with several million rows:

- import file two times using **readr** library and **data.table**'s **fread**
- when importing with **readr** do column parsing at the point of import
- when importing with **fread** force all columns to be parsed as characters (***colClasses = "character"***)
- compare execution times for each importing strategy