Introduction to R 2

Andreas Chandra

Contents

- Reading Raw Data
- Reading CSV file
- Regular Expression
- Using Apply
- Data Table
- Dplyr

Reading Raw File

Reading RAW File

```
read.table(file = <path/to/file>, header = TRUE, sep = ‘,’)

readHTMLTable(theURL, which = 1, header = FALSE, stringsAsFactors = FALSE)
```

Reading CSV File

Read CSV File

```
read.csv(file = theUrl, header = TRUE, sep = ',')
```

More

Function	Format
`read.spss`	SPSS
`read.dta`	Stata
`read.ssd`	SAS
`read.octave`	Octave
`read.mtp`	Minitab
`read.systat`	Systat

Regular Expression

Using RegEx

Take a look http://stringr.tidyverse.org/articles/regular-expressions.html

String Function

- str_split(string = <string>, pattern = <RegEx>)
- str_trim(<string>) > removing white-space
- str_sub(<string>, start = <int>, end = <int>)
- str_detect(string = <string>, pattern = <RegEx>)
- str_extract(string = <string>, pattern = <RegEx>)
- str replace(string = <vector>, pattern = <RegEx>, replacement = <string>)
- str replace all(string = <vector>, pattern = <RegEx>, replacement = <string>)

Exercise

1. Date string e.g “Jan 1 2018” “01 01 2018 10:10:10” > get desired format date
2. Email lala@lala.com, goks@geq.co.id > get the domain
3.
4. Website blog.andreas.or.id dsi.or.id

Using Apply Function

Apply Family

- apply(<matrix>, axis = <1|2>, <function>, na.rm = <T|F>) 1 for row 2 for col, return vector
- lapply(<list>, <function>), from list return list
- sapply(<list>, <function>), from list return vector
- mapply(<function>, list1, list2), from list return vector

Process Data with data.table

Creating data.table

```
require(data.table)

data.table(<data.frame>)
```

Processing data.table

In short, data.table like sql data.table[where, select, group by]

```
Ex: House[size > 10, mean(size), by = “luxury”)]
```

Exercise 1

get new metrics total calls and total minutes

Exercise 2

select only state is WV or MI and churn is False

Exercise 3

get mean, median, mode, std for each state

Exercise 4

Join table

Exercise 5

get data then order ascending area code but descending customer service call

Exercise 6

get pivot table for

Using dplyr

Usage

Data_frame %>% dplyr_funs()

E.g:

flight %>% mean()

Common functions

```
Dim_desc #get dimension
Summarise family
Arrange
Filter
Join
Group_by
mutate
```