

DATA SCIENCE WITH R

Class 4 – Data Manipulation in R

Topic 7

★ Manipulating Character Strings and Using sqldf ★

INDEX

Manipulating data using base R

Using dplyr to manipulate data

Working with date objects

Merging tables

Missing value treatment

Using reshape2() to transpose data



Manipulating Character Strings

Using sqldf

Manipulating Strings in R

Manipulating Strings in R

- We'll cover the following R functions used to manipulate character data
- `substr()`
- `nchar()`
- `tolower()`, `toupper()`
- `strsplit()`, `paste()`
- `grep()`, `grepl()`,
- `sub()`, `gsub()`

Manipulating Strings in R

- substr(), nchar(), tolower(), toupper()

```
> a<-"Batman"  
> substr(a,start=2,stop=6)  
[1] "atman"  
> nchar(a)  
[1] 6  
> tolower(a)  
[1] "batman"  
> toupper(a)  
[1] "BATMAN"
```

Manipulating Strings in R

- `strsplit()`, `paste()`, `grep()`, `grepl()`, `sub()`, `gsub()`

```
> b<-"Bat-Man"
> c<-"Bat/Man"
> strsplit(c,split="/")
[[1]]
[1] "Bat" "Man"

> paste(b,split=c)
[1] "Bat-Man Bat/Man"
> grep("-",b)
[1] 1
> grepl("/",<div data-bbox="770 953 984 973" data-label="Page-Footer">

© Jigsaw Academy Education Pvt Ltd


```

INDEX

Manipulating data using base R

Using dplyr to manipulate data

Working with date objects

Merging tables

Missing value treatment

Using reshape2() to transpose data

Manipulating Character Strings



Using sqldf

Using sql commands within R

SQL queries within R: sqldf

- One can easily use sql queries within R using a package called sqldf
- This package is available at CRAN
- `install.package(sqldf)`
- `library(sqldf)`

Manipulating Strings in R

- Using select statement

```
> library(sqldf)
> #Using SELECT statement
> oj_s<-sqldf("select brand, income, feat from oj ")
```

- Using where statement

```
> #Subsetting using where statement
> oi_s<-sqldf("select brand, income, feat from oi where price<3.8 and income<10")
```

- Using order by statement

```
> #Order by statement
> oj_s<-sqldf("select store,brand,week,logmove,feat,price, income from oj order by income asc")
```

SQL queries within R: sqldf

- Using sql functions

```
> #Demo sql functions
> sqldf("select avg(income) from oj")
  avg(income)
1    10.61673
> sqldf("select min(price) from oj")
  min(price)
1         0.52
```

RECAP

- Basic string manipulation functions: tolower, toupper, grep, grepl, sub, gsub
- Executing sql queries in R, using sqldf() package