LAB -4

Data Cleaning

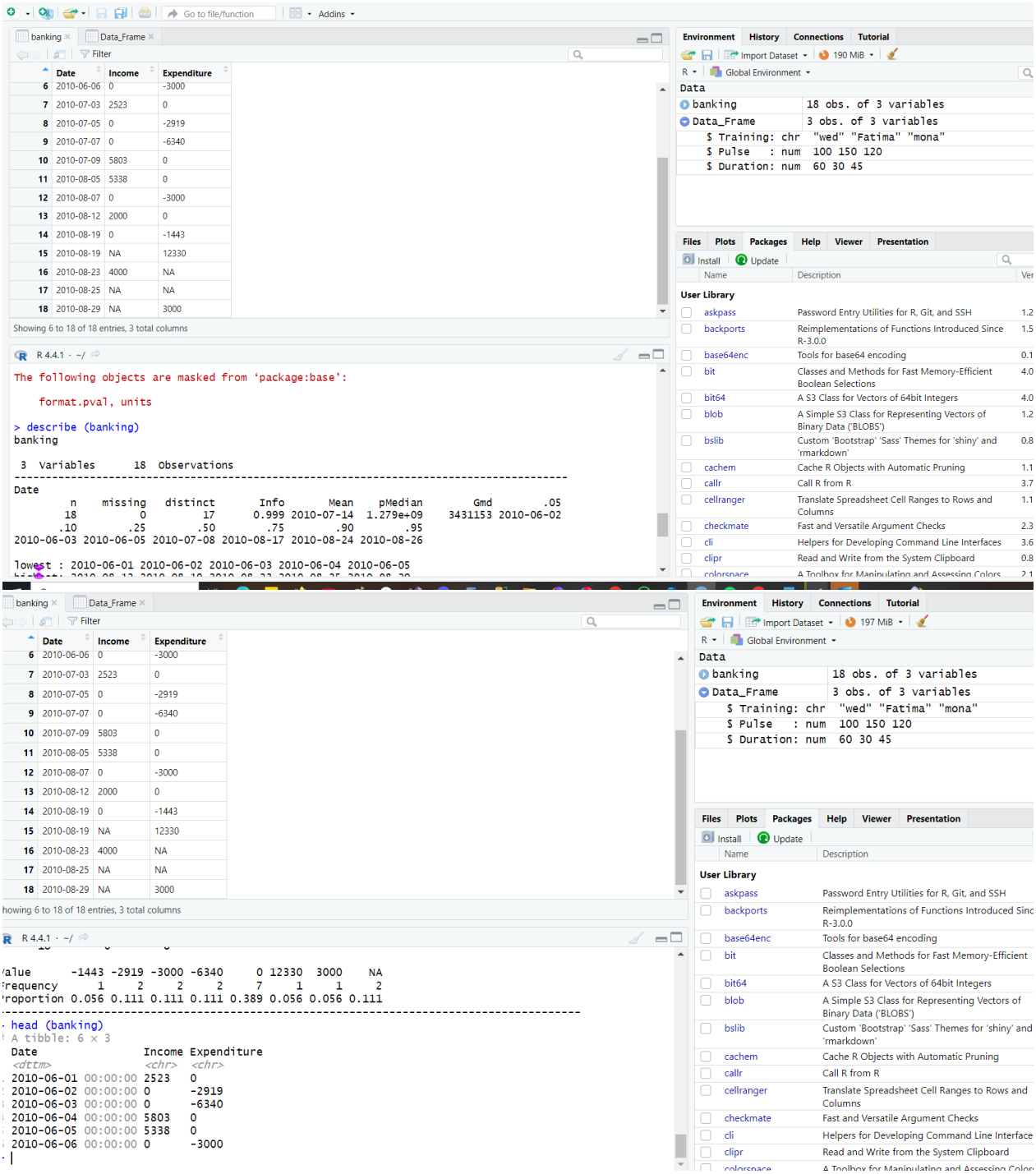
ID :442801472

Name: Wed Ghanem

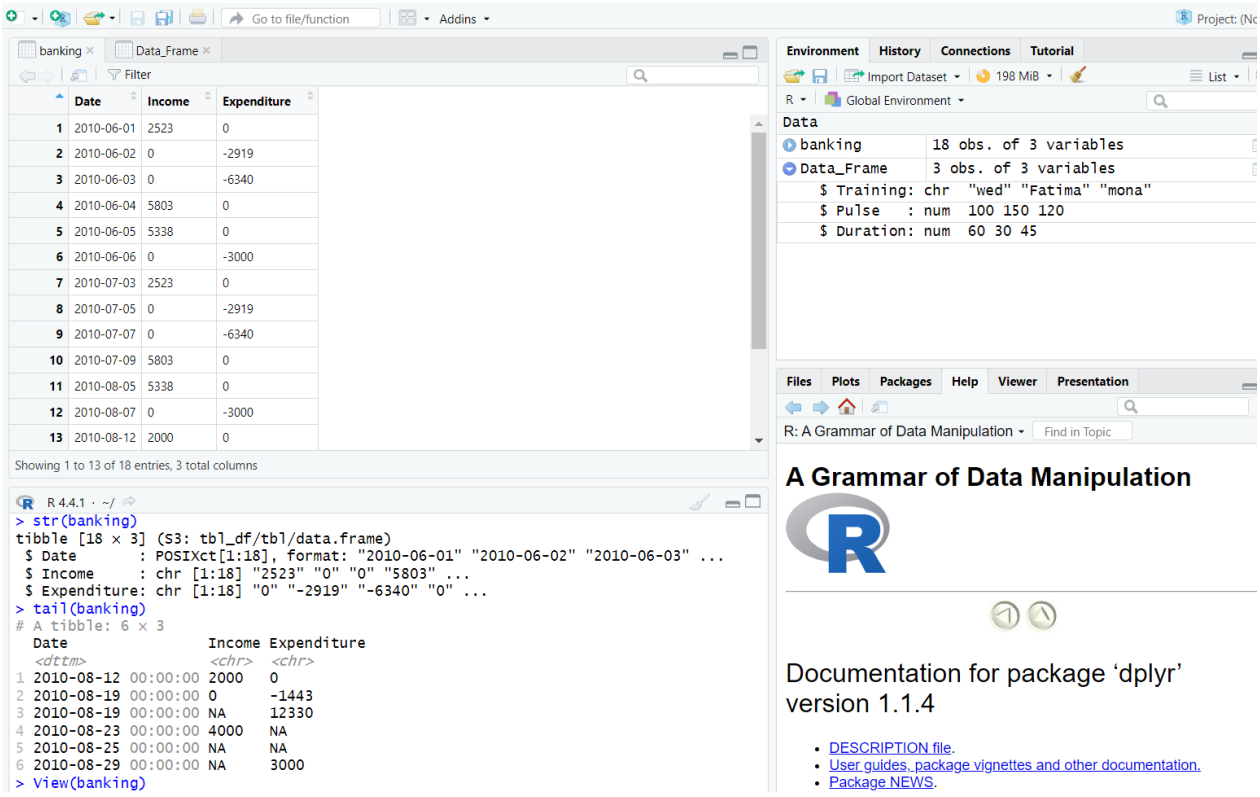
2024/11/03

Screenshot:

Step 1:



Step2:



	Date	Income	Expenditure
1	2010-06-01	2523	0
2	2010-06-02	0	-2919
3	2010-06-03	0	-6340
4	2010-06-04	5803	0
5	2010-06-05	5338	0
6	2010-06-06	0	-3000
7	2010-07-03	2523	0
8	2010-07-05	0	-2919
9	2010-07-07	0	-6340
10	2010-07-09	5803	0
11	2010-08-05	5338	0
12	2010-08-07	0	-3000
13	2010-08-12	2000	0

Showing 1 to 13 of 18 entries, 3 total columns

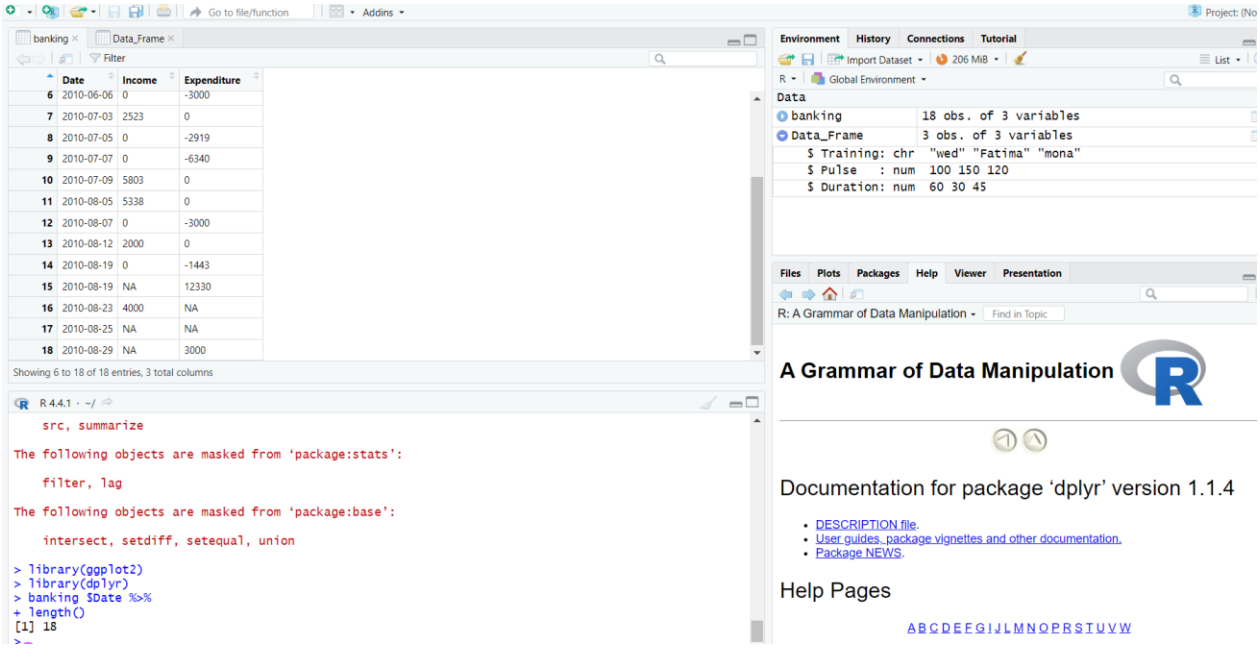
```
> str(banking)
tibble [18 × 3] (S3: tbl_df/tbl/data.frame)
 $ Date       : POSIXct[1:18], format: "2010-06-01" "2010-06-02" "2010-06-03" ...
 $ Income     : chr [1:18] "2523" "0" "0" "5803" ...
 $ Expenditure: chr [1:18] "0" "-2919" "-6340" "0" ...
> tail(banking)
# A tibble: 6 × 3
  Date                Income Expenditure
  <dttm>              <chr>  <chr>
1 2010-08-12 00:00:00 2000   0
2 2010-08-19 00:00:00 0      -1443
3 2010-08-19 00:00:00 NA     12330
4 2010-08-23 00:00:00 4000   NA
5 2010-08-25 00:00:00 NA     NA
6 2010-08-29 00:00:00 NA     3000
> View(banking)
```

```
Data
banking      18 obs. of 3 variables
Data_Frame   3 obs. of 3 variables
 $ Training: chr  "wed" "Fatima" "mona"
 $ Pulse   : num  100 150 120
 $ Duration: num  60 30 45
```

A Grammar of Data Manipulation

Documentation for package ‘dplyr’ version 1.1.4

- DESCRIPTION file.
- User guides, package vignettes and other documentation.
- Package NEWS.



	Date	Income	Expenditure
6	2010-06-06	0	-3000
7	2010-07-03	2523	0
8	2010-07-05	0	-2919
9	2010-07-07	0	-6340
10	2010-07-09	5803	0
11	2010-08-05	5338	0
12	2010-08-07	0	-3000
13	2010-08-12	2000	0
14	2010-08-19	0	-1443
15	2010-08-19	NA	12330
16	2010-08-23	4000	NA
17	2010-08-25	NA	NA
18	2010-08-29	NA	3000

Showing 6 to 18 of 18 entries, 3 total columns

```
    src, summarize

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

> library(ggplot2)
> library(dplyr)
> banking $Date %>%
+ length()
[1] 18
```

```
Data
banking      18 obs. of 3 variables
Data_Frame   3 obs. of 3 variables
 $ Training: chr  "wed" "Fatima" "mona"
 $ Pulse   : num  100 150 120
 $ Duration: num  60 30 45
```

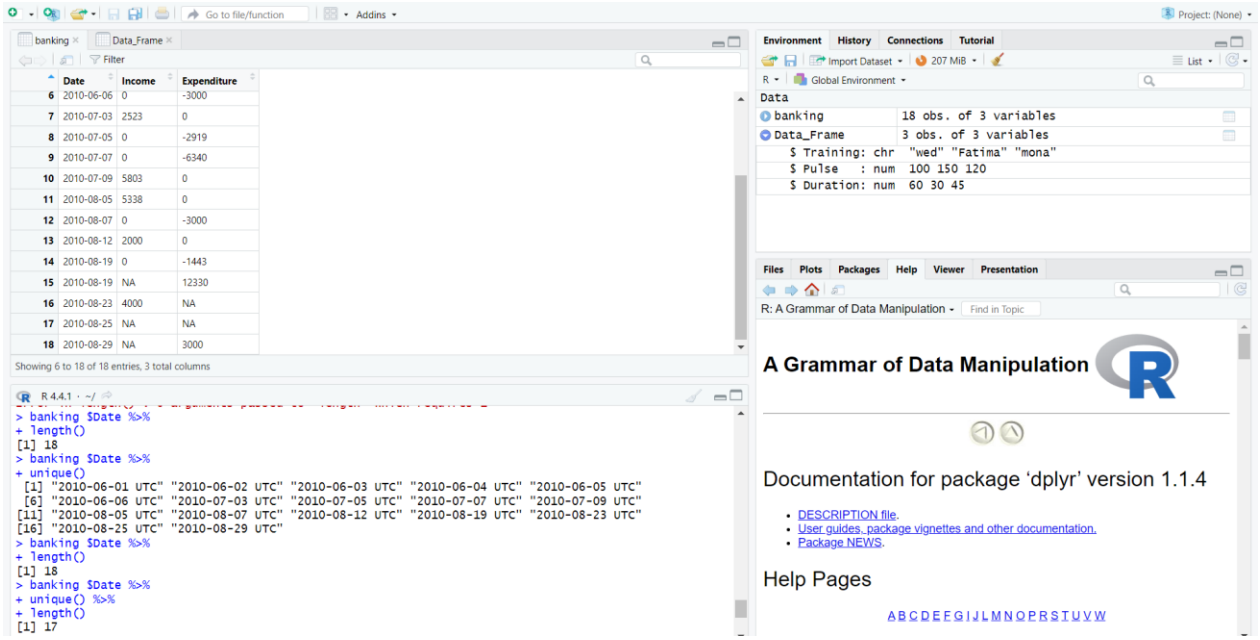
A Grammar of Data Manipulation

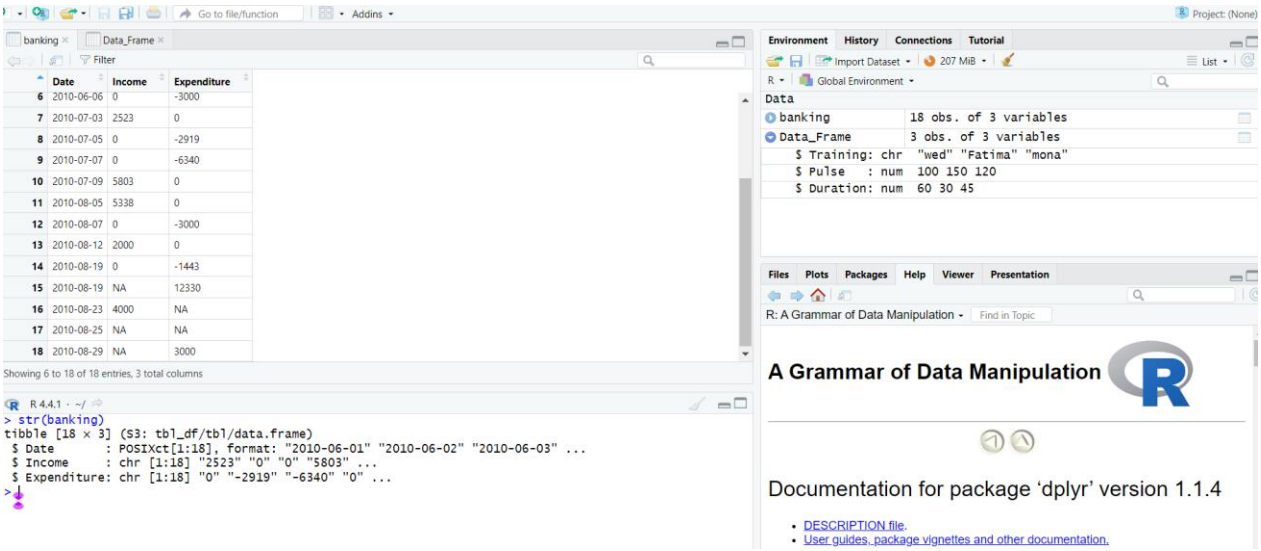
Documentation for package ‘dplyr’ version 1.1.4

- DESCRIPTION file.
- User guides, package vignettes and other documentation.
- Package NEWS.

Help Pages

A B C D E F G I J L M N O P R S T U V W


banking
Data_Frame
Filter
Date Income Expenditure
6 2010-06-06 0 -3000
7 2010-07-03 2523 0
8 2010-07-05 0 -2919
9 2010-07-07 0 -6340
10 2010-07-09 5803 0
11 2010-08-05 5338 0
12 2010-08-07 0 -3000
13 2010-08-12 2000 0
14 2010-08-19 0 -1443
15 2010-08-19 NA 12330
16 2010-08-23 4000 NA
17 2010-08-25 NA NA
18 2010-08-29 NA 3000
Showing 6 to 18 of 18 entries, 3 total columns
R 4.4.1 · ~/
> banking $Date %>%
+ length()
[1] 18
> banking $Date %>%
+ unique()
[1] "2010-06-01 UTC" "2010-06-02 UTC" "2010-06-03 UTC" "2010-06-04 UTC" "2010-06-05 UTC"
[6] "2010-06-06 UTC" "2010-07-03 UTC" "2010-07-05 UTC" "2010-07-07 UTC" "2010-07-09 UTC"
[11] "2010-08-05 UTC" "2010-08-07 UTC" "2010-08-12 UTC" "2010-08-19 UTC" "2010-08-23 UTC"
[16] "2010-08-25 UTC" "2010-08-29 UTC"
> banking $Date %>%
+ length()
[1] 18
> banking $Date %>%
+ unique() %>%
+ length()
[1] 17
Environment History Connections Tutorial
Import Dataset 207 MiB
Global Environment
Data
banking 18 obs. of 3 variables
Data_Frame 3 obs. of 3 variables
$ Training: chr "wed" "Fatima" "mona"
$ Pulse : num 100 150 120
$ Duration: num 60 30 45
Files Plots Packages Help Viewer Presentation
R: A Grammar of Data Manipulation
A Grammar of Data Manipulation
Documentation for package 'dplyr' version 1.1.4
DESCRIPTION file.
User guides, package vignettes and other documentation.
Package NEWS.
Help Pages
A B C D E F G I J L M N O P R S T U V W


banking
Data_Frame
Filter
Date Income Expenditure
6 2010-06-06 0 -3000
7 2010-07-03 2523 0
8 2010-07-05 0 -2919
9 2010-07-07 0 -6340
10 2010-07-09 5803 0
11 2010-08-05 5338 0
12 2010-08-07 0 -3000
13 2010-08-12 2000 0
14 2010-08-19 0 -1443
15 2010-08-19 NA 12330
16 2010-08-23 4000 NA
17 2010-08-25 NA NA
18 2010-08-29 NA 3000
Showing 6 to 18 of 18 entries, 3 total columns
R 4.4.1 · ~/
> str(banking)
tibble [18 × 3] (S3: tbl_df/tbl/data.frame)
$ Date : POSIXct[1:18], format: "2010-06-01" "2010-06-02" "2010-06-03" ...
$ Income : chr [1:18] "2523" "0" "0" "5803" ...
$ Expenditure: chr [1:18] "0" "-2919" "-6340" "0" ...
Environment History Connections Tutorial
Import Dataset 207 MiB
Global Environment
Data
banking 18 obs. of 3 variables
Data_Frame 3 obs. of 3 variables
$ Training: chr "wed" "Fatima" "mona"
$ Pulse : num 100 150 120
$ Duration: num 60 30 45
Files Plots Packages Help Viewer Presentation
R: A Grammar of Data Manipulation
A Grammar of Data Manipulation
Documentation for package 'dplyr' version 1.1.4
DESCRIPTION file.
User guides, package vignettes and other documentation.