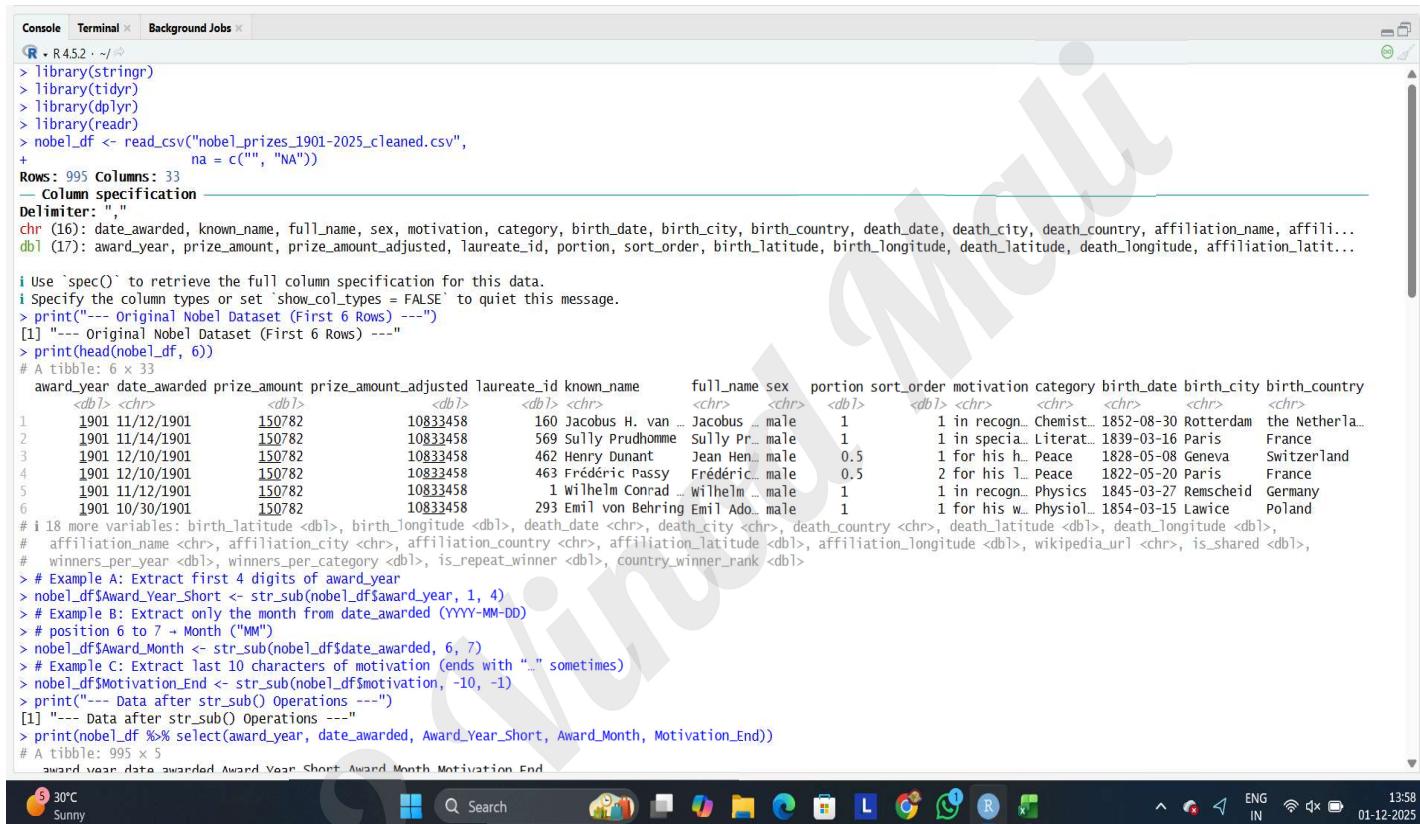


Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS /R
Practical No 9

Aim :- Performing text manipulation using str_sub(), str_split() (R). import dataset.



The screenshot shows an RStudio interface with the following session history:

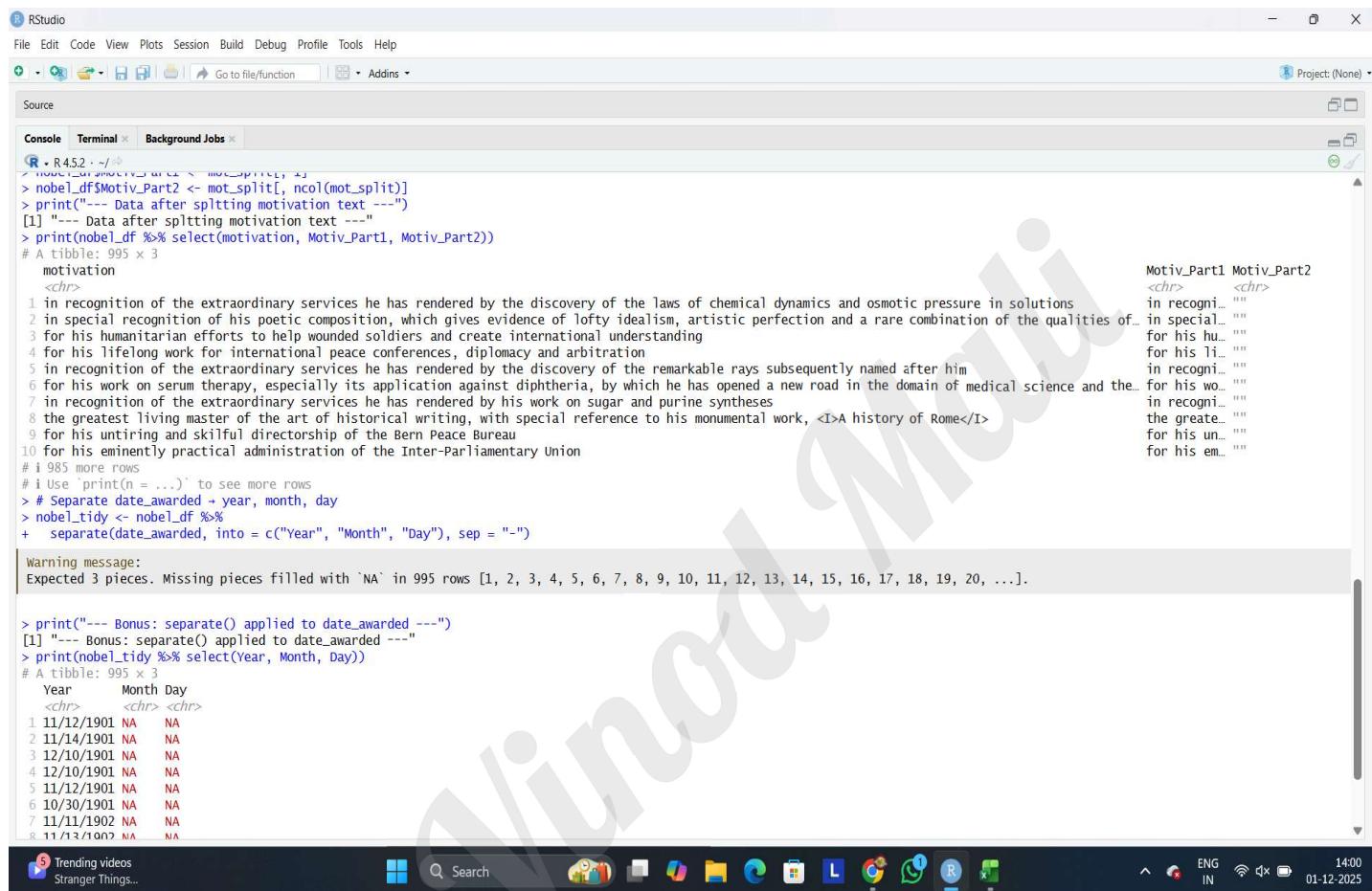
```
R 4.5.2 · ~/Desktop
> library(stringr)
> library(tidyverse)
> library(dplyr)
> library(readr)
> nobel_df <- read_csv("nobel_prizes_1901-2025_cleaned.csv",
+   na = c("", "NA"))
Rows: 995 Columns: 33
  Column specification:
  Delimiter: ","
  chr (16): date_awarded, known_name, full_name, sex, motivation, category, birth_date, birth_city, birth_country, death_date, death_city, death_country, affiliation_name, affili...
  dbl (17): award_year, prize_amount, prize_amount_adjusted, laureate_id, portion, sort_order, birth_latitude, birth_longitude, death_latitude, death_longitude, death_longitude, affiliation_lati...
  i Use `spec()` to retrieve the full column specification for this data.
  i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> print("--- Original Nobel Dataset (First 6 Rows) ---")
[1] "--- Original Nobel Dataset (First 6 Rows) ---"
> print(head(nobel_df, 6))
# A tibble: 6 × 33
  award_year date_awarded prize_amount prize_amount_adjusted laureate_id known_name      full_name sex  portion sort_order motivation category birth_date birth_city birth_country
  <dbl> <chr>        <dbl>          <dbl>       <dbl> <chr>      <chr> <chr> <dbl> <dbl> <chr> <chr> <chr> <chr>
1 1901 11/12/1901     150782        10833458    160 Jacobus H. van ... Jacobus ... male  1     1 in recogn... Chemist... 1852-08-30 Rotterdam the Netherla...
2 1901 11/14/1901     150782        10833458    569 Sully Prudhomme Sully Pr... male  1     1 in specia... Literat... 1839-03-16 Paris France
3 1901 12/10/1901     150782        10833458    462 Henry Dunant Jean Hen... male  0.5   1 for his h... Peace   1828-05-08 Geneva Switzerland
4 1901 12/10/1901     150782        10833458    463 Frédéric Passy Frédéric... male  0.5   2 for his l... Peace   1822-05-20 Paris France
5 1901 11/12/1901     150782        10833458    1 Wilhelm Conrad ... Wilhelm ... male  1     1 in recogn... Physics 1845-03-27 Remscheid Germany
6 1901 10/30/1901     150782        10833458    293 Emil von Behring Emil Ado... male  1     1 for his w... Physiol. 1854-03-15 Lawice Poland
# i 18 more variables: birth_latitude <dbl>, birth_longitude <dbl>, death_date <chr>, death_city <chr>, death_country <chr>, death_longitude <dbl>, death_latitude <dbl>, death_longitude <dbl>,
# affiliation_name <chr>, affiliation_city <chr>, affiliation_country <chr>, affiliation_latitude <dbl>, affiliation_longitude <dbl>, wikipedia_url <chr>, is_shared <dbl>,
# winners_per_year <dbl>, winners_per_category <dbl>, is_repeat_winner <dbl>, country_winner_rank <dbl>
> # Example A: Extract first 4 digits of award_year
> nobel_df$Award_Year_Short <- str_sub(nobel_df$award_year, 1, 4)
> # Example B: Extract only the month from date_awarded (YYYY-MM-DD)
> # position 6 to 7 → Month ("MM")
> nobel_df$Award_Month <- str_sub(nobel_df$date_awarded, 6, 7)
> # Example C: Extract last 10 characters of motivation (ends with "..." sometimes)
> nobel_df$Motivation_End <- str_sub(nobel_df$motivation, -10, -1)
> print("--- Data after str_sub() Operations ---")
[1] "--- Data after str_sub() Operations ---"
> print(nobel_df %>% select(award_year, date_awarded, Award_Year_Short, Award_Month, Motivation_End))
# A tibble: 995 × 5
  award_year date_awarded Award_Year_Short Award_Month Motivation_End
```

Name :- Vinod Mali
Roll No :- S092

**Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS /R
Practical No 9**

**Name :- Vinod Mali
Roll No :- S092**

Sheth L.U.J & Sir M.V College Of Science
Subject :- Data Analysis with SAS / SPSS /R
Practical No 9



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
R • R4.5.2 ~/Desktop/nobel_df.R
> nobel_df$Motiv_Part2 <- mot_split[, ncol(mot_split)]
> print("--- Data after splitting motivation text ---")
[1] "--- Data after splitting motivation text ---"
> print(nobel_df %>% select(motivation, Motiv_Part1, Motiv_Part2))
# A tibble: 995 × 3
  motivation Motiv_Part1 Motiv_Part2
  <chr>       <chr>       <chr>
1 in recognition of the extraordinary services he has rendered by the discovery of the laws of chemical dynamics and osmotic pressure in solutions
2 in special recognition of his poetic composition, which gives evidence of lofty idealism, artistic perfection and a rare combination of the qualities of...
3 for his humanitarian efforts to help wounded soldiers and create international understanding
4 for his lifelong work for international peace conferences, diplomacy and arbitration
5 in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him
6 for his work on serum therapy, especially its application against diphtheria, by which he has opened a new road in the domain of medical science and the...
7 in recognition of the extraordinary services he has rendered by his work on sugar and purine syntheses
8 the greatest living master of the art of historical writing, with special reference to his monumental work, <I>A history of Rome</I>
9 for his untiring and skilful directorship of the Bern Peace Bureau
10 for his eminently practical administration of the Inter-Parliamentary Union
# i 985 more rows
# i Use `print(n = ...)` to see more rows
> # Separate date_awarded -> year, month, day
> nobel_tidy <- nobel_df %>%
  +   separate(date_awarded, into = c("Year", "Month", "Day"), sep = "-")

Warning message:
Expected 3 pieces. Missing pieces filled with `NA` in 995 rows [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
> print("---- Bonus: separate() applied to date_awarded ----")
[1] "---- Bonus: separate() applied to date_awarded ----"
> print(nobel_tidy %>% select(Year, Month, Day))
# A tibble: 995 × 3
  Year     Month Day
  <chr>    <chr> <chr>
1 11/12/1901 NA    NA
2 11/14/1901 NA    NA
3 12/10/1901 NA    NA
4 12/10/1901 NA    NA
5 11/12/1901 NA    NA
6 10/30/1901 NA    NA
7 11/11/1902 NA    NA
8 11/13/1902 NA    NA
```

Trending videos
Stranger Things... Search

ENG IN 01-12-2025

Name :- Vinod Mali
Roll No :- S092