

SHETH L.U.J AND SIR M.V COLLEGE

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

```
R - R 4.5.2 - / ->
> install.packages("stringr")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/yukta/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/stringr_1.6.0.zip'
Content type 'application/zip' length 350430 bytes (342 KB)
downloaded 342 KB

package 'stringr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/yukta/AppData/Local/Temp/RtmpILRFgg/downloaded_packages
> install.packages("tidyverse")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/yukta/AppData/Local/R/win-library/4.5'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/tidyr_1.3.1.zip'
Content type 'application/zip' length 1276404 bytes (1.2 MB)
downloaded 1.2 MB

package 'tidyverse' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:/Users/yukta/AppData/Local/Temp/RtmpILRFgg/downloaded_packages
> library(stringr)
> library(tidyverse)
> library(dplyr)

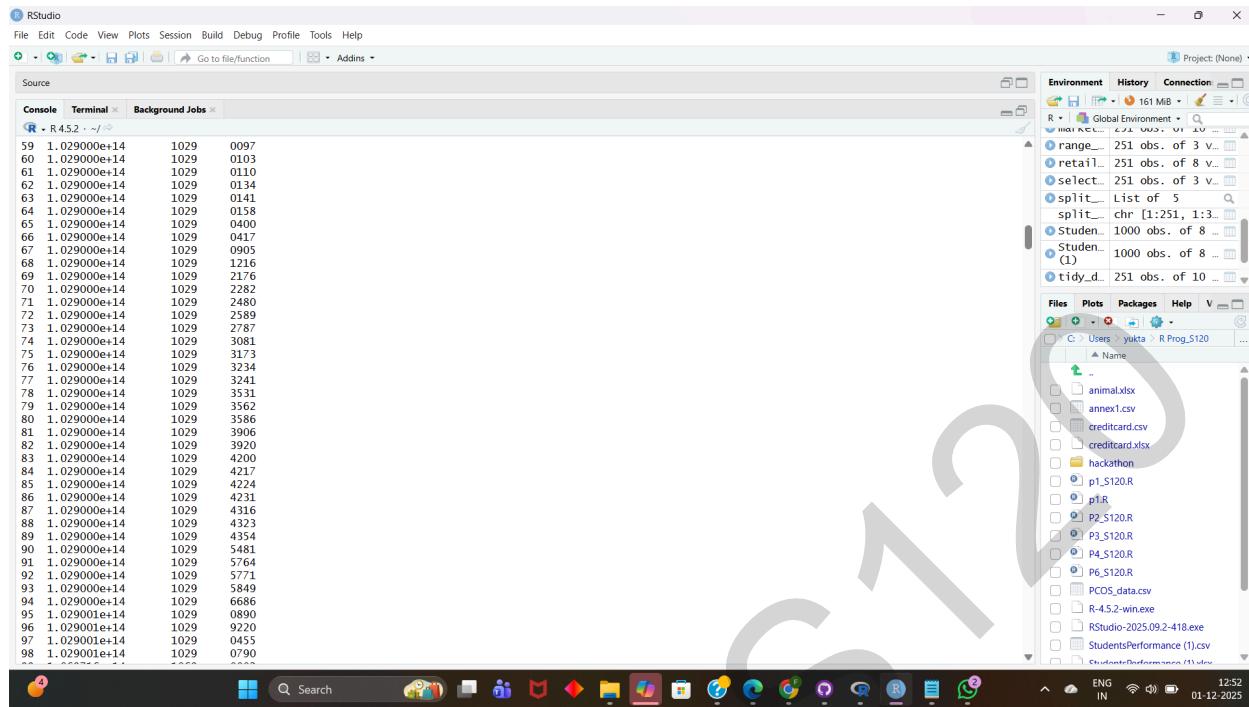
30°C Sunny Search ENG IN 12:51 01-12-2025
```



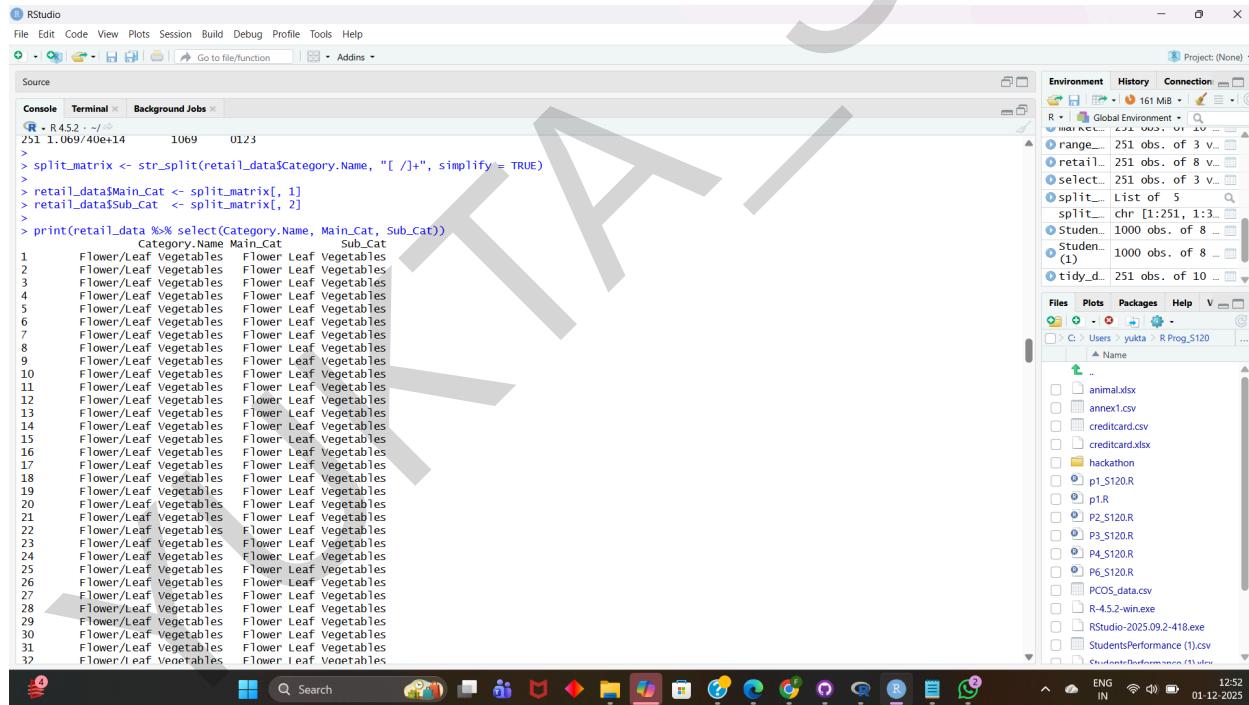
```
R - R 4.5.2 - / ->
> retail_data <- read.csv("annex1.csv", na.strings = c("", "NA"))
> print(head(retail_data))
  Item.Code      Item.Name Category.Code   Category.Name
1 1.029e+14    Niushou Shengcai 1011010101 Flower/Leaf Vegetables
2 1.029e+14    Sichuan Red Cedar 1011010101 Flower/Leaf Vegetables
3 1.029e+14 Local Xiao Mao Cabbage 1011010101 Flower/Leaf Vegetables
4 1.029e+14     White Caital 1011010101 Flower/Leaf Vegetables
5 1.029e+14   Amaranth 1011010101 Flower/Leaf Vegetables
6 1.029e+14   Yunnan Shengcai 1011010101 Flower/Leaf Vegetables
> retail_data$Code_Start <- str_sub(retail_data$Item.Code, 1, 4)
> retail_data$Code_End <- str_sub(retail_data$Item.Code, -4, -1)
>
> print(retail_data %>% select(Item.Code, Code_Start, Code_End))
  Item.Code Code_Start Code_End
1 1.029000e+14      1029      5168
2 1.029000e+14      1029      5199
3 1.029000e+14      1029      5625
4 1.029000e+14      1029      5748
5 1.029000e+14      1029      5761
6 1.029000e+14      1029      5799
7 1.029000e+14      1029      5786
8 1.029000e+14      1029      5793
9 1.029000e+14      1029      5816
10 1.029000e+14      1029      5823
11 1.029000e+14      1029      5854
12 1.029000e+14      1029      5861
13 1.029000e+14      1029      5878
14 1.029000e+14      1029      5885
15 1.029000e+14      1029      5908
16 1.029000e+14      1029      5946
17 1.029000e+14      1029      5960
18 1.029000e+14      1029      5971
19 1.029000e+14      1029      5974
20 1.029000e+14      1029      6639
21 1.029000e+14      1029      6776
22 1.029000e+14      1029      6790
23 1.029000e+14      1029      6806
24 1.029000e+14      1029      8572
25 1.029000e+14      1029      8817

animal.xlsx annex1.csv creditcard.csv creditcard.xlsx hackathon p1_S120.R p1.R P2_S120.R P3_S120.R P4_S120.R P6_S120.R PCOS_data.csv R-4.5.2-win.exe RStudio-2025.09.2-418.exe StudentsPerformance (1).csv StudentsPerformance (1).xlsx 161 MB 12:52 ENG IN 01-12-2025
```

SHETH L.U.J AND SIR M.V COLLEGE



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - 4.5.2 - />
59 1.029000e+14 1029 0097
60 1.029000e+14 1029 0103
61 1.029000e+14 1029 0110
62 1.029000e+14 1029 0134
63 1.029000e+14 1029 0141
64 1.029000e+14 1029 0158
65 1.029000e+14 1029 0400
66 1.029000e+14 1029 0417
67 1.029000e+14 1029 0905
68 1.029000e+14 1029 1216
69 1.029000e+14 1029 2176
70 1.029000e+14 1029 2282
71 1.029000e+14 1029 2480
72 1.029000e+14 1029 2589
73 1.029000e+14 1029 2787
74 1.029000e+14 1029 3081
75 1.029000e+14 1029 3173
76 1.029000e+14 1029 3234
77 1.029000e+14 1029 3241
78 1.029000e+14 1029 3531
79 1.029000e+14 1029 3562
80 1.029000e+14 1029 3586
81 1.029000e+14 1029 3900
82 1.029000e+14 1029 3920
83 1.029000e+14 1029 4200
84 1.029000e+14 1029 4217
85 1.029000e+14 1029 4224
86 1.029000e+14 1029 4231
87 1.029000e+14 1029 4316
88 1.029000e+14 1029 4323
89 1.029000e+14 1029 4354
90 1.029000e+14 1029 5481
91 1.029000e+14 1029 5764
92 1.029000e+14 1029 5771
93 1.029000e+14 1029 5849
94 1.029000e+14 1029 6800
95 1.029001e+14 1029 6890
96 1.029001e+14 1029 9220
97 1.029001e+14 1029 9455
98 1.029001e+14 1029 0790
12:52
ENG IN 01-12-2025



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R - 4.5.2 - />
251 1.069/40e+14 1069 0123
>
> split_matrix <- str_split(retail_data\$category.Name, "[/]+", simplify = TRUE)
>
> retail_data\$Main_Cat <- split_matrix[, 1]
> retail_data\$Sub_Cat <- split_matrix[, 2]
>
> print(retail_data %>% select(Category.Name, Main_Cat, Sub_Cat))
Category.Name Main_Cat Sub_Cat
1 Flower/Leaf Vegetables Flower Leaf Vegetables
2 Flower/Leaf Vegetables Flower Leaf Vegetables
3 Flower/Leaf Vegetables Flower Leaf Vegetables
4 Flower/Leaf Vegetables Flower Leaf Vegetables
5 Flower/Leaf Vegetables Flower Leaf Vegetables
6 Flower/Leaf Vegetables Flower Leaf Vegetables
7 Flower/Leaf Vegetables Flower Leaf Vegetables
8 Flower/Leaf Vegetables Flower Leaf Vegetables
9 Flower/Leaf Vegetables Flower Leaf Vegetables
10 Flower/Leaf Vegetables Flower Leaf Vegetables
11 Flower/Leaf Vegetables Flower Leaf Vegetables
12 Flower/Leaf Vegetables Flower Leaf Vegetables
13 Flower/Leaf Vegetables Flower Leaf Vegetables
14 Flower/Leaf Vegetables Flower Leaf Vegetables
15 Flower/Leaf Vegetables Flower Leaf Vegetables
16 Flower/Leaf Vegetables Flower Leaf Vegetables
17 Flower/Leaf Vegetables Flower Leaf Vegetables
18 Flower/Leaf Vegetables Flower Leaf Vegetables
19 Flower/Leaf Vegetables Flower Leaf Vegetables
20 Flower/Leaf Vegetables Flower Leaf Vegetables
21 Flower/Leaf Vegetables Flower Leaf Vegetables
22 Flower/Leaf Vegetables Flower Leaf Vegetables
23 Flower/Leaf Vegetables Flower Leaf Vegetables
24 Flower/Leaf Vegetables Flower Leaf Vegetables
25 Flower/Leaf Vegetables Flower Leaf Vegetables
26 Flower/Leaf Vegetables Flower Leaf Vegetables
27 Flower/Leaf Vegetables Flower Leaf Vegetables
28 Flower/Leaf Vegetables Flower Leaf Vegetables
29 Flower/Leaf Vegetables Flower Leaf Vegetables
30 Flower/Leaf Vegetables Flower Leaf Vegetables
31 Flower/Leaf Vegetables Flower Leaf Vegetables
32 Flower/Leaf Vegetables Flower Leaf Vegetables
12:52
ENG IN 01-12-2025

SHETH L.U.J AND SIR M.V COLLEGE

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The main area displays a console window titled "Console" with the following R code and output:

```
R - R 4.5.2 ~
215 1.029000e+14 1029 00011030929
216 1.029000e+14 1029 00011031599
217 1.029000e+14 1029 00011031841
218 1.029000e+14 1029 00011031858
219 1.029000e+14 1029 00011031926
220 1.029000e+14 1029 00011031995
221 1.029000e+14 1029 00011032619
222 1.029000e+14 1029 00011032626
223 1.029000e+14 1029 00011032633
224 1.029000e+14 1029 00011032640
225 1.029000e+14 1029 00011032647
226 1.029000e+14 1029 000110333944
227 1.029000e+14 1029 00011033968
228 1.029000e+14 1029 00011034330
229 1.029000e+14 1029 00011034538
230 1.029000e+14 1029 00011034705
231 1.029000e+14 1029 00011035740
232 1.029000e+14 1029 00011035788
233 1.029000e+14 1029 00011036068
234 1.029000e+14 1029 00011036266
235 1.029001e+14 1029 00051009336
236 1.069303e+14 1069 3027420092
237 1.069303e+14 1069 30274620090
238 1.069319e+14 1069 31885000035
239 1.069497e+14 1069 49711300068
240 1.069497e+14 1069 49711300068
241 1.069497e+14 1069 49711300167
242 1.069497e+14 1069 49711300259
243 1.069561e+14 1069 56146480197
244 1.069561e+14 1069 56146480203
245 1.069576e+14 1069 57634300010
246 1.069576e+14 1069 57634300058
247 1.069589e+14 1069 58851400125
248 1.069715e+14 1069 71533450003
249 1.069715e+14 1069 71533455008
250 1.069732e+14 1069 73223300667
251 1.069740e+14 1069 73990980123
> print("Yukta Sonawane S120")
[1] "Yukta Sonawane S120"
> |
```

The right sidebar contains tabs for Environment, History, Connection, and a file browser. The file browser shows a directory structure under "Name" with files like animal.xlsx, annex1.csv, creditcard.csv, creditcard.xlsx, hackathon, p1_S120.R, p1.R, P2_S120.R, P3_S120.R, P4_S120.R, P6_S120.R, PCOS_data.csv, R-4.5.2-win.exe, RStudio-2025.09.2-418.exe, StudentsPerformance (1).csv, and StudentPerformance (1).xlsx.

**YUKTA SONAWANE S120
R PROGRAMMING
PRACTICAL NO: 9**