

# SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

## SUBJECT : Data Analysis with SAS / SPSS / R

**AIM :** Performing text manipulation using `str_sub()`, `str_split()` (R). import dataset.

### OUTPUT

The first screenshot shows the RStudio interface with the following code in the console:

```
R - R4.5.1 - ~/r/
> library(stringr)
Warning message:
package 'stringr' was built under R version 4.5.2
> library(tidyverse)
> library(dplyr)
> traffic_df <- read_csv("local_authority_traffic.csv")
> print("--- Original Dataset (First 6 Rows) ---")
[1] "--- Original Dataset (First 6 Rows) ---"
> print(head(traffic_df))
  local_authority_id local_authority_name year link_length_km link_length_miles cars_and_taxis
1                45      Aberdeenshire 2019      6273.378      3898.10      1537817161
2                107      Lambeth 2019      377.000      234.26      397710911
3                172 Newcastle upon Tyne 2019      984.992      612.05      1006029025
4                93      Tower Hamlets 2019      287.327      178.54      465293697
5                158      St. Helens 2019      740.220      459.95      728175389
6                69      Worcestershire 2019      4273.420      2655.38      3556865608

all_motor_vehicles
1 2055244624
2 547123377
3 1209012058
4 630938804
5 930780139
6 4546565782

> traffic_df$Name_Start <- str_sub(traffic_df$local_authority_name, 1, 3)
> traffic_df$Name_End <- str_sub(traffic_df$local_authority_name, -5, -1)
> traffic_df$Year_Str <- str_sub(as.character(traffic_df$year), 1, 4)
> print("--- After str_sub() ---")
[1] "--- After str_sub() ---"
> print(traffic_df %>% select(local_authority_name, Name_Start, Name_End, Year_Str))
  local_authority_name Name_Start Name_End Year_Str
1 Aberdeenshire Aberde Aberdee 2019
2 Lambeth Lambeth Lambeth 2019
3 Newcastle upon Tyne Newcas Newcastle 2019
4 Tower Hamlets Tower Tower 2019
5 St. Helens St. Helens St. Helens 2019
6 Worcestershire Worces Worcesters 2019
```

The second screenshot shows the same RStudio interface with the following code in the console:

```
R - R4.5.1 - ~/r/
> print(traffic_df %>% select(local_authority_name, Name_Start, Name_End, Year_Str))
  local_authority_name Name_Start Name_End Year_Str
4 Newcastle upon Tyne Newcas Newcastle 2019
5 Tower Hamlets Tower Tower 2019
6 St. Helens St. Helens St. Helens 2019
7 Worcestershire Worces Worcesters 2019
8 Lewisham Lewisham Lewisham 2019
9 Camden Camden Camden 2019
10 Kingston upon Hull, City of Kingston Kingston 2019
11 Stockport Stockport Stockport 2019
12 Bedford Bedford Bedford 2019
13 Swansea Swansea Swansea 2019
14 Somerset Somerset Somerset 2019
15 Cardiff Cardiff Cardiff 2019
16 Kirklees Kirklees Kirklees 2019
17 Bath and North East Somerset Bath Bath 2019
18 Bracknell Forest Bracknell Bracknell 2019
19 Falkirk Falkirk Falkirk 2019
20 Hertfordshire Hertford Hertford 2019
21 Torbay Torbay Torbay 2019
22 Pembrokeshire Pembrok Pembrok 2019
23 Barnsley Barnsley Barnsley 2019
24 Suffolk Suffolk Suffolk 2019
25 Slough Slough Slough 2019
26 Sutton Sutton Sutton 2019
27 East Riding of Yorkshire East Riding East Riding 2019
28 Norfolk Norfolk Norfolk 2019
29 West Lothian West Lothian West Lothian 2019
30 Hampshire Hampshire Hampshire 2019
31 Southampton Southampton Southampton 2019
32 Essex Essex Essex 2019
33 North Ayrshire North Ayrshire North Ayrshire 2019
34 Wigan Wigan Wigan 2019
35 Croydon Croydon Croydon 2019
36 Rotherham Rotherham Rotherham 2019
```

# SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

## SUBJECT : Data Analysis with SAS / SPSS / R

The image displays two screenshots of the RStudio interface, showing R code execution for data analysis.

**Top Screenshot:**

```
R - R4.5.1 - ~/...
[ reached 'max' / getoption("max.print") -- omitted 5279 rows ]
> split_list <- str_split(traffic_df$local_authority_name, " ")
> print("--- Example of split list (first item) ---")
[1] "--- Example of split list (first item) ---"
> print(split_list[[1]])
[1] "Aberdeenshire"
> split_matrix <- str_split(traffic_df$local_authority_name, " ", simplify = TRUE)
> traffic_df$word1 <- split_matrix[, 1]
> traffic_df$word2 <- split_matrix[, 2]
> traffic_df$word3 <- split_matrix[, 3]
> print("--- After str_split() ---")
[1] "--- After str_split() ---"
> print(traffic_df %>% select(local_authority_name, word1, word2, word3))
```

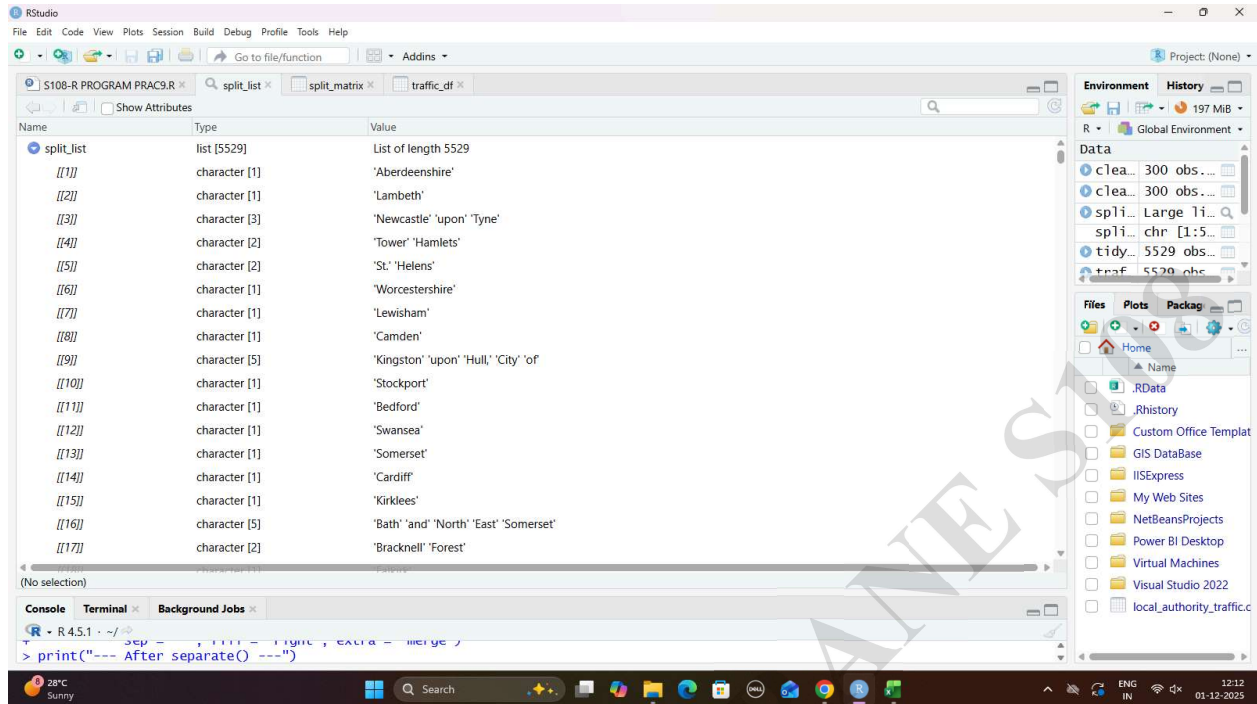
local_authority_name	word1	word2	word3
Aberdeenshire	Aberdeenshire		
Lambeth	Lambeth		
Newcastle upon Tyne	Newcastle	upon	Tyne
Tower Hamlets	Tower	Hamlets	
St. Helens	St.	Helens	
Worcestershire	Worcestershire		
Lewisham	Lewisham		
Camden	Camden		
Kingston upon Hull, City of	Kingston	upon	Hull,
Stockport	Stockport		
Bedford	Bedford		

**Bottom Screenshot:**

```
R - R4.5.1 - ~/...
[ reached 'max' / getoption("max.print") -- omitted 5279 rows ]
> tidy_name <- traffic_df %>%
+ separate(local_authority_name, into = c("Part1", "Part2", "Part3"),
+ sep = " ", fill = "right", extra = "merge")
> print("--- After separate() ---")
[1] "--- After separate() ---"
> print(tidy_name %>% select(Part1, Part2, Part3) %>% head())
```

Part1	Part2	Part3
Aberdeenshire	<NA>	<NA>
Lambeth	<NA>	<NA>
Newcastle upon Tyne	<NA>	<NA>
Tower Hamlets	<NA>	<NA>
St. Helens	<NA>	<NA>
Worcestershire	<NA>	<NA>

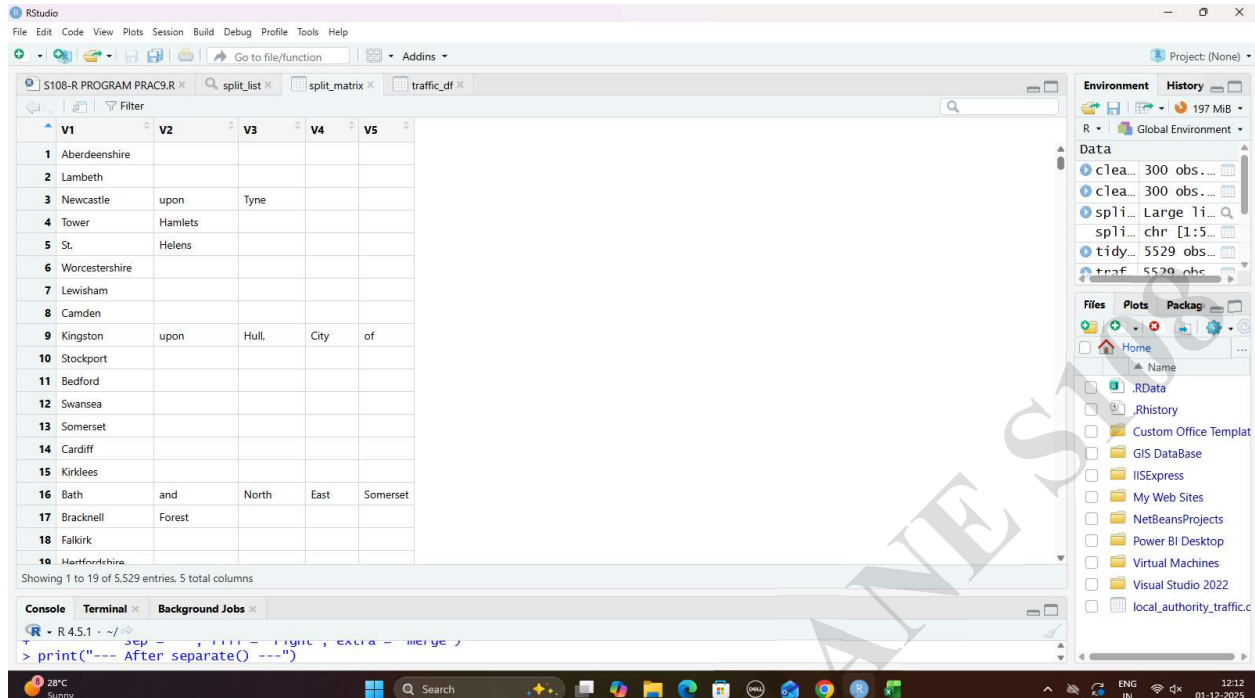
**SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE**  
**SUBJECT : Data Analysis with SAS / SPSS / R**





# SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

## SUBJECT : Data Analysis with SAS / SPSS / R



The first screenshot shows an RStudio window with a data table. The table has 5 columns labeled V1, V2, V3, V4, and V5. The rows contain data for various locations, including Aberdeenshire, Lambeth, Newcastle upon Tyne, Tower Hamlets, St. Helens, Worcestershire, Lewisham, Camden, Kingston upon Hull, City of, Stockport, Bedford, Swansea, Somerset, Cardiff, Kirklees, Bath and North East Somerset, Bracknell Forest, and Falkirk. The table is showing 1 to 19 of 5,529 entries.

V1	V2	V3	V4	V5
1	Aberdeenshire			
2	Lambeth			
3	Newcastle upon Tyne			
4	Tower Hamlets			
5	St. Helens			
6	Worcestershire			
7	Lewisham			
8	Camden			
9	Kingston upon Hull, City of			
10	Stockport			
11	Bedford			
12	Swansea			
13	Somerset			
14	Cardiff			
15	Kirklees			
16	Bath and North East Somerset			
17	Bracknell Forest			
18	Falkirk			
19	Hartfordshire			

The second screenshot shows the same RStudio window, but the data table is now expanded to show 13 columns: local\_authority\_id, local\_authority\_name, year, link\_length\_km, link\_length\_miles, cars\_and\_taxis, all\_motor\_vehicles, Name\_Start, Name\_End, and Year\_Start. The table is showing 1 to 18 of 5,529 entries.

local_authority_id	local_authority_name	year	link_length_km	link_length_miles	cars_and_taxis	all_motor_vehicles	Name_Start	Name_End	Year_Start
1	45 Aberdeenshire	2019	6273.378	3898.10	1537817161.0	2055244624	Abe	shire	2019
2	107 Lambeth	2019	377.000	234.26	397710910.8	547123377	Lam	mbeth	2019
3	172 Newcastle upon Tyne	2019	984.992	612.05	1006029025.0	1209012058	New	Tyne	2019
4	93 Tower Hamlets	2019	287.327	178.54	465293697.1	630938804	Tow	mlts	2019
5	158 St. Helens	2019	740.220	459.95	728175389.4	930780139	St.	elens	2019
6	69 Worcestershire	2019	4273.420	2655.38	3556865608.0	4546565782	Wor	shire	2019
7	104 Lewisham	2019	447.188	277.87	467457465.5	613517557	Lew	isham	2019
8	145 Camden	2019	279.342	173.58	218491001.6	297009025	Cam	amden	2019
9	169 Kingston upon Hull, City of	2019	772.361	479.92	827160827.6	1043436141	Kin	ty of	2019
10	56 Stockport	2019	1004.974	624.46	1161305176.0	1412753855	Sto	kport	2019
11	186 Bedford	2019	930.400	578.12	722473805.8	917378332	Bed	dford	2019
12	8 Swansea	2019	1167.030	725.16	983067119.3	1200222980	Swa	anseas	2019
13	5 Somerset	2019	6807.584	4230.04	3685156283.0	4704198951	Som	erset	2019
14	19 Cardiff	2019	1093.979	679.77	1734110249.0	2107567897	Car	rdiff	2019
15	197 Kirklees	2019	1909.033	1186.22	1518187481.0	1959431313	Kir	klees	2019
16	115 Bath and North East Somerset	2019	1079.621	670.85	628532401.9	780570733	Bat	erset	2019
17	180 Bracknell Forest	2019	474.714	294.97	413645269.3	497682894	Bra	orest	2019
18	30 Falkirk	2019	970.375	602.96	781875898.7	1015092641	Fal	lkirk	2019

SHUBHAM RAORANE S108

# SHETH L.U.J & SIR M.V COLLEGE OF SCIENCE

## SUBJECT : Data Analysis with SAS / SPSS / R

The screenshot displays the RStudio environment with a data frame loaded. The data frame has the following columns: year, link\_length\_km, link\_length\_miles, cars\_and\_taxis, all\_motor\_vehicles, Name\_Start, Name\_End, Year\_Str, Word1, Word2, and Word3. The data is filtered to show the first 18 entries out of 5,529 total entries.

year	link_length_km	link_length_miles	cars_and_taxis	all_motor_vehicles	Name_Start	Name_End	Year_Str	Word1	Word2	Word3
2019	6273.378	3898.10	1537817161.0	2055244624	Abe	shire	2019	Aberdeenshire		
2019	377.000	234.26	397710910.8	547123377	Lam	mbeth	2019	Lambeth		
2019	984.992	612.05	1006029025.0	1209012058	New	Tyne	2019	Newcastle	upon	Tyne
2019	287.327	178.54	465293697.1	630938804	Tow	mlets	2019	Tower	Hamlets	
2019	740.220	459.95	728175389.4	930780139	St.	elens	2019	St	Helens	
2019	4273.420	2655.38	3556865608.0	4546565782	Wor	shire	2019	Worcestershire		
2019	447.188	277.87	467457465.5	613517557	Lew	isham	2019	Lewisham		
2019	279.342	173.58	218491001.6	297009025	Cam	arden	2019	Camden		
2019	772.361	479.92	827160827.6	1043436141	Kin	ty of	2019	Kingston	upon	Hull,
2019	1004.974	624.46	1161305176.0	1412753855	Sto	kport	2019	Stockport		
2019	930.400	578.12	722473805.8	917378332	Bed	dford	2019	Bedford		
2019	1167.030	725.16	983067119.3	1200222980	Swa	ansea	2019	Swansea		
2019	6807.584	4230.04	3685156283.0	4704198951	Som	erset	2019	Somerset		
2019	1093.979	679.77	1734110249.0	2107567897	Car	rdiff	2019	Cardiff		
2019	1909.033	1186.22	1518187481.0	1959431313	Kir	klees	2019	Kirklees		
2019	1079.621	670.85	628532401.9	780570733	Bat	erset	2019	Bath	and	North
2019	474.714	294.97	413645269.3	497682894	Bra	orest	2019	Bracknell	Forest	
2019	970.375	602.96	781875898.7	1015092641	Fal	lkirk	2019	Falkirk		

Showing 1 to 18 of 5,529 entries. 13 total columns

Console Terminal Background Jobs

```
R - R 4.5.1 - /...  
> print("--- After separate() ---")
```