

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

SUBJECT : Data Analysis with SAS / SPSS / R

AIM : Selecting and dropping variables using select() in R. import dataset.

OUTPUT

The image displays two screenshots of the RStudio interface, showing the process of importing a dataset and performing variable selection and dropping operations.

Top Screenshot:

```
R 4.5.1 - ~/r/
> library(dplyr)
> nba <- read.csv("final_data.csv")
> print(head(nba, 3))
```

GAME_ID	TEAM_ID	TEAM_NAME	TEAM_ABBREVIATION	TEAM_CITY	HOME_TEAM	MIN	FGM	FGA	FG_PCT	FG3M	FG3A
21200001	1610612764	Wizards	WAS	Washington	Cavaliers	240:00	32	90	0.356	8	32
21200001	1610612739	Cavaliers	CLE	Cleveland	Cavaliers	240:00	36	79	0.456	7	20

```
> selected_basic <- nba %>%
+   select(GAME_ID, TEAM_ID, TEAM_NAME, FGM, FGA)
> print(head(selected_basic, 3))
```

GAME_ID	TEAM_ID	TEAM_NAME	FGM	FGA
21200001	1610612764	Wizards	32	90
21200001	1610612739	Cavaliers	36	79
21200002	1610612748	Heat	43	79

```
> selected_range <- nba %>%
+   select(FGM:FT_PCT)
> print(head(selected_range, 3))
```

FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT	FTM	FTA	FT_PCT
32	90	0.356	8	32	0.25	12	20	0.600
36	79	0.456	7	20	0.35	15	22	0.682
43	79	0.544	8	16	0.50	26	32	0.813

```
> selected_start_f <- nba %>%
+   select(starts_with("FG"))
> print(head(selected_start_f, 3))
```

Bottom Screenshot:

```
> print(head(selected_start_f, 3))
```

FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT
32	90	0.356	8	32	0.25
36	79	0.456	7	20	0.35
43	79	0.544	8	16	0.50

```
> dropped_team <- nba %>%
+   select(-TEAM_NAME)
> print(names(dropped_team))
```

[1]	"GAME_ID"	"TEAM_ID"	"TEAM_ABBREVIATION"	"TEAM_CITY"	"HOME_TEAM"
[6]	"MIN"	"FGM"	"FGA"	"FG_PCT"	"FG3M"
[11]	"FG3A"	"FG3_PCT"	"FTM"	"FTA"	"FT_PCT"
[16]	"OREB"	"DREB"	"REB"	"AST"	"STL"
[21]	"BLK"	"TO"	"PF"	"PTS"	"PLUS_MINUS"
[26]	"EFG_PCT"	"PIE"	"COVID_FLAG"	"RESULT"	"SEASON"
[31]	"WIN_PCT"				

```
> dropped_multiple <- nba %>%
+   select(-TEAM_NAME, -TEAM_ABBREVIATION, -HOME_TEAM)
> print(names(dropped_multiple))
```

[1]	"GAME_ID"	"TEAM_ID"	"TEAM_CITY"	"MIN"	"FGM"	"FGA"	"FG_PCT"
[8] <th>"FG3M"</th> <th>"FG3A"</th> <th>"FG3_PCT"</th> <th>"FTM"</th> <th>"FTA"</th> <th>"FT_PCT"</th> <th>"OREB"</th>	"FG3M"	"FG3A"	"FG3_PCT"	"FTM"	"FTA"	"FT_PCT"	"OREB"
[15] <th>"DREB"</th> <th>"REB"</th> <th>"AST"</th> <th>"STL"</th> <th>"BLK"</th> <th>"TO"</th> <th>"PF"</th>	"DREB"	"REB"	"AST"	"STL"	"BLK"	"TO"	"PF"
[22] <th>"PTS"</th> <th>"PLUS_MINUS"</th> <th>"EFG_PCT"</th> <th>"PIE"</th> <th>"COVID_FLAG"</th> <th>"RESULT"</th> <th>"SEASON"</th>	"PTS"	"PLUS_MINUS"	"EFG_PCT"	"PIE"	"COVID_FLAG"	"RESULT"	"SEASON"
[29] <th>"WIN_PCT"</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>	"WIN_PCT"						

```
> dropped_range <- nba %>%
+   select(-FGM:FT_PCT)
> print(names(dropped_range))
```

[1]	"GAME_ID"	"TEAM_ID"	"TEAM_NAME"	"TEAM_ABBREVIATION"	"TEAM_CITY"
[6] <th>"HOME_TEAM"</th> <th>"MIN"</th> <th>"OREB"</th> <th>"DREB"</th> <th>"REB"</th>	"HOME_TEAM"	"MIN"	"OREB"	"DREB"	"REB"
[11] <th>"AST"</th> <th>"STL"</th> <th>"BLK"</th> <th>"TO"</th> <th>"PF"</th>	"AST"	"STL"	"BLK"	"TO"	"PF"
[16] <th>"PTS"</th> <th>"PLUS_MINUS"</th> <th>"EFG_PCT"</th> <th>"PIE"</th> <th>"COVID_FLAG"</th>	"PTS"	"PLUS_MINUS"	"EFG_PCT"	"PIE"	"COVID_FLAG"
[21] <th>"RESULT"</th> <th>"SEASON"</th> <th>"WIN_PCT"</th> <th></th> <th></th>	"RESULT"	"SEASON"	"WIN_PCT"		

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

SUBJECT : Data Analysis with SAS / SPSS / R

The first screenshot shows the RStudio interface with the following data in the Environment pane:

Variable	Observations	Variables
dropped_m...	31340	29
dropped_r...	31340	23
dropped_t...	31340	31
final	3	3
final_list	5	3

The second screenshot shows the RStudio interface with the following data in the Environment pane:

Variable	Observations	Variables
dropped_m...	31340	29
dropped_r...	31340	23
dropped_t...	31340	31
final	3	3
final_list	5	3

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

SUBJECT : Data Analysis with SAS / SPSS / R

The first screenshot shows the RStudio interface with a data frame containing 18 rows of basketball statistics. The columns are FGM, FGA, FG_PCT, FG3M, FG3A, and FG3_PCT. The data is as follows:

	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT
1	32	90	0.356	8	32	0.250
2	36	79	0.456	7	20	0.350
3	43	79	0.544	8	16	0.500
4	39	75	0.520	6	13	0.462
5	38	77	0.494	3	13	0.231
6	40	85	0.471	5	15	0.333
7	33	91	0.363	6	17	0.353
8	37	78	0.474	5	11	0.455
9	33	88	0.375	4	18	0.222
10	30	85	0.353	7	25	0.280
11	39	79	0.494	10	27	0.370
12	35	79	0.443	6	16	0.375
13	34	84	0.405	6	17	0.353
14	33	79	0.418	2	9	0.222
15	35	80	0.438	4	17	0.235
16	38	80	0.475	6	20	0.300
17	32	85	0.376	11	22	0.500
18	38	89	0.427	6	14	0.429

The second screenshot shows the same RStudio interface but with a more complete data frame. The columns include GAME_ID, TEAM_ID, TEAM_CITY, MIN, FGM, FGA, FG_PCT, FG3M, FG3A, FG3_PCT, FTM, FTA, and FT_PCT. The data is as follows:

	GAME_ID	TEAM_ID	TEAM_CITY	MIN	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT	FTM	FTA	FT_PCT	OR
1	21200001	1610612764	Washington	240:00	32	90	0.356	8	32	0.250	12	20	0.600	
2	21200001	1610612739	Cleveland	240:00	36	79	0.456	7	20	0.350	15	22	0.682	
3	21200002	1610612748	Miami	240:00	43	79	0.544	8	16	0.500	26	32	0.813	
4	21200002	1610612738	Boston	240:00	39	75	0.520	6	13	0.462	23	28	0.821	
5	21200003	1610612747	Los Angeles	240:00	38	77	0.494	3	13	0.231	12	31	0.387	
6	21200003	1610612742	Dallas	240:00	40	85	0.471	5	15	0.333	14	18	0.778	
7	21200004	1610612761	Toronto	240:00	33	91	0.363	6	17	0.353	16	19	0.842	
8	21200004	1610612754	Indiana	240:00	37	78	0.474	5	11	0.455	11	16	0.688	
9	21200005	1610612743	Denver	240:00	33	88	0.375	4	18	0.222	5	11	0.455	
10	21200005	1610612755	Philadelphia	240:00	30	85	0.353	7	25	0.280	17	21	0.810	
11	21200006	1610612745	Houston	240:00	39	79	0.494	10	27	0.370	17	23	0.739	
12	21200006	1610612765	Detroit	240:00	35	79	0.443	6	16	0.375	20	26	0.769	
13	21200007	1610612758	Sacramento	240:00	34	84	0.405	6	17	0.353	13	16	0.813	
14	21200007	1610612741	Chicago	240:00	33	79	0.418	2	9	0.222	25	33	0.758	
15	21200008	1610612740	New Orleans	240:00	35	80	0.438	4	17	0.235	21	23	0.913	
16	21200008	1610612759	San Antonio	240:00	38	80	0.475	6	20	0.300	17	26	0.654	
17	21200009	1610612742	Dallas	240:00	32	85	0.376	11	22	0.500	19	26	0.731	

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

SUBJECT : Data Analysis with SAS / SPSS / R

The top screenshot shows the RStudio interface with a data frame containing 17 rows of game statistics. The columns are: GAME_ID, TEAM_ID, TEAM_NAME, TEAM_ABBREVIATION, TEAM_CITY, HOME_TEAM, MIN, OREB, DREB, REB, and AST. The data is filtered to show 17 of 31,340 entries.

GAME_ID	TEAM_ID	TEAM_NAME	TEAM_ABBREVIATION	TEAM_CITY	HOME_TEAM	MIN	OREB	DREB	REB	AST
1	21200001	1610612764	Wizards	Washington	Cavaliers	240:00	18	21	39	2
2	21200001	1610612739	Cavaliers	Cleveland	Cavaliers	240:00	18	36	54	2
3	21200002	1610612748	Heat	Miami	Heat	240:00	5	31	36	2
4	21200002	1610612738	Celtics	Boston	Heat	240:00	7	34	41	2
5	21200003	1610612747	Lakers	Los Angeles	Lakers	240:00	15	31	46	2
6	21200003	1610612742	Mavericks	Dallas	Lakers	240:00	9	31	40	2
7	21200004	1610612761	Raptors	Toronto	Raptors	240:00	15	27	42	1
8	21200004	1610612754	Pacers	Indiana	Raptors	240:00	9	37	46	2
9	21200005	1610612743	Nuggets	Denver	76ers	240:00	16	38	54	1
10	21200005	1610612755	76ers	Philadelphia	76ers	240:00	14	33	47	1
11	21200006	1610612745	Rockets	Houston	Pistons	240:00	12	33	45	2
12	21200006	1610612765	Pistons	Detroit	Pistons	240:00	10	26	36	2
13	21200007	1610612758	Kings	Sacramento	Bulls	240:00	11	29	40	1
14	21200007	1610612741	Bulls	Chicago	Bulls	240:00	14	32	46	2
15	21200008	1610612740	Hornets	New Orleans	Hornets	240:00	9	34	43	2
16	21200008	1610612759	Spurs	San Antonio	Hornets	240:00	10	33	43	2
17	21200009	1610612742	Mavericks	Dallas	Jazz	240:00	8	32	40	2

The bottom screenshot shows the same data frame with additional columns for player statistics: FGM, FGA, FG_PCT, FG3M, FG3A, and FG3_PCT. The data is filtered to show 17 of 31,340 entries.

GAME_ID	TEAM_ID	TEAM_ABBREVIATION	TEAM_CITY	HOME_TEAM	MIN	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT
1	21200001	1610612764	WAS	Washington	Cavaliers	240:00	32	90	0.356	8	32
2	21200001	1610612739	CLE	Cleveland	Cavaliers	240:00	36	79	0.456	7	20
3	21200002	1610612748	MIA	Miami	Heat	240:00	43	79	0.544	8	16
4	21200002	1610612738	BOS	Boston	Heat	240:00	39	75	0.520	6	13
5	21200003	1610612747	LAL	Los Angeles	Lakers	240:00	38	77	0.494	3	13
6	21200003	1610612742	DAL	Dallas	Lakers	240:00	40	85	0.471	5	15
7	21200004	1610612761	TOR	Toronto	Raptors	240:00	33	91	0.363	6	17
8	21200004	1610612754	IND	Indiana	Raptors	240:00	37	78	0.474	5	11
9	21200005	1610612743	DEN	Denver	76ers	240:00	33	88	0.375	4	18
10	21200005	1610612755	PHI	Philadelphia	76ers	240:00	30	85	0.353	7	25
11	21200006	1610612745	HOU	Houston	Pistons	240:00	39	79	0.494	10	27
12	21200006	1610612765	DET	Detroit	Pistons	240:00	35	79	0.443	6	16
13	21200007	1610612758	SAC	Sacramento	Bulls	240:00	34	84	0.405	6	17
14	21200007	1610612741	CHI	Chicago	Bulls	240:00	33	79	0.418	2	9
15	21200008	1610612740	NOH	New Orleans	Hornets	240:00	35	80	0.438	4	17
16	21200008	1610612759	SAS	San Antonio	Hornets	240:00	38	80	0.475	6	20
17	21200009	1610612742	DAL	Dallas	Jazz	240:00	32	85	0.376	11	22