

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

SUBJECT :- DATA ANALYSIS WITH SAS / SPSS / R

PRACTICAL – 4

AIM:- Applying conditional filters subset() or filter() in R.

OUTPUT:-

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R v4.5.2 - ~/
> olympics <- read_csv("C:\\Users\\Rohit\\Downloads\\olympics.csv")
Rows: 11115 Columns: 17
# Column specification
Delimiter: ",
chr (13): name, name_short, name_tv, gender, function, country_code, country, c...
dbl (3): code, height, weight
date (1): birth_date

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.
> head(olympics)
# A tibble: 6 x 17
  code name name_short name_tv gender function country_code country
<dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 1535420 VALENCIA Alejan... VALENCIA A Alejan... Female Athlete MEX Mexico
2 1535429 RUIZ Angela RUIZ A Angela... Female Athlete MEX Mexico
3 1535430 GRANDE Matias GRANDE M Matias... Male Athlete MEX Mexico
4 1535460 ROJAS Carlos ROJAS C Carlos... Male Athlete MEX Mexico
5 1535467 MARTINEZ WING B... MARTINEZ ... Bruno ... Male Athlete MEX Mexico
6 1535479 VAZQUEZ Ana VAZQUEZ A Ana VA... Female Athlete MEX Mexico

# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
> summary(olympics)
code name name_short name_tv
Min. :1532872 Length:11115 Length:11115 Length:11115
1st Qu.:1888186 Class:character Class:character Class:character
Median:1918842 Mode:character Mode:character Mode:character
Mean :1887426
3rd Qu.:1948950
Max. :9460001

gender function country_code country
Length:11115 Length:11115 Length:11115 Length:11115
Class:character Class:character Class:character Class:character
Mode:character Mode:character Mode:character Mode:character
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R v4.5.2 - ~/
gender function country_code country
Length:11115 Length:11115 Length:11115 Length:11115
Class:character Class:character Class:character Class:character
Mode:character Mode:character Mode:character Mode:character

country_full nationality nationality_full nationality_code
Length:11115 Length:11115 Length:11115 Length:11115
Class:character Class:character Class:character Class:character
Mode:character Mode:character Mode:character Mode:character

height weight disciplines events
Min. : 0.00 Min. : 0.000 Length:11115 Length:11115
1st Qu.: 0.00 1st Qu.: 0.000 Class:character Class:character
Median : 0.00 Median : 0.000 Mode:character Mode:character
Mean : 81.84 Mean : 2.214
3rd Qu.:177.00 3rd Qu.: 0.000
Max. :222.00 Max. :113.000
NA's :16

birth_date
Min. :1954-12-01
1st Qu.:1994-07-20
Median :1998-02-01
Mean :1997-07-07
3rd Qu.:2001-03-22
Max. :2012-08-11

> tall_athletes <- subset(olympics, height > 180)
> cat("Tall athletes (>180cm):", nrow(tall_athletes), "\n")
Tall athletes (>180cm): 2155
```

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

SUBJECT :- DATA ANALYSIS WITH SAS / SPSS / R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs

R - R4.5.2 - ~/R
> tall_athletes <- subset(olympics, height > 180)
> cat("Tall athletes (>180cm):", nrow(tall_athletes), "\n")
Tall athletes (>180cm): 2155
> summary(tall_athletes$height)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 181.0  185.0  189.0  190.1  194.0  222.0
> tall_heavy_athletes <- subset(olympics, height > 180 & weight > 80)
> cat("Tall & Heavy athletes:", nrow(tall_heavy_athletes), "\n")
Tall & Heavy athletes: 89
> head(tall_heavy_athletes)
# A tibble: 6 x 17
  code name      name_short name_tv gender `function` country_code country
<dbl> <chr>      <chr>      <chr> <chr> <chr>      <chr>      <chr>
1 1539950 MCNULTY Harry   MCNULTY H Harry   Male Athlete  IRL      Ireland
2 1539951 MOLLEN Bryan  MOLLEN B Bryan   Male Athlete  IRL      Ireland
3 1539952 KENNEDY Terry   KENNEDY T Terry   Male Athlete  IRL      Ireland
4 1540534 VERA FELD Santi... VERA FELD... Santi... Male Athlete  ARG      Argent...
5 1540536 OSADCZUK Matias OSADCZUK M Matias... Male Athlete  ARG      Argent...
6 1540537 ALVAREZ Santiago ALVAREZ S Santiago... Male Athlete  ARG      Argent...
# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
> female_or_very_tall <- subset(olympics, gender == "Female" | height > 190)
> cat("Female OR >190 cm:", nrow(female_or_very_tall), "\n")
Female OR >190 cm: 6243
> head(female_or_very_tall)
# A tibble: 6 x 17
  code name      name_short name_tv gender `function` country_code country
<dbl> <chr>      <chr>      <chr> <chr> <chr>      <chr>      <chr>
1 1535420 VALENCIA Alejan... VALENCIA A Alejan... Female Athlete  MEX      Mexico
2 1535429 RUIZ Angela    RUIZ A Angela... Female Athlete  MEX      Mexico
3 1536479 VAZQUEZ Ana    VAZQUEZ A Ana VA... Female Athlete  MEX      Mexico
4 1536632 FALLAH Mobina FALLAH M Mobina... Female Athlete  IRI      IR Iran
5 1538449 AMAISTROAIE Mad... AMAISTROA... Madali... Female Athlete  ROU      Romania
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs

R - R4.5.2 - ~/R
3 1536479 VAZQUEZ Ana    VAZQUEZ A Ana VA... Female Athlete  MEX      Mexico
4 1536632 FALLAH Mobina FALLAH M Mobina... Female Athlete  IRI      IR Iran
5 1538449 AMAISTROAIE Mad... AMAISTROA... Madali... Female Athlete  ROU      Romania
6 1541265 BARBELIN Lisa  BARBELIN L Lisa B... Female Athlete  FRA      France
# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
> mex_athletes <- olympics |>
+ filter(country_code == "MEX")
> cat("Mexican Athletes:", nrow(mex_athletes), "\n")
Mexican Athletes: 108
> heavy_male <- olympics |>
+ filter(gender == "Male", weight > 85)
> cat("Heavy Male Athletes:", nrow(heavy_male), "\n")
Heavy Male Athletes: 95
> head(heavy_male)
# A tibble: 6 x 17
  code name      name_short name_tv gender `function` country_code country
<dbl> <chr>      <chr>      <chr> <chr> <chr>      <chr>      <chr>
1 1539950 MCNULTY Harry   MCNULTY H Harry   Male Athlete  IRL      Ireland
2 1539951 MOLLEN Bryan  MOLLEN B Bryan   Male Athlete  IRL      Ireland
3 1540536 OSADCZUK Matias OSADCZUK M Matias... Male Athlete  ARG      Argent...
4 1540537 ALVAREZ Santiago ALVAREZ S Santiago... Male Athlete  ARG      Argent...
5 1540538 SCHULZ German  SCHULZ G German... Male Athlete  ARG      Argent...
6 1540603 ELIZALDE Tomas  ELIZALDE T Tomas... Male Athlete  ARG      Argent...
# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
> selected_countries <- olympics |>
+ filter(country_code %in% c("MEX", "USA", "IND"))
> cat("Athletes from selected countries:", nrow(selected_countries), "\n")
Athletes from selected countries: 840
> table(selected_countries$country_code)
```

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

SUBJECT :- DATA ANALYSIS WITH SAS / SPSS / R

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins Project: (None)

Source
Console Terminal Background Jobs

R - R4.5.2 - ~/
1 1539950 MCNULTY Harry MCNULTY H Harry Male Athlete IRL Ireland
2 1539951 MOLLEN Bryan MOLLEN B Bryan Male Athlete IRL Ireland
3 1540536 OSADCZUK Matias OSADCZUK M Matias Male Athlete ARG Argent...
4 1540537 ALVAREZ Santiago ALVAREZ S Santiago Male Athlete ARG Argent...
5 1540538 SCHULZ German SCHULZ G German Male Athlete ARG Argent...
6 1540603 ELIZALDE Tomas ELIZALDE T Tomas Male Athlete ARG Argent...
# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
> selected_countries <- olympics |>
+ filter(country_code %in% c("MEX", "USA", "IND"))
> cat("Athletes from selected countries:", nrow(selected_countries), "\n")
Athletes from selected countries: 840
> table(selected_countries$country_code)

IND MEX USA
112 108 620
> archery_athletes <- olympics |>
+ filter(grepl("Archery", disciplines))
> cat("Archery Athletes:", nrow(archery_athletes), "\n")
Archery Athletes: 128
> head(archery_athletes)
# A tibble: 6 x 17
   code name name_short name_tv gender `function` country_code country
   <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 1535420 VALENCIA Alejan... VALENCIA A Alejan... Female Athlete MEX Mexico
2 1535429 RUIZ Angela RUIZ A Angela... Female Athlete MEX Mexico
3 1535430 GRANDE Matias GRANDE M Matias... Male Athlete MEX Mexico
4 1536460 ROJAS Carlos ROJAS C Carlos... Male Athlete MEX Mexico
5 1536467 MARTINEZ WING B... MARTINEZ Bruno Male Athlete MEX Mexico
6 1536479 VAZQUEZ Ana VAZQUEZ A Ana VA... Female Athlete MEX Mexico
# 9 more variables: country_full <chr>, nationality <chr>,
# nationality_full <chr>, nationality_code <chr>, height <dbl>, weight <dbl>,
# disciplines <chr>, events <chr>, birth_date <date>
```