

PRACTICAL NO 4

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

The screenshot shows the RStudio interface with the following components:

- Console:** Displays the R code to load the 'hr' dataset from a CSV file. It shows the column specification and a preview of the first 6 rows of the data.
- Environment:** Lists the loaded objects: 'dh' (1000 obs. of 8 variables), 'high_income_subset' (281 obs. of 35 variables), 'hr' (1470 obs. of 35 variables), 'my_data' (1000 obs. of 8 variables), 'overtime_or_happy...' (733 obs. of 35 variables), 'StudentsPerforman...' (1000 obs. of 8 variables), and 'young_high_income...' (55 obs. of 35 variables).
- Files:** Shows the file explorer with various files and folders, including 'HR-Employee-Attrition.csv'.

```

R - R4.1.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # Load your HR dataset
> hr <- read_csv("Downloads/HR-Employee-Attrition.csv")
rows: 1470 columns: 35
-- Column specification -----
delimter: ","
chr (9): Attrition, BusinessTravel, Department, EducationField, Gender, Jobrole...
dbl (26): Age, DailyRate, DistanceFromHome, Education, EmployeeCount, EmployeeNu...

1 use 'spec()' to retrieve the full column specification for this data.
1 Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # quick look
> head(hr)
# A tibble: 6 x 35
  Age Attrition BusinessTravel DailyRate Department DistanceFromHome Education
<dbl> <chr> <chr> <dbl> <chr> <dbl> <dbl>
1 41 Yes Travel_Rarely 1102 Sales 8 1
2 49 No Travel_Frequently 279 Research & 8 1
3 37 Yes Travel_Rarely 1373 Research & 2 2
4 33 No Travel_Frequently 1392 Research & 3 4
5 27 No Travel_Rarely 591 Research & 2 1
6 32 No Travel_Frequently 1005 Research & 2 2
# 128 more variables: EducationField <chr>, EmployeeCount <dbl>,
# EmployeeNumber <dbl>, EnvironmentSatisfaction <dbl>, Gender <chr>,
# HourlyRate <dbl>, JobInvolvement <dbl>, JobLevel <dbl>, JobRole <chr>,
# JobSatisfaction <dbl>, MaritalStatus <chr>, MonthlyIncome <dbl>,
# MonthlyRate <dbl>, NumCompaniesWorked <dbl>, Over18 <chr>, OverTime <chr>,
# PercentSalaryHike <dbl>, PerformanceRating <dbl>,
# RelationshipsSatisfaction <dbl>, StandardHours <dbl>, ...
> # Example 1: Single condition
> # Filter employees with MonthlyIncome > 10000
> high_income_subset <- subset(hr, MonthlyIncome > 10000)
> cat("Number of high-income employees (MonthlyIncome > 10000):",
+ nrow(high_income_subset), "\n")
Number of high-income employees (MonthlyIncome > 10000): 281
> summary(high_income_subset$MonthlyIncome)

```

The screenshot shows the RStudio interface with the following components:

- Console:** Displays the R code to filter the 'hr' dataset based on multiple conditions. It shows the resulting subset and a summary of the data.
- Environment:** Lists the loaded objects: 'dh' (1000 obs. of 8 variables), 'high_income_subset' (281 obs. of 35 variables), 'hr' (1470 obs. of 35 variables), 'my_data' (1000 obs. of 8 variables), 'overtime_or_happy...' (733 obs. of 35 variables), 'StudentsPerforman...' (1000 obs. of 8 variables), and 'young_high_income...' (55 obs. of 35 variables).
- Files:** Shows the file explorer with various files and folders, including 'HR-Employee-Attrition.csv'.

```

R - R4.1.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
> # Filter employees with MonthlyIncome > 10000
> high_income_subset <- subset(hr, MonthlyIncome > 10000)
> cat("Number of high-income employees (MonthlyIncome > 10000):",
+ nrow(high_income_subset), "\n")
Number of high-income employees (MonthlyIncome > 10000): 281
> summary(high_income_subset$MonthlyIncome)
Min. 1st Qu. Median Mean 3rd Qu. Max.
10008 11245 14118 14826 17875 19999
> # Example 2: Multiple conditions (AND)
> # Employees with MonthlyIncome > 10000 AND YearsatCompany < 5
> young_high_income_subset <- subset(hr,
+ MonthlyIncome > 10000 & YearsatCompany < 5)
> cat("Employees earning >10000 AND less than 5 years in company:",
+ nrow(young_high_income_subset), "\n")
Employees earning >10000 AND less than 5 years in company: 55
> head(young_high_income_subset)
# A tibble: 6 x 35
  Age Attrition BusinessTravel DailyRate Department DistanceFromHome Education
<dbl> <chr> <chr> <dbl> <chr> <dbl> <dbl>
1 46 No Travel_Rarely 705 Sales 2 4
2 45 No Travel_Rarely 193 Research & 6 4
3 55 No Travel_Rarely 111 Sales 1 2
4 54 No Travel_Rarely 1217 Research & 2 4
5 59 No Non-Travel 1420 Human Reso- 2 4
6 51 Yes Travel_Frequently 1150 Research & 8 4
# 128 more variables: EducationField <chr>, EmployeeCount <dbl>,
# EmployeeNumber <dbl>, EnvironmentSatisfaction <dbl>, Gender <chr>,
# HourlyRate <dbl>, JobInvolvement <dbl>, JobLevel <dbl>, JobRole <chr>,
# JobSatisfaction <dbl>, MaritalStatus <chr>, MonthlyIncome <dbl>,
# MonthlyRate <dbl>, NumCompaniesWorked <dbl>, Over18 <chr>, OverTime <chr>,
# PercentSalaryHike <dbl>, PerformanceRating <dbl>,
# RelationshipsSatisfaction <dbl>, StandardHours <dbl>, ...
> # Example 3: Multiple conditions (OR)
> # Employees with OverTime == "Yes" OR JobSatisfaction > 3
> subset(hr, OverTime == "Yes" | JobSatisfaction > 3)

```

The screenshot displays the RStudio environment with the following components:

- Source Panel:** Contains R code for data analysis. The code includes comments about variables like `EducationField`, `EmployeeCount`, `Gender`, `HourlyRate`, `JobInvolvement`, `JobLevel`, `JobRole`, `JobSatisfaction`, `MaritalStatus`, `MonthlyIncome`, `MonthlyRate`, `NumCompaniesWorked`, `over18`, `overTime`, `PercentsalaryHike`, `PerformanceRating`, `Relationshipsatisfaction`, and `standardHours`. It also includes a multiple condition example for `overTime` and `JobSatisfaction`.
- Console Panel:** Shows the output of the R code, including a tibble with 6 rows and 35 columns. The tibble columns are `Age`, `Attrition`, `BusinessTravel`, `DailyRate`, `Department`, `DistanceFromHome`, and `Education`. The output shows data for employees with different travel frequencies and departments.
- Environment Panel:** Lists the objects in the R environment, including `Global Environment`, `dh` (1000 obs. of 8 variables), `high_income_subset` (281 obs. of 35 variables), `hr` (1470 obs. of 35 variables), `my_data` (1000 obs. of 8 variables), `overTime_or_happy` (733 obs. of 35 variables), `StudentsPerforman` (1000 obs. of 8 variables), and `young_high_income` (55 obs. of 35 variables).
- Files Panel:** Shows a list of files in the Downloads folder, including `Saurabh BI PRACTICAL V.pdf`, `SetupCrypTool_1_4_42_en.exe`, `Student Mental health.csv`, `StudentsPerformance - StudentsPerformance.csv`, `Unnati prac 1 pdf.odt`, `Unnati prac 1 pdf.pdf`, `Untitled document (1).pdf`, `Untitled document.pdf`, `WhatsApp Unknown 2025-11-18 at 9.41.15 AM`, `WhatsApp Unknown 2025-11-18 at 9.41.15 AM.zip`, `Whols`, `Whols.zip`, and `HR-Employee-Attrition.csv`.