

Sheth L.U.J & Sir M.V. College Of Science

Practical N0 4

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

The image displays two screenshots of the RStudio interface, demonstrating data filtering operations in R.

Top Screenshot: The console shows the loading of the `ai_job_market.csv` dataset. The data is a tibble with 2000 rows and 12 columns. The columns are: `job_id`, `company_name`, `industry`, `job_title`, `skills_required`, `experience_level`, `employment_type`, `location`, `salary_range_usd`, `posted_date`, `company_size`, and `tools_preferred`. The code filters for high-salary jobs (`salary_range_usd > 100000`) and displays the first six rows of the resulting dataset.

Bottom Screenshot: The console shows further filtering of the data. It filters for high-salary jobs with senior roles (`experience_level == "Senior"`) and displays the first six rows. It then filters for big companies (`company_size > 500`) and displays the first six rows. Finally, it filters for jobs located in the US (`location == "United States"`) and displays the first six rows.

The Environment pane on the right shows the following objects:

- `high_salary`: 1093 obs. of 12 variables
- `high_salary_senior`: 336 obs. of 12 variables
- `industry_filter`: 0 obs. of 12 variables
- `jobs`: 2000 obs. of 12 variables
- `low_occupants_filter`: 1386 obs. of 6 variables
- `my_data`: 3000 obs. of 18 variables
- `region_filter`: 3278 obs. of 6 variables
- `special_buildings`: 3007 obs. of 6 variables
- `us_jobs`: 10 obs. of 12 variables

Name :- Vinod Mali

Roll No :- S092

Subject :- Data Analysis with SAS / SPSS / R

Sheth L.U.J & Sir M.V. College Of Science

Practical N0 4

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for data manipulation, including filtering jobs by company size, location, and industry.
- Console:** Shows the output of the R code, including the number of big companies (2000) and the first 10 jobs located in the US.
- Environment:** Lists the objects created in the R session, such as `high_salary`, `high_salary_senior`, `industry_filter`, `jobs`, `low_occupants_filter`, `my_data`, `region_filter`, `special_buildings`, and `us_jobs`.
- Files:** Shows the file explorer with various files and folders, including `Practical_No_3_RR`, `Practical_No_4_RR`, and `Practical_No_5_RR`.

```
# company_size <- jobs[, tools_preferred <- chr]
> big_companies <- jobs[tools_preferred > 500,]
+ filter(company_size > 500)
> cat("Number of big companies:", nrow(big_companies), "\n")
Number of big companies: 2000
> head(big_companies)
# A tibble: 6 x 12
  job_id company_name industry job_title skills_required experience_level employment_type
  <dbl> <chr> <chr> <chr> <chr> <chr> <chr>
1 1 Foster and Sons Healthc. Data Ana. NumPy, Reinfor. Mid Full-time
2 2 Boyd, Myers and R. Tech Computer. Scikit-learn, Senior Full-time
3 3 King Inc. Tech Quant Re. MLFlow, FastAP. Entry Full-time
4 4 Cooper, Archer an. Tech AI Produ. Scikit-learn, Mid Full-time
5 5 Hall LLC Finance Data Sci. Excel, Keras, Senior Contract
6 6 Ellis PLC E-commo. AI Produ. GCP, Excel, Sc. Senior Remote

# 15 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
# company_size <chr>, tools_preferred <chr>
> us_jobs <- jobs[location == "USA",]
+ filter(grepl("USA|United States", location, ignore.case = TRUE))
> cat("Jobs located in the US:", nrow(us_jobs), "\n")
Jobs located in the US: 10
> head(us_jobs)
# A tibble: 6 x 12
  job_id company_name industry job_title skills_required experience_level employment_type
  <dbl> <chr> <chr> <chr> <chr> <chr> <chr>
1 143 Perez Ltd Healthc. AI Produ. MLFlow, LangCh. Entry Remote
2 291 Johnson, Humphrey Tech AI Resea. Excel, NumPy, Mid Contract
3 534 Davidson, Walsh a. Finance Data Ana. Keras, Python, Senior Internship
4 721 Shepard-Kelly Automot. Data Ana. TensorFlow, ML Entry Internship
5 764 Gray, Gonzalez an. Automot. AI Resea. PyTorch, Excel Mid Remote
6 778 Hall, Golden and Healthc. Quant Re. Hugging Face, Mid Full-time

# 15 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
# company_size <chr>, tools_preferred <chr>
> industry_filter <- jobs[industry %in% c("Artificial Intelligence", "Software"),]
+ filter(industry %in% c("Artificial Intelligence", "Software"))
> cat("Jobs in AI or Software industry:", nrow(industry_filter), "\n")
Jobs in AI or Software industry: 0
> table(industry_filter$industry)
<table of extent 0 >
>
```

Name :- Vinod Mali

Roll No :- S092

Subject :- Data Analysis with SAS / SPSS / R