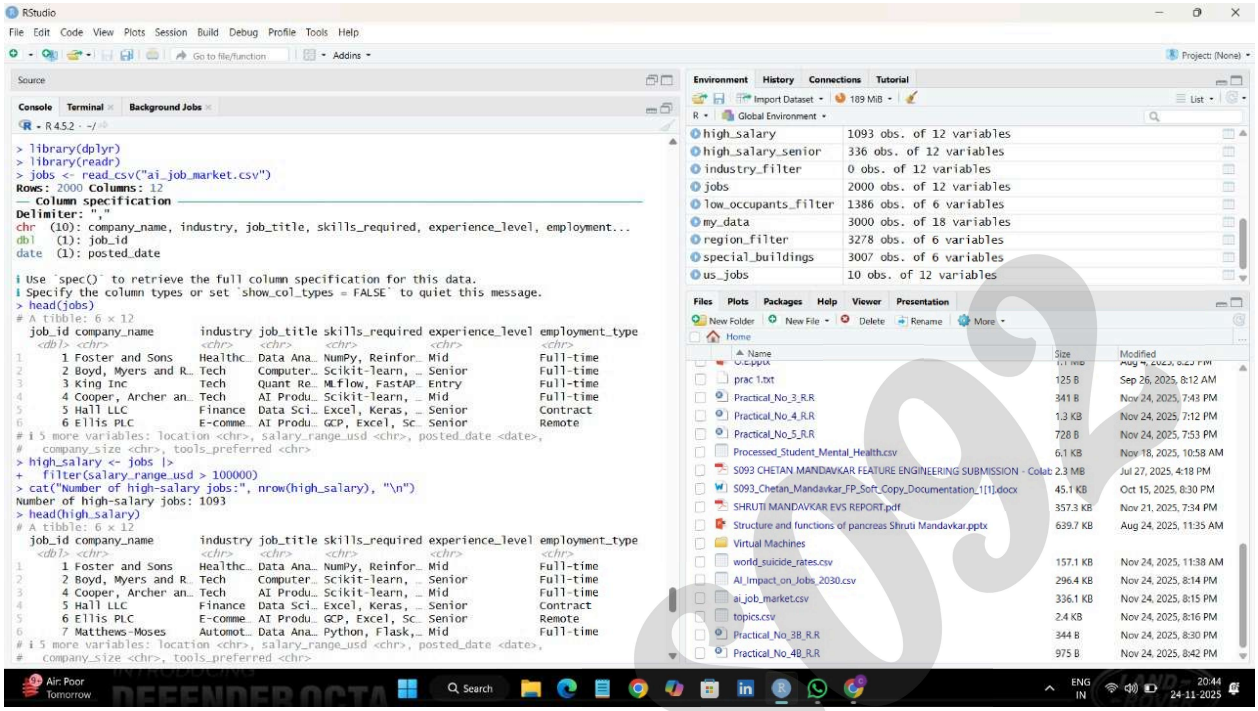
Sheth L.U.J & Sir M.V. College Of Science

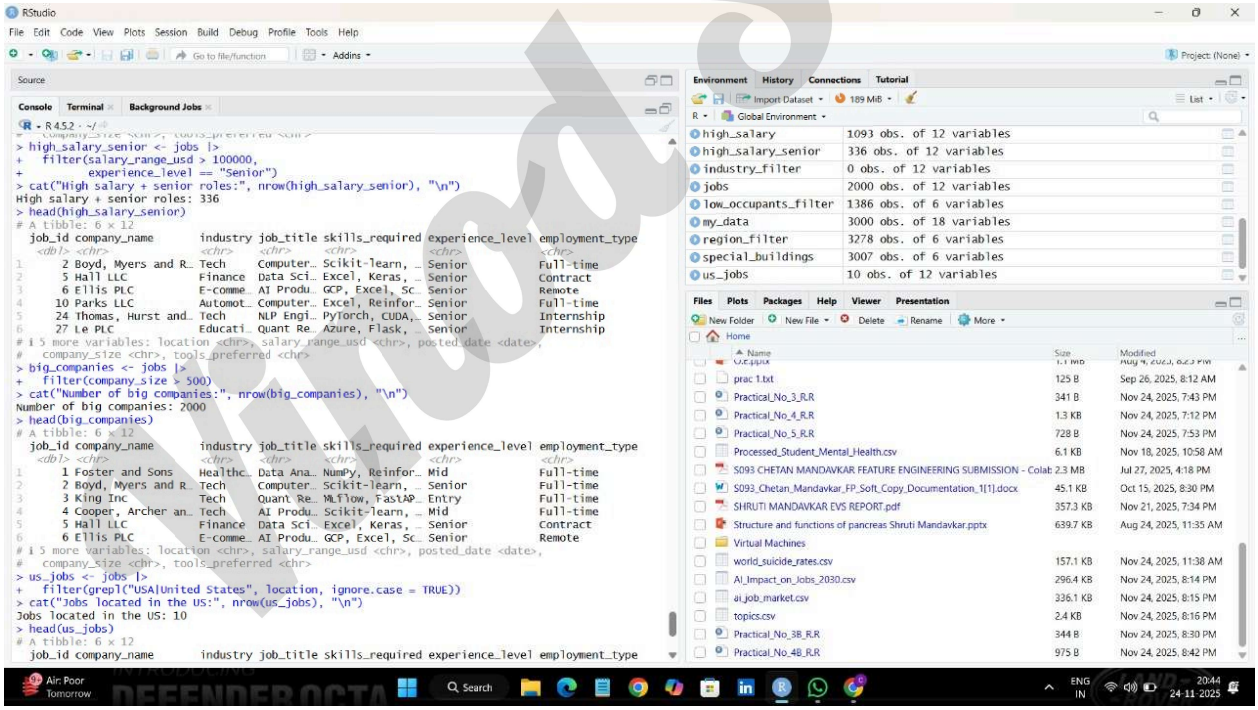
Practical N0 4

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



```
> library(dplyr)
> library(readr)
> jobs <- read_csv("ai_job_market.csv")
Rows: 2000 Columns: 12
── Column specification ──
Delimiter: ","
chr  (10): company_name, industry, job_title, skills_required, experience_level, employment...
dbl   (1): job_id
date  (1): posted_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(jobs)
# A tibble: 6 × 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-comme… AI Produ… GCP, Excel, Sc… Senior           Remote
# i 5 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
#   company_size <chr>, tools_preferred <chr>
> high_salary <- jobs |>
+   filter(salary_range_usd > 100000)
> cat("Number of high-salary jobs:", nrow(high_salary), "\n")
Number of high-salary jobs: 1093
> head(high_salary)
# A tibble: 6 × 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      4 Cooper, Archer an… Tech     AI Produ… Scikit-learn, … Mid              Full-time
4      5 Hall LLC           Finance  Data Sci… Excel, Keras, … Senior           Contract
5      6 Ellis PLC          E-comme… AI Produ… GCP, Excel, Sc… Senior           Remote
6      7 Matthews-Moses     Automot… Data Ana… Python, Flask,… Mid              Full-time
# i 5 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
#   company_size <chr>, tools_preferred <chr>
```



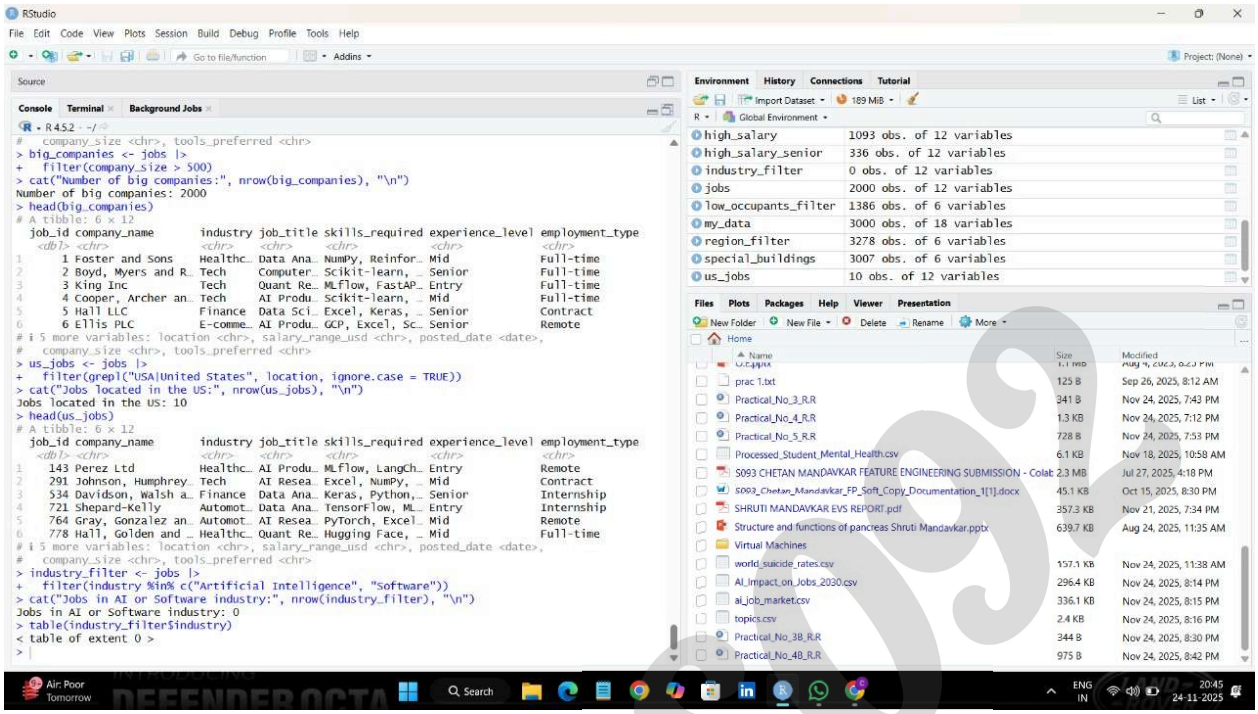
```
> high_salary_senior <- jobs |>
+   filter(salary_range_usd > 100000,
+          experience_level == "Senior")
> cat("High salary + senior roles:", nrow(high_salary_senior), "\n")
High salary + senior roles: 336
> head(high_salary_senior)
# A tibble: 6 × 12
  job_id company_name       industry job_title skills_required experience_level employment_type
   <dbl> <chr>              <chr>    <chr>     <chr>           <chr>            <chr>
1      2 Boyd, Myers and R… Tech     Computer… Scikit-learn, … Senior           Full-time
2      5 Hall LLC           Finance  Data Sci… Excel, Keras, … Senior           Contract
3      6 Ellis PLC          E-comme… AI Produ… GCP, Excel, Sc… Senior           Remote
4     10 Parks LLC          Automot… Computer… Excel, Reinfor… Senior           Full-time
5     24 Thomas, Hurst and… Tech     NLP Engi… PyTorch, CUDA,… Senior           Internship
6     27 Le PLC             Educati… Quant Re… Azure, Flask, … Senior           Internship
# i 5 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
#   company_size <chr>, tools_preferred <chr>
> big_companies <- jobs |>
+   filter(company_size > 500)
> cat("Number of big companies:", nrow(big_companies), "\n")
Number of big companies: 2000
> head(big_companies)
# A tibble: 6 × 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-comme… AI Produ… GCP, Excel, Sc… Senior           Remote
# i 5 more variables: location <chr>, salary_range_usd <chr>, posted_date <date>,
#   company_size <chr>, tools_preferred <chr>
> us_jobs <- jobs |>
+   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 × 12
  job_id company_name       industry job_title skills_required experience_level employment_type
```

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



Name :- Vinod Mali

Roll No :- S092

Subject :- Data Analysis with SAS / SPSS /R

Vinod S092