

SHETH LUJ AND SIR MV COLLEGE

The screenshot shows the RStudio interface. The script editor contains the following code:

```
library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

library(readr)

college <- read_csv("~/Nishita/data.csv")

Rows: 777 Columns: 18
#> Column specification
#> Delimiter: ";"
chr (1): private
dbl (17): apps, accept, enroll, top10perc, top25perc, f_undergrad, p_undergrad, outstate, ro...

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(college)
```

The Environment pane on the right shows the following objects:

- olympics_high_gold: 3 obs. of 5 variables
- olympics_multi_sort: 74 obs. of 5 variables
- olympics_sorted_gold: 74 obs. of 5 variables
- olympics_sorted_tota: 74 obs. of 5 variables
- pcos: 541 obs. of 45 variables
- PCOS_data: 541 obs. of 45 variables
- private_or_public: 777 obs. of 18 variables
- sales: 1000 obs. of 17 variables
- special_colleges: 578 obs. of 18 variables

The Files pane shows the following files:

- Athens 2004 Olympics Nations Medals.csv: 1 KB, Nov 24, 2025, 11:23 PM
- data.csv: 56.5 KB, Nov 24, 2025, 11:32 PM
- DS_5.py: 1.1 KB, Jul 22, 2025, 1:18 PM
- DS_P3_S080.py: 6 KB, Jul 9, 2025, 11:24 PM
- DS6_80.py: 4.3 KB, Aug 5, 2025, 9:22 AM
- ds6.py: 4.5 KB, Aug 5, 2025, 10:29 AM
- ds8.py: 674 B, Aug 20, 2025, 7:28 PM
- OS_P4.py: 1.2 KB, Jul 22, 2025, 1:37 PM
- P5_S080.R: 578 B, Nov 24, 2025, 11:28 PM
- S080_7.py: 440 B, Sep 16, 2025, 10:28 PM
- s080.py: 4.4 KB, Jul 29, 2025, 9:01 AM
- S080scala

The script editor contains the following code:

```
high_exp_subset <- subset(college, expend > 50000)

cat("Number of high-expenditure colleges:", nrow(high_exp_subset), "\n")
#> [1] 1

summary(high_exp_subset$expend)
#>   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.      <dbl>
#> 56233  56233   56233   56233   56233   56233   56233

high_exp_high_grad <- subset(college, expend > 50000 & grad_rate > 80)

cat("High expenditure & high graduation rate:", nrow(high_exp_high_grad), "\n")
#> [1] 1

head(high_exp_high_grad)
```

The Environment pane on the right shows the following objects:

- olympics_high_gold: 3 obs. of 5 variables
- olympics_multi_sort: 74 obs. of 5 variables
- olympics_sorted_gold: 74 obs. of 5 variables
- olympics_sorted_tota: 74 obs. of 5 variables
- pcos: 541 obs. of 45 variables
- PCOS_data: 541 obs. of 45 variables
- private_or_public: 777 obs. of 18 variables
- sales: 1000 obs. of 17 variables
- special_colleges: 578 obs. of 18 variables

The Files pane shows the following files:

- Athens 2004 Olympics Nations Medals.csv: 1 KB, Nov 24, 2025, 11:23 PM
- data.csv: 56.5 KB, Nov 24, 2025, 11:32 PM
- DS_5.py: 1.1 KB, Jul 22, 2025, 1:18 PM
- DS_P3_S080.py: 6 KB, Jul 9, 2025, 11:24 PM
- DS6_80.py: 4.3 KB, Aug 5, 2025, 9:22 AM
- ds6.py: 4.5 KB, Aug 5, 2025, 10:29 AM
- ds8.py: 674 B, Aug 20, 2025, 7:28 PM
- OS_P4.py: 1.2 KB, Jul 22, 2025, 1:37 PM
- P5_S080.R: 578 B, Nov 24, 2025, 11:28 PM
- S080_7.py: 440 B, Sep 16, 2025, 10:28 PM
- s080.py: 4.4 KB, Jul 29, 2025, 9:01 AM
- S080scala

SHETH LUJ AND SIR MV COLLEGE

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
P3_S120.R P4_S120.R P5_S080.R Athens_2004_Olympics_Nations_Medals
Source on Save Run Source
high_exp_high_grad <- subset(college, expend > 30000 & grad_rate > 50)
51:1 (Top Level) 2
R - R 4.5.2 - ~/
Console Terminal Background Jobs
> special_colleges <- subset(college, private == "Yes" | top10perc > 50)
> cat("Special colleges (Private OR top10perc > 50):", nrow(special_colleges), "\n")
Special colleges (Private OR top10perc > 50): 578
> head(special_colleges)
# A tibble: 6 x 18
  private apps accept enroll top10perc top25perc f_undergrad p_undergrad outstate room_board
<chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 Yes 1660 1232 721 23 52 2885 537 2440 3300
2 Yes 2186 1924 512 16 29 2683 1227 12280 6450
3 Yes 1428 1097 336 22 50 1036 99 11250 3750
4 Yes 417 349 137 60 89 510 63 12960 5450
5 Yes 193 146 55 16 44 249 869 2560 4120
6 Yes 587 479 158 38 62 678 41 13500 3335
# 18 more variables: books <dbl>, personal <dbl>, phd <dbl>, terminal <dbl>, s_f_ratio <dbl>,
# perc_alumni <dbl>, expend <dbl>, grad_rate <dbl>
>
> low_outstate <- college |>
+ filter(outstate < 10000)
>
> cat("Number of low outstate tuition colleges:", nrow(low_outstate), "\n")
Number of low outstate tuition colleges: 392
> summary(low_outstate$outstate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2340   5984   7348   7223   8684   9996
>
>
> large_undergrad <- college |>
+ filter(f_undergrad > 5000)
>
> cat("Large undergrad colleges (>5000 students):", nrow(large_undergrad), "\n")
Large undergrad colleges (>5000 students): 167
> head(large_undergrad)
# A tibble: 6 x 18
```

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
P3_S120.R P4_S120.R P5_S080.R Athens_2004_Olympics_Nations_Medals
Source on Save Run Source
high_exp_high_grad <- subset(college, expend > 30000 & grad_rate > 50)
51:1 (Top Level) 2
R - R 4.5.2 - ~/
Console Terminal Background Jobs
> private_or_public <- college |>
+ filter(private %in% c("Yes", "No"))
>
> cat("Total colleges:", nrow(private_or_public), "\n")
Total colleges: 777
> table(private_or_public$private)
No Yes
212 565
> head(college)
# A tibble: 6 x 18
  private apps accept enroll top10perc top25perc f_undergrad p_undergrad outstate room_board
<chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 No 2313 4664 1910 20 63 9940 1035 6806 2540
2 No 12809 10308 3761 24 49 22593 2585 7434 4850
3 No 7548 6791 3070 25 57 16262 1716 6300 3933
4 Yes 6075 5349 2367 34 66 9919 484 6450 3920
5 No 6773 3028 1025 15 55 5847 946 7844 2948
6 Yes 20192 13007 3810 45 80 14971 3113 18420 6810
# 18 more variables: books <dbl>, personal <dbl>, phd <dbl>, terminal <dbl>, s_f_ratio <dbl>,
# perc_alumni <dbl>, expend <dbl>, grad_rate <dbl>
>
> cat("Number of high-expenditure colleges:", nrow(high_exp_subset), "\n")
Number of high-expenditure colleges: 1
> cat("High expenditure & high graduation rate:", nrow(high_exp_high_grad), "\n")
High expenditure & high graduation rate: 1
```