

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE

SUBJECT - Data Analysis with SAS / SPSS / R

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

```
RStudio: R Console - Titanic
# Column specification
# Definition: "x"
# (3): Name, Sex, Ticket, Cabin, Embarked
# (7): PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare

# Use 'spec()' to retrieve the full column specification for this data.
# Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # quick look at the data structure
> head(titanic)
# A tibble: 6 x 12
  PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
  <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1 1 0 3 Braund, Mr. Owen Harris male 22 1 0 A/5 2. 7.25 NA S
2 2 1 1 Cumings, Mrs. John Bradley O'Leone Briggis Th. fema. 38 1 0 PC 17. 73.3 CRJ C
3 3 1 1 Heikkinen, Miss. Laina fema. 26 0 0 S309/. 51.0 NA S
4 4 1 1 Fretelle, Mrs. Jacques Heath (Lily May Peel) fema. 35 1 0 113803 53.1 C323 E
5 5 0 3 Allen, Mr. William Henry male 35 0 0 17559 8.00 NA E
6 6 0 3 Moran, Mr. James male NA 0 0 103077 8.00 NA Q

# High-fare subset
> high_fare_subset <- subset(titanic, Fare > 40)
> cat("Number of high-fare passengers (fare > 40):", nrow(high_fare_subset), "\n")
Number of high-fare passengers (fare > 40): 176
> summary(high_fare_subset$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  40.32  58.50  77.96 100.65 113.28 152.33

# Young high-fare subset
> young_high_fare_subset <- subset(titanic, Fare > 40 & Age < 30)
> cat("Number of young high-fare passengers:", nrow(young_high_fare_subset), "\n")
Number of young high-fare passengers: 62
> head(young_high_fare_subset)
# A tibble: 6 x 12
  PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
  <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1 25 0 1 Meyer, Mr. Edgar Joseph male 28 1 0 PC 17684 82.3 NA C
2 44 1 2 Laroché, Miss. Simonne Marie Rose Andree female 3 1 2 SC/Paris 21. 41.6 NA C
3 66 0 3 Goodwin, Master. William Frederick male 11 5 2 CA 2144 46.9 NA S
4 72 0 3 Goodwin, Miss. Lillian May female 16 5 2 CA 2144 46.9 NA S
5 73 0 2 Wood, Mr. Andrew Jr male 21 0 0 S.O.C. 14879 73.5 NA S

# Survivors or Firstclass
> survivors_or_firstclass <- subset(titanic, Survived == 1 | Pclass == 1)
> cat("Number of survivors or 1st-class passengers:", nrow(survivors_or_firstclass), "\n")
Number of survivors or 1st-class passengers: 422
> head(survivors_or_firstclass)
# A tibble: 6 x 12
  PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
  <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1 2 1 1 Cumings, Mrs. John Bradley O'Leone Briggis Th. fema. 38 1 0 PC 17. 71.3 CRJ C
2 3 1 1 Heikkinen, Miss. Laina fema. 26 0 0 S309/. 51.0 NA S
3 4 1 1 Fretelle, Mrs. Jacques Heath (Lily May Peel) fema. 35 1 0 113803 53.1 C323 S
4 7 0 1 McCarthy, Mr. Timothy J male 54 0 0 17463 31.9 E48 S
5 9 1 3 Johnson, Mrs. Oscar W (Elizabeth Wilhelmina Be. fema. 27 0 2 347742 51.1 NA S
6 10 1 2 Nassar, Mrs. Nicholas (Adèle Achon) fema. 14 1 0 210736 30.1 NA C

# Young passengers
> young_passengers <- titanic[ ]
> # Filter(Age < 14)
> cat("Number of passengers younger than 14:", nrow(young_passengers), "\n")
Number of passengers younger than 14: 71
> summary(young_passengers$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.420  2.008  4.800  5.082  8.000 13.808

# Female survivors
> female_survivors <- titanic[ ]
> # Filter(Sex == "female")
> cat("Number of female survivors:", nrow(female_survivors), "\n")
Number of female survivors: 231
> head(female_survivors)
# A tibble: 6 x 12
  PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked
  <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr> <chr>
1 2 1 1 Cumings, Mrs. John Bradley O'Leone Briggis Th. fema. 38 1 0 PC 17. 71.3 CRJ C
2 3 1 1 Heikkinen, Miss. Laina fema. 26 0 0 S309/. 51.0 NA S
3 4 1 1 Fretelle, Mrs. Jacques Heath (Lily May Peel) fema. 35 1 0 113803 53.1 C323 S
4 7 0 1 McCarthy, Mr. Timothy J male 54 0 0 17463 31.9 E48 S
5 9 1 3 Johnson, Mrs. Oscar W (Elizabeth Wilhelmina Be. fema. 27 0 2 347742 51.1 NA S
6 10 1 2 Nassar, Mrs. Nicholas (Adèle Achon) fema. 14 1 0 210736 30.1 NA C

# First or second class
> first_or_second_class <- titanic[ ]
> # Filter(pclass %in% c(1, 2))
> cat("Number of passengers in 1st or 2nd class:", nrow(first_or_second_class), "\n")
```