

R Packages Setup Guide for Data Science Course

Complete package installation and configuration

Installing R Packages for the Full Course

COMPLETE PACKAGE INSTALLATION:

Copy and run this code in R console:

```
``r
# Install all required packages (one-time setup)
packages <- c(
  # Core data manipulation
  "tidyverse",    # dplyr, ggplot2, tidyr, readr
  "dplyr",
  "tidyr",
  "readr",

  # Visualization
  "ggplot2",
  "gridExtra",
  "ggthemes",
  "RColorBrewer",
  "corrplot",
  "GGally",

  # Data exploration
  "skimr",
  "DataExplorer",
  "VIM",

  # Dates and times
  "lubridate",

  # Statistics
  "moments",
```

```

"lme4",
"nlme",
"car",

# Machine learning
"caret",
"rpart",
"rpart.plot",
"randomForest",
"e1071",
"class",
"cluster",
"clusterSim",

# Time series
"forecast",
"fable",
"tsibble",

# Utility
"here",
"rmarkdown",
"knitr",
"writexl",
"readxl"
)

# Install all
install.packages(packages)

# Load and check
sapply(packages, library, character.only = TRUE)
```

```

PACKAGE BY WEEK:

Week 1-5 Packages:

```

```r
install.packages(c(
  "tidyverse",
  "ggplot2",
  "dplyr",

```

```
"tidyr",  
"readr",  
"lubridate"  
)  
...
```

Week 6-8 (Statistics):

```
``r  
install.packages(c(  
  "moments",  
  "car",  
  "lmtest",  
  "effsize",  
  "tseries"  
)  
...
```

Week 9-10 (Machine Learning):

```
``r  
install.packages(c(  
  "caret",  
  "rpart",  
  "rpart.plot",  
  "randomForest",  
  "e1071",  
  "class",  
  "cluster"  
)  
...
```

Week 11 (Time Series):

```
``r  
install.packages(c(  
  "forecast",  
  "fable",  
  "tsibble"  
)  
...
```

TROUBLESHOOTING:

If package installation fails:

1. Check internet connection
2. Update R to latest version
3. Try RStudio console install button
4. Use binary installation (often faster):
``install.packages("package_name", type = "binary")``
5. For Mac, may need Xcode command line tools:
``xcode-select --install``
6. For Linux, install development headers:
``sudo apt install r-base-dev``

Common Issues:

Issue: "Package not found"

Solution: Check spelling, use quotes

Issue: "Dependency X also needs to install"

Solution: Say yes, let it install dependencies

Issue: "Permission denied"

Solution: Run RStudio as administrator

LOADING PACKAGES IN EVERY SCRIPT:

Best practice - start every analysis with:

```
``r
# =====
# Load required packages
# =====
library(tidyverse)  # Main package
library(ggplot2)    # Visualization
library(dplyr)      # Data manipulation
library(readr)      # Reading data
library(lubridate)  # Dates
library(caret)      # ML
library(forecast)   # Time series

# Set seed for reproducibility
set.seed(42)

# Your analysis code follows...
```

...

PACKAGE VERSIONS:

Check installed version:

```
``r  
packageVersion("tidyverse")  
...
```

Update package:

```
``r  
install.packages("tidyverse")  
...
```

Update all packages:

```
``r  
update.packages()  
...
```

Create reproducible environment file:

```
``r  
# Save your environment  
writeLines(  
  capture.output(sessionInfo()),  
  "session_info.txt"  
)  
...
```

QUICK REFERENCE - MAIN PACKAGES:

tidyverse:

- dplyr: Data manipulation
- ggplot2: Visualization
- tidyr: Data reshaping
- readr: Reading data

caret:

- Pre/post processing
- Model training
- Cross-validation
- Feature selection

forecast:

- Time series analysis
- ARIMA models
- Forecasting
- Accuracy metrics

rpart + rpart.plot:

- Decision trees
- Tree visualization

randomForest:

- Random forest models
- Ensemble methods
- Feature importance

OPTIONAL ADVANCED PACKAGES:

```
```r
```

```
For advanced ML
```

```
install.packages("xgboost")
```

```
install.packages("neuralnet")
```

```
For data validation
```

```
install.packages("validate")
```

```
For interactive plots
```

```
install.packages("plotly")
```

```
install.packages("shiny")
```

```
For parallel computing
```

```
install.packages("parallel")
```

```
For web scraping
```

```
install.packages("rvest")
```

```
install.packages("jsonlite")
```

```
```
```