Capacity Building in Seasonal Hydrological Forecasting

Introduction to R and Programming Basics

AGRHYMET, Climate Regional Center for West-Africa and Sahel

@Arsène KIEMA

2025-10-08

Objective

> **i** Learning outcomes
>
> This page provides a curated list of **relevant R documentation and learning resources**
> to help participants strengthen their programming and data analysis skills after the training.
> All links below are **clickable** and lead directly to the official documentation or free online books. nvb bn

1. Introduction to R Programming

> **i** Start here
>
> If you are new to R, begin with these essential resources to learn the fundamentals of programming, data manipulation, and visualization.

- **R for Data Science (2nd Edition)** — *Hadley Wickham & Garrett Grolemund*
  The most accessible and modern book to start learning R, covering programming, tidy data, and visualization.

1. Introduction to R Programming

> **i** Start here
>
> If you are new to R, begin with these essential resources to learn the fundamentals of programming, data manipulation, and visualization.

- **R for Data Science (2nd Edition)** — *Hadley Wickham & Garrett Grolemund*
  The most accessible and modern book to start learning R, covering programming, tidy data, and visualization.

- **RStudio Cheat Sheets Collection**
  Official Posit (RStudio) quick reference sheets for all major R packages (dplyr, ggplot2, tidyr, etc.).

1. Introduction to R Programming

> **i** Start here
>
> If you are new to R, begin with these essential resources to learn the fundamentals of programming, data manipulation, and visualization.

- **Hands-On Programming with R (O'Reilly)**
  A beginner-friendly introduction to R programming through practical coding exercises.

1. Introduction to R Programming

> **ℹ Start here**
>
> If you are new to R, begin with these essential resources to learn the fundamentals of programming, data manipulation, and visualization.

- **Hands-On Programming with R (O'Reilly)**
  A beginner-friendly introduction to R programming through practical coding exercises.

- **CRAN R Manuals**
  The official documentation of R from the Comprehensive R Archive Network (CRAN).

2. Data Import and Manipulation

> **Goal**
>
> Learn how to import, clean, and transform your tabular data for analysis or modeling.

- **Tidyverse Packages**
  Official website listing all packages for data wrangling (`readr`, `dplyr`, `tidyr`, etc.).

2. Data Import and Manipulation

> 💡 Goal
>
> Learn how to import, clean, and transform your tabular data for analysis or modeling.

- **Tidyverse Packages**
  Official website listing all packages for data wrangling (`readr`, `dplyr`, `tidyr`, etc.).

- **R for Data Science: Data Wrangling Chapters**
  How to reshape, summarize, and manipulate datasets efficiently.

2. Data Import and Manipulation

> 💡 Goal
>
> Learn how to import, clean, and transform your tabular data for analysis or modeling.

- **Data Import in R — RStudio Guide**
  Explains how to load data from CSV, Excel, and databases using `readr` and `readxl`.

2. Data Import and Manipulation

> 💡 Goal
>
> Learn how to import, clean, and transform your tabular data for analysis or modeling.

- **Data Import in R — RStudio Guide**
  Explains how to load data from CSV, Excel, and databases using `readr` and `readxl`.

- **Dplyr Package Documentation**
  Core tool for manipulating tabular data with readable syntax (`filter`, `mutate`, `summarise`, `group_by`).

3. Spatial and Spatio-temporal Data

i Essential for hydrology and climate data

These resources cover handling shapefiles, rasters, and NetCDF time series for hydrological applications.

- **Simple Features for R (sf)** — *Pebesma, 2018*
  The modern spatial data standard in R for vector data (points, lines, polygons).

3. Spatial and Spatio-temporal Data

> **i** Essential for hydrology and climate data
>
> These resources cover handling shapefiles, rasters, and NetCDF time series for hydrological applications.

- **Simple Features for R (sf)** — *Pebesma, 2018*
  The modern spatial data standard in R for vector data (points, lines, polygons).

- **Stars: Spatio-Temporal Arrays**
  Manage, analyze, and visualize raster and time-series gridded data (NetCDF, GeoTIFF, GRIB).

3. Spatial and Spatio-temporal Data

> **i** Essential for hydrology and climate data
>
> These resources cover handling shapefiles, rasters, and NetCDF time series for hydrological applications.

- **Geocomputation with R** — *Lovelace, Nowosad, Muenchow (2019)*
  A full open-access book covering GIS, remote sensing, and spatial analysis in R.

3. Spatial and Spatio-temporal Data

i Essential for hydrology and climate data

These resources cover handling shapefiles, rasters, and NetCDF time series for hydrological applications.

- **Geocomputation with R** — *Lovelace, Nowosad, Muenchow (2019)*
  A full open-access book covering GIS, remote sensing, and spatial analysis in R.

- **CRAN Task View: Analysis of Spatial Data**
  Overview of all CRAN packages related to spatial data analysis.

4. Data Visualization in R

> 💡 **Goal**
>
> Visualize hydrological data effectively to reveal trends, patterns, and anomalies.

- **ggplot2: Elegant Graphics for Data Analysis** — *Hadley Wickham*
  The official book describing the grammar of graphics and advanced visualization techniques.

4. Data Visualization in R

> 💡 **Goal**
>
> Visualize hydrological data effectively to reveal trends, patterns, and anomalies.

- **ggplot2: Elegant Graphics for Data Analysis** — *Hadley Wickham*
  The official book describing the grammar of graphics and advanced visualization techniques.

- **R Graphics Cookbook (2nd Edition)** — *Winston Chang*
  Practical examples for histograms, boxplots, time series, and geospatial graphics.

4. Data Visualization in R

Goal

Visualize hydrological data effectively to reveal trends, patterns, and anomalies.

- **ggplot2 Cheat Sheet (RStudio)**
  Quick reference for creating plots with layers, themes, and scales.

4. Data Visualization in R

> 💡 **Goal**
>
> Visualize hydrological data effectively to reveal trends, patterns, and anomalies.

- **ggplot2 Cheat Sheet (RStudio)**
  Quick reference for creating plots with layers, themes, and scales.

- **Interactive Visualization with plotly in R**
  Create interactive plots and dashboards directly from R.

5. Statistical Analysis and Time Series

i Core for hydrological forecasting

These materials explain how to analyze trends, correlations, and time-dependent patterns in hydrological data.

- **Forecasting: Principles and Practice (3rd Edition)** — *Hyndman & Athanasopoulos*
  The definitive free resource for time series analysis and forecasting using R.

5. Statistical Analysis and Time Series

> **i** Core for hydrological forecasting
>
> These materials explain how to analyze trends, correlations, and time-dependent patterns in hydrological data.

- **Forecasting: Principles and Practice (3rd Edition)** — *Hyndman & Athanasopoulos*
  The definitive free resource for time series analysis and forecasting using R.

- **An Introduction to Statistical Learning (ISLR)** — *James, Witten, Hastie & Tibshirani*
  Covers regression, classification, and model evaluation — foundation for Ridge, Lasso, and PCR.

5. Statistical Analysis and Time Series

> **i** Core for hydrological forecasting
>
> These materials explain how to analyze trends, correlations, and time-dependent patterns in hydrological data.

- **TidyModels Learn Center**
  Tutorials on data splitting, preprocessing (`recipes`), model training, tuning, and evaluation.

5. Statistical Analysis and Time Series

> **i** Core for hydrological forecasting
>
> These materials explain how to analyze trends, correlations, and time-dependent patterns in hydrological data.

- **TidyModels Learn Center**
  Tutorials on data splitting, preprocessing (`recipes`), model training, tuning, and evaluation.

- **R Time Series Task View (CRAN)**
  List of R packages dedicated to time series modeling and analysis.

6. Machine Learning in R

> **ℹ Modern workflows**
>
> Resources for supervised learning, cross-validation, and model deployment using R.

- **Tidy Modeling with R** — *Max Kuhn & Julia Silge (2022)*
  The official book for modern machine learning using the `tidymodels` ecosystem.

6. Machine Learning in R

> **ℹ Modern workflows**
>
> Resources for supervised learning, cross-validation, and model deployment using R.

- **Tidy Modeling with R** — *Max Kuhn & Julia Silge (2022)*
  The official book for modern machine learning using the `tidymodels` ecosystem.

- **Caret Package Guide**
  Classic machine learning package, still widely used for regression and classification.

6. Machine Learning in R

> **ℹ Modern workflows**
>
> Resources for supervised learning, cross-validation, and model deployment using R.

- **Ranger: Fast Implementation of Random Forests**
  The official documentation for the `ranger` package used in this training.

6. Machine Learning in R

> **i** Modern workflows
>
> Resources for supervised learning, cross-validation, and model deployment using R.

- **Ranger: Fast Implementation of Random Forests**
  The official documentation for the `ranger` package used in this training.

- **StatQuest (YouTube)**
  Simplified explanations of Ridge, Lasso, Random Forests, and other ML algorithms.

7. Reproducible Reports and Quarto

> 💡 **For professional reporting**
>
> Learn how to create reproducible scientific documents and presentations with R and Quarto.

- **Quarto Documentation**
  Official site — explains syntax, YAML configuration, code blocks, and rendering to HTML/PDF.

7. Reproducible Reports and Quarto

> 💡 For professional reporting
>
> Learn how to create reproducible scientific documents and presentations with R and Quarto.

- **Quarto Documentation**
  Official site — explains syntax, YAML configuration, code blocks, and rendering to HTML/PDF.

- **Quarto for Scientists Workshop**
  Short and effective introduction to Quarto for research and education.

7. Reproducible Reports and Quarto

> 💡 **For professional reporting**
>
> Learn how to create reproducible scientific documents and presentations with R and Quarto.

- **R Markdown and Quarto Cheat Sheet**
  Quick reference for formatting, code chunks, and exporting documents.

8. Advanced Topics and Good Practices

> 💡 **For further development**
>
> Recommended for participants who want to go beyond the basics and start automating or packaging their work.

- **Efficient R Programming** — *Gillespie & Lovelace*
  How to write faster and more efficient R code.

8. Advanced Topics and Good Practices

> 💡 For further development
>
> Recommended for participants who want to go beyond the basics and start automating or packaging their work.

- **Efficient R Programming** — *Gillespie & Lovelace*
  How to write faster and more efficient R code.

- **R Packages Book** — *Hadley Wickham*
  Step-by-step guide for building and maintaining R packages (useful for WASS2SHydroR).

8. Advanced Topics and Good Practices

> 💡 For further development
>
> Recommended for participants who want to go beyond the basics and start automating or packaging their work.

- **Targets Workflow Management**
  Automate and parallelize R pipelines — ideal for reproducible hydrological workflows.

8. Advanced Topics and Good Practices

💡 For further development

Recommended for participants who want to go beyond the basics and start automating or packaging their work.

- **Targets Workflow Management**
  Automate and parallelize R pipelines — ideal for reproducible hydrological workflows.

- **RStudio Community**
  Ask questions, share experiences, and get support from other R users.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.
2. Learn visualization with **ggplot2**.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.
2. Learn visualization with **ggplot2**.
3. Explore **spatial data** using sf and stars.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.
2. Learn visualization with **ggplot2**.
3. Explore **spatial data** using `sf` and `stars`.
4. Study **statistical modeling and ML** with ISLR and TidyModels.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.
2. Learn visualization with **ggplot2**.
3. Explore **spatial data** using sf and stars.
4. Study **statistical modeling and ML** with ISLR and TidyModels.
5. Practice **forecasting and time series** with FPP3.

9. Suggested Learning Path

1. Start with **R for Data Science** to master programming and data manipulation.
2. Learn visualization with **ggplot2**.
3. Explore **spatial data** using sf and stars.
4. Study **statistical modeling and ML** with ISLR and TidyModels.
5. Practice **forecasting and time series** with FPP3.
6. Document your work with **Quarto**.

10. Downloadable Resources

Resource	Link
RStudio Cheat Sheets (PDF Collection)	Download here
Geocomputation with R (Free PDF)	Download here
Forecasting: Principles and Practice (PDF)	Download here
An Introduction to Statistical Learning (PDF)	Download here

Final Note

> **!** Keep learning!
>
> R is a powerful and evolving ecosystem.
> The more you practice, the more intuitive it becomes.
> Bookmark this page and revisit it often to explore new capabilities.

THANK YOU FOR YOUR ATTENTATION