# Online Supplementary Materials

A Checklist for Deliberation Process

Table A.1: Checklist with Decision Rationale

|  |  |  |
| --- | --- | --- |
| Step | Checklist | Notes |
| 1.Clarify Conceptual Construct | Literature review and shared across the team. | Notes from team dis- cussion |
| Confirm shared understanding of theoretical con- struct. |
| Relevant theoretical dimensions are discussed and documented. |
| 2.Document Research Goals | Instructions on key variable formats and downstream analytical needs. | Update when neces- sary |
| Review initial codebook. |
| Data input training. |
| 3.Assign Data Collection | Assign datasets to team members by geography or source. | Document ambigu- ous items |
| Each team member maintains a seperate sheet for raw data collection and a log of decisions. |
| 4.Dual-Entry and Cross-Check | Conduct dual entry by second team member. | Record discrepancies found |
| Discrepancies flagged and logged for group discussion. |
| 5.Deliberation on Discrepancies | Team discussion on discrepancies. | Provide examples of key disputes and how they were resolved |
| Items with unclear mapping to conceptual dimensions are categorized or excluded. |
| Update codebook/documentation. |
| 6.Log Data and Decision | Finalize data by cross-check of all members | Mention major up- dates. |
| Log all decisions and changes in version-controlled repository (e.g., OSF, GitHub). |

Table A.1 is a checklist with notes or rationales for key decisions made during the deliberation process. The focus on each step may vary depending on the research purpose. For example, in public opinion harmonization projects like DCPO, more time is typically de- voted to conceptualization and construct development compared to administrative data projects such as SWIID. However, this general checklist can serve as a useful guide across a range of harmonization efforts.

# B R packages for data wrangling

Here are exemplary R packages that researchers can use to collect, clean, and transform data. The following tables were generated by the pkgsearch::pkg\_search() function with the keywords relating to data downloading, wrangling, and transforming. The pack- ages are ranked based on the ‘score’ metric that reflects both textual relevances with the keyword and package popularity in the last month. Only the top twenty packages and only the maintainers’ names are shown. We encourage readers to use the codes in this paper’s replication file to explore more useful packages. We also recommend readers to refer to the [“CRAN Task View: Reproducible Research”](https://cran.r-project.org/web/views/ReproducibleResearch.html) page for more useful tools to achieve the first-order and second-order opening.

Table A.2: Example packages for downloading data with API

package title maintainer

giscoR Download Map Data from GISCO API - Eurostat Diego Hernangómez rwebstat Download Data from the Webstat API Vincent Guegan

crypto2 Download Crypto Currency Data from ’CoinMarket- Sebastian Stoeckl

Cap’ without ’API’

RKaggle ’Kaggle’ Dataset Downloader ’API’ Benjamin Smith csodata Download Data from the CSO ’PxStat’ API Conor Crowley

hansard Provides Easy Downloading Capabilities for the UK Evan Odell

Parliament API

cranlogs Download Logs from the ’RStudio’ ’CRAN’ Mirror Gábor Csárdi

clinicalomicsdbR Interface with the ’ClinicalOmicsDB’ API, Allowing

for Easy DataDownloading and Importing

wdi2 Download World Development Indicators from the

World BankIndicators API

GDELTtools Download, Slice, and Normalize GDELT V1 Event

and Sentiment APIData

John Elizarraras

Christoph Scheuch Stephen R. Haptonstahl

neonUtilities Utilities for Working with NEON Data Claire Lunch piggyback Managing Larger Data on a GitHub Repository Carl Boettiger Quandl API Wrapper for Quandl.com Dave Dotson

nasapower NASA POWER API Client Adam H. Sparks

rstudioapi Safely Access the RStudio API Kevin Ushey

rdhs API Client and Dataset Management for the Demo- OJ Watson

graphic and HealthSurvey (DHS) Data

fishtree Interface to the Fish Tree of Life API Jonathan Chang ridigbio Interface to the iDigBio Data API Jesse Bennett

tradestatistics Open Trade Statistics API Wrapper and Utility Pro- Mauricio Vargas

gram

FlickrAPI Access to Flickr API Koki Ando

easycensus Quickly Find, Extract, and Marginalize U.S. Census Cory McCartan

Tables

shutterstock Access ’Shutterstock’ REST API Metin Yazici

Continued on next page

Table A.2: Example packages for downloading data with API (Continued)

wbstats Programmatic Access to Data and Statistics from the

World BankAPI

gwasrapidd ’REST’ ’API’ Client for the ’NHGRI’-’EBI’ ’GWAS’

Catalog

Jesse Piburn Ramiro Magno

ecos Economic Statistics System of the Bank of Korea Seokhoon Joo rscopus Scopus Database ’API’ Interface John Muschelli

I14Y Search and Get Data from the I14Y Interoperability Felix Luginbuhl

Platform ofSwitzerland

riingo An R Interface to the ’Tiingo’ Stock Price API Davis Vaughan

kaigiroku Programmatic Access to the API for Japanese Diet

Proceedings

PurpleAir Query the ’PurpleAir’ Application Programming In-

terface

Akitaka Matsuo Cole Brokamp

mgpStreamingSDK Interact with the Maxar MGP Streaming API Nathan Carr

GetLattesData Reading Bibliometric Data from Lattes Platform Marcelo Perlin

worldbank Client for World Banks’s ’Indicators’ and ’Poverty

andInequality Platform (PIP)’ APIs

Maximilian Mücke

BFS Get Data from the Swiss Federal Statistical Oﬀice Felix Luginbuhl

trud Query the ’NHS TRUD API’ Alasdair Warwick rinat Access ’iNaturalist’ Data Through APIs Stéphane Guillou

jsonlite A Simple and Robust JSON Parser and Generator for Jeroen Ooms

R

zen4R Interface to ’Zenodo’ REST API Emmanuel Blondel

yfinancer ’Yahoo Finance’ API Wrapper Giovanni Colitti

opendotaR Interface for OpenDota API Kari Gunnarsson Visualize.CRAN.DownloVadisualize Downloads from ’CRAN’ Packages Marcelo Ponce

PurpleAirAPI Historical Data Retrieval from ’PurpleAir’ Sensors via Heba Abdelrazzak

API

pacu Precision Agriculture Computational Utilities dos Santos Caio cbsodataR Statistics Netherlands (CBS) Open Data API Client Edwin de Jonge kosis Korean Statistical Information Service (KOSIS) Seokhoon Joo

MetaculR Analyze Metaculus Predictions and Questions Joseph de la Torre Dwyer trelloR Access the Trello API Jakub Chromec

rscorecard A Method to Download Department of Education Col- Benjamin Skinner

lege ScorecardData

inegiR Integrate INEGI’s (Mexican Stats Oﬀice) API with R Eduardo Flores naptanr Call the ’NaPTAN’ API Through R Francesca Bryden

Table A.3: Example packages for cleaning data with API

package title maintainer

discretization Data Preprocessing, Discretization for Classification HyunJi Kim

helda Preprocess Data and Get Better Insights from Ma-

chine LearningModels

recipes Preprocessing and Feature Engineering Steps for Mod-

eling

Simon Corde Max Kuhn

dunlin Preprocessing Tools for Clinical Trial Data Joe Zhu

dataprep Eﬀicient and Flexible Data Preprocessing Tools Chun-Sheng Liang smallsets Visual Documentation for Data Preprocessing Lydia R. Lucchesi

Continued on next page

Table A.3: Example packages for cleaning data with API (Continued)

rtry Preprocessing Plant Trait Data Olee Hoi Ying Lam

PupilPre Preprocessing Pupil Size Data Aki-Juhani Kyröläinen

mpactr Correction of Preprocessed MS Data Patrick Schloss bdpar Big Data Preprocessing Architecture Miguel Ferreiro-Díaz webtrackR Preprocessing and Analyzing Web Tracking Data David Schoch tsrobprep Robust Preprocessing of Time Series Data Michał Narajewski VWPre Tools for Preprocessing Visual World Data Vincent Porretta

binst Data Preprocessing, Binning for Classification and Re-

gression

PreProcessing Various Preprocessing Transformations of Numeric

Data Matrices

esmtools Preprocessing Experience Sampling Method (ESM)

Data

RobLoxBioC Infinitesimally Robust Estimators for Preprocessing -

Omics Data

shinyrecipes Gadget to Use the Data Preprocessing ’recipes’ Pack-

ageInteractively

RGCxGC Preprocessing and Multivariate Analysis of Bidimen-

sional GasChromatography Data

Chapman Siu Swamiji Pravedson Jordan Revol Matthias Kohl

Alberto Almuiña

Cristian Quiroz-Moreno

mlr3pipelines Preprocessing Operators and Pipelines for ’mlr3’ Martin Binder

EEM Read and Preprocess Fluorescence Excitation-

Emission Matrix(EEM) Data

Vipavee Trivittayasil

cobalt Covariate Balance Tables and Plots Noah Greifer

clickR Semi-Automatic Preprocessing of Messy Data with

Change Trackingfor Dataset Cleaning

SerolyzeR Reading, Quality Control and Preprocessing of MBA

(MultiplexBead Assay) Data

PvSTATEM Reading, Quality Control and Preprocessing of MBA

(MultiplexBead Assay) Data

David Hervas Marin Tymoteusz Kwiecinski Tymoteusz Kwiecinski

huge High-Dimensional Undirected Graph Estimation Haoming Jiang

klaR Classification and Visualization Uwe Ligges datawizard Easy Data Wrangling and Statistical Transformations Etienne Bacher dplyr A Grammar of Data Manipulation Hadley Wickham pagoda2 Single Cell Analysis and Differential Expression Evan Biederstedt

ggplot2 Create Elegant Data Visualisations Using the Gram-

mar of Graphics

Thomas Lin Pedersen

biclust BiCluster Algorithms Sebastian Kaiser

tidyr Tidy Messy Data Hadley Wickham

tibble Simple Data Frames Kirill Müller

prospectr Miscellaneous Functions for Processing and Sample

Selection ofSpectroscopic Data

Leonardo Ramirez-Lopez

microeco Microbial Community Ecology Data Analysis Chi Liu

pammtools Piece-Wise Exponential Additive Mixed Modeling

Tools forSurvival Analysis

Andreas Bender

ebal Entropy Reweighting to Create Balanced Samples Jens Hainmueller

ordinalRR Analysis of Repeatability and Reproducibility Studies

withOrdinal Measurements

ff Memory-Eﬀicient Storage of Large Data on Disk and Fast AccessFunctions

Ken Ryan

Jens Oehlschlägel

Continued on next page

Table A.3: Example packages for cleaning data with API (Continued)

mlr3data Collection of Machine Learning Data Sets for ’mlr3’ Marc Becker lubridate Make Dealing with Dates a Little Easier Vitalie Spinu daltoolbox Leveraging Experiment Lines to Data Analytics Eduardo Ogasawara

mlrCPO Composable Preprocessing Operators and Pipelines

for MachineLearning

Martin Binder

readr Read Rectangular Text Data Jennifer Bryan

CRMetrics Cell Ranger Output Filtering and Metrics Visualiza- Rasmus Rydbirk

tion

JointAI Joint Analysis and Imputation of Incomplete Data Nicole S. Erler HiClimR Hierarchical Climate Regionalization Hamada S. Badr simulariatools Simularia Tools for the Analysis of Air Pollution Data Giuseppe Carlino gcxgclab GCxGC Preprocessing and Analysis Stephanie Gamble

Table A.4: Example packages for transforming data with API

package title maintainer

yaml Methods to Convert R Data to YAML and Back Shawn Garbett geojsonio Convert Data from and to ’GeoJSON’ or ’TopoJSON’ Michael Mahoney

jsonlite A Simple and Robust JSON Parser and Generator for Jeroen Ooms

R

reticulate Interface to ’Python’ Tomasz Kalinowski

keyToEnglish Convert Data to Memorable Phrases Max Candocia qtl2convert Convert Data among QTL Mapping Packages Karl W Broman gtools Various R Programming Tools Ben Bolker

rmarkdown Dynamic Documents for R Yihui Xie interleave Converts Tabular Data to Interleaved Vectors David Cooley do Data Operator Jing Zhang

rio A Swiss-Army Knife for Data I/O Chung-hong Chan data.tree General Purpose Hierarchical Data Structure Christoph Glur wktmo Converting Weekly Data to Monthly Data You Li

GDPuc Easily Convert GDP Data Johannes Koch

nuts Convert European Regional Data Moritz Hennicke wearables Tools to Read and Convert Wearables Data Peter de Looff

xml2relational Converting XML Documents into Relational Data Joachim Zuckarelli

Models

TidyMultiqc Converts ’MultiQC’ Reports into Tidy Data Frames Michael Milton

odk Convert ’ODK’ or ’XLSForm’ to ’SPSS’ Data Frame Muntashir-Al-Arefin spbabel Convert Spatial Data Using Tidy Tables Michael D. Sumner exp2flux Convert Gene EXPression Data to FBA FLUXes Daniel Osorio ecocomDP Tools to Create, Use, and Convert ecocomDP Data Colin Smith

tbl2xts Convert Tibbles or Data Frames to Xts Easily Nico Katzke broom.mixed Tidying Methods for Mixed Models Ben Bolker

LAIr Converting NDVI to LAI of Field, Proximal and Satel- Francesco Chianucci

lite Data

intergraph Coercion Routines for Network Data Objects Michał Bojanowski

ILRCM Convert Irregular Longitudinal Data to Regular Inter-

vals andPerform Clustering

Atanu Bhattacharjee

vcfR Manipulate and Visualize VCF Data Brian J. Knaus

gtfs2gps Converting Transport Data from GTFS Format to GPS-Like Records

Pedro R. Andrade

Continued on next page

Table A.4: Example packages for transforming data with API (Continued)

RJSONIO Serialize R Objects to JSON, JavaScript Object Nota-

tion

MissingHandle Handles Missing Dates and Data and Converts into

Weekly andMonthly from Daily

orsk Converting Odds Ratio to Relative Risk in Cohort

Studies withPartial Data Information

Yaoxiang Li

Mr. Sandip Garai Zhu Wang

sjlabelled Labelled Data Utility Functions Daniel Lüdecke

dplyr A Grammar of Data Manipulation Hadley Wickham pack Convert Values to/from Raw Vectors Joshua M. Ulrich tidytree A Tidy Tool for Phylogenetic Tree Data Manipulation Guangchuang Yu

ggplot2 Create Elegant Data Visualisations Using the Gram-

mar of Graphics

Thomas Lin Pedersen

string2path Rendering Font into ’data.frame’ Hiroaki Yutani tdata Prepare Your Time-Series Data for Further Analysis Ramin Mojab DDIwR DDI with R Adrian Dusa CADF Customer Analytics Data Formatting Ludwig Steven

tidyr Tidy Messy Data Hadley Wickham

tibble Simple Data Frames Kirill Müller redquack Transfer ’REDCap’ Data to Database Dylan Pieper

tinytable Simple and Configurable Tables in ’HTML’, ’LaTeX’,

’Markdown’,’Word’, ’PNG’, ’PDF’, and ’Typst’ For-

mats

mergen AI-Driven Code Generation, Explanation and Execu-

tion for DataAnalysis

Vincent Arel-Bundock

Altuna Akalin

unpivotr Unpivot Complex and Irregular Data Layouts Duncan Garmonsway jsonld JSON for Linking Data Jeroen Ooms

mltools Machine Learning Tools Ben Gorman

mergenstudio ’Mergen’ Studio: An ’RStudio’ Addin Wrapper for the

’Mergen’Package

Jacqueline Jansen