

APPENDIX B

R and Python Computing Setup Guide

Every chapter's Computational Corner gives parallel R and Python code. This appendix lists the packages used throughout the book and how to install them, together with brief notes on the handful of specialised genetics packages referenced in the text.

B.1 Setting Up R

R (version 4.0 or later is recommended) can be downloaded from <https://cran.r-project.org>; RStudio Desktop is the recommended editing environment. Packages are installed with `install.packages("package_name")`, run once per machine.

Core packages used in this book:

Package	Used in	Purpose
stats (base R)	All chapters	Chi-square tests, glm, lm, optim, distributions
HardyWeinberg	Chapter 3	Hardy–Weinberg equilibrium testing and visualisation
kinship2	Chapter 6	Pedigree drawing and kinship-matrix computation
genetics	Chapters 15, 19	Genotype objects, linkage disequilibrium statistics

haplo.stats	Chapter 15	EM-based haplotype frequency estimation
qqman	Chapter 18	Manhattan and Q-Q plots for GWAS summary statistics
coala / scrm	Chapter 16, 20	Coalescent simulation (CRAN installation required; not pre-installed in all environments)
lme4 / nlme	Chapters 9, 14	Mixed-effects models for family and structured data

Install the core set with:

```
install.packages(c("HardyWeinberg", "kinship2", "genetics",
"haplo.stats", "qqman", "lme4"))
```

B.2 Setting Up Python

Python (version 3.9 or later) is available from <https://www.python.org>, or via the Anaconda distribution, which bundles most scientific packages by default. Packages are installed with *pip install package_name* (add *—break-system-packages* on some Linux systems with an externally managed environment).

Core packages used in this book:

Package	Used in	Purpose
numpy	All chapters	Array computation
scipy	All chapters	Statistical tests, distributions, optimisation
pandas	Most chapters	Tabular data handling
statsmodels	Chapters 9, 11, 12	Regression models (OLS, logistic)
scikit-learn	Chapter 14	Principal component analysis

matplotlib	Chapters 4, 18,	Plotting
b	20	
network	Chapter 6	Pedigree graph representation
x		
msprime	Chapter 20	Coalescent simulation of sequence data
scikit- allel	Chapter (further reading)	20 Population genetic analysis of variant call data

Install the core set with:

```
pip install numpy scipy pandas statsmodels scikit-learn  
matplotlib networkx msprime scikit-allel --break-system-  
packages
```

B.3 A Note on Reproducibility

Every simulation-based Computational Corner and Master Answer Key script in this book sets an explicit random seed (*set.seed()* in R, a *numpy.random.default_rng(seed)* generator, or *random.seed()* in Python) so that results are exactly reproducible on any machine running the same package versions. Readers repeating these exercises with a different seed should expect numerically different, but qualitatively similar, results — an instructive illustration, in itself, of Monte Carlo sampling variability (Chapter 20, Section 20.5).

B.4 Running the Book's Code

No special software beyond a working R or Python installation and the packages listed above is required to run any script in this book. Each Computational Corner's R and Python blocks are self-contained and can be copied directly into an R console/RStudio script, or a Python interpreter/Jupyter notebook, respectively, and run independently of any other chapter's code.