upset.hp: An R Package for Visualizing Hierarchical Partitioning and Variance Partitioning Results Using UpSet Plots

Bangken Ying[1,2,3], Yao Liu[4**], Jiangshan Lai[1,2,5*]**

[1]Co-Innovation Center for Sustainable Forestry in Southern China, College of Ecology and Environment, Nanjing Forestry University, Nanjing 210037, China

[2]Research Center of Quantitative Ecology, Nanjing Forestry University, Nanjing 210037, China

[3]Higher School of Medical and Biology, South Ural State University (National Research University), Chelyabinsk 454080, Russia

[4]School of Environment and Ecology, Xiamen University, Xiamen 361102, China

[5]University of Chinese Academy of Sciences, Beijing 100049, China

Keywords: Average shared variance; R package, glmm.hp; rdacca.hp; gam.hp; phylolm.hp

Citation: Ying B.K., Liu Y., Lai J.S.: upset.hp: An R Package for Visualizing Hierarchical Partitioning and Variance Partitioning Results Using UpSet Plots. Data Intelligence, 2026, Vol. 8, Art. No.: 2026d01. DOI: https://doi.org/10.3724/2096-7004.di.2025.0200

ABSTRACT

Traditional Venn diagrams become increasingly complex and difficult to interpret when visualizing the intersections of more than four variables or variable groups. The UpSet plot, a matrix-based, data-driven visualization technique, offers a scalable and more readable alternative for representing complex set intersections. The upset.hp R package, newly released in 2025, extends the capabilities of the previously developed UpSetVP package by providing a unified visualization tool compatible with the full suite of hp-series packages developed for hierarchical and variation partitioning, including rdacca.hp, glmm.hp, gam.hp, and phylolm.hp.

1. INTRODUCTION

Understanding the intricate relationships among multiple sets is a paramount task in data analysis, with widespread applications spanning numerous fields, from bioinformatics to social sciences. However, this endeavor is accompanied by significant challenges, as the analysis and visualization of sets become exceptionally demanding when the number of sets exceeds a mere handful. While traditional set

*Corresponding author: Jiangshan Lai (E-mail: lai@njfu.edu.cn).

**Bangken Ying and Yao Liu contributed equally to this work.

© 2026 Chinese Academy of Sciences. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

visualization methods, such as Venn diagrams and Euler diagrams, offer intuitive and easily interpretable representations for a limited number of sets (typically three to four), they depict sets and their intersections using overlapping geometric shapes, often relying on area proportionality to indicate relative sizes [1]. Nevertheless, as the number of sets increases, the potential number of set intersections proliferates combinatorially, leading to diagrams that are not only exceedingly complex and visually cluttered but also profoundly challenging to interpret accurately. Although the conceptual meaning of multiple set intersections remains intuitive, their visual depiction, especially when involving more than three or four overlapping sets and their intricate interactions, is far from trivial. Precisely identifying the specific sets involved in a particular intersection becomes exceptionally arduous. Furthermore, given that area encoding is inherently less precise than positional encoding for conveying quantitative information, accurately assessing intersection sizes—such as identifying the largest or smallest overlaps—is severely compromised.

It is precisely within this context that UpSet emerged as an innovative visualization technique, specifically engineered to overcome these inherent limitations and provide a highly scalable solution for complex set analysis [2]. At its core, UpSet employs a divide-and-conquer strategy: it meticulously decomposes the dataset into all exclusive intersections—defined as the fundamental atomic units where elements belong precisely to a specific combination of sets and only those sets. This foundational decomposition enables a precise and highly scalable representation of complex set relationships. Subsequently, UpSet introduces task-driven aggregation, which allows for the meaningful recombination of these exclusive intersections according to an analyst's specific objectives. This facilitates comprehensive quantitative analysis of intersection sizes, element counts, and their associated attributes. By combining consistent visual encodings (primarily leveraging position, widely acknowledged as the most effective visual variable for quantitative representation), intuitive aggregation and sorting mechanisms, and straightforward querying and interaction techniques, UpSet delivers an efficient, comprehensible, and user-centric solution. This approach proficiently scales to manage a substantial number of sets (e.g., 20-30 or more, depending on dataset properties), thereby supporting a broad spectrum of advanced set-related analytical tasks. A distinctive feature of UpSet is its comprehensive exploitation of the inherent duality between attribute visualization and set membership. This empowers flexible querying and filtering across both the set space (i.e., by selecting elements through their set associations) and the element space (i.e., based on their attributes). By integrating attribute visualization either directly within the set space or, for more detailed exploration, within the element space, UpSet significantly simplifies the comparison of different data partitions, ultimately furnishing analysts with unprecedented depth of insight.

Quantifying the relative contributions of multiple explanatory variables in statistical models is fundamental across diverse scientific disciplines, including ecology, environmental science, and social sciences. Such analyses are crucial for illuminating the intricate relationships among variables and achieving a profound comprehension of underlying phenomena. Traditional methods, such as Variation Partitioning (VP) and Hierarchical Partitioning (HP), serve to decompose the total explained variation into unique and shared components, thereby enabling researchers to disentangle the individual impacts of predictors and their interactive effects. However, as the number of explanatory variables increases, the visualization of shared contributions becomes increasingly complex and challenging.

A common challenge in statistical modeling, multicollinearity, exacerbates the interpretational difficulty of variable contributions. This phenomenon, where two or more predictors are highly correlated, can lead to unstable regression coefficient estimates and complicate the accurate attribution of variable importance. Extensive research has been dedicated to addressing multicollinearity [2-4], culminating in pivotal advancements such as the 'averaging over orderings' approach. Independently proposed by Lindeman et al., Cox, and Kruskal [3, 5-7], this method robustly quantifies relative importance by systematically allocating shared R^2 values to predictors across all possible model subsets. This foundational approach underpins sophisticated methods like Hierarchical Partitioning and Dominance Analysis , facilitating more nuanced insights into variable contributions [8-9].

Building upon these methodological advancements, a suite of R packages, collectively known as the hp-series, has been developed to extend these robust partitioning capabilities across diverse statistical models. As the predominant statistical analysis tool in ecology, R provides the foundational platform that supports these specialized packages [10]. The rdacca.hp package implements the Average Shared Variance (ASV) approach for canonical analyses, fairly partitioning overlapping explained variance among correlated predictors [11]. This methodology has since been extended to Generalized Linear Mixed Models (GLMMs) by glmm.hp [12-13], Generalized Additive Models (GAMs) by gam.hp [14], and most recently to Phylogenetic Generalized Linear Models (PGLMs) by phylolm.hp [15]. These packages provide comprehensive analytical outputs detailing unique and shared contributions across various model types, offering powerful tools for quantitative analyses.

While the hp-series packages provide powerful analytical capabilities, the challenge of effectively visualizing the complex structure of shared contributions remains formidable, especially as the number of variables increases. Traditional visual tools, such as Venn diagrams, often lose clarity and comprehensibility when depicting intersections among multiple sets, rapidly becoming unwieldy with larger datasets. In response to this limitation, UpSet plots have gained prominence as a superior alternative [16]. By employing a combination of matrix layouts and bar charts, UpSet plots offer a scalable and structured approach to visualize intricate intersection relationships, enabling researchers to discern patterns of shared contributions more effectively and gain clearer representations of how explanatory variables interact.

To address this critical visualization gap, this paper introduces the upset.hp R package. This package is engineered to seamlessly integrate the advanced visualization capabilities of UpSet plots with the comprehensive outputs of the hp-series packages (rdacca.hp, glmm.hp, gam.hp, and phylolm.hp). By providing a unified and intuitive solution, upset.hp's primary objective is to profoundly enhance the clarity and interpretability of hierarchical and variation partitioning results, thereby elucidating the intricate roles of multiple explanatory variables in complex statistical models. This streamlined tool, incorporating state-of-the-art visualization techniques, will significantly advance the understanding of variable relationships, empowering researchers to more effectively navigate and interpret the inherent complexities of multifactorial analyses.

In conclusion, the development of upset.hp represents a crucial advancement in bridging the gap between sophisticated statistical modeling and intuitive data visualization. This package aims to empower

researchers by providing a robust and accessible tool to effectively interpret the nuanced interplay of variables in complex scientific inquiries.

2. PACKAGE FEATURES

Key features of this package include full compatibility, which allows it to automatically recognize and process results from all supported hp-series packages; flexible input, accepting standard HP/VP output data frames without the need for additional formatting; enhanced interpretability, improving the clarity of shared versus independent variable contributions in highdimensional models; and consistent aesthetics, offering standardized color palettes(Color palette name: "nature" (default), "science", "cell", "bw" (black white), or "cvd" (color-blind friendly))., labels, and layouts to facilitate comparative analysis. These functionalities make upset.hp a valuable tool for researchers dealing with complex statistical models.

3. PACKAGE DESCRIPTION

The upset.hp package is written in the language R (R Core Team 2024) and can be installed from the Comprehensive R Archive Network (http://cran.r-project.org/web/packages/upset.hp/) or GitHub (https://github.com/laijiangshan/upset.hp). The package features one primary function, upset.hp(), which generates UpSet plots to visualize the unique, common, and individual contributions of predictors towards explained variation in various models through variation partitioning (VP) and hierarchical partitioning (HP). The upset.hp() function takes two main arguments: vp and hp, which should contain the outputs of variation partitioning and hierarchical partitioning derived from packages such as rdacca.hp, glmm.hp, gam.hp, and phylolm.hp. The function also provides various graphical parameters to customize the display, such as ordering options for predictors and the ability to show or hide effect sizes above bars.

The upset.hp() function depends on several packages, including MuMIn, vegan, ggplot2, and patchwork, to facilitate its computations and visualizations. The function returns a ggplot2 object, allowing for further customization and integration into larger plotting workflows.

The upset.hp function is designed to accommodate an unlimited number of predictor variables or groups of predictors, though it is recommended to limit the number of predictors to a maximum of seven for interpretability.

Handling missing values is automatically managed within upset.hp, with rows containing missing data excluded from the analysis. The upset.hp package provides a powerful tool for researchers looking to visualize complex interactions among predictors in their statistical models.

4. WORKING EXAMPLES

We will present four illustrative case studies demonstrating the functionality of the upset.hp function in conjunction with various hp series packages, including rdacca.hp, glmm.hp, gam.hp, and phylolm.

hp. These examples cover a range of statistical analysis scenarios, from constrained ordination with multivariate responses to generalized linear mixed-effects models, generalized additive models, and phylogenetic regression models. The aim is to comprehensively showcase the broad applicability and robust functionality of upset.hp in visualizing complex hierarchical and variance partitioning results.

4.1 Application with rdacca.hp: Constrained Ordination Analysis

The rdacca.hp package is designed for hierarchical partitioning in constrained ordination analyses, such as Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA) [17]. For this case study, we will utilize built-in datasets from the vegan package [18]. The following example demonstrates how to use upset.hp to visualize the results from rdacca.hp.

```
# Load necessary libraries
library (upset.hp)
library (vegan)
library (rdacca.hp)

# Load and prepare data
data (mite)          # Mite species abundance data
data (mite.env)      # Environmental variables data
data (mite.xy)       # Spatial coordinates (not used in this example)
data (mite.pcnm)     # Spatial variables (not used in this example)

# Perform Hellinger transformation on species abundance data
mite.hel <- decostand (mite, "hellinger")

# Execute hierarchical partitioning for Redundancy Analysis (RDA)
# method = "RDA" specifies RDA
# type = "adjR2" specifies adjusted R-squared as the metric
# var.part = TRUE indicates to perform variance partitioning (commonality analysis)
m1 <- rdacca.hp (mite.hel, mite.env, method = "RDA", type = "adjR2", var.part = TRUE)

# Visualize variance partitioning and hierarchical partitioning results using upset.hp
# m1$ Var.part contains the variance partitioning (commonality analysis) results
# m1$ Hier.part contains the hierarchical partitioning (individual R-squared contributions) results
upset.hp (m1$ Var.part, m1$ Hier.part, col = "nature", digits = 2)
```

In this case, upset.hp effectively visualizes the unique contributions of individual environmental variables (from mite.env) and their shared explanatory power in accounting for the variation in mite community structure (represented by Hellinger-transformed species abundances) (Figure 1). This is crucial for understanding the mechanisms by which environmental factors influence community composition in complex ecological systems.

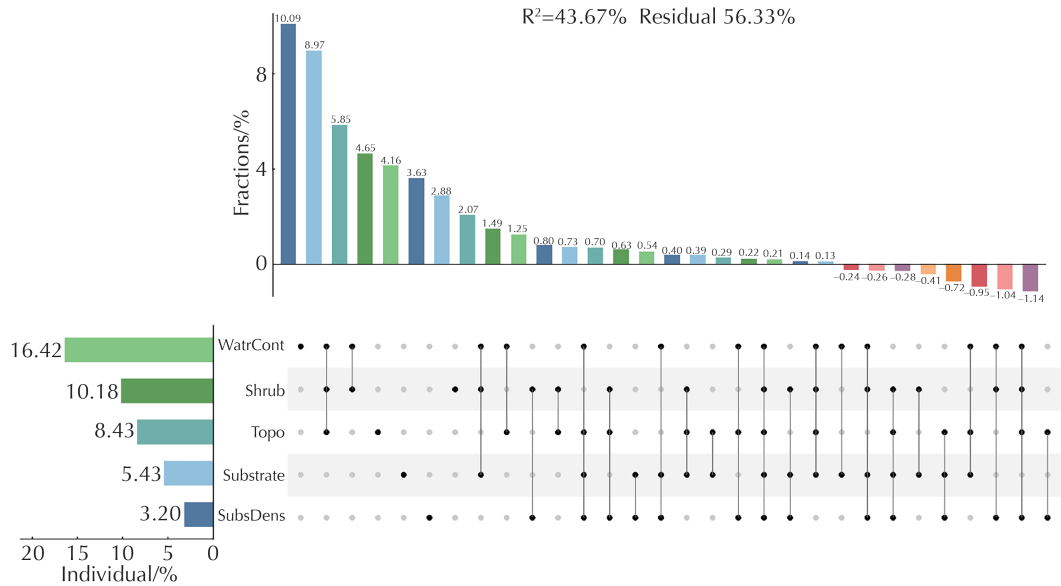


Figure 1. upset.hp application with rdacca.hp().

4.2 Application with glmm.hp: Linear, Generalized Linear, and Linear Mixed-Effects Models

The glmm.hp package primarily focuses on hierarchical partitioning for Generalized Linear Mixed-Effects Models (GLMMs) but also supports standard Linear Models (lm) and Generalized Linear Models (glm) [11-12]. This section illustrates its capabilities through three sub-examples.

4.2.1 Linear Model (lm)

```
# Load glmm.hp library
library (glmm.hp)
library (upset.hp)

# upset.hp application with lm ()
# Fit a linear model using the mtcars dataset
m2 <- lm (mpg ~ wt + carb + cyl, mtcars)

# Calculate commonality analysis results
vp <- glmm.hp (m2, commonality = TRUE)$ commonality. analysis
# Calculate hierarchical partitioning results
hp <- glmm.hp (m2)$ hierarchical.partitioning

# Visualize the results using upset.hp
upset.hp (vp, hp, col = "nature", digits = 2)
```

Here, upset.hp demonstrates the unique and shared explanatory power of car weight (wt), number of carburetors (carb), and number of cylinders (cyl) on the variation in miles per gallon (mpg) (Figure 2).

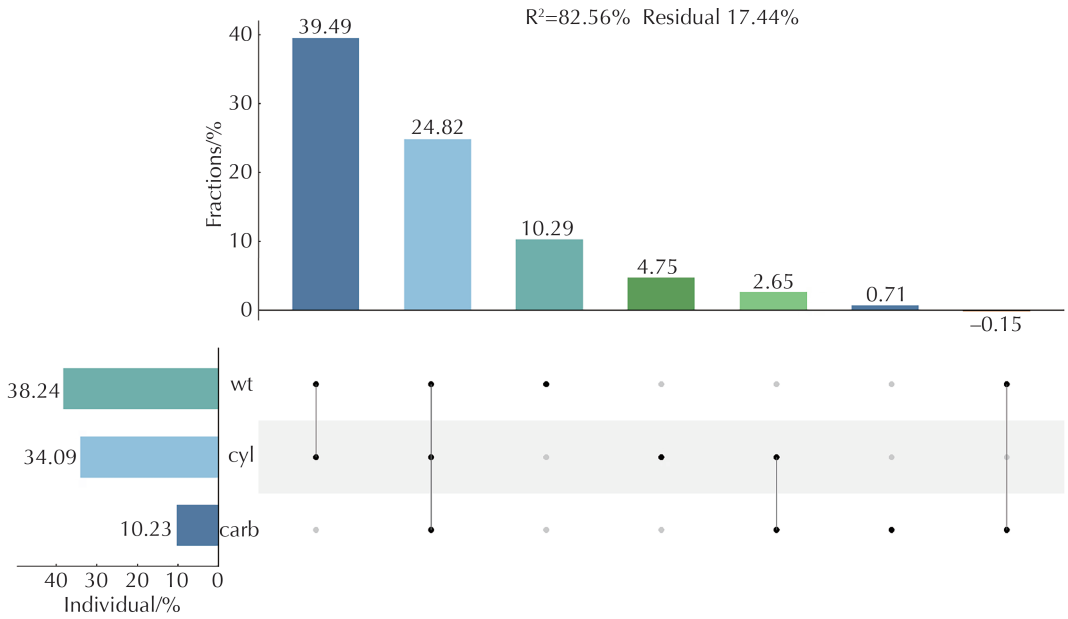


Figure 2. upset.hp application with lm().

4.2.2 Generalized Linear Model (glm)

```
# Load glmm.hp library
library (glmm.hp)
library (upset.hp)

# upset.hp application with glm ()
# Fit a generalized linear model (binomial logistic regression) using mtcars
# Predicting transmission type (am)
m3 <- glm (am~cyl + disp + wt, family = binomial, data = mtcars)

# Calculate theoretical commonality analysis results (recommended for GLMs)
vp <- glmm.hp (m3, commonality = TRUE)$ theoretical
# Calculate theoretical hierarchical partitioning results
hp <- glmm.hp (m3)$ theoretical

# Visualize the results using upset.hp
upset.hp(vp, hp, col = "nature", digits = 2)
```

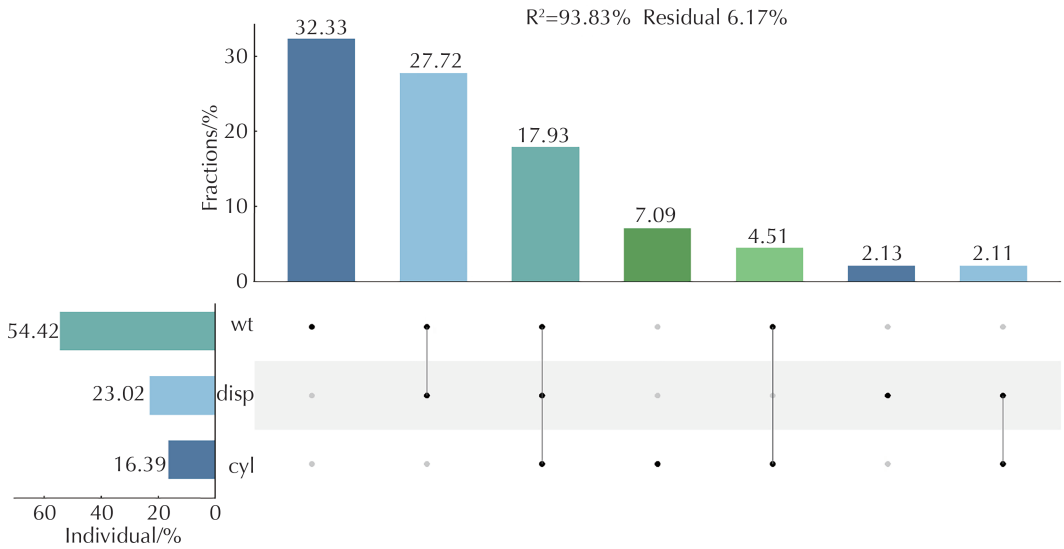


Figure 3. upset.hp application with glm().

This example illustrates how the number of cylinders (cyl), displacement (disp), and weight (wt) of a car uniquely or jointly influence the probability of its transmission type (am) (Figure 3).

4.2.3 Linear Mixed-Effects Model (lmer)

```
# Load necessary libraries
library (lme4)
library (partR2) # Contains the biomass dataset
library (upset.hp)

# upset.hp application with lmer ()
# Load the biomass dataset
data (biomass)
# Fit a linear mixed-effects model, with Population as a random effect
m4 <- lmer (Biomass ~ Year + Temperature + Precipitation + SpeciesDiversity + (1 | Population), data
      = biomass)

# Calculate commonality analysis results
vp <- glmm.hp (m4, commonality = TRUE)$ commonality. analysis
# Calculate hierarchical partitioning results
hp <- glmm.hp (m4)$ hierarchical. partitioning

# Visualize the results using upset.hp
upset.hp (vp, hp, col = "nature", digits = 2)
```

This comprehensive example demonstrates how biomass is influenced by Year, Temperature, Precipitation, and SpeciesDiversity, while accounting for random variation across different Population units. upset.hp clearly presents the unique and shared contributions of these fixed effect predictors to the total biomass variation(Figure 4).

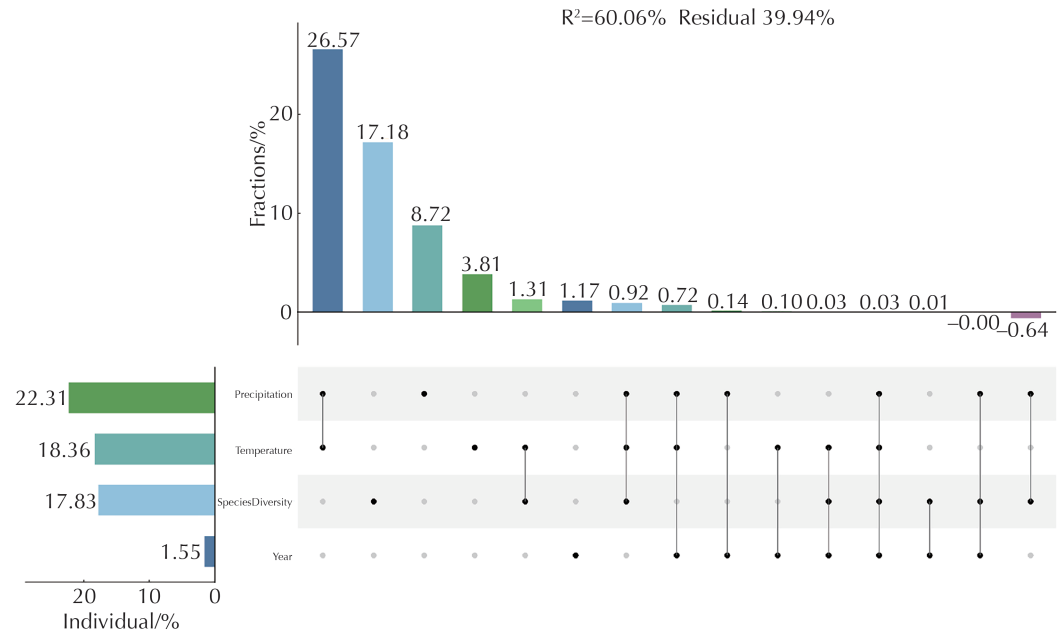


Figure 4. upset.hp application with lmer()

4.3 Application with gam.hp: Generalized Additive Models

The gam.hp package is used for hierarchical partitioning in Generalized Additive Models (GAMs), which allow for non-linear relationships between predictor and response variables [13].

```
# Load gam.hp library
library (gam.hp)
library (upset.hp)

# upset.hp application with gam.hp ()
# Fit a GAM using the iris dataset
# s() function indicates a smooth term, allowing for non-linear relationships
m5 <- gam (Sepal.Length~s (Petal.Length) + s (Petal.Width) + Sepal.Width, data = iris)

# Calculate commonality analysis results
vp <- gam.hp (m5, commonality = TRUE)$ commonality.analysis
```

```
# Calculate hierarchical partitioning results
hp <- gam.hp (m5)$ hierarchical. partitioning

# Visualize the results using upset.hp
upset.hp (vp, hp, col = "nature", digits = 2)
```

In this scenario, upset.hp visualizes how the variation in iris sepal length (Sepal.Length) is uniquely and jointly explained by petal length (Petal.Length), petal width (Petal.Width) (potentially via non-linear functions), and sepal width (Sepal.Width) (Figure 5).

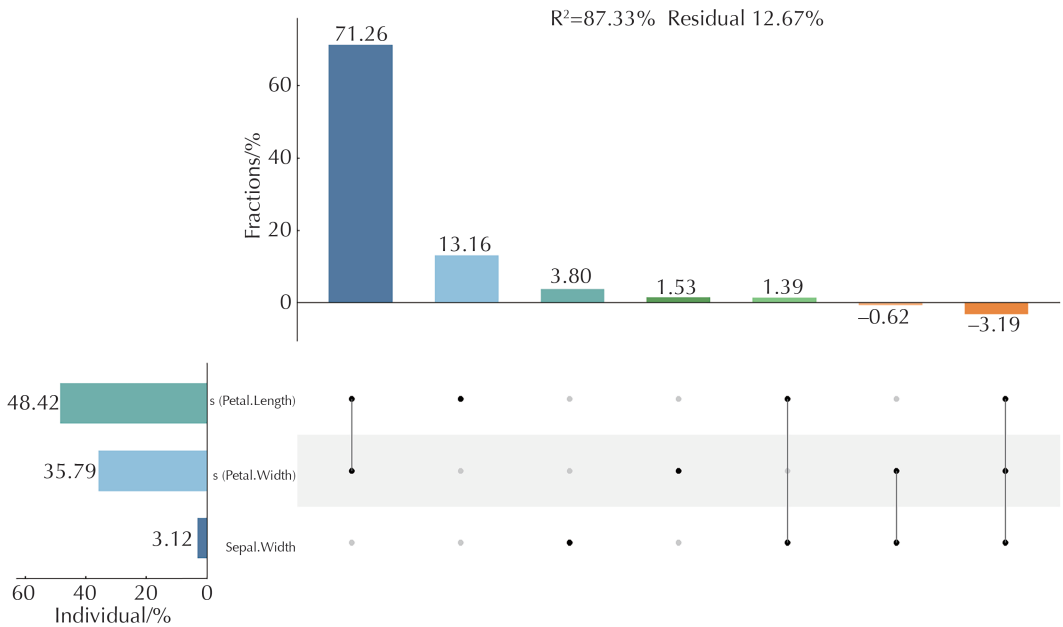


Figure 5. upset.hp application with gam.hp().

4.4 Application with phylolm.hp: Phylogenetic Regression Models

The phylolm.hp package extends hierarchical partitioning to regression models that account for phylogenetic relationships among species. This is crucial for analyzing biological data where species are not independent observations due to shared evolutionary history [14].

```
# Load phylolm.hp library
library (phylolm.hp)
library (upset.hp)

# upset.hp application with phylolm.hp ()
set.seed (231) # Set seed for reproducibility

# Generate a random phylogenetic tree
tre <- rcoal (60)
taxa <- sort (tre $ tip.label)

# Define regression coefficients
b0 <- 0
b1 <- 0.3
b2 <- 0.5
b3 <- 0.4

# Simulate predictor variables with phylogenetic signal (using a lambda model)
x <- rTrait (n = 1, phy = tre, model = "lambda", parameters = list (ancestral.state = 0, sigma2 = 15,
      lambda = 0.9) )
x2 <- rTrait (n = 1, phy = tre, model = "lambda", arameters = list (ancestral.state = 0, sigma2 = 10,
      lambda = 0.9) )
x3 <- rTrait (n = 1, phy = tre, model = "lambda", parameters = list (ancestral.state = 0, sigma2 = 13,
      lambda = 0.9) )

# Simulate a response variable with phylogenetic signal
y <- b0 + b1 * x + b2 * x2 + b3 * x3 + rTrait (n = 1, phy = tre, model = "lambda", parameters = list
      (ancestral. state = 0, sigma2 = 5, lambda = 0.9) )

# Create a data frame
dat <- data.frame (trait = y[taxa], pred = x[taxa], pred2 = x2[taxa], pred3 = x3[taxa])

# Fit a phylogenetic linear model (phylolm), incorporating the phylogenetic tree
m6 <- phylolm (trait ~ pred + pred2 + pred3, data = dat, phy = tre , model = "lambda")

# Calculate commonality analysis results
vp <- phylolm.hp (m6, commonality = TRUE)$ commonality.analysis
# Calculate hierarchical partitioning results (Individual.R2 for phylolm.hp)
hp <- phylolm.hp (m6)$ Individual.R2

# Visualize the results using upset.hp
upset.hp (vp, hp, col = "nature", digits = 2)
```

This advanced case demonstrates how upset.hp can be used to disentangle the unique and shared contributions of three simulated predictor variables (pred, pred2, pred3) to a simulated trait (trait), while explicitly accounting for the underlying phylogenetic relationships. This provides an intuitive visualization to interpret complex phylogenetic regression outcomes (Figure 6).

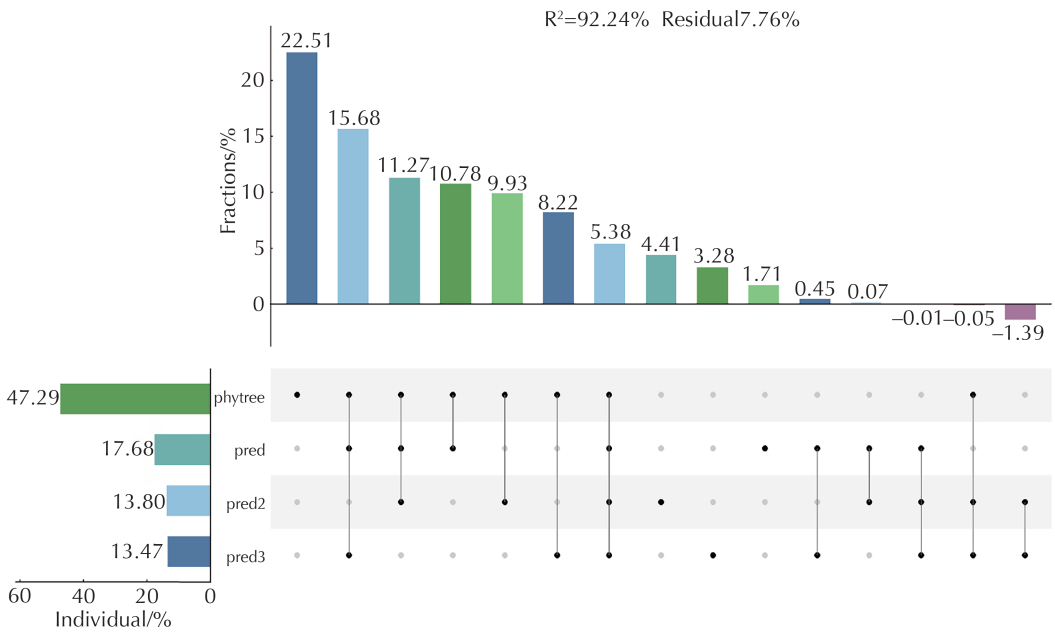


Figure 6. upset.hp application with phylolm.hp().

5. CONCLUSION

upset.hp provides a powerful and user-friendly tool for visualizing complex variable contributions in statistical models using UpSet plots. We recommend upset.hp as a general-purpose visualization package for anyone conducting hierarchical or variation partitioning in R.

DATA AVAILABILITY

The data sets generated during and/or analyzed during the current study are available in the ScienceDB repository, https://doi.org/10.57760/sciencedb.j00104.00149.

AUTHOR CONTRIBUTIONS

Bangken Ying: Writing original draft, Formal analysis. Yao Liu: Writing original draft, Formal analysis. Jiangshan Lai: Conceptualization, Writing review & editing.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 32571954, 32271551).

REFERENCES

[1] Bennett, D.: Origins of the Venn diagram. In: Research in History and Philosophy of Mathematics: The CSHPM 2014 Annual Meeting in St. Catharines, Ontario, pp. 105–119 (2015).
[2] Nathans, L. L., Oswald, F. L. and Nimon, K.: Interpreting multiple linear regression: a guidebook of variable importance. In: Practical Assessment, Research & Evaluation, Vol. 17(9), n9 (2012).
[3] Bi, J.: A review of statistical methods for determination of relative importance of correlated predictors and identifi cation of drivers of consumer liking. In: Journal of Sensory Studies, Vol. 27(2), 87–101 (2012).
[4] Grömping, U.: Variable importance in regression models. In: Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 7(2), 137–152 (2015).
[5] Lindeman, R. H., Merenda, P. F. and Gold, R. Z.: Introduction to Bivariate and Multivariate Analysis. CRC Press, Boca Raton (1980).
[6] Cox Jr, L. A.: A new measure of attributable risk for public health applications. In: Management Science, Vol. 31(7), 800–813 (1985).
[7] Kruskal, W.: Relative importance by averaging over orderings. In: The American Statistician, Vol. 41(1), 6–10 (1987).
[8] Chevan, A. and Sutherland, M.: Hierarchical partitioning. In: The American Statistician, Vol. 45(2), 90–96 (1991).
[9] Budescu, D. V.: Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. In: Psychological Bulletin, Vol. 114(3), 542 (1993).
[10] Gao, M., Ye, Y., Zheng, Y., & Lai, J.: A comprehensive analysis of R's application in ecological research from 2008 to 2023: Journal of Plant Ecology, 18(1). (2025).
[11] Lai, J., Zou, Y., Zhang, J. and Peres-Neto, P. R.: Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. In: Methods in Ecology and Evolution, pp. 782–788 (2022).
[12] Lai, J., Zou, Y., Zhang, S., Zhang, X. and Mao, L.: glmm. hp: an R package for computing individual effect of predictors in generalized linear mixed models. In: Journal of Plant Ecology, Vol. 15(6), 1302–1307 (2022).
[13] Lai, J., Zhu, W., Cui, D. and Mao, L.: Extension of the glmm. hp package to zero-inflated generalized linear mixed models and multiple regression. In: Journal of Plant Ecology, Vol. 16(6), rtad038 (2023).
[14] Lai, J., Tang, J., Li, T., Zhang, A. and Mao, L.: Evaluating the relative importance of predictors in Generalized Additive Models using the gam.hp R package. In: Plant Diversity, Vol. 46(4), 542–546 (2024).
[15] Lai, J., He, Y., Hou, M., Zhang, A., Wang, G. and Mao, L.: Evaluating the relative importance of phylogeny and predictors in Phylogenetic Generalized Linear Models using the phylolm.hp R package. In: Plant Diversity (2025).
[16] Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H.: UpSet: visualization of intersecting sets. In: IEEE Transactions on Visualization and Computer Graphics, Vol. 20(12), 1983–1992 (2014).
[17] Liu, Y., Yu, X., Yu, Y., Hu, W. H. and Lai, J. S.: Application of "rdacca.hp" R package in ecological data analysis: case and progress. In: Chinese Journal of Plant Ecology, Vol. 47(1), 134 (2023).
[18] Dixon, P.: VEGAN, a package of R functions for community ecology. In: Journal of Vegetation Science, Vol. 14(6), 927–930 (2003).