Package ‘xadmix’

July 8, 2022

Type Package
Title Subsetting and Plotting Optimized for Admixture Data
Version 1.0.0
Description A few functions which provide a quick way of subsetting genomic admixture data and generating customizable stacked barplots.
License GPL (>= 3)
URL https://github.com/SpaceCowboy-71/xadmix
BugReports https://github.com/SpaceCowboy-71/xadmix/issues
Depends R (>= 2.10)
Imports dplyr, forcats, ggplot2, magrittr, methods, rlang, stringr, tidyr, viridis
Suggests knitr, rmarkdown
VignetteBuilder knitr
Encoding UTF-8
Language en-US
LazyData true
RoxygenNote 7.2.0
NeedsCompilation no
Author Lukas Schönmann [aut, cre] (<https://orcid.org/0000-0002-3161-4270>)
Maintainer Lukas Schönmann <lukas.schoenmann@outlook.com>
Repository CRAN
Date/Publication 2022-07-08 15:20:02 UTC

R topics documented:

admix_barplot .......................................................... 2
admix_subset ............................................................ 4
xadmixture ................................................................. 5

Index 7
Description

Stacked barplot optimized for admixture data.

Usage

\[
\text{admix_barplot}( \\
\text{data,} \\
\text{K = 2:ncol(data),} \\
\text{individuals = 1,} \\
\text{sortkey = NULL,} \\
\text{grouping = NULL,} \\
\text{palette = "default",} \\
\text{names = TRUE,} \\
\text{xlab = "Individuals",} \\
\text{ylab = "Ancestry",} \\
\text{main = "Admixture Plot",} \\
\text{noclip = FALSE} \\
) 
\]

Arguments

- **data**: Data frame containing the admixture data.
- **K**: Positions of the columns containing the ancestry percentages in the provided data frame; default is second to last column.
- **individuals**: Position of the column with the names for the x-axis; default is the first column.
- **sortkey**: Name of the column containing ancestry percentages to sort the stacked barplot with.
- **grouping**: Name of the column by which the stacked bars are to be grouped.
- **palette**: Either a color palette object, or a string to use one of the predefined color palettes ("viridis", "turbo", "alternating"); default is a modified ggplot palette.
- **names**: Whether to show the x-axis bar labels or not; default is "TRUE".
- **xlab**: A label for the x-axis.
- **ylab**: A label for the y-axis.
- **main**: A main title for the plot.
- **noclip**: Directly draw the plot, with clipping removed from elements. Then function does not return an object; default is set to "FALSE". Setting to "TRUE" may require launching a new R graphics device.

Value

A ggplot object of the stacked barplot.
Examples

# load simulated admixture data
data("xadmixture")

# for data frame with ancestries (K) in fourth to last column,
# without showing bar labels
admix_barplot(xadmixture, K = 4:ncol(xadmixture), names = FALSE)

# grouping data by column "country",
# and sorting each group by ancestry column "K1"
admix_barplot(xadmixture, K = 4:ncol(xadmixture), grouping = "country",
sortkey = "K1", names = FALSE)

# changing color palette to "turbo" from package 'viridis,'
admix_barplot(xadmixture, K = 4:ncol(xadmixture), palette = "turbo",
names = FALSE)

# removing title and changing axis labels text
admix_barplot(xadmixture, K = 4:ncol(xadmixture), main = "",
 xlab = "Accessions", ylab = "Ancestry [%]", names = FALSE)

# directly output grouped plot with clipping removed from elements
# (useful if there are groups with a low number of observations)
# create a subset of the data
xadmixture_sub <- admix_subset(xadmixture, anc = c("K3", "K4"),
pct = c(0.3, 0.2))

# generate a grouped & sorted stacked barplot
# setting "noclip" to "TRUE" may require opening a new graphics device
dev.new()
admix_barplot(xadmixture_sub, K = 4:ncol(xadmixture), sortkey = "K5",
grouping = "country", palette = "viridis", names = FALSE,
noclip = TRUE)
### admix_subset

Admixture Data Subsetting

**Description**

Subset function optimized for admixture data. Filters for the percentages of any number of ancestry (K) columns and prints progress. Also allows passing additional arguments to filter columns with.

**Usage**

```r
admix_subset(
  data,
  anc = NULL,
  pct = NULL,
  comparison = "greater",
  quiet = FALSE,
  ...
)
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>Data frame containing the admixture data.</td>
</tr>
<tr>
<td>anc</td>
<td>Vector of ancestry column names to use for pairwise subsetting with percentage vector. Must be of same length as the supplied percentage vector.</td>
</tr>
<tr>
<td>pct</td>
<td>Vector of percentage values to use for pairwise subsetting with ancestry column name vector. Only ancestries with values above the percentage are kept.</td>
</tr>
<tr>
<td>comparison</td>
<td>What comparison operator to use for the subsetting. Can either be &quot;greater&quot; or &quot;less&quot;; default is &quot;greater&quot;. Also accepts &quot;gt&quot;, &quot;lt&quot;, &quot;&gt;&quot; and &quot;&lt;&quot;.</td>
</tr>
<tr>
<td>quiet</td>
<td>Whether to print progress or not; default is &quot;FALSE&quot;.</td>
</tr>
<tr>
<td>...</td>
<td>Variable number of additional vectors for subsetting. Looking at the column with argument name, keeps only those observations with values which are elements of the argument vector.</td>
</tr>
</tbody>
</table>

**Value**

A subset of the provided data frame.

**Examples**

```r
# load simulated admixture data
data("xadmixture")

# keep only observations with K1 > 0.1 and K2 > 0.01
subset1 <- admix_subset(xadmixture, anc = c("K1", "K2"),
```

```r
```
subset2 <- admix_subset(xadmixture, anc = c("K2", "K3"), pct = c(0.4, 0.1), comparison = "less")

subset3 <- admix_subset(xadmixture, country = c("GBR", "FRA"), species = c("lorem", "dolor"))

library(magrittr)
subset4 <- admix_subset(xadmixture, anc = "K1", pct = 0.1, quiet = TRUE) %>% admix_subset(anc = "K4", pct = 0.3, comparison = "less", quiet = TRUE)

xadmixture

Simulated Admixture Data

Description

A dataset containing simulated admixture data of 600 observations.

Usage

xadmixture

Format

A data frame with 600 rows and 8 variables:

acc  Accession identifier

country Country where plant material was collected

species Name of species

K1,K2,K3,K4,K5  Admixture coefficients; expresses the proportions of the respective ancestries. Sum up to 1.
Source


Examples

# load simulated admixture data
data("xadmixture")

# create a subset of the data
xadmixture_sub <- admix_subset(xadmixture,
country = c("GBR", "FRA"),
anc = c("K1", "K2"),
pct = c(0.02, 0.2))

# generate a grouped & sorted stacked barplot
admix_barplot(xadmixture_sub,
    K = 4:ncol(xadmixture),
    sortkey = "K1",
    grouping = "country",
    palette = "turbo")
Index

* datasets
  xadmixture, 5
admix_barplot, 2
admix_subset, 4
xadmixture, 5