

Package ‘survivoR’

September 21, 2021

Type Package

Title Data from all Seasons of Survivor (US) TV Series in Tidy Format

Version 0.9.4

Description

Several datasets which detail the results and events of each season of Survivor. This includes details on the cast, voting history, immunity and reward challenges, jury votes and viewers. This data is useful for practicing data wrangling, graph analytics and analysing how each season of Survivor played out. Includes 'ggplot2' scales and colour palettes for visualisation.

Depends R (>= 3.5.0)

Imports dplyr, tidyr, ggplot2, stringr, magrittr

Suggests forcats, glue

License MIT + file LICENSE

URL <https://github.com/doehm/survivoR>

BugReports <https://github.com/doehm/survivoR/issues>

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Daniel Oehm [aut, cre]

Maintainer Daniel Oehm <danieloehm@gmail.com>

Repository CRAN

Date/Publication 2021-09-21 11:30:02 UTC

R topics documented:

castaways	2
challenges	3

clean_votes	4
hidden_idols	5
immunity	6
jury_votes	7
rewards	8
season_palettes	9
season_summary	9
survivor_pal	10
tribes_pal	11
tribe_colours	12
tribe_mapping	13
viewers	14
vote_history	15

Index	17
--------------	-----------

castaways	<i>Castaways</i>
-----------	------------------

Description

A dataset containing details on the castaways for each season

Usage

castaways

Format

This data frame contains the following columns:

season Season number

season_name Season name

full_name Full name of the castaway

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano

castaway Name of castaway. Generally this is the name they were most commonly referred to or nickname e.g. no one called Coach, Benjamin. He was simply Coach

age Age of the castaway during the season they played

city City of residence during the season they played

state State of residence during the season they played

personality_type The Myer-Briggs personality type of the castaway

day Number of days the castaway survived. A missing value indicates they later returned to the game that season

order Order in which castaway was voted out e.g. 5 is the 5th person voted of the island

result Final result
 jury_status Jury status
 original_tribe Original tribe name
 swapped_tribe Swapped tribe name
 swapped_tribe2 Second swapped tribe in the event of a second tribe swap or other tribe restructure such as absorbed tribe, outcasts, etc
 merged_tribe Merged tribe name
 total_votes_received Total number of tribal votes received during the main game for a given season (not overall for those who have played more than once). This includes votes from ties
 immunity_idols_won The number of immunity idols won by a castaway for the given season

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```

library(dplyr)
library(tidyr)
castaways %>%
  filter(season == 40)
  
```

challenges

Challenges

Description

A dataset detailing the challenges played including reward and immunity challenges. Note: The intention is for this dataset to ultimately replace the individual immunity and rewards datasets.

Usage

challenges

Format

This nested data frame contains the following columns:

season_name The season name
 season The season number
 episode Episode number
 title Episode title
 day The day of the tribal council
 challenge_type The challenge type e.g. immunity, reward, etc

challenge_name [under development] Name of the challenge played (TBA)
 winners_id The ID of the winners of the challenge. Consistent with castaway_id
 winners The list of winners. Either the list of people in the tribe which won, list of people that participated on the reward or the individual winner
 winning_tribe Name of the winner tribe. NA during the merge

Details

A nested tidy data frame of immunity and reward challenge results. The winners and winning tribe of the challenge are found by expanding the 'winners' column. For individual immunity challenges the winning tribe is simply 'NA'.

Typically in the merge if a single person win a reward they are allowed to bring others along with them. The first castaway in the expanded list is likely to be the winner and the subsequent players those they brought along with them. Although, not always. Occasionally in the merge the castaways are split into two teams for the purpose of the reward, in which case all castaways win the reward rather than a single person.

The 'day' field on this data set represents the day of the tribal council rather than the day of the challenge. This is to more easily associate the reward challenge with the immunity challenge and result of the tribal council. It also helps for joining tables.

Note the challenges table is the combined immunity and rewards tables which will eventually be dropped in later releases.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```
library(dplyr)
library(tidyr)
challenges %>%
  filter(season == 40) %>%
  unnest(winners)
```

clean_votes

Cleans votes

Description

There are certain events in the game of survivor which mean someone may attend tribal council and not get the change to vote for some reason or their vote is unique e.g. when rocks are drawn. You may want to remove the votes that were not an actual vote for a person. clean_votes is a convenience function to remove these records. Can be piped.

Usage

```
clean_votes(df)
```

Arguments

df Data frame which must contain the vote data.

Value

Returns a tidy data frame

Examples

```
library(dplyr)
vh <- vote_history %>%
  filter(
    season == 40,
    episode == 10
  ) %>%
  count(vote)
vh

vh %>%
  clean_votes()
```

hidden_idols

Hidden Immunity Idols

Description

A dataset containing the history of hidden immunity idols including who found them, on what day and which day they were played.

Usage

hidden_idols

Format

This data frame contains the following columns:

season_name The season name

season The season number

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano

castaway Name of the castaway

idol_number Indicates whether it is the first, second, etc idol found in the season

idols_held The number of idols held by the castaway

votes_nullified The number of votes nullified by the idol

day_found The day the idol was found

day_played The day of the tribal council

legacy_advantage If the idol was a legacy advantage or not

Source

https://survivor.fandom.com/wiki/Hidden_Immunity_Idol

immunity

Immunity challenges (deprecated)

Description

A dataset containing details on the immunity challenges for each season. This holds the same information as the challenges dataset. (superseded by the 'challenges' dataset)

Usage

```
immunity
```

Format

This nested tidy data frame contains the following columns:

season_name The season name

season The season number

episode Episode number of the immunity challenge was played

title Episode title

voted_out The castaway voted out

day Day the castaway or tribe won the immunity challenge

order Order in which the castaway was voted off the island

immunity Winners of the immunity challenge. Nested

Details

Contains details on tribal immunity and individual immunity. Currently it does not include details on hidden immunity idols. This will be added in time.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```
library(dplyr)
library(tidyr)
immunity
immunity %>%
  unnest(immunity)
```

jury_votes	<i>Jury votes</i>
------------	-------------------

Description

A dataset containing details on the final jury votes to determine the winner for each season

Usage

```
jury_votes
```

Format

This data frame contains the following columns:

season_name The season name

season The season number

castaway Name of the castaway

finalist The finalists for which a vote can be placed

vote Vote. 0-1 variable for easy summation

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano

finalist_id The ID of the finalist for which a vote can be placed. Consistent with castaway ID

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```
library(dplyr)
jury_votes %>%
  filter(season == 40) %>%
  group_by(finalist) %>%
  summarise(votes = sum(vote))
```

rewards

Reward challenges (deprecated)

Description

A dataset containing details on the reward challenges for each season, This holds the same information as the challenges dataset. (superceded by the 'challenges' dataset)

Usage

```
rewards
```

Format

This nested tidy data frame contains the following columns:

season_name Season name

season Sesaon number

episode Episode number of when the reward challenge was played

title Episode title

day Day of the immunity challenge rather than the reward (to be updated)

Reward Winners of the reward challenge. Tidy data frame. See details for more

Details

This is a nested data frame since more than one person can win the reward. The list of castaways include all those that participated in the reward rather than simply the castaway that won the challenge. Many challenges in the merge are such that there is one winner of the challenge and they can choose a set number of people to join them. Typically the first person on the list is the person who won the challenge and other just participated in the reward. In the case where castaways were split into teams for the challenge (post merge), technically they all won.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```
library(dplyr)
library(tidyr)
rewards %>%
  unnest(reward)
```

season_palettes	<i>Season palettes</i>
-----------------	------------------------

Description

A dataset containing palettes generated from the season logos

Usage

season_palettes

Format

This nested data frame contains the following columns:

season_name The season name

season The season number

palette The season palette

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

season_summary	<i>Season summary</i>
----------------	-----------------------

Description

A dataset containing a summary of all 40 seasons of Survivor

Usage

season_summary

Format

This data frame contains the following columns:

season_name Season name

season Sesaon number

location Location of the season

country Country the season was held

tribe_setup Initial setup of the tribe e.g. heroes vs Healers vs Hustlers

full_name Full name of the winner

winner_id ID for the winner of the season (primary key)
 winner Winner of the season
 runner_ups Runner ups for the season. Either one or two runner ups as a string
 final_vote Final vote allocation. See the jury_votes dataset for better aggregation of this data
 timeslot Timeslot of the show in the US
 premiered Date the first episode aired
 ended Date the season ended
 filming_started Date the filming of the season started
 filming_ended Date the filming ended (39 or 42 days after the start)
 viewers_premier Number of viewers (millions) who tuned in for the premier
 viewers_finale Number of viewers (millions) who tuned in for the finale
 viewers_reunion Number of viewers (millions) who tuned in for the reunion
 viewers_mean Average number of viewers (millions) who tuned in over the season
 rank Season rank

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

survivor_pal

Survivor season colour palette

Description

ggplot2 scales for each season of Survivor.

Usage

```
survivor_pal(season = NULL, scale_type = "d", reverse = FALSE, ...)
```

```
scale_fill_survivor(season = NULL, scale_type = "d", reverse = FALSE, ...)
```

```
scale_colour_survivor(season = NULL, scale_type = "d", reverse = FALSE, ...)
```

Arguments

season	Season number
scale_type	Discrete or continuous. Input d or c.
reverse	Logical. Reverse the palette?
...	Other arguments passed on to methods.

Details

Palettes are created from the logo for the season.

Value

Scale functions for ggplot2

Examples

```
library(ggplot2)
library(dplyr)
mpg %>%
  ggplot(aes(x = displ, fill = manufacturer)) +
  geom_histogram(colour = "black") +
  scale_fill_survivor(40)
```

tribes_pal

Tribes colour palette

Description

To create scale functions for ggplot. Given a season of Survivor, a palette is created from the tribe colours for that season including the merged tribe.

Usage

```
tribes_pal(season = NULL, scale_type = "d", reverse = FALSE, tribe = NULL, ...)
```

```
scale_fill_tribes(season = NULL, scale_type = "d", reverse = FALSE, ...)
```

```
scale_colour_tribes(season = NULL, scale_type = "d", reverse = FALSE, ...)
```

Arguments

season	Season number
scale_type	Discrete or continuous. Input d or c.
reverse	Logical. Reverse the palette?
tribe	Tribe names. Default NULL
...	Other arguments passed on to methods.

Details

If it is intended the colours will correspond to the tribes e.g. a stacked bar chart of votes given to each finalist and the colour corresponds to their original tribe (as in the example below), the tribe vector needs to be passed to the scale function (for now). If no tribe vector is given it will simply treat the tribe colours as a colour palette.

Value

Scale functions for ggplot2

Examples

```
library(ggplot2)
library(stringr)
library(dplyr)
library(glue)
ssn <- 35
labels <- castaways %>%
  filter(
    season == ssn,
    str_detect(result, "Sole|unner")
  ) %>%
  select(castaway, original_tribe) %>%
  mutate(label = glue("{castaway} ({original_tribe}")) %>%
  select(label, castaway)
jury_votes %>%
  filter(season == ssn) %>%
  left_join(
    castaways %>%
      filter(season == ssn) %>%
      select(castaway, original_tribe),
    by = "castaway"
  ) %>%
  group_by(finalist, original_tribe) %>%
  summarise(votes = sum(vote)) %>%
  left_join(labels, by = c("finalist" = "castaway")) %>% {
  ggplot(., aes(x = label, y = votes, fill = original_tribe)) +
    geom_bar(stat = "identity", width = 0.5) +
    scale_fill_tribes(ssn, tribe = .$original_tribe) +
    theme_minimal() +
    labs(
      x = "Finalist (original tribe)",
      y = "Votes",
      fill = "Original\\ntribe",
      title = "Votes received by each finalist"
    )
  }
}
```

tribe_colours

Tribe colours

Description

A dataset containing the tribe colours for each season

Usage

```
tribe_colours
```

Format

This data frame contains the following columns:

season_name The season name

season The season number

tribe Tribe name

tribe_colour Colour of the tribe

tribe_status Tribe status e.g. original, swapped or merged. In the instance where a tribe is formed at the swap by splitting 2 tribes into 3, the 3rd tribe will be labelled 'swapped'

Source

<https://survivor.fandom.com/wiki/Tribe>

Examples

```
library(ggplot2)
library(dplyr)
library(forcats)
df <- tribe_colours %>%
  group_by(season_name) %>%
  mutate(
    xmin = 1,
    xmax = 2,
    ymin = 1:n(),
    ymax = ymin + 1
  ) %>%
  ungroup() %>%
  mutate(
    season_name = fct_reorder(season_name, season),
    font_colour = ifelse(tribe_colour == "#000000", "white", "black")
  )
ggplot() +
  geom_rect(data = df,
    mapping = aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),
    fill = df$tribe_colour) +
  geom_text(data = df,
    mapping = aes(x = xmin+0.5, y = ymin+0.5, label = tribe),
    colour = df$font_colour) +
  theme_void() +
  facet_wrap(~season_name, scales = "free_y")
```

 tribe_mapping

Tribe mapping

Description

A mapping for castaways to tribes for each day (day being the day of the tribal council) This is useful for observing who is on what tribe throughout the game.

Usage

tribe_mapping

Format

This data frame contains the following columns:

season_name The season name

season The season number

day The day of the tribal council

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano

castaway Name of the castaway

tribe Name of the tribe the castaway was on

Details

Each season by day holds a complete list of castaways still in the game and which tribe they are on. Moving through each day you can observe the changes in the tribe. For example the first day (usual day 2) has all castaways mapped to their original tribe. The next day has the same minus the castaway just voted out. This is useful for observing the changes in tribe make either due to castaways being voted off the island, tribe swaps or otherwise.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

viewers

Viewers

Description

A dataset containing the viewer history for each season and episode

Usage

viewers

Format

This data frame contains the following columns:

season_name The season name

season Season number

episode_number_overall The cumulative episode number

episode Episode number for the season

title Episode title
 episode_date Date the episode aired
 viewers Number of viewers (millions) who tuned in
 rating_18_49 TV rating for the 18-49 aged group
 share_18_49 TV share for the 18-49 aged group

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

vote_history	<i>Vote history</i>
--------------	---------------------

Description

A dataset containing details on the vote history for each season

Usage

vote_history

Format

This data frame contains the following columns:

season_name The season name
 season The season number
 episode Episode number
 day Day the tribal council took place
 tribe_status The status of the tribe e.g. original, swapped, merged, etc. See details for more
 castaway Name of the castaway
 immunity Type of immunity held by the castaway at the time of the vote e.g. individual, hidden
 (see details for hidden immunity data)
 vote The castaway for which the vote was cast
 nullified Was the vote nullified by a hidden immunity idol? Logical
 voted_out The castaway who was voted out
 order The order in which the castaway was voted off the island
 vote_order In the case of ties this indicates the order the votes took place
 castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g.
 Amber Brkich / Amber Mariano
 vote_id ID of the castaway voted for
 voted_out_id ID of the castaway voted_out

Details

This data frame contains a complete history of votes cast across all seasons of Survivor. While there are consistent events across the seasons there are some unique events such as the 'mutiny' in Survivor: Cook Islands (season 13) or the 'Outcasts' in Survivor: Pearl Islands (season 7). For maintaining a standard, whenever there has been a change in tribe for the castaways it has been recorded as swapped. swapped is used as the term since 'the tribe swap' is a typical recurring milestone in each season of Survivor. Subsequent changes are recorded with a trailing digit e.g. swapped2. This includes absorbed tribes e.g. Stephanie was 'absorbed' in Survivor: Palau (season 10) and when 3 tribes are reduced to 2. These cases are still considered 'swapped' to indicate a change in tribe status.

Some events result in a castaway attending tribal but not voting. These are recorded as

Win The castaway won the fire challenge

Lose The castaway lost the fire challenge

None The castaway did not cast a vote. This may be due to a vote steal or some other means

Immune The castaway did not vote but were immune from the vote

Where a castaway has immunity == 'hidden' this means that player is protected by a hidden immunity idol. It may not necessarily mean they played the idol, the idol may have been played for them. While the nullified votes data is complete the immunity data does not include those who had immunity but did not receive a vote. This is a TODO.

In the case where the 'steal a vote' advantage was played, there is a second row for the castaway that stole the vote. The castaway who had their vote stolen are is recorded as None.

Many castaways have been medically evacuated, quit or left the game for some other reason. In these cases where no votes were cast there is a skip in the order variable. Since no votes were cast there is nothing to record on this data frame. The correct order in which castaways departed the island is recorded on castaways.

In the case of a tie, voted_out is recorded as tie to indicate no one was voted off the island in that instance. The re-vote is recorded with vote_order = 2 to indicate this is the second round of voting. In the case of a second tie voted_out is recorded as tie2. The third step is either a draw of rocks, fire challenge or countback (in the early days of survivor). In these cases vote is recorded as the colour of the rock drawn, result of the fire challenge or 'countback'.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series))

Examples

```
# The number of times Tony voted for each castaway in Survivor: Winners at War
library(dplyr)
vote_history %>%
  filter(
    season == 40,
    castaway == "Tony"
  ) %>%
  count(vote)
```


Index

* datasets

- castaways, [2](#)
- challenges, [3](#)
- hidden_idols, [5](#)
- immunity, [6](#)
- jury_votes, [7](#)
- rewards, [8](#)
- season_palettes, [9](#)
- season_summary, [9](#)
- tribe_colours, [12](#)
- tribe_mapping, [13](#)
- viewers, [14](#)
- vote_history, [15](#)

- castaways, [2](#)
- challenges, [3](#)
- clean_votes, [4](#)

- hidden_idols, [5](#)

- immunity, [6](#)

- jury_votes, [7](#)

- rewards, [8](#)

- scale_colour_survivor (survivor_pal), [10](#)

- scale_colour_tribes (tribes_pal), [11](#)

- scale_fill_survivor (survivor_pal), [10](#)

- scale_fill_tribes (tribes_pal), [11](#)

- season_palettes, [9](#)

- season_summary, [9](#)

- survivor_pal, [10](#)

- tribe_colours, [12](#)

- tribe_mapping, [13](#)

- tribes_pal, [11](#)

- viewers, [14](#)

- vote_history, [15](#)