

Package: survivoR (via r-universe)

November 1, 2024

Type Package

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

Version 2.3.4

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 tidy, ggplot2, stringr, magrittr, glue, shiny, purrr, dplyr, crayon, readr, shinycssloaders, lubridate, DT, shinyjs

Suggests forcats, testthat (>= 3.0.0)

License MIT + file LICENSE

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

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

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Config/testthat/edition 3

Repository <https://doehm.r-universe.dev>

RemoteUrl <https://github.com/doehm/survivor>

RemoteRef HEAD

RemoteSha ad5aae149a411b5875580fefeb0e64c62b9a6ed4

Contents

advantage_details	2
advantage_movement	3
auction_details	4
boot_mapping	5
castaways	6
castaway_details	8
challenge_description	9
challenge_results	12
challenge_summary	13
confessionals	14
episodes	15
episode_summary	16
get_castaway_image	16
get_confessional_timing	17
jury_votes	18
launch_confessional_app	19
screen_time	19
season_palettes	20
season_summary	21
survivor_auction	22
survivor_pal	23
tribes_pal	24
tribe_colours	25
tribe_mapping	27
viewers	28
vote_history	29
Index	32

advantage_details	<i>Advantage Details</i>
-------------------	--------------------------

Description

A dataset containing the details and characteristics of each idol and advantage. This maps to ‘advantage_movement’

Usage

advantage_details

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

season_name The season name

season The season number

advantage_id The ID / primary key of the advantage

advantage_type Advantage type e.g. hidden immunity idol, extra vote, steal a vote, etc

clue_details Details if a clue existed for the advantage and if so where was the clue found

location_found The location the idol or advantage was found

conditions Extra details about the unique conditions of the idol or advantage

Details

There are split idols which need to be combined to be played. In these case the first one found is given an ID. The second or subsequent parts are given the same ID with a trailing letter. For example in season 40 Denise found an idol that was split (USHI4002). Later she found the other half (USHI4002b). When played the second half is considered to have 'absorbed' into the first idol. The first idol found is always considered the primary idol.

advantage_movement *Advantage Movement*

Description

A dataset containing the movement details of each advantage or hidden immunity idol. Each row is considered an event e.g. the idol was found, played, etc. If the advantage changed hands it records who received it. The logical flow is identified by the 'sequence_id'.

Usage

advantage_movement

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

season_name The season name

season The season number

castaway Name of the castaway involved in the event e.g. found, played, received, etc.

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU.
 advantage_id The ID / primary key of the advantage
 sequence_id The sequence of events. For example 'sequence_id == 1' usually means the advantage was found. Each subsequent event follows the 'sequence_id'
 day The day the event occurred
 episode The episode the event occurred
 event The event e.g. the advantage was found, played, received, etc
 played_for If the advantage or idol was played this records who it was played for
 played_for_id the ID for who the advantage or idol was played for
 success If the play was successful or not. Only relevant for advantages since playing a hidden immunity idol is always successful in terms of saving who it was played for.
 votes_nullified In the case of hidden immunity idols this is the count of how many votes were nullified when played

 auction_details

Survivor Auction Details

Description

The details of the items purchased at the Survivor Auction. survivor_auction is at the castaway level and includes all castaways whether or not they purchased an item and auction_details is at the item level.

Usage

auction_details

Format

This data frame contains the following columns:

version Country code for the version of the show
 version_season Version season key
 season_name The season name
 season The season number
 item Item number
 item_description Item description
 category The item category. See details for more.
 castaway Castaway
 castaway_id Castaway ID

covered If the item was covered or not
 cost The amount paid for the item
 money_remaining How much money the castaway has remaining
 auction_num If the same item is auctioned for a second time it has a value of 2
 participated The names of castaways that could participate in the purchased item e.g. sharing a tub of peanut butter with the tribe
 notes Additional notes
 alternative_offered If an alternative was offered to the player after purchase
 alternative_accepted If they accepted the alternative offer
 other_item Description of the refused item
 other_item_category Category of the refused item

Details

Each item has been categorised into 5 main categories: 1. Food and drink: The most common item. It may be simply food or drink, not necessarily both. 2. Comfort: Things like a shower, toothpaste, etc 3. Letters from home 4. Advantage: Could be a clue to a hidden immunity idol, advantage in the next challenge, or in the current auction 5. Bad item: The not good item, typically one of the covered items. Whether or not it's actually bad is subjective, but where someone is hoping for pizza and gets bat soup I consider it a bad item.

Source

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

boot_mapping	<i>Boot mapping</i>
--------------	---------------------

Description

A mapping table for easily filtering to the set of castaways that are still in the game after a specified number of boots.

Usage

boot_mapping

Format

This data frame contains the following columns:

version Country code for the version of the show
 version_season Version season key
 season_name The season name

season The season number
episode Episode number
order The number of boots that there have been in the game e.g. if 'order == 2' there have been 2 boots in the game so far and there are N-2 castaways left in the game
final_n The final number of castaways e.g. you can filter to the final 4 by 'filter(boot_mapping, final_n == 4)'. There are missing values where players have returned to the game. This means there are multiple stages of the game where there is a different make up of the final 8, for example. This field just takes the last set so that you can filter for 'final_n' and it will return a single set of castaways.
n_boots Similar to 'final_n' but the number of boots in the game. This is different to 'order' where order counts if someone has been booted twice. 'n_boots' is simply the number of people in the season minus the 'final_n'.
sog_id Stage of game ID for joining to vote_history and challenge_results
castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU.
castaway Name of the castaway
tribe Name of the tribe the castaway was on
tribe_status The status of the tribe e.g. original, swapped, merged, etc. See details for more
game_status Logical flag to identify if the castaway is currently in the game. If 'FALSE' the castaway is on Redemption Island or Edge of Extinction.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series)) https://survivor.fandom.com/wiki/Main_Page

castaways

Castaways

Description

A dataset containing details on the results for every castaway and season

Usage

castaways

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

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. The first two letters reference the country of the version played e.g. US, AU (TBA).

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

episode Episode number

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

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

result Final result

result_number Result number i.e. the final place. NA for castaways that were voted out but later returned e.g. Redemption Island

jury_status Jury status

original_tribe Original tribe name

finalist Logical. TRUE if the castaway was a finalists

jury Logical. TRUE if the castaway was a jury member

winner Logical. TRUE if the castaway was the winner

acknowledge Did the contestant acknowledge their teammates in one of these specific ways after snuffing — or just walk away?

ack_gesture for any physical gestures towards the tribe after torch snuffing. Types: wave, nod, wink, bow or prayer sign with hands

ack_look For making eye contact with one or more members of the tribe after torch snuffing

ack_smile For smiling at the tribe after torch snuffing

ack_speak For any verbal communication directed at the tribe after torch snuffing

ack_quote What, if anything, the contestant said. Direct quotes only.

ack_score The score is derived from the four subcategories of acknowledgment: words, look, gesture, and smile. Each true value in these categories adds 1 to the score.

Details

Note that in the seasons where castaways returned to the game e.g. Redemption Island, a castaway may appear twice.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series)); https://survivor.fandom.com/wiki/Main_Page; ack_ features from Matt Stiles <https://github.com/stiles/survivor-voteoffs>

Examples

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

castaway_details	<i>Castaway details</i>
------------------	-------------------------

Description

A dataset containing details on the castaways for each season

Usage

```
castaway_details
```

Format

This data frame contains the following columns:

`castaway_id` ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU (TBA).

`full_name` Full name of the castaway

`full_name_detailed` A detailed version of `full_name` for plotting e.g. 'Boston' Rob Mariano

`castaway` Short name of the castaway. Name typically used during the season. Sometimes there are multiple people with the same name e.g. Rob C and Rob M in Survivor All-Stars. This field takes the most verbose name used

`date_of_birth` Date of birth

`date_of_death` Date of death

`gender` Gender of castaway

`african` TRUE if African-American or African-Canadian as per https://survivor.fandom.com/wiki/Main_Page

`asian` TRUE if Asian-American or Asian-Canadian as per https://survivor.fandom.com/wiki/Main_Page

latin_american TRUE if Latin-American as per https://survivor.fandom.com/wiki/Main_Page
 native_american TRUE if Native-American as per https://survivor.fandom.com/wiki/Main_Page
 bipoc Black, Indigenous, or Person of Colour
 lgbt LGBTQIA+ status as listed on the survivor wiki.
 personality_type The Myer-Briggs personality type of the castaway
 occupation Occupation
 three_words Answer to the question "three words to describe you?"
 hobbies Answer to the question "what are your favourite hobbies?"
 pet_peeves Answer to the question "what are your pet peeves?"
 race Race (if known)
 ethnicity Ethnicity (if known)

Details

Race and ethnicity data is included if known and can point to a source, rather than making an assumption about an individual.

poc has been deprecated and replaced with bipoc which is now logical and only for the US. bipoc is TRUE if any of african, asian, latin_american, or native_american is TRUE.

Source

https://survivor.fandom.com/wiki/Main_Page, <https://www.personality-database.com/>

Examples

```
library(dplyr)
castaway_details |>
  count(gender)
```

challenge_description *Challenge Description*

Description

A dataset detailing the challenges played and the elements they include over all seasons of Survivor

Usage

```
challenge_description
```

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

season_name The season name

season The season number

episode Episode number

challenge_id Primary key

challenge_number

challenge_type

name The name of the challenge

recurring_name Challenges can go by different names but are often associated with a particular challenge or element of a challenge. Some challenges use combinations of other challenges so it's not perfect but consistent with the wiki page. Use recurring_name to analyse how often a challenge has been run.

description Description of the challenge

reward Description of the reward

additional_stipulation Some challenges come with various rules or success criteria. This states those conditions.

race If the challenge is a race between tribes, teams or individuals

endurance If the challenge is an endurance event e.g. last tribe, team, individual standing

turn_based If the challenge is turn bases i.e. conducted in rounds

puzzle If the challenge contains a puzzle element

puzzle_slide If the challenge contained a slide puzzle

puzzle_word If the challenge contained a word puzzle

precision If the challenge contains a precision element e.g. shooting an arrow, hitting a target, etc

precision_catch If the challenge featured catching a ball or similar

precision_roll_ball If the challenge featured rolling a ball

precision_slingshot If the challenge featured a slingshot, either the large version or handheld version

precision_throw_balls If the challenge featured throwing balls

precision_throw_coconuts If the challenge featured throwing coconuts

precision_throw_rings if the challenge featured throwing rings

precision_throw_sandbags if the challenge featured throwing sandbags

strength If the challenge has a strength based

balance If the challenge contains a balancing element. My refer to the player balancing on something or the player balancing an object on something e.g. The Ball Drop

balance_beam If the challenge featured a balance beam of similar they were required to balance on

balance_ball If the challenge featured balancing a ball on something

food If the challenge contains a food element e.g. the food challenge, biting off chunks of meat

knowledge If the challenge contains a knowledge component e.g. Q and A about the location

memory If the challenge contains a memory element e.g. memorising a sequence of items

fire If the challenge contains an element of fire making / maintaining

water If the challenge is held, in part, in the water

water_swim If castaways had to swim in the challenge

water_paddling If castaways were required to paddle a boat or similar

obstacle_blindfolded If the challenge required castaways to be blindfolded

obstacle_cargo_net If the challenge featured a cargo net

obstacle_chopping If castaways were required to chop a rope or similar

obstacle_combination_lock If the challenge feature a combination lock

obstacle_digging If the challenge involved digging

obstacle_knots If the challenge involved untying knots

obstacle_padlocks If the challenge featured opening padlocks

mud If the challenge required castaways to get covered in mud

Details

This data set contains the name, description, and descriptive features for each challenge where it is known. Challenges can go by different names so have included the unique name and the recurring challenge name. These are taken directly from the [Survivor Wiki](https://survivor.fandom.com/wiki/Category:Recurring_Challenges). Sometimes there can be variations made on the challenge but go by the same name, or the challenge is integrated with a longer obstacle. In these cases the challenge may share the same recurring challenge name but have a different challenge name. Even if they share the same names the description could be different.

The features of each challenge have been determined largely through string searches of key words that describe the challenge. It may not be 100 different and inconsistent descriptions but in most part they will provide a good basis for analysis.

If any descriptive features need altering please let me know in the [issues](https://github.com/doehm/survivoR/issues).

For updated data please see the git version.

Source

<https://survivor.fandom.com/wiki/Category:Challenges> https://survivor.fandom.com/wiki/Main_Page

Examples

```
library(dplyr)
library(tidyr)
challenge_description
```

challenge_results *Challenge Results*

Description

A dataset detailing the challenges played including reward and immunity challenges.

Usage

challenge_results

Format

This data frame contains the following columns

version Country code for the version of the show

version_season Version season key

season_name The season name

season The season number

episode Episode number

n_boots The number of boots that there have been in the game e.g. if 'n_boots == 2' there have been 2 boots in the game so far and there are N-2 castaways left in the game

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU (TBA).

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

outcome_type Whether the challenge is individual or tribal. Some individual reward challenges may involve multiple castaways as the winner gets to choose who they bring along

tribe Current tribe the castaway is on

tribe_status The status of the tribe e.g. original, swapped, merged, etc. See details for more

challenge_type The challenge type e.g. immunity, reward, etc

challenge_id Primary key to the challenge_description data set which contains features of the challenge

result Result of challenge

result_notes Additional notes about the result of the challenge

order_of_finish Order of finish for tribal challenges. Useful when there are 3 or more tribes to see who actually came first, second and who lost the challenge.

chosen_for_reward If after the reward challenge the castaway was chosen to participate in the reward

sit_out TRUE if they sat out of the challenge or FALSE if they participate

team Team allocation when they are split into teams

sog_id Stage of game ID for joining to boot_mapping and vote_history

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series)) https://survivor.fandom.com/wiki/Main_Page

Examples

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

challenge_summary	<i>Challenge Summary</i>
-------------------	--------------------------

Description

A dataset summarising challenge_results

Usage

```
challenge_summary
```

Format

This data frame contains the following columns

category The category of the challenge e.g. tribal, individual, individual immunity, duel, etc. This makes it easy to split out the difference types of challenges and avoid complications such as 'Team / Individual' challenges where there is a dependent outcome structure. Join to challenge_results using challenge_id, version_season and castaway_id

version_season Version season key

challenge_id Primary key to the challenge_description data set which contains features of the challenge

challenge_type The challenge type e.g. immunity, reward, etc

outcome_type Whether the challenge is individual or tribal. Some individual reward challenges may involve multiple castaways as the winner gets to choose who they bring along

tribe Current tribe the castaway is on

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

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU (TBA).

n_entities Number of entities competing for the win e.g. the number of tribes, teams, or people.

`n_winners` Number of winners (or winning entities) e.g. if there are two tribes there is only one winning tribe, if there are three tribes like the new era there are two winning tribes and one that goes to tribal council.

`n_in_team` The number of people in the tribe or team

`won` If the castaway won

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series)) https://survivor.fandom.com/wiki/Main_Page

Examples

```
library(dplyr)
library(tidyr)
challenge_summary %>%
  filter(version_season == 46)
```

confessionals	<i>Confessionals</i>
---------------	----------------------

Description

A dataset containing the count of confessionals per castaway per episode. A confessional is when the castaway is speaking directly to the camera about their game.

Usage

```
confessionals
```

Format

This data frame contains the following columns:

`version` Country code for the version of the show

`version_season` Version season key

`season_name` The season name

`season` The season number

`episode` Episode number

`castaway` Name of the castaway

`castaway_id` ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU.

`confessional_count` The count of confessionals for the castaway during the episode

`confessional_time` The total time for all confessionals for the episode for each castaway

`index_count` The index based on the confessional counts. See details.

`index_time` The index based on the confessional time. See details.

Details

Confessional data has been counted by contributors of the survivoR R package and consolidated with external sources. The aim is to establish consistency in confessional counts in the absence of official sources. Given the subjective nature of the counts and the potential for clerical error no single source is more valid than another. Therefore, it is reasonable to average across all sources.

In the case of double or extended episodes, if the episode only has one title it is considered a single episode. This means the average number of confessionals per person is likely to be higher for this episode given it's length. If there are two episode titles the confessionals are counted for the appropriate episode. This is to ensure consistency across all other datasets.

In the case of recap episodes, this episode is left blank.

The indexes are a measure of how many more confessional counts or time the castaway has received given the point in the game. For example a 'index_count' of 1 implies the castaway has received the expected number of confessionals given equal share within tribe. An index of 1.5 implies have received 50 typically receives more confessionals for the episode. Makes sense. 'index_time' is the same but using time instead of counts.

If you also count confessionals, please get in touch and I'll add them into the package.

 episodes

Episodes

Description

A dataset containing details for each episode

Usage

episodes

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

season_name The season name

season Season number

episode_number_overall The cumulative episode number

episode Episode number for the season

episode_title Episode title

episode_label A standardised episode label

episode_date Date the episode aired

episode_length Episode length in minutes

viewers Number of viewers (millions) who tuned in

imdb_rating IMDb rating for the episode on a scale of 0-10
 n_ratings The number of ratings submitted to IMDb
 episode_summary Description of the episode from wikipedia

Source

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

episode_summary	<i>Episode summary</i>
-----------------	------------------------

Description

A dataset containing a summary of all US episodes seasons of Survivor

Usage

episode_summary

Format

This data frame contains the following columns:

version Country code for the version of the show
 version_season Version season key
 episode Episode number
 episode_summary summary of the episode

Source

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

get_castaway_image	<i>Castaway images</i>
--------------------	------------------------

Description

Returns the URL for the image of the specified castaways by their 'castaway_id' and season / version they were in

Usage

get_castaway_image(castaway_ids, version_season)

Arguments

castaway_ids Castaway ID
version_season Version season key for the season they played

Value

Character vector of URLs

Examples

```
library(dplyr)

survivoR::castaways %>%
  filter(version_season == "US42") %>%
  mutate(castaway_image = get_castaway_image(castaway_id, version_season))
```

get_confessional_timing
Confessional time

Description

Takes the output of the times recorded from the Shiny app and aggregates to the final confessional times and confessional counts. `confessional_time` is the total duration in seconds for the episode. `confessional_count` is the number of confessionals recorded to be at least 10 seconds apart.

Usage

```
get_confessional_timing(x, .vs, .episode, .mda = 3)
```

Arguments

`x` Either a data frame or path(s) to the csv file containing all the time stamps from the Shiny app
`.vs` Version season
`.episode` Episode
`.mda` Missing duration adjustment (MDA) in seconds. If either start or stop is missing from the records, the missing value is imputed with a 3 second adjustment by default.

Value

data frame

Examples

```
# After running app and recording confessionals, run...
# Example from a saved timing file

library(readr)

path <- system.file(package = "survivoR", "extdata/US4412.csv")
df_us4412 <- read_csv(path)
get_confessional_timing(df_us4412, .vs = "US44", .episode = 12)
```

jury_votes

*Jury votes***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:

version Country code for the version of the show

version_season Version season key

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. The first two letters reference the country of the version played e.g. US, AU.

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))
```

launch_confessional_app
Launch Confessional App

Description

Launches the confessional timing app in either a browser or viewer. Default is set to browser. The user is required to provide a path for which the time stamps are recorded.

Usage

```
launch_confessional_app(browser = TRUE, path = NULL, write = TRUE)
```

Arguments

browser	Open in browser instead of viewer. Default TRUE
path	Parent directory for output files. Default is a sub-folder 'confessional-timing' in the current working directory.
write	Write to disc. Default TRUE.

Value

An active R shiny application

Examples

```
## Only run this example in interactive R sessions

if(interactive()) {

  # launch app
  # launch_confessional_app()

}
```

screen_time *Screen Time*

Description

A dataset summarising the screen time of contestants on the TV show Survivor. Currently only contains Season 1-4 and 42.

Usage

```
screen_time
```

Format

This data frame contains the following columns:

version_season Version season key

episode Episode number

castaway_id ID of the castaway (primary key). Also includes two special IDs of host (i.e. Jeff Probst) or unknown (the image detection couldn't identify the face with sufficient accuracy)

screen_time Estimated screen time for the individual in seconds.

Details

Individuals' screen time is calculated, at a high-level, via the following process:

1. Frames are sampled from episodes on a 1 second time interval
2. MTCNN detects the human faces within each frame
3. VGGFace2 converts each detected face into a 512d vector space
4. A training set of labelled images (1 for each contestant + 3 for Jeff Probst) is processed in the same way to determine where they sit in the vector space. **TODO:** This could be made more accurate by increasing the number of training images per contestant.
5. The Euclidean distance is calculated for the faces detected in the frame to each of the contestants in the season (+Jeff). If the minimum distance is greater than 1.2 the face is labelled as "unknown". **TODO:** Review how robust this distance cutoff truly is - currently based on manual review of Season 42.
6. A multi-class SVM is trained on the training set to label faces. For any face not identified as "unknown", the vector embedding is run into this model and a label is generated.
7. All labelled faces are aggregated together, with an assumption of 1 full second of screen time each time a face is seen.

season_palettes

Season palettes

Description

A dataset containing palettes generated from the season logos

Usage

season_palettes

Format

This nested data frame contains the following columns:

version Country code for the version of the show
 version_season Version season key
 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 seasons of Survivor

Usage

season_summary

Format

This data frame contains the following columns:

version Country code for the version of the show
 version_season Version season key
 season_name Season name
 season Season number
 n_cast Number of cast in the season
 n_tribes Number of starting tribes
 n_finalists Number of finalists
 n_jury Number of jury members
 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 data set 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_premiere 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)) https://survivor.fandom.com/wiki/Main_Page

survivor_auction	<i>Survivor Auction</i>
------------------	-------------------------

Description

A dataset showing who attended the Survivor Auction during the seasons they were held. `survivor_auction` is at the castaway level and includes all castaways whether or not they purchased an item and `auction_details` is at the item level.

Usage

```
survivor_auction
```

Format

This data frame contains the following columns:

`version` Country code for the version of the show
`version_season` Version season key
`season_name` The season name
`season` The season number
`episode` Episode number
`n_boots` The number of boots so far in the game

castaway_id ID of the castaway (primary key). Consistent across seasons and name changes e.g. Amber Brkich / Amber Mariano. The first two letters reference the country of the version played e.g. US, AU (TBA).

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

tribe_status The status of the tribe e.g. original, swapped, merged, etc. See details for more

tribe Tribe name

currency Currency

total Total amount either given to or found by the castaway

Source

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

survivor_pal	<i>Survivor season colour palette</i>
--------------	---------------------------------------

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

Scale functions for ggplot2

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
Scale functions for ggplot2
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:

version Country code for the version of the show

version_season Version season key

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	<i>Tribe mapping</i>
---------------	----------------------

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:

version Country code for the version of the show

version_season Version season key

season_name The season name

season The season number

episode Episode 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. The first two letters reference the country of the version played e.g. US, AU.

castaway Name of the castaway

tribe Name of the tribe the castaway was on

tribe_status The status of the tribe e.g. original, swapped, merged, etc. See details for more

Details

Each season by episode and 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 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, who is on Redemption Island and Edge of Extinction.

Source

[https://en.wikipedia.org/wiki/Survivor_\(American_TV_series\)](https://en.wikipedia.org/wiki/Survivor_(American_TV_series)) https://survivor.fandom.com/wiki/Main_Page

viewers	<i>Viewers</i>
---------	----------------

Description

A dataset containing the viewer history for each season and episode

Usage

viewers

Format

This data frame contains the following columns:

version Country code for the version of the show

version_season Version season key

season_name The season name

season Season number

episode_number_overall The cumulative episode number

episode Episode number for the season

episode_title Episode title

episode_label A standardised episode label

episode_date Date the episode aired

episode_length Episode length in minutes

viewers Number of viewers (millions) who tuned in

imdb_rating IMDb rating for the episode on a scale of 0-10

n_ratings The number of ratings submitted to IMDb

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:

version Country code for the version of the show

version_season Version season key

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

tribe Tribe name

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

vote_event Extra details on the vote e.g. Won or lost the fire challenge, played an extra vote, etc

vote_event_outcome The outcome of the vote event

split_vote If there was a decision to split the vote this records who the vote was split with. Helps to identify successful boots

nullified Was the vote nullified by a hidden immunity idol? Logical

tie If the set of votes resulted in a tie. Logical

voted_out The castaway who was voted out

order Boot order. Order in which castaway was voted out e.g. 5 is the 5th person voted of 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. The first two letters reference the country of the version played e.g. US, AU.

vote_id ID of the castaway voted for

voted_out_id ID of the castaway voted_out
 sog_id Stage of game ID for joining to boot_mapping and challenge_results
 challenge_id Primary key to the challenge_description data set which contains features of the challenge. The helps map the immunity challenge which result in the tribal.

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

- advantage_details, 2
 - advantage_movement, 3
 - auction_details, 4
 - boot_mapping, 5
 - castaway_details, 8
 - castaways, 6
 - challenge_description, 9
 - challenge_results, 12
 - challenge_summary, 13
 - confessionals, 14
 - episode_summary, 16
 - episodes, 15
 - jury_votes, 18
 - screen_time, 19
 - season_palettes, 20
 - season_summary, 21
 - survivor_auction, 22
 - tribe_colours, 25
 - tribe_mapping, 27
 - viewers, 28
 - vote_history, 29
-
- advantage_details, 2
 - advantage_movement, 3
 - auction_details, 4
-
- boot_mapping, 5
-
- castaway_details, 8
 - castaways, 6
 - challenge_description, 9
 - challenge_results, 12
 - challenge_summary, 13
 - confessionals, 14
-
- episode_summary, 16
 - episodes, 15
-
- get_castaway_image, 16
 - get_confessional_timing, 17
-
- jury_votes, 18
-
- launch_confessional_app, 19
-
- scale_colour_survivor (survivor_pal), 23
 - scale_colour_tribes (tribes_pal), 24
 - scale_fill_survivor (survivor_pal), 23
 - scale_fill_tribes (tribes_pal), 24
 - screen_time, 19
 - season_palettes, 20
 - season_summary, 21
 - survivor_auction, 22
 - survivor_pal, 23
-
- tribe_colours, 25
 - tribe_mapping, 27
 - tribes_pal, 24
-
- viewers, 28
 - vote_history, 29