

Package ‘whatr’

March 19, 2020

Title Scrape and Analyze the J! Archive

Version 1.0.0

Description Scrape the fan-made J! Archive <<https://www.j-archive.com/>>
for Jeopardy episode contestants, categories, clues, answers, and scores.

License GPL-3

URL <https://github.com/kiernann/whatr>

BugReports <https://github.com/kiernann/whatr/issues>

Depends R (>= 3.2)

Imports dplyr (>= 0.8.3), ggplot2 (>= 3.2.1), httr (>= 1.4.1),
magrittr (>= 1.5), methods, purrr (>= 0.3.3), rlang (>= 0.4.4),
rvest (>= 0.3.4), scales (>= 1.0.0), stringr (>= 1.4.0), tibble
(>= 2.1.3), tidyr (>= 1.0.0), xml2 (>= 1.2.2)

Suggests covr (>= 3.3.2), spelling (>= 2.1), testthat (>= 2.3.0)

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.0.2

NeedsCompilation no

Author Kiernan Nicholls [aut, cre]

Maintainer Kiernan Nicholls <kiernann@protonmail.com>

Repository CRAN

Date/Publication 2020-03-19 13:30:02 UTC

R topics documented:

boards	2
entity_clean	3
episodes	3
players	4

scores	4
synopses	5
whatr_airdate	5
whatr_answers	6
whatr_board	7
whatr_categories	8
whatr_clues	9
whatr_data	10
whatr_doubles	10
whatr_html	11
whatr_order	12
whatr_players	13
whatr_plot	14
whatr_scores	14
whatr_synopsis	15

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

boards	<i>2019 Jeopardy! game boards</i>
--------	-----------------------------------

Description

The categories, clues, and answers in the 2019 season of Jeopardy.

Usage

boards

Format

A tibble with 13,261 rows and 8 variables:

game The J! Archive game ID.

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

n The order of clue chosen.

category Category title, often humorous or with instructions.

clue The clue read to the contestants.

answer The *correct* answer to a clue.

Source

<https://www.j-archive.com/>

entity_clean	<i>Repair the text from J! Archive HTML</i>
--------------	---

Description

Repair the text from J! Archive HTML

Usage

```
entity_clean(string)
```

Arguments

string Some text, usually from `rvest::html_text()`.

Value

Normalize character vectors.

episodes	<i>2019 Jeopardy! episodes</i>
----------	--------------------------------

Description

The episodes in the 2019 season of Jeopardy.

Usage

```
episodes
```

Format

A tibble with 218 rows and 3 variables:

game The non-sequential J! Archive game ID.

show The sequential show number of an episode.

date The air date of an episode.

Source

<https://www.j-archive.com/>

players *2019 Jeopardy! contestants*

Description

The contestants in the 2019 season of Jeopardy.

Usage

players

Format

A tibble with 672 rows and 6 variables:

game The J! Archive game ID.

show The sequential episode show number.

date The date each episodes initially aired.

Source

<https://www.j-archive.com/>

scores *2019 Jeopardy! game scores*

Description

The score history of the games in the 2019 season of Jeopardy.

Usage

scores

Format

A tibble with 13,261 rows and 6 variables:

round The round a clue is chosen.

n The order of clue chosen.

name First name of player responding.

score Change in score from this clue.

double Is the clue a daily double.

Source

<https://www.j-archive.com/>

synopses	<i>2019 Jeopardy! game synopses</i>
----------	-------------------------------------

Description

Synopses of the games in the 2019 season of Jeopardy.

Usage

synopses

Format

A tibble with 654 rows and 6 variables:

- name** The contestant's given name.
- coryat** Score if all wagering is disregarded.
- final** Final score after Double Jeopardy.
- right** Number of correct answers.
- wrong** Number of incorrect answers.

Source

<https://www.j-archive.com/>

whatr_airdate	<i>What is the air date?</i>
---------------	------------------------------

Description

This date identifies when an episode was first viewed on television.

Usage

whatr_airdate(game)

Arguments

- | | |
|------|--|
| game | One of four types of input, all passed to <code>whatr_html()</code> : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd". |
|------|--|

Format

A tibble with 1 row and 3 variables:

game The non-sequential J! Archive game ID.

show The sequential show number of an episode.

date The air date of an episode.

Value

A tidy tibble of clue text.

Examples

```
whatr_airdate(game = 6304)
```

whatr_answers	<i>What are the answers?</i>
---------------	------------------------------

Description

These must be given by the contestants in the form of a question as a response to the clues given.

Usage

```
whatr_answers(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Format

A tibble with (up to) 61 rows and 5 variables:

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

i The order of clue chosen.

answer The *correct* answer to a clue.

Value

A tidy tibble of answer text.

Examples

```
whatr_answers(game = 6304)
```

whatr_board	<i>What is the board?</i>
-------------	---------------------------

Description

This grid contains all the categories, clues, and answers in a game.

Usage

```
whatr_board(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Format

A tibble with (usually) 61 rows and 4 variables:

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

n The order of clue chosen.

category Category title, often humorous or with instructions.

clue The clue read to the contestants.

answer The *correct* answer to a clue.

Value

A tidy tibble of clue text.

Examples

```
whatr_board(game = 6304)
```

whatr_categories	<i>What are categories?</i>
------------------	-----------------------------

Description

These are collections of five clues related by subject matter.

Usage

```
whatr_categories(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none">1. A numeric game ID.2. Either a 'showgame' or 'showscores' HTML document.3. A show number character starting with "#".4. An air date like "yyyy-mm-dd".
------	--

Format

A tibble with (up to) 13 rows and 3 variables:

round The round a clue is chosen.

col The column position left-to-right.

category Category title, often humorous or with instructions.

Value

Tidy tibble of category titles.

Examples

```
whatr_categories(game = 6304)
```

`whatr_clues`*What are the clues?*

Description

Confusingly, *these* are posed to contestants in the form of an answer.

Usage

```
whatr_clues(game)
```

Arguments

`game` One of four types of input, all passed to `whatr_html()`:

1. A numeric game ID.
2. Either a 'showgame' or 'showscores' HTML document.
3. A show number character starting with "#".
4. An air date like "yyyy-mm-dd".

Format

A tibble with (usually) 61 rows and 4 variables:

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

i The order of clue chosen.

clue The clue read to the contestants.

Value

A tidy tibble of clue text.

Examples

```
whatr_clues(game = 6304)
```

whatr_data	<i>What is the board?</i>
------------	---------------------------

Description

This is the name for game mechanic displaying all categories and clues.

Usage

```
whatr_data(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Value

A list of tibbles.

Examples

```
whatr_data(game = 6304)
```

whatr_doubles	<i>What are daily doubles?</i>
---------------	--------------------------------

Description

These types of clues have no dollar value. Players wager some of their score before hearing the clue. In the first round, one such clue is present; in the second round, there are two.

Usage

```
whatr_doubles(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Format

a named list

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

i The order of clue chosen.

clue The clue read to the contestants.

score The amount won or lost on the wager.

Value

a list containing the question indices of the daily doubles in the first and second rounds

Examples

```
whatr_doubles(game = 6304)
```

whatr_html

Return the right J! Archive HTML document

Description

To avoid downloading and reading the J! Archive over and over, this function can be used to return an HTML document from one of four types of input:

1. Game ID (6304)
2. Show number ("#8006")
3. Air date ("2019-06-03")
4. HTML document itself

Usage

```
whatr_html(x, out = c("showgame", "showscores"))
```

Arguments

x Any kind of J! Archive game identifier.

out One of "showscores" or "showgame" used for `whatr_scores()` or `whatr_clues()`, etc. respectively. Either document can be returned from the input of the other.

Value

A J! Archive html_document.

Examples

```

whatr_html(x = 6304, out = "showscores")
whatr_html(x = "2019-06-03", out = "showgame")
whatr_html("#8006", "showgame") %>% whatr_clues()

```

 whatr_order

What is the order?

Description

Scrapes clue order left-to-right, top-to-bottom.

Usage

```
whatr_order(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Format

A tibble with (usually) 61 rows and 4 variables:

round The round a clue is chosen.

col The column position left-to-right.

row The row position top-to-bottom.

Value

A tidy tibble of player info.

A tidy tibble of clue orders.

whatr_players	<i>Who are the players?</i>
---------------	-----------------------------

Description

These individuals compete to score the most points and win the game.

Usage

```
whatr_players(game)
```

Arguments

game	One of four types of input, all passed to <code>whatr_html()</code> : <ol style="list-style-type: none">1. A numeric game ID.2. Either a 'showgame' or 'showscores' HTML document.3. A show number character starting with "#".4. An air date like "yyyy-mm-dd".
------	---

Format

A tibble with (usually) 3 rows and 4 variables:

first The contestant's given name.

last The contestant's surname name.

occupation A short description of what the contestant does.

from The city or institution from where the contestant comes.

Value

A tidy tibble of player info.

Examples

```
whatr_players(game = 6304)
```

whatr_plot	<i>What is a plot?</i>
------------	------------------------

Description

This type of graphic presents data in a visual manner.

Usage

```
whatr_plot(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none">1. A numeric game ID.2. Either a 'showgame' or 'showscores' HTML document.3. A show number character starting with "#".4. An air date like "yyyy-mm-dd".
------	--

Value

A ggplot object.

Examples

```
whatr_plot(game = 6304)
```

whatr_scores	<i>What are player scores?</i>
--------------	--------------------------------

Description

This data describes how players performed over the course of a game.

Usage

```
whatr_scores(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none">1. A numeric game ID.2. Either a 'showgame' or 'showscores' HTML document.3. A show number character starting with "#".4. An air date like "yyyy-mm-dd".
------	--

Format

A tibble with (up to) 61 rows and 5 variables:

round The round a clue is chosen.

i The order of clue chosen.

name First name of player responding.

score Change in score from this clue.

double Is the clue a daily double.

Value

Tidy tibble of clue scores.

Examples

```
whatr_scores(game = 6304)
```

whatr_synopsis	<i>What a synopsis?</i>
----------------	-------------------------

Description

This is an overview containing various summary statistics of a game.

Usage

```
whatr_synopsis(game)
```

Arguments

game	One of four types of input, all passed to whatr_html() : <ol style="list-style-type: none"> 1. A numeric game ID. 2. Either a 'showgame' or 'showscores' HTML document. 3. A show number character starting with "#". 4. An air date like "yyyy-mm-dd".
------	---

Format

A tibble with (usually) 3 row and 5 variables:

name The contestant's given name.

coryat Score if all wagering is disregarded.

final Final score after Double Jeopardy.

right Number of correct answers.

wrong Number of incorrect answers.

Value

A tidy tibble of clue text.

Examples

```
whatr_synopsis(game = 6304)
```


Index

*Topic **datasets**

- boards, 2
- episodes, 3
- players, 4
- scores, 4
- synopses, 5

boards, 2

entity_clean, 3
episodes, 3

players, 4

rvest::html_text(), 3

scores, 4
synopses, 5

whatr_airdate, 5
whatr_answers, 6
whatr_board, 7
whatr_categories, 8
whatr_clues, 9
whatr_clues(), 11
whatr_data, 10
whatr_doubles, 10
whatr_html, 11
whatr_html(), 5–10, 12–15
whatr_order, 12
whatr_players, 13
whatr_plot, 14
whatr_scores, 14
whatr_scores(), 11
whatr_synopsis, 15