Package: ei.Datasets (via r-universe)

October 14, 2024

Title Real Datasets for Assessing Ecological Inference Algorithms

Version 0.0.1-3

Description Provides more than 550 data sets of actual election results. Each of the data sets includes aggregate party and candidate outcomes at the voting unit (polling stations) level and two-way cross-tabulated results at the district level. These data sets can be used to assess ecological inference algorithms devised for estimating RxC (global) ecological contingency tables using exclusively aggregate results from voting units. Reference: Pavía (2022) <doi:10.1177/08944393211040808>.

License EPL | CC BY 4.0 | file LICENSE

Encoding UTF-8

Imports tibble

Depends R (>= 3.5.0)

LazyData true

RoxygenNote 7.1.1

NeedsCompilation no

Author Jose M. Pavía [aut, cre] (<https://orcid.org/0000-0002-0129-726X>)

Maintainer Jose M. Pavía <jose.m.pavia@uv.es>

Repository CRAN

Date/Publication 2022-04-27 11:10:02 UTC

Contents

`ei_NZ_2002` *Ecological inference data sets of the 2002 New Zealand General Election.*

Description

This tibble contains 69 data sets corresponding to the 2002 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2002)
```

Format

A tibble containing 69 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 69 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.

- `Votes_to_candidates`: A list of 69 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.
- `District_cross_votes`: A list of 69 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 69 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in `District_cross_percentages` and `District_cross_votes`. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in `Votes_to_parties` and `Votes_to_candidates`. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

Pavia, JM (2021). ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2005 ei_NZ_2008 ei_NZ_2011 ei_NZ_2014 ei_NZ_2017 ei_NZ_2020 ei_SCO_2007

ei_NZ_2005 *Ecological inference data sets of the 2005 New Zealand General Election.*

Description

This tibble contains 69 data sets corresponding to the 2005 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2005)
```

Format

A tibble containing 69 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 69 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.
- `Votes_to_candidates`: A list of 69 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.

- District_cross_votes: A list of 69 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding Votes_to_parties tibble, the rest of the variables (columns), ordered as in the corresponding Votes_to_candidates tibble, are labelled as the candidate voting options.
- District_cross_percentages: A list of 69 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding Votes_to_parties tibble, the rest of the variables (columns), ordered as in the corresponding Votes_to_candidates tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in District_cross_percentages and District_cross_votes. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in Votes_to_parties and Votes_to_candidates. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2008 ei_NZ_2011 ei_NZ_2014 ei_NZ_2017 ei_NZ_2020 ei_SCO_2007

---

ei_NZ_2008 *Ecological inference data sets of the 2008 New Zealand General Election.*

Description

This tibble contains 70 data sets corresponding to the 2008 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2008)
```

Format

A tibble containing 70 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 70 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.
- `Votes_to_candidates`: A list of 70 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.
- `District_cross_votes`: A list of 70 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 70 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in `District_cross_percentages` and `District_cross_votes`. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in `Votes_to_parties` and `Votes_to_candidates`. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2005 ei_NZ_2011 ei_NZ_2014 ei_NZ_2017 ei_NZ_2020 ei_SCO_2007

`ei_NZ_2011` *Ecological inference data sets of the 2011 New Zealand General Election.*

Description

This tibble contains 70 data sets corresponding to the 2011 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2011)
```

Format

A tibble containing 70 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 70 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.
- `Votes_to_candidates`: A list of 70 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.
- `District_cross_votes`: A list of 70 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 70 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in `District_cross_percentages` and `District_cross_votes`. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in `Votes_to_parties` and `Votes_to_candidates`. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2005 ei_NZ_2008 ei_NZ_2014 ei_NZ_2017 ei_NZ_2020 ei_SCO_2007

---

ei_NZ_2014 *Ecological inference data sets of the 2014 New Zealand General Election.*

Description

This tibble contains 71 data sets corresponding to the 2014 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2014)
```

Format

A tibble containing 71 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the Votes_to_parties, Votes_to_candidates, District_cross_votes and District_cross_percentag variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district i:

- Votes_to_parties: A list of 71 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, City and Address inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in Votes_to_parties and Votes_to_candidates coincide.
- Votes_to_candidates: A list of 71 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, City and Address inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in Votes_to_candidates and Votes_to_parties coincide.
- District_cross_votes: A list of 71 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding Votes_to_parties tibble, the rest of the variables (columns), ordered as in the corresponding Votes_to_candidates tibble, are labelled as the candidate voting options.
- District_cross_percentages: A list of 71 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding Votes_to_parties tibble, the rest of the variables (columns), ordered as in the corresponding Votes_to_candidates tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in District_cross_percentages and District_cross_votes. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in Votes_to_parties and Votes_to_candidates. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2005 ei_NZ_2008 ei_NZ_2011 ei_NZ_2017 ei_NZ_2020 ei_SCO_2007

---

ei_NZ_2017 *Ecological inference data sets of the 2017 New Zealand General Election.*

Description

This tibble contains 71 data sets corresponding to the 2017 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2017)
```

Format

A tibble containing 71 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 71 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.
- `Votes_to_candidates`: A list of 71 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.
- `District_cross_votes`: A list of 71 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 71 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in `District_cross_percentages` and `District_cross_votes`. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in `Votes_to_parties` and `Votes_to_candidates`. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 19 January 2019.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2005 ei_NZ_2008 ei_NZ_2011 ei_NZ_2014 ei_NZ_2020 ei_SCO_2007

---

ei_NZ_2020 *Ecological inference data sets of the 2020 New Zealand General Election.*

Description

This tibble contains 72 data sets corresponding to the 2020 New Zealand General Election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (electorate) level.

Usage

```
data(ei_NZ_2020)
```

Format

A tibble containing 72 observations and 6 variables:

`Number_of_district` Number assigned to the district/electorate by the New Zealand Electoral Commission.

`District` Name of the district/electorate.

`Votes_to_parties` A tibble for each electorate/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each electorate/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each electorate/district with the parties-candidates cross-distribution of votes in the entire electorate/district.

`District_cross_percentages` A tibble for each electorate/district, with the parties to candidates voter transition probabilities (in percentages) in the entire electorate/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 72 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.

- `Votes_to_candidates`: A list of 72 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `City` and `Address` inform, respectively, about the place in the district where the voting unit is located and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.
- `District_cross_votes`: A list of 72 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 72 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Note

The New Zealand Electoral Commission had no involvement in preparing these data sets. The raw data has been pre-processed in order to guarantee their straighforward usefulness in ecological inference procedures. Some small discrepancies exist among the figures in `District_cross_percentages` and `District_cross_votes`. The percentages are a direct translation of the published data, whereas the vote counts have been adjusted using integer linear programming to make them congruents with the figures in `Votes_to_parties` and `Votes_to_candidates`. More details in Pavia (2021). For the official results, visit https://www.electionresults.govt.nz.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from data available in https://www.electionresults.govt.nz, retrieved 23 January 2021.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

ei_NZ_2002 ei_NZ_2005 ei_NZ_2008 ei_NZ_2011 ei_NZ_2014 ei_NZ_2017 ei_SCO_2007

ei_SCO_2007 *Ecological inference data sets of the 2007 Scottish National Assembly.*

Description

This tibble contains 73 data sets corresponding to the 2007 Scottish National Assembly election. Each data set includes party and candidate vote results by voting unit as well as their associate cross-distributions (for votes and percentages) at the district (constituency) level.

Usage

```
data(ei_SCO_2007)
```

Format

A tibble containing 73 observations and 6 variables:

`Number_of_district` Number assigned to the district/constituency by the New Zealand Electoral Commission.

`District` Name of the district/constituency.

`Votes_to_parties` A tibble for each constituency/district with the party votes recorded in each voting unit of the district.

`Votes_to_candidates` A tibble for each constituency/district with the candidate votes recorded in each voting unit of the district.

`District_cross_votes` A tibble for each constituency/district with the parties-candidates cross-distribution of votes in the entire constituency/district.

`District_cross_percentages` A tibble for each constituency/district, with the parties to candidates voter transition probabilities (in percentages) in the entire constituency/district.

Details

Description of the `Votes_to_parties`, `Votes_to_candidates`, `District_cross_votes` and `District_cross_percentag` variables in more detail, where N(i), R(i) and C(i) denote, respectively, the number of voting units, party voting options and candidate voting options in district `i`:

- `Votes_to_parties`: A list of 73 tibbles/data.frames, with each data.frame containing N(i) observations and 2+R(i) variables. The two first variables, `Polling` and `Address` inform, respectively, about the code in the district assigned to the voting unit and the voting unit address. The rest of the columns correspond to the votes gained by the different party voting options competing in the district. The orders of the voting units in `Votes_to_parties` and `Votes_to_candidates` coincide.
- `Votes_to_candidates`: A list of 73 tibbles/data.frames, with each data.frame containing N(i) observations and 2+C(i) variables. The two first variables, `Polling` and `Address` inform, respectively, about the code in the district assigned to the voting unit and the voting unit address. The rest of the columns correspond to the votes gained by the different candidate voting options competing in the district. The orders of the voting units in `Votes_to_candidates` and `Votes_to_parties` coincide.

- `District_cross_votes`: A list of 73 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.
- `District_cross_percentages`: A list of 73 tibbles/data.frames, with each data.frame containing R(i) rows and 1+C(i) columns (variables). The first variable, which is labelled after the name of the district, contains the names of the parties in the same order than in corresponding `Votes_to_parties` tibble, the rest of the variables (columns), ordered as in the corresponding `Votes_to_candidates` tibble, are labelled as the candidate voting options.

Author(s)

Jose M. Pavia, <pavia@uv.es>

Source

Own elaboration from raw data downloading from the Scotland Electoral Office website in 2011 by Carolina Plescia. These data are not longer available in that site.

References

ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

`ei_NZ_2002` `ei_NZ_2005` `ei_NZ_2008` `ei_NZ_2011` `ei_NZ_2014` `ei_NZ_2017` `ei_NZ_2020`

`merge_small_options` *Merge small parties and/or candidates*

Description

Merge small parties and also small candidates by, respectively, aggregating them in the options 'Other parties votes' and 'Other candidates votes'.

Usage

```
merge_small_options(x, min.party, min.candidate)
```

Arguments

`x`	A tibble with the same components and structure as the tibbles in the `ei.Datasets` package. For instance, like the ei_NZ_2020 object.
`min.party`	A number between 0 and 100. Those parties which individually did not reach at least min.party% of the election-district vote are grouped in the option 'Other parties votes'.

`min.candidate` A number between 0 and 100. Those candidates which individually did not reach at least min.candididate% of the election-district vote are grouped in the option 'Other candidates votes'.

Value

A tibble similar to `x` with small parties and candidates merged on, respectively, 'Other parties votes' and 'Other candidates votes', with `min.party` and `min.candidate` used to determine when an electoral option is small.

Author(s)

Jose M. Pavia, <jose.m.pavia@uv.es>

References

Pavia, JM (2021). ei.Datasets: Real datasets for assessing ecological inference algorithms, Social Science Computer Review, forthcoming.

See Also

`ei_NZ_2002` `ei_NZ_2005` `ei_NZ_2008` `ei_NZ_2011` `ei_NZ_2014` `ei_NZ_2017` `ei_NZ_2020` `ei_SCO_2007`

Examples

```
collapsed.tibble <- merge_small_options(x = ei_NZ_2020, min.party = 3, min.candidate = 5)
```

Index