

Package ‘historydata’

May 8, 2026

Title Data Sets for Historians

Version 0.1

Description These sample data sets are intended for historians learning R. They include population, institutional, religious, military, and prosopographical data suitable for mapping, quantitative analysis, and network analysis.

Depends R (>= 3.0.2)

Suggests ggplot2 (>= 1.0.0)

License MIT + file LICENSE

LazyData true

URL <https://github.com/ropensci/historydata>

BugReports <https://github.com/ropensci/historydata/issues>

Author Lincoln Mullen [aut, cre]

Maintainer Lincoln Mullen <lincoln@lincolnmullen.com>

NeedsCompilation no

Repository CRAN

Date/Publication 2014-12-24 06:08:09

Contents

catholic_dioceses	2
early_colleges	3
historydata	4
judges	4
naval_promotions	5
paulist_missions	6
sarna	7
tudors	8
us_national_population	8
us_state_populations	9
Index	10

catholic_dioceses *Roman Catholic dioceses in the United States, Canada, and Mexico*

Description

Dates when Roman Catholic dioceses and archdioceses in the United States were founded or made metropolitan sees, with geocoded locations. The sources cited indicate that none of these sees have been discontinued.

Usage

```
catholic_dioceses
```

Format

A data frame with 425 observations of 6 variables.

Variables

- `diocese`: name, and thus location, of the diocese.
- `event, date`: the date when the diocese was erected, that is first founded, or made a metropolitan see. Encoded as a date object.
- `rite`: the rite overseen by the diocese. Regions with ordinary jurisdiction but not episcopal character are not included.
- `lat, long`: latitude and longitude coordinates for the headquarters city of the diocese.

References

This data is compiled from several sources:

Joseph Bernard Code, *Dictionary of the American Hierarchy (1789-1964)* (New York: Joseph F. Wagner, 1964), 425-26.

For the United States since 1963, Canada, and Mexico: <http://www.catholic-hierarchy.org/>

Examples

```
head(catholic_dioceses)
```

`early_colleges`*Early colleges in the United States*

Description

This dataset contains information about the founding of colleges established before 1848 in the United States of America.

Usage`early_colleges`**Format**

A data frame with 65 observations of 6 variables.

Variables

- `college`: The name of the college or university.
- `original_name`: The name under which the institution was founded, if different.
- `city, state`: The location of the institution.
- `established`: The year that the institution was founded.
- `sponsorship`: The sponsoring religious denomination, or secular if not founded by a denomination.

Author(s)

George Oberle, George Mason University

Daniel Walker Howe

References

This data was transcribed by George Oberle from the table "Some American Institutions of Higher Education Founded Before 1848, in Daniel Walker Howe, *What Hath God Wrought: The Transformation of America, 1815-1848* (New York: Oxford University Press, 2007), 460-461.

Examples

```
head(early_colleges)
if(require(ggplot2)) {
  ggplot(early_colleges, aes(x = established)) + geom_bar(binwidth = 5) +
  ggtitle("Founding Dates of Early American Colleges")
}
```

historydata

Datasets for historians

Description

This package provides sample datasets of interest to historians, analogous to the **datasets** package for R generally and the **histdata** package for datasets from the history of statistics.

References

Citations to the sources for datasets are provided in the documentation for each dataset.

judges

Federal judges in the United States of America

Description

This dataset contains information about the appointments and careers of all federal judges in United States history since 1789. It includes judges who "judges presidentially appointed during good behavior who have served since 1789 on the U.S. District Courts, the U.S. Courts of Appeals, the Supreme Court of the United States, the former U.S. Circuit Courts, and the federal judiciary's courts of special jurisdiction." Some of the unnecessary information from the source has been excluded.

Format

Two data frames, `judges_people` and `judges_appointments`.

Data frames

The data frame `judges_people` contains information about the judges, such as names and vital information. The data frame `judges_appointments` contains information about their appointments, such as the name of the court, nominating president, and the dates of service.

References

This data is taken from the [Biographical Directory of Federal Judges, 1789-present](#).

Examples

```
data(judges_people)
data(judges_appointments)
```

naval_promotions	<i>Promotions of U.S. Navy officers, 1798-1849</i>
------------------	--

Description

This dataset contains records of the careers of officers of the line in the United States Navy who became midshipmen between 1798 and 1849. It lists the dates at which officers were promoted to the ranks of midshipment, lieutenant, master commandant, and captain.

Usage

naval_promotions

Format

A data frame with 5,705 observations of 4 variables.

Variables

- `id`: A unique numeric identifier for each officer.
- `name`: Name of the officer.
- `generation`: A numeric value to represent a judgment about the generations within the officer corp of the U.S. Navy. First generation officers entered the service before 1801, as a rank higher than midshipman. Second generation officers entered the service before the Peace Establishment Act (or by the end of 1801), but as midshipman. Thus, they essentially became adults in the service, and they learned their craft from the first generation. Third generation officers entered the service as midshipmen after the Peace Establishment Act but before the end of the War of 1812. Those officers in this generation who became captain rose to that rank in the 1830s and '40s. Fourth generation officers entered the service after the War of 1812 had ended. These officers saw almost no wartime service, and many of the ones who achieved captain found themselves having to decide whether to serve in the Union or the Confederacy during the Civil War.
- `rank`: The ranks which officers obtained, including midshipman, lieutenant, master_commandant, and captain. Also included are the dates some officers left_service.
- `date`: The date as a character string on which an officer attained the associated rank.

Author(s)

Abby Mullen, Northeastern University

Edward W. Callahan

References

This data was compiled by Abby Mullen from Edward W. Callahan, *Officers of the Continental and U.S. Navy and Marine Corps, 1775-1900* (New York: L. R. Hamersly, 1901), made available via the Naval Historical Center, <http://www.ibiblio.org/hyperwar/NHC/Callahan/index.htm>. See Mullen's blog post about compiling the data, "[Database of Officers of the Line](#)," abbymullen.org, January 5, 2013.

Examples

```
head(naval_promotions)
```

paulist_missions	<i>Records of missions held by the Paulist Fathers, 1851–1893</i>
------------------	---

Description

This dataset contains transcriptions of records of missions held in by the Paulist Fathers, a Roman Catholic missionary order, in the nineteenth-century United States. This dataset contains only the most interesting data recorded in the full Paulist mission chronicles. The founding members of the Paulist Fathers began as members of the Redemptorist order. Their chronicles include both the missions held as Redemptorists and as Paulists. This transcription only includes missions up to the year 1893; the Paulist Chronicles contain more records than are transcribed here, and the Paulists continued holding missions for many more years than are recorded in the Chronicles.

Usage

```
paulist_missions
```

Format

A data frame with 841 observations of 11 variables.

Variables

- `mission_number`: The number assigned to the mission in the Paulist Mission Chronicles.
- `church`: The name of the church or cathedral at which the mission was held.
- `city, state`: The location of the mission.
- `start_date, end_date`: The start and end dates of the mission as a character string.
- `confessions`: The number of confessions heard by the Paulists at the mission, which is a rough proxy for the number of people who attended the mission.
- `converts`: The total number of converts made during the mission and people left under instruction for conversion after the mission was over.
- `order`: Whether the mission was held under the Redemptorist or Paulist order.
- `lat, long`: The latitude and longitude of the city where the mission was held.

Author(s)

Lincoln Mullen, George Mason University

References

The Paulist missions are recorded in *Chronicle of the Missions Given by the Congregation of Missionary Priests of St. Paul the Apostle*, six manuscript volumes, Office of Paulist History and Archives, North American Paulist Center, Washington, DC. Data transcribed by Lincoln Mullen.

Examples

```
head(paulist_missions)
```

sarna	<i>Population estimates of American Jews</i>
-------	--

Description

This dataset contains estimates of the population of American Jews.

Usage

```
sarna
```

Format

A data frame with 92 observations of 3 variables.

Variables

- year: date of estimate.
- estimate: the type of estimate. `population_low` and `population_high` are the lower and upper bounds on the population of American Jews; `percentage_high` and `percentage_low` are the lower and upper bounds on the percentage of Jews among the United States population.
- value: the value of the estimate.

References

This data is taken from the appendix in Jonathan D. Sarna, *American Judaism: A History* (New Haven: Yale University Press, 2004), 375-376.

Examples

```
head(sarna)
```

tudors

Tudor dynasty

Description

Spousal and parent/child relationships among selected members of the Tudor dynasty, suitable for network analysis.

Usage

```
tudors
```

Format

A data frame with 35 observations of 3 variables.

Variables

- person_1, person_2: The two people in the relationship.
- relationship: The type of relationship.

Examples

```
data(tudors)
```

us_national_population

Population of the United States, 1790-2010

Description

Population figures for the entire United States of America from the decennial census.

Usage

```
us_national_population
```

Format

A data frame with 23 observations of 2 variables.

Variables

- year: date of the census.
- population: population of the state or territory.

References

This dataset has been gathered by the [NHGIS](#). Minnesota Population Center, *National Historical Geographic Information System: Version 2.0* (Minneapolis: University of Minnesota, 2011).

Examples

```
head(us_national_population)
if(require(ggplot2)) {
  ggplot(us_national_population,
         aes(x = year, y = population)) +
  geom_line() +
  ggtitle("Population of the United States, 1790-2010")
}
```

us_state_populations *Populations of US states and territories, 1790-2010*

Description

Population figures for US states and territories from the decennial census.

Usage

```
us_state_populations
```

Format

A data frame with 983 observations of 4 variables.

Variables

- year: date of the census.
- state: name of the state or territory.
- population: population of the state or territory.
- GISJOIN: a unique identifier for joining NHGIS data to spatial data.

References

This dataset has been gathered by the [NHGIS](#). Minnesota Population Center, *National Historical Geographic Information System: Version 2.0* (Minneapolis: University of Minnesota, 2011).

Examples

```
head(us_state_populations)
```

Index

* datasets

- catholic_dioceses, 2
- early_colleges, 3
- naval_promotions, 5
- paulist_missions, 6
- sarna, 7
- tudors, 8
- us_national_population, 8
- us_state_populations, 9

catholic_dioceses, 2

early_colleges, 3

historydata, 4

historydata-package (historydata), 4

judges, 4

judges_appointments (judges), 4

judges_people (judges), 4

naval_promotions, 5

paulist_missions, 6

sarna, 7

tudors, 8

us_national_population, 8

us_state_populations, 9