

Package ‘Guerry’

October 24, 2023

Type Package

Title Maps, Data and Methods Related to Guerry (1833) “Moral Statistics of France”

Version 1.8.3

Date 2023-10-24

Maintainer Michael Friendly <friendly@yorku.ca>

Encoding UTF-8

Language en-US

Depends R (>= 3.5.0)

Imports sp

Suggests knitr, sf, spdep, ade4, adegraphics, adespatial, RColorBrewer, corrgram, car, effects, rmarkdown, here, ggplot2, ggpcp, ggrepel, heplots, patchwork, candisc, colorspace, scales, remotes, dplyr, tidy

Description

Maps of France in 1830, multivariate datasets from A.-M. Guerry and others, and statistical and graphic methods related to Guerry's “Moral Statistics of France”. The goal is to facilitate the exploration and development of statistical and graphic methods for multivariate data in a geospatial context of historical interest.

License GPL

URL <https://github.com/friendly/Guerry>

BugReports <https://github.com/friendly/Guerry/issues>

LazyLoad yes

LazyData yes

VignetteBuilder knitr

NeedsCompilation no

Author Michael Friendly [aut, cre] (<<https://orcid.org/0000-0002-3237-0941>>),
Stephane Dray [aut] (<<https://orcid.org/0000-0003-0153-1105>>),
Roger Bivand [ctb]

Repository CRAN

Date/Publication 2023-10-24 20:20:10 UTC

R topics documented:

Guerry-package	2
Angeville	4
gfrance	6
gfrance85	7
Guerry	9
propensity	12

Index **13**

Guerry-package	<i>Maps, Data and Methods Related to Guerry (1833) "Moral Statistics of France"</i>
----------------	---

Description

Andre-Michel Guerry (1833) was the first to systematically collect and analyze social data on such things as crime, literacy and suicide with the view to determining social laws and the relations among these variables. He provided the first essentially multivariate and georeferenced spatial data on socially important questions, e.g., Is the rate of crime related to education or literacy? How does this vary over the departments of France? Are the rates of crime or suicide within departments stable over time?

In an age well before the idea of correlation had been invented, Guerry used graphics and statistical maps to try to shed light on such questions. In a later work (Guerry, 1864), he explicitly tried to entertain larger questions, but with still-limited statistical tools: Can rates of various crimes be related to multiple causes or predictors? Are the rates and ascribable causes in France similar or different to those found in England?

The **Guerry** package comprises maps of France in 1830, multivariate data from A.-M. Guerry and others (Angeville, 1836), and statistical and graphic methods related to Guerry's *Moral Statistics of France*. The goal of providing these as an R package is to facilitate the exploration and development of statistical and graphic methods for multivariate data in a geo-spatial context.

Details

The DESCRIPTION file:

```
Package:      Guerry
Type:        Package
Title:       Maps, Data and Methods Related to Guerry (1833) "Moral Statistics of France"
Version:     1.8.3
Date:       2023-10-24
Authors@R:  c( person(given = "Michael", family = "Friendly", role=c("aut", "cre"), email="friendly@yorku.ca", comm
Maintainer:  Michael Friendly <friendly@yorku.ca>
```

Encoding: UTF-8
 Language: en-US
 Depends: R (>= 3.5.0)
 Imports: sp
 Suggests: knitr, sf, spdep, ade4, adeggraphics, adespatial, RColorBrewer, corrgram, car, effects, rmarkdown, here, ggp
 Description: Maps of France in 1830, multivariate datasets from A.-M. Guerry and others, and statistical and graphic m
 License: GPL
 URL: <https://github.com/friendly/Guerry>
 BugReports: <https://github.com/friendly/Guerry/issues>
 LazyLoad: yes
 LazyData: yes
 VignetteBuilder: knitr
 Author: Michael Friendly [aut, cre] (<<https://orcid.org/0000-0002-3237-0941>>), Stephane Dray [aut] (<<https://orcid.org/0000-0002-3237-0941>>)

Index of help topics:

Angeville	Data from d'Angeville (1836) on the population of France
Guerry	Data from A.-M. Guerry, "Essay on the Moral Statistics of France"
Guerry-package	Maps, Data and Methods Related to Guerry (1833) "Moral Statistics of France"
gfrance	Map of France in 1830 with the Guerry data
gfrance85	Map of France in 1830 with the Guerry data, excluding Corsica
propensity	Distribution of crimes against persons at different ages

Data from Guerry and others is contained in the data frame [Guerry](#). Because Corsica is often considered an outlier both spatially and statistically, the map of France circa 1830, together with the Guerry data is provided as `SpatialPolygonsDataFrames` in two forms: [gfrance](#) for all 86 departments, and [gfrance85](#), for the 85 departments excluding Corsica.

Author(s)

NA

Maintainer: Michael Friendly <friendly@yorku.ca>

References

d'Angeville, A. (1836). *Essai sur la Statistique de la Population francaise*, Paris: F. Darfour.

Dray, S. and Jombart, T. (2011). A Revisit Of Guerry's Data: Introducing Spatial Constraints In Multivariate Analysis. *The Annals of Applied Statistics*, 5(4).

Brunsdon, C. and Dykes, J. (2007). Geographically weighted visualization: interactive graphics for scale-varying exploratory analysis. Geographical Information Science Research Conference (GISRUK 2007). NUI Maynooth, Ireland, April, 2007. <https://www.maynoothuniversity.ie/national-centre-geocomputation-ncg>.

Friendly, M. (2007). A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. *Statistical Science*, 22, 368-399. <http://www.datavis.ca/papers/guerry-ST5241.pdf>

Friendly, M. (2007). Supplementary materials for Andre-Michel Guerry's Moral Statistics of France: Challenges for Multivariate Spatial Analysis, <http://www.datavis.ca/gallery/guerry/>.

Friendly, M. (2022). The life and works of Andre-Michel Guerry, revisited. *Sociological Spectrum*, 42, 233–259. doi:10.1080/02732173.2022.2078450

Guerry, A.-M. (1833). *Essai sur la statistique morale de la France* Paris: Crochard. English translation: Hugh P. Whitt and Victor W. Reinking, Lewiston, N.Y.: Edwin Mellen Press, 2002.

Guerry, A.-M. (1864). *Statistique morale de l'Angleterre compar?e avec la statistique morale de la France, d'apres les comptes de l'administration de la justice criminelle en Angleterre et en France, etc.* Paris: J.-B. Bailliere et fils.

Angeville

Data from d'Angeville (1836) on the population of France

Description

Adolph d'Angeville (1836) presented a comprehensive statistical summary of nearly every known measurable characteristic of the French population (by department) in his *Essai sur la Statistique de la Population francaise*. Using the graphic method of shaded (choropleth) maps invented by Baron Charles Dupin and applied to significant social questions by Guerry, Angeville's *Essai* became the first broad and general application of principles of graphic representation to national industrial and population data.

The collection of variables in the data frame Angeville is a small subset of over 120 columns presented in 8 tables and many graphic maps.

Usage

```
data(Angeville)
```

Format

A data frame with 86 observations on the following 16 variables.

dept a numeric vector

Department Department name: a factor with levels Ain Aisne ... Vosges Yonne

Mortality Mortality: Number of births to give 100 people at age 21 (T1:13)

Marriages Number of marriages per 1000 men aged 21 (T1:15)

Legit_births Annual no. of legitimate births (T2:17)

Illeg_births Annual no. of illegitimate births (T2:18)

Recruits Number of people registered for military recruitment from 1825-1833 (T3:32)

Conscripts Number of inhabitants per military conscript (T3:33)

Exemptions Number of military exemptions per 1000 all of physical causes (T3:47)
 Farmers Number of farmers during the census in 1831 (T4:65)
 Recruits_ignorant Average number of ignorant recruits per 1000 (T5:69)
 Schoolchildren Number of schoolchildren per 1000 inhabitants (T5:71)
 Windows_doors Number of windows & doors in houses per 100 inhabitants (T5:72). This is sometimes taken as an indicator of household wealth.
 Primary_schools "Number of primary schools (T5:74)
 Life_exp Life expectancy in years (T1:9a,9b)
 Pop1831 Population in 1831

Details

ID codes for dept were modified from those in Angeville's tables to match those used in [Guerry](#). Angeville's variables are recorded in a variety of different ways and some of these were calculated from other columns in his tables not included here. As well, the variable names and labels used here were often shortened from the more complete descriptions given by d'Angeville. The notation "(Tn:k)" indicates that the variable used here came from Table n, Column k.

Source

Angeville, A. d' (1836). *Essai sur la Statistique de la Population francaise*, Paris: F. Darfour.

The data was digitally scanned from Angeville's tables using OCR software, then extensively edited to correct obvious errors and finally subjected to some consistency checks using the column totals and ranked values he provided.

References

Whitt, H. P. (2007). Modernism, internal colonialism, and the direction of violence: suicide and crimes against persons in France, 1825-1830. Unpublished ms.

Examples

```
library(Guerry)
library(sp)
library(RColorBrewer)

data(Guerry)
data(gfrance)
data(Angeville)

gf <- gfrance      # the SpatialPolygonsDataFrame

# Add some Angeville variables, transform them to ranks
gf$Mortality      <- rank(Angeville$Mortality)
gf$Marriages      <- rank(Angeville$Marriages)
gf$Legit_births   <- rank(Angeville$Legit_births)
gf$Illeg_births   <- rank(Angeville$Illeg_births)
gf$Farmers        <- rank(Angeville$Farmers)
```

```
gf$Schoolchildren <- rank(Angeville$Schoolchildren)

# plot them on map of France
my.palette <- rev(brewer.pal(n = 9, name = "PuBu"))
splot(gf,
      c("Mortality", "Marriages", "Legit_births", "Illeg_births", "Farmers", "Schoolchildren"),
      names.attr = c("Mortality", "Marriages", "Legit_births",
                    "Illeg_births", "Farmers", "Schoolchildren"),
      layout=c(3,2),
      as.table=TRUE,
      col.regions = my.palette,
      cuts = 8, # col = "transparent",
      main="Angeville variables")
```

gfrance

Map of France in 1830 with the Guerry data

Description

gfrance is a `SpatialPolygonsDataFrame` object created with the `sp` package, containing the polygon boundaries of the map of France as it was in 1830, together with the [Guerry](#) data frame.

Usage

```
data(gfrance)
```

Format

The format is: Formal class 'SpatialPolygonsDataFrame' [package "sp"] with 5 slots:

- gfrance@data,
- gfrance@polygons,
- gfrance@plotOrder,
- gfrance@bbox,
- gfrance@proj4string.

See: [SpatialPolygonsDataFrame](#) for descriptions of some components.

The analysis variables, represented in gfrance@data are described in [Guerry](#).

Details

In the present version, the PROJ4 projection is not specified.

Source

Friendly, M. (2007). Supplementary materials for Andre-Michel Guerry's Moral Statistics of France: Challenges for Multivariate Spatial Analysis, <http://www.datavis.ca/gallery/guerry/>.

References

Friendly, M. (2007). A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. *Statistical Science*, 22, 368-399.

See Also

[Guerry](#) for description of the analysis variables [Angeville](#) for other analysis variables

Examples

```
library(sp)
data(gfrance)
names(gfrance) ## list @data variables
plot(gfrance) ## just show the map outline

# Show basic choropleth plots of some of the variables
spplot(gfrance, "Crime_pers")

# use something like Guerry's palette, where dark = Worse
my.palette <- rev(RColorBrewer::brewer.pal(n = 9, name = "PuBu"))
spplot(gfrance, "Crime_pers", col.regions = my.palette, cuts = 8)

spplot(gfrance, "Crime_prop")

# Note that spplot assumes all variables are on the same scale for comparative plots
# transform variables to ranks (as Guerry did)

## Not run:
local({
  gfrance$Crime_pers <- rank(gfrance$Crime_pers)
  gfrance$Crime_prop <- rank(gfrance$Crime_prop)
  gfrance$Literacy <- rank(gfrance$Literacy)
  gfrance$Donations <- rank(gfrance$Donations)
  gfrance$Infants <- rank(gfrance$Infants)
  gfrance$Suicides <- rank(gfrance$Suicides)

  spplot(gfrance, c("Crime_pers", "Crime_prop", "Literacy", "Donations", "Infants", "Suicides"),
    layout=c(3,2), as.table=TRUE, main="Guerry's main moral variables")
})

## End(Not run)
```

gfrance85

Map of France in 1830 with the Guerry data, excluding Corsica

Description

gfrance85 is a SpatialPolygonsDataFrame object created with the sp package, containing the polygon boundaries of the map of France as it was in 1830, together with the [Guerry](#) data frame. This version excludes Corsica, which is an outlier both in the map and in many analyses.

Usage

```
data(gfrance85)
```

Format

The format is: Formal class 'SpatialPolygonsDataFrame' [package "sp"] with 5 slots: gfrance85@data, gfrance85@polygons, gfrance85@plotOrder, gfrance85@bbox, gfrance85@proj4string. See: [SpatialPolygonsDataFrame](#) for descriptions of some components.

The analysis variables are described in [Guerry](#).

Details

In the present version, the PROJ4 projection is not specified.

Source

Friendly, M. (2007). Supplementary materials for Andr -Michel Guerry's Moral Statistics of France: Challenges for Multivariate Spatial Analysis, <http://datavis.ca/gallery/guerry/>.

References

Dray, S. and Jombart, T. (2009). A Revisit Of Guerry's Data: Introducing Spatial Constraints In Multivariate Analysis. Unpublished manuscript.

Friendly, M. (2007). A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. *Statistical Science*, 22, 368-399.

Examples

```
data(gfrance85)
require(sp)
require(scales)
plot(gfrance85) # plot the empty outline map

# extract some useful components
df <- data.frame(gfrance85)[,7:12] # main moral variables
xy <- coordinates(gfrance85) # department centroids
dep.names <- data.frame(gfrance85)[,6]
region.names <- data.frame(gfrance85)[,5]
col.region <- colors()[c(149,254,468,552,26)] |>
  scales::alpha(alpha = 0.2)

# plot the map showing regions by color with department labels
op <- par(mar=rep(0.1,4))
plot(gfrance85,col=col.region[region.names])
text(xy, labels=dep.names, cex=0.6)
par(op)
```

 Guerry

Data from A.-M. Guerry, "Essay on the Moral Statistics of France"

Description

Andre-Michel Guerry (1833) was the first to systematically collect and analyze social data on such things as crime, literacy and suicide with the view to determining social laws and the relations among these variables.

The Guerry data frame comprises a collection of 'moral variables' on the 86 departments of France around 1830. A few additional variables have been added from other sources.

Usage

```
data(Guerry)
```

Format

A data frame with 86 observations (the departments of France) on the following 23 variables.

dept Department ID: Standard numbers for the departments, except for Corsica (200)

Region Region of France ('N'='North', 'S'='South', 'E'='East', 'W'='West', 'C'='Central'). Corsica is coded as NA

Department Department name: Departments are named according to usage in 1830, but without accents. A factor with levels Ain Aisne Allier ... Vosges Yonne

Crime_pers Population per Crime against persons. Source: A2 (Comptes general, 1825-1830)

Crime_prop Population per Crime against property. Source: A2 (Compte general, 1825-1830)

Literacy Percent Read & Write: Percent of military conscripts who can read and write. Source: A2

Donations Donations to the poor. Source: A2 (Bulletin des lois)

Infants Population per illegitimate birth. Source: A2 (Bureau des Longitudes, 1817-1821)

Suicides Population per suicide. Source: A2 (Compte general, 1827-1830)

MainCity Size of principal city ('1:Sm', '2:Med', '3:Lg'), used as a surrogate for population density. Large refers to the top 10, small to the bottom 10; all the rest are classed Medium. Source: A1. An ordered factor with levels 1:Sm < 2:Med < 3:Lg

Wealth Per capita tax on personal property. A ranked index based on taxes on personal and movable property per inhabitant. Source: A1

Commerce Commerce and Industry, measured by the rank of the number of patents / population. Source: A1

Clergy Distribution of clergy, measured by the rank of the number of Catholic priests in active service / population. Source: A1 (Almanach officiel du clergy, 1829)

Crime_parents Crimes against parents, measured by the rank of the ratio of crimes against parents to all crimes– Average for the years 1825-1830. Source: A1 (Compte general)

- Infanticide Infanticides per capita. A ranked ratio of number of infanticides to population— Average for the years 1825-1830. Source: A1 (Compte general)
- Donation_clergy Donations to the clergy. A ranked ratio of the number of bequests and donations inter vivos to population— Average for the years 1815-1824. Source: A1 (Bull. des lois, ordonn. d'autorisation)
- Lottery Per capita wager on Royal Lottery. Ranked ratio of the proceeds bet on the royal lottery to population— Average for the years 1822-1826. Source: A1 (Compte rendus par le ministre des finances)
- Desertion Military desertion, ratio of the number of young soldiers accused of desertion to the force of the military contingent, minus the deficit produced by the insufficiency of available billets— Average of the years 1825-1827. Source: A1 (Compte du ministere du guerre, 1829 etat V)
- Instruction Instruction. Ranks recorded from Guerry's map of Instruction. Note: this is inversely related to Literacy (as defined here)
- Prostitutes Prostitutes in Paris. Number of prostitutes registered in Paris from 1816 to 1834, classified by the department of their birth Source: Parent-Duchatelet (1836), *De la prostitution en Paris*
- Distance Distance to Paris (km). Distance of each department centroid to the centroid of the Seine (Paris) Source: calculated from department centroids
- Area Area (1000 km²). Source: Angeville (1836)
- Pop1831 1831 population. Population in 1831, taken from Angeville (1836), *Essai sur la Statistique de la Population francaise*, in 1000s

Details

Note that most of the variables (e.g., Crime_pers) are scaled so that 'more is better' morally.

Values for the quantitative variables displayed on Guerry's maps were taken from Table A2 in the English translation of Guerry (1833) by Whitt and Reinking. Values for the ranked variables were taken from Table A1, with some corrections applied. The maximum is indicated by rank 1, and the minimum by rank 86.

Source

Angeville, A. (1836). *Essai sur la Statistique de la Population francaise* Paris: F. Doufour.

Guerry, A.-M. (1833). *Essai sur la statistique morale de la France* Paris: Crochard. English translation: Hugh P. Whitt and Victor W. Reinking, Lewiston, N.Y. : Edwin Mellen Press, 2002.

Parent-Duchatelet, A. (1836). *De la prostitution dans la ville de Paris*, 3rd ed, 1857, p. 32, 36

References

Dray, S., & Jombart, T. (2011). Revisiting Guerry's data: Introducing spatial constraints in multivariate analysis. *Annals of Applied Statistics*, **5**, 2278-2299

Brunsdon, C. and Dykes, J. (2007). Geographically weighted visualization: Interactive graphics for scale-varying exploratory analysis. *Geographical Information Science Research Conference (GISRUK 07)*, NUI Maynooth, Ireland, April, 2007.

Friendly, M. (2007). A.-M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. *Statistical Science*, 22, 368-399.

Friendly, M. (2007). Data from A.-M. Guerry, Essay on the Moral Statistics of France (1833), <https://www.datavis.ca/gallery/guerry/guerrydat.html>.

See Also

[Angeville](#) for other analysis variables

Examples

```
library(car)
data(Guerry)

# Is there a relation between crime and literacy?

# Plot personal crime rate vs. literacy, using data ellipses.
# Identify the departments that stand out
set.seed(12315)
with(Guerry,{
  dataEllipse(Literacy, Crime_pers,
  levels = 0.68,
    ylim = c(0,40000), xlim = c(0, 80),
    ylab="Pop. per crime against persons",
    xlab="Percent who can read & write",
    pch = 16,
    grid = FALSE,
    id = list(method="mahal", n = 8, labels=Department, location="avoid", cex=1.2),
    center.pch = 3, center.cex=5,
    cex.lab=1.5)
  # add a 95% ellipse
  dataEllipse(Literacy, Crime_pers,
  levels = 0.95, add=TRUE,
    ylim = c(0,40000), xlim = c(0, 80),
    lwd=2, lty="longdash",
    col="gray",
    center.pch = FALSE
  )

  # add the LS line and a loess smooth.
  abline( lm(Crime_pers ~ Literacy), lwd=2)
  lines(loess.smooth(Literacy, Crime_pers), col="red", lwd=3)
}
)

# A corrgram to show the relations among the main moral variables
# Re-arrange variables by PCA ordering.

library(corrgram)
corrgram(Guerry[,4:9], upper=panel.ellipse, order=TRUE)
```

propensity

Distribution of crimes against persons at different ages

Description

This dataset comes from Plate IV, "Influence de l'age" of Guerry(1833), transcribed in Whitt & Reinking's (2002) translation as Table 9A, pp. 38-43. It gives the rank ordering of crimes against persons in seven age groups, in long form.

Usage

```
data("propensity")
```

Format

A data frame with 124 observations on the following 4 variables.

age a character vector, with 7 age groups, <21, 21-30, 30-40 ... 60-70, 70-

rank a numeric vector, rank of the crime within each age group

crime a character vector, label of the crime

share a numeric vector, share (frequency) of the crime in a population of 1000

Details

For each age group (both males and females), the 17 most frequent crimes are listed in rank order, followed by an 'Other crime' category.

Source

H. P. Whitt and V. W. Reinking (2002). A Translation of Andr'e-Michel Guerry's *Essay on the Moral Statistics of France*, Lewiston, N.Y.: Edwin Mellen Press, 2002.

References

Guerry, A.-M. (1833). *Essai sur la statistique morale de la France* Paris: Crochard.

Examples

```
data(propensity)
## maybe str(propensity) ; plot(propensity) ...
```

Index

* datasets

Angeville, [4](#)
gfrance, [6](#)
gfrance85, [7](#)
Guerry, [9](#)
propensity, [12](#)

* data

Guerry-package, [2](#)

* spatial

gfrance, [6](#)
gfrance85, [7](#)
Guerry-package, [2](#)

Angeville, [4](#), [7](#), [11](#)

gfrance, [3](#), [6](#)
gfrance85, [3](#), [7](#)
Guerry, [3](#), [5–8](#), [9](#)
Guerry-package, [2](#)

propensity, [12](#)

SpatialPolygonsDataFrame, [6](#), [8](#)