Package 'YourCast'

January 2, 2012

Version 1.1-11

Date 2010-04-03

Title YourCast

Author Federico Girosi <girosi@rand.org> Gary King <king@harvard.edu>
Jon Bischof <jbischof@fas.harvard.edu>

Maintainer Gary King <king@harvard.edu>

Depends R (>= 1.9.0), foreign, lattice

Description YourCast makes time series cross-sectional forecasts with multiple cross-sections based on your assumptions. It allows a variety of smoothing assumptions based on similarities among the levels, trends, or interactions in the expected value of the dependent variable rather than the coefficients. YourCast implements ideas in the book Federico Girosi and Gary King. DEMOGRAPHIC FORECASTING. Princeton University Press, 2008; see http://gking.harvard.edu/files/smooth/

License file LICENSE

URL http://gking.harvard.edu/yourcast

Repository CRAN

Date/Publication 2010-04-04 15:12:23

R topics documented:

adjacency *Proximity data for yourprep example*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg== Murray Research Archive

Usage

```
data(adjacency)
```

Format

A fixed width '.txt' file.

Source

WHO

References

http://gking.harvard.edu/yourcast

See Also

example(yourprep)

chp.11.1.RData *Respiratory infections, Belize*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.1 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.1)

---

chp.11.10 *Breast cancer, Chile, Cuba, Belgium, and the Netherlands*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.10)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.10 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.10)

chp.11.11 *Transportation Accidents, Argentina and Chile*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.11)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.11 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.11)

chp.11.12 *Transportation Accidents, Argentina*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.12)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.12 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.12)

`chp.11.13` *Transportation Accidents, Argentina, Chile, Canada, Colombia, Costa Rica, Cuba, and USA*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.13)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.13 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.13)

chp.11.2 *Respiratory infections, Bulgaria*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.2 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.2)

chp.11.3 *Respiratory infections, Bulgaria*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.3)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.3 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.3)

---

chp.11.4 *Lung Cancer, Peru*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.4)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.4 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.4)

chp.11.5 *Lung Cancer, Trinidad and Tobago and Ukraine*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.5)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.5 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.5)

chp.11.7.1 *Lung Cancer, 51 countries*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.7.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to calculate empirical priors for the 'map' model used in figure 11.7 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.7)

---

chp.11.7.2 *Lung Cancer, Peru*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.7.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.7 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.7)

chp.11.8.1 *Lung Cancer, 51 countries*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.8.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to calculate empirical priors for the 'map' model used in figure 11.8 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.8)

chp.11.8.2 *Lung Cancer, Ukraine*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.8.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.8 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.8)

---

chp.11.8.3 *Lung Cancer, Trinidad and Tobago*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.8.3)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.8 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.8)

chp.11.9.1 *Breast Cancer, 43 countries*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.9.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to calculate empirical priors for the 'map' model used in figure 11.9 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.9)

chp.11.9.2 *Breast Cancer, Croatia*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.11.9.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 11.9 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.11.9)

---

chp.2.6.1 *All causes of death, New Zealand and Hungary*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.2.6.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 2.6 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.2.6)

chp.2.6.2 *Transportation accidents (males), Portugal*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.2.6.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 2.6 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.2.6)

chp.2.7.1 *Suicide deaths (male), USA*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.2.7.1)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 2.7 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.2.7)

---

chp.2.7.2 *Female digestive disease, Hungary*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.2.7.2)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 2.7 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.2.7)

chp.2.7.3 *Female cervix cancer, United Kingdom*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(chp.2.7.3)
```

Format

A YourCast 'dataobj' input object: a list of a list of cross sectional dataframes, dataframe of geographic names, and string indicationg CSID coding rule. Used to create figure 2.7 in *Demographic Forecasting*.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

See Also

demo(chp.2.7)

cntry.codes *Country codes and names for yourprep example*

Description

World Health Organization data 1920-2000. Federico Girosi and Gary King, 2006, "Cause of Death Data", hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg== Murray Research Archive

Usage

```
data(cntry.codes)
```

Format

A fixed width '.txt' file.

Details

191 countries included.

Source

WHO

References

http://gking.harvard.edu/yourcast

See Also

example(yourprep)

---

csid204500 *yourprep example file for Belize, age 0*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204500)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 0. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204505 *yourprep example file for Belize, age 5*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204505)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 5. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204510 *yourprep example file for Belize, age 10*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204510)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 10. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

`csid204515` *yourprep example file for Belize, age 15*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204515)
```

Format

A fixed width '`.txt`' file.

Details

Example cross section dataframe for Belize, age 15. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204520 *yourprep example file for Belize, age 20*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204520)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 20. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204525 *yourprep example file for Belize, age 25*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204525)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 25. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

`csid204530` *yourprep example file for Belize, age 30*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204530)
```

Format

A fixed width '`.txt`' file.

Details

Example cross section dataframe for Belize, age 30. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204535 *yourprep example file for Belize, age 35*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204535)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 35. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204540 *yourprep example file for Belize, age 40*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204540)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 40. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

`csid204545` *yourprep example file for Belize, age 45*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204545)
```

Format

A fixed width '`.txt`' file.

Details

Example cross section dataframe for Belize, age 45. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204550 *yourprep example file for Belize, age 50*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204550)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 50. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204555 *yourprep example file for Belize, age 55*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204555)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 55. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

`csid204560` *yourprep example file for Belize, age 60*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204560)
```

Format

A fixed width '`.txt`' file.

Details

Example cross section dataframe for Belize, age 60. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204565 *yourprep example file for Belize, age 65*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204565)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 65. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204570 *yourprep example file for Belize, age 70*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204570)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 70. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

`csid204575` *yourprep example file for Belize, age 75*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204575)
```

Format

A fixed width '`.txt`' file.

Details

Example cross section dataframe for Belize, age 75. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

csid204580 *yourprep example file for Belize, age 80*

Description

World Health Organization data years 1920-2000. Federico Girosi and Gary King, 2006, *Cause of Death Data*, hdl:1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg==Murray Research Archive

Usage

```
data(csid204580)
```

Format

A fixed width '.txt' file.

Details

Example cross section dataframe for Belize, age 80. To be used in with Example section in user's manual.

Source

World Health Organization

References

http://gking.harvard.edu/yourcast

histograph *Histograms for model ebayes*

Description

Draws histograms for priors calculated by model ebayes.

Usage

```
histograph(d1.a, d1.t, dt.da, SD, depvar=" ",
           model="ebayes", graphics.file=NA)
```

Arguments

d1.a	Numeric vector. First derivative respect to age.
d1.t	Numeric vector. First derivative respect to time.
dt.da	Numeric vector. Second derivative respect to age and time.
SD	Numeric vector. Standard deviation.
depvar	String with the name of the dependent variable. Default: " "
model	String with the name of the model. Default: "ebayes".
graphics.file	String or NA. If string the name of a file to be appended to a directory path where the graphics will be saved, if NA it is displayed in the screen and not saved. Default: NA.

Value

Histograms of the vectors d1.a, d1.t, dt.da, and SD; see demos chp.11.7, chp.11.8, or chp.11.9.

Author(s)

Federico Girosi <girosi@rand.org>

References

http://gking.harvard.edu/yourcast

---

plot.yourcast *Plot generation tool for YourCast*

Description

Creates graphics from yourcast output for each geographical unit and prints to the device window or a .pdf file in the specified directory

Usage

```
## S3 method for class 'yourcast'
plot(x, dpath=getwd(), dvlabel=NULL,
                          family="agetime",
                          args.matplot=list(),args.wireframe=list(),
                          time.insamp.obs=TRUE, time.insamp.predict=TRUE,
                          age.insamp.predict=TRUE,
                          threedim.insamp.predict=TRUE,
                          age.xlab=NULL, age.ylab=NULL,
                          time.xlab=NULL, time.ylab=NULL,
                          threedim.xlab=NULL, threedim.ylab=NULL,
                          threedim.zlab=NULL,
                          screen=list(z=-40, x=-60, y=0),
                          age.incl=NULL,
                          print="device",filename=NULL,...)
```

Arguments

x	yourcast output object
dpath	String. Directory where '.pdf' outputs are saved. Defaults to working directory if not specified. Will be ignored if print="device". Default: getwd()
dvlabel	String. Description of dependent variable that will be used as the main title for the plots. Default: NULL
family	String. Specifies type of plot generated by the function. "time" creates a time series plot where each age cohort is plotted separately on the device. "age" creates a plot of forecasts on age where data from each each year is plotted separately. "agetime" creates a side-by-side presentation of the "age" and "time" plots. Finally, "threedim" creates a three-dimensional plot of the information in the '"age"' and '"time"' plots. Default: "agetime"
args.matplot	List. A list of arguments (must be labeled) to be passed to matplot. For example, if wanted to turn change line weight, could add args.matplot=list(lwd=2). Does not apply to plots of family threedim. In most cases users will have to launch age and time plots separately for this feature to work well. Note that some arguments to matplot such as xlab and main should be made as arguments to plot.yourcast; they will be overwritten if found in args.matplot. For the moment, main is specified via the dvlabel argument.
args.wireframe	List. A list of arguments (must be labeled) to be passed to wireframe. This only applies to plots of the threedim family. Note that some arguments to wireframe such as zlab and screen should be made as arguments to plot.yourcast; they will be overwritten if found in args.wireframe.
time.insamp.obs	Logical. For "time" and 'time' plots in "agetime" plots, specifies whether observed values should be plotted within the range of years with observations for the dependent variable. Default: TRUE
time.insamp.predict	Logical. For "time" and 'time' plots in "agetime" plots, specifies whether predicted values should be plotted within the range of years with observations for the dependent variable. For "time" plots, the default is to print both predicted and observed values; this option merely removes predicted values from this range if set to FALSE. Default: TRUE
age.insamp.predict	Logical. For "age" plots and the 'age' plots of "agetime" plots, specifies whether predicted values should be plotted within the range of years with observations for the dependent variable. If set to FALSE, predicted values are replaced by observed values in this range. Default: TRUE
threedim.insamp.predict	Logical. For "threedim" plots, specifies whether predicted values should be plotted within the range of years with observations for the dependent variable. If set to FALSE, predicted values are replaced by observed values in this range. Default: TRUE
age.xlab	String. The label for the 'x' axis of "age" plots and the 'age' plot of "agetime" plots. Will be ignored if 'age' plot not created. Default: "Age"

age.ylab	String. The label for the 'y' axis of "age" plots and the 'age' plot of "agetime" plots. Will be ignored if 'age' plot not created. Default: "Forecasts"
time.xlab	String. The label for the 'x' axis of "time" plots and the 'time' plot of "agetime" plots. Will be ignored if 'time' plot not created. Default: "Time"
time.ylab	String. The label for the 'y' axis of "time" plots and the 'time' plot of "agetime" plots. Will be ignored if 'time' plot not created. Default: "Data and Forecasts"
threedim.xlab	String. The label for the 'x' axis of "threedim" plots. Will be ignored if 'three-dim' plot not created. Default: "Year"
threedim.ylab	String. The label for the 'y' axis of "threedim" plots. Will be ignored if 'three-dim' plot not created. Default: "Age"
threedim.zlab	String. The label for the 'z' axis of "threedim" plots. Will be ignored if 'three-dim' plot not created. Default: "Forecasts"
screen	List. List with three elements 'x', 'y', and 'z' that rotate the viewing angle for three dimensional plots. Argument ignored for all other plot types. Default: list(z=-40, x=-60, y=0)
age.incl	Vector. If changed from NULL, subset of age groups to be included in the time series plot. Ages should be specified in the same way they are in the index.code argument to yourcast(). Default: NULL
print	String. Specifies whether graphical output should be displayed sequentially on a device window ("device") or saved directly to a '.pdf' file in the dpath ("pdf"). Default: "device"
filename	Vector of strings. If changed from NULL, provides filenames for pdfs created from the graphical output for each geographical area. The order of the labels should be the same as the order of the areas in dataobj$data. Filenames will be recycled if the length of the vector is less than the number of areas.
...	Arguments to be passed to par() for the purpose of setting graphical parameters for the plotting functions. See help(par) for more details.

Details

Prints sequentially to the device or saves '.pdf' files with the requested plot for each geographic unit in the sample. If requested, '.pdf' files will be saved in a specified directory or the working directory. Three-dimensional plots are created with the wireframe function from the lattice library. For space considerations, axes are labeled with the numerical versions of the 'age' and 'time' vectors regardless of whether A.names and T.names are supplied to yourcast.

Plots are titled with the dvlabel and the G.names dataframe if it was supplied to yourcast in the dataobj. For example if dvlabel="Respiratory Infections" and the geographic identifier for that region is matched with "Belize", the plot will be titled "Respiratory Infections, Belize". One or both labels will be utilized by the function if available.

Axis labels can be changed with the appropriate xlab, ylab, and zlab arguments.

It is important to note that plot.yourcast will only work if all cross sections within the same geographic unit are of the same dimensions. If, for example, a cross section for one age group has fewer yearly observations than another from the same group, these missing years must be filled in with NA, even if they occur in the beginning of the sample period. This does not hold across geographic units, however.

Finally, `plot.yourcast` handles `"agetime"` plots differently than the other families by opening a new device window for each new plot. This is done so that the size of the device can be controlled to keep the side-by-side plots from appearing distorted when launched. This convenience makes it impossible to place multiple plots on the same device, however. Users seeking to create 1x2, 2x2, 3x2, etc., plot layouts for the purposes of comparison are advised to use the separate `"age"` and `"time"` plot families and print each plot indivudally to the device.

Below is some example code for a 2x2 plot layout with, in effect, two 'agetime' plots:

```
par(mfrow=c(2,2))
plot(y.out1,family="age")
plot(y.out1,family="time")
plot(y.out2,family="age")
plot(y.out2,family="time")
```

Value

Device windows with requested plots or '`.pdf`' files saved in the `dpath`.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

`yourcast` function and documentation (`help(yourcast)`)

`print.summary.yourcast` *Print method for yourcast summary output*

Description

summary.yourcast class print function

Usage

```
print.summary.yourcast(x, ...)
```

Arguments

`x`	`summary.yourcast` output object to be printed
`...`	Arguments to be passed to or from other methods.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

yourcast,summary.yourcast

`print.yourcast` *Print method for yourcast output*

Description

yourcast class print function

Usage

```
## S3 method for class 'yourcast'
print(x, ...)
```

Arguments

x	yourcast output object to be printed
...	Arguments to be passed to or from other methods.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

yourcast

print.yourprep *Print method for yourprep output*

Description

yourprep class print function

Usage

```
print.yourprep(x, ...)
```

Arguments

x	yourprep output object to be printed
...	Arguments to be passed to or from other methods.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

yourprep

summary.yourcast *Summarize yourcast output*

Description

yourcast class summary function

Usage

```
## S3 method for class 'yourcast'
summary(object, ...)
```

Arguments

object	yourcast output object to be summarized
...	Arguments to be passed to or from other methods.

Value

A list of summary objects of class 'summary.yourcast':

`sample.frame`	Four element vector containing, in order, the start and end time periods to be used for the observed data and the start and end time periods to be forecast
`params`	Vector. Smoothing parameters used in model.
`model`	String. Estimation technique used.
`formula`	Language. Formula used in analysis.
`numcs`	Integer. Number of cross sections used in analysis.
`cntry.codes`	Vector. Unique geographical codes of cross sections used in analysis.
`cntry.names`	Vector. If `G.names` dataframe submitted to `yourcast`, the names of all unique geographical codes of cross sections used in analysis.
`coef`	List. List of matrices with estimated coefficents for each cross section. For Lee-Carter models, only one set of coefficents calculated. Standard errors are not calculated for these coefficients.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

`yourcast`,`print.summary.yourcast`

---

user.prompt *Scan user input*

Description

Gets user response to continue with the demos simulation or to quit.

Usage

```
user.prompt()
```

Details

Allows users to terminate demos early.

Value

None.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

See demos; for example: demo(chp.11.4)

---

yourcast *Time-series cross-sectional Forecasting*

Description

Runs a set of regression models to forecast time-series cross-sectional data by either considering independent regressions in each cross-sectional unit or by using a variety of techniques to smooth across units.

Usage

```
yourcast(formula=NULL, dataobj=NULL,sample.frame=c(1950,2000,2001,2030),
                      standardize=TRUE, elim.collinear=FALSE,
                      tol=0.9999, solve.tol = 1.e-10,svdtol=10^(-10),
                      userfile=NULL, savetmp = T, model.frame=FALSE,
                      debug = F,  rerun= "yourcast.savetmp",
            ### specific to models
                      model="OLS",zero.mean=FALSE,
            #### smooth over ages
                      Ha.sigma = 0.3,
                      Ha.sigma.sd= 0.1, Ha.deriv=c(0,0,1),
                      Ha.age.weight=0, Ha.time.weight=0,
            #### smooth over time
                      Ht.sigma= 0.3,
                      Ht.sigma.sd=0.1,  Ht.deriv=c(0,0,1),
                      Ht.age.weight=0, Ht.time.weight=0,
            #### smooth over age-time
                      Hat.sigma=0.2,
                      Hat.sigma.sd=0.1,Hat.a.deriv=c(0,1),Hat.t.deriv=c(0,1),
                      Hat.age.weight=0,Hat.time.weight=0,
            #### smooth over cntry-time
                      Hct.sigma=0.3, Hct.sigma.sd =0.1,
                      Hct.t.deriv=1, Hct.time.weight = 0,
                      LI.sigma.mean=0.2,LI.sigma.sd = 0.1, nsample= 500,
                      low.pow=T, verbose=TRUE)
```

Arguments

formula	A standard R formula of the form $y \sim x_1 + x_2$, except that an explanatory variable is included for a particular cross-section only if it is both listed in the formula and available in that cross-section's data set (see dataobj). Explanatory variables in the formula but not available for a cross-section (or in a cross-sectional dataset but not in the formula) are excluded. (For mortality forecasting, the specification looks like log(deaths/population) $\sim x_1 + x_2$, with deaths and population stored as separate variables in each dataframe.) (May be set to NULL if savetmp was set to TRUE on the last run, in which case the value of formula will come from the saved file.)
dataobj	A object of class 'yourcast' or equivalent. See help(yourprep) for more details. The dataobj may be supplied in one of four ways. Most commonly, the argument will specify (1) an object (in working memory) or (2) a string with the name of a file in the working directory. However, if (3) dataobj is a string referring to a directory on disk, then each element of the list above should be stored in a file in that directory, with element 'data' consisting of a subdirectory containing separate ASCII data files. (If this option is chosen, a complete data object, called 'dataobj.Rdata', will be stored in the directory named, and it will be loaded automatically if yourcast is run again with this chosen option.) (4) The last option is for dataobj to be set to NULL, after which the function will look for a 'yourcast.savetmp' file in the working directory from a previous run of the function where the argument savetmp was set to TRUE. The function yourprep is available to help construct the dataobj in the proper format from individual cross section files in the working directory or the workspace. This function also performs a number of diagnostics to ensure that the data is entered properly and can be read by yourcast. See help(yourprep) for more information
sample.frame	Vector. A four element vector containing, in order, the start and end time periods to be used for the observed data and the start and end time periods to be forecast. Years identified here that are not available for a cross-section are ignored. Default: c(1950,2000,2001,2030).
standardize	Boolean. Should the covariates in each cross-sectional unit be standardized (to zero mean and standard deviation of 1)? Standardization is performed for both the in- and out-of-sample periods. Default: TRUE.
elim.collinear	Boolean. Whether collinearity among covariates should be tested and those that are collinear shoul be eliminated. Default: FALSE.
tol	Double scalar. Tolerance to find collinearities among covariates. Default: 0.9999.
solve.tol	A real number smaller than one that is used in the argument of the R-function solve to invert matrices (see description for tol). Default: 1^{-10}.
svdtol	A scalar; the tolerance used in inverting a matrix by SVD. Default: 10^{-10}.
userfile	A string with the name of a file that contains your values for some or all of yourcast's arguments. This file contains R code that changes default values of arguments. E.g., the file might contain:

```
index.code <- 30
data <- "WHOmortalityData"
```

If an option is specified in `userfile`, it takes precidence over command line options, so it is normally best to specify each option in either the `userfile` or the command line but not both. Default: `NULL`

`savetmp` If `TRUE`, `yourcast` saves a file in the default directory (called '`yourcast.savetmp`') with preliminary calculations. If the value of `formula` or `dataobj` is missing when `yourcast` is called, `yourcast` will get their values from this file, if it exists. This saves a minute or so of computing time for large data sets and is useful for multiple runs on the same data with different formulas specified or different prior values. If `FALSE`, no file is saved. (The structure of '`yourcast.savetmp`' is for the convenience of `yourcast` and is not intended to be read by the user or saved for more than one run.) Default: `TRUE`.

`model.frame` If `TRUE`, include entire input dataobj in the output object. Default: `FALSE`.

`debug` Boolean. It puts the environment that contains parameters and arguments of the simulation in the user workspace. Default `FALSE`.

`rerun` String. The name of the file that is saved in the default directory with preliminary calculations; see `savetmp`. Default: `yourcast.savetmp`

`model` A string indicating the forecasting method, including: Bayes maximum a posteriori (`map`), Bayes with Gibbs sampling (`bayes`), Ordinary Least Squares (`ols`), Poisson (`poisson`), and Lee-Carter (`LC`). Default: `ols`. (We usually recommend `map`.)

`yourcast` also includes a procedure to help users set the sigma parameters below automatically for the case of model=`map`, and smoothing over age, time, or age and time, but for only one country. You may do this by running a preprocessing instance of `yourcast` first by setting this parameter to `ebayes` and using either the data to be analyzed or a larger data set which is likely to have similar or related parameter values. When `ebayes` is chosen, the `yourcast` output object will contain only the parameter values to feed into the next run of `yourcast`.

`zero.mean` A boolean or named vector with a value of $\bar{\mu}$ for each age group. If `TRUE`, the prior has zero mean. If `FALSE`, the prior has nonzero mean centered around the observed mean age profile (i.e., the average of Y over time and levels of the geographic index for each age group). Default: `FALSE`.

`Ha.sigma` This can be set in one of three ways: (1) a scalar which sets σ_a, the prior standard deviation of $E(Y)$, indicating how much to smooth $E(Y)$ over age groups (which may vary over geographic areas and time periods, and with the standard deviations averaged over age groups). A larger standard deviation represents more prior uncertainty, which allows the data to play a greater role. (2) `NA` to not smooth in this way. (3) To have `yourcast` search for a good value based on a target value of the derivative of $E(Y)$ with respect to age, set to a vector of elements containing the start and end of a range in sigma in which to look (such as 0.05 and 1.5), the number of values to look at within this range (such as 5), and the target value of the derivative of $E(Y)$ with respect to age (such as 0.05). The vector may also include a fifth element, which is the target value of the total standard deviation of $E(Y)$ over all dimensions of the prior (such as 0.1). (You

may choose to run yourcast with model=ebayes on a related data set to find an approximate target value of the derivative and standard deviation automatically.) Default: 0.30.

Ha.sigma.sd A scalar; the standard deviation of parameter Ha.sigma (for Gibbs sampling only). Default: 0.1.

Ha.deriv A numeric vector, each element of which is n,the degree of a (discrete) derivative of the smoothness functional with respect to the age group. Element k of this vector refers to the $(k-1)$th derivative, where 0 excludes the derviative, 1 includes it, and values in between include the derivative but weight it down proportionally. The first element of the vector corresponds to the weight on the derivative with respect to age of order 0 (the identity operator), the second to the weight on the derivative of order 1 (the 1st derivative), etc. For example, c(0, 1, 1) corresponds to a mixed functional that penalizes the first and second derivatives equally. The higher the order of derivative, the more local smoothness over age groups; and lowest specified derivative controls the form of prior indifference. Default: c(0, 0, 1), which usually works well.

Ha.age.weight A scalar or a numeric vector with weights that determine how much smoothing occurs for different age groups. If set to 0 or NA, age groups are weighted equally; if set to a nonzero scalar, the weight for age group a is set proportional to $a^H a.age.weight$; if a vector of length A, the ath element is the weight of age group a. Default: 0.

Ha.time.weight A scalar or a numeric vector with weights that determine how much smoothing occurs for different time periods when smoothing over age groups. If 0 or NA, time periods are weighted equally; if set to a nonzero scalar value, the weight for time period t in smoothing age groups is proportional to $t^H a.time.weight$; if the argument is a vector of length T, the tth element is the weight of time period t. Default: 0.

Ht.sigma This can be set in one of three ways: (1) a scalar which sets σ_t, the prior standard deviation of $E(Y)$, indicating how much to smooth $E(Y)$ over time periods (which may vary over geographic areas and age groups, and with the standard deviations averaged over time periods). A larger standard deviation represents more prior uncertainty, which allows the data to play a greater role. (2) NA to not smooth in this way. (3) To have yourcast search for a good value based on a target value of the derivative of $E(Y)$ with respect to time, set to a vector of elements containing the start and end of a range in sigma in which to look (such as 0.05 and 1.5), the number of values to look at within this range (such as 5), and the target value of the derivative of $E(Y)$ with respect to time (such as 0.05). The vector may also include a fifth element, which is the target value of the total standard deviation of $E(Y)$ over all dimensions of the prior (such as 0.1). (You may choose to run yourcast with model=ebayes on a related data set to find an approximate target value of the derivative and standard deviation automatically.) Default: 0.30.

Ht.sigma.sd A scalar; the standard deviation of parameter Ht.sigma (for Gibbs sampling only). Default: 0.1.

Ht.deriv A numeric vector, each element of which is n, the degree of a (discrete) derivative of the smoothness functional with respect to time. Element k of this vector refers to the $(k-1)$th derivative, where 0 excludes the derviative, 1 includes it,

and values in between include the derivative but weight it down proportionally. The first element of the vector corresponds to the weight on the derivative with respect to time of order 0 (the identity operator), the second to the weight on the derivative of order 1 (the 1st derivative), etc. For example, `c(0, 1, 1)` corresponds to a mixed functional that penalizes the first and second derivatives equally. The higher the order of derivative, the more local smoothness over time; and lowest specified derivative controls the form of prior indifference. Default: `c(0, 0, 1)`, which usually works well.

`Ht.age.weight` A scalar or a numeric vector with weights that determine how much smoothing occurs for different age groups when smoothing over time. If set to `0` or `NA`, age groups are weighted equally in smoothing over time; if set to a nonzero scalar, the weight for age group a is set proportional to $a^H t.age.weight$; if a vector of length A, the ath element is the weight of age group a. Default: 0.

`Ht.time.weight` A scalar or a numeric vector with weights that determine how much smoothing occurs for different time periods when smoothing over time. If `0` or `NA`, time periods are weighted equally; if set to a nonzero scalar value, the weight for time period t in smoothing time periods is proportional to $t^H t.time.weight$; if the argument is a vector of length T, the tth element is the weight of time period t. Default: 0.

`Hat.sigma` This can be set in one of three ways: (1) a scalar which sets σ_{at}, the prior standard deviation of $E(Y)$, indicating how much to smooth the time trend in $E(Y)$ over age groups. A larger standard deviation represents more prior uncertainty, which allows the data to play a greater role. (2) NA to not smooth in this way. (3) To have `yourcast` search for a good value based on a target value of the derivative of $E(Y)$ with respect to age and time, set to a vector of elements containing the start and end of a range in sigma in which to look (such as 0.05 and 1.5), the number of values to look at within this range (such as 5), and the target value of the derivative of $E(Y)$ with respect to age and time (such as 0.05). The vector may also include a fifth element, which is the target value of the total standard deviation of $E(Y)$ over all dimensions of the prior (such as 0.1). (You may choose to run `yourcast` with `model=ebayes` on a related data set to find an approximate target value of the derivative and standard deviation automatically.) Default: `0.2`.

`Hat.sigma.sd` A scalar; the standard deviation of parameter `Hat.sigma` (for Gibbs sampling only). Default: `0.1`.

`Hat.a.deriv` A numeric vector, each element of which is n, the degree of a (discrete) derivative of the smoothness functional of time trends with respect to age groups. Element k of this vector refers to the $(k-1)$th derivative of the time trend v with respect to age, where 0 excludes the derviative, 1 includes it, and values in between include the derivative but weight it down proportionally. The first element of the vector corresponds to the weight on the derivative of the time trend with respect to age of order 0 (the identity operator), the second to the weight on the derivative of order 1 (the 1st derivative), etc. For example, `c(0, 1, 1)` corresponds to a mixed functional that penalizes the first and second derivatives equally. The higher the order of derivative, the more local smoothness over time; and lowest specified derivative controls the form of prior indifference. Default: `c(0, 0, 1)`, which usually works well.

`Hat.t.deriv` A numeric vector, each element of which is n, the degree of a (discrete) derivative of the smoothness functional of age derivative with respect to time. Element k of this vector refers to the $(k-1)$th derivative of the age derivative with respect to time, where 0 excludes the derviative, 1 includes it, and values in between include the derivative but weight it down proportionally. The first element of the vector corresponds to the weight on the age derivative with respect to time of order 0 (the identity operator), the second to the weight on the derivative of order 1 (the 1st derivative), etc. For example, `c(0, 1, 1)` corresponds to a mixed functional that penalizes the first and second derivatives equally. The higher the order of derivative, the more local smoothness over time; and lowest specified derivative controls the form of prior indifference. Default: `c(0, 0, 1)`, which usually works well.

`Hat.age.weight` A scalar or a numeric vector with weights that determines how much smoothing occurs for different age groups when smoothing over age and time. If set to `0` or `NA`, age groups are weighted equally in smoothing over time; if set to a nonzero scalar, the weight for age group a is set proportional to $a^H t.age.weight$; if a vector of length A, the ath element is the weight of age group a. Default: `0`.

`Hat.time.weight`

A scalar or a numeric vector with weights that determine how much smoothing occurs for different time periods when smoothing over age and time. If `0` or `NA`, time periods are weighted equally; if set to a nonzero scalar value, the weight for time period t in smoothing time periods is proportional to $t^H t.time.weight$; if the argument is a vector of length T, the tth element is the weight of time period t. Default: `0`.

`Hct.sigma` A scalar which sets σ_t, the prior standard deviation of $E(Y)$, which indicates how to smooth $E(Y)$ over geographic areas, or NA to not smooth in this way. The parameter $\sigma_c t$ is the expected prior standard deviation of $E(Y)$ for a geographic area (varying over time periods and age groups, and with the standard deviations averaged over geographic areas). (A larger standard deviation represents more prior uncertainty, which allows the data to play a greater role.) Default: `0.3`.

`Hct.sigma.sd` A scalar; the standard deviation of parameter Ht.sigma (for Gibbs sampling only). Default: `0.1`.

`Hct.t.deriv` A numeric vector; controls whether smoothing the level or the time trend of $E(Y)$ over geographic areas (both cannot presently be done simultaneously). To smooth the level of $E(Y)$ over geographic areas, set to 1, the identity. To smooth the time trend, set this (as in `Hat.t.deriv`) to the weight of the partial derivative taken with respect to time in the standard smoothness functional for the prior. The use of the first or higher order partial derivatives are supported. Default: `1`.

`Hct.time.weight`

A scalar or a numeric vector with weights that determine how much smoothing occurs for different time periods when smoothing over geographic areas. If `0` or `NA`, time periods are weighted equally; if set to a nonzero scalar value, the weight for time period t in smoothing over areas is proportional to $t^H ct.time.weight$; if the argument is a vector of length T, the tth element is the weight of time period t. Default: `0`.

LI.sigma.mean	A scalar; used in the likelihood and in the calculation of the priors in conjunction with Ha.sigma.sd, Hat.sigma.sd, Ht.sigma.sd, and Hct.sigma.sd. Default: 0.2.
LI.sigma.sd	A scalar; the standard deviation of LI.sigma.mean used in the calculation of the priors. Default: 0.1.
nsample	A scalar; represents the number of iterations in the Gibbs algorithm bayes. Default: 500.
low.pow	Boolean. Whether to include lower-power of explanatory variables in the simulation as derived from formula. For example $y \sim x^4$, if low.pow = TRUE, then x, x^2, x^3, x^4 will be included. Default: TRUE.
verbose	Boolean. Suppress verbose output. Default: FALSE

Value

Returns a list of class 'yourcast' containing the following components:

call	The full call, including all command line options when yourcast was called.
userfile	The full userfile if it was specified.
yhat	A list with the same cross-sectional elements as the input data, but with two columns: 'y' for the observed dependent variable and 'yhat' for the predicted values. These include both in-sample and out-of-sample values, as distinguished by the values of sample.frame.
coeff	A list with the same cross-sectional elements as the input data, elements of which are the estimated coefficients if calculated by the chosen model.
sigma	A list with the same cross-sectional elements as the input data, elements of which are the estimated standard error of the estimate of the regression (the standard deviation of the dependent variable given the explanatory variables).
aux	List. A list of summary information about the yourcast analysis used by plot.yourcast
params	Vector. Smoothing parameters used in model.

Author(s)

Federico Girosi <girosi@rand.org>; Elena Villalon <evillalon@iq.harvard.edu>; Gary King <king@harvard.edu>

References

http://gking.harvard.edu/yourcast

yourprep *Data object creation wizard for YourCast*

Description

Builds the data object for yourcast function from files in working directory or other specified directory and checks for errors

Usage

```
yourprep(dpath=getwd(),tag="csid",index.code="ggggaa",
                      datalist=NULL,G.names=NULL,A.names=NULL,
                      T.names=NULL,adjacency=NULL,year.var=FALSE,
                      sample.frame=NULL,summary=FALSE,verbose=FALSE,

                      #lagging utility
                      lag=NULL,formula=NULL,vars.nolag=NULL)
```

Arguments

dpath	String. Name of the directory where data files are stored. If NULL then defaults to working directory. Default: NULL
tag	String. Group of characters placed before CSID code in filenames to indicate which files in dpath function should load. The tag can also be used to differentiate between different groups to be considered in separate analysis; for example, 'm' for male deaths and 'f' for female deaths. Default: "csid"
index.code	String indicating how the CSID index variable is coded in the input data. Between 0 and 4 of the following two characters are used in this order: g for the geographic index (such as country) and a for a grouped continuous variable like an age group. For example, ggggaa would have the function interpret '245045' by using '2450' as the country code and '45' as the age group. Default: "ggggaa"
datalist	A list of cross section dataframes already loaded into the workspace to be added to the dataobj. Names of list elements should be the numerical CSID code for each cross section, and dataframes should be formated identically to files loaded from an external directory (see Details)
A.names, G.names, T.names	String. Filename of optional two-column data files that list all valid numerical codes (in the first column) and corresponding alphanumeric names (optionally in the second column) for the indices corresponding to geographic areas in G.names, age groups in A.names, and time periods in T.names. Function will search dpath for file with specified name; please include column labels. The optional alphanumeric identifiers are most commonly only used for geographic areas since numerical values for age groups and time periods are usually meaingful on their own. However, if other grouped continuous variable used in place of ages, for example, specifying these labels will be important for output to be meaningful. NOTE: Auxiliary files will loaded automatically by yourprep() if

they are saved in the dpath and labeled with the tag specified by the user. See 'Details' section for more infromation. Default: NULL

adjacency — Data file with codes to construct the symmetric matrix (geographic region by geographic region) of proximity scores for geographic smoothing used by the 'map' and 'bayes' methods. The larger the relative score, the more proximate that pair of countries is in the prior; a zero element means the two geographic areas are unrelated (the diagonal is ignored). Each row of the proximity file has three columns, consisting of geographic codes for two countries and a score indicating the proximity or similarity of the two geographic regions; please include column labels. For convenience, geographic regions that are unrelated (and would have zero entries in the symmetric matrix) may be omitted from proximity. In addition, proximity may include rows corresponding to geographic regions not included in the present analysis. Default: NULL

year.var — Boolean. Should be TRUE if year coded as separate variable rather than as rowname for cross section data files. Function will look for year variable to use as rownames and then drop it from the dataframe. Change will only be made to dataframe if it does not already have rownames or if exisiting rownames are merely a '1...N' index of row numbers, so it is possible to apply correction even if some cross sections do not have a year variable and already have the correct rownames. Default: FALSE

sample.frame — Optional four element vector containing, in order, the start and end time periods to be used for the observed data and the start and end time periods to be forecast. All cross sections do not have to begin at starting date, but must contain all years after the first observed value. Variables to be forecasted should be coded as NA in the out-of-sample period. Note that this makes it easy to reserve a range of values of the dependent variable for out-of-sample forecasting evaluation; our summary and plot functions in yourcast will make these comparisons automatically if the out-of-sample data are included. yourprep() uses this information only to verify that cross sections are correctly constructed, but it should also be included if one wants to use the lag utility. Default: NULL

summary — Boolean. If TRUE, means for available observations on each variable are displayed for the cross sections read by yourprep(). Default: FALSE

verbose — Boolean. If TRUE, function prints name of each cross section or auxiliary file as it is read into the dataobj. Default: FALSE

lag — Number of years covariate data needs to be lagged from current position is cross section files. See 'Details' for more information. Default: NULL

formula — Formula. The formula that one will use in the subsequent run of yourcast(). This helps the lagging utility distinguish between the response variable (which will not be shifted between cross sections) and the covariates of interest that should be lagged and included in the final cross sections of the dataobj. If the covariate 'index' is included in the formula, the lagging utility will include a variable in the cross sections that starts from 1 and counts the number of time periods since the start of the cross section. If a lag is requested, the formula argument must be specified. Default: NULL

vars.nolag — Vector of strings. Vector of variables to be included in the dataobj but not lagged. These variables do not need to be included in the formula, and if found there will not ignored when the other covariates are lagged.

Details

Creates dataobj input for yourcast from files in working directory or other specified directory. Checks that all cross sections in data list titled properly and if all years up to last predicted year included in the dataframes (if sample.frame argument specified). Please note, however, that all cross sections from the same geographic area must have the same observation and prediction years in the dataframe (even if NA) for the graphing software plot.yourcast to work.

The cross section files must be named according to the CSID identifiers for country code and age group, preceeded by the specified tag (default: "csid") so that yourprep() can identify the file from other files in the dpath. For example, for the USA (country code 2450) time series of 45 year old individuals, the file name should be 'csid245045.txt' if the tag is left as the default. Files must have an extension so that the program can recognize how the data is coded. Currently, fixed width text files ('*.txt'), comma-separated values ('*.csv'), and Stata v.5-10 ('*.dta') files are supported, and multiple file types may be used in the same run of the program. '*.Rdata' objects can be included with the datalist option after they are loaded to a list in the workspace. yourprep() includes diagnostics to ensure that objects are properly named and not included accidentally, but users should examine the specified dpath before running yourprep() to minimize errors.

Each cross section file should be labeled columns of time-series data for the dependent variable(s) (e.g., disease, pop) and the covariates that will be used in the forecast. The rownames for the dataframe should be the observation year (if the year is coded as a separate variable, set year.var=TRUE). The files must contain the full time series that will be specified in the sample.frame argument in yourcast after the first observed year. For instance, if sample.frame=c(1950,2000,2001,2030), then files would have observations that start between 1950 and 2000 and include all other years (even if the entries are NA) up to the last year of prediction, i.e., 2030.

Optional auxiliary files such as G.names should be named according to the filename specified in the respective arguments. If specified, these files must have extensions and be coded in one of the three supported file types. However, these files will be automatically loaded by yourprep() if they are saved in the dpath and labeled with the tag specified by the user. The default names for these files must be used (e.g., 'G.names' and 'adjacency'). For example, if the tag is left as the default and there is a file in the dpath labeled 'csid.G.names.txt', yourprep() will load this automatically and save the input as the G.names element of the 'dataobj' list. yourprep() arguments such as G.names take precedence over 'TAG.*' files in thedpath.

yourprep() also includes a lagging utility (activated once one specifies a lag length with the 'lag' argument). This utility is useful for when the data in each cross section is, for example, the response and covariates for 50 year olds in each year but the desired content for each cross section is the response for 50 year olds and the covariates for 25 year olds 25 years prior to each year (implying a lag of 25 years). In order to have yourprep() perform this lagging automatically, include cross sections for each age group with data starting the same number of years before the first observation year as the requested lag period. Thus if lag=25 and the first observation year is 1950, then the cross sections should all start at 1925. Age groups younger than the length of the lag will not retain covariate data (except perhaps an 'index' variable) in the output object. The covariates lagged are the predictor variables specified in the formula argument.

If data for a cohort 25 years (in this case) younger is not available for some cohort over age 25, yourprep() will look for the closest cohort available and issue a warning message.

Value

dataobj A list with several components:

data A list with the cross-sectional data matrices as elements.

proximity A symmetric matrix (geographic region by geographic region) of proximity scores for geographic smoothing used by the 'map' and 'bayes' methods. The larger each element of the matrix, the more proximate that pair of countries is in the prior; a zero element means the two geographic areas are unrelated (the diagonal is ignored). Each element of the symmetric matrix is created from one row of the proximity input to `yourprep()` (which is two country codes and a proximity score).

G.names, A.names, T.names Optional two-column dataframes that list all valid numerical codes (in the first column, labeled codes) and corresponding alphanumeric names (optionally in the second column, labeled name) for the indices corresponding to the geographic areas in `G.names`, age groups in `A.names`, and time periods in `T.names`.

index.code A string indicating how the index variable is coded in the input data.

Author(s)

Jon Bischof <jbischof@fas.harvard.edu>

References

http://gking.harvard.edu/yourcast

See Also

`yourcast` function and documentation (`help(yourcast)`)

Examples

```
## Not run:
# Working directory automatically set to directory with cross
# section and auxiliary files to begin. Files for this example
# in 'data' folder of YourCast library.

#Old working directory to be restored later
oldwd <- getwd()
# Now setting wd to 'data' folder in YourCast library
setwd(system.file("data",package="YourCast"))

# Simple run of the function, using option that turns year variable
# into label in each cs. Use sample.frame argument for all diagnostics
# to work

dta <- yourprep(G.names="cntry.codes.txt",adjacency="adjacency.txt",
year.var=TRUE,verbose=TRUE,sample.frame=c(1950,2000,2001,2030))


# With summary output (means of variables in each cross section)


dta <- yourprep(G.names="cntry.codes.txt",adjacency="adjacency.txt",
```

```
year.var=TRUE,summary=TRUE)


# Function can also add datafiles already loaded into R as objects in
# the workspace with "datalist" option if put into a list and properly
# labeled. All diagnostics still performed
# 'csid204545', etc., are dataframes in workspace

# Labels changed to nonsense ones so as not to confuse with other files

data(csid204545)
data(csid204550)
data(csid204555)

datalist <- list("123456"=csid204545,"234567"=csid204550,
"345678"=csid204555)

# Verbose option turned on and datalist argument added

dta <- yourprep(G.names="cntry.codes.txt",adjacency="adjacency.txt",
year.var=TRUE,verbose=TRUE,datalist=datalist)

# Setting working directory back
setwd(oldwd)
rm(oldwd)

## End(Not run)
```

Index