Galton's data on the heights of parents and their children, by child

Description

This data set lists the individual observations for 934 children in 205 families on which Galton (1886) based his cross-tabulation.

In addition to the question of the relation between heights of parents and their offspring, for which this data is mainly famous, Galton had another purpose which the data in this form allows to address: Does marriage selection indicate a relationship between the heights of husbands and wives, a topic he called *assortative mating*? Keen [p. 297-298](2010) provides a brief discussion of this topic.

Usage

```
data(GaltonFamilies)
```

Format

A data frame with 934 observations on the following 8 variables.

`family`

>   family ID, a factor with levels `001-204`

`father`

>   height of father

`mother`

>   height of mother

`midparentHeight`

>   mid-parent height, calculated as `(father + 1.08*mother)/2`

`children`

>   number of children in this family

`childNum`

>   number of this child within family. Children are listed in decreasing order of height for boys followed by girls

`gender`

>   child gender, a factor with levels `female male`

`childHeight`

height of child

Details

Galton's notebook lists 963 children in 205 families ranging from 1-15 adult children children. Of these, 29 had non-numeric heights recorded and are not included here.

Families are largely listed in descending order of fathers and mothers height.

Source

Galton's notebook, http://www.medicine.mcgill.ca/epidemiology/hanley/galton/notebook/, transcribed by Beverley Shipley in 2001.

References

Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature *Journal of the Anthropological Institute*, 15, 246-263

Hanley, J. A. (2004). "Transmuting" Women into Men: Galton's Family Data on Human Stature. *The American Statistician*, 58, 237-243. See: http://www.medicine.mcgill.ca/epidemiology/hanley/galton/ for source materials.

Keen, K. J. (2010). *Graphics for Statistics and Data Analysis with R*, Boca Raton: CRC Press, https://www.unbc.ca/keen/graphics-for-statistics-and-data-analysis-with-r.

See Also

Galton, PearsonLee

Examples

```
data(GaltonFamilies)
str(GaltonFamilies)

## reproduce Fig 2 in Hanley (2004)
library(car)
scatterplot(childHeight ~ midparentHeight | gender, data=GaltonFamilies,
    ellipse=TRUE, levels=0.68, legend.coords=list(x=64, y=78))

# multiply daughters' heights by 1.08
GF1 <- within(GaltonFamilies,
            {childHeight <- ifelse (gender=="female", 1.08*childHeight, childHeight)} )
scatterplot(childHeight ~ midparentHeight | gender, data=GF1,
    ellipse=TRUE, levels=0.68, legend.coords=list(x=64, y=78))

# add 5.2 to daughters' heights
GF2 <- within(GaltonFamilies,
            {childHeight <- ifelse (gender=="female", childHeight+5.2, childHeight)} )
scatterplot(childHeight ~ midparentHeight | gender, data=GF2,
    ellipse=TRUE, levels=0.68, legend.coords=list(x=64, y=78))

#########################################
```

```
# relationship between heights of parents
###########################################

Parents <- subset(GaltonFamilies, !duplicated(GaltonFamilies$family))

with(Parents, {
  sunflowerplot(mother, father, rotate=TRUE, pch=16,
     xlab="Mother height", ylab="Father height")
        dataEllipse(mother, father, add=TRUE, plot.points=FALSE,
     center.pch=NULL, levels=0.68)
        abline(lm(father ~ mother), col="red", lwd=2)
        }
        )
```

[Package *HistData* version 0.8-7 <u>Index</u>]