Package 'AmesHousing'

June 23, 2020

Version 0.0.4

Title The Ames Iowa Housing Data

URL https://github.com/topepo/AmesHousing

BugReports https://github.com/topepo/AmesHousing/issues

Description Raw and processed versions of the data from De Cock (2011) <http://ww2.amstat.org/publications/jse> are included in the package.

License GPL-2

Encoding UTF-8

LazyData true

ByteCompile true

Depends R (>= 2.10)

Imports dplyr, magrittr

RoxygenNote 7.1.0.9000

Suggests covr

NeedsCompilation no

Author Max Kuhn [aut, cre],
      Dmytro Perepolkin [ctb],
      RStudio [cph]

Maintainer Max Kuhn <max@rstudio.com>

Repository CRAN

Date/Publication 2020-06-23 20:10:03 UTC

R topics documented:

---

ames_raw   *Raw Ames Housing Data*

Description

Summon the data described by De Cock (2011) where 82 fields were recored for 2,930 properties in Ames IA.

Details

From the data documentation reference, the columns include:

- `Order`: Observation number
- `PID`: Parcel identification number - can be used with city web site for parcel review.
- MS SubClass: Identifies the type of dwelling involved in the sale.
- MS Zoning: Identifies the general zoning classification of the sale.
- Lot Frontage: Linear feet of street connected to property
- Lot Area: Lot size in square feet
- `Street`: Type of road access to property
- `Alley`: Type of alley access to property
- Lot Shape: General shape of property
- Land Contour: Flatness of the property
- `Utilities`: Type of utilities available
- Lot Config: Lot configuration
- Land Slope: Slope of property
- `Neighborhood`: Physical locations within Ames city limits (map available)
- Condition 1: Proximity to various conditions
- Condition 2: Proximity to various conditions (if more than one is present)
- Bldg Type: Type of dwelling
- House Style: Style of dwelling
- Overall Qual: Rates the overall material and finish of the house
- Overall Cond: Rates the overall condition of the house
- Year Built: Original construction date
- Year Remod/Add: Remodel date (same as construction date if no remodeling or additions)
- Roof Style: Type of roof
- Roof Matl: Roof material
- Exterior 1: Exterior covering on house
- Exterior 2: Exterior covering on house (if more than one material)
- Mas Vnr Type: Masonry veneer type
- Mas Vnr Area: Masonry veneer area in square feet
- Exter Qual: Evaluates the quality of the material on the exterior
- Exter Cond: Evaluates the present condition of the material on the exterior

- `Foundation`: Type of foundation
- Bsmt Qual: Evaluates the height of the basement
- Bsmt Cond: Evaluates the general condition of the basement
- Bsmt Exposure: Refers to walkout or garden level walls
- BsmtFin Type 1: Rating of basement finished area
- BsmtFin SF 1: Type 1 finished square feet
- BsmtFinType 2: Rating of basement finished area (if multiple types)
- BsmtFin SF 2: Type 2 finished square feet
- Bsmt Unf SF: Unfinished square feet of basement area
- Total Bsmt SF: Total square feet of basement area
- `Heating`: Type of heating
- `HeatingQC`: Heating quality and condition
- Central Air: Central air conditioning
- `Electrical`: Electrical system
- 1st Flr SF: First Floor square feet
- 2nd Flr SF: Second floor square feet
- Low Qual Fin SF: Low quality finished square feet (all floors)
- Gr Liv Area: Above grade (ground) living area square feet
- Bsmt Full Bath: Basement full bathrooms
- Bsmt Half Bath: Basement half bathrooms
- Full Bath: Full bathrooms above grade
- Half Bath: Half baths above grade
- `Bedroom`: Bedrooms above grade (does NOT include basement bedrooms)
- `Kitchen`: Kitchens above grade
- `KitchenQual`: Kitchen quality
- `TotRmsAbvGrd`: Total rooms above grade (does not include bathrooms)
- `Functional`: Home functionality (Assume typical unless deductions are warranted)
- `Fireplaces`: Number of fireplaces
- `FireplaceQu`: Fireplace quality
- Garage Type: Garage location
- Garage Yr Blt: Year garage was built
- Garage Finish: Interior finish of the garage
- Garage Cars: Size of garage in car capacity
- Garage Area: Size of garage in square feet
- Garage Qual: Garage quality
- Garage Cond: Garage condition
- Paved Drive: Paved driveway
- Wood Deck SF: Wood deck area in square feet
- Open Porch SF: Open porch area in square feet
- Enclosed Porch: Enclosed porch area in square feet
- 3-Ssn Porch: Three season porch area in square feet

- Screen Porch: Screen porch area in square feet
- Pool Area: Pool area in square feet
- Pool QC: Pool quality

- Fence: Fence quality
- Misc Feature: Miscellaneous feature not covered in other categories
- Misc Val: $Value of miscellaneous feature
- Mo Sold: Month Sold
- Yr Sold: Year Sold
- Sale Type: Type of sale
- Sale Condition: Condition of sale

Value

`ames_raw`   a tibble

Source

De Cock, D. (2011). "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project," *Journal of Statistics Education*, Volume 19, Number 3.

https://ww2.amstat.org/publications/jse/v19n3/decock/DataDocumentation.txt

http://ww2.amstat.org/publications/jse/v19n3/decock.pdf

`ames_schools`   *Ames Public Schools*

Description

Locations of local schools and outline of elementary school districts. Elementary school district boundaries are returned as `sf` object in CRS 4326

Details

The data set includes the school name and the geocodes.

Value

`ames_schools_geo`
                    a tibble

---

make_ames *Create a Processed Version of the Ames Housing Data*

Description

Create a Processed Version of the Ames Housing Data

Usage

```
make_ames()

make_ames_new()

make_ordinal_ames()
```

Details

For the processed version, the exact details can be found in the code of `make_ames` but a summary of the differences between these data sets and `ames_raw` is:

- All factors are *unordered*.

- `PID` and `Order` are removed.

- Spaces and special characters in column names where changed to snake case. To be consistent, `SalePrice` was changed to `Sale_Price`.

- Many factor levels were changed to be more understandable (e.g. `Split_or_Multilevel` instead of `080`)

- Many missing values were reset. For example, if the variable `Bsmt_Qual` was missing, this implies that there is no basement on the property. Instead of a missing value, the value of `Bsmt_Qual` was changed to `No_Basement`. Similarly, numeric data pertaining to basements were set to zero where appropriate such as variables `Bsmt_Full_Bath` and `Total_Bsmt_SF`.

- `Garage_Yr_Blt` contained many missing data and was removed.

- Approximate longitude and latitude are included for the properties. Also, note that there are 6 properties with identical geotags. These are units within the same building. For some properties, updated versions of the PID identifiers were found and are replaced with new values.

`make_ordinal_ames` is the same as `make_ames` but many factor variables were changed to class `ordered` (see below).

The documentation for `ames_raw()` contains descriptions of the columns although, as noted above, the column names in `ames_raw()` are slightly different from the processed versions.

`make_ames_new()` creates a data set of new properties. These were populated using less data sources than the original and lack a number of the condition and quality. Both properties were unsold at the time of this writing.

Value

A tibble with the data.

Examples

```
ames <- make_ames()
nrow(ames)
summary(ames$Sale_Price)

ames_ord <- make_ordinal_ames()
ord_vars <- vapply(ames_ord, is.ordered, logical(1))
names(ord_vars)[ord_vars]
```

Index