Package 'r02pro'

May 31, 2023

Title R Programming: Zero to Pro

Version 0.2

Description This is a companion package of the book ``R Program-
ming: Zero to Pro'' <https://r02pro.github.io/>. It con-
tains the datasets used in the book and provides interactive exercises correspond-
ing to the book. It covers a wide range of topics including visualization, data transforma-
tion, tidying data, data input and output.

License GPL-2

URL https://r02pro.github.io/

Encoding UTF-8

Depends R (>= 3.5.0)

RoxygenNote 7.2.3

Imports learnr

LazyData true

NeedsCompilation no

Author Yang Feng [aut, cre],
Jianan Zhu [aut]

Maintainer Yang Feng <yangfengstat@gmail.com>

Repository CRAN

Date/Publication 2023-05-31 14:30:02 UTC

R topics documented:

---

ahp *Ames Housing Price data.*

Description

A dataset of 2048 houses in Ames, Iowa from 2006 to 2010, with 56 features including the sale date and price.

Usage

ahp

Format

A data frame with variables:

dt_sold Date Sold

yr_sold Year Sold

mo_sold Month Sold

yr_built Original construction date

yr_remodel Remodel date (same as construction date if no remodeling or additions)

bldg_class The building class

- 20: 1-STORY 1946 & NEWER ALL STYLES
- 30: 1-STORY 1945 & OLDER
- 40: 1-STORY W/FINISHED ATTIC ALL AGES
- 45: 1-1/2 STORY - UNFINISHED ALL AGES
- 50: 1-1/2 STORY FINISHED ALL AGES
- 60: 2-STORY 1946 & NEWER
- 70: 2-STORY 1945 & OLDER
- 75: 2-1/2 STORY ALL AGES
- 80: SPLIT OR MULTI-LEVEL
- 85: SPLIT FOYER
- 90: DUPLEX - ALL STYLES AND AGES
- 120: 1-STORY PUD (Planned Unit Development) - 1946 & NEWER
- 150: 1-1/2 STORY PUD - ALL AGES
- 160: 2-STORY PUD - 1946 & NEWER
- 180: PUD - MULTILEVEL - INCL SPLIT LEV/FOYER
- 190: 2 FAMILY CONVERSION - ALL STYLES AND AGES

bldg_type Type of dwelling

- 1Fam: Single-family Detached
- 2FmCon: Two-family Conversion; originally built as one-family dwelling
- Duplx: Duplex

- TwnhsE: Townhouse End Unit
- TwnhsI: Townhouse Inside Unit

house_style Style of dwelling

- 1Story: One story
- 1.5Fin: One and one-half story: 2nd level finished
- 1.5Unf: One and one-half story: 2nd level unfinished
- 2Story: Two story
- 2.5Fin: Two and one-half story: 2nd level finished
- 2.5Unf: Two and one-half story: 2nd level unfinished
- SFoyer: Split Foyer
- SLvl: Split Level

zoning Identifies the general zoning classification of the sale

- A: Agriculture
- C: Commercial
- FV: Floating Village Residential
- I: Industrial
- RH: Residential High Density
- RL: Residential Low Density
- RP: Residential Low Density Park
- RM: Residential Medium Density

neighborhd Physical locations within Ames city limits

- Blmngtn: Bloomington Heights
- Blueste: Bluestem
- BrDale: Briardale
- BrkSide: Brookside
- ClearCr: Clear Creek
- CollgCr: College Creek
- Crawfor: Crawford
- Edwards: Edwards
- Gilbert: Gilbert
- IDOTRR: Iowa DOT and Rail Road
- MeadowV: Meadow Village
- Mitchel: Mitchell
- Names: North Ames
- NoRidge: Northridge
- NPkVill: Northpark Villa
- NridgHt: Northridge Heights
- NWAmes: Northwest Ames
- OldTown: Old Town
- SWISU: South & West of Iowa State University
- Sawyer: Sawyer

- SawyerW: SawyerW
- Somerst: Somerset
- StoneBr: Stone Brook
- Timber: Timberland
- Veenker: Veenker

oa_cond Overall condition rating

- 10: Very Excellent
- 9: Excellent
- 8: Very Good
- 7: Good
- 6: Above Average
- 5: Average
- 4: Below Average
- 3: Fair
- 2: Poor
- 1: Very Poor

oa_qual Overall material and finish quality

- 10: Very Excellent
- 9: Excellent
- 8: Very Good
- 7: Good
- 6: Above Average
- 5: Average
- 4: Below Average
- 3: Fair
- 2: Poor
- 1: Very Poor

func Home functionality rating

- Typ: Typical Functionality
- Min1: Minor Deductions 1
- Min2Minor Deductions 2
- Mod: Moderate Deductions
- Maj1: Major Deductions 1
- Maj2: Major Deductions 2
- Sev: Severely Damaged
- Sal: Salvage only

liv_area living area square feet

1fl_area First Floor square feet

2fl_area Second floor square feet

tot_rms Total rooms

bedroom Number of bedrooms

bathroom Number of bathrooms

kit Number of kitchens

kit_qual Kitchen quality

central_air Central air conditioning

- N: No
- Y: Yes

elect Electrical system

- SBrkr: Standard Circuit Breakers & Romex
- FuseA: Fuse Box over 60 AMP and all Romex wiring (Average)
- FuseF: 60 AMP Fuse Box and mostly Romex wiring (Fair)
- FuseP: 60 AMP Fuse Box and mostly knob & tube wiring (poor)
- Mix: Mixed

bsmt_area Total square feet of basement area

bsmt_cond General condition of the basement

bsmt_exp Walkout or garden level basement walls

- Gd: Good Exposure
- Av: Average Exposure (split levels or foyers typically score average or above)
- Mn: Mimimum Exposure
- No: No Exposure
- NA: No Basement

bsmt_ht Height of the basement

- Excellent: 100+ inches
- Good: 90-99 inches
- Average: 80-89 inches
- Fair: 70-79 inches
- Poor: <70 inches
- NA: No Basement

bsmt_fin_qual Quality of basement finished area

- GLQ: Good Living Quarters
- ALQ: Average Living Quarters
- BLQ: Below Average Living Quarters
- Rec: Average Rec Room
- LwQ: Low Quality
- Unf: Unfinshed
- NA: No Basement

ext_cond Present condition of the material on the exterior

ext_cover Exterior covering on house

- AsbShng: Asbestos Shingles
- AsphShn: Asphalt Shingles
- BrkComm: Brick Common

- BrkFace: Brick Face
- CBlock: Cinder Block
- CemntBd: Cement Board
- HdBoard: Hard Board
- ImStucc: Imitation Stucco
- MetalSd: Metal Siding
- Other: Other
- Plywood: Plywood
- PreCast: PreCast
- Stone: Stone
- Stucco: Stucco
- VinylSd: Vinyl Siding
- Wd Sdng: Wood Siding
- WdShing: Wood Shingles

ext_qual Exterior material quality

fdn Type of foundation

- BrkTil: Brick & Tile
- CBlock: Cinder Block
- PConc: Poured Contrete
- Slab: Slab
- Stone: Stone
- Wood: Wood

fence Fence quality

- GdPrv: Good Privacy
- MnPrv: Minimum Privacy
- GdWo: Good Wood
- MnWw: Minimum Wood/Wire
- NA: No Fence

fp Number of fireplaces

fp_qual Fireplace quality

gar_area Size of garage in square feet

gar_car Size of garage in car capacity

gar_cond Garage condition

gar_fin Interior finish of the garage

- Fin: Finished
- RFn: Rough Finished
- Unf: Unfinished
- NA: No Garage

gar_qual Garage quality

gar_type Garage location

- 2Types: More than one type of garage
- Attchd: Attached to home
- Basment: Basement Garage
- BuiltIn: Built-In (Garage part of house - typically has room above garage)
- CarPort: Car Port
- Detchd: Detached from home
- NA: No Garage

gar_yr Year garage was built

heat_qual Heating quality and condition

land_contour Flatness of the property

- Lvl: Near Flat/Level
- Bnk: Banked - Quick and significant rise from street grade to building
- HLS: Hillside - Significant slope from side to side
- Low: Depression

land_slope Slope of property

- Gtl: Gentle slope
- Mod: Moderate Slope
- Sev: Severe Slope

lot_area Lot size in square feet

lot_config Lot configuration

- Inside: Inside lot
- Corner: Corner lot
- CulDSac: Cul-de-sac
- FR2: Frontage on 2 sides of property
- FR3: Frontage on 3 sides of property

lot_frontage Linear feet of street connected to property

lot_shape General shape of lot

- Reg: Regular
- IR1: Slightly irregular
- IR2: Moderately Irregular
- IR3: Irregular

pave_dr Paved driveway

- Y: Paved
- P: Partial Pavement
- N: Dirt/Gravel

roof_matl Roof material

- ClyTile: Clay or Tile
- CompShg: Standard (Composite) Shingle
- Membran: Membrane
- Metal: Metal

- Roll: Roll
- Tar&Grv: Gravel & Tar
- WdShake: Wood Shakes
- WdShngl: Wood Shingles

roof_style Type of roof

- Flat: Flat
- Gable: Gable
- Gambrel: Gabrel (Barn)
- Hip: Hip
- Mansard: Mansard
- Shed: Shed

op_area Open porch area in square feet

ep_area Enclosed porch area in square feet

wd_area Wood deck area in square feet

sale_price The property's sale price in thousand dollars

Source

The original data comes from <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>. Some data cleaning were applied.

---

gm   *Gapminder Global Health Data.*

Description

A dataset of 239 countries worldwide with 33 sociodemographic and public health features, some of which are the same variable but measured in dichotomized genders.

Usage

gm

Format

A data frame with variables:

country Country

year the year of 2004

smoking_female Percentage of female (over age 15) that smoke

smoking_male Percentage of male (over age 15) that smoke

lungcancer_newcases_female Number of new female cases of lung cancer in 100,000 residents, adjusting each country's age composition to the world population. Unit: person per 100,000 people

lungcancer_newcases_male Number of new male cases of lung cancer in 100,000 residents, adjusting each country's age composition to the world population. Unit: person per 100,000 people

owid_edu_idx OWID Education Index: Education index calculated based on Avg years of schooling, taking values 0 as minimum and 15 as maximum.

food_supply Calories measures the energy content of the food. The required intake varies, but it is normally in the range of 1500-3000 kilocalories per day. Unit: kilocalories per person and day

average_daily_income This is the average daily household per capita income or consumption expenditure from the survey expressed in 2011 PPP. Unit: $1,000

sanitation The percentage of people using at least basic sanitation services, that is, improved sanitation facilities that are not shared with other households.

child_mortality Death of children under five years of age per 1,000 live births. Unit: per 1000 live births

income_per_person Gross domestic product per person adjusted for differences in purchasing power (in international $, fixed 2017 prices, PPP based on 2017 ICP). Unit: $1,000

HDI Human Development Index. An index used to rank countries by the level of "human development" from three dimensions: health level, educational level, and living standard.

alcohol_male Total alcohol consumption per capita, male, liters of pure alcohol, 15+ years of age.

alcohol_female Total alcohol consumption per capita, female, liters of pure alcohol, 15+ years of age.

livercancer_newcases_male Number of new male cases of liver cancer in 100,000 residents, adjusting each country's age composition to the world population. Unit: person per 100,000 people.

livercancer_newcases_female Number of new female cases of liver cancer in 100,000 residents, adjusting each country's age composition to the world population. Unit: person per 100,000 people.

mortality_male Mortality rate, adult, male (per 1,000 male adults).

mortality_female Mortality rate, adult, female (per 1,000 female adults).

cholesterol_fat_in_blood_male The mean TC (Total Cholesterol) of the male population, counted in mmol per L.

cholesterol_fat_in_blood_female The mean TC (Total Cholesterol) of the female population, counted in mmol per L.

continent The continent that a country is part of

- Africa
- Americas
- Asia
- Europe
- Oceania

region Sub specification of the region that a country belongs to

population Total population of each country in 2004. Unit: 1,000 people

life_expectancy The average number of years a newborn child would live if current mortality patterns were to stay the same. Unit: year

sugar The quantity of food consumption of sugar and sweeteners per person. Unit: grams per person and day

BMI_female The mean BMI (Body Mass Index) of the female population; this mean is calculated as if each country has the same age composition as the world population. Unit: Kilogram per square meter

BMI_female_group Group according to `BMI_female`

- under_weight: < 18.5
- normal_weight: 18.5 - 24.9
- pre_obesity: 25.0 - 29.9
- obesity_class_I: 30.0 - 34.9
- obesity_class_II: 35.0 - 39.9

BMI_male The mean BMI (Body Mass Index) of the male population; this mean is calculated as if each country has the same age composition as the world population. Unit: Kilogram per square meter

BMI_male_group Group according to `BMI_male`

- under_weight: < 18.5
- normal_weight: 18.5 - 24.9
- pre_obesity: 25.0 - 29.9
- obesity_class_I: 30.0 - 34.9
- obesity_class_II: 35.0 - 39.9

health_spending The sum of public and private health expenditure as a percentage of GDP. Unit: percent

GDP_per_capita Inflation-adjusted gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Unit: $1,000

HDI_category Human Development Index categories

- Very high: HDI above 0.800
- High: HDI between 0.700 and 0.799
- Medium: HDI between 0.550–0.699
- Low: HDI below 0.549

Source

The original data comes from https://www.gapminder.org/data/. Some data cleaning was applied.

---

gm2004 *Gapminder Global Health Data in year 2004.*

Description

A dataset of 236 countries worldwide with 23 sociodemographic and public health features, some of which are the same variable but measured in dichotomized genders.

Usage

    gm2004

Format

A data frame with variables:

country Country

year the year of 2004

gender Gender

continent The continent that a country is part of

- Africa
- Americas
- Asia
- Europe
- Oceania

region Sub specification of region that a contry belongs to

population Total population of each country in 2004. Unit: 1,000 people

BMI The mean BMI (Body Mass Index) of the whole population; this mean is calculated as if each country has the same age composition as the world population. Unit: Kilogram per square meter

livercancer_newcases Number of new cases of liver cancer in 100,000 residents in 2004, adjusting each country's age composition to the world population. Unit: person per 100,000 people

lungcancer_newcases Number of new cases of lung cancer in 100,000 residents in 2004, adjusting each country's age composition to the world population. Unit: person per 100,000 people

cholesterol Mean TC (Total Cholesterol) of the whole population, adjusting each country's age composition to the world population. Unit: mmol/L (Millimoles per liter)

life_expectancy The average number of years a newborn child would live if current mortality patterns were to stay the same. Unit: year

sugar The quantity of good consumption of sugar and sweeteners per person. Unit: grams per person and day

health_spending The sum of public and private health expenditure as a percentage of GDP. Unit: percent

GDP_per_capita Inflation-adjusted gross domestic product divided by midyear population. GDP is the sum of gross value added by all residents producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degredation of natural resources. Unit: $1,000

HDI Human Development Index. An index used to rank countries by level of "human development" from three dimensions: health level, educational level, and living standard.

HDI_category Human Development Index categories

- Very high: HDI above 0.800
- High: HDI between 0.700 and 0.799
- Medium: HDI between 0.550–0.699
- Low: HDI below 0.549

smoking Percentage of both men and women (over age 15) that smoke

food_supply Caloreis measures the energy content of the food. The required intake varies, but it is normally in the range of 1500- 3000 kilocalories per day. Unit: kilocalories per person and day

owid_edu_idx OWID Education Index: Education index calculated based on Avg years of schooling, taking values 0 as minimum and 15 as maximum.

average_daily_income This is the average daily household per capita income or consumption expenditure from the survey expressed in 2011 PPP. Unit: $1,000

income_per_person Gross domestic product per person adjusted for differences in purchasing pwoer (in international $, fixed 2017 prices, PPP based on 2017 ICP). Unit: $1,000

sanitation The percentage of people using at least basic sanitation services, that is, improved sanitation faciliteis that are not hsared with other households.

child_mortality Death of children under five years of age per 1,000 live births. Unit: per 1000 live births

Source

The original data comes from https://www.gapminder.org/data/. Some data cleaning were applied.

---

r02pro   *Do the interactive exercises*

Description

This function provides interactive exercises for each lesson corresponding to each subsection of the book "R Programming: Zero to Pro"

Usage

```
r02pro(id)
```

Arguments

id   the index of the lesson

Value

This function is an interactive exercise. Hence, no value returned.

Examples

```
#Do the exercise for Section 1.1
## Not run: r02pro(1.1)
```

---

sahp   *Small Version of Ames Housing Price data.*

Description

The small version of Ames Housing Price data of 165 observations, with 12 features including the sale date and price.

Usage

```
sahp
```

Format

A data frame with 165 observations and 12 features:

dt_sold Date Sold

bedroom Number of bedrooms

bathroom Number of bathrooms

gar_car Size of garage in car capacity

oa_qual Overall material and finish quality

- 10: Very Excellent
- 9: Excellent
- 8: Very Good
- 7: Good
- 6: Above Average
- 5: Average
- 4: Below Average
- 3: Fair
- 2: Poor
- 1: Very Poor

liv_area living area square feet

lot_area Lot size in square feet

house_style Style of dwelling

- 1Story: One story
- 1.5Fin: One and one-half story: 2nd level finished
- 1.5Unf: One and one-half story: 2nd level unfinished
- 2Story: Two story
- 2.5Fin: Two and one-half story: 2nd level finished
- 2.5Unf: Two and one-half story: 2nd level unfinished
- SFoyer: Split Foyer
- SLvl: Split Level

kit_qual Kitchen quality

heat_qual Heating quality and condition

central_air Central air conditioning

- N: No
- Y: Yes

sale_price The property's sale price in thousand dollars

Source

The original data comes from https://www.kaggle.com/c/house-prices-advanced-regression-techniques. Some data cleaning were applied.

Index