

The Ames Housing Data Set

The Ames Housing Data Set Description

The `ames_housing.csv` file on Canvas contains a subset of the original Ames Housing Data set curated by Dean De Cock from Truman State University. You can read about the original data set here (<https://jse.amstat.org/v19n3/decock/DataDocumentation.txt>) and here (<https://jse.amstat.org/v19n3/decock.pdf>).

I took the original data set and cleaned it up to make it more appropriate for this course. In particular, I

- Removed large houses with an above ground living area of more than 1500 square feet.
- Filtered the data set to include houses that had a 'Normal' sales condition.
- Included only numeric variables (categorical variables were removed).
- Removed rows with NA values.
- Renamed a few variables to make their meaning clearer.

This was done by running the following code, where `AmesHousing.txt` came from <https://jse.amstat.org/v19n3/decock/AmesHousing.txt>.

```
library(readr)
library(janitor)
library(tidyverse)

ames_full = read_delim('AmesHousing.txt', delim='\t') |>
  clean_names(case = 'upper_camel')

exclude_vars = c("MSSubClass", "MSZoning", "MoSold", "YrSold",
                 "MiscVal", "BsmtUnfSf", "Order", "PoolArea", "LowQualFinSf")
ames = ames_full |>
  filter(GrLivArea < 1500) |>
  filter(SaleCondition == 'Normal') |>
  select_if(is.numeric) |>
  select_if(~ !any(is.na(.))) |>
  dplyr::select(-one_of(exclude_vars)) |>
  rename(FirstFlrSf = X1StFlrSf) |>
  rename(SecondFlrSf = X2NdFlrSf) |>
  rename(ThreeSsnPorch = X3SsnPorch)

write_csv(ames, "ames_housing.csv")
```

Data Dictionary

The full data set contains 1,359 homes and the following 27 variables:

- `LotArea`: Lot size in square feet.
- `OverallQual`: Overall quality of the house's material and finish. The scale ranges from 1 (Very Poor) to 9 (Very Excellent).
- `OverallCond`: Overall condition rating. The scale ranges from 1 (Very Poor) to 9 (Very Excellent).

- YearBuilt: Original construction date.
- YearRemodAdd: Remodel date.
- BsmtFinSf1: Type 1 finished square feet.
- BsmtFinType2: Type 2 finished square feet.
- TotalBsmtSf: Total square feet of basement area.
- FirstFlrSf: First floor square feet.
- SecondFlrSf: Second floor square feet.
- GrLivArea: Above grade (ground) living area square feet.
- BsmtFullBath: Number of full bathrooms in the basement.
- BsmtHalfBath: Number of half baths in the basement.
- FullBath: Number of full bathrooms above ground.
- HalfBath: Number of half baths above ground.
- BedroomAbvGr: Number of Bedrooms above ground.
- KitchenAbvGr: Number of Kitchens above ground.
- TotRmsAbvGrd: Total rooms above ground (does not include bathrooms).
- Fireplaces: Number of fireplaces.
- GarageCars: Size of garage in car capacity.
- GarageArea: Size of garage in square feet.
- WoodDeckSf: Wood deck area in square feet.
- OpenPorchSf: Open porch area in square feet.
- EnclosedPorch: Enclosed porch area in square feet.
- ThreeSsnProch: Three season porch area in square feet.
- ScreenPorch: Screen porch area in square feet.
- SalePrice: The property's sale price in dollars.