```
NAME: AmesHousing.txt
TYPE: Population
SIZE: 2930 observations, 82 variables
ARTICLE TITLE: Ames Iowa: Alternative to the Boston Housing Data Set
```

DESCRIPTIVE ABSTRACT: Data set contains information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

SOURCES:
Ames, Iowa Assessor's Office

VARIABLE DESCRIPTIONS:
Tab characters are used to separate variables in the data file. The data has 82 columns which include 23 nominal, 23 ordinal, 14 discrete, and 20 continuous variables (and 2 additional observation identifiers).

Order (Discrete): Observation number

PID (Nominal): Parcel identification number - can be used with city web site for parcel review.

MS SubClass (Nominal): Identifies the type of dwelling involved in the sale.

020	1-STORY 1946 & NEWER ALL STYLES
030	1-STORY 1945 & OLDER
040	1-STORY W/FINISHED ATTIC ALL AGES
045	1-1/2 STORY - UNFINISHED ALL AGES
050	1-1/2 STORY FINISHED ALL AGES
060	2-STORY 1946 & NEWER
070	2-STORY 1945 & OLDER
075	2-1/2 STORY ALL AGES
080	SPLIT OR MULTI-LEVEL
085	SPLIT FOYER
090	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

MS Zoning (Nominal): 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

Lot Frontage (Continuous): Linear feet of street connected to property

Lot Area (Continuous): Lot size in square feet

Street (Nominal): Type of road access to property

Grvl	Gravel
Pave	Paved

Alley (Nominal): Type of alley access to property

Grvl	Gravel
Pave	Paved
NA	No alley access

Lot Shape (Ordinal): General shape of property

       Reg   Regular
       IR1   Slightly irregular
       IR2   Moderately Irregular
       IR3   Irregular

Land Contour (Nominal): 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

Utilities (Ordinal): Type of utilities available

       AllPub   All public Utilities (E,G,W,& S)
       NoSewr   Electricity, Gas, and Water (Septic Tank)
       NoSeWa   Electricity and Gas Only
       ELO   Electricity only

Lot Config (Nominal): 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

Land Slope (Ordinal): Slope of property

       Gtl   Gentle slope
       Mod   Moderate Slope
       Sev   Severe Slope

Neighborhood (Nominal): Physical locations within Ames city limits (map available)

       Blmngtn Bloomington Heights
       Blueste Bluestem
       BrDale   Briardale
       BrkSide Brookside
       ClearCr Clear Creek
       CollgCr College Creek
       Crawfor Crawford
       Edwards Edwards
       Gilbert Gilbert
       Greens   Greens
       GrnHill Green Hills
       IDOTRR   Iowa DOT and Rail Road
       Landmrk Landmark
       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 Sawyer West
       Somerst Somerset
       StoneBr Stone Brook
       Timber   Timberland

Veenker   Veenker

Condition 1 (Nominal): Proximity to various conditions

        Artery   Adjacent to arterial street
        Feedr   Adjacent to feeder street
        Norm   Normal
        RRNn   Within 200' of North-South Railroad
        RRAn   Adjacent to North-South Railroad
        PosN   Near positive off-site feature--park, greenbelt, etc.
        PosA   Adjacent to postive off-site feature
        RRNe   Within 200' of East-West Railroad
        RRAe   Adjacent to East-West Railroad

Condition 2 (Nominal): Proximity to various conditions (if more than one is present)

        Artery   Adjacent to arterial street
        Feedr   Adjacent to feeder street
        Norm   Normal
        RRNn   Within 200' of North-South Railroad
        RRAn   Adjacent to North-South Railroad
        PosN   Near positive off-site feature--park, greenbelt, etc.
        PosA   Adjacent to postive off-site feature
        RRNe   Within 200' of East-West Railroad
        RRAe   Adjacent to East-West Railroad

Bldg Type (Nominal): 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 (Nominal): 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

Overall Qual (Ordinal): Rates the overall material and finish of the house

        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

Overall Cond (Ordinal): Rates the overall condition of the house

        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
```

Year Built (Discrete): Original construction date

Year Remod/Add (Discrete): Remodel date (same as construction date if no remodeling or additions)

Roof Style (Nominal): Type of roof

```
       Flat     Flat
       Gable    Gable
       Gambrel  Gabrel (Barn)
       Hip      Hip
       Mansard  Mansard
       Shed     Shed
```

Roof Matl (Nominal): 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
```

Exterior 1 (Nominal): 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
```

Exterior 2 (Nominal): Exterior covering on house (if more than one material)

```
       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


Mas Vnr Type (Nominal): Masonry veneer type

        BrkCmn   Brick Common
        BrkFace  Brick Face
        CBlock   Cinder Block
        None     None
        Stone    Stone


Mas Vnr Area (Continuous): Masonry veneer area in square feet

Exter Qual (Ordinal): Evaluates the quality of the material on the exterior

        Ex       Excellent
        Gd       Good
        TA       Average/Typical
        Fa       Fair
        Po       Poor


Exter Cond (Ordinal): Evaluates the present condition of the material on the exterior

        Ex       Excellent
        Gd       Good
        TA       Average/Typical
        Fa       Fair
        Po       Poor


Foundation (Nominal): Type of foundation

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

Bsmt Qual (Ordinal): Evaluates the height of the basement

        Ex       Excellent (100+ inches)
        Gd       Good (90-99 inches)
        TA       Typical (80-89 inches)
        Fa       Fair (70-79 inches)
        Po       Poor (<70 inches
        NA       No Basement

Bsmt Cond (Ordinal): Evaluates the general condition of the basement

        Ex       Excellent
        Gd       Good
        TA       Typical - slight dampness allowed
        Fa       Fair - dampness or some cracking or settling
        Po       Poor - Severe cracking, settling, or wetness
        NA       No Basement

Bsmt Exposure   (Ordinal): Refers to walkout or garden level 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
```

BsmtFin Type 1 (Ordinal): Rating 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

BsmtFin SF 1 (Continuous): Type 1 finished square feet

BsmtFinType 2   (Ordinal): Rating of basement finished area (if multiple types)

        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

BsmtFin SF 2 (Continuous): Type 2 finished square feet

Bsmt Unf SF (Continuous): Unfinished square feet of basement area

Total Bsmt SF (Continuous): Total square feet of basement area

Heating (Nominal): Type of heating

        Floor   Floor Furnace
        GasA   Gas forced warm air furnace
        GasW   Gas hot water or steam heat
        Grav   Gravity furnace
        OthW   Hot water or steam heat other than gas
        Wall   Wall furnace

HeatingQC (Ordinal): Heating quality and condition

        Ex   Excellent
        Gd   Good
        TA   Average/Typical
        Fa   Fair
        Po   Poor

Central Air (Nominal): Central air conditioning

        N   No
        Y   Yes

Electrical (Ordinal): 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

1st Flr SF (Continuous): First Floor square feet

2nd Flr SF (Continuous) : Second floor square feet

Low Qual Fin SF (Continuous): Low quality finished square feet (all floors)

Gr Liv Area (Continuous): Above grade (ground) living area square feet

Bsmt Full Bath (Discrete): Basement full bathrooms

Bsmt Half Bath (Discrete): Basement half bathrooms

Full Bath (Discrete): Full bathrooms above grade

Half Bath (Discrete): Half baths above grade

Bedroom (Discrete): Bedrooms above grade (does NOT include basement bedrooms)

Kitchen (Discrete): Kitchens above grade

KitchenQual (Ordinal): Kitchen quality

        Ex   Excellent
        Gd   Good
        TA   Typical/Average
        Fa   Fair
        Po   Poor

TotRmsAbvGrd   (Discrete): Total rooms above grade (does not include bathrooms)

Functional (Ordinal): Home functionality (Assume typical unless deductions are warranted)

        Typ   Typical Functionality
        Min1   Minor Deductions 1
        Min2   Minor Deductions 2
        Mod   Moderate Deductions
        Maj1   Major Deductions 1
        Maj2   Major Deductions 2
        Sev   Severely Damaged
        Sal   Salvage only

Fireplaces (Discrete): Number of fireplaces

FireplaceQu (Ordinal): Fireplace quality

        Ex   Excellent - Exceptional Masonry Fireplace
        Gd   Good - Masonry Fireplace in main level
        TA   Average - Prefabricated Fireplace in main living area or Masonry Fireplace in
basement
        Fa   Fair - Prefabricated Fireplace in basement
        Po   Poor - Ben Franklin Stove
        NA   No Fireplace

Garage Type (Nominal): 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

Garage Yr Blt (Discrete): Year garage was built

Garage Finish (Ordinal) : Interior finish of the garage

        Fin   Finished
        RFn   Rough Finished
        Unf   Unfinished
        NA   No Garage

Garage Cars (Discrete): Size of garage in car capacity

Garage Area (Continuous): Size of garage in square feet

Garage Qual (Ordinal): Garage quality

       Ex   Excellent
       Gd   Good
       TA   Typical/Average
       Fa   Fair
       Po   Poor
       NA   No Garage

Garage Cond (Ordinal): Garage condition

       Ex   Excellent
       Gd   Good
       TA   Typical/Average
       Fa   Fair
       Po   Poor
       NA   No Garage

Paved Drive (Ordinal): Paved driveway

       Y   Paved
       P   Partial Pavement
       N   Dirt/Gravel

Wood Deck SF (Continuous): Wood deck area in square feet

Open Porch SF (Continuous): Open porch area in square feet

Enclosed Porch (Continuous): Enclosed porch area in square feet

3-Ssn Porch (Continuous): Three season porch area in square feet

Screen Porch (Continuous): Screen porch area in square feet

Pool Area (Continuous): Pool area in square feet

Pool QC (Ordinal): Pool quality

       Ex   Excellent
       Gd   Good
       TA   Average/Typical
       Fa   Fair
       NA   No Pool

Fence (Ordinal): Fence quality

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

Misc Feature (Nominal): Miscellaneous feature not covered in other categories

       Elev   Elevator
       Gar2   2nd Garage (if not described in garage section)
       Othr   Other
       Shed   Shed (over 100 SF)
       TenC   Tennis Court
       NA   None

Misc Val (Continuous): $Value of miscellaneous feature

Mo Sold (Discrete): Month Sold (MM)

Yr Sold (Discrete): Year Sold (YYYY)

Sale Type (Nominal): Type of sale

       WD   Warranty Deed - Conventional
       CWD   Warranty Deed - Cash
       VWD   Warranty Deed - VA Loan
       New   Home just constructed and sold
       COD   Court Officer Deed/Estate
       Con   Contract 15% Down payment regular terms
       ConLw   Contract Low Down payment and low interest
       ConLI   Contract Low Interest
       ConLD   Contract Low Down
       Oth   Other

Sale Condition (Nominal): Condition of sale

       Normal Normal Sale
       Abnorml Abnormal Sale - trade, foreclosure, short sale
       AdjLand Adjoining Land Purchase
       Alloca Allocation - two linked properties with separate deeds, typically condo with a garage
unit
       Family Sale between family members
       Partial Home was not completed when last assessed (associated with New Homes)

SalePrice (Continuous): Sale price $$

SPECIAL NOTES:
There are 5 observations that an instructor may wish to remove from the data set before giving it to students (a plot of SALE PRICE versus GR LIV AREA will indicate them quickly). Three of them are true outliers (Partial Sales that likely don't represent actual market values) and two of them are simply unusual sales (very large houses priced relatively appropriately). I would recommend removing any houses with more than 4000 square feet from the data set (which eliminates these 5 unusual observations) before assigning it to students.

STORY BEHIND THE DATA:
This data set was constructed for the purpose of an end of semester project for an undergraduate regression course. The original data (obtained directly from the Ames Assessor's Office) is used for tax assessment purposes but lends itself directly to the prediction of home selling prices. The type of information contained in the data is similar to what a typical home buyer would want to know before making a purchase and students should find most variables straightforward and understandable.

PEDAGOGICAL NOTES:
Instructors unfamiliar with multiple regression may wish to use this data set in conjunction with an earlier JSE paper that reviews most of the major issues found in regression modeling:

Kuiper , S. (2008), "Introduction to Multiple Regression: How Much Is Your Car Worth?", Journal of Statistics Education Volume 16, Number 3 (2008).

Outside of the general issues associated with multiple regression discussed in this article, this particular data set offers several opportunities to discuss how the purpose of a model might affect the type of modeling done. User of this data may also want to review another JSE article related directly to real estate pricing:

Pardoe , I. (2008), "Modeling home prices using realtor data", Journal of Statistics Education Volume 16, Number 2 (2008).

One issue is in regards to homoscedasticity and assumption violations. The graph included in the article appears to indicate heteroscedasticity with variation increasing with sale price and this problem is evident in many simple home pricing models that focus only on house and lot sizes. Though

this violation can be alleviated by transforming the response variable (sale price), the resulting equation yields difficult to interpret fitted values (selling price in log or square root dollars). This situation gives the instructor the opportunity to talk about the costs (biased estimators, incorrect statistical tests, etc.) and benefits (ease of use) of not correcting this assumption violation. If the purpose in building the model is simply to allow a typical buyer or real estate agent to sit down and estimate the selling price of a house, such transformations may be unnecessary or inappropriate for the task at hand. This issue could also open into a discussion on the contrasts and comparisons between data mining, predictive models, and formal statistical inference.

A second issue closely related to the intended use of the model, is the handling of outliers and unusual observations. In general, I instruct my students to never throw away data points simply because they do not match a priori expectations (or other data points). I strongly make this point in the situation where data are being analyzed for research purposes that will be shared with a larger audience. Alternatively, if the purpose is to once again create a common use model to estimate a "typical" sale, it is in the modeler's best interest to remove any observations that do not seem typical (such as foreclosures or family sales).

REFERENCES:
Individual homes within the data set can be referenced directly from the Ames City Assessor webpage via the Parcel ID (PID) found in the data set. Note these are nominal values (non-numeric) so preceding 0's must be included in the data entry field on the website. Access to the database can be gained from the Ames site (http://www.cityofames.org/assessor/) by clicking on "property search" or by accessing the Beacon (http://beacon.schneidercorp.com/Default.aspx) website and inputting Iowa and Ames in the appropriate fields. A city map showing the location of all the neighborhoods is also available on the Ames site and can be accessed by clicking on "Maps" and then "Residential Assessment Neighborhoods (City of Ames Only)".

SUBMITTED BY:
Dean De Cock
Truman State University
100 E. Normal St., Kirksville, MO, 63501
decock@truman.edu