**Data Description:**

**Data Selected:** Ames Housing

**Number of observations:** 2,930 observations

**Number of variables:** 82 variables

**Data type of each variable and number of levels for factors:**

1 discrete identifier (order) and 1 nominal identifier (PID), in addition to:

|  |  |  |  |
| --- | --- | --- | --- |
| **23 Nominal (Factor)** | **23 Ordinal (Factor)** | **14 Discrete (Quantitative)** | **20 Continuous (Quantitative)** |
| MS Subclass (16)  MS Zoning (8)  Street (2)  Alley (3)  Land Contour (2)  Lot Config (5)  Neighborhood (28)  Condition 1 (9)  Condition 2 (9)  Bldg Type (5)  House Style (8)  Roof Style (6)  Roof Matl (8)  Exterior 1 (17)  Exterior 2 (17)  Mas Vnr Type (5)  Foundation (6)  Heating (6)  Central Air (2)  Garage Type (7)  Mic Feature (6)  Sale Type (10)  Sale Condition (6) | Lot Shape (4)  Utilities (4)  Lot Slope (3)  Overall Qual (10)  Overall Cond (10)  Exter Qual (5)  Exter Cond (5)  Bsmt Qual (6)  Bsmt Cond (6)  Bsmt Exposure (5)  BsmtFin Type 1 (7)  BsmtFin Type 2 (7)  HeatingQC (5)  Electrical (5)  Kitchen Qual (5)  Functional (8)  FireplaceQu (6)  Garage Finish (4)  Garage Qual (6)  Garage Cond (6)  Paved Drive (3)  Pool QC (5)  Fence (5) | Year Built  Year Remod/Add  Bsmt Full Bath  Bsmt Half Bath  Full Bath  Half Bath  Bedroom  Kitchen  TotRmsAbvGrd  Fireplaces  Garage Yr Blt  Garage Cars  Mo Sold  Yr Sold | Lot Frontage  Lot Area  Mas Vnr Area  BsmtFin SF 1  BsmtFin SF 2  Bsmt Unf SF  Total Bsmt SF  1st Flr SF  2nd Flr SF  Low Qual Fin SF  Gr Liv Area  Garage Area  Wood Deck SF  Open Porch SF  Enclosed Porch  3-Ssn Porch (3 season)  Screen Porch  Pool Area  Misc Val  SalePrice |

**Amount of missingness:** Total missing values is 13,960 out of 240,260 observations (about 5.81% is missing in this data)

**High-level overview of analysis:**

Over the course of this semester I hope to accomplish the following analysis using this data:

1. Determine which variables are most useful in determining home price.
2. Be able to predict home selling price after acquiring home information.
3. Examine neighborhood proximity to important local areas (school, shopping center, park, sewage treatment plant, etc.) using map information to include additional information in the data that can then be used to control for external factors. This will be done to better understand how internal home conditions determine home selling price.
4. Determine how to best improve the value of a home in this area and at this time.
   1. While more difficult, I would also like to attempt to find a way to optimize this process so a rough estimate of the improvement cost can be included in the data to find which improvements offer the greatest return on investment.