

# RFinalHYu

Haozhe (Jerry) Yu

2023-04-27

## Data Import

### Read in Raw Data

Read in the raw data directly from the url.

```
rawhouse <- read.csv("https://www4.stat.ncsu.edu/~online/ST308/Data/hyu23_house.csv")
```

### Data Subsetting

Create a tibble from the read in data table with the following modifications:

1. Remove any observations where
  - the `SaleType` variable takes the value “Other” or
  - the `BedroomAbvGr` variable takes on a value less than or equal to 2
2. Create a new variable with a name of your choosing that is the `SalePrice` variable divided by 100000.
3. The `GarageArea` and `MSZoning` variables are removed

```
House <- rawhouse %>%  
  filter(SaleType != "Other") %>%  
  filter(BedroomAbvGr > 2) %>%  
  mutate(SalePrice100k = SalePrice/100000) %>%  
  select(-GarageArea, -MSZoning)
```

Now print out the first 10 observations and first 6 variables of House.

```
House %>%  
  select(SalePrice, BsmtUnfSF, OverallQual, OpenPorchSF, BedroomAbvGr, YrSold) %>%  
  slice(1:10) %>%  
  kable()
```

SalePrice	BsmtUnfSF	OverallQual	OpenPorchSF	BedroomAbvGr	YrSold
208500	150	7	61	3	2008
181500	284	6	0	3	2007
223500	434	7	42	3	2008
140000	540	7	35	3	2006
250000	490	8	84	4	2008
307000	317	8	57	3	2007
200000	216	7	204	3	2009
279500	1494	7	33	3	2007
159000	468	5	102	3	2008
139000	525	5	0	3	2009

## Output Creation Steps

### Contingency Tables

Create a 2 way contingency table between `BsmtFinType2` and `LotShap`.

```
contingency <- table(House$BsmtFinType2,
  House$LotShap)
contingency
```

```
##
##      IR1 IR2or3 Reg
## Other  29      4  68
## Unf   229     24 361
```

The upper most value of 29 is the number of observations where `BsmtFinType2` equals “Other” and `LotShap` equals “IR1”.