

## EDA Graphs

These are the notes for the exploratory data analysis.

Please note the code chunks are what should be included to paste the image in an rmd file.

```

```

This graph shows the distribution of the numeric variables.

Normalcy of the features is not a requirement for OLS so we are not too worried about skewed features *per se*. Some variables that look problematic are the ones with very little variance because they offer little to the model.

```

```

The main issue here is that many of the variables are lopsided. This means there is little variance for our model to learn from.

```

```

The scatter of the numeric variables against `SalePrice` shows relationships between many of the variables. These are especially strong in `GrLivArea` and `OverallQual` but are apparent in many others as well.

Notice that many of these relationships look non-linear.

```

```

There are fewer trends here, but some dummy variables may be useful.

```

```

Transforming the response variable makes the relationships much linear as well as making trends more obvious.

There is not a tone to say about the correlation heat matrix except that there is little worry about multi co-linearity.