

# Analysis of Energy efficiency on buildings

*Tejovardhan Medamarti*

*24 January 2017*

## Overview and summary

This report is based on the dataset available at <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency#> and provides the raw data. Our interest in this dataset is to explore 3 interesting data points, and visualize how they are influencing the energy load. The dataset contains 8 input variables and 2 output variables (see variable information section).

We have figured that the Wall Area, Roof Area, Glazing Area are the key indicators and can influence the energy load efficiency for both (Heating and Cooling).

## Variable(s) Information:

- Relative Compactness
- Surface Area -  $\text{m}^2$
- Wall Area -  $\text{m}^2$
- Roof Area -  $\text{m}^2$
- Overall Height - m
- Orientation - 2:North, 3:East, 4:South, 5:West
- Glazing Area - 0%, 10%, 25%, 40% (of floor area)
- Glazing Area Distribution (Variance) - 1:Uniform, 2:North, 3:East, 4:South, 5:West
- Heating Load -  $\text{kWh/m}^2$
- Cooling Load -  $\text{kWh/m}^2$

**NOTE:** In order to run this file from source you need to ensure the following R packages are installed:

- ggplot2

We have a cleansed dataset available at Github from: <https://raw.githubusercontent.com/StephenElston/DataScience350/master/Lecture1/EnergyEfficiencyData.csv>.

Below are the steps followed in downloading the RAW data from the github and also, the actual dataset contains only the numerical data. So let's categorize few useful variables and convert them to a categorical for “**Orientation**”, “**Glazing Area**”, “**Glazing Area Distribution (variance)**” variables.

## Data Loading and Visualizations:

```
rm(list = ls())
```

```
SourceURL_Raw <- "https://raw.githubusercontent.com/StephenElston/DataScience350/master/Lecture1/EnergyEfficiencyData.csv"
```

```
energy.efficiency <- read.csv( SourceURL_Raw, header = TRUE)
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```

energy.efficiency$Orientation <- as.factor(energy.efficiency$Orientation)

levels(energy.efficiency$Orientation) <- c("North", "East", "South", "West")

energy.efficiency$Glazing.Area.Distribution <- as.factor(energy.efficiency$Glazing.Area.Distribution)

levels(energy.efficiency$Glazing.Area.Distribution) <- c("UnKnown", "Uniform", "North", "East", "South")

energy.efficiency$Glazing.Area <- as.factor(energy.efficiency$Glazing.Area)

levels(energy.efficiency$Glazing.Area) <- c("0%", "10%", "25%", "40%")

```

Lets look at the summary of the energy.efficiency data.

```
summary(energy.efficiency)
```

```

## Relative.Compactness Surface.Area Wall.Area Roof.Area
## Min. :0.6200 Min. :514.5 Min. :245.0 Min. :110.2
## 1st Qu.:0.6825 1st Qu.:606.4 1st Qu.:294.0 1st Qu.:140.9
## Median :0.7500 Median :673.8 Median :318.5 Median :183.8
## Mean :0.7642 Mean :671.7 Mean :318.5 Mean :176.6
## 3rd Qu.:0.8300 3rd Qu.:741.1 3rd Qu.:343.0 3rd Qu.:220.5
## Max. :0.9800 Max. :808.5 Max. :416.5 Max. :220.5
## Overall.Height Orientation Glazing.Area Glazing.Area.Distribution
## Min. :3.50 North:192 0% : 48 UnKnown: 48
## 1st Qu.:3.50 East :192 10%:240 Uniform:144
## Median :5.25 South:192 25%:240 North :144
## Mean :5.25 West :192 40%:240 East :144
## 3rd Qu.:7.00 South :144
## Max. :7.00 West :144
## Heating.Load Cooling.Load
## Min. : 6.01 Min. :10.90
## 1st Qu.:12.99 1st Qu.:15.62
## Median :18.95 Median :22.08
## Mean :22.31 Mean :24.59
## 3rd Qu.:31.67 3rd Qu.:33.13
## Max. :43.10 Max. :48.03

```

## Visualizations

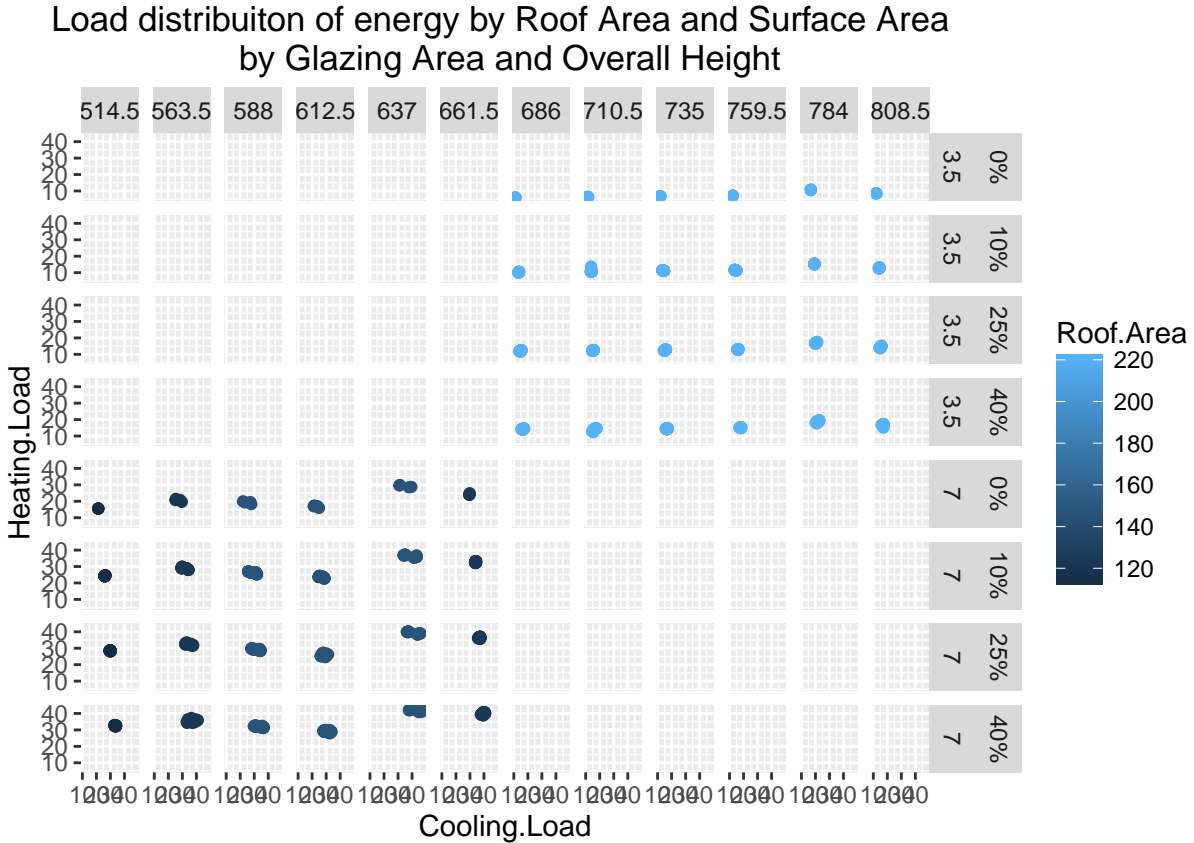
### Plot 1:

Lets visualize and find out if there is any relation between Roof Area, Surface Area and Glazing Area and how the Load is distributed using scatter plot.

```

ggplot(energy.efficiency, aes(x = Cooling.Load, y = Heating.Load), alpha = 0.3)+
  geom_point(aes(colour = Roof.Area ))+
  facet_grid(Overall.Height ~ Glazing.Area, space = "free") +
  ggtitle("Load distribuion of energy by Roof Area and Surface Area \n by Glazing Area and Overall Height")

```



#### Observations:

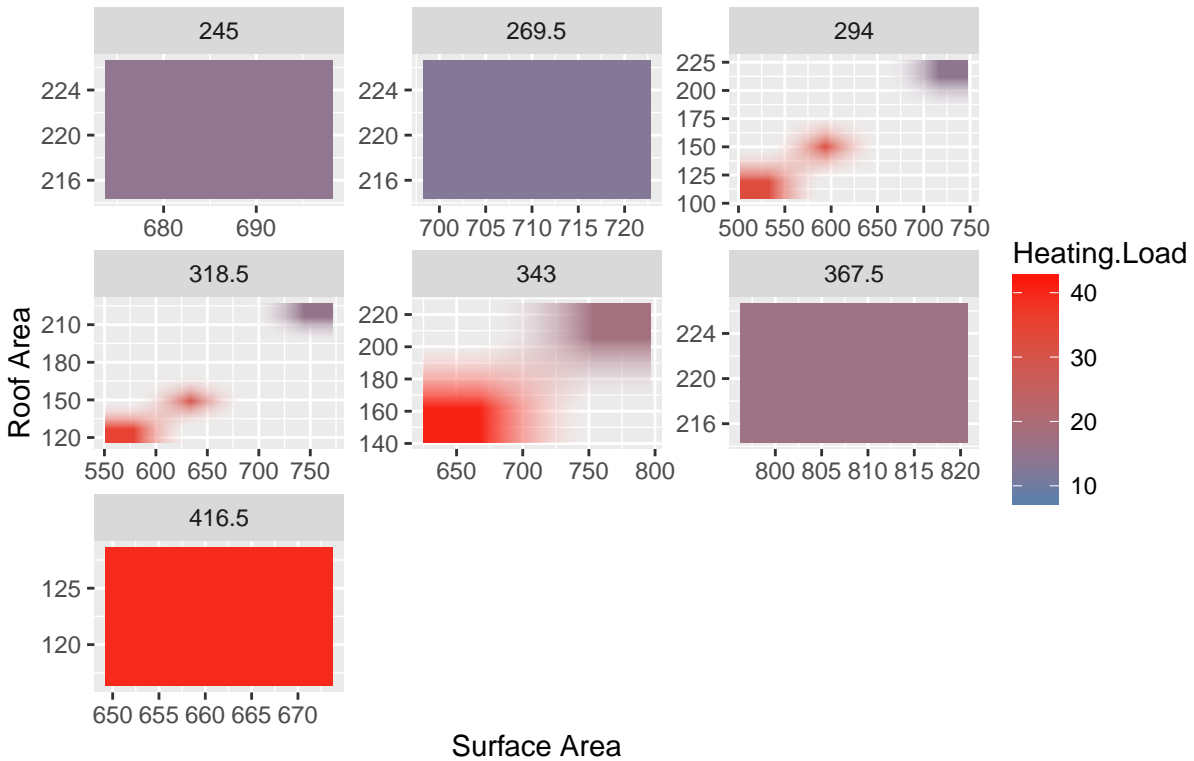
- Roof area and Surface area range is high for minimum/ lowest (3.5) over-all height and
- Roof area and Surface area range is low for maximum/ highest (7.0) over-all height.
- There are no data points when the overall height is 7 and the highest surface area range and also for low overall height 3.5, we have no data points with the low surface area range.

#### Plot 2(a)

Lets visualize again, how the Wall area is influencing heating load using raster plot.

```
ggplot(energy.efficiency, aes( Surface.Area, Roof.Area)) +
  geom_raster(aes(fill = Heating.Load), interpolate = TRUE) +
  scale_fill_gradient(low = "steelblue", high = "red")+
  facet_wrap(~Wall.Area, scales = "free" )+
  ggtitle('Measuring Heating Load distribution \n by Wall Area, Surface Area and Roof Area') +
  xlab('Surface Area') + ylab('Roof Area')
```

## Measuring Heating Load distribution by Wall Area, Surface Area and Roof Area



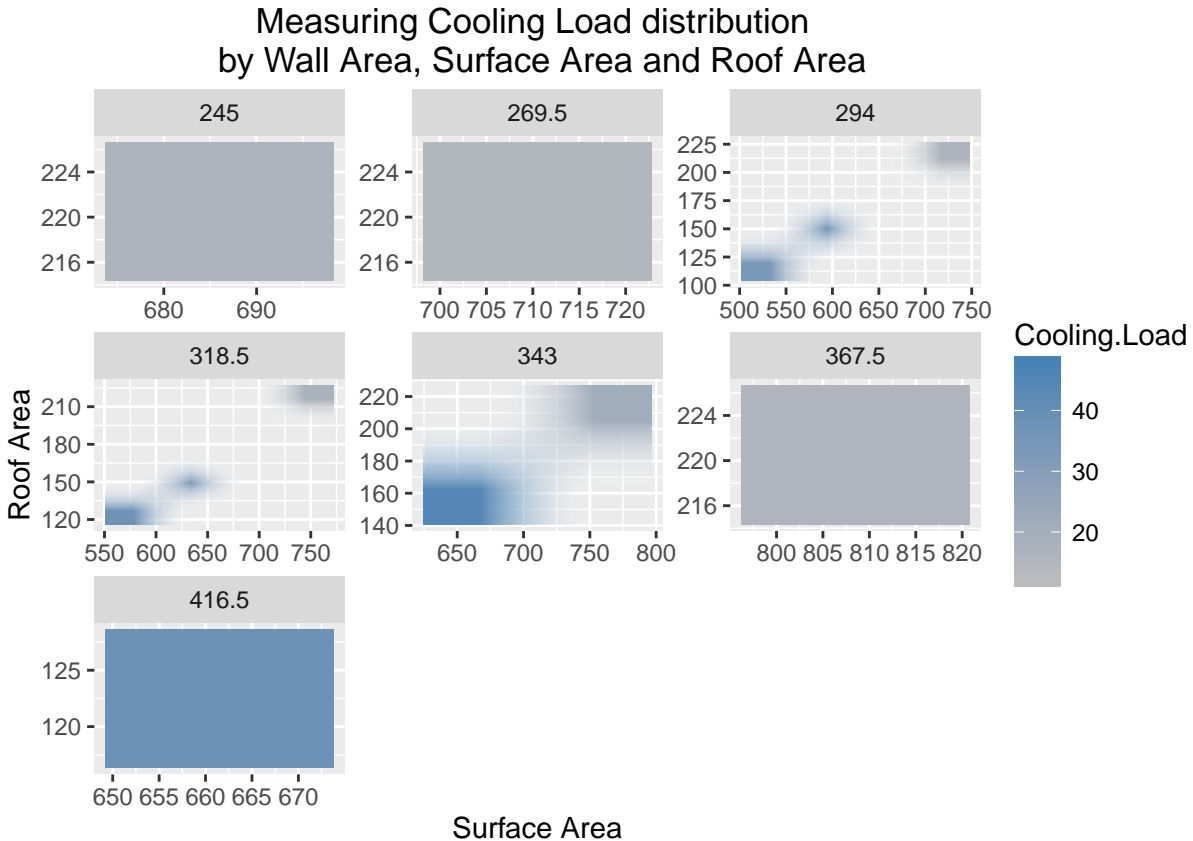
### Observations:

- By looking at the figure, we can conclude wall area is playing a significant role in heating load irrespective of Surface Area and Roof Area. (Ex: Higher the wall area, higher the heating load).

Similarly, let's figure out for the cooling load.

### Plot 2(b)

```
ggplot(energy. efficiency, aes(Surface.Area, Roof.Area)) +
  geom_raster(aes(fill = Cooling.Load), interpolate = TRUE) +
  scale_fill_gradient(low = "grey", high = "steelblue") +
  facet_wrap(~Wall.Area, scales = "free") +
  ggtitle('Measuring Cooling Load distribution \n by Wall Area, Surface Area and Roof Area') +
  xlab('Surface Area') + ylab('Roof Area')
```



So, Wall Area plays a significant role in both Heating and Cooling Load efficiency.

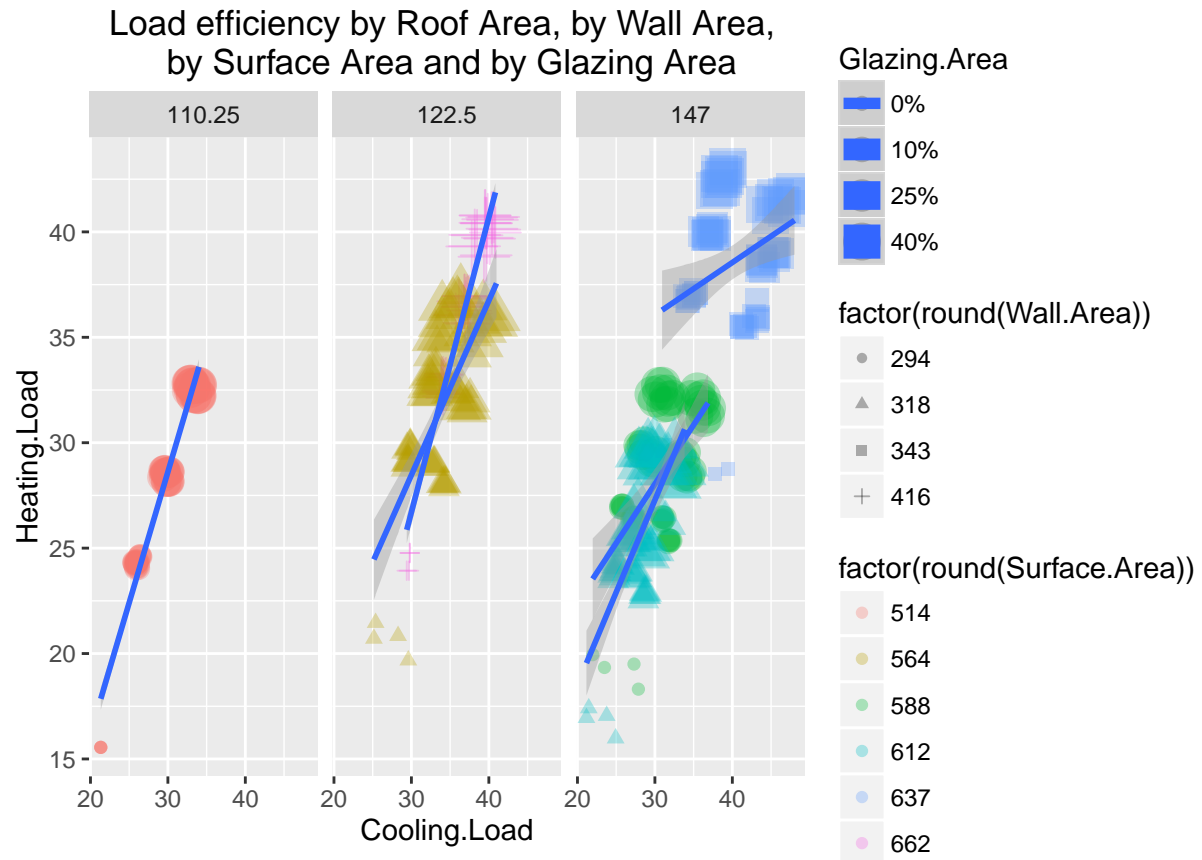
### Plot 3

We have seen more variation in load data when the overall height is (7.0). So let's create a subset named (energy.eff.sub7.0) which contains the filtered data with overall height = 7.0.

Let's visualize, if the Roof Area, Wall Area, Surface Area and Glazing Area are influencing the load efficiency.

```
energy.eff.sub7.0 <- energy.efficiency[ energy.efficiency$Overall.Height ==7.0,]
ggplot(energy.eff.sub7.0,
  aes(x = Cooling.Load, y = Heating.Load, group = factor(round(Wall.Area)),
    size = Glazing.Area,
    shape = factor(round(Wall.Area))))+
  geom_point(aes(colour= factor(round(Surface.Area))), alpha = 0.3)+
  geom_smooth(method = "lm", se = TRUE )+
  facet_grid(~ Roof.Area ) +
  ggtitle('Load efficiency by Roof Area, by Wall Area, \n by Surface Area and by Glazing Area')
```

```
## Warning: Using size for a discrete variable is not advised.
```



### Conclusion:

- It is clearly evident that the Load efficiency is influenced by the Wall Area, Roof Area, Glazing Area.
- When the Glazing Area is high, Roof Area is high and Wall Area is high, Load will be high and viceversa.