Lesson1_Assignment_EnergyEfficiency

Suman Nooney

January 18, 2017

Synopsis

This report is based on the dataset available at https://archive.ics.uci.edu/ml/datasets/Energy+efficiency#. Our interest in this data, is to find out the 3 interesting patterns how the Heating and Cooling Loads are impacted with the given 8 input variables.

Variable(s) Information:

- Relative Compactness
- Surface Area - m^2
- Wall Area - m^2
- Roof Area - 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 - kWh/m^2
- Cooling Load - kWh/m^2

Data staging

There is a cleansed dataset available from the Github at https://raw.githubusercontent.com/StephenElston/DataScience350/master/Lecture1/EnergyEfficiencyData.csv. Used it as a source, and below are the steps used in downloading and created a categorical variables for **"Orientation"**, **"Glazing Area Distribution (variance)"**.

```
rm(list = ls())

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

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

require(ggplot2)
```

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

```
#install.packages("gridExtra")
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
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.effiiency data.

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

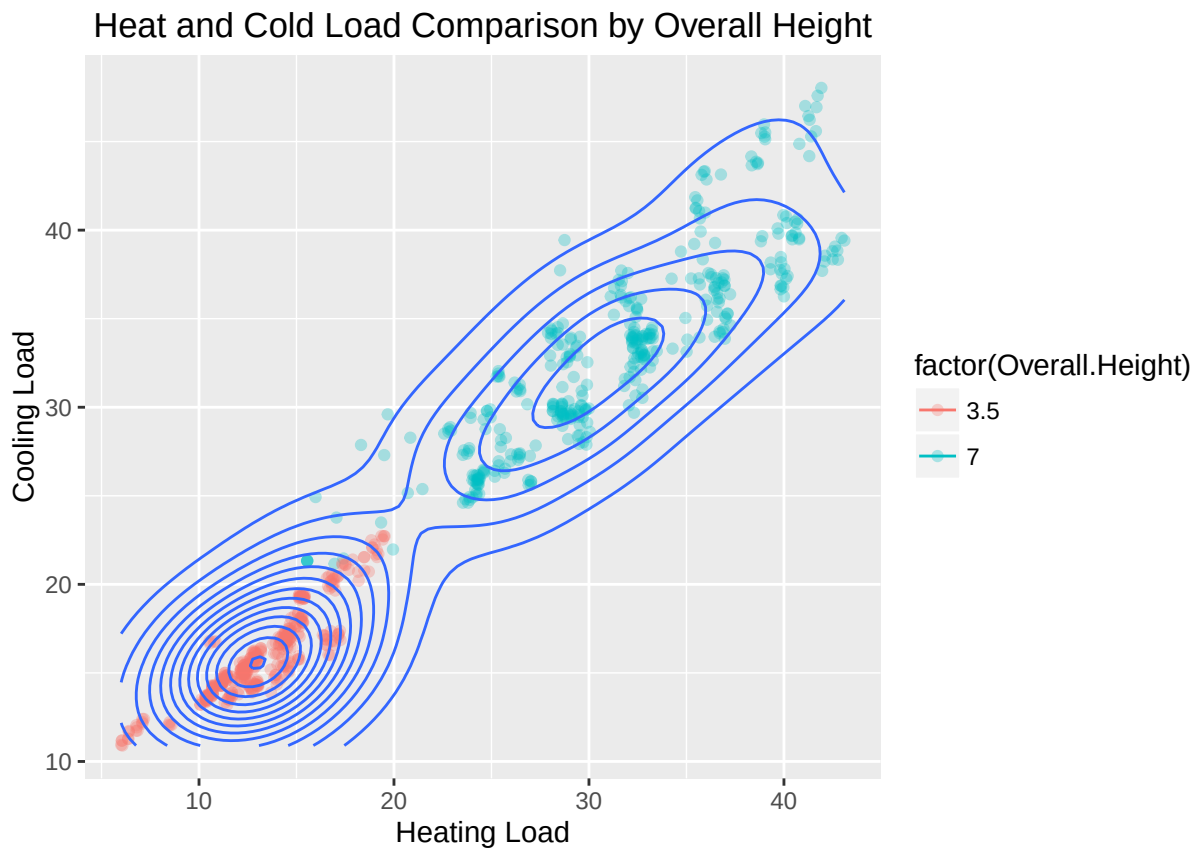
```
##  Relative.Compactness  Surface.Area     Wall.Area       Roof.Area
##  Min.   :0.6200       Min.   :514.5   Min.   :245.0   Min.   :110.2
##  1st Qu.:0.6825       1st Qu.:606.4   1st Qu.:294.0   1st Qu.:140.9
##  Median :0.7500       Median :673.8   Median :318.5   Median :183.8
##  Mean   :0.7642       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 how the overall height impacts the overall Cooling and Heating Load using density plot.

```
ggplot(energy.efficiency, aes(x = Heating.Load , y = Cooling.Load)) +
        geom_point( aes(col = factor(Overall.Height)), alpha= 0.3) +
        geom_density2d()+
        xlab('Heating Load') +
        ylab('Cooling Load') +
        ggtitle('Heat and Cold Load Comparison by Overall Height')
```

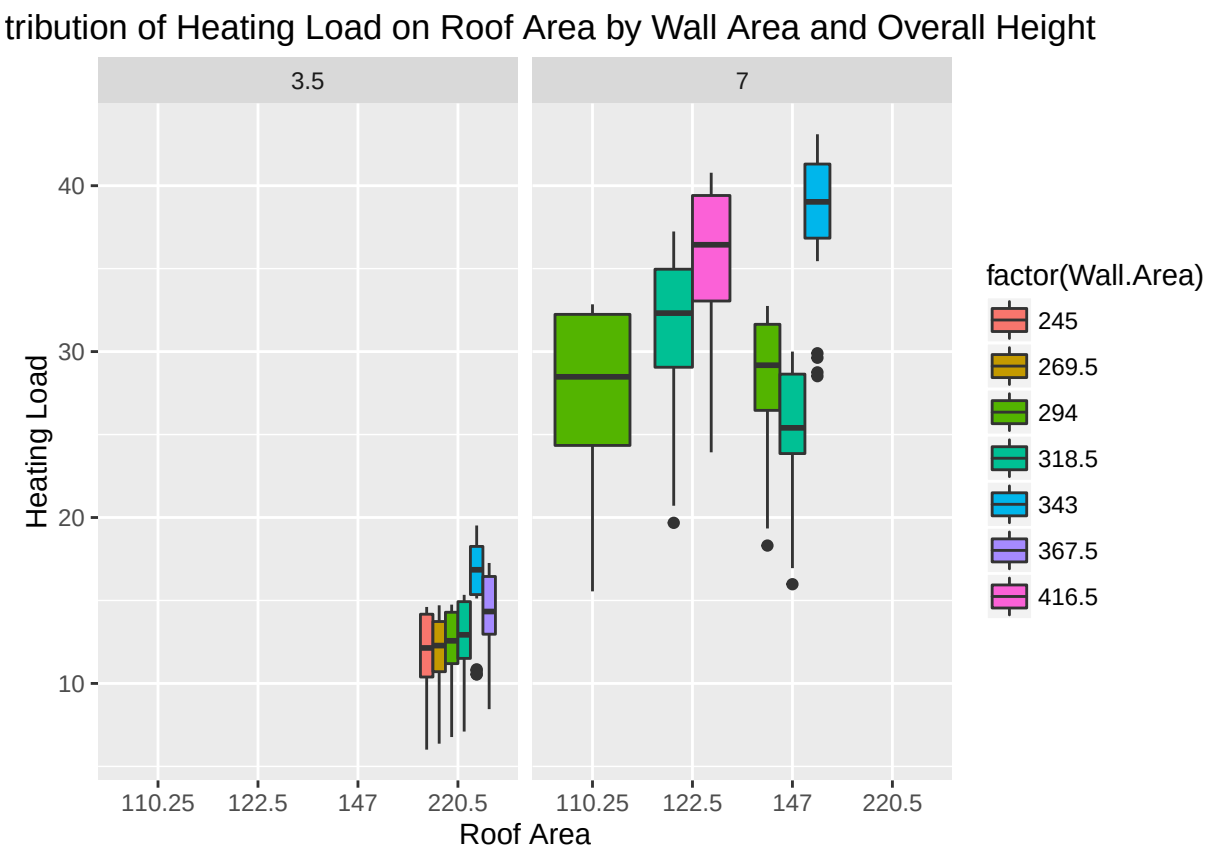


From the above plot, it is clear that the overall height plays a critical role in heating and cooling load.

Plot 2:

Lets visualize our second plot by how the roof and wall areas impacts the Heating Load using box plot.

```
ggplot(energy.efficiency, aes(x = factor(Roof.Area), y = Heating.Load,  group = Surface.Area)) +
        geom_boxplot(aes(fill = factor(Wall.Area))) +
        # geom_jitter(alpha= 0.3)+
        facet_grid(.~Overall.Height)+
        xlab('Roof Area') +
        ylab('Heating Load') +
        ggtitle('Distribution of Heating Load on Roof Area by Wall Area and Overall Height')
```

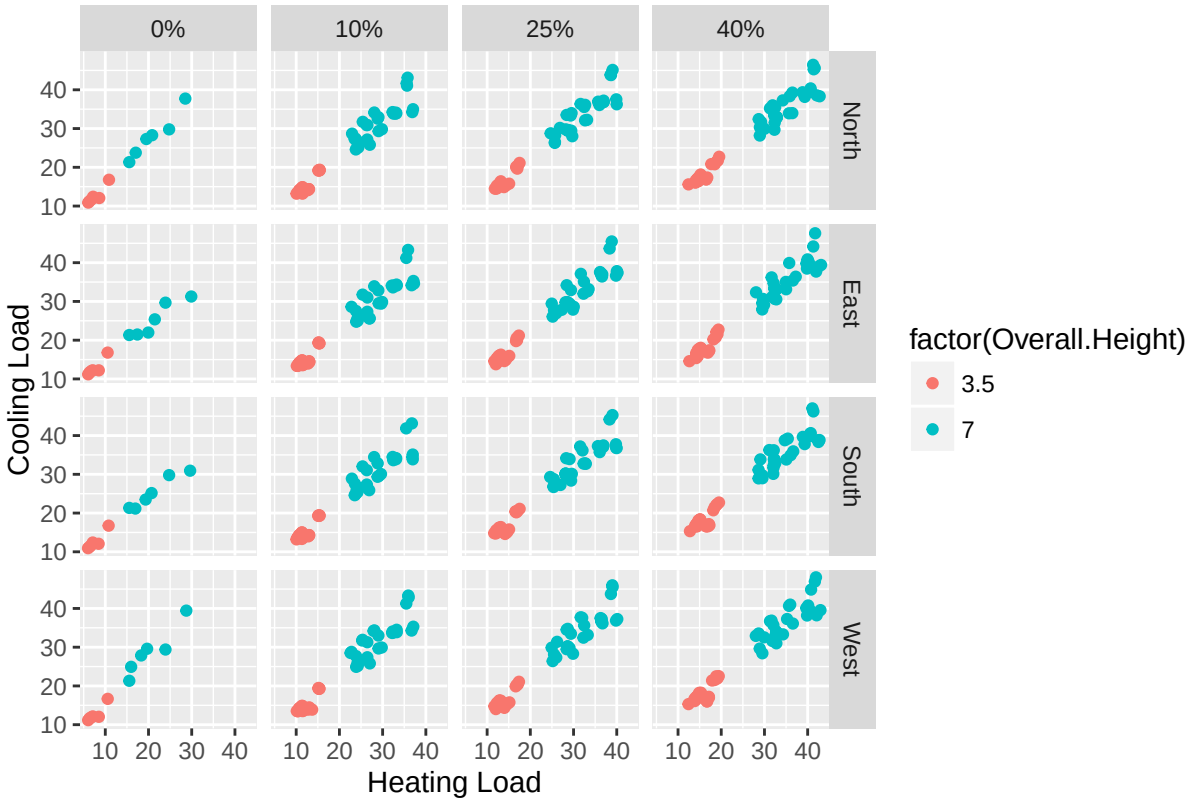


Plot: 3

Lets visualize the load distribution of Glazing Area by Orientation.

```
ggplot(energy.efficiency, aes(Heating.Load, Cooling.Load))+
  geom_point(aes(col = factor(Overall.Height)))+
  facet_grid(Orientation ~ Glazing.Area)+
  xlab('Heating Load') +
  ylab('Cooling Load') +
  ggtitle(' Heating and Cooling load distribution by Orientation and Glazing Area.')
```

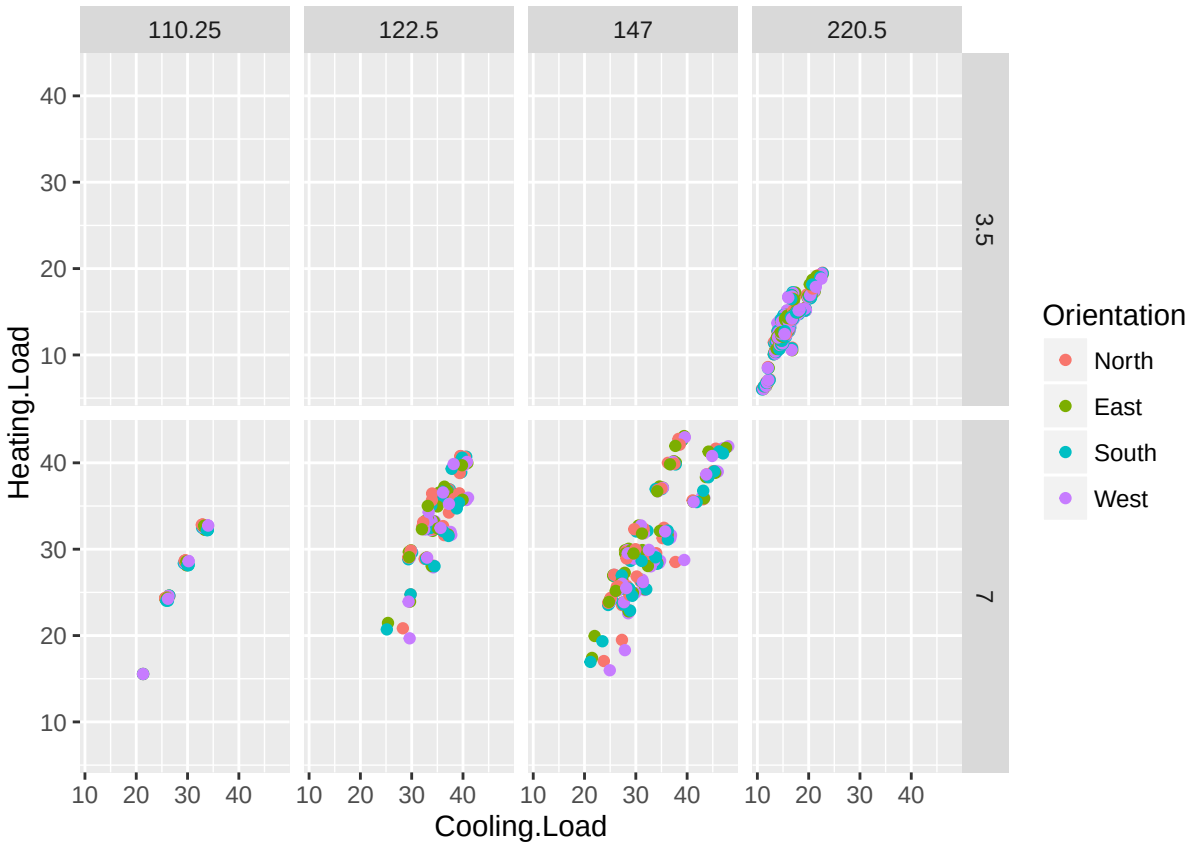
ting and Cooling load distribution by Orientation and Glazing Area.



Plot: 4

Lets visualize the Cooling and Heating load distribution by Orientation and Roof Area.

```
ggplot(energy.efficiency, aes(x = Cooling.Load, y = Heating.Load))+
        geom_point(aes(colour= Orientation))+
        facet_grid(Overall.Height ~ Roof.Area)
```



Conclusion:

From the above plots, we have clearly observed the Overall Height has a significant impact on overall heating and cooling load.