MAP 535 - Data Analysis Project

Adrien Toulouse & Paul-Antoine Girard

Introduction

Write a short introduction describing the research problem. Clearly state the research hypothesis at the end. We are given to analyse a dataset, named **House Prices: Advanced Regression Techniques**. It contains the selling price of about 1500 residential homes along with 79 explanatory variables describing (almost) every aspect of houses in Ames, Iowa. The dataset has already been preprocessed to deal with all the missing values, so ours records the houses prices along with 68 variables.

Clearly state the research hypothesis at the end.

We will start by doing a general description of the data and applying descriptive statistics to better apprehend the data. Then, we will build linear regression models and see if the asumptions are verified. To finally, present the best of them.

Exploratory Data Analysis

We start by uploading our data.

```
trainImputed <- read.csv(file='train_imputed.csv')
trainPP <- read.csv (file='train_preprocessed.csv')
```

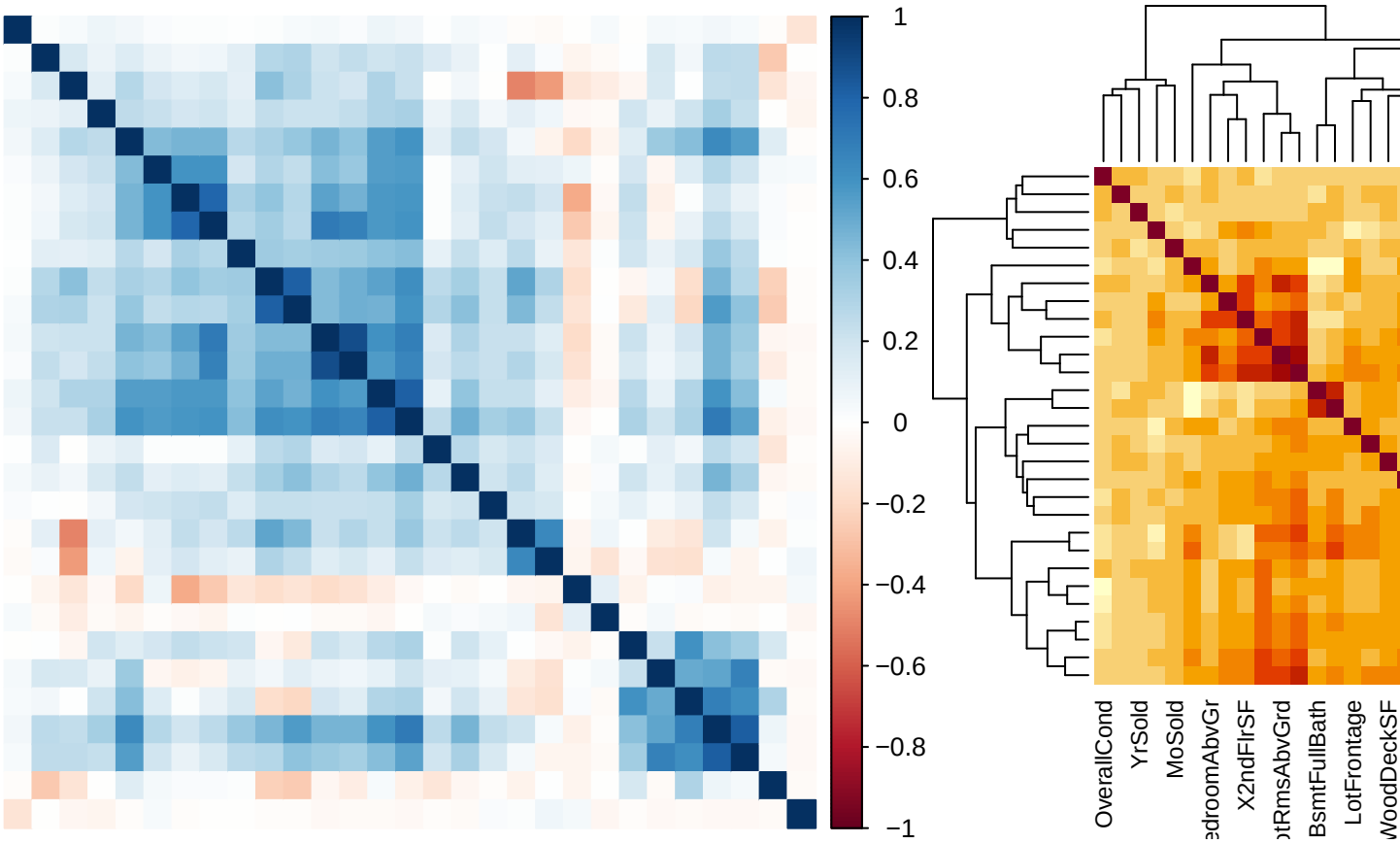
As we have seen in the presentation of the dataset, we can use the log of the house price to...

```
trainImputed$LogSalePrice <- log(trainImputed$SalePrice)
trainImputed <- select(trainImputed, -c("SalePrice", "X"))
trainPP$LogSalePrice <- log(trainPP$SalePrice)
trainPP <- select(trainPP, -c("SalePrice", "X")) #X is unrelated to our study
```

Since the dataset contains a lot of different variables, we need to focus only on a part of them. Thus, we will first analyze the correlation between the numerical ones, to create groups and reduce the size of our regressors.

```
var.numeric <- colnames(trainImputed)[sapply(trainImputed, is.numeric)]

trainImputed %>%
  select(var.numeric) %>%
  cor() %>%
  corrplot(method = 'color', order = "hclust", tl.pos = 'n') %>%
  heatmap (symm=T)
```




```
## Don't know how to print only the second graph
```

This graph shows us that the two variables related to the garage ('GarageCars' and 'GarageYrBlt') are hi

Modeling and Diagnostics