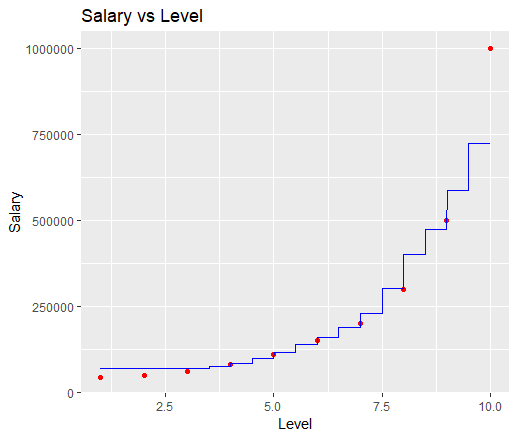
**Business problem**

Create a model that gives us the estimated salary for an employee based on the position.

**Explaining the model**

1. Import the dataset, which is in a csv format.
2. We don’t need to encode categorical data because we already have one column in our dataset that represents the position – we subset our data set, just to have one dependent variable and one independent variable.
3. Since it’s a dataset with only 10 rows, there’s not enough data to create a training and teste sets, therefore we will not split.
4. After installing the package randomForest, we create our regressor. The first argument of the rpart function is the column that has our independent variable, and since we need to use a vector in this argument, we transform our column in a vector. After that, we select our dependent variable, which is Salary, and then we choose the number of decision trees we want to implement on our random forest. Increasing the number of random trees will not increase infinitely the number of splits, but it will increase significantly the accuracy of the model. In our model, we’ve started with 10, but we finished with 500 trees, as we believe it is enough to predict an accurate result.
5. It’s now time to test our model. The plots will have to be a little different than the previous ones, as we need more definition to see the vertical splits. Because of that, we create a vector and use a sequence of 0.001 units. This vector will then be used as the dataset of our regressor’s line.

**Analyzing the results**

****

We can clearly see the splits, which were not possible using the same method we have been using on our previous models (we would see diagonals instead of vertical lines).

**Prediction**

Imagining a person has 2 years of experience as a Region Manager (level 6) and needs 2 more to jump to a Partner (7), we can consider it to be a 6.5 regarding the Level. Therefore, our model predicts a salary of $160908, against the $250.000 of the Decision Tree model, the $177.871 of the Support Vector Model and the $174.878 of the Polynomial model.