Linear Regression **PM2.5 ~ temperature**

* Built the model using data from the April 2014
* Tested it against the first 7 days of May 2014

Correlation Coefficient between the variables **0.05189936 BAD**

( A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. )

p-value of the linear regression **0.1642 BAD**

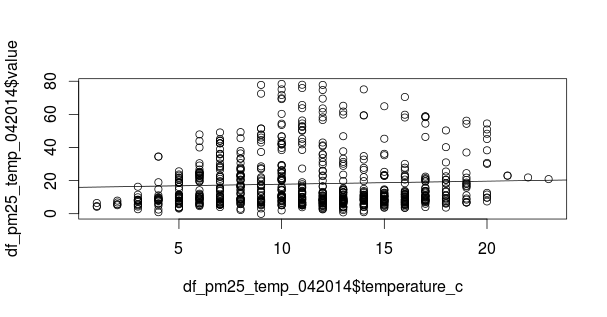
(y = a + βx + err

As the p-value is much more than 0.05, we cannot reject the null hypothesis that β = 0. Hence there is not sufficient evidence that there is a significant relationship between the variables in the linear regression model of the data.)

Goodness-of-fit of the model, p-value **0.001033 BAD**

(We reject the hypothesis that model is a good fit for our values)

* this is not the most appropriate test as it is used for categorical data
* there is a goodness-of-fit test for linear regression, need to look into it.



**Script link:** <https://github.com/sampayne/COMP3001/blob/master/implementation/RScripts/linerregpm25temp.R>

**Sources:**

1. Linear Regression with R <http://courses.statistics.com/software/R/R_Ch02.htm>
2. Chi-squared test for goodness-of-fit (although not the best option, works best with categorical data - here the ‘categorical’ one is temperature, always an integer between 0-25.

<http://www.math.csi.cuny.edu/~poje/Teach/Computer/Chi2.pdf>

<http://stackoverflow.com/questions/11408357/chi-squared-goodness-of-fit-test-in-r>

<https://en.wikipedia.org/wiki/Goodness_of_fit>

1. Correlation Coefficient <http://mathbits.com/MathBits/TISection/Statistics2/correlation.htm>
2. Plotting in R: <https://flowingdata.com/2012/12/17/getting-started-with-charts-in-r/>

<http://www.harding.edu/fmccown/r/>