**Quantitative Structure-Property Relationships Excel Activity**

1. Open the descriptor\_data.csv file that you created using your jupyter notebook in Microsoft Excel.
2. Start by making scatter plots of:
   1. Molecular weight vs Boiling point in Kelvin
   2. Wiener Index vs Boiling point in Kelvin
   3. Zagreb M1 vs Boiling point in Kelvin
   4. Zagreb M2 vs Boiling point in Kelvin

You may do each as a separate new sheet or as a region in your active sheet.

1. Add trendlines to each of your plots. Choose the best model for each. Some may be linear fit, some may be exponential or logarithmic. Display the equation and correlation coefficient on each graph.
   1. Do any of the graphs have a good correlation or predictive value?
   2. Did you have to exclude any data to make a trendline that was logarithmic?
2. View the following youtube video on performing multiple linear regression in Excel. The video shows a business example, but the process is the same for us.

<https://www.youtube.com/watch?v=HgfHefwK7VQ>

1. Since none of the individual properties provided excellent predictive value, let’s try to make a multiple linear regression to see if we can make a better predictive line. Using the method from the video create a multiple linear regression using all four descriptors. If any descriptor results in a p value of >0.05 remove it from the analysis and repeat.
2. Using the coefficients for each descriptor and intercept create an equation.

Y = 55.48788 + MW\*4.411439 +Wiener \*-0.63972 + Z1\* -4.02596

1. Create a copy of your descriptor data sheet and remove any graphs you added. Create a new column for predicted boiling points using the equation.
2. Make a new plot of predicted vs actual boiling point. Add a trendline. The better your equation, the closer to R2 = 1.00 you will be.
3. Using this equation, predict the boiling point of a molecule not in your training set. For example, 2-methylheptane is not in the training set. It has the following data:

MW = 114.232

Wiener Index = Determine using Jupyter notebook (79)

Z1 = Determine using Jupyter notebook (28)

Boiling Point = 390.6 K

For this analysis, predicted BP is 396.2K and actual is 390.6K. This is a result of 1.4% error.

**Individual Homework:**

* Using the given jupyter notebook and the original BP.csv file, you will create a new Excel spreadsheet with spreadsheet with topological and other calculated descriptors from Mordred.
* You will choose four new descriptors found in Mordred <http://mordred-descriptor.github.io/documentation/master/api/modules.html>
* Complete simple linear analysis for each of your new descriptors.
* Complete a multiple linear regression to create an equation that best represents the data boiling point data and your descriptors.
* Create a separate sheet that has your regression data.
* Make a plot of Actual vs Predicted BP for your regression.
* Choose a new molecule not in the dataset (not 2-methylheptane, be creative and use chemical intuition).
* Use your multiple linear equation to predict this molecule’s BP and look of the literature value.
* Write a short one-two page paper that includes:
  + What your descriptors mean
  + Which descriptors correlate
  + What is the overall equation calculated
  + How to choose the molecule to test.
  + How close this multiple linear regression predicts your boiling point of your molecule