1. **ChemSketch (15 minutes)**

Find a molecule of interest in PubChem use Cannonical SMILES or InChi ID and use ChemSketch to see the molecule.

1. **RStudio (30 minutes)**
2. Create two vectors of length 100. x & y are numbers from 1 to 100 and to this add a little bit of noise. Set a seed (please use 100 as seed) for gettting the same result before generating the noise vector. Hint: we can use random numbers for mimicking noise
3. calculate mean, median and sd for x and y
4. Explore x and y using histogram
5. Plot x and y
6. Build a suitable model based on step d
7. Look at the model parameters
8. Plot the predicted model, y\_hat, along with the x and y
9. **Preparation for Edirect (30 minutes)**
   1. Make sure you can log into SLURM cluster (IP: 144.175.88.21)
   2. Linux important commands (15 minutes) <https://github.com/hoodcollege/BIFX550_Spring2020/blob/master/C2/4-SR-Linuxtutorial.pdf>
   3. Why Shell programming:
      1. How to count the number of unresolved genome positions in a chromosome (say Chromosome 22)?
      2. Demonstrate how easy to answer this question using SHELL programming
10. **Break (5 minutes)**
11. **E-Direct (30 minutes)**

Motivation for e-direct tool. Compare Edirect with PubMed

(we are going to work on edirect and NCBI pubmed; open a browser tab and SLURM login window and put them side-by-side)

* 1. PubMed: clear all the filters
  2. PubMed: search for the keyword opsin gene conversion
  3. Edirect: run the following command

esearch -db pubmed -query "opsin gene conversion"

Look for the counts and compare them with the websearch

* 1. PubMed: Look for related articles

Find Related Data --> Choose DB: PubMed and option "similar articles"

Note the count (hits)

* 1. Edirect: run the following command to convince they are producing the same results

esearch -db pubmed -query "opsin gene conversion" | elink -related

* 1. PubMed: Add a filter word tetrachromacy
  2. Pubmed --> Advanced --> Choose the previous query and add additional queries like filtering. You can accomplish using the following way. In the filer box, choose

"Text Word" and enter tetrachromacy and search

* 1. PubMed: Note the count
  2. Edirect:

esearch -db pubmed -query "opsin gene conversion" | esearch -db pubmed -query "opsin gene conversion" | elink -related | efilter -query "tetrachromacy"

* 1. PubMed: Change the output format as “Abstract”
  2. Edirect: You can accomplish all the PubMed query steps using this one line code:

esearch -db pubmed -query "opsin gene conversion" | esearch -db pubmed -query "opsin gene conversion" | elink -related | efilter -query "tetrachromacy" | efetch -format abstract

Replace a query that is relevant for your gene and repeat the Edirect exercise.