1. **ChemSketch (15 minutes)**
   Find a molecule of interest in PubChem use Cannonical SMILES or InChi ID and use ChemSketch to see the molecule.

2. **RStudio (30 minutes)**
   1. 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
   2. calculate mean, median and sd for x and y
   3. Explore x and y using histogram
   4. Plot x and y
   5. Build a suitable model based on step d
   6. Look at the model parameters
   7. Plot the predicted model, y_hat, along with the x and y

3. **Preparation for Edirect (30 minutes)**
   a. Make sure you can log into SLURM cluster (IP: 144.175.88.21)
   b. Linux important commands (15 minutes)
      https://github.com/hoodcollege/BIFX550_Spring2020/blob/master/C2/4-SR-Linuxtutorial.pdf
   c. Why Shell programming:
      i. How to count the number of unresolved genome positions in a chromosome (say Chromosome 22)?
      ii. Demonstrate how easy to answer this question using SHELL programming

4. **Break (5 minutes)**
5. **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)
   a. PubMed: clear all the filters
   b. PubMed: search for the keyword opsin gene conversion
   c. Edirect: run the following command
      esearch -db pubmed -query "opsin gene conversion"
      Look for the counts and compare them with the websearch
   d. PubMed: Look for related articles
      Find Related Data --> Choose DB: PubMed and option "similar articles"
      Note the count (hits)
   e. Edirect: run the following command to convince they are producing the same results
      esearch -db pubmed -query "opsin gene conversion" | elink -related
   f. PubMed: Add a filter word tetrachromacy
   g. 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

h. PubMed: Note the count
i. Edirect:
esearch -db pubmed -query "opsin gene conversion" | esearch -db pubmed -query "opsin gene conversion" | elink -related | efilter -query "tetrachromacy"

j. PubMed: Change the output format as “Abstract”
k. 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.