

Drugs

Isolating the Drugs

```
In[524]:= Drugs = EntityList[drugs CHEMICALS];  
  
In[525]:= missingACs2 = Select[Drugs, MissingQ[#[ "AtomCount"]]&];  
  
In[526]:= drugsWithAc = Complement[Drugs, missingACs2];  
  
In[527]:= Length[drugsWithAc]  
Out[527]= 4475  
  
In[528]:= missingBps2 = Select[drugsWithAc, MissingQ[#[ "BoilingPoint"]]&];  
  
In[529]:= drugsWithBP = Complement[drugsWithAc, missingBps2];  
  
In[530]:= Length[drugsWithBP]  
Out[530]= 682  
  
In[532]:= drugMolecules = ParallelMap[Molecule, drugsWithBP];
```

Descriptors

```
In[533]:= SmilesDrugs = ParallelMap[#[ "SMILES"]&, drugsWithBP];  
  
In[534]:= BpDrugs = ParallelMap[#[ "BoilingPoint"]&, drugsWithBP];  
  
In[535]:= acDrugs = ParallelMap[#[ "AtomCount"]&, drugsWithBP];  
  
In[536]:= mmDrugs = ParallelMap[MoleculeValue[#, "MolecularMass"]&, drugMolecules];  
  
In[537]:= rgDrugs = ParallelMap[MoleculeValue[#, "RadiusOfGyration"]&, drugMolecules];  
  
In[538]:= cnDrugs = ParallelMap[MoleculeValue[#, "HeteroatomCount"]&, drugMolecules];  
  
In[539]:= KhasDrugs = ParallelMap[MoleculeValue[#, "KierHallAlphaShape"]&, drugMolecules];  
  
In[540]:= LaSaDrugs =  
          ParallelMap[MoleculeValue[#, "LabuteApproximateSurfaceArea"]&, drugMolecules];  
  
In[541]:= SasDrugs =  
          ParallelMap[MoleculeValue[#, "SyntheticAccessibilityScore"]&, drugMolecules];  
  
In[542]:= PobfDrugs =  
          ParallelMap[MoleculeValue[#, "PlaneOfBestFitDistance"]&, drugMolecules];
```

Making a dataset

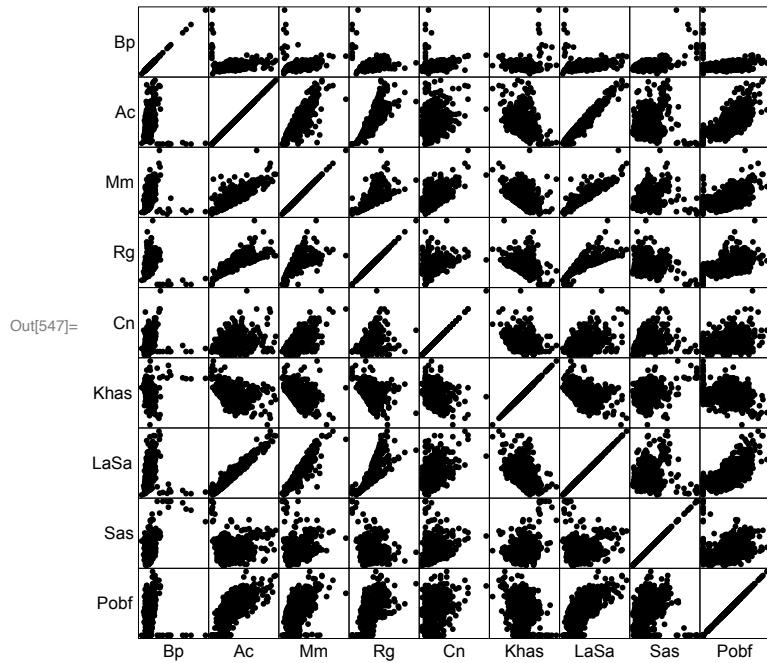
```
In[543]:= measurementsDrugs =
  Transpose[{SmilesDrugs, QuantityMagnitude[BpDrugs], QuantityMagnitude[acDrugs],
  QuantityMagnitude[mmDrugs], QuantityMagnitude[rgDrugs], QuantityMagnitude[
  cnDrugs], QuantityMagnitude[KhasDrugs], QuantityMagnitude[LaSaDrugs],
  QuantityMagnitude[SasDrugs], QuantityMagnitude[PobfDrugs]}];

In[544]:= variableNamesDrugs =
 {"Smiles", "Bp", "Ac", "Mm", "Rg", "Cn", "Khas", "LaSa", "Sas", "Pobf"};

In[545]:= drugData =
 ResourceFunction["DatasetWithHeaders"] [measurementsDrugs, variableNamesDrugs]

In[546]:= dataFitting2 = Values[Normal[
 drugData[All, {"Bp", "Ac", "Mm", "Rg", "Cn", "Khas", "LaSa", "Sas", "Pobf"}]]];

In[547]:= Graph2 = ResourceFunction["PairwiseScatterPlot"] [dataFitting2,
 "DataLabels" → {"Bp", "Ac", "Mm", "Rg", "Cn", "Khas", "LaSa", "Sas", "Pobf"}]
```



Pearson Correlation

```
In[548]:= PearsonCorrelationTest[BpDrugs, LaSaDrugs, "TestStatistic"]

Out[548]= 0.236214
```

```
In[549]:= PearsonCorrelationTest[acDrugs, LaSaDrugs, "TestStatistic"]
Out[549]= 0.973624

In[550]:= PearsonCorrelationTest[mmDrugs, LaSaDrugs, "TestStatistic"]
Out[550]= 0.929536

In[551]:= PearsonCorrelationTest[rgDrugs, LaSaDrugs, "TestStatistic"]
Out[551]= 0.843196

In[552]:= PearsonCorrelationTest[cnDrugs, LaSaDrugs, "TestStatistic"]
Out[552]= 0.408882

In[553]:= PearsonCorrelationTest[KhasDrugs, LaSaDrugs, "TestStatistic"]
Out[553]= -0.47856

In[554]:= PearsonCorrelationTest[SasDrugs, LaSaDrugs, "TestStatistic"]
Out[554]= 0.177282

In[555]:= PearsonCorrelationTest[PobfDrugs, LaSaDrugs, "TestStatistic"]
Out[555]= 0.691775
```

Machine Learning Functions

```
In[771]:= predictorVars2 = Normal@Values@drugData[All, {"Smiles", "Ac", "Mm", "Rg", "Pobf"}];
In[772]:= outputValue2 = Normal@drugData[All, "LaSa"];
In[773]:= fullDataSetPredict2 = MapThread[Rule, {predictorVars2, outputValue2}];
In[774]:= fullTrainingSet2 =
  RandomSample[fullDataSetPredict2, Round[0.8 * Length[fullDataSetPredict2]]];
In[775]:= fullTestingSet2 = Complement[fullDataSetPredict2, fullTrainingSet2];
In[776]:= trainingSet2Values = Rest /@ Keys[fullTrainingSet2];
In[777]:= trainingSet2Output = Values[fullTrainingSet2];
In[778]:= trainingSet2 = MapThread[Rule, {trainingSet2Values, trainingSet2Output}];
In[779]:= testingSet2Values = Rest /@ Keys[fullTestingSet2];
In[780]:= testingSet2Output = Values[fullTestingSet2];
In[781]:= testingSet2 = MapThread[Rule, {testingSet2Values, testingSet2Output}];
```

Linear Regression

```
In[782]:= pred1p = Predict[trainingSet2, Method → "LinearRegression"]
```

```
In[783]:= pm1p = PredictorMeasurements[pred1p, testingSet2]
In[1260]:= Dataset[AssociationMap[pm1p[#, ComputeUncertainty → True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

Out[1260]=	<table border="1"> <tr> <td>StandardDeviation</td><td>3.55 ± 0.26</td></tr> <tr> <td>RSquared</td><td>0.9950 ± 0.0011</td></tr> <tr> <td>EvaluationTime</td><td>0.0018</td></tr> </table>	StandardDeviation	3.55 ± 0.26	RSquared	0.9950 ± 0.0011	EvaluationTime	0.0018
StandardDeviation	3.55 ± 0.26						
RSquared	0.9950 ± 0.0011						
EvaluationTime	0.0018						

```
In[785]:= PredictorMeasurements[pred1p, testingSet2, "ComparisonPlot"]
```

Neural Network

```
In[788]:= pred2p = Predict[trainingSet2, Method → "NeuralNetwork"]
In[789]:= pm2p = PredictorMeasurements[pred2p, testingSet2]
In[1261]:= Dataset[AssociationMap[pm2p[#, ComputeUncertainty → True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

Out[1261]=	<table border="1"> <tr> <td>StandardDeviation</td><td>2.83 ± 0.18</td></tr> <tr> <td>RSquared</td><td>0.9968 ± 0.0007</td></tr> <tr> <td>EvaluationTime</td><td>0.0024</td></tr> </table>	StandardDeviation	2.83 ± 0.18	RSquared	0.9968 ± 0.0007	EvaluationTime	0.0024
StandardDeviation	2.83 ± 0.18						
RSquared	0.9968 ± 0.0007						
EvaluationTime	0.0024						

```
In[1307]:= y = PredictorMeasurements[pred2p, testingSet2, "ComparisonPlot"]
```

Gradient Boosted Trees

```
In[794]:= pred3p = Predict[trainingSet2, Method → "GradientBoostedTrees"]
In[795]:= pm3p = PredictorMeasurements[pred3p, testingSet2]
In[1262]:= Dataset[AssociationMap[pm3p[#, ComputeUncertainty → True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

Out[1262]=	<table border="1"> <tr> <td>StandardDeviation</td><td>3.83 ± 0.33</td></tr> <tr> <td>RSquared</td><td>0.9941 ± 0.0014</td></tr> <tr> <td>EvaluationTime</td><td>0.0038</td></tr> </table>	StandardDeviation	3.83 ± 0.33	RSquared	0.9941 ± 0.0014	EvaluationTime	0.0038
StandardDeviation	3.83 ± 0.33						
RSquared	0.9941 ± 0.0014						
EvaluationTime	0.0038						

```
In[797]:= PredictorMeasurements[pred3p, testingSet2, "ComparisonPlot"]
```

Random Forest

```
In[800]:= pred4p = Predict[trainingSet2, Method -> "RandomForest"]
In[801]:= pm4p = PredictorMeasurements[pred4p, testingSet2]
In[1263]:= Dataset[AssociationMap[pm4p[#, ComputeUncertainty -> True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

StandardDeviation	10.5 ± 0.8
RSquared	0.956 ± 0.010
EvaluationTime	0.0038

Out[1263]=

```
In[803]:= PredictorMeasurements[pred4p, testingSet2, "ComparisonPlot"]
```

Decision Tree

```
In[806]:= pred5p = Predict[trainingSet2, Method -> "DecisionTree"]
In[807]:= pm5p = PredictorMeasurements[pred5p, testingSet2]
In[1290]:= Dataset[AssociationMap[pm5p[#, ComputeUncertainty -> True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

StandardDeviation	11.5 ± 0.7
RSquared	0.947 ± 0.011
EvaluationTime	0.0015

Out[1290]=

```
In[809]:= PredictorMeasurements[pred5p, testingSet2, "ComparisonPlot"]
```

Gaussian Process

```
In[812]:= pred6p = Predict[trainingSet2, Method -> "GaussianProcess"]
In[813]:= pm6p = PredictorMeasurements[pred6p, testingSet2]
In[1265]:= Dataset[AssociationMap[pm6p[#, ComputeUncertainty -> True] &,
 {"StandardDeviation", "RSquared", "EvaluationTime"}]]
```

StandardDeviation	2.9 ± 0.4
RSquared	0.9966 ± 0.0012
EvaluationTime	0.0043

Out[1265]=

```
In[1249]:= PredictorMeasurements[pred6p, testingSet2, "ComparisonPlot"]
```

Predicting Citalopram's Molecular Surface Area

```
In[1221]:= fullTestingSet2[[128]]
```

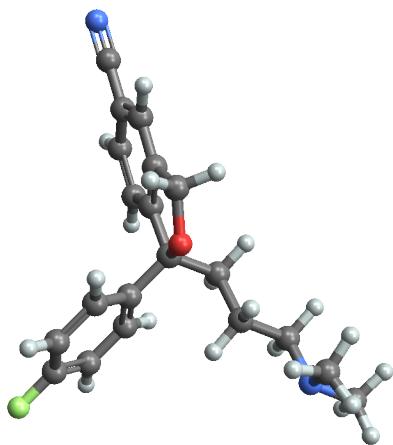
```
Out[1221]= {CN(C)CCCC1(C2=C(CO1)C=C(C=C2)C#N)C3=CC=C(C=C3)F,  
45, 324.399, 3.93421, 1.17359} → 171.338
```

```
In[1222]:= pred2p[Keys@testingSet2[[128]]]
```

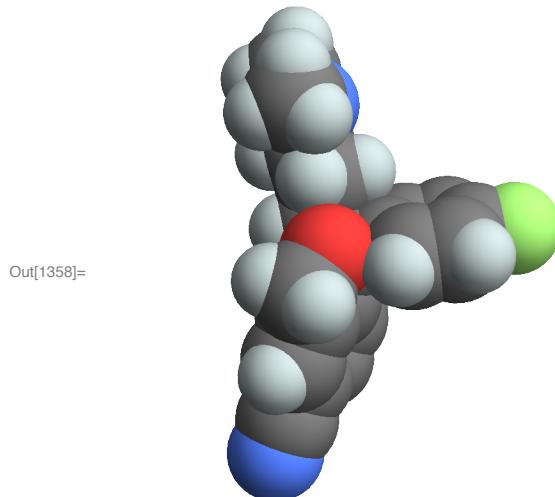
```
Out[1222]= 171.59
```

```
In[1357]:= ChemicalData["Citalopram", "MoleculePlot"]
```

```
Out[1357]=
```



```
In[1358]:= ChemicalData["Citalopram", "SpaceFillingMoleculePlot"]
```



Manipulate App

```
In[1546]:= pred2p[{45, 324.399, 3.93421, 1.17359}]
```

```
Out[1546]= 171.59
```

```
In[1547]:= pred2p[{{100, 100, 100, 0}, {20, 10, 100, 3}, {25, 23, 40, 0.2},  
{43, 129, 10, 33}, {12, 30, 52, 0.8}, {64, 12, 150, 0}, {50, 670, 1, 0.7},  
{4, 30, 10, 0.9}, {54, 264, 00, 5}, {10, 10, 100, 2}, {45, 23, 45, 0}}]
```

```
Out[1547]= {1203.32, 1224.88, 548.957, 762.331, 697.138,  
1826.79, 207.832, 153.031, 188.681, 1254.35, 593.033}
```

```
In[1545]:= Framed[Manipulate[ToString[  
  pred2p[{atomCount, molecularMass, radiusOfGyration, planeOfBestFitDistance}]],  
  Style["Drugs Surface Area Calculator", 14, Bold, Orange],  
  Delimiter, {{atomCount, 45, "Atom Count"}, 2, 120},  
  {{molecularMass, 324.399, "Molecular Mass"}, 26.038`, 822.942`},  
  {{radiusOfGyration, 3.9342068720153947`, "Radius of Gyration"},  
   0.22462`, 10.7024`}, {{planeOfBestFitDistance, 1.173585005030939`,  
   "Plane of Best Fit Distance"}, 0., 1.22643`}, ControlType -> InputField,  
  SaveDefinitions -> True, ControlPlacement -> Left, Paneled -> False], FrameMargins ->  
  5, Background -> Lighter[Gray, 0.9], RoundingRadius -> 7.5`]
```

Out[1545]=

Drugs Surface Area Calculator

Atom Count	45	
Molecular Mass	324.399	171.59
Radius of Gyration	3.93421	
Plane of Best Fit Distance	1.17359	