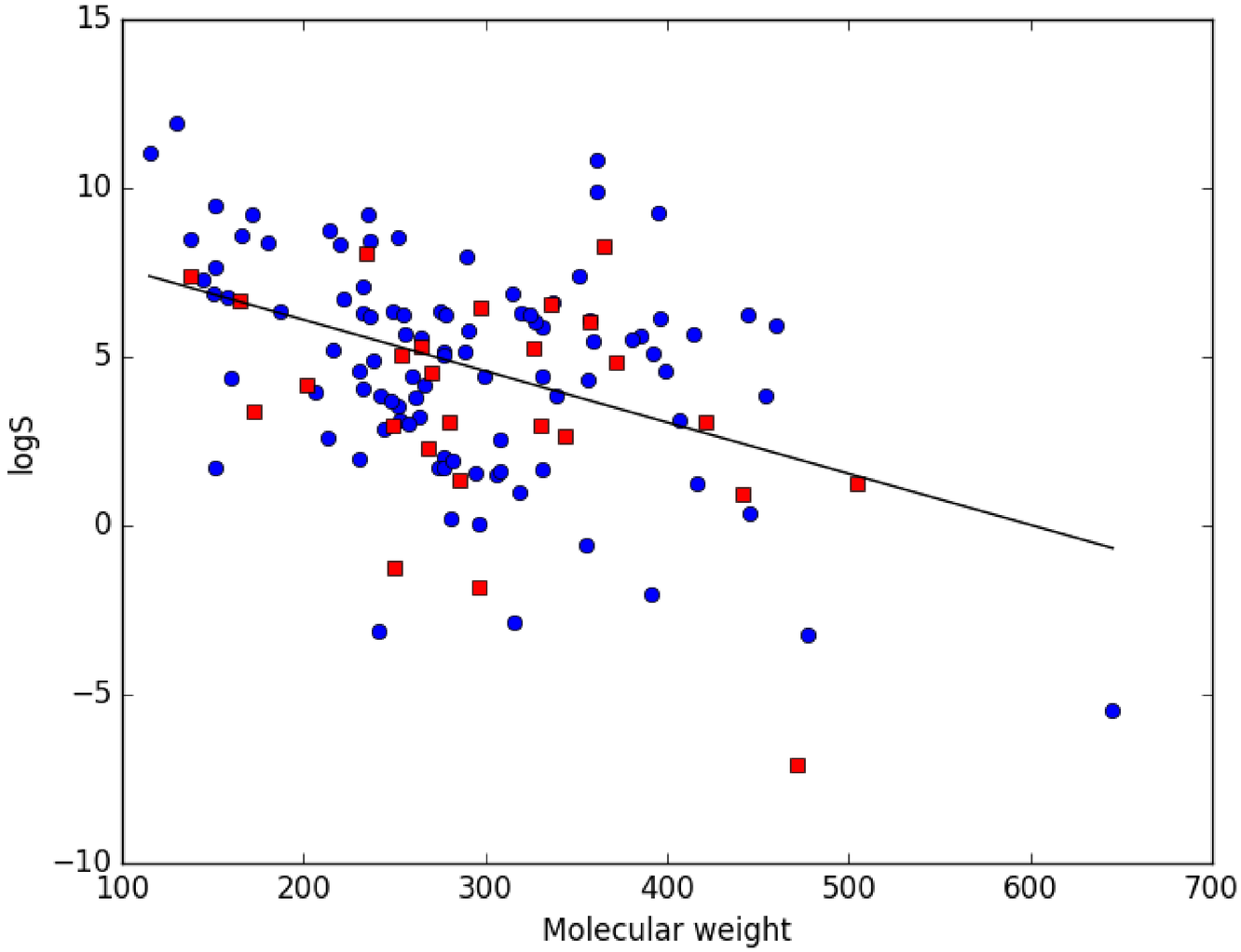
Molecular weight has some predictive power



From Ref [1]
Blue: training
Red: test
Black: predictions

QSPR in the solubility challenge

- Hewitt et al tried four QSPR models [5]
  - Multiple linear regression (426 descriptors, genetic algorithm, no more than 5 used at once)
  - Artificial neural network
  - Category-specific models based on H bond ability
  - Various commercial QSPR models
  - Consensus, mean of four models
- The best was a three-descriptor linear regression
  - based on log P, boiling point, and R maximal autocorrelation of lag 2 (related to size and connectivity).
  - $R^2_{train} = 0.74$, $R^2_{test} = 0.51$
- "none of the other modeling approaches used in this study was able to improve upon the predictions made by the MLR model"