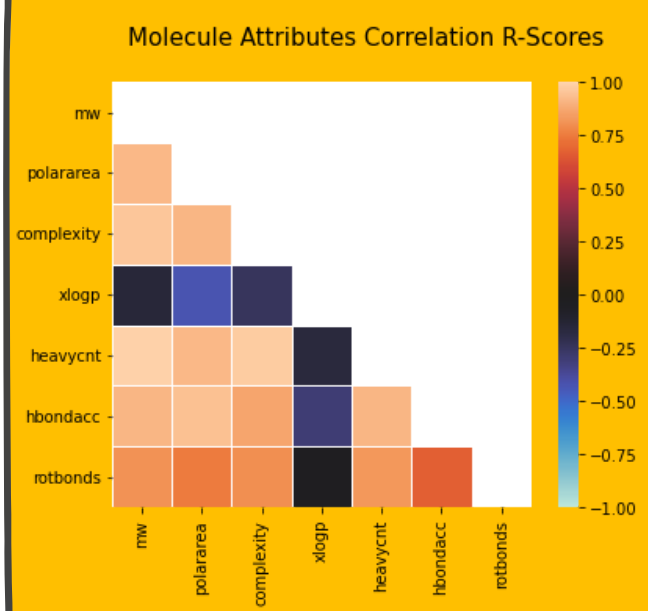
CHEMICAL COMPOUND COMPLEXITY

HOW TO RELATE IT TO OTHER CHEM STUFF USING A SIMPLE RELATIONSHIP

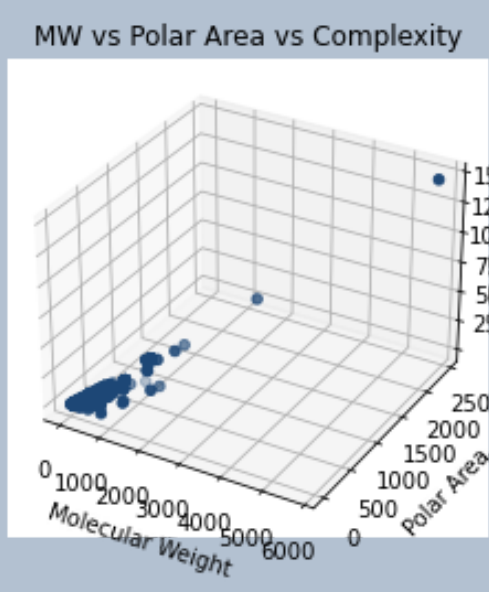


Just want to see what are the R-scores between different columns in the same dataset,

Pay special attention to complexity since it's the least defined amongst the others

SO let's use some intuitive attributes and with R scores close to ONE

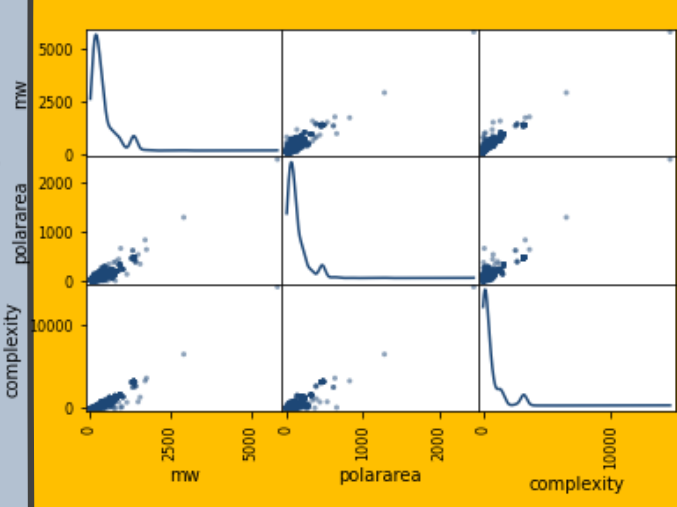
in 3D together



it's hard to tell... the few outliers makes the graph confusing

BUT we can still see some correlation

or not



the correlation is more clean when plotted separatedly

so using the .describe():

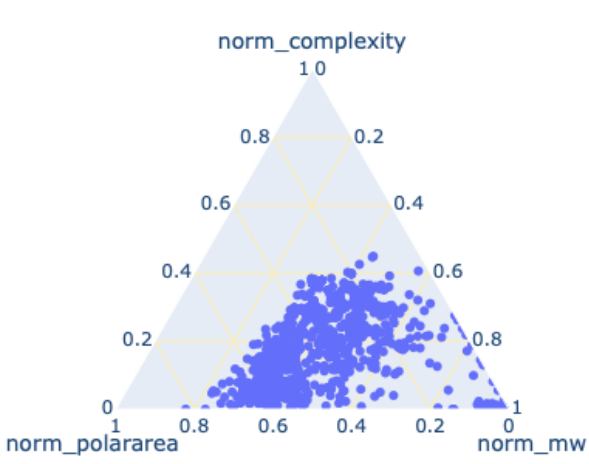
average molecular weight was 387g;

average polar area was 125;

average complexity was 600

which made sense of why the two plots above didn't make sense

although there is still a trend shown in the two plots



a Ternary plot with normalized values can show the trend more clearly---

Value of complexity can be a linear combo of weight and polar area!

Multiple Linear Regression Model:

complexity = 1.81*('mw')+1.32('polarara') -299

R-score: 0.920 (!!!, HIGHLY linearly related)