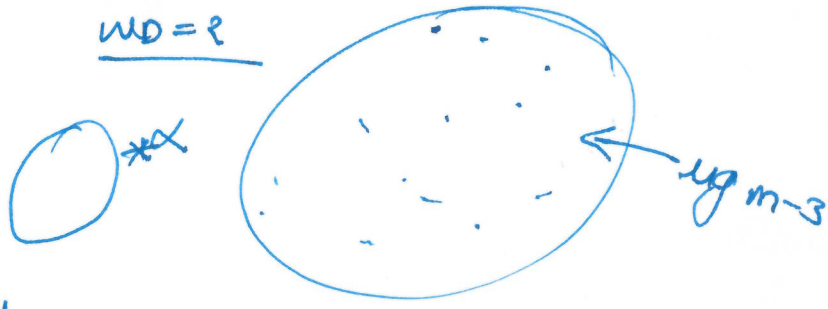
Exploratory analysis

WD = ?



1. Summary. (df)

~~1.1 Delete orig temp.~~

AT 1.2 Clean AT further. Del. 631.98. and on the lower side.

1.3 Delete ≥100 value in RH.

1.4 Keep RH: 20:80

1.5 Clean WS values $>30 ms^{-1}$ should be removed.

1.6. WD is in °

1.7. SR → remove "-ve"

1.8 BP → Make 0 as NA / Keep in 655 – 810

1.9 PM 2.5

1.9.1 Impute.

1.9.2 Real avg. from that day. ✓ four onwd. values.

1.9.3 Del. 3897.31

1.10 Del. Temp.

1.11 Code Events.

2. ~~AT~~ EDA.

2.1. Scatterplot matrix

2.2. Violin Plot – ??

2.3. Corr. Plot.

2.4 PCA Bi-Plot.

3. Models.

3.0 NULL

3.1. GLM

3.2 GAM (NL)

3.3 CART

3.4 MARS

3.5 RF ~~BR~~

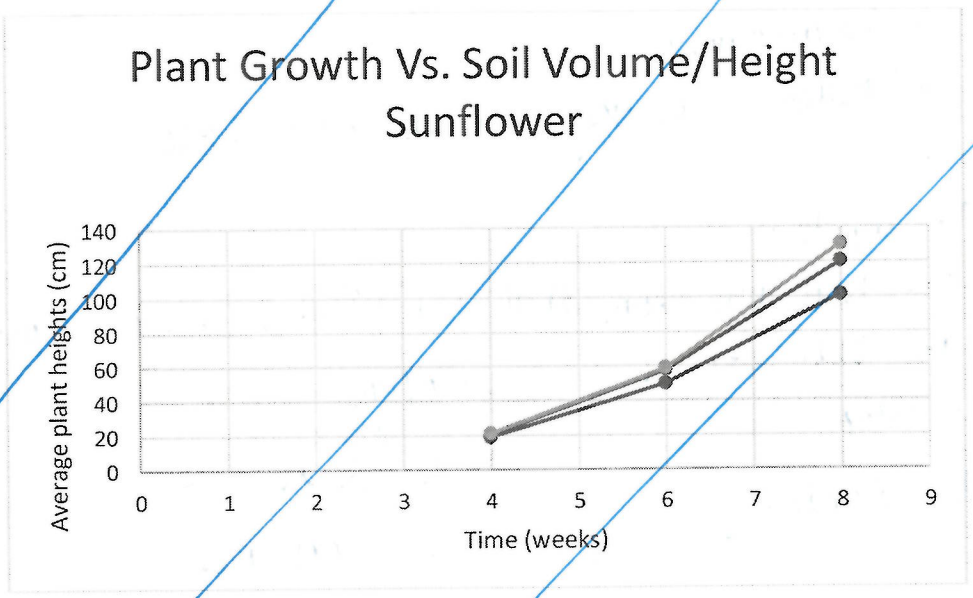
3.6 BART

3.7 SVM

3.8 NN – ??

2. **(8 points)** Prepare **XY Scatter Plots** that show how the plant height of each species was affected by the volume of soil occupied by the root system and how it changed over time. (Use a separate graph for each species. For a single species, use individual lines for each soil volume. Graphs should indicate **<u>average</u>** height [y-axis], soil volume, and time [x-axis]. Attach these graphs to your report.)

KEY
Blue = 3"
Red = 5"
Green = 8"



3.5

1. Resid plot.
   - ~~Resid vs Predfit~~ $\hat{e}$ e vs. $\hat{y}$
   - $\hat{y}$ vs. y
   - e vs. (Predictors)
   - Norm plot (e)

2. Variable Selection.

2.1 VarImp plot (where applicable)

2.2 Statistical Analysis of pred w & w/o ← Final run

2.3 Partial dependency. → Final.

3. Final ~~#~~ comparing all the CV RMSE OOB of models.