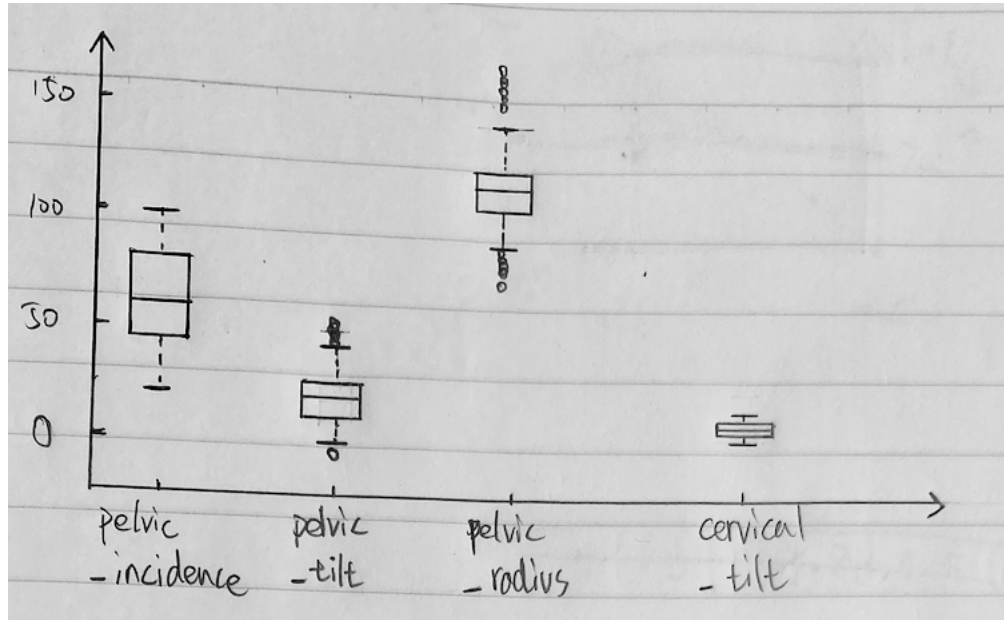


Dataset description:

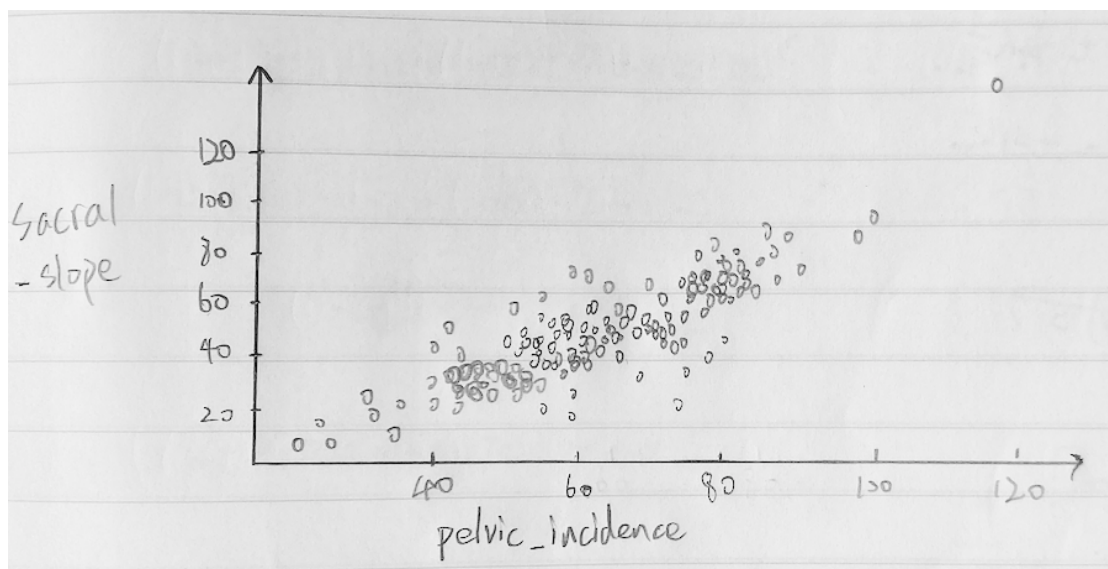
A dataset including normal and abnormal patients along with measures of cervical, thoracic and sacral spine and pelvic movement. With all these measurements, we can try to predict the last variable *class\_attribute* with a classification model.

1.



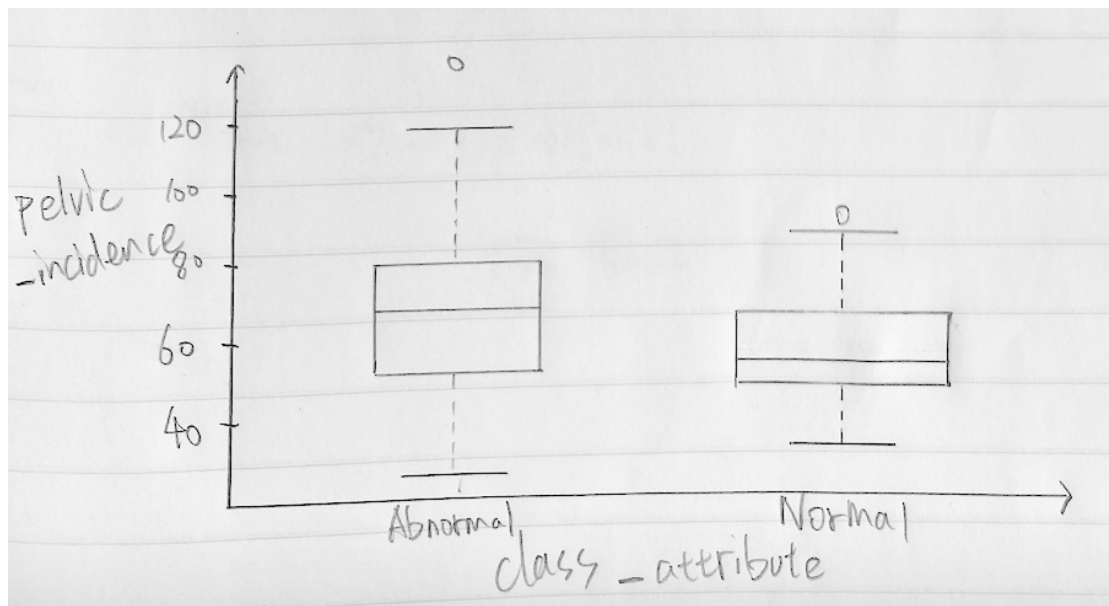
With the boxplot above, I want to address the question whether each variable resides in a reasonable range.

2.



With the scatter plot above, I want to address the question whether there is positive or negative correlation within all the variables. In this case, we can spot a positive correlation between variable *pelvic\_incidence* and *sacral\_slope*. So, when we build a classification model, we may want to keep only one of these two variables.

3.



With the boxplot above, I want to address the question do measurements differentiate between normal and abnormal patients. In this case, variable *pelvic\_incidence* has different distributions for two classes of patient.