FEDERICO RUEDA

LUCAS WHITMIRE

EDGARDO PANZA

# INTRODUCTION

Lung and breast cancers were the most common cancers worldwide, each contributing 12.3% of the total number of new cases diagnosed in 2018.

In 2019 in the US, an estimated 268,600 new cases of invasive breast cancer were diagnosed among women and approx. 2,670 cases were diagnosed in men.

For this project we choose a dataset related to breast cancer detection.

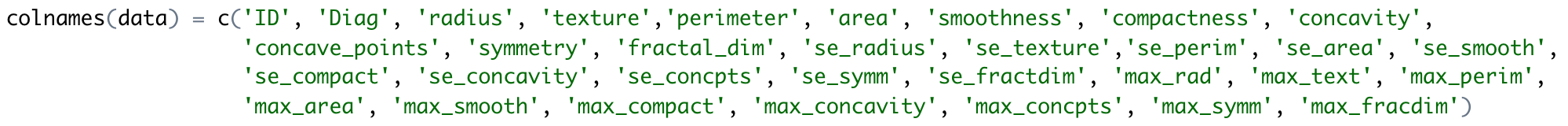
# OBJECTIVE

Apply all the concepts learned in the class of machine learning II and create the best model for the chosen dataset.

# DATASET

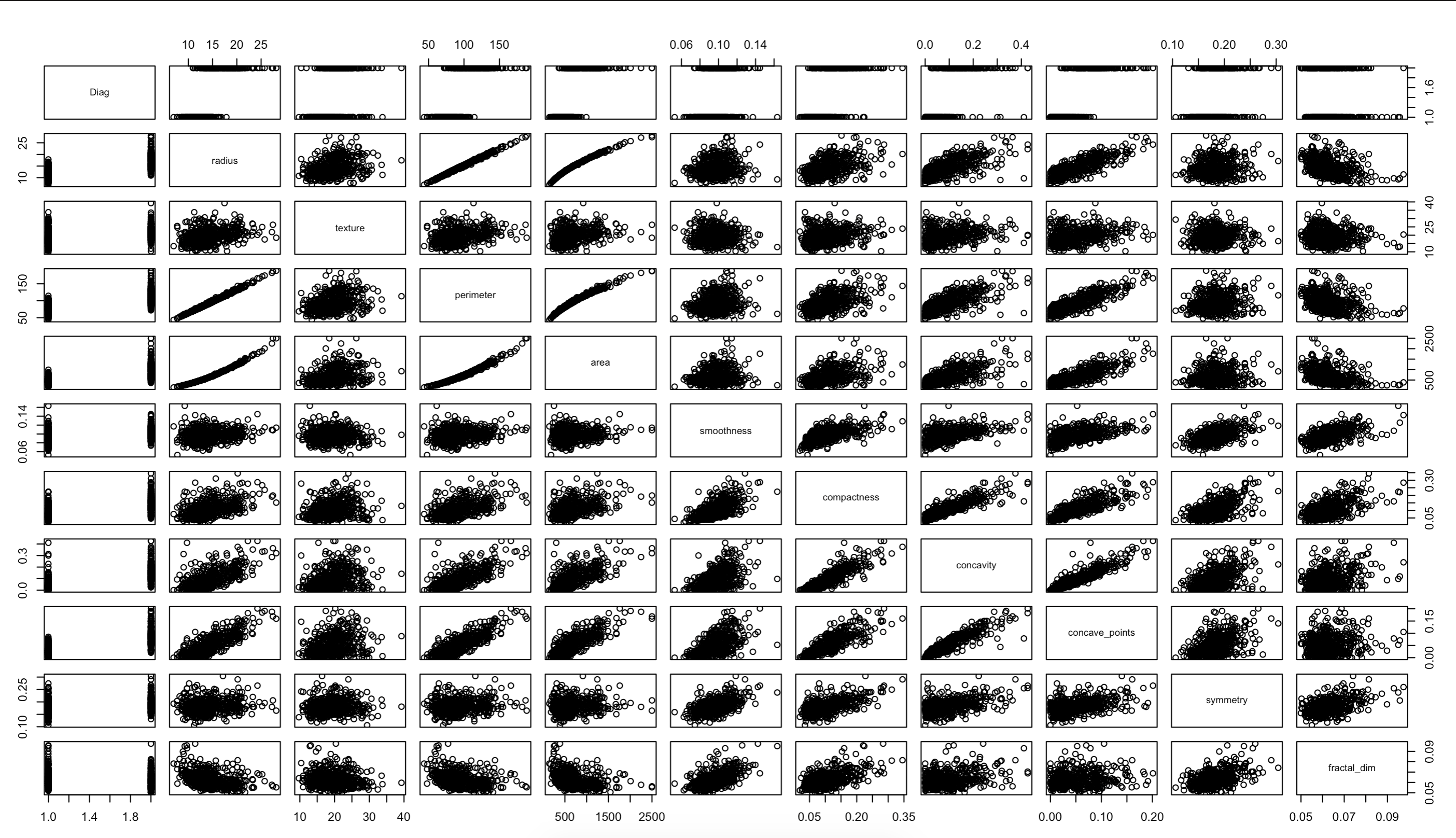
The dataset consists of 569 observations and 31 variables.

The variables are:

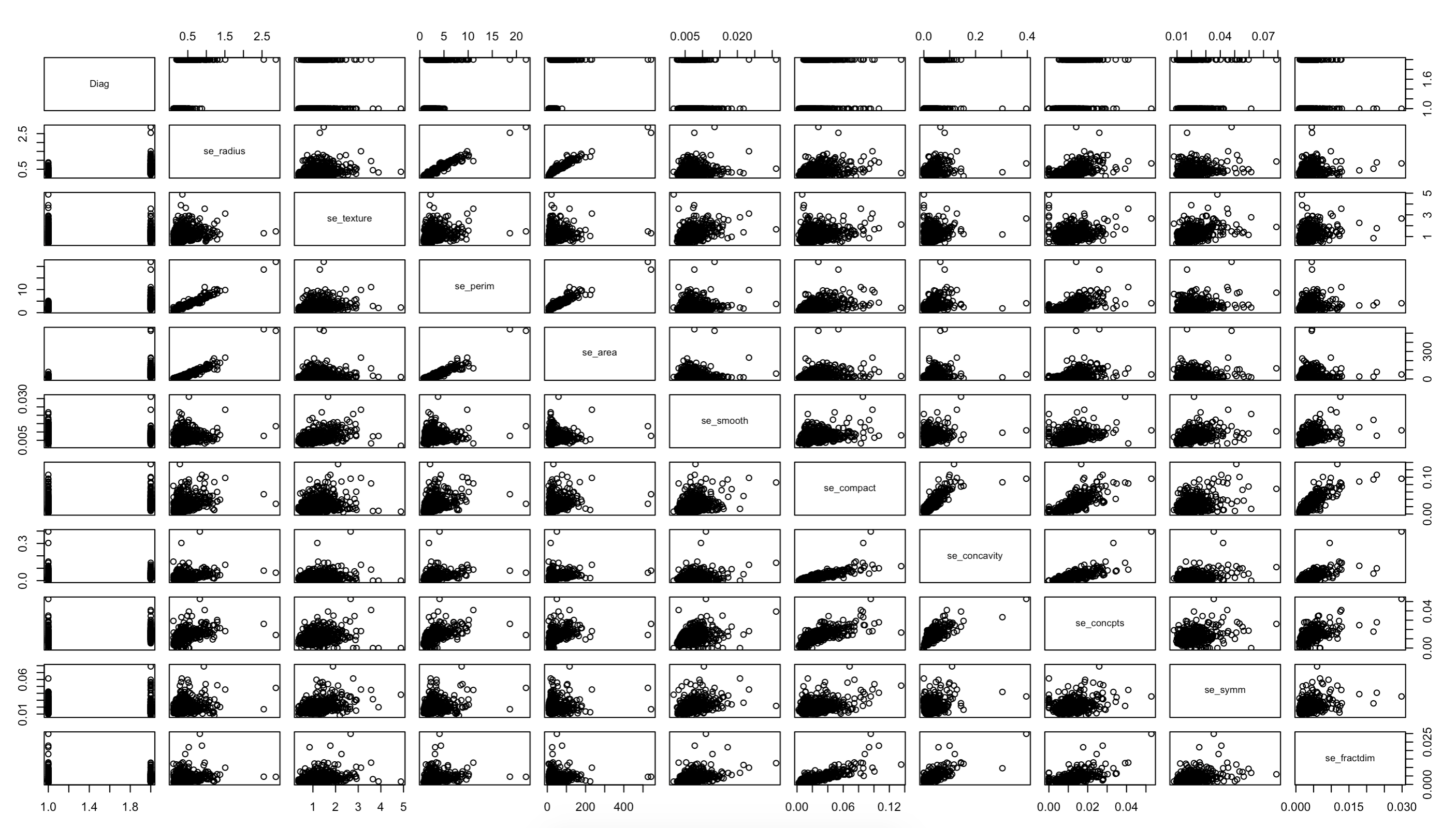


# EDA

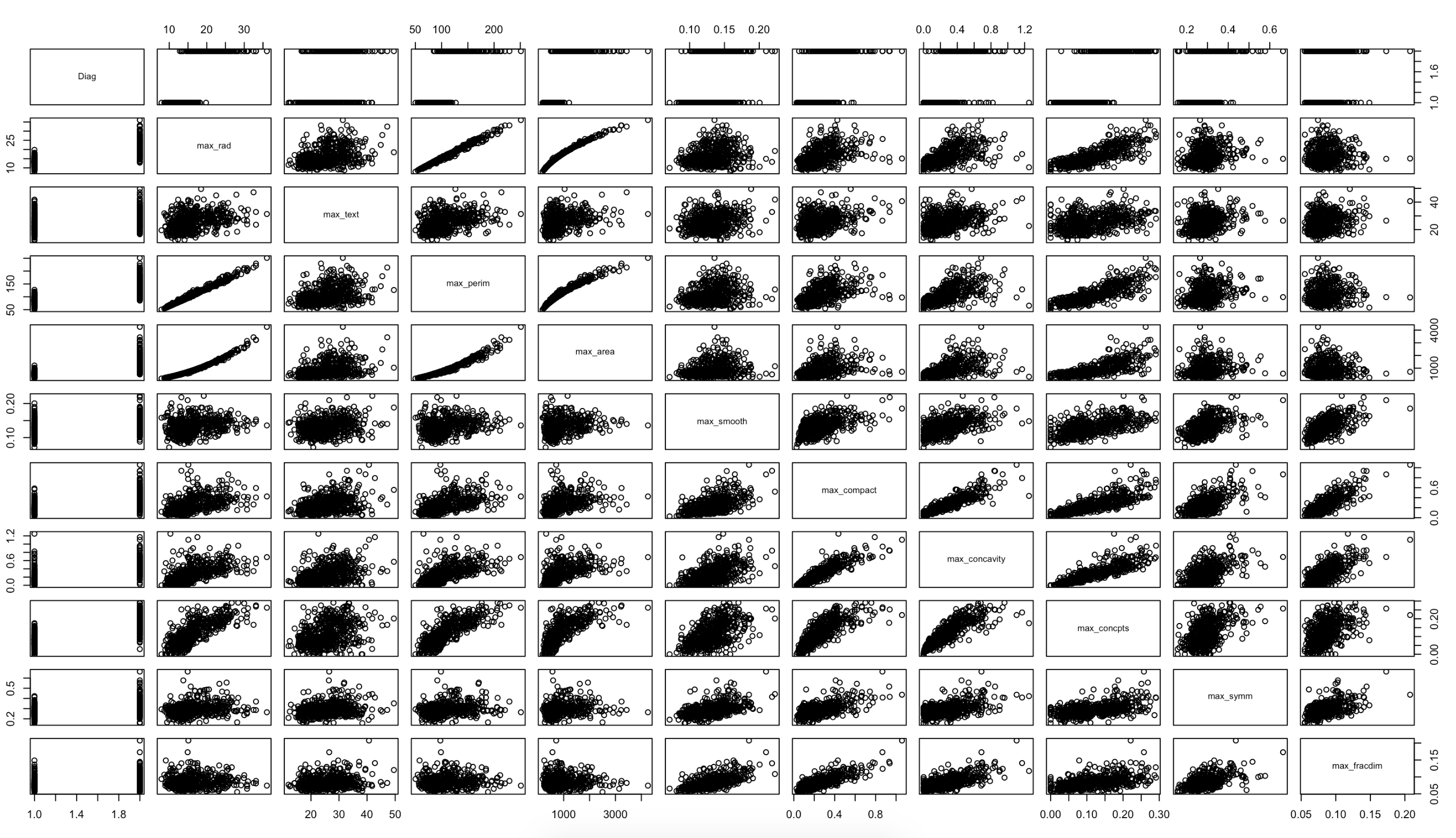
Correlation between “diagnosis” and the 2-11 variables:



Correlation between “diagnosis” and the 12-21 variables:

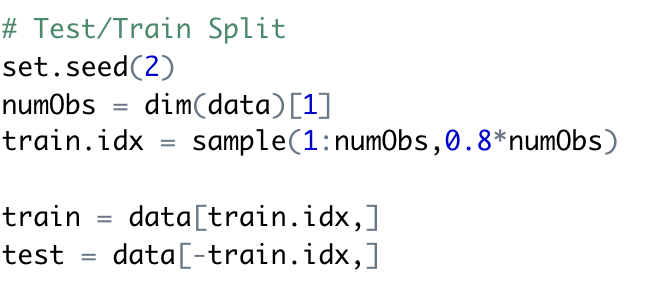


Correlation between “diagnosis” and the 22-31 variables:



# MODEL

First, we separate our data into training and test, 80% and 20%.



SVM CLASSIFIER

NEURAL NETWORK

# CONCLUSIONS