Team Cinco presents

Analyzing Breast Cancer

Presented by
Goutham Arra
Harika Putti
Jeevan Reddy Rachepalli
Pavan Kumar Madineni
Priyadarshini Vijjigiri

Problem Statement

- 1 in 8 women in the U.S suffers from Breast cancer.
- About 86% of the patients can be cured if treated in the early stages.
- Currently women can only undergo screening tests which are just physical examinations
- If any abnormalities are found, then they have to undergo a bunch of very expensive tests for diagnosis.
- Our objective is to find the main factors that can help identify cancerous tumors.
- We then build a model to identify if the tumor cell is cancerous.

Data Set

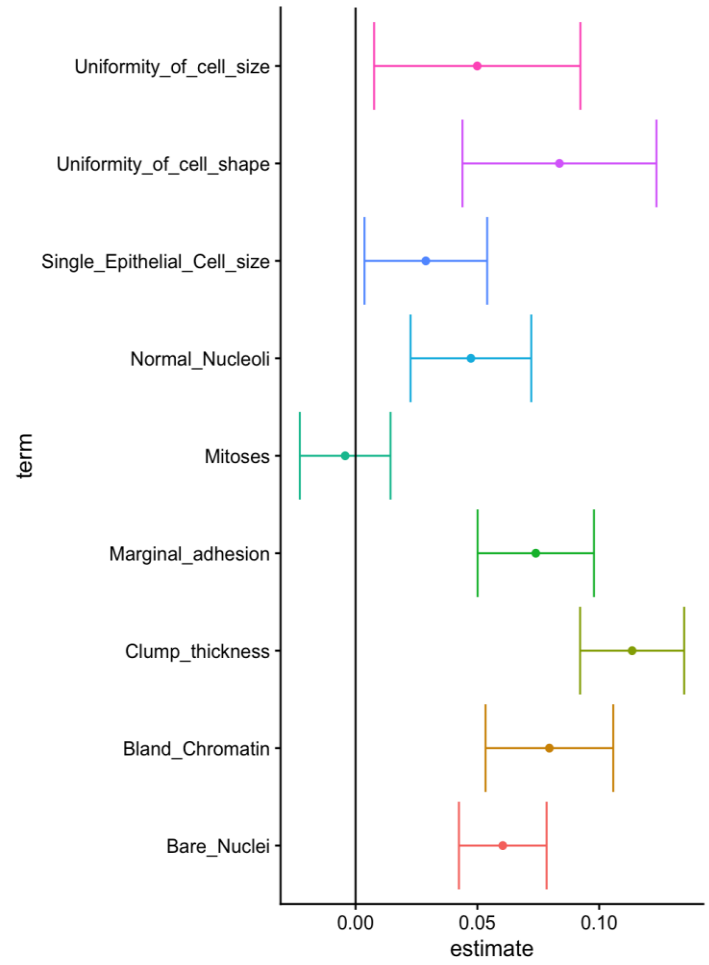
Data Set collected by:
University of Wisconsin Hospitals, Madison
from Dr. William H. Wolberg

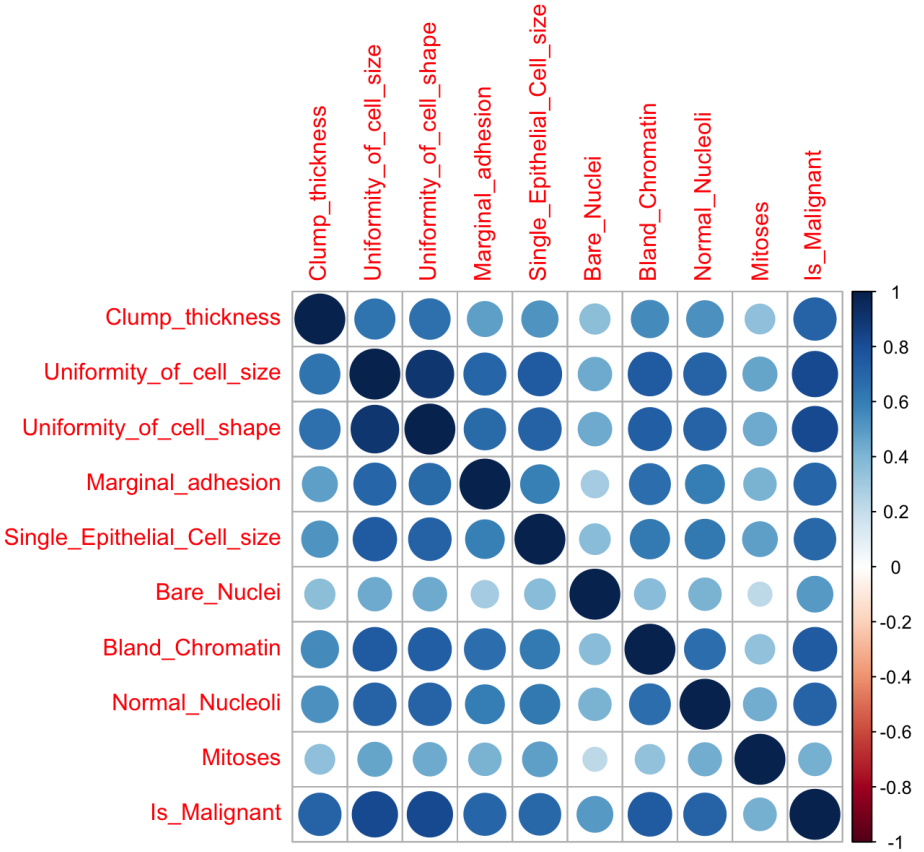
Stored in UCI Machine Learning Repository
https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/breast-cancer-wisconsin.data

Data Set Description

- The data was taken from Cytology laboratory tests
- There are 9 attributes and one class- whether the tumour is malignant or not.
- We had a total of 699 observations out of which 16 had missing values(we dropped those because they constituted < 2.5% of the data)
- The data was collected from women who had undergone laboratory test to find out whether the lump formed in their chest is a Benign or a Malignant tumor.
- Each of the 9 predictor variable is scaled from 1-10. Higher the value, greater the chance of it being malignant.

Variables

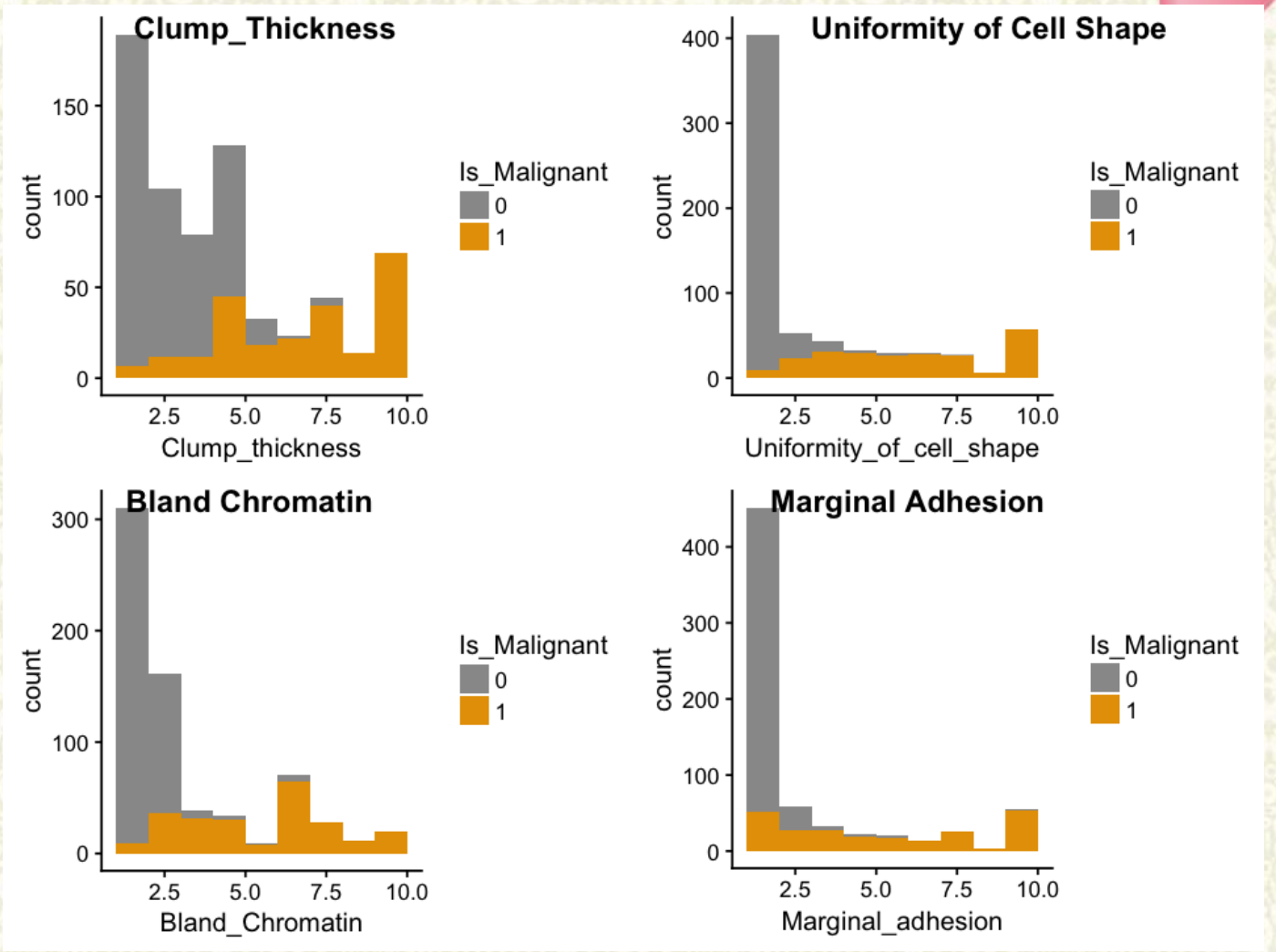

Uniformity_of_cell_size
Uniformity_of_cell_shape
Single_Epithelial_Cell_size
Normal_Nucleoli
Mitoses
Marginal_adhesion
Clump_thickness
Bland_Chromatin
Bare_Nuclei
term
0.00
0.05
0.10
estimate


Clump_thickness
Uniformity_of_cell_size
Uniformity_of_cell_shape
Marginal_adhesion
Single_Epithelial_Cell_size
Bare_Nuclei
Bland_Chromatin
Normal_Nucleoli
Mitoses
Is_Malignant
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

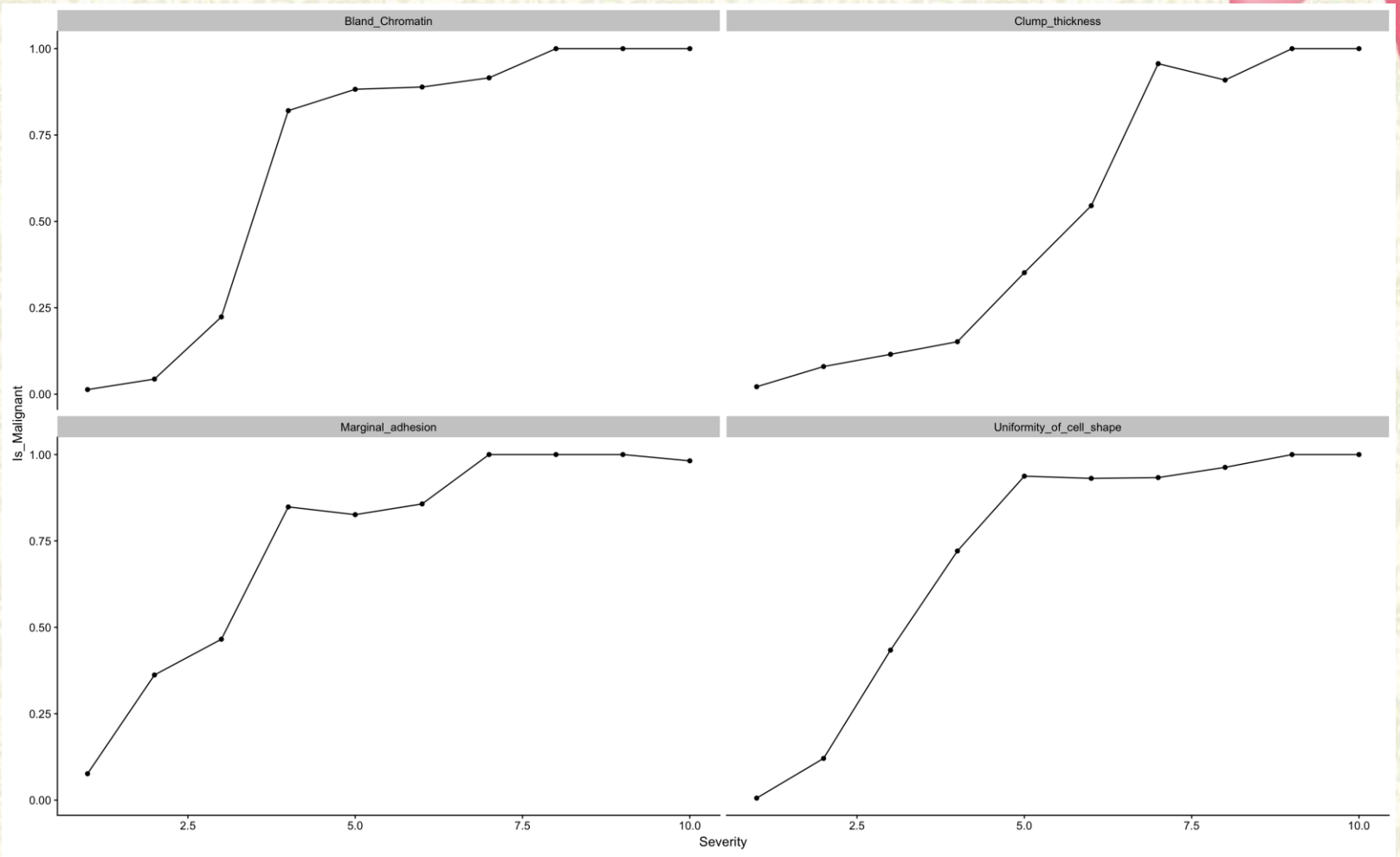
Understanding Predictors

- **Clump thickness:**
  Benign cells tend to be grouped in monolayers, while cancerous cells are often grouped in multilayers.
- **Uniformity of cell shape:**
  Cancer cells tend to vary in size and shape. So uniformity of cell size/shape points in a benign direction.
- **Bland Chromatin:**
  Describes a uniform "texture" of the nucleus seen in benign cells. In cancer cells the chromatin tends to be coarser.
- **Marginal adhesion:**
  Normal cells stick together while cancer cells lose their ability to do so. So loss of adhesion is a sign of malignancy.

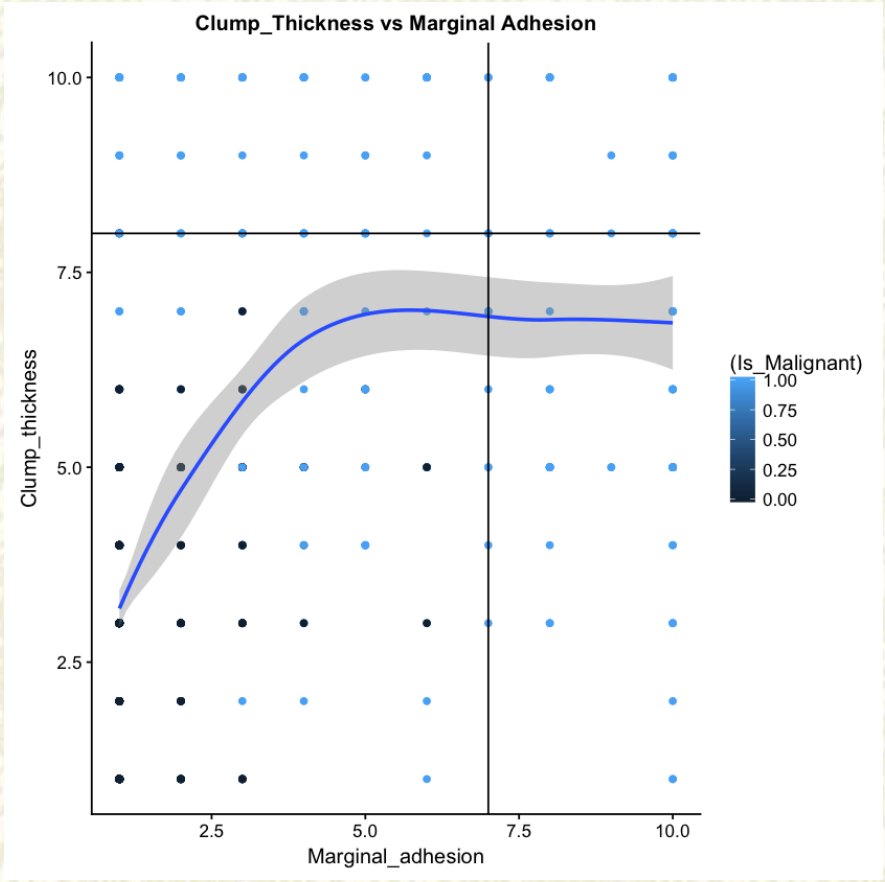
Cases of Benign and Malignant

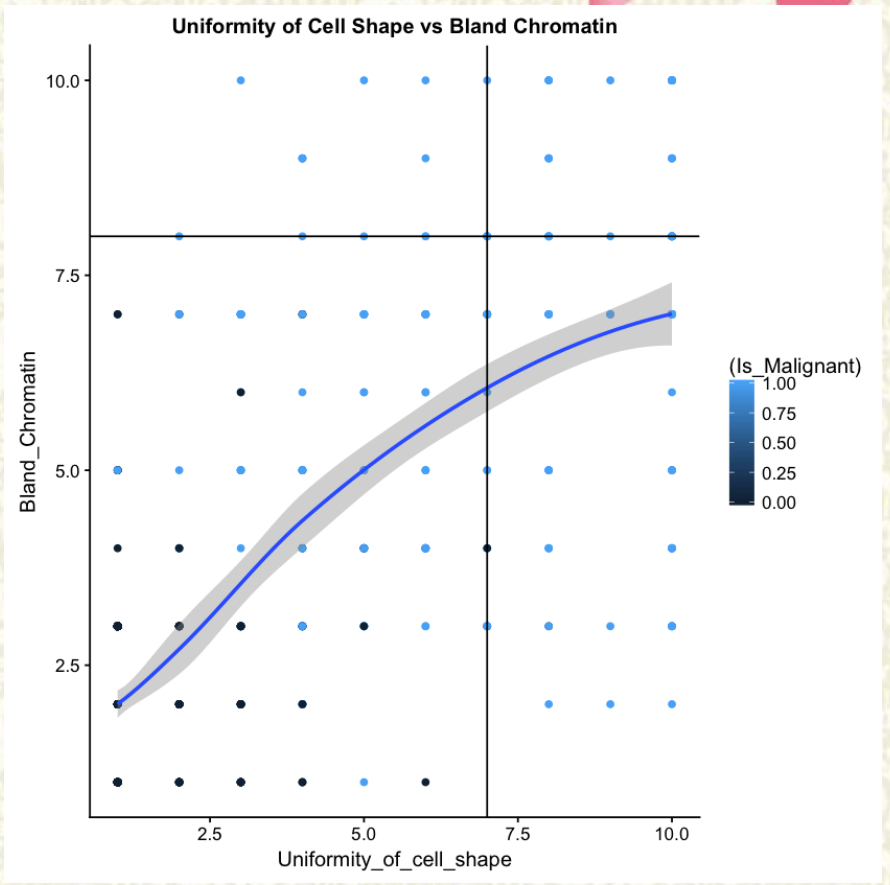


Average change in malignance

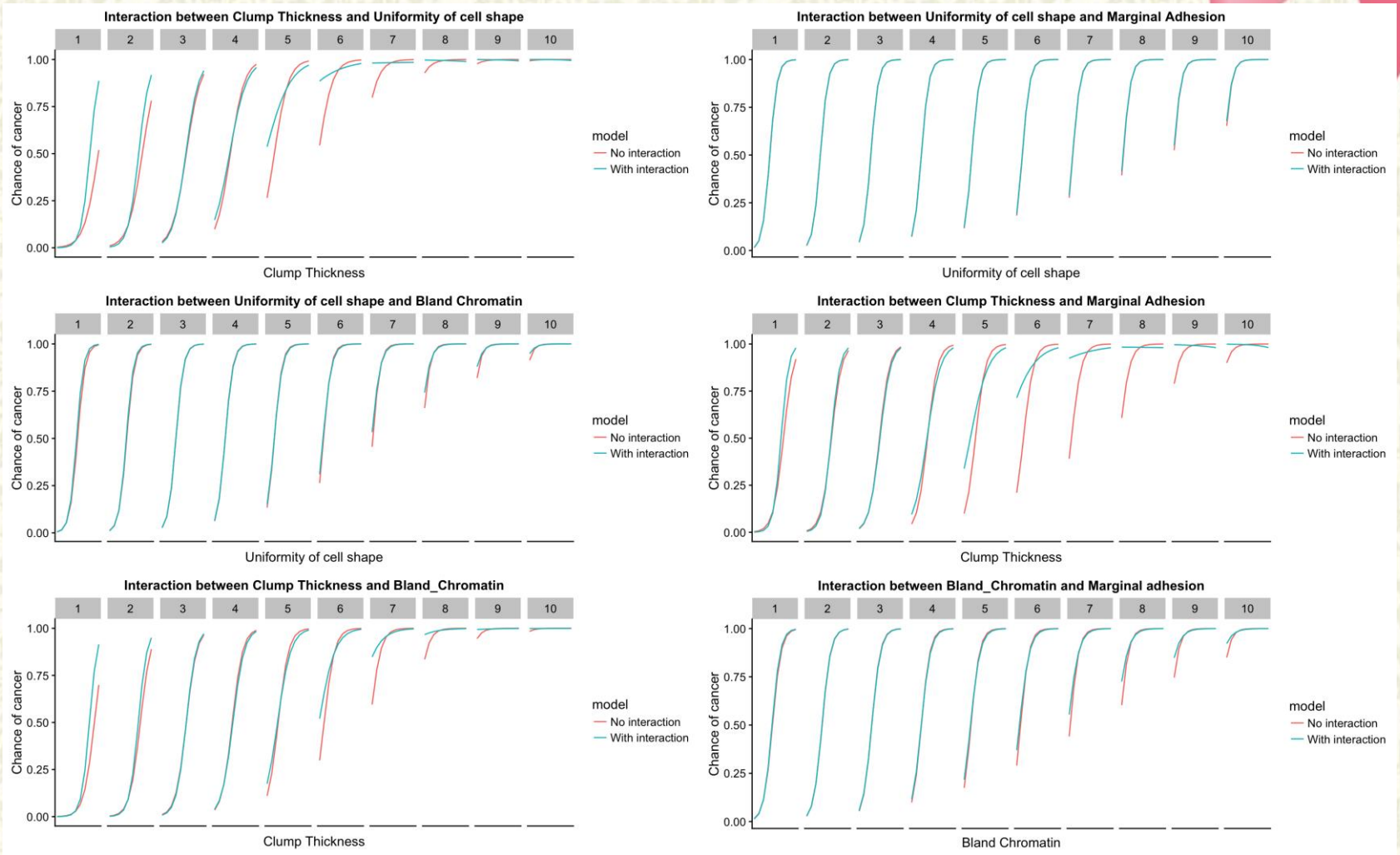


Predictor interactions

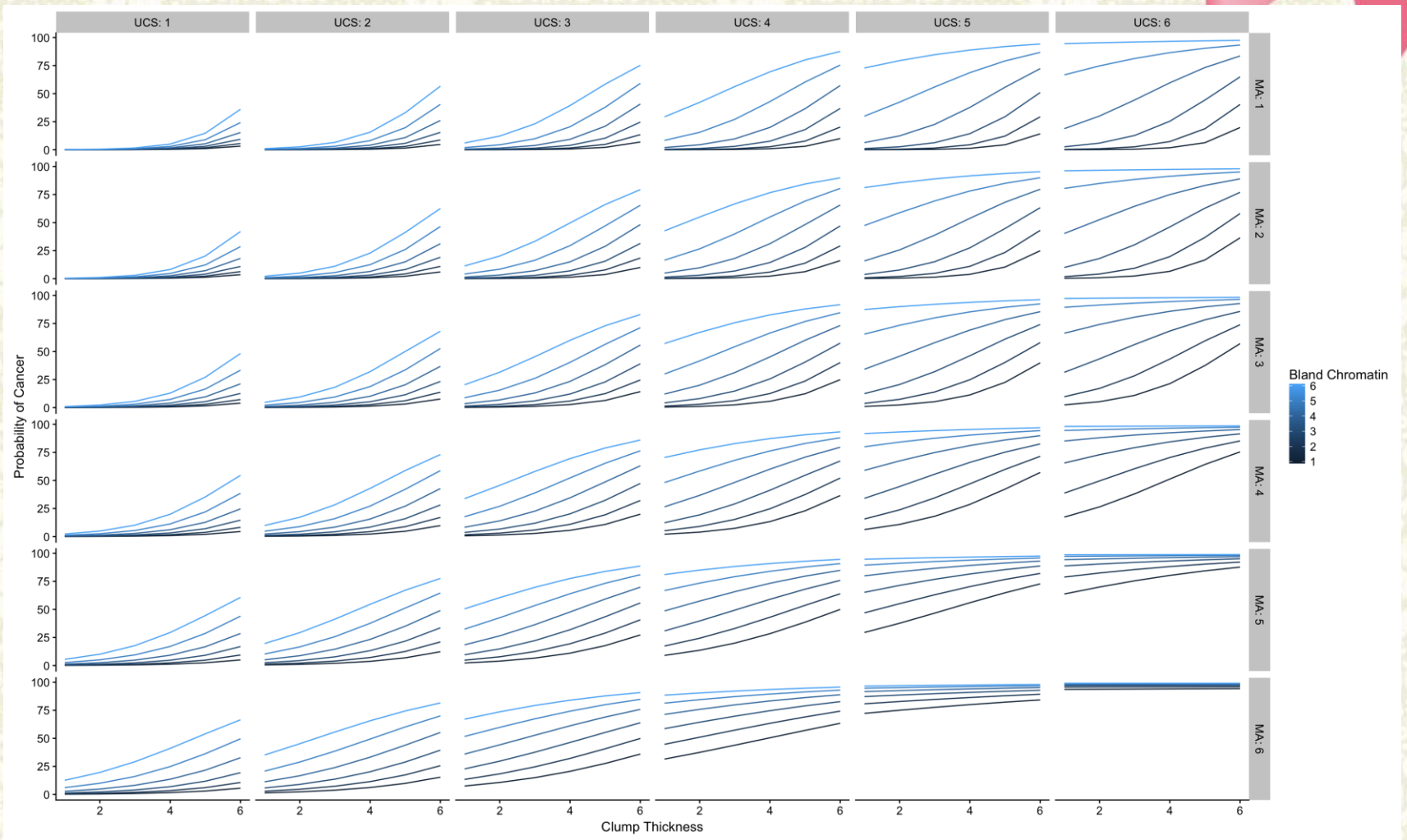




Checking for Interactions to include in the model



All Predictors' Interaction



Scope and Limitations

- We have done a detailed analysis of which of the attributes can be more helpful in determining if a tumor is cancerous.
- There is scope for R&D focusing on methods to measure these attributes precisely and with cheaper methods.
- We built a model as a preliminary test of breast cancer. We only explored laboratory data and drew inferences from that. We require real life data in-order to showcase any kind of accuracy or loss.
- This is a faster way of classifying the tumor and probably more reliable than physical examination but it is highly advised to go for further tests once a malignant tumor is detected.

Thank You