Biopsies were collected via fine needle aspiration (FNA), it is a procedure where a thin hollow needle is inserted into the patient’s tissue of question to aspirate cells to be later examined using microscopy. Eleven cytological characteristics were graded individually on how they portrayed benign or malignant characteristics. This scale was based on a 1-10 grade administered by a Dr. Wolberg[right??????] as the samples arrived periodically. A lower score corresponds to non-cancerous attributes (benign) of the particular cytological characteristic while a higher score corresponds to a more cancerous attribute (malignant).

Cytological characteristic explanations of the dataset:

Id. Unique identification number administered in chronologically order as the samples arrived and were graded.

Cl.thickness = clump thickness. Since “normal” epithelial cells have contact inhibition on many points of contact on their cellular membrane, their ability to grow on top of one another is assessed by looking at whether the sample is a monolayer of cells (exhibits contact inhibition) or is multilayered (ie. Increased clump thickness and does not exhibit contact inhibition). Increased clump thickness is thus an indicator of malignant cancerous cells present in the sample.

Cell.size = Uniformity of Cell Size. A biopsy sample is such a small tissue sample that when they are sampled you would expect nearly all of the cells sampled to be the same cell type. Because of this you would expect the cells to exhibit a uniform size since they should nearly all be the same cell type. This grade is then assessed as the lower the score the more uniform the cell sizes are (benign/”normal”) while higher scores show more differing cell sizes potentially corresponding with being considered a malignant attribute.

Cell.shape = Uniformity of Cell Shape. Same explanation as uniformity of cell size but substitute cell shape instead.

Marg.adhesion = Marginal Adhesion. This grade is based off of the level of cohesion of the periphery of the epithelial cells. Healthy (or for these grading purposes benign) epithelial cells will normally bind tightly to one another while more malignant cancerous cells tend to lose this attribute in order to move around the body.

Epith.c.size = Single Epithelial Cell Size. Similar to uniformity of cell sizes and cell shapes, single epithelial cell size however looks for individual cells that are abnormal in size. The larger an individual cell is in what should be a nearly homogenous sample would lead an analyst to believe it is more likely malignant than benign/healthy.

Bare.nuclei = Bare Nuclei. Bare nuclei is a term used to describe a cells degeneration of cytoplasm, only leaving visible nuclei.

Bl.cromatin = Bland Chromatin.

Normal.nucleoli = Normal nucleoli. The nucleolus is the dark-staining portion of the nucleus where, among other things, ribosomal subunits are assembled. Ribosomes are needed to translate transcribed RNA into proteins. Therefore in actively growing cells these nucleoli are very large (also usually one or two nucleolus per nucleus but sometimes more) and easier to see under a microscope compared to a benign/healthy cell. A large collection of healthy or benign somatic epithelial cells therefore would not have an abundance of nucleoli or abnormally large ones. However, malignant cancerous cells are constantly growing and would have large and an unusually high amount for epithelial cells.

Mitoses = Mitoses. Mitosis is the process of making two genetically identical daughter cells from one parent cell. For somatic epithelial cells and benign cells this process does not happen regularly and only occurs when the cells receive certain signals from other cells to divide. Malignant cancerous cells lose the need for these signals in order to conduct mitosis or produce the signals themselves in a perpetual self-feedback loop. Because of this when looking at a tissue sample under a microscope, malignant cancerous cells will have an abnormal amount of cells going through the stages of mitosis. So the lower the grade here the less mitosis is present, while higher grades correspond to higher amounts of mitosis visible in the sample.