



Healthcare Diagnoses: Breast Cancer Analysis



Jacquelyn Brosseau, Pranita Sutrave, Alondra Solis
Barrales, Bailee Wright

Topic

- Our project aims to recognize differences between physiological variables associated with breast cancer tumors, to gain greater insight into diagnosis patterns.
- We explore the sizing of breast cancer tumors through examination of area and radius means, to determine their role in distinguishing between benign and malignant tumors.
- Research Question -> How can the key characteristics between malignant and benign breast cancer tumors be compared to determine the diagnosis?
- The dataset used in our project is called Breast Cancer Classification and was found on Kaggle.

Data

- Source- Breast Cancer Wisconsin (Diagnostic) dataset
- Contains information about breast cancer diagnoses
- Computed from digitized images of fine needle aspirate (FNA) biopsies of breast masses.
- The data is used to predict whether a tumor is malignant or benign
- The data contains various variables related to physiological properties of tumors (e.g. radius, area, concavity, and concave points)

Highlights from EDA

- Malignant tumors are associated with higher mean values of cell measurements than benign tumors
- Suggests that tumors that are larger and that have more irregular shapes are often malignant

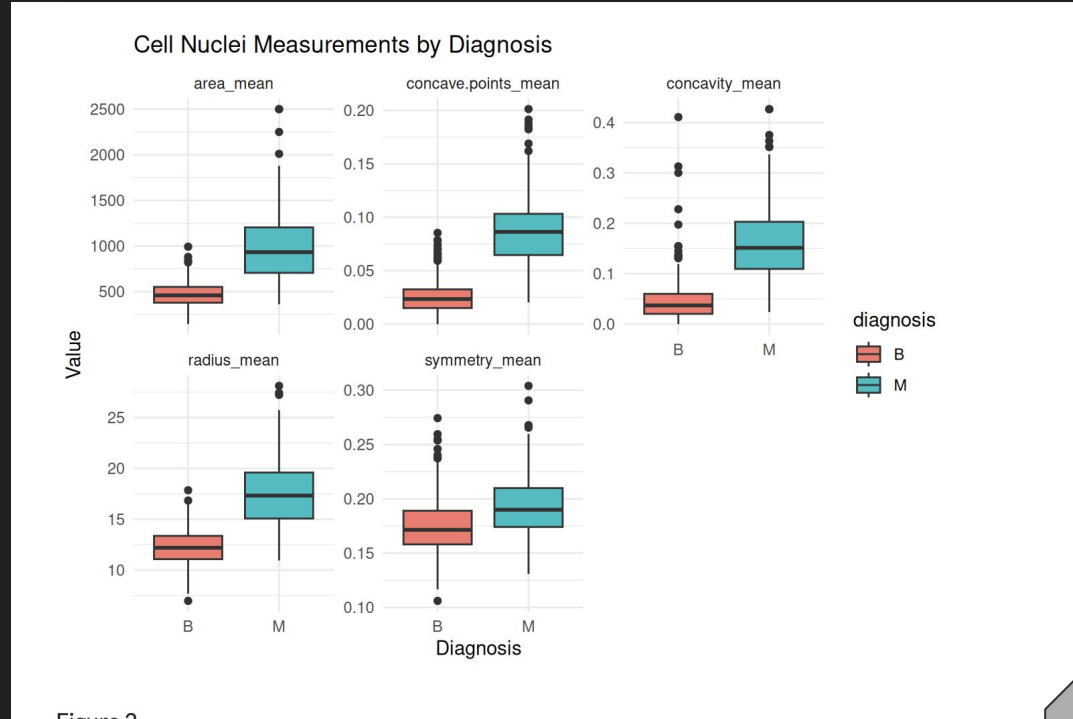
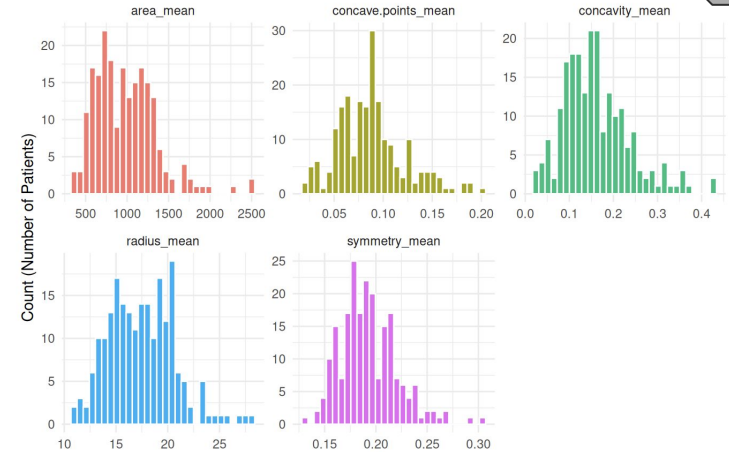


Figure 2

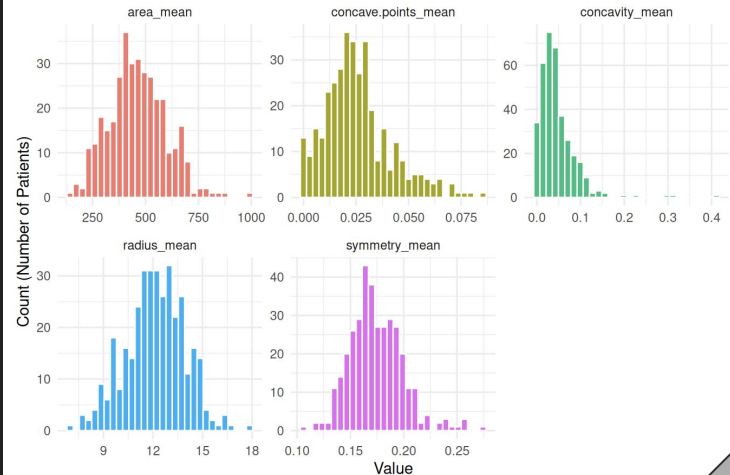
Highlights from EDA

- These separate histograms for cell nuclei measurements allow distributions to be compared between malignant and benign cells.
- Differences in values as well as distributions for some variables.
- Allows conclusions to be made about differences in size and shape of tumors and how that could contribute to diagnosis.

Distribution of Cell Nuclei Measurements (Malignant Tumors)



Distribution of Cell Nuclei Measurements (Benign Tumors)



Highlights from EDA

- Malignant tumors show higher mean values than benign tumors across all measured variables
- Malignant tumors tend to be **larger** (higher area and radius means) and more **irregularly** shaped (higher concavity and concave point means)
- Distributions for area and radius are unimodal in both groups
- Malignant tumors skew slightly left and have much larger values overall
- Concave points and concavity distributions indicate malignant tumors have more surface indentations and texture
- Symmetry mean is the exception — both groups show similar values and distributions, suggesting it is a weaker distinguishing feature between diagnoses

Conclusion, Discussion, and Critiques

Tumors classified as malignant are significantly larger and have more irregular cell nuclei measurements (higher radius, area, concavity, and concave points). These irregular measurements reflect the structural mutations that unregulated cell division induces.

Limitations/Critiques : The data's focus on patient cases from the University of Wisconsin hospital system and the exclusion of previous medical histories.

- Subsequent studies should include data from various hospitals that vary in patient populations and broad patient factors such as age, gender, family history, or existing chronic conditions.