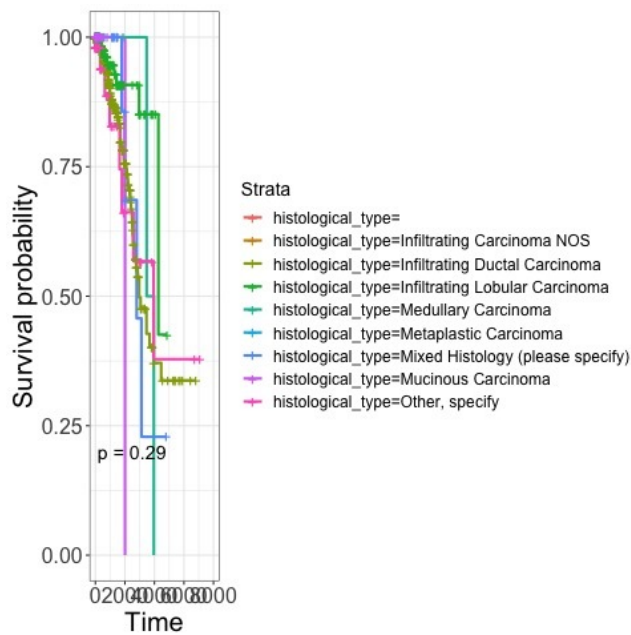


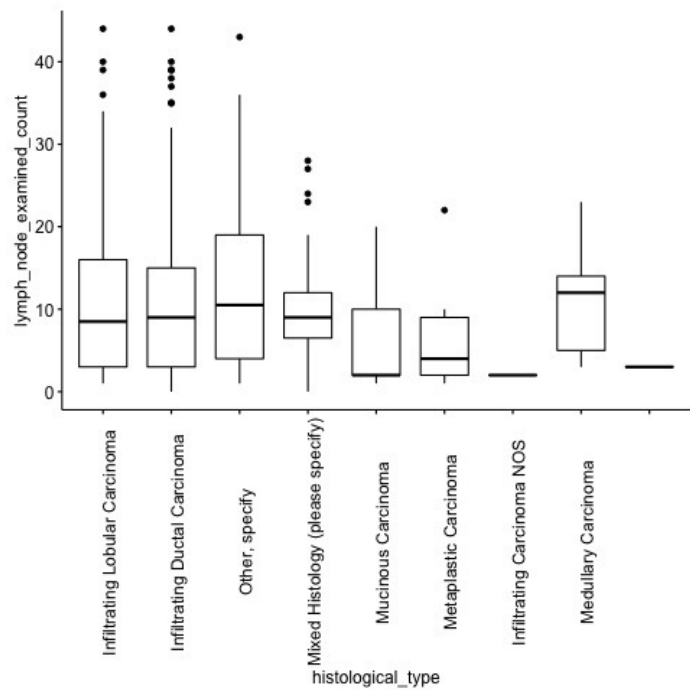
1. categorical variable: qualities of an individual; ex = sex, race
discrete variable: can only take on a finite number of real values; ex = point score, # of people

continuous variable: can take on infinite number of real values; ex = temp, weight
2. histological_type - 0 NAs
3. determined by microscopic examination of suspected tissue that has been excised by biopsy or surgical resection - categorical variable
4. <https://www.nature.com/articles/6602787> - different histological types of BC vary in their clinical and tumor characteristics, with inflammatory carcinomas more likely to be associated with poorer diagnoses
https://www.sciencedirect.com/science/article/pii/S1470204506709111?casa_token=YmoX4X2xjkoAAAAA:oEEDXZUVh5i0EZQZv8w0C4g42qhEpsWgQPvUsO9ivYK2mGuit0tVDjfk8-BJGMRq1facjC1Mi-Q - the risks associated with hormone therapy are higher for lobular and tubular cancer than ductal cancer
5. Lymph_nodes_examined_count variable = 139 NAs
#of lymph nodes in patient; lymph nodes are correlated with cancer progression, so can affect treatment consideration
6. 1- The number of lymph nodes examined in patients could depend on different histological types.
2 - More aggressive histological types result in poorer survival outcomes.
3 - More lymph nodes shouldn't have a significant impact on survival because they are examined during the diagnostic phase.
7. Boxplot: The distribution of lymph nodes is not statistically significant because the error bars are overlapping with each other for different subtypes. Thus, our hypothesis 1 is likely not true because it seems as though lymph nodes do not depend on histological subtype. I think this makes sense because lymph nodes are observed at the beginning of diagnosis when histological subtype might not be known.
Histological KM Plot: The p-value of 0.29 indicates that we cannot conclude that a statistically significant difference exists in the survival of breast cancer patients of different histological subtypes, disproving hypothesis 2.
LM KM Plot: Similarly, the high p-value of 0.46 indicates that the number of lymph nodes likely has no association with survival outcome, which supports our hypothesis 3.

Plots:



Histological KM Plot



Boxplot

Lymph node KM Plot

