Project #4 Predictors of Lung Cancer

Nefertiti Muhammad, Adrian De La Cruz, Joe Coffaro, and Alexander Walden

Lung cancer is the leading cause of cancer death worldwide. Because of late stage diagnosis, it is difficult to treat and has a low 5yr survival rate. Identifying predictors of lung cancer is key to prevention. Our goal is to create a predictive model of lung cancer, focusing on demographic and lifestyle differences between patients.

1. Are demographic factors (e.g. sex, age) and lifestyle choices associated with certain lung cancers?
2. What impact does smoking history have on the development of lung cancer?
3. What risk do non-smokers have of developing lung cancer and if so is it related to pollution?

Data Source

Using the cBioPortal we pulled a dataset focusing on lung cancer that originated from Nature Genetics.

cBioPortal for Cancer Genomics
Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet. 2016 Jun;48(6):607-16. doi: 10.1038/ng.3564. Epub 2016 May 9. PMID: 27158780; PMCID: PMC4884143.

Potential Coding Tools

Python Pandas
Python Matplotlib
HTML/CSS/Bootstrap
JavaScript Plotly
JavaScript Leaflet
SQL Database
MongoDB Database
Google Cloud SQL
Amazon AWS
Tableau