A Deeper Dive into Causes of Brain Strokes and Prediction Model

Sri Ramya Deepthi Bulusu, Isaac Ombotimbe and Karan Kaushal

2022-12-10

Abstract

A Stroke is a medical emergency that occurs when there is limitation or a complete cut off blood flow to the brain. It deprives a person's brain of oxygen and nutrients, which can cause brain cells to die. There are many factors which result in the stroke such as Age, High BMI, Hypertension, Smoking, Medical Record of Heart Diseases etc. As per the contemporary stroke statistics the range of stroke incidence is in between 41/100,000/year and 297/100,000/year in different countries. In this project, we intend to use the data science model approach with the use of the Stroke Prediction Dataset consisting of more than 5000 observations created by Federico Soriano Palacios and retrieved from Kaggle. Using the various attributes within this dataset which relate to brain stroke causes, several types of visualisations will be compiled for the purpose of developing a model which helps predict which groups of patients are more likely to experience a stroke. Some factors which will be used include the patient's age, BMI, hypertension status, smoking background, medical record of heart diseases, etc. By putting efforts into developing a model which could identify the risk indicators resulting in brain stroke, there are chances of reducing the sufferings by making necessary efforts to treat the symptoms in advance.

Introduction

In 2020, 1 in 6 deaths from cardiovascular disease was due to stroke. About 87% of all strokes are ischemic strokes, in which blood flow to the brain is blocked. Stroke-related costs in the United States came to nearly $53 billion between 2017 and 2018. This total includes the cost of health care services, medicines to treat stroke, and missed days of work. Knowing more about prevention and symptoms of stroke could save lives, including yours. By tidying, transforming and visualizing the dataset we are trying to answer some of our questions: • How the marital status of the person impacts the chances of Stroke? • Will Smoking lead to high probability of having Stroke? • Which gender of people are mostly prone to strokes? • Does the worktype have an impact on bmi and hypertension of the person? Do these influence the chances of having a stroke? • Does age has impact on strokes? • Does body mass index and glucose levels in a person, propel a stroke? • As many people say, various preventative steps such as switching to a healthy lifestyle by having a heart healthy diet, aiming for healthy weight, proper stress management, quitting smoking will help us to reduce the risk of having a stroke. Is this statement correct?

Dataset

Source: Stroke Prediction Dataset Originally created by Federico Soriano Palacios from Kaggle Dataset Description: 5110 observations with 12 variables **Key Variables**: Gender, Age, Hypertension, Heart Disease, Ever Married, Work Type, Residence Type, Avg Glucose Level, BMI, Smoking Status, Stroke, Age Category

Description

The dataset consists of 5110 observations with following 12 variables.

1. Gender
2. Age
3. Hypertension
4. heart_disease
5. ever_married
6. Work_type
7. Residence_type
8. Avg_glucose_level
9. Bmi
10. Smoking_status
11. Stroke
12. Age_category – We sub-divided the age in three categories (Children, Adult, Senior).

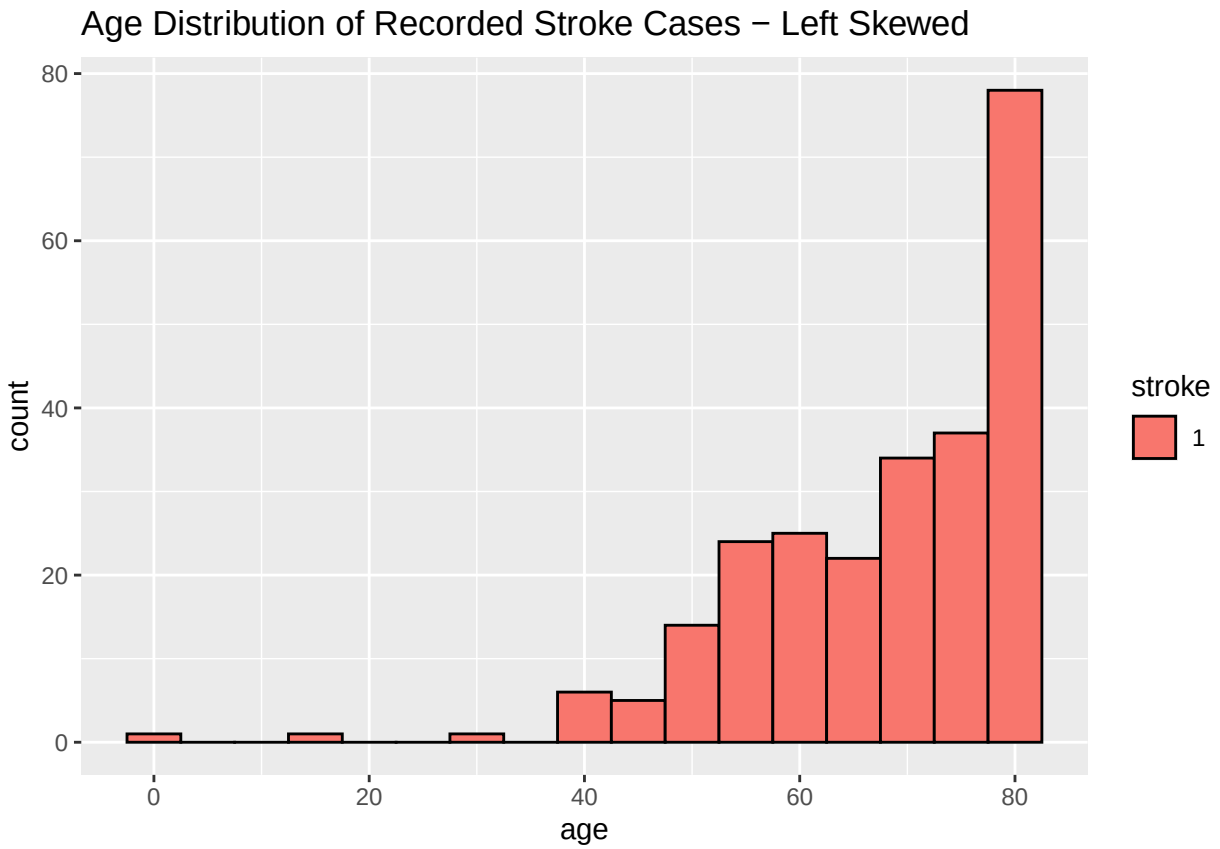
Recent research says that there are multiple factors which impacts the chances of someone experiencing a brain stroke, which are amongst the above accounted variables. Some of these factors have high probability impact on occurrence of a stoke for an individual.

Analysis/Research

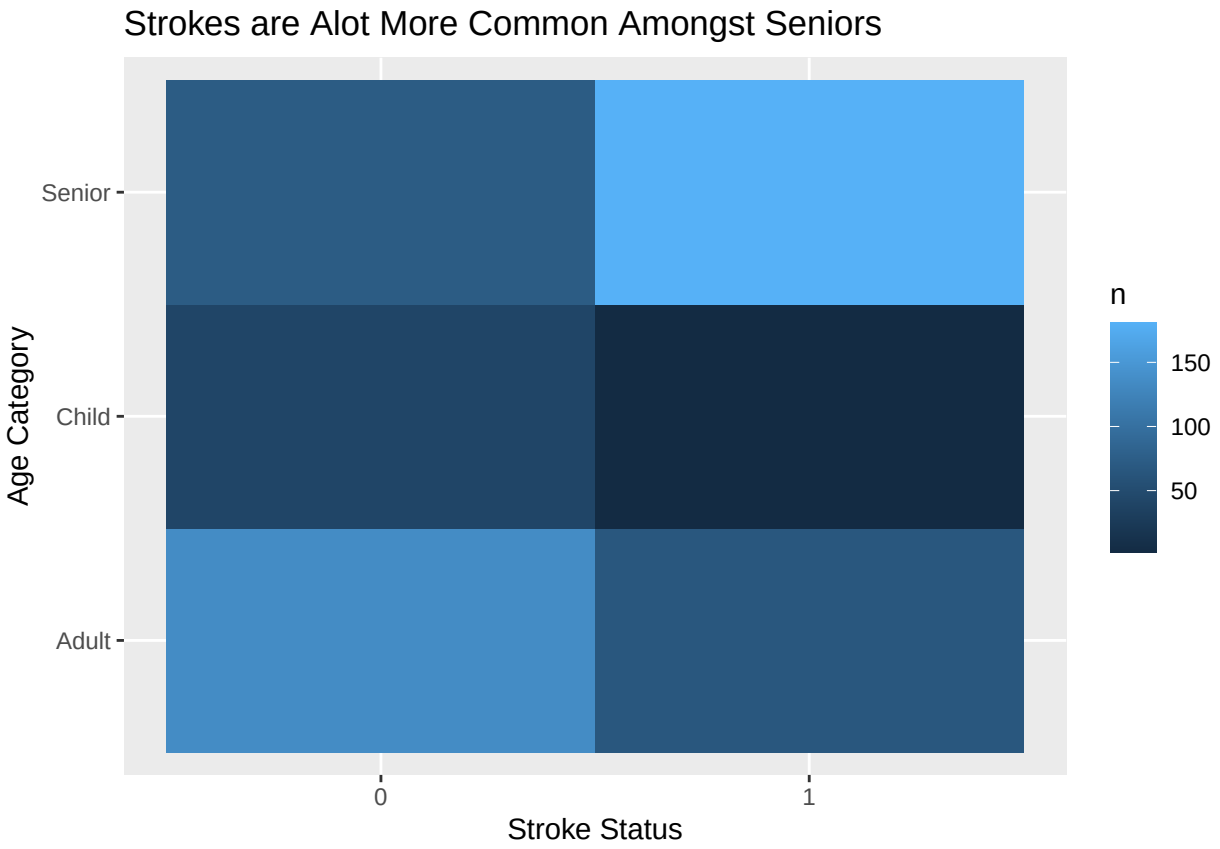
Our goal here is to build a predictive model to predict whether a person has a high risk of a stroke. Modifiable risk factors - Try to determine the factors which increase the chances of have a stroke in an individual,

Results

Age and Its Relation to Strokes

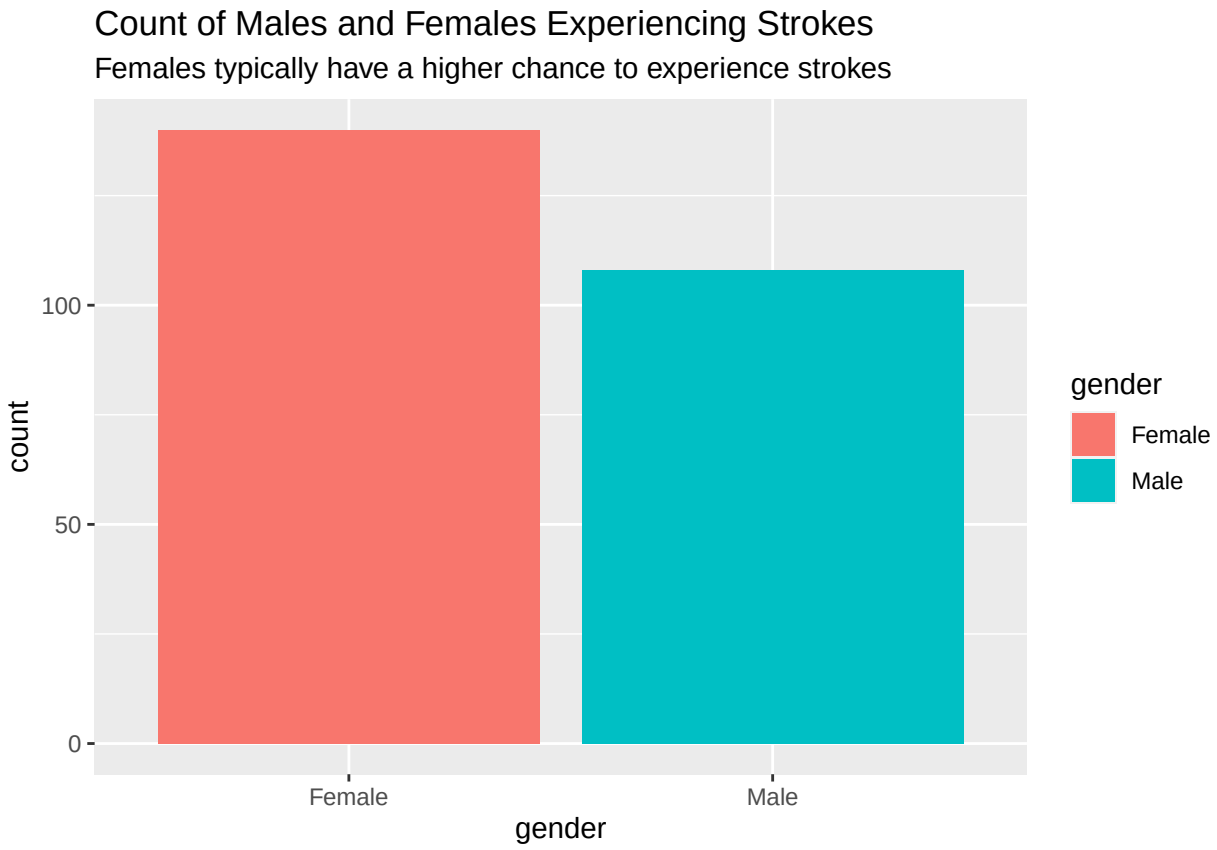


The figure above shows a histogram of the age distribution among people experiencing strokes from our data set. With the use of this visualization, we can see an abundance of stroke cases that start to rise at around the age of 40. The further right we move along the x-axis of the histogram, the see even more cases. It is at the age of around 80 years old that we see the most cases. From these observations we can conclude that the older an individual, the higher the risk of a getting a stroke. This can be linked to natural causes. From this plot as well, we can account for a few outliers. Because it is medically known that health conditions could be the result of a stroke we account these outliers as individuals born with health conditions such as prematurity as an example.

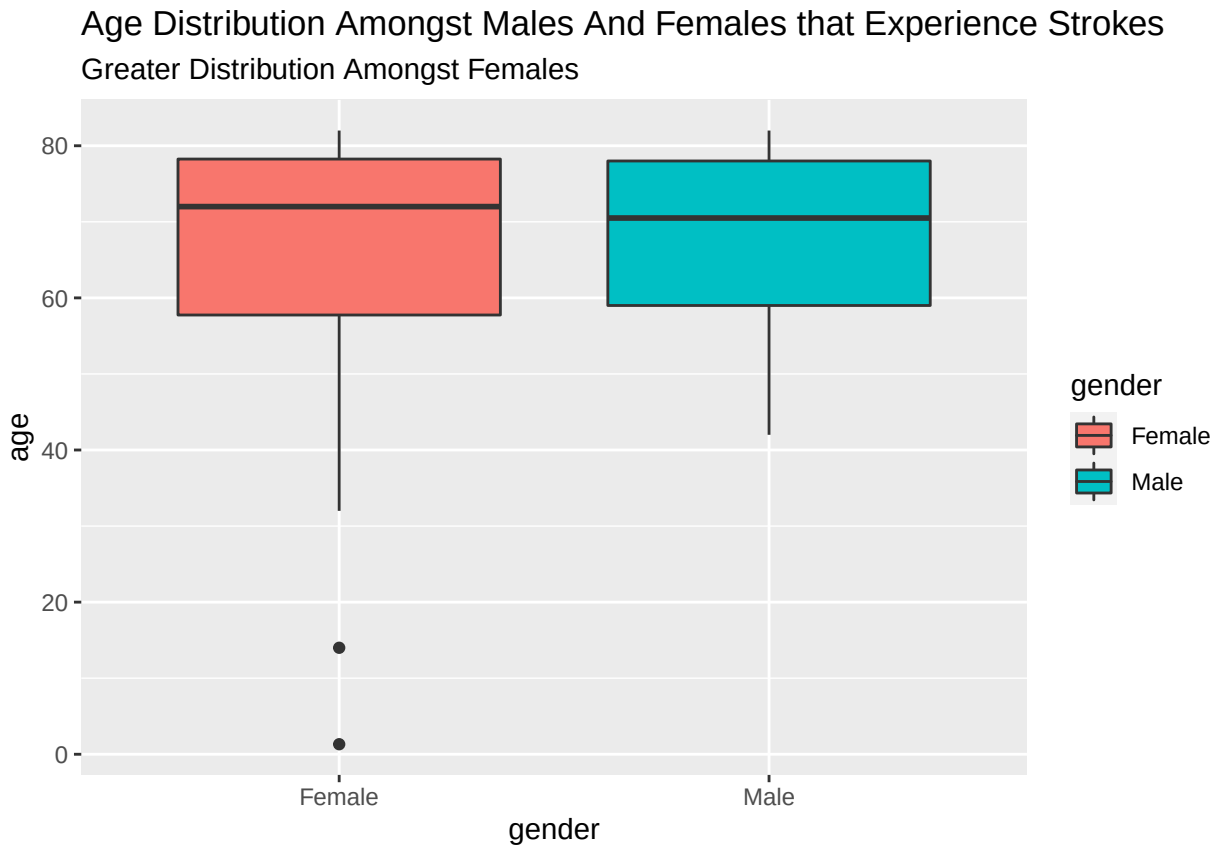


The figure above shows a heatmao and is used as another way in which we can visualize this distribution. In this heatmap, we see the distribution as from an age group perspective rather than specific ages. We conclude that seniors are most vulnerable to strokes.

Gender and its Relations to Strokes

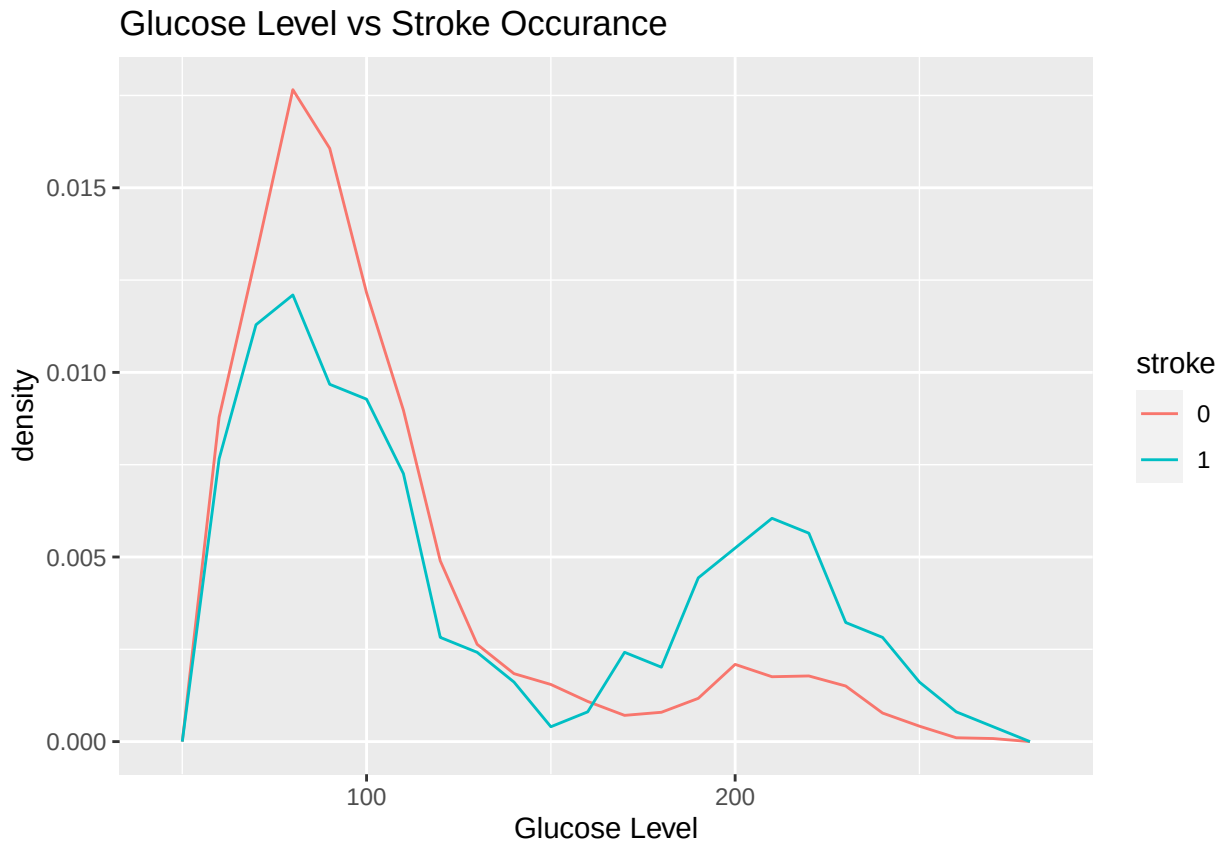


The figure above is a bar chart that shows the count of individuals experiencing strokes from our sampled data; however, grouped by gender. As it can be seen, we account for more women experiencing strokes. This could be explained because the average life span of men is greater then men.



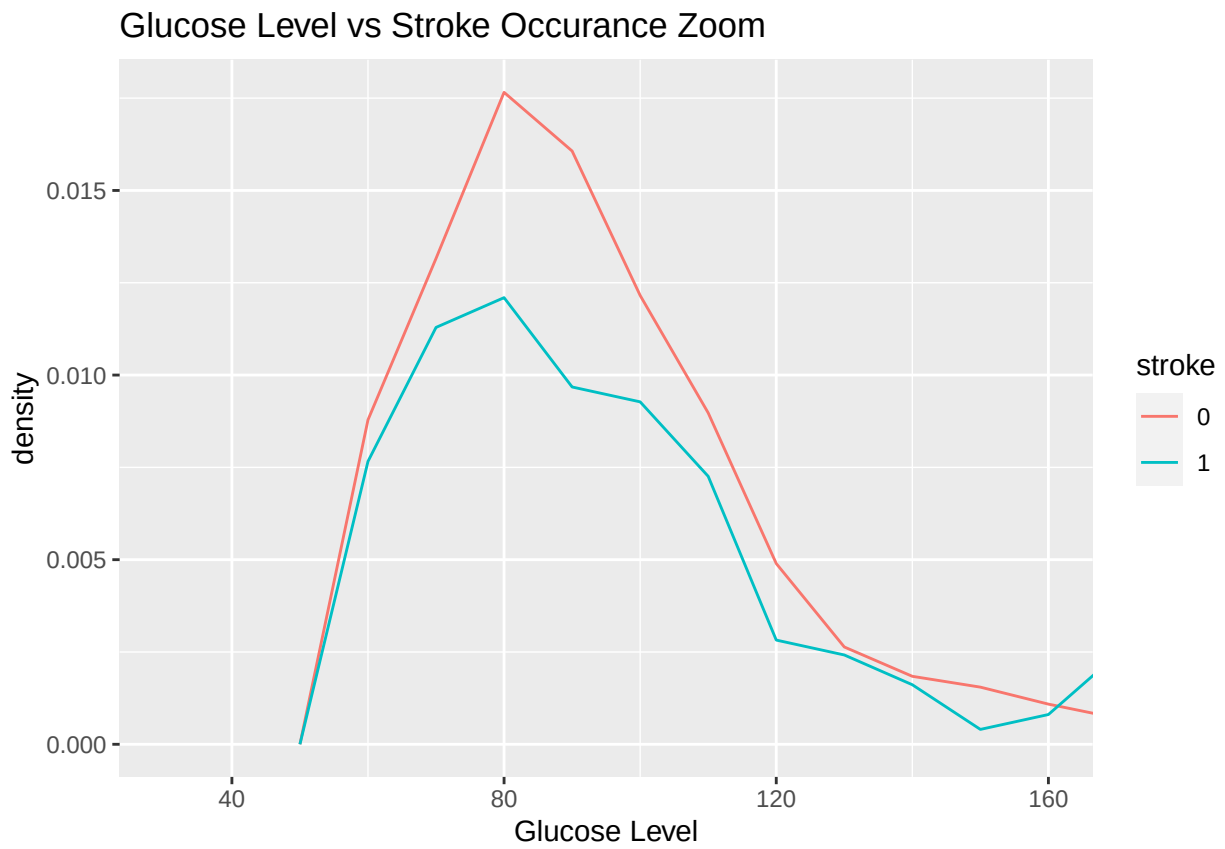
Further, we displayed a displayed of the age distribution of individuals experiencing stroke grouped by gender using a boxplot in figure 4. From this output and after further analysis, we noticed that women have a greater distribution in terms of age. The reasoning behind this is that women have unique risk factors such as having high blood pressure during pregnancy which causes younger mothers to be exposed to strokes. As a result, in increases the interquartile range on the boxplot representing womens' stroke-age distribution.

Glucose Levels vs Stroke Status



The plot above shows density graph of individuals experiencing stroke with the independent variable being the individuals. From this plot we can see two major observations:

- A big spike where glucose levels are considered normal
- Another spike where glucose levels are considered diabetic As a result, we can confirm that having a high (unhealthy) glucose level can have an impact on the probabilities on experiencing a stroke



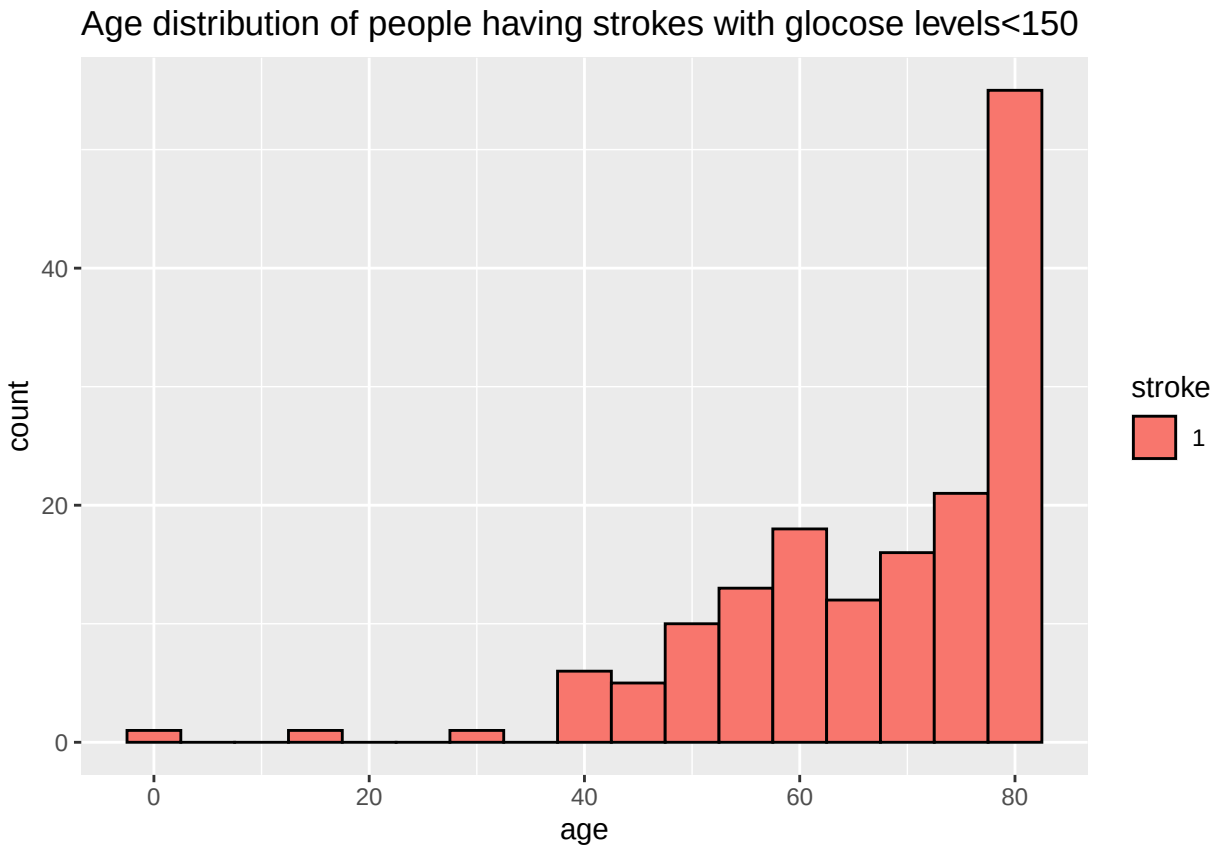
If we focus on the individuals who are experiencing stroke at normal levels of glucose (in figure 6), we can ask ourselves further questions giving us an opportunity to do further research and answer those questions

- How come people are getting strokes while having normal glucose levels
- We want to further analyze this observation

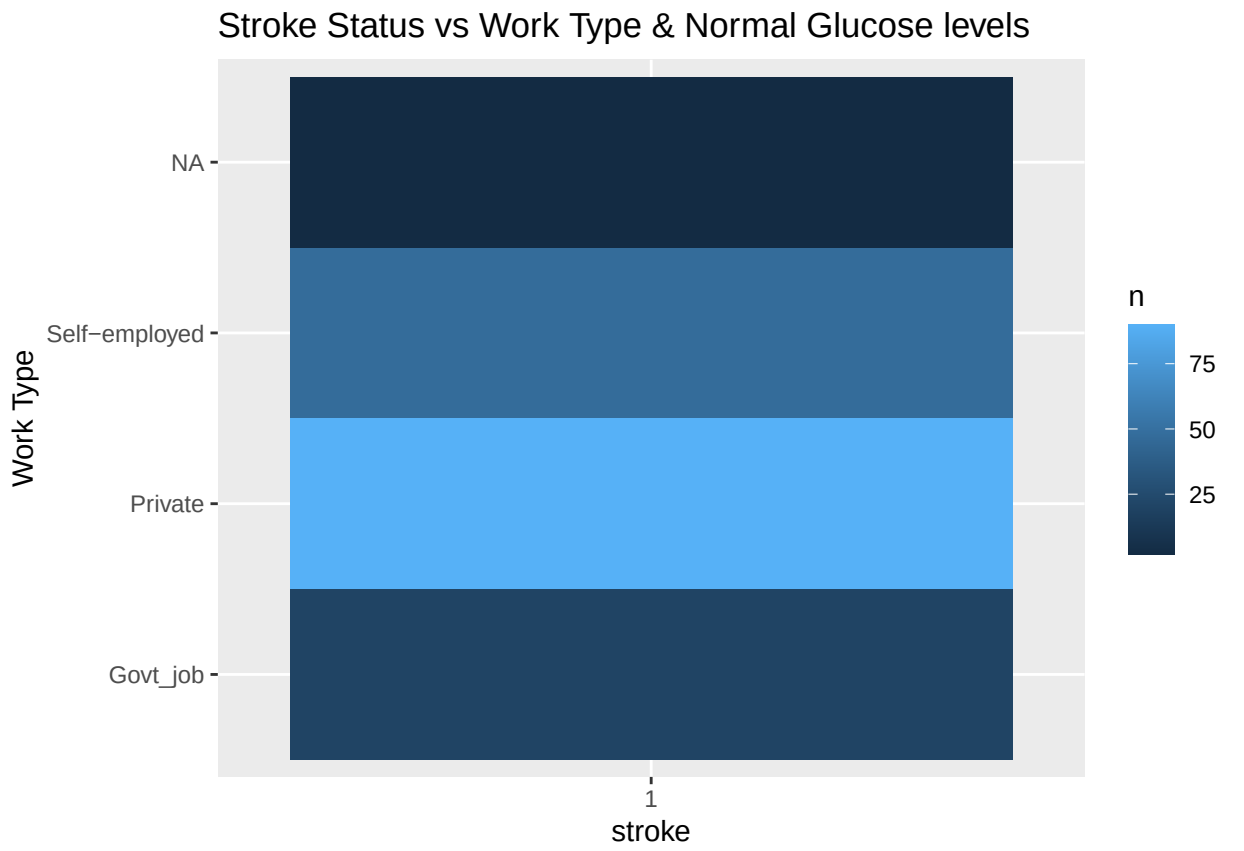
What is Causing Strokes Amongst Individuals with Healthy Glucose Levels

We filter the individuals that experiencing strokes while having glucose level less or equal to 150 (normal levels) using the following dpyr function and conduct further analysis.

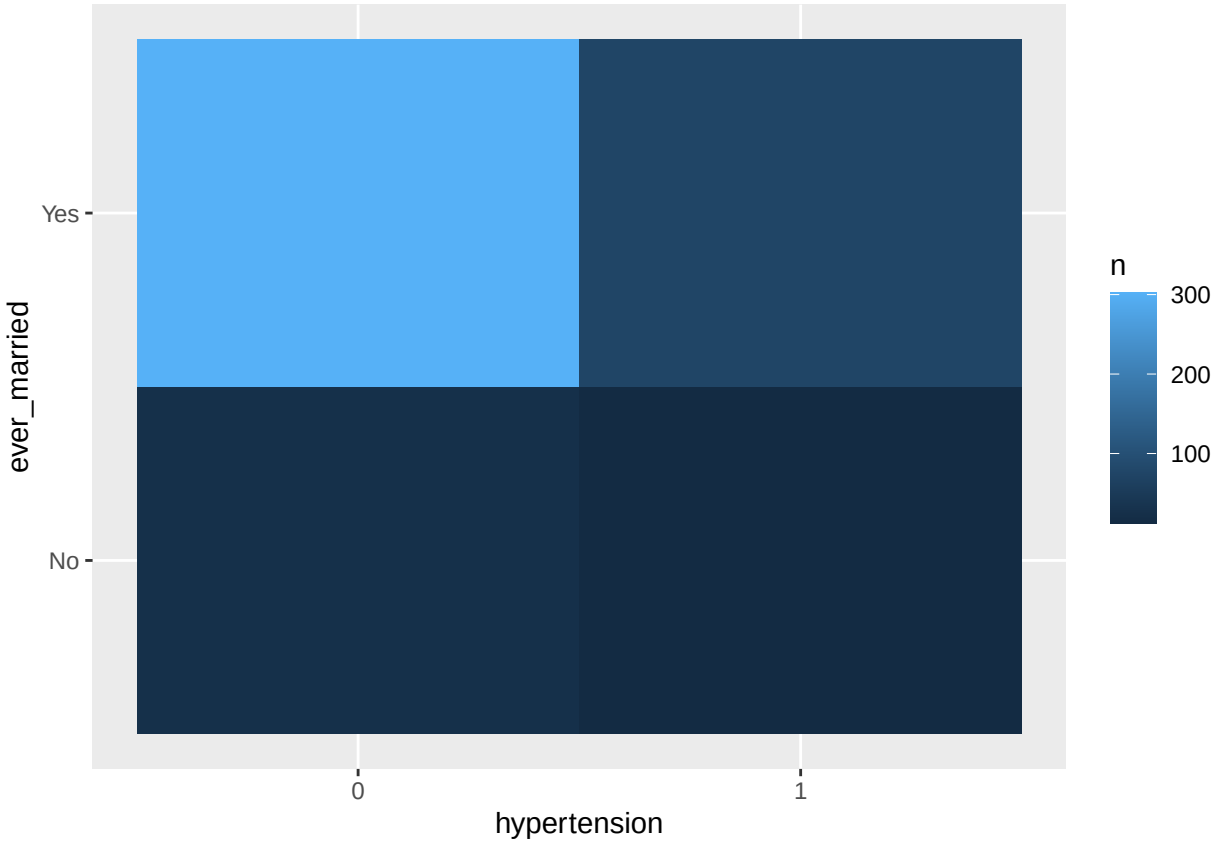
```
#-->for individuals hove have blood glucose levl below 150 and stroke==yes
strokeAndlessThan150<-filter(stroke_data, avg_glucose_level<150 & stroke==1)
```



The age distribution amongst people having normal (healthy) glucose levels while experiencing stroke is plotted again using a histogram. We can conclude that age itself is a big factor in an individual experiencing a stroke or not.

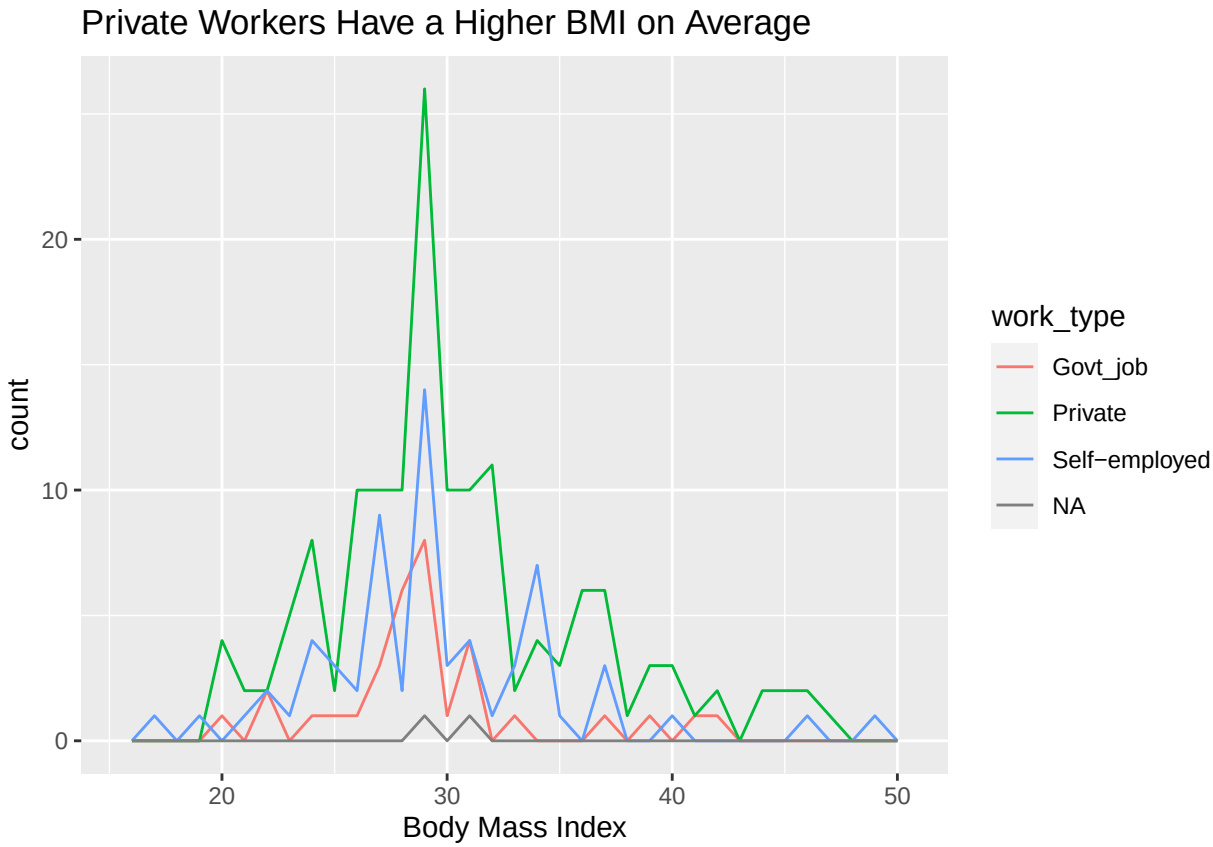


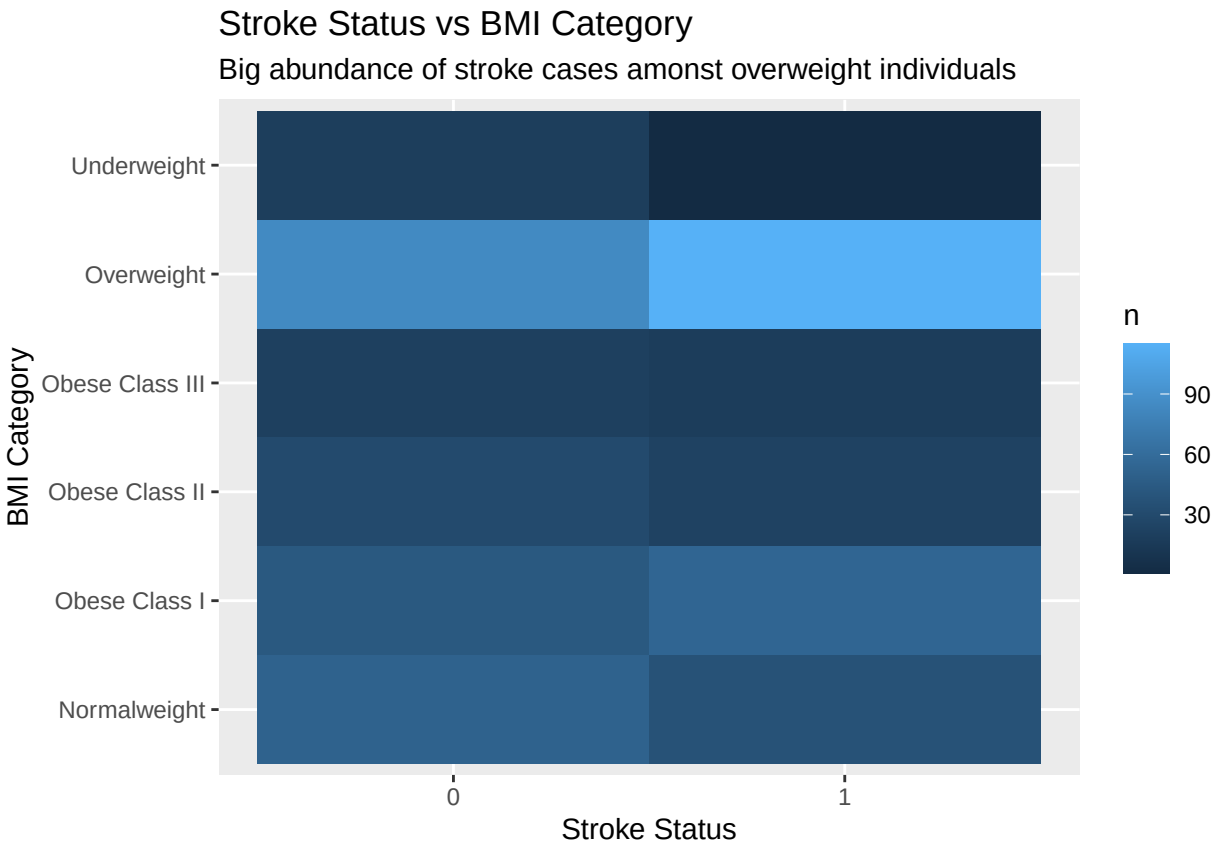
From the heatmap in figure 8, we can see that amongst people having normal (healthy) glucose levels and still experiencing strokes, a majority of them work in the private sector.



Finally, amongst people having normal (healthy) glucose levels and still experiencing strokes, a majority of them are married. Marriage could impact stress levels of some individuals, which could essentially lead to them living unhealthy lifestyles which could result in strokes.

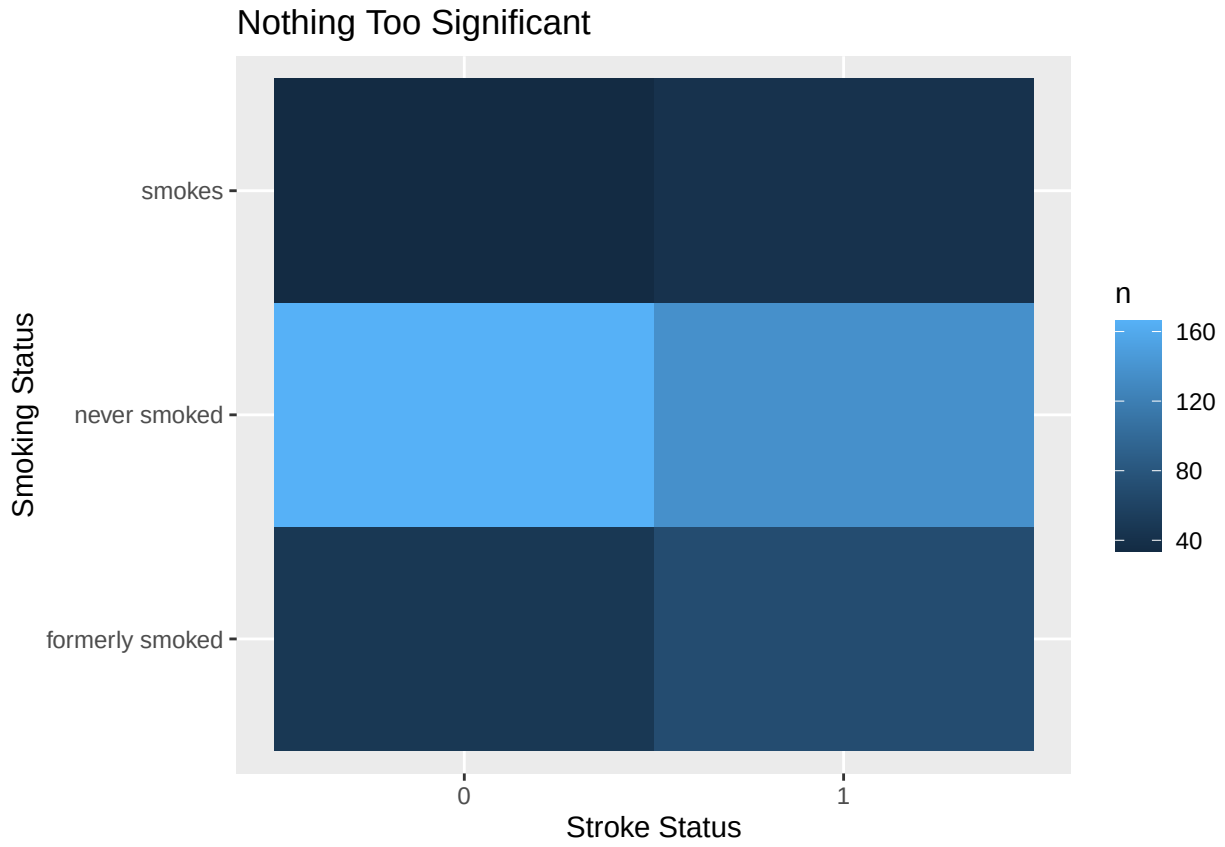
Work Type and BMI Relationship





We can see that on average, private workers seem to have a higher BMI than other workers. There is a spike of private workers at around BMI of 29 which is considered "high" for the average human being. After categorizing BMI levels, it is concluded that strokes are most common amongst overweight individuals.

Smoking & its Impact on Strokes



Based on out dataset, we cannot say that smoking increases chances of experiencing a stroke since there are frequent stroke cases amongst individual who don't smoke

Our Model Choice

- Random Forest - 92.63
- Logistic Regression - 86.28
- K Nearest Neighbors - 86.99

##Our Shiny Web Application

https://datamechanics.shinyapps.io/brainstrokepredictor