Understanding Stroke and exploring its causal factors

MOSES AMOAKU

Faculty Advisor: Dr. JOSEPH DEMAIO

INTRODUCTION

Stroke is a disease that affects the arteries, often leading to and affecting the brain.

According to the American Stroke Association, It is the No. 5 cause of death and a leading cause of disability in the United States. Research shows that, for every 40 seconds, someone in the United States is diagnosed of stroke and for every 3.5 minutes, someone dies of stroke.

The World Health Organization (W.H.O) in 2019 reported that stroke is the 2nd most cause of death in the whole world responsible for 11% of all deaths.

Risk of having a first stroke is nearly twice as high for Blacks as for Whites, and Blacks have the highest rate of death due to stroke.

The purpose of this study is to examine the various health, environmental, economical, and social factors that causes stroke, by observing some selected variables.

METHODS

The dataset for the project was obtained from Kaggle (source provided below). This study is determining factors that could potentially contribute to a person having stroke. The variables considered includes, BMI, Age, residency status, smoking status, gender, work type etc.

During the analysis, only patients above age 20 were considered. I used a proc freq in SAS to determine the percentage and number of people from each group who had stroke. I also used correlation analysis to determine the strength of these associations (example age and glucose level).

I used tableau to give a visual representation of these relationships to make it easier to understand

RESULTS

The analysis revealed that almost all the variables were associated with stroke. However, some factors (variables) did have a bigger impact on stroke than others. About 56% of women had strokes compared to 43% of the men. This translates that women risk a higher chance of getting strokes compared to men.

Individuals who had private jobs also had a higher percentage of stroke (60%) compared to those who were self-employed (26%) and even less for individuals who worked government jobs (13%).

stroke=0

smoking_status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Unknown	754	19.65	754	19.65
formerly smoked	784	20.43	1538	40.07
never smoked	1582	41.22	3120	81.29
smokes	718	18.71	3838	100.00

Residence_type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Rural	1885	49.11	1885	49.11
Urban	1953	50.89	3838	100.00

gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	2337	60.89	2337	60.89
Male	1500	39.08	3837	99.97
Other	1	0.03	3838	100.00

work_type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Govt_job	610	15.89	610	15.89
Never_worked	1	0.03	611	15.92
Private	2490	64.88	3101	80.80
Self-employed	737	19.20	3838	100.00

hypertension	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	3408	88.80	3408	88.80
1	430	11.20	3838	100.00

heart_disease	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	3610	94.06	3610	94.06
1	228	5.94	3838	100.00

ever_married	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	712	18.55	712	18.55
Yes	3126	81.45	3838	100.00

Sheet 2 — Residence type / Gender / Work Type

Sheet 1 — Ever Married / Smoking Status

Scatter Plot With 95% Prediction Ellipse
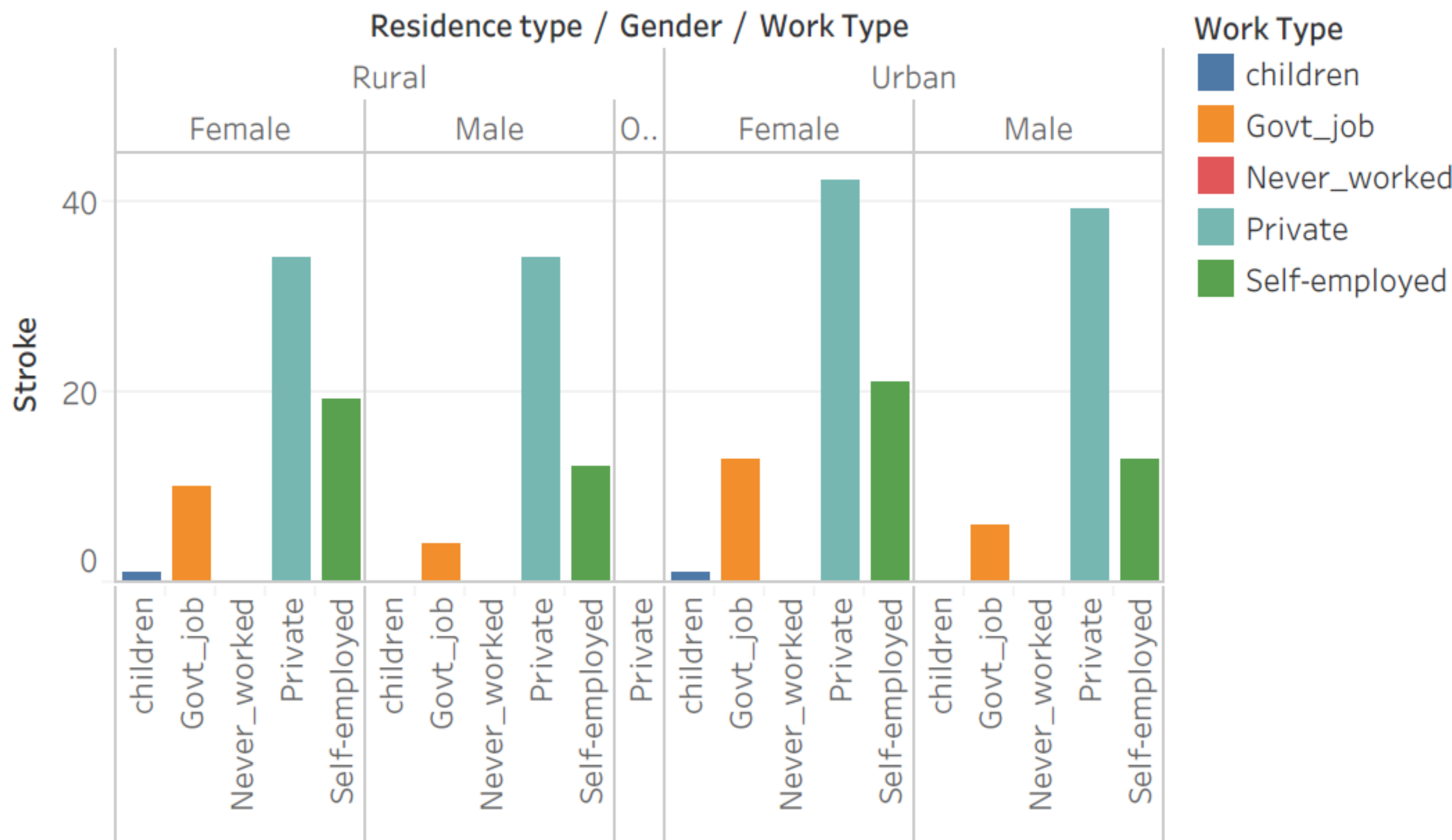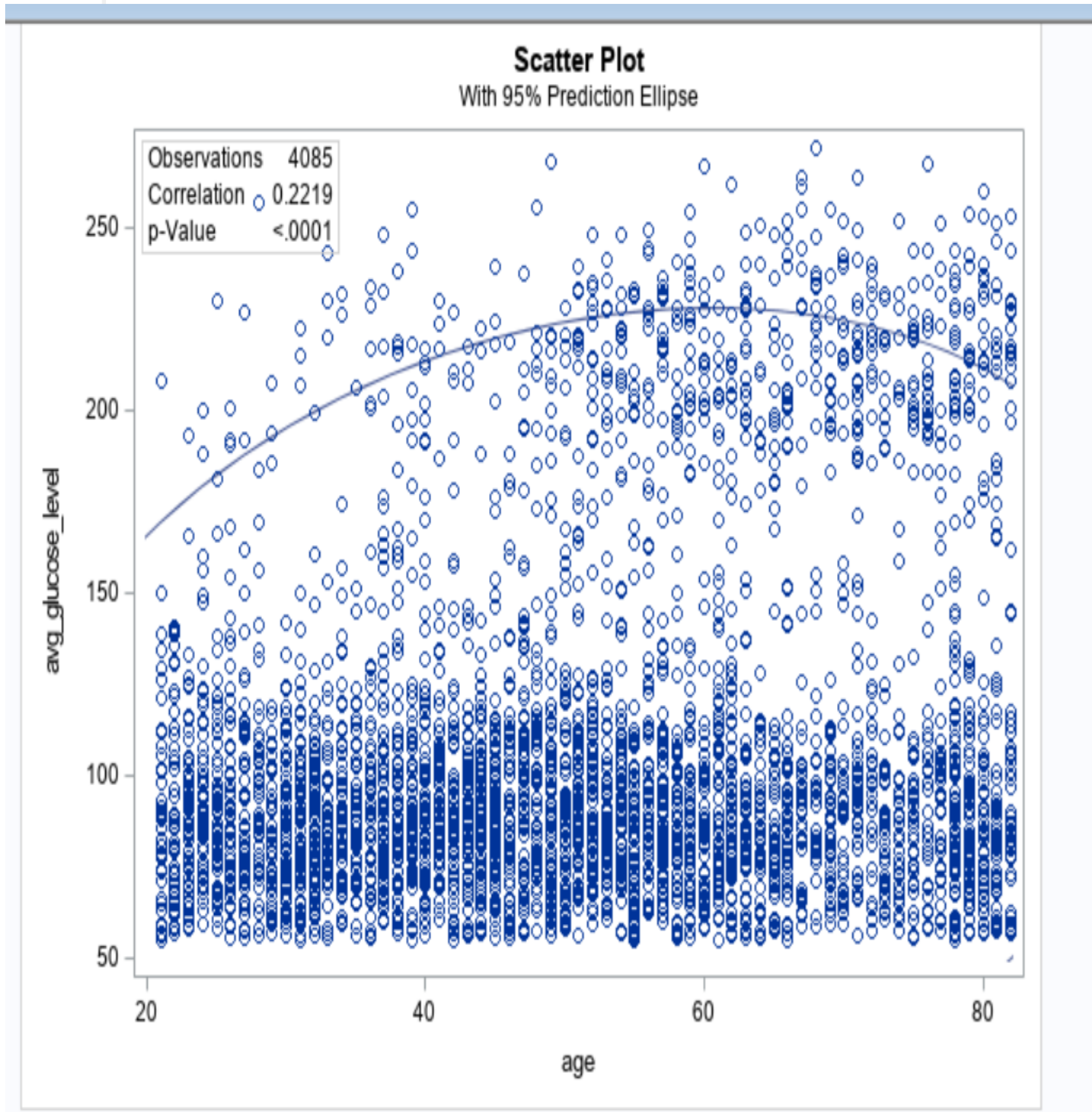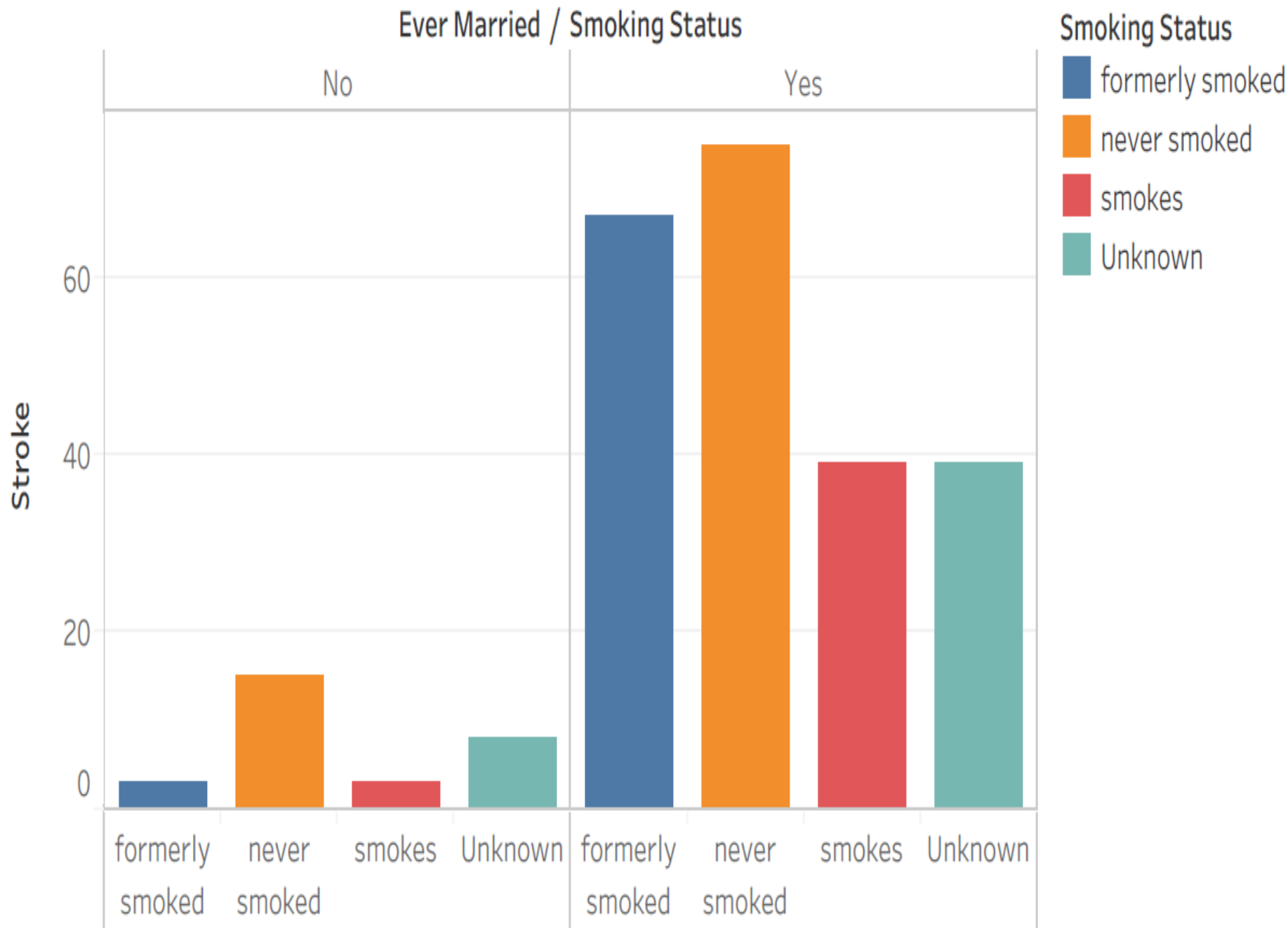
For people who reside in urban areas, 54% of them recorded having stroke compared to compared to those who lived in rural areas (45%).

I used tableau to run different tables for stroke against gender, work type, residency, smoking status, marriage status and residence type as well and it turns out that people who are married, or ever married and formerly smoked had a higher chance of getting stroke compared to individuals who never married even if they formerly smoked. The data also revealed that one can get stroke even if they haven't ever smoked. And, irrespective of one's gender or residency type, people who worked private jobs risk a very higher chance of getting stroke.

I also run a Pearson correlation analysis to understand how the individual variables are also correlated, and it turns out that a person's age is correlated with their glucose level, however, this association isn't strong (Pearson correlation of 0.239).

CONCLUSION & RECOMMENDATION

Stroke can be caused by several factors such as race and even genetics. However, some environmental, economic and health factors can also play a huge role in a person getting stroke. The data showed that even though older people stand a higher chance of getting stroke, the odds for younger people cannot be ruled out. The same ca be said for women and men as well.

It is therefore highly imperative for individuals to be educated on the need to pay attention to these environmental, health and economic factors/risks to help reduce the risk of stroke. More awareness needs to be created among the youth (especially) so they have prior knowledge and understanding of stroke and some of it causal factors.

SAS CODE

```
proc freq data=work.project;
tables smoking_status residence_type
gender work_type hypertension
heart_disease ever_married;
by stroke; Run;


PROC CORR DATA=work.project PLOTS=SCATTER(NVAR=all
    VAR age avg_glucose_level; RUN;


proc Reg data=work.project;
    model avg_glucose_level = age;
    run;
```