A STUDY ON HEALTH INSURANCE CHARGES USING R PROGRAMMING : With special reference to Customer's Demographic Profile of USA.

"We are speaking for the data"

Presented by:

R. Gobiga - Introduction

P. Jone Andrew Harris - Analysis

M. Jeevabharathi. - Findings and Conclusion

Agendas

>Introduction

>Distribution of the data

>Analysis

>Findings

>Conclusion

>References

Objectives

> To study the customer's health behaviour in health insurance
>To study the customers demographic profile.
> And give suggestions to the company.

INTRODUCTION

Health Insurance is a type of insurance that covers medical expenses that arise due to an illness.These expenses can be related to hospitalisation costs,cost of medicines or doctor consultation fees.A health insurance policy determines the types of medical services or benefits you are covered for ,which doctors you can see, and hospitals you can visit.Your plan also determines what you pay for care and services.Health insurance is for preventive and event-based care that reduces your medical expenses.

Data Introduction:

This project includes the analysis of Health Insurance charges with respect to the customer's health,family,age,region and behaviours. This data contains the information of a health insurance company . This data is collected in the year 2019-20.

Data Context:

This dataset contains 1338 rows of insured data, where the Insurance charges are given against the following attributes of the insured: Age, Sex, BMI, Number of Children, Smoker and Region. There are no missing or undefined values in the dataset.

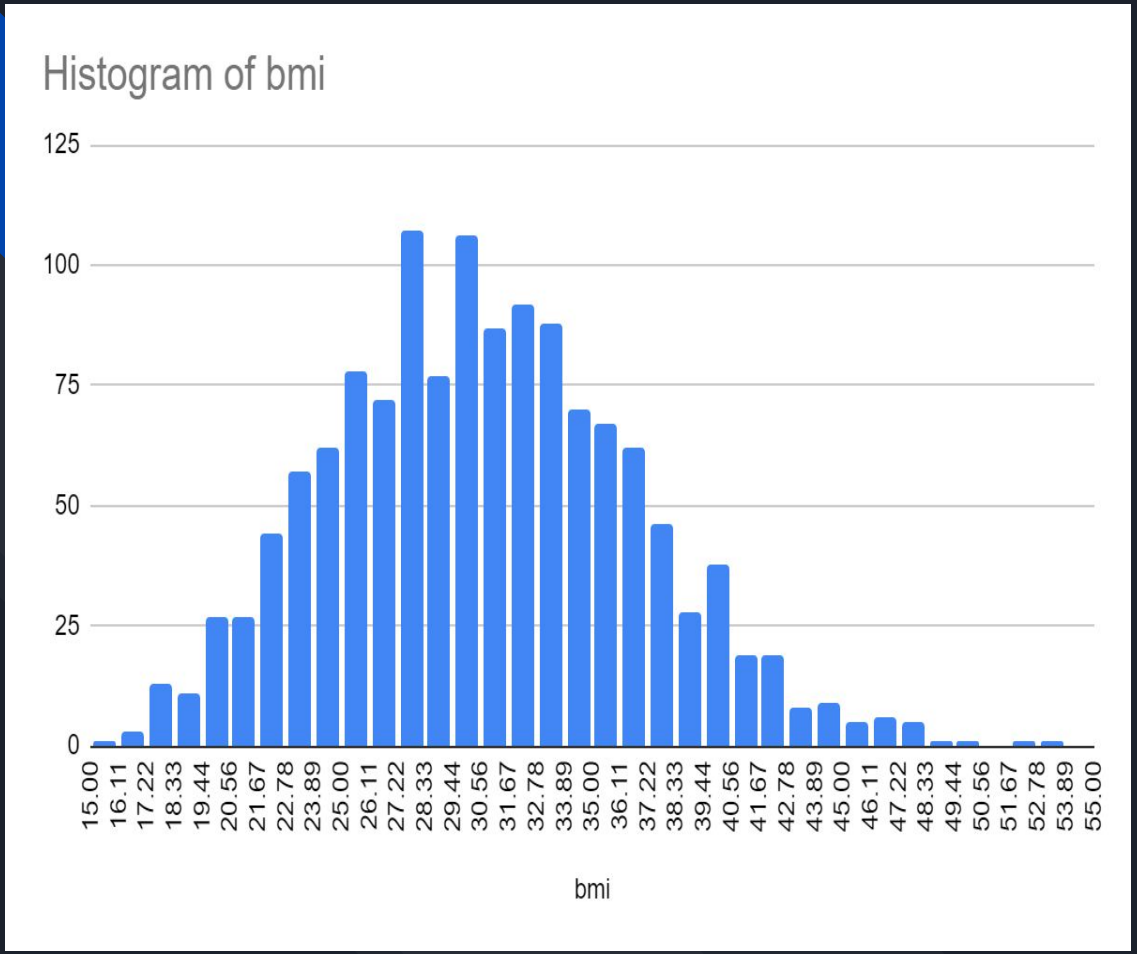
Data preview

Age	Gender	BMI	Children	Smoker	Region	Charges
19	female	27.9	0	yes	southwest	16884.924
18	male	33.77	1	no	southeast	1725.5523
28	male	33	3	no	southeast	4449.462
33	male	22.705	0	no	northwest	21984.47061
32	male	28.88	0	no	northwest	3866.8552
31	female	25.74	0	no	southeast	3756.6216
46	female	33.44	1	no	southeast	8240.5896

The total charges collected by the health insurance company 17.75M.

Number of customers 1309.

Number of children cover by the insurance 1430.



Bmi Distribution

This is the BMI distribution of a health insurance company that shows the health rate of the customers.

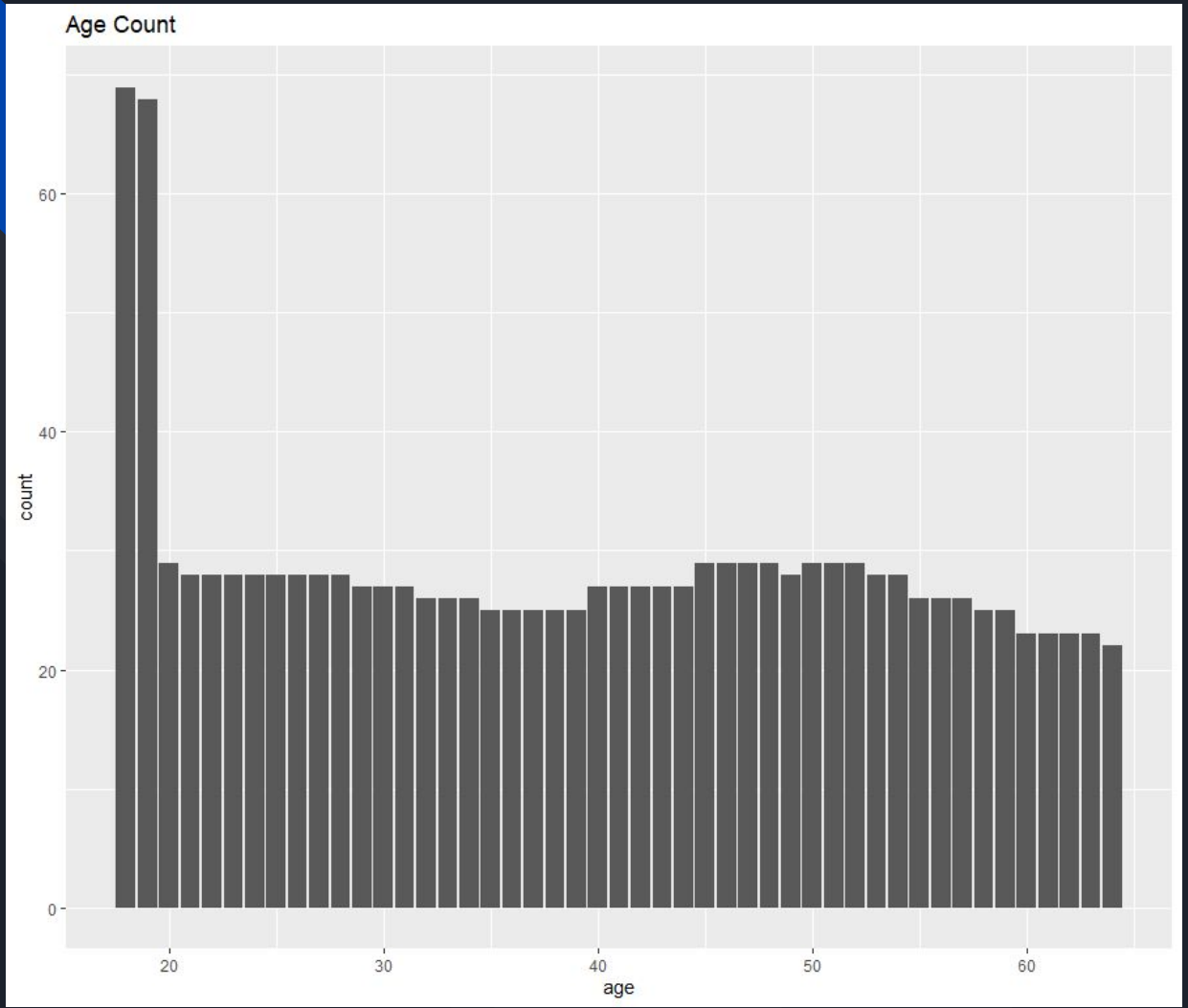
Let’s see how the data is distributed

The healthy rate of BMI is between 18.4 and 24.9

But here the the Average BMI is 30.62

The bmi rate is mostly unhealthy which exceeds the min BMI of 24.9

The formula for calculating BMI is weight in kg divided by square root of height in metres i.e, kg/m^2.

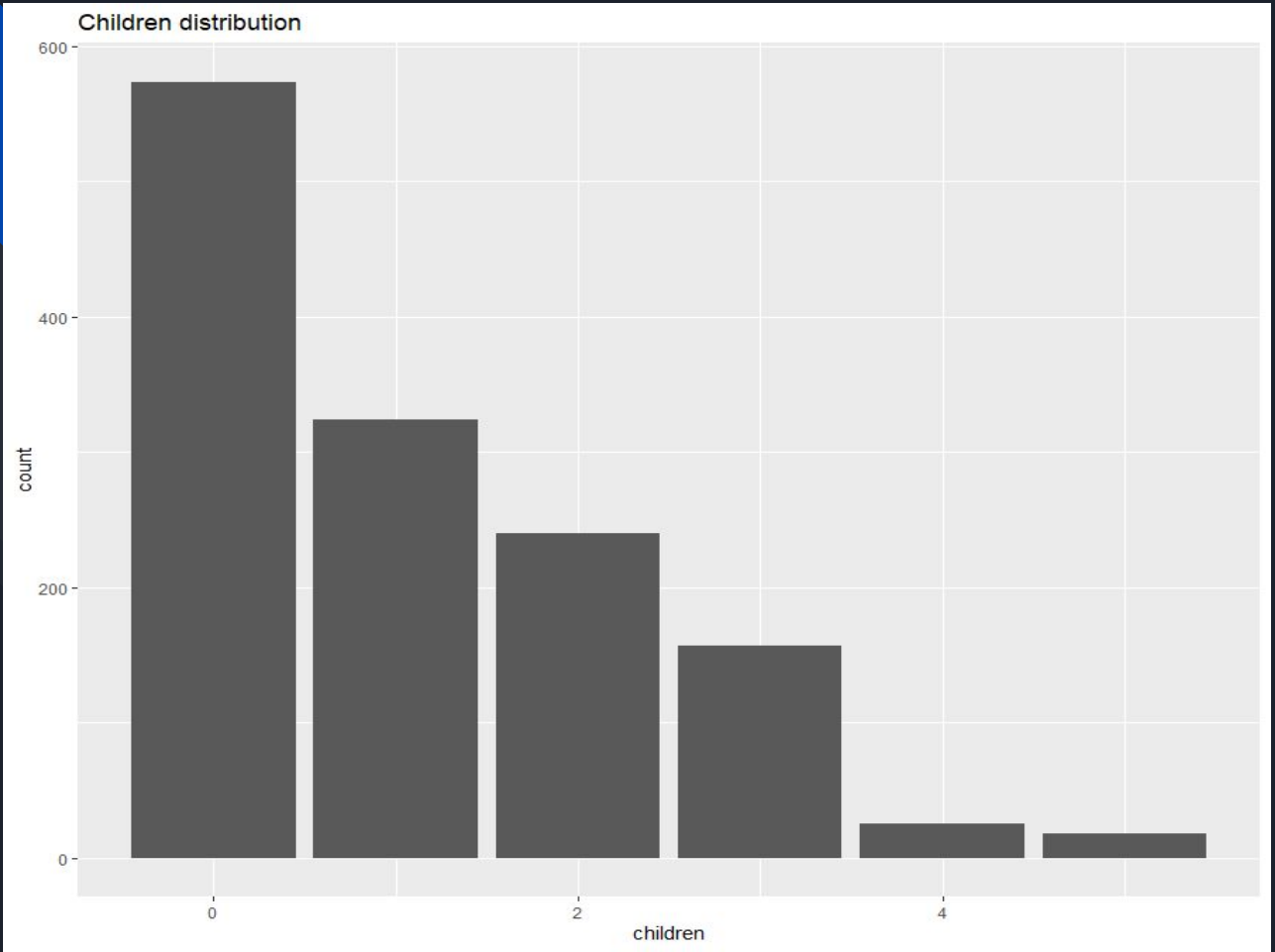


Age distribution

Here we can see most of the customers are in the age between 17 and 20 with the ratio of 1:12.

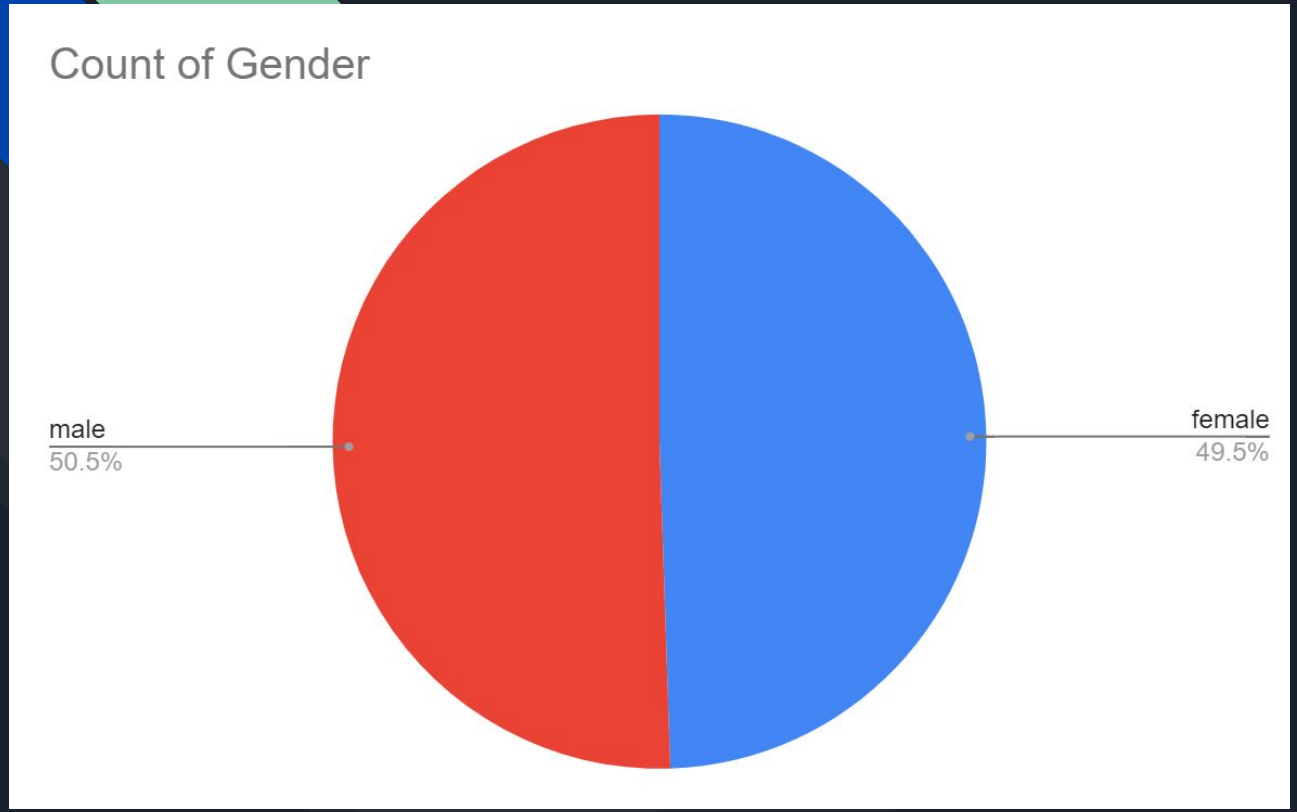
Then the rest of them are in the normally distributed.

Distribution of children



Here we can see that the people having no children are more and people having more children are low in number.

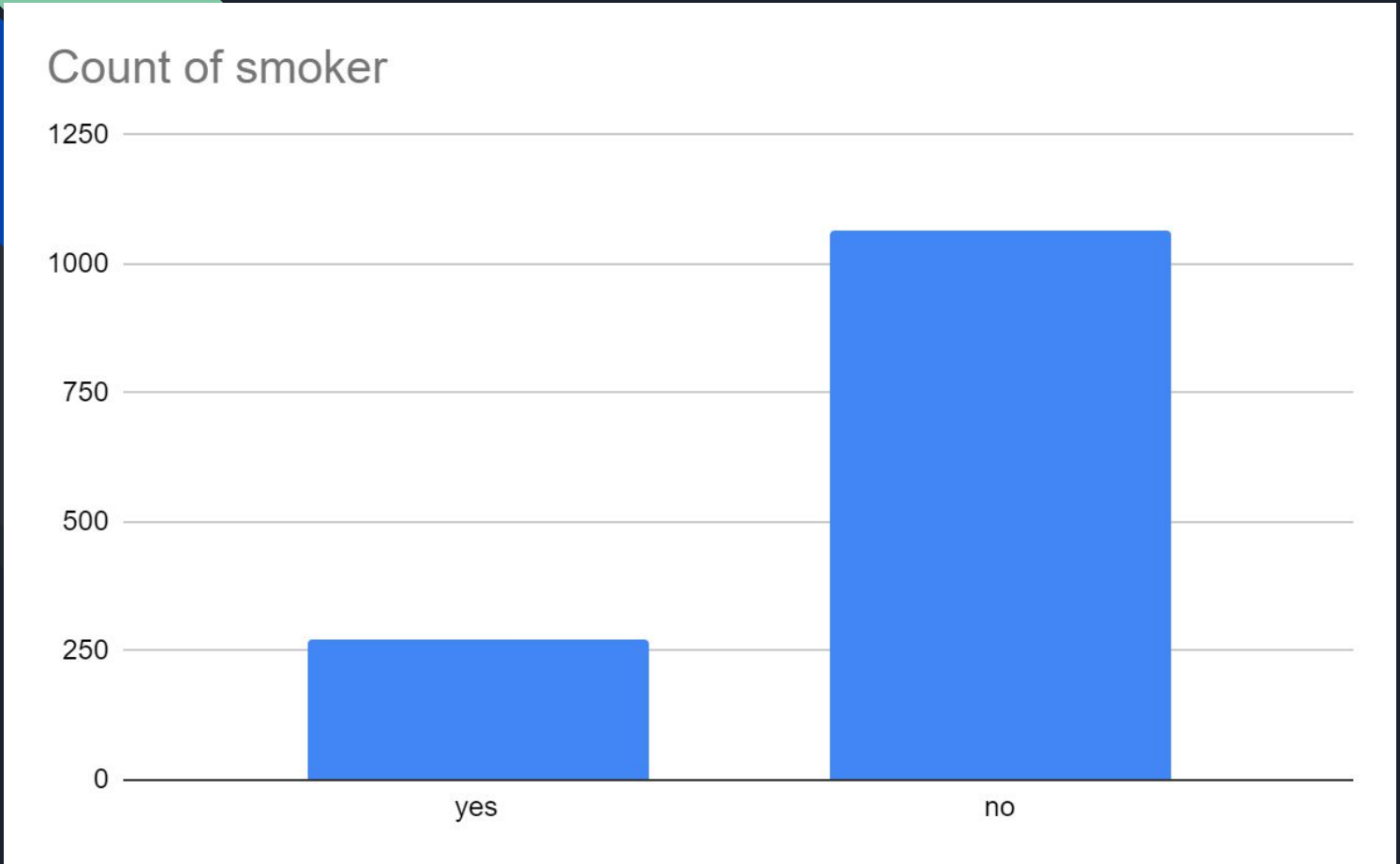
Distribution of Gender



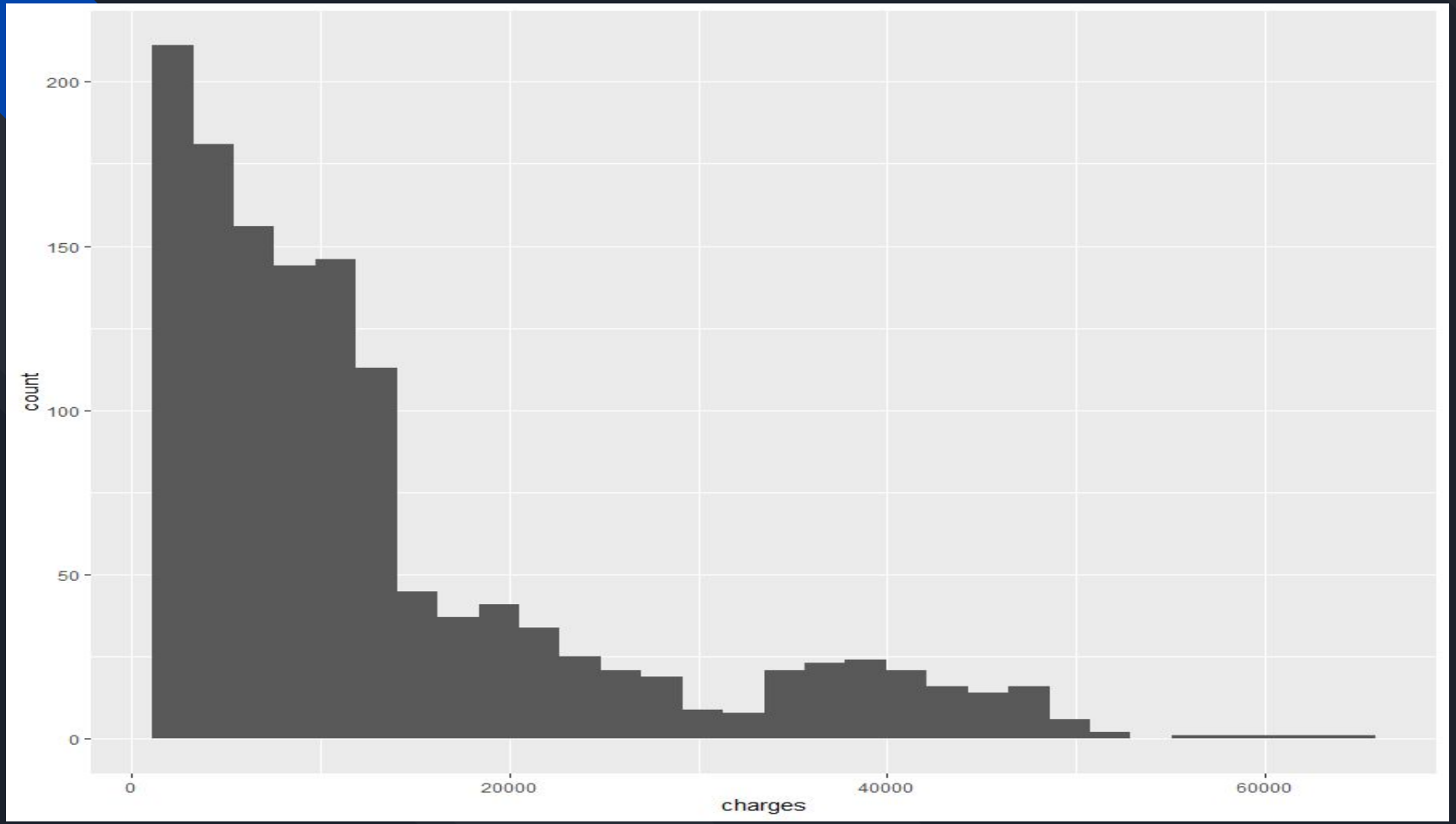
The no.of males are 662 and the no. of females are 646.

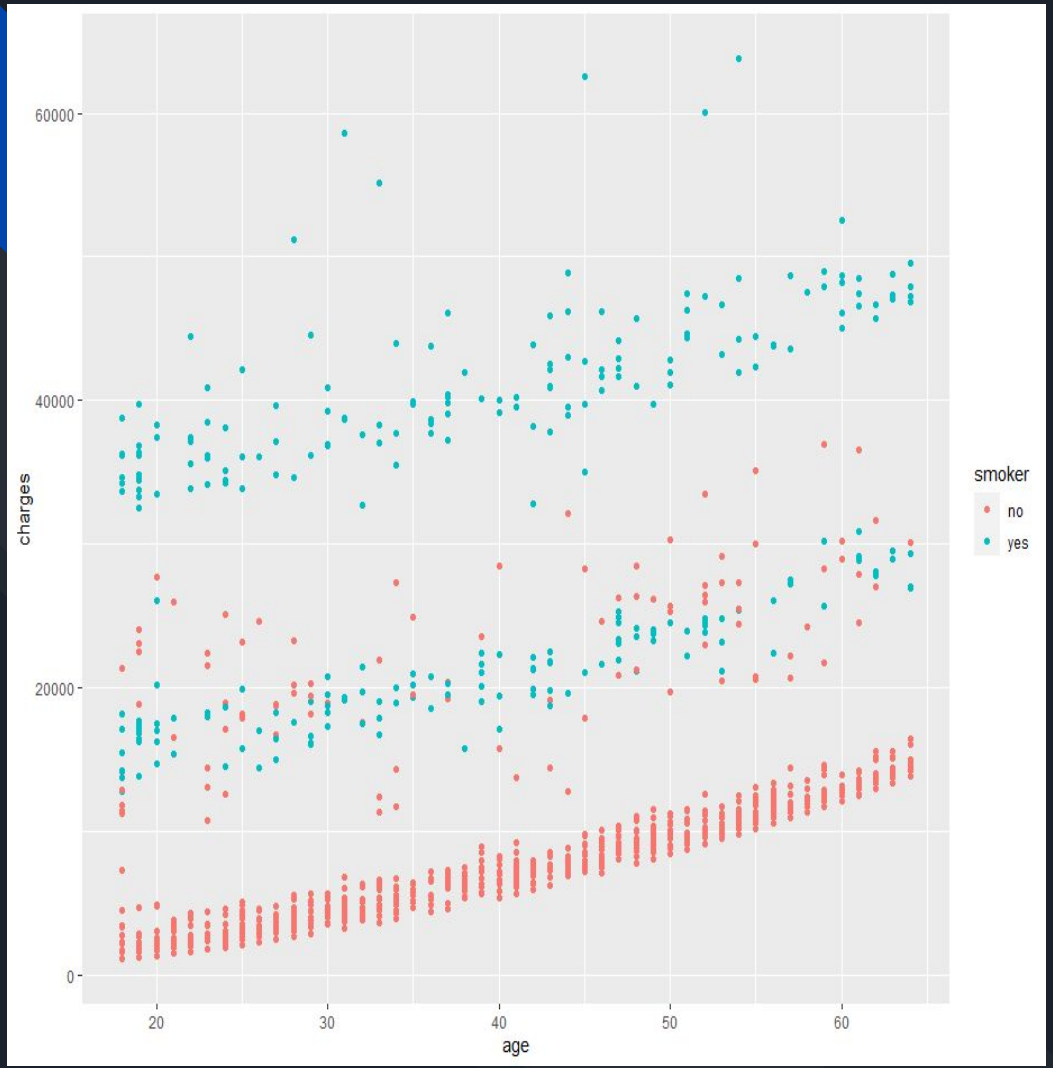
The no.of males are higher than the females by 1%.

Distribution of smokers



Distribution of charges





ANALYSIS

Insurance charges according to their behaviour

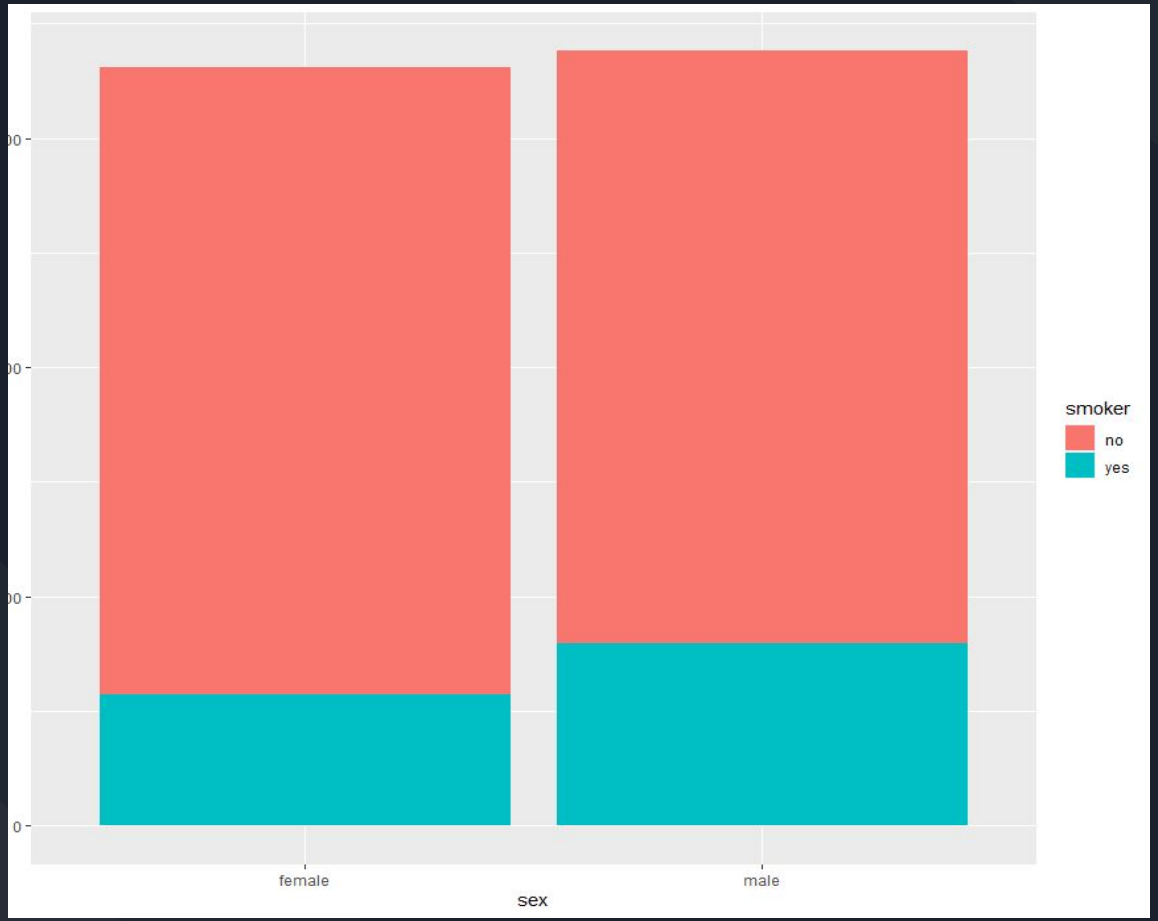
In here as you can see that most of smokers are charged more than the non-smokers

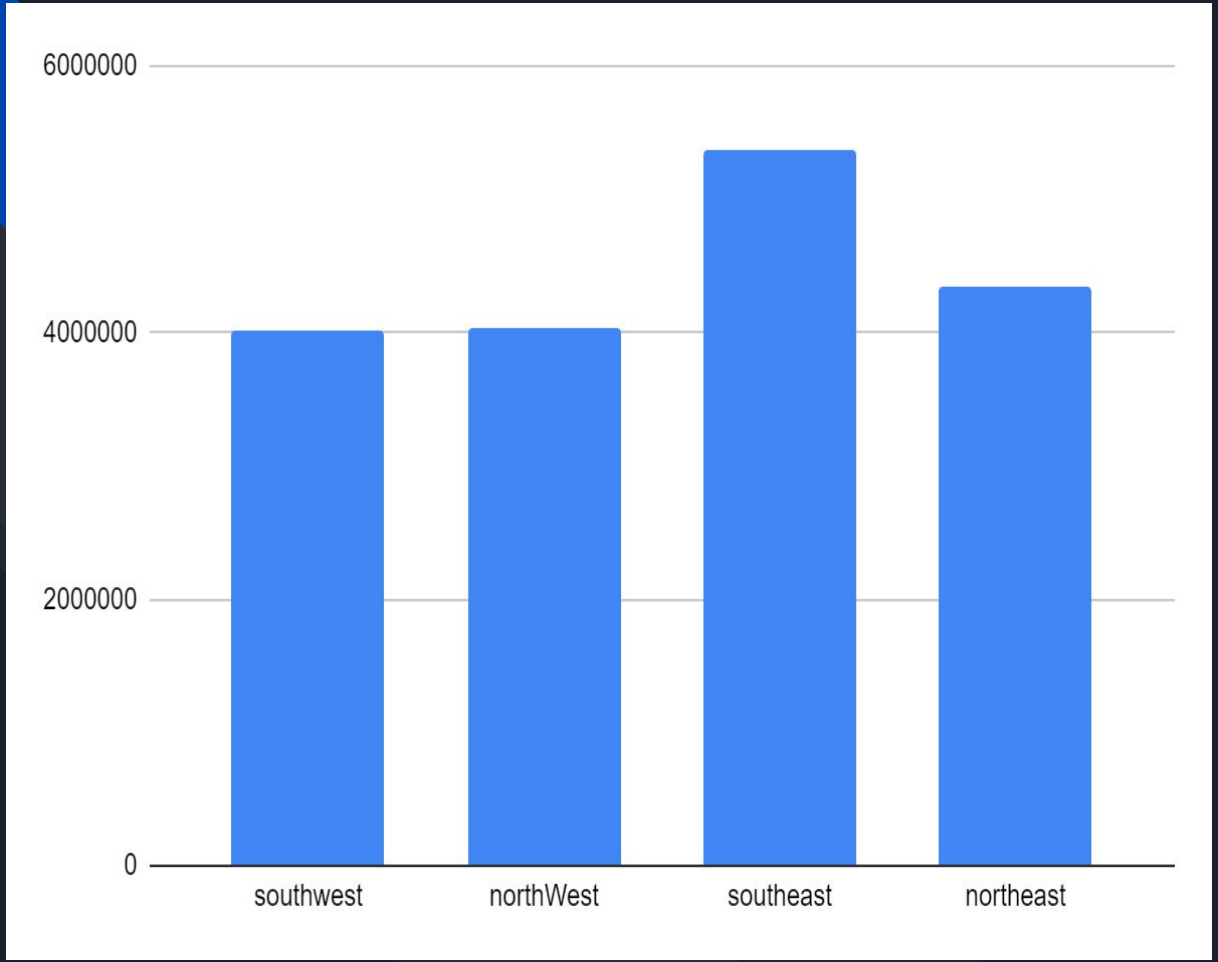
In here the charges on a non-smoker is between 4,000-18,000

Whereas for the smoker the charges is between 26,000 - 40,000

Highest smoker by Gender

The count of male smokers is slightly higher than the female by 7.84%.



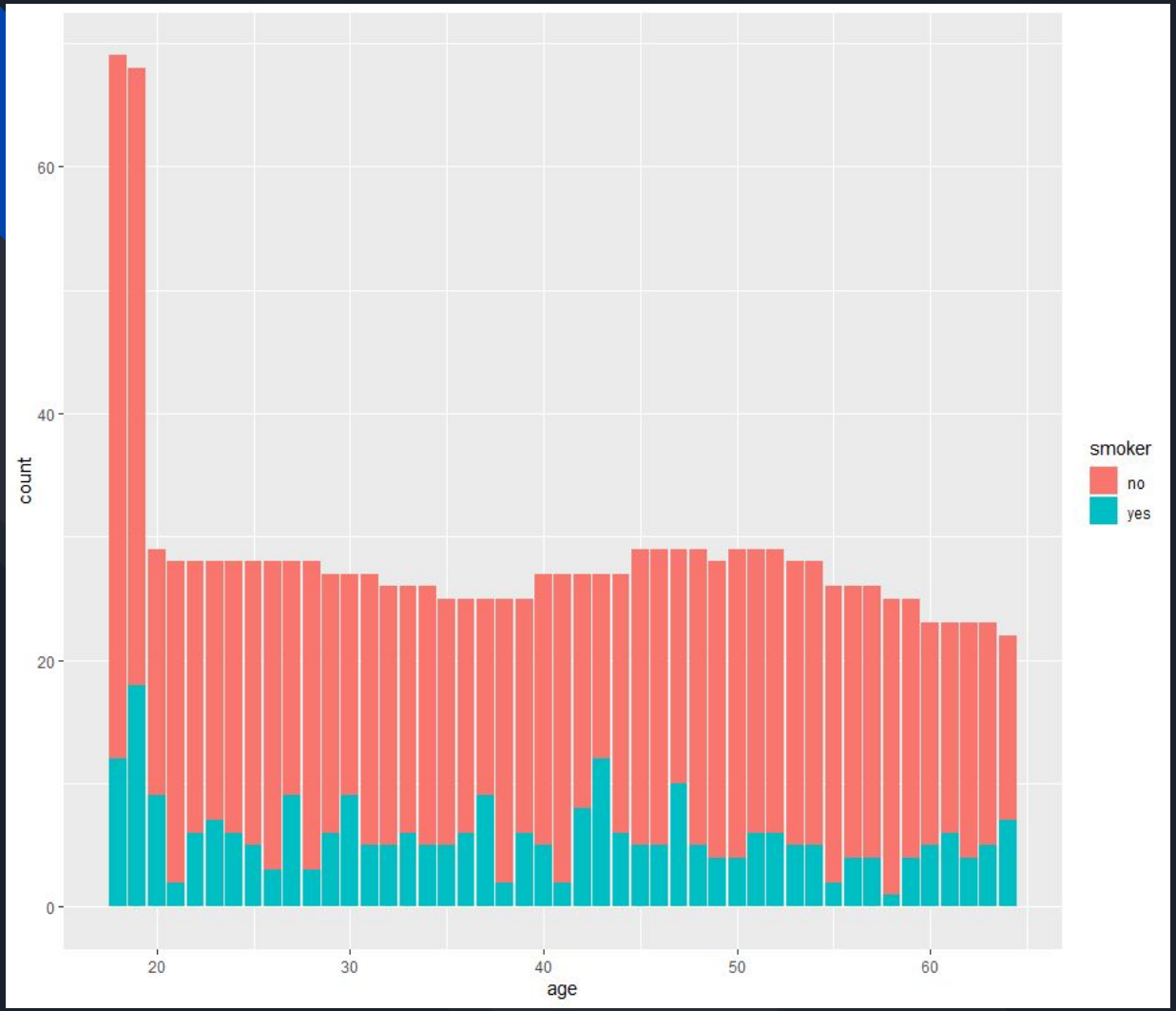


Highest charged region

The highest charged region is the Southeast region because of their High obese rate as see in fig.

The lowest charged region is south-west

Then comes the question Why? For that we need to see the relationship between charges and other factor to see which effects the charges(dependent variable).



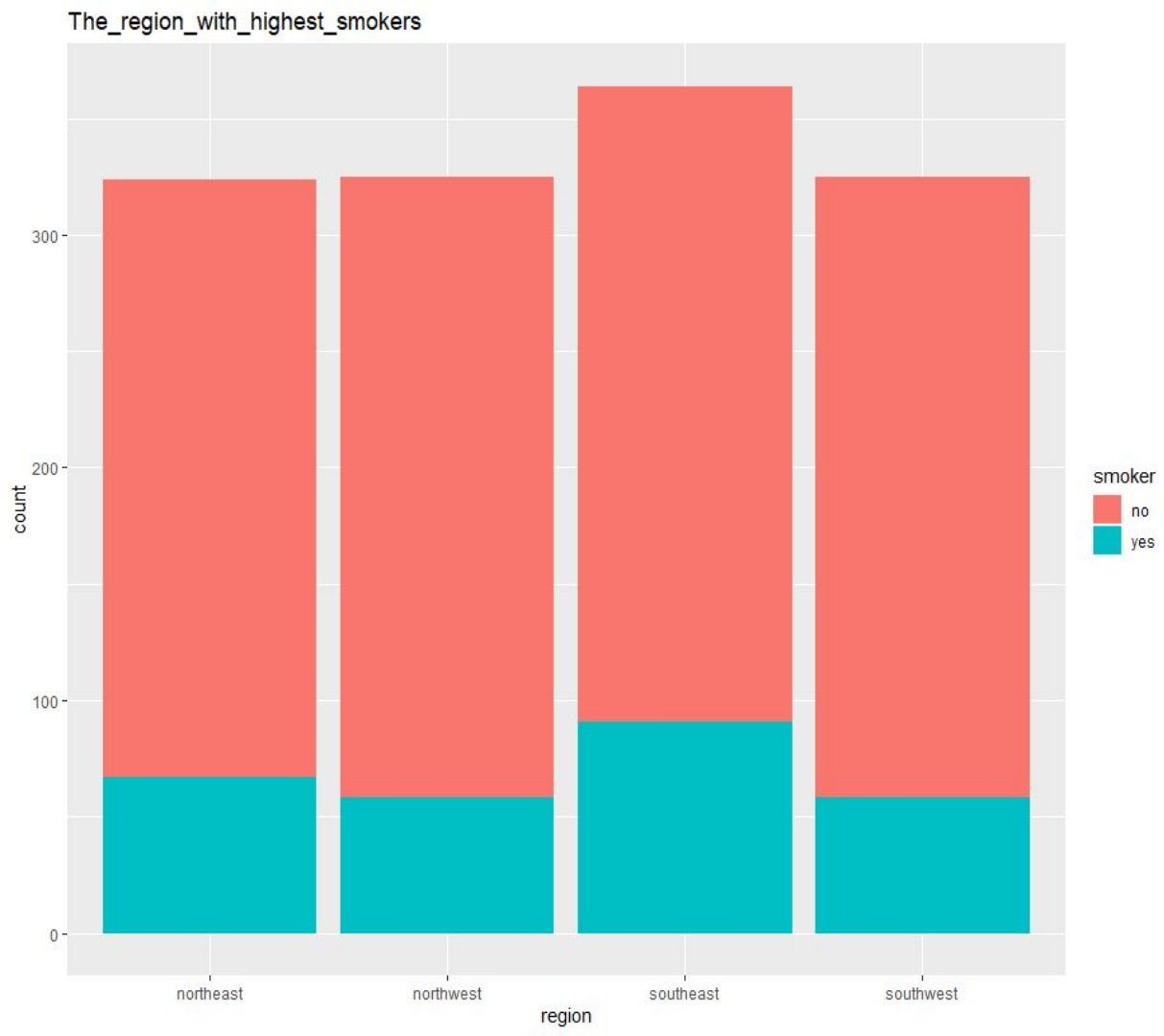
Highest smoker by Age:

Here we can see that most of the people are non-smokers .

Here the ratio of a smoker and a non-smoker is 1:4 approximately.

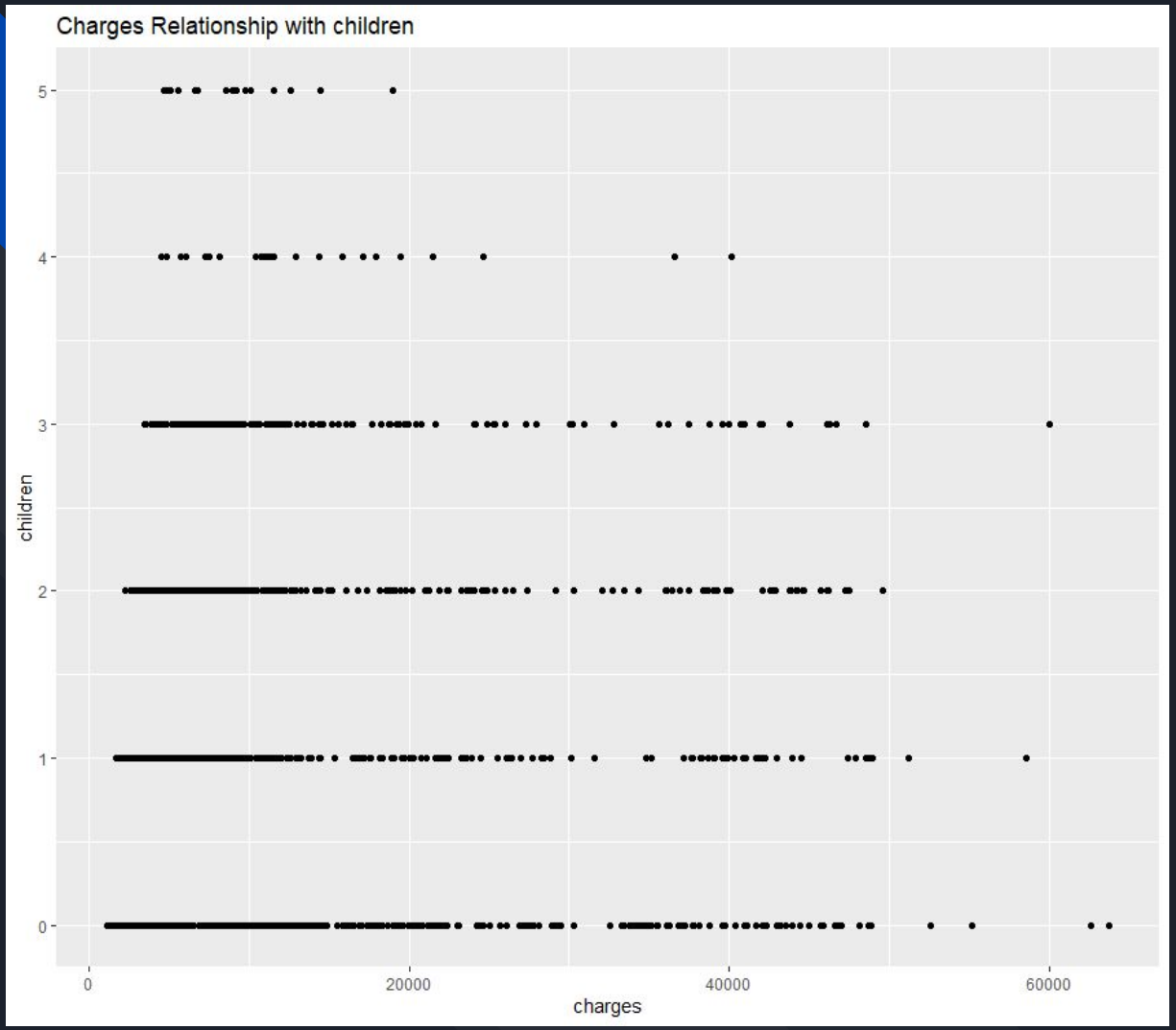
And highest non-smokers are in the age of 18-19 because it is an wrong interpretation, The data entry count of the age 18- 19 is more than the others.

But if we interpret it in the order of ratio, it is 1:12 for non-smokers with the age 18-19. So it is almost the same as the other ages.



Highest smoker by Region:

In this chart, we can see that the people's BMI is high(obese) because the no. of smokers in the south-east region is high.



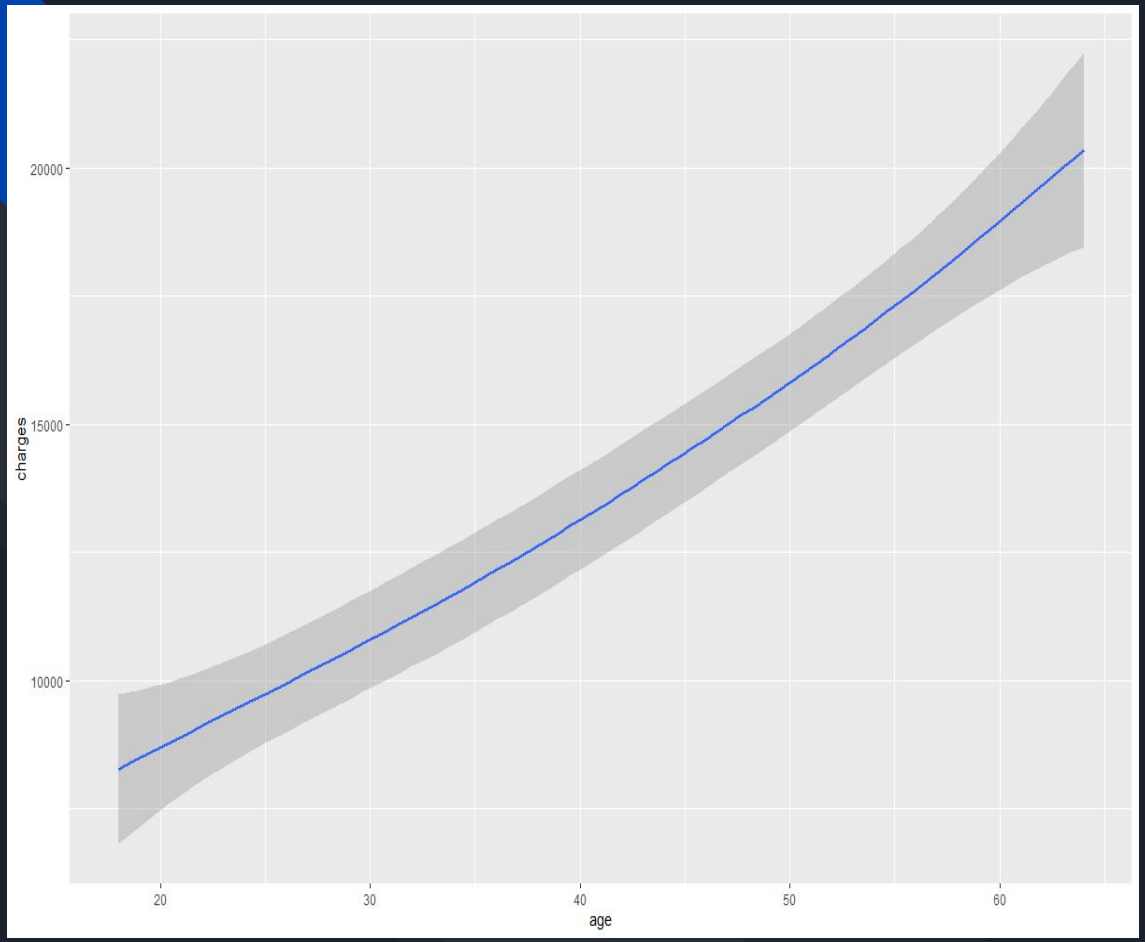
Children's impact on charges:

There is a negative correlation impact on children and charges.

As the children are increasing the charges are decreasing.

But is not clear wheather its a correlation or a causation.

For this we need more data .

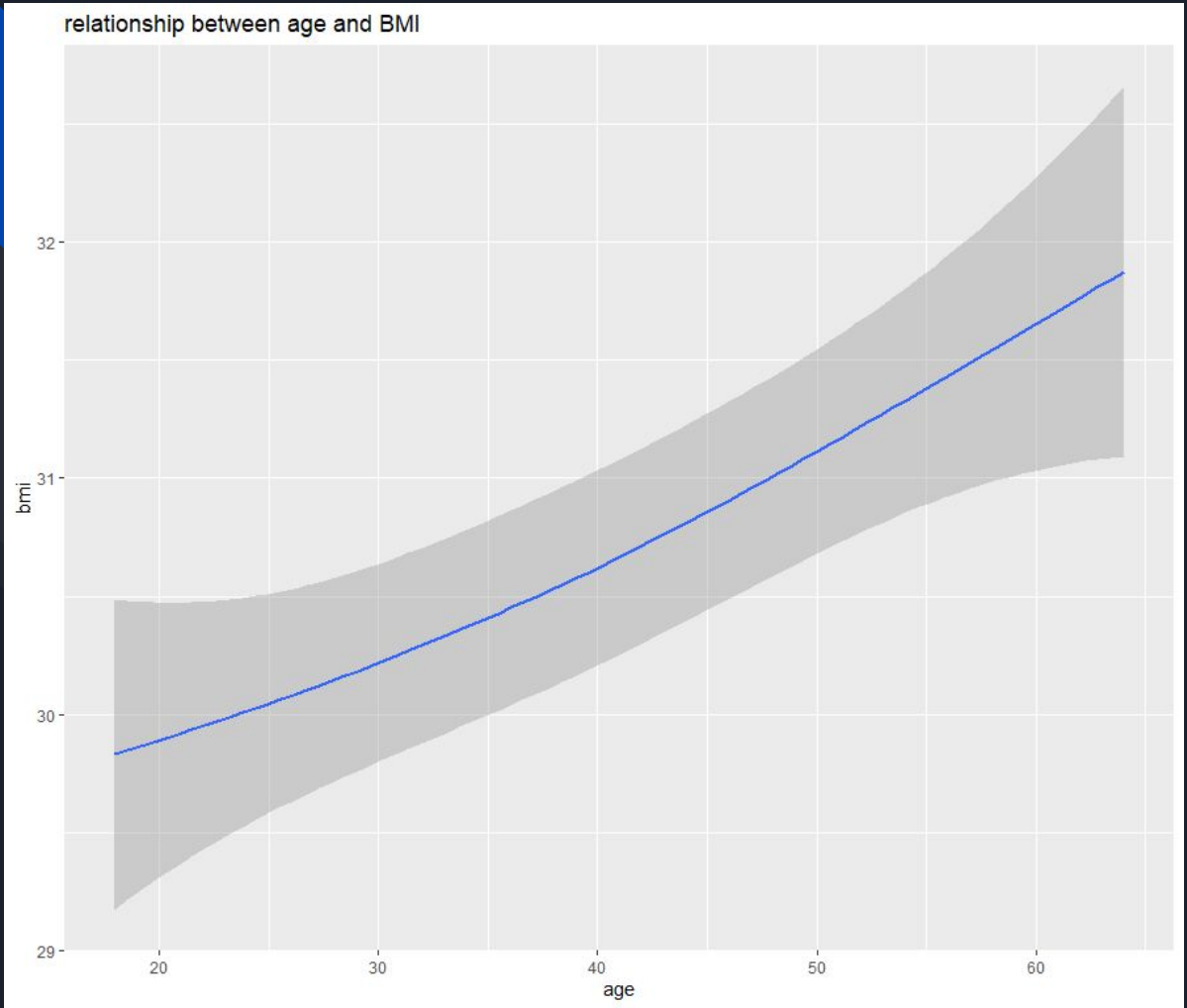


Impact on charges by age :

This particular graph contains many details

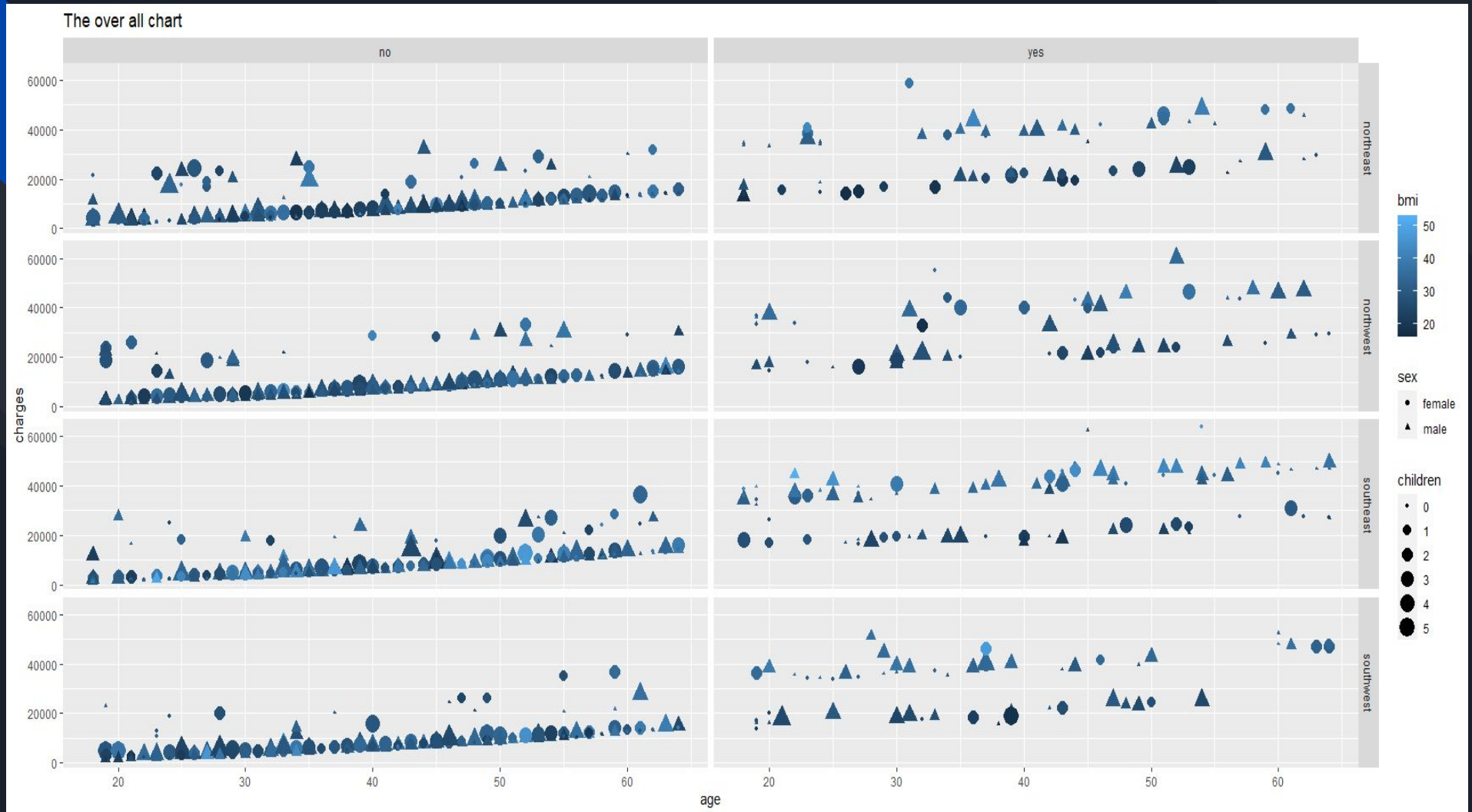
In this graph we can see that there is a positive relationship between age and charges i.e. if the age increases, the charges also increases.

This is because of the relationship between charges and BMI in the previous graph.



Age impact on BMI:

This is an average rate of relationship between age and BMI.
We can see that the values of the people increases as their age increases simultaneously.


The over all chart
no
yes
northeast
northwest
southeast
southwest
charges
age
bmi
sex
female
male
children

Findings:

> Most of the BMI rate is above 24.9, means most of them are unhealthy or obese.
>Their highest customers are among the age between 18-19.
>The most obese region is the south-east.
>The most of the customers are those who don’t have childrens.
>The positive thing is that most of them are non-smokers.
> The customers those who smoke pay the highest charges that is between 24000 - 40000.
>The charges to a non-smoker customer is between 4000 - 18000 .
>There is a positive relationship between the age and charges.
>The percentage of male is slightly greater than a female in smoking.
> The most region that have highest smokers is the south-east.
>There is a negative relationship between the children and charges.

Suggestions:

> The company should focus on non-smokers and should avoid giving insurance to smokers because the risks are high .

> The company should also focus on Obese customer and provide excercise recommendation .

Conclusion:

This project's findings can help the company to analyse the customer's Health behaviours, act accordingly and improve their services and policies.

References:

Data source: https://www.kaggle.com/teertha/ushealthinsurancedataset

Visualizations: R, Google sheets

PPT: Google slides

THANK YOU