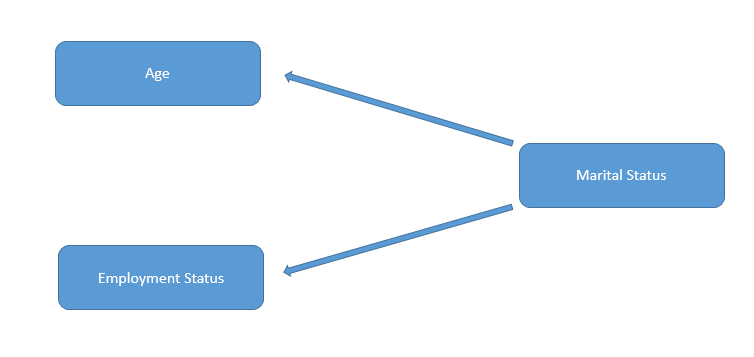
# Regression Model

**Explanation:**

GLM: Used to predict relationships among various variables. We use regression analysis for this process. It is a modelling technique to analyses several variables. It focusses on the relationship between an independent variable and several dependent variables. We can also look at the strength of the various predictors. In the below diagram we can see that how the variables like Age and Employment status affect Marital status.



For this purpose, I used conditional plots with various variables like below,

Age vs Marital status

Age vs Employment Status

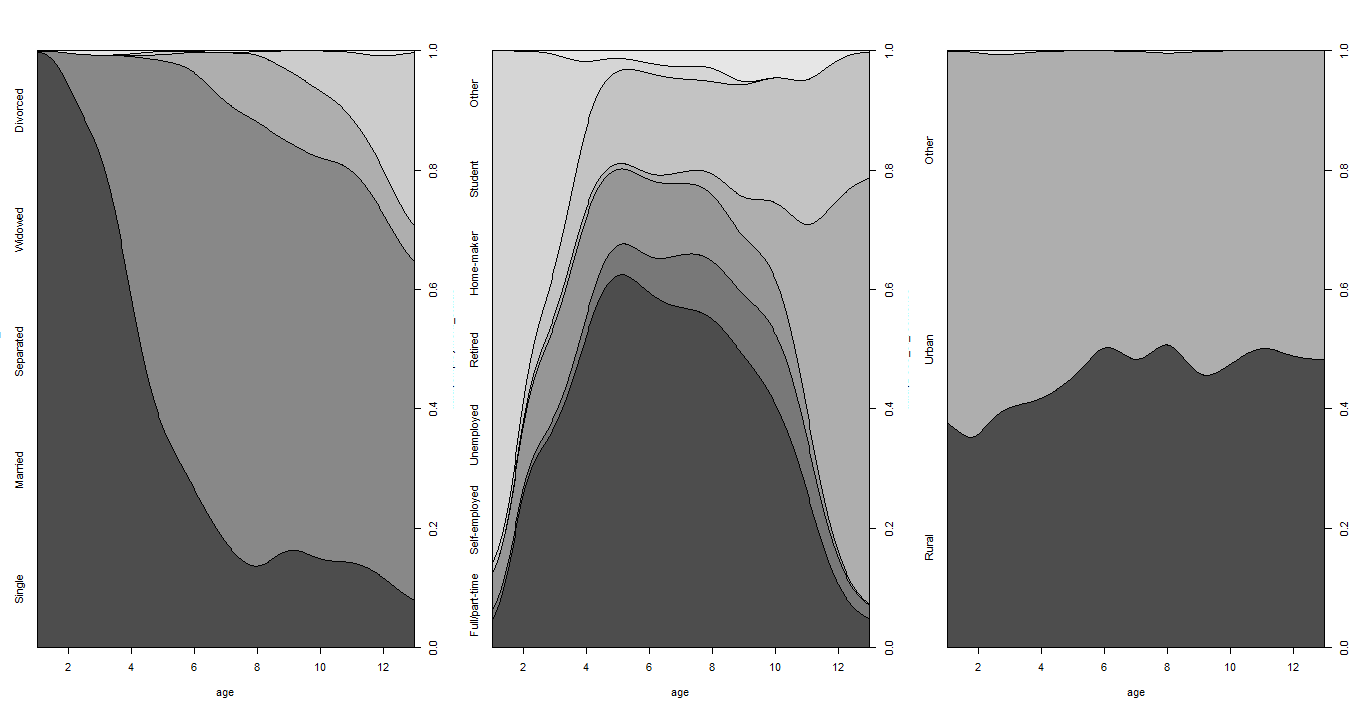
Age vs Residence

Marital status vs (Age + Employment Status)

Marital status vs (Age + Employment Status+ Residence)

This will give the relationship between dependent and independent variables. This gives

Conditional plots like the below figure.



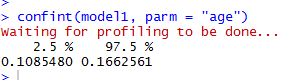
From this we can deduce that there is a smooth relationship between Age and Marital status and between Age and Employment status.

We then run the analysis of model 1:

We notice that Residual deviance is 2665.4 on 2023 degrees of freedom.

This shows that the data is not rigorous and points to invalidity of data.

After confidence interval testing ,we get



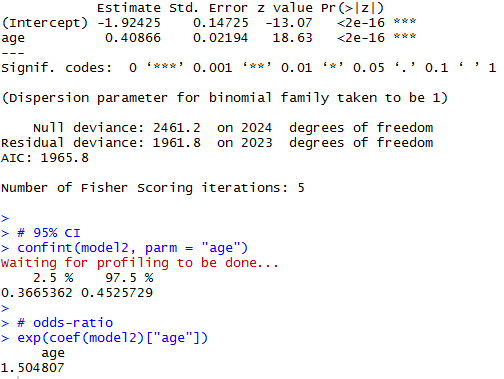
Then we check the odds ratio



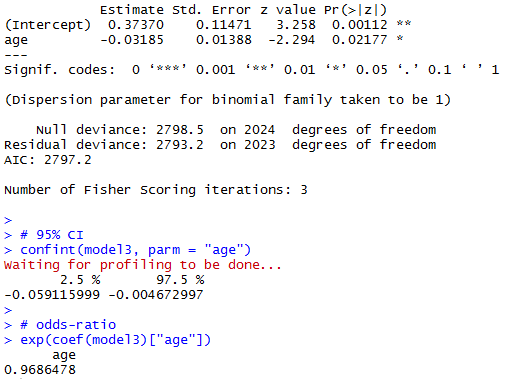
This depicts that for every 1 % increase in Age will cause 1.14 times change in Employment status.

We did this for all the 5 models,

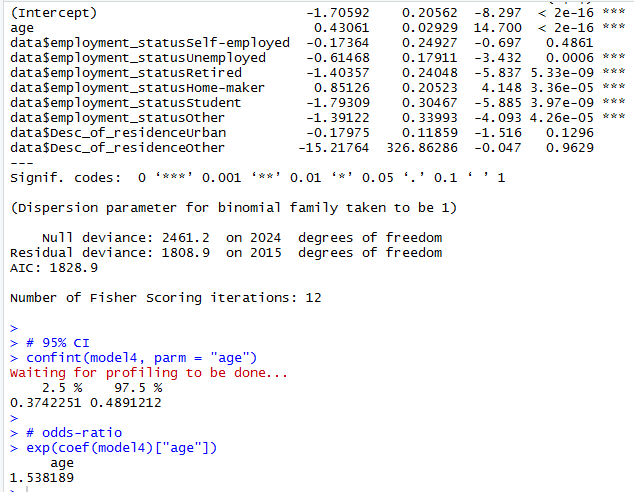
Model 2:



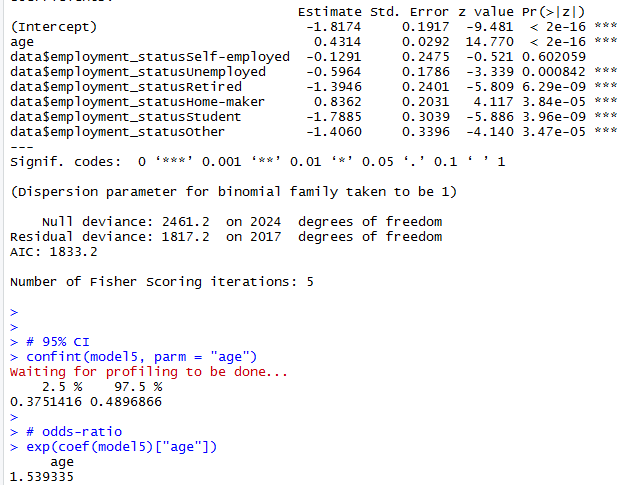
Model 3:



For model 4:



For Model 5:



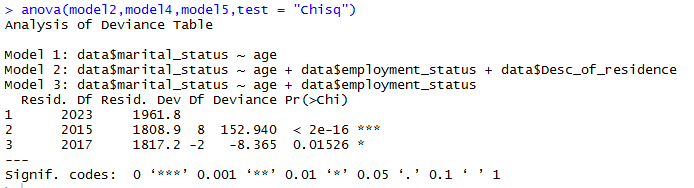
**Now which model is better?**

I reject the bad models by looking at their respective residual deviance.

This leaves us with 3 valid models namely: Model 1, model 4 and model 5.

I then run the ANOVA test (Chi-squared) on these 3 models for comparison.

The results:



This shows that Model 4 is the best as the significance level is 2e -16.

Graphical representation of Model 4 for further verification. If the regression planne fits perfectly across the 3 dimensions, that means the model is good.

**3-D scatterplot with regression plane:**

# 3D Scatterplot with Coloring and Vertical Lines

# and Regression Plane

library**(**scatterplot3d**)**

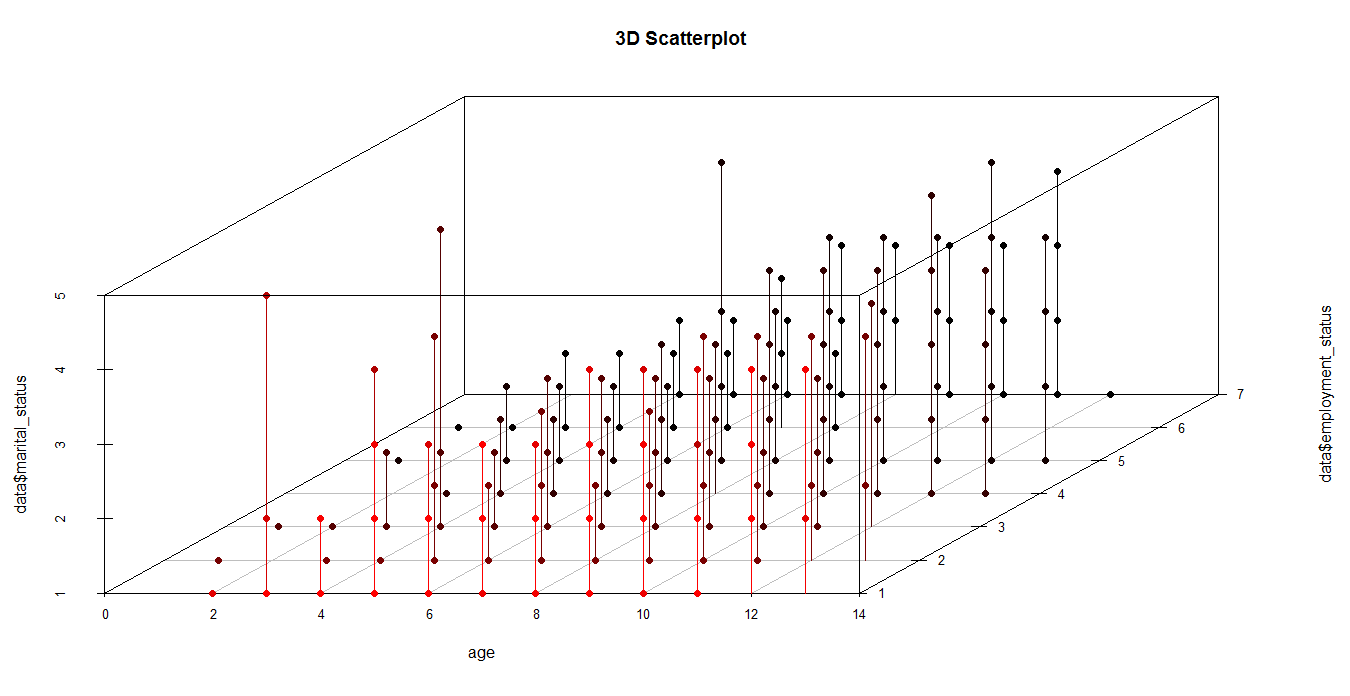
attach**(**data**)**

s3d **<-**scatterplot3d**(**age,data**$**employment\_status,data**$**marital\_status, pch**=**16, highlight.3d**=TRUE**,

type**=**"h", main**=**"3D Scatterplot"**)**

fit **<-** lm**(**data**$**marital\_status **~** age**+**data**$**employment\_status**)**

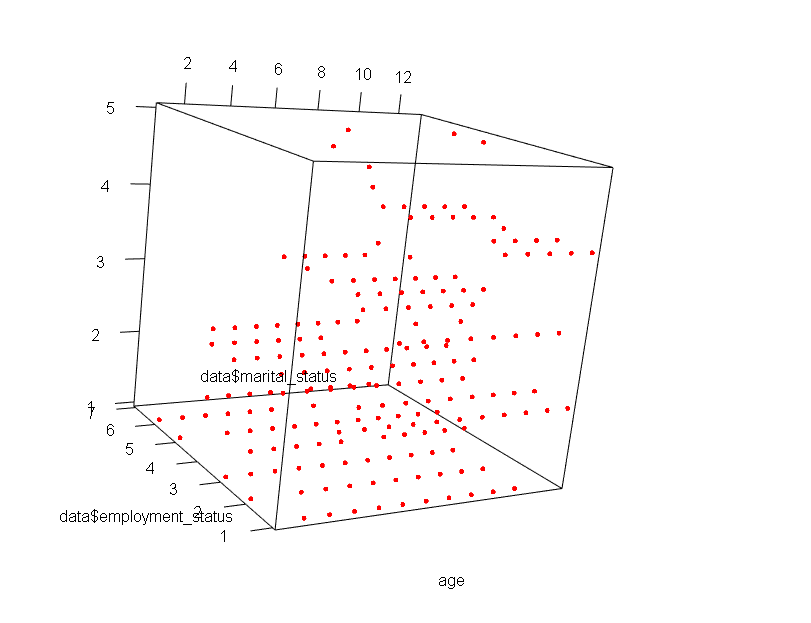
s3d**$**plane3d**(**fit**)**

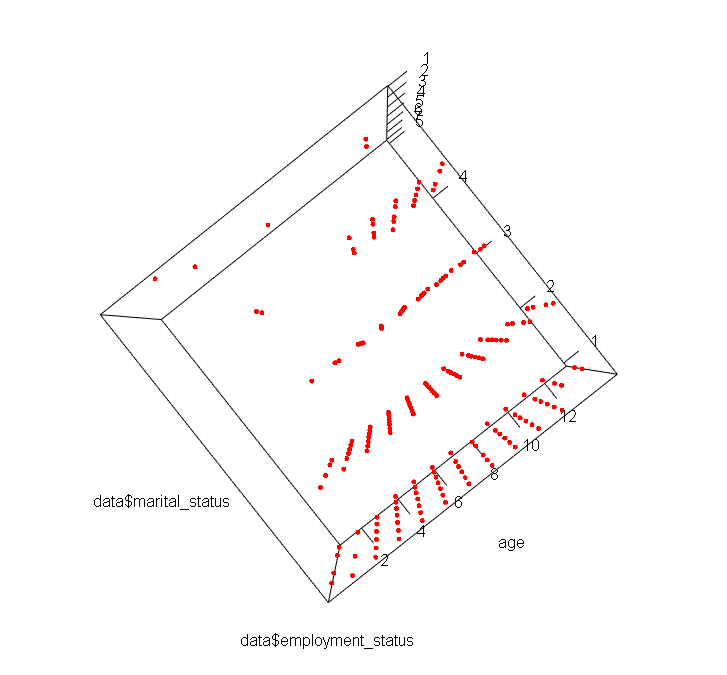


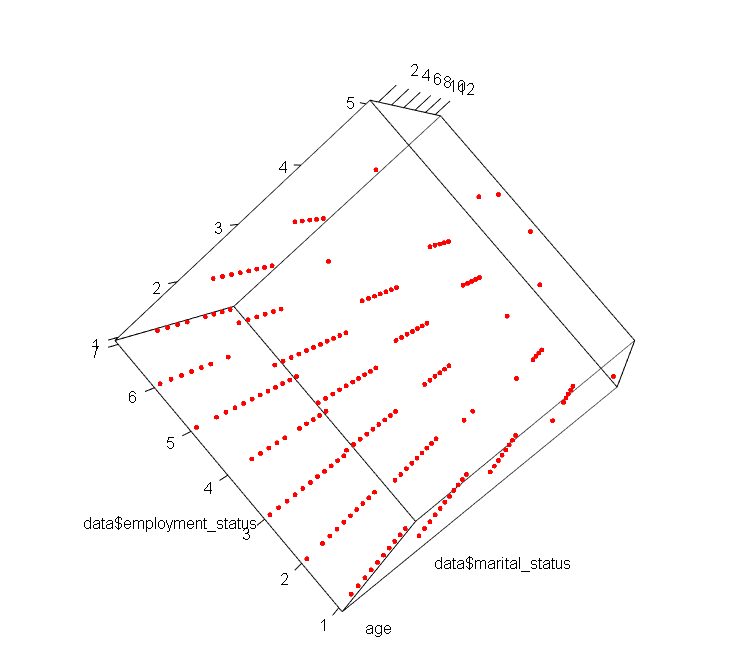
**Click the icon below for a detailed view:**



**Spinning 3D plots:**







**Conclusion:**

Thus we can see that Age and Employment status are good predictors of Marital Status.