**Q1. Simple R program for multiplication using while loop:**

c= 0

x=6

y=10

while(x>0)

{c = c + y

x = x - 1}

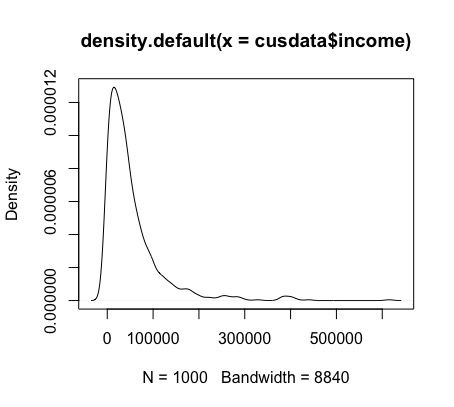
print (c)

**Q2. ​Like histogram, you can also plot density of a variable. Figure out how to plot density of income. Provide a couple of sentence description along with the plot.**

Code: > ​**options(scipen=10)**

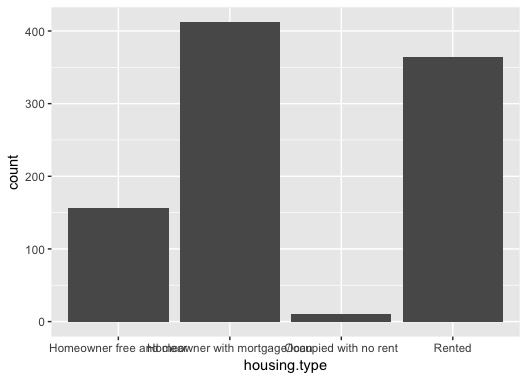
**> plot(density(cusdata$income))**

Description:  
It is evident from the density plot that the density or the number of people having income in the range of 0 and $100000 is the highest and the density is constantly low for higher incomes. Density graph is an effective way of representation of data as it is easier to identify the distribution of points and it is not affected by outliers.



**Q3. Create a bar chart for housing type. Make sure to remove the "NA" type. (Hint: you can use subset function with an appropriate condition on housing type field.) Provide your commands and the plot.  
Command:**

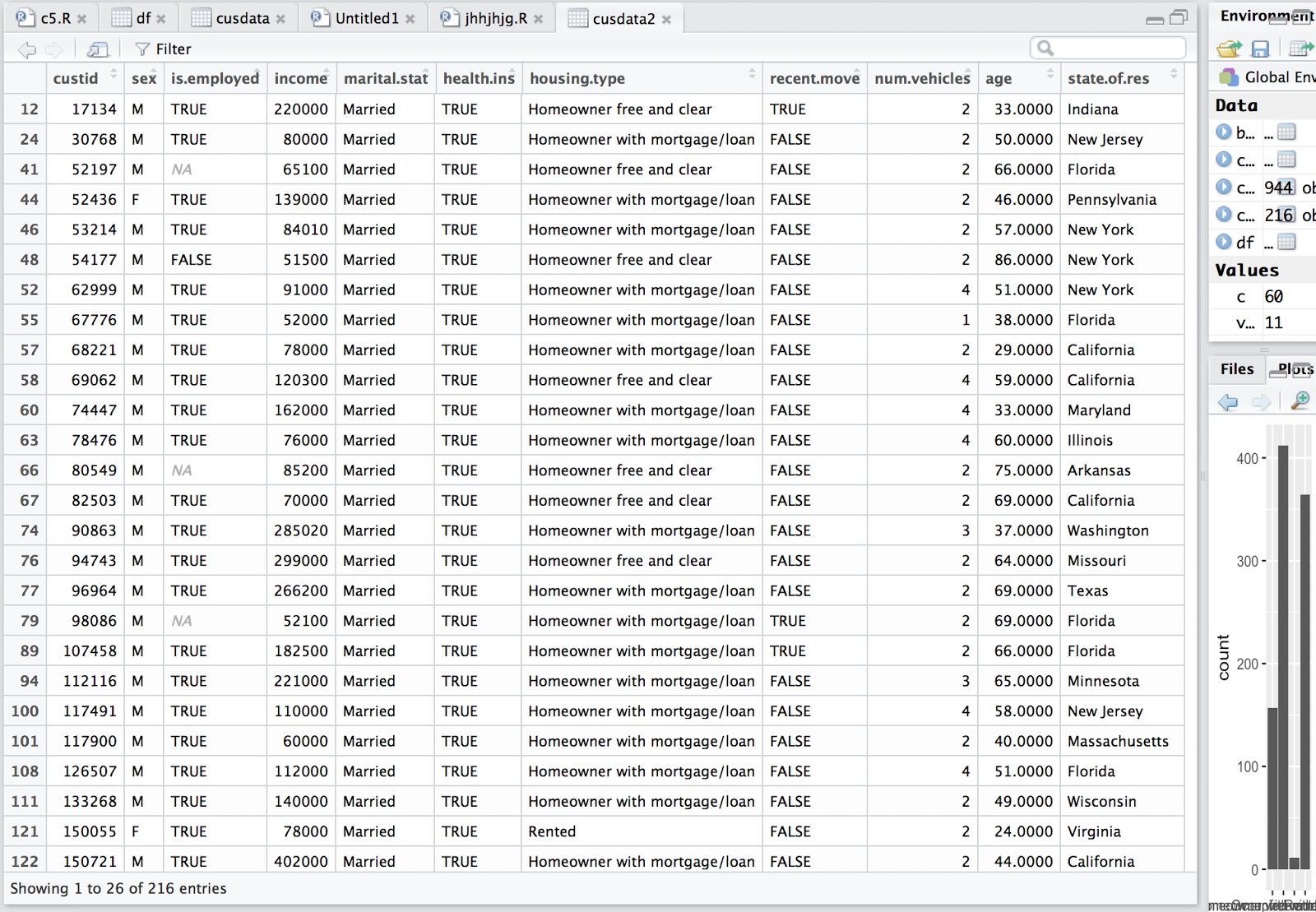
cusdata1=subset(cusdata,(cusdata$housing.type!='NA')) > ggplot(cusdata1) + geom\_bar(aes(x=housing.type))



**Q4. Extract a subset of customers that are married and have income more than $50,000. What % of these customers have health insurance? How does this % differ from that for the whole dataset?**

Subset of customers who are married and have an income more that $50000:

cusdata2= subset(cusdata,(cusdata$marital.stat=='Married' & (cusdata$income>50000))) View(cusdata2)



**What % of these customers have health insurance?**

Code: cusdata3=subset(cusdata2,(cusdata2$health.ins=='TRUE')) PercMarried= nrow(cusdata3)/nrow(cusdata2)\*100

> print(PercMarried) [1]​ **96.2963**

**How does this % differ from that for the whole dataset?**  
Code:  
We can find this by finding the percentage of all the individuals having health insurance:

cusdata4=subset(cusdata,(cusdata$health.ins=='TRUE')) PercOVER= nrow(cusdata4)/nrow(cusdata)\*100

print(PercOVER) [1]​ **84.1%**

**Q5. Do you think there is any correlationships among age, income, and number of vehicles? Report your correlation numbers and interpretations.**

Clearing the data:  
cusCor= subset(cusdata, (cusdata$age!='NA')&(cusdata$income!='NA')&(cusdata$num.vehicle!=’NA’)&(cusdata$age>17 & cusdata$age<100)&(cusdata$income>0))  
> View(cusCor)

Correlation between:  
Age and Income:  
cat('Correlation btn Age and Income', cor(cusCor$income, cusCor$age)) Correlation btn Age and Income -0.02240845

Age and No. of Vehicles:

cat('Correlation btn age and no. of vehicles', cor(cusCor1$num.vehicles, cusCor1$age)) Correlation btn age and no. of vehicles -0.07218462

Income and No. of vehicles:  
cat('Correlation btn income and no. of vehicles', cor(cusCor1$num.vehicles, cusCor1$income)) Correlation btn income and no. of vehicles 0.1477657

The variables age and no. of vehicles show a high negative correlation of -0.0722 which implies that the no. of young people having more cars is expected to be more than the number of older people.