#1. loading required library and data

library(dplyr)

library(ggplot2)

house\_1=read.csv("C:/Users/I340968/Desktop/ed-code/R programming-elective/Week\_5/housingdata\_v2.0.csv")

View(house\_1)

Console:

> #1. loading required library and data

> library(dplyr)

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

> library(ggplot2)

> house\_1=read.csv("C:/Users/I340968/Desktop/ed-code/R programming-elective/Week\_5/housingdata\_v2.0.csv")

> View(house\_1)

#2. Understanding data structure

nrow(house\_1)

summary(house\_1)

names(house\_1)

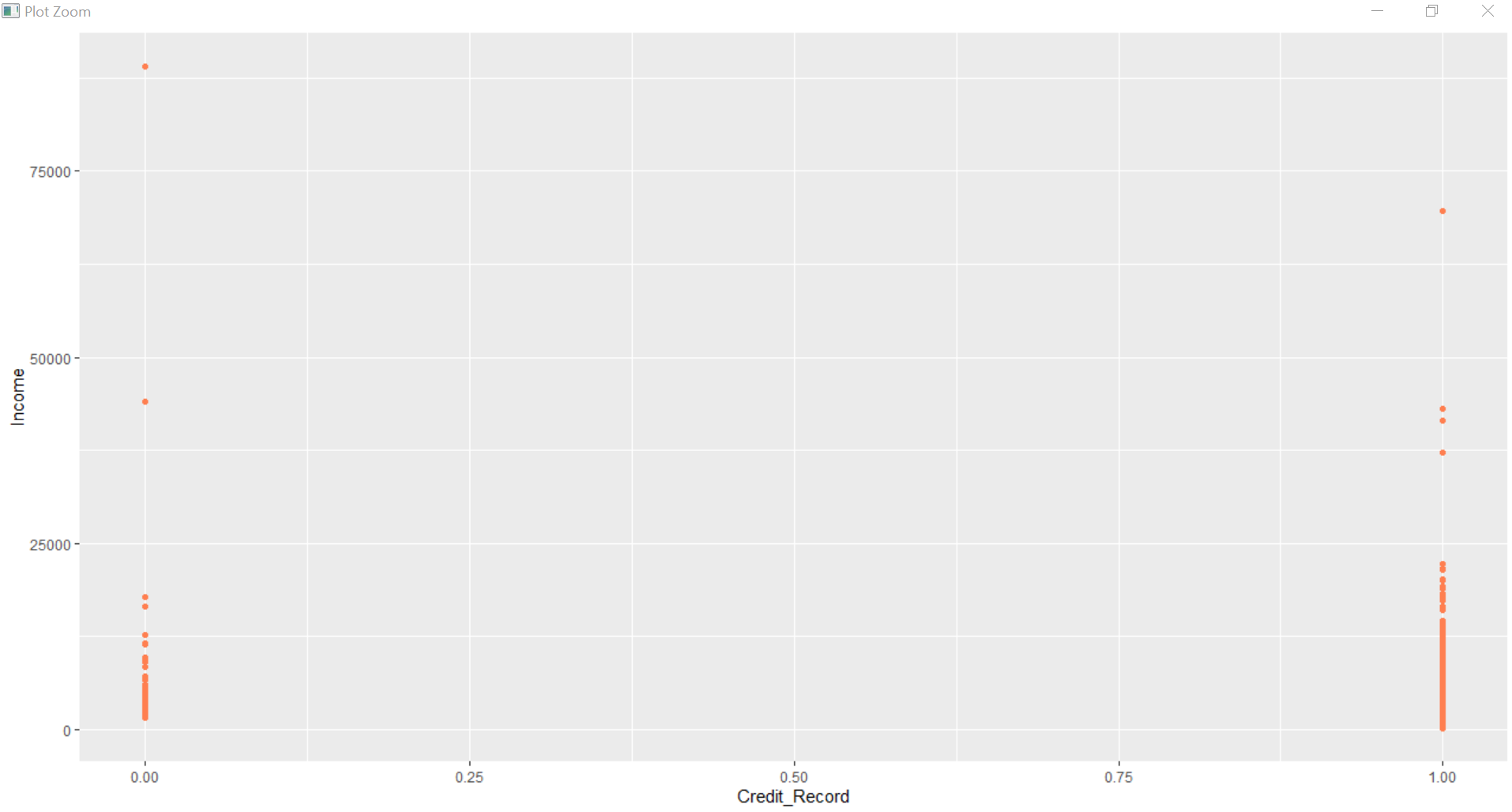
str(house\_1)

Console:

|  |
| --- |
| > #2. Understanding data structure  > nrow(house\_1)  [1] 505  > summary(house\_1)  Record Gender No\_kids Education  Record1 : 1 Female: 93 Min. :0.0000 Graduate :402  Record10 : 1 Male :412 1st Qu.:0.0000 Not Graduate:103  Record100: 1 Median :0.0000  Record101: 1 Mean :0.7723  Record102: 1 3rd Qu.:2.0000  Record103: 1 Max. :3.0000  (Other) :499  HasCar Income PropertyValue Loan\_Period  No :414 Min. : 165 Min. : 7200 Min. : 12.0  Not Answered: 25 1st Qu.: 3185 1st Qu.: 80000 1st Qu.:336.0  Yes : 66 Median : 4217 Median :102400 Median :336.0  Mean : 5953 Mean :115653 Mean :317.9  3rd Qu.: 6400 3rd Qu.:133600 3rd Qu.:336.0  Max. :89100 Max. :480000 Max. :456.0    Credit\_Record Housing\_type Property\_Purchased  Min. :0.0000 Affordable:149 N:177  1st Qu.:1.0000 Mid Range :200 Y:328  Median :1.0000 Premium :156  Mean :0.8594  3rd Qu.:1.0000  Max. :1.0000    > names(house\_1)  [1] "Record" "Gender" "No\_kids"  [4] "Education" "HasCar" "Income"  [7] "PropertyValue" "Loan\_Period" "Credit\_Record"  [10] "Housing\_type" "Property\_Purchased"  > str(house\_1)  'data.frame': 505 obs. of 11 variables:  $ Record : Factor w/ 505 levels "Record1","Record10",..: 1 484 495 2 13 24 35 57 68 90 ...  $ Gender : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 2 2 2 2 2 ...  $ No\_kids : int 0 0 0 0 0 0 0 0 0 0 ...  $ Education : Factor w/ 2 levels "Graduate","Not Graduate": 1 1 1 2 1 1 1 1 1 1 ...  $ HasCar : Factor w/ 3 levels "No","Not Answered",..: 1 1 3 1 1 1 1 3 1 1 ...  $ Income : int 710 6516 7040 4730 9167 10459 2888 10960 8692 4044 ...  $ PropertyValue : int 90400 168800 160000 155200 149600 149600 149600 144000 144000 137600 ...  $ Loan\_Period : int 456 336 336 336 336 336 336 336 336 336 ...  $ Credit\_Record : int 1 1 1 1 1 1 1 1 1 1 ...  $ Housing\_type : Factor w/ 3 levels "Affordable","Mid Range",..: 1 1 1 1 1 1 1 1 1 1 ...  $ Property\_Purchased: Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ... |
|  |
| |  | | --- | | > | |

#3. Scatter plot creadite record on x-axis and income on y-axis

ggplot(house\_1,aes(y=Income,x=Credit\_Record))+geom\_point(col="coral")

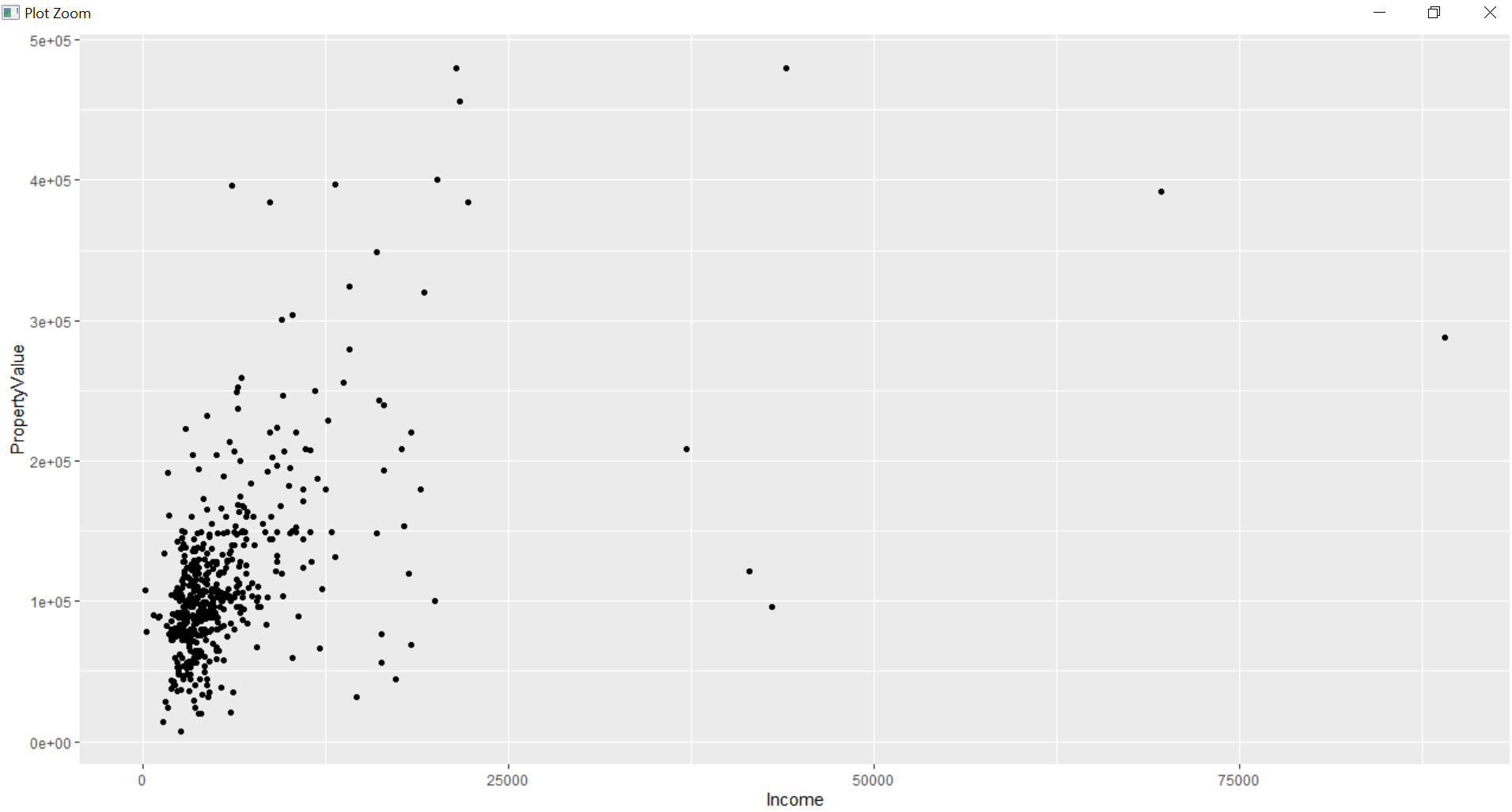


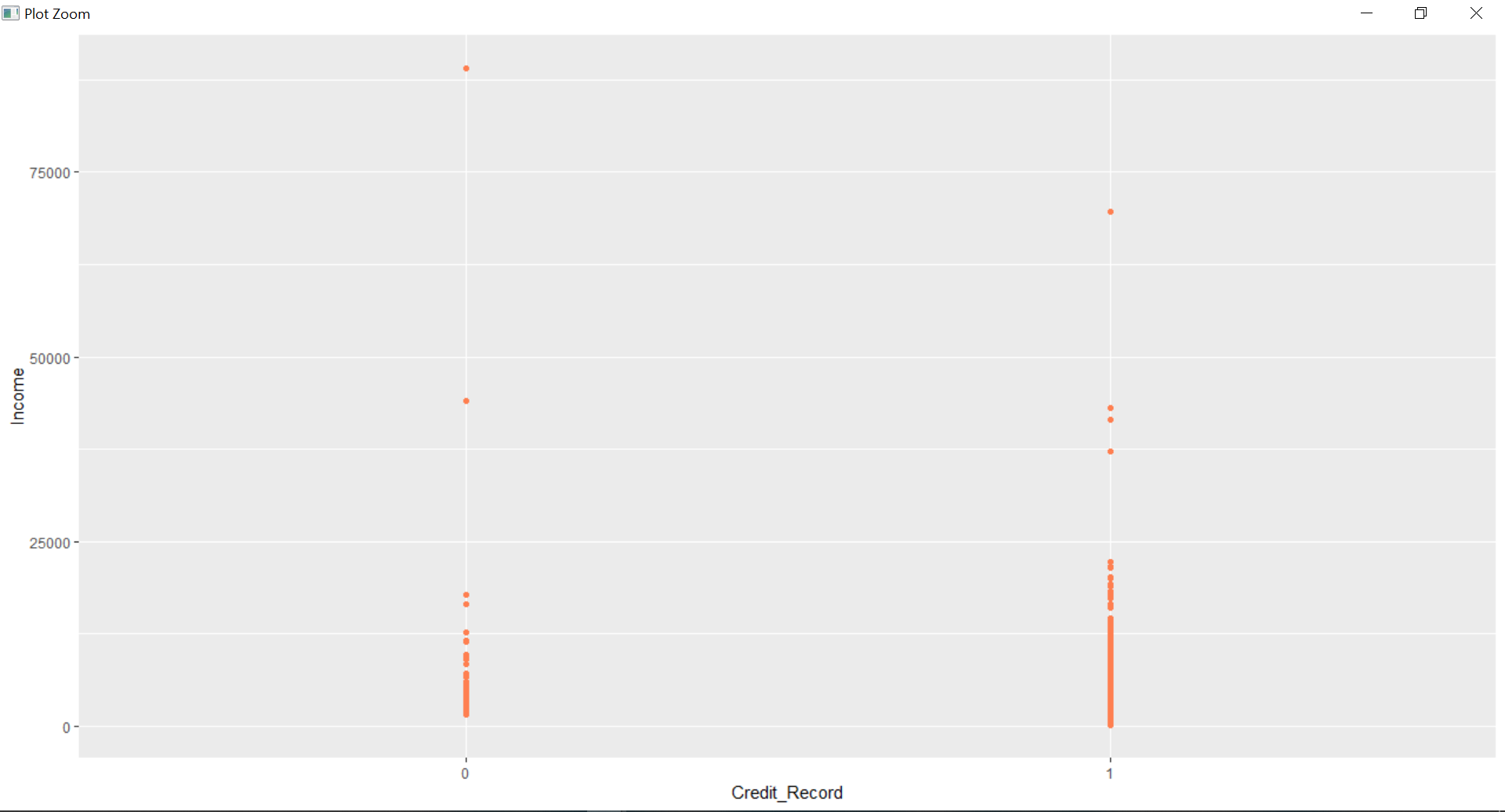
Converting credit record as factor:

==============================

house\_1$Credit\_Record=as.factor(house\_1$Credit\_Record)

ggplot(house\_1,aes(y=Income,x=Credit\_Record))+geom\_point(col="coral")





#the difference is before converting as factor the x-value has been taken

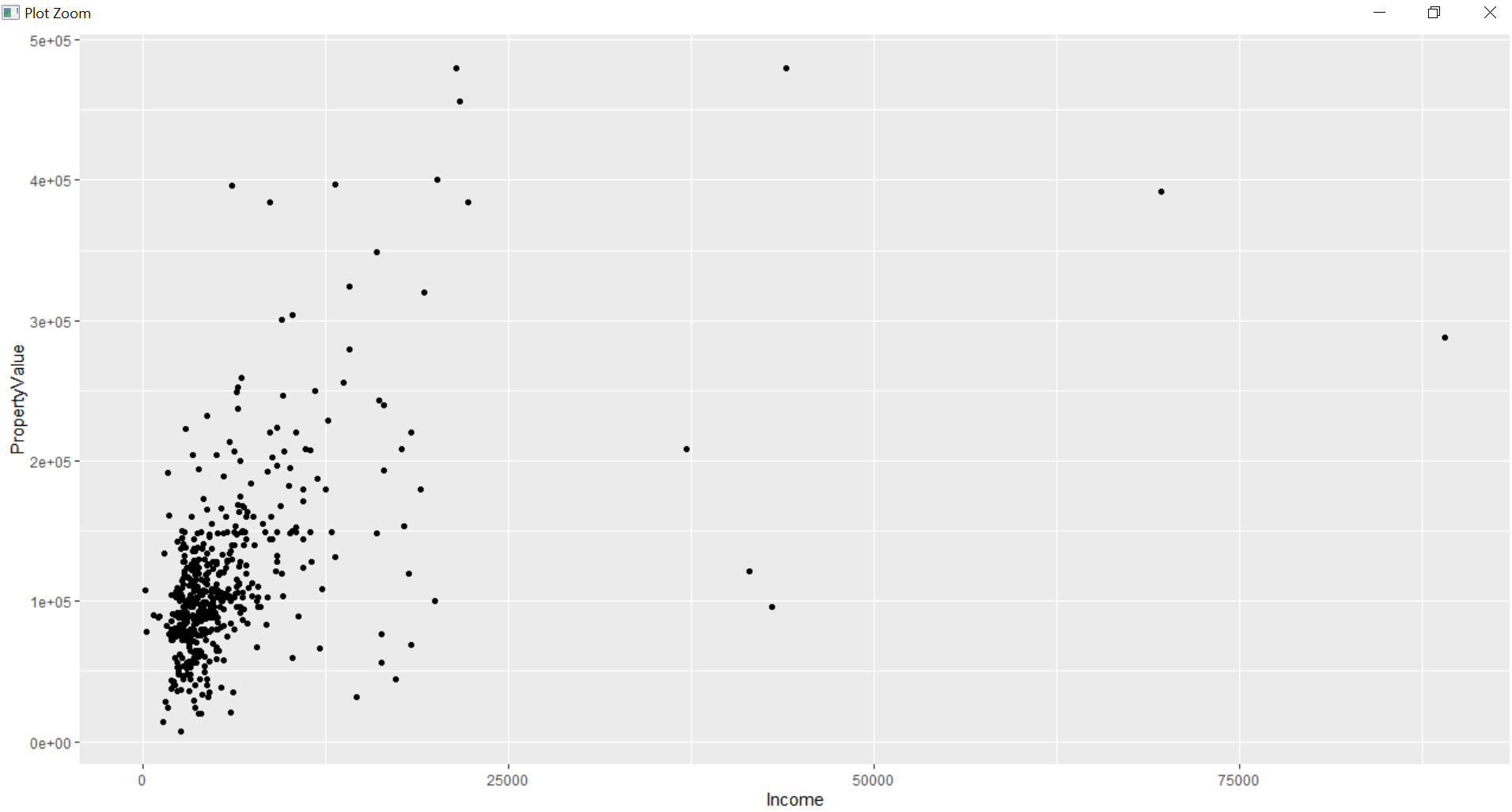
#in a range between 0 to 1 but after convertion x-axis value are 0 and 1

#4.scatter plot between in-come on x-axis and property on y-axis

ggplot(house\_1,aes(x=Income,y=PropertyValue))+geom\_point()

Console:

ggplot(house\_1,aes(x=Income,y=PropertyValue))+geom\_point()

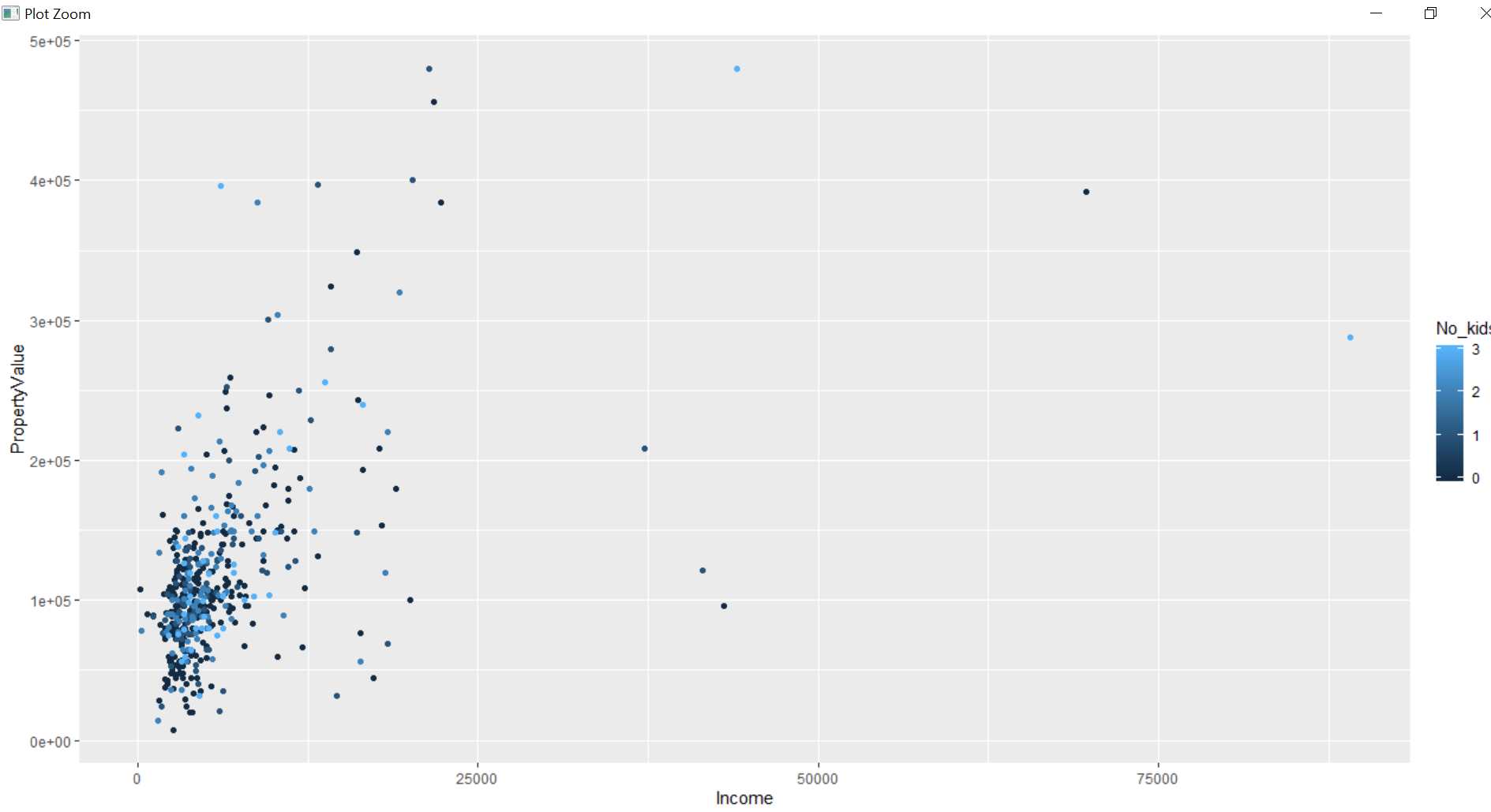


#4.1 Col argument depending on no\_kids

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point()

Console:

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point()

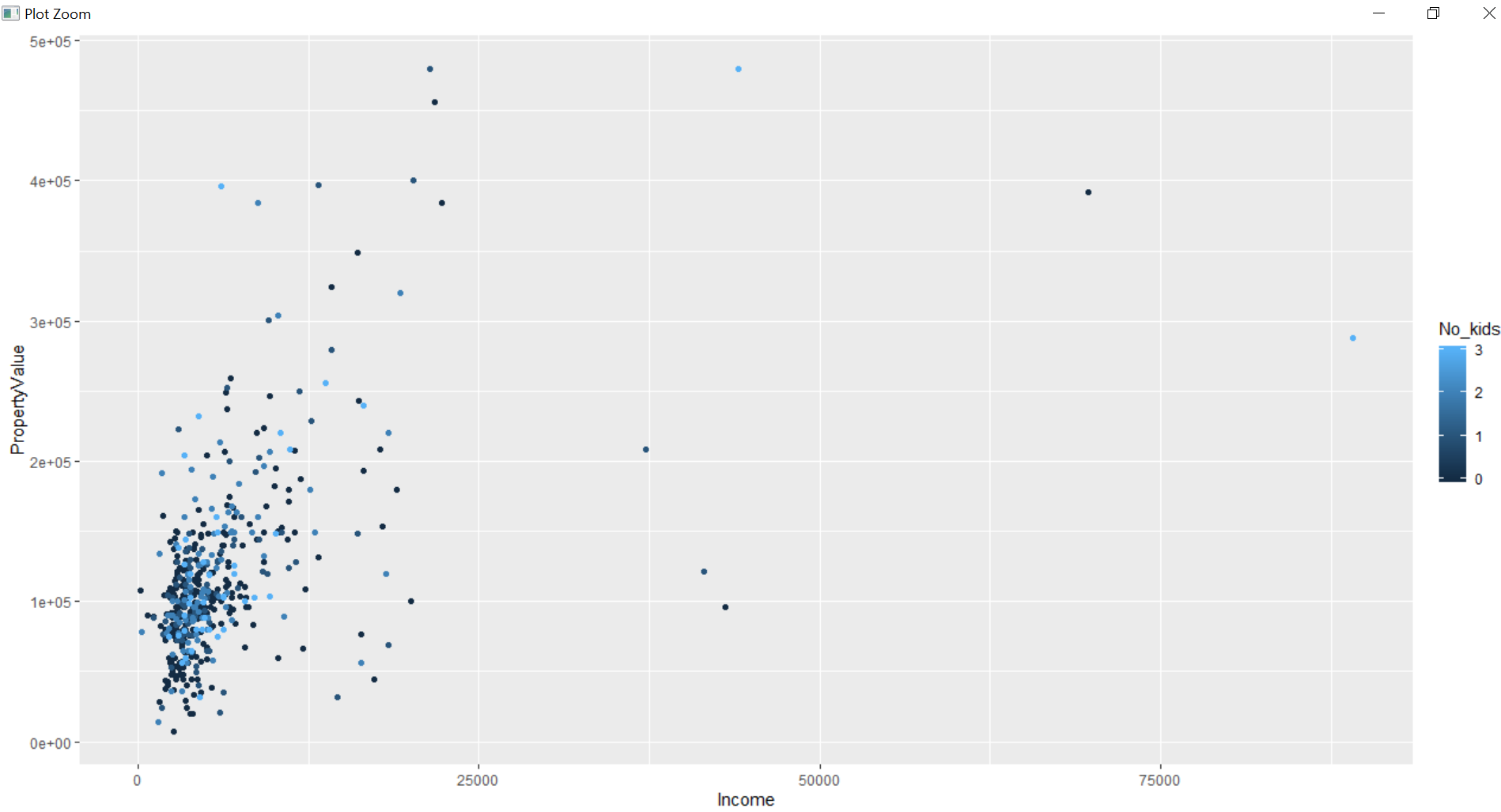


#4.2 adding size argument depends on no\_kids

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids),size=No\_kids)+geom\_point()

console:

|  |
| --- |
| > ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids),size=No\_kids)+geom\_point() |
|  |
| |  | | --- | | > | |

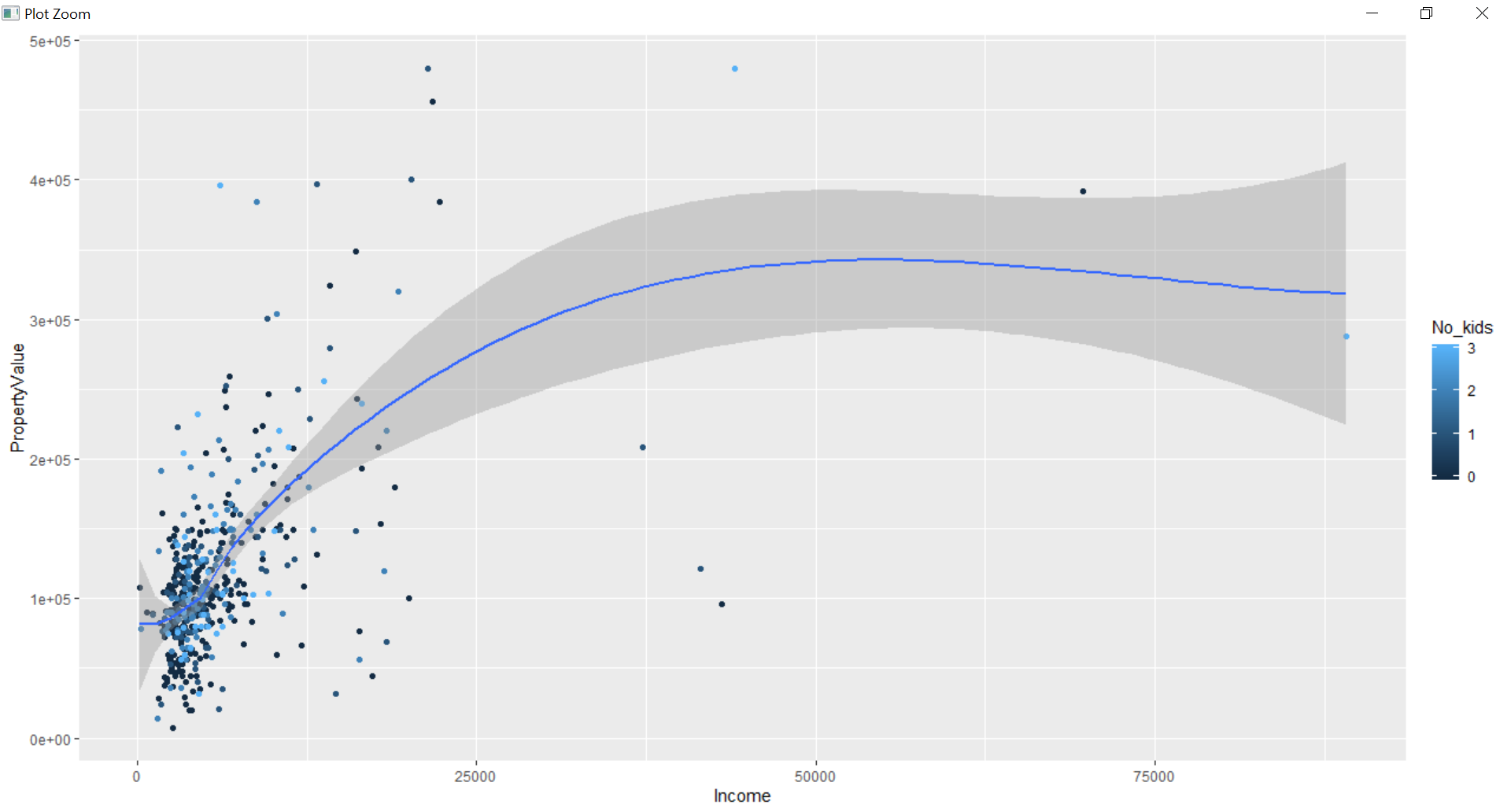


#4.3 smooth line adding

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids),size=No\_kids)+geom\_point()+geom\_smooth()

Console:

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids),size=No\_kids)+geom\_point()+geom\_smooth()



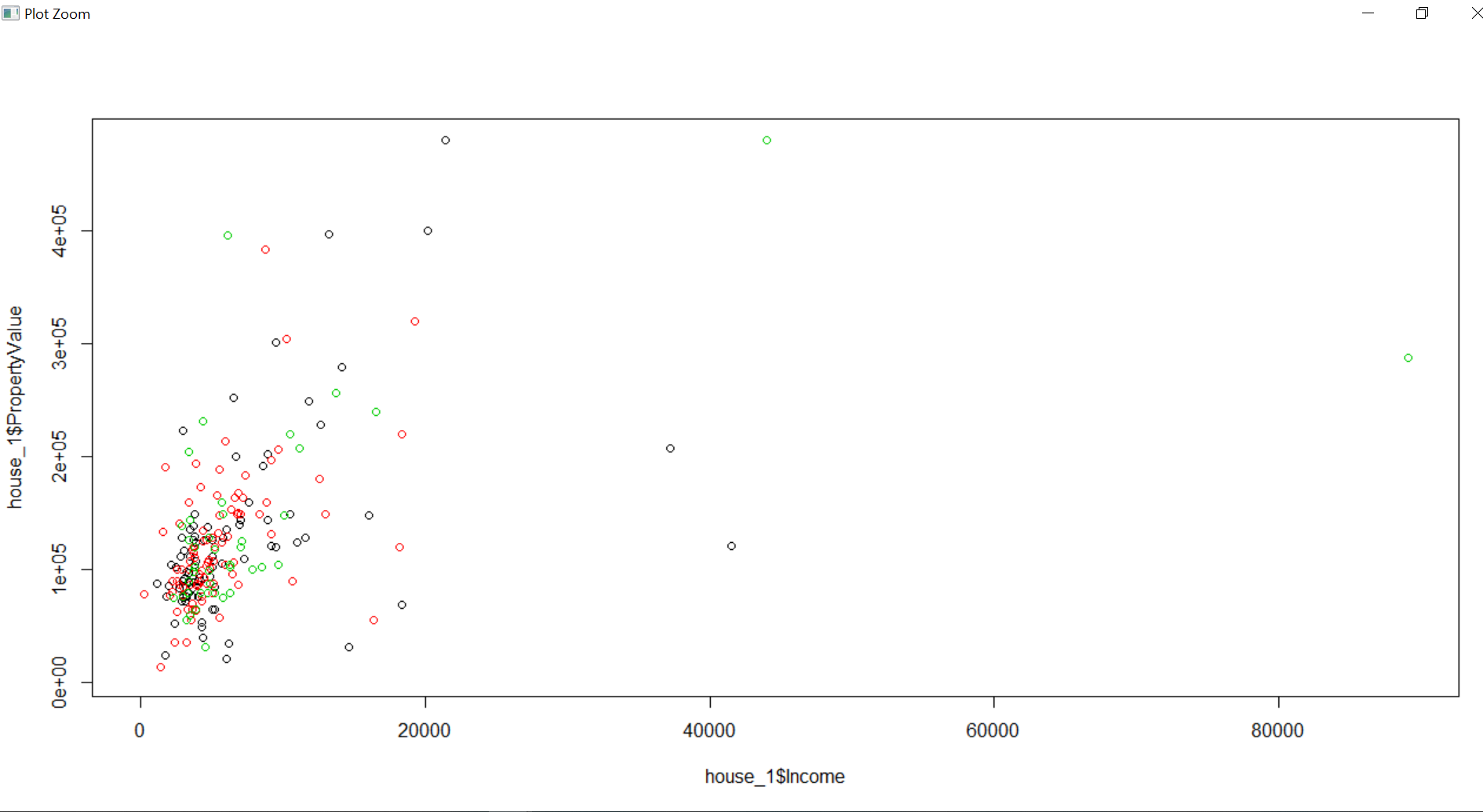
#5.ggplot comparson with base plot

#5.1 scatter plot- color according to no\_kids

plot(house\_1$PropertyValue~house\_1$Income,col=house\_1$No\_kids)

Console:

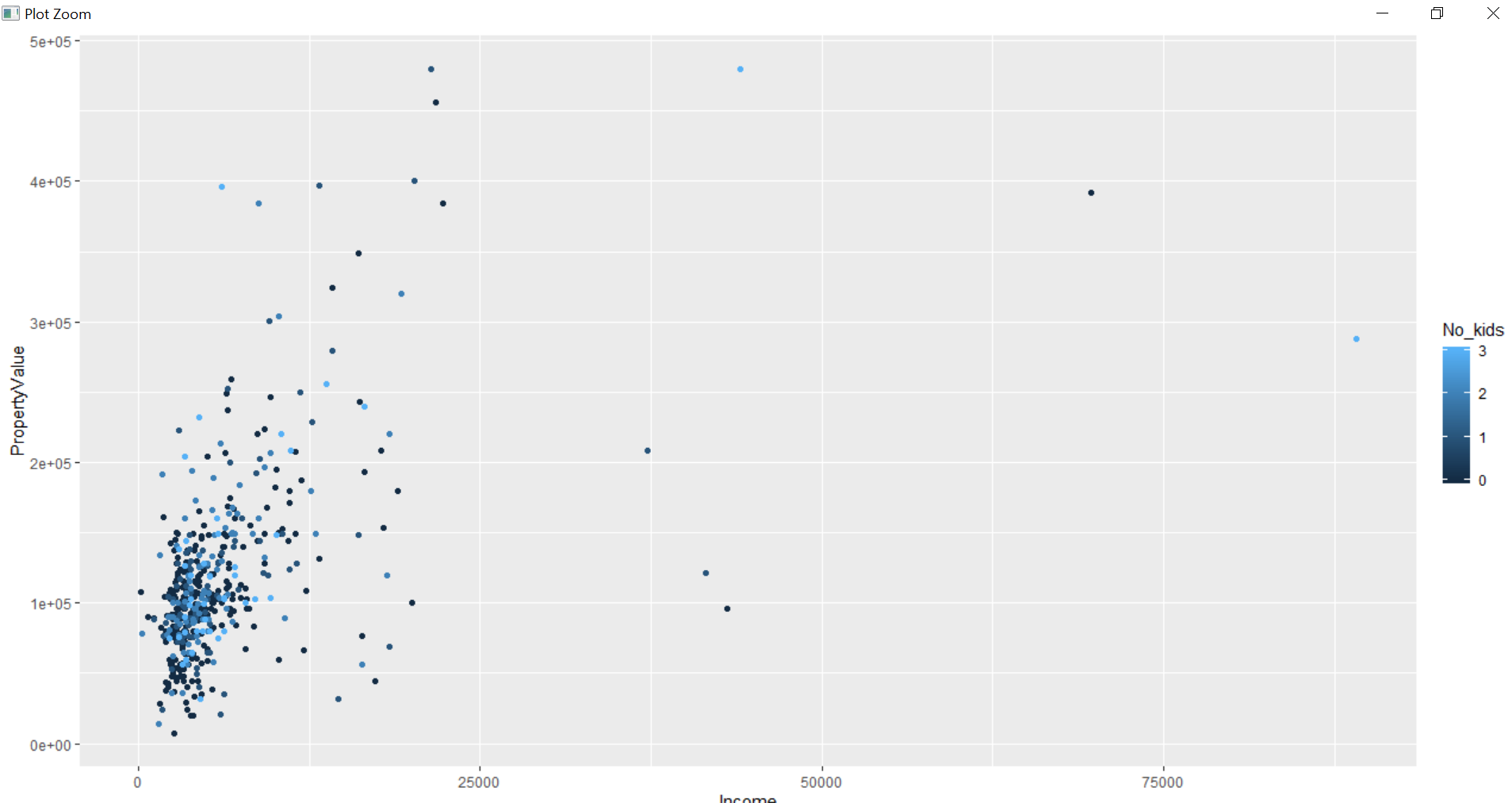
plot(house\_1$PropertyValue~house\_1$Income,col=house\_1$No\_kids)



ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point()

Console:

|  |
| --- |
| ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point() |
|  |
| |  | | --- | | > | |



#5.2 changing no of kids to factor

house\_1$No\_kids=as.factor(house\_1$No\_kids)

console:

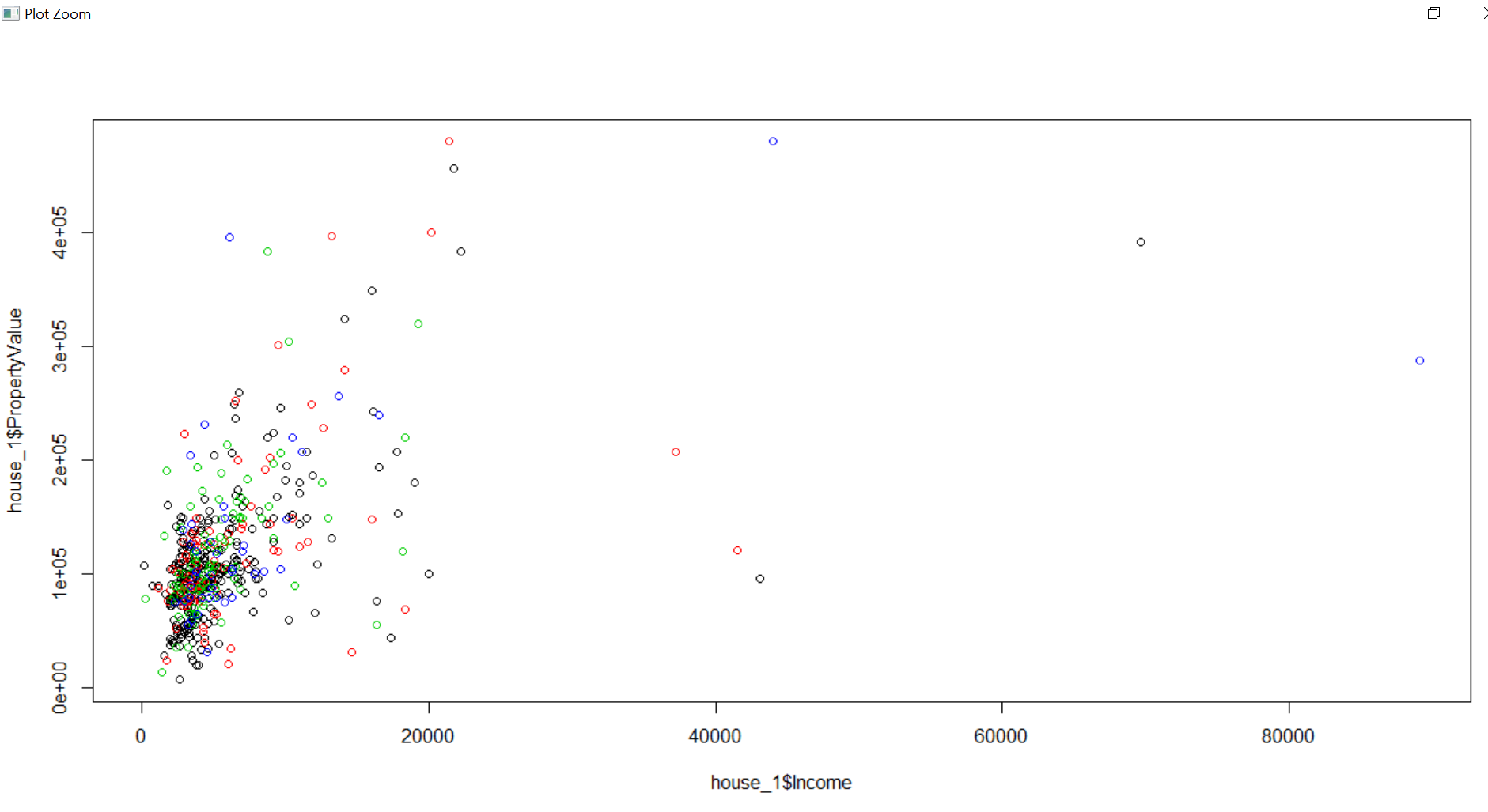
> house\_1$No\_kids=as.factor(house\_1$No\_kids)

#5.3 making the plot same as 5.1

plot(house\_1$PropertyValue~house\_1$Income,col=house\_1$No\_kids)

console:

plot(house\_1$PropertyValue~house\_1$Income,col=house\_1$No\_kids)

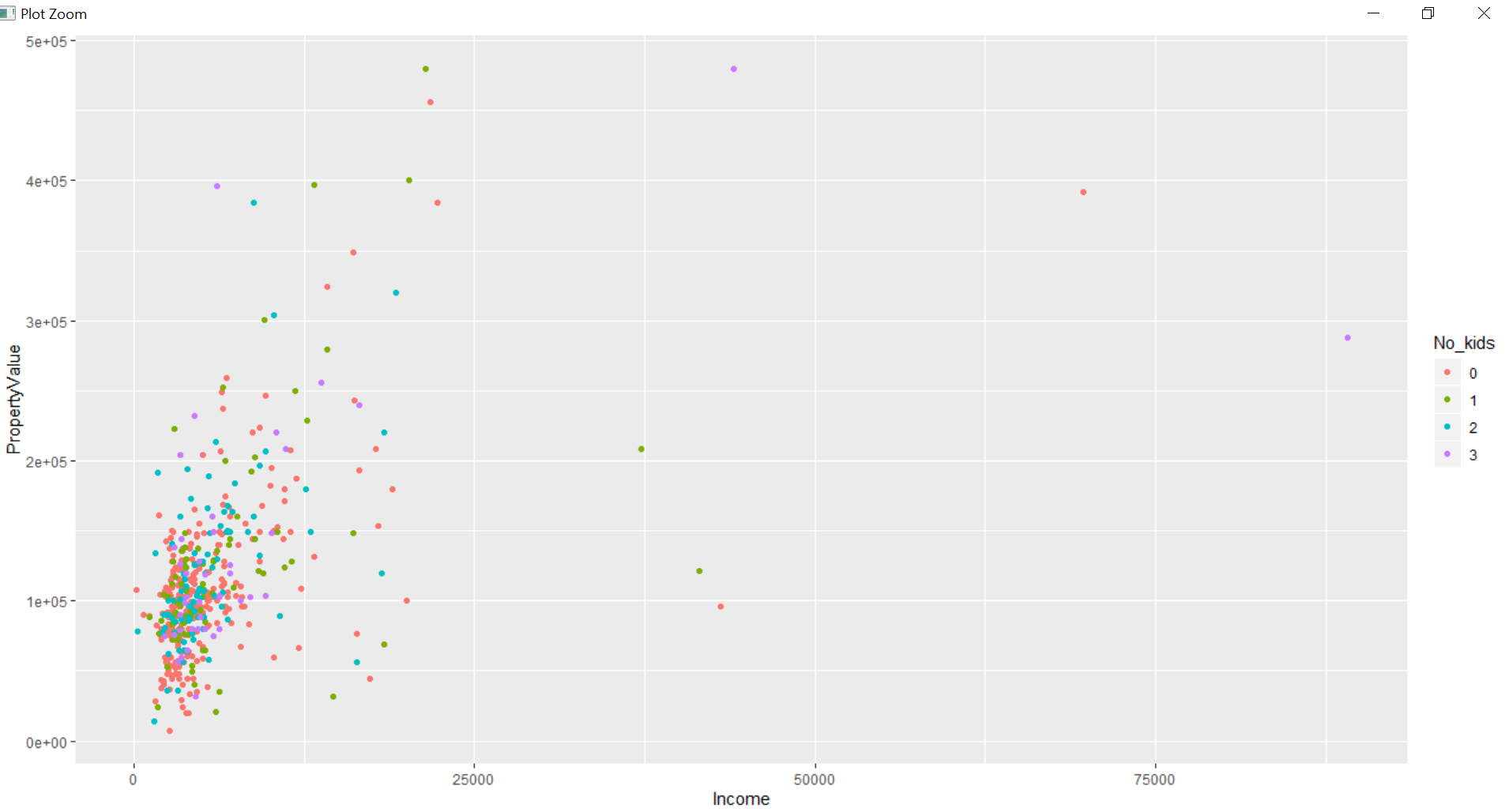


#5.4 Recreating same plot using ggplot.

ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point()

console:

|  |
| --- |
| #5.4 Recreating same plot using ggplot.  >  > ggplot(house\_1,aes(x=Income,y=PropertyValue,col=No\_kids))+geom\_point() |
|  |
| |  | | --- | | > | |



#6. Asethethics

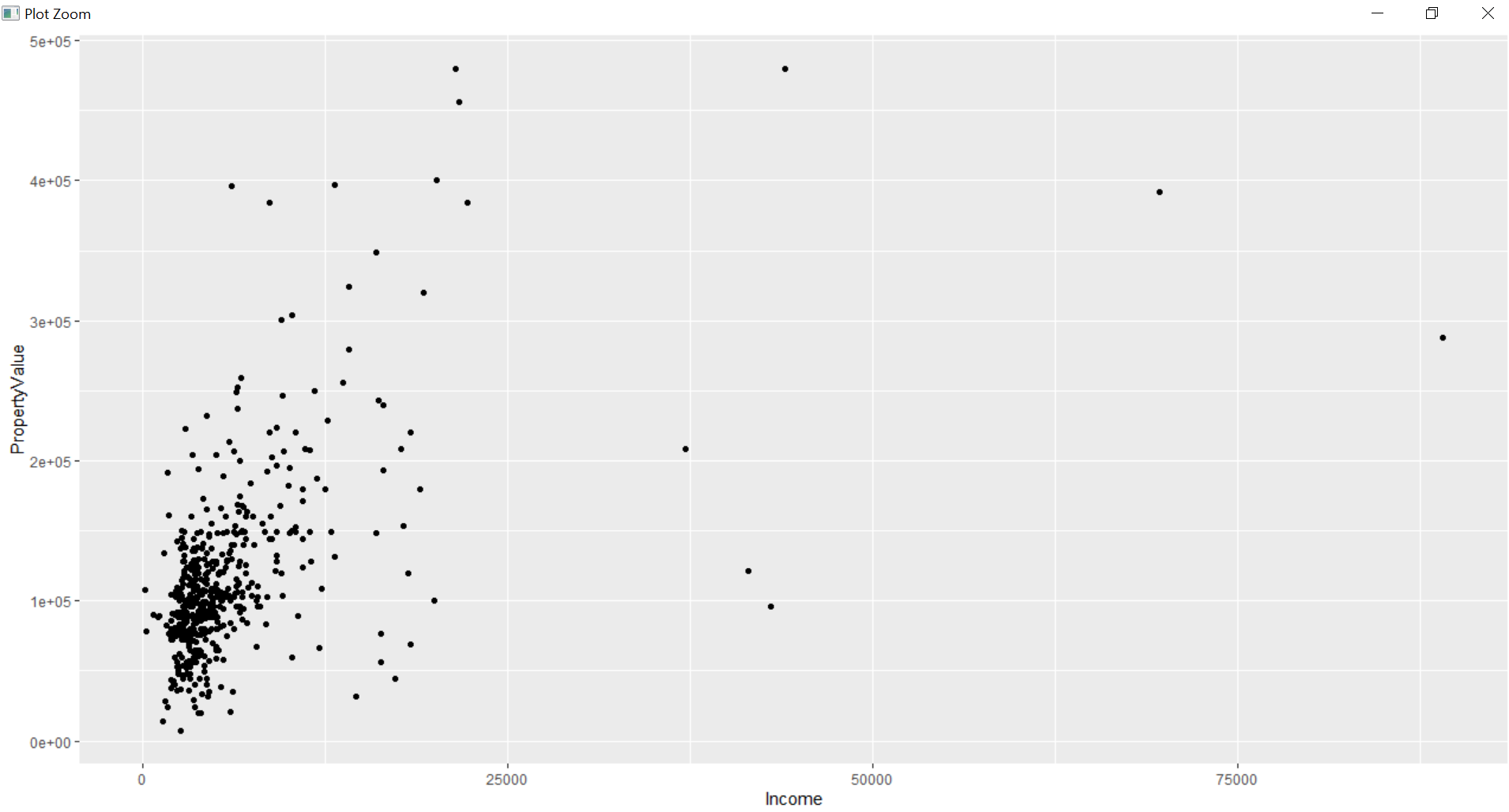
#6.1 Map incomet to x and property value to y

ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_point()

console:

> #6.1 Map incomet to x and property value to y

> ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_point()

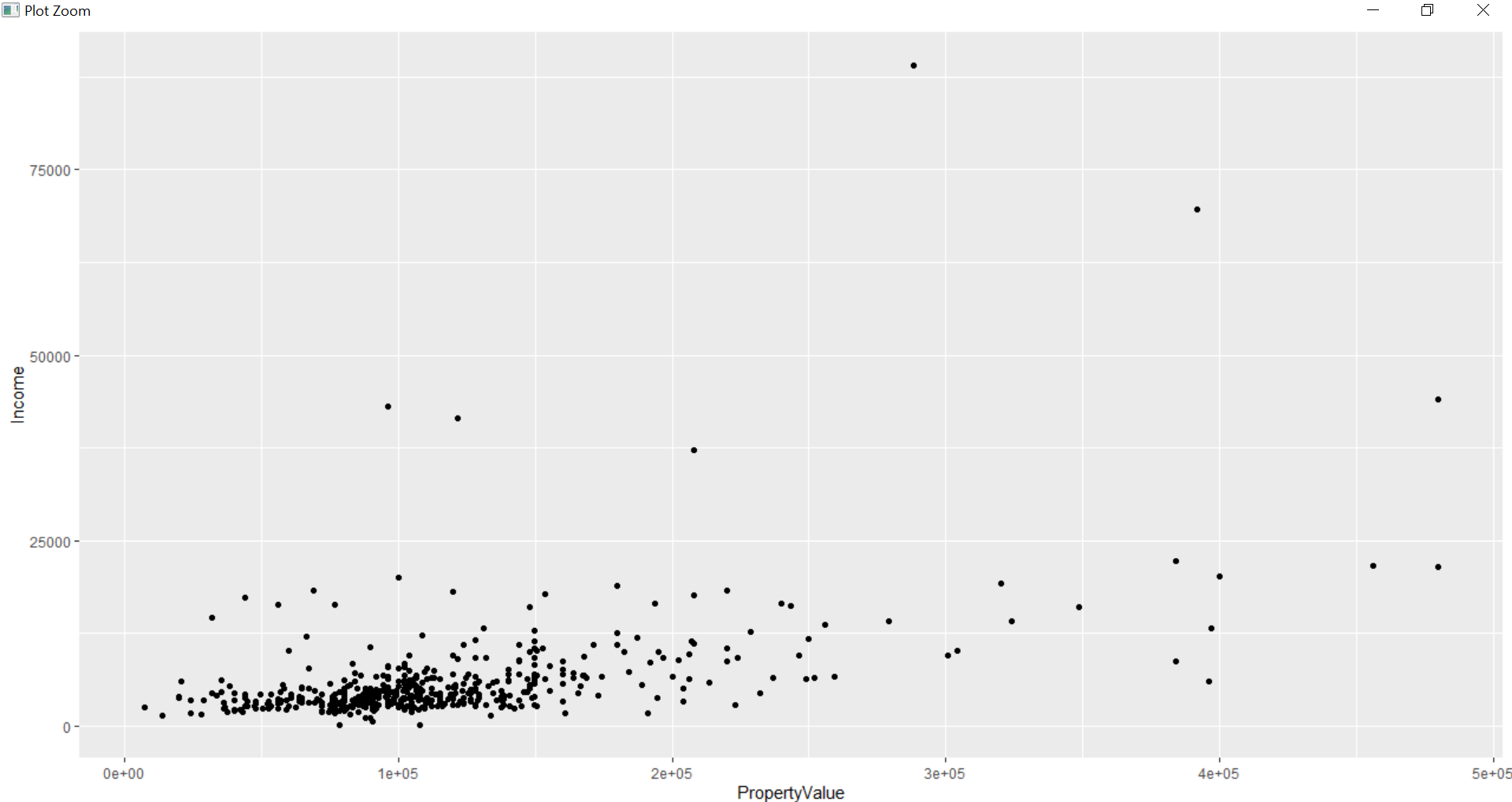


#6.2 Revese for the first

ggplot(house\_1,aes(x=PropertyValue,y=Income))+geom\_point()

Console:

|  |
| --- |
| #6.2 Revese for the first  > ggplot(house\_1,aes(x=PropertyValue,y=Income))+geom\_point() |
|  |
| |  | | --- | | > | |

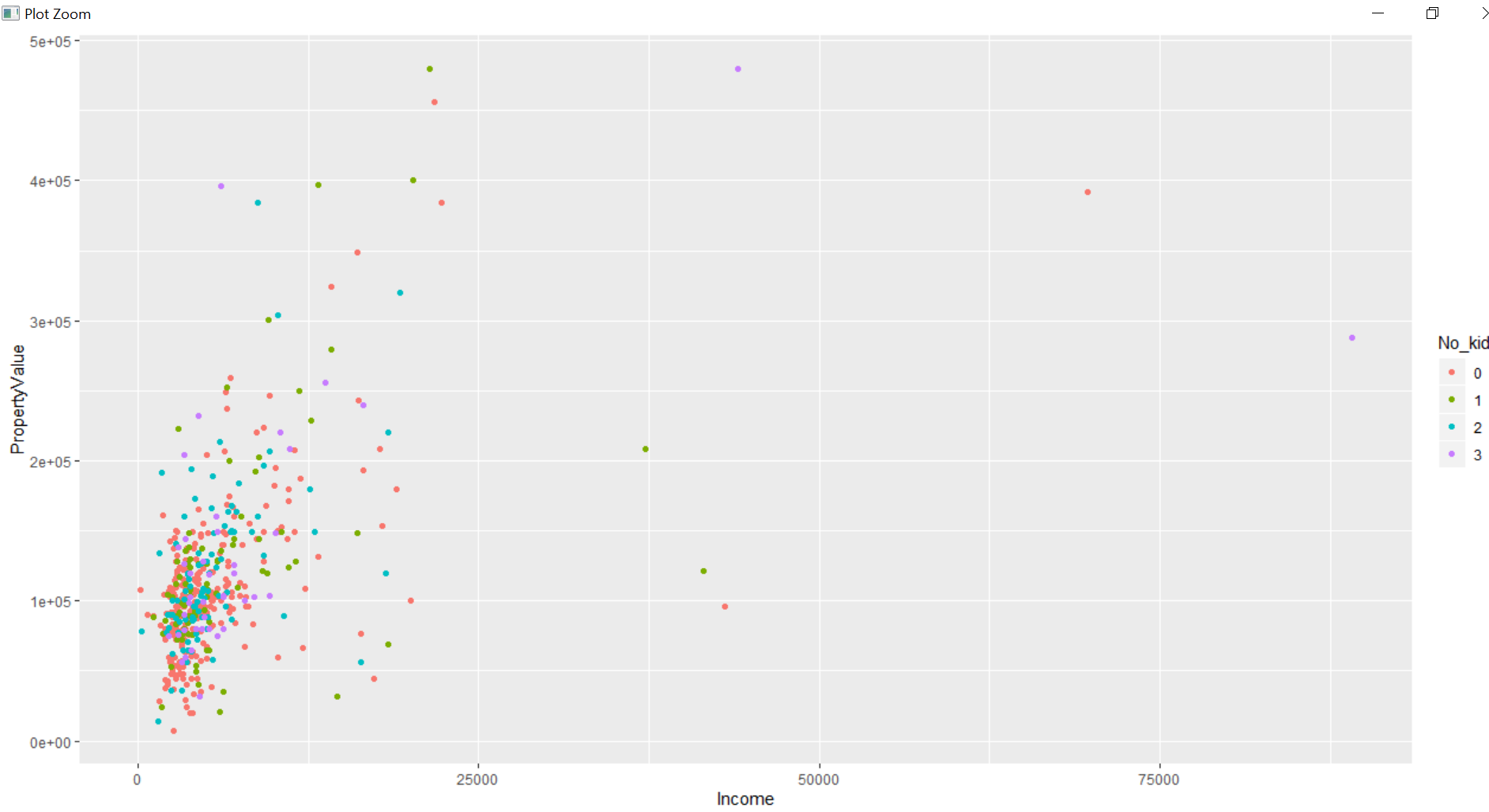


#6.3Map Income to x and Property Value to y and No of kids to col

ggplot(house\_1,aes(y=PropertyValue,x=Income,col=No\_kids))+geom\_point()

console:

|  |
| --- |
| #6.3Map Income to x and Property Value to y and No of kids to col  > ggplot(house\_1,aes(y=PropertyValue,x=Income,col=No\_kids))+geom\_point() |
|  |
| |  | | --- | | > | |



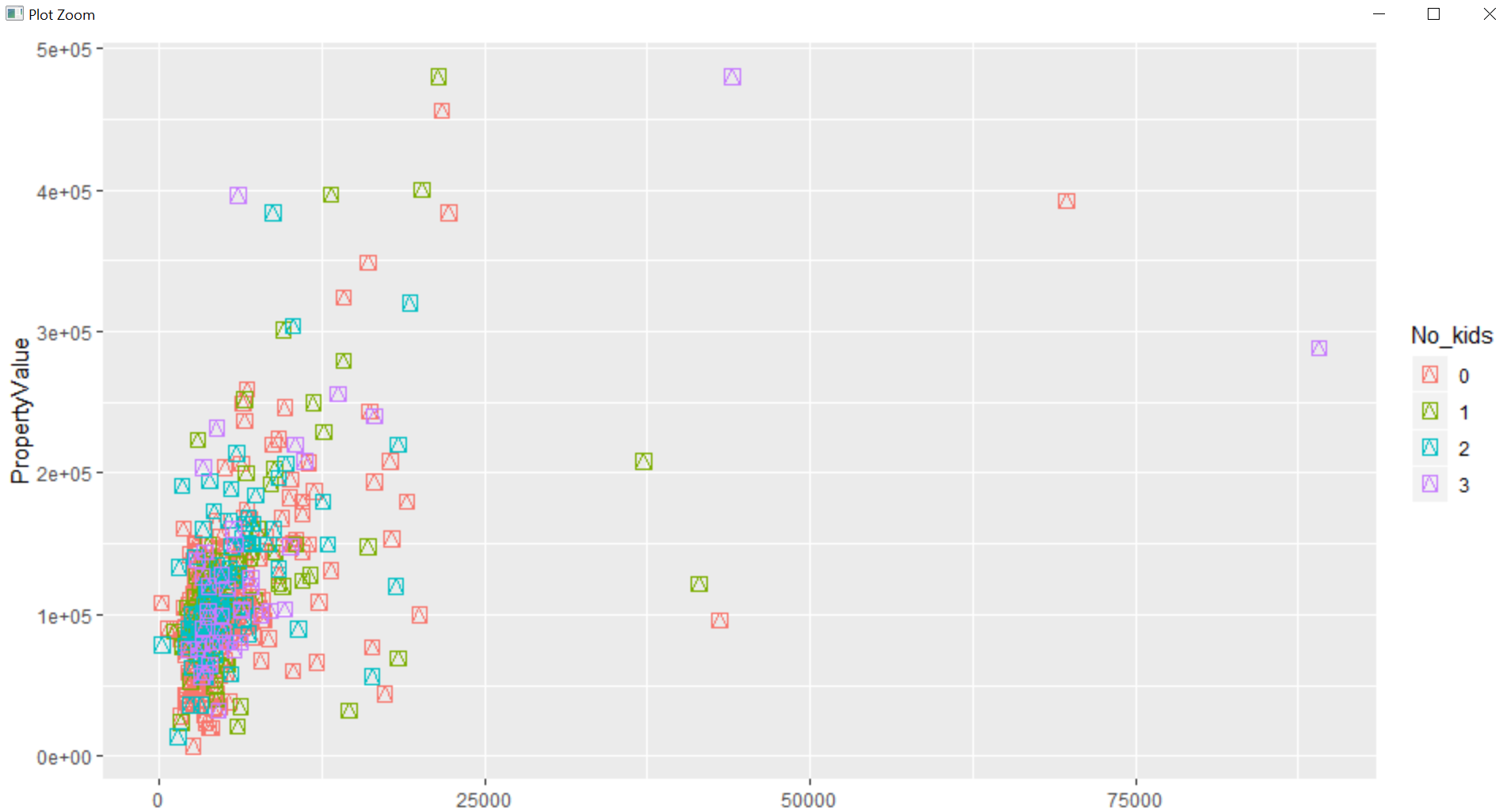
#6.4 Change on shape and size of above plot

ggplot(house\_1,aes(y=PropertyValue,x=Income,col=No\_kids))+geom\_point(shape=14,size=3)

console:

#6.4 Change on shape and size of above plot

> ggplot(house\_1,aes(y=PropertyValue,x=Income,col=No\_kids))+geom\_point(shape=14,size=3)



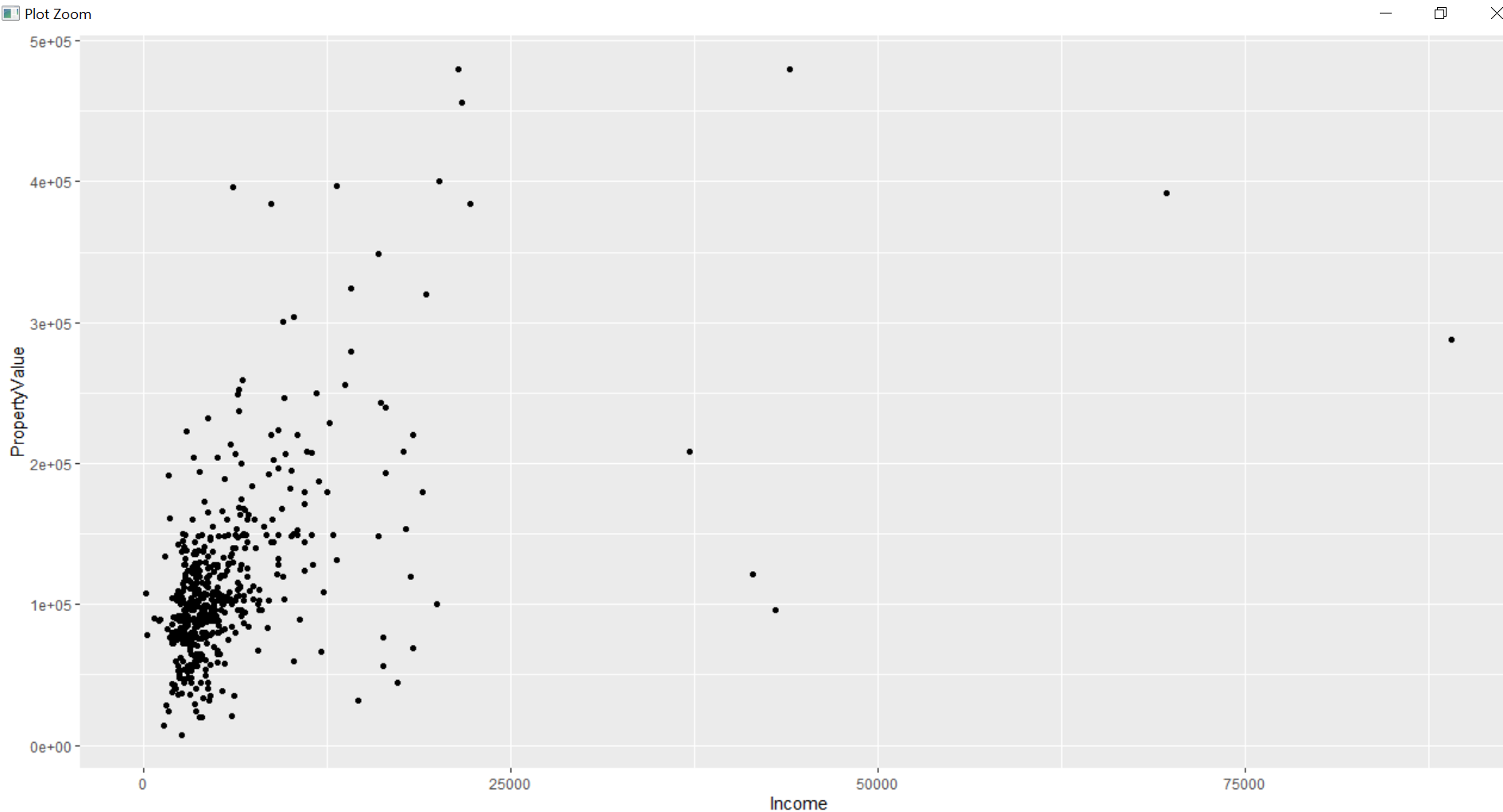
#7.Geometry

#7.1 scatter-plot x- Income and Y - PropertyValue

ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_point()

Console:

|  |
| --- |
| > #7.Geometry  > #7.1 scatter-plot x- Income and Y - PropertyValue  > ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_point() |
|  |
| |  | | --- | | > | |

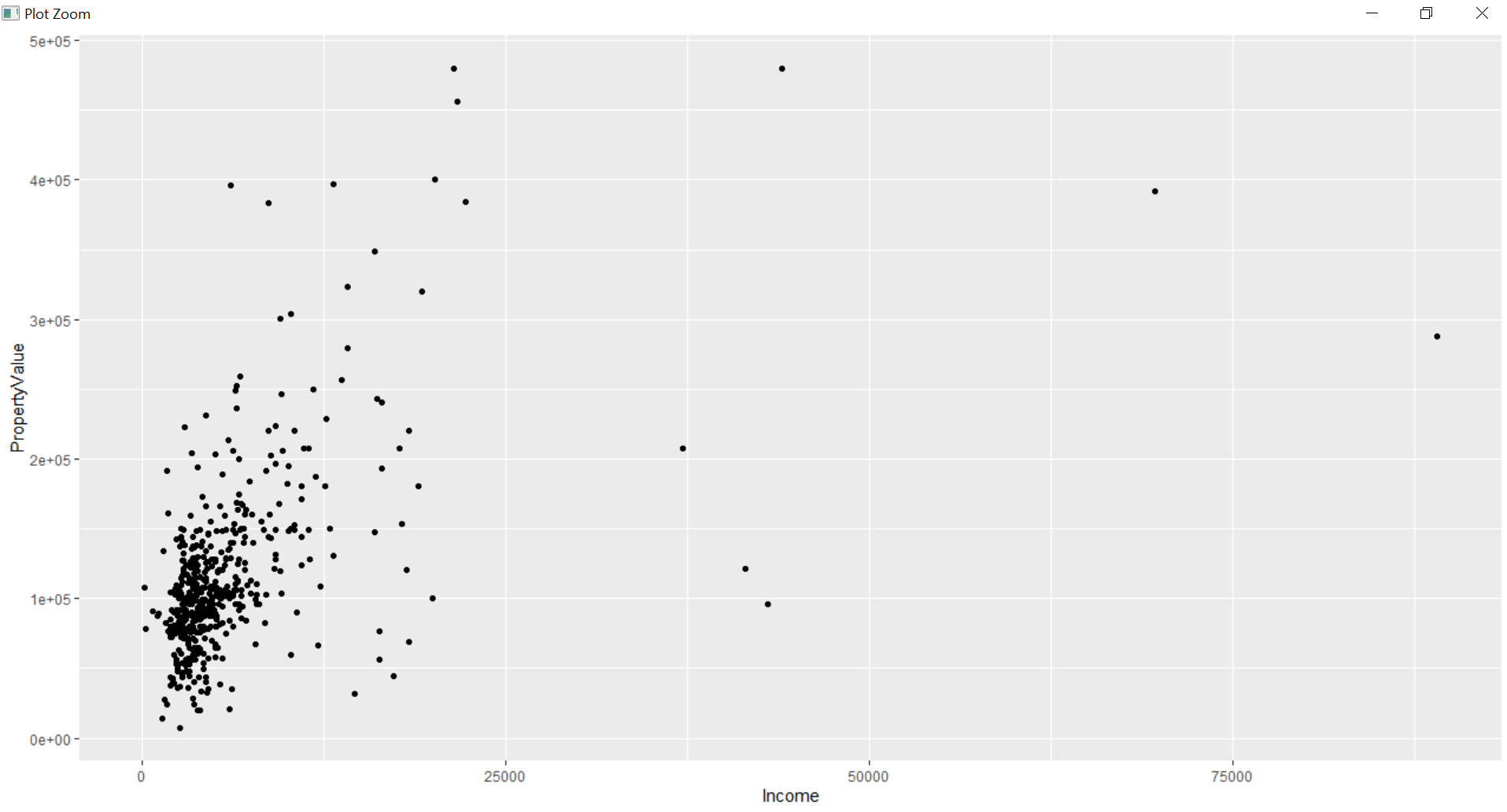


#7.2 geom\_jitter()

ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_jitter()

Console:

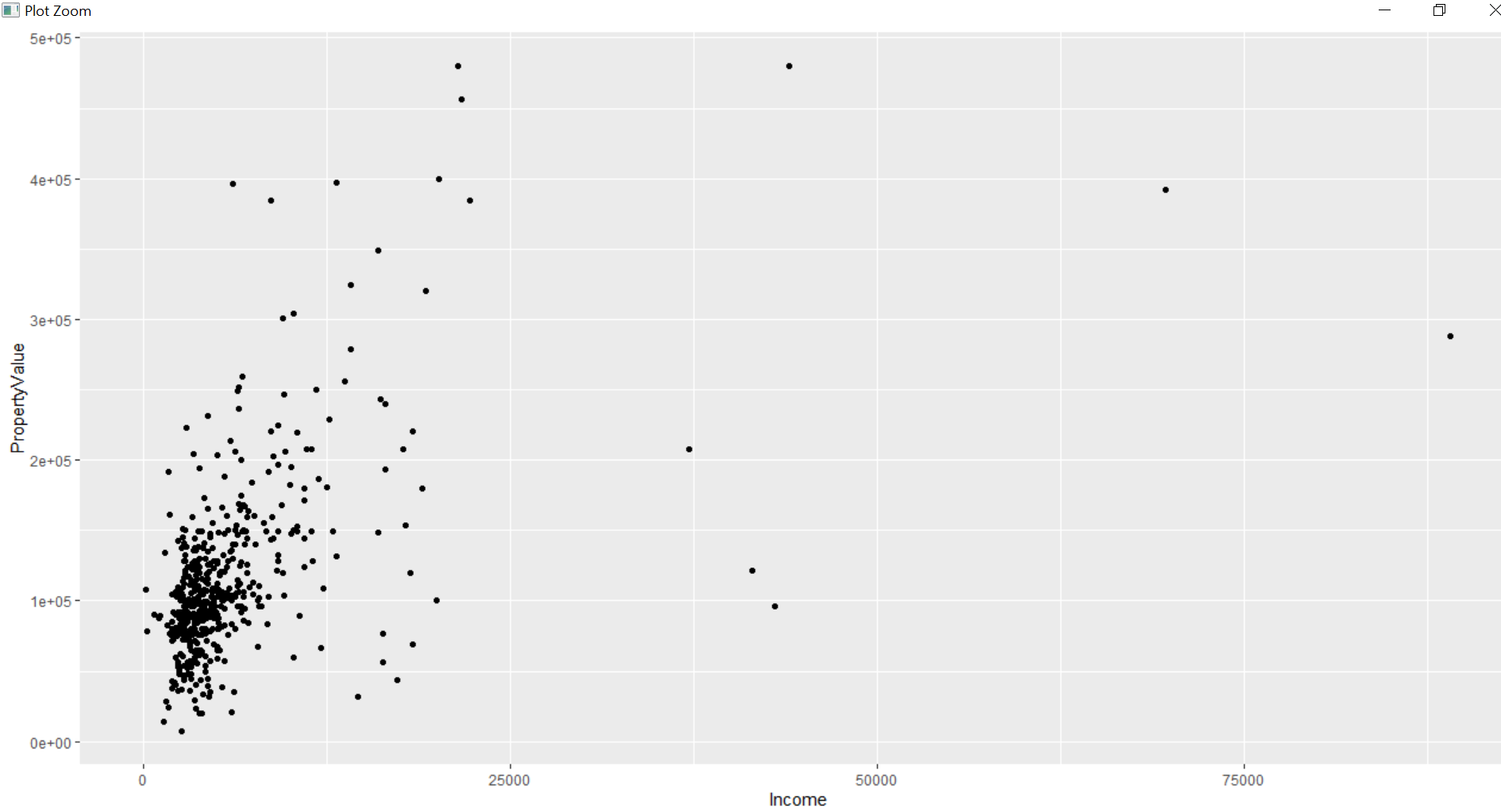
|  |
| --- |
| #7.2 geom\_jitter()  > ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_jitter() |
|  |
| |  | | --- | | > | |



#7.3 Width as 0.1 in geom\_jitter

ggplot(house\_1,aes(y=PropertyValue,x=Income))+geom\_jitter(width = 0.1)

Console:



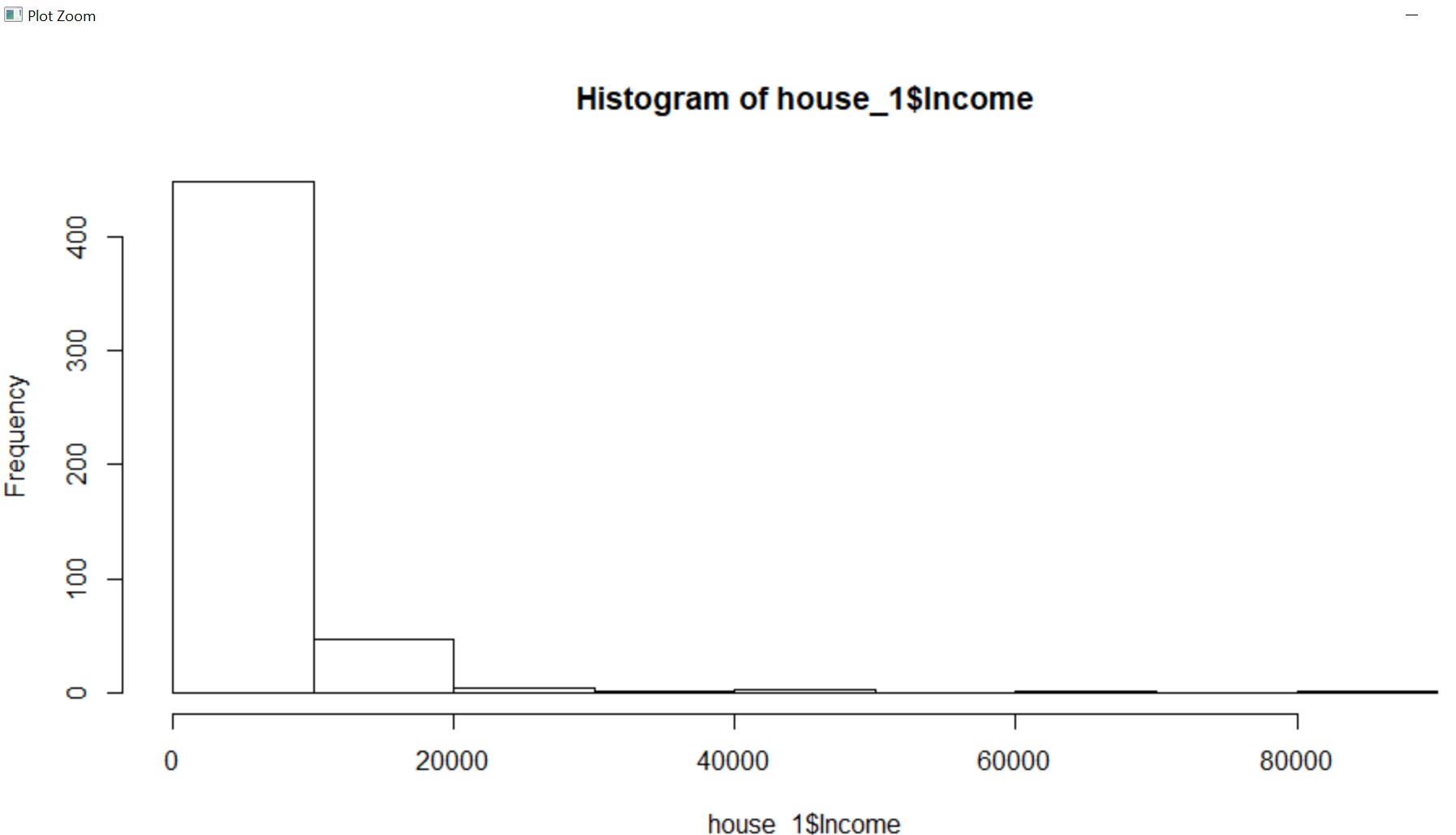
#8.Histrogram:

#8.1 Univariant histrogram on income

hist(house\_1$Income)

Console:

|  |
| --- |
| > #8.Histrogram:  > #8.1 Univariant histrogram on income  > hist(house\_1$Income) |
|  |
| |  | | --- | | > | |

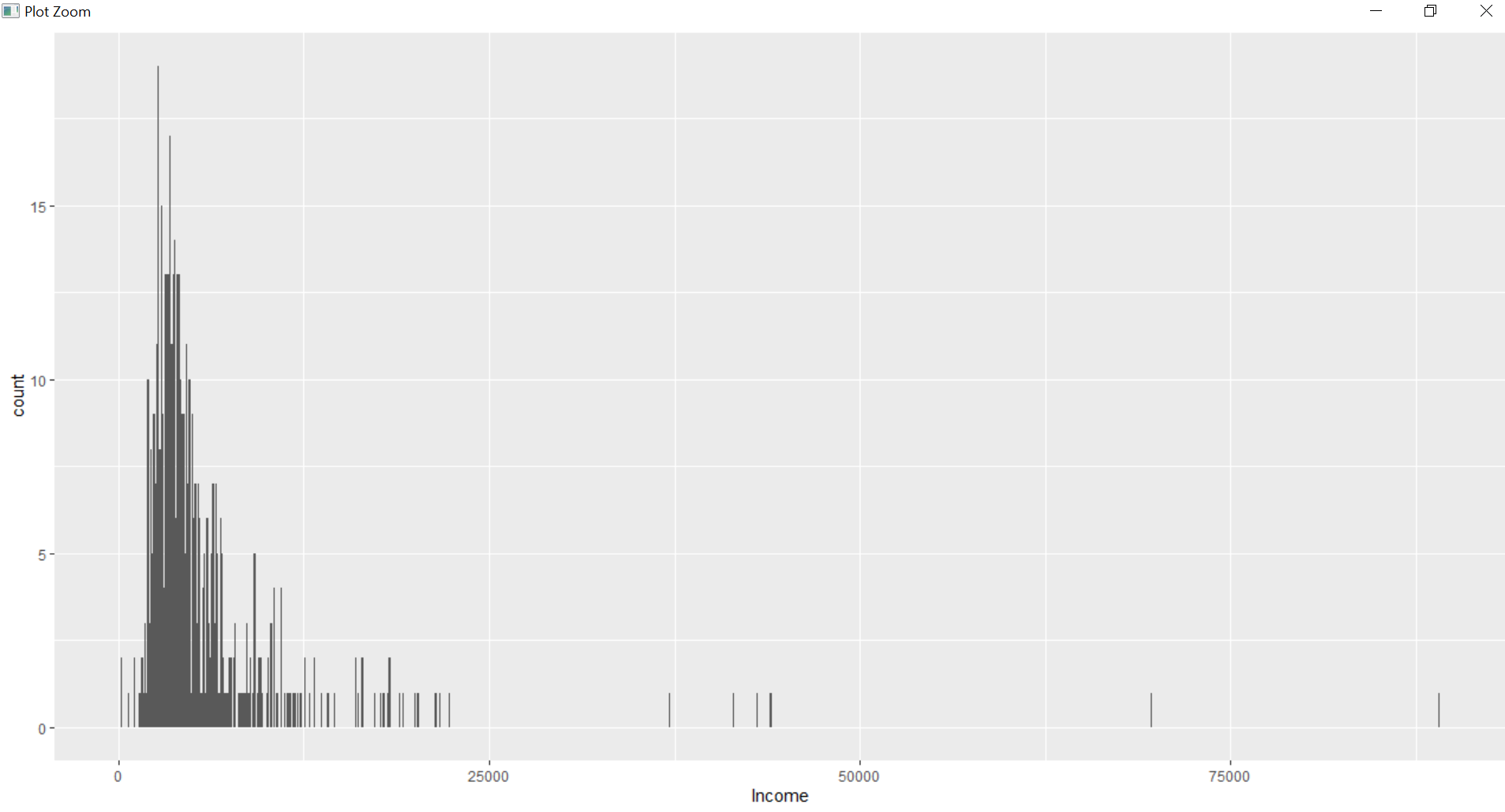


#8.2 binwidth to 100

ggplot(house\_1,aes(x=Income))+geom\_histogram(binwidth = 100)

console:

|  |
| --- |
| #8.2 binwidth to 100  > ggplot(house\_1,aes(x=Income))+geom\_histogram(binwidth = 100) |
|  |
| |  | | --- | | > | |

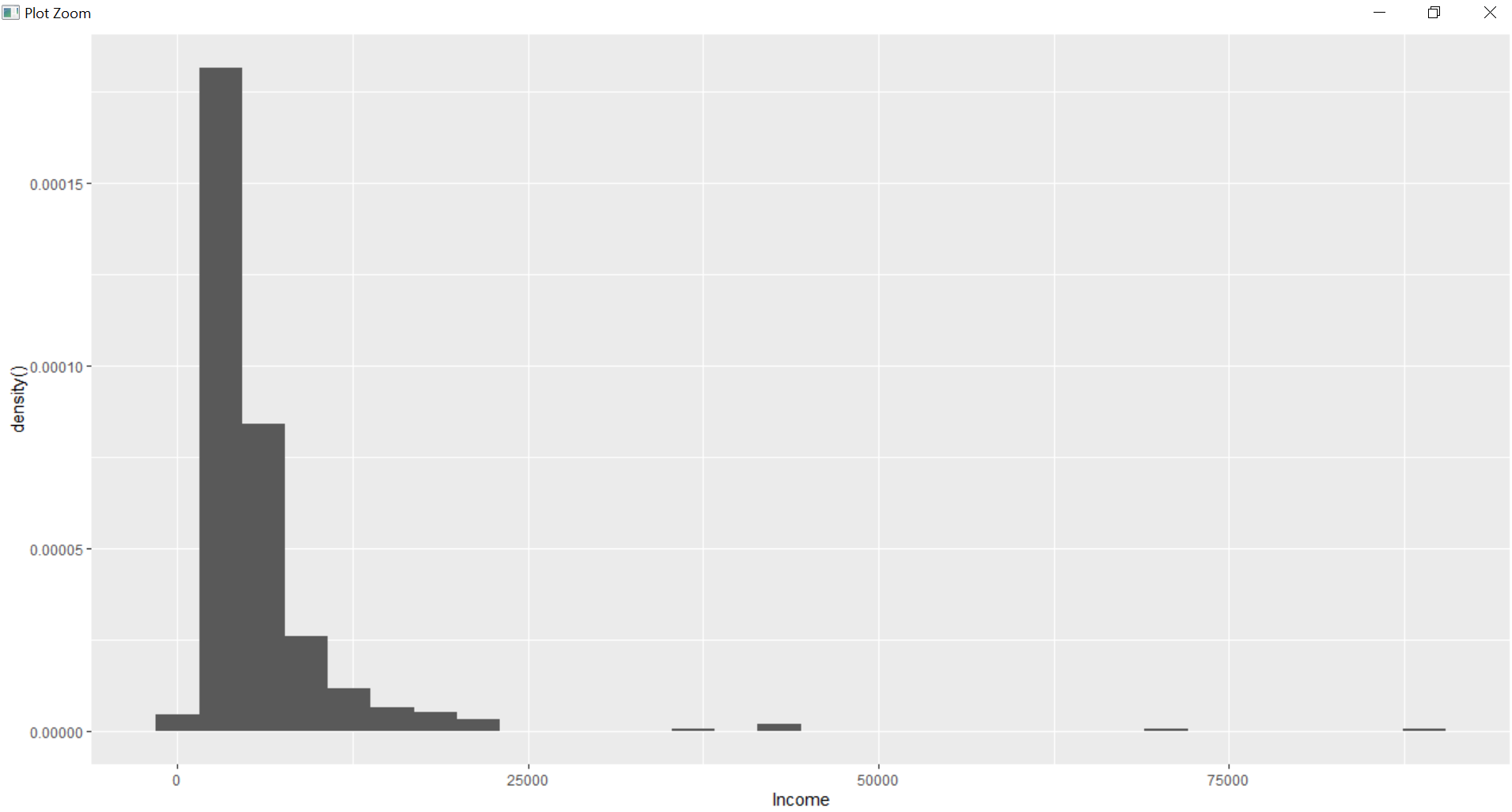


#8.3 density to y

ggplot(house\_1,aes(x=Income))+geom\_histogram(aes(y=..density..))

console:

|  |
| --- |
| > #8.3 density to y  > ggplot(house\_1,aes(x=Income))+geom\_histogram(aes(y=..density..))  `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`. |
|  |
|  |

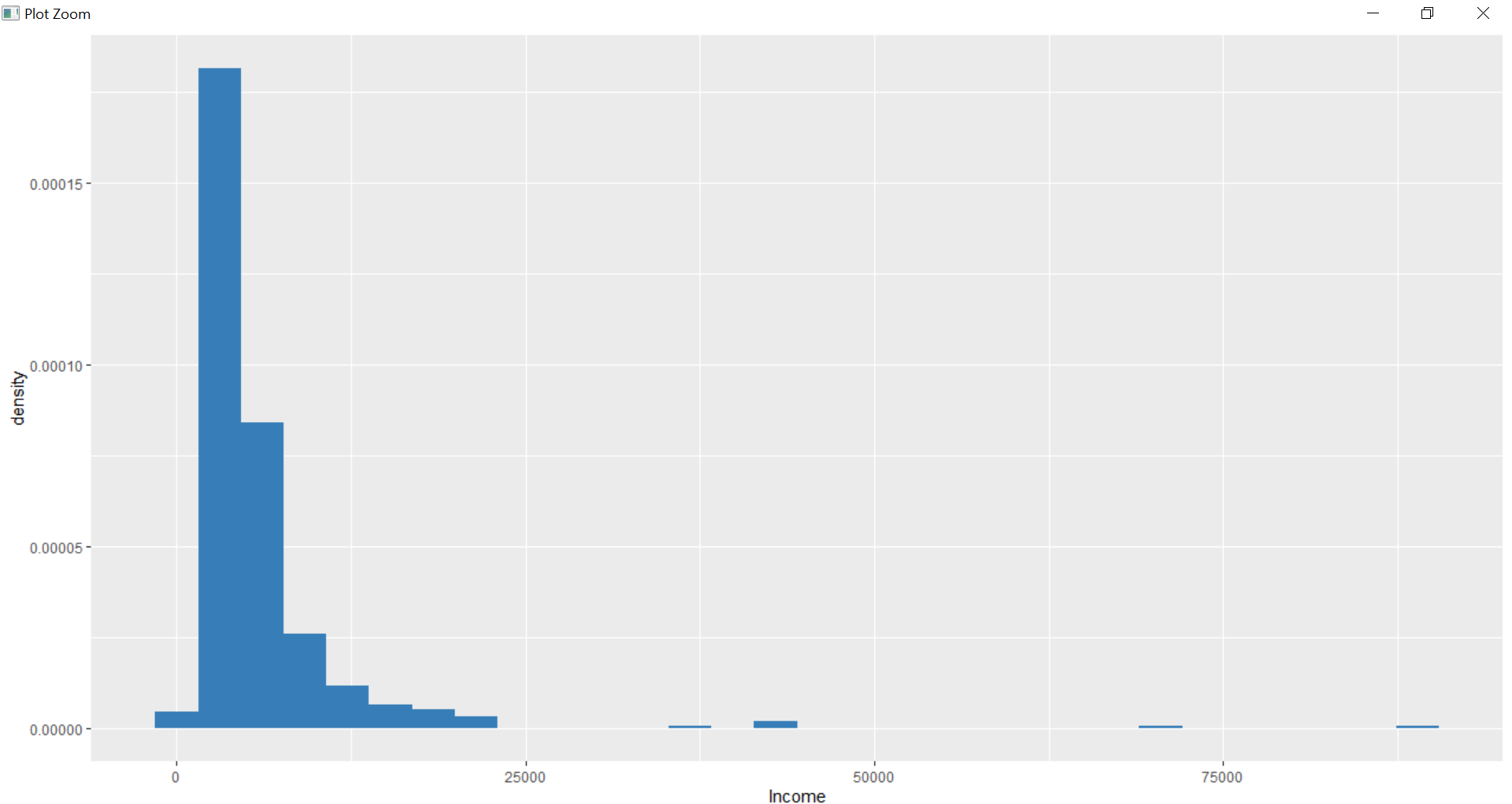


#8.4 set fill attribute

ggplot(house\_1,aes(x=Income))+geom\_histogram(aes(y=..density..),fill="#377EB8")

Console:

|  |
| --- |
| > #8.4 set fill attribute  > ggplot(house\_1,aes(x=Income))+geom\_histogram(aes(y=..density..),fill="#377EB8")  `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`. |
|  |
| |  | | --- | | > | |



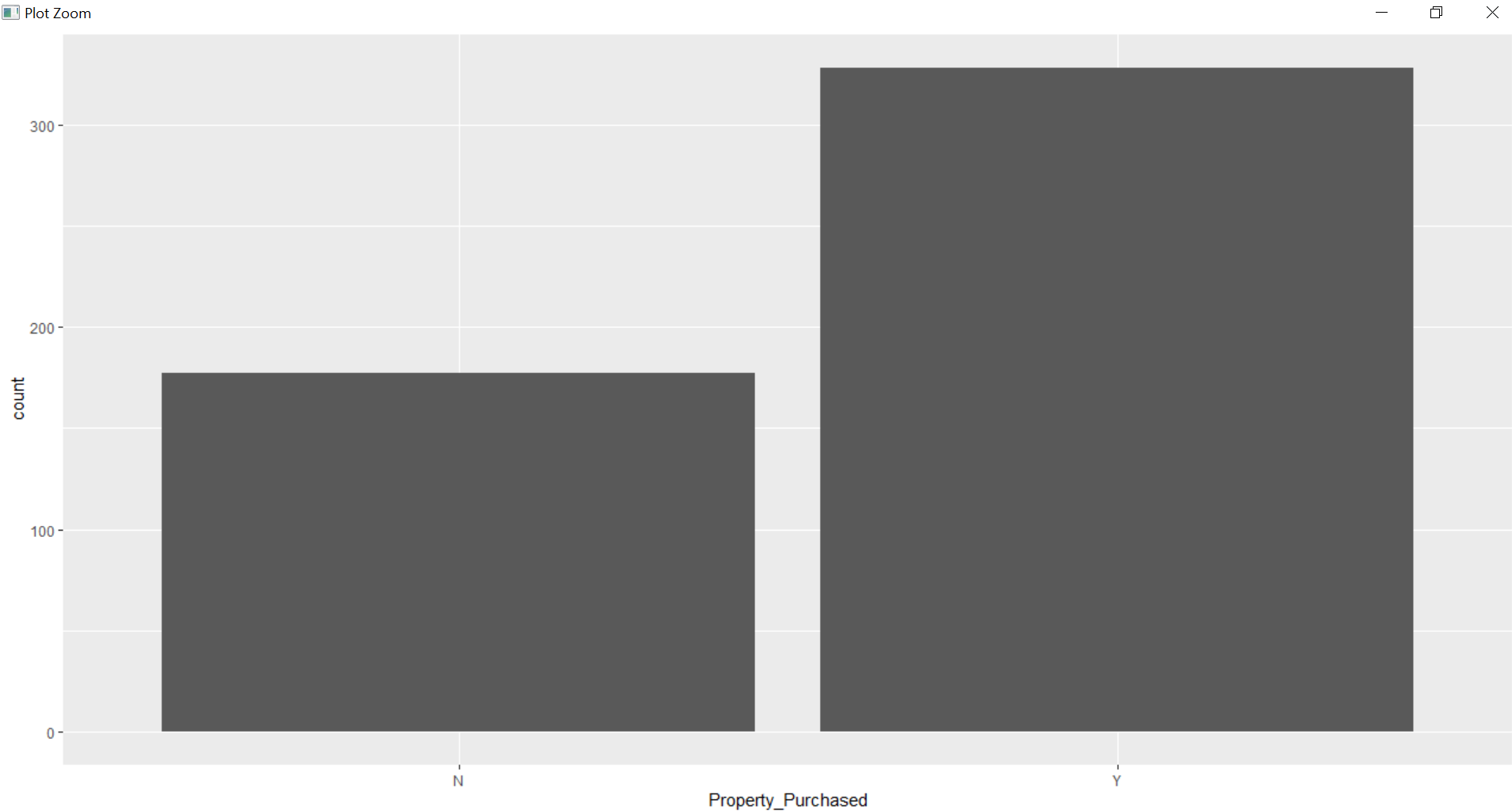
#9.Bar plot

#9.1 bar plot for property\_purchase

ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar()

Console:

|  |
| --- |
| > #9.1 bar plot for property\_purchase  > ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar() |
|  |
| |  | | --- | | > | |

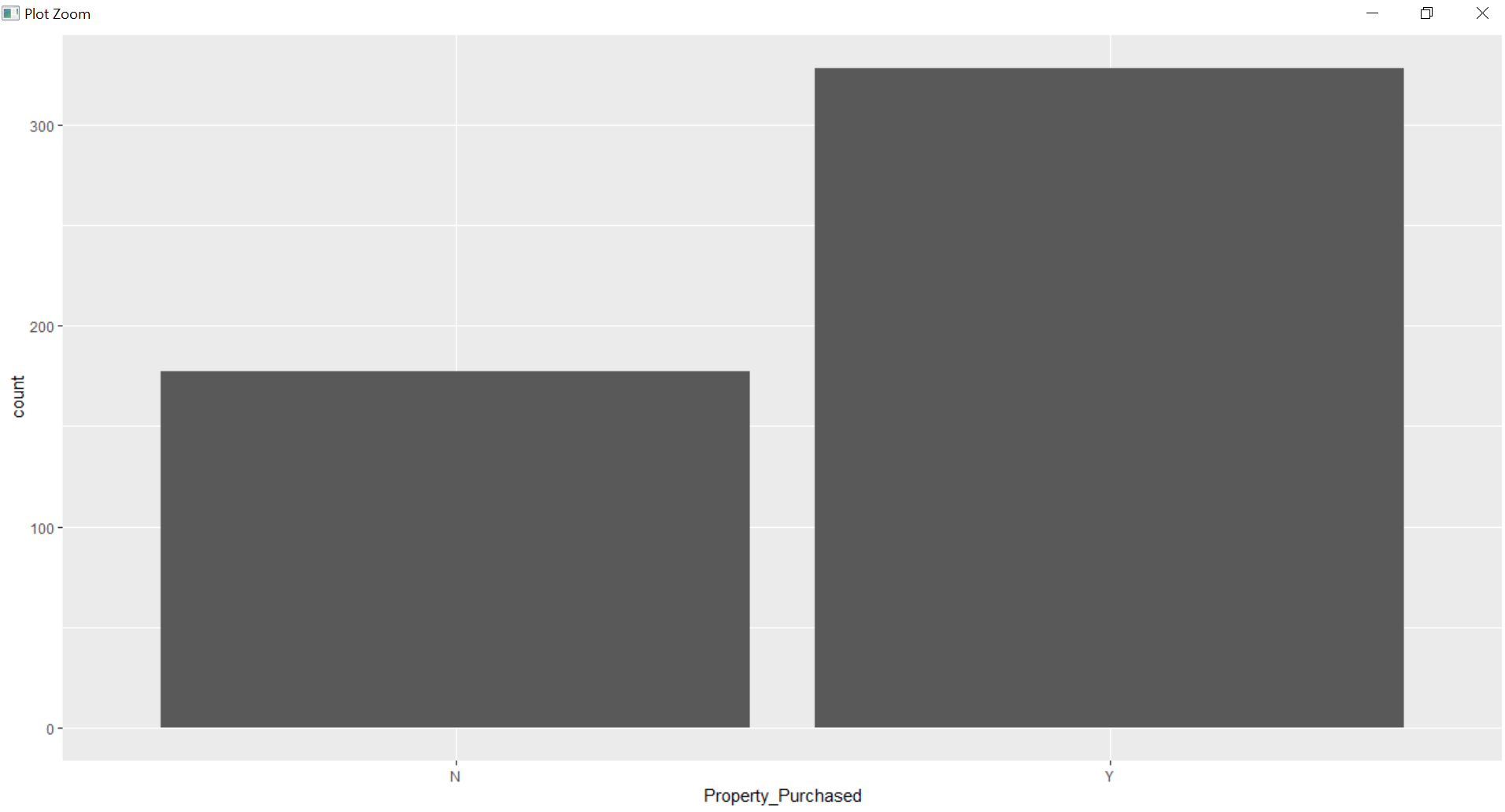


#9.2 position -stack

ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_stack())

Console:

|  |
| --- |
| #9.2 position -stack  > ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_stack()) |
|  |
| |  | | --- | | > | |

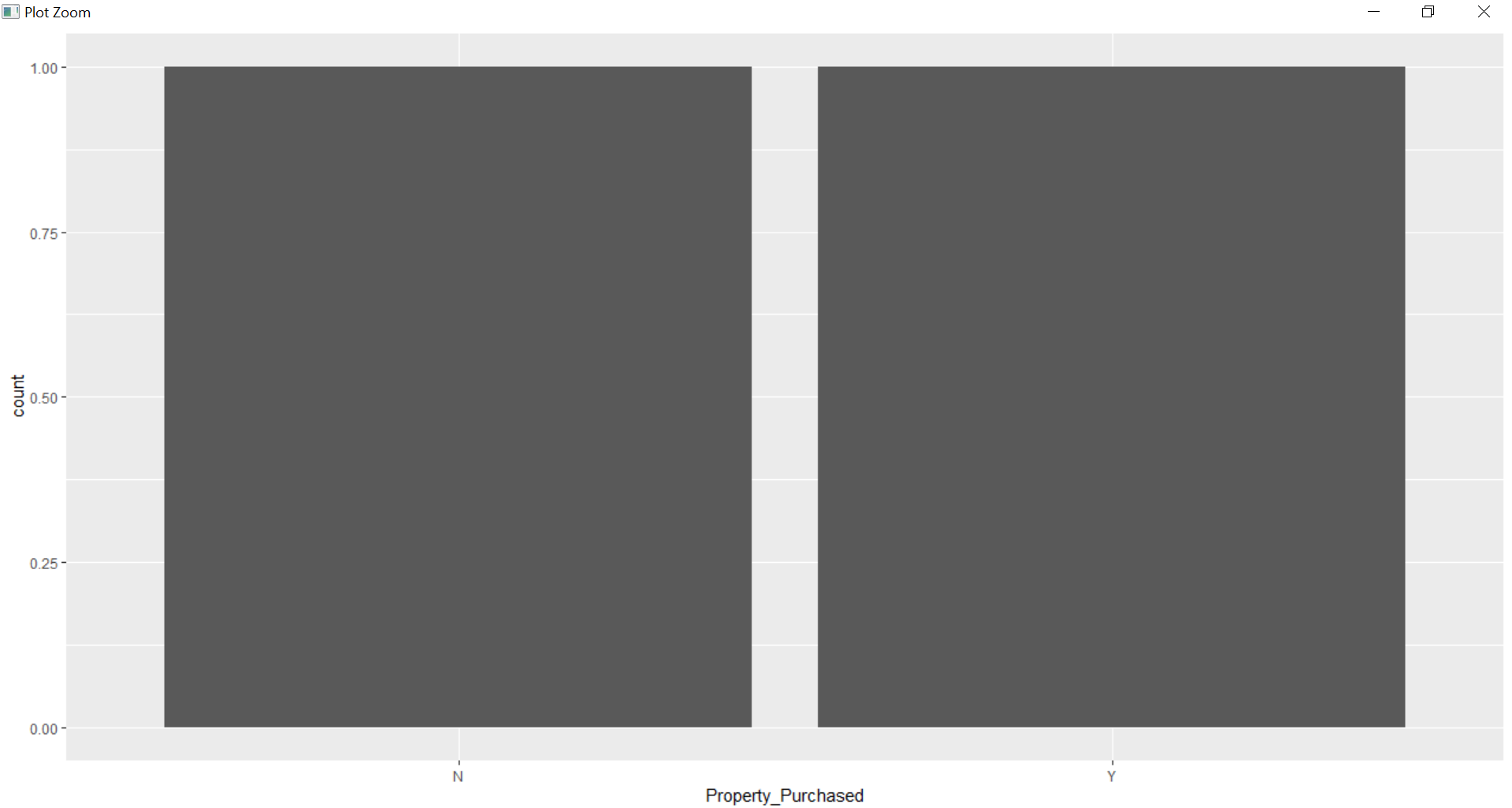


#9.2 position -fill

ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_fill())

#9.2 position -fill

> ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_fill())



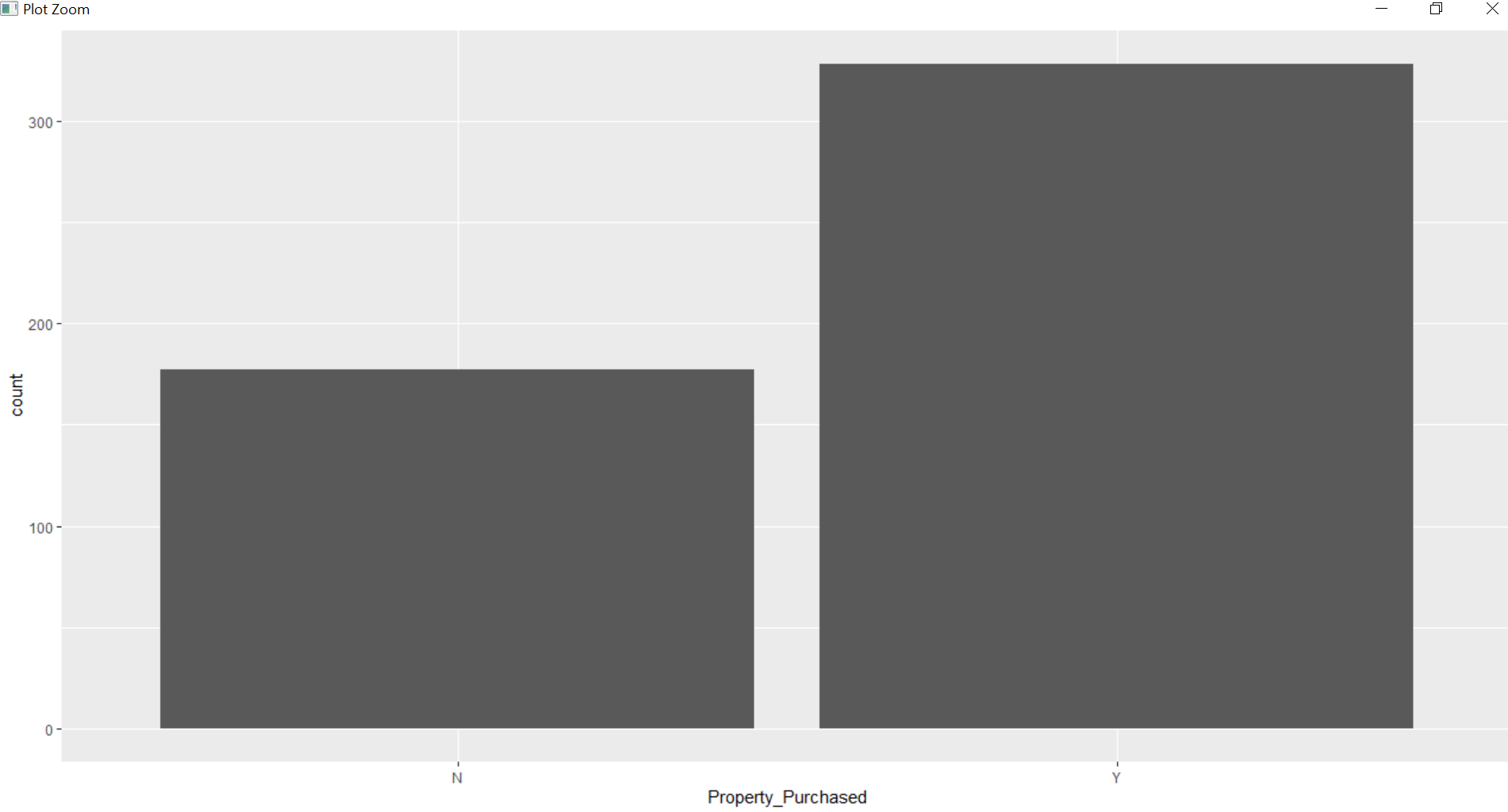
#9.4 position -dodge

ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_dodge())

Console:

> #9.4 position -dodge

> ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = position\_dodge())



#10 Overlapping bar

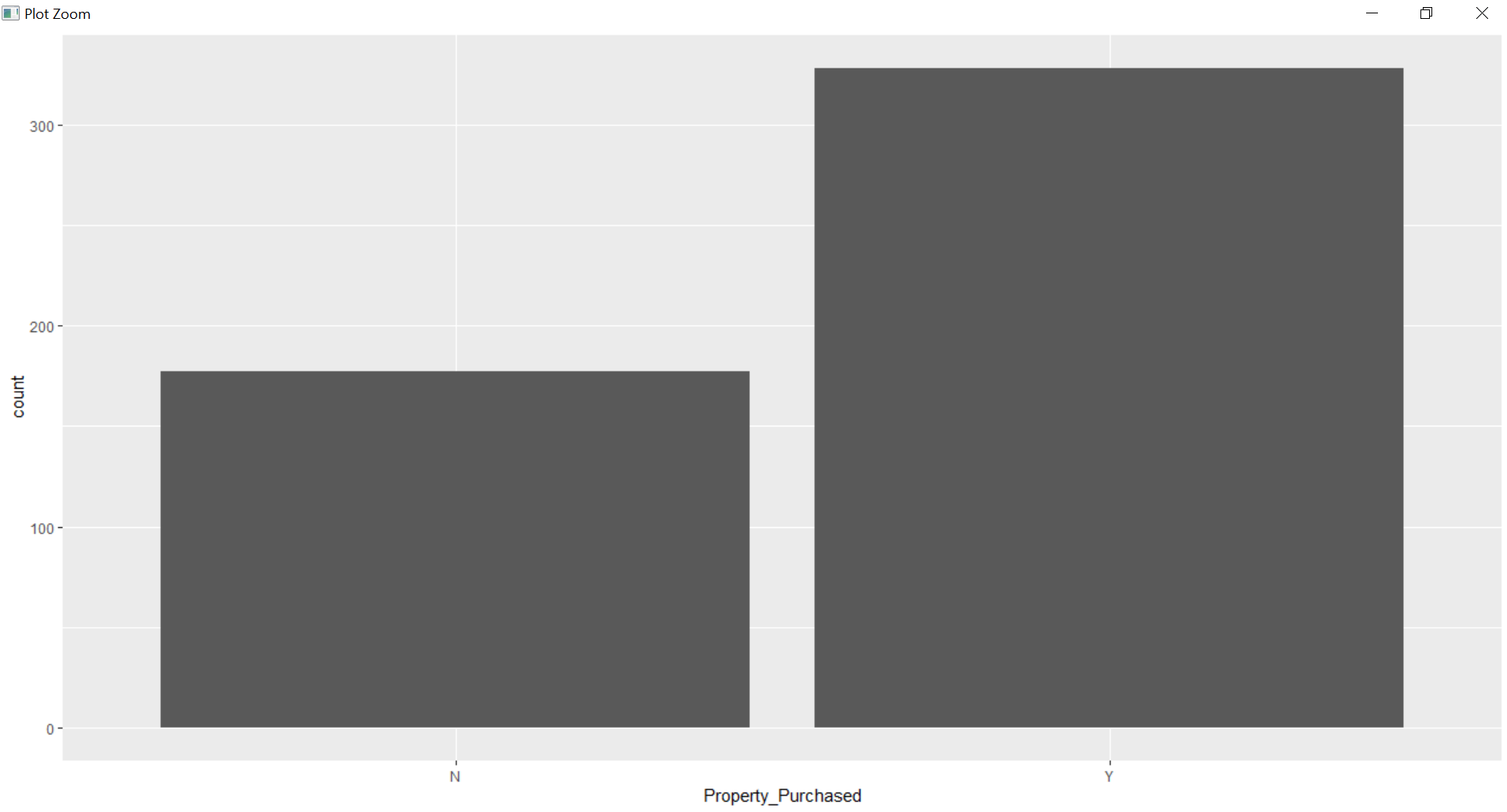
#10.2 posn\_d definintion in above plot

posn\_d <- position\_dodge(width = 0.7)

ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = posn\_d)

console:

|  |
| --- |
| #10 Overlapping bar  > #10.2 posn\_d definintion in above plot  > posn\_d <- position\_dodge(width = 0.7)  > ggplot(house\_1,aes(x=Property\_Purchased))+geom\_bar(position = posn\_d) |
|  |
| |  | | --- | | > | |



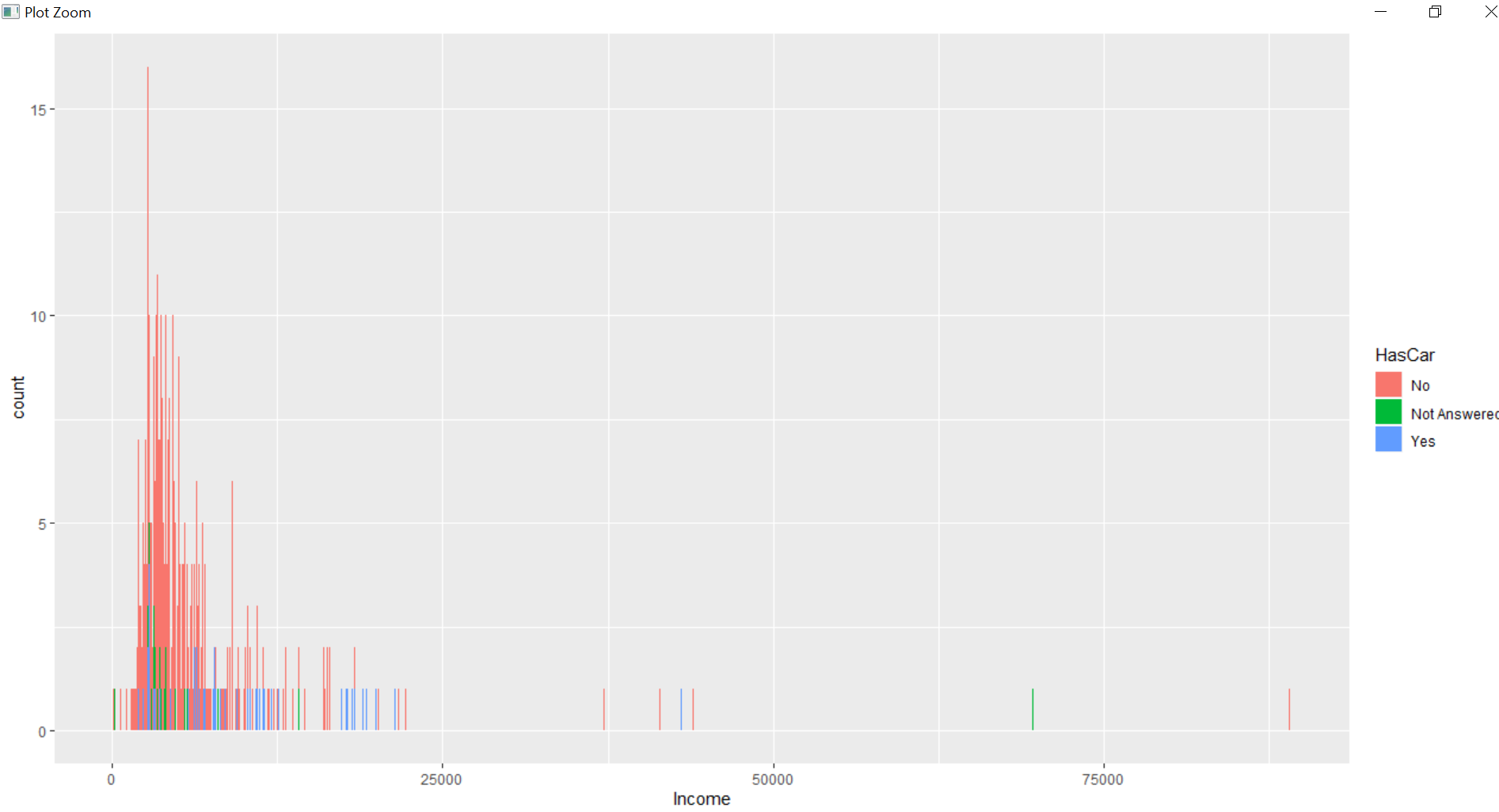
#11.overlapping histrograms

#11.1 color defined by income

ggplot(house\_1,aes(x=Income,fill=HasCar))+geom\_histogram(binwidth = 50)

Console

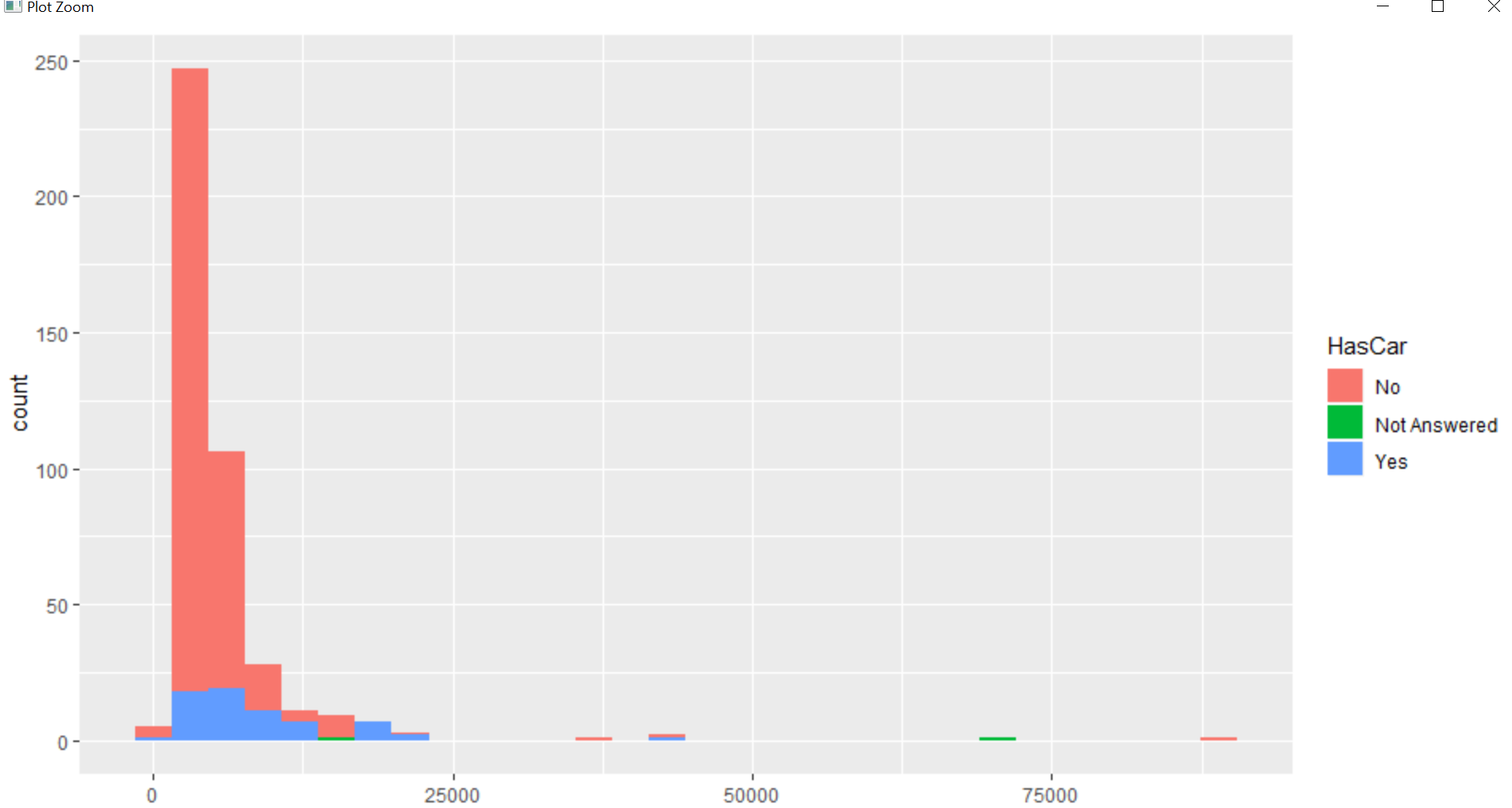
|  |
| --- |
| #11.overlapping histrograms  > #11.1 color defined by income  > ggplot(house\_1,aes(x=Income,fill=HasCar))+geom\_histogram(binwidth = 50) |
|  |
| |  | | --- | | > | |



#11.2 position is identity

ggplot(house\_1,aes(x=Income,fill=HasCar))+geom\_histogram(position = position\_identity())

|  |
| --- |
| > #11.2 position is identity  > ggplot(house\_1,aes(x=Income,fill=HasCar))+geom\_histogram(position = position\_identity())  `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`. |
|  |
| |  | | --- | | > | |



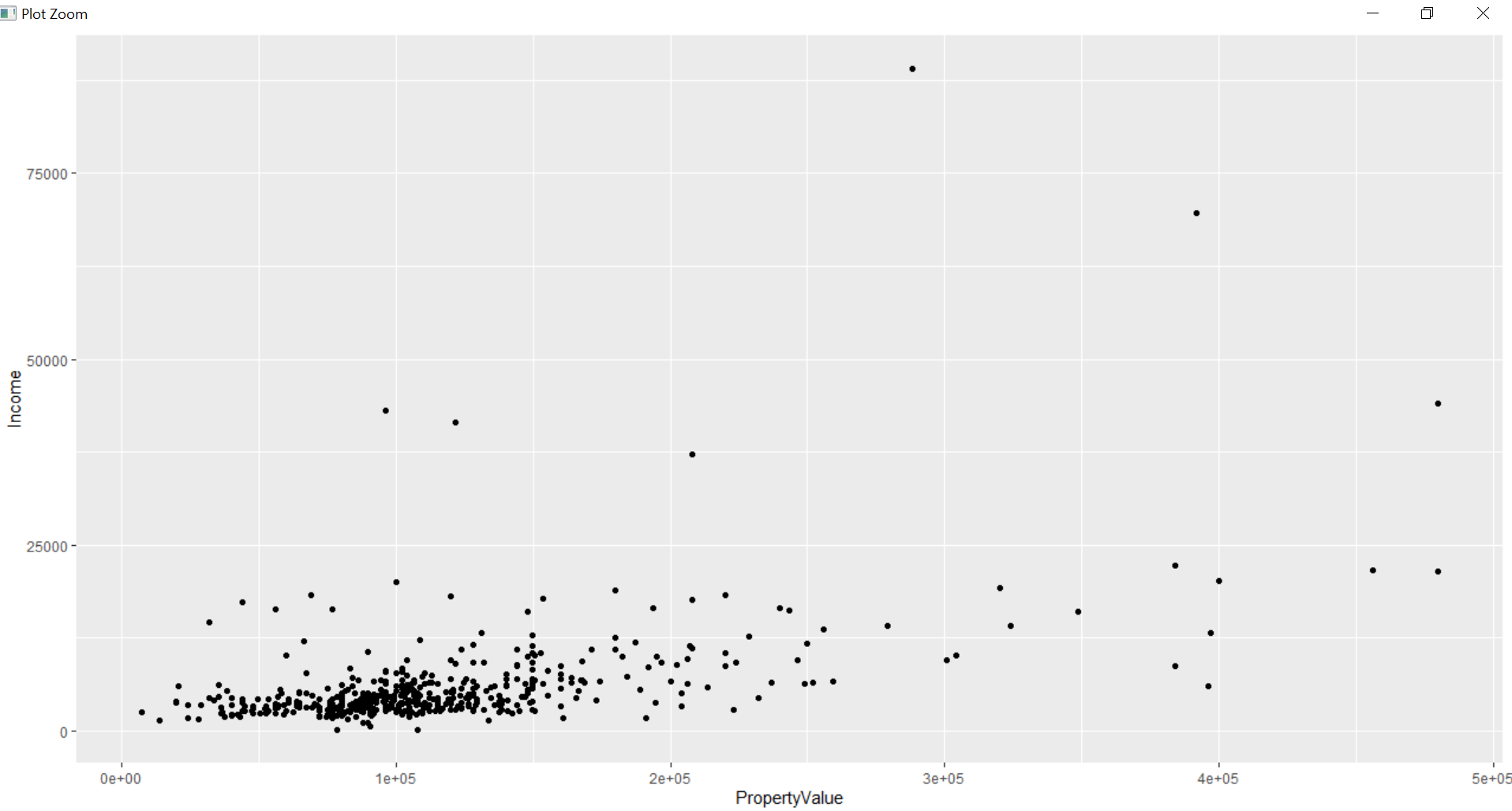
#12.Faceting

#12.1 scatter plot between income and property value

ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()

Console:

|  |
| --- |
| > #12.Faceting  > #12.1 scatter plot between income and property value  > ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point() |
|  |
| |  | | --- | |  | |



#12.2 Separate row according to HasCar

ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

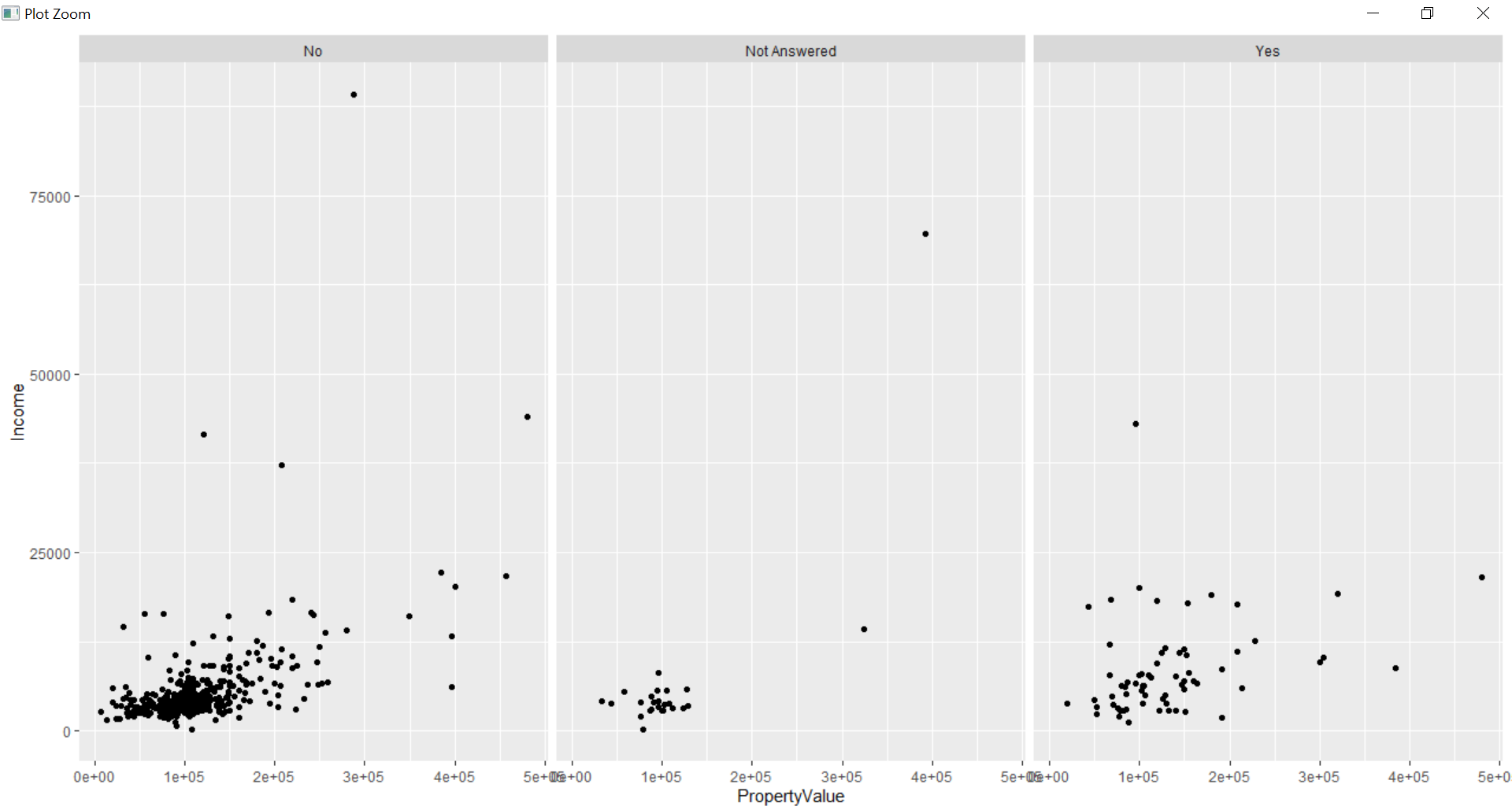
facet\_grid(.~HasCar)

Console:

> #12.2 Separate row according to HasCar

> ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

+ facet\_grid(.~HasCar)



#12.3 In 12.2 separte by no\_kids

ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

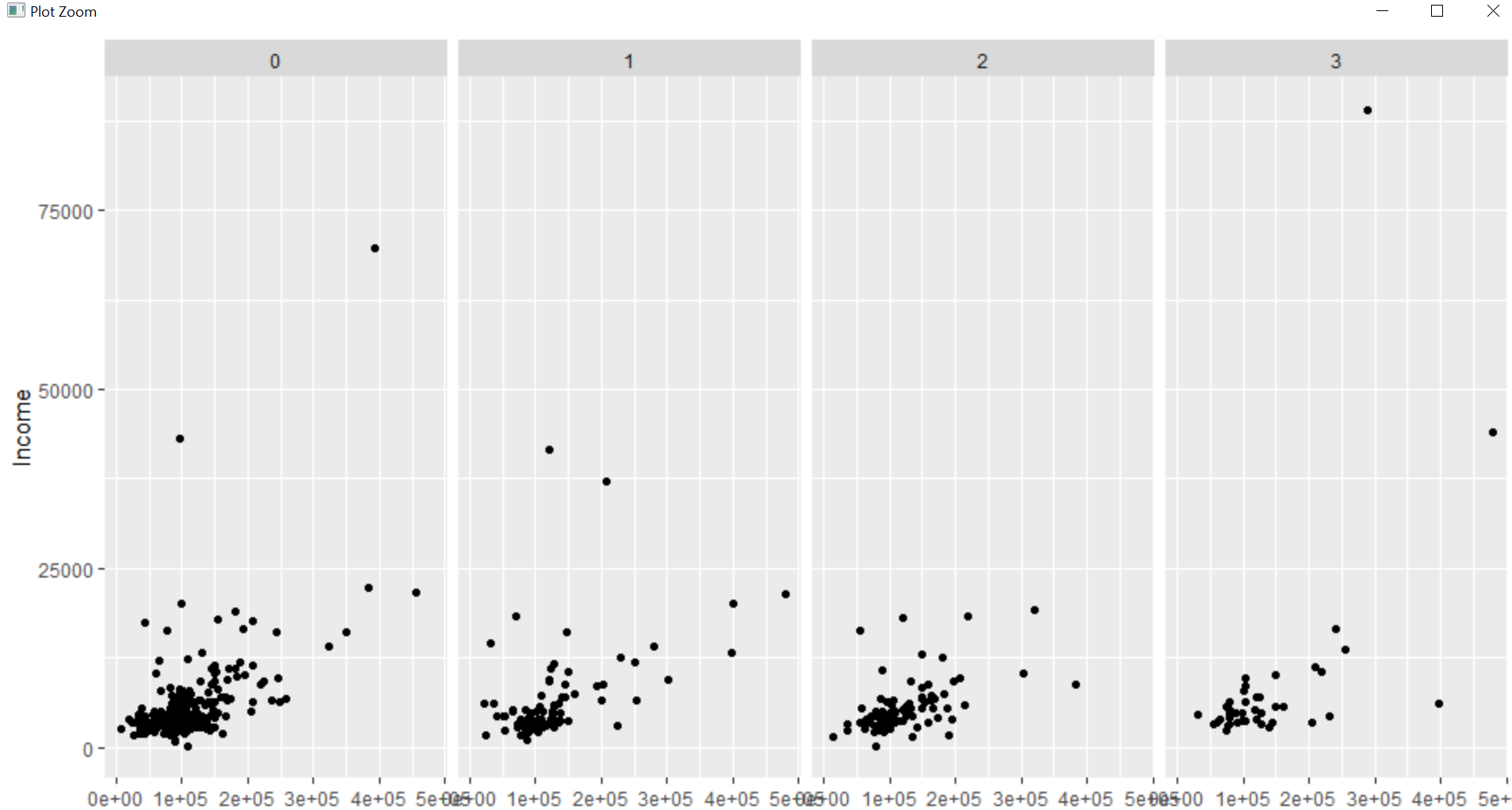
facet\_grid(.~No\_kids)

Console:

> #12.3 In 12.2 separte by no\_kids

> ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

+ facet\_grid(.~No\_kids)



#12.4 In 12.2 separte by no\_kids and hascar

ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

facet\_grid(.~HasCar)+facet\_wrap(.~No\_kids)

Console:

> #12.4 In 12.2 separte by no\_kids and hascar

> ggplot(house\_1,aes(y=Income,x=PropertyValue))+geom\_point()+

+ facet\_grid(.~HasCar)+facet\_wrap(.~No\_kids)

