Homework 2 Solution

Mohsen Nabian

5/28/2015

1.

a. Write a function that calculates the mean of any numeric vector you give it, without using the built-in mean() function.

```
calc_mean<-function(v)
{
    temp=0;
    for (i in 1:length(v))
    {
      temp=temp+v[i];
    }
    ans=temp/length(v);
    return(ans)
    print(ans)
}
calc_mean(c(1,2,3,4))
```

```
## [1] 2.5
```

b.Write a function that takes as its input a vector with four elements. If the sum of the first two elements is greater than the sum of the second two, the function returns the vector; otherwise it returns 0.

```
check_vector<-function(v)
{
  if(length(v)!=4) stop(' The vector must have 4 elements!');
  first_half=v[1]+v[2];
  second_half=v[3]+v[4];
  if (first_half>second_half)
    {
      return(v);
      print(v);
    }
  else
    {
      return(0);
      print(0);
    }
}
check_vector(c(1,2,3,4))
```

```
## [1] 0
```

```
check_vector(c(7,2,3,4))
```

```
## [1] 7 2 3 4
```

c.Write a function that calculates the Fibonacci sequence up to the nth element, where n is any number input into your function (its argument). The Fibonacci sequence is: 1, 1, 2, 3, 5, 8, 13, 21. . . , ie, each element is the sum of the previous two elements. One way to do this is to start off with the first two elements, c(1,1) and set an internal variable to this sequence. Then write a loop that counts up to n, where for each new element, you first calculate it by adding the last two elements of the growing sequence, and then stick that new number onto the growing sequence using c(). When the loop is finished, the function should return the final vector of Fibonacci numbers.

```
Fibonacci<-function(n)
{
  if (n==1 | n==2)
    {
      return(1);
    }
  else
    {
      return(Fibonacci(n-1)+Fibonacci(n-2));
    }
}
Fibonacci_write<-function(n)
{
  fibvec <- numeric(n)
  for (i in 1:n)
    {
      fibvec[i]<-Fibonacci(i);
    }
  return(fibvec);
  print(fibvec)
}
Fibonacci_write(10)
```

```
##  [1]  1  1  2  3  5  8 13 21 34 55
```

d.Create a 4x4 matrix of the numbers 1 through 16. Use apply to apply you function from (a) to each of the rows in your matrix.

```
MAT = matrix(1:16,nrow=4,ncol=4);
mean_MAT_row=apply(MAT,1,calc_mean);
print(mean_MAT_row);
```

```
## [1]  7  8  9 10
```

2.

a. Using the airquality dataset, constuct an aggregated dataset which shows the mean wind and ozone by month.

```
ds1<-aggregate(cbind(Wind,Ozone) ~ Month, data=airquality,mean)
ds1
```

```
##   Month      Wind    Ozone
## 1     5 11.457692 23.61538
## 2     6 12.177778 29.44444
## 3     7  8.523077 59.11538
## 4     8  8.565385 59.96154
## 5     9 10.075862 31.44828
```

b.Create the authors and books datasets following the example and data in the lecture, and then create a new data set by merging these two datasets by author, preserving all rows.

```
authors <- data.frame(surname = c("Tukey", "Venables", "Tierney", "Ripley", "McNeil"), nationality =
c("US", "Australia", "US", "UK", "Australia"),stringsAsFactors=FALSE);
books <- data.frame(name = c("Tukey", "Venables", "Tierney","Ripley", "Ripley", "McNeil", "R Core"),t
itle = c("Exploratory Data Analysis","Modern Applied Statistics ...","LISP-STAT","Spatial Statistic
s", "Stochastic Simulation","Interactive Data Analysis","An Introduction to R"),stringsAsFactors=FALS
E);
authors
```

```
##    surname nationality
## 1    Tukey          US
## 2 Venables   Australia
## 3  Tierney          US
## 4   Ripley          UK
## 5   McNeil   Australia
```

```
books
```

```
##       name                          title
## 1    Tukey     Exploratory Data Analysis
## 2 Venables Modern Applied Statistics ...
## 3  Tierney                      LISP-STAT
## 4   Ripley             Spatial Statistics
## 5   Ripley          Stochastic Simulation
## 6   McNeil     Interactive Data Analysis
## 7   R Core           An Introduction to R
```

```
bookmerge <- merge(authors, books, by.x="surname",by.y="name", all.x=TRUE, all.y=TRUE )
```

c.Take the following string and replace every instance of “to” or “To” with “2” “To be, or not to be – that is the question: Whether ’tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And by opposing end them. To die – to sleep – No more…”

```
gout1 <- gsub("to","2","To be, or not to be -- that is the question: Whether 'tis nobler in the mind
to suffer
The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles,
And by opposing end them. To die -- to sleep -- No more...");

gout2<-gsub("To","2",gout1)
gout2
```

```
## [1] "2 be, or not 2 be -- that is the question: Whether 'tis nobler in the mind 2 suffer\nThe slin
gs and arrows of outrageous fortune, Or 2 take arms against a sea of troubles, \nAnd by opposing end
them. 2 die -- 2 sleep -- No more..."
```

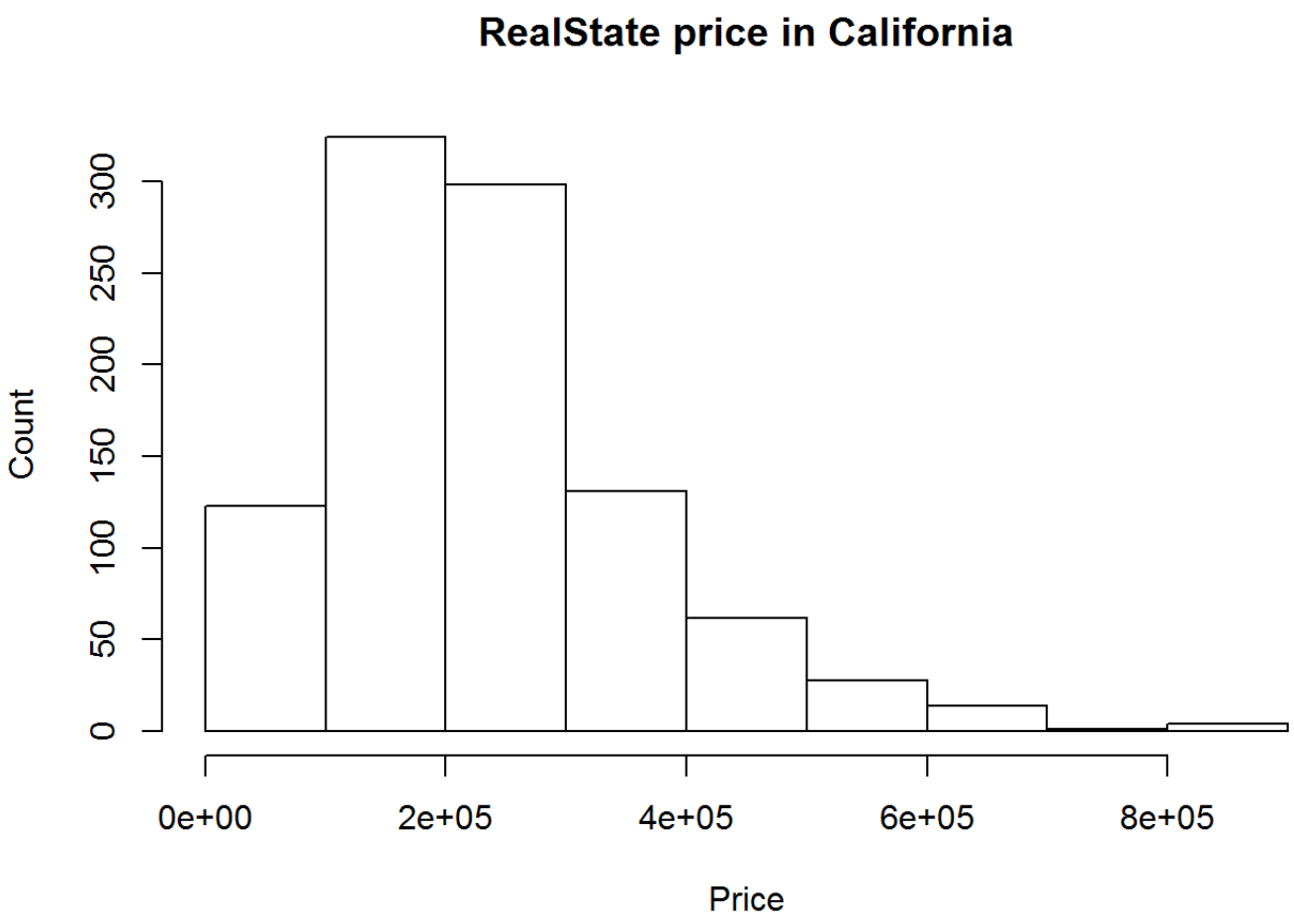
3.

a. Create a histogram using the base R graphics using some dataset or variable other than the one in the lessons. Always make sure your graph has well-labeled x and y axes and an explanatory title.

```
setwd("C:/Users/monabiyan/SkyDrive/Summer 2015/Statistics/HW2")
real_state<-read.table("realstate_CA.csv",header=TRUE,sep=",")
head(real_state)
```

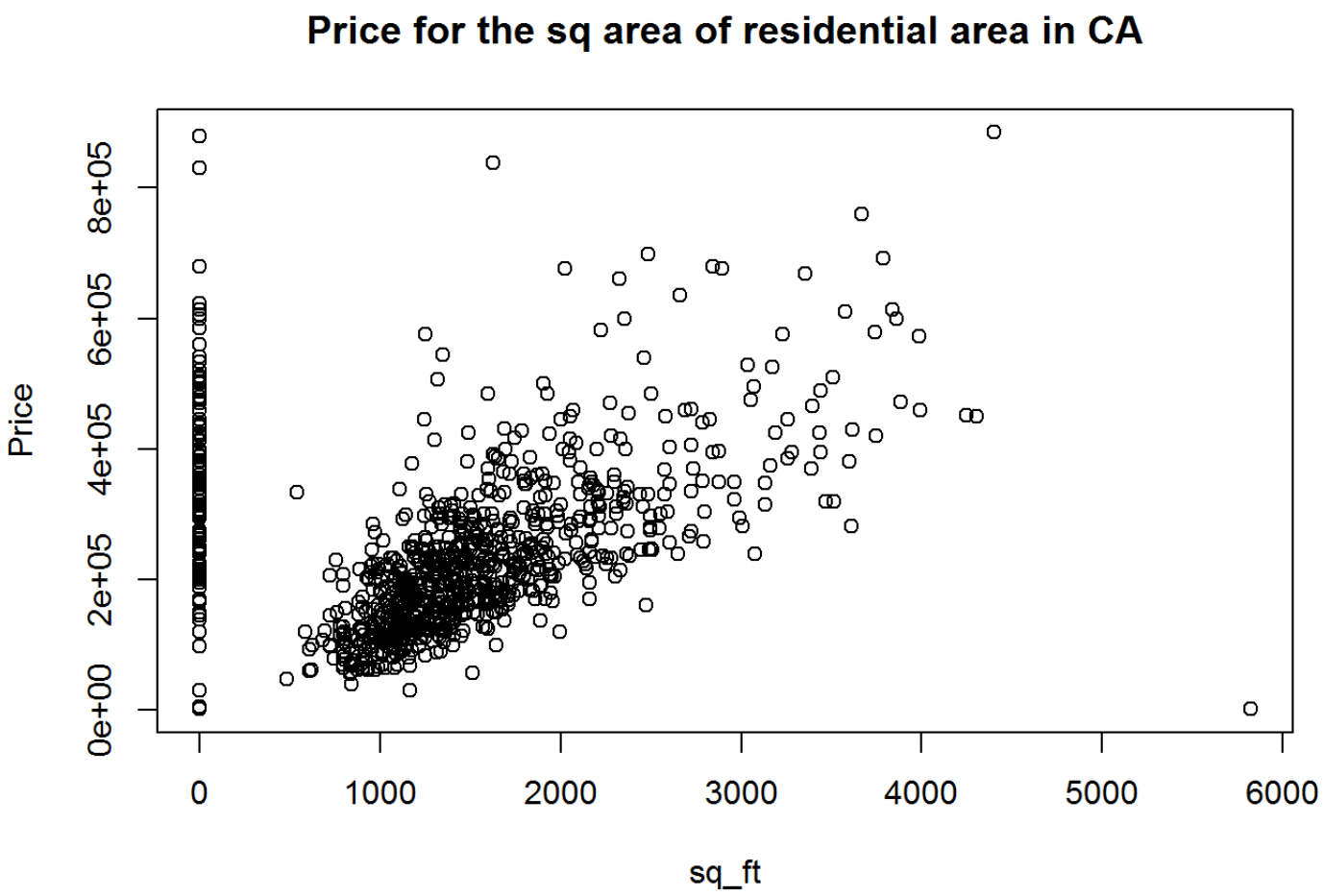
```
##               street       city   zip state beds baths sq__ft        type
## 1       3526 HIGH ST SACRAMENTO 95838    CA    2     1    836 Residential
## 2        51 OMAHA CT SACRAMENTO 95823    CA    3     1   1167 Residential
## 3     2796 BRANCH ST SACRAMENTO 95815    CA    2     1    796 Residential
## 4   2805 JANETTE WAY SACRAMENTO 95815    CA    2     1    852 Residential
## 5    6001 MCMAHON DR SACRAMENTO 95824    CA    2     1    797 Residential
## 6 5828 PEPPERMILL CT SACRAMENTO 95841    CA    3     1   1122        Condo
##                     sale_date price latitude longitude
## 1 Wed May 21 00:00:00 EDT 2008 59222 38.63191 -121.4349
## 2 Wed May 21 00:00:00 EDT 2008 68212 38.47890 -121.4310
## 3 Wed May 21 00:00:00 EDT 2008 68880 38.61830 -121.4438
## 4 Wed May 21 00:00:00 EDT 2008 69307 38.61684 -121.4391
## 5 Wed May 21 00:00:00 EDT 2008 81900 38.51947 -121.4358
## 6 Wed May 21 00:00:00 EDT 2008 89921 38.66260 -121.3278
```

```
hist(real_state$price,main="RealState price in California",xlab="Price",ylab="Count")
```



b. Create a scatter plot using the base R graphics, again with some variable other than the one in the lessons.

```
plot(real_state$sq__ft,real_state$price,xlab="sq_ft",ylab="Price",main="Price for the sq area of resi
dential area in CA")
```

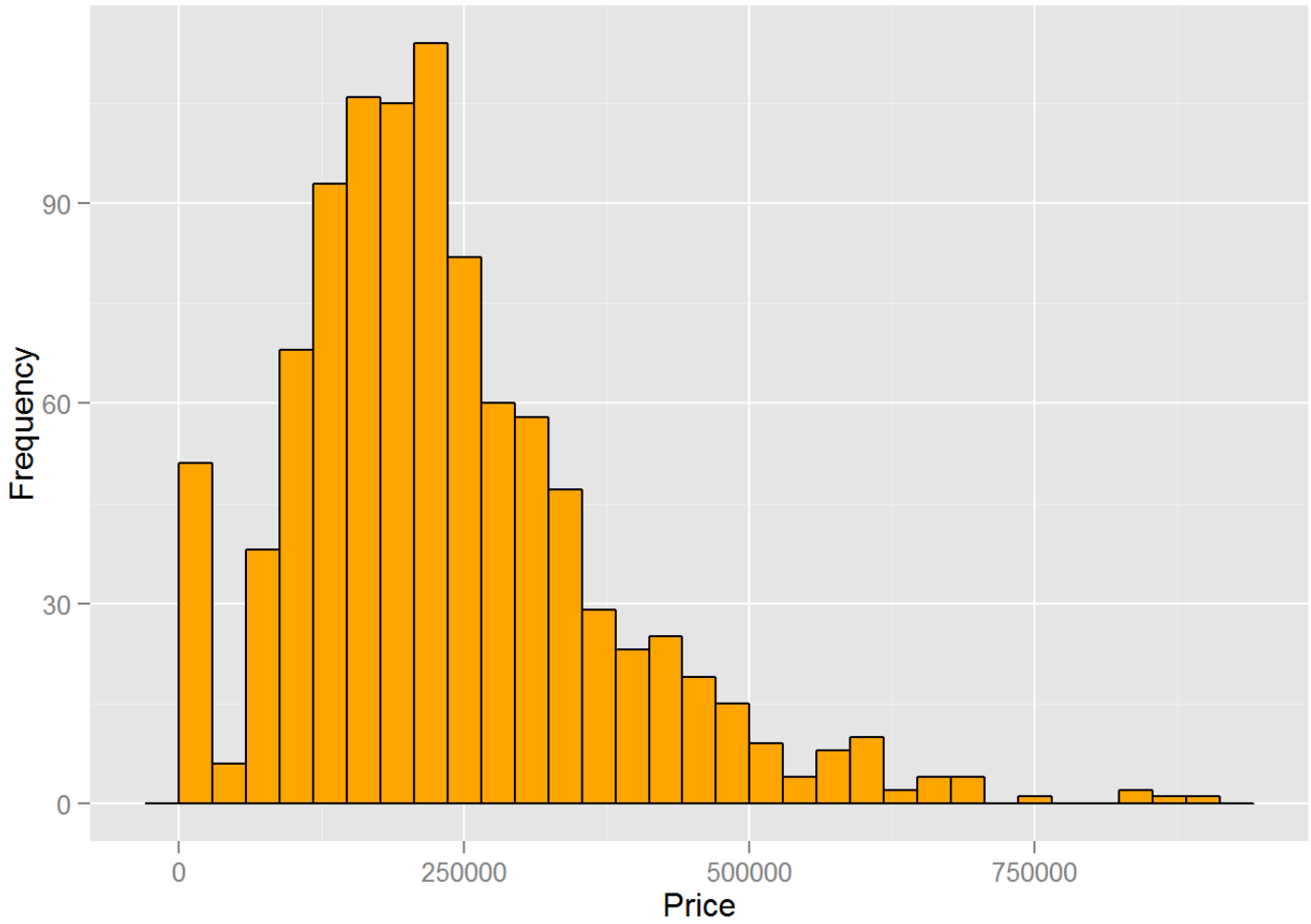


c. Create a histogram using ggplot, using some new data. In this and the later plots, feel free to tinker with the settings using the examples in http://www.cookbook-r.com/Graphs/ (http://www.cookbook-r.com/Graphs/) to make it prettier.

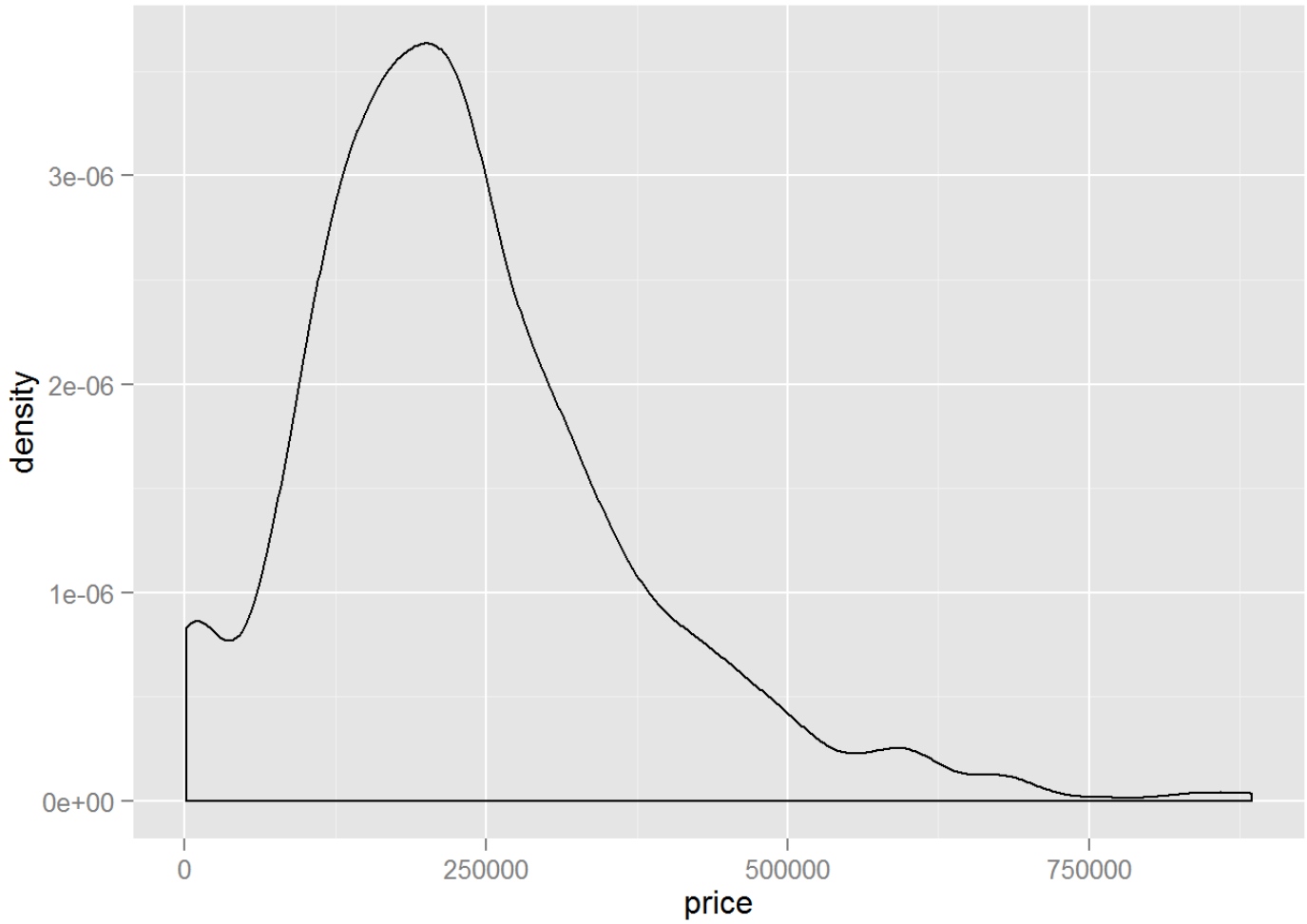
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
ggplot(data=real_state,aes(x=price)) + geom_histogram(fill="orange", colour="black") +xlab("Price") +
ylab("Frequency")
```



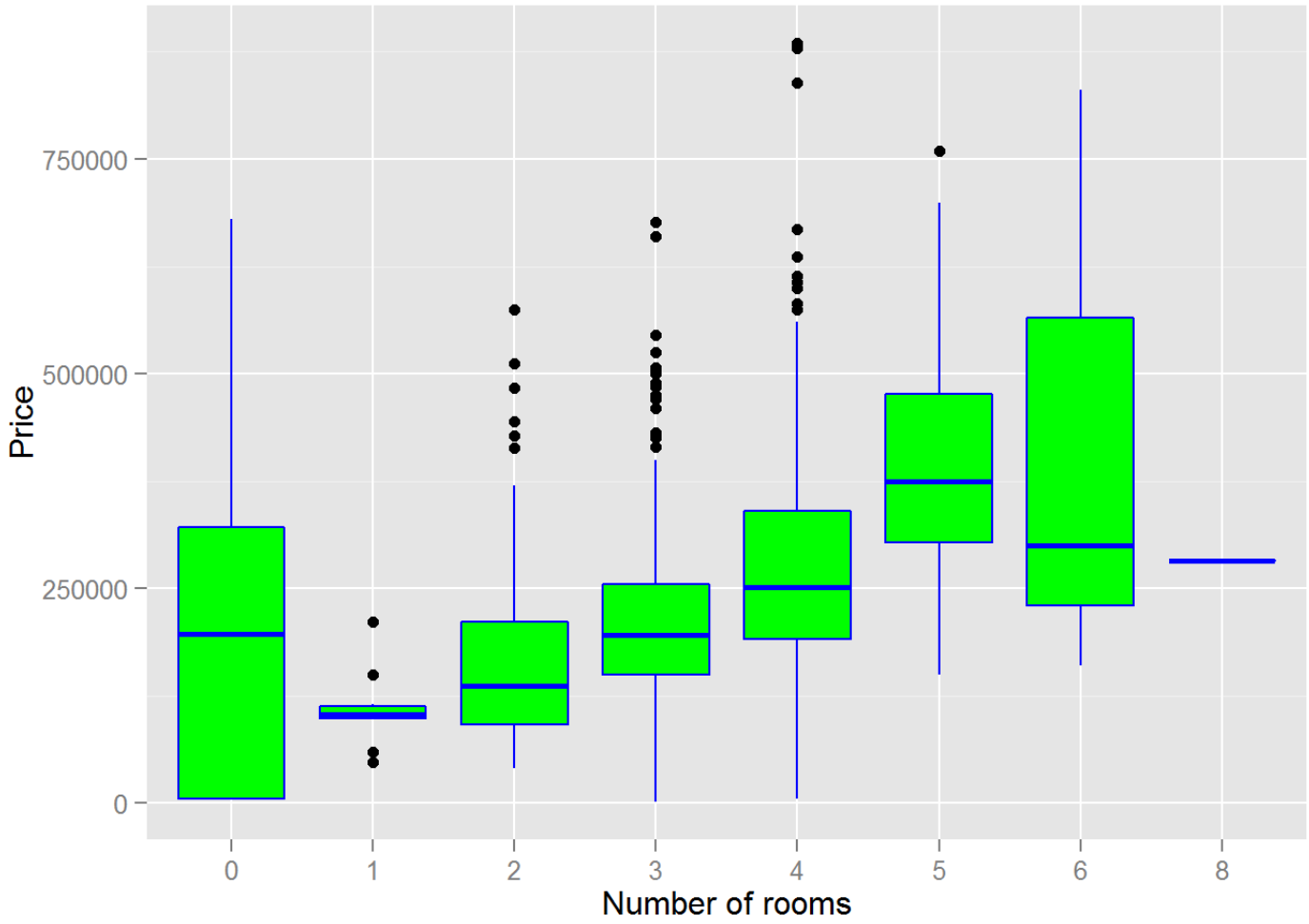

```
ggplot(data=real_state,aes(x=price)) +geom_density()
```




```
ggsave("histogram.jpg",width=6,height=4)
```

d.Create a box plot using ggplot, using some new data.

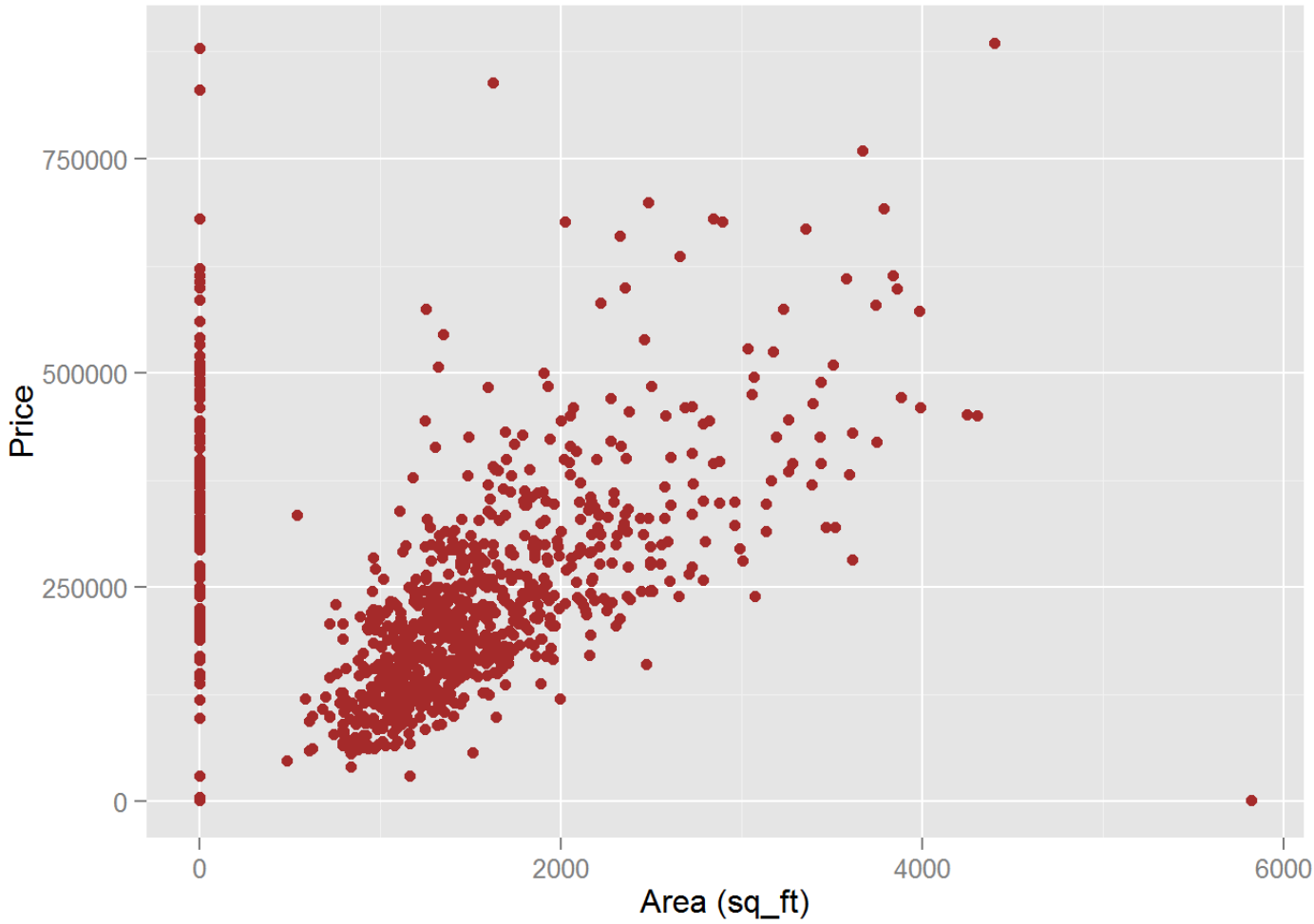
```
ggplot(data=real_state,aes(x=as.factor(beds),y=price)) + geom_boxplot(fill="green", colour="blue") +x
lab("Number of rooms") + ylab("Price")
```




```
ggsave("box_plot.jpg",width=6,height=4)
```

e.Create a scatter plot using ggplot, using some new data.

```
ggplot(data=real_state,aes(x=sq__ft,y=price),fill=cond) + geom_point(colour="brown") +xlab("Area (s
q_ft)") + ylab("Price")
```




```
ggsave("scatter_plot.jpg",width=6,height=4)
```