

Homework 2

Evan Yacek ety78

This homework is due on Feb. 2, 2016 at 11:59pm. Please submit as a PDF file on Canvas.

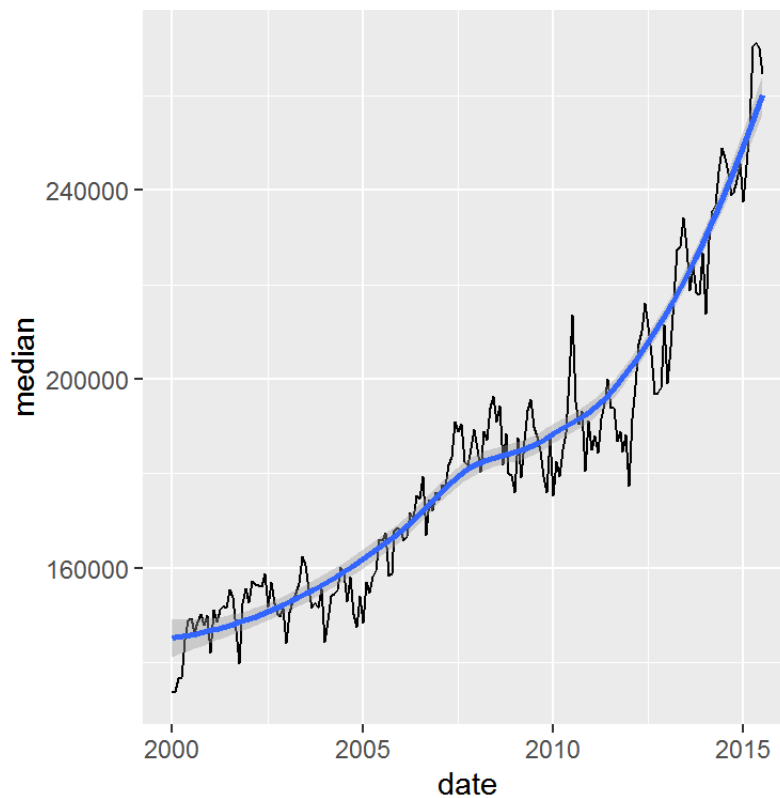
This homework uses Austin housing data from the `txhousing` data set available in R. I have already isolated the Austin housing data and placed it in a data frame called `atxhousing`. Each observation in the data frame contains information from one month of housing sales in Austin. This information includes the year, the month, number of houses sold in that month (`sales`), the total value of houses sold in dollars (`volume`), the median house price in dollars (`median`), the number of listings on the market at the end of the month (`listings`), the amount of time that it would take to sell all of the current listings at the current rate of sales (`inventory`), and the date in years in a precise decimal form (`date`).

```
atxhousing <- txhousing[txhousing$city == 'Austin', ]  
head(atxhousing)
```

##	city	year	month	sales	volume	median	listings	inventory	date
## 562	Austin	2000	1	1025	173053635	133700	3084	2.0	2000.000
## 563	Austin	2000	2	1277	226038438	134000	2989	2.0	2000.083
## 564	Austin	2000	3	1603	298557656	136700	3042	2.0	2000.167
## 565	Austin	2000	4	1556	289197960	136900	3192	2.1	2000.250
## 566	Austin	2000	5	1980	393073774	144700	3617	2.3	2000.333
## 567	Austin	2000	6	1885	368290072	148800	3799	2.4	2000.417

Question 1: (3 pts) Use `ggplot2` to create a line plot of the median housing prices in Austin over time. In the same plot, fit a curve to these data using `geom_smooth()`. In one sentence, what broad trend do you observe in median housing prices over time? **HINT:** Use the `date` column as a measurement of time.

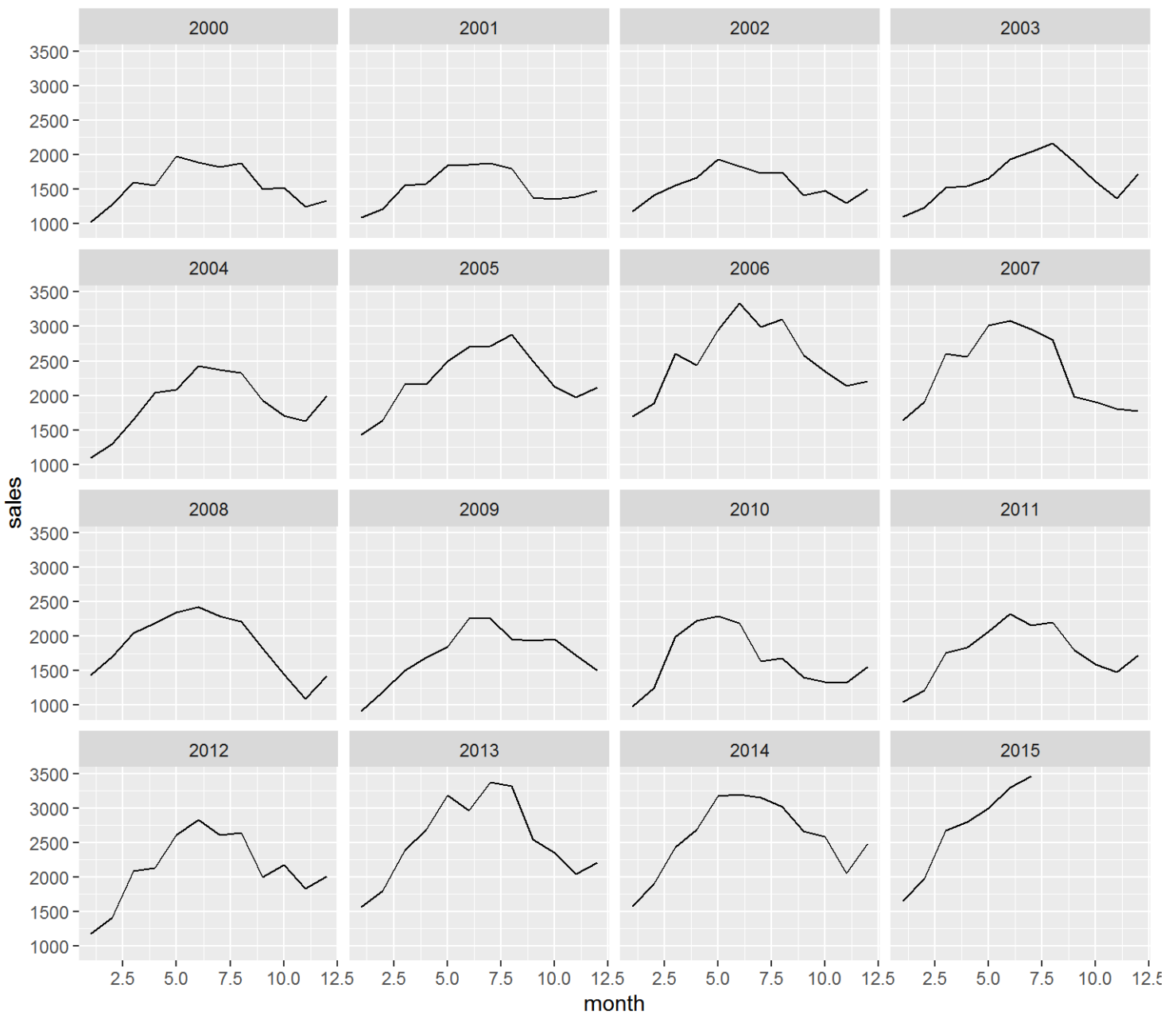
```
ggplot(data=atxhousing, aes(x=date, y=median)) +  
  geom_line()+geom_smooth()
```



The median housing price is increasing.

Question 2: (3 pts) Next, create a line plot of the number of sales per month over time, faceted by year. In general, assuming that Month 1 is January, Month 2 is February, and so on, is the number of homes sold higher in the summer or the winter? State your answer in one sentence.

```
ggplot( data = atxhousing, aes(x=month,y=sales))+geom_line()+facet_wrap(~year)
```



It appears more houses are sold in the summer months.

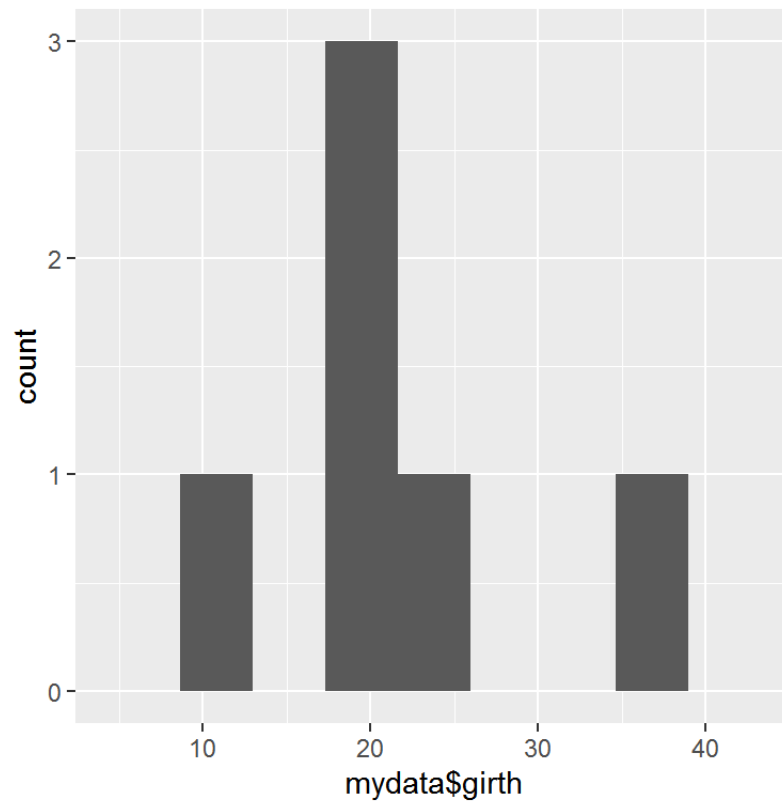
Question 3: (2 pts) Make up a simple data set (5-10 data points) and plot it twice, once using `geom_histogram()` and once using `geom_density()`.

```
mydata <- data.frame(girth = c(12, 20, 18, 38, 19, 22),
                     weight = c(138, 190, 177, 260, 180, 163),
                     size = c('M', 'L', 'M', 'L', 'L', 'M'))

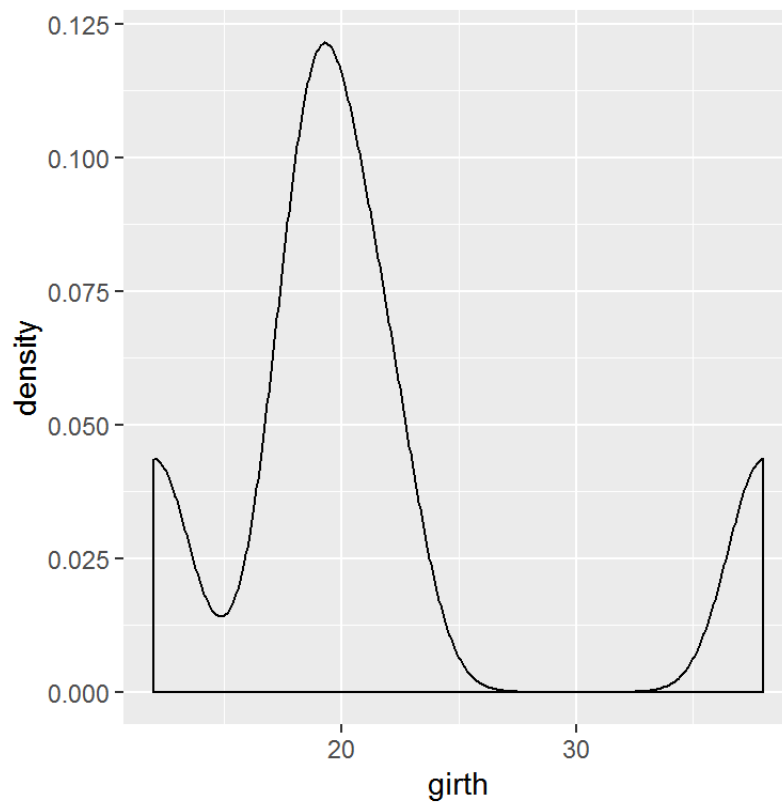
mydata
```

```
##   girth weight size
## 1   12   138    M
## 2   20   190    L
## 3   18   177    M
## 4   38   260    L
## 5   19   180    L
## 6   22   163    M
```

```
ggplot(mydata,aes(mydata$girth))+geom_histogram(bins = 6)
```



```
ggplot(mydata, aes(x=girth)) + geom_density()
```



Question 4: (2 pts) What does the y-axis in your histogram represent? In your density plot, what is the *total* area under the curve? Please give a single number as your answer. **HINT:** You do not need to do any additional calculations to determine the area under the curve.

The y axis in the histogram represented the frequency of individuals in each interval of data or bin. The total area under the curve is equal to 1.