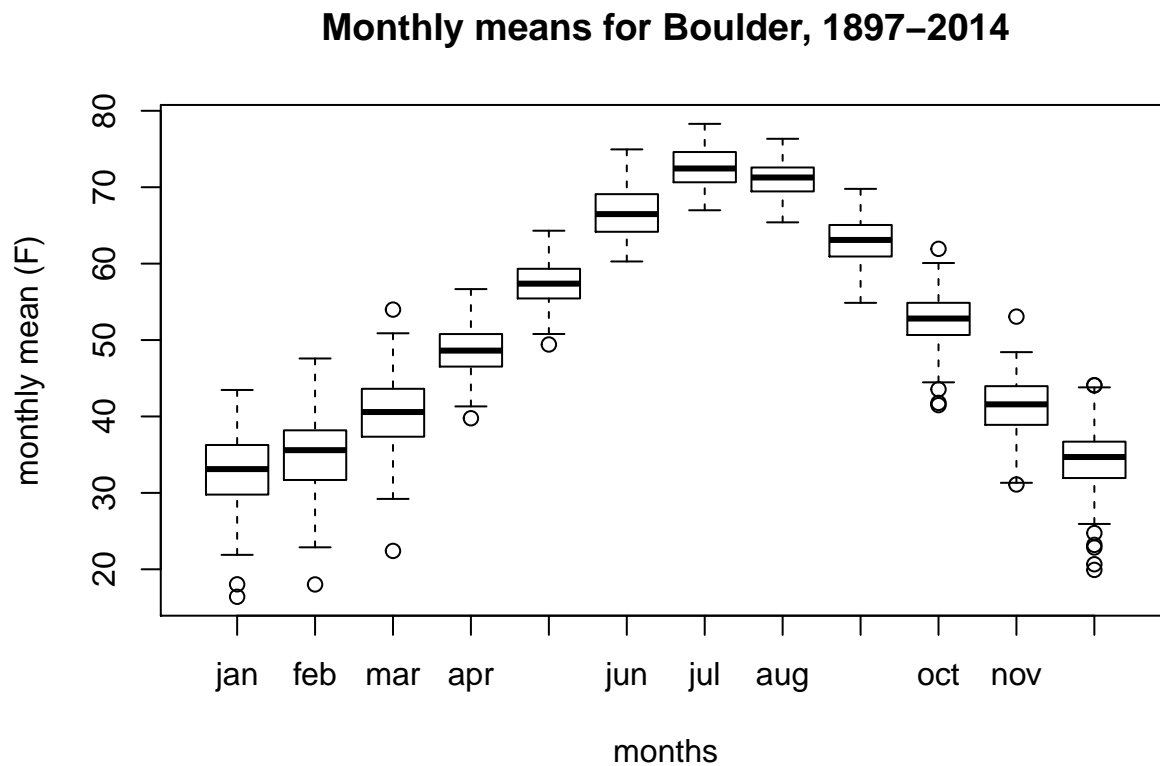


```
library( dataWorkshop)
library(fields)
```

Q1

```
data(BoulderTemperature)
boxplot( BoulderTemperature, xlab="months",
        ylab="monthly mean (F)")
title("Monthly means for Boulder, 1897-2014")
```



Which months have the most similar distributions of temperatures?

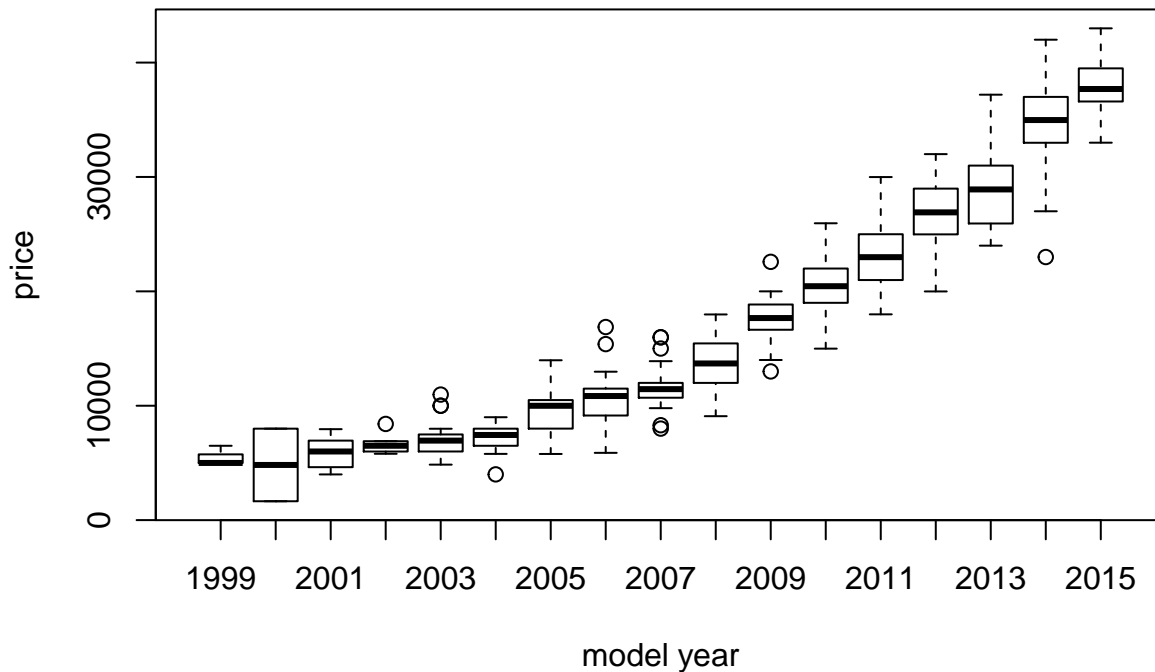
I think January and December look most similar. To shift the plot to put them side by side:

```
boxplot( BoulderTemperature[,c(12,1:11)])
```

Q2

```
data( AudiA4)
boxplot(AudiA4$price ~ AudiA4$year, xlab="model year", ylab="price" )
title("Audi A4 asking prices")
```

## Audi A4 asking prices



Which years seem to have outliers? \*\*

I think the low price in 2014 stands out. The standard boxplot rule for outliers also flags the years 2002:2004, 2006, 2007, 2009.

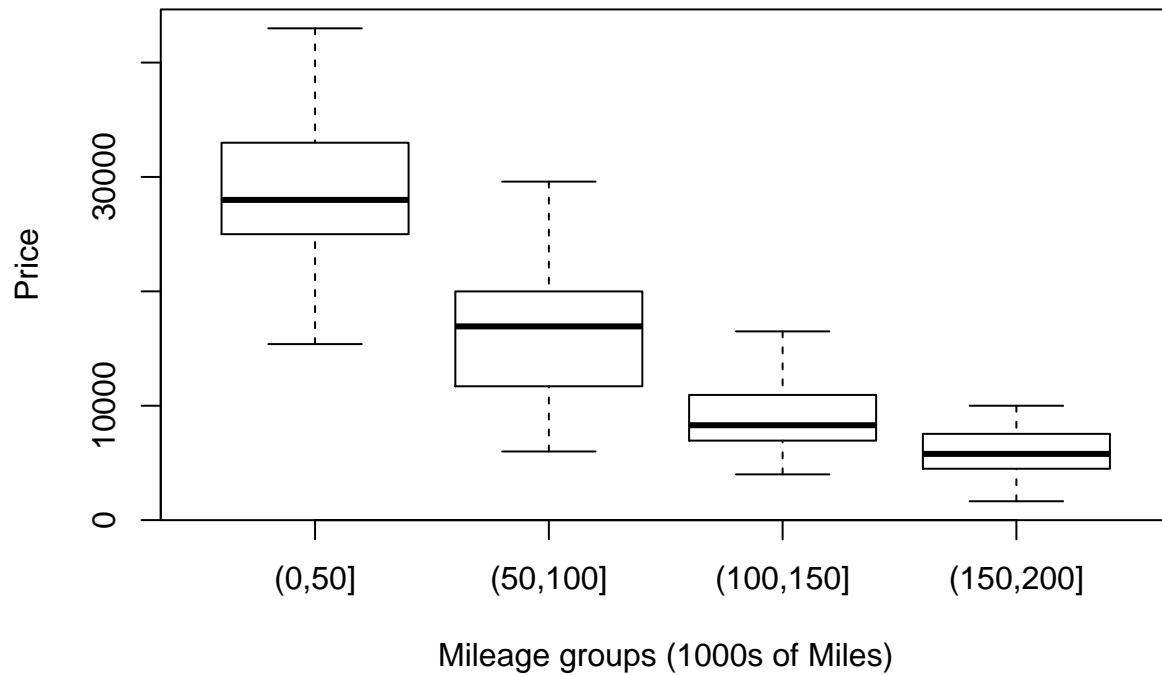
Optional: It may be helpful in judging outliers to look at boxplots that are the percentage difference from the median prices for each year. When I do this 2006 and 2003 seem to have unusual values

```
medianYear<- tapply( AudiA4$price, AudiA4$year, "median")
yearString<- as.character(AudiA4$year)
# need to convert years to characters so that
# they work for the subsetting/selection
percentPrice<- 100*AudiA4$price/medianYear[yearString]
boxplot(percentPrice ~ AudiA4$year, ylab="Percent from median" )
mileGroups<- cut(AudiA4$mileage/1000, breaks= c(0,50,100,150,200) )
```

Q3 Make boxplots of prices for these four groups

```
mileGroups<- cut(AudiA4$mileage/1000, breaks= c(0,50,100,150,200) )
boxplot( AudiA4$price ~ mileGroups,
         xlab="Mileage groups (1000s of Miles)", ylab="Price")
title( "Audi A4 asking prices")
```

## Audi A4 asking prices



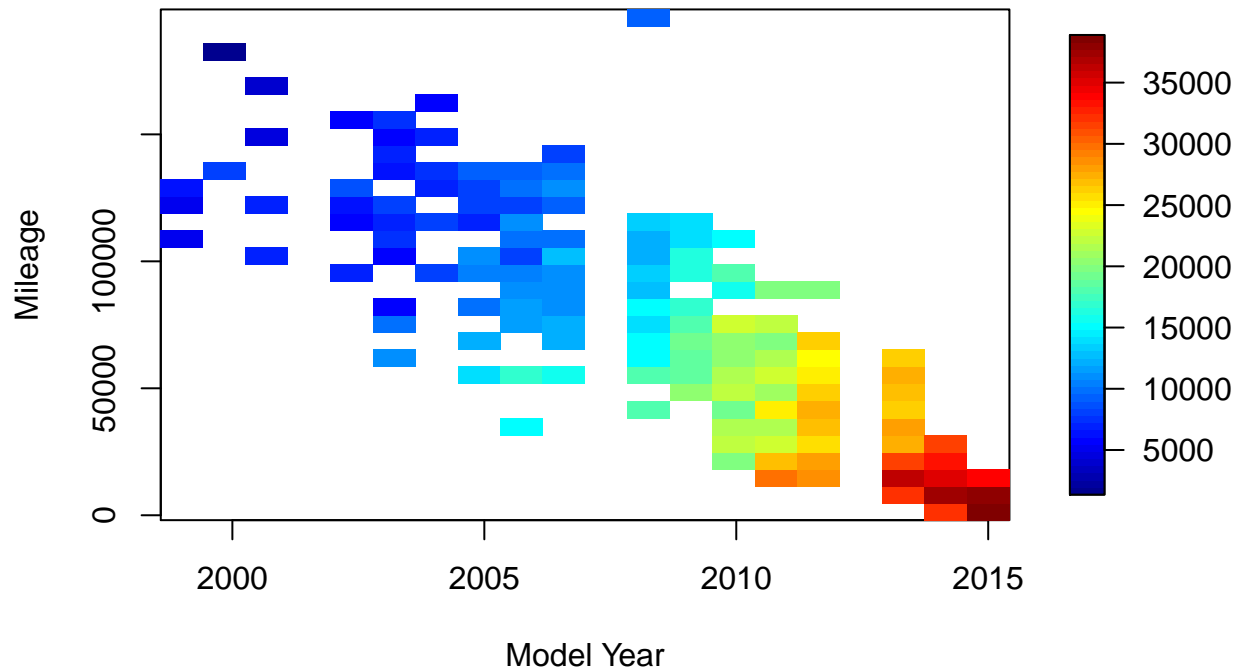
Which mileage groups overlap in their prices?

The only pair of groups that do not overlap are the lowest and highest mileage groups!

Q4

```
X<- cbind( AudiA4$year, AudiA4$mileage)
quilt.plot( X, AudiA4$price, xlab="Model Year", ylab="Mileage", nx=20, ny=30)
title("Audi A4 asking prices")
```

## Audi A4 asking prices



**you know the year of the car does it also help to know the mileage?**

**If**

Yes especially for the 2013 models where the mileage tends to separate the high and low prices. (E.g. lower prices in the 2013 group tend to have higher mileage.) The years 2005 through 2011 also seem to have this pattern to a lesser extent.

I think `nx=20` and `ny=30` gives larger and easier to interpret pixels in the quilt plot. It does make the year spacing a little goofy.