Untangling House Prices

Here's a simple sounding question:

> **How much is a fireplace worth to the value of a home?**

You might want to know the answer if you are buying or selling a house, or if you are a developer looking to build attractive but economical houses.

To support an answer, here's a set of data on houses sold in the Saratoga Springs, New York area in 2002-3. (The data have been provided by Prof. Richard De Veaux of Williams College, which is across the border from this area of New York, in Massachusetts.)

```
houses <- InstructorWorkshop::SaratogaHouses
```

Now it's easy: just compare the mean price for houses with and without fireplaces:

```
mean( Price ~ Fireplace, data=houses )

##        N        Y
## 127954.7 198454.6
```

To judge from the answer, houses with fireplaces are worth about $70,000 more than houses without fireplaces.

Question 1 : Look at the distribution of house price in the Saratoga data. Explain why the median price might be more informative than the mean.

Of course, there are factors other than a fireplace that influence a house price. Examples: the number of bedrooms, the number of bathrooms, the living area, etc.

Question 2 : Look at the median price of the homes broken down by the number of bedrooms. Estimate how much an additional bedroom is worth.

Note: If you make a box-and-windows plot of price versus the number of bedrooms, you'll have to use the `as.factor()` function to turn the quantitative variable into a categorical variable, that is:

```
bwplot( Price ~ as.factor(Bedrooms), data=houses )
```

Question 3 : Look at the median price of homes broken down by the number of bathrooms. How much is an additional bathroom worth? (Note: A "half bathroom" refers to a bathroom without a shower or bath, such as a "powder room". A typical house with 2.5 baths will have two "full baths" and one half bathroom.)

Question 4 : How does the price of a house depend on the house's living area? Note that we do not have a good tool to study median price broken down by area — you'll have to use `lm()`.

Such a simple analysis does not do justice to the determinants of house value. The problem is that the various features — number of bedrooms, number of baths, living area, fireplace — tend to be aligned. To see this, calculate the following:

- the mean or median number of baths broken down by whether there is a fireplace

- the mean or median number of bedrooms broken down by whether there is a fireplace

- the mean or median living area broken down by whether there is a fireplace

- the mean or median living area broken down by the number of baths

Question 5 : What in the results demonstrates that there is an alignment between different house features?

Because of this alignment among the features, looking at the effect of an individual feature using groupwise means or medians can be misleading. For instance, an increase in price that's really due to the number of bathrooms might show up as due to living area.

To untangle the influences of the different features, you can build a model of house price that includes several factors simultaneously. You can build such a model using `lm()`.

Question 6 : Build one or more such models and use them to estimate the influence of a fireplace on the price of a house.

When you have your estimate, search the Internet for information about how builders, remodelers, real-estate agents or appraisers relate a fireplace to price.

Some Nomenclature: In the economics literature, the method of estimating the value of a single feature by building a model that includes several features simultaneously is called "hedonic regression."