# In-class Lab 5

ECON 425 (Justin Heflin, West Virginia University)

February 10, 2023

The purpose of this in-class lab is to further practice your regression skills. The lab may be completed as a group, but each student should turn in their own work. To get credit, upload your .R script to the appropriate place on eCampus (“In-Class Labs’ ’ folder).

## For starters

Open up a new R script (named `ICL5_XYZ.R`, where `XYZ` are your initials)

## Install a new R Package

```
install.packages("car")
```

```
library(car)
```

```
## Loading required package: carData
```

For this lab, let’s use data on house prices. This data is located in the `hprice1` data set in the `wooldridge` package. Each observation is a house.

```
library(wooldridge)
house_prices <- as.data.frame(hprice1)
#View(house_prices)
```

Variable Names:

- price: house price, $1000s
- assess: assessed value, $1000s
- bdrms: number of bedrooms
- lotsize: size of lot in square feet
- sqrft: size of house in square feet
- colonial: dummy variable where =1 if home is colonial style

```
summary(house_prices)
```

## Multiple Regression

Let's estimate the following regression model:

$$price = \beta_0 + \beta_1 lotsize + \beta_2 colonial + \epsilon$$

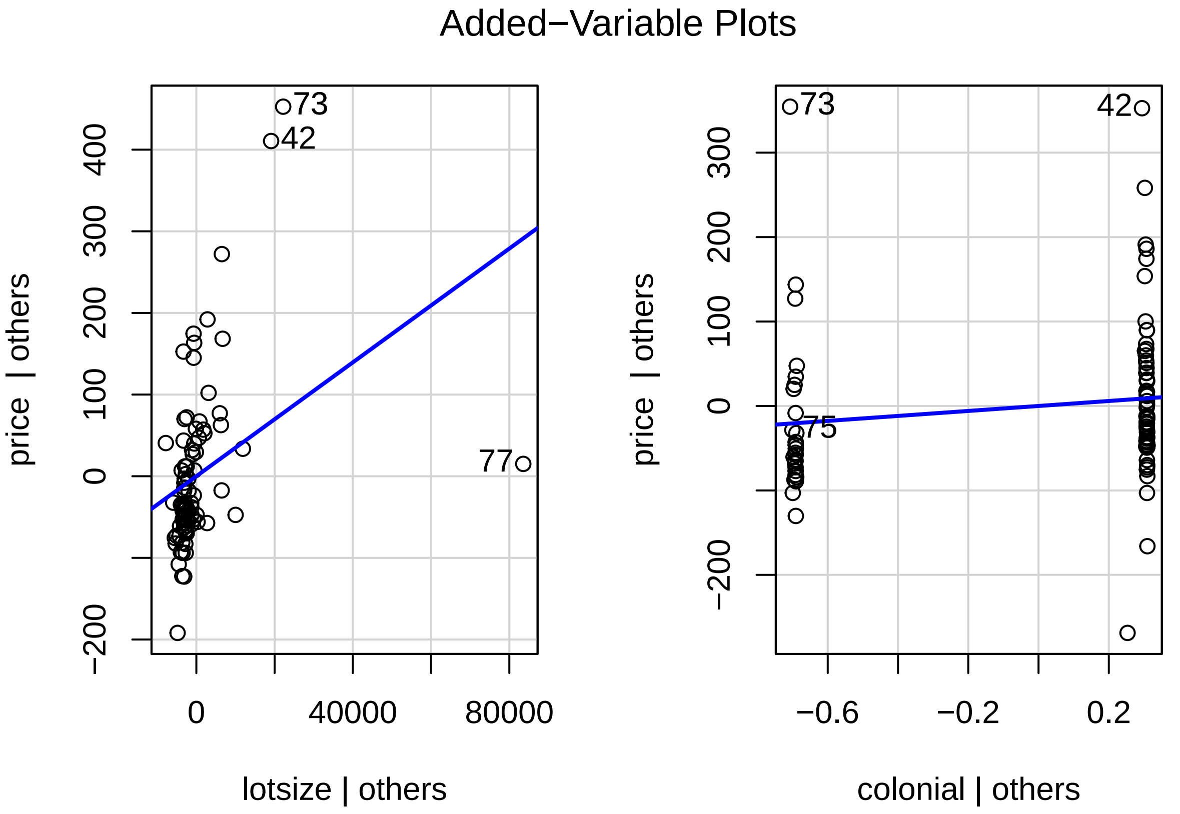
where *price* is the house price in thousands of dollars.

The code to do so:

```
regression_1 <- lm(price ~ lotsize + colonial, data = house_prices)
summary(regression_1)
```

Notice our $R^2$ is relatively small, implying our estimated model does not do a great job of explaining our dependent variable (`price`). We can also see this visually by using the `car` package we installed and loaded earlier.

```
avPlots(regression_1)
```



Here is how to interpret each plot:

- The x-axis displays a single independent variable (`lotsize`, `colonial`) and the y-axis displays the dependent variable (`price`)
- The blue line shows the association between the independent variable and the dependent variable, *while holding the value of all other independent variables constant*

- The points that are labeled in each plot represent the observations with the largest residuals and the observations with the largest partial leverage (implying those observations are outliers that are heavily influencing the fit of the model). In the case of the `lotsize` plot observations 42, 73, and 77 represent the largest residuals.
- Note that the angle of the line in each plot matches the sign of the coefficient from the estimated regression equation.

Let's estimate the following regression model:

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + \epsilon$$

where *price* is the house price in thousands of dollars.

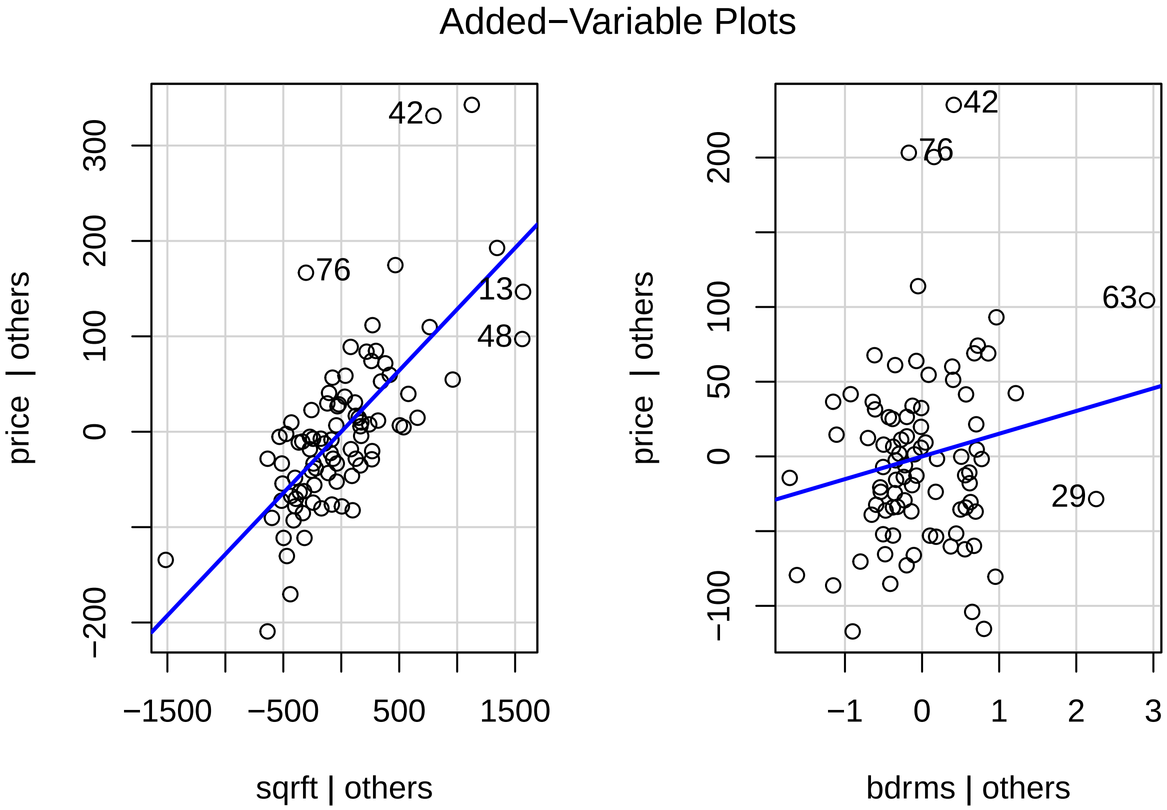
The code to do so:

```
regression_2 <- lm(price ~ sqrft + bdrms, data = house_prices)
summary(regression_2)
```

You should get a coefficient of `0.128` on `sqrft` and `15.2` on `bdrms`. Interpret these coefficients. (You can type the interpretation as a comment in your .R script) Do these numbers seem reasonable?

You should get $R^2 = 0.632$. Based on that number, do you believe this is a good model of house prices?

```
avPlots(regression_2)
```



Now let's add a third independent variable, `assess`, and see if our coefficients change, what happens to our $R^2$ (does it increase, decrease, or remain relatively the same)

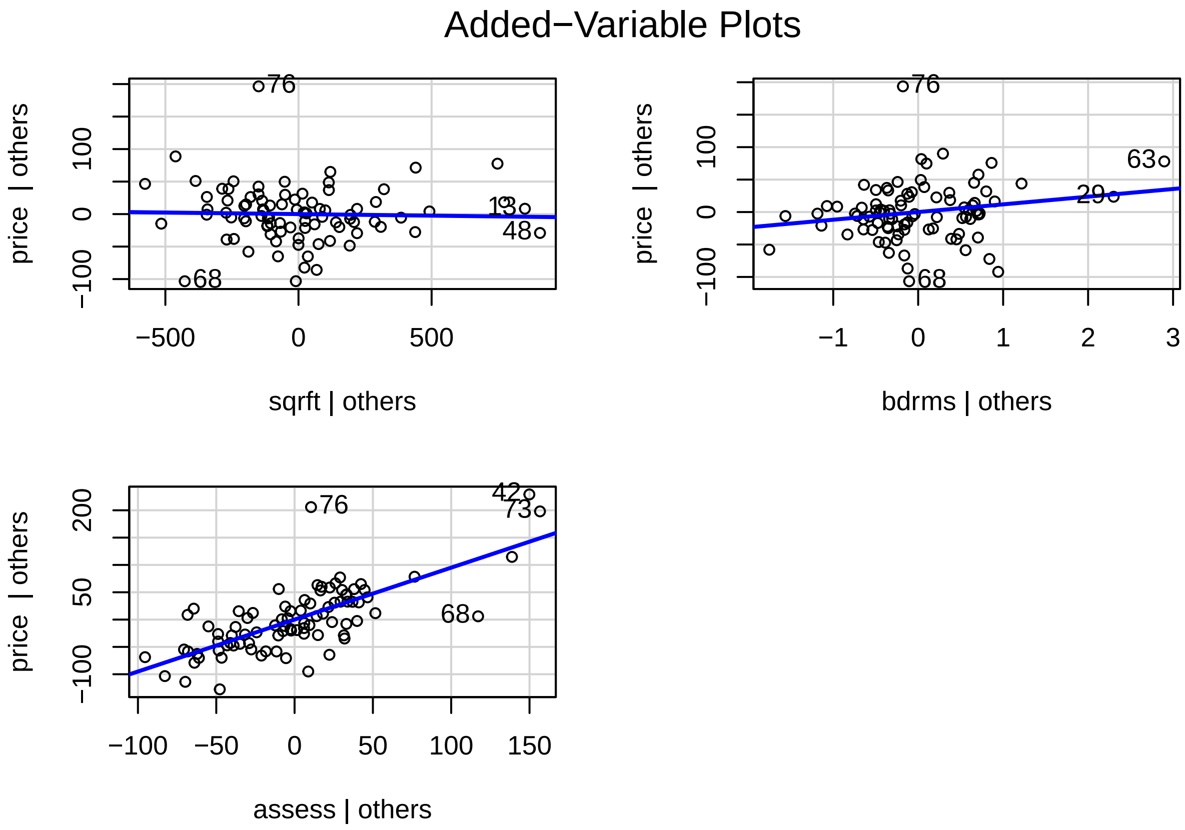
$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + \beta_3 assess + \epsilon$$

```
regression_3 <- lm(price ~ sqrft + bdrms + assess, data = house_prices)
summary(regression_3)
```

Let's first take a look at our estimated regression coefficients. Intuitively, does it make sense that as the size of the house in square feet decreases that price would increase? What about the signs (directions) for the other two independent variables?

Now look at our $R^2$, you should get 0.826, implying this estimated model is a better fit to our sample of data. Again, we can visually see this by looking at the plots of this regression.

```
avPlots(regression_3)
```


Added−Variable Plots
price | others
sqrft | others
−500
0
500
−100
0
100
76
68
13
48
bdrms | others
−1
0
1
2
3
76
68
63
29
assess | others
−100
−50
0
50
100
150
200
76
42
73
68