.

```
house = read.table("http://www.rossmanchance.com/iscam2/data/housing.txt", he
ader = T, sep = "\t")
attach(house)
names(house)

## [1] "sqft"     "price"    "City"     "bedrooms" "baths"
```

##Run Simple Linear Regression- Price being predicted by Sqft

```
summary(lm(price ~ sqft))

##
## Call:
## lm(formula = price ~ sqft)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -439654 -144256  -52040   97373  636508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 65930.31   60993.62   1.081    0.283
## sqft          202.43      26.39   7.670 3.35e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 222100 on 81 degrees of freedom
## Multiple R-squared:  0.4207, Adjusted R-squared:  0.4136
## F-statistic: 58.83 on 1 and 81 DF,  p-value: 3.349e-11
```

##The output above showed that : price being predicted by sqft. very low-p-value meant high statistical significance. This was supported by the fact that R2= 42.07%

Run Simple Linear Regression again, now price being predicted by bedroom

```
summary(lm(price ~ bedrooms))

##
## Call:
## lm(formula = price ~ bedrooms)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -454935 -206553  -76206  190930  798794
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   220612     107208   2.058  0.04283 *
## bedrooms       76865      28802   2.669  0.00919 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 279800 on 81 degrees of freedom
## Multiple R-squared:  0.08082,    Adjusted R-squared:  0.06947
## F-statistic: 7.122 on 1 and 81 DF,  p-value: 0.009195
```

Output above showed statistical significanse again. But this time, p-value for bedroom = 0.009, R2= 8.08% b

Run Linear regression –Price being predicted by BOTH SQFT and BEDROOMS

```
summary(lm (price ~ sqft + bedrooms))
```

```
##
## Call:
## lm(formula = price ~ sqft + bedrooms)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470394 -127929  -68968   87079  605609
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 150678.0    85160.0   1.769   0.0806 .
## sqft           229.4       32.4   7.080 4.98e-10 ***
## bedrooms    -39767.3    28067.3  -1.417   0.1604
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 220800 on 80 degrees of freedom
## Multiple R-squared:  0.4349, Adjusted R-squared:  0.4208
## F-statistic: 30.78 on 2 and 80 DF,  p-value: 1.217e-10
```

An interesting thing happens above. This time the output showed individual t-test(s)/p-value(s) significant for SQFT while being statistically insignificant for Bedrooms. Thus the final model can only contain the following command.

```
summary(lm(price~sqft))
```

```
##
## Call:
## lm(formula = price ~ sqft)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -439654 -144256  -52040   97373  636508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 65930.31   60993.62   1.081    0.283
```

```
## sqft          202.43      26.39   7.670 3.35e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 222100 on 81 degrees of freedom
## Multiple R-squared:  0.4207, Adjusted R-squared:  0.4136
## F-statistic: 58.83 on 1 and 81 DF,  p-value: 3.349e-11
```