Lab 2 Exercise 1

Line 11 - Displays a view of the csv file

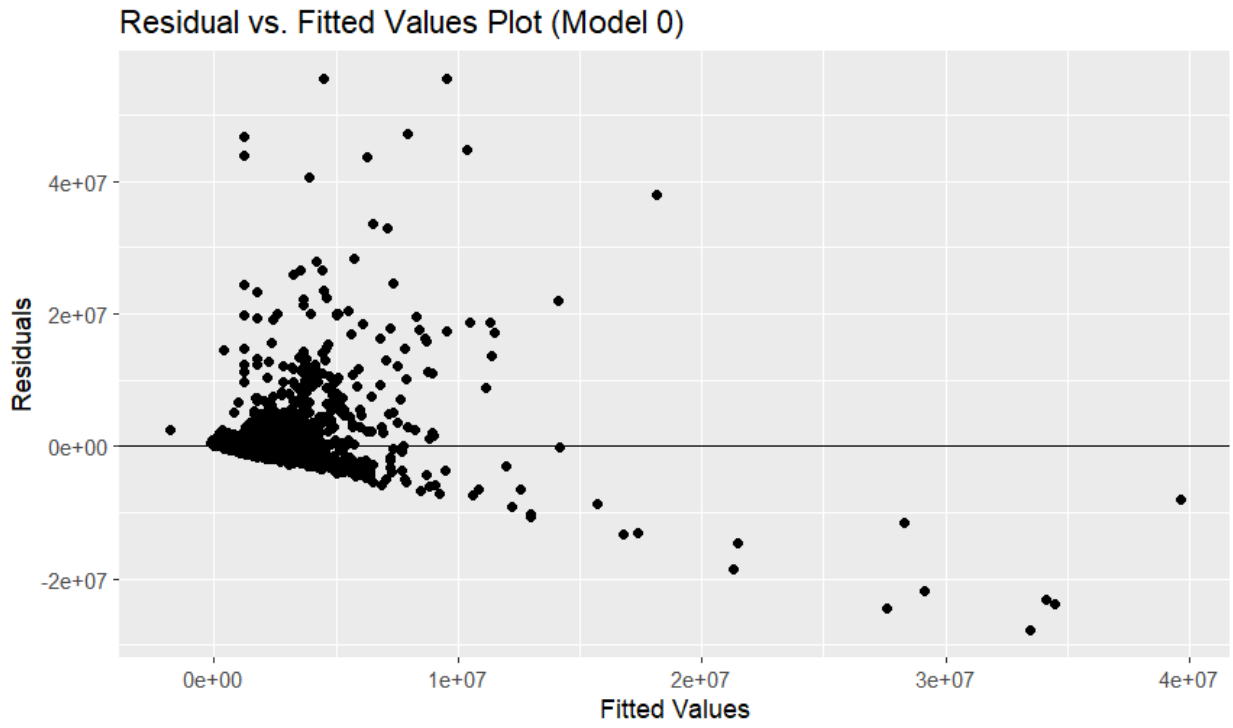
	BROKERTITLE	TYPE	PRICE	BEDS	BATH	PROPERTYSQFT	ADD
1	Brokered by Douglas Elliman -111 Fifth Ave	Condo for sale	315000	2	2.000000	1400.000	2 E 5
2	Brokered by Serhant	Condo for sale	195000000	7	10.000000	17545.000	Cen
3	Brokered by Sowae Corp	House for sale	260000	4	2.000000	2015.000	620
4	Brokered by COMPASS	Condo for sale	69000	3	1.000000	445.000	2 E 5
5	Brokered by Sotheby's International Realty - East Side Manh...	Townhouse for sale	55000000	7	2.373861	14175.000	5 E 6
6	Brokered by Sowae Corp	House for sale	690000	5	2.000000	4004.000	584
7	Brokered by Douglas Elliman - 575 Madison Ave	Condo for sale	899500	2	2.000000	2184.208	157
8	Brokered by Connie Profaci Realty	House for sale	16800000	8	16.000000	33000.000	177
9	Brokered by Pantiga Group Inc.	Co-op for sale	265000	1	1.000000	750.000	875
10	Brokered by CENTURY 21 MK Realty	Co-op for sale	440000	2	1.000000	978.000	135
11	Brokered by Engel & Volkers Americas	Co-op for sale	375000	2	1.000000	850.000	800
12	Brokered by Re/Max Edge	Townhouse for sale	689000	3	2.373861	1162.000	456
13	Brokered by COMPASS	Co-op for sale	259000	3	1.000000	2184.208	34-4
14	Brokered by Jamie & Connie Real Estate Grp	Co-op for sale	430000	2	2.000000	2184.208	91-1
15	Brokered by Corcoran Chelsea/Flatiron	Co-op for sale	895000	3	1.000000	2184.208	61 J

Line 39 - Displays a summary of variables in the first model

```
> ### Model 0: All predictors
> lin.mod0 <- lm(Price ~ PropertySqFt + Beds + Bath, house.df)
> summary(lin.mod0)

Call:
lm(formula = Price ~ PropertySqFt + Beds + Bath, data = house.df)

Residuals:
      Min        1Q    Median        3Q       Max
-27661993   -980490   -513593     -3245  55493406

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -173010.03   87670.38  -1.973   0.0485 *
PropertySqFt     597.96      24.85  24.064  < 2e-16 ***
Beds         -248685.43   31371.13  -7.927 2.77e-15 ***
Bath          670376.79   43172.38  15.528  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3571000 on 4795 degrees of freedom
Multiple R-squared:  0.229,     Adjusted R-squared:  0.2285
F-statistic: 474.8 on 3 and 4795 DF,  p-value: < 2.2e-16
```

Line 42 - 44: Create Scatter plot for the first model with best fit line

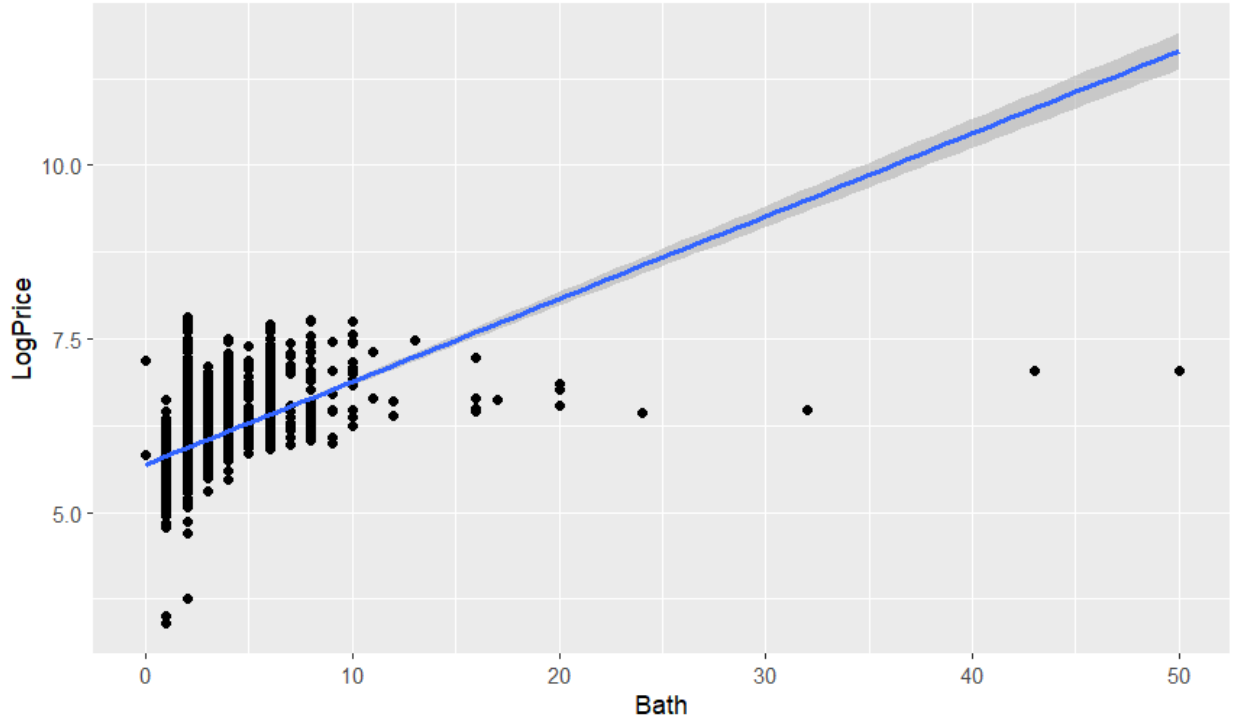
Line 46 - 51: Create residual plot for the first model



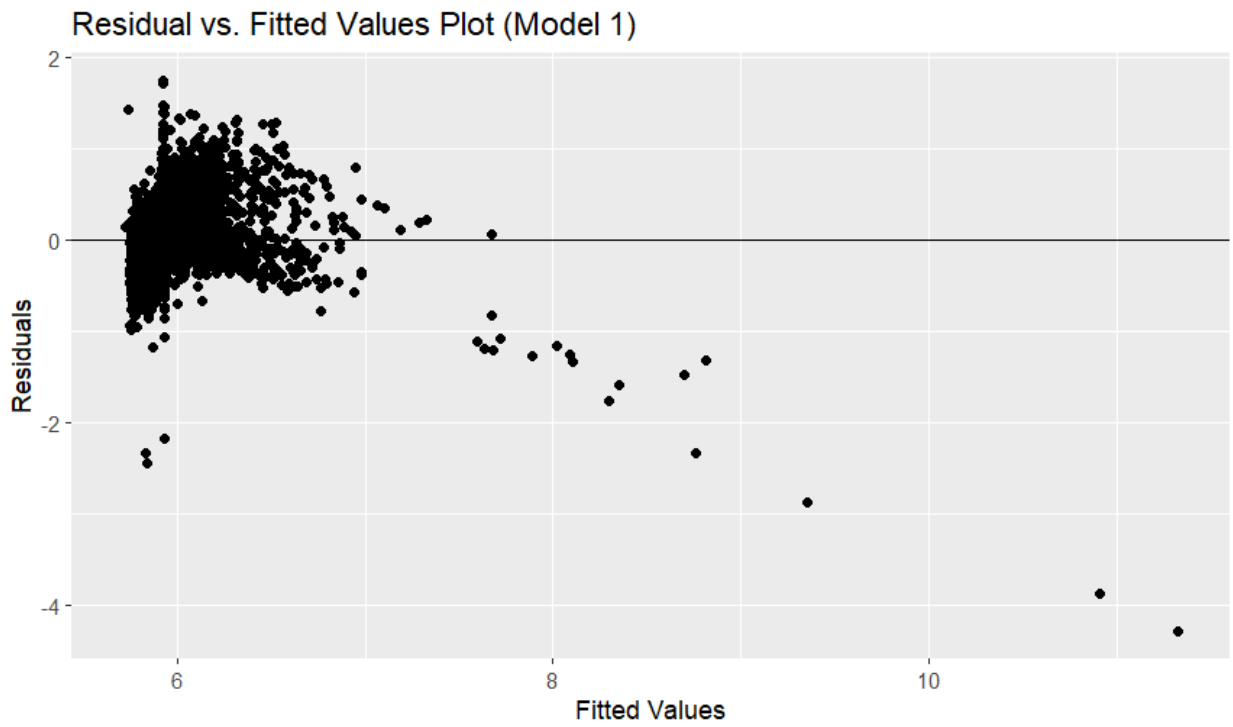
Line 58 - Displays a summary of variables in the second model

```
> ### Model 1: Log-transformed Price
> house.df$LogPrice <- log10(house.df$Price)
> lin.mod1 <- lm(LogPrice ~ PropertySqFt + Beds + Bath, house.df)
> summary(lin.mod1)

Call:
lm(formula = LogPrice ~ PropertySqFt + Beds + Bath, data = house.df)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2819 -0.2033 -0.0327  0.1691  1.7573

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.656e+00  8.988e-03 629.267   <2e-16 ***
PropertySqFt  4.546e-05  2.547e-06  17.846   <2e-16 ***
Beds         -5.952e-03  3.216e-03  -1.851   0.0643 .
Bath          9.927e-02  4.426e-03  22.428   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3661 on 4795 degrees of freedom
Multiple R-squared:  0.3187,    Adjusted R-squared:  0.3183
F-statistic: 747.7 on 3 and 4795 DF,  p-value: < 2.2e-16
```

Line 61 - 63: Create Scatter plot for the second model with best fit line



Line 66 - 70: Create residual plot for the second model



Line 75 - Displays a summary of variables in the third model

```
> ### Model 2: Reduced predictors
> lin.mod2 <- lm(Price ~ PropertySqFt + Bath, house.df)
> summary(lin.mod2)

Call:
lm(formula = Price ~ PropertySqFt + Bath, data = house.df)

Residuals:
      Min        1Q    Median        3Q       Max
-26618097   -998904   -543599    -29804  55747307

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -387730.27   83917.00   -4.62 3.93e-06 ***
PropertySqFt     578.36      24.88   23.24  < 2e-16 ***
Bath          424643.85   30242.00   14.04  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3594000 on 4796 degrees of freedom
Multiple R-squared:  0.2189,    Adjusted R-squared:  0.2186
F-statistic: 672.1 on 2 and 4796 DF,  p-value: < 2.2e-16
```

Line 78 - 80: Create Scatter plot for the third model with best fit line

Line 83 - 87: Create residual plot for the third model