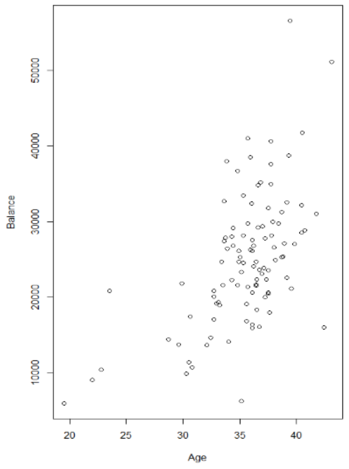
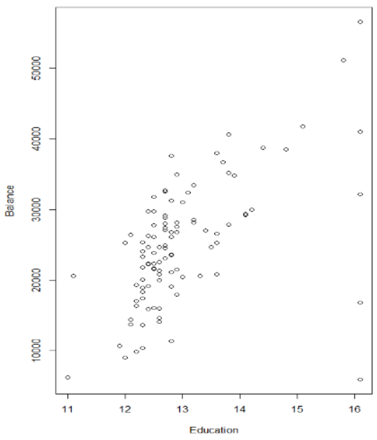
“Banking” Model Creation and Evaluation

Scatter Plots plot(dependent variable(Y)~independent variable (X), data = BANKING)

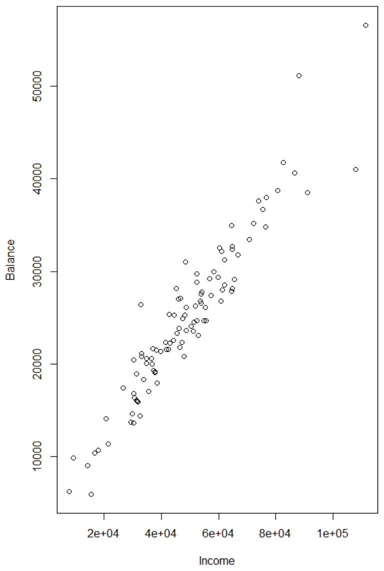
Bank Balance and Age



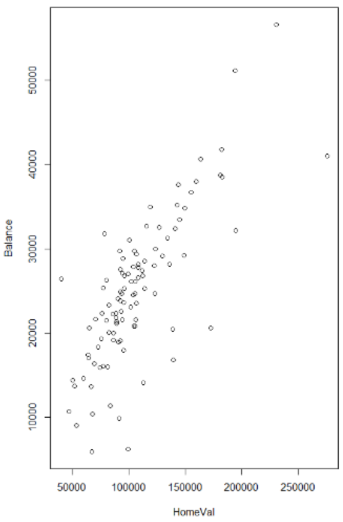
Bank Balance and Education



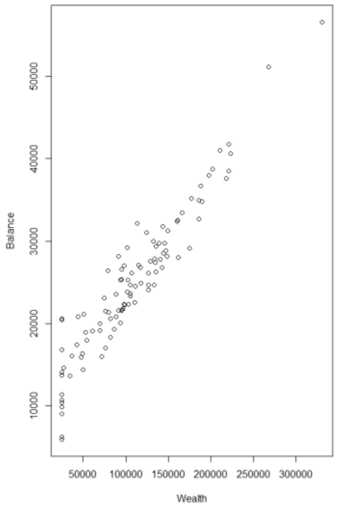
Bank Balance and Income



Banks Balance and HomeVal



Bank Balance and Wealth



Scatter Plot Evaluations

i.) Bank Balance and Age

The form that takes place is somewhat linear, with a moderate strength, and has a positive direction. There are outliers at both ends that weaken the strength.

ii.) Bank Balance and Education

The form that takes place is somewhat linear, with a moderate strength, and has a positive direction. There are outliers at both ends, but a higher amount near the end of the independent variable.

iii.) Bank Balance and Income

The form that takes place is linear, with a high amount of strength, and has a positive direction. There are few outliers.

iv.) Banks Balance and HomeVal

The form that takes place is linear, with a moderately high amount of strength, and has a positive direction. There are outliers that take away from the strength of the model.

v.) Bank Balance and Wealth

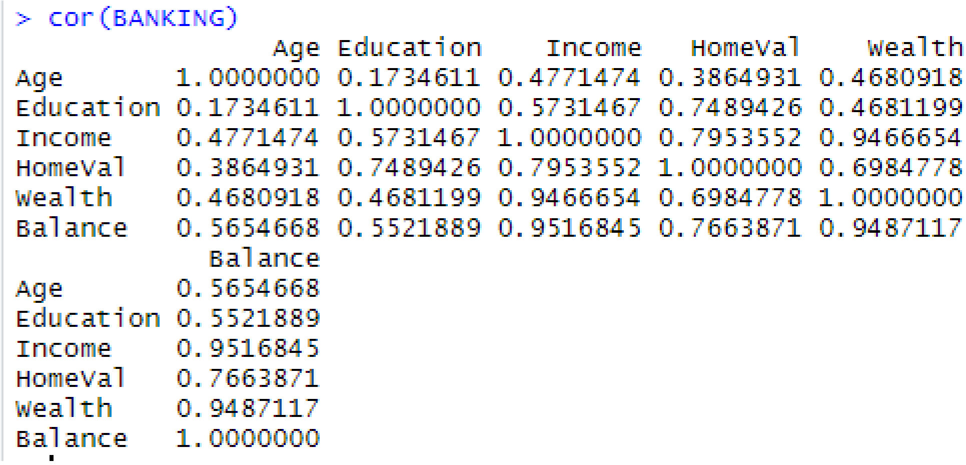
The form that takes place is linear, with a high amount of strength, and has a positive direction. There are few outliers.

Correlation Evaluation

cor(BANKING)

We can refer to the last section in the below model that refers to the balance.

Model 2



```
> cor(BANKING)
                Age Education    Income   HomeVal    Wealth
Age       1.0000000 0.1734611 0.4771474 0.3864931 0.4680918
Education 0.1734611 1.0000000 0.5731467 0.7489426 0.4681199
Income    0.4771474 0.5731467 1.0000000 0.7953552 0.9466654
HomeVal   0.3864931 0.7489426 0.7953552 1.0000000 0.6984778
Wealth    0.4680918 0.4681199 0.9466654 0.6984778 1.0000000
Balance   0.5654668 0.5521889 0.9516845 0.7663871 0.9487117
            Balance
Age       0.5654668
Education 0.5521889
Income    0.9516845
HomeVal   0.7663871
Wealth    0.9487117
Balance   1.0000000
```

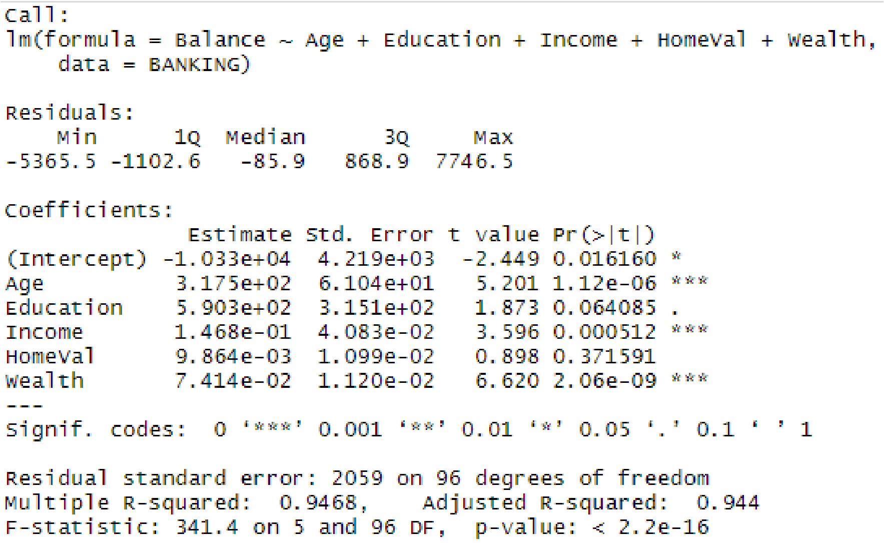
When interpreting correlation, it is important to know what makes a strong correlation, and what makes a weak correlation. A correlation of 1 means they correlate perfectly, while a correlation of - 1 means the opposite. In our cor(BANKING) formula**, we will want to focus on the correlation with Balance**. This the last segment shown in our correlation since it is the last column in our spreadsheet. As we can see in model 2, income (.952) and wealth (.948) correlate the strongest with bank balance. This logically makes sense because the higher salary as well as your wealth may be, the higher bank account balance should be. All of the independent variables have a positive correlation with bank account balance.

Looking at our Current Model

mdl <- lm(Balance~Age+Education+Income+HomeVal+Wealth, data = BANKING)

summary(mdl)

Model 3



```
Call:
lm(formula = Balance ~ Age + Education + Income + HomeVal + Wealth,
    data = BANKING)

Residuals:
    Min      1Q  Median      3Q     Max 
-5365.5 -1102.6   -85.9   868.9  7746.5 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.033e+04  4.219e+03  -2.449 0.016160 *  
Age          3.175e+02  6.104e+01   5.201 1.12e-06 ***
Education    5.903e+02  3.151e+02   1.873 0.064085 .  
Income       1.468e-01  4.083e-02   3.596 0.000512 ***
HomeVal      9.864e-03  1.099e-02   0.898 0.371591    
Wealth       7.414e-02  1.120e-02   6.620 2.06e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2059 on 96 degrees of freedom
Multiple R-squared:  0.9468,	Adjusted R-squared:  0.944 
F-statistic: 341.4 on 5 and 96 DF,  p-value: < 2.2e-16
```

In Model 3, we can interpret the f test (p-value: < 2.2e-16) tells us that at least one of the betas is not 0. This means that we can reject the null hypothesis, and accept the alternative. The adjusted R- Squared tells us that the 94.4% of variability of our dependent variable is explained by our regression model. In other words – the variability in age, education, income, homeval, and wealth are all explained by predicting the bank balance. HomeVal and Education have a p-value >.05(do not pass the T-test) , and should be considered to be taken out of the model.

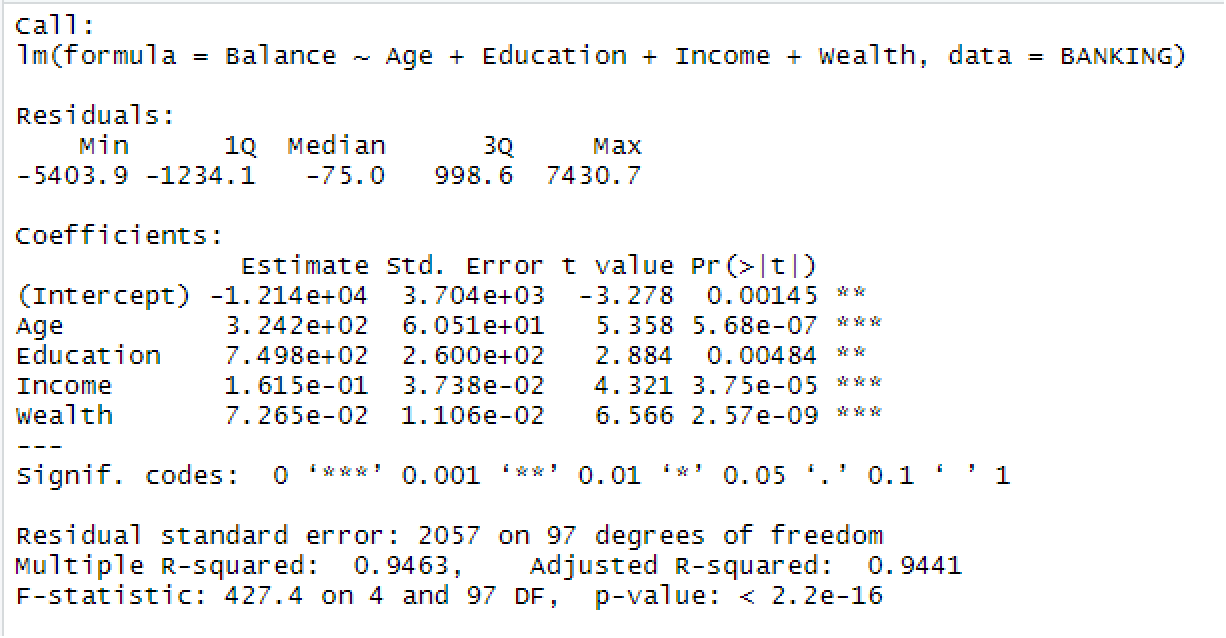
Age, Income, and Wealth all of have a significant impact on Balance. According to Model 3, all of these predictors have p-value that is <.05, which means with those variables we can accept the alternative hypothesis that the beta is not equal to zero, and use their estimates. They also have high t-Values which means that there is greater evidence against the null hypothesis.

Adjusting the Model

mdl2 <- lm(Balance~Age+Education+Income+Wealth, data = BANKING)

summary(mdl2)

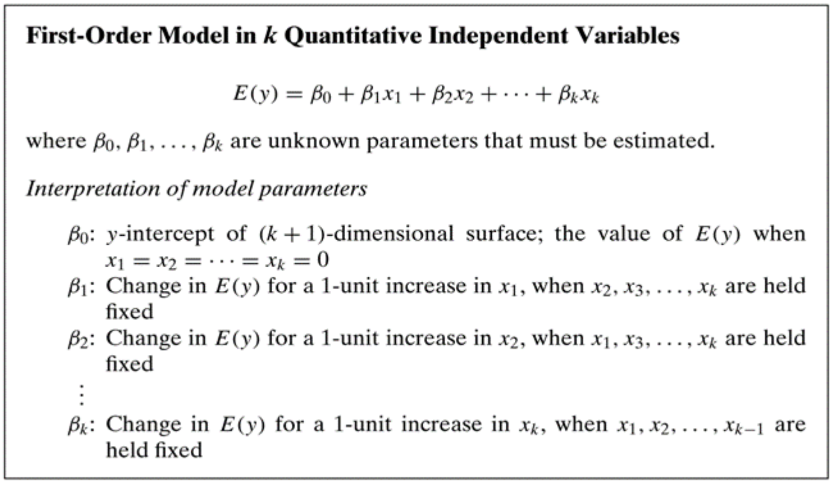
model 4



```
Call:
lm(formula = Balance ~ Age + Education + Income + Wealth, data = BANKING)

Residuals:
    Min      1Q  Median      3Q     Max 
-5403.9 -1234.1   -75.0   998.6  7430.7 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.214e+04  3.704e+03  -3.278  0.00145 ** 
Age          3.242e+02  6.051e+01   5.358 5.68e-07 ***
Education    7.498e+02  2.600e+02   2.884  0.00484 ** 
Income       1.615e-01  3.738e-02   4.321 3.75e-05 ***
Wealth       7.265e-02  1.106e-02   6.566 2.57e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2057 on 97 degrees of freedom
Multiple R-squared:  0.9463,	Adjusted R-squared:  0.9441 
F-statistic: 427.4 on 4 and 97 DF,  p-value: < 2.2e-16
```

With the removal of HomeVal, education no longer needs to be removed, since its p-value has decreased to <.05. All remaining predictors are significant.

Model 5



First-Order Model in *k* Quantitative Independent Variables

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters that must be estimated.

Interpretation of model parameters

β_0: y-intercept of $(k+1)$-dimensional surface; the value of $E(y)$ when $x_1 = x_2 = \cdots = x_k = 0$

β_1: Change in $E(y)$ for a 1-unit increase in x_1, when $x_2, x_3, \ldots, x_k$ are held fixed

β_2: Change in $E(y)$ for a 1-unit increase in x_2, when $x_1, x_3, \ldots, x_k$ are held fixed

$\vdots$

β_k: Change in $E(y)$ for a 1-unit increase in x_k, when $x_1, x_2, \ldots, x_{k-1}$ are held fixed

Answer based off of Model 5

Beta 0 = Y-Intercept = E(y) when x1 equal to x2 equal to etc… = Balance

Beta 1 = Age = Change in E(y) for a 1 unit increase in x1, when x2, x3, and x4 held

Beta 2 = Education = Change in E(y) for a 1 unit increase in x2, when x1, x3, and x4 held constant

Beta 3 = Income = Change in E(y) for a 1 unit increase in x3, when x1, x2, and x4 held constant.

Beta 4 = Wealth = Change in E(y) for a 1 unit increase in x4, when x1, x2, and x3 held constant

The adjusted R-squared in model 4 is .9441. This means that 94.41% of variation in our dependent variable is explained by our independent variables.