MTH 325, Exam #1, Spring 2023 Name KEY

Instructions: Answer each question thoroughly. For questions in Part 1, use the work you did at home to answer the questions. Be sure to answer each part of each question. In Part 2, report exact answers unless directed to round.

Part I:
Use the work you did at home to answer these questions about tax paid and the neighborhoods in our dataset.

1. Based on your correlation table, identify the variable that has the highest correlation with Annual Salary. What is the correlation value?

Total experience

0.8599

2. Based on your correlation table (or graphs), which variables (other than Annual Salary) appear to have potential collinearity problems?

prior and beta experience are highly correlated w/ total experience

3. What is the simple linear regression equation your found relating Age to Annual Salary?

$$\underset{y}{\text{Annual Salary}} = 847.4 \underset{x}{\text{Age}} + 37{,}745.7$$

4. Interpret the slope in the context of the problem.

for every year old, employees can expect on average to earn an additional $847.40

5. What percent of the variability in Annual Salary can be explained by the relationship with Age?

0.09402

$\approx$ 9%

6. Compare your machine found model with your final backwards selection model. Describe any differences in your models (variables included), any errors generated in the selection, etc.

best subset generated an error because of collinearity. Similar forward backward selection ended w/ gender, prior + beta experience + education w/ 0 intercept.

Salary = 7689.1 Education + 2599.6 Beta + 3042.9 Prior − 6691.1 Gender

7. Answer this question and the remaining questions in Part 1 using the backward selection model you found by hand. Write the equation of your model that describes your multiple regression model.

Salary = 7689.1 Education + 2599.6 Beta + 3042.9 Prior
+ (−6691.1) Gender

8. Construct a prediction interval for an employee with ~~gender 1~~, ~~age 40,~~ 15 years of prior experience, 5 years of beta experience (and thus 20 years of total experience) and education level 6.

7689.1(6) + 2599.6(5) + 3042.9(15) − 6691.1 = 98,085

est. t w/ 1.96

ME ± 1.96 * 8271 ± 16,211.16

(81,874, 114,296) approximately

9. Interpret the meaning of the gender coefficient in the context of the problem.

gender 1 is penalized by $6,691.10

10. Construct a confidence interval for the gender variable coefficient.

ME ± 1.96(1064.7) = 2086.81

(−8777.91, −4604.29) approximately

11. Test your model assumptions using your residual plots and other diagnostic plots. Do they appear to be approximately satisfied? Identify any potential outliers.

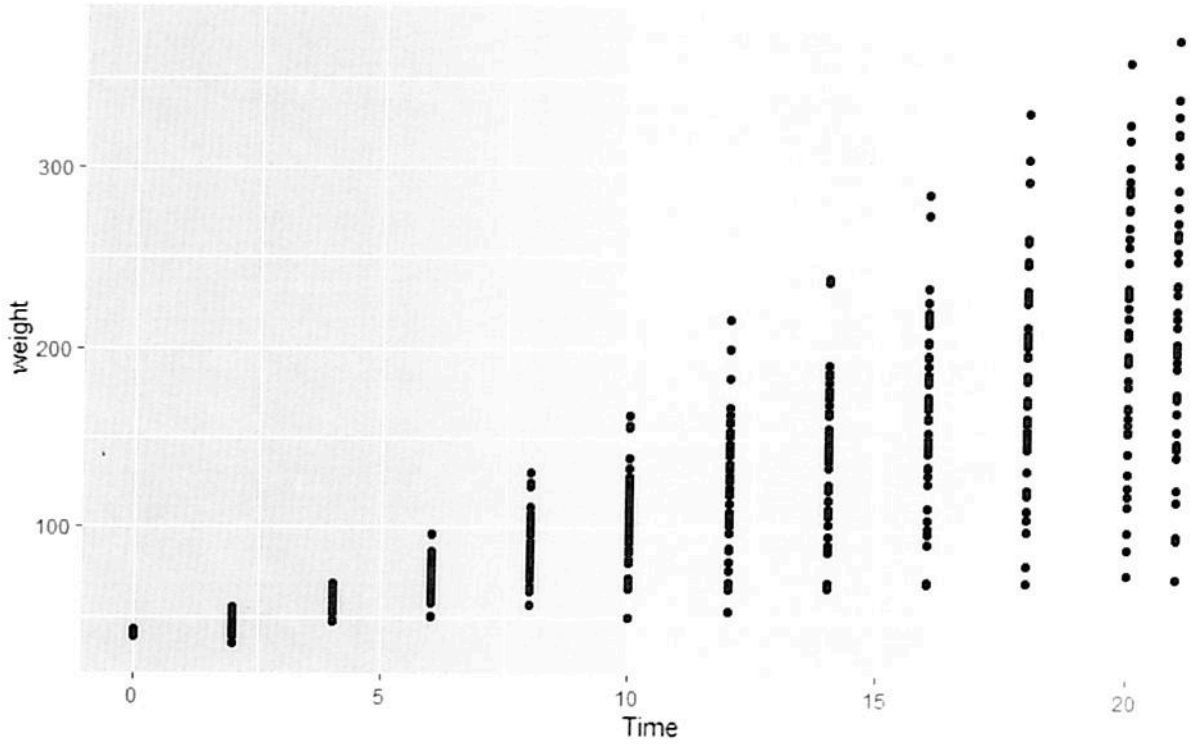
yes. even though the variables are discrete,
they do otherwise appear to be random, w/ constant variance
and residual normality plot is pretty good.

12. Based on your best model, interpret the meaning of the R^2 value.

0.9888
98.8% of variability in salary can be explained by
Gender, Education & experience (Prior+Beta).

Part II:

13. Examine the scatterplot below. Identify some potential issues with using a simple linear regression to model weight with Time.



there is a problem. The variance is not constant.

a transformation (like log) may be needed.

14. Recall that $Cov(X,Y) = E(XY) - E(X)E(Y)$. For the probability density function $f(x,y) = \frac{3}{512}x^3y^2, y \in [0,2], x \in [0,4]$, find the covariance.

$$E(XY) = \int_0^4 \int_0^2 \frac{3}{512} x^4 y^3 \, dy \, dx = \frac{3}{512} \cdot 4 \cdot \frac{1024}{5} = \frac{24}{5}$$

$$E(X) = \int_0^4 \int_0^2 \frac{3}{512} x^4 y^2 \, dy \, dx = \frac{3}{512} \cdot \frac{3}{8} \cdot \frac{1024}{5} = \frac{9}{20}$$

$$E(Y) = \int_0^4 \int_0^2 \frac{3}{512} x^3 y^3 \, dy \, dx = \frac{3}{512} \cdot 4 \cdot 64 = \frac{3}{2}$$

$$E(XY) - E(X)E(Y) = \frac{24}{5} - \frac{9}{20} \cdot \frac{3}{2} = \frac{33}{8} = 4.125$$

15. State the null and alternative hypothesis for a multiple regression model.

$H_0: \beta_i = 0 \ \forall i$

$H_a: \beta_i \neq 0$ for some i

16. Consider the small data set {(12,1), (8,3), (5,7)}. Find the value of the regression coefficients for $y = \beta_0 + \beta_1 x$, using the normal equation $(A^T A)^{-1} A^T Y = B$. Write the coefficients you find in the equation.

$$\begin{aligned} \beta_0 + 12\beta_1 &= 1 \\ \beta_0 + 8\beta_1 &= 3 \\ \beta_0 + 5\beta_1 &= 7 \end{aligned} \qquad A = \begin{bmatrix} 1 & 12 \\ 1 & 8 \\ 1 & 5 \end{bmatrix}, \; B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \; Y = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 1 & 1 \\ 12 & 8 & 5 \end{bmatrix} \qquad A^T A = \begin{bmatrix} 3 & 25 \\ 25 & 233 \end{bmatrix} \qquad A^T Y = \begin{bmatrix} 11 \\ 71 \end{bmatrix}$$

$$B = (A^T A)^{-1} A^T Y = \begin{bmatrix} 3 & 25 \\ 25 & 233 \end{bmatrix}^{-1} \begin{bmatrix} 11 \\ 71 \end{bmatrix} = \begin{bmatrix} 10.6486\ldots \\ -.8378\ldots \end{bmatrix}$$

$$\text{or } \begin{bmatrix} 394/37 \\ -31/37 \end{bmatrix}$$

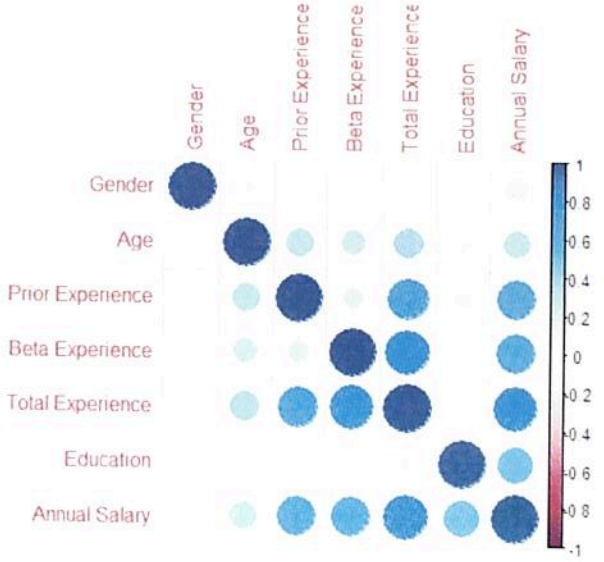
$$Y = \frac{394}{37} - \frac{31}{37} X$$

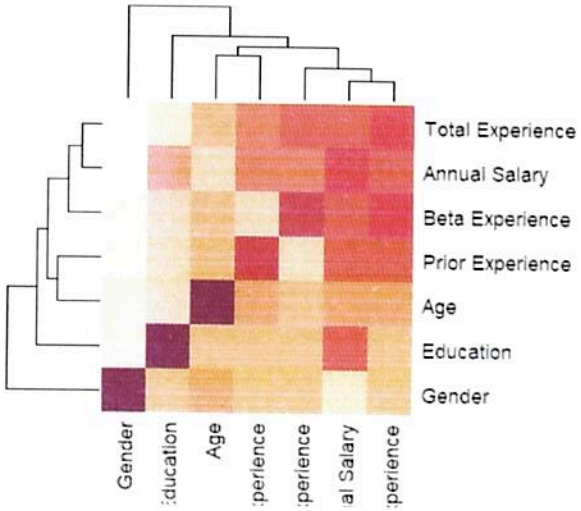
or

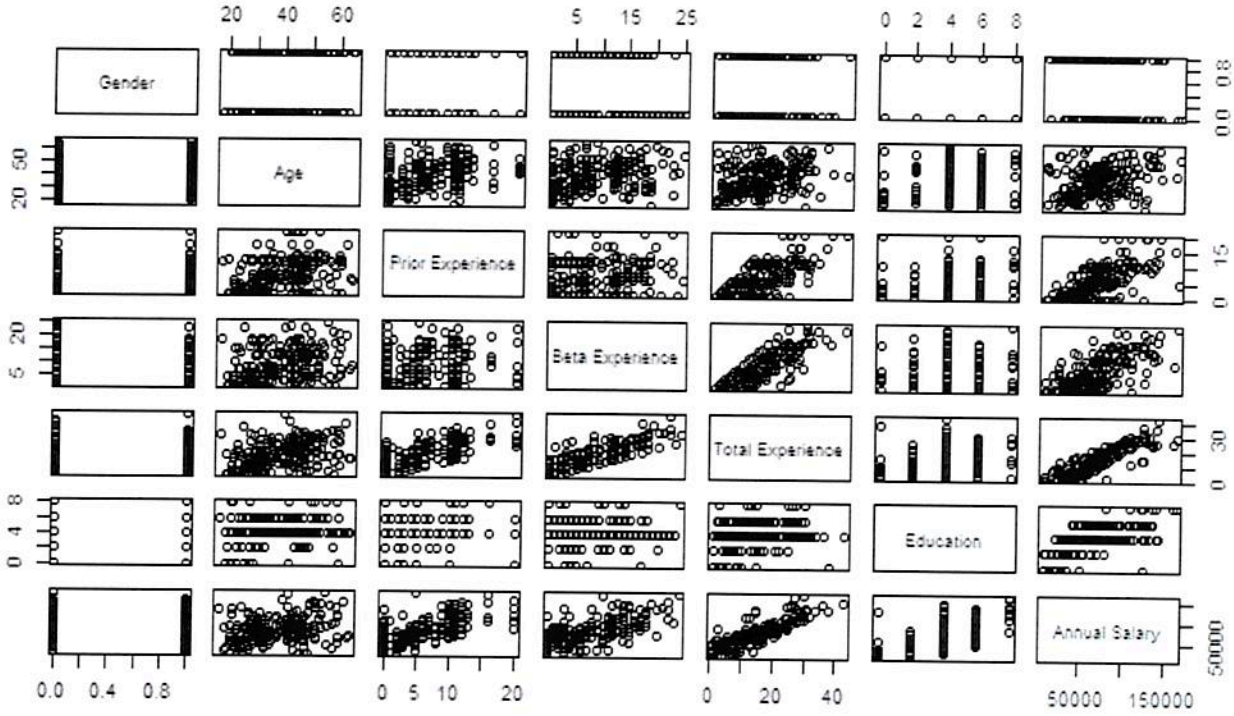
$$Y = 10.6486 - 0.8378 X$$

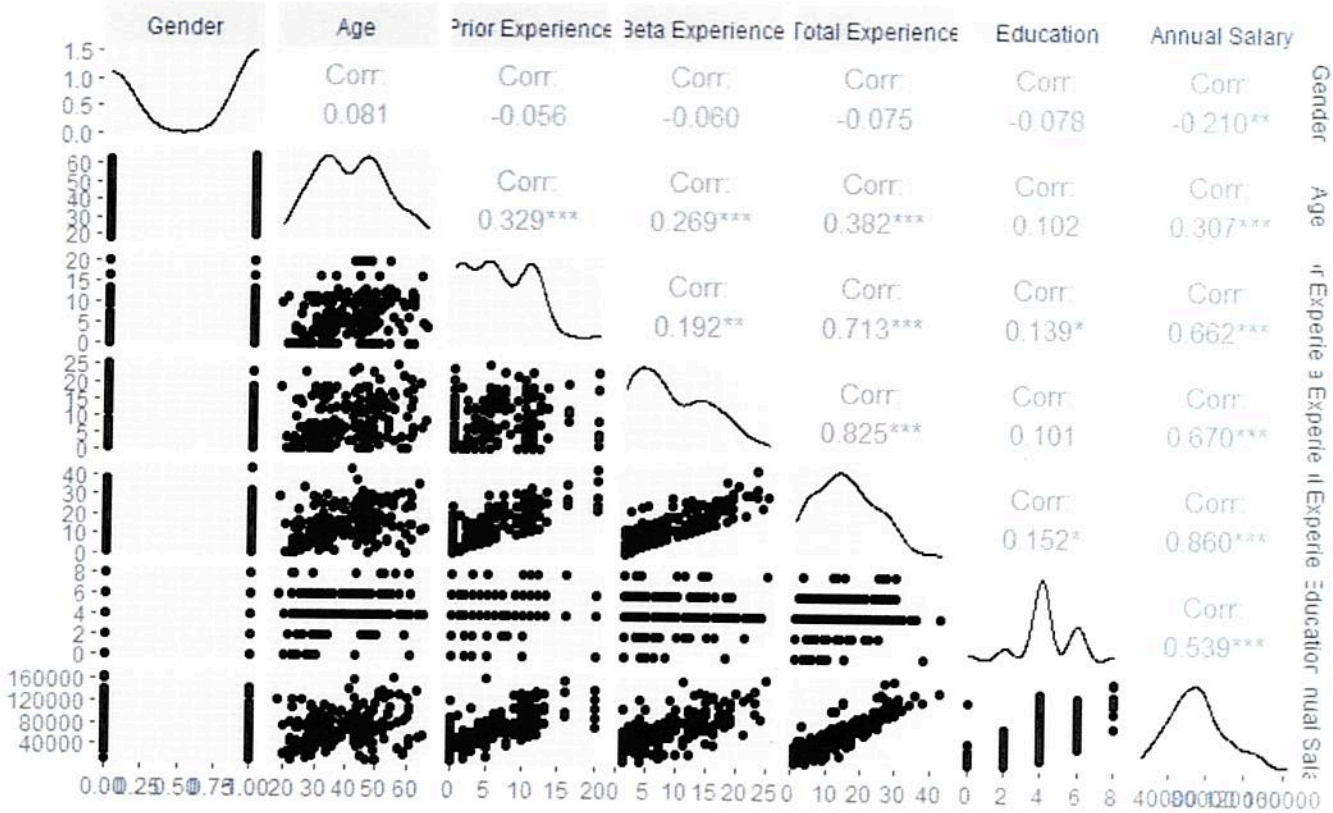
MTH 325 Exam #1 At Home analysis

	Gender	Age	Prior Experience	Beta Experience	Total Experience	Education	Annual Salary
Gender	1.00000000	0.08133774	-0.05591134	-0.05953542	-0.07472861	-0.07757681	-0.2103258
Age	0.08133774	1.00000000	0.32893803	0.26906401	0.38164829	0.10234155	0.3066272
Prior Experience	-0.05591134	0.32893803	1.00000000	0.19209036	**0.71318870**	0.13891933	0.6624701
Beta Experience	-0.05953542	0.26906401	0.19209036	1.00000000	**0.82491473**	0.10073427	0.6696936
Total Experience	-0.07472861	0.38164829	**0.71318870**	**0.82491473**	1.00000000	0.15196711	0.8599165
Education	-0.07757681	0.10234155	0.13891933	0.10073427	0.15196711	1.00000000	0.5389206
Annual Salary	-0.21032579	0.30662717	0.66247007	0.66969362	0.85991652	0.53892062	1.0000000






Gender
Age
Prior Experience
Beta Experience
Total Experience
Education
Annual Salary


Gender
Age
Prior Experience
Beta Experience
Total Experience
Education
Annual Salary
Corr: 0.081
Corr: -0.056
Corr: -0.060
Corr: -0.075
Corr: -0.078
Corr: -0.210**
Corr: 0.329***
Corr: 0.269***
Corr: 0.382***
Corr: 0.102
Corr: 0.307***
Corr: 0.192**
Corr: 0.713***
Corr: 0.139*
Corr: 0.662***
Corr: 0.825***
Corr: 0.101
Corr: 0.670***
Corr: 0.152*
Corr: 0.860***
Corr: 0.539***

Call:
lm(formula = Annual_Salary ~ Age, data = data)

Residuals:

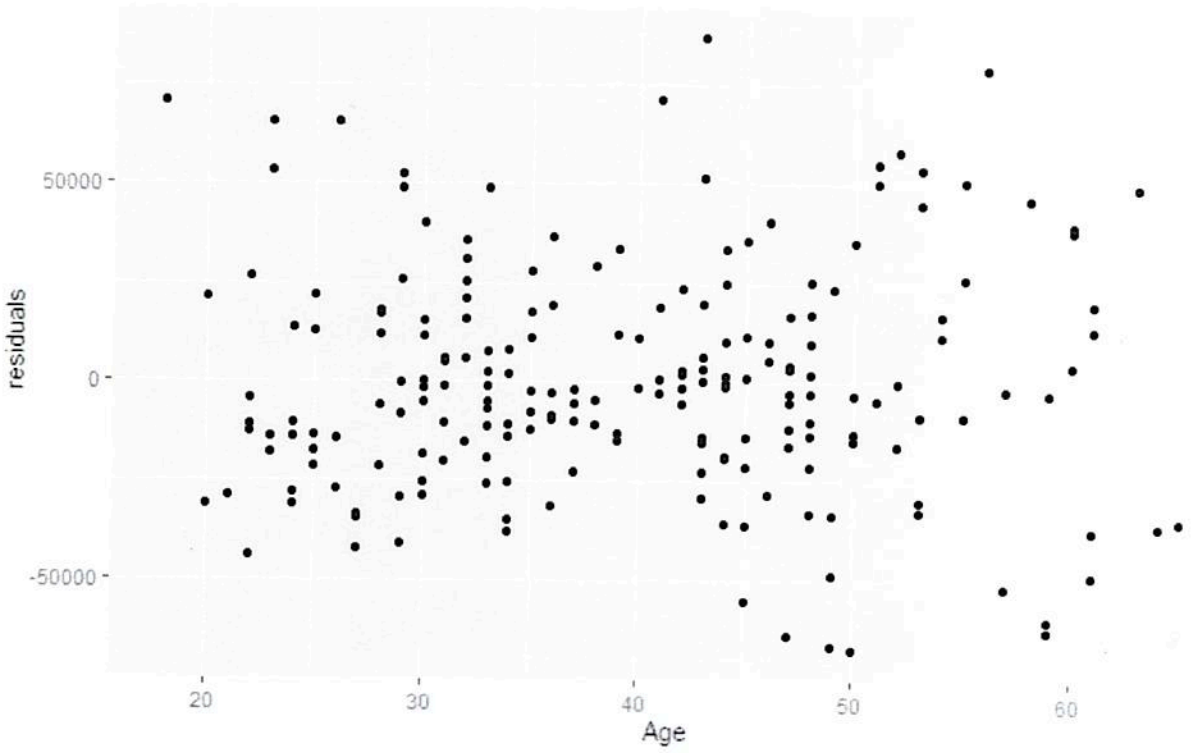
Min	1Q	Median	3Q	Max
-67614	-16564	-3006	16755	86418

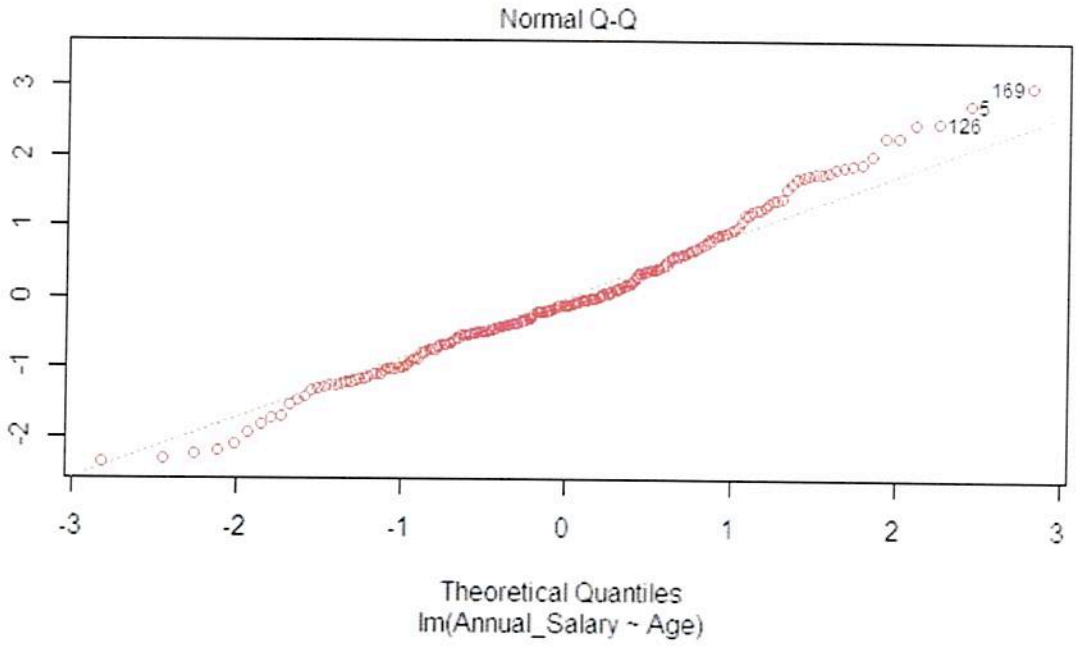
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37745.7	7596.8	4.969	1.43e-06 ***
Age	847.4	185.1	4.579	8.18e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28870 on 202 degrees of freedom
Multiple R-squared: 0.09402, Adjusted R-squared: 0.08954
F-statistic: 20.96 on 1 and 202 DF, p-value: 8.176e-06





Call:
lm(formula = Annual_Salary ~ ., data = data)

Residuals:

Min	1Q	Median	3Q	Max
-26726.3	-3595.4	-1063.6	419.3	24967.4

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5429.21	2598.12	2.090	0.0379 *
Gender	-7061.07	1181.88	-5.974	1.06e-08 ***
Age	-101.83	57.71	-1.764	0.0792 .
Prior_Experience	3074.19	123.77	24.838	< 2e-16 ***
Beta_Experience	2601.00	97.53	26.669	< 2e-16 ***
Total_Experience	NA	NA	NA	NA
Education	7413.20	350.97	21.122	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8217 on 198 degrees of freedom
Multiple R-squared: 0.928, Adjusted R-squared: 0.9262
F-statistic: 510.7 on 5 and 198 DF, p-value: < 2.2e-16

Forward selection with AIC

Call:
lm(formula = Annual_Salary ~ Gender + Age + Prior_Experience +
Beta_Experience + Total_Experience + Education, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-26726.3	-3595.4	-1063.6	419.3	24967.4

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5429.21	2598.12	2.090	0.0379 *
Gender	-7061.07	1181.88	-5.974	1.06e-08 ***
Age	-101.83	57.71	-1.764	0.0792 .
Prior_Experience	3074.19	123.77	24.838	< 2e-16 ***
Beta_Experience	2601.00	97.53	26.669	< 2e-16 ***
Total_Experience	NA	NA	NA	NA
Education	7413.20	350.97	21.122	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8217 on 198 degrees of freedom
Multiple R-squared: 0.928, Adjusted R-squared: 0.9262
F-statistic: 510.7 on 5 and 198 DF, p-value: < 2.2e-16

Best subset regression generates an error because of collinearity

Backward selection with AIC
Call:
lm(formula = Annual_Salary ~ Gender + Age + Prior_Experience +
Beta_Experience + Education, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-26726.3	-3595.4	-1063.6	419.3	24967.4

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5429.21	2598.12	2.090	0.0379 *
Gender	-7061.07	1181.88	-5.974	1.06e-08 ***
Age	-101.83	57.71	-1.764	0.0792 .
Prior_Experience	3074.19	123.77	24.838	< 2e-16 ***
Beta_Experience	2601.00	97.53	26.669	< 2e-16 ***
Education	7413.20	350.97	21.122	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8217 on 198 degrees of freedom

Multiple R-squared: 0.928, Adjusted R-squared: 0.9262
F-statistic: 510.7 on 5 and 198 DF, p-value: < 2.2e-16

Call:
lm(formula = Annual_Salary ~ Gender + Prior_Experience + Beta_Experience +
Education + 0, data = data)

Residuals:

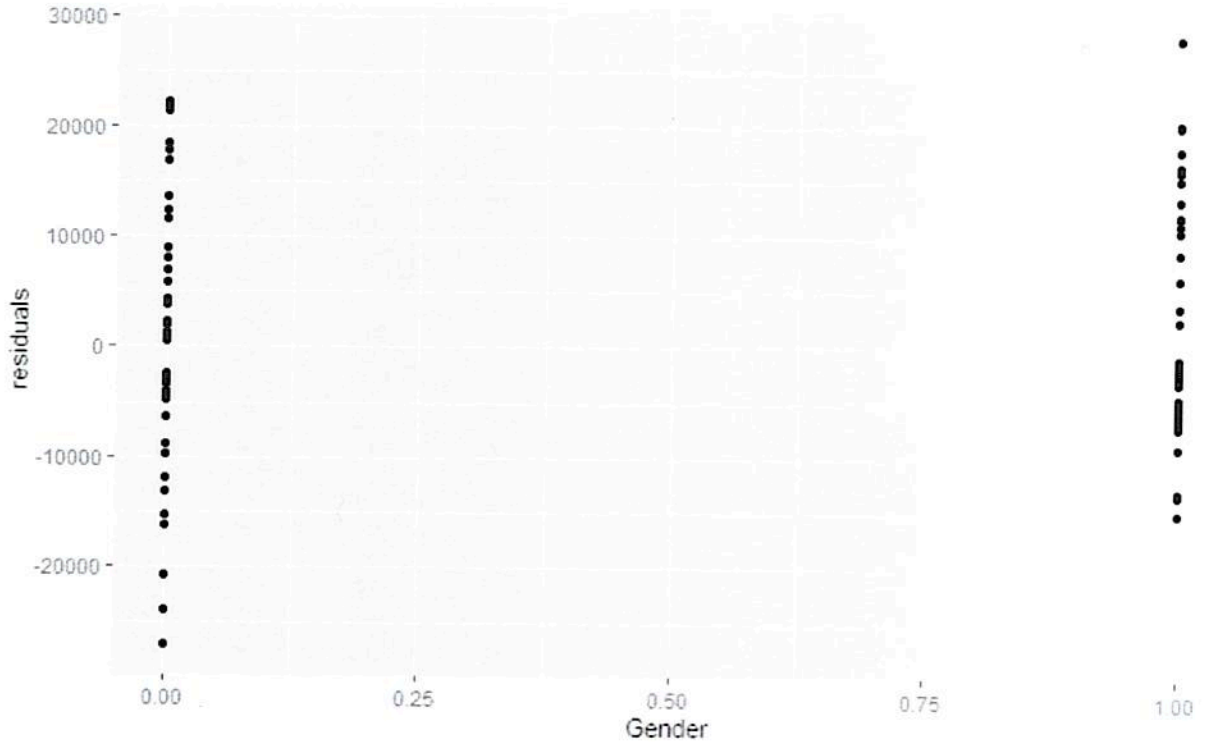
Min	1Q	Median	3Q	Max
-27045.9	-4232.3	-1240.9	977.2	27991.5

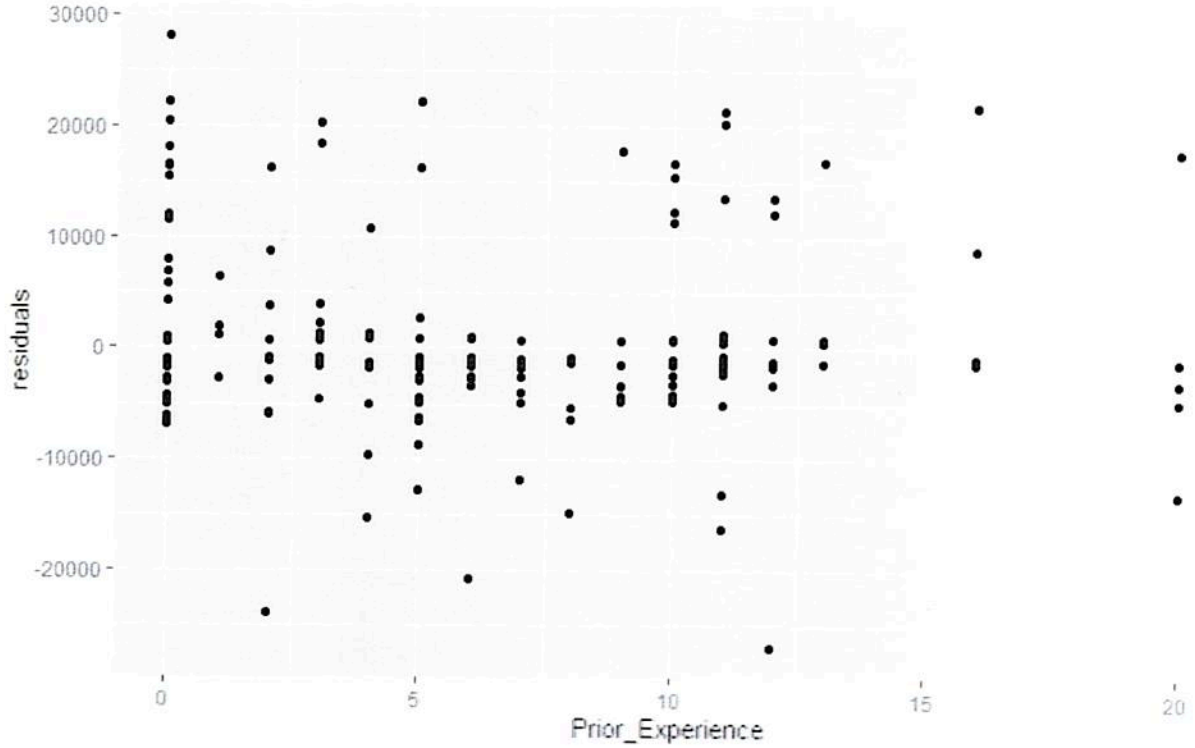
Coefficients:

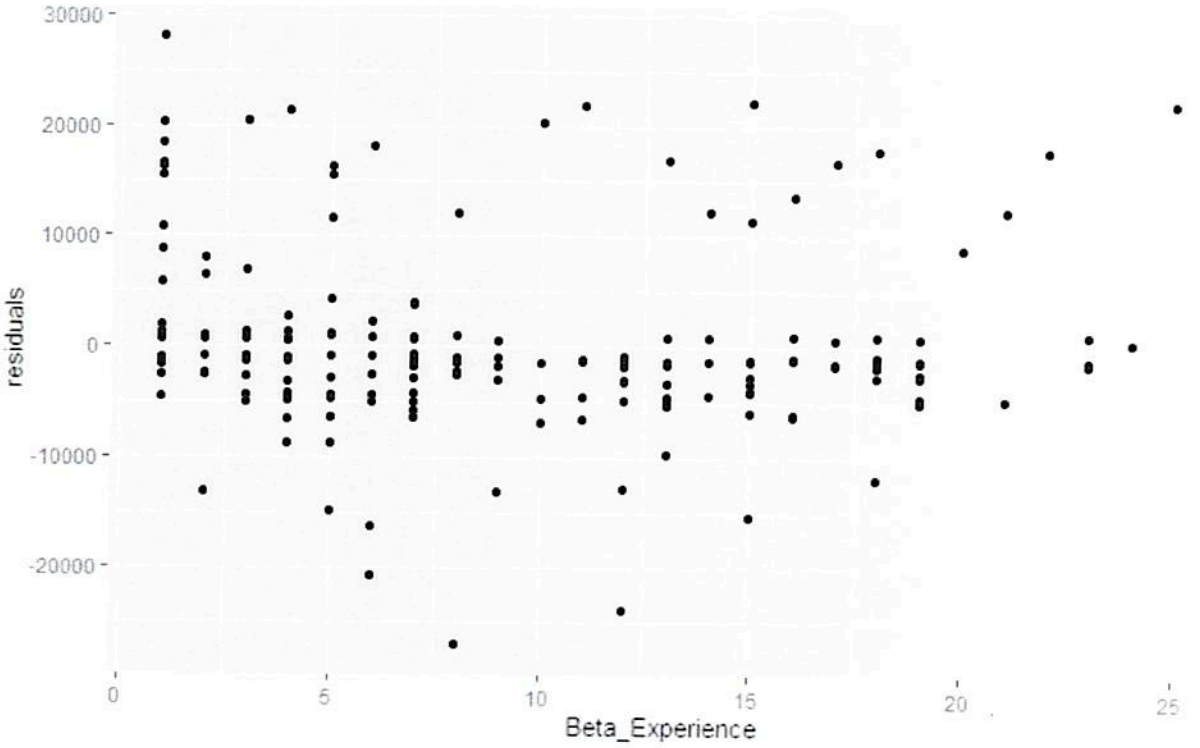
	Estimate	Std. Error	t value	Pr(>|t|)
Gender	-6691.1	1064.7	-6.285	2.02e-09 ***
Prior_Experience	3042.9	116.1	26.200	< 2e-16 ***
Beta_Experience	2599.6	90.7	28.660	< 2e-16 ***
Education	7689.1	247.6	31.059	< 2e-16 ***

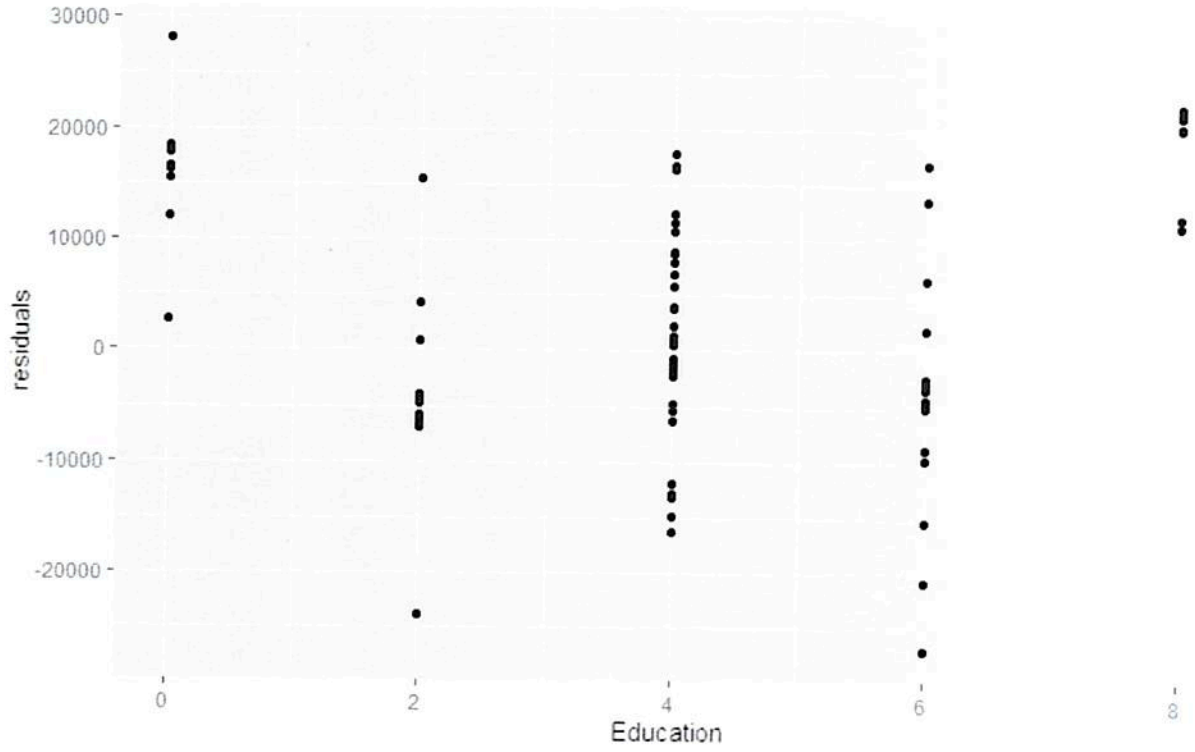
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

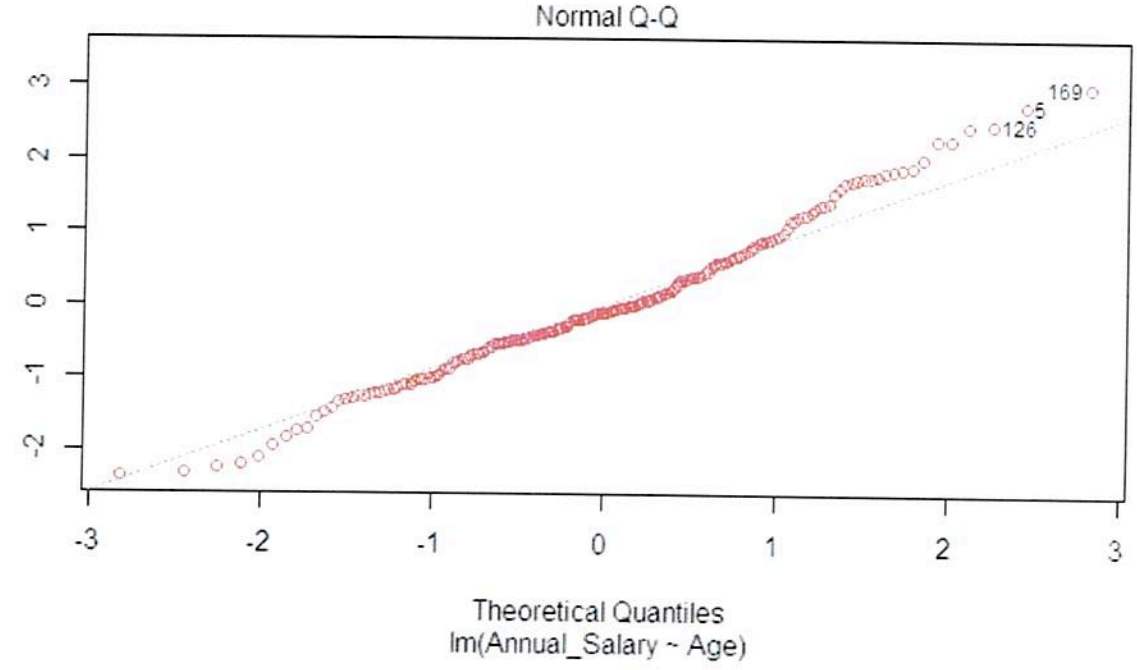
Residual standard error: 8271 on 200 degrees of freedom
Multiple R-squared: 0.9888, Adjusted R-squared: 0.9886
F-statistic: 4416 on 4 and 200 DF, p-value: < 2.2e-16




30000
20000
10000
0
-10000
-20000
residuals
0
5
10
15
20
Prior_Experience


30000
20000
10000
0
-10000
-20000
residuals
0
5
10
15
20
25
Beta_Experience


30000
20000
10000
0
-10000
-20000
residuals
0
2
4
6
8
Education


Normal Q-Q
169
5
126
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(Annual_Salary ~ Age)