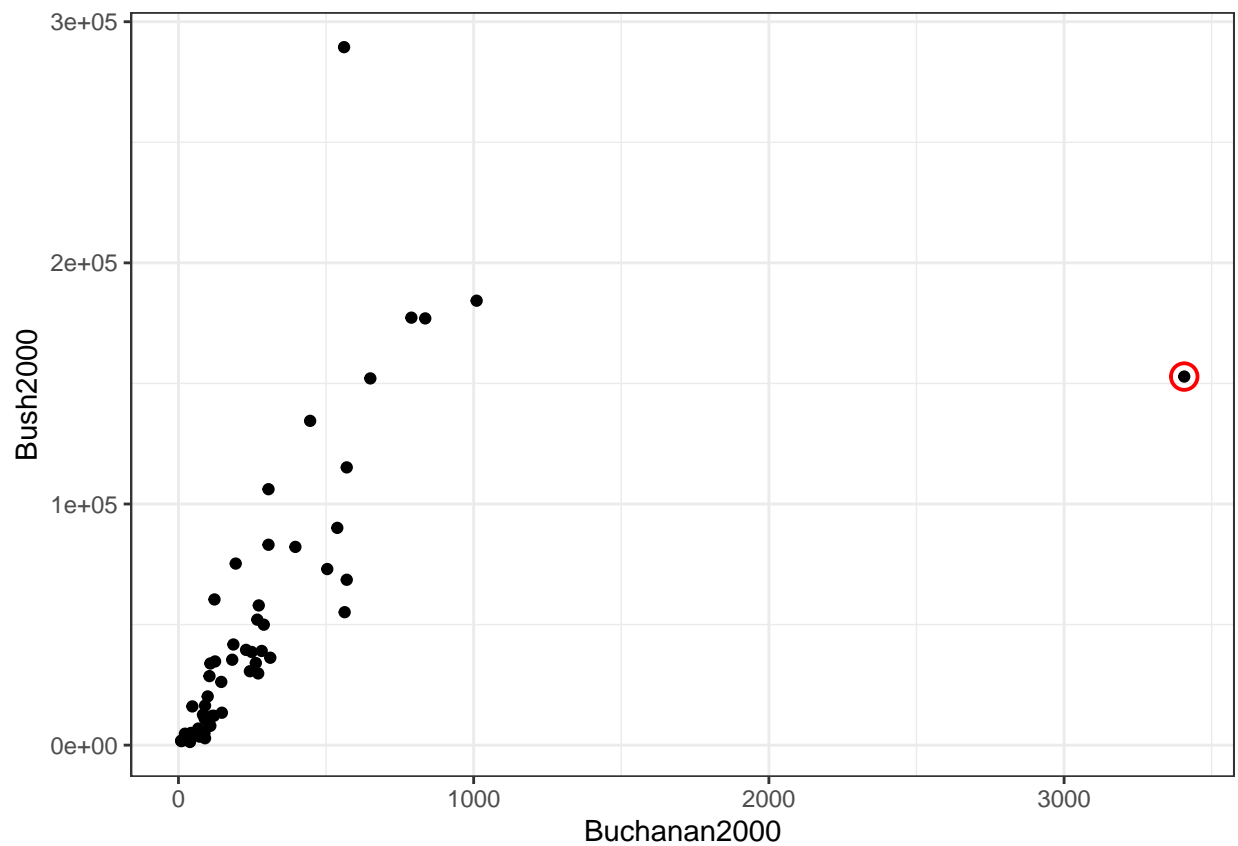


Question 2

QUESTION 1

Make a scatterplot of the variables Buchanan2000 and Bush2000. What evidence is there in the scatterplot that Buchanan received more votes than expected in Palm Beach County?



As shown in the above scatter plot, the circled data point identifies that Buchanan received more votes than expected in Palm Beach County. As the circled point is clearly an outlier, illustrating that is the only county where the votes are past 3000.

(B)

Fit a linear regression model to the data to predict Buchanan votes from Bush votes, without using Palm Beach County results. You should consider transformations for both variables if you think there's a violation of normality and/or linearity.

$$y = \hat{B}_0 + \hat{B}_1 x$$

$$\hat{\log(\text{Buchanan2000})} = -2.34149 + 0.73096 * \log(\text{Bush2000})$$

(C)

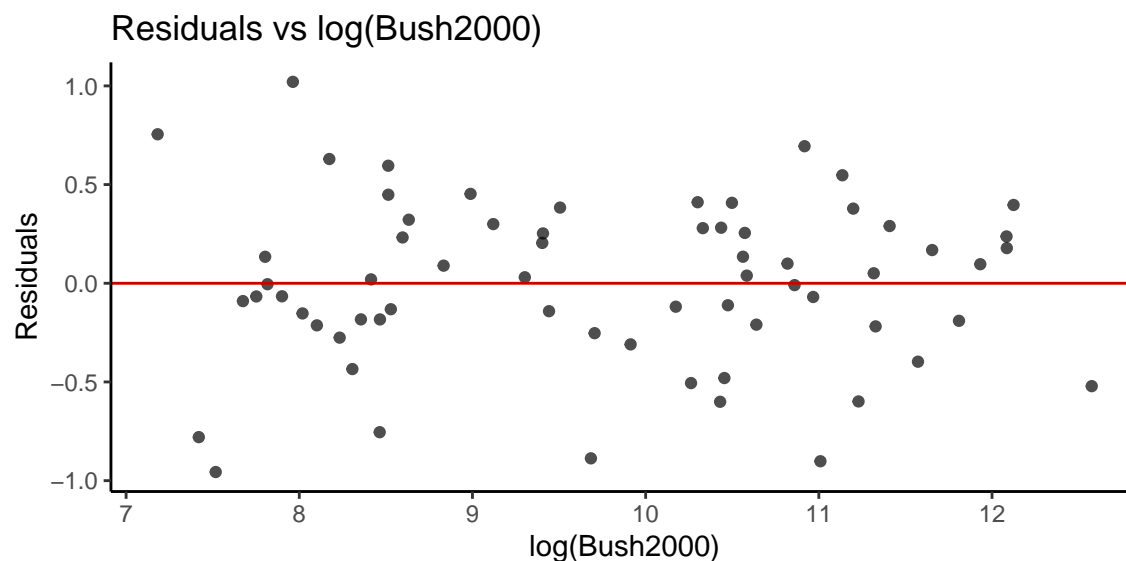
Include the output from the final regression model that you used, as well as evidence that the model fits the assumptions reasonably well.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.341	0.3544	-6.607	9.071e-09
log(Bush2000)	0.731	0.03597	20.32	1.294e-29

Table 2: Fitting linear model: $\log(\text{Buchanan2000}) \sim \log(\text{Bush2000})$

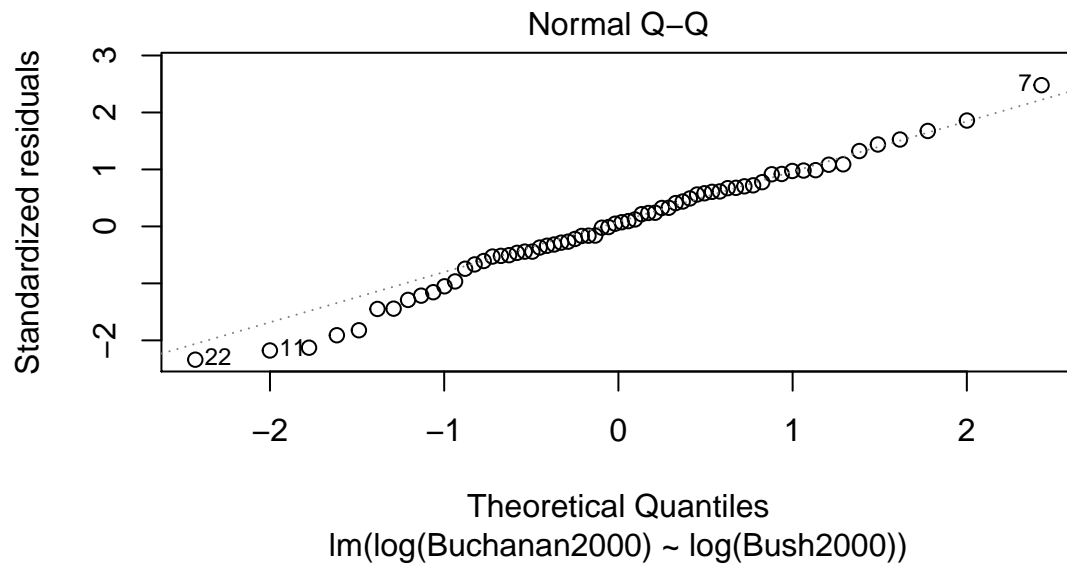
Observations	Residual Std. Error	R^2	Adjusted R^2
66	0.4198	0.8658	0.8637

Test for Linearity



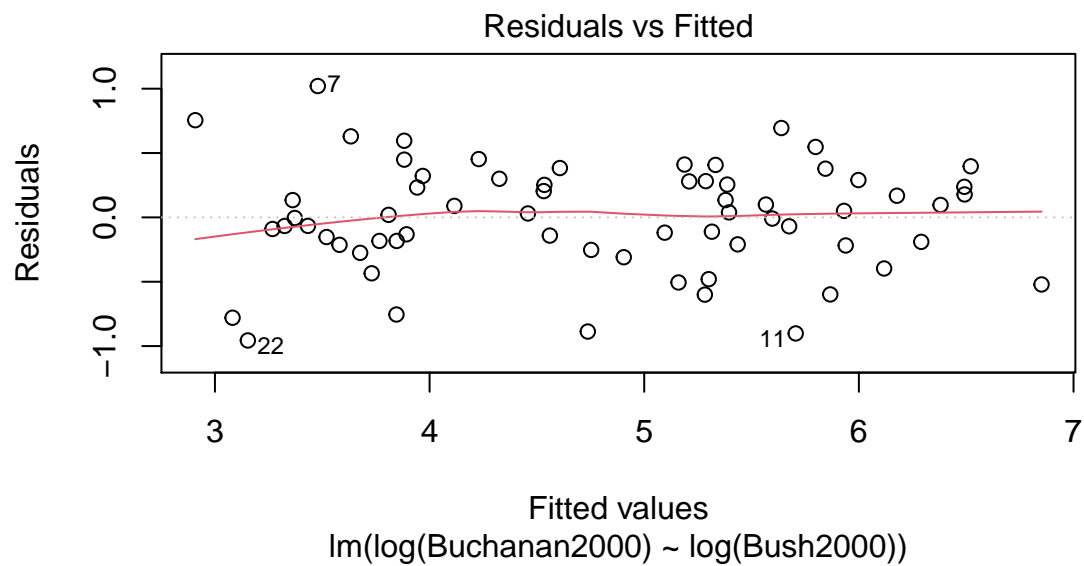
As there exists no distinct pattern, the model linearity assumption holds true.

Test for Normality



As most of data points lie on the 45 degree line, the model normality assumption holds true.

Test for Independence and Equal Variance



As the data points are spread throughout the X axis and are around the 0 residual mark on the Y axis, the model independence and equal variance holds true.

(D)

Obtain a 95% prediction interval for the number of Buchanan votes in Palm Beach from this result, assuming the relationship is the same in this county as in the others. If it is assumed that Buchanan's actual count contains a number of votes intended for Gore, what can be said about the likely size of this number from the prediction interval?

Bush2000	fit	lwr	upr
152846	592.4	250.8	1399

Initially, the votes for Buchanan at Palm Beach County were 3407. When the linear regression model was fitted after removing this point, for the same number of votes for Bush at Palm Beach County (153846), the model predicts the number of votes for Buchanan to be ~592 votes. At 95% prediction interval, the upper limit is 1399.164. As the prediction intervals are wider, the upper limit is still significantly lower than the actual votes for Buchanan (3407). This concludes that Buchanan's actual count contains a number of votes intended for Gore.