Normality Test

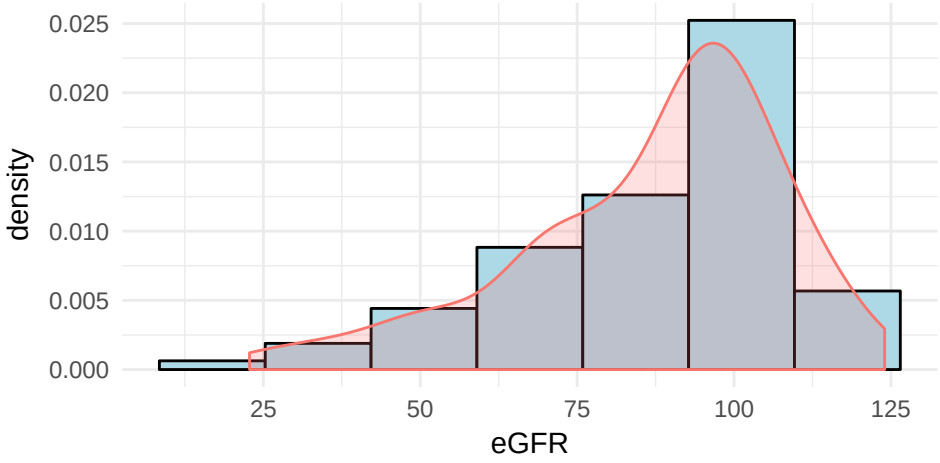
Katica Ristic

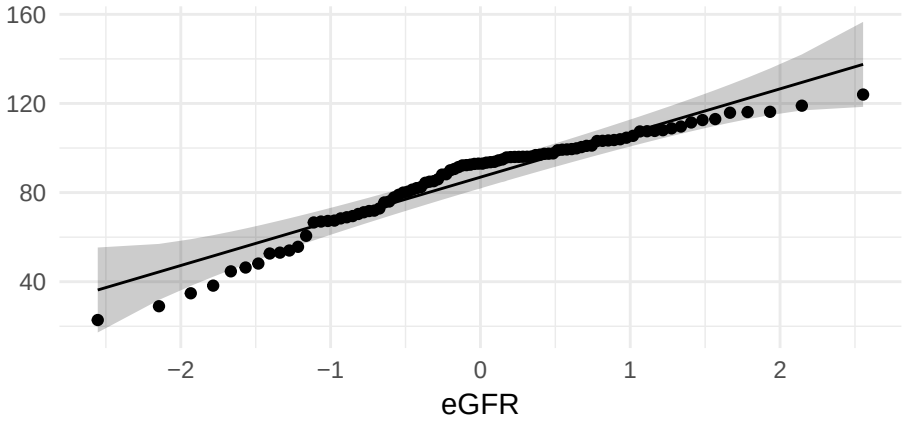
May 14, 2018

Assess the normality of the data - visual method with nomality test (comparing Shapiro-Wilk's, Anderson-Darling and D'Agostino - Pearson test)

eGFR

From the histogram and densety line of eGRF one could see the eGRF is left skewed distribution. Also on the eGRF q -q plot points are not following closely the line x = y, which indicate that the distribution of the data is not normal.





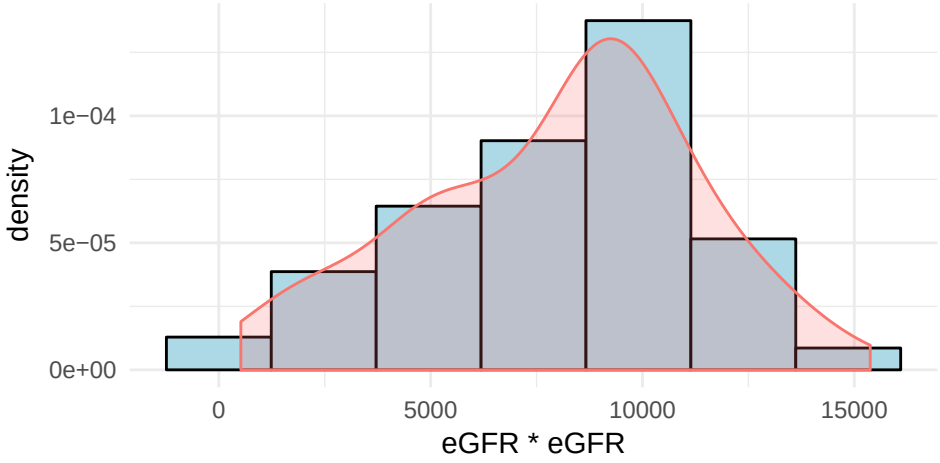
Let us now test the normallity of the data with three statistical tests.

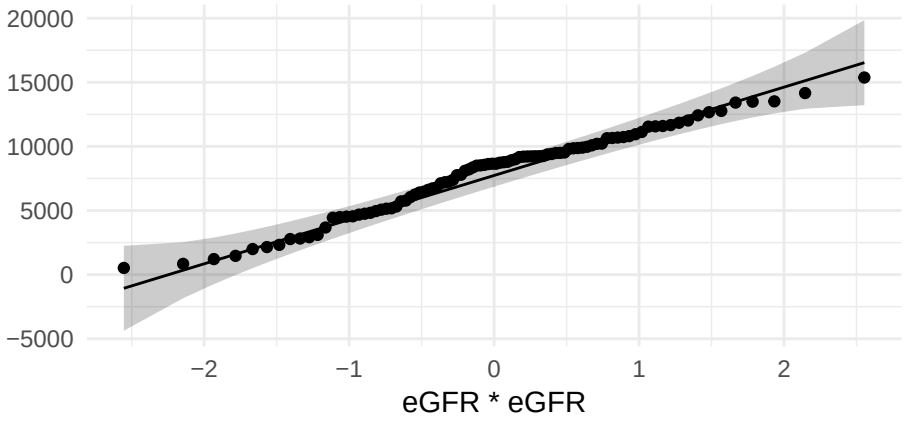
Table 1: P-values of different normality tests for eGFR.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.000129	1.87e-05	0.0011694

All have p-values less than 0.05, which means that we reject the null hypothesis:"Data is normally distributed".

Some methods of making data normally distributed is to use the multiplicative inverse (reciprocal) or the square transformation. After applaying the square transformation on eGFR the histogram and q - q plot show normal distribution.



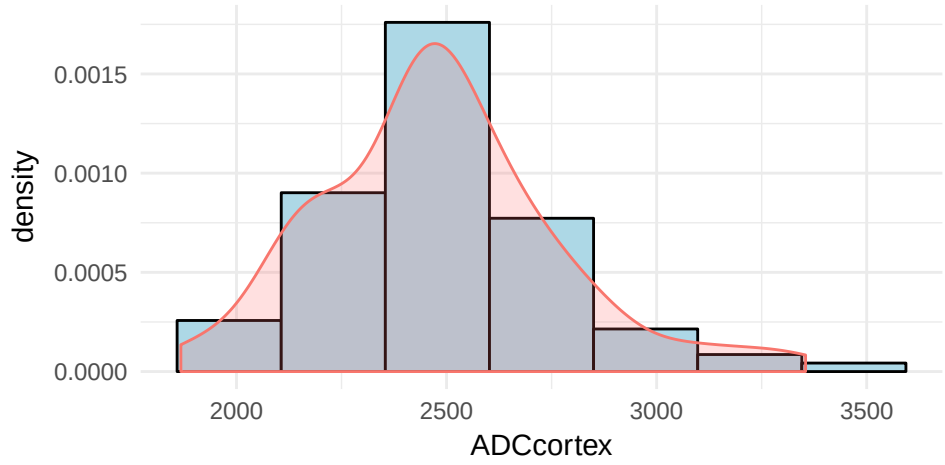


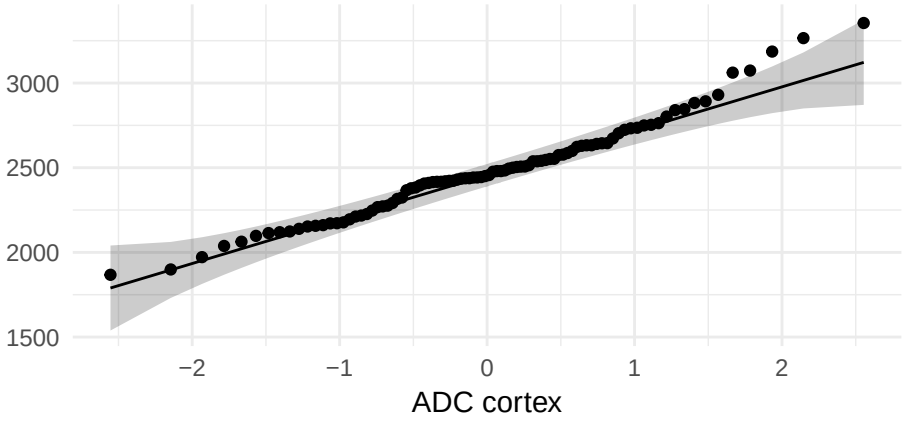
Also two out of three p - values of transformed data are now greater than 0.05, which means that we fail to reject the null hypothesis: "Data is normally distributed". These two test are good candidates for choosing the normality test. Let us see how the rest of the variables will respond to them and to the corresponding plots.

Table 2: P-values of different normality tests for eGFR * eGFR.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.1116641	0.0353241	0.3000796

ADC cortex



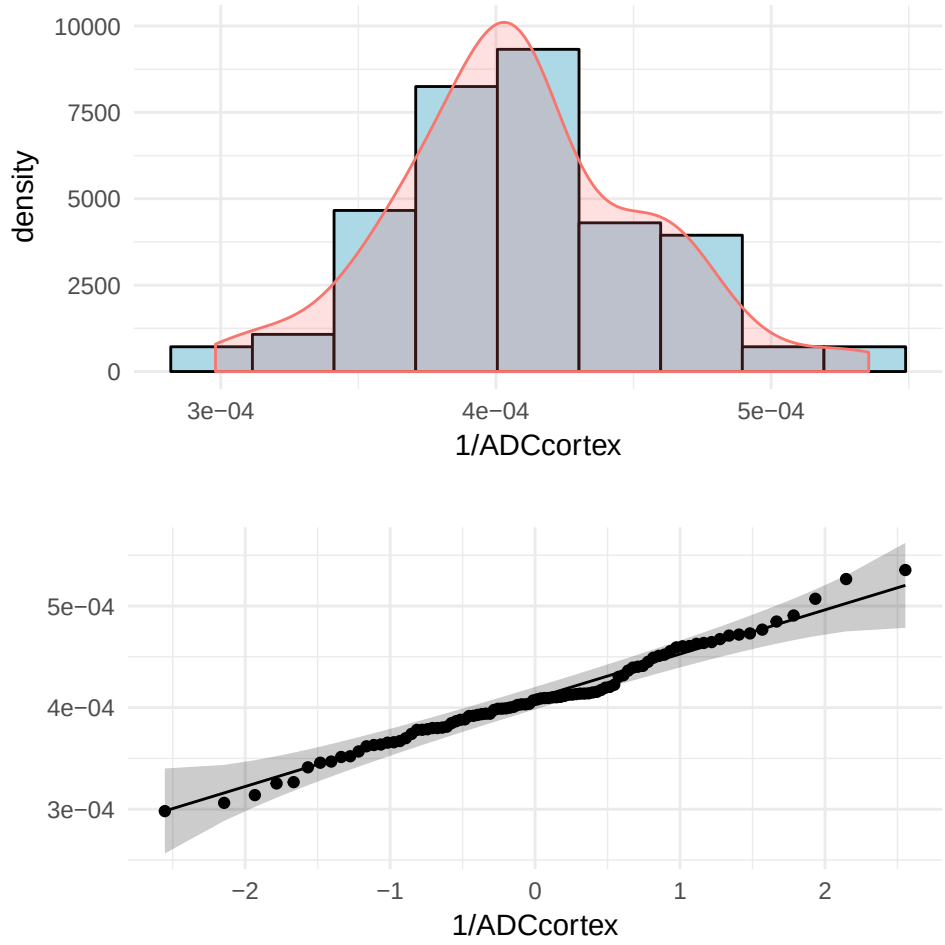


The same two out of three statistical tests for testing the normality of the data are less than 0.05, so we reject the null hypothesis: "Data is normally distributed" and transformation of the data will be preformed.

Table 3: P-values of different normality tests for ADC cortex.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.0428553	0.0518566	0.0260224

ADC cortex with multiplicative inverse (reciprocal) transformation

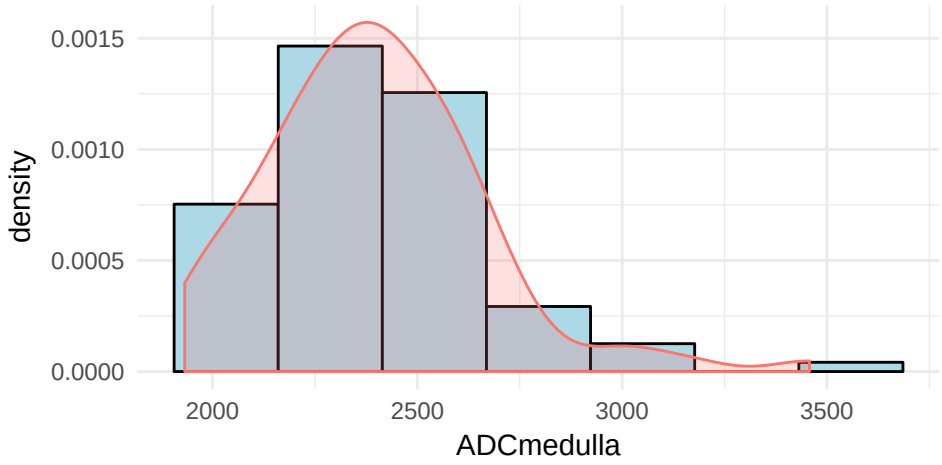


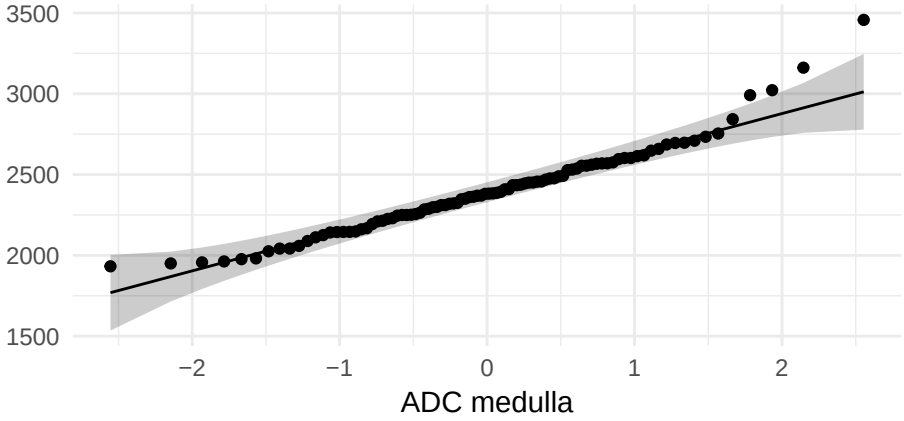
P - values of transformed data for all three tests are now greater than 0.05, which means that we fail to reject the null hypothesis: " Data is normally distributed ".

Table 4: P-values of different normality tests for 1/ADCcortex.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.3876616	0.1193183	0.4614084

ADC medulla



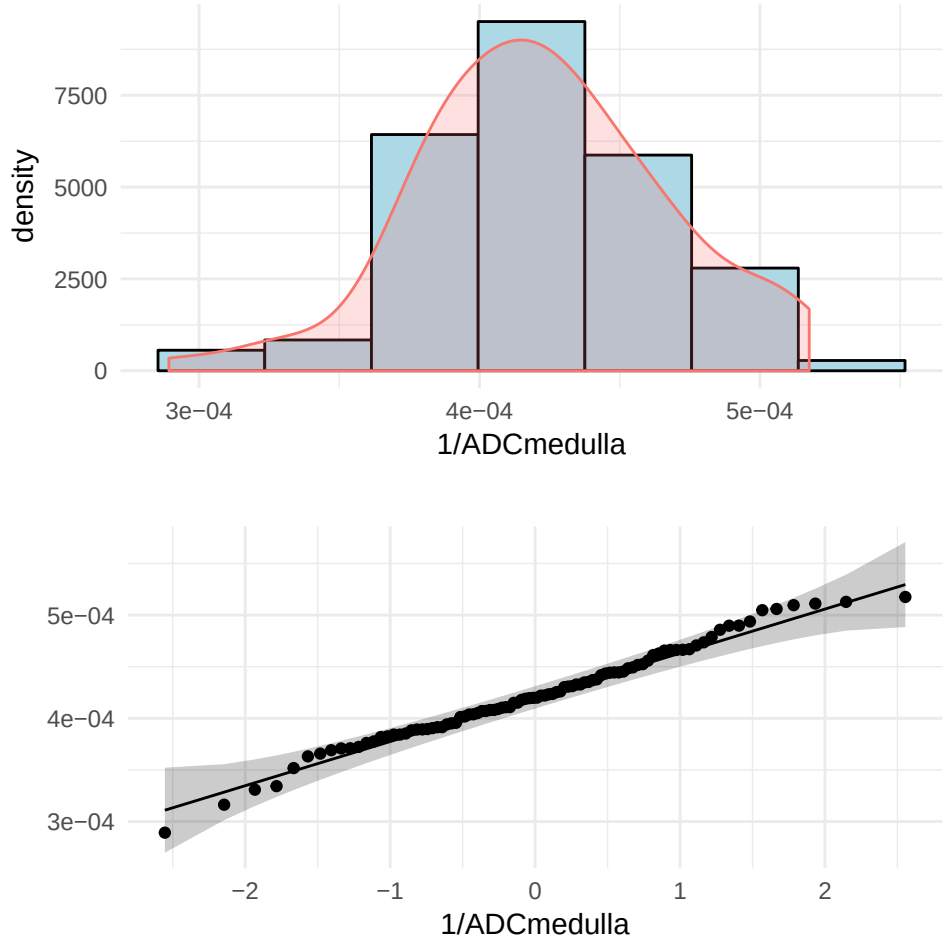


Both of the candidate tests for testing the normality of the data are far less than 0.05, so we reject the null hypothesis: "Data is normally distributed" and the transformation of the data will be preformed.

Table 5: P-values of different normality tests for ADC medulla.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.0013887	0.0899483	5.14e-05

ADC medulla with multiplicative inverse (reciprocal) transformation transformation

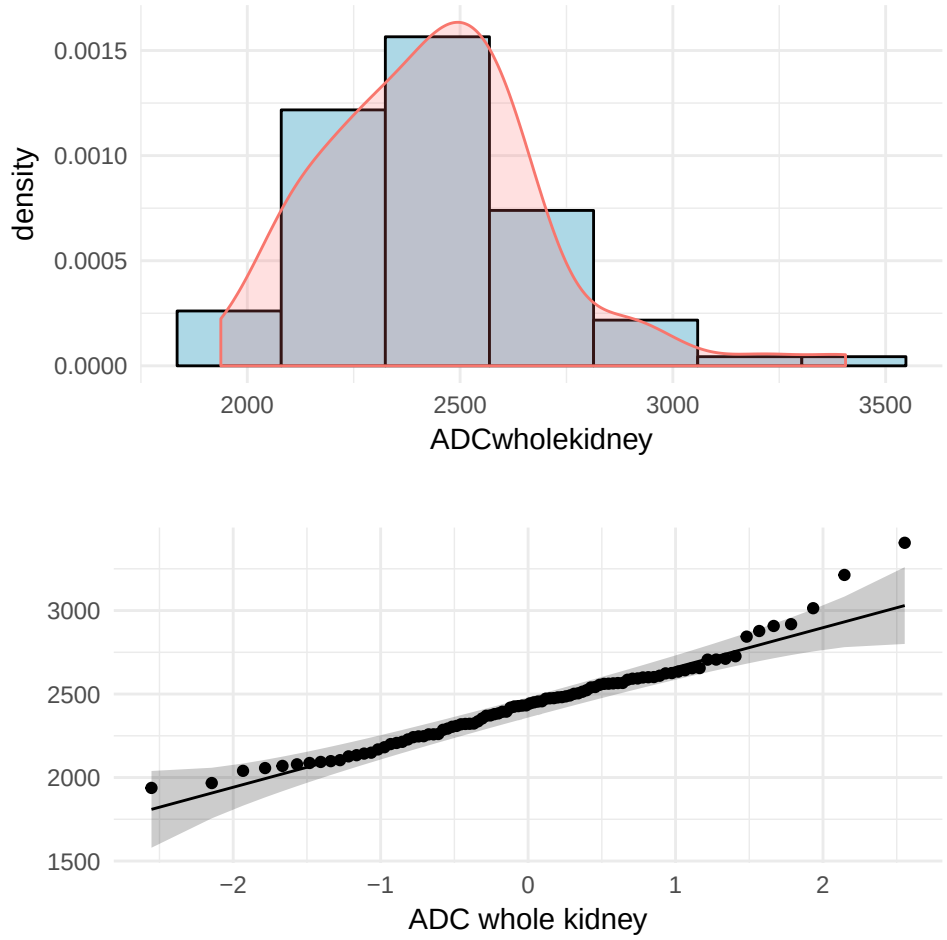


All the P - values of transformed data are now greater than 0.05, which means that we fail to reject the null hypothesis: "Data is normally distributed". So far the D'Agostino - Pearson test is the most sensitive to the data.

Table 6: P-values of different normality tests for 1/ADCmedulla.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.4639489	0.5515183	0.7432711

ADC whole kidney

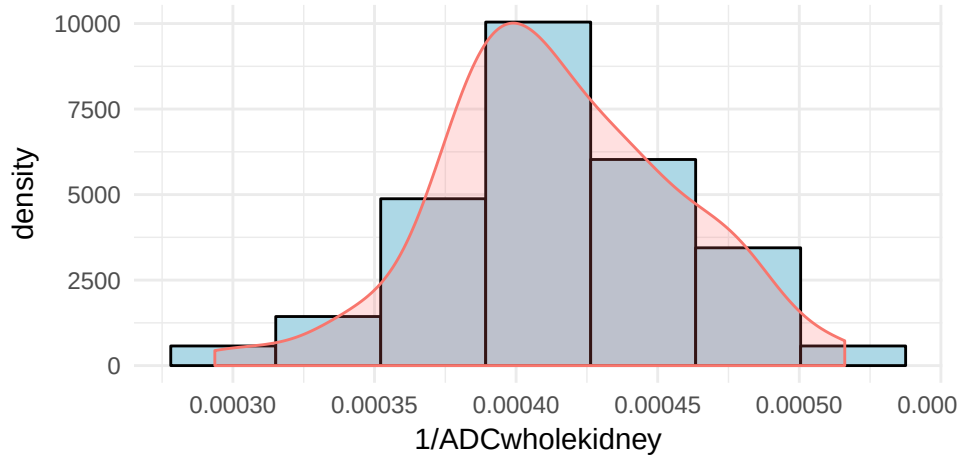


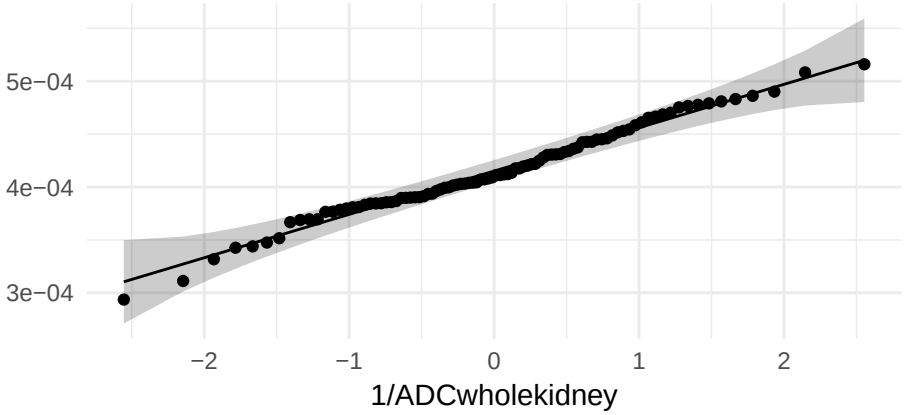
Our most promising candidates for testing the normality of the data are less than 0.05, so we reject the null hypothesis: "Data is normally distributed" and transformation of the data will be preformed.

Table 7: P-values of different normality tests for ADC whole kidney.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.0039673	0.0840606	0.0003769

ADC whole kidney with multiplicative inverse (reciprocal) transformation



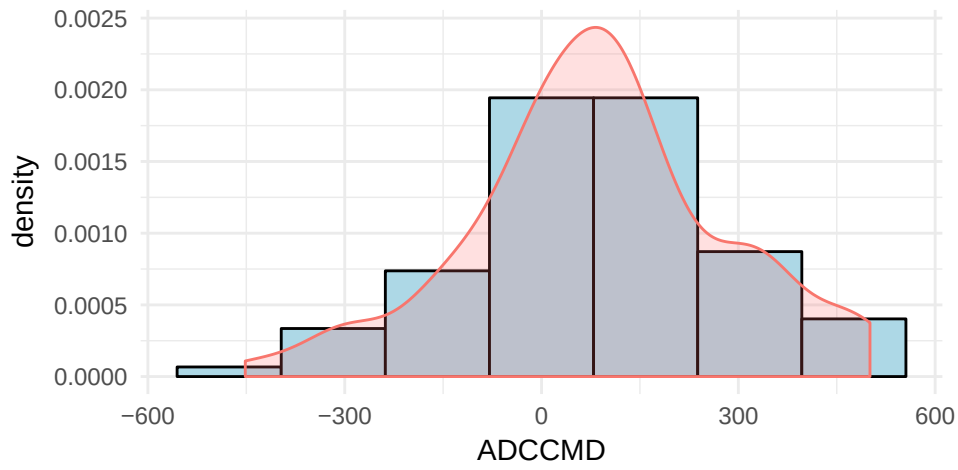


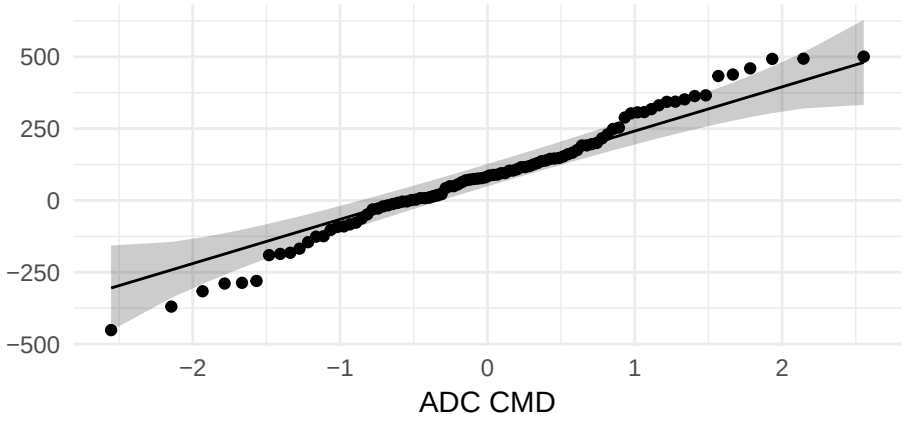
P - values of transformed data for all three tests are now greater than 0.05, which means that we fail to reject the null hypothesis: " Data is normally distributed ".

Table 8: P-values of different normality tests for 1/ADCwholekidney.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.5477818	0.2794742	0.8134307

ADC CMD





For all three tests ADC CMD p - value is greater tnan 0.05, so we fail to reject the null hypothesis: " Data is normally distributed ". So we do not need to transform the data.

Table 9: P-values of different normality tests for ADC CMD.

	Shapiro - Wilk	Anderson Darling	D'Agostino - Pearson
p-value	0.3592056	0.2253156	0.7733204

Conclusion

Comparing three normality statistical tests for this particular data the obvious choice is **D'Agostino - Pearson** test. This test is most sensitive to the data, as mentioned before, and also the most closely corresponds to the conclusions derived from the plots.