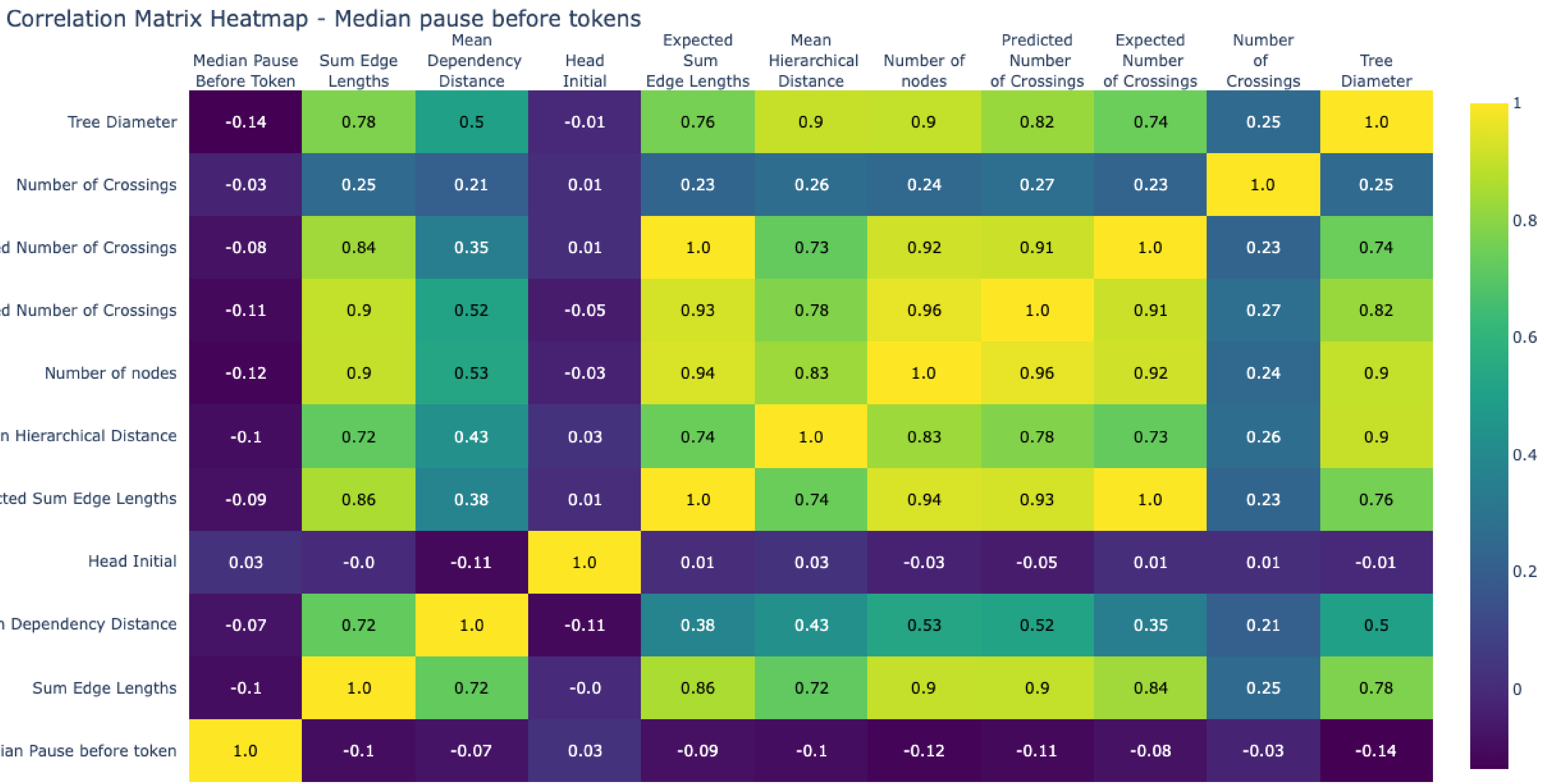
What I tried

- Redid the correlation with the median rather than average



Correlation Matrix Heatmap - Median pause before tokens

	Median Pause Before Token	Sum Edge Lengths	Mean Dependency Distance	Head Initial	Expected Sum Edge Lengths	Mean Hierarchical Distance	Number of nodes	Predicted Number of Crossings	Expected Number of Crossings	Number of Crossings	Tree Diameter
Tree Diameter	-0.14	0.78	0.5	-0.01	0.76	0.9	0.9	0.82	0.74	0.25	1.0
Number of Crossings	-0.03	0.25	0.21	0.01	0.23	0.26	0.24	0.27	0.23	1.0	0.25
Expected Number of Crossings	-0.08	0.84	0.35	0.01	1.0	0.73	0.92	0.91	1.0	0.23	0.74
Predicted Number of Crossings	-0.11	0.9	0.52	-0.05	0.93	0.78	0.96	1.0	0.91	0.27	0.82
Number of nodes	-0.12	0.9	0.53	-0.03	0.94	0.83	1.0	0.96	0.92	0.24	0.9
Mean Hierarchical Distance	-0.1	0.72	0.43	0.03	0.74	1.0	0.83	0.78	0.73	0.26	0.9
Expected Sum Edge Lengths	-0.09	0.86	0.38	0.01	1.0	0.74	0.94	0.93	1.0	0.23	0.76
Head Initial	0.03	-0.0	-0.11	1.0	0.01	0.03	-0.03	-0.05	0.01	0.01	-0.01
Mean Dependency Distance	-0.07	0.72	1.0	-0.11	0.38	0.43	0.53	0.52	0.35	0.21	0.5
Sum Edge Lengths	-0.1	1.0	0.72	-0.0	0.86	0.72	0.9	0.9	0.84	0.25	0.78
Median Pause before token	1.0	-0.1	-0.07	0.03	-0.09	-0.1	-0.12	-0.11	-0.08	-0.03	-0.14

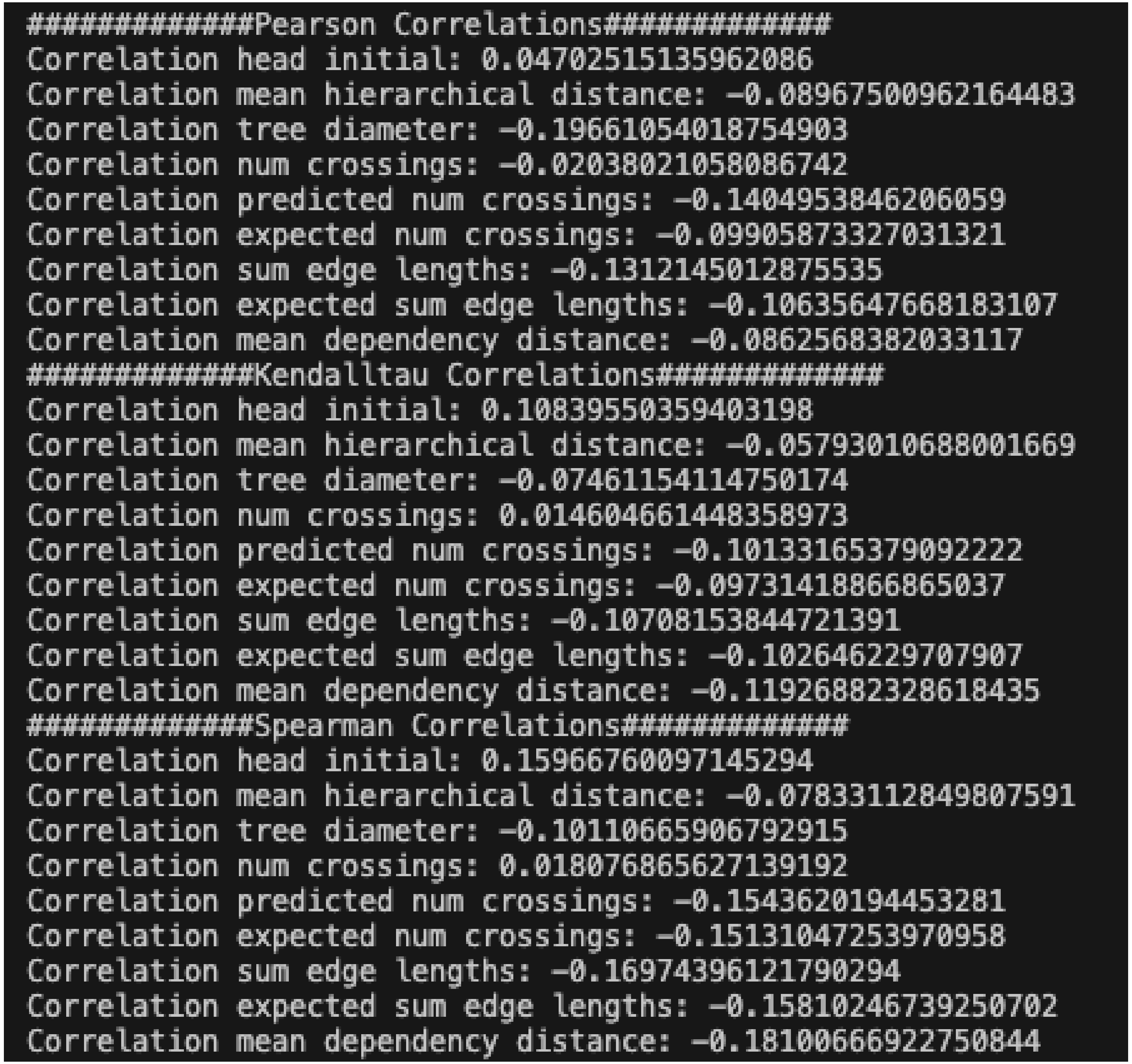
Redid all the correlations with a more robust data structure and checking an “obvious” case – number of tokens vs. number of nodes.

Integrated the data in the existing data structure to handle conll files in order to be able to track back results and check for errors.

Obtained correlation 1.0 between nb of tokens and nb of nodes confirming correctness of data / approach to correlation

Tried different correlation formulas: pearson, spearman and kendall

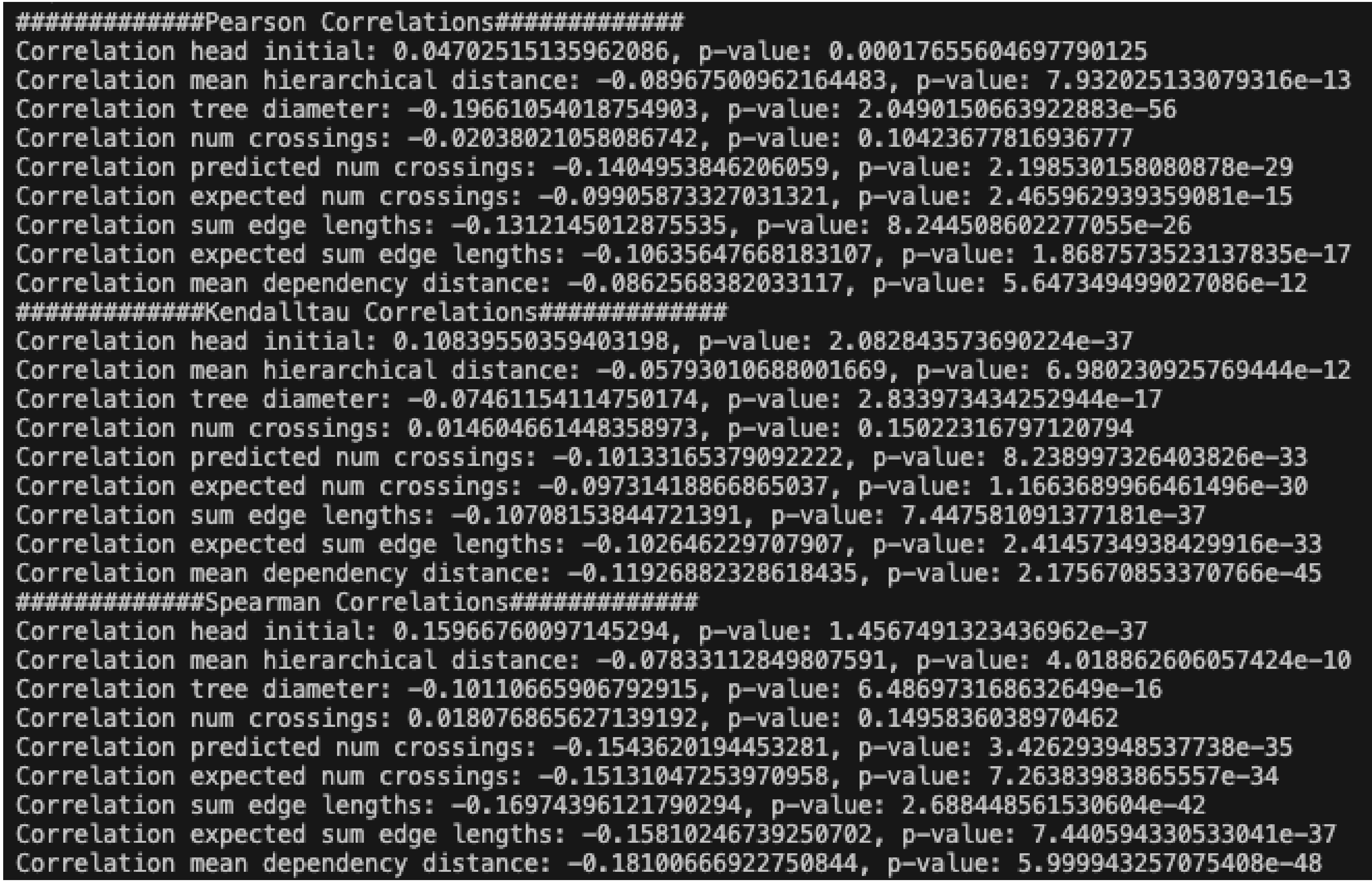
At sentence level:



```
#############Pearson Correlations#############
Correlation head initial: 0.04702515135962086
Correlation mean hierarchical distance: -0.08967500962164483
Correlation tree diameter: -0.19661054018754903
Correlation num crossings: -0.02038021058086742
Correlation predicted num crossings: -0.1404953846206059
Correlation expected num crossings: -0.09905873327031321
Correlation sum edge lengths: -0.1312145012875535
Correlation expected sum edge lengths: -0.10635647668183107
Correlation mean dependency distance: -0.0862568382033117
#############Kendalltau Correlations#############
Correlation head initial: 0.10839550359403198
Correlation mean hierarchical distance: -0.05793010688001669
Correlation tree diameter: -0.07461154114750174
Correlation num crossings: 0.01460466144835897З
Correlation predicted num crossings: -0.1013316537909222
Correlation expected num crossings: -0.0973141886865037
Correlation sum edge lengths: -0.1070815384472139
Correlation expected sum edge lengths: -0.102646229707907
Correlation mean dependency distance: -0.1192688232861843
#############Spearman Correlations#############
Correlation head initial: 0.15966760097145294
Correlation mean hierarchical distance: -0.07833112849807591
Correlation tree diameter: -0.10110665906792915
Correlation num crossings: 0.01807686562713919
Correlation predicted num crossings: -0.154362019445328
Correlation expected num crossings: -0.1513104725397095
Correlation sum edge lengths: -0.16974396121790294
Correlation expected sum edge lengths: -0.15810246739250702
Correlation mean dependency distance: -0.1810066692275084
```

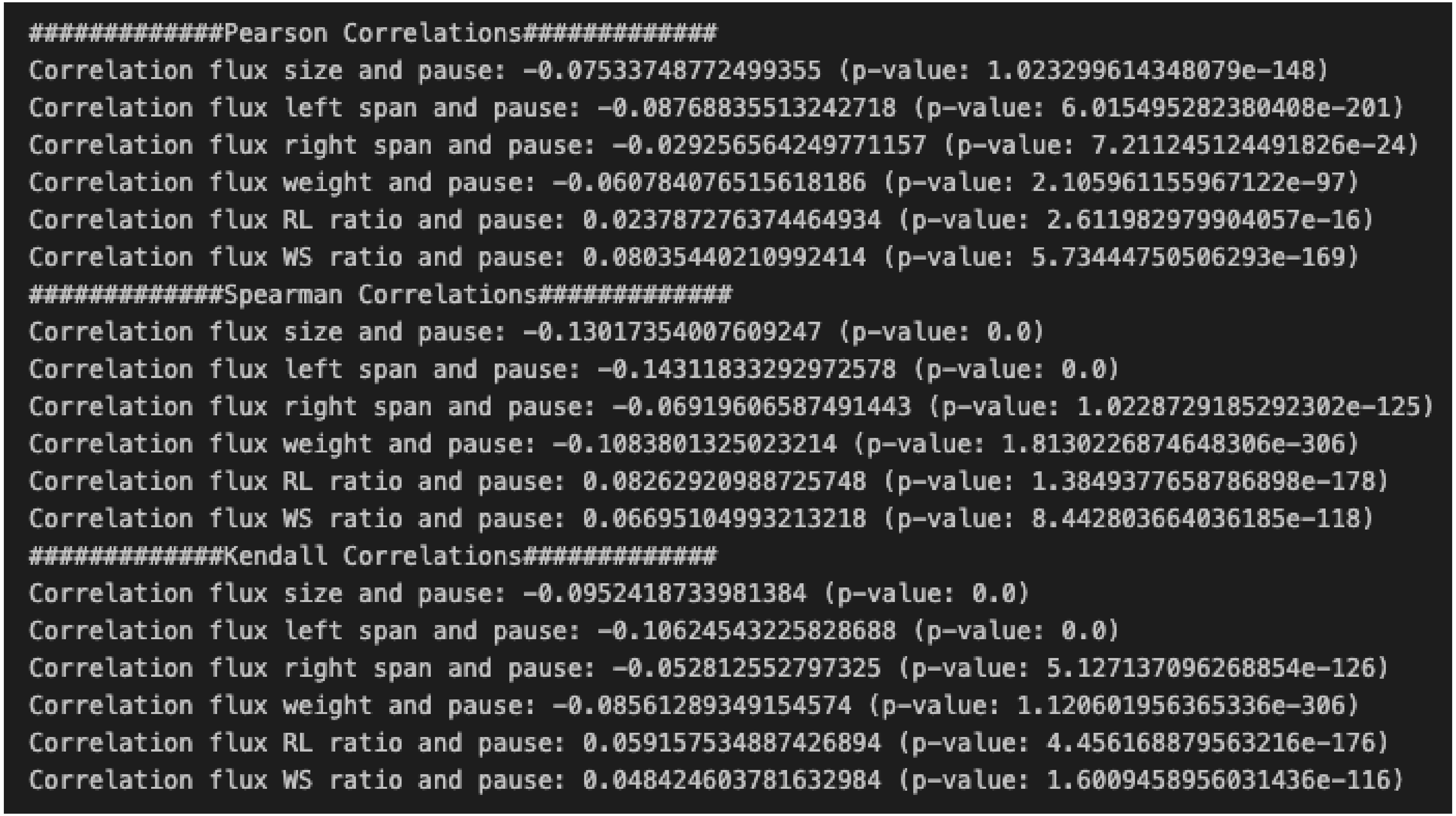
Also added p-value to check significance of results

At sentence level:



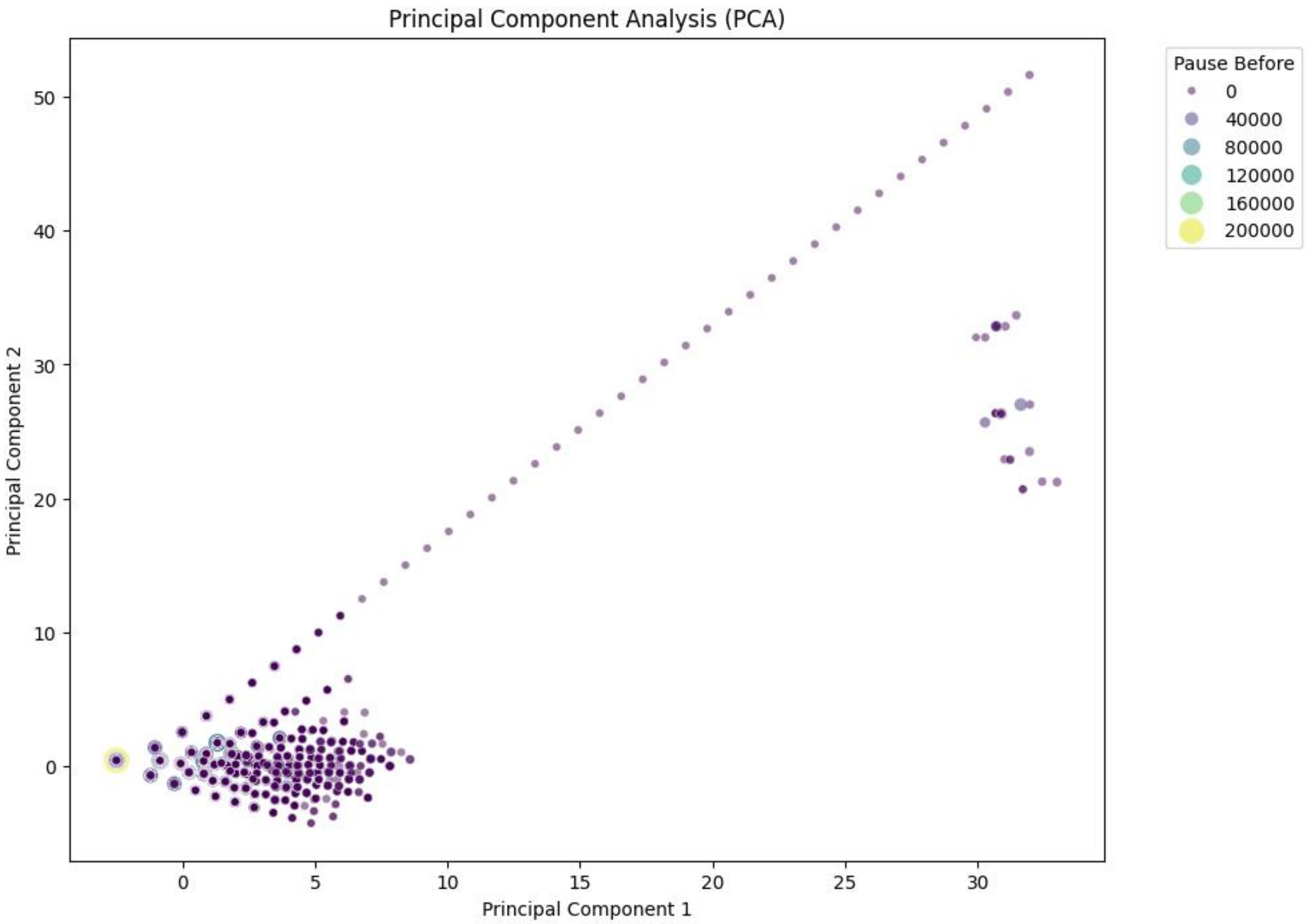
```
#############Pearson Correlations#############
Correlation head initial: 0.04702515135962086, p-value: 0.00017655604697790125
Correlation mean hierarchical distance: -0.08967500962164483, p-value: 7.932025133079316e-13
Correlation tree diameter: -0.19661054018754903, p-value: 2.0490150663922883e-56
Correlation num crossings: -0.02038021058086742, p-value: 0.10423677816936777
Correlation predicted num crossings: -0.1404953846206059, p-value: 2.198530158080878e-29
Correlation expected num crossings: -0.09905873327031321, p-value: 2.465962939359081e-15
Correlation sum edge lengths: -0.1312145012875535, p-value: 8.244508602277055e-26
Correlation expected sum edge lengths: -0.10635647668183107, p-value: 1.8687573523137835e-17
Correlation mean dependency distance: -0.0862568382033117, p-value: 5.647349499027086e-12
#############Kendalltau Correlations#############
Correlation head initial: 0.10839550359403198, p-value: 2.082843573690224e-37
Correlation mean hierarchical distance: -0.05793010688001669, p-value: 6.980230925769444e-12
Correlation tree diameter: -0.07461154114750174, p-value: 2.833973434252944e-17
Correlation num crossings: 0.014604661448358973, p-value: 0.15022316797120794
Correlation predicted num crossings: -0.10133165379092222, p-value: 8.238997326403826e-33
Correlation expected num crossings: -0.09731418866865037, p-value: 1.1663689966461496e-30
Correlation sum edge lengths: -0.10708153844721391, p-value: 7.447581091377181e-37
Correlation expected sum edge lengths: -0.102646229707907, p-value: 2.4145734938429916e-33
Correlation mean dependency distance: -0.11926882328618435, p-value: 2.175670853370766e-45
#############Spearman Correlations#############
Correlation head initial: 0.15966760097145294, p-value: 1.4567491323436962e-37
Correlation mean hierarchical distance: -0.07833112849807591, p-value: 4.018862606057424e-10
Correlation tree diameter: -0.10110665906792915, p-value: 6.486973168632649e-16
Correlation num crossings: 0.018076865627139192, p-value: 0.1495836038970462
Correlation predicted num crossings: -0.1543620194453281, p-value: 3.426293948537738e-35
Correlation expected num crossings: -0.15131047253970958, p-value: 7.26383983865557e-34
Correlation sum edge lengths: -0.16974396121790294, p-value: 2.688448561530604e-42
Correlation expected sum edge lengths: -0.15810246739250702, p-value: 7.440594330533041e-37
Correlation mean dependency distance: -0.18100666922750844, p-value: 5.999943257075408e-48
```

At token level:

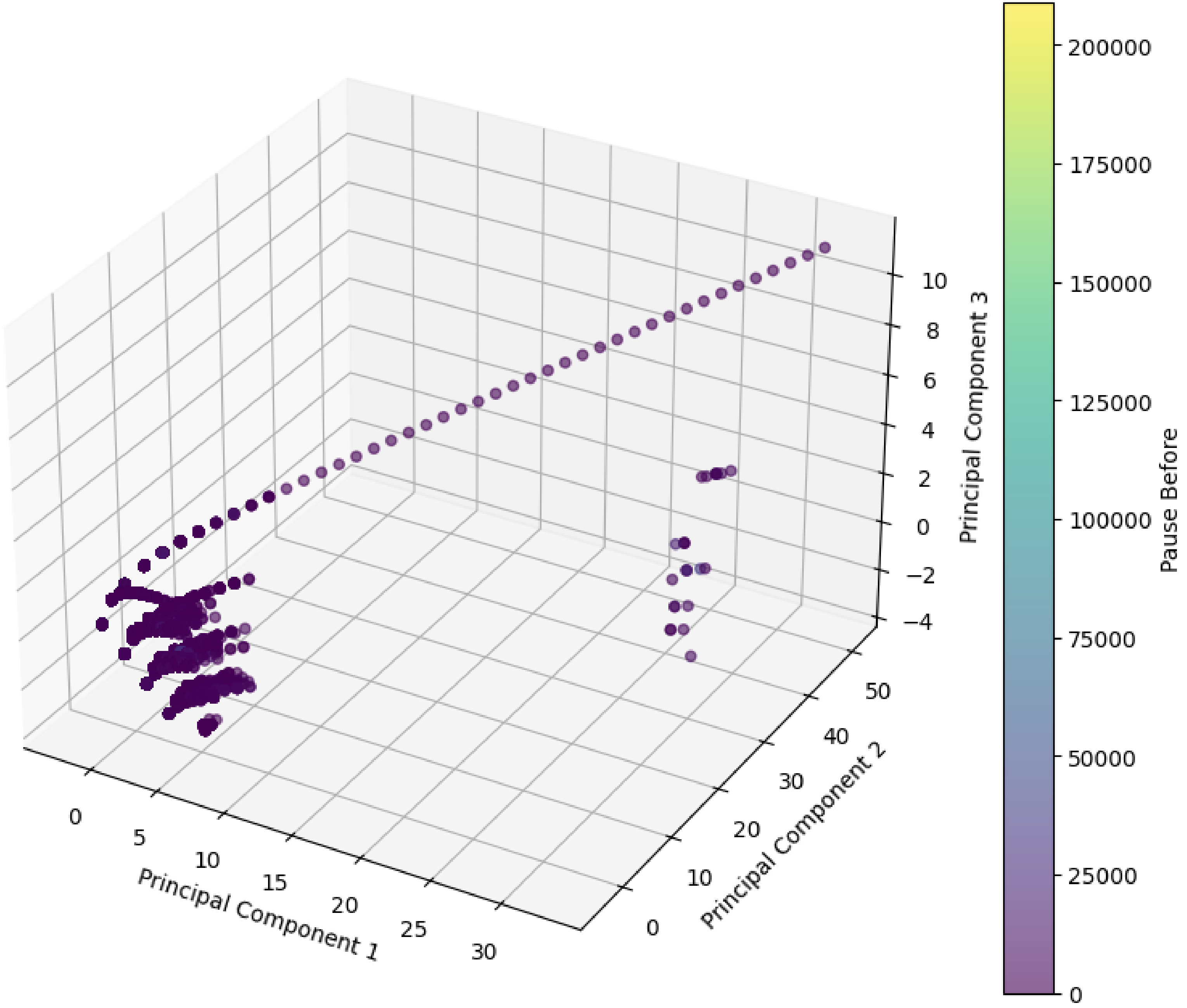


```
#############Pearson Correlations#############
Correlation flux size and pause: -0.07533748772499355 (p-value: 1.023299614348079e-148)
Correlation flux left span and pause: -0.08768835513242718 (p-value: 6.015495282380408e-201)
Correlation flux right span and pause: -0.029256564249771157 (p-value: 7.211245124491826e-24)
Correlation flux weight and pause: -0.060784076515618186 (p-value: 2.105961155967122e-97)
Correlation flux RL ratio and pause: 0.023787276374464934 (p-value: 2.611982979904057e-16)
Correlation flux WS ratio and pause: 0.08035440210992414 (p-value: 5.73444750506293e-169)
#############Spearman Correlations#############
Correlation flux size and pause: -0.13017354007609247 (p-value: 0.0)
Correlation flux left span and pause: -0.14311833292972578 (p-value: 0.0)
Correlation flux right span and pause: -0.06919606587491443 (p-value: 1.0228729185292302e-125)
Correlation flux weight and pause: -0.1083801325023214 (p-value: 1.8130226874648306e-306)
Correlation flux RL ratio and pause: 0.08262920988725748 (p-value: 1.3849377658786898e-178)
Correlation flux WS ratio and pause: 0.06695104993213218 (p-value: 8.442803664036185e-118)
#############Kendall Correlations#############
Correlation flux size and pause: -0.0952418733981384 (p-value: 0.0)
Correlation flux left span and pause: -0.10624543225828688 (p-value: 0.0)
Correlation flux right span and pause: -0.052812552797325 (p-value: 5.127137096268854e-126)
Correlation flux weight and pause: -0.08561289349154574 (p-value: 1.120601956365336e-306)
Correlation flux RL ratio and pause: 0.059157534887426894 (p-value: 4.456168879563216e-176)
Correlation flux WS ratio and pause: 0.048424603781632984 (p-value: 1.6009458956031436e-116)
```

Principal component analysis of fluxes with 2 principal components (variance ratio of 0.77)

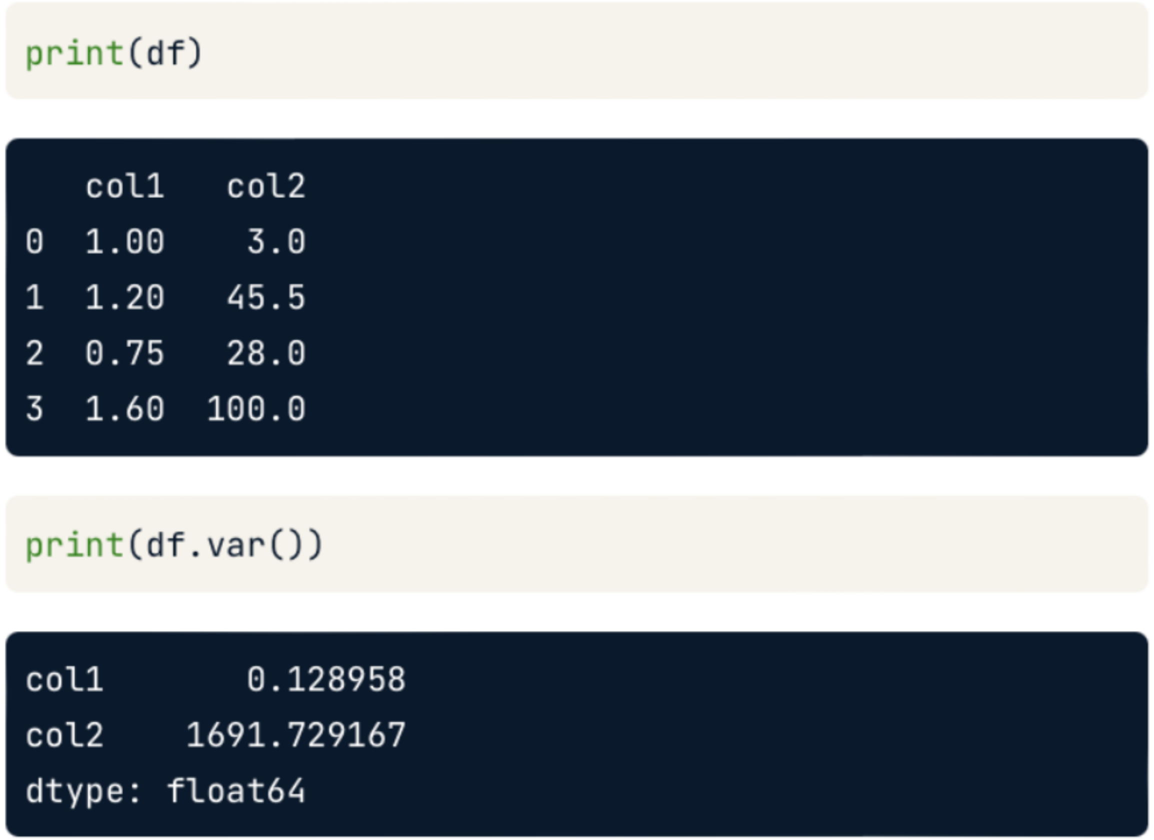


Principal component analysis of fluxes with 3 principal components (variance ratio of 0.96)



Log normalization

- The purpose is to bring variances closer together
- If we check the variance of the columns, we can see that column 2 has a significantly higher variance than column 1, which makes it a clear candidate for log normalization.



```
print(df)
```

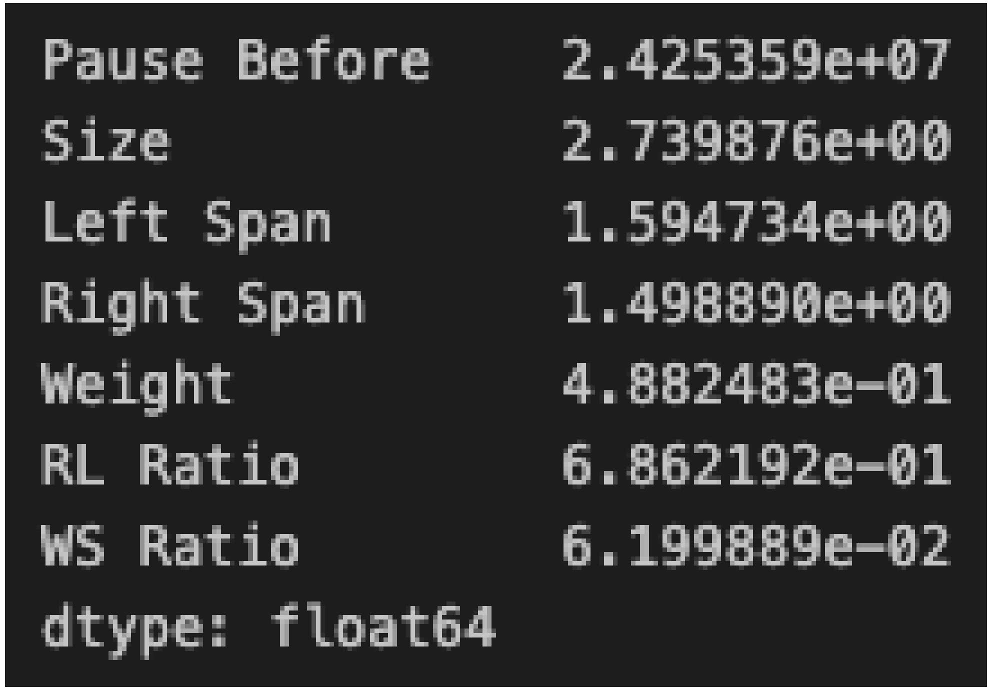
```
   col1   col2
0  1.00    3.0
1  1.20   45.5
2  0.75   28.0
3  1.60  100.0
```

```
print(df.var())
```

```
col1       0.128958
col2    1691.729167
dtype: float64
```

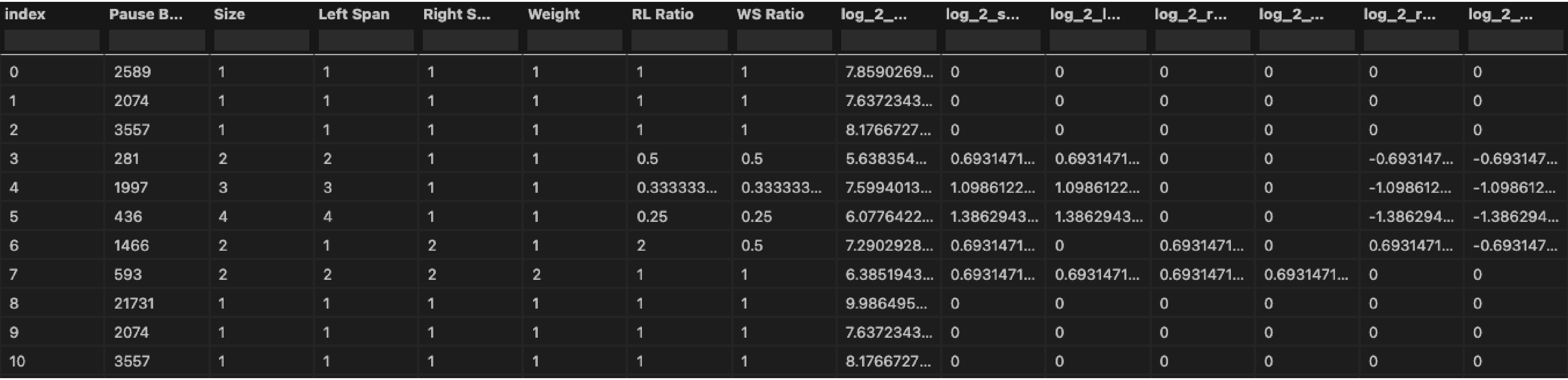
But in our case…

- Variance is very small in all cases so there isn’t a column that it would make sense to do log normalization on.

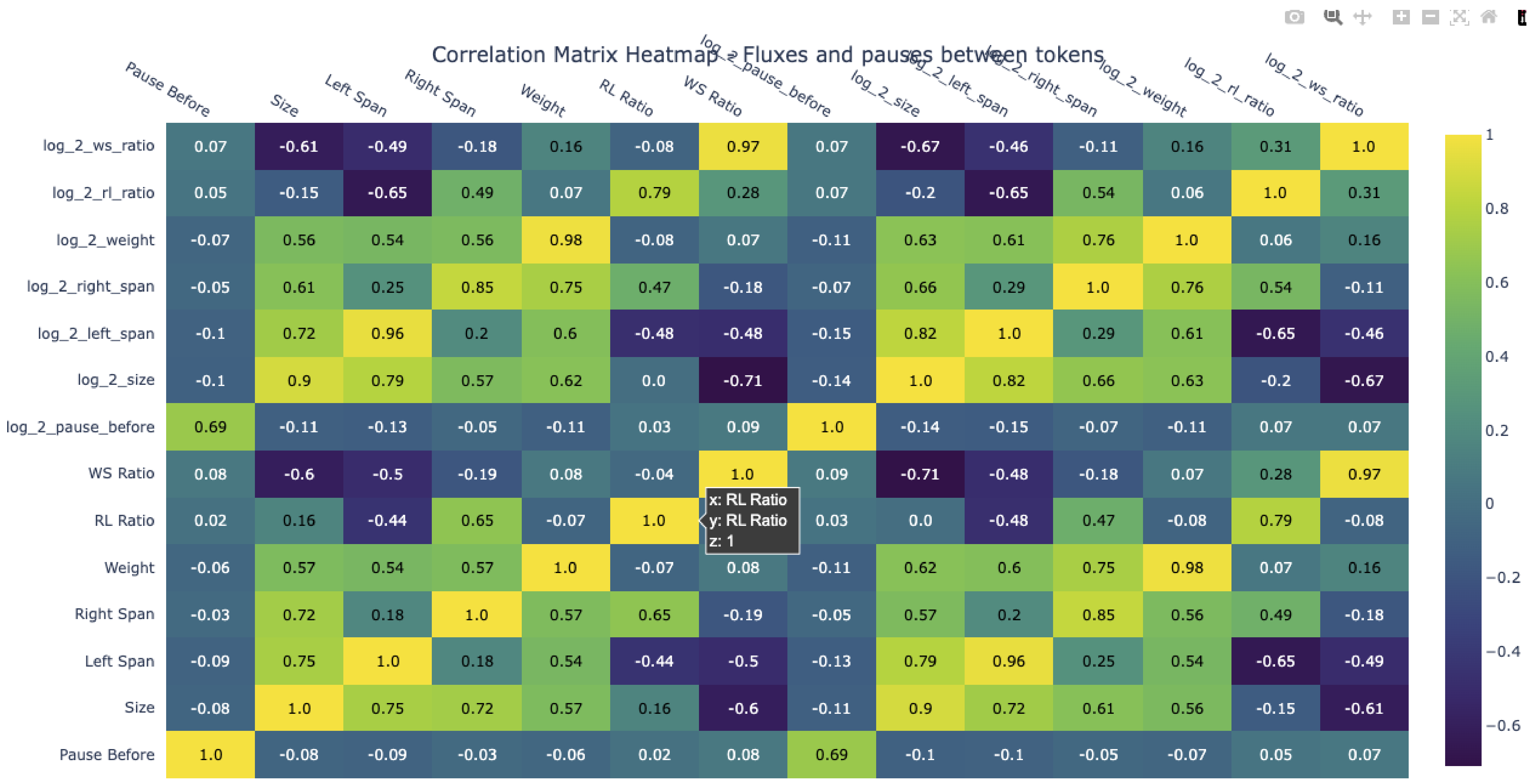


```
Pause Before    2.425359e+07
Size            2.739876e+00
Left Span       1.594734e+00
Right Span      1.498890e+00
Weight          4.882483e-01
RL Ratio        6.862192e-01
WS Ratio        6.199889e-02
dtype: float64
```

If we apply log normalization to all columns



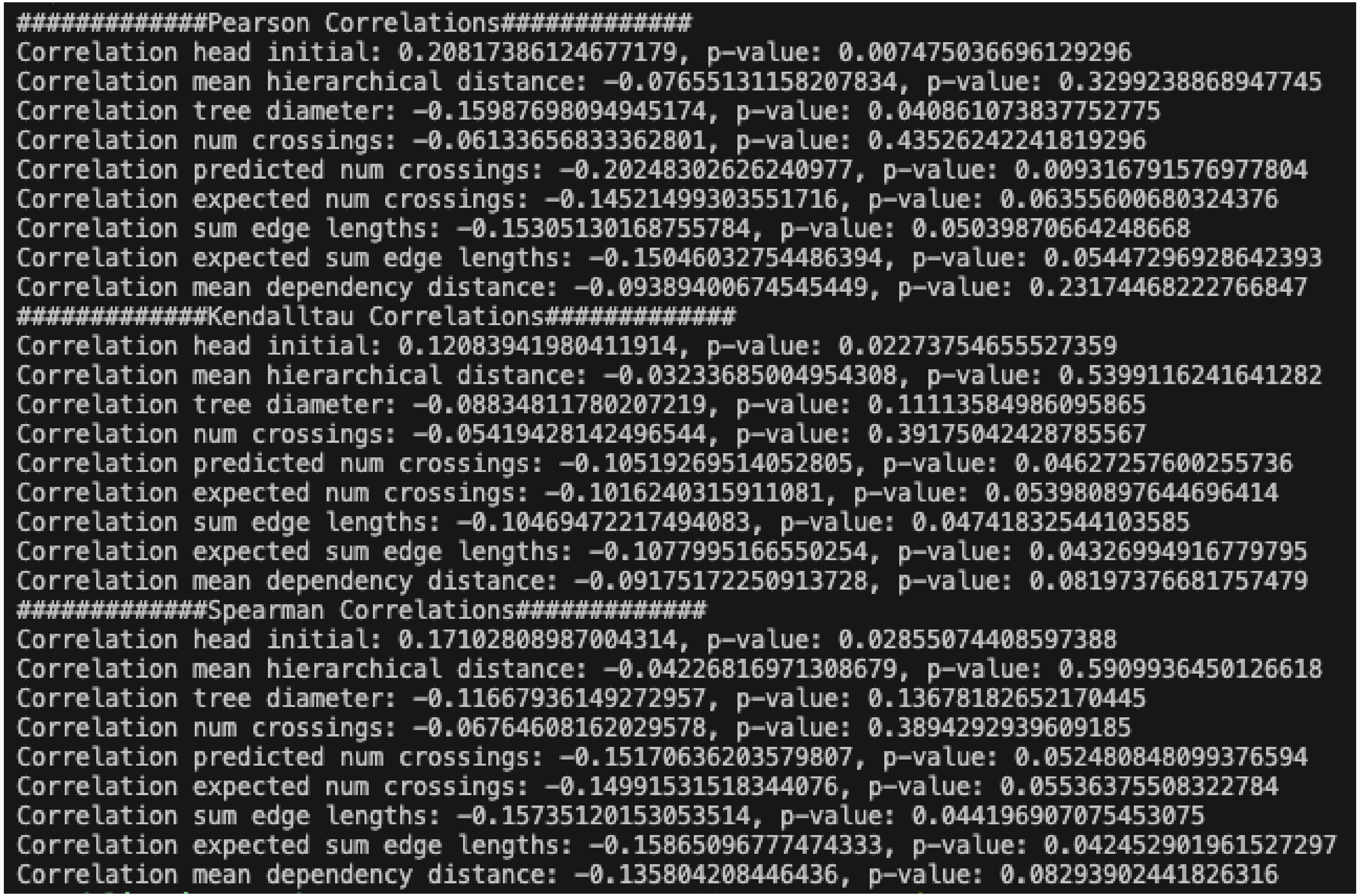
index	Pause B...	Size	Left Span	Right S...	Weight	RL Ratio	WS Ratio	log_2_...	log_2_s...	log_2_l...	log_2_r...	log_2_...	log_2_r...	log_2_...
0	2589	1	1	1	1	1	1	7.8590269...	0	0	0	0	0	0
1	2074	1	1	1	1	1	1	7.6372343...	0	0	0	0	0	0
2	3557	1	1	1	1	1	1	8.1766727...	0	0	0	0	0	0
3	281	2	2	1	1	0.5	0.5	5.638354...	0.6931471...	0.6931471...	0	0	-0.693147...	-0.693147...
4	1997	3	3	1	1	0.333333...	0.333333...	7.5994013...	1.0986122...	1.0986122...	0	0	-1.098612...	-1.098612...
5	436	4	4	1	1	0.25	0.25	6.0776422...	1.3862943...	1.3862943...	0	0	-1.386294...	-1.386294...
6	1466	2	1	2	1	2	0.5	7.2902928...	0.6931471...	0	0.6931471...	0	0.6931471...	-0.693147...
7	593	2	2	2	2	1	1	6.3851943...	0.6931471...	0.6931471...	0.6931471...	0.6931471...	0	0
8	21731	1	1	1	1	1	1	9.986495...	0	0	0	0	0	0
9	2074	1	1	1	1	1	1	7.6372343...	0	0	0	0	0	0
10	3557	1	1	1	1	1	1	8.1766727...	0	0	0	0	0	0



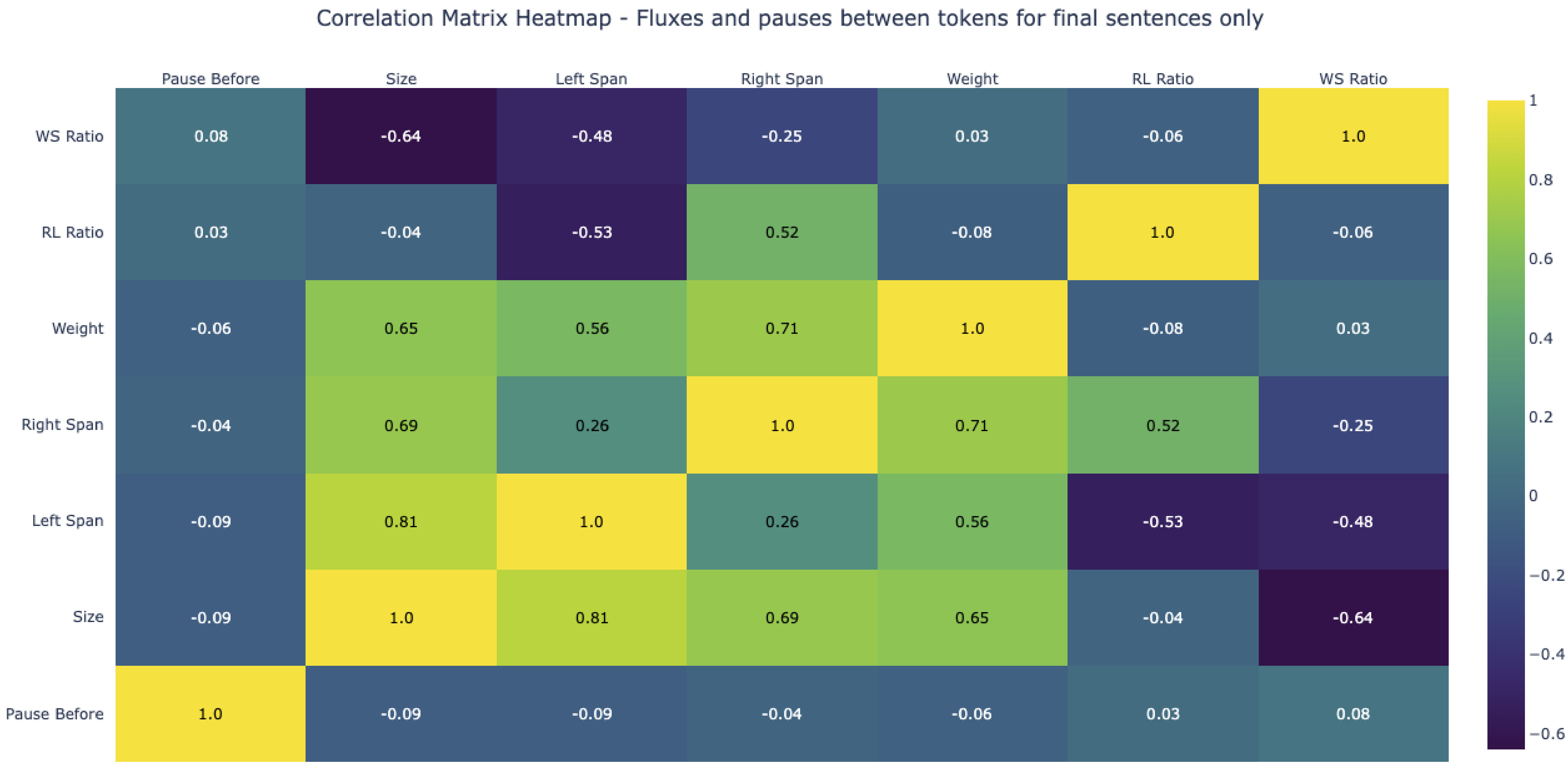
Correlation Matrix Heatmap - Fluxes and pauses between tokens

	Pause Before	Size	Left Span	Right Span	Weight	RL Ratio	WS Ratio	log_2_pause_before	log_2_size	log_2_left_span	log_2_right_span	log_2_weight	log_2_rl_ratio	log_2_ws_ratio
log_2_ws_ratio	0.07	-0.61	-0.49	-0.18	0.16	-0.08	0.97	0.07	-0.67	-0.46	-0.11	0.16	0.31	1.0
log_2_rl_ratio	0.05	-0.15	-0.65	0.49	0.07	0.79	0.28	0.07	-0.2	-0.65	0.54	0.06	1.0	0.31
log_2_weight	-0.07	0.56	0.54	0.56	0.98	-0.08	0.07	-0.11	0.63	0.61	0.76	1.0	0.06	0.16
log_2_right_span	-0.05	0.61	0.25	0.85	0.75	0.47	-0.18	-0.07	0.66	0.29	1.0	0.76	0.54	-0.11
log_2_left_span	-0.1	0.72	0.96	0.2	0.6	-0.48	-0.48	-0.15	0.82	1.0	0.29	0.61	-0.65	-0.46
log_2_size	-0.1	0.9	0.79	0.57	0.62	0.0	-0.71	-0.14	1.0	0.82	0.66	0.63	-0.2	-0.67
log_2_pause_before	0.69	-0.11	-0.13	-0.05	-0.11	0.03	0.09	1.0	-0.14	-0.15	-0.07	-0.11	0.07	0.07
WS Ratio	0.08	-0.6	-0.5	-0.19	0.08	-0.04	1.0	0.09	-0.71	-0.48	-0.18	0.07	0.28	0.97
RL Ratio	0.02	0.16	-0.44	0.65	-0.07	1.0	[illegible]	0.03	0.0	-0.48	0.47	-0.08	0.79	-0.08
Weight	-0.06	0.57	0.54	0.57	1.0	-0.07	0.08	-0.11	0.62	0.6	0.75	0.98	0.07	0.16
Right Span	-0.03	0.72	0.18	1.0	0.57	0.65	-0.19	-0.05	0.57	0.2	0.85	0.56	0.49	-0.18
Left Span	-0.09	0.75	1.0	0.18	0.54	-0.44	-0.5	-0.13	0.79	0.96	0.25	0.54	-0.65	-0.49
Size	-0.08	1.0	0.75	0.72	0.57	0.16	-0.6	-0.11	0.9	0.72	0.61	0.56	-0.15	-0.61
Pause Before	1.0	-0.08	-0.09	-0.03	-0.06	0.02	0.08	0.69	-0.1	-0.1	-0.05	-0.07	0.05	0.07

Also tried the correlation only taking into account the final sentence(s) in each file



```
#############Pearson Correlations#############
Correlation head initial: 0.20817386124677179, p-value: 0.007475036696129296
Correlation mean hierarchical distance: -0.07655131158207834, p-value: 0.3299238868947745
Correlation tree diameter: -0.15987698094945174, p-value: 0.040861073837752775
Correlation num crossings: -0.06133656833362801, p-value: 0.43526242241819296
Correlation predicted num crossings: -0.20248302626240977, p-value: 0.009316791576977804
Correlation expected num crossings: -0.14521499303551716, p-value: 0.06355600680324376
Correlation sum edge lengths: -0.15305130168755784, p-value: 0.05039870664248668
Correlation expected sum edge lengths: -0.15046032754486394, p-value: 0.05447296928642393
Correlation mean dependency distance: -0.09389400674545449, p-value: 0.23174468222766847
#############Kendalltau Correlations#############
Correlation head initial: 0.12083941980411914, p-value: 0.02273754655527359
Correlation mean hierarchical distance: -0.03233685004954308, p-value: 0.5399116241641282
Correlation tree diameter: -0.08834811780207219, p-value: 0.11113584986095865
Correlation num crossings: -0.05419428142496544, p-value: 0.39175042428785567
Correlation predicted num crossings: -0.10519269514052805, p-value: 0.04627257600255736
Correlation expected num crossings: -0.1016240315911081, p-value: 0.053980897644696414
Correlation sum edge lengths: -0.10469472217494083, p-value: 0.04741832544103585
Correlation expected sum edge lengths: -0.1077995166550254, p-value: 0.04326994916779795
Correlation mean dependency distance: -0.09175172250913728, p-value: 0.08197376681757479
#############Spearman Correlations#############
Correlation head initial: 0.17102808987004314, p-value: 0.02855074408597388
Correlation mean hierarchical distance: -0.04226816971308679, p-value: 0.5909936450126618
Correlation tree diameter: -0.11667936149272957, p-value: 0.13678182652170445
Correlation num crossings: -0.06764608162029578, p-value: 0.3894292939609185
Correlation predicted num crossings: -0.15170636203579807, p-value: 0.052480848099376594
Correlation expected num crossings: -0.14991531518344076, p-value: 0.05536375508322784
Correlation sum edge lengths: -0.15735120153053514, p-value: 0.044196907075453075
Correlation expected sum edge lengths: -0.15865096777474333, p-value: 0.042452901961527297
Correlation mean dependency distance: -0.135804208446436, p-value: 0.08293902441826316
```



Correlation Matrix Heatmap - Fluxes and pauses between tokens for final sentences only

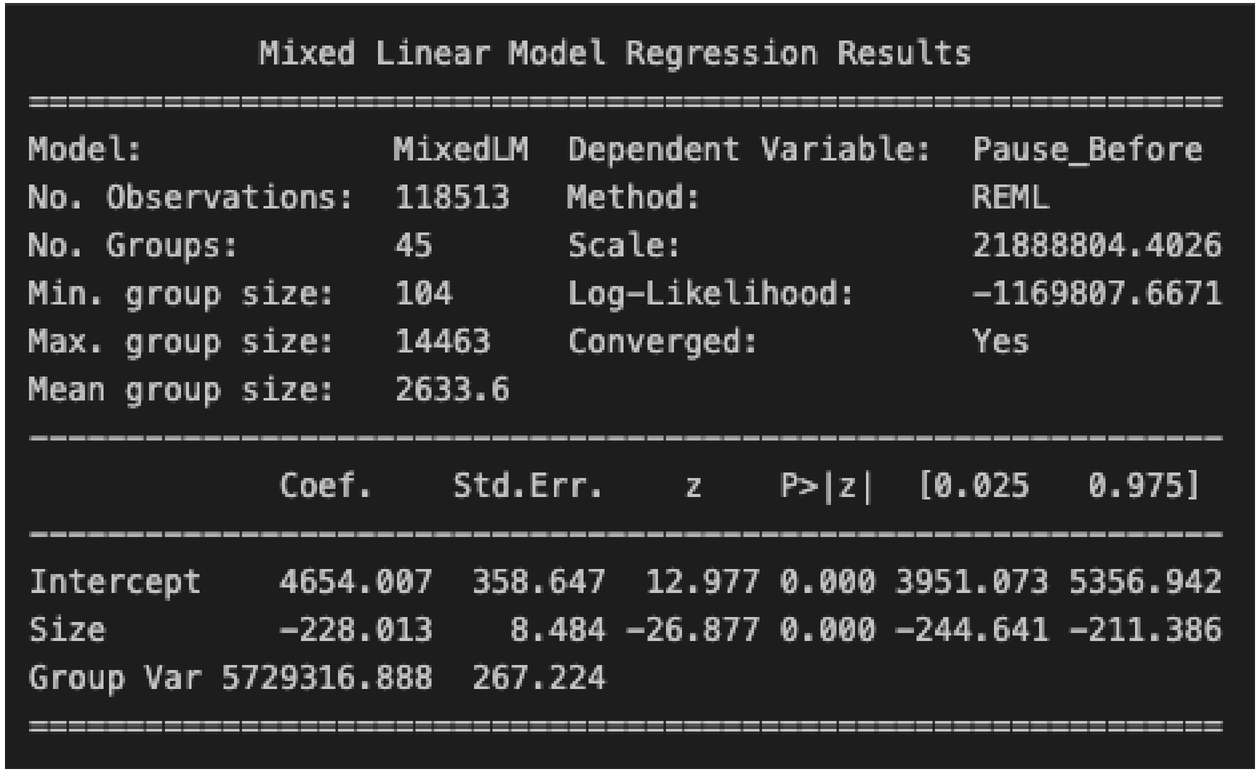
	Pause Before	Size	Left Span	Right Span	Weight	RL Ratio	WS Ratio
WS Ratio	0.08	-0.64	-0.48	-0.25	0.03	-0.06	1.0
RL Ratio	0.03	-0.04	-0.53	0.52	-0.08	1.0	-0.06
Weight	-0.06	0.65	0.56	0.71	1.0	-0.08	0.03
Right Span	-0.04	0.69	0.26	1.0	0.71	0.52	-0.25
Left Span	-0.09	0.81	1.0	0.26	0.56	-0.53	-0.48
Size	-0.09	1.0	0.81	0.69	0.65	-0.04	-0.64
Pause Before	1.0	-0.09	-0.09	-0.04	-0.06	0.03	0.08

Mixed effects analysis

- Also known as a hierarchical linear model, which is useful when you want to account for both fixed effects and random effects in your data.
- In mixed-effects models:
- **Fixed effects** are the main effects you're interested in. For example, in our case, the fixed effect might be the size of the flux.
- **Random effects** account for variations between individual subjects or groups.

When looking at the pause and size of the flux:

Interpretation:
- Intercept (4654.007): This is the estimated average value of Pause_Before when Size is zero. The positive coefficient indicates that, on average, the Pause_Before is 4654.007 ms when Size is zero. The p-value (0.000) indicates that this coefficient is statistically significant.
- Size (-228.013): This coefficient represents the change in Pause_Before for a one-unit increase in Size. The negative value (-228.013) suggests that as Size increases, Pause_Before decreases. The p-value (0.000) indicates that this relationship is statistically significant.
- Group Variance (5,729,316.888): This value represents the variance of the random intercepts for the groups (Person). A higher variance indicates more variability in Pause_Before between different persons.



```
            Mixed Linear Model Regression Results
===============================================================
Model:              MixedLM  Dependent Variable:  Pause_Before
No. Observations:   118513   Method:              REML
No. Groups:         45       Scale:               21888804.4026
Min. group size:    104      Log-Likelihood:      -1169807.6671
Max. group size:    14463    Converged:           Yes
Mean group size:    2633.6
---------------------------------------------------------------
             Coef.    Std.Err.    z    P>|z|  [0.025   0.975]
---------------------------------------------------------------
Intercept    4654.007  358.647  12.977 0.000 3951.073 5356.942
Size         -228.013    8.484 -26.877 0.000 -244.641 -211.386
Group Var 5729316.888  267.224
===============================================================
```

Next steps:

- **Model Validation**: The model should be validated by checking residuals and performing diagnostic checks to ensure the assumptions of the mixed-effects model are met.
- **Further Exploration**: Explore the effects of the other measures on Pause_Before by including them in the model.

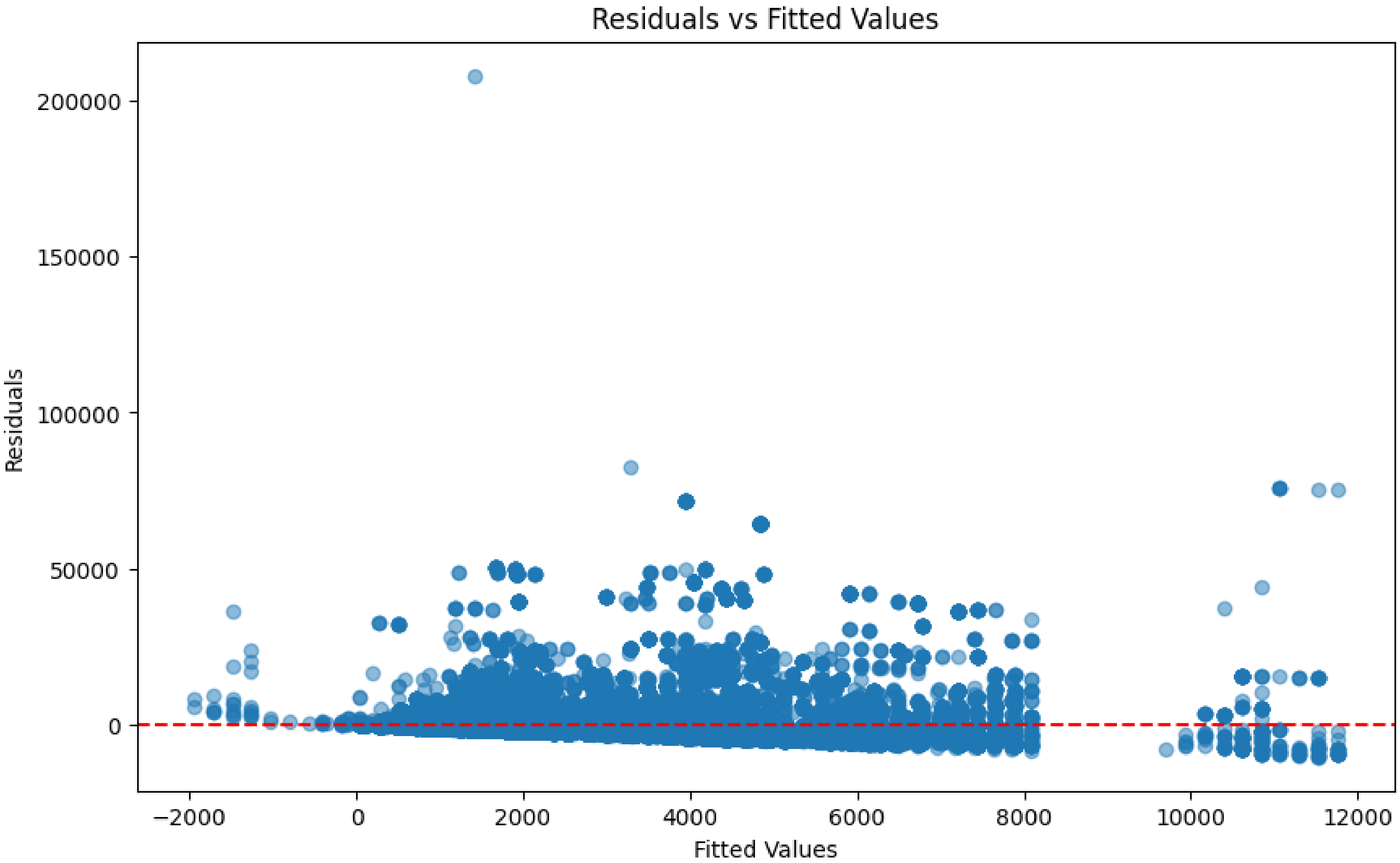
Validating the model

- Validating the model involves checking the residuals and ensuring that the assumptions of the mixed-effects model are met. Here are the key steps you can take for model validation:

1. **Plotting Residuals**: Check for normality and homoscedasticity.
2. **QQ Plot**: Check if the residuals follow a normal distribution.
3. **Random Effects**: Inspect the random effects to ensure they are normally distributed.
4. **Influence of Random Effects**: Check if random effects significantly improve the model fit.

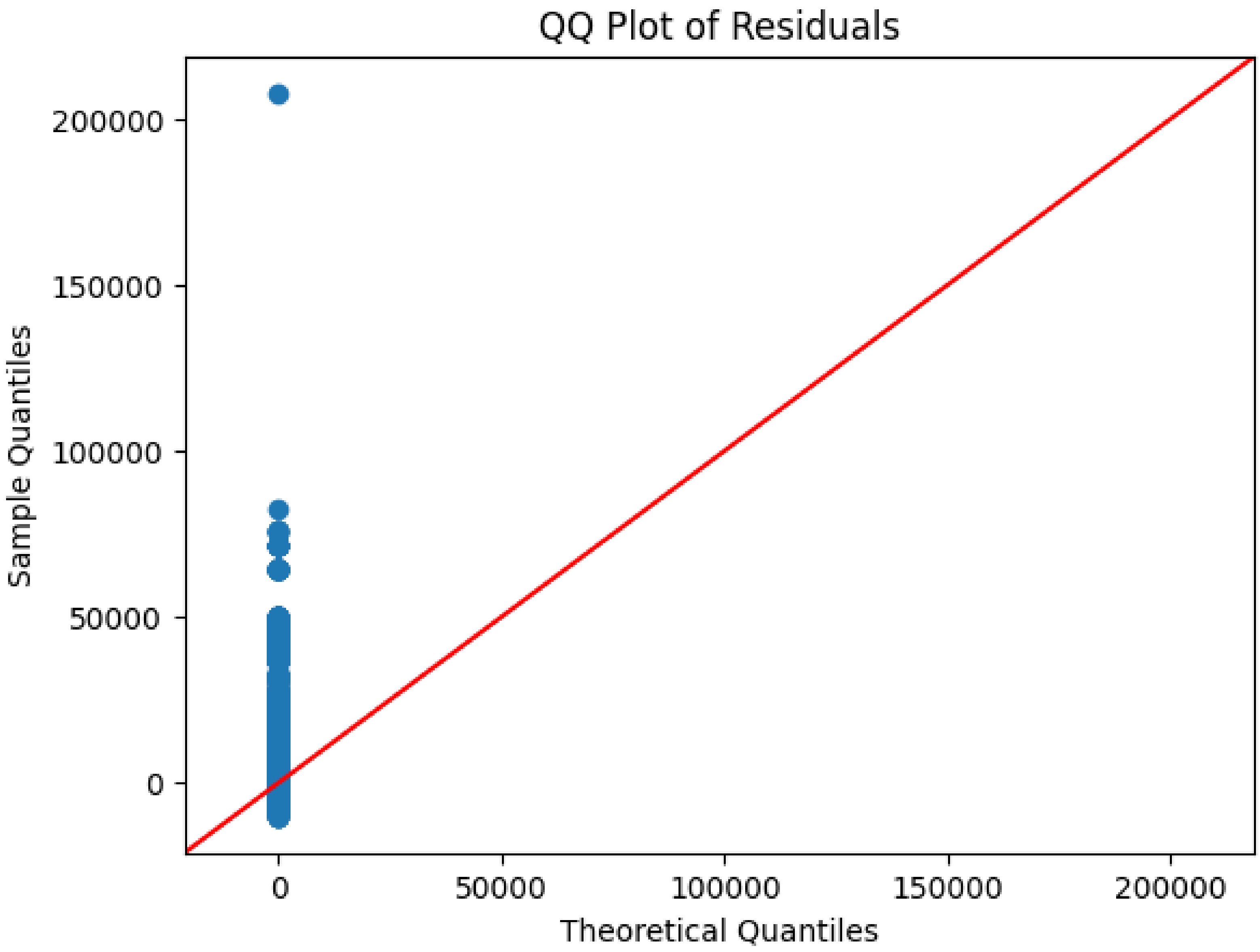
Plotting residuals

This plot helps check for homoscedasticity (constant variance of residuals). Ideally, the residuals should be randomly scattered around zero without any discernible pattern.



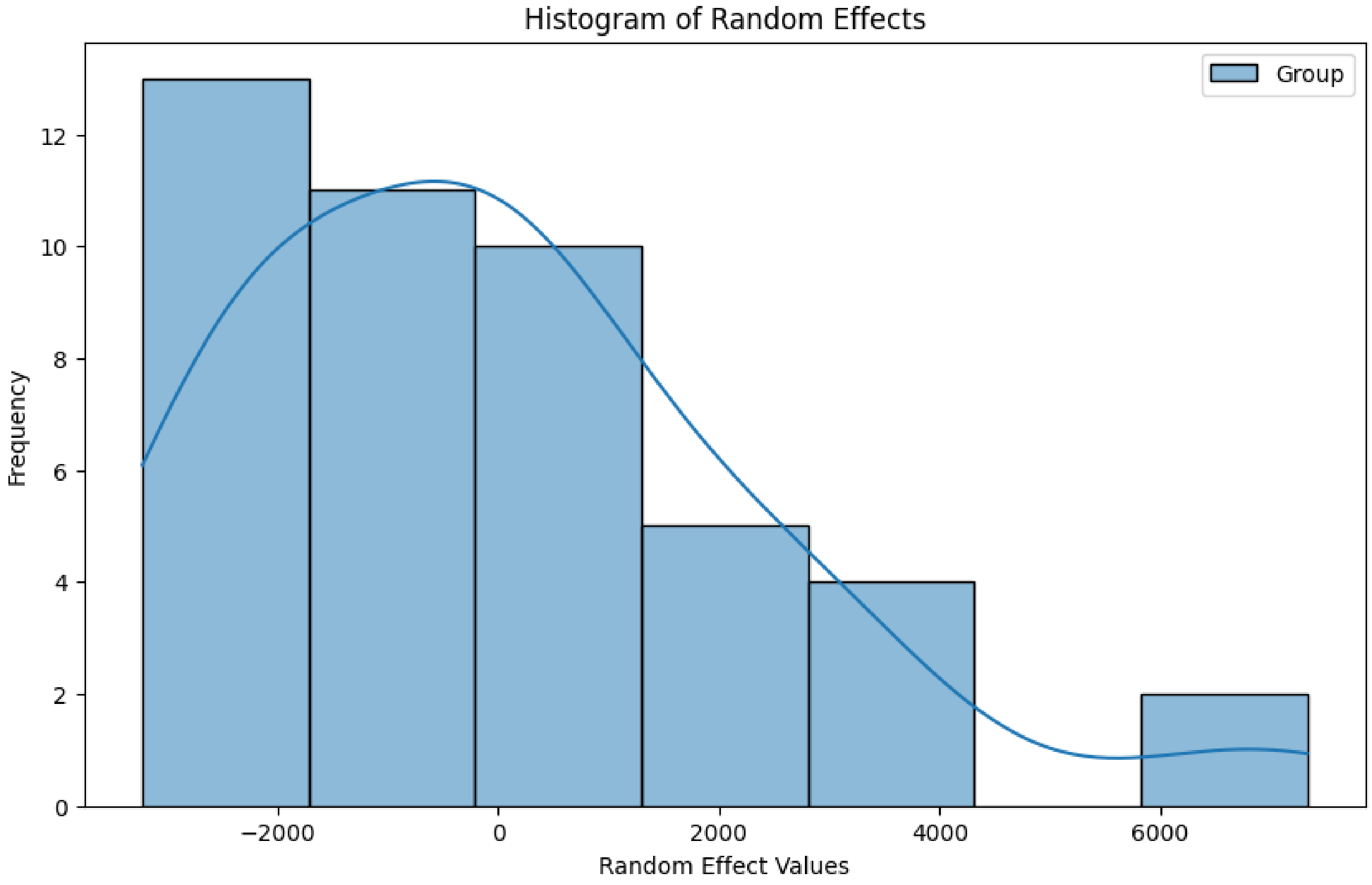
QQ Plot

- If the residuals follow the line closely, they are approximately normally distributed. Deviations from the line suggest departures from normality.



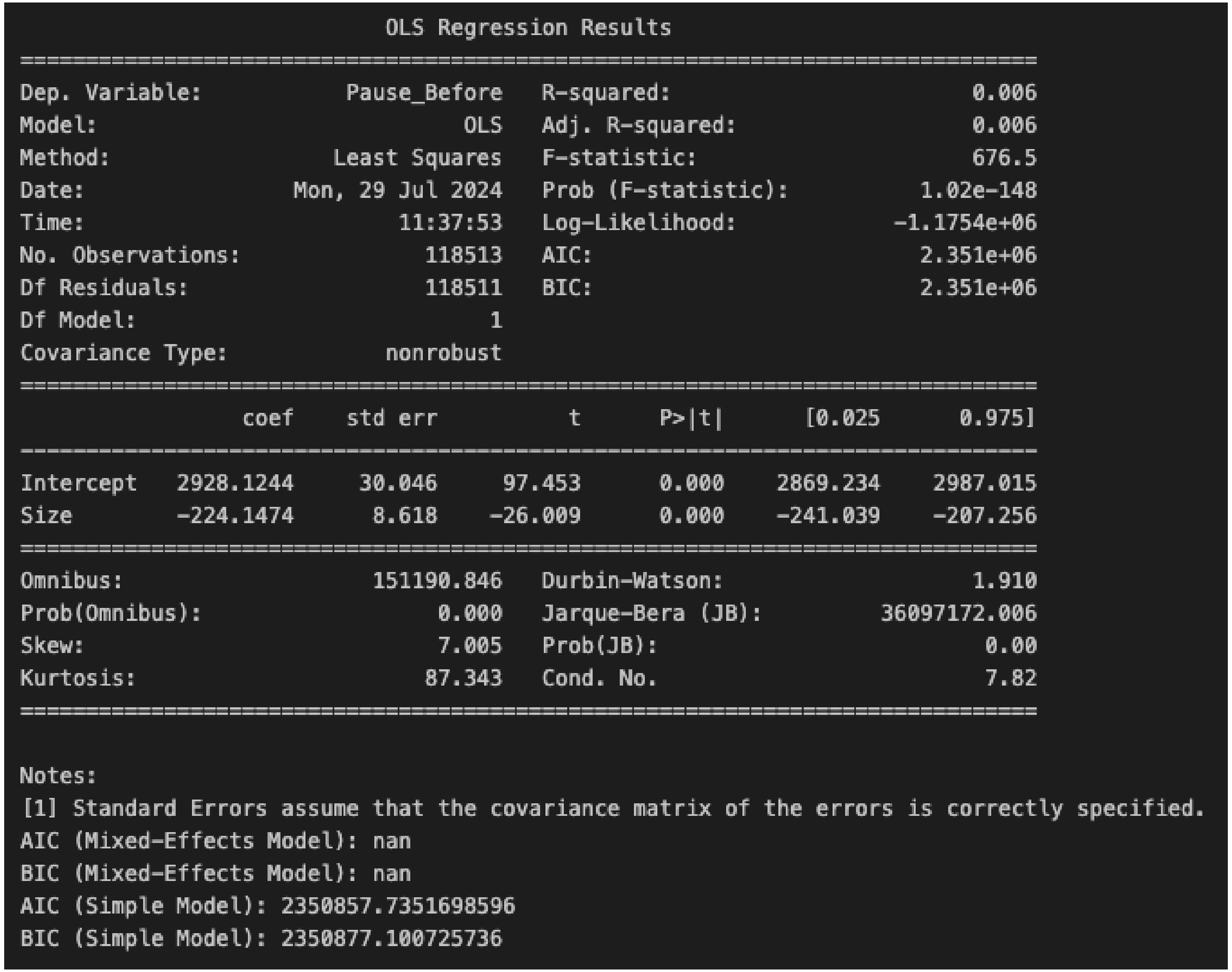
Distribution of Random Effects

The histogram should approximate a normal distribution if the random effects are normally distributed.



Model Comparison with and without Random Effects

- Comparing AIC and BIC values helps determine if including random effects improves the model. Lower AIC and BIC values indicate a better model fit.



```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Pause_Before   R-squared:                       0.006
Model:                             OLS   Adj. R-squared:                  0.006
Method:                  Least Squares   F-statistic:                     676.5
Date:                 Mon, 29 Jul 2024   Prob (F-statistic):          1.02e-148
Time:                         11:37:53   Log-Likelihood:            -1.1754e+06
No. Observations:               118513   AIC:                         2.351e+06
Df Residuals:                   118511   BIC:                         2.351e+06
Df Model:                            1
Covariance Type:             nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   2928.1244     30.046     97.453      0.000    2869.234    2987.015
Size        -224.1474      8.618    -26.009      0.000    -241.039    -207.256
==============================================================================
Omnibus:                151190.846   Durbin-Watson:                   1.910
Prob(Omnibus):               0.000   Jarque-Bera (JB):         36097172.006
Skew:                        7.005   Prob(JB):                         0.00
Kurtosis:                   87.343   Cond. No.                         7.82
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
AIC (Mixed-Effects Model): nan
BIC (Mixed-Effects Model): nan
AIC (Simple Model): 2350857.7351698596
BIC (Simple Model): 2350877.100725736
```

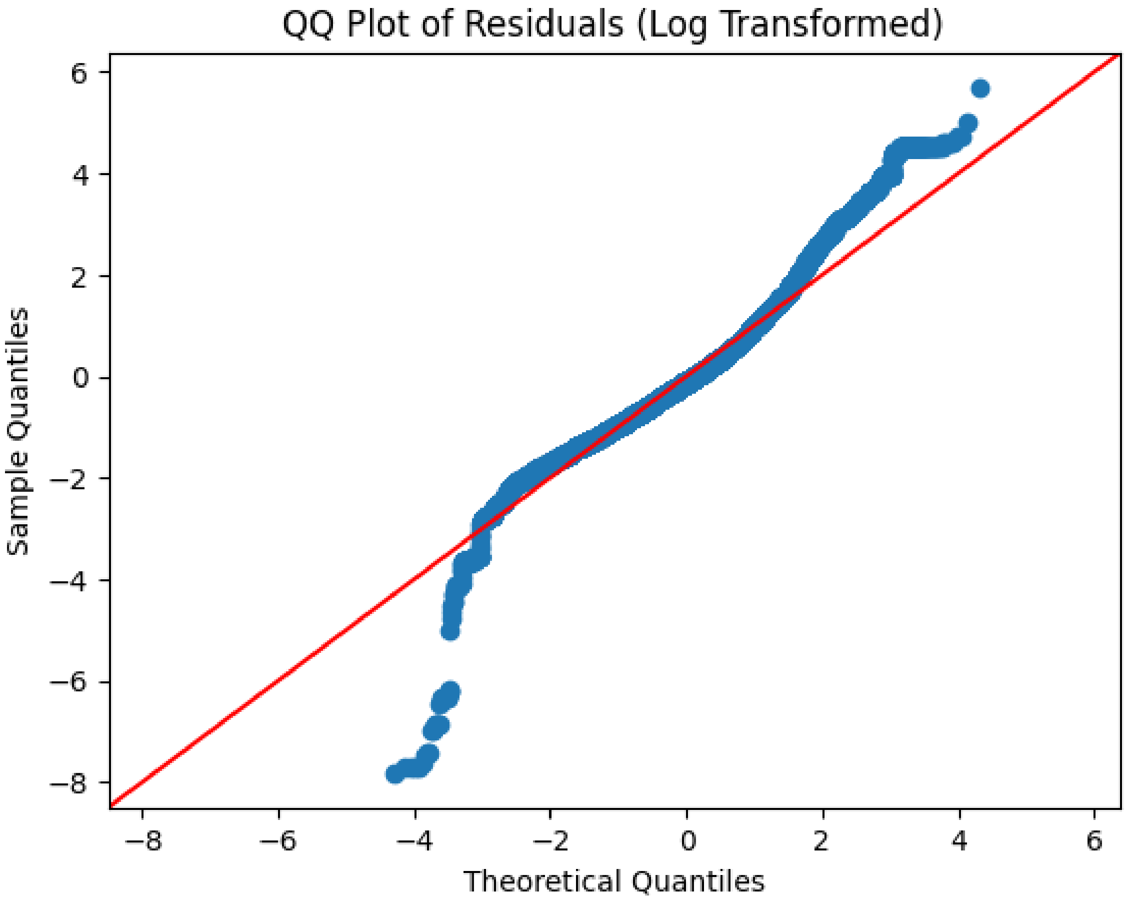
- The QQ plot of the residuals shows a significant deviation from the reference line (the red line), which indicates that the residuals do not follow a normal distribution. Here are some key observations from the plot:

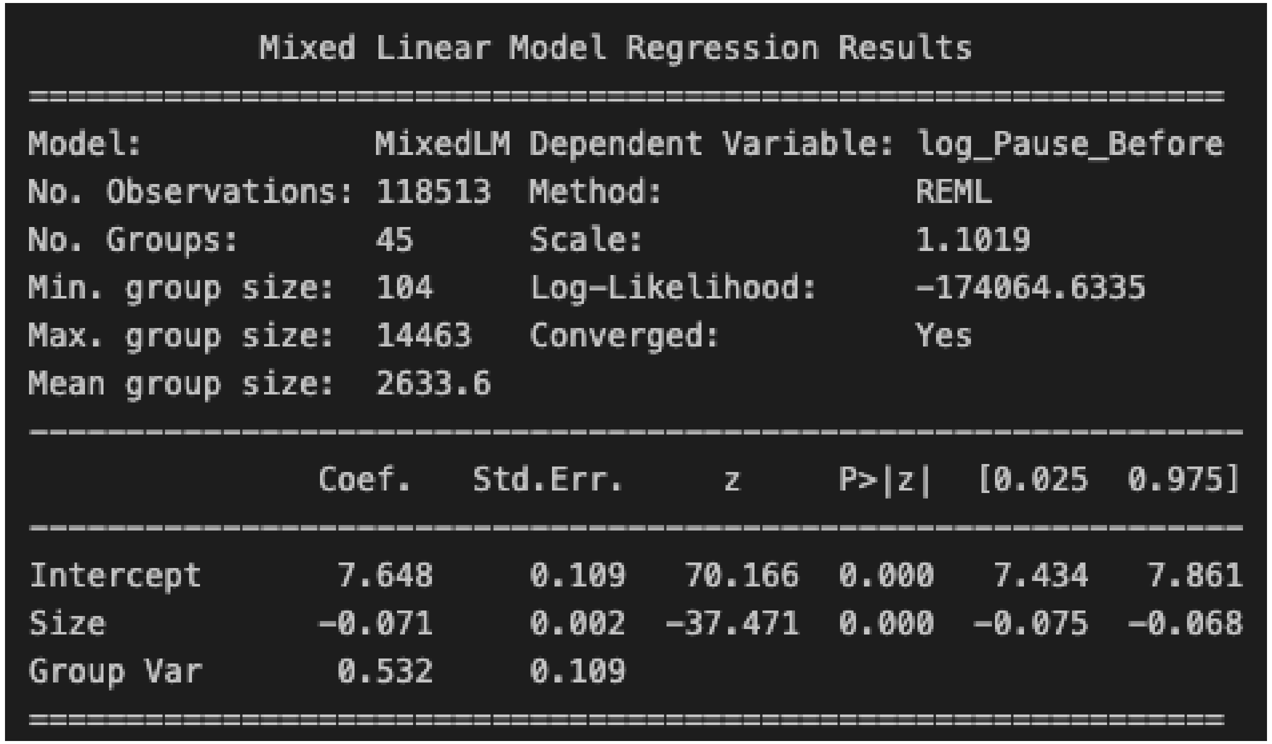
1. **Heavy Tail and Outliers**: The points are clustered near the origin, with some points deviating significantly from the line. This suggests the presence of heavy tails or outliers in the data.
2. **Non-Normality**: The extreme deviation of points from the line indicates that the residuals are not normally distributed. In a QQ plot, if the residuals were normally distributed, the points would fall along the red line.
3. **Potential Skewness**: The concentration of points at the lower end and the spread at the upper end suggest possible skewness in the residuals.

- **Steps to Address Non-Normality**

1. **Transform the Dependent Variable**: Apply transformations such as log, square root, or Box-Cox transformation to the dependent variable to stabilize variance and make the distribution more normal.
2. **Check for Outliers**: Identify and handle outliers, which can disproportionately affect the normality of residuals.
3. **Use Robust Methods**: Consider using robust regression methods that are less sensitive to violations of assumptions.
4. **Model Diagnostics**: Perform further diagnostics to understand the source of non-normality. Check for patterns or trends in residual plots.

So I applied a log transformation to the pause and refit the model



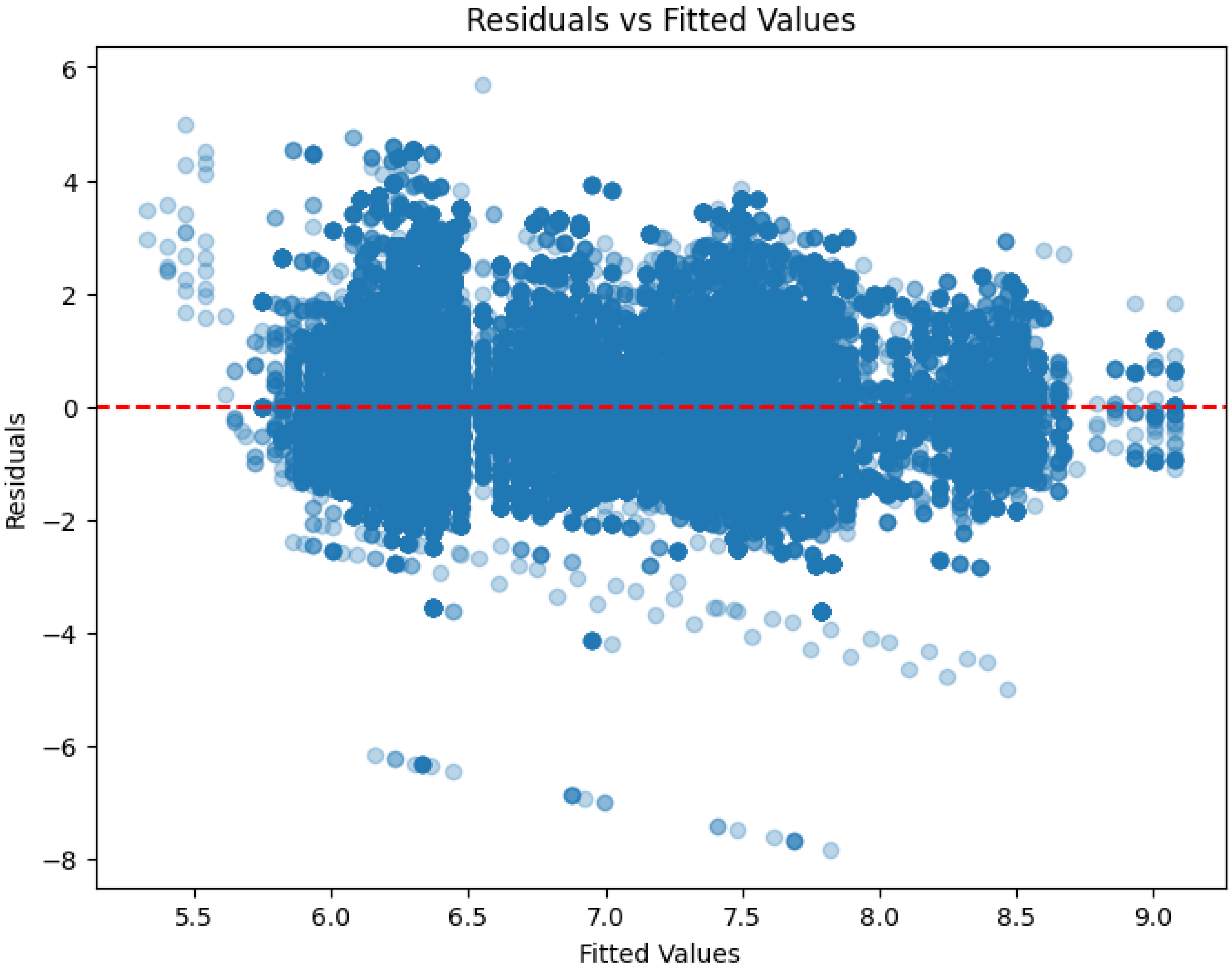


```
                 Mixed Linear Model Regression Results
==============================================================
Model:              MixedLM Dependent Variable: log_Pause_Before
No. Observations: 118513  Method:             REML
No. Groups:         45      Scale:              1.1019
Min. group size:    104     Log-Likelihood:     -174064.6335
Max. group size:    14463   Converged:          Yes
Mean group size:    2633.6
---------------------------------------------------------------
               Coef.   Std.Err.    z     P>|z|  [0.025  0.975]
---------------------------------------------------------------
Intercept      7.648     0.109   70.166  0.000   7.434   7.861
Size          -0.071     0.002  -37.471  0.000  -0.075  -0.068
Group Var      0.532     0.109
==============================================================
```

- Interpretation:
- Intercept (7.648): This is the estimated average value of log_Pause_Before when Size is zero. The positive coefficient indicates that, on average, the log-transformed Pause_Before is 7.648 units when Size is zero. The p-value (0.000) indicates that this coefficient is statistically significant.
- Size (-0.071): This coefficient represents the change in log_Pause_Before for a one-unit increase in Size. The negative value (-0.071) suggests that as Size increases, the log-transformed Pause_Before decreases. The p-value (0.000) indicates that this relationship is statistically significant.
- Group Variance (0.532): This value represents the variance of the random intercepts for the groups (Person). A higher variance indicates more variability in log_Pause_Before between different persons.
- Conclusion:
- The model suggests that log_Pause_Before is significantly influenced by Size, with larger sizes leading to shorter pauses on a logarithmic scale. There is also significant variability in log_Pause_Before across different persons, as indicated by the group variance. The model has converged, indicating that the fitting process was successful. The log transformation has likely improved the model fit, as indicated by the lower scale value (1.1019) compared to the untransformed model.

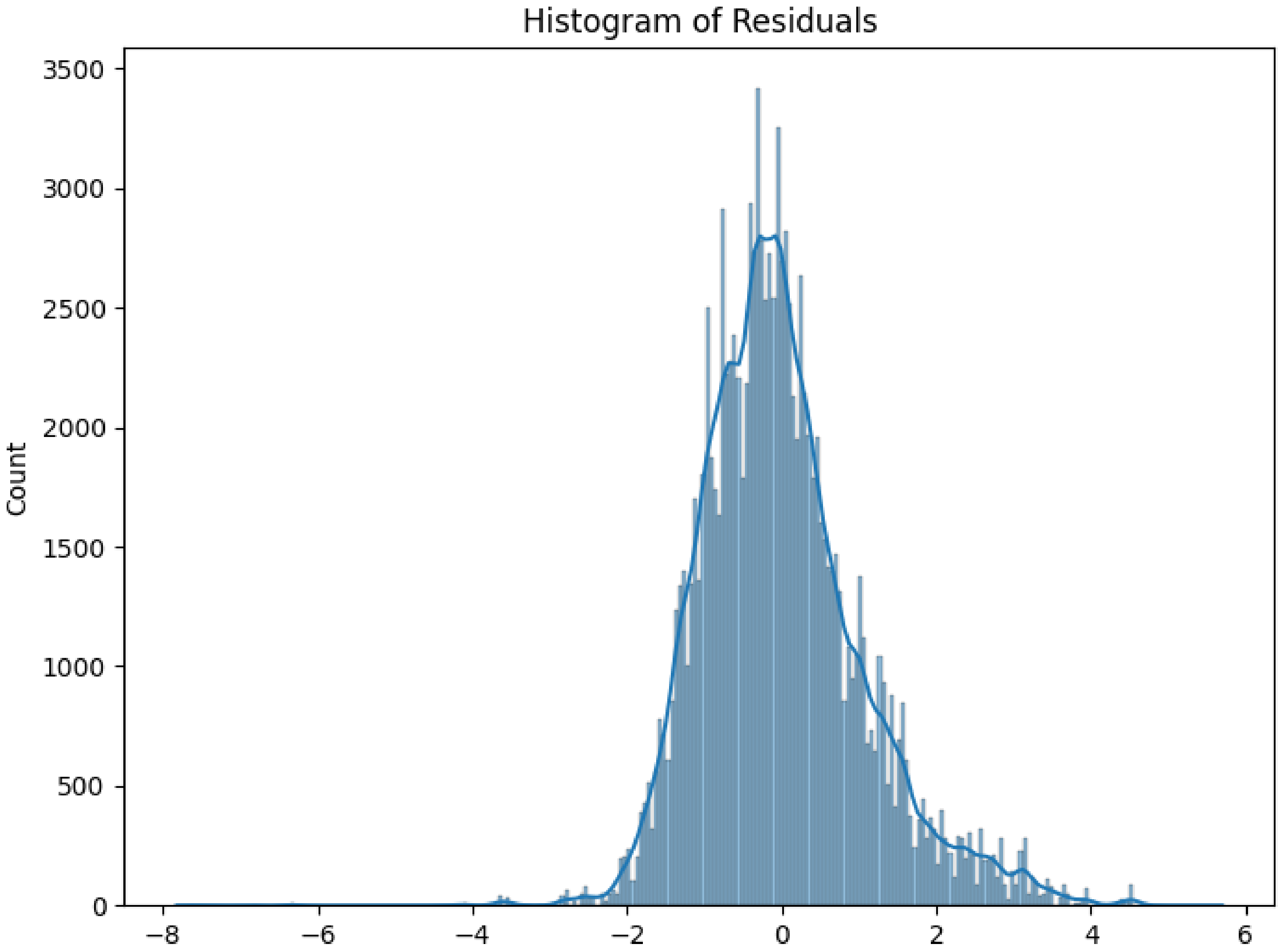
Same as before: validating the model

This plot helps check for homoscedasticity (constant variance of residuals). Ideally, the residuals should be randomly scattered around zero without any discernible pattern.



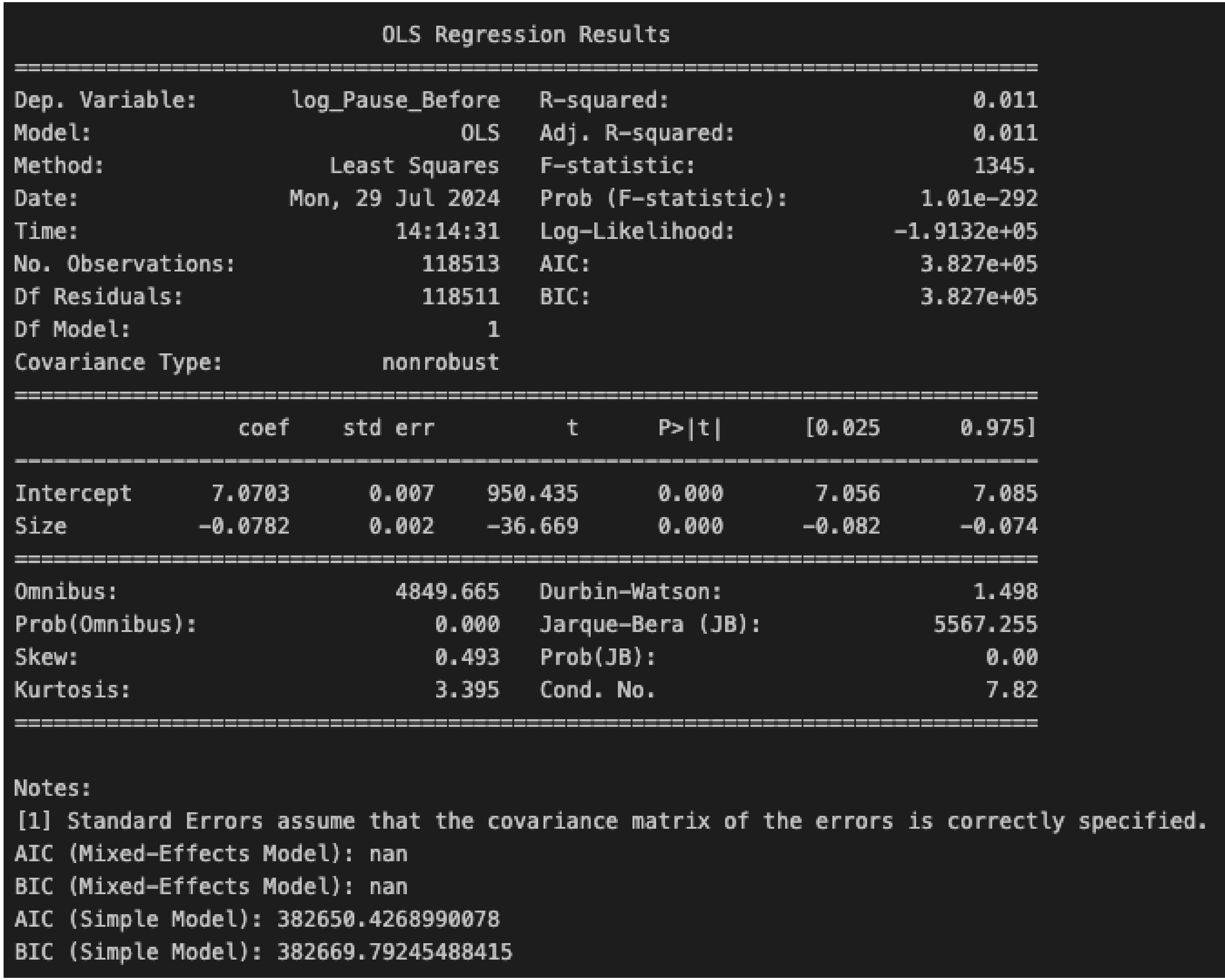
Distribution of Random Effects

The histogram should approximate a normal distribution if the random effects are normally distributed.



Model Comparison with and without Random Effects

- Comparing AIC and BIC values helps determine if including random effects improves the model. Lower AIC and BIC values indicate a better model fit.



```
                            OLS Regression Results
==============================================================================
Dep. Variable:       log_Pause_Before   R-squared:                       0.011
Model:                            OLS   Adj. R-squared:                  0.011
Method:                 Least Squares   F-statistic:                     1345.
Date:                Mon, 29 Jul 2024   Prob (F-statistic):          1.01e-292
Time:                        14:14:31   Log-Likelihood:            -1.9132e+05
No. Observations:              118513   AIC:                         3.827e+05
Df Residuals:                  118511   BIC:                         3.827e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      7.0703      0.007    950.435      0.000       7.056       7.085
Size          -0.0782      0.002    -36.669      0.000      -0.082      -0.074
==============================================================================
Omnibus:                     4849.665   Durbin-Watson:                   1.498
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5567.255
Skew:                           0.493   Prob(JB):                         0.00
Kurtosis:                       3.395   Cond. No.                         7.82
==============================================================================


Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
AIC (Mixed-Effects Model): nan
BIC (Mixed-Effects Model): nan
AIC (Simple Model): 382650.4268990078
BIC (Simple Model): 382669.79245488415
```

En conclusion:

- Intercept : 7.648 -> Cette valeur représente la moyenne de la log_pause lorsque la taille du flux (size) est à 0.

- Effet de la taille du flux (size) : -0.071 -> Ce coefficient indique que pour chaque augmentation d'une unité dans la taille du flux, la log_pause diminue en moyenne de 0.071. Comme le coefficient est négatif, cela signifie qu'une augmentation de la taille du flux est associée à une diminution de la log_pause. **En d'autres termes, une augmentation de 1 unité de size est associée à une diminution de 0.071 ms dans la valeur logarithmique de la pause.**

- Variance du groupe : 0.532 -> Cette valeur suggère qu'il existe une variabilité significative de la log_pause entre les différents sujets. Autrement dit, il y a des différences importantes entre les sujets en ce qui concerne leurs pauses.