

A Statistical Analysis of Relationship Between SUS and RTLX Scores

Kishore Shaw M R, M.Sc. HCI

University College Dublin

kishore.manaksharamani@ucdconnect.ie

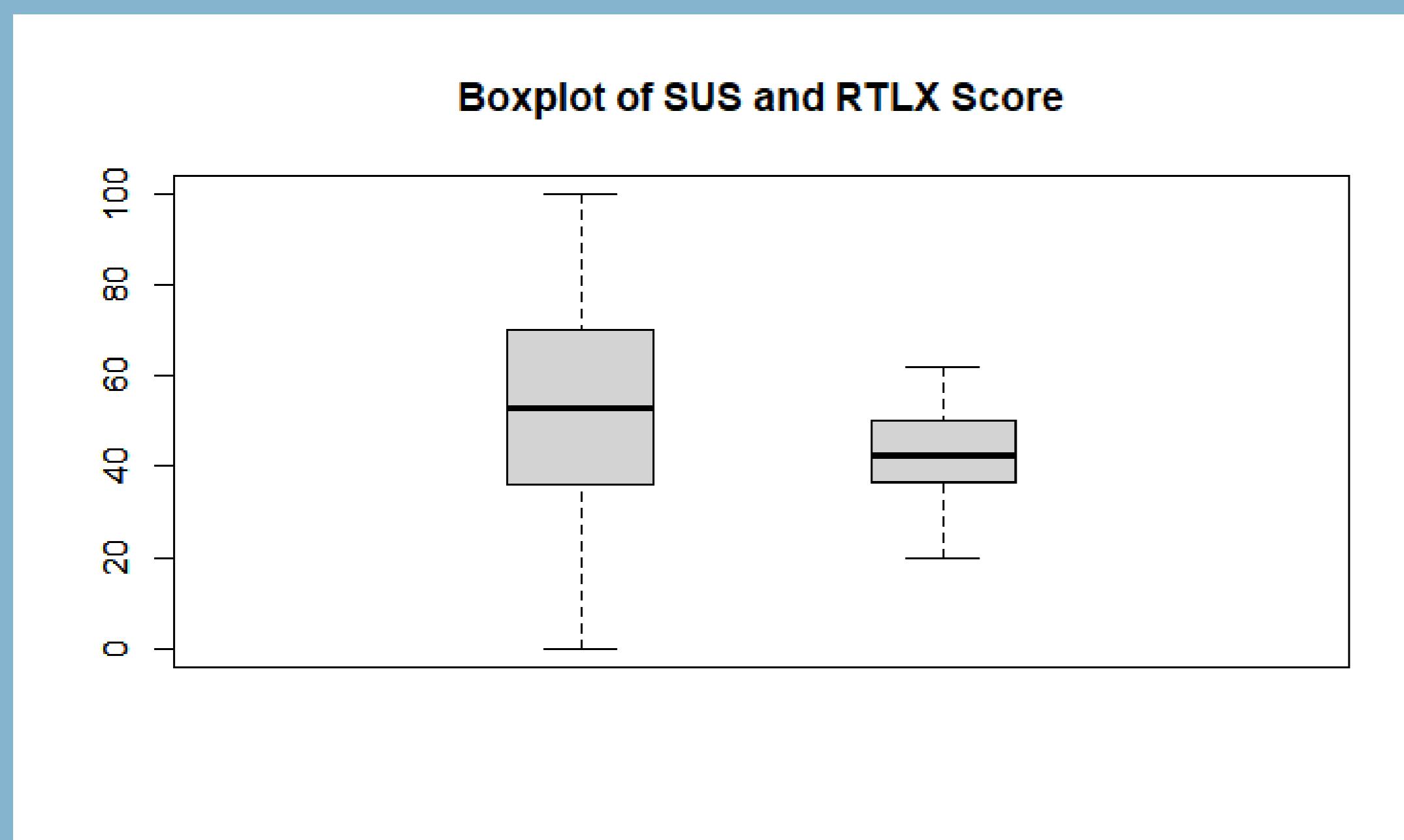
Results

Descriptive Data:

Study on the relationship between subjective work load(RTLX) and system usability(SUS) in the use of voice user interfaces by 100 participants completing a set of 10 everyday tasks with Amazon Alexa via an Echo smart speaker.

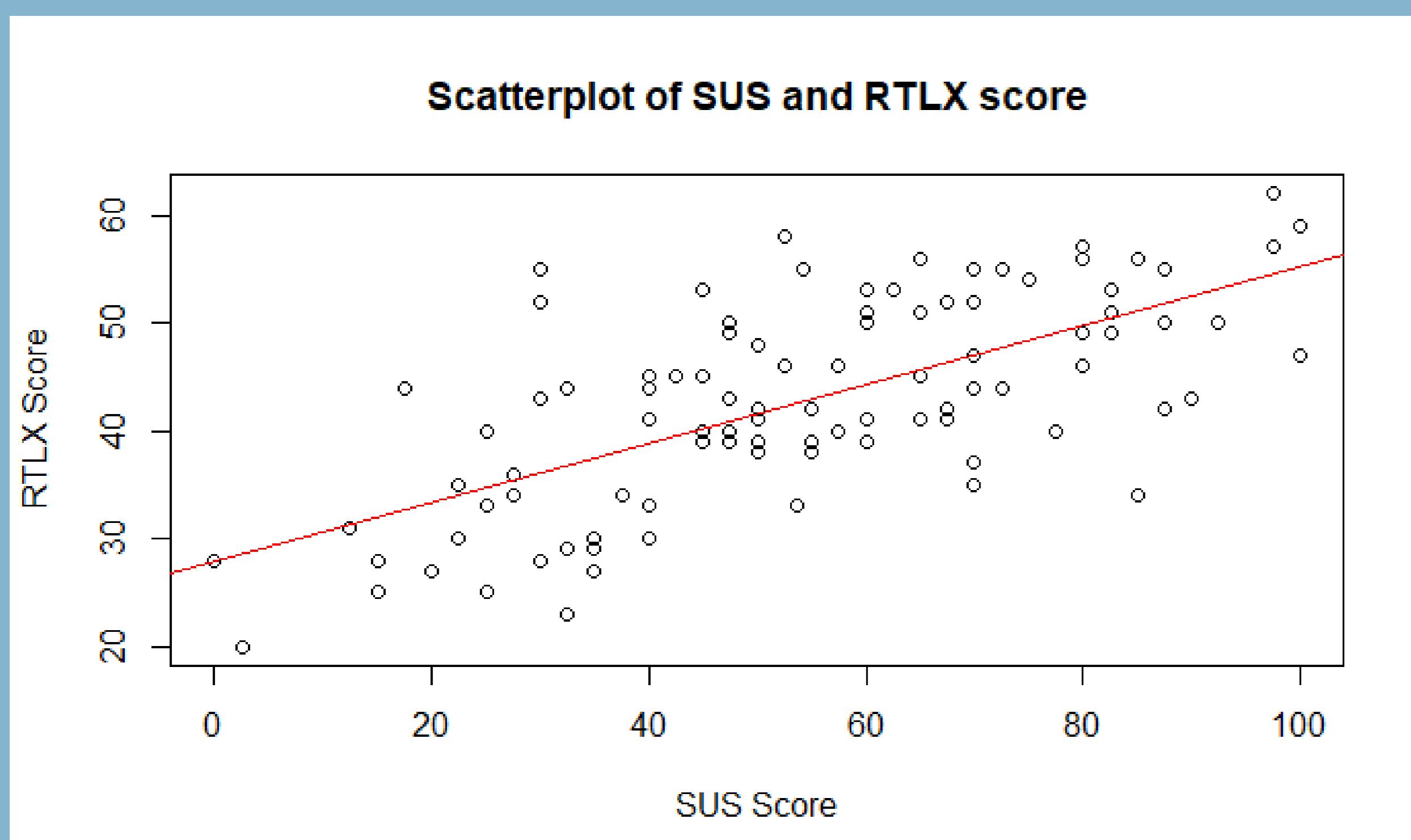
Statistical Analysis:

- Performed data cleaning for the dataset.
- Conducted descriptive statistics (Measures of Central Tendency, Measures of Dispersion) and correlational analysis.



Correlation Insights:

The strong positive correlation of 0.671614 is revealed between SUS and RTLX scores, supported by a significantly low p-value of 2.039e-14, provides robust evidence against the null hypothesis.



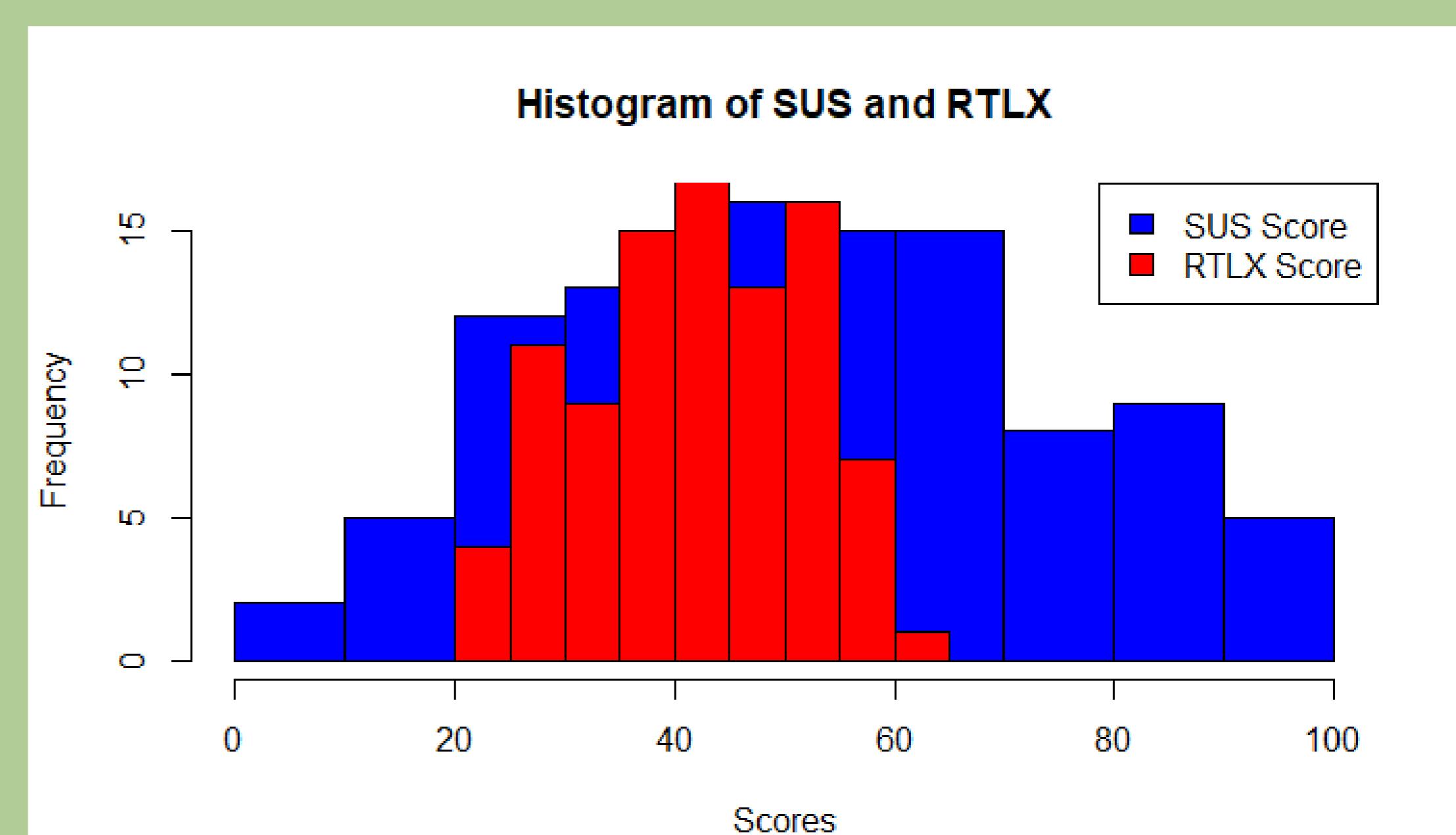
Supporting Hypothesis:

The calculated correlation coefficient of 0.671614, accompanied by p-value<0.001 and a 95% confidence interval between 0.547 and 0.767, strongly supports H1, confirming a moderate and statistically significant positive correlation between the SUS(system usability) and RTLX (mental workload) scores in the dataset.

Discussions

Key Findings:

In summary, from empirical evidence, the Hypothesis (H1) is accepted. This has been supported by obtained Pearson coefficients and p-value(Fisher's Criterion) which confirmed that there is a statistically significant relationship between RTLX and SUS. This association suggests that, in the domain of voice user interfaces, usability tends to improve as mental workload increases.



Functional Outcomes:

The significant positive association suggests that improving voice user interfaces for better usability may reduce mental strain as mental workload and usability are elements of user engagement, suggesting that a little correlation exists between the two constructs (O'Brien et al.,2008) increasing efficiency and they could be jointly used to better describe objective indicator of user performance, a dimension of user experience(Longo, L., 2017).

Limitations:

- **Sample Diversity:** Doesn't represent diverse user population and limited to a relatively constrained linguistic task(Wu et al., 2020)
- **Influence on Perceptions:** participants' characteristics could have influenced the groups' perceptions of both the methods as well as the application(Dr. M Georgsson, 2020).

Future Research:

Moderating Factors: Explore potential variables which can influence the relationship such as factors like user experience, task complexity, user characteristics.

Measurement Refinement: The precision of the measurements could be increased by validating and avoid outliers,corrupted data.

References

- O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955. <http://dx.doi.org/10.1002/as.20801>
- Wu, Y., Edwards, J., Cooney, O., Bleakley, A., Doyle, P. R., Clark, L., Rough, D., & Cowan, B. R. (2020). Mental Workload and Language Production in Non-Native Speaker IPA Interaction. In CUI '20: Proceedings of the 2nd Conference on Conversational User Interfaces (pp. 1-8). Article 3 (ACM International Conference Proceeding Series). Association for Computing Machinery (ACM). <https://doi.org/10.1145/3405755.3406118>
- Longo, L. (2017). Subjective usability, mental workload assessments and their impact on objective human performance. Proceedings Part II, 16th IFIP TC 13 International Conference, 25-29 September, Mumbai, India. doi: 10.21427/mymz-rx38
- Georgsson, M. (2020). NASA RTLX as a novel assessment tool for determining cognitive load and user acceptance of expert and user-based usability evaluation methods. *EJBI*, 16(2), 14-21. <https://doi.org/10.24105/ejbi.2020.16.2.14>