Faster keyboard typist also type fast on a mobile device

Cong Liu - 1006267200

2022-11-18

Abstract

News report shows that people can achieve similar typing speed on mobile devices as on computer keyboards. To confirms this finding, we collected typing speed measurements from students enrolled in STA490. We repeatedly measured each student's typing speed and controlled for other relevant characteristics. We show that, across individuals, the keyboard typing speed and the mobile phone type are significantly related to the typing speed on mobile devices. People who type fast on a keyboard also *tend to* type faster on a mobile device.

Introduction

A news article reported that people can now type almost as fast on their smartphones as they can on their computer keyboards. The article motivates an interest in investigating the question: **Do people who type fast on a keyboard also type faster on their mobile phones?**

The analysis leverages data on typing speed and relevant characteristics collected from students enrolled in a fourth-year statistics course (STA490) at the University of Toronto. Using the individual-level data, we employ a **multiple linear regression model** to understand the association between the computer keyboard typing speed and the typing speed on mobile devices.

The remainder of the report is organized as follows. **Data Summary** introduces the data, variables and averaging used in the analysis. **Methods** describes the study design and model considerations. **Results** reports the regression results. **Discussion** discusses limitations in the analysis and future considerations, finally **Conclusion** concludes the findings from the analysis.

Data Summary

All 39 students enrolled in STA490 tested their typing speed using an online platform. They did three test trials on both a keyboard and a mobile device, contributing 294 typing speed measurements.

The test has a time limit of 60 seconds and the test takers are to type out randomly generated words. For measurement purposes, each **word is standardized to be five characters or keystrokes long in English**, including spaces and punctuation. The online platform thus bases its WPM calculation on the following conversion rule: **5 keystrokes equal 1 WPM**. The records of the test results entail (i) the typing speed (in *correct* **Words Per Minute (WPM)**) and (ii) accuracy (percentage of the correct keystrokes).

On both a physical keyboard and a mobile device, students tested their typing speeds **three times** and recorded the **results from each trial** in a survey form. The raw data exported from the survey underwent a cleaning process.

Table 1: STA490 Typing Test Sample: Variable Definition

Variable	Definition
key_avg_wpm	Averaged keyboard typing speed (WPM)
mobile_fingers	Mobile typing posture in tests (Number of fingers)
mobile_type	Mobile device type used in tests
screen_diag_size	Screen size of mobile device used in tests (cm)
key_type	Keyboard type used in tests
age	Age at the time of tests (years)
key_freq	Keyboard usage frequency at the time of tests
gamer	Gamer at the time of tests
musician	Plays a keyboard instrument at the time of tests
physical_limitation	Physical limitation related to typing at the time of tests
english_fluency	English fluency level at the time of tests
mobile_avg_wpm	Averaged mobile device typing speed (WPM)

Data Averaging

One **key change** to the raw data is that, on each typing device (physical keyboard and mobile phone), the results from the **three test trials are averaged** so that each individual contributes only one value to the analysis. In total, we have 39 observations of individuals' average typing speed on both a keyboard and a mobile device.

We shall recall that the research question in the analysis is: **Do people who type fast on a keyboard also type faster on their mobile phones?** In essence, the research question wants to understand *do individuals with high typing speed on the keyboard also have high typing speed on mobile devices.* For this analysis, we care more about whether this positive association between keyboard typing speed and mobile device typing speed exists **across individuals** rather than validating the existence of the association **within each individual**.

Thus for each individual, the averaged typing speed is kept and used for analyses.

Initial Exploration of Variables

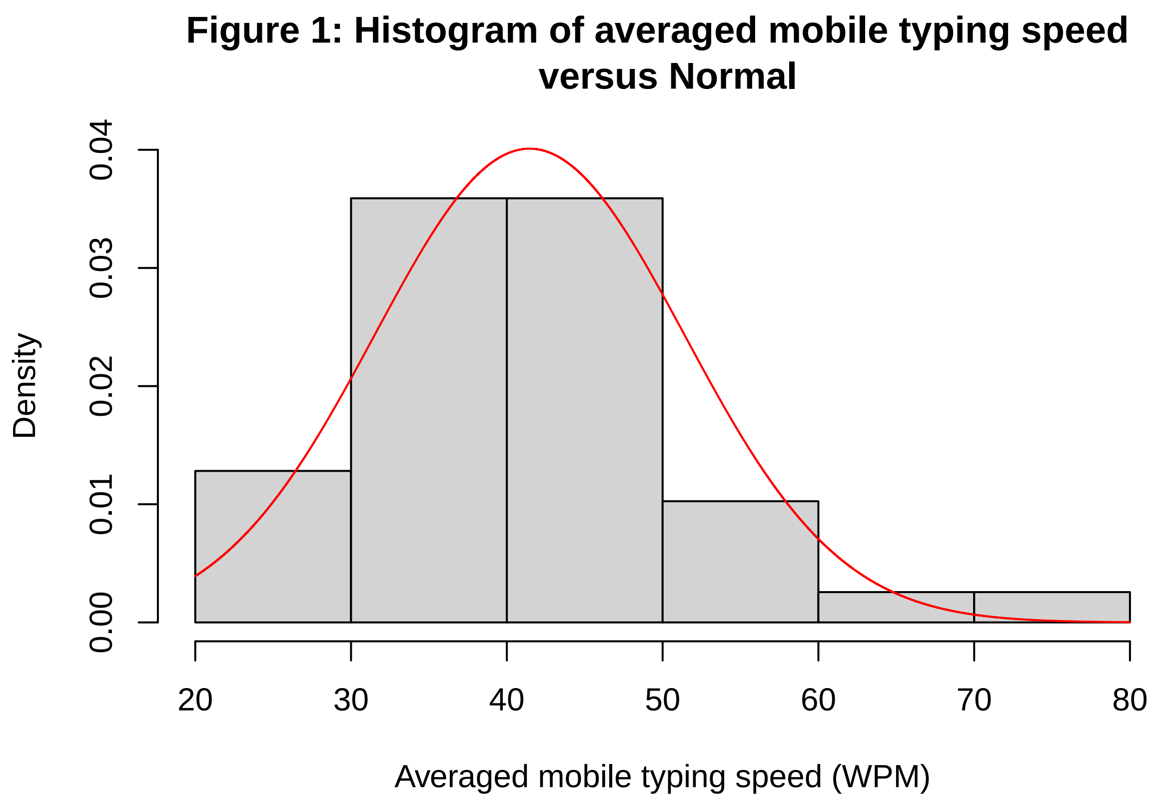
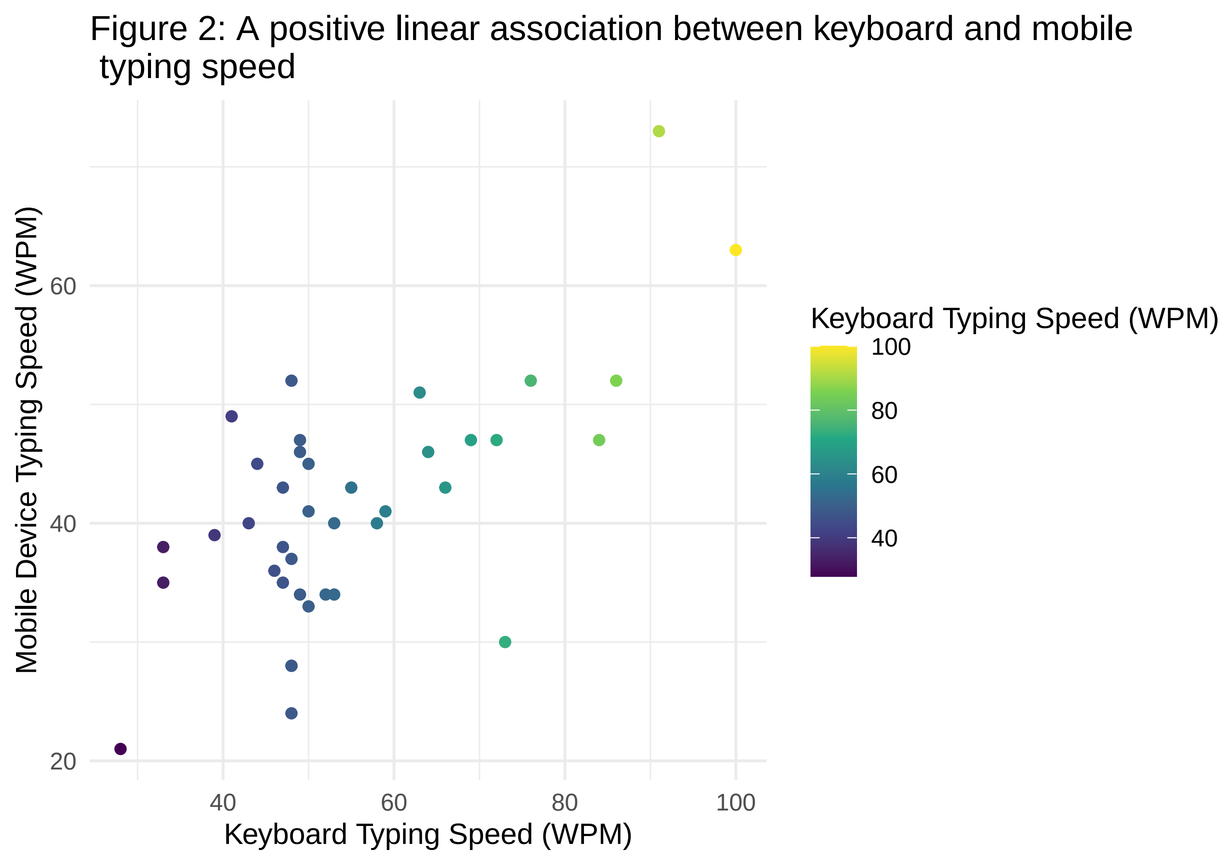


Figure 1 is a histogram of the individuals' averaged mobile typing speed, overlaying the normal density with the typing sample mean and standard deviation (in red). By visual inspection, we may conclude that the distribution of the typing speed reasonably approximates a Normal distribution.



From Figure 2 we can observe a positive trend between the typing speed on a keyboard and the typing speed on a mobile device, after averaging across the three trials on each typing device.

The above plots motivate the choice of methods in this analysis, to be elaborated in the **Methods** section.

Methods

Study Design

All 39 students from STA490 tested their typing speed on an online platform. We expect the sample students to represent other senior undergrads in Canadian universities reasonably well. On both a keyboard and a mobile device, students tested their typing speeds **three times** and recorded the **results from each trial** in a survey form. The three trials were designed for getting a more precise estimate of individuals' typing speed on the two types of devices. Using repeated measurements, we hope to eliminate irrelevant idiosyncrasies (e.g. measurement errors).

Although one would anticipate an increasing trend of typing speed across the three trials, this "learning effect" is irrelevant to the research question. One may even argue that the practice trials allow the test setting better simulate real-world typing scenarios. Thus in the scope of this analysis, we proceed with the caveat that the *number of test trials may be correlated with the typing speed outcome*. After averaging, our data for analysis contains **39 observations** of students' typing speed.

A **fixed effects multiple linear regression** is chosen to model the relationship between keyboard typing speed and mobile device typing speed.

Why multiple linear regression

From Figure 2 above, we can infer that the relationship between keyboard typing speed and mobile typing speed is linear. Observations from the additional Motivating Plots for Modelling (in **Appendix**) show that other variables also affect the mobile typing speed and some may be interacting with the keyboard typing speed. Taking account of the graphical insights, we examine the relationship between keyboard typing speed and mobile typing speed for the STA490 sample **in a multivariate context**. We are to investigate whether keyboard typing speed is a significant covariate among the features collected in the student-level data, controlling for other variables associated with individual typing speed on a mobile device.

Why fixed effects

By using a fixed effects model, data is used to estimate the parameters (especially the expected value of mobile typing speed), which are fixed but unknown constants, and their variability via standard errors.

We do not think it is appropriate to introduce random effects to the model since (i) study participants are **NOT chosen randomly** from the population of interest (senior year undergraduates in a Canadian university). STA490 as a class participates in the study as a convenient sample and no randomization is involved in this sampling process (ii) inducing correlation structure to the dependent variable may be a premature decision given the limited number of observations and the unknown time lapse between trials.

Model Assumptions

We can be assured that the multiple linear regression model gives us reliable coefficient estimates if the data satisfies, or **at least does not have a major violation** of, the assumptions listed below.

1. Independence of observations

Other than enrolling in the same course, there is **no known inter-correlation among the individuals** relevant to typing speed (e.g. the course enrollment is not based on any evaluation of typing proficiency). Thus we do have **39 independent observations**.

2. A linear relationship

The scatterplot between keyboard typing speed and mobile typing speed shows **an approximately linear trend** (Figure 2 in **Data Summary**). There is no sufficient evidence suggesting that other variables have a non-linear effect on the mobile device typing speed (Motivating Plots for Modelling in **Appendix**).

3. No perfect correlation between model predictors

There is some concern since some factors (e.g. students' age, physical limitation and English fluency) affect the typing speed on both devices simultaneously.

4. Error terms have a constant variance.

5. Error terms follow a Normal distribution.

The multiple linear regression also assumes that error terms follow a Normal distribution and have a constant variance. Error terms with a constant variance ensure the stability of our coefficient estimates. Others can be convinced by the model results if the variability of the coefficient estimates is well-controlled. The normal distribution is a nice-to-have. Mathematically, normality gives us convenient properties when we try to dig deeper into the coefficient estimates.

From Figure 1 in **Data Summary**, the approximately close-to-normal distribution of the mobile typing speed may give us some confidence that our data do not have a major violation of the assumptions. The evaluation of the assumptions on the error term's distribution is fully discussed in the **Model Diagnostics** section.

Model Selection

By careful visual inspection of the relevant plots (Motivating Plots for Modelling in **Appendix**), we build four candidate linear regression models.

In our case, the true model is unknown and may not be one of the proposed candidate models. Among the various statistical criteria to evaluate model performance, the Akaike information criterion **(AIC) is being used when trying to find the model that best describes the data**. The lower the AIC value is, the better the model is at describing the data.

After comparing the AIC of the candidate models, we have one optimal candidate model. Then we use the **likelihood ratio test (LRT)** to compare the "winner" model against the simplest model to decide which one of them fits the data better, where the simplest model uses the averaged keyboard typing speed as the only predictor. LRT can help us conclude whether at least one of the coefficients associated with the additional variables is non-zero.

Model Diagnostics

The normality and homoskedasticity of the error term

Three plots are closely examined to **check whether error terms follow a Normal distribution and have a constant variance**.

1. Plot of residuals versus fitted values
2. Plot of residuals versus leverage
3. Normal Quantile-Quantile plot of the standardized model residuals' quantile against the theoretical Normal quantiles

If the model is suitable for the data, we can see:

- From Plot 1
  - Majority of the values fall in the rough range of [-2, 2]
  - Random scatter of the values, no observable pattern/trend
  - Rough symmetry around 0
- Plot 2
  - Falls inside the boundary defined by the Cook's distance, a measurement of how influential an observation is to the fitted model
- Plot 3
  - Quantiles of the standardized residuals agree well with the theoretical Normal residual
  - The quantile points follow closely to the diagonal

Results

Model Specification

Let Y_i = Averaged Mobile device typing speed (WPM).

After comparing the candidate models' performance using the selection process described in **Model Selection**, the final model that has the lowest AIC value and "passes" the likelihood ratio test ($\chi^2_7 = 17.236$, p-value ≈ 0.01541) is

$Y_i = \beta_0 + \beta_1$ (key_avg_wpm) $+\beta_2$ (mobile_type) $+\beta_3$ (screen_diag_size) $+\beta_4$ (key_type)

$+\beta_5$ (age) $+\beta_6$ (gamer) $+\beta_7$ (Professional English fluency) $+\beta_8$ (Full English fluency) $+\epsilon_i$

For variable definitions, please consult Table 1 in **Data Summary**.

The diagnostics plots (in **Appendix**) show that the model does not have a major violation of the aforementioned assumptions (**Model Assumptions**). Thus we proceed with interpreting the results from regression analysis.

Regression Analysis

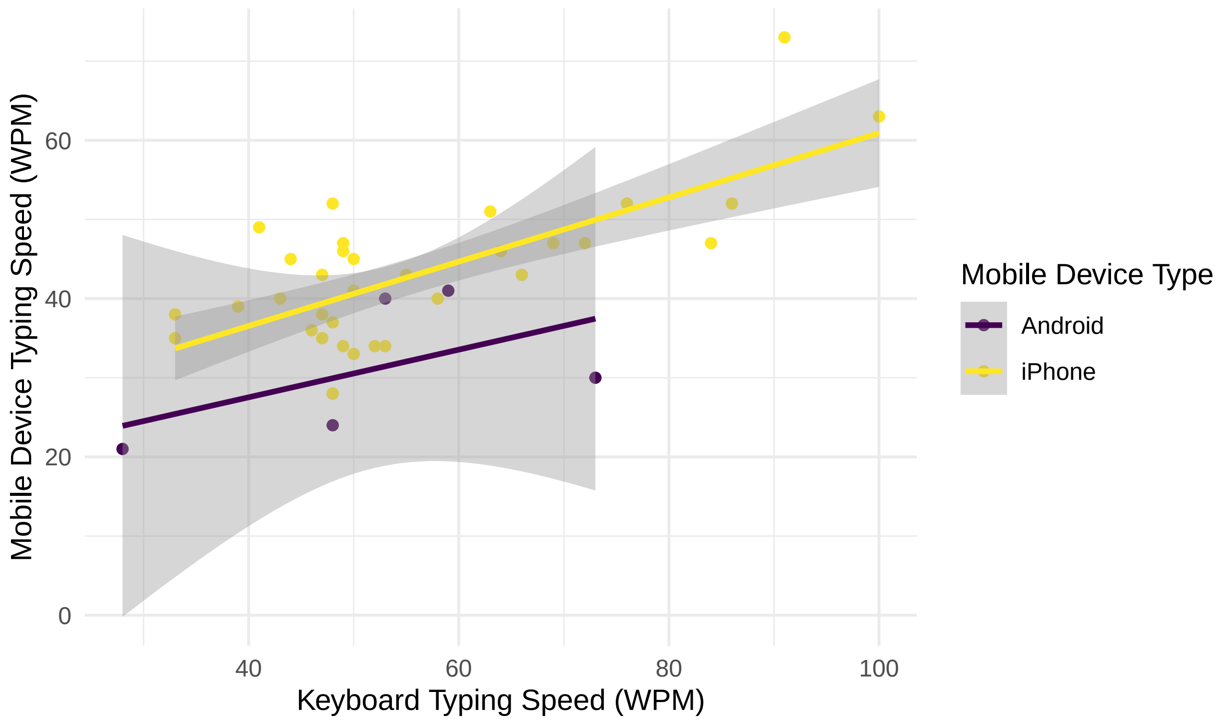
Based on the table that reports all coefficient estimates from the regression of mobile device typing speed on selected predictors (Table 2 in **Appendix**), we have the below findings.

There is a statistically **significant and positive linear association between the keyboard typing speed and the typing speed on a mobile device** (t-statistics $\approx$ 4.0186 with p-value $\approx$ 0.0004). That is, we may answer a **"yes" to our research question**: "do people who type fast on a keyboard also type faster on their mobile phones?". Holding all other model covariates constant, the mobile device typing speed is estimated to increase, on average, by at least 3 words per minute when we type 10 more words per minute on a keyboard.

Additionally, the effect of mobile device type on typing speed is also statistically significant (t-statistics $\approx$ 2.9209 with p-value $\approx$ 0.0066). It is estimated that iPhone users can type at least 10 more words per minute on their phone compared to Android users, controlling for all other predictors.

Figure 3: Keyboard versus Mobile Typing Speed across Mobile Device Types

iPhone users are going to win the typing race?



Conclusion

To conclude, our analysis would answer "Yes" to the research question: **People who type fast on a keyboard also type faster on their mobile phones**.

The analysis focuses on a dataset of typing test results collected from STA490 students, along with other student information relevant to their typing performance (e.g. students' keyboard usage frequency, whether the student has a physical injury at the time of tests). A multiple linear regression model is fit to the data to investigate the relationship between students' keyboard typing speed and mobile device typing speed. From the model, we find that the keyboard typing speed is significantly related to one's mobile device typing speed. Controlling for other model variables, someone who types 10 WPM faster on a keyboard is estimated to have an increased mobile device typing speed by at least 3 WPM on average. Another significant factor is the mobile device type. Holding all other model variables constant, on average, people who use an iPhone can type at least 10 more words per minute compared to Android users.

Discussion

Limitation 1: Inadequate Model

Although the effects of all other covariates on mobile device typing speed were not statistically significant, there is one unexpected estimate. Reading from Table 2 in **Appendix**, the direction of association between mobile device's screen size and mobile typing speed is positive, opposite to what is seen from Motivating Plots for Modelling in **Appendix**. One may also have some concern about the positive association between using a raised (mechanical) keyboard and mobile typing speed could be due to a correlation between keyboard type and keyboard typing speed rather than the type of keyboard directly affecting one's mobile typing speed.

We shall not overstate the regression results. The regression does not have a satisfactory adjusted R^2 (around 0.5451). Only 54.41% of the variation in mobile typing speed is explained by independent variables.

On top of the model's non-satisfactory performance in capturing data patterns, the choice of fixed effects also restricts the ability of findings from this analysis to be generalized to out-of-sample individuals who also belong to the population of interest.

Limitation 2: Under-representative Data

Sampling from the STA490 class is indeed convenient, yet it comes with some issues with the data. In a handful of variables, there is an evident lack of variation. Students enrolled in a fourth-year course are roughly the same age. The majority of the class types on a keyboard every day and uses two fingers when typing on a mobile phone. The prevalent use of the iPhone not only results in small variations in mobile device type, but also affects the screen sizes since iPhones have similar sizes despite the students may have different phone models.

The population of interest, by looking at the research question, can be broadly referred as "people who type fast on a keyboard". In our study, we restrict the study population to students in their senior years of undergraduate studies at a Canadian university. Yet our sample can barely represent this restricted study population. Coupled with the model inadequacies discussed above, the findings from the model cannot substantially contribute to answering the research question.

Limitation 3: Questionable Test Response

Usually, we are typing sentences that are logically coherent, rather than randomly generated words. The coherence allows us to anticipate the words coming next. Being mentally prepared, it is likely that we can type at a higher speed. The typing test's mechanism of randomly generated words does not properly simulate real-world typing. Under the stress of time constraints and test setup, students' typing performance could be downward biased.

Future consideration

In terms of the study design, it would be beneficial for the analysis to fix the time lapse between each trial of the typing tests and to increase the number of trials for each typing device. Instead of randomly generated words, we can test the typing speed using randomly generated article excerpts so that the test results can better reflect individual typing speed. When collecting the test results from study participants, all survey questions need to clearly state what information is being recorded. For example, the question about English fluency level should clarify the ranking of the levels so that the responses accurately reflect individuals' English proficiency. Most limitations discussed above can be attributed to not having a representative sample of the typing population. A model built on an inadequate sample lacks the ability to explain and generalize to the larger population. A larger sample contains more information about the population and the models have a higher chance of capturing the relationships.

Appendix

Motivating Plots for Modelling

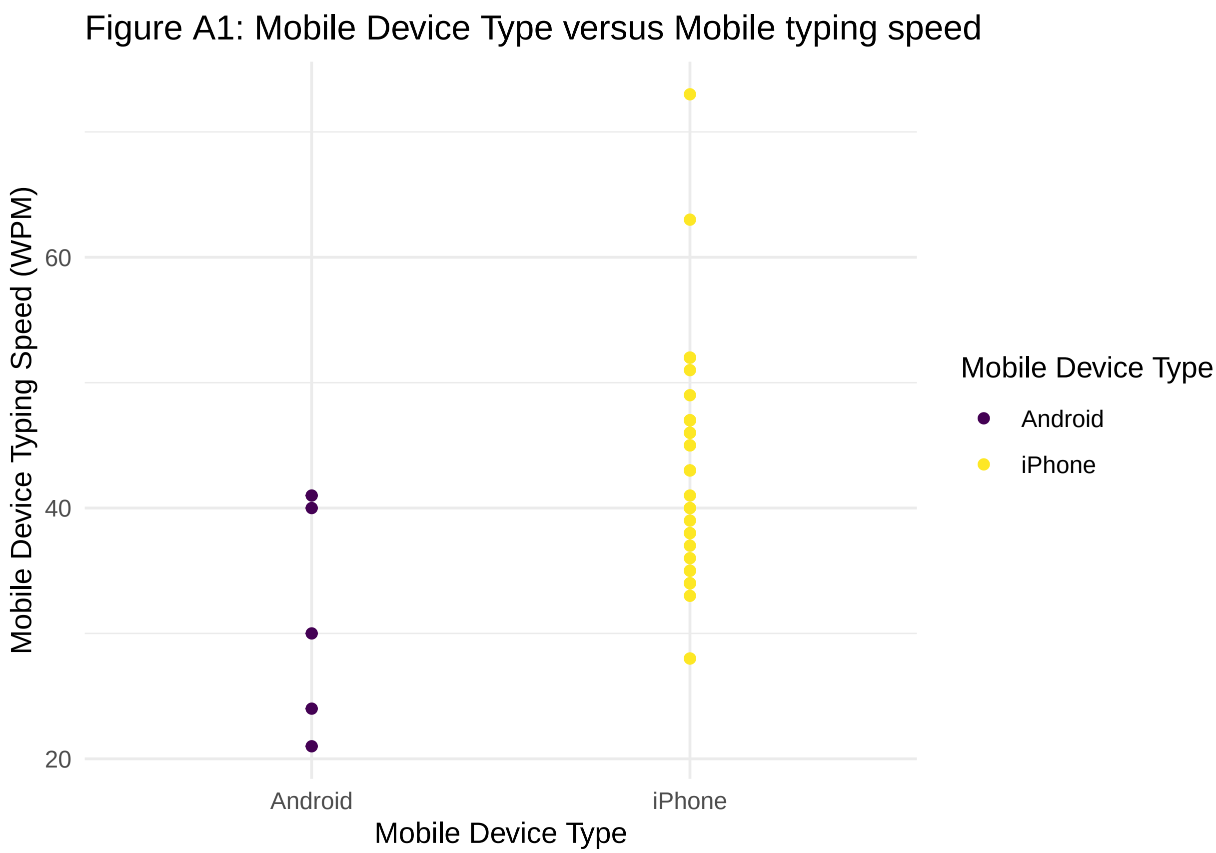


Figure A1: There is a visible difference in mobile typing speed between iPhone user and Android user. One caveat is that there are ONLY FIVE subjects using Android.

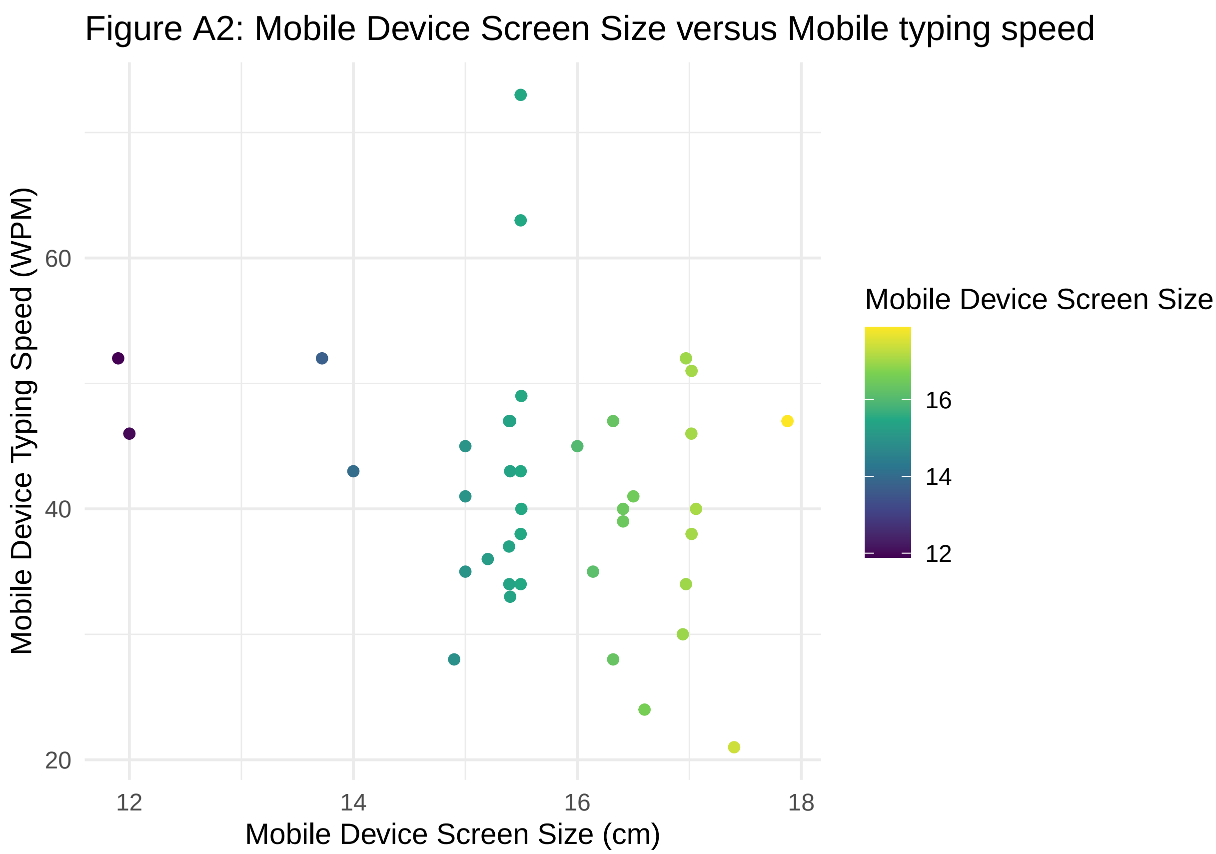
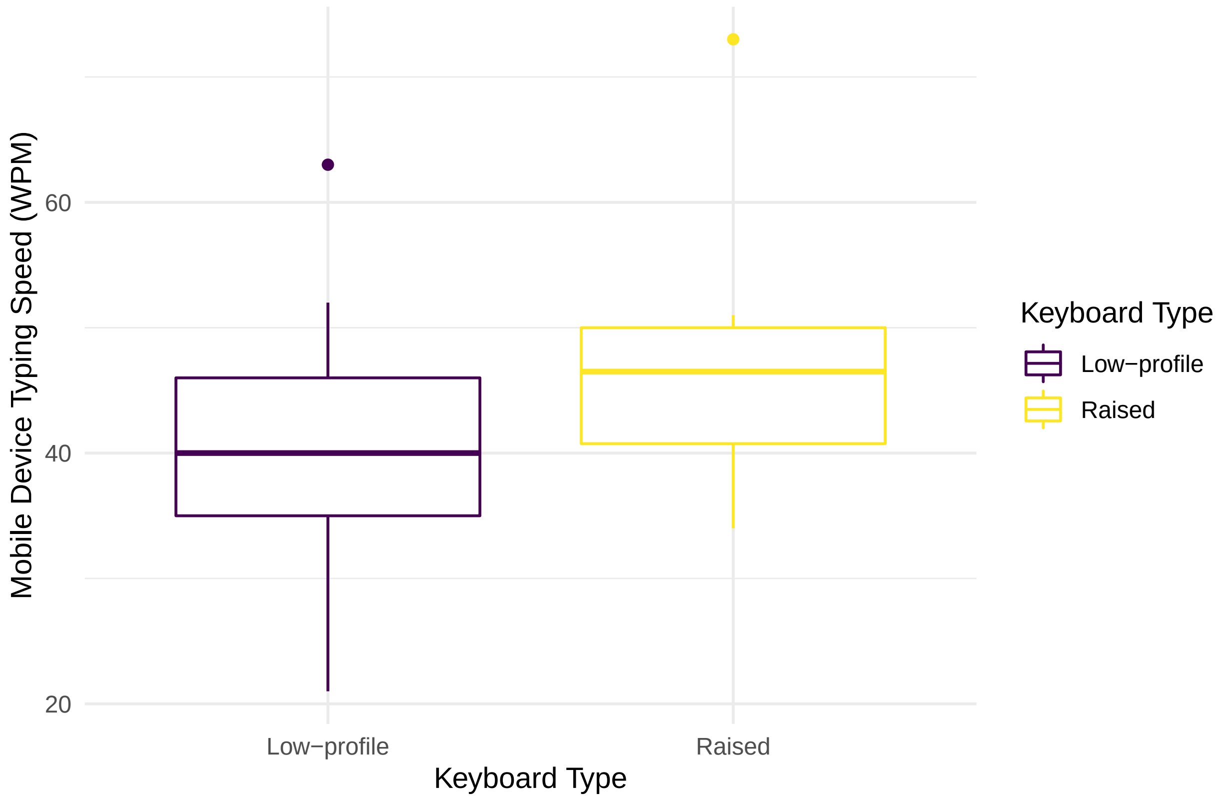


Figure A2: A negative linear association between mobile screen size and typing speed on a mobile device.

Figure A3(a): Keyboard Type versus Mobile Typing Speed



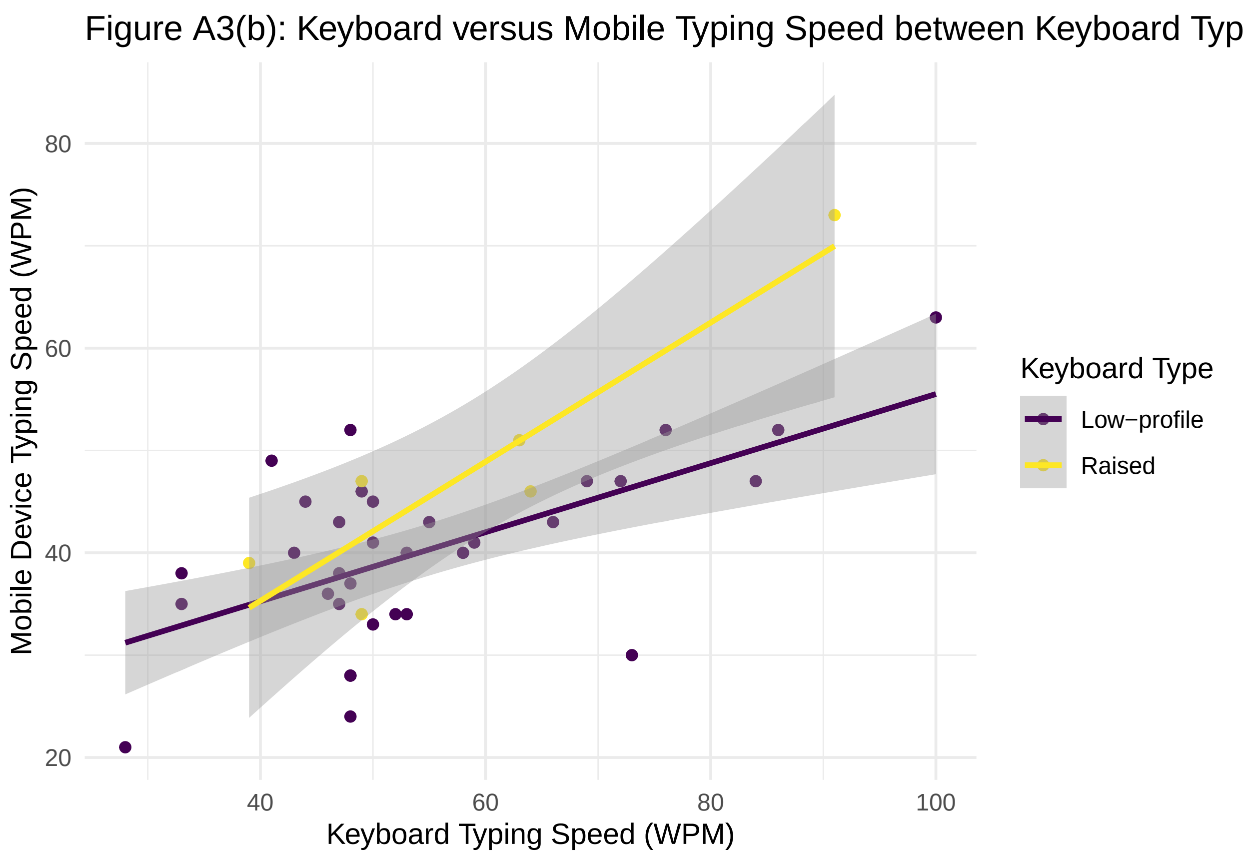


Figure A3(a), A3(b): There is a visible difference in typing speed between using a low-profile keyboard versus a raise one. **Yet** (i) there is ONLY SIX subjects using raised, mechanical keyboard (ii) **collinearity** between keyboard type and typing speed on the keyboard when including this variable in the model.

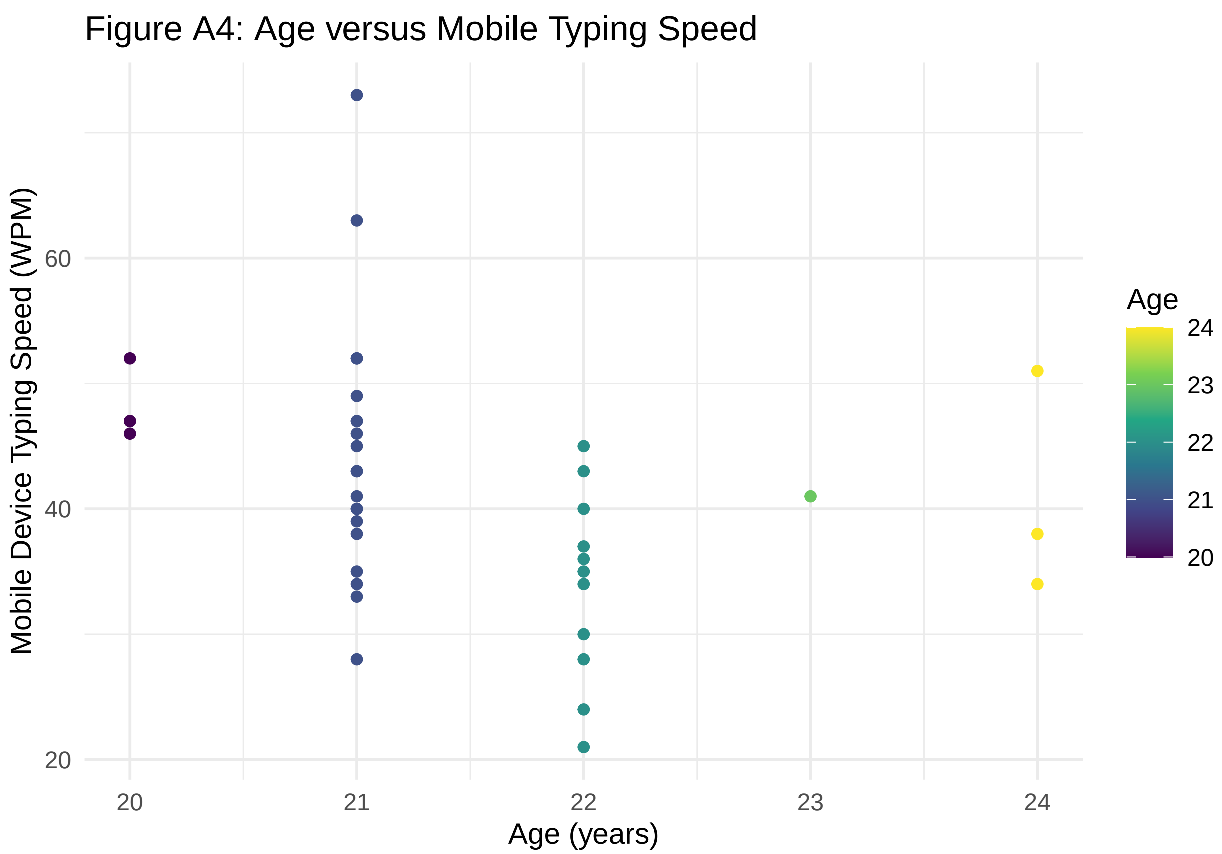


Figure A4: There is a negative linear association between age and typing speed on a mobile device, for those at the age of 20-22. **Yet** there are only three observations for age 20, three observations in total for 23-24.

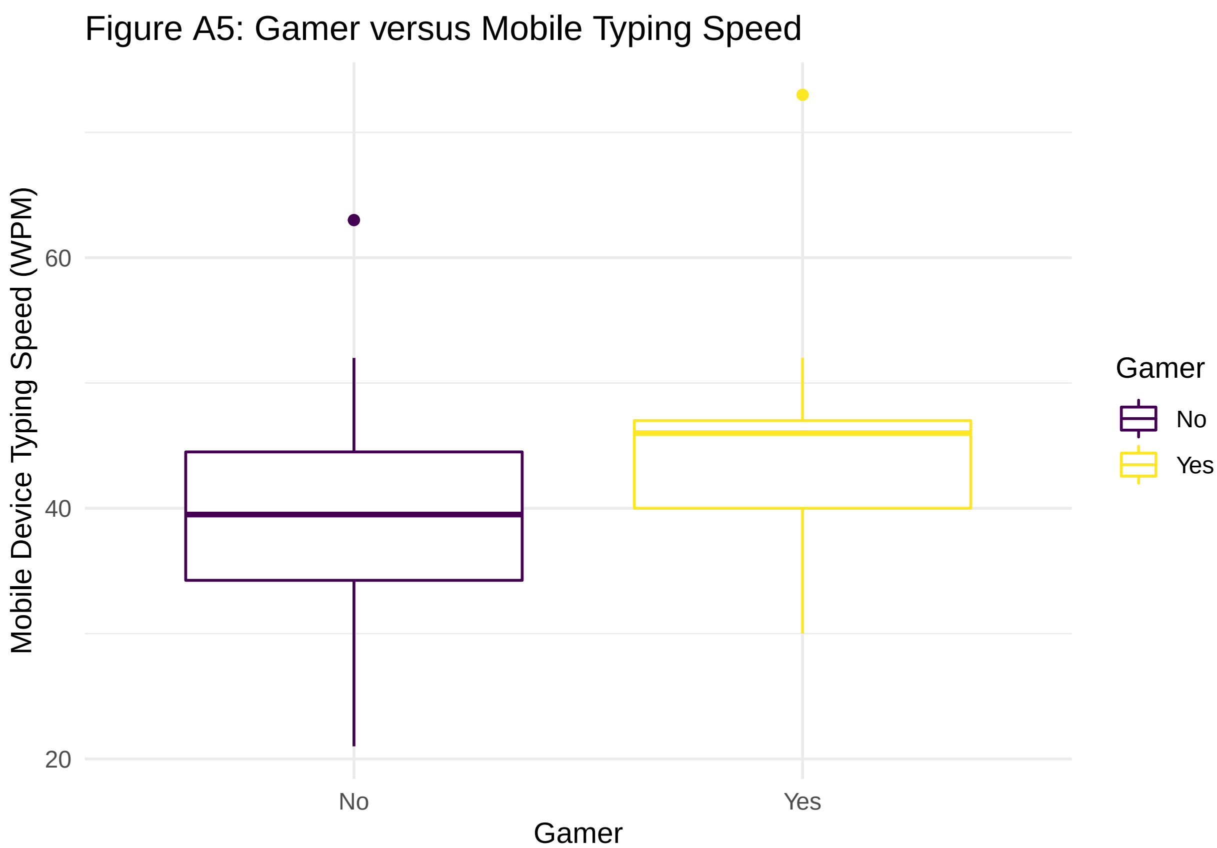


Figure A5: There is a visible difference in mobile typing speed between gamer and non-gamer.

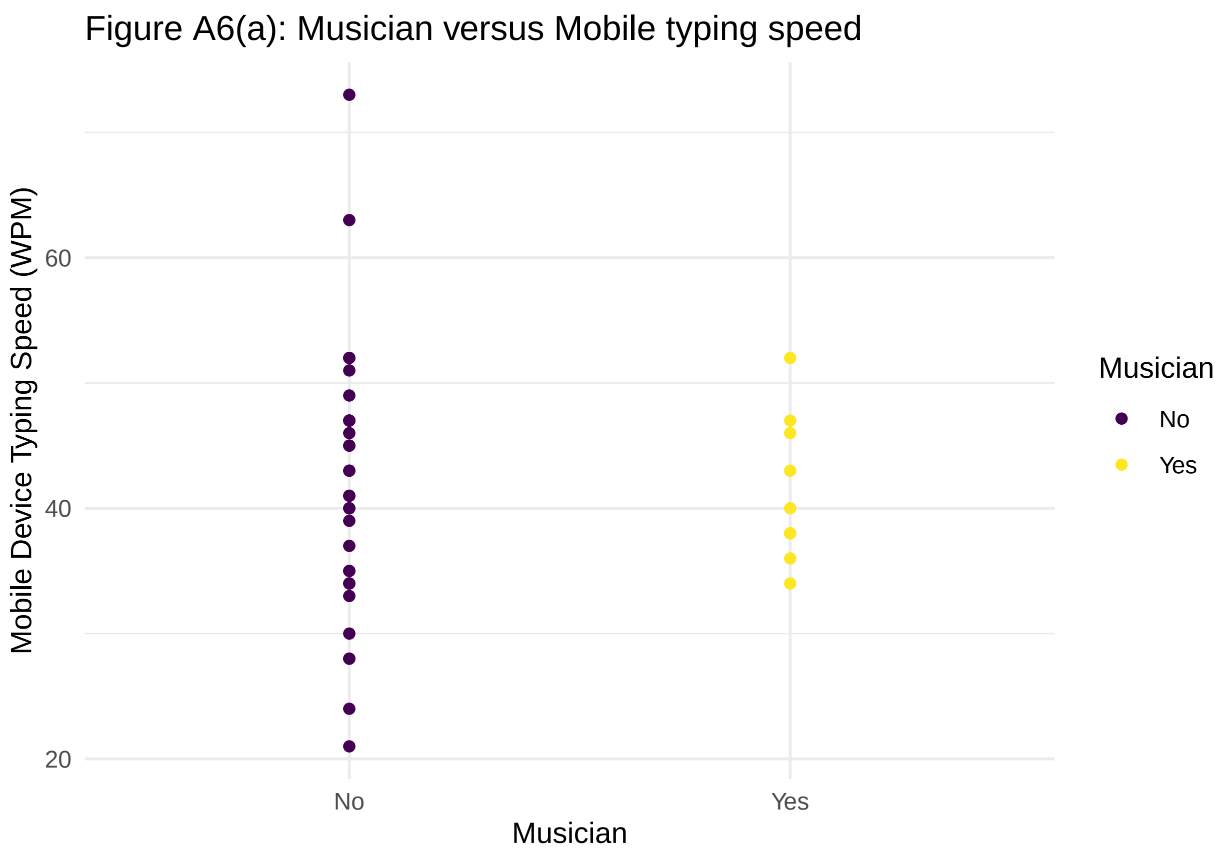


Figure A6(a): Musician versus Mobile typing speed
Mobile Device Typing Speed (WPM)
60
40
20
No
Yes
Musician
Musician
No
Yes

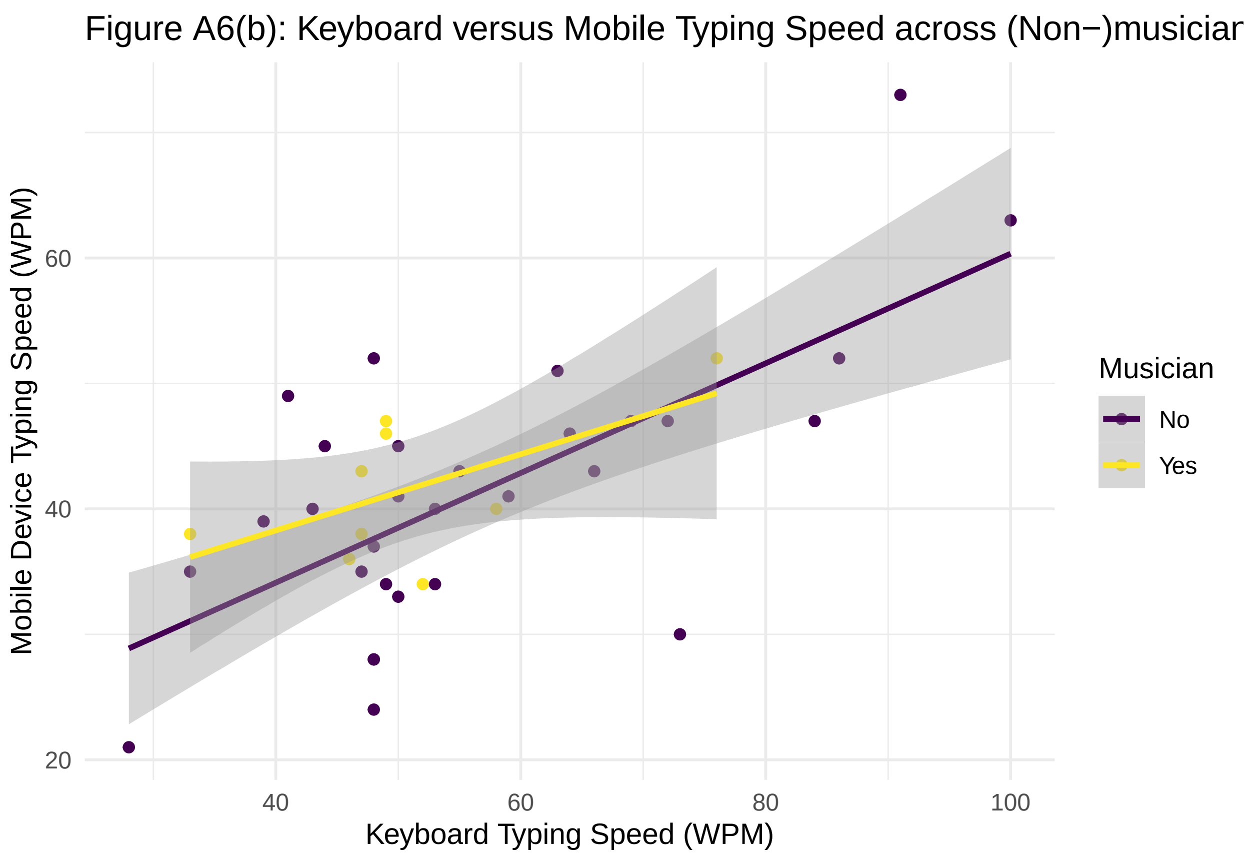


Figure A6(a), A6(b): Potential interaction between playing a keyboard instrument and keyboard typing speed.

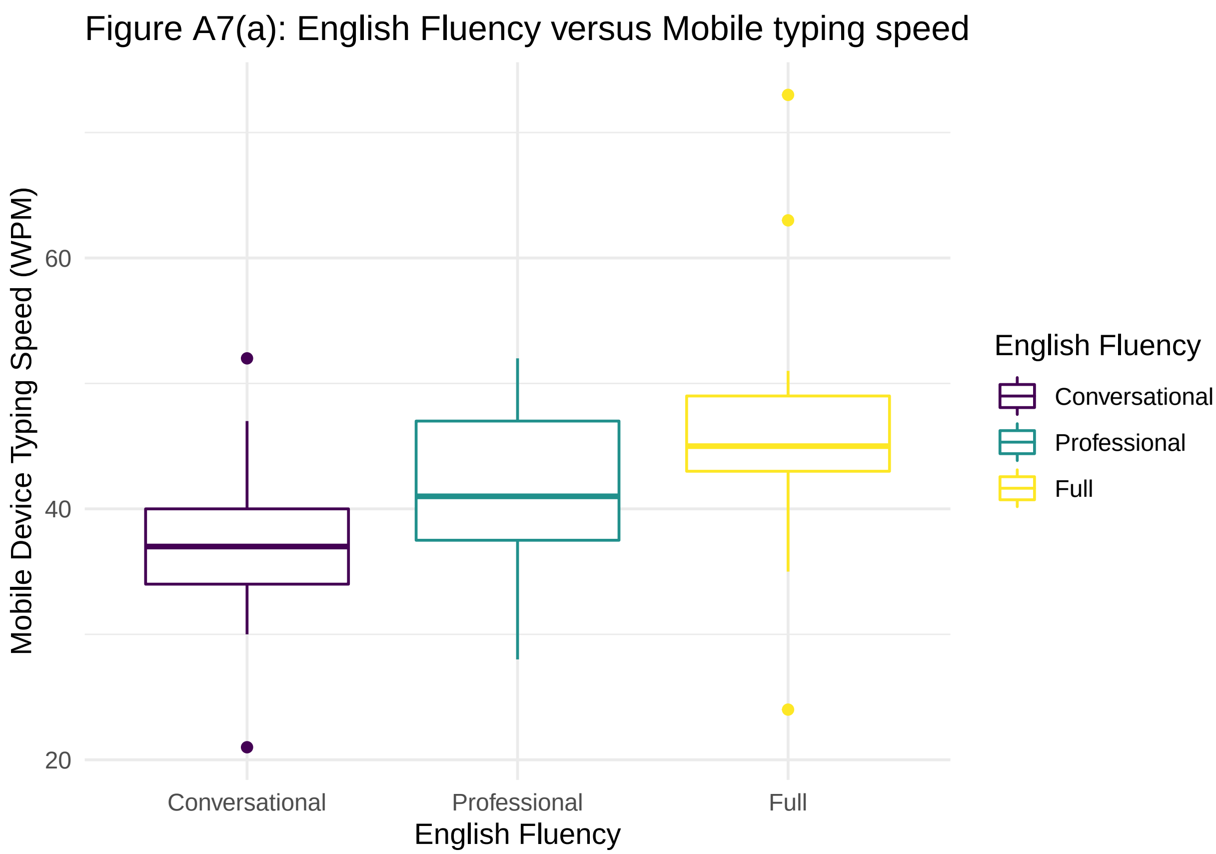


Figure A7(a): English Fluency versus Mobile typing speed
Mobile Device Typing Speed (WPM)
60
40
20
Conversational
Professional
Full
English Fluency
English Fluency
Conversational
Professional
Full

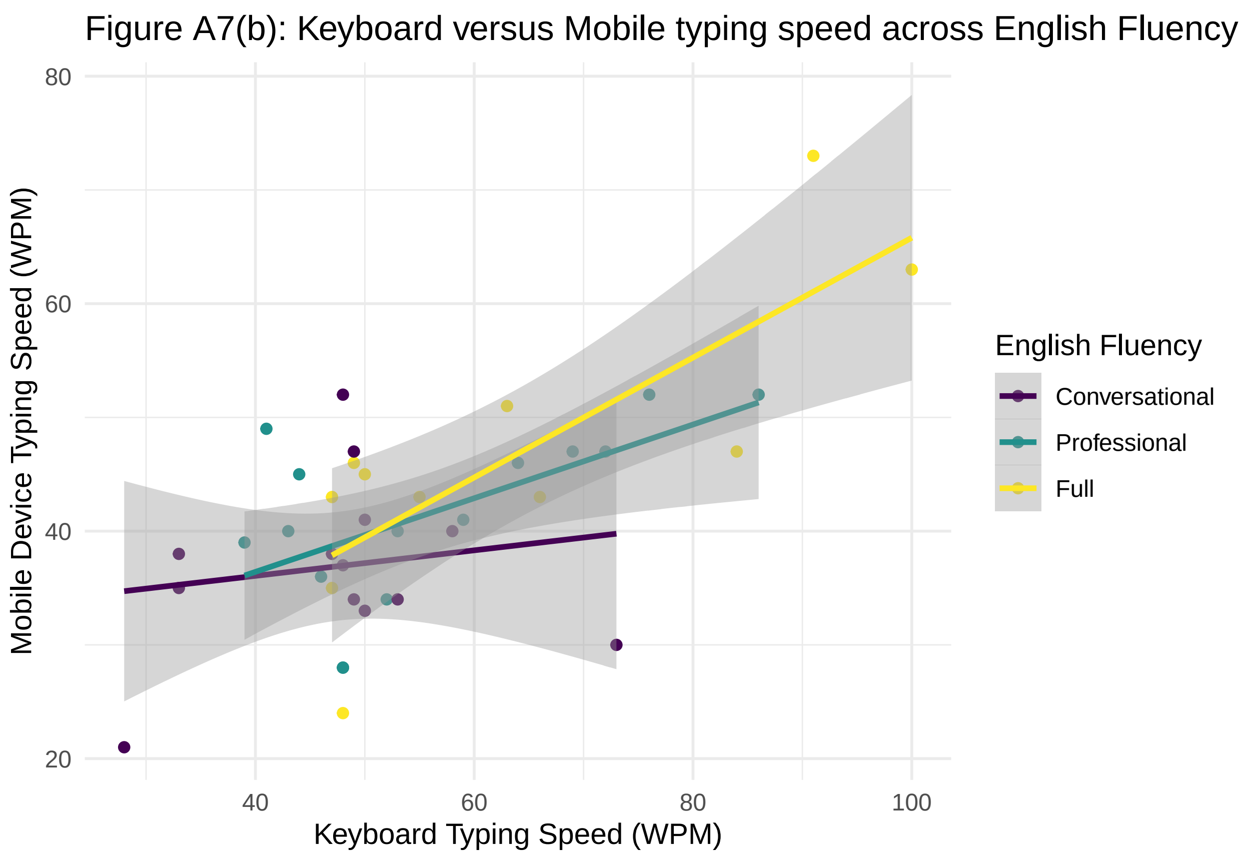


Figure A7(a), A7(b): Potential positive association between English fluency level and mobile device typing speed. Potential interaction between English fluency level and keyboard typing speed.

Regression Analysis

Table 2: Regression Analysis of Mobile Device Typing Speed on Selected Predictors

Term	Estimate	Std. Error	t-statistic	p-value (for Wald test)
Intercept	31.5007	33.8860	0.9296	0.3600
Keyboard typing speed (WPM)	0.3387	0.0843	4.0186	0.0004
Mobile device type: iPhone	10.5396	3.6083	2.9209	0.0066
Mobile device screen size (cm)	0.9399	1.0606	0.8862	0.3826
Keyboard type: Raised	4.9906	3.5977	1.3872	0.1756
Age	-1.6423	1.2787	-1.2844	0.2088
Gamer: Yes	1.1650	2.8232	0.4126	0.6828
English fluency: Professional	1.4491	2.9447	0.4921	0.6262
English fluency: Full	2.8090	3.0746	0.9136	0.3682

Diagnostics Plots

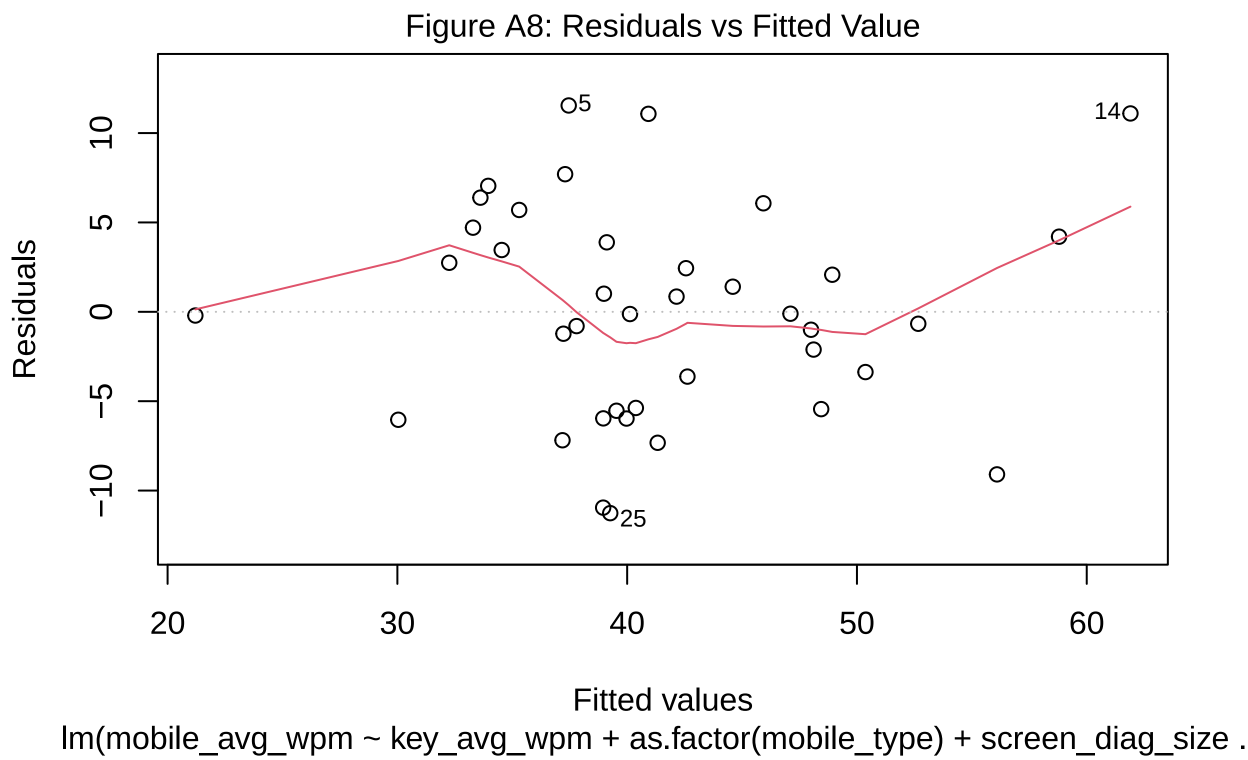


Figure A8: Plot of residuals versus fitted values

By visual inspection,

- no systematic pattern in the residuals, linearity of the relationship
- no cluster of residuals, independence of errors
- no fanning pattern, constant variance

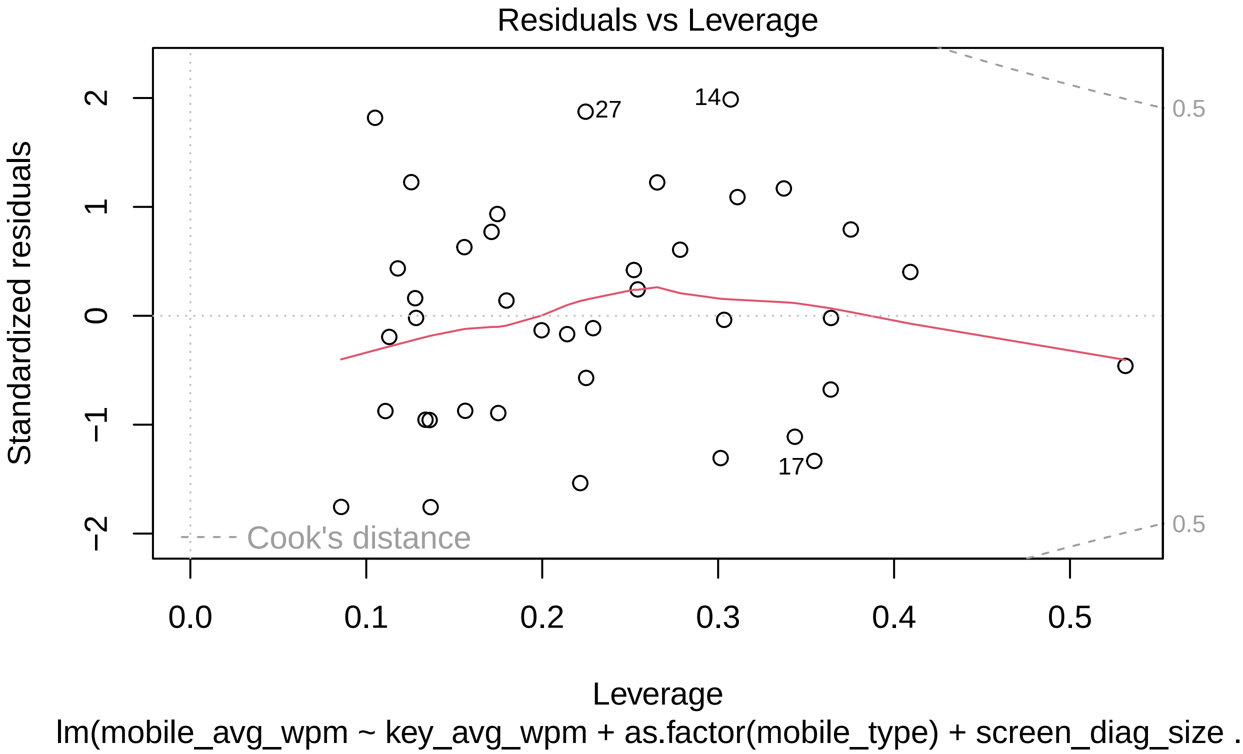


Figure A9: Plot of residuals versus leverage

Observation #17 lies closest to the border of Cook's distance, but it doesn't fall outside of the dashed line. This means there are not any influential points in our regression model.

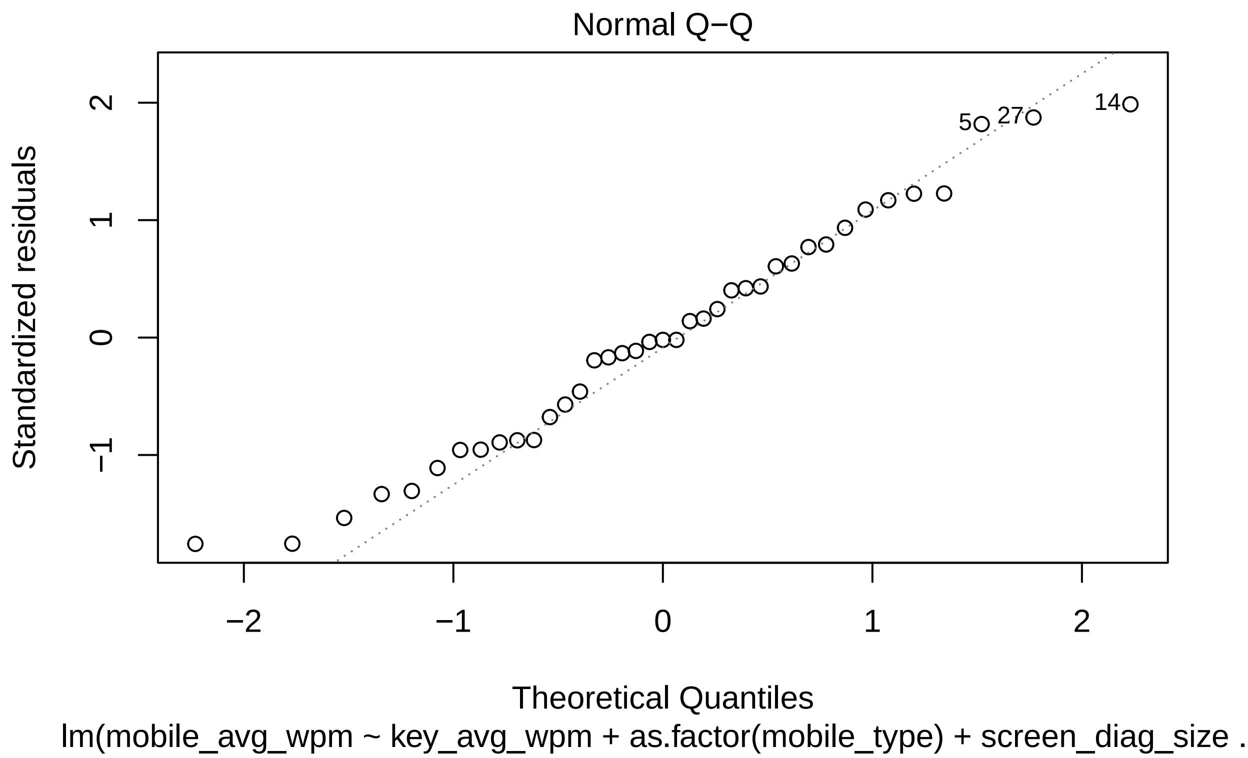


Figure A10: Normal Quantile-Quantile plot

Majority of the standardized residuals stay close enough with the theoretical normal quantiles.