Word Separator Test

Experiment 2, Experimentation & Evaluation 2021

Filippo Piloni & Alessandro Zanzi

# Abstract

Short (120-130 words) summary of your entire report. Give the reader a quick idea of what you did and what the main findings were (if you prepare this report ahead of time, leave out the findings until after you finish the analysis).

# 1. Introduction

Effective communication through written code is crucial in software development, and the readability of source code significantly impacts the efficiency of understanding and maintaining software systems. One aspect that has been explored in natural language reading studies is the use of explicit separators between words, which has been shown to improve reading speed by up to 20%. The question that arises is whether this finding holds true for source code as well. Specifically, does the use of a specific separator influence the speed of reading composed identifiers in code?

The motivation behind this experiment lies in the potential impact on code readability and developer productivity. In software development, developers often encounter composed identifiers, such as variable names or function names, that consist of more than one word. The common conventions for such identifiers include camelCase (e.g., readingThisText) and kebab-case (e.g., reading-this-text). Understanding whether these conventions affect code reading speed can provide valuable insights into best practices for code style and contribute to the ongoing discussions within the software development community.

To investigate the impact of different separators on code reading speed, our experiment will focus on three conditions: camelCase, kebab-case, and no separator. Participants will be presented with code snippets containing composed identifiers written in each of these styles. The experiment will be conducted in a controlled environment where participants will read and comprehend the provided code snippets. We will measure the time taken by participants to read and understand the code in each condition. By analyzing the reading speeds across the three conditions, we aim to draw conclusions regarding the effectiveness of separators in enhancing code readability. This experiment holds the potential to inform coding conventions and practices, contributing to the ongoing efforts to improve code quality and developer experience.

|  |
| --- |
| **Hypotheses:** |
| Write down your (falsifiable!) hypotheses here. Each hypothesis must include **independent** and your **dependent** variables. You must write down your hypotheses **before** you do your experiment!  **Null Hypothesis:** there will not be a significant difference in the response time or the number of attempts across different word formatting styles |

# 2. Method

In the following subsections, describe everything that a reader would need to replicate your experiment in all important details.

## 2.1 Variables

Explicitly identify the independent variable(s) (i.e., what you as the experimenter manipulate):

|  |  |
| --- | --- |
| **Independent variable** | **Levels** |
| Separator between the words  Number of words | Camel case, kebab case, no separator  2, 3 |

Explicitly identify the dependent variable(s) (i.e., what you measure):

|  |  |
| --- | --- |
| **Dependent variable** | **Measurement Scale** |
| Time required to click the correct word | Seconds |

Explicitly identify any important control variable(s) (i.e., what you keep constant): Note that you do *not* need to spell out items that you do not expect to make a *significant* difference! E.g., do not list room temperature unless you believe that minor differences have an impact! Or the other way around: only list variables here that you think are important to keep at a certain level.

|  |  |
| --- | --- |
| **Control variable** | **Fixed Value** |
| Words to choose | Software-related words |

Explicitly identify the blocking variable(s) (i.e., potential sources of variability you measure and will use to partition the experimental units into blocks, but that are not part of the hypothesis):

|  |  |
| --- | --- |
| **Blocking variable** | **Levels** |
| Computer abilities  Language proficiency  Prior exposure to formatting style | Low, Medium, High  Low, Medium, High  No exposure, some exposure, high exposure |

## 2.2 Design

Check off the characteristics of your experimental design:

**Type of Study** (check one):

|  |  |  |
| --- | --- | --- |
| ⃞ **Observational Study** | ⃞ **Quasi-Experiment** | X **Experiment** |

**Number of Factors** (check one):

|  |  |  |
| --- | --- | --- |
| ⃞ **Single-Factor Design** | X **Multi-Factor Design** | ⃞ Other |

**Between vs. Within** (check one):

|  |  |  |
| --- | --- | --- |
| ⃞ **Between Group Design** (independent measures) | X **Within Subject Design** (repeated measures) | ⃞ Other |

Explain, (1) *in text using terminology from the book and lectures* **and** (2) with a figure (similar to those used in Chapter 3 of the Field & Hole book), what kind of experiment you did.

The study incorporates multiple independent variables, forming a multi-factor design. This approach allows us to explore the interactions and combined effects of these variables, providing a more comprehensive understanding of the phenomena under investigation.

For the experiment, all the possible combinations of independent variables have been taken into consideration for all the control variable. Each separator (camel case, kebab case, and no separator) is tested with both two and three words.

The experiment has been conducted using the within subject design each participant serves as their control, experiencing all levels of the independent variable. In this case, each participant reads code snippets in all three conditions (camelCase, kebab-case, and no separator).

## 2.3 Participants

Describe who will take / took part in your experiment. Provide descriptive/summative statistics of their gender, age, professional backgrounds, and any other characteristics that may be relevant to your experiment. Also explain how you will recruit / recruited them (volunteers recruited through email, classmates who were asked to do this, etc) and how you will allocate / allocated them into the different study conditions, i.e., control group vs experimental group(s).

The experiment involved fifteen participants, all recruited through communication channels like messages and emails within and outside a university environment. The participants exhibited a diverse set of characteristics, encompassing various factors such as age, sex, occupation, field of study, education level, usage patterns, and engagement with coding at home.

Out of the fifteen participants, 10 were male, while the remaining 5 were female. The majority of the participants, specifically 73% (11 individuals), were currently engaged in studying computer science, reflecting a focus on individuals with a background in this field, with an high-school diploma as highest degree.

The age range of the participants varied, with individuals ranging from 17 to 38 years old. Occupations included both study and work, with participants pursuing education at different levels, from high school to a Ph.D. level. Fields of study covered a spectrum, with a notable concentration in computer science.

Education levels among the participants included bachelor's, master's, and Ph.D. degrees, indicating a diverse educational background.

Additionally, participants' coding habits at home were explored, with some indicating an habit to code at home (10 of them). This diversity in coding habits provides insights into the participants' extracurricular engagement with the subject matter outside formal educational settings.

## 2.4 Apparatus and Materials

* Computer: model Dell Precision 5550
* Web Application: Next.js 14
* Libreoffice: version 6.4.7.2 40 (Used to save the datas)
* Jasp 0.18.1 (Used for statistics and graphs)

## 2.5 Procedure

Describe how you used your props and the participants to perform your actual experiment, i.e., how you actually carried out a single experimental run. What was done to the participants? What did they have to do? How long did each session take (unless this is an actual dependent variable)? If you did not have participants, explain, e.g., what software was started by whom in what order.

In conducting the experiment, the participants have been engaged through an online survey platform. The experimental setup involved presenting participants with a series of 30 questions, each displaying a word in either camel case, kebab case, or without a separator. The main objective for each participant was to identify the correct word from a set of three distractors.

The participants accessed the survey through a designated website. Upon initiation, each question prompted the display of a word formatted in one of the specified cases. The participants were required to carefully examine the word and select the correct one from the three alternatives provided. The survey platform recorded both the time taken by the participants to respond to each question and the number of attempts made before identifying the correct word.

Participants were instructed to approach each question independently, ensuring that their selections were based on accurate identification rather than random guessing. The order of the presentation of the questions was randomized to minimize potential order effects.

The survey was designed to be user-friendly, with clear instructions provided at the beginning to guide participants through the process. Ethical considerations, such as informed consent and the protection of participant privacy, were addressed in accordance with research guidelines.

Data collected during each session, including response times and the number of attempts, were later analyzed to draw conclusions about the participants' ability to identify words in different formatting styles. The use of the online survey platform allowed for efficient data collection and streamlined the experimental process.

# 3. Results

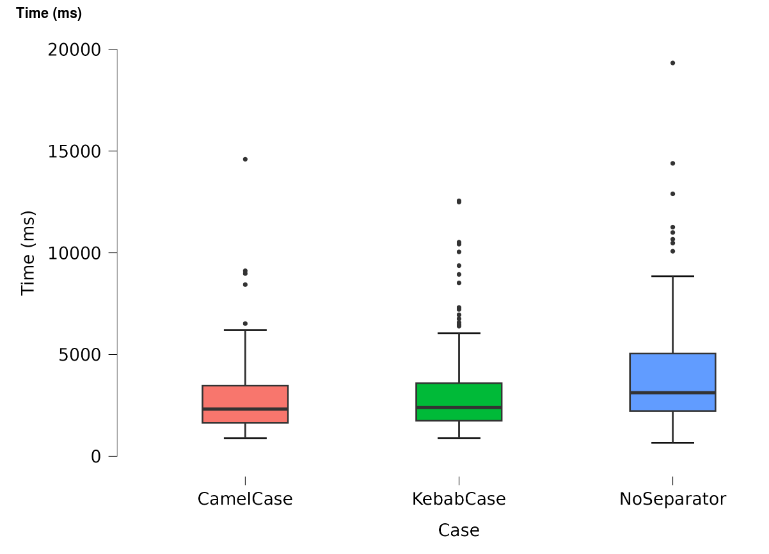
## 3.1 Visual Overview

Provide an insightful overview of the data you collected. This requires some engineering from your part, to find a good degree of summarization: On one end of the spectrum, you don't summarize, and report hundreds of raw measurement values in a block of text. On the other end of the spectrum, you report a single number (like a mean value). Both approaches are bad.

Instead, use appropriate visual summaries (such as **scatter plots**, **histograms**, **box plots**, or **empirical cumulative distribution functions**) to show the distribution of your data. If you have a very small number of measurement values, then report all of them in a **well organized table** (where rows and/or columns correspond to different levels of different factors).

Graphs for the time of each question can be found in Appendix A.

In this section, we believe the most relevant data to show is a plot of the time distribution using the various cases.

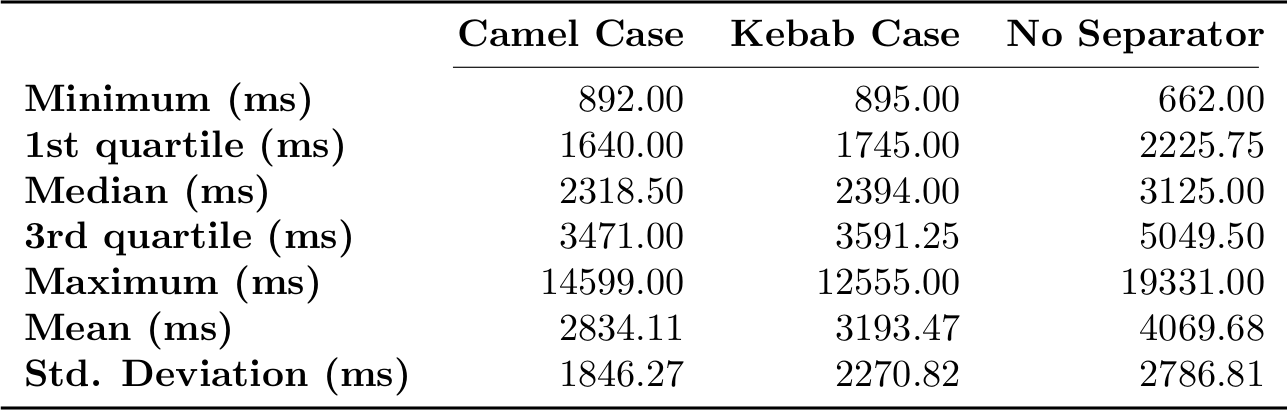


## 3.2 Descriptive Statistics

To provide a comprehensive overview of the time distribution using each separator in various scenarios, we will present five key measurements: the minimum, first quartile, median, third quartile, mean, standard deviation, and maximum. The minimum and maximum values offer insights into the lower and upper bounds of processing times with a specific separator.

The first quartile signifies the value below which 25% of the data points reside, while the third quartile indicates the value below which 75% of the data points fall. The median provides an understanding of the central value within the distribution of processing times.

In addition to these traditional measures, we will include the mean, which represents the arithmetic average of the time distribution. The mean provides a measure of the central tendency, giving an overall sense of the average performance with the chosen separator. Furthermore, the standard deviation, a measure of the dispersion or spread of the data, complements the mean by indicating the extent to which individual processing times deviate from the average.



## 3.3 Inferential Statistics

If applicable, you then follow these up with inferential statistics – i.e., the **results of statistical tests** that you did in order to decide whether there were any “real” (i.e., not by chance) differences between the conditions/groups. You should also explain what statistical test you used, and, if not immediately obvious, why.

In our test, we utilized three inferential statistics to delve deeper into the results.

The t-value is a measure that measures the difference between the means of two groups, considering the variance within each group. A higher t-value signifies a more substantial difference between the means.

The p-value, another crucial measure, represents the probability of obtaining results as extreme as the ones observed, assuming the null hypothesis is true. A lower p-value suggests stronger evidence against the null hypothesis, indicating greater statistical significance.

Finally, we also used the Cohen's d. Cohen's d acts like a ruler for measuring how big the difference is between two groups. It helps us understand how important or practical that difference is. It gives us a standard way to figure out the significance of what we observe. A higher Cohen's d indicates a larger effect size and vice versa. Together, these statistical measures contribute to a comprehensive evaluation of the study's findings.

Of course, we found these measurements for each couple: Camel case – Kebab case, Camel case – no separator, Kebab case – no separator.

|  |  |  |  |
| --- | --- | --- | --- |
|  | t-value | p-value | Cohen’s d |
| Camel Case – Kebab Case | - 1.77 | 0.08 | - 0.14 |
| Camel Case – No Separator | - 5.76 | < 0.001 | - 0. 47 |
| Kebab Case – No Separator | - 4.04 | < 0.001 | - 0.33 |

Make sure you explain – in your words – what these statistics mean “in plain English”, but don’t yet interpret them (this is for the Discussion section).

# 4. Discussion

## 4.1 Compare Hypothesis to Results

Provide a brief restatement of the main results from the previous section, and if (or if not) these support your research hypothesis.

If there is a discrepancy between your hypothesis and the results of your experiment, speculate about why you were unable to find evidence to support your hypothesis.

## 4.2 Limitations and Threats to Validity

Acknowledge any limitations and threats to validity of your study, and how seriously these affect your results. How could these be remedied in future work?

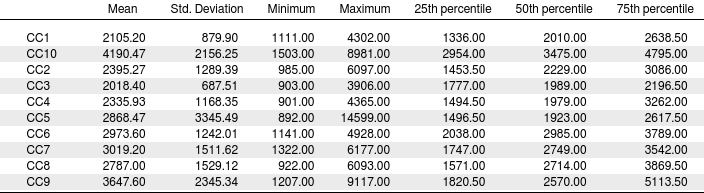
## 4.3 Conclusions

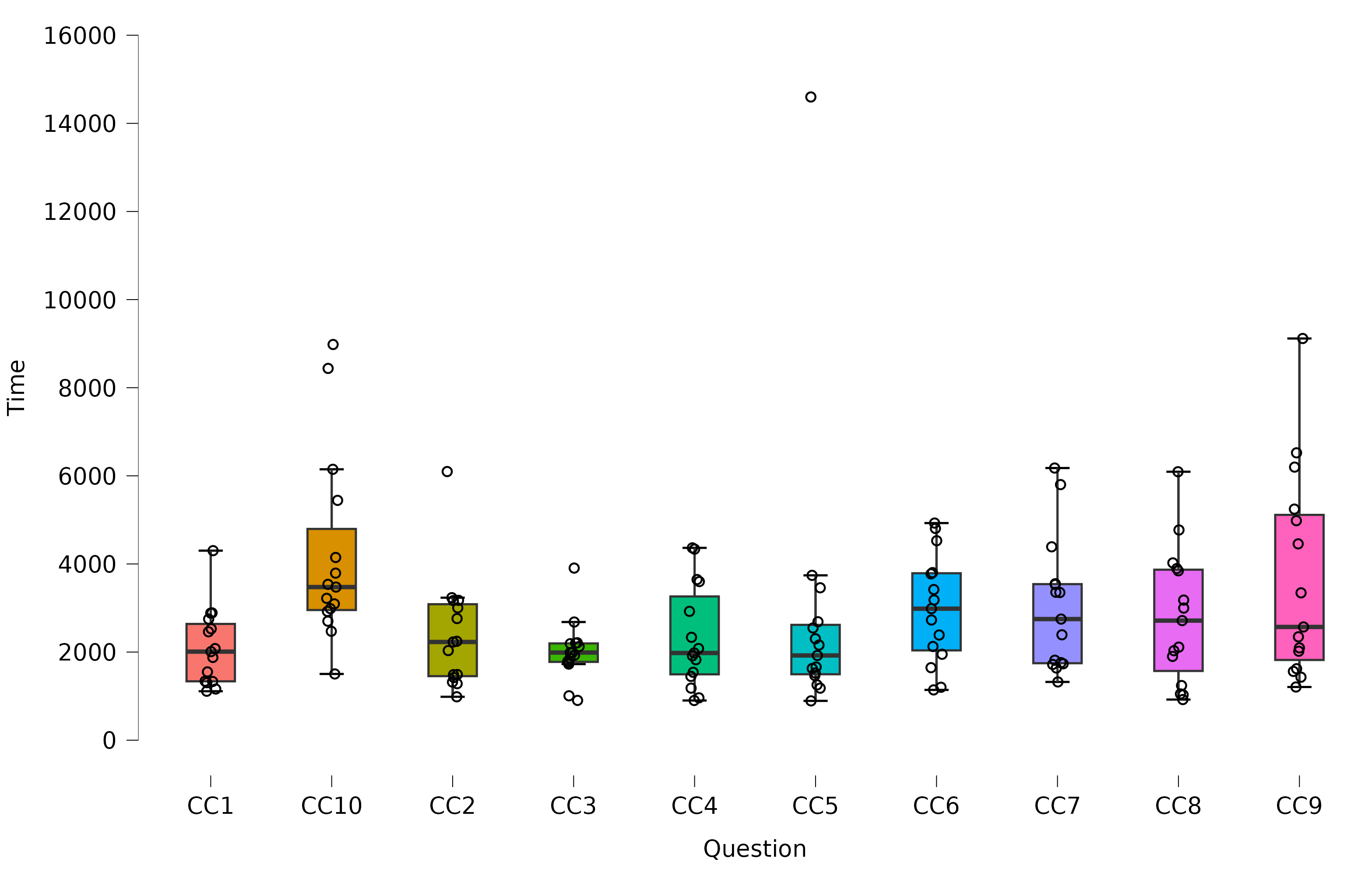
End with the main conclusions that can be drawn from your study.

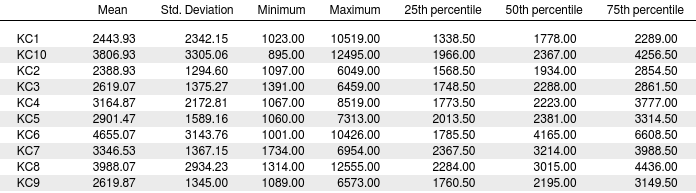
Appendix

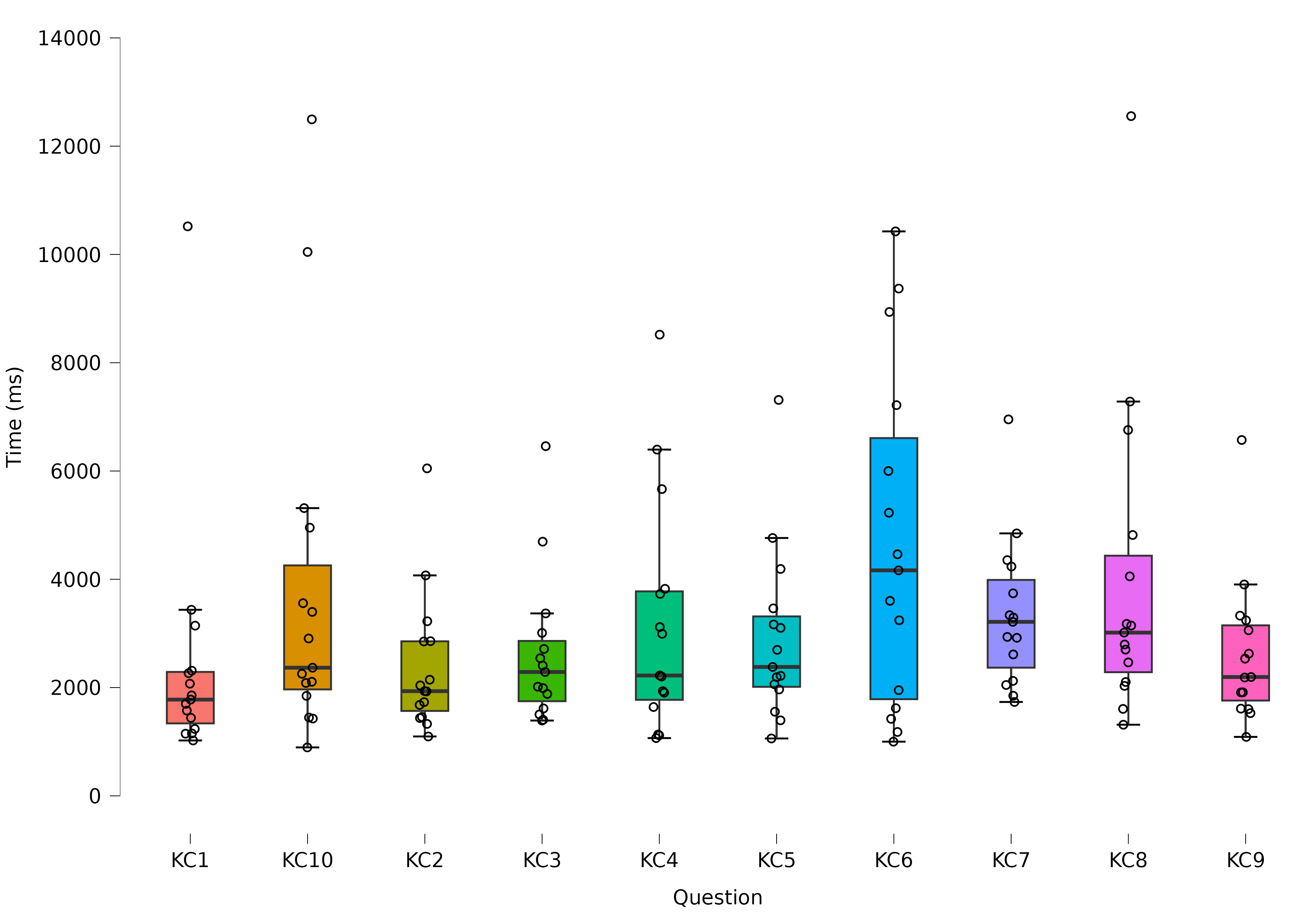
# A. Materials

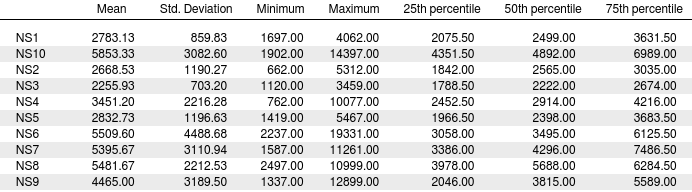
**NOTE: ALL THE VALUES IN GRAPHS AND TABLES ARE IN ms**

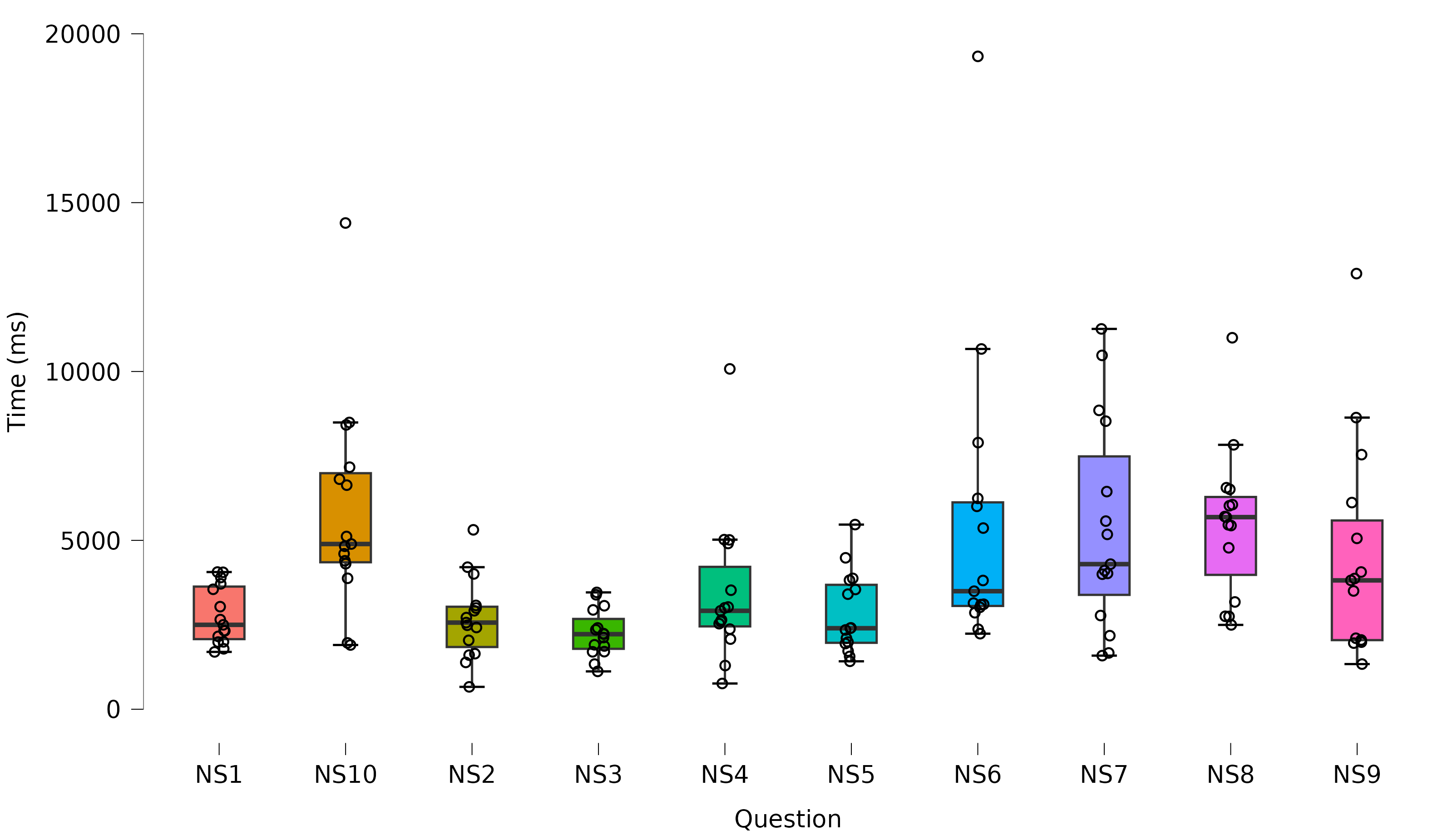
Statistics table per questions in Camel Case (all in ms)

Distribution time-question for Camel Case (ms)

Statistics table per question in Kebab Case (ms)

Distribution time-question for Kebab Case (ms)

Statistics table per question with no separator (ms)

Distribution time-question for no separator (ms)

# B. Reproduction Package (or: Raw Data)

Before, during, and after the experiment you collected all kinds of data. Don't ever throw such data away! Any plots, tables, summaries, and statistics provided in this report should be recreatable from the raw data you have.