How Do Programmers Use Optional Typing? An Empirical Study

Carlos Souza  
Software Engineering Lab  
Federal University of Minas Gerais (UFMG)  
carlosgsouza@gmail.com

Eduardo Figueiredo  
Software Engineering Lab  
Federal University of Minas Gerais (UFMG)  
figueiredo@dcc.ufmg.br

Abstract

The recent popularization of dynamically typed languages, such as Ruby and JavaScript, has brought more attention to the discussion about the impact of typing strategies on development. Types allow the compiler to find type errors earlier and potentially improve the readability and maintainability of code. On the other hand, "untyped" code may be easier to change and require less work from programmers. This paper tries to identify the programmers' point of view about these tradeoffs. An analysis of the source code of 6638 projects written in Groovy, a programming language which features optional typing, shows in which scenarios programmers prefer to type or not to type their declarations. Our results show that types are popular in the definition of module interfaces, but are less used in scripts, test classes and frequently changed code. There is no correlation between the size and age of projects and how their constructs are typed. Finally, we also found evidence that the background of programmers influences how they use types.

Categories and Subject Descriptors  D.2.3 [Software Engineering]: Cooding Tools and Techniques

General Terms  Experimentation, Language

Keywords  Type Systems; Static Analysis; Groovy

1. Introduction

Type systems are one of the most important characteristics of a programming language and also a major topic of research in software engineering [13, 29, 32, 34]. A programming language’s type system determines when the type of an expression is defined [29]. Statically typed languages, such as Java and C#, require programmers to explicitly define the type of a declaration, which can then be used by the compiler to check for type errors. On the other hand, in dynamically typed languages, such as Ruby and JavaScript, the definition of the type of an expression only happens at run time.

Discussions about what is the best type system for a particular situation have become increasingly important in recent years due to the rapid popularization of dynamically typed languages. According to the TIOBE Programming Community Index [35], a well-known ranking that measures the popularity of programming languages, 27% of the programming languages used in industry are dynamically typed. A decade ago, this number was only 17%. Among the 10 languages on top of the ranking, four are dynamically typed: JavaScript, Perl, Python and PHP. None of these languages were among the top 10 rank before 1998.

Several factors may be considered when choosing between a dynamically or statically typed language. Dynamically typed languages tend to allow programmers to code faster and to adapt their programs to frequently changing requirements more easily. Also, by removing the repetitive work of defining types, these languages allow programmers to focus on the problem to be solved rather than on the rules of the language [36].

Statically typed languages also have their advantages. They allow compilers to find type errors statically [24]. Typed declarations increase the maintainability of systems because they implicitly document the code, telling programmers about the nature of expressions [4, 25]. Systems built with these languages tend to be more efficient since they do not need to perform type checking during execution [2, 5]. Finally, modern development environments, such as Eclipse and IDEA, are able to assist programmers with functionalities such as code completion based on the information provided by statically typed declarations [3].

Some languages try to combine characteristics from both static and dynamic type systems. Groovy [18] is one of these languages. Although Groovy is mostly a dynamically typed language, it gives programmers the option to use type annotations as a means to document their code. It is also possible to turn static type checking on so the compiler can find type errors before execution. This allows developers to choose the most appropriate paradigm for each situation.

Understanding the point of view of programmers about the use of types is an important matter. Programming language developers can consider this information in their design so they can develop the most appropriate features for their target audience. Tools can be created or improved to overcome weaknesses of languages. Finally, programmers can benefit from this knowledge when choosing programming languages or typing paradigms for a given context.

This paper presents a large scale empirical study about how programmers use optional typing in Groovy. Through the analysis of a massive dataset with almost seven thousand Groovy projects, we were able to identify when programmers prefer to type or not their declarations. Our results show the point of view of programmers about the use of types. This analysis complements, with a different point of view, existing studies based on controlled experiments [9, 14, 19, 20, 22, 25, 31].

The remainder of this paper is organized as follows. Section 2 introduces the main concepts of the Groovy programming language and Section 3 presents the study settings. Section 4 describes the
results of the study, which are then discussed in Section 5. Threats
to the validity and related work are presented in Sections 6 and 7.
Finally, Section 8 concludes this study and suggests future work.

2. The Groovy Language
Groovy is a dynamically typed programming language designed to
run on the Java Virtual Machine. Its adoption has grown remarkably
over the last years. According to the TIOBE Programming Index,
Groovy is the 22nd most popular language in the software industry
[35], ahead of languages like Haskell and Scala. It builds upon the
strengths of Java, but has additional features inspired by dynamic
languages such as Python and Ruby, such as metaprogramming,
closures and script support. Like Java, Groovy code is compiled
to bytecode, allowing it to seamlessly integrate with existing Java
classes and libraries. These factors have attracted a large number of
Java programmers who want to use Groovy’s dynamic functionality
without having to learn a completely different language or change
the execution platform of their systems.

When Groovy was first launched, in 2007, it was a purely
dynamically typed language. However, it allowed programmers to
optionally type their declarations. Examples of typed and untyped
declarations combined in the same file are shown in Listing 1.
This kind of typing should not be confused with static typing
since the Groovy compiler does not use these type annotations to
look for errors. For example, the snippet of code shown in Listing
2 compiles without any errors. Nevertheless, during runtime, the
string variable references an instance of the Integer class and an
exception is thrown when the method toUpperCase is invoked since
the Integer class does not have such method.

Listing 1 Typed and untyped declarations mixed together
@TypeChecked
class DynamicTyping {
    private String typedField
    private untypedField
    DynamicTyping(typedParam){}
    def untypedMethod(untypedParam, int typedParam){
        def untypedVariable = 1.0
        return untypedVariable
    }
    int typedMethod(){
        String typedVariable = ""
        return typedVariable
    }
}

Listing 2 Types are not checked by default by the Groovy compiler
String string = new Integer(1)
string.toUpperCase()

Since version 2.0, Groovy allows programmers to explicitly
activate static typing by using the @TypeChecked annotation. This
makes Groovy a gradually typed language [15–17, 33, 34]. In this
mode, the Groovy compiler looks for type errors and fails if it finds
any. Listing 3 shows an example of static typing in Groovy. Trying
to compile the class TypeCheckedGroovyClass produces an error
since the method sum is supposed to receive two parameters of the
type int, but it is actually called with two parameters of the type
String.

Listing 3 Forcing the compiler to check types
@TypeChecked
class TypeCheckedGroovyClass {
    static int sum(int a, int b) {
        a + b
    }
    public static void main(String[] args) {
        println sum("1", "2")
    }
}

The @TypeChecked annotation is reasonably recent and most
Groovy programmers still do not use it. Typing annotations on the
other hand are very popular. Although they do not provide static
type checking, they are capable of documenting the code and aiding
in the integration with development tools. In the remainder of this
text, we refer to declarations with type annotations as "typed",
while the word "untyped" is used for declarations with no type
annotations.

3. Study Settings
The study presented in this paper consists in the static analysis of
the source code of a corpus of 6638 Groovy projects. Its goal is
to find in which contexts Groovy programmers type or do not type
their declarations. In this section, we present five research questions
that guide our analysis, the data collection and analysis procedures
and the characterization of the studied dataset.

We have shared all the artifacts of this study in our project
website.1 This includes the source code of the programs used in the
data collection and analysis procedures, the analyzed data (except
the source code of the analyzed projects) and detailed results. Our
goal is to allow this study to be easily replicated or extended.

3.1 Research Questions
We aim to answer the following research questions about the usage
of types by Groovy programmers:

• Question Q1: Do programmers use types more often in the
  interface of their modules? We believe that the benefits of the
  use of types are more clear in declarations that define the
  interface of modules. In such cases, programmers are specifying
  how the rest of the program should interact with a module, po-
  tentially improving readability and maintainability. On the other
  hand, in declarations that are hidden from external modules, we
  expect programmers to opt for the simplicity and flexibility of-
  fered by untyped declarations more often.

• Question Q2: Do programmers use types less often in test
  classes and scripts? Many studies analyze typing paradigms in
  the main classes of a program. However, little is known about
  this question in scripts or test classes. We want to understand
  if programmers consider different typing strategies in these
  scenarios.

• Question Q3: Does the experience of programmers with
  other languages influence their choice for typing their code?
  We believe that programmers familiar with a dynamically typed
  language are more comfortable with the lack of types and end up
  using types less often in Groovy.

1http://github.com/carlosgsouza/groovonomics
3.2 Data Collection Procedure

The projects used in this study were obtained from GitHub, a popular source control service based on Git. For each project, it was necessary to retrieve its source code, metadata, commit history, and the metadata of all of its developers. GitHub does not offer a listing of all hosted projects, but it offers two search mechanisms, a REST API and a Web based search page. Unfortunately, the GitHub API is too limited for our requirements. It imposes a limit of one thousand results and does not allow filtering projects by their programming language.

In order to retrieve an extensive dataset, it was necessary to write a bot to simulate human interactions with the GitHub webpage and search for projects. Some special care was necessary to make this work. For instance, because the number of results is limited to one thousand projects, we had to segment the queries. Multiple requests were made, each of them asking for the name of all projects created on a given month. Results were then combined into a single list. Another problem faced was that GitHub denies excessive requests from the same client. By adding artificial delays between requests, it was possible to overcome this limitation.

With the name of all projects in hand, it was then possible to use the GitHub REST API to query their metadata. That metadata also contains the identifiers of the developers and of the commits of that project. Using those identifiers we once again used the GitHub REST API and obtained the background of all developers and the file changes of all projects.

3.3 Dataset

Our dataset consists of 6638 projects with almost 9.8 million lines of code. Table 1 shows descriptive statistics for the size, age and number of commits of these projects. There are more than 1.5 million declarations of all types and visibilities in our dataset. Note that the value of the median of the number of lines is relatively small. Most projects have 529 lines of code or less. This was expected. Since Groovy is a very expressive language, with many traits of functional programming languages, programmers can write more concise code leading to smaller programs. We want to understand which one of these strategies is actually preferred by Groovy developers.

Table 1. Characterization of Projects

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Sd</th>
<th>Max</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (LoC)</td>
<td>1,471</td>
<td>529</td>
<td>4,545</td>
<td>149,933</td>
<td>9,770,783</td>
</tr>
<tr>
<td>Commits</td>
<td>31</td>
<td>5</td>
<td>175</td>
<td>6,545</td>
<td>203,375</td>
</tr>
<tr>
<td>Age (Days)</td>
<td>361</td>
<td>280</td>
<td>333</td>
<td>1,717</td>
<td>2,395,441</td>
</tr>
</tbody>
</table>

Table 2 shows the number of declarations by kind and visibility. and Table 3 shows the use of visibilities according to kind of declaration. Note that most fields are declared with private visibility while most methods are public. This is explained by the fact that, in Groovy, the default visibilities for fields and methods are private and public respectively.

Table 2. Number of Declarations per Project

<table>
<thead>
<tr>
<th>Kind</th>
<th>Private</th>
<th>Protected</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
<td>346,462</td>
<td>2,996</td>
<td>16,690</td>
</tr>
<tr>
<td>Constructor Parameter</td>
<td>680</td>
<td>246</td>
<td>18,030</td>
</tr>
<tr>
<td>Method Parameter</td>
<td>27,174</td>
<td>10,897</td>
<td>164,546</td>
</tr>
<tr>
<td>Method Return</td>
<td>21,460</td>
<td>28,507</td>
<td>308,030</td>
</tr>
</tbody>
</table>

Table 3. Number of Declarations by Visibility

Figure 1. Most popular languages among Groovy developers
3.4 Analysis

In order to understand where programmers use types, we developed a static code analyzer based on the Groovy metaprogramming library. This analyzer is capable of retrieving the declaration information of parameters and returns of methods, parameters of constructors, fields and local variables. In addition, the analyzer can tell if a declaration is part of a test class or a script and what is its visibility.

A relevant decision we made was not to compile projects, which would require all dependencies to be resolved. This is not feasible given the size of our dataset. Instead, we generated the AST for each file using the CONVERSION phase of the Groovy compiler. At this phase, the compiler has not tried to resolve any dependencies yet, but it is capable of generating an AST with enough information to determine whether a declaration is typed or not. This makes it possible to analyze each Groovy file separately without having to compile the whole project. The downside of this approach is that we cannot analyze Groovy code in conjunction with its dependencies. For example, it is impossible to determine whether programmers tend to type code that interacts with other typed modules since we have not resolved any dependencies to these modules. However, our choice was fundamental in order to execute a study with such an extensive dataset.

4. Results

This section presents the data obtained from our analysis and its statistical treatment. We show how the usage of types varies according to the kind and visibility of declarations, nature of code, programmers’ background, project maturity and frequency of changes. The interpretation of these results is left to Section 5, which answers the proposed research questions (Section 3.1) based on this data and raises new research questions.

4.1 Overall Result

Figure 2 shows a histogram and the descriptive statistics for the relative usage of types in declarations of projects. This value can vary from 0 (a project does not declare any types) to 1 (all declarations of a project are typed). All declarations are considered.

<table>
<thead>
<tr>
<th>Declaration Type</th>
<th>n</th>
<th>mean</th>
<th>median</th>
<th>std. dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
<td>6000</td>
<td>0.43</td>
<td>0.39</td>
<td>0.33</td>
</tr>
<tr>
<td>Constructor Parameter</td>
<td>1670</td>
<td>0.80</td>
<td>1.00</td>
<td>0.35</td>
</tr>
<tr>
<td>Method Parameter</td>
<td>4867</td>
<td>0.67</td>
<td>0.86</td>
<td>0.36</td>
</tr>
<tr>
<td>Method Return</td>
<td>5881</td>
<td>0.68</td>
<td>0.75</td>
<td>0.31</td>
</tr>
<tr>
<td>Local Variable</td>
<td>5845</td>
<td>0.29</td>
<td>0.18</td>
<td>0.32</td>
</tr>
</tbody>
</table>

**Figure 2.** Usage of types in all declarations of all projects

Note that there is a significant number of projects for which the relative usage of types is either approximately 0 or 1. These are mostly small projects. About 95% of them have less than 1000 lines of code and 22% of them have less than 100 lines of code. In such projects, it is easier to be consistent on the typing strategy since there are just a few declarations. We initially considered not including these projects in the rest of our analysis since they could not represent well the entire population of Groovy projects. However, doing so did not alter the results significantly and we decided to include all projects in our analysis regardless of their size. In the rest of this section, this data will be presented in more detail so we can understand which factors lead programmers to use types or not.

### 4.2 Kind of Declaration

This section investigates whether programmers use types differently depending on the kind of the declaration. For each project, we measured the relative usage of types in fields, constructor parameters, method returns, method parameters and local variables. These results are displayed in box plots in Figure 3 along with the corresponding descriptive statistics. Note that the size of each sample, n, is different since not all projects have all types of declarations. For instance, there are only 1670 out of 6638 projects that declare constructor parameters. On the other hand, 6000 projects have declarations of fields.

The results presented in Figure 3 suggest that programmers use types differently depending on the type of a declaration. Local variables, for example, are typed less often. Half of the projects have only 18% or less of their local variables typed. Conversely, methods and constructors are typed in most cases. Note that the median for constructor parameters is equal to 1.00, which means that at least half of the projects with constructor parameters type all declarations of this kind. Since local variables are never part of a module interface, these results suggest a positive answer for Question Q1, i.e, declarations that compose module interfaces are typed more often than other declarations.

The box plot graph and the descriptive statistics are not enough to determine whether the difference in the usage of types in any two kinds of declaration is significant. In order to do that, a significance test should be applied. We start by defining a hypothesis below, which can then be rejected or accepted by this test.

**H0** There is no difference in how programmers type different kinds of declarations

**H1** Programmers type their declarations differently depending on the kind of the declaration

The appropriate significance test should be chosen carefully. It needs to compare multiple treatments, which represent the 5 distinct kinds of declaration. We first considered applying repeated t-tests or Mann-Whitney U-tests in order to compare every two
4.3 Declaration Visibility

This section presents an analysis about how programmers use types according to the visibility of a declaration. We follow the same approach as in the previous section. Figure 4 shows the box plots for the usage of types per declaration visibility along with the descriptive statistics. The ANOVA test reported a $p$ value equal to 0 for these treatments, allowing us to strongly reject the null hypothesis. Finally, the results of the Tukey HSD Test are reported in Table 5. These results show that all treatments are different from each other since all $p$ values are equal to 0.

Protected declarations are those typed most often. Note how skewed is the distribution for these elements in Figure 4. Almost all 2387 projects which use protected visibility in their declarations have all of their protected fields, methods and constructors typed. The confidence intervals reported by the Tukey HSD test show very large differences between these declarations and those with either private or public visibility. Although public declarations are not typed as much, they are also typed very often. At least half of the projects type 75% or more of their declarations. Conversely, private declarations are those with the smallest relative use of types. These results again suggest a positive answer for Question Q1, which hypothesizes that declarations that are part of a module definition are typed more frequently.

<table>
<thead>
<tr>
<th>Declaration Visibility</th>
<th>n</th>
<th>mean</th>
<th>median</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>5852</td>
<td>0.69</td>
<td>0.75</td>
<td>0.29</td>
</tr>
<tr>
<td>Protected</td>
<td>2387</td>
<td>0.93</td>
<td>1.00</td>
<td>0.19</td>
</tr>
<tr>
<td>Private</td>
<td>6023</td>
<td>0.43</td>
<td>0.40</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Figure 4. Usage of types in all declarations by type of declaration

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protected Private</td>
<td>0</td>
<td>(-0.47, 0.52)</td>
</tr>
<tr>
<td>Public Private</td>
<td>0</td>
<td>(0.24, 0.28)</td>
</tr>
<tr>
<td>Public Protected</td>
<td>0</td>
<td>(-0.27, -0.22)</td>
</tr>
</tbody>
</table>

Table 5. Tukey Honest Significant Differences Test results for the comparison between the usage of types by visibility of declaration

4.4 Test Classes and Main Classes

We now analyze the use of types in test classes in comparison to main classes. We used a simple heuristic to determine the kind of the class. In Groovy, like in Java, it is common to organize test classes and main classes in different source folders. The convention adopted by build tools popular among Groovy programmers, such as Gradle and Maven, assumes that test classes and main classes are in the src/test/groovy and src/main/groovy directories respectively. Based on these conventions, we can assume that all classes inside a test directory, but not in a main directory, are test classes.

For every project, we measured the usage of types in test classes and main classes. Script files are not considered in this analysis. We found test classes in 4350 of the 6638 projects in our dataset. Results are displayed in Figure 5 and show the relative usage of...
In order to compare the usage of types in test and main classes, we use a slightly different approach. In this analysis, there are two independent variables, the kind of declaration and the kind of class. Thus, we are required to use Factorial ANOVA, which is the generalization of the One Way ANOVA for multiple factors. Multiple values of $p$ are calculated by this test, each one corresponding to the comparison of treatments according to one of the factors. We report the $p$ value corresponding to the factor representing the kind of class. We also apply the Tukey HSD test, for which results are displayed in Table 6. This time we want to show what is the difference of the relative usage of types between the same kinds of declaration, but in different kinds of classes. For example, the third row of Table 6 shows that there is a significant difference in how programmers type local variables in main and test classes. The difference reported by the confidence interval in Table 6 for this case is $(-0.31, -0.26)$. At least half of the projects type all of their method returns in test classes. Although counterintuitive this result can be easily explained. Automated testing frameworks usually enforce a certain method signature for test methods. JUnit for example, which is used in 2525 of the 4350 projects with test classes, requires test methods to be typed as void. Other popular test frameworks, such as TestNG, have similar requirements. This implies that, in this case, developers type their methods not because they want to, but because they have to.

![Figure 5. Usage of types by declaration type in test classes and main classes.](image)

<table>
<thead>
<tr>
<th>Declaration Type</th>
<th>Class Type</th>
<th>n</th>
<th>mean</th>
<th>median</th>
<th>std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field</td>
<td>Test</td>
<td>1769</td>
<td>0.48</td>
<td>0.47</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Main</td>
<td>5857</td>
<td>0.43</td>
<td>0.39</td>
<td>0.33</td>
</tr>
<tr>
<td>Constructor Parameter</td>
<td>Test</td>
<td>124</td>
<td>0.77</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Main</td>
<td>1623</td>
<td>0.80</td>
<td>1.00</td>
<td>0.34</td>
</tr>
<tr>
<td>Method Parameter</td>
<td>Test</td>
<td>1524</td>
<td>0.34</td>
<td>0.00</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Main</td>
<td>4593</td>
<td>0.71</td>
<td>0.91</td>
<td>0.35</td>
</tr>
<tr>
<td>Method Return</td>
<td>Test</td>
<td>4334</td>
<td>0.85</td>
<td>1.00</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Main</td>
<td>5299</td>
<td>0.54</td>
<td>0.60</td>
<td>0.39</td>
</tr>
<tr>
<td>Local Variable</td>
<td>Test</td>
<td>2842</td>
<td>0.23</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Main</td>
<td>5548</td>
<td>0.30</td>
<td>0.19</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Table 6. Tukey Honest Significant Differences Test results for the comparison between the usage of types by main and test classes

### 4.5 Script Files and Class Files

In Groovy, programmers can write code in the form of scripts, not requiring the definition of classes for simple tasks. This section investigates how programmers type their code in such scripts. Similar to what was done in the previous section, we measured the usage of types in script and class files in all projects and compared the obtained data. We do not consider test classes in this analysis. Determining whether a file corresponds to a script or a class is fairly simple since, in Groovy, scripts are compiled into a class extending `groovy.lang.Script`.

Figure 6 shows the distribution of the relative usage of types in class and script files. Note that constructors and fields are not considered since there is no way to declare those elements in scripts. Also, we do not present an analysis of declarations grouped by visibility since, although allowed, defining the visibility of a declaration inside a script does not make much sense.

The execution of the ANOVA test reported a $p$ value equal to 0, revealing that the declarations are typed significantly different in script files. Table 7 displays the results for the Tukey HSD test, which provide detailed results by the kind of declaration. There is no significant difference on the usage of types in local variables. On the other hand, declarations of parameters or returns of methods are typed much less frequently in scripts. Note however that the value for the last quartile of these declarations is very high, superior to 0.8. This indicates that, although most projects prefer not to use types in method returns, there are a few projects that consistently type most of them.

![Figure 6. Distribution of relative usage of types in class and script files.](image)

Along with the results of the analysis of test classes, the results presented in this section show that the answer for Question Q2 is positive. There are large differences in how programmers type scripts and classes. Although it is not clear from our analysis what the reason for such phenomena, we will raise some hypothesis in our discussion, presented in Section 5.

### 4.6 Programmers’ Background

In this section, we analyze how programmers use types in their declarations according to their backgrounds. Projects are distributed in three groups based on the type system of the languages their

![Image of a table and distribution graph](image)
developers have used on GitHub. The first group comprises those projects of programmers who developed only in statically typed languages, such as Java or C#. The projects of those who developed only in dynamically typed languages, such as Ruby or JavaScript, comprise the second group. Finally, the third group is formed by the projects of those programmers with both dynamically and statically typed languages in their portfolio. We refer to these three groups by the names Static Only, Dynamic Only and Static and Dynamic respectively.

Figures 7 and 8 show results by declaration type and visibility. The p value reported by the ANOVA test is equal to 0, implying that there is a significant difference in how programmers with different backgrounds type their declarations. The results of the Tukey HSD test are reported in Tables 8 and 9. These tables are divided in three parts, each one corresponding to the comparison between the data of two of the three groups.

There are significant differences in the usage of types between all groups. These differences, however, are not as clear as those found in the previous analyses. Let’s start with the comparison between projects in the Static and Dynamic and the Dynamic Only groups. There are significant differences only in private declarations and declarations of fields. Still, these differences are not very large, (0.01, 0.08) for fields and (0.02, 0.09) for private declarations. All in all, these two groups present very similar behavior when typing their declarations, apart from those two exceptions.

The comparison between the Static and the other groups shows more clear differences. Most of the p values reported by the Tukey HSD test are equal to 0. However, constructor parameters and protected declarations never present significant differences. This indicates a strong influence of these types of declarations over the programmers’ behavior. There are other two other cases that do not present significant differences, method returns and public declarations, both in the comparisons between programmers of the Dynamic and Static groups. It is important though to say that the p value reported in these two cases is relatively small, 0.01. These differences are considered not significant only because we are using a very strict confidence level, but would be considered significantly different under a confidence level of 0.05.

The results reported in this section suggest a positive answer for Question Q3, i.e., programmers use types differently depending on their background. This difference is larger when comparing
those projects of the Static group with projects of the other groups. These results are statistically strong, but should be generalized with care. The detailed results reported by the Tukey HSD test reveal many exceptions, specially in the comparison between the groups comprising the projects of those programmers who have at least one dynamically typed language in their portfolio, Static and Dynamic Only. This comparison reveals that, for all kinds and visibilities of declarations, the difference is either not significant or small.

### 4.7 Project Size, Age and Number of Commits

This section investigates whether programmers use types differently in their code depending on the project characteristics. We analyze three project metrics: age, number of lines of code and number of commits. We start by analyzing the correlation between these metrics and the relative use of types in declarations by type and visibility. The Spearman rank correlation coefficient is used for this purpose. This coefficient, which ranges from -1 to 1, is a measure of the dependence between two variables. A positive value means that two variables are correlated, i.e., as the value of one grows, so does the value of the other. A negative value means an inverse correlation. Values close to 1 or -1 indicate very strong relationships and values above 0.5 or below -0.5 can be considered strong correlations.

Table 10 shows the Spearman correlation coefficient between the usage of types and the size, age and number of commits of projects.

<table>
<thead>
<tr>
<th>Table 9. Tukey Honest Significant Differences Test results for the comparison between the usage of types by declaration visibility and programmers background</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Static and Dynamic vs. Static</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Dynamic vs. Static</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Static and Dynamic vs. Dynamic</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8. Tukey Honest Significant Differences Test results for the comparison between the usage of types by declaration type and programmers background</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Field</td>
</tr>
<tr>
<td>Constructor Parameter</td>
</tr>
<tr>
<td>Method Parameter</td>
</tr>
<tr>
<td>Method Return</td>
</tr>
<tr>
<td>Local Variable</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Protected</td>
</tr>
<tr>
<td>Private</td>
</tr>
</tbody>
</table>

Table 10. Spearman Correlation between the usage of types and the size, age and number of commits of projects.
is 100 days old or more and has, at least, 2KLoC and 100 commits. These numbers were defined by manually inspecting our dataset and finding that there are popular and mature projects that barely exceed these three metrics. According to our criteria, there are 223 mature projects in our dataset, which are characterized in Table 11.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Max</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (LoC)</td>
<td>9947</td>
<td>5627</td>
<td>14594</td>
<td>14993</td>
<td>2218189</td>
</tr>
<tr>
<td>Commits</td>
<td>487</td>
<td>213</td>
<td>800</td>
<td>6545</td>
<td>108583</td>
</tr>
<tr>
<td>Age (Days)</td>
<td>600</td>
<td>574</td>
<td>350</td>
<td>1469</td>
<td>133697</td>
</tr>
</tbody>
</table>

Table 11. Descriptive statistics for mature projects

Figures 9 and 10 show the box plots for the usage of types in mature projects and others by declaration type and visibility respectively. The ANOVA test reported a p value equal to 0.0518 for this analysis, which implies that there are no treatments that are significantly different from each other. Because of this, we do not report the results for the Tukey HSD test. This result, along with the very low correlation coefficients reported in Table 10, implies that the answer for Question Q4 is negative. There is no significant difference in how programmers type declarations in mature projects compared to declarations in other projects.

4.8 Frequency of changes

This section investigates whether programmers prefer to type their declarations in frequently changed code or not. Only the mature projects defined in the previous section are considered since we would not be able to obtain meaningful results from small and young projects.

We found that in most of the mature projects programmers prefer untyped declarations in frequently changed code. We calculated the Spearman correlation coefficient between the frequency of changes of a file and the usage of types in that file for all mature projects. In projects where types are used more often in frequently changed files, this coefficient is positive, and negative when untyped declarations are preferred. Figure 11 displays the cumulative distribution of this coefficient across the mature projects dataset. It shows that 65% of them present a negative Spearman correlation coefficient and that almost half of these present strong correlations, i.e., inferior to -0.5. On the other hand, only 10% of the mature projects present strong positive correlations.

5. Discussion

In this section, we discuss the results of our study in the light of the research questions proposed in Section 3.1. Although we were able to obtain a good understanding of the usage of types in different
contexts, the cause for such results is still unclear in many of them. We provide several hypotheses with the goal of identifying future research topics that can provide more detailed insights about such causes.

**Q1: Do programmers use types more often in the interface of their modules?**

The analysis of the usage of types by kind and visibility of declarations, presented in Sections 4.2 and 4.3, provides clear evidence that the answer for Q1 is affirmative. Private declarations are typed less often than protected and public declarations. Also, fields, methods, and constructors are typed more often than local variables. Although fields are significantly less typed than methods and constructors, this can be explained by the fact that most fields are declared privately as shown in Section 3.3. In Groovy, similar to what happens with Java, interactions with fields of other modules usually happen through accessor methods.

While it is clear that module definitions are typed more often, the cause for such phenomena is still open to discussion. We believe that the main motivation for this is the implicit documentation provided by types. In these scenarios, types provide useful hints about the behavior of modules [8] and define pre and post condition of contracts [11, 13, 26, 27, 30, 37]. Users of a well defined module learn how to use it faster and do not need to read its implementation to understand how to use it. Programmers may consider that delicate contracts, such as those defined by protected methods, should be well documented and thus are typed more often.

Constructor parameters and protected declarations are always the elements typed most often. In addition, the analyses of different contexts, i.e., main and test classes, mature and non-mature projects and programmers’ backgrounds, show no significant difference in the usage of types. This reveals that the presence of these particular declarations is considered by programmers as a more important factor in their choice whether to type a declaration than these contexts. We hypothesize that programmers consider documentation more important in these elements. Constructors usually define the dependencies of an object, at least for its creation. In addition, they might be the first element that a programmer interacts with when dealing with a new module. Protected declarations are often used as a means to delegate the implementation of a method to subclasses, which requires a well defined contract so the superclass can work properly. Moreover, they give subclasses and other classes in the same package access to internal elements of a class, forming a tightly coupled relationship [6].

We can also speculate that declarations that are not part of a module definition, which are local variables and those with private visibility, require less documentation and thus are typed less often. Programmers can easily find all the references to these elements. Local variables are only used inside the block of code where they are declared, while all the references to elements declared privately are in the same file. This makes it easier for a programmer to infer the type of such a declaration even when it is not explicitly defined.

Documentation may not be the only reason why programmers type declarations in modules interfaces. We can think of at least two other reasons. First, a programmer might type a declaration so that he or she can get code assistance from the development environment. For example, typing the declaration of a method parameter allows the development environment to provide code completion for that parameter inside the method. Another possibility is that, even though Groovy is actually a dynamically typed language, programmers might type their declarations thinking that the compiler will check for type errors, which would lead to safer interactions between modules.

**Q2: Do programmers use types less often in test classes and scripts?**

There are notable differences between the usage of types in either test classes or scripts and the main classes of a program. Sections 4.4 and 4.5 show that, in these scenarios, programmers use types less often. If we are right about our hypothesis that programmers type their modules as a means to document their code, this could explain this less frequent use of types. Scripts and test classes are usually not designed as reusable modules. Test classes have the sole goal of verifying a program’s functionality and not interfering with it while scripts cannot be instantiated or referenced by other modules. In these scenarios, programmers might perceive documentation as less important. It is curious however that test code itself is usually perceived as a form of documentation [1, 28]. Because of this, we were expecting programmers to actually use more types in test classes as a means to improve the documentation they provide. Perhaps, although programmers use test classes as documentation, they might not write them with this goal in mind.

An alternative explanation for the fact that scripts and tests are typed less frequently is that most of them are probably simpler than main classes. As found in the recent work of Hanenberg et al. [20], untyped code potentially has a positive impact on the development time of easier tasks. In such case, programmers might not type their declarations in scripts or test classes since this would allow them to finish their tasks faster.

**Q3: Does the experience of programmers with other languages influence their choice for typing their code?**

The analysis presented in Section 4.6 indicates that the answer for Question Q3 is affirmative. The choice for using types on a language with optional typing, such as Groovy, is in fact influenced by the programmers’ experience with other languages. In general, those programmers who have only statically typed languages in their portfolio type more often than the others. However, the two groups with projects written by programmers who have either statically and dynamically typed languages or only dynamically typed languages are very similar in most cases. Apparently programmers who develop in an “untyped” language get used to the lack of types, leading them to declare types less often. This hypothesis supports the work of Daly et al. which suggests that programmers have ways of reasoning about types that compensate for the lack of static type information [9].

**Q4: Does the size, age or level of activity of a project have any influence on the usage of types?**

We initially believed that, as these metrics grow, the maintenance of projects becomes more difficult, leading programmers to use more types as a means to make code more readable. However, the analysis presented in Section 4.7 shows no evidence of such behavior. We consider two hypotheses in order to explain these results. First, the considered metrics might not actually correlate to the need for maintenance of projects. Second, even if these metrics are a good indicative of the necessity of better maintainability, once projects start growing and aging, programmers might not have the opportunity or desire to make their code more maintainable.

**Q5: In frequently changed code, do developers prefer typed or untyped declarations?**

In frequently changed code, there are arguments in favor and against using types. Since types act as documentation, programmers might use them to make code more maintainable and easier to change [24]. On the other hand, untyped code is simpler and can potentially be changed faster [32]. The results presented in Section 4.8 however suggests that the latter is considered more often than the former.
In most projects, the usage of untyped declarations grows as the frequency of changes in a file increases. Apparently, programmers understand that untyped code makes maintenance tasks easier. One can argue that the causal relationship is the opposite, i.e., these files have to be changed more often due to the use of untyped declarations. However, it is very unlikely that programmers would not notice that untyped declarations would be causing such problems and would not add types to those declarations in order to fix them.

6. Threats to Validity

In this section, we discuss potential threats to validity of our study. As usual, we have arranged possible threats in two categories, internal and external validity [38].

Internal Validity

Perhaps the most relevant internal threat to our study is that in a large scale empirical study such as ours, there might be many confounds which are difficult to identify. In Section 4.6 we consider that the GitHub portfolio of a programmer represents well his experience with other languages and type systems, but this might not be true for all programmers. They may have projects in their portfolio that they have not worked on or projects hosted elsewhere written in other languages. There is also the possibility of a programmer having multiple GitHub accounts with different languages in each one, causing such a programmer to be measured twice with different inferred backgrounds. Due to the large number of programmers considered in our study, we expect these special cases not to have a large influence on the results.

There are other factors that might have influenced programmers besides the ones considered in this study. Some frameworks require programmers to use typed or untyped declarations in some cases. For example, we found that the data collected in test classes is biased by the fact that popular testing frameworks, such as JUnit, require test methods to have their returns declared as void. There might be other similar cases that we are not aware of.

External Validity

Although we have analyzed a very extensive number of Groovy projects, it cannot be said that we have covered all possible scenarios. By manually inspecting our dataset, we could find only a few projects with characteristics of software developed inside an organization. Most of them were developed by small groups of people or open source communities. Enterprise projects are probably hosted privately on GitHub or in private servers, and hence unavailable to us.

The behavior observed for Groovy projects can be very different in other languages. Most languages are not like Groovy and feature either static or dynamic typing, forcing programmers to choose a single typing strategy for all scenarios in a single project. Even a language with a hybrid typing paradigm might implement different strategies which will be perceived differently by the programmers of that language. Finally, the tools used to code in a given language might influence programmers to choose different type strategies.

7. Related Work

There are multiple studies in the literature that compare different typing strategies. Although we are not aware of any studies that analyze this question using a large scale case study as ours, we know of multiple controlled experiments with significant results. One of the goals of this paper is to complement these results by providing a different point of view to their analyses.

In a recent experiment, Hanenberg et al. [20] studied the impact of the use of types on the development time of programmers while performing tasks on an undocumented API. The experiment divided 27 people in two groups. They developed in two languages, one statically typed, Java, and the other dynamically typed, Groovy. Results revealed a positive impact of the use of types when these were used to document design decisions or when a high number of classes had to be identified by programmers. On the other hand, for easier tasks, programmers developed faster in the dynamically typed language. Our analysis suggest that programmers are aware of these tradeoffs and consider them when choosing whether or not to type their declarations. In potentially simpler scenarios, such as scripts and tests, we show a lower usage of types. Conversely, programmers type the interface of their modules very often, probably as a means to document their code, which potentially improves the development time in more complex scenarios.

Daly et al. conduct an experiment in order to compare the performance of two groups working on small development tasks [9]. One group used Ruby, a dynamically typed language, while the other used DRuby, a statically typed version of Ruby. Results showed that the DRuby compiler rarely managed to capture any errors that were not already evident for programmers. Most subjects involved in the study had previous experience with Ruby, which suggests that programmers get used to the lack of typing. Our analysis of the programmers’ backgrounds supports this argument. It shows that those programmers who have worked with dynamically typed languages in fact use types less often.

8. Conclusions and Future Work

In this paper we conducted a large scale case study in order to investigate how types are used in Groovy. Our goal is to find what are the point of view of programmers about what typing strategies are more suitable to which contexts. Our main findings are:

- Groovy programmers type declarations that define the interface of modules more often than other declarations
- Types are less popular in test classes and script files
- Those programmers who have developed in at least one dynamically typed language use types less frequently than those who have only worked with statically typed languages
- Apparently, there is no influence of the size, age or level of activity of a project on how programmers use types
- In most projects, the files that change more frequently are also those files with a lower usage of types

We believe that these results are valuable to the developers of programming languages and development tools who can base their designs on real user data. Also, programmers can understand the tradeoffs between using or not types in their projects. Our results provide a different point of view and complement previous studies, which analyzed typing strategies through the use of controlled experiments. We hope that the questions raised in our discussion can inspire other researchers to analyze our findings in more detail.

In future work we want to conduct controlled experiments and qualitative studies in order to evaluate some of our results in more details. In particular, we have started a new study with the goal of finding what is the impact of the use of the documentation provided by types and the documentation provided by unit tests on the development time of maintenance tasks. Another work in progress is a framework for the static source code analysis of massive datasets called Elastic Repository Analysis - ERA2. This framework is based on the artifacts we built to analyze the Groovy projects in this study. ERA will help researchers retrieve large numbers of projects from GitHub and process these projects quickly on the cloud using Amazon Web Services.

2 http://github.com/carlosgsouza/ERA
Acknowledgments

Many thanks to Stefan Hanenberg for the inspiration and the invaluable contributions.
This work was partially supported by FAPEMIG, grants APQ-02376-11 and APQ-02532-12, and CNPq grant 485907/2013-5.

References