[Vectors 2](#_Toc493070547)

[Creating a New Vector 2](#_Toc493070548)

[Updating a Vector 3](#_Toc493070549)

[Dropping a Vector Drops Its Elements 3](#_Toc493070550)

[Reading Elements of Vectors 4](#_Toc493070551)

[Invalid References 5](#_Toc493070552)

[Using an Enum to Store Multiple Types 6](#_Toc493070553)

[Strings 7](#_Toc493070554)

[What Is a String? 8](#_Toc493070555)

[Creating a New String 8](#_Toc493070556)

[Updating a String 10](#_Toc493070557)

[Appending to a String with push\_str and push 10](#_Toc493070558)

[Concatenation with the + Operator or the format! Macro 11](#_Toc493070559)

[Indexing into Strings 12](#_Toc493070560)

[Internal Representation 13](#_Toc493070561)

[Bytes and Scalar Values and Grapheme Clusters! Oh My! 14](#_Toc493070562)

[Slicing Strings 14](#_Toc493070563)

[Methods for Iterating Over Strings 15](#_Toc493070564)

[Strings Are Not So Simple 16](#_Toc493070565)

[Hash Maps 16](#_Toc493070566)

[Creating a New Hash Map 17](#_Toc493070567)

[Hash Maps and Ownership 18](#_Toc493070568)

[Accessing Values in a Hash Map 19](#_Toc493070569)

[Updating a Hash Map 20](#_Toc493070570)

[Overwriting a Value 20](#_Toc493070571)

[Only Insert If the Key Has No Value 20](#_Toc493070572)

[Updating a Value Based on the Old Value 21](#_Toc493070573)

[Hashing Function 22](#_Toc493070574)

[Summary 22](#_Toc493070575)

CE: In many cases we are using the term “parameter” rather than “argument”--- if you see “argument” used can you please flag it for the author to confirm whether it’s correct

8

Common Collections

Rust’s standard library includes a number of very useful data structures called collections. Most other data types represent one specific value, but collections can contain multiple values. Unlike the built-in array and tuple types, the data these collections point to is stored on the heap, which means the amount of data does not need to be known at compile time and can grow or shrink as the program runs. Each kind of collection has different capabilities and costs, and choosing an appropriate one for your current situation is a skill you’ll develop over time. In this chapter, we’ll discuss three collections that are used very often in Rust programs:

A vector allows us to store a variable number of values next to each other.

A string is a collection of characters. We’ve discussed the String type previously, but in this chapter we’ll talk about it in depth.

A hash map allows us to associate a value with a particular key. It’s a particular implementation of the more general data structure called a map.

To learn about the other kinds of collections provided by the standard library, see the documentation at https://doc.rust-lang.org/stable/std/collections/.

We’ll discuss how to create and update vectors, strings, and hash maps, as well as what makes each special.

Vectors

The first collection type we’ll look at is Vec<T>, also known as a vector. Vectors allow us to store more than one value in a single data structure that puts all the values next to each other in memory. Vectors can only store values of the same type. They are useful in situations in which you have a list of items, such as the lines of text in a file or the prices of items in a shopping cart.

Creating a New Vector

To create a new, empty vector, we can call the Vec::new function as shown in Listing 8-1:

let v: Vec<i32> = Vec::new();

Listing 8-1: Creating a new, empty vector to hold values of type i32

Note that we added a type annotation here. Because we aren’t inserting any values into this vector, Rust doesn’t know what kind of elements we intend to store. This is an important point. Vectors are implemented using generics; we’ll cover how to use generics with your own types in Chapter 10. For now, know that the Vec<T> type provided by the standard library can hold any type, and when a specific vector holds a specific type, the type is specified within angle brackets. In Listing 8-1, we’ve told Rust that the Vec<T> in v will hold elements of the i32 type.

prod: check xref

In more realistic code, Rust can often infer the type of value we want to store once we insert values, so you rarely need to do this type annotation. It’s more common to create a Vec that has initial values, and Rust provides the vec! macro for convenience. The macro will create a new Vec that holds the values we give it. Listing 8-2 creates a new Vec<i32> that holds the values 1, 2, and 3:

let v = vec![1, 2, 3];

Listing 8-2: Creating a new vector containing values

Because we’ve given initial i32 values, Rust can infer that the type of v is Vec<i32>, and the type annotation isn’t necessary. Next, we’ll look at how to modify a vector.

Updating a Vector

To create a vector and then add elements to it, we can use the push method as shown in Listing 8-3:

let mut v = Vec::new();

v.push(5);

v.push(6);

v.push(7);

v.push(8);

Listing 8-3: Using the push method to add values to a vector

As with any variable, as discussed in Chapter 3, if we want to be able to change its value, we need to make it mutable using the mut keyword. The numbers we place inside are all of type i32, and Rust infers this from the data, so we don’t need the Vec<i32> annotation.

prod: check xref

Dropping a Vector Drops Its Elements

Like any other struct, a vector will be freed when it goes out of scope, as annotated in Listing 8-4:

{

let v = vec![1, 2, 3, 4];

// do stuff with v

} // <- v goes out of scope and is freed here

Listing 8-4: Showing where the vector and its elements are dropped

When the vector gets dropped, all of its contents will also be dropped, meaning those integers it holds will be cleaned up. This may seem like a straightforward point but can get a bit more complicated when we start to introduce references to the elements of the vector. Let’s tackle that next!

Reading Elements of Vectors

Now that you know how to create, update, and destroy vectors, knowing how to read their contents is a good next step. There are two ways to reference a value stored in a vector. In the examples, we’ve annotated the types of the values that are returned from these functions for extra clarity.

Listing 8-5 shows both methods of accessing a value in a vector either with indexing syntax or the get method:

let v = vec![1, 2, 3, 4, 5];

let third: &i32 = &v[2];

let third: Option<&i32> = v.get(2);

Listing 8-5: Using indexing syntax or the get method to access an item in a vector

Note two details here. First, we use the index value of 2 to get the third element: vectors are indexed by number, starting at zero. Second, the two different ways to get the third element are by using & and [], which gives us a reference, or by using the get method with the index passed as an argument, which gives us an Option<&T>.

The reason Rust has two ways to reference an element is so you can choose how the program behaves when you try to use an index value that the vector doesn’t have an element for. As an example, what should a program do if it has a vector that holds five elements and then tries to access an element at index 100, as shown in Listing 8-6:

let v = vec![1, 2, 3, 4, 5];

let does\_not\_exist = &v[100];

let does\_not\_exist = v.get(100);

Listing 8-6: Attempting to access the element at index 100 in a vector containing 5 elements

When you run this code, the first [] method will cause a panic! because it references a nonexistent element. This method is best used when you want your program to consider an attempt to access an element past the end of the vector to be a fatal error that crashes the program.

When the get method is passed an index that is outside the vector, it returns None without panicking. You would use this method if accessing an element beyond the range of the vector happens occasionally under normal circumstances. Your code will then have logic to handle having either Some(&element) or None, as discussed in Chapter 6. For example, the index could be coming from a person entering a number. If they accidentally enter a number that’s too large and the program gets a None value, you could tell the user how many items are in the current Vec and give them another chance to enter a valid value. That would be more user-friendly than crashing the program due to a typo!

prod: check xref

Invalid References

When the program has a valid reference, the borrow checker enforces the ownership and borrowing rules (covered in Chapter 4) to ensure this reference and any other references to the contents of the vector remain valid. Recall the rule that states we can’t have mutable and immutable references in the same scope. That rule applies in Listing 8-7 where we hold an immutable reference to the first element in a vector and try to add an element to the end:

prod: check xref

let mut v = vec![1, 2, 3, 4, 5];

let first = &v[0];

v.push(6);

Listing 8-7: Attempting to add an element to a vector while holding a reference to an item

Compiling this code will result in this error:

error[E0502]: cannot borrow `v` as mutable because it is also borrowed as immutable

|

4 | let first = &v[0];

| - immutable borrow occurs here

5 |

6 | v.push(6);

| ^ mutable borrow occurs here

7 | }

| - immutable borrow ends here

The code in Listing 8-7 might look like it should work: why should a reference to the first element care about what changes at the end of the vector? The reason behind this error is due to the way vectors work: adding a new element onto the end of the vector might require allocating new memory and copying the old elements to the new space if there isn’t enough room to put all the elements next to each other where the vector was. In that case, the reference to the first element would be pointing to deallocated memory. The borrowing rules prevent programs from ending up in that situation.

Note: For more on the implementation details of the Vec<T> type, see “The Nomicon” at <https://doc.rust-lang.org/stable/nomicon/vec.html>.

Iterating Over the Values in a Vector

If we want to access each element in a vector in turn, rather than using indexing to access one element, we can iterate through all of the elements. Listing 8-8 shows how to use a for loop to get immutable references to each element in a vector of i32 values and print them out:

let v = vec![100, 32, 57];

for i in &v {

println!("{}", i);

}

Listing 8-8: Printing each element in a vector by iterating over the elements using a for loop

We can also iterate over mutable references to each element in a mutable vector if we want to make changes to all the elements. The for loop in Listing 8-9 will add 50 to each element:

let mut v = vec![100, 32, 57];

for i in &mut v {

\*i += 50;

}

Listing 8-9: Iterating over mutable references to elements in a vector

In order to change the value that the mutable reference refers to, before we can use the += operator with i, we have to use the dereference operator (\*) to get to the value.

Using an Enum to Store Multiple Types

At the beginning of this chapter, we said that vectors can only store values that are the same type. This can be inconvenient; there are definitely use cases for needing to store a list of items of different types. Fortunately, the variants of an enum are defined under the same enum type, so when we need to store elements of a different type in a vector, we can define and use an enum!

For example, let’s say we want to get values from a row in a spreadsheet where some of the columns in the row contain integers, some floating-point numbers, and some strings. We can define an enum whose variants will hold the different value types, and then all the enum variants will be considered the same type, that of the enum. Then we can create a vector that holds that enum and so, ultimately, holds different types. We’ve demonstrated this in Listing 8-8:

enum SpreadsheetCell {

Int(i32),

Float(f64),

Text(String),

}

let row = vec![

SpreadsheetCell::Int(3),

SpreadsheetCell::Text(String::from("blue")),

SpreadsheetCell::Float(10.12),

];

Listing 8-8: Defining an enum to store values of different types in one vector

The reason Rust needs to know what types will be in the vector at compile time is so it knows exactly how much memory on the heap will be needed to store each element. A secondary advantage is that we can be explicit about what types are allowed in this vector. If Rust allowed a vector to hold any type, there would be a chance that one or more of the types would cause errors with the operations performed on the elements of the vector. Using an enum plus a match expression means that Rust will ensure at compile time that we always handle every possible case, as discussed in Chapter 6.

prod: confirm xref

If you don’t know when you’re writing a program the exhaustive set of types the program will get at runtime to store in a vector, the enum technique won’t work. Instead, you can use a trait object, which we’ll cover in Chapter 17.

prod: check xref

Now that we’ve discussed some of the most common ways to use vectors, be sure to review the API documentation for all the many useful methods defined on Vec by the standard library. For example, in addition to push, a pop method removes and returns the last element. Let’s move on to the next collection type: String!

Strings

We talked about strings in Chapter 4, but we’ll look at them in more depth now. New Rustaceans commonly get stuck on strings due to a combination of three concepts: Rust’s propensity for exposing possible errors, strings being a more complicated data structure than many programmers give them credit for, and UTF-8. These concepts combine in a way that can seem difficult when you’re coming from other programming languages.

prod: check xref

This discussion of strings is in the collections chapter because strings are implemented as a collection of bytes plus some methods to provide useful functionality when those bytes are interpreted as text. In this section, we’ll talk about the operations on String that every collection type has, such as creating, updating, and reading. We’ll also discuss the ways in which String is different than the other collections, namely how indexing into a String is complicated by the differences between how people and computers interpret String data.

What Is a String?

We’ll first define what we mean by the term string. Rust has only one string type in the core language, which is the string slice str that is usually seen in its borrowed form &str. In Chapter 4, we talked about string slices, which are references to some UTF-8 encoded string data stored elsewhere. String literals, for example, are stored in the binary output of the program and are therefore string slices.

prod: check xref

The String type is provided in Rust’s standard library rather than coded into the core language and is a growable, mutable, owned, UTF-8 encoded string type. When Rustaceans refer to “strings” in Rust, they usually mean the String and the string slice &str types, not just one of those types. Although this section is largely about String, both types are used heavily in Rust’s standard library and both String and string slices are UTF-8 encoded.

Rust’s standard library also includes a number of other string types, such as OsString, OsStr, CString, and CStr. Library crates can provide even more options for storing string data. Similar to the \*String/\*Str naming, they often provide an owned and borrowed variant, just like String/&str. These string types can store text in different encodings or be represented in memory in a different way, for example. We won’t discuss these other string types in this chapter; see their API documentation for more about how to use them and when each is appropriate.

Creating a New String

Many of the same operations available with Vec are available with String as well, starting with the new function to create a string, shown in Listing 8-9:

let mut s = String::new();

Listing 8-9: Creating a new, empty String

This line creates a new empty string called s that we can then load data into. Often, we’ll have some initial data that we want to start the string with. For that, we use the to\_string method, which is available on any type that implements the Display trait, which string literals do. Listing 8-10 shows two examples:

let data = "initial contents";

let s = data.to\_string();

// the method also works on a literal directly:

let s = "initial contents".to\_string();

Listing 8-10: Using the to\_string method to create a String from a string literal

This code creates a string containing initial contents.

We can also use the function String::from to create a String from a string literal. The code in Listing 8-11 is equivalent to the code from Listing 8-10 that uses to\_string:

let s = String::from("initial contents");

Listing 8-11: Using the String::from function to create a String from a string literal

Because strings are used for so many things, we can use many different generic APIs for strings, providing us with a lot of options. Some of them can seem redundant, but they all have their place! In this case, String::from and to\_string do the same thing, so which you choose is a matter of style.

Remember that strings are UTF-8 encoded, so we can include any properly encoded data in them, as shown in Listing 8-12:

let hello = String::from("السلام عليكم");

let hello = String::from("Dobrý den");

let hello = String::from("Hello");

let hello = String::from("שָׁלוֹם");

let hello = String::from("नमस्ते");

let hello = String::from("こんにちは");

let hello = String::from("안녕하세요");

let hello = String::from("你好");

let hello = String::from("Olá");

let hello = String::from("Здравствуйте");

let hello = String::from("Hola");

Listing 8-12: Storing greetings in different languages in strings

All of these are valid String values.

Updating a String

A String can grow in size and its contents can change, just like the contents of a Vec, by pushing more data into it. In addition, we can conveniently use the + operator or the format! macro to concatenate String values together.

Appending to a String with push\_str and push

We can grow a String by using the push\_str method to append a string slice, as shown in Listing 8-13:

let mut s = String::from("foo");

s.push\_str("bar");

Listing 8-13: Appending a string slice to a String using the push\_str method

After these two lines, s will contain foobar. The push\_str method takes a string slice because we don’t necessarily want to take ownership of the parameter. For example, the code in Listing 8-14 shows that it would be unfortunate if we weren’t able to use s2 after appending its contents to s1:

let mut s1 = String::from("foo");

let s2 = "bar";

s1.push\_str(&s2);

println!("s2 is {}", s2);

Listing 8-14: Using a string slice after appending its contents to a String

If the push\_str method took ownership of s2, we wouldn’t be able to print out its value on the last line. However, this code works as we’d expect!

The push method takes a single character as a parameter and adds it to the String. Listing 8-15 shows code that adds an l to a String using the push method:

let mut s = String::from("lo");

s.push('l');

Listing 8-15: Adding one character to a String value using push

As a result of this code, s will contain lol.

Concatenation with the + Operator or the format! Macro

Often, we’ll want to combine two existing strings. One way is to use the + operator, as shown in Listing 8-16:

let s1 = String::from("Hello, ");

let s2 = String::from("world!");

let s3 = s1 + &s2; // Note that s1 has been moved here and can no longer be used

Listing 8-16: Using the + operator to combine two String values into a new String value

As a result of this code, the string s3 will contain Hello, world!. The reason s1 is no longer valid after the addition and the reason we used a reference to s2 has to do with the signature of the method that gets called when we use the + operator. The + operator uses the add method, whose signature looks something like this:

fn add(self, s: &str) -> String {

This isn’t the exact signature that’s in the standard library: in the standard library, add is defined using generics. Here, we’re looking at the signature of add with concrete types substituted for the generic ones, which is what happens when we call this method with String values. We’ll discuss generics in Chapter 10. This signature gives us the clues we need to understand the tricky bits of the + operator.

prod: confirm xref

First, s2 has an &, meaning that we’re adding a reference of the second string to the first string because of the s parameter in the add function: we can only add a &str to a String; we can’t add two String values together. But wait - the type of &s2 is &String, not &str, as specified in the second parameter to add. Why does Listing 8-16 compile? We are able to use &s2 in the call to add because the compiler can coerce the &String argument into a &str. When we call the add method, Rust uses something called a deref coercion, which you could think of here as turning &s2 into &s2[..]. We’ll discuss deref coercion in more depth in Chapter 15. Because add does not take ownership of the s parameter, s2 will still be a valid String after this operation.

Second, we can see in the signature that add takes ownership of self, because self does not have an &. This means s1 in Listing 8-16 will be moved into the add call and no longer be valid after that. So although let s3 = s1 + &s2; looks like it will copy both strings and create a new one, this statement actually takes ownership of s1, appends a copy of the contents of s2, and then returns ownership of the result. In other words, it looks like it’s making a lot of copies but isn’t: the implementation is more efficient than copying.

If we need to concatenate multiple strings, the behavior of + gets unwieldy:

let s1 = String::from("tic");

let s2 = String::from("tac");

let s3 = String::from("toe");

let s = s1 + "-" + &s2 + "-" + &s3;

At this point, s will be tic-tac-toe. With all of the + and " characters, it’s difficult to see what’s going on. For more complicated string combining, we can use the format! macro:

let s1 = String::from("tic");

let s2 = String::from("tac");

let s3 = String::from("toe");

let s = format!("{}-{}-{}", s1, s2, s3);

This code also sets s to tic-tac-toe. The format! macro works in the same way as println!, but instead of printing the output to the screen, it returns a String with the contents. The version of the code using format! is much easier to read and also doesn’t take ownership of any of its parameters.

Indexing into Strings

In many other programming languages, accessing individual characters in a string by referencing them by index is a valid and common operation. However, if we try to access parts of a String using indexing syntax in Rust, we’ll get an error. Consider the code in Listing 8-17:

let s1 = String::from("hello");

let h = s1[0];

Listing 8-17: Attempting to use indexing syntax with a String

This code will result in the following error:

error: the trait bound `std::string::String: std::ops::Index<\_>` is not

satisfied [--explain E0277]

|>

|> let h = s1[0];

|> ^^^^^

note: the type `std::string::String` cannot be indexed by `\_`

The error and the note tell the story: Rust strings don’t support indexing. But why not? To answer that question, we need to discuss how Rust stores strings in memory.

Internal Representation

A String is a wrapper over a Vec<u8>. Let’s look at some of our properly encoded UTF-8 example strings from Listing 8-12. First, this one:

let len = String::from("Hola").len();

In this case, len will be four, which means the Vec storing the string “Hola” is four bytes long. Each of these letters takes one byte when encoded in UTF-8. But what about the following line?

let len = String::from("Здравствуйте").len();

Asked how long the string is, you might say 12. However, Rust’s answer is 24: that’s the number of bytes it takes to encode “Здравствуйте” in UTF-8, because each Unicode scalar value takes two bytes of storage. Therefore, an index into the string’s bytes will not always correlate to a valid Unicode scalar value. To demonstrate, consider this invalid Rust code:

let hello = "Здравствуйте";

let answer = &hello[0];

What should the value of answer be? Should it be З, the first letter? When encoded in UTF-8, the first byte of З is 208, and the second is 151, so answer should in fact be 208, but 208 is not a valid character on its own. Returning 208 is likely not what a user would want if they asked for the first letter of this string; however, that’s the only data that Rust has at byte index 0. Returning the byte value is probably not what users want, even if the string contains only Latin letters: if &"hello"[0] was valid code that returned the byte value, it would return 104, not h. To avoid returning an unexpected value and causing bugs that might not be discovered immediately, Rust doesn’t compile this code at all and prevents misunderstandings earlier in the development process.

Bytes and Scalar Values and Grapheme Clusters! Oh My!

Another point about UTF-8 is that there are actually three relevant ways to look at strings from Rust’s perspective: as bytes, scalar values, and grapheme clusters (the closest thing to what we would call letters).

If we look at the Hindi word “नमस्ते” written in the Devanagari script, it is stored as a Vec of u8 values that looks like this:

[224, 164, 168, 224, 164, 174, 224, 164, 184, 224, 165, 141, 224, 164, 164, 224, 165, 135]

That’s 18 bytes and is how computers ultimately store this data. If we look at them as Unicode scalar values, which are what Rust’s char type is, those bytes look like this:

['न', 'म', 'स', '्', 'त', 'े']

There are six char values here, but the fourth and sixth are not letters: they’re diacritics that don’t make sense on their own. Finally, if we look at them as grapheme clusters, we’d get what a person would call the four letters that make up the Hindi word:

["न", "म", "स्", "ते"]

Rust provides different ways of interpreting the raw string data that computers store so that each program can choose the interpretation it needs, no matter what human language the data is in.

A final reason Rust doesn’t allow us to index into a String to get a character is that indexing operations are expected to always take constant time (O(1)). But it isn’t possible to guarantee that performance with a String, because Rust would have to walk through the contents from the beginning to the index to determine how many valid characters there were.

Slicing Strings

Indexing into a string is often a bad idea because it’s not clear what the return type of the string indexing operation should be: a byte value, a character, a grapheme cluster, or a string slice. Therefore, Rust asks you to be more specific if you really need to use indices to create string slices. To be more specific in your indexing and indicate that you want a string slice, rather than indexing using [] with a single number, you can use [] with a range to create a string slice containing particular bytes:

let hello = "Здравствуйте";

let s = &hello[0..4];

Here, s will be a &str that contains the first four bytes of the string. Earlier, we mentioned that each of these characters was two bytes, which means s will be Зд.

What would happen if we used &hello[0..1]? The answer: Rust will panic at runtime in the same way that accessing an invalid index in a vector does:

thread 'main' panicked at 'index 0 and/or 1 in `Здравствуйте` do not lie on

character boundary', ../src/libcore/str/mod.rs:1694

You should use ranges to create string slices with caution, because it can crash your program.

Methods for Iterating Over Strings

Fortunately, we can access elements in a string in other ways.

If we need to perform operations on individual Unicode scalar values, the best way to do so is to use the chars method. Calling chars on “नमस्ते” separates out and returns six values of type char, and we can iterate over the result to access each element:

for c in "नमस्ते".chars() {

println!("{}", c);

}

This code will print the following:

न

म

स

्

त

े

The bytes method returns each raw byte, which might be appropriate for your domain:

for b in "नमस्ते".bytes() {

println!("{}", b);

}

This code will print the 18 bytes that make up this String, starting with:

224

164

168

224

// ... etc

But be sure to remember that valid Unicode scalar values may be made up of more than one byte.

Getting grapheme clusters from strings is complex, so this functionality is not provided by the standard library. Crates are available on https://crates.io if this is the functionality you need.

Strings Are Not So Simple

To summarize, strings are complicated. Different programming languages make different choices about how to present this complexity to the programmer. Rust has chosen to make the correct handling of String data the default behavior for all Rust programs, which means programmers have to put more thought into handling UTF-8 data upfront. This trade-off exposes more of the complexity of strings than other programming languages do but prevents you from having to handle errors involving non-ASCII characters later in your development life cycle.

Let’s switch to something a bit less complex: hash maps!

Hash Maps

The last of our common collections is the hash map. The type HashMap<K, V> stores a mapping of keys of type K to values of type V. It does this via a hashing function, which determines how it places these keys and values into memory. Many different programming languages support this kind of data structure, but often use a different name, such as hash, map, object, hash table, or associative array, just to name a few.

Hash maps are useful for when you want to look up data not by an index, as you can with vectors, but by using a key that can be of any type. For example, in a game, you could keep track of each team’s score in a hash map where each key is a team’s name and the values are each team’s score. Given a team name, you can retrieve its score.

We’ll go over the basic API of hash maps in this section, but many more goodies are hiding in the functions defined on HashMap<K, V> by the standard library. As always, check the standard library documentation for more information.

Creating a New Hash Map

We can create an empty hash map with new and add elements with insert. In Listing 8-18, we’re keeping track of the scores of two teams whose names are Blue and Yellow. The Blue team will start with 10 points, and the Yellow team starts with 50:

use std::collections::HashMap;

let mut scores = HashMap::new();

scores.insert(String::from("Blue"), 10);

scores.insert(String::from("Yellow"), 50);

Listing 8-18: Creating a new hash map and inserting some keys and values

Note that we need to first use the HashMap from the collections portion of the standard library. Of our three common collections, this one is the least often used, so it’s not included in the features imported automatically in the prelude. Hash maps also have less support from the standard library; there’s no built-in macro to construct them, for example.

Just like vectors, hash maps store their data on the heap. This HashMap has keys of type String and values of type i32. Like vectors, hash maps are homogeneous: all of the keys must have the same type, and all of the values must have the same type.

Another way of constructing a hash map is by using the collect method on a vector of tuples, where each tuple consists of a key and its value. The collect method gathers data into a number of collection types, including HashMap. For example, if we had the team names and initial scores in two separate vectors, we can use the zip method to create a vector of tuples where “Blue” is paired with 10, and so forth. Then we can use the collect method to turn that vector of tuples into a HashMap as shown in Listing 8-19:

use std::collections::HashMap;

let teams = vec![String::from("Blue"), String::from("Yellow")];

let initial\_scores = vec![10, 50];

let scores: HashMap<\_, \_> = teams.iter().zip(initial\_scores.iter()).collect();

Listing 8-19: Creating a hash map from a list of teams and a list of scores

The type annotation HashMap<\_, \_> is needed here because it’s possible to collect into many different data structures, and Rust doesn’t know which you want unless you specify. For the type parameters for the key and value types, however, we use underscores, and Rust can infer the types that the hash map contains based on the types of the data in the vectors.

Hash Maps and Ownership

For types that implement the Copy trait, like i32, the values are copied into the hash map. For owned values like String, the values will be moved and the hash map will be the owner of those values as demonstrated in Listing 8-20:

use std::collections::HashMap;

let field\_name = String::from("Favorite color");

let field\_value = String::from("Blue");

let mut map = HashMap::new();

map.insert(field\_name, field\_value);

// field\_name and field\_value are invalid at this point, try using them and

// see what compiler error you get!

Listing 8-20: Showing that keys and values are owned by the hash map once they’re inserted

We aren’t able to use the variables field\_name and field\_value after they’ve been moved into the hash map with the call to insert.

If we insert references to values into the hash map, the values won’t be moved into the hash map. The values that the references point to must be valid for at least as long as the hash map is valid. We’ll talk more about these issues in the “Validating References with Lifetimes” section in Chapter 10.

prod: confirm xref

Accessing Values in a Hash Map

We can get a value out of the hash map by providing its key to the get method as shown in Listing 8-21:

use std::collections::HashMap;

let mut scores = HashMap::new();

scores.insert(String::from("Blue"), 10);

scores.insert(String::from("Yellow"), 50);

let team\_name = String::from("Blue");

let score = scores.get(&team\_name);

Listing 8-21: Accessing the score for the Blue team stored in the hash map

Here, score will have the value that’s associated with the Blue team, and the result will be Some(&10). The result is wrapped in Some because get returns an Option<&V>; if there’s no value for that key in the hash map, get will return None. The program will need to handle the Option in one of the ways that we covered in Chapter 6.

prod: confirm xref

We can iterate over each key/value pair in a hash map in a similar manner as we do with vectors, using a for loop:

use std::collections::HashMap;

let mut scores = HashMap::new();

scores.insert(String::from("Blue"), 10);

scores.insert(String::from("Yellow"), 50);

for (key, value) in &scores {

println!("{}: {}", key, value);

}

This code will print each pair in an arbitrary order:

Yellow: 50

Blue: 10

Updating a Hash Map

Although the number of keys and values is growable, each key can only have one value associated with it at a time. When we want to change the data in a hash map, we have to decide how to handle the case when a key already has a value assigned. We could replace the old value with the new value, completely disregarding the old value. We could keep the old value and ignore the new value, and only add the new value if the key doesn’t already have a value. Or we could combine the old value and the new value. Let’s look at how to do each of these!

Overwriting a Value

If we insert a key and a value into a hash map, and then insert that same key with a different value, the value associated with that key will be replaced. Even though the code in Listing 8-22 calls insert twice, the hash map will only contain one key/value pair because we’re inserting the value for the Blue team’s key both times:

use std::collections::HashMap;

let mut scores = HashMap::new();

scores.insert(String::from("Blue"), 10);

scores.insert(String::from("Blue"), 25);

println!("{:?}", scores);

Listing 8-22: Replacing a value stored with a particular key

This code will print {"Blue": 25}. The original value of 10 has been overwritten.

Only Insert If the Key Has No Value

It’s common to check whether a particular key has a value, and if it doesn’t, insert a value for it. Hash maps have a special API for this called entry that takes the key we want to check as a parameter. The return value of the entry function is an enum called Entry that represents a value that might or might not exist. Let’s say we want to check whether the key for the Yellow team has a value associated with it. If it doesn’t, we want to insert the value 50, and the same for the Blue team. Using the entry API, the code looks like Listing 8-23:

use std::collections::HashMap;

let mut scores = HashMap::new();

scores.insert(String::from("Blue"), 10);

scores.entry(String::from("Yellow")).or\_insert(50);

scores.entry(String::from("Blue")).or\_insert(50);

println!("{:?}", scores);

Listing 8-23: Using the entry method to only insert if the key does not already have a value

The or\_insert method on Entry is defined to return the value for the corresponding Entry key if that key exists, and if not, inserts the parameter as the new value for this key and returns the modified Entry. This technique is much cleaner than writing the logic ourselves, and in addition, plays more nicely with the borrow checker.

Running the code in Listing 8-23 will print {"Yellow": 50, "Blue": 10}. The first call to entry will insert the key for the Yellow team with the value 50 because the Yellow team doesn’t have a value already. The second call to entry will not change the hash map because the Blue team already has the value 10.

Updating a Value Based on the Old Value

Another common use case for hash maps is to look up a key’s value and then update it based on the old value. For instance, Listing 8-24 shows code that counts how many times each word appears in some text. We use a hash map with the words as keys and increment the value to keep track of how many times we’ve seen that word. If it’s the first time we’ve seen a word, we’ll first insert the value 0:

use std::collections::HashMap;

let text = "hello world wonderful world";

let mut map = HashMap::new();

for word in text.split\_whitespace() {

let count = map.entry(word).or\_insert(0);

\*count += 1;

}

println!("{:?}", map);

Listing 8-24: Counting occurrences of words using a hash map that stores words and counts

This code will print {"world": 2, "hello": 1, "wonderful": 1}. The or\_insert method actually returns a mutable reference (&mut V) to the value for this key. Here we store that mutable reference in the count variable, so in order to assign to that value we must first dereference count using the asterisk (\*). The mutable reference goes out of scope at the end of the for loop, so all of these changes are safe and allowed by the borrowing rules.

Hashing Function

By default, HashMap uses a cryptographically secure hashing function that can provide resistance to Denial of Service (DoS) attacks. This is not the fastest hashing algorithm available, but the trade-off for better security that comes with the drop in performance is worth it. If you profile your code and find that the default hash function is too slow for your purposes, you can switch to another function by specifying a different hasher. A hasher is a type that implements the BuildHasher trait. We’ll talk about traits and how to implement them in Chapter 10. You don’t necessarily have to implement your own hasher from scratch; https://crates.io has libraries shared by other Rust users that provide hashers implementing many common hashing algorithms.

prod: confirm xref

Summary

Vectors, strings, and hash maps will provide a large amount of functionality that you need in programs where you need to store, access, and modify data. Here are some exercises you should now be equipped to solve:

Given a list of integers, use a vector and return the mean (average), median (when sorted, the value in the middle position), and mode (the value that occurs most often; a hash map will be helpful here) of the list.

Convert strings to pig latin. The first consonant of each word is moved to the end of the word and “ay” is added, so “first” becomes “irst-fay.” Words that start with a vowel have “hay” added to the end instead (“apple” becomes “apple-hay”). Keep in mind the details about UTF-8 encoding!

Using a hash map and vectors, create a text interface to allow a user to add employee names to a department in a company. For example, “Add Sally to Engineering” or “Add Amir to Sales.” Then let the user retrieve a list of all people in a department or all people in the company by department, sorted alphabetically.

The standard library API documentation describes methods that vectors, strings, and hash maps have that will be helpful for these exercises!

We’re getting into more complex programs in which operations can fail; so, it’s a perfect time to discuss error handling next!