

Computational Thinking for Actuaries and Financial Professionals

With Applications in Julia

Alec Loudenback and Yun-Tien Lee

2024-12-16

Table of contents

Preface	1
1. Draft of Cover	3
I. Introduction	5
The approach	8
What you will learn	8
Prerequisites	9
The Contents of This Book	10
Notes on formatting	11
2. Why Program?	13
2.1. In this Chapter	13
2.2. The Long View	13
2.3. What's coding got to do with this?	15
2.4. The 10x Actuary	16
2.5. Risk Governance	17
2.6. Managing and Leading the Transformation	18
2.7. Outlook	18
3. Why use Julia?	19
3.1. Expressiveness and Syntax	20
3.1.1. Example: Retention Analysis	20
3.1.2. Example: Random Sampling	22
3.2. The Speed	23
3.3. More of Julia's benefits	24
3.4. The Tradeoff	25
3.5. Package Ecosystem	26
3.6. Conclusion	27

II. Conceptual Foundations	29
4. Elements of Financial Modeling	31
4.1. In this Chapter	31
4.2. What is a model?	31
4.3. What is a <i>Financial</i> Model?	32
4.4. The Process of Building a Financial Model	33
4.5. Predictive versus Explanatory Models	34
4.5.1. A Historical Example	34
4.5.2. Examples in the Financial Context	37
4.6. What makes a good model?	38
4.6.1. Achieving original purpose	38
4.6.2. Usability	38
4.6.3. Performance	39
4.7. What makes a good modeler?	40
4.7.1. Domain Expertise	40
4.7.2. Model Theory	40
4.7.3. Curiosity	40
4.7.4. Rigor	40
4.7.5. Toolset	40
5. Elements of Programming	41
5.1. In this section	41
5.2. Computer Science, Programming, and Coding	41
5.3. Assignment and Variables	42
5.4. Data Types	43
5.4.1. Numbers	44
5.4.2. Type Hierarchy	46
5.4.3. Arrays	47
5.4.4. Characters, Strings, and Symbols	51
5.4.5. Tuples	53
5.4.6. Parametric Types	54
5.4.7. Creating User Defined Types	55
5.5. Expressions and Control Flow	55
5.5.1. More on Assignment and Variables	55
5.5.2. Operators	55
5.6. Functions	55
5.6.1. Broadcasting	55
5.7. Scope	55
5.7.1. Local Scope	55
5.7.2. Global Scope	55

Table of contents

5.7.3. Modules	55
6. Elements of Computer Science	57
6.1. In this section	57
6.2. Computer Science for Financial Professionals . .	57
6.3. Algorithms	59
6.3.1. Computational Complexity	59
6.4. Data Structures	59
6.5. Information Theory	59
6.5.1. Signal vs Noise	59
6.6. Formal Verification	59
6.7. The Discipline of Software Engineering	59
6.7.1. Patterns	59
7. Hardware and It's Implications	61
7.1. In this section	61
8. Applying Software Engineering Principles	63
8.1. In this section	63
III. Computational Thinking in an Actuarial and Financial Context	65
9. Modeling	67
9.1. In This Chapter	67
9.2. Parsimony	67
10. Optimization	69
10.1. In This Chapter	69
11. Sensitivity Analysis	71
11.1. In This Chapter	71
12. Stochastic Modeling	73
13. Visualizations	75
13.1. In This Chapter	75
14. Matrices and Their Uses	77
14.1. In This Chapter	77

Table of contents

15. Learning from Data	79
15.1. In this chapter	79
 IV. Applications in Practice	 81
16. Stochastic Mortality Projections	83
16.1. In This Chapter	83
17. Scenario Generation	85
17.1. In This Chapter	85
18. Similarity Analysis	87
18.1. In This Chapter	87
19. Portfolio Optimization	89
19.1. In This Chapter	89
20. Other Useful Techniques	91
20.1. In this chapter	91
20.2. Taking things to the Extreme	91
20.3. Range Bounding	91
 V. Appendices	 93
21. Set up Julia and the Computing Environment	95
21.1. Installation	95
21.2. Package Management	95
21.3. Editors	96
21.3.1. Visual Studio Code	96
21.3.2. Notebooks	97
 22. The Julia Ecosystem Today	 99
22.0.1. Data	100
22.0.2. Plotting	100
22.0.3. Statistics	100
22.0.4. Machine Learning	101
22.0.5. Differentiable Programming	101
22.0.6. Utilities	102
22.0.7. Other packages	103
22.0.8. Actuarial packages	103

Table of contents

References	105
-------------------	------------

Preface

This book is intended to enable practitioners and advanced students of financial disciplines to utilize the tools, language, and ideas of computational sciences in their own discipline.

1. Draft of Cover

1. Draft of Cover

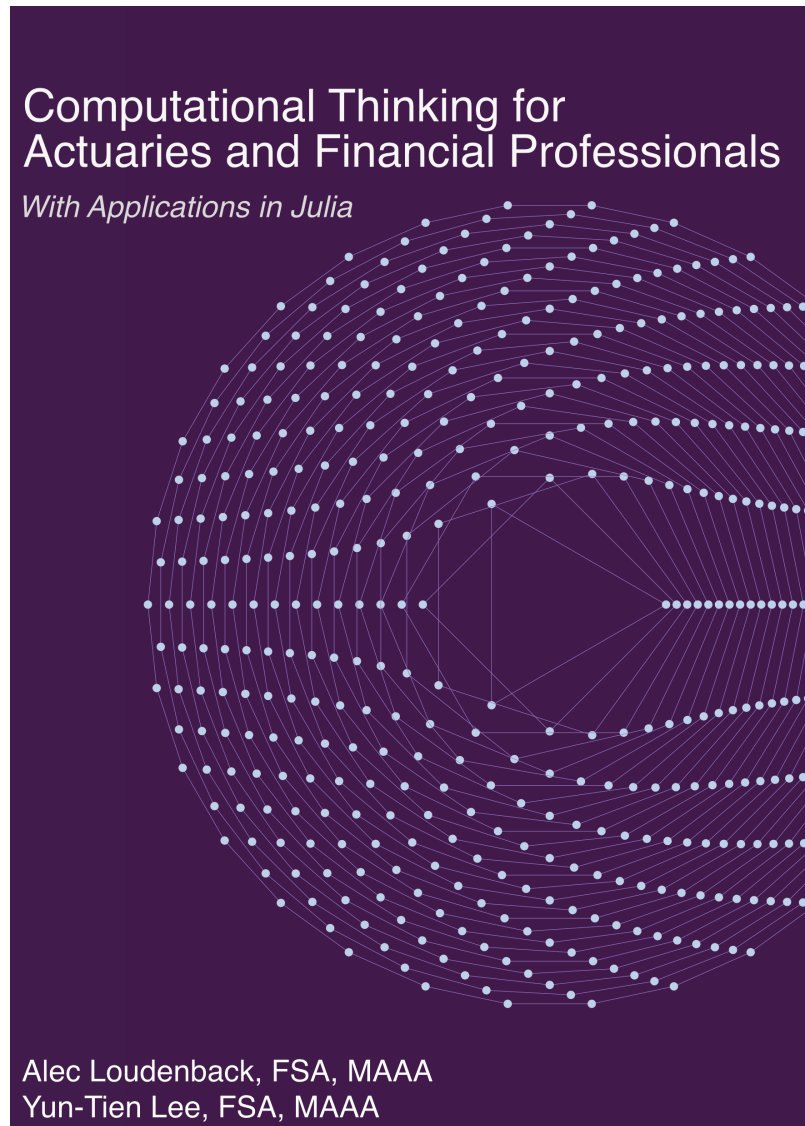


Figure 1.1.: Draft of Cover

Part I.

Introduction

“I think one of the things that really separates us from the high primates is that we’re tool builders. I read a study that measured the efficiency of locomotion for various species on the planet. The condor used the least energy to move a kilometer. And, humans came in with a rather unimpressive showing, about a third of the way down the list. It was not too proud a showing for the crown of creation. So, that didn’t look so good. But, then somebody at Scientific American had the insight to test the efficiency of locomotion for a man on a bicycle. And, a man on a bicycle, a human on a bicycle, blew the condor away, completely off the top of the charts.

And that’s what a computer is to me. What a computer is to me is it’s the most remarkable tool that we’ve ever come up with, and it’s the equivalent of a bicycle for our minds.” - Steve Jobs (1990)

The world of financial modeling is incredibly complex and variegated. It, along with many of the sciences, is a place where practical goals harness computational tools to arrive at answers that (we hope) are meaningful in a way that tells us more about the world we live in. What this usually means specifically is that practitioners utilize computers to do the heavy work of processing data or running simulations which reveal the something about the complex systems we seek to represent. In this way, then, financial modelers must also be a craftsman who seeks not only to design new products, but must also think carefully about the tools and the process used therein.

This book seeks to aid the practitioner in developing that workmanship: we will develop new ways to look at the *process*, think about how to most clearly represent ideas, dive into details about computer hardware and bring it back up to the most abstract levels, and develop a vocabulary to more clearly express and communicate these concepts. The book contains a large number of practical examples to demonstrate that the end result is better for the journey we will take.

This book looks at programming for the applied financial professional and we will start by answering a very basic question:

“why is this relevant for financial modeling?”. The answer is simple: financial modeling is complex, data intensive, and often very abstract. Programming is the best tool humans have so far developed for rigorously transforming ideas and data into results. A builder may be the most skilled person in the world with a hammer but another with some basic training in a richer set of tools will build a better house. This book will enhance your toolkit with experience with multiple tools: a specific programming language, yes, but much more than that: a language to talk about solving problems, a deeper understanding of specific problem solving techniques, how to make decisions about what the architecture of a solution looks like, and practical advice from experienced practitioners.

The approach

The authors of the book are practicing actuaries, but we intend for the content to be applicable to nearly all practitioners in the financial industry. The discussion and examples may have an orientation towards insurance topics, but the concepts and patterns are applicable to a wide variety of related disciplines.

We will pull from examples on both sides of the balance sheet: the left (assets) and right (liabilities). We may also take the liberty to, at times, abuse traditional accounting notions: a liability is just an asset with the obligor and obligee switched. When the accounting conventions are important (such as modeling a total balance sheet) we will be mindful in explaining the accounting perspective. In practice, this means that we’ll take examples that use examples of assets (fixed income, equity, derivatives) or liabilities (life insurance, annuities, long term care) and show that similar modeling techniques can be used for both.

What you will learn

It is our hope that with the help of this book, you will find it more efficient to discuss aspects of modeling with colleagues,

borrow problem solving language from computer science, spot recurring structural patterns in problems that arise, and understand how best to make use of the “bicycle for your mind” in the context of financial modeling.

It is the experience of the authors that many professionals that do complex modeling as a part of their work have gotten to be very proficient *in spite of* not having substantive formal training on problem solving, algorithms, or model architecture. This book serves to fill that gap and provide the “missing semester” (or “years of practical learning”!). After reading this book, we hope that you will *appreciate* the attributes of Microsoft Excel that made it so ubiquitous, but that you *prefer* to use a programming language for the ability to more naturally express the relevant abstractions which make your models simpler, faster, or more usable by others.

Prerequisites

Basic experience with financial modeling is not strictly required, but it will benefit the reader to be familiar so that the examples will not be attempting to teach both financial maths and computer science simultaneously.

Advanced financial maths (e.g. stochastic calculus) is *not* required. Indeed, this book is not oriented to the advanced technicalities of Wall Street “quants” and is instead directed at the multitudes of financial practitioners focused on producing results that are not measured in the microseconds of high-frequency trading.

Prior programming experience is *not* required either: Chapter 5 introduces the basic syntax and concepts while Chapter 21 covers setting up your environment to follow along. For readers with background in programming, we recommend skimming Chapter 5 and reading in full the sections which have a symbol in the margin, which is our way of highlighting Julia-specific content to be aware of.

i TODO

Create a venn diagram showing financial modeling at the intersection of statistics, financial math, computer science.

The Contents of This Book

Part 1 of the book addresses the theoretical and technical foundations of programming, as well as the conceptual basis for financial modelling. It familiarizes the readers with key functional programming principles, alongside introducing important aspects of software engineering relevant to financial modelling.

Parts 2 and 3 bridge the gap between theory and practical applications, underlining the features of Julia that make it a robust tool for real-world financial and actuarial contexts. Through a careful exploration of topics like sensitivity analysis, optimization, stochastic modeling, visualization, and practical financial applications, the book demonstrates how Julia's high-level, high-performance programming capabilities can enhance accuracy and efficiency in financial modelling. As an up-and-coming language loved for its speed and simplicity, Julia is ripe for wide adoption in the financial sector. The time for this book is ripe, as it will satiate the growing demand for professionals who want to blend programming skills with financial modelling acumen.

While we have chosen to use Julia for the examples in this book, the vast majority of the concepts presented are not Julia-specific. We will attempt to motivate why Julia works so well as a language for financial modeling but like mathematics and applied mathematics, the concepts are portable even if the numbers (language) changes. Readers are encouraged to follow along the examples on their own computer (see instructions for Julia in Chapter 21) and the entire book is available on GitHub at [#TODO: determine book URL].

Notes on formatting

When a concept is defined for the first time, the term will be **bold**. Code, or references to pieces of code will be formatted in inline code style like `1+1` or in separate code blocks:

```
"This is a code block that doesn't show any results"
```

```
"This is a code block that does show output"
```

```
"This is a code block that does show output"
```


2. Why Program?

“Humans are allergic to change. They love to say, ‘We’ve always done it this way.’ I try to fight that. That’s why I have a clock on my wall that runs counterclockwise.” - Grace Hopper (1987)

[Drafting Note: This chapter is pulled from the article published in 2020 and needs to be adapted for the book’s audience. Also to include: why not low-code solutions?]

2.1. In this Chapter

We motivate why a financial professional should adopt programming skills which will improve their own capabilities and enjoyment of the discipline, whilst allowing themselves to better themselves and the industry we work in.

2.2. The Long View

It might be odd to say that technology and its use in insurance is on a one-hundred-year cycle, but that seems to be the case.

130 years ago, actuaries crowded into a room at a meeting of the Actuarial Society of America to watch a demonstration that would revolutionize the industry: Herman Hollerith’s tabulating punch card machine¹.

For the next half-century, the increasing automation — from tabulating machines to early-adopting mainframes and computers — was a critical competitive differentiator. Companies like Prudential, MetLife, and others partnered with technology companies in the development of hardware and software².

¹ Co-evolution of Information Processing Technology and Use: Interaction Between the Life Insurance and Tabulating Industries

² From Tabulators to Early Computers in the U.S. Life Insurance Industry

2. Why Program?

The dramatic embodiment of this information-driven cycle was portrayed in the infamous Billion Dollar Bubble movie, which showcased the power and abstraction of the computer to commit millions of dollars of fraud by creating and maintaining fake insurance policies.

The movie also starts to hint at the oscillation away from the technological-competitive focus of insurance companies. I argue that the focus on technology was lost over the last 50 years with the rise of Wall Street finance, investment-oriented life insurance, industry consolidation, and the explosion of financial structuring like derivatives, reserve financing, or other advanced forms of reinsurance.

Value-add came from the C-Suite, not from the underlying business processes, operations, and analysis. The result is, e.g., ever-more complicated reinsurance treaties layered into mainframes and admin systems older than most of the actuaries interfacing with them.

The pace of *strategic value-add* isn't slowing, though it must stretch further (in complexity and risk) to find comparable opportunities as the past. Having more agile, data-oriented operations enables companies to be able to react to and implement those opportunities. *Technological value-add* can improve a company's bottom line through lower expenses and higher top-line growth, but often with a more favorable risk profile than some of the "strategic" opportunities.

Today, there is a trend reverting back to technological value-creation and is evident across many traditional sectors. Tesla claims that it's a technology company; Amazon is the #1 product retailer because of its vehement focus on internal information sharing³; Airlines are so dependent on their systems that the skies become quieter on the rare occasion that their computers give way.

Why is it, that companies that are so involved in *things* (cars, shopping) and *physical services* (flights) are so much more focused on improving their technological operations than insurance companies *whose very focus is 'information-based'*? **The market has rewarded those who have prioritized their internal technological solutions.**

³ Have you had your Bezos moment?
What you can learn from Amazon

2.3. What's coding got to do with this?

Commoditized investing services and low yield environments have reduced insurance companies' comparative advantage to "manage money". Yield compression and the explosion of consumer-oriented investment services means a more competitive focus on the ability to manage the entire policy lifecycle efficiently (digitally), perform more real-time analysis of experience and risk management, and handle the growing product and regulatory complexity.

These are problems that have technological solutions and are waiting for insurance company adoption.

Companies that treat data like coordinates on a grid (spreadsheets) *will get left behind*. Two main hurdles have prevented technology companies from breaking into insurance:

1. High regulatory barriers to entry, and
2. Difficulty in selling complex insurance products without traditional distribution.

Once those two walls are breached, traditional insurance companies without a strong technology core will struggle to keep up. The key to thriving is not just adding "developers" to an organization; it's going to be **getting domain experts like actuaries to be an integral part of the technology transformation.**

2.3. What's coding got to do with this?

Everything. Programming is the optimal way to interact between the computer and actuary — and importantly between computer and computer. Programming is the actionable expression of ideas, math, analysis, and information. Think of programming as the 21st-century leap in the actuary's toolkit, just as spreadsheets were in the preceding 40 years. Versus a spreadsheet-oriented workflow:

- More natural automation of, and between processes
- Better reproducibility
- Scaling to fit any size dataset and workload
- Statistics and machine learning capabilities

2. Why Program?

- Advanced visualizations to garner new views into your data

This list isn't comprehensive and some benefits are subtle — when you are code-oriented instead of spreadsheet-oriented, you tend to want to structure your data in a portable and shareable way. For example, relying more on data warehouses instead of email attachments. This, in turn, enables data discovery and insights that otherwise wouldn't be there. Investing in a code-oriented workflow is playing the long-game.

The actuary of the future needs to have coding as one of their core skills. Already today, the advances of business processes, insurance products, and financial ingenuity are written with lines of code — *not* spreadsheets. Not being able to code *necessarily* means that you are *following* what others are doing today.

It's commonly accepted now that to gather insights from your data, you need to know how to code. Similar to your data, your business architecture, modeling needs, and product peculiarities are often better suited to customized solutions. Why stop at data science when learning how to solve problems with a computer?

2.4. The 10x Actuary

As we swing back to a technological focus, we do not leave the finance-driven complexity behind. The increasingly complex business needs will highlight a large productivity difference between an actuary who can code and one who can't — simply because the former can react, create, synthesize, and model faster than the latter. From the efficiency of transforming administration extracts, summarizing and aggregating valuation output, to analyzing claims data in ways that spreadsheets simply can't handle, you can become a “**10x Actuary**”⁴.

⁴ The 10x [Rockstar] developer is NOT a myth

Flipping switches in a graphical user interface versus being able to *build models* is the difference between having a surface-level familiarity and having full command over the analysis and the

2.5. Risk Governance

concepts involved — with the flexibility to do what your software can't.

Your current software might be able to perform the first layer of analysis but be at a loss when you want to visualize, perform sensitivity analysis, statistics, stochastic analysis, or process automation. Things that, when done programmatically, are often just a few lines of additional code.

Do I advocate dropping the license for your software vendor? No, not yet anyway. But the ability to supplement and break out of the modeling box has been an increasingly important part of most actuaries' work.

Additionally, code-based solutions can leverage the entire-technology sector's progress to solve problems that are *hard* otherwise: scalability, data workflows, integration across functional areas, version control and versioning, model change governance, reproducibility, and more.

30-40 years ago, there were no vendor-supplied modeling solutions and so you had no choice but to build models internally. This shifted with the advent of vendor-supplied modeling solutions. Today, it's never been better for companies to leverage open source to support their custom modeling, risk analysis/monitoring, and reporting workflows.

2.5. Risk Governance

Code-based workflows are highly conducive to risk governance frameworks as well. If a modern software project has all of the following benefits, then why not a modern insurance product and associated processes?

- Access control and approval processes
- Version control, version management, and reproducibility
- Continuous testing and validation of results
- Open and transparent design
- Minimization of manual overrides, intervention, and opportunity for user error
- Automated trending analysis, system metrics, and summary statistics

2. Why Program?

- Continuously updated, integrated, and self-generating documentation
- Integration with other business processes through a formal boundary (e.g. via an API)
- Tools to manage collaboration in parallel and in sequence

2.6. Managing and Leading the Transformation

The ability to understand the concepts, capabilities, challenges, and lingo is not a dichotomy, it's a spectrum. Most actuaries, even at fairly high levels, are still often involved in analytical work. Still above that, it's difficult to lead something that you don't understand.

Conversely, the skill and practice of coding enhances managerial capabilities. When you are really skilled at pulling apart a problem or process into its constituent parts and designing optimal solutions; that's a core attribute of leadership: having the vision of where the organization *should be* instead of thinking about where it is now.

Nor is the skillset described here limiting in any other aspect of career development any more than mathematical ability, project collaboration, or financial acumen — just to name a few.

2.7. Outlook

It will increasingly be essential for companies to modernize to remain competitive. That modernization isn't built with big black-box software packages; it will be with domain experts who can translate the expertise into new forms of analysis - doing it faster and more robustly than the competition.

SpaceX doesn't just hire rocket scientists - they hire rocket scientists who code.

Be an actuary who codes.

3. Why use Julia?

[Drafting Note: This chapter is pulled from the article published in 2021 and needs to be adapted for the book’s audience.]

Julia is relatively new⁵, and *it shows*. It is evident in its pragmatic, productivity-focused design choices, pleasant syntax, rich ecosystem, thriving communities, and its ability to be both very general purpose and power cutting edge computing.

With Julia: math-heavy code looks like math; it’s easy to pick up, and quick-to-prototype. Packages are well-integrated, with excellent visualization libraries and pragmatic design choices.

Julia’s popularity continues to grow across many fields and there’s a growing body of online references and tutorials, videos, and print media to learn from.

Large financial services organizations have already started realizing gains: BlackRock’s Aladdin portfolio modeling, the Federal Reserve’s economic simulations, and Aviva’s Solvency II-compliant modeling⁶. The last of these has a great talk on YouTube by Aviva’s Tim Thornham, which showcases an on-the-ground view of what difference the right choice of technology and programming language can make. Moving from their vendor-supplied modeling solution was **1000x faster, took 1/10 the amount of code, and was implemented 10x faster**.

The language is not just great for data science — but also modeling, ETL, visualizations, package control/version management, machine learning, string manipulation, web-backends, and many other use cases. Julia is well suited for financial modeling work: easy to read and write and very performant.

⁵ Python first appeared in 1990. R is an implementation of S, which was created in 1976, though depending on when you want to place the start of an independent R project varies (1993, 1995, and 2000 are alternate dates). The history of these languages is long and substantial changes have occurred since these dates.

⁶ Aviva Case Study

3. Why use Julia?

3.1. Expressiveness and Syntax

Expressiveness is the *manner in which* and *scope of* ideas and concepts that can be represented in a programming language. **Syntax** refers to how the code *looks* on the screen and its readability.

In a language with high expressiveness and pleasant syntax, you:

- Go from idea in your head to final product faster.
- Encapsulate concepts naturally and write concise functions.
- Compose functions and data naturally.
- Focus on the end-goal instead of fighting the tools.

Expressiveness can be hard to explain, but perhaps two short examples will illustrate.

3.1.1. Example: Retention Analysis

This is a really simple example relating Cessions, Policys, and Lives to do simple retention analysis.

First, let's define our data:

```
# Define our data structures
struct Life
    policies
end

struct Policy
    face
    cessions
end

struct Cession
    ceded
end
```

3.1. Expressiveness and Syntax

Now to calculate amounts retained. First, let's say what retention means for a `Policy`:

```
# define retention
function retained(pol::Policy)
    pol.face - sum(cession.ceded for cession in pol.cessions)
end
```

And then what retention means for a `Life`:

```
function retained(l::Life)
    sum(retained(policy) for policy in life.policies)
end
```

It's almost exactly how you'd specify it English. No joins, no boilerplate, no fiddling with complicated syntax. You can express ideas and concepts the way that you think of them, not, for example, as a series of dataframe joins or as row/column coordinates on a spreadsheet.

We defined `retained` and adapted it to mean related, but different things depending on the specific context. That is, we didn't have to define `retained_life(...)` and `retained_pol(...)` because Julia can be *dispatch* based on what you give it. This is, as some would call it, unreasonably effective.

Let's use the above code in practice then.

The `julia>` syntax indicates that we've moved into Julia's interactive mode (REPL mode):

```
# create two policies with two and one cessions respectively
julia> pol_1 = Policy( 1000, [ Cession(100), Cession(500)] )
julia> pol_2 = Policy( 2500, [ Cession(1000) ] )

# create a life, which has the two policies
julia> life = Life([pol_1, pol_2])

julia> retained(pol_1)
400
```

3. Why use Julia?

```
julia> retained(life)
1900
```

And for the last trick, something called “broadcasting”, which automatically vectorizes any function you write, no need to write loops or create if statements to handle a single vs repeated case:

```
julia> retained.(life.policies) # retained amount for each policy
[400 , 1500]
```

3.1.2. Example: Random Sampling

As another motivating example showcasing multiple dispatch, here’s random sampling in Julia, R, and Python.

We generate 100:

- Uniform random numbers
- Standard normal random numbers
- Bernoulli random number
- Random samples with a given set

Table 3.1.: A comparison of random outcome generation in Julia, R, and Python.

Julia	R	Python
<pre>using Distributions rand(100) rand(Normal(), 100) rand(Bernoulli(0.5), 100) rand(["Preferred", "Standard"], 100)</pre>	<pre>runif(100) rnorm(100) rbern(100, 0.5) sample(c("Preferred", "Standard"), 100, replace=TRUE)</pre>	<pre>import scipy.stats as sps import numpy as np sps.uniform.rvs(size=100) sps.norm.rvs(size=100) sps.bernoulli.rvs(p=0.5, size=100) np.random.choice(["Preferred", "Standard"], size=100)</pre>

3.2. The Speed

By understanding the different types of things passed to `rand()`, it maintains the same syntax across a variety of different scenarios. We could define `rand(Cession)` and have it generate a random `Cession` like we used above.

3.2. The Speed

As the journal *Nature* said, “Come for the Syntax, Stay for the Speed”.

Recall the Solvency II compliance which ran 1000x faster than the prior vendor solution mentioned earlier: what does it mean to be 1000x faster at something? It’s the difference between something taking 10 seconds instead of 3 hours — or 1 hour instead of 42 days.

**What analysis would you like to do if it took less time?
A stochastic analysis of life-level claims? Machine
learning with your experience data? Daily valuation
instead of quarterly?**

Speaking from experience, speed is not just great for production time improvements. During development, it’s really helpful too. When building something, I can see that I messed something up in a couple of seconds instead of 20 minutes. The build, test, fix, iteration cycle goes faster this way.

Admittedly, most workflows don’t see a 1000x speedup, but 10x to 1000x is a very common range of speed differences vs R or Python or MATLAB.

Sometimes you will see less of a speed difference; R and Python have already circumvented this and written much core code in low-level languages. This is an example of what’s called the “two-language” problem where the language productive to write in isn’t very fast. For example, more than half of R packages use C/C++/Fortran and core packages in Python like Pandas, PyTorch, NumPy, SciPy, etc. do this too.

Within the bounds of the optimized R/Python libraries, you can leverage this work. Extending it can be difficult: what if

3. Why use Julia?

you have a custom retention management system running on millions of policies every night?

Julia packages you are using are almost always written in pure Julia: you can see what's going on, learn from them, or even contribute a package of your own!

3.3. More of Julia's benefits

Julia is easy to write, learn, and be productive in:

- It's free and open-source
 - Very permissive licenses, facilitating the use in commercial environments (same with most packages)
- Large and growing set of available packages
- Write how you like because it's multi-paradigm: vectorizable (R), object-oriented (Python), functional (Lisp), or detail-oriented (C)
- Built-in package manager, documentation, and testing-library
- Jupyter Notebook support (it's in the name! **Julia-Python-R**)
- Many small, nice things that add up:
 - Unicode characters like α or β
 - Nice display of arrays
 - Simple anonymous function syntax
 - Wide range of text editor support
 - First-class support for missing values across the entire language
 - Literate programming support (like R-Markdown)
- Built-in `Dates` package that makes working with dates pleasant
- Ability to directly call and use R and Python code/packages with the `PyCall` and `RCall` packages
- Error messages are helpful and tell you *what line* the error came from, not just the type of error
- Debugger functionality so you can step through your code line by line

3.4. The Tradeoff

For power-users, advanced features are easily accessible: parallel programming, broadcasting, types, interfaces, metaprogramming, and more.

These are some of the things that make Julia one of the world’s most loved languages on the StackOverflow Developer Survey.

For those who are enterprise-minded: in addition to the liberal licensing mentioned above, there are professional products from organizations like Julia Computing that provide hands-on support, training, IT governance solutions, behind-the-firewall package management, and deployment/scaling assistance.

3.4. The Tradeoff

Julia is fast because it’s compiled, unlike R and Python where (loosely speaking) the computer just reads one line at a time. Julia compiles code “just-in-time”: right before you use a function for the first time, it will take a moment to pre-process the code section for the machine. Subsequent calls don’t need to be re-compiled and are very fast.

A hypothetical example: running 10,000 stochastic projections where Julia needs to precompile but then runs each 10x faster:

- Julia runs in 2 minutes: the first projection takes 1 second to compile and run, but each 9,999 remaining projections only take 10ms.
- Python runs in 17 minutes: 100ms of a second for each computation.

Typically, the compilation is very fast (milliseconds), but in the most complicated cases it can be several seconds. One of these is the “time-to-first-plot” issue because it’s the most common one users encounter: super-flexible plotting libraries have a lot of things to pre-compile. So in the case of plotting, it can take several seconds to display the first plot after starting Julia, but then it’s remarkably quick and easy to create an animation of your model results. The time-to-first plot is a solvable problem

3. *Why use Julia?*

that's receiving a lot of attention from the core developers and will get better with future Julia releases.

For users working with a lot of data or complex calculations (like actuaries!), the runtime speedup is worth a few seconds at the start.

3.5. Package Ecosystem

Using packages as dependencies in your project is assisted by Julia's bundled package manager.

For each project, you can track the exact set of dependencies and replicate the code/process on another machine or another time. In R or Python, dependency management is notoriously difficult and it's one of the things that the Julia creators wanted to fix from the start.

Packages can be one of the thousands of publicly available, or private packages hosted internally behind a firewall.

Another powerful aspect of the package ecosystem is that due to the language design, packages can be combined/extended in ways that are difficult for other common languages. This means that Julia packages often interop without any additional coordination.

For example, packages that operate on data tables work without issue together in Julia. In R/Python, many features tend to come bundled in a giant singular package like Python's Pandas which has Input/Output, Date manipulation, plotting, resampling, and more. There's a new Consortium for Python Data API Standards which seeks to harmonize the different packages in Python to make them more consistent (R's Tidyverse plays a similar role in coordinating their subset of the package ecosystem).

In Julia, packages tend to be more plug-and-play. For example, every time you want to load a CSV you might not want to transform the data into a dataframe (maybe you want a matrix or a plot instead). To load data into a dataframe, in Julia the practice is to use both the CSV and DataFrames packages, which

help separate concerns. Some users may prefer the Python/R approach of less modular but more all-inclusive packages.

3.6. Conclusion

Looking at other great tools like R and Python, it can be difficult to summarize a single reason to motivate a switch to Julia, but hopefully this article piqued an interest to try it for your next project.

That said, Julia shouldn't be the only tool in your tool-kit. SQL will remain an important way to interact with databases. R and Python aren't going anywhere in the short term and will always offer a different perspective on things!

In an earlier article, I talked about becoming a **10x Actuary** which meant being proficient in the language of computers so that you could build and implement great things. In a large way, the choice of tools and paradigms shape your focus. Productivity is one aspect, expressiveness is another, speed one more. There are many reasons to think about what tools you use and trying out different ones is probably the best way to find what works best for you.

It is said that you cannot fully conceptualize something unless your language has a word for it. Similar to spoken language, you may find that breaking out of spreadsheet coordinates (and even a dataframe-centric view of the world) reveals different questions to ask and enables innovated ways to solve problems. In this way, you reward your intellect while building more meaningful and relevant models and analysis.

Part II.

Conceptual Foundations

4. Elements of Financial Modeling

“Truth ... is much too complicated to allow anything but approximations” - John von Neumann

4.1. In this Chapter

We explain what constitutes a financial model and what are common uses of a model. We explain what makes an adept practitioner.

4.2. What is a model?

A **model** represents aspects of the world around us distilled down into simpler, more tractable components. It is impossible to fully capture the everything that may affect the objects of our interest.

For example, say we want to simulate the returns for the stocks in our retirement portfolio. It would be impossible to try to build a model which would capture all of the individual people working jobs and making decisions, weather events that damage property, political machinations, etc. Instead, we try to capture certain fundamental characteristics. For example, it is common to model equity returns as cumulative pluses and minuses from random movements where those movements have certain theoretical or historical characteristics.

Whether we are using this model of equity returns to estimate available retirement income or replicate an exotic option price, a key aspect of the model is the **assumptions** used therein.

4. *Elements of Financial Modeling*

For the retirement income scenario we might *assume* a healthy eight percent return on stocks and conclude that such a return will be sufficient to retire at age 53. Alternatively, we may assume that future returns will follow a stochastic path with a certain distribution of volatility and drift. These two assumption sets will produce **output** - results from our model that must be inspected, questioned, and understood in the context of the “small world” of the model’s mechanistic workings. Lastly, to be effective practitioners we must be able to contextualize the “small world” results withing the “large world” that exists around us.

More on the “small world” vs “large world”: say that our model is one that discounts a fixed set of future cashflows using the US Treasury rate curve. If I run my model using current rates today, and then re-run my model tomorrow with the same future cashflows and the present value of those cashflows has increased by 5% I may ask why the result has changed so much in such a short period of time! In the “small”, mechanistic world of the model I may be able to see that the rates I used to discount the cashflows with have fallen substantially. The “small world” answer is that the inputs have changed which produced a mechanical change in the output. The “big world” answer may be that the Federal Reserve lowered the Federal Funds Rate to prevent the economy from entering a deflationary recession. Of course, we can’t completely explain the relation between our model and the real world (otherwise we could capture that relationship in our model!). An effective practitioner will always try to look up from the immediate work and take stock of how the world at large *is* or *is not* reflected in the model.

4.3. What is a *Financial Model*?

Financial models are those used extensively to ascertain better understanding of complex contracts, perform scenario analysis, and inform market participants’ decisions related to perceived value (and therefore price). It can’t be quantified directly, but it is likely not an exaggeration that many billions of dollars is transacted each day as a result of decisions made from the output of financial models.

4.4. The Process of Building a Financial Model

Most financial models can be characterized with a focus on the first or both of:

1. Attempting to project pattern of cashflows or obligations at future timepoints
2. Reducing the projected obligations into a current value

Examples of this:

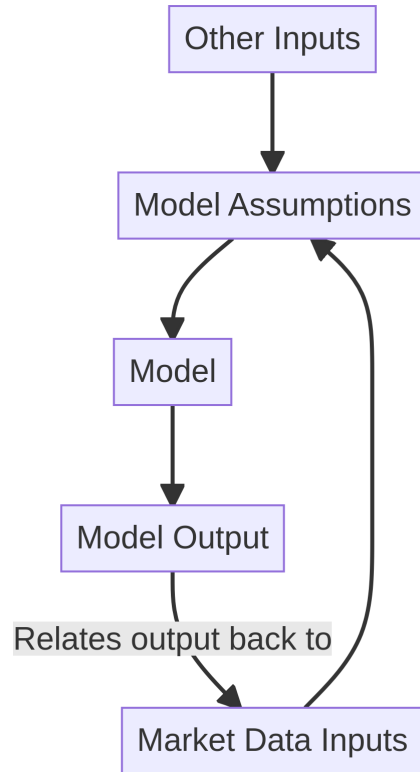
- Projecting a retiree's savings through time (1), and determining how much they should be saving today for their retirement goal (2)
- Projecting the obligation of an exotic option across different potential paths (1), and determining the premium for that option (2)

Models are sometimes taken a step further, such as transforming the underlying **economic view** into an accounting or regulatory view (such as representing associated debits and credits, capital requirements, or associated intangible, capitalized balances).

We should also distinguish a financial model from a purely statistical model, where often the inputs and output data are known and the intention is to estimate relationships between variables (example: linear regressions). That said, a financial model may have statistical components and many aspects of modeling is shared between the two kinds.

4.4. The Process of Building a Financial Model

i TODO: Describe model building process and make associated diagram



4.5. Predictive versus Explanatory Models

Given a set of inputs, our model will generate an output and we are generally interested in its accuracy. *The model need not have a realistic mechanism for how the world works.* That is, we may primarily be interested in accurately calculating an output value without the model having any scientific, explanatory power of how different parts of the real-world system interact.

4.5.1. A Historical Example

Consider the classic underdog story where Copernicus overthrew the status quo when he proposed (correctly) that the earth orbited the sun instead of the other way around⁷.

⁷ Prof. Richard Fitzpatrick has excellent coverage of the associated mathematics and implications in “A Modern Almagest”: <https://farside.ph.utexas.edu/books/Syntaxis/Almagest/Almagest.html>

4.5. Predictive versus Explanatory Models

The existing Ptolemaic model used a geocentric view of the solar system in which the planets and sun orbited the Earth in perfect circles with an epicycle used to explain retrograde motion (as see in Figure 4.1). Retrograde motion is the term used to describe the apparent, temporarily reversed motion of a planet as viewed from Earth when the Earth is overtaking the other planet in orbit around the sun. This was accurate enough to match the observational data that described the position of the planets in the sky.

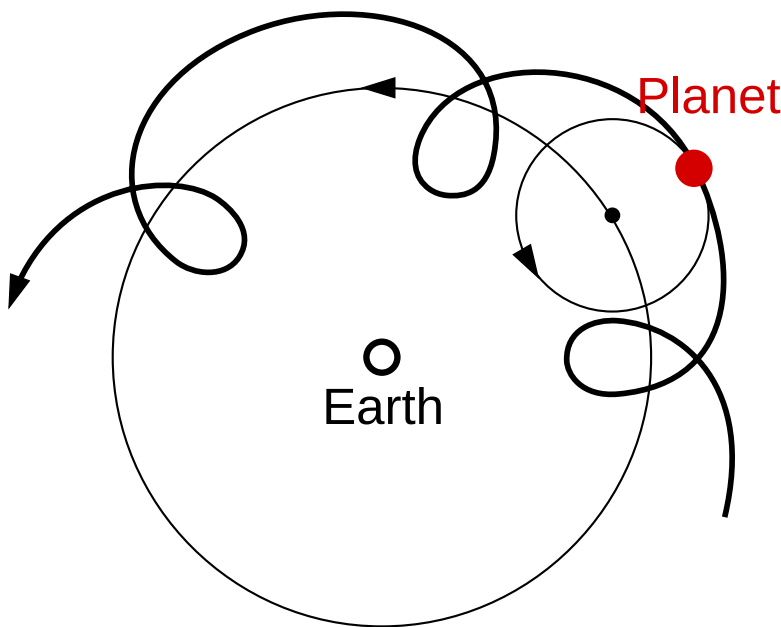


Figure 4.1.: In the Ptolemaic solar model, the retrograde motion of the planets was explained by adding an epicycle to the circular orbit around the earth.

Famously, Copernicus came along and said that the sun, not the Earth, should be at the center (a heliocentric model). Earth revolves around the sun! Today, we know this to be a much better description of reality than one in which the Earth arrogantly sits at the center of the universe. However the model was actually slightly *less* accurate in predicting the apparent position of the planets (to the limits of observational precision

4. Elements of Financial Modeling

at the time)! Why would this be?

First, the Copernican proposal still used perfectly circular orbits with an epicycle adjustment, which we know today to be inaccurate (in favor of an elliptical orbit consistent with the theory of gravity). *Despite being more scientifically correct, it was still not the complete picture.*

Second, the geocentric model was already very accurate because it was essentially a Taylor-series approximation which described to sufficient observational accuracy the apparent position of the planet relative to the Earth. *The heliocentric model was effectively a re-parameterization of the orbital approximation.*

Third, we have considered a limited criteria for which we are evaluating the model for accuracy, namely apparent position of the planets. *It's not until we contemplate other observational data that the Copernican model would demonstrate greater modeling accuracy:* apparent brightness of the planets as they undergo retrograde motion and angular relationship of the planets to the sun.

For modelers today, this demonstrates a few things to keep in mind:

1. Predictive models need not have a scientific, causal structure to make accurate predictions.
2. It is difficult to capture the complete scientific inter-relationships of a system and much care and thought needs to be given in what aspects are included in our model.
3. We should look at, or seek out, additional data that is related to our model because we may accurately fit (or overfit) to one outcome while achieving an increasingly poor fit to other related variables.

Striving to better understand the world is a *good thing to do* but trying to include more components into the model is not always going to help achieve our goals.

4.5. Predictive versus Explanatory Models

4.5.2. Examples in the Financial Context

4.5.2.1. Home Prices

American home prices which have a strong degree of seasonality and have the strongest prices around April of each year. We may find that including a simple oscillating term in our model captures the variability in prices *better* than if we tried to imperfectly capture the true market dynamics of home sales: supply and demand curves varying by personal (job bonus payment timing, school calendars), local (new homes built, company relocation), and national (monetary policy, tax incentives for home-ownership). In other words, one could likely predict a stable pattern like this with a model that contains a simple sinusoidal periodic component. One could likely spend months trying to build a more scientific model and not achieve as good of fit, *even though the latter tries to be more conceptually accurate*.

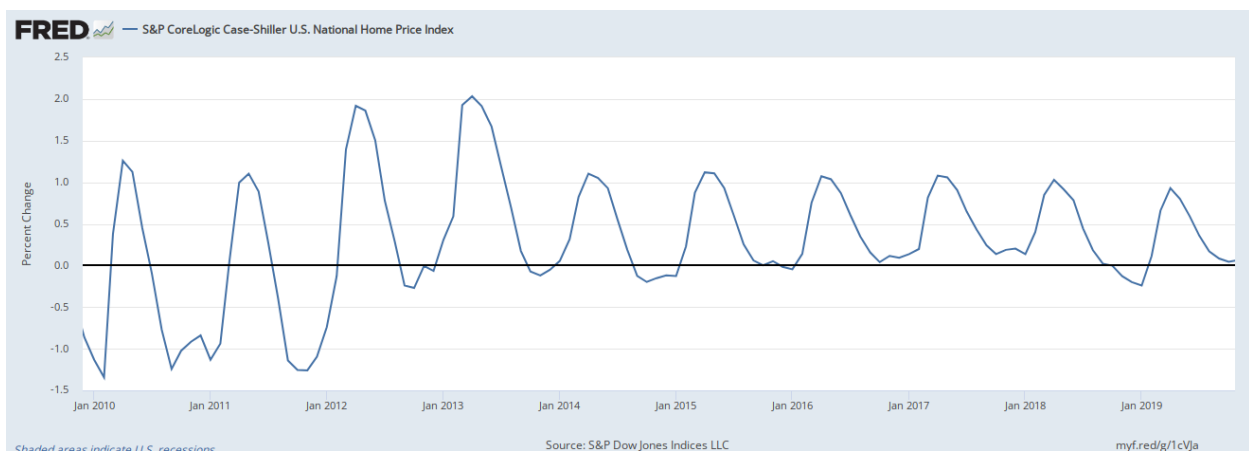


Figure 4.2.: House prices peak around April of each year.

4.5.2.2. Replicating Portfolio

Another example in the financial modeling realm: in attempting to value a portfolio of insurance contracts a **replicating portfolio** of hypothetical assets will sometimes be constructed⁸. The point of this is to create a basket of assets that can be more

⁸ See, e.g., SOA Investment Symposium March 2010. *Replicating Portfolios in the Insurance Industry* (Curt Burmeister Mike Dorsel Patricia Matson)

4. Elements of Financial Modeling

quickly (minutes to hours) valued in response to changing market conditions than it would take to run the actuarial model (hours to days). This is an example where the basket of assets has no ability to explain why the projected cashflows are what they are - but retains strong predictive accuracy.

4.6. What makes a good model?

The answer is: *it depends*.

4.6.1. Achieving original purpose

A model is built for a specific set of reasons and therefore we must evaluate a model in terms of achieving that goal. We should not critique a model if we want to use it outside of what it was intended to do. This includes: contents of output and required level of accuracy.

A model may have been created to for scenario analysis to value all assets in a portfolio to within half a percent of a more accurate, but much more computationally expensive model. If we try to add a never-before-seen asset class or use the model to order trades we may be extending the design scope of the original model.

4.6.2. Usability

How easy is it for someone to use? Does it require pages and pages of documentation, weeks of specialized training and an on-call help desk? *All else equal*, it is an indicator of how usable the model is by the amount of support and training. However, one may sometimes wish to create a highly capable, complex model which is known to require a high amount of experience and expertise. An analogy here might be the cockpit of a small Cessna aircraft versus a fighter jet: the former is a lot simpler and takes less training to master but is also more limited.

Figure 4.3 illustrates this concept and shows that if your goal is very high capability that you may need to expect to develop

4.6. What makes a good model?

training materials and support the more complex model. On this view, a better model is one that is able to have a shorter amount of time and experience to achieve the same level of capability.

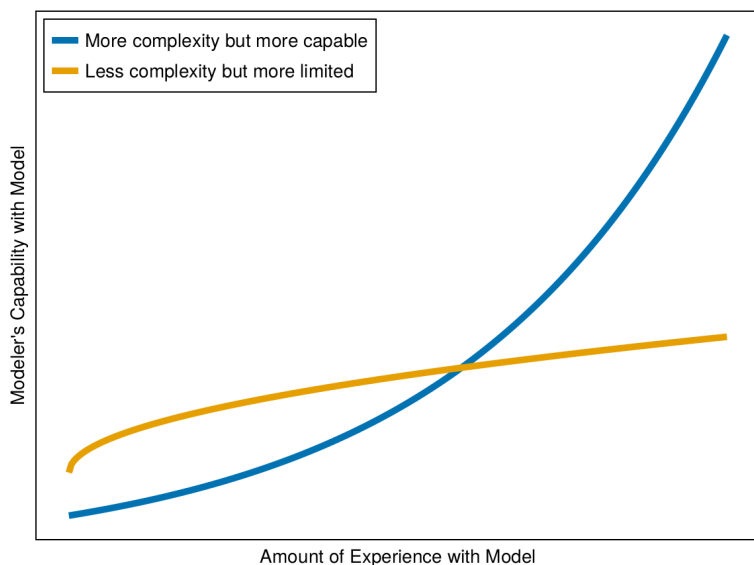


Figure 4.3.: Tradeoff between complexity and capability

4.6.3. Performance

Financial models are generally not used for their awe-inspiring beauty - users are results oriented and the faster a model returns the requested results, the better. Aside from direct computational costs such as server runtime, a shorter model runtime means that one can iterate faster, test new ideas on the fly, and stay focused on the problem at hand.

Many readers may be familiar with the cadence of (1) try running model overnight, (2) see results failed in the morning, (3) spend day developing, (4) repeat step 1. It is preferred if this cycle can be measured in minutes instead of hours or days.

Of course, requirements must be considered here too: needs for high frequency trading, daily portfolio rebalancing, and quarterly valuations are different when it comes to performance.

4.7. What makes a good modeler?

A model is nothing without it's operator, and a skilled practitioner is worth their weight in gold. What elements separate a good modeler from a mediocre modeler?

4.7.1. Domain Expertise

4.7.2. Model Theory

4.7.3. Curiosity

4.7.4. Rigor

4.7.5. Toolset

... The skills in this book!

5. Elements of Programming

“Programming is not about typing, it’s about thinking.” — Rich Hickey (2011)

5.1. In this section

Start building up computer science concepts by introducing tangible programming essentials. Data types, variables, control flow, functions, and scope are introduced.

5.2. Computer Science, Programming, and Coding

Computer Science is the study of computing and information. As a science, it is distinct from programming languages which are merely coarse implementations of specific computer science concepts⁹. Programming (or “coding”) is the art and science of writing code in programming languages to have the computer perform desired tasks. While this may sound mechanistic, programming truly is one of the highest forms of abstract thinking and the design space of potential solutions is so large and potentially complex that much art and experience is needed to create a well-made program.

The language of computer science also provides a lexicon so that financial practitioners can discuss model architecture and problem characteristics. Having the language to describe a concept will also help see aspects of the problem in new ways, opening one up to more innovative solutions.

In the context of this financial modeling that we do, we can consider a financial model to be a type of computer program.

⁹ Said differently, computer science may contemplate ideas and abstractions more generally than a specific implementation, as in mathematics where a theorem may be proved ($a^2 + b^2 = c^2$) without resorting to specific numeric examples ($3^2 + 4^2 = 5^2$).

5. *Elements of Programming*

It takes as input abstract information (data), performs calculations (an algorithm), and returns new data as an output. In this context, we generally do not need to consider many things that a software engineer may contemplate such as a graphical user interface, networking, or access restrictions. But there are many similarities: a good financial modeler must understand data types, algorithms, and some hardware details.

We will build up the concepts over this and the following chapter:

- This chapter will provide a survey of important concepts in computer science that will prove useful for our financial modeling. First, we will talk about data types, boolean logic, and basic expressions. We'll build on those to discuss algorithms (functions) which perform useful work and use control flow and recursion.
- The following chapter will step back and discuss higher level concepts: the “schools of thought” around organizing the relationship between data and functions (functional versus object-oriented programming), design patterns, computational complexity, and compilation.

Tip

There will be brief references to hardware considerations for completeness, but hardware knowledge is not necessary to understand most programming languages (including Julia). It's impossible to completely avoid talking about hardware when you care about the performance of your code, so feel free to gloss over the reference to hardware details on the first read and come back later after Chapter 7.

5.3. Assignment and Variables

One of the first things it will be convenient to understand is the concept of variables. In virtually every programming language, we can assign values to make our program more organized and meaningful to the human reader. In the following example, we

5.4. Data Types

assign values to intermediate symbols to benefit us humans as we convert (silly!) American distance units:

```
feet_per_yard = 3
yards_per_mile = 1760

feet = 3000
miles = feet / feet_per_yard / yards_per_mile
```

0.5681818181818182

Beyond readability, variables are a form of **abstraction** which allows us to think beyond specific instances of data and numbers to a more general representation. For example, the last line in the prior code example is a very generic computation of a unit conversion relationship and `feet` could be any number and the expression remains a valid calculation.

We will return to this subject in more detail in (**ref-assignment?**).

?quarto-cite:ref-assignment

5.4. Data Types

Data types are a way of categorizing information by intrinsic characteristics. We instinctively know that 13.24 is different than "this set of words" and types are how we will formalize this distinction. This is a key conceptual point, and mathematically it's like we have different sets of objects to perform specialized operations on. Beyond this set-like abstraction is implementation details related to computer hardware. You probably know that computers only natively "speak" in binary zeros and ones. Data types are a primary way that a computer can understand if it should interpret 01000010 as B or as 66¹⁰.

Each 0 or 1 within a computer is called a **bit** and eight bits in a row form a **byte** (such as 01000010). This is where we get terms like "gigabytes" or "kilobits per second" as a measure of the quantity or rate of bits something can handle¹¹.

¹⁰ This binary representations correspond to B and 66 with the *ASCII character set* and 8-bit integer encodings, discussed later in this chapter.

¹¹ Some distinctions you may encounter: in short-form, "kb" means *kilobits* while the upper-case "B" in "kB" means *kilobytes*. Also confusingly, sometimes the "k" can be binary or decimal - because computers speak in binary, a binary "k" means 1024 (equal to 2^{10}) instead of the usual decimal 1000. In most computer contexts, the binary (multiples of 1024) is more common.

5.4.1. Numbers

¹² The term floating point refers to the fact that the number's radix (decimal) point can “float” between the significant digits of the number.

¹³ That is, it *reads* the code input from the user, *evaluates* what code was given to it, *prints* the result of the input to the screen, and *loops* through the process again.

Numbers are usually grouped into two categories: **integers** and **floating-point**¹² numbers. Integers are like the mathematical set of integers while floating-point is a way of representing decimal numbers. Both have some limitations since computers can only natively represent a finite set of numbers due to the hardware (more on this in Chapter 7). Here are three integers that are input into the **REPL** (Read-Eval-Print-Loop)¹³ and the result is **printed** below the input:

```
2
```

```
2
```

```
423
```

```
423
```

```
1929234
```

```
1929234
```

And three floating-point numbers:

```
0.2
```

```
0.2
```

```
-23.3421
```

```
-23.3421
```

```
14e3      # the same as 14,000.0
```

5.4. Data Types

14000.0

On most systems, `0.2` will be interpreted as a 64-bit floating point type called `Float64` in Julia since most architectures these days are 64-bit¹⁴, while on a 32-bit system `0.2` would be interpreted as a `Float32`. Given that there are a finite amount of bits attempting to represent a continuous, infinite set of numbers means that some numbers are not able to be represented with perfect precision. For example, if we ask for `0.2`, the closest representations in 64 and 32 bit are:

- `0.200000000298023223876953125` in 32-bit
- `0.2000000000000000011102230246251565404236316680908203125` in 64-bit

This leads to special considerations that computers take when performing calculations on floating point maths, some of which will be covered in more detail in Chapter 7. For now, just note that floating point numbers have limited precision and even if we input `0.2`, your computations will use the above decimal representations even if it will print out a number with fewer digits shown:

```
x = 0.2 ①  
big(x) ②
```

- ① Here, we **assign** the value `0.2` to a **variable** `x`. More on variables/assignments in Section 5.5.1.
- ② `big(x)` is a arbitrary precision floating point number and by default prints the full precision that was embedded in our variable `x`, which was originally `Float64`.

0.2000000000000000011102230246251565404236316680908203125

Note

Note the difference in what printed between the last example and when we input `0.2` earlier in the chapter. The former had the same (not-exactly equal to `0.2`) *value*, but

¹⁴ This means that their central processing units (CPUs) use instructions that are 64 bits long.

5. Elements of Programming

it printed an abbreviated set of digits as a nicety for the user, who usually doesn't want to look at floating point numbers with their full machine precision. The system has the full precision (0.20...3125) but is truncating the output.

In the last example, we've converted the normal `Float64` to a `BigFloat` which will not truncate the output when printing.

Integers are similarly represented as 32 or 64 bits (with `Int32` and `Int64`) and are limited to exact precision:

- -32,767 to 32,767 for `Int32`
- -2,147,483,647 to 2,147,483,647 for `Int64`

Additional range in the positive direction if one chooses to use “unsigned”, non-negative numbers (`UInt32` and `UInt64`). Unlike floating point numbers, the integers have a type `Int` which will use the system bit architecture by default (that is, `Int{30}` will create a 64 bit integer on 64-bit systems and 32-bit on 32-bit systems)

5.4.2. Type Hierarchy

We can describe a *hierarchy* of types. Both `Float64` and `Int64` are examples of `Real` numbers (here, `Real` is an **abstract** Julia type which corresponds to the mathematical set of real numbers commonly denoted with \mathbb{R}). Both `Float64` and `Int32` are `Real` numbers, so why not just define all numbers as a `Real` type? Because for performant calculations, the computer must know in advance how many bits each number is represented with.

Figure 5.1 shows the type hierarchy for most built-in Julia number types.

The integer and floating point types described in the prior section are known as **concrete** types because there are no possible sub types (child types). Further, a concrete type can be a **bit type** if the data type will always have the same number of bits in memory: a `Float32` will always be 32 bits in memory, for

5.4. Data Types

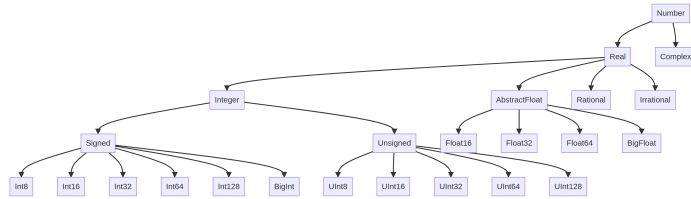


Figure 5.1.: Numeric Type Hierarchy in Julia. Leafs of the tree are concrete types.

example. Contrast this with strings (described below) which can contain an arbitrary number of characters.

5.4.3. Arrays

Arrays are the most common way to represent a collection of similar data. For example, we can represent a set of integers as follows:

```
[1, 10, 300]
```

3-element Vector{Int64}:

```
1
10
300
```

And a floating point array:

```
[0.2, 1.3, 300.0]
```

3-element Vector{Float64}:

```
0.2
```

Julia has very powerful and friendly array types.

5. Elements of Programming

```
1.3
300.0
```

Note the above two arrays are different types of arrays. The first is `Vector{Int64}` and the second is `Vector{Float64}`. These are arrays of concrete types and so Julia will know that each element of an array is the same amount of bits which will enable more efficient computations. With the following set of mixed numbers, Julia will **promote** the integers to floating point since the integers can be accurately represented¹⁵ in floating point.

¹⁵ Accurate only to a limited precision, as described in Section 5.4.1.

```
[1, 1.3, 300.0, 21]
```

4-element `Vector{Float64}`:

```
1.0
1.3
300.0
21.0
```

However, if we explicitly ask Julia to use a `Real`-typed array, the type is now `Vector{Real}`. Recall that `Real` is an abstract type. Having heterogeneous types within the array is conceptually fine, but in practice limits performance. Again, this will be covered in more detail in Chapter 7.

In Julia, arrays can be multi-dimensional. Here are two three-dimensional arrays with length three in each dimension:

```
rand(3, 3, 3)
```

3×3×3 `Array{Float64, 3}`:

```
[:, :, 1] =
 0.184476  0.301779  0.954145
 0.0850162 0.580595  0.419575
 0.401745  0.599289  0.993595
```

```
[:, :, 2] =
 0.23618  0.999788  0.354967
 0.845709 0.52029  0.242072
```


5.4. Data Types

```
0.0895747 0.853431 0.992184
```

```
[:, :, 3] =  
0.697881 0.488533 0.198338  
0.933285 0.885645 0.453481  
0.454158 0.209277 0.111265
```

```
[x + y + z for x in 1:3, y in 11:13, z in 21:23]
```

```
3×3×3 Array{Int64, 3}:
```

```
[:, :, 1] =  
33 34 35  
34 35 36  
35 36 37
```

```
[:, :, 2] =  
34 35 36  
35 36 37  
36 37 38
```

```
[:, :, 3] =  
35 36 37  
36 37 38  
37 38 39
```

The above example demonstrates **array comprehension** syntax which is a convenient way to create arrays in Julia.

A two-dimensional array has the rows by semi-colons (;):

```
x = [1 2 3; 4 5 6]
```

```
2×3 Matrix{Int64}:
```

```
1 2 3  
4 5 6
```

5. Elements of Programming

i Note

In Julia, a `Vector{Float64}` is simply a one-dimensional array of floating points and a `Matrix{Float64}` is a two-dimensional array. More precisely, they are **type aliases** of the more generic `Array{Float64,1}` and `Array{Float64,2}` names.

5.4.3.1. Array indexing

Array elements are accessed with the integer position, starting at 1 for the first element¹⁶ ¹⁷:

```
v = [10, 20, 30, 40, 50]
v[2]
```

20

We can also access a subset of the vector's contents by passing a range:

```
v[2:4]
```

3-element Vector{Int64}:

```
20
30
40
```

And we can generically reference the array's contents, such as:

```
v[begin+1:end-1]
```

3-element Vector{Int64}:

```
20
30
40
```

50

¹⁶ Whether an index starts at 1 or 0 is sometimes debated. Zero-based indexing is natural in the context of low-level programming which deal with bits and positional *offsets* in computer memory. For higher level programming one-based indexing is more natural: in a set of data stored in an array, it is much more natural to reference the *first* (through n^{th}) datum instead of the *zeroth* (through $(n-1)^{th}$) datum.

¹⁷ Arrays in Julia can actually be indexed with an arbitrary starting point: see the package `OffsetArrays.jl`

5.4. Data Types

We can assign values into the array as well, as well as combine arrays and push new elements to the end:

```
v[2] = -1
push!(v, 5)
vcat(v, [1, 2, 3])
```

9-element Vector{Int64}:

```
10
-1
30
40
50
5
1
2
3
```

5.4.4. Characters, Strings, and Symbols

Characters are represented in most programming languages as letters within quotation marks. In Julia, individual characters are represented using single quotes:

```
'a'
```

`'a': ASCII/Unicode U+0061 (category Ll: Letter, lowercase)`

Letters and other characters present more difficulties than numbers to represent within a computer (think of how many languages and alphabets exist!), and it essentially only works because the world at large has agreed to a given representation. Originally **ASCII** (American Standard Code for Information Interchange) was used to represent just 95 of the most common English characters (“a” through “z”, zero through nine, etc.). Now, UTF (Unicode Transformation Format) can encode more than a million characters and symbols from many human languages.

5. Elements of Programming

Strings are a collection¹⁸ of characters, and can be created in Julia with double quotes:

```
"hello world"
```

```
"hello world"
```

It's easy to ascertain how 'normal' characters can be inserted into a string, but what about things like new lines or tabs? They are represented by their own characters but are normally not printed in computer output. However, those otherwise invisible characters do exist. For example, here we will use a **string literal** (indicated by the `"""`) to tell Julia to interpret the string as given, including the invisible new line created by hitting return on the keyboard between the two words:

```
"""  
hello  
world  
"""
```

```
"hello\nworld\n"
```

The output above shows the `\n` character contained within the string.

Symbols are a way of representing an identifier which cannot be seen as a collection of individual characters. `:helloworld` is distinct from `"helloworld"` - you can kind of think of the former as an un-executed bit of code - if we were to execute it (with `eval(:helloworld)`), we would get an error `UndefVarError: `a` not defined`. Symbols can *look* like strings but do not behave like them. For now, it is best to not worry about symbols but it is an important aspect of Julia which allows the language to represent aspects of itself as data. This allows for powerful self-reference and self-modification of code but this is a more advanced topic generally out of scope of this book.

¹⁸ Under the hood, strings are essentially a vector of characters. There are considerations about how they are encoded (UTF-8) and how they are converted to a uniform bit representation for computational purposes. This avoids making strings larger than necessary.

5.4.5. Tuples

Tuples are a set of values that belong together and are denoted by a values inside parenthesis and separated by a comma. An example might be x-y coordinates in 2 dimensional space:

```
x = 3
y = 4
location = (x, y)
```

(3, 4)

Tuple's values can be accessed like arrays:

```
{location[1]}
```

Tuples fill a middle ground between scalar types and arrays in more ways that one:

- Tuples have no problem having heterogeneous types in the different slots.
- Tuples are **immutable**, meaning that you cannot overwrite the value in memory.
- It's generally expected that within an array, you would be able to apply the same operation to all the elements (e.g. square each element) or do something like sum all of the elements together which isn't generally case for a tuple.
- Tuples are generally stack allocated instead of being heap allocated like arrays¹⁹, meaning that a lot of times they can be faster than arrays.

¹⁹ What this means will be explained in Chapter 7 .

5.4.5.1. Named Tuples

Named tuples provide a way to give each field within the tuple a specific name. For example, our x-y coordinate example above could become:

```
location2 = (x=3, y=4)
```

5. Elements of Programming

```
(x = 3, y = 4)
```

The benefit is that we can give more meaning to each field and access the values in a nicer way. Previously, we used `location[1]` to access the x-value, but with the new definition we can access it by name:

```
location2.x
```

3

5.4.6. Parametric Types

We just saw how tuples can contain heterogeneous types of data inside a common container. Let's look at this a little bit closer by looking at the full type:

```
typeof(location)
```

```
Tuple{Int64, Int64}
```

`location` is a `Tuple{Int64,Int64}` type, which means that its first and second elements are both `Int64`. Contrast this with:

```
typeof(("hello", 1.0))
```

```
Tuple{String, Float64}
```

These tuples are both of the form `Tuple{T,U}` where `T` and `U` are both types. Why does this matter? We and the compiler can distinguish between a `Tuple{Int64,Int64}` and a `Tuple{String,Float64}` which allows us to reason about things (“I can add the first element of tuple together only if both are numbers”) and the compiler to optimize (sometimes it can know exactly how many bits in memory a tuple of a certain kind will need and be more efficient about memory use). Further, we will see how this can become a powerful force

in writing appropriately abstracted code and more logically organize our entire program when we encounter “multiple dispatch” later on.

5.4.7. Creating User Defined Types

We’ve talked about some built-in types but so much additional capabilities come from being able to define our own types. For example

5.5. Expressions and Control Flow

5.5.1. More on Assignment and Variables

5.5.2. Operators

unary/binary

5.6. Functions

5.6.1. Broadcasting

5.7. Scope

5.7.1. Local Scope

5.7.2. Global Scope

5.7.3. Modules

Note

Summarizing related terminology:

- A **module** is a block of code such as `module MySimulation ... end`

5. *Elements of Programming*

- A **package** is a module that has a specific set of files and associated metadata. Essentially, it's a module with a `Project.toml` file that has a name and unique identifier listed, and a file in a `src/` directory called `MySimulation.jl`
 - **Library** is just another name for a package, and the most common context this comes up is when talking about the packages that are bundled with Julia itself called the **standard library** (`stdlib`).

6. Elements of Computer Science

“Fundamentally, computer science is a science of abstraction—creating the right model for a problem and devising the appropriate mechanizable techniques to solve it. Confronted with a problem, we must create an abstraction of that problem that can be represented and manipulated inside a computer. Through these manipulations, we try to find a solution to the original problem.” - Al Aho and Jeff Ullman (1992)

6.1. In this section

Adapting computer science concepts to work for financial professionals. Concepts like computability, computational complexity, the language of algorithms and problem solving, looking for and using patterns, and adopting digital-first practices to automate the boring parts of the job.

6.2. Computer Science for Financial Professionals

Computer science as a term can be a bit misleading because of the overwhelming association with the physical desktop or laptop machines that we call “computers”. The discipline of computer science is much richer than consumer electronics: at its core, computer science concerns itself with areas of research and answering tough questions:

6. Elements of Computer Science

- **Algorithms and Optimization.** How can a problem be solved efficiently? How can that problem be solved *at all*? Given constraints, how can one find an optimal solution?
- **Information Theory.** Given limited data, what *can* be known or inferred from it?
- **Theory of Computation.** What sorts of questions are even answerable? Is an answer easy to computer or will resolving it require more resources than the entire known universe? Will a computation ever stop calculating?
- **Data Structures.** How to encode, store, and use data? How does that data relate to each other and what are the trade-offs between different representations of that data?

For a reader in the twenty-first century we hope that's it's patently obvious how impactful the *applied* computer science has been as an end-user of the internet, artificial intelligence, computational photography, safety control systems, etc., etc. have been to our lives. It is a testament to the utility of being able to harness some of the ideas of this science is. Many of the most impactful advances occur at the boundary between two disciplines. It's here in this chapter that we desire to bring together the financial discipline together with computer science and to provide the financial practitioner with the language and concepts to leverage some of computer science's most relevant ideas.

6.3. Algorithms

6.3.1. Computational Complexity

6.4. Data Structures

6.5. Information Theory

6.5.1. Signal vs Noise

6.6. Formal Verification

6.7. The Discipline of Software Engineering

6.7.1. Patterns

7. Hardware and It's Implications

7.1. In this section

A discussion of why a cursory understanding of modern computing hardware and architecture is important for making the right design decisions within a modeling context. Stack vs heap allocations, pointers. A discussion of parallelism and the different kinds of parallelism.

8. Applying Software Engineering Principles

“Programs must be written for people to read,
and only incidentally for machines to execute.” —
Harold Abelson and Gerald Jay Sussman (1984)

8.1. In this section

We describe modern software engineering practices such as testing, documentation, and pipelines which can be utilized by the financial professional to make their own work more robust and automated.

Part III.

Computational Thinking in an Actuarial and Financial Context

9. Modeling

9.1. In This Chapter

We discuss how to approach a problem and identify the key attributes to include in the model, what are the inherent trade-offs with different approaches, and how to work with data that feeds your model.

9.2. Parsimony

10. Optimization

10.1. In This Chapter

Optimization as root finding or minimization/maximization of defined objectives. Differentiable programming and the benefits to optimization problems. Model fitting as an optimization problem.

11. Sensitivity Analysis

11.1. In This Chapter

Different approaches to understanding the sensitivity of a model to changes in its inputs: derivatives, finite differences, global sensitivity analysis approaches, and statistical approaches.

12. Stochastic Modeling

The Monte Carlo Method: (i) A last resort when doing numerical integration, and (ii) a way of wastefully using computer time. - Malvin H. Kalos²⁰ (c. 1960)

²⁰ Kalos was a pioneer in Monte Carlo techniques, quoted via https://doi.org/10.1007/978-3-540-74686-7_3

13. Visualizations

13.1. In This Chapter

The evolved brain and pattern recognition, recommended principles for looking at data, and avoiding common mistakes. Exploratory visualization versus visualizations intended for an audience.

14. Matrices and Their Uses

14.1. In This Chapter

Matrices and their myriad uses: reframing problems through the eyes of linear algebra, an intuitive refreshing on applicable maths, and recurring patterns of matrix operations in financial modeling.

15. Learning from Data

15.1. In this chapter

Using data to inform a model: fitting parameters, forecasting, and fundamental limitations on prediction. Also covered are elements of practical review such as static and dynamic validations, and implied rate analysis.

Part IV.

Applications in Practice

16. Stochastic Mortality Projections

16.1. In This Chapter

A term life insurance policy is used to illustrate: selecting key model features, design tradeoffs between a few different approaches, and a discussion of the performance impacts of the different approaches to parallelism.

17. Scenario Generation

17.1. In This Chapter

How to generate synthetic data for your model using sub-models, with applications to economic scenario generation and portfolio composition.

18. Similarity Analysis

18.1. In This Chapter

Given a set of interest, understanding the relative similarity (or not) of features of interest is useful in classification and data compression techniques.

19. Portfolio Optimization

19.1. In This Chapter

Optimization in a portfolio context with examples of asset selection under different constraints and objectives.

20. Other Useful Techniques

20.1. In this chapter

Other useful techniques are surveyed, such as: memoization to avoid repeated computations, psuedo-monte carlo, creating a model office, and tips on modeling a complete balance sheet.

20.2. Taking things to the Extreme

Consider what happens if something is taken to an extreme. For example, what happens in the model if we input negative rates? Where should negative rates be allowed and can the model handle them?

20.3. Range Bounding

Sometimes you just need to know that an outcome is within a certain range - if you can develop a “high” and “low” estimate by making assumptions that you know are outside of feasible ranges, then you can determine whether something is reasonable or within tolerances.

To take an example from the pages of interview questions: say you need to determine if a mortgaged property’s value is greater than the amount of the outstanding loan (say \$100,000). You don’t have an appraisal, but know that it’s in reasonable condition and that (1) a comparable house with many more issues sold for \$100 per square foot. You also don’t know the square footage of the house, but know from the number of rooms and layout that it must be at least 1000 square feet. Therefore you know that the value should at least be greater than:

20. *Other Useful Techniques*

$$\frac{\$100}{\text{sq. ft}} \times 1000\text{sq. ft} = \$100,000$$

We'd then conclude that the value of the house very likely exceeds the outstanding balance of the loan and resolves our query without complex modeling or expensive appraisals.

Part V.

Appendices

21. Set up Julia and the Computing Environment

21.1. Installation

Julia is open source and can be downloaded from [JuliaLang.org](https://julialang.org) and is available for all major operating systems. After you download and install, then you have Julia installed and can access the **REPL**, or Read-Eval-Print-Loop, which can run complete programs or function as powerful day-to-day calculator. However, many people find it more comfortable to work in a text editor or **IDE** (Integrated Development Environment).

If you are looking for managed installations with a curated set of packages for use within an organization, there are ways to self-host package repositories and otherwise administratively manage packages. Julia Computing offers managed support with enterprise solutions, including push-button cloud compute capabilities.

21.2. Package Management

Julia comes with `Pkg`, a built-in package manager. With it, you can install packages, pin certain versions, recreate environments with the same set of dependencies, and upgrade/remove/develop packages easily. It's one of the things that *just works* and makes Julia stand out versus alternative languages that don't have a de-facto way of managing or installing packages.

Package installation is accomplished interactively in the REPL or executing commands.

21. Set up Julia and the Computing Environment

- In the REPL, you can change to the Package Management Mode by hitting `]` and, e.g., add `DataFrames` `CSV` to install the two packages. Hit `[backspace]` to exit that mode in the REPL.
- The same operation without changing REPL modes would be: using `Pkg`; `Pkg.add(["DataFrames", "CSV"])`

Related to packages, are **environments** which are a self-contained workspaces for your code. This lets you install only packages that are relevant to the current work. It also lets you ‘remember’ the exact set of packages and versions that you used. In fact, you can share the environment with others, and it will be able to recreate the same environment as when you ran the code. This is accomplished via a `Project.toml` file, which tracks the direct dependencies you’ve added, along with details about your project like its version number. The `Manifest.toml` tracks the entire dependency tree.

Reproducibility via the environment tools above is a really key aspect that will ensure Julia code is consistent across time and users, which is important for financial controls.

21.3. Editors

Because Julia is very extensible and amenable to analysis of its own code, you can typically find plugins for whatever tool you prefer to write code in. A few examples:

21.3.1. Visual Studio Code

Visual Studio Code is a free editor from Microsoft. There’s a full-featured Julia plugin available, which will help with auto-completion, warnings, and other code hints that you might find in a dedicated editor (e.g. PyCharm or RStudio). Like those tools, you can view plots, search documentation, show datasets, debug, and manage version control.

21.3.2. Notebooks

Notebooks are typically more interactive environments than text editors - you can write code in cells and see the results side-by-side.

The most popular notebook tool is Jupyter (“Julia, Python, R”). It is widely used and fits in well with exploratory data analysis or other interactive workflows. It can be installed by adding the `IJulia.jl` package.

`Pluto.jl` is a newer tool, which adds reactivity and interactivity. It is also more amenable to version control than Jupyter notebooks because notebooks are saved as plain Julia scripts. Pluto is unique to Julia because of the language’s ability to introspect and analyze dependencies in its own code. Pluto also has built-in package/environment management, meaning that Pluto notebooks contains all the code needed to reproduce results (as long as Julia and Pluto are installed).

22. The Julia Ecosystem Today

A tour of relevant available packages as of 2023.

The Julia ecosystem favors composability and interoperability, enabled by multiple dispatch. In other words, because it's easy to automatically specialize functionality based on the type of data being used, there's much less need to bundle a lot of features within a single package.

As you'll see, Julia packages tend to be less vertically integrated because it's easier to pass data around. Counterexamples of this in Python and R:

- Numpy-compatible packages that are designed to work with a subset of numerically fast libraries in Python
- special functions in Pandas to read CSV, JSON, database connections, etc.
- The Tidyverse in R has a tightly coupled set of packages that works well together but has limitations with some other R packages

Julia is not perfect in this regard, but it's neat to see how frequently things *just work*. It's not magic, but because of Julia features outside the scope of this article it's easy for package developers (and you!) to do this.

Julia also has language-level support for documentation, so packages can follow a consistent style of help-text and have the docs be auto-generated into web pages available locally or online.

The following highlighted packages were chosen for their relevance to typical actuarial work, with a bias towards those used regularly by the authors. This is a small sampling of the over 6000 registered Julia Packages^[2]

22.0.1. Data

Julia offers a rich data ecosystem with a multitude of available packages. Perhaps at the center of the data ecosystem are `CSV.jl` and `DataFrames.jl`. `CSV.jl` is for reading and writing files text files (namely CSVs) and offers top-class read and write performance. `DataFrames.jl` is a mature package for working with dataframes, comparable to Pandas or dplyr.

Other notable packages include `ODBC.jl`, which lets you connect to any database (given you have the right drivers installed), and `Arrow.jl` which implements the Apache Arrow standard in Julia.

Worth mentioning also is `Dates`, a built-in package making date manipulation straightforward and robust.

Check out [JuliaData.org](https://juliadata.org) for more packages and information.

22.0.2. Plotting

`Plots.jl` is a meta-package providing an interface to consistently work with several plotting backends, depending if you are trying to emphasize interactivity on the web or print-quality output. You can very easily add animations or change almost any feature of a plot.

`StatsPlots.jl` extends `Plots.jl` with a focus on data visualization and compatibility with dataframes.

`Makie.jl` supports GPU-accelerated plotting and can create very rich, beautiful visualizations, but its main downside is that it has not yet been optimized to minimize the time-to-first-plot.

22.0.3. Statistics

Julia has first-class support for missing values, which follows the rules of three-valued logic so other packages don't need to do anything special to incorporate missing values.

`StatsBase.jl` and `Distributions.jl` are essentials for a range of statistics functions and probability distributions respectively.

Others include:

- `Turing.jl`, a probabilistic programming (Bayesian statistics) library, which is outstanding in its combination of clear model syntax with performance.
- `GLM.jl` for any type of linear modeling (mimicking R's `glm` functionality).
- `LsqFit.jl` for fitting data to non-linear models.
- `MultivariateStats.jl` for multivariate statistics, such as PCA.

You can find more packages and learn about them [here](#).

22.0.4. Machine Learning

`Flux`, `Gen`, `Knet`, and `MLJ` are all very popular machine learning libraries. There are also packages for `PyTorch`, `Tensorflow`, and `SciKitML` available. One advantage for users is that the Julia packages are written in Julia, so it can be easier to adapt or see what's going on in the entire stack. In contrast to this design, `PyTorch` and `Tensorflow` are built primarily with C++.

Another advantage is that the Julia libraries can use automatic differentiation to optimize on a wider range of data and functions than those built into libraries in other languages.

22.0.5. Differentiable Programming

Sensitivity testing is very common in actuarial workflows: essentially, it's understanding the change in one variable in relation to another. In other words, the derivative!

Julia has unique capabilities where almost across the entire language and ecosystem, you can take the derivative of entire functions or scripts. For example, the following is real Julia code to automatically calculate the sensitivity of the ending account value with respect to the inputs:

22. The Julia Ecosystem Today

```
julia> using Zygote

julia> function policy_av(pol)
    COIs = [0.00319, 0.00345, 0.0038, 0.00419, 0.0047, 0.00532]
    av = 0.0
    for (i,coi) in enumerate(COIs)
        av += av * pol.credit_rate
        av += pol.annual_premium
        av -= pol.face * coi
    end
    return av                # return the final account value
end

julia> pol = (annual_premium = 1000, face = 100_000, credit_rate = 0.06)

julia> policy_av(pol)        # the ending account value
4048.08

julia> policy_av'(pol)       # the derivative of the account value with
                             # respect to the parameters
(annual_premium = 6.802, face = -0.0275, credit_rate = 10972.52)
```

When executing the code above, Julia isn't just adding a small amount and calculating the finite difference. Differentiation is applied to entire programs through extensive use of basic derivatives and the chain rule. **Automatic differentiation**, has uses in optimization, machine learning, sensitivity testing, and risk analysis. You can read more about Julia's autodiff ecosystem [here](#).

22.0.6. Utilities


There are also a lot of quality-of-life packages, like `Revise.jl` which lets you edit code on the fly without needing to re-run entire scripts.

`BenchmarkTools.jl` makes it incredibly easy to benchmark your code - simply add `@benchmark` in front of what you want to test, and you will be presented with detailed statistics. For example:

```
julia> using ActuaryUtilities, BenchmarkTools

julia> @benchmark present_value(0.05,[10,10,10])

BenchmarkTools.Trial: 10000 samples with 994 evaluations.
 Range (min ... max): 33.492 ns ... 829.015 ns | GC (min ... max): 0.00% ... 95.40%
 Time (median):       34.708 ns                | GC (median):      0.00%
 Time (mean ± σ):     36.599 ns ± 33.686 ns    | GC (mean ± σ):   4.40% ± 4.55%


 33.5 ns          Histogram: log(frequency) by time          45.6 ns <

Memory estimate: 112 bytes, allocs estimate: 1.
```

Test is a built-in package for performing testsets, while Documenter.jl will build high-quality documentation based on your inline documentation.

ClipData.jl lets you copy and paste from spreadsheets to Julia sessions.

22.0.7. Other packages

Julia is a general-purpose language, so you will find packages for web development, graphics, game development, audio production, and much more. You can explore packages (and their dependencies) at <https://juliahub.com/>.

22.0.8. Actuarial packages

Saving the best for last, the next article in the series will dive deeper into actuarial packages, such as those published by JuliaActuary for easy mortality table manipulation, common actuarial functions, financial math, and experience analysis.

References

