

3SA03 Scientific Programming with Julia – DataFrames.jl

Rattachai Wongtanawijit

Semester: 1/2568, Revision: 10 June 2025

คำชี้แจง:

1. แนะนำให้ใช้คอมพิวเตอร์ Laptop ของตนเองสำหรับแลปนี้ (หากมี)
2. มี Quiz ในกระดานชั้นชั่วโมง โปรดศึกษารายละเอียดในเอกสารฉบับนี้ก่อนถึงคาบเรียน

สิ่งที่ต้องทำก่อนเข้าคาบเรียน

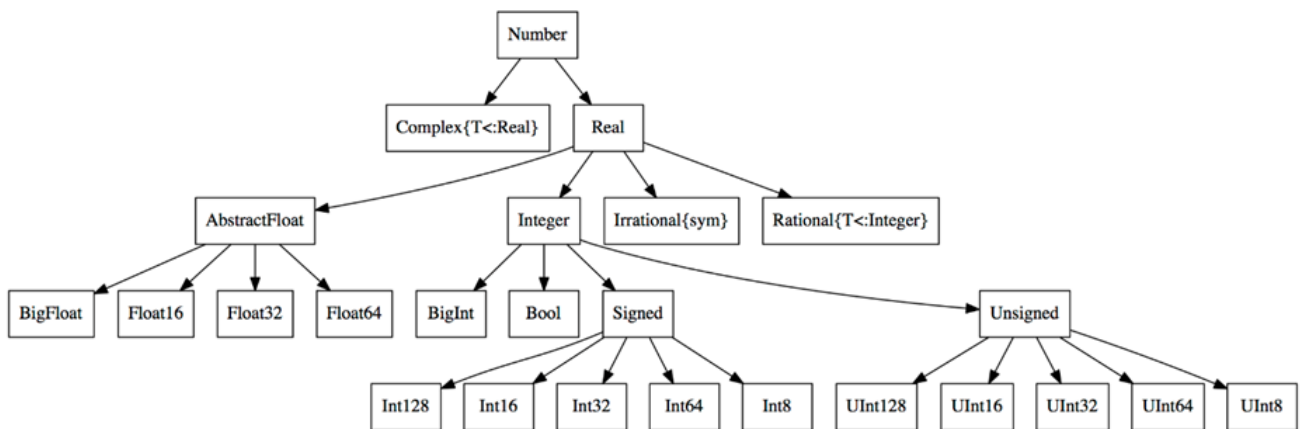
1. ติดตั้ง Julia executable version 1.11.x ในเครื่องคอมพิวเตอร์ของตนเอง ด้วยเครื่องมือที่ชื่อว่า Juliaup (<https://github.com/JuliaLang/juliaup>)
2. ทำความเข้าใจภาษา Julia เบื้องต้น <https://carpentries-incubator.github.io/julia-novice/instructor/aio.html> ในเรื่อง Primitive Types, Function, Array
3. ติดตั้งโปรแกรม Visual Studio Code พร้อมส่วนขยาย Julia, Jupyter extensions
 - Julia Language Support Extension (by JuliaLang)
<https://marketplace.visualstudio.com/items?itemName=julialang.language-julia>
 - Jupyter notebook (by Microsoft)
<https://marketplace.visualstudio.com/items?itemName=ms-toolsai.jupyter>
4. ศึกษาการทำงานของโค้ดในไฟล์ demo.ipynb ดาวน์โหลดจาก LMS

Introduction

JuliaLang หรือ Julia Programming Language เป็นภาษาโปรแกรมระดับสูงแบบมีการคอมไพล์ แบบก่อนการเรียกใช้งานโค้ดโปรแกรม (Just-in-Time Compile) ภาษา Julia ถูกออกแบบมาเป็น General Purpose Programming Language แต่การนำไปใช้งานจากกลุ่มผู้ใช้จะเน้นการเขียนโปรแกรมเชิงคำนวณ การทดลอง หรือการพัฒนาอัลกอริทึมต่างๆ งาน Machine learning รวมไปถึงการวิเคราะห์หรือแสดงผลข้อมูล มีจุดเด่นในเรื่องความเร็วในการทำงาน และรูปแบบไวยากรณ์ภาษา (Syntax) ที่อ่านเข้าใจได้ง่าย ภาษา Julia เป็นภาษาประเภท Strong-dynamic type นั้นหมายถึงข้อมูลที่ถูกอ้างอิงด้วยตัวแปรของภาษาจะมีชนิดกำกับเสมอ โดยโปรแกรมจะไม่รองรับ

การดำเนินการระหว่างตัวแปร (Operation) หากชนิดข้อมูลนั้นไม่ถูกรองรับภาษา และชนิดของตัวแปรมีความ Dynamic หมายถึงชนิดของตัวแปรสามารถเปลี่ยนแปลงได้ขณะ Runtime ชนิดข้อมูลพื้นฐานของภาษา Julia แสดงด้วยกราฟตามรูปที่ 1 และนอกจากนี้ภาษา Julia ยังมีฟีเจอร์การทำงานในรูปแบบอื่น มาโคร (Macros), Multiple-Dispatch เป็นต้น

DataFrames.jl เป็นไลบรารีที่ใช้ประมวลผลข้อมูลในรูปแบบตาราง (Tabular Data) ถูกพัฒนาขึ้นในภาษา Julia เพื่อใช้งานทดแทนหรือเทียบเท่า Pandas ในภาษา Python มีฟังก์ชันการทำงานหลักๆ เช่น การเลือกข้อมูล (querying) ในรูปแบบที่กำหนด, การจัดเรียง (Sorting) และ Dataframe สามารถใช้งานร่วมกับฟังก์ชันพื้นฐานหรือแพ็คเกจอื่นในภาษา Julia ได้เช่นกัน โดยข้อมูลในแต่ละคอลัมน์ของตารางจะมีชนิดข้อมูลกำกับ (Type)



รูปที่ 1 กราฟแสดงชนิดข้อมูลพื้นฐาน (Primitive Types) ของภาษา Julia

ทดลองใช้งาน DataFrames.jl ร่วมกับไลบรารีอื่นๆ

ให้ทดลองเพื่อให้ได้ผลลัพธ์แต่ละข้อต่อไปนี้ นักศึกษาอาจใช้โค้ดตัวอย่างจากไฟล์ demo.ipynb หรืออาจใช้คำสั่งอื่นๆ นอกเหนือ ในการทำโจทย์แต่ละข้อ โดยบังคับให้นักศึกษาใช้ไลบรารี DataFrames.jl, CSV.jl, Plots.jl, StatsPlots.jl เป็นอย่างน้อย

1. ให้นักศึกษาดาวน์โหลดไฟล์ CSV (Comma Separated Value) [personality_dataset.csv](#) ซึ่งเป็นข้อมูล *Social Behaviors and Personality Types* จากเว็บไซต์ Kaggle พร้อมทำความเข้าใจความหมายของข้อมูลในแต่ละคอลัมน์ <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data> และโหลด `personality_dataset.csv` เข้าสู่โปรแกรมด้วย DataFrames.jl

2. หาค่าเฉลี่ยของจำนวนครั้งของการเข้าสังคม (Social_event_attendance) ของกลุ่มที่เป็น Introvert เปรียบเทียบกับกลุ่มที่เป็น Extrovert พร้อมแสดง Density Plot เปรียบเทียบทั้งสองกลุ่ม
3. แสดง 2D Scatter Plot ระหว่างขนาดกลุ่มเพื่อนสนิท (Friends_circle_size) กับเวลาที่ชอบอยู่ตามลำพังคนเดียว (Time_spent_Alone) ของกลุ่มคนที่เป็น Extrovert และ Introvert โดยให้นักศึกษานำแถวที่ไม่มีข้อมูล (Missing Data) ออกก่อน
4. แสดง Box Plot ของค่าการโพลสนสื่อสังคมออนไลน์ ของกลุ่มคนที่เป็น Introvert เทียบกับ Extrovert
5. สร้าง DataFrame ใหม่ 1 Dataframe จากการสุ่มข้อมูลมาจาก DataFrame เดิม ให้มีจำนวนคนที่เป็น Introvert 20 คน และ Extrovert 20 คน พร้อมแสดง DataFrame ที่ได้แยกตามกลุ่ม Introvert, Extrovert

การส่งงาน

1. ให้นักศึกษาเขียนโค้ดในภาษา Julia ในไฟล์ Jupyter notebook พร้อมแสดงผลการรันโค้ดในแต่ละข้อ (ข้อละ 1 Code Section พร้อมเขียนหมายเลขข้อกำกับในคอมเมนต์ก่อนเริ่ม)
2. เขียนคอมเมนต์อธิบายการทำงานของโค้ดในแต่ละ Section อย่างละเอียด
3. ส่งงานด้วยไฟล์ .ipynb โดยเขียนชื่อ-นามสกุลใน Section แรกของไฟล์ และตั้งชื่อไฟล์ด้วยรหัสนักศึกษา ตัวอย่างเช่น 6810110xxx.ipynb
4. ส่งงานผ่าน LMS ภายในเวลาที่กำหนด (จะประกาศวัน-เวลาดำหนดส่งในคาบเรียน)