

ศุภชัย สมพานิช
ผู้เขียนหนังสือ Bestseller ด้าน Programming



PROVISION

เรียนรู้ Data Science ด้วย ภาษา R

หนึ่งในภาษายอดนิยมที่ใช้ตั้งแต่วิเคราะห์ข้อมูล (Data Analysis) ไปจนถึง Machine Learning



DOWNLOAD
provision.co.th/R

ศุภชัย สมพานิช

ผู้เขียนหนังสือ Bestseller ด้าน Programming

เรียนรู้ **Data Science** ด้วย
ภาษา R



เรียนรู้ Data Science ด้วย

ภาษา R

ผู้เขียน : ศุภชัย สมพานิช
ออกแบบปก : ธนกฤต เกียรติศักดิ์ากุล
จัดรูปเล่ม : อภิญญา สุทธิประภา

ISBN : 978-616-204-783-1

978-616-204-794-7 (E-book)

ราคา : 350 บาท

ข้อมูลบรรณานุกรมของหอสมุดแห่งชาติ

ศุภชัย สมพานิช

เรียนรู้ Data Science ด้วยภาษา R.

--กรุงเทพฯ : โปรวิชั่น, 2564. 320 หน้า.

1. โปรแกรมประยุกต์บนเว็บ. I. ชื่อเรื่อง.

006.78

ชื่อและเครื่องหมายการค้าอื่น ๆ ที่อ้างอิงในหนังสือฉบับนี้ และในเนื้อหาหรือสื่อบันทึกข้อมูลอื่น ๆ ที่เกี่ยวข้อง (ถ้ามี) เป็นสิทธิโดยชอบด้วยกฎหมายของเจ้าของแต่ละราย บริษัท โปรวิชั่น จำกัด มิได้อ้างความเป็นเจ้าของ ตัวแทน หรือมีส่วนเกี่ยวข้องแต่อย่างใด นอกจากนี้ทางบริษัทได้ใช้ความพยายามอย่างเต็มที่ในการตรวจสอบความถูกต้องของเนื้อหาเท่าที่จะทำได้ แต่มิได้ยอมรับหรือรับประกันความถูกต้องหรือทันสมัยของข้อมูลต่างๆ ในทุกกรณี ผู้อ่านพึงใช้วิจารณญาณของตนเอง หากพบข้อผิดพลาดหรือสงสัยใดๆ โปรดแจ้งที่อีเมล editor@provision.co.th เพื่อที่ทางบริษัทจะได้ตรวจสอบและแก้ไขตามความเหมาะสมในการพิมพ์ครั้งต่อไป รวมทั้งประกาศให้ทราบผ่านทางเว็บไซต์ www.provision.co.th/update ด้วย

สงวนสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 โดย บริษัท โปรวิชั่น จำกัด ห้ามนำส่วนใดส่วนหนึ่งของหนังสือเล่มนี้ไปทำซ้ำ ดัดแปลงหรือเผยแพร่ต่อสาธารณชนไม่ว่ารูปแบบใดๆ นอกจากจะได้รับอนุญาตเป็นลายลักษณ์อักษรล่วงหน้าจากทางบริษัทเท่านั้น ชื่อผลิตภัณฑ์และเครื่องหมายการค้าต่างๆที่อ้างถึงเป็นสิทธิโดยชอบด้วยกฎหมายของบริษัทนั้นๆ

PROVISION

จัดพิมพ์โดย : บริษัท โปรวิชั่น จำกัด

11/8 ซอยแจ้งวัฒนะ 14

แขวงทุ่งสองห้อง เขตหลักสี่ กทม. 10210

โทร: 0-2077-4058

ผู้ก่อตั้ง/ประธานกรรมการ
กรรมการบริหาร : วศิน เพิ่มทรัพย์
วงศ์ประชา จันทร์สมวงศ์
อนิรุทธ์ รัชตะวราห์
บรรณาธิการบริหาร : บุญญาดา ช้อนขุนทด
ประสานงานกอง บก. : นนทพร ตันดีเยี่ยมสกุล

ติดต่อสั่งซื้อหนังสือ

✉ sales@provision.co.th

โทร. 0-2077-4058 หรือ 08-1928-5299

💻 www.dplusshop.com f dplusshop

ติดต่อโฆษณาและการตลาด

✉ marketing@provision.co.th

สอบถามปัญหา/ติดต่อกอง บก.

✉ editor@provision.co.th

ติดตามข่าวสาร

💻 www.provision.co.th, www.dplusguide.com

f provision1991, dplusguide

🐦 @Provision1991, @dplusguide

📺 provision1991, dplusguide

🗣️ @dplusguide

จัดจำหน่ายโดย :

บริษัท ซีเอ็ดดูเคชั่น จำกัด (มหาชน)

1858/87-90 ถ.เทพรัตน์

แขวงบางนาใต้ เขตบางนา กรุงเทพฯ 10260

โทร. 0-2826-8000 FAX: 0-2826-8999

ที่ปรึกษากฎหมาย : คุณไพบุลย์ อมรภิญโญเกียรติ

สำนักงานกฎหมาย P&P LAW FIRM โทร. 0-2651-2121

พิมพ์ที่ บริษัท พิมพ์ดี จำกัด

นายเสริม พูนพนิช ผู้พิมพ์ผู้โฆษณา พ.ศ. 2564

INTRODUCTION

ในยุคปัจจุบัน ข้อมูล ข่าวสารกลายเป็นสิ่งที่มีค่ามีความหมายมากมาย จึงถือเป็นอีก 1 แขนงที่ในทุกๆ สายอาชีพต้องมีสภินี้ติดตัวไว้ เนื้อหาในหนังสือเล่มนี้ เป็นการนำเสนอการใช้งานภาษา R เป็นภาษาที่ถูกสร้างขึ้นมาเพื่อทำงานกับข้อมูลโดยเฉพาะ

องค์ประกอบหนึ่งของภาษา R นั่นคือ การนำ Package ต่างๆ ที่มีความสำคัญในระดับต้นๆ เข้ามาช่วยการทำงานกับข้อมูลให้มีประสิทธิภาพดียิ่งขึ้น

หากคุณผู้อ่านท่านใดต้องการแนะนำ, ข้อติติง, เสนอแนะ รวมไปถึงข้อผิดพลาดต่างๆ สามารถแจ้งผู้เขียนได้ 4 ช่องทาง คือ

1. แฟนเพจ Thaivb.NET ที่ <https://www.facebook.com/thaivb.net/>
2. กลุ่มชื่อว่า Thaivb.NET ที่ <https://www.facebook.com/groups/Thaivb.NET>
3. ช่อง YouTube ชื่อว่า Thaivb.NET
4. แฟนเพจ Thaivb.NET ที่ <https://www.blockdit.com/thaivb.net>

ศุภชัย สมพานิช

มกราคม 2564



DOWNLOAD
provision.co.th/R



Contents

CHAPTER 01 เตรียมความพร้อมก่อนใช้งานภาษา R

การดาวน์โหลดและติดตั้ง R SDK และโปรแกรม R Studio.....	12
การปรับฟอนต์และสีธีมของโปรแกรม R Studio.....	16
ทำความเข้าใจกับสภาพแวดล้อมของโปรแกรม R Studio.....	17
การกำหนดให้ R Studio สามารถแสดงผลภาษาไทย.....	18
การเขียนโปรแกรมภาษา R แบบออนไลน์ด้วย Microsoft Azure Notebooks.....	18
พื้นฐานการใช้งานโปรแกรม Excel.....	23

CHAPTER 02 พื้นฐานการเขียนโปรแกรมด้วยภาษา R

การสร้างตัวแปร (Variable) ขึ้นมาทำหน้าที่เก็บข้อมูล.....	27
การลบบรรายการตัวแปรที่ค้างอยู่ในระบบ.....	30
ชนิดข้อมูลพื้นฐาน (Basic Data Types).....	30
ชนิดข้อมูลเลขจำนวนเต็ม integer และเลขทศนิยม numeric.....	31
ชนิดข้อมูลประเภทตัวอักษร หรือข้อความ character.....	32

ชนิดข้อมูลทางตรรกะ logical.....	33
ชนิดข้อมูลจำนวนเชิงซ้อน complex.....	33
ชนิดข้อมูลวันที่ Date.....	34
ชนิดข้อมูลสำหรับจัดกลุ่ม factor.....	35
การทำหมายเหตุ (Comment).....	36
การจัดรูปแบบให้กับข้อมูลประเภทตัวเลข ด้วยฟังก์ชัน sprintf().....	36
พื้นฐานการคำนวณด้านคณิตศาสตร์.....	37
ฟังก์ชันสำหรับตรวจสอบชนิดข้อมูล.....	40
การตรวจสอบว่าเป็นข้อความ character หรือไม่.....	40
การตรวจสอบว่าเป็นตัวเลข numeric หรือไม่.....	42
การตรวจสอบว่าเป็นข้อมูลทางตรรกะ logical หรือไม่.....	43
การตรวจสอบเงื่อนไขด้วยคำสั่ง if...else.....	44
การสั่งให้ทำงานซ้ำด้วยการวนลูปแบบ for.....	47
การวนลูปแบบ while.....	50
การออกจากลูปทันทีด้วยคำสั่ง break.....	51
การเว้นวรรคหรือข้ามการทำงานในลูปด้วยคำสั่ง next.....	52
การวนลูปแบบ repeat.....	53



CHAPTER
03 † ทำความรู้จักกับโครงสร้าง
ข้อมูลเวกเตอร์ (Vector)

โครงสร้างข้อมูลแบบ Vector	56
การอ่านข้อมูลจากเวกเตอร์.....	61
การวนลูปอ่านค่าในเวกเตอร์.....	68
การทำงานกับข้อมูลเวกเตอร์หลายชุด	70
การคำนวณยอดรวมทั้งหมด ด้วยฟังก์ชัน sum()	72
การคำนวณสัดส่วนเปอร์เซ็นต์ ของยอดขาย	73
การอ่านข้อมูลจากเวกเตอร์หลายชุด แบบมีเงื่อนไข.....	76
การนำเวกเตอร์มารวมกัน	78
กรณีเวกเตอร์เลขจำนวนเต็ม กับเลขจำนวนเต็ม.....	78
กรณีเวกเตอร์เลขจำนวนเต็ม กับเลขทศนิยม	79
กรณีเวกเตอร์ข้อความกับตัวเลข	80
การรวมเวกเตอร์ด้วยฟังก์ชัน paste().....	82
การเปรียบเทียบข้อมูลในเวกเตอร์ เท่ากันหรือไม่ด้วยเครื่องหมาย ==	83
การบวกกันของเวกเตอร์ตัวเลข.....	84
พื้นฐานการหาค่าสูงสุด, ต่ำสุด, ผลรวม และค่าเฉลี่ยจากเวกเตอร์ตัวเลข	86

CHAPTER
04 † กรอบข้อมูล
(Data Frame)

การสร้างกรอบข้อมูลจากเวกเตอร์.....	90
การอ่านข้อมูลจาก Data Frame	92
พื้นฐานการอ่านข้อมูลใน Data Frame แบบมีเงื่อนไข.....	94
การอ่านข้อมูลจาก Data Frame แบบมีเงื่อนไขด้วยฟังก์ชัน subset()	96
พื้นฐานการแก้ไขข้อมูลใน Data Frame.....	98
การเพิ่มหรือลบคอลัมน์ใน Data Frame..	100
การเพิ่มหรือลบแถวใน Data Frame	104
การทำงานกับแถวข้อมูล ที่มีค่าผิดพลาด NA	105
การสลับตำแหน่งคอลัมน์.....	106
การเปลี่ยนชื่อคอลัมน์ใน Data Frame ด้วยฟังก์ชัน names().....	107
การทำงานกับ Data Frame แบบถาวรด้วยฟังก์ชัน attach().....	108
การเรียงลำดับข้อมูลใน Data Frame ด้วยฟังก์ชัน order()	109
การทำงานหลายอย่างกับ Data Frame ด้วยฟังก์ชัน with()	111

Contents

CHAPTER 05 การสร้างฟังก์ชัน (Function) ขึ้นมาใช้งาน

พื้นฐานการสร้างฟังก์ชัน (Function) ใน R.....	114
การสร้างฟังก์ชันที่มีพารามิเตอร์หลายตัว.....	116
การสร้างฟังก์ชันรับค่าจากผู้ใช้งานโดยตรง.....	118
การสร้างฟังก์ชันสำหรับคำนวณพื้นที่.....	119
การสุ่มข้อมูลด้วยฟังก์ชัน sample().....	120
การสุ่มข้อมูลแบบถ่วงน้ำหนัก.....	124

CHAPTER 06 การ Import และ Export ข้อมูลจากไฟล์ Excel, CSV และไฟล์ข้อความ

การ Import ข้อมูลจากไฟล์ข้อความ (Text File).....	127
การ Import ข้อมูลจากไฟล์ CSV.....	131
การ Import ข้อมูลจากไฟล์ Excel.....	133
การ Export ข้อมูลจาก Data Frame เป็นไฟล์ CSV.....	135

CHAPTER 07 พื้นฐานการใช้ R ร่วมกับข้อมูลในระบบฐานข้อมูล SQL Server

การดาวน์โหลดและติดตั้งฐานข้อมูล SQL Server.....	141
การติดตั้งตัวจัดการฐานข้อมูล Microsoft SQL Server Management Studio (SSMS).....	146
การเชื่อมต่อฐานข้อมูล SQL Server ด้วย SSMS.....	148
การกำหนดให้ SSMS สามารถแก้ไขโครงสร้างฐานข้อมูลได้.....	148
การสร้างฐานข้อมูลใหม่ด้วย SSMS.....	149
การเพิ่ม/แก้ไขข้อมูลในตาราง.....	152
การดูข้อมูลที่อยู่ในตาราง.....	153
การดูโครงสร้างตาราง.....	153
การกำหนดให้ฐานข้อมูล SQL Server เข้ารหัสภาษาไทย.....	154
การดาวน์โหลดและติดตั้ง ODBC Driver สำหรับฐานข้อมูล SQL Server.....	155
การดาวน์โหลดและติดตั้ง RODBC สำหรับภาษา R.....	156
การเชื่อมต่อเข้ากับฐานข้อมูล SQL Server ด้วย R.....	157



CHAPTER
08 ▶ การทำงานกับข้อมูลใน
ฐานข้อมูล SQL Server

พื้นฐานการคิวรีข้อมูลจาก
ฐานข้อมูล SQL Server แบบ 1 ตาราง..... 159

การทำงานกับข้อมูลภาษาไทย
จากฐานข้อมูล SQL Server
ด้วยวิธี Export ข้อมูลเป็น Excel..... 161

การทำงานกับข้อมูลตั้งแต่ 2 ตารางขึ้นไป..... 166

 การสร้าง View ในฐานข้อมูล SQL Server
 สำหรับข้อมูลตั้งแต่ 2 ตารางขึ้นไป..... 169

CHAPTER
09 ▶ ภาษา R กับข้อมูลใน
ฐานข้อมูล MongoDB

การดาวน์โหลดและติดตั้ง
ฐานข้อมูล MongoDB 177

การเชื่อมต่อกับฐานข้อมูล MongoDB..... 178

การสร้างฐานข้อมูลและคอลเล็กชัน
จัดเก็บข้อมูล 179

 การปรับมุมมองข้อมูล..... 182

การดาวน์โหลดและติดตั้ง mongolite 183

การเพิ่มข้อมูลในฐานข้อมูล MongoDB
ด้วยภาษา R..... 184

การแสดงข้อมูลจากฐานข้อมูล MongoDB
ด้วยฟังก์ชัน View() 185

การเพิ่มข้อมูลที่มีโครงสร้างแตกต่างกัน 186

การแก้ไขข้อมูลที่มีโครงสร้างเหมือนกัน 187

การแก้ไขข้อมูลที่มีโครงสร้างแตกต่างกัน 189

การลบข้อมูล..... 190

CHAPTER
10 ▶ สถิติเบื้องต้นกับภาษา R

ทำความเข้าใจกับ “ประชากร” (Population)
และ “ตัวอย่าง” (Sample)..... 192

ค่ากลางของข้อมูล..... 193

ค่าเฉลี่ย (Mean)..... 193

 ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)
 แบบไม่แจกแจงความถี่..... 193

 การหาค่าเฉลี่ยเลขคณิต
 กรณีมีข้อมูลผิดพลาด (NA)..... 195

 ค่าเฉลี่ยเลขคณิต
 กับค่าผิดปกติ (Outliers)..... 197

 ค่าเฉลี่ยถ่วงน้ำหนัก (Weighted Mean)
 แบบไม่แจกแจงความถี่..... 199

 ค่าเฉลี่ยเรขาคณิต (Geometric Mean)
 แบบไม่แจกแจงความถี่..... 201

Contents

การคำนวณค่าเฉลี่ยเรขาคณิต โดยอาศัย psych	206
ค่าเฉลี่ยฮาร์โมนิก (Harmonic Mean) แบบไม่แจกแจงความถี่ของข้อมูล	208
มัธยฐาน (Median)	210
การหาค่ากลางมัธยฐาน (Median) แบบไม่แจกแจงความถี่ของข้อมูล	210
การหาค่าฐานนิยม (Mode) แบบไม่แจกแจงความถี่.....	213

CHAPTER 11 การหาค่าแห่งของข้อมูล (Measures Of Variability)

ควอร์ไทล์ (Quartile), เดไซล์ (Decile) และ เปอร์เซ็นไทล์ (Percentile)	217
ควอร์ไทล์ (Quartile).....	217
เดไซล์ (Decile).....	218
เปอร์เซ็นไทล์ (Percentile)	218
วิธีการหาควอร์ไทล์, เดไซล์ และเปอร์เซ็นไทล์แบบไม่แจกแจงความถี่.....	219
การคำนวณควอร์ไทล์, เดไซล์ และ เปอร์เซ็นไทล์ด้วยฟังก์ชัน quantile().....	223

CHAPTER 12 การวัดกระจายของข้อมูล (Measure of Dispersion)

ปัญหาของค่ากลางของข้อมูล.....	226
การวัดการกระจายสัมบูรณ์ (Absolute Variation)	228
พิสัย (Range) แบบไม่แจกแจงความถี่.....	228
ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) แบบไม่แจกแจงความถี่.....	229
ส่วนเบี่ยงเบนเฉลี่ย (Mean Deviation) แบบไม่แจกแจงความถี่.....	232
ส่วนเบี่ยงเบนควอร์ไทล์ (Quartile Deviation) แบบไม่แจกแจงความถี่.....	235
การวัดการกระจายสัมพัทธ์ (Relative Variation)	237
สัมประสิทธิ์ของพิสัย แบบไม่แจกแจงความถี่.....	237
สัมประสิทธิ์ของการแปรผัน แบบไม่แจกแจงความถี่.....	239
สัมประสิทธิ์ของส่วนเบี่ยงเบนเฉลี่ย แบบไม่แจกแจงความถี่.....	240
สัมประสิทธิ์ของส่วนเบี่ยงเบนควอร์ไทล์ แบบไม่แจกแจงความถี่.....	242
การแปลความหมายของ การวัดการกระจายสัมพัทธ์	244



CHAPTER

13 ▶ เมทริกซ์ (Matrix)

โครงสร้างข้อมูลแบบเมทริกซ์ (Matrix).....	248
การแปลงเวกเตอร์เป็นเมทริกซ์ ด้วยฟังก์ชัน dim().....	249
พื้นฐานการจัดการข้อมูลในเมทริกซ์ ด้วยฟังก์ชัน sweep().....	251
การทำงานกับข้อมูลในเมทริกซ์ แนวแถวหรือคอลัมน์.....	252
การหาผลรวมคอลัมน์ ด้วยฟังก์ชัน colSums().....	255
การหาผลรวมแถว ด้วยฟังก์ชัน rowSums().....	256
การหาค่าเฉลี่ยเลขคณิตตาม แนวคอลัมน์และแถวด้วยฟังก์ชัน colMeans() กับ rowMeans().....	257
การเพิ่มคอลัมน์ใหม่ในเมทริกซ์ ด้วยฟังก์ชัน cbind().....	258
การเพิ่มแถวใหม่ในเมทริกซ์ ด้วยฟังก์ชัน rbind().....	260
การหาค่ามัธยฐาน (Median) ในเมทริกซ์.....	261

CHAPTER

14 ▶ การทำงานกับข้อมูลที่น่าสนใจ

การหาค่าเฉลี่ยเลขคณิตแบบจัดกลุ่ม.....	264
พื้นฐานการหาค่าเฉลี่ยถ่วงน้ำหนัก กับข้อมูลด้วยฟังก์ชัน weighted.mean().....	266
การหาค่าเฉลี่ยถ่วงน้ำหนักของข้อมูล ที่ถูกจัดกลุ่ม.....	267
การตรวจสอบค่าในช่วงที่ระบุ ด้วยฟังก์ชัน between() ของ dplyr.....	268
การคำนวณค่าทางสถิติตามแถวหรือคอลัมน์ ด้วยฟังก์ชัน apply().....	270
การคำนวณค่าทางสถิติเบื้องต้น แบบแบ่งกลุ่มด้วยฟังก์ชัน tapply().....	272
การใส่หมายเลขบรรทัด ด้วยฟังก์ชัน row_number() ของ dplyr.....	273
การจัดลำดับแบบแบ่งกลุ่มข้อมูล ด้วยฟังก์ชัน ntile() ของ dplyr.....	274
การเรียงลำดับข้อมูลแบบ Percent Rank ด้วยฟังก์ชัน percent_rank() ของ dplyr.....	275
การคำนวณค่าสถิติเบื้องต้น ด้วยฟังก์ชัน summary().....	277
การคำนวณค่าสถิติเบื้องต้น ด้วยฟังก์ชัน summarise() ของ dplyr.....	279

Contents

CHAPTER 15 ▸ พื้นฐานการนำเสนอข้อมูลด้วยกราฟ (Data Visualization)

การสร้างกราฟแท่ง (Bar Chart) ด้วยฟังก์ชัน <code>barplot()</code>	284
การ Export กราฟ.....	285
การใส่ป้าย Label ให้กับแกน x.....	286
การสร้างกราฟแท่งจาก Data Frame.....	287
การปรับแต่งกราฟแท่ง	288
การแสดงผลข้อมูลแบบแบ่งกลุ่มด้วยกราฟแท่ง	291
การสร้างกราฟแท่งจากข้อมูลประเภทความถี่ ...	294

CHAPTER 16 ▸ การสร้างกราฟแบบ Box Plot

การใช้งานกราฟ Box Plot แสดงการกระจายของข้อมูลด้วยฟังก์ชัน <code>boxplot()</code>	298
การกำหนดข้อความกำกับ Box Plot	304
การเปลี่ยนสี Box Plot.....	305

CHAPTER 17 ▸ การสร้างกราฟแบบ Heat Map และกราฟเส้น

การดาวน์โหลดและติดตั้ง <code>ggplot2</code> และ <code>tidyr</code>	307
การสร้างกราฟแบบ Heat Map	307
การปรับแต่งโทนสีใน Heat Map.....	312
การสร้างกราฟเส้น (Line Chart)	313
การแสดงผลค่าข้อมูลใน Heat Map.....	315

CHAPTER 18 ▸ การจัดการข้อมูลสูญหาย (NA)

พื้นฐานการจัดการค่า NA ด้วยฟังก์ชัน <code>coalesce()</code> ของ <code>dplyr</code>	318
การแก้ไขค่า NA ด้วยเวกเตอร์ร่วมกับฟังก์ชัน <code>coalesce()</code>	319
การตรวจสอบค่า NA ด้วยฟังก์ชัน <code>complete.cases()</code>	320

CHAPTER

01



เตรียมความพร้อม ก่อนใช้งานภาษา R

ในปัจจุบัน “ข้อมูล (Data)” กลายเป็นสิ่งที่มีค่าเป็นอย่างมาก เรารับข้อมูลหลายรูปแบบจากหลายทิศทาง ในช่วงเวลาต่อเนื่องกัน ใครก็ตามที่สามารถใช้ประโยชน์จากข้อมูลที่มีอยู่ได้มากที่สุด ย่อมได้เปรียบมากกว่าคนที่มีข้อมูลน้อยกว่า เพราะความที่ข้อมูลที่เราสนใจมีหลายรูปแบบ มาจากหลายแหล่งหลายช่วงเวลาที่แตกต่างกัน จึงต้องมีกระบวนการหรือวิธีการ “ย่อย” ข้อมูลเหล่านั้นออกมาก่อน กระบวนการย่อยข้อมูลที่กำลังกล่าวถึง นี้คือสิ่งที่เราต้องเรียนรู้ ผลลัพธ์ของการย่อยข้อมูลมีหลายรูปแบบ ก็จะได้ข้อมูลในอีกรูปแบบหนึ่งที่แตกต่างกันไปจากข้อมูลต้นทางเดิมหรือข้อมูลดิบ

วิธีการย่อยข้อมูลมีหลายวิธี หนังสือเล่มนี้นำเสนอพื้นฐานการทำงานกับข้อมูลและการวิเคราะห์ข้อมูลเบื้องต้นด้วยภาษา R และมีการใช้งานโปรแกรม Excel เข้ามาช่วยในการทำงานด้วยบ้างเล็กน้อย

ภาษา R มีหน้าที่เหมือนกับโปรแกรม Excel มีข้อแตกต่างที่ชัดเจนที่สุดคือ โปรแกรม Excel มีส่วนแสดงผล มีเครื่องมือ มีฟังก์ชันต่างๆ ช่วยให้เราสามารถทำงานกับข้อมูลในด้านต่างๆ ส่วนภาษา R ทำงานกับข้อมูลในรูปแบบของการเขียนโปรแกรม

รูปแบบการนำเสนอในหนังสือเล่มนี้

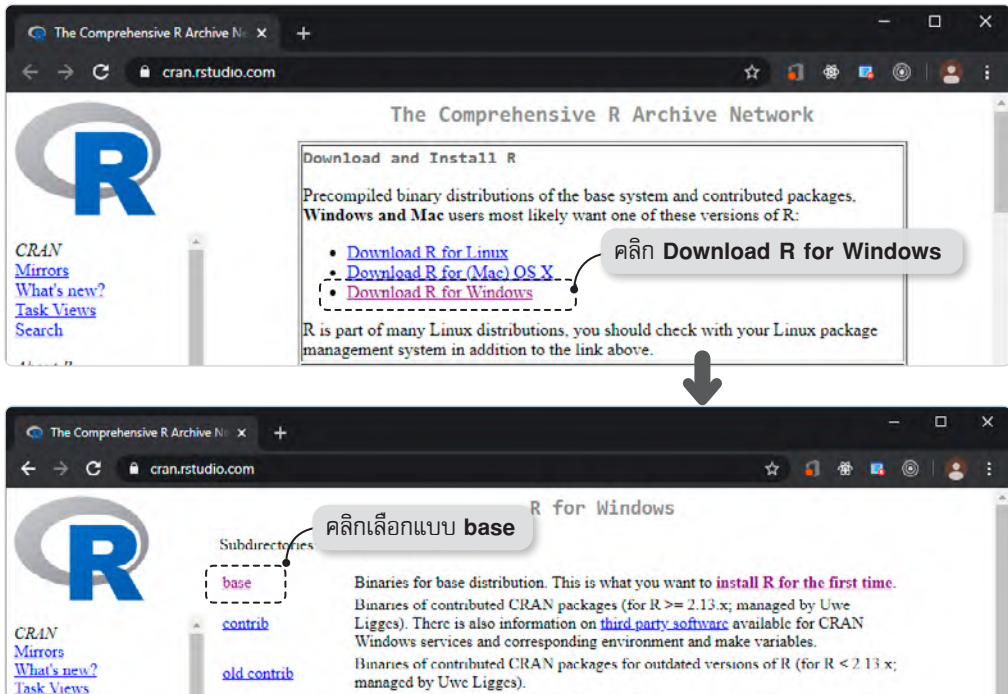
เนื้อหาหนังสือเล่มนี้นำเสนอวิธีการทำงานและวิเคราะห์ข้อมูลในขั้นต้นด้วยภาษา R และมีการใช้งานโปรแกรม Excel เข้ามาช่วยเสริมการนำเสนอเนื้อหา เพื่อให้คุณผู้อ่านสามารถทำความเข้าใจเนื้อหาได้ง่ายขึ้น

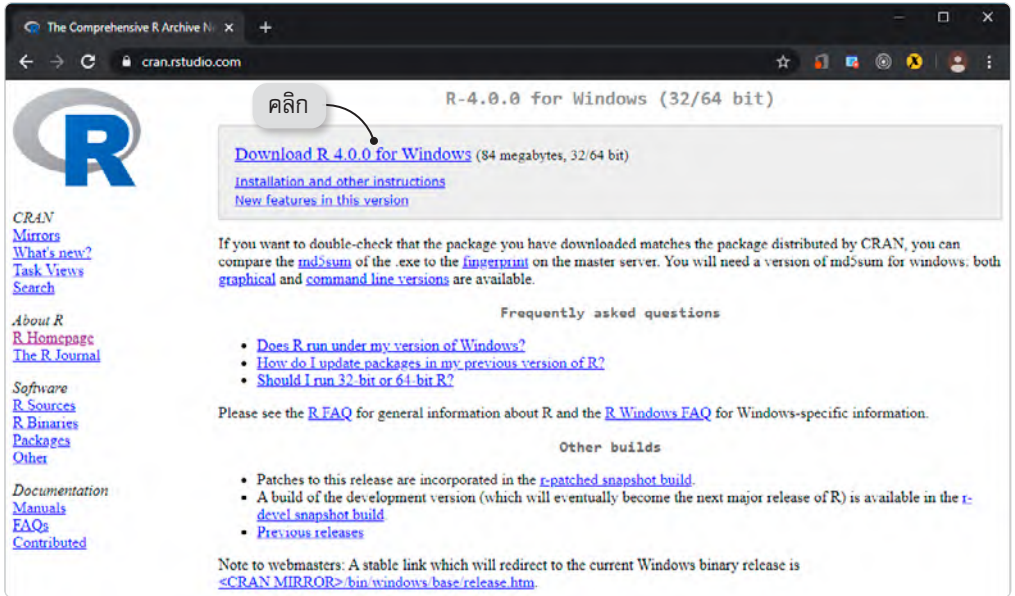
คุณผู้อ่านที่ไม่เคยใช้โปรแกรม Excel มาก่อน ไม่มีปัญหาแต่อย่างใด เพราะว่าจะเป็นการศึกษาการใช้งานโปรแกรม Excel ไปด้วย มีคำอธิบายประกอบทุกหัวข้อ ส่งผลให้คุณผู้อ่านเรียนรู้ทั้งการเขียนโปรแกรมภาษา R และใช้งานโปรแกรม Excel ในเบื้องต้นเท่านั้น

การดาวน์โหลดและติดตั้ง R SDK และโปรแกรม R Studio

เมื่อคุณต้องการใช้งานภาษา R ให้คุณติดตั้ง R SDK ก่อน สามารถดาวน์โหลดได้ที่เว็บไซต์ <https://cran.rstudio.com/> มีขั้นตอนดังนี้

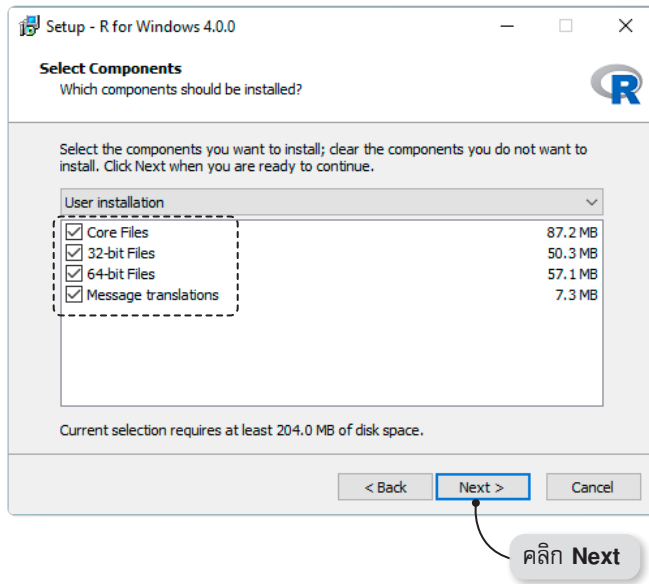
1. ให้คุณดาวน์โหลด R SDK ตามระบบปฏิบัติการที่คุณใช้งาน ในกรณีนี้ผู้เขียนใช้งาน Windows จึงเลือกดาวน์โหลดแบบ **R for Windows** แบบ **base** ดังรูปที่ 1-1

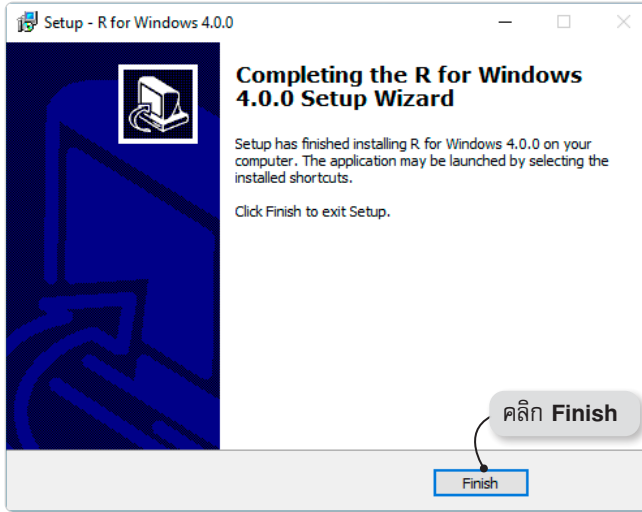




รูปที่ 1-1 แสดงการดาวน์โหลด R SDK

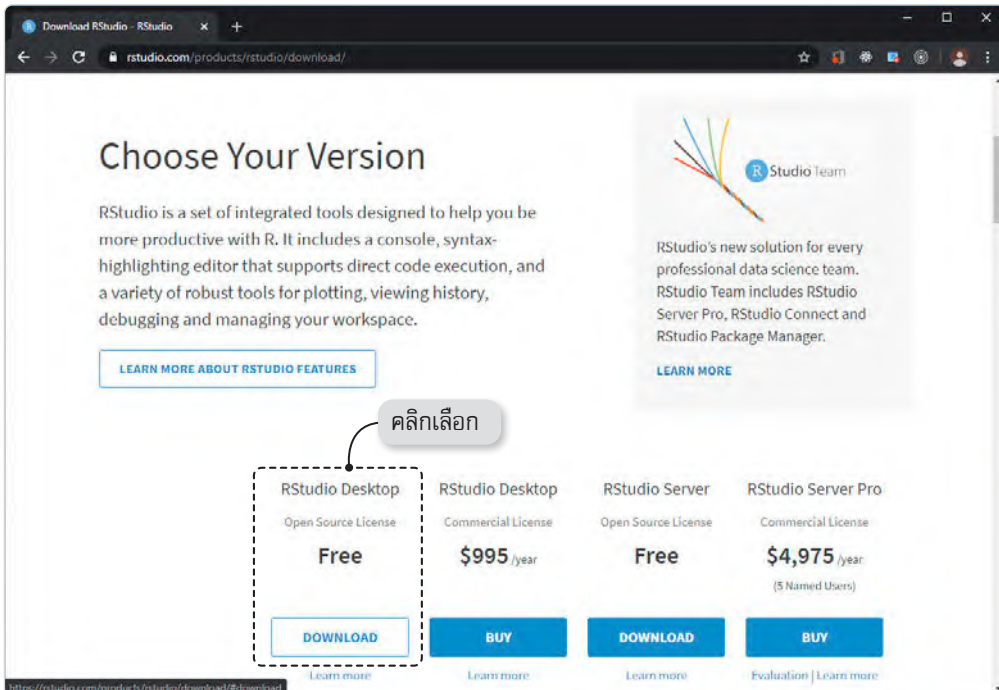
2. การติดตั้ง R SDK จะใช้ค่าเริ่มต้นที่มากับตัวติดตั้งทั้งหมด ดังรูปที่ 1-2

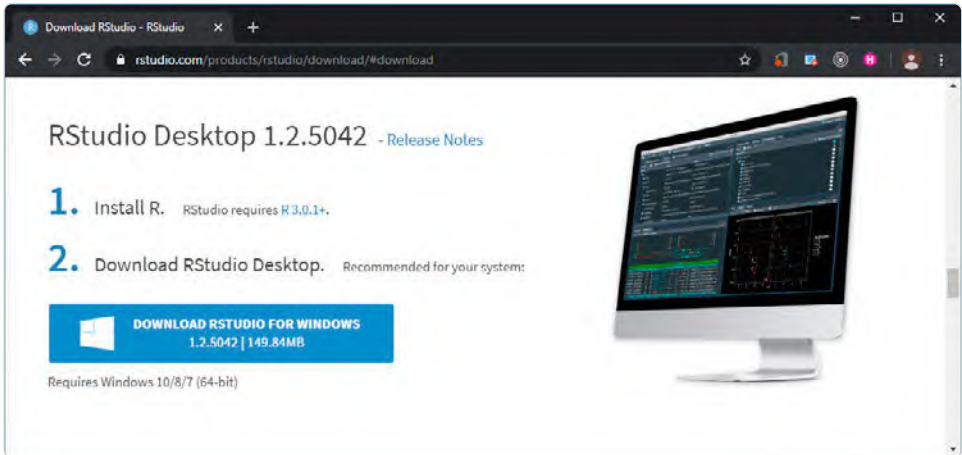




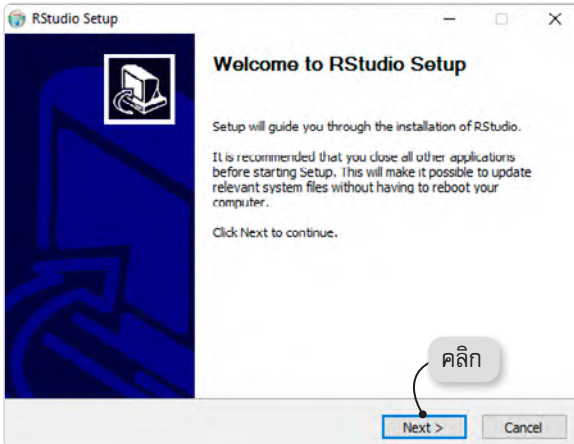
รูปที่ 1-2 แสดงการติดตั้ง R SDK เสร็จสมบูรณ์

3. ให้คุณดาวน์โหลดโปรแกรม R Studio มาใช้งานได้ฟรีที่เว็บไซต์ <https://rstudio.com/products/rstudio/download/> ดังรูปที่ 1-3

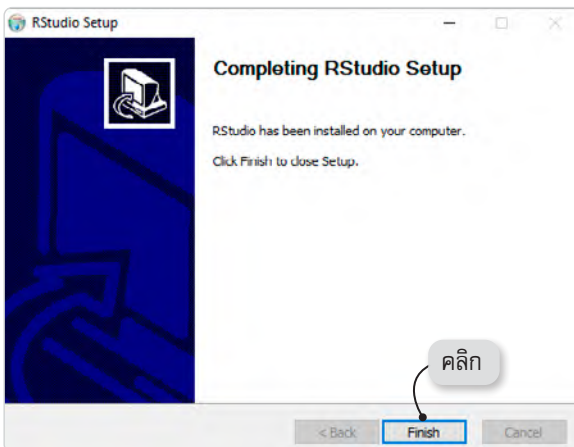




รูปที่ 1-3 แสดงการดาวน์โหลดไฟล์ติดตั้งโปรแกรม R Studio



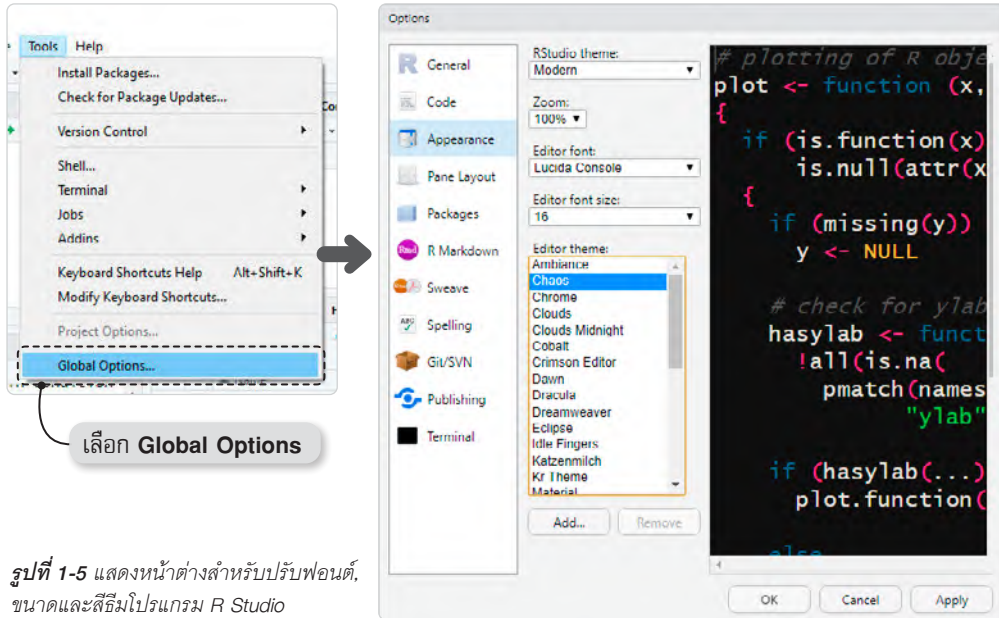
4. การติดตั้งโปรแกรม R Studio ใช้ค่าเริ่มต้นที่มากับตัวติดตั้งทั้งหมด ดังรูปที่ 1-4



รูปที่ 1-4 แสดงการติดตั้งโปรแกรม R Studio เสร็จสมบูรณ์

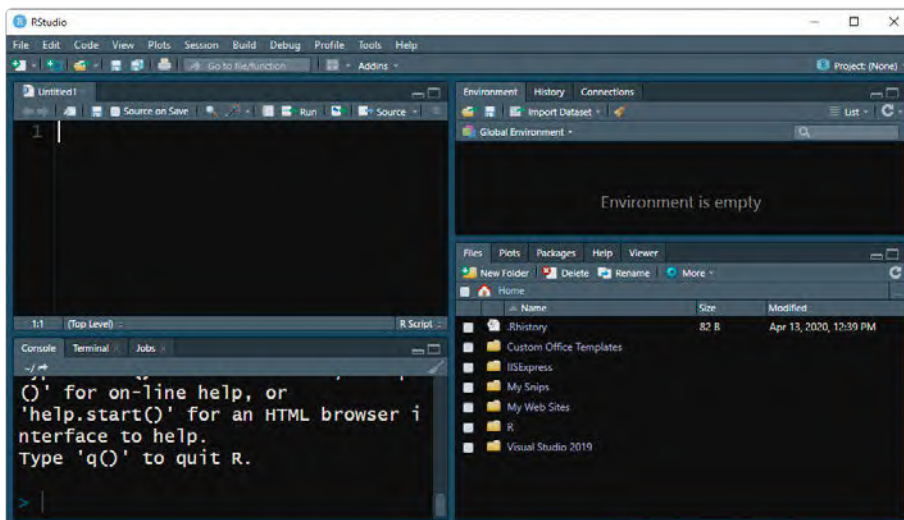
การปรับฟอนต์และสีธีมของโปรแกรม R Studio

คุณสามารถปรับแต่งฟอนต์, ขนาดและสีธีมให้กับตัวโปรแกรม R Studio โดยการเลือกเมนู **Tools ▶ Global Options...** เพื่อแสดงหน้าต่าง Options ขึ้นมา ดังรูปที่ 1-5



รูปที่ 1-5 แสดงหน้าต่างสำหรับปรับฟอนต์, ขนาดและสีธีมโปรแกรม R Studio

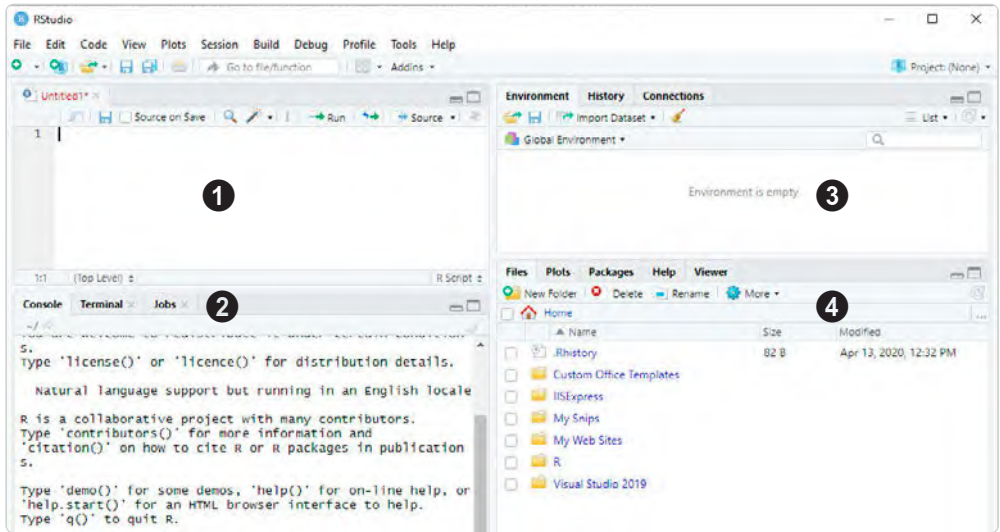
จากรูปที่ 1-5 ที่รายการ **Appearance** คุณสามารถเลือกฟอนต์, ขนาดและสีธีมตามความต้องการของคุณ ในปัจจุบัน การใช้ธีมที่มีโทนสีดำได้รับความนิยมเป็นอย่างมาก ดังรูปที่ 1-6



รูปที่ 1-6 แสดงโปรแกรม R Studio แบบใช้ธีมที่มีโทนสีดำ

ทำความรู้จักกับสภาพแวดล้อมของโปรแกรม R Studio

หลังจากที่ติดตั้งโปรแกรม R Studio + R SDK เรียบร้อยแล้ว ให้คุณเปิดโปรแกรม R Studio ขึ้นมา ให้คลิกที่ปุ่ม  เพื่อแบ่งพื้นที่เขียนโค้ดและหน้าต่าง Console ออกมา ดังรูปที่ 1-7



รูปที่ 1-7 แสดงสภาพแวดล้อมของโปรแกรม R Studio

- หมายเลข ❶ พื้นที่สำหรับเขียนโปรแกรมภาษา R เปรียบเสมือนกับเป็นแป้นพิมพ์ตัวเลขบนเครื่องคิดเลข เราจะป้อนภาษา R ตรงพื้นที่แห่งนี้
- หมายเลข ❷ แท็บ Console คือ ทำหน้าที่ทั้งเขียนโค้ดภาษา R และแสดงผลการทำงาน (Output) ของโค้ดที่เราสั่งให้รัน
- หมายเลข ❸ แท็บ Environment คือ รายการตัวแปรและค่า Output ที่ได้
- หมายเลข ❹ เป็นพื้นที่สำหรับแสดงผลหรือ Output พิเศษ เช่น ในแท็บ Plots ทำหน้าที่แสดงกราฟ เป็นต้น

การกำหนดให้ R Studio สามารถแสดงผลภาษาไทย

เพื่อให้สามารถทำงานกับข้อมูลที่มีภาษาไทยรวมอยู่ด้วย จึงต้องมีการกำหนดให้ R Studio เข้ารหัสภาษาไทยด้วย โดยการพิมพ์คำสั่งต่อไปนี้ในหน้าต่าง Console แล้วกดปุ่ม **Enter** ดังรูปที่ 1-8

R

```
Sys.setlocale("LC_CTYPE", "Thai")
```

```

> Sys.setlocale("LC_CTYPE", "Thai")
[1] "Thai_Thailand.874"
> |
  
```

Note

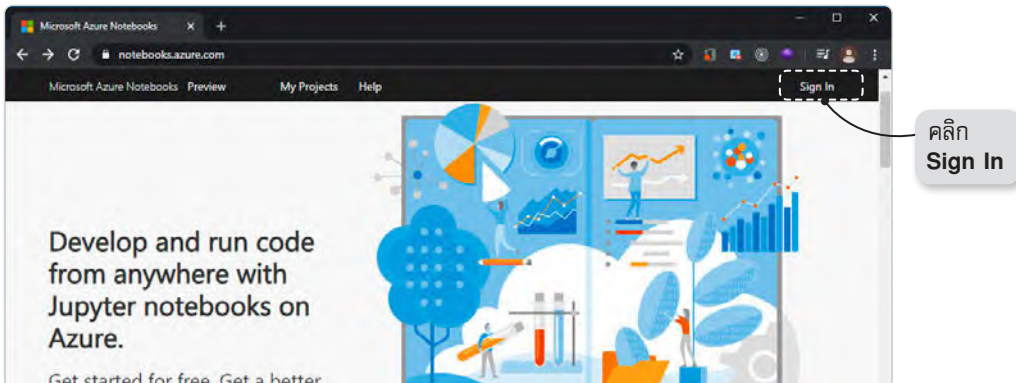
ต้องรันคำสั่งนี้เสมอ เมื่อทำงานกับข้อมูลภาษาไทย

รูปที่ 1-8 แสดงการกำหนดให้แสดงผลภาษาไทย

การเขียนโปรแกรมภาษา R แบบออนไลน์ด้วย Microsoft Azure Notebooks

การเขียนโปรแกรมภาษา R อีกวิธีหนึ่งที่ไม่จำเป็นต้องติดตั้งโปรแกรมใดๆ ในเครื่องของคุณ นั่นคือ อาศัย Microsoft Azure Notebooks เป็นบริการของไมโครซอฟท์ คุณสามารถใช้งานได้ฟรี มีขั้นตอนดังนี้

1. ไปที่เว็บไซต์ <https://notebooks.azure.com/> คุณต้องมี Microsoft Account ก่อน สามารถสมัครได้ฟรีเช่นกัน โดยการเลือกเมนู **Sign In**

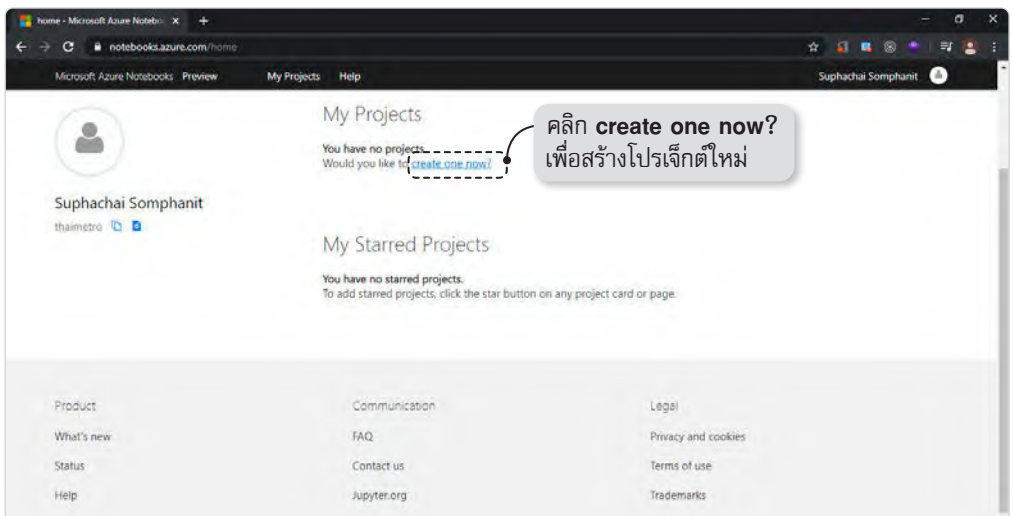


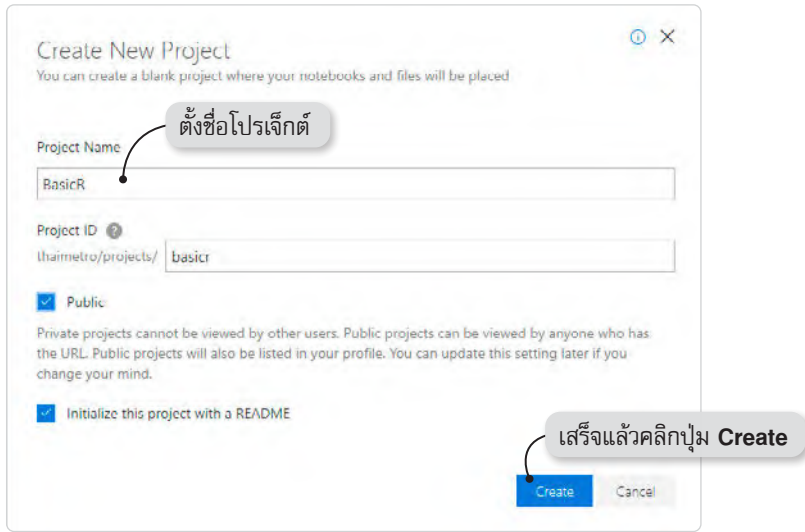


รูปที่ 1-9 แสดงการตอบรับข้อตกลงการใช้งาน

จากรูปที่ 1-9 เมื่อคุณ Login เข้ามาขอใช้บริการ Azure Notebooks ก็ให้ยอมรับสิทธิ์และข้อตกลงการใช้งานก่อน

2. การใช้งานครั้งแรกของคุณ ยังไม่เคยมีการสร้างโปรเจกต์ใดๆ ใน Azure Notebooks มาก่อน ให้คลิกที่ **create one now?** เพื่อสร้างโปรเจกต์ใหม่ ดังรูปที่ 1-10

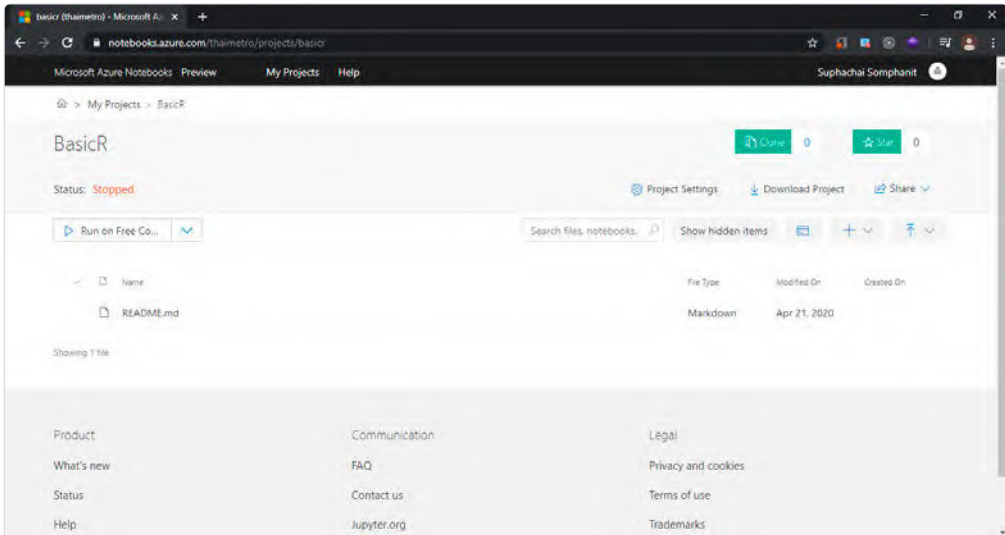




รูปที่ 1-10 แสดงการสร้างโปรเจกต์ที่ชื่อว่า BasicR

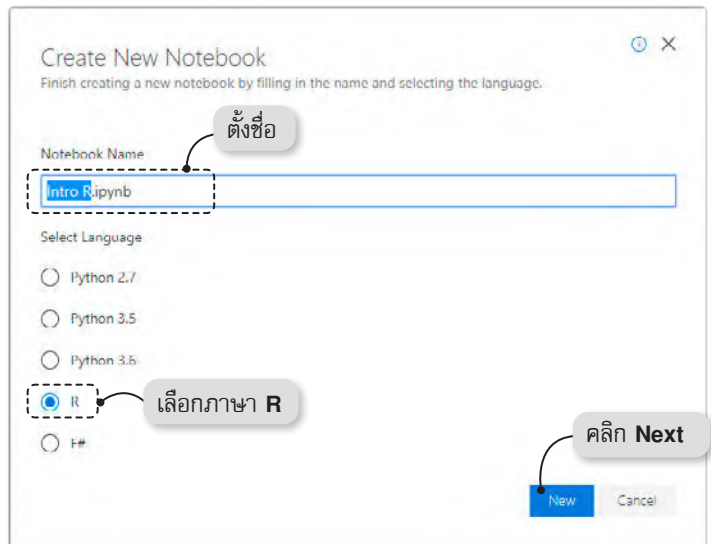
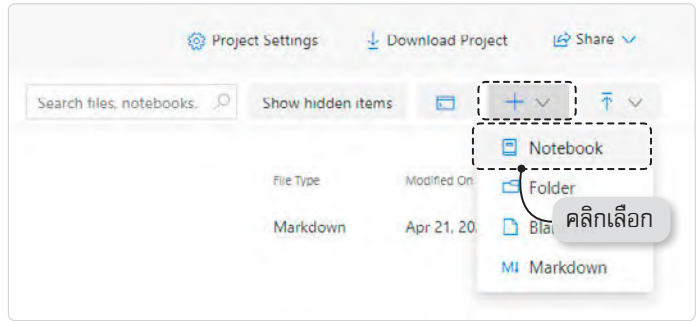
จากรูปที่ 1-10 ในกรณีนี้ ผู้เขียนสร้างโปรเจกต์ที่ชื่อว่า BasicR (หรือชื่อโปรเจกต์ตามที่คุณต้องการ) โดยการคลิกที่ปุ่ม [Create](#) เพื่อทดสอบและเขียนภาษา R

3. ภายในโปรเจกต์ BasicR ของเรายังว่างอยู่ เพราะว่า ยังไม่มีการสร้างไฟล์สำหรับเขียนโค้ดภาษา R ดังรูปที่ 1-11



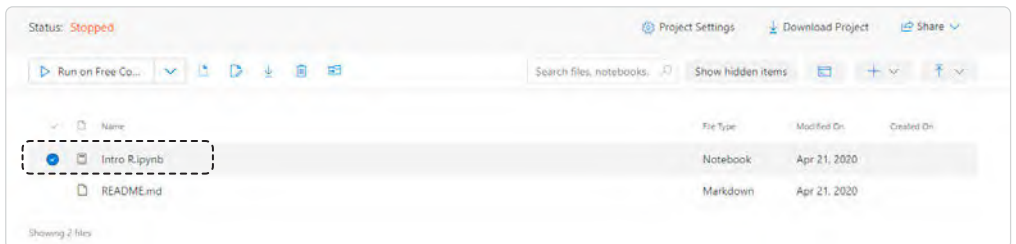
รูปที่ 1-11 แสดงโปรเจกต์ว่าง ๆ ที่ชื่อว่า BasicR

4. เป็นขั้นตอนการสร้างไฟล์ขึ้นมาในโปรเจกต์ BasicR โดยการคลิกที่ปุ่ม **+** เลือกคำสั่ง **Notebook** ในกรณีนี้ ผู้เขียนอยากให้ไฟล์นี้ทำหน้าที่ทดสอบภาษา R เบื้องต้น เลือกใช้งานภาษา R ตั้งชื่อว่า **Intro R.ipynb** จากนั้นคลิกที่ปุ่ม **New** ดังรูปที่ 1-12



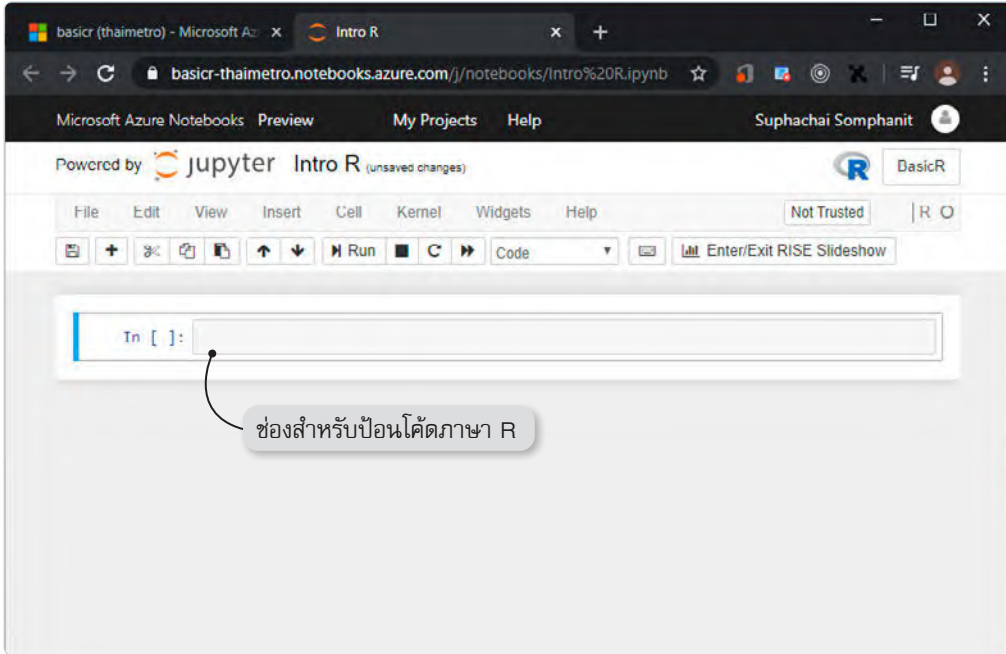
รูปที่ 1-12 แสดงการสร้างไฟล์สำหรับทดสอบภาษา R

5. คุณจะได้ไฟล์ที่ชื่อว่า Intro R.ipynb สำหรับเขียนภาษา R ดังรูปที่ 1-13



รูปที่ 1-13 แสดงไฟล์ Intro R.ipynb สำหรับเขียนโค้ดภาษา R

6. ให้คุณเข้ามาในไฟล์ Intro R.ipynb ที่นี่คือ พื้นที่สำหรับเขียนโค้ดภาษา R ดังรูปที่ 1-14



รูปที่ 1-14 แสดงตัวเขียนภาษา R

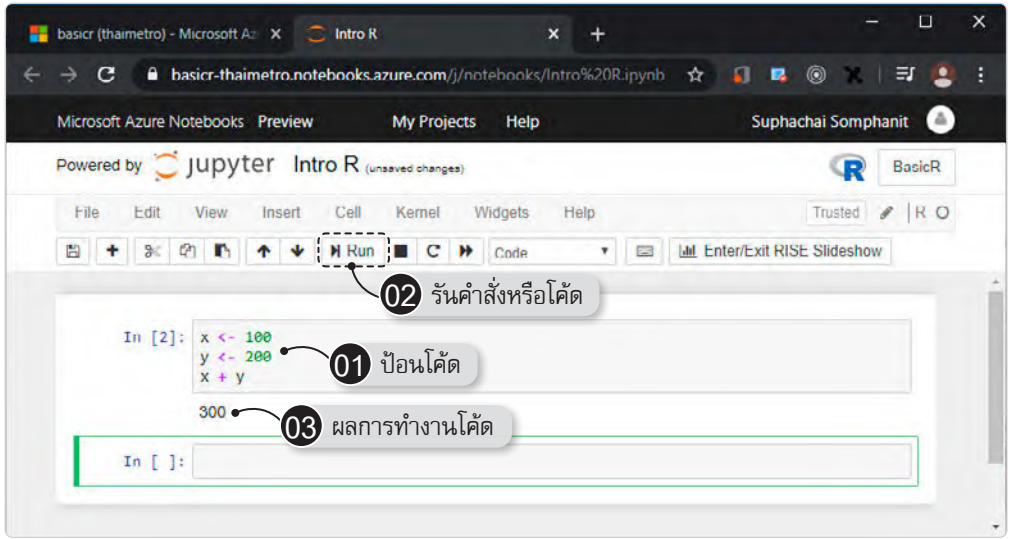
7. ที่ช่องป้อนโค้ด ให้คุณป้อนโค้ดภาษา R ต่อไปนี้เพื่อทดสอบการทำงาน กล่าวคือ

- สร้างตัวแปร x ทำหน้าที่เก็บเลขจำนวนเต็ม 100
- สร้างตัวแปร y ทำหน้าที่เก็บเลขจำนวนเต็ม 200

จากนั้นกำหนดให้ $x + y$ แล้วคลิกปุ่ม **Run** ก็จะพบกับคำตอบ 300 ถือว่า คุณสามารถใช้งานภาษา R ใน Azure Notebooks ได้แล้ว ดังรูปที่ 1-15

R

```
x <- 100
y <- 200
x + y
```

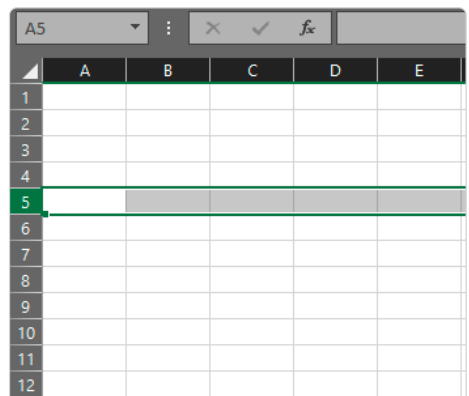
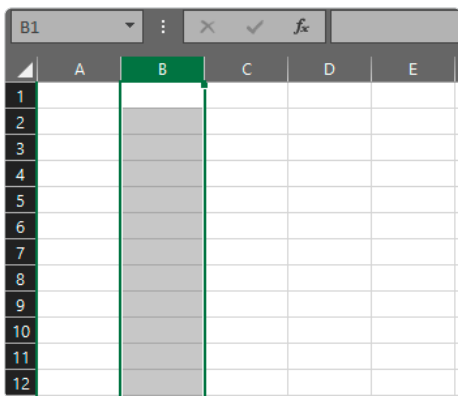


รูปที่ 1-15 แสดงผลการทำงานของภาษา R ในขั้นต้น

พื้นฐานการใช้งานโปรแกรม Excel

โปรแกรม Excel เป็นโปรแกรมที่ทำหน้าที่แสดงข้อมูลในรูปแบบตาราง และมีฟังก์ชันสำเร็จรูปด้านต่างๆ ให้คุณเรียกใช้งาน สำหรับคุณผู้อ่านที่ยังไม่เคยใช้งานมาก่อน ผู้เขียนมีข้อเสนอเบื้องต้นให้คุณทราบ เพราะว่าการนำเสนอเนื้อหาของหนังสือเล่มนี้ มีการใช้งานควบคู่กันระหว่างการเขียนโปรแกรมด้วยภาษา R และ Excel ประกอบความเข้าใจด้วย

การมองตารางในแนวดัง เรียกว่า คอลัมน์ (Column) เช่น คอลัมน์ A, B, C... ส่วนการมองข้อมูลในแนวนอน เรียกว่า แถว (Row) ดังรูปที่ 1-16



รูปที่ 1-16 แสดงแนวคอลัมน์ (Column) และแถว (Row)

ช่องสี่เหลี่ยมแต่ละช่อง เรียกว่า เซล (Cell) โดยปกติแล้วเราจะระบุตำแหน่งของแต่ละช่อง ด้วยวิธีการเรียกชื่อคอลัมน์ก่อนแล้วตามด้วยชื่อแถว เช่น เซล B2 แสดงดังรูปที่ 1-17

	A	B	C	D
1				
2				
3				
4				
5				

รูปที่ 1-17 แสดงตำแหน่งเซลล์ B2

ข้อมูลที่ปรากฏในแต่ละเซลล์ มี 2 รูปแบบคือ

1. การป้อนข้อมูลเข้าไปตรงๆ เช่น ป้อนตัวเลขจำนวนเต็ม 1, 2 และ 3 เข้าไปที่เซลล์ A1, A2 และ A3 ตามลำดับ ดังรูปที่ 1-18

	A	B	C	D
1	1			
2	2			
3	3			
4				
5				

รูปที่ 1-18 แสดงการป้อนข้อมูลตรงๆ เข้าไปในเซลล์

2. ข้อมูลที่เกิดจากการคำนวณ ไม่ได้มาจากการป้อนด้วยตัวเอง ตัวเลขที่ปรากฏในแต่ละเซลล์ อาจเกิดมาจากการคำนวณต่างๆ เช่น คุณป้อนตัวเลข 10 และ 20 ลงในเซลล์ A1 กับ B1 ตามลำดับ

ส่วนที่เซลล์ C1 เกิดจากการนำค่าของเซลล์ A1 บวกกับเซลล์ B1 โดยการพิมพ์สูตรดังต่อไปนี้ แล้วกดคีย์ **Enter** ดังรูปที่ 1-19 ก็จะได้ผลบวก 30 ถือเป็นตัวเลขที่เกิดจากการคำนวณ

เซลล์ C1

```
=A1+B1
```

	A	B	C	D	E
1	10	20			
2					
3					
4					
5					

	A	B	C	D	E
1	10	20	=A1+B1		
2					
3					
4					
5					

	A	B	C	D	E
1	10	20	30		
2					
3					
4					
5					

Note
ใช้สูตรหรือฟังก์ชันต่างๆ
ใน Excel ต้องเริ่มต้น
ด้วยเครื่องหมาย =

รูปที่ 1-19 แสดงตัวเลขที่เกิดจากการคำนวณในเซลล์ C1

ข้อมูลที่เกิดจากการคำนวณ จะมีความสัมพันธ์กันตามสูตรที่ใส่ไว้ เช่น ผู้เขียนแก้ไขตัวเลขในเซลล์ B1 ให้กลายเป็นตัวเลขใหม่ 100 พบว่า ผลบวกในเซลล์ C1 ก็จะไปเปลี่ยนไปโดยอัตโนมัติ นี่คือการคำนวณพื้นฐานของ Excel

	A	B	C	D	E
1	10	100	110		
2					
3					
4					
5					

รูปที่ 1-20 ผลบวกที่เปลี่ยนแปลงไปตามค่าปัจจุบัน

สรุปท้ายบท

เนื้อหาในบทนี้เป็นเตรียมความพร้อมก่อนใช้งานภาษา R และโปรแกรม Excel เข้ามาช่วยทำงานกับข้อมูลที่เราสนใจ ก็จะเป็นเนื้อหาตั้งแต่บทที่ 2 เป็นต้นไป

CHAPTER

02



พื้นฐานการเขียนโปรแกรม ด้วยภาษา R

เนื้อหาในบทนี้ เป็นบทแรกที่จะนำเสนอเนื้อหาการเขียนโปรแกรมด้วยภาษา R โดยที่ผู้เขียนจะเริ่มต้นจากพื้นฐานลำดับแรกสุดเพื่อใช้เป็นฐานในบทต่อไป

การสร้างตัวแปร (Variable) ขึ้นมาทำหน้าที่เก็บข้อมูล

ตัวแปร (Variable) ในภาษา R ทำหน้าที่เหมือนกับภาษาอื่นๆ นั่นคือ เก็บข้อมูล คุณสามารถสร้างตัวแปรและสั่งให้ตัวแปรดังกล่าว เก็บข้อมูลได้ 2 รูปแบบคือ

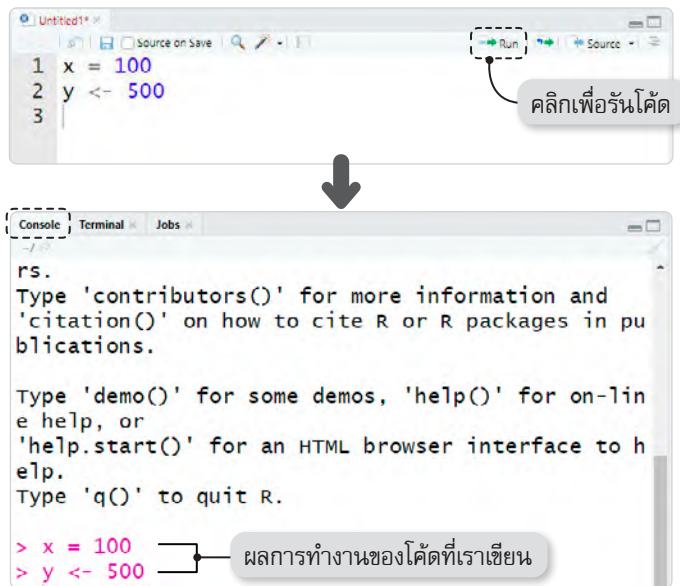
1. ใช้เครื่องหมาย = ทำหน้าที่กำหนดให้ตัวแปรเก็บค่า
2. ใช้เครื่องหมาย <- ทำหน้าที่กำหนดให้ตัวแปรเก็บค่าเช่นกัน โดยที่ภาษา R ใช้วิธีนี้เป็นหลัก

ต้องการสร้างตัวแปร x เก็บเลขจำนวนเต็ม 100 โดยใช้เครื่องหมาย = ส่วนตัวแปร y เก็บเลขจำนวนเต็ม 500 โดยใช้เครื่องหมาย <- ได้ผลเหมือนกัน


การนำเสนอเนื้อหาในหนังสือเล่มนี้ จะใช้เครื่องหมาย <- เข้ามาทำหน้าที่กำหนดค่าให้กับตัวแปรเป็นหลัก

R

```
x = 100  
y <- 500
```



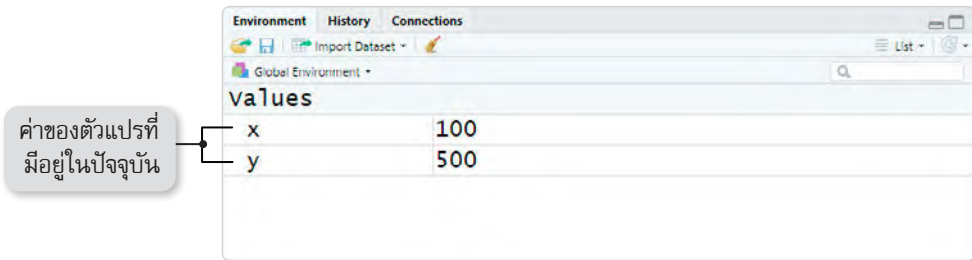
รูปที่ 2-1 แสดงการสร้างตัวแปร x กับ y เก็บเลขจำนวนเต็ม

จากรูปที่ 2-1 คุณสามารถคลิกที่ปุ่ม  เพื่อสั่งให้รันคำสั่งข้างต้น ผลการทำงานแสดงในแท็บ Console ด้านล่าง

Note

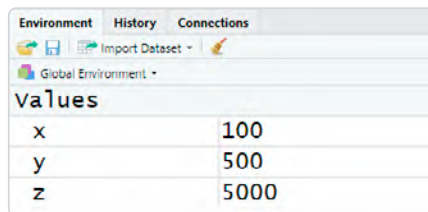
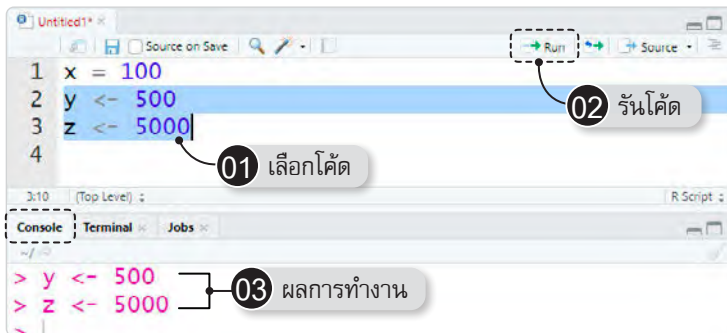
คุณสามารถล้างผลการทำงานเดิมในหน้าต่าง Console โดยการกดคีย์ **Ctrl+L**

ส่วนที่แท็บ **Environment** จะแสดงค่าของตัวแปรเท่าที่มีอยู่ในช่วงเวลาปัจจุบัน นั่นคือ x เก็บค่า 100 ส่วนตัวแปร y เก็บค่า 500 ดังรูปที่ 2-2



รูปที่ 2-2 แสดงค่าของตัวแปร x กับ y ในปัจจุบัน

ในกรณีที่คุณต้องการรันโค้ดมากกว่า 1 บรรทัดในเวลาเดียวกัน ให้คุณเลือกโค้ดที่ต้องการรันก่อน ในกรณีนี้คือ ครบคลุมตัวแปร y กับ z จากนั้นคลิกที่ปุ่ม **Run** ก็จะได้ค่าของตัวแปร y กับ z ทั้ง 2 ตัวในหน้าต่าง Console ดังรูปที่ 2-3



รูปที่ 2-3 กรณีต้องการรันโค้ดมากกว่า 1 บรรทัดในเวลาเดียวกัน

ในกรณีที่คุณต้องการลบตัวแปรออก คุณสามารถใช้ฟังก์ชันที่ชื่อว่า `rm()` เช่น ต้องการลบตัวแปร `z` ออกจากระบบ ให้พิมพ์คำสั่ง `rm(z)` และรันโค้ด พบว่าที่แท็บ Environment ตัวแปร `z` จะถูกนำออกไปด้วย ดังรูปที่ 2-4

```

1 x = 100
2 y <- 500
3 z <- 5000
4 rm(z)
5

```

values	
x	100
y	500

รูปที่ 2-4 กรณีลบตัวแปร `z` ออกจากระบบ

ภาษา R ถือว่า ตัวอักษรพิมพ์เล็กกับพิมพ์ใหญ่แตกต่างกัน ขอให้คุณระมัดระวังเรื่องนี้เป็นพิเศษด้วย ผู้เขียนสร้างตัวแปร `x` (ตัวพิมพ์เล็ก) กับตัวแปร `X` (ตัวพิมพ์ใหญ่) ถือว่า เป็นตัวแปรคนละตัว ดังรูปที่ 2-5

R

```

x <- 100
X <- 500

```

```

1 x <- 100
2 X <- 500
3 |

```

```


> x <- 100
> X <- 500
> |

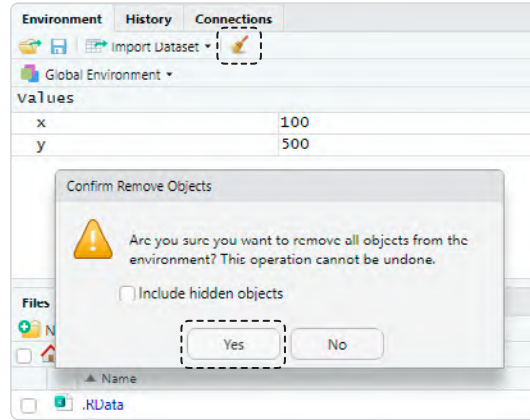
```

values	
x	100
X	500

รูปที่ 2-5 ค่าของตัวแปร `x` (ตัวพิมพ์เล็ก) กับ `X` (ตัวพิมพ์ใหญ่)

การลบรายการตัวแปรที่ค้างอยู่ในระบบ

เมื่อคุณมีการใช้งานตัวแปร ในภาษา R ถือเป็น object) ต่างๆ มากมาย รายการตัวแปรเหล่านี้ ยังคงค้างอยู่ในแท็บ Environment คุณสามารถคลิกที่ปุ่ม  เพื่อลบรายการตัวแปรเหล่านี้ ออกไปได้เช่นกัน ดังรูปที่ 2-6



รูปที่ 2-6 แสดงการลบรายการตัวแปร (หรือรายการ objects)

ชนิดข้อมูลพื้นฐาน (Basic Data Types)

ภาษา R มีชนิดข้อมูลพื้นฐาน (Basic Data Types) ทำหน้าที่เก็บข้อมูลเบื้องต้น ได้แก่

- **character** หมายถึง ตัวอักษรหรือข้อความ
- **numeric** หมายถึง ตัวเลขจำนวนเต็มและตัวเลขทศนิยม
- **integer** หมายถึง ตัวเลขจำนวนเต็ม
- **logical** หมายถึง ค่าทางตรรกะ
- **complex** หมายถึง จำนวนเชิงซ้อน
- **Date** หมายถึง ข้อมูลประเภทวันที่
- **factor** หมายถึง ข้อมูลที่ใช้จัดกลุ่ม

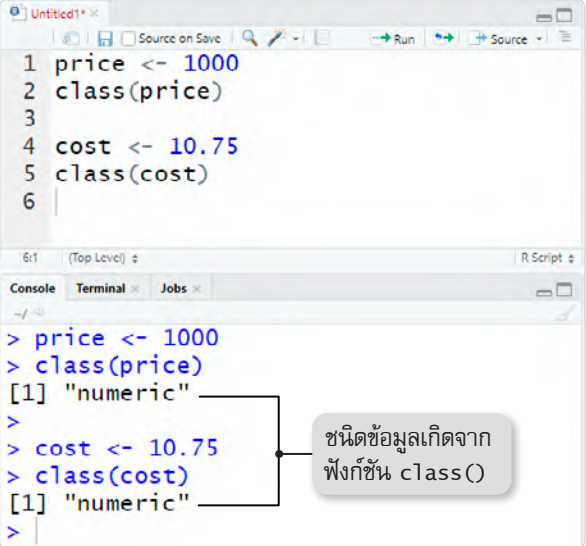
ชนิดข้อมูลเลขจำนวนเต็ม integer และเลขทศนิยม numeric

สร้างตัวแปรที่ชื่อว่า price เก็บตัวเลข 1000 (เลขจำนวนเต็ม) และตัวแปร cost เก็บตัวเลข 10.75 (เลขทศนิยม) เมื่อใช้ฟังก์ชัน class() เข้ามาทำหน้าที่ตรวจสอบตัวแปรทั้ง 2 ตัว พบว่า เป็นข้อมูลประเภท numeric ดังรูปที่ 2-7

R

```
price <- 1000
class(price)

cost <- 10.75
class(cost)
```



```
1 price <- 1000
2 class(price)
3
4 cost <- 10.75
5 class(cost)
6
```

Console

```
> price <- 1000
> class(price)
[1] "numeric"
>
> cost <- 10.75
> class(cost)
[1] "numeric"
>
```

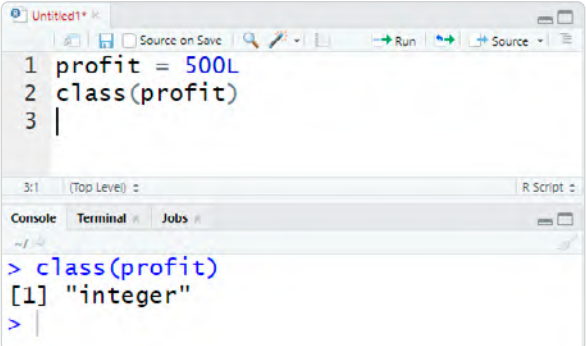
ชนิดข้อมูลเกิดจากฟังก์ชัน class()

รูปที่ 2-7 แสดงชนิดข้อมูลประเภทเลขทศนิยม numeric

ในกรณีที่คุณต้องการสร้างตัวแปรประเภทเลขจำนวนเต็ม integer ให้คุณใช้ตัวอักษร L ต่อท้ายตัวเลข เช่น สร้างตัวแปรที่ชื่อว่า profit ทำหน้าที่เก็บเลขจำนวนเต็ม 500 เมื่อตรวจสอบด้วยฟังก์ชัน class() พบว่า เป็นชนิดข้อมูล integer ดังรูปที่ 2-8

R

```
profit = 500L
class(profit)
```



```
1 profit = 500L
2 class(profit)
3
```

Console

```
> class(profit)
[1] "integer"
>
```

รูปที่ 2-8 แสดงชนิดข้อมูลประเภทเลขจำนวนเต็ม integer