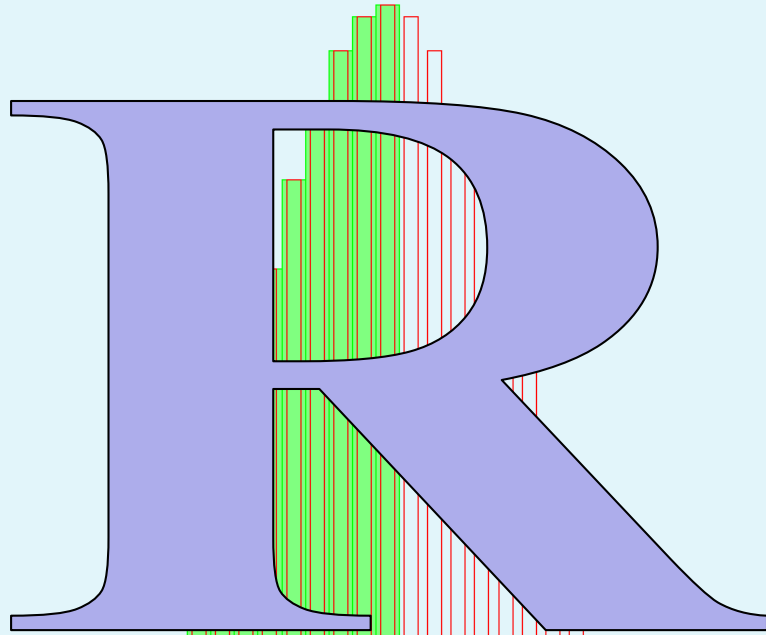


การวิเคราะห์ข้อมูลทางสถิติ ด้วยโปรแกรม R



R

R Gui

R Studio

R Commander

รองศาสตราจารย์ ดำรงค์ ทิพย์โยธา
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

การวิเคราะห์ข้อมูลทางสถิติด้วย โปรแกรม R

รองศาสตราจารย์ดำรงค์ ทิพย์โยธา

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ชื่อหนังสือ : การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรม R

ชื่อผู้แต่ง : รองศาสตราจารย์ ดำรงค์ ทิพย์โยธา

จัดทำโดย : รองศาสตราจารย์ ดำรงค์ ทิพย์โยธา

พิมพ์ครั้งที่ : 1

เดือนและปีที่ พ.ศ. ที่จัดพิมพ์ : มกราคม พ.ศ. 2561

เลขมาตรฐานสากลประจำหนังสืออิเล็กทรอนิกส์

ISBN (e-book) : 978-616-455-137-4

พิมพ์ที่ : ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

คำนำ

หนังสือ การวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรม R เป็นคู่มือใช้งานโปรแกรม R RGui RStudio R Commander เพื่อวิเคราะห์ข้อมูลทางสถิติ เนื้อหาภายในเล่มประกอบด้วย

การคำนวณค่าสถิติเบื้องต้น เช่น ค่าเฉลี่ย ค่าความแปรปรวน

การสร้างแฟ้มข้อมูล การนำเสนอข้อมูลในรูปแบบกราฟ

การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของค่าพารามิเตอร์

การทดสอบสมมติฐาน เช่น $H_0 : \mu = \mu_0$, $H_0 : \mu_1 - \mu_2 = d_0$, $H_0 : \sigma^2 = \sigma_0^2$

การทดสอบภาวะสารูปสนิทธิ (Test of Goodness Of Fit)

การทดสอบสมมติฐานว่าข้อมูลเป็นอิสระต่อกันหรือไม่

การถดถอยเชิงเส้นตรง สหสัมพันธ์ สัมประสิทธิ์สหสัมพันธ์ และสัมประสิทธิ์การกำหนด

การเลือกรูปแบบความสัมพันธ์ที่ดีที่สุดสำหรับข้อมูล

การวิเคราะห์ความแปรปรวน (ANOVA)

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม เช่น ทวินาม ทวินามลบ Z, t, F, chi-square

เนื้อหาภายในเล่มจะมีขั้นตอนการทำงานทางสถิติ มีเหตุผลของทฤษฎีความน่าจะเป็นและสถิติประกอบการทำงาน แสดงสูตรสถิติซึ่งเป็นที่มาของค่าสถิติที่โปรแกรม R คำนวณมาให้ และสรุปผลการวิเคราะห์ข้อมูล โปรแกรม R และข้อมูลของหนังสือบางส่วน download ได้ดังนี้

โปรแกรม R : <https://cran.r-project.org/bin/windows/base/>

โปรแกรม RStudio : <https://www.rstudio.com/products/rstudio/download/>

ตัวอย่าง R Script ของหนังสือ :

<http://pioneer.netserv.chula.ac.th/~tdumrong/2301286data/Rprogramming.rar>

ผู้เขียนหวังว่าหนังสือเล่มนี้จะช่วยให้ผู้อ่านทุกท่านสามารถนำโปรแกรม R ไปใช้วิเคราะห์ข้อมูลได้เป็นอย่างดี และขอขอบคุณผู้อ่านทุกท่านที่ได้ติดตามผลงานของผู้เขียนมาโดยตลอด

รองศาสตราจารย์ ดำรงค์ ทิพย์โยธา

มกราคม 2561

สารบัญ

หน้า

บทที่ 1	ความสามารถเบื้องต้นของโปรแกรม R	1 – 158
----------------	--	----------------

1.1	การติดตั้งโปรแกรม R (RGui)	2
1.2	การติดตั้งโปรแกรม RStudio	15
1.3	ความสามารถเบื้องต้นของโปรแกรม R	24
1.4	การใช้งานโปรแกรม R ผ่าน RGui	43
1.5	การใช้งานโปรแกรม R ผ่าน RStudio	79
1.6	การใช้งานโปรแกรม R Commander	111
	ผ่านทางโปรแกรม RGui	

บทที่ 2	การสร้างแฟ้มข้อมูล	159 – 216
----------------	---------------------------	------------------

บทที่ 3	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของค่าพารามิเตอร์	217 – 332
----------------	--	------------------

3.1	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของค่าเฉลี่ย μ	217
3.2	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของผลต่างของ ค่าเฉลี่ย $\mu_1 - \mu_2$	239
3.3	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของค่าเฉลี่ยของ ผลต่าง μ_D	287
3.4	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของความแปรปรวน σ^2	309
3.5	การหาช่วงความเชื่อมั่น $(1 - \alpha)100\%$ ของสัดส่วน ความแปรปรวน $\frac{\sigma_1^2}{\sigma_2^2}$	315

บทที่ 4	การทดสอบสมมติฐาน	333 – 542
4.1	การทดสอบสมมติฐาน $H_0 : \mu = \mu_0$	334
4.2	การทดสอบสมมติฐาน $H_0 : \mu_1 - \mu_2 = d_0$ กรณีประชากร 2 ชุดเป็นอิสระต่อกัน	356
4.3	การทดสอบสมมติฐาน $H_0 : \mu_D = d_0$ กรณีประชากร 2 ชุดไม่เป็นอิสระต่อกัน (ข้อมูลเป็นคู่)	425
4.4	การทดสอบสมมติฐาน $H_0 : \sigma^2 = \sigma_0^2$	489
4.5	การทดสอบสมมติฐาน $\frac{\sigma_1^2}{\sigma_2^2} = k$	499
4.6	การทดสอบภาวะสารูปสนิทธิ (Test of Goodness Of Fit)	520
4.7	การทดสอบสมมติฐานว่าข้อมูลเป็นอิสระต่อกันหรือไม่	530
บทที่ 5	สหสัมพันธ์และการถดถอย	543 – 674
5.1	การหาสมการเส้นถดถอยเชิงเส้นตรงเชิงเดียว (Simple Linear Regression) และสหสัมพันธ์ (Correlation)	545
5.2	ช่วงความเชื่อมั่นของระยะตัดแกน สัมประสิทธิ์การถดถอย และสัมประสิทธิ์สหสัมพันธ์	578
5.3	การทดสอบสมมติฐานประชากร 2 ชุดมีความสัมพันธ์เชิงเส้นตรงโดยใช้ค่าสถิติที (t)	602
5.4	การทดสอบสมมติฐานประชากร 2 ชุดมีความสัมพันธ์เชิงเส้นตรงโดยใช้ค่าสถิติเอฟ (F)	614
5.5	การทดสอบสมมติฐานค่าสัมประสิทธิ์การถดถอยของสมการถดถอยเชิงเส้นตรง $H_0 : \beta = \beta_0$	627

5.6 การทดสอบสมมติฐานระยะตัดแกนของสมการถดถอยเชิง	642
---	-----

เส้นตรง $H_0 : \alpha = \alpha_0$

5.7 การเลือกรูปแบบความสัมพันธ์ที่ดีที่สุดสำหรับข้อมูล	650
---	-----

บทที่ 6	การวิเคราะห์ความแปรปรวน	675 – 744
----------------	--------------------------------	------------------

6.1 การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว	676
---	-----

(One-Way ANOVA, Simple-Factor ANOVA)

6.2 การวิเคราะห์ความแปรปรวนแบบที่มีการสุ่มอย่างสมบูรณ์ใน	702
--	-----

แต่ละกลุ่ม (Randomized Complete Block Designs,
Mutiple-Factor ANOVA)

บทที่ 7	การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม	745 – 808
----------------	--	------------------

7.1 ตัวแปรสุ่มปกติ (X) และ ตัวแปรสุ่มปกติมาตรฐาน (Z)	746
--	-----

7.2 ตัวแปรสุ่มที (t)	756
----------------------	-----

7.3 ตัวแปรสุ่มไคสแควร์ (χ^2)	762
-------------------------------------	-----

7.4 ตัวแปรสุ่มเอฟ (F)	769
-----------------------	-----

7.5 ตัวแปรสุ่มทวินาม (Binomial)	775
---------------------------------	-----

7.6 ตัวแปรสุ่มทวินามลบ (Negative binomial) และ	780
--	-----

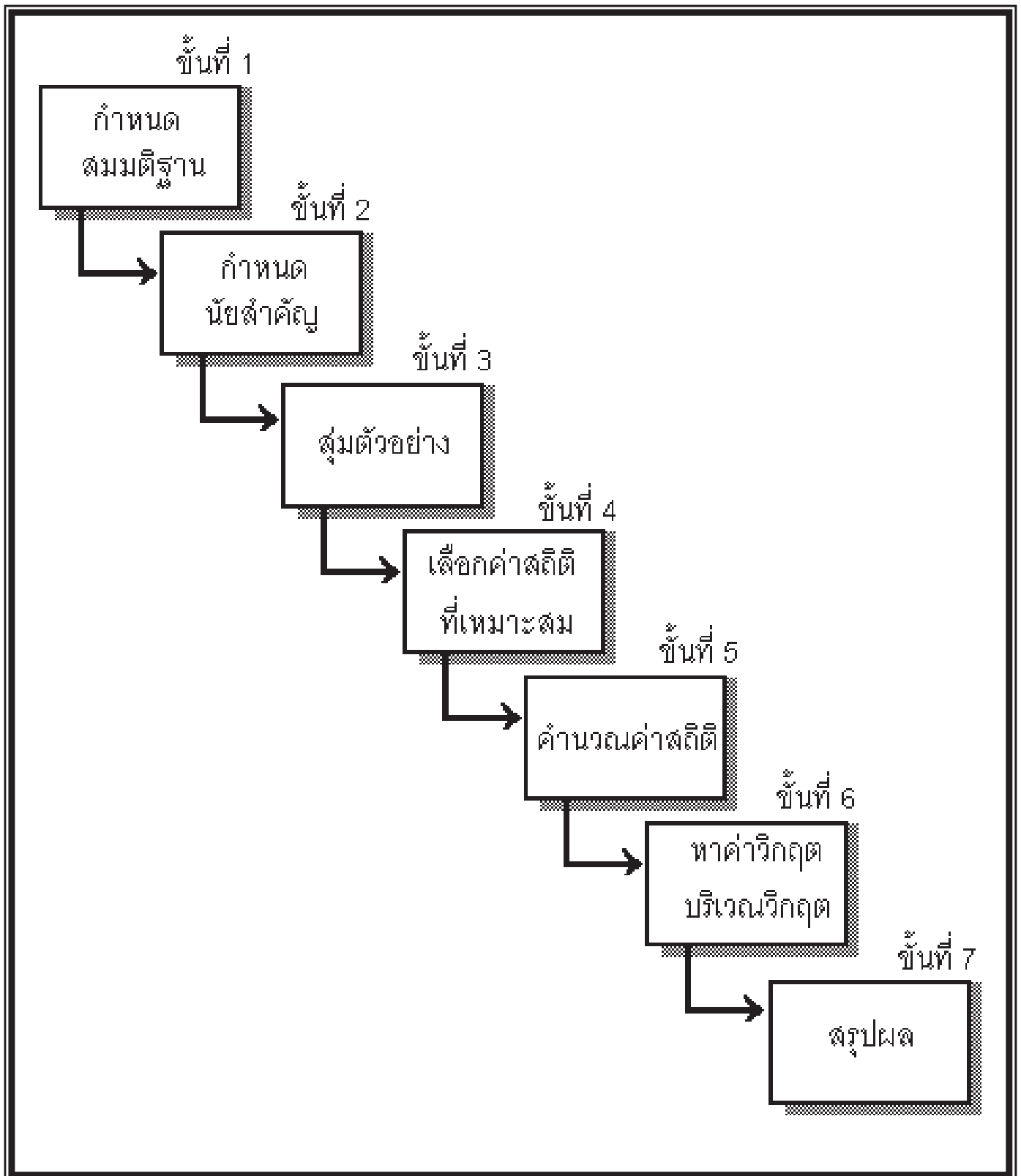
ตัวแปรสุ่มเรขาคณิต (Geometric)

7.7 ตัวแปรสุ่มไฮเพอร์จีอเมตริก (Hypergeometric)	787
---	-----

7.8 ตัวแปรสุ่มปัวส์ซอง (Poisson)	791
----------------------------------	-----

7.9 โค้งปฏิบัติการแผนการสุ่มตัวอย่าง (OC-curve)	797
---	-----

การทดสอบสมมติฐาน



หมายเหตุ ในขั้นที่ 4 ค่าสถิติที่เหมาะสมและนิยมใช้กันมากคือค่า Z , t , F และ χ^2

ขั้นที่ 7 การสรุปผลทำได้ 2 แบบ

- นำค่าสถิติจากตัวอย่าง t จำนวน, z จำนวน, f จำนวน, χ^2 จำนวน เปรียบเทียบกับค่าวิกฤต
- นำค่า Sig. (1-tailed), Sig. (2-tailed) เปรียบเทียบกับระดับนัยสำคัญ α

ความสามารถเบื้องต้นของโปรแกรม R

โปรแกรม R เป็นโปรแกรมช่วยในการคำนวณและวิเคราะห์ข้อมูลทางสถิติ ในบทที่ 1 จะแนะนำการติดตั้งโปรแกรมและนำความสามารถเบื้องต้นของโปรแกรม R ที่คำนวณได้มาให้ดูก่อนเพื่อผู้อ่านจะได้เห็นความสามารถทางการคำนวณที่สำคัญของโปรแกรม R ส่วนการประยุกต์ต่าง ๆ จะกล่าวในบทต่อไป ตัวอย่างภาพหน้าจอแรกของการทำงานบน Window ของโปรแกรม R ซึ่งเรียกว่า RGui

```

RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

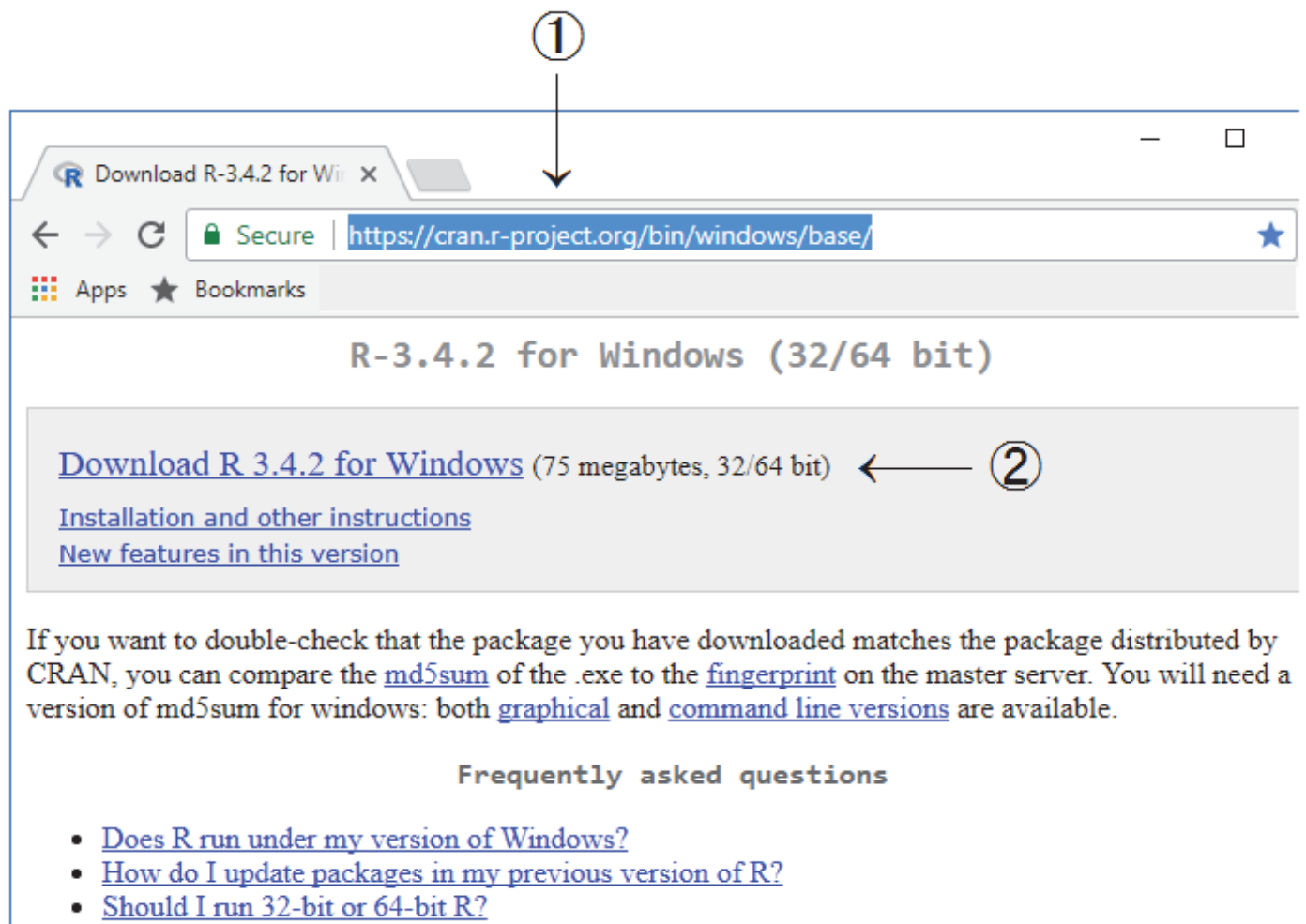
[Previously saved workspace restored]

> 1+2
[1] 3
> |
  
```

1.1 การติดตั้งโปรแกรม R (RGui)

เริ่มต้นให้ Download โปรแกรม R จาก Website

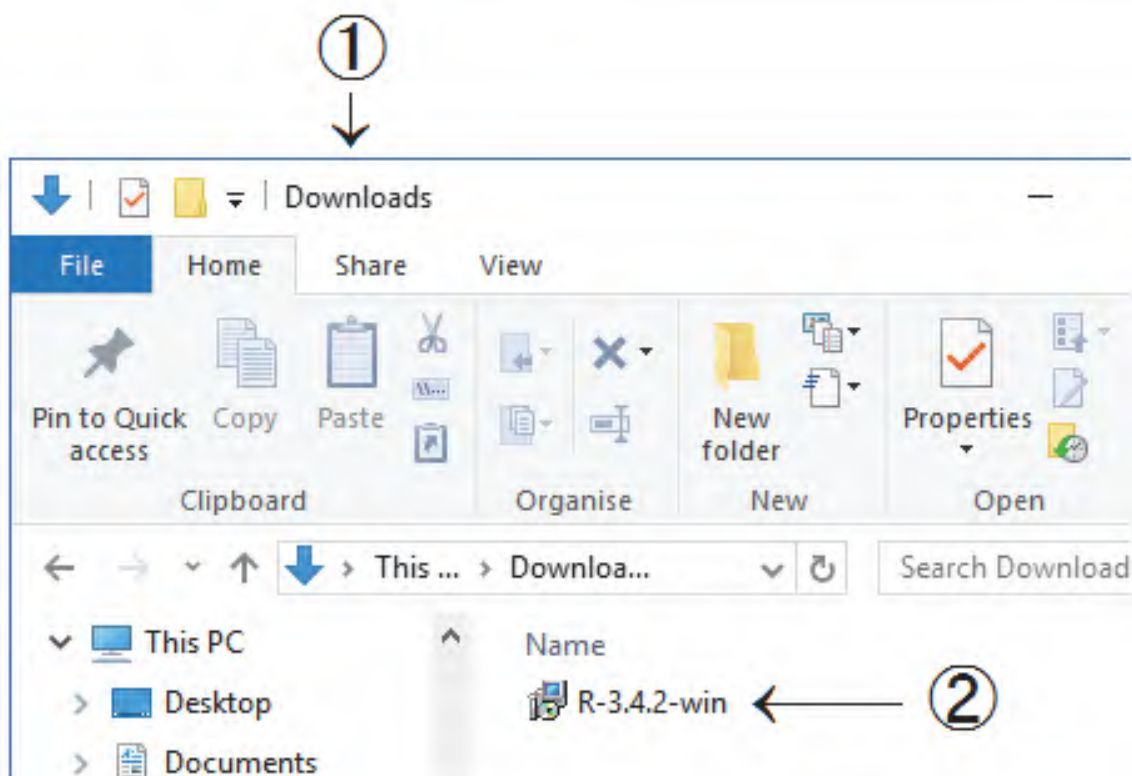
1. <https://cran.r-project.org/bin/windows/base/>
2. Download R 3.4.2 for Windows



หมายเหตุ Version ของโปรแกรม R อาจมีการเปลี่ยนแปลงเป็น Version ที่ใหม่ขึ้น

เมื่อ Download เสร็จแล้วจะได้โปรแกรม R-3.4.2-win ในคอมพิวเตอร์ที่ Directory ชื่อ Downloads

หมายเหตุ ชื่อ Directory Defaults คือ Directory ชื่อ Downloads แต่ในคอมพิวเตอร์บางเครื่อง อาจจะ Download โปรแกรม R-3.4.2-win ไว้ใน Directory ชื่ออื่นตามที่ผู้ใช้กำหนด

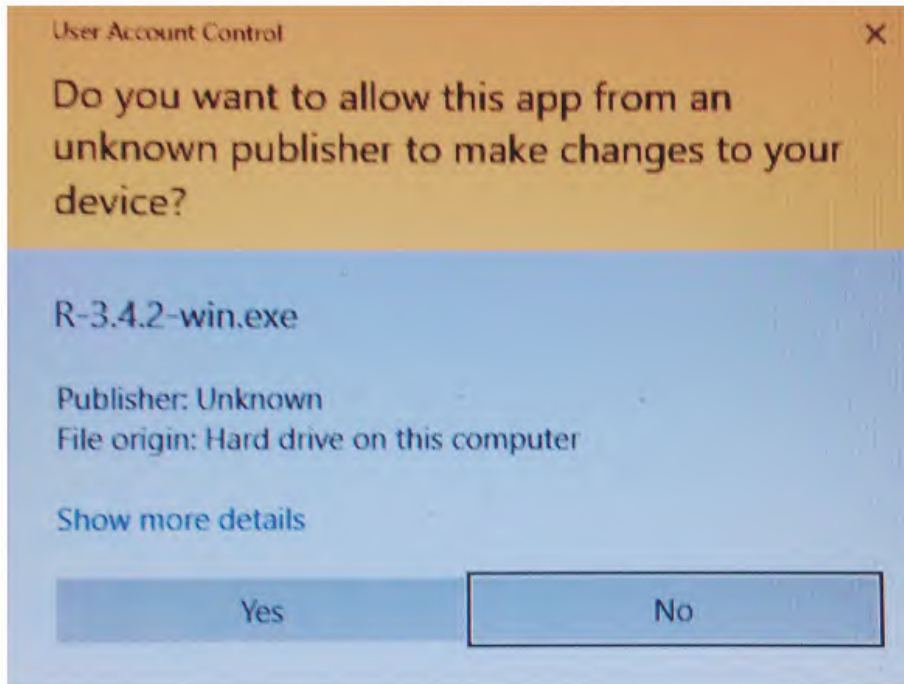


การติดตั้งโปรแกรม R (RGui) มีขั้นตอนดังนี้

ขั้นที่ 1

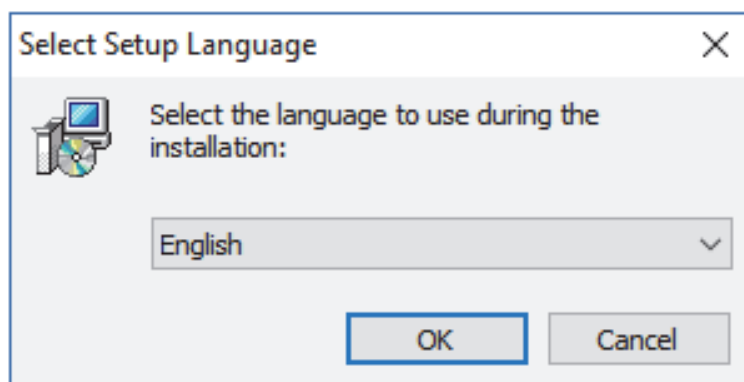
1. ให้ไปที่ Directory ชื่อ Downloads
2. สั่งให้โปรแกรม R-3.4.2-win ทำงาน (กดดับเบิลคลิกที่ชื่อโปรแกรม R-3.4.2-win)

ขั้นที่ 2 บนจอภาพจะมีคำถามว่าต้องการติดตั้งโปรแกรมใช้หรือไม่



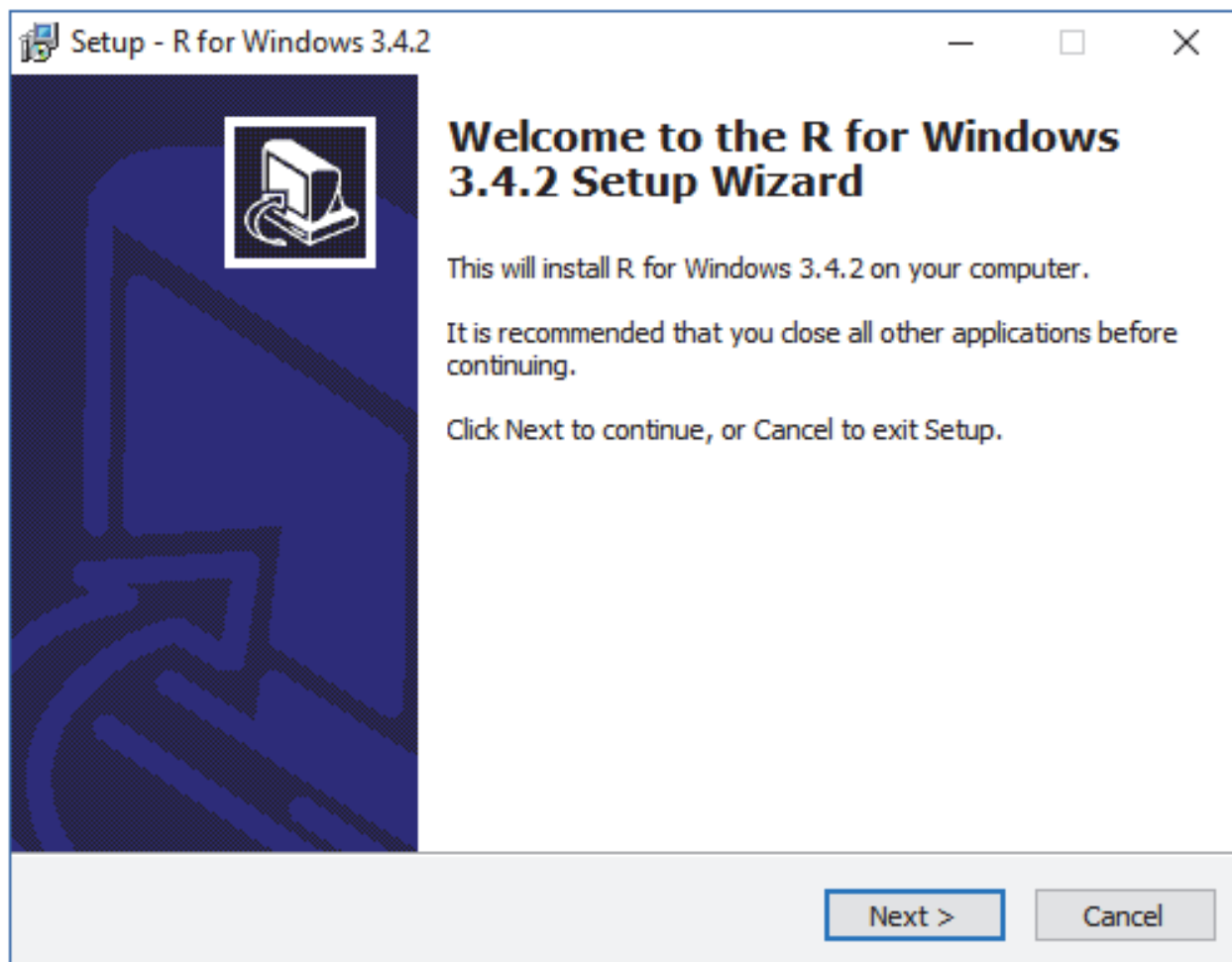
ให้ตอบ Yes เพื่อทำการติดตั้งโปรแกรมต่อไป

ขั้นที่ 3 บนจอภาพจะมีเมนูย่อย Select Setup Language ให้เลือกภาษา



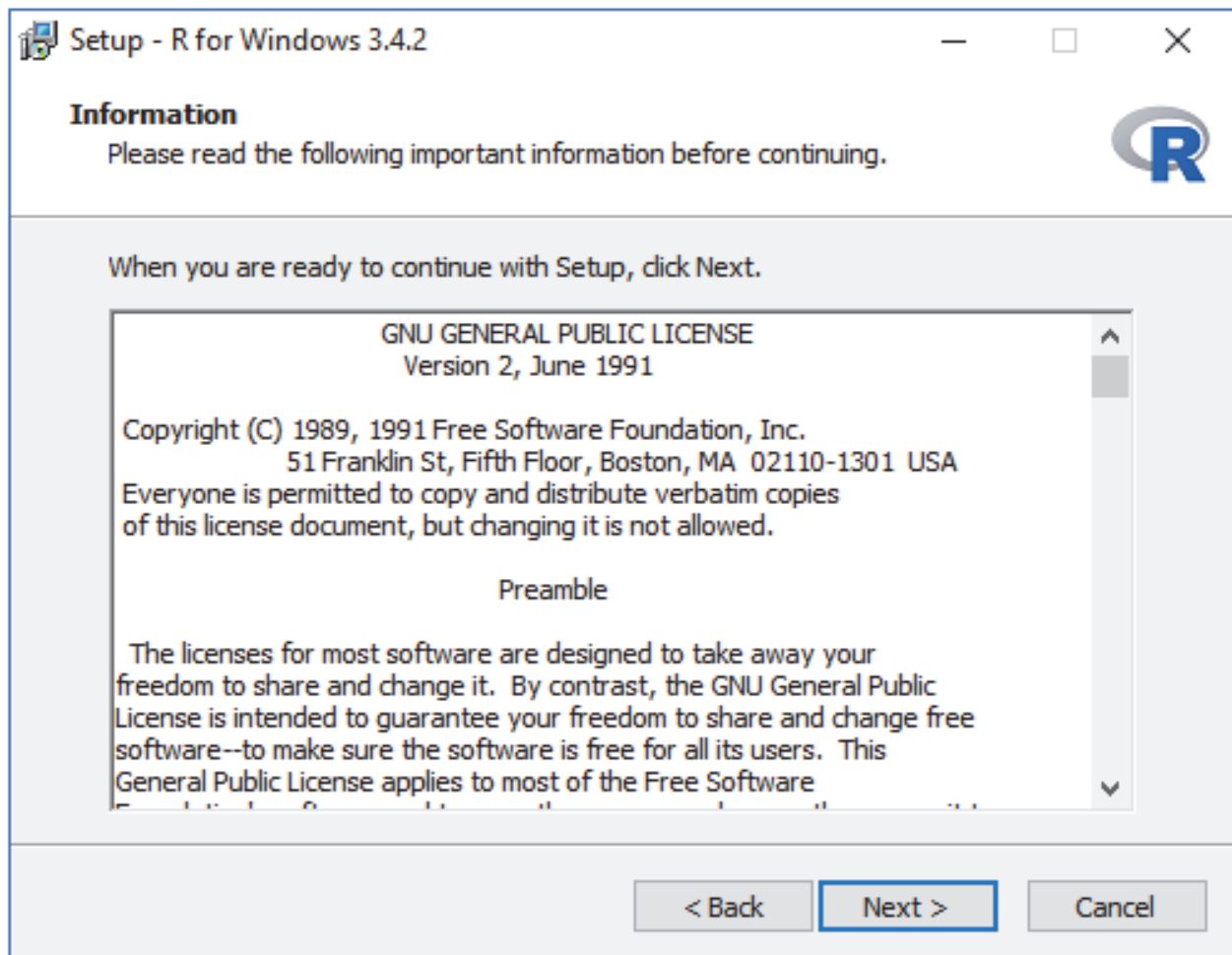
ให้เลือกภาษาอังกฤษ แล้วกด OK

ขั้นที่ 4 บนจอภาพจะมีข้อความยินดีต้อนรับพร้อมที่จะติดตั้งโปรแกรม R



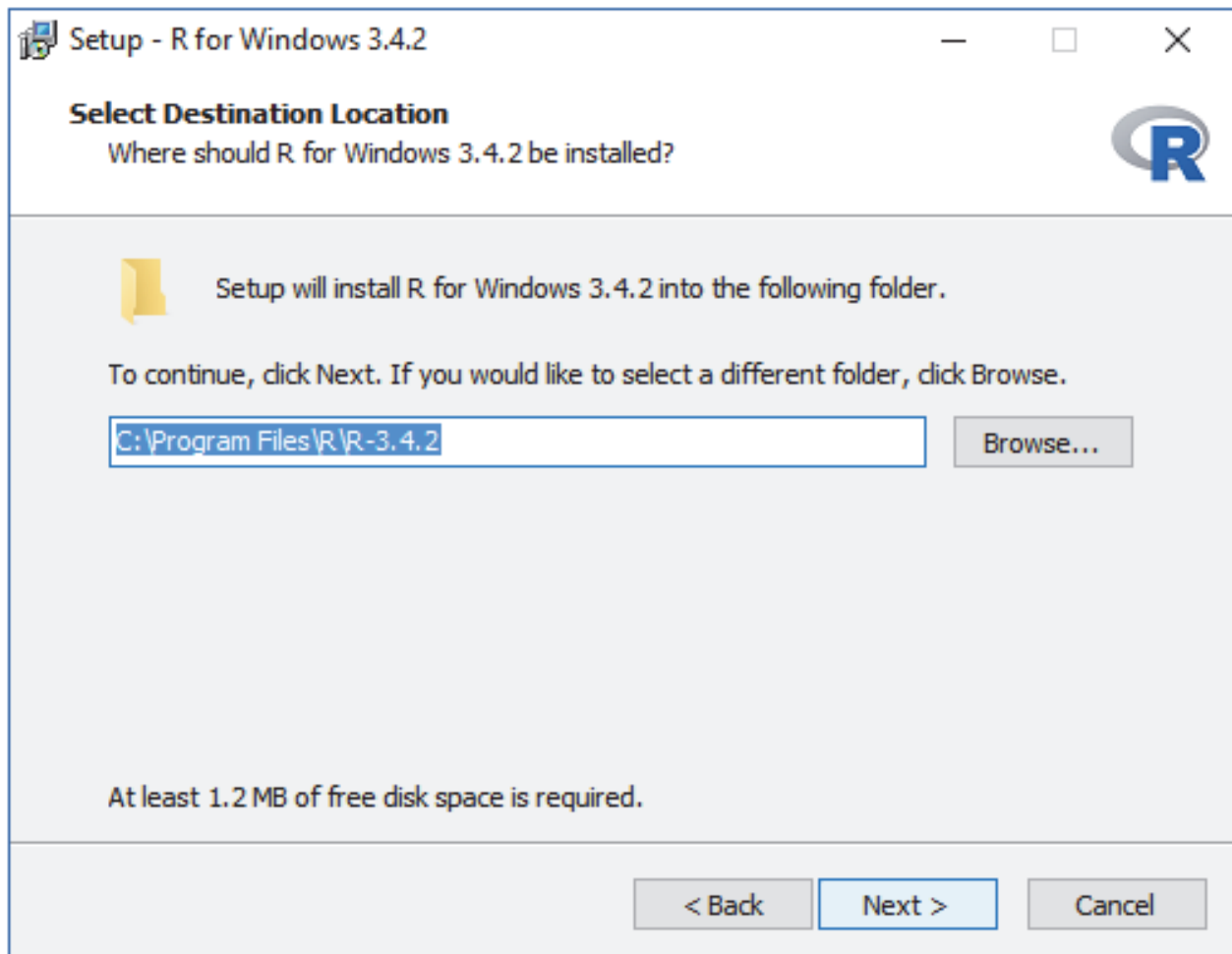
ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนต่อไป

ขั้นที่ 5 บนจอภาพจะมีข้อความเบื้องต้นเกี่ยวกับลิขสิทธิ์การใช้งาน



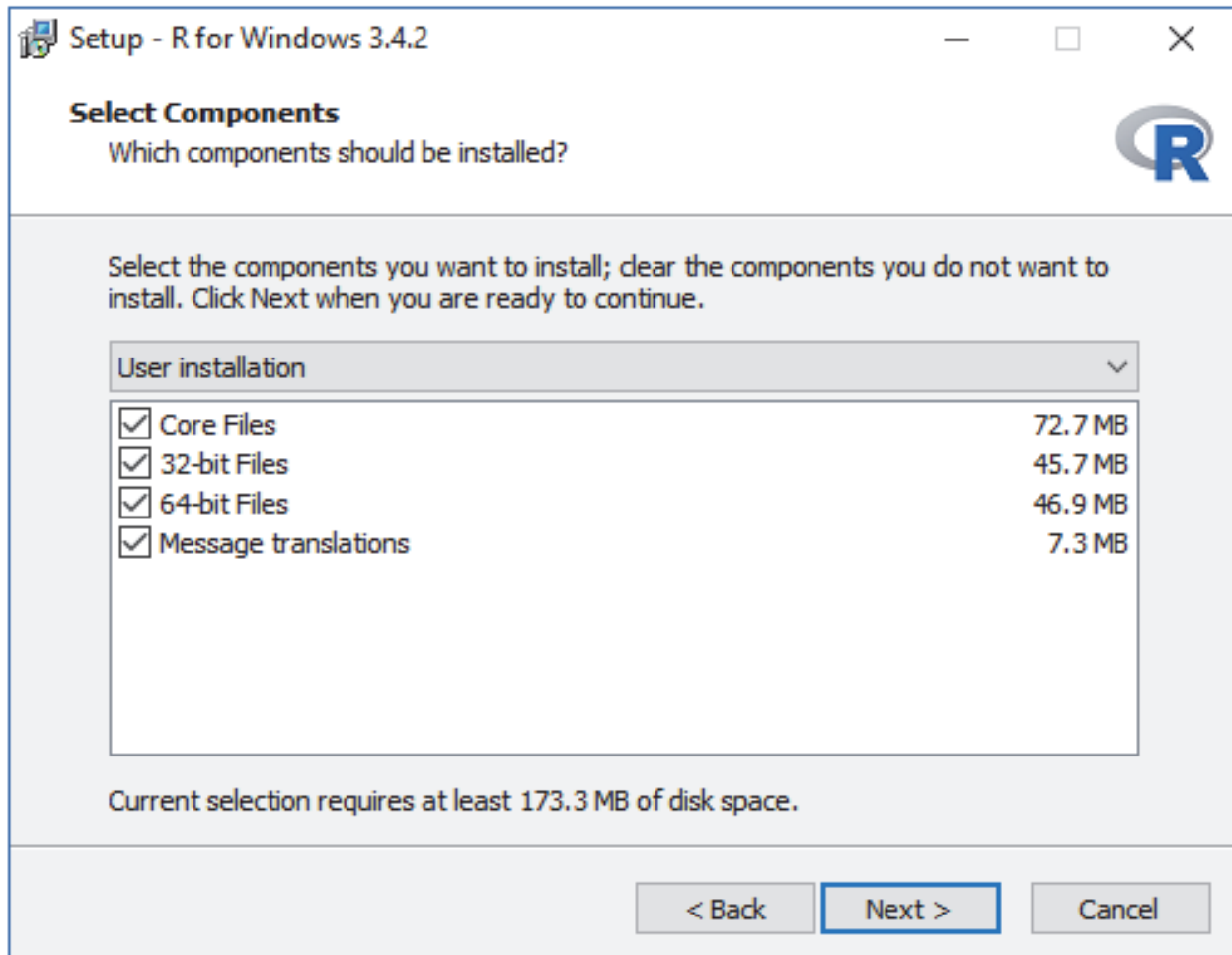
ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนต่อไป

ขั้นที่ 6 บนจอภาพจะมีเมนูย่อย Select Destination Location ให้เราเลือก Directory ที่จะติดตั้งโปรแกรม ในที่นี้เราเลือกใช้ตามคำแนะนำที่พิมพ์บนจอภาพ



ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนต่อไป

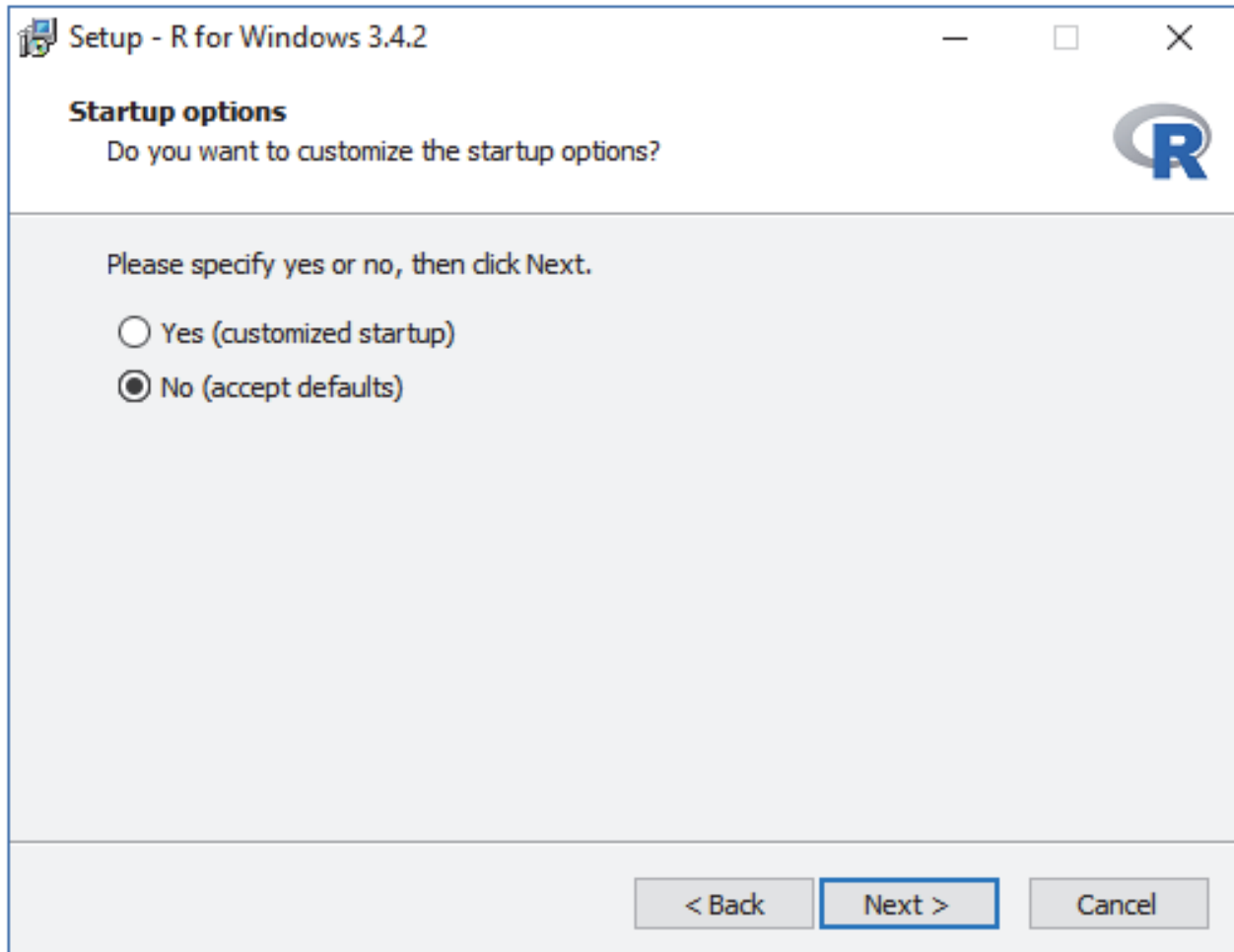
ขั้นที่ 7 บนจอภาพจะมีเมนูย่อย Select Components ให้เลือก Options การติดตั้ง ให้เราเลือกเพื่อความเหมาะสมในการใช้งาน



ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นต่อไป

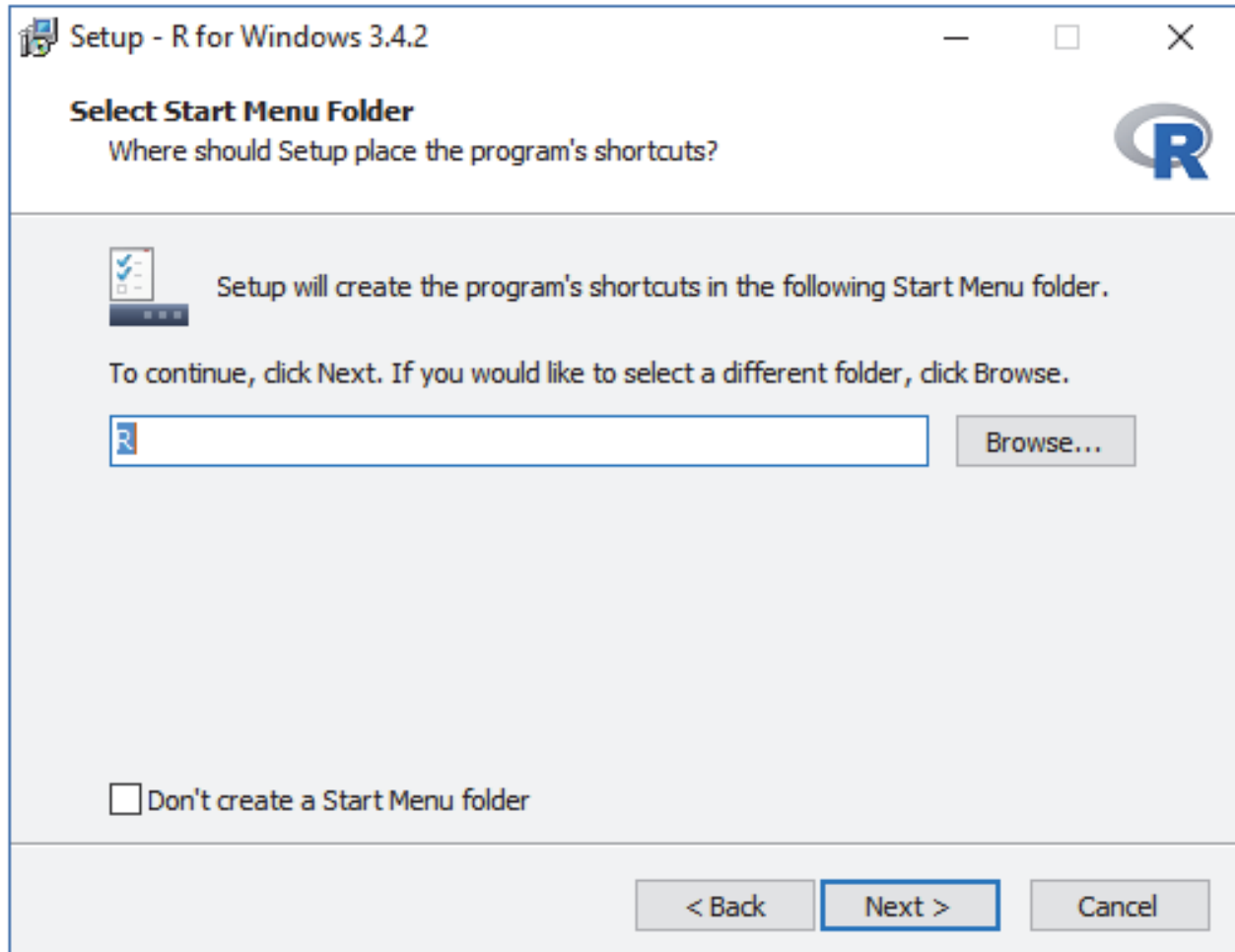
หมายเหตุ เลือกทั้งหมดหรือเลือกบาง Options ให้เหมาะสมกับเครื่องคอมพิวเตอร์ก็ได้

ขั้นที่ 8 บนจอภาพจะมีเมนูย่อย Startup options ให้เราเลือก Options ของการติดตั้งโปรแกรม ในที่นี้เราเลือก No (accept defaults)



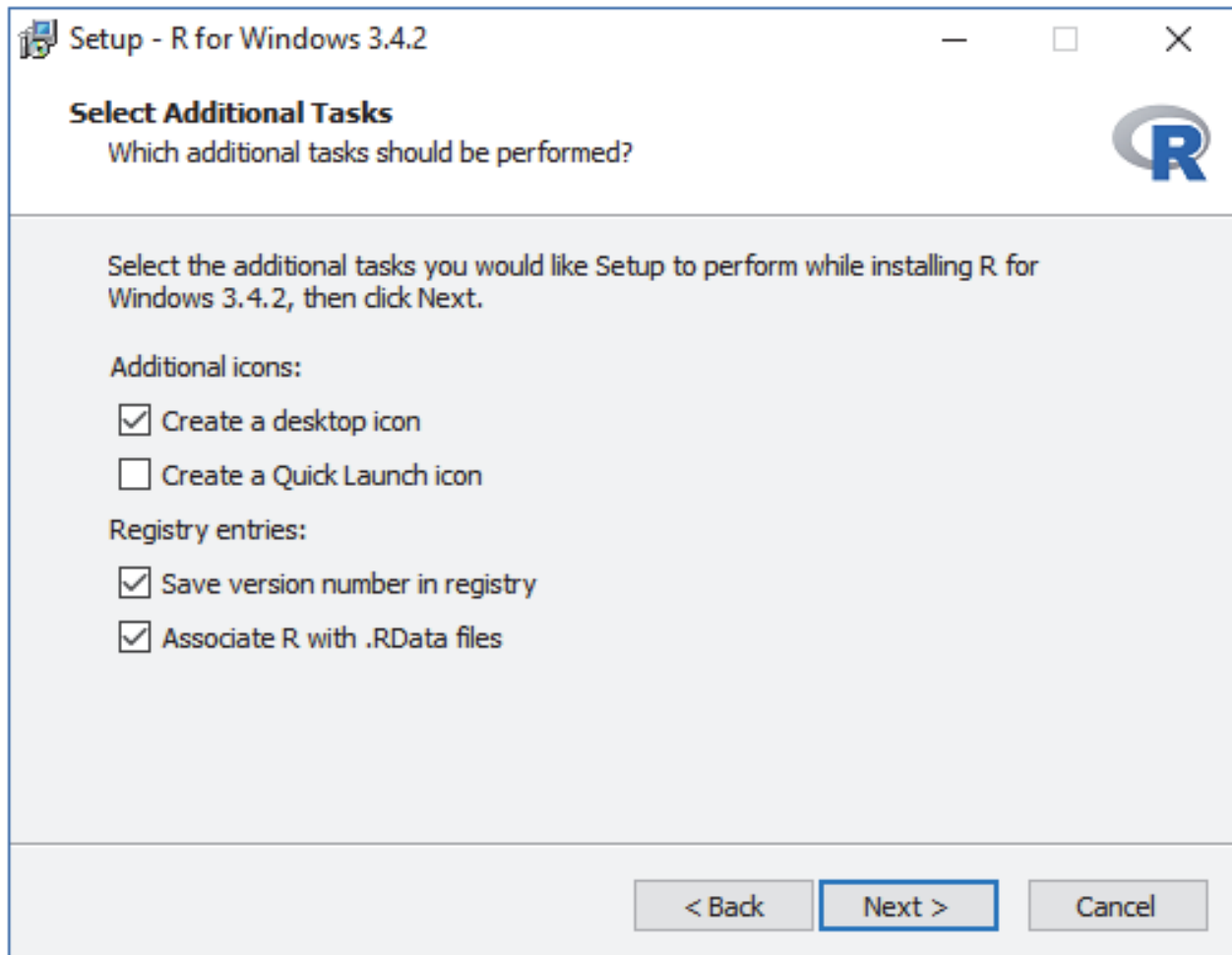
ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนต่อไป

ขั้นที่ 9 บนจอภาพจะมีเมนูย่อย Select Start Menu Folder เพื่อติดตั้ง Shortcuts ต่าง ๆ ของโปรแกรม R ในที่นี้เราเลือกตามคำแนะนำที่พิมพ์ไว้แล้วของโปรแกรม



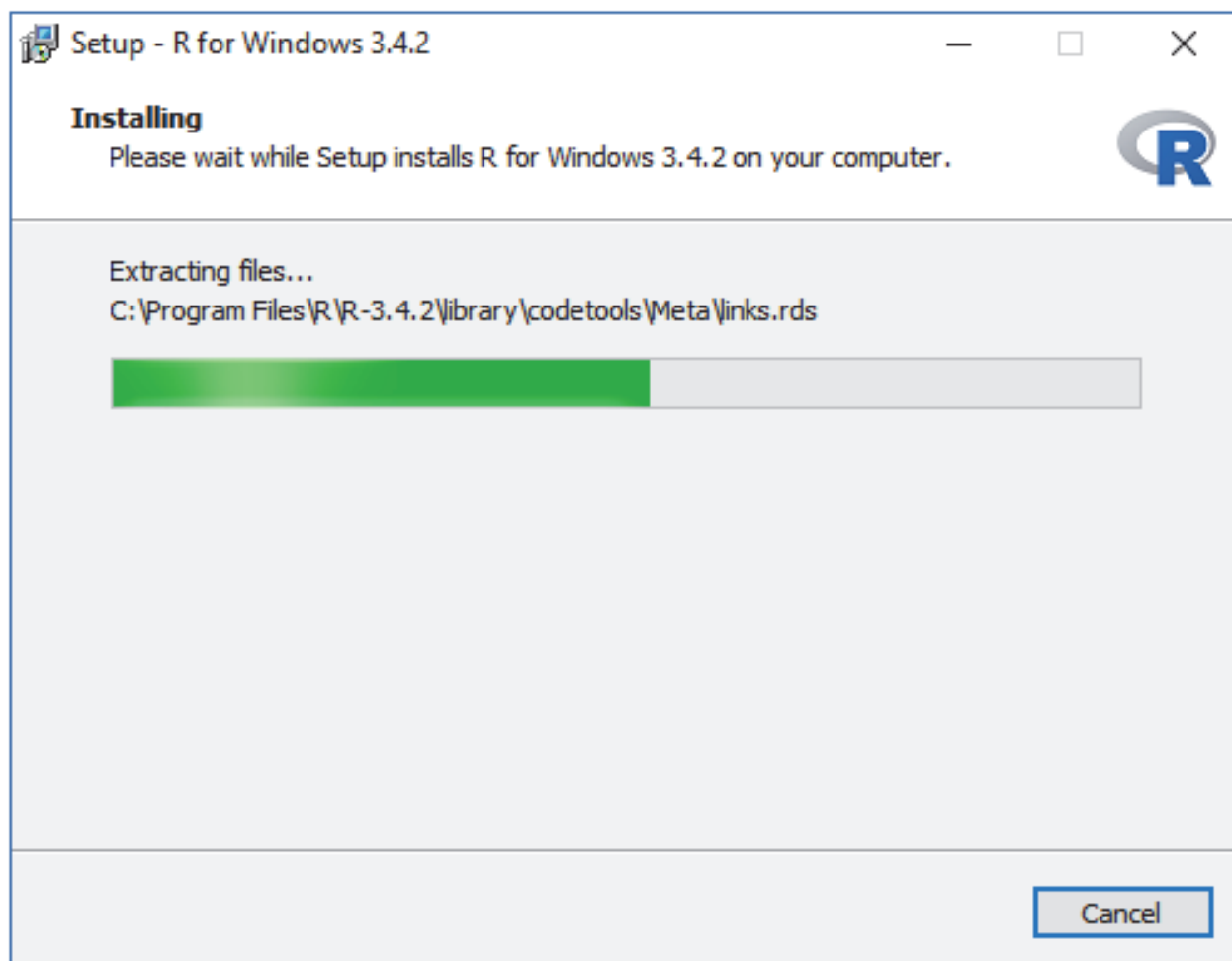
ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนต่อไป

ขั้นที่ 10 บนจอภาพจะมีเมนูย่อย Select Additional Tasks ให้เราเลือก Options ว่าต้องการแสดง icon shortcuts ของโปรแกรม R บนจอภาพเป็น desktop icon หรือ Quick Launch icon ให้ผู้ใช้เลือกตามความต้องการ



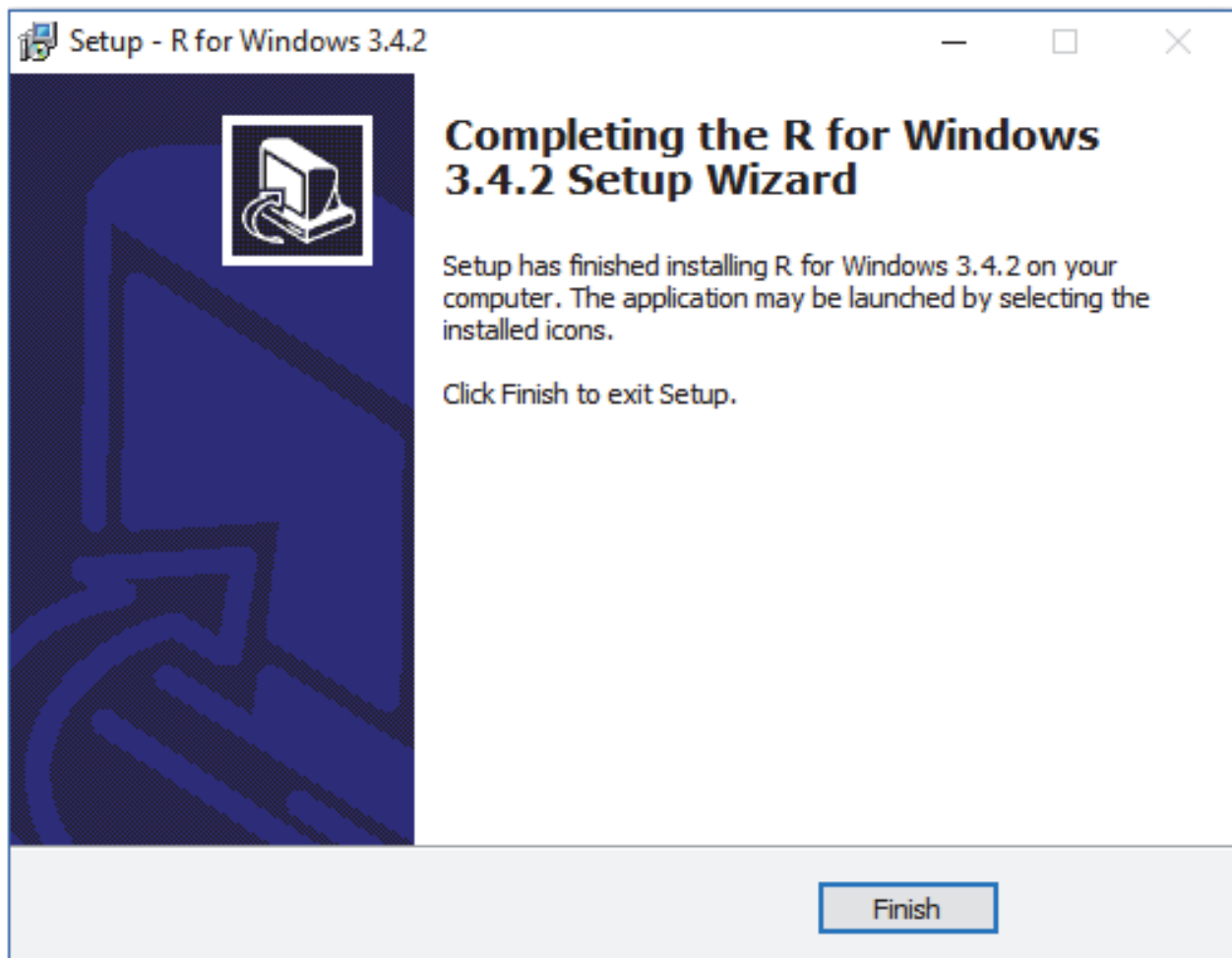
ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม R ขั้นตอนที่ต่อไป

ขั้นที่ 11 โปรแกรม Setup-R for Windows 3.4.2 จะทำการติดตั้งโปรแกรม R ให้



ต้องรอสักครู่จนติดตั้งโปรแกรม R เสร็จ

ขั้นที่ 12 เมื่อติดตั้งโปรแกรม R เสร็จ บนจอภาพจะบอกว่าการติดตั้งเสร็จสมบูรณ์



ให้คลิก Finish

ขณะนี้การติดตั้งโปรแกรม R เสร็จสมบูรณ์ พร้อมใช้งานโปรแกรม R แล้ว

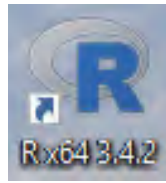
ขั้นตอนต่อไป ให้ไปที่จอภาพ Desktop จะพบ icon ของโปรแกรม R ดังนี้



หมายเหตุ หากไม่พบ Icon ให้สร้างเองขึ้นมาใหม่ หรือ Run โปรแกรม R ผ่านเมนู Application ของ Window

การเรียกโปรแกรม R ขึ้นมาใช้งาน

ให้ไปที่จอภาพ Desktop จะพบ icon

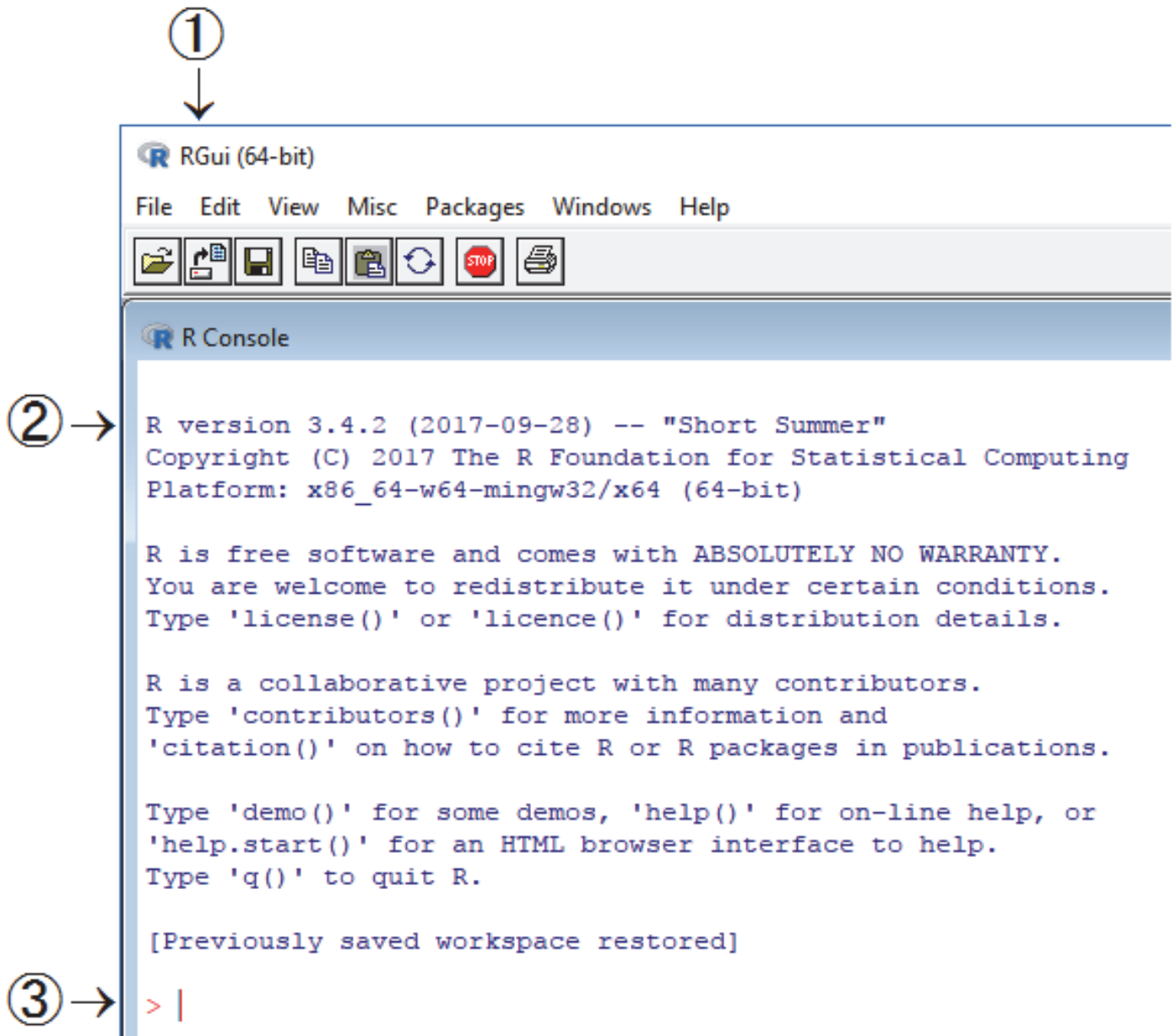


ให้ดับเบิลคลิกที่ icon



เพื่อเรียก

โปรแกรม R ขึ้นมาใช้งาน จอภาพแรกเมื่อเข้าสู่การใช้งานของโปรแกรม R เป็นดังนี้

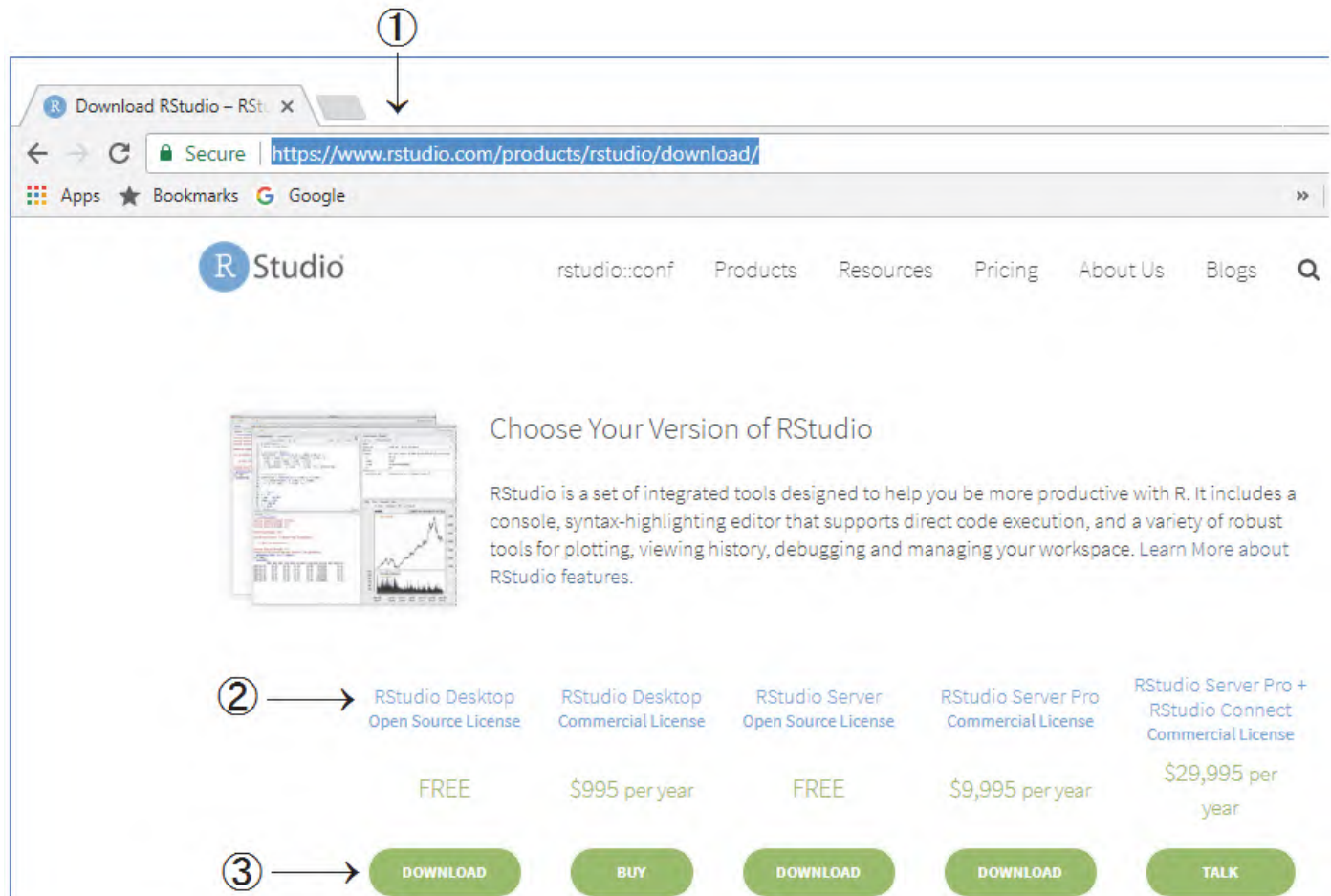


1. ชื่อโปรแกรม RGui (64-bit)
2. ระบุว่าเป็น R version 3.4.2 (2017-09-28)
3. ตำแหน่งเริ่มต้น (Prompt) ของการพิมพ์คำสั่งต่าง ๆ ในการใช้โปรแกรม R
4. ออกจากโปรแกรม R ให้พิมพ์ q() แล้วกด Enter

1.2 การติดตั้งโปรแกรม RStudio

เริ่มต้นให้ Download โปรแกรม RStudio จาก

Website <https://www.rstudio.com/products/rstudio/download/>



1. <https://www.rstudio.com/products/rstudio/download/>

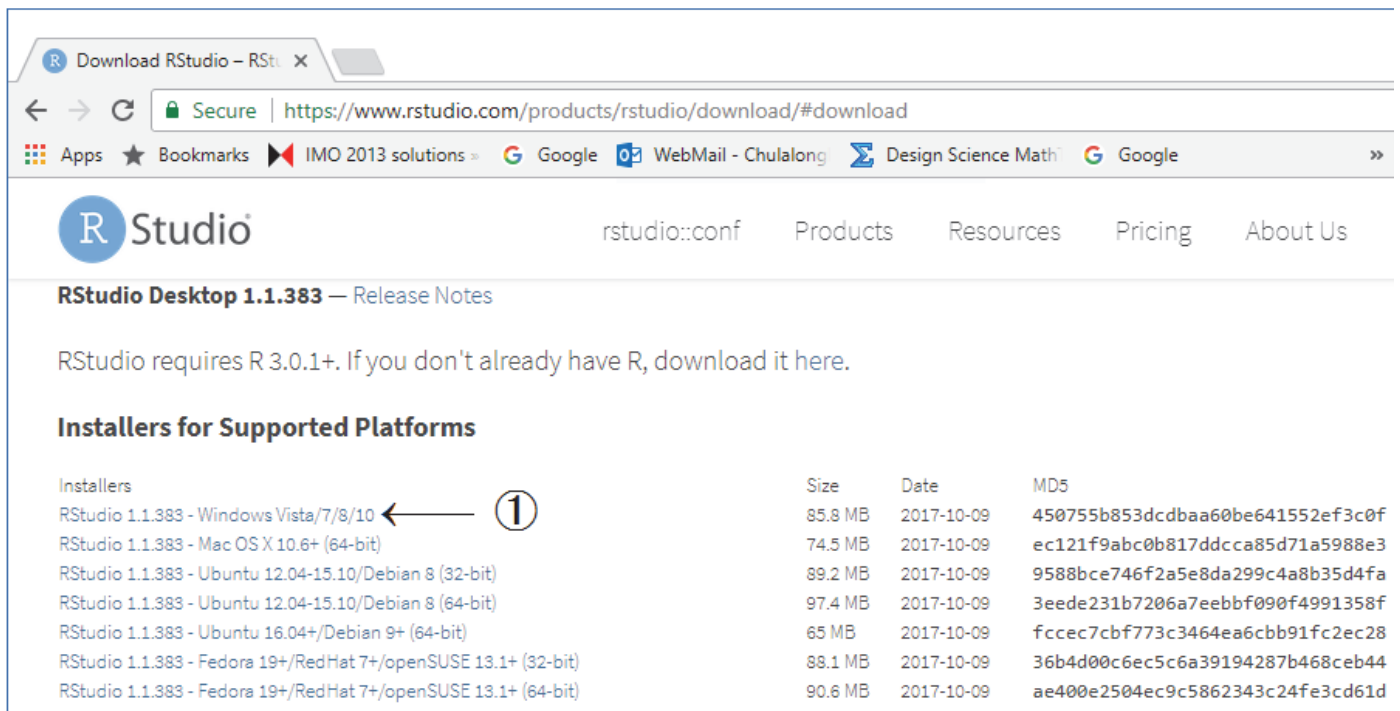
2. Download RStudio Desktop Open Source License

3. คลิกที่ **DOWNLOAD** เพื่อไปหน้าจอของรายชื่อโปรแกรมที่ต้องการ Download

คลิกที่


 DOWNLOAD

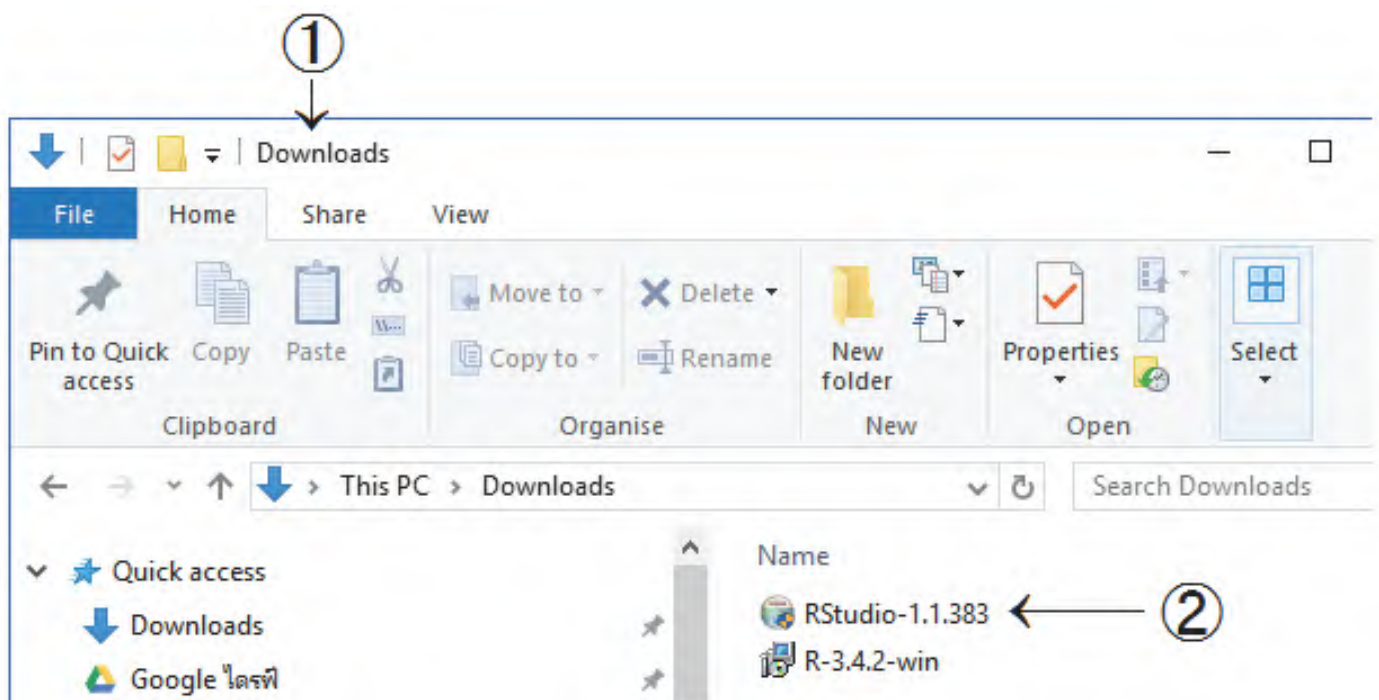
จะได้ผลบนจอภาพดังนี้



1. Download โปรแกรม RStudio 1.1.383 Windows Vista/7/8/10

เมื่อ Download เสร็จ จะได้โปรแกรม RStudio 1.1.383 ในคอมพิวเตอร์ที่ Directory ชื่อ Downloads

หมายเหตุ ชื่อ Directory Defaults คือ Directory Downloads แต่ในคอมพิวเตอร์บางเครื่องอาจจะ Download โปรแกรม RStudio 1.1.383 ไว้ใน Directory ชื่ออื่นตามที่ผู้ใช้งานกำหนด

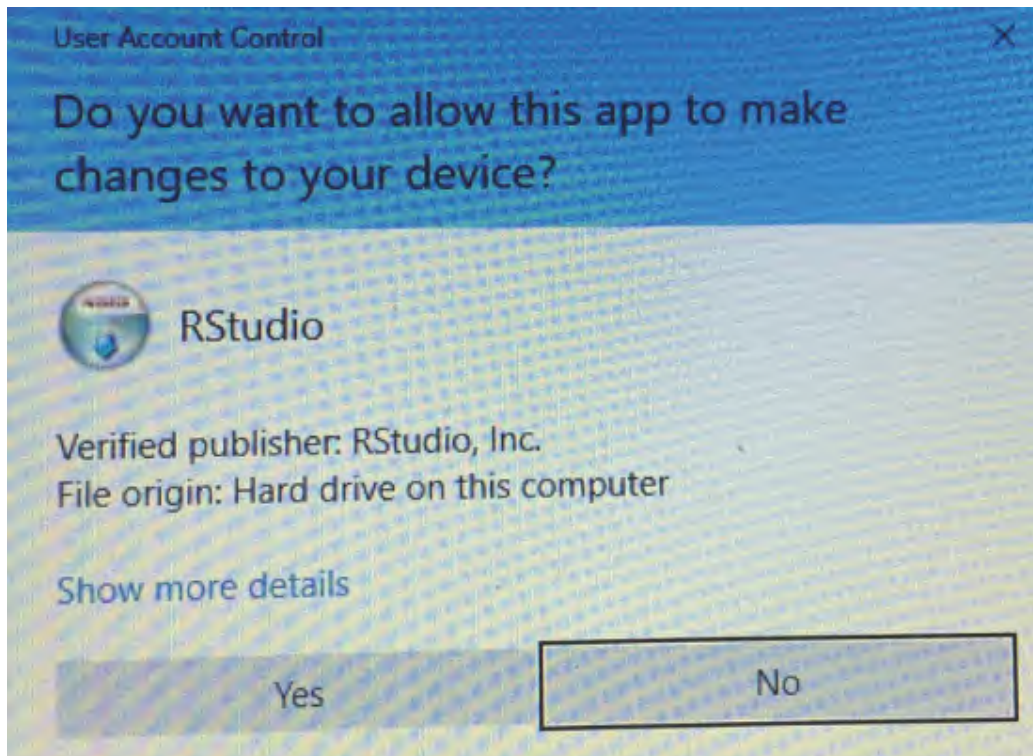


การติดตั้งโปรแกรม RStudio มีขั้นตอนดังนี้

ขั้นที่ 1

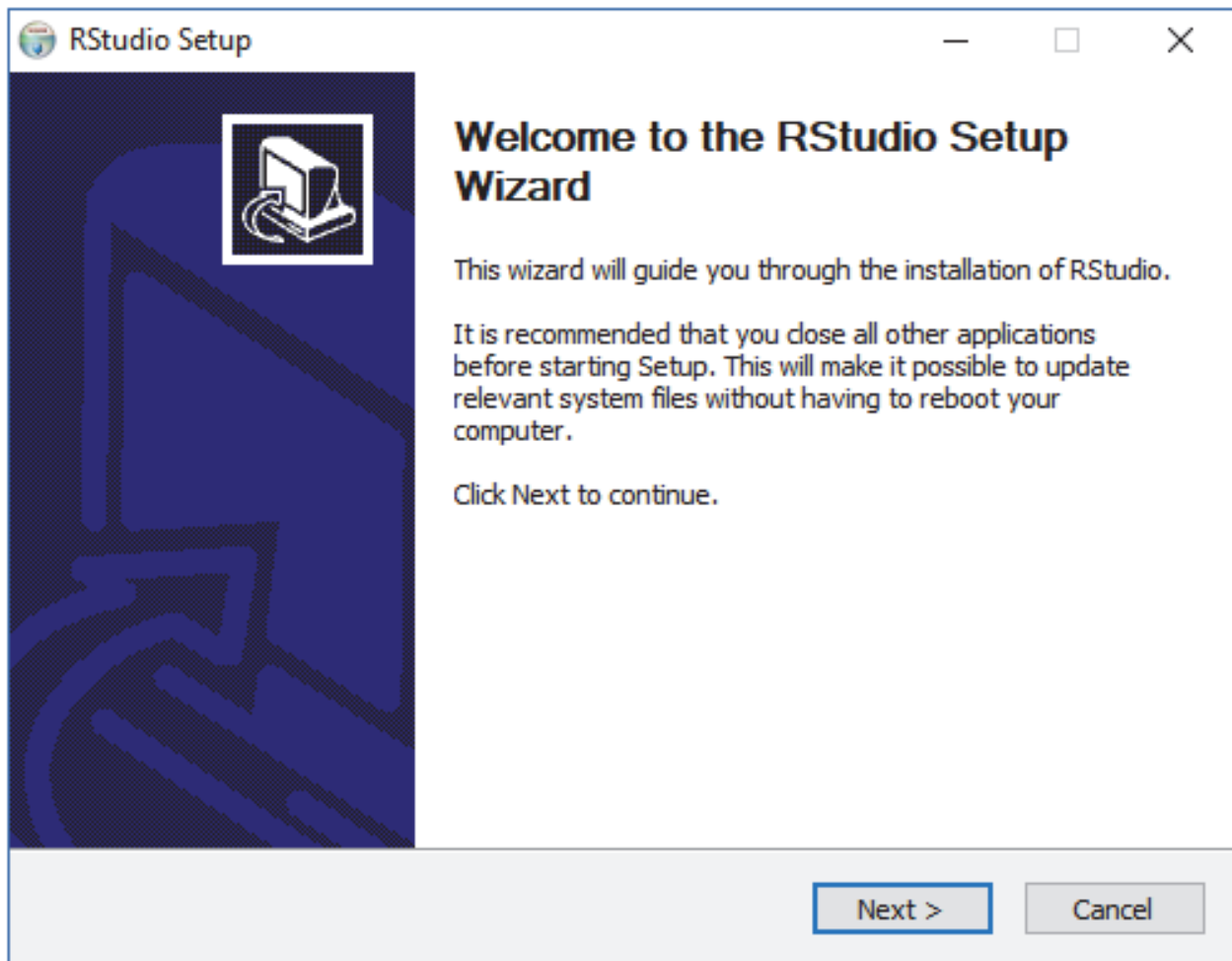
1. ให้ไปที่ Directory ชื่อ Downloads
2. สั่งให้โปรแกรม RStudio 1.1.383 ทำงาน (กดดับเบิลคลิกที่ชื่อโปรแกรม RStudio 1.1.383)

ขั้นที่ 2 บนจอภาพจะมีคำถามว่าต้องการติดตั้งโปรแกรมใช้หรือไม่



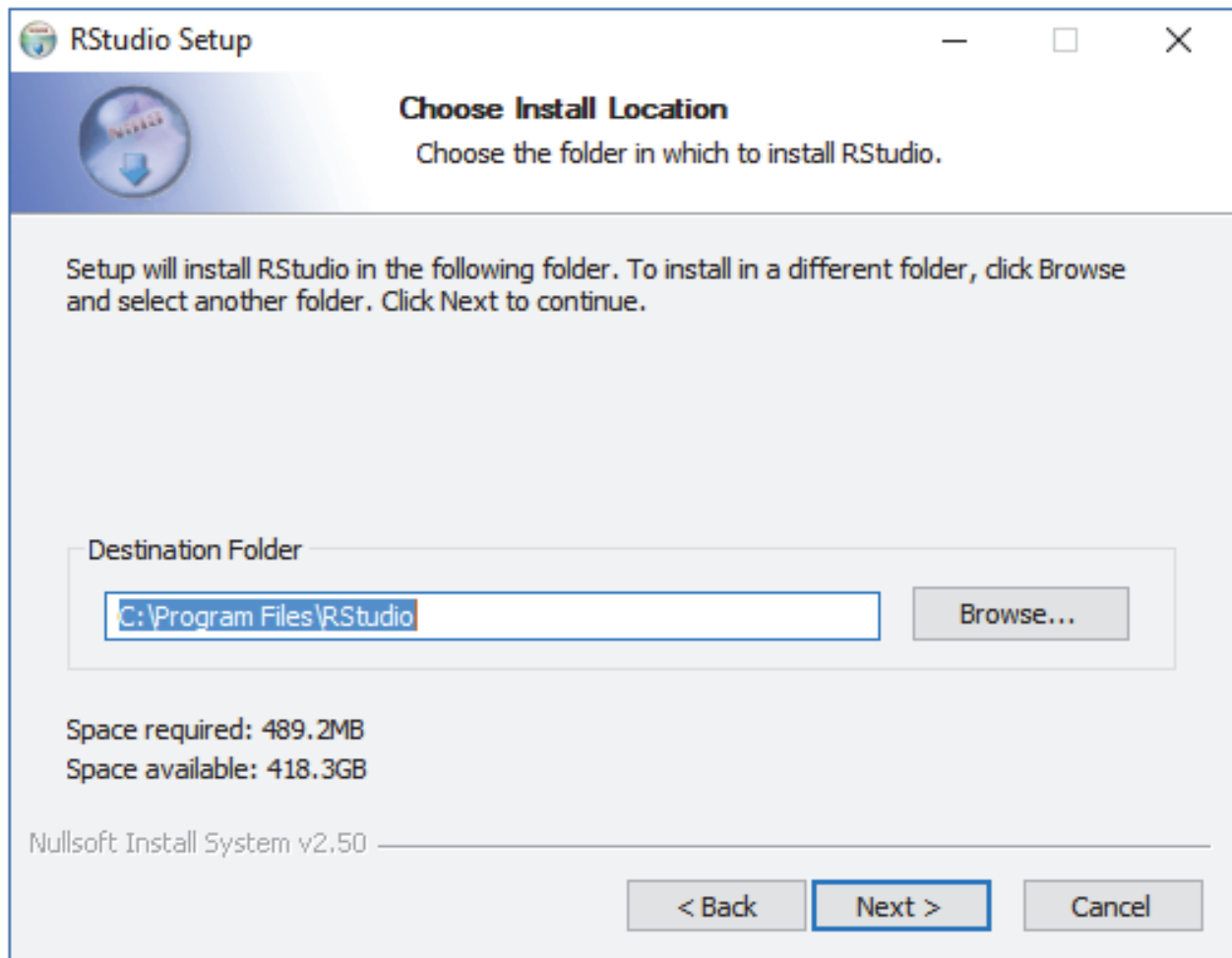
ให้ตอบ Yes เพื่อทำการติดตั้งโปรแกรมต่อไป

ขั้นที่ 3 บนจอภาพจะมีข้อความยินดีต้อนรับพร้อมที่จะติดตั้งโปรแกรม RStudio



ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม RStudio ขั้นต่อไป

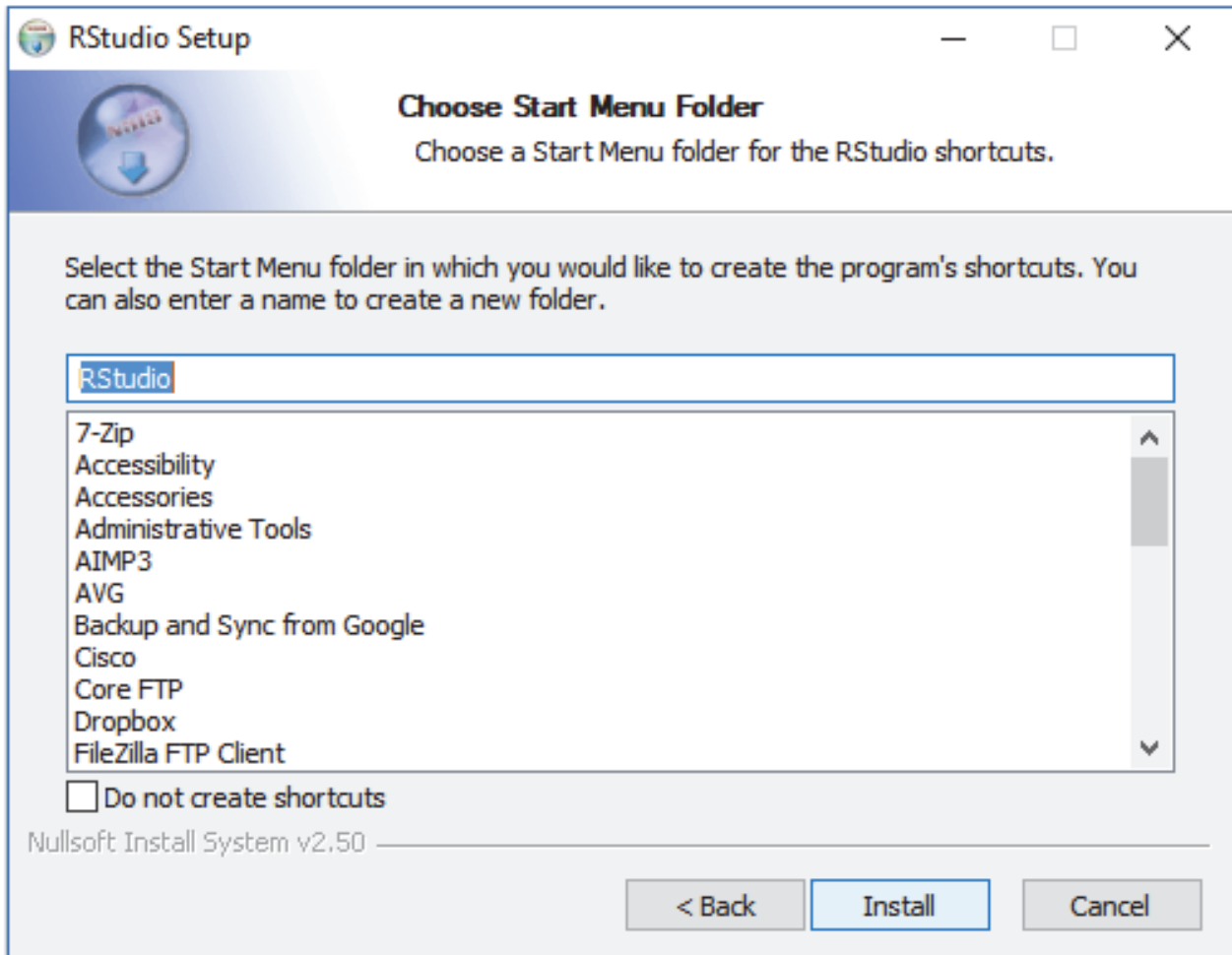
ขั้นที่ 4 บนจอภาพจะมีเมนูย่อย Choose Install Location ให้เราเลือก Directory ที่จะติดตั้งโปรแกรม RStudio ในที่นี้เราเลือกใช้ตามคำแนะนำที่พิมพ์บนจอภาพ



ให้คลิก Next เพื่อเข้าสู่การติดตั้งโปรแกรม RStudio ขั้นต่อไป

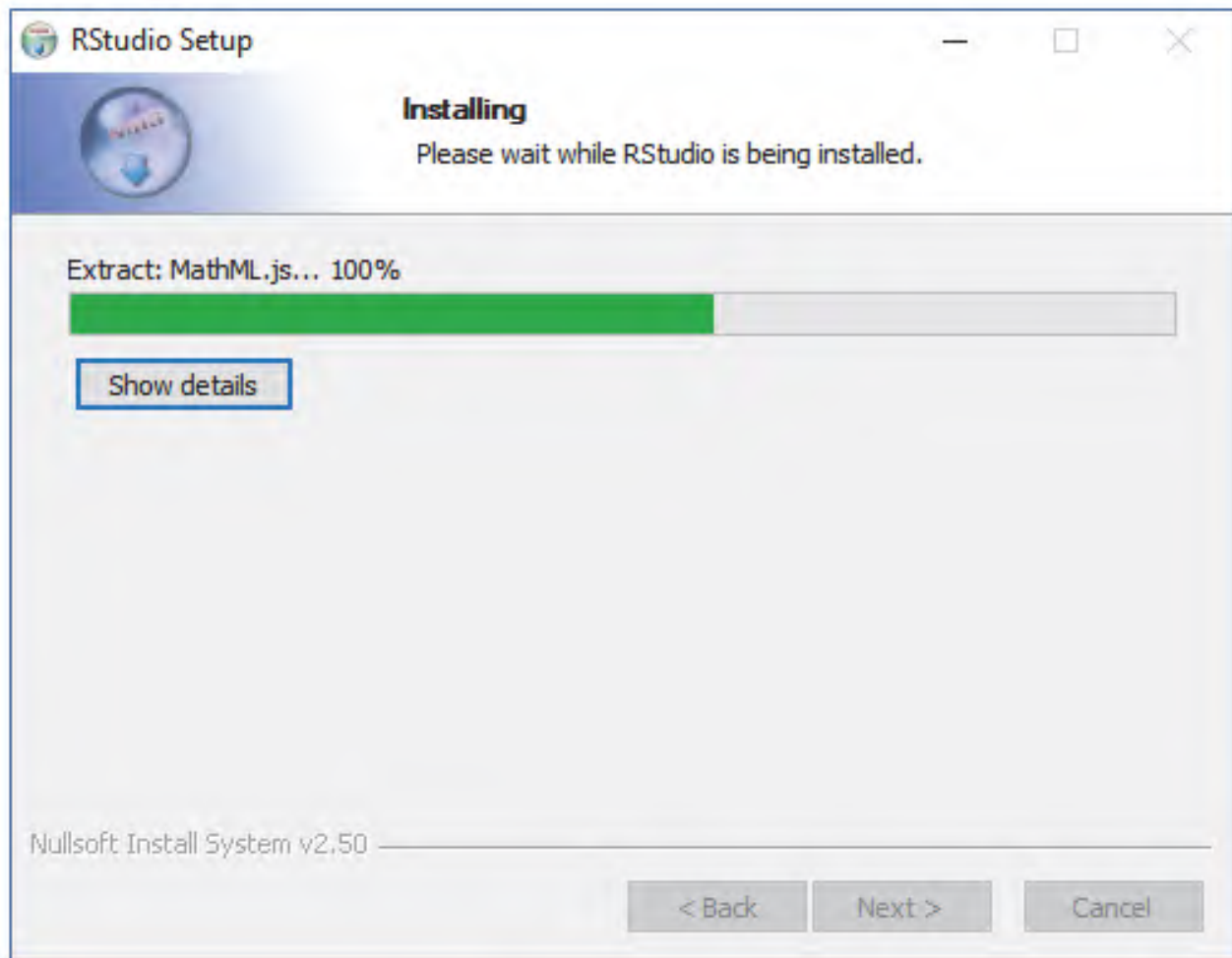
ขั้นที่ 5 บนจอภาพจะมีเมนูย่อย Choose Start Menu Folder

在这里我们选择按照程序的指引选择 Folder 名为 RStudio



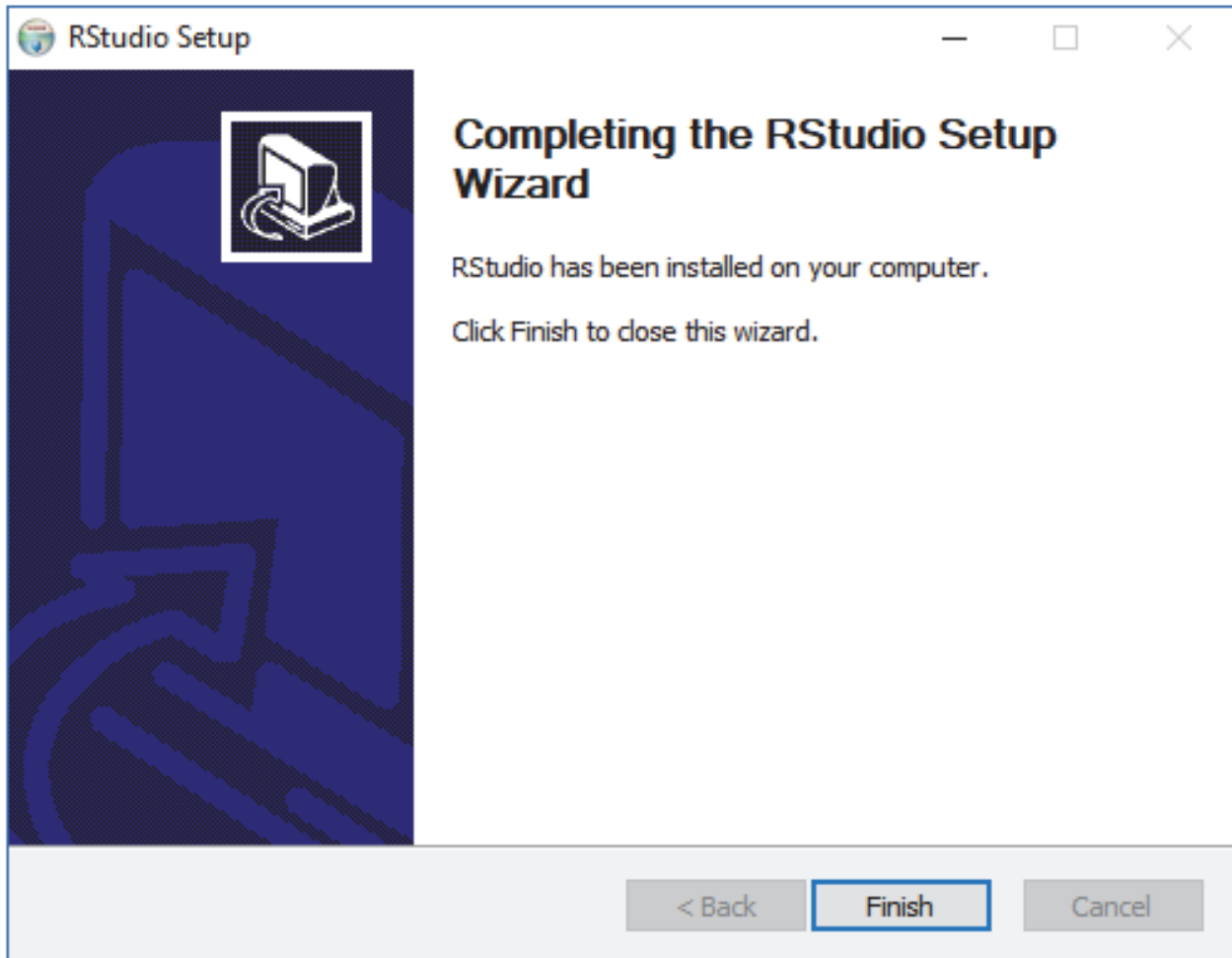
ให้คลิก Install เพื่อเข้าสู่การติดตั้งโปรแกรม RStudio ขั้นตอนต่อไป

ขั้นที่ 6 โปรแกรม RStudio Setup จะทำการติดตั้งโปรแกรม RStudio ให้



ต้องรอสักครู่จนติดตั้งโปรแกรม RStudio เสร็จ

ขั้นที่ 7 เมื่อติดตั้งโปรแกรม RStudio เสร็จ บนจอภาพจะบอกว่าการติดตั้งเสร็จสมบูรณ์



ให้คลิก Finish

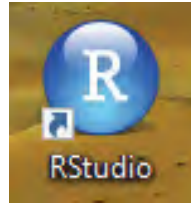
ขณะนี้การติดตั้งโปรแกรม RStudio เสร็จสมบูรณ์ พร้อมใช้งานโปรแกรม RStudio แล้ว
ขั้นตอนต่อไป ให้ไปที่จอภาพ Desktop จะพบ icon ของโปรแกรม RStudio เป็นดังนี้



หมายเหตุ หากไม่พบ Icon ให้สร้างเองขึ้นมาใหม่ หรือ Run โปรแกรม RStudio ผ่านเมนู Application ของ Window

การเรียกโปรแกรม RStudio ขึ้นมาใช้งาน

ให้ไปที่จอภาพ Desktop จะพบ icon

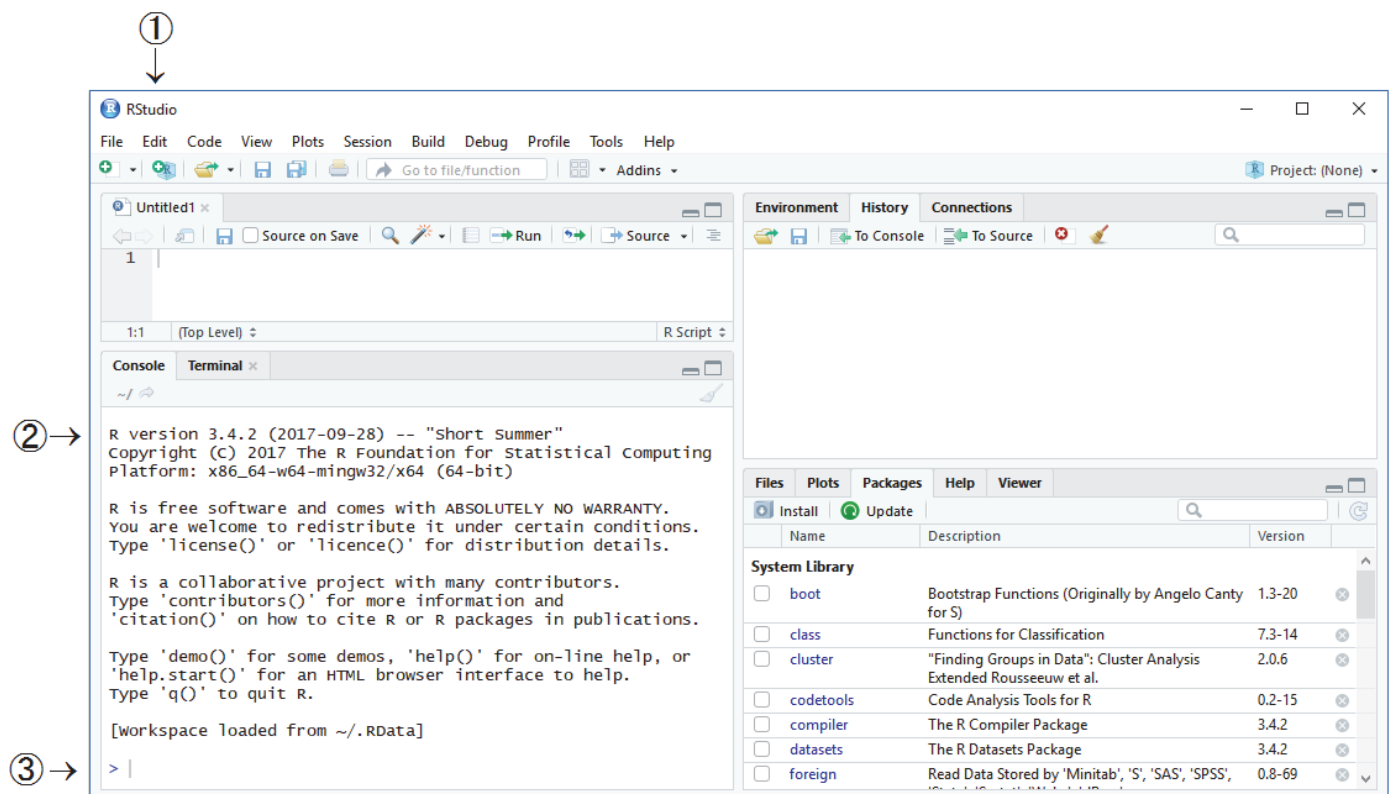


ให้ดับเบิลคลิกที่ icon



เพื่อเรียกโปรแกรม RStudio ขึ้นมาใช้งาน

จอภาพแรกเมื่อเข้าสู่การใช้งานของโปรแกรม RStudio เป็นดังนี้



1. ชื่อโปรแกรม RStudio
2. ระบุว่า เป็น R version 3.4.2 (2017-09-28)
3. ตำแหน่งเริ่มต้น (Prompt) ของการพิมพ์คำสั่งต่าง ๆ ในการใช้โปรแกรม RStudio
4. ออกจากโปรแกรม RStudio ให้พิมพ์ q() แล้วกด Enter

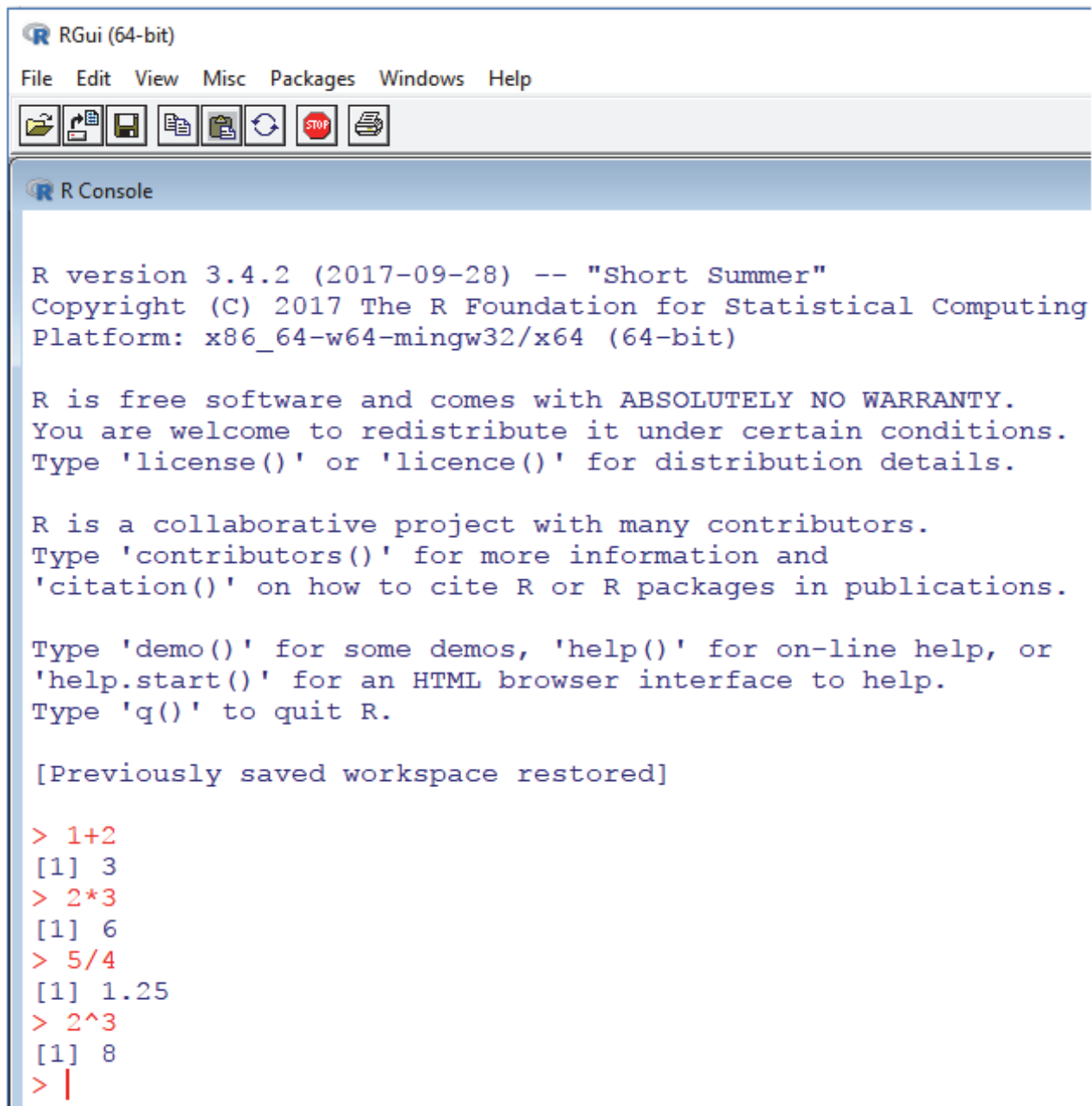
1.3 ความสามารถเบื้องต้นของโปรแกรม R

การทำงานของโปรแกรม R มีบรรทัดของคำสั่งที่เราต้องพิมพ์คำสั่งของการคำนวณเข้าไปตรงตำแหน่งที่มีสัญลักษณ์ `>` (เรียกว่า R Prompt) และแสดงผลการคำนวณในบรรทัดต่อไป

การใช้งานโปรแกรม R มีให้เลือกได้หลายแบบเช่น

1. ใช้โปรแกรม R ผ่านทาง RGui
2. ใช้โปรแกรม R ผ่านทาง RStudio
3. ใช้โปรแกรม R ผ่านทาง RGui และ R Commander

ตัวอย่างต่อไปนี้เป็นการทำงานบน Window R Console ของ RGui (บน RStudio สามารถคำนวณและแสดงผลได้เหมือนกัน)



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> 1+2
[1] 3
> 2*3
[1] 6
> 5/4
[1] 1.25
> 2^3
[1] 8
> |
```


1.3.1 การคำนวณเบื้องต้น บวก ลบ คูณ หาร และเลขยกกำลัง

```
> 12.5+7.25
[1] 19.75
> 12.5-7.25
[1] 5.25
> 12.5*4
[1] 50
> 17/5
[1] 3.4
> 4^3
[1] 64
> sqrt(2)
[1] 1.414214
```

หมายเหตุ โปรแกรม R สามารถคำนวณค่าของจำนวนเชิงซ้อนได้

```
> (3+3i)+(4-5i)
[1] 7-2i
> 4*(3+4i)
[1] 12+16i
> (3+4i)*(3-4i)
[1] 25+0i
> sqrt(-1+0i)
[1] 0+1i
> sqrt(3+4i)
[1] 2+1i
```

คำสั่งแนะนำ `getwd()` เป็นคำสั่งให้ Working directory ที่ทำงานกับโปรแกรม R ขณะนั้น
ตัวอย่างเช่น

```
> getwd()
[1] "C:/Rtemp"
```

แสดงว่า Working directory ที่ทำงานกับโปรแกรม R ขณะนั้นคือ C:/Rtemp

1.3.2 มีฟังก์ชันตรีโกณมิติ ฟังก์ชันลอการิทึม และฟังก์ชันเอกซ์โพเนนเชียล

```
> pi
[1] 3.141593
> sin(pi/4)
[1] 0.7071068
> sin(45)
[1] 0.8509035
> sin(pi/2)
[1] 1
> cos(pi/4)
[1] 0.7071068
> tan(pi/6)
[1] 0.5773503
> log(10)
[1] 2.302585
> log(10,base=10)
[1] 1
> log(10,10)
[1] 1
> log(100,10)
[1] 2
> log(8,2)
[1] 3
> exp(1)
[1] 2.718282
> exp(-1)
[1] 0.3678794
```

หมายเหตุ 1. หน่วยของมุมที่ใช้กับฟังก์ชันตรีโกณมิติ sin, cos, tan มีหน่วยเป็นเรเดียน

2. $e = 2.718282$

3. $\log(x)$ เป็นค่าลอการิทึมของ x ฐาน e

4. $\log(x, k)$ หรือ $\log(x, \text{base} = k)$ เป็นค่าลอการิทึมของ x ฐาน k

1.3.3 มีความสามารถในการกำหนดค่าให้กับตัวแปร

```
> x=12
> x+5
[1] 17
> y=25
> x+y
[1] 37
> x/y
[1] 0.48
> x^2+2*x+4
[1] 172
> x^2+y^2
[1] 769
```

ตัวอย่างการเขียนโปรแกรมภาษา R เพื่อหาพื้นที่สามเหลี่ยมที่มีความยาวด้านเป็น a, b และ c

```
> a=3
> b=4
> c=5
> s=(a+b+c)/2
> Area=sqrt(s*(s-a)*(s-b)*(s-c))
> Area
[1] 6
```

คำสั่งแนะนำ setwd เป็นคำสั่งกำหนด Working directory ที่ต้องการทำงานร่วมกับโปรแกรม R
ขณะนั้น ตัวอย่างเช่น

```
> setwd("C:/Windows")
> getwd()
[1] "C:/Windows"
> setwd("C:/Rtemp")
> getwd()
[1] "C:/Rtemp"
```

1.3.4 มีความสามารถในการกำหนดข้อมูลให้กับตัวแปร

```
> x=c(2,3,5,7,12)
> x
[1] 2 3 5 7 12
> y<-c(15,22,37,43,56,74)
> y
[1] 15 22 37 43 56 74
> w=c(x,y)
> w
[1] 2 3 5 7 12 15 22 37 43 56 74
> mean(x)
[1] 5.8
> mean(y)
[1] 41.16667
> sd(x)
[1] 3.962323
> sd(y)
[1] 21.77537
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.0     3.0     5.0     5.8     7.0    12.0
```

หมายเหตุ 1. คำสั่ง $x = c(\dots)$ หรือ $x \leftarrow c(\dots)$ เป็นคำสั่งกำหนดข้อมูลให้กับตัวแปร x

ตัวแปร x มีข้อมูล 5 ตัว คือ 2, 3, 5, 7, 12

ตัวแปร y มีข้อมูล 6 ตัว คือ 15, 22, 37, 43, 56, 74

2. ตัวแปร w มีข้อมูล 11 ตัว ได้จากการรวมข้อมูลในตัวแปร x และ y

3. mean เป็นคำสั่งหาค่าเฉลี่ย และ sd เป็นคำสั่งหาส่วนเบี่ยงเบนมาตรฐานตัวอย่าง

4. summary เป็นคำสั่งหาค่าสถิติเบื้องต้น ซึ่งจะได้ค่าต่าง ๆ ดังนี้

ค่าต่ำสุด ควอไทล์ที่ 1 มัธยฐาน ค่าเฉลี่ย ควอไทล์ที่ 3 ค่าสูงสุด

1.3.5 มีความสามารถในการกำหนดข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพให้กับตัวแปร

```
> status=c("Single","Married","Divorced","Widow")
> status
[1] "Single"    "Married"    "Divorced"    "Widow"
> age=c(25,32,26,41,37,49)
> age
[1] 25 32 26 41 37 49
> summary(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  25.0   27.5   34.5   35.0   40.0   49.0
```

หมายเหตุ 1. ตัวแปร status เป็นตัวแปรเก็บข้อมูลเชิงคุณภาพ

2. ตัวแปร age เป็นตัวแปรเก็บข้อมูลเชิงปริมาณ

3. สามารถนำข้อมูลหลายตัวแปรมารวมกันเป็นกลุ่มของข้อมูลได้

```
> no=c(1,2,3,4,5)
> score1=c(12,15,22,17,14)
> score2=c(8,7,5,9,4)
> no
[1] 1 2 3 4 5
> score1
[1] 12 15 22 17 14
> score2
[1] 8 7 5 9 4
> data.frame(no,score1,score2)
  no score1 score2
1  1      12      8
2  2      15      7
3  3      22      5
4  4      17      9
5  5      14      4
```

1.3.6 สามารถกำหนดข้อมูล คำนวณหาผลบวกของข้อมูลในตัวแปรได้

```

> x=c(2,5,7,8,9)
> y=c(3,5,4,7,8)
> x
[1] 2 5 7 8 9
> y
[1] 3 5 4 7 8
> sum(x)
[1] 31
> sum(y)
[1] 27
> x^2
[1] 4 25 49 64 81
> sum(x^2)
[1] 223
> y^2
[1] 9 25 16 49 64
> sum(y^2)
[1] 163
> x*y
[1] 6 25 28 56 72
> sum(x*y)
[1] 187

```

ตัวอย่างการคำนวณค่าเฉลี่ยและความแปรปรวนของ x

```

> sum(x)/5
[1] 6.2
> mean(x)
[1] 6.2
> sum((x-mean(x))^2)/4
[1] 7.7
> var(x)
[1] 7.7

```

หมายเหตุ $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ และ $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

1.3.7 การคำนวณค่าสถิติเบื้องต้นของข้อมูล

```
> x=c(3,7,2,5,12)
> x
[1] 3 7 2 5 12
> min(x)
[1] 2
> max(x)
[1] 12
> range(x)
[1] 2 12
> sum(x)
[1] 29
> x^2
[1] 9 49 4 25 144
> sum(x^2)
[1] 231
> mean(x)
[1] 5.8
> sum(x)/5
[1] 5.8
> sd(x)
[1] 3.962323
> var(x)
[1] 15.7
> sum((x-mean(x))^2)/4
[1] 15.7
> y=sort(x)
> y
[1] 2 3 5 7 12
```

หมายเหตุ 1. min, max เป็นคำสั่งหาค่าต่ำสุดและค่าสูงสุดของข้อมูล

2. range เป็นคำสั่งหาค่าพิสัยของข้อมูล

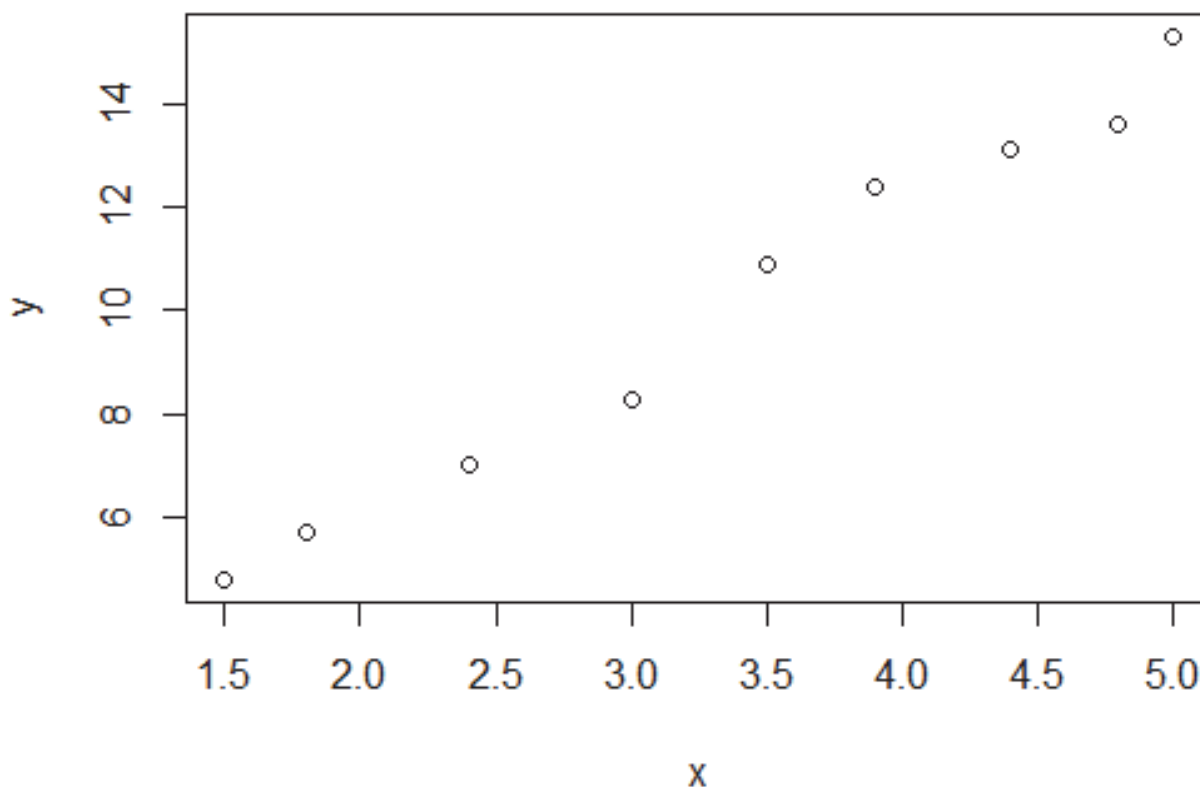
3. sort เป็นคำสั่งเรียงลำดับข้อมูล

4. var(x) คำนวณมาจากสูตรความแปรปรวนตัวอย่าง $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

1.3.8 สามารถเขียนกราฟแผนภาพการกระจาย

```
> x=c(1.5,1.8,2.4,3.0,3.5,3.9,4.4,4.8,5.0)
> y=c(4.8,5.7,7.0,8.3,10.9,12.4,13.1,13.6,15.3)
> x
[1] 1.5 1.8 2.4 3.0 3.5 3.9 4.4 4.8 5.0
> y
[1] 4.8 5.7 7.0 8.3 10.9 12.4 13.1 13.6 15.3
> plot(x,y)
```

ได้แผนภาพการกระจายเป็นดังนี้



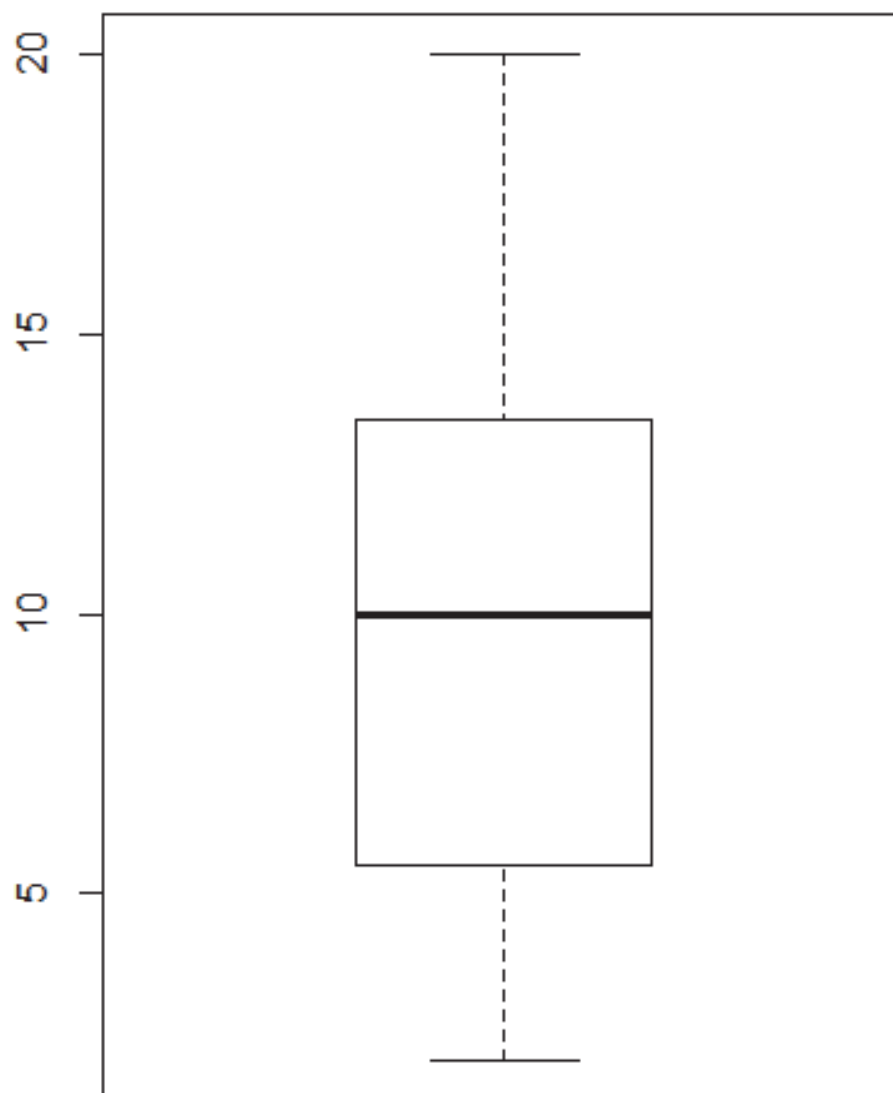
คำสั่งแนะนำ `options(scipen = k)` เป็นคำสั่งกำหนดการแสดงผลตัวเลขที่มีค่ามากเกินไปหรือน้อยเกินไปจากรูปแบบ scientific notation ให้แสดงผลเป็นเลขทศนิยมปกติ ตัวอย่างเช่น

```
> 12.25^(-8.74)
[1] 3.088095e-10
> options(scipen=15)
> 12.25^(-8.74)
[1] 0.0000000003088095
```


1.3.9 สามารถเขียนกราฟในรูปแบบ boxplot

```
> x=c(2,4,7,9,11,12,15,20)
> x
[1]  2  4  7  9 11 12 15 20
> boxplot(x)
> summary(x)
      Min. 1st Qu.  Median    Mean
      2.00   6.25   10.00   10.00
3rd Qu.    Max.
 12.75   20.00
```

กราฟ boxplot ที่ได้เป็นดังนี้



1.3.10 สามารถคำนวณค่าของการแจกแจงปกติและการแจกแจงปกติมาตรฐาน

```

> pnorm(1.96)
[1] 0.9750021
> pnorm(1.96, mean=0, sd=1)
[1] 0.9750021
> pnorm(5, mean=3, sd=2)
[1] 0.8413447
> qnorm(0.975)
[1] 1.959964
> qnorm(0.975, mean=0, sd=1)
[1] 1.959964
> qnorm(0.8413447, mean=3, sd=2)
[1] 5
> pnorm(1.96) - pnorm(1.645)
[1] 0.02498701

```

หมายเหตุ

1. $\text{pnorm}(k, \text{mean} = 0, \text{sd} = 1)$ หรือ $\text{pnorm}(k)$ มีค่าเท่ากับ $P(-\infty < Z < k)$
2. $\text{qnorm}(A)$ หรือ $\text{qnorm}(A, \text{mean} = 0, \text{sd} = 1)$ คือจำนวนจริง k ที่ทำให้ $P(-\infty < Z < k) = A$
3. $\text{pnorm}(k, \text{mean} = \mu, \text{sd} = \sigma) = P(-\infty < X < k)$
ของตัวแปรสุ่มปกติ X ที่มีค่าเฉลี่ย μ ส่วนเบี่ยงเบนมาตรฐาน σ
4. $\text{qnorm}(A, \text{mean} = \mu, \text{sd} = \sigma)$ คือจำนวนจริง k ที่ทำให้ $P(-\infty < X < k) = A$
ของตัวแปรสุ่มปกติ X ที่มีค่าเฉลี่ย μ ส่วนเบี่ยงเบนมาตรฐาน σ
5. $P(a < Z < b) = \text{pnorm}(b) - \text{pnorm}(a)$
6. $\text{pnorm}(k, \text{mean} = \mu, \text{sd} = \sigma)$ และ $\text{pnorm}(k, \mu, \sigma)$ เหมือนกัน
7. $\text{pnorm}(k)$ และ $\text{pnorm}(k, 0, 1)$ และ $\text{pnorm}(k, \text{mean} = 0, \text{sd} = 1)$ เหมือนกัน
8. $\text{qnorm}(A)$ และ $\text{qnorm}(A, 0, 1)$ และ $\text{qnorm}(A, \text{mean} = 0, \text{sd} = 1)$ เหมือนกัน

1.3.11 สามารถคำนวณค่าของการแจกแจงทวินาม

```
> dbinom(2,size=5,prob=0.25)
[1] 0.2636719
> pbinom(2,size=5,prob=0.25)
[1] 0.8964844
> x=seq(0,5,by=1)
> Prob.x=dbinom(x,10,0.25)
> Cum.Prob.x=pbinom(x,10,0.25)
> table=data.frame(x,Prob.x,Cum.Prob.x)
> table
```

	x	Prob.x	Cum.Prob.x
1	0	0.05631351	0.05631351
2	1	0.18771172	0.24402523
3	2	0.28156757	0.52559280
4	3	0.25028229	0.77587509
5	4	0.14599800	0.92187309
6	5	0.05839920	0.98027229

หมายเหตุ ให้ $X = 0, 1, 2, \dots, n$ เป็นตัวแปรสุ่มทวินาม

1. $P(X = x) = b(x, n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$
2. $\text{dbinom}(x, \text{size} = n, \text{prob} = p) = \binom{n}{x} p^x (1 - p)^{n-x} = P(X = x)$
3. $\text{pbinom}(k, \text{size} = n, \text{prob} = p) = \sum_{x=0}^k \binom{n}{x} p^x (1 - p)^{n-x} = P(X \leq k)$
4. $\text{dbinom}(x, \text{size} = n, \text{prob} = p)$ และ $\text{dbinom}(x, n, p)$ เหมือนกัน
5. $\text{pbinom}(p, \text{size} = n, \text{prob} = p)$ และ $\text{pbinom}(k, n, p)$ เหมือนกัน

คำสั่งแนะนำ `options(digits = k)` เป็นคำสั่งกำหนดการแสดงผลตัวเลข k ตัว ตัวอย่างเช่น

```
> 22/7
[1] 3.142857
> options(digits=12)
> 22/7
[1] 3.14285714286
> options(digits=4)
> 22/7
[1] 3.143
```

1.3.12 สามารถหาช่วงความเชื่อมั่นของค่าพารามิเตอร์

ตัวอย่างเช่นการหาช่วงความเชื่อมั่นของค่าเฉลี่ยจากตัวอย่างที่สุ่มได้ 2, 3, 5, 7, 12

```
> x=c(2,3,5,7,12)
> x
[1] 2 3 5 7 12
> mean(x)
[1] 5.8
> sd(x)
[1] 3.962323
> t.test(x)
```

One Sample t-test

```
data: x
t = 3.2731, df = 4, p-value = 0.0307
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.8801267 10.7198733
sample estimates:
mean of x
 5.8

> lower=mean(x)-qt(0.975,4)*(sd(x)/(sqrt(5)));lower
[1] 0.8801267
> upper=mean(x)+qt(0.975,4)*(sd(x)/(sqrt(5)));upper
[1] 10.71987
```

ช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยคือ (0.8801267, 10.71987)

หมายเหตุ สูตรที่ใช้ในการหาช่วงความเชื่อมั่น 95% ของค่าเฉลี่ยคือ

$$\bar{x} - t_{0.025, 4} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.025, 4} \frac{s}{\sqrt{n}}$$

1.3.13 สามารถนิยามสูตรฟังก์ชันใช้เองได้

การนิยามสูตร $f(x) = x^2 + 2x + 4$ และหาค่า $f(0)$, $f(1)$

```
> f=function(x) x^2+2*x+4
> f(0)
[1] 4
> f(1)
[1] 7
```

การนิยามสูตร $f(x, y) = x^2 + 2y^2$ และหาค่า $f(1, 2)$, $f(2, 3)$

```
> f=function(x,y) x^2+2*y^2
> f(1,2)
[1] 9
> f(2,3)
[1] 22
```

การนิยามสูตรสำหรับหาค่าความน่าจะเป็น $P(a < Z < b)$

```
> P=function(a,b) {return(pnorm(b)-pnorm(a))}
> P(0,1.96)
[1] 0.4750021
> P(-1.96,0)
[1] 0.4750021
```

หมายเหตุ การกำหนดสูตร ชื่อฟังก์ชัน = function(ตัวแปร) สูตรในเทอมของตัวแปร

หรือ ชื่อฟังก์ชัน = function(ตัวแปร) { return(สูตรในเทอมของตัวแปร) }

ตัวอย่างเช่น

$f = \text{function}(x) \ x^2 + 2x + 4$ หมายถึงการกำหนดสูตร $f(x) = x^2 + 2x + 4$

$f(0)$ และ $f(1)$ เป็นการสั่งให้คำนวณค่าฟังก์ชัน

$f = \text{function}(x, y) \ x^2 + 2y^2$ หมายถึงการกำหนดสูตร $f(x, y) = x^2 + 2y^2$

$f(1, 2)$ และ $f(2, 3)$ เป็นการสั่งให้คำนวณค่าฟังก์ชัน

1.3.14 สามารถหาความสัมพันธ์เชิงเส้นตรงของข้อมูล

```
> x=c(4.3,4.5,5.9,5.6,6.1,5.2,3.8,2.1,7.5)
> y=c(126,121,116,118,114,118,132,141,108)
> model.linear=lm(y~x)
> summary(model.linear)
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7175	-0.5992	0.1360	1.1049	2.8557

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	153.1755	2.6147	58.58	1.11e-10 ***
x	-6.3240	0.5019	-12.60	4.58e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.203 on 7 degrees of freedom

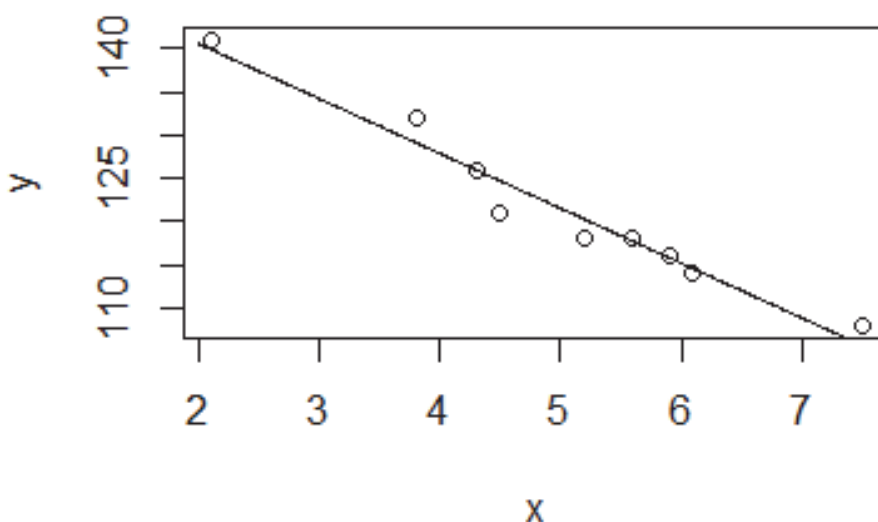
Multiple R-squared: 0.9578, Adjusted R-squared: 0.9517

F-statistic: 158.8 on 1 and 7 DF, p-value: 4.579e-06

ผลการวิเคราะห์จะได้ $\hat{y} = 153.1755 - 6.3240x$ และสัมประสิทธิ์การกำหนด = 0.9578

กราฟของแผนภาพการกระจายและเส้นถดถอย

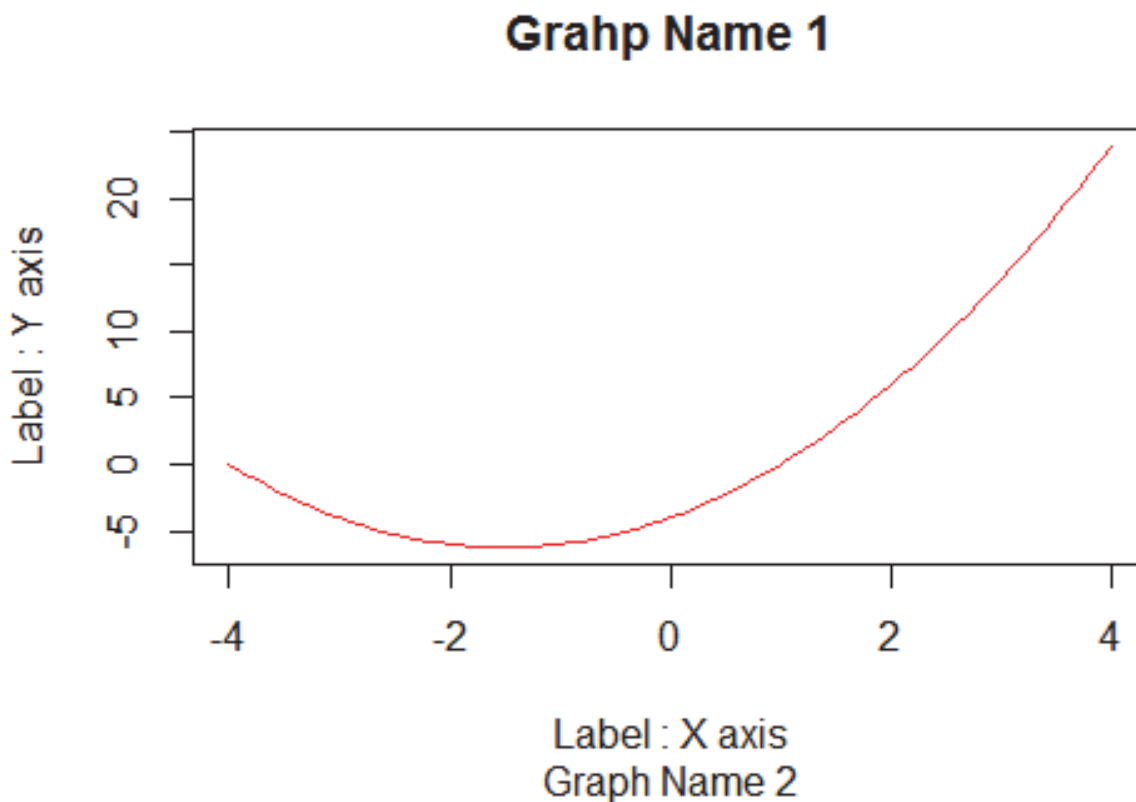
```
> ycap=function(t) 153.1755-6.3240*t
> t=seq(2,8,by=0.01)
> plot(x,y)
> lines(t,ycap(t))
```



1.3.15 สามารถเขียนกราฟได้หลายรูปแบบ

1.3.15.1 กราฟของเส้นโค้ง 1 เส้น

```
> f=function(x) x^2+3*x-4
> x=seq(-4,4,by=0.1)
> plot(x,f(x),type="l",
+       main="Grahp Name 1",
+       sub="Graph Name 2",
+       xlab="Label : X axis",
+       ylab="Label : Y axis",
+       col = "red")
```

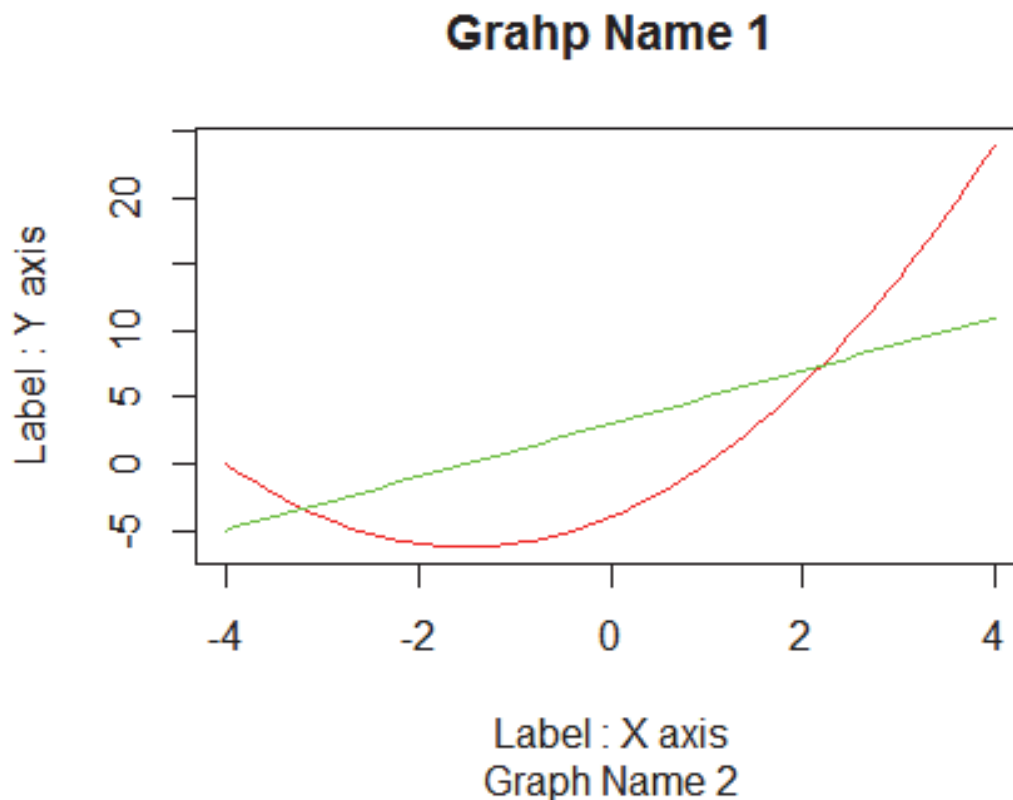


หมายเหตุ

1. การเขียนคำสั่งให้อ่านง่าย ให้ขึ้นบรรทัดใหม่ภายในคำสั่งเดียวกันให้กด Shift + Enter
2. การกด Shift + Enter บนจอภาพจะมีสัญลักษณ์ + เพื่อบอกการพิมพ์คำสั่งอื่น ๆ ต่อไป
3. การกด Enter หมายถึง สั่งให้คำนวณคำสั่งในบรรทัดนั้น

1.3.15.2 กราฟของเส้นโค้ง 2 เส้น

```
> f=function(x) x^2+3*x-4
> x=seq(-4,4,by=0.1)
> plot(x,f(x),type="l",
+       main="Grahp Name 1",
+       sub="Graph Name 2",
+       xlab="Label : X axis",
+       ylab="Label : Y axis",
+       col = "red")
> g=function(x) 2*x+3
> lines(x,g(x),type="l",col="green")
```

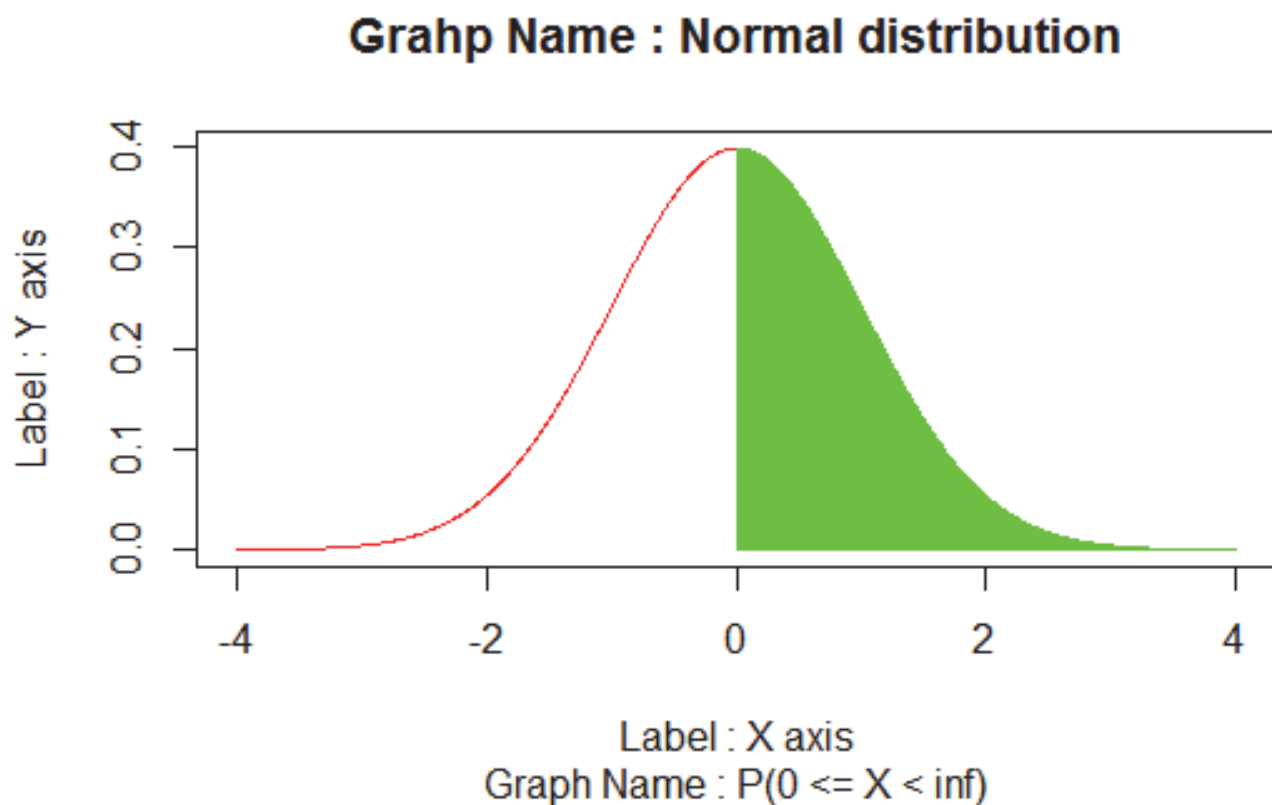


คำสั่งแนะนำ ls() เป็นคำสั่งแสดงตัวแปรหรือ object ที่มีอยู่ในขณะนั้น ตัวอย่างเช่น

```
> x=2.5
> y=3.25
> z=x*y
> z
[1] 8.125
> ls()
[1] "x" "y" "z"
```


1.3.15.3 กราฟแรเงา

```
> f=function(z) dnorm(z)
> z=seq(-4,4,by=0.01)
> plot(z,f(z),type="l",
+       main="Grahp Name : Normal distribution",
+       sub="Graph Name : P(0 <= X < inf)",
+       xlab="Label : X axis",
+       ylab="Label : Y axis",
+       col = "red")
> t=seq(0,4,by=0.01)
> g=function(t) dnorm(t)
> lines(t,g(t),type="h",col="green")
```



1.3.16 สามารถทำการวิเคราะห์ความแปรปรวน (ANOVA)

ตัวอย่างเช่น วิศวกรผู้หนึ่งสนใจว่าค่าเฉลี่ยของการดูดความชื้นของคอนกรีตอัดแรง 5 ชนิดเท่ากันหรือไม่ ได้นำตัวอย่างมาทดลองชนิดละ 6 หน่วย นาน 48 ชั่วโมง บันทึกผลการดูดความชื้นของคอนกรีตได้ค่าการดูดความชื้นของคอนกรีต (หน่วย : น้ำหนัก %) ดังนี้

ชนิดที่ 1	ชนิดที่ 2	ชนิดที่ 3	ชนิดที่ 4	ชนิดที่ 5
551	595	639	417	563
457	580	615	449	631
450	508	511	517	522
731	583	573	438	613
499	633	648	415	656
632	517	677	555	679

การกำหนดข้อมูลและสร้างตาราง ANOVA สามารถทำได้ดังนี้

```
> t1=c(1,1,1,1,1,1)
> w1=c(551,457,450,731,499,632)
> t2=c(2,2,2,2,2,2)
> w2=c(595,580,508,583,633,517)
> t3=c(3,3,3,3,3,3)
> w3=c(639,615,511,573,648,677)
> t4=c(4,4,4,4,4,4)
> w4=c(417,449,517,438,415,555)
> t5=c(5,5,5,5,5,5)
> w5=c(563,631,522,613,656,679)
> t=c(t1,t2,t3,t4,t5)
> w=c(w1,w2,w3,w4,w5)
> t=factor(t)
> report=aov(w~t)
> summary(report)
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
t           4  85356    21339     4.302 0.00875 **
Residuals  25 124020     4961

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

การแปลความหมายเราจะได้ศึกษาในบทที่ 6 การวิเคราะห์ความแปรปรวน

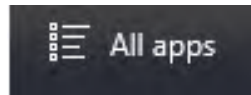
1.4 การใช้งานโปรแกรม R ผ่าน RGui

การวิเคราะห์ข้อมูลโดยใช้โปรแกรม R สามารถใช้ RGui, RStudio และ R Commander ในหัวข้อนี้ เราจะใช้ RGui มาศึกษาในส่วนของกำหนัดตัวแปร การสร้างแฟ้มข้อมูล

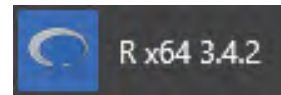
การนำโปรแกรม RGui มาใช้งานมีขั้นตอนดังนี้

ขั้นที่ 1 เปิดโปรแกรม RGui (R Graphical user interface)

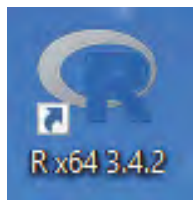
แบบที่ 1 คลิกที่



แล้วเลือก Application



แบบที่ 2 คลิกที่



ที่หน้าจอ Desktop

หมายเหตุ สัญลักษณ์ของ icon R programming มีให้ใช้งานหลังเราสั่งติดตั้งโปรแกรม R รอสักครู่จะได้ Window R Console ของ RGui (64-bit) เป็นดังนี้

```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

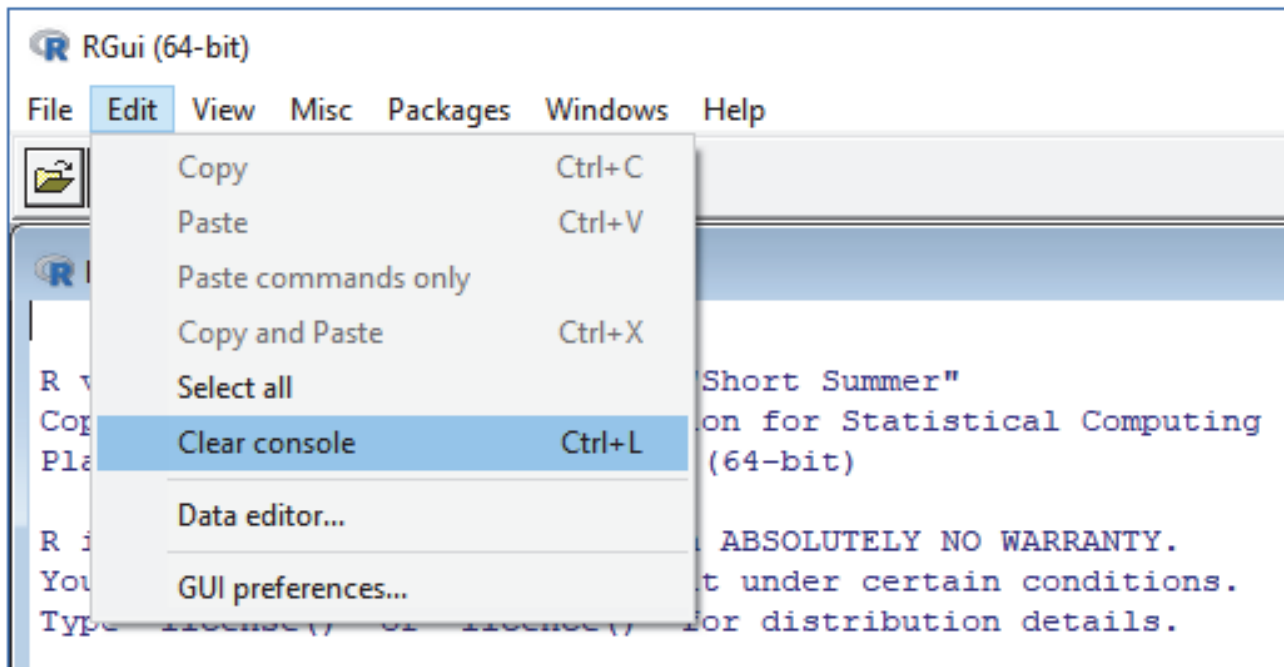
[Previously saved workspace restored]

> |
```

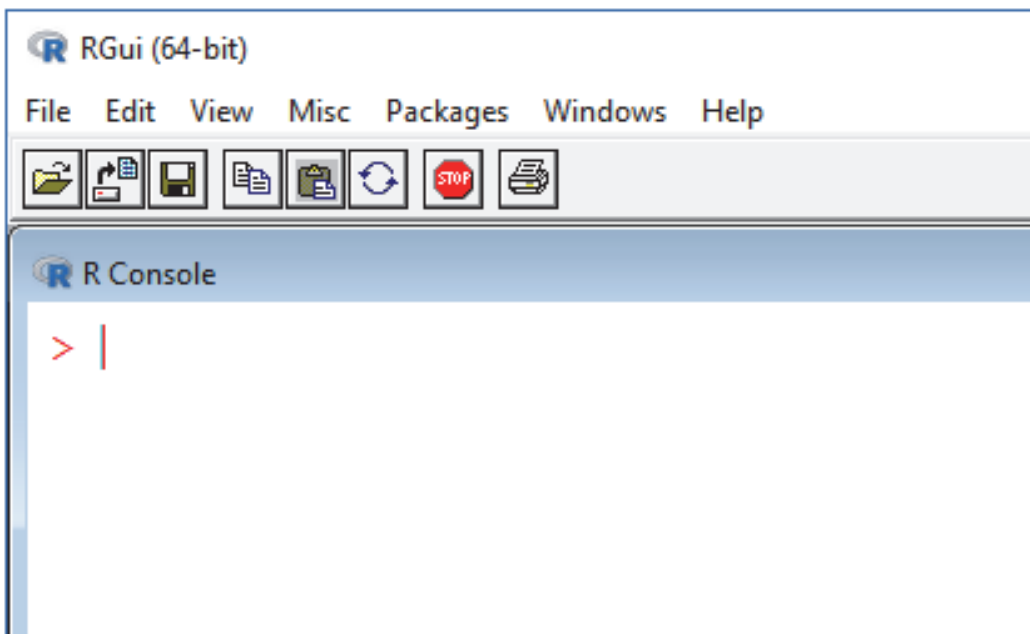
1. Window RGui (64-bit) เป็นหน้าต่างของโปรแกรม R ที่พร้อมให้ใช้งาน
2. Window R Console ของ RGui

เป็นบริเวณทำงาน กำหนดตัวแปร รับข้อมูล แสดงผล ของโปรแกรม R

คำแนะนำ ใช้เมนูคำสั่ง Edit/Clear console เพื่อลบข้อความบน Window R Console



หรือ กด Ctrl + L เพื่อลบข้อความบน Window R Console



ข้อมูลที่จะนำมาวิเคราะห์หากเป็นข้อมูลที่มี 1 ตัวแปร ให้กำหนดในรูปแบบเวกเตอร์แถว

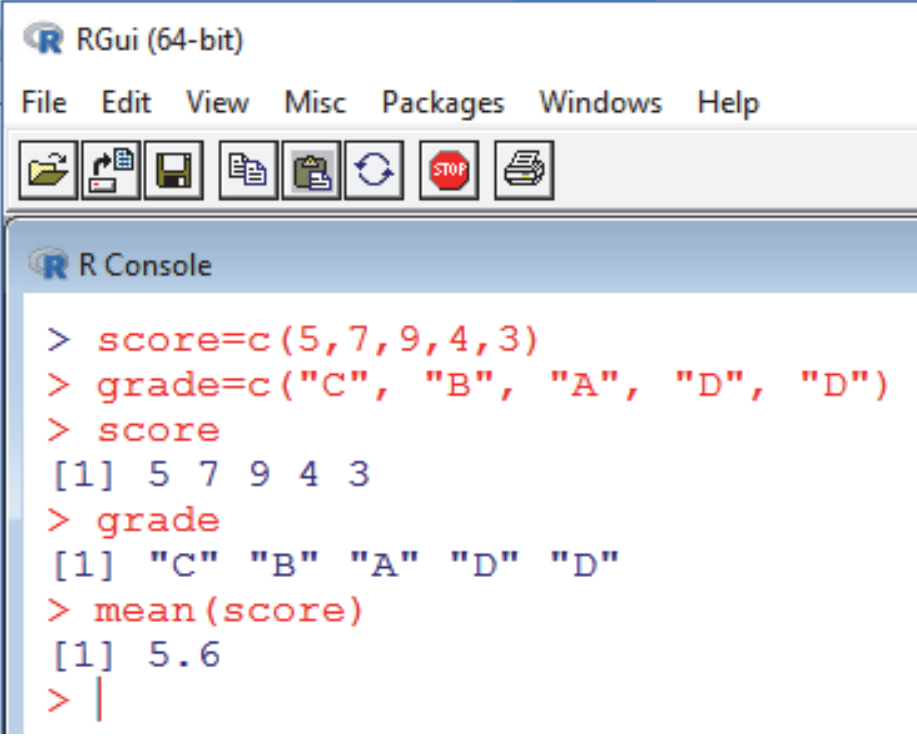
ตัวอย่างเช่น คะแนนสอบของนักเรียน 5 คนที่สอบได้ คือ 5, 7, 9, 4, 3

กำหนดเป็น `score = c(5, 7, 9, 4, 3)`

เกรดผลการเรียนของนักเรียน 5 คนที่สอบได้ คือ C, B, A, D, D

กำหนดเป็น `grade = c("C", "B", "A", "D", "D")`

ตัวอย่างการพิมพ์ใน R Console



```
> score=c(5,7,9,4,3)
> grade=c("C", "B", "A", "D", "D")
> score
[1] 5 7 9 4 3
> grade
[1] "C" "B" "A" "D" "D"
> mean(score)
[1] 5.6
> |
```

`score = c(5, 7, 9, 4, 3)`

กำหนดข้อมูล Numeric

`grade = c("C", "B", "A", "D", "D")`

กำหนดข้อมูล Character

`score`

แสดงค่าที่เก็บในตัวแปร `score`

`grade`

แสดงค่าที่เก็บในตัวแปร `grade`

`mean(score)`

หาค่าเฉลี่ยของ `score`

ลักษณะของข้อมูลทางสถิติที่นิยมใช้งานกันจะมีหลายตัวแปร เช่น ข้อมูลพนักงานของบริษัท ประกอบด้วย เลขประจำตัว (id), เพศ (sex), อายุ (age), ระดับการศึกษา (educ), สถานะภาพ (status), เงินเดือน (income), ระดับคะแนนความสามารถ (grade) และ เงินตอบแทนประจำปี (bonus)

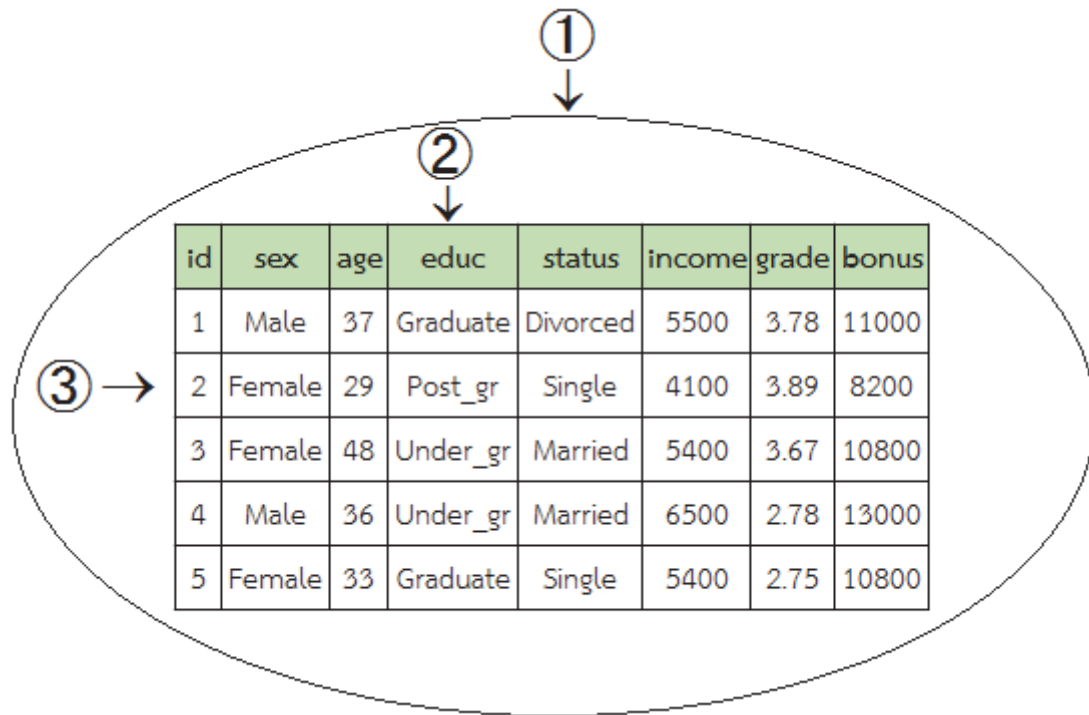
id	sex	age	educ	status	income	grade	bonus
1	Male	37	Graduate	Divorced	5500	3.78	11000
2	Female	29	Post_gr	Single	4100	3.89	8200
3	Female	48	Under_gr	Married	5400	3.67	10800
4	Male	36	Under_gr	Married	6500	2.78	13000
5	Female	33	Graduate	Single	5400	2.75	10800

ตัวแปรแบบ Numeric เป็นตัวแปรสำหรับเก็บข้อมูลเชิงปริมาณ เช่น เลขประจำตัว (id), อายุ (age), รายได้ (income), ระดับคะแนนความสามารถ (grade), เงินตอบแทนประจำปี (bonus)

ตัวแปรแบบ Character เป็นตัวแปรสำหรับเก็บข้อมูลเชิงคุณภาพ เช่น เพศ (sex), ระดับการศึกษา (educ), สถานะภาพ (status)

ตัวแปรแบบ Logical เป็นตัวแปรสำหรับเก็บข้อมูลเชิงตรรก เช่น ความชอบในสินค้า (ชอบ กับ ไม่ชอบ) สีของสินค้า (สีดำ กับ ไม่ใช่สีดำ)

ข้อมูลในลักษณะตารางที่มีหลายตัวแปร โปรแกรม R จะกำหนดโครงสร้างและการเรียกชื่อดังนี้



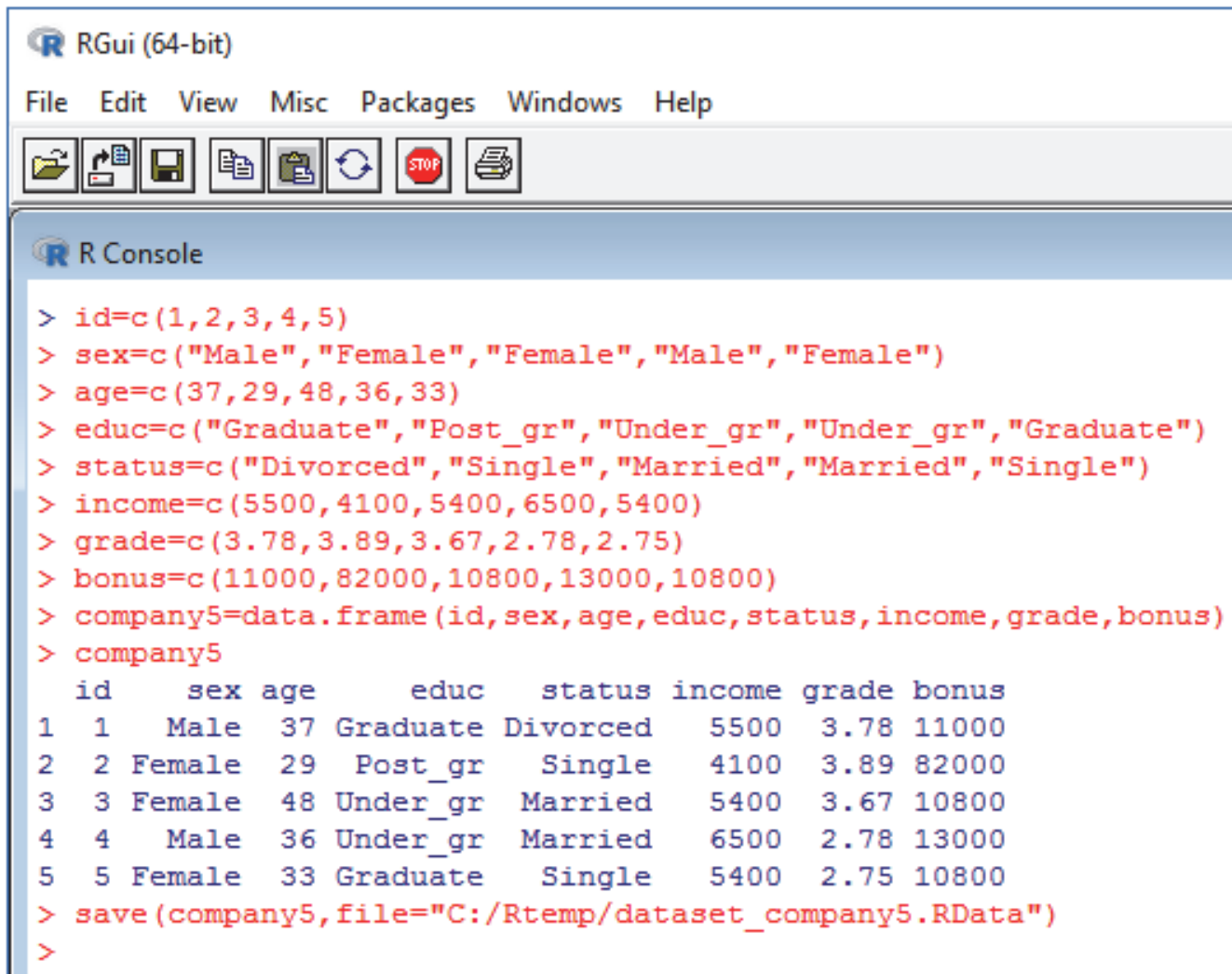
- ข้อมูลทั้งหมดในตารางนิยมเรียกว่า dataset, dataframe, datatable, object หรือ ข้อมูล
หมายเหตุ ขณะนี้จะขอเรียกว่า **dataframe** จะได้ช่วยในการจำคำสั่ง data.frame ได้หนึ่งคำสั่ง
data.frame(x, y) เป็นคำสั่งรวมตัวแปร x และ y ให้เป็นข้อมูลแบบ dataframe
- ข้อมูลในแต่ละ Column เราเรียกว่า **ตัวแปร (variabe)** หรือ **ตัวแปรหลัก (column variable)**
โดยที่ค่าในตัวแปรต้องเป็น numeric เหมือนกันทั้งหมด หรือต้องเป็น Character เหมือนกัน
ทั้งหมด
หมายเหตุ ต่อไปจะขอเรียกข้อมูลในแต่ละ Column ว่า **ตัวแปร** หรือ **variable**
- ข้อมูลในบรรทัด 1, 2, 3, ... จะเรียกว่า **ค่าสังเกต (Observation)** ตัวที่ 1, 2, 3, ... ตามลำดับ

การกำหนด dataframe ข้างต้นเราทำได้ 2 วิธีคือ

วิธีที่ 1 พิมพ์ทุกคำสั่งใน Window R Console ของ RGui แล้วนำมารวมกันเป็น dataframe โดยใช้คำสั่ง `data.frame`

ขั้นที่ 1 พิมพ์ข้อมูลของตัวแปร `id`, `sex`, ... , `bonus` แต่ละตัวแปร

ต่อไปจึงนำตัวแปรทุกตัวมารวมกันด้วยคำสั่ง `data.frame(id, sex, ... , bonus)`



```

> id=c(1,2,3,4,5)
> sex=c("Male","Female","Female","Male","Female")
> age=c(37,29,48,36,33)
> educ=c("Graduate","Post_gr","Under_gr","Under_gr","Graduate")
> status=c("Divorced","Single","Married","Married","Single")
> income=c(5500,4100,5400,6500,5400)
> grade=c(3.78,3.89,3.67,2.78,2.75)
> bonus=c(11000,82000,10800,13000,10800)
> company5=data.frame(id,sex,age,educ,status,income,grade,bonus)
> company5
  id  sex age  educ status income grade bonus
1  1 Male  37 Graduate Divorced   5500   3.78 11000
2  2 Female 29 Post_gr  Single   4100   3.89 82000
3  3 Female 48 Under_gr Married   5400   3.67 10800
4  4 Male  36 Under_gr Married   6500   2.78 13000
5  5 Female 33 Graduate  Single   5400   2.75 10800
> save(company5,file="C:/Rtemp/dataset_company5.RData")
>
  
```

คำสั่งแนะนำ `setwd` เป็นคำสั่งกำหนด Working directory ที่ต้องการทำงานร่วมกับโปรแกรม R ขณะนั้น ตัวอย่างเช่น

```

> setwd("C:/Windows")
> getwd()
[1] "C:/Windows"
> setwd("C:/Rtemp")
> getwd()
[1] "C:/Rtemp"
  
```


ความหมายของแต่ละบรรทัดคำสั่ง

```
id = c(1, 2, 3, 4, 5)
```

กำหนดค่าตัวแปร id

```
sex = c("Male", "Female", "Female", "Male", "Female")
```

กำหนดค่าตัวแปร sex

```
age = c(37, 29, 48, 36, 33)
```

กำหนดค่าตัวแปร age

```
educ = c("Graduate", "Post_gr", "Under_gr", "Under_gr", "Graduate")
```

กำหนดค่าตัวแปร educ

```
status = c("Divorced", "Single", "Married", "Married", "Single")
```

กำหนดค่าตัวแปร status

```
income = c(5500, 4100, 5400, 6500, 5400)
```

กำหนดค่าตัวแปร income

```
grade = c(3.78, 3.89, 3.67, 2.78, 2.75)
```

กำหนดค่าตัวแปร grade

```
bonus = c(11000, 82000, 10800, 13000, 10800)
```

กำหนดค่าตัวแปร bonus

```
company5 = data.frame(id, sex, age, educ, status, income, grade, bonus)
```

รวมตัวแปร id, sex, ... , bonus ด้วยคำสั่ง data.frame แล้วเก็บที่ dataset ชื่อ company5
company5

แสดงข้อมูลใน dataset ชื่อ company5

```
save(company5, file = "C:/Rtemp/dataset_company5.RData")
```

บันทึก dataset ชื่อ company5 ไว้ที่แฟ้มชื่อ dataset_company5.RData

และเก็บไว้ที่ Directory C:\Rtemp

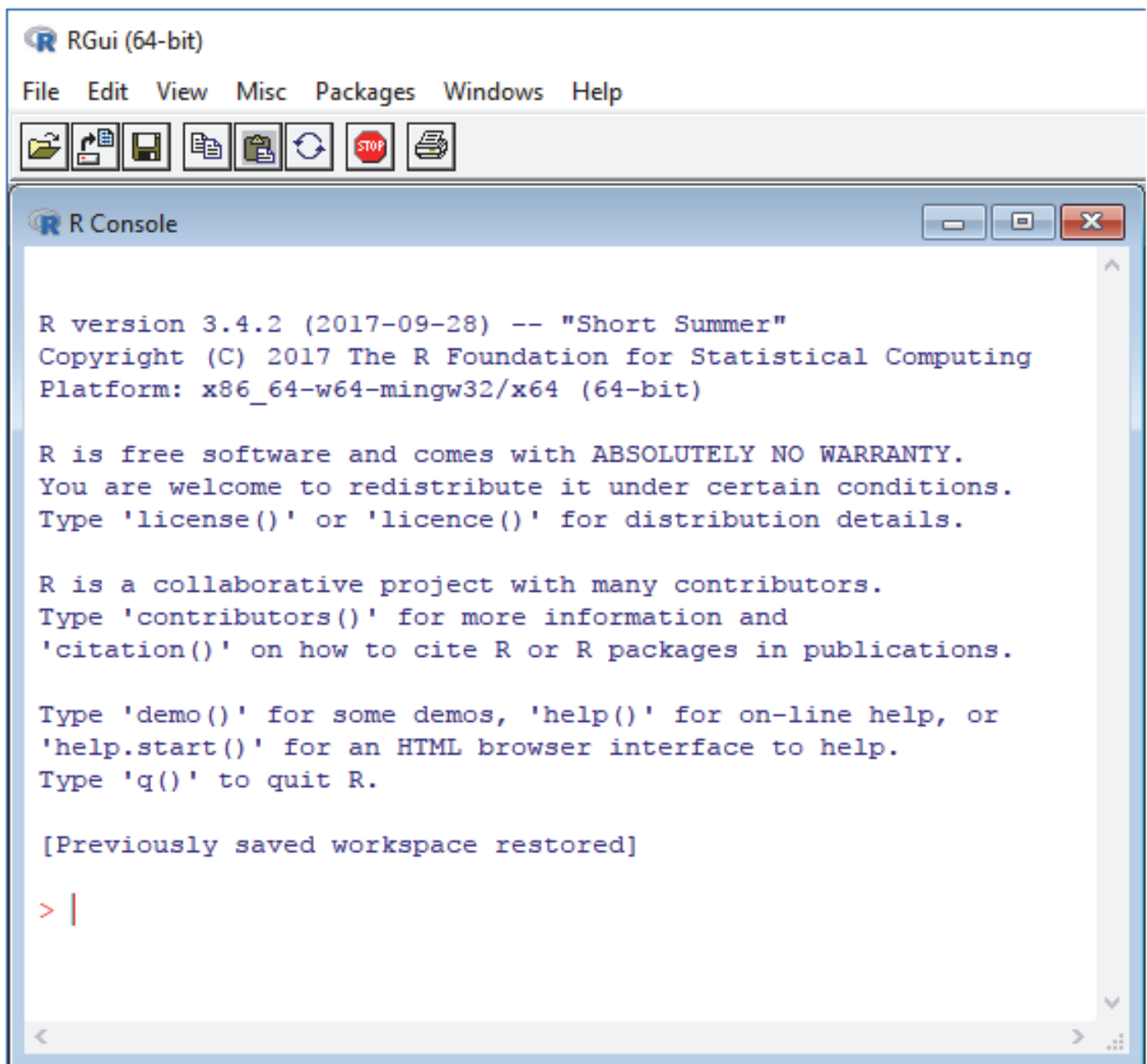
หมายเหตุ dataset_company5.RData เป็นชื่อแฟ้มข้อมูล

company5 เป็นชื่อของ dataset ที่อยู่ในแฟ้ม

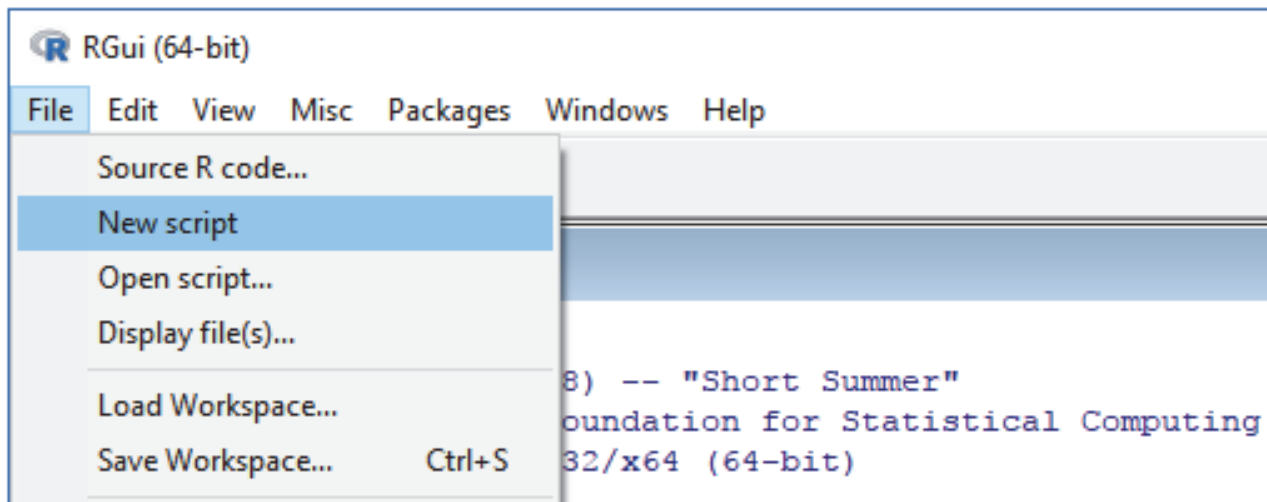
ชื่อแฟ้มข้อมูลกับชื่อ dataset ในแฟ้มข้อมูล มีชื่อเหมือนกัน หรือ ต่างกัน ก็ได้

คำสั่งแนะนำ ls() เป็นคำสั่งให้แสดง Object ที่มีอยู่ในขณะนั้น

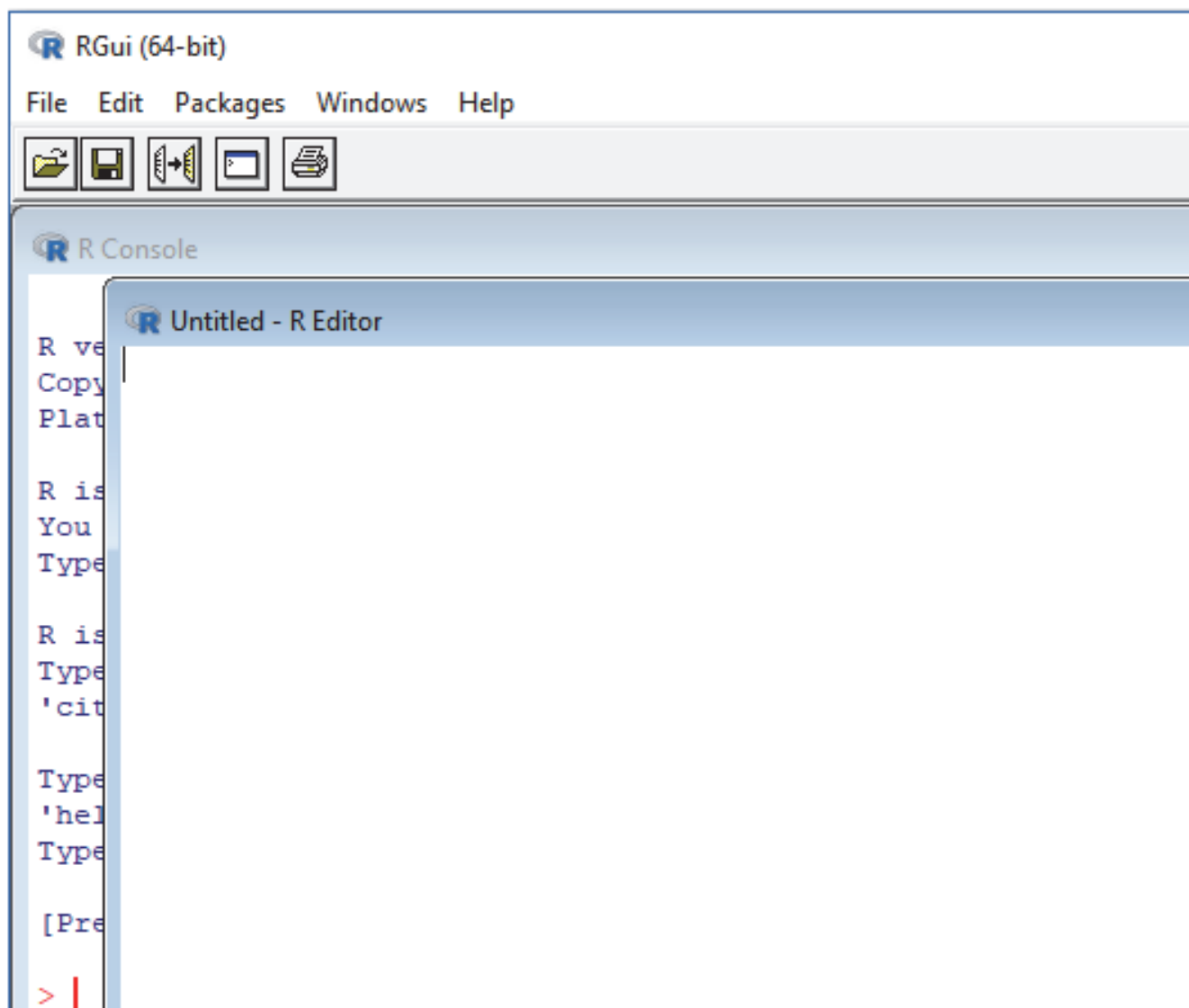
วิธีที่ 2 เขียนคำสั่งการกำหนดตัวแปรเป็นโปรแกรมใน R Script Editor ก่อน แล้วสั่ง Run ที่หลัง
ขั้นที่ 1 เริ่มที่ Window RGui



ขั้นที่ 2 เลือกคำสั่ง File/New script



ขั้นที่ 3 คลิกที่คำสั่ง New script จะได้ Window R Editor เพิ่มขึ้นอีกหนึ่ง Window เป็นดังนี้



หมายเหตุ ชื่อของ R Script ขณะนี้คือ Untitled เป็นชื่อ R Script ชั่วคราวที่โปรแกรม R กำหนดให้

ขั้นที่ 4 พิมพ์คำสั่งต่อไปนี้ใน Window R Editor

```
id = c(1, 2, 3, 4, 5)
```

```
sex = c("Male", "Female", "Female", "Male", "Female")
```

```
age = c(37, 29, 48, 36, 33)
```

```
educ = c("Graduate", "Post_gr", "Under_gr", "Under_gr", "Graduate")
```

```
status = c("Divorced", "Single", "Married", "Married", "Single")
```

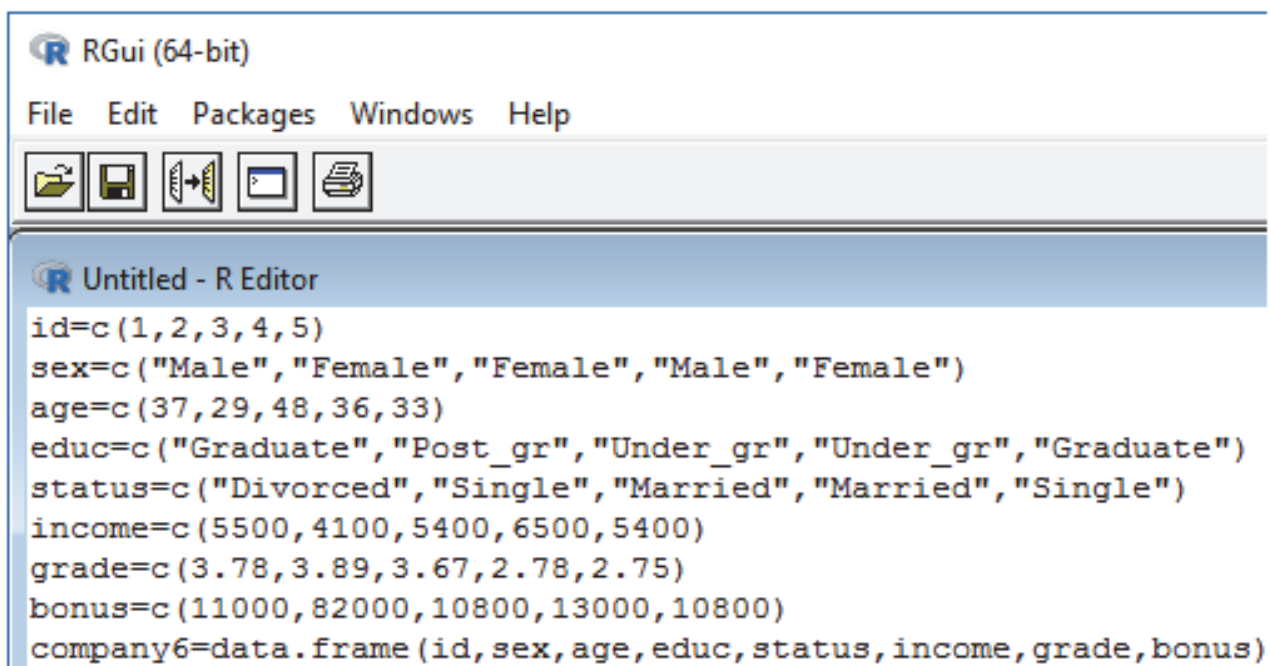
```
income = c(5500, 4100, 5400, 6500, 5400)
```

```
grade = c(3.78, 3.89, 3.67, 2.78, 2.75)
```

```
bonus = c(11000, 82000, 10800, 13000, 10800)
```

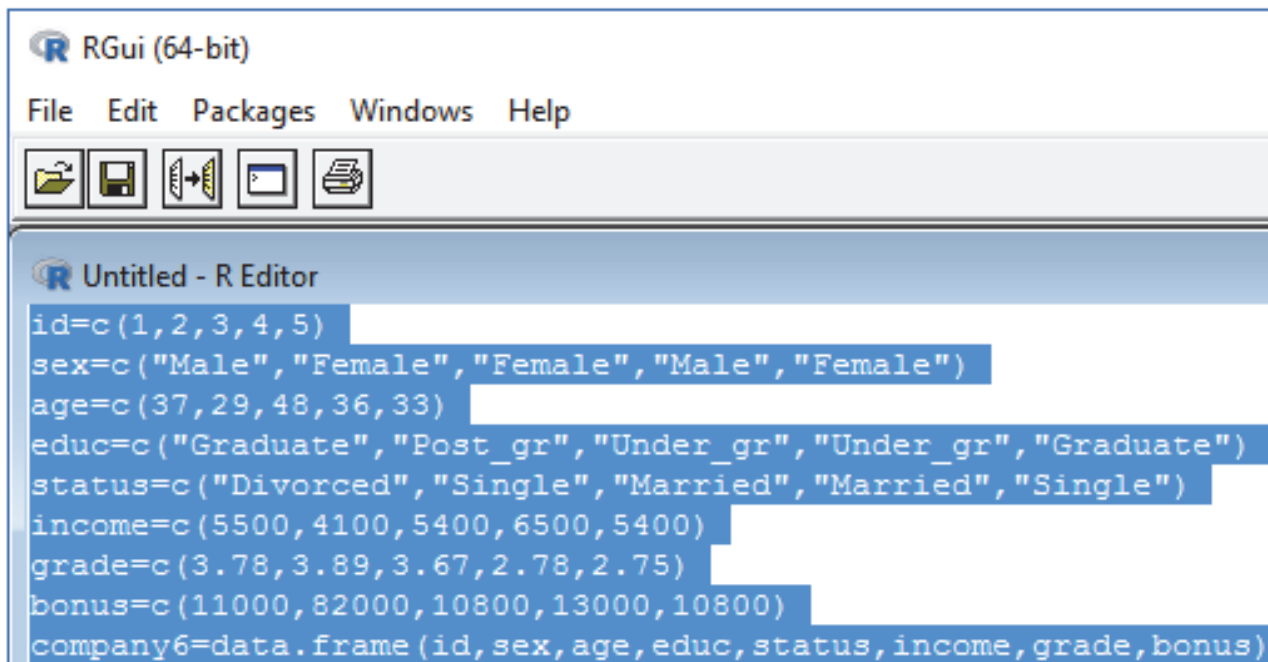
```
company6 = data.frame(id, sex, age, educ, status, income, grade, bonus)
```

ผลบนจอภาพเป็นดังนี้



หมายเหตุ บันทึก Script ชื่อ script_ch_01_02_001_12Nov2017

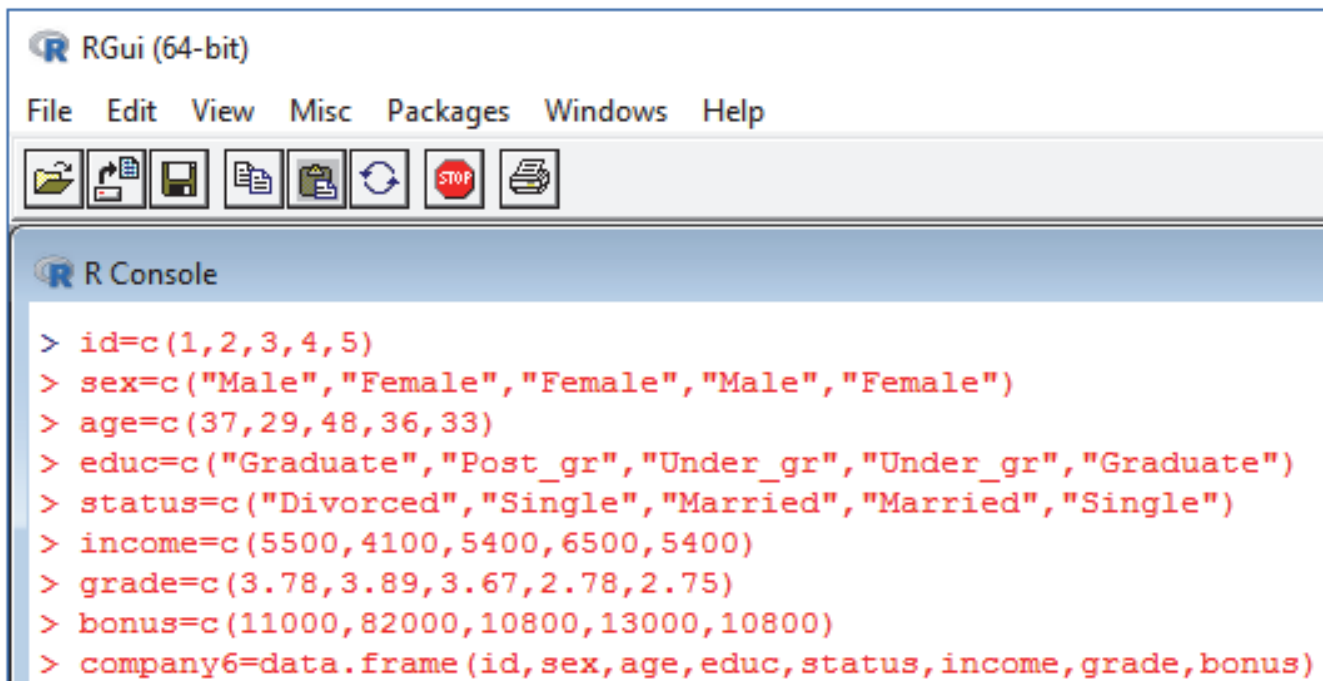
ขั้นที่ 5 กด Ctrl + A เพื่อเลือกคำสั่งทั้งหมด ผลบนจอภาพของ Window R Editor จะเป็นดังนี้



The screenshot shows the RGui (64-bit) interface. The menu bar includes File, Edit, Packages, Windows, and Help. The toolbar contains icons for opening, saving, undo, redo, and printing. The main window is titled 'Untitled - R Editor' and contains the following R code:

```
id=c(1,2,3,4,5)
sex=c("Male","Female","Female","Male","Female")
age=c(37,29,48,36,33)
educ=c("Graduate","Post_gr","Under_gr","Under_gr","Graduate")
status=c("Divorced","Single","Married","Married","Single")
income=c(5500,4100,5400,6500,5400)
grade=c(3.78,3.89,3.67,2.78,2.75)
bonus=c(11000,82000,10800,13000,10800)
company6=data.frame(id,sex,age,educ,status,income,grade,bonus)
```

ขั้นที่ 6 กด Ctrl + R เพื่อสั่ง Run คำสั่งที่เลือกทั้งหมดข้างต้น
การทำงานของโปรแกรมจะไปปรากฏที่ Window R Console เป็นดังนี้



The screenshot shows the RGui (64-bit) interface with the R Console window active. The menu bar includes File, Edit, View, Misc, Packages, Windows, and Help. The toolbar contains icons for opening, saving, undo, redo, running, and printing. The R Console displays the following commands and their output:

```
> id=c(1,2,3,4,5)
> sex=c("Male","Female","Female","Male","Female")
> age=c(37,29,48,36,33)
> educ=c("Graduate","Post_gr","Under_gr","Under_gr","Graduate")
> status=c("Divorced","Single","Married","Married","Single")
> income=c(5500,4100,5400,6500,5400)
> grade=c(3.78,3.89,3.67,2.78,2.75)
> bonus=c(11000,82000,10800,13000,10800)
> company6=data.frame(id,sex,age,educ,status,income,grade,bonus)
```

ขณะนี้เราได้สร้าง dataset ชื่อ company6 เสร็จแล้ว

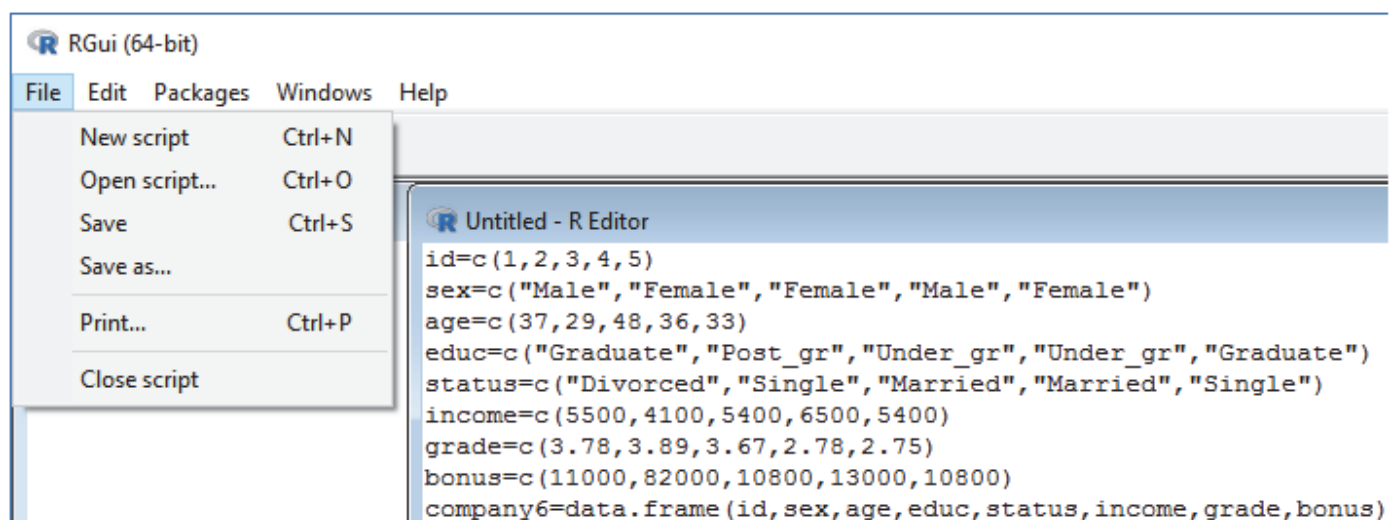
พิมพ์ company6 ใน Window R Console เพื่อแสดงค่าต่าง ๆ ที่มีใน dataset ชื่อ company6 และ บันทึก dataset ชื่อ company6 ไว้ที่แฟ้มชื่อ dataset_company6.RData ใน directory C:/Rtemp

```
> company6
  id    sex age   educ   status income grade bonus
1  1  Male  37 Graduate Divorced   5500   3.78 11000
2  2 Female  29 Post_gr  Single   4100   3.89 82000
3  3 Female  48 Under_gr Married   5400   3.67 10800
4  4  Male  36 Under_gr Married   6500   2.78 13000
5  5 Female  33 Graduate  Single   5400   2.75 10800
> save(company6,file="C:/Rtemp/dataset_company6.RData")
```

ต่อไปแนะนำให้บันทึก R Script ที่เราพิมพ์มาข้างต้น เพื่อเก็บข้อความต่าง ๆ ที่อยู่ใน R Script ไว้ศึกษาภายหลัง

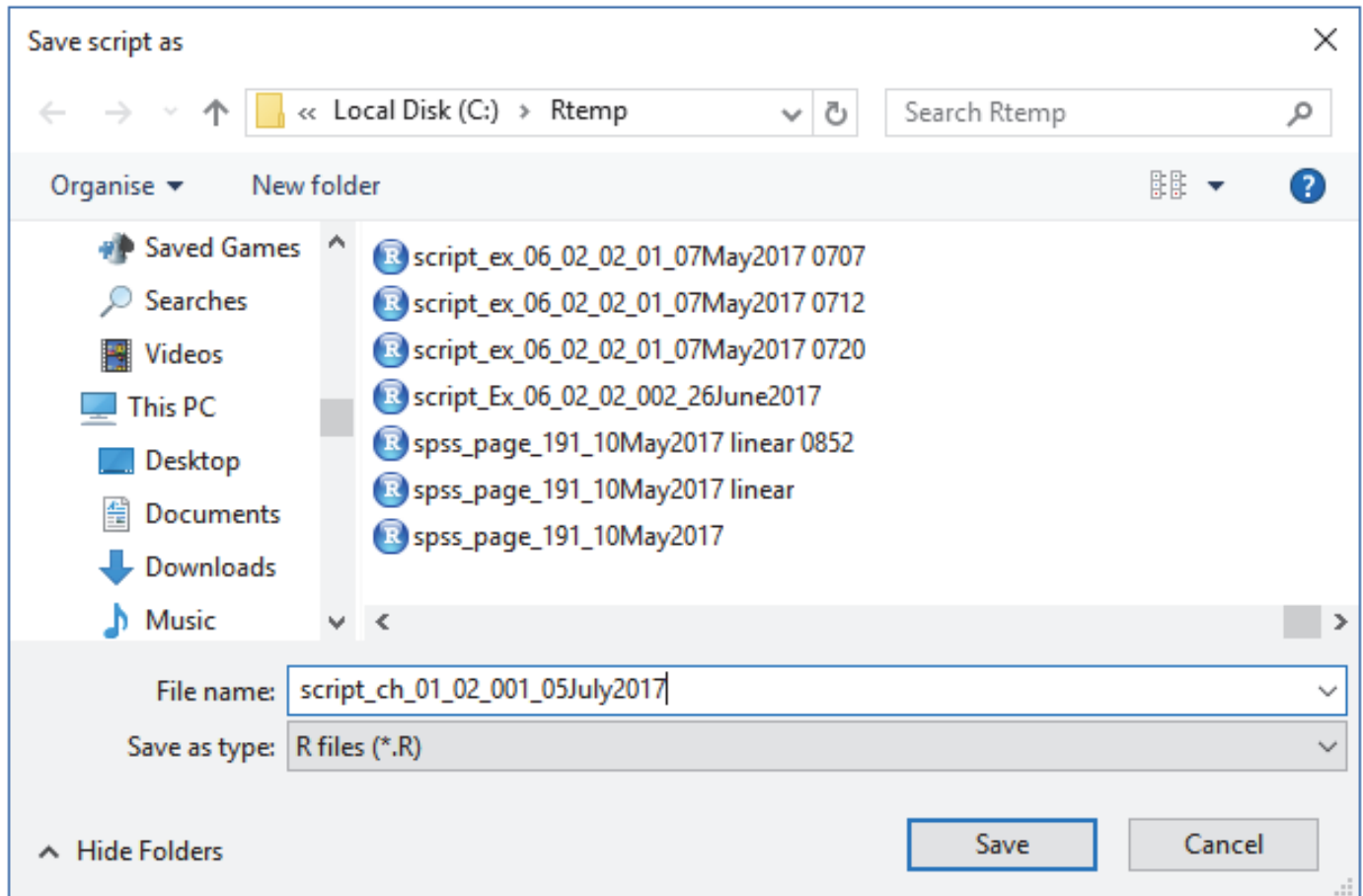
การบันทึก R Script มีขั้นตอนดังนี้

ขั้นที่ 1 คลิกที่ Window R Editor เพื่อให้เมนูคำสั่งเป็นของ Window R Editor
เลือกคำสั่ง File/Save



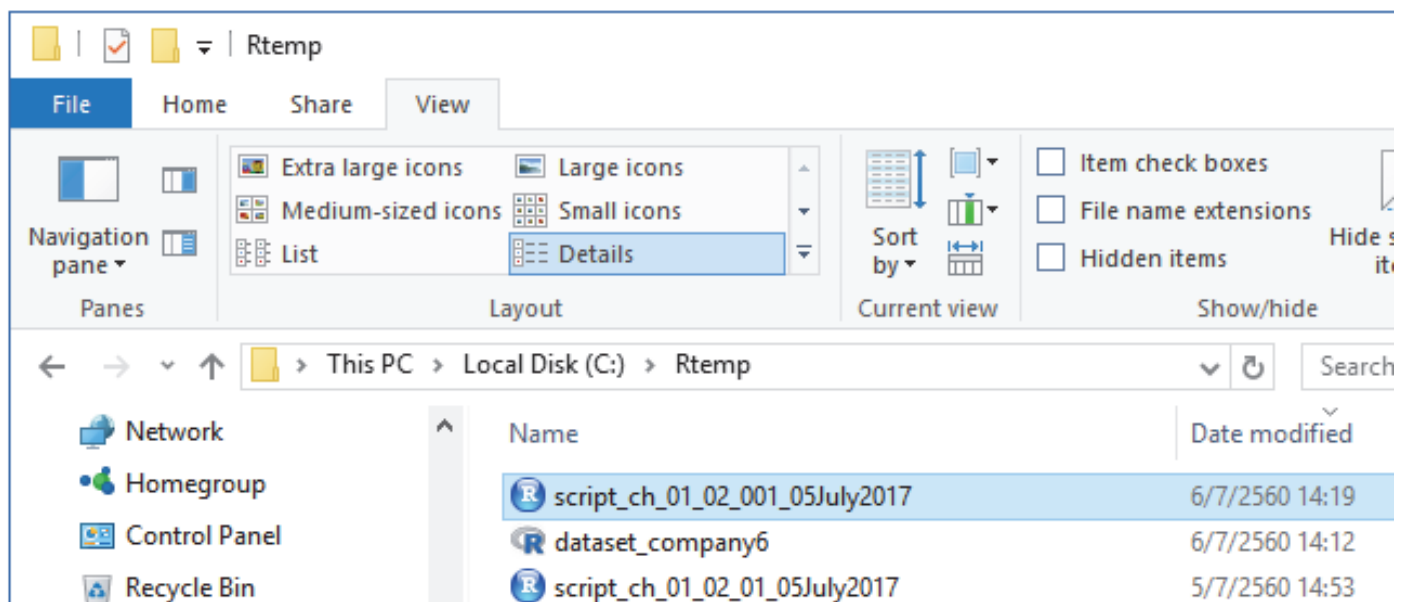
ขั้นที่ 2 คลิกที่คำสั่ง Save จะได้เมนูย่อย Save script as ต่อไปให้ทำดังนี้

1. เลือก Directory ที่ต้องการเก็บ R Script (ขอให้เลือกเป็น C:/Rtemp)
2. พิมพ์ชื่อที่ต้องการ เช่น script_ch_01_02_001_05July2017.R
3. คลิกปุ่ม Save



ขอให้ลองตรวจสอบที่ Directory C:/Rtemp ที่เราเลือกไว้

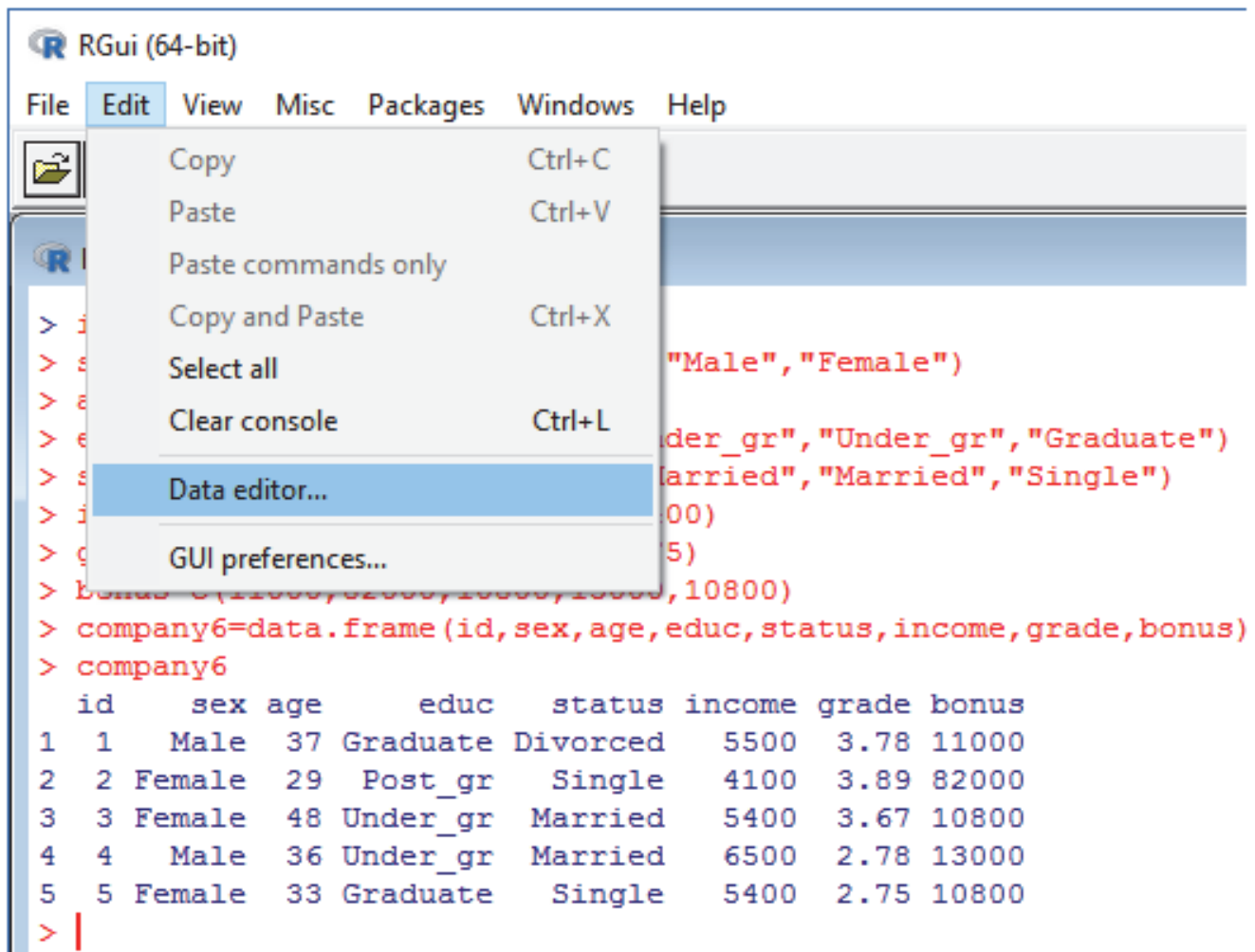
จะพบ Script ชื่อ script_ch_01_02_001_05July2017.R เป็นดังนี้



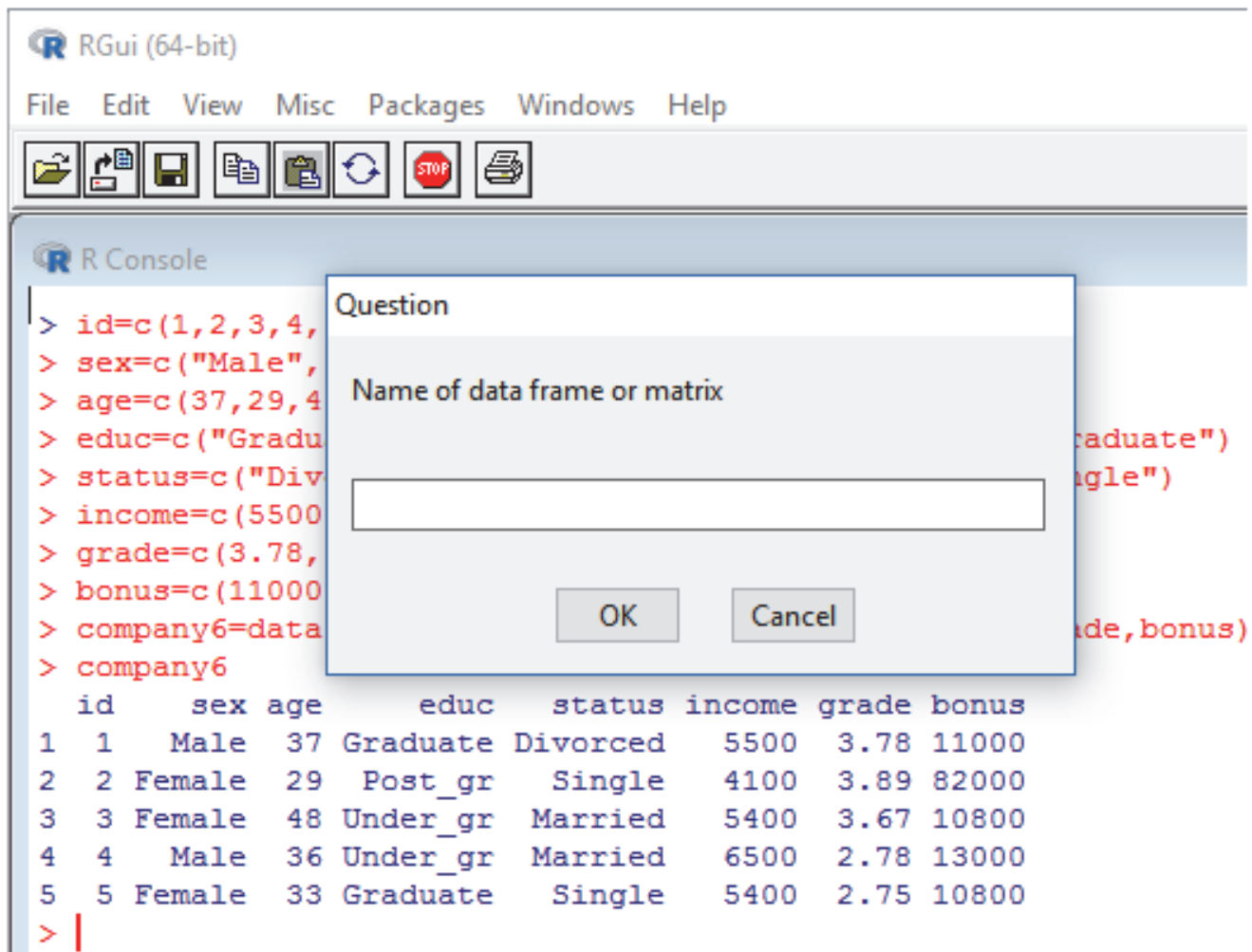
การตรวจสอบ ปรับปรุง และ แก้ไข dataset

หมายเหตุ ขณะนี้ถือว่ามี dataset ชื่อ company6 ใน RGui แล้ว

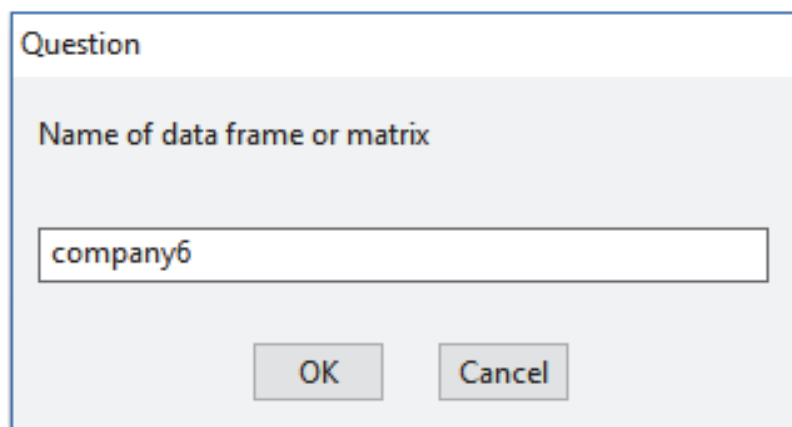
ขั้นที่ 1 เลือกคำสั่ง Edit/Data editor



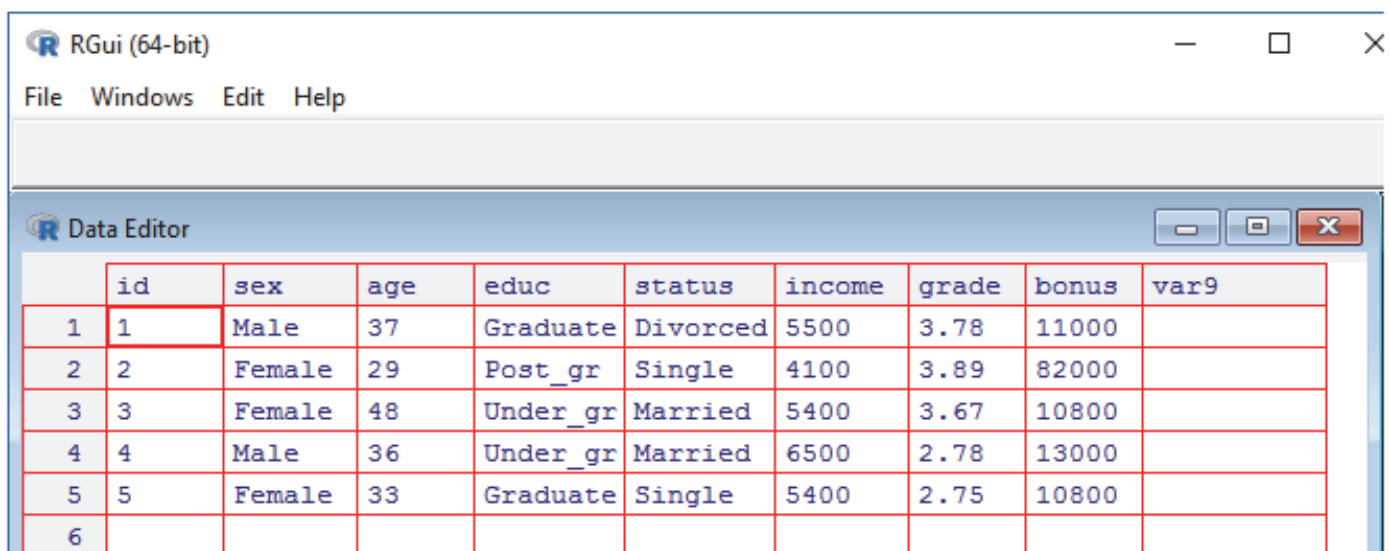
ขั้นที่ 2 คลิกคำสั่ง Data editor จะได้เมนูย่อย Question เป็นดังนี้



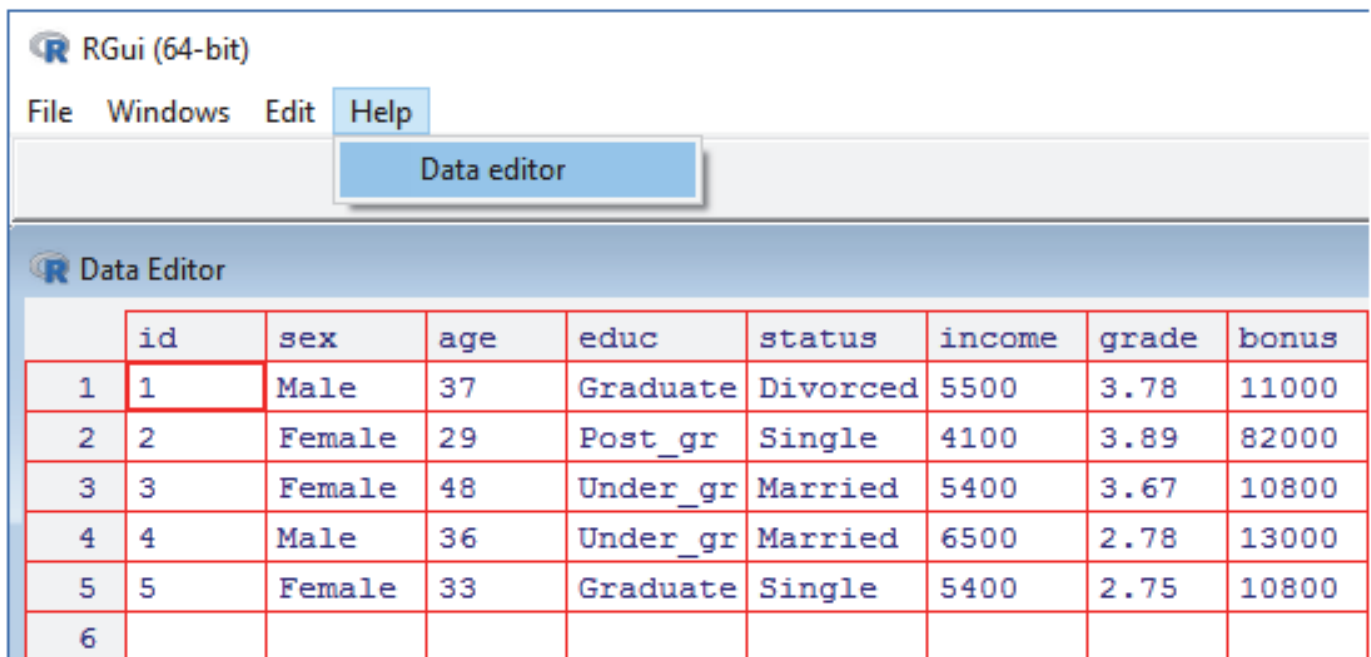
ขั้นที่ 3 พิมพ์ชื่อ dataset ที่ต้องการแก้ไขคือ company6



บนจอภาพจะเปิด Window Data Editor ให้ตรวจสอบ เพิ่ม/ลด ตัวแปร และ แก้ไขข้อมูล



คำสั่ง Help/Data editor



	id	sex	age	educ	status	income	grade	bonus
1	1	Male	37	Graduate	Divorced	5500	3.78	11000
2	2	Female	29	Post_gr	Single	4100	3.89	82000
3	3	Female	48	Under_gr	Married	5400	3.67	10800
4	4	Male	36	Under_gr	Married	6500	2.78	13000
5	5	Female	33	Graduate	Single	5400	2.75	10800
6								

จะได้คำแนะนำในการปรับปรุง เพิ่ม/ลด ตัวแปร และแก้ไขข้อมูล

Information



Navigation.

Keyboard: cursor keys move selection
Tab move right, Shift+Tab moves left
PgDn or Ctrl+F: move down one screenful
PgUp or Ctrl+B: move up one screenful
Home: move to (1,1) cell
End: show last rows of last column.
Mouse: left-click in a cell, use the scrollbar(s).

Editing.

Type in the currently hightlighted cell
Double-click in a cell for an editable field

Misc.

Ctrl-L redraws the screen, auto-resizing the columns
Ctrl-C copies selected cell
Ctrl-V pastes to selected cell
Right-click menu for copy, paste, autosize currently selected column

OK

การนำเข้าข้อมูล *.xlsx หรือ *.xls ของ Excel มาเป็นข้อมูล dataset ของ RGui
ให้ข้อมูลของพนักงานในบริษัทเป็นดังนี้

id	sex	age	educ	status	income	grade	bonus
1	Male	37	Graduate	Divorced	5500	3.78	11000
2	Female	29	Post_gr	Single	4100	3.89	8200
3	Female	48	Under_gr	Married	5400	3.67	10800
4	Male	36	Under_gr	Married	6500	2.78	13000
5	Female	33	Graduate	Single	5400	2.75	10800
6	Female	45	Post_gr	Divorced	8300	3.45	16600
7	Female	38	Under_gr	Divorced	7700	3.89	15400
8	Female	23	Post_gr	Single	3900	3.67	7800
9	Male	34	Graduate	Divorced	4500	2.56	9000
10	Male	50	Graduate	Married	6700	2.69	13400
11	Female	43	Graduate	Married	4700	3.56	9400
12	Female	37	Post_gr	Married	3900	3.15	7800
13	Male	24	Graduate	Single	3300	2.45	6600
14	Male	46	Graduate	Married	4900	2.45	9800
15	Male	32	Under_gr	Single	4200	3.87	8400
16	Male	42	Graduate	Widow	6600	3.67	13200
17	Male	38	Doctor	Married	8000	3.23	16000
18	Female	41	Graduate	Widow	7000	3.45	14000
19	Female	32	Under_gr	Married	3500	3.21	7000
20	Male	54	Graduate	Married	7400	3.25	14800
21	Female	32	Post_gr	Divorced	6200	2.56	12400
22	Male	43	Under_gr	Married	4700	2.45	9400
23	Female	22	Under_gr	Single	3400	3.78	6800
24	Male	40	Graduate	Married	5900	2.67	11800
25	Male	37	Doctor	Widow	7500	3.45	15000

สร้างแฟ้มข้อมูลข้างต้นใน Excel เป็นดังนี้

	A	B	C	D	E	F	G	H	I
1	id	sex	age	educ	status	income	grade	bonus	
2	1	Male	37	Graduate	Divorced	5500	3.78	11000	
3	2	Female	29	Post_gr	Single	4100	3.89	8200	
4	3	Female	48	Under_gr	Married	5400	3.67	10800	
5	4	Male	36	Under_gr	Married	6500	2.78	13000	
6	5	Female	33	Graduate	Single	5400	2.75	10800	
7	6	Female	45	Post_gr	Divorced	8300	3.45	16600	
8	7	Female	38	Under_gr	Divorced	7700	3.89	15400	
9	8	Female	23	Post_gr	Single	3900	3.67	7800	
10	9	Male	34	Graduate	Divorced	4500	2.56	9000	
11	10	Male	50	Graduate	Married	6700	2.69	13400	
12	11	Female	43	Graduate	Married	4700	3.56	9400	
13	12	Female	37	Post_gr	Married	3900	3.15	7800	
14	13	Male	24	Graduate	Single	3300	2.45	6600	
15	14	Male	46	Graduate	Married	4900	2.45	9800	
16	15	Male	32	Under_gr	Single	4200	3.87	8400	
17	16	Male	42	Graduate	Widow	6600	3.67	13200	
18	17	Male	38	Doctor	Married	8000	3.23	16000	
19	18	Female	41	Graduate	Widow	7000	3.45	14000	
20	19	Female	32	Under_gr	Married	3500	3.21	7000	
21	20	Male	54	Graduate	Married	7400	3.25	14800	
22	21	Female	32	Post_gr	Divorced	6200	2.56	12400	
23	22	Male	43	Under_gr	Married	4700	2.45	9400	
24	23	Female	22	Under_gr	Single	3400	3.78	6800	
25	24	Male	40	Graduate	Married	5900	2.67	11800	
26	25	Male	37	Doctor	Widow	7500	3.45	15000	

หมายเหตุ บันทึกแฟ้มข้อมูลไว้ทั้ง 3 แบบที่ Directory C:/Rtemp

1. บันทึกแฟ้ม excel ชื่อ File_001.xlsx
2. บันทึกแฟ้มแบบ Text (Tab Delimited) ชื่อ File_001_tabdelimited.txt
3. บันทึกแฟ้มแบบ CSV ชื่อ File_001_commadelimited.csv

หมายเหตุ แฟ้มข้อมูลแบบ *.csv และ *.txt สามารถสร้างหรือแก้ไขได้ด้วยโปรแกรม Notepad

ในกรณีที่แฟ้มที่สร้างด้วย Excel และบันทึกเป็นแฟ้มแบบ *.txt (และเลือกค้นข้อมูลด้วย Tab)

ตัวอย่างเช่น C:/Rtemp/File_001_tab.txt

หมายเหตุ แฟ้มข้อมูลที่บันทึกในรูปแบบ Tab Delimited

สามารถสร้าง อ่าน และ แก้ไขได้ด้วยโปรแกรม Notepad

File_001_tab - Notepad									
File Edit Format View Help									
id	sex	age	educ	status	income	grade	bonus		
1	Male	37	Graduate	Divorced			5500	3.78	11000
2	Female	29	Post_gr	Single	4100	3.89	8200		
3	Female	48	Under_gr	Married	5400		3.67	10800	
4	Male	36	Under_gr	Married	6500		2.78	13000	
5	Female	33	Graduate	Single	5400		2.75	10800	
6	Female	45	Post_gr	Divorced		8300	3.45	16600	
7	Female	38	Under_gr	Divorced			7700	3.89	15400
8	Female	23	Post_gr	Single	3900	3.67	7800		
9	Male	34	Graduate	Divorced			4500	2.56	9000
10	Male	50	Graduate	Married	6700		2.69	13400	
11	Female	43	Graduate	Married	4700		3.56	9400	
12	Female	37	Post_gr	Married	3900	3.15	7800		
13	Male	24	Graduate	Single	3300		2.45	6600	
14	Male	46	Graduate	Married	4900		2.45	9800	
15	Male	32	Under_gr	Single	4200		3.87	8400	
16	Male	42	Graduate	Widow	6600		3.67	13200	
17	Male	38	Doctor	Married	8000	3.23	16000		
18	Female	41	Graduate	Widow	7000		3.45	14000	
19	Female	32	Under_gr	Married	3500		3.21	7000	
20	Male	54	Graduate	Married	7400		3.25	14800	
21	Female	32	Post_gr	Divorced		6200	2.56	12400	
22	Male	43	Under_gr	Married	4700		2.45	9400	
23	Female	22	Under_gr	Single	3400		3.78	6800	
24	Male	40	Graduate	Married	5900		2.67	11800	
25	Male	37	Doctor	Widow	7500	3.45	15000		

ในกรณีที่แฟ้มที่สร้างด้วย Excel และบันทึกเป็นแฟ้มแบบ *.csv (และเลือกคั่นข้อมูลด้วย comma)

ตัวอย่างเช่น C:/Rtemp/File_001_comma.csv

หมายเหตุ แฟ้มข้อมูลที่บันทึกในรูปแบบ *.csv (comma separate variable)

สามารถสร้าง อ่าน และ แก้ไขได้ด้วย โปรแกรม Notepad

```

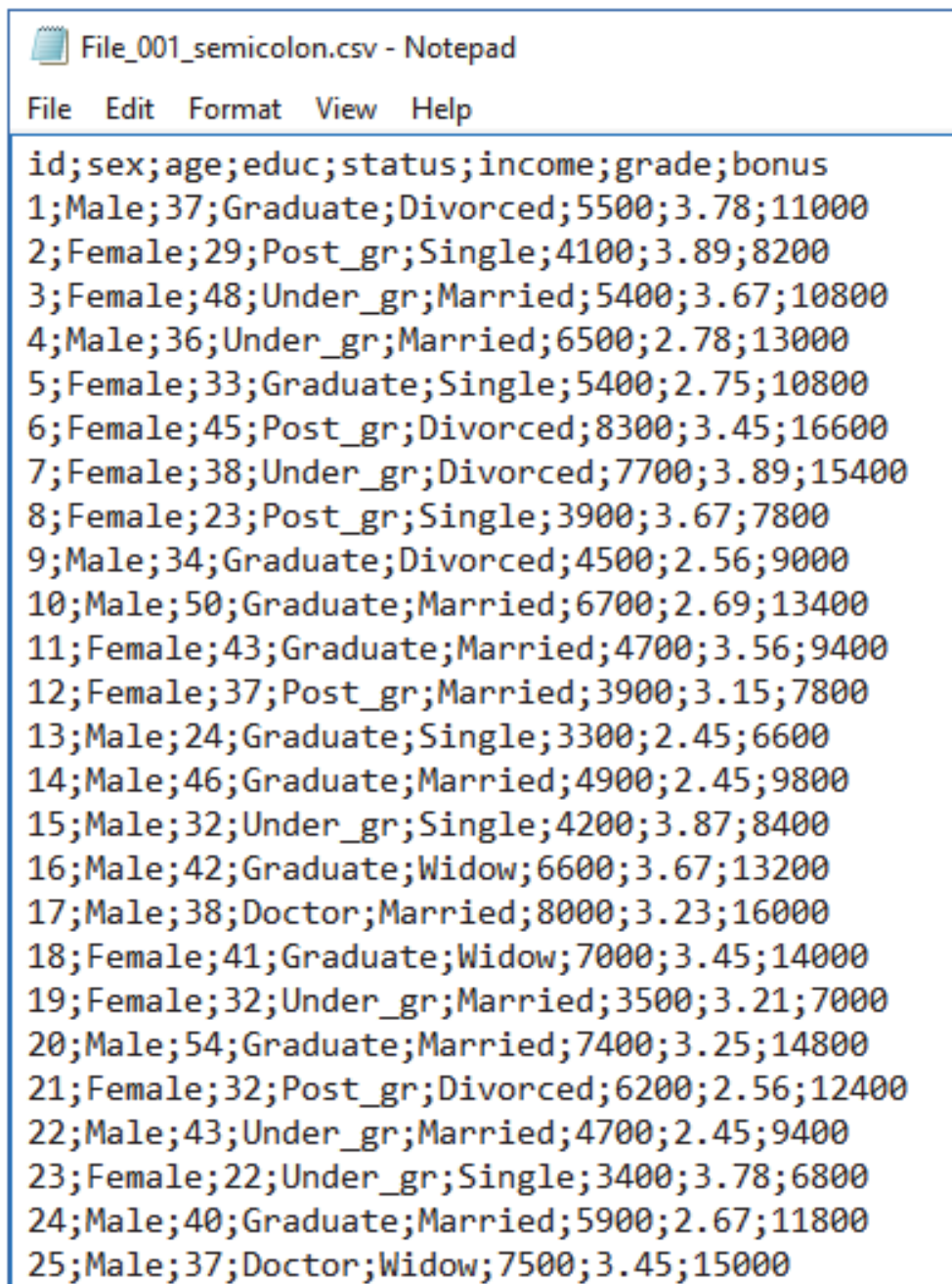
File_001_comma.csv - Notepad
File Edit Format View Help

id,sex,age,educ,status,income,grade,bonus
1,Male,37,Graduate,Divorced,5500,3.78,11000
2,Female,29,Post_gr,Single,4100,3.89,8200
3,Female,48,Under_gr,Married,5400,3.67,10800
4,Male,36,Under_gr,Married,6500,2.78,13000
5,Female,33,Graduate,Single,5400,2.75,10800
6,Female,45,Post_gr,Divorced,8300,3.45,16600
7,Female,38,Under_gr,Divorced,7700,3.89,15400
8,Female,23,Post_gr,Single,3900,3.67,7800
9,Male,34,Graduate,Divorced,4500,2.56,9000
10,Male,50,Graduate,Married,6700,2.69,13400
11,Female,43,Graduate,Married,4700,3.56,9400
12,Female,37,Post_gr,Married,3900,3.15,7800
13,Male,24,Graduate,Single,3300,2.45,6600
14,Male,46,Graduate,Married,4900,2.45,9800
15,Male,32,Under_gr,Single,4200,3.87,8400
16,Male,42,Graduate,Widow,6600,3.67,13200
17,Male,38,Doctor,Married,8000,3.23,16000
18,Female,41,Graduate,Widow,7000,3.45,14000
19,Female,32,Under_gr,Married,3500,3.21,7000
20,Male,54,Graduate,Married,7400,3.25,14800
21,Female,32,Post_gr,Divorced,6200,2.56,12400
22,Male,43,Under_gr,Married,4700,2.45,9400
23,Female,22,Under_gr,Single,3400,3.78,6800
24,Male,40,Graduate,Married,5900,2.67,11800
25,Male,37,Doctor,Widow,7500,3.45,15000

```


ในกรณีที่แฟ้มที่สร้างด้วย Excel และบันทึกเป็นแฟ้มแบบ *.csv (และเลือกคั่นข้อมูลด้วย semicolon) ตัวอย่างเช่น C:/Rtemp/File_001_semicolon.csv

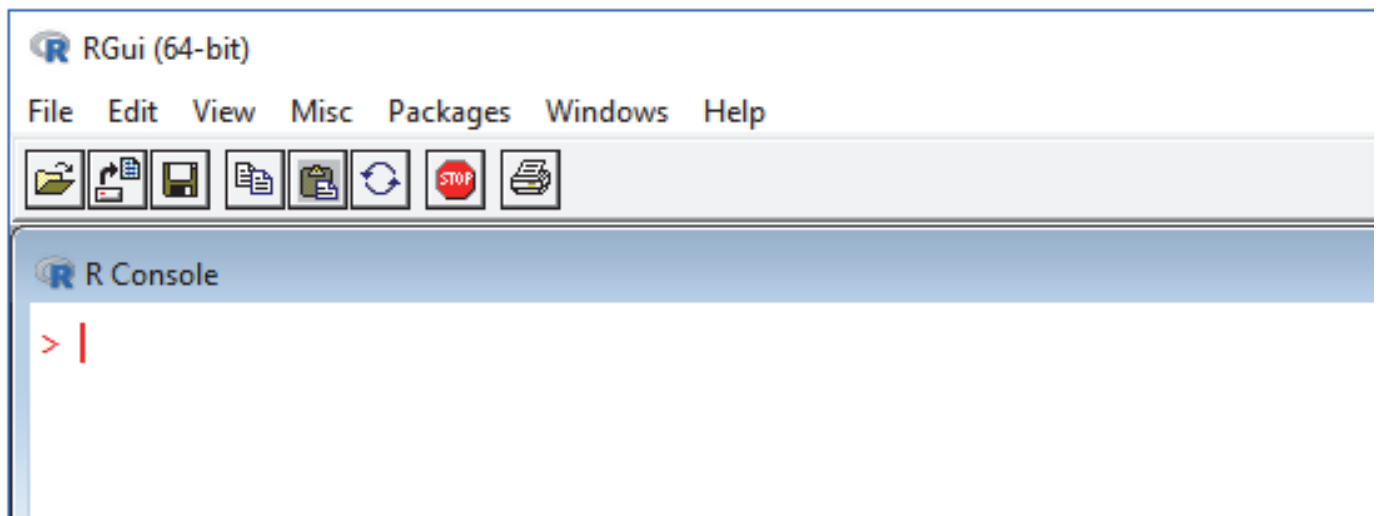
หมายเหตุ แฟ้มข้อมูลที่บันทึกในรูปแบบ *.csv และคั่นข้อมูลด้วย semicolon สามารถสร้าง อ่าน และ แก้ไขได้ด้วย โปรแกรม Notepad



```
id;sex;age;educ;status;income;grade;bonus
1;Male;37;Graduate;Divorced;5500;3.78;11000
2;Female;29;Post_gr;Single;4100;3.89;8200
3;Female;48;Under_gr;Married;5400;3.67;10800
4;Male;36;Under_gr;Married;6500;2.78;13000
5;Female;33;Graduate;Single;5400;2.75;10800
6;Female;45;Post_gr;Divorced;8300;3.45;16600
7;Female;38;Under_gr;Divorced;7700;3.89;15400
8;Female;23;Post_gr;Single;3900;3.67;7800
9;Male;34;Graduate;Divorced;4500;2.56;9000
10;Male;50;Graduate;Married;6700;2.69;13400
11;Female;43;Graduate;Married;4700;3.56;9400
12;Female;37;Post_gr;Married;3900;3.15;7800
13;Male;24;Graduate;Single;3300;2.45;6600
14;Male;46;Graduate;Married;4900;2.45;9800
15;Male;32;Under_gr;Single;4200;3.87;8400
16;Male;42;Graduate;Widow;6600;3.67;13200
17;Male;38;Doctor;Married;8000;3.23;16000
18;Female;41;Graduate;Widow;7000;3.45;14000
19;Female;32;Under_gr;Married;3500;3.21;7000
20;Male;54;Graduate;Married;7400;3.25;14800
21;Female;32;Post_gr;Divorced;6200;2.56;12400
22;Male;43;Under_gr;Married;4700;2.45;9400
23;Female;22;Under_gr;Single;3400;3.78;6800
24;Male;40;Graduate;Married;5900;2.67;11800
25;Male;37;Doctor;Widow;7500;3.45;15000
```

การเปิดแฟ้มข้อมูลแบบ *.txt (Tab Delimited)

ขั้นที่ 1 เปิดโปรแกรม RGui



ขั้นที่ 2

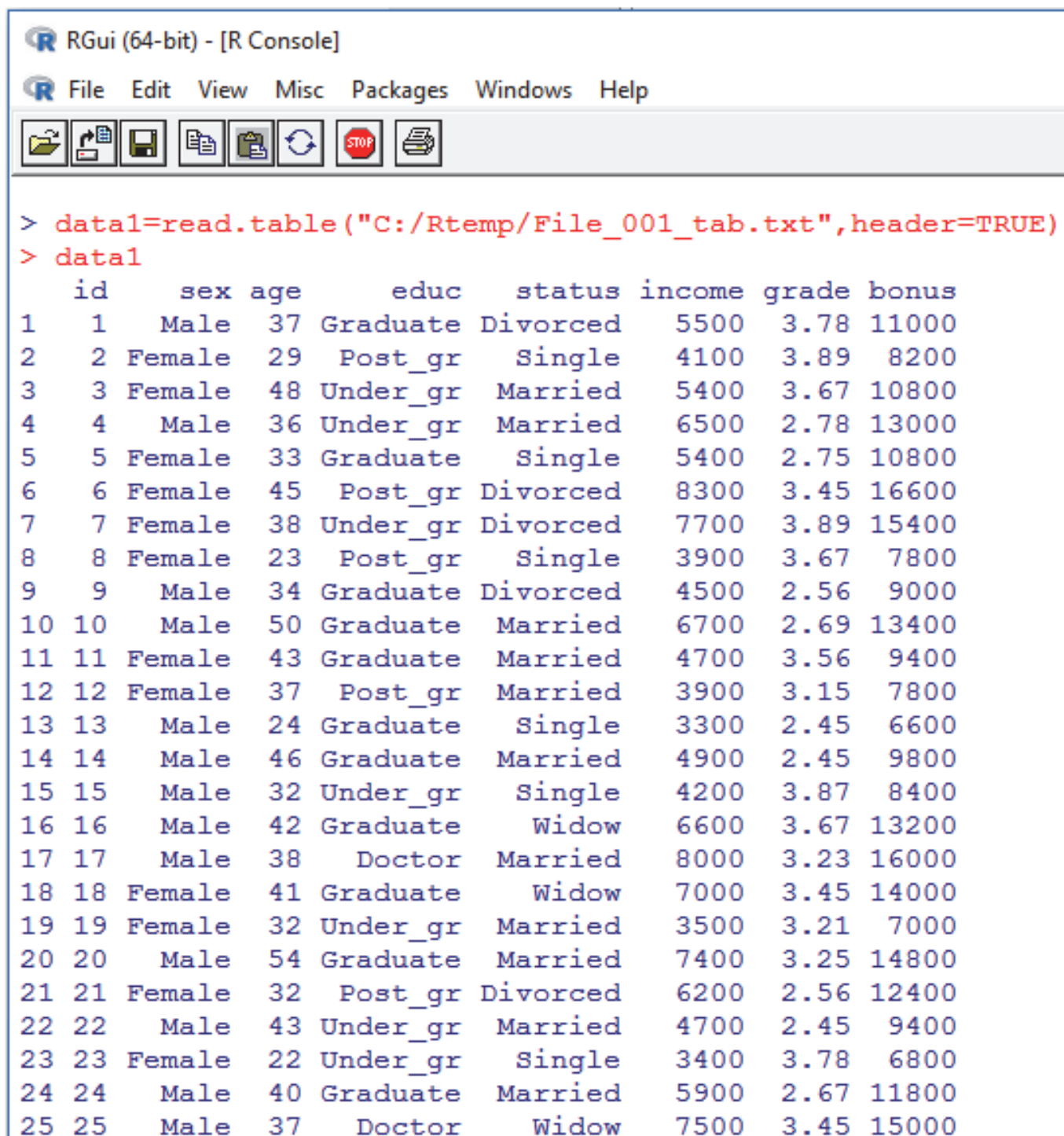
1. ใช้คำสั่งอ่านข้อมูล *.txt (tabdelimited) เข้ามาและเก็บไว้ที่ dataset ชื่อ data1

```
data1 = read.table("C:/Rtemp/File_001_tab.txt", header = TRUE)
```

หมายเหตุ คำสั่งข้างต้น เราระบุ directory ที่หน้าชื่อแฟ้ม *.txt

ถ้าเราใช้คำสั่ง `setwd("C:/Rtemp")` ก่อน แล้ว พิมพ์เฉพาะชื่อแฟ้ม File_001_tab.txt ก็พอ

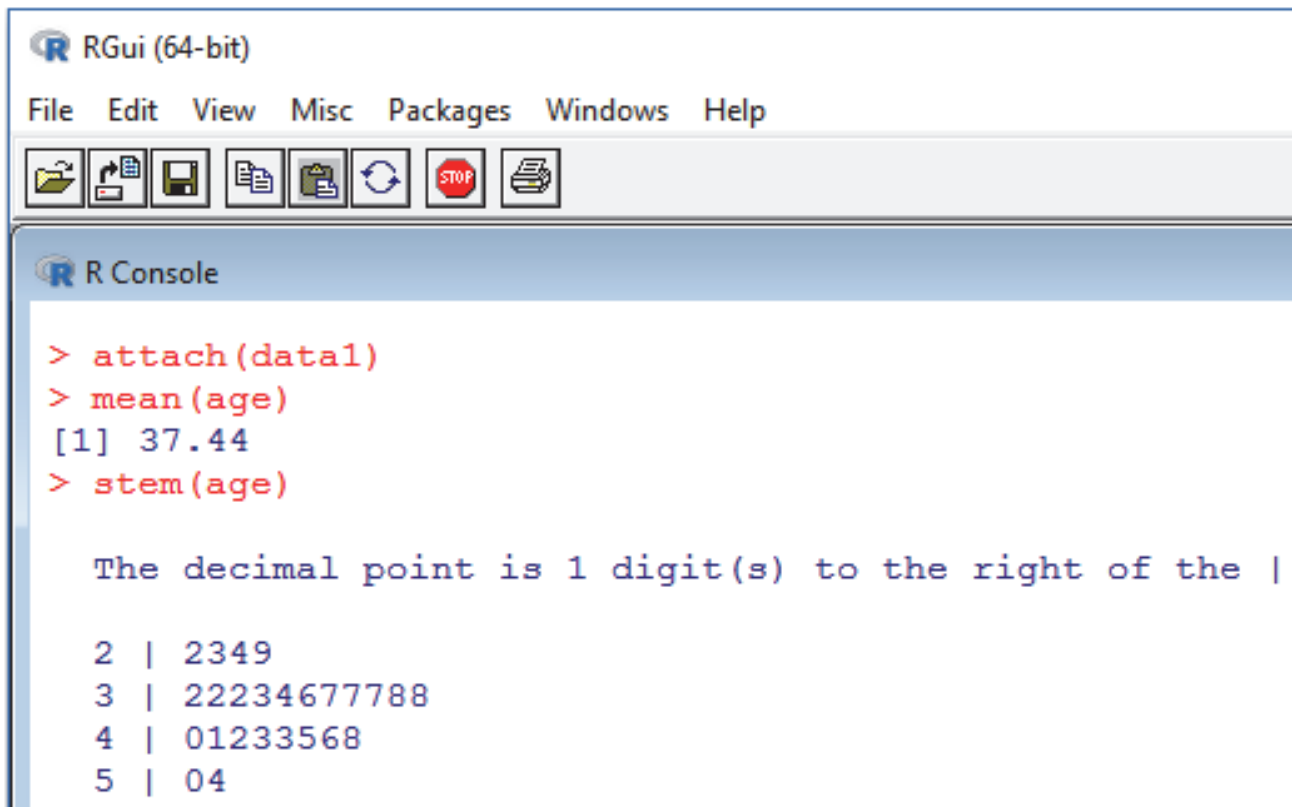
2. พิมพ์ data1 เพื่อแสดงข้อมูลใน dataset ชื่อ data1



```
> data1=read.table("C:/Rtemp/File_001_tab.txt",header=TRUE)
> data1
```

	id	sex	age	educ	status	income	grade	bonus
1	1	Male	37	Graduate	Divorced	5500	3.78	11000
2	2	Female	29	Post_gr	Single	4100	3.89	8200
3	3	Female	48	Under_gr	Married	5400	3.67	10800
4	4	Male	36	Under_gr	Married	6500	2.78	13000
5	5	Female	33	Graduate	Single	5400	2.75	10800
6	6	Female	45	Post_gr	Divorced	8300	3.45	16600
7	7	Female	38	Under_gr	Divorced	7700	3.89	15400
8	8	Female	23	Post_gr	Single	3900	3.67	7800
9	9	Male	34	Graduate	Divorced	4500	2.56	9000
10	10	Male	50	Graduate	Married	6700	2.69	13400
11	11	Female	43	Graduate	Married	4700	3.56	9400
12	12	Female	37	Post_gr	Married	3900	3.15	7800
13	13	Male	24	Graduate	Single	3300	2.45	6600
14	14	Male	46	Graduate	Married	4900	2.45	9800
15	15	Male	32	Under_gr	Single	4200	3.87	8400
16	16	Male	42	Graduate	Widow	6600	3.67	13200
17	17	Male	38	Doctor	Married	8000	3.23	16000
18	18	Female	41	Graduate	Widow	7000	3.45	14000
19	19	Female	32	Under_gr	Married	3500	3.21	7000
20	20	Male	54	Graduate	Married	7400	3.25	14800
21	21	Female	32	Post_gr	Divorced	6200	2.56	12400
22	22	Male	43	Under_gr	Married	4700	2.45	9400
23	23	Female	22	Under_gr	Single	3400	3.78	6800
24	24	Male	40	Graduate	Married	5900	2.67	11800
25	25	Male	37	Doctor	Widow	7500	3.45	15000

ขั้นที่ 3 ทดลองหาค่าเฉลี่ย เขียนกราฟแผนภาพต้นไม้และใบ



```

RGui (64-bit)
File Edit View Misc Packages Windows Help

> attach(data1)
> mean(age)
[1] 37.44
> stem(age)

The decimal point is 1 digit(s) to the right of the |

2 | 2349
3 | 22234677788
4 | 01233568
5 | 04
  
```

หมายเหตุ

1. `attach(data1)` เป็นการกำหนดให้ตัวแปร `id`, `sex`, ... , `bonus` ของ dataset ชื่อ `data1` สามารถนำมาวิเคราะห์ข้อมูลได้
2. `mean(age)` คำสั่งหาค่าเฉลี่ยของตัวแปร `age`
3. `stem(age)` คำสั่งเขียนกราฟแผนภาพต้นไม้และใบ

ความหมายจากกราฟ จะได้ข้อมูลในช่วงอายุต่าง ๆ มีพนักงานจำแนกรายบุคคลดังนี้

20 – 29 มี 4 คน และมีอายุดังนี้ 22, 23, 24, 29

30 – 39 มี 11 คน และมีอายุดังนี้ 32, 32, 32, 33, 34, 36, 37, 37, 37, 38, 38

40 – 49 มี 8 คน และมีอายุดังนี้ 40, 41, 42, 43, 43, 45, 46, 48

50 – 54 มี 2 คน และมีอายุดังนี้ 50, 54

ในกรณีที่เพิ่มข้อมูลจาก Excel บันทึกเป็นแฟ้มแบบ *.csv (และเลือกคั่นข้อมูลด้วย comma) ให้ใช้คำสั่งอ่านข้อมูลดังนี้

```
data2 = read.table("C:/Rtemp/File_001_comma.csv", header = TRUE, sep = ",")
```

กล่าวคือต้องมี sep = "," เพื่อระบุว่าข้อมูลแต่ละบรรทัดแยกกันด้วย comma

หมายเหตุ แฟ้มข้อมูลแบบ *.csv สามารถสร้างหรือแก้ไขได้ด้วยโปรแกรม Notepad

```
File_001_comma.csv - Notepad
File Edit Format View Help
id,sex,age,educ,status,income,grade,bonus
1,Male,37,Graduate,Divorced,5500,3.78,11000
2,Female,29,Post_gr,Single,4100,3.89,8200
3,Female,48,Under_gr,Married,5400,3.67,10800
4,Male,36,Under_gr,Married,6500,2.78,13000
5,Female,33,Graduate,Single,5400,2.75,10800
6,Female,45,Post_gr,Divorced,8300,3.45,16600
7,Female,38,Under_gr,Divorced,7700,3.89,15400
8,Female,23,Post_gr,Single,3900,3.67,7800
9,Male,34,Graduate,Divorced,4500,2.56,9000
10,Male,50,Graduate,Married,6700,2.69,13400
11,Female,43,Graduate,Married,4700,3.56,9400
12,Female,37,Post_gr,Married,3900,3.15,7800
13,Male,24,Graduate,Single,3300,2.45,6600
14,Male,46,Graduate,Married,4900,2.45,9800
15,Male,32,Under_gr,Single,4200,3.87,8400
16,Male,42,Graduate,Widow,6600,3.67,13200
17,Male,38,Doctor,Married,8000,3.23,16000
18,Female,41,Graduate,Widow,7000,3.45,14000
19,Female,32,Under_gr,Married,3500,3.21,7000
20,Male,54,Graduate,Married,7400,3.25,14800
21,Female,32,Post_gr,Divorced,6200,2.56,12400
22,Male,43,Under_gr,Married,4700,2.45,9400
23,Female,22,Under_gr,Single,3400,3.78,6800
24,Male,40,Graduate,Married,5900,2.67,11800
25,Male,37,Doctor,Widow,7500,3.45,15000
```

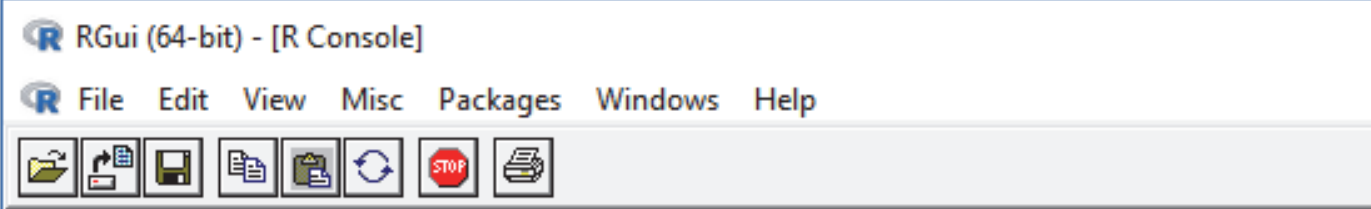
ตัวอย่างการเปิดแฟ้มแบบ *.csv

1. ใช้คำสั่ง `setwd("C:/Rtemp")` เพื่อกำหนดว่าแฟ้มข้อมูลถูกเก็บไว้ที่ directory C:/Rtemp
2. `data2 = read.table("File_001_Comma_delimited.csv", header = TRUE, sep = ",")`
อ่านข้อมูลมาเก็บที่ dataset ชื่อ data2

หมายเหตุ เพราะที่กำหนด directory แล้วจึงไม่ต้องพิมพ์ C:/Rtemp หน้าชื่อแฟ้ม

ตัวอย่างการนำเข้าข้อมูลแบบ *.csv (และเลือกคั่นข้อมูลด้วย comma)

ชื่อ File_001_comma.csv มาใช้ใน Window R Console และเก็บไว้ที่ dataset ชื่อ data2



```

> setwd("C:/Rtemp")
> data2=read.table("File_001_comma.csv",header=TRUE,sep=",")
> data2

```

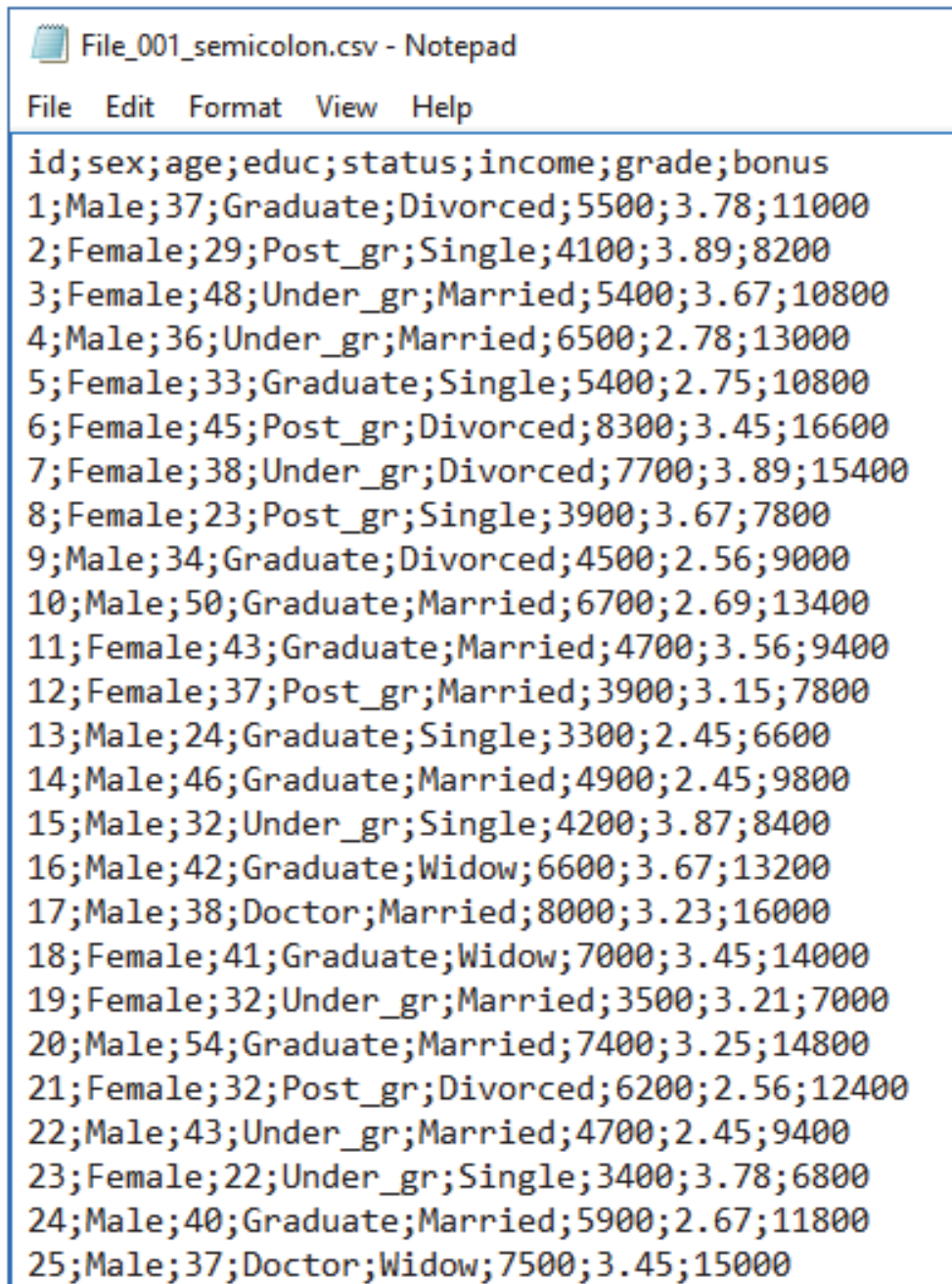
	id	sex	age	educ	status	income	grade	bonus
1	1	Male	37	Graduate	Divorced	5500	3.78	11000
2	2	Female	29	Post_gr	Single	4100	3.89	8200
3	3	Female	48	Under_gr	Married	5400	3.67	10800
4	4	Male	36	Under_gr	Married	6500	2.78	13000
5	5	Female	33	Graduate	Single	5400	2.75	10800
6	6	Female	45	Post_gr	Divorced	8300	3.45	16600
7	7	Female	38	Under_gr	Divorced	7700	3.89	15400
8	8	Female	23	Post_gr	Single	3900	3.67	7800
9	9	Male	34	Graduate	Divorced	4500	2.56	9000
10	10	Male	50	Graduate	Married	6700	2.69	13400
11	11	Female	43	Graduate	Married	4700	3.56	9400
12	12	Female	37	Post_gr	Married	3900	3.15	7800
13	13	Male	24	Graduate	Single	3300	2.45	6600
14	14	Male	46	Graduate	Married	4900	2.45	9800
15	15	Male	32	Under_gr	Single	4200	3.87	8400
16	16	Male	42	Graduate	Widow	6600	3.67	13200
17	17	Male	38	Doctor	Married	8000	3.23	16000
18	18	Female	41	Graduate	Widow	7000	3.45	14000
19	19	Female	32	Under_gr	Married	3500	3.21	7000
20	20	Male	54	Graduate	Married	7400	3.25	14800
21	21	Female	32	Post_gr	Divorced	6200	2.56	12400
22	22	Male	43	Under_gr	Married	4700	2.45	9400
23	23	Female	22	Under_gr	Single	3400	3.78	6800
24	24	Male	40	Graduate	Married	5900	2.67	11800
25	25	Male	37	Doctor	Widow	7500	3.45	15000

3. ในกรณีที่แฟ้มข้อมูลจาก Excel บันทึกเป็นแฟ้มแบบ *.csv (และเลือกคั่นข้อมูลด้วย semicolon)

```
data3 = read.table("C:/Rtemp/File_001_semicolon.csv", header = TRUE, sep = ";")
```

อ่านข้อมูลมาเก็บที่ dataset ชื่อ data3

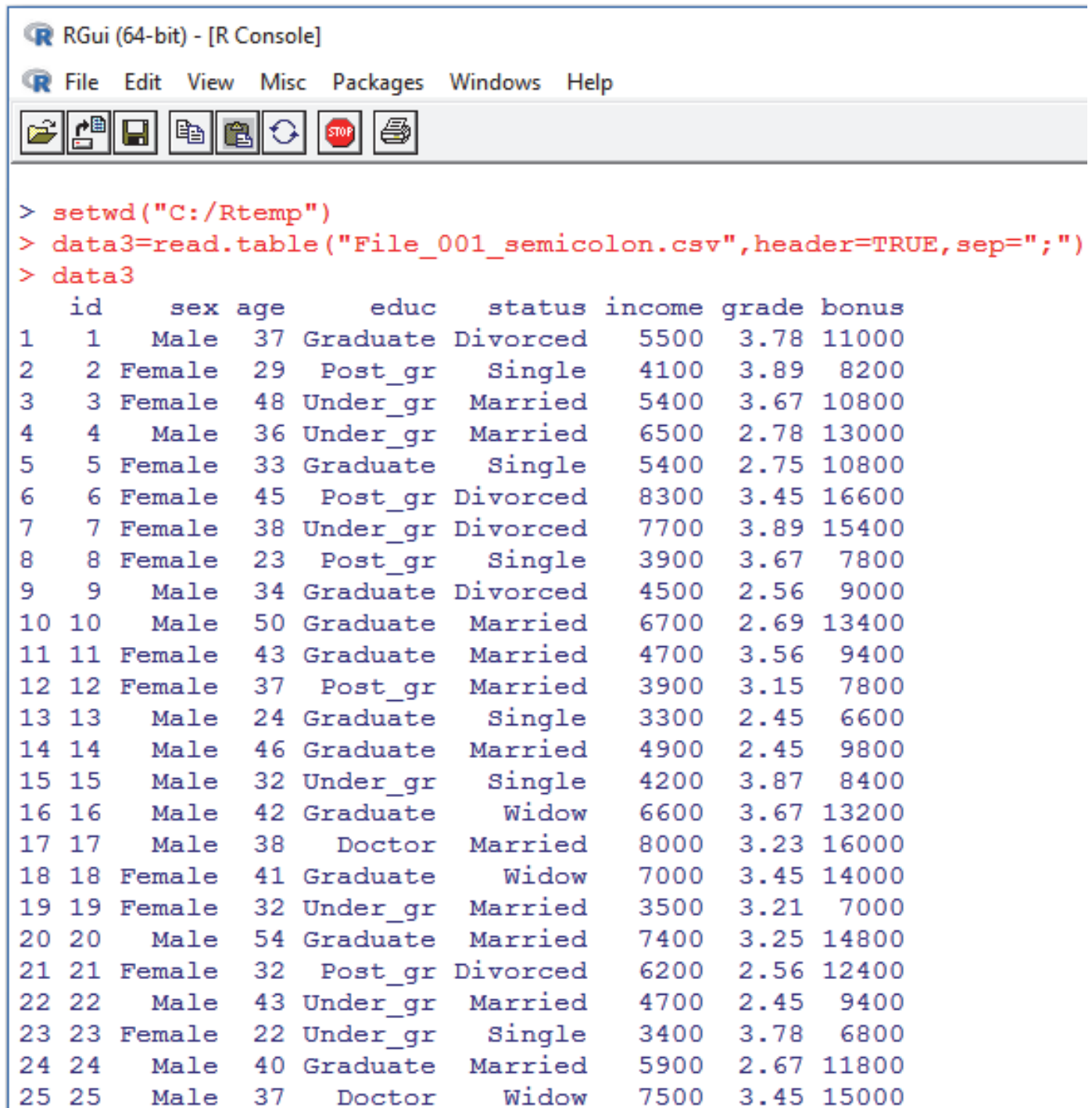
กล่าวคือต้องมี sep = ";" เพื่อระบุว่าข้อมูลแต่ละบรรทัดแยกกันด้วย semicolon



```
File_001_semicolon.csv - Notepad
File Edit Format View Help
id;sex;age;educ;status;income;grade;bonus
1;Male;37;Graduate;Divorced;5500;3.78;11000
2;Female;29;Post_gr;Single;4100;3.89;8200
3;Female;48;Under_gr;Married;5400;3.67;10800
4;Male;36;Under_gr;Married;6500;2.78;13000
5;Female;33;Graduate;Single;5400;2.75;10800
6;Female;45;Post_gr;Divorced;8300;3.45;16600
7;Female;38;Under_gr;Divorced;7700;3.89;15400
8;Female;23;Post_gr;Single;3900;3.67;7800
9;Male;34;Graduate;Divorced;4500;2.56;9000
10;Male;50;Graduate;Married;6700;2.69;13400
11;Female;43;Graduate;Married;4700;3.56;9400
12;Female;37;Post_gr;Married;3900;3.15;7800
13;Male;24;Graduate;Single;3300;2.45;6600
14;Male;46;Graduate;Married;4900;2.45;9800
15;Male;32;Under_gr;Single;4200;3.87;8400
16;Male;42;Graduate;Widow;6600;3.67;13200
17;Male;38;Doctor;Married;8000;3.23;16000
18;Female;41;Graduate;Widow;7000;3.45;14000
19;Female;32;Under_gr;Married;3500;3.21;7000
20;Male;54;Graduate;Married;7400;3.25;14800
21;Female;32;Post_gr;Divorced;6200;2.56;12400
22;Male;43;Under_gr;Married;4700;2.45;9400
23;Female;22;Under_gr;Single;3400;3.78;6800
24;Male;40;Graduate;Married;5900;2.67;11800
25;Male;37;Doctor;Widow;7500;3.45;15000
```

ตัวอย่างการนำแฟ้มข้อมูลแบบ *.csv (และเลือกคั่นข้อมูลด้วย semicolon)

ชื่อ File_001_semicolon.csv มาใช้ใน Window R Console และเก็บไว้ที่ dataset ชื่อ data3

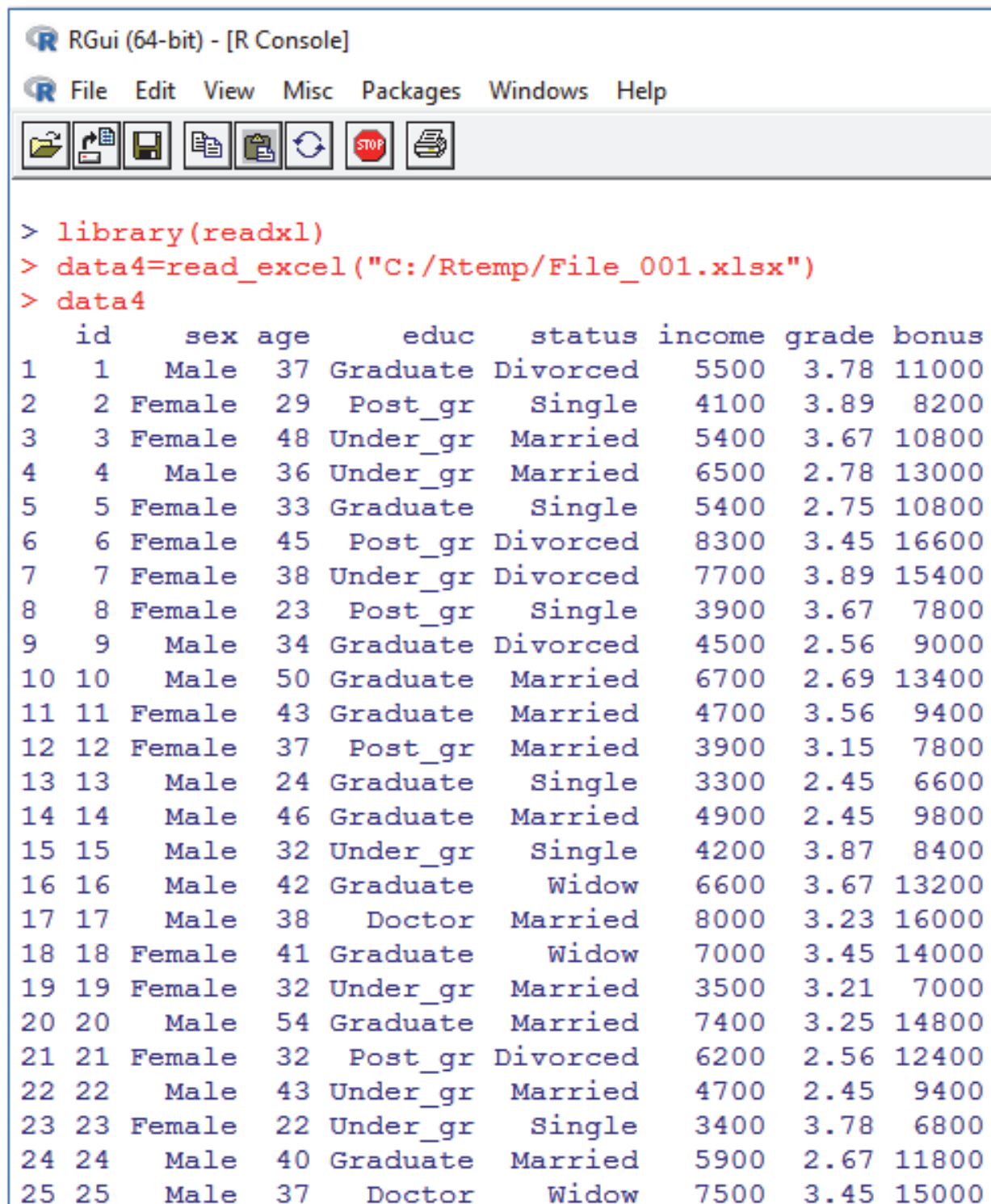


```
> setwd("C:/Rtemp")
> data3=read.table("File_001_semicolon.csv",header=TRUE,sep=";")
> data3
```

	id	sex	age	educ	status	income	grade	bonus
1	1	Male	37	Graduate	Divorced	5500	3.78	11000
2	2	Female	29	Post_gr	Single	4100	3.89	8200
3	3	Female	48	Under_gr	Married	5400	3.67	10800
4	4	Male	36	Under_gr	Married	6500	2.78	13000
5	5	Female	33	Graduate	Single	5400	2.75	10800
6	6	Female	45	Post_gr	Divorced	8300	3.45	16600
7	7	Female	38	Under_gr	Divorced	7700	3.89	15400
8	8	Female	23	Post_gr	Single	3900	3.67	7800
9	9	Male	34	Graduate	Divorced	4500	2.56	9000
10	10	Male	50	Graduate	Married	6700	2.69	13400
11	11	Female	43	Graduate	Married	4700	3.56	9400
12	12	Female	37	Post_gr	Married	3900	3.15	7800
13	13	Male	24	Graduate	Single	3300	2.45	6600
14	14	Male	46	Graduate	Married	4900	2.45	9800
15	15	Male	32	Under_gr	Single	4200	3.87	8400
16	16	Male	42	Graduate	Widow	6600	3.67	13200
17	17	Male	38	Doctor	Married	8000	3.23	16000
18	18	Female	41	Graduate	Widow	7000	3.45	14000
19	19	Female	32	Under_gr	Married	3500	3.21	7000
20	20	Male	54	Graduate	Married	7400	3.25	14800
21	21	Female	32	Post_gr	Divorced	6200	2.56	12400
22	22	Male	43	Under_gr	Married	4700	2.45	9400
23	23	Female	22	Under_gr	Single	3400	3.78	6800
24	24	Male	40	Graduate	Married	5900	2.67	11800
25	25	Male	37	Doctor	Widow	7500	3.45	15000

การเปิดแฟ้มข้อมูลแบบ *.xlsx

1. ใช้คำสั่ง `library(readxl)` เพื่อ Load Package สำหรับอ่านแฟ้มข้อมูล
2. ใช้คำสั่ง `data4 = read_excel("C:/Rtemp/File_001.xlsx")`
อ่านข้อมูลมาเก็บที่ dataset ชื่อ data4
3. พิมพ์ data4 เพื่อแสดงข้อมูลใน data4



```

RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help

> library(readxl)
> data4=read_excel("C:/Rtemp/File_001.xlsx")
> data4
  id  sex age  educ  status income grade bonus
1  1  Male 37 Graduate Divorced   5500  3.78 11000
2  2 Female 29 Post_gr Single    4100  3.89  8200
3  3 Female 48 Under_gr Married   5400  3.67 10800
4  4  Male 36 Under_gr Married   6500  2.78 13000
5  5 Female 33 Graduate Single    5400  2.75 10800
6  6 Female 45 Post_gr Divorced   8300  3.45 16600
7  7 Female 38 Under_gr Divorced   7700  3.89 15400
8  8 Female 23 Post_gr Single    3900  3.67  7800
9  9  Male 34 Graduate Divorced   4500  2.56  9000
10 10  Male 50 Graduate Married    6700  2.69 13400
11 11 Female 43 Graduate Married    4700  3.56  9400
12 12 Female 37 Post_gr Married    3900  3.15  7800
13 13  Male 24 Graduate Single    3300  2.45  6600
14 14  Male 46 Graduate Married    4900  2.45  9800
15 15  Male 32 Under_gr Single    4200  3.87  8400
16 16  Male 42 Graduate Widow    6600  3.67 13200
17 17  Male 38 Doctor Married    8000  3.23 16000
18 18 Female 41 Graduate Widow    7000  3.45 14000
19 19 Female 32 Under_gr Married    3500  3.21  7000
20 20  Male 54 Graduate Married    7400  3.25 14800
21 21 Female 32 Post_gr Divorced    6200  2.56 12400
22 22  Male 43 Under_gr Married    4700  2.45  9400
23 23 Female 22 Under_gr Single    3400  3.78  6800
24 24  Male 40 Graduate Married    5900  2.67 11800
25 25  Male 37 Doctor Widow    7500  3.45 15000
  
```