วารสารวิชาการสาธารณสุข
ปีที่ 33 ฉบับที่ 6 พฤศจิกายน - ธันวาคม 2567

Journal of Health Science of Thailand
Vol. 33 No. 6, November - December 2024

มุมวิธีวิจัย

Methodology Corner

การแสดงภาพข้อมูลในโปรแกรม R

จิรนันท์ ทิพงษ์ วท.บ.
หทัยรัตน์ โกษียาภรณ์ M.S.
มูลนิธิเพื่อการพัฒนานโยบายสุขภาพระหว่างประเทศ

ระพีพงศ์ สุพรรณไชยมาตย์ พ.บ., Ph.D.

การแสดงภาพข้อมูลเป็นวิธีที่ใช้ในการแสดงข้อมูลเชิงลึกโดยใช้ภาพ เช่น แผนภาพ แผนภูมิ แผนที่ ซึ่งจะช่วยให้สามารถเข้าใจข้อมูลปริมาณมากได้อย่างง่ายขึ้น และนำไปสู่การตัดสินใจที่ดีขึ้นเกี่ยวกับข้อมูลนั้น ๆ ในปัจจุบัน โปรแกรม R เป็นหนึ่งในเครื่องมือที่นิยมใช้ในการสร้างข้อมูลภาพ โดยเป็นโปรแกรมสำหรับการคำนวณทางสถิติและกราฟิก เช่น การสร้างแบบจำลองเชิงเส้นและไม่เชิงเส้น รวมถึงวิธีการสร้างกราฟิกต่าง ๆ อย่างไรก็ตาม มีหลากหลายวิธีในการแสดงภาพข้อมูลใน R โดยทั่วไปจะนิยมติดตั้งแพ็กเกจ 'RBase', 'dplyr', 'ggplot2' เป็นต้น โดยก่อนที่จะมีการแสดงภาพ ข้อมูลเหล่านั้นจะต้องอยู่ในรูปแบบไฟล์สกุล 'CSV' ที่มีเฉพาะข้อมูลที่ต้องการแสดงเป็นภาพและควรตรวจสอบว่าชื่อตัวแปรต่าง ๆ มีความสอดคล้องกัน[1,2]

ประเภทของการแสดงข้อมูลภาพโดย R สามารถแบ่งได้ตามข้อมูลที่ต้องการแสดงภาพ ดังนี้

1. แผนภาพแสดงภาพตัวแปรเดียว (univariate graphs) โดยอาจจะเป็นตัวแปรเดียว ตัวแปรแบบหมวดหมู่หรือตัวแปรเชิงปริมาณ เช่น แผนภาพแท่ง แผนภูมิวงกลม หรือ histogram

2. แผนภาพแสดงภาพความสัมพันธ์หรือเปรียบเทียบระหว่างตัวแปรสองตัว (bivariate graphs) เช่น พล็อตเส้น พล็อตกล่อง (box plot) หรือ พล็อตกระจาย (scatterplot)

3. แผนภาพแสดงความสัมพันธ์ระหว่างตัวแปรตามหนึ่งตัวกับตัวแปรต้นหลาย ๆ ตัว (multivariate analysis)

4. แผนภาพแสดงภาพแผนที่

5. แผนภาพแสดงภาพเวลาเพื่อแสดงการเปลี่ยนแปลงเมื่อเวลาผ่านไป โดยแผนภาพที่พบได้บ่อยที่สุดคืออนุกรมเวลา (time series)

6. แผนภาพแบบจำลองทางสถิติ

7. แผนภาพรูปแบบอื่น ๆ เช่น แผนที่ความร้อน (heatmaps) หรือพล็อตกระจายแบบ 3 มิติ

ซึ่งในที่นี้ ได้แสดงตัวอย่างไว้ 2 รูปแบบคือ

1. แผนภาพ histogram เป็นแนวทางที่ใช้กันทั่วไปในการแสดงภาพตัวแปรเชิงปริมาณ โดยทั่วไปจะใช้เพื่อตรวจสอบการกระจายข้อมูลที่เท่ากันและสมมาตร รวมถึงระบุความเบี่ยงเบนไปจากค่าที่คาดหวัง โดยมีตัวอย่างฟังก์ชัน ดังนี้

```
ggplot(df, aes(x = xx, fill=yy)) +
  geom_histogram(color="blue", binwidth=2)+
  xlab("aa")+ylab("bb")+ ggtitle("cc")
```

df ในที่นี้คือชุดข้อมูลที่ต้องการแสดง

xx ในที่นี้คือตัวแปรที่ต้องการแสดงในแกน X

fill ในที่นี้คือการกำหนดให้สีของแถบในฮิสโตแกรมถูกแยกตามค่าของตัวแปร yy

color ในที่นี้คือสีของเส้นขอบของแถบในฮิสโตแกรม โดยพิมพ์ชื่อสีในภาษาอังกฤษที่ต้องการ

binwidth ในที่นี้คือค่าของการกำหนดความกว้างของแต่ละช่วงข้อมูล

aa, bb ในที่นี้คือชื่อของแกน X และ Y

cc ในที่นี้คือชื่อหัวข้อของแผนภาพ

2. พล็อตกล่อง คือใช้เพื่อให้คำอธิบายทางสถิติที่ครอบคลุมของข้อมูลผ่านภาพ และระบุจุดผิดปกติที่ไม่อยู่ในช่วงระหว่างควอไทล์ของข้อมูล โดยจะแสดงข้อมูลจุดข้อมูลต่ำสุดและสูงสุด ค่ามัธยฐาน ควอไทล์ที่หนึ่งและสามและช่วงระหว่างควอไทล์ โดยมีตัวอย่างฟังก์ชัน ดังนี้

```
ggplot(df, aes(x =xx, y = yy), fill=zz) +
  geom_boxplot(color="black") +xlab("aa")+
  ylab("bb")+ggtitle("cc")
```

โดย df ในที่นี้คือชุดข้อมูลที่ต้องการแสดง

xx, yy คือตัวแปรที่ต้องการแสดงในแกน X, Y

aa, bb ในที่นี้คือชื่อของแกน X และ Y

cc ในที่นี้คือชื่อหัวข้อของแผนภาพ

color ในที่นี้คือสีของกล่องแผนภาพ โดยพิมพ์ชื่อสีในภาษาอังกฤษที่ต้องการ

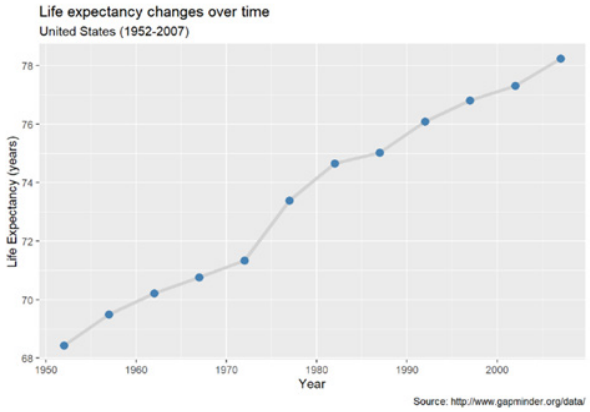
ตัวอย่างการแสดงแผนภาพเส้นและการปรับแต่งเบื้องต้นของข้อมูลความสัมพันธ์ระหว่างเวลา (ปี) และอายุขัยในประเทศสหรัฐอเมริการะหว่างปี 1952 ถึง 2007 (ภาพที่ 1) โดยมีรายละเอียดคำสั่ง ดังนี้

```
library(ggplot2)
ggplot(df, aes(x = year, y = lifeExp)) +
  geom_line(size = 1.5, color = "grey" +
  geom_point(size = 3, color = "lightblue") +
  labs(y = "Life Expectancy (years)", x = "Year",
  title = "Life Expectancy changes over time")
```

โดย size คือ ขนาดเส้นแผนภาพ color คือ สีของเส้นแผนภาพ line คือ เส้นข้อมูลในแผนภาพ point คือ จุดข้อมูลในแผนภาพ labs คือ ชื่อในแนวแกน x และ y และ title คือ ชื่อแผนภาพ

โดยสรุปชุดคำสั่งมักจะนิยมใช้แพ็กเกจ 'ggplot2' ซึ่งจะต้องมีการเรียกใช้ชุดข้อมูล (library) และกำหนดตัวแปรในแกน X และ Y ในชุดข้อมูล (data frame; df) เพื่อแสดงภาพข้อมูลนั้น ๆ โดยในรายละเอียดอาจจะมีความแตกต่างกันไปตามประเภทของการแสดงข้อมูลภาพ

ภาพที่ 1 แผนภาพความสัมพันธ์ระหว่างเวลาและอายุขัยในประเทศสหรัฐอเมริการะหว่างปี 1952 ถึง 2007



อย่างไรก็ตามสามารถใช้คำสั่ง 'function_name' เพื่อขอรับความช่วยเหลือหรือดูรายละเอียดอื่น ๆ เพิ่มเติมของแต่ละชุดคำสั่ง สำหรับประโยชน์ของการแสดงภาพข้อมูลในโปรแกรม R คือ สามารถนำเสนอการแสดงภาพในรูปแบบที่หลากหลายตามลักษณะของข้อมูลและวัตถุประสงค์ในการสื่อสาร อีกทั้งยังสามารถปรับแต่งการแสดงข้อมูลได้ตามความต้องการ เช่น การเปลี่ยนแกนแบบอักษร คำอธิบายประกอบหรือป้ายกำกับ นอกจากนี้แพ็กเกจในการใช้งานยังมีการพัฒนาเป็นระยะ รวมถึงมีคำแนะนำจากชุมชนออนไลน์เกี่ยวกับปัญหาการใช้งาน อย่างไรก็ตามการแสดงข้อมูลภาพในโปรแกรม R ยังมีข้อจำกัดที่ผู้วิจัยต้องทราบแพ็คเกจที่ถูกต้อง และแต่ละแพ็คเกจมีลักษณะการเขียนคำสั่งที่จำเพาะและใช้เวลาค่อนข้างมากสำหรับข้อมูลจำนวนมากเมื่อเปรียบเทียบกับโปรแกรมอื่น ๆ ที่สร้างมาเพื่อการนำเสนอข้อมูลโดยตรง

เอกสารอ้างอิง

1. Zhou A. Data Visualization in R [Internet]. 2022 [cited 2023 Oct 20]. Available from: https://www.hsph.harvard.edu/mcgoldrick/wp-content/uploads/sites/2488/2022/09/Data-Visualization-in-R.pdf
2. Kabacoff R. Modern data visualization with R [Internet]. 2015 [cited 2023 Dec 4]. Available from: https://rkabacoff.github.io/datavis/