

תרגיל בית 1ג' (תרגול ב-R)

היכרות ראשונית עם R

בתרגיל זה תתקינו את R, את חבילת tidyverse ותתרגלו מעט עבודה עם נתונים.

חלק ראשון

התקינו את תוכנת R ואת RStudio.

את תוכנת R ניתן להוריד בקישור הבא:

(<https://cran.r-project.org/bin/windows/base/>) (<https://cran.r-project.org/bin/windows/base/>)

את RStudio ניתן להוריד בקישור הבא:

(<https://www.rstudio.com/products/rstudio/download/#download>) (<https://www.rstudio.com/products/rstudio/download/#download>)

שימו לב להשתמש בגרסאות המתאימות לכם (יש גרסאות שונות ל-Linux, Windows, macOS)

כעת לאחר שהתקנתם, הפעילו את RStudio והתקינו את חבילת tidyverse על ידי הקלדה ב-Console של:

```
install.packages("tidyverse")
```

זה עשוי לקחת קצת זמן, אז התאזרו בסבלנות.

אם קיבלתם הודעת שגיאה כגון: *Error in install.packages : cannot open file* זה כנראה נובע מבעיות הרשאה לתיקייה אליה R מנסה להתקין את החבילה. אפשר לפתור את הבעיה הזו על ידי העלאת RStudio עם כפתור ימני של העכבר, ואז Run as Administrator (ואז לנסות להתקין מחדש).

חלק שני

התקינו את חבילת palmerpenguins. השתמשו באותה הפקודה בה השתמשנו בחלק הראשון לצורך התקנת חבילה (רק שנו את השם tidyverse לשם של החבילה palmerpenguins).

כעת השתמשו בקוד הבא ותארו מהם הנתונים הקיימים בטבלה penguins:

```
library(tidyverse)
library(palmerpenguins)

glimpse(penguins)
```

הריצו גם את הפקודות הבאות וכתבו מה ההבדל בין הפונקציה view, tail, head, glimpse, ולרשום penguins ב-console.

```
head(penguins)
tail(penguins)
view(penguins)
penguins
```

• מה מתארת כל שורה בטבלה?

• מה מתאר המשתנה island?

- מה מתאר המשתנה species?

השתמשו בפונקציה count בדומה לאופן המתואר להלן, וענו על השאלות:

- כמה איים מתועדים בטבלה?
- כמה זכרים מתועדים בטבלה?
- לכמה תצפיות חסרים נתונים זכר/נקבה?

דוגמה לשימוש בפונקציה count (תצטרכו להרחיב את הדוגמה הזו על ידי שינוי הקוד)

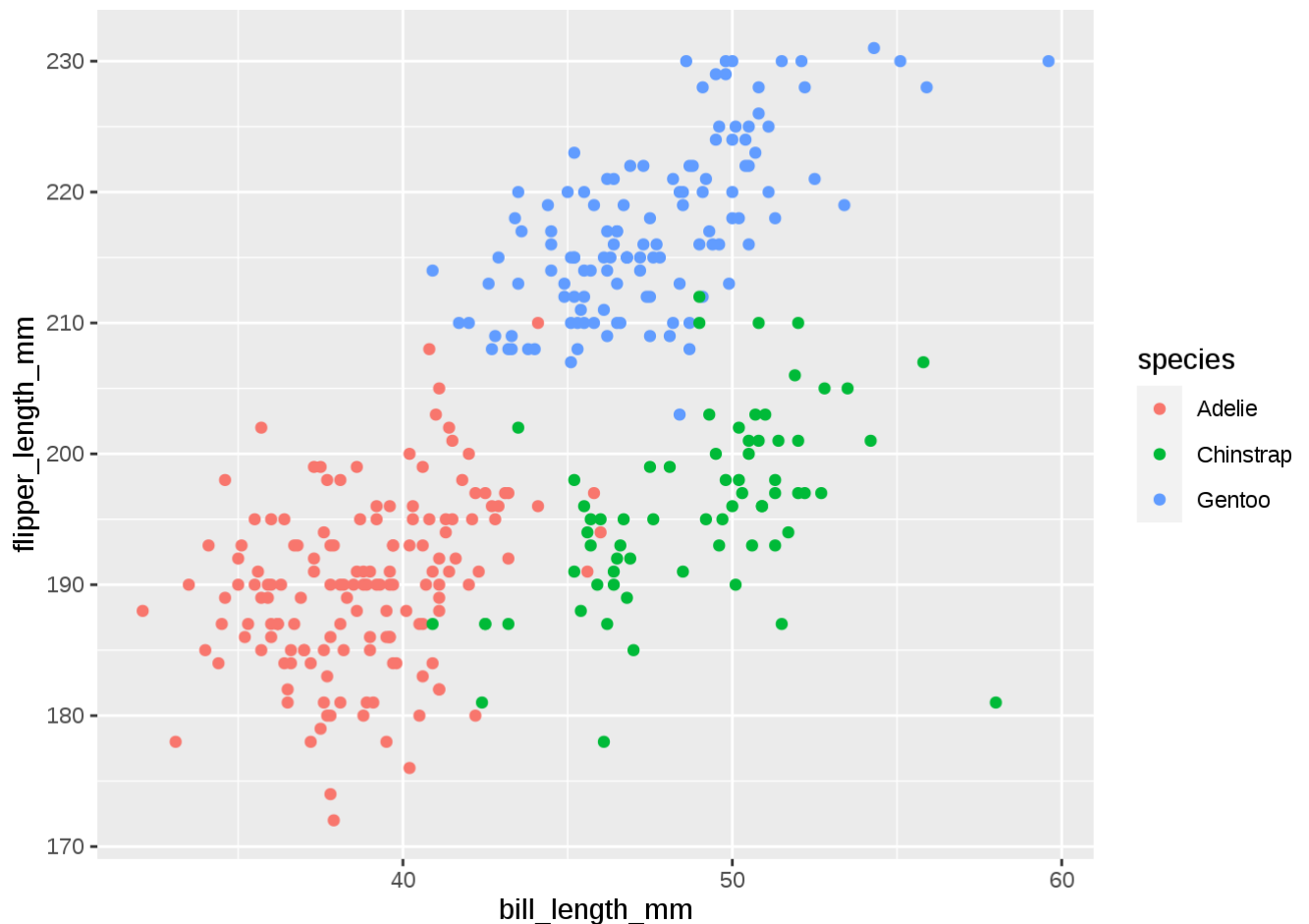
```
penguins %>%  
  count(species)
```

חלק שלישי

הקוד הבא מצייר תרשים של הקשר בין אורך מקור ואורך כנף.

```
ggplot(data = penguins,  
       aes(x = bill_length_mm,  
           y = flipper_length_mm,  
           color = species)) +  
  geom_point()
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



1. מה אומרת הודעת האזהרה של R? מה המשמעות שלה? (...Removed 2 rows containins)
2. הסבירו מה הקשר שמתאר התרשים בין שלושת המשתנים שבו (שני הצירים וצבע הנקודות).
3. עבור כל פונקציה שהשתמשנו בה, הסבירו מה התפקיד שלה ביצירת התרשים (להלן הפונקציות שבהן השתמשנו, `ggplot`, `aes`, `geom_point`). באפשרותכם להיעזר בעמוד הזה (<https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf>).
4. כתבו קוד אשר יצייר תרשים שמתאר את הקשר בין אורך המקור, עומק המקור, והאי שבו נמדדה הדגימה (לשם כך היעזרו בקוד של התרשים הקודם, ושנו אותו כך שיתאים לצרכים).
5. נתחו את התרשים: מה הקשר שאתם מבחינים בו? מדוע האי שבו נמדדה הדגימה משפיע על הקשר בין האורך והעומק של המקור?