



R you ready?

היכרות קצרה עם שפת R

הצגה לקבוצת InsightsIL



אוקטובר 2020

מה היא R?

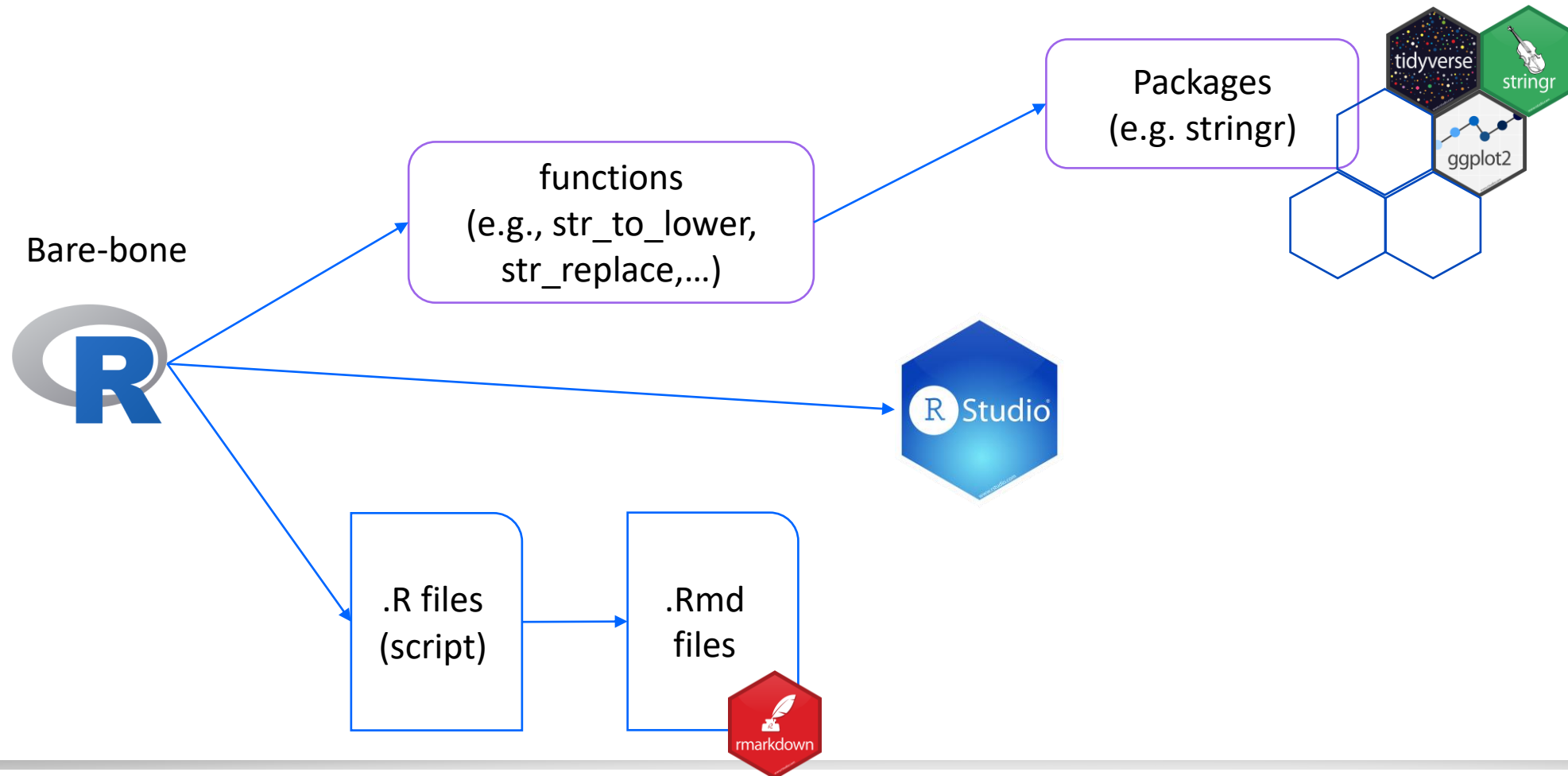
- שפת תכנות, בקוד פתוח, המיועדת לניתוח נתונים וסטטיסטיקה
 - בפועל זו שפה כל כך מגוונת, שאפשר לעשות בה כמעט הכל
- ההיסטוריה של R טמונה בשפה קודמת שנקראה S (משנות ה-70)
- R החלה להתפתח בשנות ה-90 כ-Spinoff של S, אבל עברה אותה ביכולות ובפופולריות

- עד סוף הוובינר נבין:

- איפה נמצאת R "במרחב" הכלים הזמינים לניתוח נתונים
- מה היתרונות של R ביחס לכלים אחרים
- מה היכולות של R (ואיך היא יכולה לשרת אתכם) + הדגמה!
- איך מתחילים, איפה ואיך ללמוד עוד



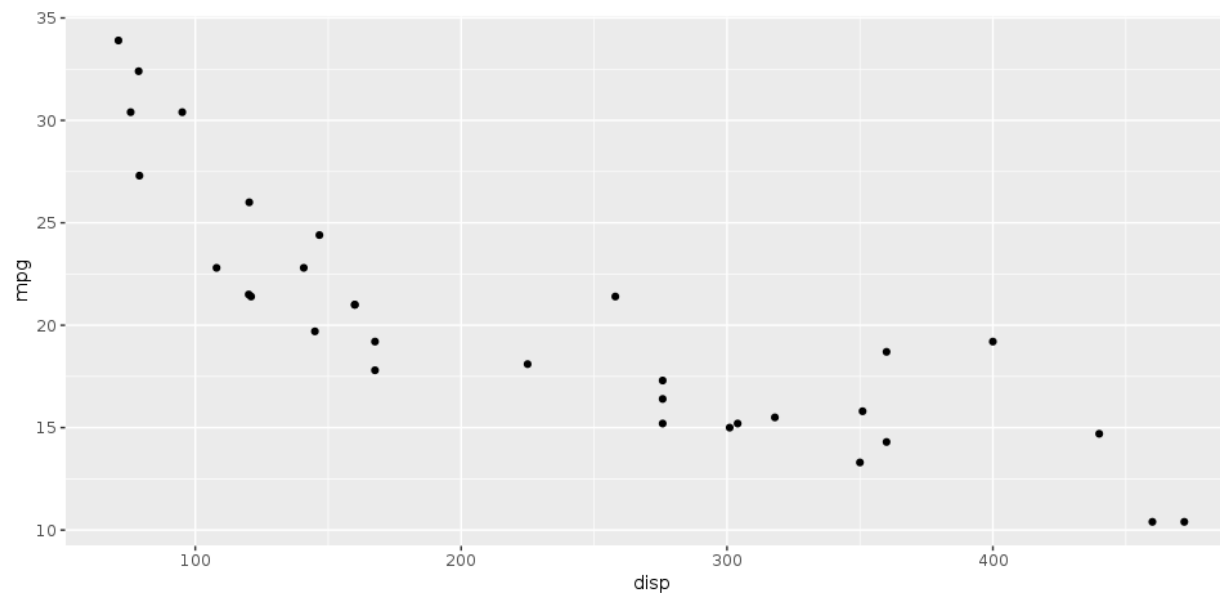
מה נמצא בפנים?



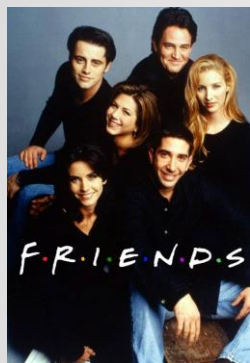
זה אולי נשמע מפחיד, אבל זה ממש לא! נחשוב על זה רגע כמו "מחשבון מתוכם"

```
> a <- 5
> b <- 6
> a + b
[1] 11
> sqrt(2)
[1] 1.414214
> pi
[1] 3.141593
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> mean(1:10)
[1] 5.5
```

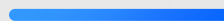
```
> mtcars
...
> library(ggplot2)
> ggplot(mtcars, aes(x = disp, y = mpg)) + geom_point()
```



הדגמה מספר 1



מי מדבר הכי הרבה ב-Friends?



איפה נמצאת שפת R ביחס לכלים אחרים לניתוח נתונים? (אילוסטרציה)

גודל קהילה

• הבעיה עם כלים סטטיסטיים "מהדור הישן" כמו SPSS, SAS, STATA (ובמידת מה גם Matlab) היא שהם:

- קוד סגור (פיתוח סגור שאין אליו גישה) – לא יכולים לעמוד בקצב בהשוואה לכלים פתוחים
- רובם מעודדים "עצלנות" - שימוש ב"ממשק משתמש" שמייצר בעיה קשה של Reproducibility

• הבעיה עם אקסל היא שהוא לא מתאים לניתוח נתונים

• גם R וגם Python הם כלים נפוצים, בקוד פתוח, עם קהילה מאוד עשירה, שמתאימים לניתוח נתונים ו-Data Science



התאמה לניתוח סטטיסטי

למה אקסל לא מתאים לניתוח נתונים?

• הוא מעודד עצלנות – הכל קורה בממשק משתמש, הנוסחאות בתוך התאים

• קשה לתעד בצורה מסודרת

• קשה להתחקות אחרי חישובים בקבצים מסובכים

• נהיה "כבד" מהר מאוד

• וחוץ מזה יש גם את זה:

• How Excel may have caused **loss of 16,000 test result**

• מגבלה של מספר שורות בקובץ גרמה לכך שתוצאות "הלכו לאיבוד"

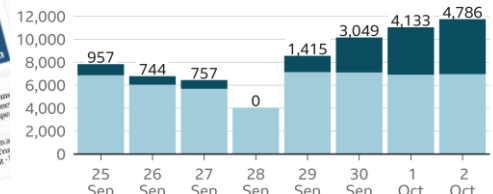
• (בעצם זה לא בעיה באקסל אלא בחירה לקויה של המשתמשים לעבוד עם אקסל)



Thousands of missing coronavirus cases added after reporting problem

Number of new coronavirus cases by date reported

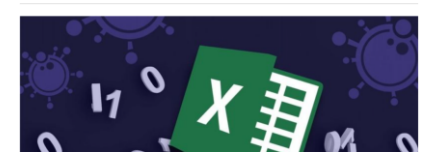
■ Missing cases added ■ Previously announced cases



Source: Gov.uk dashboard, Public Health England

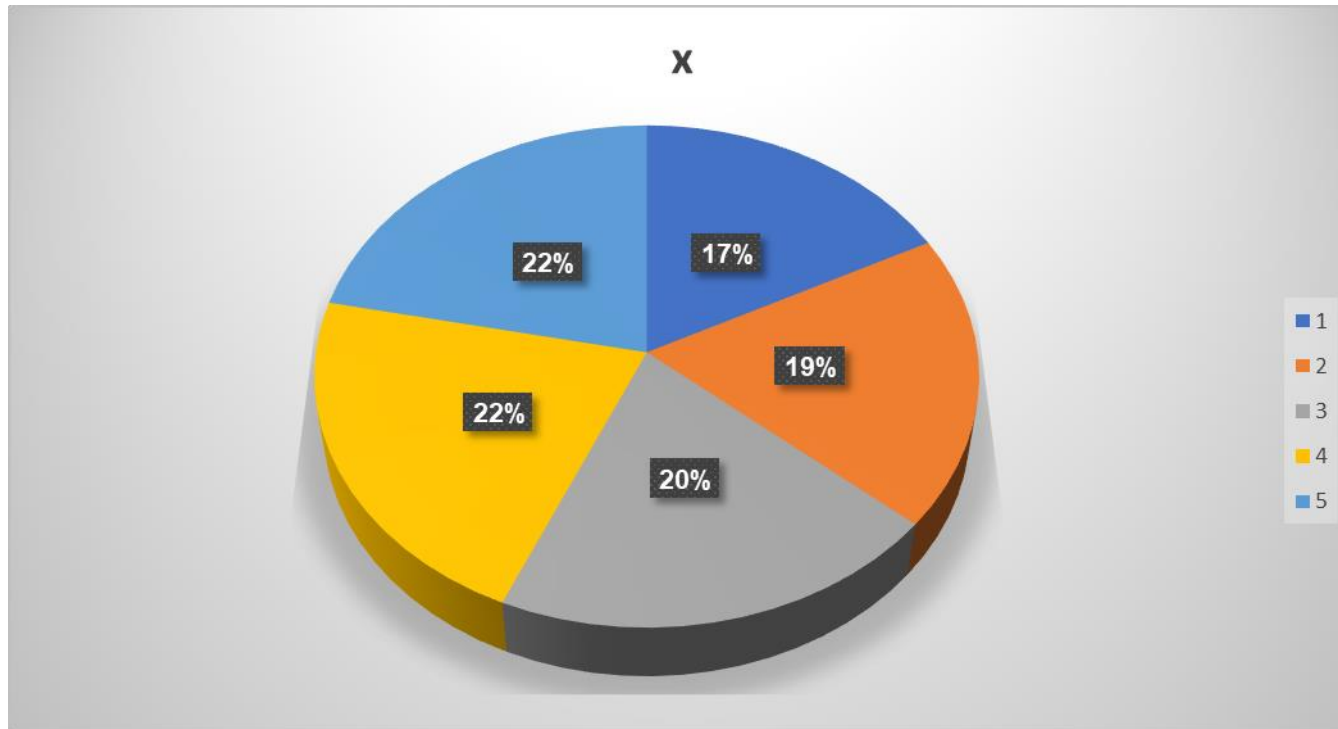


Tech
Excel: Why using Microsoft's tool caused Covid-19 results to be lost
By Leo Kroll
Technology desk editor
4 days ago
Coronavirus pandemic



למה אקסל לא מתאים לניתוח נתונים?

ואת זה:



ויש גם את זה:

A	
קבוצת גיל	1
61+	2
46-60	3
36-45	4
26-35	5
13-25	6
05-דצמ	7
01-אפר	8
	9

יתרונות בשימוש ב-R ביחס ל...

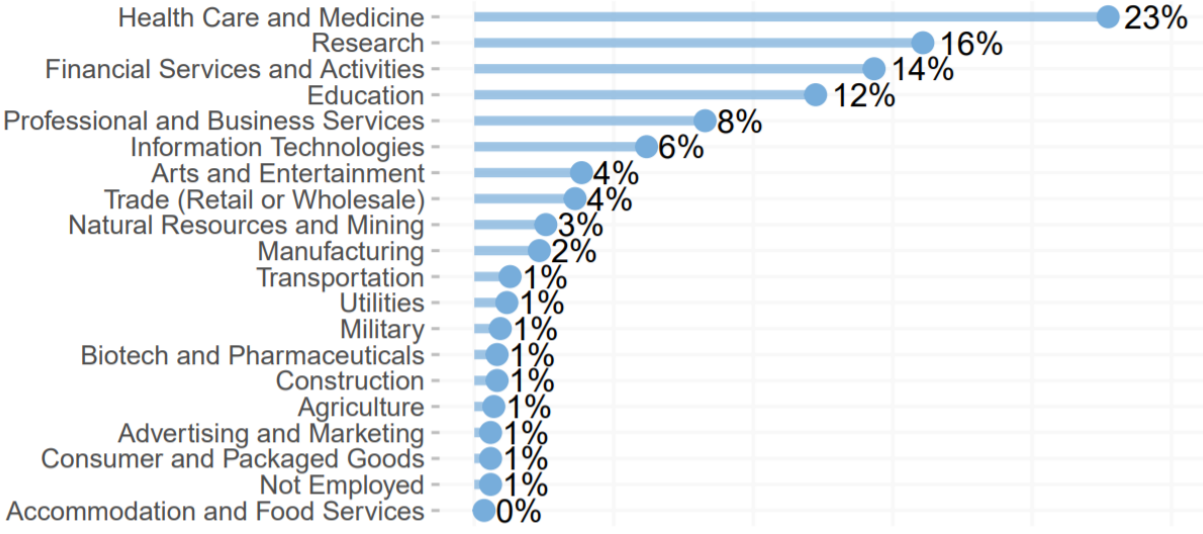
- בכל מקרה של ניתוח נתונים יש לפחות 2 אנשים שיעברו עליו
 - אני ואני בעוד כמה חודשים/שנים
 - כתיבה בקוד עוזרת לתעד ולחזור על הפעולה באופן מדויק וזהה
- במילים אחרות – התוצר הוא Reproduceable
 - יותר קל לעשות בקרה
 - התוצרים הם Scalable
- Open Source לעומת פיתוח סגור
 - קהילה עצומה – תמיד יש עם מי לדבר, איפה לשאול שאלות,
 - כל השיטות החדישות מגיעות כחבילות יחסית מהר
- אבל... עבודה עם קוד היא יותר דורשת (הכשרה והדרכה, עקומת למידה)

למי R מתאימה?

- לכולם! אבל אני לא אומר את זה סתם...
- בתחילת השנה (עוד לפני הקורונה) העברתי סדנא בכנס RStudio בסן פרנסיסקו
- הנה התפלגות המשתתפים בכנס מבחינת תפקידים ותעשייה

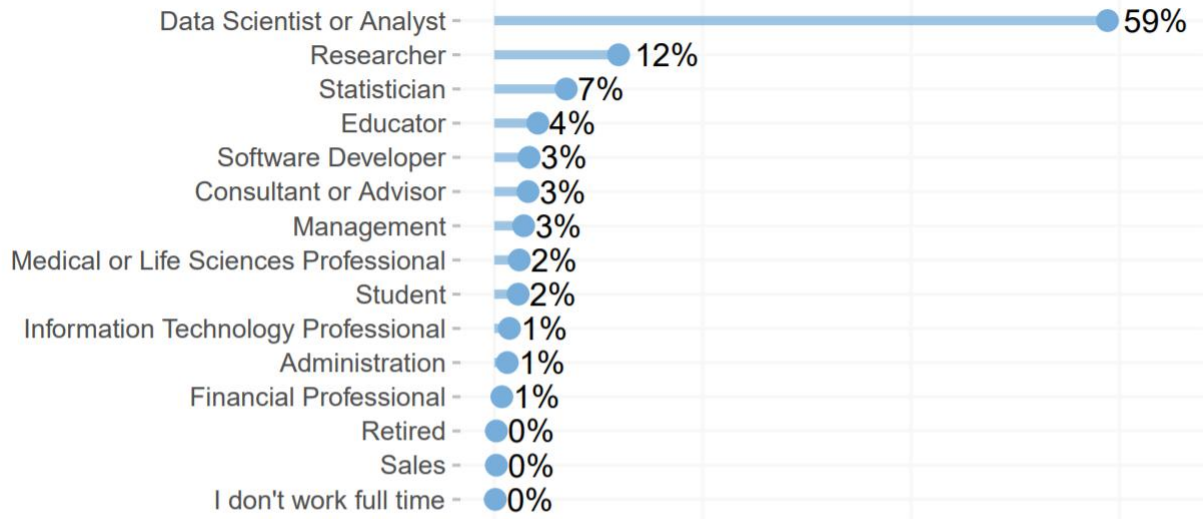
All Workshops

"What industry do you work in?"



Attendee survey responses, n = 858

"What category best describes the work you do?"



Attendee survey responses, n = 955

מה אפשר לעשות ב-R?

- ראינו דוגמה לניתוח נתונים פשוט
- Data Science
- ייצור של מצגות וכתיבת דוחות
- ייצור מסיבי של תוצרים (אוטומציות)
- פיתוח אפליקציות שמנגישות את התוצרים



איך מתקינים

- קודם כל – הכל בחינם (קוד פתוח)
- כל מה שנראה לא עולה גרוש ומותר לשימוש מסחרי
- [מתקינים R](#) (הגרסה האחרונה היא 4.0.3)
- [מתקינים Rstudio Desktop](#) (סביבת פיתוח ל-R, גרסה עדכנית 1.3)
- יש גם אפשרות להתנסות עם הממשק בענן [rstudio.cloud](#) (לא ממליץ על זה)



הדגמה מספר 2 אוטומציה של מצגות



הדגמה מספר 3

אפליקציה

איפה לומדים עוד?

- אני כותב ספר בעברית (כרגע ממש בהתחלה שלו) פרקים מתעדכנים פה:
<https://hebrewr.co.il>
- אנחנו נותנים קורסים מקצועיים בארגונים שמעוניינים להטמיע את R
 - הדרכות, שיטות, וליווי מקצועי
 - בקרוב תיפתח הרשמה לקורס הקרוב, בנובמבר, פרטים יופצו בקבוצה
- מקורות בעברית – קבוצת "R להמונים" בפייסבוק
- מקורות באנגלית
 - ספר מצויין <https://r4ds.had.co.nz/>
 - הקלטות של ניתוח נתונים ב-Youtube
 - [דיוויד רובינסון](#)
 - הדרכות Tutorials בתוך Rstudio
 - בלוגים [r-bloggers](#)



Thanks!

Adi Sarid

@SaridResearch 

<https://www.sarid-ins.co.il>