

# DATA SCIENCE

## R for Data Science III

**Tomas Karpati MD**

[tc.datascience@gmail.com](mailto:tc.datascience@gmail.com)

054-2002430



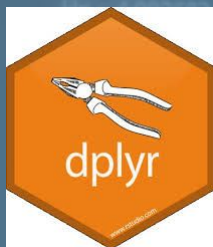
# פונקציות חיצוניות: R

dplyr

regex

יצירת פונקציות

יבוא/יצוא



select



filter



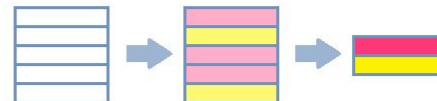
arrange



mutate



summarise



# תרגול

## תרגיל:

1. תריצו את הפקודה הבאה

```
nasa1 <- as_data_frame(nasa)
```

1. אם הטבלה שנוצרה תייצרו טבלה חדשה ברצף פקודות של dplyr לפי התנאים הבאים:

a. תבחרו רק את הרשומות שה-lat בין 29.56 ל-33.09

ו-long בין -110.93 ל- -90.55

b. תייצרו משתנה חדש המראה את היחס בין

temperature לבין ה-surftemp

c. תייצרו אגרגציה לפי שנה שמראה את הממוצע וסטיית

התקן של ozone, pressure ו-המשתנה שייצרתם ב-b

d. תסדרו את התוצאות לפי סדר יורד של ממוצעת ה-ozone

<code/>



# פונקציות חיצוניות R:

dplyr

regex

יצירת פונקציות

יבוא/יצוא

`grep(<exp>,<obj>,value=T)`

`grepl(<exp>,<obj>,value=T)`

`sub(<exp>,<substitute>,<obj>)`

`gsub(<exp>,<substitute>,<obj>)`

`regexpr(<exp>,<obj>)`

`gregexpr(<exp>,<obj>)`

`regexexec(<exp>,<obj>)`

## Regular- Expression Patterns

- `^` Matches beginning of line.
- `$` Matches end of line.
- `.` Matches any single char except newline.
- `[...]` Matches any single char in brackets.
- `[^...]` Matches any single char not in brackets.
- `\w` Matches word characters.
- `\W` Matches nonword characters.
- `\s` Matches whitespace.
- `\S` Matches nonwhitespace.
- `\d` Matches digits.
- `\D` Matches nondigits.
- `\A` Matches beginning of string.
- `\Z` Matches end of string.
- `\z` Matches end of string.
- `\G` Matches point where last match finished.
- `x|y` Matches either x or y.
- `[0-9]` Match any digit; same as `[0123456789]`
- `[a-z]` Match any lowercase ASCII letter
- `[A-Z]` Match any uppercase ASCII letter
- `[a-zA-Z0-9]` Match any of the above
- `[^aeiou]` Match any other than a lowercase vowel
- `[^0-9]` Match anything other than a digit.

# פונקציות חיצוניות: R

dplyr

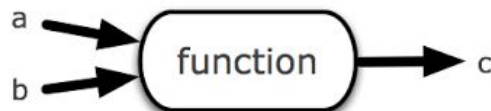
regex

יצירת פונקציות

יבוא/יצוא

פונקציות הם עצמים הניתנים לקריאה כדי לבצע פעולה כלשהו. המשתמש של הפונקציה צריך לקרוא אותה בשמה ולהוסיף את הפרמטרים הנדרשים.

- כדי ליצור פונקציה בעצמנו, עלינו לוודא ש:
1. אנו מוסיפים בתוך הפונקציה קריאה לספריות הנדרשות
  2. לדאוג על טיפול בפרמטרים הנדרשות לפונקציה
  3. לדאוג על טיפול בטעויות שעלולות לקרוא



# פונקציות חיצוניות: R

dplyr

regex

יצירת פונקציות

יבוא/יצוא

סוגי מקורות הנתונים הפופולריים הם:

a. CSV

b. אקסל

c. SPSS/SAS

d. Html / XML

e. Json

f. SQL

