כריית נתונים ב- R // תרגיל בית מספר 2

בתרגיל זה נשתמש בקובץ נתונים BostonHousing.csv המכיל מידע על שכונות בבוסטון, ובכללו חציון מחיר בתים בשכונה, רמת פשיעה בשכונה, מצב סוציו-אקונומי, ועוד.

הקובץ נלקח מתוך המאמר:

Harrison and Rubinfeld, "Hedonic prices and the demand for clean air", *Journal of Environmental Economics & Management*, vol. 5, pages 81-102, 1978.

להלן תיאור משתני המסד:

שם משתנה	תיאור
CRIM	אחוז פשיעה
ZN	אחוז שטחים פרטיים הגדולים מ 25,000 ft^2
INDUS	אחוז שטח שאינו מסחרי
CHAS	האם ממוקם ליד נהר צ'רלס (1 אם כן)
NOX	ריכוז תחמוצת החנקן
RM	מספר חדרים ממוצע לבית
AGE	אחוז דירות שנבנו לפני 1940
DIS	מרחק משוקלל ממרכז בוסטון
RAD	גישה לכבישים היקפיים
TAX	ערך ארנונה
LSTAT	אחוז אוכלוסייה במצב סוציו אקונומי נמוך
MEDV	חציון ערך בתים באלפי דולרים
CAT.MEDV	האם MEDV גדול מחציון הבתים הכללי שערכו 30,000$ (1 אם כן)

עבור כל אחד מהקשרים הבאים, בנה את הגרף המתאים, **והסבר את אופי הקשר בין המשתנים**.
יש לעבוד עם גרפים מספריית ggplot2. וודא כי הגרף מסוגנן היטב: **שמות צירים, כותרת לגרף, log-scale במידת הצורך** (2 נקודות לסעיף, תורד נקודה אחת על שימוש בגרפים שאינם מספריית ggplot2):

1. הקשר בין MEDV ל- NOX.
2. הקשר בין CHAS **לממוצע** MEDV.
3. היסטוגרמה של מספר החדרים הממוצע לבית (RM).
4. **התפלגות (distribution)** של CRIM לפי CAT.MEDV (למדנו מספר דרכים להציג התפלגות יש להשתמש באחת מהן).

אופן הגשה:

- ✓ הגשה דרך אתר למידה
- ✓ הגשה בזוגות או ביחידים (רק אחד מבני הזוג צריך להגיש. על הקובץ יופיעו השמות של שני המגישים)
- ✓ יש להגיש קובץ R מתועד, וקובץ פלט עם הגרפים **והסבר קצר על כל גרף**.