כריית נתונים ב- R // תרגיל בית מספר 1

בתרגיל זה נשתמש במסד ה- IMDB, המכיל את שני קבצי הנתונים שהוצגו בכיתה:

IMDB_movies.csv

IMDB_players

1. (ללא נקודות) המר את משתנה budget למשתנה מספרי. כאשר ערך ה- budget לא קיים, השאר את הערך NA.

דוגמא:

id	budget	numericBudget
10itemsorless.htm	N/A	N/A
10thandwolf.htm	$8 million	8000000
127hours.htm	$18 million	18000000
12rounds.htm	N/A	N/A
13assassins.htm	N/A	N/A

- ✓ אפשר להמיר באקסל
- ✓ למתקדמים: השתמש בפונקציות מספריית stringr (או כל ספרייה שימושית אחרת).

2. (2 נק') חשב קורלציה בין budget ל- total gross, עבור סרטים להם התקציב מדווח.
הדפס את הקורלציה לפלט.

3. (3 נק') חשב מספר בעלי תפקידים (player.id במסד IMDB_players) מכל סוג (actor, director וכו) לפי סרט. הכנס את החישוב למשתנה מסוג data.frame בשם n.actors.
אתגר למעוניינים: השתמש בפונקצית cast בסיפרית reshape לצורך החישוב.
הדפס את 6 השורות הראשונות של n.actors.

דוגמא לשורת פלט:

```
> head(n.actors)
                id Actor Cinematographer Composer Director Producer Writer
1       10000bc.htm     0               0        0        1        2      2
```

4. (5 נק') נרצה לחשב האם יש הבדלים (בממוצע) בין רווחי סרטים (total gross) להם יש דיווח של תקציב (budget), לאלו שאין להם.
   - ✓ איזה מבחן יש להריץ? הדפס את שם המבחן לפלט.
   - ✓ הרץ את המבחן המתאים והדפס לפלט את תוצאת המבחן.

- ✓ הדפס לפלט את המסקנה במילותיך.

<u>אופן הגשה</u>:

- ✓ הגשה דרך אתר למידה
- ✓ יש להגיש קובץ R **מתועד**, וקובץ **פלט** בפורמט word/ pdf המכיל את התשובות לשורות המסומנות בצהוב