מעבדה לסטטיסטיקה 2024, מטלה 1
הצגה ב21.5

במשימה זו נחקור מקטע של 10 מיליון בסיסים בreference genome של כרומוזום אנושי. המטרה היא לראות כיצד מפוזרים הבסיסים השונים בכרומוזום.

א. קראו את הכרומוזום 1 כפי שהראינו בכיתה. עבדו על האזורים הבאים:
   a. קבוצות A בין בסיס 10 מיליון ל30 מיליון
   b. קבוצות B בין בסיס 30 ל50 מיליון
   c. קבוצות C בין בסיס 50 ל70 מיליון

ב. כתבו פונקציה שמקבלת אזור של הגנום גודל תא ומסכמת את מספר הבסיסים בכל תא. עבור גודל תא של 1000 בסיסים, חשבו את מספר הופעות כל אחד מהבסיסים בתאים.

ג. הציגו את ההתפלגויות עבור כל אחד מהבסיסים (A,C,G,T) בעזרת היסטוגרמה.

ד. ציירו את השכיחות של כל אחד מהבסיסים מול המיקום בכרומוזום. האם הקווים דומים או שונים? נסו להציג ביחד בצורה שמאפשרת להשוות.

ה. ציירו scatter-plot עבור השכיחות של זוג בסיסים באותו התא. חזרו עבור כל זוג בסיסים.
   a. מהי המגמה?
   b. האם יש קבוצת תאים יוצאת דופן?
   c. מצאו איפה תאים אלו נמצאים על הכרומוזום; האם יש משהו מיוחד ברצפים באזורים אלו?

ו. אם הייתם צריכים לחלק את 4 סוגי הבסיסים {A,C,G,T} לשני זוגות בסיסים שמתנהגים "דומה", כיצד הייתם מחלקים? הסבירו.

<u>הערות:</u>
חשבו על עיצוב הגרפים. תנו כותרת לצירים, שימו לב לאורך הצירים.
השתמשו בצבעים, עובי נקודה, וכו' כדי להדגיש נקודות חשובות.