

# Car Accidents

Victor Khodjaev, Eyal Amar, Roei Zaady, Shalev Ezran, Omer Okun

בחרנו לבצע את עבודתנו על תאונות דרכים במדינת ישראל. ברשותנו בסיס נתונים מהרשות הלאומית לבטיחות בדרכים. הנתונים נאספו בחמשת השנים האחרונות לא כולל שנת 2020. כמו כן, השתמשנו בנתונים מרשות האוכלוסין והאגירה בכדי לבדוק את כמות התושבים לכל ישוב. בנוסף, השתמשנו במדד האשכול הסוציו-אקונומי לכל ישוב שנלקח מהלשכה המרכזית לסטטיסטיקה.

כל הנתונים מצורפים בקישורים מטה:

[https://data.gov.il/dataset/accidents\\_municipal\\_Safety](https://data.gov.il/dataset/accidents_municipal_Safety) ([https://data.gov.il/dataset/accidents\\_municipal](https://data.gov.il/dataset/accidents_municipal))



[https://data.gov.il/dataset/residents\\_in\\_israel\\_by\\_communities\\_and\\_age\\_groups](https://data.gov.il/dataset/residents_in_israel_by_communities_and_age_groups) (<https://data.gov.il/dataset>



[/residents\\_in\\_israel\\_by\\_communities\\_and\\_age\\_groups](#))

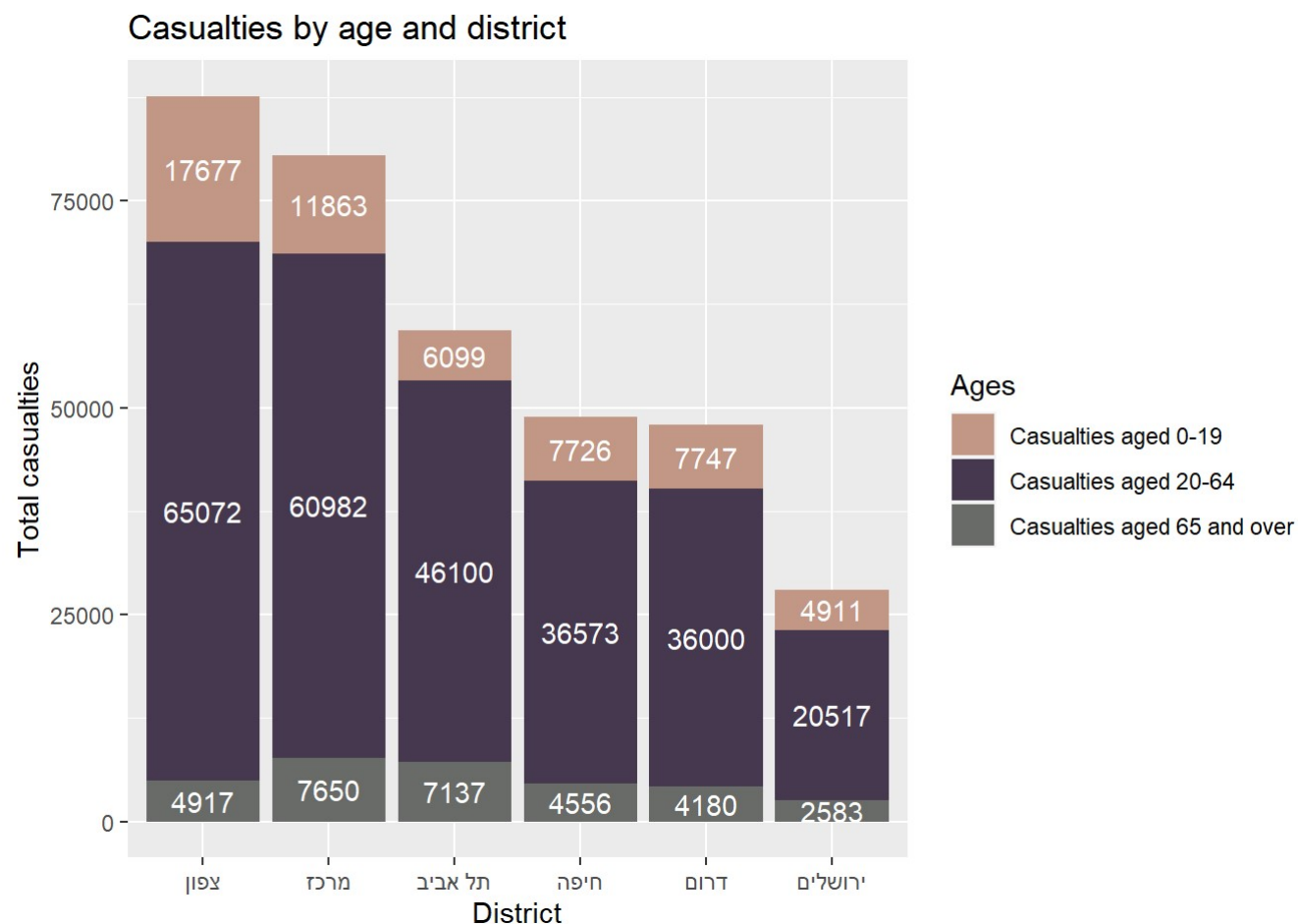
<https://www.cbs.gov.il/he/mediarelease/pages>) ... <https://www.cbs.gov.il/he/mediarelease/pages/2020/2017>



2020//אפיון-יחידות-גאוגרפיות-וסיווגן-לפי-הרמה-החברתית-כלכלית-של-האוכלוסייה-2017.aspx)

**שאלת מחקר: מה הוא ההבדל בין כמות הנפגעים בתאונות הדרכים באזורים שונים בישראל?**

בחלק זה של המחקר רצינו לבדוק את ההבדלים בין כמות הנפגעים בתאונות הדרכים במחוזות שונים במדינת ישראל. בגרף הבא ניתן לראות את כלל הנפגעים בתאונות הדרכים, לפי מחוזות ולפי שלושה טווחי גילאים: נפגעים בגילאי 0-19, נפגעים בגילאי 20-64, נפגעים גילאי 65 ומעלה.



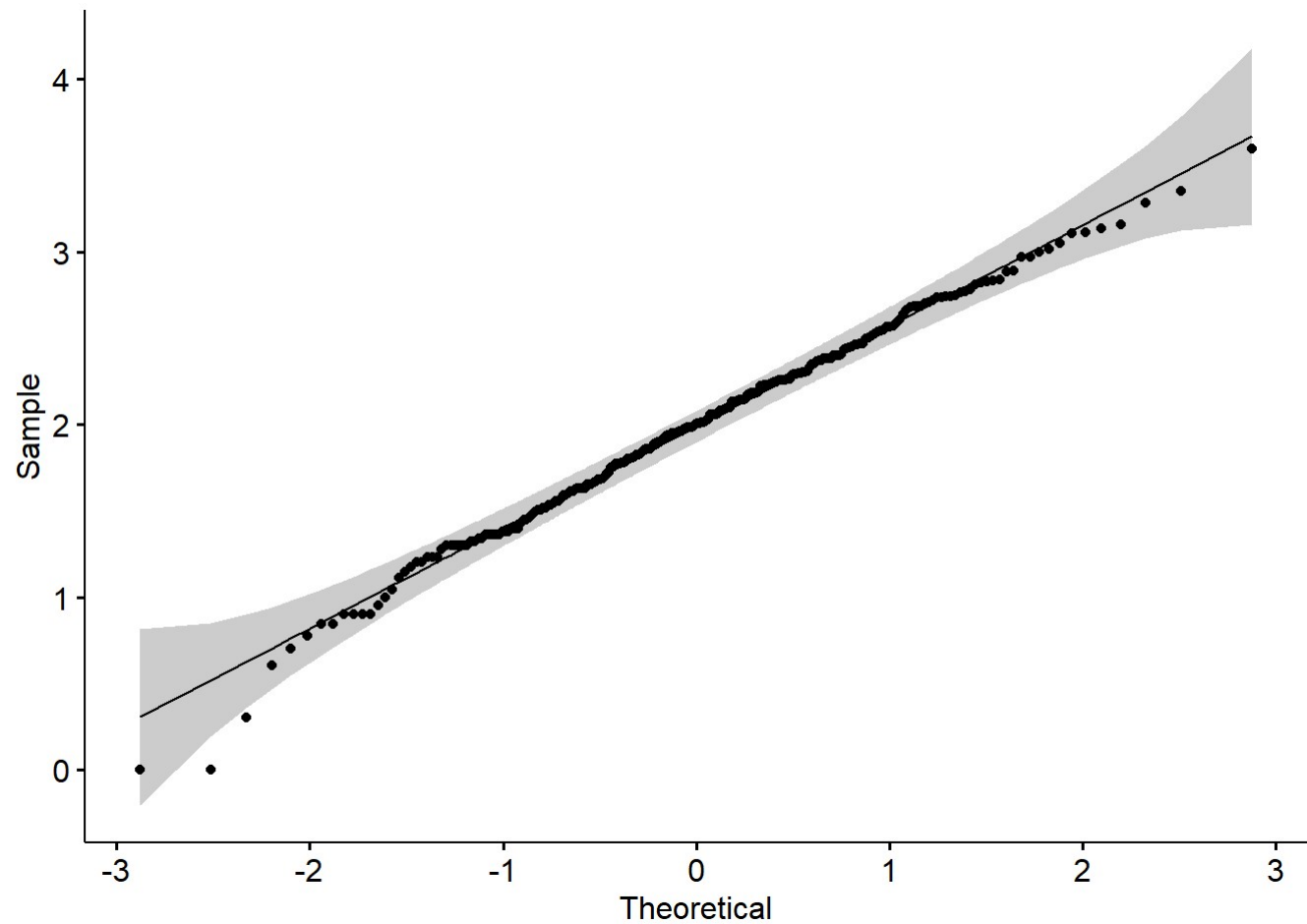
ניתן לראות שבמחוז צפון כמות הנפגעים בגילאי 0-19 ובגילאי 20-64 היא הגבוהה ביותר, כאשר כמות הנפגעים הגבוהה ביותר בגילאים 65 ומעלה היא במחוז מרכז. במחוז ירושלים כמות הנפגעים היא הקטנה ביותר.

בכדי לנתח את ההבדלים בין המחוזות בחרנו את המודל ניתוח שונות חד כיווני (ANOVA) וזאת כיוון שישנן 6 קבוצות (כמות המחוזות) בניתוח. נבצע את המבחן עבור כל טווח גילאים בנפרד. אחת ההנחות של מבחן ניתוח שונות חד כיווני היא שהנתונים מתפלגים נורמלית. בדקנו עבור כל קבוצת הנתונים את ההנחה ע"י פלט QQ-plot וע"י מבחן שפירו-Wilk לנורמליות. במבחן זה כאשר לא דוחים את השערת ה-0, ניתן לומר שהנתונים מתפלגים נורמלית.

תחילה נתבונן בפלט של המבחן שפירו-Wilk עבור קבוצת הגילאים 0-19:

```
##  
## Shapiro-Wilk normality test  
##  
## data: df_fixed_normality$Ing0_19  
## W = 0.99277, p-value = 0.264
```

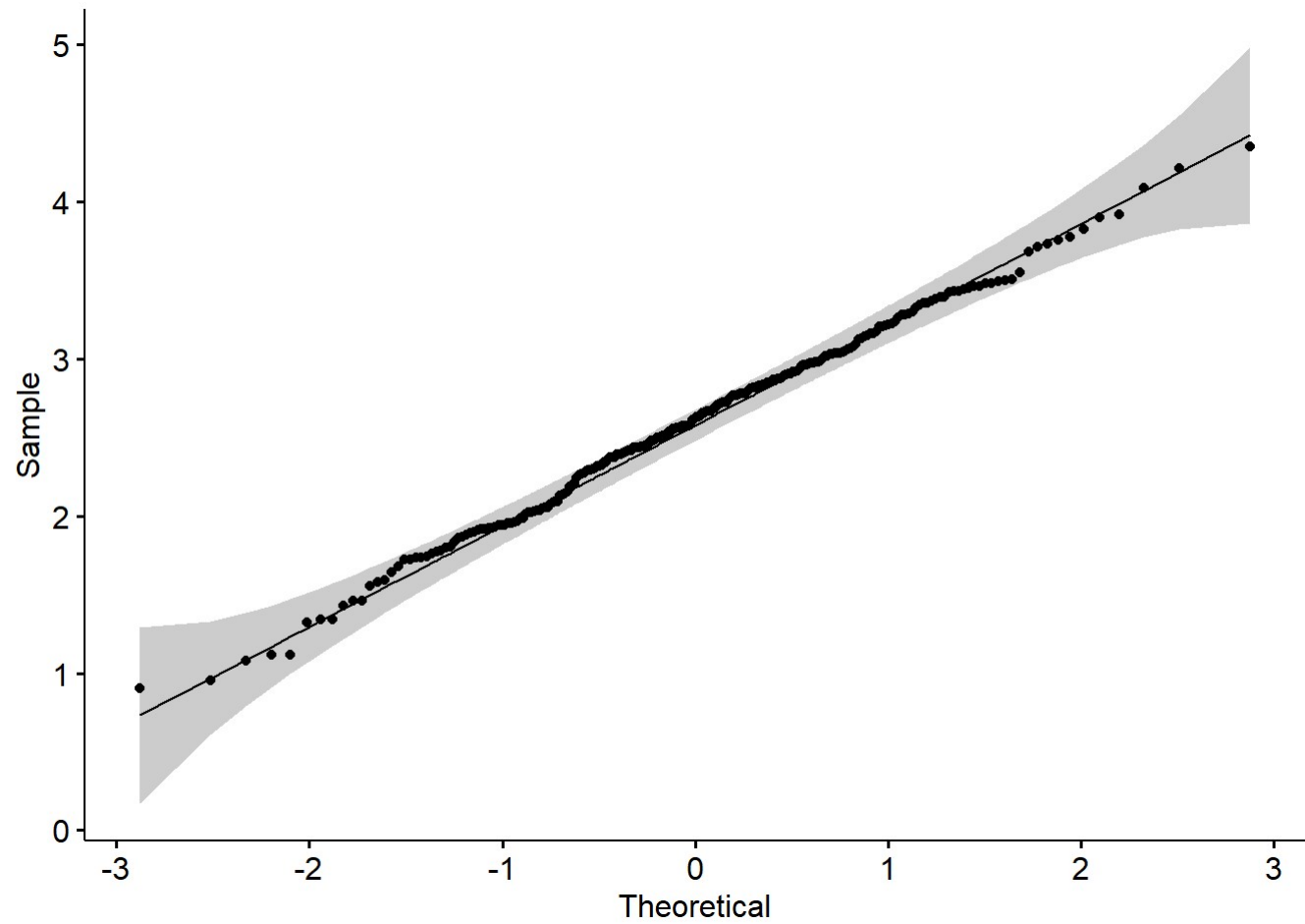
הערך p-value גדול מ- 5% ולכן לא נדחה את השערת ה-0 ונסיק שהנתונים במקרה זה מתפלגים נורמלית. כמו כן ניתן לראות בפלט QQ-plot שנתונים אכן מתפלגים נורמלית:



נמשיך ונבדוק עבור שני קבוצות הגילאים הנותרות את ההנחה. עבור קבוצת גילאים 20-64:

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  df_fixed_normality$Ing20_64  
## W = 0.99631, p-value = 0.8247
```

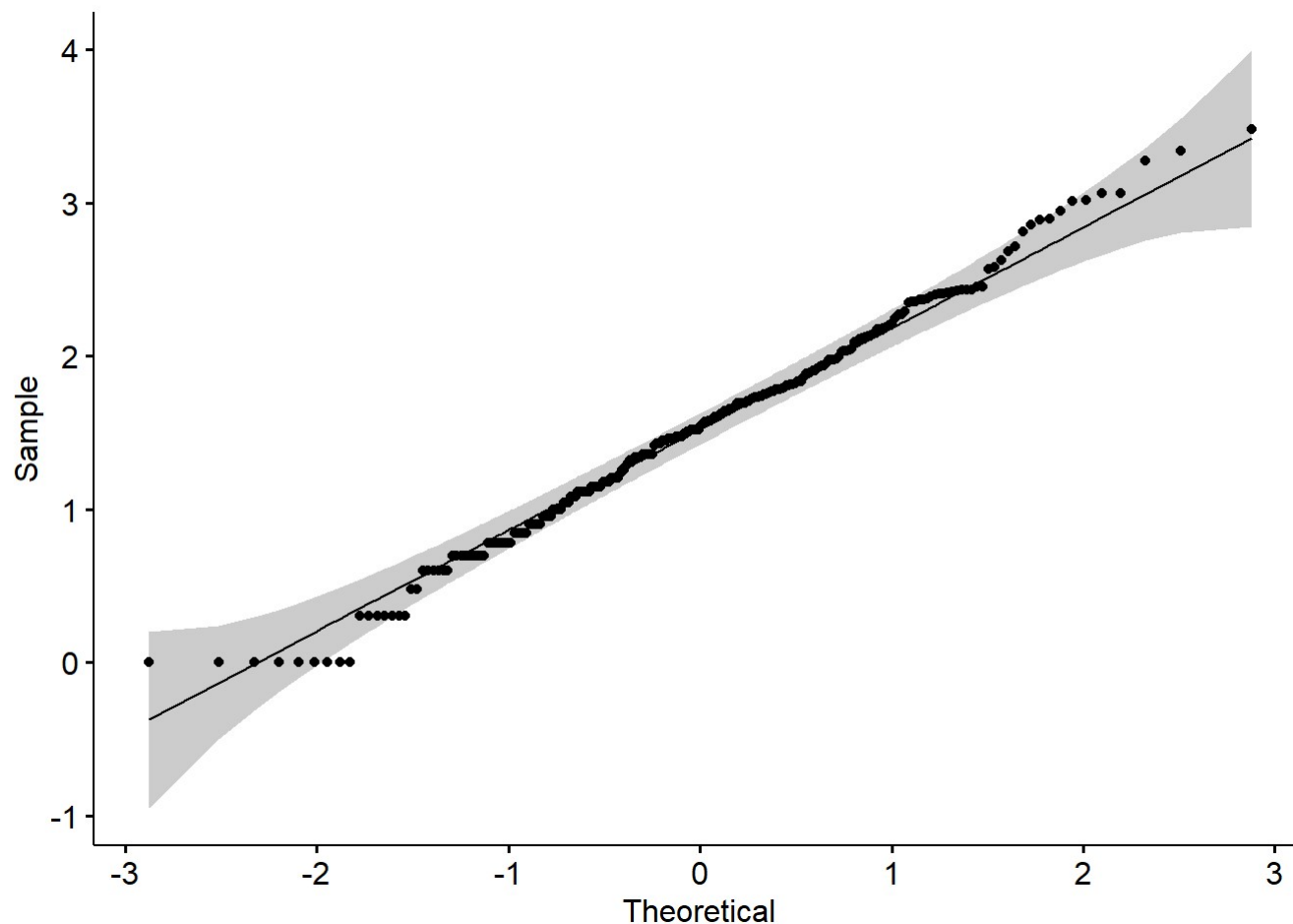
גם כאן הערך p-value גדול מ-5% ולכן לא נדחה את השערת ה-0 ונסיק שהנתונים במקרה זה מתפלגים נורמלית. מצורף פלט של QQ-plot גם עבור הקבוצה השנייה:



עבור קבוצת גילאים 65 ומעלה:

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  df_fixed_normality$Ing65  
## W = 0.99193, p-value = 0.1874
```

גם כאן הערך p-value גדול מ- 5% ולכן לא נדחה את השערת ה-0 ונסיק שהנתונים במקרה זה מתפלגים נורמלית. מצורף פלט של QQ-plot גם עבור הקבוצה השלישית:



נדגיש שעבור כל שלושת הקבוצות לפני בדיקת נורמליות ביצענו טרנספורמציה לוגריתמית לנתונים עקב זיהוי תבנית זו בנתונים שיש ברשותנו.

מאחר וכל הנתונים מתפלגים נורמלית ניתן לעבור למבחן חד כיווני – Anova. כאשר מבחן ההשערות מוגדר באופן הבא:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu$

$H_1: \text{else}$

כאשר השערת ה-0 אומרת שלא קיים הבדל בכמות תאונות הדרכים במחוזות וההשערה האלטרנטיבית אומרת שקיים הבדל בכמות תאונות הדרכים לפחות במחוז אחד.

כאמור לעיל ביצענו את המבחן עבור כל טווח הגילאים 0-19 התקבל פלט הבא:

```
##              Df    Sum Sq Mean Sq F value    Pr(>F)
## data.complete$DISTRICT    5  3014263   602853    4.471 0.000634 ***
## Residuals                257 34649831   134824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ניתן לראות שהערך p-value קטן מ-5% ולכן ניתן לדחות את השערת ה-0 ולהסיק שקיים הבדל בין כמות תאונות הדרכים בטווח גילאי 0-19 לפחות במחוז אחד (ברמת מובהקות 5%).

עבור טווח הגילאים 20-64 התקבל פלט הבא:

```
##              Df    Sum Sq Mean Sq F value    Pr(>F)
## dfforG2$DISTRICT    5 1.162e+08 23244017    5.363 0.000104 ***
## Residuals          257 1.114e+09  4333805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 40 observations deleted due to missingness
```

גם כאן ניתן לראות שהערך p-value קטן מ-5% ולכן ניתן לדחות את השערת ה-0 ולהסיק שקיים הבדל בין כמות תאונות הדרכים בטווח גילאי 20-64 לפחות במחוז אחד (ברמת מובהקות 5%).

עבור טווח הגילאים 65 ומעלה התקבל פלט הבא:

```
##              Df    Sum Sq Mean Sq F value    Pr(>F)
## dfforG2$DISTRICT    5  3071894   614379    7.39 1.71e-06 ***
## Residuals          257 21365016    83132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 40 observations deleted due to missingness
```

גם כאן ניתן לראות שהערך p-value קטן מ- 5% ולכן ניתן לדחות את השערת ה-0 ולהסיק שקיים הבדל בין כמות תאונות הדרכים בטווח גילאי 65 ומעלה לפחות במחוז אחד (ברמת מובהקות 5%).

כיוון שגילינו שבמחוז צפון כמות הנפגעים בגילאי 0-19 ובגילאי 20-64 היא הגבוהה ביותר החלטנו לבדוק מה היא הסיבה לכך. בקורסים קודמים למדנו שניתן להשתמש במודל רגרסיה לינארי פשוטה כדי לבדוק את טיב ההתאמה וכיוונה (השפעה שלילית או חיובית) בין משתנה מוסבר (כמות הנפגעים בתאונה) למשתנה מסביר (אשכול חברתי-כלכלי) ולכן נשתמש בה לביצוע הניתוח. מודל מוגדר באופן הבא:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

כאשר  $\beta_1$  הוא שיפוע עבור משתנה מסביר (אשכול חברתי כלכלי).

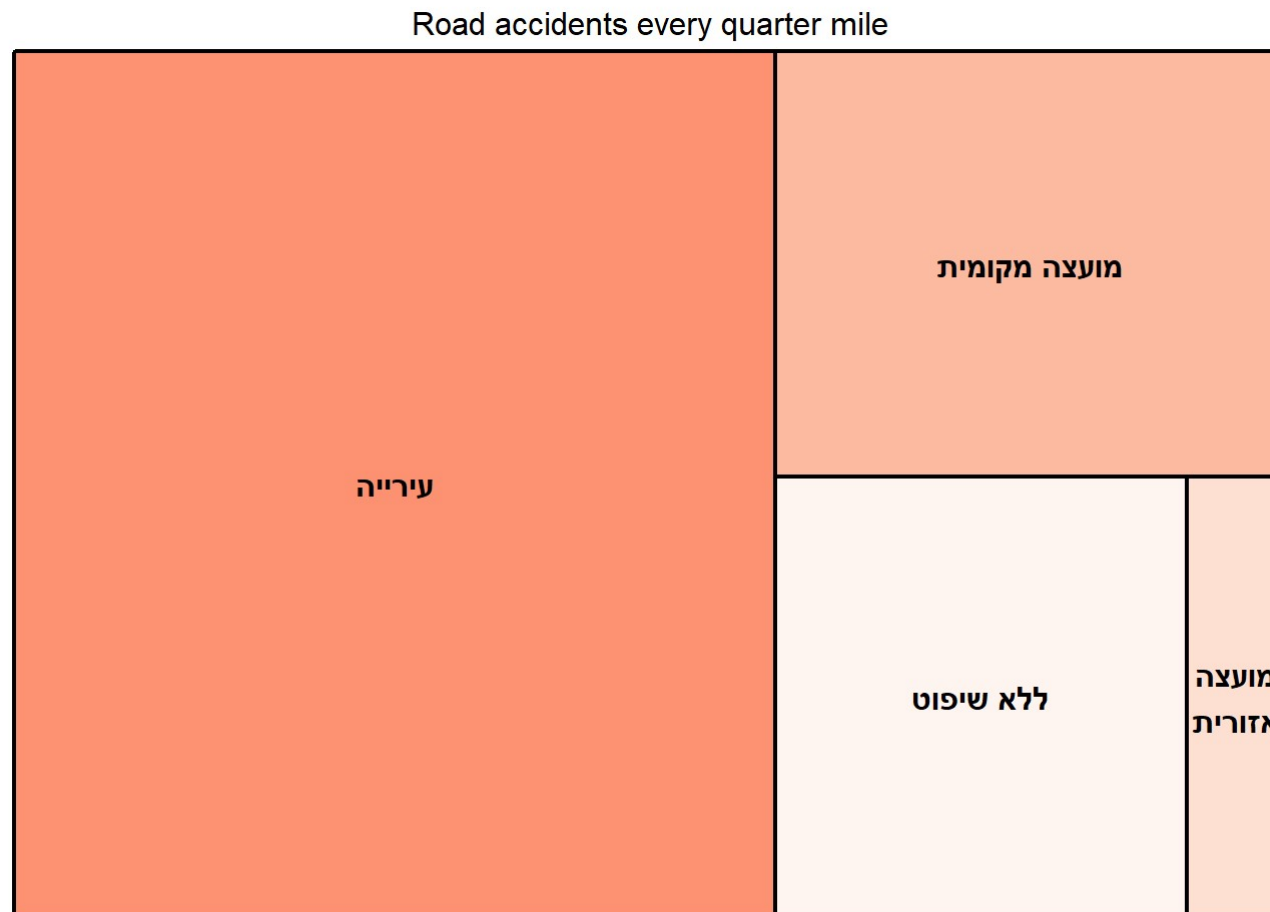
```
##
## Call:
## lm(formula = Casualties_20_64 ~ Socio_economic_cluster, data = df_linearmodel)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -960.2  -405.8  -236.2   133.3  2383.1
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         438.3      147.5   2.972  0.00405 **
## Socio_economic_cluster3    161.6      213.2    0.758  0.45105
## Socio_economic_cluster4    214.8      255.4    0.841  0.40318
## Socio_economic_cluster5    244.8      282.4    0.867  0.38901
## Socio_economic_cluster6    590.9      390.2    1.514  0.13441
## Socio_economic_cluster7    160.2      531.7    0.301  0.76414
## Socio_economic_cluster8   -336.1      390.2   -0.861  0.39198
## Socio_economic_cluster9   -343.3      737.4   -0.466  0.64293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 722.5 on 70 degrees of freedom
## Multiple R-squared:  0.06569,    Adjusted R-squared:  -0.02774
## F-statistic: 0.703 on 7 and 70 DF,  p-value: 0.6693
```

ניתן לראות שהערך p-value גדול מ- 5% עבור כל רמת אשכול חברתי-כלכלי ולכן ניתן להסיק כי אין קשר בין אשכול חברתי-כלכלי של ישובים במחוז צפון, לבין כמות תאונות

הדרכים (ברמת מובהקות 5%).

**שאלת מחקר:** מה הוא ההבדל בין כמות התאונות המתרחשות באזורים שונים בישראל?

בחלק זה של המחקר רצינו לבדוק את ההבדלים בין כמות הנפגעים בתאונות הדרכים במחוזות שונים במדינת ישראל את השאלה הזאת בחנו מכמה הביטים. נציג תחילה את כמות תאונות הדרכים שהתרחשו בממוצע לכל ק"מ רבוע בחמש שנים האחרונות, לפי סוג רשות מקומית:



ניתן לראות שבסוג רשות מקומית "עירייה" היו בממוצע הכי הרבה תאונות לכל ק"מ רבוע. בטבלה הבאה ניתן לראות את כמות התאונות לכל ק"מ רבוע עבור כל סוג רשות מקומית:



```
## Type of local authority Accidents per square kilometer
## 1 30.59038 ללא שיפוט
## 2 7.28366 מועצה אזורית
## 3 36.63887 מועצה מקומית
## 4 111.28618 עירייה
```

ניתן לראות שבסוג רשות מקומית "עירייה" היו בממוצע כ-111 תאונות לכל ק"מ רבוע. בנוסף ניתן שבסוג רשות "מועצה מקומית" היו יותר מ-36 תאונות לכל ק"מ רבוע.

ברצוננו לבדוק ע"י מבחן סטטיסטי שבסוג רשות "עירייה" היו יותר תאונות לכל ק"מ רבוע מאשר בסוג רשות "מועצה מקומית". לצורך בדיקה זו ביצענו מבחן וילקוקסון לבדיקת השערות על ההפרש בין התוחלות של שתי אוכלוסיות. נדגש שתחילה ביצענו מבחן t, אך במבחן זה קיימת הנחה שהנתונים מתפלגים נורמלית וכאשר בדקנו את ההנחה, גילינו שהנתונים לא מתפלגים נורמלית, לכן עברנו למבחן וילקוקסון כי במבחן זה הנחה זו מתבטלת. מבחן ההשערות מוגדר באופן הבא:

$H_0: \mu_1 - \mu_2 = 0$

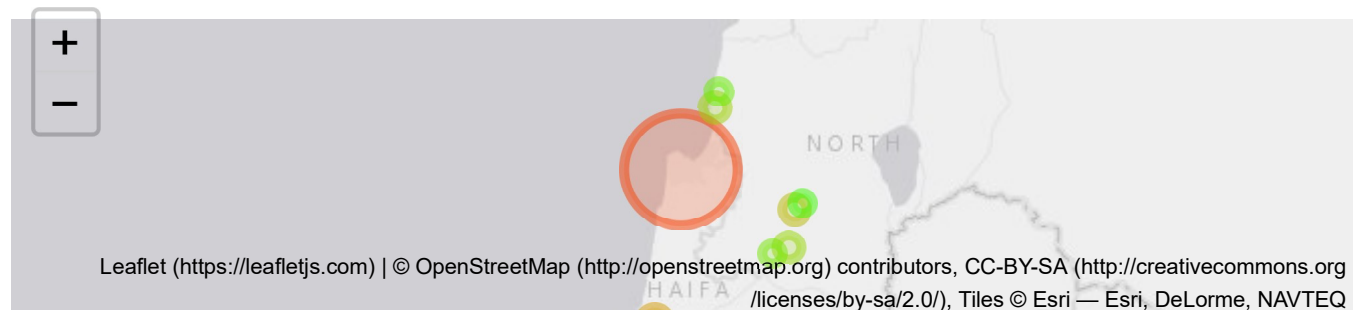
$H_1: \mu_1 - \mu_2 > 0$

השערת ה-0 אומרת שאין הבדל בין ההפרש של שתי אוכלוסיות כאשר ההשערה האלטרנטיבית אומרת שקיים הבדל בהפרש תוחלות של שתי הקבוצות.

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: acc_per_km_1$Accidents_per_square_kilometer and acc_per_km_2$Accidents_per_square_kilometer
## W = 6750, p-value = 1.142e-14
## alternative hypothesis: true location shift is greater than 0
```

ניתן לראות שהערך p-value קטן מ-5% ולכן ניתן לדחות את השערת ה-0 ולהסיק כי קיים הפרש בין שני תוחלות, כלומר שבסוג רשות "עירייה" היו יותר תאונות לכל ק"מ רבוע מאשר בסוג רשות "מועצה מקומית" (ברמת מובהקות 5%).

**שאלת מחקר: מה הוא ההבדל בין כמות הנפגעים בתאונות הדרכים באזורים שונים בישראל?**





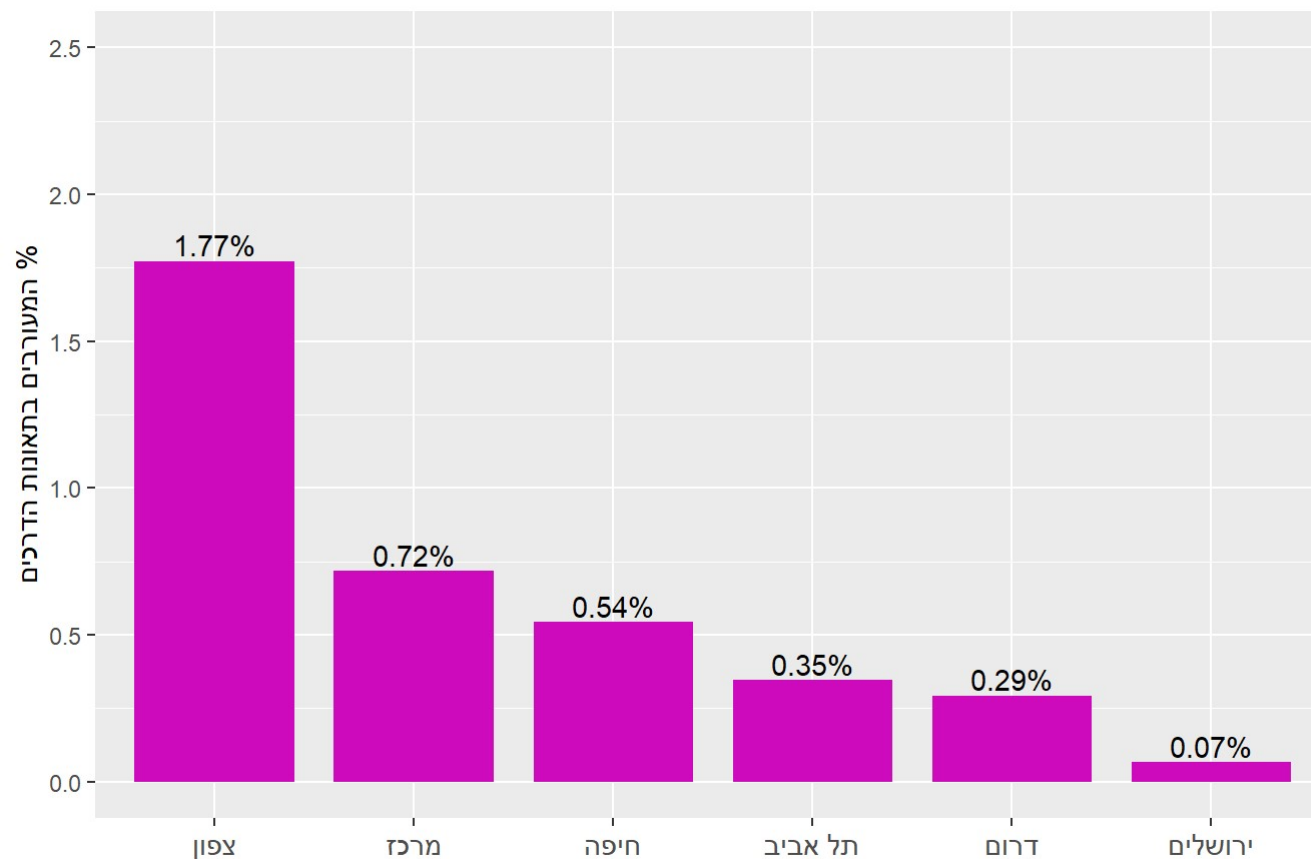
המפה האינטראקטיבית הזאת מראה לנו את 30 היישובים שבהם התרחשו הכי הרבה תאונות דרכים בשנים האחרונות לפי הנתונים שלנו. ניתן ללחוץ על הסימונים של כל יישוב ויופיע שם היישוב ומספר התאונות שהתרחשו בו.

צבוע וגודלו של כל סימון מייצג את מספר התאונות שהתרחשו ביישוב, כך שככל שהעיגול גדול יותר ואדום יותר כך מספר הטעונות היה גדול יותר.

ניתן להבחין שמספר רב של היישובים הללו נמצאים במחוז המרכזי של הארץ.

**שאלת מחקר: מה הוא ההבדל בין כמות התאונות המתרחשות באזורים שונים בישראל?**

### שיעור המעורבים בתאונות דרכים ביחס לאוכלוסיה לפי מחוז



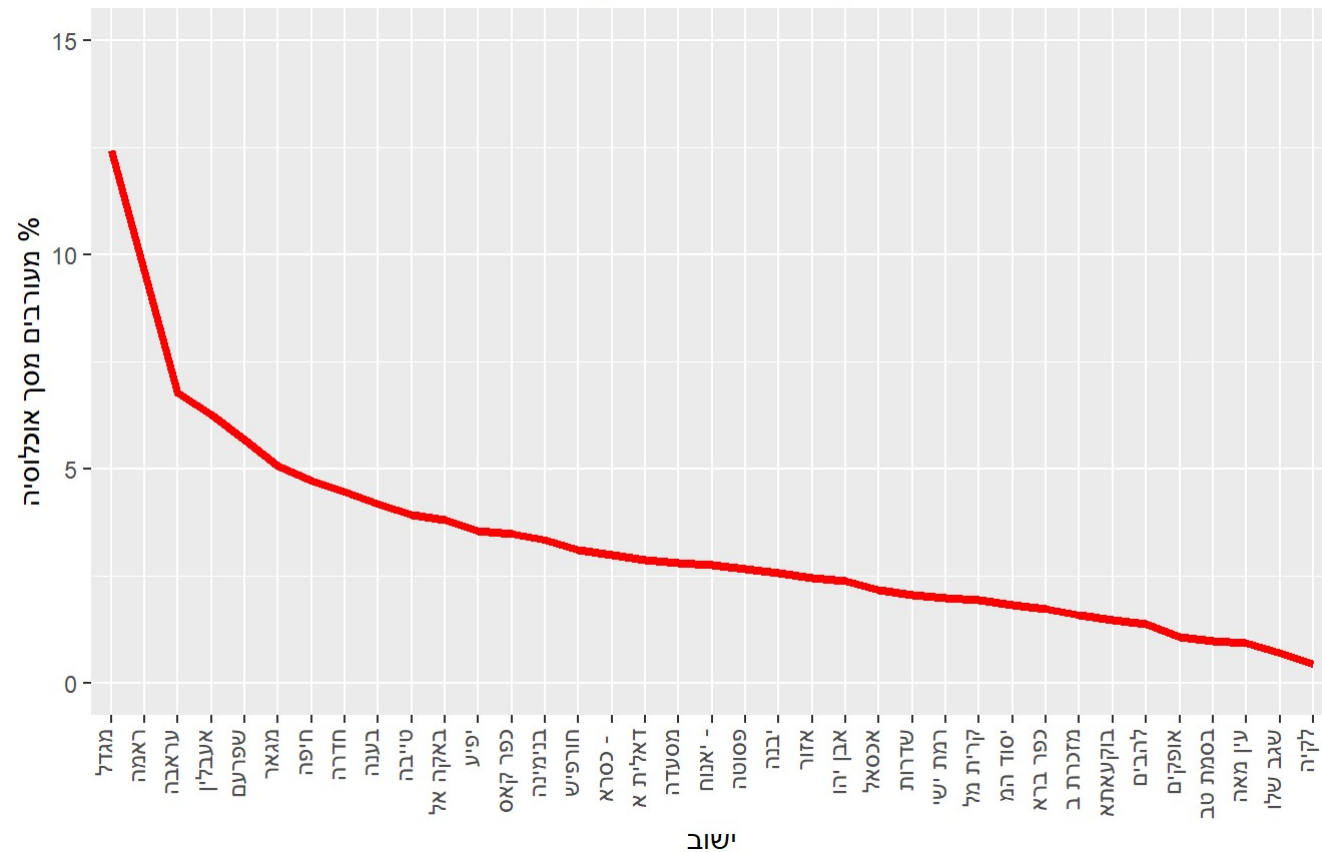
בעזרת הגרף הזה ניתן לקבל תמונה התחלתית על הנתונים.

הגרף מציג את השיעור הממוצע של אנשים המעורבים בתאונה בחמש השנים האחרונות מנורמל לפי אוכלוסיה באותו מחוז.

ניתן לראות בגרף שבממוצע שנתי על פי החמש שנים האחרונות כמעט 1 מתוך 50 אנשים במחוז צפון היה מעורב בתאונה.

בנוסף, ניתן לראות שבמחוז צפון שיעור המעורבים בתאונה ביחס לאוכלוסיה הוא הרבה יותר גובה משאר המחוזות, מעל פי 2 מהמחוז הבא בגרף.

### אחוז מוערבים בתאונה לפי ישוב (מדגם מייצג n=39)



בגרף זה אנחנו נכנסים עמוק יותר לנתונים כדי לבדוק אם נוכל לזהות סיבות או דפוסים משותפים שיסבירו למה לישוב מסויים יש יותר או פחות מעורבים בתאונה ביחס לאוכלוסיה באותו הישוב.

ניתן לראות בגרף כי חמשת הישובים עם השיעור הכי נמוך הם ישובים עם דירוג אשכול חברתי כלכלי נמוך.

אם זאת ישנם ישובים מדירוגים חברתיים כלכליים שונים מפוזרים במקומות שונים בגרף ועל כן נרצה לבדוק את ההתאמה בין דירוג אשכול חברתי-כלכלי ושיעור המוערבים בתאונה ביחס לאוכלוסיה בעזרת מבחן סטטיסטי.

בקורסים קודמים למדנו שניתן להשתמש במודל רגרסיה לינארי פשוטה כדי לבדוק את טיב ההתאמה וכיוונה (השפעה שלילית או חיובית) בין משתנה מוסבר (שיעור המוערבים בתאונה ביחס לאוכלוסיה) למשתנה מסביר (אשכול חברתי-כלכלי) ולכן נשתמש בה לביצוע הניתוח:

$$H_0: \beta_1 = 0$$

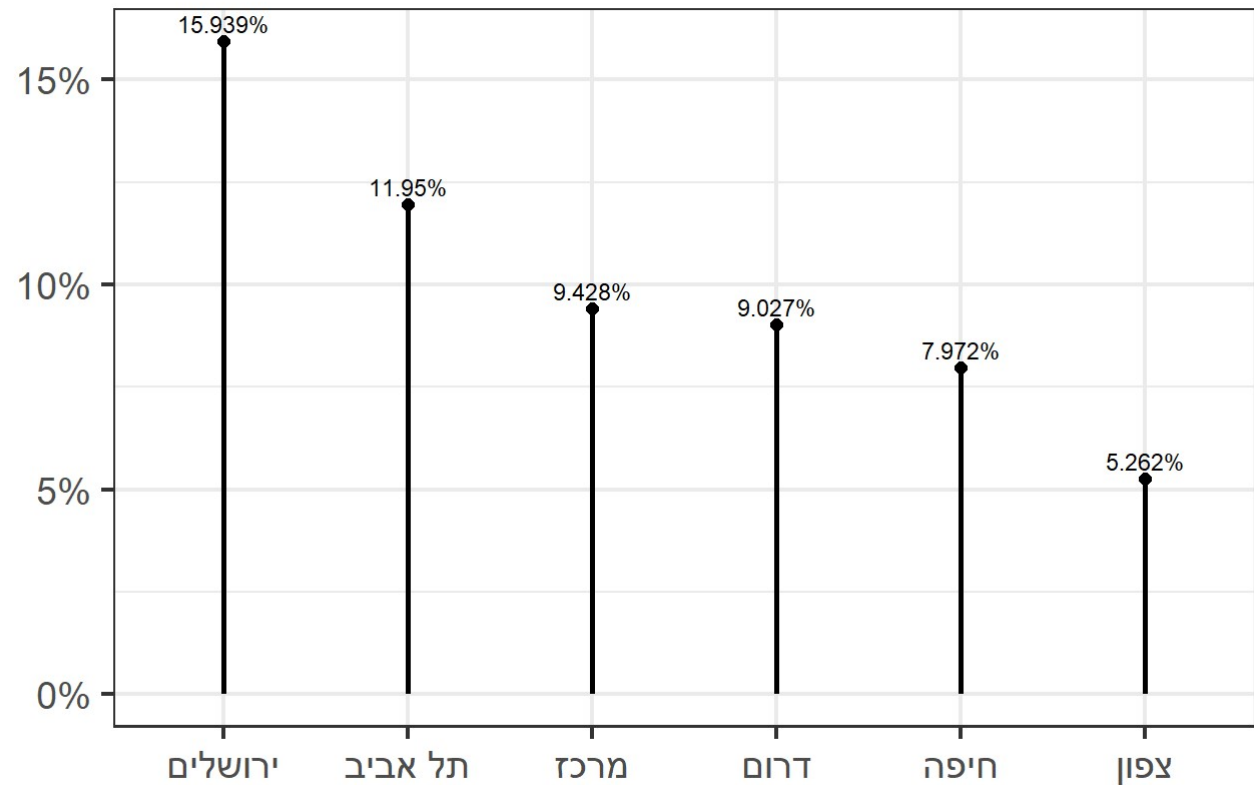
## H1: $\beta_1 \neq 0$

```
##  
## Call:  
## lm(formula = y ~ x)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -2.9523 -1.3702 -0.4729  0.7375  9.2315   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  3.232220   0.359430   8.993 3.11e-16 ***  
## x           -0.008333   0.066128  -0.126    0.9      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 2.205 on 182 degrees of freedom  
## Multiple R-squared:  8.724e-05, Adjusted R-squared:  -0.005407  
## F-statistic: 0.01588 on 1 and 182 DF,  p-value: 0.8999
```

ניתן לראות כי ה-Pvalue גדול מ-5% לכן לא נדחה  $H_0$  ברמת מובהקות 5% ונסיק כי אין קשר בין שיעור המעורבים בתאונה ביחס לאוכלוסיה בישוב מסויים לדירוג החברתי-כלכלי של אותו ישוב.

**שאלת מחקר: האם קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין המחוזות?**

## שיעור הולכי הרגל שנפגעו בתאונות דרכים לפי מחוז



בחנו את השאלה האם קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין המחוזות?  
על-פי הגרף ניתן לראות הבדלים בין המחוזות וההבדל המהותי ביותר הינו בין מחוז צפון לבין ירושלים.  
לכן, החלטנו לנתח הבדלים אלו באמצעות ניתוח שונות חד-כיווני.

### שלב 1 -

בדיקת הנחת התפלגות נורמלית של הנתונים על-ידי מבחן Shapiro-Wilk.

$$H_0: Y \sim N(\mu, \sigma^2)$$

**H1: else**

```
##  
## Shapiro-Wilk normality test  
##  
## data: My_Data_By_Districts$Percentage  
## W = 0.17379, p-value < 2.2e-16
```

כפי שניתן לראות קיבלנו שה - P-Value שואף ל - 0 ולכן דחינו את השערת האפס. כלומר, הנתונים אינם מתפלגים נורמלית.

## שלב 2 -

בדיקת הנחת שוויון שוניות על-ידי מבחן Levene.

השתמשנו במבחן זה מכיוון שאינו דורש שהנחת הנורמליות תתקיים על-מנת שנוכל לבצעו (ההנחה שדחינו בשלב 1).

**H0:  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma^2$**

**H1: else**

```
## Levene's Test for Homogeneity of Variance (center = median)  
##           Df F value    Pr(>F)  
## group      5  10.025 7.053e-09 ***  
##           296  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

כפי שניתן לראות קיבלנו שה - P-Value שואף ל - 0 ולכן דחינו את השערת האפס. כלומר, הנחת שוויון שוניות אינה מתקיימת.

## שלב 3 -

ביצענו ניתוח שונות חד-כיווני באמצעות מבחן Kruskal-Wallis שהינו מבחן א-פרמטרי.

בחרנו במבחן זה מכיוון שהוא מאפשר ביצוע של ניתוח שונות חד-כיווני ללא קיומם של ההנחות שמבחן Anova דורש (ההנחות שדחינו בשלבים 1 ו - 2).

**H0:  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu$**

**H1: else**

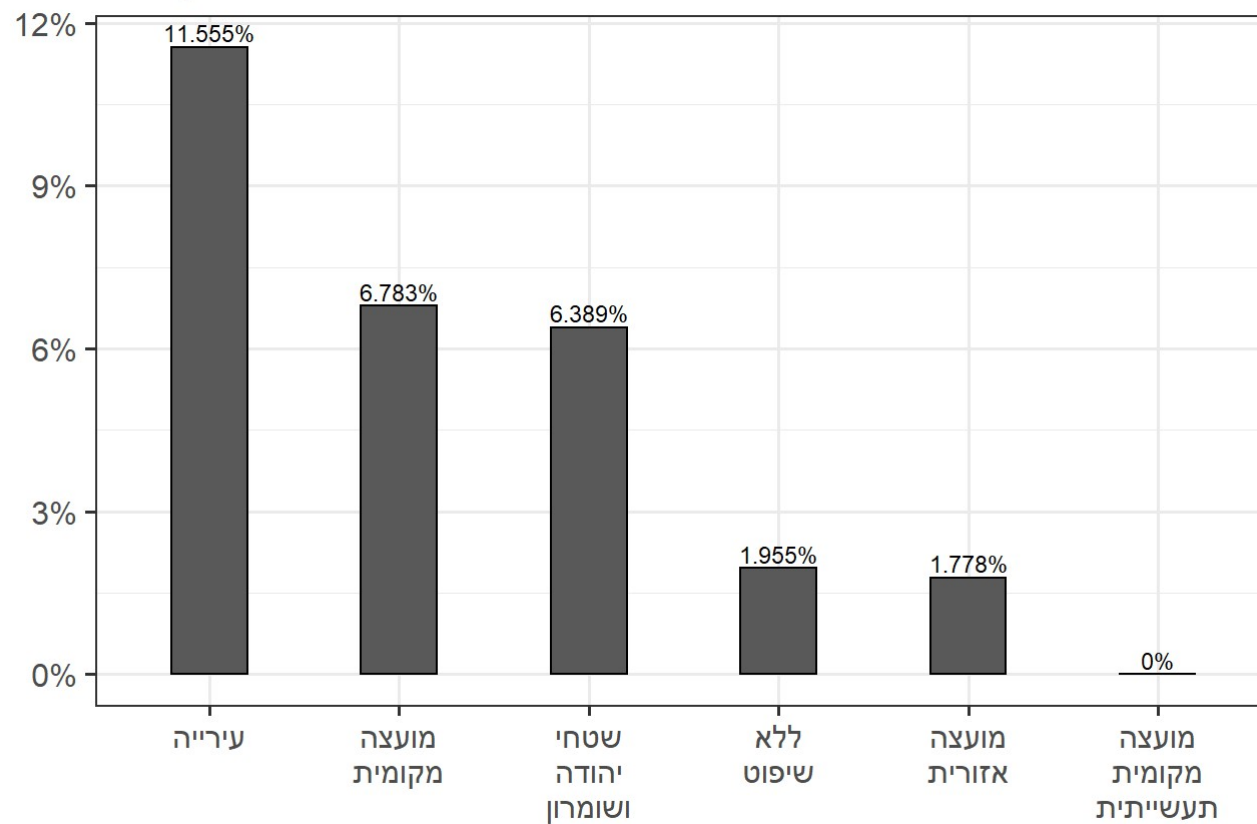
```
##  
## Kruskal-Wallis rank sum test  
##  
## data: Percentage by District  
## Kruskal-Wallis chi-squared = 33.908, df = 5, p-value = 2.483e-06
```

כפי שניתן לראות קיבלנו שה - P-Value משמעותית קטן יותר מרמת המובהקות ( $\alpha = 5\%$ ) ולכן דחינו את השערת האפס. כלומר, הגענו למסקנה כי קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין המחוזות.

**שאלת מחקר: האם קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין רשויות מקומיות?**



## שיעור הולכי הרגל שנפגעו בתאונות דרכים לפי סוג רשות מקומית



כמו כן, בחנו את השאלה האם קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין רשויות מקומיות?

על-פי הגרף ניכר כי קיימים הבדלים, כמו למשל: בין עירייה לבין מועצה מקומית תעשייתית.

לכן, גם כאן החלטנו לנתח הבדלים אלו באמצעות ניתוח שונות חד-כיווני.

### שלב 1 -

בדיקת הנחת התפלגות נורמלית של הנתונים על-ידי מבחן Shapiro–Wilk.

$$H_0: Y \sim N(\mu, \sigma^2)$$

**H1: else**

```
##  
## Shapiro-Wilk normality test  
##  
## data: My_Data_By_Type_Of_Local_Authority$Percentage  
## W = 0.16234, p-value < 2.2e-16
```

כפי שניתן לראות קיבלנו שה - P-Value שואף ל - 0 ולכן דחינו את השערת האפס. כלומר, הנתונים אינם מתפלגים נורמלית.

## שלב 2 -

בדיקת הנחת שוויון שוניות על-ידי מבחן Levene.

השתמשנו במבחן זה מכיוון שאינו דורש שהנחת הנורמליות תתקיים על-מנת שנוכל לבצעו (ההנחה שדחינו בשלב 1).

**H0:  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma^2$**

**H1: else**

```
## Levene's Test for Homogeneity of Variance (center = median)  
##      Df F value    Pr(>F)  
## group  5  4.1286 0.001219 **  
##      297  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

כפי שניתן לראות קיבלנו שה - P-Value קטן יותר מרמת המובהקות ( $\alpha = 5\%$ ) ולכן דחינו את השערת האפס. כלומר, הנחת שוויון שוניות אינה מתקיימת.

## שלב 3 -

ביצענו ניתוח שונות חד-כיווני באמצעות מבחן Kruskal-Wallis שהינו מבחן א-פרמטרי.

בחרנו במבחן זה מכיוון שהוא מאפשר ביצוע של ניתוח שונות חד-כיווני ללא קיומם של ההנחות שמבחן Anova דורש (ההנחות שדחינו בשלבים 1 ו-2).

**H0:  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu$**

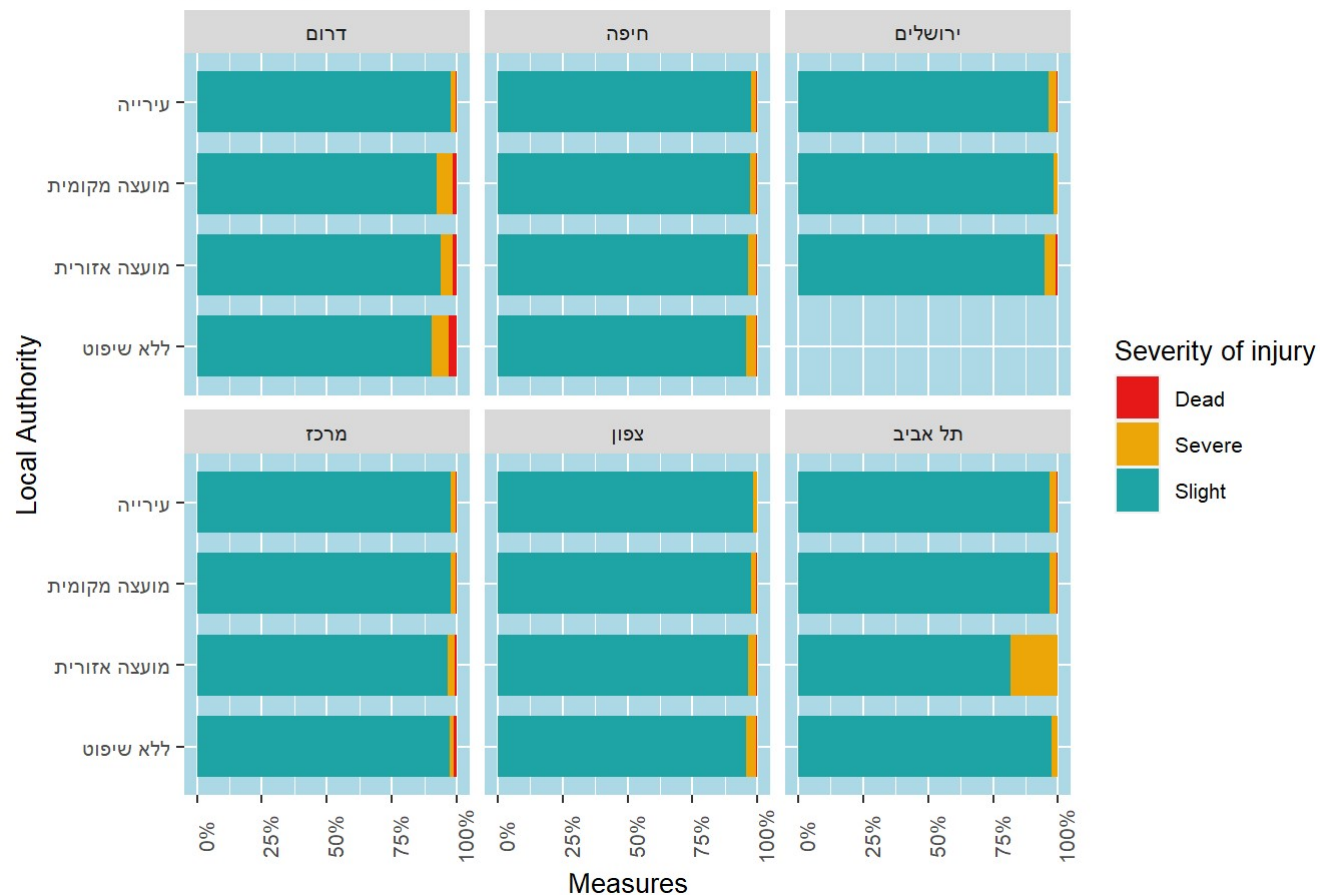
**H1: else**

```
##  
## Kruskal-Wallis rank sum test  
##  
## data: Percentage by Type_Of_Local_Authority  
## Kruskal-Wallis chi-squared = 126.57, df = 5, p-value < 2.2e-16
```

כפי שניתן לראות קיבלנו שה - P-Value שואף ל - 0 ולכן דחינו את השערת האפס. כלומר, הגענו למסקנה כי קיימים הבדלים בשיעור הולכי הרגל שנפגעו בתאונות דרכים בין רשויות מקומיות.

**שאלת מחקר: האם לסוג הרשות המקומית של יישוב או המחוז עליו היישוב שייך יש השפעה על רמות הפציעה של האנשים הנפגעים בתאונות הדרכים באותו יישוב?**

Segmentation of severity of injury



הגרף הזה מציג לנו את הפילוח של רמות הפציעה השונות מתאונות דרכים ביישובים לפי רשות מקומית ומחוז.

ניתן לראות מהגרף באופן כללי שלאזורים ללא שיפוט יש אחוז גבוהה יותר של פצועים קשה ואף מתים מיישובים בעלי סוגי רשויות אחרות, כאשר בעיריות יש את האחוזים הכי נמוכים של מתים ופצועים קשה. בנוסף נראה שמכל המחוזות, במחוז דרום יש את אחוז המתים הכי גדול.

כדי לבדוק את השארתנו מהגרף נשתמש במבחן הסטטיסטי הבא:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

```
##
## Welch Two Sample t-test
##
## data: noAuthorDead and muniDead
## t = 1.4461, df = 5.0261, p-value = 0.1037
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.002849402      Inf
## sample estimates:
## mean of x mean of y
## 0.009939455 0.002668088
```

השתמשנו במבחן t-test דו מדגמי כדי לבדוק האם, מבחינה סטטיסטית, אחוז המתים מתאונות דרכים באזורים ללא שיפוט הוא גבוהה מאחוז המתים מתאונות דרכים ביישובים עם עירייה.

לפי ה-p-value של המבחן (0.1037) לא דחינו את השערת האפס ברמת מובהקות של 5% והסקנו שלמרות מה ששיערנו מהגרף, אין הבדל משמעותי בין אחוז המתים באזורים ללא רשות לאחוז המתים ביישובים עם עירייה.

**הנושא בו בחרנו לחלק הניתוח הכלכלי הנו בדיקת כדאיות פרויקטים**

**לצורך הניתוח בחרנו בפרויקטים הבאים:**

**פרויקט 1:**

**השקעה התחלתית:**

60

**תזרים המזומנים לפי שנים:**

1. 10. 2. 20. 3. 20. 4. 20. 5. 40

**מחיר ההון של החברה:**

10%

פרויקט 2:

השקעה התחלתית:

100

תזרים המזומנים לפי שנים:

1. 12 2. 40 3. 30 4. 15 5. 28

מחיר ההון של החברה:

6%

**בדיקת כדאיות הפרויקטים:**

בודקת למעשה את הכדאיות הכלכלית של השקעת הפירמה בפרויקט מסוים, בהתחשב באורך חיי הפרויקט, בתזרים המזומנים של הפרויקט ובמחיר ההון של הפירמה.

קבלת ההחלטות בנוגע לביצוע או אי ביצוע הפרויקט תלויה בערך הנוכחי שלו, אותו מחשבים באמצעות היוון תזרים המזומנים לנקודת זמן של היום.

עבור ערך נוכחי נקי הגדול מאפס, ההשקעה בפרויקט כדאית.

עבור ערך נוכחי נקי השווה לאפס, קיימת אדישות בנוגע לכדאיות ההשקעה בפרויקט.

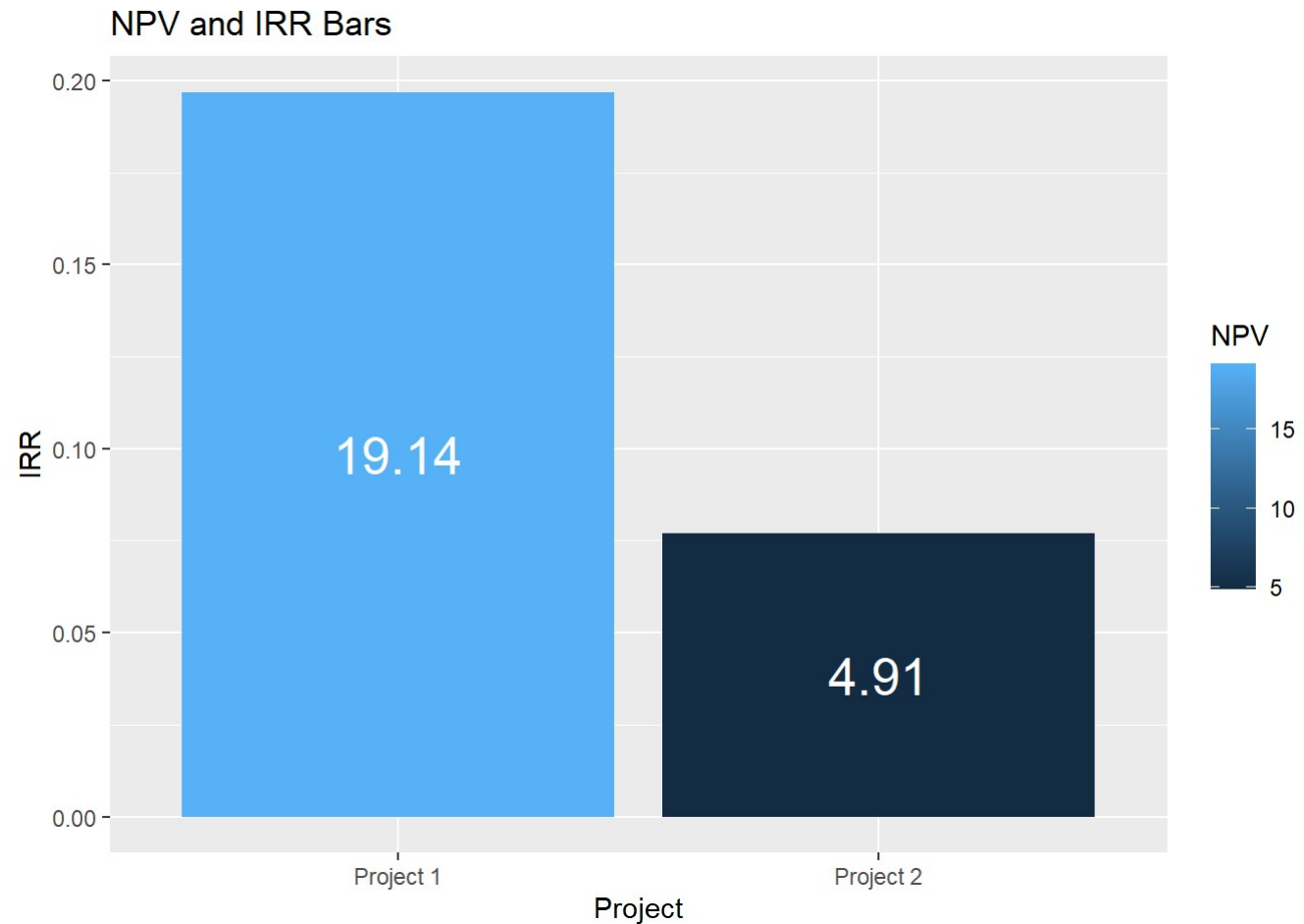
עבור ערך נוכחי נקי הקטן מאפס, ההשקעה בפרויקט אינה כדאית.

IRR – שיעור ההיוון שמאפס את הערך הנוכחי הנקי של הפרויקט.

עבור מחיר הון הגדול מ-IRR, ההשקעה בפרויקט אינה כדאית.

עבור מחיר הון השווה ל-IRR, קיימת אדישות בנוגע לכדאיות ההשקעה בפרויקט.

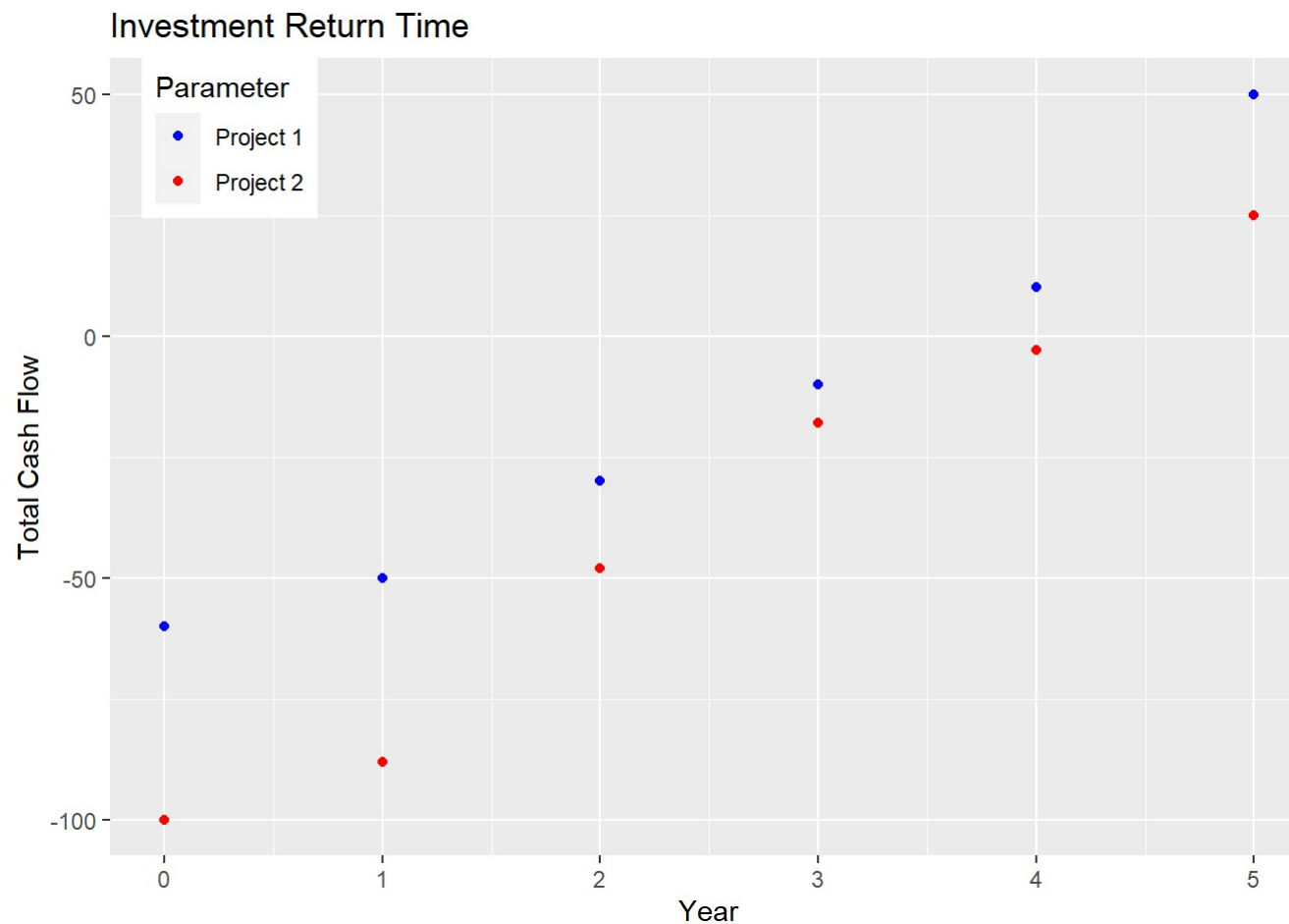
עבור מחיר הון הקטן מ-IRR, ההשקעה בפרויקט כדאית.



בגרף זה ניתן להבחין כי שני הפרויקטים כדאיים לפירמות המתאימות, אך פרויקט 1 כדאי במידה רבה יותר עבור הפירמה המבצעת אותו מאשר פרויקט 2 עבור הפירמה המתאימה לו, הן מבחינת הענ"נ והן מבחינת השת"פ

#### זמן החזר השקעה:

הפרויקטים בהם בחרנו לצורך ביצוע הניתוח הנם פרויקטים השקעתיים, משמע תזרים המזומנים נובע מהשקעה ראשונית של הפירמה בפרויקט (למשל: הקמת תשתיות, תשלום לצוותי פיתוח וכד') חישוב זמן החזר ההשקעה הנו כלי חשוב בבדיקת כדאיות פרויקטים, שכן החל מזמן זה ואילך למעשה הפירמה תתחיל להניב רווחים מהפרויקט.



גרף זה מציג את סה"כ תזרים המזומנים של הפירמה כתוצאה מהשקעה בפרויקט המתאים במהלך כל שנה, בנוסף, בגרף זה ניתן לראות מתי תזרים המזומנים של הפרויקט חוצה את רף ה-0. הנקודה הראשונה שחוצה רף זה מצביעה על השנה בה הפירמה החזירה את השקעתה בפרויקט ולמעשה החל ממנה ואילך כל תזרים מזומנים נוסף יהווה רווח עבודה.

#### לסיכום:

**פרויקט 1:** פרויקט כדאי - ענ"נ: 19.14325, שת"פ: 19.69%, זמן החזר השקעה: 4 שנים

**פרויקט 2:** פרויקט כדאי - ענ"נ: 4.913825, שת"פ: 7.74%, זמן החזר השקעה: 5 שנים