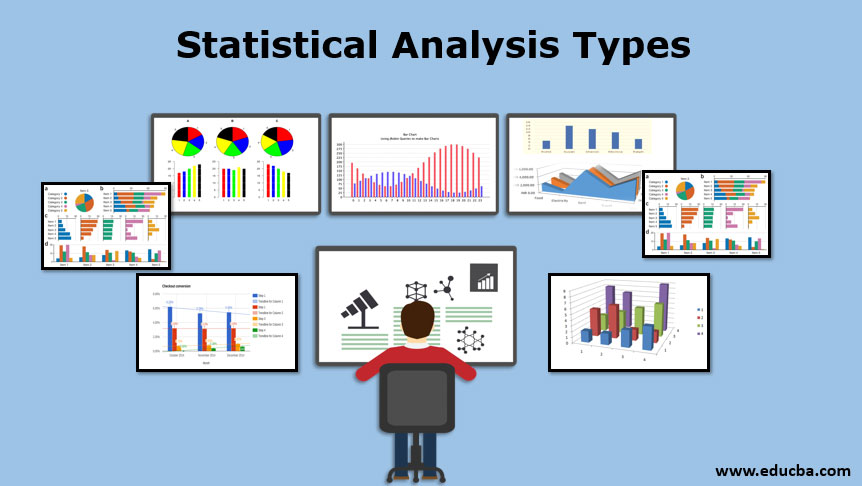
פרויקט במודלים של רגרסיה לינארית חלק א



מגישים :   
316320779  
316164177

קבוצה 41

**תוכן עניינים:**

1. [הData base שלנו הוא House Price. 3](#_Toc122215678)
2. [יצירת טבלת משתנים: 3](#_Toc122215679)
3. [תיאור המשתנים: 3](#_Toc122215680)
4. [תיאור קשרים העיקריים בין המשתנים המסבירים ובין המשתנים המסבירים למוסבר: 4](#_Toc122215681)
5. [ניתוח תיאורי של המשתנים הרציפים: 7](#_Toc122215682)
6. [ניתוח חריגים: 9](#_Toc122215683)
7. [פונקציית צפיפות והתפלגות מצטברת עבור שלושה מסבירים: 12](#_Toc122215684)
8. [ייצוג קשרים בעזרת תרשימים: 14](#_Toc122215685)
9. [טבלאות שכיחות : 17](#_Toc122215686)

# חלק א': הData base שלנו הוא House Price.

# יצירת טבלת משתנים:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| שם המשתנה | סוג המשתנה מוסבר/מסביר | סימון במודל | יחידת מידה | סוג המשתנה – רציף / קטגוריאלי | הסבר קצר על המשתנה |
| rooms | מסביר |  | מספרים | רציף | מספר החדרים בבית |
| sqft\_house | מסביר |  | מספרים | רציף | גודל הבית בSquare Feet |
| sqft\_lot | מסביר |  | מספרים | רציף | גודל המגרש בSquare Feet |
| sqft\_basement | מסביר |  | מספרים | רציף | גודל המרתף בSquare Feet |
| Floors | מסביר |  | מספרים | רציף | מספר הקומות בנכס |
| Condition | מסביר |  | 1-לא טוב 2- כמעט טוב 3- טוב 4- כמעט טוב מאוד 5-טוב מאוד | קטגוריאלי | מצב הנכס |
| yr\_built | מסביר |  | מספרים | רציף | השנה בה הנכס נבנה |
| Renovated | מסביר |  | 1-כן 0-לא | קטגוריאלי | האם הנכס שופץ (1-כן, 0-לא) |

# תיאור המשתנים:

1. מספר החדרים בבית ((:

אנו מאמינים כי ישנו קשר בין מספר החדרים בנכס לבין מחירו. מספר החדרים הוא אחד הפרמטרים הנבדקים ביותר בבואו של רוכש פוטנציאלי. ככל שמספר החדרים גדול יותר סביר כי תהיינה השפעה על מחיר הנכס וכן נכס זה עשוי להיות אטרקטיבי למגוון רב יותר של רוכשים.

1. גודל הבית ((:  
   אנו חושבים כי גודל הבית נמנה מבין הפרמטרים הבולטים שעשויים להשפיע על מחיר הנכס. ככל שהבית גדול יותר סביר כי מחירו יושפע בהתאם. ככל שהבית גדול יותר ניתן למשל לאכלס בו יותר דיירים ולגבות מחיר בהתאם.
2. גודל המגרש ((:

אנו סבורים כי גודל המגרש הינו אחד מהפרמטרים הבולטים שעשויים להשפיע על מחיר הנכס. ככל שהנכס נפרש על שטח גדול יותר סביר כי מחירו יושפע בהתאם. במגרש גדול יש פוטנציאל גדול יותר לבצע התאמות לרצונותיו של הרוכש, כגון בניית בריכה, חניה וכו'.

1. גודל המרתף ((:

ייתכן וגודל המרתף משפיע על מחיר הנכס שכן במידה ורוכשים נכס עם מרתף, הרי שהגודל שלו מעיד על האפשרויות הטמונות לניצולו ועל כן גם עשוי להשפיע על מחירו בהתאם.

1. מספר הקומות ((:

אנו מאמינים שישנה השפעה של מספר הקומות על מחיר הנכס. מספר הקומות עשוי להצביע על סגנונות בניה שונים וסגנון בניית הבית יכול להשפיע בצורה חיובית או שלילית על מחיר הנכס ועל האטרקטיביות שלו. כמו כן, מספר הקומות בבית יכול להעיד בנוסף על הנגישות/חוס הנגישות של הנכס לבעלי מוגבלויות ועקב כך גם להשפיע על האטרקטיביות של הנכס בעיני הרוכשים הפוטנציאליים ועל מחירו.

1. מצב הנכס ((:

לדעתנו מצב הנכס הוא משתנה שעשוי להשפיע רבות על מחיר הנכס שכן ביקוש לנכס במצב טוב עשוי להיות גבוה יותר מאשר לנכס במצב לא טוב הדורש תיקונים. רוכש פוטנציאלי עשוי לקחת בחשבון את ההוצאות הכלולות ברכישת נכס במצב לא טוב ודבר זה יכול להשפיע על המחיר של הנכס.

1. שנת בניית הנכס ((:

שנת הבניה של הנכס יכולה להשפיע על סגנון הבניה של הנכס, על התשתיות ועל רגולציה (לדוגמה פינוי בינוי עתידי). כל אלו שיקולים שעשויים להשפיע על מחירו של הנכס.

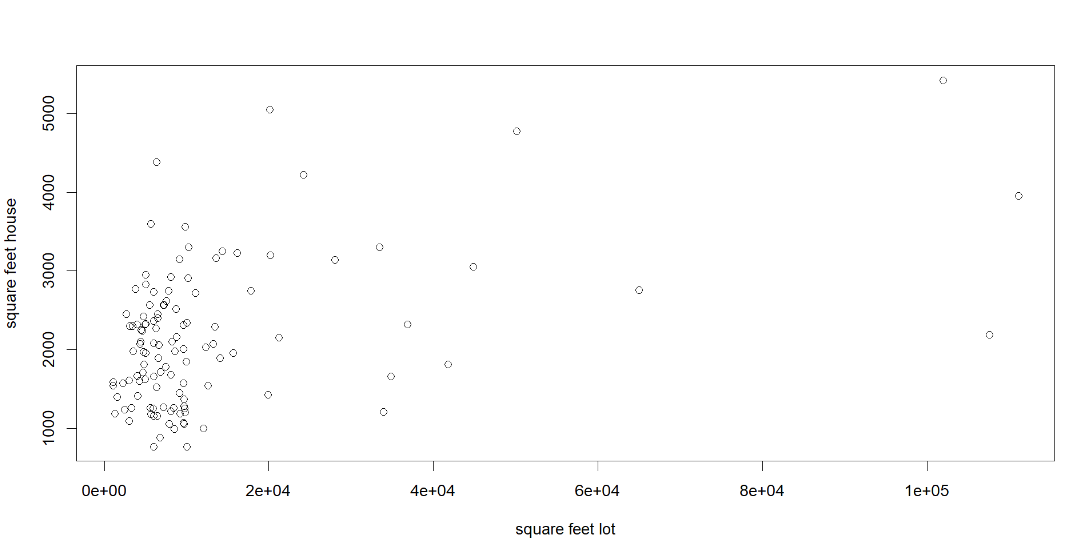
1. האם הנכס שופץ((:

שיפוץ נכס עשוי לקרות בעקבות מגוון סיבות, החל בשדרוג וכלה בתיקונים בשל בעיות חוזרות. על כן היותו של נכס משופץ עשויה להוביל לעליה או ירידה במחירו של הנכס.

# תיאור קשרים העיקריים בין המשתנים המסבירים ובין המשתנים המסבירים למוסבר:

הקשר בין גודל המגרש וגודל הבית:

בדקנו האם וכיצד משפיע גודל המגרש על גודל הבית. הנחנו כי ככל שהמגרש גדול יותר כך גם שטח הבית יהיה גדול בהתאם. הנחה זו נובעת מהעבודה שבשטח גדול יותר ניתן לבנות נכס גדול יותר לעומת שטח קטן אשר מגביל את הבונה ועל כן מדובר בקשר סיבתי. אך בניגוד לציפייה שלנו לקשר חזק, מצאנו כי הקורלציה בין המשתנים היא בינונית וערכה 0.4276966, וכן למרות שגדל שטח המגרש, שטח הבית לא גדל בסדר גודל דומה ונותר יחסית באותו טווח ערכים. אנו חושבים כי עשוי להיות גורם שלישי המגביל את שטח הנכס (לדוגמה היתרי בנייה המגבילים את גודל הנכס).



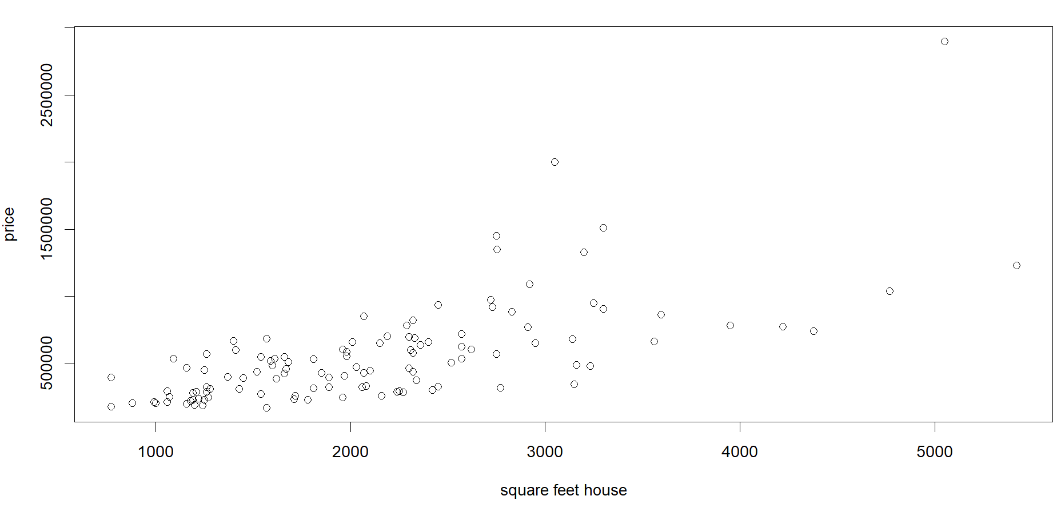
הקשר בין גודל הבית למספר החדרים:

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיבדקנו האם גודל הבית משפיע על מספר החדרים בו. הנחנו כי ככל ששטח הנכס גדול יותר, הרי שהוא מכיל בתוכו מספר רב יותר של חדרים. הנחה זו נובעת מכך שכמות חדרים גדולה גם דורשת שטח גדול יותר בנכס מאשר מספר נמוך של חדרים ועל כן מדובר בקשר סיבתי. לאחר ניתוח הגרף נוכחנו לראות כי גם כאן להפתעתנו מתקבלת קורלציה בינונית שערכה 0.4696185. לדעתנו קורלציה זו היא בינונית מפני שככל שהשטח גדל יש פחות תצפיות של כמות חדרים קטנה, אך מנגד לא ניתן למצוא עליה בכמות החדרים מעבר ל5 חדרים בשטחים הגדולים יותר. בנוסף בשטחים הקטנים יחסית, נראה שכלל לא קיימים בתים עם 4 חדרים ומעלה.

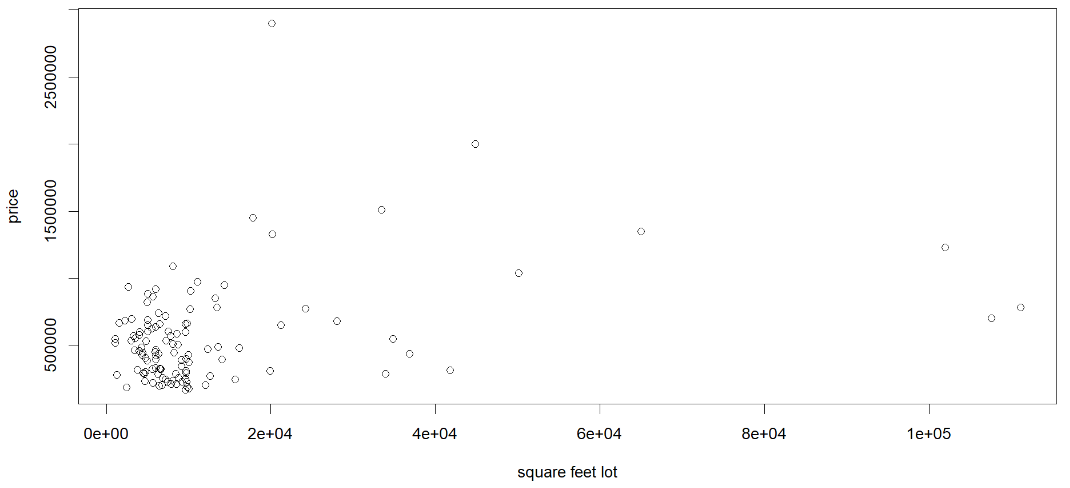
הקשר בין גודל הנכס למחירו:

בדקנו כיצד משפיע שטח הנכס על מחירו. הנחנו כי תהיה קורלציה גבוהה בין השניים מאחר וככל שהשטח גדול יותר, הבית יהיה אטרקטיבי יותר ומחירו יגדל בהתאם. מתקבלת קורלציה של 0.6796191 והיא יחסית מתאימה להשערתנו. נראה כי מדובר בקשר סיבתי מאחר והנכס, כמוצר, נחשב לאטרקטיבי יותר כאשר הוא גדול יותר ולכן ניתן לדרוש מחיר גבוה יותר.



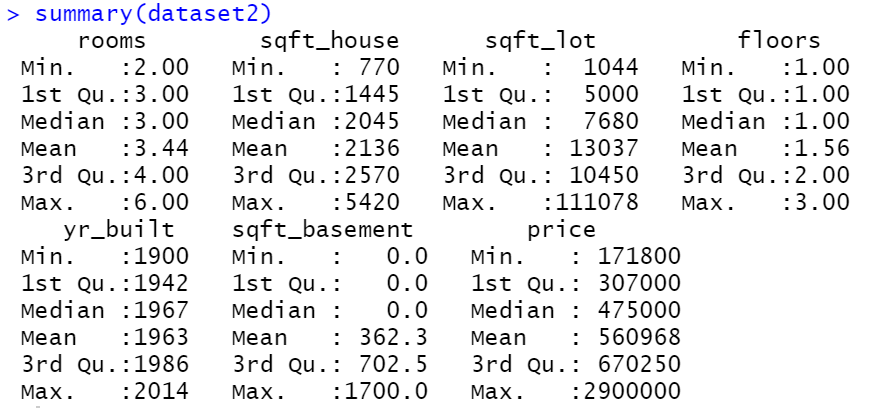
הקשר בין גודל השטח לבין מחיר הנכס:

בדקנו האם ישנו קשר בין גודל השטח עליו בנוי הנכס לבין מחיר הנכס. הנחנו, בדומה להשפעה של גודל הבית על מחירו, שתהיה קורלציה גבוהה גם בין שטח הנכס לבין המחיר. מדובר בקשר סיבתי מפני שאם מתמחרים נכס באמצעות השטח עליו הוא מוקם, הרי שככל שהשטח יהיה גדול יותר, משלמים על יותר שטח ולכן מחירו עולה בהתאם. אך מתקבלת קורלציה של 0.3405972 המתארת קורלציה יחסית נמוכה, שאיננה תואמת את השערותינו. הסיבה לכך טמונה כנראה במשתנה נוסף שעשוי להיות מיקום השטח. כך למשל שטח גדול במדבר עשוי להיות מתומחר נמוך יותר משטח קטן במרכז העיר.

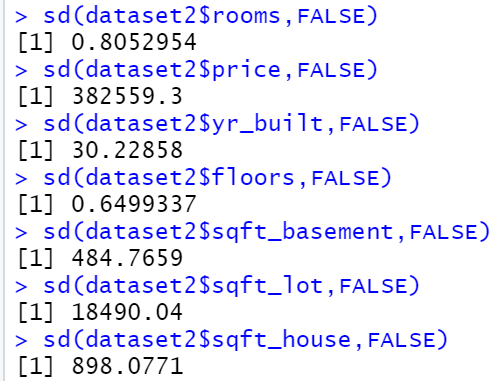


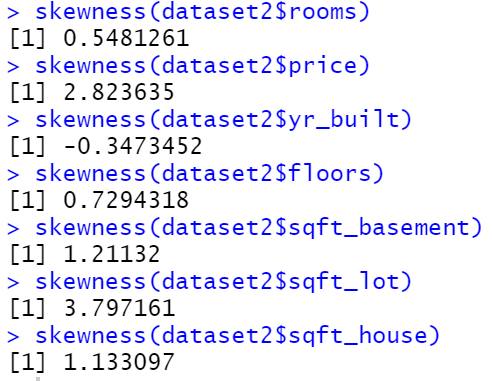
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| mean | median | SD | Interquartile range | 1st Q-25% | 3rd Q-75% | skewness |  |
| 560968 | 475000 | 382559.3 | 363250 | 307000 | 670250 | 2.823635 | Price($) |
| 1963 | 1967 | 30.22858 | 44 | 1942 | 1986 | 0.3473452- | yr\_built(years) |
| 1.56 | 1 | 0.6499337 | 1 | 1 | 2 | 0.7294318 | Floors(number) |
| 362.3 | 0 | 484.7659 | 702.5 | 0 | 702.5 | 1.21132 | sqft\_basement (Square Feet) |
| 13037 | 7680 | 18490.04 | 5450 | 5000 | 10450 | 3.797161 | sqft\_lot )Square Feet) |
| 2136 | 2045 | 898.0771 | 1125 | 1445 | 2570 | 1.133097 | sqft\_house (Square Feet) |
| 3.44 | 3 | 0.8052954 | 1 | 3 | 4 | 0.5481261 | Rooms(number) |

# ניתוח תיאורי של המשתנים הרציפים:

לצורך ניתוח המשתנים נעזרנו בפונקציית summary עבור הממוצע, החציון, והתחום בין-רבעוני:

לטובת חישוב סטיית התקן השתמשנו בפונקציית SD:



לטובת חישוב אסימטריה השתמשנו בפונקציית skewness:

א. מחיר:  
כפי שאנו רואים הממוצע והחציון אינם מאוד קרובים, כמו גם הרבעון הראשון והשלישי. סטיית התקן גם היא גדולה דבר שמצביע על שונות גדולה בנתונים, כלומר טווח המחירים עבור הנכסים השונים הוא יחסית רחב. התחום הבין רבעוני והאסימטריה החיובית בנתונים מצביעים על כך שיש לנו זנב ימני בהתפלגות המחיר.

ב. שנת הבניה:

במשתנה זה הממוצע והחציון כבר קרובים יותר אחד לשני, אך עדיין ישנה שונות שאינה קטנה במיוחד

בנתונים. מזאת נגזר שישנם נכסים שנבנו בשנים מוקדמות מאוד ביחס לאחרים. גם כאן ניתן לראות כי התחום הבין רבעוני יחסית גדול, אך הפעם הוא עם מדד אסימטריה שלילי שמסביר את הזנב השמאלי בהתפלגות.

ג. מספר הקומות:  
גם במשתנה זה הממוצע והחציון קרובים וסטייה לא גדולה .לומדים מכך שרוב הבתים נבנו במספר דומה של קומות ללא תנודות גדולות יחסית. מדד אסימטריה חיובי שמסביר את הזנב הימני בהתפלגות.

ד. שטח המרתף:

כאן ניתן לראות כי ההפרש בין הממוצע לחציון גדול יחד עם סטיית תקן גדולה בנתונים, דבר שמסביר את העובדה כי לפחות למחצית אין מרתף כלל (חציון = 0). בנוסף ניתן לראות שמדד האסימטריה חיובי, דבר שמצביע על הימצאות זנב ימני.

ה. גודל המגרש

שוב ניתן לראות כי ההפרש בין הממוצע לחציון גדול יחד עם סטיית תקן גדולה בנתונים, דבר שמסביר את העובדה שיש כמות נכסים בעלי שטח מגרש גדול. בנוסף ניתן לראות שמדד האסימטריה חיובי וגדול יחסית דבר שמצביע על הימצאות זנב ימני ארוך.

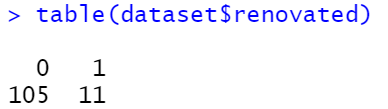
ו. גודל הנכס:

כאן דווקא אנו יכולים לראות שההפרש בין הממוצע לחציון אינו גדול במיוחד, אך סטיית התקן בנתונים היא די גבוהה. תופעה זו יכולה להיגרם מכך שהקצוות מאוד מנוגדים באופן כזה הגורם לכך שהם מקזזים את הממוצע ומקרבים אותו לחציון אך יוצרים שונות גבוהה. בנוסף ניתן גם כאן לראות שמדד האסימטריה חיובי, דבר שמצביע על הימצאות זנב ימני.

ז. מספר חדרים:  
גם במשתנה זה הממוצע והחציון קרובים וסטייה לא גדולה .ניתן להבין שרוב הבתים נבנו עם מספר חדרים דומה ללא תנודות גדולות יחסית. מדד אסימטריה חיובי שמסביר את הזנב הימני בהתפלגות.

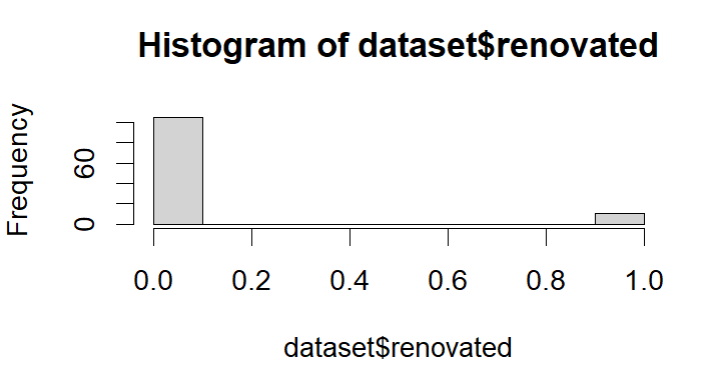
ניתוח תיאורי של המשתנים הקטגוריאליים:

לטובת משתנה renovated המקבל ערך 1 אם הנכס שופץ או 0 במידה ולא, הפקנו היסטוגרמה המראה את כמות התוצאות מכל קטגוריה:



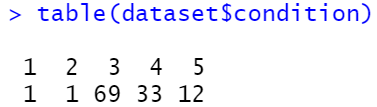
כלומר הפרופורציה עבור ערך 1 היא 11/116

והפרופורציה עבור ערך 0 היא 105/116.



לטובת משתנה condition המקבל ערך בין 1-אם הנכס במצב לא טוב לבין 5 אם הנכס במצב טוב.

הפקנו היסטוגרמה המראה את כמות התוצאות מכל קטגוריה:

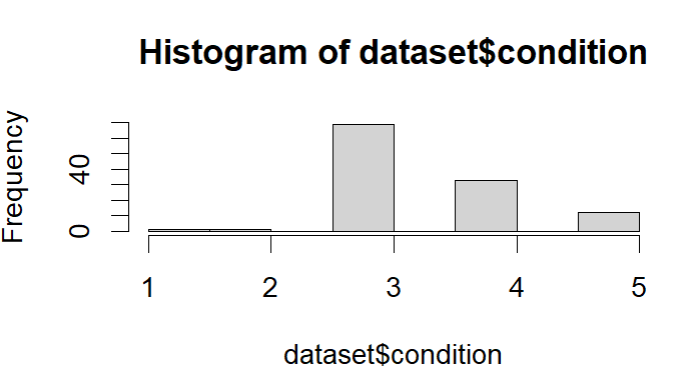


כלומר הפרופורציה עבור ערך 1 היא 1/116

כלומר הפרופורציה עבור ערך 2 היא 1/116

כלומר הפרופורציה עבור ערך 3 היא 69/116

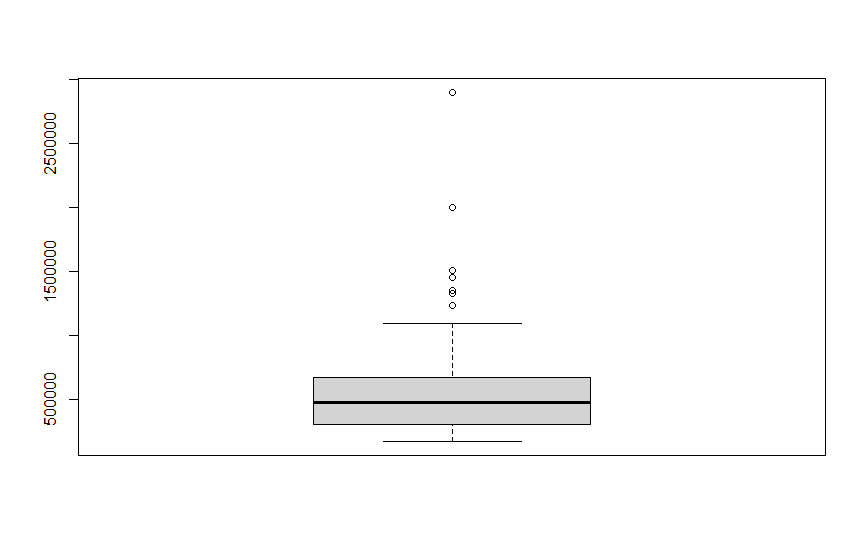
כלומר הפרופורציה עבור ערך 4 היא 33/116

כלומר הפרופורציה עבור ערך 5 היא 12/116

# ניתוח חריגים:

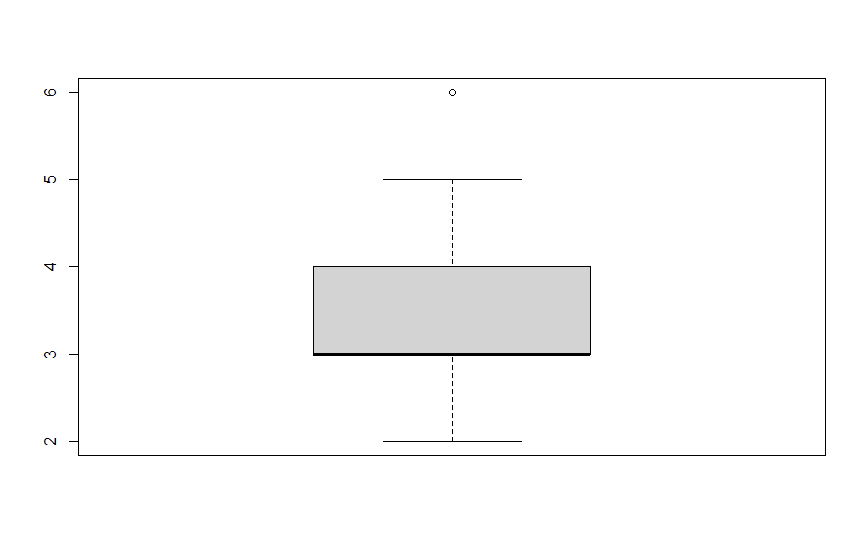
מחיר ((:

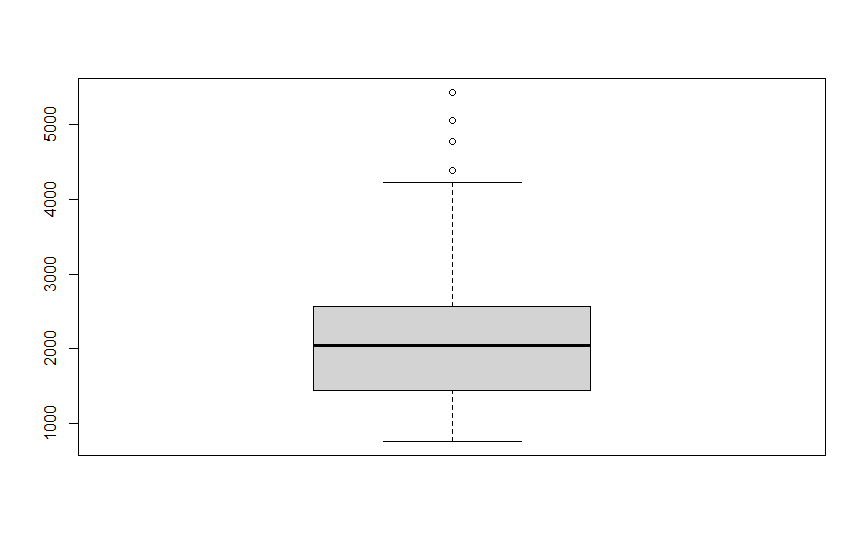
ניתן לראות כי ישנם 7 חריגים במדגם אשר מחירם גבוה בהרבה משאר הנכסים. החלטנו לא לחריג את נתונים אלו וזאת מכיוון שהגיוני שיהיו נכסים שיהיו יקרים בהרבה מנכסים אחרים בערים או שכונות מסוימות, שכוללות מאות נכסים מסוגים ומחירים שונים למשל יכול להיות שהתצפיות האלה מייצגות וילות יוקרה לעומת דירות אחרות שנחשבות דירות ממוצעות.



מספר החדרים בבית ((:

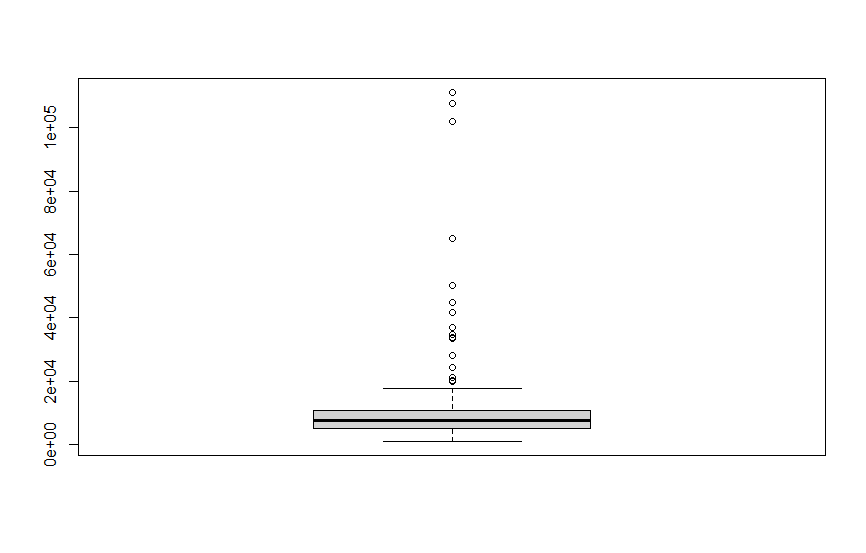
ניתן לראות כי ישנה תצפית חריגה אחת במדגם עבור נכס בעל מספר חדרים גבוה משאר הנכסים. החלטנו לא לחריג את נתונים אלו וזאת מכיוון שהגיוני שיהיו נכסים בעלי מספר רב של חדרים בהרבה מנכסים אחרים בערים או שכונות מסוימות, שכוללות מאות נכסים מסוגים שונים למשל יכול להיות שהתצפיות האלה מייצגות נכס המיועד למשפחות מרובות ילדים עם הרבה חדרים לעומת דירות מגורים המאופיינות בפחות חדרים למשפחות בגודל ממוצע.

  
גודל הבית ((:

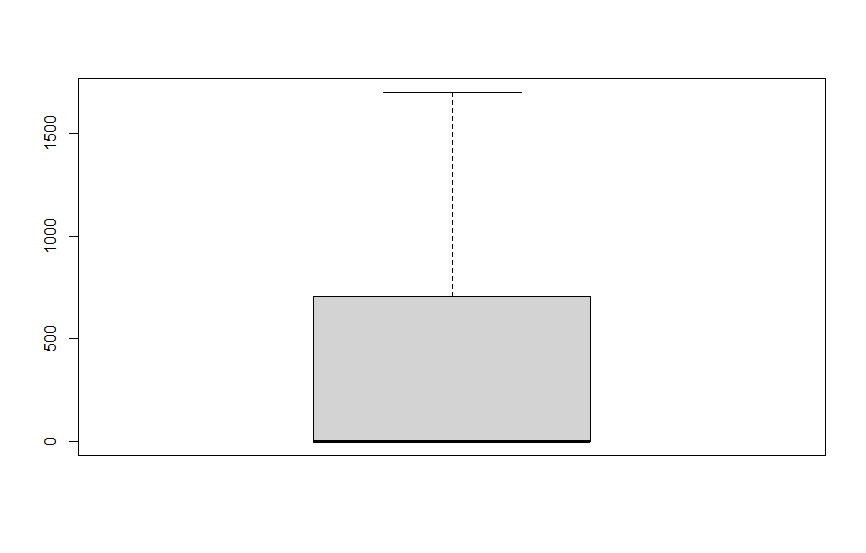
ניתן לראות כי ישנן 4 תצפיות חריגות במדגם עבור נכסים בעלי שטח גדול משאר הנכסים.  
החלטנו שלא לחריג נתונים אלו מפני שהגיוני שיהיו נכסים בעלי שטח בית בהרבה מנכסים אחרים בערים או שכונות מסוימות, שכוללות מאות נכסים מסוגים שונים למשל יכול להיות שהתצפיות האלה מייצגות נכסים בעלי שטח בית גדול.  


גודל המגרש ((:

ניתן לראות כי ישנן 14 תצפיות חריגות במדגם עבור נכסים בעלי שטח מגרש גדול משאר הנכסים.  
החלטנו לא לחריג את נתונים אלו וזאת מכיוון שהגיוני שיהיו נכסים שיהיו בעלי שטח מגרש גדול בהרבה מנכסים אחרים בערים או שכונות מסוימות, למשל יכול להיות שהתצפיות הללו מייצגות נכסים בעלי חצר גדולה וחניה.

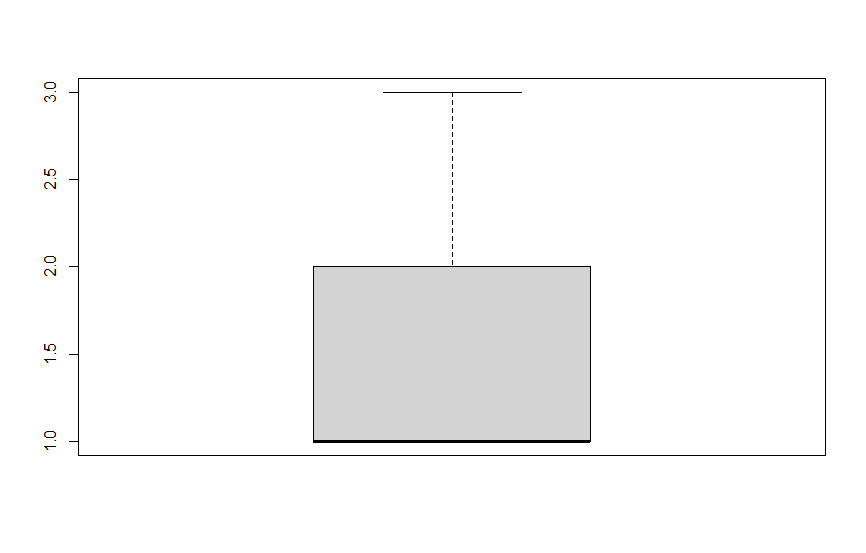


גודל המרתף ((:  
לא התקבלו כלל תוצאות חריגות.

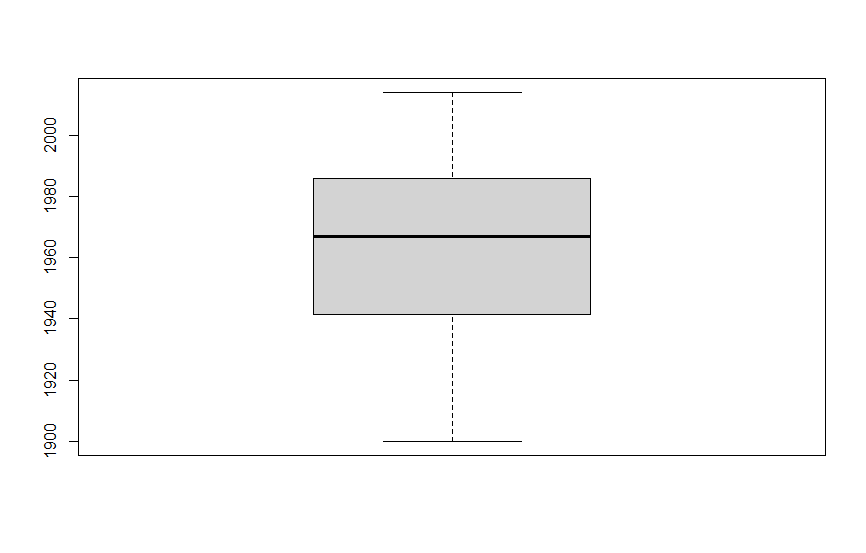


מספר הקומות ((:

לא התקבלו כלל תוצאות חריגות.

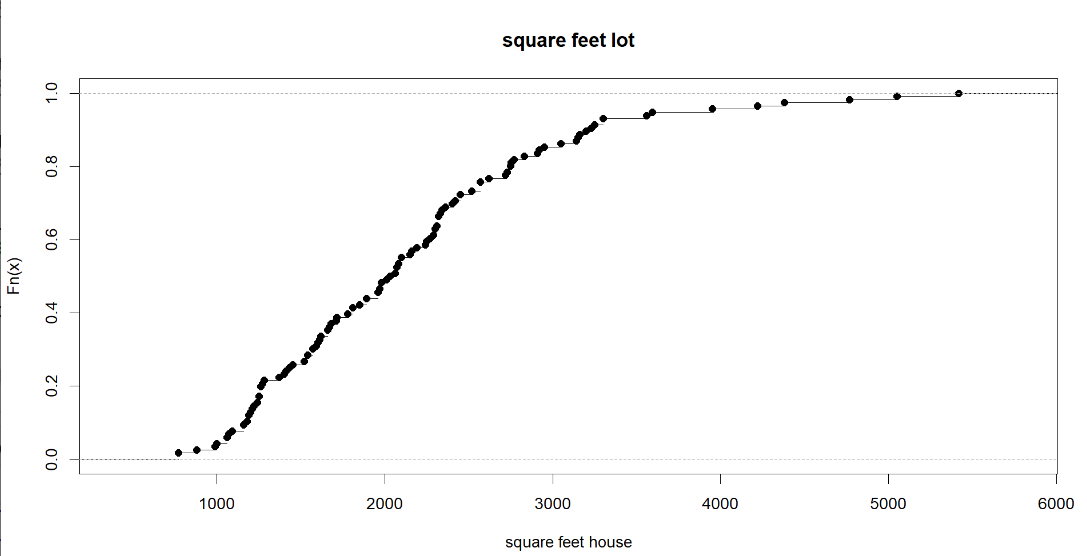
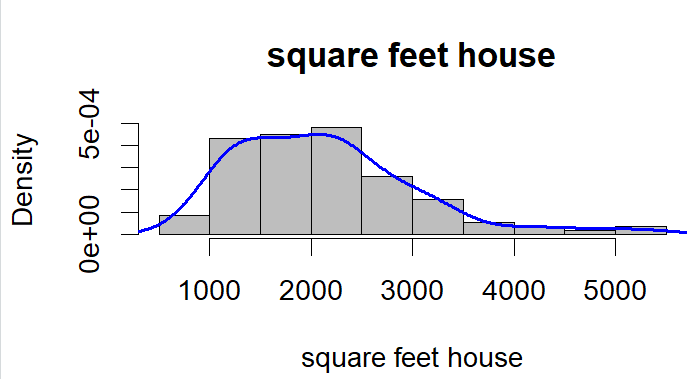


שנת בניית הנכס ((:

לא התקבלו כלל תוצאות חריגות.

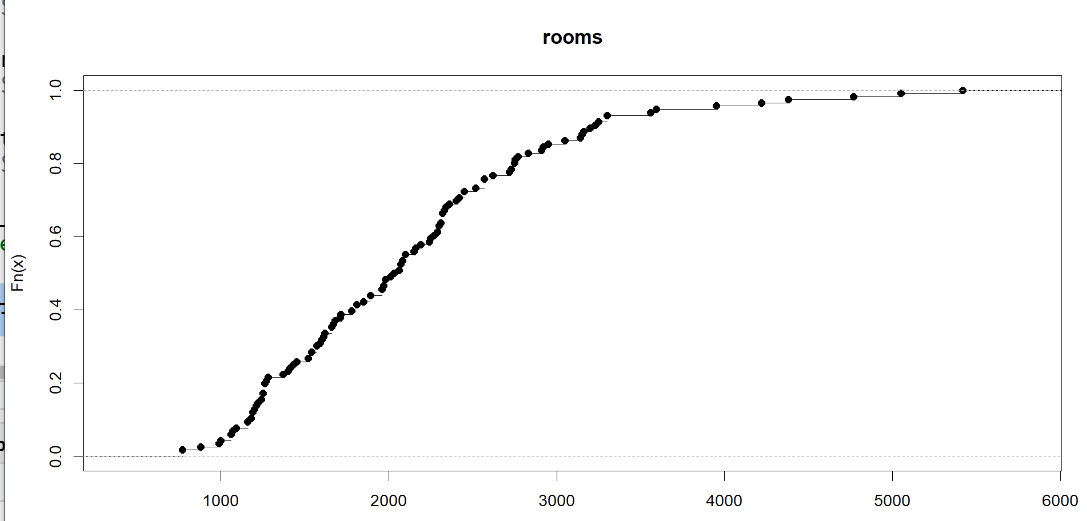
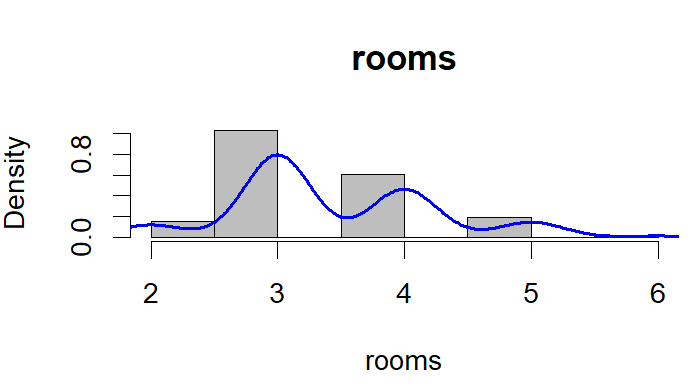
# פונקציית צפיפות והתפלגות מצטברת עבור שלושה מסבירים:

**עבור המסביר הראשון גודל הבית ((:**  
ניתן לראות כי השיא התצפיות נמצא בטווח של בין 1000-2500 כמו כן ניתן לראות שבפונקציית הצפיפות שככל שגודל הבית גדל, כך כמות התצפיות של בתים בעלי שטח גדול יורדת ולכן נוצר זנב ימני. באופן הגיוני, מספר הבתים הבולטים בגודלם הוא נמוך יותר ממספר הבתים בעלי גודל ממוצע מאחר ולרוב בתים מותאמים למספר הנפשות המשתכנות בהם ועל כן משפחה סטנדרטית מסתפקת בבית סטנדרטי. בנוסף, ניתן לראות שהנתונים תואמים את מגמת פונקציית ההתפלגות, יש צפיפות גבוהה בגודל בית הנפרש על שטח קטן לעומת הצפיפות הנמוכה בשטחים גדולים.



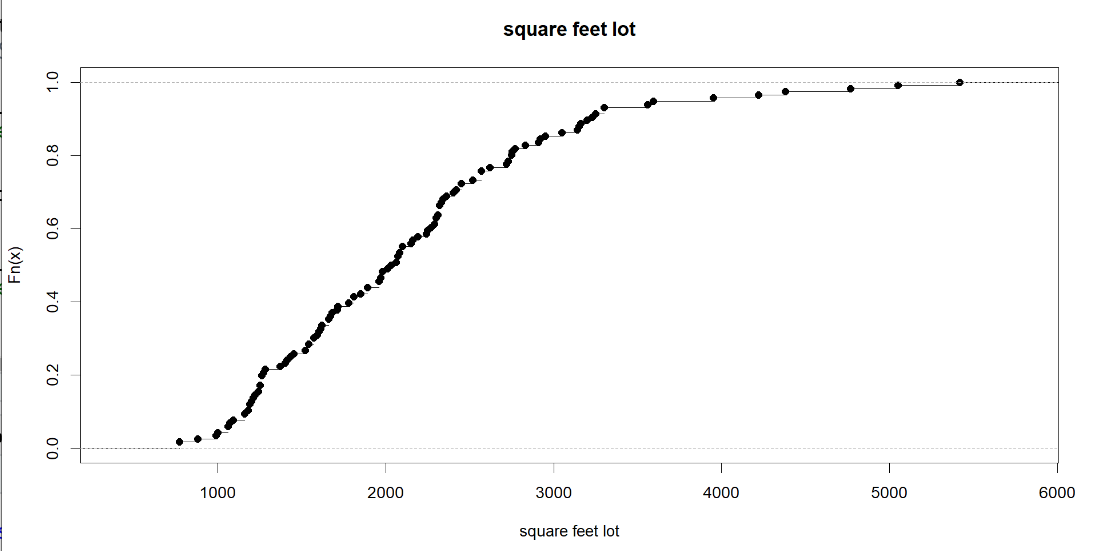
**מספר החדרים בבית ((:**

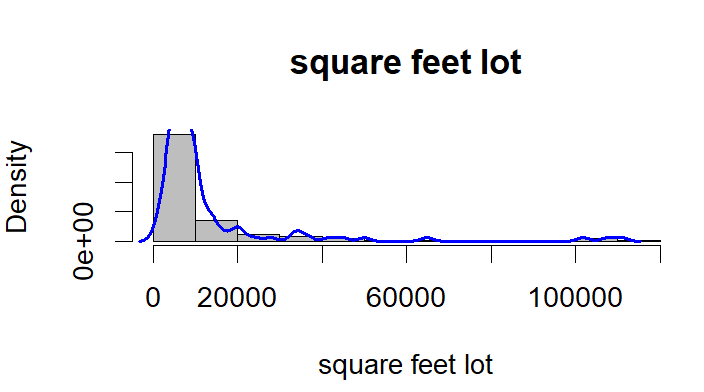
ניתן לראות כי שיא הגרף מתייחס 3 חדרים לנכס , ניתן לראות מפונקציית הצפיפות שככל שמספר החדרים עולה מספר התצפיות יורד אך גם לכיוון השמאלי כמעט ואין תצפיות של נכסים בעלות 2 חדרים.  
הצפיפות הקטנה בצד הימני יוצר זנב ימנית.  
ניתן לראות שהנתונים תואמים את מגמת פונקציית ההתפלגות, באזור 2 החדרים הצפיפות נמוכה ולאחר מכן גדלה לפיק של 3 ולאחר מכן מגמת הצפיפות קטנה.



**גודל המגרש ((:**

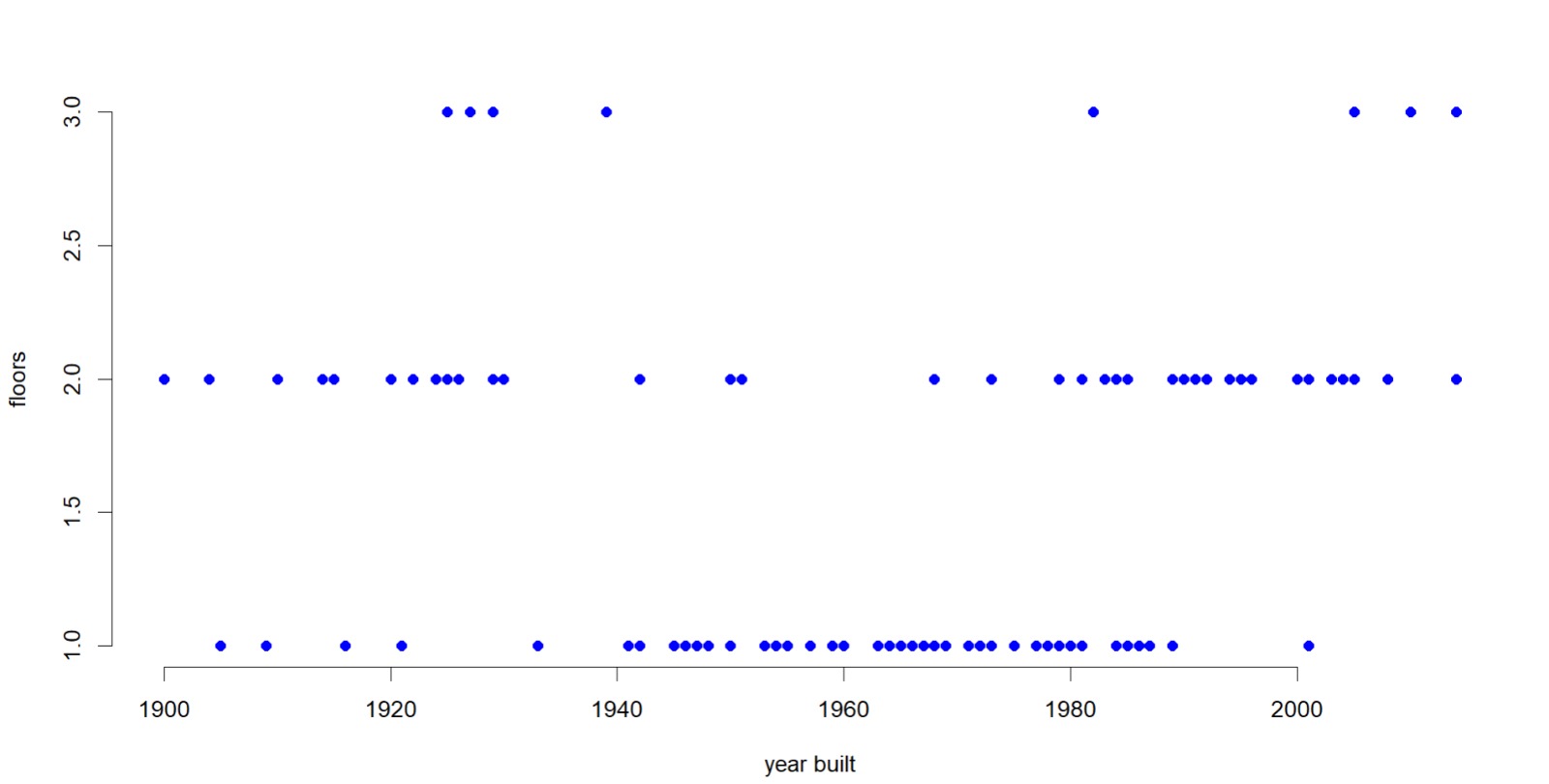
ניתן לראות כי השיא התצפיות נמצא בטווח של בין 0-10,000 כמו כן ניתן לראות שבפונקציית הצפיפות שככל שגודל השטח גדל, כך כמות התצפיות של בתים בעלי שטח גדול יורדת ולכן נוצר זנב ימני. באופן הגיוני, מספר השטחים הבולטים בגודלם הוא נמוך יותר ממספר הבתים בעלי גודל ממוצע בנוסף, ניתן לראות שהנתונים תואמים את מגמת פונקציית ההתפלגות, יש צפיפות גבוהה בגודל מגרש הנפרש על שטח קטן לעומת הצפיפות הנמוכה בשטחים גדולים.

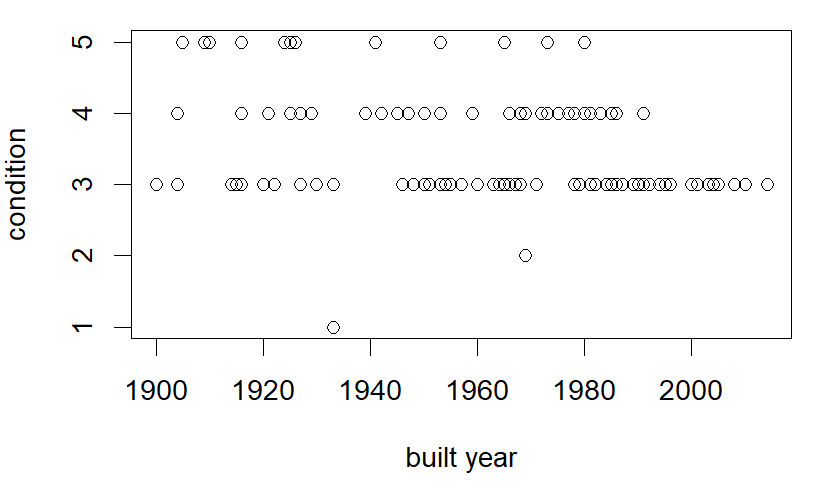




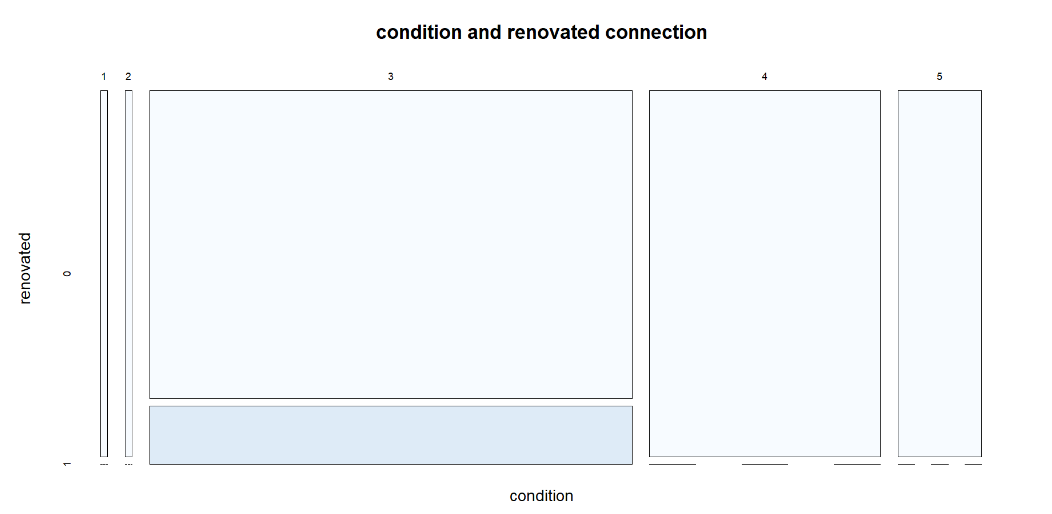
# ייצוג קשרים בעזרת תרשימים:

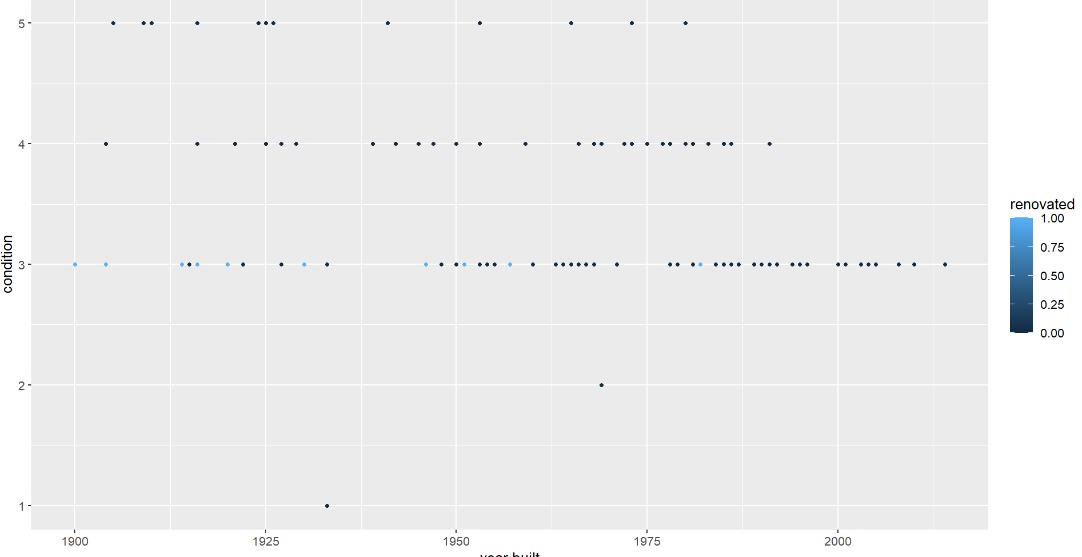
רצינו לבחון האם יכול להיות קשר בין שנת בניית הנכס לבין מספר הקומות של אותו הנכס.  
שיערנו כי בשנים שונות הבנייה תושפע מסגנונות ואופנה שונה או לחלופין מיכולות טכנולוגיות ואחד מהפרמטרים שיוכלו להשפיע הוא מספר הקומות.  
ניתן לראות שאכן שסביבות שנות ה-20 נבנו בעיקר 2 קומות לנכס אם מעט יוצא דופן בעלי 3 קומות או דירות קרקע, אך משנות ה-40 עד ל-80 הסגנון משתנה לגמרי וכמעט כל הבתים הם דירות קרקע.  
בשנות ה-80 והלאה האופנה משתנה שוב וחוזרים לבנות נכסים בעלי 2 קומות.



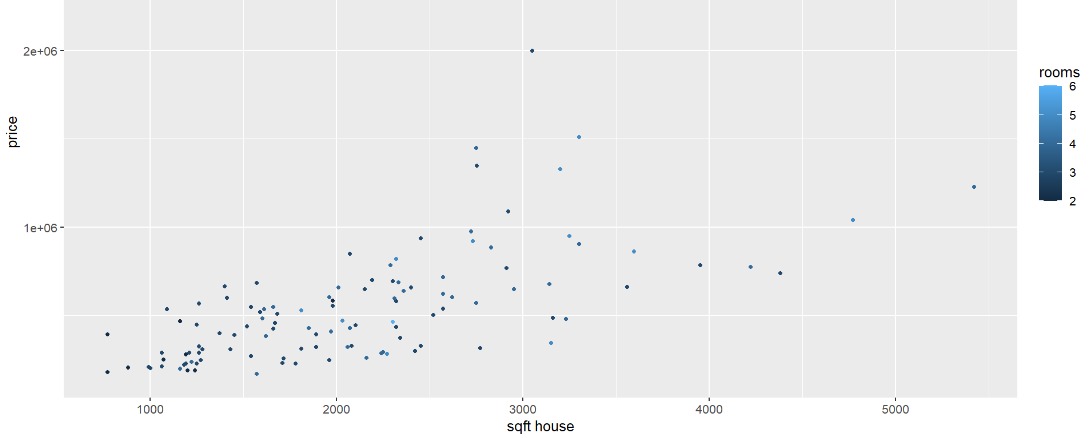
שיערנו כי יהיה קשר בין מצב הנכס לבית שנת הבנייה שלו , הגיוני שנכסים ישנים יותר יהיו במצב פחות טוב ושנכסים חדשים יחסית יהיו במצב יחסית טוב יותר.  
להפתעתנו ברוב השנים לא היה שינוי משמעותי ורוב הנכסים נמצאים במצב 3.  
דווקא עבור הנכסים החדשים ביותר משנות ה90 והלאה כלל לא נרשמו תצפיות במצב 4 או 5 דבר שאולי מצביע על בנייה באיכות ירודה.  
נתון מפתיע נוסף הוא ש-7 תצפיות מתוך 12 בעלות דירוג 5 הגבוה ביותר נמצאות בין שנות 10-30 נתון זה יכול לנבוע מהעובדה שבתים אלה מאוד ישנים וכנראה דרשו שיפוץ שהעלה את דירוגם ל-5.

בנוסף, מאוד עניין אותנו לבדוק האם דירוגם של בתים שעברו שיפוץ בהכרח יהיה גבוה יותר. מפני ששיפוץ של בית יכול לנבוע ממספר סיבות, החל משיפוץ בעקבות מצב רעוע המחייב שיפוץ בתשתית, בקירות הבית ועוד. או שיפוץ לטובת שדרוג הבית, הגדלת שטח, שפצור, התחדשות. אך למרבה התדהמה, נראה שרוב הבתים כלל לא עברו שיפוץ ואלו שכן עברו שיפוץ מדורגים בציון 3. דבר שמנע מאיתנו להסיק האם שיפוץ הבית מוביל לעלייה או לירידה בדירוג. קשה אף להצביע על קיום קשר כלשהו בין השניים, אך קל לומר שרוב הבתים כלל לא עוברים שיפוץ.



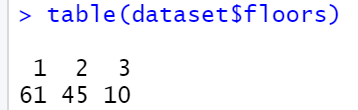
  
  
לאחר מכן הסתקרנו לגלות כיצד והאם שנת הבנייה של הבית משפיעה על מצבו, תוך אבחנה האם הבית שופץ או לא. מאוד סקרן אותנו לגלות האם בתים ישנים יותר עברו שיפוץ ומה מצבם. נראה שבתים משנות ה2000 נמצאים במצב 3 ולא עברו שיפוץ. בעוד שבתים ישנים מאוד (שנת בנייה 1900) באותו המצב דווקא כן עברו שיפוץ.

Year built

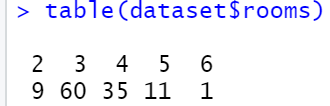
קשר מעניין נוסף שרצינו לבחון הוא הקשר בין גודל הבית לבית לבין מחירו, אך תוך מתן דגש לבחינת מספר החדרים. רצינו לראות האם יש שוני במחיר בין בתים בעלי אותו השטח אך עם מספר חדרים שונה. אך לא גילינו קשר בולט ביניהם.

# טבלאות שכיחות :

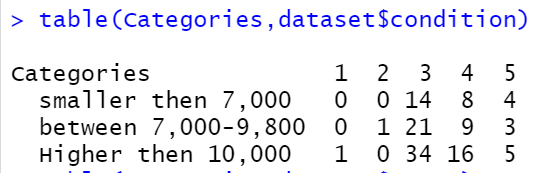
טבלאות שכיחות חד ממדיות-  
טבלה ראשונה- מספר הקומות ((:

ניתן לראות מהטבלה כי התקבלה כמות התצפיות הגבוהה ביותר עבור בתים בעלי קומה אחת, העומדת על 61 תצפיות כלומר מתוך המדגם שלנו כמות הקומות השכיחה ביותר היא קומה אחת. לאחר מכן הופיעו 45 תצפיות של בתים בעלי שתי קומות ולבסוף התקבלו 10 תצפיות בלבד של בתים בני 3 קומות. בתים בעלי 3 קומות הם הכי פחות נפוצים. ניתן להניח שרוב האוכלוסיה מעדיפה בתים בני קומה אחת.

טבלה שניה - מספר החדרים בבית ((:

כפי שמוצג בטבלה, ניתן לראות כי כמות החדרים השכיחה ביותר היא 3 חדרים עם *60 תצפיות, 35 תצפיות עבור בתים בהם 4 חדרים. כמו כן, נצפה רק נכס אחד בעל 6 חדרים, נתון בולט המדגיש את העובדה שכנראה האוכלוסייה הנדגמת בממוצע מסתפקת ב3-4 חדרים, ייתכן בעקבות מספר הנפשות החיות בבית, מכאן עולה שכנראה מספר הילדים במשפחה ממוצעת הינו 2-3 ילדים.*

טבלאות שכיחות דו ממדיות-  
לצורך הטבלאות הבאות, יצרנו משתנה חדש וסכמנו לתוכו שלושה משתנים – גודל הבית, גודל השטח וגודל המרתף. כך שמשתנה זה מהווה את הגודל הכולל של הנכס. לאחר מכן יצרנו טבלת שכיחות דו ממדית ואיחדנו את הערכים שהתקבלו ל3 קטגוריות:  
גודל נכס עד גודל 7000 (sqft)  
גודל נכס בתחום 7000-9800 (sqft)  
גודל נכס מעל 10000 (sqft)

עבור הטבלה הראשונה, שילבנו בין המשתנה שיצרנו לבין מצב הנכס ((:

נוצרה טבלת שכיחות באמצעותה בדקנו כמה תצפיות מכל מצב התקבלו בהפרדה לקטגוריות המבוססות על גודל הנכס הכולל. כך למשל ניתן לראות שעבור הקטגוריה הראשונה, נמצאו 14 בתים שמצבם מדורג – 3, 21 בתים בציון זהה מהקטגוריה השנייה ו34 בתים מהקטגוריה השלישית. כבר בחלקים קודמים הסקנו כי מצב הבית השכיח הוא 3, אך טבלה זו אפשרה לנו לפלח את מצבי הבתים לפי גודלם.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטיעבור הטבלה השנייה, יצרנו טבלת שכיחות באמצעותה בדקנו כמה תצפיות התקבלו עבור מספר חדרים שונה. ניתן לראות שנצפה רק בית אחד בעל 6 חדרים, וכן גם ניתן לאפיין את גודלו ולומר שהוא קטן מ7000(sqft). כמות החדרים השכיחה ביותר היא 3 חדרים. וניתן למצוא הכי הרבה תצפיות מעין אלו דווקא בבתים שגודלם מעל 10,000(sqft). אלו התוצאות מכיוון שמאוד סביר שבית המתפרש על שטח גדול יותר, גם יכול כמות גדולה יותר של חדרים. לכן גם אנו נתקלים ב6 תצפיות של בתים בעלי 5 חדרים, בעוד שבבתים הקטנים יותר נראו פחות תצפיות.