**פרויקט רגרסיה**

**חלק א**



Cancer mortality rate

קבוצה 18

**315407114 | 208080879 | 209060334**

**תוכן עניינים:**

טבלת משתנים .................................................................................................................................3

תיאור המשתנים ...........................................................................................................................3-4

תיאור קשרים בין משתנים ..............................................................................................................4-7

ניתוח תיאורי ................................................................................................................................7-9

ניתוח חריגים ............................................................................................................................10-13

פונקציית צפיפות והתפלגות מצטברת ..........................................................................................12-13

תרשים פיזור מורכב ....................................................................................................................... 14

ייצוג קשרים באמצעות תרשימים .................................................................................................14-15

טבלת שכיחויות ........................................................................................................................ 16-17

נספחים ....................................................................................................................................18-23

1. **בחירת מאגר נתונים-**

מאגר הנתונים שלנו הינו "Cancer Mortality Rate"

1. **טבלת משתנים:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **שם המשתנה** | **הסבר קצר על  המשתנה** | **סוג משתנה מוסבר/מסביר** | **סימון ב-R** | **יחידת מידה** | **סוג המשתנה -רציף/ קטגוריאלי** |
| **Country Index** | - | - | Country Index | - | - |
| **Cancer mortality rate** | שיעור התמותה ממחלת הסרטן | מוסבר | Cancer mortality rate | אחוז | רציף |
| **Obesity rate** | שיעור השמנת יתר | מסביר | Obesity rate | אחוז | רציף |
| **Smoke rate** | שיעור המעשנים | מסביר | Smoke rate | אחוז | רציף |
| **Air pollution index** | מדד זיהום אוויר | מסביר | Air pollution index | חסר יחידות | קטגוריאלי |
| **State Development Index** | מדד התפתחות במדינה | מסביר | State Development Index (0-1) | ללא יחידות | רציף |
| **Median age** | חציון הגיל | מסביר | Median age | שנים | רציף |
| **Average temperature** | טמפרטורה ממוצעת | מסביר | Average temperature | מעלות צלזיוס | רציף |
| **Percentage of residents in cities** | שיעור התושבים בעיר | מסביר | Percentage of residents in cities | אחוז | רציף |
| **Depression rate** | שיעור הלוקים בדיכאון | מסביר | Depression rate (%) | אחוז | רציף |

1. **תיאור המשתנים:**

**שיעור השמנת יתר:** שיעור זה מתאר את אחוז האנשים במדינה הסובלים ממשקל עודף. שיעור זה עשוי לתרום להבנת מגמות השינוי במשתנה המוסבר ע"י כך שיכול לשקף דפוסי סגנון אורח חיים ואיזון תזונתי לקוי אשר יובילו לחוסרים או עודפים ברכיבי תזונה. המצב עלול להחליש את מערכת החיסון ולהגביר את הסיכון לתאים סרטניים וכתוצאה מכך - את התמותה מסרטן.

**שיעור המעשנים:** שיעור המעשנים במדינה מתאר את אחוז האנשים במדינה אשר צורכים מוצרי סיגריות וטבק. משתנה זה עשוי להשפיע על מגמות בריאותיות, ומשמש להבנת השפעות ארוכות טווח על שיעורי תחלואה ותמותה מסרטן. עישון ידוע כגורם סיכון מרכזי לסוגי סרטן, במיוחד סרטן הריאות והלוע. החשיפה לרעלנים בסיגריות עלולה לגרום לנזק תאי מצטבר שמגביר את הסיכון לגידולים, ובכך להשפיע ישירות על שיעורי התמותה ממחלת הסרטן.

**מדד זיהום אוויר:** מדד זה עשוי לשקף את רמת החשיפה של האוכלוסייה לגורמים סביבתיים מזיקים. זיהום אוויר נמדד על ידי קטגוריות שונות על פי סדר עולה- . Low, Minor, Medium, High, Extreme זיהום האוויר עשוי להשפיע על הבריאות על ידי החדרת חלקיקים וחומרים מסרטניים לגוף. חומרים אלו עלולים לגרום לנזק לרקמות, דלקות ופגיעה בתאים, דבר המעלה את הסיכון להתפתחות מחלות סרטן, במיוחד הריאות, ובכך מסייע להבנת הקשר בין תנאים סביבתיים לשיעורי התמותה מסרטן.

**מדד התפתחות המדינה:** מדד זה מתאר את רמת הפיתוח הכלכלי, החברתי והבריאותי במדינה, ומדרג מדינות בין מתפתחות למפותחות (ערך בין 0 ל-1 בהתאמה). מדינות בעלות מדד פיתוח גבוה עשויות להציע גישה טובה יותר לשירותי בריאות מתקדמים, אבחון מוקדם, וטיפולים יעילים למחלות, כולל סרטן. לעומת זאת, מדינות בעלות מדד פיתוח נמוך עשויות להציג אתגרים בנגישות לטיפולים רפואיים ובאיכותם, מה שיכול להשפיע על שיעורי התמותה מסרטן. לכן, מדד זה עשוי להוות אינדיקציה למגמות בשיעורי התמותה בהתבסס על זמינות השירותים ותשתיות הבריאות במדינה.

**גיל חציוני:** משקף את הגיל שבו מחצית מהאוכלוסייה מבוגרת יותר ומחצית צעירה יותר. מדד זה יכול לתרום להבנת מגמות בשיעור התמותה מסרטן, שכן מחלות סרטן נפוצות יותר בגילאים מתקדמים. במדינות שבהן הגיל החציוני גבוה, נצפה ששיעור התמותה מסרטן יושפע ויהיה גבוה יותר באופן יחסי.

**טמפרטורה ממוצעת:** פרמטר משמעותי מתנאי האקלים השוררים בה. משתנה זה עשוי לתרום להבנת מגמות בשיעור התמותה מסרטן על ידי כך שמשקף את החשיפה לתנאי אקלים מסוימים. לדוגמה, אקלים חם עם חשיפה מוגברת לקרינת שמש עשוי להגביר את הסיכון לסרטן העור. בנוסף, טמפרטורה עשויה להשפיע באופן עקיף על הבריאות הכללית דרך שינויים בתזונה, ותהליכי חילוף חומרים, אשר יכולים להשפיע על סיכון לחלות בסרטן וכתוצאה מכך לתמותה מהמחלה.

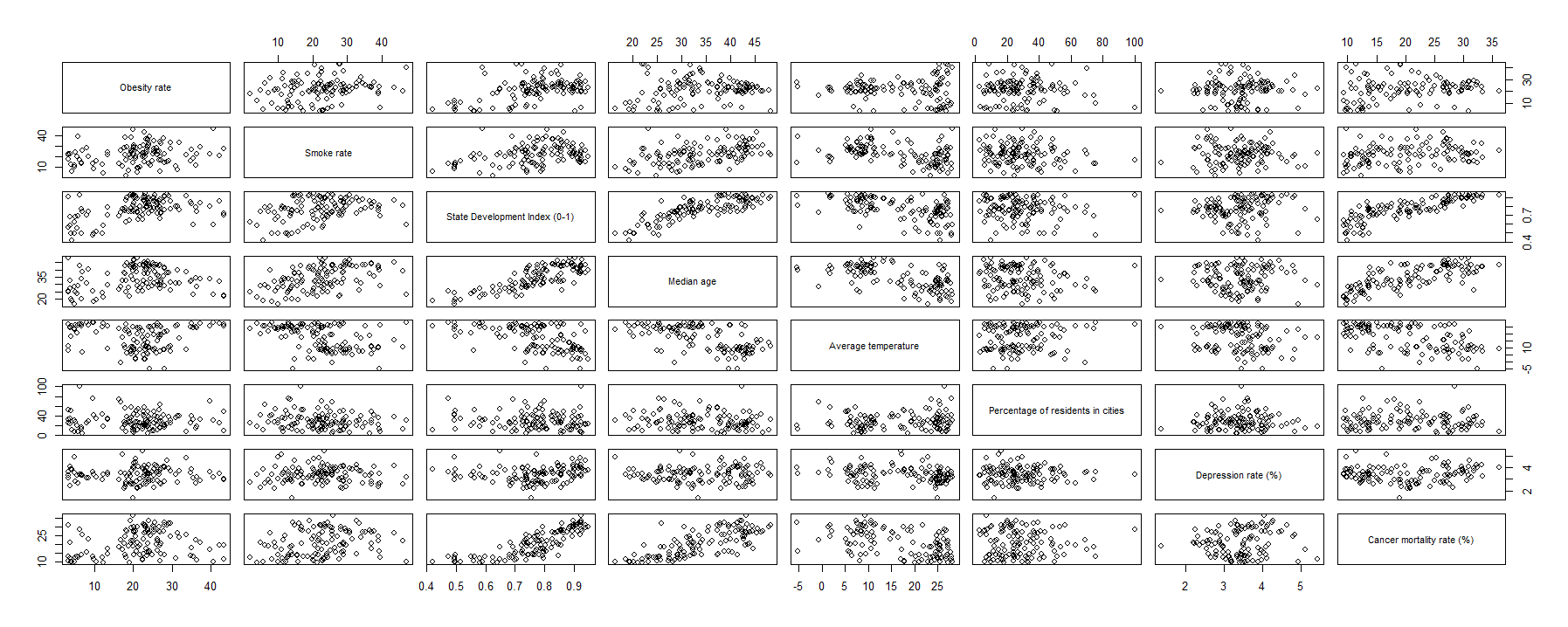
**שיעור התושבים בעיר:** מתאר את שיעור האוכלוסייה המתגוררת באזורים עירוניים. משתנה זה עשוי לתרום להבנת מגמות בשיעור התמותה מסרטן מכיוון שמשקף את דפוסי העיור במדינה. מצד אחד, אוכלוסיות עירוניות עשויות להיות חשופות לזיהום אוויר גבוה יותר, (מעלה את הסיכון לסרטן וכן לתמותה ממנו). מצד שני, בערים קיימת נגישות גבוהה לשירותי בריאות, אבחון וטיפול, אשר עשויים להפחית את שיעורי התמותה. לכן, יכול להוות אינדיקציה להיבטים שונים בהקשר מגמות תמותה מסרטן.

**שיעור הלוקים בדיכאון:** מתאר את שיעור האוכלוסייה המאובחנת עם דיכאון. משתנה זה עשוי לתרום להבנת מגמות בשיעור התמותה מסרטן בכך שהוא משקף את מצב הבריאות הנפשית במדינה. דיכאון עלול להוביל להזנחת בדיקות רפואיות ואבחון מוקדם, וכן לפגיעה באורח החיים הבריא, מה שעשוי להשפיע על עלייה בשיעור התחלואה והתמותה ממחלות כרוניות, כולל סרטן. ניתוח קשר זה יכול לספק תובנות על קשר חיובי בין תפקוד לקוי של הבריאות הנפשית על תמותה מסרטן.

**יבשת:** היבשת שבה ממוקמת המדינה לעיתים יכולה לבחון קשר בין מגמה אזורית לבין שיעור התמותה מסרטן על סמך היבטים תרבותיים, גנטיים, וכלכליים. לדוגמה, ההבדלים בין יבשות יכולים לכלול גישה לשירותי בריאות, חשיפה לסיכונים סביבתיים ותשתיות רפואיות. בנוסף, קיימת שונות גנטית בין אוכלוסיות מאזורים שונים בעולם, אשר יכולה להשפיע על הנטייה לפתח סוגי סרטן שונים. דוגמה יכולה להיות יבשת כמו אפריקה, שבה רמת החשיפה לשמש גבוהה, דבר היכול לעלות באופן מובהק את הסיכויים לסרטן העור בקרב האוכלוסייה במדינה. מנגד, בכל יבשת ישנן הרבה מדינות שונות אשר מתנהלות כל אחת בסגנון חיים אחר ובאקלים שונה ולכן, ייתכן שלא תמיד תהיה מגמה ברורה בין היבשת לשיעור התמותה מסרטן במדינה.

1. **תיאור קשרים בין משתנים:**בניתוח הקשרים בין משתנים המסבירים הרציפים, נבחין בין שלושה סוגי קשרים עיקריים:

* **קשר סיבתי**: מתאר מצב שבו משתנה X משפיע באופן ישיר על משתנה אחר Y.
* **קשר מדגמי**: קשר שמבוסס על נתוני המדגם בלבד ואינו בהכרח משקף קשר אמיתי או סיבתי. ייתכן שהקשר נובע ממשתנה שלישי Z שמשפיע על שני המשתנים.
* **קורלציה**: מדד סטטיסטי שמתאר את מידת העקביות והכיוון בקשר בין שני משתנים. לדוגמה, קורלציה חיובית מצביעה על כך ששני משתנים נוטים לעלות יחד, אך אין בכך בהכרח עדות לקשר סיבתי.

**תרשים plot על כלל המשתנים הרציפים:**

**הקשר בין גיל חציוני לשיעור המעשנים:**

תמונה שמכילה טקסט, צילום מסך, תרשים

התיאור נוצר באופן אוטומטי

בהתבסס על הגרף, ניתן לראות כי הקשר בין הגיל החציוני (Median age) לבין שיעור העישון (Smoke rate) הוא קשר מדגמי בינוני עם מגמה חיובית. הקורלציה בין המשתנים עומדת על 0.463. אנחנו משערות שמדובר בקשר מדגמי, כיוון שיתכן ששני המשתנים מושפעים מגורמים אחרים כמו מאפיינים חברתיים או תרבותיים. ייתכן שבמדינות בהם הגיל החציוני גבוה יותר שיעור המעשנים גבוה יותר כי עישון נפוץ יותר בקרב אוכלוסיות ותיקות כי זה היה מקובל יותר תרבותית בעבר.

**הקשר בין מדד הפיתוח המדיני שיעור הדיכאון:**

תמונה שמכילה טקסט, צילום מסך, תרשים, קו

התיאור נוצר באופן אוטומטי

גרף מראה קשר חיובי בין מדד הפיתוח המדינתי לשיעור הדיכאון. ככל שמדד הפיתוח עולה (בין 0 ל-1), יש נטייה לעלייה בשיעור הדיכאון, אך הקשר אינו חזק במיוחד. למרות שהקשר חיובי, לא ניתן להסיק בהכרח שמדד הפיתוח הוא הגורם הישיר לדיכאון. במדינות עם רמות פיתוח גבוהות יותר, יש גישה גבוהה יותר לשירותי בריאות ותמיכה נפשית, מה שמוביל לעיתים לגילוי של יותר מקרים של דיכאון. מצד שני, במדינות מפותחות יש גם יותר לחצים חברתיים ופסיכולוגיים, שעשויים להגביר את שיעור הדיכאון.  לכן אנו מניחות שמדובר בקשר מדגמי, בקורלציה חיובית (0.118), כיוון שאין מנגנון סיבתי ברור, וייתכן שגורמים נוספים, כמו גישה לשירותי בריאות או אורח חיים עירוני, משפיעים על שני המשתנים. במקרה זה ציפינו לראות מגמת עלייה גבוהה יותר בשיעור הדיכאון ככל שהמדינה מפותחת יותר, בגלל הלחצים הקיימים במדינות מפותחות ותחושת ניכור שעשויה להתפתח בהן בקרב הפרט. אך ייתכן שמודעות וטיפול זמין במדינות המפותחות דווקא מפחיתות את שיעור הדיכאון ביחס למצופה.

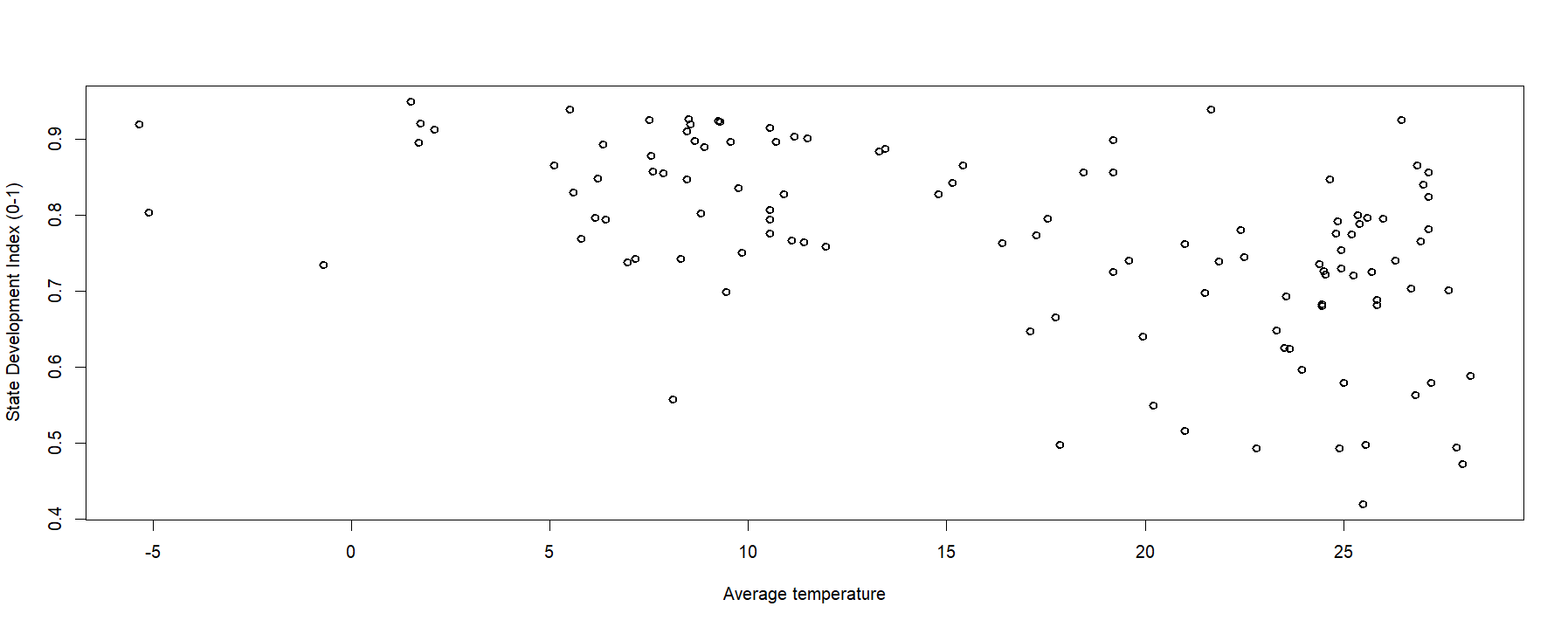
**הקשר בין מידת הפיתוח לגיל החציוני:**

**תמונה שמכילה טקסט, תרשים, צילום מסך, קו

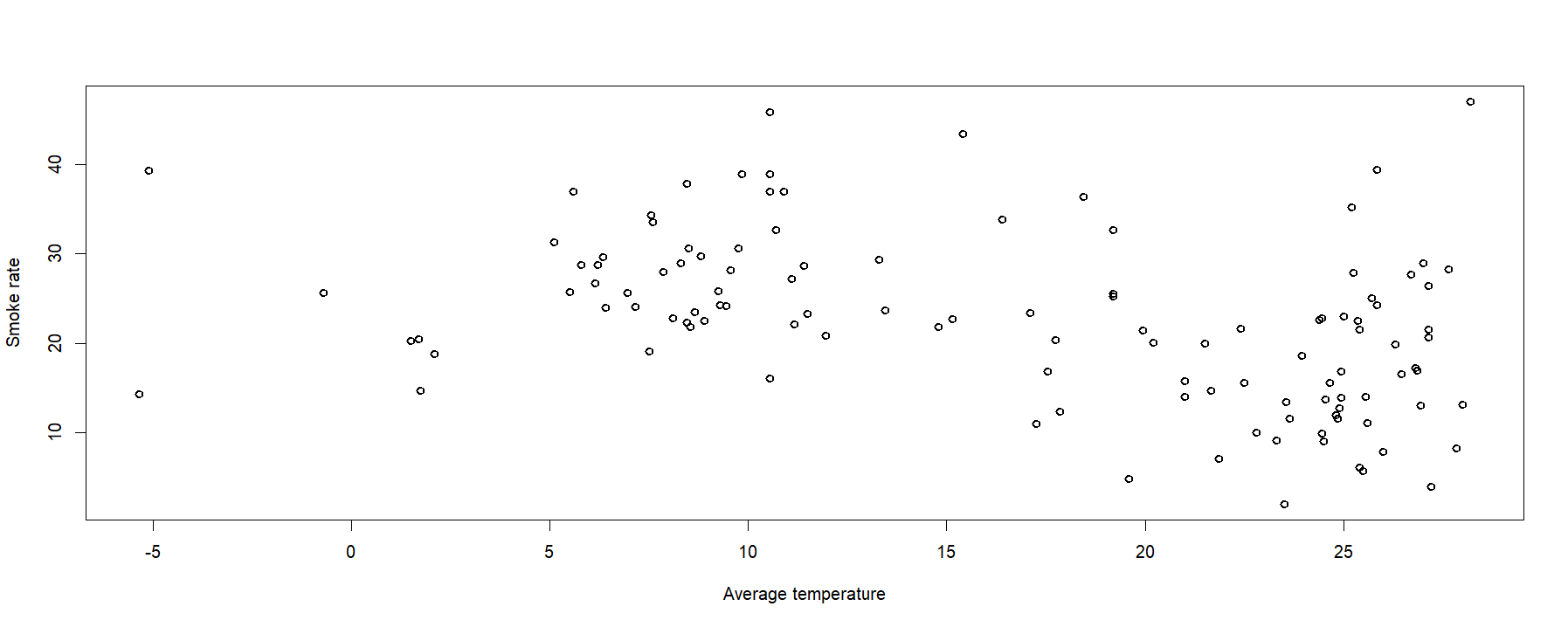
התיאור נוצר באופן אוטומטי**

בהתבסס על הגרף והקורלציה של 0.81, ניתן להסיק כי יש קשר חיובי חזק בין מדד הפיתוח המדינתי ל-הגיל החציוני. כלומר, ככל שמדד הפיתוח במדינה גבוה יותר, כך הגיל החציוני במדינה נוטה להיות גבוה יותר. ציפינו לתוצאות אלו מתוך הנחה שבמדינות עם רמות פיתוח גבוהות יותר יש אוכלוסייה מבוגרת יותר, כנראה כתוצאה מתוחלת חיים גבוהה יותר ושירותי בריאות מתקדמים. המדדים הסוציו-אקונומיים שמאפיינים מדינות מפותחות, כמו תנאי חיים טובים, חינוך בריאותי ותנאים כלכליים טובים, יכולים להוביל לכך שבמדינות אלו יהיו יותר אנשים בגילאים מבוגרים, מה שמגדיל את הגיל החציוני. אנו משערות שמדובר בקשר סיבתי כיוון שתנאי חיים טובים כתוצאה מפיתוח מדיני, גורמים להארכת תוחלת החיים והעלאת הגיל החציוני במדינה.

**הקשר בין הטמפרטורה הממוצעת למדד הפיתוח של המדינה:**

  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
הגרף מציג את הקשר בין הטמפרטורה הממוצעת למדד הפיתוח המדיני .(State Development Index) הקשר בין המשתנים הוא שלילי עם קורלציה של -0.52, כלומר, ככל שהטמפרטורה הממוצעת עולה, מדד הפיתוח המדיני נוטה להיות נמוך יותר. אנחנו משערות שייתכן שבמדינות עם תנאים אקלימים קשים ישנם אתגרים רבים בתחומים כמו חקלאות, אשר עלולים להאט את הפיתוח הכלכלי והחברתי שלהן. עם זאת לא נוכל לקבוע שמדובר בקשר סיבתי אלא קשר מדגמי. הפיזור הרחב של הנתונים מצביע על כך שהקשר אינו מוחלט, וכי ייתכן שישנם גורמים נוספים המשפיעים על שני המשתנים.

**הקשר בין טמפרטורה לאחוז המעשנים:**



בהתבסס על הגרף ומקדם הקורלציה, ניתן לראות קשר שלילי בין טמפרטורה ממוצעת ל-שיעור העישון. מקדם הקורלציה של -0.413 מצביע על קשר שלילי ביניהם, כלומר, ככל שהטמפרטורה הממוצעת עולה, יש נטייה לשיעור העישון לרדת.

הקשר הזה יכול להיות מוסבר מכך שבאזורים חמים יותר, אנשים נוטים להימנע משהייה בחוץ ולכן מעישון בחוץ במהלך הקיץ החם, או שהתנאים החיצוניים משפיעים על אורח החיים.

לכן אנחנו מניחות שהקשר בין טמפרטורה ממוצעת לשיעור העישון הוא קשר מדגמי, עם קורלציה שלילית, וייתכן שהוא מושפע ממאפיינים כמו הרגלים חברתיים ותרבותיים, ולכן אינו מעיד על קשר סיבתי ישיר.

1. **ניתוח תיאורי של המשתנים**

**משתנה רציף: אחוז מעשנים (Smoke rate)**

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור רחב באחוז המעשנים, כפי שמשתקף מסטיית התקן הגבוהה. התחום הבין-רבעוני מעיד על ריכוז מרבית הערכים בטווח מתון ויציב. קרבת הממוצע לחציון והאסימטריה הקלה ימינה מצביעים על התפלגות כמעט סימטרית עם השפעה מוגבלת של ערכים גבוהים. שילוב זה מדגיש ריכוז מרכזי לצד השפעה מסוימת של ערכים קיצוניים בקצוות.

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 22.41 |
| חציון | 22.5 |
| סטיית תקן | 9.426261 |
| רבעון 25% | 15.6 |
| רבעון 75% | 28.3 |
| תחום בין רבעוני | 12.7 |
| אסימטריה | 0.2146442 |

**משתנה רציף: טמפרטורה ממוצעת במדינה (Average temperature)**

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור נרחב בטמפרטורה הממוצעת, כפי שעולה מסטיית התקן. התחום הבין-רבעוני מצביע על ריכוז הערכים המרכזיים בטווח רחב יחסית, המשקף שונות ניכרת במדגם. קרבת הממוצע לחציון והאסימטריה הקלה לערכים הנמוכים מצביעות על התפלגות כמעט סימטרית עם השפעה קלה של ערכים נמוכים. השילוב בין הטווח הבין-רבעוני לסטיית התקן ממחיש את נוכחותם של ערכים קיצוניים המשפיעים על פיזור הנתונים.

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 16.73 |
| חציון | 18.45 |
| סטיית תקן | 8.698227 |
| רבעון 25% | 8.9 |
| רבעון 75% | 24.95 |
| תחום בין רבעוני | 16.05 |
| אסימטריה | -0.389412 |

**משתנה רציף: מדד התפתחות במדינה State Development))**

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור מתון במדד הפיתוח, כפי שמשתקף מסטיית התקן הנמוכה והתחום הבין-רבעוני הצר, המעידים על אחידות יחסית בערכים המרכזיים. קרבת הממוצע לחציון מצביעה על כך שרוב הערכים מרוכזים במרכז ההתפלגות, בעוד האסימטריה המשמעותית לשלילה מעידה על נטייה לערכים נמוכים שמשפיעים על מבנה ההתפלגות. השילוב בין התחום הבין-רבעוני לסטיית התקן ממחיש ריכוז ברור בערכים המרכזיים לצד השפעה ניכרת של ערכים נמוכים.

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 0.7673 |
| חציון | 0.7810 |
| סטיית תקן | 0.122528 |
| רבעון 25% | 0.704 |
| רבעון 75% | 0.858 |
| תחום בין רבעוני | 0.154 |
| אסימטריה | -0.7567567 |

**משתנה רציף: אחוז האנשים שגרים בעיר (Percentage of residents in cities)**

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 28.99 |
| חציון | 26.1 |
| סטיית תקן | 17.16613 |
| רבעון 25% | 17.2 |
| רבעון 75% | 36.8 |
| תחום בין רבעוני | 19.6 |
| אסימטריה | 1.131652 |

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור רחב יחסית באחוז האנשים שגרים בערים, כפי שמשתקף מסטיית התקן הגבוהה. התחום הבין-רבעוני מעיד על שונות ניכרת גם בערכים המרכזיים במדגם. קרבת הממוצע לחציון מצביעה על כך שרוב הערכים במדגם מרוכזים סביב המרכז, אך האסימטריה הגבוהה לחיוב מעידה על השפעה של ערכים חריגים בקצה העליון של ההתפלגות. במילים אחרות, בעוד שרוב המדינות נמצאות בטווחים קרובים יחסית, קיימות מדינות עם אחוז גבוה במיוחד של תושבים עירוניים, שמושכות את ההתפלגות כלפי מעלה.

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור רחב יחסית באחוז האנשים שגרים בערים, כפי שמשתקף מסטיית התקן הגבוהה. התחום הבין-רבעוני, העומד על, מעיד על שונות ניכרת גם בערכים המרכזיים במדגם. קרבת הממוצע לחציון מצביעה על כך שרוב הערכים במדגם מרוכזים סביב המרכז, אך האסימטריה הגבוהה לחיוב מעידה על השפעה של ערכים חריגים בקצה העליון של ההתפלגות. במילים אחרות, בעוד שרוב המדינות נמצאות בטווחים קרובים יחסית, קיימות מדינות עם אחוז גבוה במיוחד של תושבים עירוניים, שמושכות את ההתפלגות כלפי מעלה.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| נתון סטטיסטי | Extreme pollution | High pollution | Low pollution | Medium pollution | Minor pollution |
| ממוצע | 13.28 | 15.44 | 17.826 | 15.550 | 23.03 |
| חציון | 11.96 | 16.90 | 16.333 | 12.682 | 24.75 |
| סטיית תקן | 4.162299 | 3.236928 | 6.479793 | 6.301803 | 7.029739 |
| רבעון 25% | 10.41 | 14.31 | 12.613 | 10.553 | 17.40 |
| רבעון 75% | 14.83 | 17.29 | 22.014 | 21.739 | 29.09 |
| תחום בין רבעוני | 4.42 | 2.98 | 9.401 | 11.186 | 11.683 |
| אסימטריה | 0.5027055 | -0.3587992 | 0.4518545 | 0.4918286 | -0.2053371 |

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור רחב יחסית באחוז האנשים שגרים בערים, כפי שמשתקף מסטיית התקן הגבוהה. התחום הבין-רבעוני, העומד על, מעיד על שונות ניכרת גם בערכים המרכזיים במדגם. קרבת הממוצע לחציון מצביעה על כך שרוב הערכים במדגם מרוכזים סביב המרכז, אך האסימטריה הגבוהה לחיוב מעידה על השפעה של ערכים חריגים בקצה העליון של ההתפלגות. במילים אחרות, בעוד שרוב המדינות נמצאות בטווחים קרובים יחסית, קיימות מדינות עם אחוז גבוה במיוחד של תושבים עירוניים, שמושכות את ההתפלגות כלפי מעלה.

**פירוש תוצאות הניתוח:** הנתונים מצביעים על פיזור רחב יחסית באחוז האנשים שגרים בערים, כפי שמשתקף מסטיית התקן הגבוהה. התחום הבין-רבעוני, העומד על, מעיד על שונות ניכרת גם בערכים המרכזיים במדגם. קרבת הממוצע לחציון מצביעה על כך שרוב הערכים במדגם מרוכזים סביב המרכז, אך האסימטריה הגבוהה לחיוב מעידה על השפעה של ערכים חריגים בקצה העליון של ההתפלגות. במילים אחרות, בעוד שרוב המדינות נמצאות בטווחים קרובים יחסית, קיימות מדינות עם אחוז גבוה במיוחד של תושבים עירוניים, שמושכות את ההתפלגות כלפי מעלה.

**משתנה רציף: אחוז השמנה (Obesity rate)**

**משתנה רציף: שיעור התמותה מסרטן ((Cancer mortality rate**

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 20.508 |
| חציון | 20.342 |
| סטיית תקן | 7.349432 |
| רבעון 25% | 13.676 |
| רבעון 75% | 27.167 |
| תחום בין רבעוני | 13.491 |
| אסימטריה | 0.1361792 |

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 21.3 |
| חציון | 22.5 |
| סטיית תקן | 9.438476 |
| רבעון 25% | 17.9 |
| רבעון 75% | 26.6 |
| תחום בין רבעוני | 8.7 |
| אסימטריה | -0.1173986 |

**משתנה רציף: גיל ממוצע במדינה (Median age)**

**משתנה רציף: שיעור הדיכאון במדינה ((Depression rate**

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 3.4 |
| חציון | 3.434 |
| סטיית תקן | 0.6739245 |
| רבעון 25% | 2.905 |
| רבעון 75% | 3.849 |
| תחום בין רבעוני | 0.944 |
| אסימטריה | 0.1843397 |

|  |  |
| --- | --- |
| נתון סטטיסטי | תוצאה |
| ממוצע | 34.19 |
| חציון | 33.7 |
| סטיית תקן | 7.794011 |
| רבעון 25% | 28.6 |
| רבעון 75% | 41.4 |
| תחום בין רבעוני | 12.8 |
| אסימטריה | -0.2029641 |

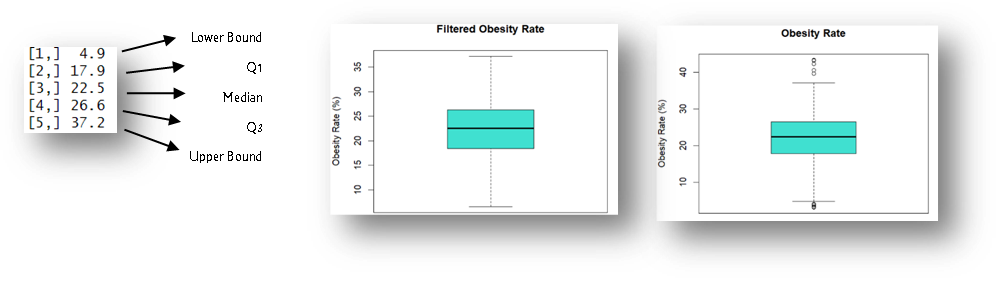
**משתנה קטגוריאלי: מדד זיהום אוויר (Air pollution index)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| נתון סטטיסטי | Extreme pollution | High pollution | Low pollution | Medium pollution | Minor pollution |
| ממוצע | 13.28 | 15.44 | 17.826 | 15.550 | 23.03 |
| חציון | 11.96 | 16.90 | 16.333 | 12.682 | 24.75 |
| סטיית תקן | 4.162299 | 3.236928 | 6.479793 | 6.301803 | 7.029739 |
| רבעון 25% | 10.41 | 14.31 | 12.613 | 10.553 | 17.40 |
| רבעון 75% | 14.83 | 17.29 | 22.014 | 21.739 | 29.09 |
| תחום בין רבעוני | 4.42 | 2.98 | 9.401 | 11.186 | 11.683 |
| אסימטריה | 0.7739657 | -0.6591562 | 0.4762758 | 0.5674176 | -0.2098171 |

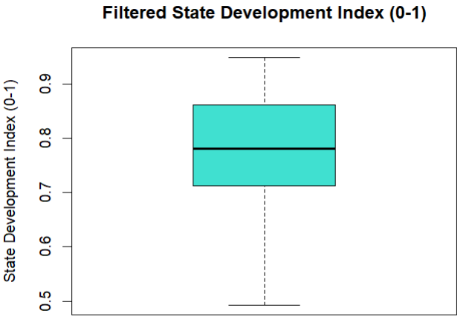
**ניתוח והצלבת נתונים**- ניתוח רמות זיהום האוויר מצביע על מגמות ברורות בקשר לאחוז התמותה מסרטן. בקטגוריות עם זיהום נמוך, הממוצע והחציון של התמותה גבוהים יותר, בעוד שבקטגוריות עם זיהום גבוה, הם נמוכים משמעותית. תחום בין-רבעוני רחב יותר בזיהום נמוך מעיד על שונות גדולה בין המדינות, לעומת אחידות רבה יותר בזיהום גבוה. גם סטיית התקן עוקבת אחר מגמה זו, כאשר בזיהום נמוך היא גבוהה יותר ומצביעה על פיזור רחב של הנתונים, לעומת פיזור צר בזיהום גבוה. מגמות אלו מצביעות על כך שבזיהום נמוך, ייתכן שגורמים נוספים כמו גיל ממוצע גבוה או מדד פיתוח משפיעים על שיעורי התמותה מפאת חוסר אחידות בנתונים, בעוד שבזיהום גבוה, ישנה אחידות רבה יותר בנוגע להשפעה על שיעורי התמותה ומדד זה מהווה גורם מרכזי בקביעת שיעור זה.

1. **ניתוח חריגים:**

**אחוז השמנה (Obesity Rate)**



ניתוח התצפיות החריגות**:** התצפיות החריגות, כפי שניתן לראות בתרשים הראשון (Obesity Rate), הן הערכים מתחת ל-4.9 ומעל 37.2%. מדוע הערכים יצאו חריגים?מתחת ל-4.9%: ייתכן שמדובר במדינות או קבוצות אוכלוסייה קטנות עם שיעורי השמנה נמוכים מאוד, כמו מדינות עם מודעות גבוהה לאורח חיים בריא או תנאי תזונה מיוחדים. ערכים אלו רחוקים מהטווח המרכזי ולכן מוגדרים חריגים.מעל ל-37.2%: מדינות עם שיעורי השמנה גבוהים במיוחד, ייתכן בשל אורח חיים לקוי, גישה תכופה למזון מהיר או בעלי רמה נמוכה למודעות לאורח חיים בריא. ערכים אלו רחוקים משאר התצפיות ולכן מוגדרים חריגים.האם להוריד חריגים?התצפיות החריגות הללו (פחות מ-4.9% ויותר מ-37.2%) כנראה מייצגות קבוצות יוצאות דופן מבחינת דפוסי ההשמנה שלהן. קבוצות אלו הן מועטות במיוחד ביחס לשאר המדינות באוכלוסייה, ולכן לא נרצה שכמה מדינות בודדות ישפיע על התוצאות כולן לכן במקרה זה נבחר להסיר.

**תמונה שמכילה טקסט, צילום מסך, תרשים, קו

התיאור נוצר באופן אוטומטימדד התפתחות במדינה (State Development Index)**

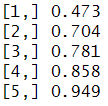
Lower Bound

Q1

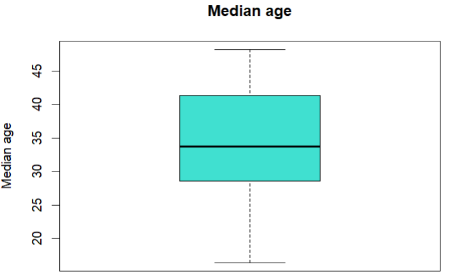
Median

3Q

Upper Bound



במקרה כאן, יצאה תצפית חריגה אחת בלבד – 0.42. מדינה אחת מתוך 117 מדינות הניבה תוצאה חריגה ועל כן, אין בעיה להוריד אותה (גם אם לא היינו מורידים, לא הייתה לה השפעה רבה). המרחק של התצפית מהגבול התחתון הוא 0.053 ובכל מקרה, לא הייתה לתצפית זאת השפעה גדולה על התוצאות.



**אחוז מעשנים (Smoke Rate)**לא נמצאו תצפיות חריגות במשתנה אחוזי המעשנים. טווח הערכים מתפלג עם גבולות תחתונים וגבוהים ברורים (2.0 ו-47.0, בהתאמה), ללא ערכים שמוגדרים כחריגים על פי טווח ה-Boxplot.

תמונה שמכילה טקסט, צילום מסך, תרשים, מלבן

התיאור נוצר באופן אוטומטי

**גיל ממוצע (Median Age)**לא נמצאו תצפיות חריגות במשתנה Median Age. טווח הערכים מתפלג עם גבולות תחתונים וגבוהים ברורים (16.4 ו-48.2, בהתאמה), ללא ערכים שמוגדרים כחריגים על פי טווח ה-BoxPlot. ניתן לראות שקו החציון נמצא מעט למטה ממרכז הקופסא.

**תמונה שמכילה צילום מסך, טקסט, תרשים, מלבן

התיאור נוצר באופן אוטומטי**

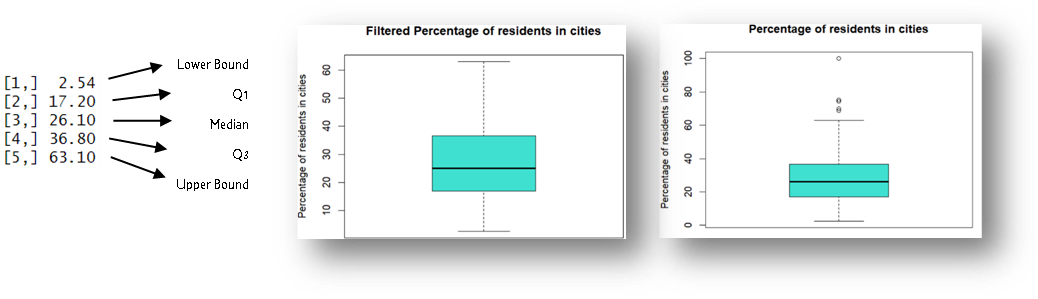
**טמפרטורה ממוצעת (Average Temperature)**לא נמצאו תצפיות חריגות במשתנה Average temperature. טווח הערכים מתפלג עם גבולות תחתונים וגבוהים ברורים (-5.85 ו-28.2, בהתאמה), ללא ערכים שמוגדרים כחריגים על פי טווח ה-Boxplot. ניתן לראות שקו החציון נמצא מעט למעלה ממרכז הקופסא.

תמונה שמכילה טקסט, צילום מסך, תרשים, מלבן

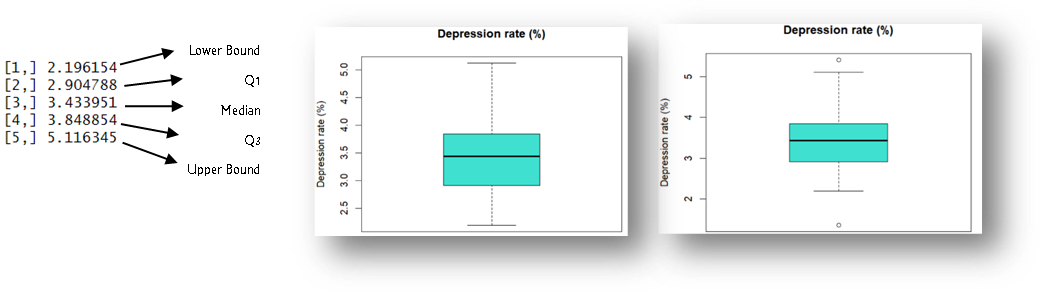
התיאור נוצר באופן אוטומטי

**אחוז התמותה מסרטן (Cancer mortality rate)**לא נמצאו תצפיות חריגות במשתנה אחוז תמותה מסרטן. טווח הערכים מתפלג עם גבולות תחתונים וגבוהים ברורים (9.29 ו-36.14, בהתאמה), ללא ערכים שמוגדרים כחריגים על פי טווח ה-Boxplot.

**אחוז התושבים שגרים בעיר** (**(Percentage of residents in cities**



**ניתוח התצפיות החריגות:** התצפיות החריגות, כפי שניתן לראות בתרשים הראשון (Percentage of residents in cities), הן הערכים מתחת ל-2.54% ומעל 63.10%.   
**מדוע הערכים יצאו חריגים?** מעל ל-63.10%: מדובר במדינות או אזורים שבהם אחוז האוכלוסייה המתגוררת בערים גבוה במיוחד: מדינות עם אורבניזציה גבוהה (שבהן רוב מוחלט של האוכלוסייה מתגורר בערים גדולות). ייתכן ומדובר במדינות קטנות גיאוגרפית שבהן אין כפרים רבים או אזורי מגורים כפריים. ערכים אלה רחוקים מאוד מהטווח הבין-רבעוני (IQR), ולכן מוגדרים כחריגים סטטיסטיים. **האם להוריד חריגים?** התצפיות החריגות הללו כנראה מייצגות קבוצות יוצאות דופן מבחינת רמת העיור במדינה. קבוצות אלו הן מועטות במיוחד (כ-5 מדינות מתוך 117) ביחס לשאר המדינות באוכלוסייה, ולכן לא נרצה שכמה מדינות בודדות ישפיעו על התוצאות כולן.

**אחוז הדיכאון (Depression rate)**

**ניתוח התצפיות החריגות:** יצאו 2 תצפיות חריגות – אחת מעל הגבול העליון ואחת מתחתיו, כפי שניתן לראות בתרשים הראשון (Depression rate (%), הן הערכים מתחת ל-2.19% ומעל 5.12%.   
**מדוע הערכים יצאו חריגים?** מתחת ל-2.19%: מדובר במדינה בה שיעור הדיכאון נמוך במיוחד, ייתכן בשל גורמים כמו אורח חיים בריא, תמיכה חברתית גבוהה, או תנאים תרבותיים ייחודיים המפחיתים לחץ נפשי. ערך זה נדיר במיוחד, ומייצגים מדינה יוצאת דופן בתוך המדגם. מעל ל-5.12%: מדובר במדינה עם שיעורי דיכאון גבוהים במיוחד, ייתכן בשל גורמים כמו עוני, חוסר גישה לטיפול רפואי, או משברים פוליטיים/כלכליים חמורים. ערכים אלו חורגים משמעותית מהטווח הבין-רבעוני, ולכן מוגדרים כחריגים סטטיסטיים.   
**האם להוריד חריגים? כן -** התצפיות החריגות הללו מייצגות קבוצות ייחודיות. קבוצות אלו מהוות 2 מתוך 117 מדינות במדגם, ולכן אפשר להסירן כדי לא לעוות את הממצאים הכלליים, למרות שלא ישפיעו כמעט על התוצאות.

1. **פונקציית צפיפות והתפלגות מצטברת:**

תמונה שמכילה תרשים, טקסט, עלילה, קו

התיאור נוצר באופן אוטומטי**עבור Obesity Rate (%)**גרף 1: פונקציית צפיפות ההסתברות (היסטוגרמה) ניתן לראות **שיא** משמעותי בסביבות הערכים 20%-25%, דבר המעיד על כך שחלק גדול מהמדינות במדגם מציגות שיעור השמנת יתר בטווח זה. יש שיא נוסף קטן יותר בסביבות 10%, שמעיד על קבוצה קטנה של מדינות עם שיעורי השמנת יתר נמוכים יותר. מבחינת **הזנבות**, הזנב הימני ארוך יותר, מעיד על כך שישנן מעט מדינות עם שיעור השמנת יתר גבוה במיוחד (מעל 40%) . הזנב השמאלי מציין את הערכים הנמוכים יותר של שיעור ההשמנה, שיעורים אלו נמוכים מאוד (מתחת ל-10%) והם פחות נפוצים, אך ישנם כמה מדינות בהם חל שיעור זה. מבחינת **פיזור** הנתונים הם מתפרסים על טווח רחב (מ-0% עד מעל 40%), עם ריכוז גבוה בטווח האמצעי של 10%-30%. **קו הצפיפות** מחזק את ההבחנה שיש שיא עיקרי בטווח 20%-25% ושיא קטן יותר בטווח 10%-15%. ההתפלגות מזכירה התפלגות של התפלגות נורמלית אסימטרית (דבר המאשש את מדד האסימטריה בסעיפים קודמים), אך התפלגות זו איננה נורמלית באופן מושלם

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטיגרף 2: פונקציית התפלגות מצטברת   
גרף זה בעל **שיפוע** תלול, המציג עלייה תלולה בסביבות 10%-30%, דבר המעיד על כך שרוב המדינות (כ-70%) נמצאות בטווח זה של שיעור השמנת היתר. מעל 30%, העלייה נהיית מתונה יותר, מה שמעיד על כך שיש פחות מדינות עם שיעורי השמנת יתר גבוהים במיוחד. מבחינת **פיזור** האחוזים ניתן להסיק מכאן ש-50% מהמדינות (נקודת **החציון**) נמצאות מתחת ל-25% שיעור השמנת יתר, ואילו 75% מהמדינות נמצאות מתחת ל-35%.

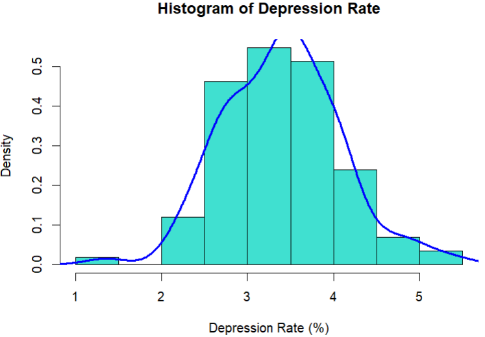
**לסיכום,** רוב המדינות מתרכזות בטווח 10%-30% של שיעור השמנת היתר. ישנן מספר מדינות עם שיעורי השמנת יתר גבוהים מאוד (מעל 40%), כפי שניתן לראות מהזנב הימני הארוך. פונקציית ההתפלגות המצטברת מראה שפיזור הנתונים אינו אחיד, עם ריכוז גבוה בטווח האמצעי.

תמונה שמכילה תרשים, עלילה, קו, טקסט

התיאור נוצר באופן אוטומטי**עבור Smoke Rate (%)**   
גרף 1: פונקציית צפיפות ההסתברות (היסטוגרמה)   
ניתן לראות **שיא** משמעותי בסביבות הערכים 20%-25%, המעיד על כך שרוב המדינות במדגם מציגות שיעור מעשנים בטווח זה. יש שיא קטן יותר בטווח של 10%-15%, המעיד על קבוצת מדינות עם שיעור מעשנים נמוך יחסית. הזנב הימני ארוך יותר, מה שמעיד על כך שישנן מדינות עם שיעור מעשנים גבוה במיוחד (עד 50%). ההתפלגות נראית כמו התפלגות נורמלית אסימטרית, עם נטייה קלה ימינה בשל **הזנב** הימני הארוך, דבר המאשש את נתון אסימטריה בסעיפים הקודמים. הדבר מעיד על קבוצה קטנה יחסית של מדינות עם שיעורי עישון חריגים גבוהים. הזנב השמאלי מעיד על כך ששיעורי עישון נמוכים מאוד (0%-10%) נדירים יחסית, אך קיימים .**פיזור** הנתונים מצביע על ריכוז גבוה בטווח 10%-30%, עם **קו צפיפות** המחזק את ההבחנה ששיא עיקרי נמצא בטווח 20%-25%, ושיא נוסף קטן בטווח 10%-15%.

תמונה שמכילה טקסט, צילום מסך, עלילה, תרשים

התיאור נוצר באופן אוטומטיגרף 2: פונקציית התפלגות מצטברת:   
גרף זה מציג **שיפוע** תלול בסביבות 10%-30%, המצביע על כך שרוב המדינות (בקירוב 70%) נמצאות בטווח זה של שיעור המעשנים. מעל 30%, העלייה נהיית מתונה יותר, מה שמדגיש שמספר המדינות עם שיעורי מעשנים גבוהים במיוחד קטן. מהגרף ניתן לראות ש-50% מהמדינות (נקודת **החציון**) נמצאות מתחת ל-25% שיעור מעשנים וש-75% מהמדינות נמצאות מתחת ל-35%. ניתן לראות שהגרף מתקרב להסתברות של הערך 1 בערכים סביב 40%-50%. דבר המראה שכל שיעורי העישון באוכלוסייה מרוכזים עד טווח זה. **לסיכום,** רוב המדינות מתרכזות בטווח 10%-30% של שיעור המעשנים. ישנן מדינות עם שיעור מעשנים גבוה מאוד (עד 50%), כפי שניתן לראות מהזנב הימני הארוך. פונקציית ההתפלגות המצטברת מציינת שפיזור הנתונים אינו אחיד, עם ריכוז משמעותי בטווח האמצעי.

**עבור Depression Rate (%)**   
גרף 1: פונקציית צפיפות ההסתברות (היסטוגרמה)

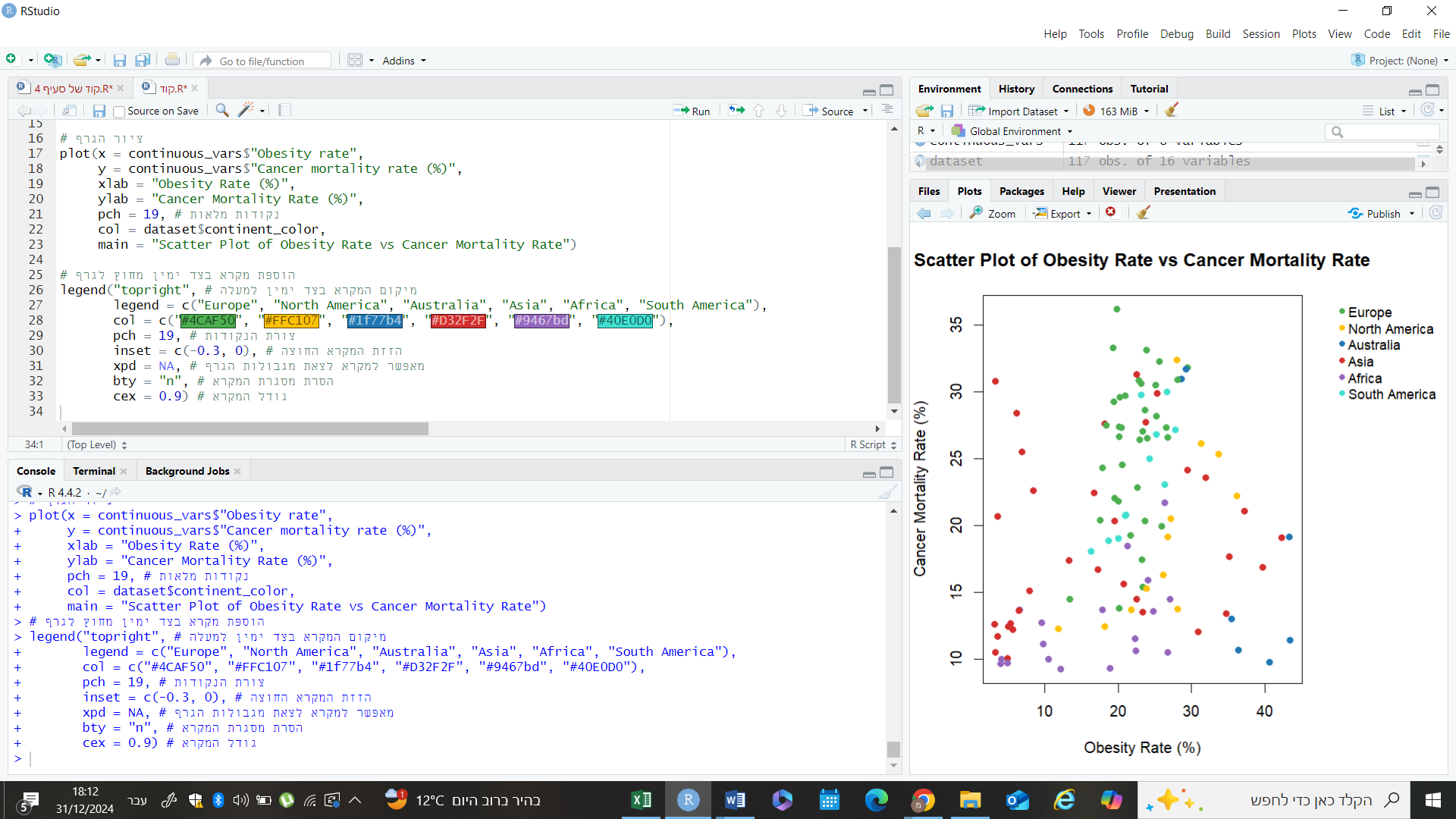
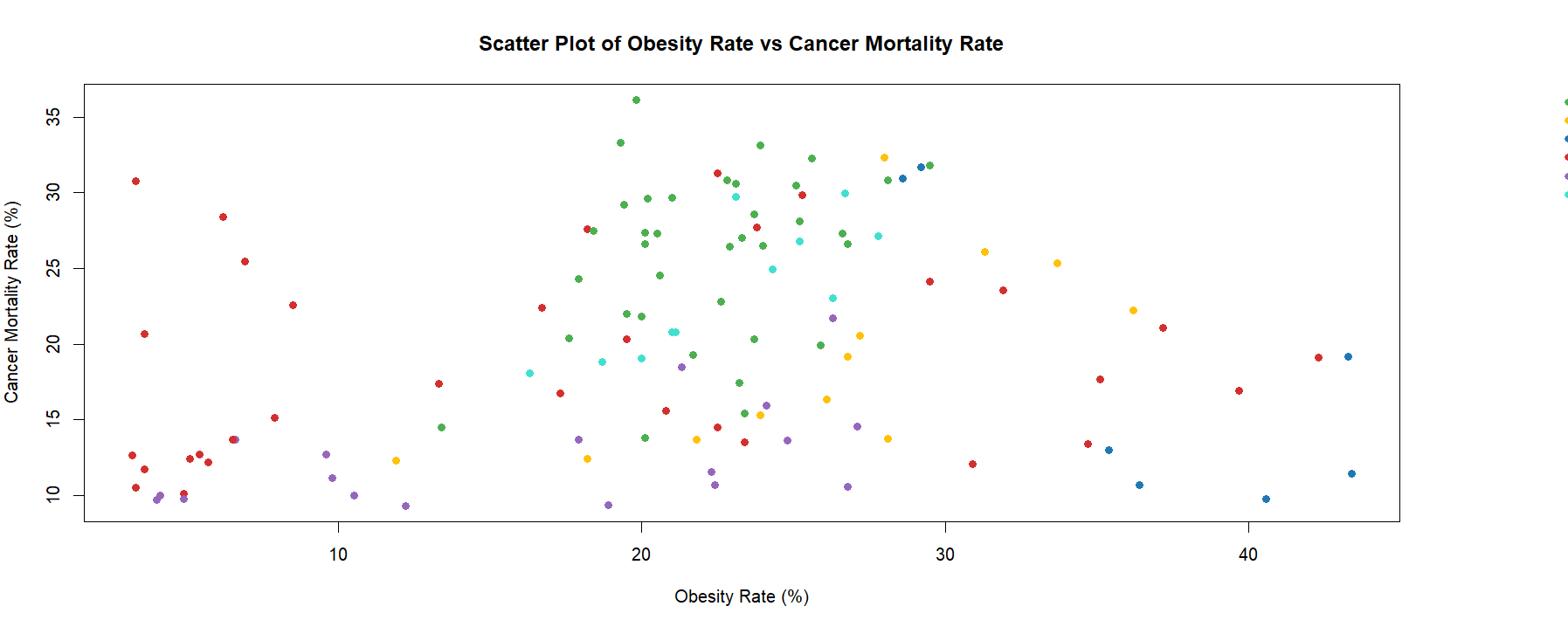
ניתן לראות **שיא** משמעותי בטווח הערכים 3%-4%, דבר המעיד על כך שרוב המדינות במדגם מציגות שיעור דיכאון בטווח זה. קיימת כמעט התפלגות נורמלית, עם מעט מדינות בעלות שיעור דיכאון נמוך מ-2% ומעט מדינות בעלות שיעור גבוה מ-5%. מבחינת **הזנבות**, ההתפלגות סימטרית יחסית ודי צרה, עם פיזור דומה לשני הכיוונים. במבט על הזנב ימני שיעורי דיכאון גבוהים מאוד (5% ומעלה) הם נדירים יחסית, וכן גם לגבי הזנב השמאלי בו שיעורי דיכאון נמוכים מאוד (1%-2%). **הירידה חדה בזנבות** מראה שהשכיחות של שיעורי דיכאון קיצוניים (נמוכים מאוד או גבוהים מאוד) היא קטנה.

תמונה שמכילה טקסט, צילום מסך, תרשים, עלילה

התיאור נוצר באופן אוטומטיגרף 2: פונקציית התפלגות מצטברת

גרף זה מציג **עלייה** תלולה בין 2%-4%, דבר המעיד על כך שכ-80% מהמדינות נמצאות בטווח זה של שיעור הדיכאון. לאחר 4%, העלייה מתונה יותר, דבר המצביע על כך שמספר המדינות עם שיעורי דיכאון גבוהים במיוחד הוא קטן. בנוסף, ניתן לראות כי **החציון** נמצא בטווח 3.5%, כלומר מחצית מהמדינות מציגות שיעור דיכאון מתחת לערך זה. מהגרף אפשר להסיק כי 75% מהמדינות נמצאות מתחת ל-4.5% שיעור דיכאון. **לסיכום,** רוב המדינות מתרכזות בטווח 2%-4% של שיעור הדיכאון, כאשר ישנן מעט מדינות עם שיעורי דיכאון גבוהים מאוד (מעל 5%) או נמוכים מאוד (מתחת ל-2%). פונקציית ההתפלגות המצטברת מדגישה את הפיזור האחיד יחסית של הנתונים, עם ריכוז גבוה בטווח האמצעי.

1. **קשר בין שיעור ההשמנה לבין גיל חציוני במדינות שונות:**

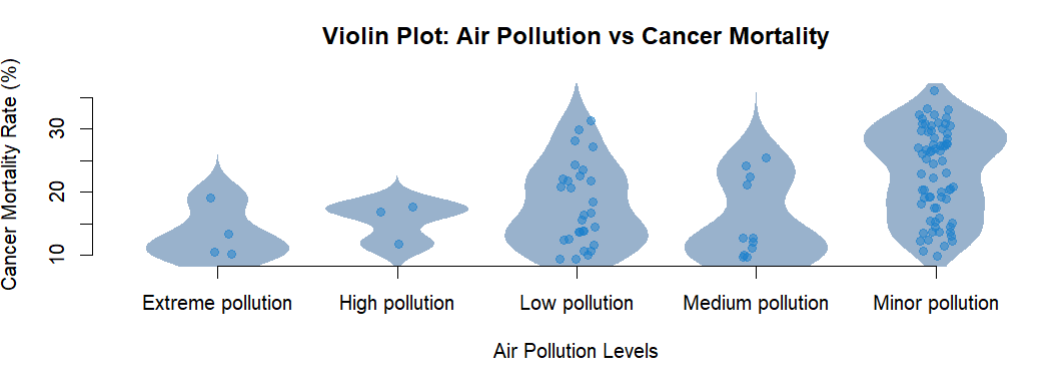


הגרף מציג קשר בין שיעור ההשמנה (Obesity rate) לבין שיעור התמותה מסרטן (Median age) במדינות שונות, כאשר הצבעים מאפשרים לפלח את הנקודות לפי יבשות. כל נקודה מייצגת מדינה, וכל צבע מייצג יבשת. הצבעים הם: ירוק לאירופה, צהוב לצפון אמריקה, אדום לאסיה, סגול לאפריקה, כחול לאוסטרליה וטורקיז לדרום אמריקה. כפי שנראה, אין קשר חזק וברור בין שני המשתנים, כאשר הנתונים מפוזרים בצורה יחסית אחידה על פני הצירים השונים.

עם זאת, כשמתבוננים על כל יבשת בנפרד, ניתן להבחין בממצאים מעניינים ובמגמות שונות. לדוגמא, בצפון אמריקה ודרום אמריקה ניתן לראות קשר חיובי בין שיעור ההשמנה לתמותה מסרטן. באירופה ניתן לראות שהנקודות מרוכזות באזור מסוים בגרף, ועבור אותם ערכי שיעור השמנה, ישנו פיזור רחב של שיעור התמותה מסרטן, כלומר אפשר להסיק שככל הנראה קיימים שם גורמים נוספים המשפיעים על שיעור התמותה מסרטן. באפריקה ניתן לראות שמגמת העלייה מינורית, מה שעשוי להצביע על כך שייתכן שביבשת זו ההשמנה היא לא גורם משפיע על תמותה מסרטן, או שייתכן שאין מספיק נתונים בנושא כיוון שבמדינות אפריקה הפיתוח הרפואי נמוך יותר ולכן גם אבחוני הסרטן.

לסיכום, התרשים הדגים באופן ויזואלי את הקשר בין שיעור ההשמנה לשיעור התמותה מסרטן, תוך פילוח הנתונים לפי יבשות. התרשים מסייע לנו לזהות מגמות ותופעות התלויות ביבשת שבה המדינה נמצאת.

1. **ייצוג קשרים בעזרת תרשימים:**
2. **גרף כינור משולב – הקשר בין רמות זיהום האוויר לשיעור תמותה מסרטן**

****

פענוח תוצאות הגרף: הגרף הוא Violin Plot שמציג את הקשר בין רמות זיהום האוויר (Air Pollution Levels) לבין שיעורי התמותה מסרטן (Cancer Mortality Rate (%).

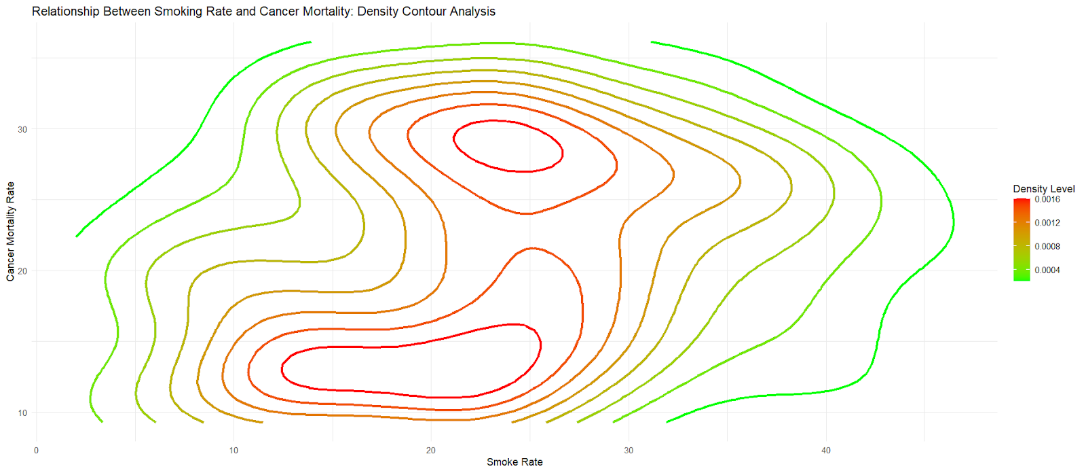
* חלקים רחבים יותר => יש יותר מדינות עם שיעור תמותה באותו תחום.
* חלקים צרים יותר => יש פחות מדינות עם שיעור תמותה באותו תחום.
* הנקודות: התצפיות בפועל לכל קטגוריית זיהום.

מסקנות מהגרף:

* הכינור של Minor pollution: רמת הזיהום הקלה ביותר (Minor pollution) מציגה התפלגות רחבה מאוד בשיעורי התמותה מסרטן. רוב הנתונים נמצאים בטווח 10%-20%, אך יש גם מדינות עם שיעור גבוה יותר.
* הכינור של Medium pollution: רמת זיהום זו מציגה צפיפות גבוהה סביב 15%-20%, עם התפלגות סימטרית יחסית. רוחב הכינור מלמד שזו אחת הקטגוריות עם צפיפות נתונים גבוהה.
* הכינור של Low pollution: רמת הזיהום הנמוכה מראה פיזור רחב יחסית בשיעורי התמותה. חלק מהמדינות נמצאות בשיעור תמותה נמוך מאוד, אך ישנן מדינות עם תמותה גבוהה יחסית (20%-30%).
* הכינור של High pollution: ברמת זיהום גבוהה ישנה פיזור קטן יחסית, עם צפיפות גבוהה בשיעורי תמותה של כ-15%-20%. הצורה הצרה מלמדת על מספר קטן של מדינות בקבוצה זו.
* הכינור של Extreme pollution: רמת זיהום קיצונית כוללת מדינות עם פיזור נמוך מאוד בשיעורי התמותה. הנתונים מראים ריכוז מסוים בשיעורי תמותה נמוכים יחסית (10%-15%).

לסיכום: זיהום אוויר נמוך (Minor/Medium pollution): ההתפלגות רחבה מאוד, מה שמעיד על שונות גבוהה בשיעורי התמותה במדינות עם רמות זיהום אלו. זיהום אוויר גבוה (High/Extreme pollution): ההתפלגות צרה יחסית, עם ריכוזים ברורים בשיעורי תמותה ממוצעים-נמוכים. מדינות עם זיהום אוויר בינוני עד קל מציגות פיזור רחב הרבה יותר, דבר שיכול להעיד על גורמים נוספים שמשפיעים על שיעורי התמותה מלבד זיהום האוויר.

1. **תרשים צפיפות- הקשר בין שיעור העישון לשיעור התמותה מסרטן**



הגרף הוא תרשים צפיפות , (Density Contour Plot) שמציג קשר בין שיעור העישון (Smoke Rate) לשיעור התמותה מסרטן (Cancer Mortality Rate). **הקווים** בגרף מייצגים גבולות בין רמות צפיפות שונות של הנתונים. כל קו מגדיר רמת צפיפות מסוימת, כאשר המרחק בין הקווים משקף את השינוי ברמת הצפיפות. כאשר הקווים קרובים זה לזה, המשמעות היא שיש שינוי חד בצפיפות באותו אזור. **הצבעים** ממחישים את רמת הצפיפות בתוך האזורים התחומים על ידי הקווים. הצבעים נעים מירוק לצהוב, כתום ועד אדום. צבע אדום מסמן את האזורים עם צפיפות הנתונים הגבוהה ביותר (ריכוז גבוה של ערכים דומים), ואילו ירוק מסמן אזורים עם צפיפות נמוכה יותר. הקווים עוזרים להבין את המבנה הכללי של הצפיפות, והצבעים מעשירים את המידע וממחישים בצורה ויזואלית איפה נמצאים ריכוזי הנתונים הגבוהים ביותר. במקרה הזה, אזורים אדומים מצביעים על שיעור עישון בינוני (10%-20%) שמלווה בשיעור תמותה מסרטן בינוני-גבוה (10%-30%).

הקווים הצפופים מסמנים שינויים חדים בצפיפות, והם נמצאים בעיקר באזורים שבהם יש התפלגות נתונים ברורה (למשל סביב ריכוזי הצפיפות האדומים). ירוק ואזורי קווים רחוקים מראים מקומות בהם הנתונים מפוזרים יותר, עם פחות חזרה על ערכים דומים.

1. **טבלאות שכיחות:**

**טבלת שכיחות חד ממדית- רמת זיהום אוויר**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Frequency** | | | | | |
| **רמת זיהום האוויר** | Extreme pollution | High pollution | Low pollution | Medium pollution | Minor pollution |
|  | 4 | 3 | 29 | 11 | 70 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Percentage** | | | | | |
| **רמת זיהום האוויר** | Extreme pollution | High pollution | Low pollution | Medium pollution | Minor pollution |
|  | 3.418803 | 2.564103 | 24.786325 | 9.401709 | 59.829060 |

הטבלה מציגה את שכיחות רמות זיהום האוויר במדגם. הרמה השכיחה ביותר היא "Minor pollution", המהווה 59.83% מהמדגם עם 70 רשומות. אחריה "Low pollution", המהווה 24.79% עם 29 רשומות. רמות "Medium pollution", "High pollution"ו-"Extreme pollution" נדירות יותר. הנתונים מצביעים על כך שרוב המדינות במדגם נמצאות ברמות זיהום אוויר נמוכות יחסית, בעוד שמספר קטן של מדינות נמצא ברמות זיהום גבוהות.

**טבלת שכיחות חד ממדית- יבשות**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Frequency** | | | | | | |
| **יבשת** | Europe | North America | Australia | Asia | South America | Africa |
|  | 36 | 12 | 7 | 33 | 11 | 18 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Percentage** | | | | | | |
| **יבשת** | Europe | North America | Australia | Asia | South America | Africa |
|  | 30.769231 | 10.256410 | 5.982906 | 28.205128 | 9.401709 | 15.384615 |

הטבלה מציגה את התפלגות המדינות במדגם לפי יבשות, ומראה כי היבשת עם מספר המדינות הגדול ביותר היא אירופה, המונה 36 מדינות ומהווה 30.78% מהמדגם. לאחריה נמצאת אסיה עם 33 מדינות (28.2%), ואפריקה במקום השלישי עם 18 מדינות (15.38%). צפון אמריקה כוללת 12 מדינות (10.26%), ודרום אמריקה כוללת 11 מדינות (9.4%). אוסטרליה היא היבשת עם המספר הקטן ביותר של מדינות במדגם, עם 7 מדינות בלבד (5.98%). הנתונים מצביעים על כך שהמדגם מתמקד בעיקר במדינות מאירופה ואסיה, בעוד שתרומתן של היבשות האחרות, ובעיקר אוסטרליה, קטנה יותר.

**טבלת שכיחות דו מימדית- רמת זיהום ליבשת**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Percentage** | | | | | |
|  | **Extreme pollution** | **High pollution** | **Low pollution** | **Medium pollution** | **Minor pollution** |
| **Africa** | 0 | 0 | 8.547009 | 4.273504 | 2.564103 |
| **Asia** | 3.418803 | 2.564103 | 6.837607 | 5.128205 | 10.256410 |
| **Australia** | 0 | 0 | 0 | 0 | 5.982906 |
| **Europe** | 0 | 0 | 4.273504 | 0 | 26.495726 |
| **North America** | 0 | 0 | 3.418803 | 0 | 6.837607 |
| **South America** | 0 | 0 | 1.709402 | 0 | 7.692308 |

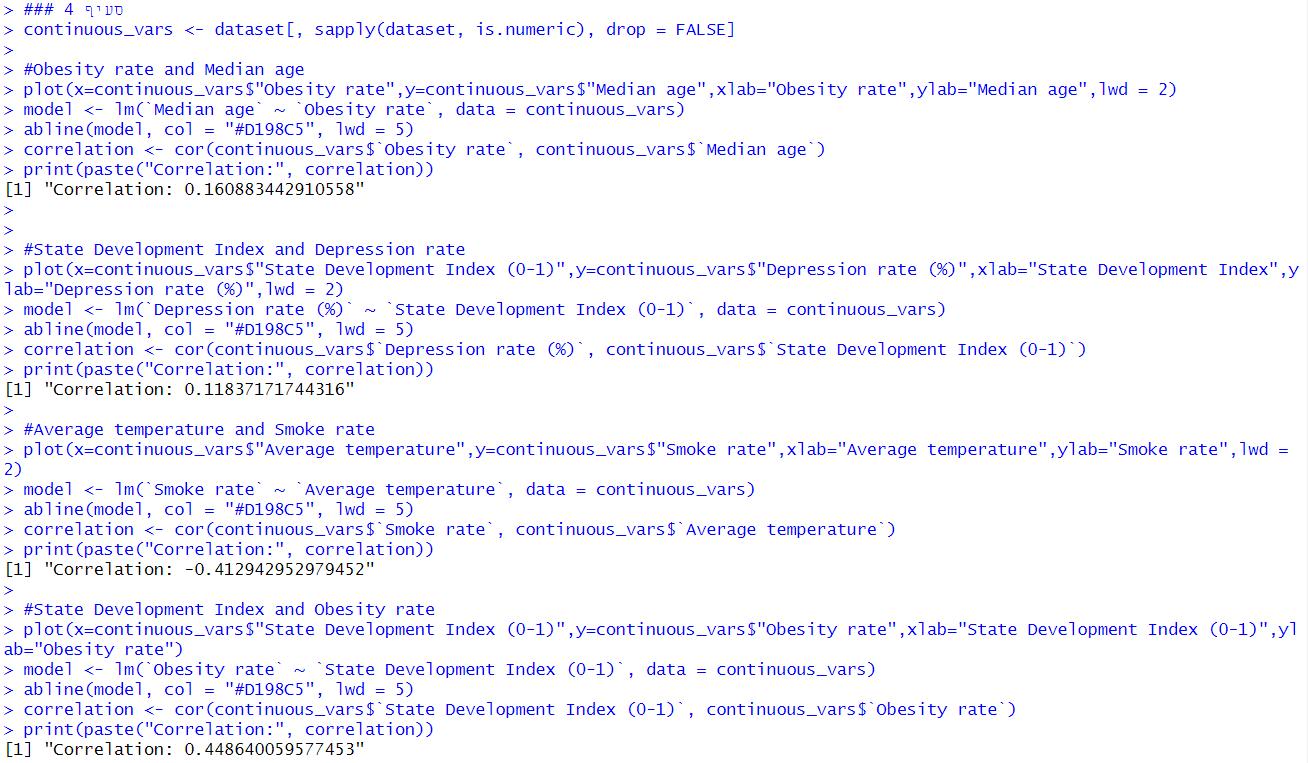
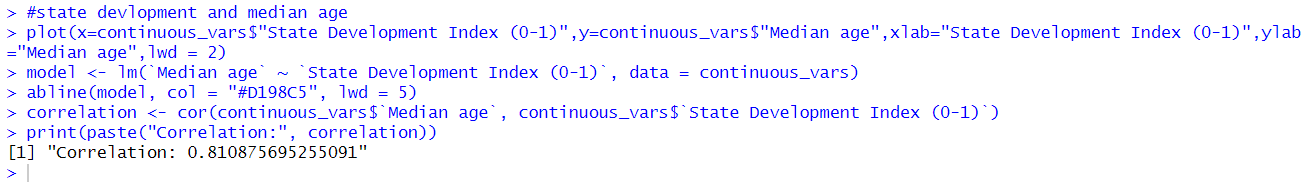
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Frequency** | | | | | |
|  | **Extreme pollution** | **High pollution** | **Low pollution** | **Medium pollution** | **Minor pollution** |
| **Africa** | 0 | 0 | 10 | 5 | 3 |
| **Asia** | 4 | 3 | 8 | 6 | 12 |
| **Australia** | 0 | 0 | 0 | 0 | 7 |
| **Europe** | 0 | 0 | 5 | 0 | 31 |
| **North America** | 0 | 0 | 4 | 0 | 8 |
| **South America** | 0 | 0 | 2 | 0 | 9 |

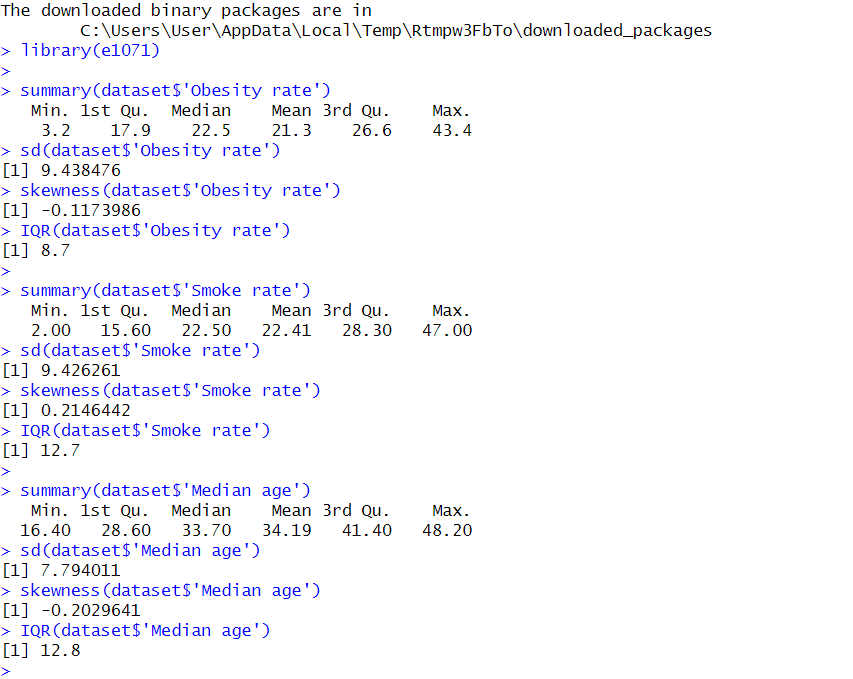
הטבלה מציגה את התפלגות רמות זיהום האוויר לפי יבשות, ומצביעה על שונות ניכרת בין האזורים. רמת זיהום נמוכה מאוד ("Minor pollution") נפוצה במיוחד באירופה, שם 31 מדינות (26.5%) משתייכות לקטגוריה זו, ואחריה אסיה עם 12 מדינות (10.26%). לעומת זאת, זיהום קיצוני ("Extreme pollution") קיים רק באסיה, עם 4 מדינות (3.42%), בעוד שביבשות אחרות לא נרשמו מקרים כאלה. רמת זיהום בינוני נפוצה בעיקר באפריקה ובאסיה, אך אינה קיימת באירופה או אוסטרליה.

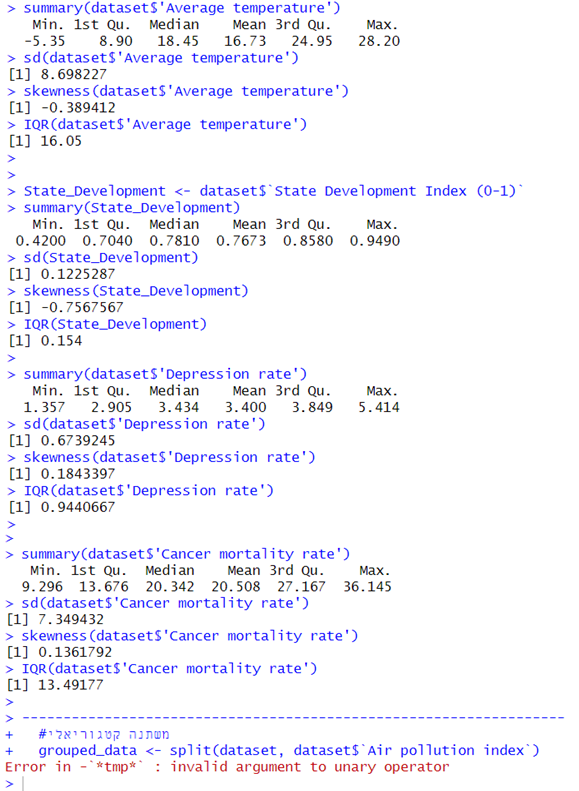
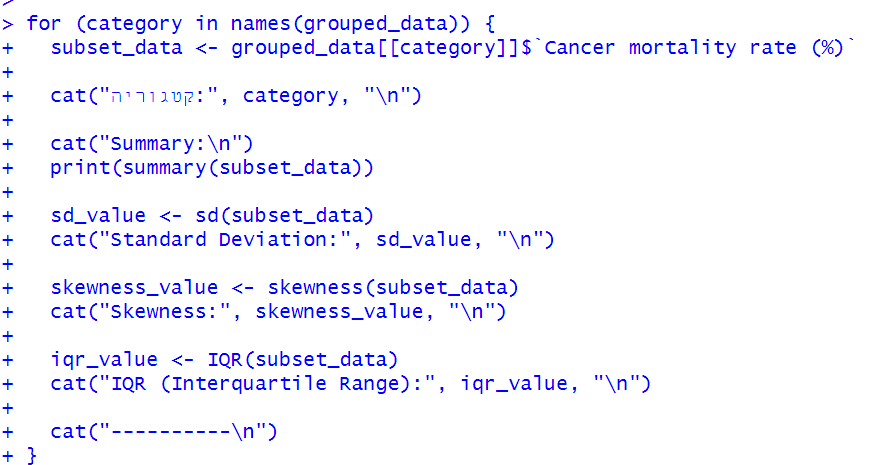
הנתונים מצביעים על כך שאירופה שומרת על זיהום אוויר נמוך יחסית, בעוד שאסיה כוללת מדינות גם ברמות זיהום נמוכות וגם בזיהום קיצוני, מה שעשוי להעיד על פערים סביבתיים גדולים בין מדינות היבשת. אפריקה בולטת ברמות זיהום נמוכות עד בינוניות, בעוד שאוסטרליה כמעט ולא מציגה גיוון ברמות הזיהום, עם התמקדות בזיהום נמוך בלבד. התפלגות זו משקפת את ההבדלים בתשתיות, מדיניות סביבתית ומאפיינים גאוגרפיים בין היבשות.

נספחים:

**נספח 4**



  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
**נספח 5**



תמונה שמכילה טקסט, צילום מסך, גופן, מסמך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטי

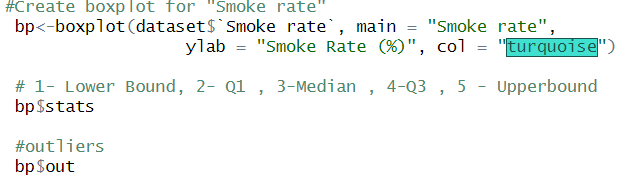
**נספח 6**

**Obesity Rate 6.1**

תמונה שמכילה טקסט, צילום מסך, גופן, אלגברה

התיאור נוצר באופן אוטומטי  
  
  
  
  
  
  
  
  
  
  
פלט**תמונה שמכילה טקסט, גופן, צילום מסך

התיאור נוצר באופן אוטומטי:**

**Smoke Rate 6.2**

תמונה שמכילה טקסט, גופן, צילום מסך, עיצוב

התיאור נוצר באופן אוטומטי  
  
  
  
  
  
פלט:

**תמונה שמכילה טקסט, גופן, צילום מסך, מספר

התיאור נוצר באופן אוטומטיתמונה שמכילה טקסט, צילום מסך, גופן

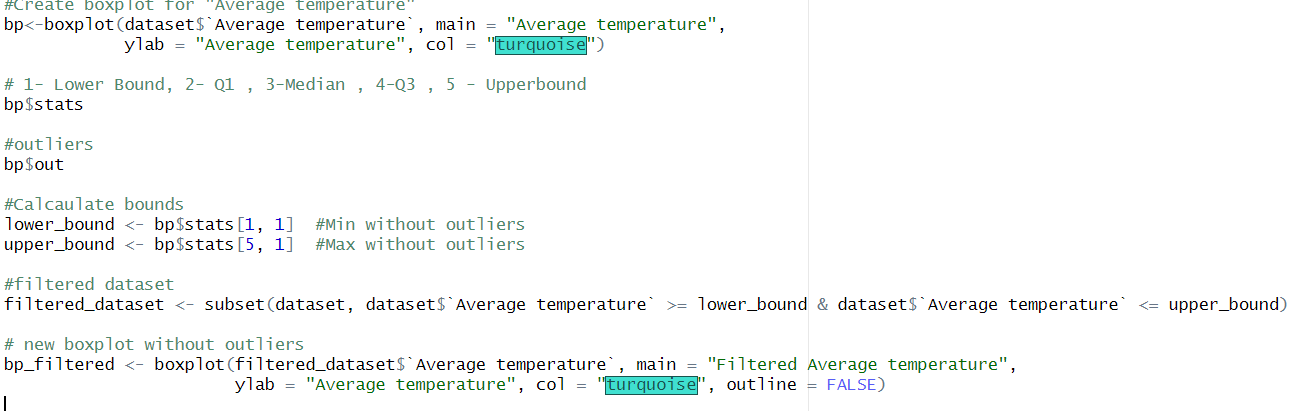
התיאור נוצר באופן אוטומטי**  
  
  
  
  
**State Development Index 6.3 (0-1)**  
  
  
הפלט:

תמונה שמכילה טקסט, צילום מסך, גופן, קו

התיאור נוצר באופן אוטומטי  
  
  
  
**Median Age 6.4**

תמונה שמכילה טקסט, גופן, צילום מסך, מספר

התיאור נוצר באופן אוטומטי  
  
  
  
  
הפלט:

**Average Temperature 6.5**

תמונה שמכילה טקסט, גופן, צילום מסך, מספר

התיאור נוצר באופן אוטומטיהפלט:

**Cancer mortality rate (%) 6.6**

תמונה שמכילה טקסט, גופן, צילום מסך, קו

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, גופן, צילום מסך, מספר

התיאור נוצר באופן אוטומטיהפלט

**Percentage of residents in cities 6.7**

תמונה שמכילה טקסט, צילום מסך, גופן

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטיהפלט

**Depression rate (%) 6.8**

תמונה שמכילה טקסט, גופן, צילום מסך

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, גופן, צילום מסך, עיצוב

התיאור נוצר באופן אוטומטי

הפלט

תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטיסעיף 10:

**תמונה שמכילה טקסט, צילום מסך, גופן, מספר

התיאור נוצר באופן אוטומטי**