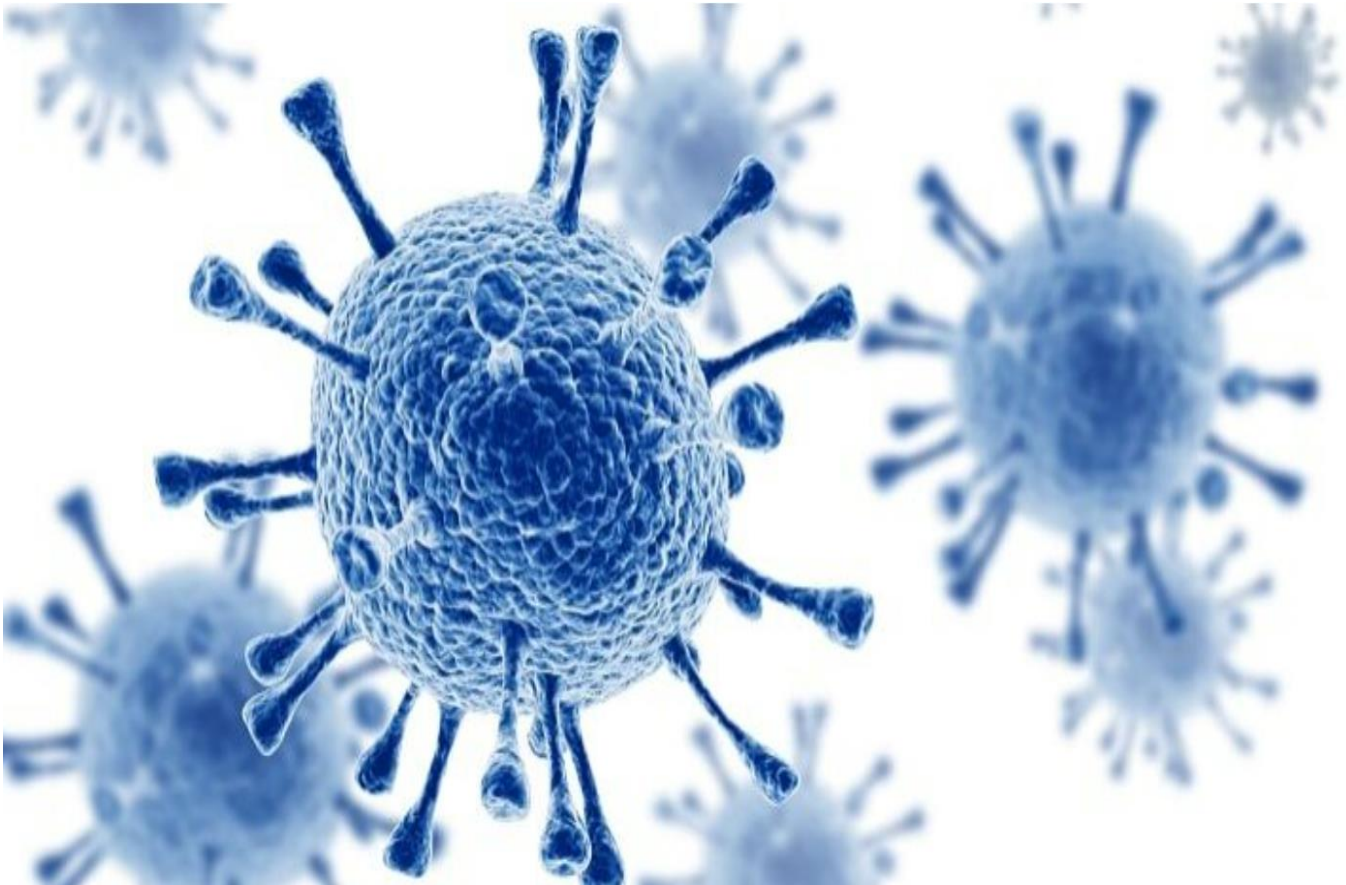


## **פרויקט בוויזואליזציה של מידע (382.1.3203)**

**אילו גורמים משפיעים על התקדמות ה-HIV בחולים העוברים טיפולי  
ART?**



**שמות המגישות + ת"ז:**

סתיו ברק 206477952

נעמה מימון 207547183

**תאריך ההגשה:**

21/07/2024

## חלק 2:

### 1. מבוא:

המוקד העיקרי של פרויקט הדמיה זה הוא לחקור את הגורמים המשפיעים על התקדמות האיידס בחולים העוברים פרוטוקולים שונים של טיפול אנטי-רטרו-ויראלי (ART). הבנת הגורמים הללו חיונית לשיפור תוצאות הטיפול ולפיתוח משטרי ART מותאמים שיכולים לנהל ולהאט ביעילות את התקדמות האיידס בחולים.

### **השאלה המרכזית שפרויקט זה שואף לענות היא:**

### **אילו גורמים משפיעים על התקדמות האיידס בחולים העוברים פרוטוקולים שונים של ART?**

כדי להתייחס לשאלה המרכזית, נבחן מספר שאלות משנה:

1. כיצד משתנה הכיסוי של טיפול אנטי-רטרו-ויראלי (ART) ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות?
2. כיצד משתנים המדדים הקליניים (CD4 ו-CD8) בתגובה לפרוטוקולים השונים של טיפול אנטי-רטרו-ויראלי (ART) בנקודת התחלה ולאחר 20 שבועות?
3. מה הקשר בין פרוטוקולי ART שונים ושיעור התמותה בקרב חולי איידס לאורך זמן?
4. אילו נתונים דמוגרפיים של מטופלים (מין, גזע) משפיעים על ספירת ה-CD4 וה-CD8 בקרב חולי איידס מטופלים בפרוטוקולי ART שונים.
5. כיצד משתנים שיעורי ההדבקה בנגיף HIV בהתאם לפקטורים שונים בפרוטוקולים שונים של טיפול אנטי-רטרו-ויראלי (ART)? (Drugs, Hemo, Homo).

מערך הנתונים המשמש לפרויקט זה מקורו בשני מקורות מ-Kaggle:

#### 1. AIDS Virus Infection Prediction:

מערך נתונים זה כולל מידע מפורט על מטופלים כגון: גיל, מין, ספירת CD4 (תאים של מערכת החיסון המושפעים מהנגיף), עומסים ויראליים (כמות הנגיף בדם), מצבים נלווים (כגון מחלות נוספות), פרטי פרוטוקולי ART. הנתונים האלו מאפשרים הבנה מעמיקה של השפעת פרוטוקולי ART שונים על התקדמות האיידס במטופלים, כולל מעקב אחר התקדמות המחלה לאורך זמן.

#### 2. HIV AIDS Dataset:

מערך נתונים זה מספק מידע נוסף כגון: מספר האנשים החיים עם HIV, מספר האנשים המדווחים שמקבלים טיפול ART, אחוז הכיסוי של ART בקרב האנשים החיים עם HIV, מספר האנשים החיים עם HIV/AIDS באומדן (כולל מינימום ומקסימום), אזור ארגון הבריאות העולמי (WHO Region)

שני מערכי הנתונים האלו משולבים בפרויקט כדי לספק תמונה כוללת ומעמיקה של מצב הטיפול ב-HIV/AIDS והשפעותיו השונות על האוכלוסייה.

## 2. נתונים:

המשתנים שהשתמשנו בהם לטובת הפרויקט ממערך נתונים זה:

משתנים בינאריים: infected, gender, race, trt, hemo, homo, drugs,

משתנים כמותיים: age, time, cd40, cd420, cd80, cd820

קרדינליות הנתונים והמפתחות:

מספר רשומות (מטופלים): 2139

מפתח ראשי: אין מפתח ראשי מפורש, אבל כל רשומה מייצגת מטופל ייחודי.

Data types, Attributes

תכונות	סוג הדאטה	תיאור
Time	מספרי	זמן לכישלון או לצנזור
Trt	קטגורי	אינדיקטור לטיפול (0 = ZDV only; 1 = ZDV + ddl, 2 = ZDV + Zal, 3 = ddl only)
Age	מספרי	גיל בשנים
Hemo	בינארי	מצב המופיליה(מחלה תורשתית המתבטאת בתפקוד לקוי של מנגנון קרישת הדם) (0=no, 1=yes)
Homo	בינארי	פעילות הומוסקסואלית (0=no, 1=yes)
Drugs	בינארי	היסטוריה של שימוש בתרופות IV (0=no, 1=yes)
Race	בינארי	גזע (0=White, 1=non-white)
Gender	בינארי	מגדר (0=Female, 1=Male)
cd40	מספרי	ספירת CD4 (סוג של תאי דם לבנים)
cd420	מספרי	ספירת CD4 ב-20 +/- 5 שבועות

cd80	מספרי	ספירת CD8 (סוג אחר של תאי דם לבנים)
cd820	מספרי	ספירת CD8 ב-20 +/- 5 שבועות
Infected	בינארי	נגוע באיידס (0=No, 1=Yes)

כמו כן, מערך הנתונים "HIV AIDS Dataset" מכיל סטטיסטיקות בריאות ומידע קטגורי על חולים עם HIV ואיידס. הוא כולל תכונות שונות הקשורות לדמוגרפיה של המטופל, מספר האנשים החיים עם HIV, מספר האנשים שמקבלים טיפול אנטי-רטרו-ויראלי (ART), ואחוזי הכיסוי של ART באזורים שונים בעולם.

DataSet Types: טבלה רב-משתנית: מערך הנתונים מורכב ממספר תכונות (עמודות) המתארות היבטים שונים של כל מדינה (שורות).

המשתנים שהשתמשו בהם לטובת הפרויקט ממערך נתונים זה:

משתנים קטגוריאליים: Country

משתנים כמותיים: Reported number of people receiving ART

קרדינליות ומפתחות:

מספר רשומות (מטופלים): 170

מפתח ראשי: עמודת country.

Data types, Attributes

תכונות	סוג הדאטה	תיאור
Country	קטגוריאלי	מסמן את שם המדינה
Reported number of people receiving ART	כמותי	מספר האנשים שמקבלים טיפול ART

### 3. מטלות:

1. תחום הרפואה ודמוגרפיה - כיצד משתנה הכיסוי של טיפול אנטי-רטרו-ויראלי (ART) ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות?
2. תחום הרפואה – כיצד משתנים מדדים קליניים בתגובה לפרוטוקולים שונים של טיפול ART בנק התחלה ולאחר 20 שבועות
3. תחום הרפואה - כיצד משפיעים פרוטוקולי ART שונים על שיעור התמותה של חולי האיידס לאורך זמן?
4. תחום הרפואה - אילו נתונים דמוגרפיים של מטופלים משפיעים על ספירת ה-CD4 וה-CD8-בקרוב חולי איידס מטופלים בפרוטוקולי ART שונים.
5. תחום הרפואה (מחלות נלוות, היסטוריה של המטופל) - כיצד משפיעים מצבים נלווים על התקדמות האיידס בחולים עם ART? (המופילה, שימוש בסמים, פעילות הומוסקסואלית)

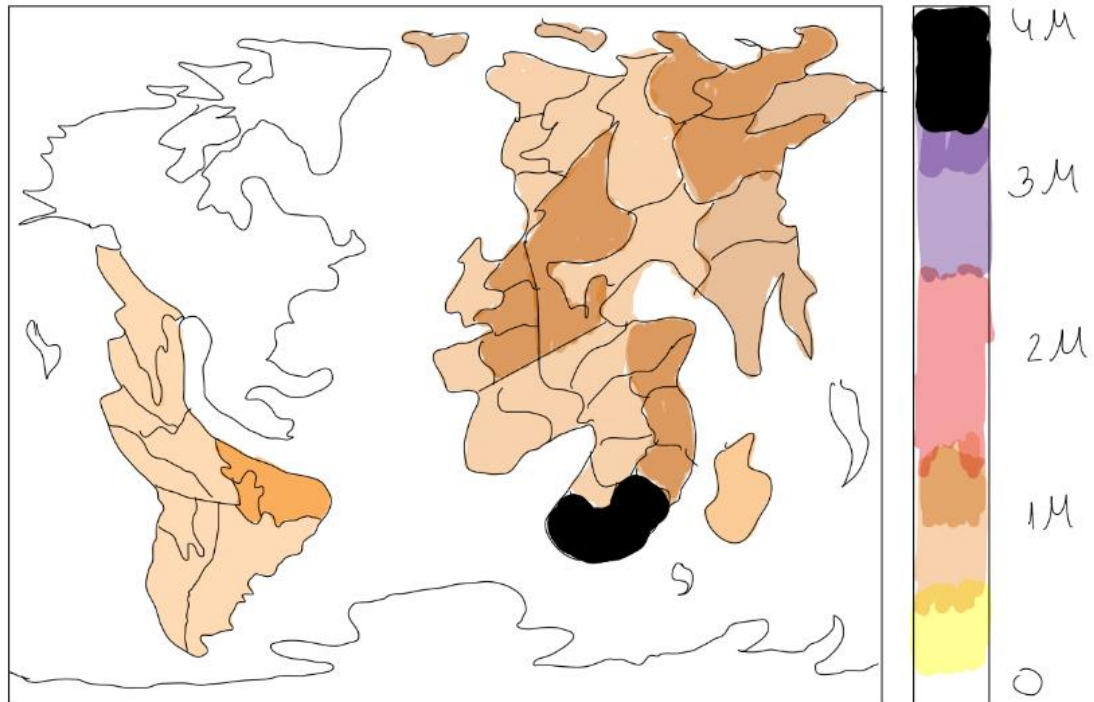
### 3. מיפוי למונחים מופשטים לפי הטיפולוגיה של Munzner:

מספר שאלה	Action	Target
1 - לחקור את הכיסוי של טיפול ART ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות.	identity	התפשטות גיאוגרפית של היסטוריית הטיפולים על התקדמות המחלה
2 -לנתח את השינויים במדדים הקליניים CD4 ו-CD8- בתגובה לפרוטוקולי ART שונים בבסיס ולאחר 20 שבועות.	Compare	סמנים קליניים כמו ספירת CD4/CD8 בבסיס ולאחר 20 שבועות
3 - להעריך את ההשפעה של פרוטוקולי ART שונים על פני זמן.	Compare	סדרות זמן - פרוטוקולי ART על פני זמן
4 - לנתח נתונים דמוגרפיים של מטופלים והשפעתם על ספירת CD4 ו-CD8-בקרוב חולי איידס מטופלים בפרוטוקולי ART שונים.	Compare	מאפיינים דמוגרפיים של מטופלים כמו מגדר, וגזע והשפעתם על תוצאות
5 – לנתח את שיעור ההדבקה בנגיף HIV בהתאם לפקטורים שונים בפרוטוקולים שונים של טיפול ART.	Discover	מחלות נלוות והשפעתן על התקדמות האיידס

### חלק 3 - רעיונות ועיצובים חלופיים:

שאלת מחקר משנית: כיצד משתנה הכיסוי של טיפול אנטי-רטרו-ויראלי (ART) ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות?

אלטרנטיבה 1 לשאלה משנית זו: Choropleth Map



Channels:

- Color: צבע Channel משמש לייצוג מספר האנשים המקבלים טיפול ART במדינות שונות. סולם הצבעים נע בין צהוב בהיר (ערכים נמוכים) לשחור (ערכים גבוהים). צבע לבן מסמן על כך שהמדינה לא נמצאת בדאטה.
- Spatial Position: המיקום המרחבי משמש לייצוג המיקום הגיאוגרפי של כל מדינה על מפת העולם.

Marks:

- Areas: המדינות במפה מסומנות בצבעים שונים לפי מספר האנשים שמקבלים טיפול ART.

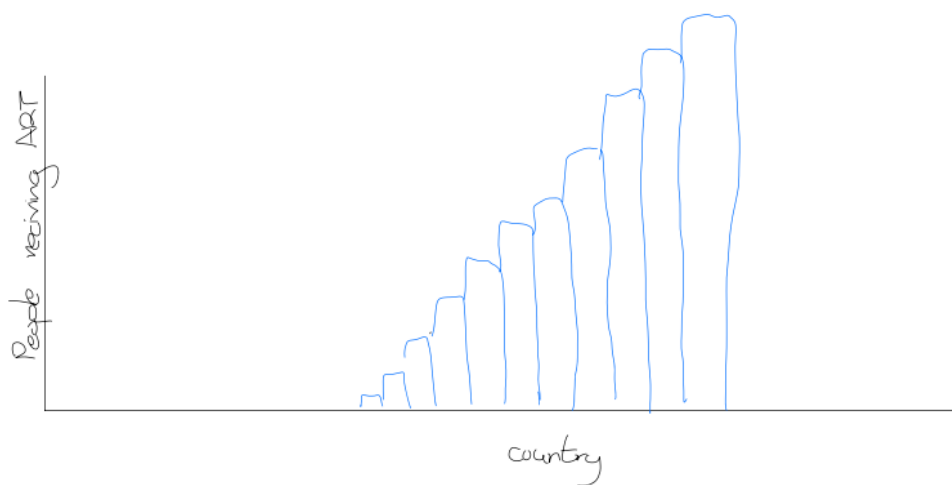
יתרונות:

המפה מספקת דרך אינטואיטיבית להמחיש נתונים גיאוגרפיים, מה שמקל על ההבנה באילו אזורים יש כיסוי גבוה יותר או נמוך יותר של טיפול ART. סולם הצבעים יעיל להצגת שיפוע של ערכים, המאפשר השוואה חזותית מהירה בין מדינות. המפה יכולה לבטא מידע גיאוגרפי מפורט ולהראות דפוסים או אשכולות רלוונטיים גיאוגרפית. המפה מאפשרת השוואה קלה בין מדינות ואזורים שונים, ועוזרת לזהות פערים בכיסוי טיפול ART.

## חסרונות:

סולם הצבעים עשוי שלא להיות נגיש לצופים עיוור צבעים, מה שמקשה להבחין בהבדלים עדינים בכיסוי ART. מדינות גדולות יותר עלולות למשוך תשומת לב מיותרת, גם אם כיסוי ה-ART שלהן נמוך יותר, דבר העלול להטעות את הצופים. המפה עשויה לפשט נתונים מורכבים יתר על המידה, מכיוון שאינה יכולה לתפוס שינויים אזוריים בתוך מדינות גדולות.

## אלטרנטיבה 2 לשאלה משנית זו: Bar chart



## :Channels

- Color: צבע יחיד (כחול) משמש עבור כל העמודות.
- Spatial Position: ציר ה-x מייצג מדינות שונות. ציר ה-y מייצג את מספר האנשים המקבלים טיפולי ART.

## :Marks

- bars: כל פס מייצג את מספר האנשים המקבלים ART במדינה מסוימת. אורך הפס מתאים לערך על ציר ה-y.

## :יתרונות:

תרשים העמודות יעיל להשוואת מספר האנשים המקבלים ART במדינות שונות. הוא מספק ייצוג ברור וישיר של הנתונים. העמודות מסודרות לפי ערך, מה שמקל לראות באילו מדינות יש את המספר הגבוה ביותר והנמוך ביותר של אנשים שמקבלים ART.

## חסרונות:

האפקטיביות מוגבלת על ידי היעדר הקשר גיאוגרפי והעומס הפוטנציאלי של תוויות. הוא אינו מספק מידע על מגמות אזוריות או התפלגות מרחבית. כאשר מדינות רבות מופיעות על ציר ה-X, התוויות עלולות להיות עמוסות וקשות לקריאה, במיוחד אם הן מסובבות או חופפות. תרשים העמודות מספק תמונת מצב סטטית של הנתונים, אשר עשויה שלא לתפוס את התמונה המלאה של מגמות כיסוי ART לאורך זמן.

## למה אלטרנטיבה 1 עדיפה על אלטרנטיבה 2?

מפת Choropleth מתאימה יותר למענה על שאלת המחקר מכיוון שהיא מספקת הקשר גיאוגרפי ומראה תפוצה מרחבית. זה מאפשר הדמיה ברורה של מגמות אזוריות והבדלים בכיסוי ART בין מדינות. תרשים העמודות פחות יעיל עבור שאלת מחקר זו, למרות שהוא מספק השוואה ברורה של ערכים, הוא אינו מעביר את התמונה המלאה של מגמות סיקור ART העולמיות. בסך הכל, מפת choropleth היא בחירה מתאימה ויעילה יותר לניתוח כיצד כיסוי ART משתנה ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות. הוא מספק תובנות מפורטות ומדגיש דפוסים גיאוגרפיים חשובים, מה שהופך אותו לכלי בעל ערך עבור שאלת מחקר זו.

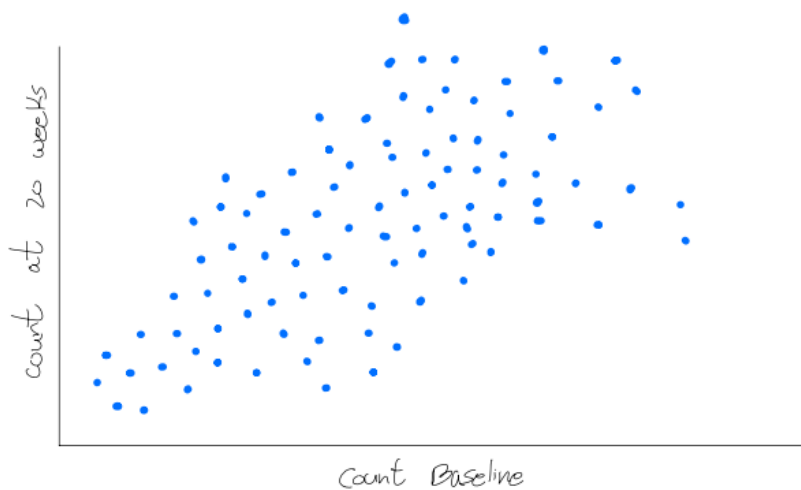


שאלת מחקר משנית: כיצד משתנים המדדים הקליניים (CD4 ו-CD8) בתגובה לפרוטוקולים השונים של טיפול אנטי-רטרו-ויראלי (ART) ולאחר 20 שבועות?

אלטרנטיבה 1 לשאלה משנית זו: Scatter Plot

Select ART Protocol: zdv only / zdv + ddI / zdv + zai / ddI only

Choose a type of white blood cell: CD4 / CD8



#### Channels

- Color - כחול למדידות מספר תאי הדם הלבנים מסוג CD4/CD8 בקרב כל מטופל למדידות בסיס ומדידות נוספות לאחר 20 שבועות.
- Spatial Position - ציר X מייצג את ספירת CD4/CD8 הבסיסית וציר Y מייצג את ספירתן לאחר 20 שבועות.
- Scatter Markers - משמשים למדידות מטופל בודד.

#### :Marks

- Points: כל נקודה מייצגת מדידה ספציפית של ספירת CD8 או ספירת CD4 עבור מטופל בתחילת הטיפול ולאחר 20 שבועות.

#### יתרונות:

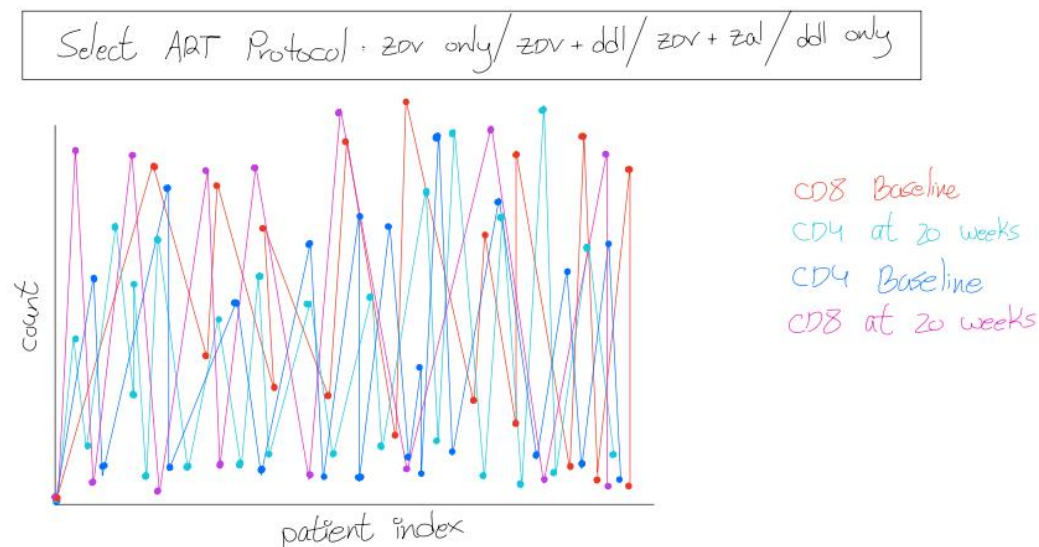
עלילת הפיזור מאפשרת הדמיה מפורטת של מטופלים בודדים, המראה את הקשר בין ספירת CD4/CD8 הבסיסית לספירתן לאחר 20 שבועות. הצבע הכחול הכהה המובהק למדידות הבסיס מספק תובנה חזותית מיידית לגבי ערכי ההתחלה, בעוד שהמגמה הכוללת לכיוון הימין העליון מציגה את העלייה לאחר 20 שבועות. הגרף הוא אקספרסיבי בהעברת מידע מפורט על נתוני המטופל, המאפשר זיהוי של מגמות וחריגים.

## חסרונות:

עם מספר רב של נקודות נתונים, הגרף יכול להיות עמוס, מה שמקשה על הבחנה בין נקודות נתונים ומגמות בודדות. עלילת הפיזור אינה מספקת מטבעה הקשר נוסף כגון התפלגות נקודות נתונים (למשל, צפיפות) או משתנים אחרים שעלולים להיות רלוונטיים.

## אלטרנטיבה 2 לשאלה משנית זו:

### Line Chart with scatter markers, also known as a Line-Scatter Plot



## :Channels

- Color: צבע channel משמש להבדיל בין מדידות הבסיס ל-20 שבועות של ספירת CD8 ו-CD4.
  - CD8: אדום משמש לקו הבסיס, ורוד משמש למדידות לאחר 20 שבועות.
  - CD4: כחול כהה משמש לקו הבסיס, תכלת משמש למדידות לאחר 20 שבועות.
- Spatial Position: ציר ה-x מייצג את אינדקס המטופל כלומר, כל ערך בציר ה-x מתאים למטופל אחר, וציר ה-y מייצג את ערכי ספירת ה-CD8 ו-CD4.
- Shape/Mark: סמני פיזור משמשים לציון נקודות נתונים בודדות, וקווים מחברים נקודות אלה כדי להציג מגמות לאורך זמן.

## :Marks

- Points: כל נקודה מייצגת מדידה ספציפית של ספירת CD8 או ספירת CD4 עבור מטופל בתחילת הטיפול או לאחר 20 שבועות.
- Lines: קווים מחברים את הנקודות כדי להראות את השינוי בספירת CD8 וספירת ה-CD4 על פני אינדקס המטופל.

## יתרונות:

השימוש בסמני פיזור יחד עם קווים מאפשר הדמיה מפורטת של נקודות נתונים בודדות תוך הצגת מגמות כוללות. ההפרדה הברורה לפי צבע בין מדידות התחלה ל-20 שבועות אחרי מספקת תובנה חזותית מיידיית לגבי ההבדלים והמגמות. הגרף הוא אקספרסיבי בהעברת מידע מפורט על נתוני מטופל בודדים ומאפשר זיהוי של מגמות וחריגים. הקווים המחברים עוזרים להמחיש את המגמה על מדד המטופל, מה שמקל לראות כיצד ספירות CD4 ו-CD8 משתנות בהתאם לכל מטופל.

## חסרונות:

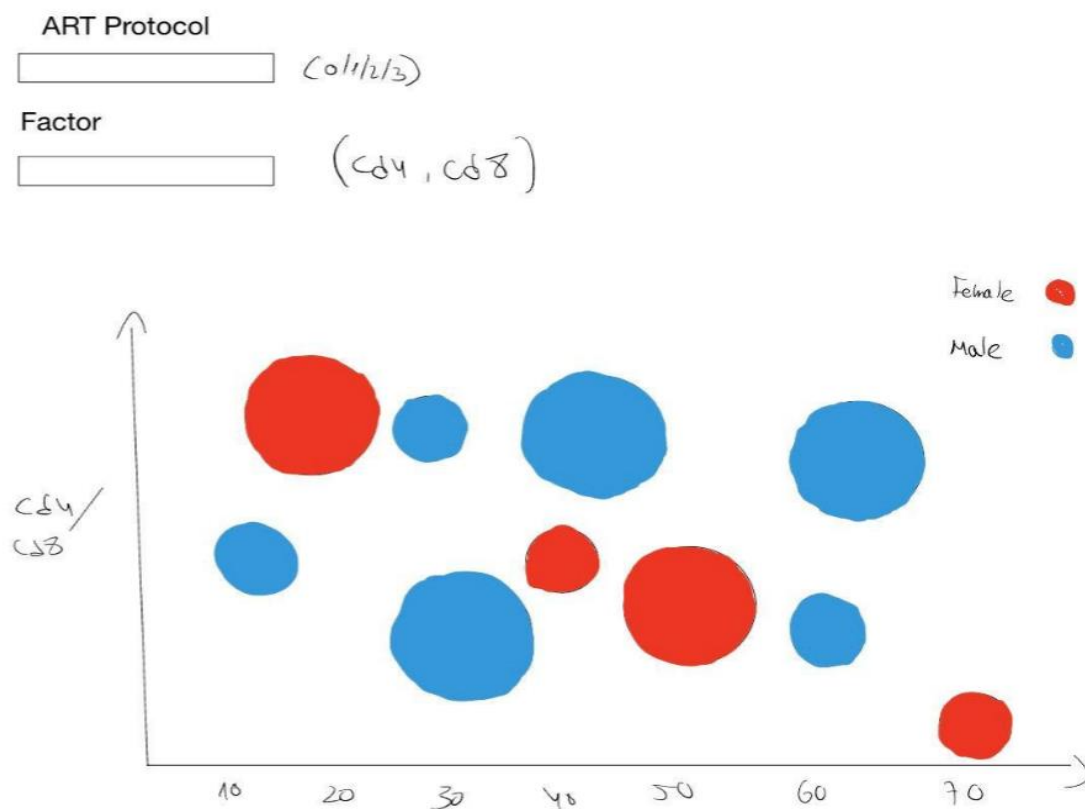
עם מספר רב של מטופלים, הגרף יכול להיות עמוס בנקודות נתונים חופפות, מה שמקשה על הבחנה בין נקודות נתונים ומגמות בודדות. השימוש באדום ובגוון בהיר יותר של אדום עשוי שלא להבדיל בקלות עבור חלק מהצופים, במיוחד אלה עם עיוורון צבעים. המורכבות של הגרף עשויה להקשות על חלק מהצופים לפרש אותו.

## למה אלטרנטיבה 1 עדיפה על אלטרנטיבה 2?

Scatter plot מועדף על פני תרשים קווים עם סמני פיזור בשל הבהירות והפשטות שלו, המדגיש למעשה את הקשר בין ספירת הבסיס לספירת CD4/CD8 לאחר 20 שבועות, ניתן לראות קורלציה חיובית לאחר 20 שבועות מקבלת הטיפול כמו שציפנו. זה שומר על הפרדה חזותית קלה יותר לפירוש, במיוחד עם מערכי נתונים גדולים. Scatter plot מונע עומס ומספק תובנות ברורות לגבי מגמות וחריגים, מה שהופך אותו לנגיש ופשוט יותר לניתוח. לעומת Line-Scatter Plot, שמחבר בין מטופלים (נקודות) על ידי קו ישר מכיוון שאין קשר בין מטופל אחד לשני והקו לא מייצג לנו מידע נוסף על ההשוואה בין הספירה שלהם. בנוסף, התצוגה לא ברורה ולא רואים מספיק את ההבדלים בין מצב נוכחי למצב לאחר 20 שבועות.

שאלת מחקר משנית: אילו נתונים דמוגרפיים של מטופלים (מין, גזע) משפיעים על ספירת ה-CD8 או ה-CD4 בקרב חולי איידס המטופלים בפרוטוקולי ART שונים?

אלטרנטיבה 1 לשאלה משנית זו: Bubble Chart



#### :Channels

- Color - אדום משמש לייצג משתתפות נשים. כחול משמש לייצג משתתפים גברים.
- Spatial Position: ציר ה-x מייצג את גיל המשתתפים. ציר ה-y מייצג את היחס CD4/CD8.
- Size - גודל הבועות מייצג את הגזע של המשתתפים.

#### :Marks

- Points: כל נקודה (בועה) מייצגת משתתף בודד.

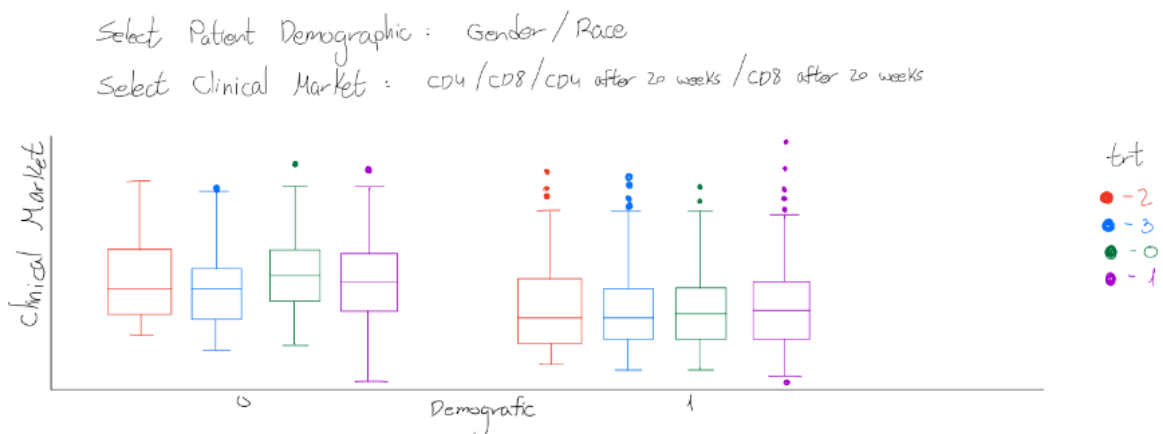
#### :יתרונות

השימוש בצבע (אדום לנקבה וכחול לזכר) מקל על הבחנה בין המינים במבט מהיר. גודל הבועה מייצג למעשה גזע, ומאפשר הערכה ויזואלית מהירה של פרופורציות יחסיות. מיקום על ציר x ו-y מספק ייצוג ברור ודו מימדי של גיל ויחס CD4/CD8, ומקל על זיהוי מגמות.

## חסרונות:

בועות חופפות יכולות לטשטש נקודות נתונים בודדות, מה שמקשה על זיהוי ערכים ספציפיים. המשמעות המדויקת של גדלי בועות עשויה שלא להיות ברורה מיד ללא מידע נוסף. ייתכן שלא ניתן להבחין בקלות בין אדום וכחול לאנשים עם ליקויים בראיית צבעים. שימוש בדפוסים או בצורות שונות עשוי לשפר את הנגישות. ללא קווי רשת או סמנים, פירוש ערכים מדויקים לגיל ויחס CD4/CD8 יכול להיות מאתגר.

## אלטרנטיבה 2 לשאלה משנית זו: BOX-PLOT



## Channels

- **Color**: ערוץ הצבע משמש להבדיל בין פרטוקולי ART שונים. אדום, כחול, ירוק וסגול מייצגים פרטוקולי ART שונים.
- **Spatial Position**: ציר ה-x מייצג את הדמוגרפיה שנבחרה (למשל, מגדר או גזע). ציר ה-y מייצג את הערך של הסמן הקליני שנבחר (ספירת CD8 או CD4).

## Marks

- **Boxes**: כל תיבה מייצגת את הטווח הבין-רבעוני (IQR) של הנתונים, ומציגה את ה-50% האמצעיים מהערכים. הקו בתוך התיבה מייצג את הערך החיצוני. קווי הטווח הנמשכים מהקופסאות מציגים את טווח הנתונים, בדרך כלל לא כולל חריגים.
- **Points**: נקודות בודדות מחוץ לתיבות מייצגות חריגים.

## יתרונות:

Box plots היא מאוד אקספרסיבית בהצגת הסיכום הסטטיסטי של הנתונים. יעילה להצגת התפלגות וטווח של ספירות CD4 ו-CD8 בין קבוצות דמוגרפיות שונות (מין, גזע). מאפשר השוואות חזותיות מהירות בין הקבוצות הללו וכיצד הם שונים תחת פרטוקולי ART שונים.

## חסרונות:

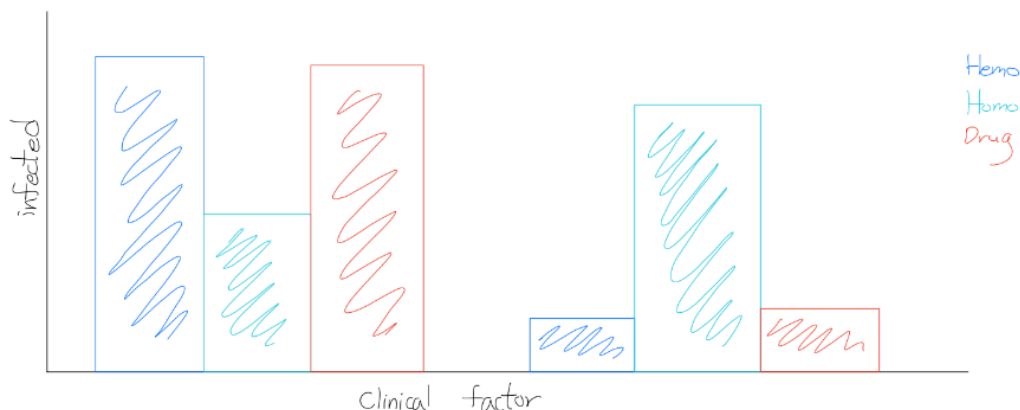
Box plots אינן מציגות נקודות נתונים בודדות, מה שעשוי להסתיר פרטים מסוימים של התפלגות הנתונים. עם מספר Box plots זו לצד זו, במיוחד עם ערכים דומים, זה יכול להיות מאתגר להבחין ביניהם בבירור. ניתן להפחית את האפקטיביות אם הצופה אינו בקיא בפרשנות Box plots. יש לשקול היטב את השימוש בצבע כדי להבטיח שהוא מובחן וברור. בעוד שחריגים מוצגים, הערכים המדויקים והסיבות לסטייה שלהם אינם בחרים.

## למה אלטרנטיבה 1 עדיפה על אלטרנטיבה 2?

לאחר שיצרנו את ה-box plot שאין מספיק הבדל בין הקבוצות ביחס לכמות ה-CD4/CD8 ושתרשים הבועות מעביר מימדים מרובים של נתונים (גיל, מגדר, גזע, יחס CD4/CD8) בתצוגה אחת, ומספק מערך נתונים עשיר יותר במבט אחד. כל בועה מייצגת נקודת נתונים בודדת, המאפשרת תצוגה מפורטת ומפורטת יותר של הנתונים בהשוואה לסטטיסטיקות הסיכום המוצגות בתרשים. תרשים הבועות מאפשר הדמיה טובה יותר של מגמות ודפוסים על פני משתנים מרובים, מה שמקל על זיהוי מגמות וחריגים.

שאלת מחקר משנית: כיצד משתנים שיעורי ההדבקה בנגיף HIV בהתאם לפקטורים שונים בפרוטוקולים שונים של טיפול אנטי-רטרו-ויראלי (ART)? (Drugs, Hemo, Homo)

### אלטרנטיבה 1 לשאלה משנית התחתונה: Grouped Bar Chart



#### :Channels

- Color: צבעים שונים משמשים לייצג את הגורמים הקליניים השונים: כחול עבור "Infected vs Hemo" (סטטוס המופיליה), כחול בהיר עבור "Infected vs Homo" (פעילות הומוסקסואלית), אדום עבור "Infected vs Drugs" (שימוש בסמים)
- Spatial Position: ציר ה-x מייצג את המצב הבינארי של הגורמים הקליניים (0 = לא, 1 = כן). ציר ה-y מייצג את מספר החולים הנגועים.

#### :Marks

- Bars: כל פס מייצג את ספירת החולים הנגועים התואמת לגורם הקליני הספציפי והאם חולה או לא (0 או 1).

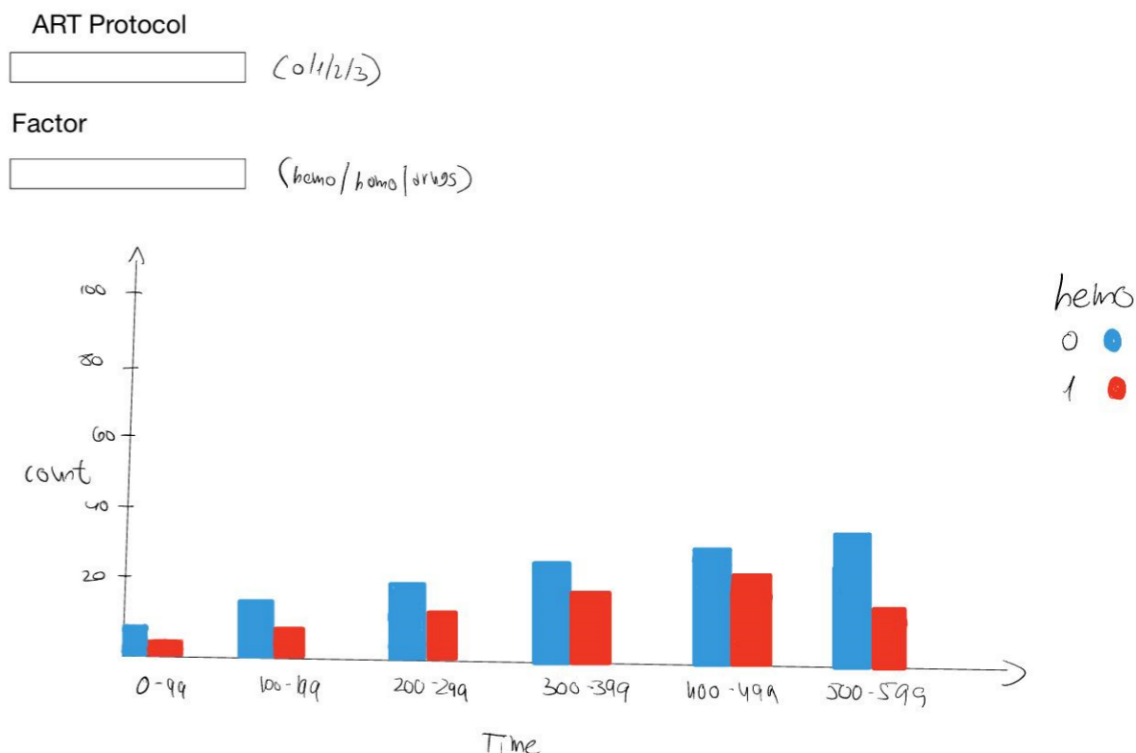
#### :יתרונות

תרשים העמודות המקובץ יעיל להצגת הבדלים בשיעורי ההידבקות ב-HIV בין גורמים קליניים שונים (המופיליה, פעילות הומוסקסואלית ושימוש בסמים). השימוש בצבעים שונים עבור גורמים קליניים שונים מסייע בהבחנה והשוואה של השפעתם על שיעורי ההידבקות ב-HIV. הפרדה ויזואלית זו מאפשרת לראות בקלות את ההבדלים בין הקבוצות. מה שהופך אותם לנגישים לקהל רחב, כולל אלה ללא רקע סטטיסטי.

#### חסרונות

תרשים העמודות מספק סקירה ברמה גבוהה אך חסר פירוט. הגרף יכול להיות אקספרסיבי יותר על ידי הכללת הקשר נוסף, כגון המספר הכולל של חולים בכל קבוצה או אחוז החולים הנגועים, כדי לספק יותר תובנה. האפקטיביות עשויה להיות מופחתת אם הגרף הופך עמוס ביותר מדי עמודות או אם ההבחנה בצבע לא ברורה.

## אלטרנטיבה 2 לשאלה משנית זו: Bar chart



### :Channels

- **Color**: ערוץ הצבע משמש לייצוג ספירת המשתנים עבור המשתנה הקליני שנבחר (Hemo/ Homo/ Drugs).
- **Spatial Position**: ציר ה-x מייצג את הזמן בימים, מחולק למרווחים. ציר ה-y מייצג את מספר האנשים שנדבקו באידס.

### :Marks

- **Bars**: כל פס מייצג את מספר האנשים שנדבקו באידס עבור כל קטגוריה של המשתנה הקליני שנבחר (למשל: 0 עבור לא הומוסקסואל, 1 עבור הומוסקסואל) בתור מרווח זמן נתון.

### :יתרונות:

Bar plots פשוטות וקלות לפירוש, מה שהופך אותה לנגישה לצופים עם רמות שונות של ידע סטטיסטי. הגרף מאפשר השוואה חזותית ישירה בין הקטגוריות, תוך הדגשת הבדלים בשיעורי ההדבקה, מה שמבהיר כמה אנשים נגועים באידס בכל קטגוריה.

### :חסרונות:

Bar plots מספק סקירה כללית ברמה גבוהה אך חסר עומק מבחינת הבנת ההתפלגות או גורמי השפעה פוטנציאליים בכל קטגוריה. ייתכן שהמשמעות המדויקת של המרווחים בציר ה-x לא תהיה ברורה מיד ללא הקשר או תוויות נוספות. ייתכן שלא ניתן להבחין בקלות בין אדם וכחול לאנשים עם ליקויים בראיית צבעים.

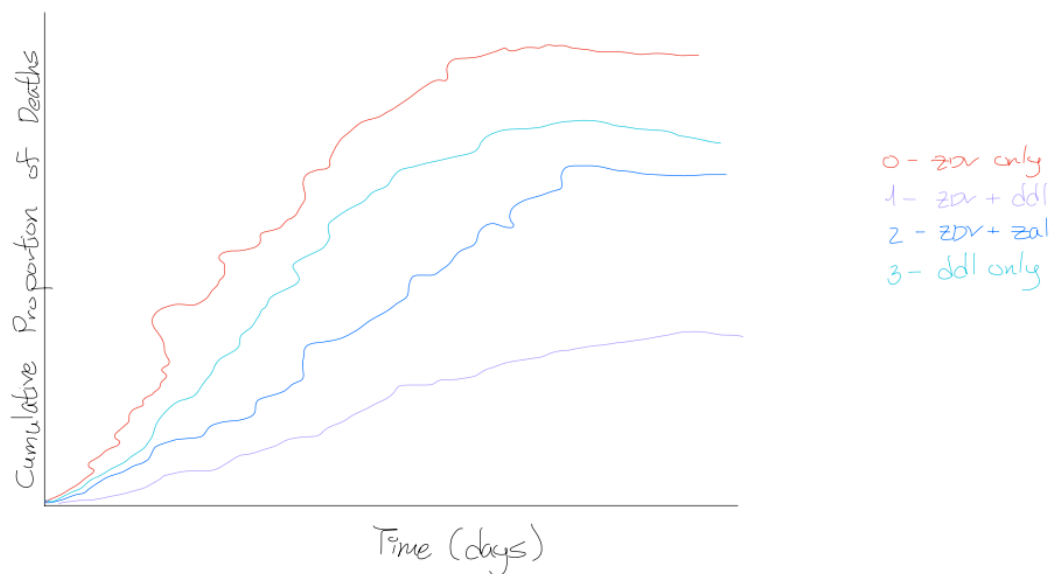


## למה אלטרנטיבה 1 עדיפה על אלטרנטיבה 2?

תרשים העמודות המקובץ עדיף על תרשים העמודות הפשוט מכיוון שהוא מאפשר השוואה בו-זמנית של מספר משתנים, ומספק תצוגה מקיפה יותר של הנתונים. על ידי הצגת גורמים קליניים שונים (כגון Hemo, Drugs ו-Homo) זה לצד זה, זה מקל על השוואות מפורטות וחושף קשרים מורכבים בין משתנים אלה. בנוסף, השימוש בצבעים ודפוסים שונים בתרשים העמודות המקובץ הופך אותו לאינפורמטיבי ועוזר להעביר מידע מורכב בצורה יעילה יותר.

שאלת מחקר משנית: מה הקשר בין פרטוקולי ART שונים ושיעור התמותה בקרב חולי איידס לאורך זמן?

אלטרנטיבה 1 לשאלה משנית זו: cumulative incidence plot



:Channels

- Color - ערוץ הצבע מבדיל בין פרטוקולי ART שונים, כאשר כל פרטוקול מיוצג על ידי צבע מובחן.
- Spatial Position: ציר ה-x מייצג את הזמן בימים, בעוד שציר ה-y מציין את השיעור המצטבר של מקרי המוות.

:Marks

- Lines - כל קו מייצג את עקומת ההיארעות המצטברת עבור פרטוקול ART ספציפי, המראה כיצד השיעור המצטבר של מקרי המוות משתנה לאורך זמן.

:יתרונות:

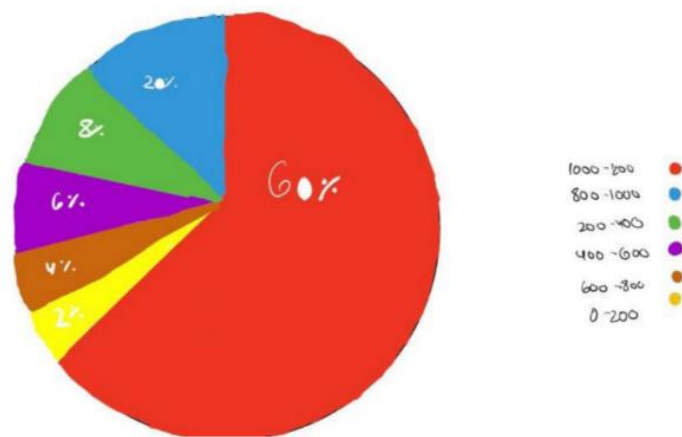
עקומות השכיחות המצטברות מציגות בבירור את התקדמות מקרי המוות לאורך זמן עבור כל פרטוקול ART, מה שמקל על השוואה בין היעילות של טיפולים שונים. השימוש בצבעים שונים עבור כל פרטוקול מאפשר השוואה חזותית מהירה ביניהם.

## חסרונות:

עבור צופים שאינם מכירים עקומות שכיחות מצטברות, הגרף עשוי להיות מאתגר יותר לפירוש בהשוואה לעלילות פשוטות יותר. הגרף אינו מסביר את הסיבות הבסיסיות להבדלים בין פרוטוקולי ה-ART, הדורשים ניתוח נוסף או נתונים נוספים לפרשנות מקיפה.

## אלטרנטיבה 2 לשאלה משנית זו: Pie Chart

Select ART Protocol : zdv only / zdv + ddl / zdv + zal / ddl only



## :Channels

- **Color:** צבעים שונים משמשים לייצוג טווחי זמנים (בימים) במהלך הניסוי.
- **Angle:** הזווית של כל פרוסה מתאימה לשיעור החולים שמתו בכל טווח זמן עבור כל פרוטוקול ART.

## :Marks

- **Slices:** כל פרוסה מייצגת חלק מהמטופלים שמתו בטווח זמן מסוים עבור פרוטוקול ART ספציפי.

## יתרונות:

תרשימי עוגה מספקים השוואה חזותית מהירה של שיעור מקרי המוות עבור כל פרוטוקול ART, מה שמקל לראות באילו פרוטוקולים יש שיעורי תמותה גבוהים או נמוכים יותר ובאיזו תדירות לאחר קבלת הטיפול. התרשים כולל לשונית בחירה המאפשרת למשתמשים לבחור פרוטוקולי ART ספציפיים לתצוגה. תכונה זו עוזרת למקד את ההשוואה בפרוטוקולים נבחרים, מה שהופך את התרשים לפחות עמוס ומותאם יותר לשאלות ספציפיות.

## חסרונות:

תרשים העוגה פחות אקספרסיבי להצגת מידע מפורט או השוואות קבוצות מרובות. היא לא מעבירה מגמות לאורך זמן או מאפשרת השוואות מדויקות. התרשים אינו מסביר את הסיבות הבסיסיות להבדלים בין

פרוטוקולי ART, הדורשים ניתוח נוסף או נתונים נוספים לפרשנות מקיפה. היעילות של קידוד הצבע יכולה להיות מופחתת עבור צופים עם עיוורון צבעים. הבטחת ניגודיות צבע ומתן תיאורי טקסט חלופיים יכולים לעזור להפחית זאת. בני אדם אינם טובים בתפיסת זוויות כפי שהם בתפיסת אורכים, מה שעלול להוביל לפירוש שגוי של הפרופורציות.

#### למה אלטרנטיבה 1 עדיפה על אלטרנטיבה 2?

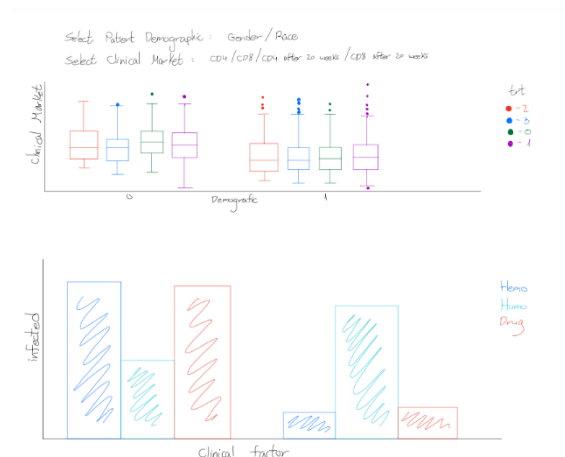
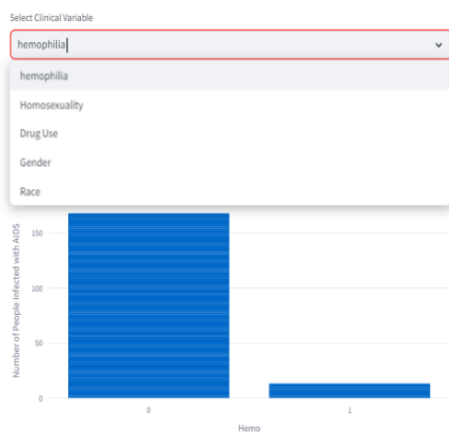
עלילות שכיחות מצטברות עדיפות מכיוון שהן מציגות מגמות ומידע מפורט לאורך זמן, מה שמאפשר השוואות ברורות ומדויקות יותר מאשר תרשימי עוגה.

## חלק 4 – הסבר על העיצוב שנבחר ועל יישומו

### Pre-processing of data

- טעינת מערכי הנתונים - טענו את מערכי הנתונים לתוך Panda DataFrames לצורך ניתוח
  - hiv\_df מכיל מידע על כיסוי ART לפי מדינה.
  - df מכיל מידע על סיווג AIDS ופרוטוקולי טיפול.
- בדיקת ערכים חסרים - בוצעה בדיקה עבור ערכים חסרים בעמודות הרלוונטיות של כל אחד ממערכי הנתונים:
  - במערך הנתונים hiv\_df עמודת Reported number of people receiving ART.
  - כל העמודות במערך הנתונים df נבדקו עבור ערכים חסרים.קיבלנו שאין בכלל ערכים חסרים בעמודות הרלוונטיות.
- יחידא שכל המדינות כלולות במפה - כדי להבטיח שכל המדינות מיוצגות במערך הכיסוי של ART: רשימה של כל המדינות התקבלה ממערך הנתונים gapminder של Plotly. רשימה זו מוזגה עם hiv\_df כדי לכלול את כל המדינות, גם אלה ללא נתוני ART. ערכים חסרים מולאו ב-0.
- יצירת עמודות בינאריות לסוגי טיפול - יצרנו משתנים בינאריים עבור עמודות ttt (טיפולים) במערך הנתונים של df, המאפשרים ניתוח קל יותר של פרוטוקולי טיפול שונים.

### ההבדלים בין העיצוב המקורי לעיצוב שמומש:



בתכנון הראשוני, התכווננו להשתמש בשני גרפים נפרדים כדי לענות על שתי השאלות:

1. אילו נתונים דמוגרפיים של חולים (מגדר, גזע) משפיעים על ספירת CD4 ו-CD8 בקרב חולי איידס שטופלו בפרוטוקולים שונים של ART?

שאלה זו תוכננה להיות מוצגת על ידי box plot המציגה את הקשר בין ספירת CD4/CD8 וגורמים דמוגרפיים (מגדר, גזע) לפי פרטוקולי ART השונים.

2. כיצד משתנים שיעורי ההדבקה ב-HIV בהתאם לגורמים שונים בפרטוקולים שונים של טיפול אנטי-רטרו-ויראלי (ART)? (Homo, Hemo, Drugs).

שאלה זו תוכננה להיות מוצגת על ידי bar plot המראה את שיעורי ההדבקה על פני גורמים קליניים שונים.

עם זאת, ביישום הסופי, החלטנו לשלב את ייצוג הנתונים לגרף אחד לניתוח מקיף והוליסטי יותר. העיצוב הממומש כולל כעת bar plot המציג את שיעורי ההדבקה על פני חמשת הגורמים השונים (Homo, Hemo, Gender, Race, Drugs) בתצוגה אחת.

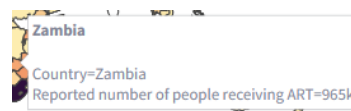
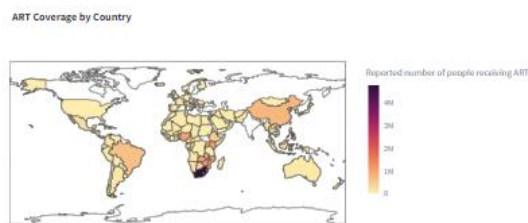
היתרון בגרף שהחלטנו לממש מספר משתנים לגרף אחד (חמשת המשתנים הם בינאריים), מה שמפשט את ההדמיה ואת ממשק המשתמש ומקל על השוואת שיעורי ההדבקה בין גורמים קליניים שונים מבלי לעבור בין שני גרפים שונים. בנוסף הגרף מספק תצוגה אחידה שעוזרת להבין כיצד גורמים שונים משפיעים יחד על שיעורי ההידבקות.

באמצעות השוואה נוכל לזהות יותר לאיזה גורם יש קשר יותר משמעותי עם שיעורי ההדבקה.

### צילומי מסך של הוויזואליזציה פועלת:

גרף לשאלה: כיצד משתנה הכיסוי של טיפול אנטי-רטרו-ויראלי (ART) ברחבי העולם וכמה אנשים מקבלים את הטיפול במדינות שונות?

ART Coverage by Country

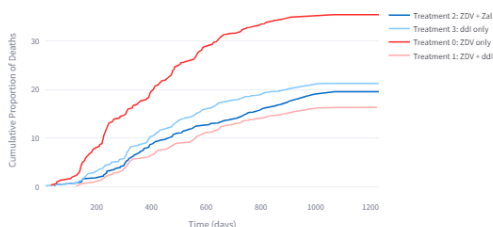


הגרף מציג את כיסוי טיפולי ה-ART במדינות שונות ברחבי העולם, כאשר הצבעים מייצגים את מספר האנשים המקבלים טיפול בכל מדינה (צבע לבן במפה מסמל על כך שהמדינה לא נמצאת בדאטה). האינטראקציה מאפשרת למשתמש "לרחף" מעל מדינות שונות ולקבל מידע מפורט על כמות המטופלים בכל מדינה ספציפית. ניתן לבצע זום אין וזום אאוט במפה כך ניתן לזהות את האזורים עם כיסוי הטיפול הגבוה ביותר והנמוך ביותר בקלות.

## גרף לשאלה: כיצד משתנים המדדים הקליניים (CD8 ו-CD4) בתגובה לפרוטוקולים השונים של טיפול אנטי-רטרו-ויראלי (ART) בנקודת התחלה ולאחר 20 שבועות?

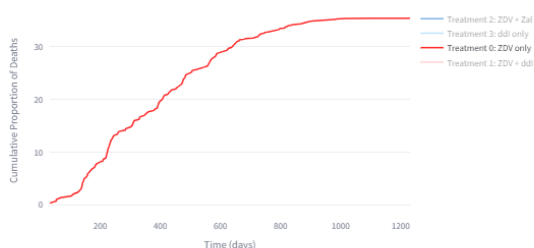
### Cumulative Incidence Analysis

Cumulative Incidence Curves by ART Protocol



### Cumulative Incidence Analysis

Cumulative Incidence Curves by ART Protocol



הגרף מציג את ערכי ה-CD8\CD4 לעומת הערכים לאחר 20 שבועות עבור פרוטוקול ART מסוים. כל נקודה מייצגת מטופל, כאשר ניתן לראות את הקשר בין ערכי ה-CD8\CD4 בשני זמני המדידה. האינטראקציה מאפשרת לבחור פרוטוקול ART וסוג תא דם לבן כדי לבחון את ההשפעה על מדדי ה-CD4 וה-CD8 בכל טיפול.

## גרף לשאלה: מה הקשר בין פרוטוקולי ART שונים ושיעור התמותה בקרב חולי איידס לאורך זמן?

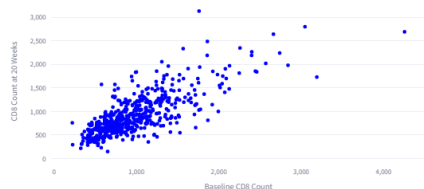
הגרף מציג את העקומות המצטברות של התמותה (Cumulative Incidence) עבור כל אחד מפרוטוקולי ה-ART לאורך זמן. כל קו מייצג פרוטוקול שונה ומציג את הפחופרציה המצטברת של מקרי התמותה לאורך זמן עבור כל פרוטוקול. האינטראקציה מאפשרת לבחור פרוטוקול ART ולהשוות בין ההשפעות השונות של כל פרוטוקול על התמותה. ניתן ללחוץ על האפשרויות בצד כדי להפעיל או לכבות כל פרוטוקול בנפרד.

### AIDS Progression Analysis

Select ART Protocol  
ZDV only

Choose a type of white blood cell  
CD8

Baseline vs 20 Weeks CD8 for Protocol 0

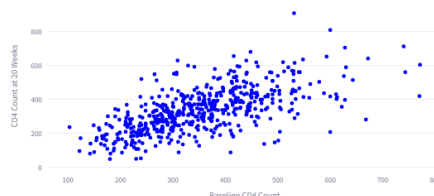


### AIDS Progression Analysis

Select ART Protocol  
ZDV only

Choose a type of white blood cell  
CD4

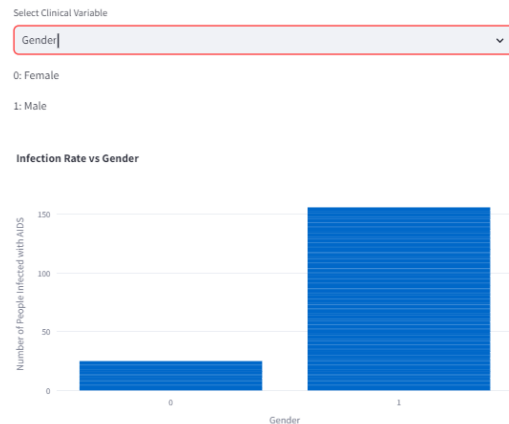
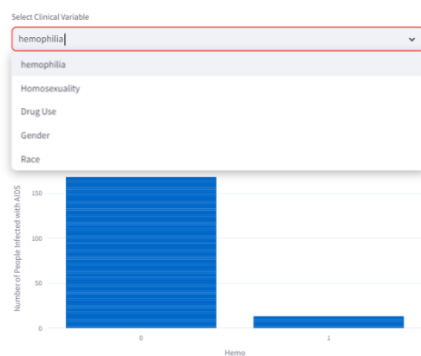
Baseline vs 20 Weeks CD4 for Protocol 0



## גרף לשאלות:

- אילו נתונים דמוגרפיים של מטופלים (מין, גזע) משפיעים על ספירת ה-CD4 וה-CD8 בקרב חולי איידס מטופלים בפרוטוקולי ART שונים?

- כיצד משתנים שיעורי ההדבקה בנגיף HIV בהתאם לפקטורים שונים בפרוטוקולים שונים של טיפול אנטי-רטרו-ויראלי (ART)? (Drugs, Hemo, Homo)



הגרף מציג את שיעור ההידבקות ב HIV ביחס למשתנה קליני נבחר. ניתן לבחור בין חמישה משתנים קליניים: הומוסקסואליות, המופיליה, שימוש בסמים, מין וגזע. האינטראקציה מאפשרת לבחור משתנה קליני באמצעות תיבת בחירה ולהציג את הנתונים בהתאם למשתנה הנבחר.

### כתיבת קוד:

לצורך כתיבת הקוד ויצירת הוויזואליזציה השתמשנו בחבילות שונות.

להלן הסבר קצר על מה שעשינו:

### ספריות:

streamlit: ליצירת אפליקציה אינטראקטיבית להצגת נתונים.

pandas: לטעינת הנתונים ולטיפול בהם.

plotly.graph\_objs ו-plotly.express: ליצירת גרפים ווויזואליזציות אינטראקטיביות.

יצרנו רשימת כל המדינות מ-plotly בשביל להבטיח שכל המדינות ייכללו בקובץ הנתונים (ולא רק המדינות הקיימות בדאטה).

הוספנו עמודות בינאריות לכל סוגי הטיפולים בנתונים.

### השתמשנו בפונקציות ליצירת גרפים שונים:

plot\_map: ליצירת מפה להצגת כיוון הטיפול לפי מדינה.

update\_cumulative\_incidence\_curve: ליצירת גרף של עקומות המציגות את שיעור התמותה המצטבר לאורך זמן, לפי פרוטוקולי ART שונים.

יצירת גרפים נוספים להצגת נתוני CD4 ו-CD8 לפני ואחרי 20 שבועות בהתאם לפרוטוקול ART שנבחר, וגרף בר להצגת שיעור ההידבקות לפי משתנים קליניים שונים.



### אינטראקציה ותצוגה:

אפשרנו למשתמש לבחור פרוטוקול ART, סוג תאי דם לבנים ומשתנים קליניים דרך תיבות בחירה אינטראקטיביות.

הצגנו גרפים אינטראקטיביים שהתעדכנו בהתאם לבחירות המשתמש.

לחץ [כאן](#) לקישור ל-Dashboard.