



ויזואליזציה של מידע 364-2-1721

סמסטר א' תשפ"ד

ארד פלג 206068801

עדי יוגב גולברי 307849869

קישור לתיקיית GITHUB

תיקיית GitHub מכילה את הקבצים הרלוונטיים (הנתונים, מחברת של הקוד ולינק לGoogle collab לגרפים אינטראקטיביים)

על מנת להריץ את הקוד יש לקרוא תחילה את קובץ הREADME

תוכן עניינים

2	הקדמה
2	רקע על הנתונים
2	פעולות טיוב נתונים שבוצעו:
2	מבנה הנתונים בתוך הקובץ לאחר השינויים
3	גרף מספר 1 (אנימציה)
3	השמות הנפוצים ביותר בכל שנה – מ1948 ועד 2021 (בחלוקה לפי מגדר)
4	גרף 1.1 (אינטראקטיבי)
4	השמות הנפוצים ביותר בכל שנה – מ1948 ועד 2021 (בחלוקה לפי מגדר)
5	גרף 2
5	השינויים בשמות הנפוצים ביותר (בחלוקה לפי מגדר)
7	גרף 3 (אינטראקטיבי)
7	חפיפה בשמות שיש לבנים ולבנות ושינויים לאורך השנים
8	גרף 4
8	הדתות שחולקות את המתאם הגבוה ביותר בין השמות
9	גרף 5
9	השמות עם החפיפה הכי גדולה בכל הדתות
9	גרף 5.1 (אינטראקטיבי)
9	השמות שנמצאים בכל הדתות (ללא משמעות לגודל החפיפה)
10	גרף 6
10	כמות אותיות בשם לפי דת
11	עובדות לא חשובות (נכון עד לשנת 2021)
12	עובדות לא חשובות - שמות ומלחמות



הקדמה

רקע על הנתונים

- הנתונים מכילים שמות פרטיים של תושבי ישראל, לפי שנת לידה, דת ומין, 1948 – 2021.
- הנתונים מבוססים על קובץ מרשם האוכלוסין המתקבל ממשרד הפנים.
- הנתונים עוברים בלמ"ס תהליך של טיוב ועיבוד.
- הנתונים מגיעים בקובץ Excel אחד ובתוכו עשרה גיליונות שונים (ע"פ חלוקה לדת ומגדר)

פעולות טיוב נתונים שבוצעו:

- איחוד כל השמות לגיליון אחד
- הוספת 2 עמודות (דת ומגדר)
- לא נלקחו השמות מהגיליונות 'אחרים' ו'אחרות'
- תיקון עמודת השם הפרטי – לצורך שאיבת השם בצורה נכונה (שלא יופיע הפוך בפלטים)

מבנה הנתונים בתוך הקובץ לאחר השינויים

Religion	Gender	2021	1948	SumYears	Firstup	Firstname
Jew	Male	..	.	179	אבא	אבא
Jew	Male	10	34	9,047	יבא	אבי
Jew	Male	.	.	29	סהרבא-יבא	אבי-אברהם
Jew	Male	32	.	523	רואיבא	אביאור
Jew	Male	180	..	5,734	לאיבא	אביאל

- לדוגמא בשנת 1948 נולדו בישראל 34 בנים בשם אבי
- נקודה אחת (.) – חוסר מקרים
- שתי נקודות (..) – תועדו פחות מ-5 מקרים
- לצורכי חישוב תא אשר הכיל (.) או (..) הוחלף ב-0
- לצורך נוחות הטבלה המוצגת כאן מכילה רק את השנים 1948 ו-2021

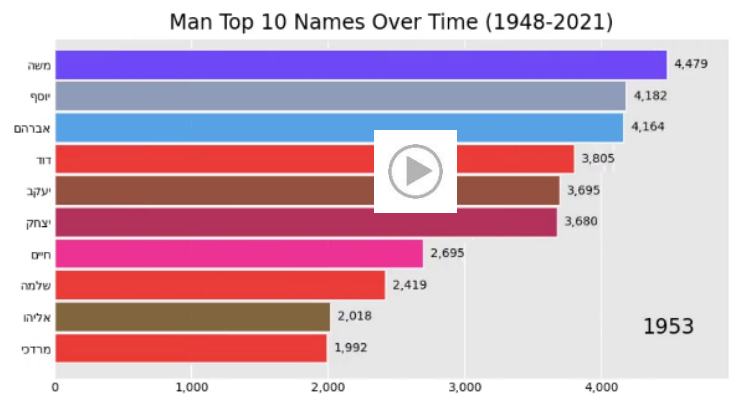
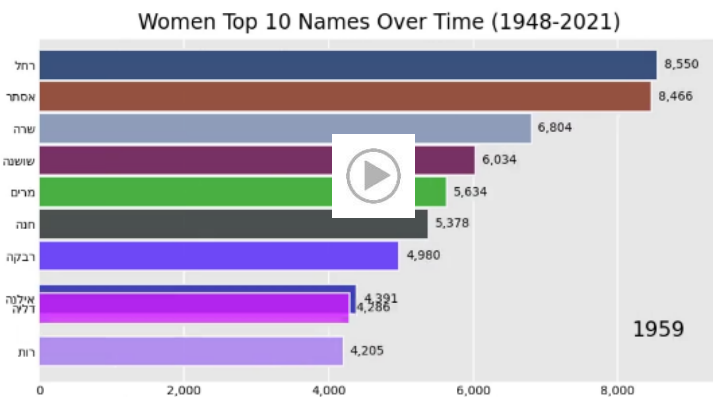


גרף מספר 1 (אנימציה)

השמות הנפוצים ביותר בכל שנה – מ1948 ועד 2021 (בחלוקה לפי מגדר)

idiom	Bar chart
What Data	משתנה כמותי (כמות האנשים שנקראו בשם מסויים) ומשתנה קטגורי (השמות)
How Encode	סימוני עמודות מיוצגות בצורה רוחבית מיושרת לפי משתנה כמות של השם. השימוש באנימציה מאפשר להתמקד ב10 השמות בצורה נוחה יותר
Why Task	השוואת של עשרת השמות הנפוצים בכל שנה (בחלוקה לפי מגדר)
Scale	ציר X - כמות האנשים בשם מסויים ממאות עד עשרות אלפים ציר Y - שמות האנשים בצד מימד הזמן בשנים מ1948 עד 2021

****יש ללחוץ על הנגן של כל גרף****



הבחירה בBar Chart מאפשרת להשוות מספר גדול של משתנים קטגוריאליים, השימוש באנימציה מאפשר לקבל תמונה ברורה עבור עשרת השמות הכי נפוצים בכל שנה.

החישובים שנעשו:

לקיחת עשרת שמות הנפוצים ביותר בשנה מסוימת ולא בוצע סינון של דת.

מסקנות (גברים):

עד 1970 עשרת השמות הנפוצים ביותר בישראל היו של יהודים, החל מ1970 חלה עליה משמעותית בכמות הבנים בשם 'מוחמד' ולמעשה זהו השם הנפוץ ביותר עבור בנים משנת 1993 עד שנת 2021.

מרבית מהשמות הנפוצים לאורך השנים הם שמות אשר לקוחים מהתנ"ך.

מסקנות (נשים):

משנת 1948 ועד שנת 2008 השמות הנפוצים ביותר היו או אסתר או רחל.

בדומה לגברים גם השמות הנפוצים ביותר לנשים נלקחו מהתנ"ך.



גרף 1.1 (אינטראקטיבי)

השמות הנפוצים ביותר בכל שנה – מ1948 ועד 2021 (בחלוקה לפי מגדר)

בנוסף לגרף 1, רצינו לאפשר משחק עם הנתונים עבור השמות הנפוצים ביותר בכל שנה בעזרת ממשק אינטראקטיבי.

דוגמא עבור **בנות** בשנת 2021 (השנה האחרונה בנתונים)

Top 10 Girls' Names in 2021



דוגמא עבור **בנים** בשנת 2021 (השנה האחרונה בנתונים)

Top 10 Boys' Names in 2021





גרף 2

השינויים בשמות הנפוצים ביותר (בחלוקה לפי מגדר)

idiom	Line chart
What Data	משתנה כמותי (כמות האנשים שנקראו בשם מסוים) ומשתנה אורדינלי (השנים)
How Encode	סימון ע"י נקודות אשר מיושרת עם הצייר האנכי (אחוז האנשים שנקראו בשם מסוים מסך השמות באותה השנה) ומחוברות בקו ביניהן ע"י סדר רוחבי (השנים)
Why Task	על מנת להראות את מגמת השינוי באחוז האנשים בשם מסוים מכלל האוכלוסייה לאורך השנים
Scale	ציר X - זמן - בשנים מ-1948 עד 2021 ציר Y - אחוז השם מכלל השמות באותה שנה מ-0% ועד 100% (בגרפים עד 2.5%)

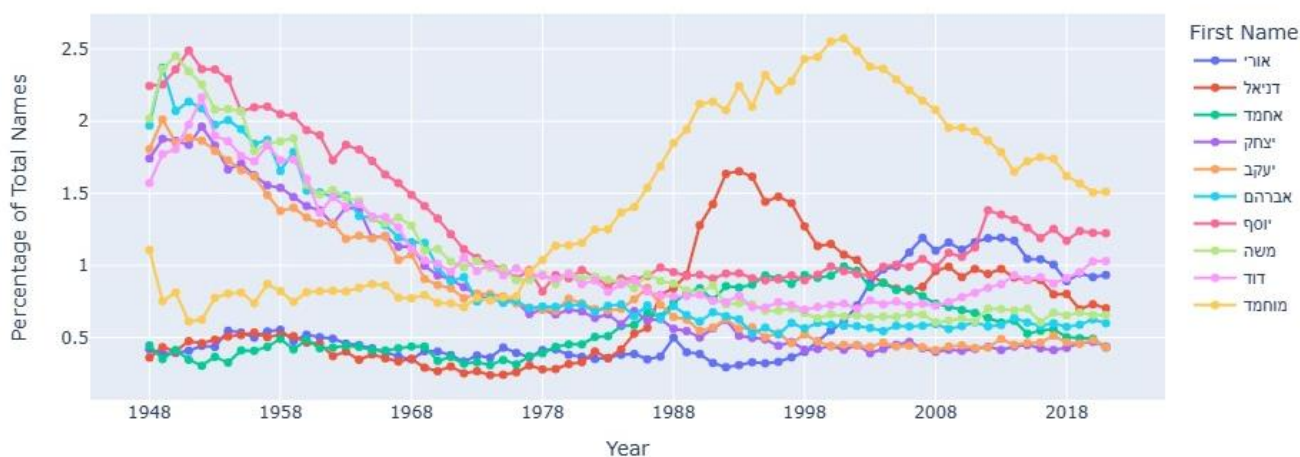
בשונה מהגרף הקודם, בו בכל פעם ניתן לעקוב אחרי עשרת השמות הכי נפוצים **בכל שנה** בגרף הנ"ל מצוינים עשרת השמות הנפוצים ביותר **בכל תקופת הזמן** (משנת 1948 ועד שנת 2021 סה"כ).

הבחירה ב-Line Chart מאפשרת לבדוק בצורה נוחה האם השמות הנפוצים ביותר בכל תקופת הזמן תמיד היו הכי נפוצים או שמא חל שינוי משמעותי בהם.

החישובים שנעשו:

על מנת לנטרל את הגידול הטבעי באוכלוסייה שעלול להצביע רק על עליות בשמות הבדיקה נעשתה על אחוז של השם מכלל השמות באותה השנה ולא על הכמות.

Top Ten Most Popular Names Of All Time (Male)

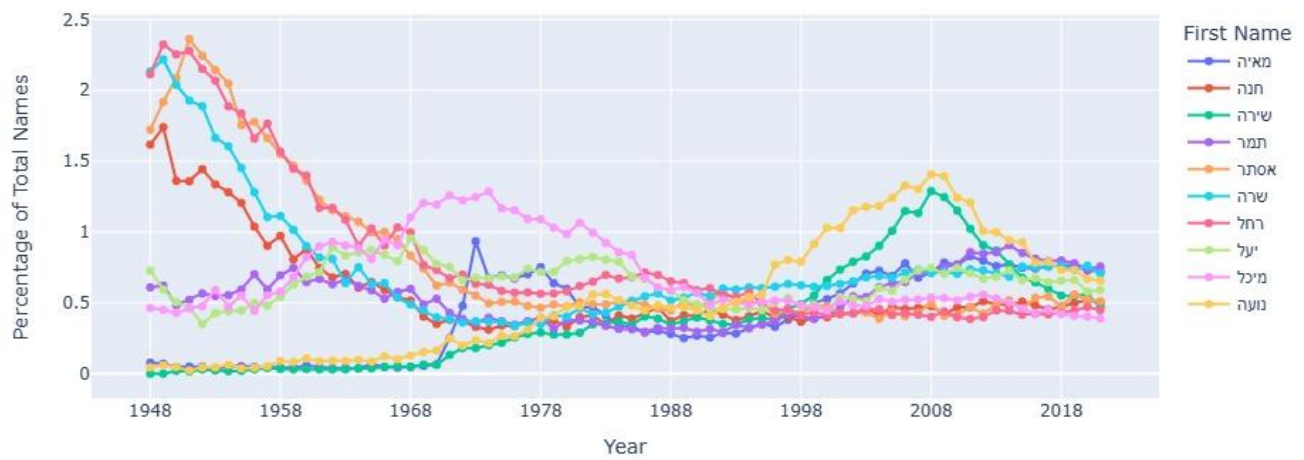


מסקנות (גברים):

הגרף מאפשר לנו לצפות בשתי מגמות, שמות שהיו מאוד נפוצים פעם וכיום נמצאים במגמת ירידה (מרבית השמות בגרף) והמגמה ההפוכה בה שמות שהיו פחות נפוצים פעם הפכו לנפוצים היום.



Top Ten Most Popular Names Of All Time (Female)



מסקנות (נשים):

ניתן להבחין בשלוש קבוצות, האחת במגמת עליה, השנייה במגמת ירידה כמו אצל הגברים, אך קיימת קבוצה נוספת בה השמות שומרת על אחוז פופולריות יחסית יציב ביחס לאוכלוסייה.



גרף 3 (אינטראקטיבי)

חפיפה בשמות שיש לבנים ולבנות ושינויים לאורך השנים

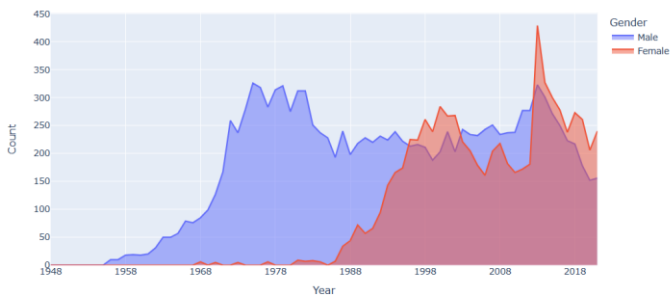
idiom	Stream graph
What Data	גרף רב מימדי, משתנה כמותי (כמות האנשים שנקראו בשם מסויים), משתנה אורדינלי (השנים) ומשתנה קטגורי (מגדר)
How Encode	סימון ע"י התפלגות משתנה כמותי (כמות האנשים שנקראו בשם מסויים) לאורך המשתנה האורדינלי (השנים), כאשר כל התפלגות בצבע שונה ע"י המשתנה הקטגורי (מגדר)
Why Task	מאפשר להשוות בקלות יחסית בין שני טרנדים במקביל
Scale	ציר X - זמן - בשנים מ-1948 עד 2021 ציר Y - כמות אנשים באותו שם ממאות עד עשרות אלפים מימד המגדר - שני צבעים: בנים-כחול, בנות-אדום

הבחירה ב-Stream Graph מאפשרת להשוות בין שני טרנדים של שני משתנים קטגוריאליים במקביל.

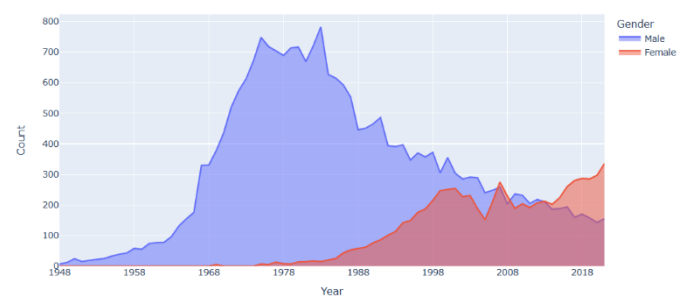
החישובים שנעשו:

- 1) בניית רשימת עבור שמות נפוצים בשני המינים
- 2) יצירת גרף עבור כל שם
- 3) בחירת הגרפים בעלי המגמות המשמעותיות ביותר

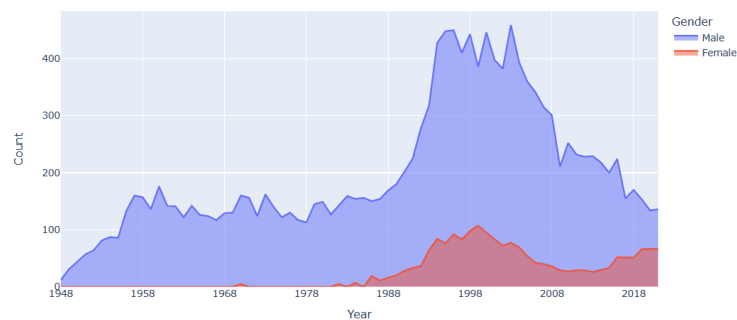
Name Popularity Over Time by Gender (1948–2021): 'אופיר'



Name Popularity Over Time by Gender (1948–2021): 'שי'



Name Popularity Over Time by Gender (1948–2021): 'רון'



מסקנות:

שמות רבים של גברים הפכו להיות פופולריים בקרב נשים (בעיקר מתחילת שנות השמונים).

במקרים מסוימים שמות שהיו מזוהים בעיקר עם גברים הפכו לשמות שמזוהים בעיקר עם נשים כפי שניתן לראות בגרפים העליונים של 'שי' ו'אופיר'.

במידה והמגמות ימשיכו כפי שהן היום, השם 'רון' (הגרף התחתון) יהפוך להיות שם נפוץ יותר בקרב נשים.



גרף 4

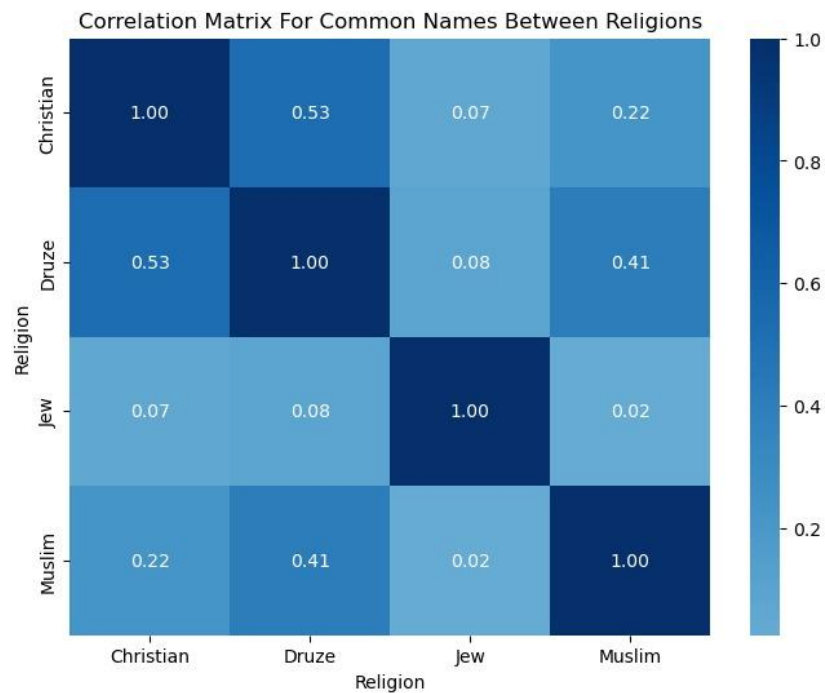
הדתות שחולקות את המתאם הגבוה ביותר בין השמות

idiom	Heat map
What Data	משמש עבור 2 משתנים קטגוריאליים, במקרה הזה אותו משנה מופיע בשני צידי המטריצה (דת) ומשתנה כמותי (המתאם בין השמות)
How Encode	מטריצה דו מימדית
Why Task	מאפשרת למצוא קשרים בין הדתות, כלומר בין אילו דתות יש את התפלגות השמות הדומה ביותר
Scale	המשתנה הכמותי - מתאם בין 0 ל1

הבחירה בHeat Map מאפשרת להשוות בצורה נוחה בין שני משתנים קטגוריאליים ולבדוק האם קיים מתאם ביניהם או האם הם מסווגים לאותה קבוצה.

החישובים שנעשו:

- יצירת טבלה חדשה שתכיל את תוצאת פעולת `group by (sum)` לכל שם לפי דת.
- יצירת מטריצת קורלציה כאשר שני הצירים מכילים אותם המשתנים (ארבעת הדתות) ע"י שימוש בפעולת `pivot` לטבלה החדשה שיצרנו.



מסקנות:

המתאם הגדול ביותר בשמות בין הדתות הוא בין הנוצרים לדרוזים (0.53) ולאחר מכן בין הדרוזים למוסלמים (0.43) ואילו המתאם הנמוך ביותר הוא בין המוסלמים ליהודים (0.02)

רעיון לבדיקת המשך, להניח שאברהם (יהודים) ואיברהים (מוסלמים) זה אותו השם בקירוב ולבדוק האם המתאם עולה באופן משמעותי בין שתי הדתות.



גרף 5

השמות עם החפיפה הכי גדולה בכל הדתות

idiom	Spider/Radar Graph
What Data	משתנה כמותי (כמות האנשים שנקראו בשם מסויים), משתנה קטגורי (דתות)
How Encode	עבור כל שם ממקמים את הערכים הכמותיים בהתאמה למשתנה הקטגורי באופן רדיאלי ומחברים את הערכים ע"י קו
Why Task	מאפשר להשוות התפלגות של שם מסויים בין הדתות, ועד כמה הוא נפוץ.
Scale	המשתנה הכמותי - מאחדות עד מאות

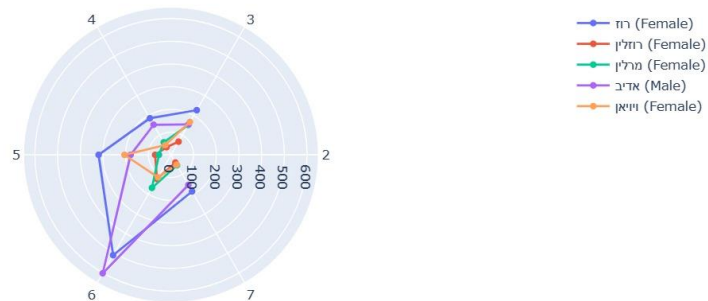
הבחירה ב Spider Graph מאפשרת להציג פיזור של משתנים קטגוריאליים ע"י תכונות קטגוריאליות.

החישובים שנעשו:

- 1) בדיקה של השמות שמופיעים בכל הדתות
- 2) עבור כל שם חושב היחס בין הערך המינימלי (כמות השמות הנמוכה ביותר באותה השורה לפי דת) לערך המקסימלי עבור כל שם.
- 3) דירוג על פי תוצאת סעיף 2
- 4) לקיחת 5 השמות בעלי הדירוג הגבוה ביותר

הסיבה שבה ביצענו את החישוב כך היא כי רצינו לבדוק מהו השם הנפוץ ביותר בכל הדתות אבל גם שההתפלגות שלו (כמות האנשים בשם זה) בין הדתות יחסית דומה.

Top Five Names With The Greatest Overlap Between Religions



מסקנות:

רוב השמות החופפים הם בעיקר של בנות.
רוב השמות החופפים נפוצים על פי אותו יחס בכל דת (מלבד 'וויאן')

גרף 5.1 (אינטראקטיבי)

השמות שנמצאים בכל הדתות (ללא משמעות לגודל החפיפה)

בדומה לגרף הקודם, אך ללא סינון של 5 השמות בעלי החפיפה הגדולה ביותר.



גרף 6

כמות אותיות בשם לפי דת

idiom	Spider/Radar Graph
What Data	משתנה אורדינלי (כמות אותיות בשם), משתנה כמותי (כמות האנשים עם כמותיות אותיות מסוימת בשם) ומשתנה קטגורי (דתות)
How Encode	ממקמים את הערכים הכמותיים בהתאמה למשתנה האורדינלי באופן רדיאלי ומחברים את הערכים ע"י קו, הקו מסמן את המשתנה הקטגורי
Why Task	מאפשר להשוות התפלגות של שם מסויים בין הדתות, ועד כמה הוא נפוץ.
Scale	המשתנה האורדינלי מ-2-7 המשתנה הכמותי - מיליונים (ובהתאמה בסקאלה הלוגריתמית)

הבחירה ב Spider Graph מאפשרת להציג פיזור של משתנים קטגוריאליים ע"י תכונות קטגוריאליים.

החישובים שנעשו:

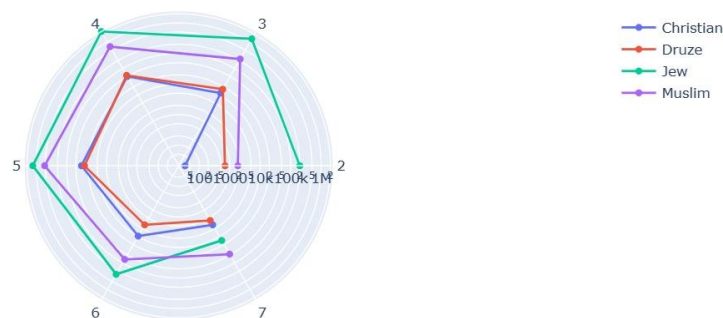
- חישוב כמות האותיות בכל שם
- יצירת טבלה חדשה שתכיל את תוצאת פעולת `group by (sum)` לכל קבוצת אותיות לפי דת.

Letters Counts Per Name Across Religions



- מכיוון שיש כמות משמעותית יותר של שמות של יהודים נאלצנו לבצע נרמול של הנתונים (LOG)

Letters Counts Per Name Across Religions (Log Scale)



מסקנות:

התפלגות כמות האותיות בשם דיי דומה בין הדתות

ביהדות רוב השמות הם בין 3 ל-6 אותיות

בנצרות יש מעט שמות עם 2 אותיות בלבד



עובדות לא חשובות (נכון עד לשנת 2021)

שמות ומלחמות

בשנת 1967 השם הנפוץ ביותר היה 'משה' ייתכן בשל העובדה שהוא מסמל את ראשי התיבות של מלחמת ששת הימים שנערכה באותה השנה.

מסיבה דומה בשנת 1973 השם הנפוץ ביותר היה 'מיכל', ככל הנראה בגלל הקרבה לראשי התיבות של מלחמת יום כיפור (בתוספת האות 'ל')

'נצחון' / 'ניצחיה' (26 אנשים)

16 נשים בשם ניצחיה ו10 גברים בשם נצחון

מחציתם נקראו כך בשנת 1967, ככל הנראה בעקבות הניצחון במלחמת ששת הימים.

שנים מיוחדות

'עשור' / 'עשורה' (24 אנשים)

5 נשים בשם עשורה ו19 גברים בשם עשור

16 מתוכם נקראו כך בשנת 1958 (בחגיגות העשור למדינה)

'יובל'

בשנת 1998, בחגיגות ה50 (היובל) למדינה,

1,823 אנשים נקראו בשם יובל

מספר זה היה גבוה ב45% מאנשים שנקראו יובל בשנת 1997 (1,261)

וגבוה ב48% מאנשים שנקראו יובל בשנת 1999 (1,230)

ואיך אתם כותבים?

חווה (8,151) מול חווה (78)

יהונתן (26,969) מול יונתן (23,296) מול ג'ונתן (104)

ערד (167) מול ארד (917)

טום (2,000) מול תום (6,247) מול תם (828)

ארץ - עיר

21 אנשים נקראו דימונה מרביתם בין השנים 1955-1957 בסמוך להקמת העיר בשנת 1955

16 נשים נקראו בשם ירושלמית

3,875 אנשים נקראו בשם בארי (89% גברים)

167 אנשים נקראו ערד כולם גברים

'רועי'

למרות שרועי הוא אחד מהשמות הכי נפוצים היום,

לא תועדו מקרים של אנשים שנקראו כך

מקום המדינה ועד שנת 1954

וואלה מפתיע

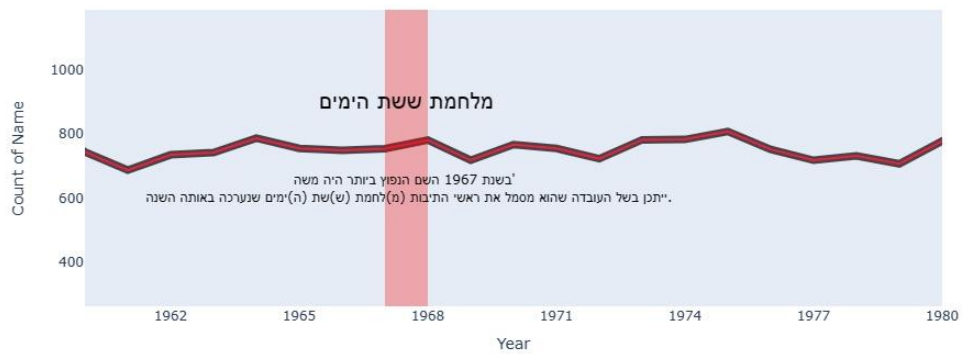
ישנם 462 אנשים בשם 'יום'

בשמות יהודים ישנן 57 קומבינציות לשמות שמורכבים משני שמות

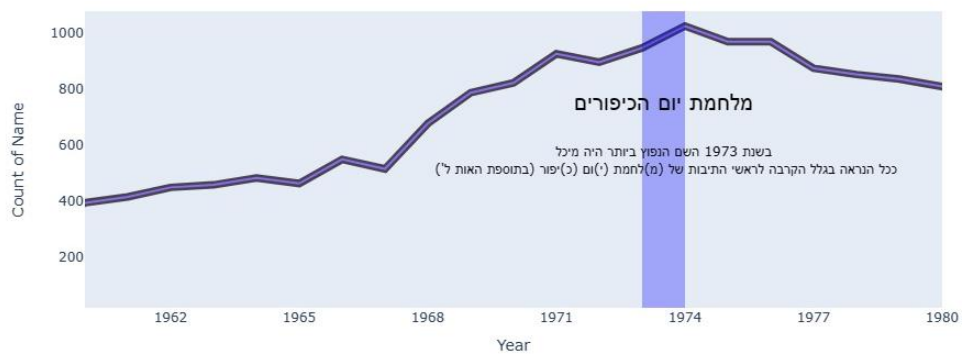


עובדות לא חשובות - שמות ומלחמות

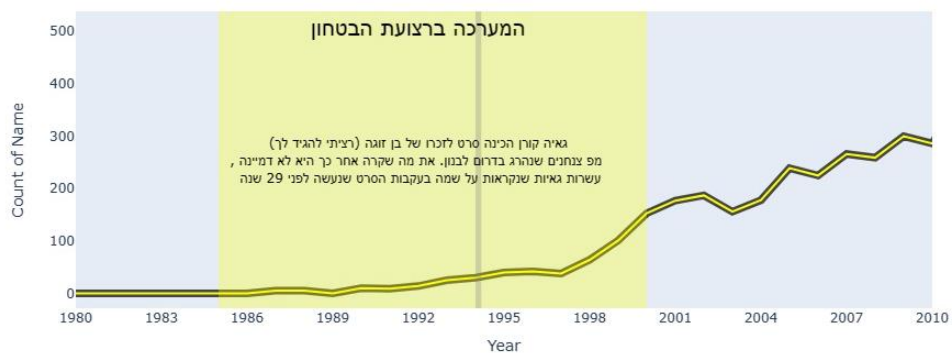
Names through war stories: 'משה'



Names through war stories: 'מיכל'



Names through war stories: 'גאיה'



Names through war stories: 'איתן'

