

– ויזואליזציה של מידע –
"1000 songs to hear
before you die"

מגישים:

יובל ברוה: 201531043

ראני מוסללם: 203622568

הקדמה

בפרויקט הזה לקחנו דאטה שמצאנו באינטרנט – "1000 שירים לשמוע לפני שתמות", השירים רק באנגלית והיוצרים הם אמריקאים. הדאטה מתחיל בשנת 1916 עד 2008. הדאטה אינו סוג של מסמך מדעי מדויק, הוא בחירה של בן אדם מסוים שמשקף את רוח המוזיקה במאה האחרונה.

קצת על הדאטה, הדאטה מחולק ל-6 קטגוריות: אהבה, פוליטיקה, אנשים ומקומות, שובר לב, סקס, חיים ומוות, מסיבות.

כל שורה בדטה מכילה: שם האמן, נושא, שם השיר, שנה וחלק מהשורות יש לינק לשיר.

הדאטה שקיבלנו הוא בפורמט CSV במהלך הפרויקט עשינו עליו כל מיני תהליכים בהם המרה ל-TSV, JSON ועוד כל מיני עיבודים.

העיבודים נעשו בשפת פייתון, על מנת למצות את החלקים בדאטה שמצאנו מעניינים.

הכנו מספר ויזואליזציות בפרויקט את כולן אפשר למצוא תחת הלינקים הבאים:

דוגמא לויזואליזציה גרועה - Pie chart:

<http://ranimsallam.github.io/vis/PieChart>

ויזואליזציה בנושא פוליטיקה Bar Chart:

<http://ranimsallam.github.io/vis/barChart>

ויזואליזציה עבור האמנים הטובים ביותר (ויזואליזציה לא טובה שלמדנו ממנה):

<http://ranimsallam.github.io/vis/index.html>

ויזואליזציה עבור האמנים הטובים ביותר Tree map:

<http://ranimsallam.github.io/vis/VisioNew>

ויזואליזציה עבור האמנים הטובים ביותר Circle Map:

<http://ranimsallam.github.io/vis/circleViso>

ויזואליזציה ראשונה – Pie Chart

כחלק מלמידת הפלטפורמה D3 התחלנו בדוגמא פשוטה. עיבדנו את הדאטה ומצאנו כמה שירים יש מכל נושא. הנה התוצאה:

<http://ranimsallam.github.io/vis/PieChart>



Learning D3 - Pie chart

This Chart shows the partition of the themes in the data. Love, Sex, Politics, heart break, people and places, life and death, party as you can see this is a bad visualization of the data. you cant really understand anything valuable from this visualization.

אחרי הויזואליזציה הזאת, הבנו בעצם, איך לא עושים ויזואליזציה.

אין שום מידע מעניין שאפשר להבין דרך הויזואליזציה. הגדלים דומים, הדבר היחיד שאפשר להבין זה שמספר השירים מכל נושא דומה במספרו (חוץ מנושא אחד - סקס).

דבר זה הוביל אותנו להבנה, שקודם כל נחשוב על מה אנחנו רוצים להבין מהדאטה ואז לייצר את הויזואליזציה.

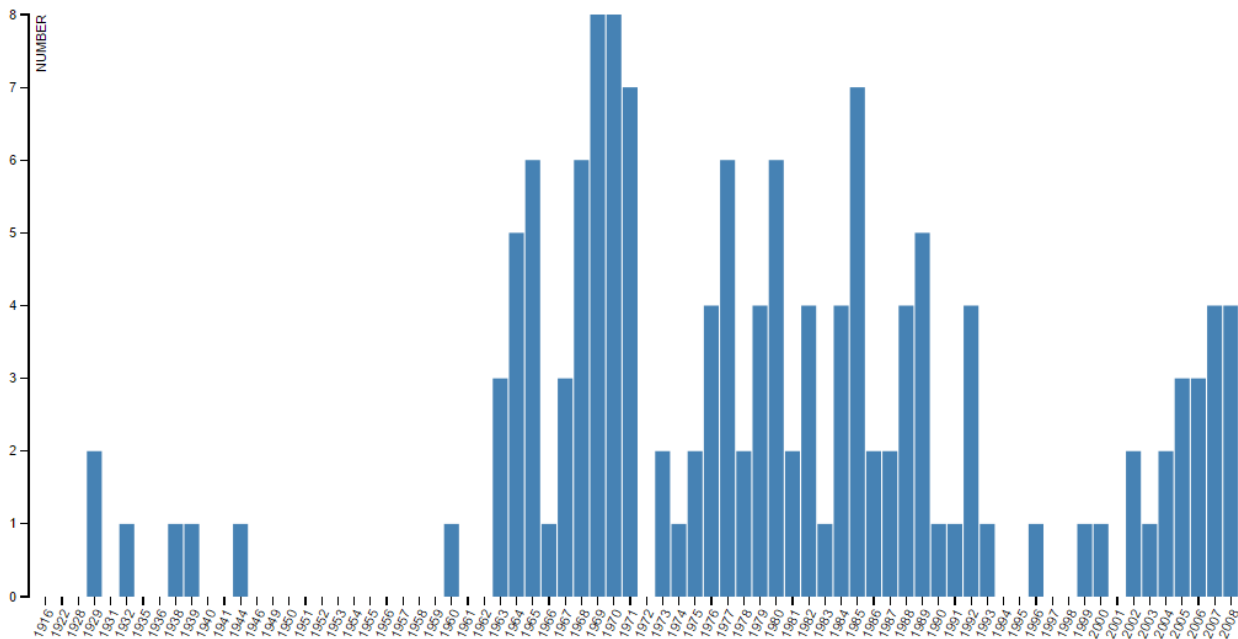
ויזואליזציה שניה – Connection between wars and songs about politics

חשבנו שיהיה מעניין לבדוק: "האם קיים קשר בין מספר השירים שיש בנושא פוליטיקה בשנה מסוימת ובין מצבה המלחמתי של ארצות הברית.

כלומר אם בתקופה שבה ארצות הברית הייתה במלחמה יש גידול במספר השירים בנושא פוליטיקה על פי הדאטה – התוצאות היו מעניינות.

הנה הויזואליזציה:

<http://ranimsallam.github.io/vis/barChart>



קצת על הויזואליזציה – בציר ה-X יש את השנים בין 1916 עד 2008 בציר ה-Y מספר השירים בדאטה בנושא פוליטיקה בכל שנה.

ההשערה שלנו הייתה שבאזור תקופת מלחמת וייטנאם תהיה עליה במספר השירים בנושא פוליטיקה.

ארצות הברית התערבה במלחמת וייטנאם ב-1964 עד 1975. באמת אפשר לראות עלייה במספר השירים לקראת 1964 ושיא ב-1969 ו-1970.

הוספנו לויזואליזציה את היכולת לעשות "האבר" מעל כל שנה ולראות על איזה שירים מדובר.

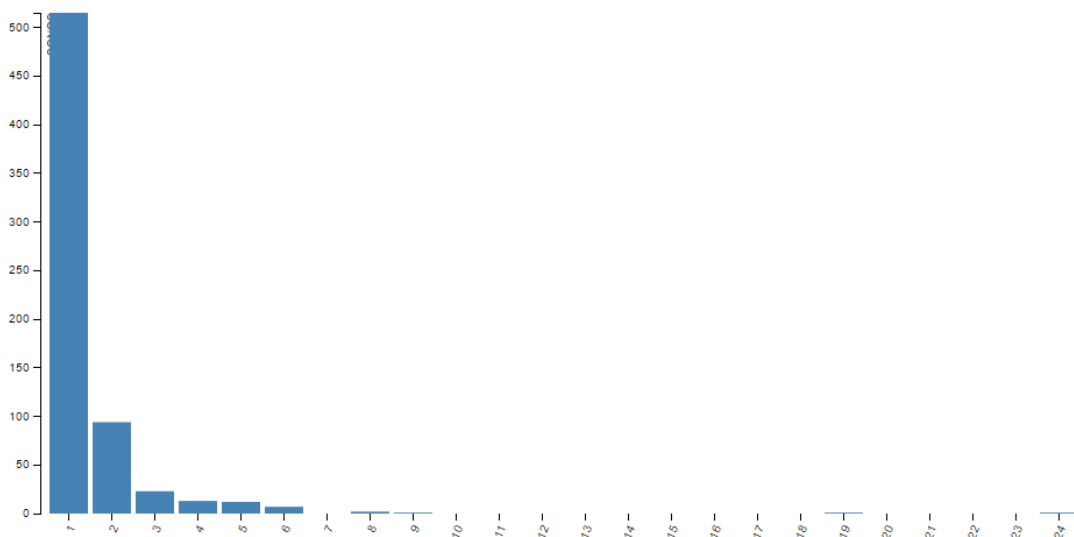
מהסתכלות על השירים באמצעות "Hover" בכל שנה גילינו דבר מעניין שלא חשבנו עליו בהתחלה. בשנות ה-60 וה-70 אכן היו הרבה שירים על מלחמת ויטנאם, אך גם חלק מהשירים היו על מלחמה אחרת לגמרי שלא חשבנו עליה – מלחמת הזכויות של האפרו-אמריקאים בארצות הברית. היה מאוד מעניין לגלות איך שתי התופעות השפיעו וקיבלו ביטוי בדאטה ובוויזואליזציה. שמחנו לראות איך הויזואליזציה שייצרנו "גילתה" לנו דבר חדש שלא חשבנו עליו מראש.

שאלה נוספת – "מי הם האמנים הבולטים בדאטה"

אז הדאטה מחולק ככה שאמן מסויים יכול להיות מוזכר במספר כניסות בתוך הדאטה. ומעניין לדעת האם האמנים המוזכרים ביותר בדאטה באמת "נחשבים" אמנים משפיעים בתרבות האמריקאית.

עבדנו קצת על עיבוד הדאטה כך שנוכל לראות כמה שירים יש לכל אמן, וייצרנו את הויזואליזציה הבאה (זאת עוד דוגמא לויזואליזציה לא טובה שלמדנו ממנה):

<http://ranimsallam.github.io/vis/index.html>



מה אפשר להבין מהויזואליזציה הזאת? כלום! אנחנו רואים שיש המון אמנים עם שיר אחד, פחות עם שניים וכך הלאה.

דווקא הקווים החשובים בקושי נראים, הקווים במספר 24 ובמספר 19 – בוב דילן והביטלס בהתאמה.

מייד היה ברור לנו שאנחנו מחפשים ויזואליזציה חדשה. כמה מילים על זאת בכל מקרה.

אז מהסתכלות על הדאטה ראינו שבמקום הראשון עם 24 שירים נמצא בוב דילן ובמקום השני נמצאים הביטלס עם 19 שירים בדאטה. העובדות האלו באמת אישרו לנו את ההשערה שקיימת קורולציה בין הדאטה לאמנים משפיעים בתרבות האמריקאית.

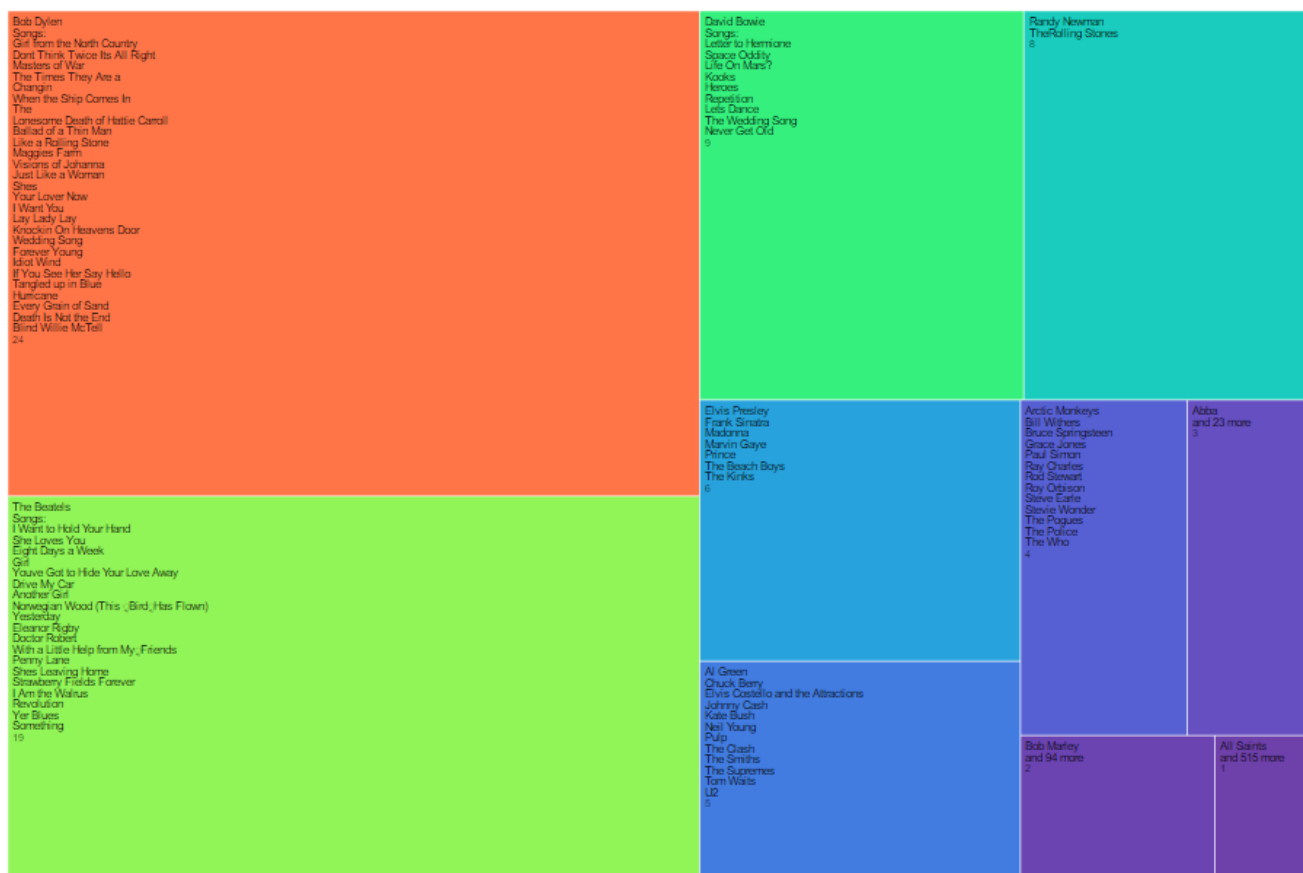
הבנו שאנחנו רוצים להדגיש לצופה מי הם אמנים הבולטים, שממבט ראשון מייד יהיה ברור שזה בוב דילן והביטלס.

חיפשנו ויזואליזציה שתתאים לפרמטרים שלנו. מצאנו כמה אפשרויות.

Tree Map

עיבדנו את הדאטה בצורה של עץ, כך שכל מסלול מהשורש עד העלה מתאר את האמן והעלה מכיל את מספר השירים שיש לו בדאטה.

<http://ranimsallam.github.io/vis/VisioNew>



Tree Map

בויזאוליזציה הזאת ככל שיש יותר שירים לאמן כך גם המידע עליו גדל, עבור האמנים עם מספר שירים הגדול מ-8 הוספנו גם את רשימת השירים הטובים שלו (שמוזכרים בדאטה). כך הצופה מייד רואה מי הם האמנים הבולטים בדאטה ושיריהם.

בנוסף הצבעים בדאטה מעידים כך שכל הצבע חם יותר כך מספר השירים עבור האומן גדול יותר.

החסרונות של הויזאוליזציה היא שהיא קצת משעממת – מרובעים, חסר אסטטיקה ואין משהו שמושך את העין לחקור את הדאטה.

המשכנו לחפש דרך נוספת לייצג את הדאטה שהצופה יכול להסיק ממנה יותר פרטים.

עוד דבר שבעייתי בויזאוליזציה הוא המעבר ברמת הפירוט בכל ריבוע. אצל האמנים עם למעלה מ-8 שירים הוספנו לריבוע את רשימת השירים. ולאמנים עם פחות, רק רשמנו לאיזה קבוצה הם משתייכים. כלומר לדוגמא: בריבוע הראשון אנחנו רואים את הביטלס ואז רשימת שירים, אבל אם נסתכל על אחד הריבועים הכחולים יותר, אז שם יש שמות זמרים בלי השירים. המעבר הזה יכול להיות בעייתי לצופה, שלא יבין את השינוי שהתרחש בתוך הויזאוליזציה. הצופה עלול להתבלבל בין שירים של זמרים לשמות של להקות ולחשוב ששיר מסוים הוא להקה כלשהי.

בכל זאת למרות הקושי הזה רצינו לשמור על המבנה הזה של הדאטה מיכיון שרצינו לתת לצופה במבט אחד את כל המידע הכי חשוב מהדטה שזה בשבילנו – מי הזמרים הכי טובים ומה שיריהם.

Circle Packing

בויזואליזציה הזאת רצינו לספר גם על איזה נושא האמן נטע יותר לכתוב עליו.

<http://ranimsallam.github.io/vis/circleViso>



Top Artist use Circle Packing visualization

In this visualization we are showing the best artist with most entries in the data. for the Artist with more than 8 songs you can check out how many songs he/she has from each theme. the bigger the circle mean the artist has more songs in the data. The closer the circle to the center of the page the bigger it is. feel free to hover over the page to find out amazing artists.

בויזואליזציה הזאת שני פרמטרים מרכזיים מייצגים את הדאטה.

ראשית גודל העיגול – גודל העיגול מתאר את מספר השירים שיש לאמן בדאטה. עבור אמנים עם מספר שירים גדול מ-8 (בתוך הדאטה) מיינו את השירים על פי הנושאים. כך אפשר לראות עבור כל אמן באיזה נושא יותר התרכז בכתביה שלו.

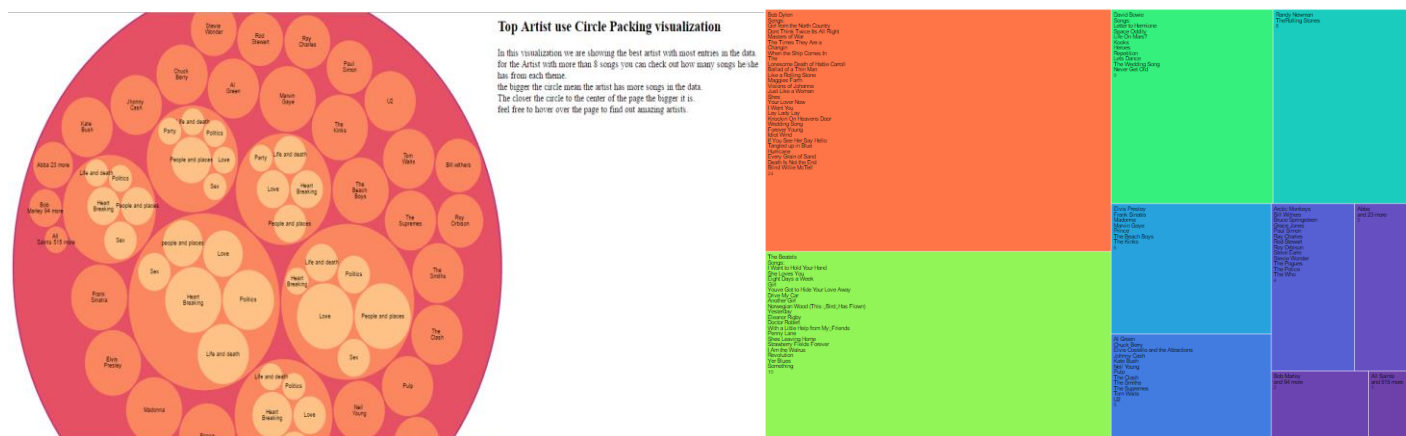
שנית – מיקום העיגול, ככל שהעיגול נמצא במרכז הדף כך מספר השירים בו גדול יותר.

עיגולים בצבע הבהיר מעידים על חלוקת השירים של אמן מסוים לקטגוריות, עיגולים בצבע הכהה מעידים על מספר השירים הכולל עבור האמן המסוים.

ניתן לעשות "האבר" עם העכבר בשביל לראות עבור כל עיגול מי האמן המדובר וכמה שירים יש בעיגול.

Tree Map VS Circle Packing

אם אנחנו משווים בין שתי הויזואליזציות האחרונות, אפשר לדבר על ההיתרנות והחסרונות של כל ויזואליזציה.



נתחיל במבט מהיר על שתי הויזואליזציות, זה ברור שהויזואליזצית העיגולים נראית הרבה יותר טוב והרבה יותר מעניינת. בנוסף בויזואליזצית העיגולים ניתן להסיק יותר מידע על הדאטה, כך שגם נקבל את מספר השירים יחסית לשאר האמנים וגם את חלוקת השירים שלו לקטגוריות השונות, לדוגמא ניתן לראות בצורה יחסית פשוטה שה"ביטלס" נהגו לכתוב הרבה שירי אהבה יחסית לשירים על פוליטיקה.

חסרון מרכזי של ה"עיגולים" יחסית ל"מפה" היא שבמפה יכלנו להוסיף עבור האמנים המרכזיים את רשימת השירים שלהם, כך צופה יכול מתוך הויזואליזציה לקבל מידע עבור איזה שירים מתוך ה-1000 הוא באמת חייב לשמוע!

חסרון אחר שקיים ב"עיגולים" היא הקושי להשוות בין אמנים עם 5 שירים לאמנים עם 4 או בין 5-6, ניסינו לפתור זאת על ידי פרמטר המיקום של עיגול ביחס למרכז הדף והוספת האבר מעליו שנותן את מספר השירים.

לעומת זאת ב"מפה" מאוד קל לראות לאיזה קבוצת זמרים יש מספר שירים גדול יותר (על פי גודל הריבוע).

סיכום

בפרויקט זה יצא לנו להתנסות ביצירת ויזואליזציות שונות עבור דאטה מסוים. למדנו שעבור שאלה מסויימת צריך להתאים ויזואליזציה מתאימה, כך לדוגמא מספר השירים על פוליטיקה בכל שנה התאים "bar chart" אבל עבור מציאת האמנים החשובים ביותר "bar chart" היה הויזואליזציה הגרועה ביותר. למדנו שעל מנת לתת לצופה "להבין משהו" מהויזואליזציה היא חייבת להיות ברורה, בהירה, פשוטה ומכילה שינוי או מגמה שהצופה יכול לעקוב אחריה, רצוי שגם תהיה מושכת את העין שצופה ירצה לחקור אותה. כך לדוגמא בויזואליזציית העיגולים מגמת גודל העיגול משתנה, ב"מפה" גודל הריבוע וב-bar chart השינוי לאורך הזמן של גודל הריבוע. שינויים אלו הם מה שנותנים לצופה להבין את הויזואליזציה ומעבירים את המסר שאליו היוצרים התכוונו להעביר. למדנו להכיר את D3 ויתרונותיו – יכולת לעבור על כל הדאטה במספר שורות קטן יחסית, להוסיף מאפיינים ומבנים גרפים בצורה מאוד פשוטה. ובכללי למדנו להעריך ולבקר ויזואליזציות שונות.