

ויזואליזציה של מידע- סמסטר ב' 2017  
מרצה: ד"ר פיטר בק

"A picture is worth a  
thousand words..."

מגישים:

\*ענבר כהן- ת.ז: 208637769.

Email: [inbar2859@gmail.com](mailto:inbar2859@gmail.com)

\*שגיא פנצל- ת.ז: 314625328.

Email: [sagipend@gmail.com](mailto:sagipend@gmail.com)

# אולימפיאדות קיץ 1896-2012

במסגרת הפרויקט בקורס "ויזואליזציה של מידע", יצרנו מספר ויזואליזציות. הויזואליזציות שלנו מסתכלות אחורה, אל עבר 116 שנות אולימפיאדה היסטוריות, אשר חודשו לראשונה באתונה בשנת 1896. במסגרת הפרויקט אנחנו סוקרים את השתלשלות ההישגים האולימפיים של מדינות העולם בשנים אלו ומספקים סקירה כללית של המדינות המשתתפות, המדליות שהושגו על ידן וענפי הספורט בהיבטים שונים ומגוונים. כל זאת נעשה תוך כוונה למשוך את עינו של הצופה לממצאים הרבים אשר ניתן להסיק מהויזואליזציות שלנו.

## **הרעיון:**

בתחילת הסמסטר חשבנו על לעבוד עם DataSet של המשחקים האולימפיים. ראשית, זהו אירוע הספורט השני בגודלו בעולם (אחרי גביע העולם בכדורגל) וחשבנו שסטטיסטיקות על אירוע בסדר גודל שכזה יכולות להיות מרתקות. שנית, כשישבנו יחד וחשבנו על שאלות ונושאים שבהם אנו רוצים לסקר את Datan שלנו, גילינו שיש הרבה מאוד מאפיינים וכיוונים מעניינים במשחקים האולימפיים שיכולים לשמש אותנו בבסיס ליצירת ויזואליזציות טובות ואיכותיות. תחילה, אחת מהשאלות עליהן רצינו לענות היתה מהם הגילאים השכיחים ביותר לקבלת מדליות בכל אחד מסוגי הספורט השונים. כשהתחלנו לעבוד על הפרויקט ולחפש DataSet מתאים מצאנו קובץ csv שמכיל מידע על האולימפיאדות בשנים 1896-2012 (WHAT) אך לצערנו המידע הנתון על גילאי הספורטאים לא היה מלא (במיוחד בתחום האתלטיקה) לכן נאלצנו לוותר על שאלות הקשורות בגילאי הספורטאים. לא היינו מוטרדים מכך מאחר והיו לנו בכיס עוד עשרות שאלות מעניינות (WHY).

## **Get Started:**

יש להריץ את הקובץ index.html שנמצא בתיקיית הגשת הפרויקט ולפתוח אותו באמצעות Firefox.

תחילה, browser ישאל אם ברצונך להתקין את ה-plug-in בשם: FireShot (במקרה שהוא לא הותקן אצלך קודם לכן). plug-in זה נחוץ לנו על מנת שהכפתור עם הרקע הסגול "SNAPSHOT" שנמצא בNavbar יהיה פונקציונלי. מטרתו, צילום המסך בצורה נוחה וידידותית למשתמש ושמירת התמונה כקובץ png. אם אינך מעוניין להתקין plug-in זה אפשר לצאת מהModal ויפתח עמוד הבית הראשי של הפרויקט שלנו. את המעבר בין הטאבים השונים, בין ויזואליזציה אחת לשנייה, ניתן לעשות באמצעות Navbar שבנוי כך:

# Project Documentation

## ?WHAT

החלטנו להשתמש בData מ Kaggle (לינק: <https://www.kaggle.com/the-guardian/olympic-games>) בתור הDataSet שלנו. האתר והData היו מאורגנים היטב וניתנים להורדה בקבצי CSV. לאחר כמה טסטים והורדת כפילויות קיבלנו את הקובץ הסופי אשר מכיל מידע על משחקי אולימפיאדות הקיץ שנערכו בשנים 1896-2012:

## Data type

טבלה בעלת 31,159 שורות.

## Attributes

-Year- השנה בה התרחשה האולימפיאדה.

-City- העיר אשר אירחה את האולימפיאדה.

-Sport- סוג הספורט (למשל אתלטיקה).

-Discipline- ענף הספורט (למשל ריצה).

-Athlete- שם מלא של הספורטאי המתחרה.

-Country- המדינה ממנה מגיע הספורטאי המתחרה (מיוצג בקוד IOC - International Olympic Committee, קוד בעל 3 אותיות לייצוג מדינה).

-Gender- מין הספורטאי המתחרה.

-Event- אירוע ספציפי (מיקצה).

-Medal- סוג המדליה שבה זכה הספורטאי.

## Ordering Direction

רציף - Sequential - אולימפיאדה כל 4 שנים למעט מספר אולימפיאדות חריגות שבוטלו, מפורט בהערות (\*). על כך בהרחבה.

## Attribute type

קטגוריות - Categorical - כל עמודה נותנת מידע שונה לגבי קטגוריה מסוימת.

## Dataset Availability

סטטי - Static.



2 שורות ראשונות לדוגמה מהDataSet שלנו:

	year	city	sport	discipline	athlete	country	gender	event	medal
▶	1896	Athens	Aquatics	Swimming	ANDREOU, Joannis	GRE	Men	1200M Freestyle	Silver
	1896	Athens	Aquatics	Swimming	CHASAPIS, Spiridon	GRE	Men	100M Freestyle For Sailors	Silver

\*הערות:

- בכל התחרויות נשים וגברים מתחרים בנפרד מלבד בתחום רכיבת סוסים.
- במהלך השנים התבטלו 3 אולימפיאדות עקב מלחמת העולם הראשונה (אולימפיאדת 1920) ועקב מלחמת העולם השנייה (אולימפיאדות 1940, 1944).
- באולימפיאדת 1980 התחרו 80 משלחות בלבד השתתפו במשחקים, המספר הנמוך ביותר מאז אולימפיאדת מלבורן (1956). מדינות רבות הצטרפו לחרם שהובילה ארצות הברית על המשחקים בעקבות הפלישה הסובייטית לאפגניסטן, ובהן סין העממית (אשר הייתה אמורה להשתתף בפעם הראשונה), יפן, ישראל קנדה ועוד.

#### ?WHY

- **השוואה** בניסיון למצוא **מגמה** בין מספר הנשים הזוכות במדליות לאורך השנים.
- **השוואה** בין מספר המדליסטיות לבין מספר המדליסטים ולגלות את הדמיון וההבדלים (**similarity**) ביניהם.
- **הצגת התפלגות** המדליות בקרב מדינות שונות:
  - לפי סוג הספורט
  - לפי סוג המדליה (זהב, כסף וארד)
- **סיכום וסקירה של התפלגות** סוגי הספורט ששוחקו במשחקים האולימפיים.
- **השוואה** בין מדינות שונות לפי כמות המדליות ועל ידי כך לחקור את **ההתפלגות** של המדליות בקרב מדינות העולם.

## HOW?

עבור כל אחת מהויזואליזציות נדרש Data pre-processing אשר כלל הרצת שאילתות על הדטה המקורי. כעת נפרט את שלבי העבודה והניתוח עבור כל אחת מהויזואליזציות לפי הסדר שבו הן יצרנו אותן:

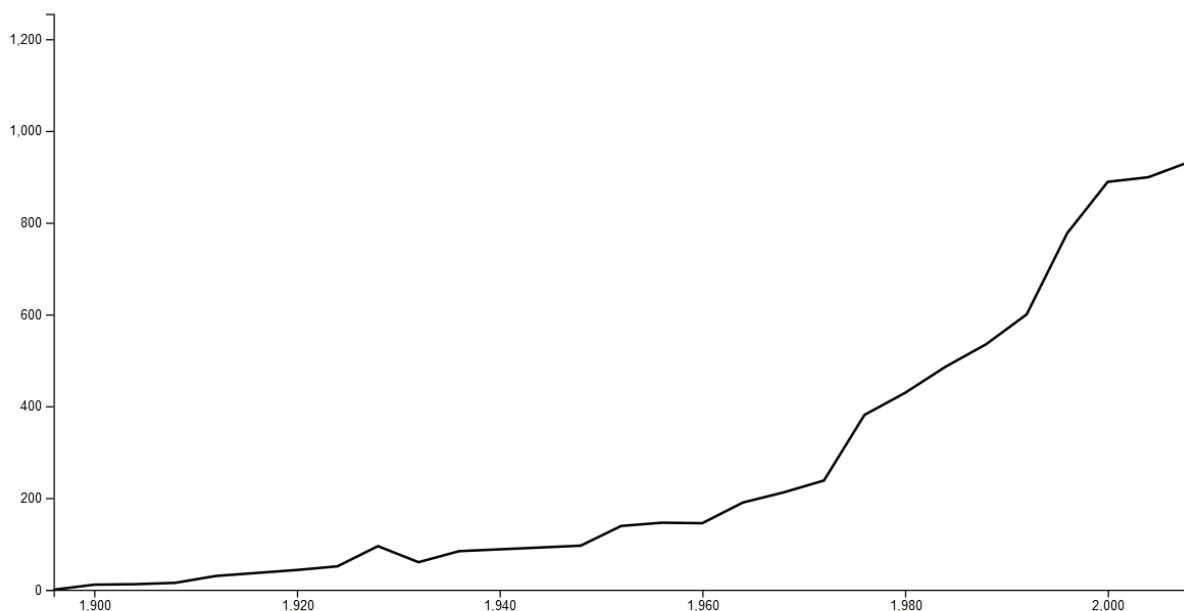
### Difference Chart (1

לפני שניגשנו ליצירת הויזואליזציה הראשונה, קראנו שנשים השתתפו לראשונה בתחרויות האולימפיות רק משנת 1900 ומספר הנשים בשנים הראשונות אשר זכו במדליה היה מזערי (11 נשים ב1900, 12 נשים ב1904).

עניין אותנו מאוד לבחון את מגמת השינוי בכמות הנשים הזוכות במדליות, ולהשוות זאת לכמות הגברים הזוכים במדליות ולנסות להגיע למסקנות.

ניתחנו את אתנת Data המלא באמצעות אחת מהפונקציות של d3 הלא היא d3.nest ובאופן זה יצרנו מערך מקונן עבור המדליסטיות(אותן נשים שהתחרו וזכו במדליה) המורכב ממערכים הבנויים בצורה הבנויים: [שנה , מספר מדליות].

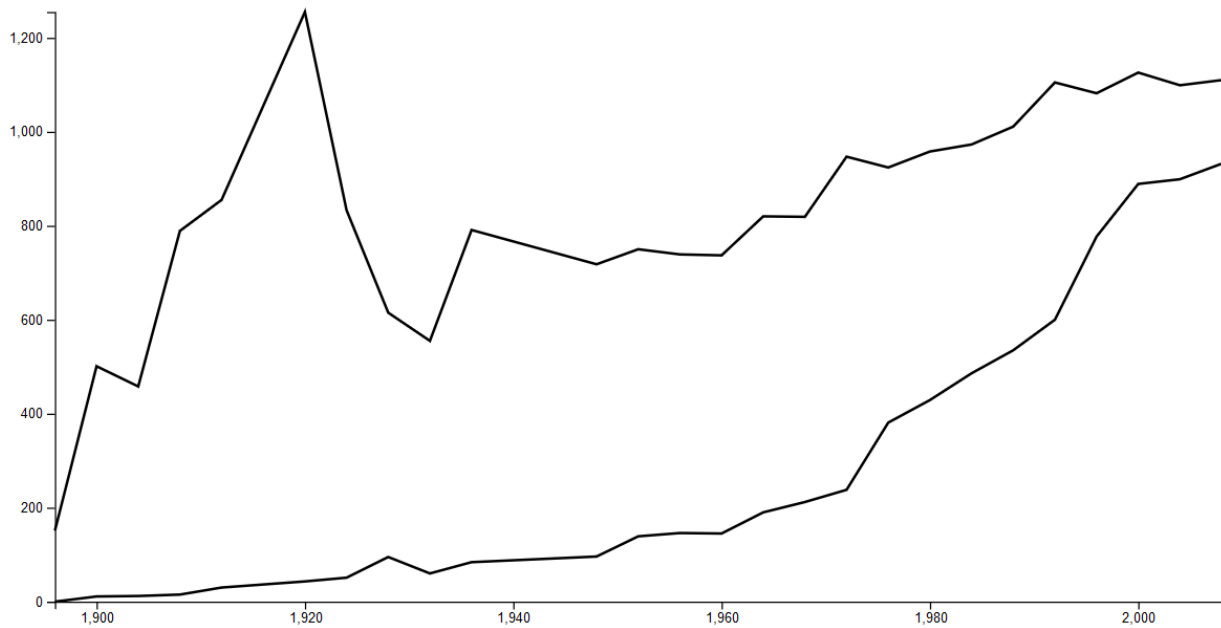
לאחר שקלול נתוני הנשים הגענו לגרף הבא:



ניתן להסיק מהגרף כי מספר הנשים הזוכות במדליות אולימפיות לאורך השנים עולה בעלייה כמעט מתמדת וחדה.

בהמשך, רצינו לאפיין בצורה הברורה ביותר את הפער הקיים בין כמות המדליות אשר נשים זוכות לבין כמות המדליות אשר גברים זוכים בהם.

לכן באותו אופן שיצרנו מערך מקונן עבור המדליסטיות, יצרנו עתה מערך מקונן גם עבור המדליסטים(אותם גברים שהתחרו וזכו במדליה) והוספנו line עבור הגברים המדליסטים. קיבלנו את התוצאה הבאה:

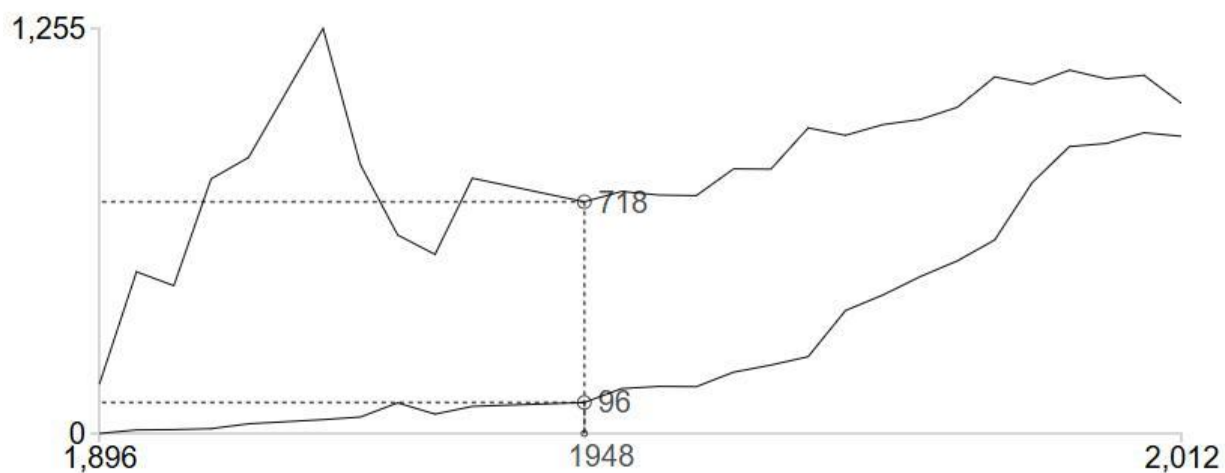


והסקנו מיידית מספר מסקנות חותכות:

- תמיד מספר הגברים הזוכים במדליות גדול יותר ממספר הנשים הזוכות במדליות.
- לאורך השנים ניתן לראות כיצד הפער בין הנשים לגברים מצטמצם באופן רציף ואפשר לעשות מעין "תחזית עתידית" ולשער שהפער רק ילך ויצטמצם.
- בשנת 1932 חלה ירידה חדה במגמת המדליות במספר הגברים הזוכים באולימפיאדה, לאחר שחקרנו את הנושא אנחנו יכולים לשער שזה בעקבות ה world wide recession.

### Evaluation

בכדי לשפר את הסתכלות המשתמש על הויזואליזציה הוספנו את האפשרות הבאה:  
כשהמשתמש מצביע על נקודה בתרשים מופיעים ערך הא וצרכי ה- $\gamma$  של הנקודה.



בזכות אפשרות זו המשתמש יוכל באופן מהיר ונוח יותר להסיק מסקנות אודות הפער במספר המדליות בין גברים לנשים.

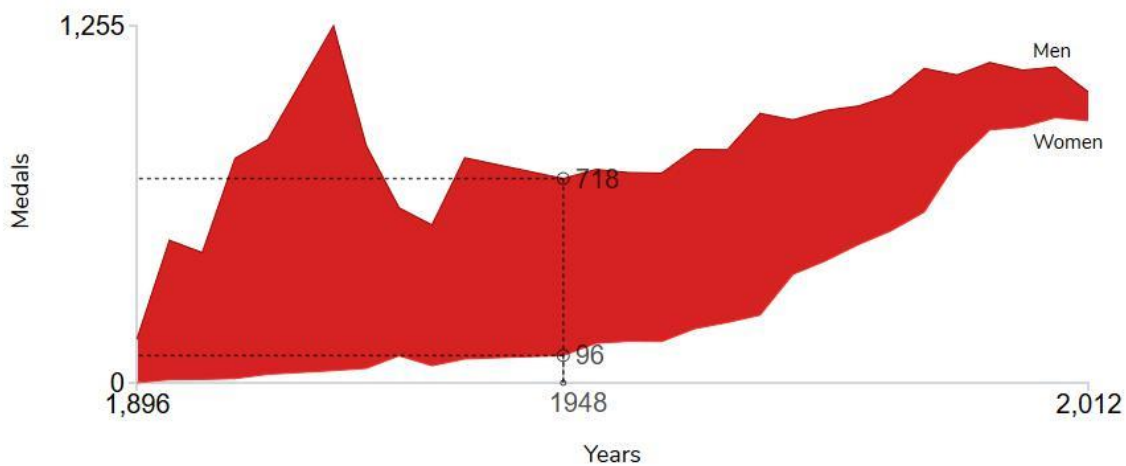
כך כאשר משתמש מתבונן בויזואליזציה הוא יוכל לקבל ממנה יותר אינפורמציה ולהתרשם מהמגמות המשתנות של כל Line בנפרד ושל המרחק בין Lines.

בהמשך, מילאנו בצבע את השטח שבין ה-Lines. באמצעות פעולה זו הדגשנו עוד יותר את הפער בין נשים לגברים מבחינת מספר מדליות בתחרויות האולימפיות.

כעת המשתמש יכול להגיע למסקנות תוך רגעים אחדים של הסתכלות בויזואליזציה זו.

התוצאה הסופית הינה:

Trends Of Medals - Men vs. Women



## :Multi Area Chart (2

בויזואליזציה זאת רצינו לנתח את ההבדלים בכמות המדליות בין מדינה למדינה לאורך השנים.

חיפשנו תרשימים שיכולים להתאים לנו ומצאנו באחת ההרצאות את הסוג הזה.

בחרנו אותו מכיוון שהוא מתאים לנו למטרת השוואה בין כמות המדליות שמשיגה כל מדינה והסקת מסקנות

לגבי מגמות הזכייה במדליות בקרב המדינות.

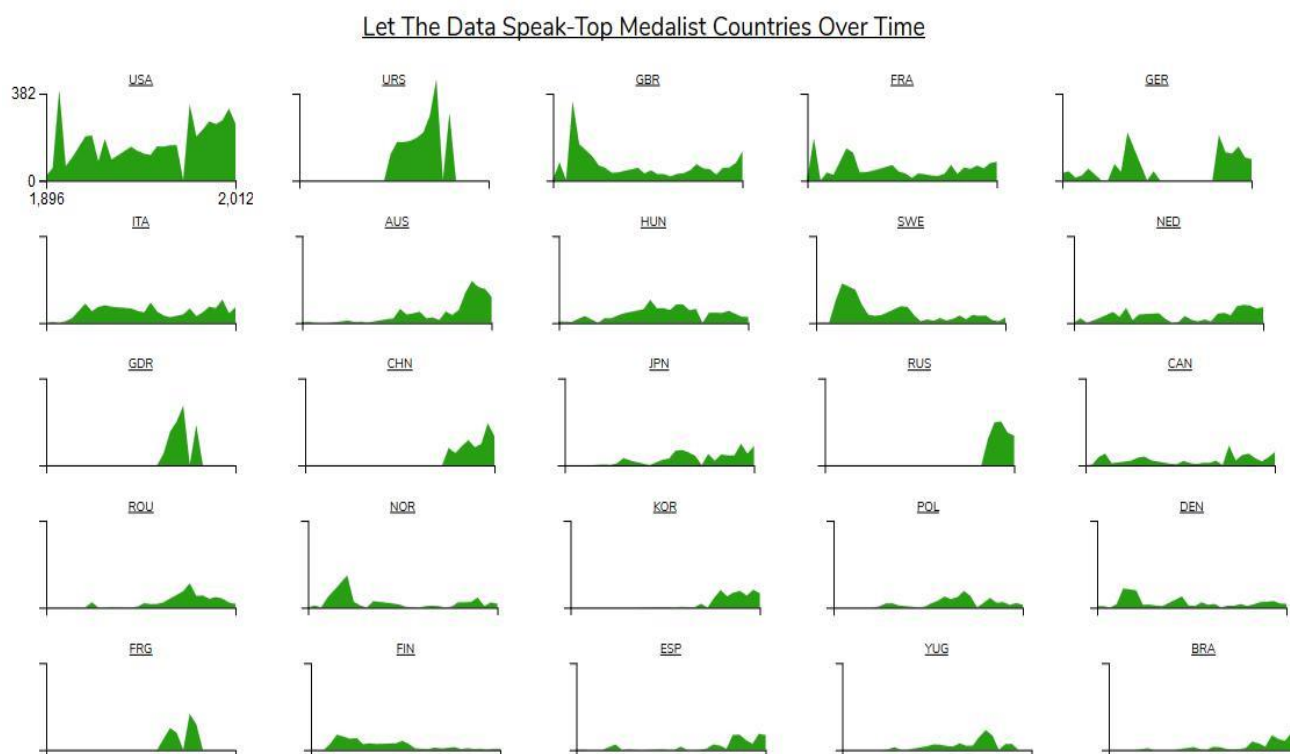
תחילה, חישבנו את הערכים הרלוונטיים עבור כל אחת מהמדינות המדליסטיות וקיבלנו את התוצאה הבאה:





לא היינו מרוצים מהתוצאה משתי סיבות. הראשונה, קשה להבחין בהבדלים בין מדינה למדינה (אולי למעט המדינות המובילות, שבהן כן ניתן לראות "כתמי צבע" מסויימים). שנית, תרשימים אלו תופסים את רוב תצוגת המסך והיא לא קומפקטית ביחס למידע אשר ניתן להסיק ממנה ולכן, בהשראת ההרצאות צמצמנו את מספר התרשימים ל-25. כלומר, אנו מציגים את 25 המדינות המובילות (מבחינת מספר המדליות).

לאחר השינוי והוספת צבעים קיבלנו את התוצאה הבאה:



ניתן לראות בויזואליזציה זו קודם כל את המדינות המובילות מבחינת מספר המדליות שהן השיגו. שנית, ניתן לראות כי קיימות מדינות (לדוגמה: סין-CHN, רוסיה-RUS, וגרמניה המזרחית-GDR) שאף על פי שהתחילו להשתתף בשלב יחסית מאוחר בתחרויות האולימפיות הם הצליחו להגיע לרשימת ה-25 מדינות המובילות.

דבר נוסף שניתן לראות הוא מדינות (לדוגמה: בריטניה-GBR ויפן-JPN) שהיו להן תקופות מצליחות יותר ותקופות מצליחות פחות וגם הן הצליחו להיכנס לרשימת ה-25 המובילות.

זאת אומרת, באמצעות ויזואליזציה זו ניתן לקבל מידע רב אודות מגמות הזכייה במדליות בקרב ה-25 המדינות המובילות.

הערה חשובה- ניתן לאתר בויזואליזציה זו שנים חריגות בהם היו נפילות משמעותיות במספר המדליות: 1916-: מלחמת העולם הראשונה.

1940+1944-: מלחמת העולם השנייה.

1980-: השתתפו 80 מדינות בלבד (הסבר מורחב נמצא בעמוד 2).

גרמניה: גרמניה השתתפה במשחקים האולימפיים שנערכו מאז חידושם בעת החדשה, באולימפיאדת אתונה (1896) ועד מלחמת העולם השנייה, למעט שתי היעדרויות: האולימפיאדות שנערכו לאחר מלחמת העולם הראשונה נערכו בבליה ובצרפת, שהיו חלק ממדינות ההסכמה, וגרמניה לא הוזמנה אליהן. לאחר פירוקה של גרמניה ועד לאיחודה מחדש ב-1990 לא התחרתה במשחקים. האולימפיאדה הראשונה שגרמניה המאוחדת מחדש השתתפה בה, התקיימה ב-1992 בברצלונה.

### **Evaluation:**

כאשר המשתמש יסתכל על הויזואליזציה הזו הדבר הראשון שימשוך לו את העין זו ארה"ב, המדינה הזוכה במספר המדליות הגבוה ביותר. ולאחר מכן אנחנו משערים שענינו יוטו לעבר הויזואליזציה הבאה שמימין לה וכן הלאה.

בסופו של דבר, המשתמש יקבל בין רגע מידע אודות 116 שנים של תחרויות אולימפיות. המידע הזה מגיע בצורה מדורגת, המדינות מסודרות מהזוכה בהכי הרבה מדליות עד לזוכה בהכי מעט מדליות (משמאל לימין) מה שמקל על המשתמש בהסקת מסקנות אודות ההבדלים שקיימים בין מדינה למדינה, למשל, מבחינת התקופות בהן הרוויחו יותר מדליות והתקופות בהם הרוויחו פחות מדליות. כמו כן ניתן באופן מיידי לראות באילו שנים כל מדינה הרוויחה את עיקר המדליות שלה. כלומר, לויזואליזציה זו יש המון נקודות הסתכלות שונות (השוואה בין מדינות שונות, חקירת ההתפלגות של המדליות לאורך השנים במדינה ספציפית או בין מדינות)

### Bubbles Chart (3)

את דיאגרמה זו יצרנו בעקבות מספר מטרות. ראשית, רצינו ליצור ויזואליזציה שמשקפת בצורה המהירה ביותר את היחסים בין מספרי המדליות אשר זכו מדינות העולם ושיהיה ניתן לשלוף, כהרף עין, את ה-5 מדינות הטובות ביותר, בהתחשבות במספר המדליות בלבד. בנוסף, נאמר כי התלהבנו מאוד מהרעיון שעלה לנו והוא להפוך את הדיאגרמה הזאת למעין "חלון" לויזואליזציה הבאה אשר מרחיבה אותה (נפרט בהמשך). כמו כן, זאת ויזואליזציה אשר מאוד מושכת את העין של המשתמש ומאפשרת לו להחליט באיזה מדינה מעניין אותו יותר להתעמק.

חישוב הגודל של כל בועה:

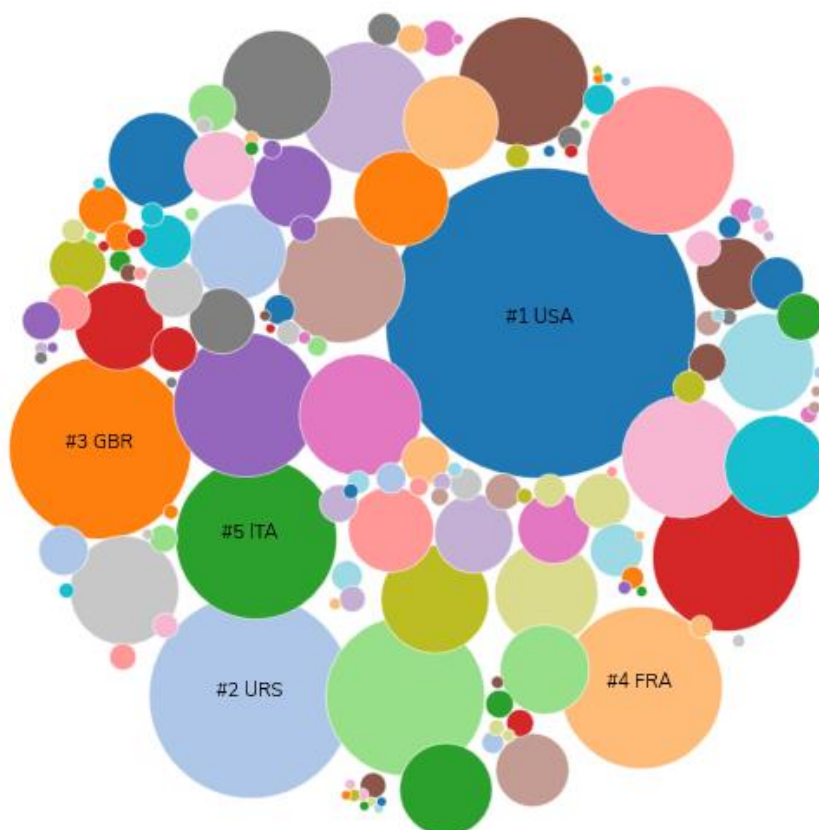
לצורך זה השתמשנו ב-Ranking system שמונהגת על ידי ה-IOC (International Olympic Committee), השיטה הפופולרית ביותר לחישוב הניקוד עבור כל מדינה הוצעה ב-2004 והיא הפורמולה הבאה: (3:2:1) Fibonacci weighted point system. כלומר:

$$\text{Total} = \#gold * 3 + \#silver * 2 + \#bronze * 1$$

כאשר Total=ניקוד המדינה בדירוג המדליות הכללי.

לאחר הוספת צבעים קיבלנו את התוצאה הזו:

### Bubbles



אנו מרוצים ממנה מהסיבה שהיא עונה בדיוק על המטרה שלנו בויזואליזציה זו.

### האינטראקציה של User יכול לבצע:

כאשר יצביע עם העכבר על אחת מהבועות יופיע Tooltip עם הנתונים הבאים:  
-שם המדינה  
-מספר המדליות שהיא זכתה בהן.

לדוגמא, כשנעביר את העכבר מעל לUSA:



### :Evaluation

המשתמש יכול לראות במהירות מיהן 5 המדינות המובילות (מבחינת מספר מדליות שחושב לפי הנוסחא), אך במידה והוא רוצה לבחור מדינה ספציפית, במידה ומספר המדליות שלה לא גדול במיוחד, הוא ימצא את משימה זו כלא פשוטה בכלל.

עם זאת, ויזואליזציה זו כן עונה על המטרה של מתן מידע כללי לגבי התפלגות המדליות בין מדינות העולם. וניתן באמצעותה להבין עד כמה ההבדל בין כמות המדליות שמקבלת מדינות שונות מטרף.

עכשיו, חשבנו באיזה מידע אנחנו מעוניינים להתרכז כאשר User ילחץ על בועה מסוימת.  
רצינו לחקור את ההתפלגות (Distribution) של כמות המדליות על פני:

1. סוגי הספורט השונים.
2. סוגי מדליות שונות (זהב, כסף, ארד).

(1)

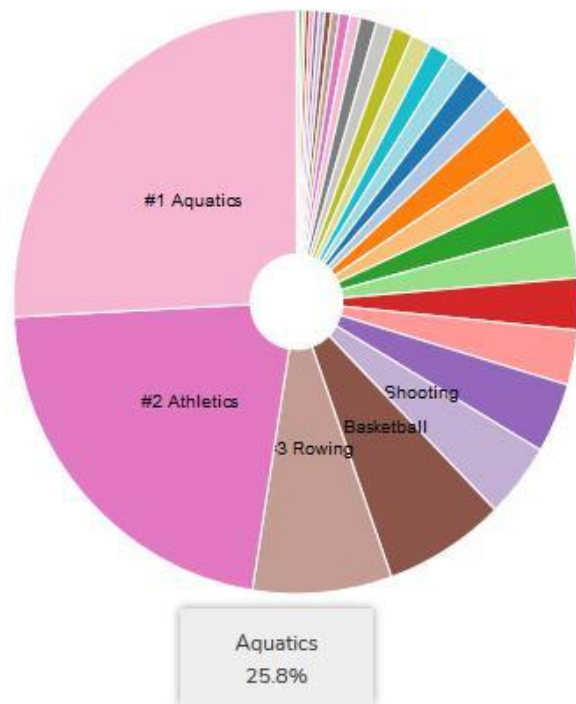
- המעניין ביותר בעינינו הוא להראות באילו סוגי ספורט מדינה מסוימת משיגה את כמות המדליות הגדולה ביותר ובאילו סוגי ספורט פחות. במילים אחרות, סוגי הספורט שבהם מתמחה המדינה.
- מעניין לראות את כמות סוגי ספורט השונים שבהם מדינה מסוימת מדליות.
- מעניין לראות את התפלגות מספר המדליות של המדינה במספר סוגי הספורט השונים שבהם היא זכתה לאורך כל השנים.

ולכן החלטנו לעבוד עם Donut Chart  
הDonut Chart נותן לנו מענה עבור כל אחת מהשאלות האלו.  
הDatan צריך להיות עם מורכב מהAttributes הבאים: סוג הספורט(sport), מספר המדליות אשר זכו בסוג ספורט זה(count).  
כדי ליצור Data המורכב בצורה זאת, הפעלנו על DataSet שלנו שאילתא.

### האינטראקציה של User יכול לבצע:

הוספנו Tooltip שמופיע כאשר User מצביע באמצעות העכבר על ספורט מסוים מתוך חלקי Donut והמידע שנגלה הוא בעל הנתונים הבאים:

-סוג הספורט  
-האחוז שמהוות המדליות בספורט זה מסך כל המדליות של המדינה.



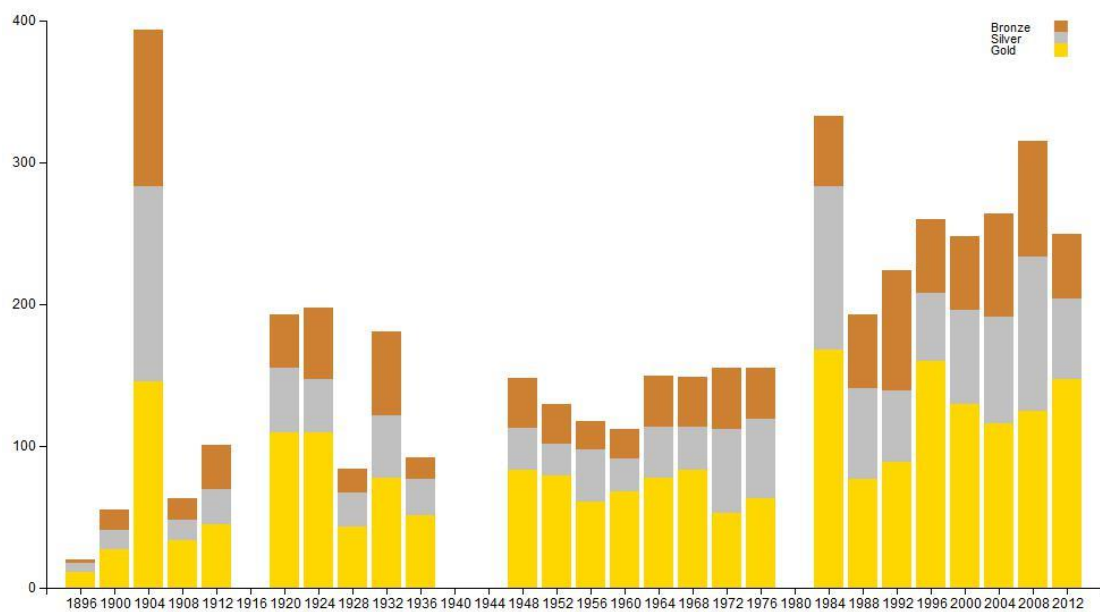
(2)

- עניין אותנו להבחין בין מדליות מסוגים שונים שכל מדינה מקבלת, ההבחנה היא כמובן בין מדליות מהסוגים הבאים: זהב, כסף וארד. רצינו לקבל מידע אודות סוג המדליה הדומיננטית בקרב אותה מדינה, האם היא זוכה לרוב במדליות זהב, כסף או ארד.
- כל זאת הופך למעניין יותר שמסתכלים על הנתונים לאורך השנים.

הגענו למסקנה שאת ה"ל נוכל להציג ב-Stacked Bar Chart, אשר מהווה ויזואליזציה שבאמצעותה ניתן לראות מגמות שינוי לאורך הזמן כפי שלמדנו בהרצאה.

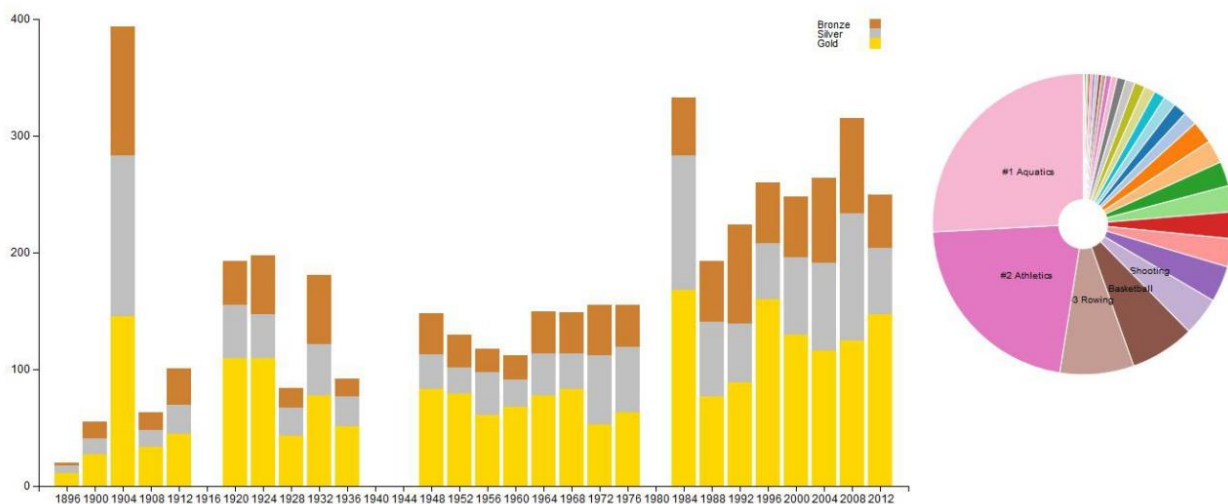
Datan צריך להיות מורכב מהקטגוריות הבאות: שנה, מספר מדליות זהב, מספר מדליות כסף, מספר מדליות ארד.

לשם כך הפעלנו על DataSet שלנו שאילתות. וקיבלנו תוצאה זאת:



חיברנו את שתי הוויזואליזציות האלו (סעיפים 1 ו-2 לעיל) וקישרנו אותן לויזואליזציית Bubbles הנזכרת למעלה. קיבלנו את התוצאה הבאה:

Medals-distribution of USA



תמונה 1

תוצאה זו התקבלה על ידי לחיצה על הבועה של USA בוויזואליזציית Bubbles.

### Evaluation:

Donut Chart •

המשתמש יכול להסיק במהירות מהם חמשת סוגי הספורט המובילים עבור המדינה אותה הוא בחר.

הוא יכול לעבור עם העכבר מעל הקטעים הצבעוניים ב Donut Chart ולראות את הנתונים של הספורט עליו הוא מצביע ב Tooltip שנמצא מתחת ל Donut Chart. אומנם כשמדובר בהרבה סוגי ספורט כמו למשל בארה"ב יש בעיה עם לחפש סוג ספורט ספציפי. אבל, בשביל להשיג את המטרה שלנו החלטנו כי אין צורך בכך ושהוויזואליזציה בדיוק כמו שהיא עכשיו עונה על המטרה

שלנו ומציגה היטב את סוגי הספורט בהם המדינה משיגה הכי הרבה מדליות ואת ההבדל (הפער) בין כמות המדליות אשר היא זוכה בסוגי ספורט שונים.

- Stacked Bar Chart:

כשהמשתמש מתבונן בויזואליזציה הוא יכול במהירות לראות מה הצבע השולט (זהב, כסף או ארד).

- חסרון: לא ניתן לעשות זאת במדינות שאין להם סוג מדליה דומיננטי יותר.

בנוסף המשתמש יכול לראות לכל שנה באיזה סוג מדליה זכו יותר.

אפשר גם לראות את השינוי לאורך המן במסר המדליות מכל סוג.

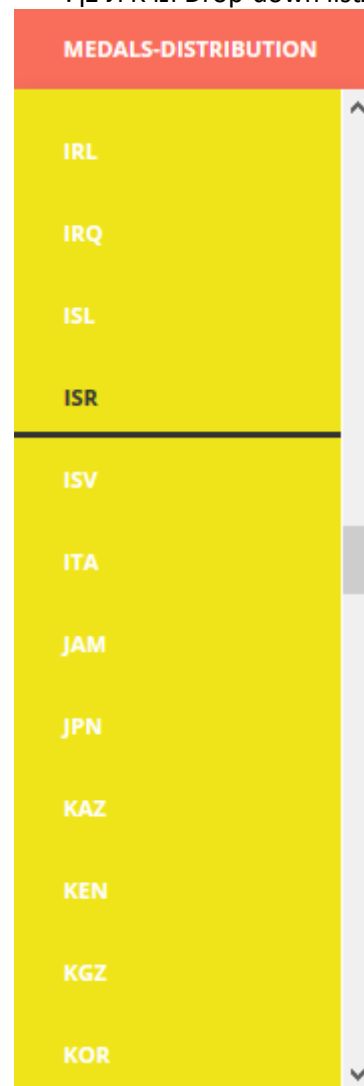
- סדרת שלושת הויזואליזציות הללו מאפשרת למשתמש לבחור באיזו מדינה הוא הכי מתעניין, והוא יכול לעשות זאת תוך שיקול של גודל הבועה (ככל שהבועה גדולה יותר המדינה מובילה בדירוג הכללי של המשחקים האולימפיים). לאחר בחירת המדינה נפתח ל-User חלון שבו הוא יכול להתעמק במדינה בה בחר ולחקור אותה לעומק מבחינת התפלגויות המדליות בין סוגי הספורט וסוגי המדליות.

החיסרון והקושי שבו נתקלנו בויזואליזציה Bubbles הוא שאין ביכולתנו לשבץ את שמות כל המדינות לתוך הבועות מאחר שגודל הבועה אינו מספיק גדול לכך. לכן מוצגים השמות רק עבור חמשת המדינות המובילות בדירוג המדליות הכללי. קושי זה פתרנו באופנים הבאים:

במידה ו-User מעוניין לחפש מדינה ספציפית, הוא יכול לחפש אותה על ידי Tooltip. אך כמובן שזהו אינו הפתרון המיטבי ולכן יצרנו אפשרות ב-Navbar שלנו ל**Medals Distribution** (אשר נפרט עליה בפסקה הבאה)

## Medals Distribution (4)

מאחר ובדיאגרמת Bubbles קשה באופן יחסי לחפש מדינה ספציפית (במיוחד אם מספר המדליות של מדינה זו הוא לא רב) החלטנו להוסיף ל User את האופציה לבחור מדינה מבין רשימת המדינות שמוצגות בד Drop-down list ונראית כך:



בצורה זו User מגיע לעמוד המכיל את ויזואליזציה 1) ו-2) המתוארות לעיל ( Donut Chart, Stacked Bar Chart), ראה: "**תמונה 1**" ויכול לחקור את התפלגות המדליות באותה מדינה.



**כעת רצינו להציג ולנתח מזוויזות שונות את סוגי הספורט השונים שבהם מתחרים הספורטאים במשחקים האולימפיים. לשם כך נעזרנו בויזואליזציות הבאות:**

## **Dots Matrix (5**

רצינו לעשות ויזואליזציה שתהווה סקירה של כל סוגי הספורט בכל השנים ותפיק מידע העונה לשאלות הבאות:

- אילו סוגי ספורט לקחו חלק באולימפיאדה במספר הפעמים הרב ביותר?
- אילו סוגי ספורט לקחו חלק באולימפיאדה במספר פעמים המועט ביותר?
- כמה סוגי ספורט היו בכל אולימפיאדה?
- האם יש סוגי ספורט שלא הוסרו מהמשחקים האולימפיים לעולם?
- מהם המשחקים החדשים ביותר שנוספו למשחקים האולימפיים?

בהתחלה היה לנו קושי בחשיבה על ויזואליזציה טובה שתשקף היטב את מטרתנו ותיתן מענה לכל השאלות הנזכרות למעלה מכיוון שאנחנו מתעסקים עם מידע רב (כלומר, יש ברשותנו הרבה שנות משחקים והרבה סוגים שונים של ענפי ספורט).

רצינו למצוא את הויזואליזציה שתשקף את כל מטרותינו ונוכל להסיק ממנה מסקנות במהירות. דרישה נוספת, רצינו שהויזואליזציה תהיה קומפקטית ככל האפשר.

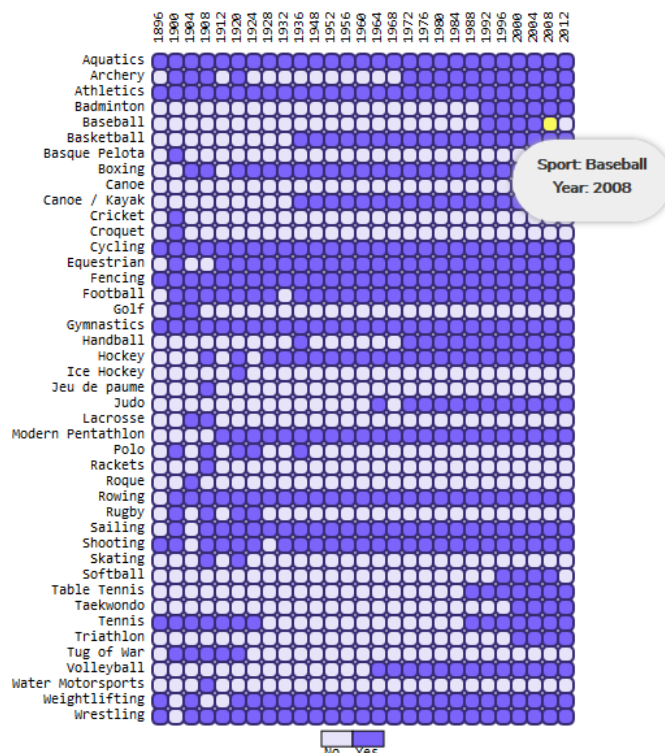
את ה Data שבו השתמשנו עבור ויזואליזציה זו היה מורכב מהקטגוריות הבאות: סוג ספורט, שנה, ערך בינארי(1-אם הספורט תפס חלק מהמשחקים האולימפיים באותה שנה ו-0 אחרת). בכדי להגיע למבנה זה הפעלנו על ה DataSet שלנו שאילתא.

דיאגרמת Dots Matrix משתמשת בעקרון הנלמד בביתה, העקרון הינו Multidimensional scaling והוא מוצג באופן הבא:

לדיאגרמה בעצם קיימים שלושה צירים:

- שנים.
  - סוגי ספורט.
  - התקיים או לא התקיים - מידע אשר אנחנו מסיקים לפי הצבע.
- התוצאה נראית כך:

## Olympic Sports



- דיאגרמה זו מציגה מידע רב אודות ענפי הספורט השונים במשחקים האולימפיים לאורך הזמן:
- קודם כל, ניתן לראות בה עבור כל סוג ספורט באילו שנים הוא תפס חלק מהמשחקים.
  - אפשר לאתר מבין כל סוגי הספורט את סוג הספורט ששוחק הכי הרבה שנים ובעצם לא הפסיק להיות משוחק מעולם (Cycling, Fencing, Gymnastics, Aquatics, Athletics) ואת הספורט ששוחק במספר שנים המועט ביותר (Cricket, Croquet, Jeu de paume, Rackets, Water Motorsports) אשר לקחו חלק באולימפיאדה בודדת).
  - כמו כן ניתן לראות כי קיימים סוגי ספורט ("ניסיוניים") אשר לאחר מספר שנים יצאו מהתחרויות האולימפיות (למשל Tug of War) וגם סוגי ספורט כמו Tennis ששוחקו ברצף עד 1924 ולאחר מכן הופסקו כל התחרויות בענף ספורט זה עד שחודשו ב-1988. כמו כן, ניתן לשים לב לתופעה מעניינת שמתרחשת בענף הספורט Polo - לא קיימת תקופה רציפה של זמן שענף זה תפס חלק מהמשחקים.
  - בנוסף, בהסתכלות על הויזואליזציה משמאל לימין, כלומר, משנות המשחקים הראשונות ועד לאחרונות, ניתן לראות בצורה מובהקת כי בשנים הראשונות היו הרבה פחות סוגי ספורט שבהם התחרו במשחקים האולימפיים מאשר בשנים האחרונות. (ייתכן שנתון זה נובע מהעובדה שמספר המדינות המשתתפות באולימפיאדה הלך וגדל עם השנים וכך גם תחומי הספורט מתרחבים).

### האינטראקציה של User יכול לבצע:

ממשק אינטראקטיבי ונוח ל-User: כפי שניתן לראות הוספנו Tooltip אשר מופיע כאשר העכבר עובר מעל לנקודה בצבע הכהה(כלומר, בשנה שבה הספורט כן התקיים כחלק מהמשחקים). הנקודה מעליה עבר הופכת לצבע צהוב וב- Tooltip מוצגים ערך השורה והעמודה של הנקודה (סוג הספורט והשנה):



בצורה זו יותר נוח ל־User להסתכל על הויזואליזציה מכיוון שהוא אינו צריך לחפש מה ערך הנקודה בציר הא ומה ערכה בציר ה־y (מבין עשרות הא־ים וה־י־ים).

### **Evaluation**

אנחנו מאוד מרוצים מהתוצאה ובמיוחד מגודלה הקומפקטי וכמות המסקנות שניתן להסיק ממנה. לדעתנו הצלחנו באמצעותה לשקף בדיוק את מה שרצינו. כאשר המשתמש יסתכל על הויזואליזציה הוא יוכל להסיק במהירות מסקנות במהירות אודות סוגי הספורט אשר לקחו חלק באולימפיאדות השונות. זאת בזכות contrast הקיים בין שני הצבעים הבודדים שהשתמשנו בהם בויזואליזציה ובעזרת ה־Tooltip.

## Chord Chart (6)

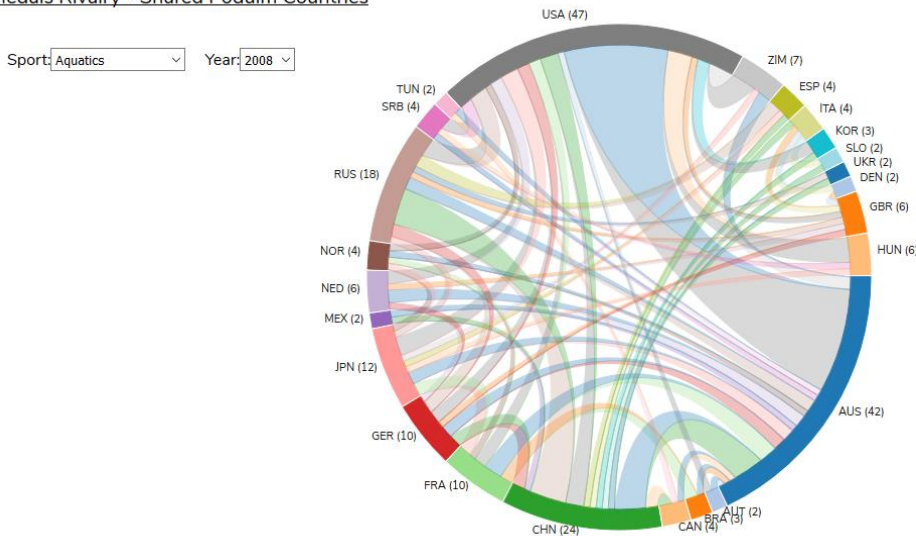
חיפשנו דרך מיוחדת ושונה להציג את התפלגות מספר המדליות בין המדינות המדליסטיות ורצינו למצוא קשרים (מבחינת אירועי הספורט המשותפים) בין המדינות. מכיוון שהDataSet שלנו די גדול, החלטנו שכדי שהUser יוכל להגיע למסקנות כלשהן יש צורך לפרק את הבעיה לבעיות קטנות יותר, כלומר, להציג מידע עבור דרישת הUser. יש באפשרותו לבחור את הספורט שבו הוא מעוניין ואת השנה (מבין 2008 ו-2012). החלטנו להתמקד בשנים אלו כי לדעתנו שנות האולימפיאדות האחרונות הן המעניינות ביותר מכיוון שבהן השתתפו מספר המדינות הרב ביותר מבין שנות המשחקים ובהן שיחקו את המספר הרב ביותר של ענפי הספורט) שבה הוא מעוניין לקבל את התוצאות. הסתכלנו באתר d3js.org ומצאנו ויזואליזציה מסוג Chord שבה ממחישים קשר בין שני nodes באמצעות מיתר, וככל שהnode יותר עבה כך ניתן להסיק שיש יותר קשרים בינו לnodes אחרים. זו ויזואליזציה שמשקפת את הקשרים הרבים בצורה כל כך אסתטית וברורה ולכן היא מתאימה לנו. במקרה שלנו:

- nodes הם המדינות.
- קיים מיתר בין שני nodes רק אם שתי המדינות הללו מקיימות את התנאים הבאים:
  1. מדינות שונות.
  2. השתתפו באותו אירוע, באותה שנה.
  3. כל אחת מהמדינות זכתה במדליה אחרת.

נקרא לקשר זה - Shared Podium.

קיבלנו את התוצאה הבאה:

Medals Rivalry - Shared Podium Countries

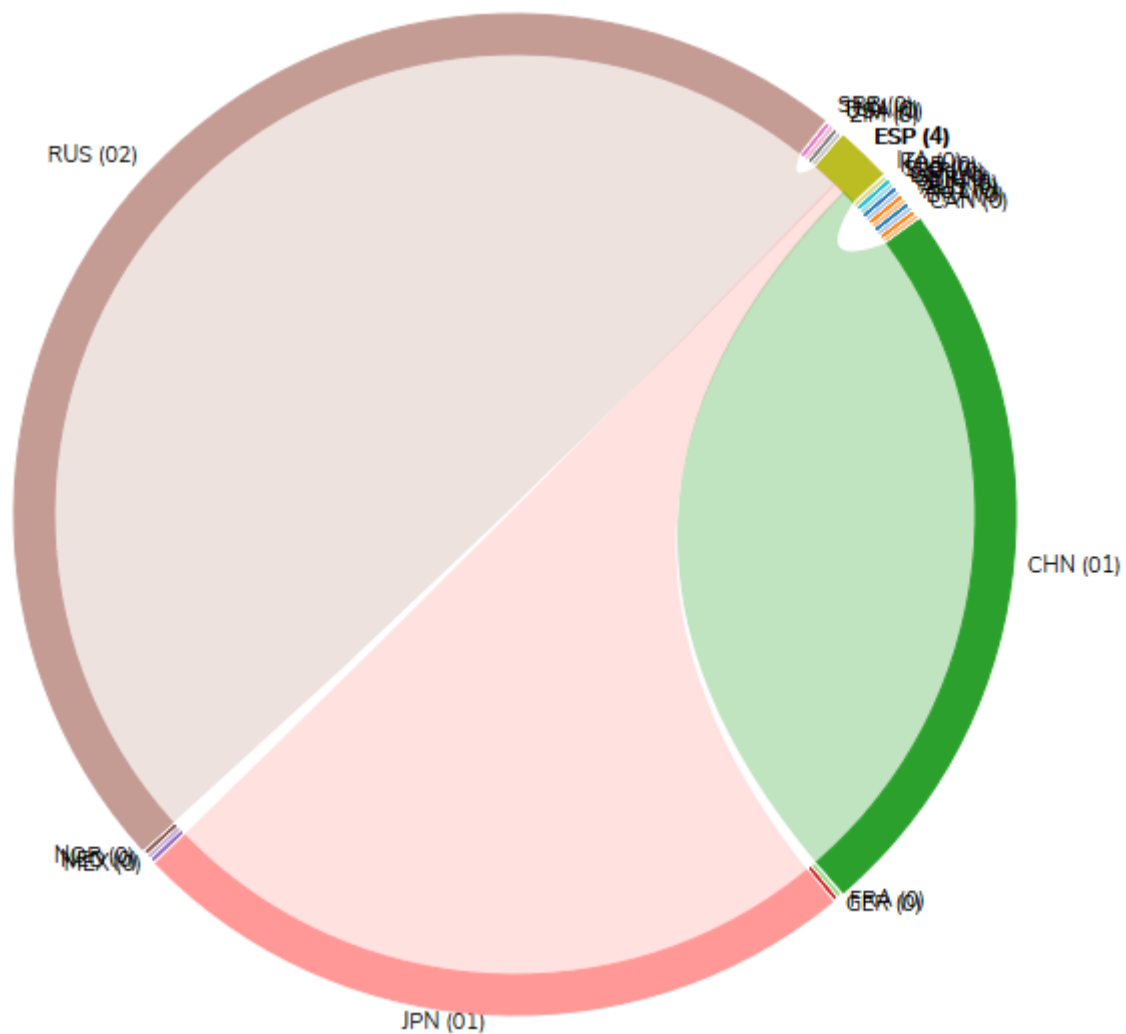


ניתן לראות בתמונה לעיל מצד שמאל שני תפריטים, האחד עבור סוגי הספורט והשני עבור השנה. User יבחר באיזה שנה ובאיזה סוג ספורט הוא מעוניין להתמקד.

לאחר מכן מוצגת הויזואליזציה במלואה ודבר ראשון שאפשר לראות הוא מהי המדינה חזקה ביותר בענף הנבחר ובשנה הנבחרת (בדוגמא הנוכחית זוהי אוסטרליה-AUS) ואת המדינות החזקות פחות (דנמרק-DEN, קוריאה-KOR ואוקראינה-UKR).

### האינטראקציה של User יכול לבצע:

כאשר נצביע עם העכבר מעל לאחת המדינות למשל מעל ספרד (ESP) נקבל-



ניתן לראות שלספרד יש 4 קשרי Shared Podium בשנת 2008 בספורט Aquatics ומתוכם 1 עם סין, 1 עם יפן ו-2 נוספים עם רוסיה.  
ניתן לראות שכל המדינות שלא רלוונטיות(כלומר, שלא חלקו פודיום עם אותה המדינה שכרגע מוצבעת).  
מצטמצמות באופן דינאמי עד שבמעט ונעלמות. מיד לאחר שהUser מזיז את העכבר בחזרה, הדיאגרמה חוזרת באופן אוטומטי למצב ה"טבעי" שלה(כלומר, מוצגים כל קשרי הShared Podium).

### **Evaluation**

לאור האמור לעיל, בויזואליזציה זו ניתן בצורה נוחה מאוד ומהירה לגלות קשרים בין מדינות בסוג ספורט מסויים בשנה מסויימת.

## Olympic Radar Rings (8)

מהרגע שסיימנו לעבוד על ה-Difference Chart - Man vs. Women על פי שהגענו למסקנות חזקות לגבי הגידול במספר הנשים הזוכות במדליות אולימפיות עדיין הייתה בנו התחושה כי זו מסקנה כללית ורצינו לרדת יותר לעומק העניין.

את ההשראה לויזואליזציה זו קיבלנו מהסמל האולימפי המורכב מ-5 טבעות שמצד אחד מייצגות את 5 היבשות - אסיה, אירופה, אפריקה, אוקיאניה ואמריקה (הצפונית והדרומית) ומצד שני טבעות אלו מייצגות ערכים חשובים בחיים בכלל ובספורט בפרט. חשבנו כיצד נוכל להציג ויזואליזציה בהשראת הסמל האולימפי.

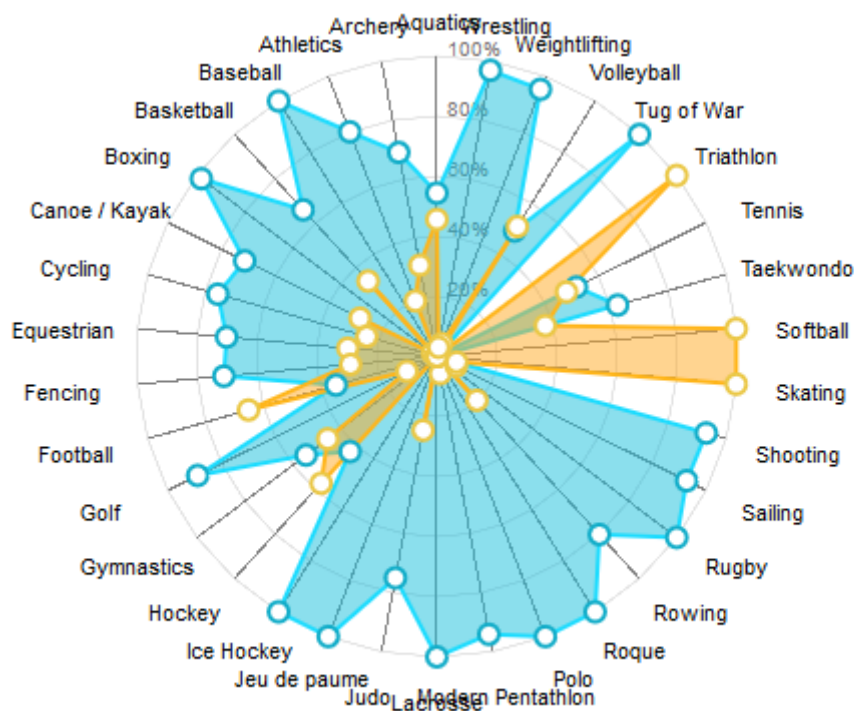
כמו כן, בהשראת עצם היותה של האולימפיאדה תחרות אשר מאחדת ומזמינה את כל מדינות העולם להשתתף ולהתחרות בה, החלטנו ליצור ויזואליזציה אשר תשדר כוח, עוצמה, אחדות של העולם כולו דרך המשחקים האולימפיים.

כזכור, רצינו לבדוק במדינות ספציפיות, באילו סוגי ספורט נשים משיגות יותר מדליות מגברים. לאחר חיפושים ממושכים בהרצאות, במאמרים המפורסמים במערכת Moodle ובאינטרנט החלטנו שנשתמש ב-Radar Chart.

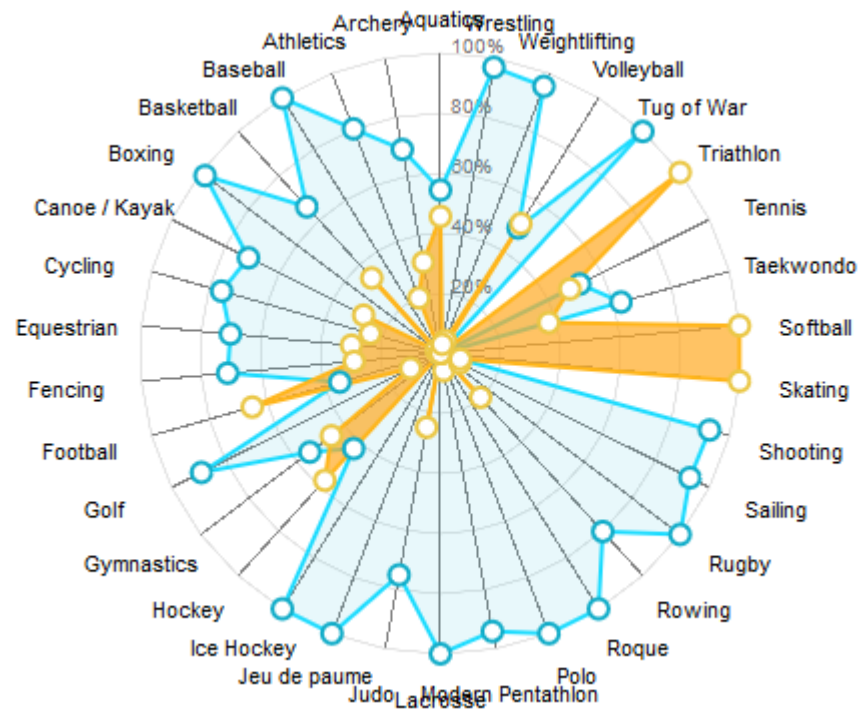
החלטנו שנעשה ויזואליזציה מסוג זה עבור מדינה ספציפית - תחילה בחרנו את מובילת הדירוג הכללי, את ארצות הברית.

-על היקף העיגול יישבו כל סוגי הספורט שהמדינה אי פעם זכתה בהם במדליות.  
- המעגל מחולק למעין "רשת קורי עכביש" עם מס' רמות (יפורט עליהן בהמשך).

התוצאה אותה קיבלנו עבור United States:

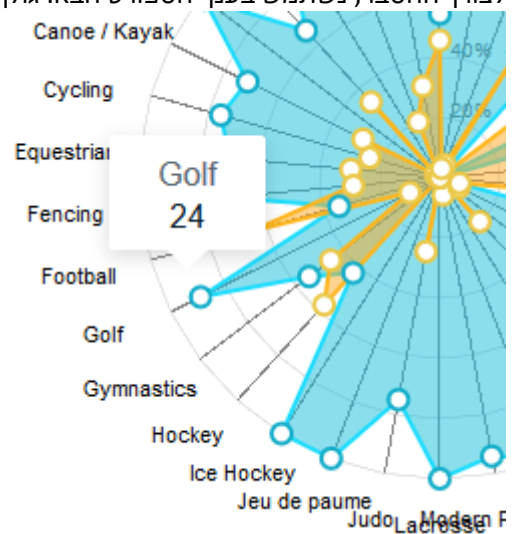


בכדי להפוך את הויזואליזציה ליותר קלה להסתכלות ולהבנה - כאשר מעבירים את העכבר מעל האזור הכתום:



ניתן להבחין שהצבע הכחול נעשה שקוף יותר על מנת להבליט את הצבע הכתום.  
(כנ"ל בשמעבירים את העבר מעל לאזור הכחול)

כל רדיוס של המעגל שווה ל-100% מהמדליות באותו ספורט אשר מצוין על היקף המעגל בקצהו של הרדיוס.  
לצורך ההסבר, נשתמש בענף הספורט הבא: גולף (Golf).



נסתכל על שני הקטעים:

1. ממרכז המעגל לנקודה הכתומה  $\Leftrightarrow$  כמות המדליות באחוזים שהשיגו נשים בגולף.
  2. ממרכז המעגל לנקודה הכחולה  $\Leftrightarrow$  כמות המדליות באחוזים שהשיגו גברים בגולף.
- סכום אורכם של שני הקטעים הנ"ל שווה לאורך הרדיוס כולו (100%).

באמצעות ויזואליזציה זו של ארה"ב אפשר לראות נתונים מעניינים במיוחד:

1. המדליות של ארה"ב, Triathlon ו-Skating, אחת מכל סוג, נשים בלבד.
2. ניתן להסיק מי השיג עבור המדינה יותר מדליות, הגברים או הנשים בכל סוג ספורט, נעשה זאת על ידי הסתכלות ברדיוס מסוים על הנקודה הכחולה (מסמלת את הגברים) והנקודה הכתומה (מסמלת את הנשים). כאשר נבחין שאחד מהצבעים הללו מרוחק יותר ממרכז המעגל מאשר הצבע השני, למשל, בענף הכדורסל:



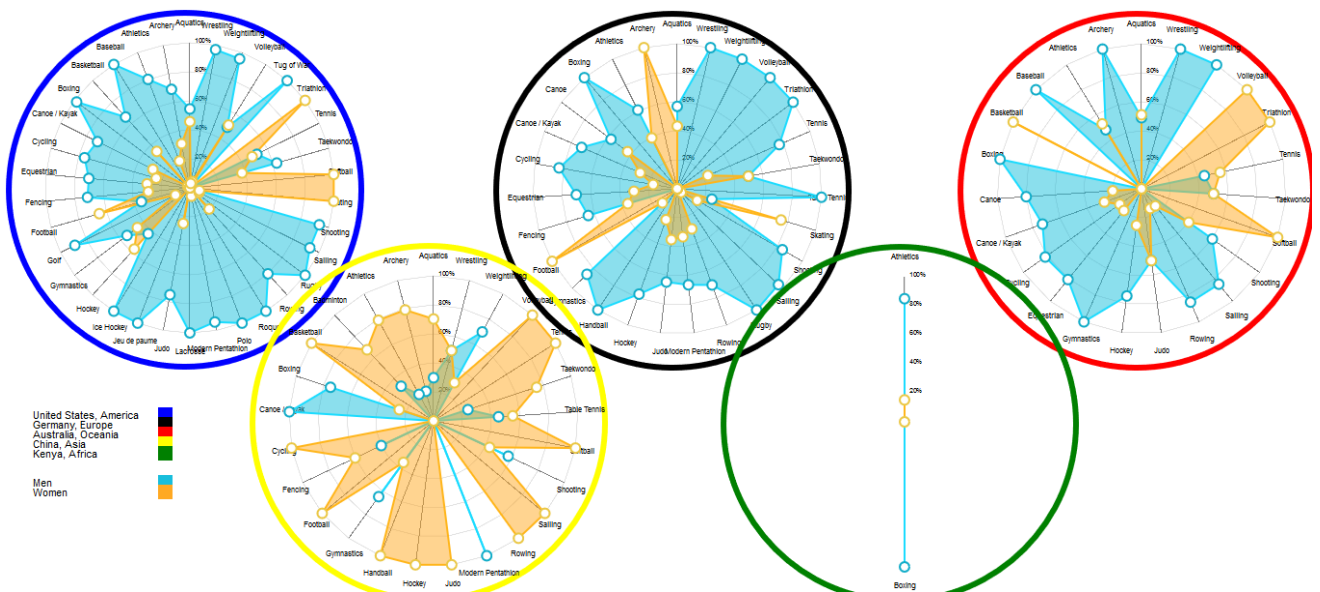
ניתן לראות שהנקודה הכחולה רחוקה יותר מהמרכז מאשר הנקודה הכתומה. מכאן, שהגברים השיגו יותר מדליות מהנשים בתחום זה.

3. ישנם סוגי ספורט כמו Gymnastic, Aquatics, Hockey שמספר המדליות של הגברים והנשים מתפלגים באופן די שווה.

לאור הקונספט וההשראה שהייתה לנו בתחילת המימוש של ויזואליזציה זו:  
בחרנו 5 מדינות, המובילות בדירוג הכללי העולמי של כל יבשת:

- אמריקה: ארצות הברית (USA)
- אירופה: גרמניה (GER)
- אוקיאניה: אוסטרליה (AUS)
- אסיה: סין (CHN)
- אפריקה: קניה (KEN)

לאחר הרבה מאוד חשיבה יצירתית הגענו לתוצאה הבאה:





### **:Evaluation & Conclusions**

כאשר המשתמש מסתכל על הויזואליזציה הזו ישנם מספר דברים אשר מושכים מיד את עינו ומובילים אותו למסקנות כמו למשל:

- המשתמש יכול לראות בבירור את הדומיננטיות של הנשים הזוכות במדליות בסין, ניתן לראות שליטה של הנשים במספר סוגי ספורט אשר הן מביאות למדינה את רוב/כל המדליות:
  - \* Volleyball-כדורעף 64 מדליות לעומת 0 לגברים,
  - \* Judo-ג'ודו 20 מדליות לעומת 0 לגברים,
  - \* Badminton-בדמינגטון 39 מדליות לעומת 19 לגברים, וכן הלאה.
- ניתן גם לראות בקניה שבאפריקה מצב מיוחד מאוד. הם זכו עד כה במדליות רק בשני סוגי ספורט:
  1. ב- Athletics זכו הגברים ב-73 מדליות והנשים זכו ב-13 מדליות.
  2. ב- Boxing זכו הגברים ב-7 מדליות והנשים לא זכו במדליות כלל.ובכל זאת הם נחשבים למדינה המובילה ביבשת אפריקה מבחינת כמות המדליות, אף על פי שמספר ענפי הספורט שבהם השיגו מדליות הוא קטן מאד.
- המשתמש יכול לראות במהירות איזו מדינה מתמחה במספר תחומי הספורט השונים הרב ביותר ואיזו מדינה במספר המועט ביותר.
- בכל המדינות חוץ מסין, ברוב ענפי הספורט, הגברים משיגים את מרבית המדליות.

לסיכום הפרוייקט,

אנו מרגישים כי הצלחנו להגשים את ציפיותינו מכל הויזואליזציות שיצרנו במסגרת הפרוייקט, הגענו למסקנות רבות ומעניינות ואנחנו מרוצים מאד מהתוצאה.