

17.01.2021

השפעות וירוס הקורונה על מגמות במוזיקה הפופולרית (ארה"ב)

האם וכיצד התבטאה מגיפת הקורונה במדדים הכמותיים של המוזיקה
המושמעת ביותר *לשנת 2020 בארצות הברית
רדיו וסטרימינג, לפי נתוני הבילבורד *

איתן ציטקין 313267791
אלקי סנדור 207047259

חלק א – תכנון וביצוע

רקע

כמה שונים הנתונים הממוצעים של הבילבורד 100 (נספח 1) ב2020 לעומת המקבילים משנים קודמות? ספציפית במדדי הקולניות (LOUDNESS), רקידות (DANCABILITY), מקצב ממוצע (BPM).

4. **Beats Per Minute (BPM)** - The tempo of the song.
6. **Danceability** - The higher the value, the easier it is to dance to this song.
7. **Loudness (dB)** - The higher the value, the louder the song.

במחקר שלנו ניסינו לתת מענה כמותי לשאלה – איך השפיעה מגיפת הקורונה על המוזיקה הפופולרית בארה"ב. מוזיקה היא אלמנט מרכזי בחיים שלנו – ברוב הפעולות שנבצע ביום יום תלויה אותנו המוזיקה שאנחנו אוהבים, לכן החלטנו לבצע את עבודת המחקר בתחום. מעבר להשפעה הלוגיסטית של מגיפת הקורונה על אורח החיים שלנו, הייתה למגפה השפעה תרבותית נרחבת ביותר שנוגעת בכולנו – תופעת הריחוק החברתי (נספח 2), ואנחנו רצינו למדוד האם הריחוק המדובר מורגש בהרגלי ההאזנה למוזיקה של אוכלוסיית ארצות הברית. בחרנו לאסוף את נתוני הבילבורד – חברה המסכמת את נתוני הסטרימינג והרדיו של כל שבוע, ומפרסמת רשימות של השירים המושמעים ביותר (בארה"ב). שיערנו כי ההשפעה של הקורונה תתבטא בהאזנה למוזיקה שונה (מהמוזיקה בתקופה ללא מגפה עולמית), וכתוצאה מכך ישתנו פני הבילבורד.

בחרנו להתמקד בכל אחד מהמדדים הנ"ל כי בעינינו הם משקפים בצורה טובה עד כמה התחושות ששיר מסוים מעורר בבני אדם הן של שמחה מול עצב – שירים ב BPM גבוה יותר נוטים להיות "שמחים" יותר: יש יותר מוזיקה "עצובה" ואיטית מאשר "שמחה" ואיטית. שירים "עצובים" יותר הם בדרך כלל פחות קולניים. בנוסף, אנו מניחים כי ככל ששיר מסוים הוא חיובי יותר – יהיה קל יותר לרקוד/לזוז אליו. על מנת לכמת את ההנחה הזו, השתמשנו בממד הרקידות, לאחר שבדקנו אותו על מספר שירים שאנחנו מכירים ולדעתנו משקפים בצורה יפה את שני קצוות הספקטרום (נספח 3).

איסוף הנתונים

הרעיון המקורי שלנו היה לקחת את טבלת הבילבורד טופ 100 של כל שבוע (כמה שיותר שבועות אחורה מהיום) ולהמיר אותה לפלייליסט באפליקציה spotify. לאחר מכן, תכננו לקבל מדדים על כל טבלה בעזרת [Sort Your Music \(playlistmachinery.com\)](https://playlistmachinery.com/):

#	TITLE	ARTIST	RELEASE	BPM	ENERGY	DANCE	LOUD	VALENCE	LENGTH	ACOUSTIC	POP	RND
1	SAD	xxxtentacion	2018-03-16	75	81	74	-6	47	2:47	35	88	215
2	Like It	Drake	2018-04-06	106	73	82	-4	45	4:13	10	94	872
3	Houdini	Drake	2018-06-29	93	91	89	-6	76	3:31	9	90	907
4	Loud Dreams	Juice WRLD	2018-10-10	84	57	51	-7	22	4:00	35	91	909
5	Sick Like You	Travis Scott	2018-04-16	60	53	61	-7	47	3:58	55	90	772
6	Psycho (feat. T. Dolla Bop)	Post Malone	2018-04-27	140	58	75	-6	46	3:41	55	82	827
7	Build Up	Drake	2017-02-22	82	79	56	-5	24	4:16	9	74	953
8	God's Plan	Drake	2018-09-29	77	45	75	-6	35	3:19	3	95	873
9	no tears left to cry	Ariana Grande	2018-08-17	122	71	70	-6	35	3:28	4	93	943

ייצאנו את הטבלאות האלה לExcel ומשם לR.

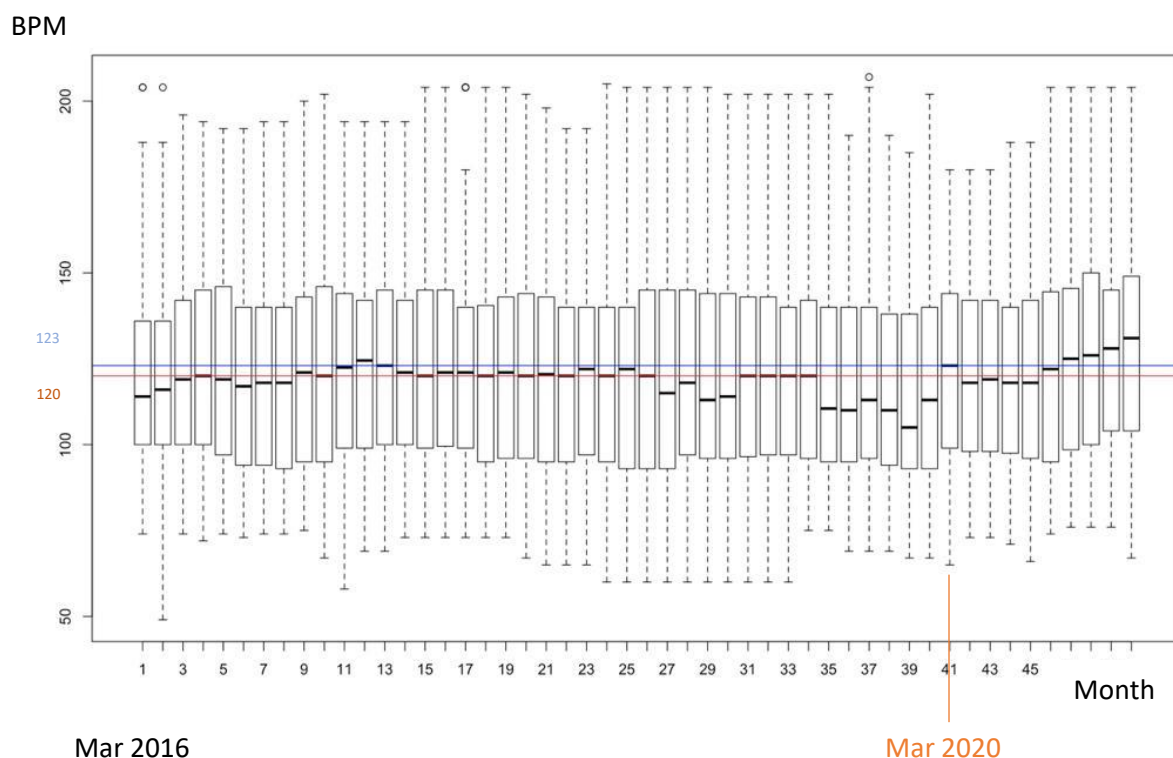
מהר מאוד נתקלנו בקושי טריוויאלי – לוקח כשעתיים בשביל להמיר פלייליסט יחיד (נתונים על שבוע בודד) לטבלת אקסל, והרי נדרשים לנו לפחות 108 פלייליסטים (שנה מול שנה). מאחר ואין לנו 216 שעות עודפות, ניגשנו לפתרון אלטרנטיבי – תוכנית פייתון שסורקת את הבילבורד וממירה את הפלייליסטים עבורינו (מצורף כקובץ נפרד). בתוך כ-26 שעות ריצה האלגוריתם המיר 5 שנים של פלייליסטים (עד ינואר 2015). משיקולים של היכולת שלנו לעבד כמויות נתונים גדולות, החלטנו לוותר על הנתונים של שנת 2015 והשתמשנו בנתונים מתחילת 2016 ועד סוף 2020.

חלק ב - תמצות וניתוח הנתונים

הערה: בניתוח הנתונים בחרנו להשמיט את הנתונים מהחודשים ינואר ופברואר מכל שנה, שכן בינואר-פברואר 2020 השפעות הקורונה עדיין לא היו ניכרות בארה"ב.

בחרנו להציג את הנתונים שלנו בעזרת גרפי תרשים קופסה. כאשר ניסינו להתייחס לממוצע החודשי (של כל מדד) גילינו שהתצפיות החריגות של כל חודש משפיעות עליו בצורה דרסטית. מספר מצומצם של שירים עם מדדים חריגים השפיעו מאוד על הממוצע, בעוד ששאלת המחקר שלנו עוסקת בשינויים במדדים של התצפיות הלא חריגות – אלה המהוות את רוב התצפיות. לפיכך, חיפשנו דרך שתאפשר לנו להתייחס בעיקר לתצפיות הרגילות – לכן בחרנו להשתמש בתרשים קופסה, הנותן פחות משקל לתצפית חריגה ויותר משקל לפיזור התצפיות. התוצאה הטבעית היא שימוש בגרף של תרשימי קופסה על פני זמן, דרך התייחסות שאיפשרה לנו לזהות מגמות שונות במדדים יחסית בקלות.

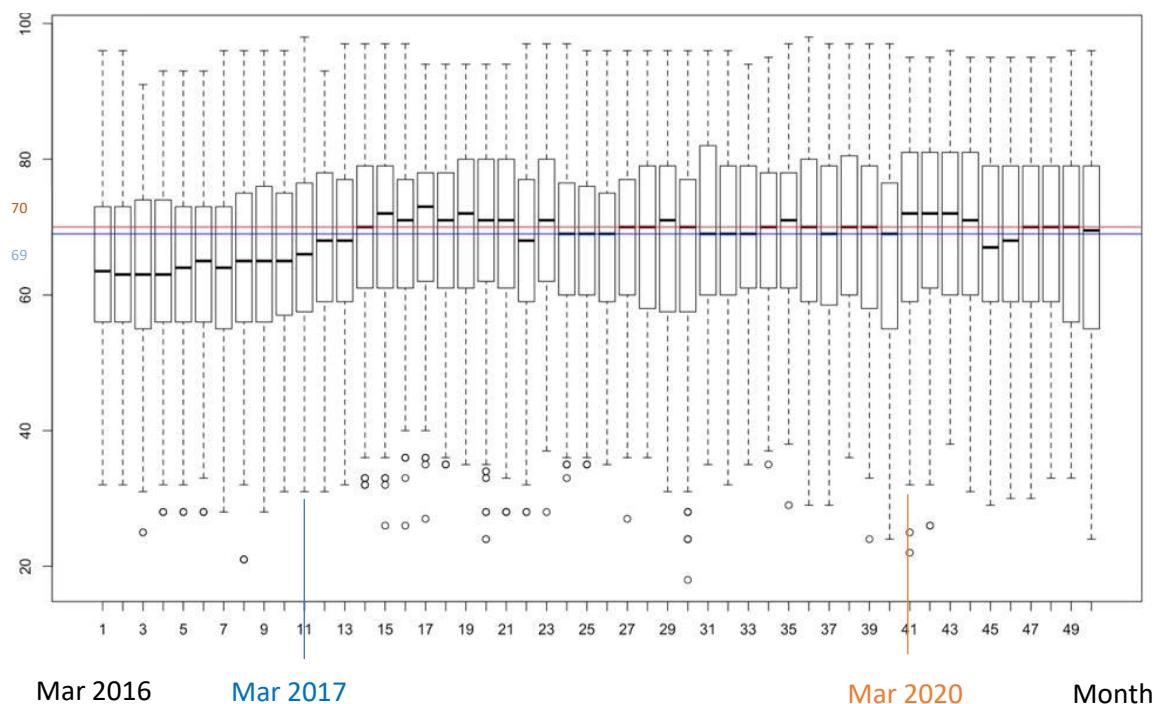
מדד המקצב BPM



נשים לב כי הפיזור של IQR הוא כעמט אחיד על פני כל התקופה – גם מבחינת הגודל (כ-45 ערכים) וגם מבחינת המדד (100-150). בחרנו להתייחס לחציון של כלל הנתונים מהחודשים 2016-12-3 (באדום) לעומת החציון של 2020 (בכחול). ניתן לראות מהתבוננות בתרשים כי החציון לשנת 2020 גבוה מהחציון של כל התקופה שקדמה לו. בנוסף, החציון החודשי נמצא במגמת עליה החל מחודש יולי 2020, והחציון השנתי של 2020 (123) גבוה יותר מהחציון השנתי של כל שנה אחרת שנמדדה (המקסימום היה 121 ב-2017). אם כן, ניתן לראות כי המגמה בפועל היא הפוכה להשערתנו המקדימה – **החציון של חודשי הדגימה לאחר הקורונה גם גבוה יותר מהחציון לפני הקורונה וגם במגמת עליה.**

מדד הרקידות Dancability

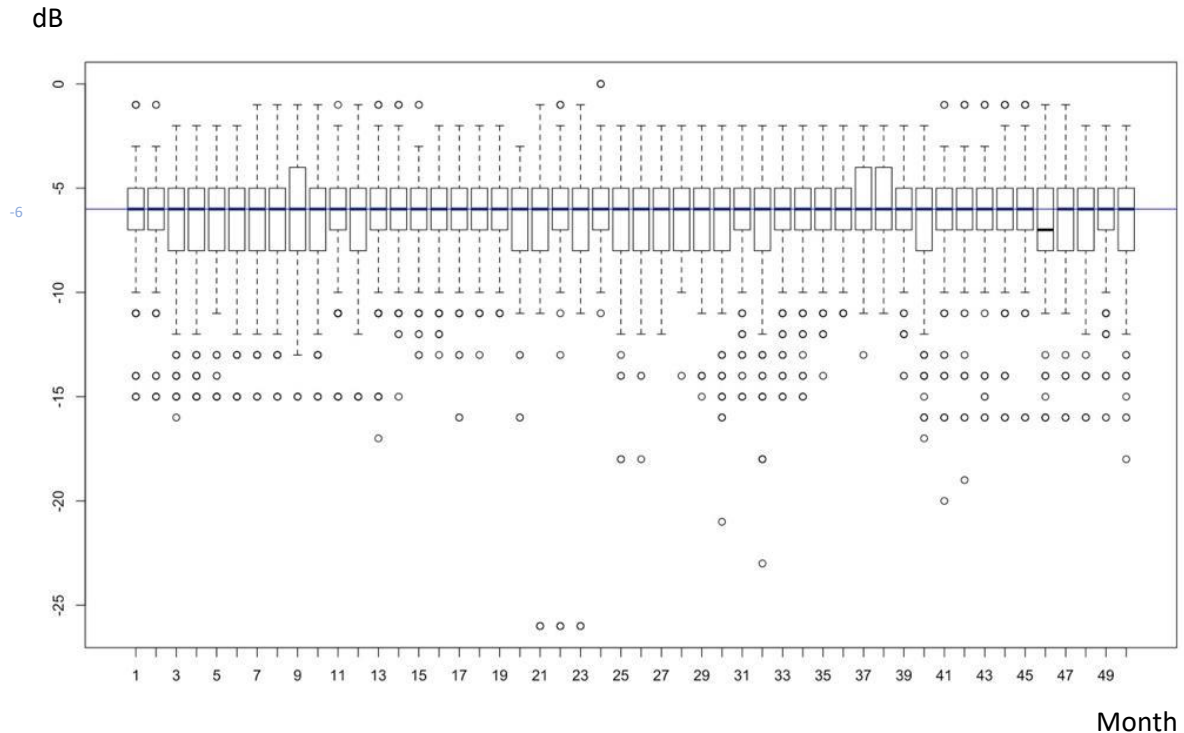
DANCABILITY



נתייחס ל IQR - ניתן לראות כי התרחשו שתי תופעות במקביל: עלייה של הערך התחתון (רבעון ראשון), וגידול בטווח הערכים – ככל שהחודשים מתקדמים ה IQR גדל מבחינת טווח הערכים שהוא מקבל: מ-74 ב-2016 ל-80 ב-2020, גידול של 16% בטווח הערכים וגם עלייה של 16% בערך הרבעון הראשון. כאשר בחנו את ה IQR של 2016-2019 קיבלנו את הערכים 59-77, לעומת הערכים 59-78 ב IQR של 2016-2020, כלומר – התרחבות קטנה כלפי מעלה. **בעינינו, השינוי קטן מכדי לייחס אותו למגמת עלייה כללית.**

נתייחס לחציונים של כל החודשים שקדמו ל-2020. במבט ראשון, ניתן לחשוב שהמדד היה גבוה יותר ב-2020, אך מהבדיקה שלנו עולה כי החציון של 2017-2020 זהה לחציון 2017-2019! (70). החציון המסומן בכחול הוא החציון של 2016-2019. ניתן להבחין בבירור כי המדד של שנת 2016 (חציון 64) נמוך באופן משמעותי מהמדדים של השנים הבאות (מבחינת הרבעונים ראשון+שלישי וגם מבחינת חציון), ומשם מגיע ההבדל. מכיוון ש-2017-2020 דומות מאוד גם מבחינת פיזור וגם מבחינת חציון, הגענו למסקנה כי להתייחס לנתוני 2016 יפספס את המטרה – המדד נותר יציב מאז שנת 2017 גם מבחינת הפיזור וגם מבחינת החציון, כלומר לא היה שינוי מהותי מאז תקופת הקורונה. **הנתונים עומדים בסתירה להשערה שלנו – שיערנו כי תהיה ירידה מובחנת במדד לשנת 2020, אך ניתן לראות כי המדד נותר ללא שינוי מהותי.**

מדד הקולניות Loudness



בניגוד למדדים הקודמים, כאן הספיק לנו חציון יחיד בשביל לרכז את כלל הנתונים. למעט חודש אוגוסט 2020, עם חציון נמוך באופן מובהק מהשאר, לא רק שיש אותו חציון (וממוצע דומה) בין 2016-2019 ו-2016, ניתן לראות בבירור כי החציון של כל החודשים זהה. **לפיכך, המדד עומד בסתירה להשערתינו – המדד נותר ללא שינוי (גם מבחינת פיזור וגם מבחינת חציון) לפני ואחרי תקופת הקורונה.** עם זאת, המדד הנ"ל הוביל אותנו להשתמש בתרשימי קופסה וחציונים: בתחילה ניסינו לבצע השוואה בין האוכלוסייה של לפני ואחרי הקורונה, לכן הנחנו כי השערת האפס היא שהתוחלות זהות, ורצינו לבדוק את ההשערה שהפרש גדול מ-0.05. רמת המובהקות של המבחן (ערך P) היא קטנה מ-0.001, כלומר יש הבדל מהותי בין תוחלת האוכלוסיות במדד. עם התוצאה הזו חזרנו אל הנתונים, וגילינו כי השוני בתוחלות נובע ממספר תצפיות חריגות מאוד שהתרחשו בשנת 2020 (הנקודות החלולות מתחת לתרשימי הקופסה), ממצא אשר הוביל אותנו לחפש דרך אחרת להסתכל על הנתונים. לראיה, כאשר התייחסנו לתרשימי הקופסה של המדד לאורך השנים, קיבלנו זהות כמעט מושלמת הן מבחינת פיזור והן מבחינת חציון, וכן רואים בבירור כי יש כמות גבוהה של תצפיות חריגות בשנת 2020.

סיכום

לסיכום הממצאים,

נראה כי מאז תחילת תקופת הקורונה אוכלוסיית ארה"ב מאזינה לשירים עם מקצב גבוה יותר משנים קודמות, בעוד שהקולניות והרקידות שלהם לא השתנתה. אם כן, לא מצאנו שינוי בשניים מתוך שלושת המדדים שבחרנו להתמקד בהם, בעוד שהמדד השלישי הציג התנהגות הפוכה להשערתינו – עליה במקום ירידה. ההשערות שלנו (לראות ירידה בשלושת המדדים) נבעו מתוך הנתונים על שיעורי הדיכאון הגוברים באוכלוסיית ארה"ב (נספח 4) מאז חודש מרץ 2020. הנחנו כי המוזיקה שהאמריקאים בחרו להאזין אליה מאז תחילת המגיפה **תשקף** את הלך הרוח שלהם, אך נראה כי הממצאים מעידים על כך שהאמריקאים השתמשו במוזיקה ככלי **מעצב** – המדדים נותרו ללא שינוי או עלו מאז תחילת המגיפה בארה"ב.

עם זאת, אחת החולשות המרכזיות של המחקר שלנו מתבטאת בכך ששאלת המחקר מתייחסת לנתונים אורדינליים – מידת ה"שמחה" של שירים פופולריים, בעוד שאמצעי החקר הוא בבסיסו כמותי – השוואה של חציונים ופיזורים של תצפיות. לפיכך, קשה להשליך מהממצאים על היחס של המאזין האמריקאי הממוצע למוזיקה (מעצבת מול משקפת).

החוזקה המרכזית של המחקר היא כמות הנתונים שנבדקה (גודל המדגם) – בדקנו את השינויים במדדים לאורך 4 שנים, בצורה שמפחיתה את ההשפעה של תצפיות חריגות על הממצאים. לכן אנו מאמינים כי המגמות שתיארנו בעמודים 2-4 אכן מבוססות.

לקחים עיקריים שהפקנו הם:

- קשה להשוות נתונים מסוגים שונים (אורדינלי מול כמותי) ולהפיק מכך מידע חד משמעי.
- יש דרכים רבות להציג את אותם הנתונים, ולהסיק מהם מסקנות שונות. יש להתאים את אופן ההצגה לשאלת המחקר על מנת להסיק מסקנות קוהרנטיות.
- שליטה בכלי המחקר (R, Excel, Python, Word) יכולה לחסוך שעות עבודה רבות, על אחת כמה וכמה כאשר יש נתונים רבים.

נספח

1.

[The Hot 100 Chart | Billboard](#)

2.

Test Dencability

BACK

Minimum BPM Maximum BPM Include Doub

#	TITLE	ARTIST	RELEASE	BPM	ENERGY	DANCE	LOUD
1	Uptown Funk (feat. Bruno Mars)	Mark Ronson	2015-01-12	115	61	86	-7
2	Happy - From "Despicable Me 2"	Pharrell Williams	2014-03-03	160	82	65	-5
3	Black Hole Sun	Soundgarden	1994-03-09	105	83	35	-5
4	Fade To Black - Remastered	Metallica	1984-07-26	114	92	26	-7

המדד המסומן משקף את הנחת היסוד "עד כמה שיר רקיד" באופן הגיוני יותר ממדדים אחרים בטבלה.

3.

ריחוק חברתי:

[Social Distancing \(cdc.gov\)](#)

4.

דיכאון, חרדה ומחשבות אובדניות בארה"ב – 2019 לעומת 2020:

[Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic — United States, June 24–30, 2020 | MMWR \(cdc.gov\)](#)