

ויזואליזציה של מידע

סמסטר א', תשפ"ה

פרויקט

משקל: 56% מציון הקורס

אמ;לק

1. בחירת סט נתונים.
2. ניתוח נתונים ומטלות.
3. עיצוב סכמטי (טיוטות) של 2 חלופות ויזואליות לכל מטלה.
4. מימוש דאשבורד של החלופות הנבחרות.
5. כתיבת דו"ח מסכם של כל התהליך.
6. הגשת המימוש והדו"ח.

מבוא

מטרת פרויקט הסיום היא לבנות ויזואליזציה מקורית משלכם תוך יישום החומר התיאורטי הנלמד בקורס. לשם כך, עליכם לבחור dataset בתחום שמעניין אתכם (ניתן, כמובן, ולעתים רצוי, להשתמש בנתונים מיותר מדאטאסט אחד), לחשוב על שאלות מעניינות שניתן לענות עליהן באמצעות הדאטאסט, לעצב ולממש ויזואליזציה בהתאם.

ניתן לפתח רעיון מקורי לוויזואליזציה, או לבחור שיטת ויזואליזציה קיימת וליישם אותה ברמה גבוהה. אינטראקציה היא מפתח חשוב בוויזואליזציה, וכמעט תמיד נדרשת רמה כלשהי של אינטראקציה (לדוגמה, שימוש במסננים וב-Brushing). התוצאה הסופית צריכה לעזור למתבונן להגיע לתובנות לגבי הנתונים או הבעיה שלא ניתן היה להגיע אליהם (או היה לוקח הרבה זמן) ללא הוויזואליזציה. למשל, להראות מגמות או דפוסים או מקרי קצה.

התוצר הסופי של הוויזואליזציה יהיה מעין דאשבורד, או דף אינטרנטי, הכולל כמה ויזואליזציות ומלל רלבנטי (הסבר כללי, כותרות והסברים לתרשימים השונים) שבאמצעותם ניתן לחקור את הנתונים.

הפרויקט יתבצע בשלבים בהתאם ללוח הזמנים הבא:

חלק	בחירת סט נתונים	תאריך	הערות
1	בחירת סט נתונים	18.11.24	חובה - תאריך אחרון לקבלת אישור
2	ניתוח נתונים ומטלות	2.12.24	מומלץ לסיים עד התאריך הנקוב
3	עיצובים חלופיים	30.12.24	מומלץ לסיים עד התאריך הנקוב
4	מימוש ודו"ח מסכם	2.2.25	חובה - הגשת מימוש + דו"ח מסכם

חלק 1: בחירת Dataset

יש לבחור מערך נתונים (Dataset) אחד או יותר עם נתונים שיכולים לענות על שאלות שמעניינות אתכם. עליכם להירשם ולבקש אישור למערך הנתונים שבחרתם [כאן](#). חשוב לוודא שבדאטאסט(ים) יש מספיק "בשר" בשביל כמה ויזואליזציות, למשל כמויות סבירות של תכונות (פיצ'רים) ורשומות/פריטים. בנספח תוכלו למצוא כמה מאגרי מידע מתוכם ניתן לבחור את ה datasets לפרויקט שלכם. רשימה זו אינה ממצה ואתם יכולים לבחור ממקורות אחרים. בכל מקרה עליכם לבקש אישור בקישור הנ"ל.

חלק 2: ניתוח נתונים ומטלות

בחלק זה עליכם לנתח את הנתונים שבחרתם ואת מטלות המשתמשים. הניתוח יתבסס על החומר שלמדנו בשיעור 3. ניתן להיעזר בספר של Munzner, פרקים 3 ו-4, ובהרצאות שלה (ראו קישור בסילבוס). הדו"ח המסכם של הפרויקט יכלול פרק של שניים עד שלושה עמודים המבוססים על פעילות זו. הפרק יחולק לשלושה סעיפים:

1. **מבוא:** מהו נושא הוויזואליזציה, מהי השאלה המרכזית בתחום עליה אתם מבקשים לענות ומהן שאלות המשנה. על אילו נתונים אתם מסתמכים, מה מקורם ומה היקפם.

2. נתונים:

א. תארו באופן כללי את הנתונים לפרויקט שבחרתם, כולל קרדינליות הנתונים, מפתחות, ומשמעות העמודות והרשומות.

ב. מפו את הנתונים בהתאם לטיפולוגיה האבסטרקטית של Munzner (Data types, dataset type, attributes).

3. מטלות:

א. תארו את מטלות המשתמש (השאלות עליהן הוויזואליזציה אמורה לענות) במונחי התחום (למשל, צרכנות, רפואה, דמוגרפיה, כדורסל)

ב. מפו אותם למונחים האבסטרקטיים בהתאם לטיפולוגיה של Munzner תוך שימוש בצמדי {action, target}. שימו לב שלכול מטלה יש פעולות בשלוש רמות שונות. בתהליך המיפוי, השתמשו במונחים שלמדנו בכיתה (באנגלית, בנוסף, אפשר לנסות לתרגם לעברית) רצוי מאד לארגן את סעיפים 2 ו-3 באמצעות טבלה.

קריטריונים להערכה:

- האם הנושא מעניין והשאלות הנשאלות אינן טריוויאליות לגמרי?
- האם המיפוי מנתוני ומטלות התחום למטלות ונתונים אבסטרקטיים נכון?

חלק 3: עיצובים חלופיים

בחלק זה עליכם לעצב רעיונות שונים שלדעתכם יכולים להתאים לוויזואליזציה הסופית. עליכם להכין סקיצות של הרעיונות, ולהעריך את התאמתן לנתונים ולמטלות כפי שתוארו בשלב הקודם וכן ליכולתכם לממש אותן. כפי שנלמד בקורס, תחום העיצוב מלא ברירות פשרה, ובחלק זה של הפרויקט עליכם לבצע הערכה של ברירות פשרה אלו: מהם היתרונות והחסרונות של כל אחד מרעיונות העיצוב השונים שאתם מציעים? למשל, חלופת עיצוב אחת תענה על כל מטלות המשתמשים, אבל תהיה קשה מאד למימוש; חלופה אחרת מתאימה לחלק מהמטלות אבל לא לכולן, וכו'.

הדו"ח המסכם של הפרויקט יכלול פרק המתאר את תוצרי הפעילות בחלק 3:

1. עליכם לעצב ולהציג שתי טיוטות לוויזואליזציות חלופיות לכל אחת מהמטלות שהצגתם בחלק 2, סעיף 3. **הציגו כל חלופה בנפרד על פי הסעיפים הבאים:**

א. ציינו מה היא המטלה עליה עונות החלופות. יש להציג את המטלה במונחים אבסטרקטיים ובמונחי התחום (העתיקו מחלק 2, סעיף 3).

ב. סקיצות (low fidelity) של כל אחת מהוויזואליזציות. הסקיצות יכולות להיעשות בשרטוט על נייר (צלמו את השרטוטים ושכצו בדו"ח), באמצעות תכנת PowerPoint או תוכנות שרטוט ועיצוב שונות.

ג. תיאור כל חלופת עיצוב במונחים של marks & channels.

ד. הסבירו בקצרה כיצד כל אחת מהחלופות עומדת בעקרונות האקספרסיביות והאפקטיביות (שימו לב שלא כל עקרונות האפקטיביות רלבנטיים בהכרח). במידה וחלופה לא עומדת בהם באופן מוחלט, מהי ההצדקה לכך?

ה. הסבר קצר של היתרונות והחסרונות של כל חלופה. ההסברים צריכים להתייחס גם לתכונות של הערוצים הוויזואליים (visual channels) בהם השתמשתם. בסעיף זה ובסעיף 2 להלן, אני מצפה להסברים

מהותיים שנשענים על מה שלמדנו בקורס ולא למריחות. בבקשה לא להשתמש במונחים כמו "קל להבנה", "אינטואיטיבי", "פשוט וברור", וכו'.

2. מיד לאחר הצגת וניתוח שתי החלופות, הסבירו את הסיבות לבחירת אחת מהן למימוש בשלב האחרון של הפרויקט (כלומר, מדוע העדפתם אותה על פני החלופה האחרת).

אני לא מכתוב לכם את אורך הפרק הזה בדו"ח כי הוא תלוי במידה רבה בגודל וכמות הסקיצות שתציגו. בכל אופן, כמות המלל נטו צריכה להיות בין שניים לשלושה עמודים.

קריטריונים להערכה:

- האם לפחות אחת החלופות יצירתית? לא חובה להמציא שיטות ויזואליזציה חדשות. יצירתיות יכולה להתבטא גם בפתרון מקומי, צורת הצגה, או אפילו בזיהוי בעיה מעניינת.
- האם תיאור האלטרנטיבות במונחים של marks & channels נכון?
- האם המיפוי למשתנים ויזואליים מוצדק ועונה על עקרונות האפקטיביות והאקספרסיביות?
- האם ההסברים לגבי היתרונות והחסרונות של החלופות ברורים וסבירים?
- האם הצדקת החלופה המועדפת מוסברת היטב?

חלק 4: מימוש ודו"ח מסכם

בחלק זה עליכם להגיש שני תוצרים. I. ויזואליזציה עובדת. II. דו"ח מסכם שייכתב בהתאם למבנה והסעיפים המפורטים בהמשך. צוות ההוראה של הקורס יריץ ויבדוק את הוויזואליזציה על דפדפן **כרום**, ויעריך אותה ואת הדו"ח שתכתבו בהתאם לקריטריונים המופיעים בהמשך.

א. מימוש הוויזואליזציה

המימוש יתבסס על החלופה שבחרתם בחלק 3 של הפרויקט, אם כי ניתן לעדכן אותה בהתאם לתובנות שנצברו עם הזמן או לאור ההתקלות במציאות הטכנולוגית (למשל, אילוצים או הזדמנויות במימוש). במימוש אני מבקש שתקפידו על ויזואליזציה שמסייעת להבנת הנתונים ושתיקחו בחשבון שפיצ'רים מיותרים לא רק שלא מסייעים לכך אלא אפילו מפריעים. עליכם לממש את הוויזואליזציה ולהגיש אותה באופן שניתן יהיה לבדוק אותה מבלי שאנחנו נצטרך לבצע הרצות או פעולות כלשהן.

עליכם להניח שאת הדאשבורד יראו אנשים שלא קראו את הדו"ח. לכן, הדאשבורד צריך לכלול הסברים קצרים ומדויקים על מהות הנתונים המוצגים, על השאלות עליהן עונה כל אחת מהוויזואליזציות, ולוודא שניתן להבין מה מציגה כל ויזואליזציה.

הגשה:

א. קבצים (קישורים ל- github או קבצים ב- zip):

- 1) קבצי קוד: python, Ipython, R
- 2) קבצי ה dataset

ב. קישור לאפליקציות דאשבורד כמו Streamlit, Dash, Shiny. יש לוודא שהוויזואליזציה ניתנת להרצה לפחות חודשיים לאחר מועד ההגשה.

II. דו"ח

הדו"ח חשוב כדי שאנשים יבינו מה עשיתם - כולל אנשים שאינם מהתחום. הדו"ח יוגש בעברית ויכלול כותרת וארבעה פרקים.

0. כותרת: שם הקורס, נושא הוויזואליזציה, שמות המגישים, תאריך הגשה, קישור לוויזואליזציה.

1. **מבוא:** מהו נושא הוויזואליזציה, על איזו שאלה (או שאלות) היא מנסה לענות. על אילו נתונים אתם מסתמכים, מה מקורם ומה היקפם. אורך הסעיף – כחצי עמוד.

2. **נתונים ומטלות:** תיאור וניתוח של הנתונים ומטלות המשתמש. זהו הפרק המתואר בחלק 2 של הפרויקט.

3. **רעיונות ועיצובים חלופיים:** הצגת העיצובים החלופיים לוויזואליזציה, בהתאם להנחיות בחלק 3.

4. **הסבר על העיצוב שנבחר ועל יישומו**

א. אם ביצעתם עיבוד מוקדם כלשהו של נתונים (pre-processing), תארו והסבירו את הפעולות שנעשו.
ב. במידה ונאלצתם לשנות את מימוש הוויזואליזציה של החלופה הנבחרת (בגלל אילוצים או תובנות חדשות תוך כדי תהליך המימוש), הסבירו בקצרה את ההבדלים בין העיצוב המקורי לעיצוב שמומש.
ב. הראו בדו"ח בעזרת תצלומי מסך (בצבע!) איך הוויזואליזציה פועלת. עבור תרשימים אינטראקטיביים, הראו מספר תצלומים והסבירו איך האינטראקציה אמורה לעבוד. זה יעזור לנו להעריך את המימוש.
ד. הסבירו בקצרה מה עשיתם בכתיבת הקוד – באילו ספריות השתמשתם? האם השתמשתם בדוגמה או קוד קיימים? אם כן, מאיפה לקחתם אותם? **ניתן להשתמש בקוד ודוגמאות קיימים, אבל חובה לציין זאת בדו"ח** (אם השתמשתם בדוגמה, מה בדיוק הוספתם, ואיזה התאמות עשיתם).

אורך הטקסט בסעיף זה: 1 - 2 עמודים.

קריטריונים להערכה:

- **התאמת הוויזואליזציה (20%).** האם הוויזואליזציה מציגה בצורה אפקטיבית את המידע. למשל:
 - האם הוויזואליזציה עונה על מטלות המשתמש?
 - האם היא מקלה על המשתמשים למצוא תשובות ולחקור בצורה טובה את השאלות?
 - **נכונות (20%)**
 - האם הוויזואליזציה בנויה ומציגה נתונים על פי הכללים שנלמדו בקורס?
 - האם השימוש בצבעים נכון?
 - האם יש אינטראקציה והאם היא בנויה בצורה נכונה?
 - האם הצירים מוצגים נכון, האם יש מקרא, כותרות, וכו'? האם הם ברורים וקריאים?
 - האם הוויזואליזציה נקיה מאלמנטים מיותרים שמקשים על הבנתה או על האינטראקציה עימה?
 - **היקף (15%).** היקף צריך להתבטא בוויזואליזציה שאינה פשוטה מדי. למעט במקרים חריגים, המשמעות היא הצגת הנתונים בעזרת כמה גרפים לא טריוויאליים תוך שימוש באינטראקציה לחקר הנתונים על ידי המשתמש. במידת הצורך, מומלץ לקשר בין התצוגות באמצעות linking and brushing.
 - **יצירתיות (10%).** ראו קריטריונים בחלק 3. מקוריות ויצירתיות יכולות לפצות על היקף עבודה קטן יותר.
 - **אסתטיות (10%).** הדף והוויזואליזציה שלכם צריכים להראות מקצועיים ואסתטיים.
 - **דו"ח (25%).** כאמור, הדו"ח חשוב כדי לאפשר לאנשים שונים להבין את הוויזואליזציה ואת הסיבות לעיצובה הסופי. לשם כך הדו"ח צריך להיות ברור ומובן, ובהערכתו נתייחס לשלמות ולבהירות בה מוצגים הנושאים השונים. הדו"ח יורכב מארבעת הפרקים הרשומים בסעיף II לעיל.
- יש להקפיד על הגשה מסודרת של הדו"ח:** יש להקליד באופן ברור את כותרת העבודה, שם הקורס, תאריך ההגשה, שמות ותעודות הזהות של המגישים, יש להקפיד על שוליים של 1 אינץ' (2.5 ס"מ) מכל צד, ועל ריווח של שורה וחצי בין השורות. יש לכתוב בגופן "אריאל" בגודל של 11 נקודות.

בהצלחה!

נספח

להלן כמה מאגרי מידע מתוכם ניתן לבחור את ה datasets לפרויקט שלכם. רשימה זו אינה ממצה ואתם יכולים לבחור ממקורות אחרים. בכל מקרה עליכם לבקש אישור בקישור שמופיע בחלק 1.

1. מאגרי מידע ממשלתיים בישראל:

<https://data.gov.il/organization>

2. נתונים על מערכת החינוך בישראל (קישור). למשל, תוצאות ההערכה החיצונית, תשפ"ג (קישור).

3. FiveThirtyEight (538)

אתר המוקדש לכתבות המבוססות על נתונים מתחומים מגוונים – פוליטיקה, ספורט, כלכלה מדע וטכנולוגיה, ועוד. ניתן למצוא נתונים עליהם התבססו הכתבות ב- <https://github.com/fivethirtyeight/data>. רוב הכתבות כוללות ויזואליזציות. עליכם לבנות פרויקט ויזואלי יותר מקיף ממה שהוצג בכל כתבה.

4. Kaggle

<https://www.kaggle.com/datasets>

5. נתוני כדורסל מה-NBA

ניתן להוריד נתונים סטטיסטיים מכאן:

<https://www.mysportsfeeds.com/data-feeds/>

או להעתיק נתונים לאקסל מכאן:

<https://www.basketball-reference.com/leagues>