

# **Degrees of Distinction: Visualizing the Data of Academic Majors**

**וויזואליזציה של מידע**

**פרויקט סוף**

רובן שוקרון - 327064358

אלון פרידנטל - 209370774

30.08.2024

## מבוא

הפרויקט הנוכחי עוסק בויזואליזציה של נתונים אודות תארים אקדמיים בארצות הברית. הנתונים לקוחים מקובץ הנתונים "recent-grads.csv" שהורד מאתר GitHub של FiveThirtyEight, המכיל מידע מקיף על תחומי לימוד שונים בארצות הברית, לרבות מספר בוגרים, התפלגות מגדרית, שיעורי תעסוקה ושכר חציוני. העיצובים שנבחרו מנסים לענות על שאלות מרכזיות כמו: כיצד מתפלג המגדר בין תחומי לימודים שונים, אילו סוגי משרות זמינות לבוגריהם, ומהם תחומי הלימודים שמציעים את השכר החציוני הגבוה ביותר – כל זאת במטרה לספק למשתמש הבנה מעמיקה של הקשר בין בחירת תואר אקדמי, מגדר ותנאי העסקה. מטרת הפרויקט היא להציג תובנות חדשות על העולם האקדמי ושוק העבודה לבוגריו באמצעות שלושה גרפים מרכזיים, וכל עיצוב בפרויקט נועד להציג חלק אחר של הנתונים: הראשון מציג את התפלגות המגדר והשכר לפי תארים ומאפשר השוואה בין קטגוריות אקדמיות שונות, השני מציג את השכר החציוני של כל תחום לימודים, והשלישי מציג את סוגי המשרות הקיימות בכל קטגוריה של תואר (משרות הדורשות תואר, שאינן דורשות תואר ומשרות בשכר נמוך). הפרויקט שם דגש על עיצובים ייחודיים שמטרתם להעשיר את חוויית המשתמש ולספק זוויות חדשות לתובנות שניתן להפיק מהנתונים, תוך שאיפה להציג כמה שיותר מידע בצורה ברורה ומובנת.

## ניתוח נתונים

קובץ הנתונים כולל טבלה אחת בשם 'recent-grads' שבה 173 רשומות, כאשר כל אחת מהן מייצגת תחום לימודים שונה (תואר). השדה "Major\_code" משמש כמפתח ראשי בטבלה, ומזהה באופן ייחודי כל תואר. הנתונים כוללים מידע כמותי המאפשר להעריך את מצבם המקצועי והכלכלי של בוגרי התארים השונים. להלן תיאור עמודות הטבלה ומיפוי שלהם לפי הטופולוגיה של Munzner:

Dataset Type: Flat table Item: תואר מסוים		
Attribute	Attribute Type	Description
Major_code	Categorical	קוד ייחודי המזהה את התואר
Rank	Ordinal	דירוג התואר לפי שכר חציוני
Major_category	Categorical	קטגוריית התחום האקדמי שאליו משתייך התואר (לדוגמה: הנדסה, מדעי הרוח)

Major	Categorical	שם התואר
Total	Quantitative	סך כל הבוגרים לתואר
Sample_size	Quantitative	גודל המדגם שבו השתמשו לצורך החישובים של השכר
Men	Quantitative	מספר הבוגרים הגברים של התואר
Women	Quantitative	מספר הבוגרים הנשים של התואר
ShareWomen	Quantitative	אחוז הנשים מתוך כלל הבוגרים של התואר
Employed	Quantitative	מספר הבוגרים המועסקים
Full_time	Quantitative	מספר הבוגרים המועסקים במשרה מלאה (יותר מ 35 שעות שבועיות)
Part_time	Quantitative	מספר הבוגרים המועסקים במשרה חלקית (פחות מ 35 שעות שבועיות)
Full_time_year_round	Quantitative	מספר הבוגרים המועסקים במשרה מלאה לאורך כל השנה (לפחות 50 שבועות בשנה)

Unemployed	Quantitative	מספר הבוגרים שאינם מועסקים
Unemployment_rate	Quantitative	אחוז האבטלה בקרב בוגרי התואר
Median	Quantitative	השכר החציוני של הבוגרים בתואר
P25th	Quantitative	האחוזון ה 25 של השכר לבוגרי התואר
P75th	Quantitative	האחוזון ה 75 של השכר לבוגרי התואר
College_jobs	Quantitative	מספר הבוגרים של התואר המועסקים בעבודות הדורשות תואר אקדמי
Non_college_jobs	Quantitative	מספר הבוגרים של התואר המועסקים בעבודות שלא דורשות תואר אקדמי
Low_wage_jobs	Quantitative	מספר הבוגרים של התואר המועסקים בעבודות שירות עם שכר נמוך

## ניתוח מטלות

### 1. ניתוח הרכב מגדרי של תארים/ קטגוריות של תארים

a. במונחים של התחום:

**נרצה להבין מהו ההרכב המגדרי של תארים אקדמיים שונים ולהשוות בין הרכבים אלו.**  
המטלה כוללת שתי רמות של השוואה: ברמה הראשונה, ניתן להשוות את הרכב המגדר של בוגרי תארים שונים בתוך כל קטגוריית תארים (למשל, להשוות בין תארים שונים בתוך קטגוריה כמו "הנדסה" או "מדעי החברה"). ברמה השנייה, ניתן להשוות את הרכב המגדר בין קטגוריות התארים השונות עצמן (למשל, לבדוק את ההבדלים בהרכב המגדרי בין קטגוריית התארים של "הנדסה" לעומת "מדעי הרוח"). המטרה היא לזהות מגמות ברמה המגדרית בתוך שדה הלימודים, כמו גם להבין כיצד תארים שונים נבדלים זה מזה מבחינה זו. בשתי הרמות נוכל לראות בדיוק כמה גברים או נשים מרכיבים את ההתפלגות ואת האחוז המדויק שהם מייצגים מהאוכלוסייה.

b. במונחים של Munzner:

Action	Target	Explanation
Lookup	Distributions	נרצה לבחור עבור איזה קטגוריות של תארים מציגים את התפלגות המגדר. ניתן לבחור כמה. כאן, המטרה (קטגוריה) ידועה וגם במיקום ידוע (מוצגת רשימה ממוינת של קטגוריות אפשריות).
Compare	Distributions	נרצה להשוות בין ההתפלגויות.
Discover	Distributions/ Trends	לאחר בחירת קטגוריות, נרצה 1. לבחור בקטגוריה ספציפית על מנת לגלות את התפלגות המגדר של התארים שלה 2. לזהות מגמות בין תארים או קטגוריות, איפה למדו יותר או פחות מכל מגדר
Browse	Distributions	אנו נוכל לעבור בין מספר התפלגויות של תארים השייכים לקטגוריה מסוימת. כאן, המיקום ידוע אך המטרה אינה ידועה, ונוכל להסתכל על מה שמוצג. יש עזרי ניווט בגרף זה (scrolling) כדי לעבור בין ההתפלגויות המוצגות.

### 2. ניתוח שכרים חציונים לפי תואר

a. במונחים של התחום:

**נרצה להציג את השכר החציוני של בוגרי תארים שונים.** המטרה המרכזית של הוויזואליזציה היא להנגיש מידע על פוטנציאל השכר של תחומים אקדמיים שונים, ולתת למשתמשים לבחון ולהשוות בין תארים לפי גובה השכר החציוני שלהם. המשתמשים יוכלו למיין את הנתונים לפי שכר או לפי קטגוריית תארים, ולבחור בסדר עולה או יורד, מה שיאפשר השוואה נוחה וברורה בין תארים ובין קטגוריות של תארים.

b. במונחים של Munzner:

Action	Target	Explanation
Identify	Extremes	נרצה לזהות את התארים עם המשכורות הגבוהות והנמוכים ביותר, באופן כללי או בתוך קטגוריה מסוימת.
Summarize	Features	נרצה להציג את כלל המשכורות עבור כלל התארים, למעשה להציג סיכום של כל המידע.
Compare	Features/ Trends	נרצה שהמידע המוצג יהיה ניתן להשוואה, בין תארים וגם עם אופציה להשוות בין קטגוריות. הצגה מסוג זה תאפשר גם זיהוי מגמות על פי קטגוריית התואר.
Browse	Features	המשתמש יוכל לחפש באופן חופשי (יעד לא ידוע) באזורי שכר שמעניינים אותו (מיקום ידוע).

### 3. ניתוח סוגי משרות לפי קטגוריה של תואר

a. במונחים של התחום:

נרצה להציג את מספר המשרות השונות לפי קטגוריות תארים אקדמיים, כשהמשרות מחולקות לשלושה סוגים: משרות הדורשות תואר אקדמי (College Jobs), משרות שאינן דורשות תואר אקדמי (Non-College Jobs), ומשרות שירות בשכר נמוך (Low Wage Service Jobs). המטרה היא לאפשר למשתמש להבין את הדינמיקה של שוק העבודה עבור כל קטגוריה של תארים ולהשוות בין סוגי המשרות הזמינות לבוגרי התחומים השונים. עיצוב זה יסייע לזהות אילו תחומים מספקים יותר משרות הדורשות תואר, דבר רצוי עבור אדם שסיים ללמוד תואר אקדמי, לעומת תחומים שבהם רוב המשרות הן שירותיות או אינן דורשות תואר.

b. במונחים של Munzner:

Action	Target	Explanation
Present	Distributions	נציג את התפלגות המשרות עבור כל קטגוריות התארים.
Compare	Distributions	נציג את ההתפלגויות עבור כל קטגוריות התארים כך שתתאפשר השוואה בין תארים שונים.
Identify	Trends	נוכל לזהות מגמות בין סוגי המשרות (לדוגמה: בכל הקטגוריות, יש פחות משרות שירות מאשר משרות אחרות)
Identify	Similarity	נוכל לזהות דמיון בהתפלגויות המשרות בין קטגוריות שונות של תארים

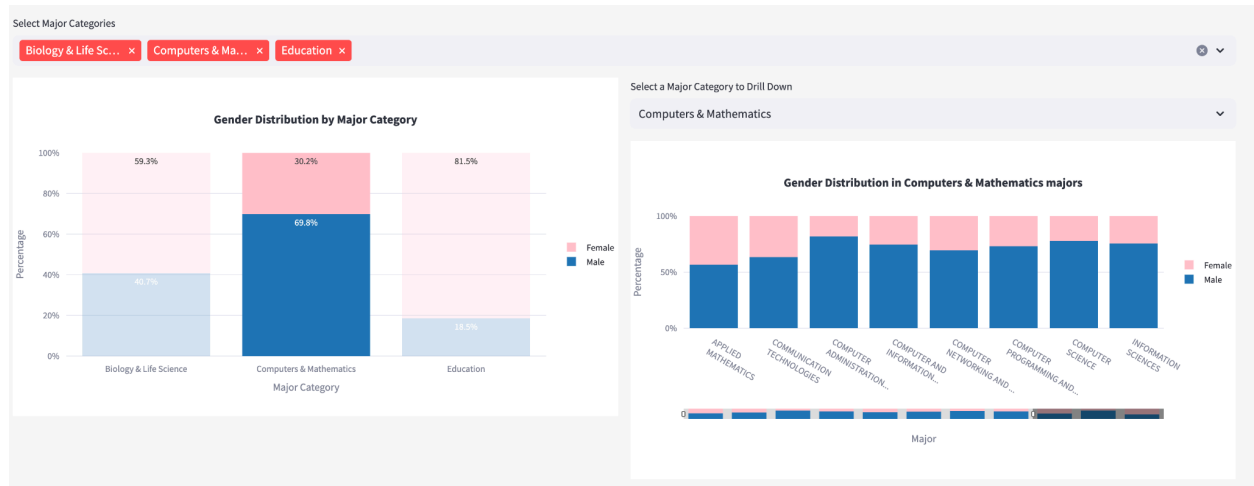
## עיצובים חלופיים

ויזואליזציה ראשונה: ניתוח הרכב מגדרי של קטגוריות של תארים ושל תארים המשתייכים לקטגוריה

מסוימת

העיצוב הראשון:

- סקיצה של הגרף - normalized stacked bar chart



- Marks

- vertical stack of line marks

- Channels

- Color hue (per gender)
- Spatial region (one per major category / major on x-axis)
- Length of internal bar component (for the gender)

○ אקספרסיביות

■ המידע המוצג מכיל 2 תכונות קטגוריות ששתיהן מוצגות ע"י ערוצי זהות (מיקום וצבע) ועוד תכונה כמותית המוצגת ע"י ערוץ עוצמה (אורך).

○ אפקטיביות

■ הגרף מאפשר זיהוי והשוואה ברורה של מגמות מגדריות בין תארים שונים. הערוצים שנבחרו הם האפקטיבים ביותר לערוצי זהות, וגם אורך הוא ערוץ עוצמה אפקטיבי.

• יתרונות

○ כיוון שיש רק 2 מינים, החלוקה בתוך כל BAR היא ברורה ואין עומס בתוך כל BAR.

○ טוב להצגת פרופורציה וגם להשוואת פרופורציות

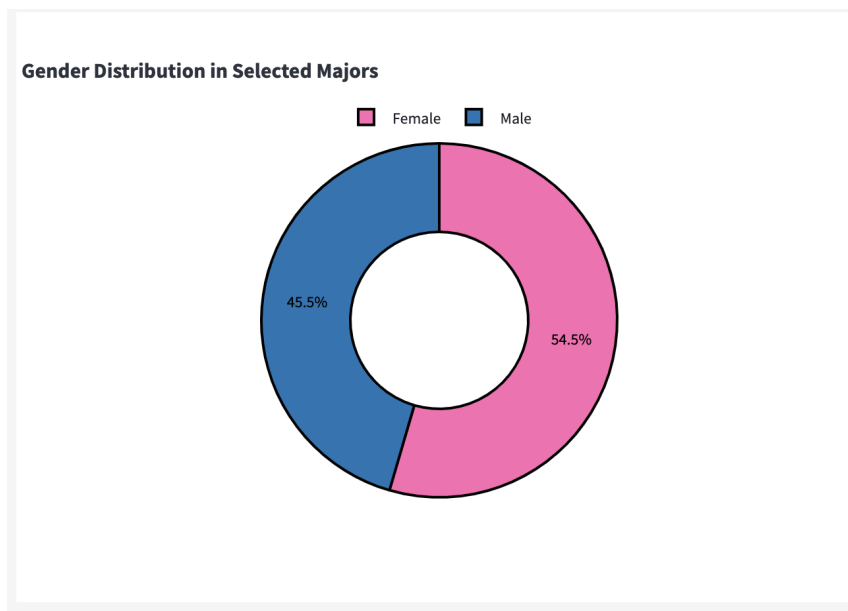
- מאפשר להציג הרבה התפלגויות במקום אחד ולהשוות ביניהם.
- מימדים שהם separable במידה מספיקה (צבע + מיקום)
- מיקום על aligned scale והשוואה קלה יותר

#### ● חסרונות

- מוגבל בהצגת גודלה המוחלט של האוכלוסייה בכל BAR.
- קשה להבחין בין שינוי קטן מאוד בין התפלגויות.

#### העיצוב השני:

- סקיצה של הגרף - pie chart



- Marks
  - interlocking area (proportion for male/female)
- Channels
  - Color (per gender)
  - Angle

- אקספרסיביות

■ הגרף מציג בצורה מדויקת את ההתפלגות המגדרית לתואר בודד. המידע המוצג מכיל 2 תכונות קטגוריות ששתייהן מוצגות ע"י ערוץ זהות (צבע) ועוד תכונה כמותית המוצגת



ע"י ערוץ עוצמה (זווית). עם זאת, כיוון שהגרף רדיאלי, שימוש בזווית כערוץ הוא באופן אינהרנטי לדאטה מחזורי, ולכן פחות מתאים לנתונים המדוברים.

○ אפקטיביות

■ הגרף מאפשר זיהוי והשוואה ברורה של מגמות מגדריות לתואר בודד. הערוץ של צבע הוא האפקטיבי ביותר לערוצי זהות, אבל זווית אינו בערוץ עוצמה הכי אפקטיבי ונתפש פחות טוב ע"י בני אדם.

● יתרונות

- פשוט ומהיר להבין.
- טוב להצגת פרופורציה.
- כיוון שיש רק 2 מינים, החלוקה ברורה וניתן להבדיל בקלות בין כל חלק והצבע שלו.

● חסרונות

- תפיסת מקום רב בדשבורד, כל פאי מייצג רק תואר/קטגוריה בודד/ת.
- שטח הוא פחות מדויק מאורך של קו rectilinear.

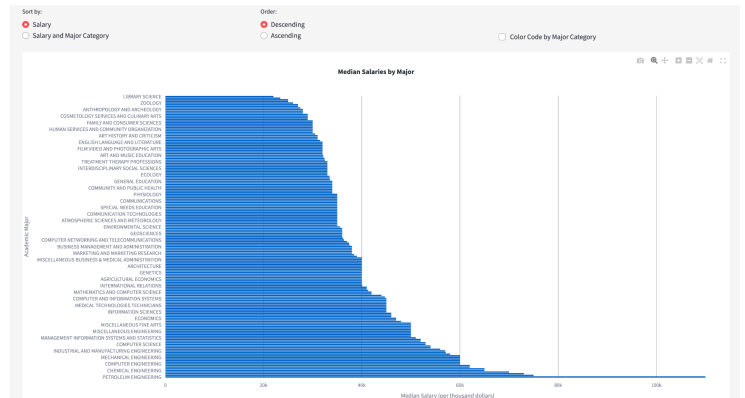
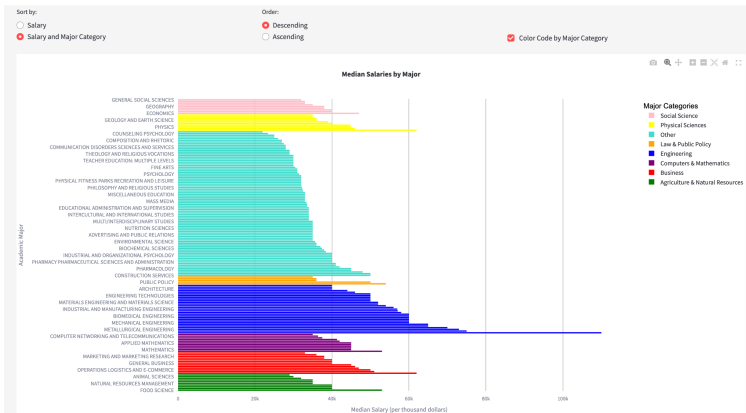
### סיכום ובחירה של עיצוב

בחרנו בעיצוב הראשון ( **Normalized Stacked Bar Chart** ) בשל האפקטיביות הגבוהה שלו והחיסכון שלו במקום. גרף מסוג זה מאפשר הצגה ברורה ומדויקת של התפלגויות ויחסים בין קטגוריות ותארים. בניגוד ל-Pie Chart, ה-Normalized Stacked Bar Chart מציג את הנתונים בצורה מסודרת על ציר רקטילינארי עם בסיס משותף, מה שמקל על השוואה בין חלקים שונים ומגמות בין קטגוריות רבות. בנוסף, אילוצי המקום והרצון להציג כמה שיותר רכיבים של קטגוריות או תארים הפכו את הבחירה ב-Pie Chart לפחות מתאימה, שכן גרף זה מציג רק item אחד, ועל מנת להציג יותר מאחד היינו צריכים להקריב הרבה מקום בדשבורד.

## ויזואליזציה שנייה: ניתוח משכורות לפי תואר ולפי קטגוריה של תואר

### העיצוב הראשון:

- סקיצה של הגרף - horizontal bar chart



- Marks
  - Lines
- Channels
  - Spatial region (one per major on y-axis)
  - Color (per major category)
  - Length of line (for the median salary)

- אקספרסיביות

■ העיצוב הנבחר הוא מתאים לנתונים. Major ו-Major Category הם משתנים קטגוריים ולכן נבחר עבורם ערוצי זהות (מיקום, צבע). השכר החיצוני הוא משתנה כמותי ולכן נבחר עבורו ערוץ עוצמה (אורך)

- אפקטיביות

■ הגרף מאפשר זיהוי והשוואה ברורה של השכר בין תארים ומגמות בין קטגוריות של תארים. הערוצים שנבחרו הם האפקטיבים ביותר לערוצי זהות, וגם אורך הוא ערוץ עוצמה אפקטיבי. הצבע מאפשר להבדיל בקלות בין קטגוריות שונות של תארים, והאורך של ה-BAR גם מייצג בצורה ברורה מאוד את גודל השכר החיצוני של אותו התואר.

- יתרונות

- גמישות במיון הנתונים - המשתמש יכול לבחור לסדר את הנתונים לפי שכר בלבד או לפי שכר בשילוב עם קטגוריה.
- מסודר וממוין -מסייע למשתמשים לזהות במהירות את התארים המשתלמים ביותר
- מאפשר להציג את כלל המידע בגרף בודד
- מימדים שהם separable במידה מספיקה (צבע + מיקום)

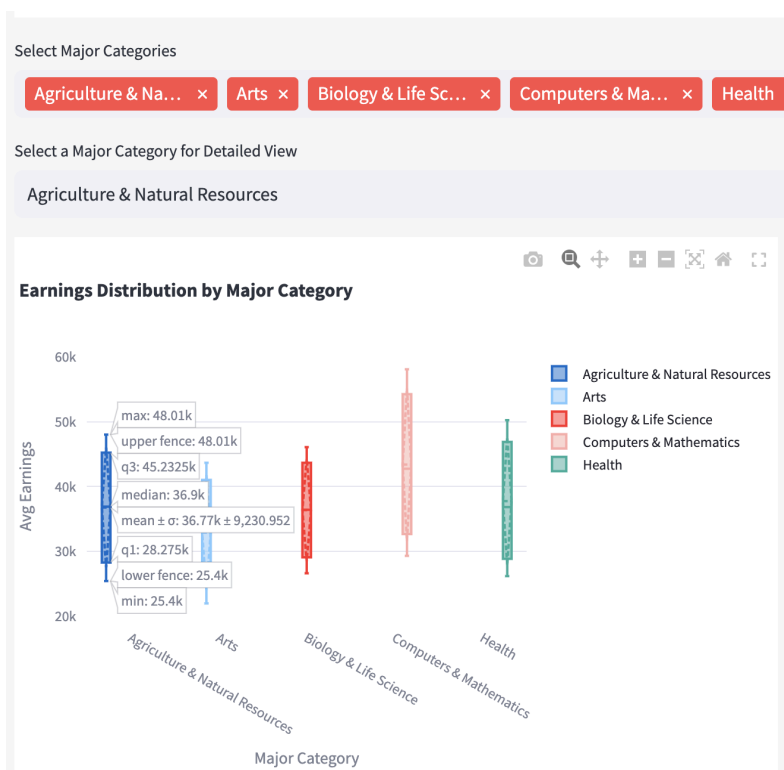
○ מיקום על aligned scale והשוואה קלה יותר בין אורכי הקווים

- חסרונות

- צפיפות בגרף, קושי בזיהוי תארים ספציפיים
- דורש פעולה נוספת של המשתמש (hover) כדי לגלות פרטים מדויקים על אותו תואר
- הגבלה במספר ה bins של צבעים שניתן להגדיר ולכן קיבוץ כמה קטגוריות לקטגוריית 'other'

### העיצוב השני:

- סקיצה של הגרף - box plot



- Marks
  - Points
  - Lines
- Channels
  - Spatial Region (one per major on x-axis)
  - Color (per major)
  - Length of bar, length of bar + whiskers

#### ○ אקספרסיביות

- העיצוב מתאים לנתונים. Major Category הוא משתנה קטגורי ולכן נבחר עבורו ערוצי זהות (מיקום וגם בנוסף צבע). הנתונים של ההתפלגות (אחוזונים, קצוות) משתנים כמותיים ולכן מסומנים ע"י ערוצי עוצמה כמו מיקום על הסקאלה ואורך הקווים.

#### ○ אפקטיביות

- האורך של הקווין מעיד באופן ברור מאוד את הטווחים של כל קטגוריה. הצבע אינו חובה, אך עוזר להבדיל בין קטגוריות כיוון שהשמות בציר ה'x' יכולים להיות צפופים עבור הרבה items. החצון ברור, אבל אינו מה שהכי בולט בגרף, כפי שהיינו רוצים לפי המטלה, ולכן מוריד את אפקטיביות עיצוב.

#### ● יתרונות

- גמישות במיון הנתונים - המשתמש יכול לבחור לסדר את הנתונים לפי שכר
- מאפשר להציג הרבה נתונים על item באובייקט יחסית פשוט
- גרף מוכר וקל להבנה
- מנצל הרבה מהנתונים הזמינים

#### ● חסרונות

- מציג מידע פחות רלוונטי למשתמש - לדוגמה, לרוב לא מתעניינים באחוזון ה'75' של טווח של משכורות
- את הקצוות ניתן לזהות ע"י גרף זה, אך לא לדעת לאיזה תואר הן שייכות
- לא מאפשר להשוות בין תארים ספציפיים, רק ביו קטגוריות
- עומס קוגניטיבי בפעולת hover, מוצגים הרבה מאוד נתונים

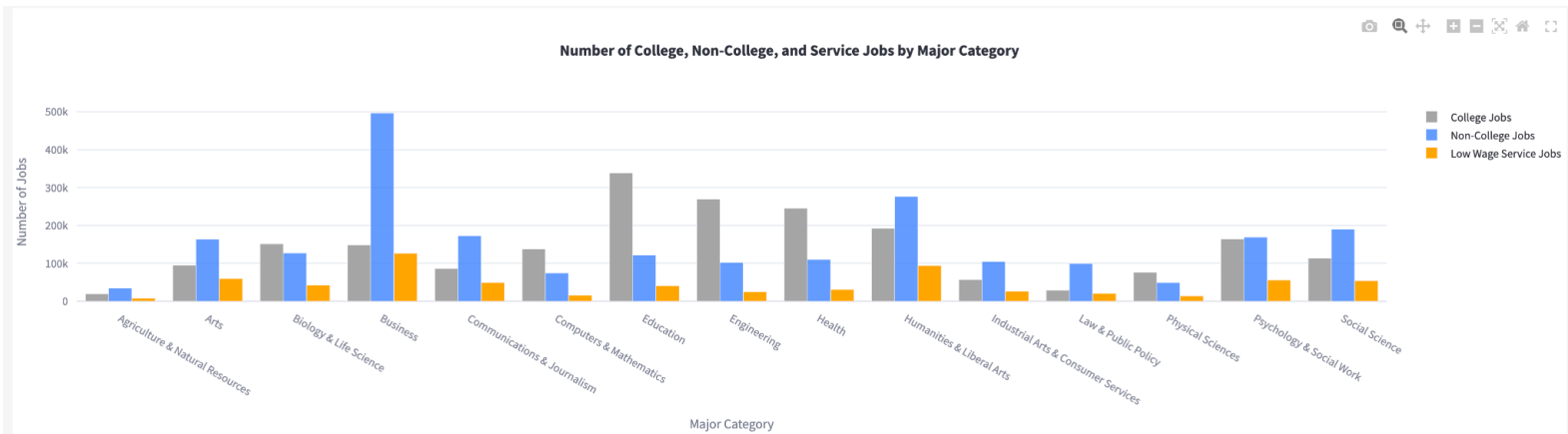
#### סיכום ובחירה של עיצוב

בחרנו בעיצוב הראשון שהוצג (horizontal bar chart) ולא ב-Box Plot, מכיוון שהוא מאפשר להציג את כל הנתונים בגרף אחד בצורה מסודרת וברורה, תוך בליטה של המטלה העיקרית - הצגת השכר החציון. בעיצוב זה, ניתן להשוות גם בין תארים ספציפיים וגם בין קטגוריות תארים שונות, מה שמספק מבט רחב ומעמיק יותר על הנתונים. הגרף מעניק גמישות גבוהה יותר למשתמש בזכות היכולת לבחור את סדר המיון, לסנן לפי שכר או קטגוריה, ולהתאים את הצפייה בהתאם לצרכיו. יתרה מזאת, הוא מנצל את הערוצים שלו בצורה יותר אפקטיבית. Box Plot, לעומת זאת, מציג מגבלות בתצוגת מבנה הנתונים הפנימי ואינו מספק את אותה רמת גמישות אינטראקטיבית, מה שמגביל את היכולת של המשתמש לקבל תובנות מעמיקות ורחבות יותר ממבט ראשון. לסיים, המידע המוצג ב-Box Plot פחות רלוונטי למשתמשים שמחפשים תובנות ברורות ופשוטות על שכר ותעסוקה.

## ויזואליזציה שלישית: ניתוח סוגי משרות לפי קטגוריה של תואר

### העיצוב הראשון:

- סקיצה של הגרף - grouped bar chart



- Marks

- Lines

- Channels

- Spatial Region (one per major category on x-axis)
- Color (per job type)
- Length of bar (for number of jobs for that job type)

- אקספרסיביות

- העיצוב הנבחר הוא מתאים לנתונים. Job Type ו-Major Category משתנים קטגוריים ולכן נבחר עבורם ערוצי זהות (מיקום, צבע). מספר המשרות הוא משתנה כמותי ולכן נבחר עבורו ערוץ עוצמה (אורך).

- אפקטיביות

- הגרף מאפשר זיהוי והשוואה ברורה של כמות המשרות מכל סוג לכל קטגוריה של תואר. הערוצים שנבחרו הם האפקטיבים ביותר לערוצי זהות, וגם אורך הוא ערוץ עוצמה אפקטיבי. המיקום והשם של הקטגוריה מובנים כאשר מסתכלים על הגרף, והצבע מאפשר להבדיל בקלות בין סוגי משרות שונים, והאורך של ה-BAR גם מייצג בצורה ברורה מאוד את כמות המשרות מסוג זה, כך שקו ארוך יותר מייצג יותר משרות וההיפך.

- יתרונות

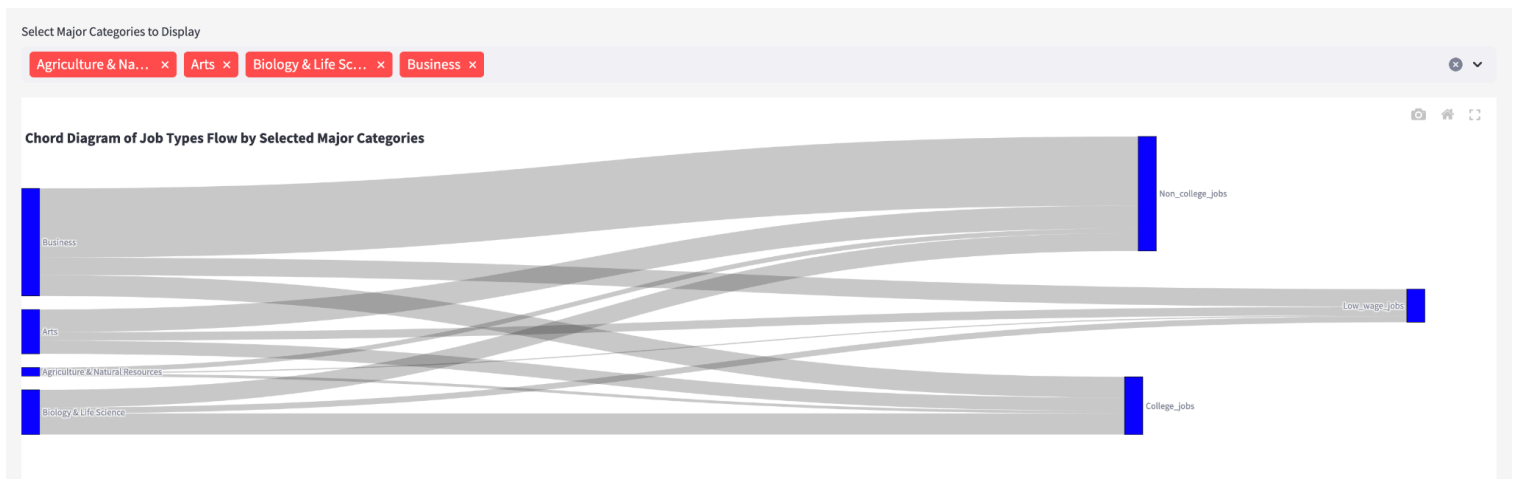
- גרף פשוט ואינטואיטיבי
- מאפשר להציג את כלל המידע (כל הקטגוריות וכל סוגי המשרה) בגרף אחד

- חסרונות

- צפיפות מסוימת בגרף
  - לא מאפשר ניתוח מעמיק יותר (drill down) ברמה של תואר ספציפי
  - קשה לזהות הבדלים קטנים במספרי המשרות
- הצגת הכמות המדויקת של משרות על ידי פעולת hover מקטינה את חומרת בעיה זו

### העיצוב השני:

- סקיצה של הגרף - sankey diagram



- Marks
  - Nodes (per major category and job type, in blue in the diagram above)
  - Links (connecting between category and type of job)
- Channels
  - Position
  - Direction (left to right)
  - Width of link

#### ○ אקספרסיביות

- העיצוב מתאים לנתונים Job Type ו-Major Category משתנים קטגוריים ולכן נבחר עבורם ערוצי זהות (מיקום, כיוון). מספר המשרות הוא משתנה כמותי ולכן נבחר עבורו ערוץ עוצמה (עובי).

#### ○ אפקטיביות

- הגרף מאפשר זיהוי החלוקה של סוגי משרות עבור על קטגוריות של תואר. ערוץ העובי פחות אפקטיבי מערוץ האורך בזיהוי כמות המשרות ולכן קשה יותר להבדיל בין עובי הקווים. הקשרין בין הקודקודים חופפים לכן מאבדים אפקטיביות באבחנה של כל קשר בודד.

#### ● יתרונות

- מאפשר להציג את הרכב סוגי המשרות פר קטגוריה, וגם את הרכב קטגוריות שונות פר סוג משרה.
- הצגה טובה בהצגת קבוצות שונות
- מאפשר למשתמש להציג את המידע שהוא רוצה בלבד (אינטראקטיבי)

#### ● חסרונות

- אינו מאפשר להציג את כלל המידע (כל הקטגוריות וכל סוגי המשרה) בגרף אחד (צפוף מדי)
- לא מאפשר ניתוח מעמיק יותר (drill down) ברמה של תואר ספציפי
- קשה להבחין באופן מדויק את עובי הקווים ולהבחין בין מקור ויעד של קווים ספציפיים
- יחסית מורכב להבנה

#### סיכום ובחירה של עיצוב

בחרנו בעיצוב הראשון של **Grouped Bar Chart** מאחר והוא מציג את הנתונים בצורה פשוטה, אינטואיטיבית וברורה, ומאפשר להשוות בקלות את כמות המשרות לפי סוג וקטגוריית תואר. העיצוב מנצל בצורה מיטבית את ערוצי המידע, כאשר המיקום והצבע משמשים לזיהוי קטגוריות וסוגי משרות, ואורך הבר מייצג בצורה ברורה את מספר המשרות, מה שמקל על המשתמש להבין את הדינמיקה בשוק העבודה בכל קטגוריה של תואר. בעוד שה-Sankey Diagram מציע דרך אסתטית להציג את הקשרים בין סוגי משרות וקטגוריות תארים, הוא פחות אפקטיבי בהצגת כמות מדויקת של משרות וקשה להבחין בעובי הקווים, מה שמקשה על השוואה בין סוגי משרות. בנוסף, ה-Sankey דורש הבנה מורכבת יותר ועלול להיות צפוף ולא קריא כשהמידע רב. ה-Grouped Bar Chart, לעומת זאת, מציג את כל המידע הדרוש בגרף אחד, שומר על קריאות גם כשיש ריבוי קטגוריות ומספק הבנה ברורה ומיידית, מה שהופך אותו לעיצוב המועדף לפרויקט זה.

## הסבר על העיצובים שנבחרו ועל יישומם

לפרויקט זה לא נעשה כל עיבוד מקדים של הנתונים.

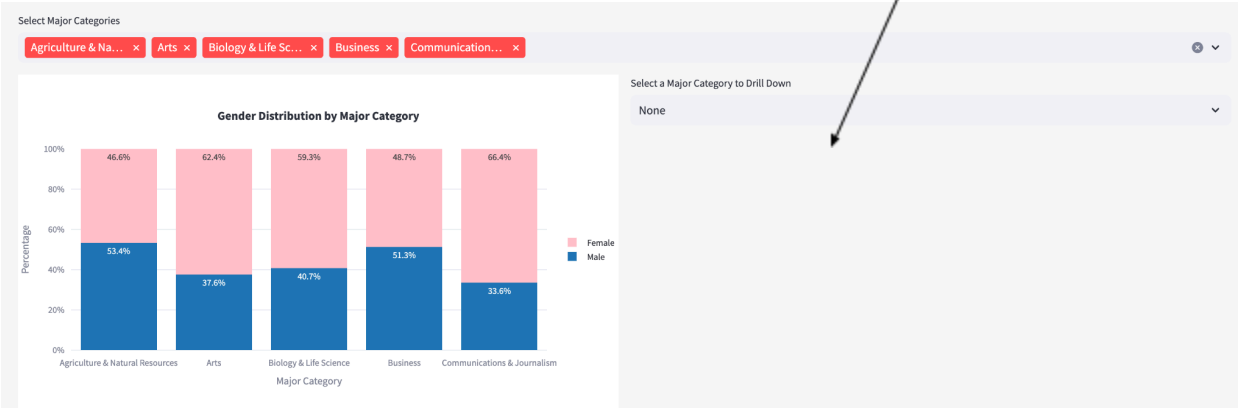
### הגרף הראשון

בעיצוב זה, המשתמש מתחיל עם גרף המציג את הרכב המגדרי בין קטגוריות של תארים שונים. כברירת מחדל מוצגים חמישה קטגוריות, אך המשתמש יכול לבחור עד שמונה קטגוריות שונות של תארים מתוך רשימה ממוינת בסדר לקסיקוגרפי הכוללת את כל הקטגוריות. לאחר בחירת הקטגוריות הרצויות, המערכת מציגה Normalized Stacked Bar Chart, הממחיש את התפלגות המגדר והשכר בתוך כל קטגוריה נבחרת. גרף זה מציג בצורה יחסית את החלק של כל מגדר בתארים השונים, ומאפשר למשתמש לראות את ההבדלים והמגמות בתוך הקטגוריות שנבחרו, תוך השוואה ברורה ומובנית של כל תואר בתוך הקבוצה שנבחרה.

בתחילת התהליך, לא הגבלנו את מספר הקטגוריות שניתן להציג, אך גילינו שהצגת קטגוריות רבות מדי יצרה צפיפות בגרף ופגעה בקריאות המידע. בעקבות זאת, החלטנו להגביל את הבחירה לעד שמונה קטגוריות בלבד. בזכות הבחירה להגביל את מספר הקטגוריות, הוספנו גם תצוגת אחוזים על גבי הבר עצמו, מה שמאפשר למשתמש להבין בצורה מהירה יותר את היחס חוסך לו פעולת hover כדי לגלות את האחוז המדויק. מבחינת הצבעים, בחרנו בצבעים בטוחים לעיוור צבעים וגם התאימו לנושא המגדר. גרפים בשתי הרמות מאפשרים הצגה של פרטים מדויקים (מספר אנשים באוכלוסייה, והאחוז המדויק) על ידי hover.

### הוספת קטגוריות להצגה

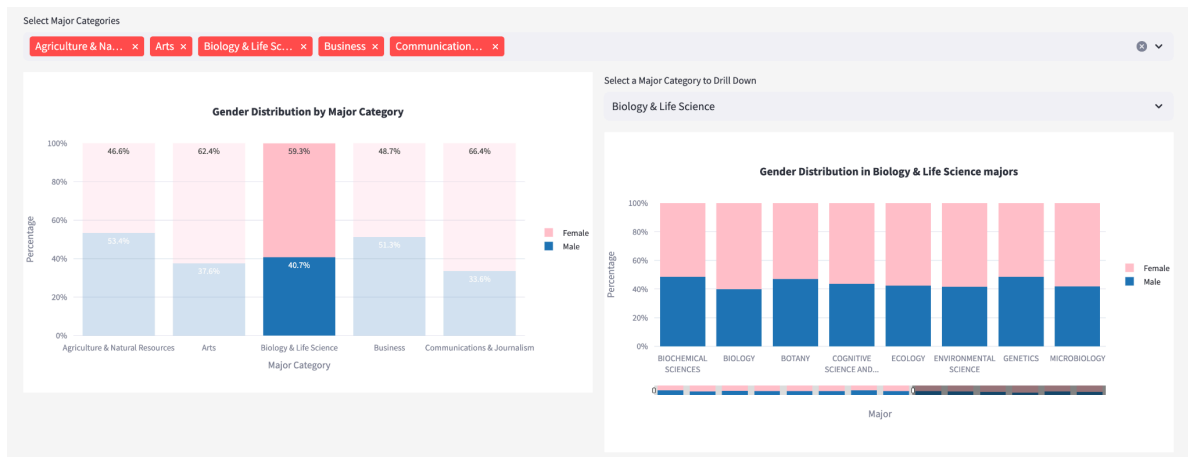
### בחירת קטגוריה אחת לביצוע drill down



Drill Down כדי להציג את התפלגות המגדר גם עבור תארים ספציפיים המשתייכים לקטגוריה מסוימת שנבחרה. מכיוון שיש קטגוריות מסוימות עם מספר רב של תארים, בחרנו לממש **slider** בציר האיקס המאפשר למשתמשים לשלוט בכמות התארים הנראים בכל רגע. הגדרה זו מגיעה עם ברירת מחדל שתוכננה להבטיח תצוגה ברורה



ונוחה להבנה. בשל הצפיפות הגבוהה המתקבלת בתצוגות רחבות יותר, הפרטים המדויקים של כל תואר נגישים רק באמצעות **hover**, המאפשר חשיפה נוחה של מידע נוסף לפי הצורך. הוספנו גם מימוש של **linking and brushing** כאשר נבחר לעשות drill down על קטגוריה מסוימת - דיכוי הצבעים בשאר הקטגוריות בגרף המקורי.



## הגרף השני

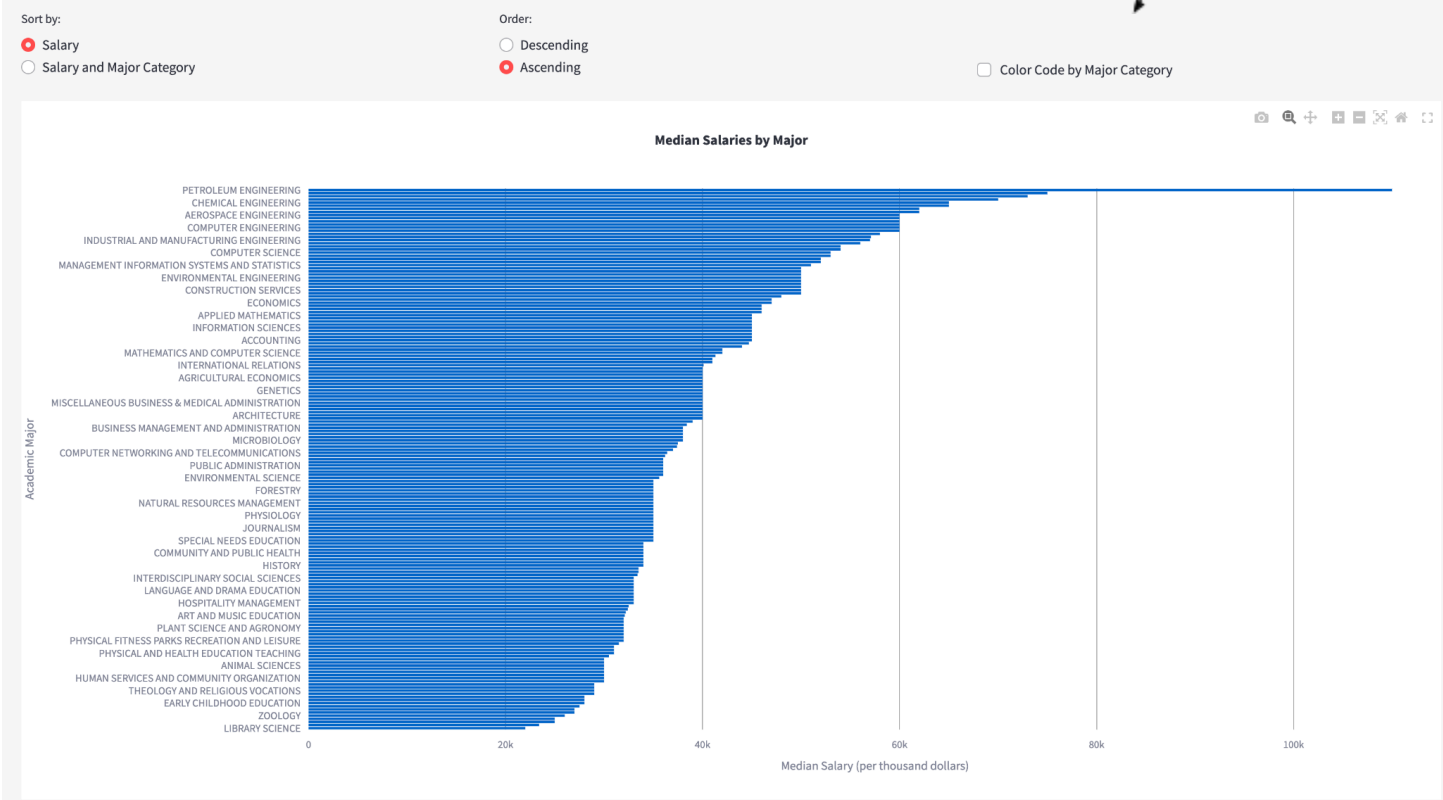
הגרף מציג את השכר החציוני לפי תארים אקדמיים שונים ומאפשר למשתמש לחקור את הקשר בין סוג התואר והשכר החציוני של בוגריו. כל בר מייצג תואר ומראה את גובה השכר החציוני לאותו תואר. המשתמש יכול להשתמש באפשרויות מיון הממוקמות בחלק העליון של הגרף כדי לסדר את הנתונים בסדר עולה או יורד. כמו כן, לבחור האם למיין לפי שכר בלבד או לפי שכר וקטגוריית התואר במקביל. מיונים אלו מאפשרים למשתמש לזהות את המשכורות הגבוהות/נמוכות בקלות.

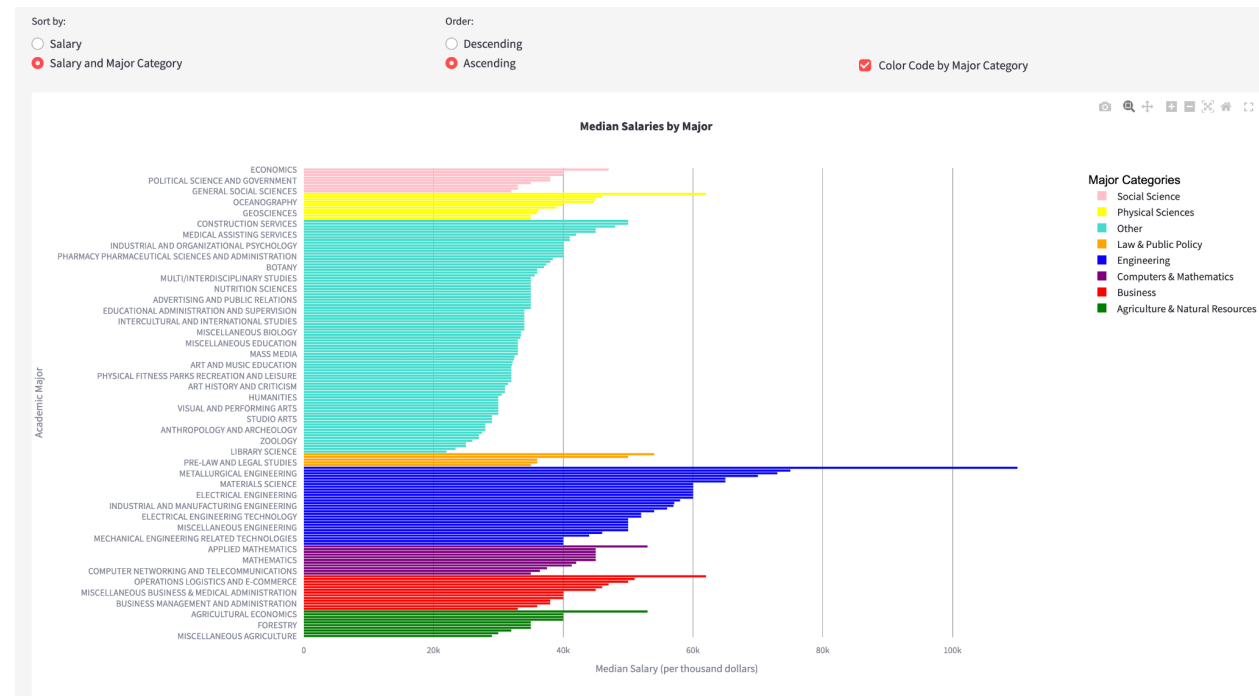
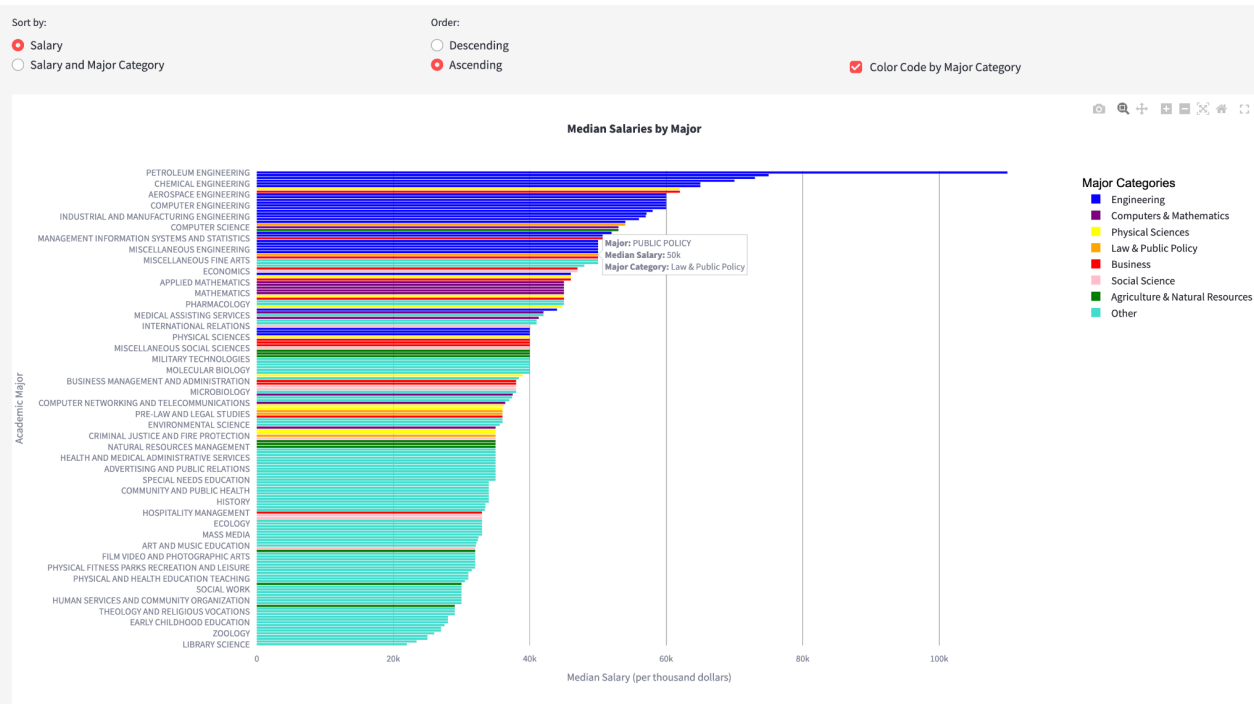
בנוסף, ניתן להפעיל אפשרות של קידוד צבעים לפי קטגוריה, מה שמקל על זיהוי חזותי של תחומים שונים, לדוגמה, הבחנה בין תארים בתחום ההנדסה, מדעי החברה, או מדעי הרוח. השימוש בגרף מאפשר למשתמש לקבל הבנה מעמיקה ומיידית לגבי אילו תארים הם המשתלמים ביותר מבחינת שכר ולהשוות בין הקטגוריות השונות בצורה ברורה ונגישה.

בחירת מיון על פי שכר  
בלבד או שכר וגם קטגוריה  
של תואר

בחירת מיון בסדר עולה או  
יורד

הפעלת קידוד צבעים

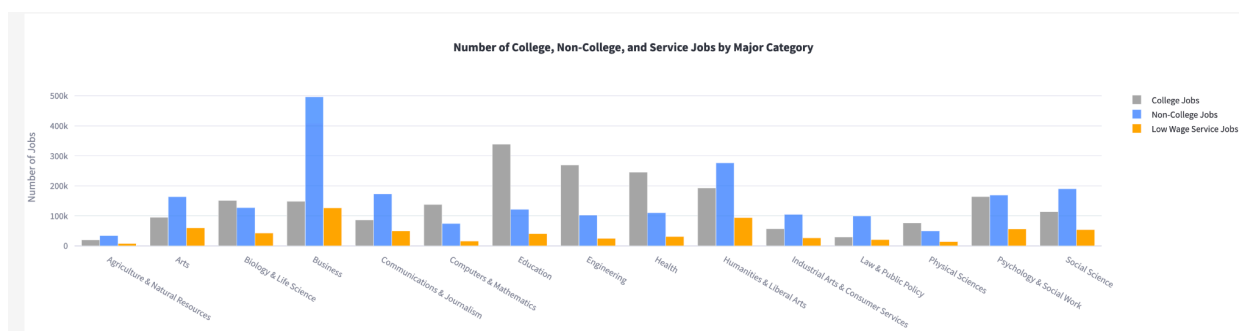




במהלך בניית הדאשבורד, נתקלנו בקושי לצמצם את גודל הגרף כך שייכנס כולו בדף אחד בצורה נוחה לקריאה. החלטנו לבצע את הצמצום הזה במקום להוסיף אופציה של גרף עם slider, מתוך רצון לשמור על פשטות הממשק ולמנוע עומס ויזואלי. בנוסף, מאחר שהגרף מציג תחומי לימודים רבים, נתקלנו בבעיה במספר הצבעים שניתן להוסיף כדי לייצג כל תואר בצורה ייחודית. כדי להתמודד עם מגבלה זו, החלטנו לקבץ מספר תארים (אלו עם הכי פחות בוגרים) תחת קטגוריה אחת של "Other", מה שהקל על הקושי להבדיל בין צבעים שונים. במקביל, הוספנו מידע ב-hover כך שהמשתמש יוכל לראות את קטגוריית התואר המדויקת גם כאשר היא מקובצת תחת "Other". הפתרון הזה שומר על קריאות ויזואלית תוך מתן גישה למידע מפורט בעת הצורך. בנוסף, ניתן לראות את השכר המדויק באמצעות hover.

### הגרף השלישי

הגרף הזה מציג את מספר המשרות השונות לפי קטגוריות תארים אקדמיים, תוך חלוקה לשלושה סוגי משרות: **College Jobs** (משרות הדורשות תואר אקדמי, מסומנות באפור), **Non-College Jobs** (משרות שאינן דורשות תואר אקדמי, מסומנות בכחול), ו-**Low Wage Service Jobs** (משרות שירות בשכר נמוך, מסומנות בכתום). כל קבוצת עמודות מייצגת קטגוריית תואר מסוימת, והעמודות בתוך כל קבוצה מציגות את מספר המשרות הקיים בכל סוג בתוך הקטגוריה. הקריאה בגרף נעשית על ידי בחינת גובה כל עמודה, המציין את מספר המשרות באליפים, מה שמאפשר למשתמש להשוות בין סוגי המשרות השונים בתוך כל קטגוריה ובין קטגוריות תארים שונות.



במקור, שקלנו להשתמש בעיצוב של **Small Multiples**, שבו כל סוג משרה וכל קטגוריית תואר הייתה מוצגת בגרף נפרד, מה שמאפשר ניתוח מפורט של סוגי המשרות בתוך כל קטגוריה. עם זאת, נתקלנו בבעיות במימוש גרף מסוג זה בצורה שתעמוד בסטנדרטים האסתטיים והקריאות שהצבנו לעצמנו. בנוסף, הבנו שבמקום להדגיש כל סוג משרה בנפרד, היה לנו חשוב יותר להציג את ההבדלים ברמת הקטגוריה של התארים עצמם, כדי שהמשתמשים יוכלו להשוות בקלות בין התחומים השונים ולראות את ההתפלגות הכוללת של המשרות. הבחירה בגרף הנוכחי אפשרה לנו לספק השוואה ברורה בין קטגוריות תארים בצורה ויזואלית שמדגישה את הדינמיקה בין סוגי המשרות בתוך כל תחום, תוך שמירה על עיצוב אסתטי ונגיש.

במהלך הניתוח המקדים של הנתונים, מצאנו כי קיימת קורלציה שלילית בין אחוז הנשים בחברה והשכר החציוני של בוגרי תארים שונים, תובנה שמצאנו חשובה להצגה ויזואלית. במקור תכננו להציג זאת באמצעות גרף **Scatter Plot**, כדי להמחיש בצורה ברורה את הקשר בין מגדר ושכר. עם זאת, גילינו שגרף כזה כבר קיים במערכת אחרת בקישור [הזה](#), והחלטנו להתרכז בהצגת תובנות ייחודיות נוספות בממשק שלנו.

#### הסבר על הקוד:

הקוד נועד ליצור אפליקציה אינטראקטיבית באמצעות Streamlit להצגת נתונים על תארים אקדמיים, תוך שימוש ב-Plotly ליצירת גרפים אינטראקטיביים. בתחילת הקוד, הוגדר עיצוב העמוד והתווספו כותרות ותיאורים לממשק המשתמש. בכל הקוד אנחנו משתמשים בספריות הבאות:

- **Streamlit**: לבניית ממשק משתמש אינטראקטיבי להצגת נתונים וגרפים.
- **Pandas**: לטעינת ועיבוד נתונים, כולל סינון וקיבוץ לצורך ניתוח.
- **Plotly Express & Graph Objects**: ליצירת גרפים אינטראקטיביים כמו Bar, Scatter ו-Pie Charts עם התאמה אישית.
- **NumPy**: לחישובים מספריים ויצירת נתונים סינתטיים.
- **Textwrap**: לעיטוף ועיצוב תוויות טקסטיות לגרפים.
- ספריות אלו מספקות כלים לניתוח והצגת נתונים באופן אינטראקטיבי וברור.