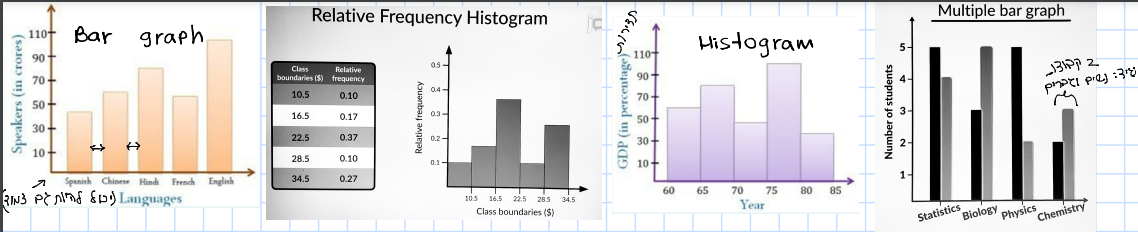
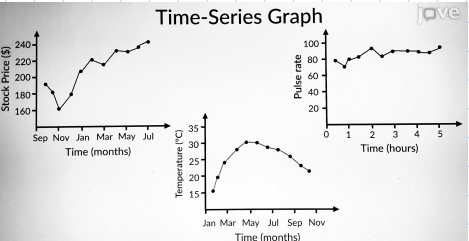
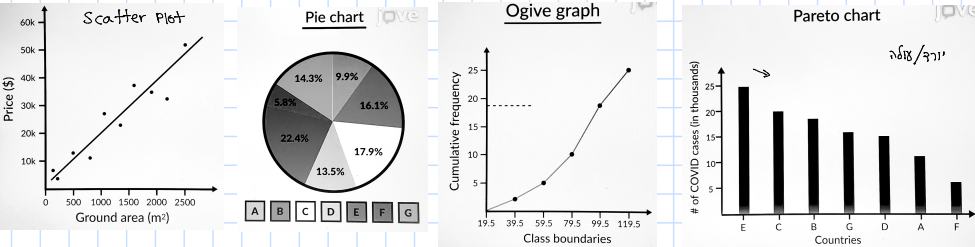
נתונים כמותיים מייצגים את המדידות או הספירות של ערכים מספריים, כגון גבהים משתנים של תלמידים בכיתה. לעומת זאת נתונים איכותיים- ידועים גם בשם נתונים קטגוריים, מייצגים משתנים לא מספריים, כמו צבעי שיער שונים. בשביל ניתוח סטטיסטי יעיל, מערכי נתונים לא מאורגנים, וגדולים אלה, מסוכמים ומיוצגים מספרית בצורת טבלה או בצורה גרפית ויזואלית. נטייה מרכזית- היא מדד סטטיסטי כדי לזהות ערך בודד בדר"כ, או נציג יחיד של הנתונים. ממוצע, חציון, שכיח ואמצע טווח הם ארבעת המדדים החיוניים של נטייה מרכזית. ממוצע אריתמטי- משמש לממוצע מערכי נתונים מספריים, מחושב ע"י חיבור כל הערכים במערך הנתונים וחלוקתם במספר הכולל של הערכים (n) (למשל זמן מחשב לחלק למספר תלמידים) .ממוצע נתוני המדגם= איקס גג, ממוצע נתוני כל האוכלוסיה=מיו.  
ממוצע גאומטרי- משמש לניתוח של נתונים הקשורים לכלכלה/ביולוגיה כאשר הערכים משתנים באופן אקספוננציאלי. מחושב כשורש מסדר n של מכפלת הערכים בתוך השורש. לא ניתן להשתמש אם יש ערך שלילי או אפס. ממוצע הרמוני- מחושב ע"י n לחלק להופכי של הנתונים (1 חלקי x )- משמש לחישוב הממוצע של יחסים,או קצבים-כמו מהירות רכב,יחס המחיר לרווח של חברה.אסור להשתמש אם אחד מהערכים אפס.

מתעלם מסטייה בסנתונים ע"י משקל רב יותר לערכים הקטנים יותר.ממוצע קטום- ללא הערך הגבוה ביותר והערך הנמוך ביותר. ממוצע משוקלל- כל ערך מוכפל במשקלו ומחלקים כל כלל המכפלות בסכום המשקלים. שורש ממוצע ריבועים/ממוצע ריבועי- משמש כאשר יש גם ערכים חיוביים וגם שליליים. כדי לחשב מעלים את כל הערכים בריבוע ואז מחברים אותם ומחלקים במספר הנתוניםn ועל התוצאה עושים שורש. הממוצע תמיד שווה או גדול יותר מהממוצע האריתמטי של ערכי הנתונים.ממוצע לטבלת שכיחויות-הכפלת כל ערך בתדירות שלו ולחבר את כל הערכים ולאחר מכן לחלק בסכום התדירויות. אם בטבלת השכיחויות יש טווחים אז מחשבים אמצע טווח ע"י הערך התחתון ועוד הערך העליון חלקי 2 ואת זה כופלים בתדירות.. שכיחmode-מדד של נטייה מרכזית, מוגדר כערך בעל השכיחות הגבוהה ביותר בנתונים. יכולים להיות כמה שכיחים(מולטי מודל),שכיח אחד, או ללא. טבלת שכיחויות(Frequency Distribution)- למשל טווחים של גובה וכמה אנשים יש בכל טווח. טבלה זאת עוזרת בהדמיית התפלגות של תלמידים לפי גובה. קל להבחין בשינויים בטבלה זו. חלוקה של התדירות של כל קטגוריה בסה"כ הכולל של המשתתפים = יחס השכיחות. הסכום של כל השכיחויות יהיה קרוב ל1. השכיחויות עוזרות לדעת כמה הטווח משפיע על מערך הנתונים. שכיחות האחוזים= מכפלה של השכיחות כפול 100. שכיחות מצטברת= שכיחות של כל קבוצה + השכיחות של הקבוצות לפנייה. למשל עבור הקבוצה השלישית השכיחות המצטברת היא 4+14+31  
גרף אוגיב((Ogive Graph - הצגה ויזואלית של השכיחויות המצטברות. גבולות המחלקה נמצאים לאורך ציר ה-X והשכיחויות המצטברות מופיעות לאורך ציר ה-Y. בגרף כל נקודה שמחוברת ע"י קו מייצגת את השכיחות המצטברת מתחת לגבול המחלקה המתאים. היסטוגרמיה ((Histogram- טבלאות שכיחות עוזרות לארגן נתונים כמותיים עם מספר קטגוריות, ניתנות לייצוג ויזואלי באמצעות היסטוגרמות. גרף המורכב מעמודות ללא רווחים. ציר הY מייצג את השכיחויות בכל מחלקה, ציר הX מייצג את הגבולות של המחלקה. מה הן גבולות המחלקה? מוסיפים חצי לכל גבול עליון. היסטוגרמת שכיחות יחסית(Relative Frequency Histogram)- טבלת התפלגות שכיחות של מספר השעונים שנמכרו בטווחי מחיר שונה המיוצגים כמרווחי מחלקה וגבולות בקרה(גבול עליון +0.5) טבלה זו מייצגת את הפרופורציות היחסיות של כל ערך כמותי במערך הנתונים. טבלאות כאלה מוצגות באמצעות גרף זה. כאשר ציר הY מייצג את השכיחויות היחסיות של כל מחלקה(יכול להיות כי יינתן גם באחוזים) וציר הX מייצג את גבולות המחלקה או נקודת האמצע של המחלקה. מהיסטוגרמות כאלה אפשר להבין באיזה תדירות מתרחש ערך כלשהו יחסית לאחרים במערך הנתונים. דיאגרמת פיזור (Scatter Plot)- נתונים כמותיים עם שני משתנים נקראים דו-משתנים(Bivariate Data) המשתנה שפועל **כגורם** נקרא ה**משתנה הבלתי תלוי**, והמשתנה האחר אשר מראה את ה**תגובה** נקרא ה**משתנה התלוי** . התלות הזאת של משתנה אחד במשתנה השני ניתנת להמחשה באמצעות דיאגרמת פיזור. את הקו מגמה יש לצייר שכמעט מספר זהה של נקודות נמצא מעל ומתחת לקו. נקודות אלה יחד יוצרות את התבנית כדי לזהות את המתאם בין שני המשתנים. מגמת עלייה מצביעה על מתאם חיובי, מגמת ירידה על מתאם שלילי, אם אין מגמה אז אין מתאם. גרף נתוני סדרת זמן (Time-Series Graph)-נתונים כמותיים שנאספים בנקודות זמן שונות נקראים נתוני סדרות זמן. מציר גרף הזמן יהיה על ציר הX, ועל ציר הY יהיה מה שאנו בודקים (למשל טמפ'). נסמן בנקודות ונחבר אותם בקו. דוגמאות נוספות -מחיר מניה של חברה, דופק, דוחות מזג אוויר. גרף עמודות (Bar Graph)-הצגת נתונים איכותיים באופן ויזואלי. הקטגוריה של הנתונים האיכותיים (למשל- קורסים שונים) מוצגים על ציר הX והשכיחויות מוצגות על ציר הY . ציור עמודות לכל קטגוריה בנפרד, ניתן עם רווחים וניתן בלי. גרף מרובה עמודות(Multiple Bar Graph)-השוואה של מערכי נתונים עם קטגוריות שונות (למשל בין מספר גברים ונשים בכל קורס) גרפים מרובי עמודות עוזרים להשוואת ולנתח דפוסים בקרב מספר קטגוריות. דיאגרמת פארטו (Pareto Chart)- כדי לדעת גבוה ביותר או נמוך ביותר, תרשים זה הוא גרף עמודות שונה, העמודות מסודרות בסדר **יורד** משמאל לימין. הקטגוריות מוצגות על הציר הX והשכיחויות על ציר הY. כלי שימושי להדמייה של מערכי נתונים גדולים. דיאגרמת עוגה (Pie Chart)- ייצוג נתונים איכותיים עם ספירת שכיחויות, גרף בצורת עיגול, בדוגמא שלנו-מספר מקרים קורנה בכל מדינה, לוקחים את מקרי הקורונה ומחשבים סהכ מקרי קורנה בכלל המדינות יחד, ומחשבים יחס של כל מדינה לסה"כ וכופלים ב360 מעלות כדי לדעת את הזווית של המדינה המתאימה.



סטטיסטיקה תיאורית-מאחר ולא ניתן יהיה אפשרי לצאת ולמדוד את כל קצב הלב של כל בני האדם נבחן רק תת קבוצה של האנשים או חלק מהאוכלוסייה המושגת=מדגם. הדופק הממוצע של המדגם נקרא הסטטיסטי. נתונים- מונח מדעי המשמש עבור אוסף תצפיות ומדידות מהווה את הבסיס לכל הניתוחים וההסקות הסטטיסטיות. ניתן לסווג נתונים על בסיס של האם ניתן למדוד אותם או שלא.למשל צבע שיער- לא ניתן למדוד אבל ניתן לחלק לקטגוריות (חום, שחור..) מערכי נתונים כאלה נקראים נתונים קטגורים או איכותיים: כלומר לא ניתנים למדידה או ספירה אך ניתנים לסיווג. נתונים שניתנים לספירה או למדידה ביחידות ספציפיות נקראים נתונים מספריים או כמותיים. הנתונים מסווגים לארבעה סולמות מדידה- נומינלי, סדר(סולם) , רווח ויחס. נומינלי- \* נתונים שלא ניתן למדוד או לסדר אך ניתן לקבץ לקטגוריה \* תשובות לסקר כמו – אוהב לא אוהב, כן לא, השתייכות למפלגה פוליטית  
סדר(אורדינלי)- \* אם ניתן לסדר על פי סדר מסוים (למשל דירוג סרטים מ1 עד 5 יוצר מערך נתונים מסודר של הציון הסופי) \*חולה/בריא \*אשם/חף מפשע  
רווח(אינטרוולי)- \*נתונים שאפשר לסדר ולהבדל בין הערכים יש משמעות . למשל הטמפ' במדבר נמדדת במעלות צלזיוס ויכולה להיות ממוינת מגבוה לנמוך והשוני מציין כמה הטמפרטורה משתנה ביום. יחס- הרחבה של קטגוריית רווח: עוסק בנתונים שיש להם נקודת אפס טבעית, ההבדל בין הערכים, שניהם בעלי משמעות למדידה על סולם רווח. למשל השוואה בין מחיר של ספר סטטיסטיקה לחוברת קומיקס. היחס בין הערכים הוא גם משמעותי- המחיר של ספר הסטטיסטיקה(יקר) הוא פי עשרה מזה של ספר הקומיקס.(הבדל מכפלה).בניתוח סטטיסטי, מאפייני האוכלוסייה נקבעים על סמך הנתונים שנאספו מהמדגם. ישנם מספר דרכים לאסוף נתונים ממדגם ולהסיק מהנתונים לכלל האוכלוסייה. המחקר התצפיתי הוא אחת השיטות הנפוצות ביותר של איסוף נתונים. מחקר עוקב- עוקב אחרי משתתפים תקופה ארוכה. המחקר לא מפריע לשגרה של המשתתפים. המחקר הניסויי הוא שיטה נפוצה לאיסוף נתונים. למשל ניסוי של השפעת השמש על הצמחים קבוצה אחת נחשפת לאור והשנייה לא. הדגימות עוברות מניפולציה לפני איסוף נתונים (עם שמש בלי שמש).סקר הוא אחד השיטות הנפוצות לאיסוף נתונים-בדרך כלל משתמשים בסקר על מנת לקבל את הפידבק של הלקוחות, להעריך את התקדמות הפרוייקט, להעריך את ציפיות הצרכנים או לקבל שביעות רצון של עובדים. שיטות דגימה משמשות לשליפת דגימות מאוכלוסייה, אשר מבטיחות שהדגימות נבחרות ללא הטיה וכי הן מייצגות במדויק את האוכלוסייה. שיטת הדגימה האקראית היא אחת השיטות הנפוצות לאיסוף נתונים. שיטות דגימה מתאימות מבטיחות שהדגימות מוצאות ללא הטיה ומייצגות את האוכלוסייה. דגימה שיטתית היא אחת מהשיטות הפשוטות ביותר של איסוף נתונים. הדגימה מבוססת אך ורק על המיקום ולא על מאפייני המדגם כגון צבע, מין. בדגימה שיטתית ניתן להשתמש רק כאשר החוקרים יודעים את המספר המדויק של האינדיבידואלים באוכלוסייה, כמו כך האוכלוסייה חייבת להתפלג באופן אקראי כדי להימנע מהטיה. דגימת נוחות היא שיטה נפוצה לאיסוף דגימות בה הדוגמאות נבחרות מאחר והן קלות להשגה. למשל- סקרים שנערכים רנדומלית בקמפוס, בקניון. יתרונות לשיטה: דוגמאות קלות להשגה, חסכונית, גוזלת פחות זמן. אפשר להשתמש בקבוצה זו כדי ליצור הערה או בכדי לבצע מחקר פיילוט. עם זאת, הדוגמאות מוטות ביותר ולא מייצגות את האוכלוסייה באופן טוב. שיטת הדגימה בשכבות משמשת בדרך כלל כשחוקרים אוכלוסייה הטרוגנית- אוכלוסייה עם וריאציות גדולות. ההטרוגניות של הקבוצה משפיעה על מה שאנו רוצים לבדוק. חסרון: לא ניתן להשתמש בה לשיטות מרוכבות.

דגימת אשכולות- כאשר מסגרת הדגימה גדולה במיוחד ומפזרת גאוגראפית או אין רשימה מסודרת. מחלקים את הקבוצה לאשכולות(תתי-קבוצות) ודוגמים אקראית חלק מהאשכולות .