Practice Assignment #1
Graphs

General Instructions

Read the following four scenarios and answer the questions. Please include your visualizations and narrative in this assignment document. Data for the questions are also attached in the Excel files. You may use any software to create the visualizations. Each of your visualizations will be evaluated based on the appropriateness of the chosen type, accuracy of information display, thoughtfulness of visual design, and inclusion of necessary graph components (e.g., title, labels, legends).

When you submit your assignment to Canvas, please make sure to submit both *this assignment document* with your answers, and *your visualization file* (e.g., Excel file, Tableau workbook, codes).

Scenarios and Questions:

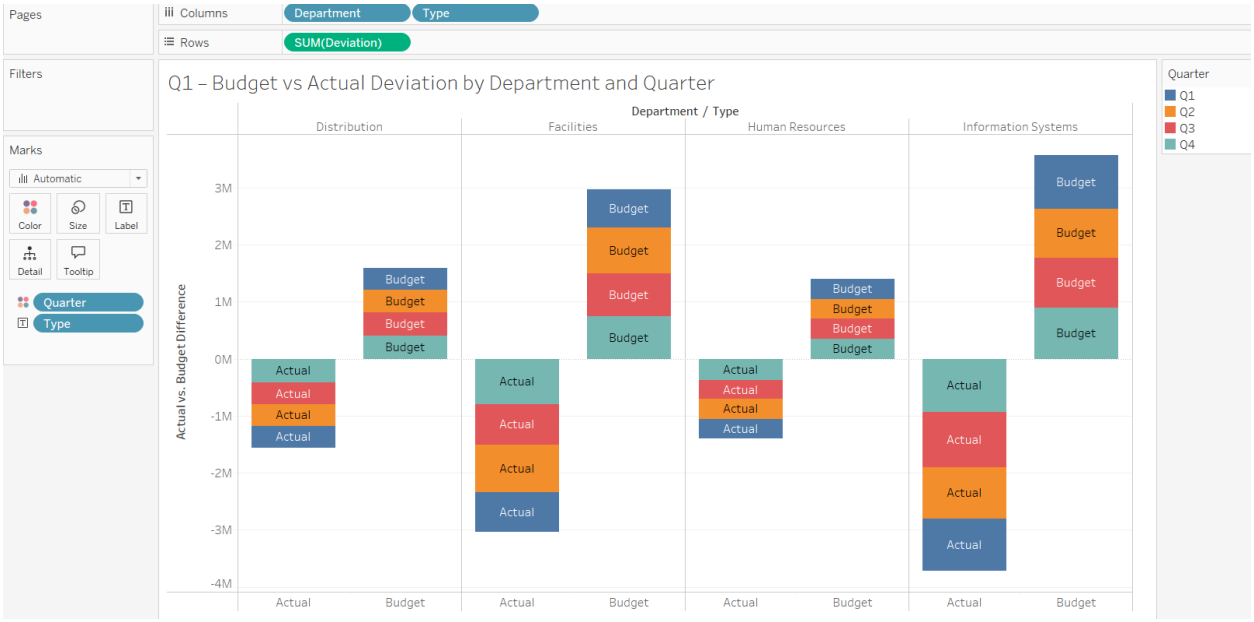
Q1. Imagine that you work in the Financial Planning and Analysis department. It is your job to report how actual quarterly expenses during the previous year compared to the budget for each of the five departments that report to the Vice President (VP) of Operations. The VP is only interested in the degree to which actual expenses deviated from budget, not the actual dollar amount. Here is the dataset:

	Q1		Q2		Q3		Q4	
Department	Budget	Actual	Budget	Actual	Budget	Actual	Budget	Actual
Distribution	390,000	375,000	395,000	382,000	400,000	390,000	410,000	408,000
Facilities	675,000	693,000	800,000	837,000	750,000	713,000	750,000	790,000
Human Resources	350,000	346,000	350,000	342,000	350,000	340,000	350,000	367,000
Information Systems	950,000	925,000	850,000	890,000	875,000	976,000	900,000	930,000

(Data also available in the dataset_1 Excel file)
Create one graph to display this information. Please include all necessary components for a graph, such as a title, legends, and labels.
After you complete the graph, use a couple of sentences to describe its design, including your rationale for each design feature, in the space below.

Ans: For this visualization, I created a **grouped stacked bar chart** to show the **deviation between actual and budgeted expenses across four quarters for each department.** More than wanting actual dollar figures, the Vice President wanted to know just how much the difference was, so I defined it and built a bar chart as a visualization with a y-axis showing budgeted amounts vs. actual amounts. Group by department and separate the bars indicating whether they are actual or budget to allow for an easy visual comparison side by side within each grouping.

Vaishnavi Pawar
vpawar@iu.edu



For easy identification of quarterly trends, I paid special attention to **color coding** for Q1 through Q4, making use of the "Quarter" field on the color mark. This facilitates the quick recognition of which quarters had high or low deviations. I labeled each stack with its type (Actual or Budget), picked a clear axis title, and included an overall title for orientation. This design powerfully showcases how deviations change not just with the departments but also the across-the-board time period - just what the VP was looking at for performance monitoring without throwing raw numbers at the viewer.

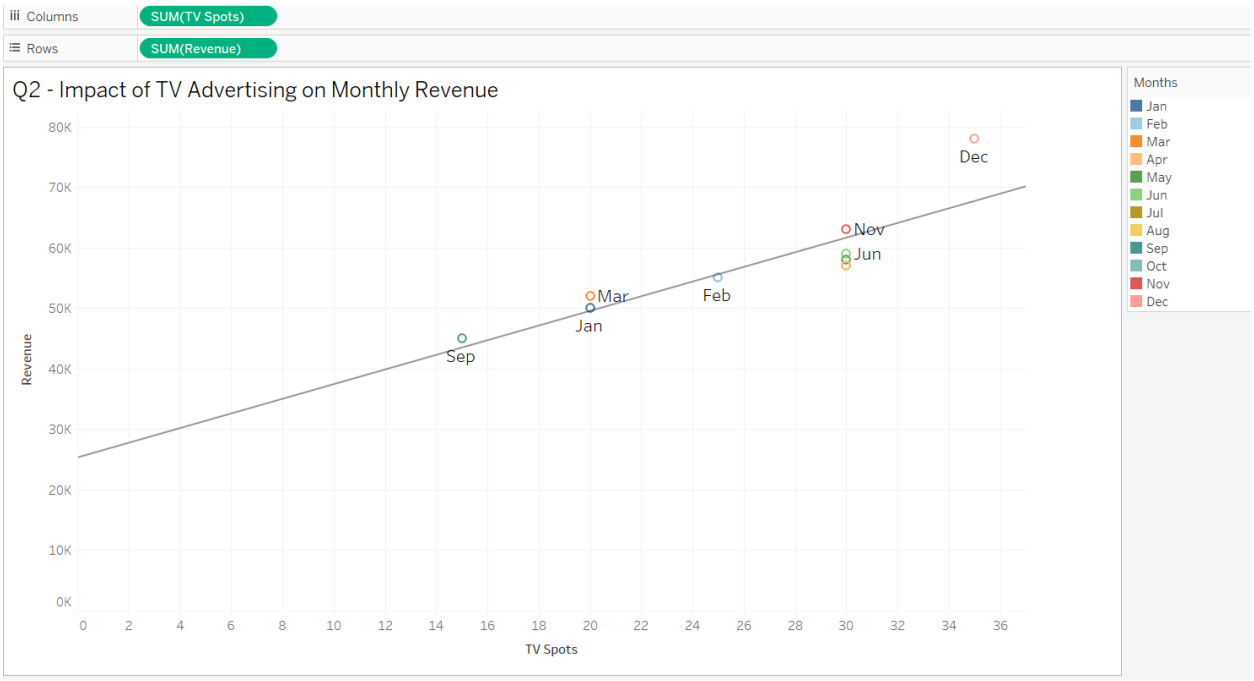
Q2. You are the company's marketing analyst. You have discovered through an analysis of last year's revenues that sales have significantly benefitted from the use of television ads. You need to present this benefit to the Chief Executive Officer (CEO). Here is the data:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
TV Spots	20	25	20	30	30	30	30	20	15	25	30	35
Revenue	50,000	55,000	52,000	57,000	58,000	59,000	57,000	50,000	45,000	55,000	63,000	78,000

(Data also available in the dataset_1 Excel file)
Create one graph to display this information. Please include all necessary components for a graph, such as a title, legends, and labels.
After you complete the graph, use a couple of sentences to describe its design, including your rationale for each design feature, in the space below.

Ans: I created **a scatter plot with Tableau** and superimposed **a trend line** on it to show the relationship between monthly revenues and expenditures on TV advertising. A scatter plot was chosen here specifically since it serves an effective purpose of testing the correlation between two quantitative variables, which in this case are the number of TV spots and the corresponding monthly revenues. Since the appeal of having the CEO comprehend the advantages of TV commercials, this visualization displays the positive correlation between ad frequency and sales in an intuitive manner.

Vaishnavi Pawar
vpawar@iu.edu

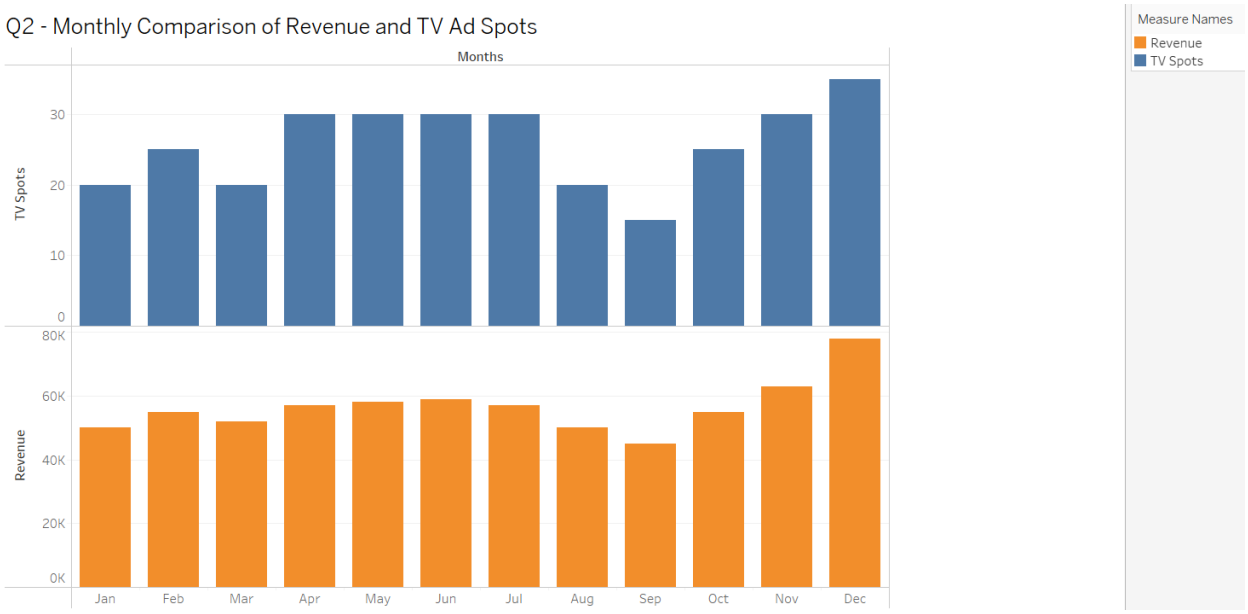


Here, each data point represents a month, which is labeled and color-coded for easy identification. This makes it easy to spot standout months and analyze yearly trends. The trend line reinforces the key message by portraying an undeniable upward path, validating what was seen: higher ad volume means generally higher revenue.

While interpreting the scatter plot, I discovered that some months appeared to be missing at first glance. After carefully studying the data, I understood that this was not an error, but rather the **result of overlapping data points**. Several months, including April, May, and July, had the same number of TV advertisements and very comparable revenue figures. Because scatter plots place points based on precise coordinates, those months wound up stacked on top of one another in the same position. Understanding this allowed me to validate that the visualization was correct and that the apparent "missing" months were merely due to how the data organically clustered, which is crucial to remember when presenting the research to others.

Before creating the scatter plot, I built a **table view (Q2-Table view in Workbook)** to ensure that the data was structured and accurate. I also created a **grouped bar chart (Q2-Bar chart in Workbook)** to compare monthly earnings and ad slots side by side. These exploratory visualizations helped to ensure that the data was clean, understandable, and ready for executive presentation.

Vaishnavi Pawar
vpawar@iu.edu



The final plot has all the necessary visual elements: title, labeled axes, and legend, with which it conveys the information clearly and gives a complete visual impression. The design intends to impart the information straightforwardly to assist in executive decision-making, consistent with the expectations of the assignment.

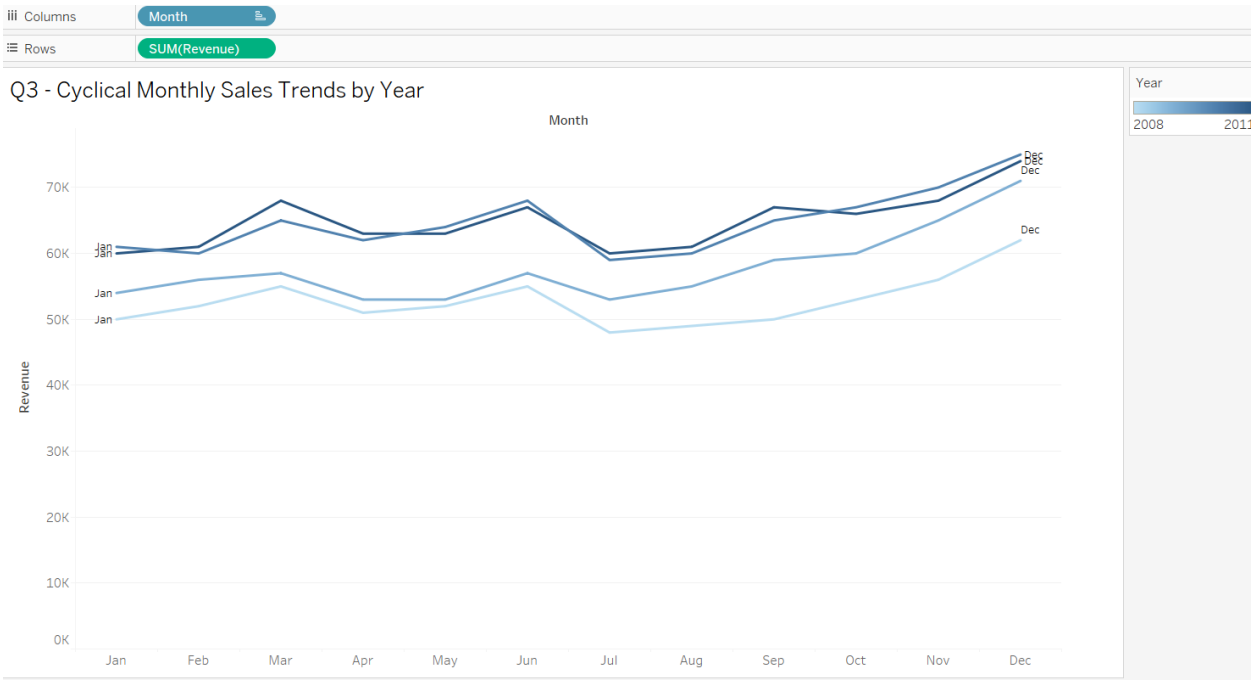
Q3. For this scenario, you are a Sale Analyst. Based on analysis of sales revenues for the past four years, you have noticed that there is a clear cyclical trend. The highest revenues are always generated during the last month of each quarter and the last quarter of each year, without exception. You want to use this information to point out the need for distributing sales activity more evenly across the year so that sales operations are not hit with such spikes in activity at quarter ends, especially at year ends. The following table contains the data in US dollars:

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	50,000	52,000	55,000	51,000	52,000	55,000	48,000	49,000	50,000	53,000	56,000	62,000
2009	54,000	56,000	57,000	53,000	53,000	57,000	53,000	55,000	59,000	60,000	65,000	71,000
2010	61,000	60,000	65,000	62,000	64,000	68,000	59,000	60,000	65,000	67,000	70,000	75,000
2011	60,000	61,000	68,000	63,000	63,000	67,000	60,000	61,000	67,000	66,000	68,000	74,000

(Data also available in the dataset_1 Excel file)
Create one graph to display this information. Please include all necessary components for a graph, such as a title, legends, and labels.
After you complete the graph, use a couple of sentences to describe its design, including your rationale for each design feature, in the space below.

Ans: To clearly demonstrate the cyclical nature of monthly sales over the past four years, I constructed a **multi-line chart** in Tableau with **Month on the x-axis and Revenue on the y-axis**, with Year separating each line. This style successfully captures seasonal revenue trends and allows for easy month-by-month comparisons across years, making it ideal for the sales force and management audience.

Vaishnavi Pawar
vpawar@iu.edu



I chose a line chart because it effectively depicts change over time and highlights cyclical patterns. Each line is color-coded by year using a diverging blue palette, with deeper colors indicating more recent years. This visual differentiation enables the spectator to rapidly discern annual trends and recognize consistent peaks in specific months. I added data labels for January and December to highlight the beginning and end of each year, which are often the times of greatest change.

Revenue consistently surges in the last month of each quarter, particularly in **March, June, September, and December.** This supports the business insight that sales activity is heavily weighted toward quarter ends, particularly at year-end. The visual makes this insight easy to observe without needing to dig into numbers manually.

A clear title, labeled axes, and a color legend indicating different years are all there in the graph, thus fulfilling the accreditations for effective communication. This really supports arguments for better spreading out of sales activities during the year instead of stressing operations, enabling the team to follow proactive strategies for their work.

Q4. Use the Indiana county-level statistics dataset (available in the dataset_2 Excel file) to create a dual-axis map that shows the total population change between 2010 and 2020 in Indiana and the housing utilization rate (i.e., the proportion of housing units that are occupied) in 2020 Indiana. Please keep in mind that you may need to clean or transform the original data for analysis and visualization purposes.

After you complete the visualization, use a couple of sentences to reflect on your map design, including your practices of cleaning or transforming data, and answer the two questions:
(1) which county had the largest positive population change between 2010 and 2020?;
(2) which counties had the highest and lowest housing utilization rates in 2020?

Ans:

To generate the dual-axis map, I first imported the **Data_2010 and Data_2020** sheets from the **dataset_2.xlsx file** into Tableau. Because the data I needed, total population and housing statistics, were distributed across both sheets, I merged **them using an inner join** on **the Name areaname field**, which was shared by both.

After joining, I created two calculated fields for the metrics I wanted to visualize: **Population Change (2010–2020):** Total Population (2020) – Total Population (2010) **Housing Utilization Rate (2020):** Occupied Housing Units / Total Housing Units

When the computations were complete, I began constructing the map. I modified the geographic function of the county name field to "County" and added it to the view to start creating a populated map. However, I soon encountered a problem: **numerous counties** were listed as **"unknown."**

To debug, I looked at the underlying data more attentively. I discovered that Tableau was failing to recognize numerous counties due to geographic recognition errors. At this point, I decided to **construct a new calculated field named County FIPS Full**, which integrated the **State FIPS and County FIPS codes** into a single unique geographic identifier:

RIGHT("00" + STR([State FIPS]), 2) + RIGHT("000" + STR([County FIPS]) 3)

I then changed the geographic role of this new County FIPS Full field to **"County",** which allowed Tableau to properly recognize and plot all 92 Indiana counties on the map.

While studying the map, I saw that Tableau had **wrongly included "Indiana County" from Pennsylvania**. Because this county is not in the state of Indiana, I manually **excluded it** using a filter, ensuring that only Indiana's counties were shown.

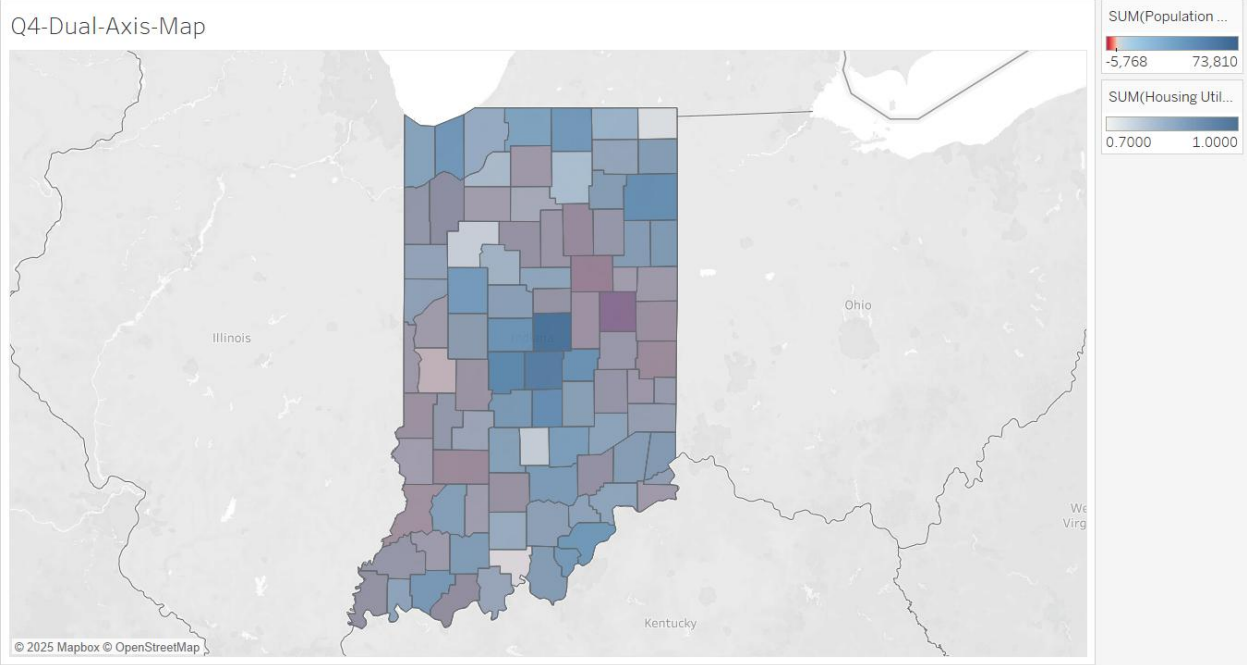
Once the geographic issues were resolved, I built the dual-axis map:

- The **first layer depicts Population Change** with a **diverging Red-Blue color palette, with red denoting population fall and blue indicating expansion.**
- The **second layer depicts the Housing Utilization Rate (2020)** as a filled map with **60% opacity,** so both values are visible and interpretable together.

Initially, I sought a way to manually synchronize the two maps, but I discovered that Tableau automatically synchronizes dual-axis maps as long as they use the same geographic field - in this example, the generated County FIPS Full. I verified this by hovering over several counties and confirming that both values appeared accurately in the tooltip.

- To improve readability and interactivity, I applied **SUM() aggregations** to both metrics.
- Tooltips are included to display the county name, population change, and housing usage rate.
- To separate the two data layers, I added a distinct title, legends and used intelligent color formatting.

Vaishnavi Pawar
vpawar@iu.edu



Using a dual-axis map, this visualization effectively illustrates demographic and housing trends while also demonstrating the procedures I took to clean, transform, and repair the data in Tableau. From merging sheets and constructing calculated fields to overcoming geographic recognition challenges and ensuring synchronization, every design decision contributes to the final result's clarity and correctness.

4.1] Which county had the largest positive population change between 2010 and 2020?
Ans:

Marion County (FIPS: 18097) experienced the greatest positive population growth, with a total increase of **73,810 individuals** between 2010 and 2020.

To validate this result, I used two methods:

- A simple bar chart (Q4-1 in workbook) rated all counties based on population change.
- The dual-axis map shows Marion County in the deepest shade on the population layer, signifying the greatest rise.

4.2] Which counties had the highest and lowest housing utilization rates in 2020?
Ans:

Hancock County (FIPS: 18059) has the **highest housing utilization rate, at 0.9599** (95.99%).
Steuben County (FIPS: 18151) has the **lowest housing utilization rate at 0.7308** (73.08%).

I double-checked these figures using an analysis of the housing utilization color layer on the dual-axis map a bar chart (Q4-2 in workbook), which showed Hancock and Steuben at opposite extremes of the spectrum.