

# DSCI 5360 – Data Visualization

## Assignment 2

---

### Instructions

1. This assignment is to be submitted online through Canvas, on or before the due date.
2. This assignment must be completed using the version of Tableau specified in the syllabus. I will not grade assignments completed with newer or older versions of the software.
3. Please name your file “Assignment2\_FirstName\_LastName.twbx” or “Assignment1\_Group\_GroupNumber” as appropriate. Please note the “.twbx” extension. I will be unable to grade files with the “.twb” extension.
4. If this assignment has been identified as an individual assignment, you are to work alone, without the help of other individuals. If this assignment has been identified as a group assignment, you are only to work the members of your group. In the latter case, I **STRONGLY SUGGEST** that each team member complete the assignment and then meet with other group members to identify the best possible answers for submission.
5. Please keep in mind course policies on academic dishonesty.
6. Your Tableau workbook should have separate worksheets clearly labeled as “Question 1,” “Question 2,” ... etc.
7. If a question asks you to do more than simply create a visualization (i.e., answer a specific question in text), please do so in the title of the worksheet or in a caption/annotation.
8. When asked to recreate a visualization, you should attempt to replicate it exactly in terms of data, color, marks, annotations, etc. The exception is axis labels, which are dynamically altered by Tableau based on the display size of the screen.
9. This assignment is to be completed without external assistance. You are welcome to have general discussions with your classmates about Tableau functionality, but there should be no specific reference to the assignment.

---

### Datasets

- HR-Employee-Attrition.csv
- GDP\_annual\_growth.csv
- Sample – Superstore (Saved Data Source in Tableau)

---

### Case 1: Attrition

Attrition is the loss of employees for reasons other than termination such as retirement, resignation, or health issues. Attrition is very costly to organizations because not only must they find, hire, and train a replacement, the lost employee may have critical organizational knowledge not formally documented. As such, attrition is a very important concern for companies and is typically closely monitored.

1. We want to begin by trying to understand how attrition is related to departments within the organization. **Recreate the visualization shown in Figure 1.**

**HINTS** Notice that the number of employees in the attrition and non-attrition categories is prefixed by “n=.” To accomplish this, once the label has been applied, click the Label card and then the ellipses next to “Text.”

2. Next, we want to examine how attrition relates to some factors we believe may encourage employees to leave. **Recreate the visualization shown in Figure 2.**

**HINTS:**

- a. This chart is tricky as the coloration is determined by a combination of a parameter and two calculated fields.
  - i. You need to begin by creating a parameter. Right clicking in the white space around your tables should give you the option to create a parameter. Call it “Attrition Threshold” and set it to integer. This will allow the visualization user to set a threshold value.
  - ii. Next, we need a calculated field equal to the number of employees for which attrition occurred. Create a calculated field called “Attrition Count” and use the following formula:

```
IF [Attrition] = "Yes"
THEN 1
ELSE 0
END
```

- iii. Finally, we need a calculated field that indicates whether the attrition count exceeds the attrition threshold. Create a calculated field called “Exceeds Threshold” and use the following formula:

```
SUM([Attrition Count])>[Attrition Threshold]
```

- iv. Drag Exceeds Threshold to the color card

## Case 2: GDP

The GDP\_annual\_growth.csv dataset is a superset of the data you analyzed in Assignment 1. In this case, we want to do a comparative analysis among countries for the year 2019.

**HINT:** To add the new data sources, create a new sheet and then select Data>New Data Source. Click “More” under To a File. Navigate to GDP\_annual\_growth.csv. Ensure it is properly formatted, then return to your sheet.

3. **Recreate the visualization shown in Figure 3.** This view shows the top 3 countries by GDP within each continent. Importantly, this figure excludes (filters out) values that are null for 2019 and groups of countries (situations where Aggregate is true).

**HINTS:**

- a. Start by building a table as you will need to create a rank table calculation to build this visualization. Rows will be Continent, Country Name, and 2019.
  - b. Your SUM(2019) rank table calculation will be: For each Continent, compute the rank, by country name (refer to slide 37 of the PowerPoint on table calculations to work out the details here).
  - c. The table calculation will have to be changed to discrete to be placed on the Rows shelf.
  - d. Once you have the table calculation correct, add your bars and labels. Then “copy” your table calculation by holding CTRL and dragging it to filters. There you should be able to select ranks of 1-3.
4. **Recreate the visualization shown in Figure 4. Please create the figure as a single visualization comprised of two views.** This visualization has the same aggregation and non-null filters as visualization #3 above.

**HINTS:**

- a. Start with the view on the left as it is easiest.
- b. For the right hand visualization, start by dragging a new copy of SUM(2019) to the columns shelf
  - a. You will convert this into a percent of total running sum. To do so you will need to create a table calculation with primary and secondary calculations
  - b. The primary calculation is a Running Total by Country Name
  - c. The secondary calculation is a percent of total by Country Name (note the “Add secondary calculation” checkbox on the Edit table calculation dialog box)
  - d. Once created, sort descending
- c. The countries associated with the top 50% of GDP are indicated with color. The color will require you to create a calculated field called “Top 50%”. The code for that field will be:

```
IF RUNNING_SUM(SUM([2019])) / TOTAL(SUM([2019])) <= .5 THEN 'YES'
ELSE 'NO'
END
```

**Case 3: Sales**

The Sales – Superstore dataset is a dataset packaged with Tableau. We are interested in how our sales have grown by state over time. While there are several ways to visually depict these trends, we have chosen an animation.

**HINT:** To add the new data sources, create a new sheet and then select Data>New Data Source. Click “Sample - Superstore” under Saved Data Sources.

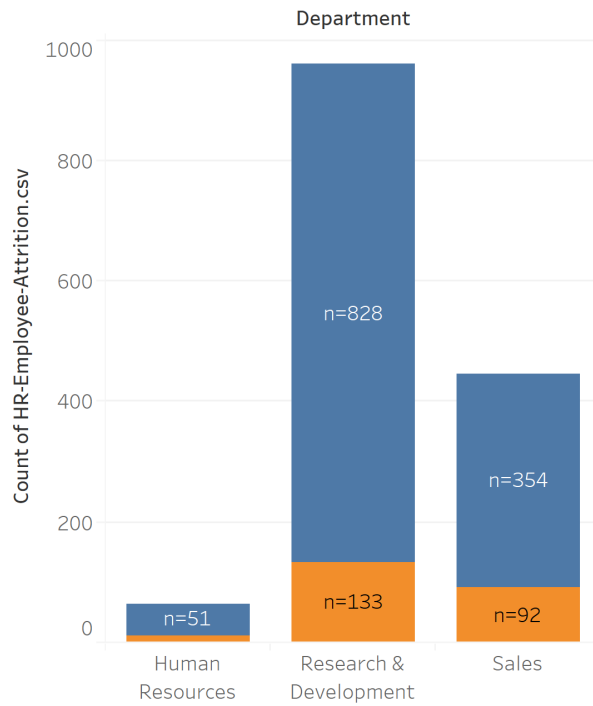
5. **Recreate the visualization shown in Figure 5.** This view shows a map of the united states where color indicates the running total of the sum of sales. As the animation plays, the states where sales have grown significantly should become progressively darker.

**HINTS:**

- a. Begin with a simple map of states. Expand location and double click on State.
- b. Color the map by dragging sales to the color card.
- c. Now drag Order Date to the pages shelf. Click the down arrow on the pill and change it from YEAR to Quarter (Q2 2015). If you play the animation, you will notice that the states don't get progressively darker... the color jumps around. That is because we are plotting each quarter's sales independent of what has happened previously in the state.
- d. To correct that issue, we need to change SUM(Sales) to a table calculation. Click the down arrow on the Sales pill you previously moved to the color card. Add a running total table calculation. Your SUM(Sales) running total table calculation will be: For each Country/Region and State, calculate a running total by quarter (refer to slide 37 of the PowerPoint on table calculations to work out the details here).
- e. Make sure to test your visualization by running the animation.

## Figures

### Q1 - Attrition by Department



**Figure 1**

### Q2 - Attrition Against Targets

Department	Business Travel	Over Time	Attrition	
			No	Yes
Human Resources	Non-Travel	No	5	
		Yes	1	
	Travel_Frequently	No	5	2
		Yes	2	2
	Travel_Rarely	No	29	5
		Yes	9	3
Research & Development	Non-Travel	No	71	3
		Yes	18	5
	Travel_Frequently	No	104	17
		Yes	41	20
	Travel_Rarely	No	456	39
		Yes	138	49
Sales	Non-Travel	No	34	2
		Yes	9	2
	Travel_Frequently	No	49	14
		Yes	7	14
	Travel_Rarely	No	191	28
		Yes	64	32

Attrition Threshold

16

AGG(Exceeds Threshold)

False

True

**Figure 2**

Q3 - Top 3 GDP Countries by Continent

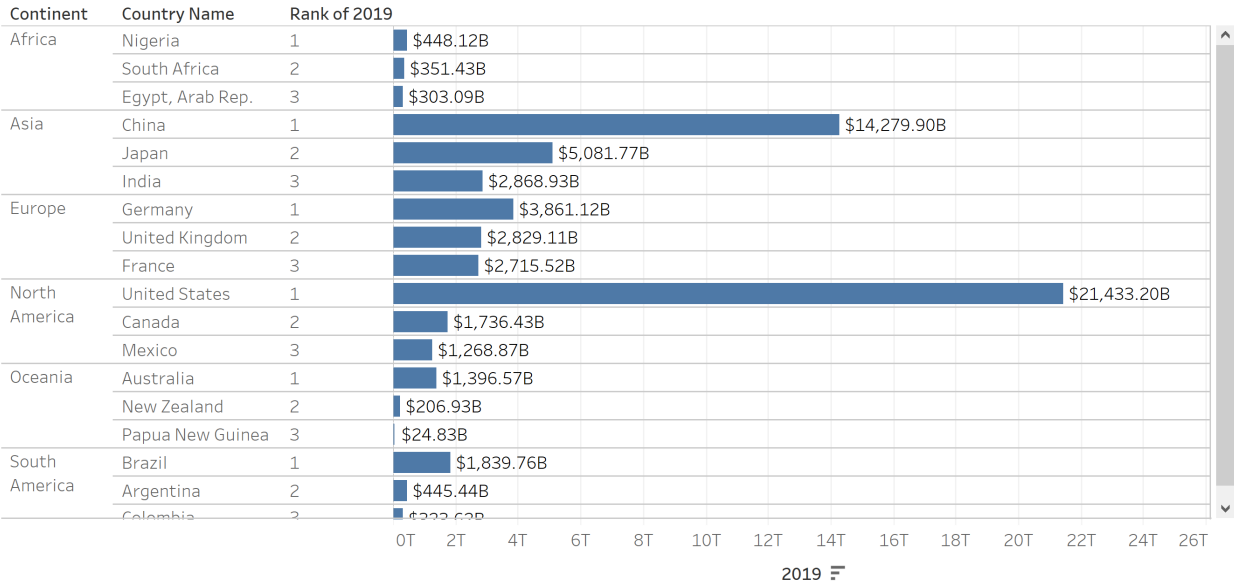


Figure 3

Q4 - Countries Accounting for Top 50% of GDP

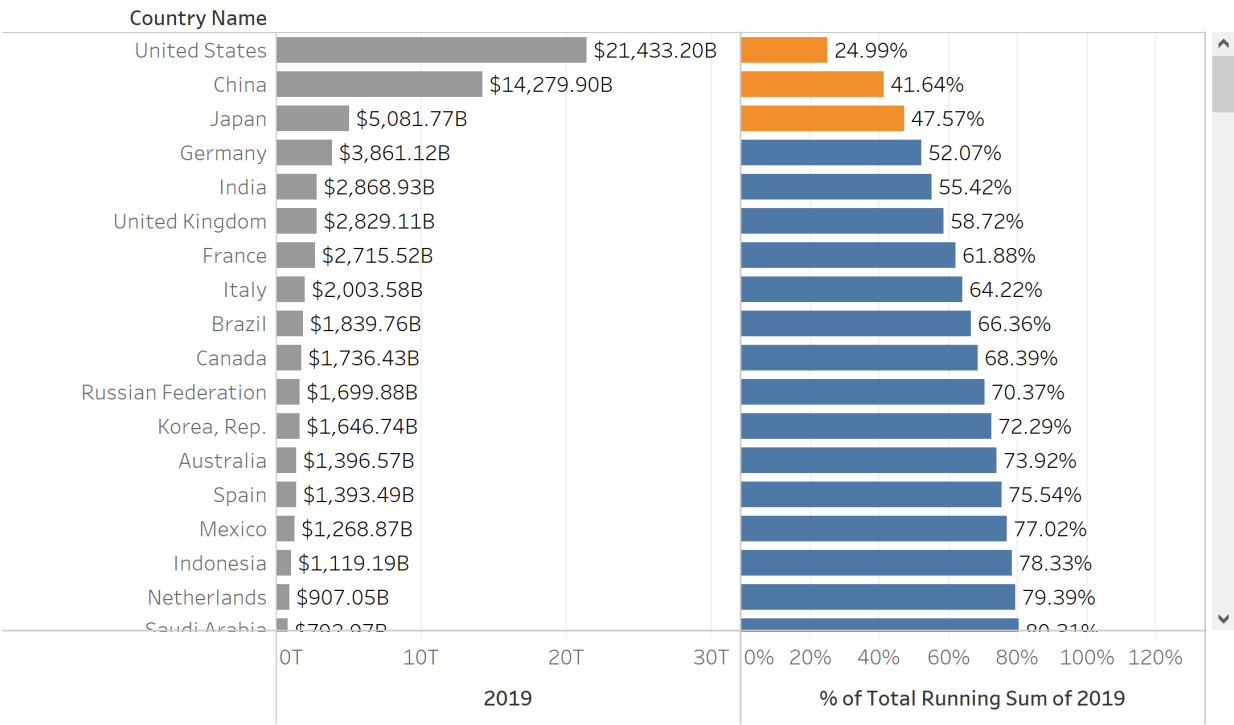
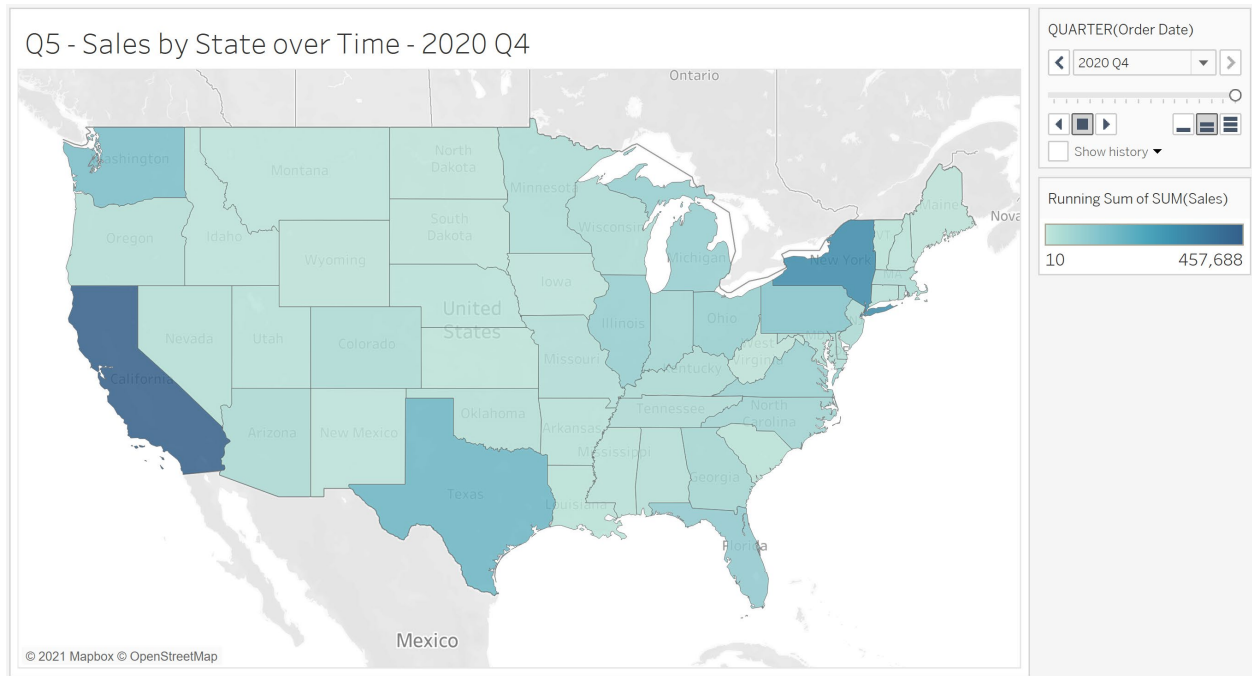


Figure 4



**Figure 5**