Effective data stories

In data analytics, **data storytelling** is communicating the meaning of a dataset with visuals and a narrative that is customized for a particular audience. In data journalism, journalists engage their audience of readers by combining visualizations, narrative, and context into data-driven articles. It turns out that data analysts and data journalists have a lot in common! As a junior data analyst, you might learn a few things about effective storytelling from data journalism. Read further to explore the role and work of a data journalist in telling a good story.

Note: This reading refers to an article published in *The New Yorker*. Non-subscribers may access several free articles each month. If you already reached your monthly limit on free articles, bookmark the article and come back to this reading later.

Take a tour of a data-driven article

[Ben Wellington](), a contributing writer for *The New Yorker* and a professor at the Pratt Institute, used New York City's [open data portal]() to track down noise complaints from logged service requests. He analyzed the data to gain a more quantitative understanding of where the noise was coming from and which neighborhoods were the noisiest. Then, he presented his findings in the [Mapping New York's Noisiest Neighborhoods]() article.

First, click the link above to skim the article and familiarize yourself with the data visualizations. Then, join the bus tour of the data! You will be directed to three visualizations (tour stops) to observe how each visualization helped strengthen the overall storytelling in the article.

Tour stop 1: setting context

Earlier in the training, you learned how context is important to understand data. **Context** is the condition in which something exists or happens. Based on the categorization of noise complaints, the data journalist set the context in the article by defining what people considered to be noise.

In the article, review the [combo table and bar chart]() that categorizes the noise complaints. Evaluate the visualization:

How does the visualization help set the context?

- The combo table and bar chart is effective in summarizing the noise categories as percentages of the logged complaints. This helps set the context by answering the question, "what is noise?" Notice that the data journalist created a combo table and bar chart instead

of a pie chart. With 11 noise categories, a list with a bar chart showing relative proportions is an elegant representation. A pie chart with 11 slices would have been harder to read.

How does the visualization help clarify the data?
- If you add the percentages in the categories in the combo table and bar chart, the total is ninety-eight percent. There is a difference of two percent that can't be accounted for in the visualization. So, rather than clarifying the data, the visualization actually causes a little confusion. One lesson is to always make sure that your percentages add up correctly. Sometimes rounding decimal places up or down causes percentages to be off so they don't add up to 100%.

Do you notice a data visualization best practice?
You learned that a companion table in Tableau shows data in a different way in case some in your audience prefer tables. It appears that the data journalist had the same idea by using a combo table and bar chart.

- **Note:** As a refresher, a companion table in Tableau is displayed right next to a visualization. A companion table displays the same data as the visualization, but in a table format. You may replay the Getting Creative video which includes an example of a companion table.

Tour stop 2: analyzing variables

After setting the context by identifying the noise categories, the data journalist describes his analysis of the noise data. One interesting analysis is the distribution of noise complaints versus the time of day.
In the article, review the stacked area chart for the distribution of noise complaints by hour of the day. Evaluate the visualization:

How does the visualization perform against the five-second rule?
- Recall that the five-second rule states that you should understand what is being conveyed within the first five seconds of seeing a chart. We are guessing that this visualization performs quite well! The area charts for loud music and barking dogs help the audience understand that more of these types of noise complaints were made during late night and early morning hours (between 10:00 PM and 2:00 AM). Notice also that the color coding in the legend aligns with the colors in the chart. A chart legend normally has the largest category at the top, but the data journalist chose to order the legend so the largest category, "Loud music or party" appears at the bottom instead. How much time do you think this alignment saved readers?

How does the visualization help clarify the data?
- Unlike the visualization from the previous tour stop, this visualization does a better job of clearly showing that all percentages add up to 100%.

Do you notice a data visualization best practice?
- As a best practice, both the x-axis and y-axis should be labeled. But, the data journalist chose to include % or A.M. and P.M. with each tick on an axis. As a result, labeling the x-axis "Time of Day'' and the y-axis "Percentage of Noise Complaints" isn't required. This demonstrates that a little creativity with labeling can help you achieve a cleaner chart.

Tour stop 3: drawing conclusions

After describing how the data was analyzed, the data journalist shares which neighborhoods are the noisiest using a variety of visualizations: [combo table and bar chart](#), [density map](#), and [neighborhood map](#).

In the article, review the [neighborhood map](#) for how close a noisy neighborhood is to a quiet neighborhood. Evaluate the visualization:

How does the visualization help make a point?
- The data journalist observed that one of the noisiest neighborhoods was right next to one of the quietest neighborhoods. The neighborhood map is effective in emphasizing this observation as a dark blue area versus a white area.

How does the visualization help clarify the data?
- The visualization classifies the data by neighborhood and allows the audience to follow along when the journalist focuses specifically on the Williamsburg, East Williamsburg, and North Side/South Side neighborhoods.

Do you notice a data visualization best practice?
- Each neighborhood is directly labeled so a legend isn't necessary.

End of the tour: being inspired

We hope you enjoyed your tour of a data journalist's work! May this inspire your data storytelling to be as engaging as possible. For additional information about effective data storytelling, read these articles:
- [What is Data Storytelling?](#)
- [The Art of Storytelling in Analytics and Data Science | How to Create Data Stories?](#)
- [Use Data and Analytics to Tell a Story](#)
- [Tell a Meaningful Story With Data](#)

Fill in the blank: A data analyst wants to pinpoint the most relevant data derived from their analysis and eliminate the less important details. They use _____ to scan the data and quickly identify the most important insights.

- ○ engaging
- ● spotlighting
- ○ messaging
- ○ annotating

⊘ **Correct**
A data analyst uses spotlighting to pinpoint the most relevant data derived from their analysis and eliminate the less important details. Spotlighting involves scanning the data to quickly identify the most important insights.

Activity Overview

You've learned that you can create charts in spreadsheets and in Tableau and that information can be customized using filters. Before working in Tableau, you will work with a spreadsheet chart editor. This will allow you to create visualizations to draw insights from immediately—and will help you start thinking about how to make design choices in more complex tools.

By the time you complete this activity, you will be able to create a chart in a spreadsheet and customize it by using filters and applying different styles. This will enable you to use spreadsheets in another helpful way, which is important for using all the tools at your disposal in your career as a data analyst.

Step-By-Step Instructions

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.

Step 1: Access the template

To use the template for this course item, click the following link and select Use Template.

Link to template: Cosmetics, Inc

OR

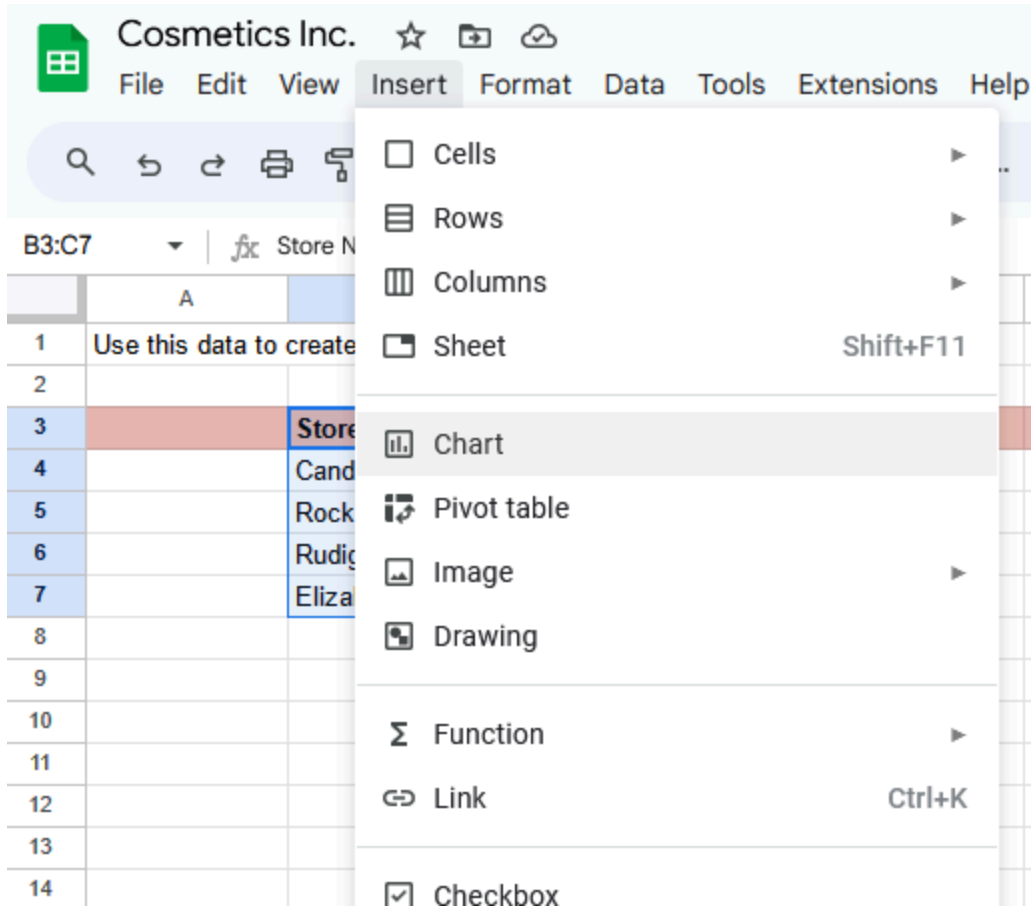If you don't have a Google account, download the template directly from the following attachment.

Cosmetics Inc.

XLSX File

Step 2: Create a chart in a spreadsheet

Important note: The following steps are for Google Sheets, but they can help guide you to similar features in Excel. Refer to the Additional resources section if you need more specific instructions for creating charts in Excel.

1. In the spreadsheet that you copied or downloaded, make sure you're working in the first tab named Create your chart here. This tab contains the total purchases calculated for each store from the second tab. All of the original data is preserved in the second tab so you can explore chart creation in the first tab without changing or deleting any of the original data.

2. Select to highlight cells B3 through C7 as the data for your chart.

3. From the main menu, select Insert, and select Chart.

Cosmetics Inc. ☆ ⊡ ☁

File   Edit   View   Insert   Format   Data   Tools   Extensions   Help

🔍 ↺ ↻ 🖨 ▦

B3:C7   ▾ | fx Store N

	A	
1	Use this data to create
2	
3		Store
4		Cand
5		Rock
6		Rudig
7		Eliza
8		
9		
10		
11		
12		
13		
14		

Insert menu:

☐ Cells   ▶
▤ Rows   ▶
▥ Columns   ▶
▭ Sheet   Shift+F11

▥ Chart
⇄ Pivot table
▣ Image   ▶
▣ Drawing

Σ Function   ▶
🔗 Link   Ctrl+K

☑ Checkbox

4. A chart is created automatically, and the Chart editor pane opens on the right so you can change the chart type. If a pie chart is displayed, select the Chart type dropdown list and then select a Column chart so your chart matches the one below. Based on the data selected, you should also get a title like the one shown.
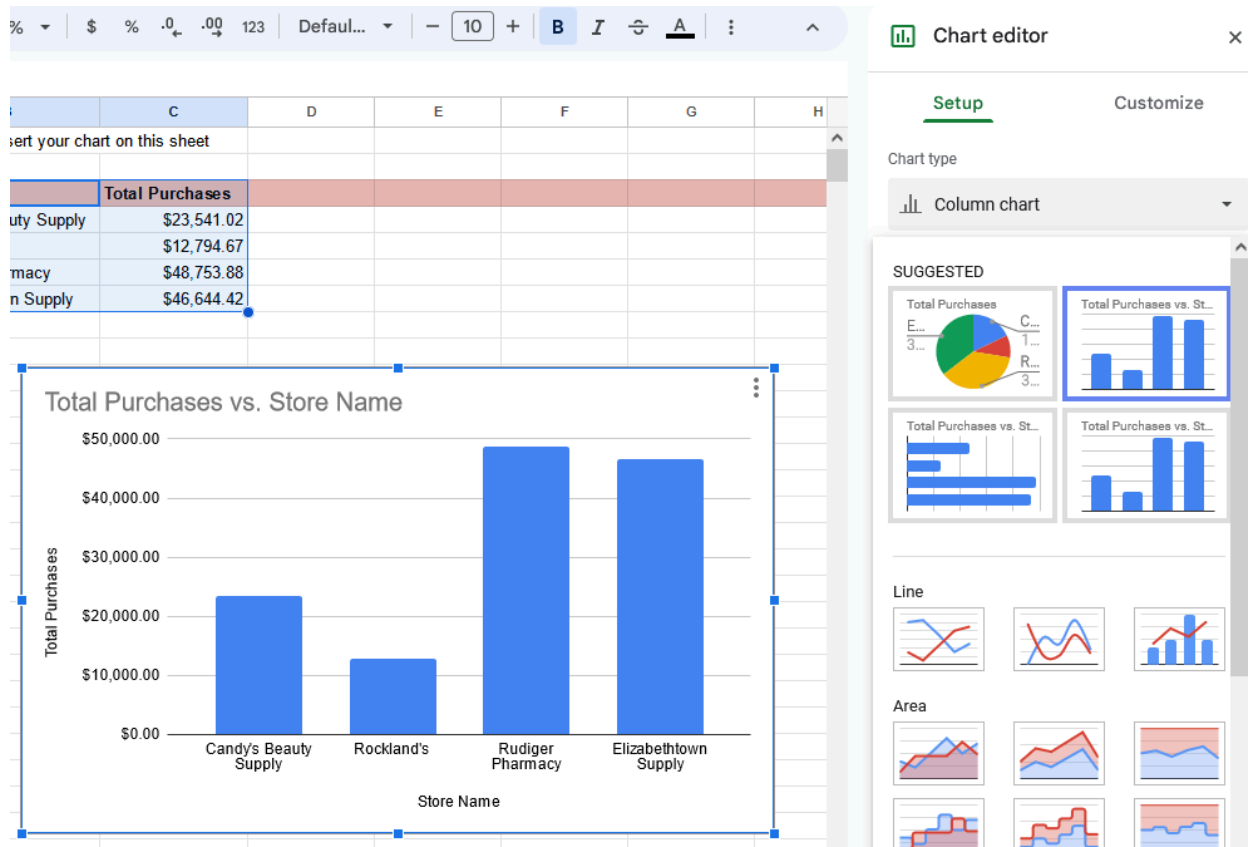
Bars represent Candy's Beauty Supply, Rockland's, Rudiger Pharmacy, and Elizabethtown Supply.

5. In the Setup section of the Chart editor, below X-axis, select the three dots menu and then select Add labels. Make sure the range B3:B7 is indicated to the left of the three dots, and after you select the Add labels option, verify that the data columns display a vertical label on each column.

6. To filter the data shown in the column chart, select the pink header row (row 3) and then select the filter icon in the toolbar. This adds a filter icon on each column of your data.

Alternatively, you can use the Data menu and select Create a filter after selecting cells B3:C3.
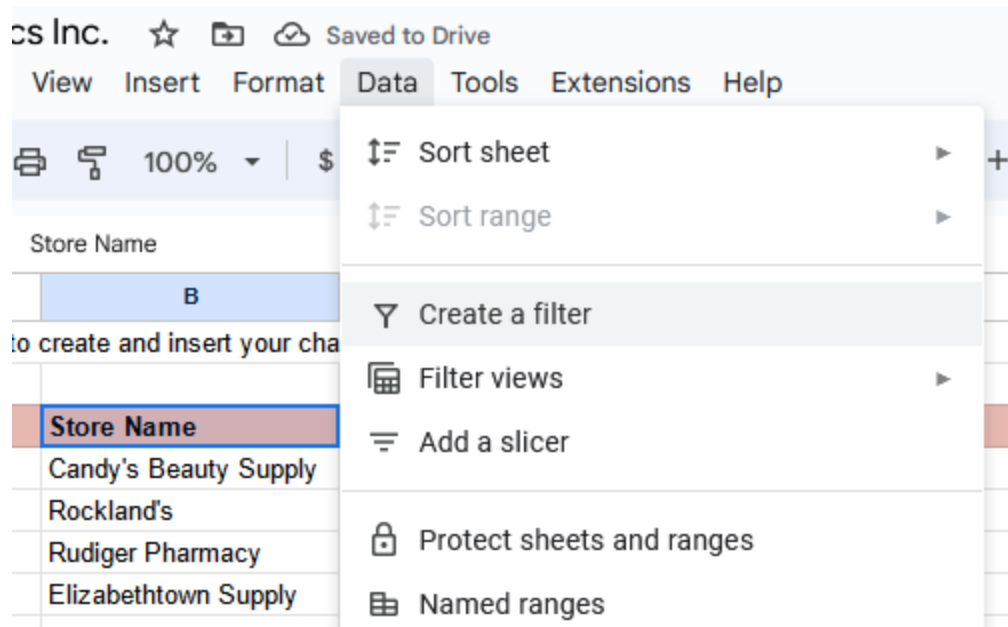
7. Select the filter icon for the Store Name column and uncheck Candy's Beauty Supply.
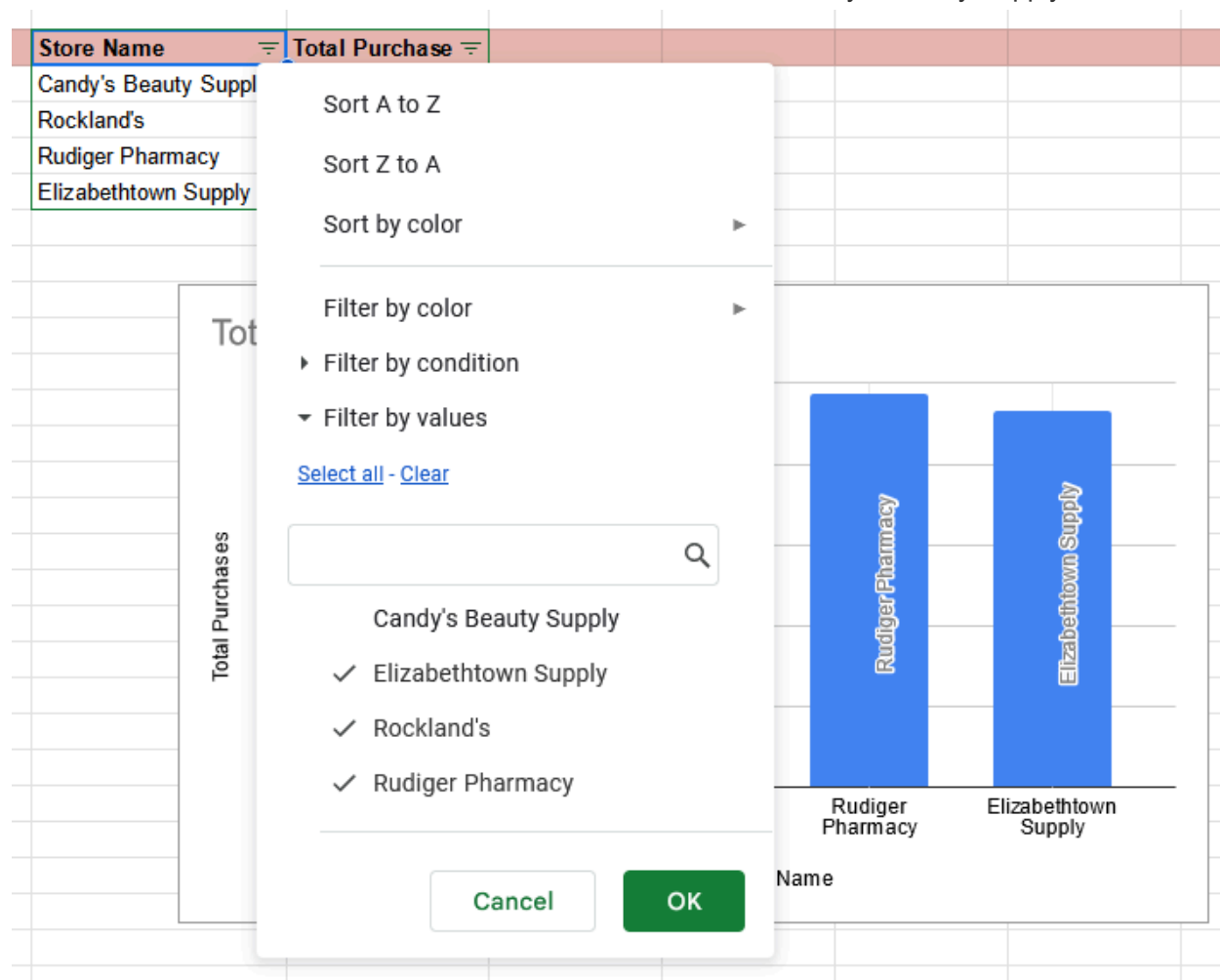
Below these options, a list of the store names allows you to check or uncheck each, allowing you to show each on the chart or not.

8. Select OK and your chart will automatically update to exclude the data for Candy's Beauty Supply, as shown below.

Store Name	Total Purchase
Rockland's	$12,794.67
Rudiger Pharmacy	$48,753.88
Elizabethtown Supply	$46,644.42
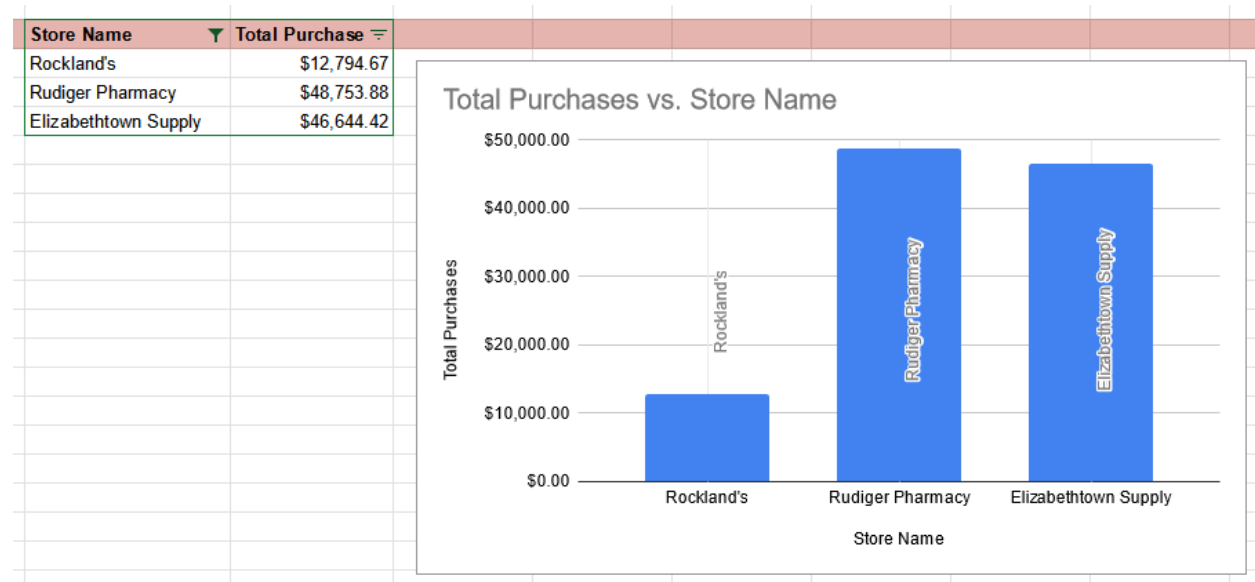
Note: To include the data for Candy's Beauty Supply again, simply select the filter icon for Store Name and then select the Candy's Beauty Supply checkbox to add it back to the chart.

Step 3: Apply styles

You can also apply different styles to the chart. For example, the default blue in the chart isn't the most appropriate for accessibility when the background is white.

1. Double-click the chart to open the Chart editor pane.

2. If necessary, in the Chart editor pane, select Series to expand the options.
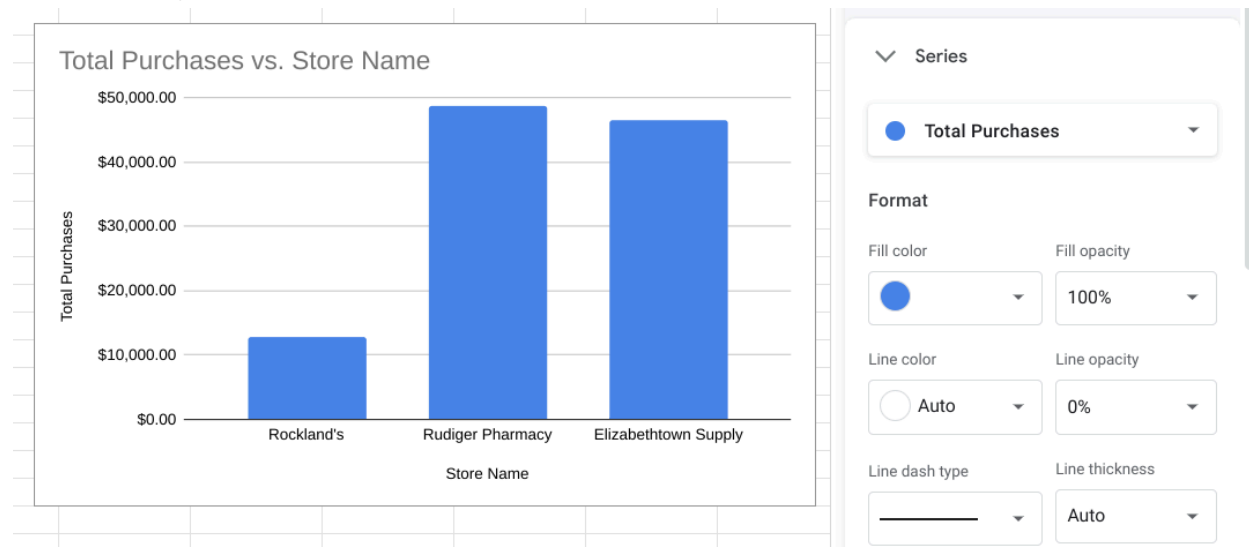
3. Below Format, from the Fill color dropdown list, select Custom.

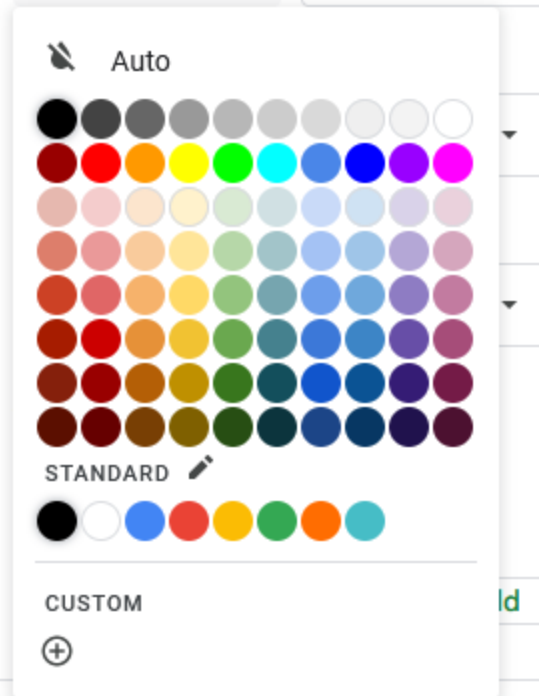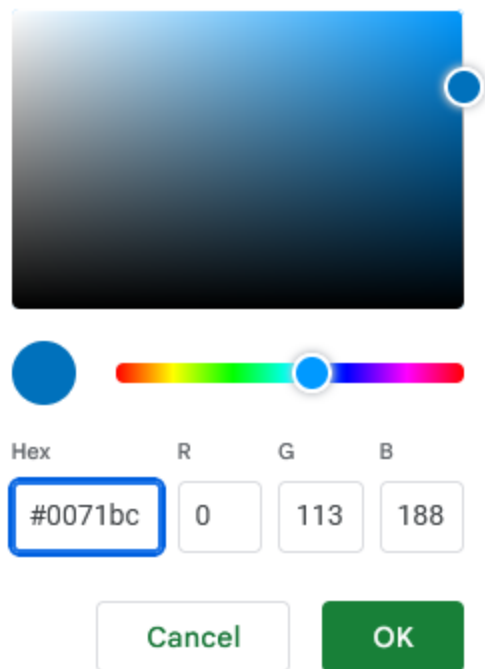You can also add a custom color at the bottom of the menu.

4. Change the Hex value for the color to one of the blue hues recommended for use on white backgrounds for accessibility. Carnegie Museums of Pittsburgh's Web accessibility guidelines are an example of web accessibility guidelines.

- #0071bc
- #046b99
- #205493

Fields below allow you to enter a Hex value, in this case 0071bc. There are also three fields that allow you to enter values for red, blue, and green to specify a color.

5.Select OK, and the chart will automatically update with the newly assigned color.

Step 4: Review resources

Here are a few more resources you can reference as you learn more about charts in spreadsheets:

- Graphs in Google Sheets: This resource contains a detailed example of chart creation in spreadsheets and provides downloadable sample data you can use to practice. As you have learned throughout this course, practicing these skills helps you learn more about the tools you are using. This example data is a great way to start!
- Add and edit a chart or graph in Google Sheets: This article includes steps for creating, editing, and changing charts in Google Sheets with how-to videos. It also has a more in-depth guide to editing and customizing your chart after you have created it.
- Create a Microsoft Excel chart from start to finish: This how-to guide from Microsoft's support site includes instructions and a video tutorial for adding charts to Excel spreadsheets. This is a useful resource if you are working specifically with Excel spreadsheets. It also links to other useful articles about creating charts in Excel.
- Microsoft Excel: Creating and modifying charts: This is an explanation of Excel charts with downloadable handouts. This resource is especially useful because it has downloadable content that you can save to reference later when you start creating charts in your own spreadsheets.

Pro Tip: Save the blank activity template

Be sure to save a blank copy of the template you used to complete this activity. You can use it for further practice or to help you work through your thought processes for similar tasks in a future data analyst role.

Create your first Tableau dashboard

You're going to take what you've learned about data visualizations and create a dashboard. Tableau contains tons of other functionality to build dashboards that update in real time and include interactive visualizations.

Dashboards are important in data analysis because they enable people to visualize data in dynamic and interactive ways, which can help enhance what you can do with your data presentations. Data visualizations are most useful when they are presented in a dashboard style format to stakeholders because dashboards put all the key information in the same place, making it easier to understand what's really important. Many dashboards also constantly update to reflect new data and can be interactive. No matter what style of dashboard you choose, they can help you present your data to stakeholders in an impactful way.

Open Tableau

You'll need to navigate and sign-in to the [Tableau Public](#) online app. You may also refer back to the reading on how to create an account, a profile, loading, and linking datasets using the Tableau public app - [Using Tableau Public](#).

Accessing the Dataset

Click on the link to create your own copy of the [CO2 Dataset](#).
If you do not have a Google account, you may directly download the CO2 dataset by clicking below:
[CO2 Dataset](#)
[XLSX File](#)

Load the Dataset

1. Now that you have logged into Tableau Public, access the data source dashboard by clicking on the **Create tab**, then the **Web Authoring** option at the top of the landing page. Next, click on the center button, **Upload from computer**, and select the CO2 Dataset that has been downloaded to your device.
2. Once the dataset has been loaded, direct your attention to the bottom of the data source tabs in the bottom left corner of the window. This opens the data sources folder Tableau public has created on your machine by default. Going forward, you should save any datasets you're working with to the data sources folder. Keeping your data files in one place is a best practice that will keep you organized.
3. Make a note that all of the tabs within the dataset are arranged into a vertical toolbar on the left side of the page.
4. Drag the Sheet titled **CO2 Data Cleaned** from the left bar to the middle of the page and click on the button **Update Automatically** option after dragging the sheet icon.
5. Navigate back to the tab titled **Sheet 1** at the bottom of the page and click on it. At this point, a new window with a vertical dashboard titled **Tables** will appear.
6. Drag the *green* **# Year** icon to the **Columns** box located at the top of the page.
7. Drag the *green* **# CO2 (kt)** icon to the **Rows** box located at the top of the page.
8. Drag the *blue* **Abc Region** icon to the **Colors** icon containing the 4-colored dots.

9. You have now created your first basic timeline dashboard, which indicates the level of CO2 (kilotons) emitted by the various parts of the globe from 1960 to 2012.
10. You may also do a preliminary analysis of the number of categories, rows, data ranges, and any null values that appear in the data source page. This will give you a sense of more dashboard types and styles that can be generated from the dataset.

Create Additional Dashboards

Directions on how to create additional dashboard types with Tableau are [here](#).
Great, you've created your first basic dashboard!

Dashboards put all the pertinent information in the same place, making it easier to understand the important takeaways. Many dashboards are also constantly updating to reflect new data, and some are even interactive. No matter what style of dashboard you choose, they can help you deliver the work you've done when creating visualizations.

After you download the Dashboards Starter Template, find the file in your storage and open it in Tableau Public Desktop.

Upon opening the Tableau project template, your screen should showcase the C02 Production by Region sheet.
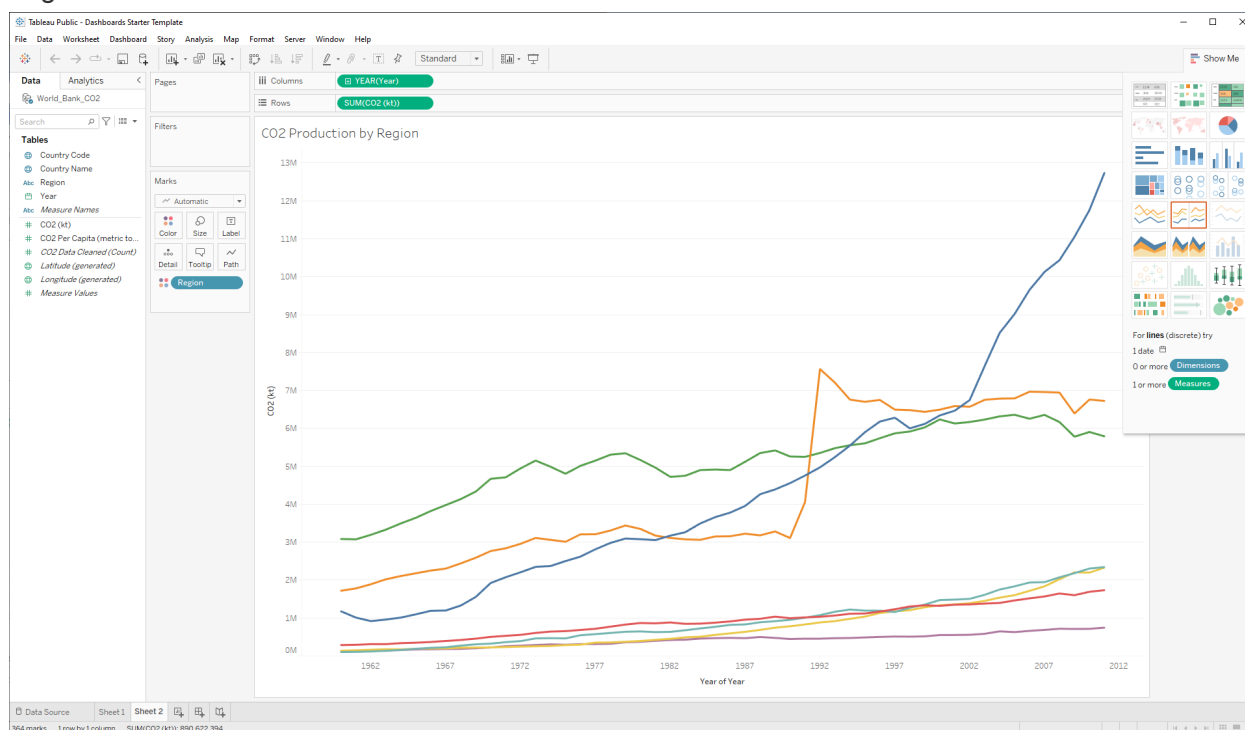
A line chart of CO2 emissions in KT by year shows six trendlines. Some lines rise more dramatically than others. All trend upwards.

The Dashboards Starter Template workbook allows you to explore and manipulate the visualizations found in two sheets: Sheet 1 and Sheet 2. However, the Tableau workbook does not contain the actual dataset.

# Step 3: Load the data

Next, you will load the dataset.

1. Select the Data Source tab in the sidebar. This will open the Data Sources folder Tableau Public has created on your computer by default.
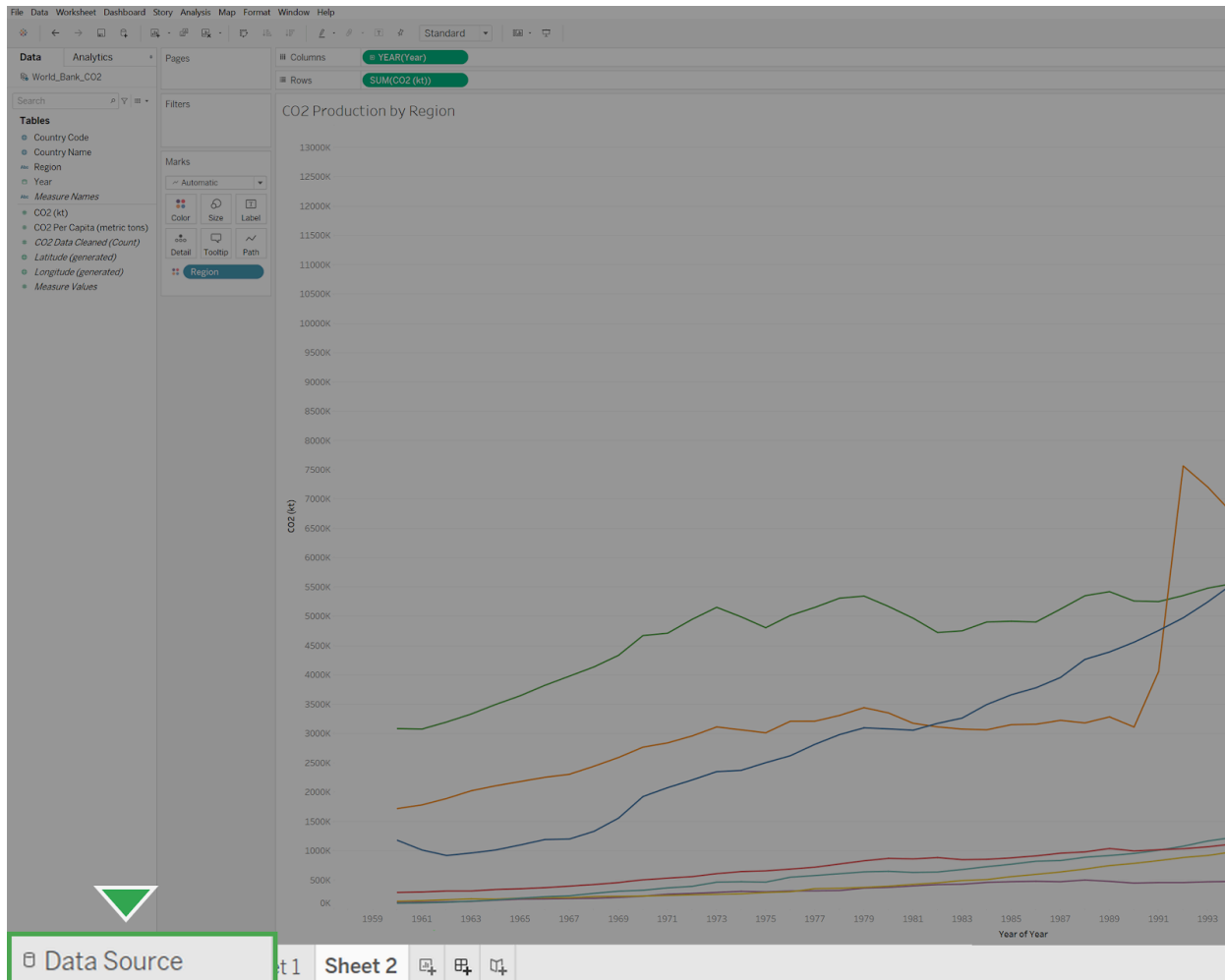
2. Navigate to the location on your computer where you downloaded the CO2 dataset (filename = "CO2 Dataset.xlsx") and open it.

3. Locate the My Tableau Repository folder on your computer. This is usually placed in the Documents folder of your local files. If you cannot find the folder, use the search bar in your computer's file explorer.

4. Open the folder My Tableau Repository, then open the folder Datasources.

5. Drag your datasets for Tableau from where you downloaded them into the Datasources folder. This will help you keep track of your datasets for various projects and stay organized.

Note: As a best practice, you should always move your datasets for Tableau into the Datasources folder.

Step 4: Create a dashboard

The example project contains the CO2 dataset which has two separate visualizations. Select Sheet 1. This visualization shows the average CO2 per capita of each country. Now, select Sheet 2. This visualization is a line chart of the CO2 production of each global region over time.

You will use these visualizations to create a dashboard. Select the Add Dashboard button, which is the middle button on the bottom row with a symbol that appears like a spreadsheet with a plus sign.
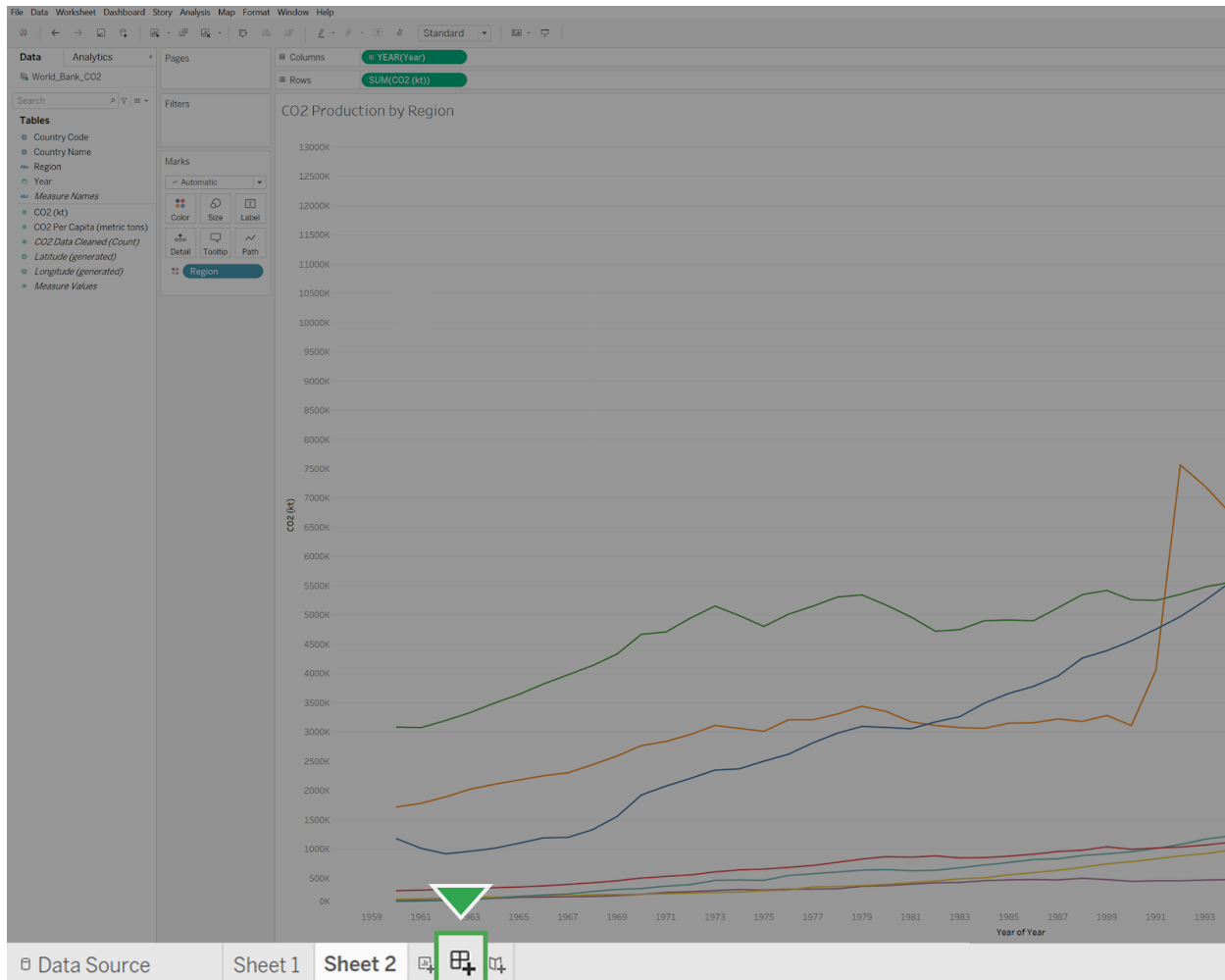
This will open a new dashboard. Your screen should appear like this:
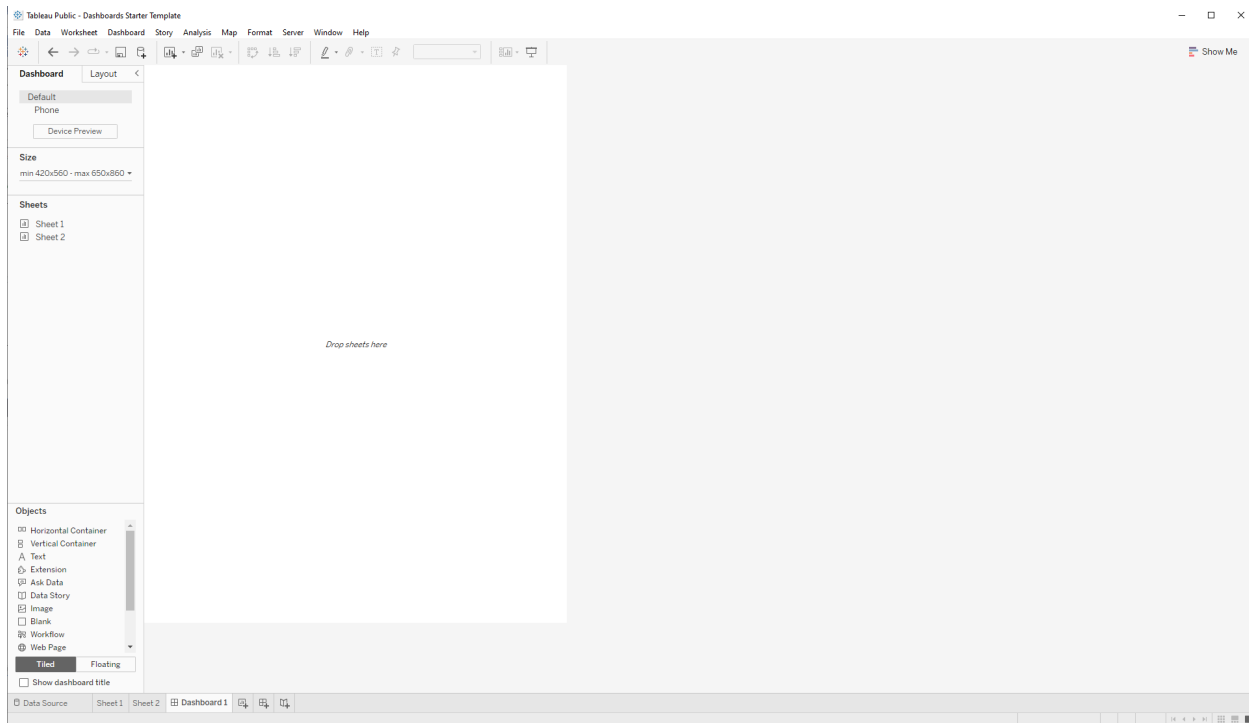
Menus across the top are File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, and Help. The Dashboard tab is selected on the left, showing size and a list of the sheets. Below this, an Objects menu lists Horizontal Container, Vertical Container, Test, Extension, Ask Data, Data Story, Image, Blank, Workflow, and Web Page. Below this are options for tiled or floating, and a checkbox to select show dashboard tile.

Step 5: Add visualizations

To add visualizations, drag the appropriate sheets onto the dashboard in the layout that you prefer. In this case, you'll add the map visualization from Sheet 1 on top of the line graph from Sheet 2.

1. Start by finding Sheet 1 in the Sheets section in the sidebar. Select and drag Sheet 1 onto the area that says Drop sheets here. Sheet 1 should appear under the Sheets tab:
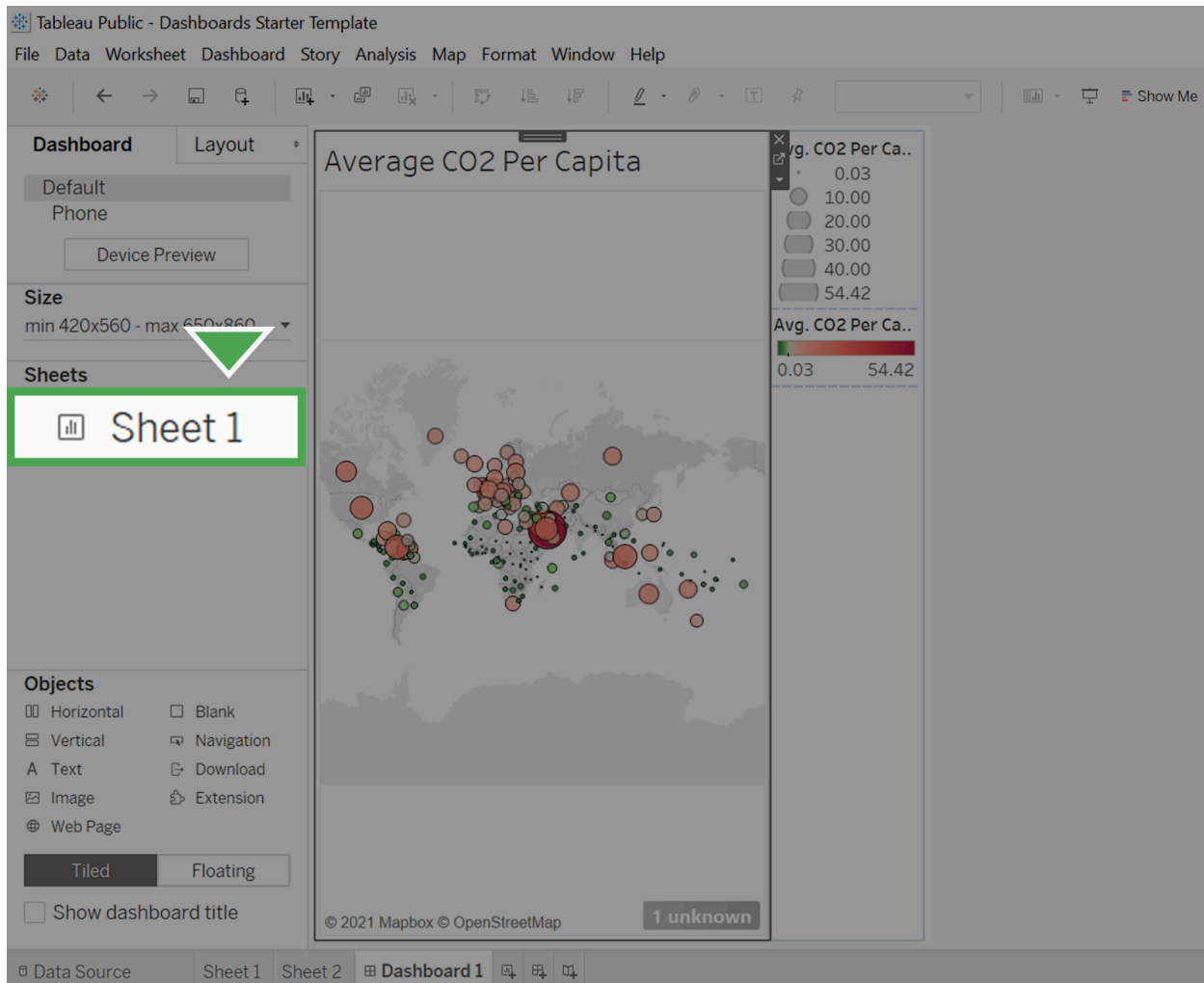
2. Select and drag Sheet 2 onto the visualization. You'll notice that the visualization adjusts to show the layout depending on where you drag the sheet. Place Sheet 2 so that it takes up the bottom half.
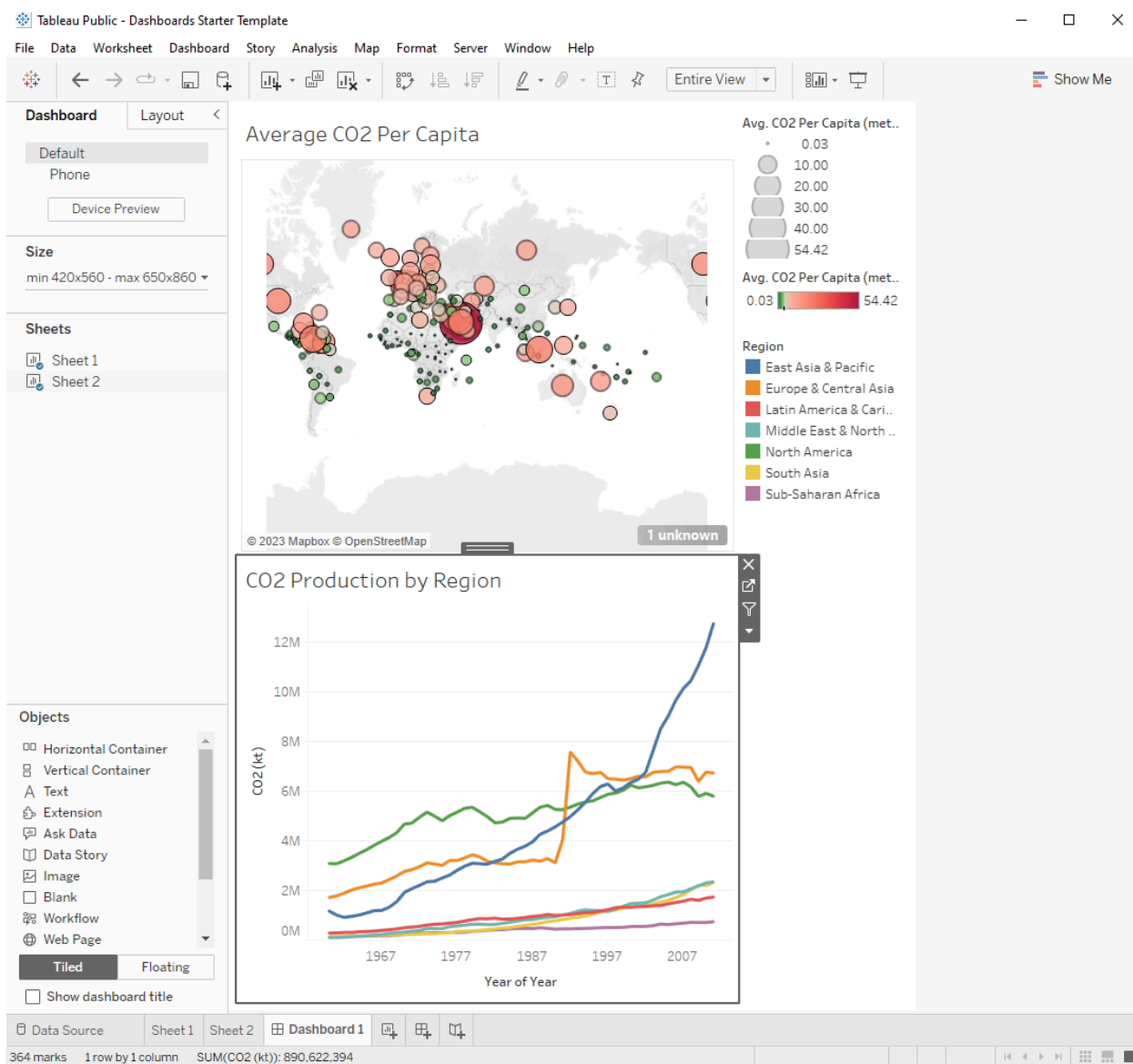
Step 6: Clean the dashboard

The dashboard currently contains three legends, but only two of them are needed. The legend of grayscale values represents the CO2 Per Capita by size.

CO2 per capita is represented by size and color. As such, Tableau creates two legends. To simplify the visualization, your best choice is to delete the topmost legend that corresponds to size.

The relationship between small and large emissions can be interpreted by the relative sizes of the circles. However, the color representing the number of emissions per capita is not interpretable without the legend.
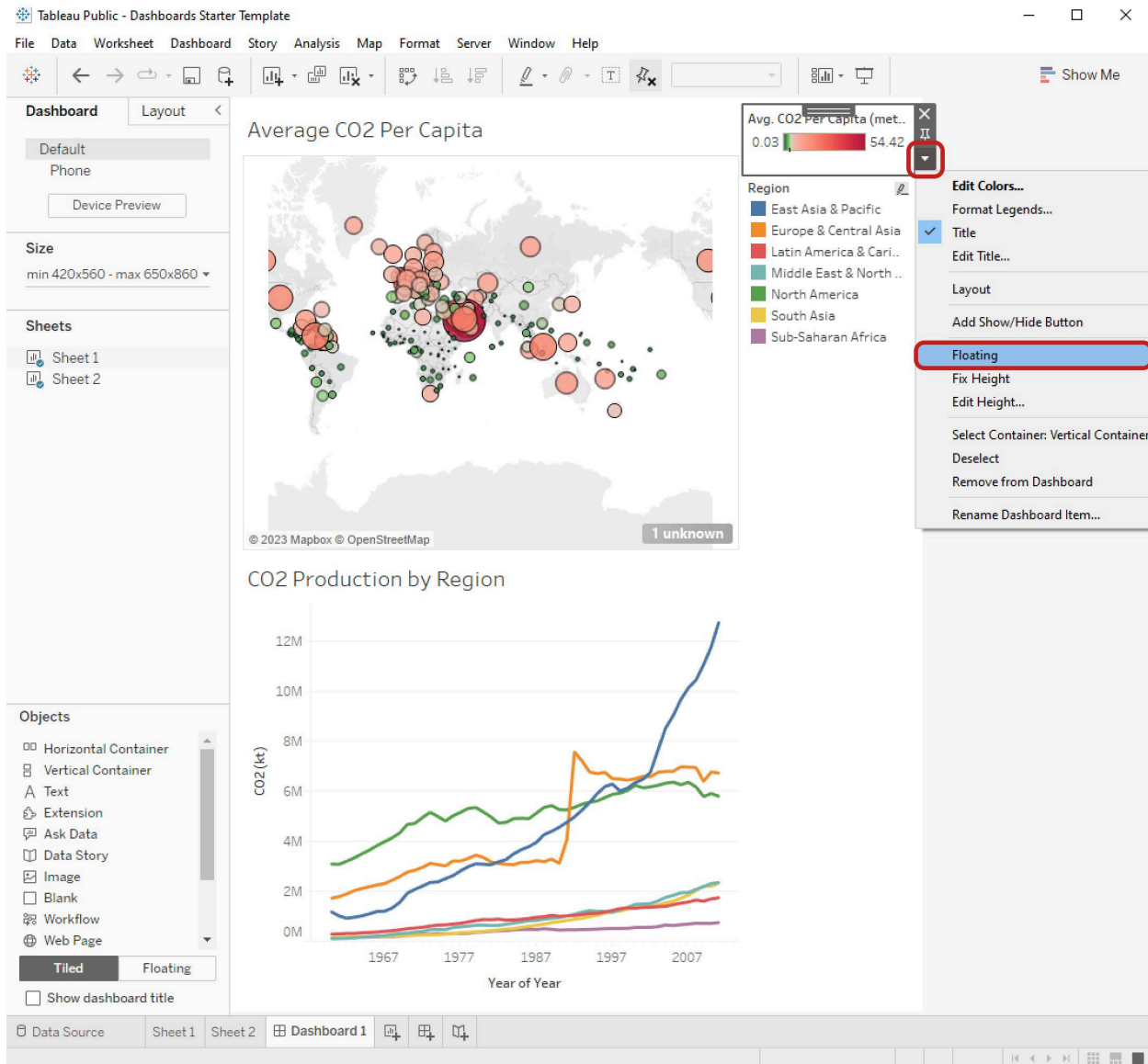
1. Delete the topmost legend. To do this, select it and then select the close button (X) to remove it from the dashboard.
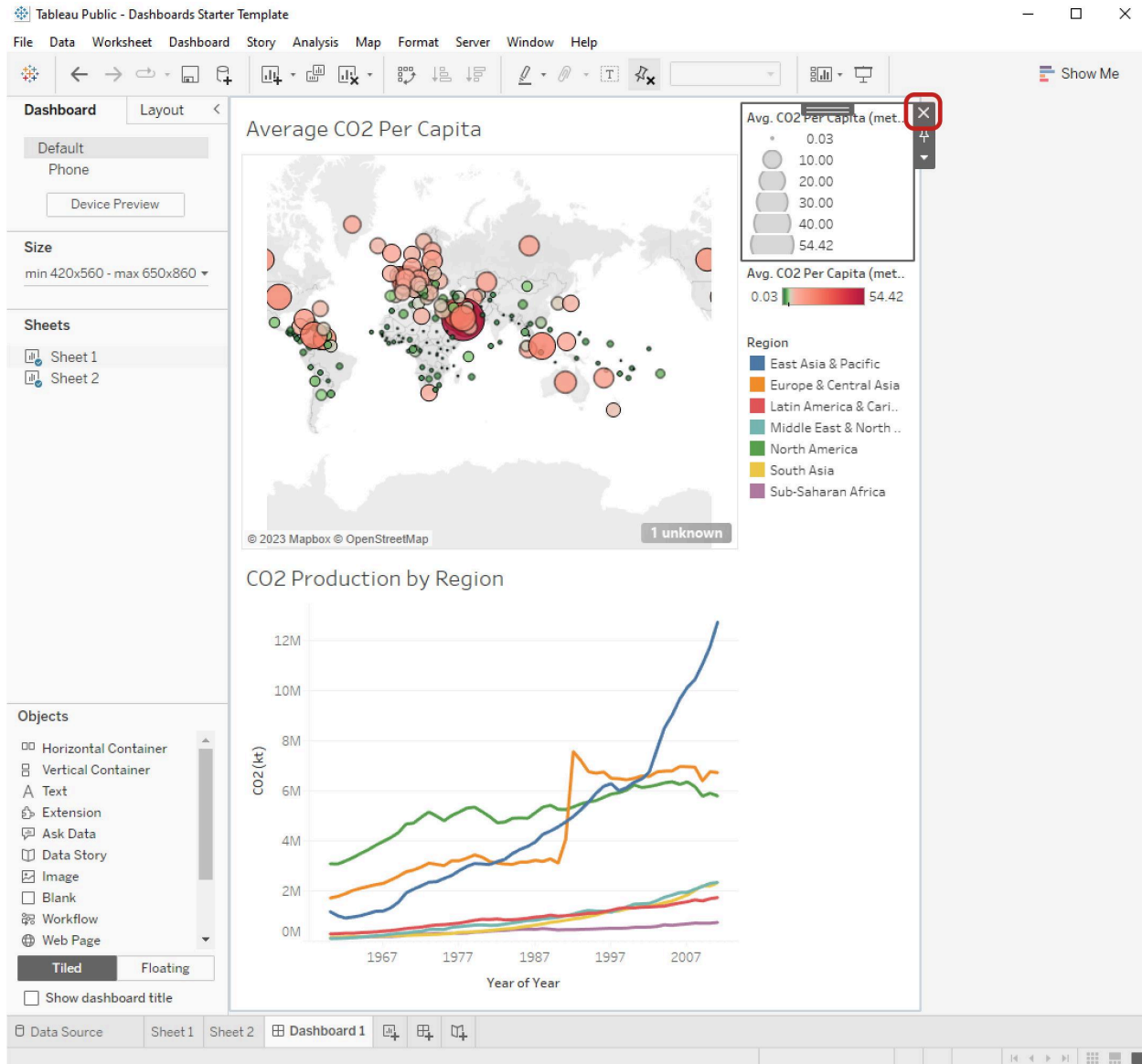
Now that it's been removed, you'll set the remaining legends to float. This will allow you to drag the legends where you want to place them.

2. Select a legend.

3. Select the dropdown list for More Options. From there, select Floating.

This arrow opens a menu including Edit Colors, Format Legends, Title, Edit Title, Layout, Add Show/Hide Button, Floating, Fix Height, Edit Height, Select Container: Vertical Container, Deselect, Remove from Dashboard, Rename Dashboard Item. Floating is highlighted.

4. Drag the legend onto the top-right corner of the map visualization.

5. Repeat steps 2-4 and float the remaining legend onto the top-right corner of the bottom graph. When you're finished, your dashboard should appear like this:
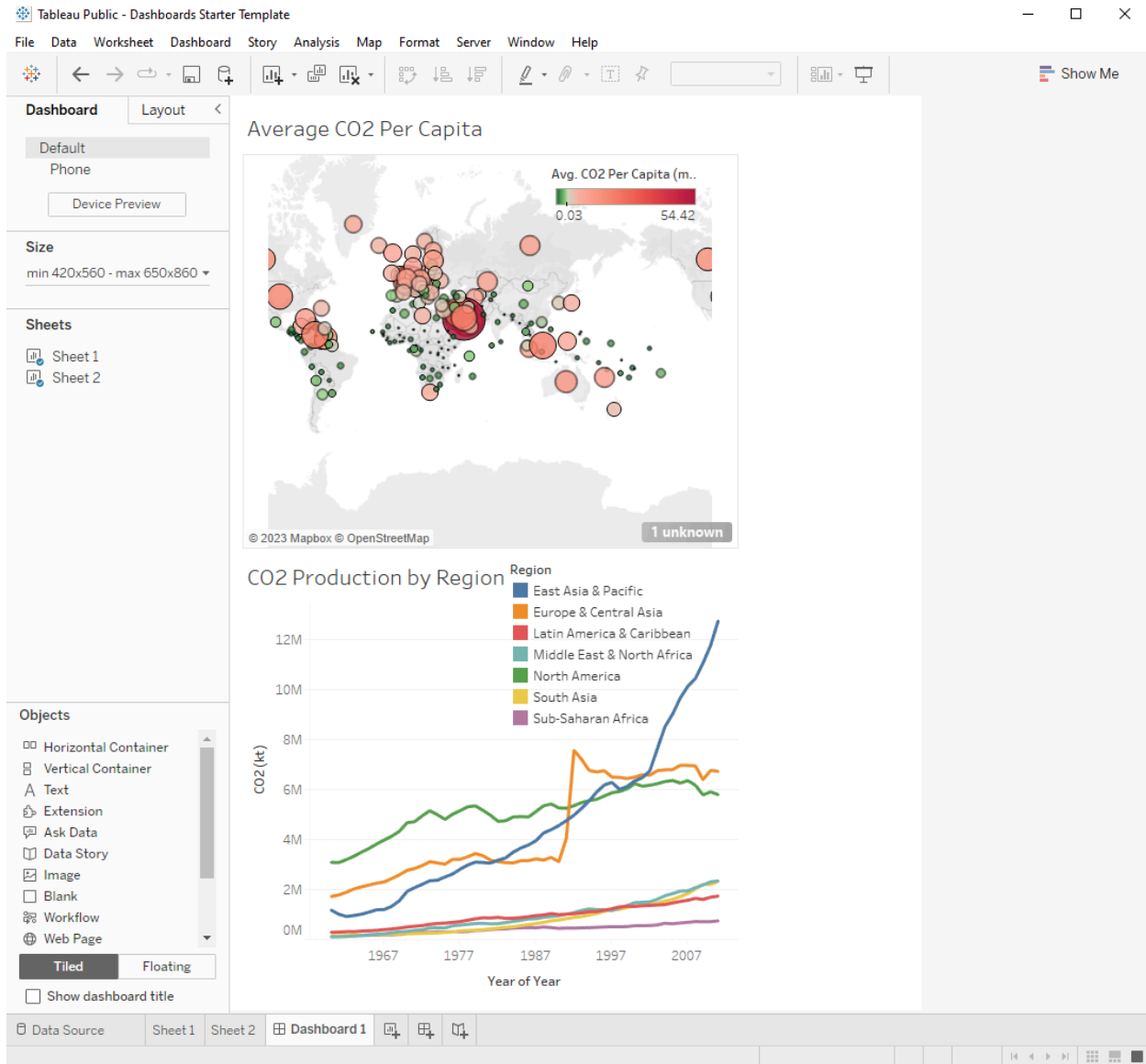
You've now created a basic dashboard. Tableau contains tons of other functionality that allows for dashboards that update automatically or interactive dashboards and visualizations.

Fill in the blank: Dashboard visuals should be _____, which means that they are balanced and make good use of space.

○ interactive

◉ cohesive

○ symmetrical

○ informative

⊘ **Correct**

Dashboard visuals must be cohesive, which means that they are balanced and make good use of space.

What type of layout should you choose if you want your Tableau dashboard to automatically resize itself based on the overall dashboard size?

○ Summarized

○ Filtered

◉ Tiled

○ Floating

✓ **Correct**
If you want your Tableau dashboard to automatically resize itself based on the overall dashboard size, choose a tiled layout.

1. A new challenge from a competitor, a process that needs to be improved, or a potential business opportunity could all represent which aspect of data storytelling?

○ Aha moment

○ Setting

○ Big reveal

◉ Plot

✓ **Correct**
A new challenge from a competitor, a process that needs to be improved, or a potential business opportunity could all be a plot in the data story. The plot reveals the problem and compels the characters to act.

2. What slideshow tool is used to control color, font types and sizes, formating, and positioning of text and visuals?

○ Motifs

○ Patterns

◉ Themes

○ Schemes

✓ **Correct**
The slideshow tool themes is used to control color, font types and sizes, formating, and positioning of text and visuals.

3. Fill in the blank: To ensure your audience is focused on what is being said, rather than reading slides, keep text to fewer than _____ per slide.

○ 10 lines and 100 words

○ 15 lines and 15 words

◉ five lines and 25 words

○ three lines and 10 words

✓ **Correct**

4. What method should be used to add a presentation visual that will automatically update to reflect the latest information whenever the original dataset changes?

○ Embedding

○ Copying and pasting

◉ Linking

○ Screenshotting

✓ **Correct**
To add a presentation visual that will automatically update to reflect the latest information whenever the original dataset changes, link the visual.

1. A data analyst at a software company analyzes sales data to identify top-performing sales representatives. While reviewing a scatterplot of sales volume versus individual representatives, the analyst notices a data point far above the average sales volume. What should they do with this outlier?

   ○ Hide it from the scatterplot

   ○ Move it closer to the main trendline

   ○ Drop it from the scatterplot

   ⦿ Determine whether it is an important observation

2. Fill in the blank: Charts, graphs, and other dashboard elements should be _____, which means they are balanced and make good use of available space.

   ☑ cohesive

   ☐ clean

   ☐ complete

   ☐ constant

3. You are building a dashboard in Tableau. To create a design in which items can be layered over other objects, which layout should you choose?

   ○ Floating

   ○ Tiled

   ⦿ Layered

   ○ Itemized

4. A data analyst adds a visual to a presentation. They want to ensure that any changes made to the original file within its source location are automatically updated in the presentation. Which method should they use when adding the visual to their presentation?

- ○ Screenshot
- ◉ Link
- ○ Embed
- ○ Copy and paste

5. A data analyst at an e-commerce company uses data to identify which customers are most likely to make a purchase. The analyst shares this information with the marketing team in a report that uses data visualizations and storytelling techniques to make the findings clear and actionable. What aspect of data storytelling does this scenario describe?

- ○ Takeaways
- ○ Narrative
- ○ Setting
- ◉ Spotlighting

6. What Tableau tool can be used to limit the data displayed on a dashboard in order to save stakeholders time and effort?

- ○ Pre-building
- ◉ Pre-sizing
- ○ Pre-sorting
- ○ Pre-filtering

7. Which of the following statements correctly describe presentation best practices? Select all that apply.

- ☐ The plot includes recommendations from the data insights.
- ☑ The characters are the people affected by the story, such as stakeholders, customers, or clients.
- ☑ The aha moment involves sharing recommendations.
- ☑ The setting describes background information about the data project and the current situation.

8. You prepare to communicate to clients about an analysis project. You consider how to connect with them and convince them to view the insights as you do. What aspect of data storytelling does this scenario describe?

- ○ Theme
- ◉ Engagement
- ○ Discussion
- ○ Takeaways