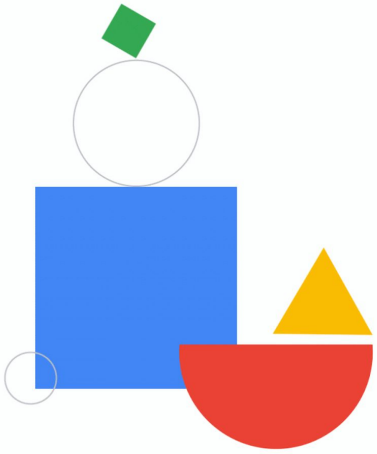
Data Visualization



Next we will cover one of the key outputs and deliverables that data analysts create - the insightful reports that you present to your audience.

Data visualization

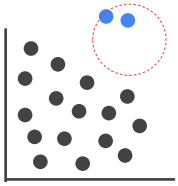
01 Overview of data visualization principles

02 Exploratory vs explanatory analysis approaches

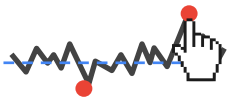
03 Introduction to Looker Studio

This data visualization module covers a little visualization theory and best practices. It then introduces Looker Studio as one of the visualization tools in your toolkit for creating actionable reports.

Use visualization to clearly and concisely present insights



Visualizing a dataset lets you spot hidden trends



Interacting with a dataset visually is often faster than writing SQL

Deliver powerful insights to your audience through reports

Get scalable performance as your dataset grows with BigQuery-backed visualization tools

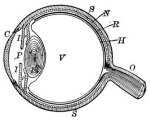
Visualizing data is a key part of the data analyst skill set. Building interactive deliverables like charts and graphs is a great way to explore datasets. A visualization tool can also help a more visual thinker uncover insights slightly more clearly than just poking around a dataset with SQL!

So four quick things that we are going to highlight.

1. Visualization provides a unique view of data that often lets you spot hidden trends.
2. You can interact with the dataset, enabling you to tell a story that has multiple layers. For example, you could start with time series data, click into one of the anomalous spikes or troughs in the data, and then drill down into those details visually. So it's a natural tool to let your audiences follow a cohesive story throughout the flow of your explanation.
3. Dashboards and reports are a fast and effective way to reach a broad target audience. They can be visually aesthetically pleasing as well!
4. If much of your data is already in BigQuery, integrating with a visualization tool like Looker and Looker Studio will naturally get all the performance benefits of having your data and your queries processed in BigQuery. It’s then displayed and rendered quickly on the frontend in a visualization tool.

Visualization theory: Perception

Stimulus



Eye

Brain

Preattentive processing
Attentive processing

Sensation
Physical process

Perception
Cognitive process

No lecture on data visualization will be complete without mentioning visualization theory. Visualization is both an art and a science. So here we are going to get some of the science of what our brains perceive when you look at a beautiful visualization.

In this particular case, we have a cat as a stimulus. Your eye immediately recognizes it and says, "Hey, I've seen that before, that's a cute little kitten." And your brain automatically says, "Hey, I've seen a thousand of these before, I immediately know without thinking that it's a cat."

It's harder for machines to have that intuition. Humans have what we call pre-attentive processing, where we can immediately recognize things.

So what does that mean for your data visualizations? It means you can effectively cheat the brain by using common human intuition to not have the brain do much work. We want to be able to make snap judgments very quickly, and then only tap into that really focused thought power when it's required. Look at an example.

Visualization theory: Count the fives

```
69750429347493732418605783578
58728294974654487818676453214
24439684634233529867321903875
65878893745390932975659391732
14725920189374476564722175652
```


So take a few seconds and count all the fives that are present.

Did you get them, or are you still looking?

Now, if you counted 16 fives in the 10 seconds that I gave you that's amazing! It's very hard to pick out those fives from this noise of numbers here.

Most of us will read from left to right, serially scan every row, count out all those fives, and keep track of them. There has to be an easier way to do that, and naturally you might expect we could do something like this.

Visualization theory: Count the **fives**

697**5**042934749373241860**5**783**5**78
5872829497465448781867645**5**3214
244396846342335298673219038**5**
6**5**878893745390932975**5**9391732
14725920189374476**5**64722175652

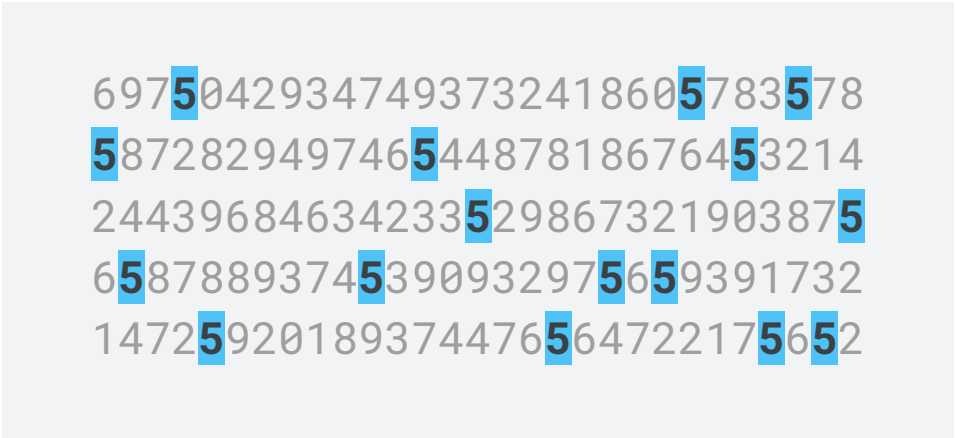
Now count the fives again.

How about this time around, was it much easier? Think about what your brain actually did to count those 16 fives and why it was much faster for you.

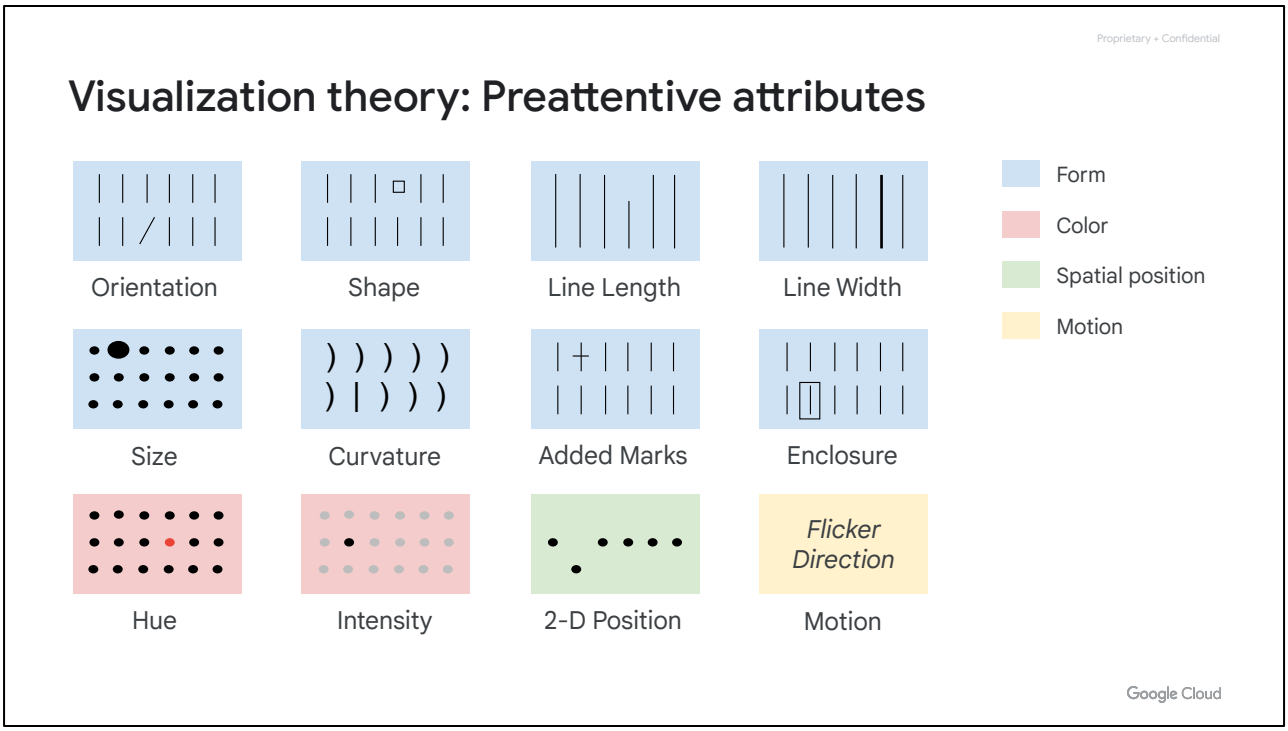
So you might say, well, you bolded the fives and that made them a lot easier to see. That's true. Think of the theory behind what makes it easier.

Bolding probably enabled you to visually cluster those fives into smaller chunks and quickly count them. From a visualization theory perspective, when you contrast certain elements you highlight the focus. In this example, it allowed us to treat all the numbers that aren't fives as background noise that we can safely ignore.

Visualization theory: Count the fives



We can take it a step further and add what we call visual encoding to these particular elements that we want to focus on. So here we introduce an element of color where you can further highlight the elements that you want your audience to focus on. It gives greater attention to what you're calling out as key points.



There are various methods that you can use to "cheat" the brain, and then really tap into that instant decision making that we have developed over the years.

You can play around with things like, the orientation, the shape and the skew, the length of a certain attribute, or the size of the mark on the page. Maybe some elements are curved and some aren't. You could add a box around it, change the intensity or the hue, move the position around, or even add a motion element.

All of these methods can help the brain focus in on what's important, making your visualization stand out which will efficiently and effectively convey the key message to your audience.

You choose: Effective or ineffective (or wrong) visuals

Okay, now we're going to play a little game. There's many different ways to visualize the same data.

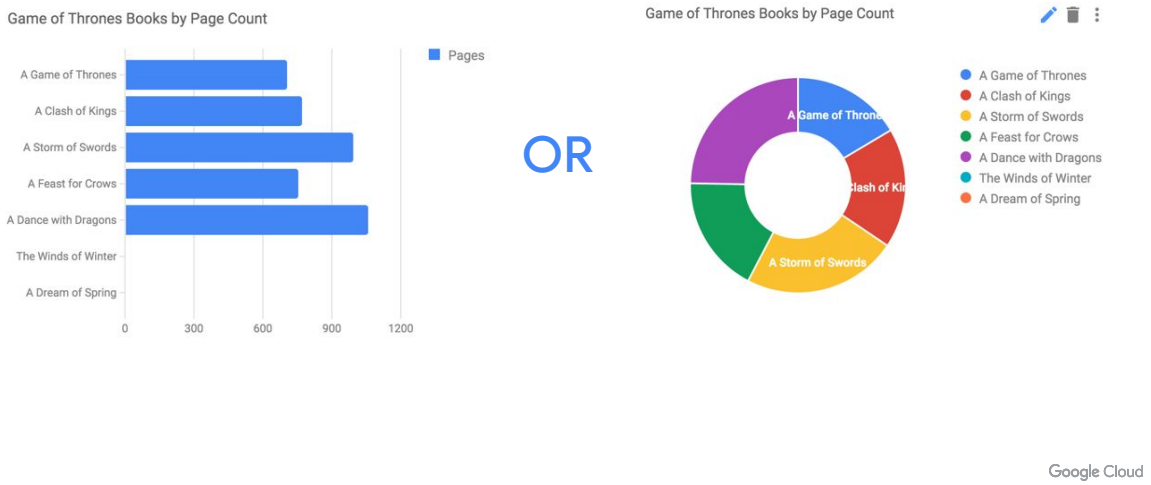
Here's an example and I want you to think which visual presents this dataset better. The dataset is Game of Thrones characters by gender.

Most would probably say the one on the left. It provides a much clearer picture of the division between the count of male characters and count of female characters. However, the graph on the left doesn't have any type of quantitative details to show what the actual values are. The graph on the right does give you the actual numeric count of those characters.

So both of these graphs could definitely be improved.

This exercise is similar to the count-to-fives exercise where we added the bolding and blue coloring. Adding additional encoding measures like data labels, a color, a good use of whitespace, and so on, can progressively help to make your message clear. But at the same time, you don't want to overload the audience with too much happening on the screen. Try another one.

You choose: Effective or ineffective visuals

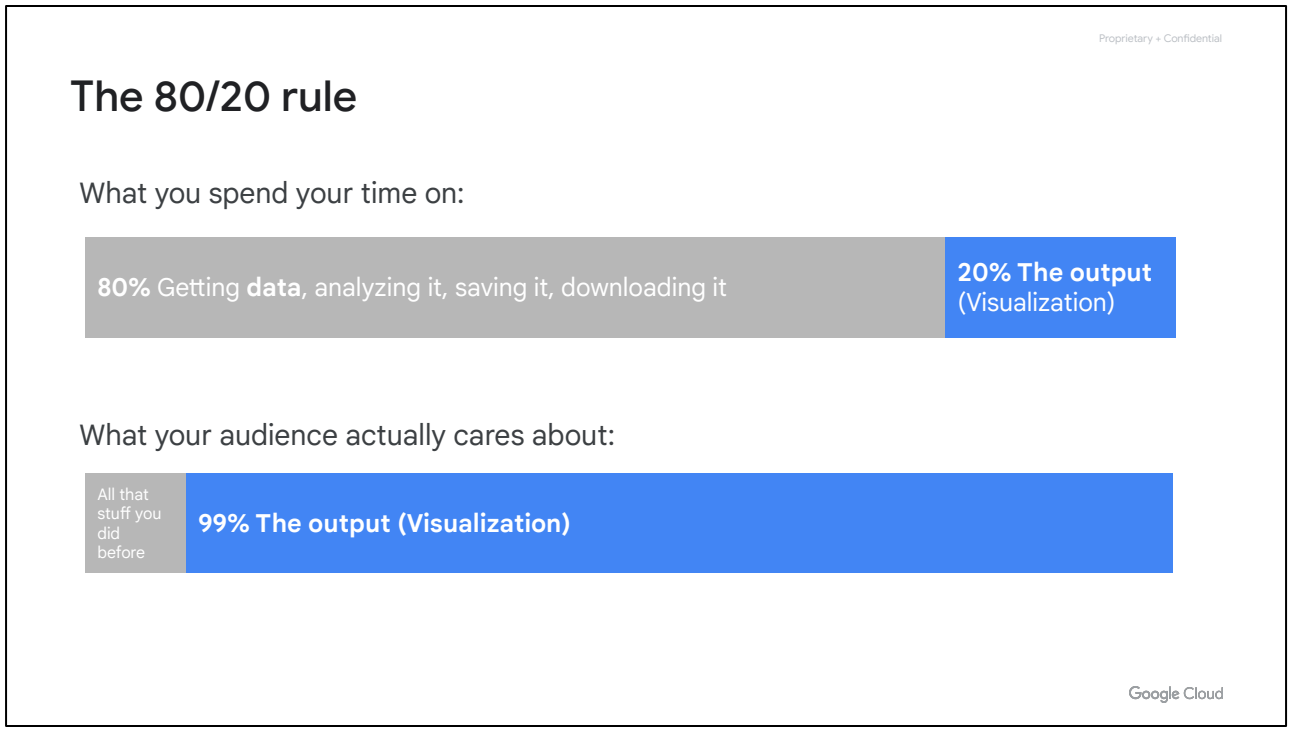


Now we have the same doughnut chart on the right and a horizontal bar chart on the left. This is the dataset of Game of Thrones books by page count.

This doughnut chart has labels this time, but is a bit hard to read because you have white text on a potentially white background. And then you have two books which aren't shown, The Winds of Winter and A Dream of Spring. Because you can't display negative or zero values on a doughnut chart, these books exist in the legend but are not represented in the chart itself.

The bar chart on the left does show the last two books: The Winds of Winter, and A Dream of Spring. Why would those have zero values for the page count? Either they haven't been written yet or there's some data missing!

Consider what the message is that you want to convey. If it's comparing page count between the books, the best may be something like a bar graph with an extra label saying how many pages are in the actual book.



The 80/20 rule is one of my favorite slides when talking about visualization theory. If we had to do a pie chart on how you actually spend your time, maybe about 80 percent of that pie chart is going to be filled with sourcing, analyzing, and deriving insights from the data. Maybe 20 percent of your time is actually going to be spent on building those beautiful visualizations and sharing them with your peers.

However, a beautiful picture or visualization of those insights is largely what the audience is actually going to care about. How you're delivering that message is just as important, if not more important, than the actual methods used to get those insights in the first place!

Visualization core concept: Dimensions and measures

	Description	Examples
1. **Dimensions**	Independent variable Categorical information	• Name • Location • Part number # • Job title
2. **Measures**	Dependent variable Any field containing quantitative information	• Revenue • Salary • Expenses • Count of errors

One of the core concepts that we're going to dive into is dimensions versus measures. You can think of the data fields and your datasets as ingredients that you can use to create beautiful visualizations.

A dimension is a field that you can classify, that is an independent variable. I like to think of dimensions as your qualitative or categorical information, anything that's not necessarily mathematical.

Whereas your measure is going to be dependent. Measures are generally going to be quantitative information, things like revenue metrics, expenses, salaries, counts, or sums. And that doesn't mean that one particular field can't serve as a potential dimension and measure. For example, if you had the employer identification number, what would that be? Would that be a dimension or measure? If you said dimension, you're correct. Just because it's an integer, or a number, doesn't necessarily mean it's a measure, because that's a qualitative piece of information. But if you had the count of all those employer identification numbers, you could then treat that dimension as a measure.

Class question

Which of these are measures?

1. Phone number
2. Employee ID
3. Age
4. Date of birth
5. Tenure at work (in years)
6. Job title

Remember, measures are usually quantitative fields

Which of these are measures? Once again Measures are quantitative fields. Which ones can you perform math on? So, from top to bottom.

Number one, phone number: is definitely a dimension. You're unlikely to perform math using a phone number.

Employee ID: no, not technically. But again, if you're doing things like counts of employees, you can apply functions on top of your dimensions to treat them as measures.

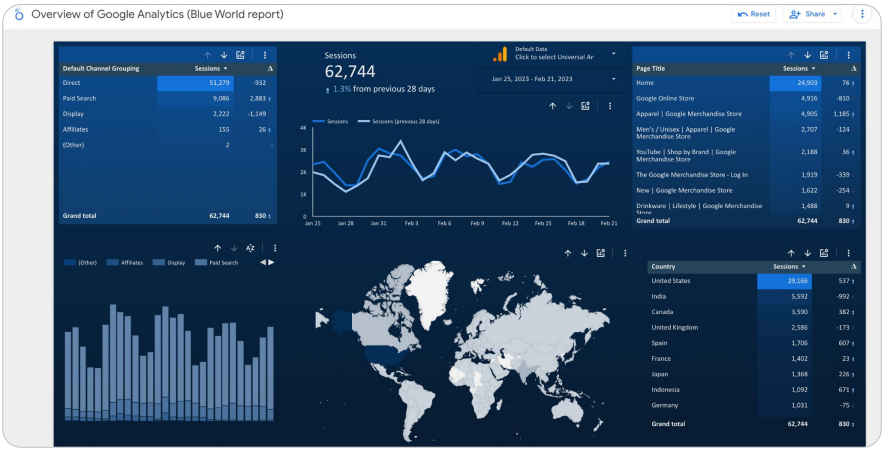
Age: absolutely. You could do average age, you could do mean, max, and so on.

Date of birth: I wouldn't really consider that as a measure.

Tenure at Work: in years, definitely. That's something that would make sense to do math on.

And Job Title: I wouldn't consider as a measure either, because again that's qualitative information.

Reports transform data into information



- Tell a clear story with your data
- Share and collaborate on reports with others

Google Cloud

Let's talk about one of the primary outputs you're going to produce, the report.

The report is a canvas for you to tell a message or story about the insights that you've gathered. And to tell it in a clear, effective, and well laid out manner, so that your audience gets the message very quickly.

And again, think back to the count of fives example. You want your audience to hone in on those really key insights that they care about within the first 10 to 15 seconds.

You can also collaborate on and share these reports with your peers as you're developing them.

Data visualization

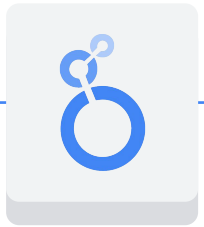
01 Overview of data visualization principles

02 Exploratory vs explanatory analysis approaches

03 Introduction to Looker Studio

Now it's time to look at Looker Studio, one of the ways that you can visualize the insights I've been talking about.

Looker versus Looker Studio



Looker

Platform for BI, data applications, and embedded analytics

Looker Studio

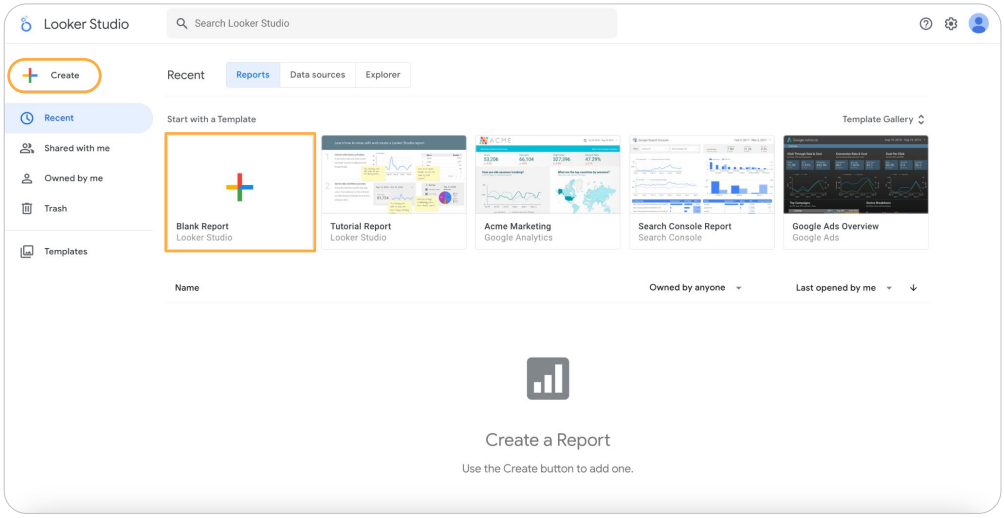
Interactive data suite for dashboarding, reporting, and analytics

First, I'll differentiate between Looker and Looker Studio.

Looker is a business intelligence and big data analytics platform that helps users explore, analyze, and share real-time business analytics.

Looker Studio, formerly known as Data Studio, is a web interface that makes it easy to create interactive dashboards and reports from a wide variety of sources, driving smarter business decisions. Data sources act as pipes to connect a Looker Studio report to underlying data.

Create new reports in Looker Studio

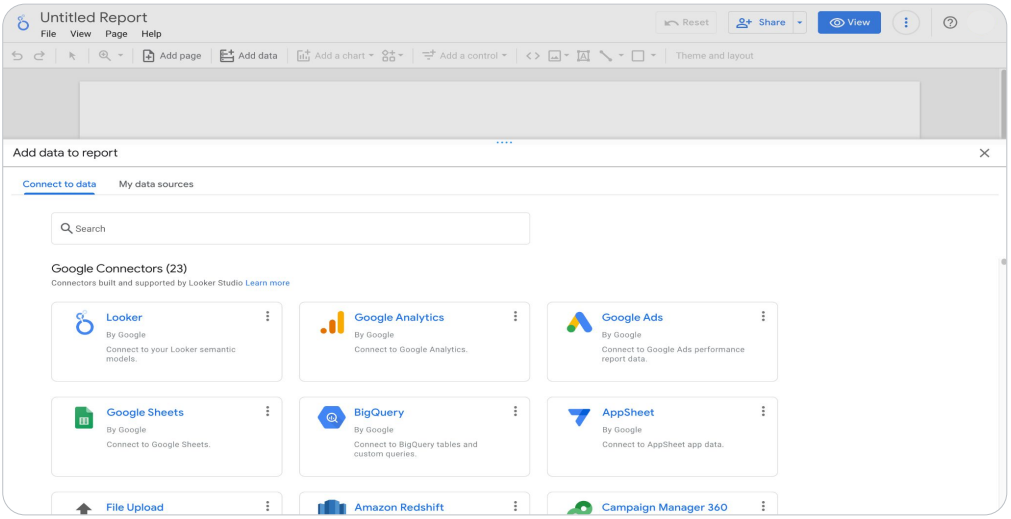


Here's the Looker Studio home screen. You can see the reports that you created previously, you can create new reports, or copy existing reports and use them as a starting source.

There are two ways to create a new report from scratch.

- Select **Blank Report** in the templates panel in the middle of the screen.
- Or click the **Create** button in the navigation pane on the left of the screen.

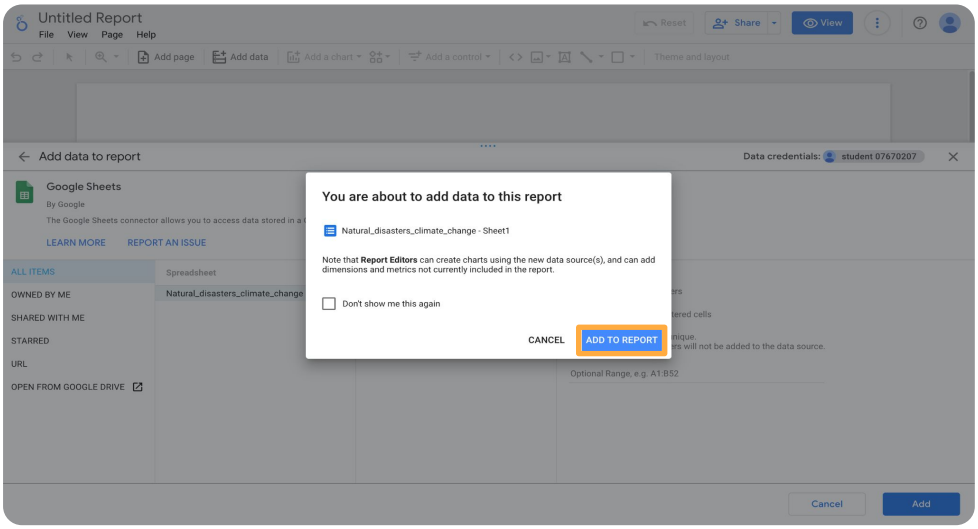
Connect to multiple different types of data sources



You can have any or all of these data sources in a single Looker Studio report.

In addition to the Google Connectors, there is an increasing list of Partner Connectors to choose from as well.

Add the data source to your report



Since Looker Studio reports can be shared, be aware of the ramifications of adding a data source.

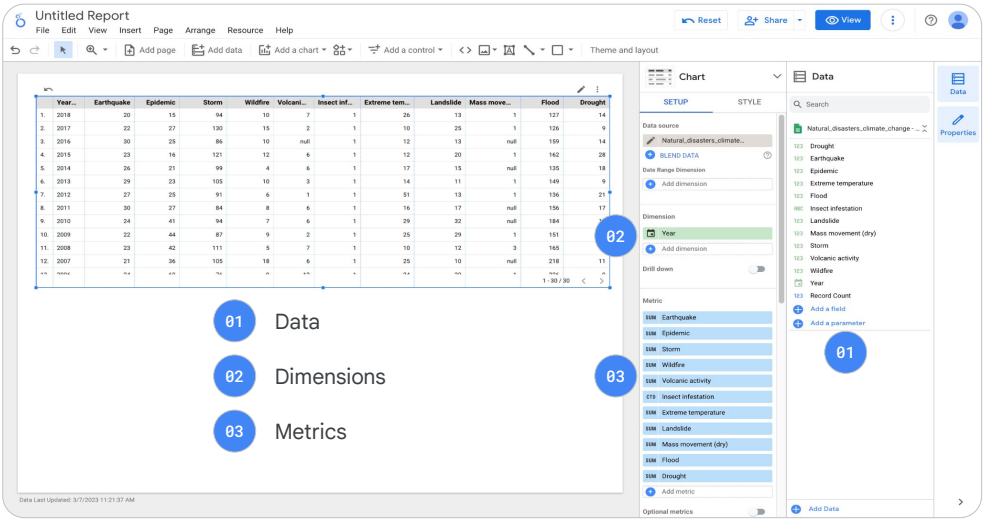
When you add a data source to a report, other people who can view the report can potentially see all the data in that data source.

And anyone who can edit the report can use all the fields from any added data sources to create new charts with them.

I'll talk about how to control access to data and sharing later on in this course.

Click **Add to report**.

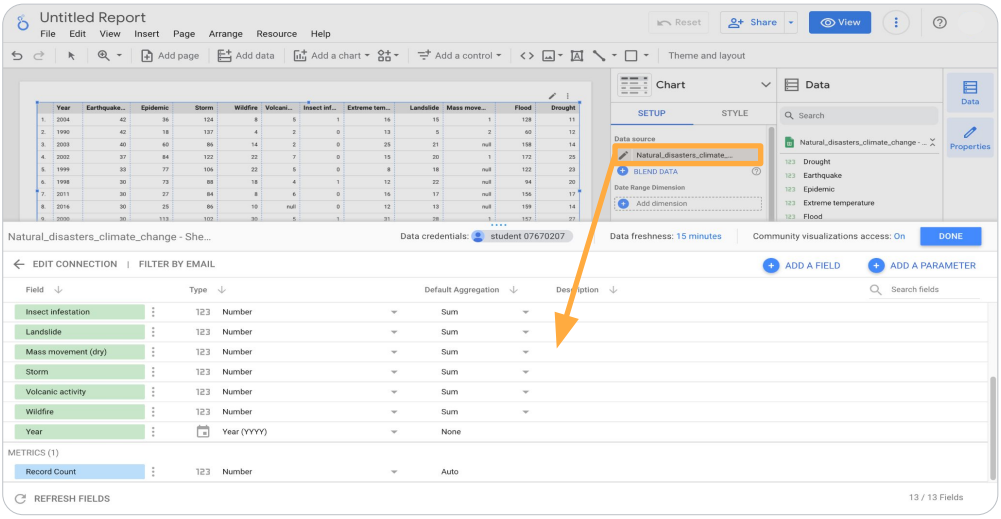
Select your data fields to build your visualizations



Having selected a dataset, you can specify what elements of the dataset you want to visualize.

Selections include the **Dimensions** and **Metrics** that you want to use from the **Data** of your dataset.

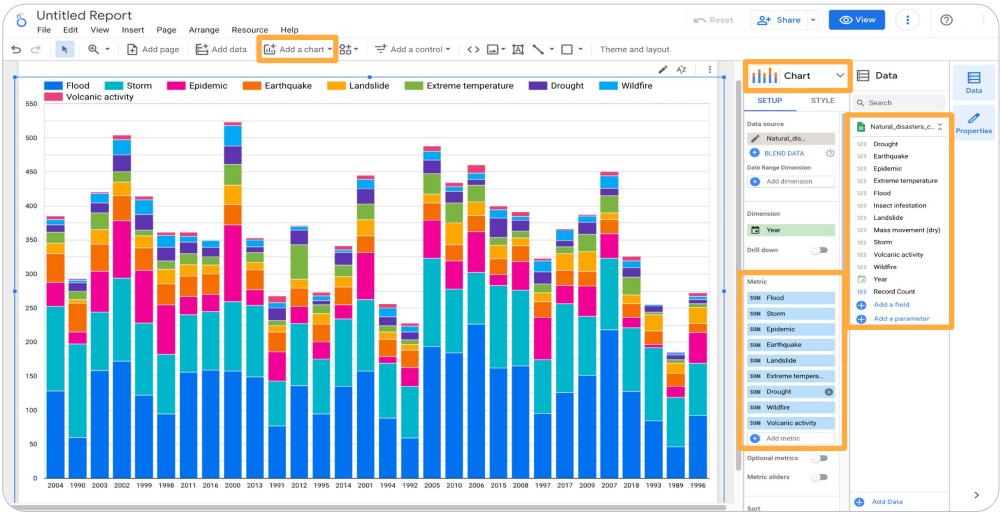
Edit your data source fields, if necessary



Google Cloud

To edit the dataset fields, select the **Edit data source picker** in front of the **Data source** name.

Create charts to visualize data relationships



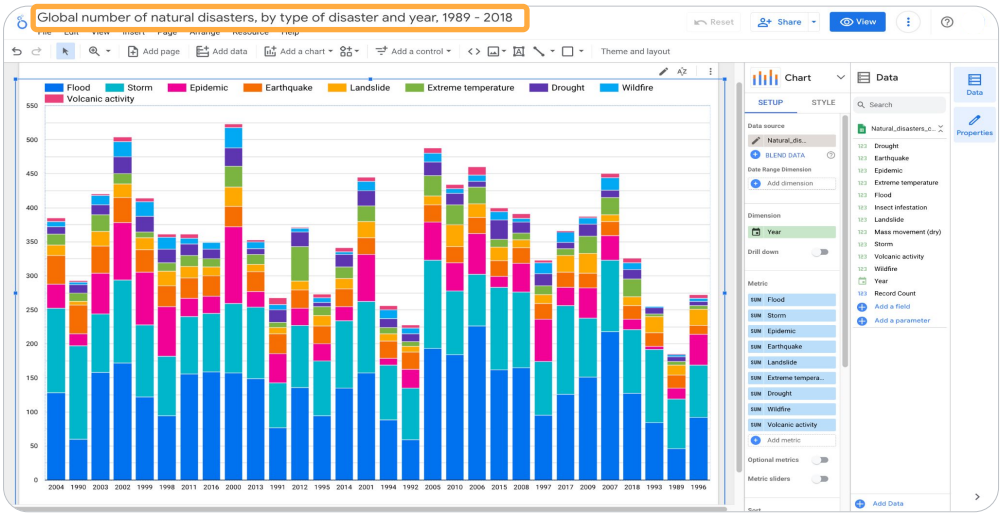
Easily change your **data table** view to a **chart** by clicking **Chart** in the properties panel and selecting a chart type from the options provided.
You can edit the style of your chart, or even change your chart type selection.
You can also revert back to a **data table** view.

You can also add separate charts by selecting Add a chart from the toolbar.
To arrange data tables and different chart types as required, resize the components on the canvas.

In the same way that you defined **Dimensions** and **Metrics** earlier,
you can do the same for your chart by adding selections from the **Data** list.

Here's a useful tip: The sequence of the fields under **Metric** will determine the order in which the data is displayed in the chart.
Use the drag feature to easily change the sequence of the fields.

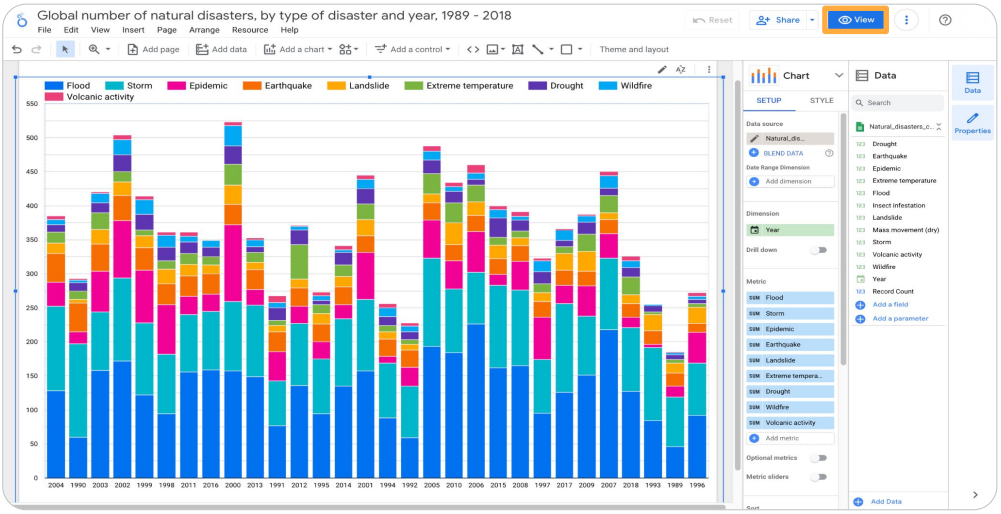
Add a descriptive name to your report



Give your report a name.

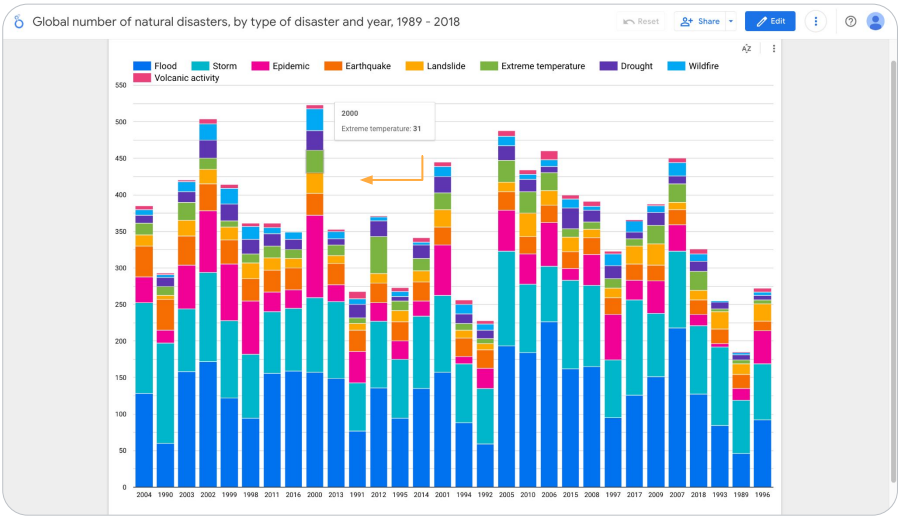
Since Looker Studio is based on Google Drive, you can have duplicate filenames.

View the end-user version of the report



Click the **View** toggle button to view the end-user version of the report.

View your report as an end-user



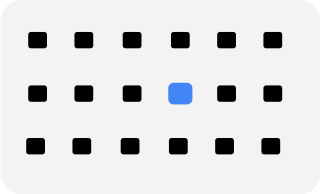
And here is your report. Notice it looks very similar to when you were editing it, but as a viewer, you cannot modify the report.

When a viewer mouses over the chart, they are able to view live data.

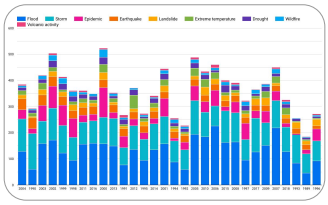
In this example, the viewer is able to see that in the year 2000, there were 31 natural disasters related to extreme temperature.

Users cannot edit your reports unless you give them permission.

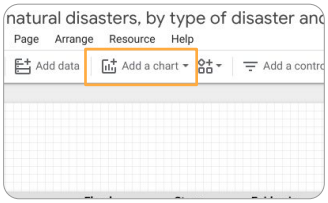
Summary: Explore and present your insights visually



Guide the eye of your user with preattentive attributes



Use the right visual to convey the right message



Create new report charts inside Looker Studio

As we wrap up this module, we see that visualizing data is both an art, and a science. We've barely scratched the surface of visualization theory, and we discussed concepts like …

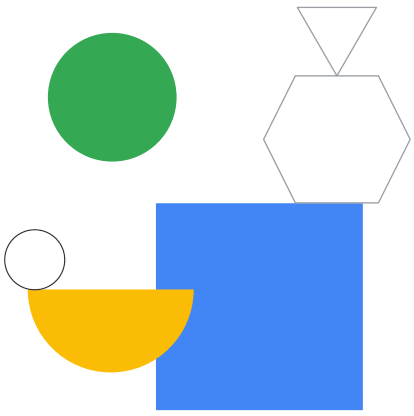
… pre-attentive versus post-attentive processing for that quick eye-to-brain understanding.

Along the way, we saw some bad ways to visualize data, as well as a few practices to do it right.

Lastly, we looked at Looker Studio, which is the visualization platform we'll be exploring in more depth in our next lab.

Lab Intro

Explore and Create Reports with Looker Studio



Now it's time for you to explore an ecommerce dataset that has millions of Google Analytics records for the Google Merchandise Store loaded into BigQuery.

In this lab, you'll pair your knowledge of the ecommerce dataset with what you just learned about visualizing dimensions and measures. You'll create a new report and add visuals and interactive filters for your reporting users.