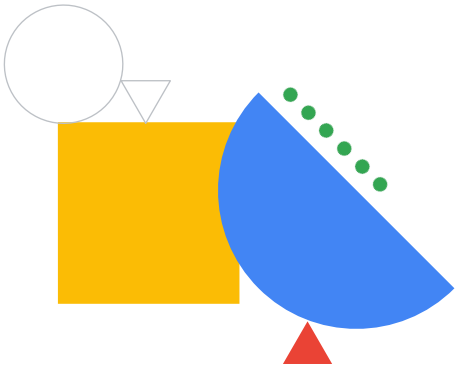
Google Cloud

Visualizing Insights and Creating Scheduled Queries



Agenda

01 Data Visualization Principles

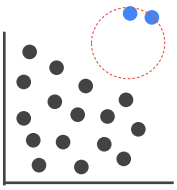
02 Common Data Visualization Pitfalls
- Demo: Creating a Scheduled Query

03 Google Data Studio
- Demo: Google Data Studio Walkthrough
- Lab: How to Build a BI Dashboard Using Google Data Studio and BigQuery

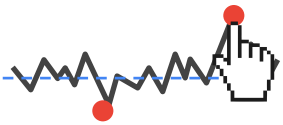
One of the key outputs data analysts create are insightful reports for your audience. In this data visualization module we will cover a little visualization theory and best practices and then we'll introduce Google Data Studio as one of the visualization tools in your toolkit for creating actionable reports.

Data Visualization Principles

Use visualization to clearly and concisely present insights



Visualizing a dataset allows you to 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

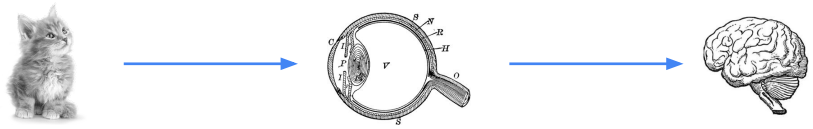
Visualization helps make sense of data.

Visualization tools make it easy to “see” data interactively.

Visualization tools that support BigQuery can give “small dataset” performance against billion+ row data.

Build visualizations without writing SQL.

Visualization Theory: Perception



We don’t see things with our eyes, we see things with our brains.
Eyes record light, translate it into electrical signals, and pass it to your brains, where the message is understood.

Preattentive processing:

- Iconic memory: <1s, automatic, unconscious

Attentive processing:

- Short-term memory: 4 chunks of information at most (color, shape, size, number, etc.)
- Long-term memory: holds more information, not our focus today

Visualization Theory: Count the fives

```
69750429347493732418605783578
58728294974654487818676453214
24439684634233529867321903875
65878893745390932975659391732
14725920189374476564722175652
```


Answer: 16
Took so long because you had to use focused attention system and scan every single row and value.

Visualization Theory: Count the **fives**

697**5**042934749373241860**5**783**5**78
5872829497465448781867643214
244396846342335298673219038**7**5
6**5**878893745390932975659391732
14725920189374476564722175652

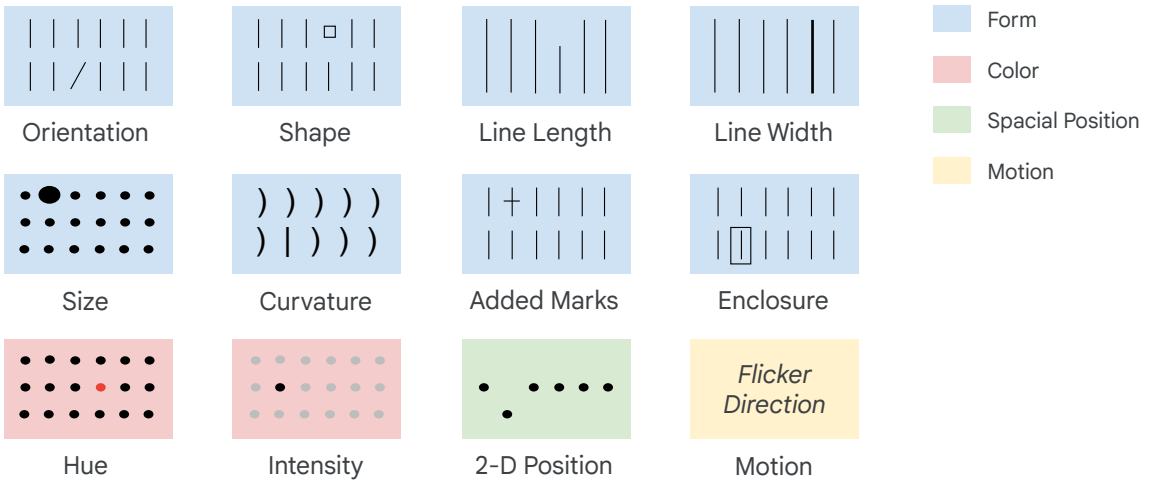
Much easier with the help of preattentive processing - features extracted quickly, effortlessly, in parallel without any attention focused on it.

Visualization Theory: Count the **fives**

697**5**042934749373241860**5**783**5**78
5872829497465448781867645**5**3214
244396846342335298673219038**75**
6**5**878893745390932975**5**6**5**9391732
14725920189374476**5**6472217**5**6**5**2

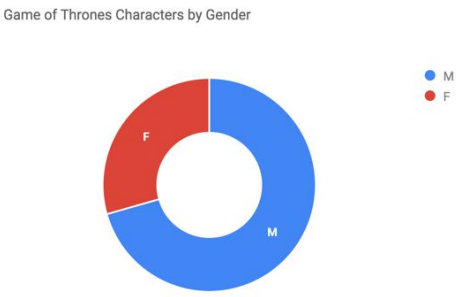
Even easier with visual encoding.

Visualization Theory: Preattentive attributes

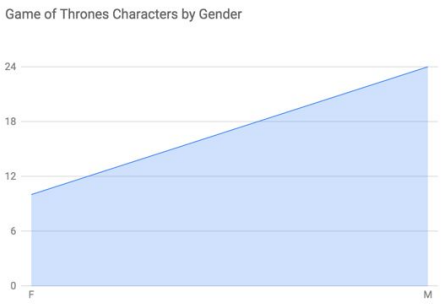


Use encoding to highlight the relevant information and help your audience understand your message in their preattentive space. That leaves the attentive space for them to listen to what you're saying!

You choose: Effective or ineffective (or wrong) visuals



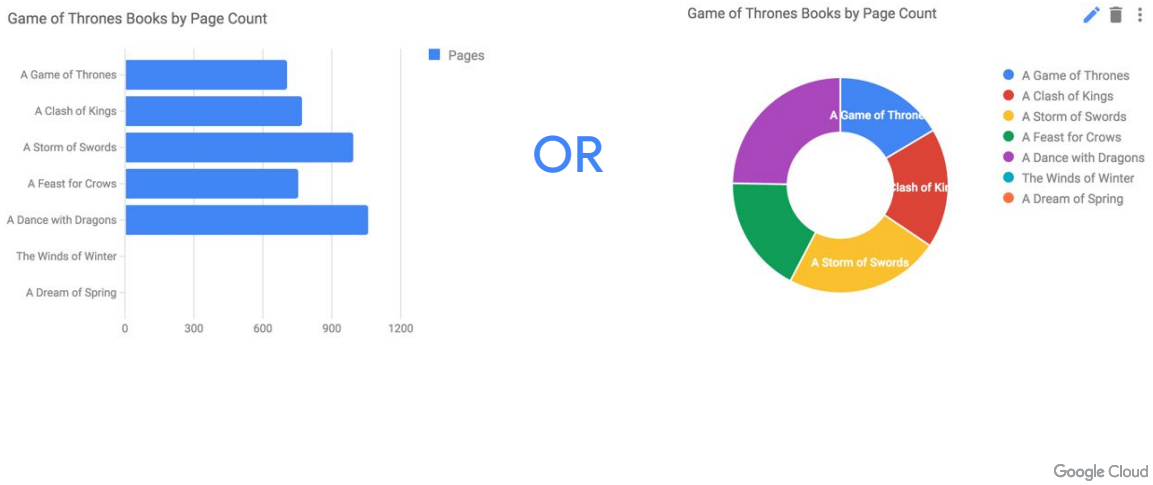
OR



Donut Chart for quick comparison of book characters by gender is likely the better visual.

Line graphs are mainly for time series and does not make sense to use here.

You choose: Effective or ineffective visuals



Horizontal bar chart allows you to have dimensions with valid 0 values (like two books that have not been written yet).

Donut charts can get crowded with multiple dimensions.

The 80/20 rule

What you spend your time on:

What your audience actually cares about:

Your audience is likely only to see the end product of your work. Your visuals and reports often carry more visual weight than your SQL scripts and analysis.

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

Class question

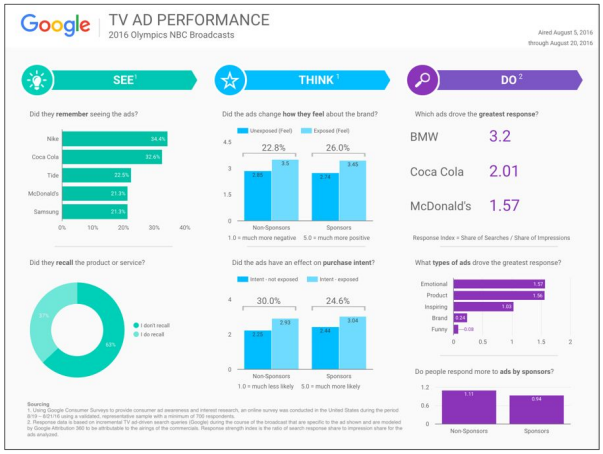
Which of the below are measures?

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

Age and Tenure (you can do math on them)... in reality, you can do COUNTs of unique Job titles and that would become a measure.

Reports transform data into information



Tell a clear story with your data

Share and collaborate on reports with others

Here's an example of an actual Data Studio report. You don't have to understand the numbers, but just on the surface, what is this report trying to do?

Role of the report:

- Visualize data
- Restrict data
- Share with viewers and editors

Common Data Visualization Pitfalls

Common pitfall: “My dashboard takes too long to load!”

Where is your dashboard pulling data from?

Dashboard powered by **Tables** (stored data)

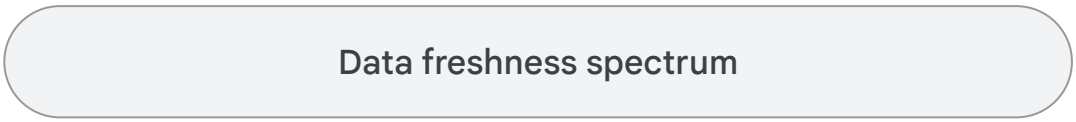
vs

Dashboard powered by **Views** (stored query)

Common Pitfall: “My dashboard takes too long to load!”

Create permanent reporting **tables** that pre-calculate fields for you to visualize

Power your dashboard entirely off of SQL **views** which query the raw data each time



Can be stale but really fast to visualize

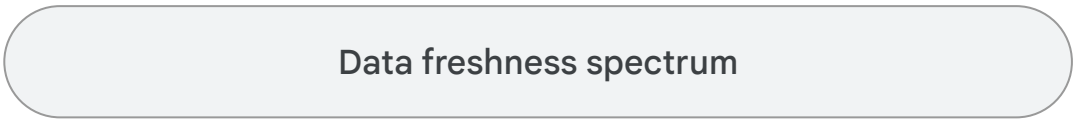
Fresh but can be performance intensive

Common solution: Create reporting tables that refresh on a regular cadence

Create permanent reporting **tables** that pre-calculate fields for you to visualize

Create reporting **tables that automatically refresh** at regular intervals

Power your dashboard entirely off of SQL **views** which query the raw data each time

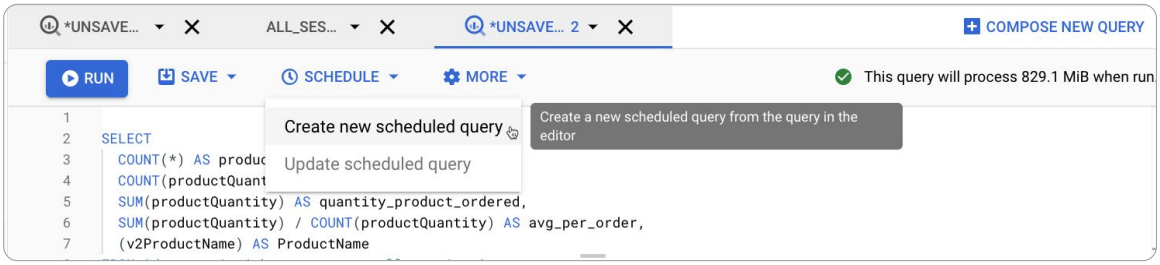


Can be stale but really fast to visualize

Create automatically updated reporting tables

Fresh but can be performance intensive

In BigQuery, you can create scheduled queries that will save your results periodically into a table



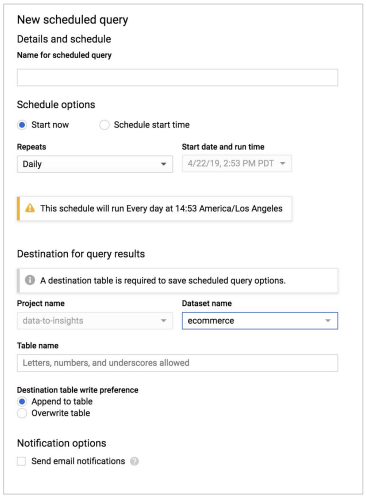
https://cloud.google.com/bigquery/docs/scheduling-queries

You can have multiple saved scheduled queries

Reporting weekly query:
Runs **Friday** at 4am

Reporting daily query:
Runs **daily** at 4am

Reporting archival query:
Runs **monthly** at 4am

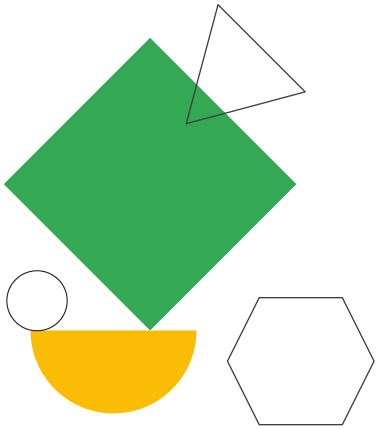


https://cloud.google.com/bigquery/docs/scheduling-queries

Demo

Creating a Scheduled Query

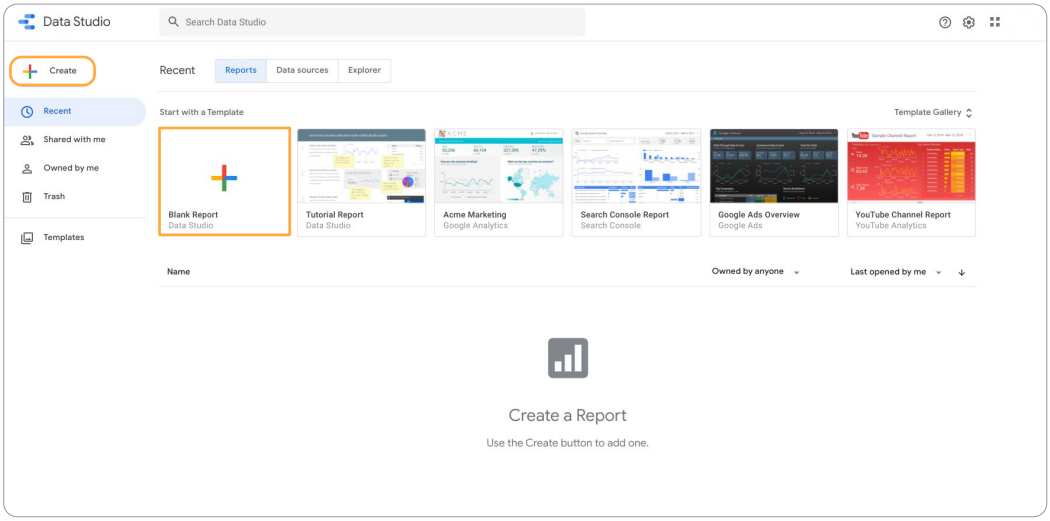
Automatically updating a data table at a set interval



Refer to
https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/data-to-insights/demos/scheduled-query.sql

Google Data Studio

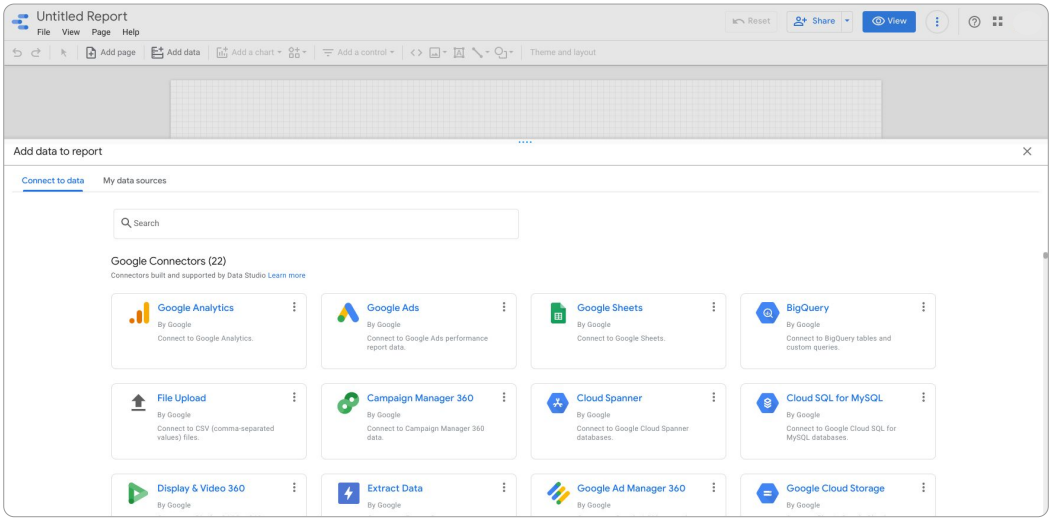
Create new reports in the Data Studio UI



This is the Data Studio Home page. There are two ways to create a new report from scratch.

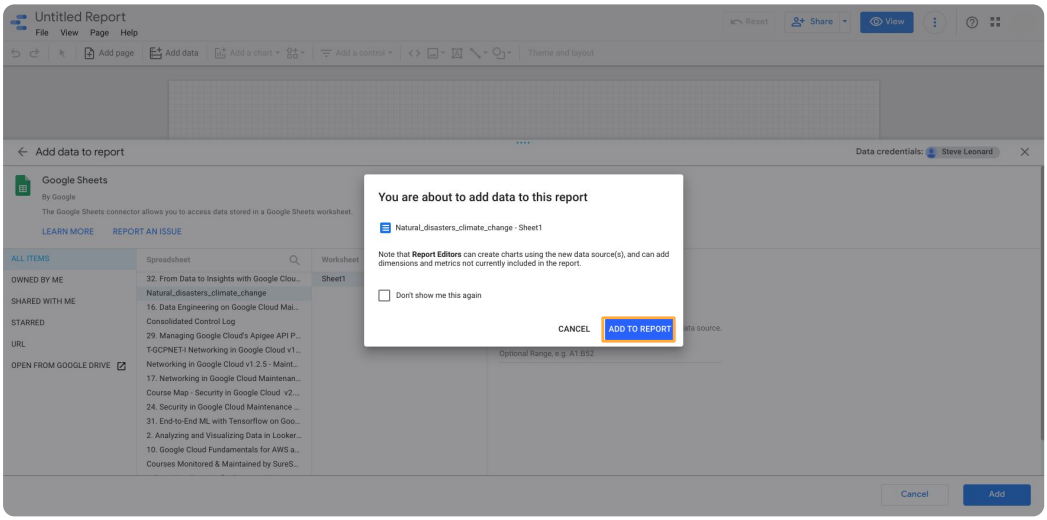
1. Select **Blank Report** in the templates panel in the middle of the screen.
2. Click the **Create** button in the navigation pane on the left of the screen.

Connect to multiple different types of data sources



Note that you can have any or all of these data sources in a single Data 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



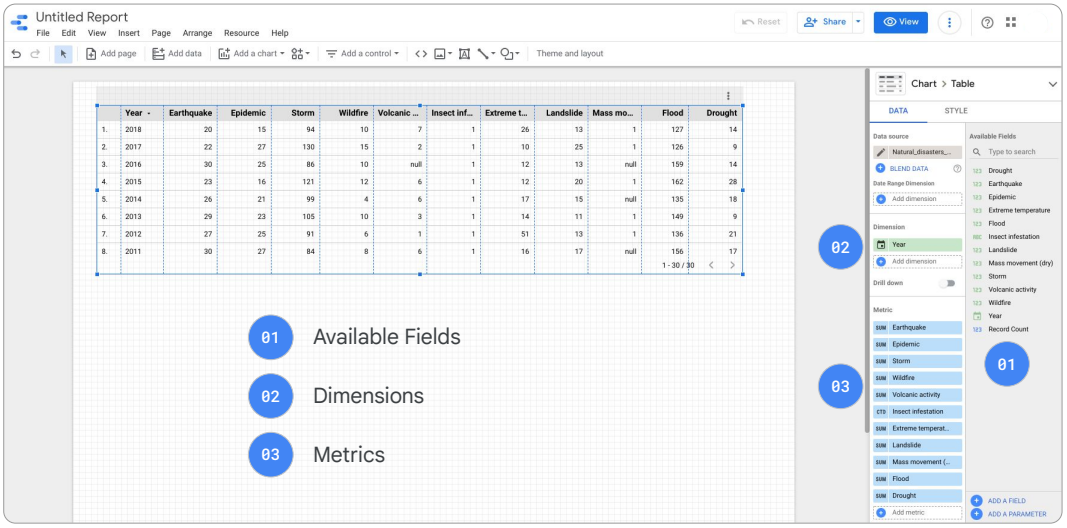
Since Data Studio reports can be shared, you should be aware of the ramifications of adding a data source.

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.

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

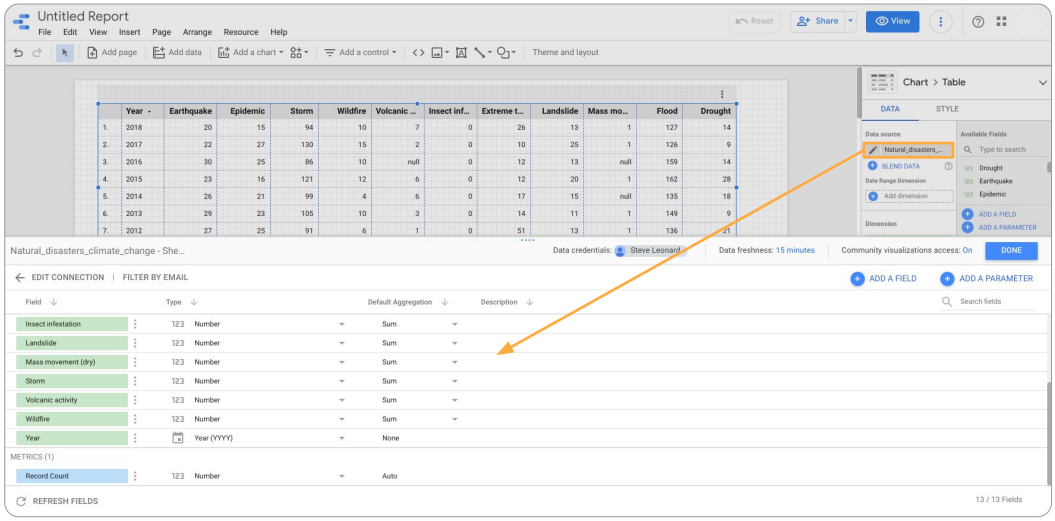
Click ADD TO REPORT

Select your data fields to build your visualizations



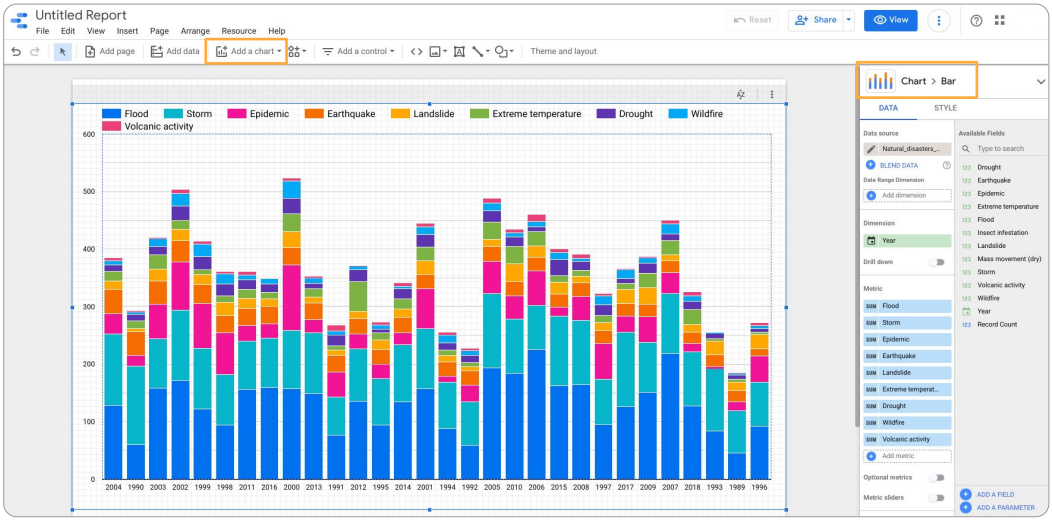
Having selected a dataset, you can specify what elements of the dataset you wish to visualize. This includes selecting the **Dimensions** and **Metrics** that you want to use from the **Available Fields** of your dataset.

Edit your data source fields, if necessary



The **Edit data source picker** in front of the **Data source** name can be selected to edit the dataset fields.

Create charts to visualize data relationships



Google Cloud

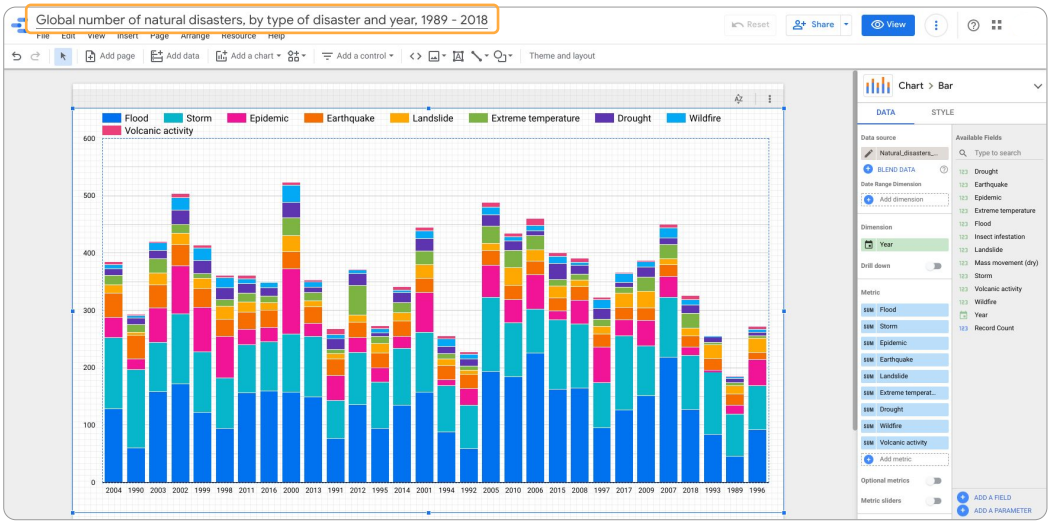
Easily change your **data table** view to a **chart** by clicking **Chart** in the properties panel and selecting a chart type from the options provided. You can edit the style of your chart, or even change your chart type selection, later. You can also revert back to a **data table** view.

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

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

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.

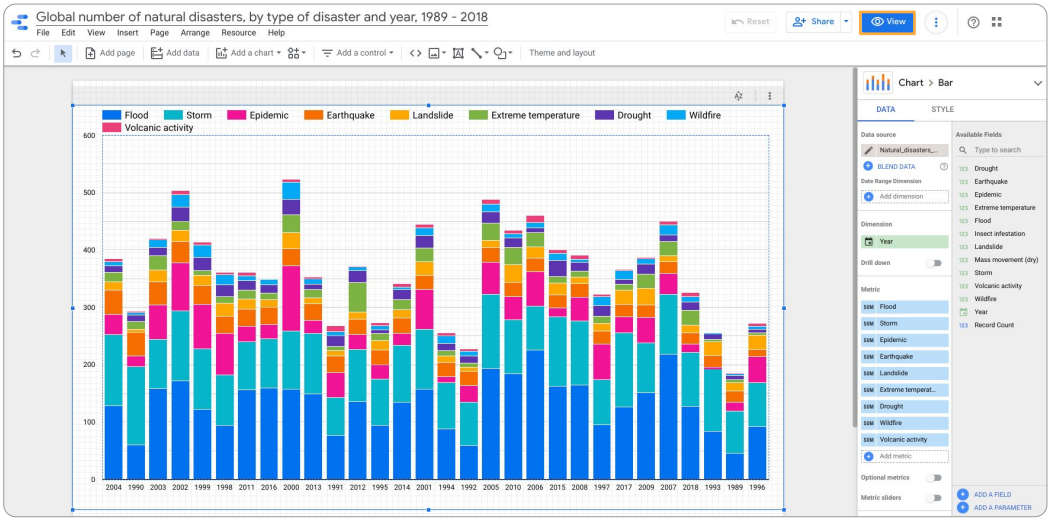
Add a descriptive name to your report



Give your report a name.

Since Data Studio is based on Google Drive, note that you can have duplicate file names.

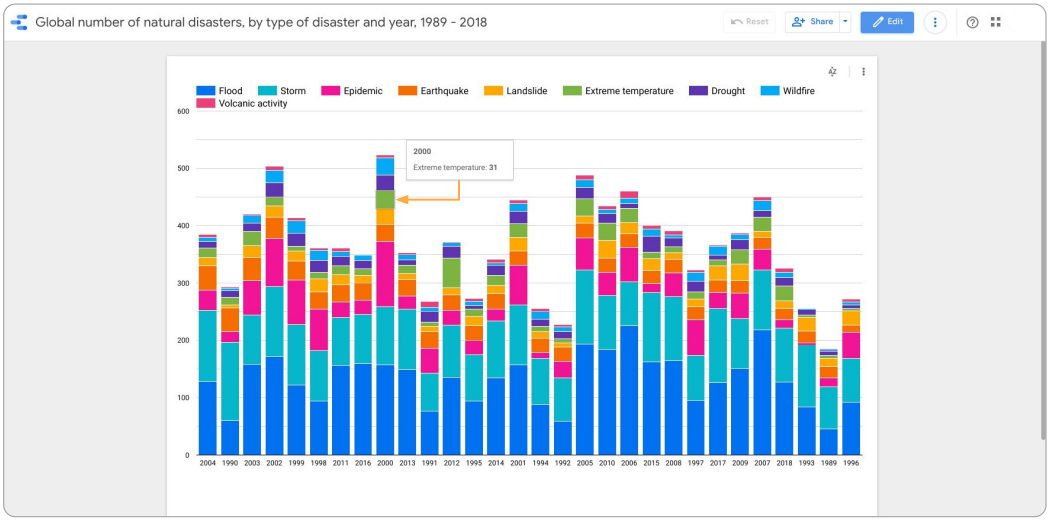
View the end-user version of the report



Google Cloud

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 can't modify the report.

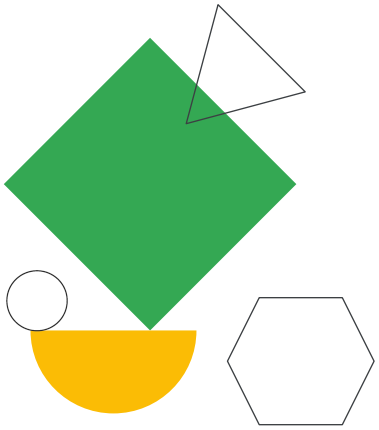
When a viewer mouses over the chart, they are able to view live data. In this example, the viewer is able to see that in the year 2000, there were 31 natural disasters attributable to extreme temperature.

Note that users cannot edit your reports unless you give them permission.

Demo

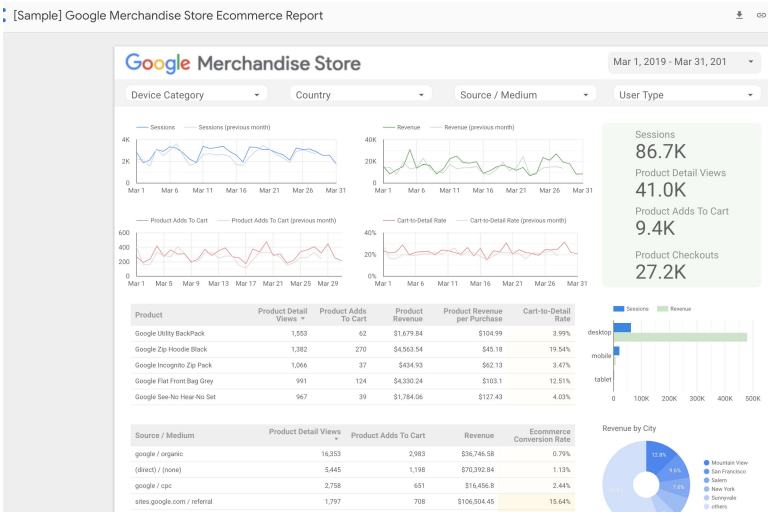
Google Data Studio Walkthrough

Data Studio basics, templates, copying and editing charts



Refer to
https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/data-to-insights/demos/data-studio.md

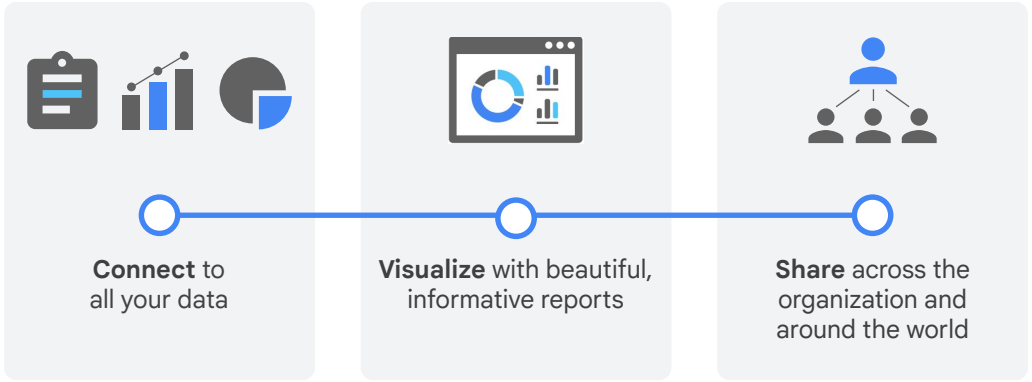
Demo: Google Data Studio walkthrough



Ecommerce Dashboard

Refer to https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/data-to-insights/demos/data-studio.md

Use Data Studio to explore and share your data insights



Data Studio's vision as a product is to simplify each of the critical steps in producing reports and dashboards. The Data Studio mantra is "Connect, visualize, share."

Creating **case statements** and calculated fields

```
CASE
    WHEN Country IN ("USA","Canada","Mexico") THEN "North America"
    WHEN Country IN ("England","France") THEN "Europe"
    ELSE "Other"
END
```

...can also be created directly in SQL functions upstream ...

Create calculated fields in Google Data Studio
https://support.google.com/datastudio/answer/7020724?hl=en

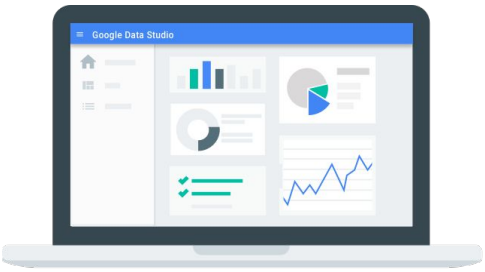
Demo:

1. Edit your data source
2. Click + to create a new calculated field
3. Enter a name for the new field
4. In the formula field, enter the CASE statement.

Avoiding visualization performance and cache pitfalls

- There are 2 parts to Data Studio cache: the responsive cache, and the predictive cache
- When all the charts in the report are being **served from the cache**, a **lightning bolt** icon appears on the dashboard
- Break both caches in Edit mode by using **Refresh Cache**
- You can (and should) **turn off responsive cache** if your data **changes frequently.**

  File → Report Settings → Enable Cache checkbox



Data Studio performance:
https://support.google.com/datastudio/answer/7020039?hl=en

Query cache

The query cache remembers the queries (requests for data) issued by the components in a report. When a person viewing the report requests the exact same query (i.e., the same dimensions, metrics, filter conditions, and date range) as a previously received query, then the data is served from the cache.

If the response can't be served from the query cache, Data Studio next looks to the prefetch cache.

Prefetch cache

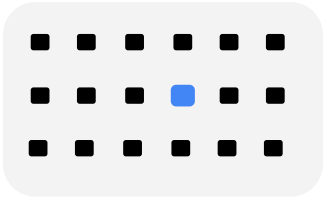
The prefetch cache (A.K.A. the "Smart cache") predicts the data that a component *could* request by analyzing the dimensions, metrics, filters, and date range properties and controls on the report. Data Studio then stores (prefetches) as much of the data as possible that could be used to answer the predicted queries. When a query can't be answered by the query cache, Data Studio tries to answer it using this prefetched data. If the query can't be answered by the prefetch cache, the data will come from the underlying data set.

Sharing and collaborating on dashboards

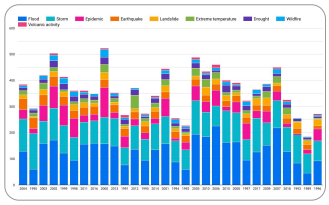
- Data Studio uses Google Drive for sharing and storing files.
- When you share a report with view permission, no login is required to view the report. A Google login is required to edit a report.
- **Sharing a report does NOT share direct access to any added data sources.**
- Data sources must be shared separately from reports.

Sharing Google Data Studio reports:
https://support.google.com/datastudio/answer/6287179?hl=en

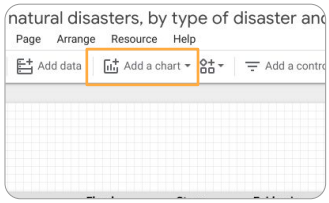
Summary: Explore and present your insights visually



Guide the eye of your user with preattentive attributes



Use the right visual to convey the right message



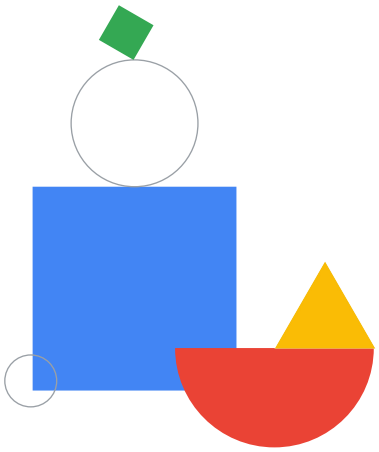
Create new report charts inside Google Data Studio

Visualizing data is both an art and a science. We've just scratched the surface of visualization theory when we discussed pre-attentive vs post-attentive processing for quick eye-to-brain understanding. Along the way we saw some terrifically bad ways to visualize data with the wrong charts and picked up some best practices along the way.

Lastly, we looked at Google Data Studio which is the visualization platform you will be exploring in more depth in your next lab.

Lab Intro

How to Build a BI Dashboard Using Google Data Studio and BigQuery



Lab objectives

01 Find the San Francisco Trees dataset

02 Explore the dataset with SQL

03 Create a reporting query that runs daily

04 Link that table into a Data Studio dashboard