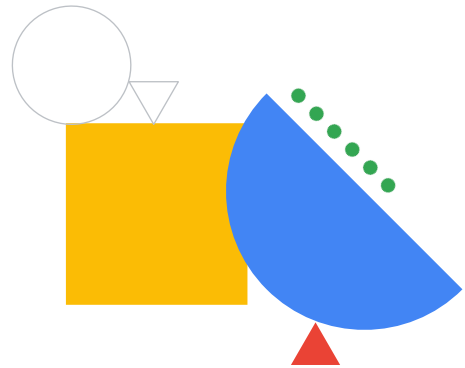
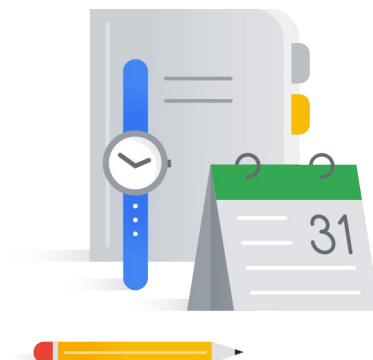


Visualizing Insights and Creating Scheduled Queries



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

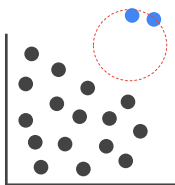


One of the key outputs data analysts create are insightful reports for your audience. In this data visualization module we will cover a little visualization theory and best practices and then we'll introduce Google Data Studio as one of the visualization tools in your toolkit for creating actionable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

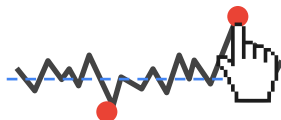


Data Visualization Principles

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

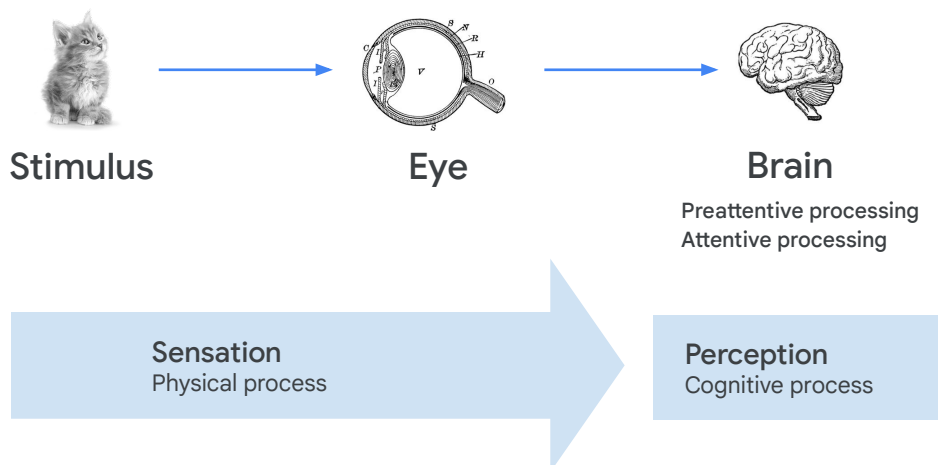
Visualization helps make sense of data.

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
587282949746**5**44878186764**5**3214
24439684634233**5**2986732190387**5**
6**5**87889374**5**39093297**5**6**5**9391732
1472**5**920189374476**5**6472217**5**6**5**2

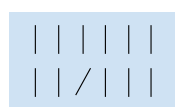
Much easier with the help of preattentive processing - features extracted quickly, effortlessly, in parallel without any attention focused on it.

Visualization Theory: Count the **fives**

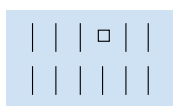
697**5**042934749373241860**5**783**5**78
587282949746**5**44878186764**5**3214
24439684634233**5**2986732190387**5**
6**5**87889374**5**39093297**5**6**5**9391732
1472**5**920189374476**5**6472217**5**6**5**2

Even easier with visual encoding.

Visualization Theory: Preattentive attributes



Orientation



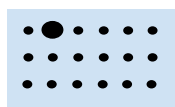
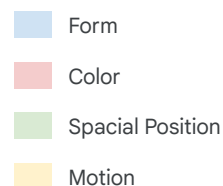
Shape



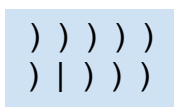
Line Length



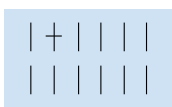
Line Width



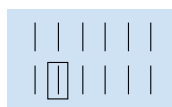
Size



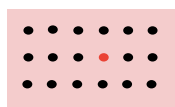
Curvature



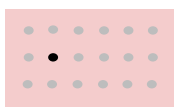
Added Marks



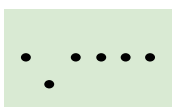
Enclosure



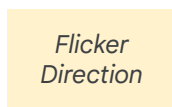
Hue



Intensity



2-D Position

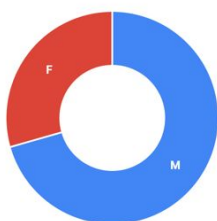


Motion

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

Game of Thrones Characters by Gender



OR

Game of Thrones Characters by Gender

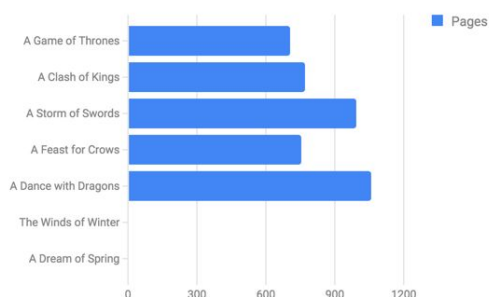


Donut Chart for quick comparison of book characters by gender is likely the better visual.

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

You choose: Effective or ineffective visuals

Game of Thrones Books by Page Count



OR

Game of Thrones Books by Page Count



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:

80% Getting **data**, analyzing it, saving it, downloading it

20% **The output**
(Visualization)

What your audience actually cares about:

All that
stuff you
did
before

99% **The output (Visualization)**

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	<ul style="list-style-type: none">• Name• Location• Part number #• Job title
2. Measures	Dependent variable Any field containing quantitative information	<ul style="list-style-type: none">• Revenue• Salary• Expenses• Count of errors

Class question

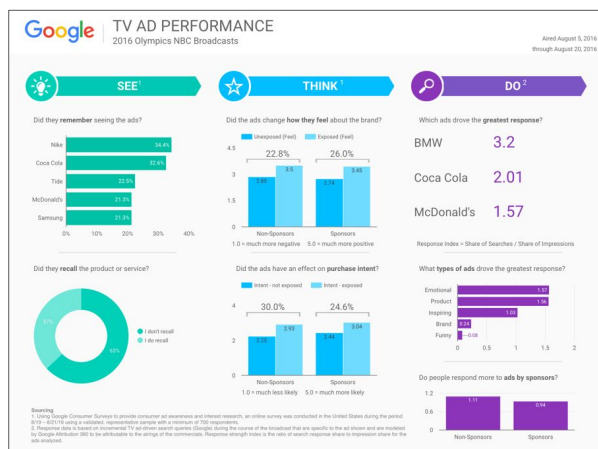
Which of the below are measures?

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

Data freshness spectrum

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

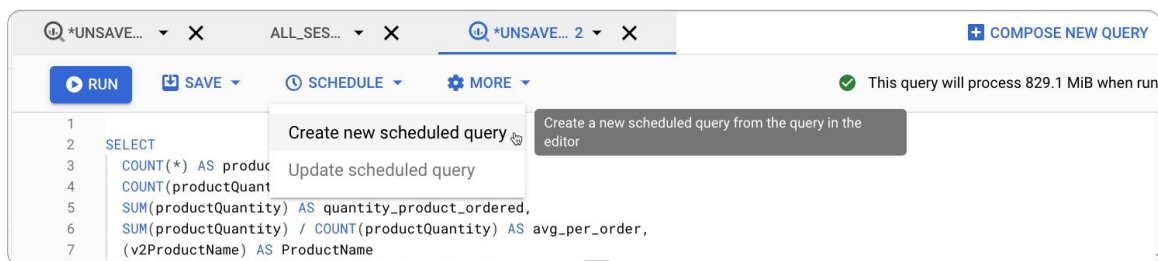
Data freshness spectrum

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

New scheduled query

Details and schedule

Name for scheduled query

Schedule options

☒ Start now ☐ Schedule start time

Repeats: Daily Start date and run time: 4/22/19, 2:53 PM PDT

⚠ This schedule will run Every day at 14:53 America/Los Angeles

Destination for query results

ⓘ A destination table is required to save scheduled query options.

Project name: data-to-insights Dataset name: ecommerce

Table name: Letters, numbers, and underscores allowed

Destination table write preference

☒ Append to table ☐ Overwrite table

Notification options

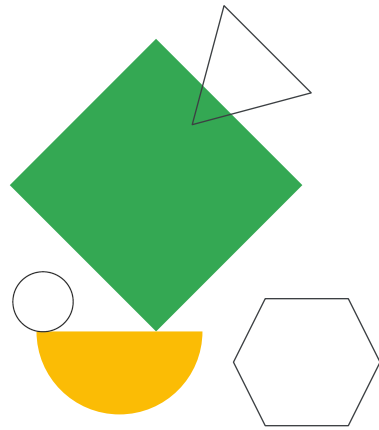
☐ Send email notifications ⓘ

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

Demo

Creating a Scheduled Query

Automatically updating a data table at
a set interval



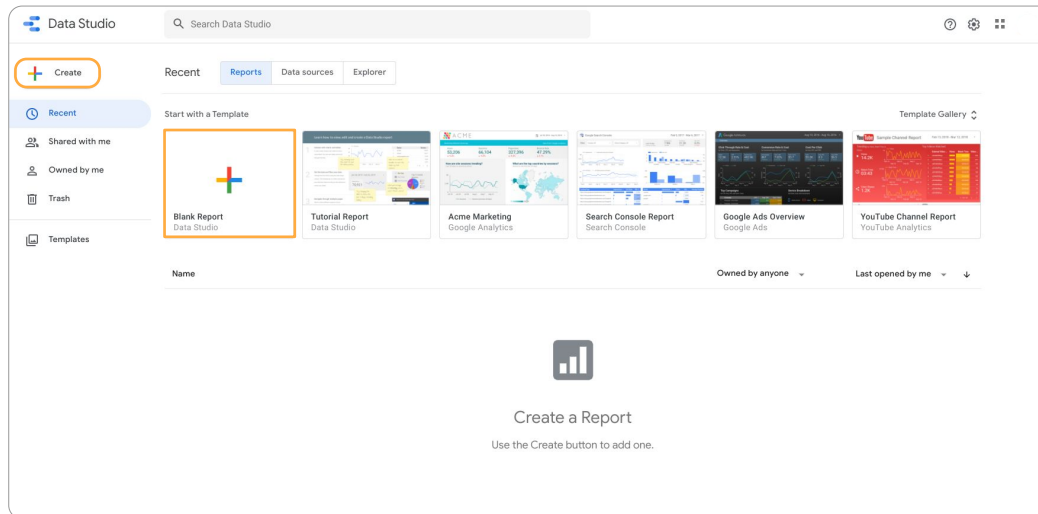
Refer to

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



Google Data Studio

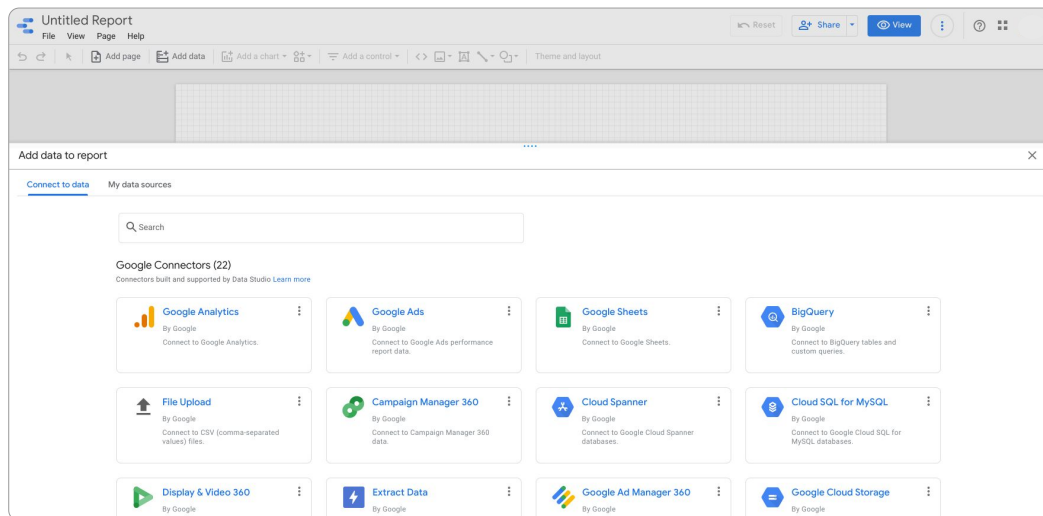
Create new reports in the Data Studio UI



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.

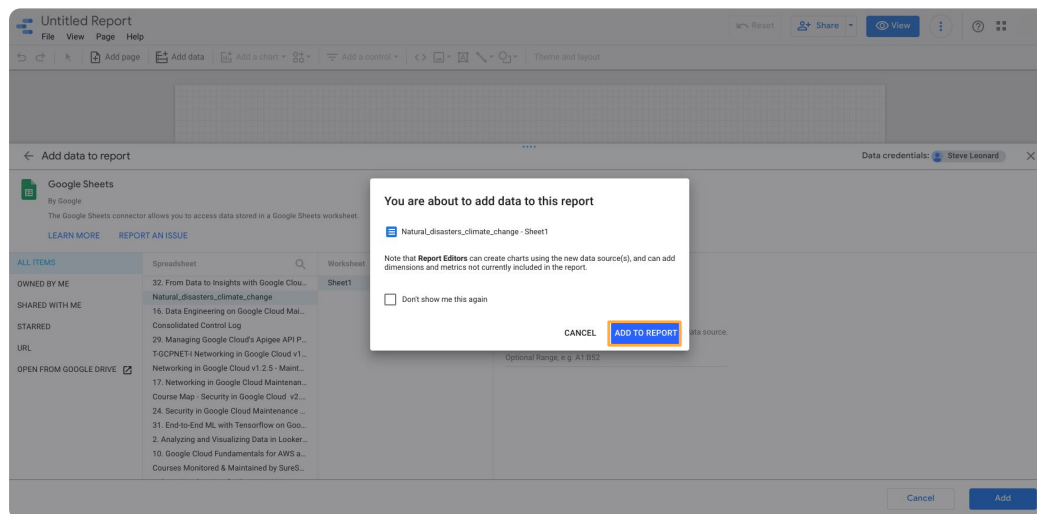
Connect to multiple different types of data sources



Google Cloud

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

The screenshot displays the Google Cloud Data Studio interface. At the top, there's a menu bar with options like File, Edit, View, Insert, Page, Arrange, Resource, and Help. Below the menu is a toolbar with icons for adding pages, data, charts, and controls. The main area shows a table of data with columns: Year, Earthquake, Epidemic, Storm, Wildfire, Volcanic, Insect inf., Extreme t., Landslide, Mass mo., Flood, and Drought. The right-hand panel is titled 'Chart > Table' and has two tabs: 'DATA' and 'STYLE'. Under the 'DATA' tab, there's a 'Data source' section with 'Natural_disasters...' selected. Below that is a 'Dimensions' section with a list of fields: Year, Earthquake, Epidemic, Storm, Wildfire, Volcanic activity, Insect infestation, Extreme temperature, Landslide, Mass movement (dry), Storm, Volcanic activity, Wildfire, Year, and Record Count. The 'Metrics' section is currently empty. Three numbered callouts are present: 01 points to the 'Available Fields' list in the 'Dimensions' section, 02 points to the 'Dimensions' section header, and 03 points to the 'Metrics' section header.

	Year	Earthquake	Epidemic	Storm	Wildfire	Volcanic ...	Insect inf...	Extreme t...	Landslide	Mass mo...	Flood	Drought
1.	2018	20	15	94	10	7	1	26	13	1	127	14
2.	2017	22	27	130	15	2	1	10	25	1	126	9
3.	2016	30	25	86	10	null	1	12	13	null	159	14
4.	2015	23	16	121	12	6	1	12	20	1	162	28
5.	2014	26	21	99	4	6	1	17	15	null	135	18
6.	2013	29	23	105	10	3	1	14	11	1	149	9
7.	2012	27	25	91	6	1	1	51	13	1	136	21
8.	2011	30	27	84	8	6	1	16	17	null	156	17

01 Available Fields

02 Dimensions

03 Metrics

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

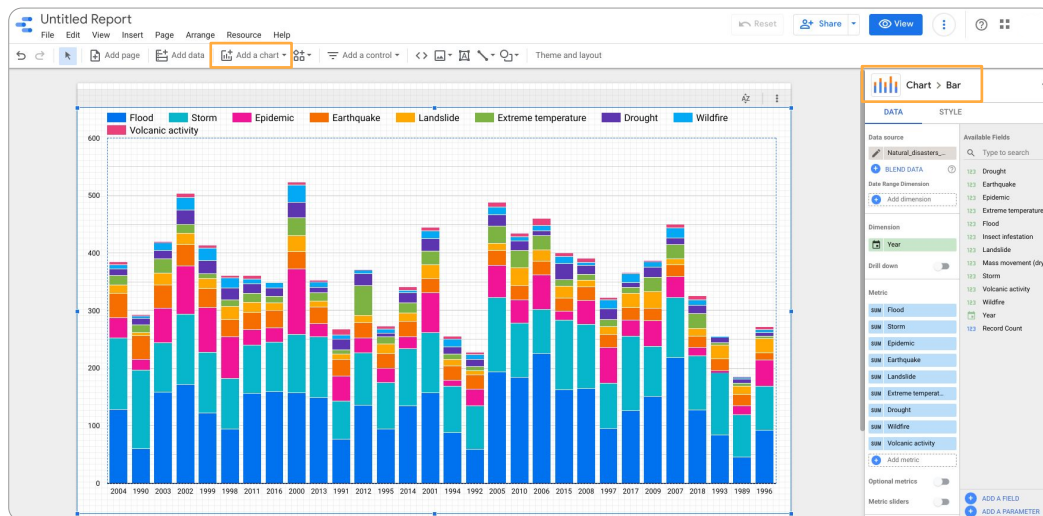
The screenshot shows the Google Cloud Data Studio interface. At the top, there's a menu bar with options like File, Edit, View, Insert, Page, Arrange, Resource, and Help. Below the menu bar, there's a toolbar with icons for adding data, charts, and controls. The main area displays a report titled 'Untitled Report' with a table of natural disaster data. The table has columns for Year, Earthquake, Epidemic, Storm, Wildfire, Volcanic, Insect inf..., Extreme L..., Landslide, Mass mo..., Flood, and Drought. The data is organized into rows, with the first row showing data for 2018. An orange arrow points from the 'Edit data source picker' icon in the top right of the report to the 'Edit data source picker' icon in the 'Data source' dropdown menu on the right sidebar.

	Year	Earthquake	Epidemic	Storm	Wildfire	Volcanic	Insect inf...	Extreme L...	Landslide	Mass mo...	Flood	Drought
1.	2018	20	15	94	10	7	0	26	13	1	127	14
2.	2017	22	27	130	15	2	0	10	25	1	126	9
3.	2016	30	25	86	10	null	0	12	13	null	159	14
4.	2015	23	16	121	12	6	0	12	20	1	162	28
5.	2014	26	21	99	4	6	0	17	15	null	135	18
6.	2013	29	23	105	10	3	0	14	11	1	149	9
7.	2012	27	25	91	6	1	0	51	13	1	136	21

Below the table, there's a section for 'Natural_disasters_climate_change - She...' with data credentials 'Steve Leonard', data freshness '15 minutes', and community visualizations access 'On'. There's a 'DONE' button. Below this, there's a section for 'EDIT CONNECTION | FILTER BY EMAIL' with a table of fields. The table has columns for Field, Type, Default Aggregation, and Description. The fields listed are Insect infestation, Landslide, Mass movement (dry), Storm, Volcanic activity, Wildfire, and Year. The 'Year' field is highlighted. Below the table, there's a section for 'METRICS (1)' with a table of metrics. The metrics listed are Record Count. The 'Record Count' metric is highlighted. At the bottom, there's a 'REFRESH FIELDS' button and a '13 / 13 Fields' indicator.

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.

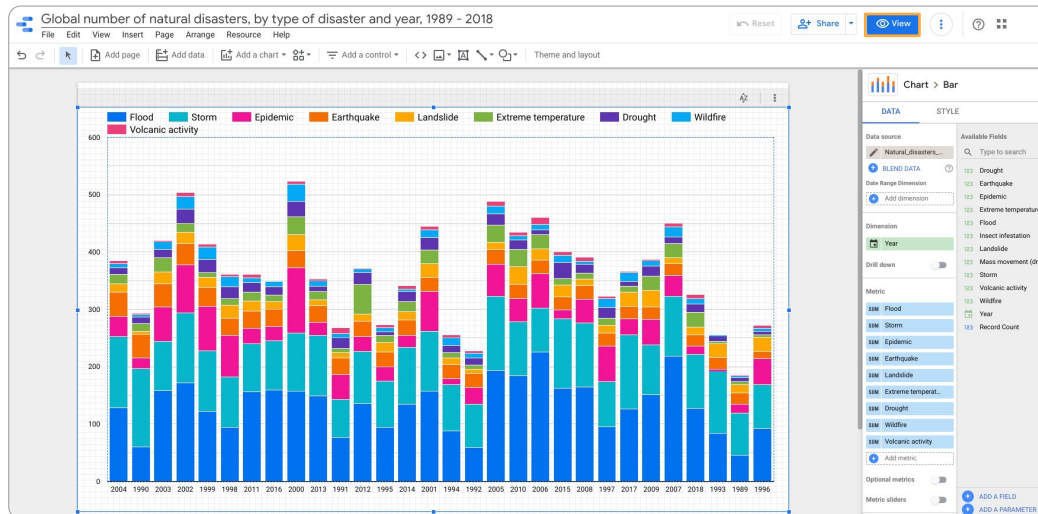
In the same way that you defined **Dimensions** and **Metrics** earlier, do the same for your chart by adding selections from the **Available Fields** list.

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.



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

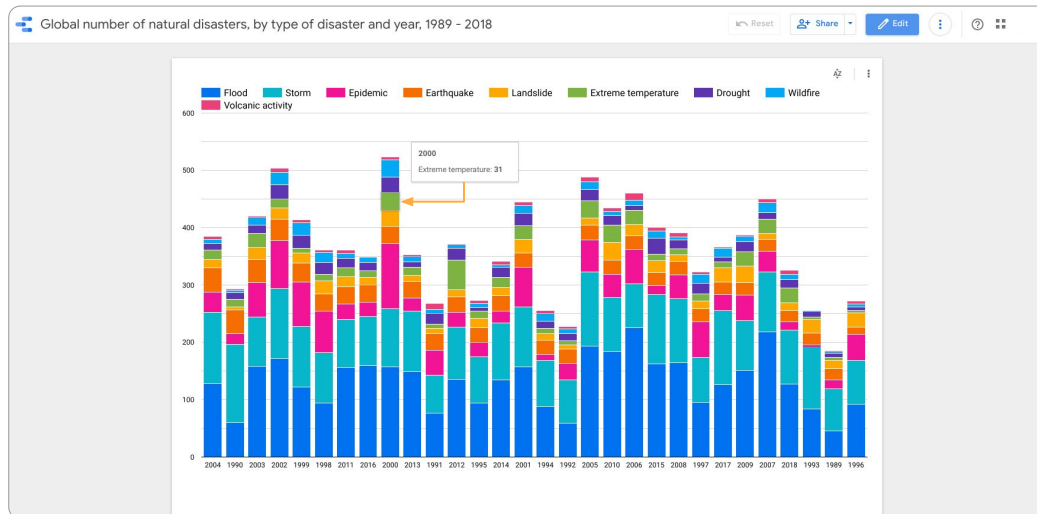
View the end-user version of the report



Google Cloud

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

View your report as an end-user



Google Cloud

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.

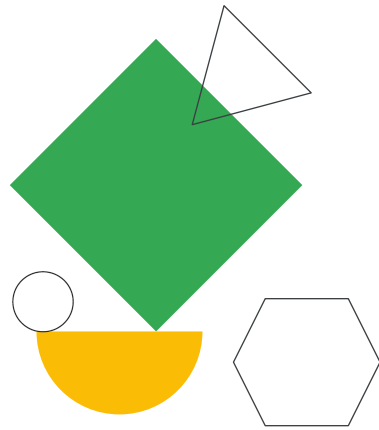
When a viewer mouses over the chart, they are able to view live data. In this example, the viewer is able to see that in the year 2000, there were 31 natural disasters attributable to extreme temperature.

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



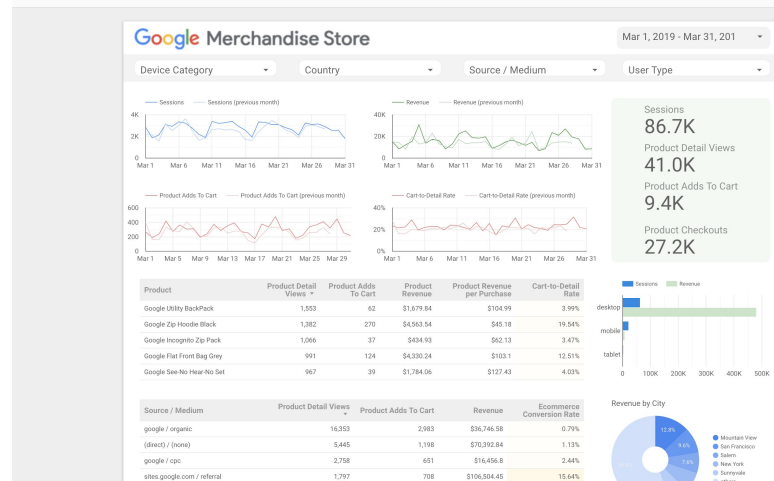
Google Cloud

Refer to

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

Demo: Google Data Studio walkthrough

[Sample] Google Merchandise Store Ecommerce Report



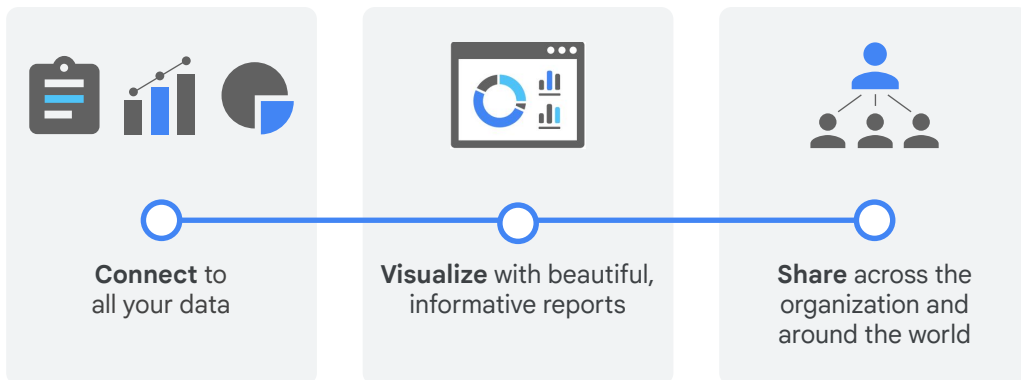
Ecommerce Dashboard

Google Cloud

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

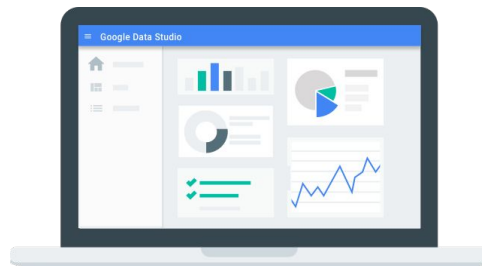
<https://support.google.com/datastudio/answer/7020724?hl=en>

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.

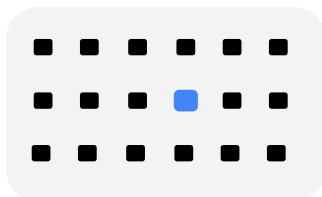


Google Cloud

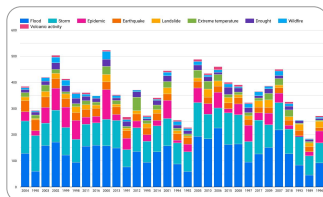
Sharing Google Data Studio reports:

<https://support.google.com/datastudio/answer/6287179?hl=en>

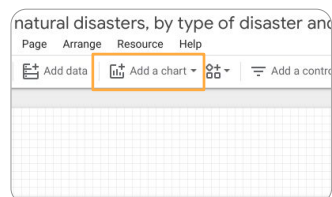
Summary: Explore and present your insights visually



Guide the eye of your user with preattentive attributes



Use the right visual to convey the right message



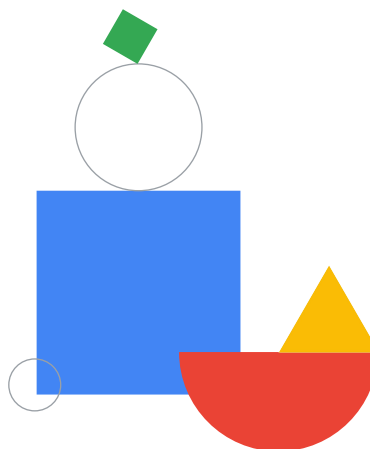
Create new report charts inside Google Data Studio

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



