Overview

Data visualization tools can help you make sense of your BigQuery data and help you analyze the data interactively. You can use visualization tools to help you identify trends, respond to them, and make predictions using your data. In this lab, you use Google Data Studio to visualize data in the BigQuery table populated by your Dataflow pipeline in the previous exercise.

At the time of this writing, streaming pipelines are not available in the DataFlow Python SDK. So the streaming labs are written in Java.
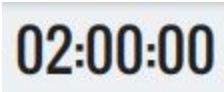
Objectives

In this lab, you will perform the following tasks:

- Connect to a BigQuery data source
- Create reports and charts to visualize BigQuery data

This lab uses Google Data Studio to visualize data in BigQuery using the BigQuery connector. In subsequent tasks, you will create a data source, a report, and charts that visualize data in the sample table.

Setup

For each lab, you get a new GCP project and set of resources for a fixed time at no cost.

1. Make sure you signed into Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example, 02:00:00 and make sure you can finish in that time block.

There is no pause feature. You can restart if needed, but you have to start at the beginning.

3. When ready, click START LAB .

4. Note your lab credentials. You will use them to sign in to Cloud Platform Console.

Open Google Console

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. **Learn more.**

Username

google2876526_student@qwiklabs.n

Password

TG959yrKDX

GCP Project ID

qwiklabs-gcp-0855e773352d3560

New to labs? View our introductory video!

5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

If you use other credentials, you'll get errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

Task 1: Preparation

You will be running a sensor simulator from the training VM. There are several files and some setup of the environment required.

Open the SSH terminal and connect to the training VM

1. In the Console, on the **Navigation menu** ( ), click **Compute Engine** > **VM instances**.
2. Locate the line with the instance called **training-vm**.
3. On the right, under 'Connect' column, click on **SSH** to open a terminal window.
4. In this lab, you will enter CLI commands on the **training-vm**.

Verify initialization is complete

5. The **training-vm** is installing some software in the background. Verify that setup is complete by checking the contents of the new directory.

```
ls /training
```

The setup is complete when the result of your list (ls) command output appears as in the image below. If the full listing does not appear, wait a few minutes and

try again. **Note**: It may take 2 to 3 minutes for all background actions to complete.

```
student-04-2324a1e71896@training-vm:~$ ls /training
bq_magic.sh  project_env.sh  sensor_magic.sh
student-04-2324a1e71896@training-vm:~$
```

Download Code Repository

6. Next you will download a code repository for use in this lab.

```
git clone https://github.com/GoogleCloudPlatform/training-data-analyst
```

Set environment variables

7. On the **training-vm** SSH terminal enter the following:

```
source /training/project_env.sh
```

This script sets the $DEVSHELL_PROJECT_ID and $BUCKET environment variables.

Task 2: Creating a data source in Data Studio

1. Google Data Studio is a separate service from the Google Cloud environment. Open a new browser tab preferably in an incognito window. Navigate to: datastudio.google.com or click on this link: Google Data Studio

The first step in creating a report in Data Studio is to create a data source for the report. A report may contain one or more data sources. When you create a BigQuery data source, Data Studio uses the BigQuery connector.
You must have the appropriate permissions in order to add a BigQuery data source to a Data Studio report. Lab initialization steps created a BigQuery Dataset and tables for use in your exercises. The permissions applied to BigQuery datasets will apply to the reports, charts, and dashboards you create in Data Studio. When a Data Studio report is shared, the report components are visible only to users who have appropriate permissions.

2. On the **Reports** page, in the **Start with a Template** section, click the **Blank Report** template. This starts the account setup process.
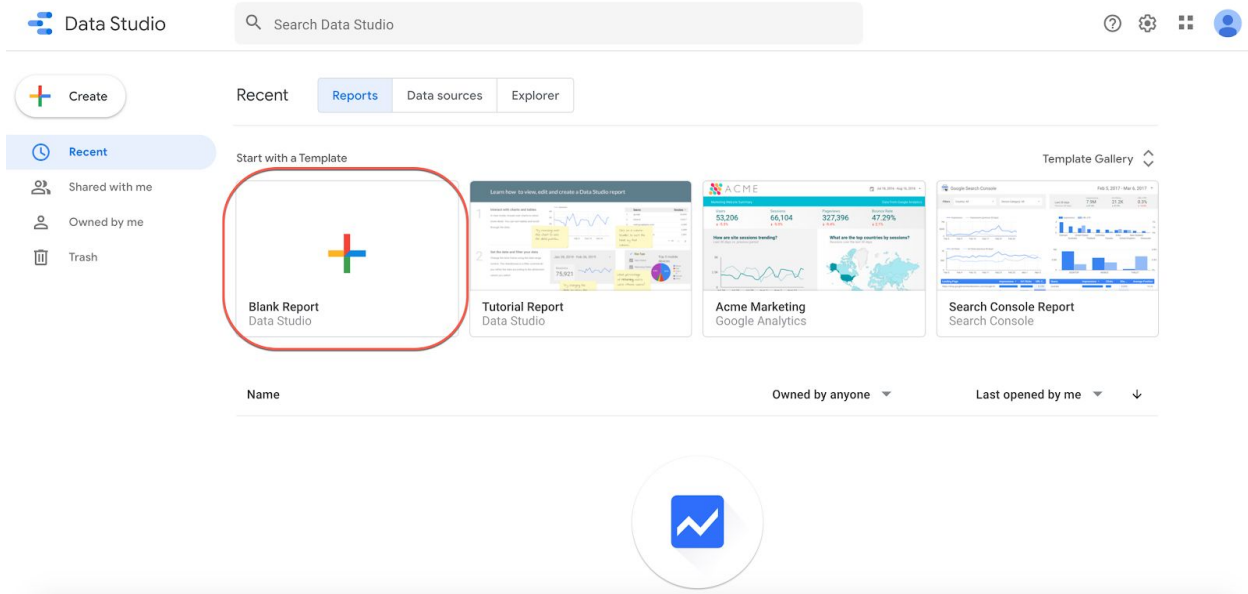
3. On the **Welcome** page, click on **GET STARTED**.

4. On the **Terms** page, click on the **checkbox** to acknowledge the terms. And click **ACCEPT**.

5. On the **Preferences** page, select **No, thanks** for each option to receive email notifications, and click **DONE**.

6. Now that the account is initialized, you need to start the process again.

7. On the **Reports** page, in the **Start with a Template** section, click the **Blank Report** template. This time it will take you to a new page and begin an **Untitled Report**.

8. The **Add data to report** panel will load from the bottom of the page.

9. In the **Google Connectors** section, select **BigQuery**.
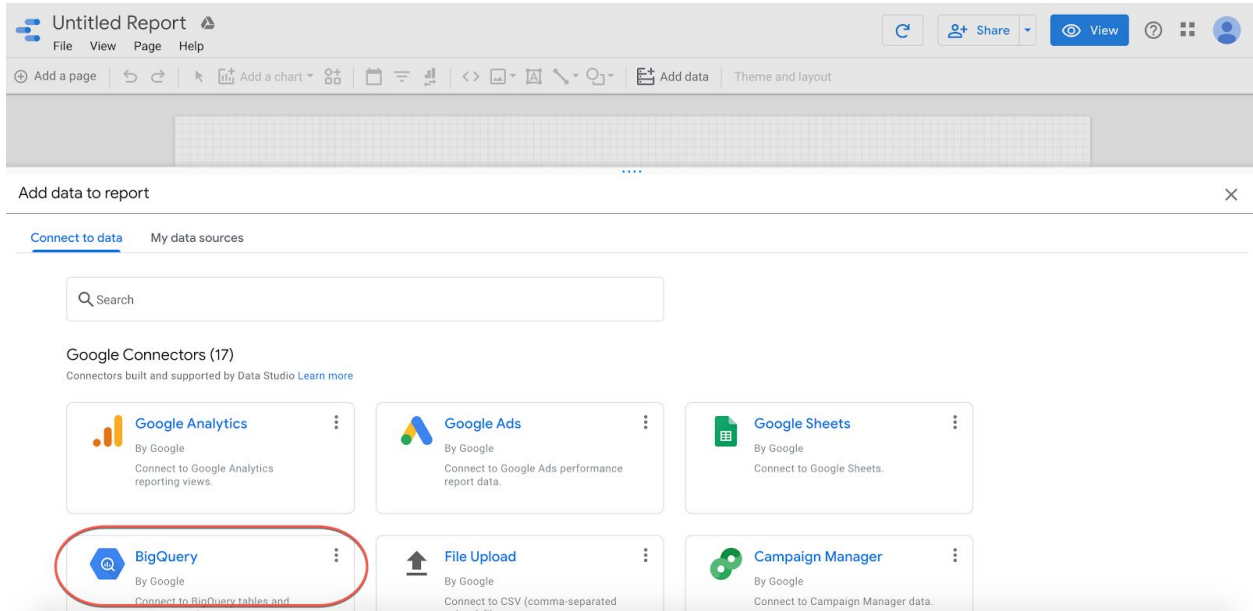
10. Click on **AUTHORIZE** for the notice that "Data Studio requires authorization to connect to your BigQuery projects."

11. If prompted, in the **Sign in** dialog, select your Qwiklabs student account.

12. If prompted, click **ALLOW** to give Google Data Studio permission to view the BigQuery resources in your lab account.

13. Then select **My Projects**.

14. In the **Project** column, click on your project name.

15. In the **Dataset** column, click on **demos**.

16. In the **Table** column, click **current_conditions**.

17. Lastly click **Add**.
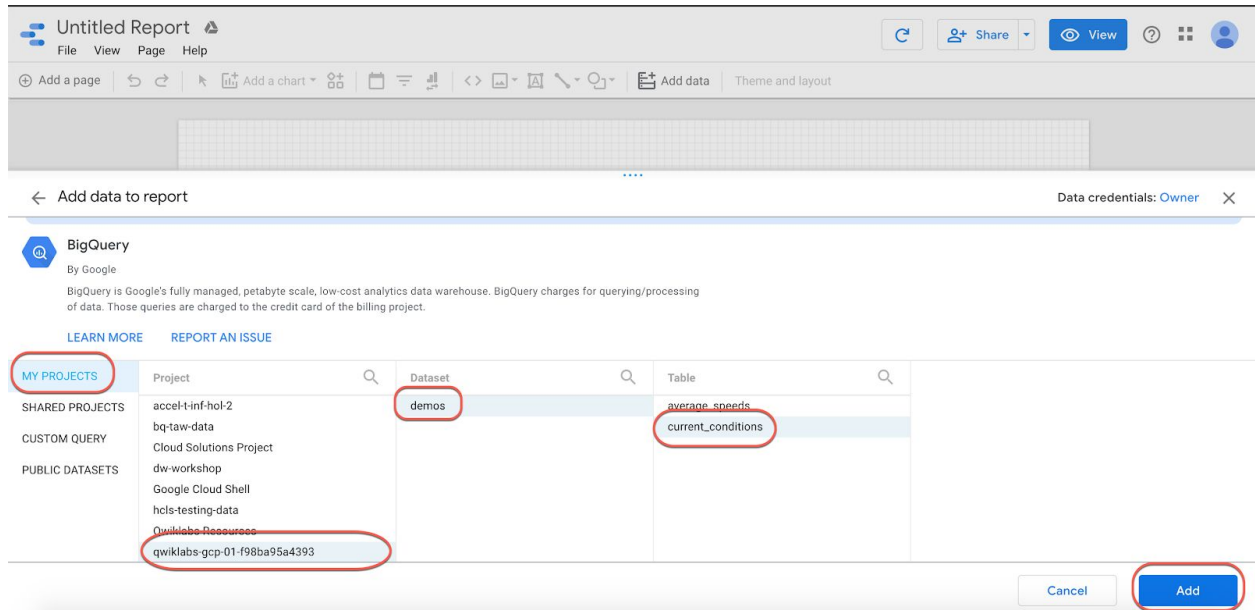
18.   You will be prompted with a "You are about to add data to this report"
notice. Check "Don't show me this again" and click **ADD TO REPORT**.
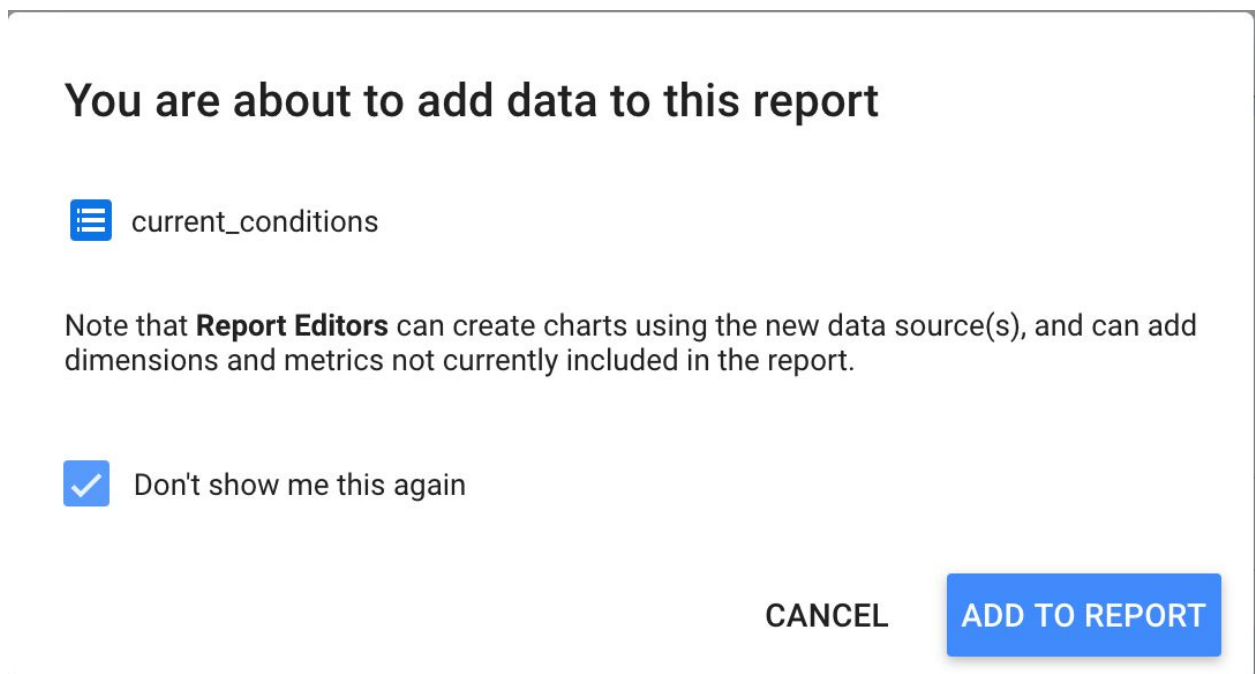
19.   Once complete, a simple default tabular report will appear. This confirms that you can see your BigQuery data in Data Studio.

	highway	Record Count ▾
1.	5	10,621
2.	8	4,601
3.	805	4,214
4.	163	817
5.	15	258

Giving Data Studio permission to Google Cloud account resources is typically a first-time activity and not something you would need to do every time you create a report.

Click *Check my progress* to verify the objective.

Creating a data source

Task 3: Creating a bar chart using a calculated field

Once you have added the **current_conditions** data source to the report, the next step is to create a visualization. Begin by creating a bar chart. The bar chart will display the total number of vehicles captured for each highway.

1. (Optional) At the top of the page, click **Untitled Report** to change the report name. For example, type **<PROJECTID>-report1-yourname**.
2. Delete the pre-populated tabluar report. You can do this by simply selecting it and pressing delete.
3. Next, from the **Add a chart** menu select the first **Bar chart**.
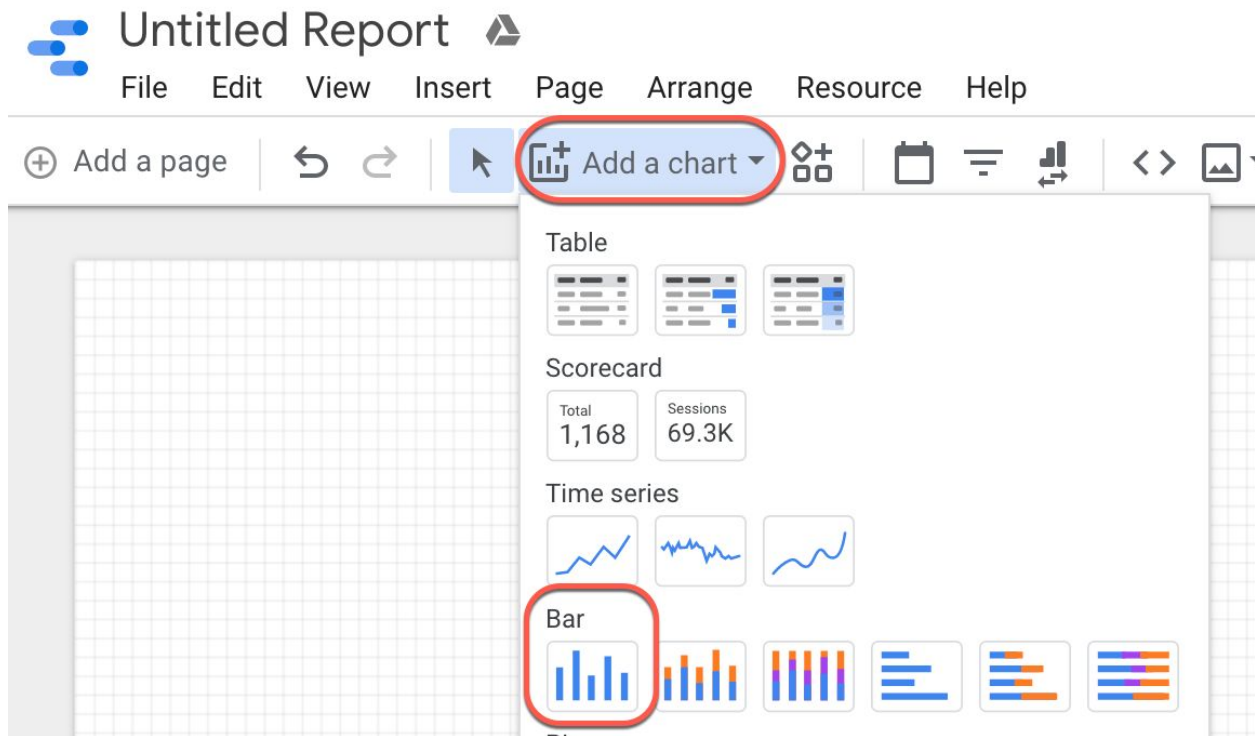
4. In the **Bar chart** properties window, on the **Data** tab, notice the value for **Data Source** (**current_conditions**) and the default values for **Dimension** and **Metric**.

5. If **Dimension** is not set to **highway**, then change **Dimension** to **highway**. In the **Dimension** section, click the existing dimension.

6. In the **Dimension picker**, select **highway**.

7. In the **Metric** section, click **Add metric** and add **latitude**.

8. In the **Metric** section, mouse over **Record Count** and click the **(x)** to remove it.

Example:

Bar chart Properties

DATA   STYLE

Data Source   >

current_conditions ✏️

Date Range Dimension   >

timestamp

Dimension   >

highway

+ Add a Dimension

Metric   >

latitude

+ Add a Metric

9. To gain insight on vehicle volume you need to add a metric for each vehicle detected.

10. In the **Metric** section, click **Add metric** and add **sensorId**.

11. A count distinct for this column is automatically created. This metric set as a count distinct does not give you a true sense of traffic volume. Click on the **CTD** text and on the popup window choose **Count** . Type the name **vehicles** in the name box. Click in the report space off the popup to close it. The change is saved automatically.

Metric

CTD   sensorId

SUM   latitude

\+   Add metric

Optional metrics

CT vehicles

Name

vehicles

Source field: **sensorId**

Aggregation

⦿ Count

◯ Count Distinct

Type

Number ▼

Comparison calculation

None ▼

Running calculation

None ▼ ⑦

12. In the **Metric** section, mouse over **latitude** and click the **(x)** to remove it.

13. The Dimension should be set to highway and the Metric should be set to sensorId. Notice the chart is sorted in Descending order by default. The highway with the most vehicles is displayed first.
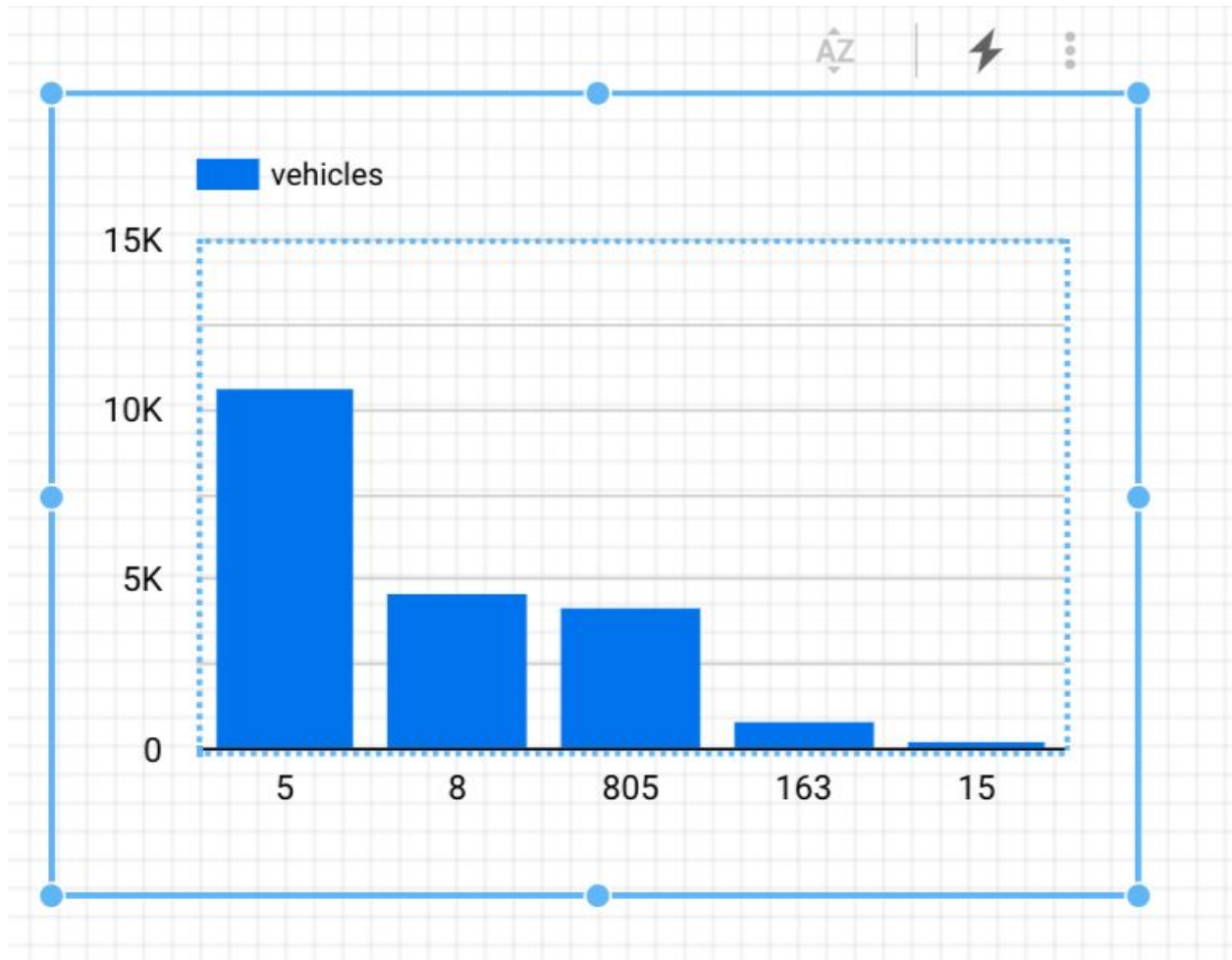
14. To enhance the chart, change the bar labels. In the **Bar chart** properties window, click the **STYLE** tab.

15. In the **Bar chart** section, check **Show data labels**.

16. The total number of vehicles is displayed above each bar in the chart.

Task 4: Creating a chart using a custom query

You may find that it is easier to work with an existing query to produce the desired reports and visualizations in Data Studio. The Custom Query option lets you leverage BigQuery's full query capabilities such as joins, unions, and analytical functions.

Alternatively, you can leverage BigQuery's full query capabilities by creating a view. A view is a virtual table defined by a SQL query. You can query data in a view by adding the dataset containing the view as a data source.

When you specify a SQL query as your BigQuery data source, the results of the query are in table format, which becomes the field definition (schema) for your data source. When you use a custom query as a data source, Data Studio uses your SQL as an inner select statement for each generated query to BigQuery. For more information on custom queries in Data Studio, consult the online help.

1. To add a bar chart to your report that uses a custom query data source:
2. From the **Add a chart** menu select the first **Bar chart**.
3. In the **Bar chart** properties window, on the **Data** tab, notice the value for Data Source (current_conditions) and the default values for Dimension and Metric are the same as the previous chart. In the **Data Source** section, click on the **current_conditions** data source. At the bottom of the pane choose **ADD DATA**.
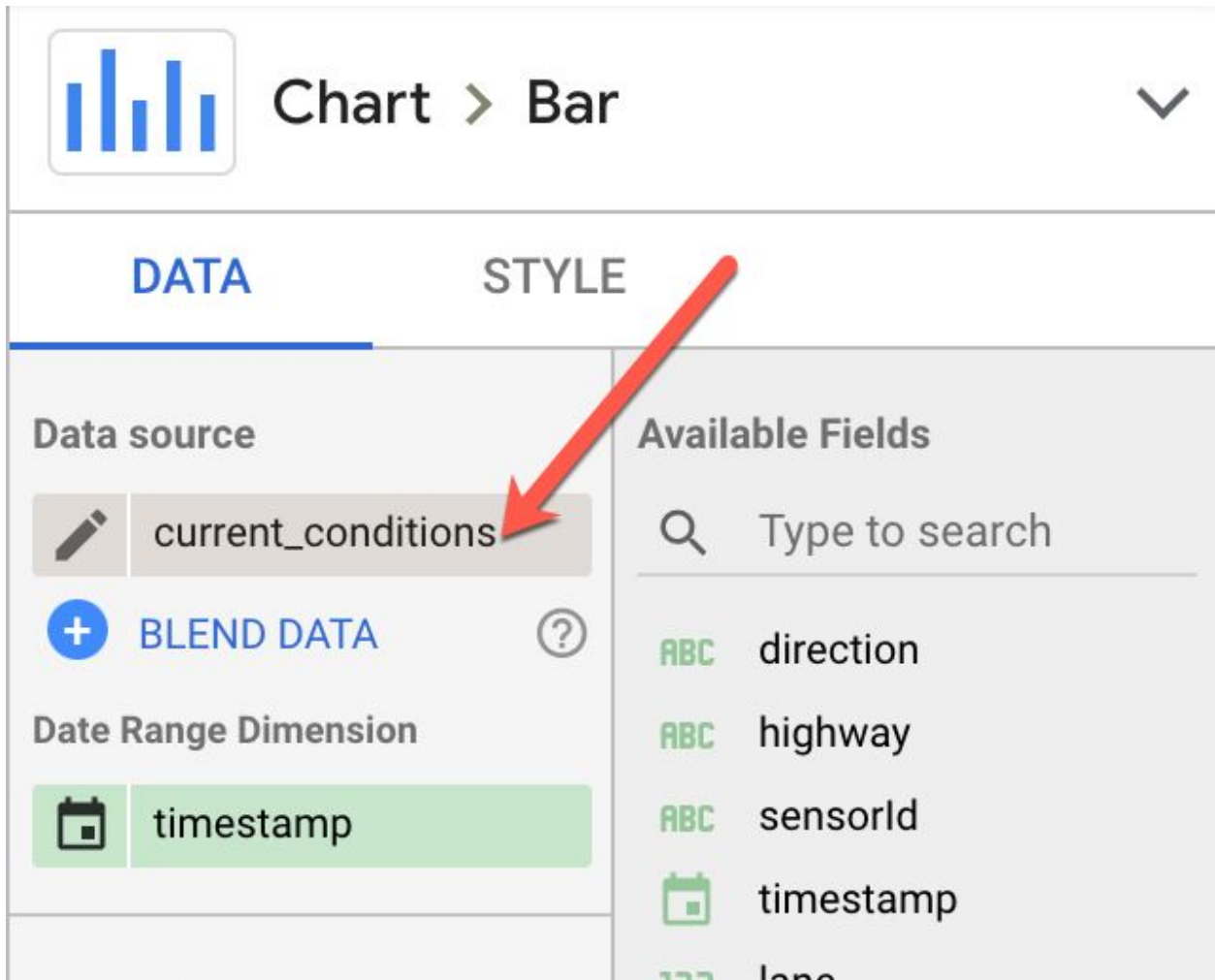
4. Under **Google Connectors**, select **BigQuery**.

5. Select **CUSTOM QUERY** in the first grouping.

6. For **Billing Project**, select your project.

7. Type the following in the **Enter custom query** window and replace the `<PROJECTID>` with your Project ID.

```
SELECT max(speed) as maxspeed, min(speed) as minspeed,
avg(speed) as avgspeed, highway
FROM `<PROJECTID>.demos.current_conditions`
group by highway
```

This query uses max/min/avg functions to give you the corresponding speed for each highway.

8. Click **ADD**.
9. When prompted, click **ADD TO REPORT**.

Data Studio may be unable to determine the appropriate Dimension and Metrics for the chart. This requires you to adjust the graph options.
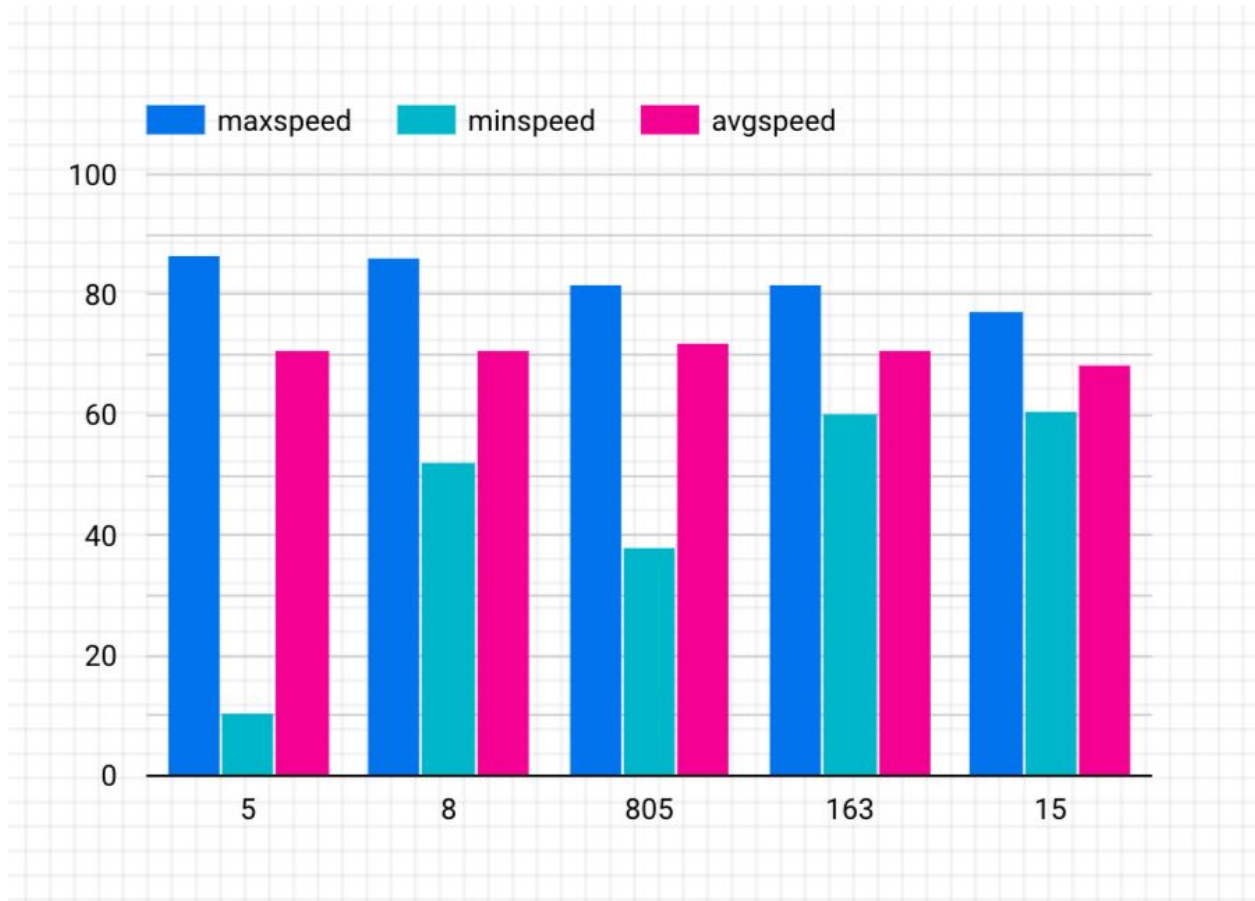
10. In the **Bar chart** properties, on the **Data** tab, in the **Metric** section, click **Record count**.
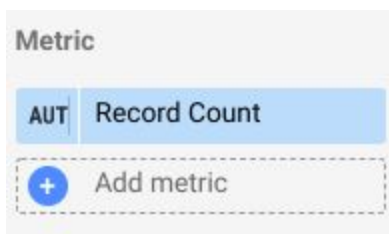
11. In the **Metric picker,** select **maxspeed**.
12. In the **Metric** section, click **Add metric**.
13. In the **Metric picker**, select **minspeed**.
14. In the **Metric** section, click **Add metric**.
15. In the **Metric picker**, select **avgspeed**.
16. Remove the metric other than maxspeed, minspeed and avgspeed, if exist.

Your chart now displays the maximum speed, minimum speed, and average speed for each highway.

Notice each bar has a default color based on the order the metrics were added to the chart.

17. For readability, change the chart styles. In the **Bar chart** properties, click the **Style** tab.

Chart > Bar

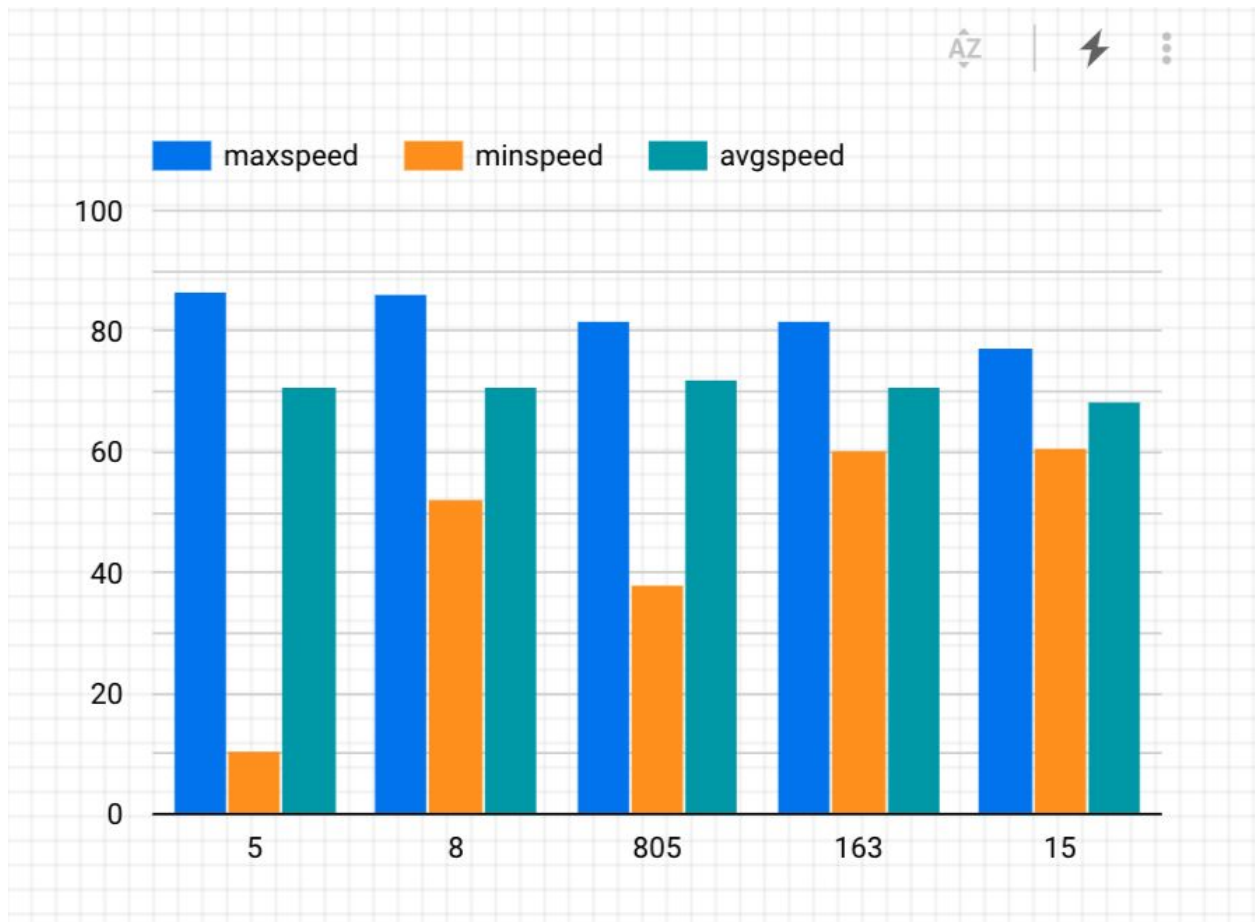DATA   **STYLE**

Bar chart

Bars

10

☐ Stacked Bars

☐ Show data labels

Color by

18. In the **Color By** section, click on the boxes to select different colors.

Click *Check my progress* to verify the objective.

Create a chart using a custom query

Check my progress

Task 5: Viewing your query history

You can view queries submitted via the BigQuery Connector by examining your query history in the BigQuery web interface. Using the query history, you can estimate query costs, and you can save queries for use in other scenarios.

Open BigQuery Console

In the Google Cloud Console, select **Navigation menu** > **BigQuery**:
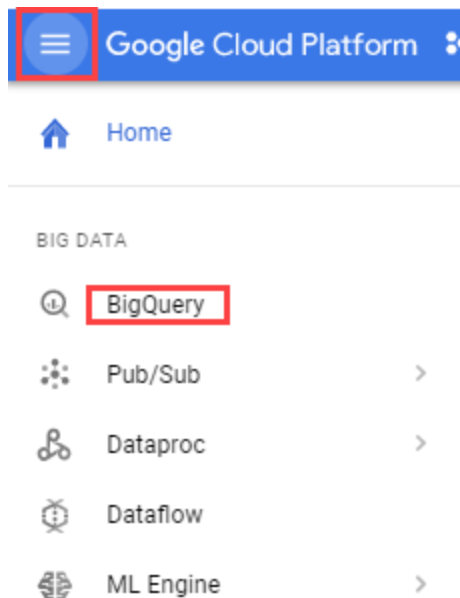
The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and lists UI updates.

Click **Done**.

1. In the left pane the first item in the list will be **Query history**. On your initial visit to the page the query history should appear on the bottom right underneath the Query editor pane. If it is not loaded click the **Query history** link.

2. The list of queries is displayed with the most recent queries first. Click on any Query to view details on the query such as Job ID and Bytes Processed.

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied

- 2 stars = Dissatisfied

- 3 stars = Neutral

- 4 stars = Satisfied

- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.