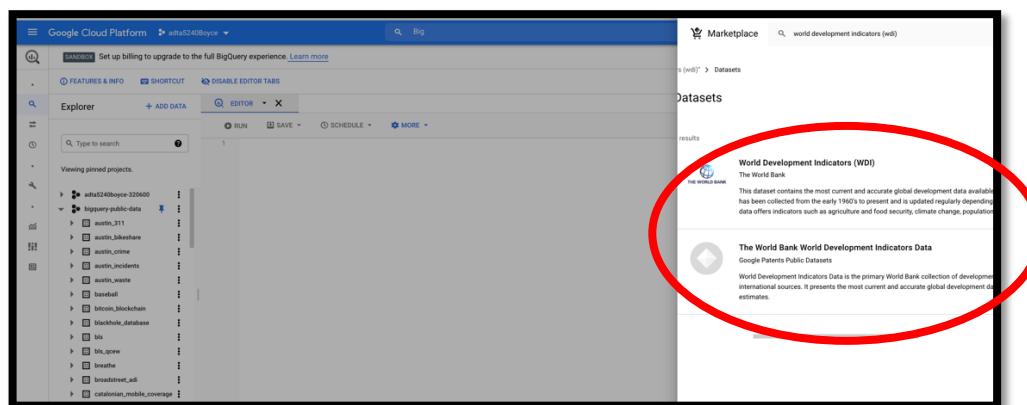
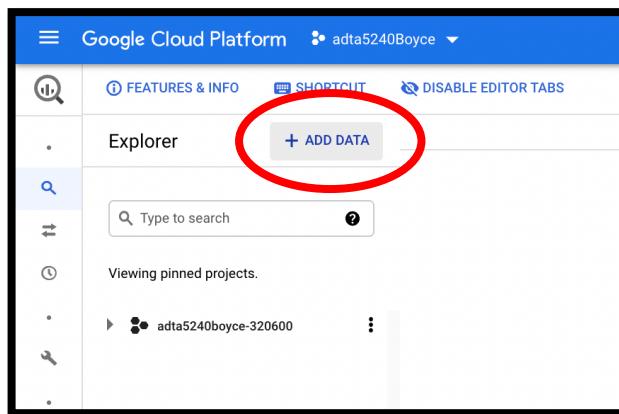


Joining data from two different publicly available datasets

We will use 2 publicly datasets for this exercise. These databases are available through Google console.

- Sustainable Development Goals (UN SDG) Indicators (UN Statistics Division). For more information on the variables please reference: <https://unstats.un.org/sdgs/metadata/>
- World Development Indicators (WDI). These databases are available through Google console. For more information on the dataset please reference: <https://datatopics.worldbank.org/world-development-indicators/>

Step 1: View the public datasets through Marketplace in GCP. For this exercise, we will only choose one variable from each dataset. Open BigQuery and Click on “+ADD DATA”. You will then see a drop-down menu, choose “Explore public datasets.” This will take you to the “Marketplace.” Type in the names of the two public datasets: Sustainable Development Goals Indicators and World Development Indicators (one at a time). This allows you to view the datasets.



Step 2: Decide what data you want to ingest from each dataset and create a table with that data.

- From UN SDG we will use the Annual Growth Rate of Real GDP per capita (%)
- From WDI, we will use Population
- We will only use the data from 2016

You will create a table from the SDG using the following SQL commands:

- For UN_SDG:

```
SELECT geoareaname, timeperiod, value FROM `bigquery-public-data.un_sdg.indicators` as UN_SDG WHERE seriesdescription = 'Annual growth rate of real GDP per capita (%)' AND timeperiod = '2016'
```

→ Hit Run

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there's a sidebar with sections like Analysis, Migration, Administration, and BI Engine. The main area shows a query editor with a red circle highlighting the 'RUN' button. Below the editor, the 'Query results' section displays the output of the query. The results table has columns: Row, geoareaname, timeperiod, and value. The data shows three rows: World (2016, 1.29588), Sint Maarten (Dutch part) (2016, -1.56664), and Haiti (2016, 0.16554). At the bottom, there are tabs for Job history, Query history, and Saved queries.

Row	geoareaname	timeperiod	value
1	World	2016	1.29588
2	Sint Maarten (Dutch part)	2016	-1.56664
3	Haiti	2016	0.16554

- Click on “Compose New Query” to add the data from WDI
- For WDI:

```
SELECT country_name, year, value FROM `bigquery-public-data.world_bank_wdi.indicators_data` as WB_WDI WHERE indicator_name = 'Population, total' AND year = 2016
```

→ Hit Run

The screenshot shows the Google Cloud Platform BigQuery interface. A red circle highlights the 'RUN' button in the top navigation bar. Another red circle highlights the results table below, which displays data for various countries in 2016. The results table has columns: Row, country.name, year, and value. The data includes:

Row	country.name	year	value
1	Caribbean small states	2016	7269385.0
2	Togo	2016	7509952.0
3	Bosnia and Herzegovina	2016	3386263.0
4	Rwanda	2016	1.1668829E7
5	Trinidad and Tobago	2016	1377563.0
6	Malta	2016	455356.0
7	Barbados	2016	285798.0

Step 3: We will now join the two datasets together based on the previous selection using SQL using the following code:

- Click on “Compose New Query” to add the data from WDI

```
SELECT UN_SDG.geoareaname, UN_SDG.timeperiod, UN_SDG.value as GDP_per_Capita_growth, WB_WDI.country_name, WB_WDI.year, WB_WDI.value as WB_Population FROM `bigquery-public-data.un_sdg.indicators` as UN_SDG JOIN `bigquery-public-data.world_bank_wdi.indicators_data` as WB_WDI on WB_WDI.country_name = UN_SDG.geoareaname WHERE UN_SDG.seriesdescription = 'Annual growth rate of real GDP per capita (%)' AND UN_SDG.timeperiod = '2016' AND WB_WDI.indicator_name = 'Population, total' AND WB_WDI.year = 2016
```

```

SELECT UN_SDG.geoareaname, UN_SDG.timeperiod, UN_SDG.value as GDP_per_Capita_growth, WB_WDI.country_name, WB_WDI.year, WB_WDI.value as WB_Population
FROM `bigquery-public-data.un_sdg.indicators` as UN_SDG
JOIN `bigquery-public-data.world_bank_wdi.indicators_data` as WB_WDI on WB_WDI.country_name = UN_SDG.geoareaname
WHERE UN_SDG.seriesdescription = 'Annual growth rate of real GDP per capita (%)'
AND UN_SDG.timeperiod = '2016'
AND WB_WDI.indicator_name = 'Population, total'
AND WB_WDI.year = 2016

```

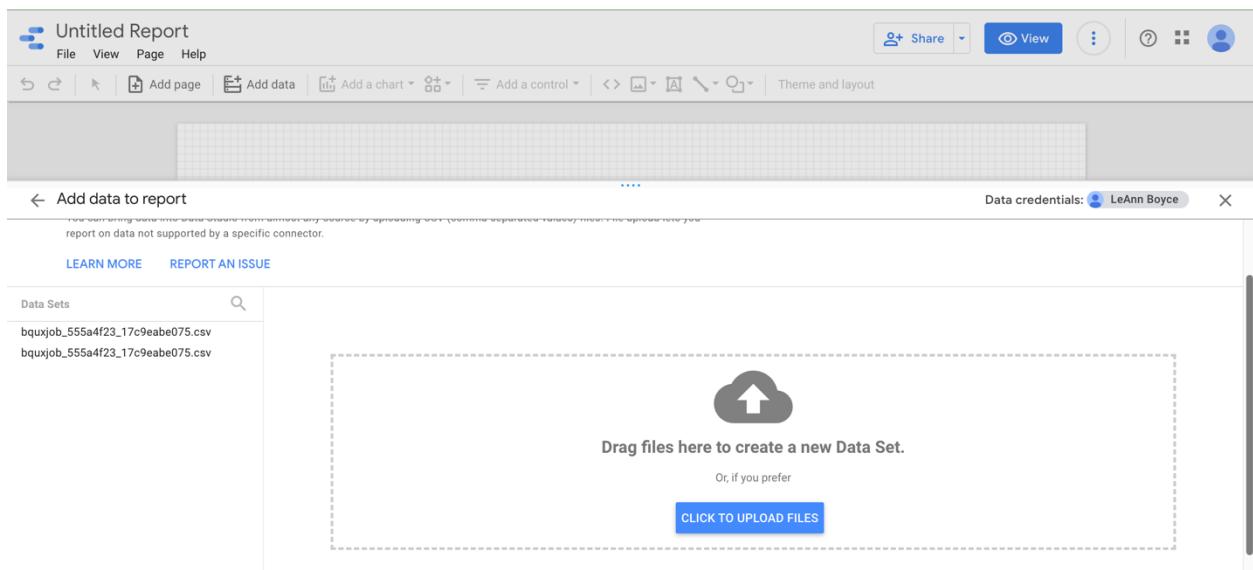
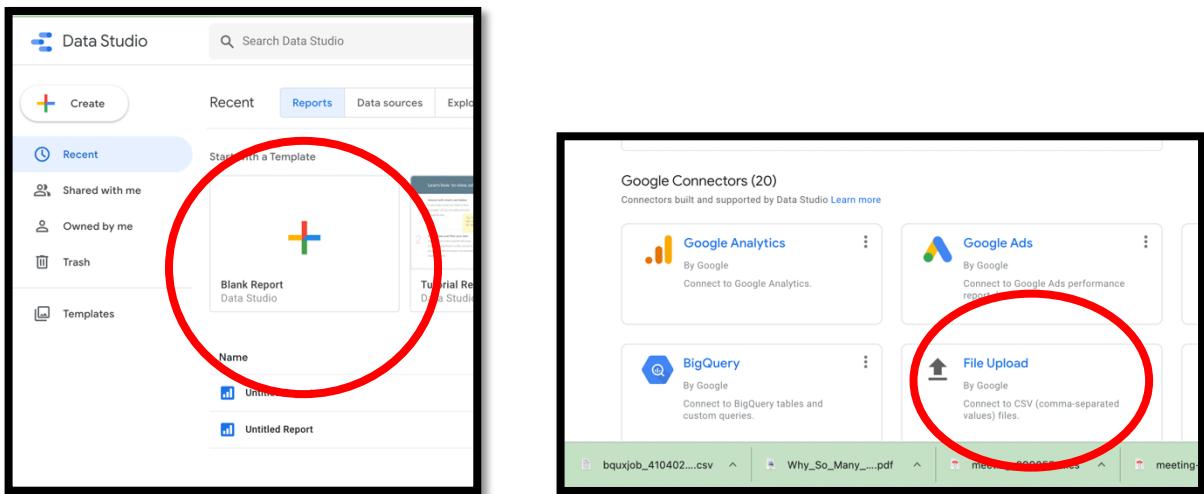
Row	geoareaname	timeperiod	GDP_per_Capita_growth	country_name	year	WB_Population
1	Bahrain	2016	-0.61191	Bahrain	2016	1425793.0
2	Samoa	2016	5.07584	Samoa	2016	194540.0
3	Chad	2016	-6.36186	Chad	2016	1.4561658E7
4	Marshall Islands	2016	2.77816	Marshall Islands	2016	57723.0
5	Maldives	2016	3.84187	Maldives	2016	475505.0
6	Nicaragua	2016	3.54631	Nicaragua	2016	6303970.0
7	Kuwait	2016	-0.49445	Kuwait	2016	3956862.0

Step 3: Let's take it one step further and make a visualization of the data Click on "SAVE RESULTS" and use Google Data Studio as a visualization tool for the homework assignment. I also challenge you to include any other type of visualization tool you have worked with in the past (Excel, Tableau, PowerBI, etc.) for 2 extra credit points. Below is some more information on Google Data Studio.

Explode with Data Studio
Visualize results and create live dashboards from your data.

Row	geoareaname	timeperiod	GDP_per_Capita_growth	country_name	year	WB_Population
1	Bahrain	2016	-0.61191	Bahrain	2016	1425793.0
2	Samoa	2016	5.07584	Samoa	2016	194540.0
3	Chad	2016	-6.36186	Chad	2016	1.4561658E7

Step 4: Google has Data Studio that can be accessed through <https://datastudio.google.com/>. Click on “Blank Report” and then “File Upload (required).” If you’d like to, you can also connect from BigQuery. Give it a try (optional).



Walk through the other steps to upload your data. Lastly, view the video for more tips on Google Data Studio: <https://youtu.be/FkxaBRiXIVc> You can, of course, find other tutorials to help you improve your visualization.

Here is an example of charts created in Google Data Studio with the data from above. I would like to see you play around with the charts, color, and text. See what you can create. However, you are only required to show two visualizations, like below, for the assignment.

