
SOCIAL DATA SCIENCE

# Accessing

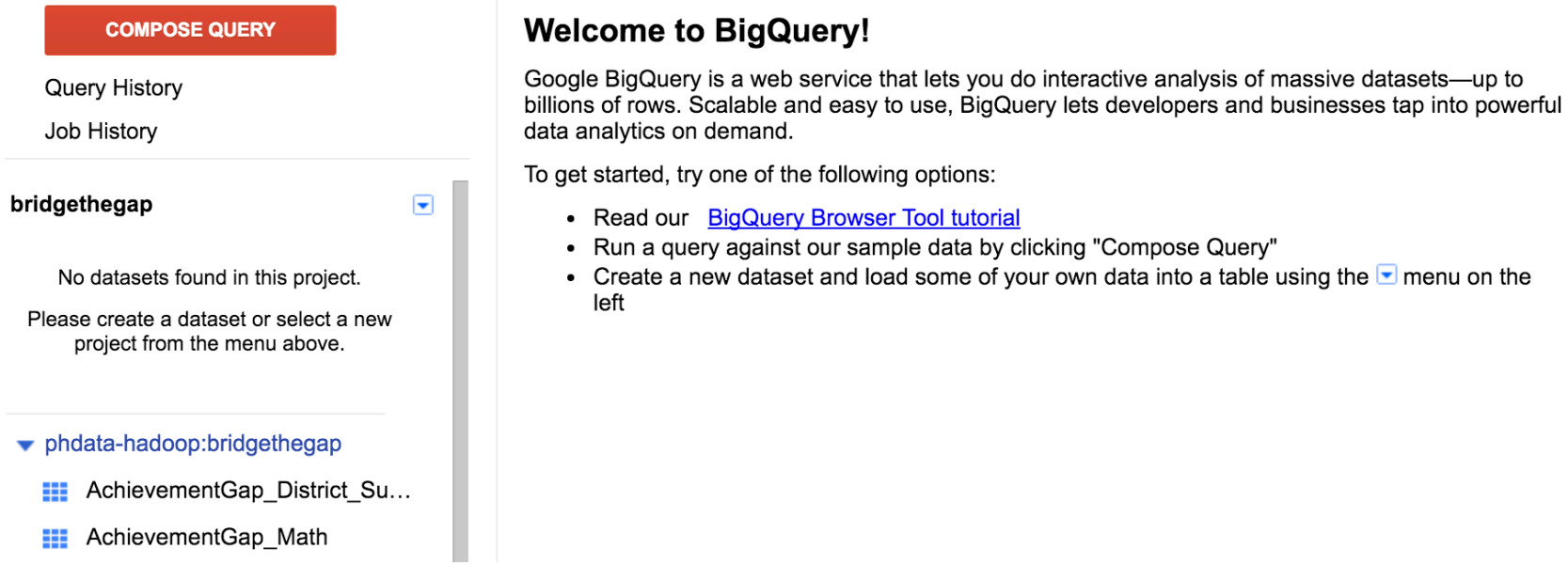
Google BigQuery

Presented by Brock Noland & John Hogue

# Accessing via Browser

# Accessing via Browser

- Navigate your browser to:
  - https://bigquery.cloud.google.com/dataset/phdata-hadoop:bridgethegap



# Accessing via Browser

- Click on Table to View Table Schema

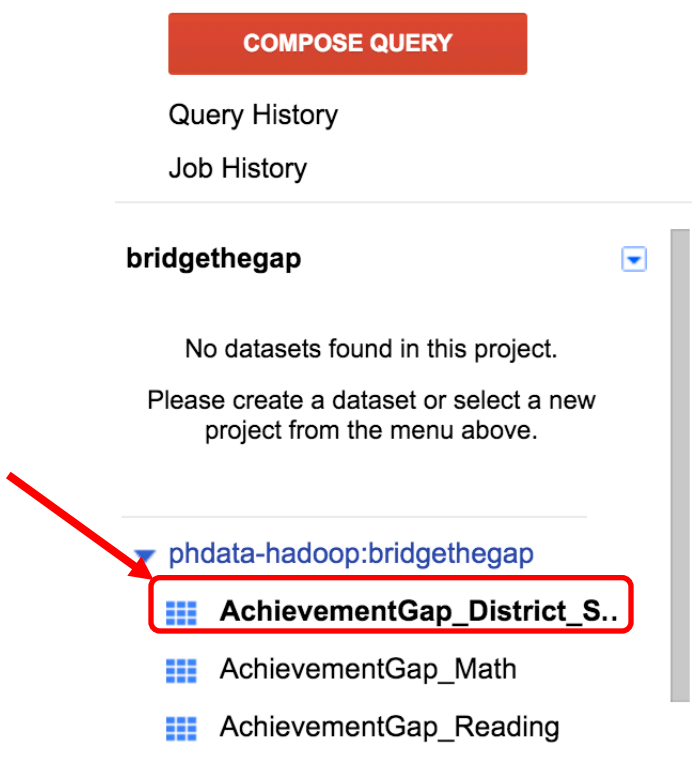



**Table Details: AchievementGap_District_Summary**

Schema | Details | Preview

| | | | |
|---|---|---|---|
| **index** | INTEGER | NULLABLE | Describe this field... |
| **Year** | INTEGER | NULLABLE | Describe this field... |
| **NCLBID** | INTEGER | NULLABLE | Describe this field... |
| **districtNumber** | INTEGER | NULLABLE | Describe this field... |
| **districtType** | INTEGER | NULLABLE | Describe this field... |
| **DistrictName** | STRING | NULLABLE | Describe this field... |
| **Subject** | STRING | NULLABLE | Describe this field... |
| **CountofStudentGroupsIdentified** | INTEGER | NULLABLE | Describe this field... |

# Accessing via Browser

- Run a Query

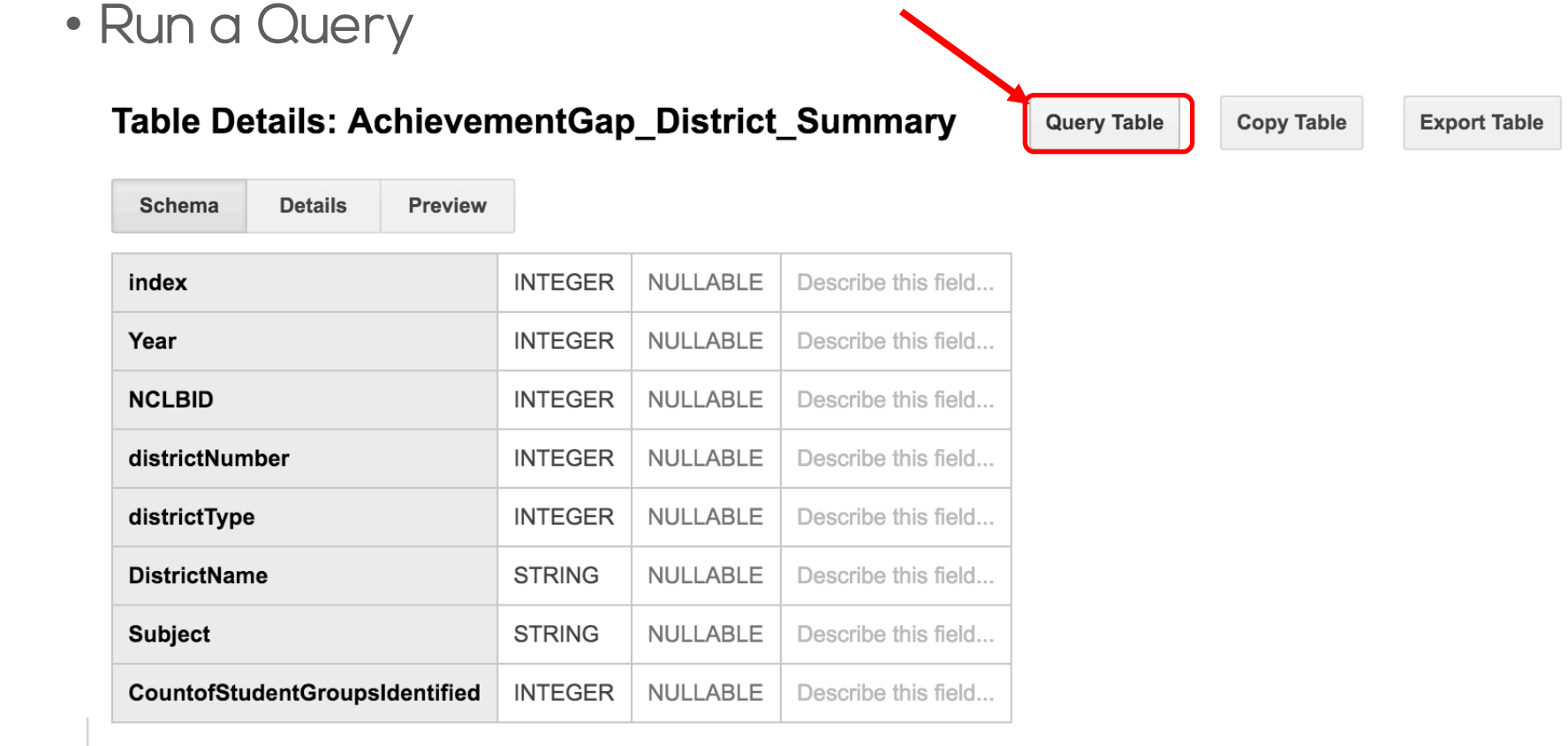


**Table Details: AchievementGap_District_Summary**

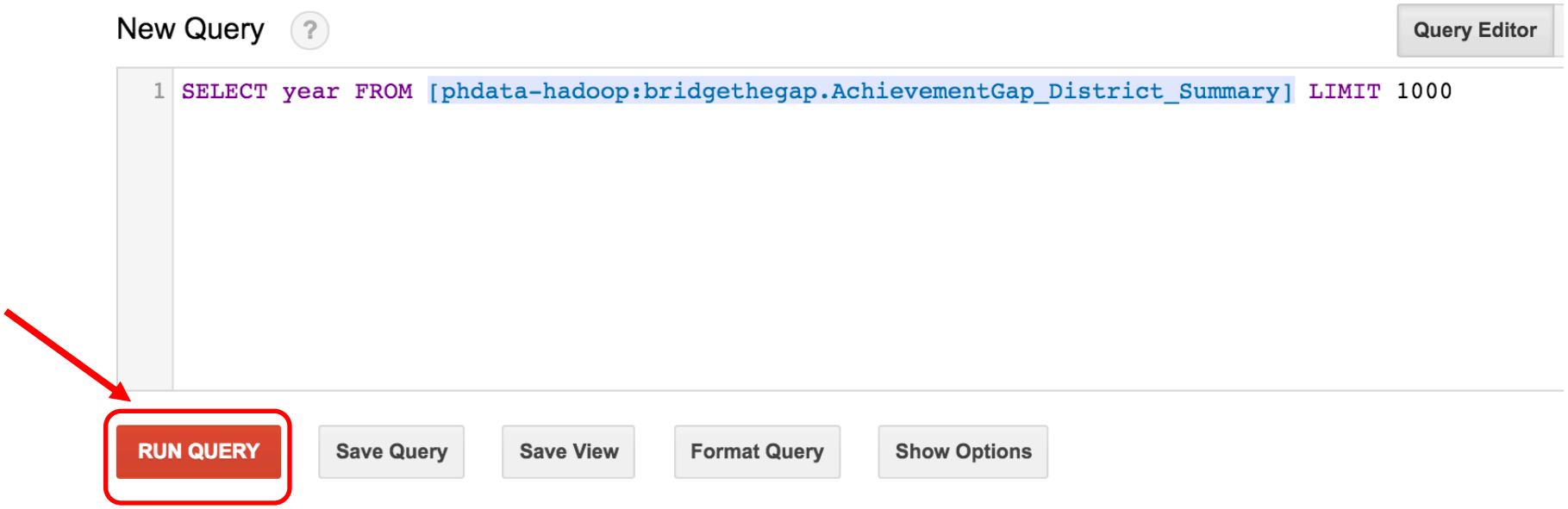
Query Table | Copy Table | Export Table

Schema | Details | Preview

| | | | |
|---|---|---|---|
| **index** | INTEGER | NULLABLE | Describe this field... |
| **Year** | INTEGER | NULLABLE | Describe this field... |
| **NCLBID** | INTEGER | NULLABLE | Describe this field... |
| **districtNumber** | INTEGER | NULLABLE | Describe this field... |
| **districtType** | INTEGER | NULLABLE | Describe this field... |
| **DistrictName** | STRING | NULLABLE | Describe this field... |
| **Subject** | STRING | NULLABLE | Describe this field... |
| **CountofStudentGroupsIdentified** | INTEGER | NULLABLE | Describe this field... |

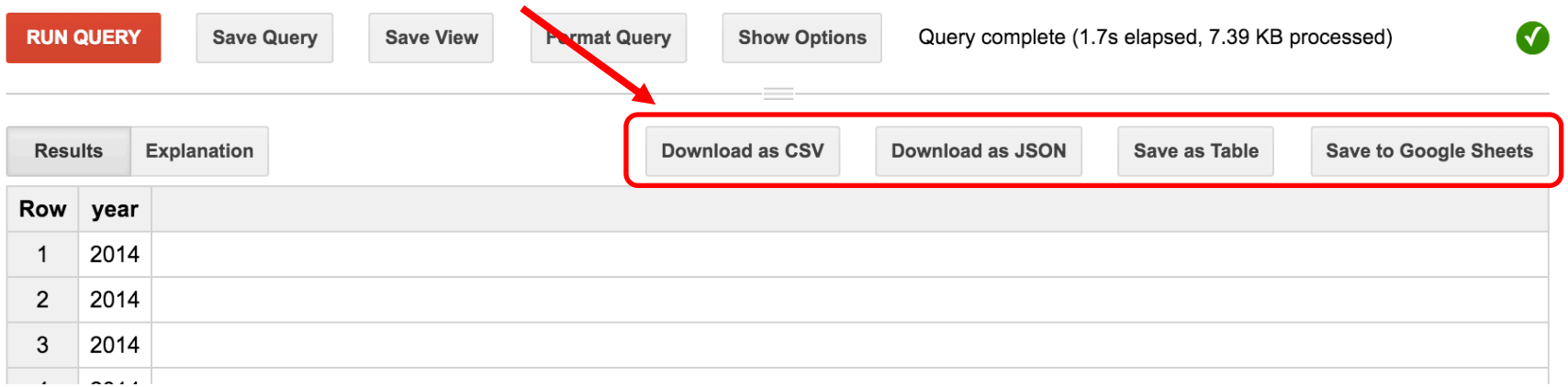
# Accessing via Browser

- Run a Query



# Accessing via Browser

- Export or Materialize Results



# Accessing via Browser

- Export Whole Table Data

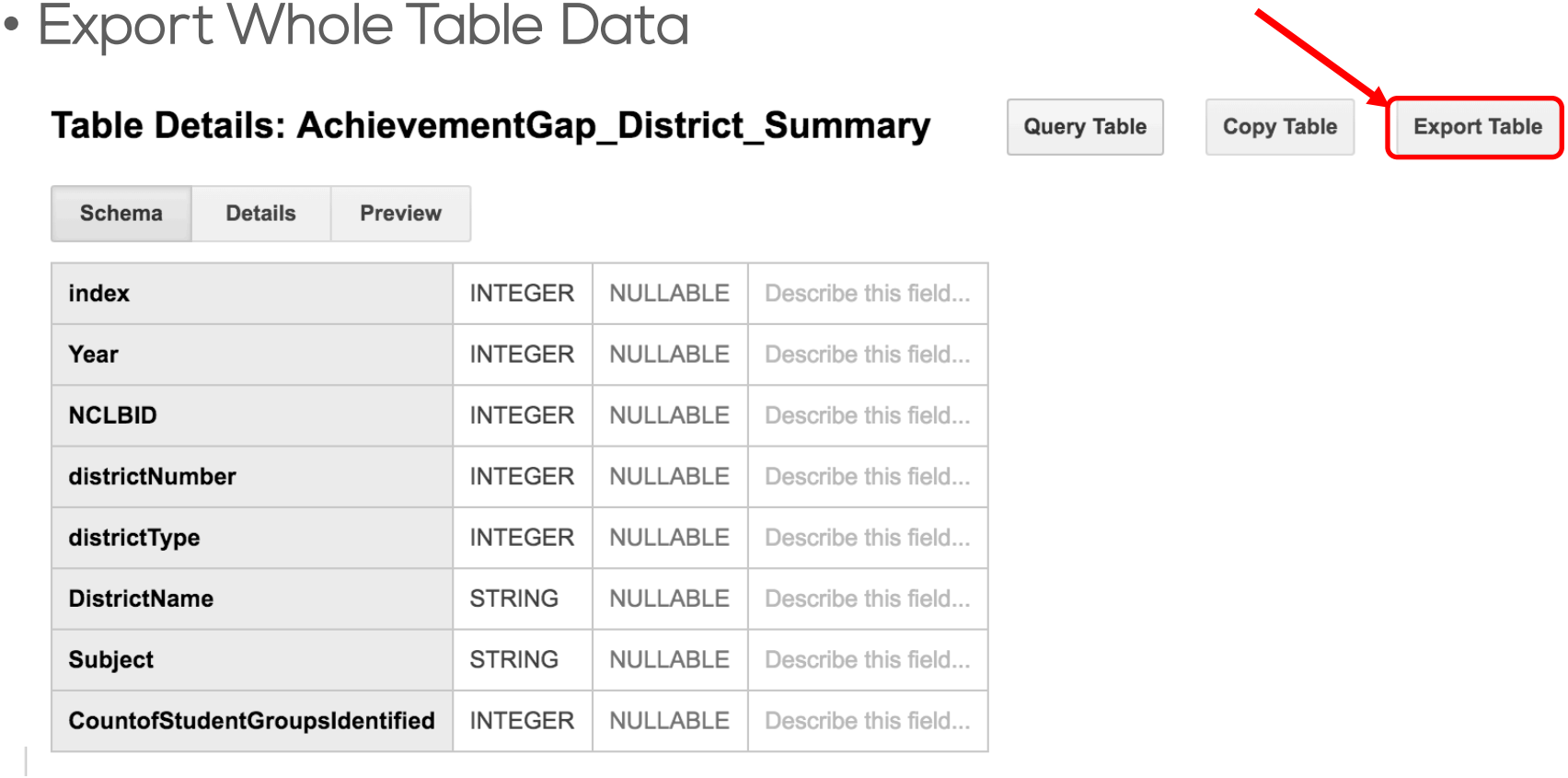


**Table Details: AchievementGap_District_Summary**

Query Table | Copy Table | Export Table

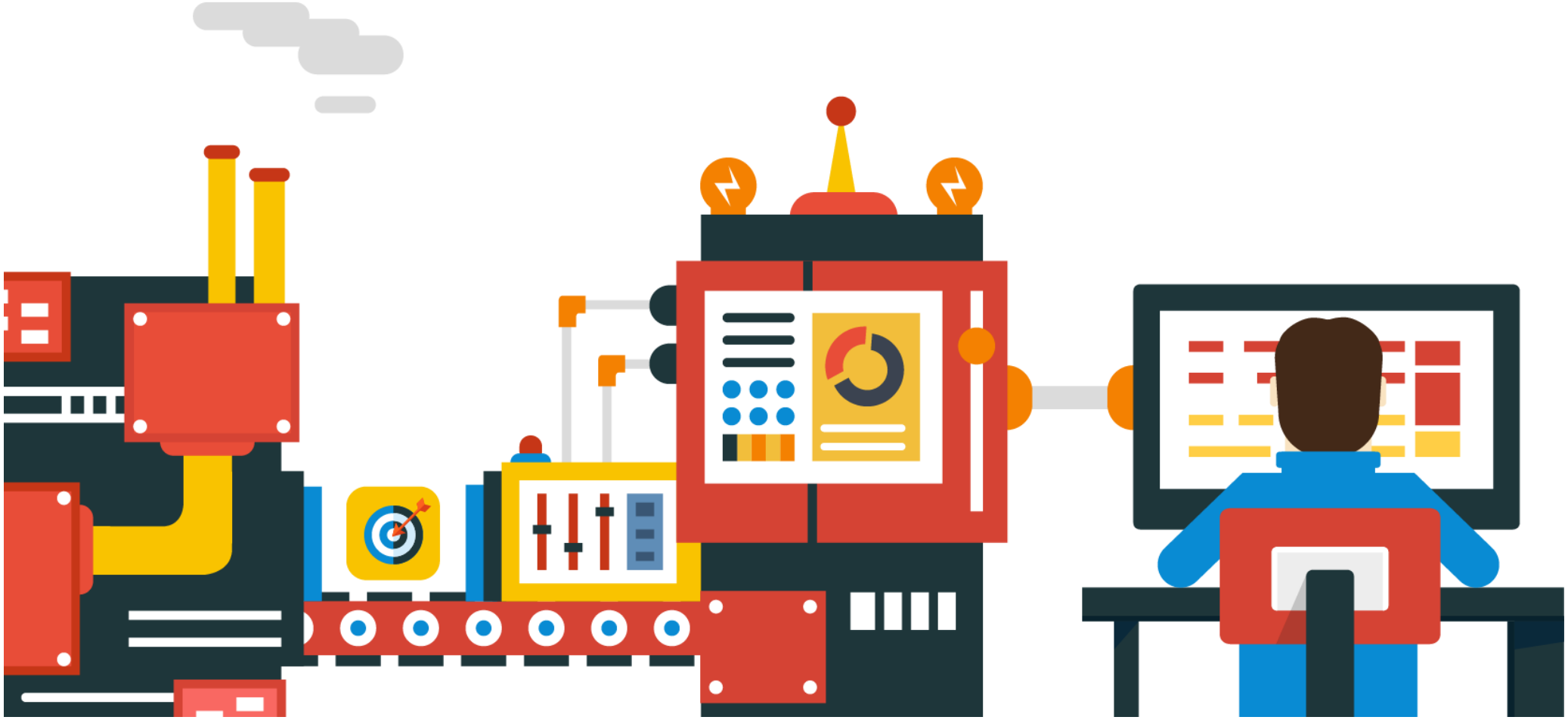
Schema | Details | Preview

| | | | |
|---|---|---|---|
| **index** | INTEGER | NULLABLE | Describe this field... |
| **Year** | INTEGER | NULLABLE | Describe this field... |
| **NCLBID** | INTEGER | NULLABLE | Describe this field... |
| **districtNumber** | INTEGER | NULLABLE | Describe this field... |
| **districtType** | INTEGER | NULLABLE | Describe this field... |
| **DistrictName** | STRING | NULLABLE | Describe this field... |
| **Subject** | STRING | NULLABLE | Describe this field... |
| **CountofStudentGroupsIdentified** | INTEGER | NULLABLE | Describe this field... |

# Accessing via Browser

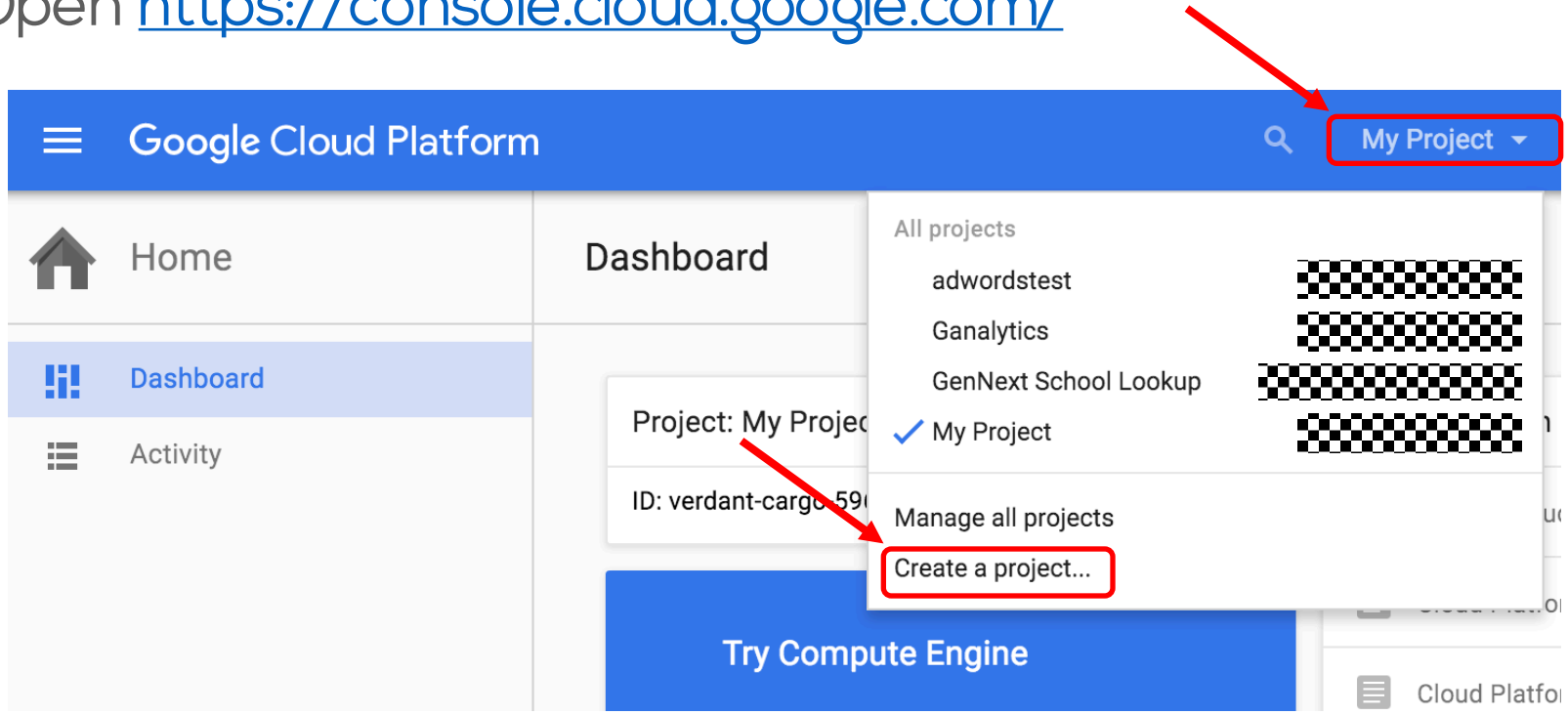
- Other things you can do
  - Copy Tables to your own project (and make your own changes)
  - Load your own data
  - Make your own query functions (UDFs)

# PROGRAMMATIC



# Setting Up Programmatic Access...

- Open https://console.cloud.google.com/

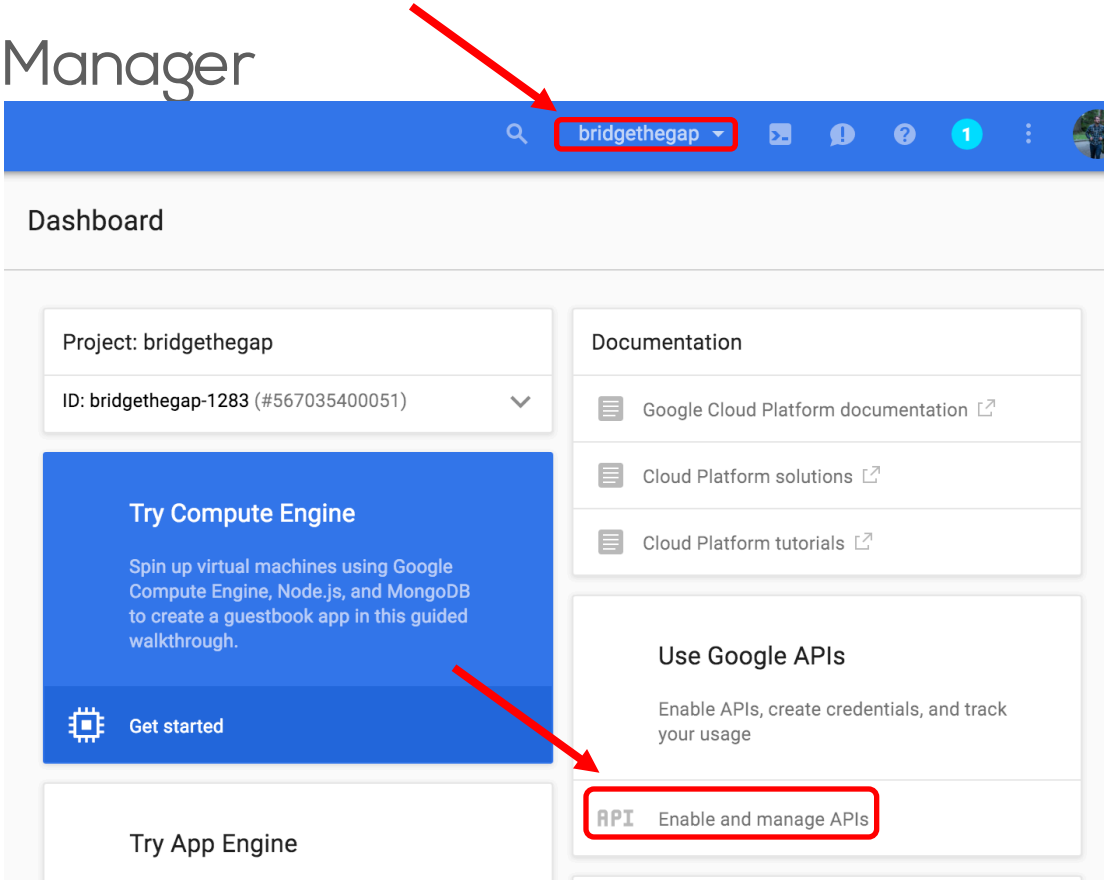


# Setting Up Programmatic Access...

- Enter a Project Name
- Copy these down in your code comments

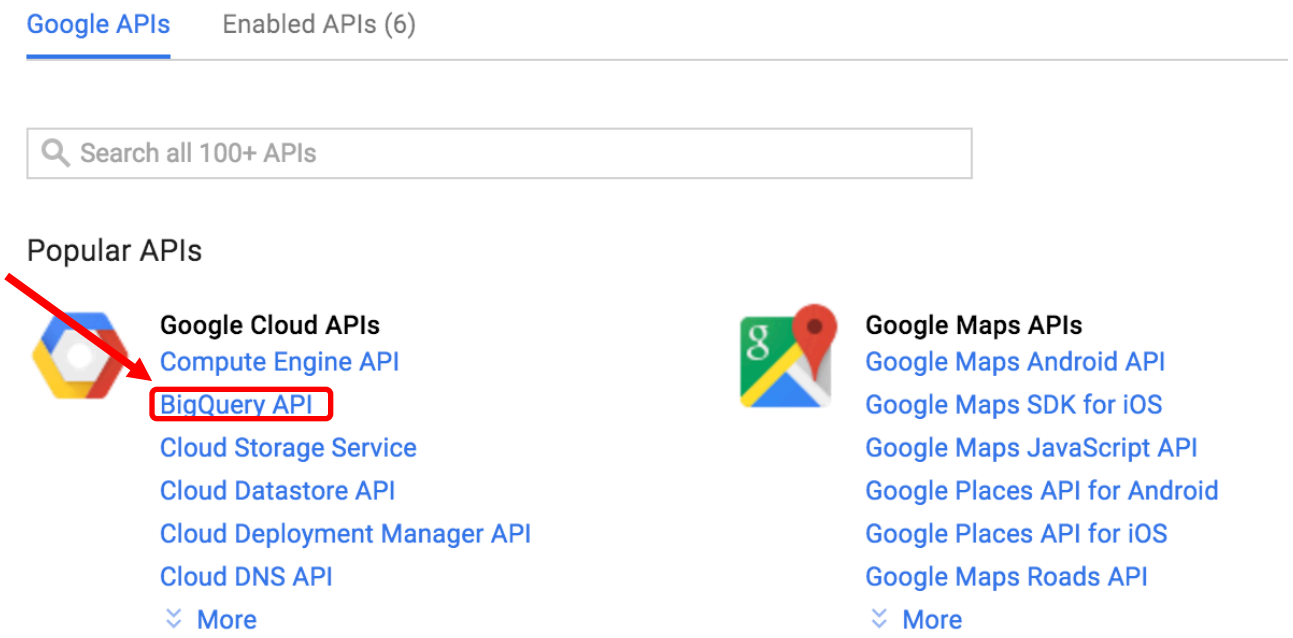
# Setting Up Programmatic Access...

- Go to API Manager



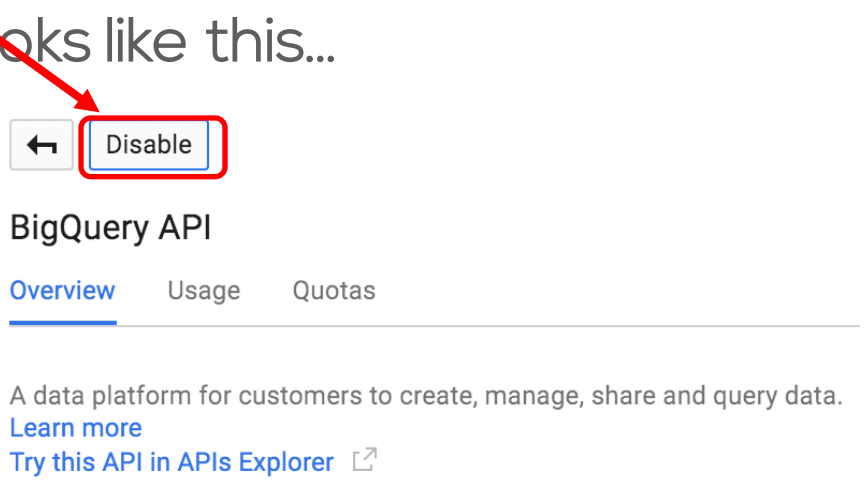
# Setting Up Programmatic Access...

- Select BigQuery API



# Setting Up Programmatic Access...

- Ensure BigQuery API is Enabled (enables on accessing this page)
- If enabled, it looks like this...



# Accessing via R

# Accessing via R

- Code examples in R section of:
  - https://github.com/SocialDataSci/AccessingBigQuery

- BigrQuery Repository
  - https://github.com/rstats-db/bigrquery

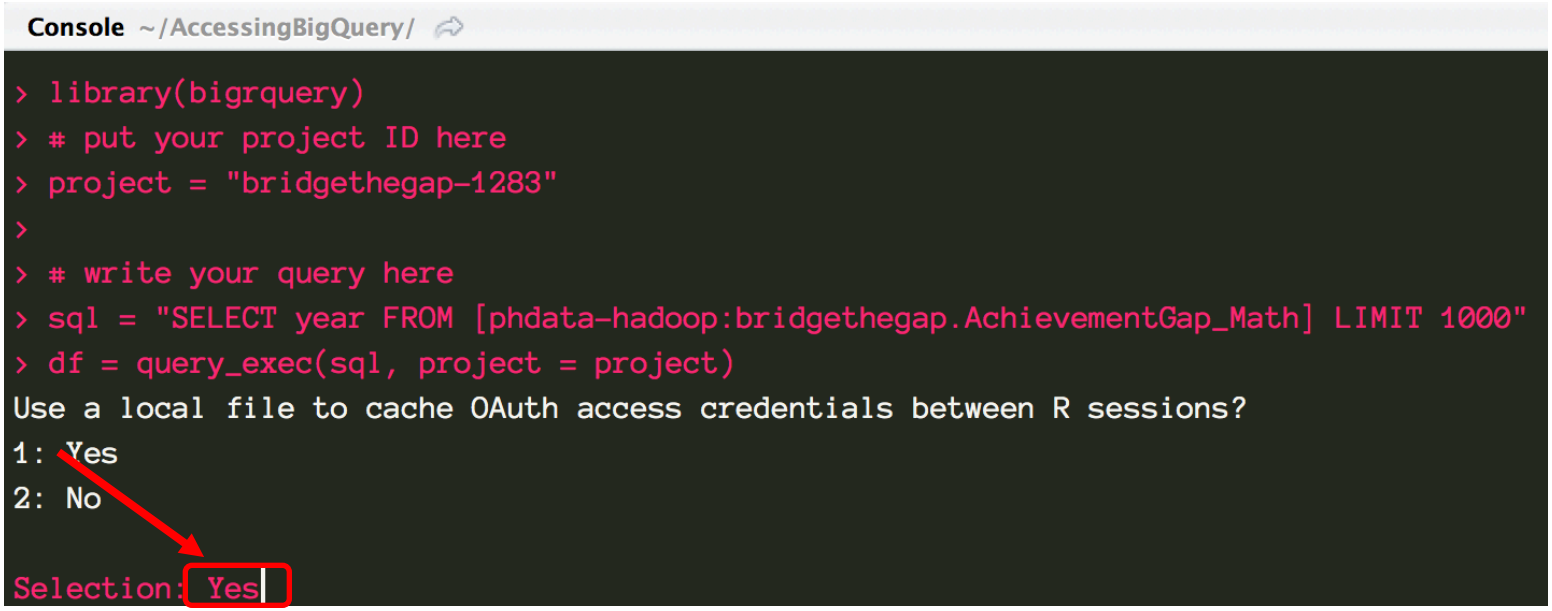
- BigrQuery Documentation
  - https://cran.r-project.org/web/packages/bigrquery/bigrquery.pdf

# Accessing via R

- Install BigrQuery
  - install.packages("bigrquery")

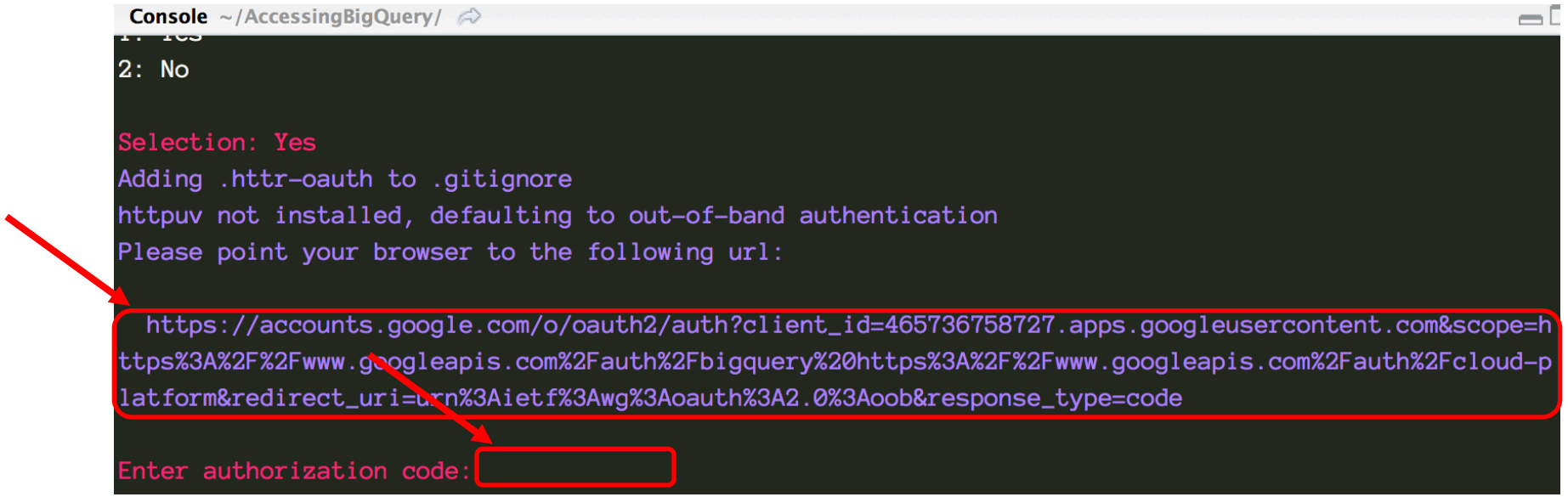
# Accessing via R

- First time Authentication
  - R will ask to cache OAuth credentials
  - Make sure to add .httr-oauth to .gitignore if using VCS



# Accessing via R

- First time Authentication
  - Copy link to browser
  - Enter code that shows up after logging into to Google



# Accessing via R

- Dumping Query Results to Data.Frame

```
# import the bigrquery package for use
library(bigrquery)

# put your project ID here
project = "bridgethegap-1283"

# write your query here
sql = "SELECT year FROM [phdata-hadoop:bridgethegap.AchievementGap_Math] LIMIT 1000"
df = query_exec(sql, project = project)
```

# Accessing via R

- Best Practices
  - Rate Limited
    - Don't hammer the system
    - Sample data first before pulling all of it
    - https://cloud.google.com/bigquery/quota-policy
  - Joins, Filters, Group Bys
    - Best to let BigQuery do to the work
      - Our data is pretty small and shouldn't be an issue

# Accessing via Python

# Accessing via Python

- Code examples in Python section of:
  - https://github.com/SocialDataSci/AccessingBigQuery

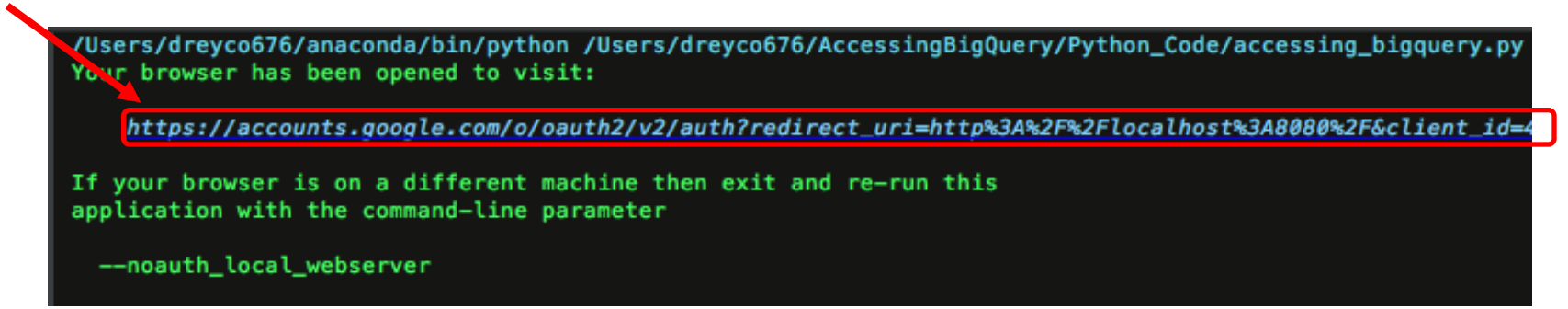
- BigQuery API Client Library for Python
  - https://developers.google.com/api-client-library/python/apis/bigquery/v2
  - Official & supported by Google

# Accessing via Python

- Install google-api-python-client
  - Go to terminal, CMD or IDE console
  - pip install --upgrade google-api-python-client

# Accessing via Python

- First time Authentication
  - Python will ask to cache OAuth credentials
  - Browser should open, allow access.



# Accessing via Python

- Dumping Query Results to Pandas Dataframe

```
from pandas import io

# put your project ID here
project_id = "bridgethegap-1283"

# write your query here
query = "SELECT year FROM [phdata-hadoop:bridgethegap.AchievementGap_Math] LIMIT 1000"
df = io.gbq.read_gbq(query, project_id=project_id)
```

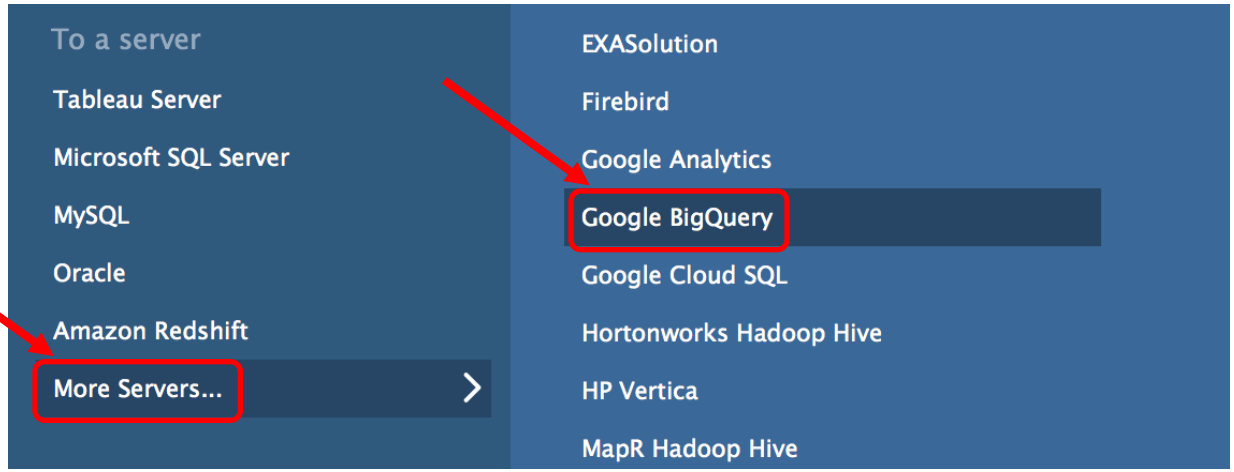
# Accessing via Python

- Best Practices
  - Rate Limited
    - Don't hammer the system
    - Sample data first before pulling all of it
    - https://cloud.google.com/bigquery/quota-policy
  - Joins, Filters, Group Bys
    - Best to let BigQuery do to the work
      - Our data is pretty small and shouldn't be an issue

# Accessing via Tableau

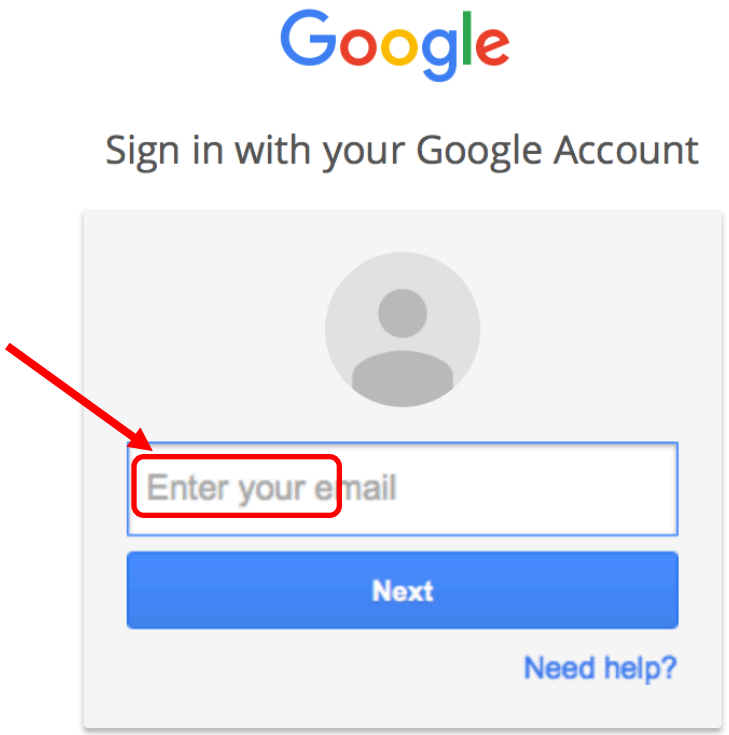
# Accessing via Tableau

- Go to Connect Screen and Find Google BigQuery



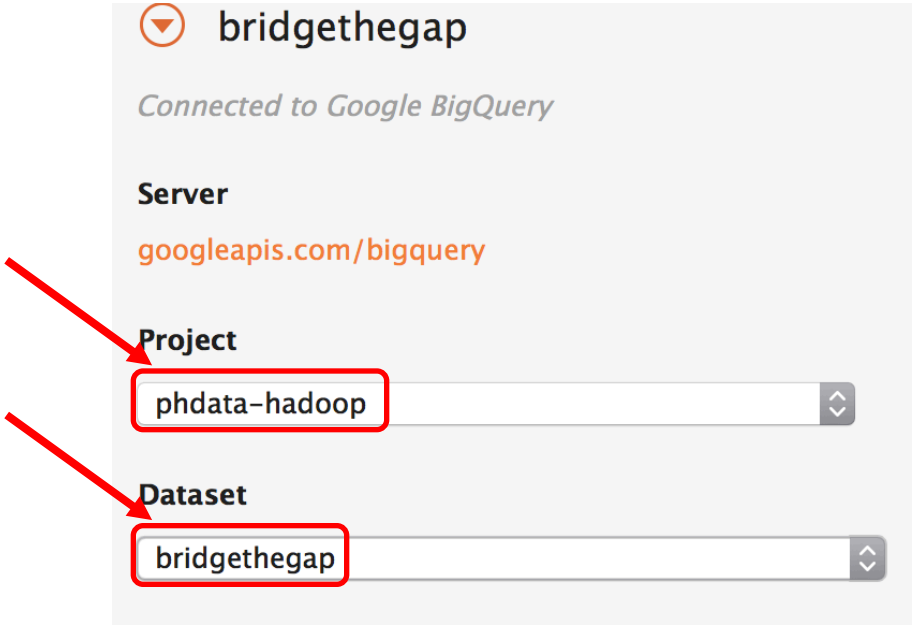
# Accessing via Tableau

- Login to Google
- Allow Access



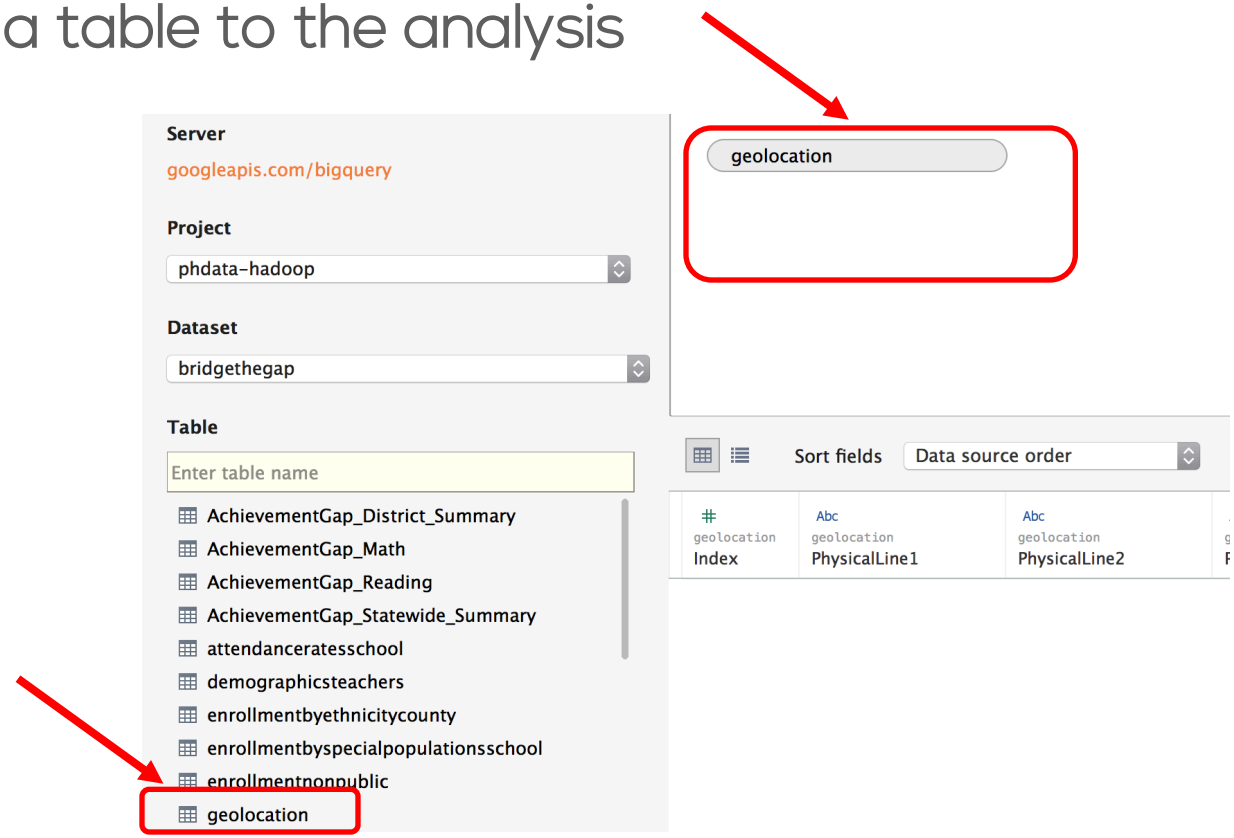
# Accessing via Tableau

- Type 'phdata-hadoop' for project
  - (this is a shared project)



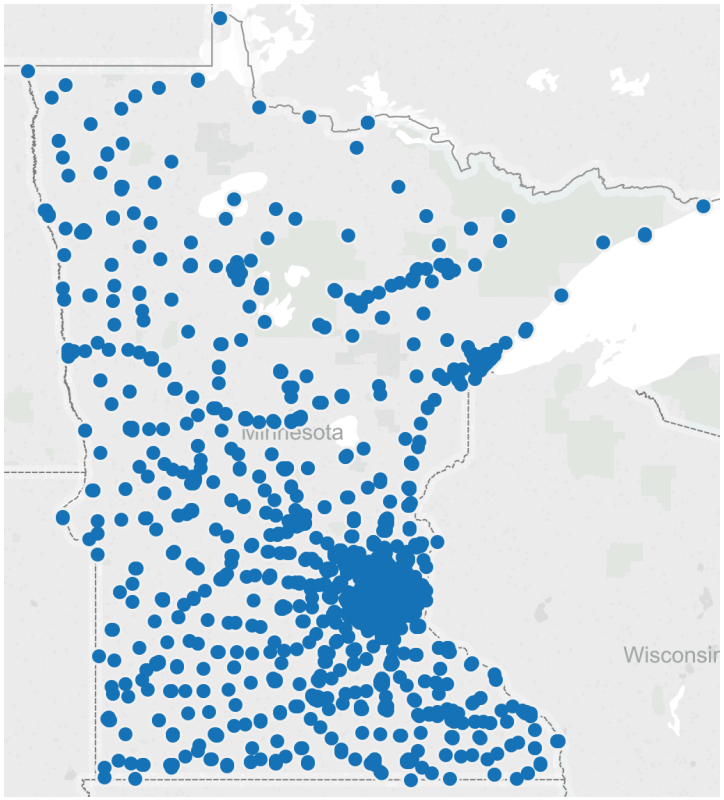
# Accessing via Tableau

- Drag a table to the analysis



# Accessing via Tableau

- Start Analysis



# Questions?