



# 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>

**COMPOSE QUERY**

Query History  
Job History

**bridgethegap** ☐

No datasets found in this project.  
Please create a dataset or select a new project from the menu above.

☒ phdata-hadoop:bridgethegap

☒ AchievementGap\_District\_Su...

☒ AchievementGap\_Math

## Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our [BigQuery Browser Tool tutorial](#)
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the ☐ menu on the left



# Accessing via Browser

- Click on Table to View Table Schema

COMPOSE QUERY

Query History

Job History

bridgethegap

No datasets found in this project.

Please create a dataset or select a new project from the menu above.

phdata-hadoop:bridgethegap

 AchievementGap\_District\_S..

 AchievementGap\_Math

 AchievementGap\_Reading

## 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

New Query ?

Query Editor

```
1 SELECT year FROM [phdata-hadoop:bridgethegap.AchievementGap_District_Summary] LIMIT 1000
```

**RUN QUERY**

Save Query

Save View

Format Query

Show Options



# Accessing via Browser

- Export or Materialize Results

**RUN QUERY** Save Query Save View Format Query Show Options Query complete (1.7s elapsed, 7.39 KB processed) ✓

ResultsExplanation

Download as CSV Download as JSON Save as Table Save to Google Sheets

Row	year	
1	2014	
2	2014	
3	2014	
4	2014	





# Accessing via Browser

- Export Whole Table Data

## Table Details: AchievementGap\_District\_Summary

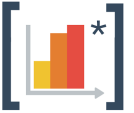
Query Table

Copy Table

Export Table

Schema	Details	Preview
--------	---------	---------

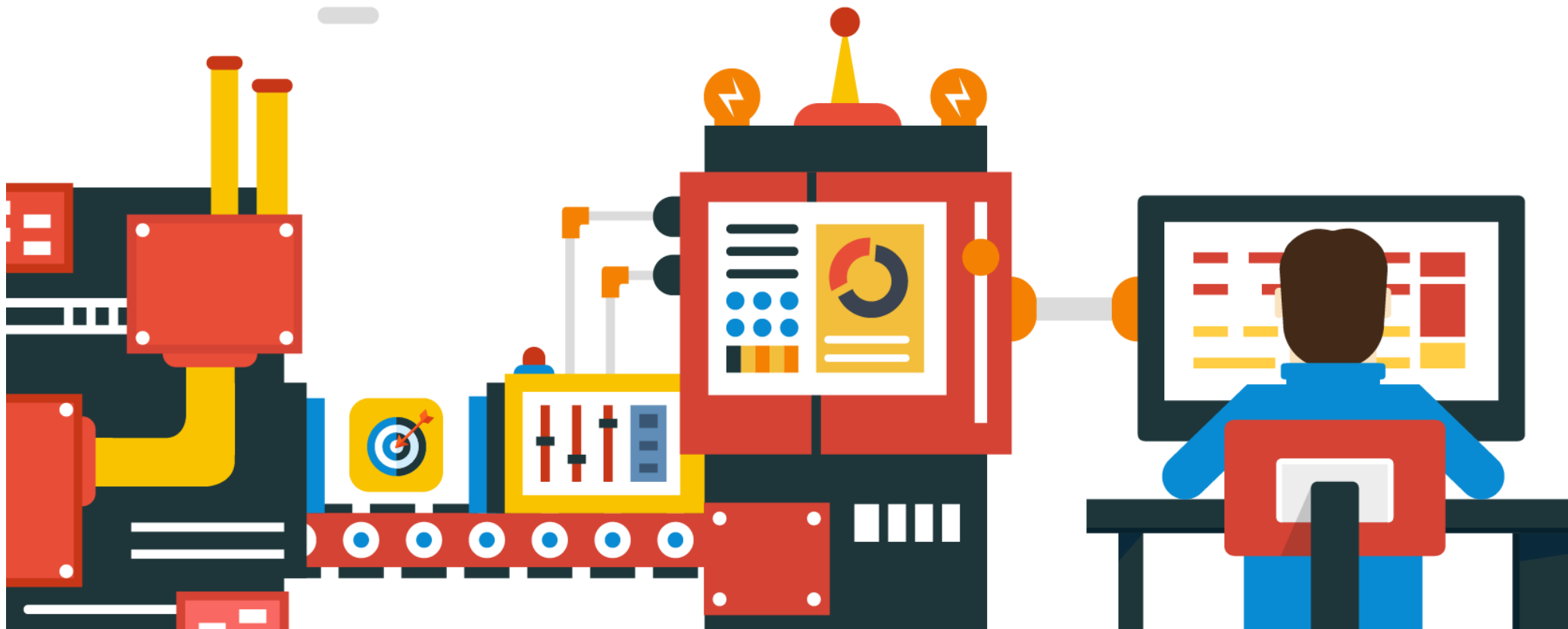
index	INTEGER	NULLABLE	Describe this field...
Year	INTEGER	NULLABLE	Describe this field...
NCLBID	INTEGER	NULLABLE	Describe this field...
districtNumber	INTEGER	NULLABLE	Describe this field...
districtType	INTEGER	NULLABLE	Describe this field...
DistrictName	STRING	NULLABLE	Describe this field...
Subject	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)

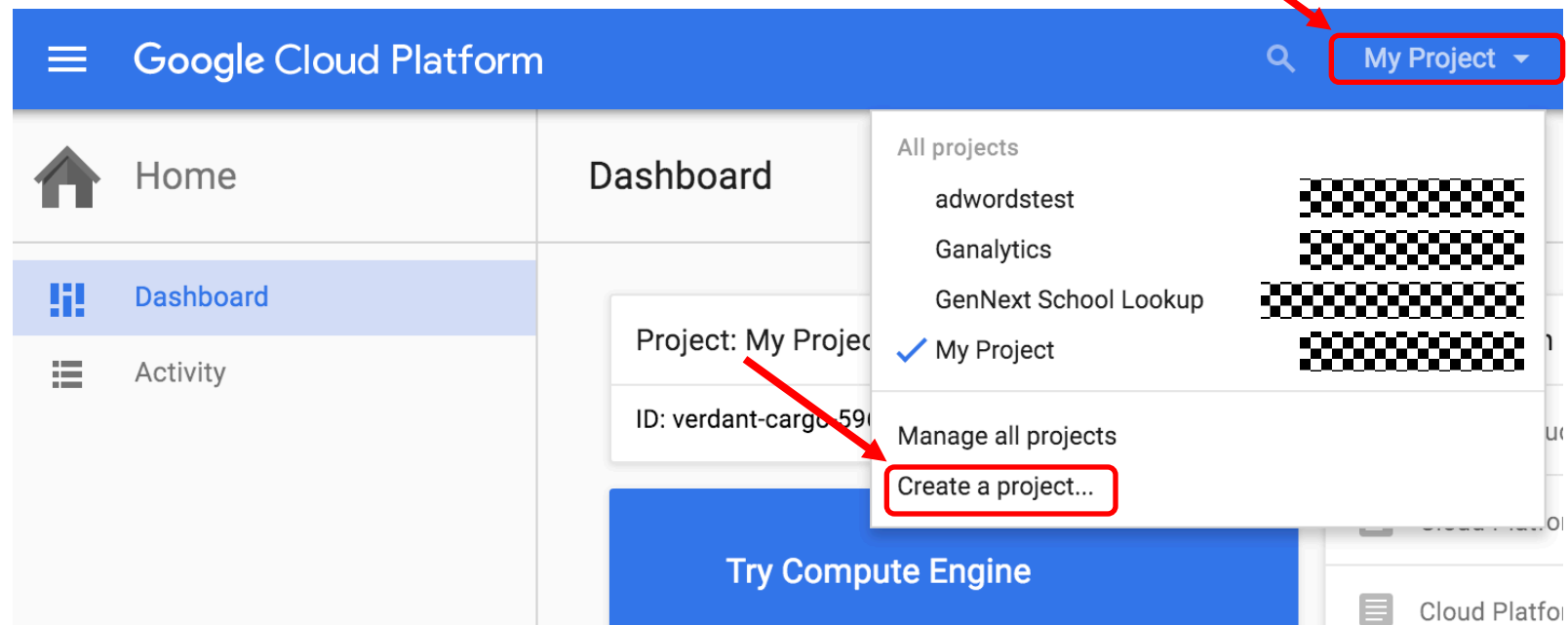
# PROGRAMMATIC





# 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

New Project

Project name ?

bridgethegap

Your project ID will be bridgethegap-1283 ? Edit

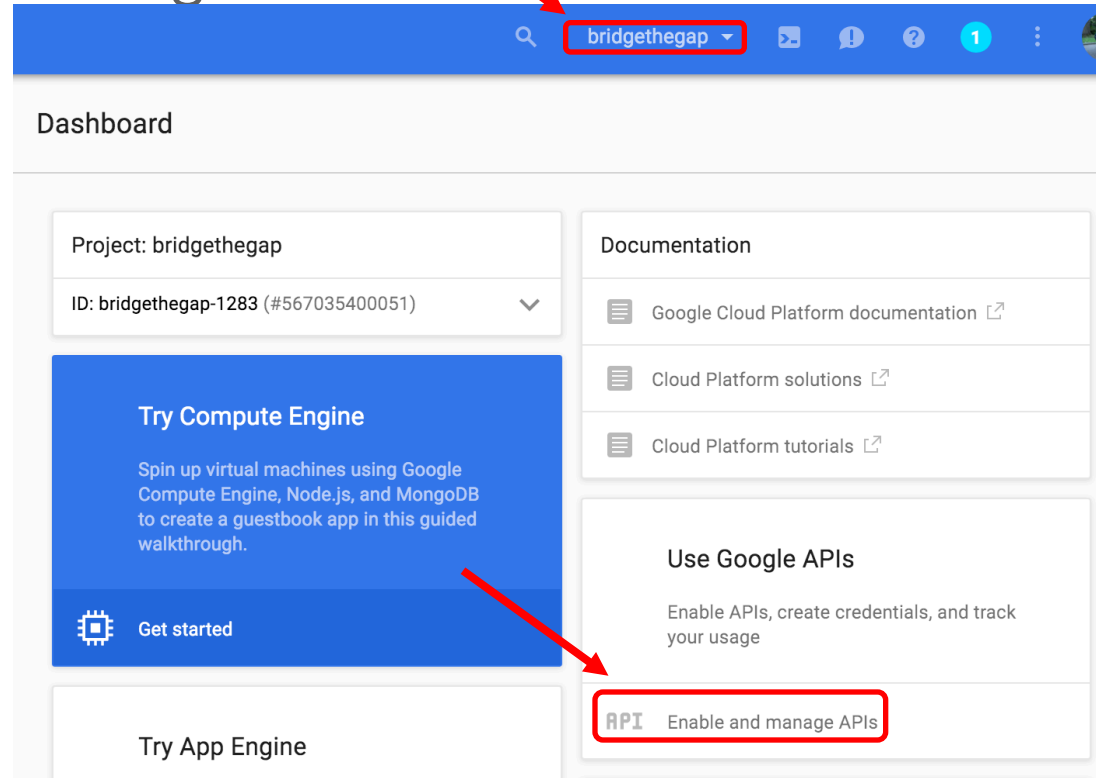
[Show advanced options...](#)

Create Cancel



# Setting Up Programmatic Access...

- Go to API Manager






# Setting Up Programmatic Access...

- Select BigQuery API


[Google APIs](#) Enabled APIs (6)

---

Popular APIs



- Google Cloud APIs
  - [Compute Engine API](#)
  - [BigQuery API](#)**
  - [Cloud Storage Service](#)
  - [Cloud Datastore API](#)
  - [Cloud Deployment Manager API](#)
  - [Cloud DNS API](#)
  - [More](#)



- Google Maps APIs
  - [Google Maps Android API](#)
  - [Google Maps SDK for iOS](#)
  - [Google Maps JavaScript API](#)
  - [Google Places API for Android](#)
  - [Google Places API for iOS](#)
  - [Google Maps Roads API](#)
  - [More](#)



# Setting Up Programmatic Access...

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



BigQuery API

[Overview](#) Usage Quotas

A data platform for customers to create, manage, share and query data.

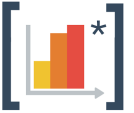
[Learn more](#)

[Try this API in APIs Explorer](#) [↗](#)



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

```
Console ~/AccessingBigQuery/ ↵  
  
> 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)  
Use a local file to cache OAuth access credentials between R sessions?  
1: Yes  
2: No  
Selection: Yes
```



# Accessing via R

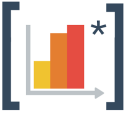
- First time Authentication
  - Copy link to browser
  - Enter code that shows up after logging into to Google

```
Console ~/AccessingBigQuery/
1: Yes
2: No

Selection: Yes
Adding .httr-oauth to .gitignore
httpuv not installed, defaulting to out-of-band authentication
Please point your browser to the following url:

https://accounts.google.com/o/oauth2/auth?client_id=465736758727.apps.googleusercontent.com&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fbigquery%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%3Aaob&response_type=code

Enter authorization code: 
```



# 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"

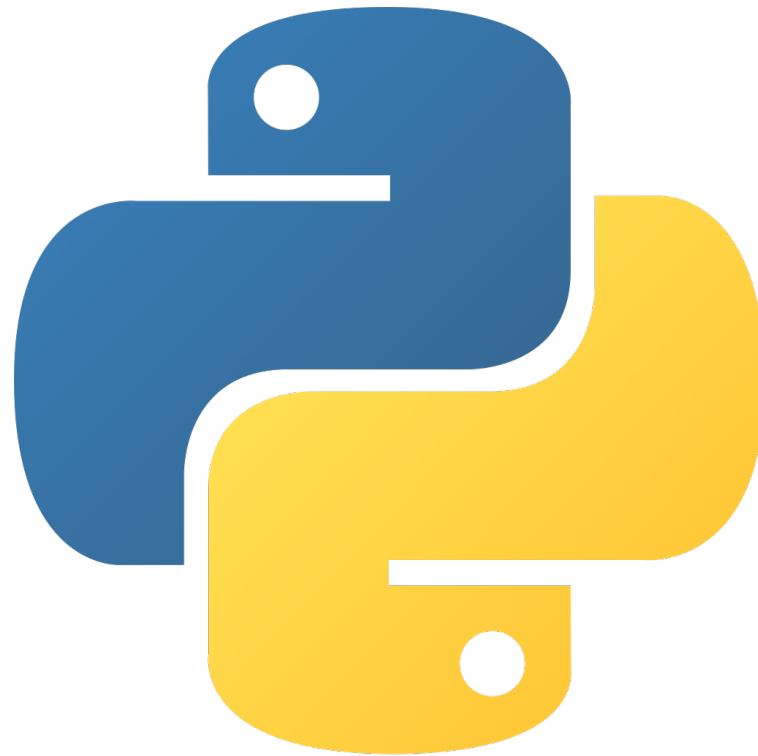
# write your query here
sql = "SELECT year FROM [phdata-hadoop:bridgethegap.AchievementGap_Math] LIMIT 1000"
df = query_exec(sql, project = project)
```



# 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

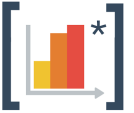






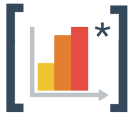
# Accessing via Python

- 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.

```
/Users/dreyco676/anaconda/bin/python /Users/dreyco676/AccessingBigQuery/Python_Code/accessing_bigquery.py  
Your browser has been opened to visit:  
https://accounts.google.com/o/oauth2/v2/auth?redirect\_uri=http%3A%2F%2Flocalhost%3A8080%2F&client\_id=4  
  
If your browser is on a different machine then exit and re-run this  
application with the command-line parameter  
  
--noauth_local_webserver
```



# Accessing via Python

- Dumping Query Results to Pandas Dataframe

```
from pandas import io

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

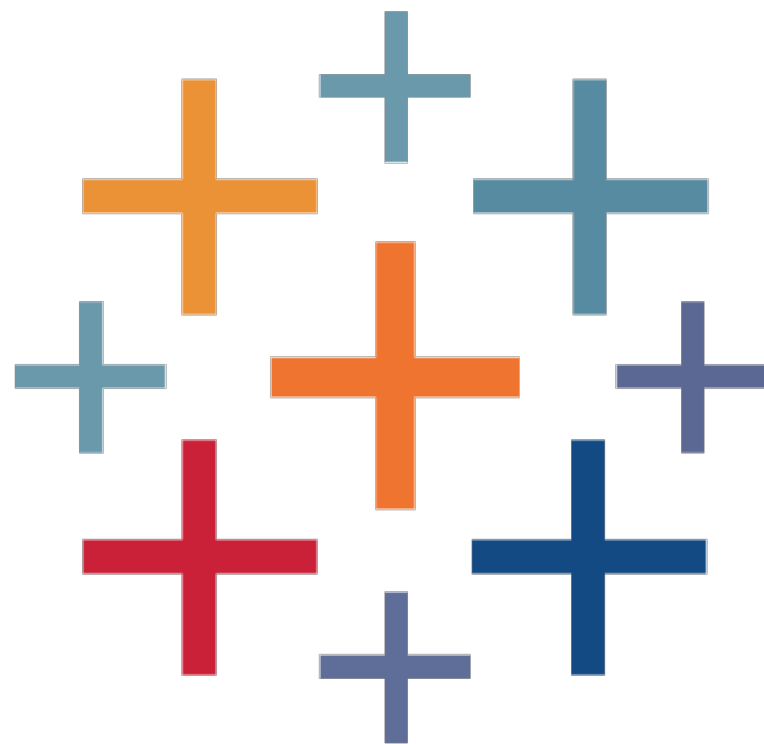
# write your query here
query = "SELECT year FROM [phdata-hadoop:bridgethegap.AchievementGap_Math] LIMIT 1000"
df = io.gbq.read_gbq(query, project_id=project_id)
```

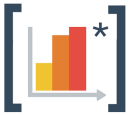


# Accessing via Python

- 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

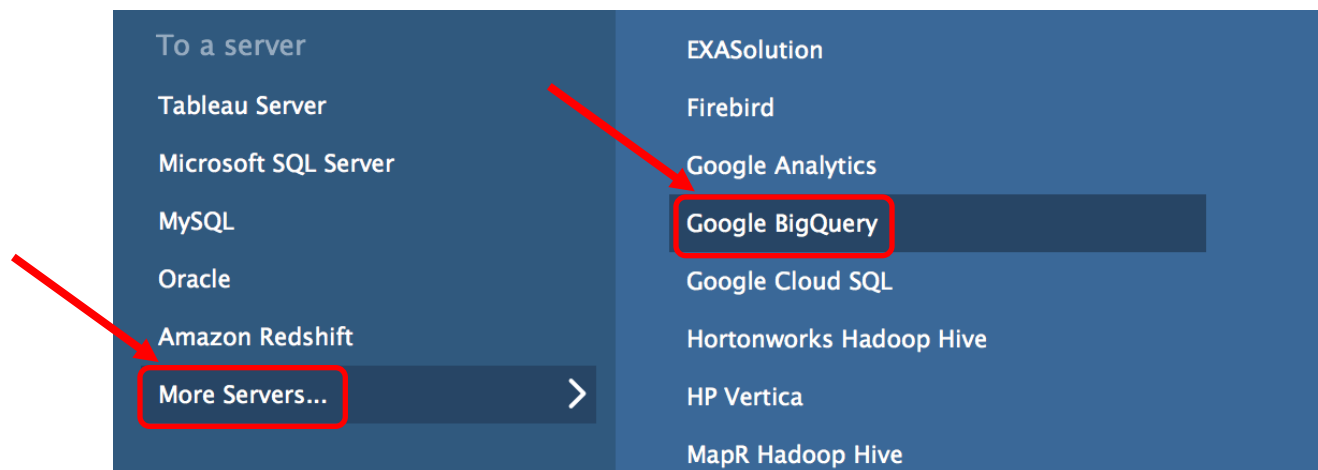
# Accessing via Tableau





# Accessing via Tableau

- Go to Connect Screen and Find Google BigQuery





# Accessing via Tableau

- Login to Google
- Allow Access



Sign in with your Google Account

A screenshot of the Google sign-in interface. It features a grey profile icon placeholder at the top. Below it is a text input field with the placeholder text "Enter your email". A red arrow points from the left towards this input field. Below the input field is a blue button labeled "Next". At the bottom right of the form is a link that says "Need help?".





# Accessing via Tableau

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

▼ bridgethegap

*Connected to Google BigQuery*

**Server**

[googleapis.com/bigquery](https://googleapis.com/bigquery)

**Project**

phdata-hadoop

**Dataset**

bridgethegap



# Accessing via Tableau

- Drag a table to the analysis

The screenshot shows the Tableau interface with the following components:

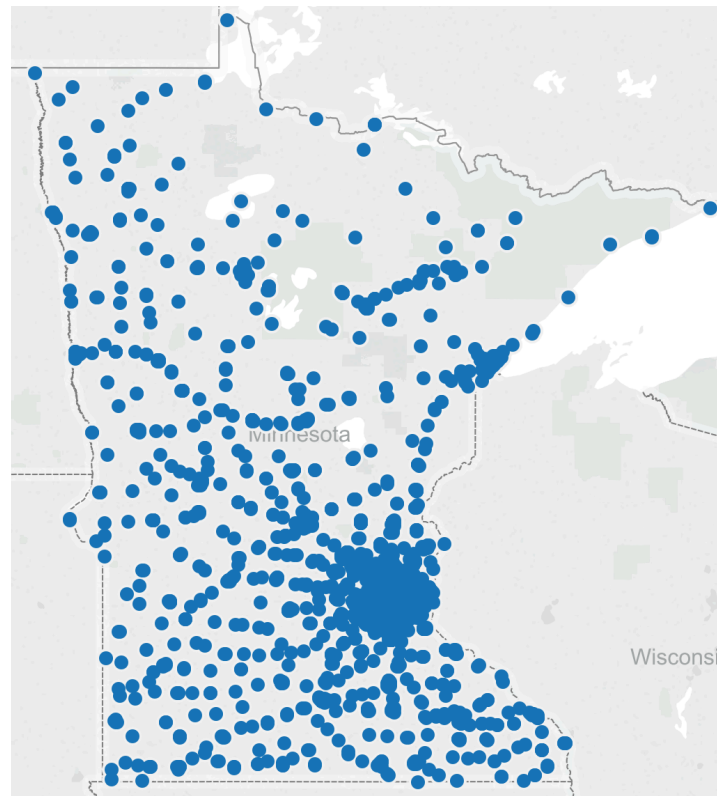
- Server:** googleapis.com/bigquery
- Project:** phdata-hadoop
- Dataset:** bridgethegap
- Table:** A list of tables is displayed, with 'geolocation' highlighted at the bottom. A red arrow points to this table.
- Analysis Area:** A red box highlights the 'geolocation' table being dragged into the analysis area. A red arrow points from the 'geolocation' table in the list to this box.
- Table Structure:** A table with 4 columns is shown below the analysis area. The first column is labeled 'geolocation Index' and the other three are labeled 'PhysicalLine1', 'PhysicalLine2', and 'PhysicalLine3'.

#	Abc	Abc	Abc
geolocation Index	geolocation PhysicalLine1	geolocation PhysicalLine2	geolocation PhysicalLine3



# Accessing via Tableau

- Start Analysis



Questions?

