

# Use R to Control Google BigQuery

Chunhsiang Chang

2017-10-17

# Get Authentication

```
install.packages("bigQueryR")  
library(bigQueryR)  
bqr_auth()
```

# Pulling Data from BigQuery to R

- 1 Data sent to R should be less than 10000 rows
- 2 Check the data type limitation for big integers

```
pull_data = bqr_query(projectId = "ntufbdata",  
                      datasetId = "politician_info",  
                      query = "SELECT page_name,  
                                PC1_std AS ideology_score,  
                                party  
                                FROM politician_pca  
                                WHERE PC1_std > 0 AND page_name IS NOT null")  
length(which(pull_data$party=="Republican"))  
length(which(pull_data$party=="Uncertain"))
```

# Save to Another Table

```
job_1 = bqr_query_async(projectId = "ntufbdata",  
                        datasetId = "politician_info",  
                        query = "SELECT PC1_std,  
                                page_id  
                                FROM politician_pca",  
                        destinationTableId = "politician_PC1_std")
```

check the result of your query

```
bqr_get_job("ntufbdata", job_1$jobReference$jobId)
```

# Using For Loop

```
ideology_vec = c(-3, -2, -1, 0, 1, 2)
for (i in 1:5) {
  job = bqr_query_async(projectId = "ntufbdata",
                        datasetId = "bigQueryR_try",
                        query = paste0("SELECT PC1_std,
                                         page_id
                                         FROM politician_PC1_std_copy
                                         WHERE PC1_std BETWEEN",
                                         ideology_vec[i-1],
                                         "AND",
                                         ideology_vec[i]),
                        destinationTableId = paste0("politician_group", i-1),
                        writeDisposition = "WRITE_TRUNCATE")

  bqr_wait_for_job(job)
}
```

# Export to Google Cloud Storage

- 1 Must use a Google Cloud Storage path
- 2 This overwrites table of same name automatically

```
for (i in 1:5) {  
  job_extract = bq_extract_data(projectId = "ntufbdata",  
                                datasetId = "bigQueryR_try",  
                                tableId = paste0("politician_group", i-1),  
                                cloudStorageBucket =  
                                  paste0("gs://ntuusfb/bigQueryR_try/politician_group", i-1, ".csv"))  
  bq_wait_for_job(job_extract)  
}
```