

# SM 727 HW4

Jay Kim

[Github](#)

```
project <- "surv-727-test-476319"
```

```
con <- dbConnect(  
  bigrquery::bigrquery(),  
  project = "bigquery-public-data",  
  dataset = "chicago_crime",  
  billing = project  
)
```

```
dbListTables(con)
```

```
[1] "crime"
```

```
SELECT count(primary_type) AS primary_count, count(*) AS overall_count  
FROM crime  
WHERE year = 2015  
LIMIT 10;
```

Table 1: 1 records

primary_count	overall_count
264874	264874

```
SELECT count(*) AS arrest_count, primary_type AS type  
FROM `bigquery-public-data.chicago_crime.crime`  
WHERE year = 2016 and arrest = true  
GROUP BY primary_type  
ORDER BY arrest_count DESC  
LIMIT 10;
```

Table 2: Displaying records 1 - 10

arrest_count	type
13327	NARCOTICS
10334	BATTERY
6522	THEFT
3724	CRIMINAL TRESPASS
3494	ASSAULT
3416	OTHER OFFENSE
2510	WEAPONS VIOLATION
1669	CRIMINAL DAMAGE
1116	PUBLIC PEACE VIOLATION
1098	MOTOR VEHICLE THEFT

### Time of day associated with most arrests

Hours 18-20 are associated with the most arrests.

```
SELECT
    EXTRACT(HOUR from date) AS hourofday,
    COUNT(*) AS arrest_count
FROM `bigquery-public-data.chicago_crime.crime`
WHERE year = 2016
    AND arrest = true
GROUP BY EXTRACT(HOUR from date)
ORDER BY arrest_count DESC
LIMIT 10;
```

Table 3: Displaying records 1 - 10

hourofday	arrest_count
19	3843
18	3482
20	3303
21	2962
16	2933
22	2896
11	2894
17	2821
12	2788

hourofday	arrest_count
14	2775

```
SELECT COUNT(*) AS arrest_count, year
FROM `bigquery-public-data.chicago_crime.crime`
WHERE arrest = true
    AND primary_type = 'HOMICIDE'
GROUP BY year
ORDER BY arrest_count DESC
LIMIT 10;
```

Table 4: Displaying records 1 - 10

arrest_count	year
431	2001
428	2002
386	2003
356	2020
321	2022
296	2021
294	2004
292	2016
288	2008
284	2005

## Highest Arrests District 2015/2016

District 11 has the highest numbers of arrest in 2015 and 2016.

```
SELECT COUNT(*) AS arrest_count, year, district
FROM `bigquery-public-data.chicago_crime.crime`
WHERE arrest = true
    AND (year = 2015 OR year = 2016)
GROUP BY district, year
ORDER BY arrest_count DESC
LIMIT 10;
```

Table 5: Displaying records 1 - 10

arrest_count	year	district
8975	2015	11
6578	2016	11
5549	2015	7
4514	2015	15
4476	2015	6
4451	2015	25
4326	2015	4
4115	2015	8
3656	2016	7
3628	2015	10

```

query <- "
SELECT COUNT(*) AS arrest_count
FROM `bigquery-public-data.chicago_crime.crime`
WHERE arrest = true
    AND year = 2016
    AND district = 11
GROUP BY primary_type
ORDER BY arrest_count DESC
"

result <- DBI::dbGetQuery(con, query)
print(result)

```

```

# A tibble: 27 x 1
  arrest_count
  <int>
1     3634
2      635
3      511
4      303
5      255
6      207
7      205
8      135
9      119
10     106
# i 17 more rows

```

```

crime <- tbl(con, "crime")

arrests_by_district <- crime %>%
  filter(arrest == TRUE, year == 2016, district == 11) %>%
  group_by(primary_type, year) %>%
  summarise(arrest_count = n(), .groups = "drop") %>%
  arrange(year)

head(arrests_by_district, 10)

# Source:      SQL [?? x 3]
# Database:    BigQueryConnection
# Ordered by:  year
# 
#   primary_type      year arrest_count
#   <chr>            <int>        <int>
# 1 INTIMIDATION    2016          2
# 2 CRIMINAL TRESPASS 2016        205
# 3 CRIMINAL SEXUAL ASSAULT 2016         3
# 4 OTHER OFFENSE   2016        255
# 5 PROSTITUTION    2016        511
# 6 NARCOTICS       2016       3634
# 7 OFFENSE INVOLVING CHILDREN 2016         27
# 8 PUBLIC INDECENCY 2016          2
# 9 BURGLARY        2016         22
# 10 THEFT          2016         98

dbDisconnect(con)

```