**SQL-Mongo Project – Spatial Data of US Wildfires**

BUAN 6320

Done by:

Mahalakshmi Sundaramurthy

|  |  |
| --- | --- |
| **Activity** | **Member 1** |
| Prepared Data Model and Created Physical DB | x |
| Loaded Data into Database | x |
| Wrote SQL Queries | x |
| Prepared Mongo Database | x |
| Loaded data into Mongo DB | x |
| Wrote Mongo Queries | x |
| Prepared Report | x |
| Reviewed Report | x |

Contents

Data Model 5

Assumptions/Notes About Data Entities and Relationships 5

Entity-Relationship Diagram 5

Physical Database 6

Assumptions/Notes About Data Set 6

Screen shot of Physical Database objects 6

Data in the Database 6

SQL Queries 7

Query 1 7

Question 7

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 7

Translation 7

Screen Shot of SQL Query and Results 7

Query 2 8

Question 8

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 8

Translation 8

Screen Shot of SQL Query and Results 8

Query 3 9

Question 9

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 9

Translation 9

Screen Shot of SQL Query and Results 9

Query 4 10

Question 10

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 10

Translation 10

Screen Shot of SQL Query and Results 10

Query 5 11

Question 11

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 11

Translation 11

Screen Shot of SQL Query and Results 11

Query 6 12

Question 12

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 12

Translation 12

Screen Shot of SQL Query and Results 12

Data Review for MongoDB 13

Assumptions/Notes About Data Collections, Attributes and Relationships between Collections 13

Physical Mongo Database 14

Assumptions/Notes About Data Set 14

Screen shot of Physical Database objects (Database, Collections and Attributes) 14

Data in the Database 14

MongoDB Queries/Code 15

Query 1 15

Question 15

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 15

Translation 15

Screen Shot of MongoDB Query/Code and Results 15

Query 2 16

Question 16

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 16

Translation 16

Screen Shot of MongoDB Query/Code and Results 16

Query 3 17

Question 17

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 17

Translation 17

Screen Shot of MongoDB Query/Code and Results 17

Query 4 18

Question 18

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 18

Translation 18

Screen Shot of MongoDB Query/Code and Results 18

Query 5 19

Question 19

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 19

Translation 19

Screen Shot of MongoDB Query/Code and Results 19

Query 6 20

Question 20

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 20

Translation 20

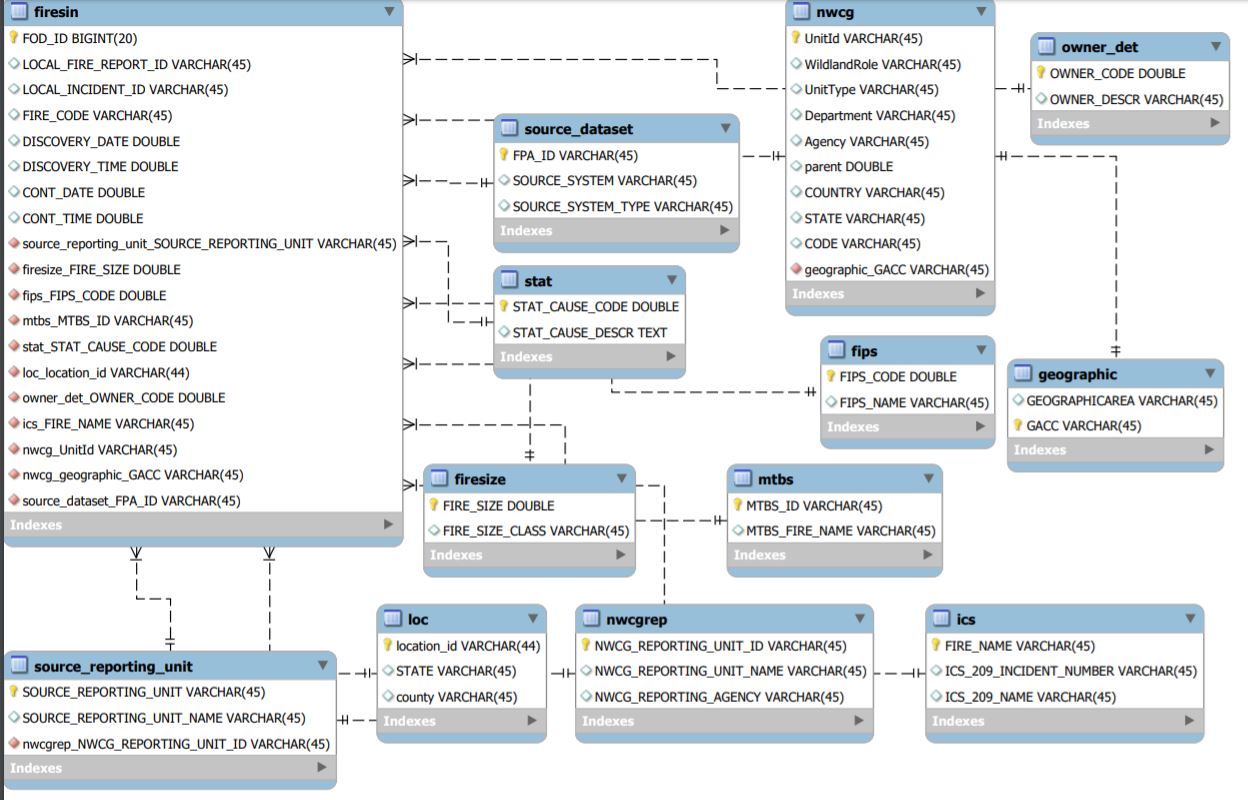
Screen Shot of MongoDB Query/Code and Results 20

# Data Model

## Assumptions/Notes About Data Entities and Relationships

* There are many non-identifying relationships between fires and other tables as listed below.
* For every record in the tables Nwcg, Owner\_det, Source\_dataset, stat, firesize, fips, mtbs, ics, source\_reporting\_unit and Loc, there are one to many records in the firesin table.
* The values of latitude and longitude have been concatenated to location id which is used as a primary key in loc table.
* DISCOVERY\_DOY and CONT\_DOY have been created in firesin as primary key.
* A new table called geographic has been created to take it into account to ensure 3nf.
* Any column which might not follow 3nf has been taken out made into separate table.
* All non-key columns depend on key columns, thus no transitive dependency.
* Primary key is present in all tables.
* All values are single celled.

## Entity-Relationship Diagram

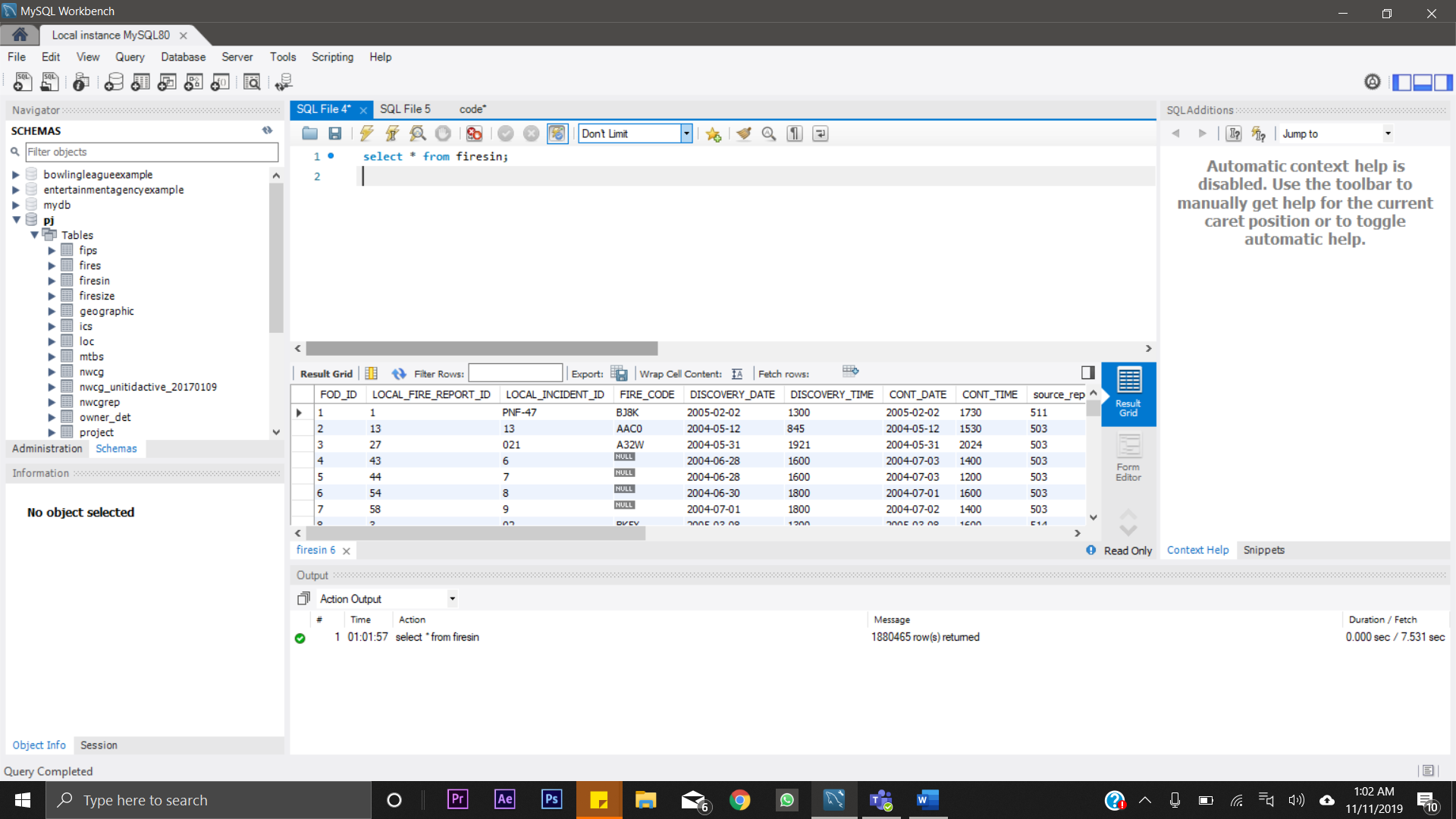


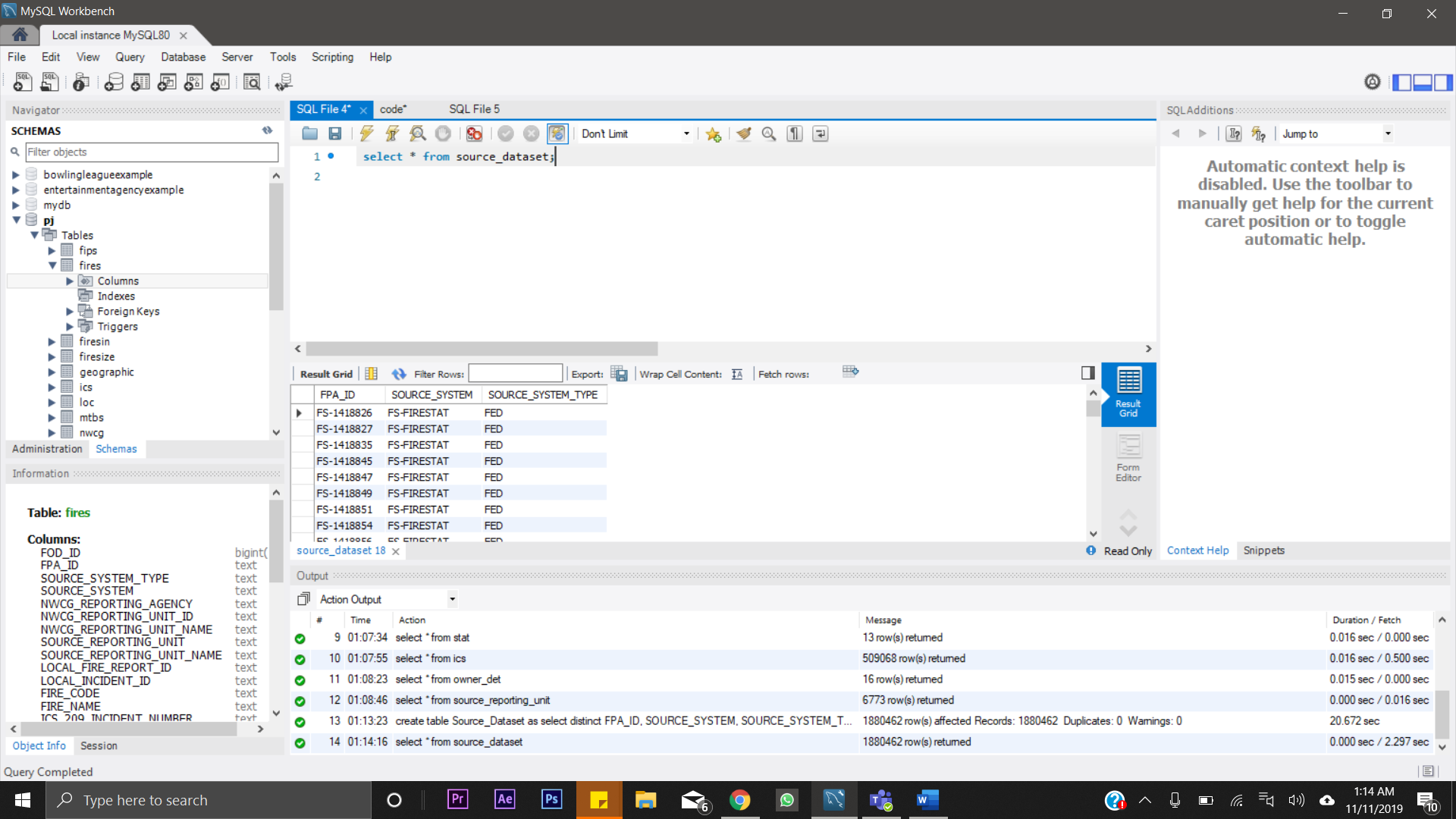
# Physical Database

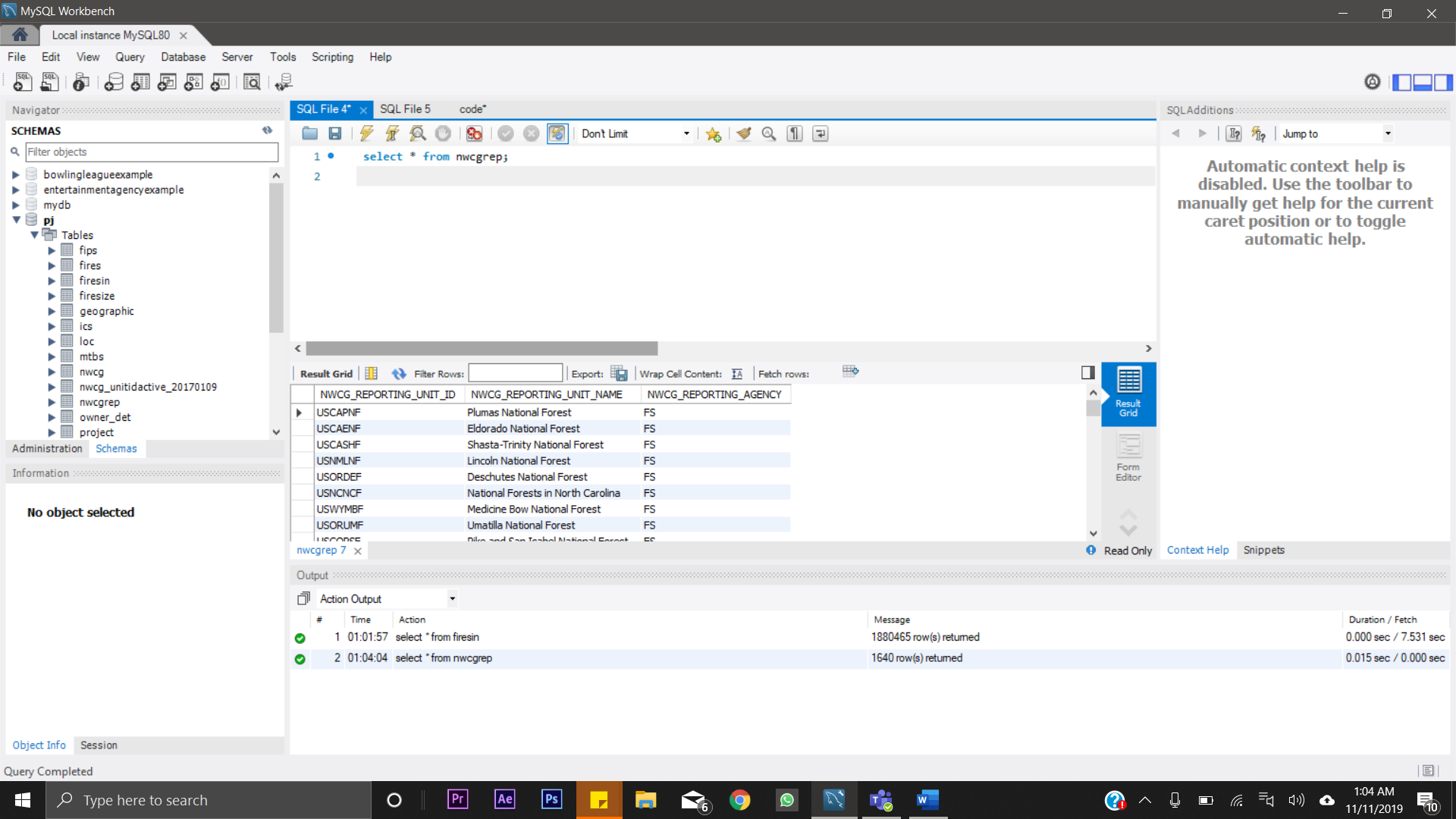
## Assumptions/Notes About Data Set

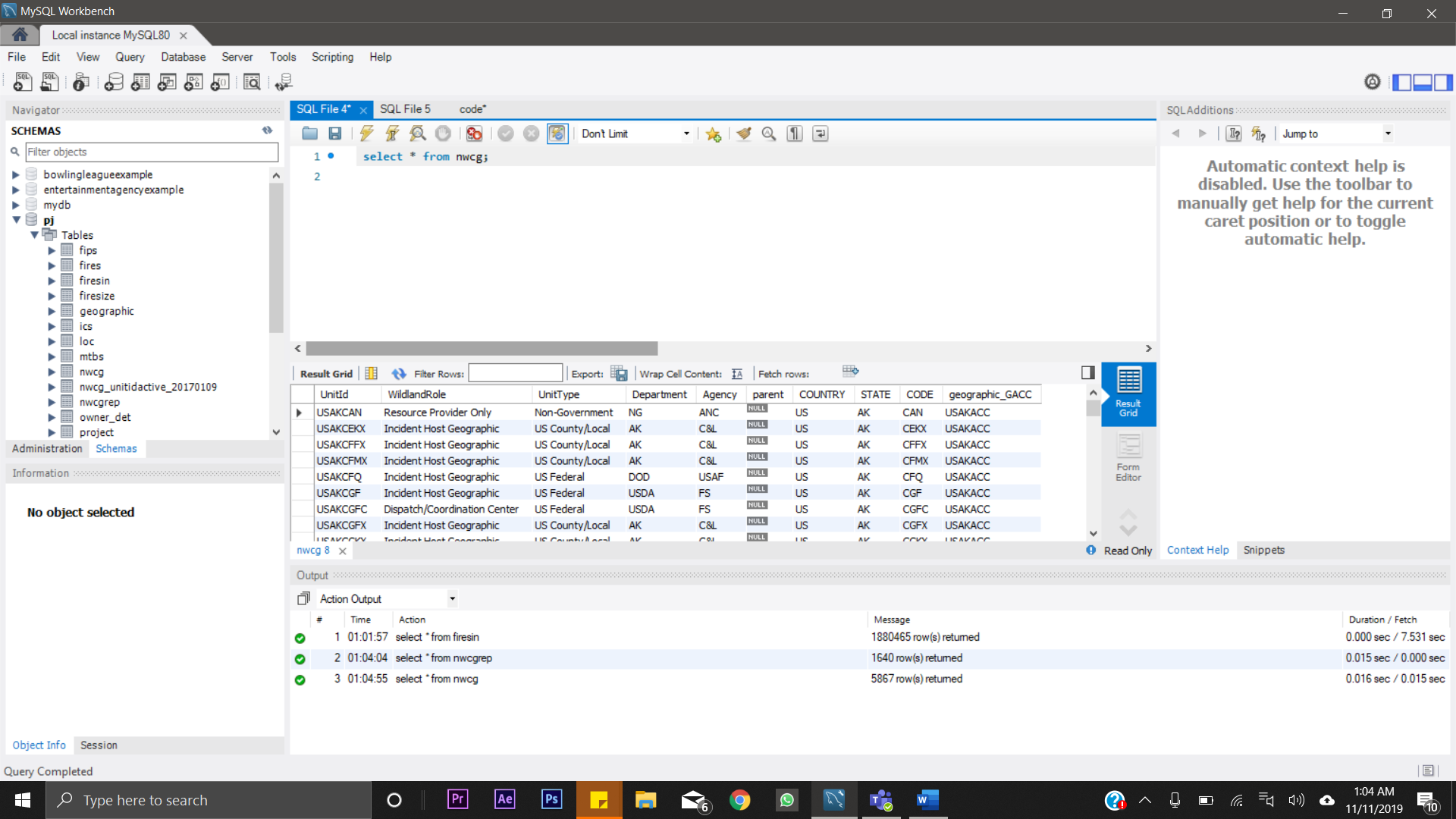
* Null values may be present in tables.

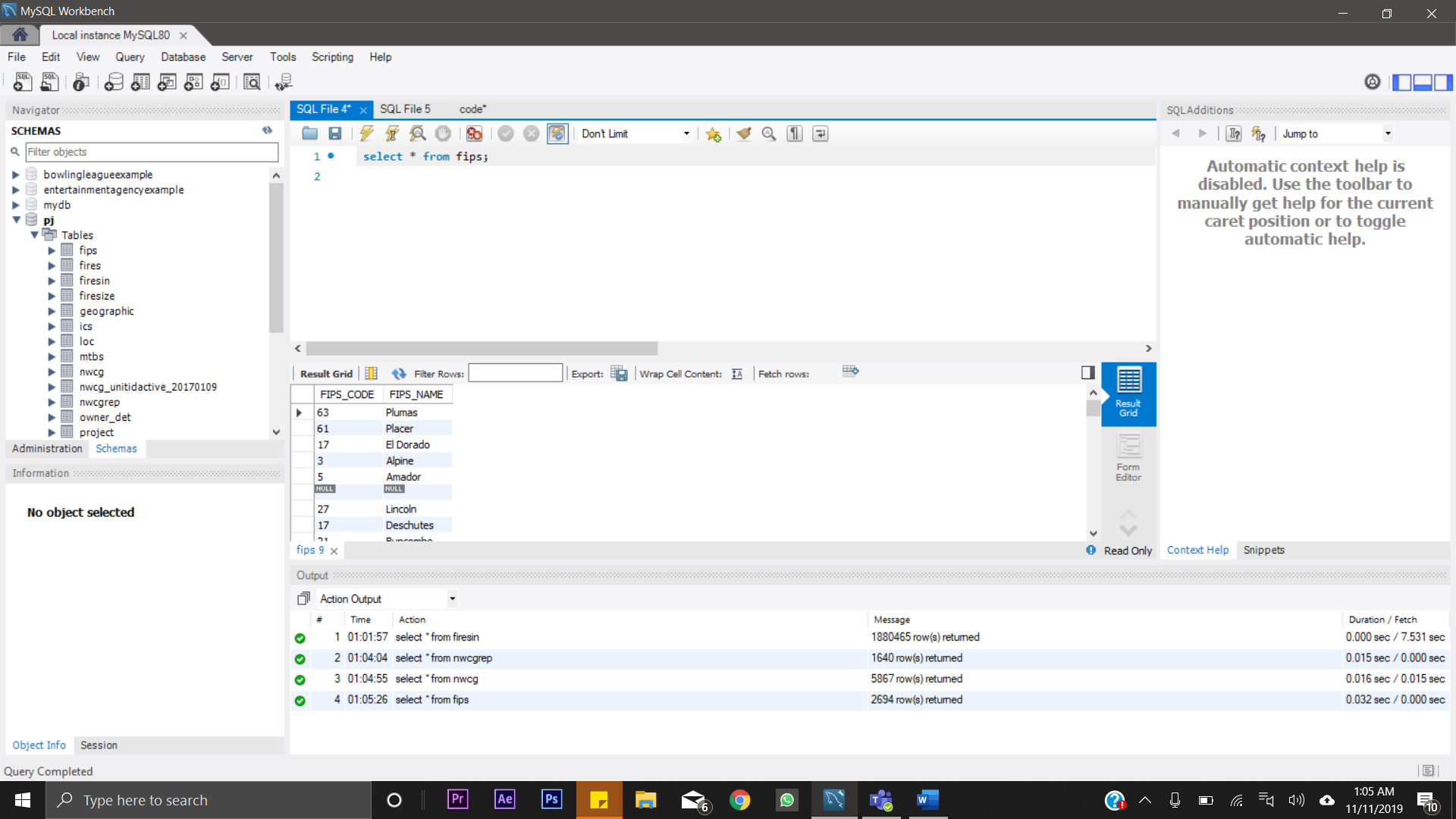
## Screen shot of Physical Database objects

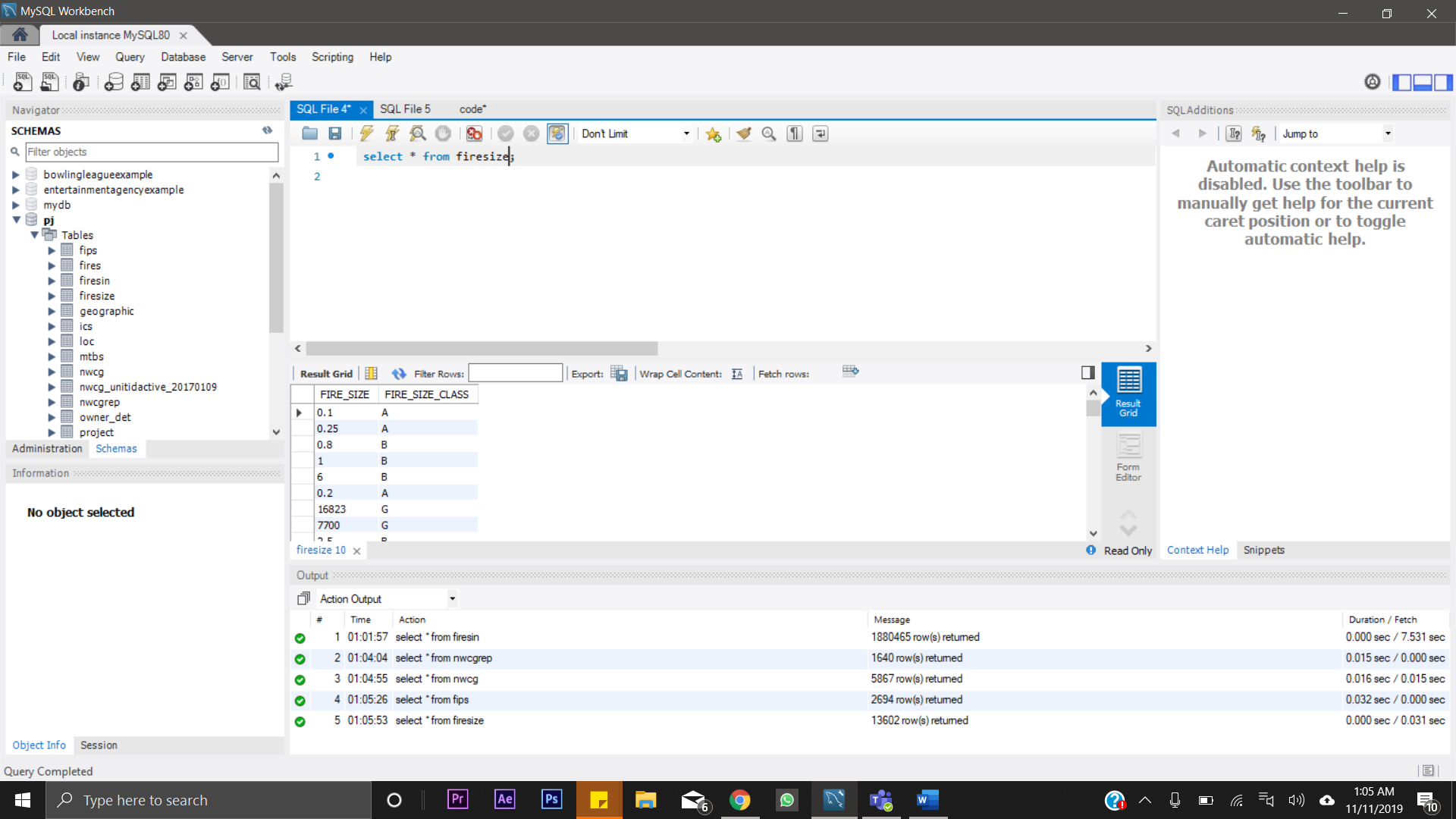








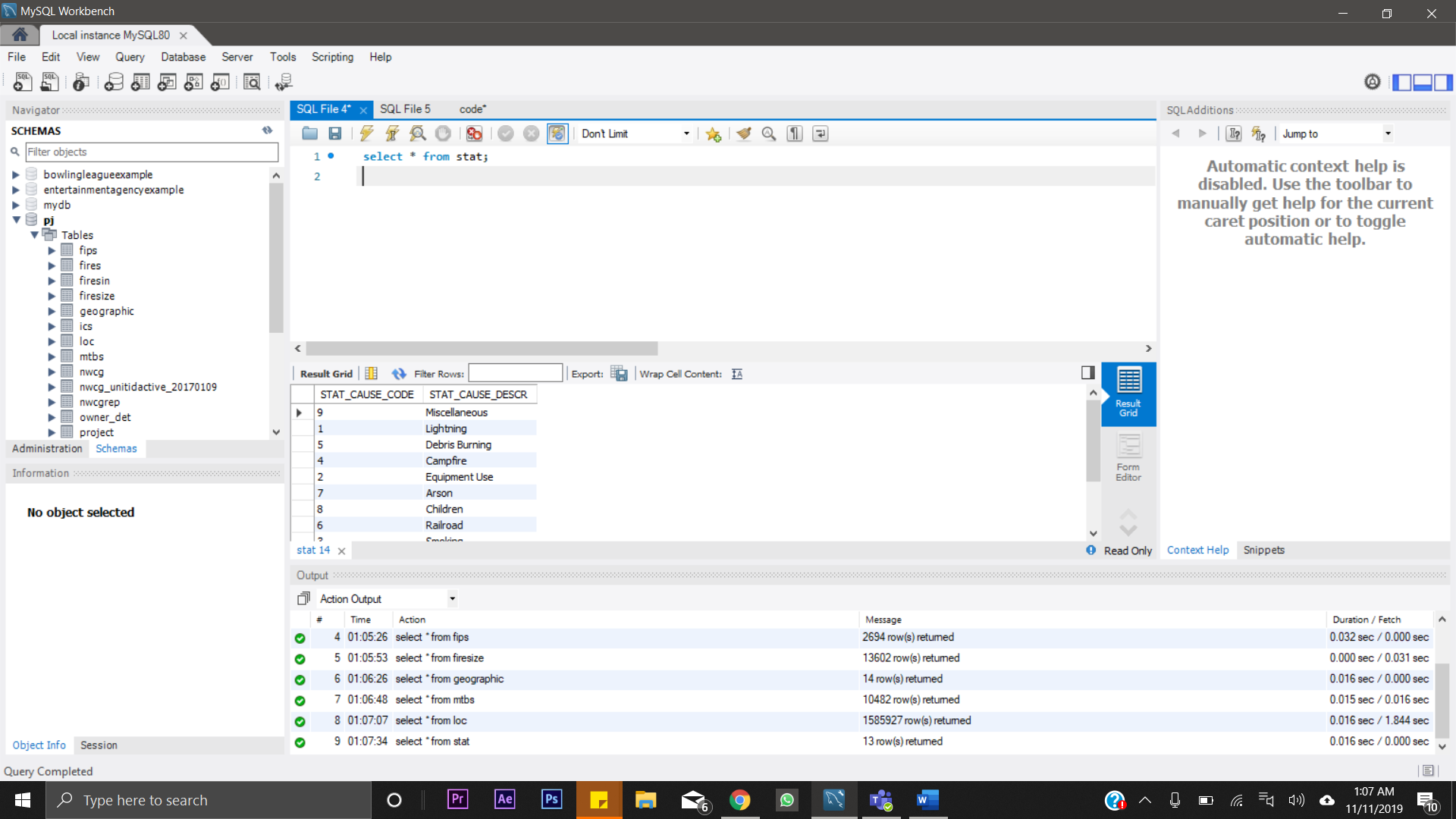


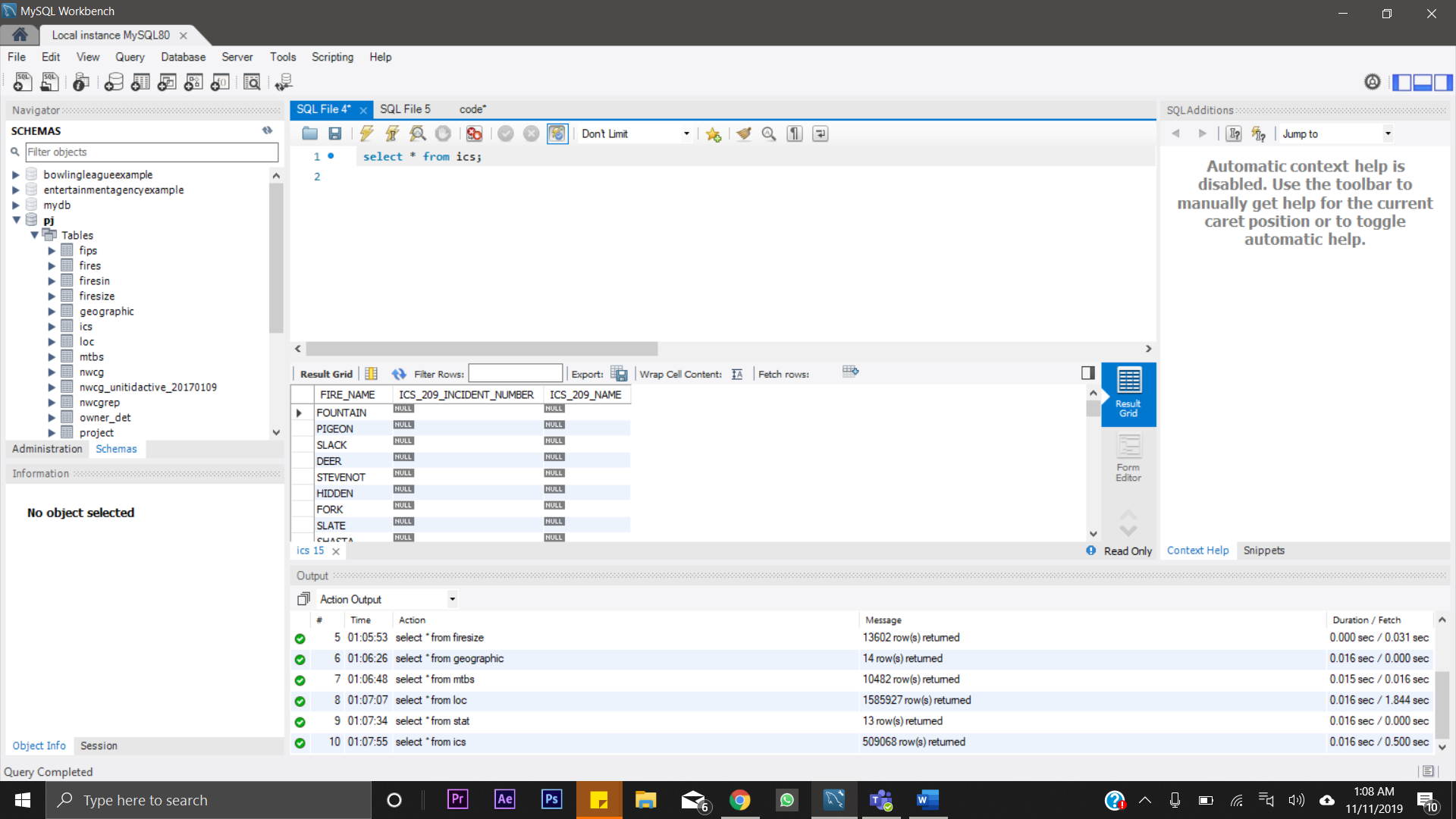


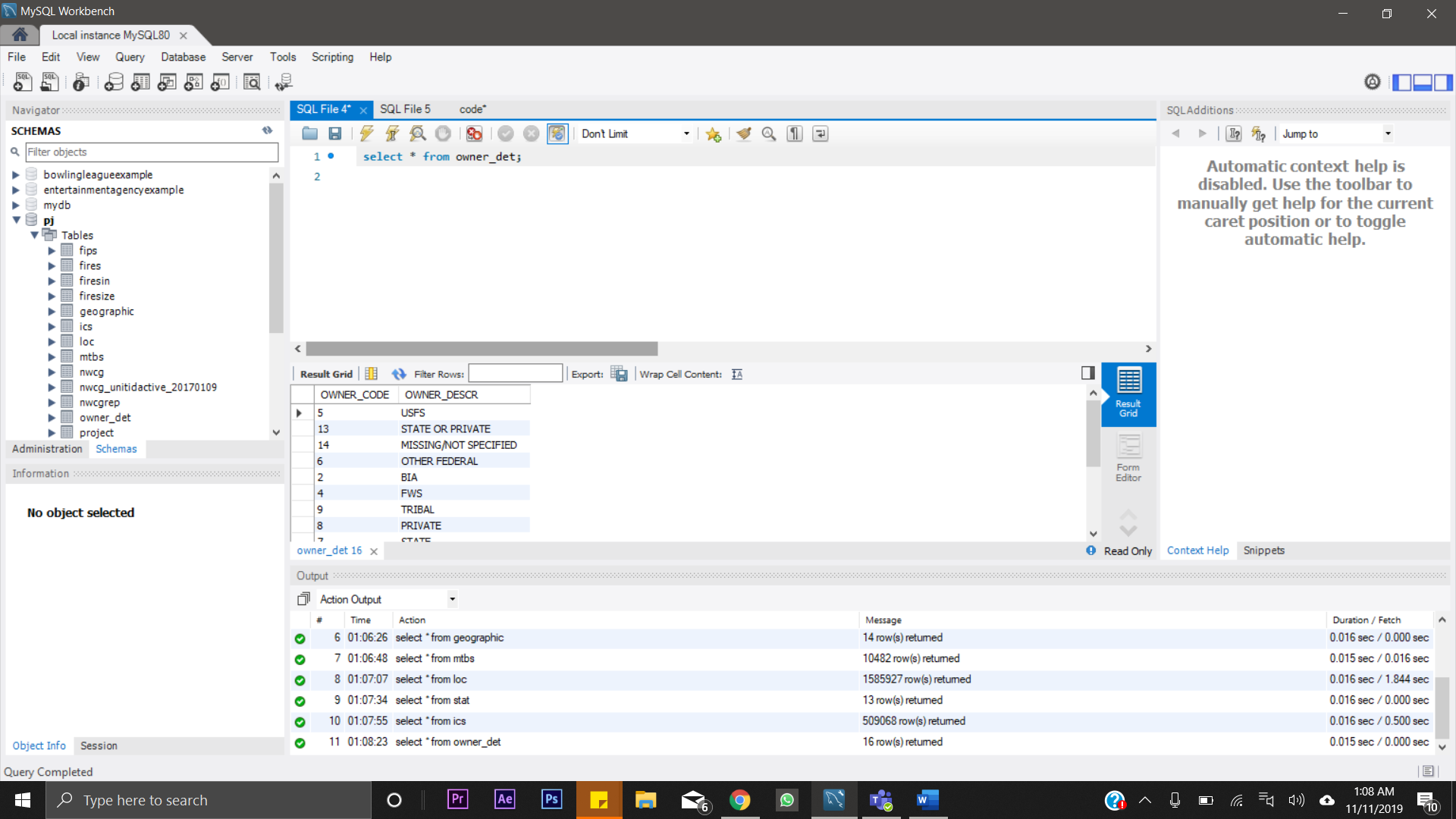


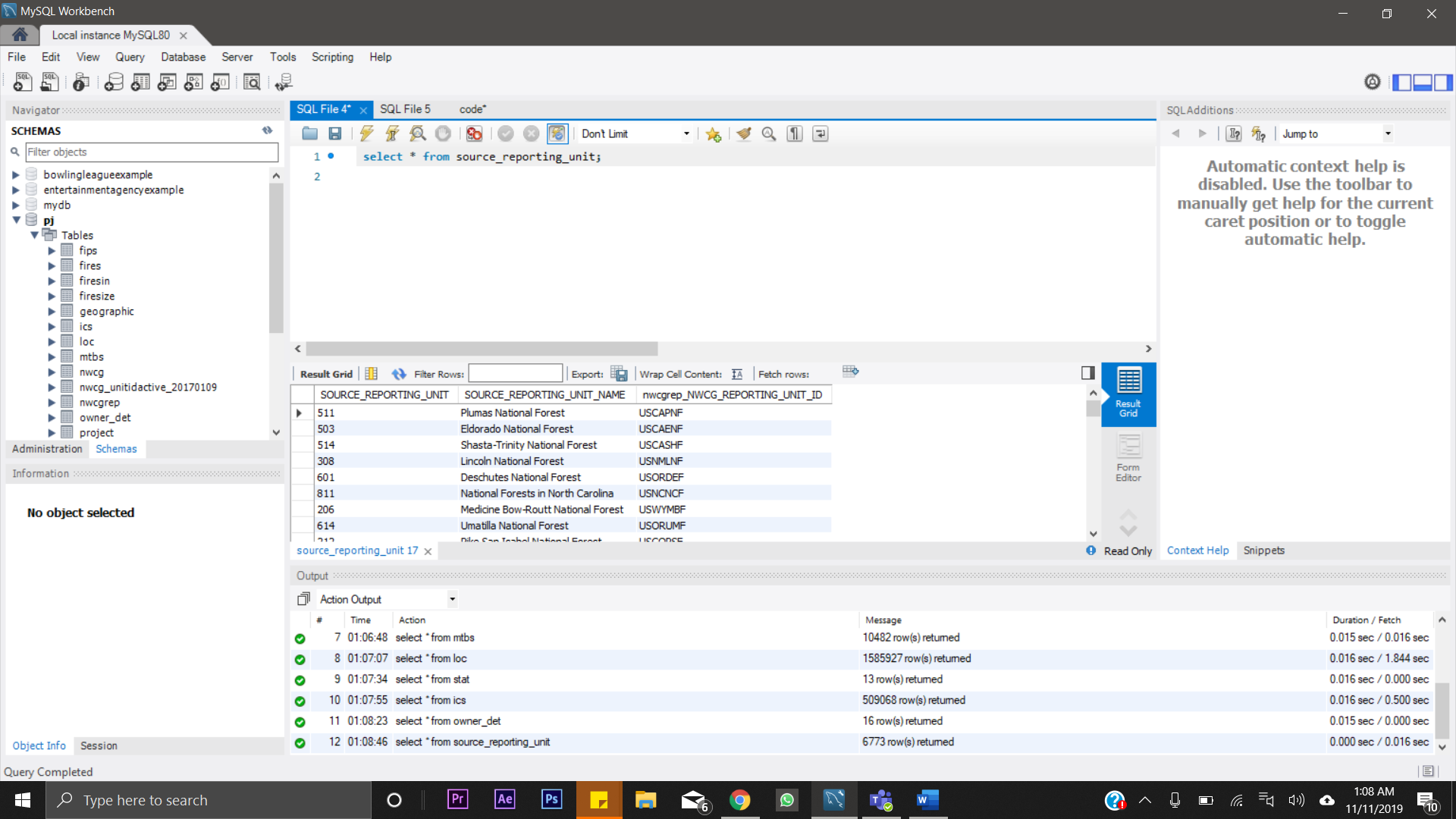












## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| firesin | FOD\_ID | SOURCE\_REPORTING\_UNIT  FIRE\_SIZE  FIPS\_CODE  MTBS\_ID  STAT\_CAUSE\_CODE  LOCATION\_ID  OWNER\_CODE  FIRE\_NAME  UNITID  GACC FPA\_ID | 18,80,465 |
| source\_dataset | FPA\_ID |  | 18,80,462 |
| stat | STAT\_CAUSE\_CODE |  | 13 |
| nwcg | UNITID | GACC | 5867 |
| owner\_det | OWNER\_CODE |  | 16 |
| fips | FIPS\_CODE |  | 2694 |
| geographic | GACC |  | 14 |
| firesize | FIRE\_SIZE |  | 13,602 |
| mtbs | MTBS\_ID |  | 10,482 |
| source\_reporting\_unit | SOURCE\_REPORTING\_UNIT |  | 6773 |
| loc | LOCATION\_ID |  | 15,85,927 |
| nwcgrep | NWCG\_REPORTING\_UNIT\_ID |  | 1640 |
| ics | FIRE\_NAME |  | 5,09068 |

# SQL Queries

## Query 1

### Question

A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the best chance to win a share of the fund?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Number of rows: 2

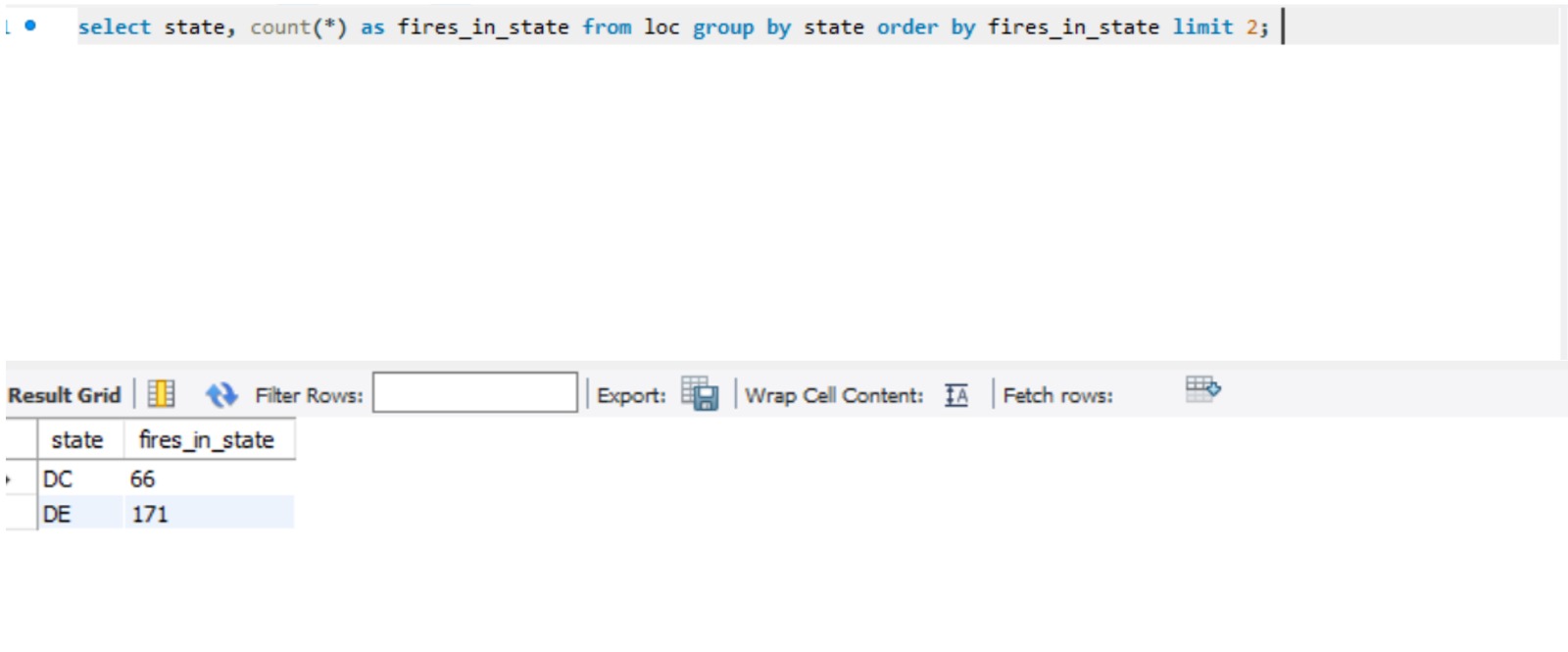
### Translation:

Select top 2 states and the number of occurrences of fires in each such states from the loc table while ordering it ascendingly.

Clean Up:

Select state and count of states from loc ordering by fires inside the state, limit 2

### Screen Shot of SQL Query and Results



## Query 2

### Question

One advocacy group says that cause of most wildfires is unknown. Write a query that can help determine the truth of this statement.

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Number of rows: 13

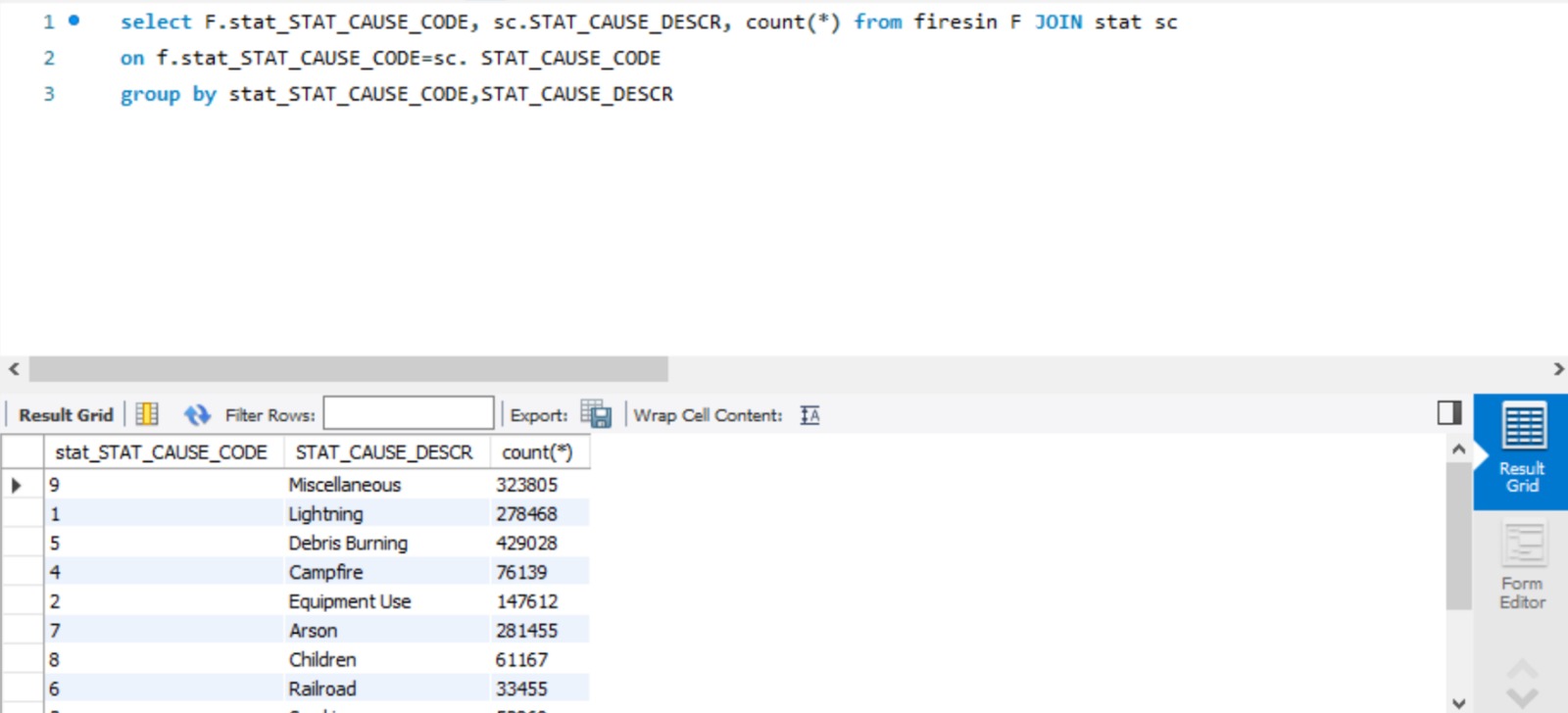
### Translation:

Select the cause code, it’s description and the number of occurrences from firesin and stat tables while joining the code for each cause

Clean Up:

Select Cause code, description and count from the firesin join stat on code

### Screen Shot of SQL Query and Results



## Query 3

### Question

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 4 children in a group visiting a forest. Name top 5 forests where this would be the most appropriate.

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Number of rows: 5

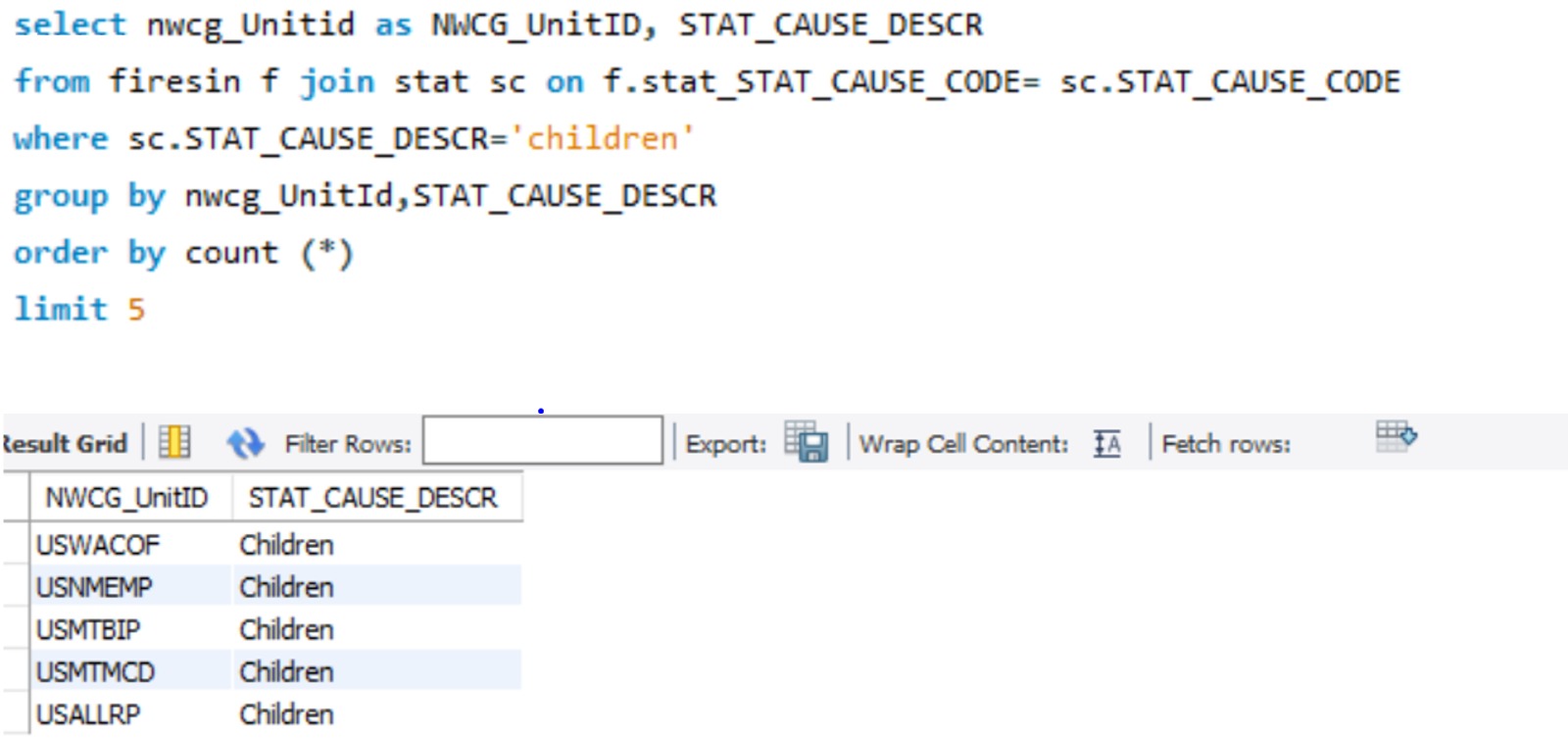
### Translation:

Select the top 5 Nwcg unit ID and description from the firesin and stat on code where cause is the children while ordering by the number of occurrences

Clean up:

Select nwcg unit id and description from firesin and stat on code where cause is children, order by count and limit 5

### Screen Shot of SQL Query and Results



## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

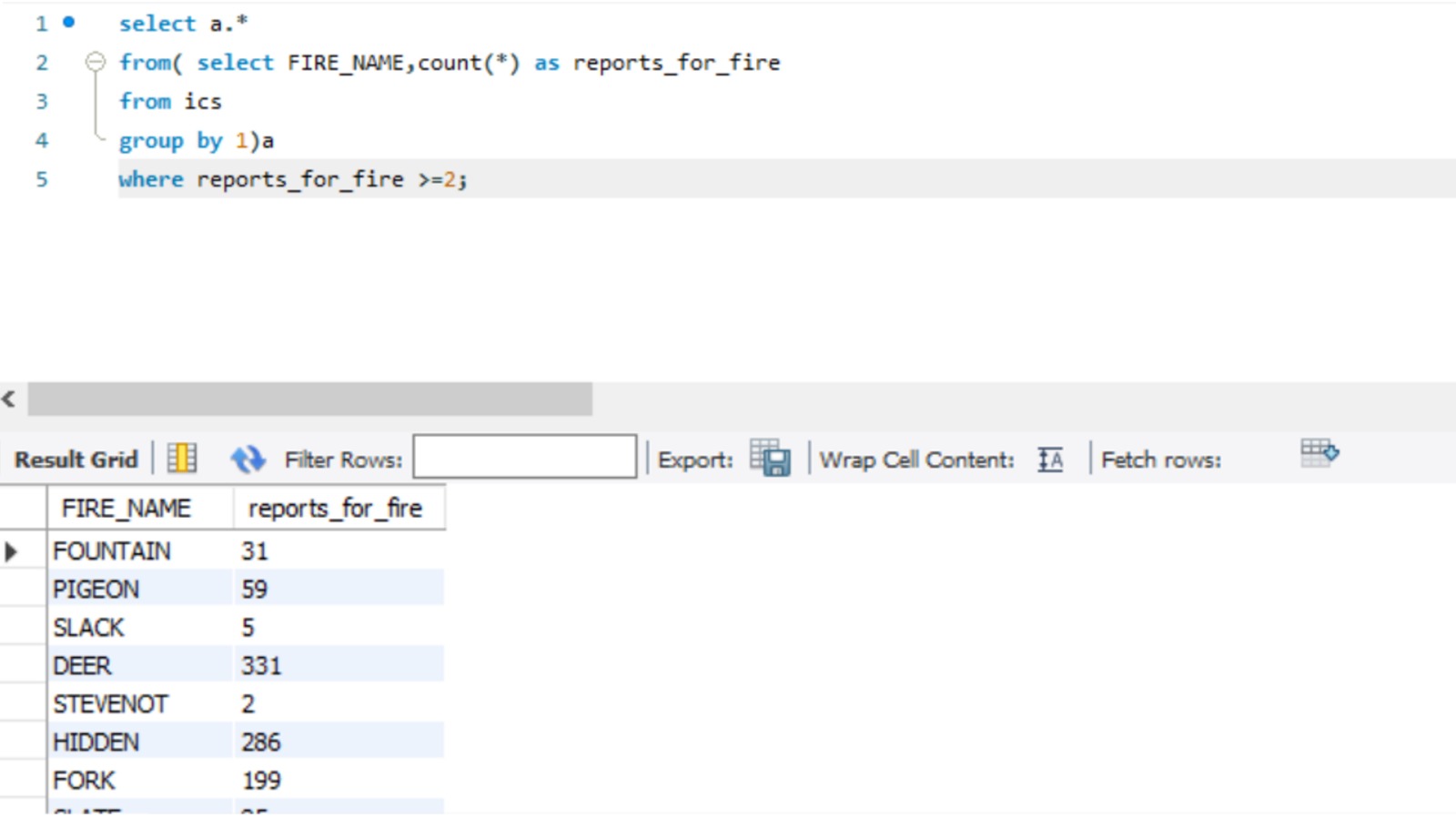
### Translation:

Get the count of rows from each fire\_name table which indicates when a fire accident happend . This count indicates the units which have reported that incident. Filter this result to get those which have more than 2.

Cleanup:

Select all rows from ics where reports\_for\_fire >=2

### Screen Shot of SQL Query and Results



### Query 5

### Question

What were the forests that had only one fire that lasted more than two days?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Number of rows: 416

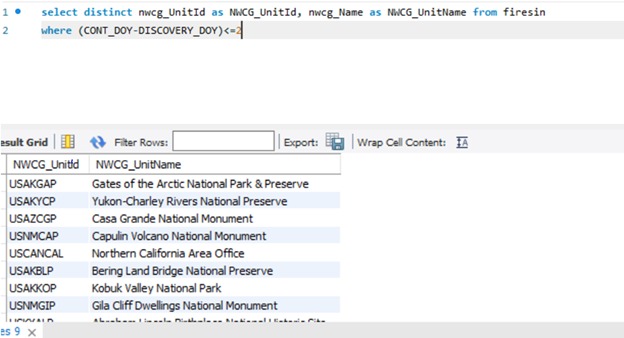
### Translation:

Select the distinct NWCG UnitId and NWCG Unit name from firesin table where the fire lasted more than two days

Cleanup:

Select distinct nwcg\_unitid and nwcg\_name from firesin where count\_coy-DISCOVERY<=2

### Screen Shot of SQL Query and Results



## Query 6

### Question

Which state had fires only in the second half of the calendar years?

### Notes/Comments About SQL Query and Results (Include # of Rows in Result)

Number of rows: 0

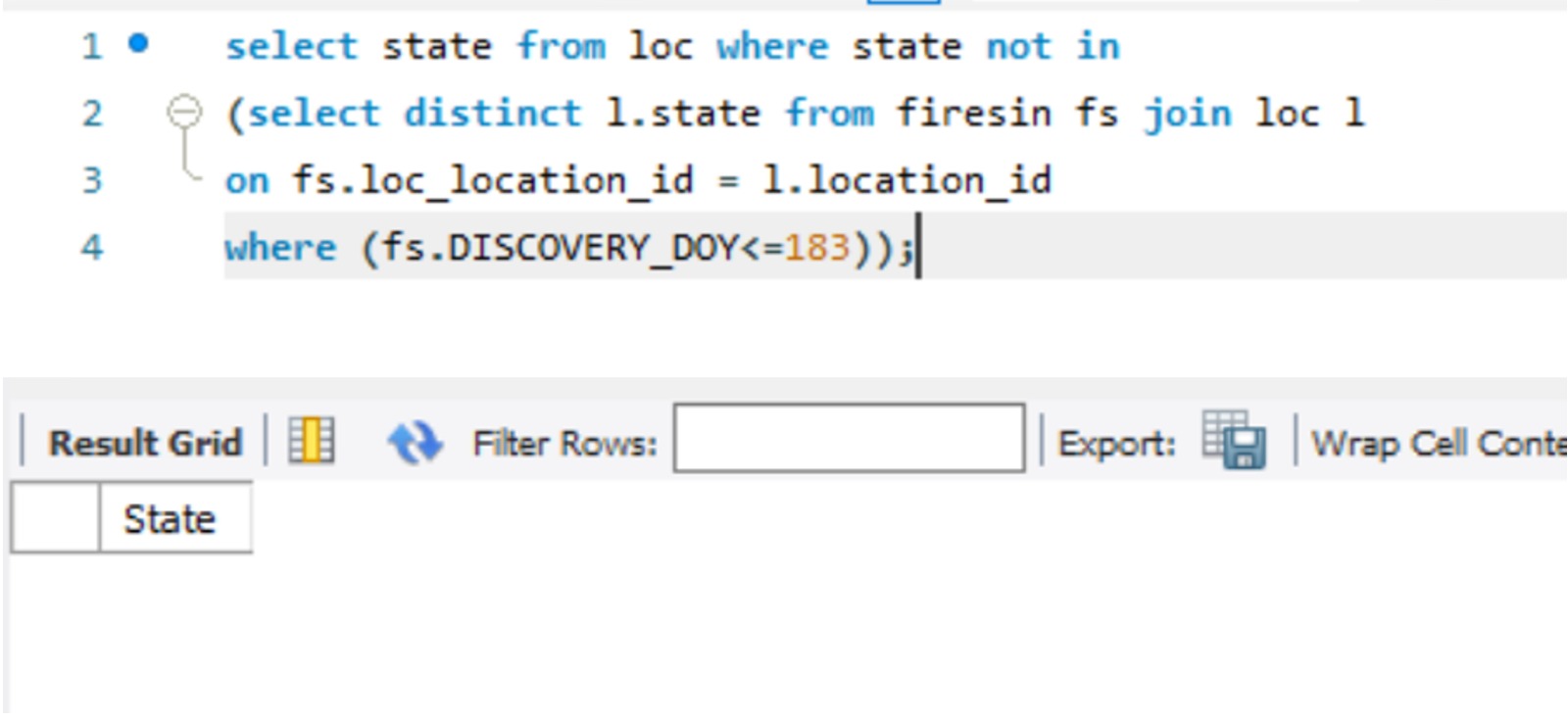
### Translation:

Select the state from the location table where the fires occured only in the second half of the calendar years (last 183 days)

Cleanup:

Select state from loc where DISCOVERY\_DOY<=183

### Screen Shot of SQL Query and Results



# Data Review for MongoDB

## Assumptions/Notes About Data Collections, Attributes and Relationships between Collections

* Row structure and normalization is not required in MongoDB as they are not enforced due to document based nature of MongoDB.

# Physical Mongo Database

## Assumptions/Notes About Data Set

## Screen shot of Physical Database objects (Database, Collections and Attributes)



## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationships With Other Collections (if any)** | **# of Documents in Collection** |
| Fires |  | 1,880,465 |
| nwcg |  | 5667 |

# MongoDB Queries/Code

## Query 1

### Question

A leading beverage company has announced a billion-dollar fund for removing debris from forests, rivers and mountains in the US. All states are interested. Which 2 states have the best chance to win a share of the fund?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

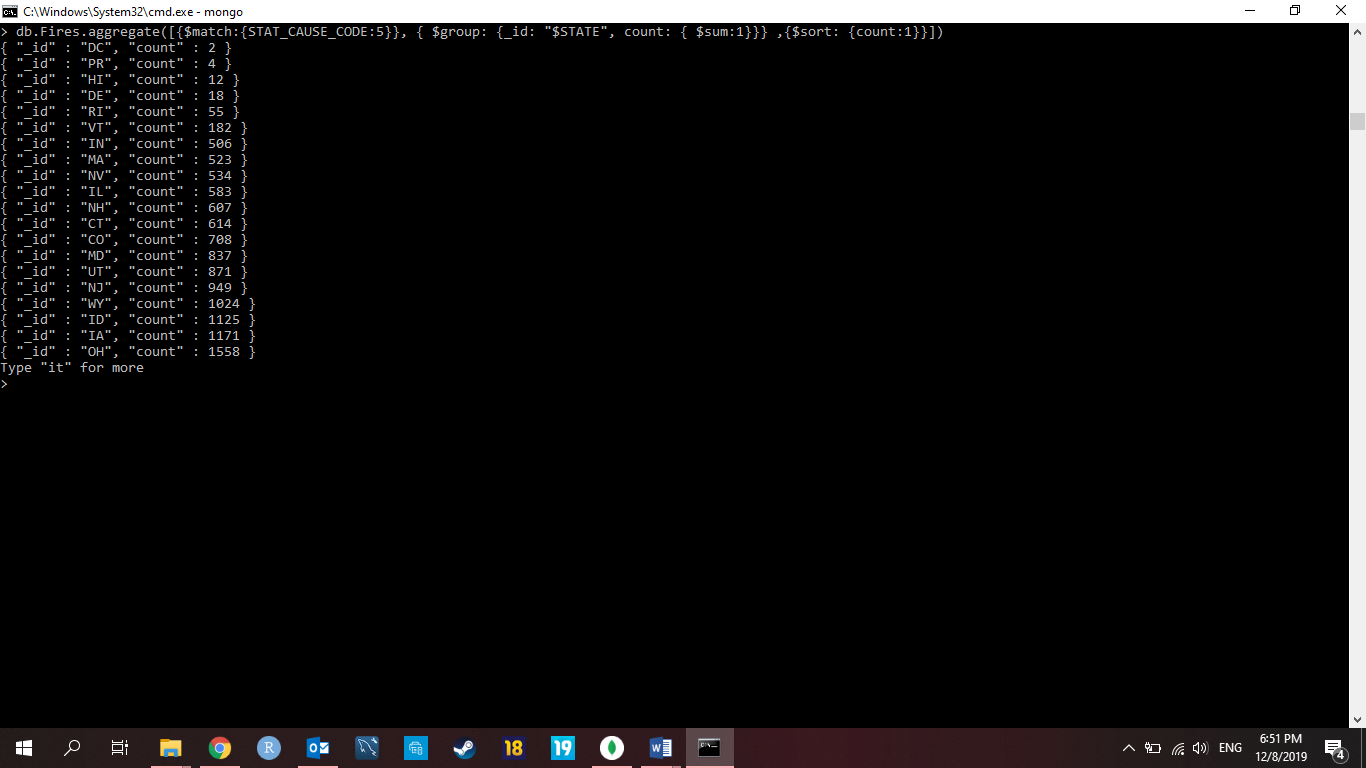
We first used the “MATCH” function to filter the data by Debris followed by the “GROUP” function to find the least number of aggregates for the state which has the least chance to win the share of the fund.

db.Fires.aggregate([{$match:{STAT\_CAUSE\_CODE:5}}, { $group: {\_id: "$STATE", count: { $sum:1}}} ,{$sort: {count:1}}])

### Translation:

Count the number of fires for each for each state using state and order by the count of number of fires to get 2 states with the least chance of winning.

### Screen Shot of MongoDB Query/Code and Results



## Query 2

### Question

One advocacy group says that cause of most wildfires is unknown. Write a query that can help determine the truth of this statement.

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) The group function is used to find the total of incidents for each individual category.

db.Fires.aggregate([{ $group: {\_id: "$STAT\_CAUSE\_DESCR", count: { $sum:1}}}])

Using this, we were able to find the total number of incidents which were caused by “Humans” and “Nature” (STAT\_CAUSE\_DESCR= “Lightning”) respectively.

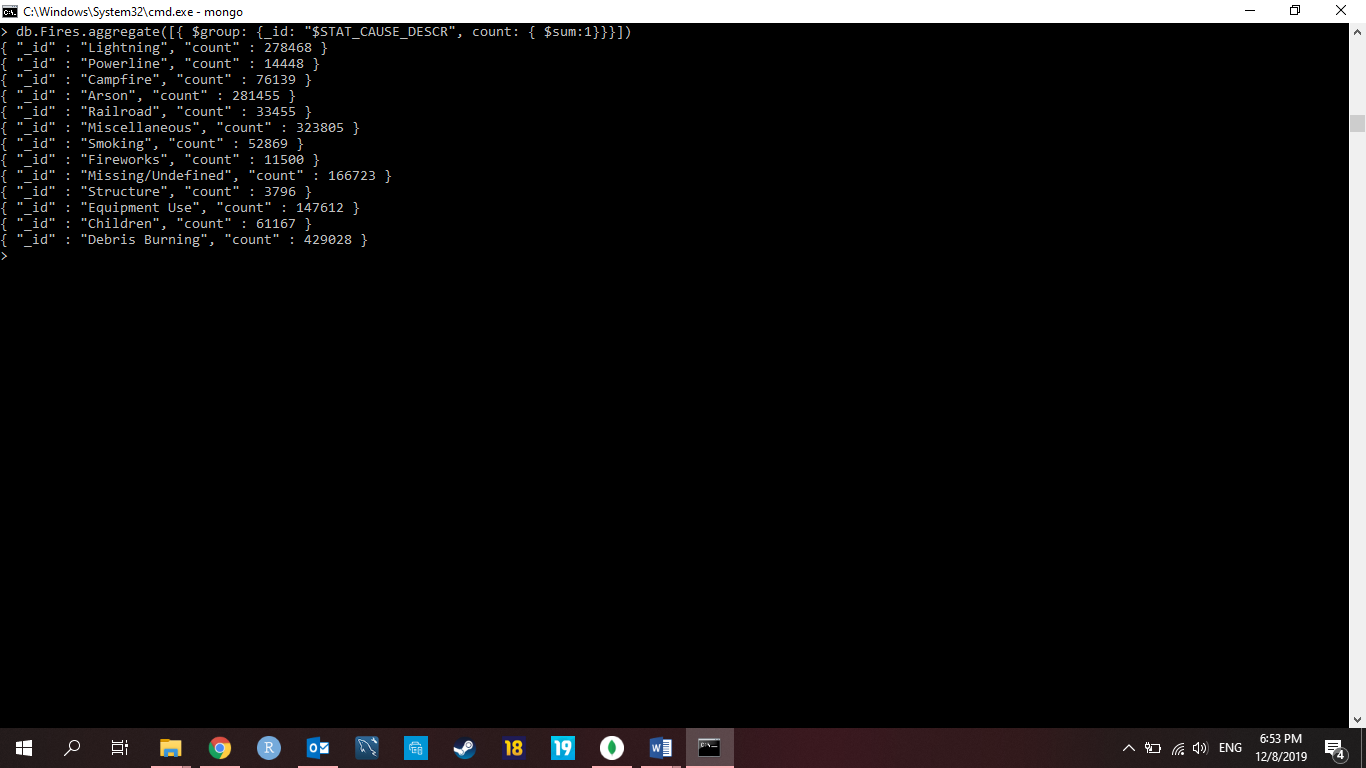
From the results, we infer that the number of wildfires caused by humans is more than the number of wildfires caused by nature.

Total number of documents: 13

### Translation:

Select the statistical cause description and count of number of fires to get the count of fires caused by different from fires and group by cause, followed by ordering by fires.

### Screen Shot of MongoDB Query/Code and Results



## Query 3

### Question

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 4 children in a group visiting a forest. Name top 5 forests where this would be the most appropriate.

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

We grouped the forest name which matches with the children category and order it with minimum and in ascending order. Then we limited the number of outputs to 5.

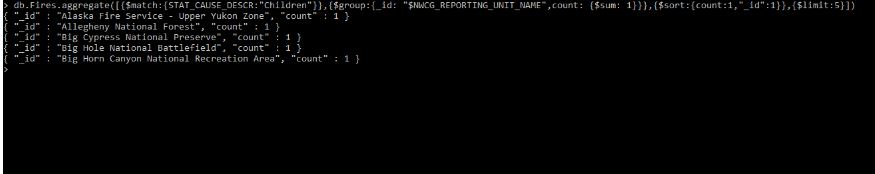
db.Fires.aggregate([{$match:{STAT\_CAUSE\_DESCR:"Children"}},{$group:{\_id: "$NWCG\_REPORTING\_UNIT\_NAME",count: {$sum: 1}}},{$sort:{count:1,"\_id":1}},{$limit:5}])

Total number of documents: 5

### Translation:

Select nwcg reporting unit name and cases due child from fires where cause is children and group them by unit name and limit the result to 5.

### Screen Shot of MongoDB Query/Code and Results



## Query 4

### Question

How many wildfires were reported by at least two units/agencies?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) We have used the “Below” command to get the count of Fires reported by at least 2 agencies.

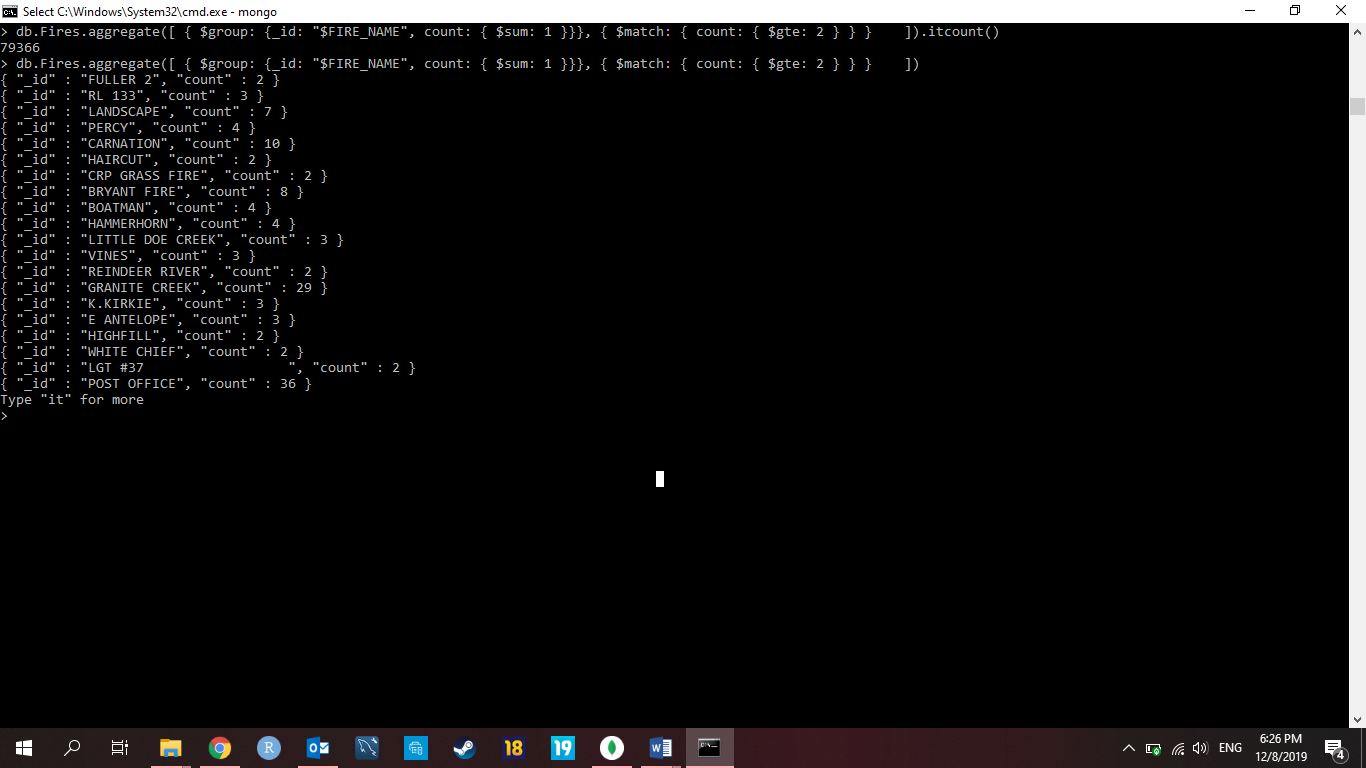
Total number of documents: 79,366

db.Fires.aggregate([ { $group: {\_id: "$FIRE\_NAME", count: { $sum: 1 }}}, { $match: { count: { $gte: 2 } } } ]).itcount()

### Translation:

Get the count of rows from each fire\_name table which indicates when a fire accident happend . This count indicates the units which have reported that incident. Filter this result to get those which have more than 2.

### Screen Shot of MongoDB Query/Code and Results



## Query 5

### Question

What were the forests that had only one fire that lasted more than two days?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

We used the below code to find the number of forest that had only one fire that lasted more than 2 days.

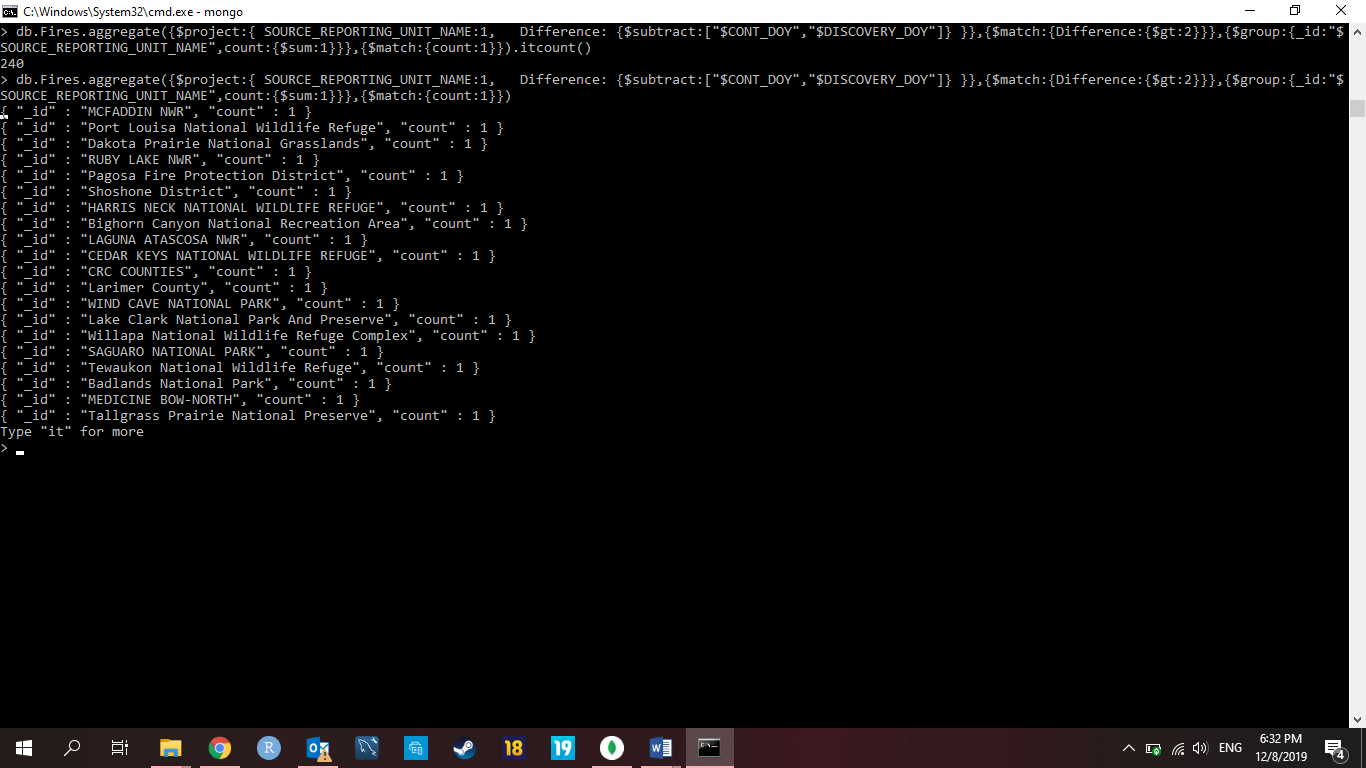
Total number of documents: 240

db.Fires.aggregate({$project:{ SOURCE\_REPORTING\_UNIT\_NAME:1, Difference: {$subtract:["$CONT\_DOY","$DISCOVERY\_DOY"]} }},{$match:{Difference:{$gt:2}}},{$group:{\_id:"$SOURCE\_REPORTING\_UNIT\_NAME",count:{$sum:1}}},{$match:{count:1}}).itcount()

### Translation

For each forest calculate the count of fires that lasted more than 2 days by taking the difference of discovery and control dates then the select only those forests where the count is excatly 1.

### Screen Shot of MongoDB Query/Code and Results



## Query 6

### Question

Which state had fires only in the second half of the calendar years?

### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

We have used the below query to find the number of fires in the second half of the calendar year only.

Total number of documents: 0

db.Fires.aggregate([{$match:{DISCOVERY\_DOY:{$lte:182}}},{$group:{\_id:"$STATE",count:{$sum:1}}},{$match:{DISCOVERY\_DOY:{$gte:182}}},{$group:{\_id:"$STATE",count:{$sum:1}}}]).itcount()

### Translation:

For each state calculate the total fires and fires that happend in the second half of the year then whichever states where total fire count is equal to fire count which happend in second half of the year are the ones to be considered.

### Screen Shot of MongoDB Query/Code and Results

