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

Contents

[Assumptions/Notes About Data Entities and Relationships 3](#_Toc29469709)

[Entity-Relationship Diagram 3](#_Toc29469710)

[Data in the Database 4](#_Toc29469711)

[Query 1 7](#_Toc29469712)

[Question 7 7](#_Toc29469713)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 7](#_Toc29469714)

[Translation 7](#_Toc29469715)

[Cleanup 7](#_Toc29469716)

[Screen Shot of SQL Query and Results 7](#_Toc29469717)

[Query 2 8](#_Toc29469718)

[Question 1 8](#_Toc29469719)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 8](#_Toc29469720)

[Translation 8](#_Toc29469721)

[Screen Shot of SQL Query and Results 8](#_Toc29469722)

[Query 3 9](#_Toc29469723)

[Question 3 9](#_Toc29469724)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 9](#_Toc29469725)

[Translation 9](#_Toc29469726)

[Screen Shot of SQL Query and Results 9](#_Toc29469727)

[Query 4 10](#_Toc29469728)

[Question 6 10](#_Toc29469729)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 10](#_Toc29469730)

[Translation 10](#_Toc29469731)

[Screen Shot of SQL Query and Results 10](#_Toc29469732)

[Query 5 11](#_Toc29469733)

[Question 8 11](#_Toc29469734)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 11](#_Toc29469735)

[Translation 11](#_Toc29469736)

[Screen Shot of SQL Query and Results 11](#_Toc29469737)

[Query 6 12](#_Toc29469738)

[Question 2 12](#_Toc29469739)

[Notes/Comments About SQL Query and Results (Include # of Rows in Result) 12](#_Toc29469740)

[Translation 12](#_Toc29469741)

[Screen Shot of SQL Query and Results 12](#_Toc29469742)

[Assumptions/Notes About Data Collections, Attributes and Relationships between Collections 13](#_Toc29469743)

[Assumptions/Notes About Data Set 14](#_Toc29469744)

[Screen shot of Physical Database objects (Database, Collections and Attributes) 14](#_Toc29469745)

[Data in the Database 14](#_Toc29469746)

[Query 1 15](#_Toc29469747)

[Question 7 15](#_Toc29469748)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 15](#_Toc29469749)

[Translation 15](#_Toc29469750)

[Screen Shot of MongoDB Query/Code and Results 15](#_Toc29469751)

[-Query 2 16](#_Toc29469752)

[Question 1 16](#_Toc29469753)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 16](#_Toc29469754)

[Translation 16](#_Toc29469755)

[Screen Shot of MongoDB Query/Code and Results 16](#_Toc29469756)

[Query 3 17](#_Toc29469757)

[Question 3 17](#_Toc29469758)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 17](#_Toc29469759)

[Translation 17](#_Toc29469760)

[Screen Shot of MongoDB Query/Code and Results 17](#_Toc29469761)

[Query 4 18](#_Toc29469762)

[Question 6 18](#_Toc29469763)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 18](#_Toc29469764)

[Translation 18](#_Toc29469765)

[Screen Shot of MongoDB Query/Code and Results 18](#_Toc29469766)

[Query 5 19](#_Toc29469767)

[Question 8 19](#_Toc29469768)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 19](#_Toc29469769)

[Translation 19](#_Toc29469770)

[Screen Shot of MongoDB Query/Code and Results 19](#_Toc29469771)

[Query 6 20](#_Toc29469772)

[Question 2 20](#_Toc29469773)

[Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 20](#_Toc29469774)

[Translation 20](#_Toc29469775)

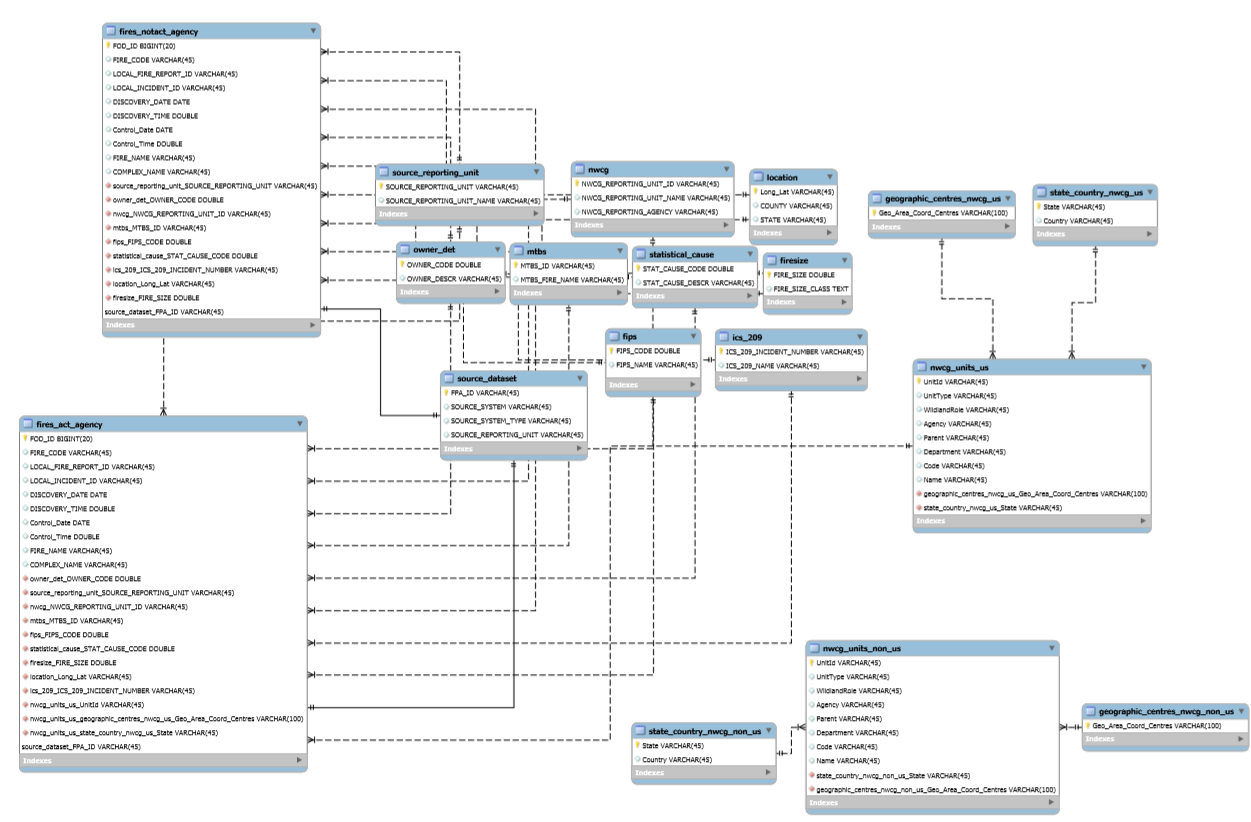
[Screen Shot of MongoDB Query/Code and Results 20](#_Toc29469776)

Data Model

## Assumptions/Notes About Data Entities and Relationships

1. The main fires table has been split into two tables as Fire incidents reported from Active agency(fire\_act\_agency) and Non-Active agency(fire\_notact\_agency). The creation of two separate tables was based on the availability of agency details in NWCG\_UnitIDActive\_20170109 table . On analysis we found USDESFM, USIASFM, USILSFM, USINSFM, USKSSFM, USKSJA409, USMASFM, USMOSFM, USNDSFM, USOHSFM, USVTSFM, USHIHIS to be notactive agencies and rest as active.
2. The NWCG table is also split into two tables based on US and non-US countries from NWCG\_UnitIDActive\_20170109 table
3. The latitude and longitude columns have been concatenated as Long\_lat and taken as a Primary key.
4. The Gacc and GeographicArea have also been concatenated as Geo\_Area\_Coord\_Centres and taken as Primary key.
5. All non-key columns depend only on primary key of the tables.
6. All tables have a primary key

## Entity-Relationship Diagram



Physical Database

## Data in the Database

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Key** | **# of Rows in Table** |
| Fires\_notact\_agency | FOD\_ID | Owner\_det\_OWNER\_CODE, Source\_reporting\_unit\_SOURCE\_REPORTING\_UNIT, Nwcg\_NWCG\_REPORTING\_UNIT\_ID,Mtbs\_MTBS\_ID,  Location\_Long\_Lat,Statistical\_cause\_STAT\_CAUSE\_CODE, Fips\_FIPS\_CODE,Firesize\_FIRE\_SIZE,Ics\_209\_ICS\_209\_INCIDENT\_NUMBER | 1877095 |
| Fires\_act\_agency | FOD\_ID | Owner\_det\_OWNER\_CODE, Source\_reporting\_unit\_SOURCE\_REPORTING\_UNIT, Nwcg\_NWCG\_REPORTING\_UNIT\_ID,Mtbs\_MTBS\_ID,  Location\_Long\_Lat,Statistical\_cause\_STAT\_CAUSE\_CODE, Fips\_FIPS\_CODE,Firesize\_FIRE\_SIZE,Ics\_209\_ICS\_209\_INCIDENT\_NUMBER,nwcg\_units\_us\_UnitId,nwcg\_units\_us\_geographic\_centres\_nwcg\_us\_Geo\_Area\_Coord\_Centres,nwcg\_units\_us\_State\_Country\_nwcg\_us\_State | 3370 |
| Source\_reporting\_unit | SOURCE\_REPORTING\_UNIT |  | 6767 |
| Owner\_det | OWNER\_CODE |  | 16 |
| Source\_dataset | FPA\_ID |  | 1880463 |
| Nwcg | NWCG\_REPORTING\_UNIT\_ID |  | 1640 |
| Mtbs | MTBS\_ID |  | 10482 |
| Location | Long\_Lat |  | 1585927 |
| Statistical\_cause | STAT\_CAUSE\_CODE |  | 13 |
| Fips | FIPS\_CODE |  | 2694 |
| Firesize | FIRE\_SIZE |  | 13602 |
| Ics\_209 | ICS\_209\_INCIDENT\_NUMBER |  | 23314 |
| Geographic\_centres\_nwcg\_us | Geo\_Area\_Coord\_Centres |  | 13 |
| State\_country\_nwcg\_us | State |  | 55 |
| Nwcg\_units\_us | UnitId | Geographic\_centres\_nwcg\_us\_Geo\_Area\_Coord\_Centres,State\_country\_nwcg\_us\_State | 5852 |
| State\_country\_nwcg\_non\_us | State |  | 13 |
| Nwcg\_units\_non\_us | UnitId | Geographic\_centres\_nwcg\_non\_us\_ Geo\_Area\_Coord\_Centres, State\_country\_nwcg\_non\_us\_State | 15 |
| Geographic\_centres\_nwcg\_non\_us | Geo\_Area\_Coord\_Centres |  | 2 |

SQL Queries

## Query 1

### Question 7

Which state had more fires in the first half of a calendar year than the second half of the calendar year?

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

This shows the states that have more fires in the first half of the calendar year compared to the second half of the calendar year (40 rows)

### Translation

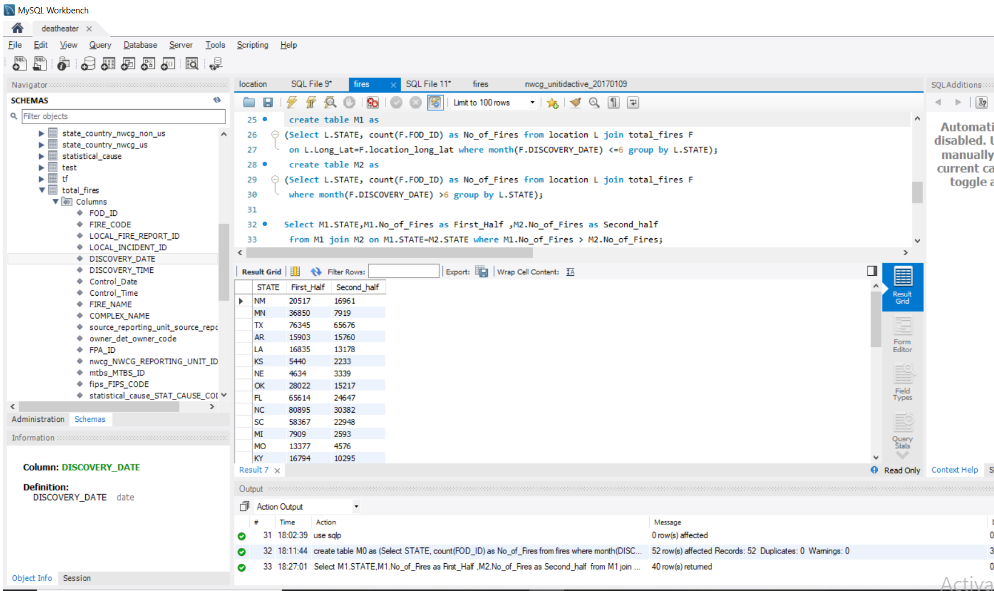
Create two tables M1 and M2 such that it contains the fires incidents and state details for the First half of the calendar year and second half of the calendar year respectively.

Select the state, the number of fires in the first half of the calendar year from M1 table and number of fires in second half of the table from M2 where number of fires in M1 is greater than M2

### Cleanup

* Create table M1 as Select STATE, count of FOD.ID where month is less than equal to 6 and group by STATE
* Create table M2 as Select STATE, count of FOD.ID where month is greater than 6 and group by STATE
* Select M1.STATE , M1.No\_of\_Fires as First\_Half and M2.No\_of\_Fires as Second\_Half from M1 and join M2 on M1.STATE equal to M2.STATE where M1.No\_of\_Fires is greater than M2.No\_of\_Fires

### Screen Shot of SQL Query and Results



## Query 2

### Question 1

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 state has the least chance to win a share of the fund?

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

This shows that state DC has the least chance to win the share of the fund. (1 Row)

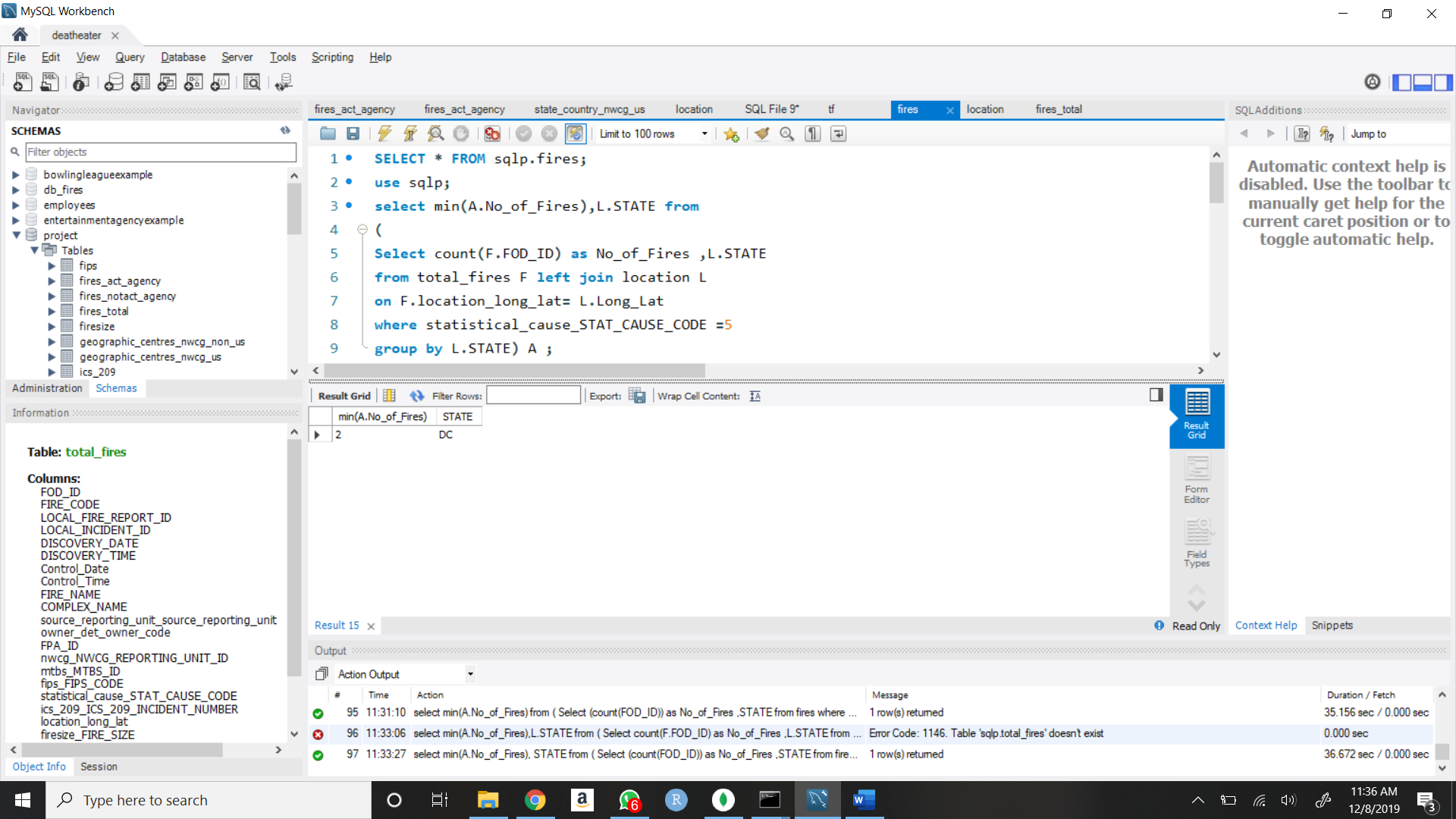
### Translation

Select the State where the cause of forest fire is due to debris burning

Cleanup

* Select minimum of No\_of\_Fires which is in subquery A and STATE from
* Select the count of FOD\_ID and STATE from total\_fires table
* where statistical\_Cause\_STAT\_CAUSE\_CODE is equal to 5 (debris burning) and group by STATE

### Screen Shot of SQL Query and Results



## Query 3

### Question 3

One advocacy group says human actions and not Nature is to blame for most wildfires. Write a query that supports this statement.

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

Shows the list of all incidents which are due to human actions is more than the number of incidents due to natural causes like lightning.(13 Rows)

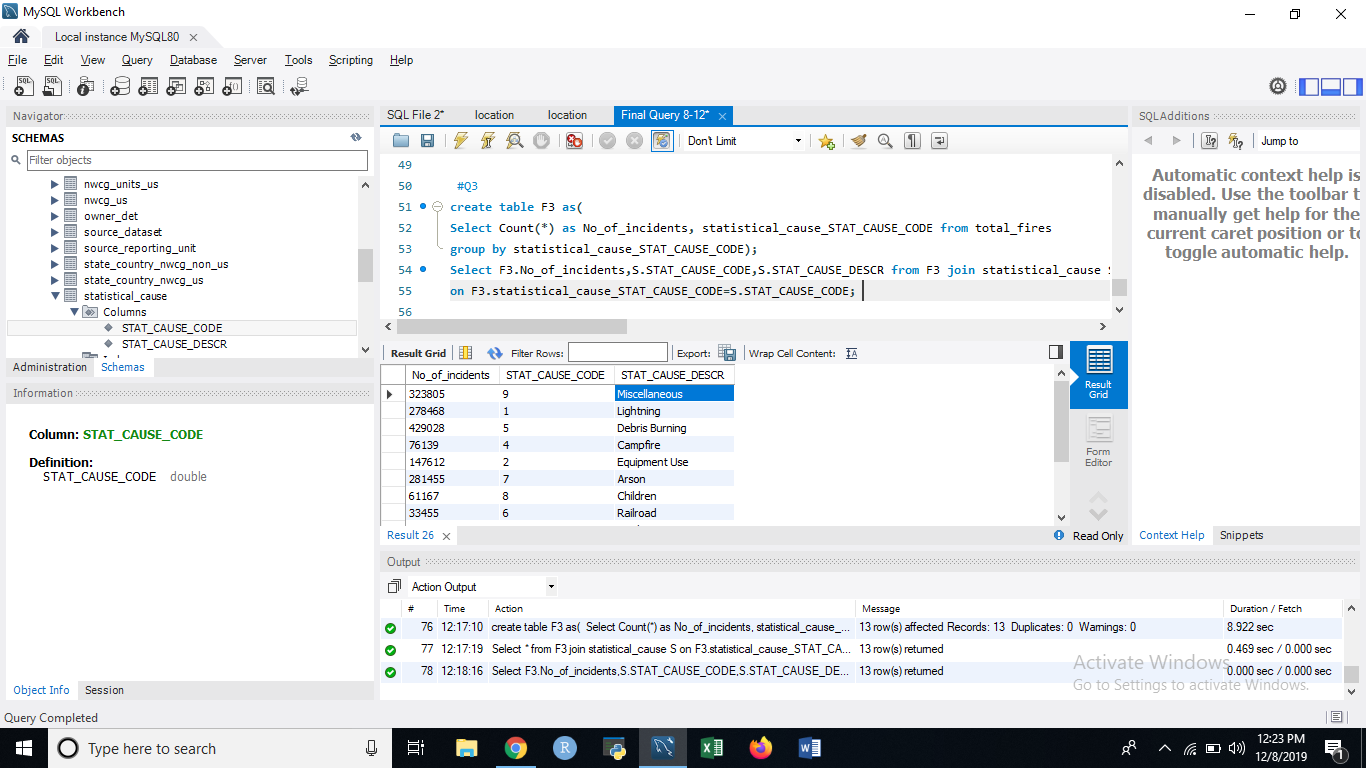
### Translation

Select the number of incidents and the statistical cause code and description

Cleanup

* Select total row count as No\_of\_incidents, STAT\_CAUSE\_CODE and STAT\_CAUSE\_DESCR in a subquery and grouped by statistical\_Cause\_STAT\_CAUSE\_CODE.
* From created table F3 joined based on STAT\_CAUSE\_CODE

### Screen Shot of SQL Query and Results



## Query 4

### Question 6

What were the forests that had no fires that lasted more than two days?

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

Shows the list of forests that had no fires that lasted for more than 2 days (1220 Rows)

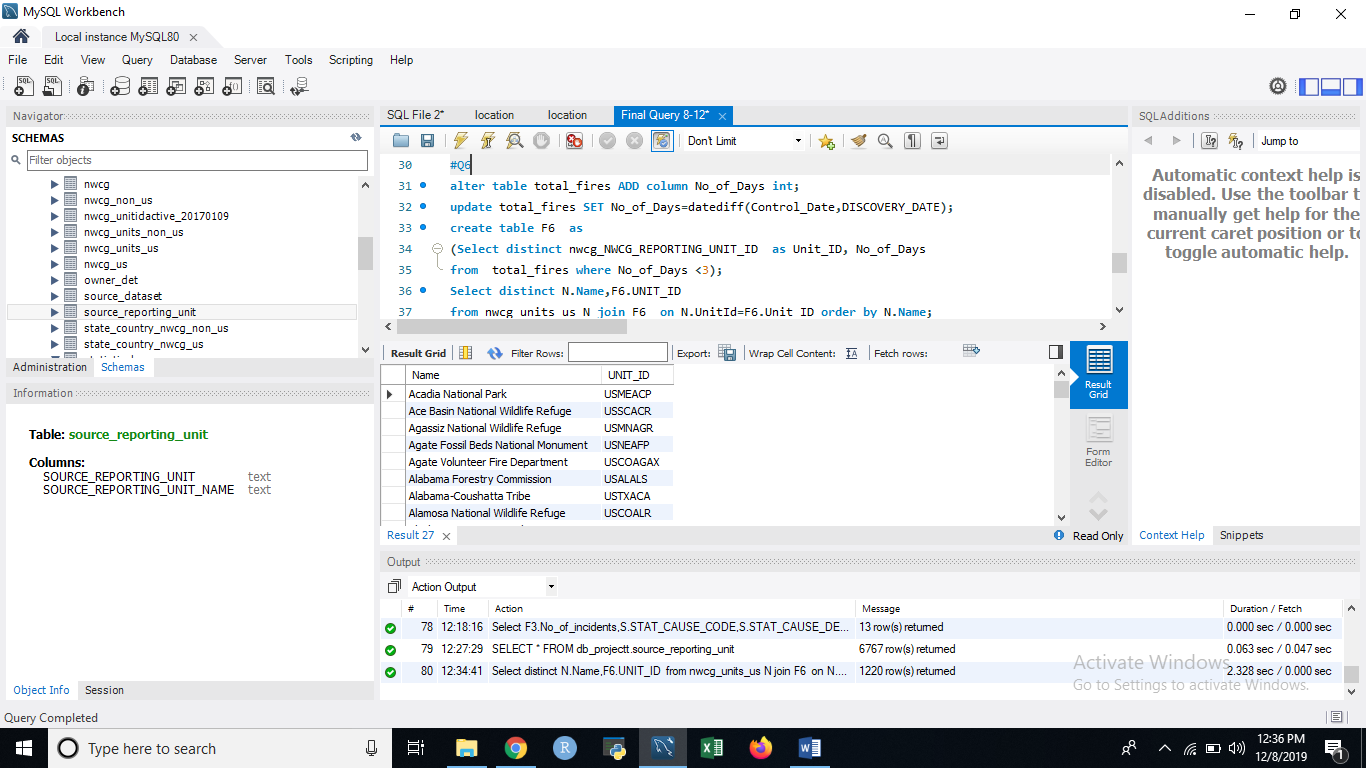
### Translation

Select the Name of the forest and nwcg reporting unit id where fires had lasted less than 3 days

Cleanup

* Add a column No\_of\_Days to the total\_fires table
* Update total\_fires as No\_of\_Days is the difference between Control\_Date and DISCOVERY\_DATE
* Select distinct values of nwcg\_NWCG\_REPORTING\_UNIT\_ID as Unit\_Id and No\_of\_Days from total\_fires table where No\_of\_Days is less than 3 in a subquery in the created table F6
* Select distinct Name and UNIT\_ID and order by Name

### Screen Shot of SQL Query and Results



## Query 5

### Question 8

Which forest had the least number of fires?

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

Shows the list of forest that had the least fire incidents (197 Rows)

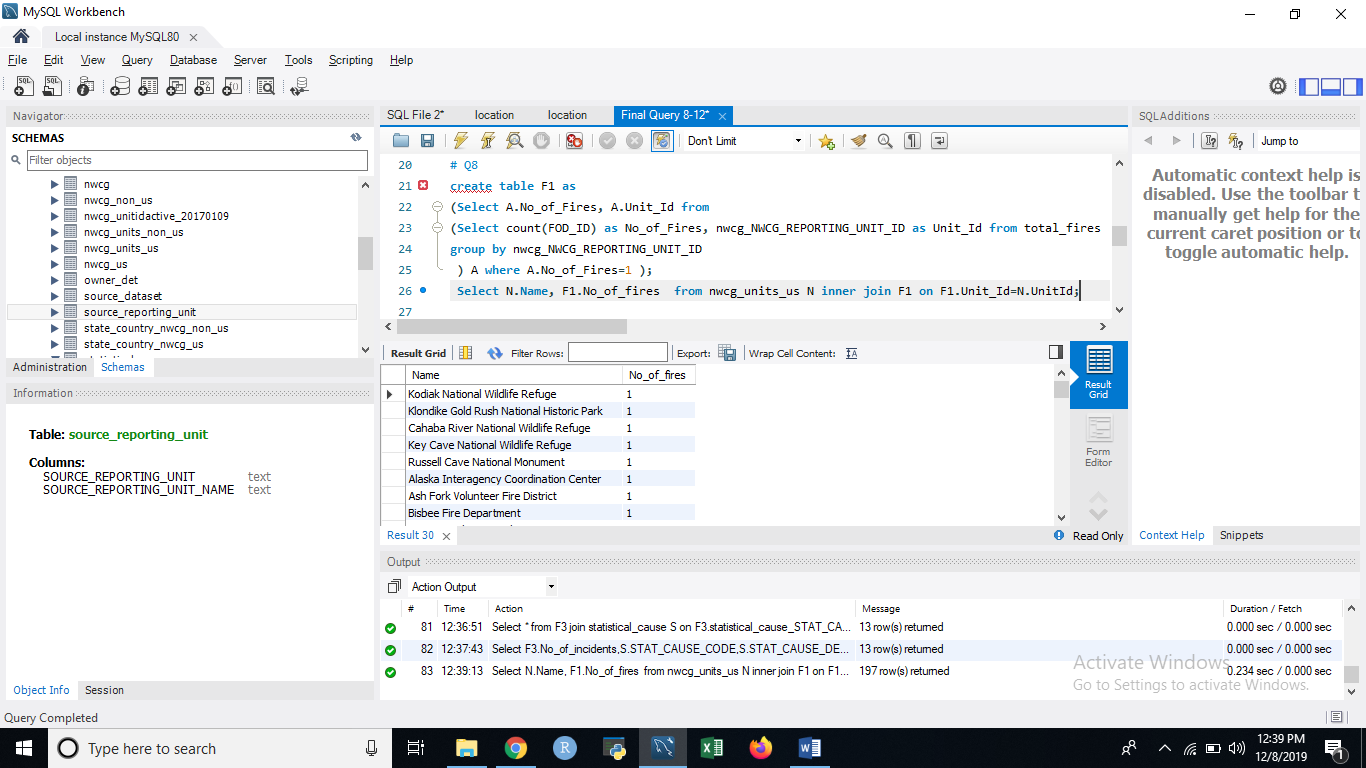
### Translation

Select the name and number of fires from nwcg units in US where the number of fires is equal to 1

Cleanup

* Select No\_of\_Fires and Unit\_Id from subquery Select count of FOD\_ID as No\_of\_Fires, nwcg\_NWCG\_REPORTING\_UNIT\_ID as Unit\_Id from total\_fires table and group by nwcg\_NWCG\_REPORTING\_UNIT\_ID
* Where the No\_of\_fires is equal to 1
* Select Name and No\_of\_Fires

### Screen Shot of SQL Query and Results



## Query 6

### Question 2

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 3 children in a group visiting a forest. Name 3 forests where this would be the least appropriate

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

Shows the names of the three forests where it would be least appropriate to suggest banning children unless there is one adult for every 3 children (3 Rows)

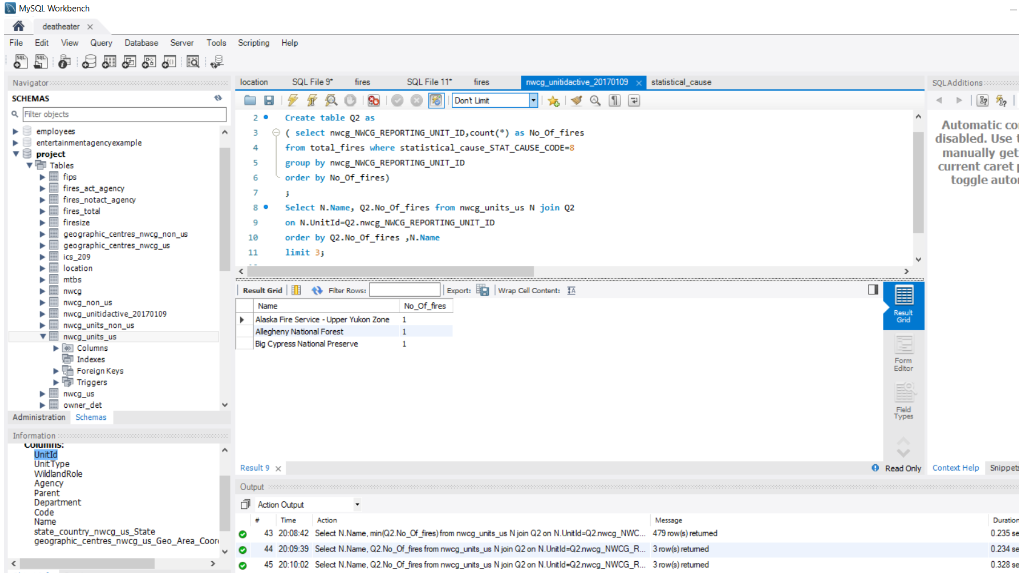
### Translation

Select the name of the forest and no of fires where statistical\_Cause\_STAT\_CAUSE\_CODE is equal to 8 and limit the list to 3

Cleanup

* Create a table Q3 as Select nwcg\_NWCG\_REPORTING\_UNIT\_ID and count of all records as No\_of\_fires from total\_fires table where Statical\_cause\_STAT\_CAUSE\_CODE is equal to 8 and group by nwcg\_NWCG\_REPORTING\_UNIT\_ID and order by No\_of\_fires
* Select Name and No\_of\_fires from nwcg\_units\_us table and on the UnitId column and order by No\_of\_fires and Name and limit the result to 3

### Screen Shot of SQL Query and Results



Data Review for MongoDB

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

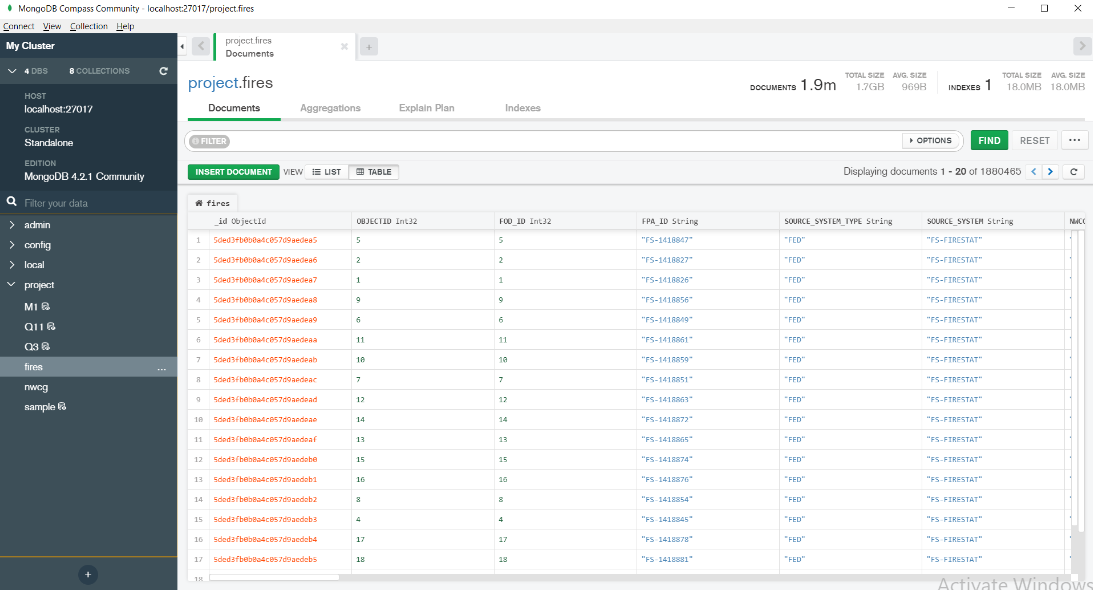
Since MongoDB is document based and hence does not require any normalization and row structure is not enforced by the table.

Physical Mongo Database

## Assumptions/Notes About Data Set

* Shape column not considered
* The order of the questions answered is jumbled. Kindly refer to question number indicated which corresponds to the questions set

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



## Data in the Database

|  |  |  |
| --- | --- | --- |
| **Collection Name** | **Relationshps With Other Collections (if any)** | **# of Documents in Collection** |
| Fires |  | 1880465 |
| Nwcg |  | 5867 |

MongoDB Queries/Code

## Query 1

### Question 7

Which state had more fires in the first half of a calendar year than the second half of the calendar year?

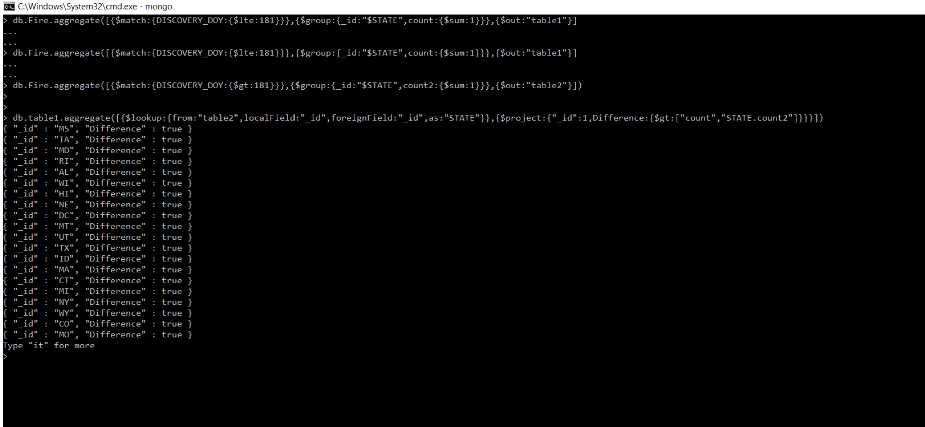
### Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result)

Showa the states for which the fires in the first half of the calendar year is more than that of the second half of the calendar year (40 Documents)

### Translation

* Select ids and value greater than difference between count and count2 from table1 joined on table2 based on id.

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



## -Query 2

### Question 1

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 state has the least chance to win a share of the fund?

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

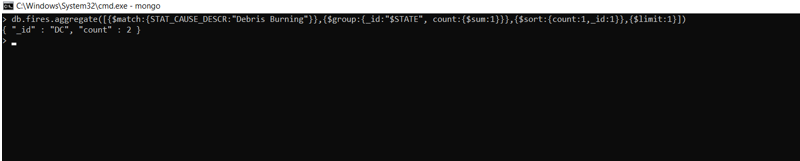
Db.fires.aggregate([{$match:{STAT\_CAUSE\_DESCR:”Debris Burning”}},{$group:{\_id:”$STATE”,count:{$sum:1}}},{$sort:{count:1,\_id:1}},{$limit:1}])

(1 Document)

### Translation

* Select state and count of occurrence where stat\_cause\_descr is debris burning

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



## Query 3

### Question 3

One advocacy group says human actions and not Nature is to blame for most wildfires. Write a query that supports this statement.

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

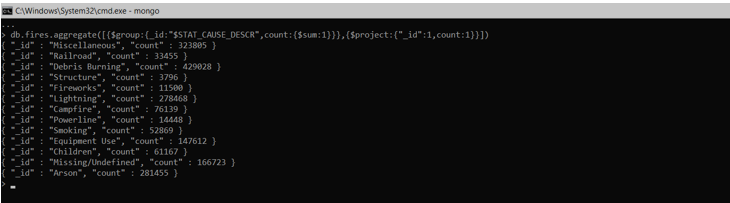
Db.fires.aggregate([{$group:{\_id:”$STAT\_CAUSE\_DESCR”,count:{$sum:1}}},{$project:{“\_id”:1,count:1}}])

(13 Documents)

### Translation

* Select id and count from fires grouped on STAT\_CAUSE\_DESCR

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



## Query 4

### Question 6

What were the forests that had no fires that lasted more than two days?

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

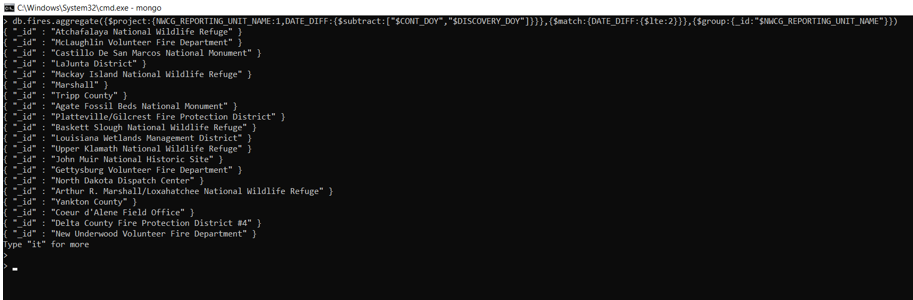
Shows the list of forests that had no fires that lasted for more than 2 days (1220 Documents)

Db.fires.aggregate({$project:{NWCG\_REPORTING\_UNIT\_NAME:1,DATE\_DIFF:{$subtract:{[“$CONT\_DOY”,”$DISCOVERY\_DOY”]}}},{$match:{DATE\_DIFF:{$lte:2}}},{$group:{\_id:”$NWCG\_REPORTING\_UNIT\_NAME”}})

### Translation

* Select NWCG\_REPORTING\_UNIT\_NAME, DATE\_DIFF from fires where date difference is less than 2 and then grouped on NWCG\_REPORTING\_UNIT\_NAME

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



## Query 5

### Question 8

Which forest had the least number of fires?

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

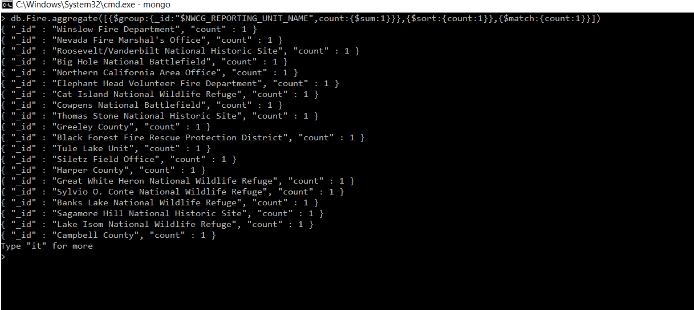
Shows the forests with the least number of fires (197 Documents)

Db.fires.aggrergate([$group:{\_id:”$NWCG\_REPORTING\_UNIT\_NAME”,COUNT:{$sum:1}}},{$sort:{COUNT:1}},{$match:{count:1}}])

### Translation

* Select forest count from fires where count is 1 sorted on count grouped on NWCG\_REPORTING\_UNIT\_NAME

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



## Query 6

### Question 2

One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 3 children in a group visiting a forest. Name 3 forests where this would be the least appropriate

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

Name of the three forests where it would be least appropriate to suggest banning children unless there is one adult for every 3 children. (3 Documents)

Db.fires.aggregate9[{$match:{STAT\_CAUSE\_DESCR:”Children”}},{$group:{\_id:”$NWCG\_REPORTING\_UNIT\_NAME”,count:{$sum:1}}},{$sort:{count:1,”\_id”:1}},{$limit:3}])

### Translation

* Select forest names from fires where STAT\_CAUSE\_DESCR is Children grouped on NWCG\_REPORTING\_UNIT\_NAME limited to 3 records

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

