SQL-Mongo Project – Spatial Data of US Wildfires

BUAN 6320

Group Members
Garima Gupta
Guru Prasad Kumar
Rohitashva Jajoo
Pratik Gupta

Group 15

Contents

Data Model

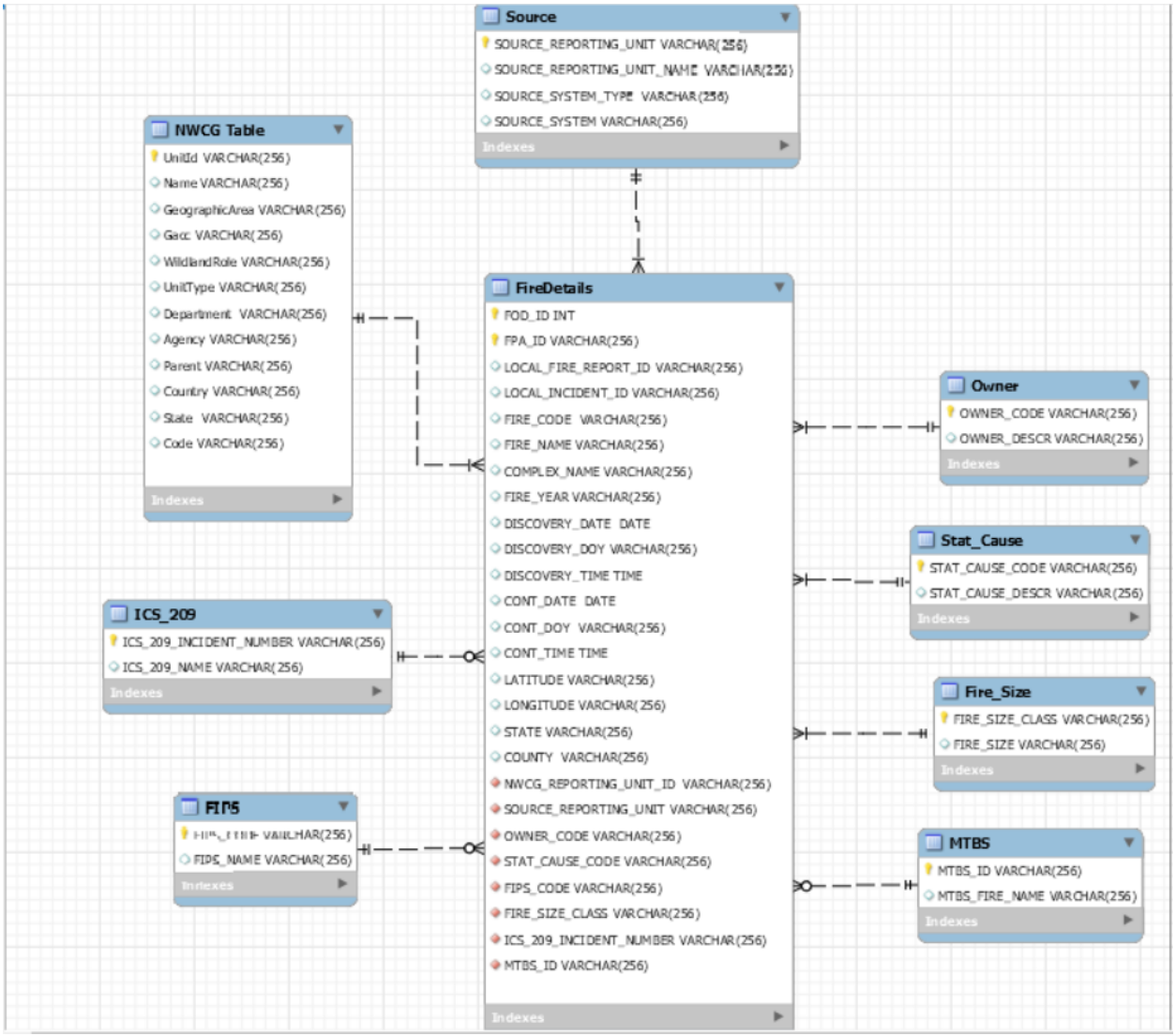
Assumptions/Notes About Data Entities and Relationships

1. A fire originating will have a single source and each fire will have a fire report and a single unit may report multiple fires.
2. Each fire may be reported by the NWGC agency and an agency unit may report one to many fires.
3. Each fire will have a statistical cause and each statistical cause denote the cause of one to many fires.
4. Each fire will have an origin and owner of the land or property from where the fire originated, and each property may have one to many instances of fire.
5. Each fire is classified based on size and each fire class may represent one to many fires.
6. Each fire may or may not have an mtbs id and each mtbs id may denote zero to many fires.
7. Each fire may or may not have an incident number from the secondary ICS209 report and each report may denote zero to many fires.
8. Each fire may have a FIPS state/county code and zero to many fires may occur in each state or county.
9. Each NWCG reporting unit or source reporting unit may be located near a forest and hence therefore represents a forest.
10. Each forest may have one to many fires.

Reasons for 3NF:

1. To eliminate any undesirable data anomalies that may be present in the data.
2. To reduce the need for restructuring over time.
3. To make the data model more informative.
4. To make the data model neutral to different kinds of query.
5. To ensure referential integrity.
6. To make sure there are no transitive functional dependencies between any column in table which satisfies the condition of 3NF.

Entity-Relationship Diagram



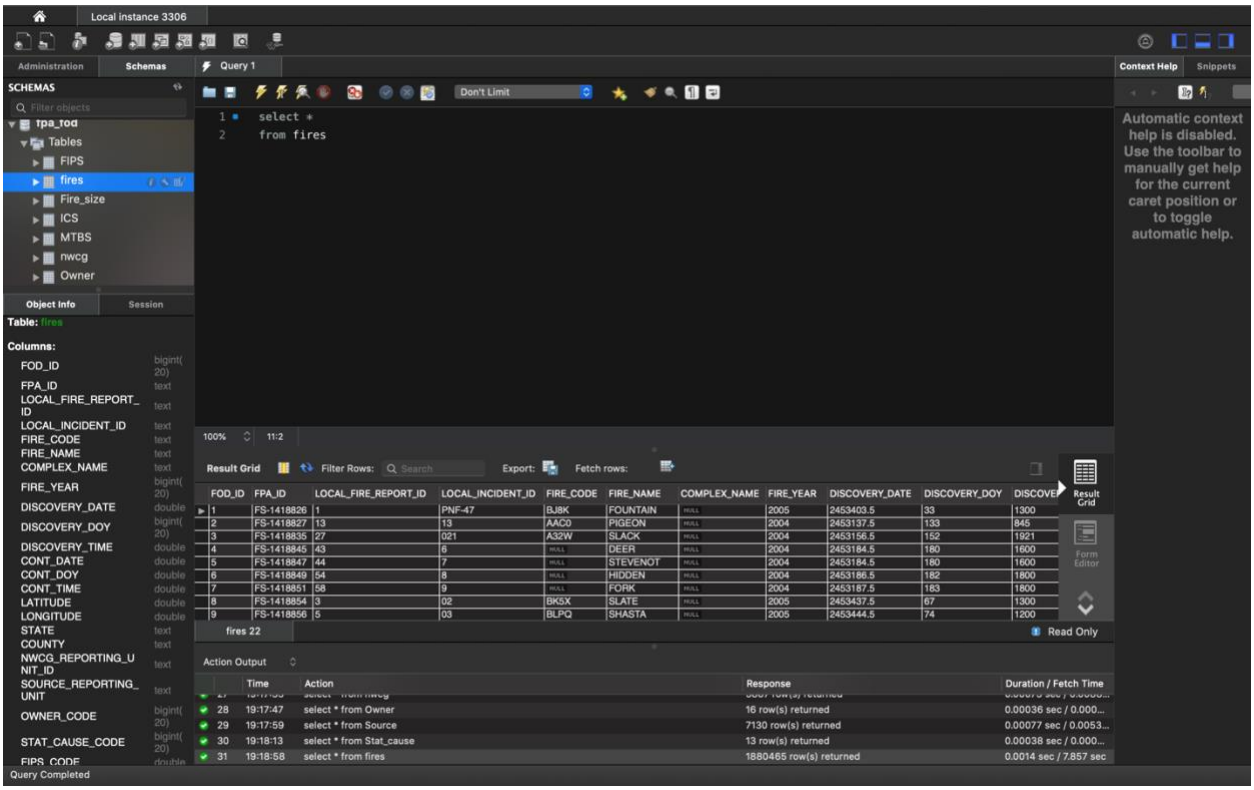
Physical Database

Assumptions/Notes About Data Set

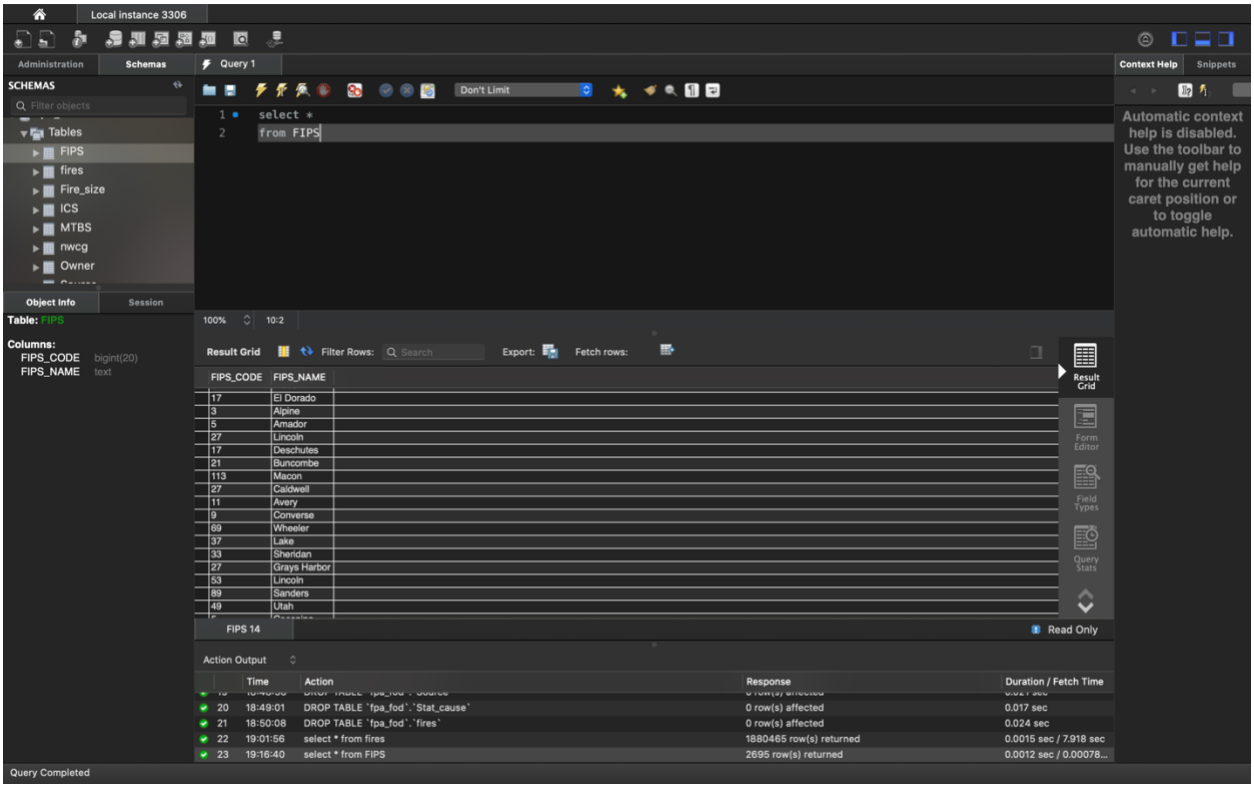
1. All records on fire data are contained in Fire Details table
2. All records on NWCG reporting agency are kept on NWCG table
3. NWGC agency and its individual units generates the fire report for the fires originating.
4. Fire report is generated by the source reporting unit indicating the source of the fire.
5. Owner table contains the data about primary owner or entity responsible for managing the land at the point of origin of the fire at the time of the incident.
6. Stat cause table contains the data about statistical cause of fire.
7. Fire size table contains the data about the size of the fire.
8. MTBS table contains the data about the burn severity of caused by the fire.
9. ICS table contains the fire id and fire number of the secondary ICS_209 fire report.
10. FIPS table contains data about the federal information process standards.

Screen shot of Physical Database objects

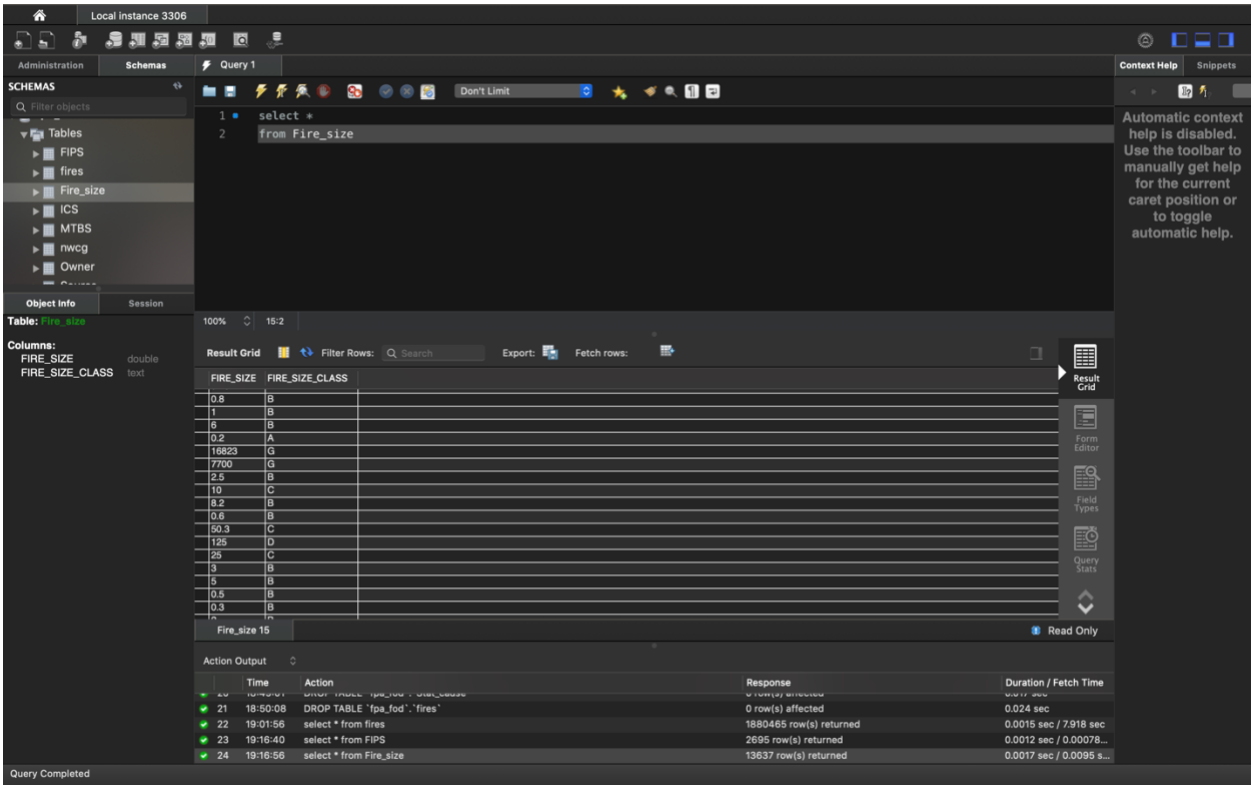
Fires Table:



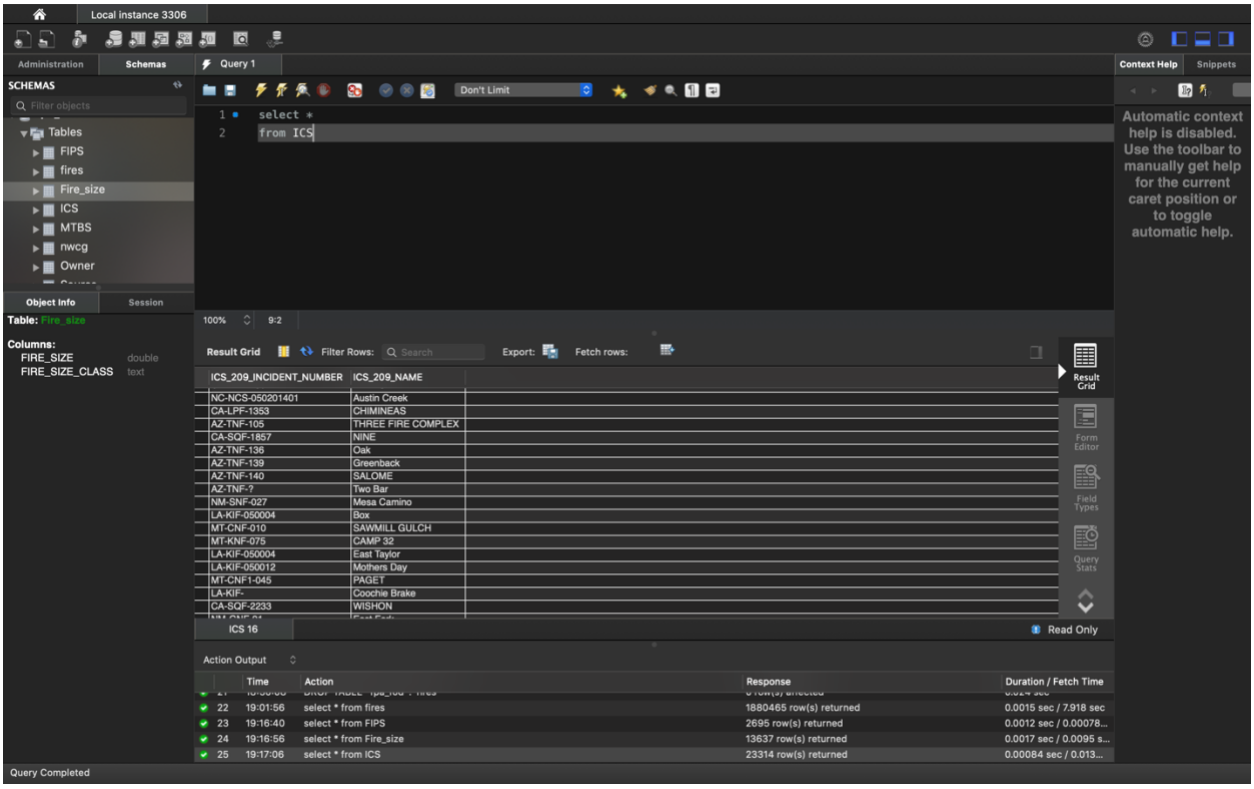
FIPS Table



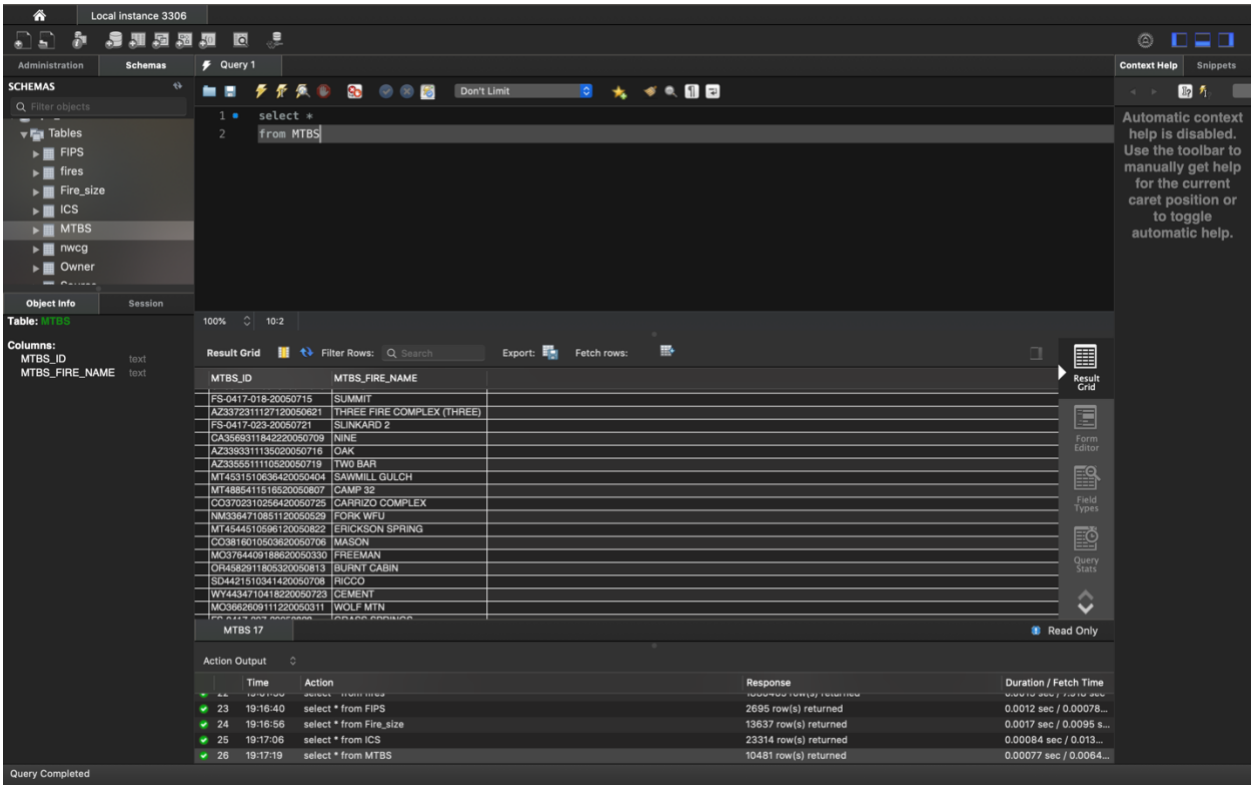
Fire Size Table



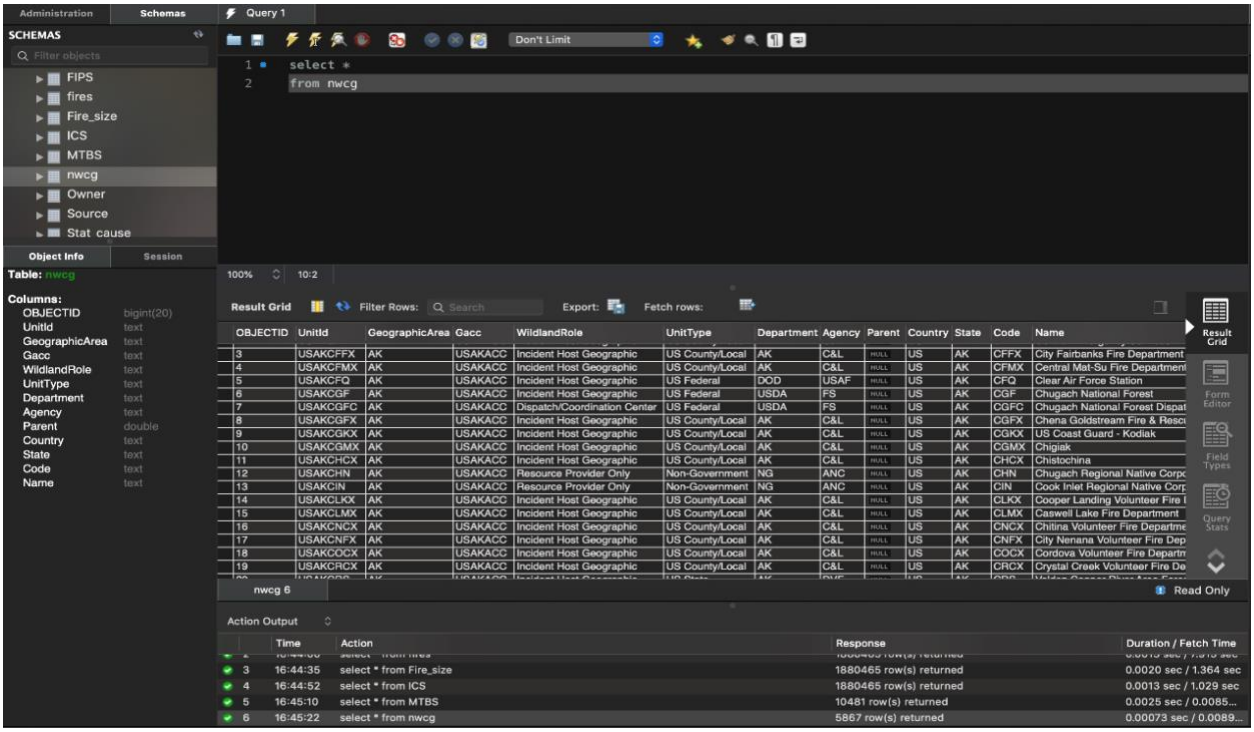
ICS Table



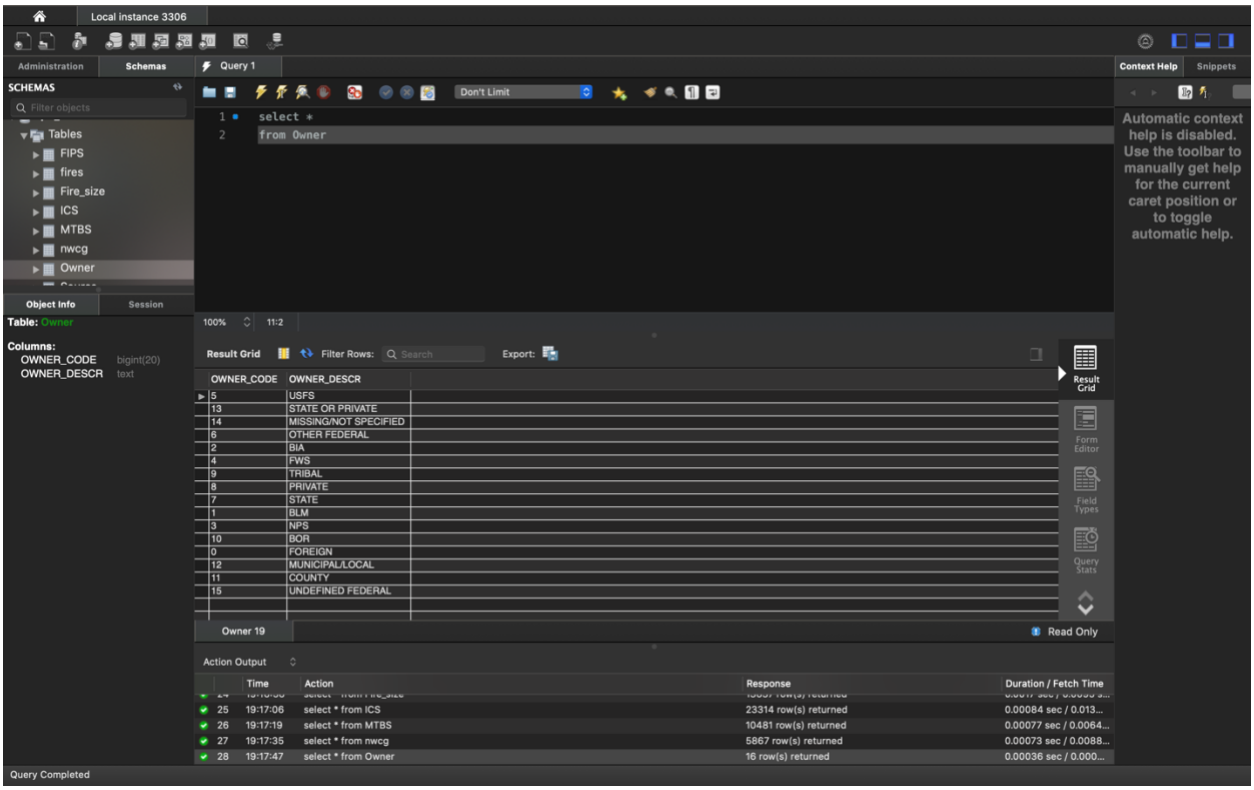
MTBS Table



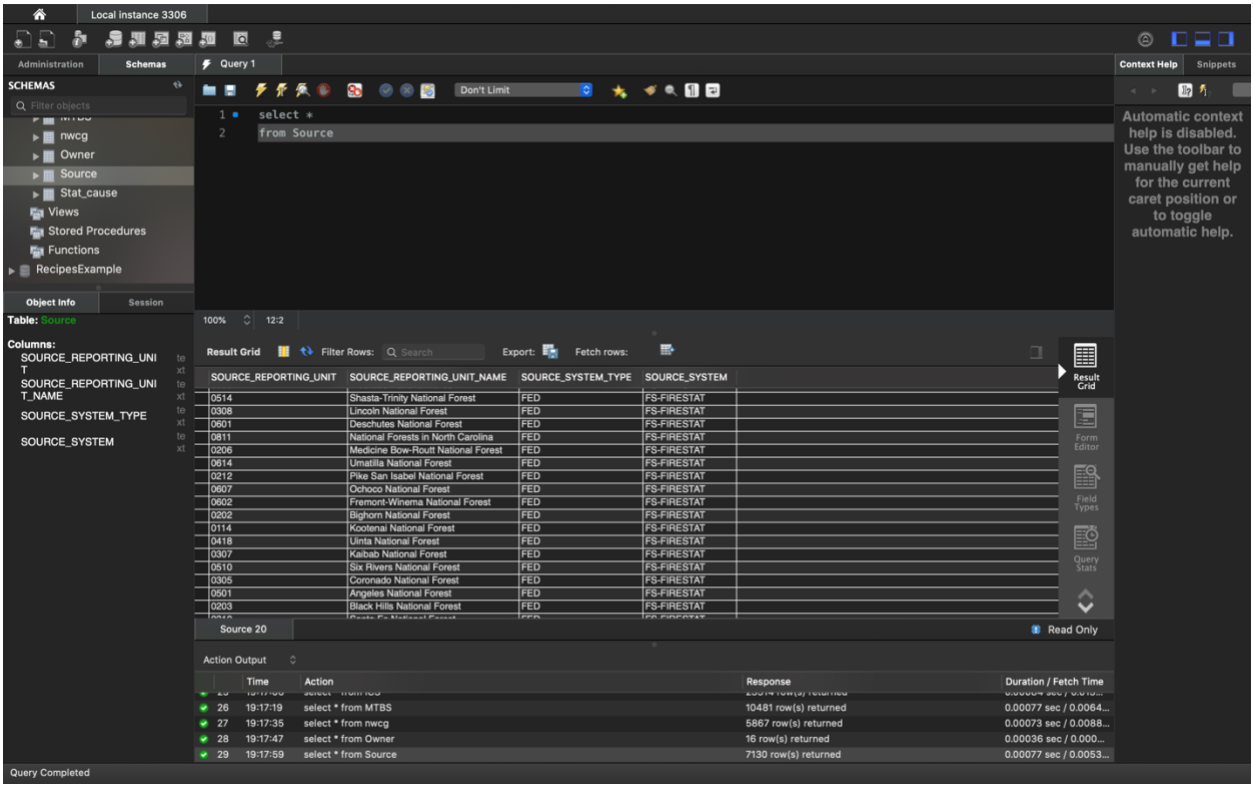
NWCG Table



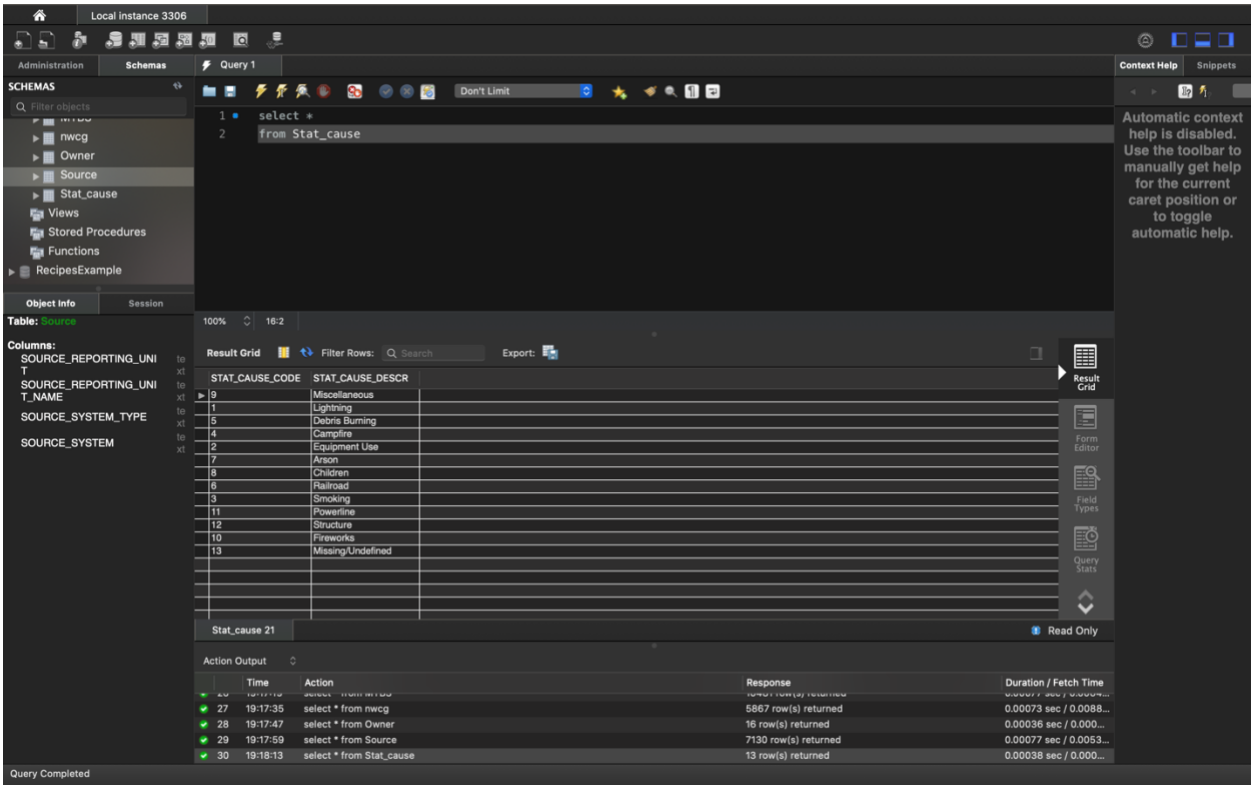
Owner Table



Source Table



Stat_Cause Table



Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
FireDetails	• FOD_INT • FPA_ID	• NWCG_REPORTING_UNIT_ID • SOURCE_REPORTING_UNIT • OWNER_CODE • STAT_CAUSE_CODE • FIPS_CODE • FIRE_SIZE_CLASS • ICS_209_INCIDENT_NUMBER • MTBS_ID	1880465
NWCG Table	UnitId		5867
Source	SOURCE_REPORTING_UNIT		7130
Owner	OWNER_CODE		16
Stat_Cause	STAT_CAUSE_CODE		13
Fire_Size	FIRE_SIZE_CLASS		13637
MTBS	MTBS_ID		10481
FIPS	FIPS_CODE		2695
ICS_209	ICS_209_INCIDENT_NUMBER		23314

SQL Queries

Query 1

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?

Assumptions

The state with fires caused by debris burning is the least, will have the least chance of winning the fund.

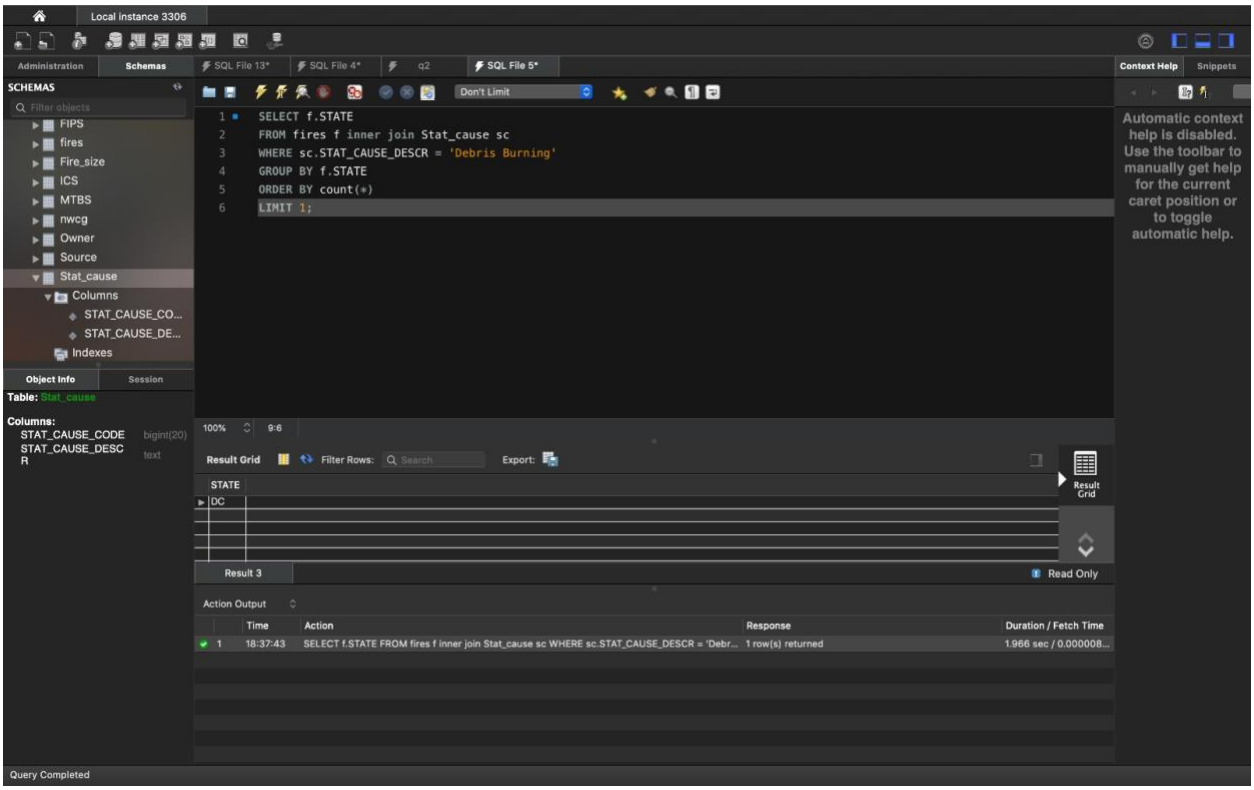
Translation

Select the state where the cause of fire due to Debris Burning is the least

Cleanup

Select STATE, COUNT (*) from fires table inner join Stat_cause where STAT_CAUSE_DESCR = 'Debris Burning' GROUP BY STATE ORDER BY FIRE_BY_DEBRIS

Screen Shot of SQL Query and Results



Result

DC state has the least chance of fire caused by Debris Burning, and hence has the least chance of winning the fund. (Rows returned:1)

Query 2

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.

Assumptions

In forests where children are the likely cause of fire, the forest with least number of fires is safer for children to be accompanied by an adult, and hence considered least appropriate. We have considered the Source_reporting_unit_name as the forest name(agencies, wildlife refuge etc) in our dataset.

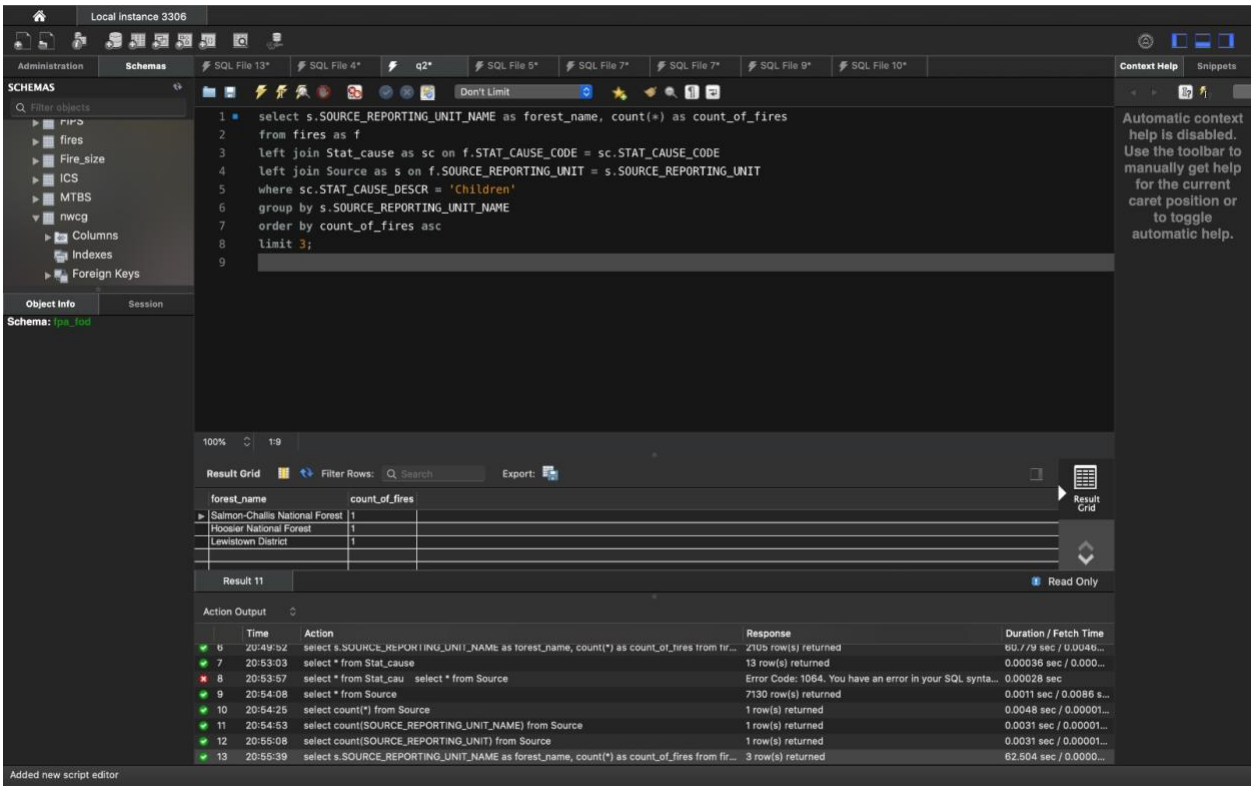
Translation

Select the forest (SOURCE_REPORTING_UNIT_NAME) where the ban for children to be accompanied with the adults is least appropriate

Cleanup

Select three SOURCE_REPORTING_UNIT_NAME (forest names) where STAT_CAUSE_DESCR = 'Children' and count of fires are least.

Screen Shot of SQL Query and Results



Result

Rows returned are 3. These are the three forests in which the ban is least appropriate.

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

Assumptions

Natural causes of fire ('Lightning', 'Structure') and Human Causes ('Debris Burning', 'Campfire', 'Equipment Use', 'Arson', 'Children', 'Railroad', 'Smoking', 'Powerline', 'Fireworks')

Translation

Select different causes with respective fire counts

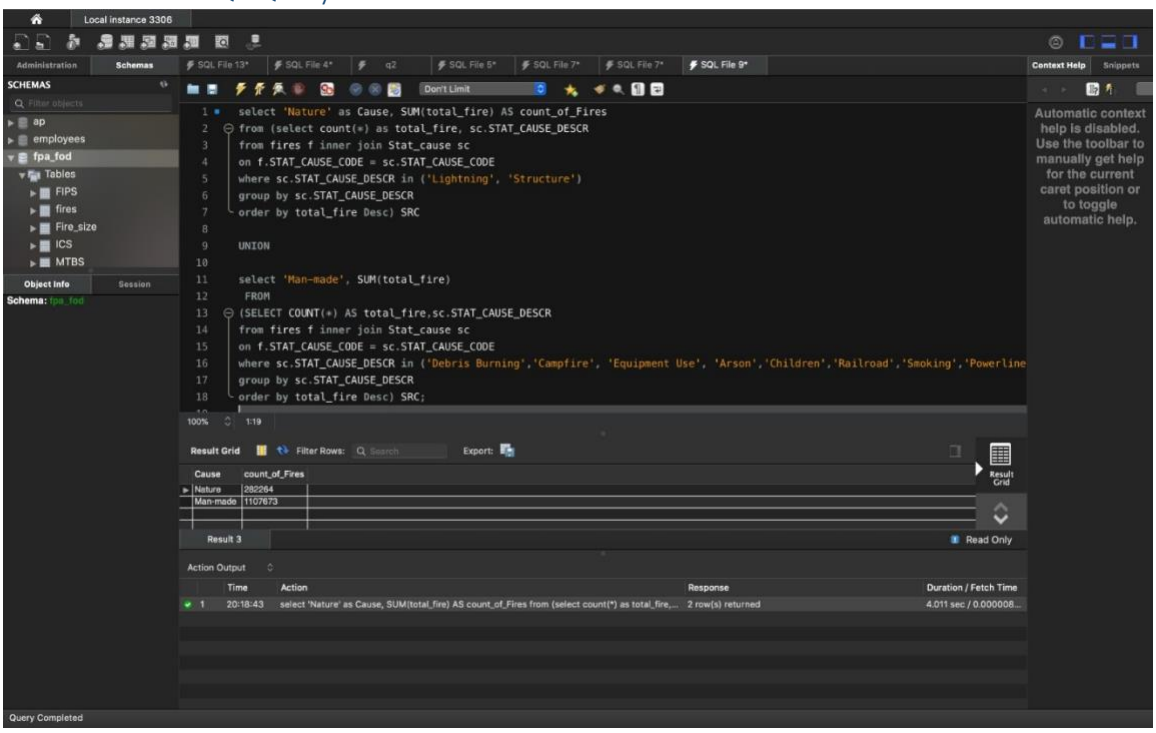
Cleanup

Select count (*), STAT_CAUSE_DESCR from fires table inner join Stat_cause where STAT_CAUSE_DESCR = ('Lightning', 'Structure')

Union

Select count (*), STAT_CAUSE_DESCR from fires table inner join Stat_cause where STAT_CAUSE_DESCR = ('Debris Burning', 'Campfire', 'Equipment Use', 'Arson', 'Children', 'Railroad', 'Smoking', 'Powerline', 'Fireworks')

Screen Shot of SQL Query and Results



Result

Rows returned are 2. Results show that, the human actions are to be blamed for most wildfires (11,87,673)

Query 4

Question 6:

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

Assumptions

We have considered the Source_reporting_unit_name (agencies, wildlife refuge etc) as the forest name in our dataset.

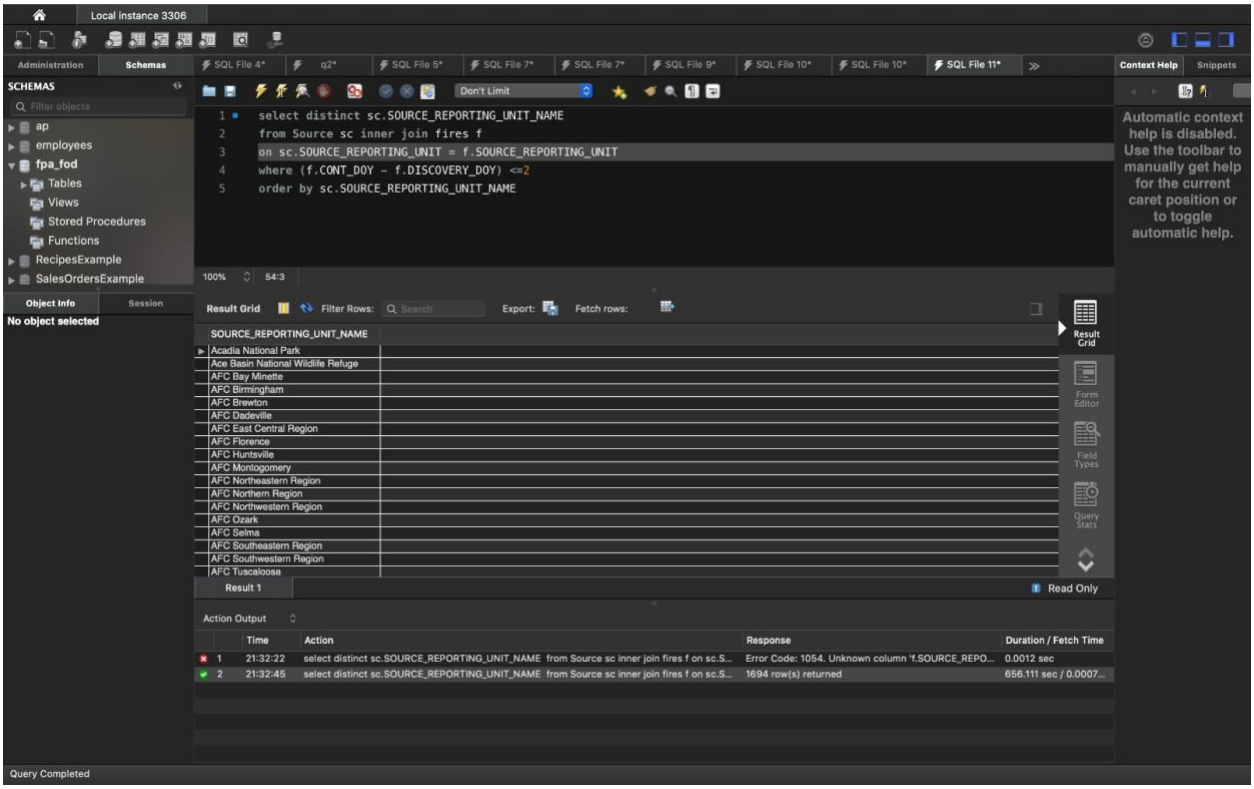
Translation

Select the forests where the fire was discovered and controlled in no more than 2 days

Cleanup

Select SOURCE_REPORTING_UNIT_NAME from source table inner join fire where (CONT_DOY - DISCOVERY_DOY) <=2

Screen Shot of SQL Query and Results



Result

Rows returned are 1694. The returned rows have the forest names in which the fire was discovered and contained within two days.

Query 5

Question 4:

What are the bottom two unit types that reported wildfires in each county in the US?

Assumptions

We have considered the Source_reporting_unit_name as the forest name in our dataset.

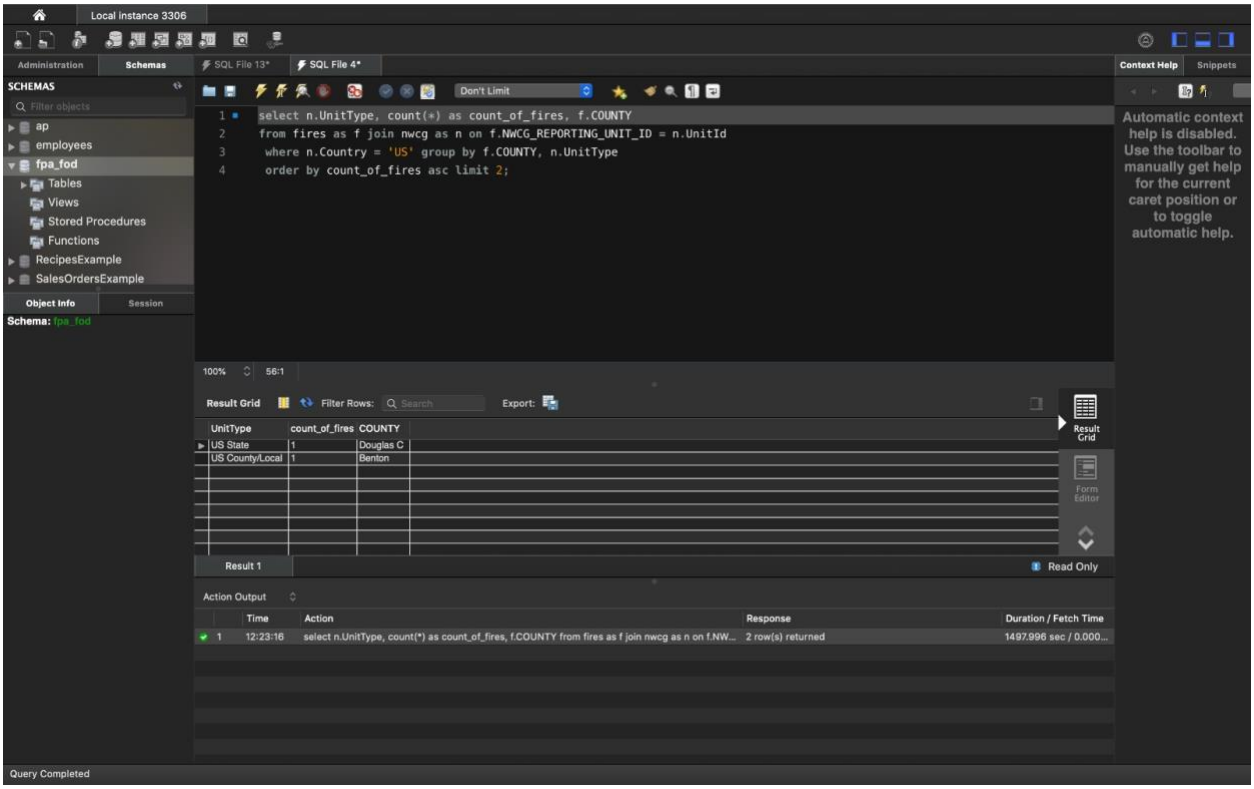
Translation

Select the bottom 2 underperforming unit types which reported the least number of fires in each county in US

Cleanup

Select UnitType, Count, County from fires join nwcg where county = US and order by count.

Screen Shot of SQL Query and Results



Result

Rows returned are 2. These are the most underperforming unit types.

Query 6

Question 8:

Which forest had the least number of fires?

Assumptions

We have considered the Source_reporting_unit_name (agencies, wildlife refuge etc) as the forest name in our dataset.

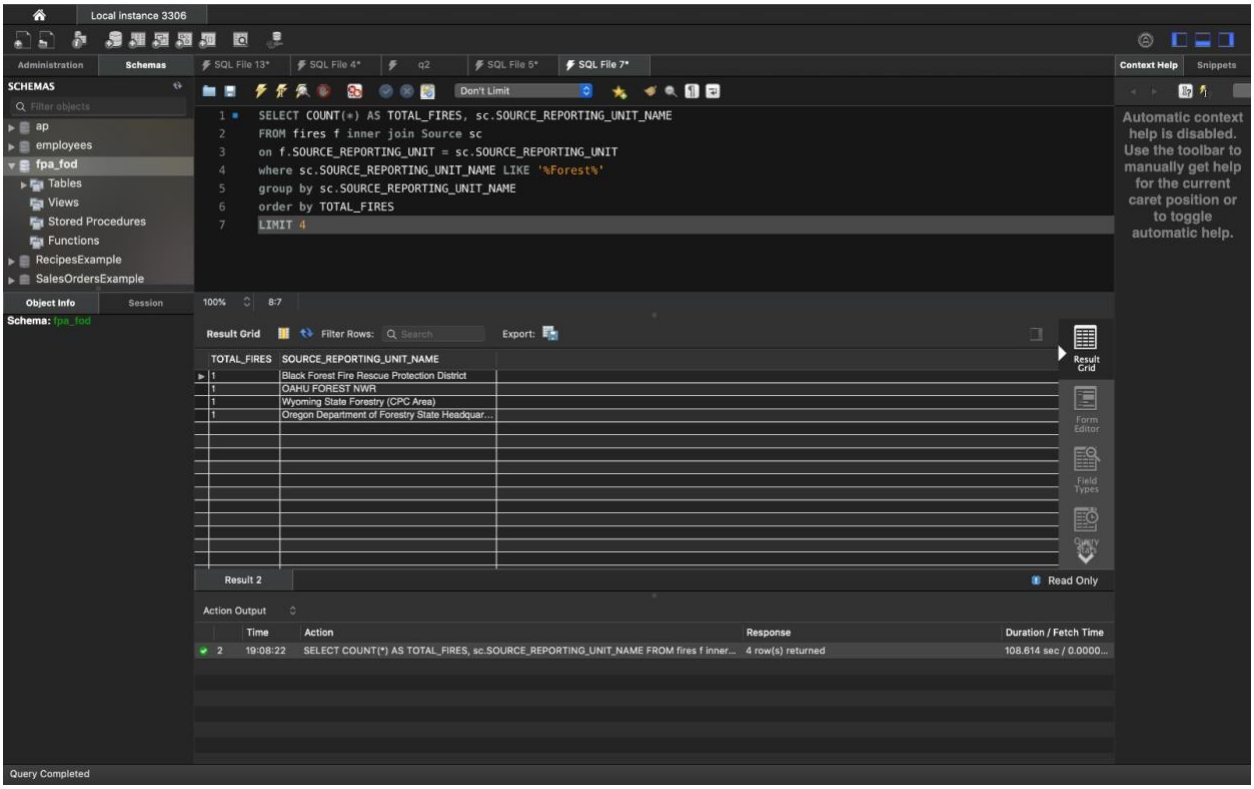
Translation

Select the forest which has least number of fires.

Cleanup

SELECT COUNT(FOD-ID), SOURCE_REPORTING_UNIT_NAME from fires tables group by SOURCE_REPORTING_UNIT_NAME

Screen Shot of SQL Query and Results



Result

Rows returned are 4. These are the forests with least number of fires.

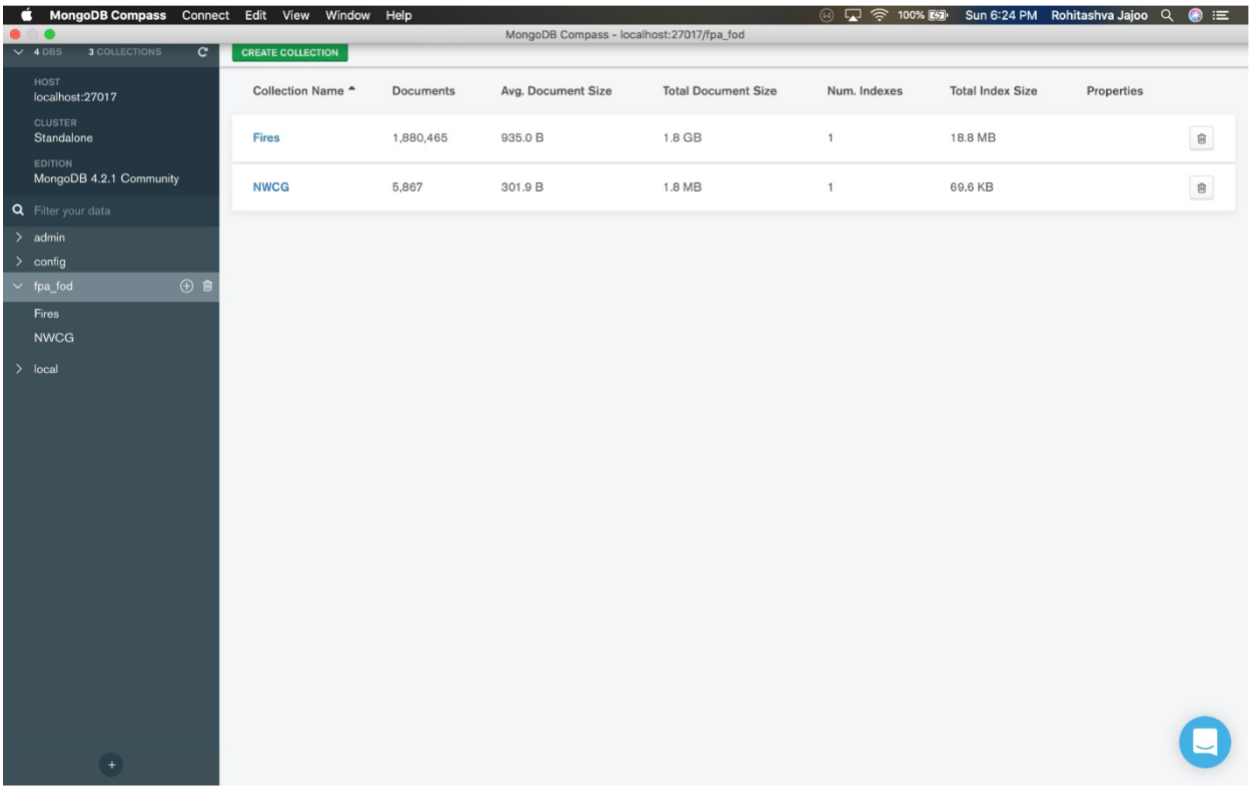
Data Review for MongoDB

Physical Mongo Database

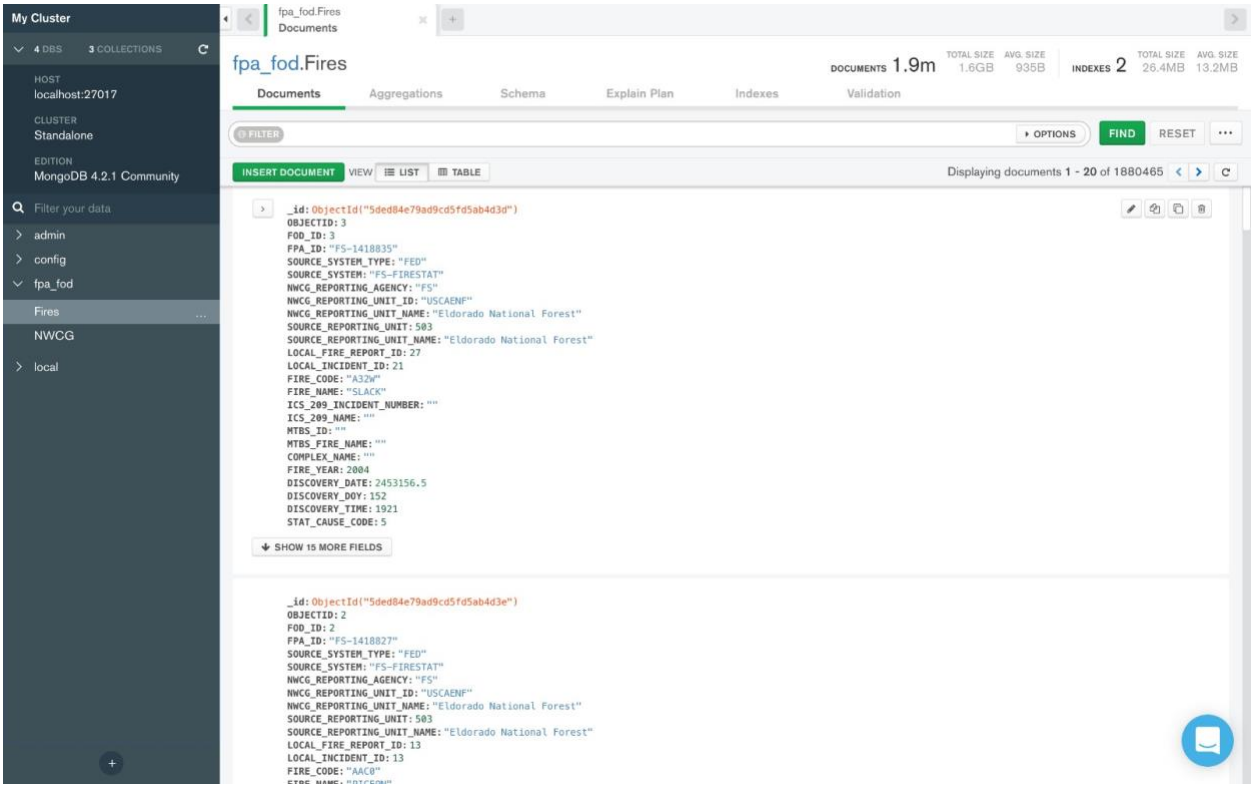
Assumptions/Notes About Data Set

1. All records on fire details are contained in fires document.
2. All records on NWCG reporting agency are kept on NWCG document

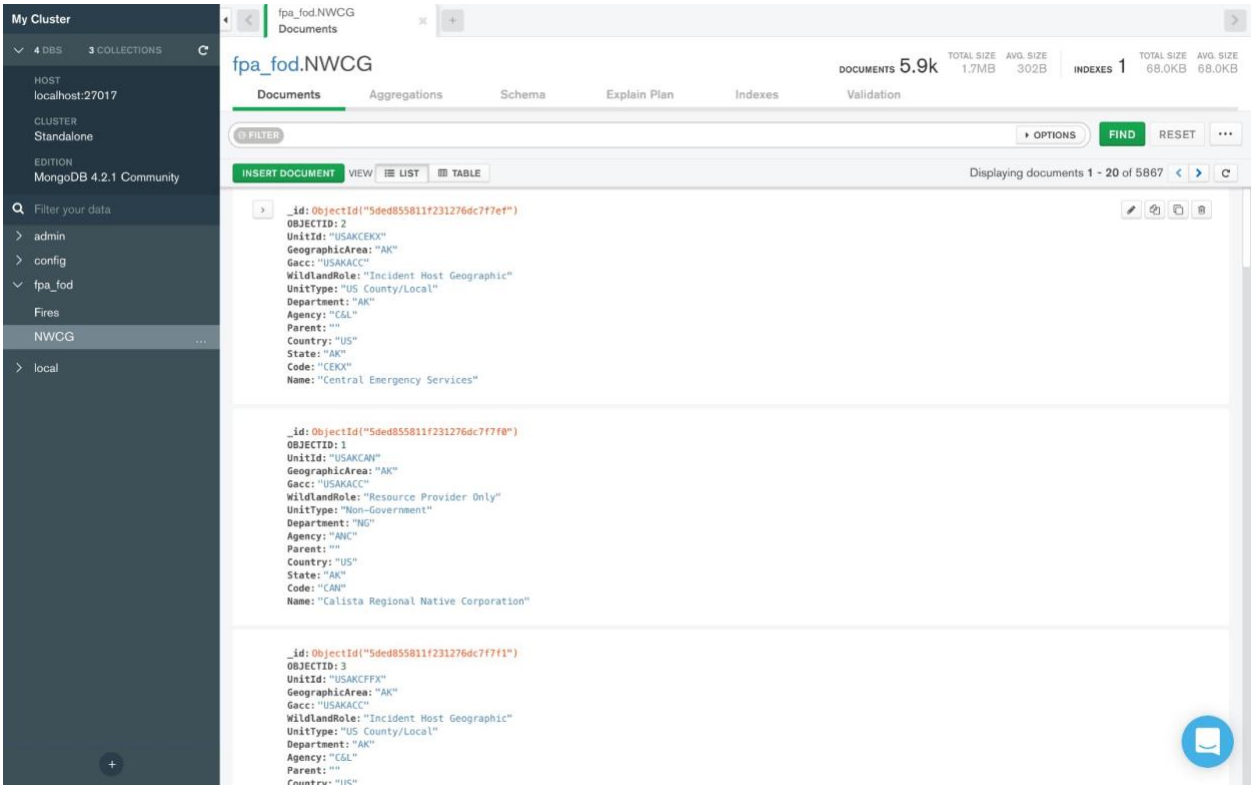
Screen shot of Physical Database objects (Database, Collections and Attributes)



Fires Document



NWCG Document



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 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?

Assumptions

The state with fires caused by debris burning is the least, will have the least chance of winning the fund.

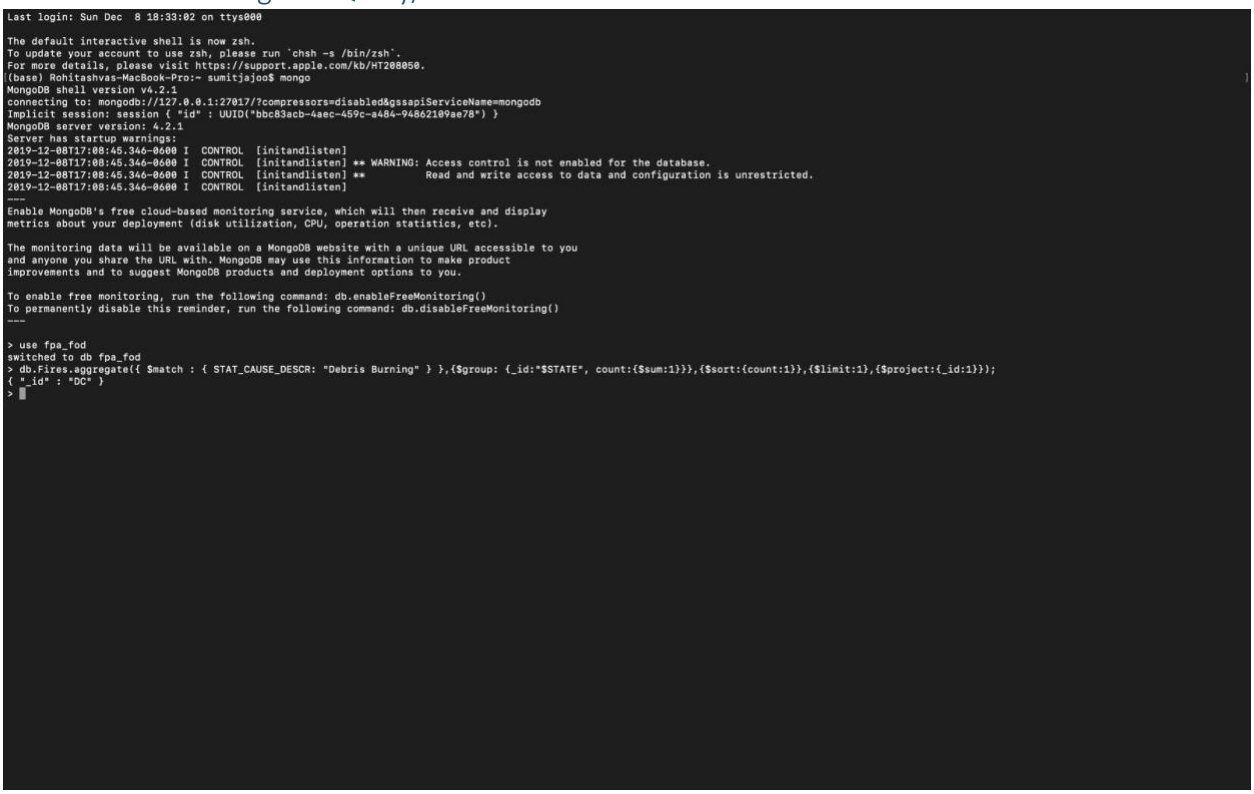
Translation

Select the state where the cause of fire due to Debris Burning is the least

Cleanup

Match stat_cause_descr with ‘debris burning’ and group by state, sort by ascending order, limit to 1.

Screen Shot of MongoDB Query/Code and Results



Result

DC state has the least chance of fire caused by Debris Burning, and hence has the least chance of winning the fund.

Query 2

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.

Assumptions

In forests where children are the likely cause of fire, the forest with least number of fires is safer for children to be accompanied by an adult, and hence considered least appropriate. We have considered the Source_reporting_unit_name as the forest name in our dataset.

Translation

In forests where children are the likely cause of fire, the forest with least number of fires is safer for children to be accompanied by an adult, and hence considered least appropriate. We have considered the Source_reporting_unit_name as the forest name in our dataset.

Cleanup

Match STAT_CAUSE_DESCR = 'Children', group by Source_reporting_unit_name and display first 3 forest names.

Screen Shot of MongoDB Query/Code and Results

```
Last login: Sun Dec  8 21:06:40 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
(base) Rohitashvas-MacBook-Pro:~ sumitjajoo$ mongo
MongoDB shell version v4.2.1
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("c8b825b6-4a91-40b5-a9c5-fc3f7e0a6664") }
MongoDB server version: 4.2.1
Server has startup warnings:
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---

> use fpa_fod
switched to db fpa_fod
> db.Fires.aggregate({ $match : { STAT_CAUSE_DESCR: "Children" } },{$group: {_id:"$SOURCE_REPORTING_UNIT_NAME", count:{$sum:1}}},{$sort:{SOURCE_REPORTING_UNIT_NAME:1, count:1}},{$limit:3});
{ "_id" : "Quitman", "count" : 1 }
{ "_id" : "Pueblo County", "count" : 1 }
{ "_id" : "OPERATIONS SALEM", "count" : 1 }
>
```

Result

These are the three forests in which the ban is least appropriate.

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.

Assumptions

Natural causes of fire ('Lightning', 'Structure') and Human Causes ('Debris Burning', 'Campfire', 'Equipment Use', 'Arson', 'Children', 'Railroad', 'Smoking', 'Powerline', 'Fireworks')

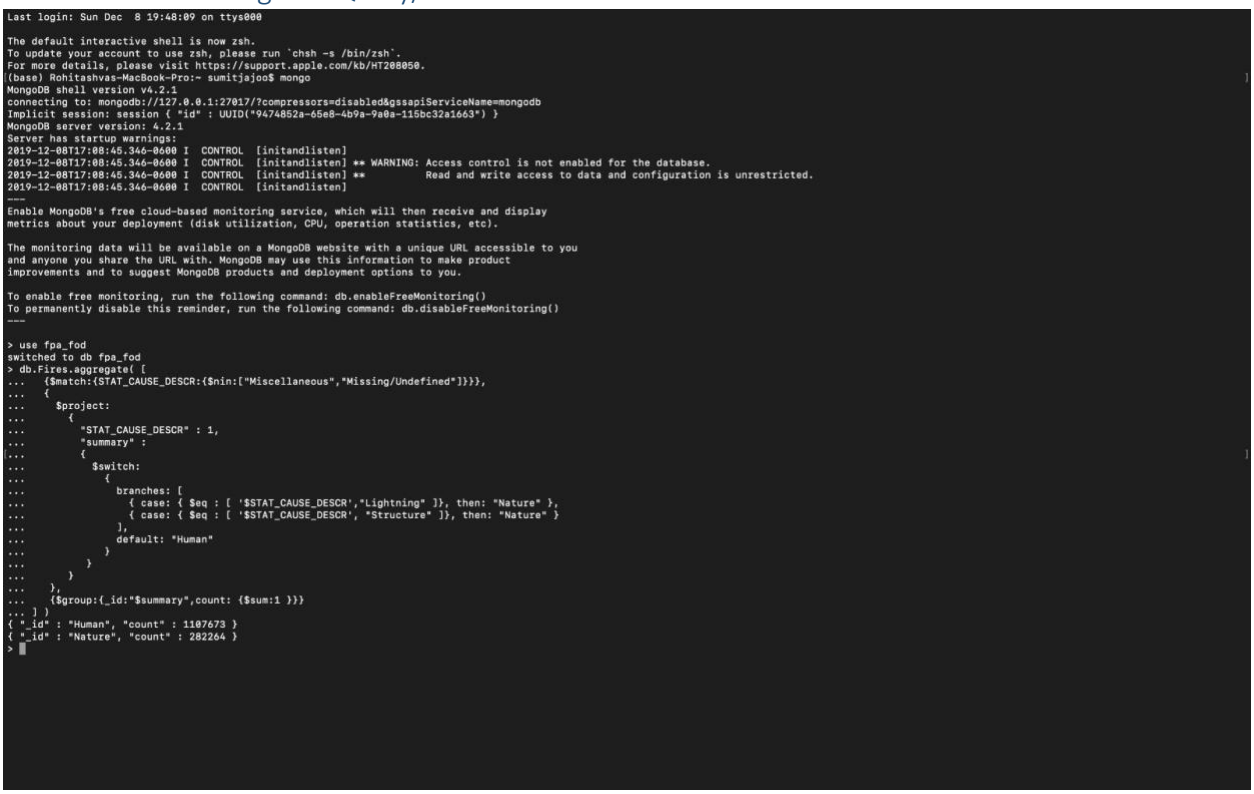
Translation

Select different causes with respective fire counts

Cleanup

Match stat_cause_descr with lighting or structure as natural cause of fire, default cause of fire as human and count no of fires.

Screen Shot of MongoDB Query/Code and Results



```
Last login: Sun Dec  8 19:48:09 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
(base) Rohitashvas-MacBook-Pro:~ sumitjajoo$ mongo
MongoDB shell version v4.2.1
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("9474852a-65e8-4b9a-9a0a-115bc32a1663") }
MongoDB server version: 4.2.1
Server has startup warnings:
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---

> use fpa_fod
switched to db fpa_fod
> db.Fires.aggregate( [
...   {$match:{STAT_CAUSE_DESCR:{$nin:["Miscellaneous","Missing/Undefined"]}}},
...   {
...     $project:
...       {
...         "STAT_CAUSE_DESCR" : 1,
...         "summary" :
...         {
...           $switch:
...             {
...               branches: [
...                 { case: { $eq : [ '$STAT_CAUSE_DESCR',"Lightning" ]}, then: "Nature" },
...                 { case: { $eq : [ '$STAT_CAUSE_DESCR', "Structure" ]}, then: "Nature" }
...               ],
...               default: "Human"
...             }
...         }
...       }
...   },
...   {$group:{_id:"$summary",count: {$sum:1 }}}
... ] )
{ "_id" : "Human", "count" : 1107673 }
{ "_id" : "Nature", "count" : 282264 }
>
```

Result

Results show that, the human actions are to be blamed for most wildfires (11,87,673)

Query 4

Question 4:

What are the bottom two unit types that reported wildfires in each county in the US?

Assumptions

We have considered the Source_reporting_unit_name(agencies, wildlife refuge etc) as the forest name in our dataset.

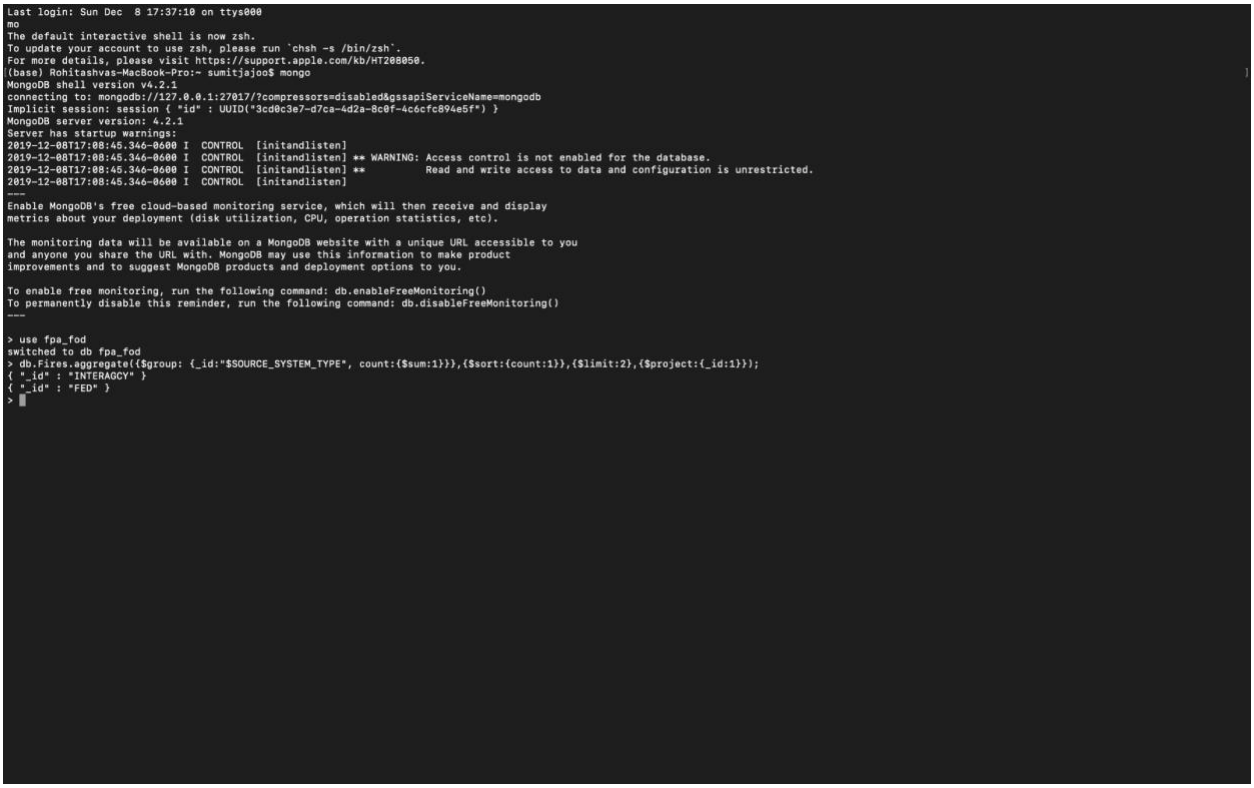
Translation

Select the bottom 2 underperforming unit types which reported the least number of fires in each county in US

Cleanup

Project the Source_system_type, count the fires, sort by ascending, display the 1st two

Screen Shot of MongoDB Query/Code and Results



Result

These are the most underperforming unit types.

Query 5

Question 5:

How many wildfires were not reported by more than one unit/agency?

Assumptions

We have considered the Source_reporting_unit_name (agencies, wildlife refuge etc) as the forest name in our dataset.

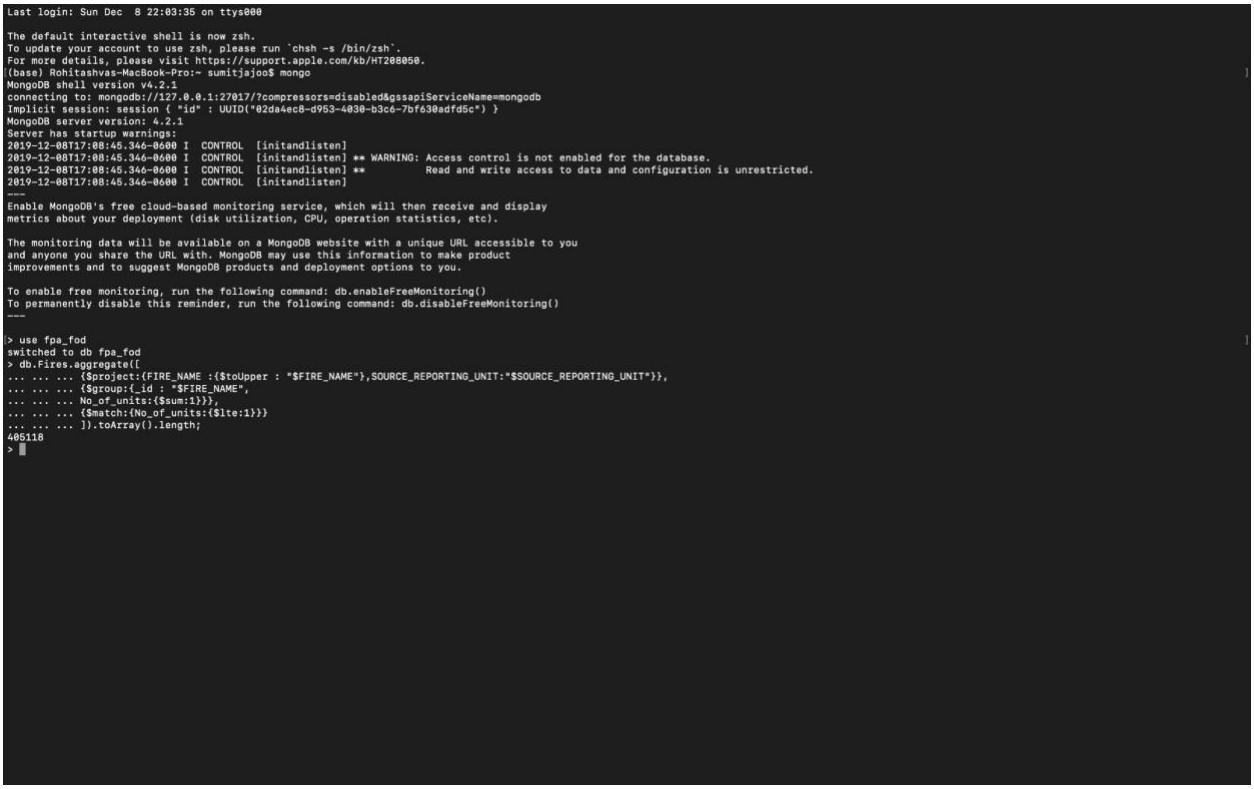
Translation

Select the fires names that were not reported by any source_reporting_unit/agency

Cleanup

Project fire name, where the fire was not reported by any of the source_reporting_unit/agency

Screen Shot of MongoDB Query/Code and Results



Result

4,05,118 wildfires were not reported by more than one agency/unit.

Query 6

Question 8:

Which forest had the least number of fires?

Assumptions

We have considered the Source_reporting_unit_name (agencies, wildlife refuge etc) as the forest name in our dataset

Translation

Select the forest which has least number of fires.

Cleanup

Groupby Source_reporting_unit_name, where fire count =1, in ascending order

Screen Shot of MongoDB Query/Code and Results

```
Last login: Sun Dec  8 21:08:35 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
(base) Rohitashvas-MacBook-Pro:~ sumitjajoo$ mongo
MongoDB shell version v4.2.1
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("bbf22910-2420-4af1-aca5-43ee8863829d") }
MongoDB server version: 4.2.1
Server has startup warnings:
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2019-12-08T17:08:45.346-0600 I  CONTROL  [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---

> use fpa_fod
switched to db fpa_fod
> db.Fires.aggregate({$group: {_id:"$SOURCE_REPORTING_UNIT_NAME", count:{$sum:1}}},{$sort:{count:1}},{$limit:1},{$project:{_id:1}});
{ "_id" : "Sanborn County" }
> db.Fires.aggregate({$group: {_id:"$SOURCE_REPORTING_UNIT_NAME", count:{$sum:1}}},{$sort:{count:1}});
{ "_id" : "Corson County", "count" : 1 }
{ "_id" : "BENTON LAKE WMD", "count" : 1 }
{ "_id" : "COUNCIL GROVE FIRE DEPT", "count" : 1 }
{ "_id" : "OVERFLOW NWR", "count" : 1 }
{ "_id" : "San Marcos Fire Protection Department", "count" : 1 }
{ "_id" : "Essex Junction Fire Department", "count" : 1 }
{ "_id" : "Port Mansfield Volunteer Fire Department", "count" : 1 }
{ "_id" : "James C. Campbell National Wildlife Refuge", "count" : 1 }
{ "_id" : "ST. CATHERINE CREEK NATIONAL WILDLIFE REFUGE", "count" : 1 }
{ "_id" : "PICTURED ROCKS NATIONAL LAKESHORE", "count" : 1 }
{ "_id" : "CENTERVILLE VFD", "count" : 1 }
{ "_id" : "MTN LONGLEAF NWR", "count" : 1 }
{ "_id" : "Sheffield VFD & EMS", "count" : 1 }
{ "_id" : "Sagamore Hill National Historic Site", "count" : 1 }
{ "_id" : "Marathon VFD", "count" : 1 }
{ "_id" : "KEYTESVILLE FD", "count" : 1 }
{ "_id" : "Southwest Highway 115 Fire Protection District", "count" : 1 }
{ "_id" : "Moran Rural VFD", "count" : 1 }
{ "_id" : "COLE CAMP CITY & RFD", "count" : 1 }
{ "_id" : "Cowpens National Battlefield", "count" : 1 }
Type "it" for more
> it
{ "_id" : "Garfield Township Fire Department", "count" : 1 }
{ "_id" : "Colonial National Historical Park", "count" : 1 }
{ "_id" : "Pleasant Township Fire Department", "count" : 1 }
{ "_id" : "Roosevelt/Vanderbilt National Historic Site", "count" : 1 }
{ "_id" : "Arizona - Tucson District", "count" : 1 }
{ "_id" : "Brockton Fire Department", "count" : 1 }
{ "_id" : "Sharon Fire Department", "count" : 1 }
{ "_id" : "Melissa VFD", "count" : 1 }
{ "_id" : "Logan County Fire Department #1", "count" : 1 }
```

Result

Above are the forests with least number of fires.