**Group 15**

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

BUAN 6320

**Group Members**

Garima Gupta

Guru Prasad Kumar

Rohitashva Jajoo

Pratik Gupta

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Activity** | **Garima** | **Guru Prasad** | **Rohitashva** | **Pratik** |
| Prepared Data Model and Created Physical DB | x | x | x | x |
| Loaded Data into Database | x | x | x | x |
| Prepared Report | x | x | x | x |
| Reviewed Report | x | x | x | x |

Contents

Data Model 5

Assumptions/Notes About Data Entities and Relationships 5

Entity-Relationship Diagram 6

Physical Database 7

Assumptions/Notes About Data Set 7

Screen shot of Physical Database objects 7

Data in the Database 12

SQL Queries 13

Query 1 13

Question 13

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 13

Translation 13

Screen Shot of SQL Query and Results 13

Query 2 14

Question 14

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 14

Translation 14

Screen Shot of SQL Query and Results 14

Query 3 15

Question 15

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 15

Translation 15

Screen Shot of SQL Query and Results 15

Query 4 16

Question 16

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 16

Translation 16

Screen Shot of SQL Query and Results 16

Query 5 17

Question 17

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 17

Translation 17

Screen Shot of SQL Query and Results 17

Query 6 18

Question 18

Notes/Comments About SQL Query and Results (Include # of Rows in Result) 18

Translation 18

Screen Shot of SQL Query and Results 18

Data Review for MongoDB 19

Physical Mongo Database 19

Assumptions/Notes About Data Set 19

Screen shot of Physical Database objects (Database, Collections and Attributes) 19

Data in the Database 21

MongoDB Queries/Code 22

Query 1 22

Question 22

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 22

Translation 22

Screen Shot of MongoDB Query/Code and Results 22

Query 2 23

Question 23

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 23

Translation 23

Screen Shot of MongoDB Query/Code and Results 23

Query 3 24

Question 24

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 24

Translation 24

Screen Shot of MongoDB Query/Code and Results 24

Query 4 25

Question 25

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 25

Translation 25

Screen Shot of MongoDB Query/Code and Results 25

Query 5 26

Question 26

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 26

Translation 26

Screen Shot of MongoDB Query/Code and Results 27

Query 6 28

Question 28

Notes/Comments About MongoDB Query/Code and Results (Include # of Documents in Result) 28

Translation 28

Screen Shot of MongoDB Query/Code and Results 28

# Data Model

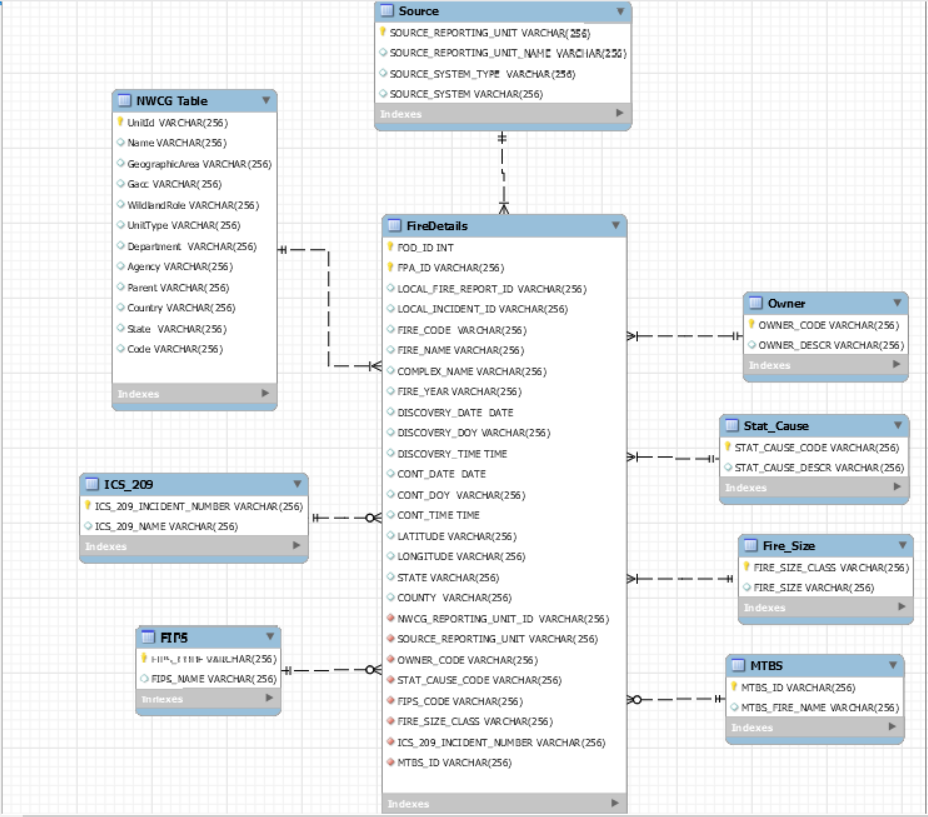
## Assumptions/Notes About Data Entities and Relationships

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



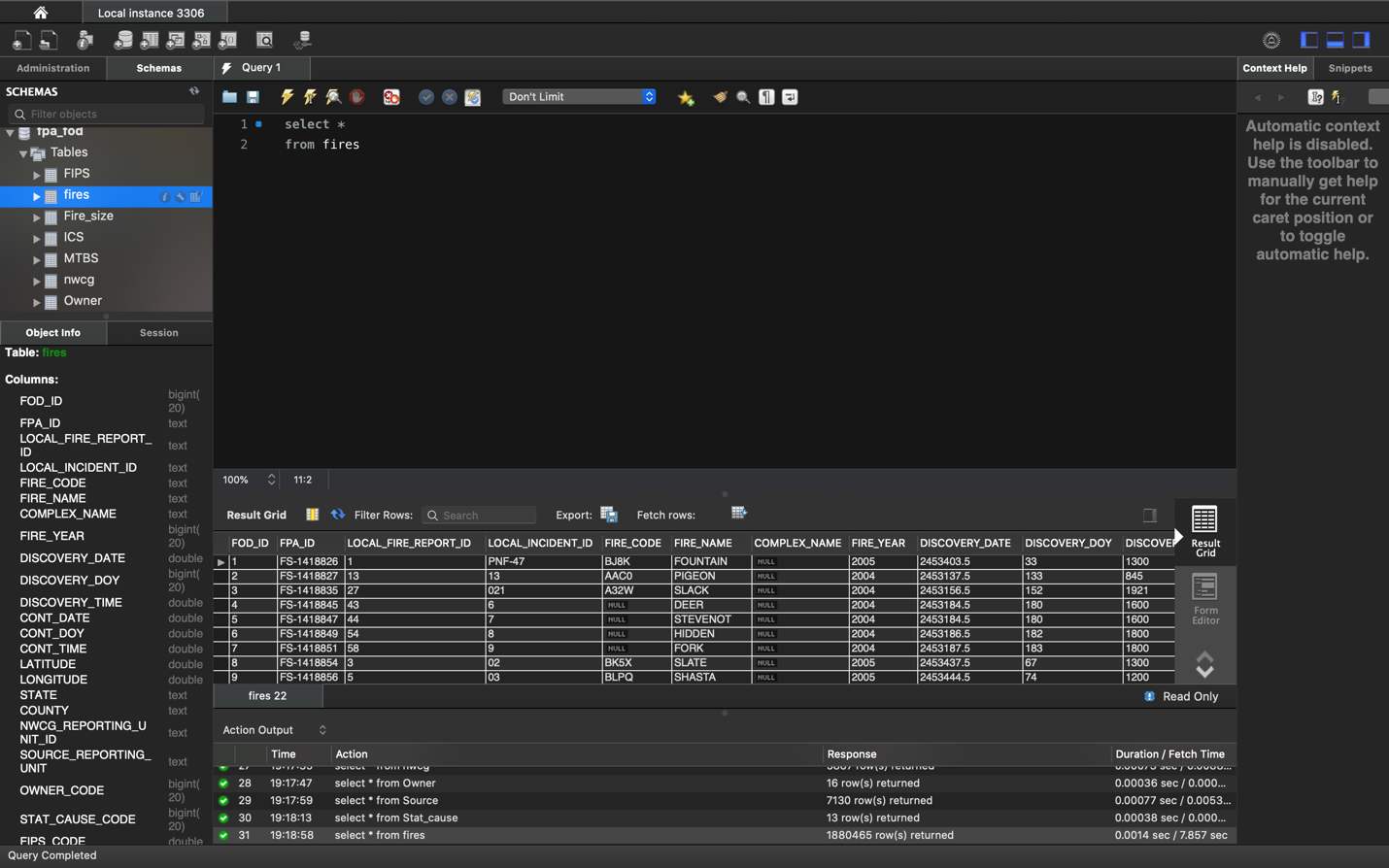
# Physical Database

## Assumptions/Notes About Data Set

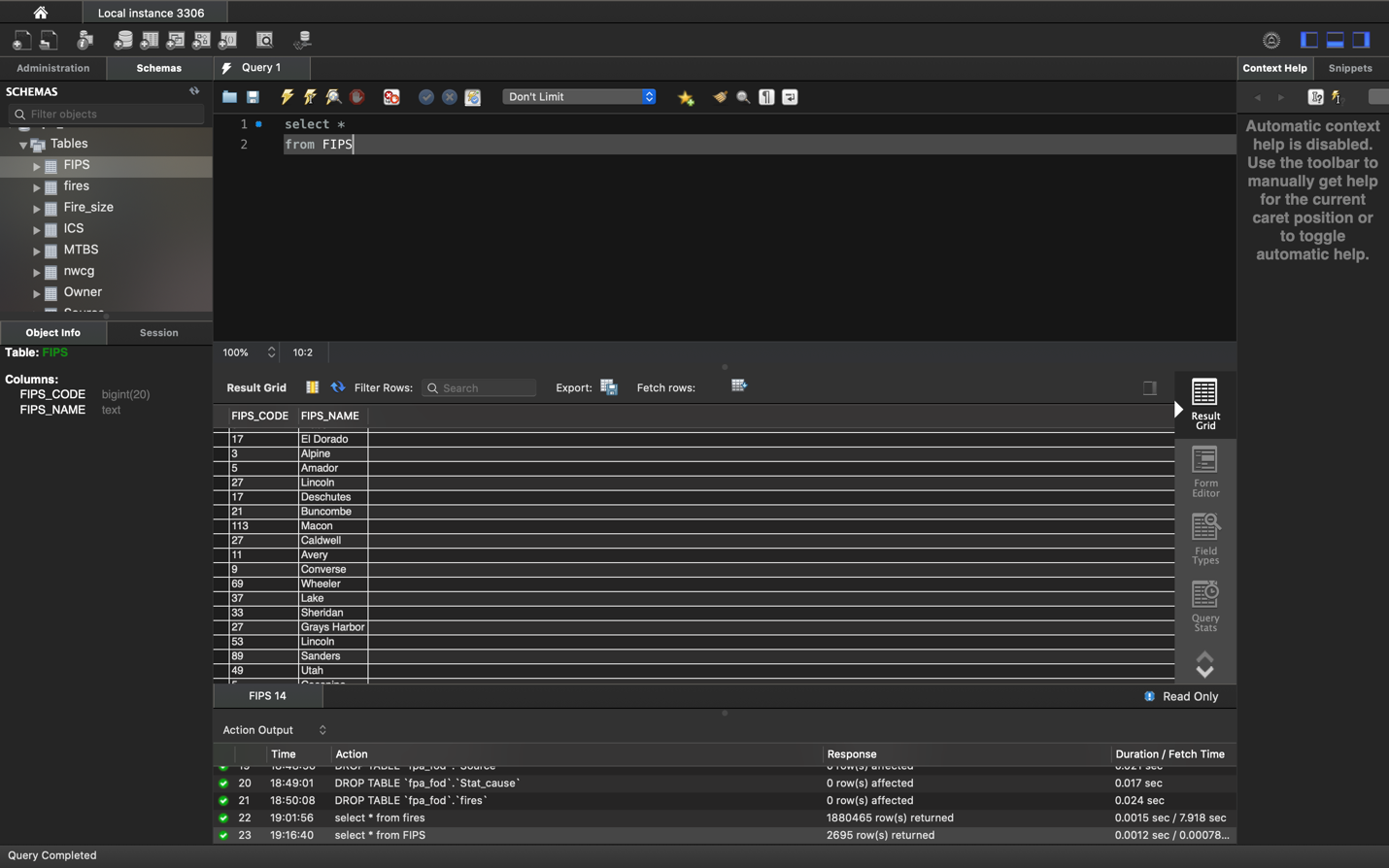
1. All records on fire data are contained in Fire Details table
2. All records on NWCG reporting agency are kept on NWCG table
3. NWGC agency and its individual units generates the fire report for the fires originating.
4. Fire report is generated by the source reporting unit indicating the source of 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

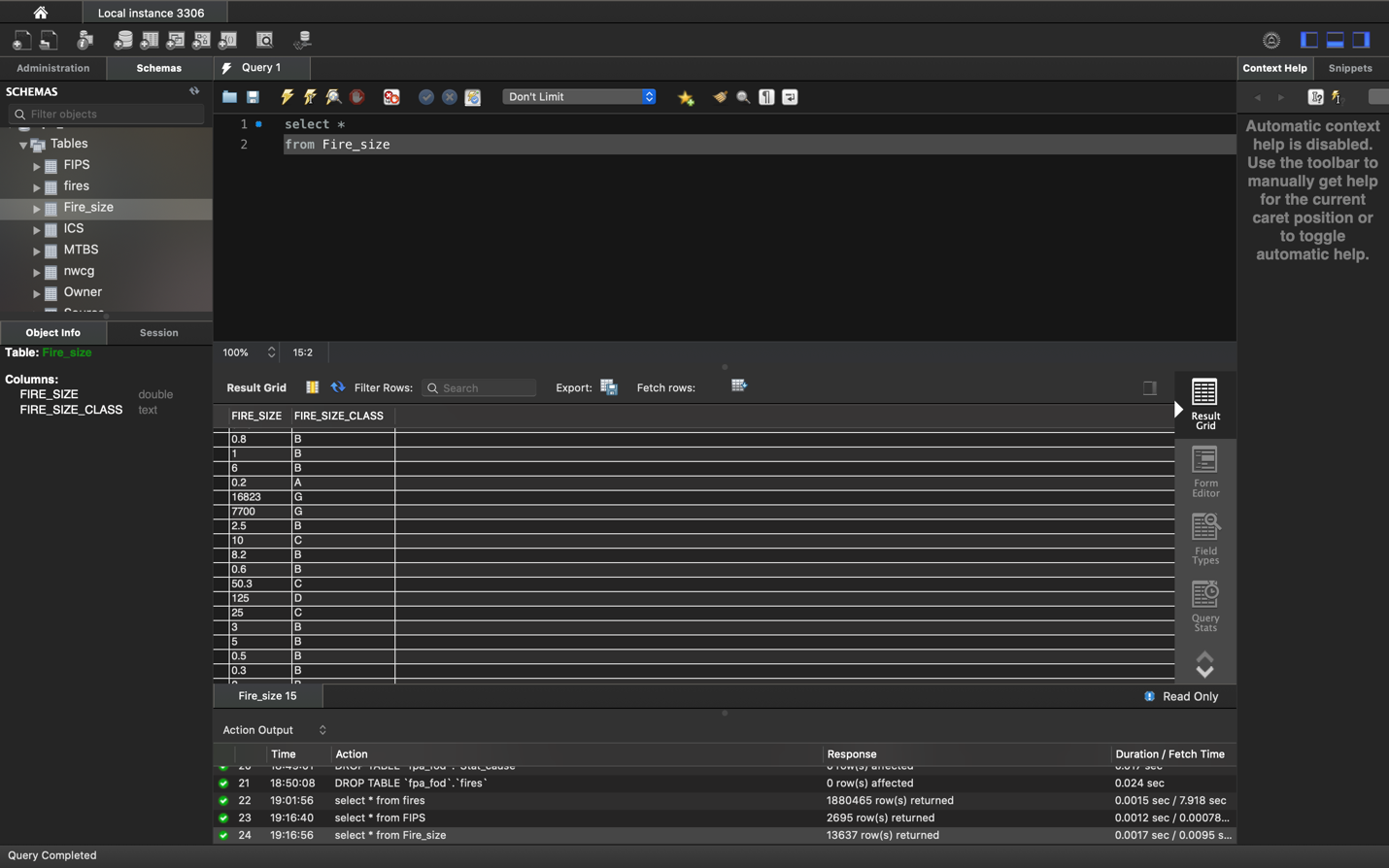
Fires Table:



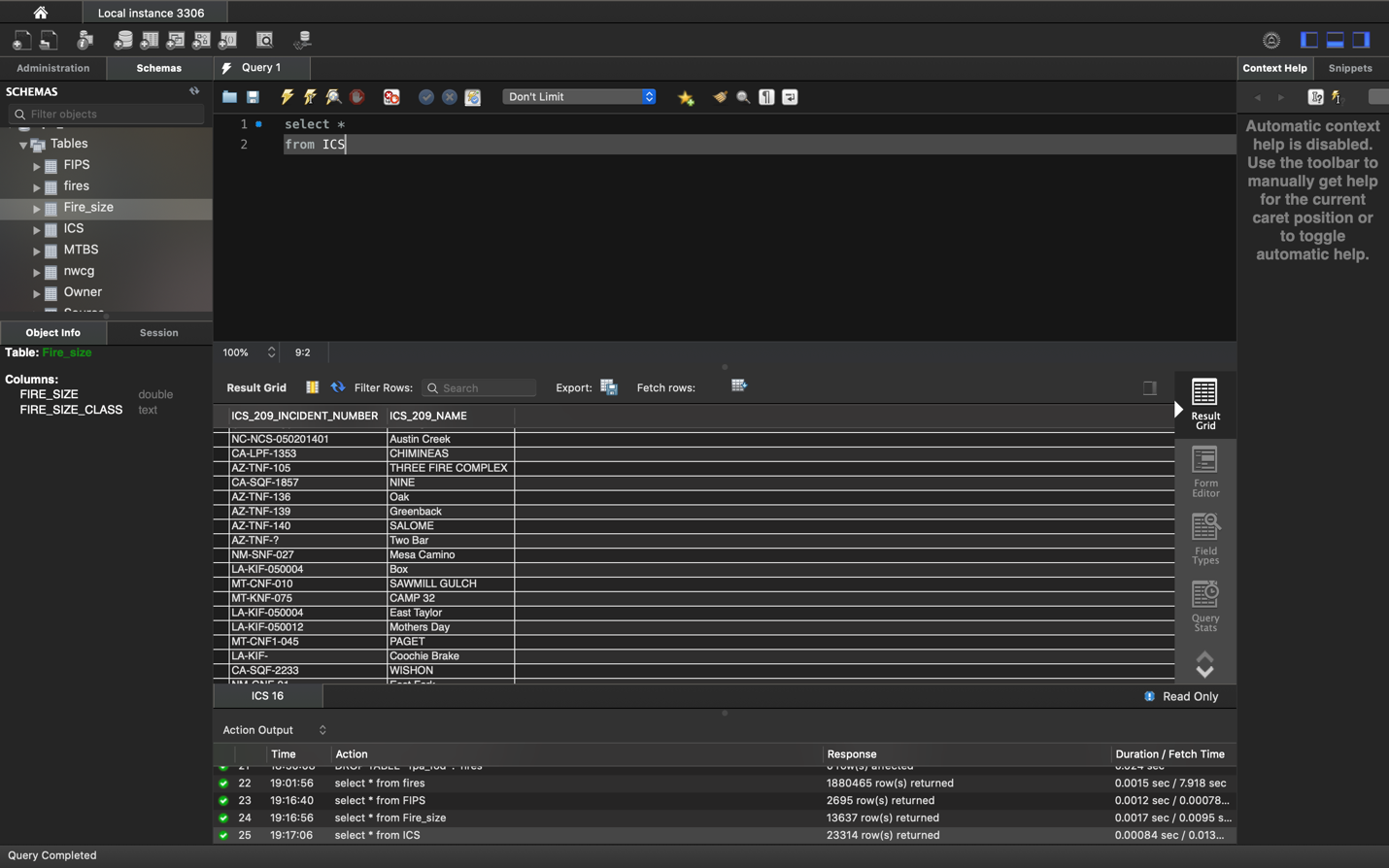
FIPS Table



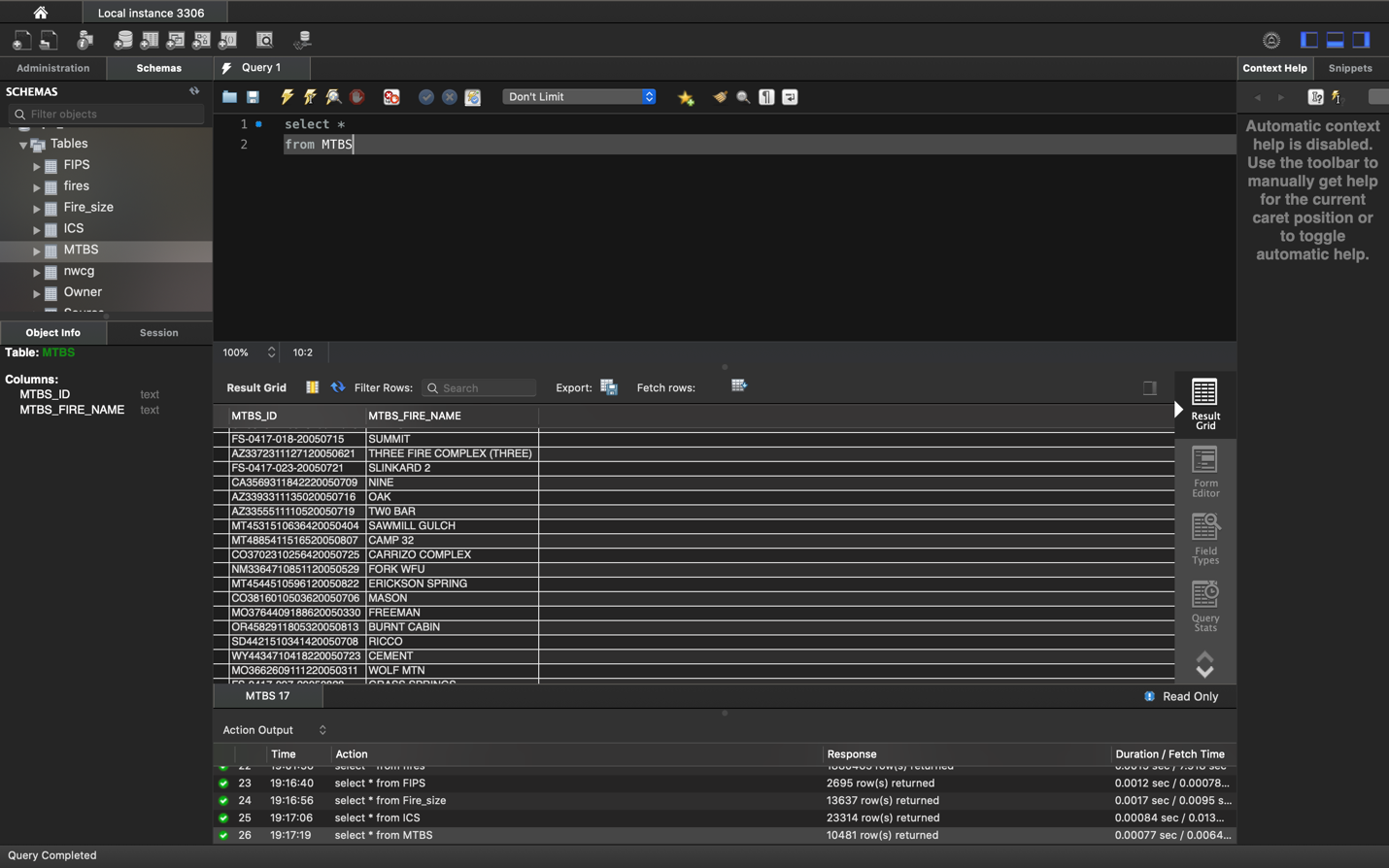
Fire Size Table



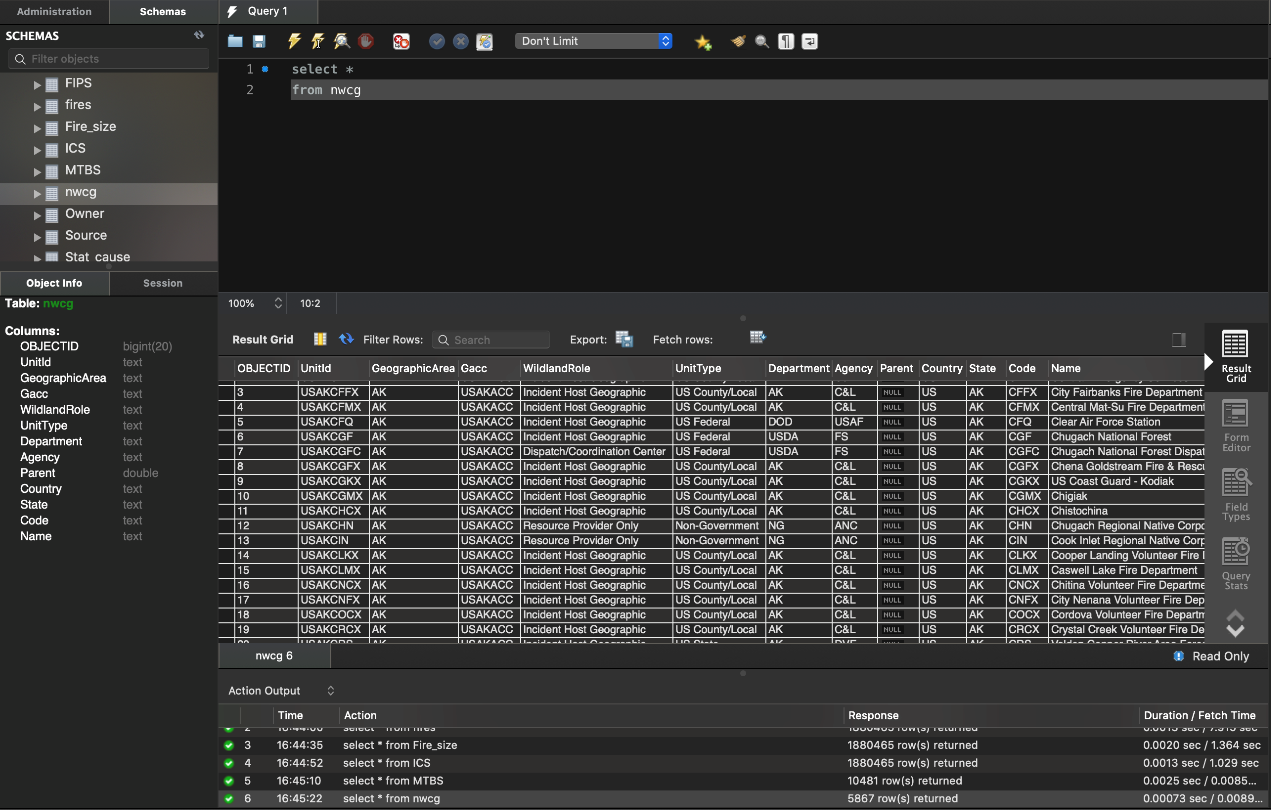
ICS Table



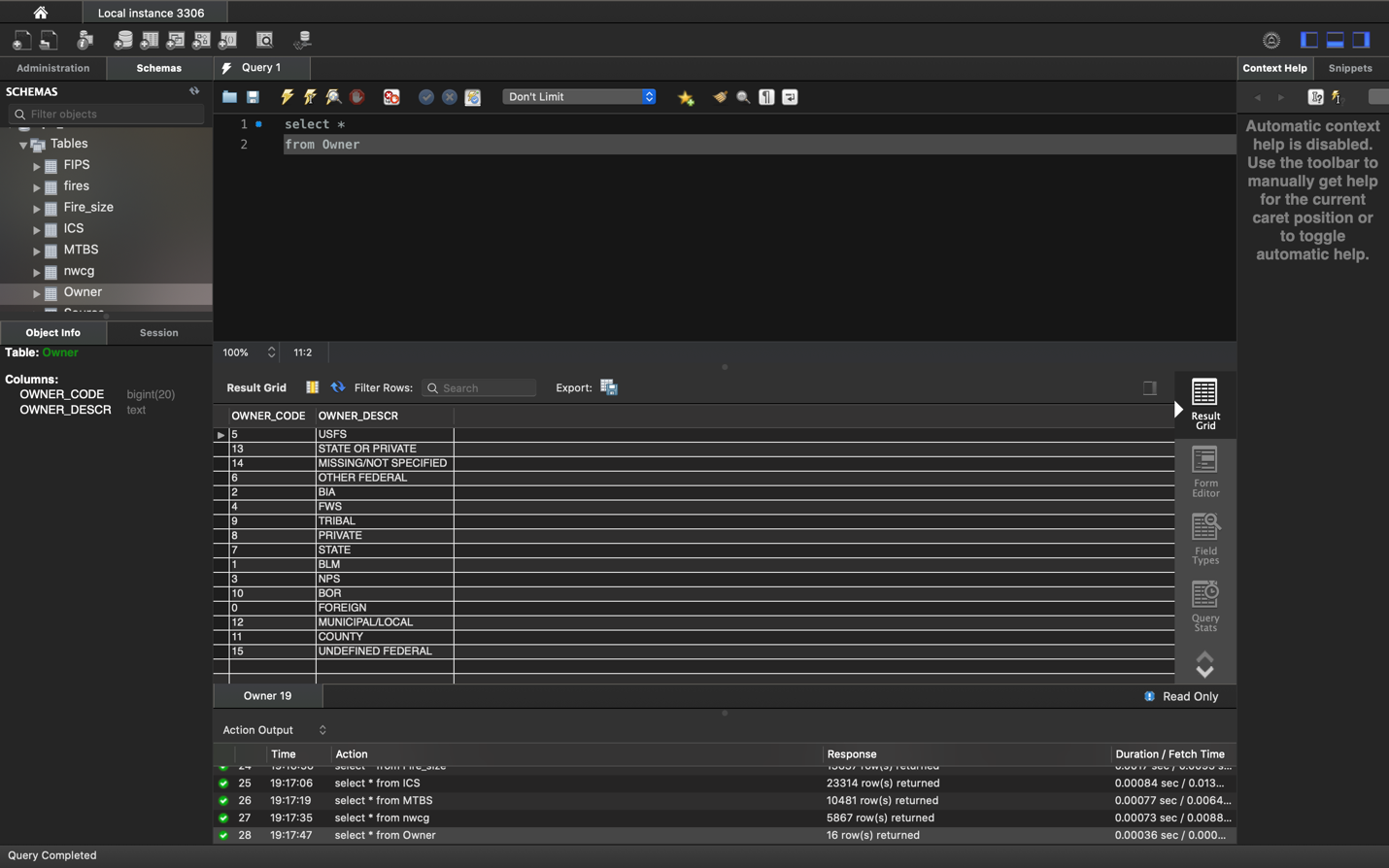
MTBS Table



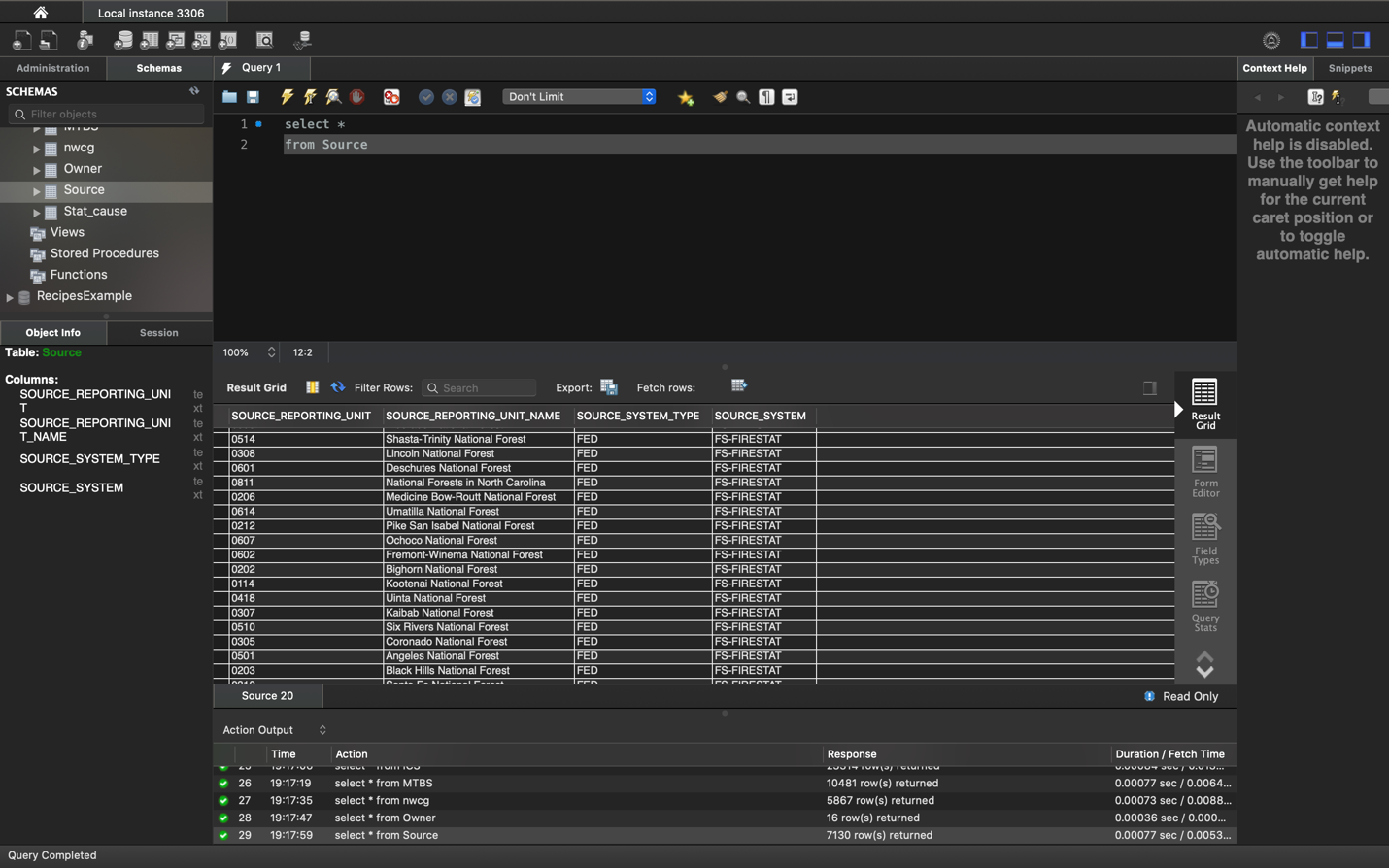
NWCG Table



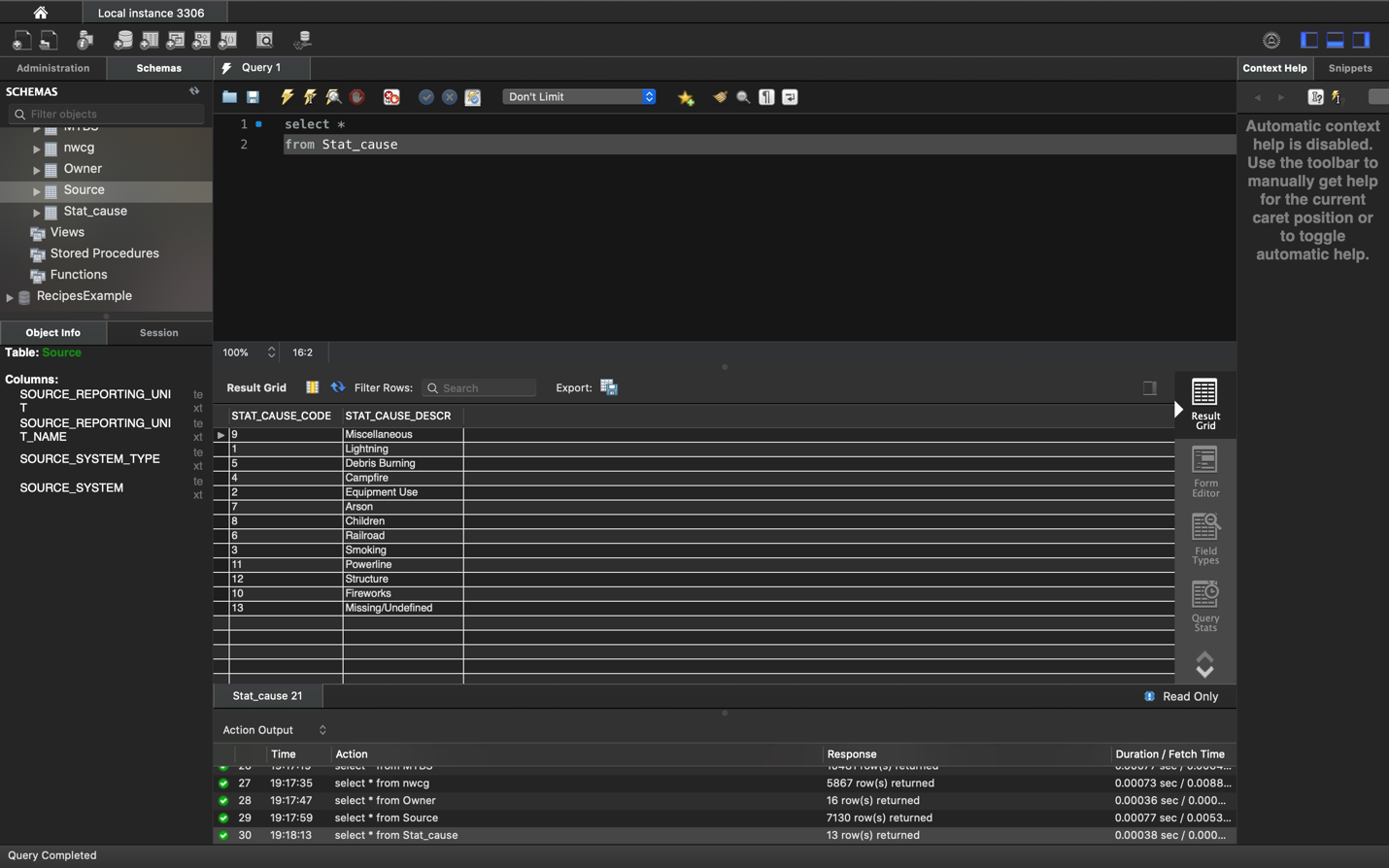
Owner Table



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.

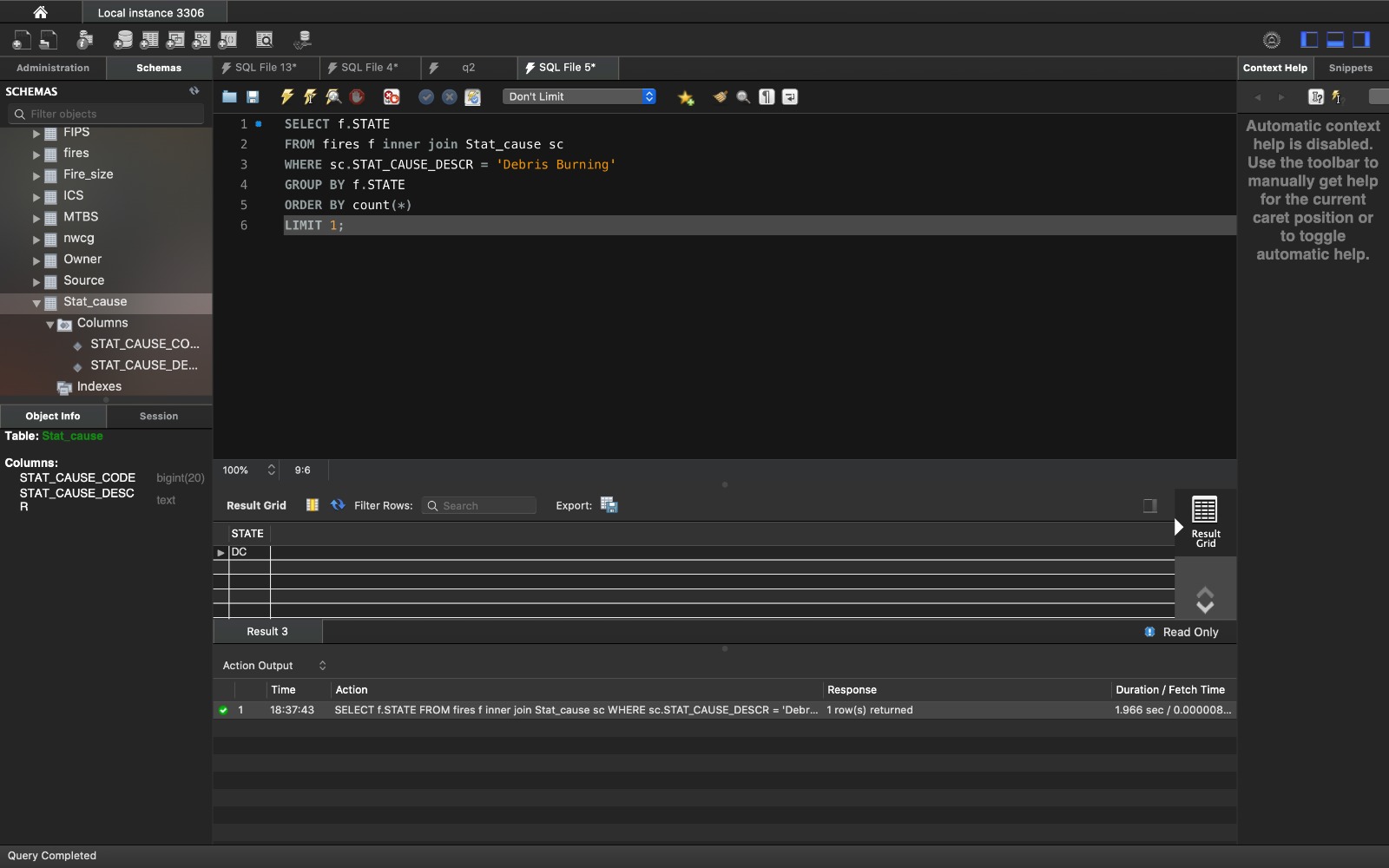
### Translation

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.

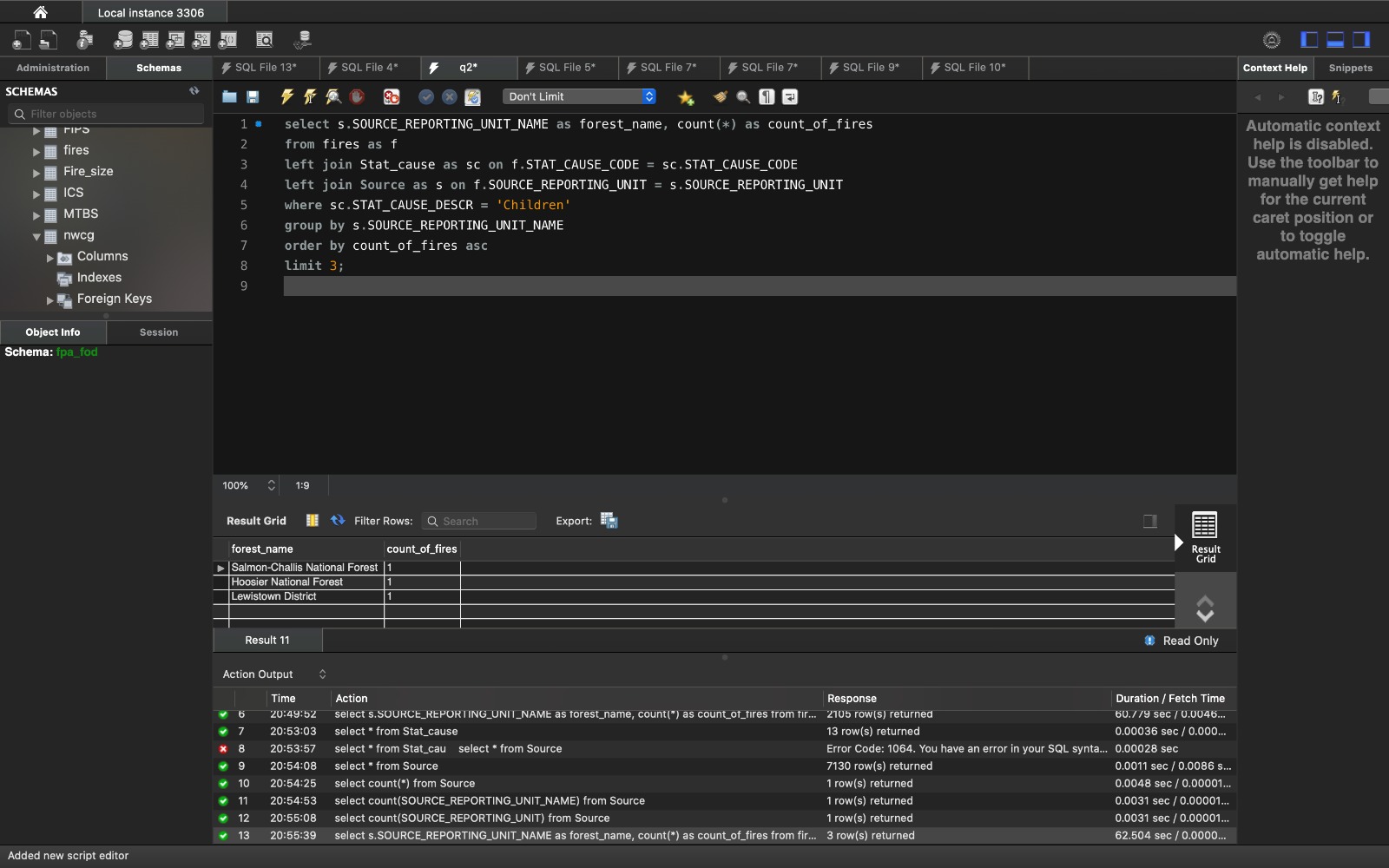
### Translation

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

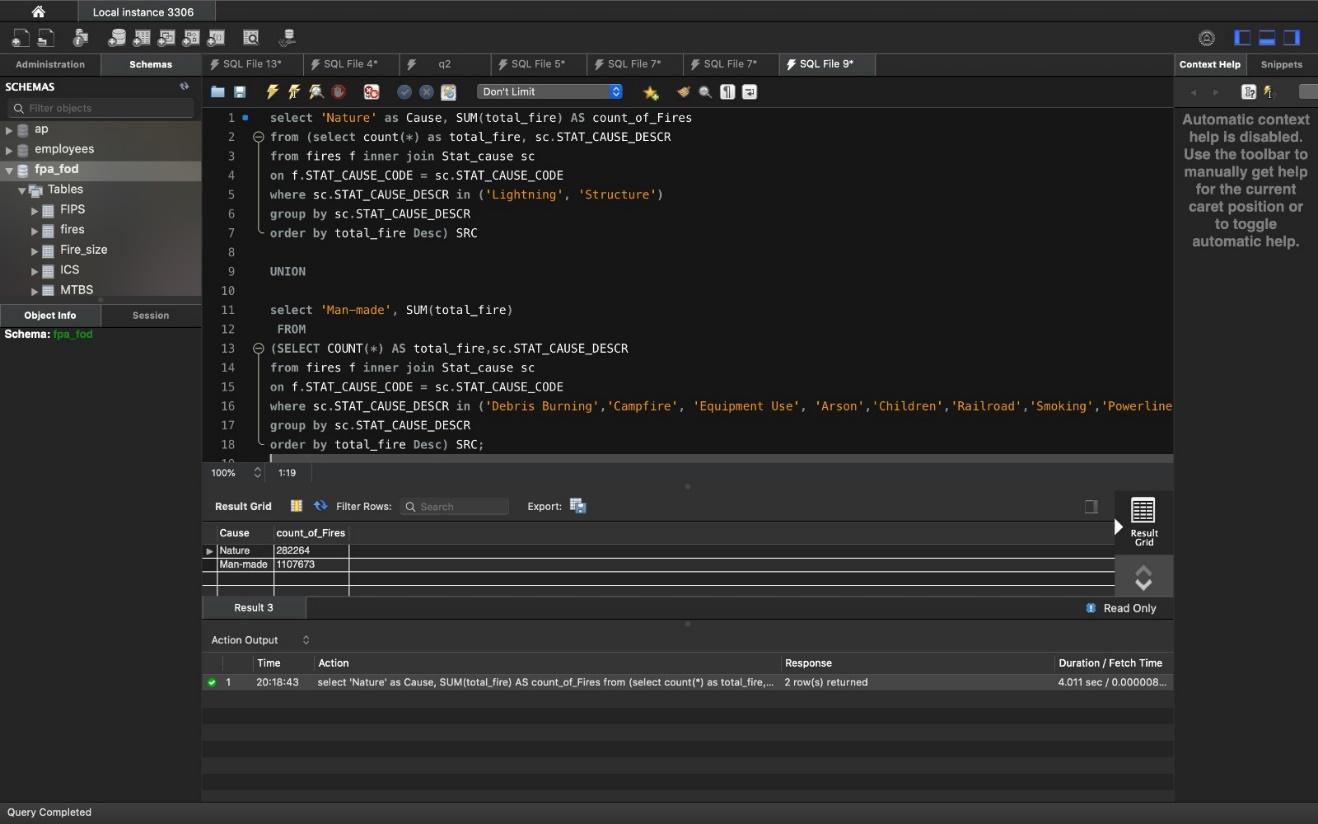
### Cleanup

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.

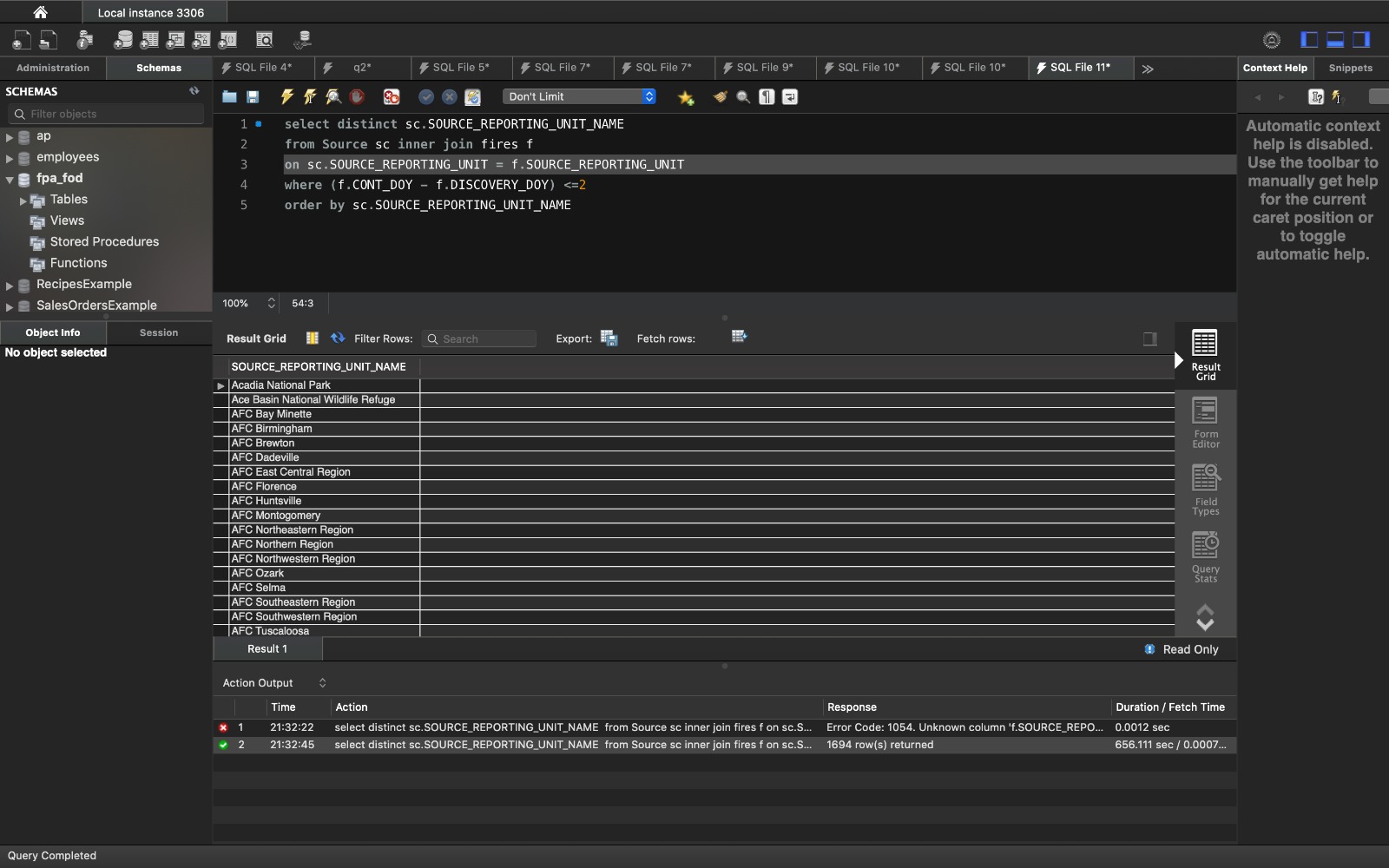
### Translation

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.

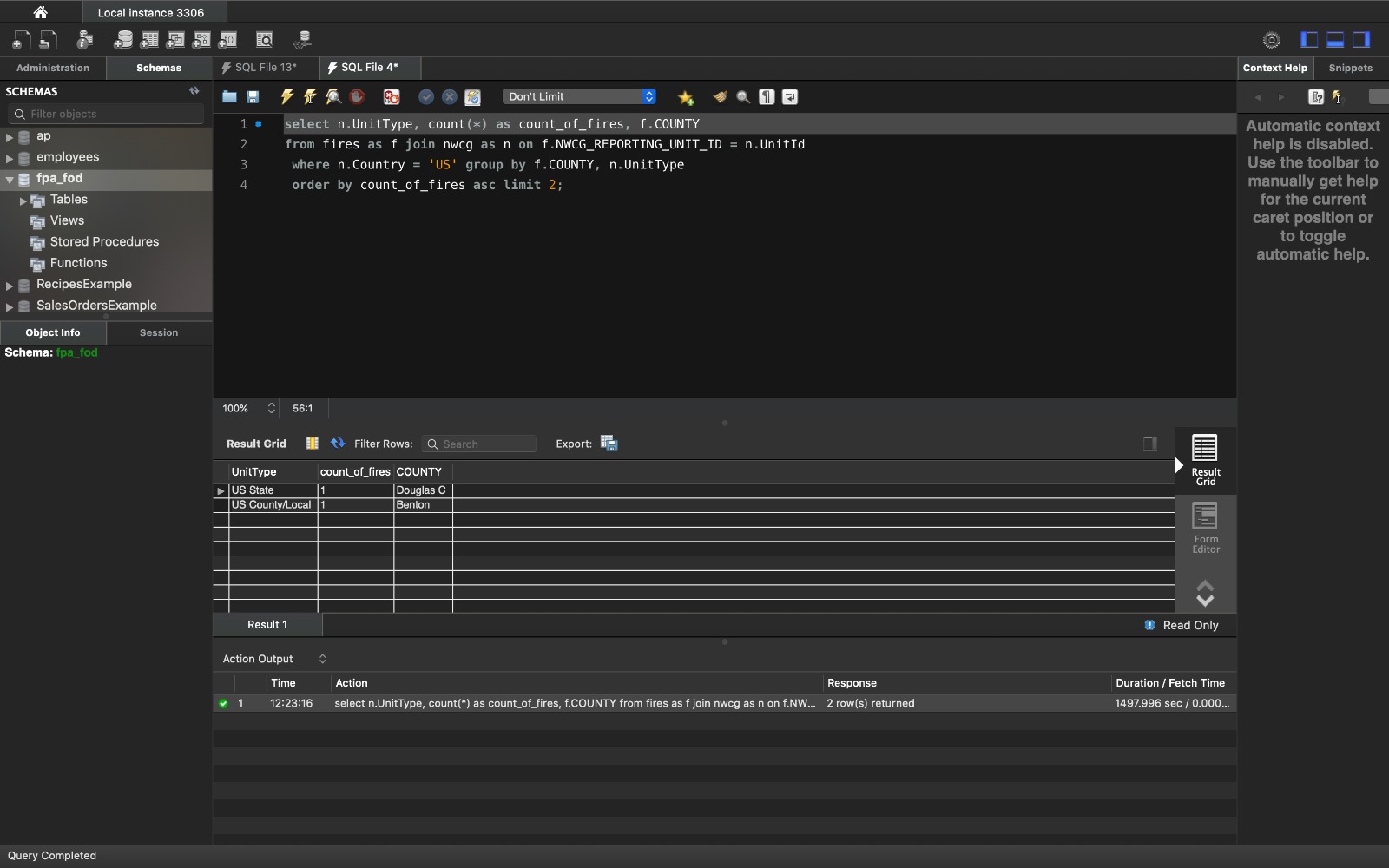
### Translation

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.

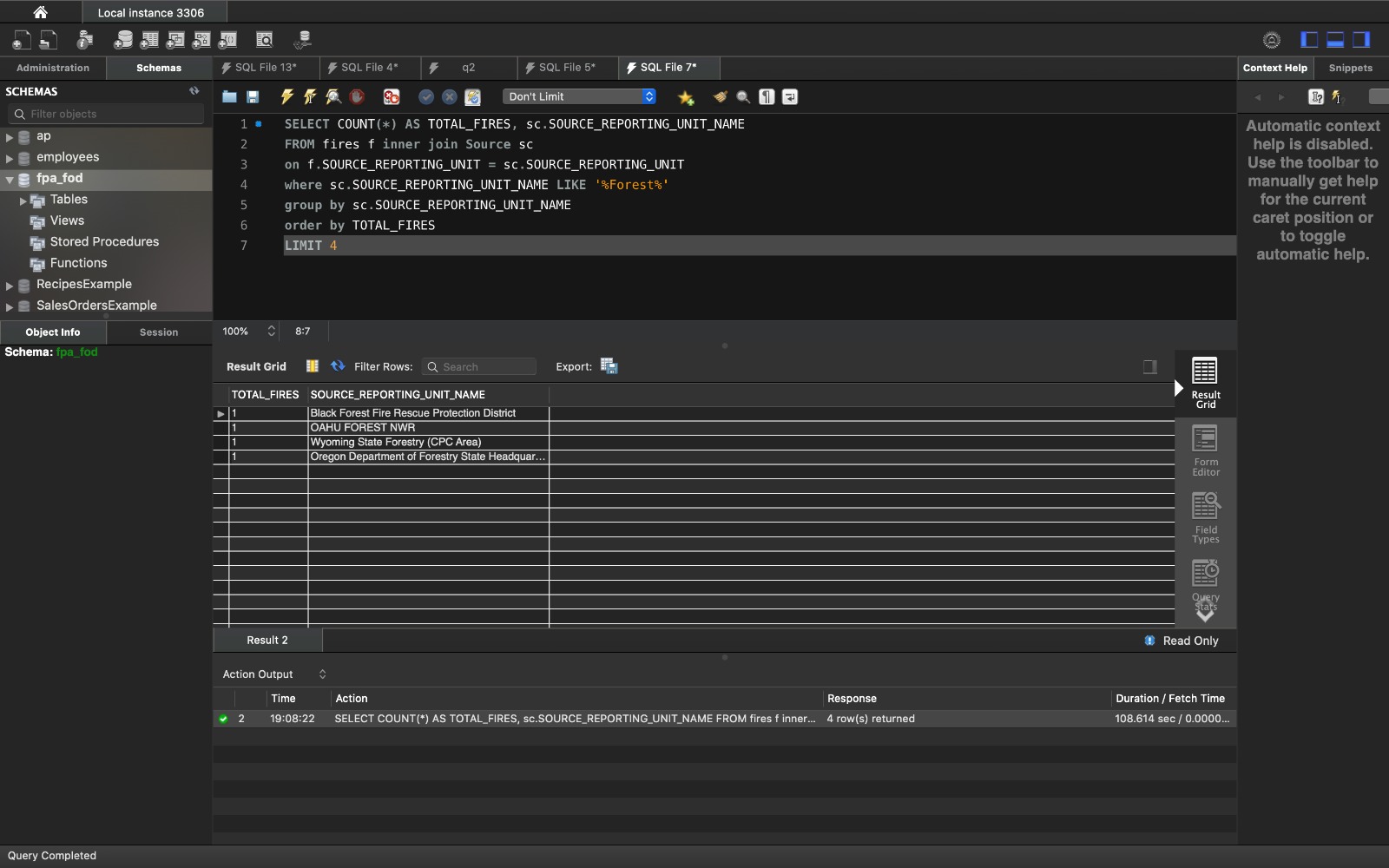
### Translation

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.

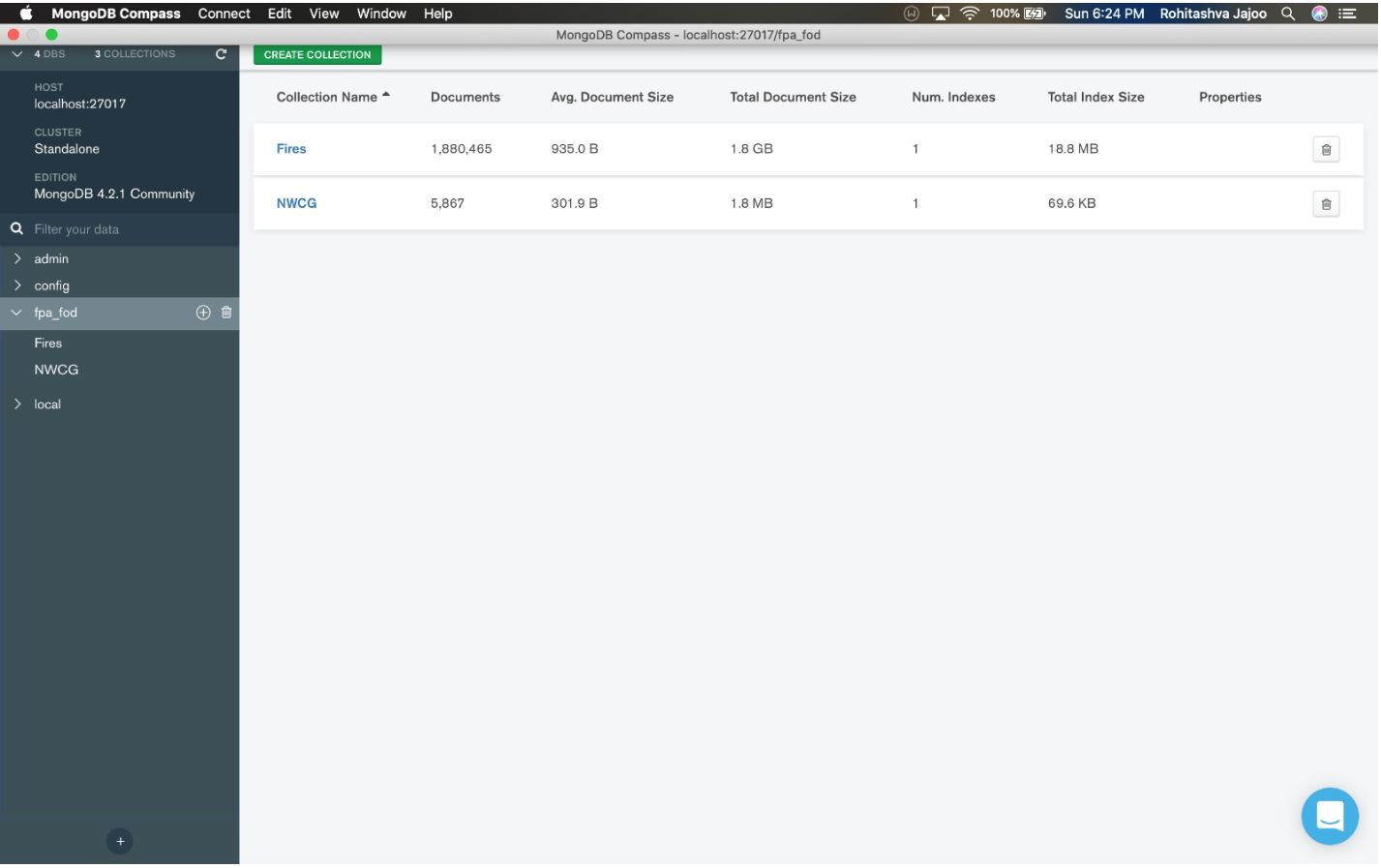
# Data Review for MongoDB

## Physical Mongo Database

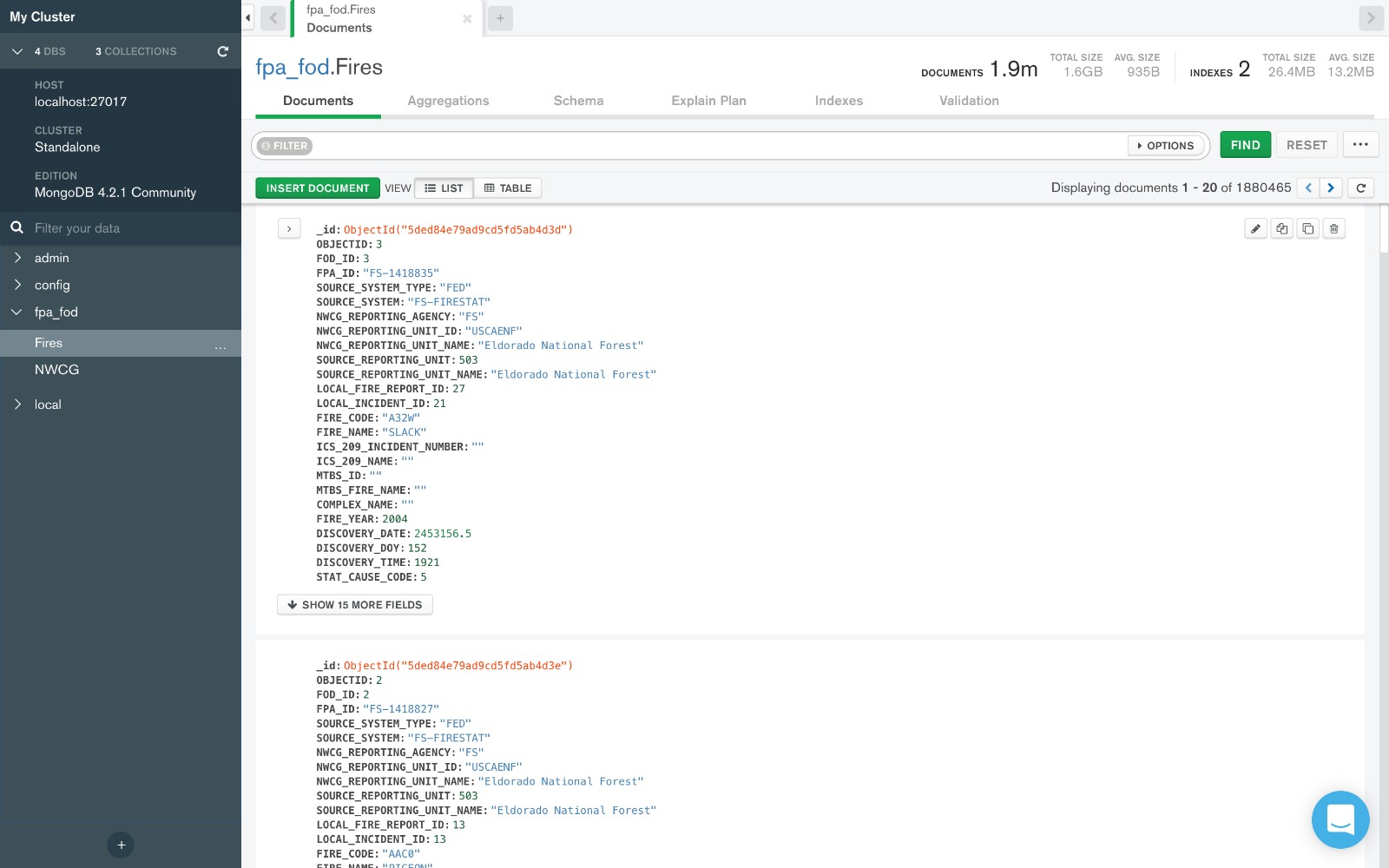
## Assumptions/Notes About Data Set

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

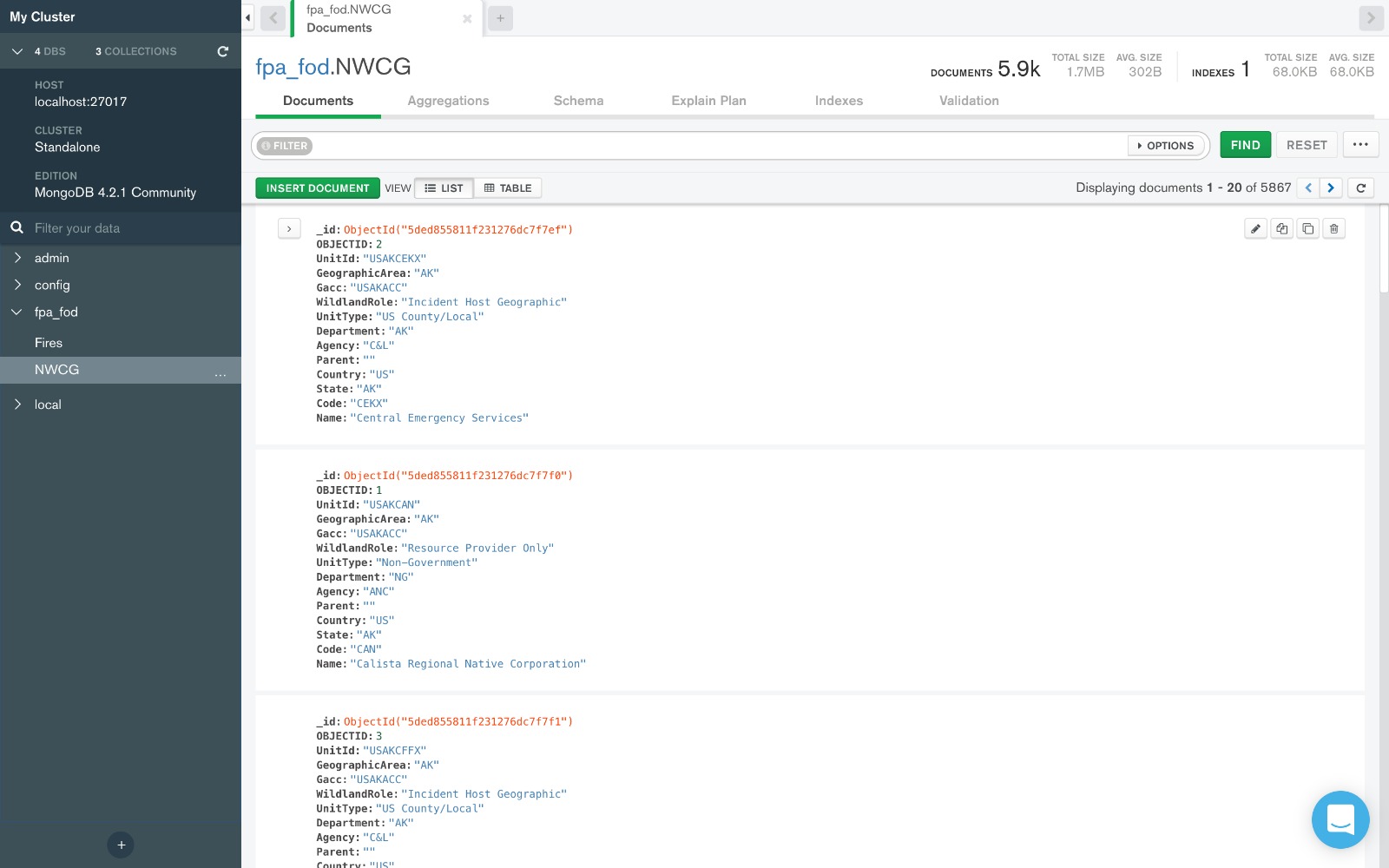
## Screen shot of Physical Database objects (Database, Collections and Attributes)



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.

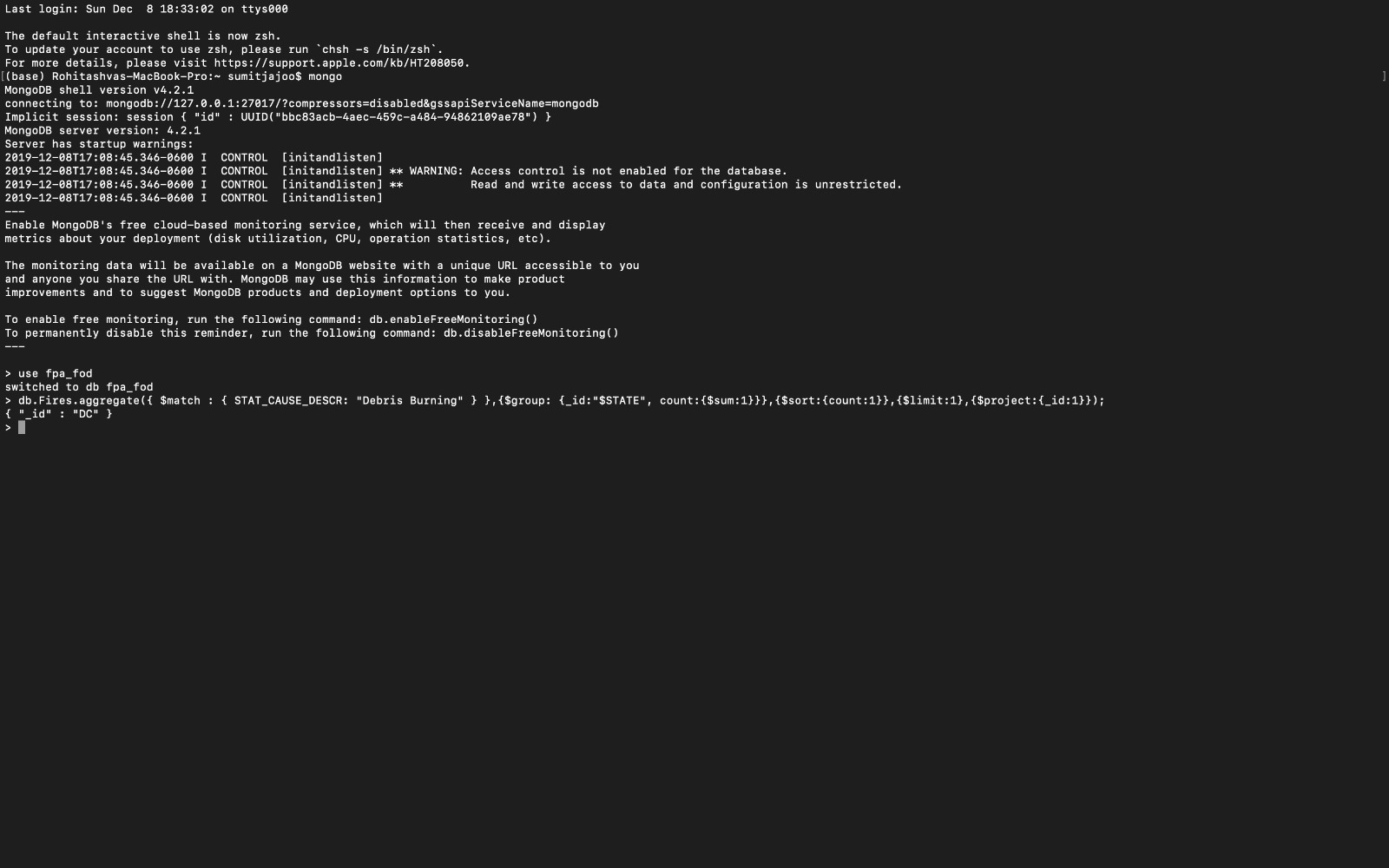
### Translation

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.

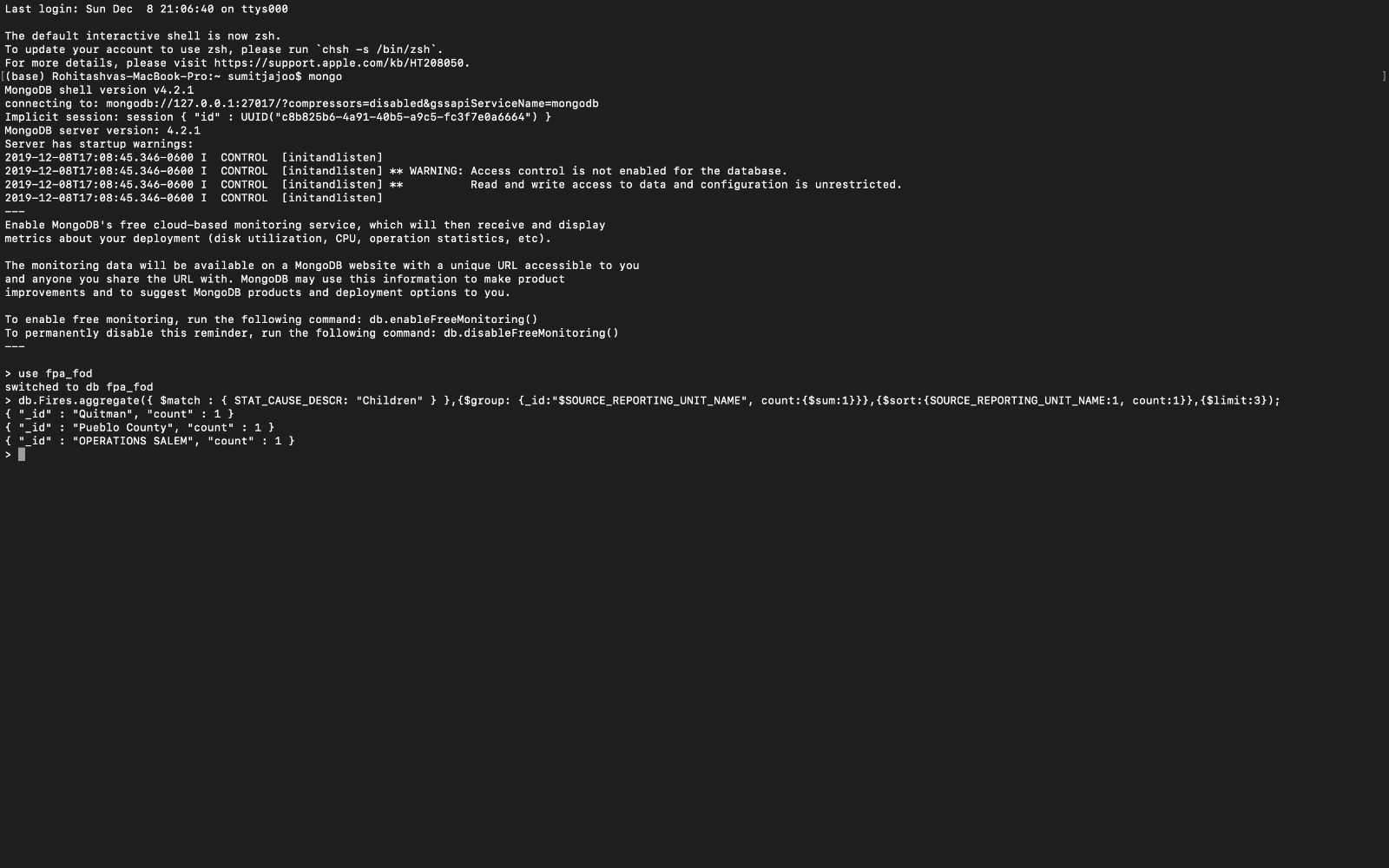
### Translation

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



### 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')

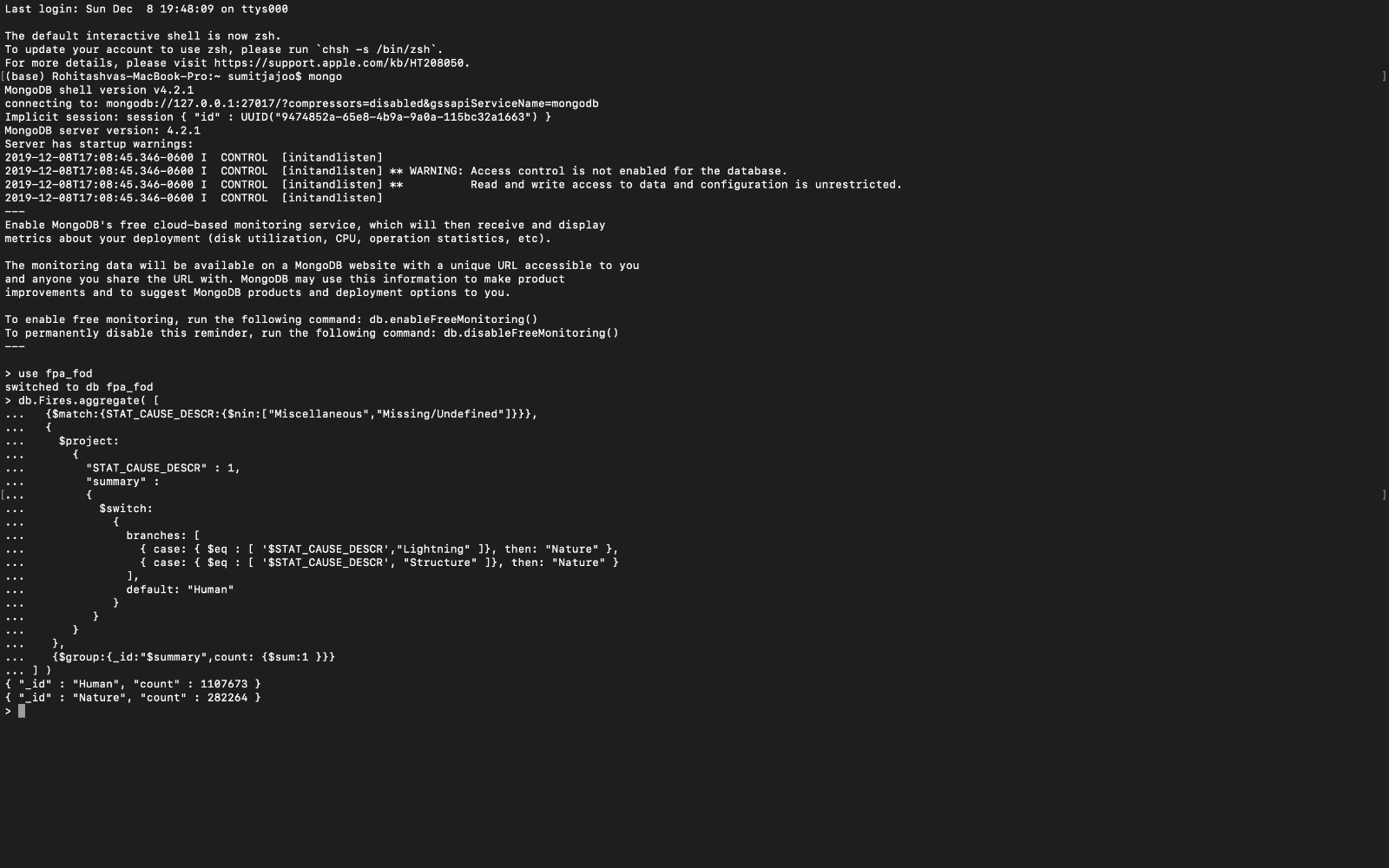
### Translation

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



### 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.

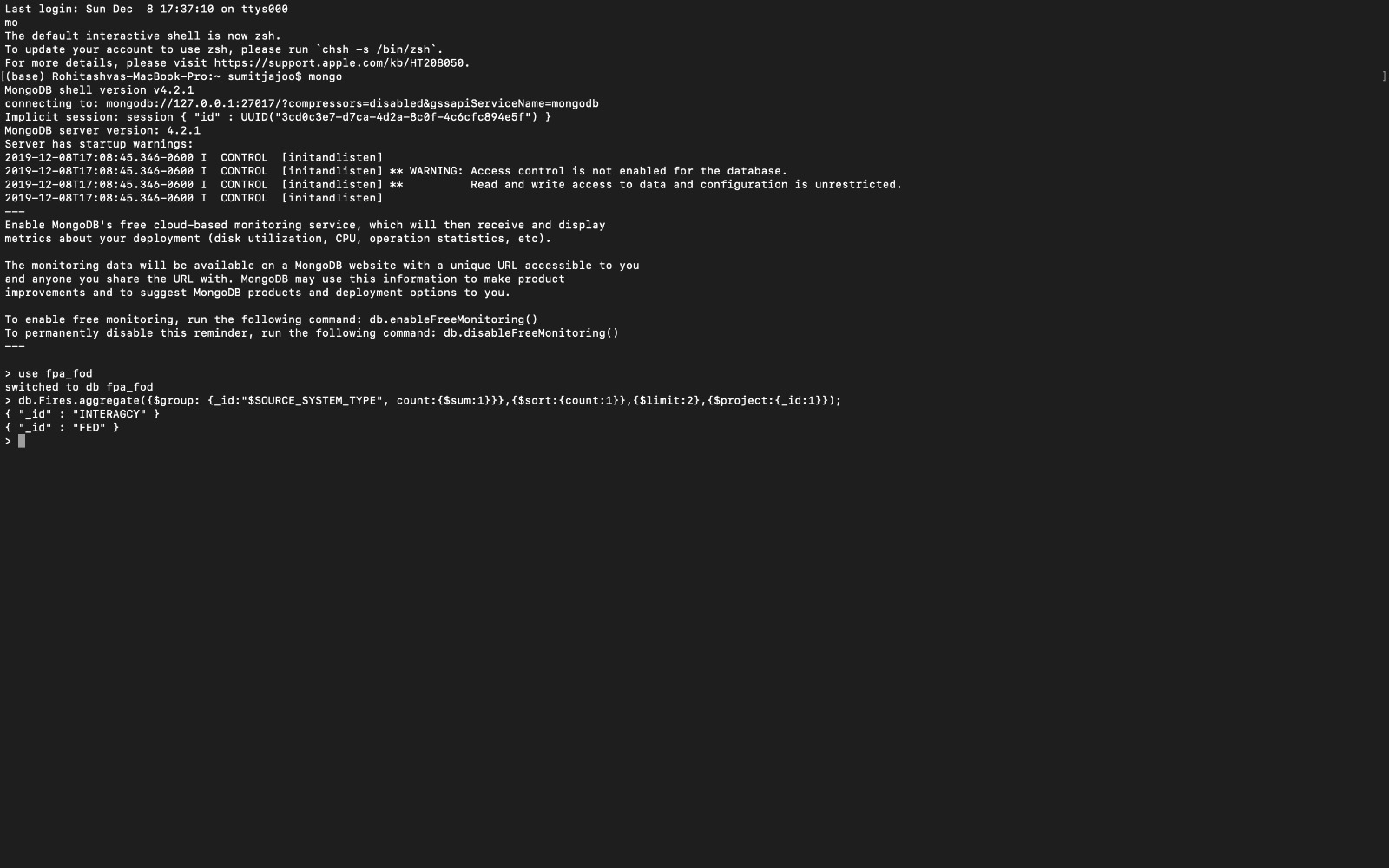
### Translation

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.

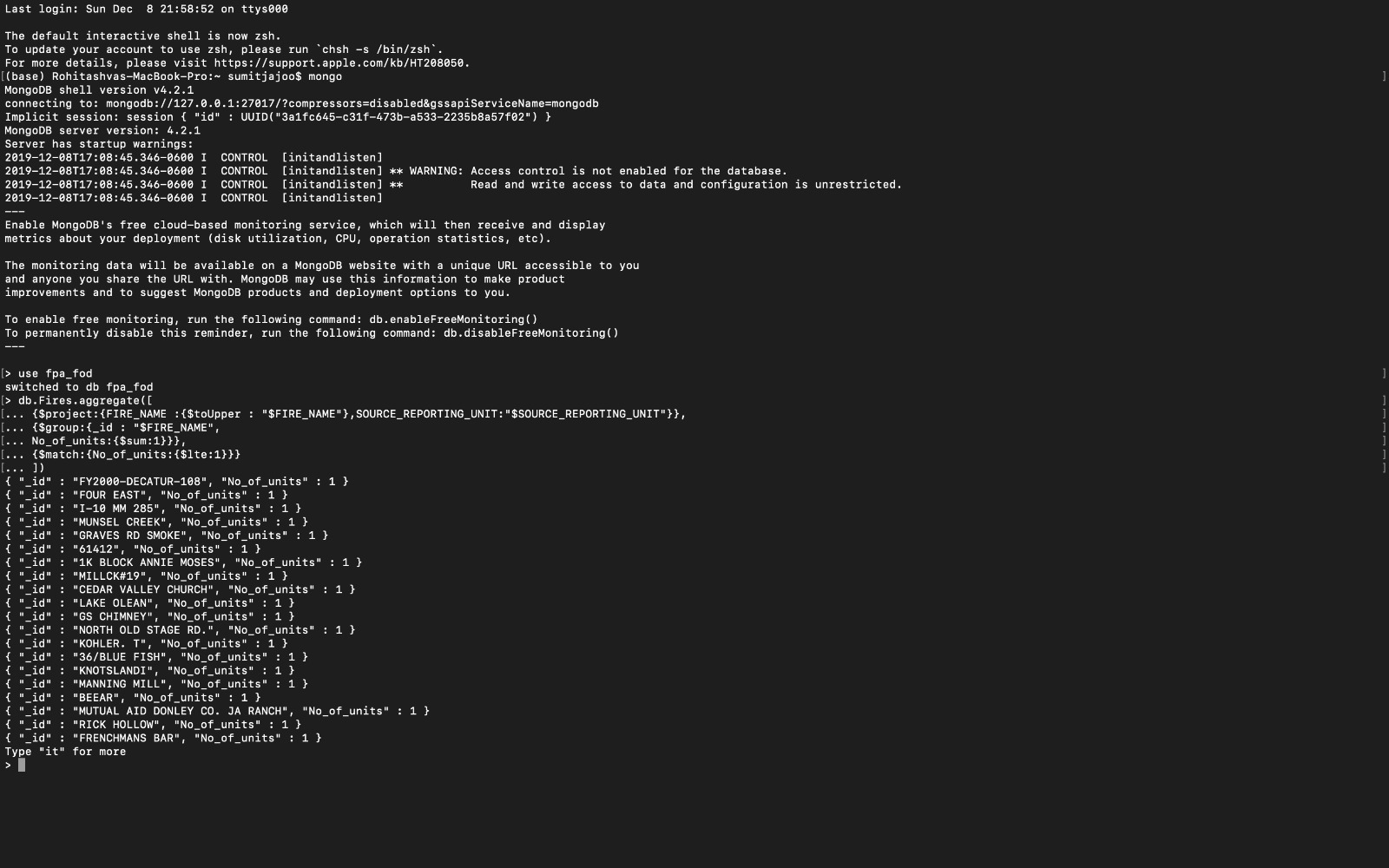
### Translation

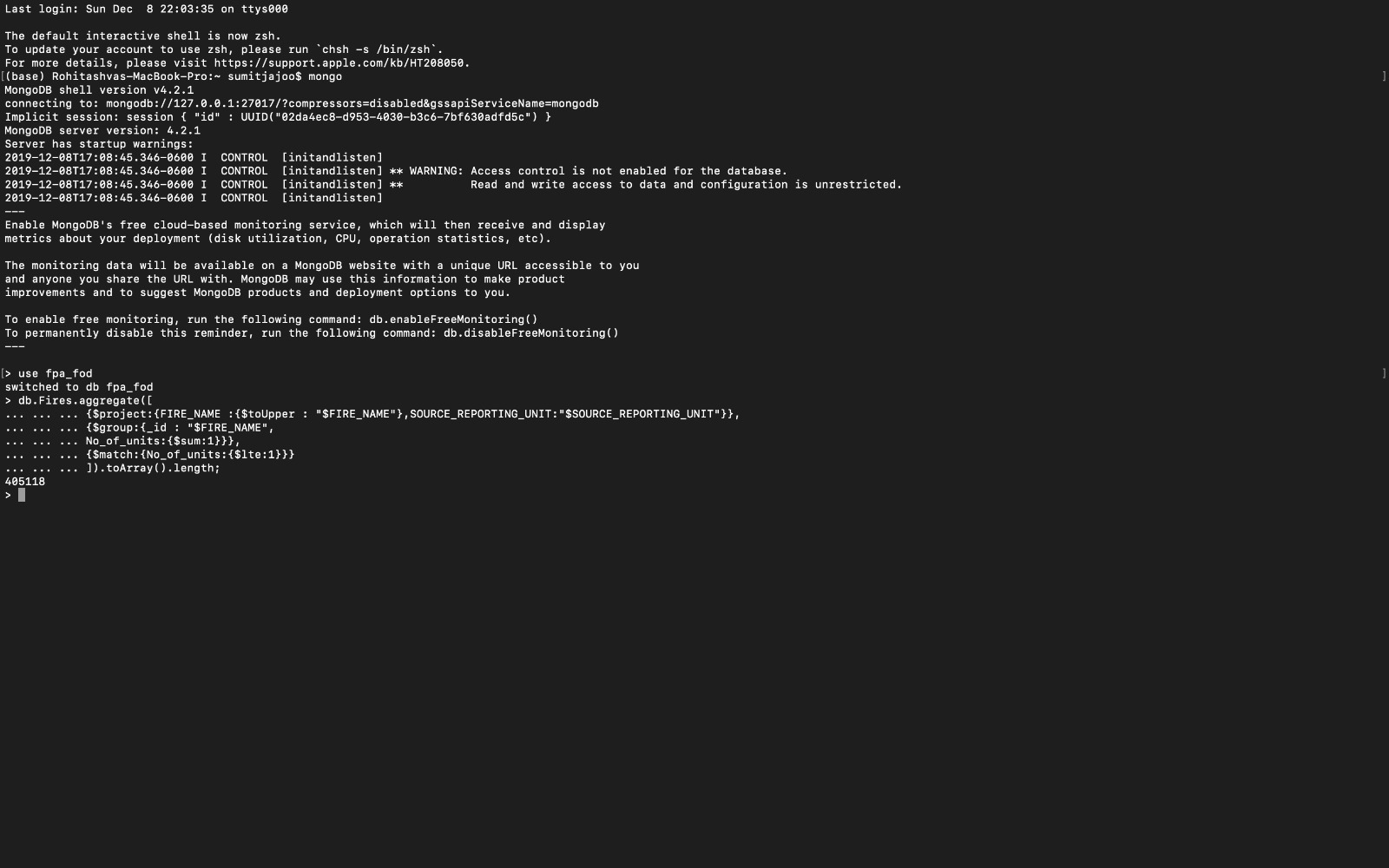
Select the fires names that were not reported by any source\_reporting\_unit/agency

### 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

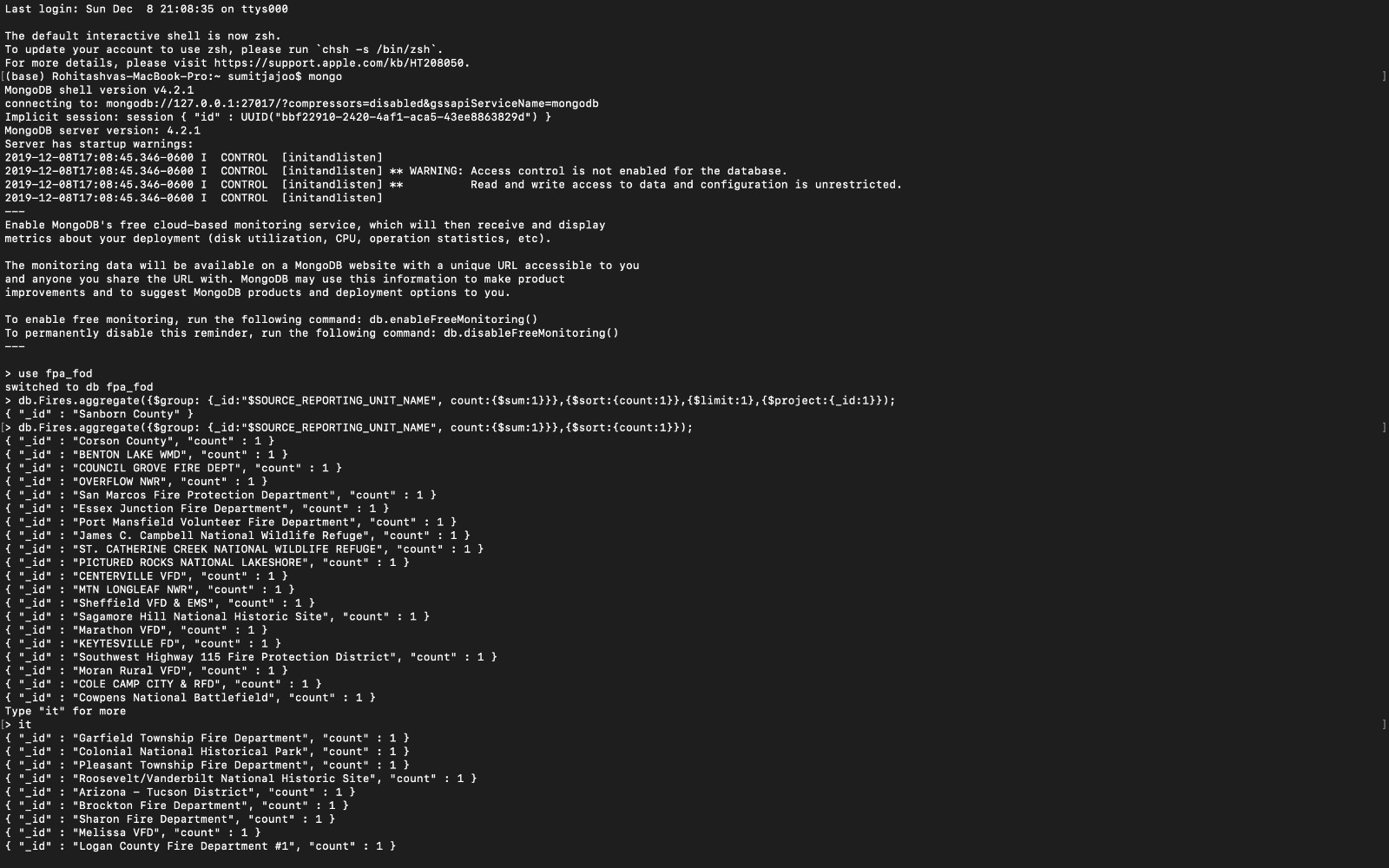
### Translation

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



### Result

Above are the forests with least number of fires.