SQL-Mongo Project – Spatial Data of US Wildfires

BUAN 6320

Submitted by:

Prem Kumar Pulluri

Activity
Prepared Data Model and Created Physical DB
Loaded Data into Database
Wrote SQL Queries
Prepared Mongo Database
Loaded data into Mongo DB
Wrote Mongo Queries
Prepared Report
Reviewed Report

Contents

Data Model

Assumptions/Notes About Data Entities and Relationships

Include assumptions about data entities and their relationships with each other.

1. The relationship between the primary key in the above-mentioned tables with the **fire_table** main table is **one to many relationships** other than Source_System_info as the FPA_ID, which is primary key has single record mapping with the foreign key, all other tables have multiple values repeated in foreign key column.

2. The **nwcg_agency_identifier_table** has the primary key unitID which has a **one to many relationships** to the **Fire_data_Table**.

3. nwcg_location_info has one to many relationship with **nwcg_agency_identifier_table** with State and Gacc Columns as primary composite key.

4. stat_cause_table has one to many relationship with fire_table with stat_cause_code as primary key.

5. location_info has one to many relationship with fire_table with Latitude and Longitude as primary composite key.

6. fips_info has one to many relationship with location_info with FIPS_CODE and FIPS_NAME as composite primary key.

7. mtbs_info has one to many relationship with fire_table with MTBS_ID as primary key.

8. Source_reporting_info has one to many relationship with fire_table with Source_Reporting_Unit as the primary key.

9. Owner_info has one to many relationship with fire_table with Owner_ID as the primary key.

10. Ics_209_info has one to many relationships with fire_table with ICS_209_Incident_number as the primary key with Non Identifying relationship.

11. Each owner may own land on many locations, while every single location can have only one land owner.
12. Many fire incidents may occur due to one statistical cause, each fire incident is caused by only one statistical cause.
13. Each MTBS identifier may identify many fire incidents, each fire is identified by only one MTBS identifier.
14. Each ICS-209 report identifier may identify many fire incidents, each fire is identified by only one MTBS identifier.
15. Each FIPS may represent many locations, each location is represented by one FIPS.

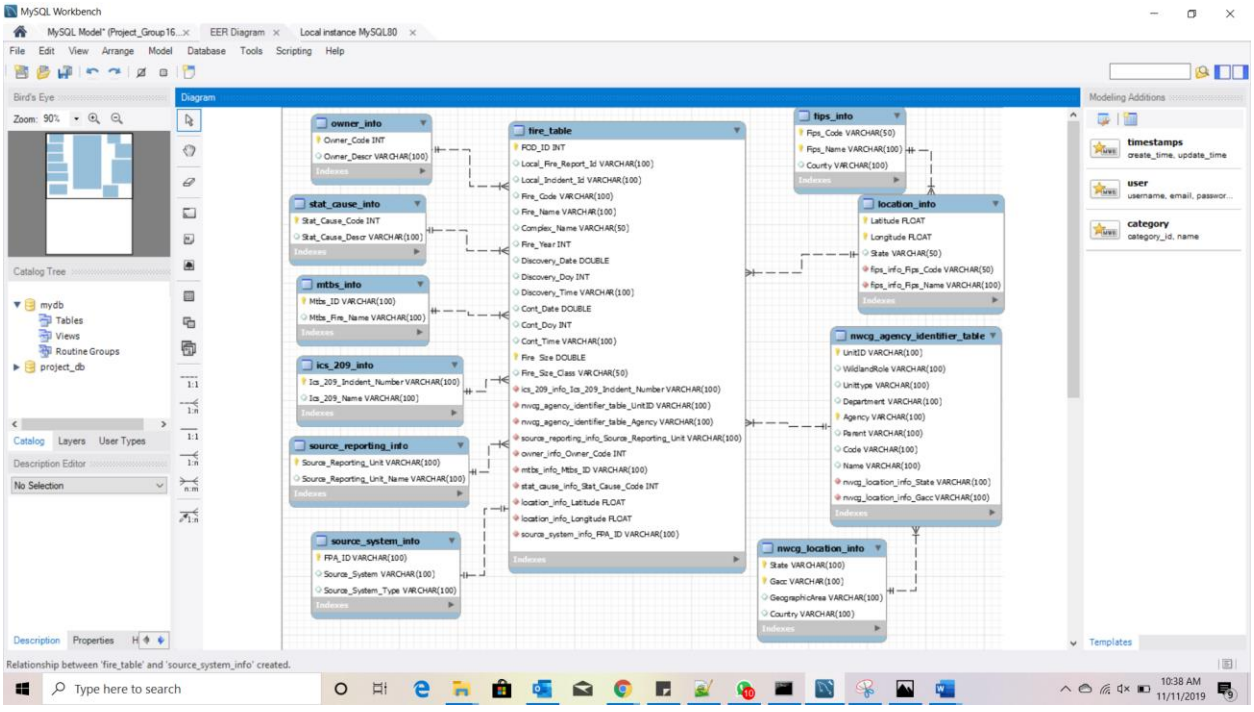
16. On one location many times a fire may occur, one fire would occur only on one location
17. Each location may have many NWCG Units, Each NWCG unit would be on only one location.
18. For each Active NWCG unit many fire incidents may occur, for each fire there was one unique NWCG unit active.
19. Each source reporting unit may prepare many fire reports, each fire report is prepared by only one source reporting unit.

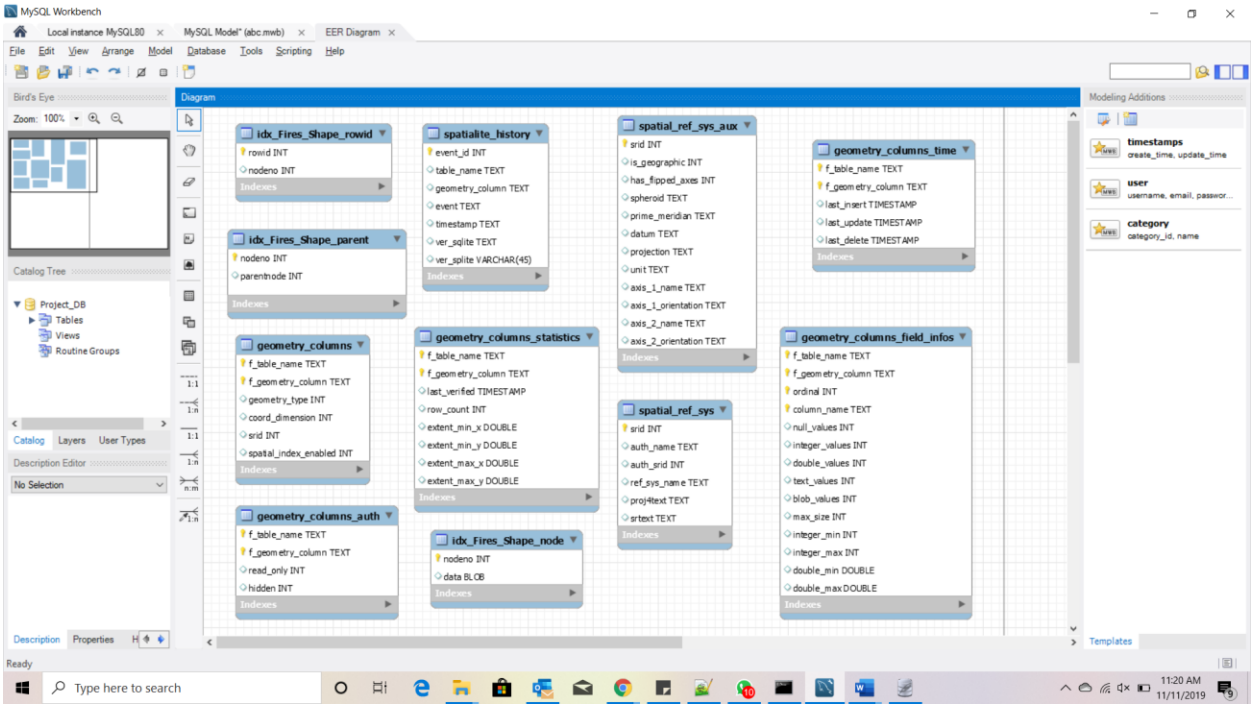
Include reasons why the data model is in 3NF.

- Each table has its own functionality and serves its own purpose.
- Every table has a primary key and each table contains only atomic values instead of group values which satisfies the condition of 1NF.
- Every column in each table is dependent on a single primary key and not on a composite key which satisfies the condition of 2NF.
- There are no transitive functional dependencies between any column in table which satisfies the condition of 3NF

Entity-Relationship Diagram

ER DIAGRAM





Physical Database

Assumptions/Notes About Data Set

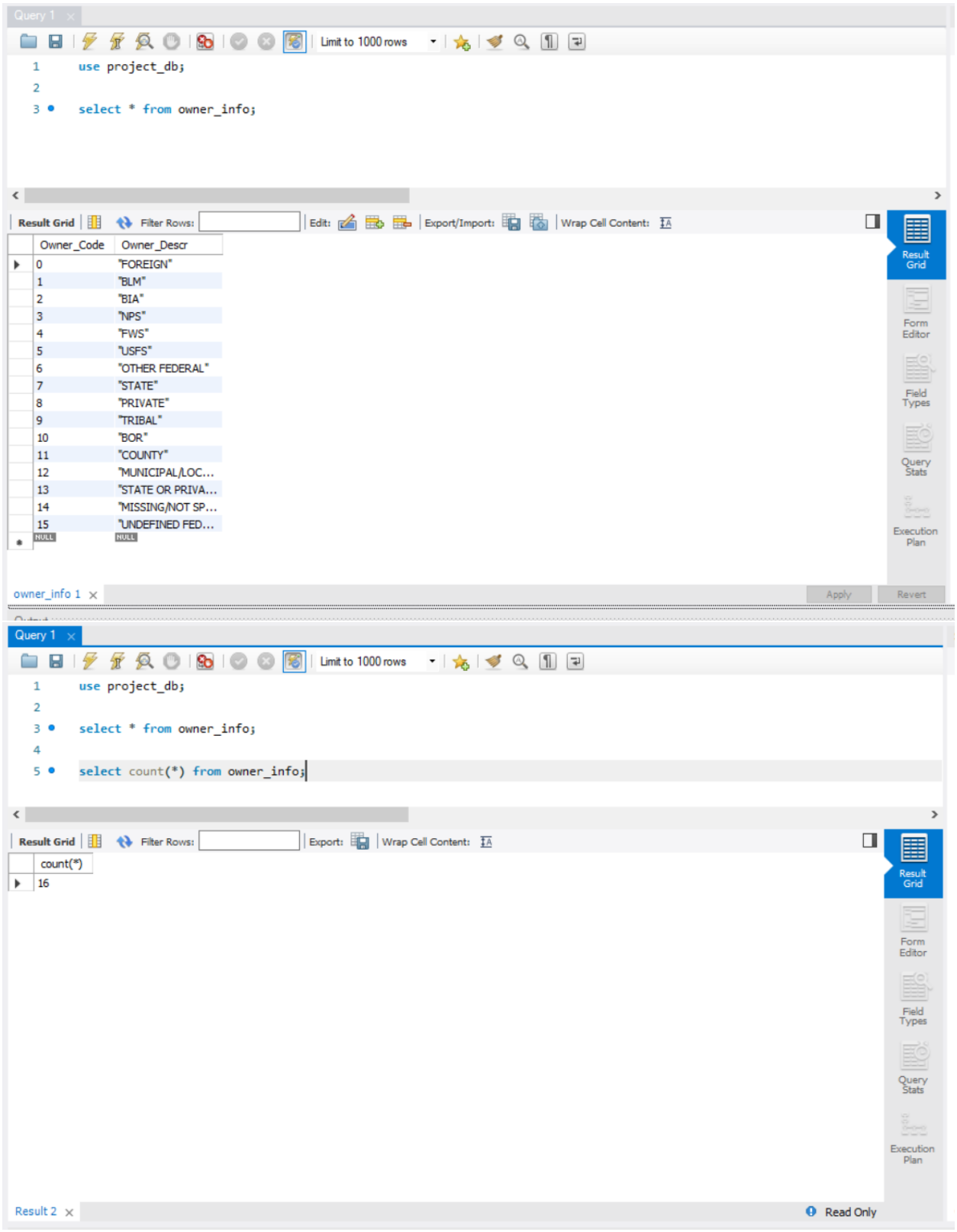
Include any assumptions made about data such as empty fields, sparse data, bad data, etc.

Bad Data:

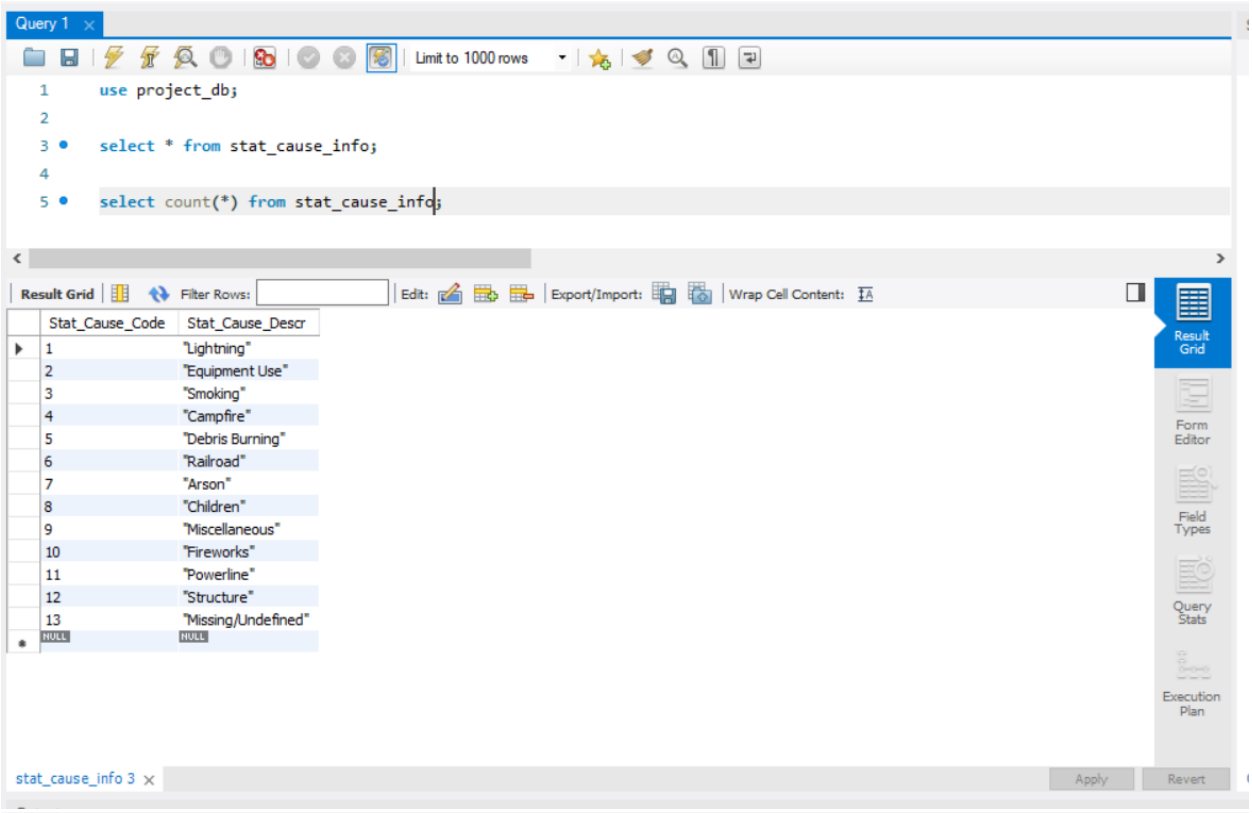
1. The column county in the original fire table does not have a **normalized entry** for example the county name harding is entered as Harding or Harding-county or 39 in different rows because of which we are unable to take the combination of state and county as unique values
2. The combination of columns fips_code,state should be unique value but due to no normalized entry of fips_name we are unable to use the combination as a primary key.

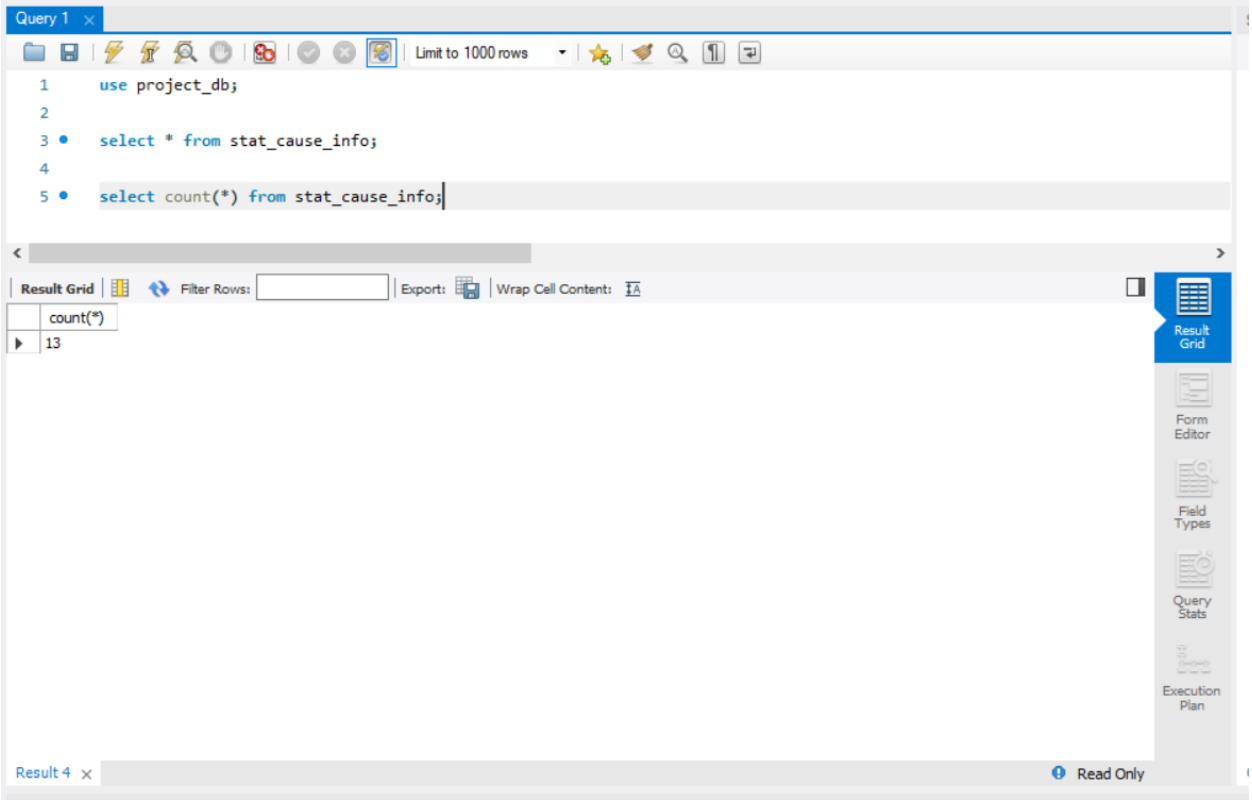
Screen shot of Physical Database objects

- owner_info

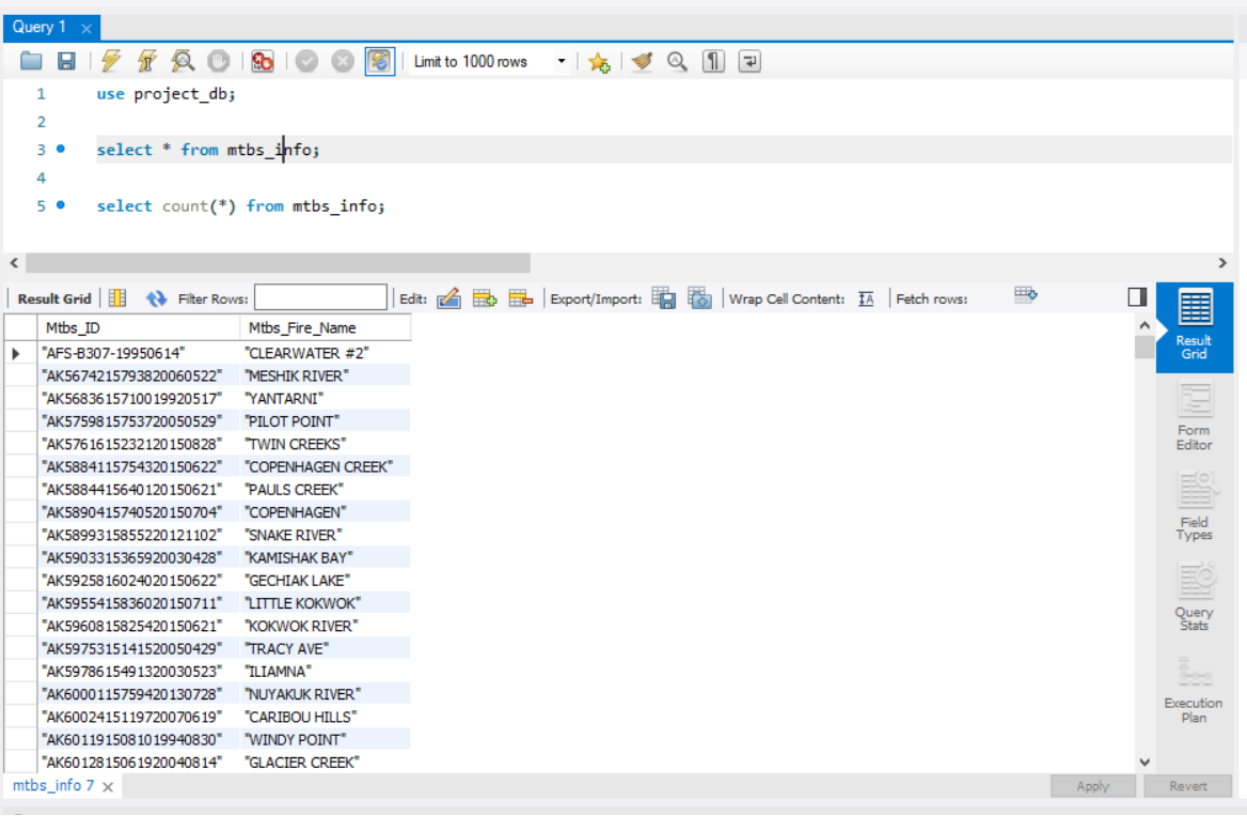


- stat_cause_info



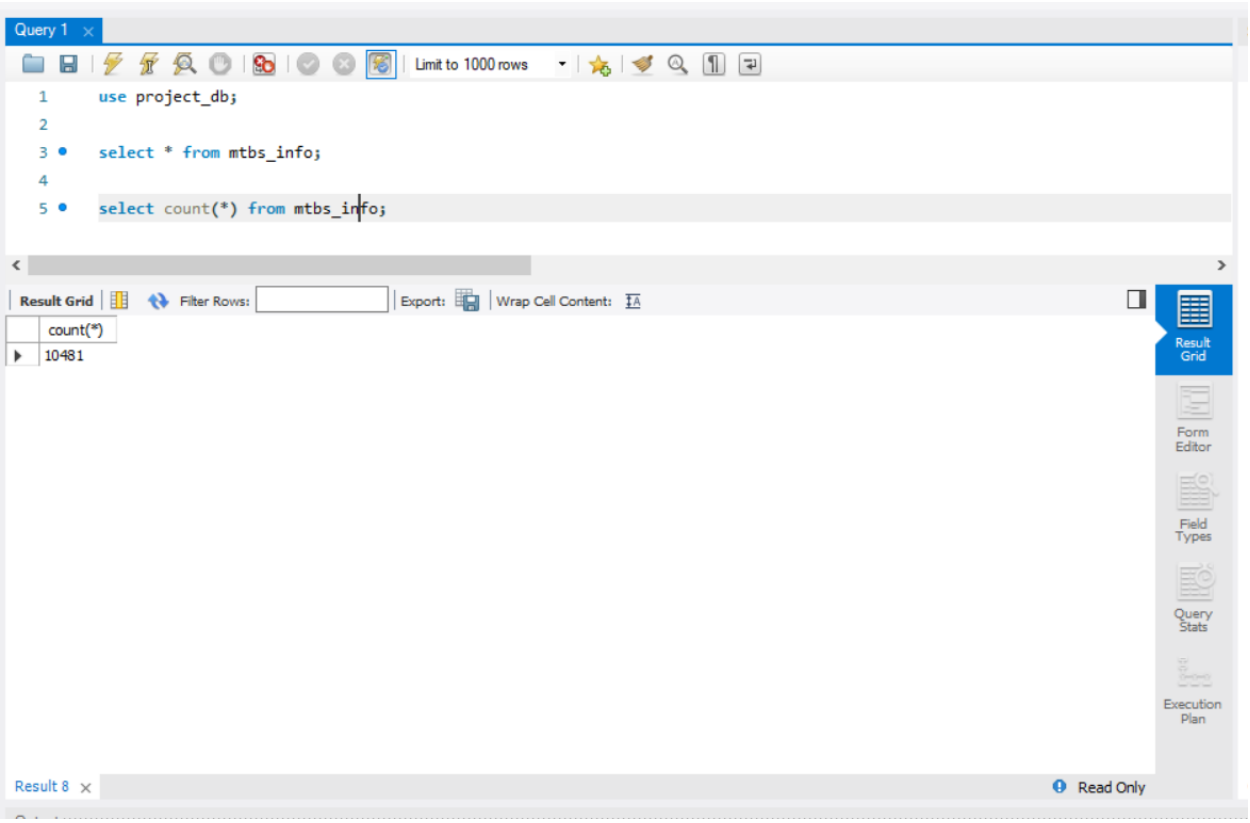


- mtbs_info



```
use project_db;

select * from mtbs_info;

select count(*) from mtbs_info;
```

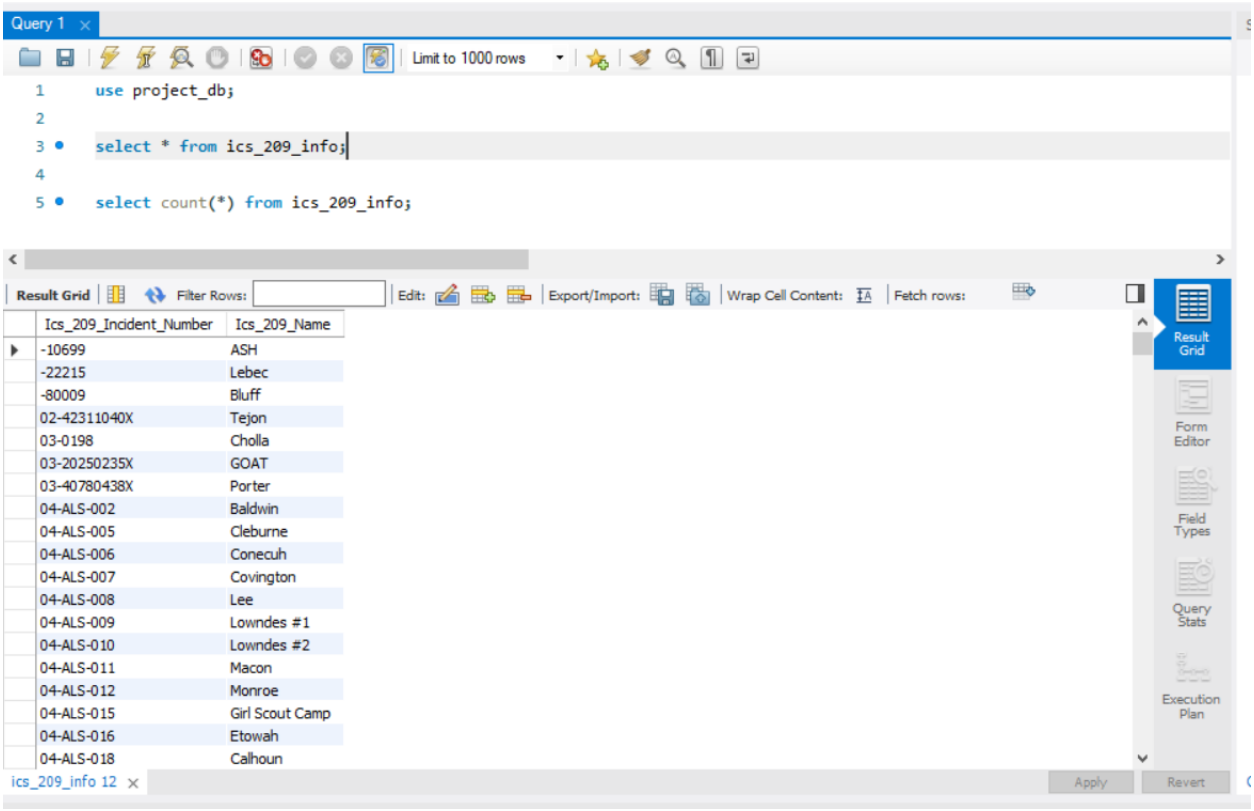
Mtbs_ID	Mtbs_Fire_Name
"AFS-B307-19950614"	"CLEARWATER #2"
"AK567421579382006052 2"	"MESHIK RIVER"
"AK568361571001992051 7"	"YANTARNI"
"AK575981575372005052 9"	"PILOT POINT"
"AK576161523212015082 8"	"TWIN CREEKS"
"AK588411575432015062 2"	"COPENHAGEN CREEK"
"AK588441564012015062 1"	"PAULS CREEK"
"AK589041574052015070 4"	"COPENHAGEN"
"AK589931585522012110 2"	"SNAKE RIVER"
"AK590331536592003042 8"	"KAMISHAK BAY"
"AK592581602402015062 2"	"GECHIAK LAKE"
"AK595541583602015071 1"	"LITTLE KOKWOK"
"AK596081582542015062 1"	"KOKWOK RIVER"
"AK597531514152005042 9"	"TRACY AVE"
"AK597861549132003052 3"	"ILIAMNA"
"AK600011575942013072 8"	"NUYAKUK RIVER"
"AK600241511972007061 9"	"CARIBOU HILLS"
"AK601191508101994083 0"	"WINDY POINT"
"AK601281506192004081 4"	"GLACIER CREEK"

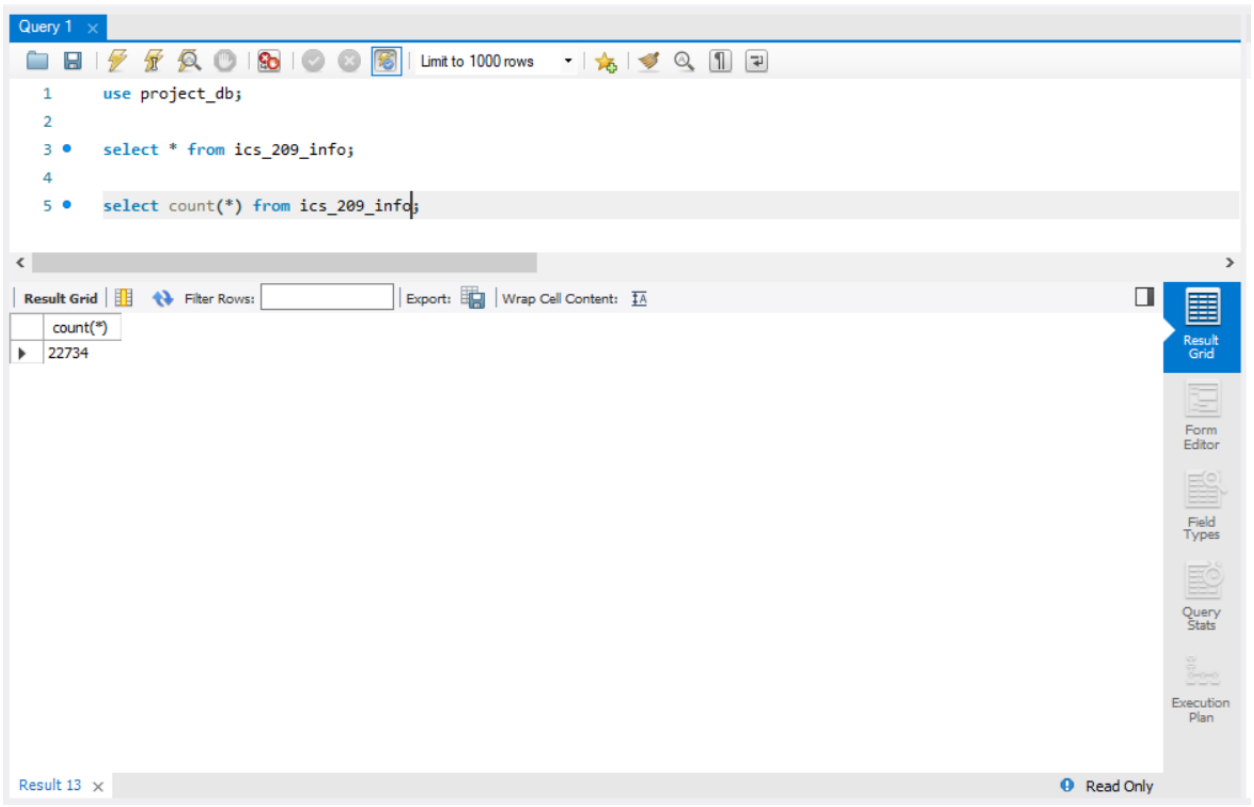


```
use project_db;

select * from mtbs_info;

select count(*) from mtbs_info;
```

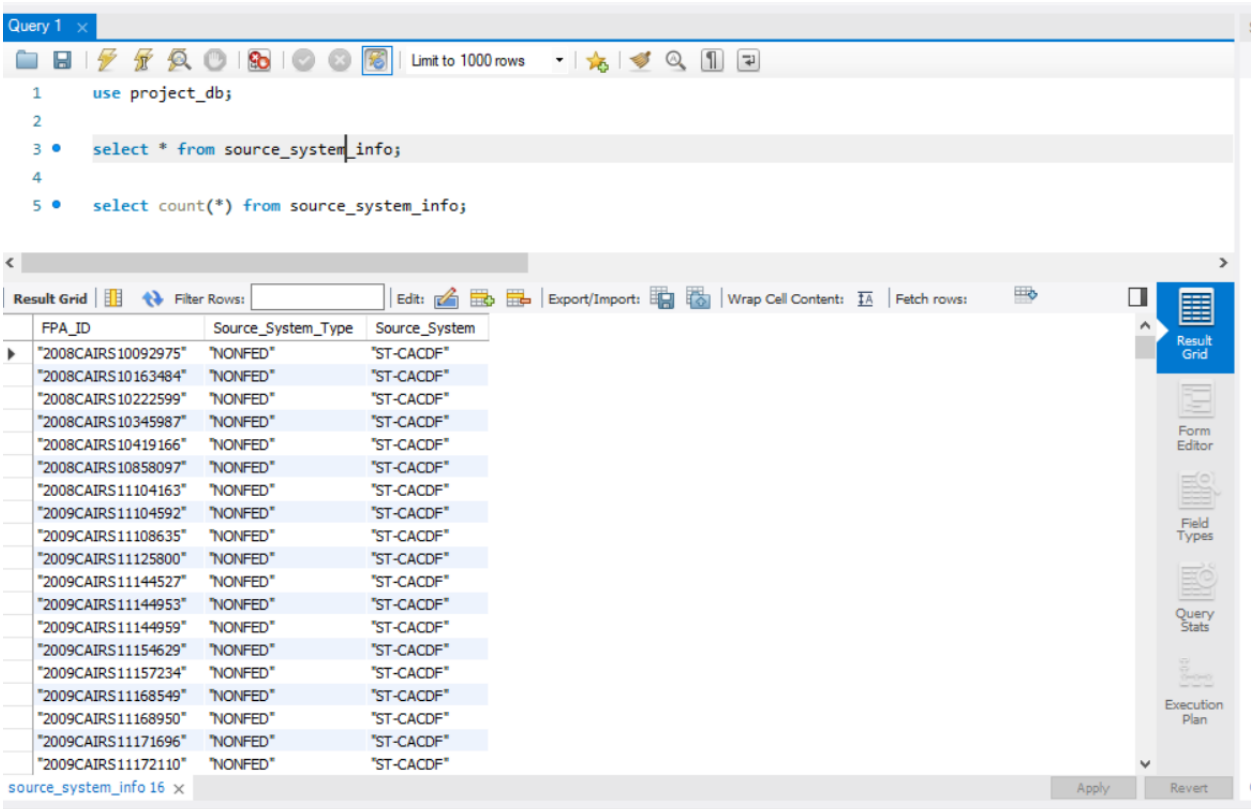
count(*)
10481

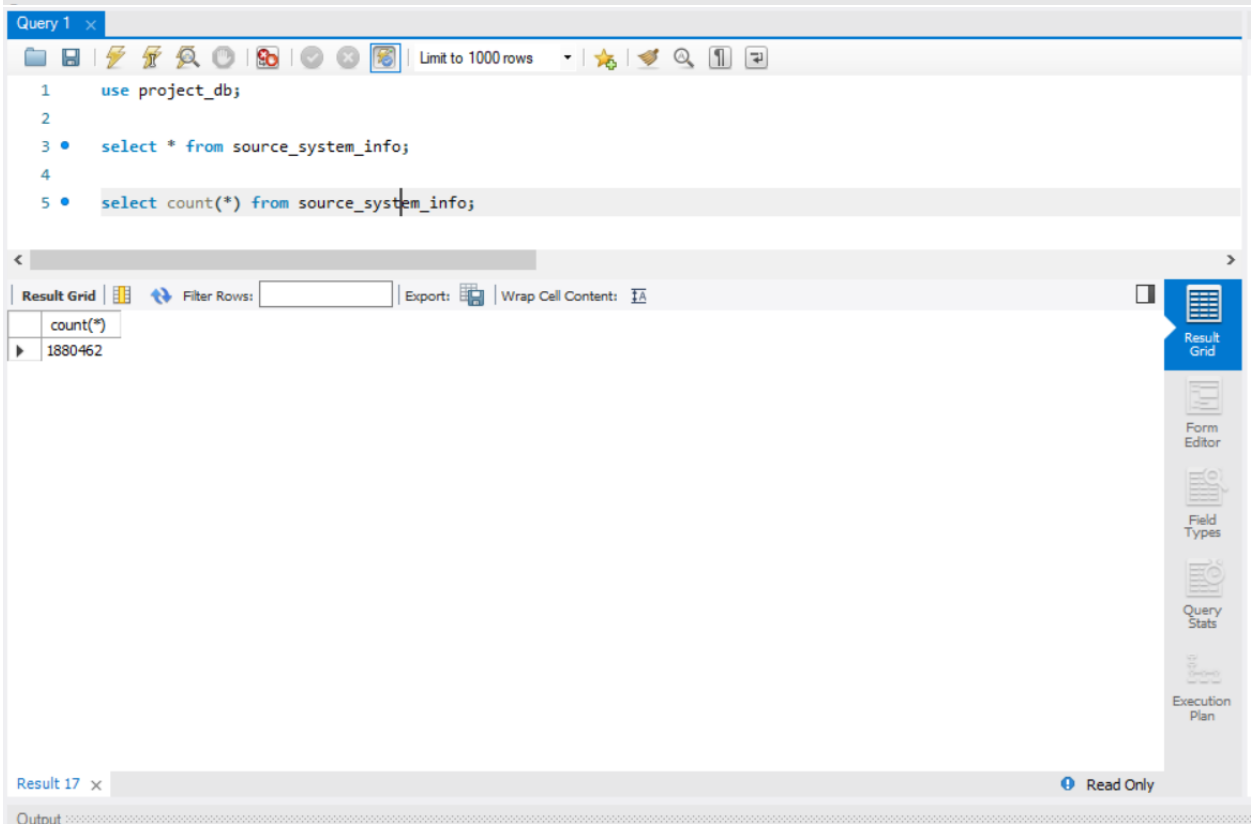
- ics_209_info


Query 1
Limit to 1000 rows
use project_db;
select * from ics_209_info;
select count(*) from ics_209_info;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
Fetch rows:
Ics_209_Incident_Number
Ics_209_Name
-10699 ASH
-22215 Lebec
-80009 Bluff
02-42311040X Tejon
03-0198 Cholla
03-20250235X GOAT
03-40780438X Porter
04-ALS-002 Baldwin
04-ALS-005 Cleburne
04-ALS-006 Conecuh
04-ALS-007 Covington
04-ALS-008 Lee
04-ALS-009 Lowndes #1
04-ALS-010 Lowndes #2
04-ALS-011 Macon
04-ALS-012 Monroe
04-ALS-015 Girl Scout Camp
04-ALS-016 Etowah
04-ALS-018 Calhoun
Result Grid
Form Editor
Field Types
Query Stats
Execution Plan
ics_209_info 12
Apply
Revert

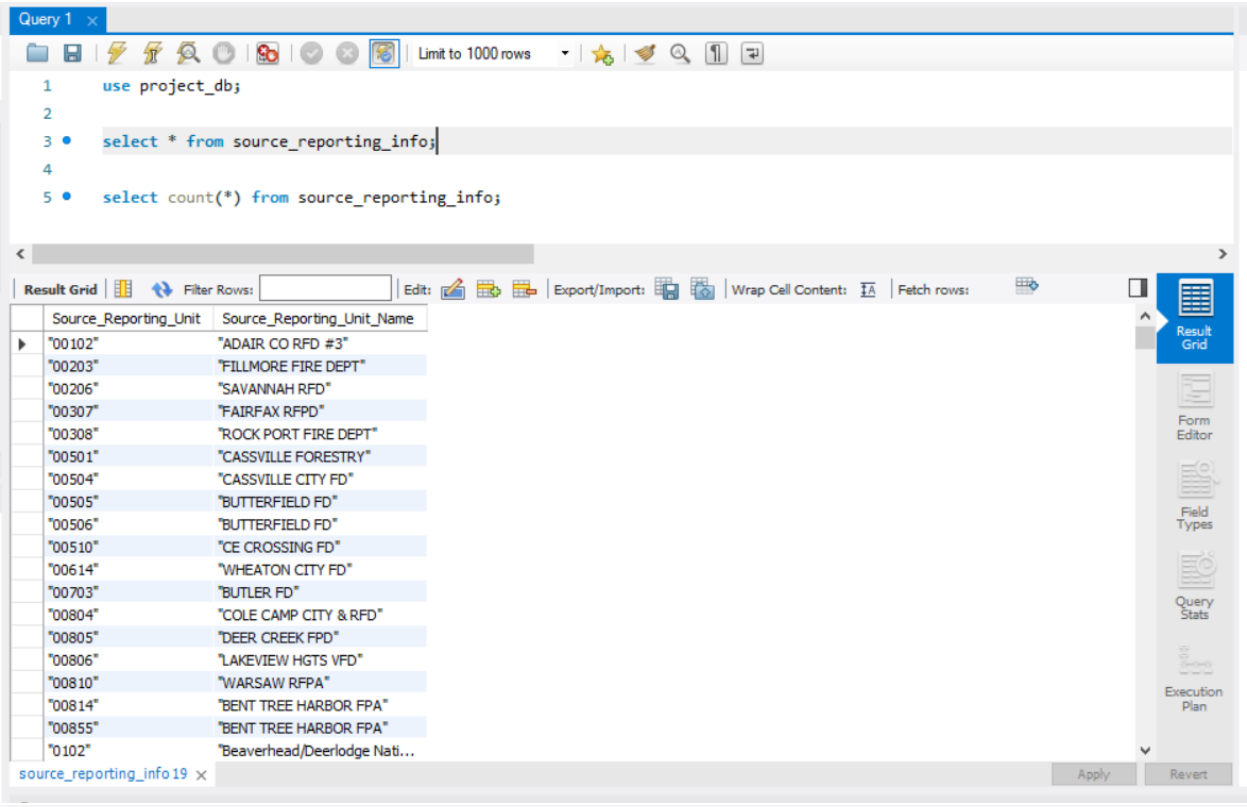

Query 1
Limit to 1000 rows
use project_db;
select * from ics_209_info;
select count(*) from ics_209_info;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
count(*)
22734
Result Grid
Form Editor
Field Types
Query Stats
Execution Plan
Result 13
Read Only

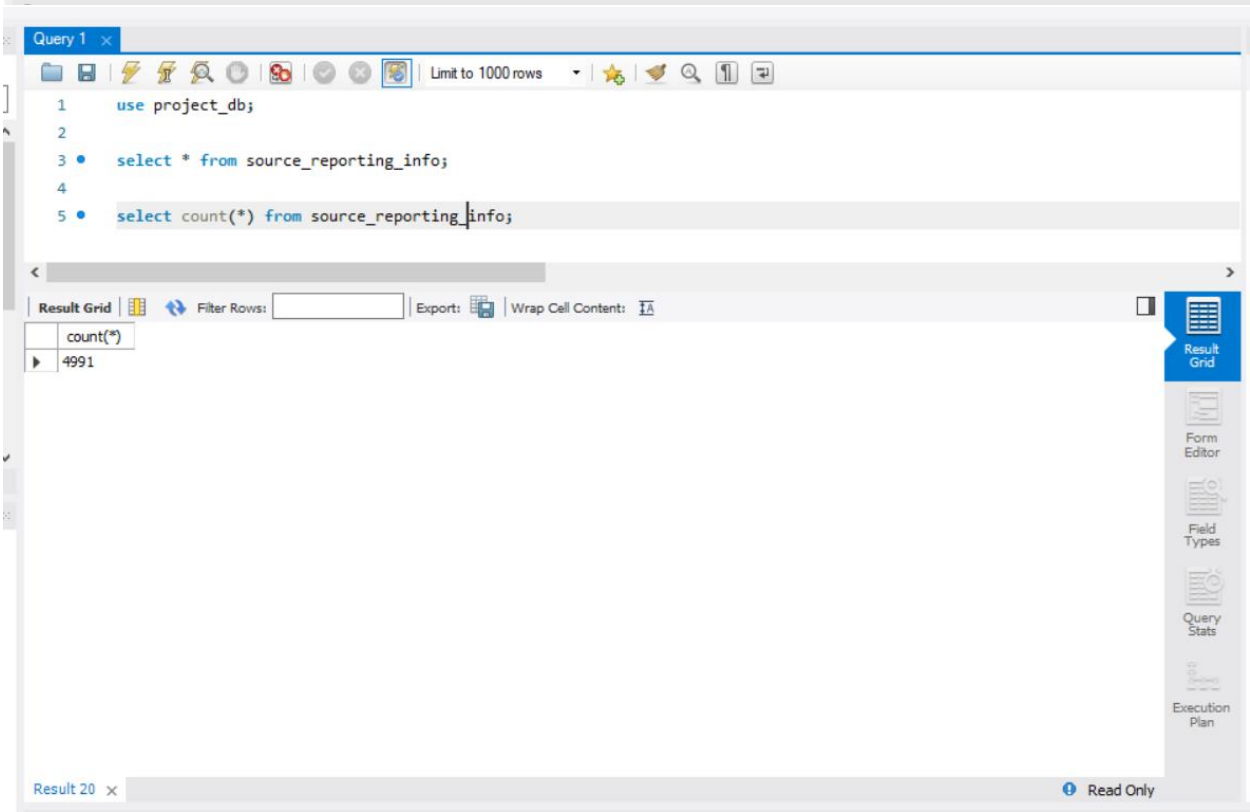
- source_system_info



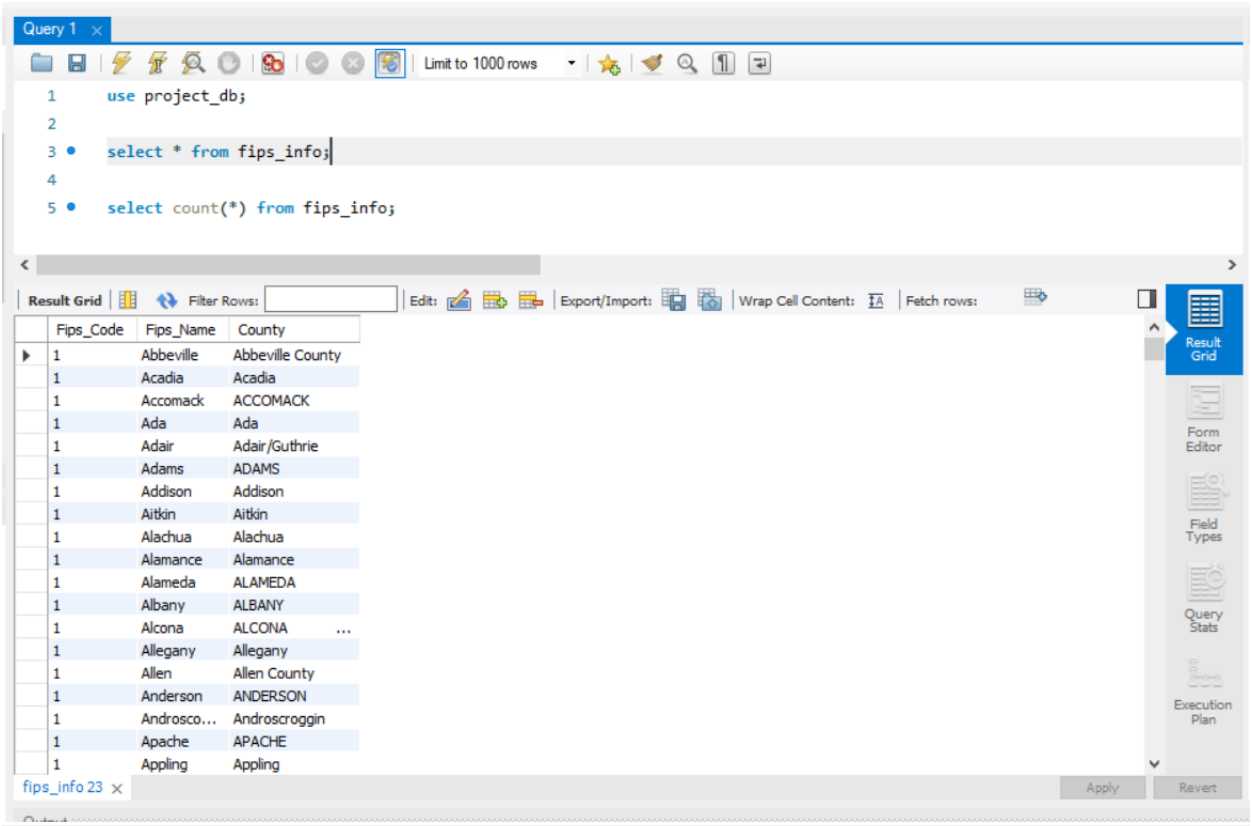


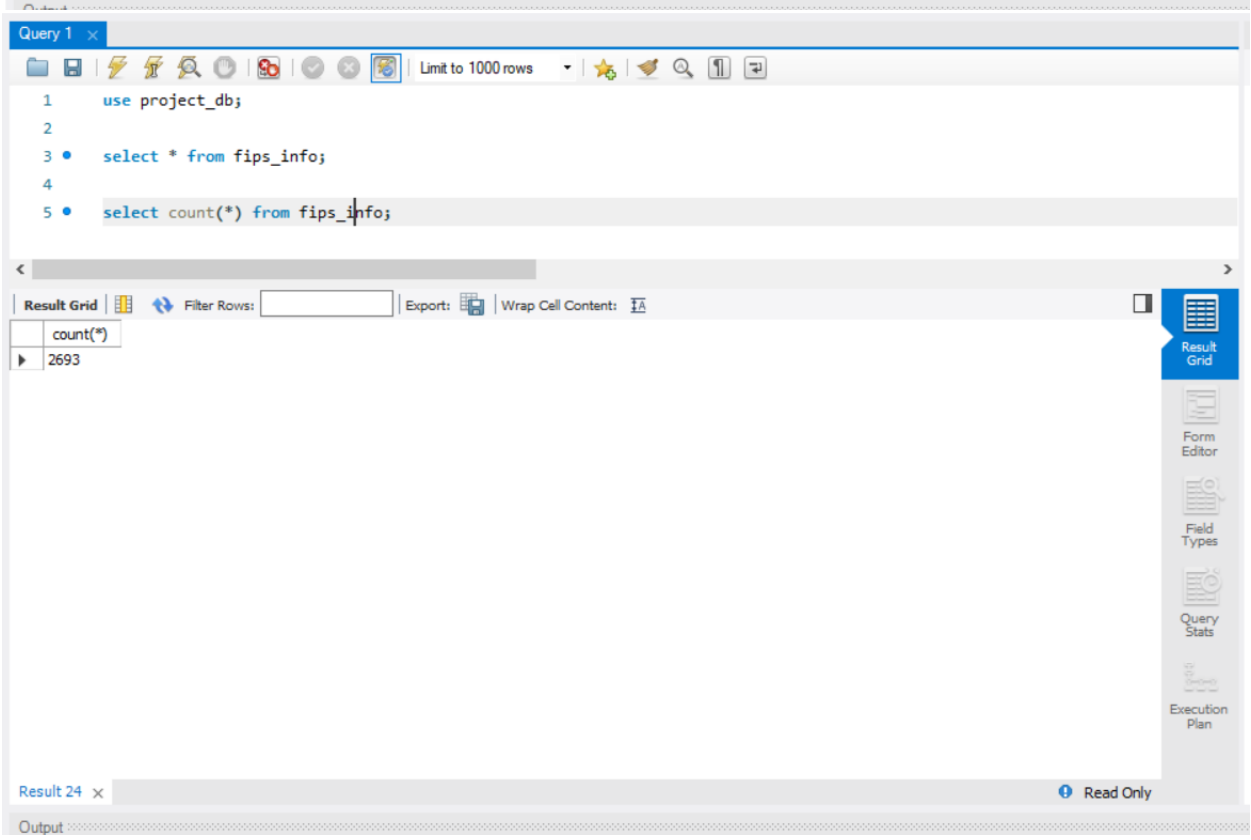
- source_reporting_info



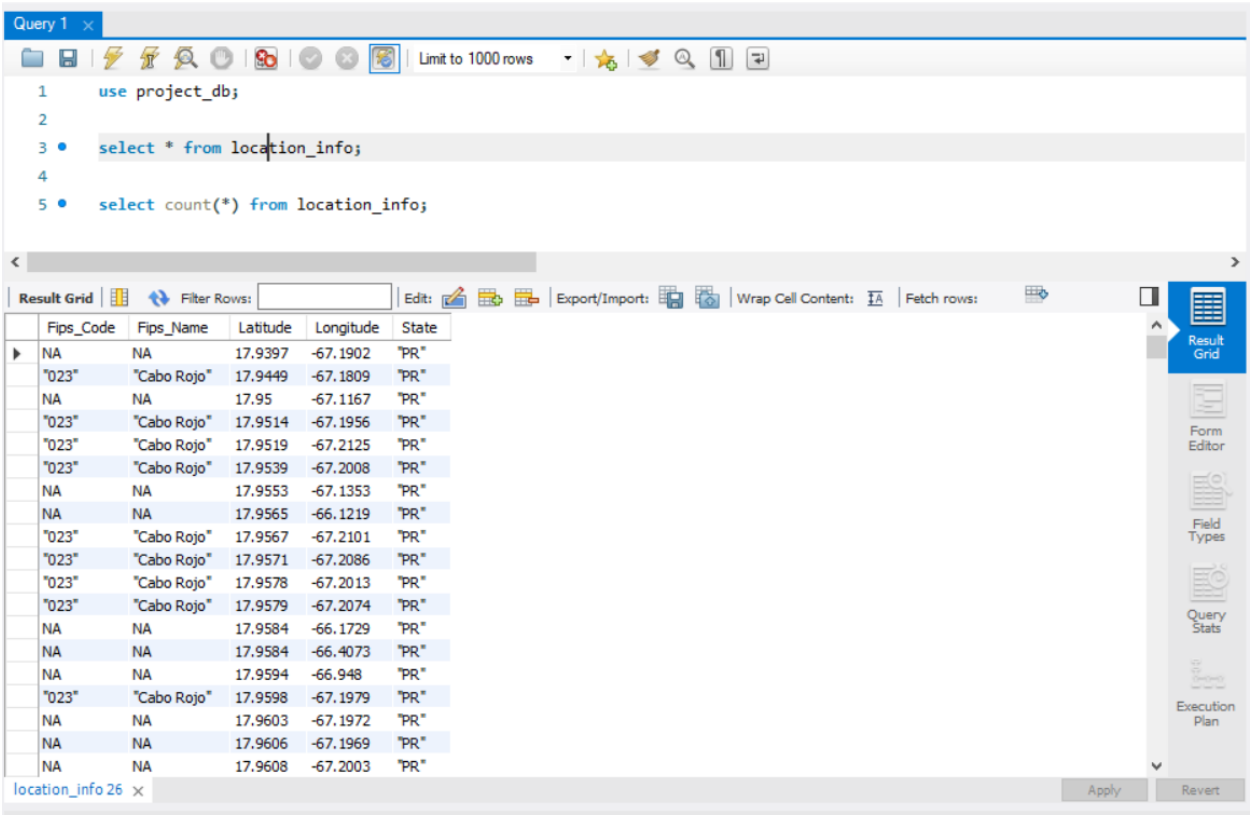


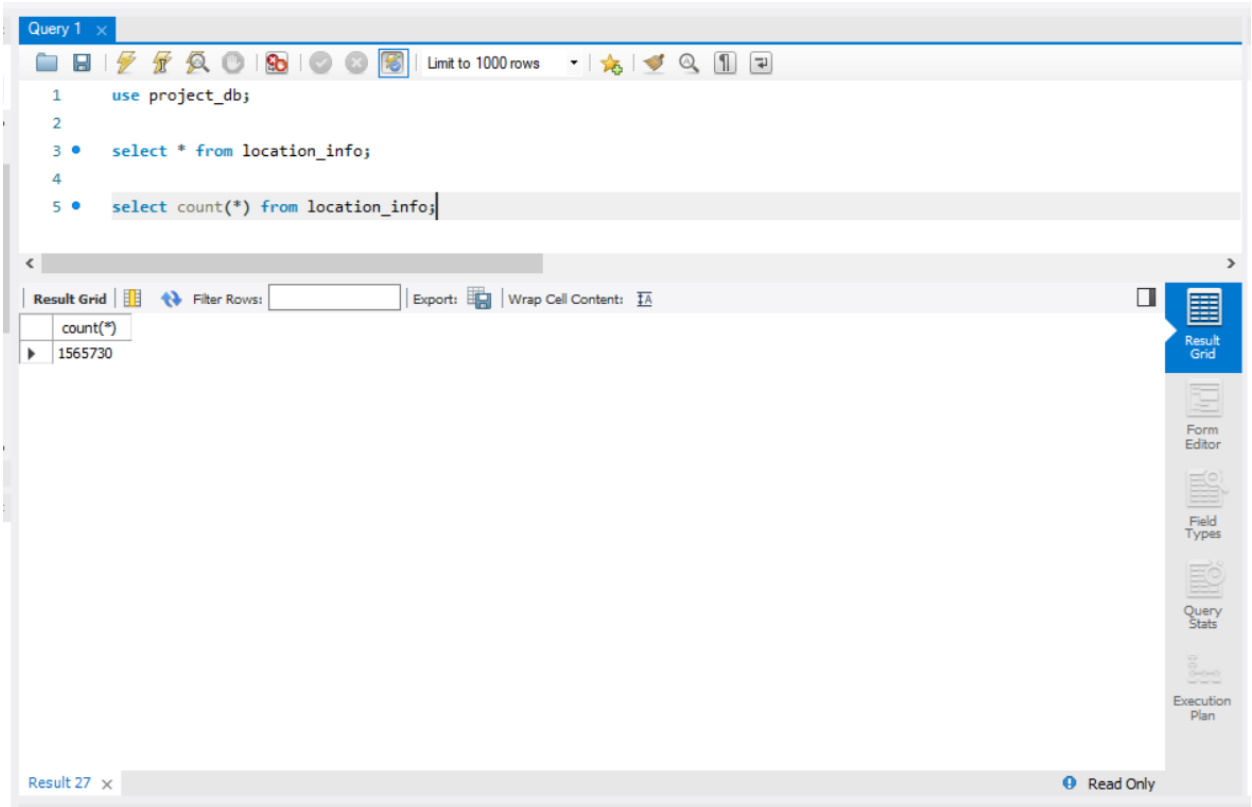
- fips_info


Query 1
Limit to 1000 rows
use project_db;
select * from fips_info;
select count(*) from fips_info;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
Fetch rows:
Fips_Code Fips_Name County
1 Abbeville Abbeville County
1 Acadia Acadia
1 Accomack ACCOMACK
1 Ada Ada
1 Adair Adair/Guthrie
1 Adams ADAMS
1 Addison Addison
1 Aitkin Aitkin
1 Alachua Alachua
1 Alamance Alamance
1 Alameda ALAMEDA
1 Albany ALBANY
1 Alcona ALCONA
1 Allegany Allegany
1 Allen Allen County
1 Anderson ANDERSON
1 Androsco... Androscroggin
1 Apache APACHE
1 Appling Appling
Result Grid
Form Editor
Field Types
Query Stats
Execution Plan
fips_info 23
Apply
Revert

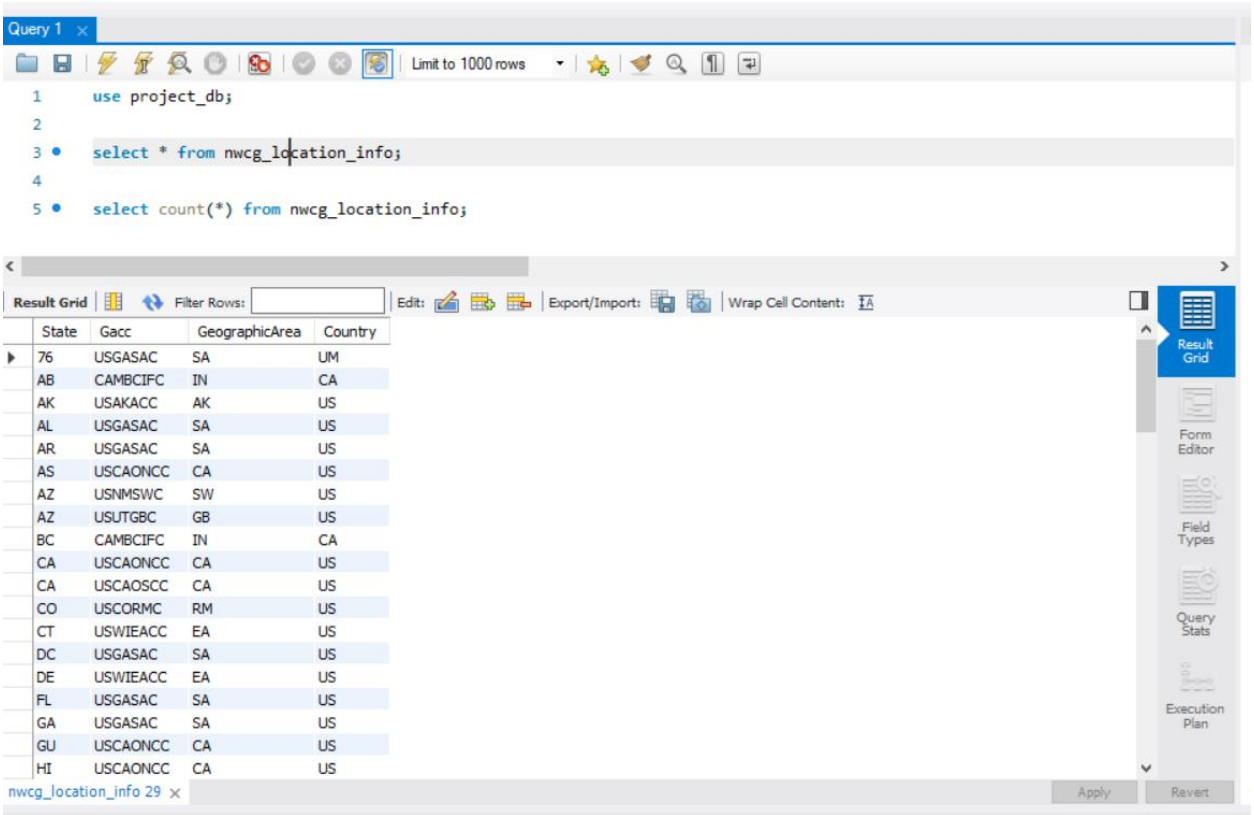

Query 1
Limit to 1000 rows
use project_db;
select * from fips_info;
select count(*) from fips_info;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
count(*)
2693
Result Grid
Form Editor
Field Types
Query Stats
Execution Plan
Result 24
Read Only
Output

- location_info



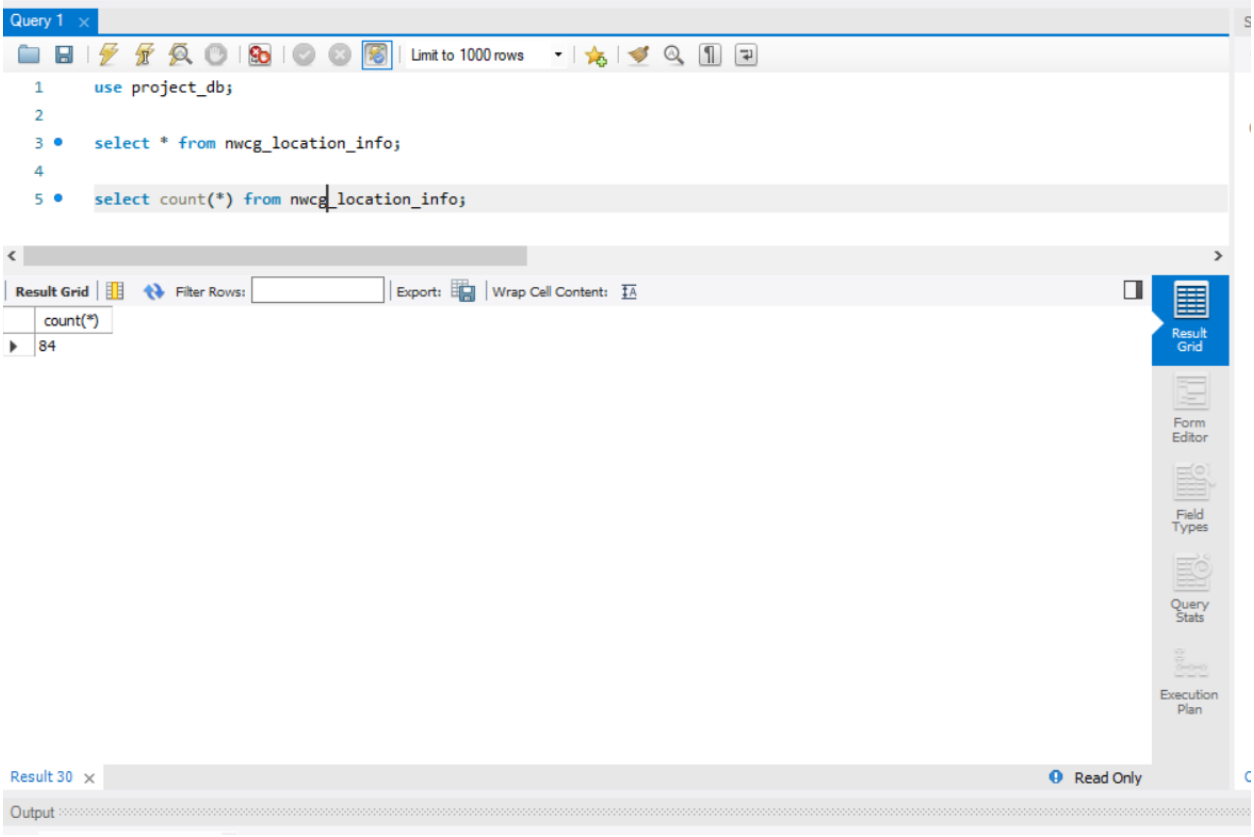


- nwcg_location_info



```
use project_db;

select * from nwcg_location_info;

select count(*) from nwcg_location_info;
```

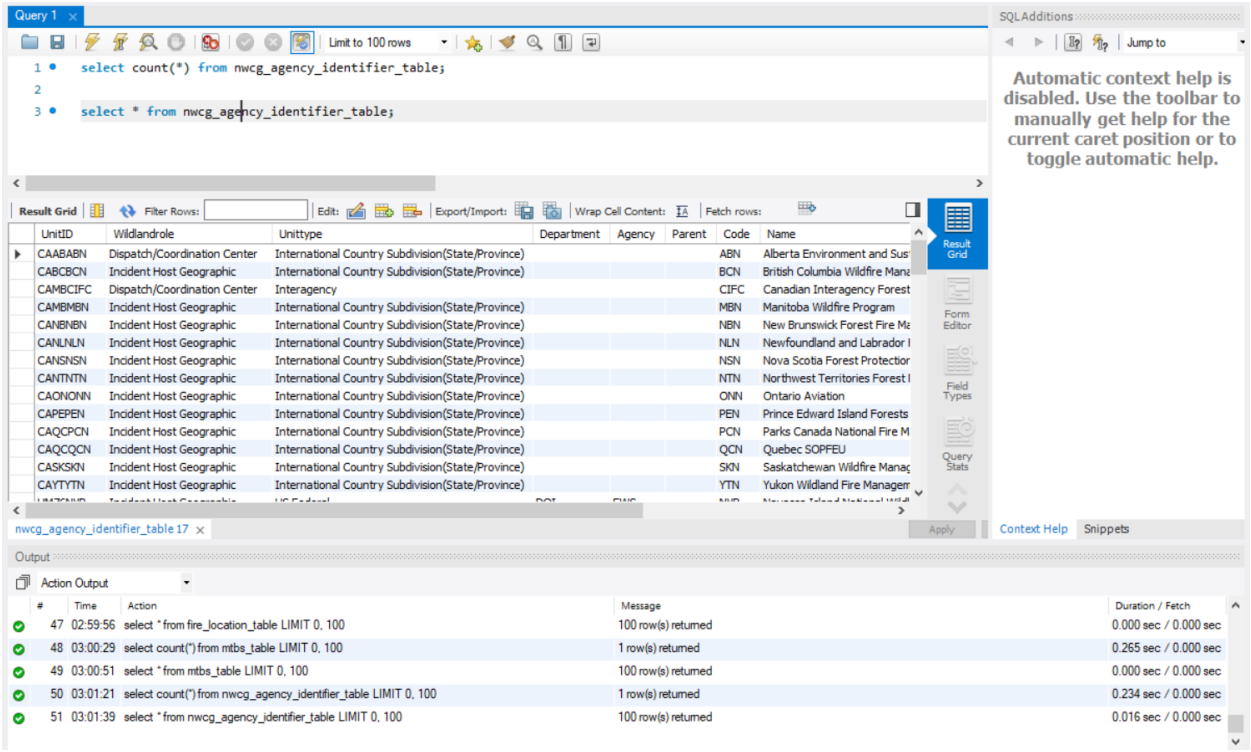
State	Gacc	GeographicArea	Country
76	USGASAC	SA	UM
AB	CAMBCIFC	IN	CA
AK	USAKACC	AK	US
AL	USGASAC	SA	US
AR	USGASAC	SA	US
AS	USCAONCC	CA	US
AZ	USNMSWC	SW	US
AZ	USUTGBC	GB	US
BC	CAMBCIFC	IN	CA
CA	USCAONCC	CA	US
CA	USCAOSCC	CA	US
CO	USCORMC	RM	US
CT	USWIEACC	EA	US
DC	USGASAC	SA	US
DE	USWIEACC	EA	US
FL	USGASAC	SA	US
GA	USGASAC	SA	US
GU	USCAONCC	CA	US
HI	USCAONCC	CA	US

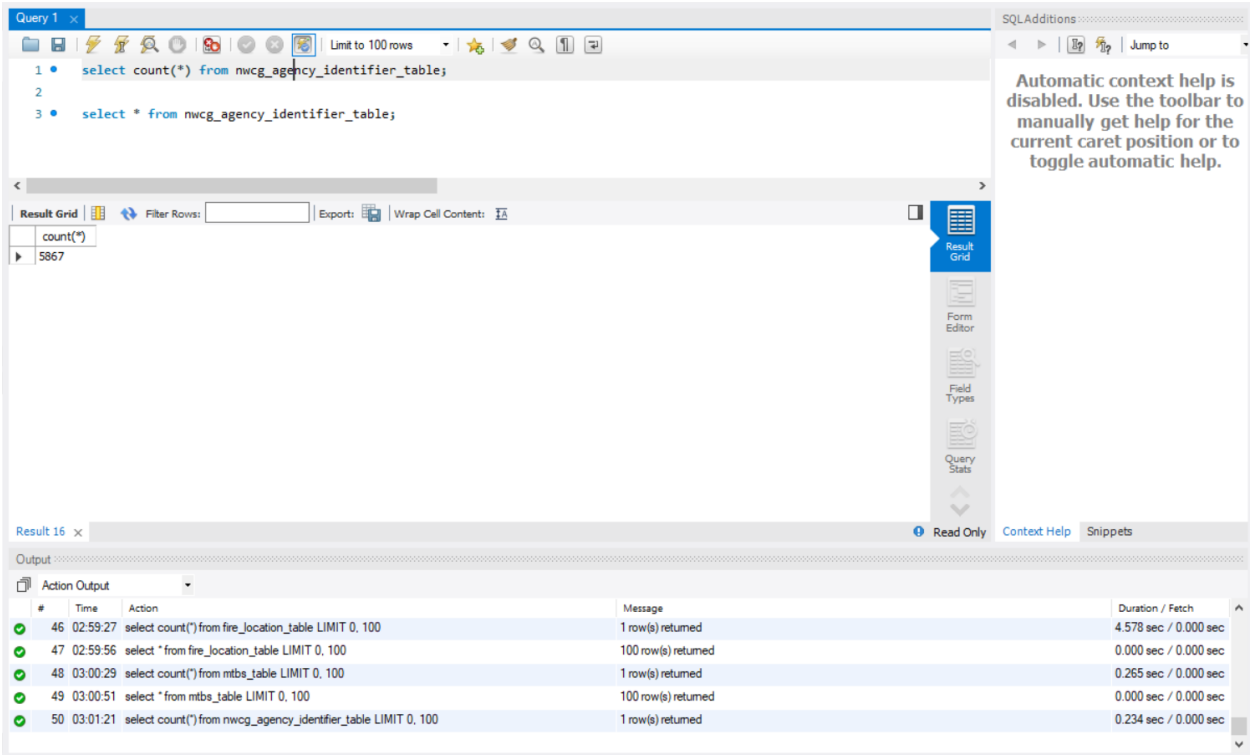


```
use project_db;

select * from nwcg_location_info;

select count(*) from nwcg_location_info;
```

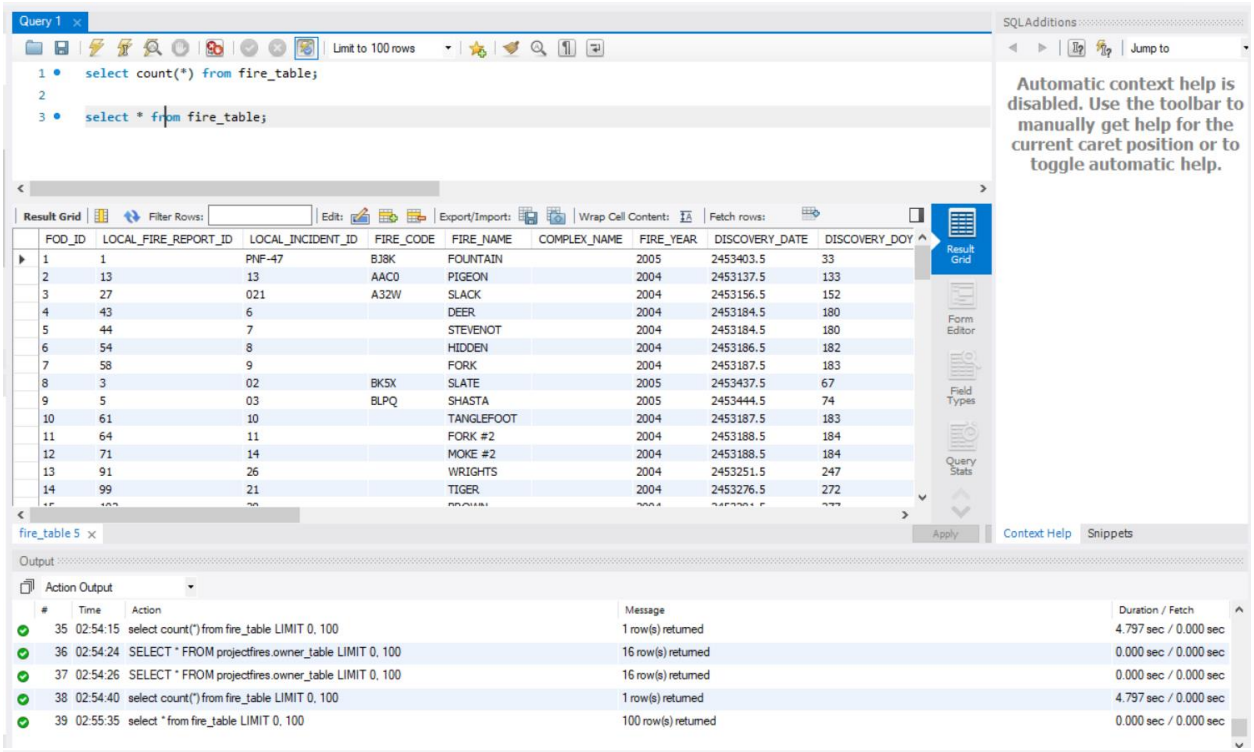
count(*)
84

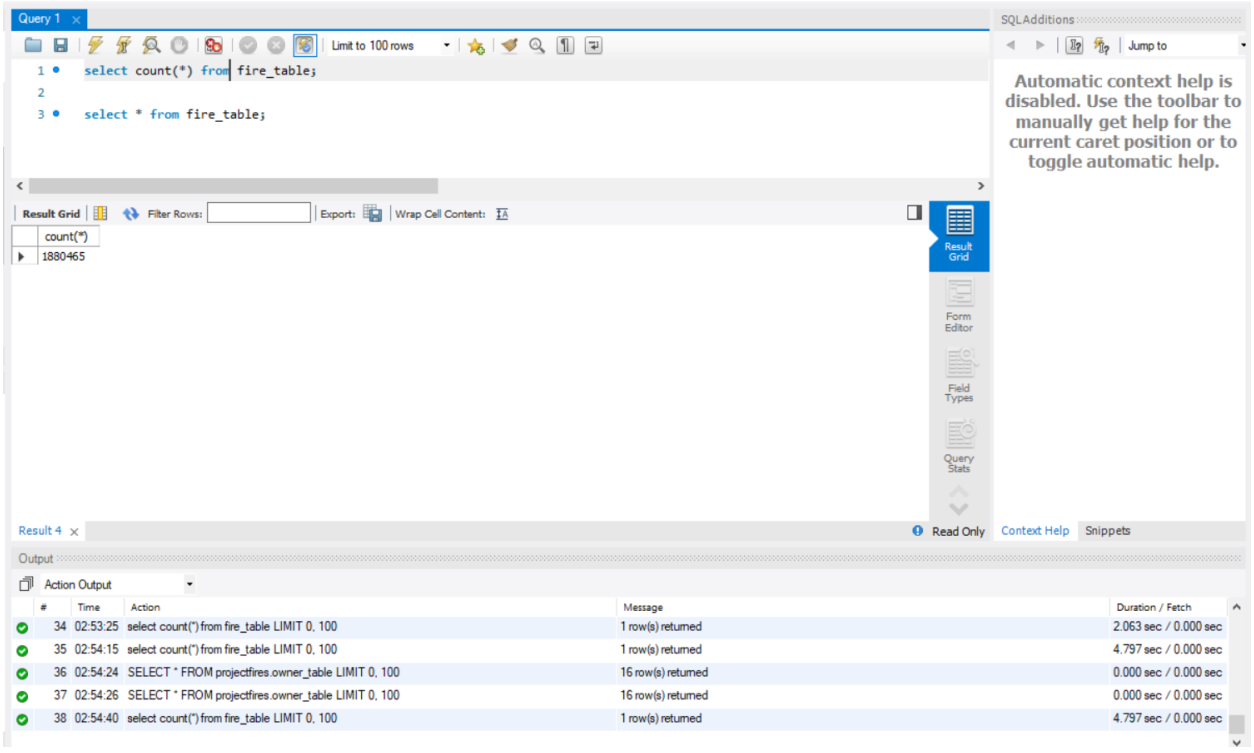
- nwcg_agency_identifier_table





- fire_table


Query 1
Limit to 100 rows
1 • select count(*) from fire_table;
2
3 • select * from fire_table;
Result Grid
Filter Rows:
Edit:
Export/Import:
Wrap Cell Content:
Fetch rows:
FOD_ID | LOCAL_FIRE_REPORT_ID | LOCAL_INCIDENT_ID | FIRE_CODE | FIRE_NAME | COMPLEX_NAME | FIRE_YEAR | DISCOVERY_DATE | DISCOVERY_DOY
1 | 1 | PNF-47 | BJ8K | FOUNTAIN | | 2005 | 2453403.5 | 33
2 | 13 | 13 | AAC0 | PIGEON | | 2004 | 2453137.5 | 133
3 | 27 | 021 | A32W | SLACK | | 2004 | 2453156.5 | 152
4 | 43 | 6 | | DEER | | 2004 | 2453184.5 | 180
5 | 44 | 7 | | STEVENOT | | 2004 | 2453184.5 | 180
6 | 54 | 8 | | HIDDEN | | 2004 | 2453186.5 | 182
7 | 58 | 9 | | FORK | | 2004 | 2453187.5 | 183
8 | 3 | 02 | BK5X | SLATE | | 2005 | 2453437.5 | 67
9 | 5 | 03 | BLPQ | SHASTA | | 2005 | 2453444.5 | 74
10 | 61 | 10 | | TANGLEFOOT | | 2004 | 2453187.5 | 183
11 | 64 | 11 | | FORK #2 | | 2004 | 2453188.5 | 184
12 | 71 | 14 | | MOKE #2 | | 2004 | 2453188.5 | 184
13 | 91 | 26 | | WRIGHTS | | 2004 | 2453251.5 | 247
14 | 99 | 21 | | TIGER | | 2004 | 2453276.5 | 272
fire_table 5
Apply
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help
Snippets
Output
Action Output
| Time | Action | Message | Duration / Fetch
35 | 02:54:15 | select count(*) from fire_table LIMIT 0, 100 | 1 row(s) returned | 4.797 sec / 0.000 sec
36 | 02:54:24 | SELECT * FROM projectfires.owner_table LIMIT 0, 100 | 16 row(s) returned | 0.000 sec / 0.000 sec
37 | 02:54:26 | SELECT * FROM projectfires.owner_table LIMIT 0, 100 | 16 row(s) returned | 0.000 sec / 0.000 sec
38 | 02:54:40 | select count(*) from fire_table LIMIT 0, 100 | 1 row(s) returned | 4.797 sec / 0.000 sec
39 | 02:55:35 | select * from fire_table LIMIT 0, 100 | 100 row(s) returned | 0.000 sec / 0.000 sec


Query 1
Limit to 100 rows
1 • select count(*) from fire_table;
2
3 • select * from fire_table;
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
count(*)
1880465
Result 4
Read Only
SQLAdditions
Jump to
Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.
Context Help
Snippets
Output
Action Output
| Time | Action | Message | Duration / Fetch
34 | 02:53:25 | select count(*) from fire_table LIMIT 0, 100 | 1 row(s) returned | 2.063 sec / 0.000 sec
35 | 02:54:15 | select count(*) from fire_table LIMIT 0, 100 | 1 row(s) returned | 4.797 sec / 0.000 sec
36 | 02:54:24 | SELECT * FROM projectfires.owner_table LIMIT 0, 100 | 16 row(s) returned | 0.000 sec / 0.000 sec
37 | 02:54:26 | SELECT * FROM projectfires.owner_table LIMIT 0, 100 | 16 row(s) returned | 0.000 sec / 0.000 sec
38 | 02:54:40 | select count(*) from fire_table LIMIT 0, 100 | 1 row(s) returned | 4.797 sec / 0.000 sec

Data in the Database

Table Name	Primary Key	Foreign Key	# of Rows in Table
stat_cause_info	Stat_Cause_Code		13
owner_info	Owner_Code		16
ics_209_info	Ics_209_Incident_Index		22734
mtbs_info	Mtbs_ID		10481
source_reporting_info	Source_Reporting_Unit		4991
source_system_info	FPA_ID		1880462
location_info	Latitude, Longitude	fips_info_Fips_Code fips_info_Fips_Name	1565730
fips_info	Fips_Code, Fips_Name		2693
nwcg_location_info	State,Gacc		84
nwcg_agency_identifier_table	UnitID, Agency	nwcg_location_info_State nwcg_location_info_Gacc	5867
fires_table	FOD_ID, Fire_Size	ics_209_info_Ics_209_Incident_Number source_system_info_FPA_ID nwcg_agency_identifier_table_UnitID nwcg_agency_identifier_table_Agency source_reporting_info_Source_Reporting_Unit owner_info_Owner_Code mtbs_info_Mtbs_ID stat_cause_info_Stat_Cause_Code location_info_Latitude location_info_Longitude	1880465

SQL Queries

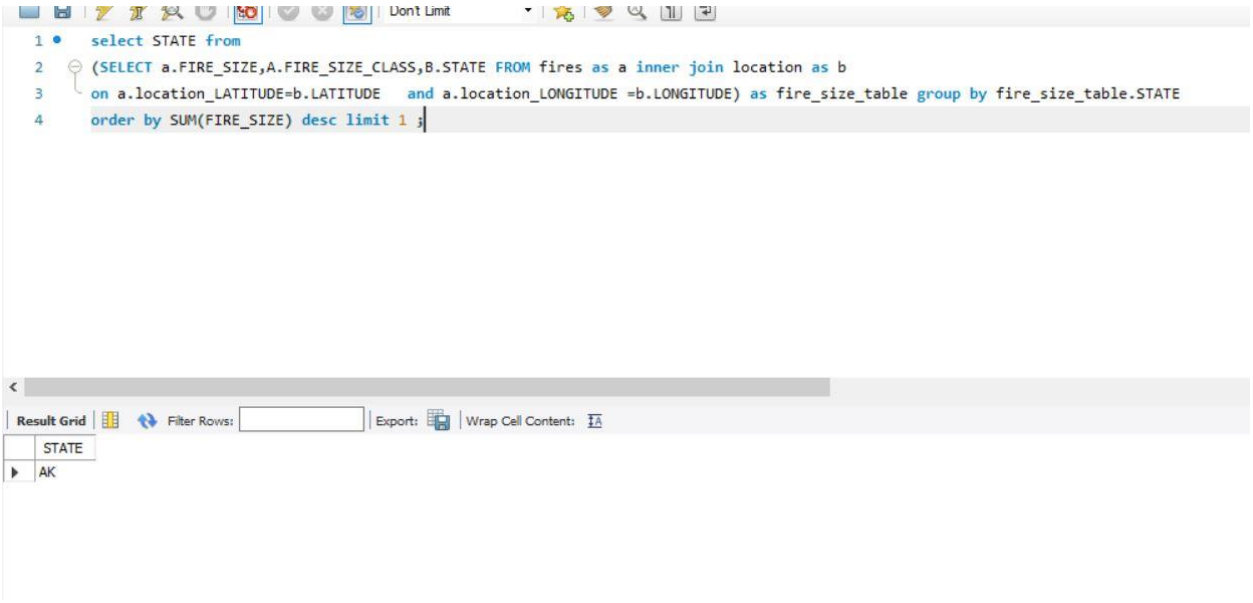
Query 1

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 best chance to win a share of the fund?

As all the states are interested in the billion-dollar fund we group by state, but the state with which has highest number of fires will receive the fund. AK has the highest fires.

Translation : Select states from location by joining it with fires on latitude and longitude and group by state in descending order.

Screen Shot of SQL Query and Results:



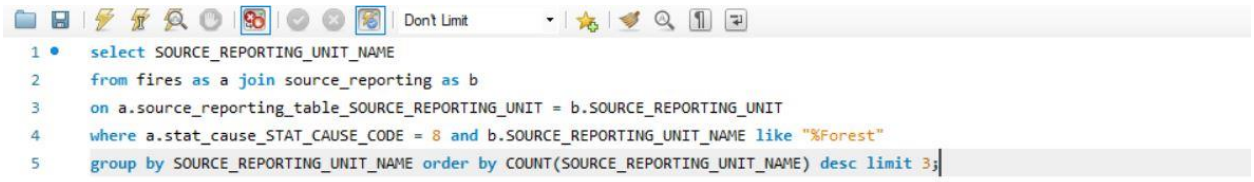
Query 2

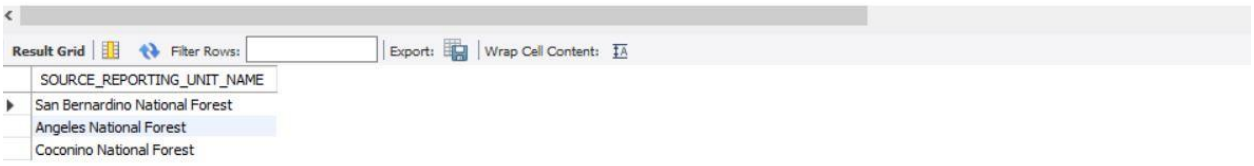
One of the reporting agencies has suggested that children be banned from its forests unless there is one adult for every 3 children in a group visiting a forest. Name 3 forests where this would be the most appropriate.

As the reporting agencies has suggested every 3 children should be accompanied by one adult, we group by source reporting unit in descending order and selecting first three forests will give us 3 rows.

Translation: select source _reporting_unit_name from source_reporting table by joining on source reporting unit where stat_cause_code is 8 and source_reporting_unit name is like forest and grouping them by source _reporting_unit_name in descending order.

Screen Shot of SQL Query and Results:





Query 3

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

As the statement states human actions are to be blamed for most wildfires, we need the count the number of fires that are caused by human and nature. If the count of fires caused by human are more this proves the above statement is true. 2 rows are retrieved.

Translation: select cause and count of fires from fires table where the stat_cause_descr is lightning, nature and human and group by cause.

Screen Shot of SQL Query and Results:

```
1
2 •  SELECT cause,count(*) as count
3    from (select IF(STAT_CAUSE_DESCR= 'Lightning' ,'Nature','Human') as cause
4    from fires
5    where STAT_CAUSE_DESCR not in ('Miscellaneous','Missing/Undefined')) as type_damage group by cause;
6
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

cause	count
Nature	278451
Human	1110963

Query 4

How many wildfires were reported by more than one unit/agency?

Retrieve the number of wildfires reported in more than one unit/agency. One row is retrieved.

Translation : select count of source_reporting_name from fires and source_reporting table by them on source _reporting_unit group by fire_name and having no of units greater than 1.

Screen Shot of SQL Query and Results:

```
1  select count(*) as no_of_fires from
2  (select a.FIRE_NAME, count(SOURCE_REPORTING_UNIT) as No_of_units_reported
3  from fires as a join source_reporting as b
4  on a.source_reporting_table_SOURCE_REPORTING_UNIT=b.SOURCE_REPORTING_UNIT
5  group by a.FIRE_NAME having No_of_units_reported>1) as fires;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

no_of_fires
76634

Query 5

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

Retrieve the forests where the fires lasted for more than 2 days. 128 rows were retrieved.

Translation: select source reporting unit name from source_reporting table by joining it with fires on source reporting unit where the difference between controlled date and discovery date is greater than 2 and source reporting unit name like forest and grouping by source reporting unit having count of source reporting unit name is greater than 1.

Screen Shot of SQL Query and Results:

```
1 •  select SOURCE_REPORTING_UNIT_NAME
2    from fires as a join source_reporting as b
3    on a.source_reporting_table_SOURCE_REPORTING_UNIT = b.SOURCE_REPORTING_UNIT
4    where (a.CONT_DATE-a.DISCOVERY_DATE)>2 and SOURCE_REPORTING_UNIT_NAME like "%Forest"
5    group by SOURCE_REPORTING_UNIT_NAME having count(SOURCE_REPORTING_UNIT_NAME)>=1;
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

SOURCE_REPORTING_UNIT_NAME
Beaverhead-Deerlodge National Forest
Bitterroot National Forest
Idaho Panhandle National Forest
Clearwater National Forest
Custer National Forest
Deerlodge National Forest
Flathead National Forest
Gallatin National Forest
Helena National Forest

Query 6

Which forest had the most number of fires?

Report the forest which has most number of fires. One row is retrieved

Translation: select source_reporting_unit_name from source_reporting_unit table by joining it with fires table on soure_reporting_unit where source_reporting_unit_name like forest and grouping it by source_reporting_unit_name in descending order.

Screen Shot of SQL Query and Results:

```
1  select SOURCE_REPORTING_UNIT_NAME,COUNT(SOURCE_REPORTING_UNIT_NAME) as Min_Fire_count
2  from fires as a join source_reporting as b
3  on a.source_reporting_table_SOURCE_REPORTING_UNIT = b.SOURCE_REPORTING_UNIT
4  where b.SOURCE_REPORTING_UNIT_NAME like "%Forest"
5  group by SOURCE_REPORTING_UNIT_NAME order by COUNT(SOURCE_REPORTING_UNIT_NAME) desc limit 1;
6
```

esult Grid | Filter Rows: | Export: | Wrap Cell Content:

SOURCE_REPORTING_UNIT_NAME	Min_Fire_count
Coconino National Forest	8812

Data Review for MongoDB

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

The wildfires database in MongoDB contains 2 collections:

1)Fires

2) NWCG

Relationship: NWCG_REPORTING_UNIT_ID field in Fires collection is linked with UnitId field in NWCG Collection

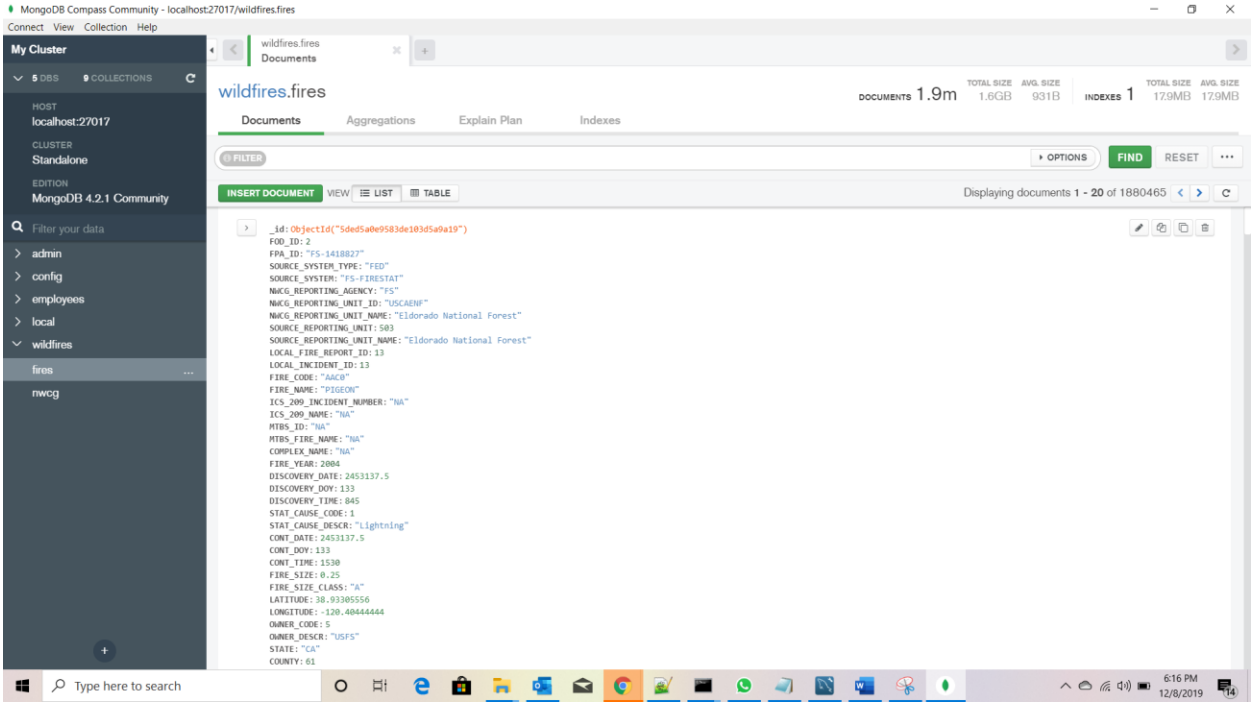
Physical Mongo Database

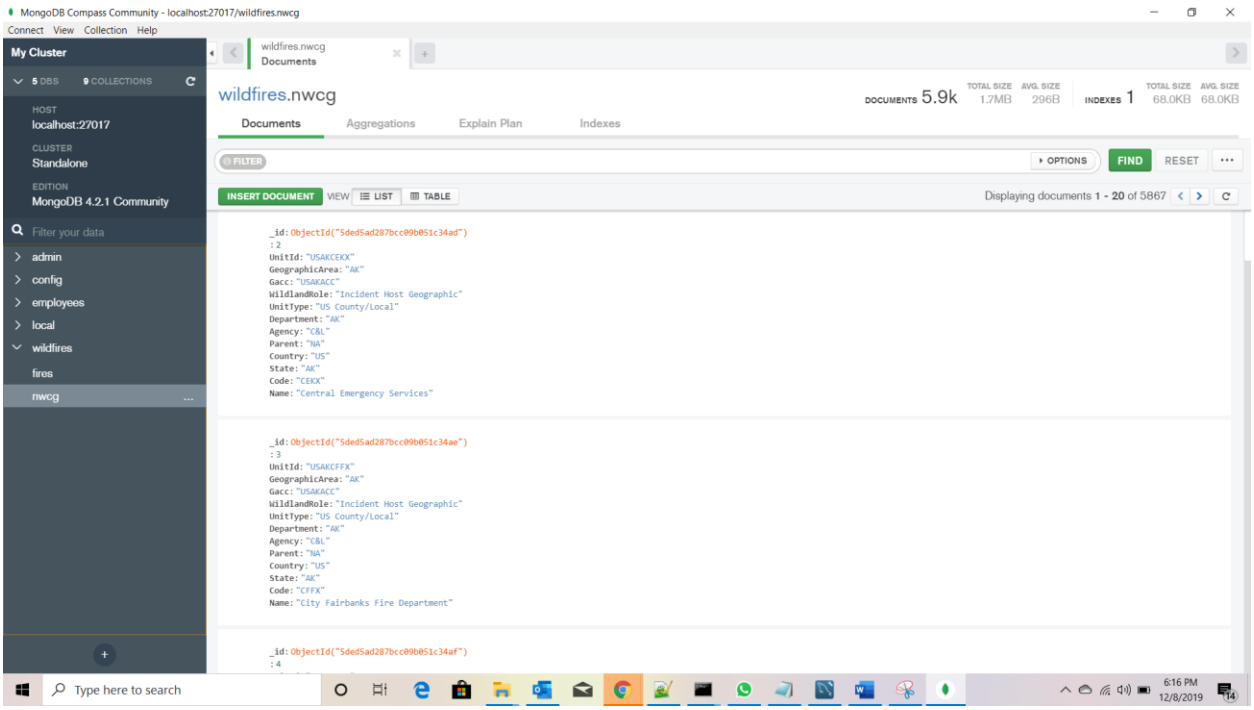
Assumptions/Notes About Data Set

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

```
C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db=wildfires --collection=fires --drop --headerline --type=csv --file="D:\Fall'19\BUAN 6320 - DB foundations - Ravi N\Project\tables\fires_info.csv"
2019-12-08T14:16:14.319-0600    connected to: mongodb://localhost/
2019-12-08T14:16:14.320-0600    dropping: wildfires.fires
2019-12-08T14:16:17.321-0600    [#.......................] wildfires.fires     32.8MB/539MB (6.1%)
2019-12-08T14:16:20.321-0600    [##......................] wildfires.fires     64.4MB/539MB (12.0%)
2019-12-08T14:16:23.321-0600    [####....................] wildfires.fires     96.2MB/539MB (17.9%)
2019-12-08T14:16:26.320-0600    [#####...................] wildfires.fires     132MB/539MB (24.5%)
2019-12-08T14:16:29.321-0600    [#######.................] wildfires.fires     168MB/539MB (31.2%)
2019-12-08T14:16:32.320-0600    [########................] wildfires.fires     203MB/539MB (37.6%)
2019-12-08T14:16:35.320-0600    [#########...............] wildfires.fires     237MB/539MB (43.9%)
2019-12-08T14:16:38.320-0600    [##########..............] wildfires.fires     268MB/539MB (49.7%)
2019-12-08T14:16:41.321-0600    [############............] wildfires.fires     302MB/539MB (56.1%)
2019-12-08T14:16:44.320-0600    [#############...........] wildfires.fires     338MB/539MB (62.7%)
2019-12-08T14:16:47.320-0600    [###############.........] wildfires.fires     376MB/539MB (69.8%)
2019-12-08T14:16:50.320-0600    [################........] wildfires.fires     408MB/539MB (75.8%)
2019-12-08T14:16:53.319-0600    [##################......] wildfires.fires     442MB/539MB (81.9%)
2019-12-08T14:16:56.319-0600    [###################.....] wildfires.fires     472MB/539MB (87.5%)
2019-12-08T14:16:59.321-0600    [####################....] wildfires.fires     507MB/539MB (94.1%)
2019-12-08T14:17:01.964-0600    [########################] wildfires.fires     539MB/539MB (100.0%)
2019-12-08T14:17:01.964-0600    1880465 document(s) imported successfully. 0 document(s) failed to import.

C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db=wildfires --collection=nwcg --drop --headerline --type=csv --file="D:\Fall'19\BUAN 6320 - DB foundations - Ravi N\Project\tables\nwcg (1).csv"
2019-12-08T14:19:30.901-0600    connected to: mongodb://localhost/
2019-12-08T14:19:30.902-0600    dropping: wildfires.nwcg
2019-12-08T14:19:30.991-0600    5867 document(s) imported successfully. 0 document(s) failed to import.
```

```
> show collections
fires
nwcg
> show dbs
admin      0.000GB
config     0.000GB
employees  0.147GB
local      0.000GB
wildfires  0.264GB
> use wildfires
switched to db wildfires
> show collections
fires
nwcg
>
```


MongoDB Compass Community - localhost:27017/wildfires.fires
Connect View Collection Help
My Cluster
5 DBS 9 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.2.1 Community
Filter your data
admin
config
employees
local
wildfires
fires
nwcg
wildfires.fires
Documents
wildfires.fires
DOCUMENTS 1.9m
TOTAL SIZE 1.6GB
AVG. SIZE 931B
INDEXES 1
TOTAL SIZE 17.9MB
AVG. SIZE 17.9MB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 1880465
_id: ObjectId("5ded5a0e9583de103d5a9a19")
FOD_ID: 2
FPA_ID: "FS-1418827"
SOURCE_SYSTEM_TYPE: "FED"
SOURCE_SYSTEM: "FS-FIRESTAT"
NWCG_REPORTING_AGENCY: "FS"
NWCG_REPORTING_UNIT_ID: "USCAENF"
NWCG_REPORTING_UNIT_NAME: "Eldorado National Forest"
SOURCE_REPORTING_UNIT: 503
SOURCE_REPORTING_UNIT_NAME: "Eldorado National Forest"
LOCAL_FIRE_REPORT_ID: 13
LOCAL_INCIDENT_ID: 13
FIRE_CODE: "AAC0"
FIRE_NAME: "PIGEON"
ICS_209_INCIDENT_NUMBER: "NA"
ICS_209_NAME: "NA"
MTBS_ID: "NA"
MTBS_FIRE_NAME: "NA"
COMPLEX_NAME: "NA"
FIRE_YEAR: 2004
DISCOVERY_DATE: 2453137.5
DISCOVERY_DOY: 133
DISCOVERY_TIME: 845
STAT_CAUSE_CODE: 1
STAT_CAUSE_DESCR: "Lightning"
CONT_DATE: 2453137.5
CONT_DOY: 133
CONT_TIME: 1530
FIRE_SIZE: 0.25
FIRE_SIZE_CLASS: "A"
LATITUDE: 38.93305556
LONGITUDE: -120.40444444
OWNER_CODE: 5
OWNER_DESCR: "USFS"
STATE: "CA"
COUNTY: 61
Type here to search
6:16 PM
12/8/2019


MongoDB Compass Community - localhost:27017/wildfires.nwcg
Connect View Collection Help
My Cluster
5 DBS 9 COLLECTIONS
HOST
localhost:27017
CLUSTER
Standalone
EDITION
MongoDB 4.2.1 Community
Filter your data
admin
config
employees
local
wildfires
fires
nwcg
wildfires.nwcg
Documents
wildfires.nwcg
DOCUMENTS 5.9k
TOTAL SIZE 1.7MB
AVG. SIZE 296B
INDEXES 1
TOTAL SIZE 68.0KB
AVG. SIZE 68.0KB
Documents
Aggregations
Explain Plan
Indexes
FILTER
OPTIONS
FIND
RESET
INSERT DOCUMENT
VIEW
LIST
TABLE
Displaying documents 1 - 20 of 5867
_id: ObjectId("5ded5ad287bcc09b051c34ad")
: 2
UnitId: "USAKCEKX"
GeographicArea: "AK"
Gacc: "USAKACC"
WildlandRole: "Incident Host Geographic"
UnitType: "US County/Local"
Department: "AK"
Agency: "C&L"
Parent: "NA"
Country: "US"
State: "AK"
Code: "CEKX"
Name: "Central Emergency Services"
_id: ObjectId("5ded5ad287bcc09b051c34ae")
: 3
UnitId: "USAKCFFX"
GeographicArea: "AK"
Gacc: "USAKACC"
WildlandRole: "Incident Host Geographic"
UnitType: "US County/Local"
Department: "AK"
Agency: "C&L"
Parent: "NA"
Country: "US"
State: "AK"
Code: "CFFX"
Name: "City Fairbanks Fire Department"
_id: ObjectId("5ded5ad287bcc09b051c34af")
: 4
Type here to search
6:16 PM
12/8/2019

Data in the Database

Collection Name	Relationshps With Other Collections (if any)	# of Documents in Collection
Fires	NWCG_REPORTING_UNIT_ID field in Fires is linked with UnitId field in NWCG Collection	1880465
NWCG	NWCG_REPORTING_UNIT_ID field in Fires is linked with UnitId field in NWCG Collection	5867

MongoDB Queries/Code

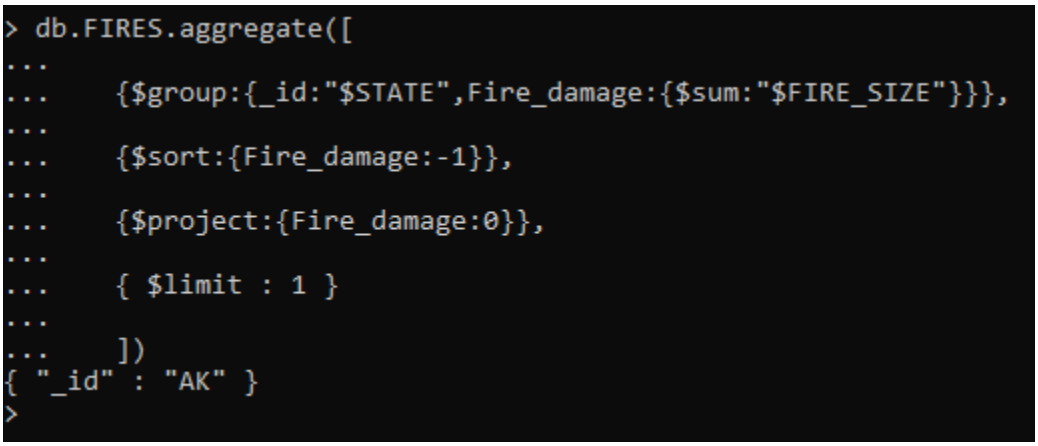
Query 1

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 best chance to win a share of the fund?

As all the states are interested in the billion-dollar fund we group by state, but the state with which has highest number of fires will receive the fund. AK has the highest fires.

Translation: Project the highest fire_damage grouping by state in descending order.

Screen Shot of MongoDB Query/Code and Results:



Query 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 most appropriate.

As the reporting agencies has suggested every 3 children should be accompanied by one adult, we group by source reporting unit in descending order and selecting first three forests will give us 3 documents.

Translation: Report the source_reporting_unit_name where the source_reporting_unit_name contains forest and stat_cause_code is 8 by sorting them in descending order.

Screen Shot of MongoDB Query/Code and Results:

```
> db.FIRES.aggregate([
...
...     {$match: {"SOURCE_REPORTING_UNIT_NAME":/Forest$/, STAT_CAUSE_CODE: 8}},
...
...     {$group: {_id:{SOURCE_REPORTING_UNIT_NAME: "$SOURCE_REPORTING_UNIT_NAME"}, Firescount:{$sum:1}}},
...
...     {$sort: {Firescount:-1, _id:-1}},
...
...     {$limit:3}
...
... ])
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "San Bernardino National Forest" }, "Firescount" : 162 }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Angeles National Forest" }, "Firescount" : 156 }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Huron-Manistee National Forest" }, "Firescount" : 94 }
>
```

Query 3

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

As the statement states human actions are to be blamed for most wildfires, we need the count the number of fires that are caused by human and nature. If the count of fires caused by human are more this proves the above statement is true. 2 documents are retrieved.

Translation: Projecting the summary which contains human and nature. If the stat_cause_code is 1 then its nature else human based on which the count of each stat_cause_code is projected.

Screen Shot of MongoDB Query/Code and Results:

```
> db.FIRES.aggregate( [
...
...   {$match:{STAT_CAUSE_CODE:{$nin:[9,13]}}},
...
...   {
...
...     $project:
...       {
...         "STAT_CAUSE_CODE" : 1,
...         "summary" :
...         {
...           $switch:
...             {
...               branches: [
...
...                 { case: { $eq : [ '$STAT_CAUSE_CODE',1 ]}, then: "Nature" }
...
...               ],
...
...               default: "Human"
...             }
...           }
...       }
...    },
...
...    {$group:{_id:"$summary",count: {$sum:1 }}}
...
... ])
{ "_id" : "Human", "count" : 1111469 }
{ "_id" : "Nature", "count" : 278468 }
```

Query 4

How many wildfires were reported by more than one unit/agency?

Retrieve the number of wildfires reported in more than one unit/agency. One document is retrieved.

Translation: Projecting the fire_name and the source_reporting_unit where the no_of_units_reported is greater than 1.

Screen Shot of MongoDB Query/Code and Results:

```
> db.FIRES.aggregate([
... {$project:{FIRE_NAME:1,SOURCE_REPORTING_UNIT:1}},
... {$group:{_id:"$FIRE_NAME",No_of_units_reported:{$sum:"$SOURCE_REPORTING_UNIT"}}},
... {$match:{No_of_units_reported:{$gte:1}}}
... ]).toArray().length
76442
```

Query 5

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

Retrieve the forests where the fires lasted for more than 2 days. 128 documents were retrieved.

Translation: Projecting the source_unit_reporting_name where the difference between controlled date and discovery date is greater than 2 and source_unit_reporting_name contains “forest“ and firecount is greater than 1.

Screen Shot of MongoDB Query/Code and Results:

```
> db.FIRES.aggregate([
...   { $addFields: { date_diff: { $subtract: ["$CONT_DATE", "$DISCOVERY_DATE"]} } },
...   { $match: { date_diff: { $gte : 2}} },
...   { $project: { SOURCE_REPORTING_UNIT_NAME : 1 }  },
...   { $match:  { "SOURCE_REPORTING_UNIT_NAME":/Forest$/ } },
...   { $group:  { _id:{SOURCE_REPORTING_UNIT_NAME: "$SOURCE_REPORTING_UNIT_NAME"}, Firescount:{$sum:1} } },
...   { $match: { Firescount: { $gte : 1}} },
...   { $project: { SOURCE_REPORTING_UNIT_NAME : 1 } } ])
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Caribou National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Modoc National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Uinta-Wasatch-Cache National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Homochitto National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Nebraska National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Colville National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Manti-LaSal National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Custer National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Chequamegon-Nicolet National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "George Washington and Jefferson National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Ottawa National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Inyo National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Santa Fe National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Caribou-Targhee National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Umatilla National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Kootenai National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Savannah River Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Boise National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Mark Twain National Forest" } }
{ "_id" : { "SOURCE_REPORTING_UNIT_NAME" : "Hoosier National Forest" } }
Type "it" for more
```

Query 6

Which forest had the most number of fires?

Report the forest which has most number of fires. One document is retrieved

Translation: Project nwcg_reporting_unit_name which contains " forest" by sorting in descending order.

Screen Shot of MongoDB Query/Code and Results:

```
> db.FIRES.aggregate([
... {$match: {"NWCG_REPORTING_UNIT_NAME":/Forest$/}},
... {$group: {_id:{NWCG_REPORTING_UNIT_NAME: "$NWCG_REPORTING_UNIT_NAME"}, Firescount:{$sum:1}}},
... {$sort: {Firescount:-1, _id:-1}},
... {$limit:1}])
{ "_id" : { "NWCG_REPORTING_UNIT_NAME" : "Coconino National Forest" }, "Firescount" : 8812 }
```