Name: Sunita Inderjit

CS 623 Database Management

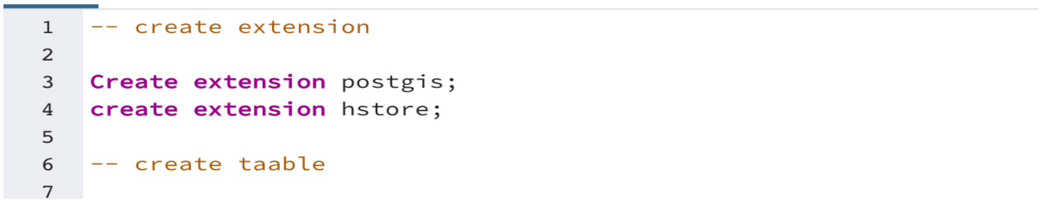
Project 2  Part 1 Analyzing Geographic Data

Topic:  Earthquake Data From USGS.GOV

Data Source: Earthquake data

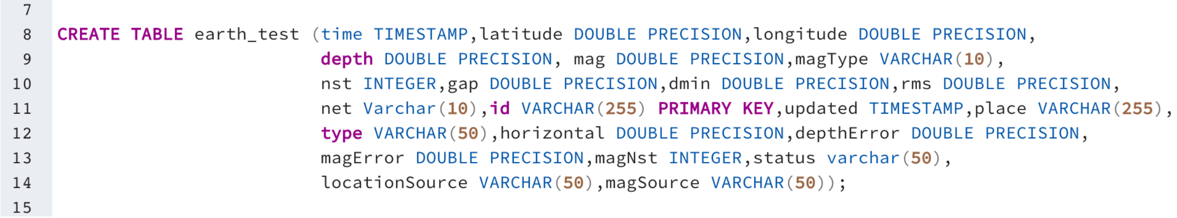
Objective:  Analyzing earthquake data, to find insight on magnitude, place and time.

## Create the necessary extension to process geographic data



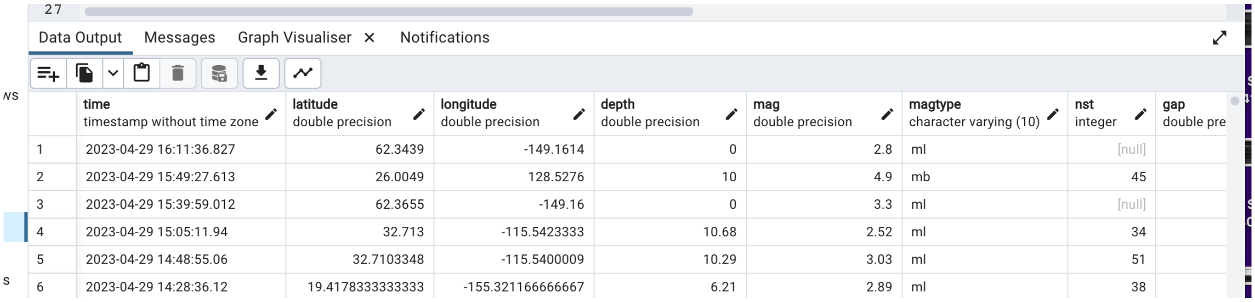
```
-- create extension

Create extension postgis;
create extension hstore;

-- create taable

```

## Create table called earth_test to load data



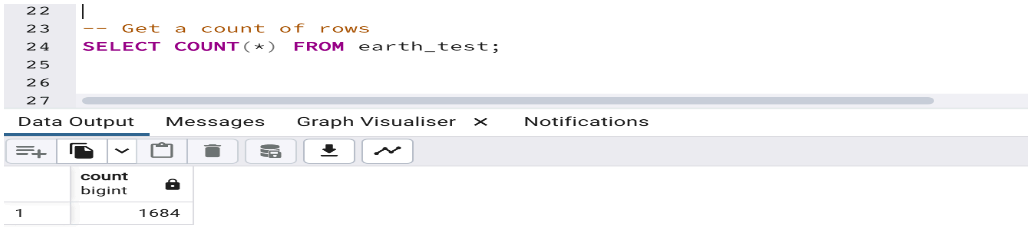
```

CREATE TABLE earth_test (time TIMESTAMP,latitude DOUBLE PRECISION,longitude DOUBLE PRECISION,
                         depth DOUBLE PRECISION, mag DOUBLE PRECISION,magType VARCHAR(10),
                         nst INTEGER,gap DOUBLE PRECISION,dmin DOUBLE PRECISION,rms DOUBLE PRECISION,
                         net Varchar(10),id VARCHAR(255) PRIMARY KEY,updated TIMESTAMP,place VARCHAR(255),
                         type VARCHAR(50),horizontal DOUBLE PRECISION,depthError DOUBLE PRECISION,
                         magError DOUBLE PRECISION,magNst INTEGER,status varchar(50),
                         locationSource VARCHAR(50),magSource VARCHAR(50));

```

## View the table



| | time<br>timestamp without time zone | latitude<br>double precision | longitude<br>double precision | depth<br>double precision | mag<br>double precision | magtype<br>character varying (10) | nst<br>integer | gap<br>double pre |
|---|---|---|---|---|---|---|---|---|
| 1 | 2023-04-29 16:11:36.827 | 62.3439 | -149.1614 | 0 | 2.8 | ml | [null] | |
| 2 | 2023-04-29 15:49:27.613 | 26.0049 | 128.5276 | 10 | 4.9 | mb | 45 | |
| 3 | 2023-04-29 15:39:59.012 | 62.3655 | -149.16 | 0 | 3.3 | ml | [null] | |
| 4 | 2023-04-29 15:05:11.94 | 32.713 | -115.5423333 | 10.68 | 2.52 | ml | 34 | |
| 5 | 2023-04-29 14:48:55.06 | 32.7103348 | -115.5400009 | 10.29 | 3.03 | ml | 51 | |
| 6 | 2023-04-29 14:28:36.12 | 19.4178333333333 | -155.321166666667 | 6.21 | 2.89 | ml | 38 | |

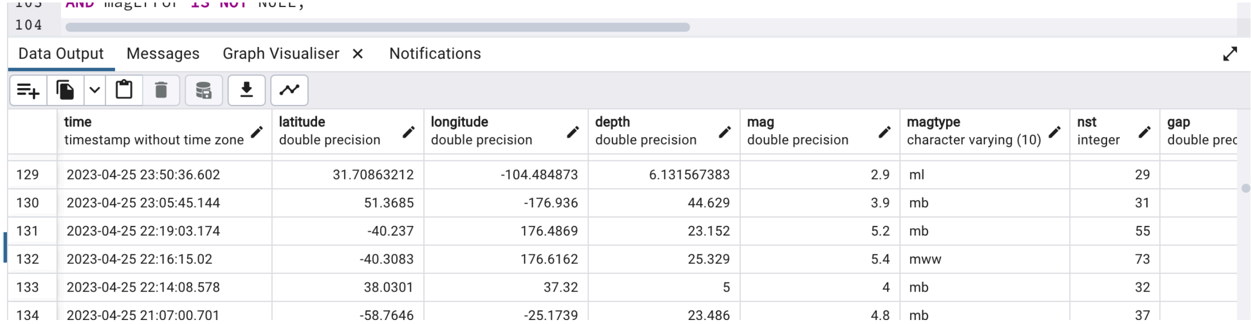
## Get count of total rows in table



```

-- Get a count of rows
SELECT COUNT(*) FROM earth_test;


```

| | count<br>bigint |
|---|---|
| 1 | 1684 |

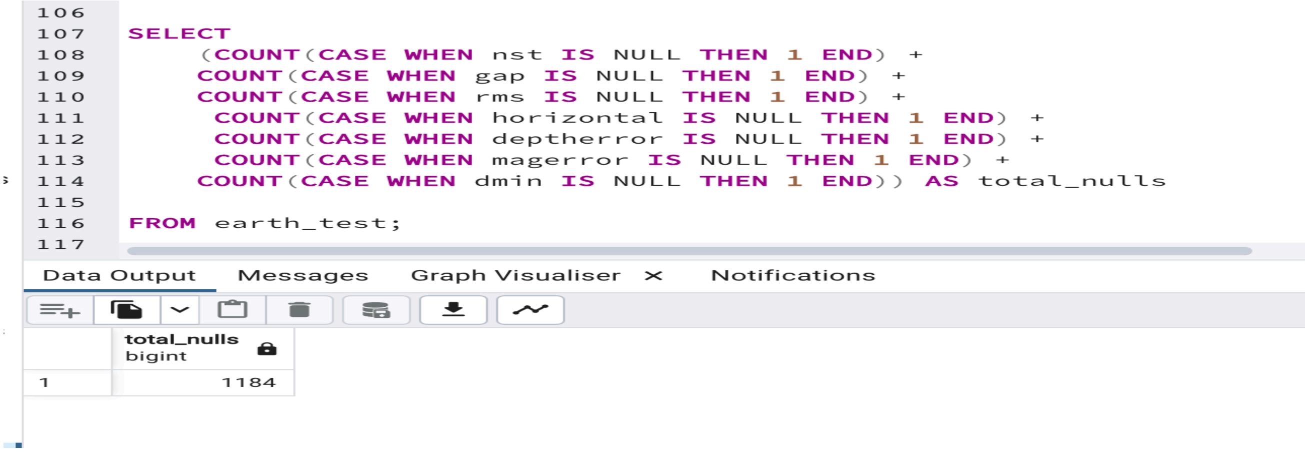
Select the all columns without nulls values using  ‘NOT IN”



```
AND magError IS NOT NULL;
```

| | time<br>timestamp without time zone | latitude<br>double precision | longitude<br>double precision | depth<br>double precision | mag<br>double precision | magtype<br>character varying (10) | nst<br>integer | gap<br>double prec |
|---|---|---|---|---|---|---|---|---|
| 129 | 2023-04-25 23:50:36.602 | 31.70863212 | -104.484873 | 6.131567383 | 2.9 | ml | 29 | |
| 130 | 2023-04-25 23:05:45.144 | 51.3685 | -176.936 | 44.629 | 3.9 | mb | 31 | |
| 131 | 2023-04-25 22:19:03.174 | -40.237 | 176.4869 | 23.152 | 5.2 | mb | 55 | |
| 132 | 2023-04-25 22:16:15.02 | -40.3083 | 176.6162 | 25.329 | 5.4 | mww | 73 | |
| 133 | 2023-04-25 22:14:08.578 | 38.0301 | 37.32 | 5 | 4 | mb | 32 | |
| 134 | 2023-04-25 21:07:00.701 | -58.7646 | -25.1739 | 23.486 | 4.8 | mb | 37 | |

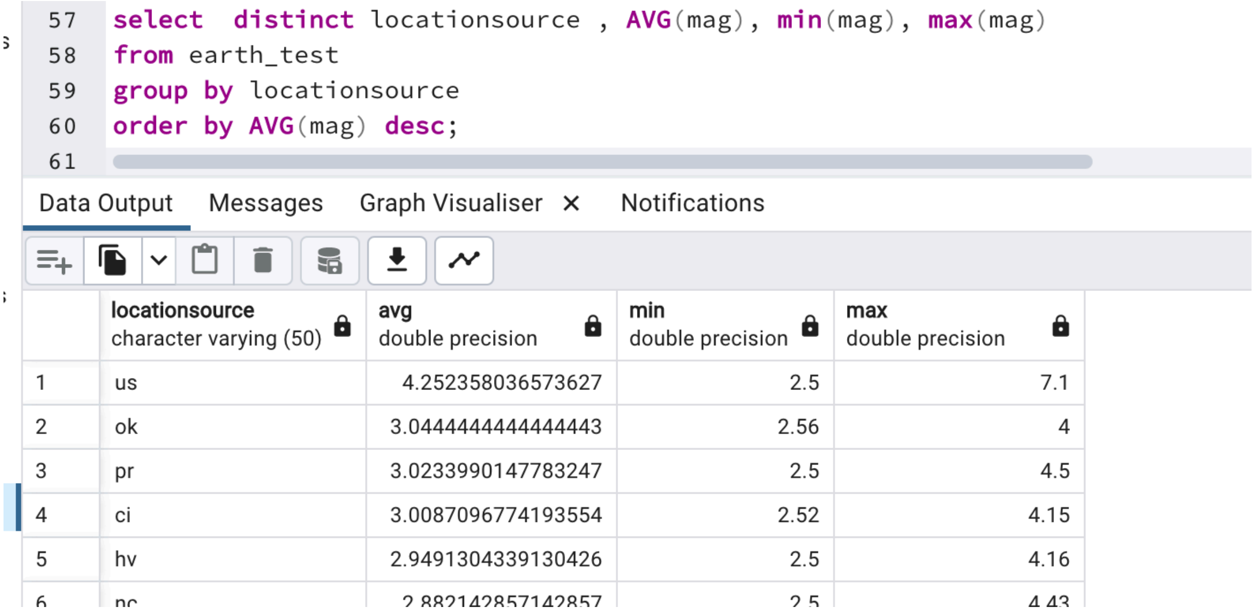
There are many null value in the table. Total null values in table is 1184



```
SELECT
    (COUNT(CASE WHEN nst IS NULL THEN 1 END) +
    COUNT(CASE WHEN gap IS NULL THEN 1 END) +
    COUNT(CASE WHEN rms IS NULL THEN 1 END) +
     COUNT(CASE WHEN horizontal IS NULL THEN 1 END) +
     COUNT(CASE WHEN deptherror IS NULL THEN 1 END) +
     COUNT(CASE WHEN magerror IS NULL THEN 1 END) +
    COUNT(CASE WHEN dmin IS NULL THEN 1 END)) AS total_nulls

FROM earth_test;
```

| | total_nulls<br>bigint |
|---|---|
| 1 | 1184 |

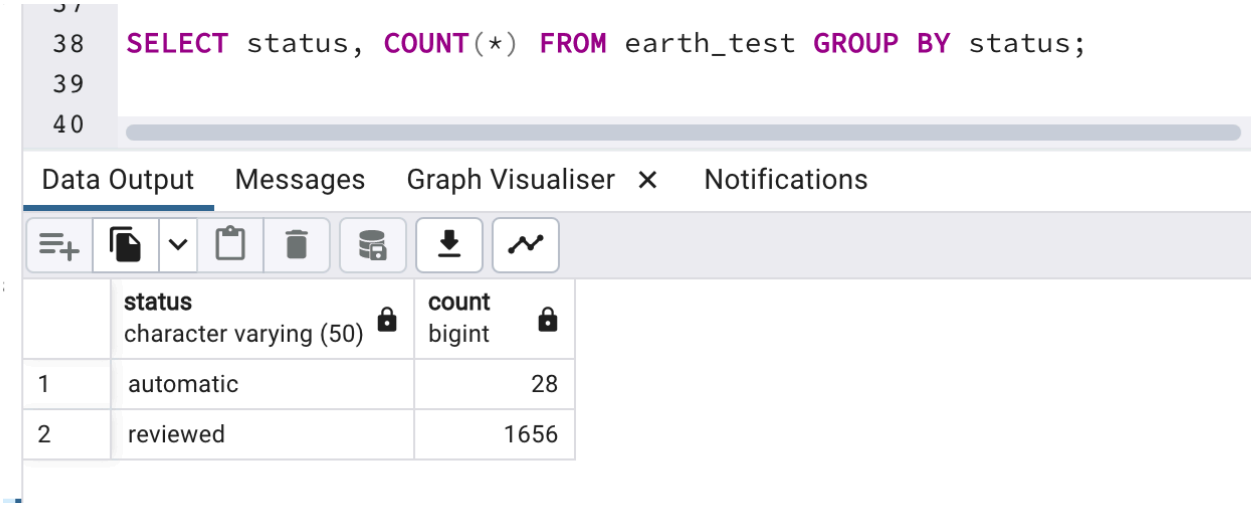
Using aggregate function to get  the Avg, min and max magnitude base on location source



```
select  distinct locationsource , AVG(mag), min(mag), max(mag)
from earth_test
group by locationsource
order by AVG(mag) desc;
```

| | locationsource<br>character varying (50) | avg<br>double precision | min<br>double precision | max<br>double precision |
|---|---|---|---|---|
| 1 | us | 4.252358036573627 | 2.5 | 7.1 |
| 2 | ok | 3.0444444444444443 | 2.56 | 4 |
| 3 | pr | 3.0233990147783247 | 2.5 | 4.5 |
| 4 | ci | 3.0087096774193554 | 2.52 | 4.15 |
| 5 | hv | 2.9491304339130426 | 2.5 | 4.16 |
| 6 | nc | 2.882142857142857 | 2.5 | 4.43 |

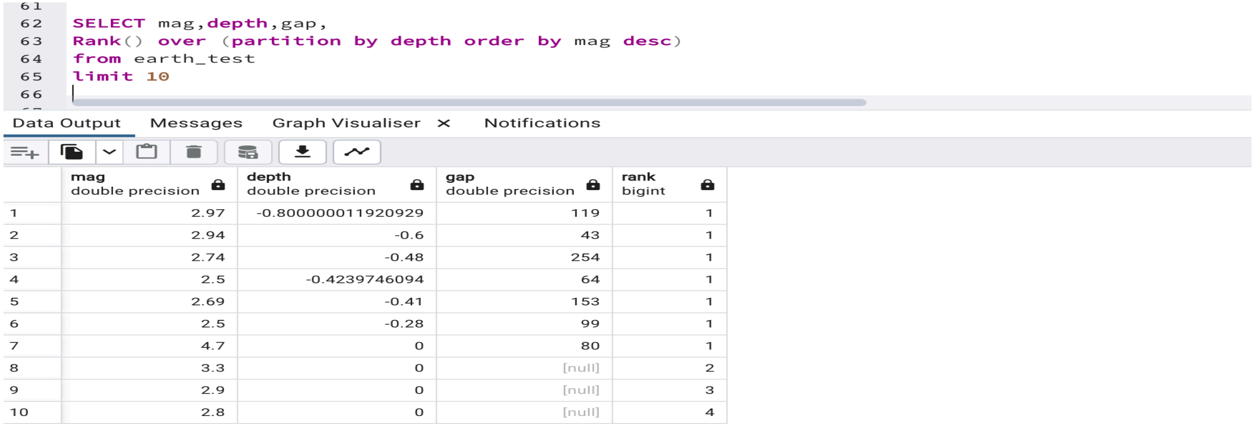
Get of  status: Automatic means that computer system detection  and review indicate that analyst reviews incidents. 1656 status was reviewed by analyst



```sql
SELECT status, COUNT(*) FROM earth_test GROUP BY status;
```

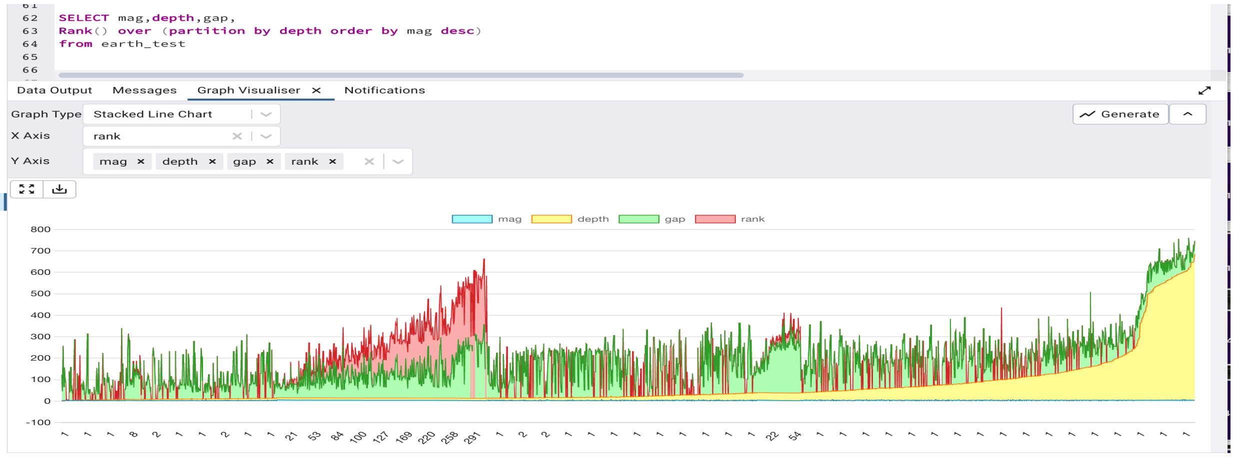
| | status<br>character varying (50) | count<br>bigint |
|---|---|---|
| 1 | automatic | 28 |
| 2 | reviewed | 1656 |

Using rank function method  to  rank mag, dep and gap



```sql
SELECT mag,depth,gap,
Rank() over (partition by depth order by mag desc)
from earth_test
limit 10
```

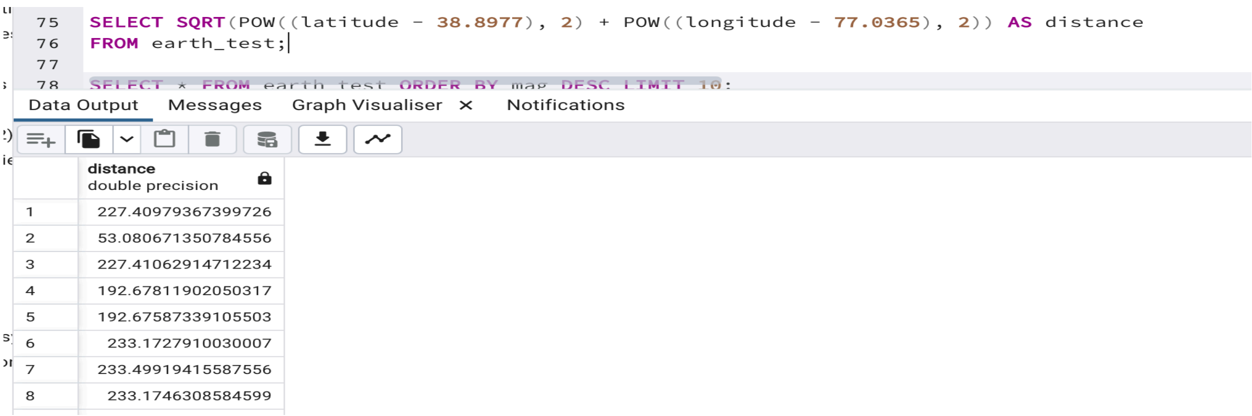
| | mag<br>double precision | depth<br>double precision | gap<br>double precision | rank<br>bigint |
|---|---|---|---|---|
| 1 | 2.97 | -0.800000011920929 | 119 | 1 |
| 2 | 2.94 | -0.6 | 43 | 1 |
| 3 | 2.74 | -0.48 | 254 | 1 |
| 4 | 2.5 | -0.4239746094 | 64 | 1 |
| 5 | 2.69 | -0.41 | 153 | 1 |
| 6 | 2.5 | -0.28 | 99 | 1 |
| 7 | 4.7 | 0 | 80 | 1 |
| 8 | 3.3 | 0 | [null] | 2 |
| 9 | 2.9 | 0 | [null] | 3 |
| 10 | 2.8 | 0 | [null] | 4 |



```sql
SELECT mag,depth,gap,
Rank() over (partition by depth order by mag desc)
from earth_test
```

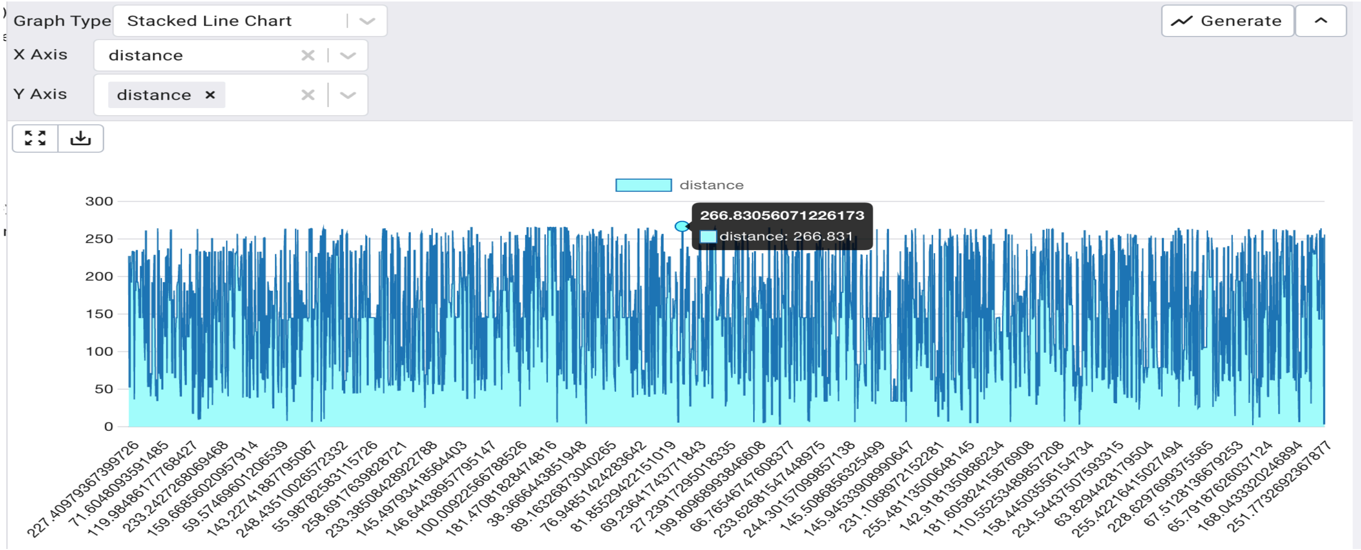
Graph Type: Stacked Line Chart
X Axis: rank
Y Axis: mag × depth × gap × rank ×

## Calculate distant between point



```
SELECT SQRT(POW((latitude - 38.8977), 2) + POW((longitude - 77.0365), 2)) AS distance
FROM earth_test;

SELECT * FROM earth_test ORDER BY mag DESC LIMIT 10;
```

|  | distance<br>double precision |
|---|---|
| 1 | 227.40979367399726 |
| 2 | 53.080671350784556 |
| 3 | 227.41062914712234 |
| 4 | 192.67811902050317 |
| 5 | 192.67587339105503 |
| 6 | 233.1727910030007 |
| 7 | 233.49919415587556 |
| 8 | 233.1746308584599 |

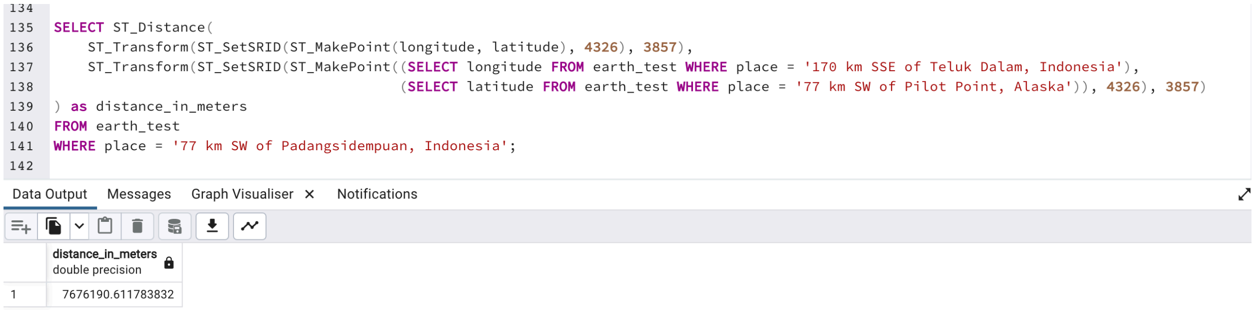


Graph Type: Stacked Line Chart

X Axis: distance

Y Axis: distance

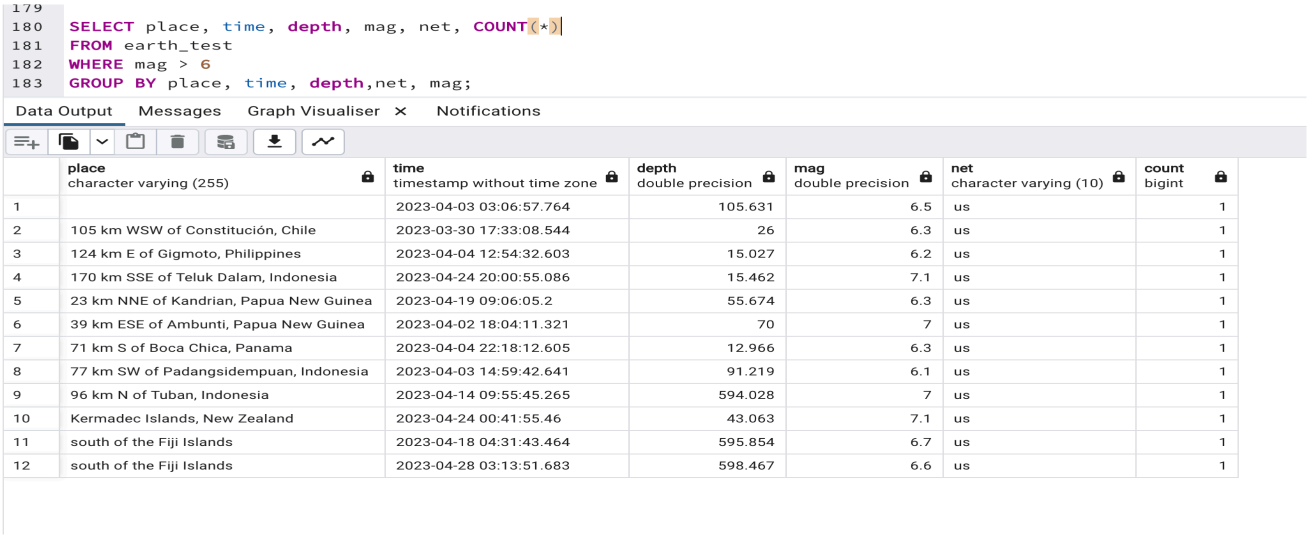
## Calculating distant between point of interest : The two location has 7.1 mag earthquake in April of 2023



```
SELECT ST_Distance(
    ST_Transform(ST_SetSRID(ST_MakePoint(longitude, latitude), 4326), 3857),
    ST_Transform(ST_SetSRID(ST_MakePoint((SELECT longitude FROM earth_test WHERE place = '170 km SSE of Teluk Dalam, Indonesia'),
                                         (SELECT latitude FROM earth_test WHERE place = '77 km SW of Pilot Point, Alaska')), 4326), 3857)
) as distance_in_meters
FROM earth_test
WHERE place = '77 km SW of Padangsidempuan, Indonesia';
```

|  | distance_in_meters<br>double precision |
|---|---|
| 1 | 7676190.611783832 |

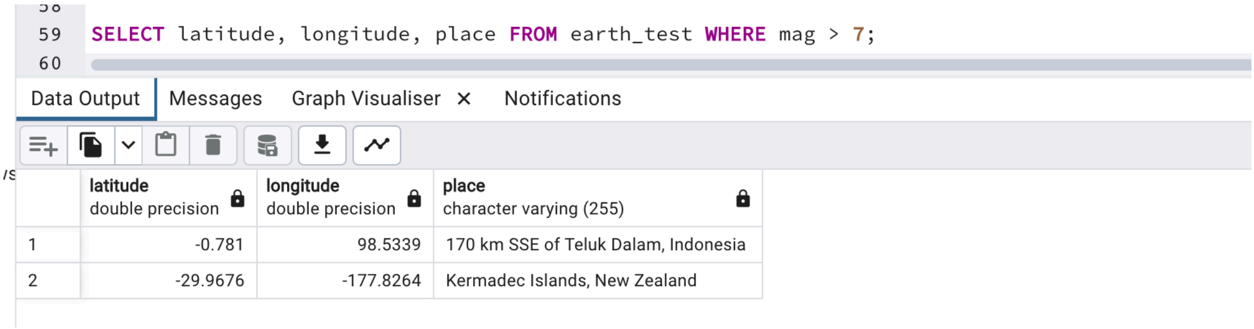
Using the group by method to group the highest mag by dept, time  and place. Here you can see the Indonesia  has two  big earthquake in  April of 2023



```
SELECT place, time, depth, mag, net, COUNT(*)
FROM earth_test
WHERE mag > 6
GROUP BY place, time, depth,net, mag;
```

| | place<br>character varying (255) | time<br>timestamp without time zone | depth<br>double precision | mag<br>double precision | net<br>character varying (10) | count<br>bigint |
|---|---|---|---|---|---|---|
| 1 | | 2023-04-03 03:06:57.764 | 105.631 | 6.5 | us | 1 |
| 2 | 105 km WSW of Constitución, Chile | 2023-03-30 17:33:08.544 | 26 | 6.3 | us | 1 |
| 3 | 124 km E of Gigmoto, Philippines | 2023-04-04 12:54:32.603 | 15.027 | 6.2 | us | 1 |
| 4 | 170 km SSE of Teluk Dalam, Indonesia | 2023-04-24 20:00:55.086 | 15.462 | 7.1 | us | 1 |
| 5 | 23 km NNE of Kandrian, Papua New Guinea | 2023-04-19 09:06:05.2 | 55.674 | 6.3 | us | 1 |
| 6 | 39 km ESE of Ambunti, Papua New Guinea | 2023-04-02 18:04:11.321 | 70 | 7 | us | 1 |
| 7 | 71 km S of Boca Chica, Panama | 2023-04-04 22:18:12.605 | 12.966 | 6.3 | us | 1 |
| 8 | 77 km SW of Padangsidempuan, Indonesia | 2023-04-03 14:59:42.641 | 91.219 | 6.1 | us | 1 |
| 9 | 96 km N of Tuban, Indonesia | 2023-04-14 09:55:45.265 | 594.028 | 7 | us | 1 |
| 10 | Kermadec Islands, New Zealand | 2023-04-24 00:41:55.46 | 43.063 | 7.1 | us | 1 |
| 11 | south of the Fiji Islands | 2023-04-18 04:31:43.464 | 595.854 | 6.7 | us | 1 |
| 12 | south of the Fiji Islands | 2023-04-28 03:13:51.683 | 598.467 | 6.6 | us | 1 |

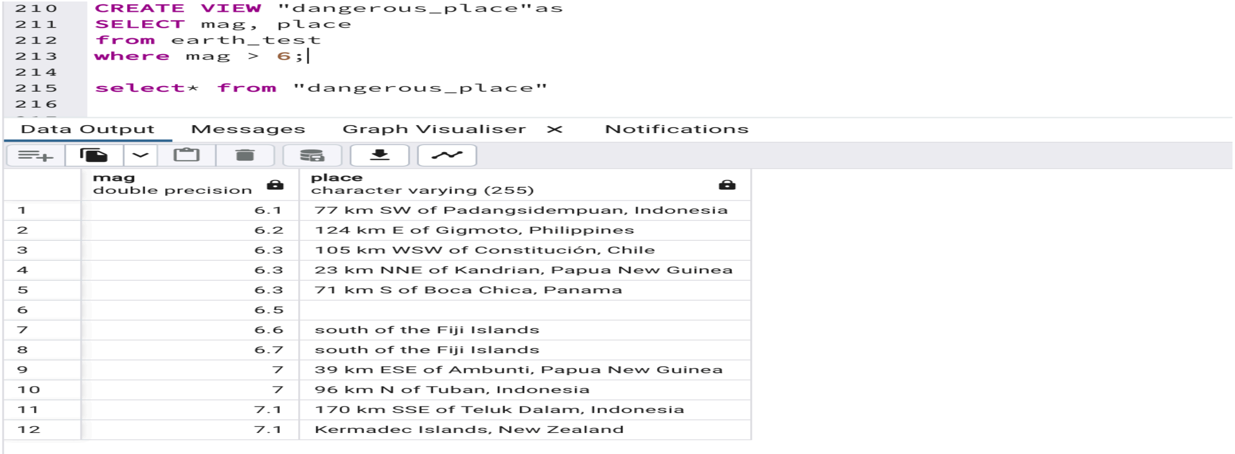
Country with mag > 7 was Indonesia and New Zealand



```
SELECT latitude, longitude, place FROM earth_test WHERE mag > 7;
```

| | latitude<br>double precision | longitude<br>double precision | place<br>character varying (255) |
|---|---|---|---|
| 1 | -0.781 | 98.5339 | 170 km SSE of Teluk Dalam, Indonesia |
| 2 | -29.9676 | -177.8264 | Kermadec Islands, New Zealand |

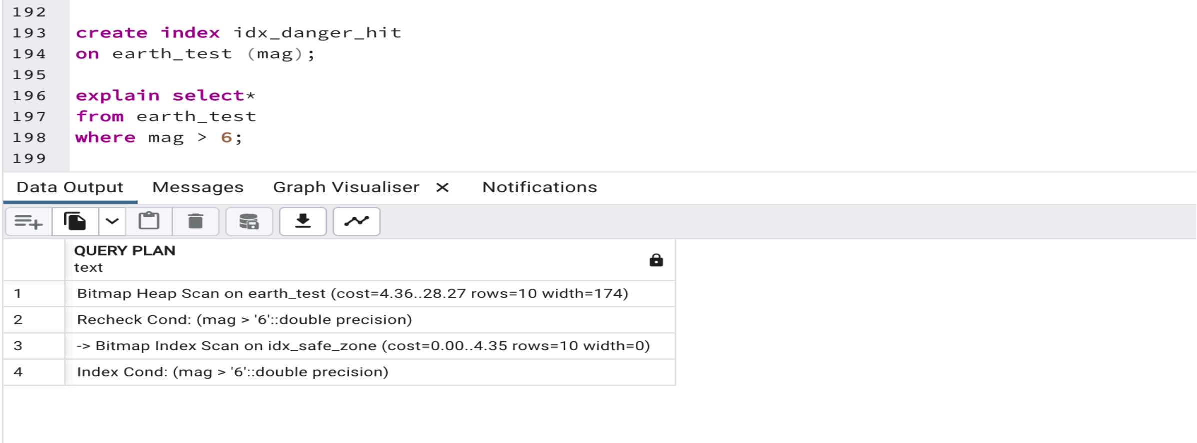
Create view   called most dangerous place



```
CREATE VIEW "dangerous_place"as
SELECT mag, place
from earth_test
where mag > 6;

select* from "dangerous_place"
```

| | mag<br>double precision | place<br>character varying (255) |
|---|---|---|
| 1 | 6.1 | 77 km SW of Padangsidempuan, Indonesia |
| 2 | 6.2 | 124 km E of Gigmoto, Philippines |
| 3 | 6.3 | 105 km WSW of Constitución, Chile |
| 4 | 6.3 | 23 km NNE of Kandrian, Papua New Guinea |
| 5 | 6.3 | 71 km S of Boca Chica, Panama |
| 6 | 6.5 | |
| 7 | 6.6 | south of the Fiji Islands |
| 8 | 6.7 | south of the Fiji Islands |
| 9 | 7 | 39 km ESE of Ambunti, Papua New Guinea |
| 10 | 7 | 96 km N of Tuban, Indonesia |
| 11 | 7.1 | 170 km SSE of Teluk Dalam, Indonesia |
| 12 | 7.1 | Kermadec Islands, New Zealand |

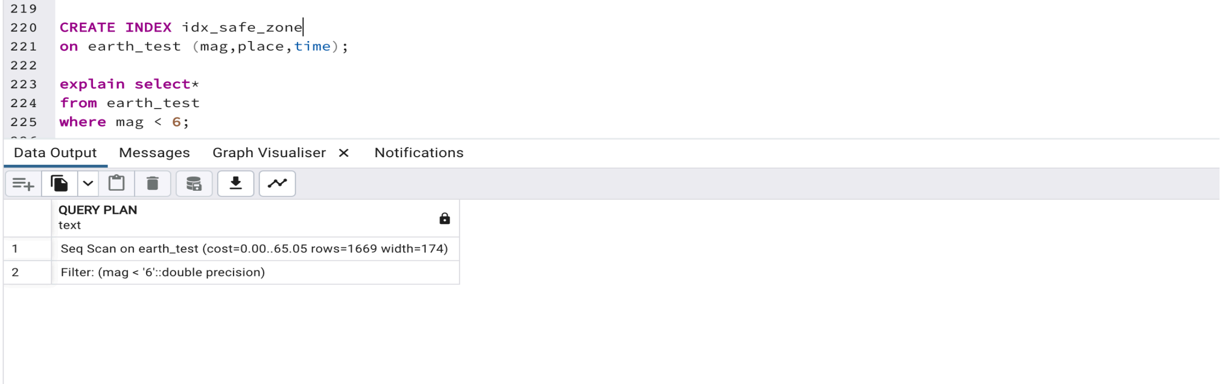
Create index  on name danger_hit. Give overview of cost of query



```
create index idx_danger_hit
on earth_test (mag);

explain select*
from earth_test
where mag > 6;
```

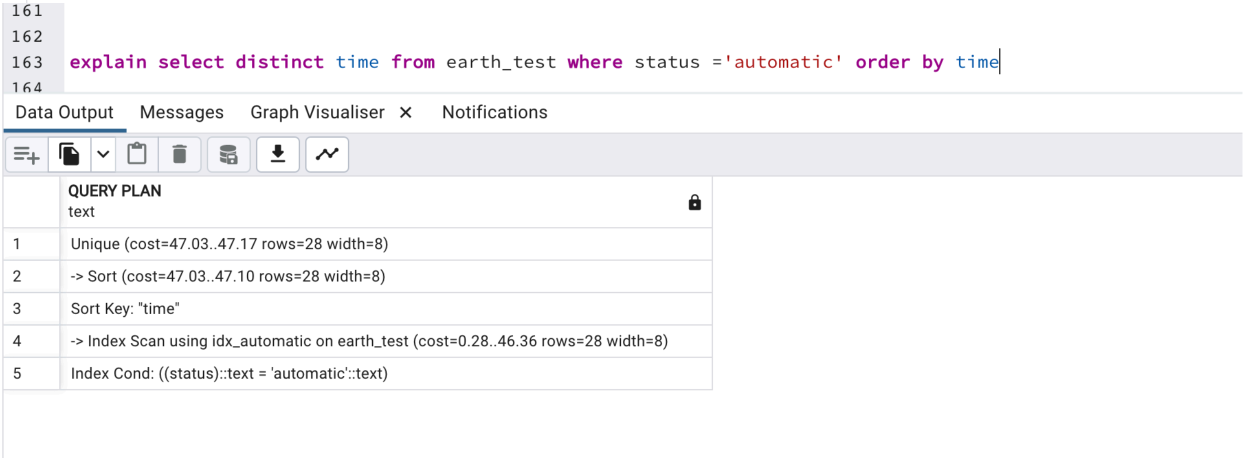
|   | QUERY PLAN<br>text |
|---|---|
| 1 | Bitmap Heap Scan on earth_test (cost=4.36..28.27 rows=10 width=174) |
| 2 | Recheck Cond: (mag > '6'::double precision) |
| 3 | -> Bitmap Index Scan on idx_safe_zone (cost=0.00..4.35 rows=10 width=0) |
| 4 | Index Cond: (mag > '6'::double precision) |

Create an index called safe zone.  Below we can see the detail of the sequential scan



```
CREATE INDEX idx_safe_zone
on earth_test (mag,place,time);

explain select*
from earth_test
where mag < 6;
```

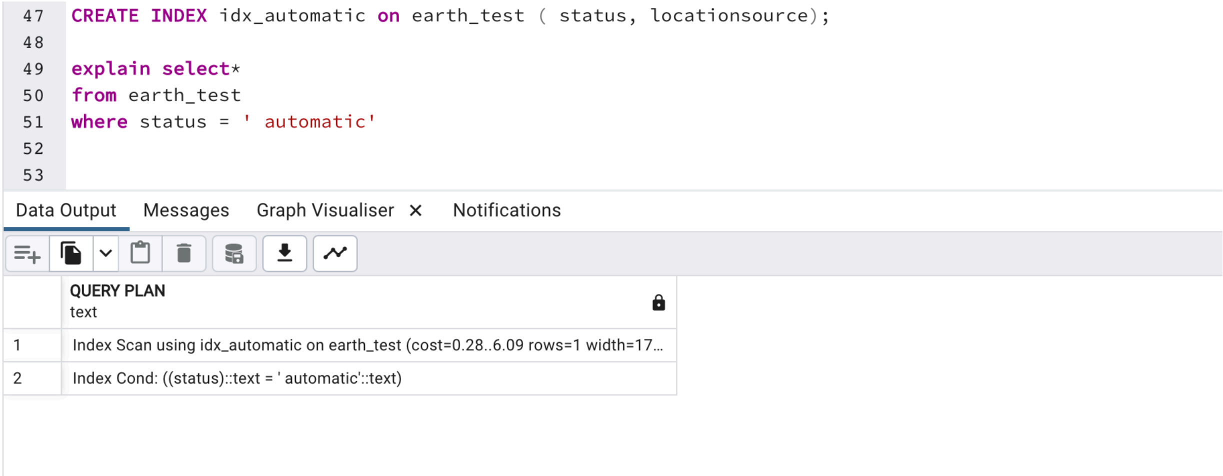
|   | QUERY PLAN<br>text |
|---|---|
| 1 | Seq Scan on earth_test (cost=0.00..65.05 rows=1669 width=174) |
| 2 | Filter: (mag < '6'::double precision) |



```
explain select distinct time from earth_test where status ='automatic' order by time
```

|   | QUERY PLAN<br>text |
|---|---|
| 1 | Unique (cost=47.03..47.17 rows=28 width=8) |
| 2 | -> Sort (cost=47.03..47.10 rows=28 width=8) |
| 3 | Sort Key: "time" |
| 4 | -> Index Scan using idx_automatic on earth_test (cost=0.28..46.36 rows=28 width=8) |
| 5 | Index Cond: ((status)::text = 'automatic'::text) |

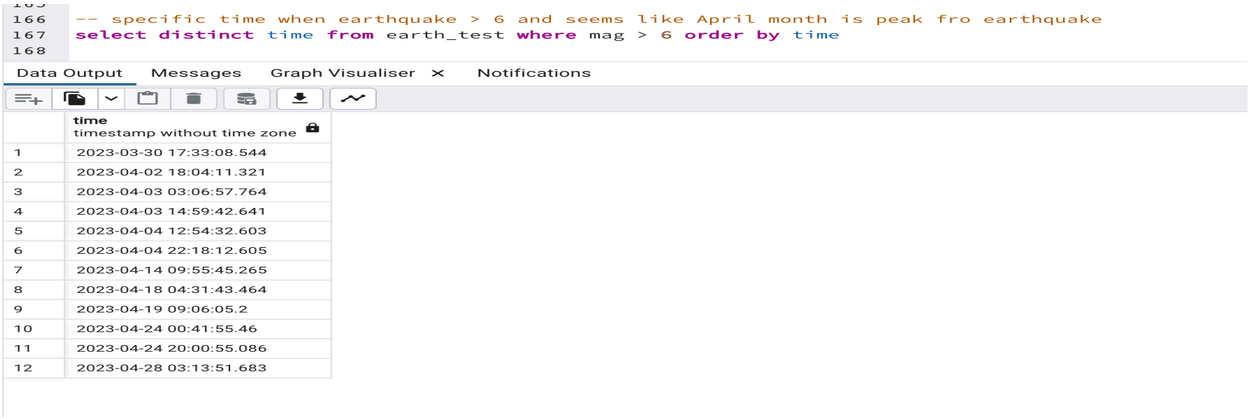
Create another index where status = automatic. Meaning the computer detect earthquake and here we can see the sequential scan and cost of query



```
CREATE INDEX idx_automatic on earth_test ( status, locationsource);

explain select*
from earth_test
where status = ' automatic'
```

| | QUERY PLAN<br>text |
|---|---|
| 1 | Index Scan using idx_automatic on earth_test (cost=0.28..6.09 rows=1 width=17… |
| 2 | Index Cond: ((status)::text = ' automatic'::text) |

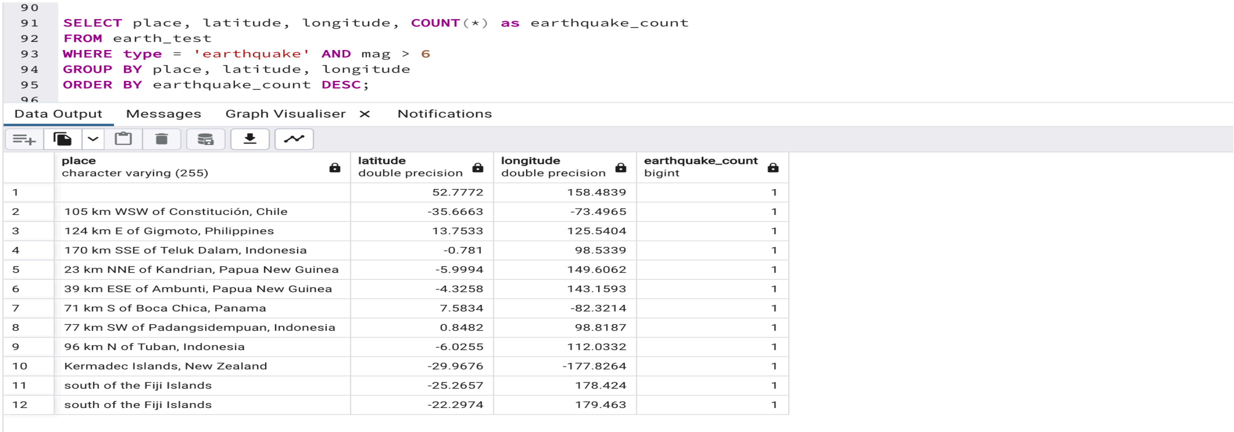
The distinct function to get the mag > 6 by time , as we can see here that April is the peak season for earthquake



```
-- specific time when earthquake > 6 and seems like April month is peak fro earthquake
select distinct time from earth_test where mag > 6 order by time
```

| | time<br>timestamp without time zone |
|---|---|
| 1 | 2023-03-30 17:33:08.544 |
| 2 | 2023-04-02 18:04:11.321 |
| 3 | 2023-04-03 03:06:57.764 |
| 4 | 2023-04-03 14:59:42.641 |
| 5 | 2023-04-04 12:54:32.603 |
| 6 | 2023-04-04 22:18:12.605 |
| 7 | 2023-04-14 09:55:45.265 |
| 8 | 2023-04-18 04:31:43.464 |
| 9 | 2023-04-19 09:06:05.2 |
| 10 | 2023-04-24 00:41:55.46 |
| 11 | 2023-04-24 20:00:55.086 |
| 12 | 2023-04-28 03:13:51.683 |

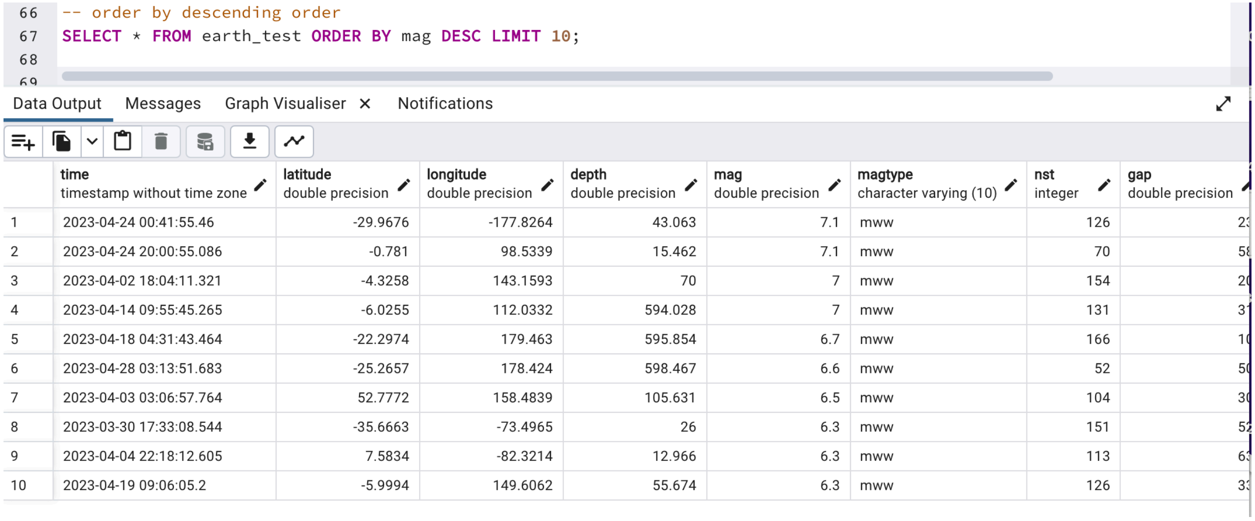
Using the count method to get a count by latitude and longitude with mag > 6



```
SELECT place, latitude, longitude, COUNT(*) as earthquake_count
FROM earth_test
WHERE type = 'earthquake' AND mag > 6
GROUP BY place, latitude, longitude
ORDER BY earthquake_count DESC;
```

| | place<br>character varying (255) | latitude<br>double precision | longitude<br>double precision | earthquake_count<br>bigint |
|---|---|---|---|---|
| 1 | | 52.7772 | 158.4839 | 1 |
| 2 | 105 km WSW of Constitución, Chile | -35.6663 | -73.4965 | 1 |
| 3 | 124 km E of Gigmoto, Philippines | 13.7533 | 125.5404 | 1 |
| 4 | 170 km SSE of Teluk Dalam, Indonesia | -0.781 | 98.5339 | 1 |
| 5 | 23 km NNE of Kandrian, Papua New Guinea | -5.9994 | 149.6062 | 1 |
| 6 | 39 km ESE of Ambunti, Papua New Guinea | -4.3258 | 143.1593 | 1 |
| 7 | 71 km S of Boca Chica, Panama | 7.5834 | -82.3214 | 1 |
| 8 | 77 km SW of Padangsidempuan, Indonesia | 0.8482 | 98.8187 | 1 |
| 9 | 96 km N of Tuban, Indonesia | -6.0255 | 112.0332 | 1 |
| 10 | Kermadec Islands, New Zealand | -29.9676 | -177.8264 | 1 |
| 11 | south of the Fiji Islands | -25.2657 | 178.424 | 1 |
| 12 | south of the Fiji Islands | -22.2974 | 179.463 | 1 |

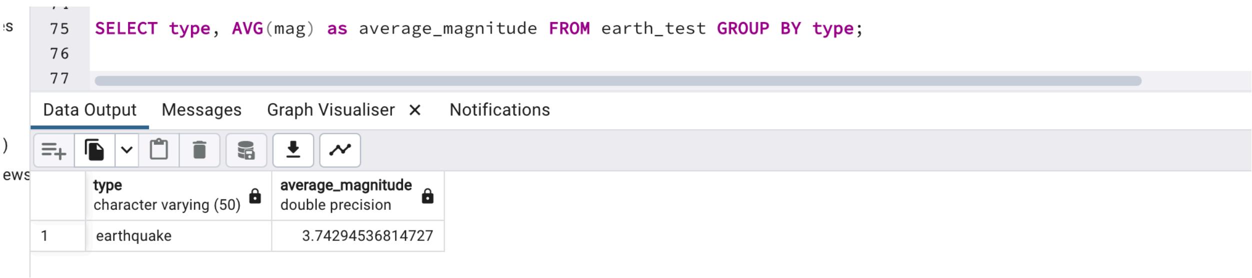
Order by the  descending  order limit 10



```
-- order by descending order
SELECT * FROM earth_test ORDER BY mag DESC LIMIT 10;
```

| | time<br>timestamp without time zone | latitude<br>double precision | longitude<br>double precision | depth<br>double precision | mag<br>double precision | magtype<br>character varying (10) | nst<br>integer | gap<br>double precision |
|---|---|---|---|---|---|---|---|---|
| 1 | 2023-04-24 00:41:55.46 | -29.9676 | -177.8264 | 43.063 | 7.1 | mww | 126 | 23 |
| 2 | 2023-04-24 20:00:55.086 | -0.781 | 98.5339 | 15.462 | 7.1 | mww | 70 | 58 |
| 3 | 2023-04-02 18:04:11.321 | -4.3258 | 143.1593 | 70 | 7 | mww | 154 | 20 |
| 4 | 2023-04-14 09:55:45.265 | -6.0255 | 112.0332 | 594.028 | 7 | mww | 131 | 31 |
| 5 | 2023-04-18 04:31:43.464 | -22.2974 | 179.463 | 595.854 | 6.7 | mww | 166 | 10 |
| 6 | 2023-04-28 03:13:51.683 | -25.2657 | 178.424 | 598.467 | 6.6 | mww | 52 | 50 |
| 7 | 2023-04-03 03:06:57.764 | 52.7772 | 158.4839 | 105.631 | 6.5 | mww | 104 | 30 |
| 8 | 2023-03-30 17:33:08.544 | -35.6663 | -73.4965 | 26 | 6.3 | mww | 151 | 52 |
| 9 | 2023-04-04 22:18:12.605 | 7.5834 | -82.3214 | 12.966 | 6.3 | mww | 113 | 63 |
| 10 | 2023-04-19 09:06:05.2 | -5.9994 | 149.6062 | 55.674 | 6.3 | mww | 126 | 33 |

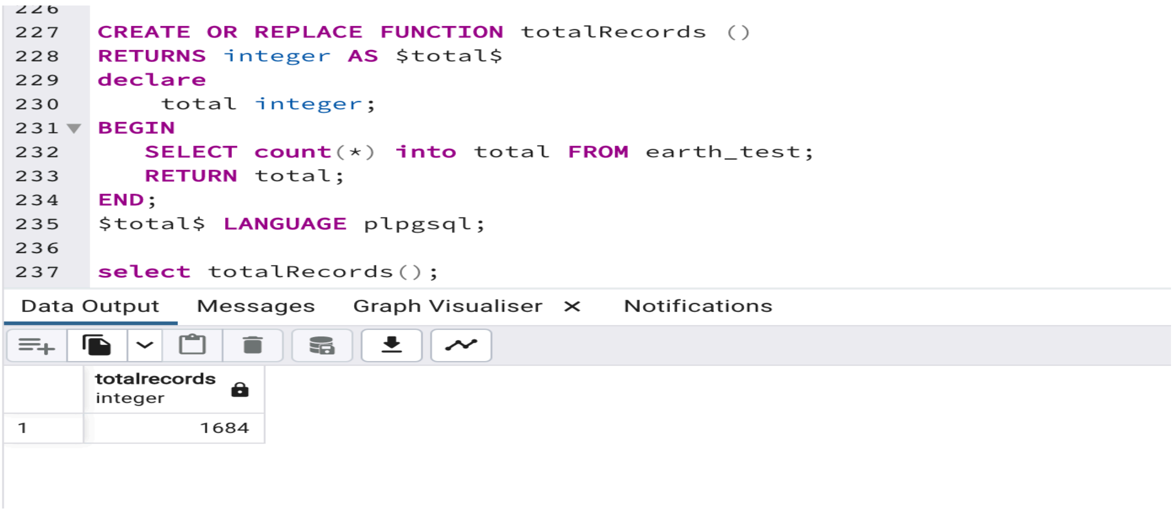
Avg earthquake by type. Here we can see that  avg is 3.7



```
SELECT type, AVG(mag) as average_magnitude FROM earth_test GROUP BY type;
```

| | type<br>character varying (50) | average_magnitude<br>double precision |
|---|---|---|
| 1 | earthquake | 3.74294536814727 |

Create function to get total records



```
CREATE OR REPLACE FUNCTION totalRecords ()
RETURNS integer AS $total$
declare
    total integer;
BEGIN
    SELECT count(*) into total FROM earth_test;
    RETURN total;
END;
$total$ LANGUAGE plpgsql;

select totalRecords();
```

| | totalrecords<br>integer |
|---|---|
| 1 | 1684 |

Conclusion:  observation on this analysis is Indonesia and New Zealand has the highest earthquake magnitude and during the month of April most of the earth occurred.