Hello Professor, I have tried couple of ways to get the desired result but I guess, somewhere, I am still making some mistakes I am getting different results. I look forward to getting your feedback on my weaknesses.

Question 1. The avg and max land temperatures for the 18th centuries

The first try:

9.62s Database default Type text

```
SELECT max(landaveragetemperature) as max_temp, avg(landaveragetemperature) as avg_temp
FROM globaltemperatures
where dt< '1800'
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200704174934_e0262646-9774-4e25-869f-8c57d1158842);
Time taken: 4.917 seconds
INFO  : OK
```

application_1589089424423_0351

Query History | Saved Queries | Query Builder | Results (1)

	max_temp	avg_temp
1	9.996	8.20303231292517

The second attempt

Hive — Add a name... — Add a description...

8.48s Database default Type text

```
SELECT avg(landaveragetemperature) as avg_temp,
     max(landaveragetemperature) as max_temp,
          extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year<'1800'
LIMIT 5
```

```
INFO  : Map 1: 0/1      Reducer 2: 0/1
INFO  : Map 1: 0(+1)/1  Reducer 2: 0/1
INFO  : Map 1: 1/1      Reducer 2: 0/1
```

application_1589089424423_0349

Query History | Saved Queries | Query Builder | Results (5)

	avg_temp	max_temp	dt_year
1	8.719363636363639	8.49	1750
2	7.9761428571428565	7.67	1751
3	5.7798333333333325	8.264999999999999	1752
4	8.388083333333332	8.962	1753
5	8.469333333333335	9.582	1754

The avg and max land temperatures for the 19th centuries

The first try

```
SELECT max(landaveragetemperature) as max_temp, avg(landaveragetemperature) as avg_temp
FROM globaltemperatures
where dt BETWEEN '1800' and '1900'
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200704175909_9237d7fd-f788-46ff-9166-ee1e92b589f4);
Time taken: 10.215 seconds
INFO  : OK
```

application_1589089424423_0352

Query History | Saved Queries | Query Builder | Results (1)

	max_temp	avg_temp
1	9.985	8.009105

The next one

```
SELECT avg(landaveragetemperature) as avg_temp,
       max(landaveragetemperature) as max_temp,
          extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year between '1800' and '1901'
LIMIT 5
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200704172532_fad9e1e8-7d04-4101-af3b-b267306f209b);
Time taken: 4.559 seconds
INFO  : OK
```

application_1589089424423_0350

Query History | Saved Queries | Query Builder | Results (5)

	avg_temp	max_temp	dt_year
1	8.484250000000001	9.469	1800
2	8.589666666666668	9.099	1801
3	8.578166666666666	9.747	1802
4	8.503	9.291	1803
5	8.844833333333334	9.282	1804

The avg and max land temperatures for the 20th centuries

```
SELECT  avg(landaveragetemperature) as avg_temp,
        max(landaveragetemperature) as max_temp,
        extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year between 1901 and 2000;
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200704180752_7f1ua1c8-ecc1-4511-a471-0fbea54bade3);
Time taken: 5.241 seconds
INFO  : OK
```

application_1589089424423_0352

Query History | Saved Queries | Query Builder | Results (100+)

	avg_temp	max_temp	dt_year
1	8.541916666666665	9.003	1901
2	8.304416666666667	8.801	1902
3	8.220166666666668	8.812000000000001	1903

Question 2. The highest temperatures for the same centuries.

It is a matter of terminology (max, largest, highest in some cases) as discussed in the class. And I am assuming that I have already performed these results. However, I am going to add here as well by limiting by 1.

The highest temp in 18 th century.

```
SELECT max(landaveragetemperature) as Max_temp,
       extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year < 1801
ORDER BY Max_temp DESC
LIMIT 1
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200704180407_9eae2528-56ba-4136-a86e-1664cdb27184);
Time taken: 5.187 seconds
INFO  : OK
```

application_1589089424423_0352

Query History | Saved Queries | Query Builder | Results (1)

	max_temp	dt_year
1	9.996	1770

The highest temp in 19 th century.

```
SELECT max(landaveragetemperature) as Max_temp,
       extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year between 1801 and 1900
ORDER BY Max_temp DESC
LIMIT 1
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200704181401_5d7bd41d-f144-4268-a592-b2cd43d2de69);
Time taken: 10.963 seconds
INFO  : OK
```

application_1589089424423_0353

Query History | Saved Queries | Query Builder | Results (1)

	max_temp	dt_year
1	9.985	1838

The highest temp in 20 th century.

```
SELECT max(landaveragetemperature) as Max_temp,
       extract(year from dt) as dt_year
FROM globaltemperatures
GROUP BY extract(year from dt)
HAVING dt_year between 1901 and 2000
ORDER BY Max_temp DESC
LIMIT 1
```

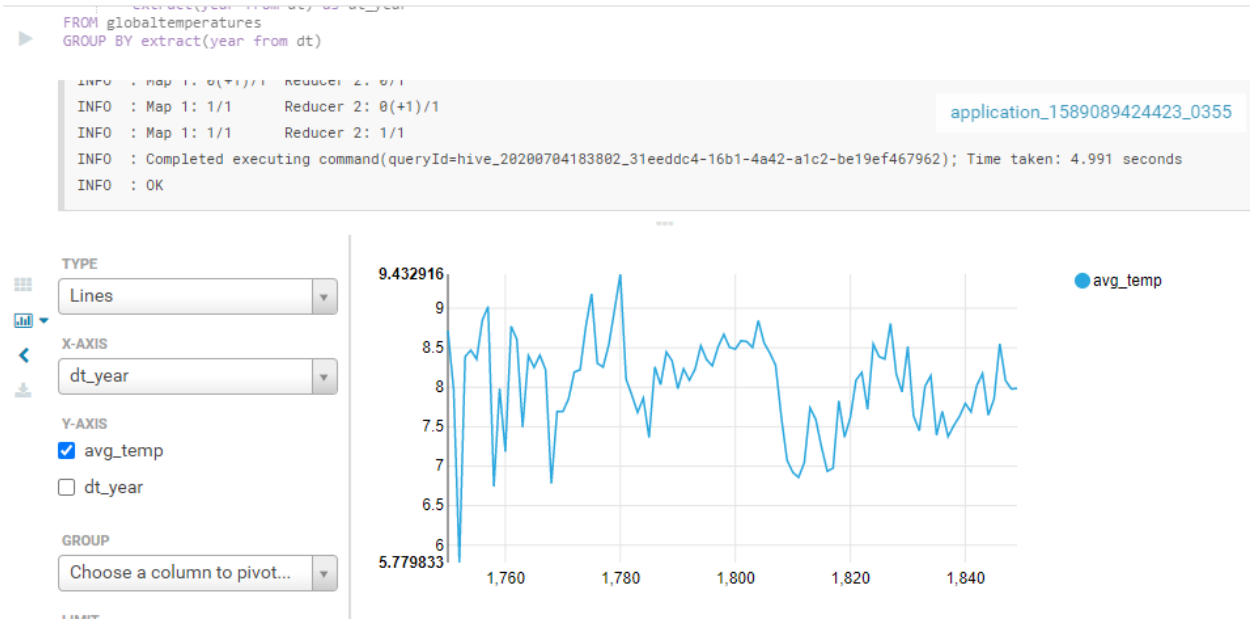
```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200704181554_be2d180a-7d78-41ee-88cb-f342cf159e43);
Time taken: 4.949 seconds
INFO  : OK
```

application_1589089424423_0353

Query History | Saved Queries | Query Builder | Results (1)

	max_temp	dt_year
1	9.997	1999

Question 3. The overall trend in the avg land temperatures?



I was expecting an upward trend in the line graph but it seems there is a fluctuation.
One thing, I am wondering about this graph is that did I include all the data from 18 th to 20 th century? In the graph, there is only data till 1850.

I would need your support with this platform, I am really new to it. Can you please share the recorded video lesson where you have explained this particular query language? .