NAME : Chanda Dunani

SUBJECT : Introduction to Data Science

15/02/2022

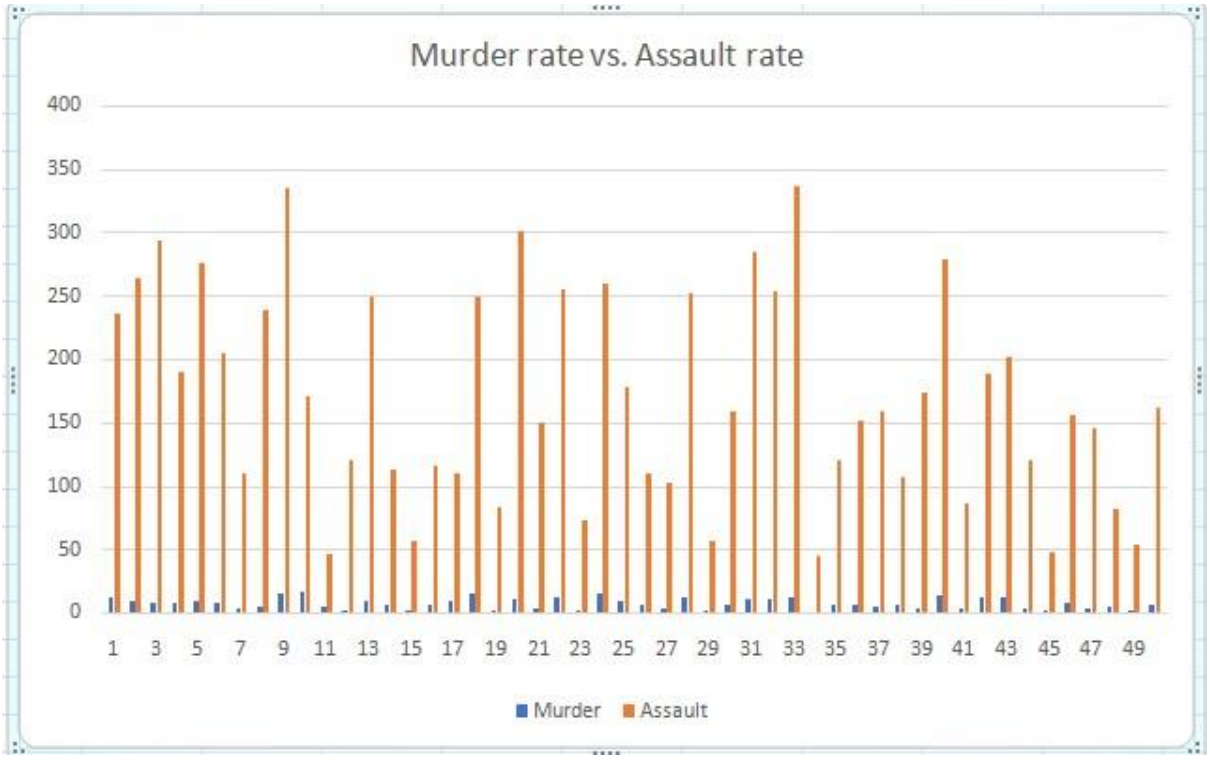
https://github.com/chandadunani/ids7

Problem 1

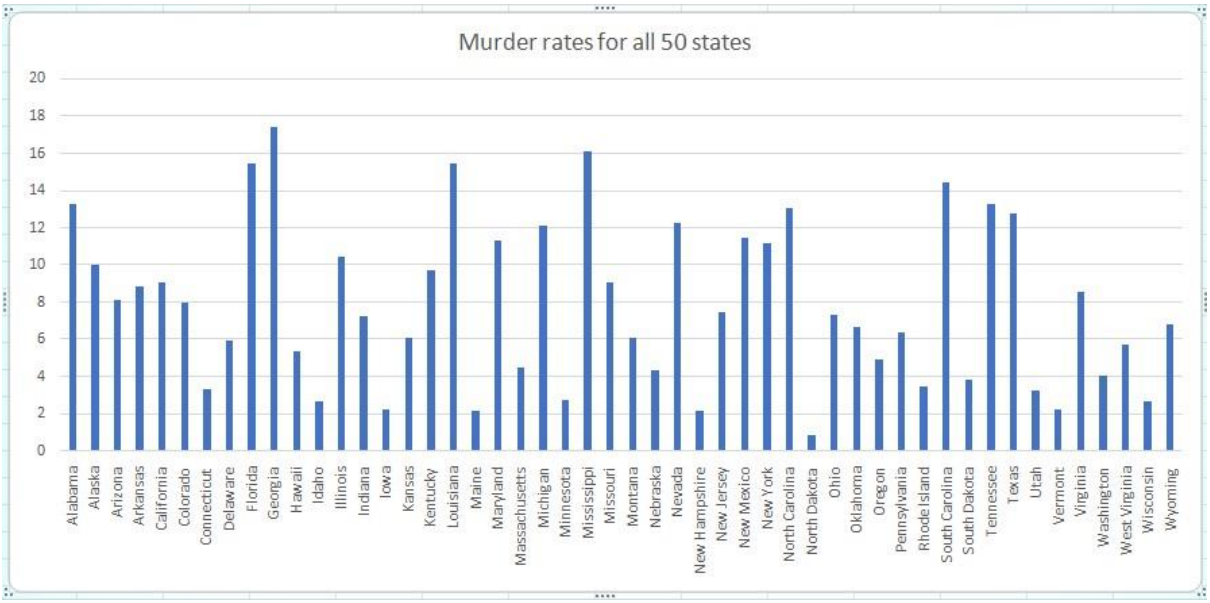
- **Excel Work**
- **Prepare the dataset to establish a relation between an urban population category and a crime type.**

State	Murder	Assault	UrbanPop	Urban_pop_cat	crime_type
Alabama	13.2	236	58	medium	Very High
Alaska	10	263	48	small	Very High
Arizona	8.1	294	80	Extra large	Very High
Arkansas	8.8	190	50	medium	High
California	9	276	91	Extra large	Very High
Colorado	7.9	204	78	Extra large	Very High
Connecticut	3.3	110	77	Extra large	less
Delaware	5.9	238	72	Extra large	Very High
Florida	15.4	335	80	Extra large	Very High
Georgia	17.4	169.94	60	large	Very High
Hawaii	5.3	46	83	Extra large	less
Idaho	2.6	120	54	medium	less
Illinois	10.4	249	83	Extra large	Very High
Indiana	7.2	113	65	large	less
Iowa	2.2	56	57	medium	least
Kansas	6	115	66	large	less
Kentucky	9.7	109	52	medium	less
Louisiana	15.4	249	66	large	Very High
Maine	2.1	83	51	medium	least
Maryland	11.3	300	67	large	Very High
Massachusetts	4.4	149	85	Extra large	less
Michigan	12.1	255	74	Extra large	Very High
Minnesota	2.7	72	66	large	least
Mississippi	16.1	259	44	small	Very High
Missouri	9	178	70	Extra large	High
Montana	6	109	53	medium	less
Nebraska	4.3	102	62	large	less
Nevada	12.2	252	81	Extra large	Very High
New Hampshire	2.1	57	56	medium	least
New Jersey	7.4	159	89	Extra large	High
New Mexico	11.4	285	70	Extra large	Very High
New York	11.1	254	86	Extra large	Very High
North Carolina	13	337	45	small	Very High
North Dakota	0.8	45	44	small	least
Ohio	7.3	120	75	Extra large	less
Oklahoma	6.6	151	68	large	High
Oregon	4.9	159	67	large	High
Pennsylvania	6.3	106	72	Extra large	less
Rhode Island	3.4	174	87	Extra large	High
South Carolina	14.4	279	48	small	Very High
South Dakota	3.8	86	45	small	least
Tennessee	13.2	188	59	medium	High
Texas	12.7	201	80	Extra large	Very High
Utah	3.2	120	80	Extra large	less
Vermont	2.2	48	32	small	least
Virginia	8.5	156	63	large	High
Washington	4	145	73	Extra large	less
West Virginia	5.7	81	39	small	less
Wisconsin	2.6	53	66	large	least
Wyoming	6.8	161	60	large	High

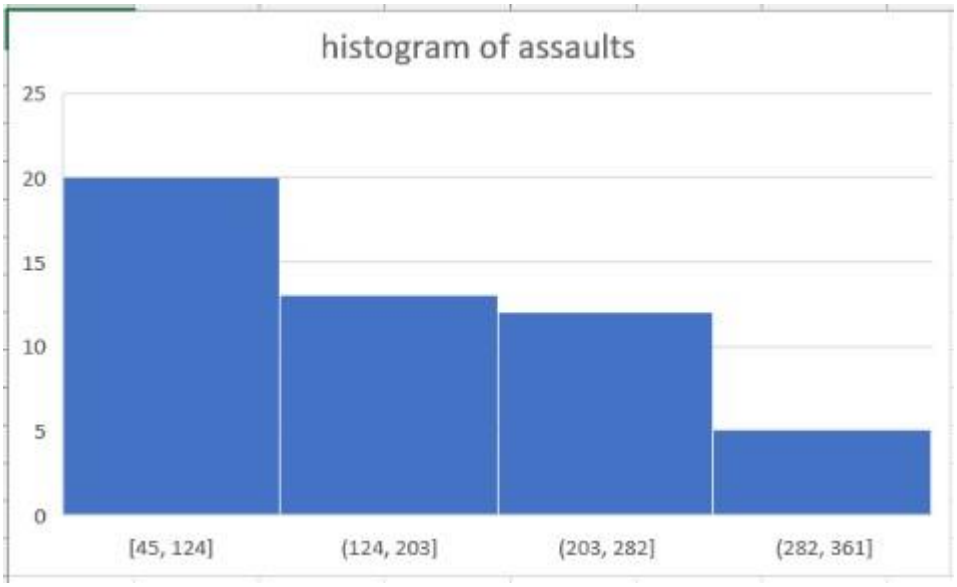
✞ Plot Murder rate vs. Assault rate.



✞ **Plot Murder rates for all 50 states.**



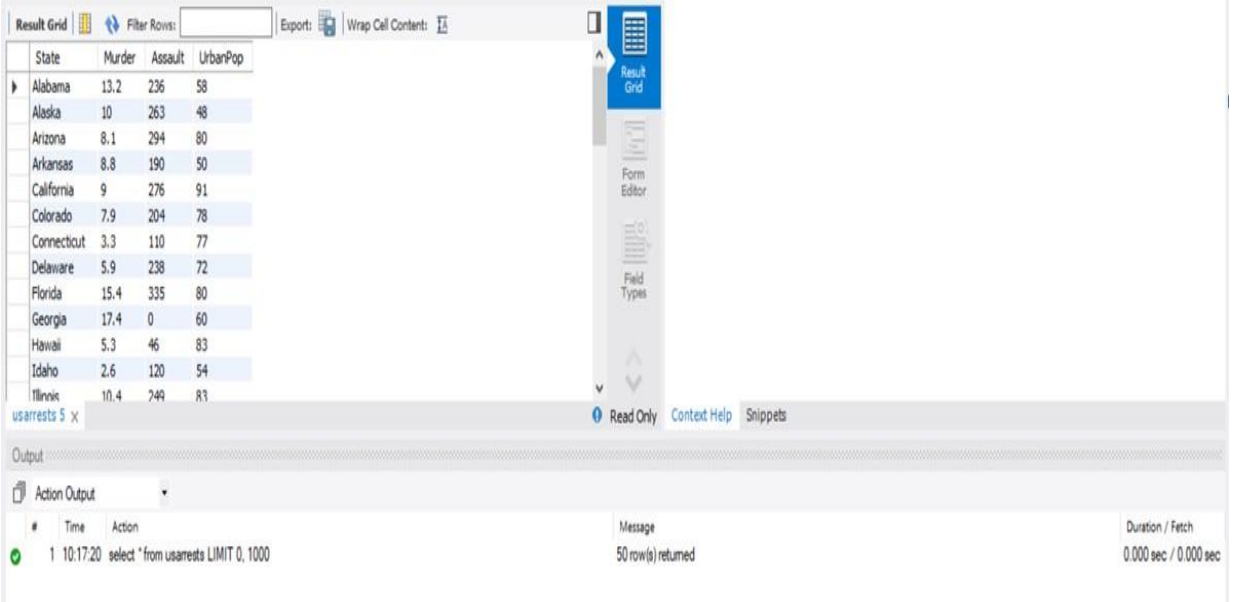
✞ **Plot histogram of assaults.**



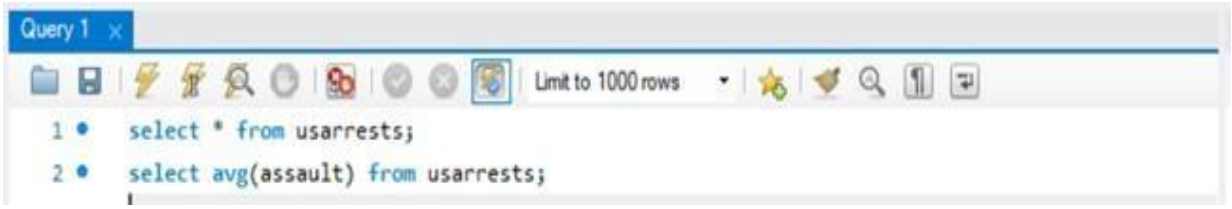
MySQL Work

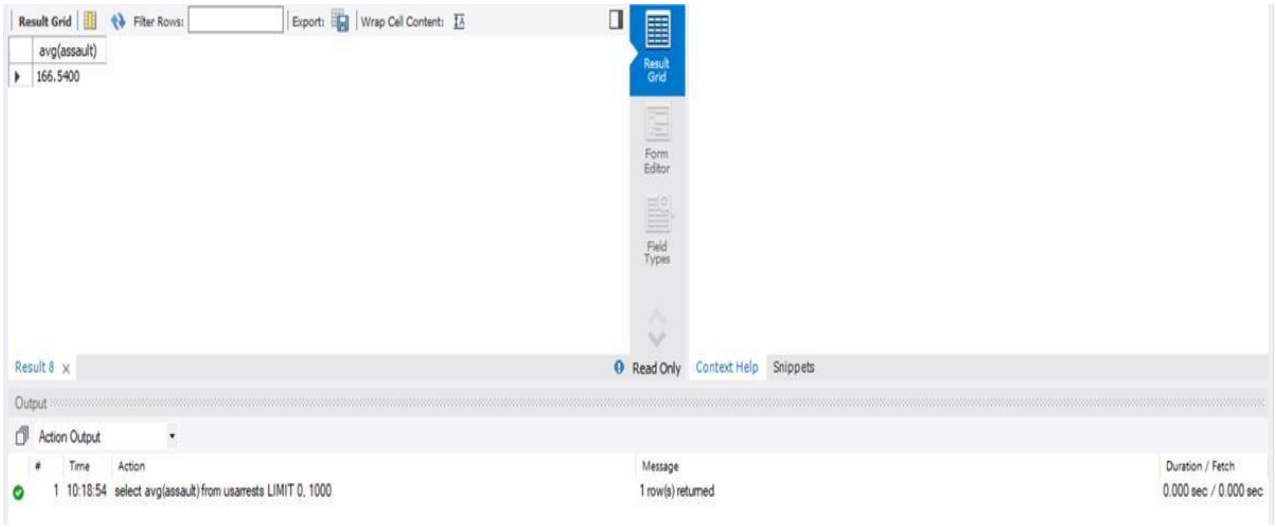
Data Discovery

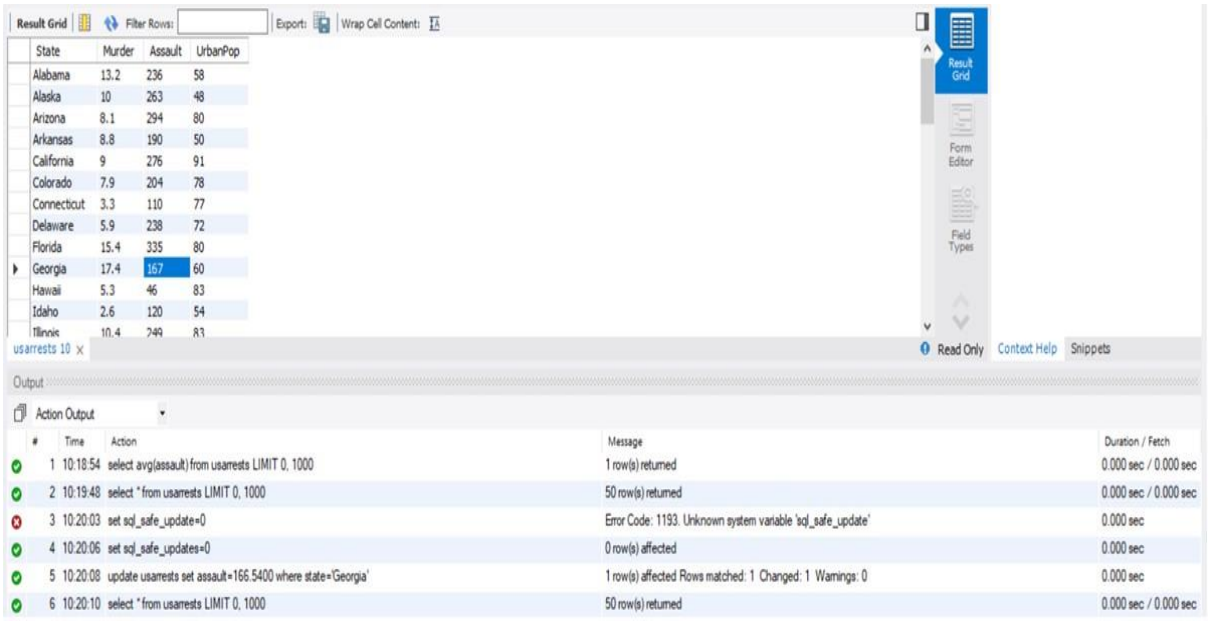
In this part, we discover the variable types and their summary statistics in the data. First, we upload the USArrests dataset.



✞ Detect Missing values







State	Murder	Assault	UrbanPop
Alabama	13.2	236	58
Alaska	10	263	48
Arizona	8.1	294	80
Arkansas	8.8	190	50
California	9	276	91
Colorado	7.9	204	78
Connecticut	3.3	110	77
Delaware	5.9	238	72
Florida	15.4	335	80
Georgia	17.4	167	60
Hawaii	5.3	46	83
Idaho	2.6	120	54
Illinois	10.4	249	83

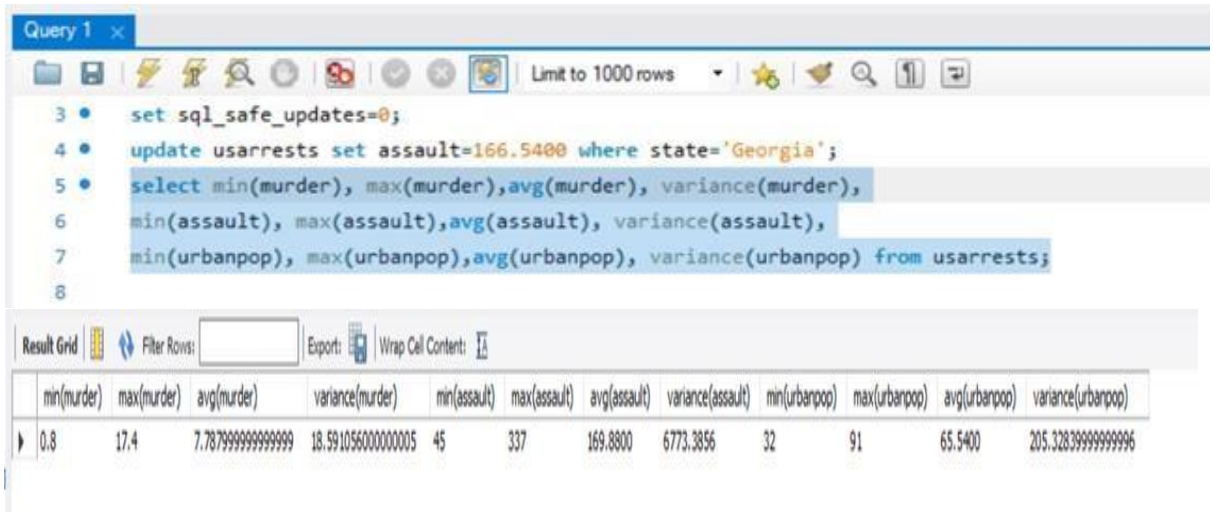
Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Result Grid | Form Editor | Field Types | usarrests 10 | Read Only | Context Help | Snippets

Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	10:18:54	select avg(assault) from usarrests LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
2	10:19:48	select * from usarrests LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec
3	10:20:03	set sql_safe_update=0	Error Code: 1193. Unknown system variable 'sql_safe_update'	0.000 sec
4	10:20:06	set sql_safe_updates=0	0 row(s) affected	0.000 sec
5	10:20:08	update usarrests set assault=166.5400 where state='Georgia'	1 row(s) affected Rows matched: 1 Changed: 1 Warnings: 0	0.000 sec
6	10:20:10	select * from usarrests LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec

✞ Find min, max, mean, and variance of all numeric attributes in SQL



Query 1

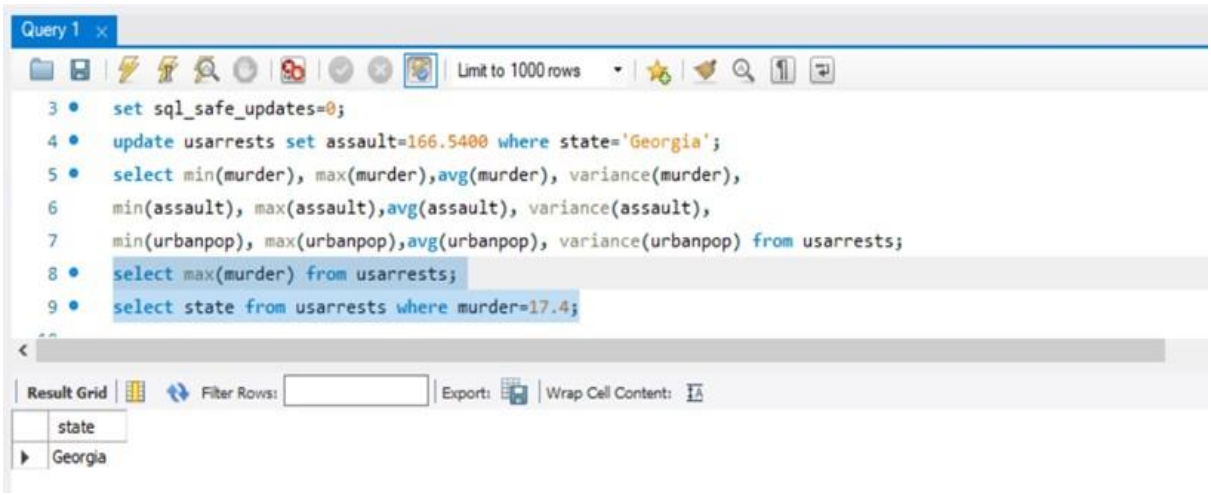
Limit to 1000 rows

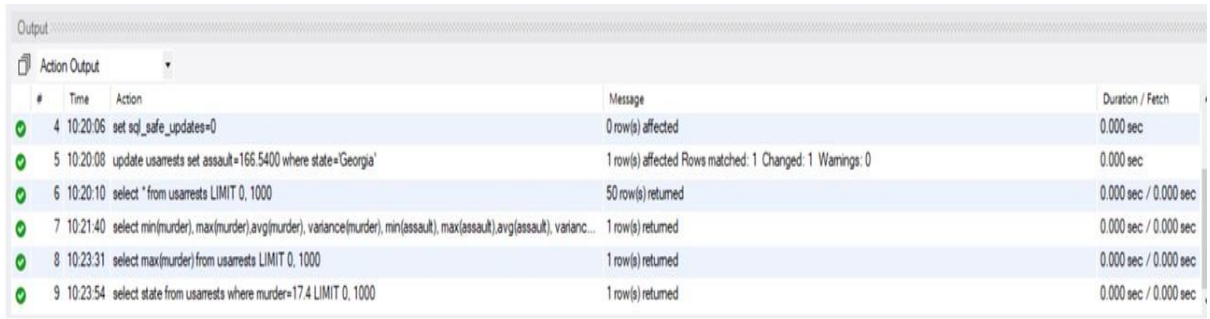
```
3  set sql_safe_updates=0;
4  update usarrests set assault=166.5400 where state='Georgia';
5  select min(murder), max(murder),avg(murder), variance(murder),
6  min(assault), max(assault),avg(assault), variance(assault),
7  min(urbanpop), max(urbanpop),avg(urbanpop), variance(urbanpop) from usarrests;
8
```

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

min(murder)	max(murder)	avg(murder)	variance(murder)	min(assault)	max(assault)	avg(assault)	variance(assault)	min(urbanpop)	max(urbanpop)	avg(urbanpop)	variance(urbanpop)
0.8	17.4	7.787999999999999	18.591056000000005	45	337	169.8800	6773.3856	32	91	65.5400	205.32839999999996

✞ Which state has the maximum murder rate?





✞ List of states in ascending order of urban population percentages.

Query 1

```
update usarrests set assault=166.5400 where state='Georgia';
select min(murder), max(murder),avg(murder), variance(murder),
min(assault), max(assault),avg(assault), variance(assault),
min(urbanpop), max(urbanpop),avg(urbanpop), variance(urbanpop) from usarrests;
select max(murder) from usarrests;
select state from usarrests where murder=17.4;
select state from usarrests order by urbanpop;
```

Result Grid

state
Vermont
West Virginia
North Dakota
Mississippi
North Carolina
South Dakota
Alaska
South Carolina
Arkansas
Maine
Kentucky
Montana
Idaho

Output

Action Output

#	Time	Action	Message	Duration / Fetch
6	10:20:10	select * from usarrests LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec
7	10:21:40	select min(murder), max(murder),avg(murder), variance(murder), min(assault), max(assault),avg(assault), varianc...	1 row(s) returned	0.000 sec / 0.000 sec
8	10:23:31	select max(murder) from usarrests LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
9	10:23:54	select state from usarrests where murder=17.4 LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
10	10:25:37	select state from usarrests order by urban_pop LIMIT 0, 1000	Error Code: 1054. Unknown column 'urban_pop' in 'order clause'	0.000 sec
11	10:25:47	select state from usarrests order by urbanpop LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec

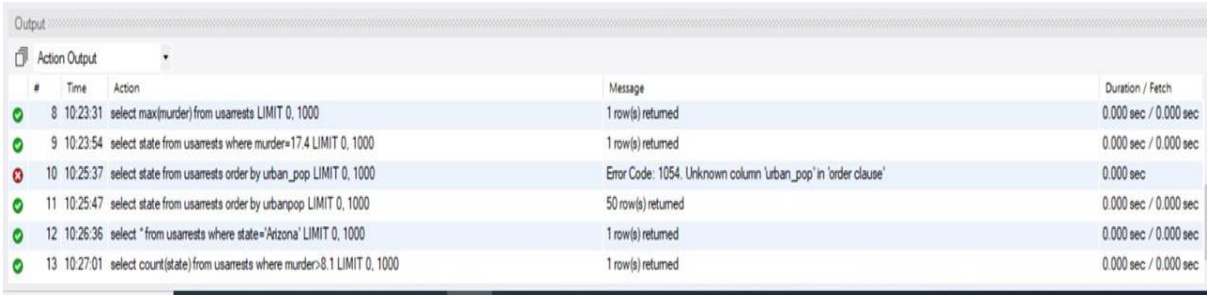
✞ **How many states have higher murder rates than Arizona? List those states.**

Query 1

```
min(assault), max(assault),avg(assault), variance(assault),
min(urbanpop), max(urbanpop),avg(urbanpop), variance(urbanpop) from usarrests;
select max(murder) from usarrests;
select state from usarrests where murder=17.4;
select state from usarrests order by urbanpop;
select * from usarrests where state='Arizona';
select count(state) from usarrests where murder>8.1;
```

Result Grid

count(state)
21

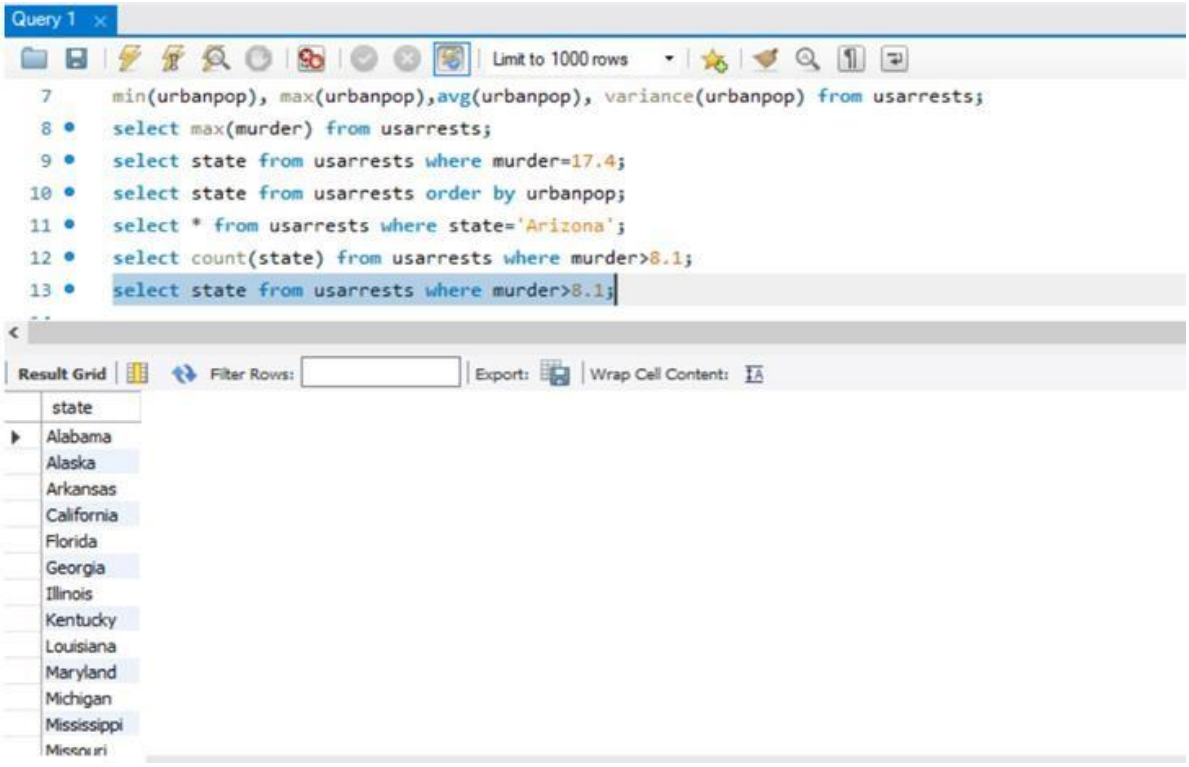


Output

Action Output

#	Time	Action	Message	Duration / Fetch
8	10:23:31	select max(murder) from usarrests LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
9	10:23:54	select state from usarrests where murder=17.4 LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
10	10:25:37	select state from usarrests order by urban_pop LIMIT 0, 1000	Error Code: 1054. Unknown column 'urban_pop' in 'order clause'	0.000 sec
11	10:25:47	select state from usarrests order by urbanpop LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec
12	10:26:36	select * from usarrests where state='Arizona' LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
13	10:27:01	select count(state) from usarrests where murder>8.1 LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec

List of States:



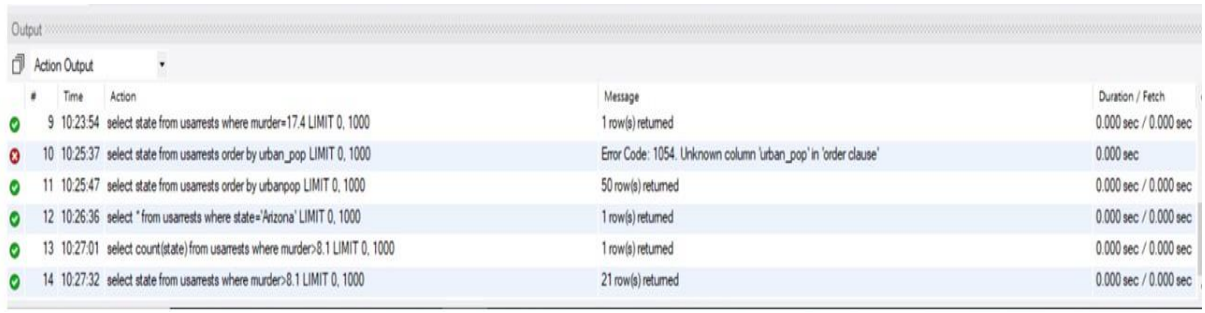
Query 1

Limit to 1000 rows

```
7   min(urbanpop), max(urbanpop),avg(urbanpop), variance(urbanpop) from usarrests;
8   select max(murder) from usarrests;
9   select state from usarrests where murder=17.4;
10  select state from usarrests order by urbanpop;
11  select * from usarrests where state='Arizona';
12  select count(state) from usarrests where murder>8.1;
13  select state from usarrests where murder>8.1;
```

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

state
Alabama
Alaska
Arkansas
California
Florida
Georgia
Illinois
Kentucky
Louisiana
Maryland
Michigan
Mississippi
Missouri



Output

Action Output

#	Time	Action	Message	Duration / Fetch
9	10:23:54	select state from usarrests where murder=17.4 LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
10	10:25:37	select state from usarrests order by urban_pop LIMIT 0, 1000	Error Code: 1054. Unknown column 'urban_pop' in 'order clause'	0.000 sec
11	10:25:47	select state from usarrests order by urbanpop LIMIT 0, 1000	50 row(s) returned	0.000 sec / 0.000 sec
12	10:26:36	select * from usarrests where state='Arizona' LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
13	10:27:01	select count(state) from usarrests where murder>8.1 LIMIT 0, 1000	1 row(s) returned	0.000 sec / 0.000 sec
14	10:27:32	select state from usarrests where murder>8.1 LIMIT 0, 1000	21 row(s) returned	0.000 sec / 0.000 sec

Conclusion:

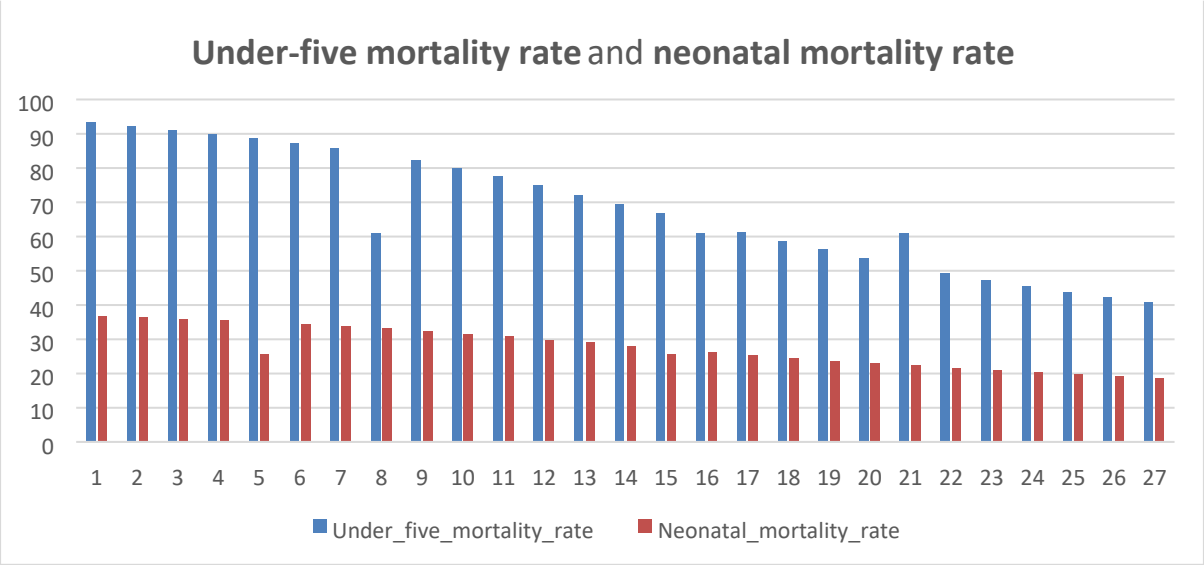
To conclude, in this article, we examine the explanatory data analysis and which visualization types we can use for the explanatory data analysis. As stated above it is a very crucial step and it must be done before future engineering and model building to better understand data. This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.This is a systematic approach for identifying and analyzing patterns and trends in crime using USArrest dataset.

https://github.com/chandadunani/ids7

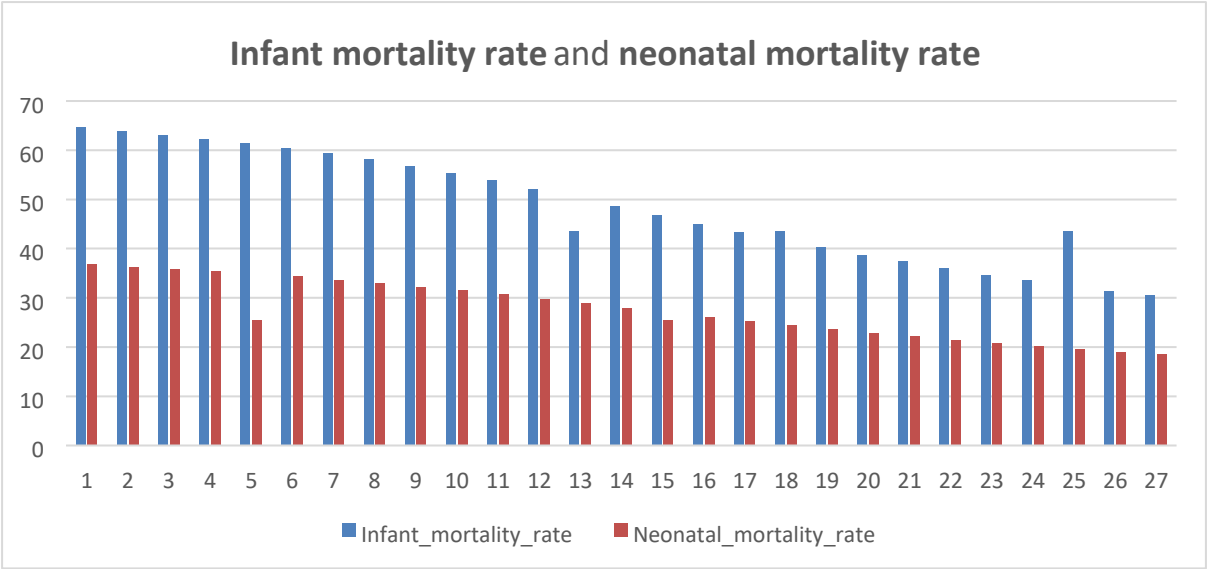
Problem 2

Excel Work

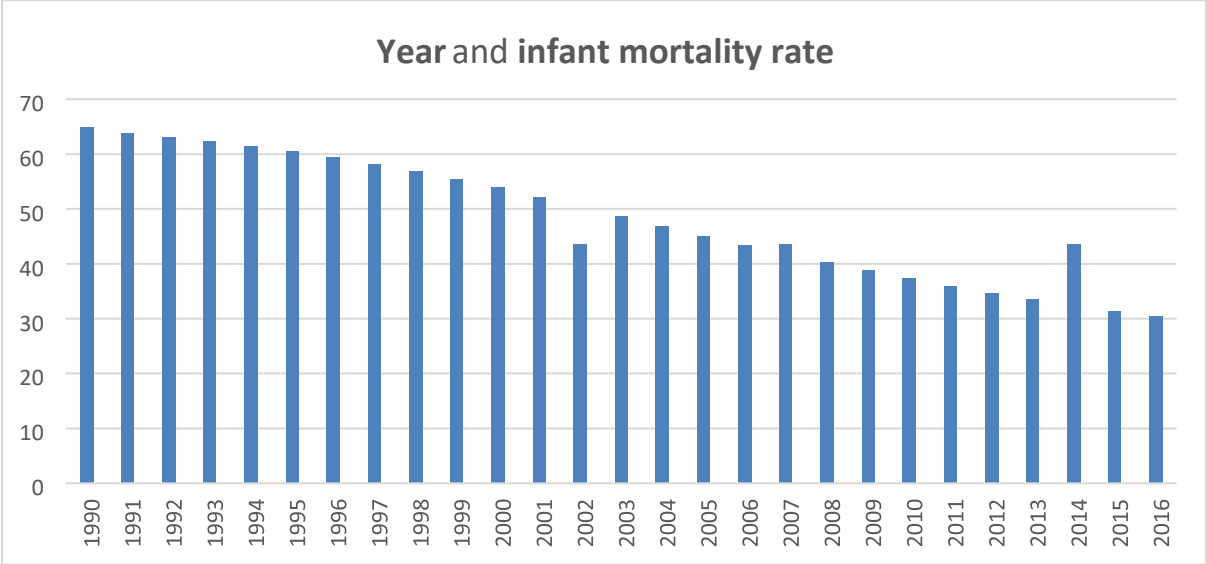
✞ Under-Five mortality rate and neonatal mortality rate.



✞ Infant mortality rate and neonatal mortality rate.



- **Year and infant mortality rate.**



MySQL Work

- **Display the entire table.**

Query 1

```
select * from child_mortality;
```

Result Grid

Year	Under_five_mortality_rate	Infant_mortality_rate	Neonatal_mortality_rate
1990	93.4	64.8	36.8
1991	92.1	63.9	36.3
1992	90.9	63.1	35.9
1993	89.7	62.3	35.4
1994	88.7	61.4	0
1995	87.3	60.5	34.4
1996	85.6	59.4	33.7
1997	0	58.2	33.1
1998	82.1	56.9	32.3
1999	79.9	55.4	31.5
2000	77.5	53.9	30.7
2001	74.8	52.1	29.8
2002	72	0	28.9

Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	10:41:00	select * from child_mortality LIMIT 0, 1000	27 row(s) returned	0.000 sec / 0.000 sec

Query 1

```
select * from child_mortality;
select avg(under_five_mortality_rate),avg(infant_mortality_rate),avg(neonatal_mortality_rate) from child_mortality;
```

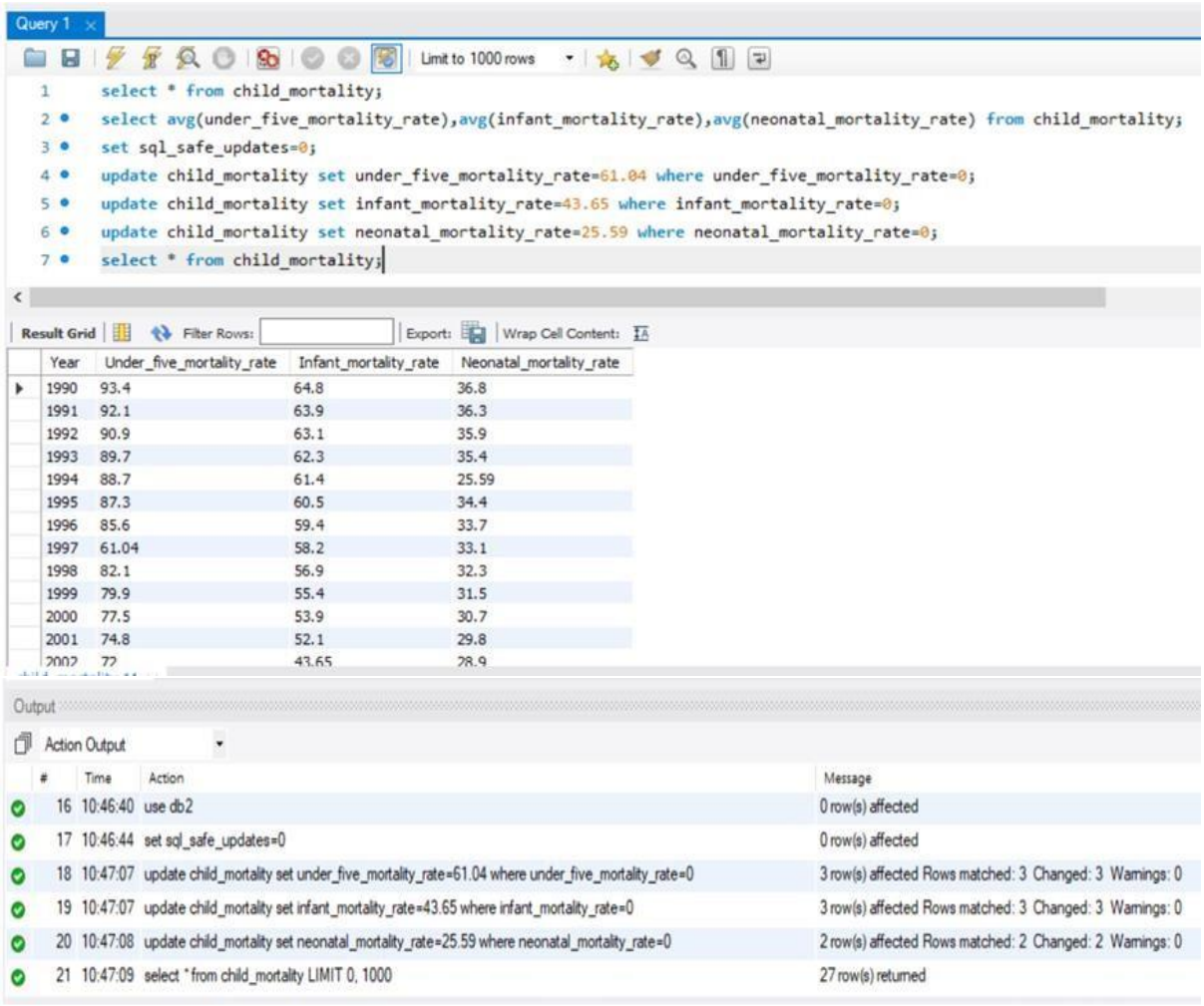
Result Grid

avg(under_five_mortality_rate)	avg(infant_mortality_rate)	avg(neonatal_mortality_rate)
61.0444444444444	43.65185185185184	25.5962962962963

Output

Action Output

#	Time	Action	Message
1	10:44:29	select avg(under_five_mortality_rate),avg(infant_mortality_rate),avg(neonatal_mortality_rate) from child_mortality L...	1 row(s) returned



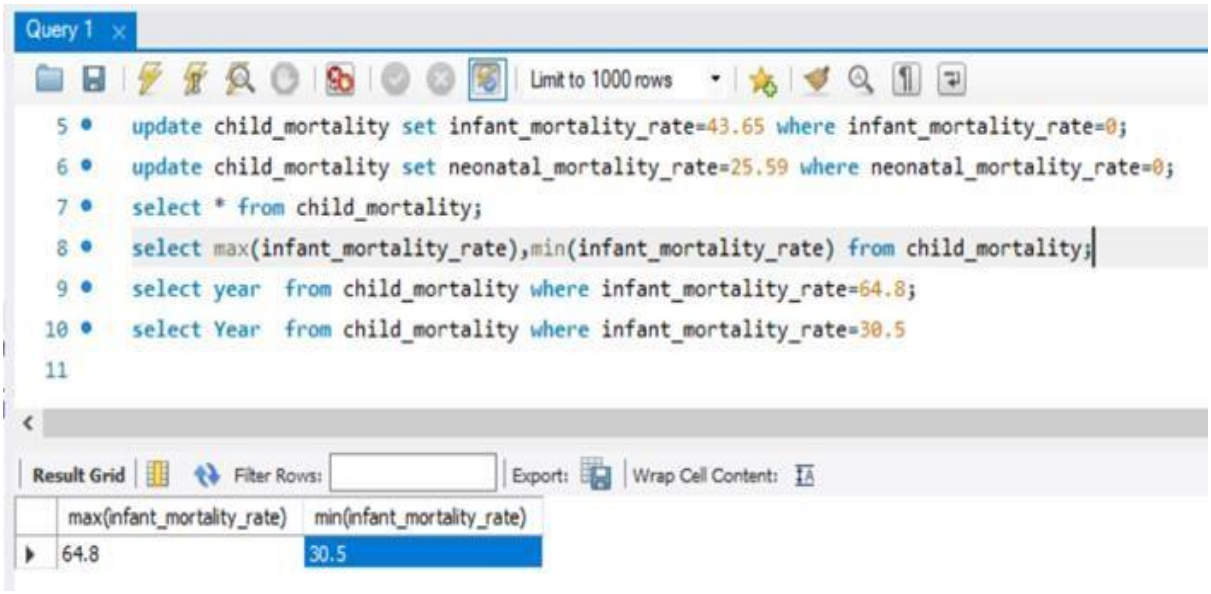
```
select * from child_mortality;
select avg(under_five_mortality_rate),avg(infant_mortality_rate),avg(neonatal_mortality_rate) from child_mortality;
set sql_safe_updates=0;
update child_mortality set under_five_mortality_rate=61.04 where under_five_mortality_rate=0;
update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0;
update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0;
select * from child_mortality;
```

Year	Under_five_mortality_rate	Infant_mortality_rate	Neonatal_mortality_rate
1990	93.4	64.8	36.8
1991	92.1	63.9	36.3
1992	90.9	63.1	35.9
1993	89.7	62.3	35.4
1994	88.7	61.4	25.59
1995	87.3	60.5	34.4
1996	85.6	59.4	33.7
1997	61.04	58.2	33.1
1998	82.1	56.9	32.3
1999	79.9	55.4	31.5
2000	77.5	53.9	30.7
2001	74.8	52.1	29.8
2002	72	43.65	28.9

Output

#	Time	Action	Message
16	10:46:40	use db2	0 row(s) affected
17	10:46:44	set sql_safe_updates=0	0 row(s) affected
18	10:47:07	update child_mortality set under_five_mortality_rate=61.04 where under_five_mortality_rate=0	3 row(s) affected Rows matched: 3 Changed: 3 Warnings: 0
19	10:47:07	update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0	3 row(s) affected Rows matched: 3 Changed: 3 Warnings: 0
20	10:47:08	update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0	2 row(s) affected Rows matched: 2 Changed: 2 Warnings: 0
21	10:47:09	select * from child_mortality LIMIT 0, 1000	27 row(s) returned

✞ **Which years have the lowest and highest infant mortality years, respectively?**



```
update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0;
update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0;
select * from child_mortality;
select max(infant_mortality_rate),min(infant_mortality_rate) from child_mortality;
select year  from child_mortality where infant_mortality_rate=64.8;
select Year  from child_mortality where infant_mortality_rate=30.5
```

max(infant_mortality_rate)	min(infant_mortality_rate)
64.8	30.5

Output

Action Output

#	Time	Action	Message
19	10:47:07	update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0	3 row(s) affected Rows matched: 3 Changed: 3 Warnings: 0
20	10:47:08	update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0	2 row(s) affected Rows matched: 2 Changed: 2 Warnings: 0
21	10:47:09	select * from child_mortality LIMIT 0, 1000	27 row(s) returned
22	10:48:46	select year from child_mortality where infant_mortality_rate=max(infant_mortality_rate) LIMIT 0, 1000	Error Code: 1111. Invalid use of group function
23	10:49:24	select * from child_mortality where infant_mortality_rate=max(infant_mortality_rate) LIMIT 0, 1000	Error Code: 1111. Invalid use of group function
24	10:51:27	select max(infant_mortality_rate),min(infant_mortality_rate) from child_mortality LIMIT 0, 1000	1 row(s) returned

Query 1

Limit to 1000 rows

```
3  set sql_safe_updates=0;
4  update child_mortality set under_five_mortality_rate=61.04 where under_five_mortality_rate=0;
5  update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0;
6  update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0;
7  select * from child_mortality;
8  select max(infant_mortality_rate),min(infant_mortality_rate) from child_mortality;
9  select year  from child_mortality where infant_mortality_rate=64.8; -- max
```

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

year
1990

Output

Action Output

#	Time	Action	Message
21	10:47:09	select * from child_mortality LIMIT 0, 1000	27 row(s) returned
22	10:48:46	select year from child_mortality where infant_mortality_rate=max(infant_mortality_rate) LIMIT 0, 1000	Error Code: 1111. Invalid use of group function
23	10:49:24	select * from child_mortality where infant_mortality_rate=max(infant_mortality_rate) LIMIT 0, 1000	Error Code: 1111. Invalid use of group function
24	10:51:27	select max(infant_mortality_rate),min(infant_mortality_rate) from child_mortality LIMIT 0, 1000	1 row(s) returned
25	10:52:10	select year from child_mortality where infant_mortality_rate=64.8 LIMIT 0, 1000	1 row(s) returned
26	10:52:38	select year from child_mortality where infant_mortality_rate=64.8 LIMIT 0, 1000	1 row(s) returned

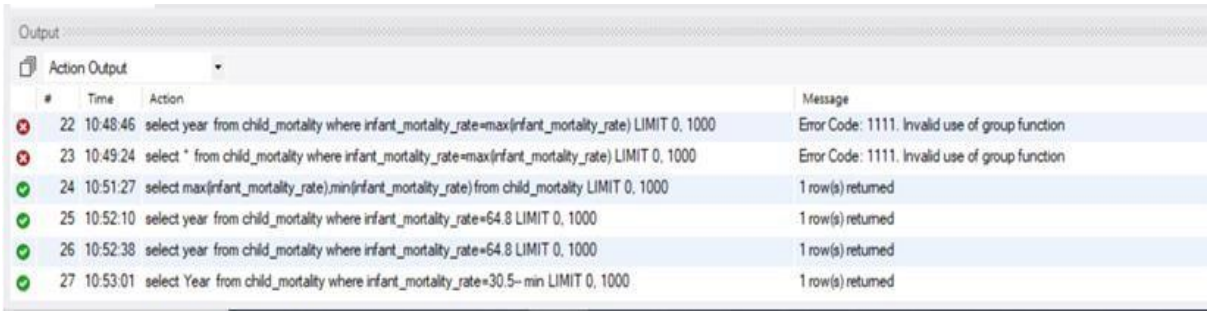
Query 1

Limit to 1000 rows

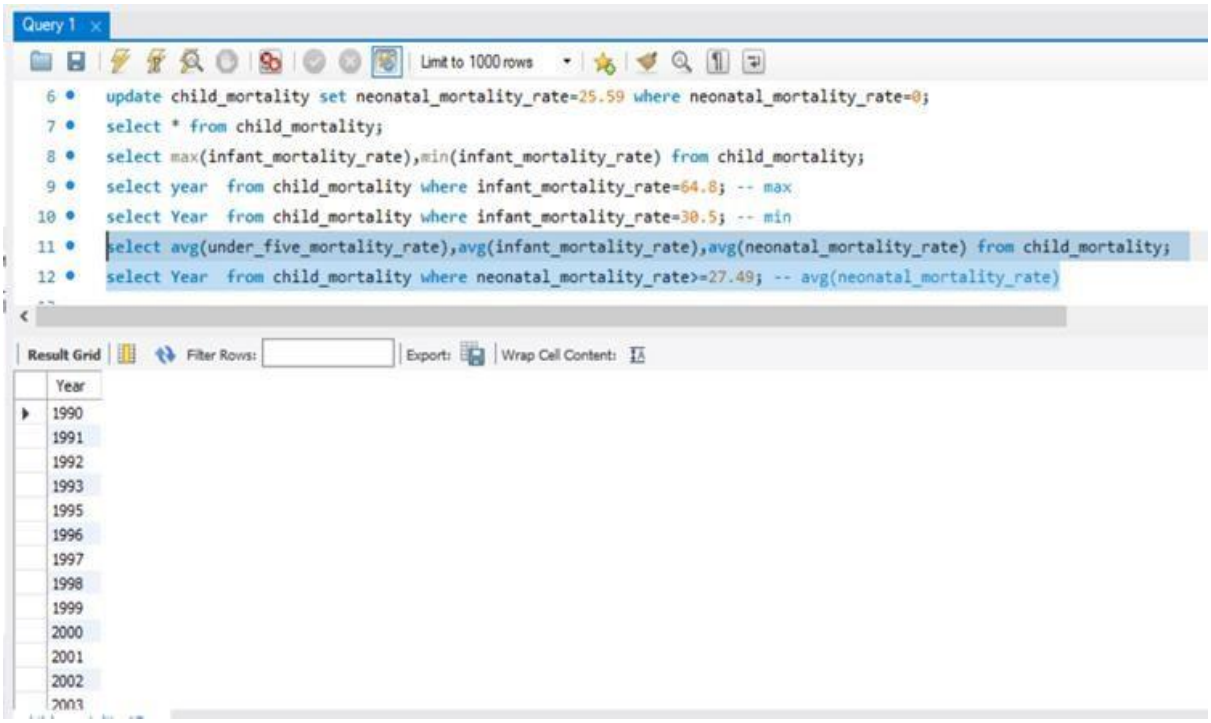
```
5  update child_mortality set infant_mortality_rate=43.65 where infant_mortality_rate=0;
6  update child_mortality set neonatal_mortality_rate=25.59 where neonatal_mortality_rate=0;
7  select * from child_mortality;
8  select max(infant_mortality_rate),min(infant_mortality_rate) from child_mortality;
9  select year  from child_mortality where infant_mortality_rate=64.8; -- max
10 select Year  from child_mortality where infant_mortality_rate=30.5  -- min
11
```

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

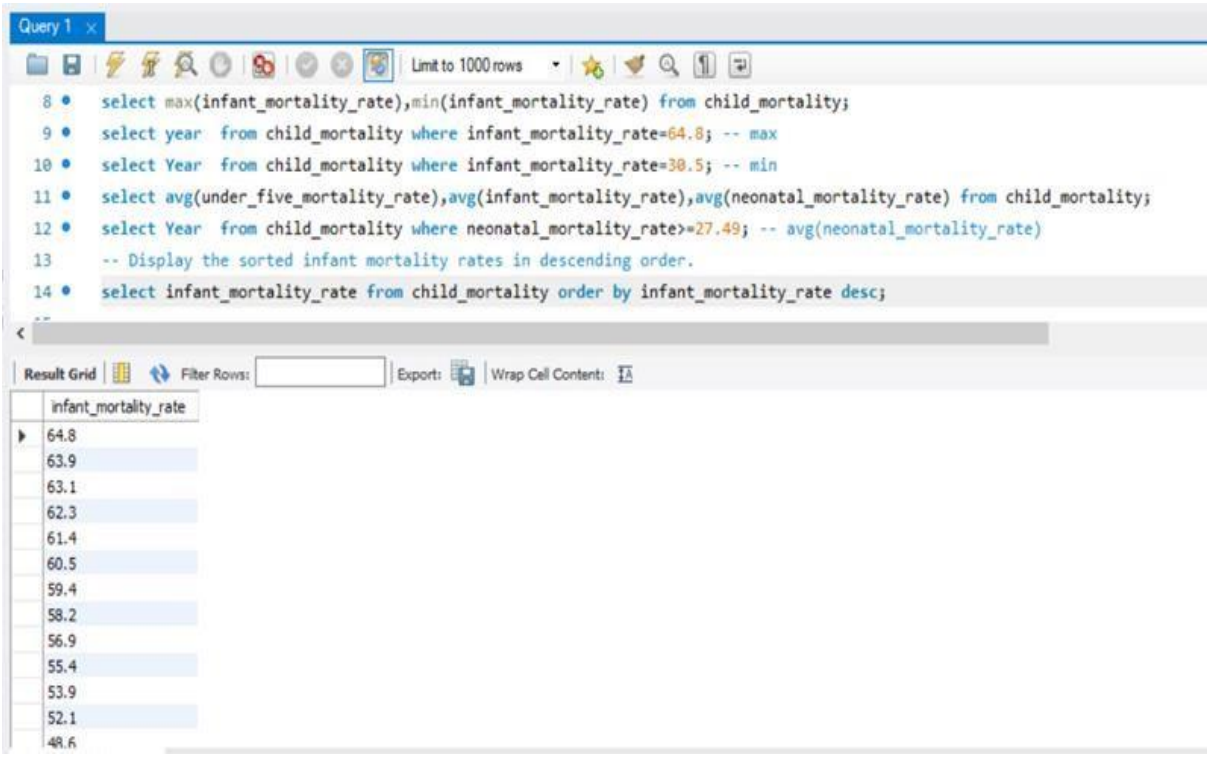
Year
2016

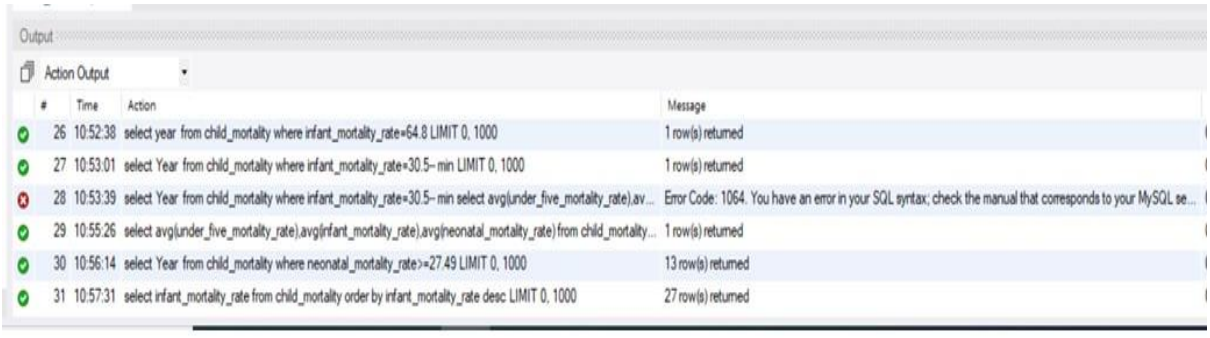


✞ In what years the neonatal mortality rates were above average?

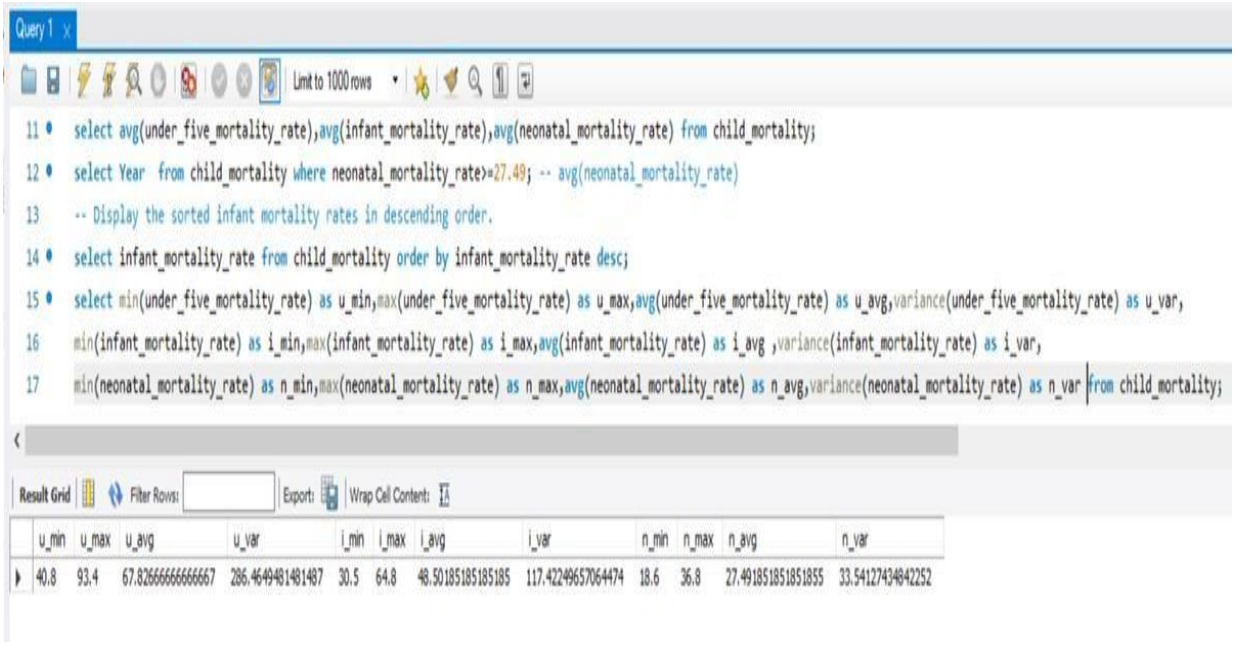


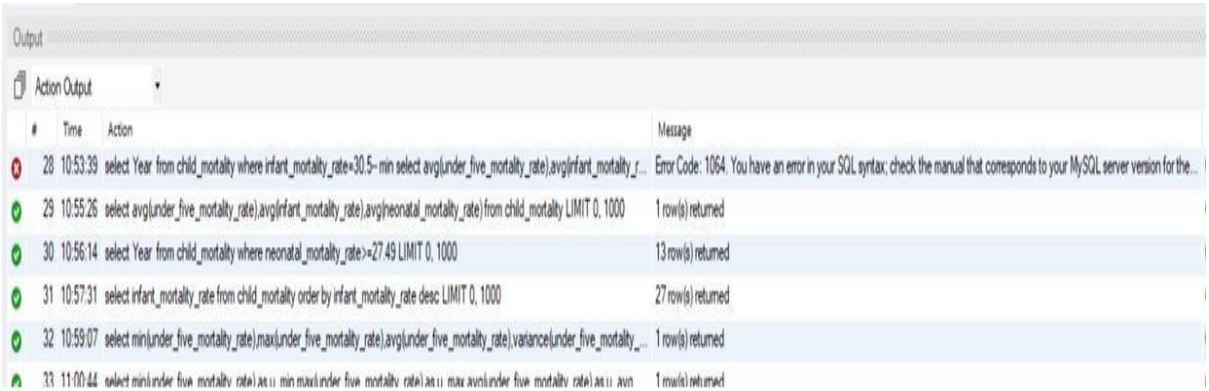
✞ **Display the sorted infant mortality rates in descending order.**





- **Display min, max, mean, variance, and standard deviation for each mortality rate.**





- **Add a new column called above-five mortality rate and populate it with appropriate values. Hint: use alter table add column.**

```
Query 1
Limit to 1000 rows
14  select infant_mortality_rate from child_mortality order by infant_mortality_rate desc;
15  select min(under_five_mortality_rate) as u_min,max(under_five_mortality_rate) as u_max,avg(under_five_mortality_rate) as u_avg,variance(under_five_mortality_rate) as u_var,
16  min(infant_mortality_rate) as i_min,max(infant_mortality_rate) as i_max,avg(infant_mortality_rate) as i_avg ,variance(infant_mortality_rate) as i_var,
17  min(neonatal_mortality_rate) as n_min,max(neonatal_mortality_rate) as n_max,avg(neonatal_mortality_rate) as n_avg,variance(neonatal_mortality_rate) as n_var from child_mortality;
18  desc child_mortality;
19  alter table child_mortality add column Above_Five_Mortality_Rate double;
20  desc child_mortality;
```

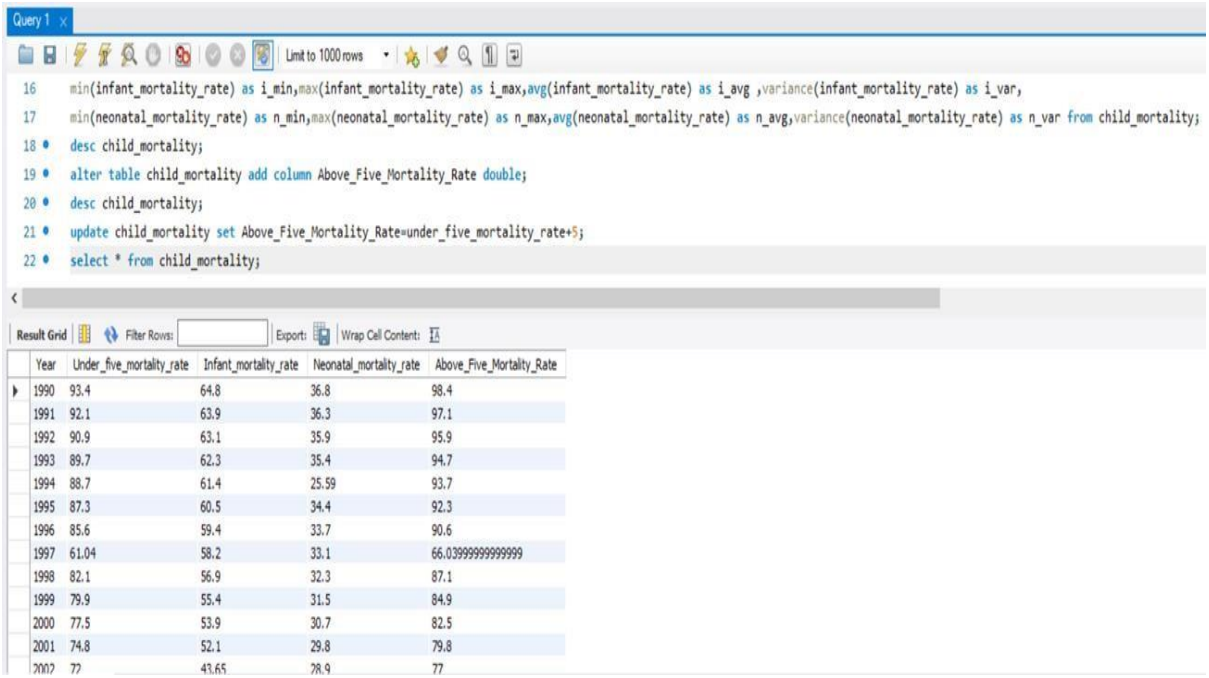
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Field	Type	Null	Key	Default	Extra
Year	int	YES		NULL	
Under_five_mortality_rate	double	YES		NULL	
Infant_mortality_rate	double	YES		NULL	
Neonatal_mortality_rate	double	YES		NULL	
Above_Five_Mortality_Rate	double	YES		NULL	

Output

Action Output

#	Time	Action	Message
31	10:57:31	select infant_mortality_rate from child_mortality order by infant_mortality_rate desc LIMIT 0, 1000	27 row(s) returned
32	10:59:07	select min(under_five_mortality_rate),max(under_five_mortality_rate),avg(under_five_mortality_rate),variance(under_five_mortality_...	1 row(s) returned
33	11:00:44	select min(under_five_mortality_rate) as u_min,max(under_five_mortality_rate) as u_max,avg(under_five_mortality_rate) as u_avg,...	1 row(s) returned
34	11:02:01	desc child_mortality	4 row(s) returned
35	11:02:12	alter table child_mortality add column Above_Five_Mortality_Rate double	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0
36	11:02:18	desc child_mortality	5 row(s) returned



```
Query 1
Limit to 1000 rows
16  min(infant_mortality_rate) as i_min,max(infant_mortality_rate) as i_max,avg(infant_mortality_rate) as i_avg ,variance(infant_mortality_rate) as i_var,
17  min(neonatal_mortality_rate) as n_min,max(neonatal_mortality_rate) as n_max,avg(neonatal_mortality_rate) as n_avg,variance(neonatal_mortality_rate) as n_var from child_mortality;
18  desc child_mortality;
19  alter table child_mortality add column Above_Five_Mortality_Rate double;
20  desc child_mortality;
21  update child_mortality set Above_Five_Mortality_Rate=under_five_mortality_rate+5;
22  select * from child_mortality;
```

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

Year	Under_five_mortality_rate	Infant_mortality_rate	Neonatal_mortality_rate	Above_Five_Mortality_Rate
1990	93.4	64.8	36.8	98.4
1991	92.1	63.9	36.3	97.1
1992	90.9	63.1	35.9	95.9
1993	89.7	62.3	35.4	94.7
1994	88.7	61.4	25.59	93.7
1995	87.3	60.5	34.4	92.3
1996	85.6	59.4	33.7	90.6
1997	61.04	58.2	33.1	66.03999999999999
1998	82.1	56.9	32.3	87.1
1999	79.9	55.4	31.5	84.9
2000	77.5	53.9	30.7	82.5
2001	74.8	52.1	29.8	79.8
2002	72	43.65	28.9	77

Output

Action Output

#	Time	Action	Message
33	11:00:44	select min(under_five_mortality_rate) as u_min,max(under_five_mortality_rate) as u_max,avg(under_five_mortality_rate) as u_avg,...	1 row(s) returned
34	11:02:01	desc child_mortality	4 row(s) returned
35	11:02:12	alter table child_mortality add column Above_Five_Mortality_Rate double	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0
36	11:02:18	desc child_mortality	5 row(s) returned
37	11:04:08	update child_mortality set Above_Five_Mortality_Rate=under_five_mortality_rate+5	27 row(s) affected Rows matched: 27 Changed: 27 Warnings: 0
38	11:04:18	select * from child_mortality LIMIT 0, 1000	27 row(s) returned

Conclusion

Childhood is a significant stage of life and deprivation during this period can have a long-term adverse impact on the wellbeing of children. Reduction in infant and child morality is likely the most important of the millennium development goals, as children are the most important assets of a nation. The focus of this paper is to examine the determinants of childhood mortality and child health in India and the factors explaining the differential performance of the child immunization and treatment of childhood diseases. For this purpose data are taken from the three rounds of the National Family Health Survey of India (NFHS) conducted in 1992-93, 199899 and 2005-06.