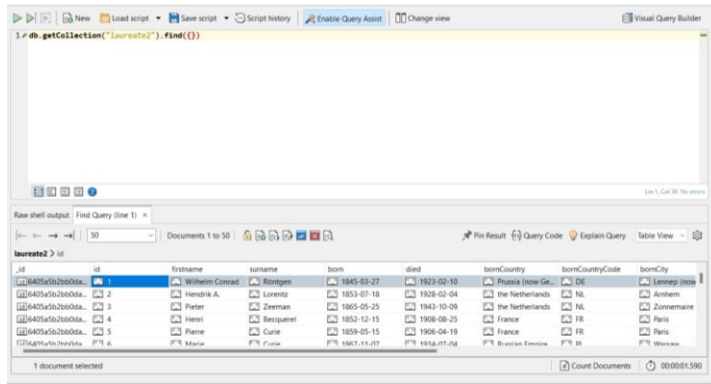


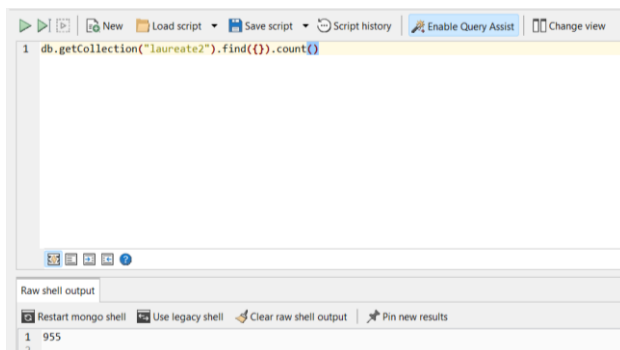
HW6 - NoSQL

Connection successful:



_id	id	firstname	surname	born	died	bornCountry	bornCountryCode	bornCity
6405a5b2b0da...	1	Wilhelm Conrad	Röntgen	1845-02-27	1923-02-19	Prussia (now Ger.)	DE	Leipzig
6405a5b2b0da...	2	Hendrik A.	Loewy	1853-07-18	1928-02-04	the Netherlands	NL	Amsterdam
6405a5b2b0da...	3	Pieter	Zeebman	1865-05-25	1943-10-09	the Netherlands	NL	Zonnebeke
6405a5b2b0da...	4	Henri	Bequerel	1852-12-15	1908-08-25	France	FR	Paris
6405a5b2b0da...	5	Pierre	Curie	1859-05-15	1906-04-19	France	FR	Paris
6405a5b2b0da...	6	Marie	Curie	1867-11-27	1935-07-28	Poland, Russian Empire	PL	Warsaw

1. the count of total number of records in the collection

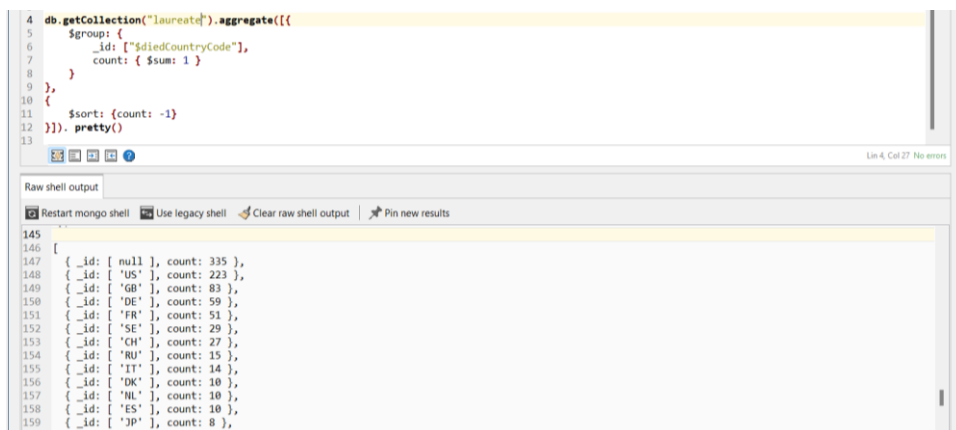


```
1 db.getCollection("laureate2").find({}).count()
```

count
955

There is a total of 955 records in the laureate.json file which the count query returned. This means that a total of 955 Nobel awards have been given since 1906 till 2016.

2. the count of records for each diedCountryCode in descending order of count.



```
4 db.getCollection("laureate2").aggregate([
5   $group: {
6     _id: "$diedCountryCode",
7     count: { $sum: 1 }
8   },
9 ],
10 { $sort: { count: -1 }
11 },
12 ).pretty()
```

_id	count
null	335
US	223
GB	83
DE	59
FR	51
SE	29
CH	27
RU	15
IT	14
DK	10
NL	10
ES	10
JP	8

3. There are 355 null values in the diedCountryCode after which US has been ranked highest in the deadCountryCodes with a count of 223, followed by GB, DE, Fr, and so on. The counts of each

are shown in descending order. The death country of most of the Nobel prize winners is unknown.

- the count of records for each prizes.category in descending order of count.

```
3 db.getCollection("laureate2").aggregate([
4   {
5     $unwind: "$prizes"
6   },
7   {
8     $group: {
9       _id: {PrizesCategory: "$prizes.category"},
10      count: {$sum: 1}
11    }
12  },
13  {
14    $sort: {count: -1}
15  }
16 ])
```

Raw shell output: Aggregate Query (line 3) x

laureate2 > count

_id	count
(1 fields)	222.0
(1 fields)	216.0
(1 fields)	186.0
(1 fields)	135.0
(1 fields)	117.0
(1 fields)	86.0

1 document selected

-

```
Raw shell output: Aggregate Query (line 3) x
```

laureate2 > _id

(Document id)	PrizesCategory
(1 fields)	medicine
(1 fields)	physics
(1 fields)	chemistry
(1 fields)	peace
(1 fields)	literature
(1 fields)	economics

1 document selected

The table view shows the count of each category of prizes, with medicine being the most frequent one (222), followed by physics, then chemistry, peace, literature, and then economics in the last with a count of 86.

To get a better view of the output with the count specified in front of each category name, the view can be changed to json which shows the following output:

```
Raw shell output: Aggregate Query (line 3) x
```

```
1 {
2   "_id": {
3     "PrizesCategory": "medicine"
4   },
5   "count": 222.0
6 }
7 {
8   "_id": {
9     "PrizesCategory": "physics"
10  },
11  "count": 216.0
12 }
13 {
14   "_id": {
15     "PrizesCategory": "chemistry"
16   },
17   "count": 186.0
18 }
19 {
20   "_id": {
21     "PrizesCategory": "peace"
22   },
23   "count": 135.0
24 }
25 {
26   "_id": {
```

1 document selected

We can see that 222 people gained a Nobel prize award in the field of medicine.

- the count of records for each gender, diedCountryCode, prize.category when prize.category is "physics". Order the output by diedCountryCode.

```
1 db.getCollection("laureate").aggregate(
2 [
3   {
4     $match: { prizes: { $elemMatch: { category: "physics" }}}
5   },
6   {
7     $unwind: "$prizes"
8   },
9   {
10    $group: {
11      _id: { Gender: "$gender", diedCountryCode: "$diedCountryCode", PrizeCategory: "$prizes.category",
12        count: { $sum: 1 }
13    }
14  },
15  {
16    $sort: { diedCountryCode: 1 }
17  }
18 ]
19 )
```

Raw shell output: Aggregate Query (line 1) x

1 document selected

```
1 {
2   "_id": {
3     "Gender": "female",
4     "PrizeCategory": "physics"
5   },
6   "count": 2.0
7 }
8 {
9   "_id": {
10    "Gender": "male",
11    "PrizeCategory": "physics"
12  },
13  "count": 76.0
14 }
15 {
16  "_id": {
17    "Gender": "male",
18    "diedCountryCode": "IN",
19    "PrizeCategory": "physics"
20  },
21  "count": 1.0
22 }
23 {
24  "_id": {
25    "Gender": "male",
26    "diedCountryCode": "IT",
27    "PrizeCategory": "physics"
28  },
29  "count": 1.0
30 }
```

1 document selected

```
62 }
63 {
64   "_id": {
65     "Gender": "male",
66     "diedCountryCode": "US",
67     "PrizeCategory": "physics"
68   },
69   "count": 60.0
70 }
```

The output shows Gender, diedCountrycode and count of each category that have prizecategory of Physics. We can interpret from the results that there are 60 male Nobel prize winners that died in the US. There were 76 males and 2 females who won Physics Nobel Prize but their death place was unknown.

- Come up with your own query to show any interesting insight. Use at least two fields for match and two fields for group.

```
1 db.getCollection("laureate2").aggregate(
2 [
3   {
4     $match: { prizes: { $elemMatch: {category: "medicine"}}, bornCountry: "USA" }
5   },
6   {
7     $unwind: "$prizes"
8   },
9   {
10    $group: {
11      _id: { Gender: "$gender", bornCountry: "$bornCountry", PrizeCategory: "$prizes.category" },
12      count: { $sum: 1 }
13    }
14  },
15  {
16    $sort: { count: 1 }
17  }
18 ]
19 )
```

Raw shell output: Aggregate Query (line 1) x

laureate2 > count

_id	count
{ "Gender": "female", "bornCountry": "USA", "PrizeCategory": "medicine" }	5.0
{ "Gender": "male", "bornCountry": "USA", "PrizeCategory": "medicine" }	73.0

1 document selected

```
1 {
2   "_id": {
3     "Gender": "female",
4     "bornCountry": "USA",
5     "PrizeCategory": "medicine"
6   },
7   "count": 5.0
8 }
9 {
10  "_id": {
11    "Gender": "male",
12    "bornCountry": "USA",
13    "PrizeCategory": "medicine"
14  },
15  "count": 73.0
16 }
17 }
```

Raw shell output: Aggregate Query (line 1) x

1 document selected

From the above query, it can be seen that out of all the Nobel prize winners in the field of medicine and born in the US, there were 5 females and 73 males. When we see for Canada, the results show that only 4 males received Nobel prize awards in medicine, with no females. This is shown below:

```
1 db.getCollection("laureate2").aggregate(
2 [
3   {
4     $match: { prizes: { $elemMatch: {category: "medicine"}}, bornCountry: "Canada" }
5   },
6   {
7     $unwind: "$prizes"
8   },
9   {
10    $group: {
11      _id: { Gender: "$gender", bornCountry: "$bornCountry", PrizeCategory: "$prizes.category" },
12      count: { $sum: 1 }
13    }
14  },
15  {
16    $sort: { count: 1 }
17  }
18 ]
19 )
```

Raw shell output: Aggregate Query (line 1) x

```
1 {
2   "_id": {
3     "Gender": "male",
4     "bornCountry": "Canada",
5     "PrizeCategory": "medicine"
6   },
7   "count": 4.0
8 }
```

1 document selected