Ali Hashir   DATA SCIENCE
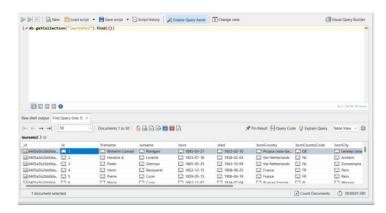Ah05433

HW6 - NoSQL

Connection successful:

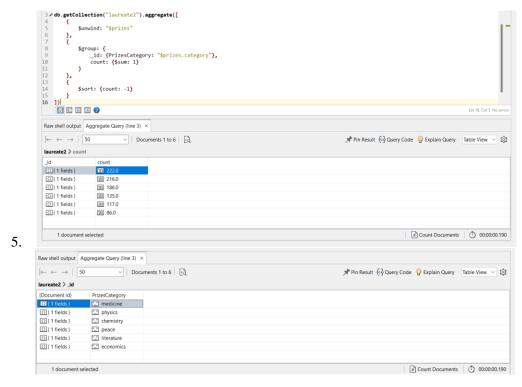1. the count of total number of records in the collection

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("laureate").aggregate([{
5      $group: {
6          _id: ["$diedCountryCode"],
7          count: { $sum: 1 }
8      }
9  },
10 {
11     $sort: {count: -1}
12 }]).pretty()
13
```

```
145
146 [
147   { _id: [ null ], count: 335 },
148   { _id: [ 'US' ], count: 223 },
149   { _id: [ 'GB' ], count: 83 },
150   { _id: [ 'DE' ], count: 59 },
151   { _id: [ 'FR' ], count: 51 },
152   { _id: [ 'SE' ], count: 29 },
153   { _id: [ 'CH' ], count: 27 },
154   { _id: [ 'RU' ], count: 15 },
155   { _id: [ 'IT' ], count: 14 },
156   { _id: [ 'DK' ], count: 10 },
157   { _id: [ 'NL' ], count: 10 },
158   { _id: [ 'ES' ], count: 10 },
159   { _id: [ 'JP' ], count: 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.

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

5.

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:

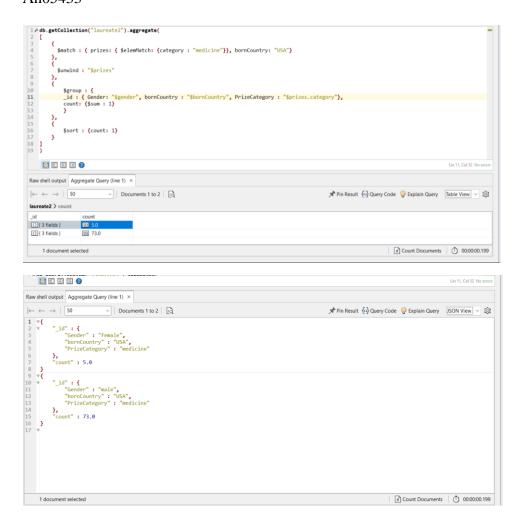We can see that 222 people gained a Nobel prize award in the field of medicine.

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

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.

7. 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) ×

laureate2 > count

_id	count
{ 3 fields }	5.0
{ 3 fields }	73.0

1 document selected   Count Documents   00:00:00.199

```
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
```

1 document selected   Count Documents   00:00:00.199

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 ]
```

Raw shell output | Aggregate Query (line 1) ×

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

1 document selected   Count Documents   00:00:00.202