Mongo Assignment 2

MongoDB - Aggregation Exercises

Import the zips.json file into your MongoDB. Database name is "population" and collection name is "zipcodes".

```
mongoimport --db population --collection zipcodes --file zips.json
```

Atlanta Population

1. use db.zipcodes.find() to filter results to only the results where city is ATLANTA and state is GA.

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.find({"city":"ATLANTA","state":"GA"});
[
  {
    _id: '30303',
    city: 'ATLANTA',
    loc: [ -84.388846, 33.752504 ],
    pop: 1845,
    state: 'GA'
  },
```

2. use db.zipcodes.aggregate with $match to do the same as above.

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{$match:{"city":"ATLANTA","state":"GA"}}]);
[
  {
    _id: '30303',
    city: 'ATLANTA',
    loc: [ -84.388846, 33.752504 ],
    pop: 1845,
    state: 'GA'
  },
```

3. use $group to count the number of zip codes in Atlanta.

```
Type "it" for more
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{$group:{_id:{"city": "ATLANTA"},count:{$sum:1}}}]);
[ { _id: { city: 'ATLANTA' }, count: 29353 } ]
Atlas atlas-b1w4mn-shard-0 [primary] population>
```

4. use $group to find the total population in Atlanta.

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{$group:{_id:{"city":"ATLANTA"},total:{$sum:"$pop"}}}]);
[ { _id: { city: 'ATLANTA' }, total: 248408400 } ]
Atlas atlas-b1w4mn-shard-0 [primary] population>
```

Populations By State

1. use aggregate to calculate the total population for each state

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: "$state", totalPop: { $sum: "$pop" }}}]);
[
  { _id: 'PA', totalPop: 11881643 },
  { _id: 'MT', totalPop: 798948 },
  { _id: 'NE', totalPop: 1578139 },
  { _id: 'NY', totalPop: 17990402 },
  { _id: 'WV', totalPop: 1793146 },
  { _id: 'VA', totalPop: 6181479 },
  { _id: 'NH', totalPop: 1109252 },
  { _id: 'MN', totalPop: 4372982 },
  { _id: 'ME', totalPop: 1226648 },
  { _id: 'SC', totalPop: 3486703 },
  { _id: 'OH', totalPop: 10846517 },
  { _id: 'FL', totalPop: 12686644 },
  { _id: 'IL', totalPop: 11427576 },
  { _id: 'MA', totalPop: 6016425 },
  { _id: 'AR', totalPop: 2350725 },
  { _id: 'ID', totalPop: 1006749 },
  { _id: 'TX', totalPop: 16984601 },
  { _id: 'CA', totalPop: 29754890 },
  { _id: 'IA', totalPop: 2776420 },
  { _id: 'OR', totalPop: 2842321 }
]
```

2. sort the results by population, highest first

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{ $group: { _id: "$state", totalPop: { $sum: "$pop" } } }, { $sort: { totalPop: -1 } }])
[
  { _id: 'CA', totalPop: 29754890 },
  { _id: 'NY', totalPop: 17990402 },
```

3. limit the results to just the first 3 results. What are the top 3 states in population?

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{$group:{_id:"$state",totalPop:{$sum:"$pop"}}},{$sort:{totalPop:-1}},{$limit: 3}])
[
  { _id: 'CA', totalPop: 29754890 },
  { _id: 'NY', totalPop: 17990402 },
  { _id: 'TX', totalPop: 16984601 }
]
```

Populations by City

1. use aggregate to calculate the total population for each city (you have to use city/state combination). You can use a combination for the _id of the $group: { city: '$city', state: '$state' }

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } } }
... },{ $group: { _id: "$_id.state",CityPop: { $sum: "$pop" } } }] )
[
  { _id: 'MN', CityPop: 4372982 },
  { _id: 'VA', CityPop: 6181479 },
```

2. sort the results by population, highest first

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } }
... },{ $group: { _id: "$_id.state",CityPop: { $sum: "$pop" } } },{$sort:{CityPop:-1}}] )
[
  { _id: 'CA', CityPop: 29754890 },
  { _id: 'NY', CityPop: 17990402 },
  { _id: 'TX', CityPop: 16984601 },
  { _id: 'FL', CityPop: 12686644 },
```

3. limit the results to just the first 3 results. What are the top 3 cities in population?

```
Type "it" for more
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } }
... },{ $group: { _id: "$_id.state",CityPop: { $sum: "$pop" } } },{$sort:{CityPop:-1}},{$limit:3}] )
[
  { _id: 'CA', CityPop: 29754890 },
  { _id: 'NY', CityPop: 17990402 },
  { _id: 'TX', CityPop: 16984601 }
]
```

4. What are the top 3 cities in population in Texas?

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate([{$match:{state:"TX"}},{$group:{_id:"$city",pop:{$max:"$pop"}}},{$sort:{pop:-1}},{$limit:3}])
[
  { _id: 'BROWNSVILLE', pop: 79463 },
  { _id: 'ALTON', pop: 67604 },
  { _id: 'SAN ANTONIO', pop: 58355 }
]
```

Bonus

1. Write a query to get the average city population for each state.

```
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } }
... },{ $group: { _id: "$_id.state", avgCityPop: { $avg: "$pop" } } }] )
[
  { _id: 'VA', avgCityPop: 8526.177931034483 },
  { _id: 'MN', avgCityPop: 5372.21375921376 },
  { _id: 'OR', avgCityPop: 8262.561046511628 },
  { _id: 'WA', avgCityPop: 12258.670025188916 },
  { _id: 'PA', avgCityPop: 8679.067202337472 },
  { _id: 'NH', avgCityPop: 5232.320754716981 },
  { _id: 'NE', avgCityPop: 3034.882692307692 },
  { _id: 'NY', avgCityPop: 13131.680291970803 },
  { _id: 'WV', avgCityPop: 2771.4775888717154 },
  { _id: 'MA', avgCityPop: 14855.37037037037 },
  { _id: 'DC', avgCityPop: 303450 },
  { _id: 'ID', avgCityPop: 4320.811158798283 },
  { _id: 'ME', avgCityPop: 3006.4901960784314 },
  { _id: 'CA', avgCityPop: 27756.42723880597 },
  { _id: 'OH', avgCityPop: 12700.839578454332 },
  { _id: 'AR', avgCityPop: 4175.355239786856 },
  { _id: 'FL', avgCityPop: 27400.958963282937 },
  { _id: 'IL', avgCityPop: 9954.334494773519 },
  { _id: 'SC', avgCityPop: 11139.626198083068 },
  { _id: 'VT', avgCityPop: 2315.8765432098767 }
]
```

2. What are the top 3 states in terms of average city population?

```
Type "it" for more
Atlas atlas-b1w4mn-shard-0 [primary] population> db.zipcodes.aggregate( [{ $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } }
... },{ $group: { _id: "$_id.state", avgCityPop: { $avg: "$pop" } } },{$sort:{avgCityPop:-
..... 1}},{$limit:3}] )
[
  { _id: 'DC', avgCityPop: 303450 },
  { _id: 'CA', avgCityPop: 27756.42723880597 },
  { _id: 'FL', avgCityPop: 27400.958963282937 }
]
```