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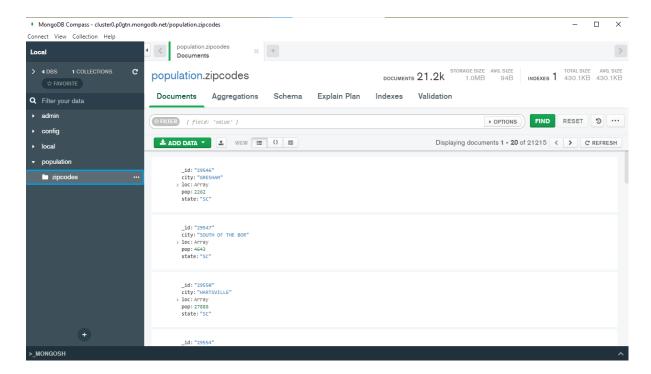
Import the zips.json file into your MongoDB. Database name is "population" and collection name is "zipcodes".

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.
2. use db.zipcodes.aggregate with $match to do the same as above.
3. use $group to count the number of zip codes in Atlanta.
4. use $group to find the total population in Atlanta.

```
> db.zipcodes.find({city:"ATLANTA",state:"GA"})
{ "_id" : "30303", "city" : "ATLANTA", "loc" : [ -84.388846, 33.752504 ], "pop" : 1845, "state" : "GA" }
{ "_id" : "30305", "city" : "ATLANTA", "loc" : [ -84.385145, 33.831963 ], "pop" : 19122, "state" : "GA" }
{ "_id" : "30306", "city" : "ATLANTA", "loc" : [ -84.351418, 33.786027 ], "pop" : 20081, "state" : "GA" }
{ "_id" : "30307", "city" : "ATLANTA", "loc" : [ -84.335957, 33.769138 ], "pop" : 16330, "state" : "GA" }
{ "_id" : "30308", "city" : "ATLANTA", "loc" : [ -84.375744, 33.771839 ], "pop" : 8549, "state" : "GA" }
{ "_id" : "30309", "city" : "ATLANTA", "loc" : [ -84.388338, 33.798407 ], "pop" : 14766, "state" : "GA" }
{ "_id" : "30310", "city" : "ATLANTA", "loc" : [ -84.423173, 33.727849 ], "pop" : 34017, "state" : "GA" }
{ "_id" : "30311", "city" : "ATLANTA", "loc" : [ -84.470219, 33.722957 ], "pop" : 34880, "state" : "GA" }
{ "_id" : "30312", "city" : "ATLANTA", "loc" : [ -84.378125, 33.746749 ], "pop" : 17683, "state" : "GA" }
{ "_id" : "30313", "city" : "ATLANTA", "loc" : [ -84.39352, 33.76825 ], "pop" : 8038, "state" : "GA" }
{ "_id" : "30314", "city" : "ATLANTA", "loc" : [ -84.425546, 33.756103 ], "pop" : 26649, "state" : "GA" }
{ "_id" : "30315", "city" : "ATLANTA", "loc" : [ -84.380771, 33.705062 ], "pop" : 41061, "state" : "GA" }
{ "_id" : "30316", "city" : "ATLANTA", "loc" : [ -84.333913, 33.721686 ], "pop" : 34668, "state" : "GA" }
{ "_id" : "30317", "city" : "ATLANTA", "loc" : [ -84.31685, 33.749788 ], "pop" : 16395, "state" : "GA" }
{ "_id" : "30318", "city" : "ATLANTA", "loc" : [ -84.445432, 33.786454 ], "pop" : 53894, "state" : "GA" }
{ "_id" : "30319", "city" : "ATLANTA", "loc" : [ -84.335091, 33.868728 ], "pop" : 32138, "state" : "GA" }
{ "_id" : "30324", "city" : "ATLANTA", "loc" : [ -84.354867, 33.820609 ], "pop" : 15044, "state" : "GA" }
{ "_id" : "30326", "city" : "ATLANTA", "loc" : [ -84.358232, 33.848168 ], "pop" : 125, "state" : "GA" }
{ "_id" : "30327", "city" : "ATLANTA", "loc" : [ -84.419966, 33.862723 ], "pop" : 18467, "state" : "GA" }
{ "_id" : "30329", "city" : "ATLANTA", "loc" : [ -84.321402, 33.823555 ], "pop" : 17013, "state" : "GA" }
Type "it" for more
> db.zipcodes.aggregate([{$match: {city:"ATLANTA", state: "GA"}}])
{ "_id" : "30303", "city" : "ATLANTA", "loc" : [ -84.388846, 33.752504 ], "pop" : 1845, "state" : "GA" }
{ "_id" : "30305", "city" : "ATLANTA", "loc" : [ -84.385145, 33.831963 ], "pop" : 19122, "state" : "GA" }
{ "_id" : "30306", "city" : "ATLANTA", "loc" : [ -84.351418, 33.786027 ], "pop" : 20081, "state" : "GA" }
{ "_id" : "30307", "city" : "ATLANTA", "loc" : [ -84.335957, 33.769138 ], "pop" : 16330, "state" : "GA" }
{ "_id" : "30308", "city" : "ATLANTA", "loc" : [ -84.375744, 33.771839 ], "pop" : 8549, "state" : "GA" }
{ "_id" : "30309", "city" : "ATLANTA", "loc" : [ -84.388338, 33.798407 ], "pop" : 14766, "state" : "GA" }
{ "_id" : "30310", "city" : "ATLANTA", "loc" : [ -84.423173, 33.727849 ], "pop" : 34017, "state" : "GA" }
{ "_id" : "30311", "city" : "ATLANTA", "loc" : [ -84.470219, 33.722957 ], "pop" : 34880, "state" : "GA" }
{ "_id" : "30312", "city" : "ATLANTA", "loc" : [ -84.378125, 33.746749 ], "pop" : 17683, "state" : "GA" }
{ "_id" : "30313", "city" : "ATLANTA", "loc" : [ -84.39352, 33.76825 ], "pop" : 8038, "state" : "GA" }
{ "_id" : "30314", "city" : "ATLANTA", "loc" : [ -84.425546, 33.756103 ], "pop" : 26649, "state" : "GA" }
{ "_id" : "30315", "city" : "ATLANTA", "loc" : [ -84.380771, 33.705062 ], "pop" : 41061, "state" : "GA" }
{ "_id" : "30316", "city" : "ATLANTA", "loc" : [ -84.333913, 33.721686 ], "pop" : 34668, "state" : "GA" }
{ "_id" : "30317", "city" : "ATLANTA", "loc" : [ -84.31685, 33.749788 ], "pop" : 16395, "state" : "GA" }
{ "_id" : "30318", "city" : "ATLANTA", "loc" : [ -84.445432, 33.786454 ], "pop" : 53894, "state" : "GA" }
{ "_id" : "30319", "city" : "ATLANTA", "loc" : [ -84.335091, 33.868728 ], "pop" : 32138, "state" : "GA" }
{ "_id" : "30324", "city" : "ATLANTA", "loc" : [ -84.354867, 33.820609 ], "pop" : 15044, "state" : "GA" }
{ "_id" : "30326", "city" : "ATLANTA", "loc" : [ -84.358232, 33.848168 ], "pop" : 125, "state" : "GA" }
{ "_id" : "30327", "city" : "ATLANTA", "loc" : [ -84.419966, 33.862723 ], "pop" : 18467, "state" : "GA" }
{ "_id" : "30329", "city" : "ATLANTA", "loc" : [ -84.321402, 33.823555 ], "pop" : 17013, "state" : "GA" }
type "it" for more
> db.zipcodes.aggregate([{$match:{city:"ATLANTA"}},{$group:{_id:"$city", count:{$sum: 1}}}])
{ "_id" : "ATLANTA", "count" : 41 }
> db.zipcodes.aggregate([{$match:{city:"ATLANTA"}},{$group:{_id:"$city", count:{$sum:"$pop"}}}])
{ "_id" : "ATLANTA", "count" : 630046 }
```

Populations By State

1. use aggregate to calculate the total population for each state
2. sort the results by population, highest first
3. limit the results to just the first 3 results. What are the top 3 states in population?

```
> db.zipcodes.aggregate([{$group:{_id:"$state", TotPopulation:{$sum:"$pop"}}}])
{ "_id" : "NV", "TotPopulation" : 1201833 }
{ "_id" : "CO", "TotPopulation" : 3293755 }
{ "_id" : "OR", "TotPopulation" : 2842321 }
{ "_id" : "MD", "TotPopulation" : 4781379 }
{ "_id" : "AZ", "TotPopulation" : 3665228 }
{ "_id" : "CT", "TotPopulation" : 3287116 }
{ "_id" : "MO", "TotPopulation" : 5110648 }
{ "_id" : "TX", "TotPopulation" : 16984601 }
{ "_id" : "WY", "TotPopulation" : 453528 }
{ "_id" : "NJ", "TotPopulation" : 7730188 }
{ "_id" : "TN", "TotPopulation" : 4876457 }
{ "_id" : "DE", "TotPopulation" : 666168 }
{ "_id" : "DC", "TotPopulation" : 606900 }
{ "_id" : "SC", "TotPopulation" : 3486703 }
{ "_id" : "OH", "TotPopulation" : 10846517 }
{ "_id" : "MA", "TotPopulation" : 6016425 }
{ "_id" : "KY", "TotPopulation" : 3675484 }
{ "_id" : "IL", "TotPopulation" : 11427576 }
{ "_id" : "WV", "TotPopulation" : 1793146 }
{ "_id" : "ID", "TotPopulation" : 1006749 }
Type "it" for more
> db.zipcodes.aggregate([{$group:{_id:"$state", TotPopulation:{$sum:"$pop"}}}, {$sort:{TotPopulation:-1}}])
{ "_id" : "CA", "TotPopulation" : 29754890 }
{ "_id" : "NY", "TotPopulation" : 17990402 }
{ "_id" : "TX", "TotPopulation" : 16984601 }
{ "_id" : "FL", "TotPopulation" : 12686644 }
{ "_id" : "PA", "TotPopulation" : 11881643 }
{ "_id" : "IL", "TotPopulation" : 11427576 }
{ "_id" : "OH", "TotPopulation" : 10846517 }
{ "_id" : "MI", "TotPopulation" : 9295297 }
{ "_id" : "NJ", "TotPopulation" : 7730188 }
{ "_id" : "NC", "TotPopulation" : 6628637 }
{ "_id" : "GA", "TotPopulation" : 6478216 }
{ "_id" : "VA", "TotPopulation" : 6181479 }
{ "_id" : "MA", "TotPopulation" : 6016425 }
{ "_id" : "IN", "TotPopulation" : 5544136 }
{ "_id" : "MO", "TotPopulation" : 5110648 }
{ "_id" : "WI", "TotPopulation" : 4891769 }
{ "_id" : "TN", "TotPopulation" : 4876457 }
{ "_id" : "WA", "TotPopulation" : 4866692 }
{ "_id" : "MD", "TotPopulation" : 4781379 }
{ "_id" : "MN", "TotPopulation" : 4372982 }
Type "it" for more
> db.zipcodes.aggregate([{$group:{_id:"$state", TotPopulation:{$sum:"$pop"}}}, {$sort:{TotPopulation:-1}}, {$limit: 3}])
{ "_id" : "CA", "TotPopulation" : 29754890 }
{ "_id" : "NY", "TotPopulation" : 17990402 }
{ "_id" : "TX", "TotPopulation" : 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' }
2. sort the results by population, highest first
3. limit the results to just the first 3 results. What are the top 3 cities in population?
4. What are the top 3 cities in population in Texas?

```
> db.zipcodes.aggregate([{$group:{_id:"$city", TotPopulation:{$sum:"$pop"}}}])
{ "_id" : "CAMBRIDGE CITY", "TotPopulation" : 4998 }
{ "_id" : "PHOENIX", "TotPopulation" : 902249 }
{ "_id" : "FORT SUPPLY", "TotPopulation" : 1222 }
{ "_id" : "MOUNTAIN MESA", "TotPopulation" : 6285 }
{ "_id" : "LENZBURG", "TotPopulation" : 1316 }
{ "_id" : "TOPONAS", "TotPopulation" : 697 }
{ "_id" : "EASTSOUND", "TotPopulation" : 2259 }
{ "_id" : "LIBERTY FARMS", "TotPopulation" : 14536 }
{ "_id" : "PALISADES", "TotPopulation" : 1032 }
{ "_id" : "COUNTRY CLUB HIL", "TotPopulation" : 16225 }
{ "_id" : "MULLENS", "TotPopulation" : 2590 }
{ "_id" : "NEWHALL", "TotPopulation" : 26413 }
{ "_id" : "TOKELAND", "TotPopulation" : 891 }
{ "_id" : "ROCKLEDGE", "TotPopulation" : 20736 }
{ "_id" : "HILL AIR FORCE B", "TotPopulation" : 5432 }
{ "_id" : "HYTOP", "TotPopulation" : 16934 }
{ "_id" : "GUYSVILLE", "TotPopulation" : 359 }
{ "_id" : "FIFTY SIX", "TotPopulation" : 346 }
{ "_id" : "HALEDON", "TotPopulation" : 19757 }
{ "_id" : "MINDEN", "TotPopulation" : 30621 }
Type "it" for more
> db.zipcodes.aggregate([{$group:{_id:"$city", TotPopulation:{$sum:"$pop"}}}, {$sort:{TotPopulation:-1}}])
{ "_id" : "CHICAGO", "TotPopulation" : 2452177 }
{ "_id" : "BROCKLYN", "TotPopulation" : 2341387 }
{ "_id" : "HOUSTON", "TotPopulation" : 2123053 }
{ "_id" : "LOS ANGELES", "TotPopulation" : 2102295 }
{ "_id" : "PHILADELPHIA", "TotPopulation" : 1639862 }
{ "_id" : "NEW YORK", "TotPopulation" : 1476790 }
{ "_id" : "BRONX", "TotPopulation" : 1209548 }
{ "_id" : "SAN DIEGO", "TotPopulation" : 1054316 }
{ "_id" : "DALLAS", "TotPopulation" : 999042 }
{ "_id" : "DETROIT", "TotPopulation" : 967468 }
{ "_id" : "PHOENIX", "TotPopulation" : 902249 }
{ "_id" : "MIAMI", "TotPopulation" : 848436 }
{ "_id" : "COLUMBUS", "TotPopulation" : 825448 }
{ "_id" : "SAN JOSE", "TotPopulation" : 817497 }
{ "_id" : "SAN ANTONIO", "TotPopulation" : 813188 }
{ "_id" : "WASHINGTON", "TotPopulation" : 780954 }
{ "_id" : "BALTIMORE", "TotPopulation" : 738846 }
{ "_id" : "JACKSONVILLE", "TotPopulation" : 735505 }
{ "_id" : "SAN FRANCISCO", "TotPopulation" : 723993 }
{ "_id" : "CLEVELAND", "TotPopulation" : 687451 }
```

```
> db.zipcodes.aggregate([{$group:{_id:"$city", TotPopulation:{$sum:"$pop"}}}, {$sort:{TotPopulation:-1}}, {$limit: 3}])
{ "_id" : "CHICAGO", "TotPopulation" : 2452177 }
{ "_id" : "BROCKLYN", "TotPopulation" : 2341387 }
{ "_id" : "HOUSTON", "TotPopulation" : 2123053 }
> db.zipcodes.aggregate([{$match:{state:"TX"}},{$group:{_id:"$city", TotPopulation:{$sum:"$pop"}}}, {$sort:{TotPopulation:-1}}, {$limit: 3}])
{ "_id" : "HOUSTON", "TotPopulation" : 2095918 }
{ "_id" : "DALLAS", "TotPopulation" : 940191 }
{ "_id" : "SAN ANTONIO", "TotPopulation" : 811792 }
```

Bonus

1. Write a query to get the average city population for each state.
2. What are the top 3 states in terms of average city population?

```
> db.zipcodes.aggregate([[$group:{_id:"$state", AvgPopulation:{$avg:"$pop"}}]])
{ "_id" : "NC", "AvgPopulation" : 9402.321985815603 }
{ "_id" : "WV", "AvgPopulation" : 2733.454268292683 }
{ "_id" : "VA", "AvgPopulation" : 7575.341911764706 }
{ "_id" : "MI", "AvgPopulation" : 10611.069634703197 }
{ "_id" : "IL", "AvgPopulation" : 9238.137429264349 }
{ "_id" : "OH", "AvgPopulation" : 10771.119165839125 }
{ "_id" : "DE", "AvgPopulation" : 12569.207547169812 }
{ "_id" : "DC", "AvgPopulation" : 25287.5 }
{ "_id" : "KY", "AvgPopulation" : 4543.243510506799 }
{ "_id" : "MA", "AvgPopulation" : 12692.879746835442 }
{ "_id" : "SC", "AvgPopulation" : 9962.00857142857 }
{ "_id" : "TN", "AvgPopulation" : 8378.792096219931 }
{ "_id" : "ID", "AvgPopulation" : 4126.020491803279 }
{ "_id" : "MO", "AvgPopulation" : 5141.495981891348 }
{ "_id" : "CT", "AvgPopulation" : 12498.539923954373 }
{ "_id" : "TX", "AvgPopulation" : 10164.333333333334 }
{ "_id" : "WY", "AvgPopulation" : 3239.485714285714 }
{ "_id" : "UT", "AvgPopulation" : 8404.148341463415 }
{ "_id" : "NJ", "AvgPopulation" : 14315.182962962962 }
{ "_id" : "AZ", "AvgPopulation" : 13574.918518518518 }
Type "it" for more
> db.zipcodes.aggregate([[$group:{_id:"$state", AvgPopulation:{$avg:"$pop"}}], {$sort:{AvgPopulation:-1}}, {$limit: 3}])
{ "_id" : "DC", "AvgPopulation" : 25287.5 }
{ "_id" : "CA", "AvgPopulation" : 19627.236147757256 }
{ "_id" : "FL", "AvgPopulation" : 15779.407960199005 }
```