

The new per-state documents have two fields: the `_id` field and the `totalPop` field. The `_id` field contains the value of the `state`; i.e. the group by field. The `totalPop` field is a calculated field that contains the total population of each state. To calculate the value, `$group` uses the `$sum` operator to add the population field (`pop`) for each state.

After the `$group` stage, the documents in the pipeline resemble the following:

```
{
  "_id" : "AK",
  "totalPop" : 550043
}
```

- The `$match` stage filters these grouped documents to output only those documents whose `totalPop` value is greater than or equal to 10 million. The `$match` stage does not alter the matching documents but outputs the matching documents unmodified.

The equivalent *SQL* for this aggregation operation is:

```
SELECT state, SUM(pop) AS totalPop
FROM zipcodes
GROUP BY state
HAVING totalPop >= (10*1000*1000)
```

See also:

`$group`, `$match`, `$sum`

## Return Average City Population by State

The following aggregation operation returns the average populations for cities in each state:

```
db.zipcodes.aggregate( [
  { $group: { _id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } } },
  { $group: { _id: "$_id.state", avgCityPop: { $avg: "$pop" } } }
] )
```

In this example, the *aggregation pipeline* (page 440) consists of the `$group` stage followed by another `$group` stage:

- The first `$group` stage groups the documents by the combination of `city` and `state`, uses the `$sum` expression to calculate the population for each combination, and outputs a document for each `city` and `state` combination.<sup>14</sup>

After this stage in the pipeline, the documents resemble the following:

```
{
  "_id" : {
    "state" : "CO",
    "city" : "EDGEWATER"
  },
  "pop" : 13154
}
```

- A second `$group` stage groups the documents in the pipeline by the `_id.state` field (i.e. the `state` field inside the `_id` document), uses the `$avg` expression to calculate the average city population (`avgCityPop`) for each state, and outputs a document for each state.

The documents that result from this aggregation operation resembles the following:

---

<sup>14</sup> A city can have more than one zip code associated with it as different sections of the city can each have a different zip code.