

Erin Zahner

Dr. F

CSC 351

22 November 2024

Assignment 5

1. Over how many years was the unemployment data collected?

```
db.HW5.aggregate([
  { $group: { _id: "$Year" } },
  { $count: "numYears" }
])
```

```
> db.HW5.aggregate([
  { $group: { _id: "$Year" } },
  { $count: "numYears" }
])
< {
  numYears: 27
}
```

To find how many years the unemployment data was collected, we need to count how many distinct groups of years we have in the data. To do this we \$group based on year and then \$count the result. From doing this, we find that the unemployment data was collected for a span of 27 years.

2. How many states were reported on in this dataset?

```
db.HW5.aggregate([
  { $group: { _id: "$State" } },
  { $count: "numStates" }
])
```

```
> db.HW5.aggregate([
  { $group: { _id: "$State" } },
  { $count: "numStates" }
])
< {
  numStates: 47
}
```

This query involved the same exact methodology as the previous one. To find the number of states reported on in this dataset, we first had to group by distinct states and then count the number of state groups. From this, we found that there were 47 states reported on in this dataset.

3. What does this query compute?:

```
db.unemployment.find({Rate : {$lt: 1.0}}).count()
```

```
> db.HW5.find({Rate : {$lt: 1.0}}).count()
< 657
```

This query counts the number of documents (entries) where the unemployment rate is less than 1.0%. As we can see, there are 657 states with an unemployment rate of less than 1. I had to change db.unemployment to db.HW5 because I named my collection “HW5”, however, everything else remained the same.

4. Find all counties with unemployment rate higher than 10%

```
db.HW5.aggregate([
  { $match: { Rate: { $gt: 10.0 } } },
  { $group: { _id: "$County" } }
])
```

```
db.HW5.aggregate([
  { $match: { Rate: { $gt: 10.0 } } },
  { $group: { _id: "$County" } }
])
{
  _id: 'Crane County'
}
{
  _id: 'Attala County'
}
```

To find the counties with an unemployment rate higher than 10%, we first have to write a \$match query to find the data where the rate is greater than 10. Once we have that, we use \$group to group the distinct counties together. After running the query, a sample of the data was returned, listing 10 of the counties with an unemployment rate of greater than 10%.

5. Calculate the average unemployment rate across all states.

```
db.HW5.aggregate([
  { $group: { _id: null, averageRate: { $avg: "$Rate" } } }
])
```

```
> db.HW5.aggregate([
  { $group: { _id: null, averageRate: { $avg: "$Rate" } } }
])
< {
  _id: null,
  averageRate: 6.1750097115006755
}
```

To calculate the average unemployment rate across all states, we first use the \$group stage to group all the documents together. By setting _id: null, we ensure that all documents are treated as a single group. Then, we use the \$avg operator on the Rate field to compute the average unemployment rate across all records in the collection. As shown in the screenshot above, the average rate of unemployment across all states is 6.175.

6. Find all counties with an unemployment rate between 5% and 8%.

```
db.HW5.aggregate([
  { $match: { Rate: { $gte: 5.0, $lte: 8.0 } } },
  { $group: { _id: "$County" } }
])
```

```
> db.HW5.aggregate([
  { $match: { Rate: { $gte: 5.0, $lte: 8.0 } } },
  { $group: { _id: "$County" } }
])
< {
  _id: 'Neshoba County'
}
{
  _id: 'Day County'
}
```

This query is almost identical to the one in question 4, the only difference is that we have two criteria we are matching on. Here, we had to check if the unemployment rates were greater than or equal to 5% and less than or equal to 8%. After doing this, we get a sample of results that fit this criteria, such as Neshoba County.

7. Find the state with the highest unemployment rate. Hint. Use { \$limit: 1 }

```
db.HW5.aggregate([
  { $sort: { Rate: -1 } },
  { $limit: 1 }
])
```

```
> db.HW5.aggregate([
  { $sort: { Rate: -1 } },
  { $limit: 1 }
])
< {
  _id: ObjectId('672ce67106ae8647a20c1f61'),
  Year: 1992,
  Month: 'January',
  State: 'Colorado',
  County: 'San Juan County',
  Rate: 58.4
}
```

test>

To find the state with the highest unemployment rate, we first need to \$sort the data in descending order, with the highest value at the top. Then, to just select the state with the highest unemployment rate, we \$limit the results by 1. From this, we can see that Colorado had the highest rate of unemployment in 1992, with a rate of 58.4%.

8. Count how many counties have an unemployment rate above 5%.

```
db.HW5.aggregate([
  { $match: { Rate: { $gt: 5.0 } } },
  { $group: { _id: "$County" } },
  { $count: "countiesAbove5Percent" }
])
```

```
> db.HW5.aggregate([
  { $match: { Rate: { $gt: 5.0 } } },
  { $group: { _id: "$County" } },
  { $count: "countiesAbove5Percent" }
])
< {
  countiesAbove5Percent: 1736
}
```

To find the number of counties that have an unemployment rate above 5%, we use the same method as questions 4 and 6. We \$match on rate greater than 5, \$group by state, and then count the results. As a result, we see that 1736 counties that have unemployment rates above 5%.

9. Calculate the average unemployment rate per state by year.

```
db.HW5.aggregate([
  { $group: {
    _id: { state: "$State", year: "$Year" },
    averageRate: { $avg: "$Rate" }
  }
}]
)
```

```
> db.HW5.aggregate([
  { $group: {
    _id: { state: "$State", year: "$Year" },
    averageRate: { $avg: "$Rate" }
  }
}]
)
< {
  _id: {
    state: 'Indiana',
    year: 2003
  },
  averageRate: 5.483967391304348
}
{
  _id: {
    state: 'Pennsylvania',
    year: 2010
  },
  averageRate: 8.92089552238806
}
```

To find the average unemployment rate per state by year, we had to \$group by state and year, and then \$avg the rate. By doing this, we get results that neatly display the state and year of the data, followed by the average unemployment rate.

10. (Extra Credit) For each state, calculate the total unemployment rate across all counties (sum of all county rates).

```
db.HW5.aggregate([
  { $group: {
    _id: { state: "$State", year: "$Year" },
    totalRate: { $sum: "$Rate" }
  }
}]
)
```

```
db.HW5.aggregate([
  { $group: {
    _id: { state: "$State", year: "$Year" },
    totalRate: { $sum: "$Rate" }
  }
}]
)
{
  _id: {
    state: 'Alabama',
    year: 1997
  },
  totalRate: 5561.7
}
{
  _id: {
    state: 'Pennsylvania',
    year: 2010
  },
  totalRate: 7172.4
}
```

To sum up the unemployment rate in each county, I was able to do everything using \$group. First, I grouped by state and year, to ensure that we were not adding up all counties of every year. Then, I found the total rate using \$sum. The results were quite high considering there are a lot of counties with high unemployment rates during certain months.

11. (Extra Credit) The same as Query 10 but for states with data from 2015 onward

```
db.HW5.aggregate([
  { $match: { Year: { $gte: 2015 } } },
  { $group: {
    _id: { state: "$State", year: "$Year" },
    totalRate: { $sum: "$Rate" }
  }
}]
)
```

}
D)

```
> db.HW5.aggregate([
  { $match: { Year: { $gte: 2015 } } },
  { $group: {
    _id: { state: "$State", year: "$Year" },
    totalRate: { $sum: "$Rate" }
  }
}]
< {
  _id: {
    state: 'Hawaii',
    year: 2016
  },
  totalRate: 163.9
}
{
  _id: {
    state: 'Ohio',
    year: 2016
  },
  totalRate: 5712.8
}
```

To this, I added \$match first to only look for years greater than or equal to 2015. Everything else from Question 10 remained the same. From this, we were able to see the total unemployment rates for each state by year from 2015 onward.