DB Assignment 5
Katrina Cwiertniewicz
11/22/2024

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

This query groups by unique year and counts the number of years labeled Amount_of_Years.

```
> db.unemployment.aggregate([
    {
      $group: {
        _id: "$Year"                // Group by Year to get unique Years
      }
    },
    {
      $count: "Amount_of_Years"     // Counts the number of Years (distinct Years)
    }
  ])
< {
    Amount_of_Years: 27
  }
```

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

This query groups by unique state and counts the number of states labeled Unique_State_Count.

```
> db.unemployment.aggregate([
    {
      $group: {
        _id: "$State"               // Group by State to get unique States
      }
    },
    {
      $count: "Unique_State_Count"  // Count the number of States (distinct States)
    }
  ])
< {
    Unique_State_Count: 47
  }
```

3. What does this query compute?
db.unemployment.find({Rate : {$lt: 1.0}}).count()
This query finds all entries of unemployment rates lower than 1%.

```
> db.unemployment.find({Rate : {$lt: 1.0}}).count() //Count rates that unemployment rate is under 1.0%
< 657
```

4. Find all counties with unemployment rate higher than 10%
This query matches rates that are greater than 10% and groups by county.

```
> db.unemployment.aggregate([
    {
      $match: { Rate: { $gt: 10 } }  // Filter documents where unemployment rate > 10%
    },
    {
      $group: { _id: "$County" }  // Group by county to get unique counties
    }
  ])
< {
    _id: 'Alcona County'
  }
```

```
  _id: 'Glacier County'
}
{
  _id: 'Haskell County'
}
{
  _id: 'Clearwater County'
}
{
  _id: 'Slope County'
}
{
  _id: 'Grant Parish'
}
{
  _id: 'Tazewell County'
}
{
  _id: 'Ogemaw County'
}
{
  _id: 'Berkeley County'
}
{
  _id: 'Bell County'
}
{
  _id: 'Faribault County'
}
```

```
{
  _id: 'Kittitas County'
}
{
  _id: 'Neshoba County'
}
{
  _id: 'Craig County'
}
{
  _id: 'Abbeville County'
}
{
  _id: 'Somerset County'
}
{
  _id: 'Mohave County'
}
{
  _id: 'Door County'
}
{
  _id: 'Catron County'
}
{
  _id: 'Rush County'
}
Type "it" for more
```

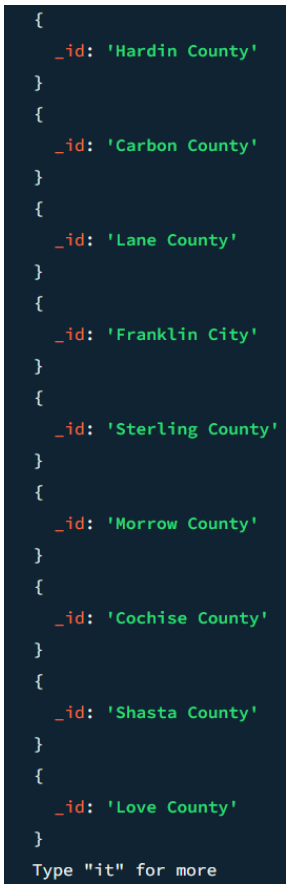
5. Calculate the average unemployment rate across all states.

This query projects the rate and groups the average unemployment rate. It rounds the average unemployment rate to two decimal places.

```
> db.unemployment.aggregate([
    {
      $project: {
        _id:0, Rate: {$ifNull: ["$Rate", 0]}          // Project Rate. If rate is null rate = 0
      }
    },
    {
      $group: {
        _id: null,
        average_unemployment_rate: {$avg: "$Rate"}    // Group by average unemployment rate
        }
    },
    {
      $project: {
        roundedAvg: {
          $round: ["$average_unemployment_rate",2]    // Round average unemployment rate
        }
      }
    }
  ])
< {
    _id: null,
    roundedAvg: 6.18
  }
```

6. Find all counties with an unemployment rate between 5% and 8%.
This query filters Rate between 5% and 8% and then groups by county.

```
> db.unemployment.aggregate([
    {
      $match: {
        Rate: { $gte: 5, $lte: 8 }  // Filter documents where unemployment rate is between 5% and 8%
      }
    },
    {
      $group: { _id: "$County" }  // Group by county to get unique counties
    }
  ])
< {
    _id: 'Trinity County'
  }
  {
    _id: 'Beckham County'
  }
```



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

This query displays State and Rate, sorts rate descending, and shows the top result using limit:1

```
> db.unemployment.aggregate([
    {
      $project: {
        State: 1, Rate: 1, _id: 0}      // Project State and Rate
    },
    {
      $sort: {Rate: -1}                 // Sort Rate descending
    },
    {
      $limit:1                          // Show top result
    }
  ])
< {
    State: 'Colorado',
    Rate: 58.4
```

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

This query filters rates above 5%, groups by county, and counts the number of counties with an unemployment rate above 5%.

```
> db.unemployment.aggregate([
    {
      $match: {
        Rate: {$gt:5}}            // Filter documents where unemployment rate is greater than 5%
    },
    {
      $group: {_id: "$County"}    // Group by county to get unique counties
    },
    {
      $count: 'County_Rates'      // Count counties with unemployment rates greater than 5%
    }
  ])

< {
    County_Rates: 1736
  }
```

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

This query displays State, Year, and Rate and groups by State, Year, and Average Unemployment Rate.

```
> db.unemployment.aggregate([
    {
      $project: {
        _id:0, State:1, Year:1, Rate: {$ifNull: ["$Rate", 0]}
      }
    },
    {
      $group: {
        _id: {state: "$State", year: "$Year"}, avg_rate: {$avg: "$Rate"}
      }
    }
  ])
< {
    _id: {
      state: 'North Dakota',
      year: 1993
    },
    avg_rate: 4.782704402515724
  }
  {
    _id: {
      state: 'Nebraska',
      year: 2013
    },
    avg_rate: 3.657437275985663
```

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

This query displays State, County, and Rate and groups sum_of_all_county_rates by State by averaging county rates and adding them together using sum.

```
> db.unemployment.aggregate([
    {
      $project: {
        _id:0, State:1, County:1,  Rate: {$ifNull: ["$Rate", 0]}                  // Project State, County and Rate. If rate is null rate=0
      }
    },
    {
      $group: {
        _id: {State: "$State"}, sum_of_all_county_rates: {$sum: {$avg: "$Rate"}}  //Average and Sum all unemployment rates, group by State
      }
    }
  ])
```

```
< {
    _id: {
      State: 'Montana'
    },
    sum_of_all_county_rates: 96261.5
  }
  {
    _id: {
      State: 'California'
    },
    sum_of_all_county_rates: 152661.6
  }
  {
    _id: {
      State: 'South Carolina'
    },
    sum_of_all_county_rates: 118915.1
  }
  {
    _id: {
      State: 'Vermont'
    },
    sum_of_all_county_rates: 22427.6
  }
  {
    _id: {
      State: 'Delaware'
    },
    sum_of_all_county_rates: 4899.9
```

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

This query displays State, County, and Rate, filters by Year >= 2015 and groups sum_of_all_county_rates by State by averaging county rates and adding them together using sum.

```
> db.unemployment.aggregate([
    {
      $project: {
        _id:0, State:1, County:1, Year:1,  Rate: {$ifNull: ["$Rate", 0]}           // Project State, County, Year, and Rate. If rate is null rate=0
      }
    },
    {
      $match: {
        Year: {$gte: 2015}                                                        // Filter Years that are >= 2015
      }
    },
    {
      $group: {
        _id: {State: "$State"}, sum_of_all_county_rates: {$sum: {$avg: "$Rate"}}   //Average and Sum all unemployment rates, group by State
      }
    }
  ])
```


```
< {
    _id: {
      State: 'South Carolina'
    },
    sum_of_all_county_rates: 7167.7
  }
  {
    _id: {
      State: 'Oklahoma'
    },
    sum_of_all_county_rates: 9286.7
  }
  {
    _id: {
      State: 'Rhode Island'
    },
    sum_of_all_county_rates: 628.4
  }
  {
    _id: {
      State: 'California'
    },
    sum_of_all_county_rates: 9679.2
  }
```