Normalization:

The initial schema is as follows:

Nation(State, TotalCases, TotalDeaths, TotalTests, TestsRatio, DeathRatio)
Population(State, NumOfPeople, Capita, CaseDensity, DeathDensity, TestsDensity)
Infectionrate(State, StartRate, EndRate)
Workforce(State, Workers, Unemployment, Lockdown)

In order to normalize the data, the functional dependencies and multivalued dependencies must be evaluated. The functional dependencies for this schema is as follows:

Functional Dependencies:

- State -> (TotalCases, TotalDeaths, TestsRatio, DeathRatio)
- (TotalCases, TotalTests) -> TestsRatio
- (TotalCases, TotalDeaths) -> DeathRatio
- (TotalCases, TotalDeaths, TotalTests) -> (TestsRatio, DeathRatio)
- State -> (Population, density, CaseDensity, DeathDensity, TestsDensity)
- State -> (StartRate, EndRatee)
- State -> (Workforce, Unemployment, Lockdown)

There are no multivalued dependencies in this schema because swapping between tuples will drastically manipulate the data because each state yields specific data based on unique environmental, political, and societal conditions.

In order to decompose the relations into BCNF, it is important to understand the definition of BCNF. There are three requirements that must be satisfied in order to normalize a relation into BCNF:

1. The table must be in 2NF. This means that each tuple in the relation must contain unique data so that it can be identified in the table and each non-key attribute must be functionally dependent on the composite key.
2. The table must be in 3NF. This means that not only does it satisfy 2NF, but it also satisfies a requirement in which no non-key attribute determines other non-key attributes.
3. Every determinant in the relation must be a candidate key. A candidate key is an attribute that determines the rest of the values in the relation.

The **Population** relation, **Infectionrate** relation, and **Workforce** relation are in BCNF because the only determinants within this relation is the candidate key. The **Nation** relation is not in BCNF. Therefore, in order to normalize the data, we must recursively decompose the tables into separate tables which separate the offending attributes until there are only 2 attributes remaining in the table or if the table has finally been normalized into BCNF. We do not need to decompose the **Population, Infectionrate**, and **Workforce** relations because they are already in BCNF. Therefore, we only need to decompose the **Nation** relation.

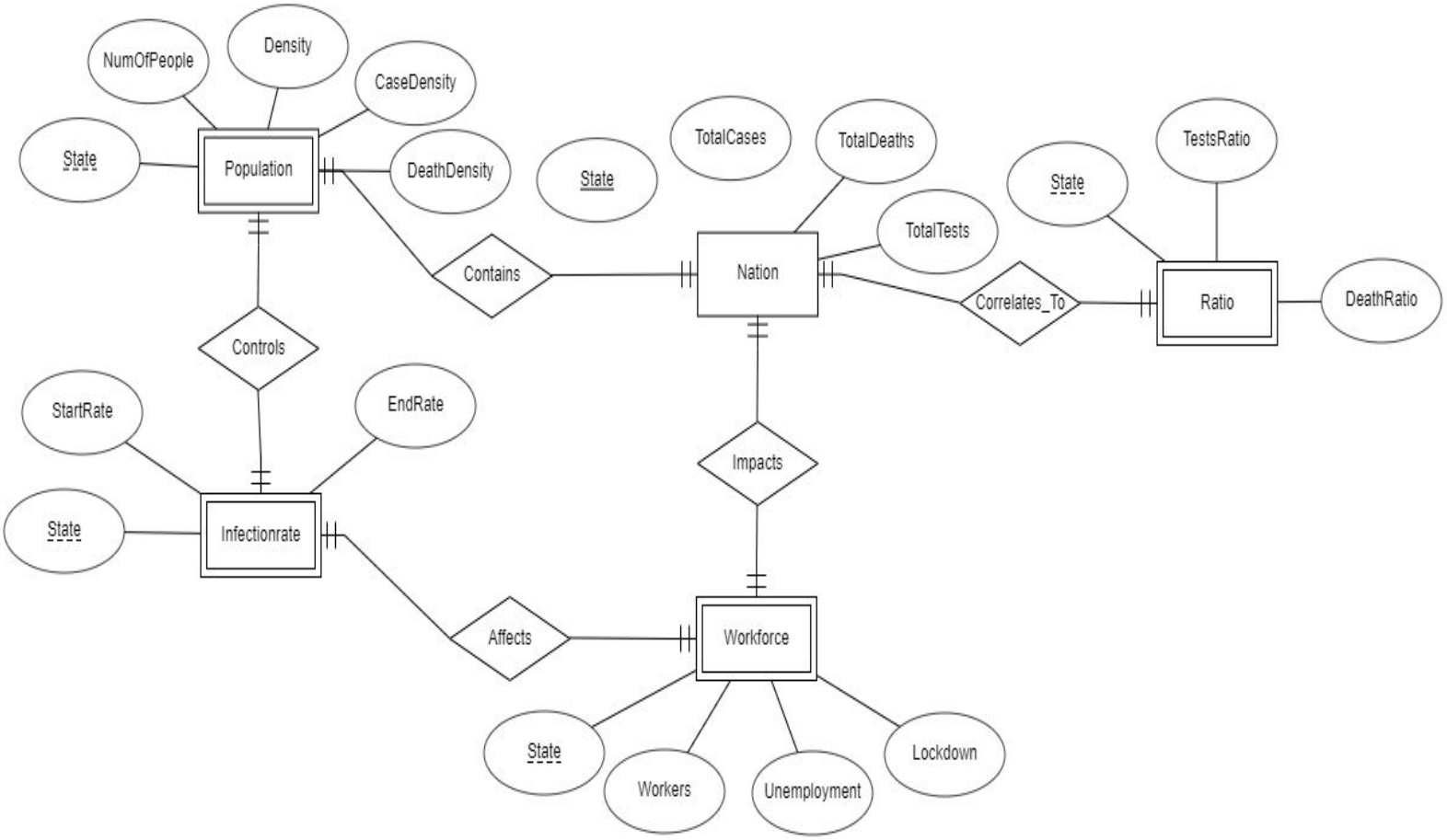
Decompose **Statistics** on (TotalCases, TotalDeaths, TotalTests) -> (TestsRatio, DeathRatio)

> **Nation(State, TotalCases, TotalDeaths, TotalTests)**
> **Ratio(State, TestsRatio, DeathRatio)**
>
> The superkey is State in both tables.
>
> The decomposition is over as both tables are now in BCNF. the only determinant in these relations is the State attribute which is the candidate key.

Since there are no multivalued dependencies, because this schema is now normalized into BCNF, it must be in 4NF as well.
As such, the final schema is shown below:

> **Nation(<u>State</u>, TotalCases, TotalDeaths, TotalTests)**
> **Ratio(<u>State</u>, TestsRatio, DeathRatio)**
> **Population(<u>State</u>, NumOfPeople, Capita, CaseDensity, DeathDensity, TestsDensity)**
> **Infectionrate(<u>State</u>, StartRate, EndRate)**
> **Workforce(<u>State</u>, Workforce, Unemployment, Lockdown)**

ER Diagram:


NumOfPeople
Density
CaseDensity
State
Population
DeathDensity
State
TotalCases
TotalDeaths
TestsRatio
State
TotalTests
Contains
Nation
Correlates_To
Ratio
DeathRatio
Controls
StartRate
EndRate
Impacts
State
Infectionrate
Affects
Workforce
State
Workers
Unemployment
Lockdown