Kimone Walker

Overview:

Originally I worked with a data set on chronic disease indicators
https://www.kaggle.com/datasets/cdc/chronic-disease because I wanted to track the
relationships between different chronic diseases and their symptoms. However, after cleaning
and preparing the dataset I created the graph but the relationships between nodes and edges
were not compatible with the tests I wanted to run. I decided to switch to a data set on the
prevalence of different vaccine preventable diseases
https://www.kaggle.com/datasets/rishidamarla/vaccine-preventable-diseases based on county
and year.

I tested six degrees of separation between Diphtheria and Alameda which was 0. This makes
sense since Diphtheria was a preventable disease that occurred in Alameda, if this wasn't the
case then the degree of separation would be a non 0 integer. The average distance between
pairs of vertices is 2.25. The diseases and counties both appear as nodes in the graph and the
edges represent the prevalence of diseases within the connected county. For future use graphs
like this can be used to track the spread of a preventable disease overtime and which ones
pose the highest risk of being contractible

"dot -Tpng graph.dot -o graph.png" in the terminal creates a visual representation of the graph

```
    Finished test [unoptimized + debuginfo] target(s) in 0.85s
     Running unittests src/main.rs (target/debug/deps/eTest-9c8b61207820fb82)

running 3 tests
test tests::test_shortest_path_length ... ok
test tests::test_degrees_of_separation ... ok
test tests::test_specific_connections ... ok

test result: ok. 3 passed; 0 failed; 0 ignored; 0 measured; 0 filtered out; finished in 0.
00s

kw@crc-dot1x-nat-10-239-134-110 mainproject %
```

```
County:  San Mateo  - Min connections to diseases: 141
 Disease: "Hepatitis C, Acute" - Min connections to counties: 472
 Disease: "Hepatitis B, Acute" - Min connections to counties: 472
 Disease: "Hepatitis A" - Min connections to counties: 472
 Disease: "Pertussis" - Min connections to counties: 1062
 Disease: "Rubella" - Min connections to counties: 531
 Disease: "Varicella Hospitalizations" - Min connections to counties: 531
 Disease: "Invasive Meningococcal Disease" - Min connections to counties: 1062
 Disease: "Measles" - Min connections to counties: 1062
 Disease: "Diphtheria" - Min connections to counties: 1062
 Disease: "Mumps" - Min connections to counties: 531
 Disease: "Tetanus" - Min connections to counties: 1062
 The distribution fits a power-law
 Degrees of separation between Diphtheria and Alameda: 0
 kw@crc-dot1x-nat-10-239-134-110 mainproject %
```

```
     Running   target/debug/eTest
Average distance between pairs of vertices: 2.2526964560862868
Number of pairs: 649
Number of edges: 8319
County: "Glenn" - Min connections to diseases: 141
County: "Marin" - Min connections to diseases: 141
County: "Kings" - Min connections to diseases: 141
County: "Santa Barbara" - Min connections to diseases: 141
County: "Yolo" - Min connections to diseases: 141
County: "Sutter" - Min connections to diseases: 141
```