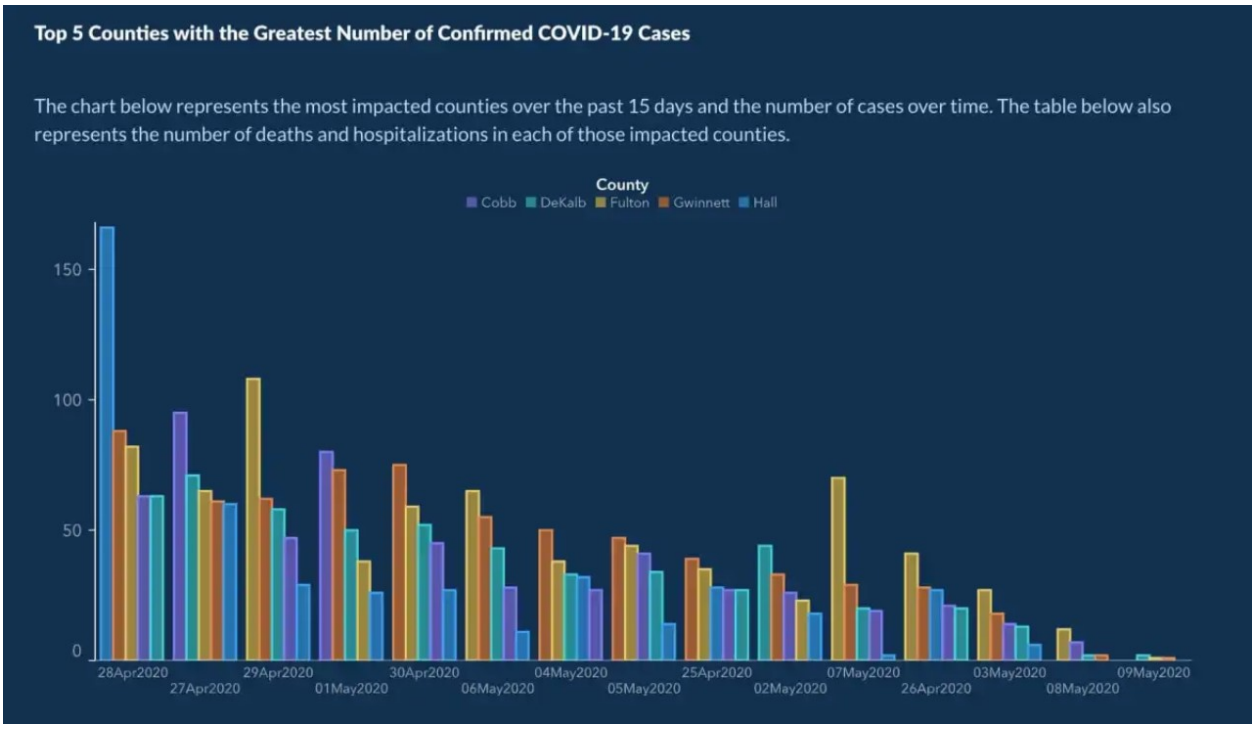
DATA 202 Midterm step 1

Adham, Aubrey, Naomi

2022-10-14

My Data set



https://www.covid-georgia.com/archive/did-georgia-graph-cases-with-the-dates-out-of-order/

```
potential_data <- read.csv("~/data202/Mid 1/owid-covid-data (1).csv")
glimpse(potential_data)
```

```
Rows: 189,999
Columns: 34
$ iso_code                  <chr> "AFG", "AFG", "AFG", "AFG", "AFG", "AFG", "~
$ continent                 <chr> "Asia", "Asia", "Asia", "Asia", "Asia", "As~
$ location                  <chr> "Afghanistan", "Afghanistan", "Afghanistan"~
$ date                      <chr> "24-02-2020", "25-02-2020", "26-02-2020", "~
$ total_cases               <int> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 8, 8, 8~
$ new_cases                 <int> 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0~
$ new_cases_smoothed        <dbl> 0.000, 0.000, 0.000, 0.000, 0.000, 0.714, 0~
$ total_deaths              <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ new_deaths                <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ new_deaths_smoothed       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ total_cases_per_million   <dbl> 0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0~
```

```
$ new_cases_per_million      <dbl> 0.125, 0.000, 0.000, 0.000, 0.000, 0.000, 0~
$ total_deaths_per_million   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ reproduction_rate          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ hosp_patients              <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ total_tests                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ positive_rate              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ tests_units                <chr> "0", "0", "0", "0", "0", "0", "0", "0", "0"~
$ total_vaccinations         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ people_vaccinated          <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ people_fully_vaccinated    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ total_boosters             <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ new_vaccinations           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ stringency_index           <dbl> 8.33, 8.33, 8.33, 8.33, 8.33, 8.33, 27.78, ~
$ population                 <dbl> 40099462, 40099462, 40099462, 40099462, 400~
$ population_density         <dbl> 54.422, 54.422, 54.422, 54.422, 54.422, 54.~
$ aged_65_older              <dbl> 2.581, 2.581, 2.581, 2.581, 2.581, 2.581, 2~
$ aged_70_older              <dbl> 1.337, 1.337, 1.337, 1.337, 1.337, 1.337, 1~
$ gdp_per_capita             <dbl> 1803.987, 1803.987, 1803.987, 1803.987, 180~
$ cardiovasc_death_rate      <dbl> 597.029, 597.029, 597.029, 597.029, 597.029~
$ diabetes_prevalence        <dbl> 9.59, 9.59, 9.59, 9.59, 9.59, 9.59, 9.59, 9~
$ hospital_beds_per_thousand <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5~
$ life_expectancy            <dbl> 64.83, 64.83, 64.83, 64.83, 64.83, 64.83, 6~
$ human_development_index    <dbl> 0.511, 0.511, 0.511, 0.511, 0.511, 0.511, 0~
```

Claim

Does Covid-19 differ in infection rates from country to country? (Egypt vs Jordan vs Israel?)

Structure of data

This data consists of plenty of numerical and categorical variables (34 to be precise). The data set also consists of plenty of observations which amount to 189999!

- I do not intend to use all the variables in my graph but they are very valuable and understandable!*