

# **Nigeria COVID-19 Data Analysis Using Python.**

Ustacky Data Scientist Microdegree Capstone Project

**Khairat Ayinde**

# Introduction

- Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus, and it has affected major parts of the world. Nigeria, a West-African country, has also been affected by the COVID-19 pandemic after recording its first case on 27th February 2020.
- In this project, data science & analytics skills are being employed to collect data, explore the data, perform analysis, create visualizations, and generate insights.

# Data Overview

- Different data was collected from different sources and combined to perform analysis and provide insights.
  - i. Data was scraped from the NCDC (The Nigeria Centre for Diseases Control) COVID-19 official website; they monitor the country's COVID-19 situation, and releases data on metrics across all the 37 states in the country
  - ii. The Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) publishes daily data on confirmed, death and recovered cases across different countries. Data is read in using Pandas from its GitHub repository

iii. Nigeria Community

Vulnerability Index data, Real Domestic Gross Product Data and State Budget Data is downloaded and read in from current working directory

iv. Another external data from Our World in Data website for COVID cases in the world is downloaded and read in also.

# Methods

- STEP 1: The first step was the collection of the various listed data, then importing for analysis.
- STEP 2: View the Data - To get the basic information about all the data, the `head()`, `shape` and `info()` method were obtained. All data showed no missing value and were in their appropriate datatype except the NCDC scraped data.

- STEPP 3: Data Cleaning & Preparation - From the information obtained viewing the datasets, there appeared a need to fix the data format of the scraped data. The numerical columns were converted to appropriate data type; integer and comma(,) removed. Columns of the scraped data were also renamed. The daily data for Nigeria was extracted from the Global daily cases data.
- STEP 4: Data Analysis - Some analysis on the datasets are performed and findings communicated in charts and summarised.

# Analysis Questions

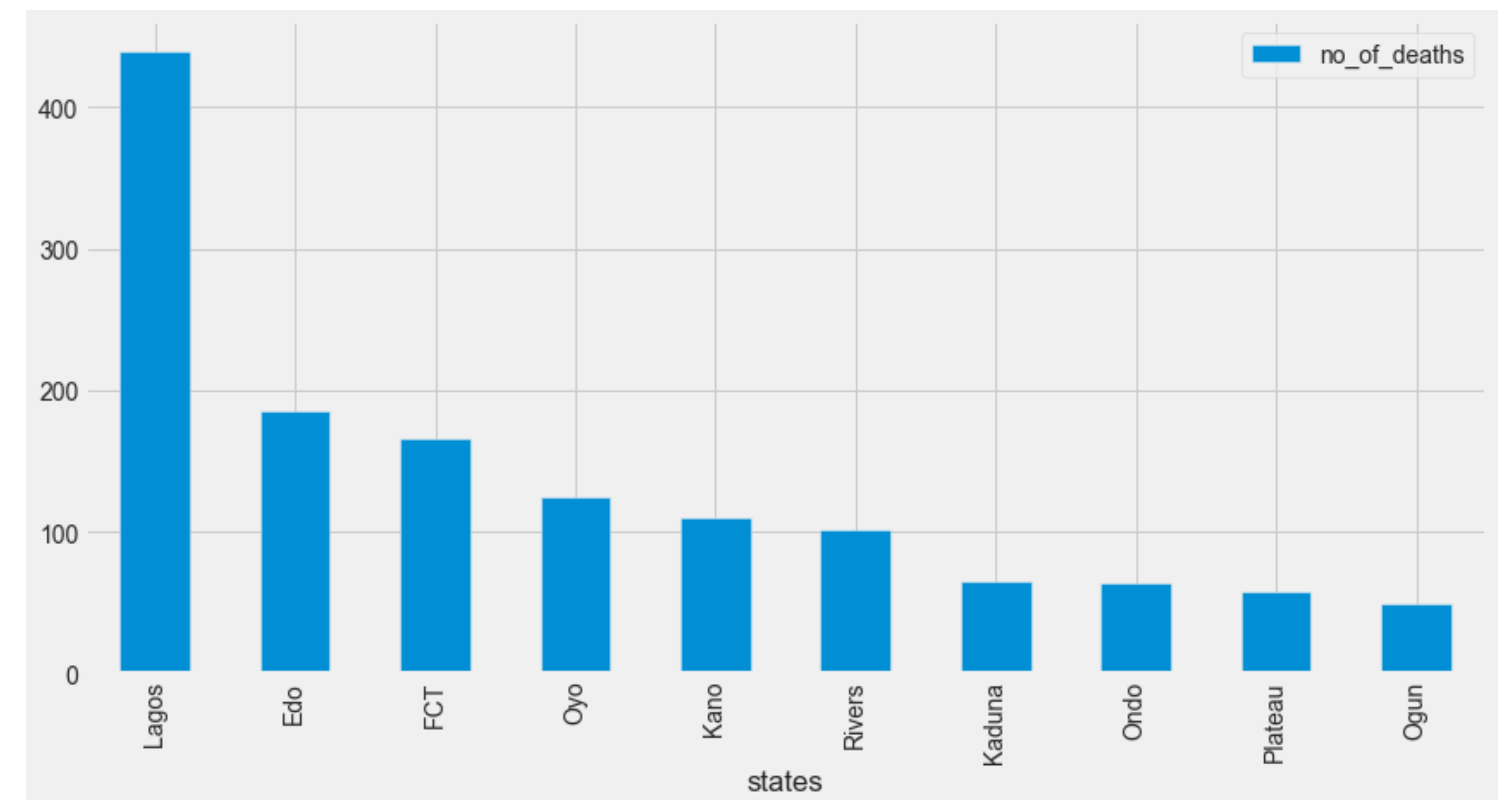
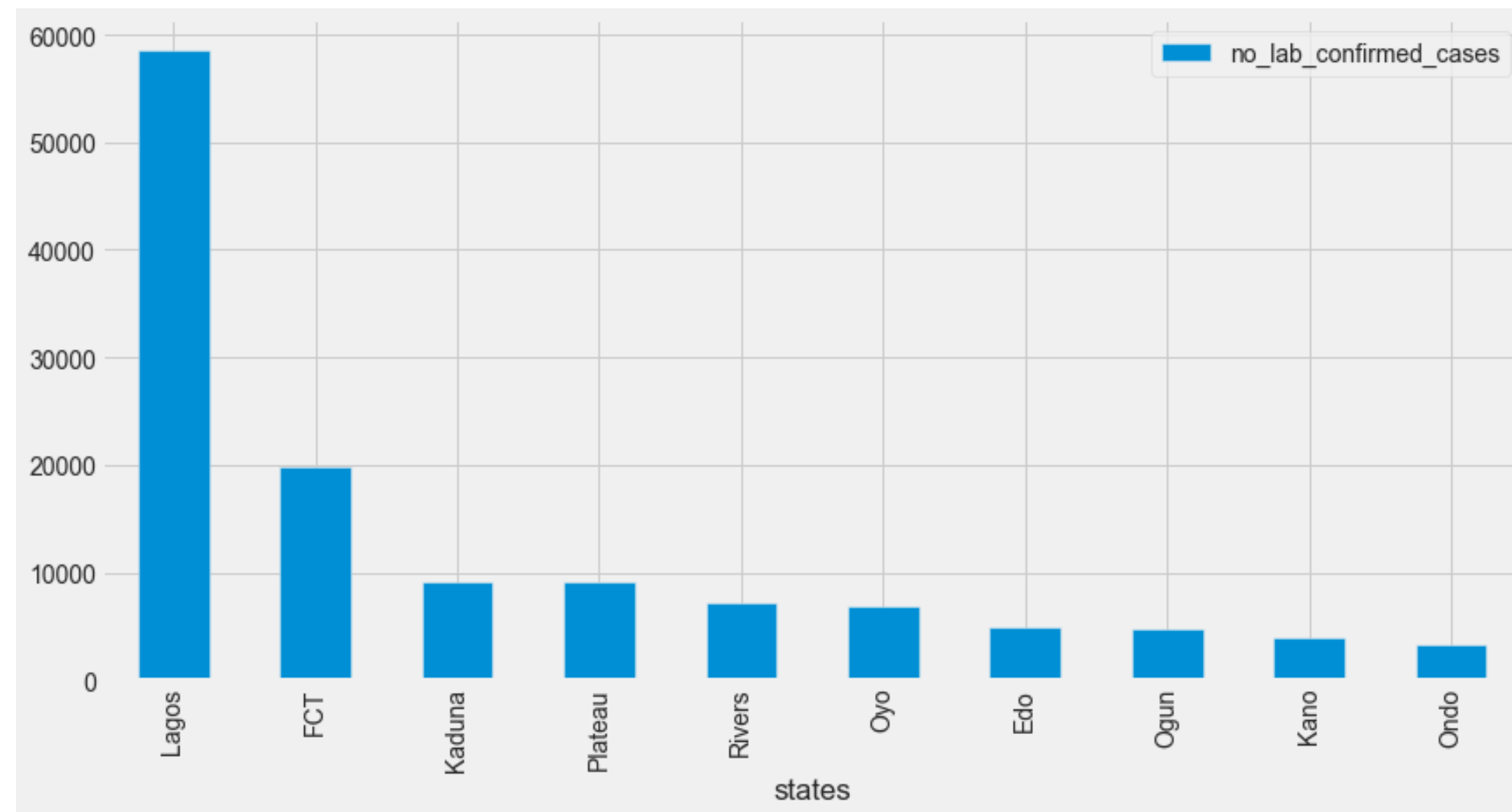
- The top 10 states in terms of confirmed Covid cases by Laboratory test, Discharged Covid cases, cases on admission and Death cases
- The total daily confirmed, recovered and death cases in Nigeria
- The maximum infection rate for a day in Nigeria and the date
- Relationship between the top 10 confirmed cases and the overall community vulnerability index.
- The linear relationship between Confirmed Cases and Population Density
- The GDP values for each year & quarters

- Summary of the initial and revised budgets
- The number of lab confirmed cases and the number of death according to their regions
- Total cases for years 2020 and 2021 in Nigeria
- The total death per month and new cases per day.
- The number of people vaccinated each month
- Record of the total tests of both years in Nigeria etc.



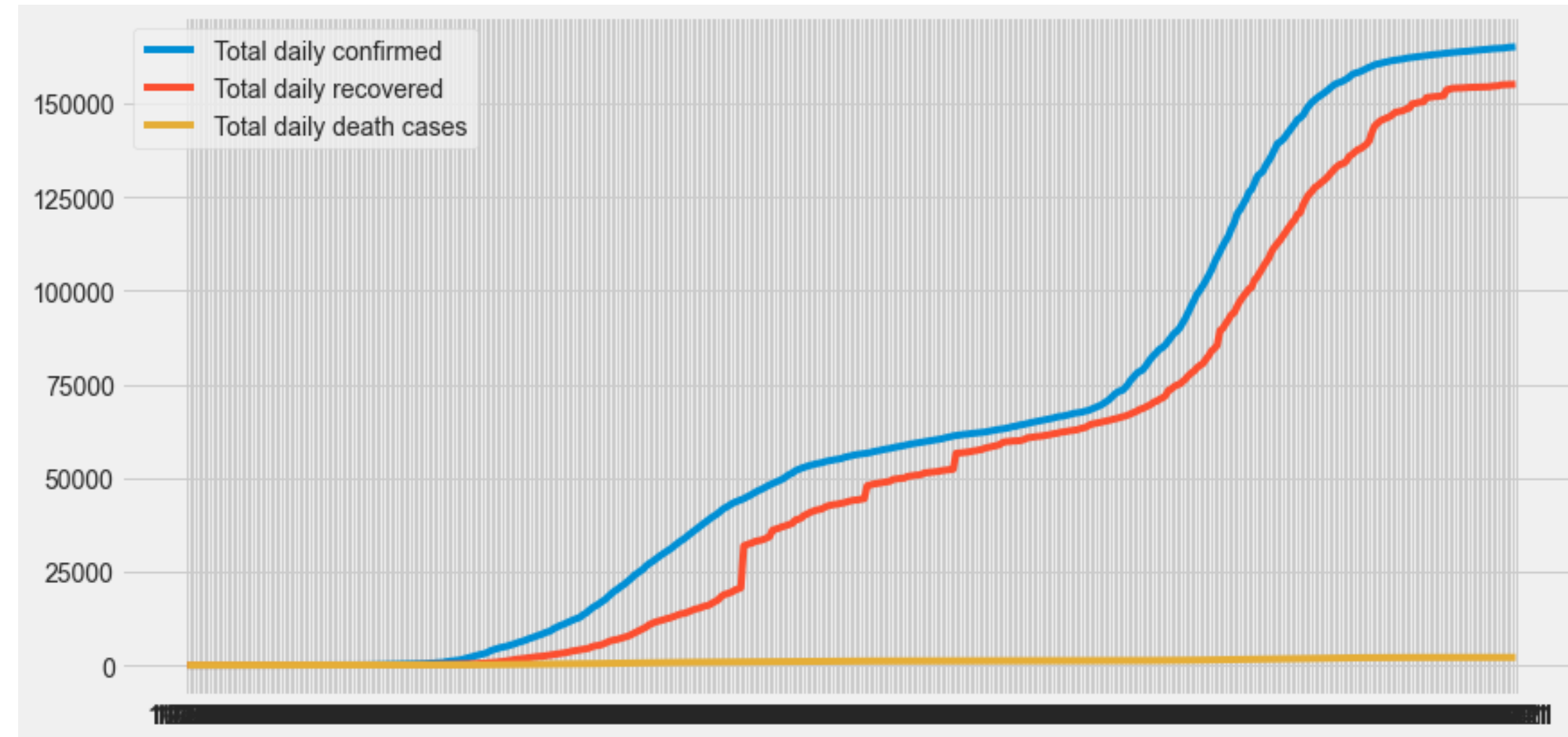
# Insights

- Lagos is the top state with confirmed COVID cases by laboratory test. The top 10 states are shown in the graph below.

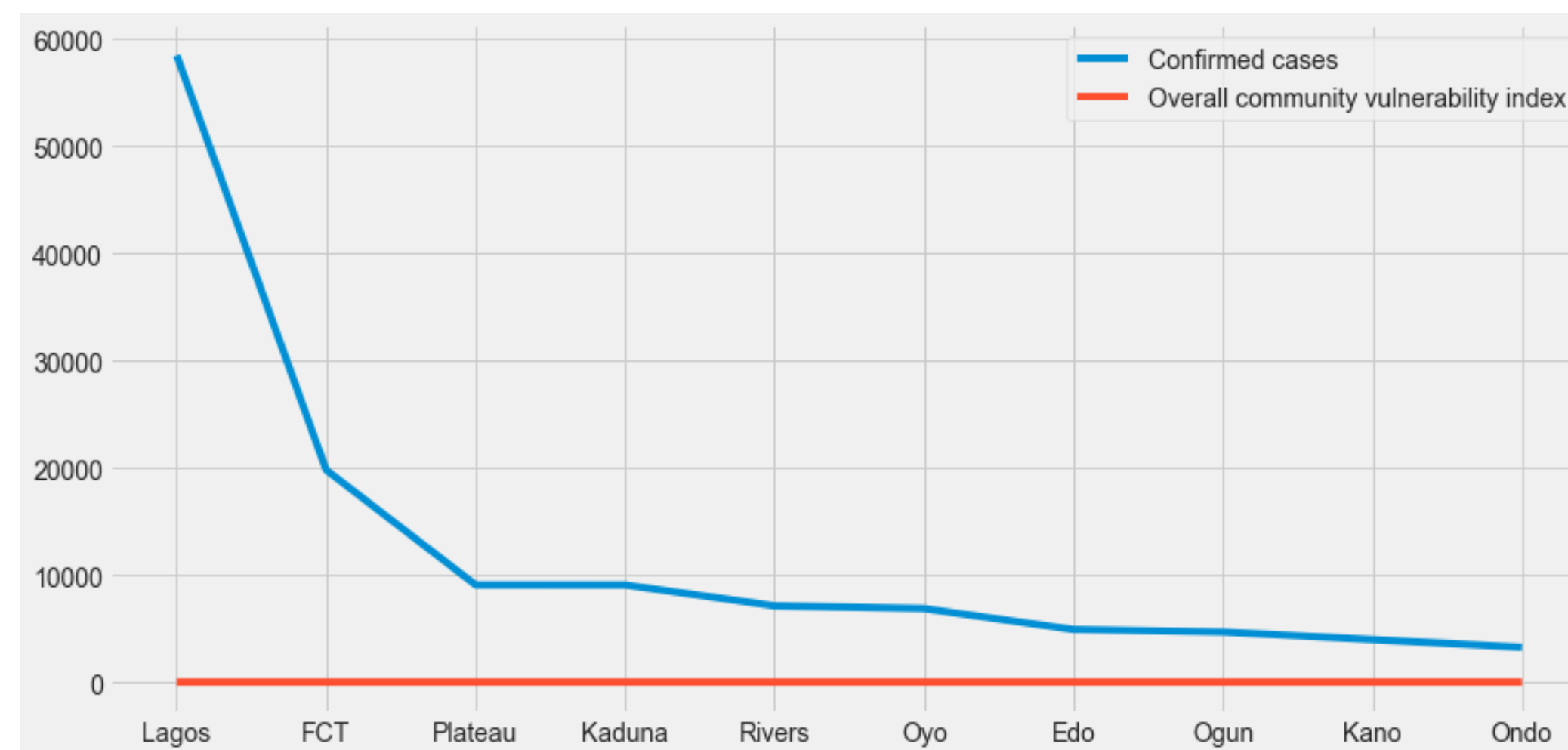


- Lagos, Edo, FCT, Oyo, Kano, Rivers, Kaduna, Ondo, Plateau and Ogun are the top 10 states with recorded death cases. the chart above shows the number of each state.

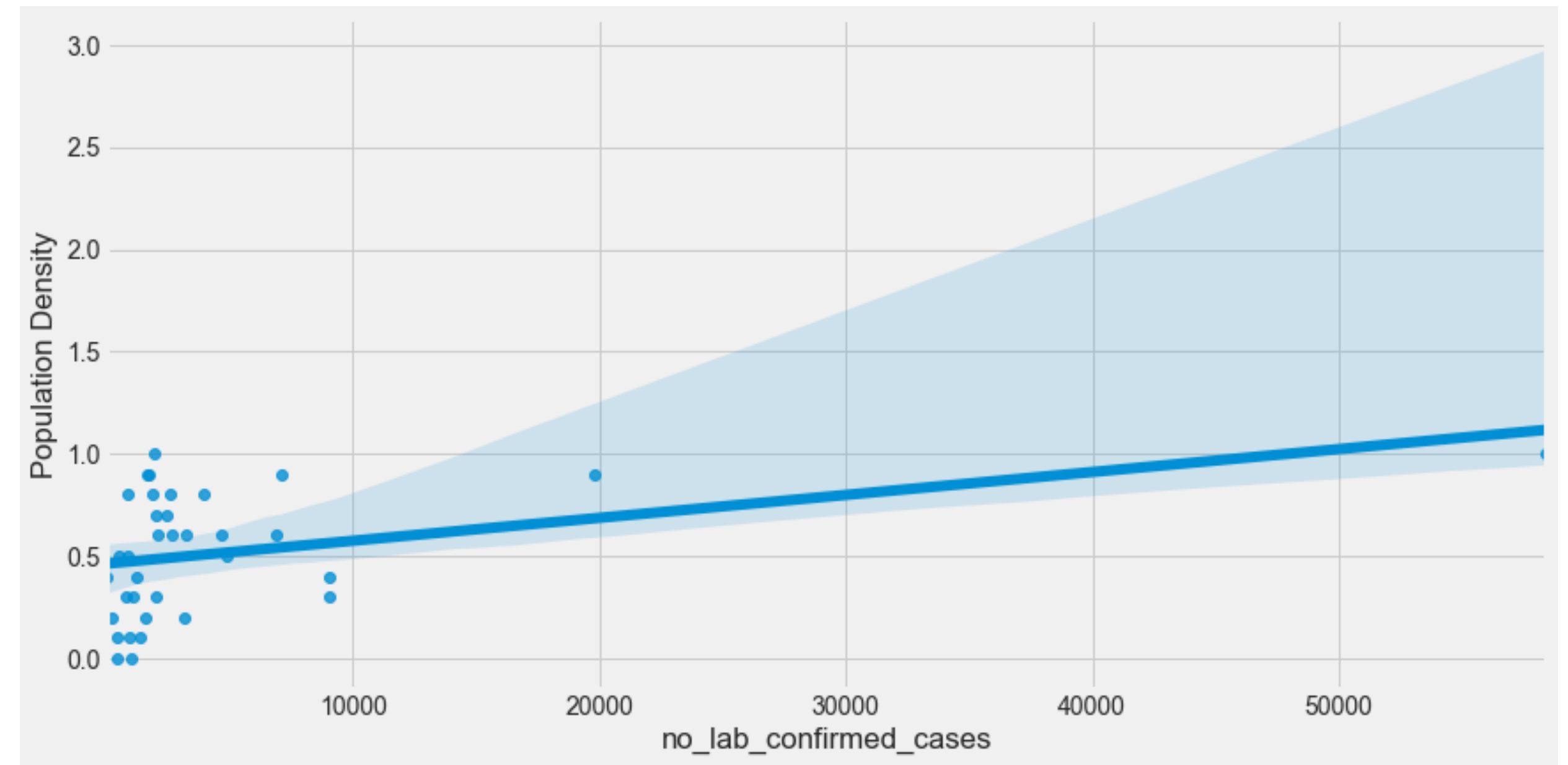
- Line plot for the total daily confirmed, recovered and death cases in Nigeria



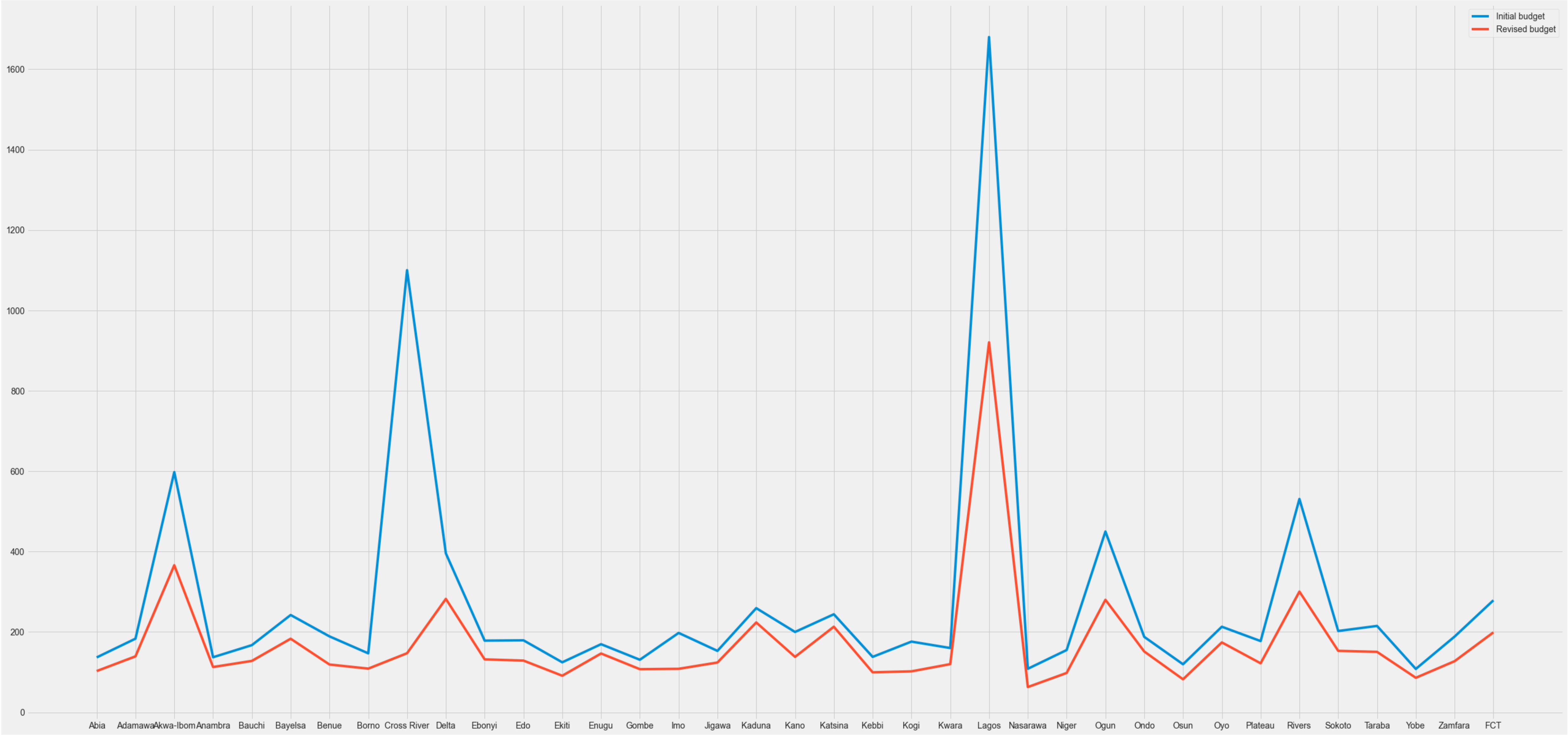
- The vulnerability index of each state is a little above 0. It is not affected by the number of confirmed cases.



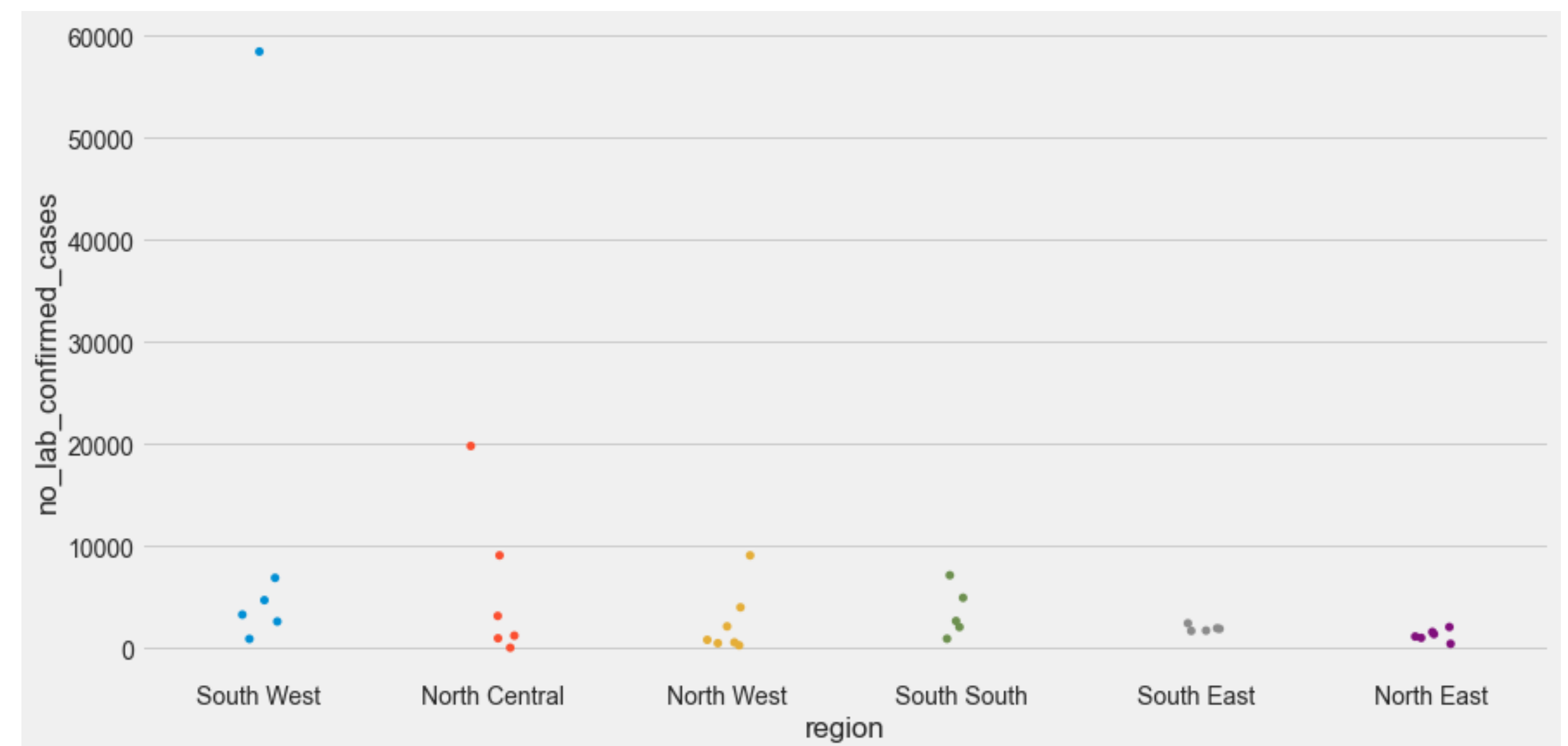
- The relationship between the Population density and confirmed cases from the graph doesn't follow a linear relation and is clustered when Confirmed cases is below 10000 but very disperse above 10000 cases.



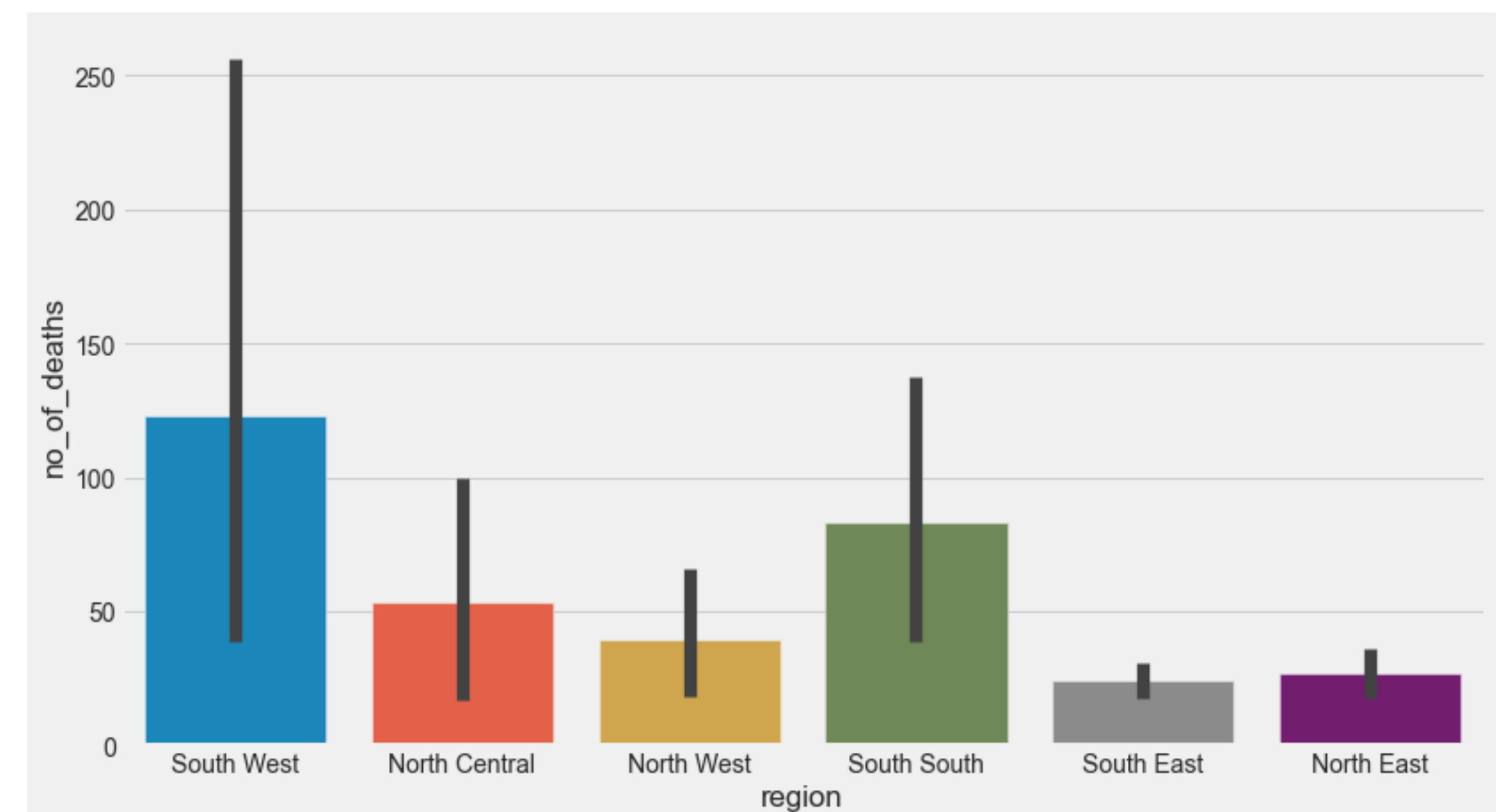
- Lagos has the highest initial budget of over 1600.



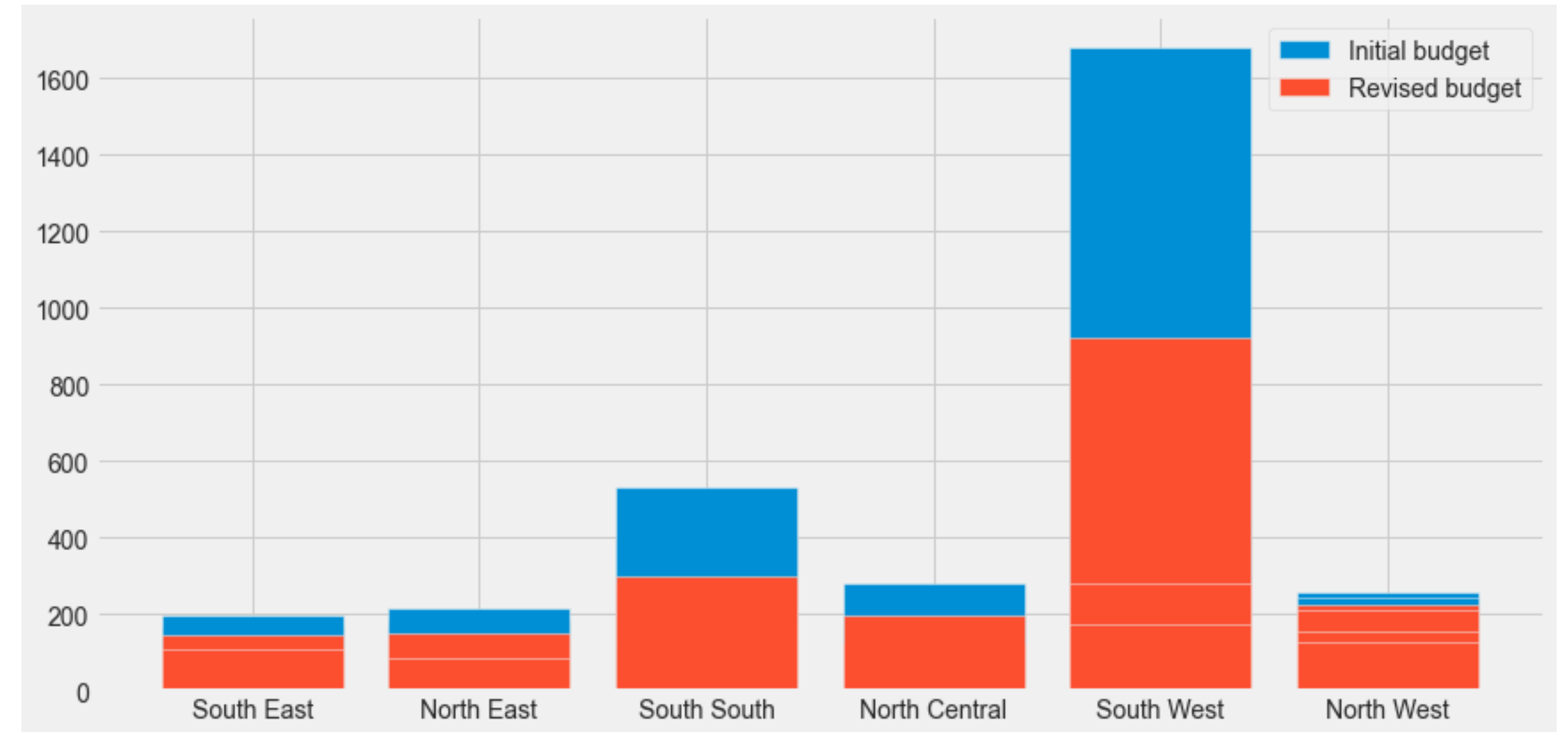
- South west region has the highest number of laboratory confirmed cases.



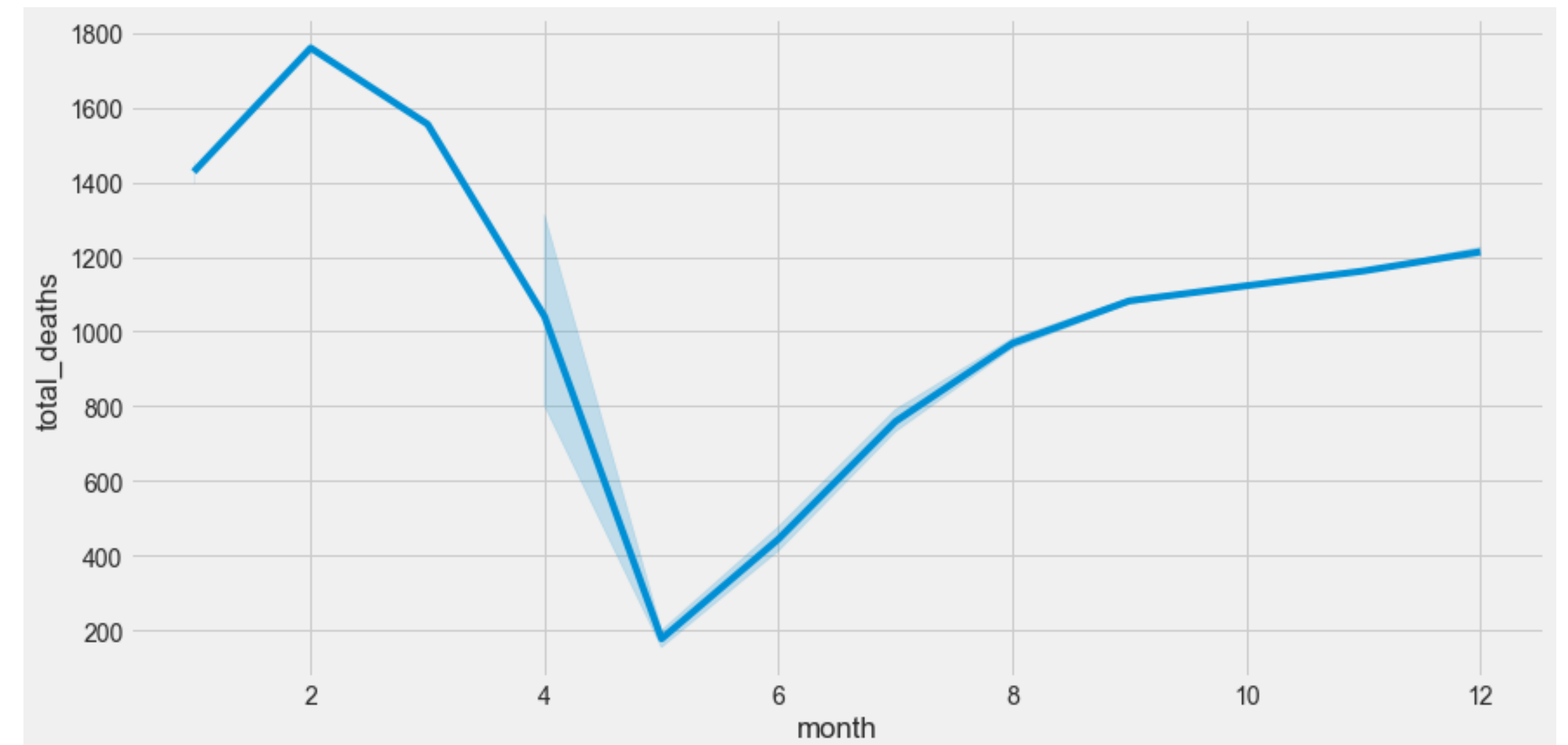
- South west also has the highest death cases with over 400 deaths, followed by the South south, North central, North west, North east and then South east regions.



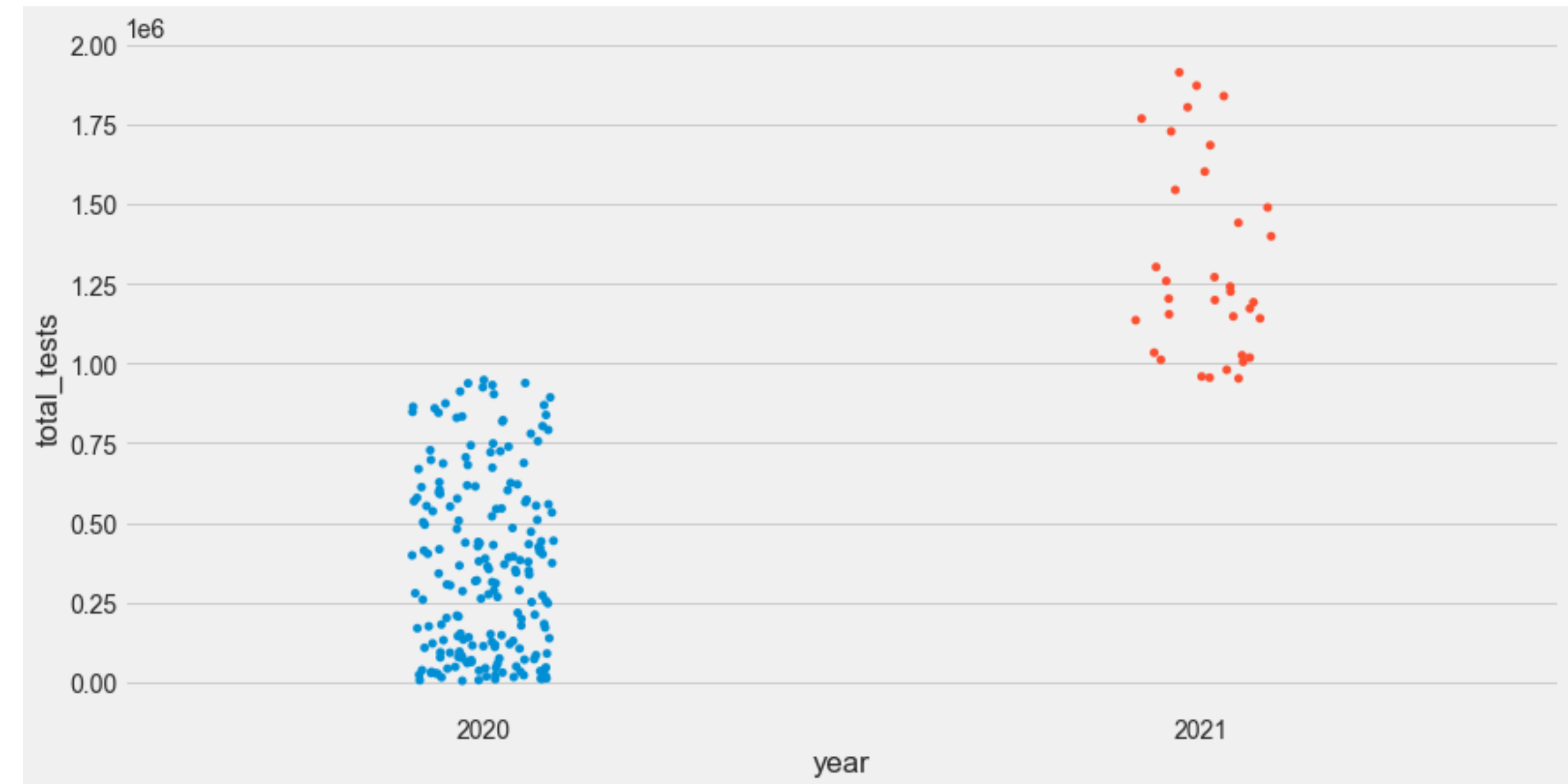
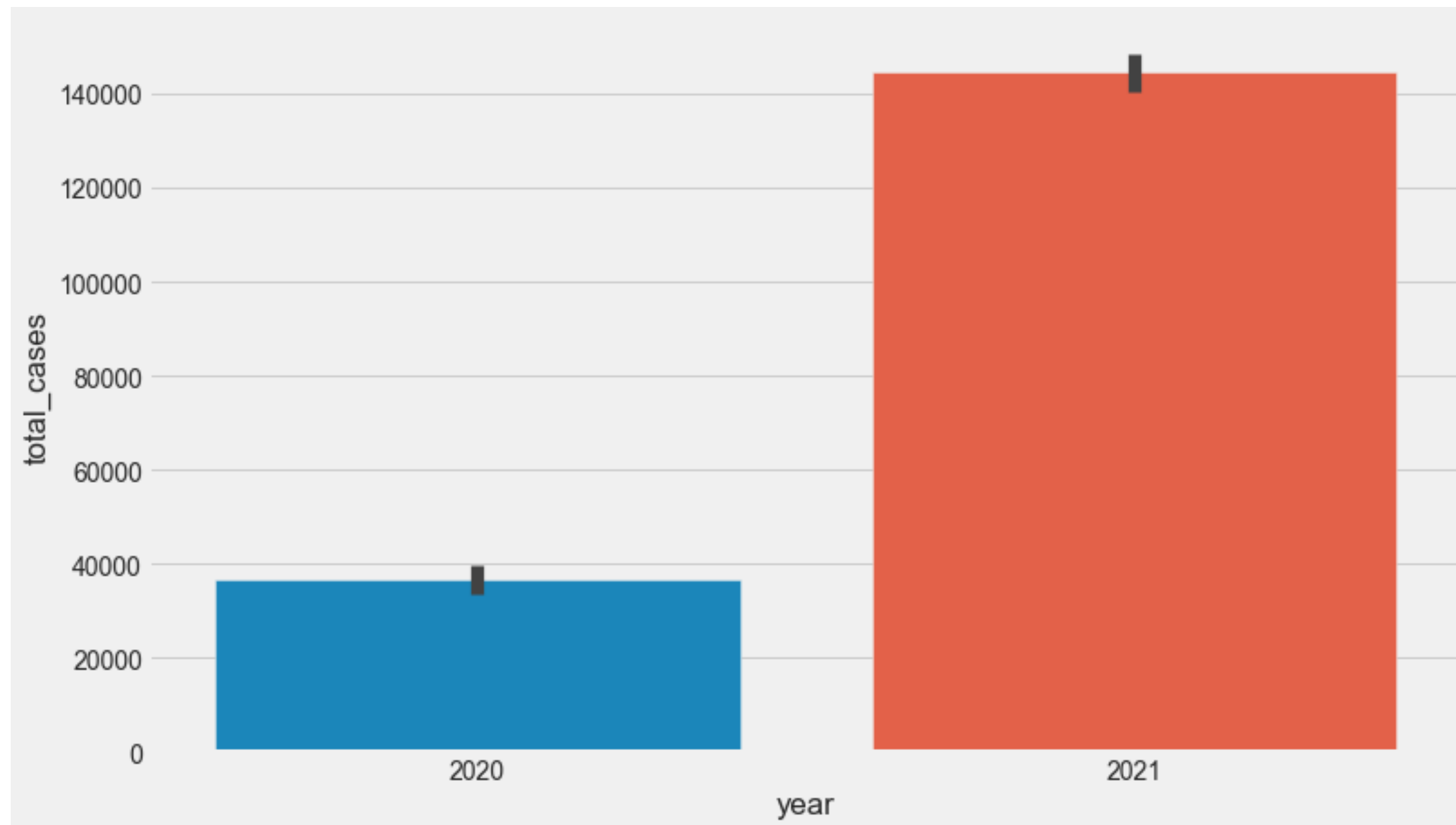
- Barplot of the total initial and revised budget of each region



- The most recorded death cases is in February



- 2021 has the highest total cases of 140000, while 2020 with over 20000. More COVID tests were also performed in 2021



# Suggestions on future work/Improvement plans

- Importation of more COVID datasets for better comparison and more diverse analysis.
- Also, building a machine learning model to predict rate of new cases, death etc. in Nigeria.



## **Information about OWID external data -**

The external data downloaded from Our World in Data holds the record of Corona virus pandemic country profile updated daily on this site.

website: <https://ourworldindata.org/coronavirus>

Thank You.