يوسف خميس محمد محمد علي 23011637

يوسف طارق محمد حسن 23011641

هادي شريف محمد نجاتي 23010157

ساجد سامح عبدالله 23011278

يوسف محمد قباري ابراهيم متولي 23011656

COVID-19 Data Analysis: Global Insights and Visualizations

A data-driven exploration of the COVID-19 pandemic. We will uncover key trends, correlations, and insights from global data. This presentation includes analysis of cases, deaths, and recovery trends, visualized through various charts and maps.

by Youssef Khamis

Data Acquisition: Web Scraping

Data Source

Data was scraped from Worldometer using Python and BeautifulSoup.

- Cases
- Deaths
- Recovered
- Population
- Tests

Date Range

The date range spans from January 2020 to the present.

Code Example

```
soup.find('table',
id='main_table_countries_today')
```

This code snippet demonstrates how the data was extracted. Web scraping provided a comprehensive dataset for analysis.

Data Cleaning and Preparation

1. **Missing Values**

   Missing values were handled through imputation and removal.

2. **Duplicates**

   Duplicates and inconsistencies were removed to ensure data integrity.

3. **Active Cases**

   Active cases were calculated as Cases - Deaths - Recovered.

4. **Data Types**

   Data types were converted to numeric and datetime formats.

Cleaned data was stored in MongoDB using `db.covid_data.insert_many(data)`. This prepared the data for effective analysis and visualization.

Exploratory Data Analysis: Recovery Rates

Calculation

Recovery Rate = Recovered / Cases

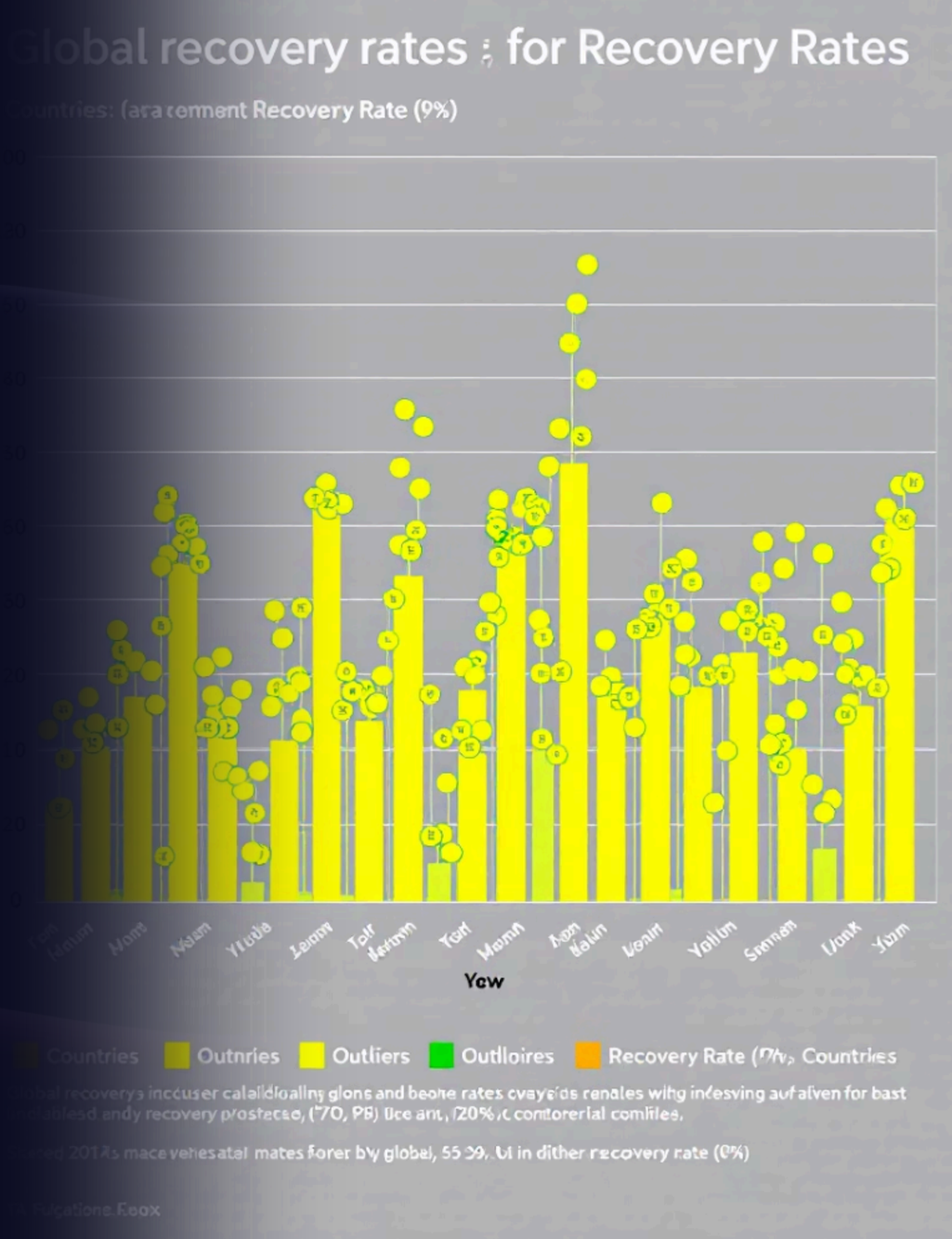
Global Rate

~97% as of October 2024

Top Countries

Singapore (99.8%) is a leader in recovery rates.

Recovery rates provide insights into healthcare effectiveness. A boxplot visualizes the distribution across countries.



EDA: Death Rates

Calculation

Death Rate = Deaths / Cases

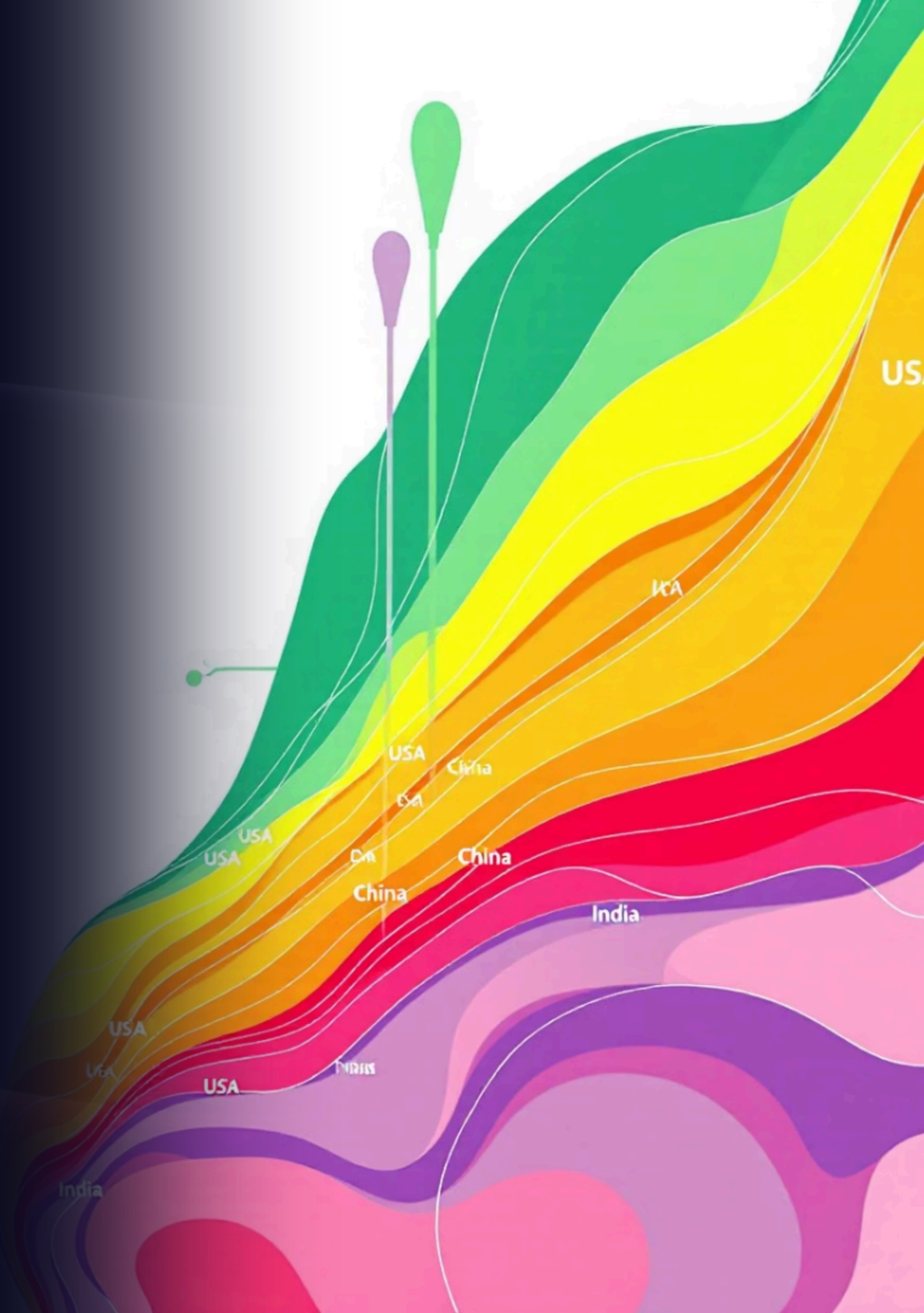
Global Rate

~2% as of October 2024

Top Countries

Yemen (18.2%) had significantly higher death rates.

Death rates are influenced by healthcare infrastructure. A violin plot illustrates the distribution of death rates globally.



EDA: Correlations

Cases vs Deaths

r = 0.85 (strong positive correlation)

Tests vs Cases

Increased testing correlates with higher case detection.

Correlation analysis reveals relationships between variables. The heatmap visualizes the correlation matrix. Scatterplot shows the relationship of tests vs cases on a logarithmic scale.

Visualizations: Global Cases and Deaths

1

Line Chart

Global cases and deaths over time

2

Spikes

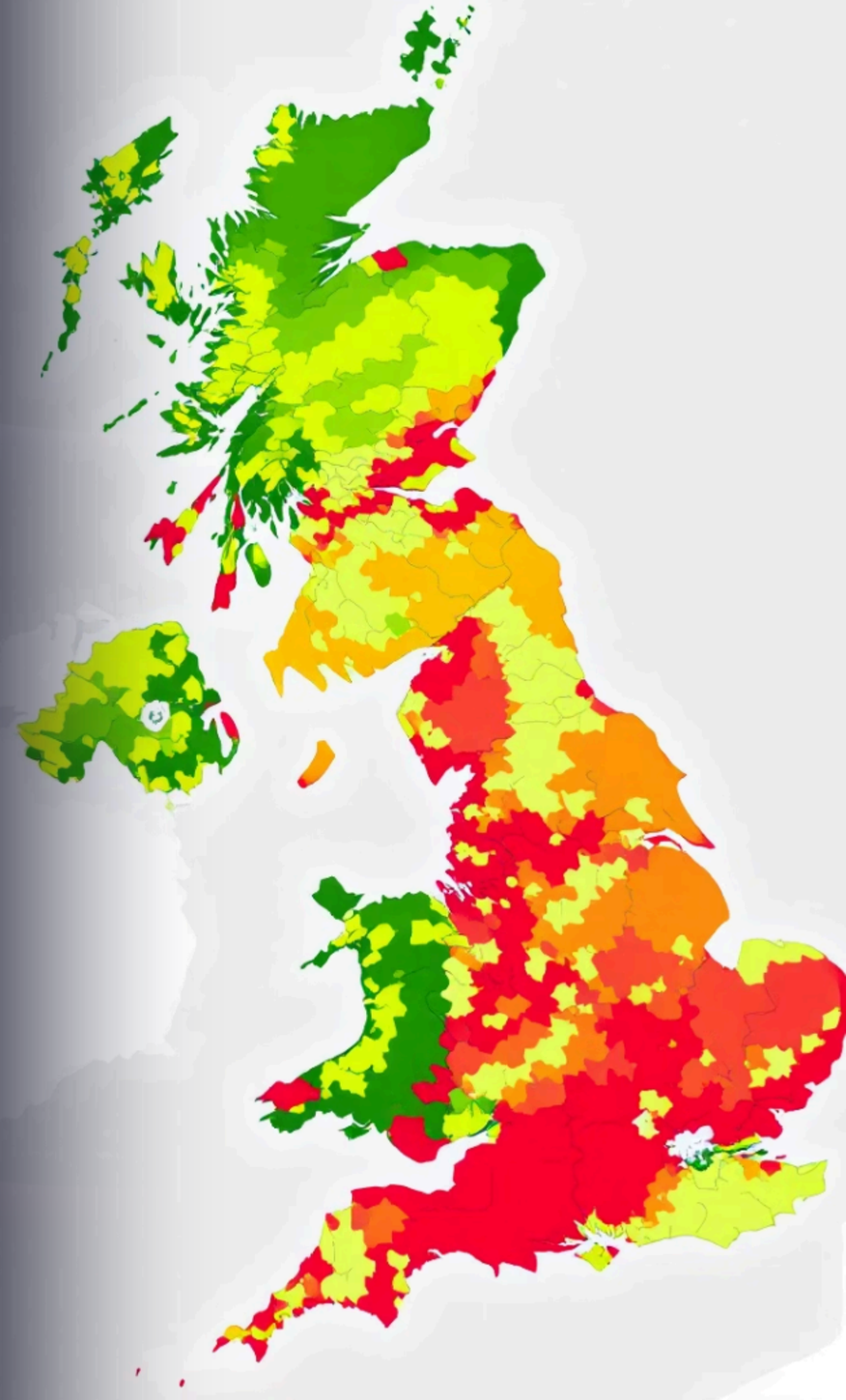
Significant spikes during major waves (Delta, Omicron)

3

Choropleth Map

Total cases per country

Visualizing cases and deaths over time reveals trends. The choropleth map shows regional disparities in case distribution.



Visualizations: Active Cases

Top Countries

United States, India, Brazil

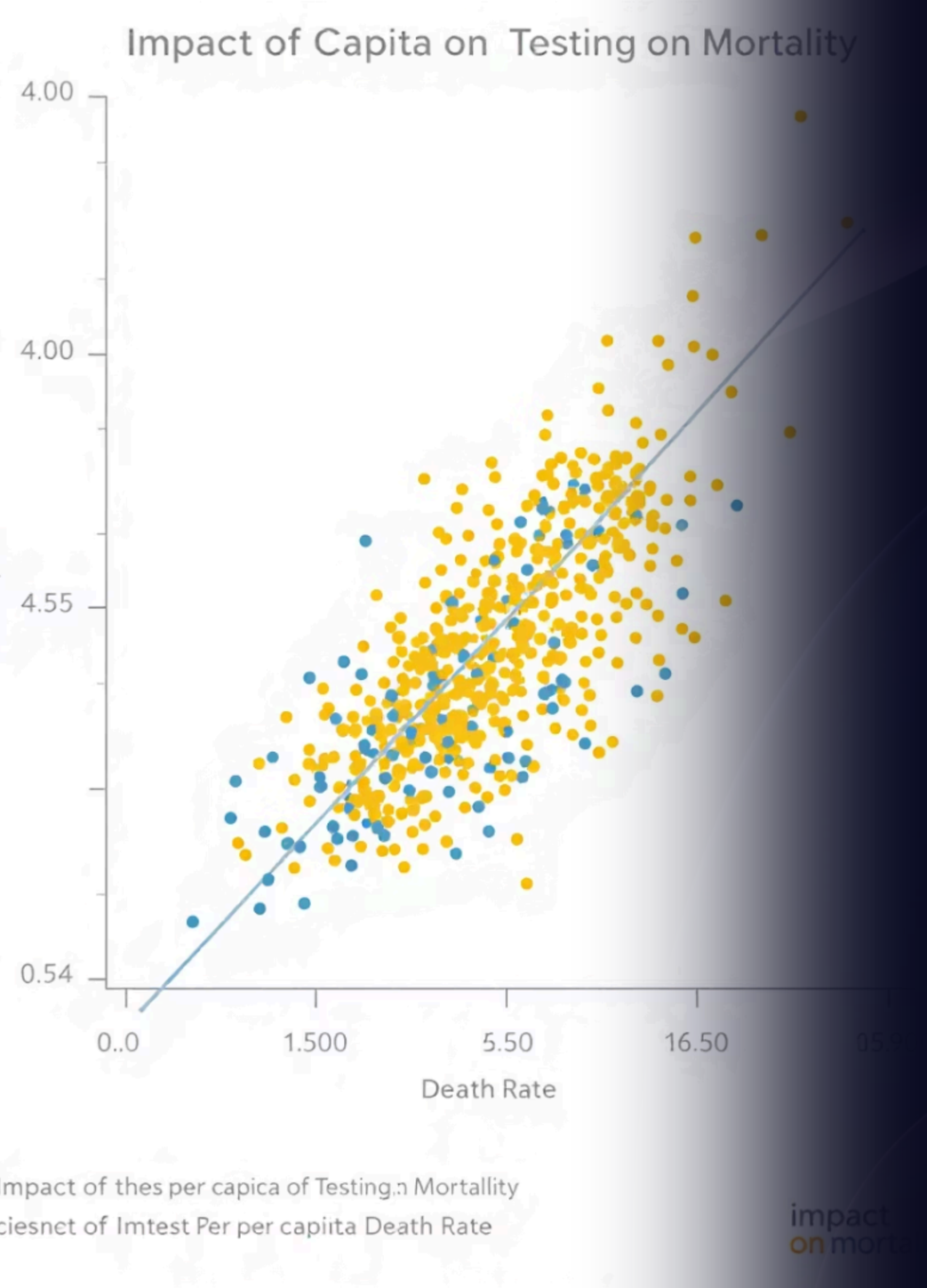
Time Series

Analysis of active cases in specific regions

Trends

Identifying resurgence and trends

Tracking active cases is crucial for monitoring the pandemic. This bar chart shows the top 10 countries with the highest active cases.



Visualizations: Testing and Impact

1. **High Testing**

   South Korea: high testing rates, low death rates.

2. **Low Testing**

   Various African nations: low testing rates, high death rates.

The scatterplot reveals the impact of testing on mortality. Visualizing this impact helps inform public health strategies.

Conclusion: Key Findings and

Key findings highlight global trends in COVID-19. The importance of testing and healthcare infrastructure is clear.

Regional disparities provide insights for public health strategies. Future research should focus on variant analysis and vaccination impact.