IMC503 – Data Science Toolkit

Lab Assessment (18/11/2025)

Q1: Visualisation & Insights Report

Name: Rakshita Rama Kubal
Roll No: 2319
Dataset Used: *country_wise_latest – country_wise_latest.csv*
Tools Used: R, ggplot2, Base R Graphics
Script: *analysis.R*

---

1. Introduction

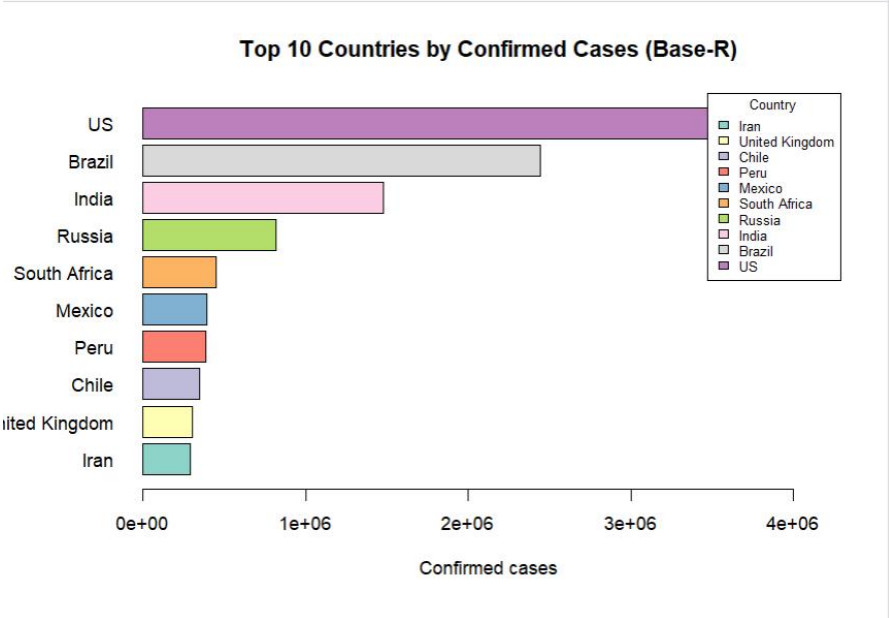
This report analyses the global COVID-19 dataset containing country-wise information on confirmed cases, recoveries, deaths, and active cases.
The goals of the analysis were to:

- Load and clean the dataset using R
- Compute **Case Fatality Rate (CFR)** and **Recovery Rate**
- Generate multiple visualisations using both base R and ggplot2
- Produce insights based on the plotted patterns
- Export all plots into a single PDF

---

2. Visualisations & Insights

Plot 1 — Top 10 Countries by Confirmed Cases *(Bar Chart)*

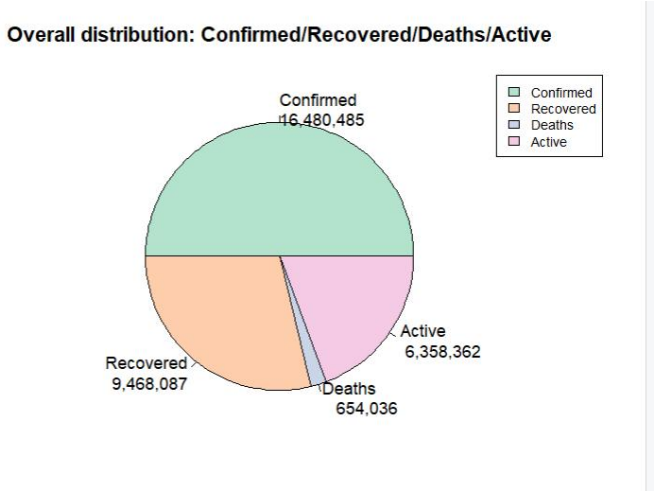


What the plot shows:
Countries with the highest number of confirmed COVID-19 infections.

Insights:

- A small number of countries dominate global confirmed cases.
- These nations acted as large hotspots with heavy healthcare load.
- The distribution is strongly skewed, showing inequality in outbreak spread.

Plot 2 — Global Outcome Distribution *(Pie Chart)*

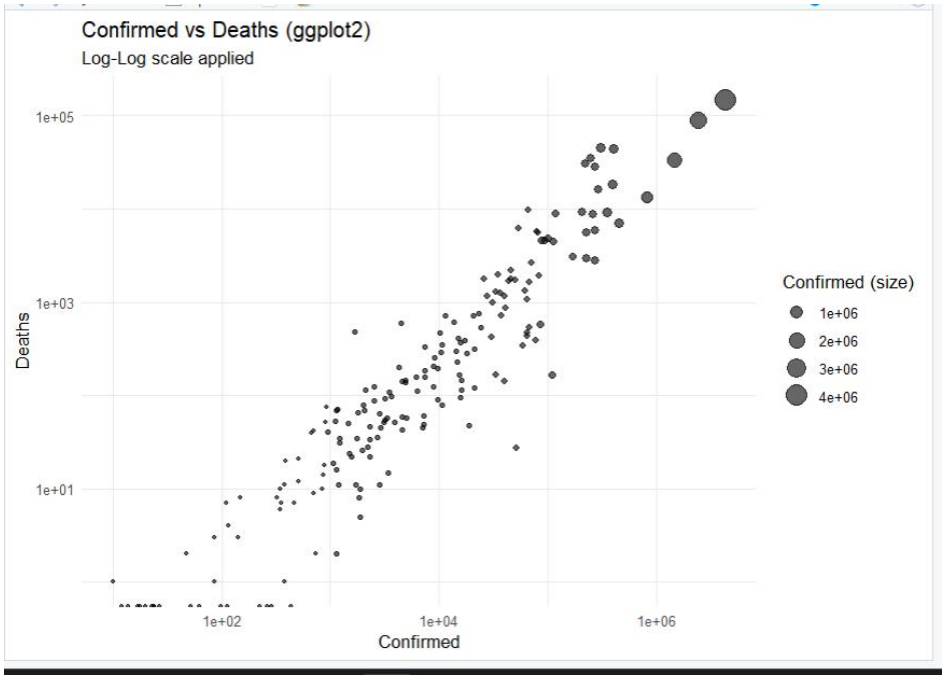


What the plot shows:
Overall distribution of Confirmed, Recovered, Deaths, and Active cases.

Insights:

- Recovered cases form the largest proportion.
- Deaths form the smallest proportion but remain significant globally.
- Active cases reveal ongoing community transmission.
- Provides a quick snapshot of the global pandemic stage.

Plot 3 — Confirmed vs Deaths *(Scatter Plot, log-scale applied when appropriate)*

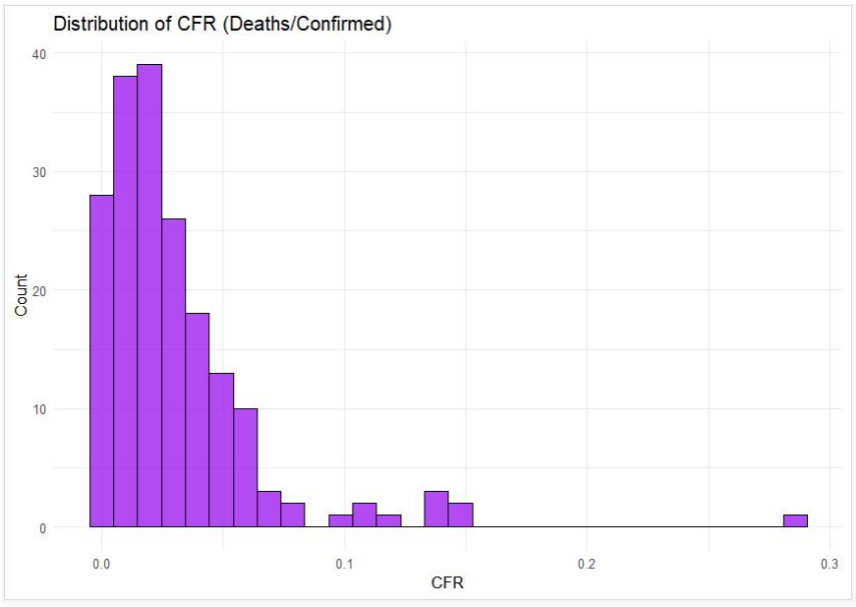


What the plot shows:
Relationship between total confirmed cases and total deaths for each country.

Insights:

- A strong positive relationship is visible: higher confirmed → higher deaths.
- Log scaling allows comparison across countries of different population sizes.
- Countries above the trend may have high fatality risk or stressed healthcare systems.
- Countries far below the trend may have under-reporting or strong medical support.

Plot 4 — Case Fatality Rate (CFR) Distribution *(Histogram)*

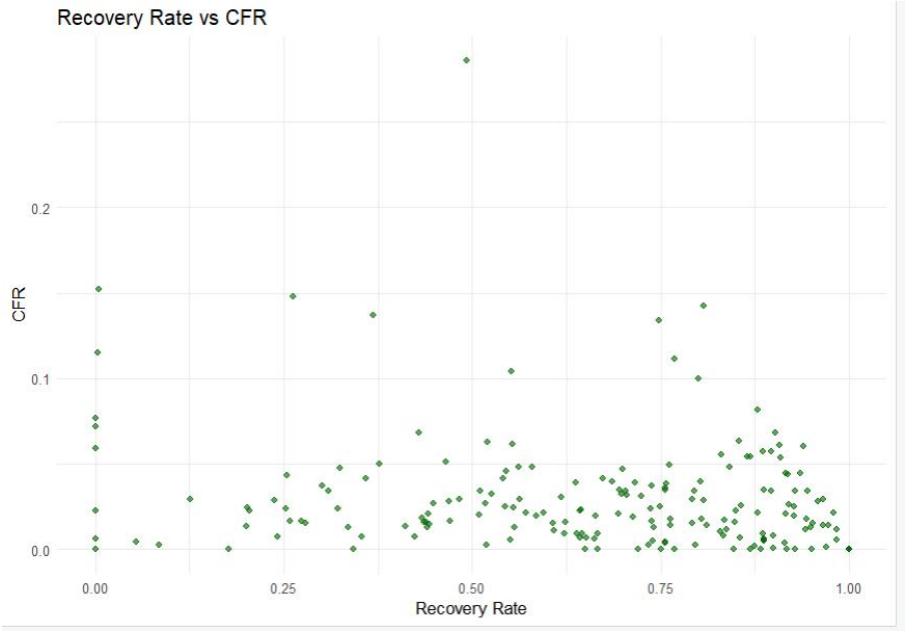


What the plot shows:
Distribution of CFR across countries.

Insights:

- Most countries show low CFR values (<5%).
- A few outliers show very high CFR → possible testing gaps or healthcare challenges.
- High-CFR regions may require targeted public-health attention.

Plot 5 — Recovery Rate vs CFR *(Scatter Plot)*

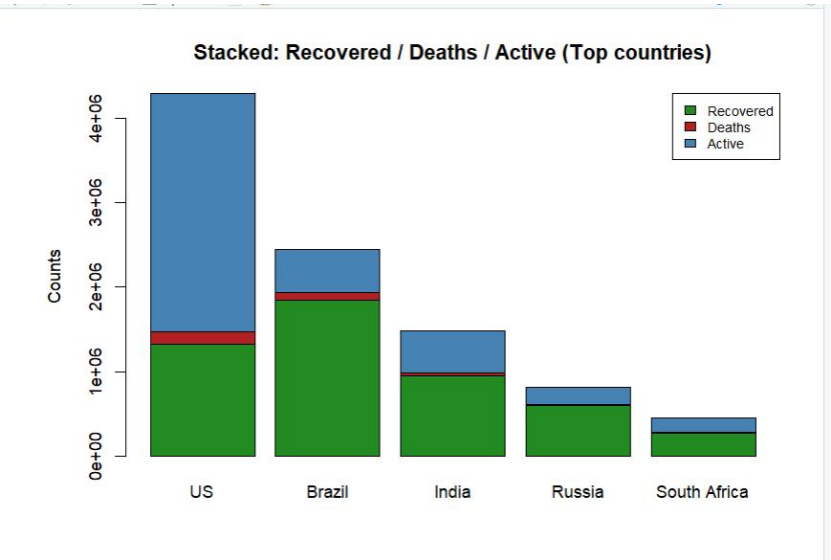


What the plot shows:
Relationship between countries' Recovery Rate and CFR.

Insights:

- Higher recovery rates correlate with lower CFR.
- Outliers with low recovery and high CFR highlight areas needing intervention.
- This plot provides a comparative look at health system performance across countries.

Plot 6 — Outcome Composition for Top Countries *(Stacked Bar Chart)*



What the plot shows:
Stacked values of Recovered, Deaths, and Active for top affected countries.

Insights:

- Many top countries are transitioning to recovery with high recovery bars.
- Others still show large active caseloads, indicating ongoing transmission.
- Death proportions help compare severity across nations with similar confirmed totals.

---

3. Overall Summary of Findings

- COVID-19 burden was **not evenly distributed**; a few nations dominated global infections.
- Countries with higher confirmed cases generally showed higher deaths.
- Recovery performance was strong in most countries, but several showed worrying fatality levels.
- CFR and recovery differences highlight variations in healthcare quality and reporting systems.
- Some top-burden countries still had high active cases, indicating continued spread.

---

4. Recommendations Based on the Analysis

- **Prioritise support** for high-burden countries with both high CFR and high active cases.
- **Investigate high-CFR outliers** for healthcare stress or under-testing.
- **Use top-10 confirmed ranking** for policy planning, aid, and vaccination strategy.
- **Add time-series data** to monitor rising or falling trends in future analysis.
- **Encourage standardised reporting** across countries for better comparison.

---

5. Limitations

- Dataset is a **single-day snapshot**; no time trends are visible.
- CFR and Recovery Rate depend on testing accuracy, which varies by country.
- Some countries may under-report deaths or cases.
- Missing or inconsistent data required assumptions during cleaning.
- No demographic or hospital capacity variables included.

---

6. Conclusion

This analysis uses R to successfully transform the COVID-19 dataset into clear, meaningful visual insights.
The visualisations highlight differences in global infection load, fatality rates, recovery performance, and active spread.
The produced PDF (visualizations_report_R.pdf) includes all graphs required for submission.
