

Statistical Analysis & Visualization on Indian Development

Reading India's growth story — and its regional divides — through Population, Income, Literacy, Health & Industry data, built entirely in R

Prakhar Srivastava | 24514, B.Tech CSE (Data Science), Indian Institute of Information Technology Una, Himachal Pradesh, India

Supervisor: Dr. Prince Sharma · DSMC203 – Data Analysis & Visualization

✓ Literacy → Income link significant ($p=0.037$) ✗ Multivariate model not significant ($p=0.157$) ✓ India overtakes China in population, 2023 ✓ Extreme poverty (\$3/day) down from 60% to 12%
✗ Male-Female literacy gap remains ~12 points ✓ 3 distinct state development clusters identified

INTRODUCTION & MOTIVATION

India, home to over 1.4 billion people, is among the world's fastest-growing economies. Yet national averages conceal sharp regional contrasts — states like Kerala and Tamil Nadu lead on human development, while others lag in literacy, healthcare access and income.

This study uses the **R programming language** to turn public state-level datasets into statistical evidence and visual stories about where India has progressed — and where gaps persist.

33

States / UTs analysed

8

Indicators studied

DATA & METHODOLOGY

- Sources:** Census of India, RBI/NSO, NITI Aayog, World Bank, National Health Profile, MoSPI (2000–2023)
- Tools:** R (≥ 4.2), RStudio — ggplot2, dplyr, tidyr, corplot, ggbeeswarm
- Descriptive stats:** mean, median, SD across literacy, income, GDP, sex ratio
- Regression:** simple & multiple linear models predicting per-capita income
- Clustering:** K-means ($k=3$, elbow-validated) + PCA to profile states

REGRESSION SNAPSHOT

Literacy → Income (simple OLS): +1% literacy = + ₹6,037 per-capita income.

$R^2=0.132$

Simple model, $p=0.037^*$

$R^2=0.245$

Multi-var, $p=0.157$ (n.s.)

Adding gender gap, sex ratio, GDP and population density explains more variance, but no single predictor reaches significance — pointing to multicollinearity among structural factors ($N=33$).

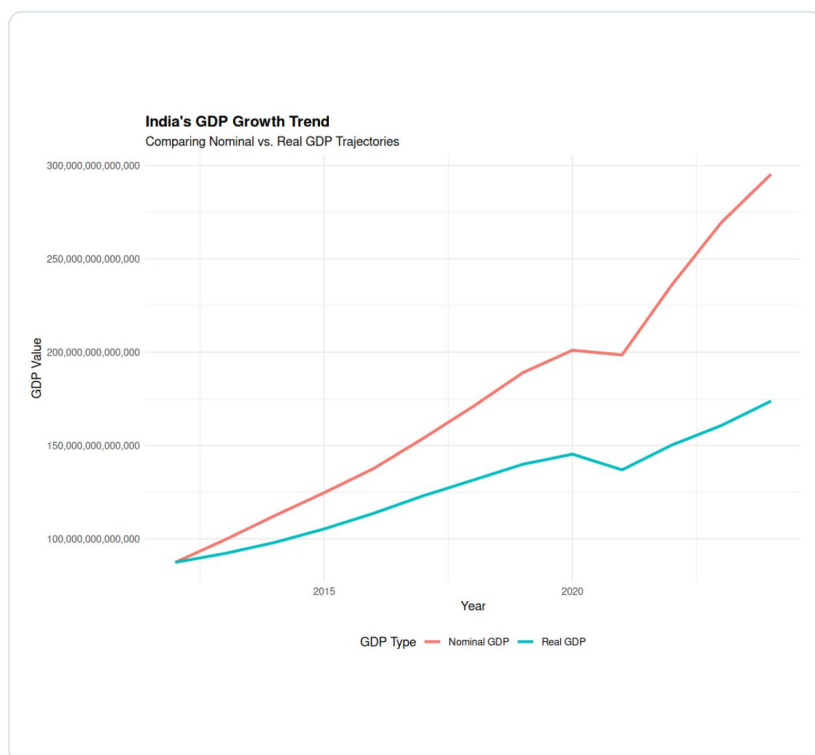
KEY CONCLUSIONS

- ✓ Literacy is a consistent, significant driver of income — but explains only ~13% of its variance alone.
- ! Healthcare staff and factories are heavily concentrated in a handful of industrialised / urbanised states.
- ✓ Poverty has fallen sharply at every PPP threshold since the 1970s, but the pace differs by region.
- ★ K-means clustering surfaces 3 development profiles: mega-economies (MH, UP), NE/resource states, and high-HDI urban leaders (Kerala, Delhi).

REFERENCES & FUTURE WORK

- Census of India (2011). *Primary Census Abstract*. Registrar General & Census Commissioner, India.
- World Bank (2023). *World Development Indicators*. Washington, DC.
- NITI Aayog (2021). *India National Multidimensional Poverty Index*.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation.
- National Health Profile, CBHI & Population Projections 2011–2036, RGI.

Where this goes next: district-level data for finer spatial resolution, panel-data & causal inference models, predictive modelling with **tidymodels**, and satellite night-light intensity as an alternative proxy for regional economic activity.



01 CENTERPIECE INDICATOR · GDP

GDP Growth: India's Economic Engine

Nominal GDP has nearly **tripled** since the early 2010s, climbing from roughly ₹85 lakh crore toward ₹290 lakh crore. Real, inflation-adjusted GDP tells a more modest story — rising steadily but at a noticeably slower pace, with the widening gap between the two lines reflecting sustained inflation over the period.

Both series show a visible dip in **2020–21**, the COVID-19 disruption, followed by a sharp stimulus-driven rebound. This trajectory underpins nearly every other indicator on this poster: it is the growth engine against which literacy, healthcare, poverty and industrial gains are ultimately measured.

~3×

Nominal GDP growth,
2011–2023

2020–21

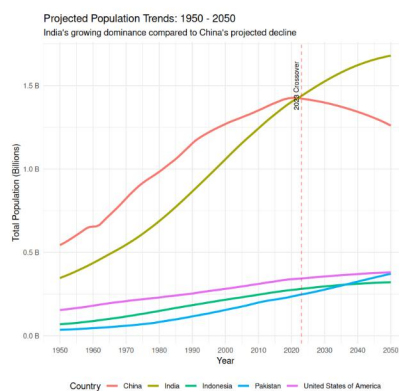
Only COVID-era
contraction

$r = 0.36$

Literacy correlates with
income, not raw GDP

02 Population Crossover

2023

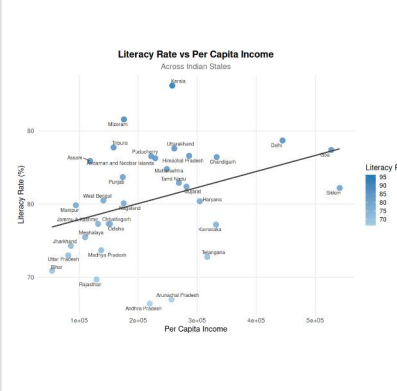


India overtook China as the world's most populous nation around 2023 and continues rising toward ~1.7B by 2050, while China's population is projected to decline.

India vs. China, Indonesia, Pakistan, USA

03 Literacy vs Income

$r > 0$

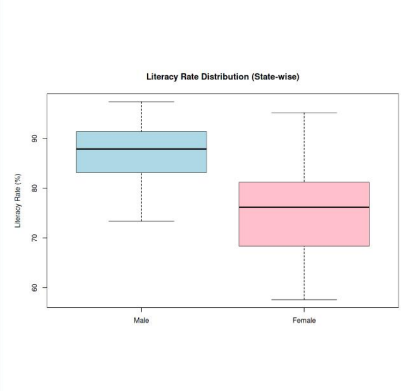


States with higher literacy generally report higher per-capita income (Goa, Delhi, Sikkim), while low-literacy states (Bihar, UP, Rajasthan) cluster at low income — a positive but noisy trend.

Trend line: positive slope

04 Gender Literacy Gap

Persistent

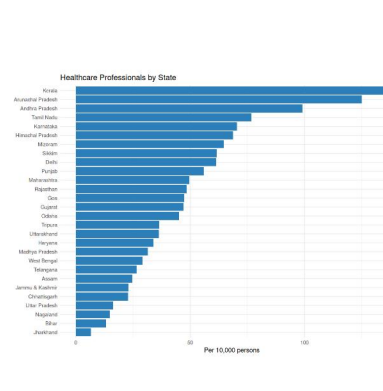


Male literacy medians (~88%) sit well above female literacy (~76%) across states, with wider spread on the female side — evidence of an enduring gender education gap.

Median gap ≈ 12 points

05 Healthcare Professionals

Uneven



Kerala, Arunachal Pradesh and Andhra Pradesh lead in healthcare staff per 10,000 people; Jharkhand, Bihar and Nagaland trail far behind — a stark infrastructure divide.

~6x gap between top & bottom states

06 Poverty Trends (1977–2022)

Declining

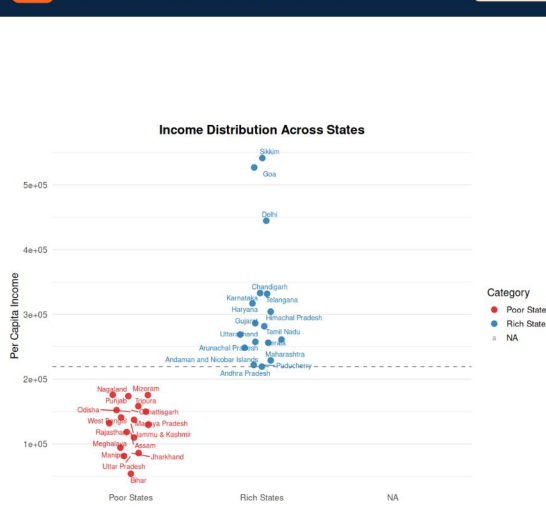


Across all three PPP poverty lines, the share of India's population in poverty has fallen steadily, with the sharpest declines after 2010.

\$3/day line: 60% → ~12%

07 Income: Rich vs Poor States

Divided

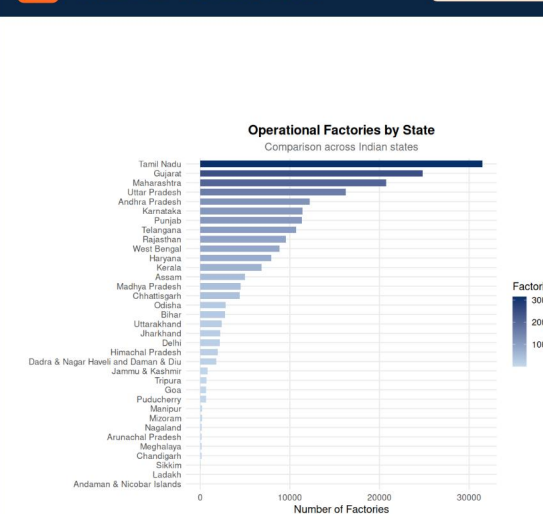


A clear split emerges around the national median: Sikkim, Goa and Delhi pull far ahead, while Bihar, UP and Jharkhand remain clustered at the bottom.

Split at median per-capita income

08 Industrial Concentration

Concentrated



Tamil Nadu, Gujarat and Maharashtra dominate operational factory counts; most north-eastern states and UTs have negligible industrial bases.

Top 3 states > combined bottom 20

Indicator Scorecard	LITERACY → INCOME $p = 0.037^*$	MULTI-VAR MODEL $R^2 = 0.245$ (n.s.)	EXTREME POVERTY 60% → 12%	GENDER GAP ≈ 12 pts	HEALTH STAFF RANGE ~6x gap	FACTORIES, TOP-3 TN · GJ · MH	STATE CLUSTERS $k = 3$
---------------------	------------------------------------	--	------------------------------	------------------------	-------------------------------	----------------------------------	---------------------------