Electoral Turnout in India | Barriers to Voting in Andhra Pradesh

Overview

More than 850 million eligible voters are currently electing the next leader of the world's biggest democracy: India. But do the disenfranchised, illiterate, rural, or other disadvantaged communities get a say in who becomes the next leader of India? Or are they kept away from the polls by long commutes, lack of infrastructure, or other voting barriers?

I analyze electoral turnout for the 2009 general elections for the assembly of Andhra Pradesh. By analyzing polling station locations, distance to transport infrastructure, sociodemographic variables, and their relation voter turnout in 294 assembly constituencies I examine the relation between voting barriers and electoral turnout.

Counter to the initial hypothesis, distance to infrastructure has no statistically discernible impact on voter turnout. Population density and polling station density are even inversely related to voter turnout (i.e. the denser the population the lower the turnout). Urban voters are on average and ceteris paribus 7.2% less likely to vote than rural voters.

Rather than infrastructure barriers, sociodemographic factors seem to drive electoral turnout. More rural, female, and literate constituencies have significantly higher turnout than others. Illiteracy is one of the biggest barriers to voting. Literacy rates as low as 27% in some constituencies keep voters may lower turnout by as much as 20 percentage points ceteris paribus.
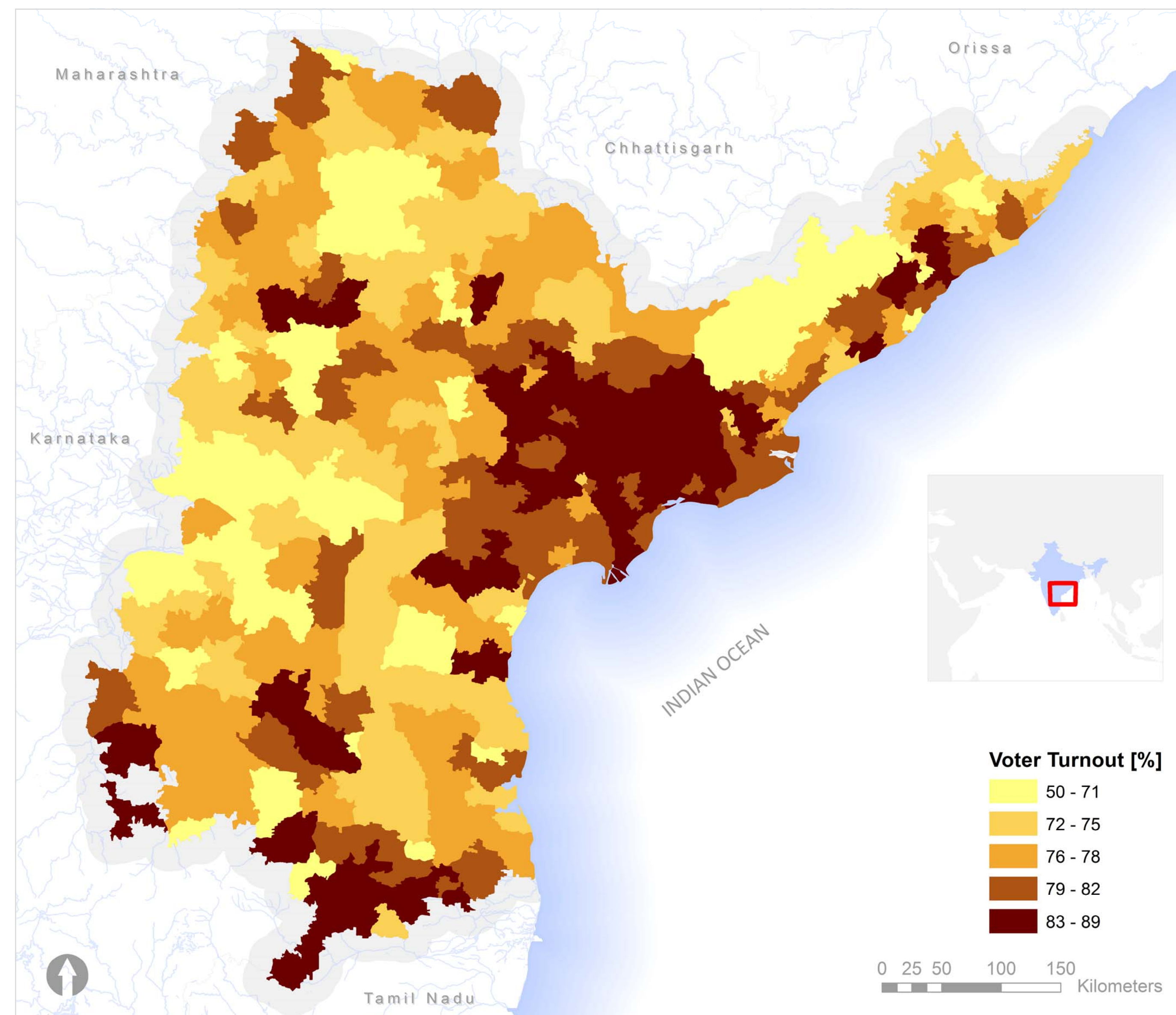
Methodology

To examine potential voting barriers this poster proceeds in four major steps:

1. Identification of Voting Barriers
2. Data Extraction & Geospatial Analysis
3. Data Aggregation on Constituency-Level
4. Regression Analysis of Barrier-Turnout Relationship

First, I identified potential voting barriers by drawing on the literature in this realm. The location of polling stations is well known to drive turnout. I also used data from the Indian Census of 2001 for sociodemographic characteristics like literacy, urbanity, agricultural labor, etc. Some seats in Indian elections are reserved for minorities (scheduled castes and scheduled tribes) and have to be treated separately in the analysis of turnout.

Second, I performed data extraction and geospatial analysis. I extracted polling station estimates from a Google Maps and PDFs. Based on this data I computed geospatial characteristics of the polling stations. For in-

stance, I calculated distances to roads, railways, airports, and other infrastructure items.

Third, I aggregate the data on the assembly constituency level. Since no valid assembly constituency shape files are available I recreate assembly constituencies from sub-districts. I then dissolved sub-districts into assembly constituencies to compute census estimates for each assembly constituency. Next, I spatially joined polling station data to the respective assembly constituencies.

Fourth and last, I performed multivariate regression analysis to identify the impact of each of the barriers on voter turnout.

I employed an Ordinary Least Squares approach with ro-

bust standard errors:

$$Voter\ Turnout = \alpha + \beta_1*DistAirport + \beta_2*DistNatHwy + \beta_3*DistStateHwy + \beta_4*DistMajRoad + \beta_5*ElectPerPollStat + \beta_6*PollStatDens + \beta_7*PopDens + \beta_8*AgricultWork + \beta_9*Rural + \beta_{10}*Female + \beta_{11}*Liter + \beta_{12}*SchedCaste + \beta_{13}*SchedTribe + u$$

While this simple linear specification could be refined further, it explains more than 52% of the variation of voter turnout even after adjusting for the high number of estimated coefficients.

Results

Voter Turnout	Coefficient	SE	t	p
Distance Airport	-0.00044	0.0001	-4.86	0%
Distance National Highway	0.000094	0.0002	0.38	71%
Distance State Highway	0.00036	0.001	0.59	56%
Distance Major Road	0.0015	0.006	0.25	80%
Electors Per Polling Station	-0.00021	0.0001	-3.06	0%
Polling Station Density	-0.0025	0.003	-0.85	39%
Population Density	-0.000021	0.00002	-1.32	19%
Share Agricultural Workers	0.52	0.076	6.87	0%
Rural	0.072	0.027	2.68	1%
Female	1.18	0.626	1.88	6%
Literacy	0.26	0.051	5.09	0%
Scheduled Caste	-0.028	0.008	-3.45	0%
Scheduled Tribe	-0.027	0.019	-1.44	15%
Constant	0.13	0.345	0.38	70%

The share of agricultural workers, literacy, and the distance to airports are the most statistically significant drivers of voter turnout. Surprisingly, agricultural work is strongly associated with turnout. On average and ceteris paribus, a 10% higher share of agricultural workers leads to a 5.2 % points higher turnout. Equally, an increase of literacy by 10% points is associated with an increase of turnout by 2.6% points. The impact of the distance to airports is economically insignificant and no other distance estimates are statistically significant.

With caveats for lack of detailed data, these results indicate that infrastructure barriers play a minor role in keeping the Indian electorate from casting its vote. Higher rural turnout indicates that infrastructure likely plays a small role in keeping voters away from the polls.

Limitations

Four challenges make geospatial analysis of Indian voter turnout data difficult:

1. Lack of detailed and consistent data,
2. High level of aggregation for voter turnout,
3. Unavailability of assembly constituency shapefiles, &
4. Absence of a good estimate of the time between voter homes and polling stations.

No detailed maps of Indian infrastructure are freely available. Thus, I had to resort to low resolution but consistent road networks provided for the year of 2001 by ML Infomaps. This lack of detail might cause omitted variables bias. Also, data collection dates do not match.

Because voter turnout is aggregated on the assembly constituency level, variation of geospatial characteristics between different polling stations within an assembly constituency cannot be exploited

Assembly constituency shapefiles are not provided by

the Election Commission of India. This analysis side-stepped this issue by reconstructing assembly constituencies from sub-districts. This might limit accuracy of the analysis.

The literature indicates that voter turnout is driven by the cost of voting. This cost is tightly linked with the time it takes a voter to get to the voting station. In the absence of this data, it is premature to infer that relocation or increase of voting stations will not increase turnout.
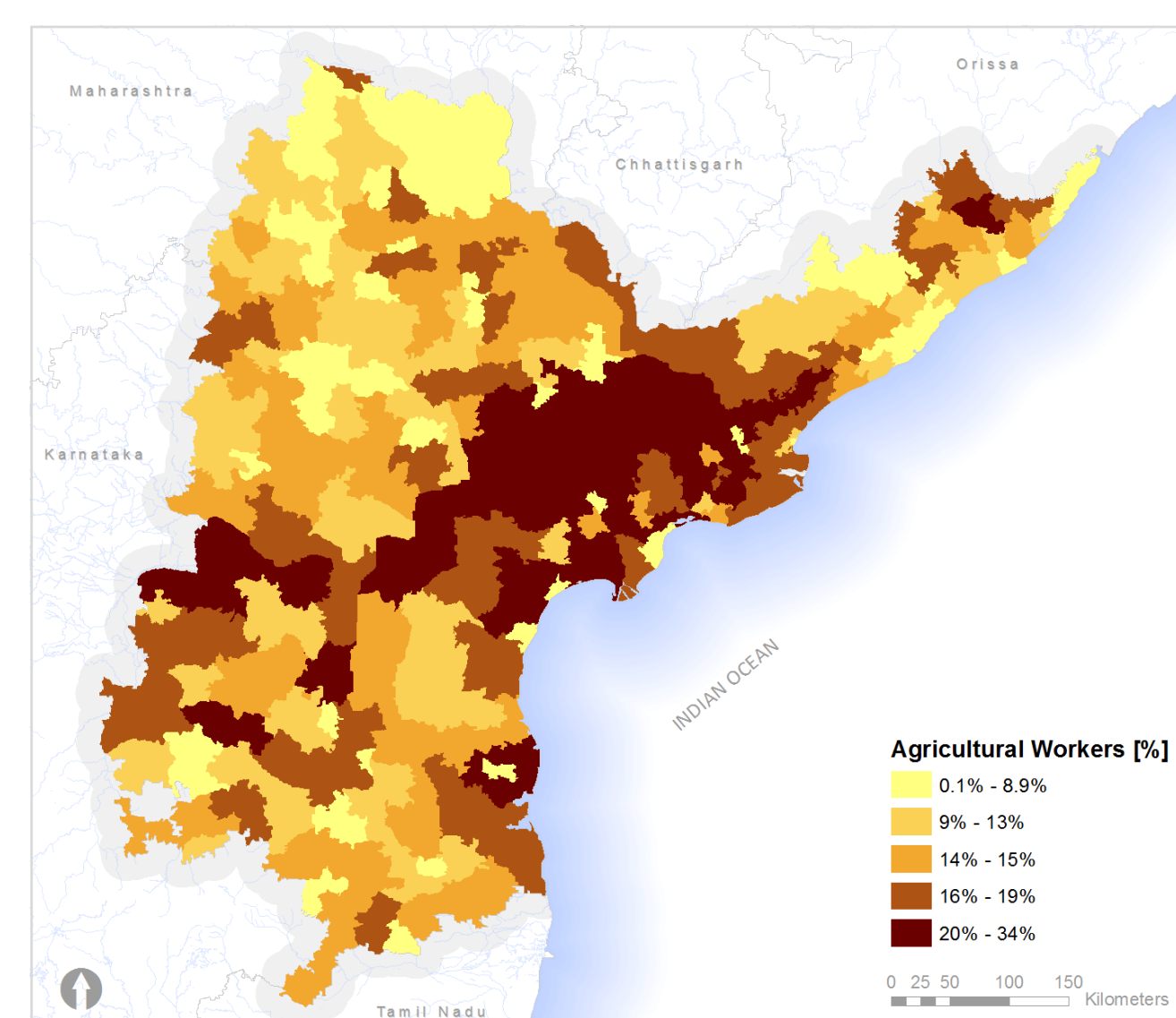
Conclusion

This analysis has a cautiously positive result: I find no evidence for an infrastructure barrier to voting in Andhra Pradesh. Sociodemographic and not geospatial characteristics seem to drive voter turnout – literacy and the share of agricultural workers stand out in this respect. The lack of detailed and disaggregated data, however, indicates that this result should be taken with caution.

In context, the results seem plausible. Between 2004 and 2009 India raised the number of polling stations by 20% and maybe this increase was effective. Also, compared to levels in the Western world, voter turnout in Andhra Pradesh is outstandingly high with 72%. This might also indicate that barriers to voting are not extremely high.
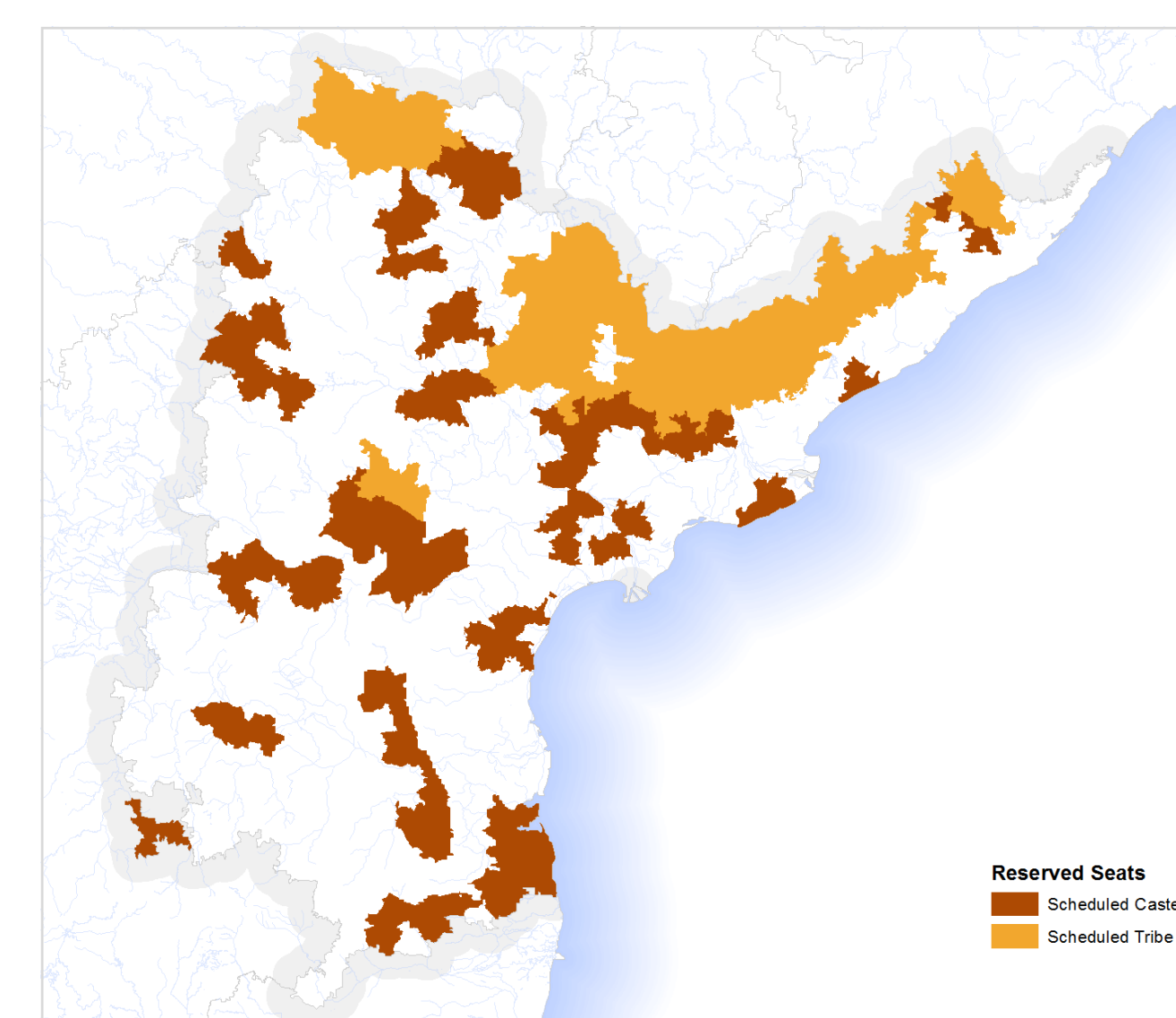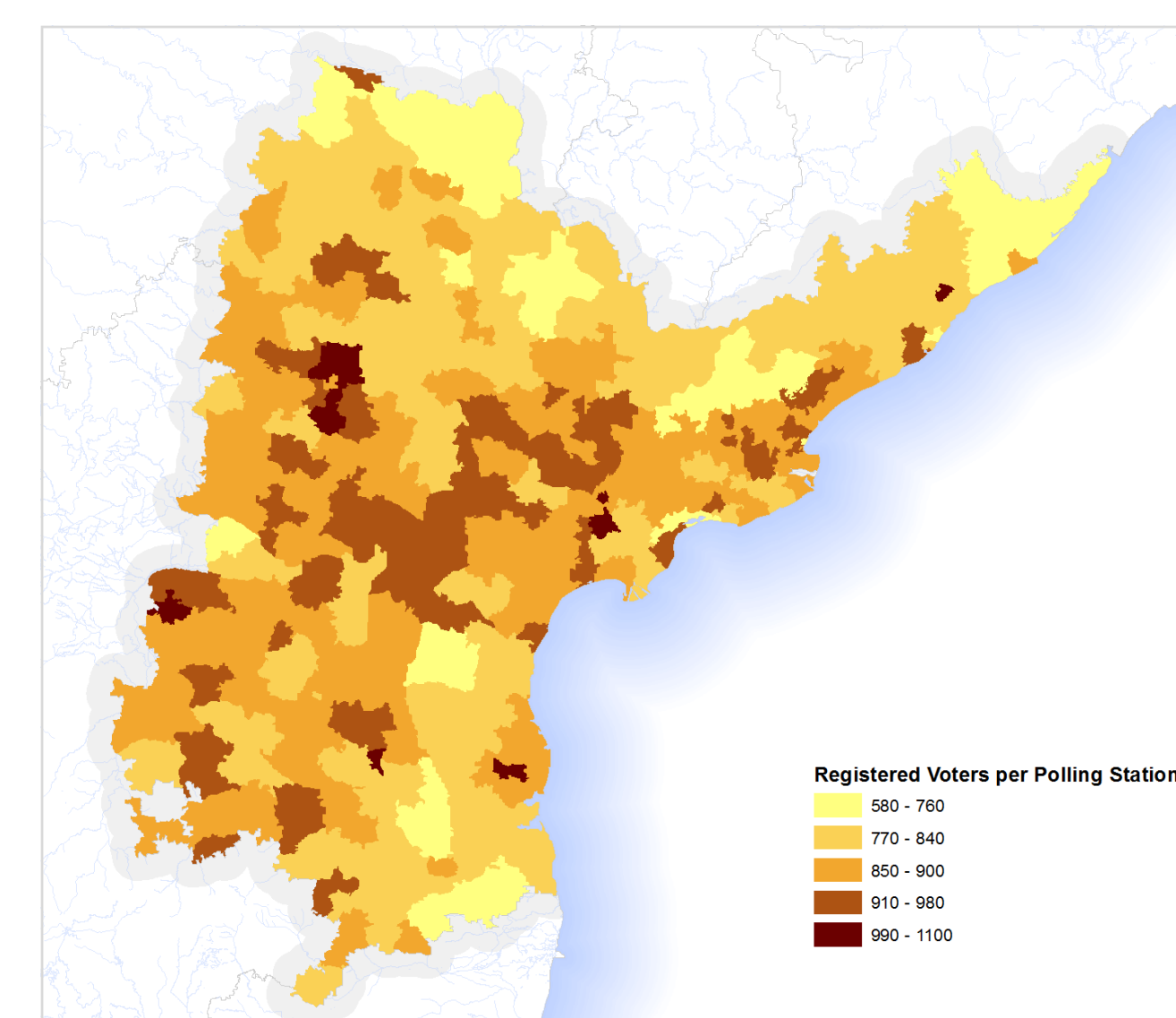
Data Sources & Information

Agricultural Workers - Most Significant Driver
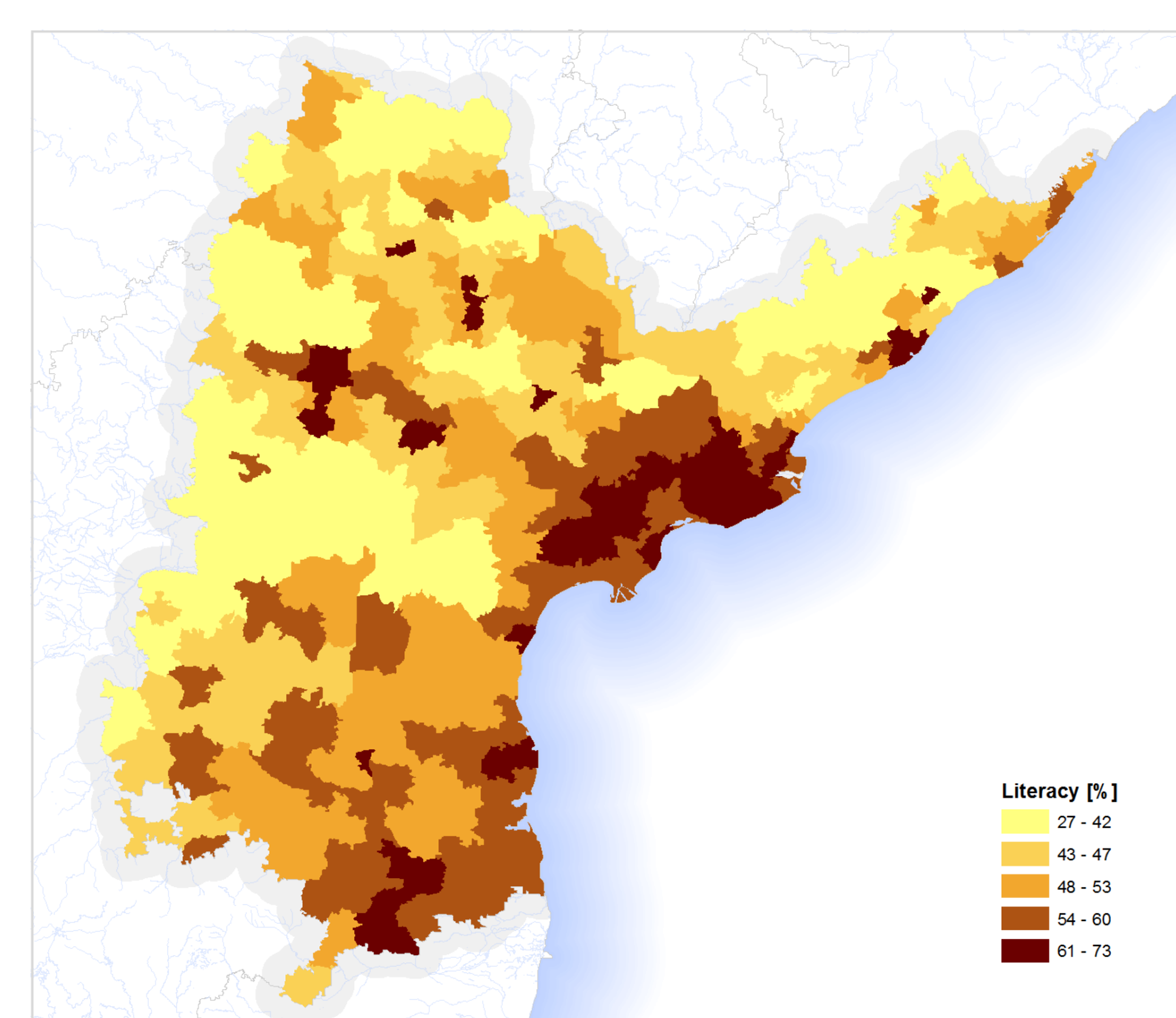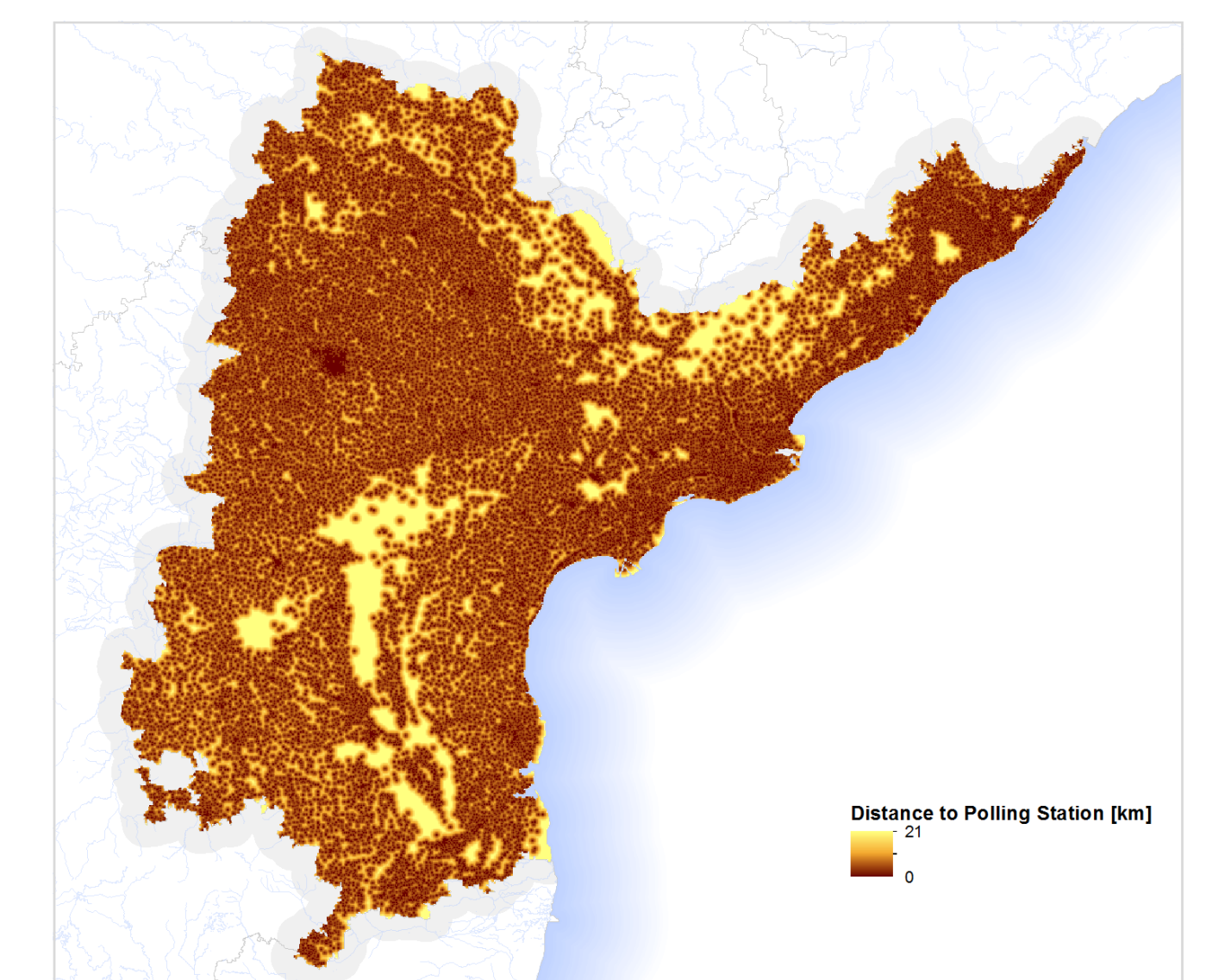

Literacy - 2nd Most Significant Driver

Reserved Seats for Schedules Castes - 3rd Most Significant Driver

Registered Voters per Voting Station - 5th Most Significant Driver

Distance to Voting Station - Hypothesized but Insignificant Driver