Group Challenge 1

DHS

Our data comes from the Demographic Health Survey Program.

- The Demographic Health Survey (DHS) is a national level survey conducted in many countries around the globe.
- DHS collects and distributes a plethora of data on aspects of developing nations.

DHS

- Our particular analysis will be conducted on the **birth** related survey from 2017-2018 in Bangladesh.
- For this survey, households were sampled in particular clusters (villages or town centers) around the country.
- If a mother was surveyed in a particular cluster location, they were asked if they had any children under the age of **5**. If they did, then a particular question was about their children under the age of 5.
- For this particular survey, 2383 observations (births) were recorded across Bangladesh.

Analysis of Covariates - Summary of Main Dataset and Cluster Metadata

The main dataset contains 29 covariates. One of these covariates is cluster_id, a unique numeric key identifying a site of measurement. Another dataset contains metadata pertaining to each individual cluster, detailing:

- Geographic location (latitude, longitude)
- ID for the DHS survey
- Year of the survey
- Coordinate reference system (CRS: WGS84)

Analysis of Covariates - Summary of Main Dataset and Cluster Metadata

Although none of the metadata fields will be used for analysis, we must join on cluster_id to recover geographic locations.

Household and Demographic Variables

- Cluster ID from DHS data
- Household number within the cluster
- Rank of individual within household
- Number of household members
- Number of children under 5 in household
- Gender of household head

Household and Demographic Variables

- Urban or rural residence
- Wealth index category
- Survey weighting variable
- Stratification variable for survey design
- PSU identifier for survey design

Maternal Variables

- Age of the mother
- Education level of the mother
- Occupation type
- Mother's height (cm)
- Mother's BMI
- Marital status of the mother
- Number of antenatal visits attended
- Diagnosed with high blood pressure
- Diagnosed with diabetes

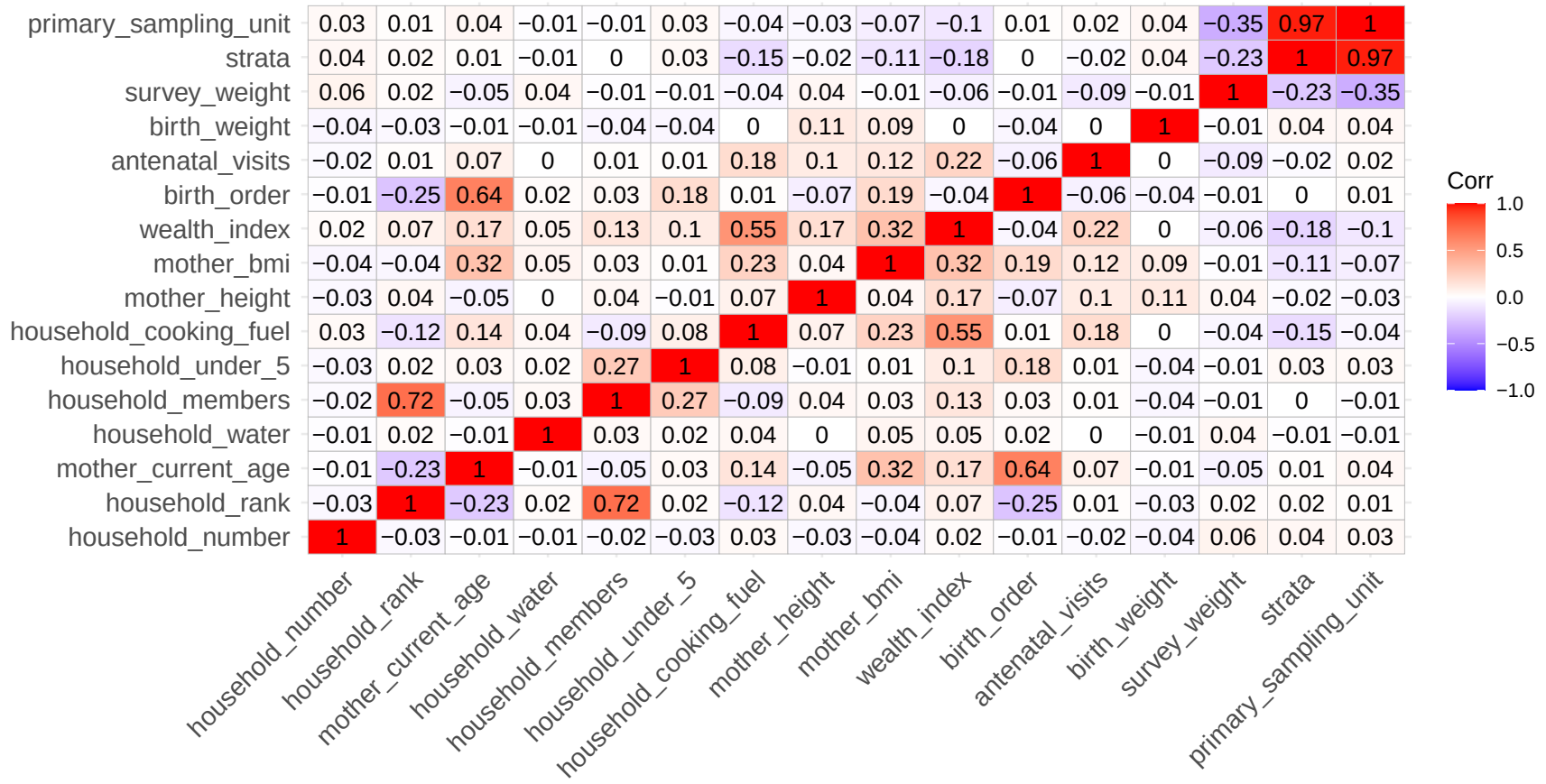
Child Variables

- Birth order of the child
- Sex of child
- Source of birth weight measurement
- Birth weight in grams
- Categorized birth weight
- Date of birth

Household Infrastructure Variables

- Type of drinking water source
- Type of toilet facility
- Availability of electricity
- Floor material
- Type of cooking fuel

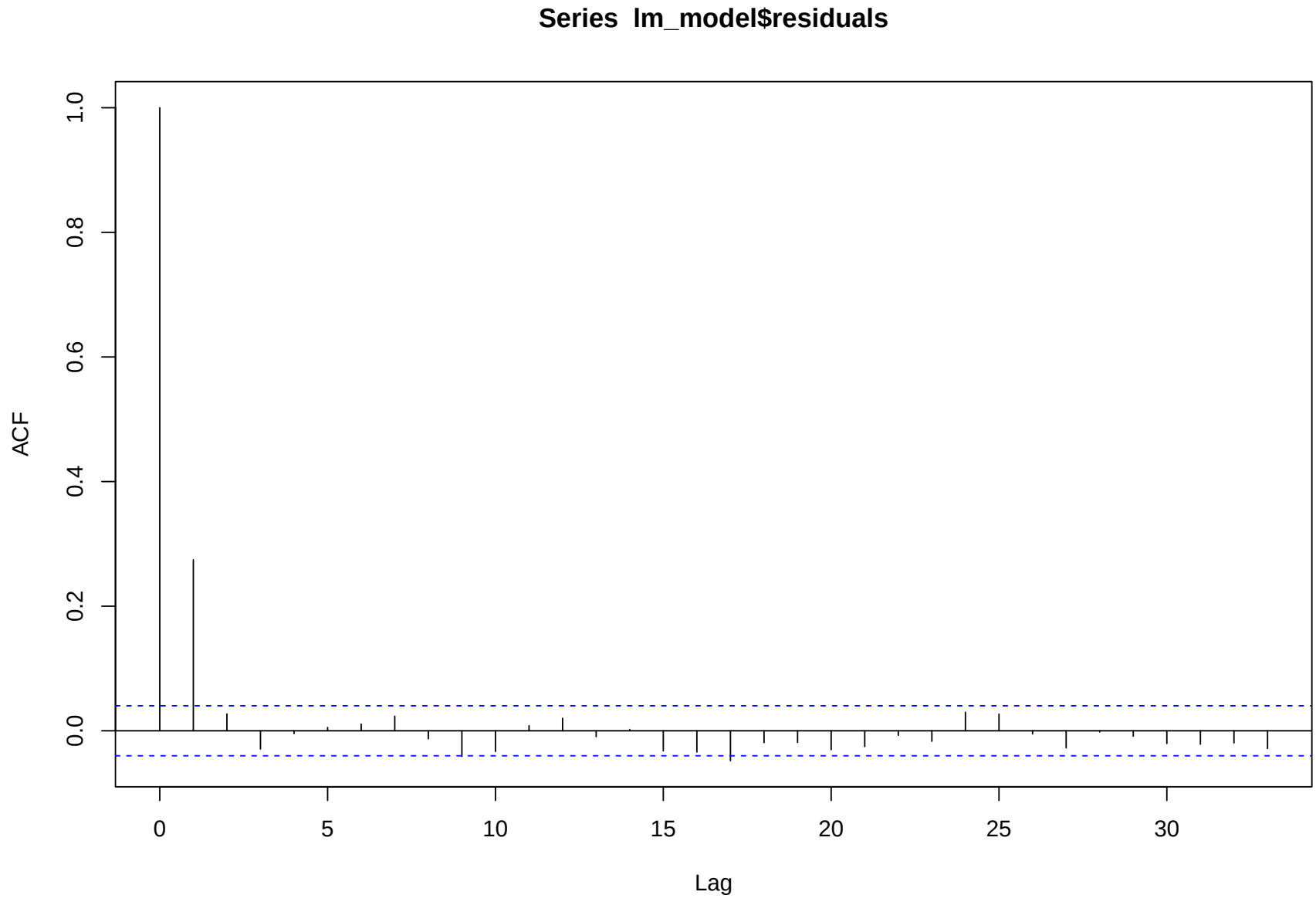
Analysis of Covariates - Pairwise Correlation of Numeric Variables



Analysis of Covariates - Variance Inflation Factor

household_rank	mother_current_age	househ
2.555389	1.983368	
household_members	household_under_5	household_coo
2.615649	1.213228	
mother_height	mother_bmi	wea
1.069034	1.248646	
birth_order	antenatal_visits	bir
2.011993	1.090437	
survey_weight	strata	primary_samp
1.552240	31.586843	

Analysis of Covariates - Autocorrelation



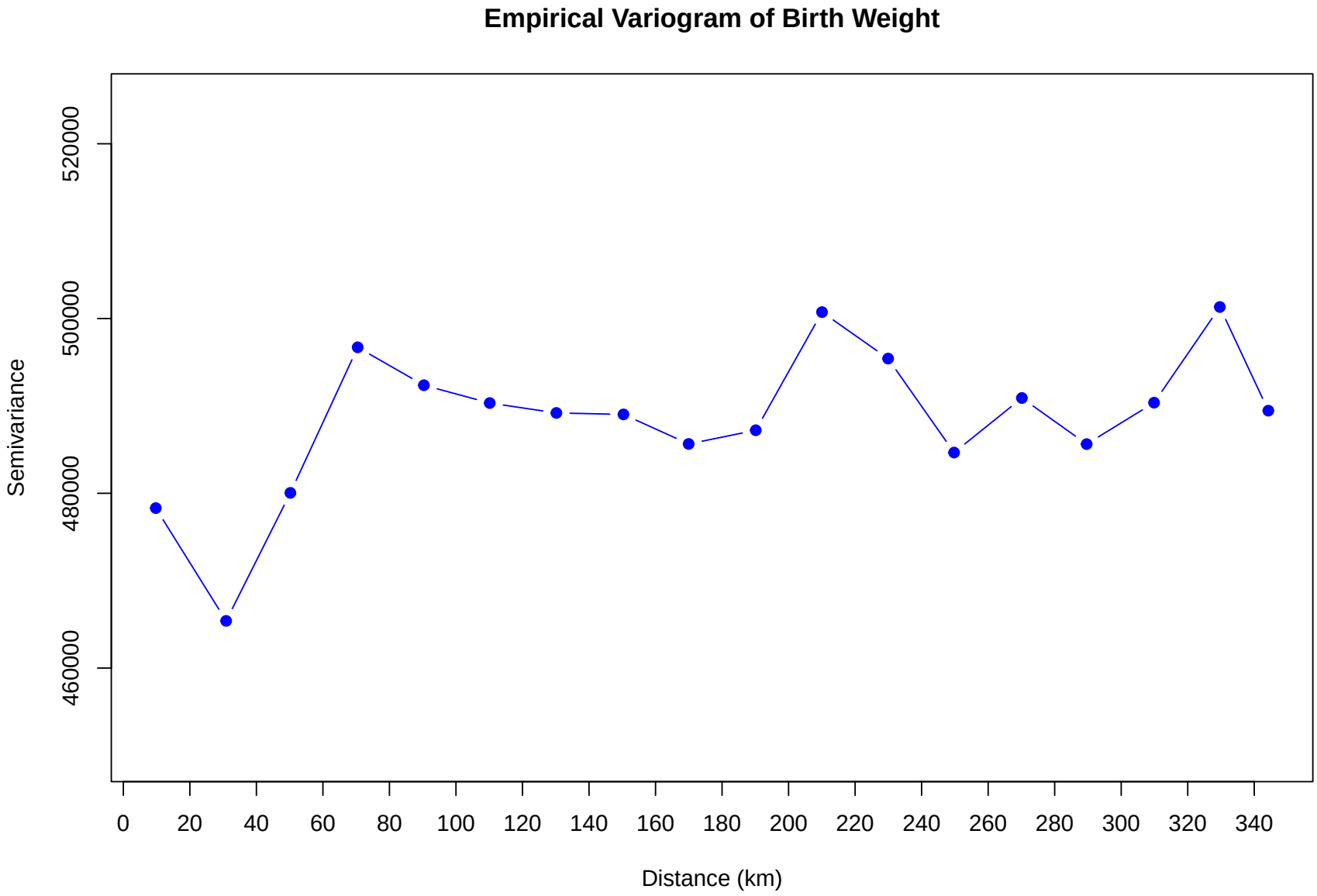
Spatial Domain

- The study covers **Bangladesh**, with observations from **2,383 participants** distributed across geographic coordinates (latitude and longitude).
- Each participant's location was masked by assigning them to a **cluster centroid** (via `clusterid`).
- To create pseudo-unique locations, small **i.i.d. Gaussian noise** (SD $= 0.001$ degrees) was added to latitude/longitude.
- Observations are **spatially indexed** using these randomized coordinates.

Spatial Dependence

- **Covariance decay with distance** is a reasonable assumption due to spatial dependence of birth weight on environmental factors (e.g., local temperature).
- **Spatial correlation** may arise from:
  - Shared environmental exposures (e.g., regional climate, pollution).
  - Socioeconomic conditions that vary geographically.
- Using **distance-based covariance** assumes that **mothers in close proximity experience similar environmental influences** affecting birth outcomes.

Empirical Variogram



Empirical Variogram

The empirical variogram for birth weight displays three key features:

1. A decrease in semivariance at very short distances ($<$20 km), likely reflecting the artificial noise added to cluster centroids. This "dip" represents reduced variability within clusters due to the imposed randomization, as participants from the same original cluster are now treated as spatially distinct but still share similar environmental conditions.

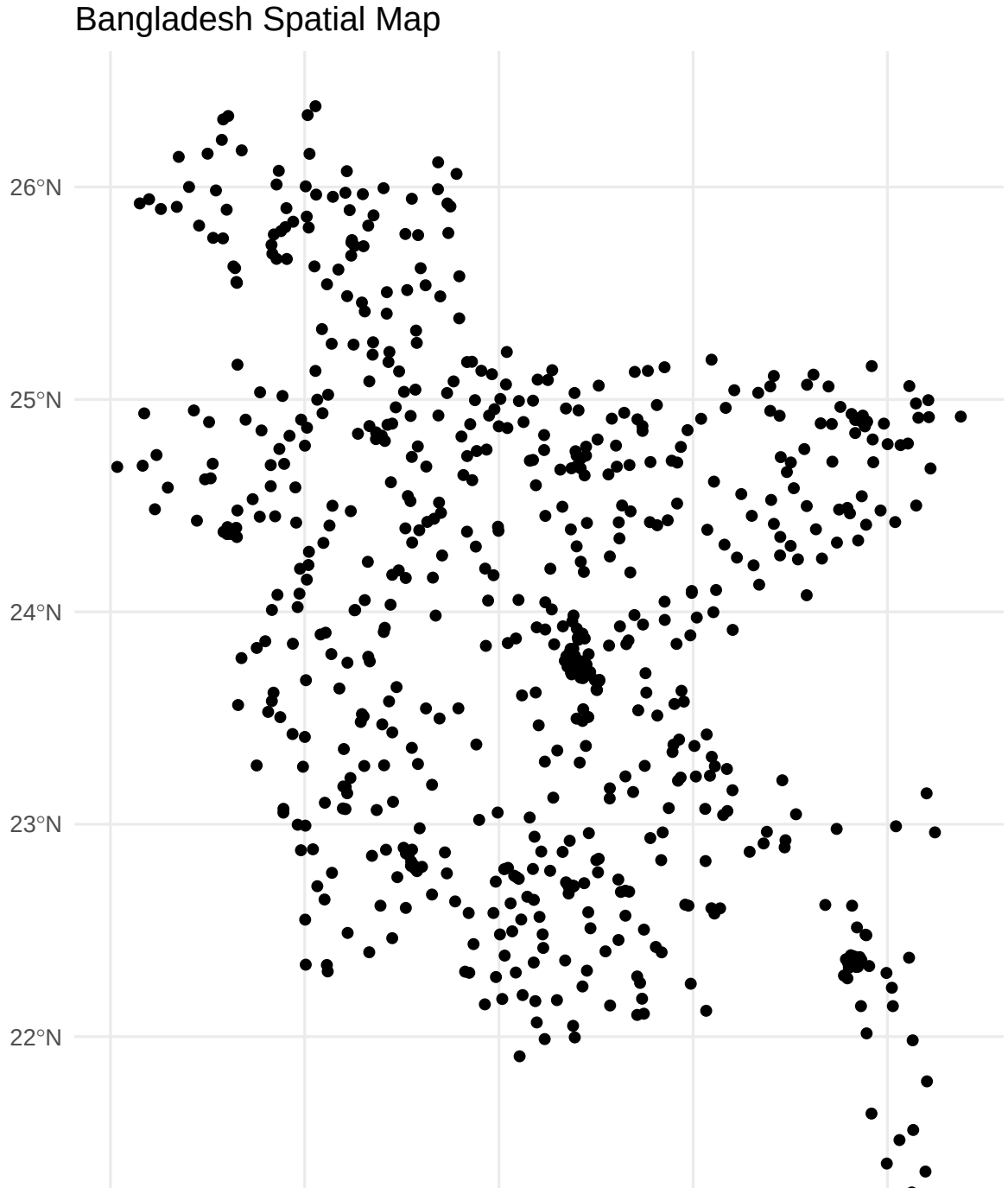
Empirical Variogram

2. An increasing trend between 20–80 km, indicating spatial dependence. Rising semivariance with distance suggests that birth weights become less similar as geographic separation increases, consistent with spatial autocorrelation.
3. A plateau beyond 80 km, implying the spatial correlation range (where covariance stabilizes) is reached.

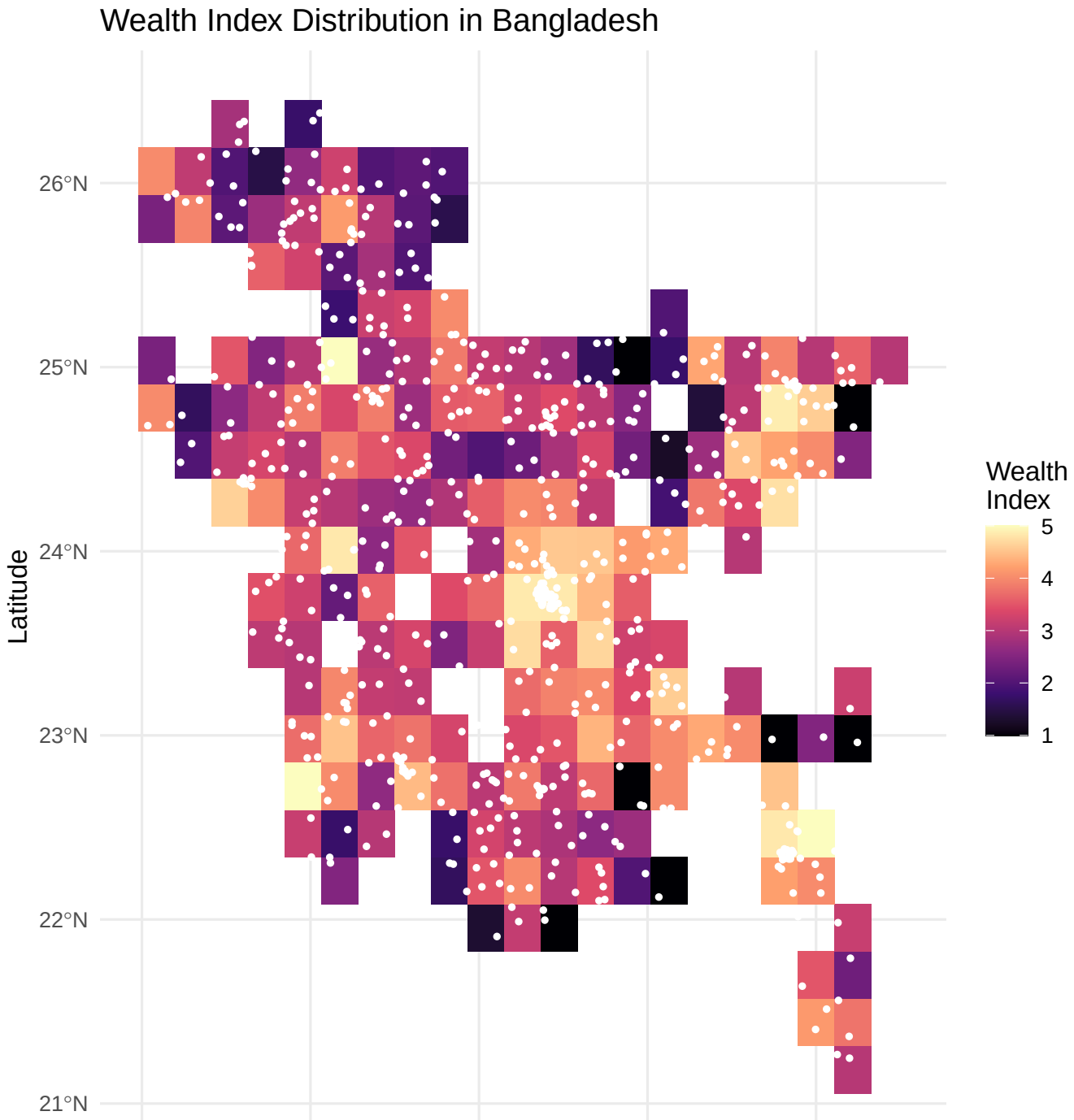
Empirical Variogram

Overall, the spatial pattern aligns with the hypothesis that environmental factors like temperature—which vary regionally—may contribute to geographic disparities in birth weight. Further investigation into the covariance structure and model diagnostics would strengthen these conclusions.

Cluster Locations



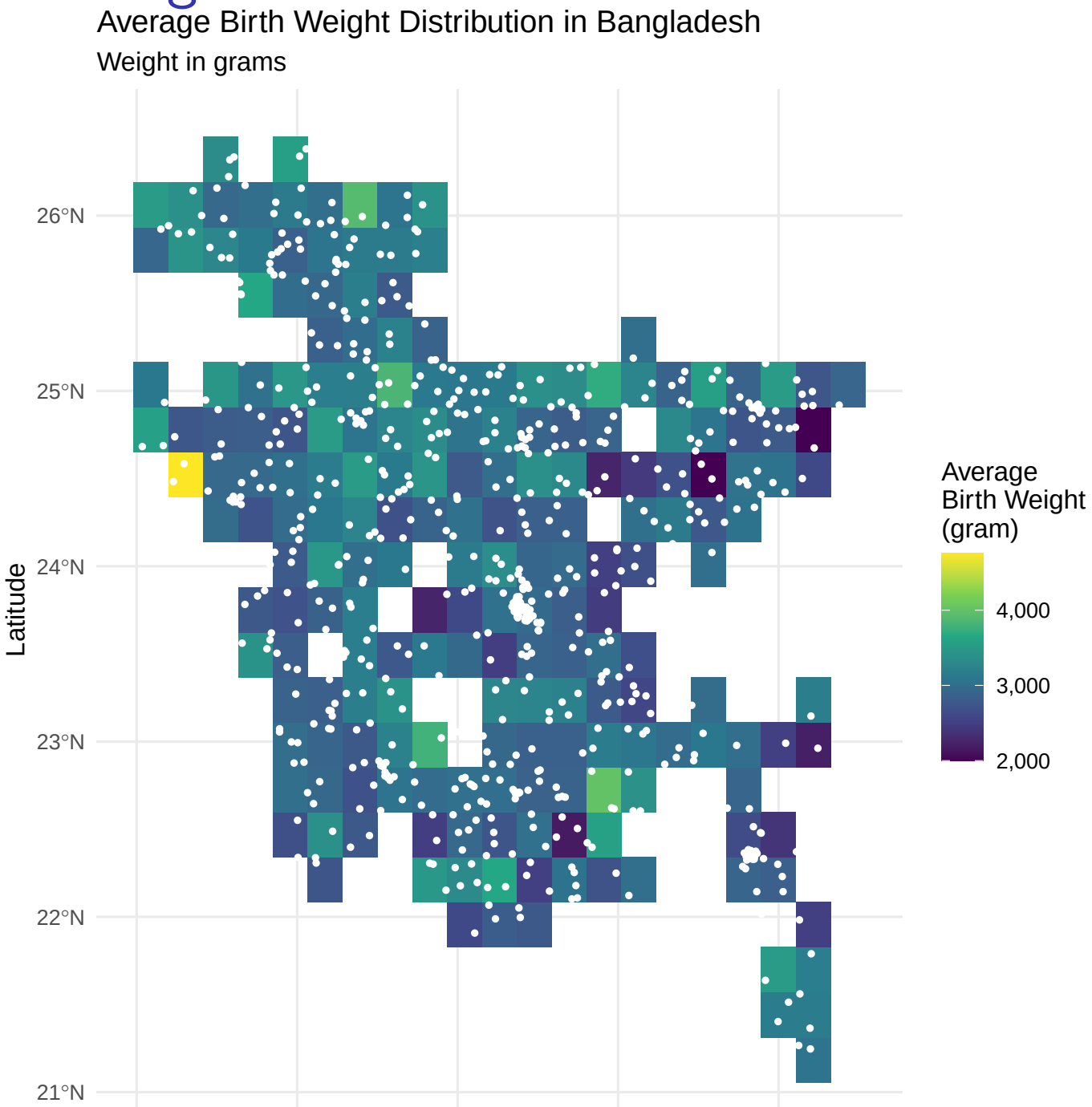
Wealth Index



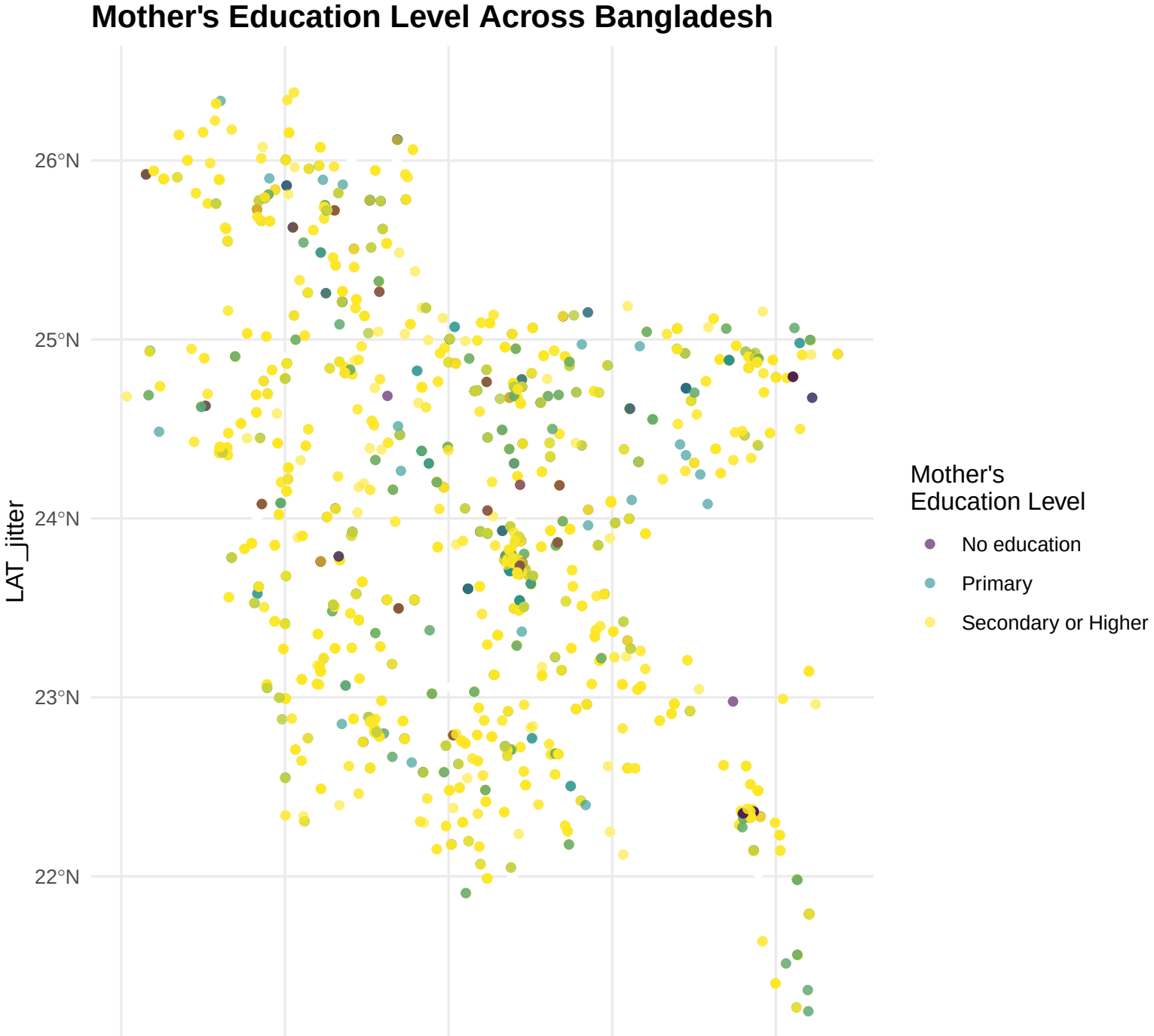
Wealth Index

The data shows both large-scale trends (visible in heat maps) and local variation Urban-rural differences create discontinuities in the spatial pattern River systems in Bangladesh might create natural boundaries/corridors for spatial correlation The jittering adds some noise to the spatial relationships but shouldn't significantly affect large-scale patterns

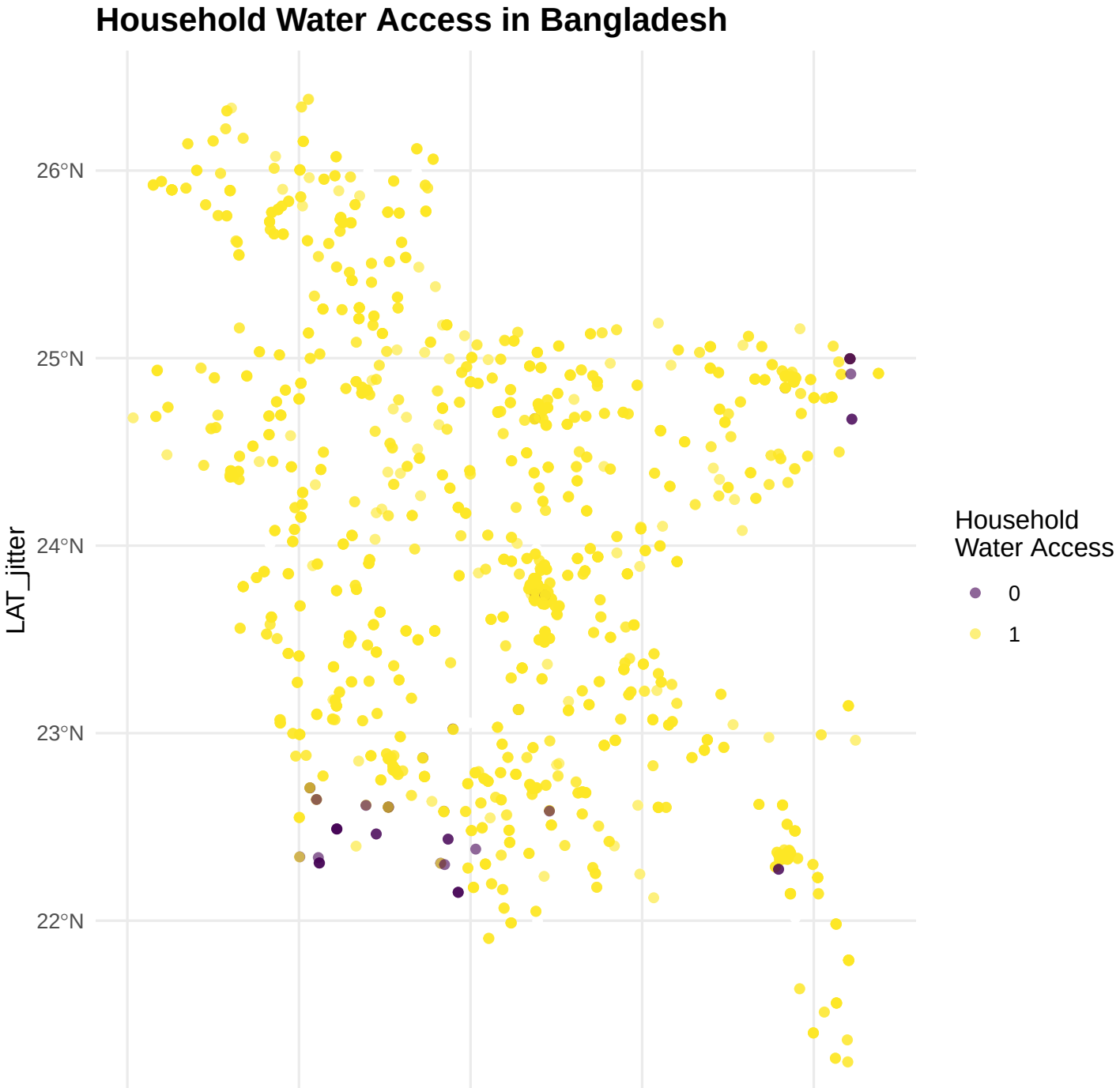
Birth Weight



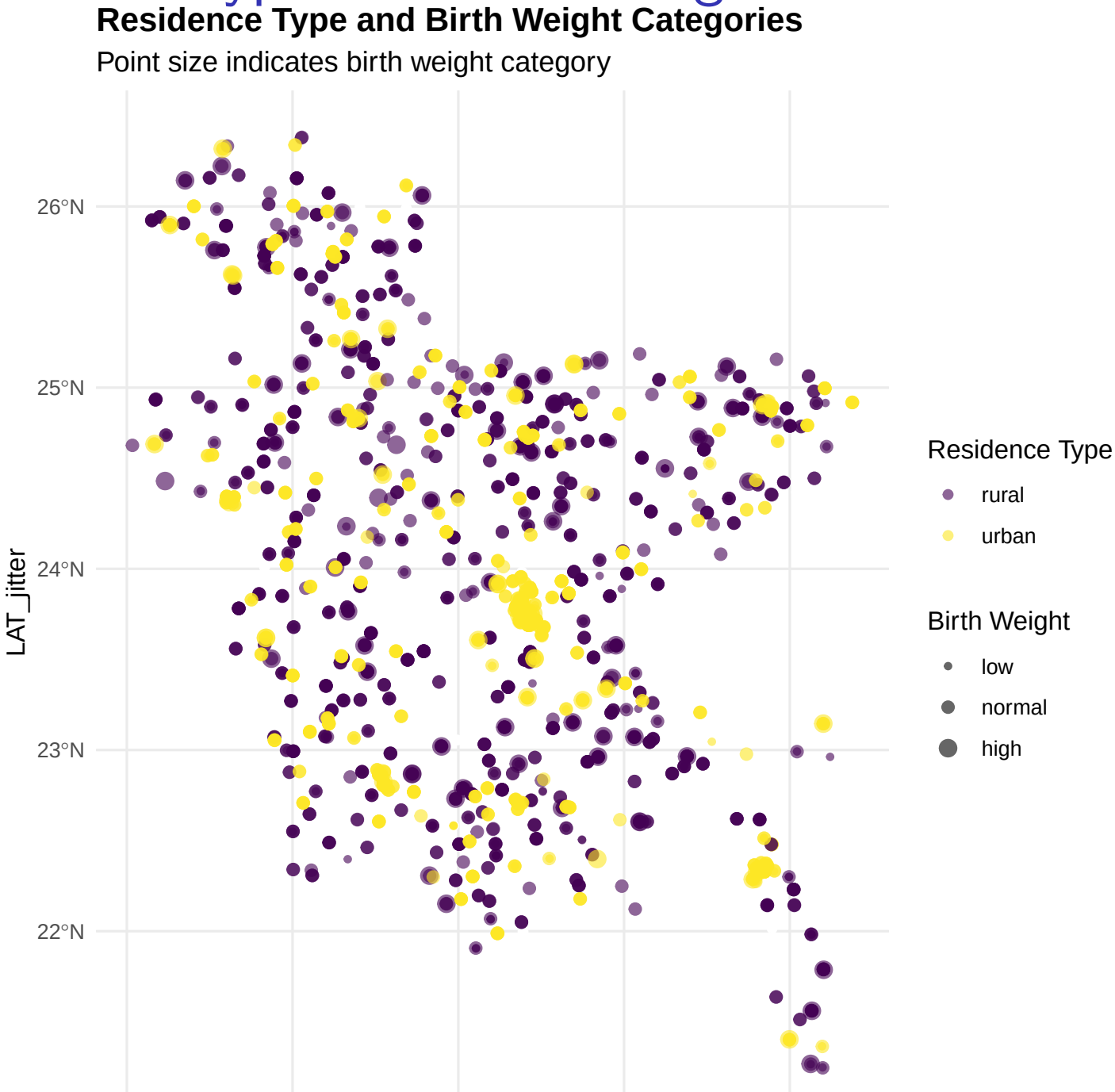
Maternal Education



Household Water Access



Residence Type and Birth Weight



Possible Research Questions

How do birth weights vary between urban and rural areas in Bangladesh? Is there a relationship between household wealth and birth weight outcomes? Do regions with better water access show improved birth weight outcomes? How does mother's education level correlate with birth weight across different regions?

Possible Research Questions

Are there distinct regional clusters of high/low birth weights? Do coastal regions show different patterns compared to inland areas? Is there evidence of spillover effects from urban centers to surrounding rural areas? How do major river systems influence the spatial distribution of birth weights?

Possible Research Questions

How strongly does wealth index predict birth weight when controlling for spatial correlation? Is the relationship between mother's education and birth weight consistent across regions? Do areas with better infrastructure (water access, healthcare facilities) show more consistent birth weights

References

National Institute of Population Research and Training (NIPORT), and ICF. (2020). *Bangladesh Demographic and Health Survey 2017-18*. Dhaka, Bangladesh, and Rockville, Maryland, USA: NIPORT and ICF.