# Undergraduate Independent Research Project Plan

**Student Name**: Namit Shrivastava
**Faculty Mentor**: John Kubale
**Project Duration**: 6/2/2025-
**Project Title**: *Exploring the potential drivers of COVID-19 burden in rural communities*

---

## 1. Project Overview and Research Questions

This independent research project will describe the neighborhood-level characteristics of rural communities during COVID-19 and assess which were associated with greater COVID-19 burden.

**Core Research Questions:**

1. What is known about the impact of COVID-19 in US rural communities and these communities' neighborhood-level characteristics?
2. How can CONEP and NaNDA data be leveraged to examine the neighborhood-level characteristics which are associated with greater COVID-19 burden in these communities, and what analytic approaches are feasible?
3. What do preliminary analyses suggest about these patterns?

---

## 2. Project Timeline and Goals

| Month | Goals | Deliverables |
|---|---|---|
| **Month 1** | **Literature review** on COVID-19 burden in rural communities; **exploration of data sources** used to measure burden (e.g., cases, hospitalizations, mortality), spatial granularity, and neighborhood-level characteristics examined.<br><br>**Develop research questions and analytic plan** based on data available from the COVID Neighborhood Project (CONEP) and the National Neighborhood Data Archive (NaNDA) and any other potential data sources identified | 1. Annotated bibliography with summary of key findings, gaps, and details of data used and methods employed.<br>2. Short memo (~2 pages) summarizing insights from literature and data review<br>3. Draft analytic plan (2-3 pages), outlining research questions, data sources, key variables, and proposed methods. |

| | | |
|---|---|---|
| | in literature review. Identify outcomes of interest, states to examine (and at what spatial level), neighborhood-level characteristics, and analytic approaches (e.g., descriptive analyses, mapping, and statistical modeling). | |
| **Month 2** | **Clean and prepare data** from CONEP and NaNDA for data exploration, visualization, and statistical modeling.<br><br>**Conduct initial data exploration**, descriptive analyses, preliminary data visualization/mapping, and begin fitting preliminary statistical models. | 4. Well commented R code for cleaning and preparing necessary data deposited in GitHub repository.<br>5. Clean preliminary dataset(s).<br>6. Short report 3-5 pages summarizing descriptive analyses, preliminary visualizations, and preliminary results from statistical modeling. The report should also include reflections on feasibility and next steps where applicable.<br>7. Present work in progress at ICPSR Research Forum (could also occur in month 3). |
| **Month 3** | **Statistical modeling and writing:** Build on preliminary analyses and data exploration moving towards final model(s). Compile results from final models and begin drafting into manuscript for submission to peer-reviewed journal. | 8. Cleaned "final" dataset(s).<br>9. Fit final model(s) saving each as a .rds file.<br>10. Decide how best to communicate results from final model(s) (i.e., tables, figures, specific comparisons, etc.)<br>11. Begin drafting manuscript. |

---

## 3. Detailed Activities by Month

**Month 1: Literature Review, Data Exploration, and Analytic Plan**

- Conduct a targeted review of academic and policy literature on:
    - Burden of COVID-19 in rural communities in the US.
        - Data sources used.
        - How rurality defined.
        - States included.

        - Geographic granularity (e.g., state, zip code, county, census tract, etc.
    - What neighborhood or community-level characteristics were described assessed?
        - Broadband access
        - Wealth/disadvantage
        - Race/ethnicity
        - Access to services
    - What gaps remain?
- Review data sources:
    - **Contextual data sources:**
        - Potential data sources:
            - NaNDA
            - Other ACS data
            - Home Mortgage Disclosure Act (HMDA) data (available on NaNDA)
    - **COVID-era data**: infection/hospitalization/mortality rates, unemployment, eviction.
        - Potential data sources:
            - CONEP (we have for 21 states) – 5 at census tract level and the rest at zip code
    - Demographic context data (race/ethnicity, income, housing) from NaNDA.
- Evaluate data by geographic level (census tract, ZIP code, county), time coverage, strengths, and limitations.

- Develop analytic plan specifying:

    - Research questions (including background on why they are needed)

    - Proposed analytic approach:

        - Data sources to use

        - Exposures/outcomes of interest

        - Modeling strategy

        - Anticipated limitations and challenges

**Deliverables**: Annotated bibliography, dataset table, summary memo, analytic plan.

---

**Month 2: Data Cleaning and Exploration**

- Clean and prepare NaNDA and CONEP data needed based on analytic plan.

- Code deposited in GitHub repo.
- Perform initial data exploration including mapping and visualizations where relevant. May include:
    - Summary statistics of neighborhood characteristics.
    - Mapping of key variables to identify patterns.
    - Initial comparison of neighborhoods with different characteristics, over time, levels of urbanicity/rurality.
- Begin fitting initial statistical models (again tracking code using git/GitHub).
- Complete short report summarizing initial findings and any challenges/limitations.
- Start creating slides for ICPSR Research Forum presentation.

**Deliverables**: Clean R code in GitHub repo, clean preliminary dataset(s), short report on exploration and visualizations, present at ICPSR Research Forum (could also occur later).

---

**Month 3: Data Assembly and Manuscript drafting**

- Building on preliminary models begin honing in on final modeling approach.
- Begin summarizing findings and organizing into a coherent narrative for the manuscript.
- Outline manuscript and begin drafting
- Consider most effective way to communicate results
    - What comparisons to make
    - What tables to include
    - What figures to include

**Deliverables**: Preliminary cleaned dataset (if available), descriptive outputs (tables/maps), short findings report.

---

## 4. Mentorship and Skill Development

**Key Skills to Develop:**

- Conducting structured literature reviews and identifying research gaps.
- Locating and evaluating neighborhood-level datasets.
- Conceptualizing a research and analytic plan.
- Dataset assembly and analysis.

- Data visualization
- Communicating research findings and challenges clearly.

- Scientific writing.

**Mentorship Plan:**

- **Weekly 30-60 minute meetings** to review progress, provide guidance, and troubleshoot issues.
- Feedback on drafts of all deliverables (literature summary, analytic plan, analysis reports).
- Hands-on support with data access and preliminary analysis steps as needed.

---

## 5. Final Deliverables (End of Month 3)

1. **Annotated bibliography** and literature synthesis (~2-3 pages).
2. **Dataset and measures table** summarizing potential data sources.
3. **Draft analytic plan** (2-3 pages) including study questions, data, methods, and anticipated challenges.
4. **Preliminary dataset(s)**.
5. **Initial descriptive tables, maps, or figures** visualizing neighborhood-level patterns.
6. **Short summary report** (~3-5 pages) synthesizing preliminary work.

7. **Slide deck** for ~30 minute presentation at ICPSR Research Forum.

8. **Clean, well-documented code** for dataset cleaning/preparation, data visualizations, and statistical analyses.

9. **Draft manuscript** outline that is being filled in (i.e., you don't need to have the entire manuscript drafted).

---