UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

BSD2333 DATA WRANGLING

TITLE: HEALTHCARE IN INDIA

STUDENT ID	NAME	SECTION
SD22002	ALMIRA DAMIA BINTI SYAHNIZAM	02G
SD22019	NUR SABIHAH BINTI ANUAR	
SD22030	TUAN NURSHAFIEKA WAHIDA BINTI TUAN NADIN	
SD22038	NURUL SYAFIQAH NATASHA BINTI MOHD RAZI	
SD22039	NURIN NADHIRAH IZZAH BINTI ZAINUDDIN	

TABLE OF CONTENTS

1.0 SYNOPSIS

1.1 Description of The Assignment

The India Annual Health Survey (AHS) 2012-13 is a significant initiative aimed at comprehensively evaluating the health and welfare of populations in nine states: Uttarakhand, Rajasthan, Uttar Pradesh, Bihar, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, and Assam. These states are recognized as Empowered Action Group (EAG) states due to their notable contribution to crucial demographic metrics such as births, infant mortality, under-5 mortality, and maternal mortality nationwide.

This survey covers a substantial sample size, reaching over 21 million individuals across about 4.32 million households in both urban and rural areas of the aforementioned states. Its primary objective is to compile a comprehensive dataset containing fundamental health indicators, including key statistics like Infant Mortality Rate, Maternal Mortality Ratio, and Total Fertility Rate. Moreover, it aims to shed light on various factors influencing population health, with a specific emphasis on Reproductive and Child Health.

The dataset comprises detailed information on 26 key indicators spanning diverse aspects of health, demographic attributes, education, employment status, disability, illnesses, fertility patterns, family planning practices, maternal and child healthcare, and mortality rates. These indicators play a pivotal role in understanding the health status of the population, monitoring trends over time, and crafting evidence-based policies and interventions to address health inequalities and enhance overall well-being.

The inclusion of unit-level mortality data, confidence intervals for significant indicators, and recognition from the Department of Health and Family Welfare, Government of India, highlight the credibility and importance of this dataset for public health research, policy formulation, and program assessment.

1.2 Problem To Be Solved

The project aims to conduct a comprehensive analysis of the India Annual Health Survey (AHS) 2012-13 dataset, focusing on three main areas: trend analysis, regional disparities, and correlates of health outcomes. Firstly, we will analyze trends in vital health indicators over the survey years to identify patterns, improvements, or regressions in health outcomes. Secondly, we will investigate regional disparities within and across the nine states surveyed, highlighting regions or districts with the highest and lowest health indicators and exploring potential reasons for the disparities. Lastly, we will dive into the factors associated with variations in health outcomes across districts, including socio-economic, demographic, and environmental factors, to understand their correlation with higher or lower health indicators. By addressing these aspects, the project seeks to provide insights into public health challenges and inform evidence-based interventions for improved health outcomes in the surveyed regions.

1.3 Research Questions

1. How have key health indicators such as vaccination coverage, disease prevalence, and treatment patterns evolved across different states as reported in the India Annual Health Survey?
2. How do environmental factors, such as access to clean water, sanitation facilities, and healthcare services, correlate with health outcomes in the nine EAG states?
3. What are the differences in health outcomes and healthcare access between rural and urban populations in the nine Empowered Action Group (EAG) states?
4. What is the impact of public health interventions on vaccination rates, disease management, and overall health awareness among women in the EAG states?

1.4 Objectives

1. To analyze the evolution of key health indicators such as vaccination coverage, disease prevalence, and treatment patterns across the nine EAG states as reported in the India Annual Health Survey 2012-2013.

2. To investigate the correlation between environmental factors such as access to clean water, sanitation facilities, healthcare services, and health outcomes in the nine EAG states.

3. To compare health outcomes and healthcare access between rural and urban populations in the nine EAG states.

4. To evaluate the impact of public health interventions on vaccination rates, disease management, and health awareness among women in the EAG states.

1.5 Basic description of data

This dataset contains 9 states, 286 districts, and other attributes. Below is the description of the attributes in the dataset:

No	Attributes	Description	Data Type
1	State_Name	The names of states in India	Qualitative
2	District_Name	The names of districts in India	Qualitative
3	Population Rural	Population people in rural area of India	Quantitative
4	Population Urban	Population people in Urban area of India	Quantitative
5	Children Received Polio Birth Rural	Children received poliomyelitis in rural area	Quantitative

6	Children Received Polio Birth Urban	Children received poliomyelitis in urban area	Quantitative
7	Children Did Not Receive Any Vaccination Rural	Children did not received any vaccination in rural area	Quantitative
8	Children Did Not Receive Any Vaccination Urban	Children did not received any vaccination in urban area	Quantitative
9	Children Received Measles Vaccine Rural	Children received measles vaccine in rural area	Quantitative
10	Children Received Measles Vaccine Urban	Children received measles vaccine in urban area	Quantitative
11	Children Fully Immunized Rural	Children fully immunized in rural area	Quantitative
12	Children Fully Immunized Urban	Children fully immunized in urban area	Quantitative
13	Children Received Bcg Rural	Children received Bacille Calmette-Guerin (BCG) in rural area	Quantitative
14	Children Received Bcg Urban	Children received Bacille Calmette-Guerin (BCG) in urban area	Quantitative
15	Diarrhea Dysentery Male Rural	Diarrhea dysentery for male in rural area	Quantitative
16	Diarrhea Dysentery Male Urban	Diarrhea dysentery for male in urban area	Quantitative
17	Diarrhea Dysentery Female Rural	Diarrhea dysentery for female in rural area	Quantitative
18	Diarrhea Dysentery Female Urban	Diarrhea dysentery for female in urban area	Quantitative
19	Respiratory Infection Male Rural	Respiratory infection for male in rural area	Quantitative
20	Respiratory Infection Male Urban	Respiratory infection for male in urban area	Quantitative
21	Respiratory Infection Female Rural	Respiratory infection for female	Quantitative

		in rural area	
22	Respiratory Infection Female Urban	Respiratory infection for female in urban area	Quantitative
23	Fever Male Rural	Fever for male in rural area	Quantitative
25	Fever Male Urban	Fever for male in urban area	Quantitative
25	Fever Female Rural	Fever for female in rural area	Quantitative
26	Fever Female Urban	Fever for female in urban area	Quantitative
27	Acute Illness Male Rural	Acute illness for male in rural area	Quantitative
28	Acute Illness Male Urban	Acute illness for male in urban area	Quantitative
29	Acute Illness Female Rural	Acute illness for female in rural area	Quantitative
30	Acute Illness Female Urban	Acute illness for female in urban area	Quantitative
31	Acute Illness And Treatment From Any Source Rural	Acute illness and treatment from any source in rural area	Quantitative
32	Acute Illness And Treatment From Any Source Urban	Acute illness and treatment from any source in urban area	Quantitative
33	Acute Illness And Taking Treatment From Any Source Male Rural	Acute illness and taking treatment from any source for male in rural area	Quantitative
34	Acute Illness And Taking Treatment From Any Source Male Urban	Acute illness and taking treatment from any source for male in urban area	Quantitative
35	Acute Illness And Taking Treatment From Any Source Female Rural	Acute illness and taking treatment from any source for female in rural area	Quantitative
36	Acute Illness And Taking Treatment From Any Source Female Urban	Acute illness and taking treatment from any source for female in urban area	Quantitative
37	Acute Illness And Taking Treatment From Government Source Person	Acute illness and taking treatment from government	Quantitative

	Rural	source person in rural area	
38	Acute Illness And Taking Treatment From Government Source Person Urban	Acute illness and taking treatment from government source person in urban area	Quantitative
39	Acute Illness And Taking Treatment From Government Male Rural	Acute illness and taking treatment from government for male in rural area	Quantitative
40	Acute Illness And Taking Treatment From Government Male Urban	Acute illness and taking treatment from government for male in urban area	Quantitative
41	Acute Illness And Taking Treatment From Government Female Rural	Acute illness and taking treatment from government for female in rural area	Quantitative
42	Acute Illness And Taking Treatment From Government Female Urban	Acute illness and taking treatment from government for female in urban area	Quantitative
43	Women Aware Of Hiv And Aids Rural	Women who are aware of HIV and AIDS in rural area	Quantitative
44	Women Aware Of Hiv And Aids Urban	Women who are aware of HIV and AIDS in urban area	Quantitative
45	Women Aware Of Rti And Sti Rural	Women who are aware of reproductive tract infections (RTI) and sexually transmitted infections (STI) in rural area	Quantitative
46	Women Aware Of Rti And Sti Urban	Women who are aware of reproductive tract infections (RTI) and sexually transmitted infections (STI) in urban area	Quantitative
47	Women Aware Of Haf Ors And Ort And Zinc And Rural	Women who are aware of Home Available Fluids (HAF), Oral Rehydration Solution (ORS), Oral Rehydration Therapy (ORT) and zinc in rural area	Quantitative
48	Women Aware Of Haf Ors And Ort And Zinc And Urban	Women who are aware of Home Available Fluids (HAF), Oral Rehydration Solution (ORS), Oral Rehydration Therapy (ORT)	Quantitative

		and zinc in urban area	
49	Women Aware Of Ari And Pneumonia Rural	Women who are aware of acute respiratory infection (ARI) and Pneumonia in rural area	Quantitative
50	Women Aware Of Ari And Pneumonia Urban	Women who are aware of acute respiratory infection (ARI) and Pneumonia in rural area	Quantitative

2.0 Packages Required

Pandas
Description: Pandas is a powerful and flexible open-source data analysis and manipulation library for Python. It provides data structures and functions needed to work with structured data seamlessly.

Usage: Pandas is used for data manipulation and cleaning, reading and writing data from various file formats such as csv handling and performing time series analysis.

NumPy
Description: NumPy is a fundamental package for scientific computing in Python. It provides support for arrays, matrices, and many mathematical functions to operate on these data structures.

Usage: NumPy is used for performing mathematical and logical operations on arrays.

Matplotlib
Description: Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension, NumPy. It is used to generate visualizations such as scatter plot, boxplot and bar chart.

Usage: Matplotlib is used for creating static, interactive, and animated visualizations in Python. It can generate plots, histograms, bar charts and scatter plots.

Seaborn
Description: Seaborn is a Python visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. It integrates closely with Pandas data structures.

Usage: Seaborn is used for making complex plots easier to create, enhancing visualizations with themes, color palettes, and data-specific plots such as heatmaps.

Plotly
Description: Plotly is an interactive graphing library for Python. It is well-suited for creating elaborate and interactive visualizations that can be displayed in a web browser.

Usage: Plotly is used for creating interactive plots, dashboards, and visualizations. It supports a wide variety of chart types with interactive capabilities like zooming and hover effects.

ipywidgets
Description: ipywidgets is a library that provides interactive HTML widgets for Jupyter notebooks and the IPython kernel.

Usage: ipywidgets is used for adding interactive controls, such as sliders, buttons, and dropdowns to Jupyter notebooks. This allows users to create interactive data visualizations and applications within notebooks.

plotly-express
Description: plotly-express is a high-level data visualization library for Plotly. It simplifies the process of creating complex visualizations with less code.

Usage:plotly-express is used for quickly generating various types of visualizations with a minimal amount of code.

ipython
Description: IPython is an interactive command-line interface for Python. It offers enhanced introspection, rich media, shell syntax, tab completion, and more.

Usage: IPython is used for interactive computing.

Random
Description: random is a built-in Python module that implements pseudo-random number generators for various distributions.

Usage: This function can be used when you need to assign random colors to different elements in a visualization, such as points in a scatter plot, bars in a bar chart, or segments in a pie chart.

3.0 Data Preparation

3.1 Data Cleaning

Data cleaning is a crucial step in order to get accurate and valid results. In this process we identified columns that we need for our project from the dataset we obtained from kaggle (link kaggle). Firstly, we structure the data by joining the columns using a technique called joins. Next we normalize the data by reducing inconsistency and cleaning unused data in the excel. After that, we uploaded the dataset into jupyter and started our cleaning process and checked if there are any null values that we need to imply imputation. The figures will show the steps we did in order to get a clean dataset.

Step 1: Importing dataset in jupyter

[4]:
```
import pandas as pd
```

[5]:
```
health=pd.read_csv("DataWrangling_datasets.csv")
```

	State_Name	State_District_Name	Population_Rural	Population_Urban	Children Received Polio Birth Rural	Children Received Polio Birth Urban	Children Did Not Receive Any Vaccination Rural	Children Did Not Receive Any Vaccination Urban	Children Received Measles Vaccine Rural	Children Received Measles Vaccine Urban	...	Acute Illness An Taking Treatment From Governmen FemaleRural	Ac Illness Tal Treatm F Governm Fen Ur
0	Assam	Barpeta	60293	4313	67.20	74.50	4.20	5.50	76.70	71.80	...	34.60	2
1	Assam	Bongaigaon	73899	10113	71.90	91.70	3.50	2.10	79.00	84.40	...	45.50	3
2	Assam	Cachar	112171	14987	80.10	91.30	2.60	2.90	80.30	79.90	...	31.10	3
3	Assam	Darrang	28256	792	75.70	NaN	5.40	NaN	71.10	NaN	...	54.70	I
4	Assam	Dhemaji	66537	8914	78.50	78.60	7.60	6.00	77.00	83.00	...	55.40	4
...	...	...	...	...	...	...	...	...	...	...	...	...	
279	Uttarakhand	Pithoragarh	84627	18372	66.41	93.56	1.43	0.40	92.45	97.18	...	51.39	2
280	Uttarakhand	Rudraprayag	138266	759	59.35	NaN	1.68	NaN	91.24	NaN	...	48.39	I
281	Uttarakhand	Tehri Garhwal	181918	23657	76.63	93.27	6.13	2.66	81.18	89.05	...	27.89	2
282	Uttarakhand	Udham Singh Nagar	64614	26904	74.89	79.77	2.60	3.25	89.53	86.79	...	18.58	1
283	Uttarakhand	Uttarkashi	67984	8664	86.48	94.09	7.35	3.76	83.17	89.25	...	24.62	3

Figure 3.1.1 Dataset of project

Step 2: Checking the sum of null values in each columns

[7]:
```
health.isnull().sum()
```

```
[7]: State_Name                                                         0
     State_District_Name                                                0
     Population_Rural                                                   0
     Population_Urban                                                   0
     Children Received Polio Birth Rural                                0
     Children Received Polio Birth Urban                               45
     Children Did Not Receive Any Vaccination Rural                     0
     Children Did Not Receive Any Vaccination Urban                    45
     Children Received Measles Vaccine Rural                            0
     Children Received Measles Vaccine Urban                           45
     Children Fully Immunized Rural                                     0
     Children Fully Immunized Urban                                    45
     Children Received Bcg Rural                                        0
     Children Received Bcg Urban                                       45
     Diarrhoea Dysentery Male Rural                                     0
     Diarrhoea Dysentery Male Urban                                    45
     Diarrhoea Dysentery Female Rural                                   0
     Diarrhoea Dysentery Female Urban                                  45
     Respiratory Infection Male Rural                                   0
     Respiratory Infection Male Urban                                  44
     Respiratory Infection Female Rural                                 0
     Respiratory Infection Female Urban                                44
     Fever Male Rural                                                   0
     Fever Male Urban                                                  44
     Fever Femal Rural                                                  0
     Fever FemaleUrban                                                 44

     Acute Illness Male Rural                                           0
     Acute Illness Male Urban                                          44
     Acute Illnes Female Rural                                          0
     Acute Illness Female Urban                                        44
     Acute Illness And Treatment From Any Source Rural                  0
     Acute Illness And Taking Treatment From Any Source Urban          44
     Acute Illness And Taking Treatment From Any Source Male Rural      0
     Acute Illness And Taking  Treatment From Any Source  Male Urban   44
     Acute Illness And Taking Treatment From Any Source Female Rural    0
     Acute Illness And Taking Treatment From Any Source Female Urban   44
     Acute Illness And Taking Treatment From Government Source Person Rural    0
     Acute Illness And Taking Treatment From Government Person Urban   44
     Acute Illness And Taking Treatment From Government Male Rural      0
     Acute Illness And Taking Treatment From Government Sourc Male Urban      44
     Acute Illness An Taking Treatment From Governmen FemaleRural       0
     Acute Illness And Taking Treatment From Government Female Urban   44
     Women Aware Of Hiv And Aids Rural                                  0
     Womene Aware Of Hiv And Aids Urban                                45
     Women Aware Of Rti And Sti Rural                                   0
     Women Aware Of Rti And Sti Urban                                  45
     Women Aware Of Haf Ors And Ort And Zinc And Rural                  0
     Women Aware Of Haf Ors And Ort And Zinc And Urban                 45
     Women Aware Of Ari And  Pneumonia Rural                            0
     Women Aware Of Ari And  Pneumonia Urban                           45
     dtype: int64
```

Figure 3.1.2 Checking null values

Step 3: Dropping the rows with null values

```
[8]: health.dropna(inplace=True)
```

Step 4: Checking for duplicate values

```
[9]: health.duplicated().sum()

[9]: 0
```

Step 5: Renaming the columns

```
: health.rename(columns={'State_Name': 'State', 'State_District_Name': 'District',
                         'Population_Rural':'Population Rural',
                         'Population_Urban':'Population Urban',
                        'Fever FemaleUrban':'Fever Female Urban',
                        'Womene Aware Of Hiv And Aids Urban':'Women Aware Of Hiv And Aids Urban',
                         'Fever Femal Rural':'Fever Female Rural'}, inplace=True)
```

Figure 3.1.3: *Changing column names*

Step 6: Checking after drop null values

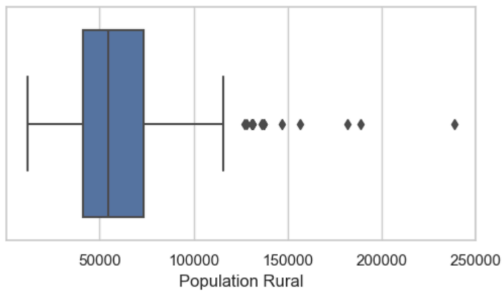
```
: health.isnull().sum()
```

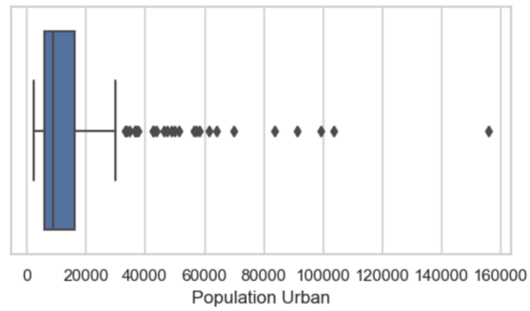
```
State                                          0
District                                       0
Population Rural                               0
Population Urban                               0
Children Received Polio Birth Rural            0
Children Received Polio Birth Urban            0
Children Did Not Receive Any Vaccination Rural 0
Children Did Not Receive Any Vaccination Urban 0
Children Received Measles Vaccine Rural        0
Children Received Measles Vaccine Urban        0
Children Fully Immunized Rural                 0
Children Fully Immunized Urban                 0
Children Received Bcg Rural                    0
Children Received Bcg Urban                    0
Diarrhoea Dysentery Male Rural                 0
Diarrhoea Dysentery Male Urban                 0
Diarrhoea Dysentery Female Rural               0
Diarrhoea Dysentery Female Urban               0
```

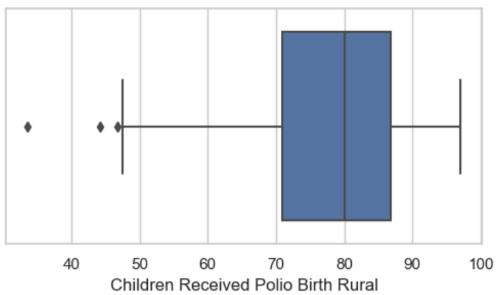
```
Respiratory Infection Male Rural                                      0
Respiratory Infection Male Urban                                      0
Respiratory Infection Female Rural                                    0
Respiratory Infection Female Urban                                    0
Fever Male Rural                                                      0
Fever Male Urban                                                      0
Fever Female Rural                                                    0
Fever Female Urban                                                    0
Acute Illness Male Rural                                              0
Acute Illness Male Urban                                              0
Acute Illnes Female Rural                                             0
Acute Illness Female Urban                                            0
Acute Illness And Treatment From Any Source Rural                     0
Acute Illness And Taking Treatment From Any Source Urban              0
Acute Illness And Taking Treatment From Any Source Male Rural         0
Acute Illness And Taking  Treatment From Any Source  Male Urban       0
Acute Illness And Taking Treatment From Any Source Female Rural       0
Acute Illness And Taking Treatment From Any Source Female Urban       0
Acute Illness And Taking Treatment From Government Source Person Rural    0
Acute Illness And Taking Treatment From Government Person Urban       0
Acute Illness And Taking Treatment From Government Male Rural         0
Acute Illness And Taking Treatment From Government Sourc Male Urban   0
Acute Illness An Taking Treatment From Governmen FemaleRural          0
Acute Illness And Taking Treatment From Government Female Urban       0
Women Aware Of Hiv And Aids Rural                                     0
Women Aware Of Hiv And Aids Urban                                     0
Women Aware Of Rti And Sti Rural                                      0
Women Aware Of Rti And Sti Urban                                      0
Women Aware Of Haf Ors And Ort And Zinc And Rural                     0
Women Aware Of Haf Ors And Ort And Zinc And Urban                     0
Women Aware Of Ari And  Pneumonia Rural                               0
Women Aware Of Ari And  Pneumonia Urban                               0
```

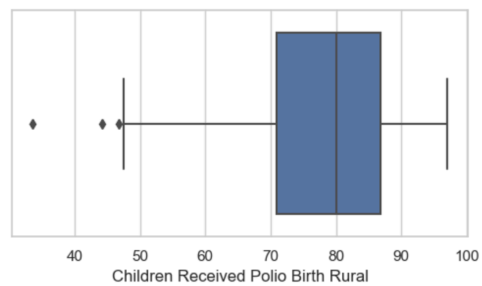
Figure 3.1.4 Checking null values after removing

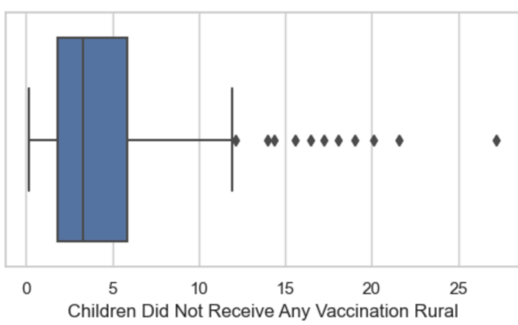
Step 7: Checking outliers for each variable

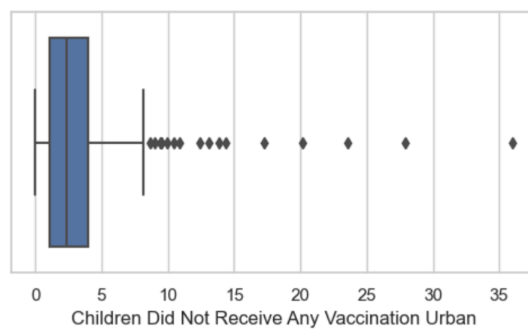


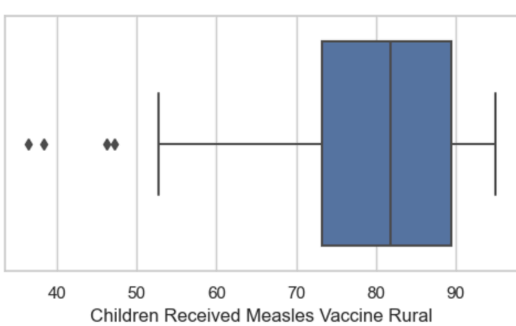


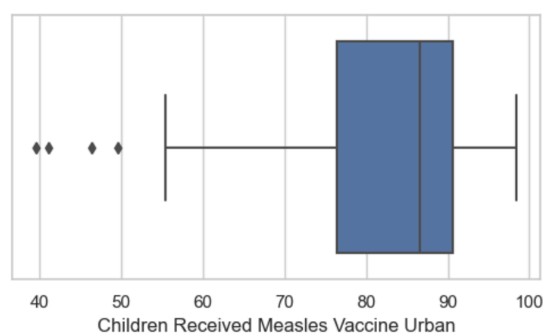


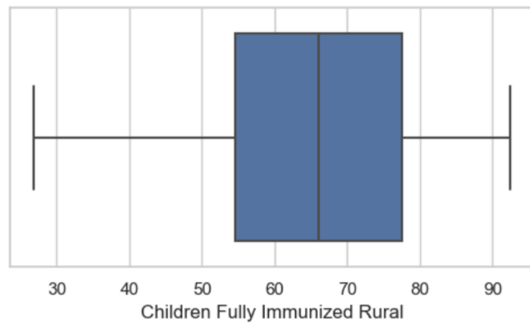


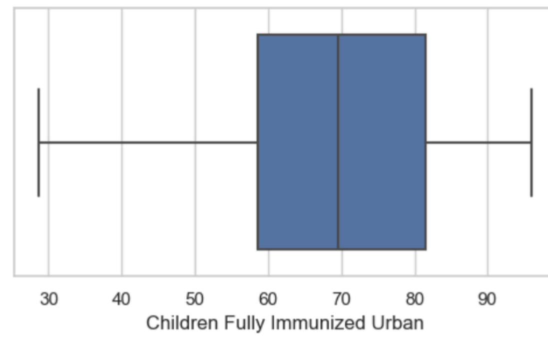


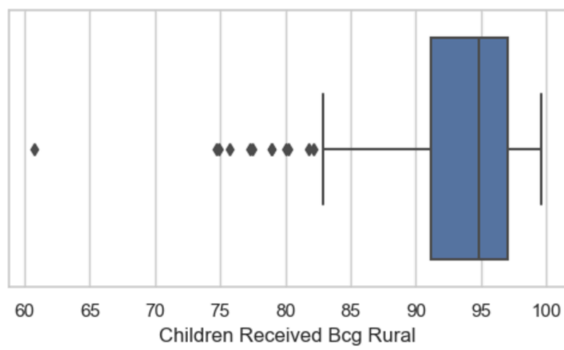


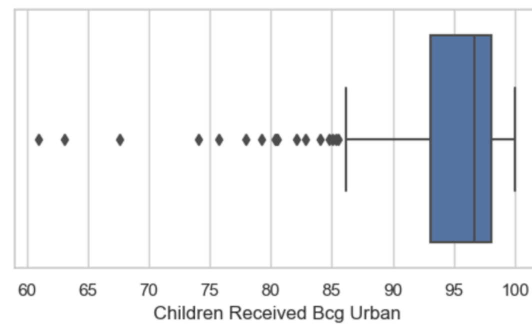


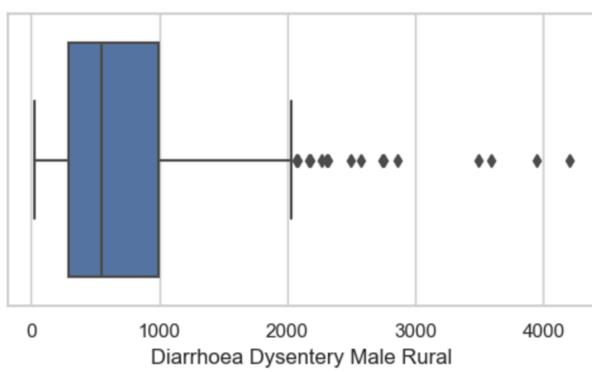


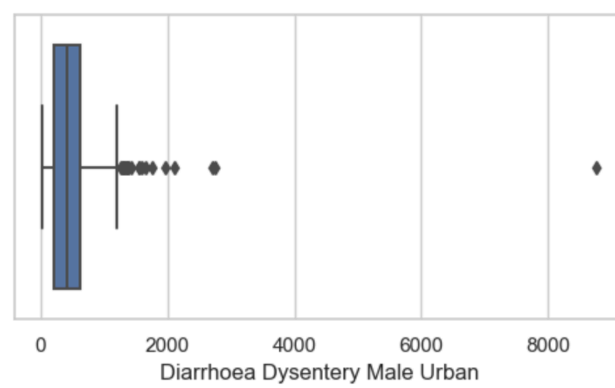

30 40 50 60 70 80 90
Children Fully Immunized Rural

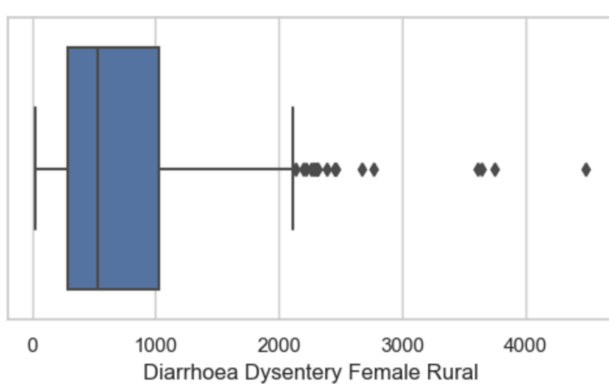

30 40 50 60 70 80 90
Children Fully Immunized Urban


60 65 70 75 80 85 90 95 100
Children Received Bcg Rural

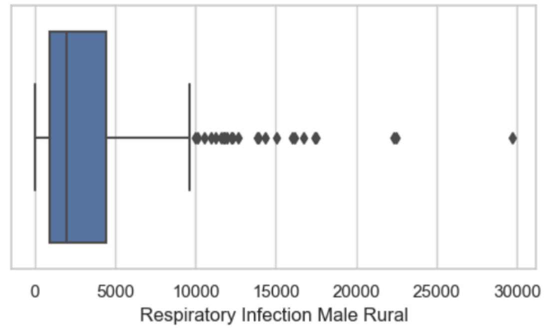

60 65 70 75 80 85 90 95 100
Children Received Bcg Urban

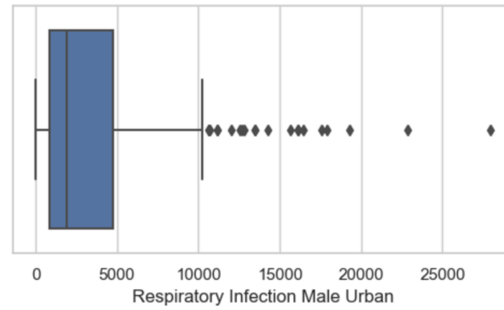

0 1000 2000 3000 4000
Diarrhoea Dysentery Male Rural

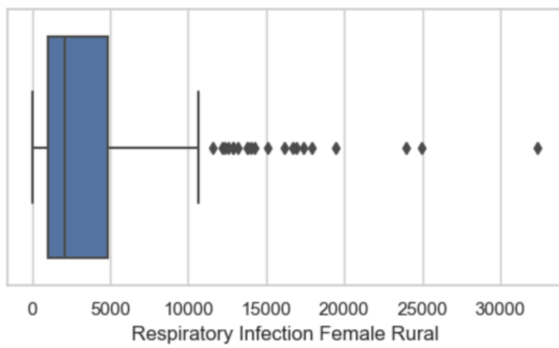

0 2000 4000 6000 8000
Diarrhoea Dysentery Male Urban

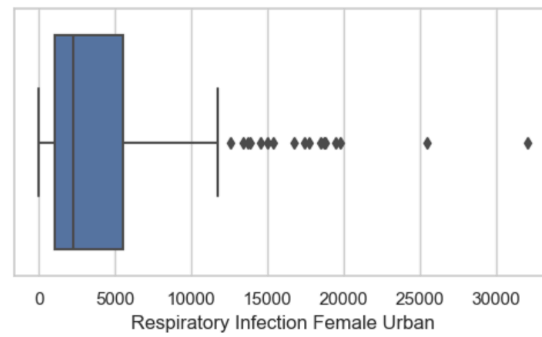

0 1000 2000 3000 4000
Diarrhoea Dysentery Female Rural

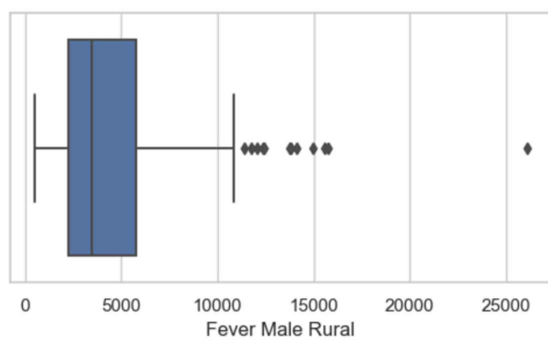

0 1000 2000 3000 4000 5000 6000
Diarrhoea Dysentery Female Urban

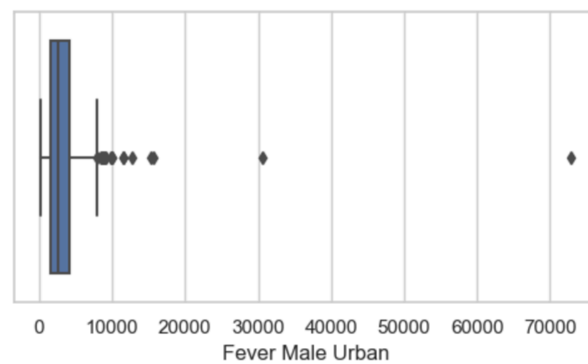

0 5000 10000 15000 20000 25000 30000
Respiratory Infection Male Rural

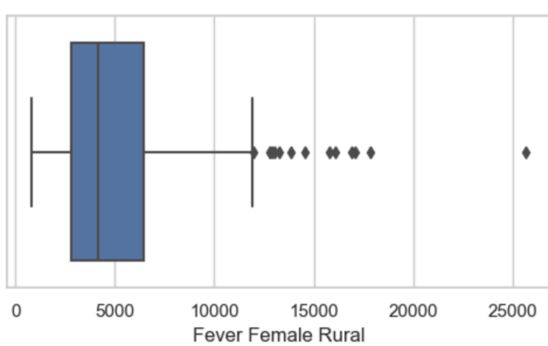

0 5000 10000 15000 20000 25000
Respiratory Infection Male Urban

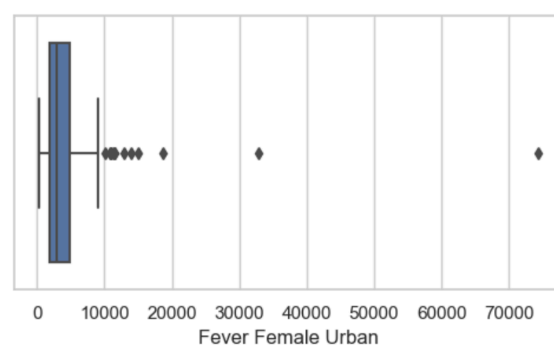

0 5000 10000 15000 20000 25000 30000
Respiratory Infection Female Rural

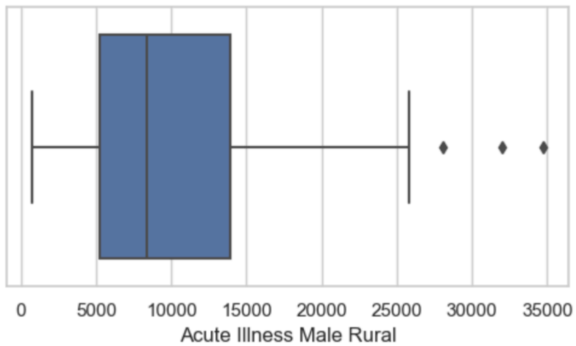

0 5000 10000 15000 20000 25000 30000
Respiratory Infection Female Urban

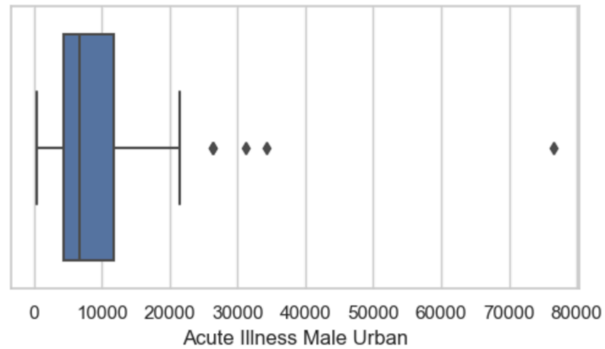

0 5000 10000 15000 20000 25000
Fever Male Rural

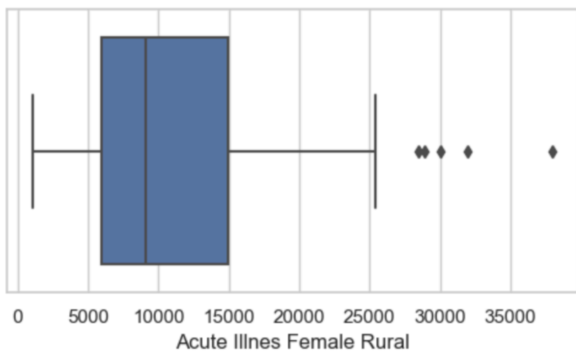

0 10000 20000 30000 40000 50000 60000 70000
Fever Male Urban

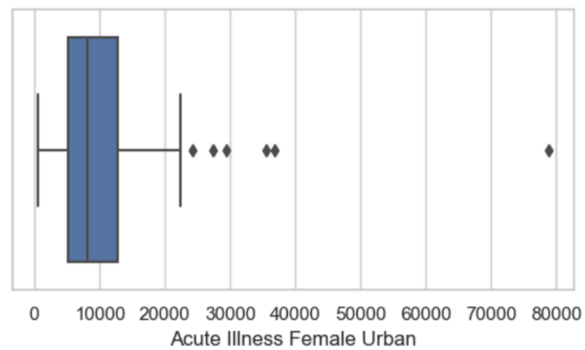

0 5000 10000 15000 20000 25000
Fever Female Rural

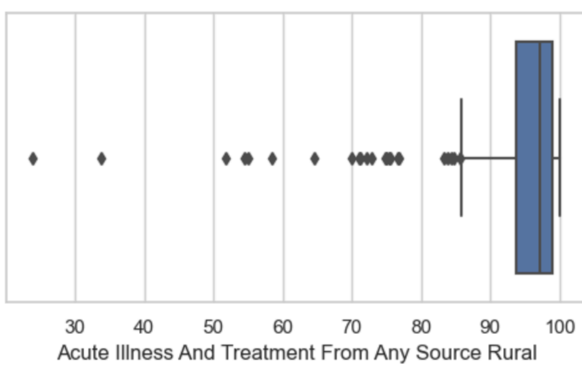

0 10000 20000 30000 40000 50000 60000 70000
Fever Female Urban

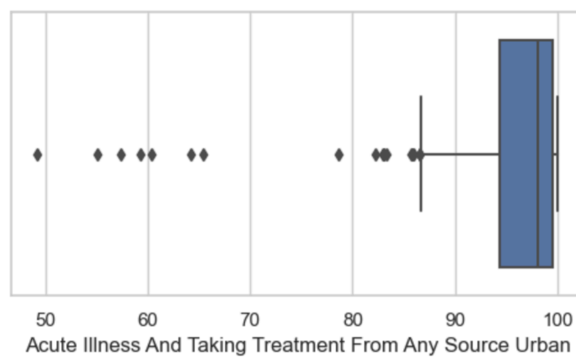

0 5000 10000 15000 20000 25000 30000 35000
Acute Illness Male Rural

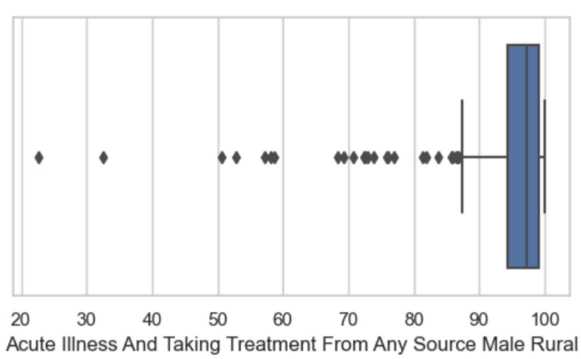

0 10000 20000 30000 40000 50000 60000 70000 80000
Acute Illness Male Urban

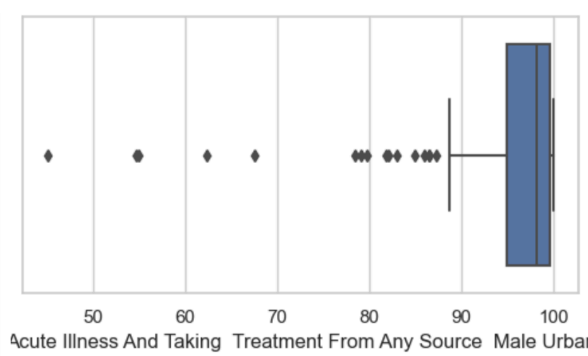

0 5000 10000 15000 20000 25000 30000 35000
Acute Illnes Female Rural

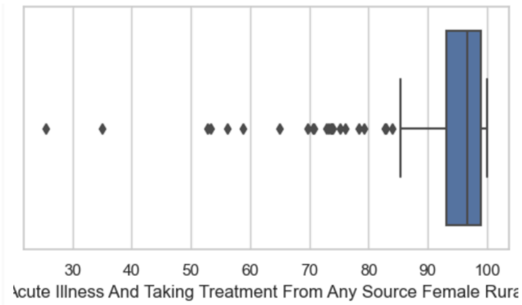

0 10000 20000 30000 40000 50000 60000 70000 80000
Acute Illness Female Urban

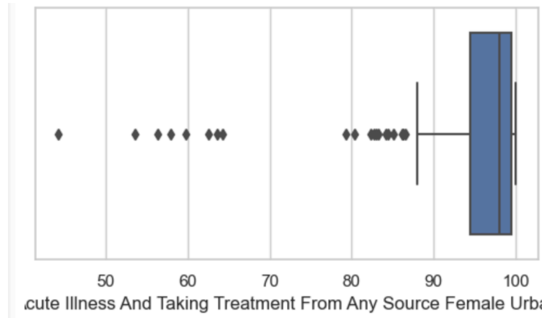

30 40 50 60 70 80 90 100
Acute Illness And Treatment From Any Source Rural

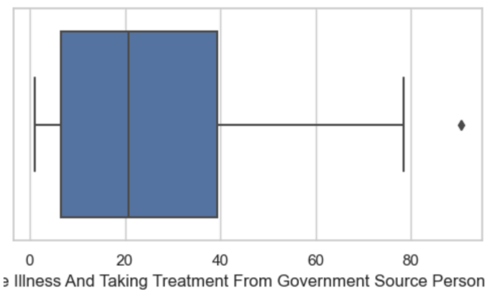

50 60 70 80 90 100
Acute Illness And Taking Treatment From Any Source Urban

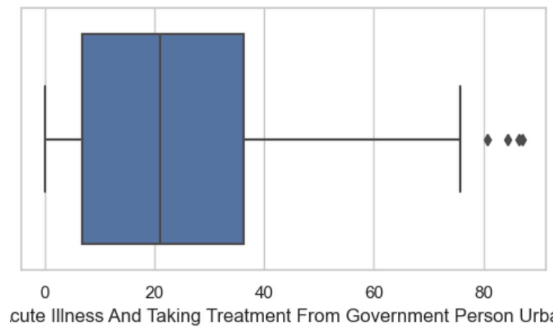

20 30 40 50 60 70 80 90 100
Acute Illness And Taking Treatment From Any Source Male Rural

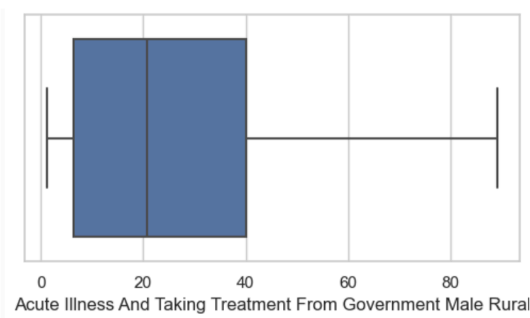

50 60 70 80 90 100
Acute Illness And Taking Treatment From Any Source Male Urbai

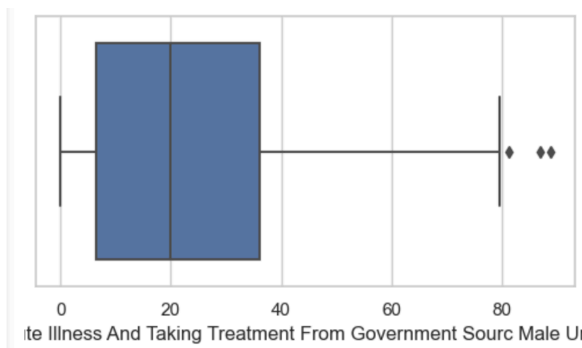

30 40 50 60 70 80 90 100
\cute Illness And Taking Treatment From Any Source Female Rura

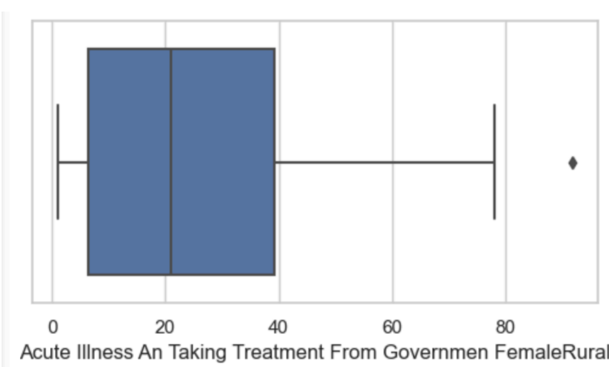

50 60 70 80 90 100
.cute Illness And Taking Treatment From Any Source Female Urba

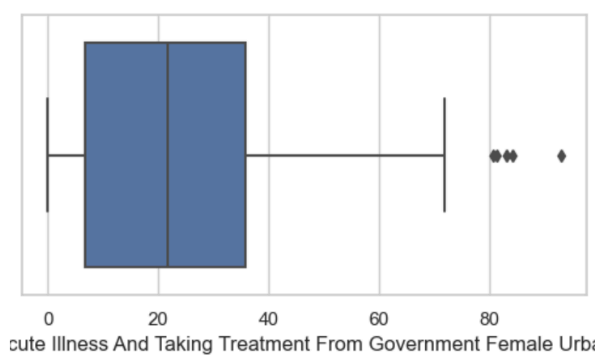

0 20 40 60 80
e Illness And Taking Treatment From Government Source Person

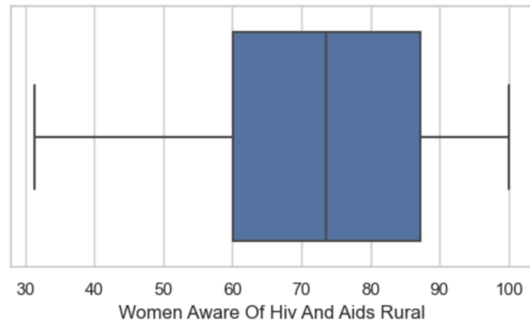

0 20 40 60 80
cute Illness And Taking Treatment From Government Person Urba

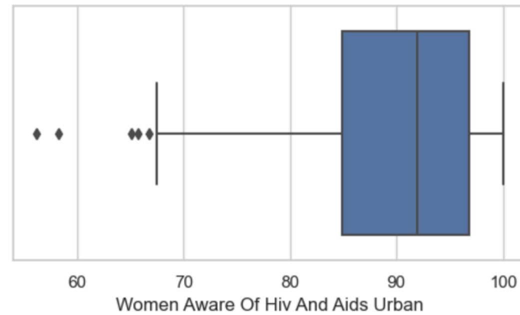

0 20 40 60 80
Acute Illness And Taking Treatment From Government Male Rural

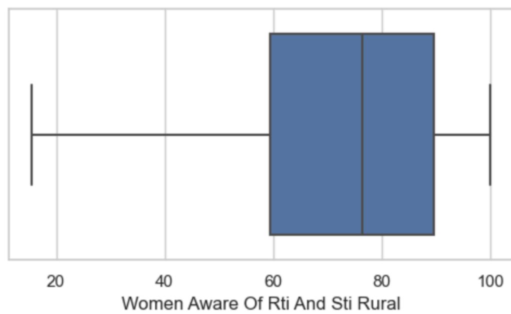

0 20 40 60 80
ite Illness And Taking Treatment From Government Sourc Male Ur

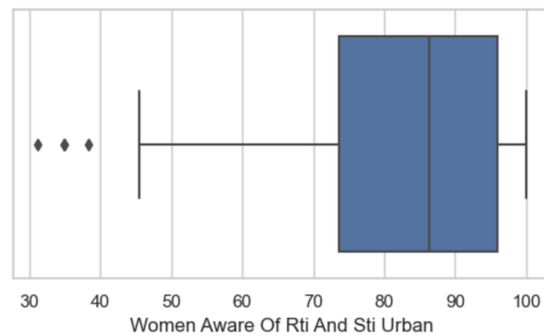

0 20 40 60 80
Acute Illness An Taking Treatment From Governmen FemaleRural

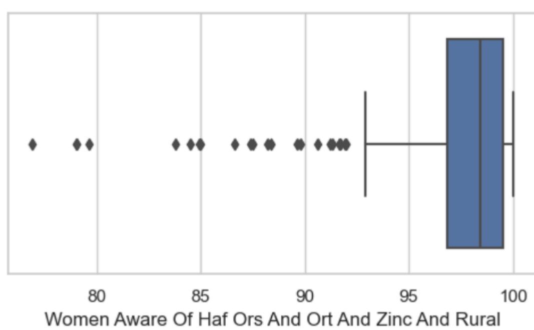

0 20 40 60 80
cute Illness And Taking Treatment From Government Female Urba

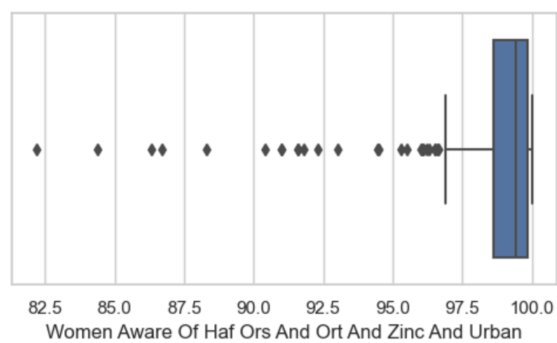

30 40 50 60 70 80 90 100
Women Aware Of Hiv And Aids Rural

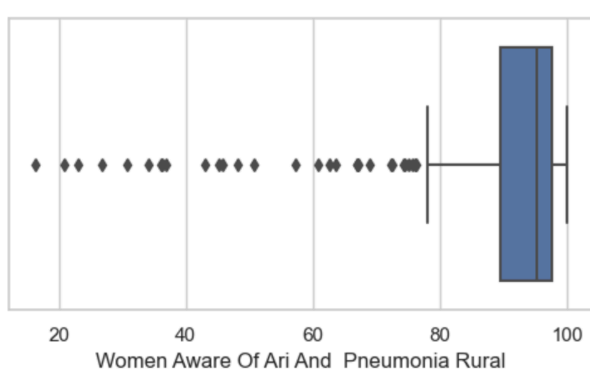

60 70 80 90 100
Women Aware Of Hiv And Aids Urban

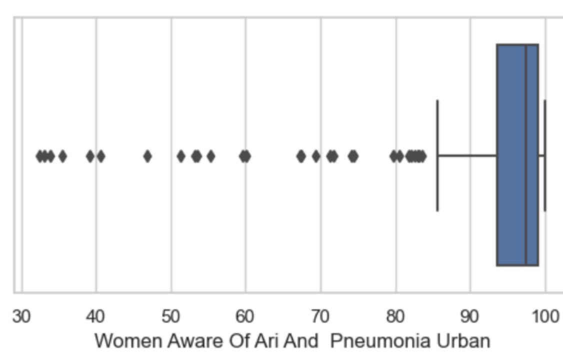

20 40 60 80 100
Women Aware Of Rti And Sti Rural


30 40 50 60 70 80 90 100
Women Aware Of Rti And Sti Urban


80 85 90 95 100
Women Aware Of Haf Ors And Ort And Zinc And Rural


82.5 85.0 87.5 90.0 92.5 95.0 97.5 100.0
Women Aware Of Haf Ors And Ort And Zinc And Urban


20 40 60 80 100
Women Aware Of Ari And Pneumonia Rural


30 40 50 60 70 80 90 100
Women Aware Of Ari And Pneumonia Urban

From the boxplot that has been provided, we can see that there are a few that have outliers. However, in this project we do not remove the outliers because we are not going to make any prediction, we only want to see the relationship between the variables.

Step 8: Uploading the clean file into laptop

```
[16]: health.to_csv(r"D:\UMP\SEMESTER 4\DATA WRANGLING\health_cleaned.csv", index=False)
```

3.2 Data Preview

After cleaning the dataset, it is crucial to conduct a data preview to understand its structure, key characteristics, and distribution before analysis. This involves examining the number of rows and columns, data types, and presence of missing values using methods like info(). Previewing the first few rows with head() ensures the data has been cleaned properly. Analyzing unique values in categorical columns and descriptive statistics for numerical columns provides insights into data diversity and distribution. Descriptive statistics offer a quantitative summary. Overall, a data preview post-cleaning ensures the dataset's integrity and informs subsequent analysis decisions efficiently.

Step 1 : Upload cleaned dataset in Jupyter

```
[1]: import pandas as pd
```

```
[4]: file_path = 'health_cleaned.csv'
     df = pd.read_csv(file_path)
```

[6]:

	State	District	Population Rural	Population Urban	Children Received Polio Birth Rural	Children Received Polio Birth Urban	Children Did Not Receive Any Vaccination Rural	Children Did Not Receive Any Vaccination Urban	Children Received Measles Vaccine Rural	Children Received Measles Vaccine Urban	...	Acute Illness An Taking Treatment From Governmen FemaleRural	Acute Illness And Taking Treatment From Government Female Urban	Women Aware Of Hiv And Aids Rural	Wome Aw Of I A A Urb
0	Assam	Barpeta	60293	4313	67.20	74.50	4.20	5.50	76.70	71.80	...	34.60	26.00	63.50	83
1	Assam	Bongaigaon	73899	10113	71.90	91.70	3.50	2.10	79.00	84.40	...	45.50	32.50	59.80	91
2	Assam	Cachar	112171	14987	80.10	91.30	2.60	2.90	80.30	79.90	...	31.10	36.70	71.20	87
3	Assam	Dhemaji	66537	8914	78.50	78.60	7.60	6.00	77.00	83.00	...	55.40	44.10	82.20	92
4	Assam	Dhubri	46337	4655	59.10	88.30	8.40	2.90	59.90	84.50	...	24.10	50.00	46.20	97
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
234	Uttarakhand	Pauri Garhwal	188897	24187	87.38	87.72	1.24	1.10	90.00	90.55	...	25.72	23.46	92.40	96
235	Uttarakhand	Pithoragarh	84627	18372	66.41	93.56	1.43	0.40	92.45	97.18	...	51.39	24.86	95.41	99
236	Uttarakhand	Tehri Garhwal	181918	23657	76.63	93.27	6.13	2.66	81.18	89.05	...	27.89	21.68	86.19	96
237	Uttarakhand	Udham Singh Nagar	64614	26904	74.89	79.77	2.60	3.25	89.53	86.79	...	18.58	15.84	91.55	98
238	Uttarakhand	Uttarkashi	67984	8664	86.48	94.09	7.35	3.76	83.17	89.25	...	24.62	33.22	86.32	98

239 rows × 50 columns

Figure 3.2.1 shows a cleaned dataset

Step 2 : Review the structure and info of dataset

[9]:
```
print("Data Info:")
df.info()
```

```
Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 239 entries, 0 to 238
Data columns (total 50 columns):
 #   Column                                      Non-Null Count  Dtype  
---  ------                                      --------------  -----  
 0   State                                       239 non-null    object 
 1   District                                    239 non-null    object 
 2   Population Rural                            239 non-null    int64  
 3   Population Urban                            239 non-null    int64  
 4   Children Received Polio Birth Rural         239 non-null    float64
 5   Children Received Polio Birth Urban         239 non-null    float64
 6   Children Did Not Receive Any Vaccination Rural  239 non-null    float64
 7   Children Did Not Receive Any Vaccination Urban  239 non-null    float64
 8   Children Received Measles Vaccine Rural     239 non-null    float64
 9   Children Received Measles Vaccine Urban     239 non-null    float64
 10  Children Fully Immunized Rural              239 non-null    float64
 11  Children Fully Immunized Urban              239 non-null    float64
 12  Children Received Bcg Rural                 239 non-null    float64
 13  Children Received Bcg Urban                 239 non-null    float64
 14  Diarrhoea Dysentery Male Rural              239 non-null    float64
 15  Diarrhoea Dysentery Male Urban              239 non-null    float64
 16  Diarrhoea Dysentery Female Rural            239 non-null    float64
 17  Diarrhoea Dysentery Female Urban            239 non-null    float64
 18  Respiratory Infection Male Rural            239 non-null    float64
 19  Respiratory Infection Male Urban            239 non-null    float64
 20  Respiratory Infection Female Rural          239 non-null    float64
 21  Respiratory Infection Female Urban          239 non-null    float64
 22  Fever Male Rural                            239 non-null    float64
 23  Fever Male Urban                            239 non-null    float64
 24  Fever Femal Rural                           239 non-null    float64
 25  Fever FemaleUrban                           239 non-null    float64
 26  Acute Illness Male Rural                    239 non-null    float64
 27  Acute Illness Male Urban                    239 non-null    float64
 28  Acute Illnes Female Rural                   239 non-null    float64
 29  Acute Illness Female Urban                  239 non-null    float64
```

Figure 3.2.2 shows info about the dataset

Step 3 : Preview of first 5 rows

```
[13]: print("\nData Preview (First 5 rows):")
      df.head()
```

```
Data Preview (First 5 rows):
```

[13]:

	State	District	Population Rural	Population Urban	Children Received Polio Birth Rural	Children Received Polio Birth Urban	Children Did Not Receive Any Vaccination Rural	Children Did Not Receive Any Vaccination Urban	Children Received Measles Vaccine Rural	Children Received Measles Vaccine Urban	...	Acute Illness An Taking Treatment From Governmen FemaleRural	Acute Illness And Taking Treatment From Government Female Urban	Women Aware Of Hiv And Aids Rural	Womene Aware Of Hiv And Aids Urban	Wc A C Ar I
0	Assam	Barpeta	60293	4313	67.2	74.5	4.2	5.5	76.7	71.8	...	34.6	26.0	63.5	83.2	
1	Assam	Bongaigaon	73899	10113	71.9	91.7	3.5	2.1	79.0	84.4	...	45.5	32.5	59.8	91.1	
2	Assam	Cachar	112171	14987	80.1	91.3	2.6	2.9	80.3	79.9	...	31.1	36.7	71.2	87.1	
3	Assam	Dhemaji	66537	8914	78.5	78.6	7.6	6.0	77.0	83.0	...	55.4	44.1	82.2	92.7	
4	Assam	Dhubri	46337	4655	59.1	88.3	8.4	2.9	59.9	84.5	...	24.1	50.0	46.2	97.9	

5 rows × 50 columns

Figure 3.2.3 shows the top 5 from the cleaned dataset

Step 4 : Review the descriptive statistics in the dataset including count,mean,standard deviation, minimum, maximum and various percentiles.

```
[14]: print("Descriptive Statistics:")
      df.describe()
```

[14]:

	Population Rural	Population Urban	Children Received Polio Birth Rural	Children Received Polio Birth Urban	Children Did Not Receive Any Vaccination Rural	Children Did Not Receive Any Vaccination Urban	Children Received Measles Vaccine Rural	Children Received Measles Vaccine Urban	Children Fully Immunized Rural	Children Fully Immunized Urban	...	Acute Illness An Taking Treatment From Governmen FemaleRural	Acute Illness And Taking Treatment From Government Female Urban
count	239.000000	239.000000	239.000000	239.000000	239.000000	239.000000	239.000000	239.000000	239.000000	239.000000	...	239.000000	239.000000
mean	61457.000000	15700.966527	78.096402	84.623305	4.464770	3.433556	79.894519	82.589958	65.015690	68.761967	...	24.800209	24.644059
std	32011.940477	18590.856396	11.085063	9.944992	4.020532	4.230919	11.448330	11.123917	15.064928	14.717695	...	20.536949	19.685091
min	11722.000000	2501.000000	33.600000	50.000000	0.170000	0.000000	36.390000	39.630000	26.860000	28.680000	...	1.000000	0.000000
25%	40984.000000	5924.000000	70.850000	79.555000	1.800000	1.100000	73.280000	76.395000	54.550000	58.530000	...	6.305000	6.715000
50%	54794.000000	9078.000000	80.100000	87.200000	3.300000	2.400000	81.900000	86.540000	66.000000	69.600000	...	20.900000	21.700000
75%	73493.000000	16320.500000	86.750000	92.275000	5.865000	4.000000	89.430000	90.570000	77.575000	81.440000	...	39.200000	35.940000
max	238744.000000	155980.000000	96.990000	98.170000	27.190000	35.980000	95.000000	98.390000	92.400000	95.970000	...	91.900000	93.060000

8 rows × 48 columns

Figure 3.2.4 shows preview of descriptive statistics

3.4 Flow Chart

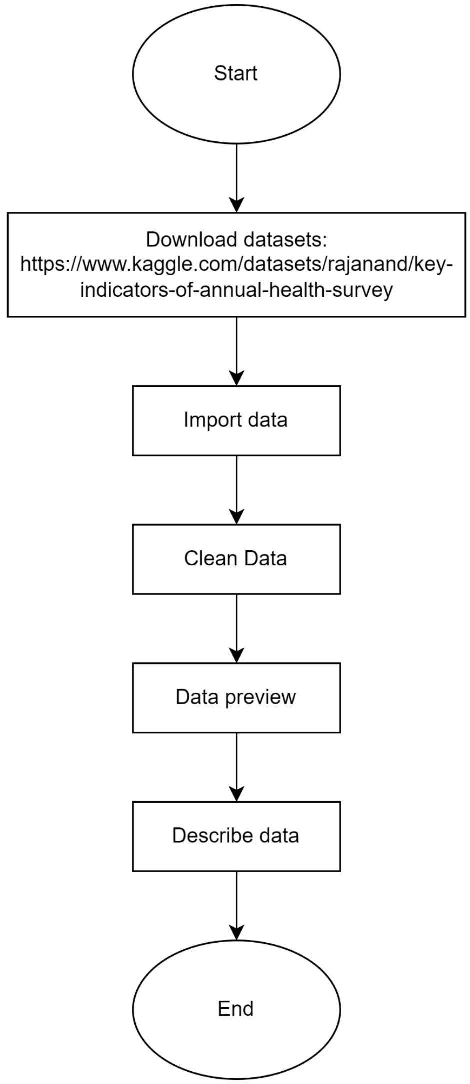


Figure 3.4.1 Data Preparation Process

4.0 Exploratory Data Analysis

4.1 Treatment patterns in rural and urban areas

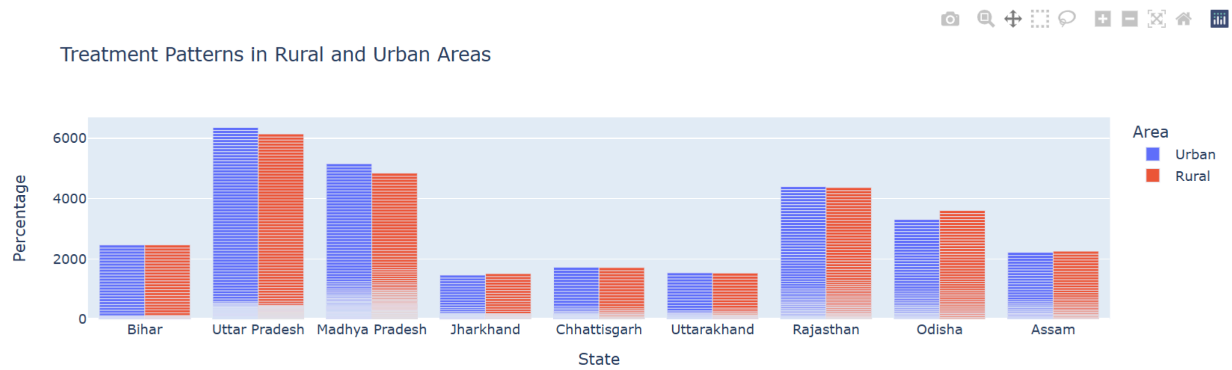


Figure 4.1 : Bar chart of treatment patterns

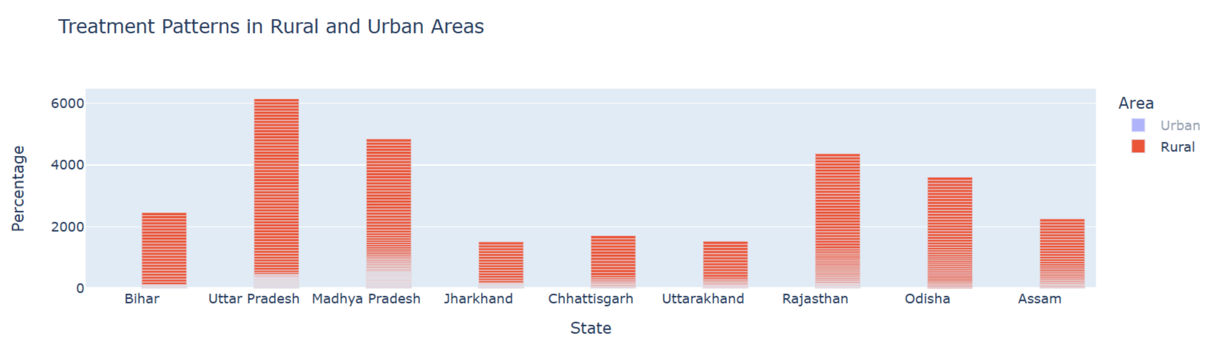


Figure 4.1.2: Filtering to display rural areas only

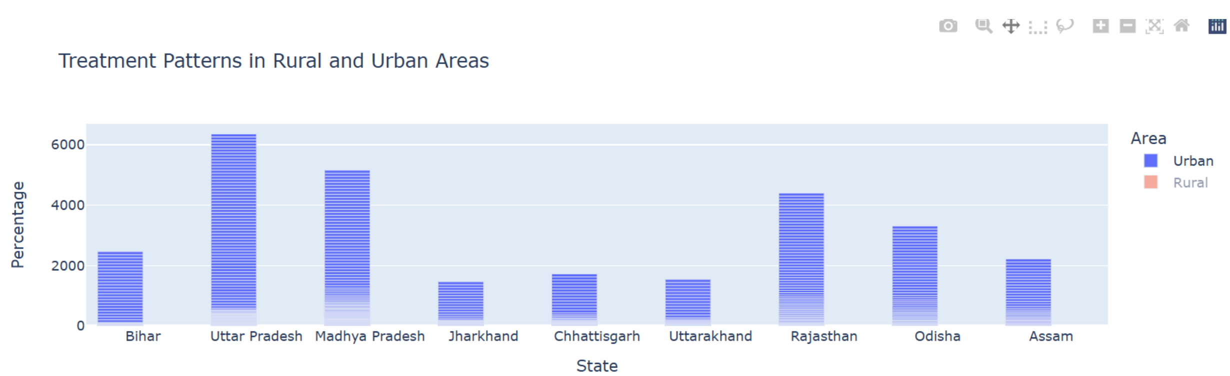


Figure 4.1.3: Filtering to display urban areas only

The interactive bar chart visualizes treatment patterns in rural and urban areas across different states using the following variables: 'Acute Illness And Treatment From Any Source Rural', 'Acute Illness And Taking Treatment From Any Source Urban', 'Acute Illness And Taking Treatment From Government Source Person Rural', and 'Acute Illness And Taking Treatment From Government Person Urban'. Specifically, it displays the percentage of individuals receiving treatment for acute illness from various sources, including any source and government sources. The data is segmented by area (rural vs. urban) and shows the distribution of treatment rates in a comparative manner. By grouping the bars based on areas, the chart makes it easy to compare the accessibility and uptake of health services between rural and urban populations within each state. This visualization can highlight disparities in healthcare access and effectiveness of public health interventions in different geographic settings.

The chart reveals distinct patterns in healthcare utilization, indicating potential gaps and areas for improvement. In Madhya Pradesh and Uttar Pradesh, urban areas tend to have higher treatment rates compared to rural areas, reflecting better accessibility and availability of healthcare services in urban regions. Conversely, Odisha may show relatively high treatment rates in rural areas, which could be attributed to targeted public health initiatives or effective community health programs. Analyzing these patterns helps identify regions where healthcare access is limited and supports the development of strategies to enhance healthcare delivery in underserved areas. Overall, this visualization provides valuable insights for policymakers and healthcare providers to address inequalities and improve health outcomes across diverse populations and archives our objective number one to analyze pattern trends across each nine states.

4.2 Women Awareness of HIV/AIDS/RTI/STI/HAF/ORS/ORT/Zinc/ARI/Pneumonia

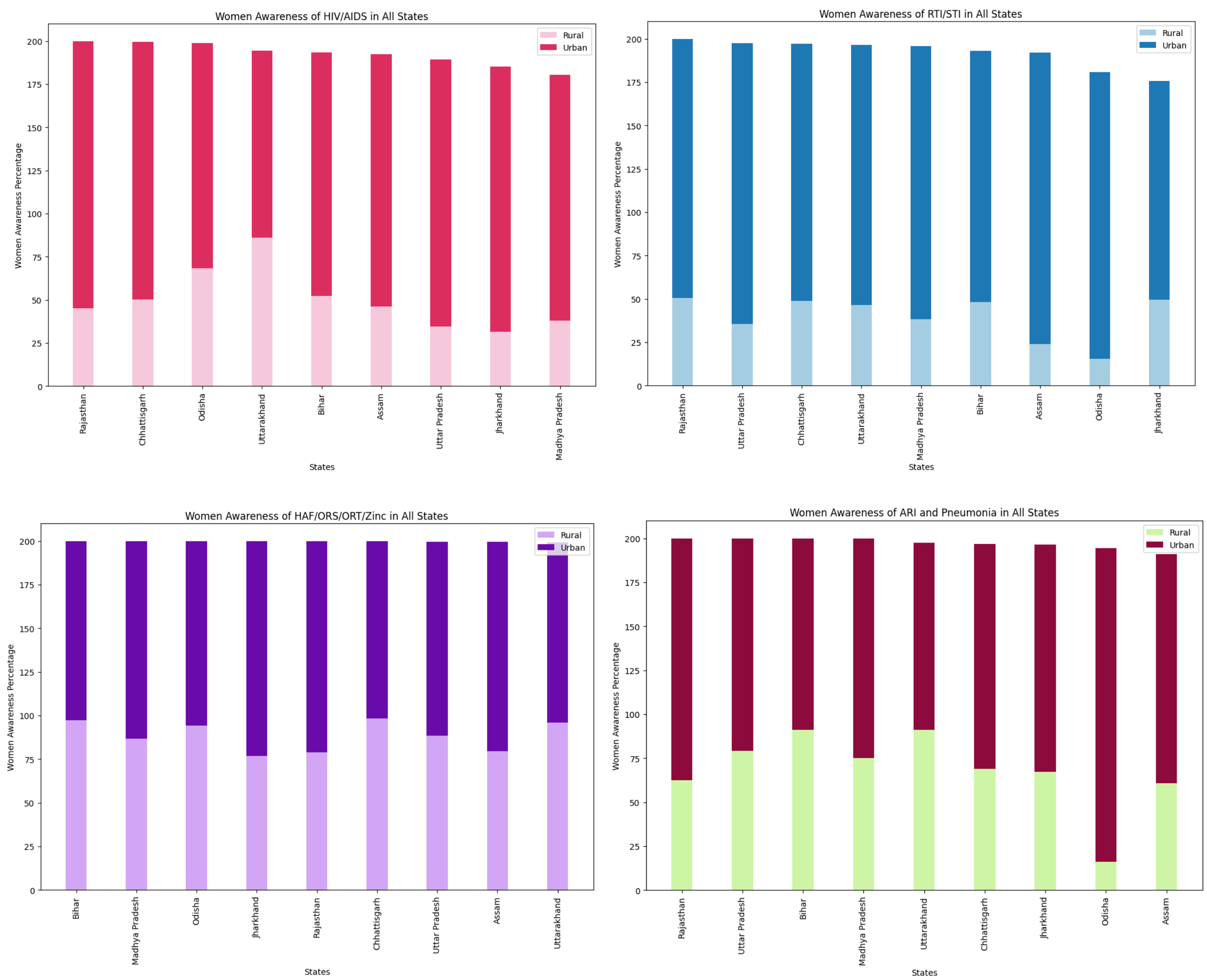


Figure 4.2.1 Women Awareness of HIV/AIDS/RTI/STI/HAF/ORS /ORT/Zinc/ARI/Pneumonia in EAG states of India

This stacked bar chart compares awareness of HIV, AIDS, RTI, STI, HAF, ORS, ORT, Zinc, ARI and Pneumonia among women in rural and urban areas of India. Bars colored in light show the percentage of women in outdoor areas while dark bars show the percentage in urban areas. This chart is arranged from highest to lowest, highlighting the importance of awareness in dealing with these health issues. Bars are divided into two segments: rural and urban.

The first chart is about the awareness of HIV and AIDS among women in high focus states in India. Rajasthan has the highest total percentage of women who are aware of HIV and AIDS followed by others. This is because Rajasthan has better facilities and health services. So, women have easier access to healthcare, and receive information about HIV and AIDS unlike Madhya Pradesh that has the lowest total percentage. Overall, rural areas have the lowest awareness of HIV and AIDS among women compared to urban areas.

The second chart represents data on the awareness of Reproductive Tract Infections and Sexually Transmitted Infections (RTI/STI) among women in all high focus states in India. Rajasthan too has the highest total percentage of women who are aware of RTI and STI. This is because Rajasthan also has effective use of media to spread information about RTI/STI among women compared to Jharkhand that has the lowest total percentage of women who are aware of RTI and STI. Overall, rural areas still have the lowest awareness of RTI and STI among women in both areas.

The third chart is about the awareness of HAF, ORS, ORT and zinc among women in all high focus states in India. This time Bihar has the highest total percentage of women who are aware of HAF, ORS, ORT and zinc. This is because Bihar has implemented public health campaigns that focus on HAF, ORS, ORT and zinc for women unlike Uttarakhand that has the lowest total percentage of women who are aware of HAF, ORS, ORT and zinc maybe because of lack of access to information. General awareness of HAF, ORS, ORT and zinc is lower among women in rural areas compared to urban areas.

The last chart represents data on the awareness of ARI and Pneumonia among women in all high focus states in India. Rajasthan again shows a high level of awareness in both areas with urban slightly higher than rural. This is because schools, community centers, and healthcare facilities in Rajasthan implement educational programs that teach the importance of high focus ARI and Pneumonia compared to Assam. Awareness of ARI and Pneumonia among women is less prevalent in rural places compared to urban areas.

To summarize, urban areas generally have higher levels of health awareness due to better access to information and health care facilities. However, states such as Madhya Pradesh, Jharkhand, Uttarakhand, and Assam show significant disparities between urban and rural areas, highlighting the need for targeted health education and outreach programs. States such as Rajasthan and Bihar show high levels of health information dissemination, indicating the success of health information dissemination. These data are important to evaluate the impact of a public health intervention on health awareness among women in EAG states.

4.3 Urban vs Rural Non-Vaccination by District

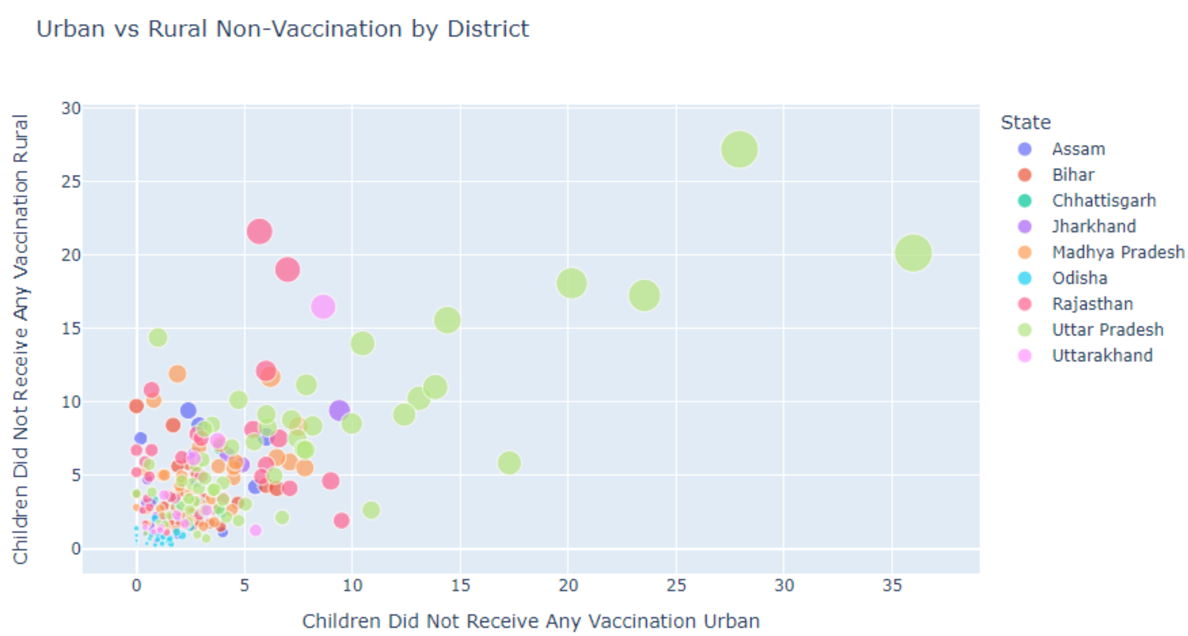


Figure 4.3.1 shows Children Urban vs Rural Non-Vaccination by District

The "Urban vs. Rural Non-Vaccination by District" scatter plot shows the connection between the number of kids who did not get any shots in cities and rural places across different districts. Different states are shown by different colors on the plot, which lets you compare these areas. Here's a more in-depth explanation.The x-axis shows the number of urban children who did not get any vaccinations.On the y-axis is the number of children in rural places who did not get any vaccinations.Some general thoughts:Most of the data points are close together at the bottom of both axes. This means that in most districts, both in cities and rural areas, the percentage of children who did not get any vaccinations is smaller.Some districts, mostly in rural areas, have higher numbers of people who have not been vaccinated, which suggests that vaccination coverage is not the same everywhere.Different colors on the plot show different states, which makes it easy to find and compare the rates of people who haven't been vaccinated in these states.

Some states, like Uttar Pradesh (shown in green), have a wider range of non-vaccination rates, which means that they vary from district to district. Urban and rural areas side by side there are a lot of data points above the 45-degree line, which means that the number of children who have not been vaccinated is higher in rural areas than in urban areas in those districts.This shows a common trend: less people get vaccinated in rural areas compared to urban areas. This could be because of things like lack of access to healthcare facilities, lower socioeconomic status, and less knowledgeThere are clear outliers, especially in the upper right corner of the plot, where some districts have much higher rates of people who haven't been vaccinated in both cities and remote areas.These outliers show which areas might need specific help to increase the number of people who have been vaccinated.This means the plot shows that most districts have pretty low rates of people who haven't been vaccinated, but there are big differences, especially in rural places.he clear difference between the number of people vaccinated in cities and those in rural areas calls for targeted public health policies to close these gaps and make sure that everyone has equal access to health care. This scatter plot does a good job of showing how different the rates of vaccination are between cities and rural areas in different districts. It also shows that these gaps need to be filled with specific public health interventions.

Numerous interconnected variables might be blamed for Uttar Pradesh's unpredictability and lower immunization rates, especially in rural regions. First off, vaccination attempts are hampered by rural Uttar Pradesh's often inadequate healthcare infrastructure, which includes fewer clinics, hospitals, and immunization centers as well as a scarcity of medical professionals with the necessary training. Socioeconomic considerations are also important; high rates of poverty and poor literacy lead to both financial troubles and a lack of knowledge about the value of immunisations. The influence of cultural and social norms on vaccination rates is noteworthy, since vaccinations may be viewed with suspicion and mistrust due to old beliefs that run counter to contemporary medical procedures. Certain societies' gender dynamics might limit women's autonomy when it comes to healthcare choices, which can impact the vaccination rates of both themselves and their children.

Healthcare providers find it challenging to administer vaccinations, and communities find it challenging to get to vaccination centers due to logistical and geographic obstacles such rough terrain and inadequate road systems. Additionally, vaccine deterioration and decreased effectiveness might result from inconsistent power and inadequate refrigeration in rural regions. In Uttar Pradesh, public health efforts often encounter uneven execution, poor training for healthcare personnel, insufficient oversight, and a deficiency of follow-up. Rural communities are sometimes neglected when resources are allocated, favoring metropolitan areas. The distribution and administration of vaccinations may also be impacted by political and administrative variables, such as corruption and inefficient bureaucracy. Lower vaccination rates are a result of inconsistent or inadequate health awareness programmes that are unable to connect with or reach rural people. A comprehensive strategy is needed to address these problems, one that involves boosting the execution of public health initiatives, guaranteeing fair resource allocation, raising education and awareness, and upgrading healthcare facilities.

4.4 Children received polio birth by State both in rural and urban area

Figure 4.4.1 shows boxplot of children received polio birth rural

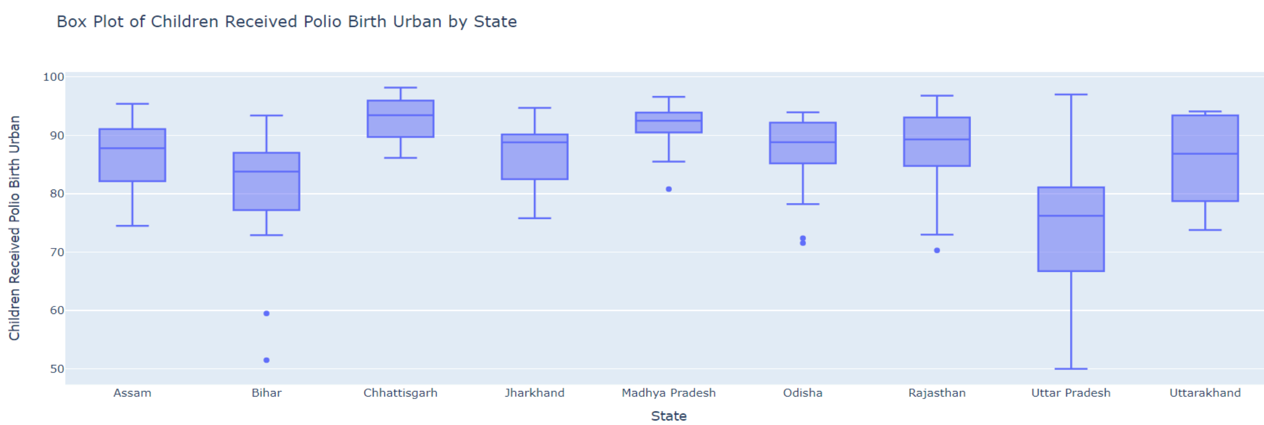


Figure 4.4.2 shows boxplot of children received polio birth urban

This boxplot shows the relationship between the children who received polio in both urban and rural by state. In Uttar Pradesh, the analysis of vaccination coverage between urban and rural areas reveals notable differences. For urban areas, the maximum observed vaccination coverage is 97, with a third quartile (Q3) value of 81.095, a median of 76.22, a first quartile (Q1) value of 66.75, and a minimum of 50. This indicates that urban areas generally have higher vaccination coverage, with 75% of the values below 81.095 and a central tendency (median) of 76.22. In contrast, rural areas have a maximum value of 93.42, Q3 of 78.345, a median of 69.8, Q1 of 65.0225, and a minimum of 48.5.

The higher median and quartile values in urban areas suggest better overall vaccination coverage compared to rural areas. The urban areas also exhibit greater variability in vaccination coverage, as evidenced by the wider spread between Q3 and Q1. The highest observed value in urban areas is greater than in rural areas, indicating that some urban districts have exceptionally high coverage. Despite urban areas showing better vaccination coverage, there is still variability, and certain regions within both urban and rural areas may need targeted public health interventions to improve health outcomes. This analysis underscores the need for tailored strategies to address specific regional disparities in vaccination coverage within Uttar Pradesh.

Next, In rural areas, the maximum observed vaccination coverage is 93.4, with the third quartile (Q3) at 86.725, a median of 81.7, and the first quartile (Q1) at 78.45. The lower fence, which helps identify outliers, is set at 72.4, and the minimum observed value is 46.5. Notable outliers include values of 66 from Bikaner, 63.6 from Jodhpur, 60.8 from Jalor, 58.5 from Jaisalmer, and 47.5 from Barmer, indicating significant deviations from the central tendency. These outliers suggest potential issues such as inadequate healthcare accessibility, lower health literacy, economic constraints, and logistical challenges in vaccine distribution.In urban areas, the maximum vaccination coverage is 96.8, with Q3 at 93.05, a median of 89.3, and Q1 at 84.75. The lower fence is set at 73, and the minimum observed value is 70.3, which is also the single outlier from the district of Jalor. The higher median and quartile values in urban areas indicate generally better vaccination coverage compared to rural areas. Urban areas benefit from higher accessibility to healthcare facilities, better health literacy, and more effective public health interventions. In contrast, rural areas exhibit greater variability and significant outliers, reflecting challenges such as inadequate infrastructure, lower socio-economic conditions, and logistical difficulties.

This difference can be attributed to several factors, for example is access to healthcare facilities. The urban area generally has better access to healthcare facilities including hospitals, clinics and vaccination centers. This increased accessibility makes it easier for urban residents to

receive vaccinations. While in rural areas, they often suffer from lack of healthcare infrastructure, which can hinder the availability and accessibility of vaccination services. Long distances to healthcare facilities and fewer healthcare providers contribute to lower vaccination rates. Next is maybe because of socio economic factors. For urban areas, they often have higher socio-economic status, which is associated with better health outcomes. People with higher income and education levels are more likely to prioritize and afford healthcare, including vaccinations. While for rural areas, they tend to have higher poverty rates, which can limit the ability to access healthcare services. Economic constraints may force families to prioritize immediate needs over preventive health measures like vaccinations.

4.5 Fever Cases by Gender and Area

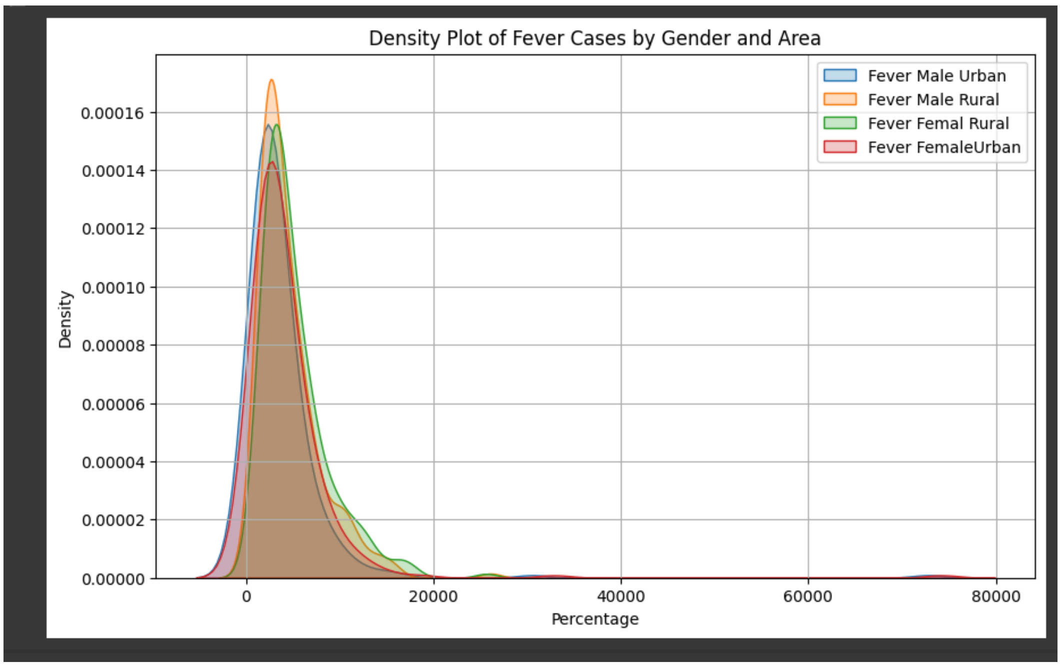
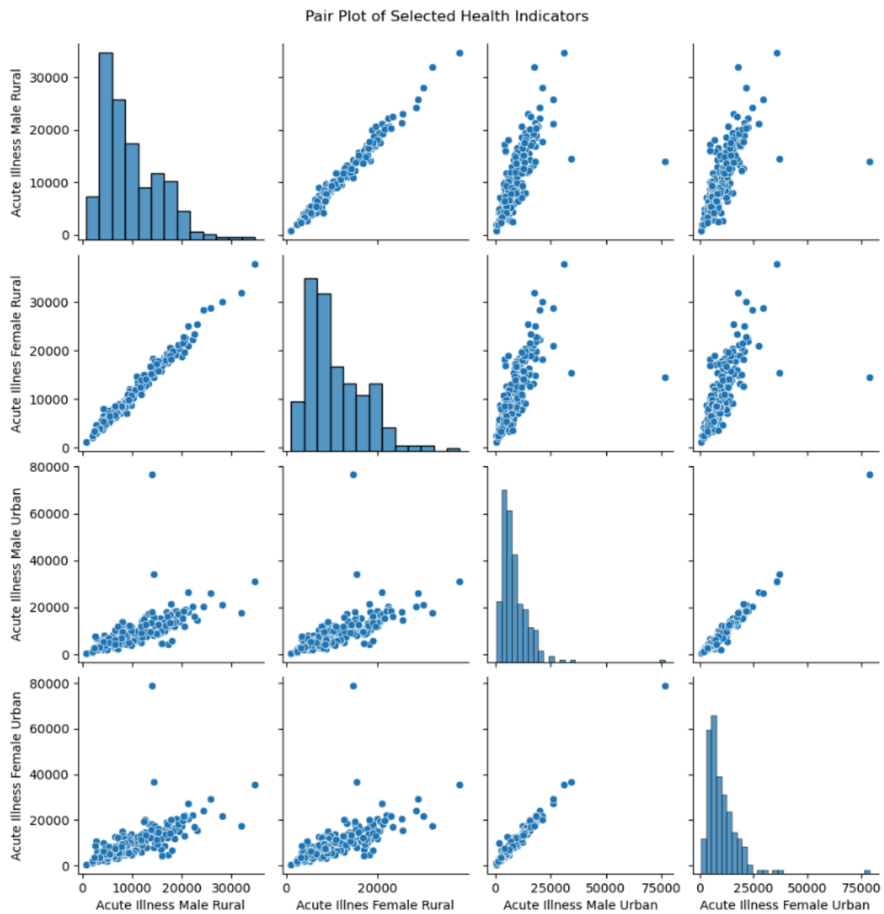


Figure 4.5.1 Density Plot of Fever Cases by Gender and Area

The density figure illustrates how fever cases are distributed throughout the four categories of "Fever Male Urban," "Fever Male Rural," "Fever Female Rural," and "Fever Female Urban." The predicted probability density of fever cases for each category is shown by a curve on the figure, which highlights the areas with the highest concentration of values. Taller peaks imply higher occurrences of a certain value. The height of the peaks reflects the most common values. The curves' breadth sheds light on the diversity of fever cases; broader curves imply higher variability, while smaller curves show values are concentrated around a central point. While unique, independent curves show different distributions of fever cases, overlapping curves reflect comparable distributions of fever cases between categories. The core trends and variability for each category may be identified by looking at the peaks and spreads of these curves. For example, comparable curves reflect similar fever case distributions for boys and females living in metropolitan areas. With the use of this visualization, it is possible to determine which populations have higher rates of fever cases. This information can then be used to influence targeted public health actions and stimulate more research into potential reasons, such as socioeconomic situations, healthcare access, or environmental variables.

There are a number of reasons for the variations in the distributions of fever cases in India among "Fever Male Urban," "Fever Male Rural," "Fever Female Rural," and "Fever Female Urban". First, access to Healthcare. Urban areas usually have more hospitals, clinics, and medical personnel in addition to improved access to healthcare services. In comparison to rural regions, this may result in more accurate reporting and treatment of fever cases and different distribution patterns. Longer travel times to medical facilities, fewer healthcare practitioners, and a lack of infrastructure for providing healthcare are just a few of the issues that rural areas frequently encounter. This may lead to fever cases being underreported or treated later than necessary, which would impact the distribution. Second, socioeconomic elements People living in cities typically have greater access to healthcare, dietary options, and hygienic conditions due to their higher socioeconomic position. On the other hand, rural communities can have a poorer socioeconomic position, which would increase their susceptibility to illnesses and change the distribution of fever cases. In cities, more educational attainment might result in improved health behaviors and awareness, which can affect the frequency and reporting of fever episodes. The lower levels of education in rural regions may have an impact on the awareness and behavior of health seekers. Last, Gender norms and expectations can have an impact on healthcare utilization and access in various regions of India. For instance, compared to men or women in cities, women living in rural regions may have less freedom to seek medical attention. This may lead to differences in how fever cases are reported and handled depending on the location and gender. On the other hand, depending on the sorts of job they do, men and women may be exposed to environmental elements that produce fever at varying amounts. For example, compared to jobs in cities, agricultural employment in rural areas may expose people to different health hazards.

4.6 Relationship Between Acute Illness Male Rural and Urban & Acute Illness Female Rural & Urban

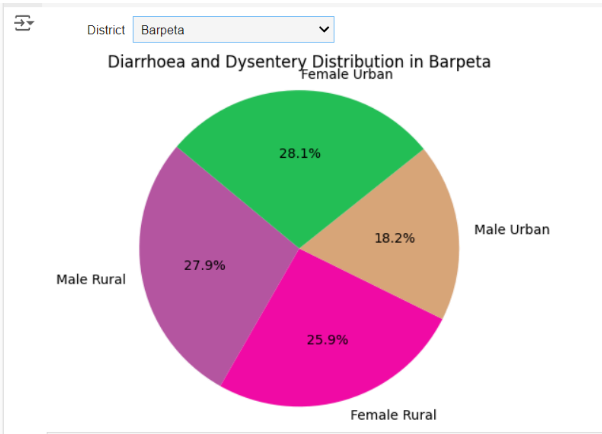


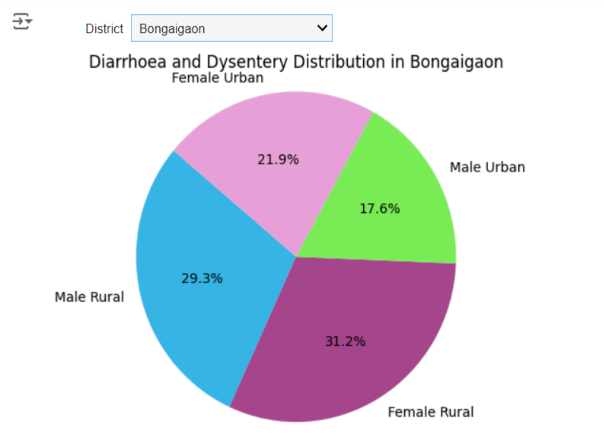
The pair plot presents a comprehensive view of the relationship between acute illness cases among males and females in both rural and urban settings. The distribution of cases of acute illness is shown by the diagonal histograms for each variable. The distributions are right-skewed for both males and females in rural areas, with most cases falling below 30,000. This skewness implies that most rural areas have lower rates of acute illness cases, but there are some areas with extremely high numbers. On the other hand, the urban distributions of both genders are more widely dispersed, with values reaching up to 75,000 cases. But the majority of urban cases are still below 40,000, suggesting a more widespread but still focused pattern of acute illness prevalence in urban areas compared to rural ones.

The scatter plots in the pair plot reveal the correlations between these variables, illustrating how the acute illness cases in one demographic group relate to those in another. In rural settings, there is a significant positive correlation between the number of acute illness cases among males and females. This implies that areas with a high incidence of acute illnesses among rural males also frequently have high incidence among rural females. In urban settings, acute illness cases among males and females also show strong correlations, suggesting that areas with high rates of illness in one gender also tend to have high rates in the other. These robust correlations within the same context suggest that both genders may be affected by similar environmental or socioeconomic factors that could affect the prevalence of acute illnesses.

Furthermore, the scatter plots also show positive correlations between rural and urban settings for both genders. Areas with high rural cases typically also have high urban cases, according to the correlation between rural males and urban males and, albeit slightly weaker, between rural females and urban females. This implies that factors causing acute illnesses may be consistent across regions, impacting people in both rural and urban areas in a comparable way. Acute illness cases across various settings and demographics are significantly correlated, as indicated by the overall pattern of correlations, which raises the possibility that these trends are influenced by systemic or regional health issues. This graphic shows possible areas for focused public health interventions and offers insightful information about the relationships between acute illness prevalence in various populations.

4.7 Diarrhoea and Dysentery by District





The data on diarrhea and dysentery cases across various districts reveals notable disparities in the distribution between rural and urban populations, as well as between genders. In rural areas, females account for 31.2% of the cases, while males make up 29.3%, together representing 60.5% of the total cases. In urban areas, females represent 21.9% and males 17.6%, contributing to 39.5% of the total cases. This substantial difference indicates that these illnesses are more common in rural areas, which may be explained by things like poor sanitation infrastructure, contaminated water sources, restricted access to healthcare services, and lower levels of knowledge about disease prevention and hygiene.

Analyzing the gender distribution across all districts, we observe that females account for 53.1% of the cases, while males make up 46.9%. This suggests that females are more affected by diarrhea and dysentery compared to males. There could be a variety of causes for this, such as biological variations, socioeconomic factors, and environmental exposures. For example, women and girls are more likely to be exposed to contaminated water and inadequate sanitation because they are frequently in charge of maintaining household hygiene and gathering water in rural areas. Targeted interventions that improve access to clean water, sanitation facilities, healthcare services, and hygiene education are necessary to address these disparities, especially in underserved and rural areas.

The interactive feature of the data visualization tool is a valuable addition that allows users to select a specific district from a dropdown menu to view the distribution of diarrhea and dysentery cases across four demographic categories: Female Rural, Male Rural, Female Urban, and Male Urban. This interactive functionality enhances the comprehensibility and utility of the data by providing a dynamic and clear visual representation for each district. It facilitates detailed analysis and comparison, enabling users to identify regional patterns and disparities more effectively. By allowing district-wise comparisons, this helps pinpoint specific areas that may require more focused public health interventions and resources. Such insights are crucial for policymakers, healthcare providers, and researchers aiming to develop targeted strategies to mitigate the impact of these diseases and improve health outcomes across different populations.

5.0 SUMMARY

Interactive bar chart comparing treatment patterns for acute illness in rural and urban areas in nine EAG states. It shows disparities in healthcare access, with urban areas in Madhya Pradesh and Uttar Pradesh having higher rates of treatment, indicating better accessibility. In contrast, Odisha showed relatively high treatment rates in rural areas, suggesting an effective public health initiative. This visualization fulfills the objective of comparing health outcomes and healthcare access between rural and urban populations in the nine EAG states, providing valuable insights to address regional healthcare inequities.

Stacked bar chart comparing health awareness among rural and urban women in nine EAG states, focusing on HIV, AIDS, RTI, STI, HAF, ORS, ORT, zinc, ARI and pneumonia. Urban areas show higher awareness due to better access to healthcare information and facilities, with Rajasthan and Bihar showing the highest levels. This visualization fulfills the objective of evaluating the impact of public health interventions on health awareness among women in the EAG states.

The scatter plot "Urban vs. Rural Non-Vaccination by District" shows disparities in vaccination rates between urban and rural areas. Most districts have low non-vaccination rates, but rural areas have higher unvaccinated children. Uttar Pradesh has significant variability in non-vaccination rates due to inadequate healthcare infrastructure, socioeconomic challenges, cultural norms, and logistical barriers. This visualization fulfills the objective of analyzing key health indicators, particularly vaccination coverage, and underscores the need for targeted public health interventions to address disparities between urban and rural populations in the nine EAG states.

The density figure shows the distribution of fever cases across four categories which is "Urban Male Fever," "Rural Male Fever," "Rural Female Fever" and "Urban Female Fever." Urban areas have better access to healthcare, while rural areas face under-reporting and delayed treatment. Socioeconomic factors, education level, and gender norms also influence the distribution. This visualization fulfills the objective of comparing health outcomes and healthcare access between rural and urban populations in the nine EAG states, providing insights for targeted public health interventions.

The paired plot shows a right-skewed distribution of acute illness cases among males and females in rural and urban areas. Rural areas have fewer cases than cities, with 75,000 cases per year. A positive correlation was observed between males and females, indicating areas of high incidence for one gender tend to have high incidence for the other. This visualization fulfills the objective of analyzing the evolution of disease prevalence and treatment patterns across the nine EAG states, highlighting areas for targeted public health interventions.

Data visualization shows that cases of diarrhea and dysentery are more prevalent in rural areas (60.5%) than in urban areas (39.5%), with women accounting for a higher percentage of cases. Women accounted for 31.2% of cases in rural areas and 29.3% in urban areas. This higher incidence may be due to their role in home hygiene and water collection. Interactive features help identify regional disparities and the need for targeted public health interventions. This visualization fulfills the objective of comparing health outcomes between rural and urban populations and investigating the correlation between environmental factors and health outcomes in the nine EAG states.

The Indian government can improve health outcomes by investing in a comprehensive public health strategy that addresses health care access, vaccination coverage, disease management, environmental factors and gender disparities. This includes improving rural health care infrastructure, implementing targeted interventions, promoting health awareness, and addressing socioeconomic and cultural barriers.

REFERENCES

GeeksforGeeks. (2022, November 30). Using Plotly for interactive data visualization in Python. GeeksforGeeks. https://www.geeksforgeeks.org/using-plotly-for-interactive-data-visualization-in-python/

Majumder, P. (2023, September 25). *Guide to Create Interactive Plots with Plotly Python*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/10/interactive-plots-in-python-with-plotly-a-complete-guide/

Kasthuri, A. (2018). Challenges to healthcare in India - The five A's. *Indian Journal of Community Medicine/Indian Journal of Community Medicine*, *43*(3), 141. https://doi.org/10.4103/ijcm.ijcm_194_18

Kumar, A. (2023). The transformation of the Indian healthcare system. *Curēus*. https://doi.org/10.7759/cureus.39079

Mhtadmin. (2024, April 29). *The significance of women empowerment in India*. Mahila Housing Trust. https://www.mahilahousingtrust.org/the-significance-of-women-empowerment-in-india/

APPENDIX

Google drive Link:
https://drive.google.com/drive/folders/1flbQ5E5UYIeN8AnaCWUzT4cZLko71d9f?usp=sharing

The google drive link has access to
- Original dataset (DataWrangling_datasets)
- Cleaned dataset (health_cleaned)
- Google Colab Visualization (DataWrangling_Project_Visualization)