Analyzing Women Migration Patterns in India

SATABDA MAJUMDER

M.Sc (STATISTICS-SEM 4)

STD ID - 2221209002003

REGI NO - 220020035064

STATISTICS - DISSERTATION

UNDER GUIDENCE OF

Dr. ANINDITA GHOSAL
HOD, Department of Statistics
Sister Nivedita University

CONTENTS

PROJECT TITLE

Analyzing Women Migration Patterns in India:

The Impact of Employment, Education, Business and Marriage, Across States

ABSTRACT

This study investigates the patterns and determinants of female migration across the 14 largest states in India, focusing on the effects of employment, business, education, and marriage as primary reasons for migration. Using data from the 1991, 2001, and 2011 census years, we performed a comprehensive analysis that includes subdivided bar diagrams, Generalized Linear Models (GLM), clustering, dendrograms, and QQ plots. The analysis is divided into rural and urban migration to capture the distinct dynamics in different geographical contexts. Our findings reveal significant insights into the temporal and spatial trends of female migration, highlighting the predominant reasons for migration and the variations among different states. We also assess model fit and distributional assumptions using heteroskedasticity tests and QQ plots. This study aims to provide a nuanced understanding of female migration patterns in India, aiding policymakers and researchers in addressing the socio-economic challenges and opportunities associated with migration.

INTRODUCTION

Migration is a complex phenomenon influenced by various socio-economic factors. In India, female migration has historically been driven by marriage, but recent trends indicate a growing influence of employment, education, and business opportunities. Understanding these dynamics is crucial for policymakers aiming to improve socio-economic conditions and gender equality. This study focuses on female migration across the 14 geographically largest states in India, analyzing data from the 1991, 2001, and 2011 census years.

Significance of Women Migration

Social Implications:

1. **Gender Roles and Empowerment**: Migration often challenges traditional gender roles and can lead to greater empowerment for women. Women who migrate for employment or education can gain financial independence and increased social status.

2. **Family Dynamics**: Women's migration, particularly for marriage, can significantly alter family structures and dynamics. It can impact caregiving roles, decision-making processes, and the distribution of household responsibilities.

Economic Implications:

1. **Labor Market Dynamics**: Women migrating for work contribute to the labor force in destination areas, often filling crucial gaps in sectors such as healthcare, education, domestic work, and manufacturing.

2. **Remittances**: Women migrants often send remittances back to their families in their place of origin. These financial transfers can be vital for household survival, education, and healthcare, thus contributing to poverty alleviation.

3. **Economic Independence:** Employment-related migration can enhance women's economic independence, leading to increased savings, investment in businesses, and overall economic growth in their communities.

Educational Implications:

1. **Access to Education**: Migration for education purposes enables women to access better educational opportunities, which can lead to improved career prospects and personal development.

2. **Skill Development**: Educational migration facilitates skill acquisition and professional training, contributing to the development of a skilled workforce.

Health Implications:

1. **Healthcare Access**: Migration can affect women's access to healthcare services. In some cases, migrating to urban areas can improve access to better healthcare facilities.

2. **Mental Health**: The migration process can be stressful and impact women's mental health, especially if they face challenges such as discrimination, isolation, or exploitation in the destination areas.

The primary objectives of this study are to:

1. **Identify and analyze the major reasons for female migration**: We categorize migration into four main reasons—employment, business, education, and marriage—and examine their impact on female migration patterns.

2. **Examine temporal changes**: By comparing data from three different census years, we observe how the reasons for migration have evolved over time.

3. **Compare rural and urban migration**: Migration dynamics often differ significantly between rural and urban areas, necessitating a separate analysis for each context.

4. **Utilize advanced statistical methods**: We employ Generalized Linear Models (GLM) to understand the relationships between migration reasons and total migration. Clustering and dendrogram techniques are used to uncover patterns among states, and QQ plots along with heteroskedasticity tests ensure the robustness of our models

The data used in this study is sourced from detailed census tables, which are meticulously organized by reason for migration, gender, and rural-urban distinction for the 14 largest states. Multiple subdivided bar diagrams are generated to visualize migration trends. We further analyze these trends using GLM on datasets segregated by total, rural, and urban migration for each census year. Clustering and dendrograms provide insights into state-wise migration patterns, and diagnostic plots such as QQ plots and heteroskedasticity tests assess the fit and assumptions of our models.

OBJECTIVES

- Proportion of women migration over men on different sectors like employment, business, education and marriage.
- Individual effects of employment, business, education and marriage on total female migration of India.
- Understanding the Similarities of Migration Patterns among the States.

DATA SOURCE

- The data on Number of migrants by their last residence and reason for migration for 2011 has been collected from census website. **https://censusindia.gov.in**
- The data on Number of migrants by their last residence and reason for migration for 2001 has been collected from census website. **https://censusindia.gov.in**
- The data on Number of migrants by their last residence and reason for migration for 1991 has been collected from census website. **https://censusindia.gov.in**

METHODOLOGY

The methodology for analyzing women migration and its various influencing factors, such as employment, education, marriage, and business, involves a systematic approach that includes data collection, data analysis, and data visualization. Here's an overview of the steps taken in the methodology:

Data Collection:

- Gathered census data from 2011, 2001, and 1991 for the 14 largest states of India.
- Collected data on women migration, including figures on migration due to employment, business, education, and marriage, as well as total migration.
- Organized the data into tables, dividing it by census year, migration reason, gender (total, male, female), and rural/urban classification.

Data Analysis and Visualization:

- Generated charts (such as bar diagrams) to visualize the figures of migration due to employment, business, education, and marriage for the 14 states.
- Created multiple subdivided bar diagrams to compare migration patterns across different states and census years.
- Used statistical tools and techniques to analyze the data, including GLM (Generalized Linear Models) for Poisson regression analysis.

Generalized Linear Model (GLM):

- Performed GLM on the data tables for each census year (2011, 2001, 1991) and for different migration categories (total migration, rural migration, urban migration).
- Set up GLM models to assess the effects of employment, education, marriage, and other factors on women migration.
- Examined the coefficients, significance levels, and other diagnostic measures provided by the GLM output to understand the relationships between variables.

Diagnostic Checks:

- Used QQ plots to visually assess the normality of residuals from the GLM models.
- Checked for heteroscadasticity by examining residual plots or performing formal tests.
- Addressed any issues identified through diagnostic checks, such as transforming variables or using robust standard errors.

Clustering and Dendrogram:

- Performed hierarchical clustering to group states based on migration patterns.
- Generated dendrograms to visualize the clustering results.

Interpretation and Reporting:

- Interpreted the results of GLM models to draw conclusions about the effects of employment, education, marriage, and literacy on women migration.
- Discuss any significant findings or patterns observed in the data, considering both overall trends and variations across states and census years.

IMPORTANT STAISTICAL TOOLS USED IN THIS PROJECT

Multiple Subdivided Bar Diagram:

- A multiple subdivided bar diagram is a graphical representation used to compare the distribution of a categorical variable across multiple groups.
- It consists of bars divided into segments, with each segment representing a subgroup within a larger category.
- This type of diagram is useful for visualizing relationships and patterns within complex categorical data sets, allowing for easy comparison between different groups or categories.

Generalized Linear Model (GLM):

- The Generalized Linear Model (GLM) is a statistical framework used for modeling relationships between a response variable and one or more predictor variables.
- It extends traditional linear regression by accommodating non-normal distributions of the response variable and non-linear relationships between predictors and the response.
- GLM allows for the modeling of a wide range of data types and response distributions, making it a flexible and powerful tool for statistical analysis in various fields.

.

QQ Plot (Quantile-Quantile Plot):

- A QQ plot is a graphical tool used to assess whether a given data set follows a specific probability distribution.
- It compares the quantiles of the observed data to the quantiles of a theoretical distribution (such as the normal distribution) that the data is assumed to follow.
- A QQ plot that closely follows a straight line indicates that the data closely matches the assumed distribution, while deviations from the line suggest departures from the assumed distribution.

Heteroscedasticity Test:

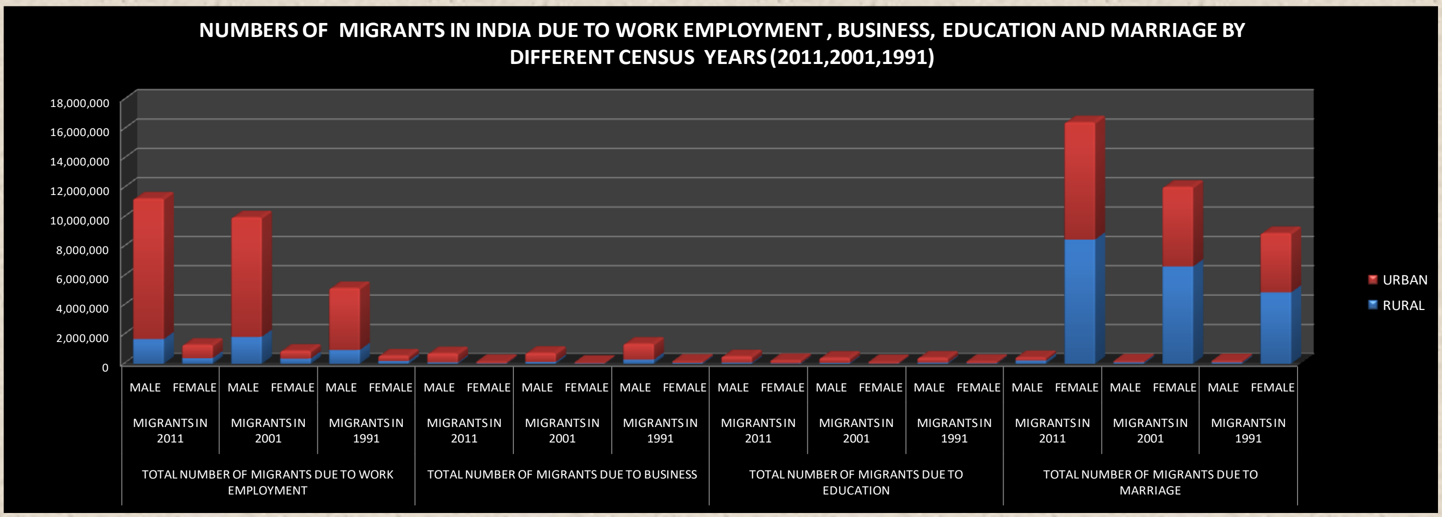
- Heteroscedasticity refers to the unequal variance of errors or residuals in a regression model.
- Heteroscedasticity can lead to biased estimates and incorrect inferences in regression analysis.
- Various statistical tests, such as the Breusch-Pagan test or White test, can be used to detect heteroscedasticity in regression models.
- Addressing heteroscedasticity may involve transforming variables, using robust standard errors, or applying alternative modeling techniques.

Clustering and Dendrogram:

- Clustering is a statistical technique used to group similar objects or data points into clusters based on their characteristics.
- Helps identify natural groupings within a dataset for exploratory data analysis and pattern recognition.
- A dendrogram is a tree-like diagram that records the sequences of merges or splits in hierarchical clustering.

DATA ANALYSIS FROM THE GRAPHICAL REPRESENTATION

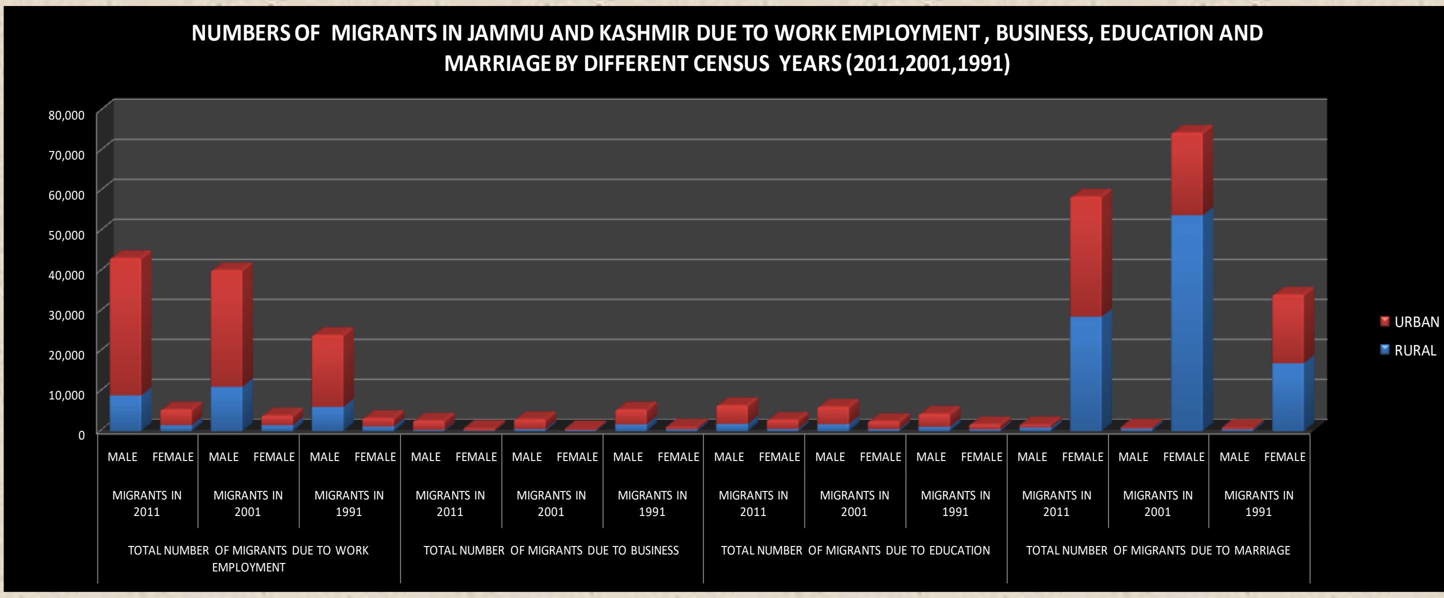
FOR INDIA SCENARIO:



KEY FINDINGS:

- In all three census years (2011, 2001, 1991), male migrants vastly outnumber female migrants for all three sections (employment, business, education)with urban migration being more prevalent than rural migration.
- The number of male migrants for work employment and education increased significantly over the years, reaching its peak in 2011 while the number of business migrants remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While rural migration due to marriage is higher compared to urban migration, though urban migration is also significant.
- Migration for work employment and marriage are the two largest categories, with work-related migration being male-dominated and marriage-related migration being female-dominated.

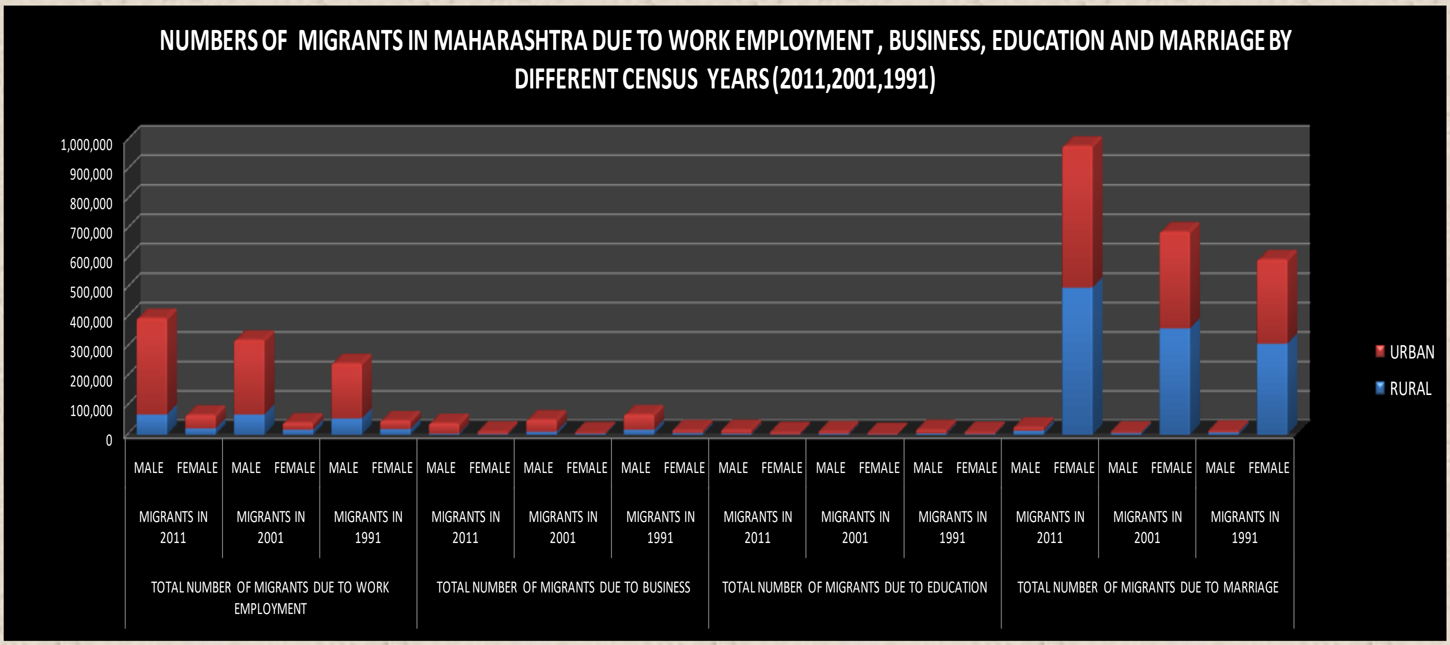
FOR JAMMU AND KASHMIR SCENARIO:



KEY FINDINGS:

- The pattern of Jammu and Kashmir's employment and education migration is quite same as India's pattern.
- But in case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- While in case of marriage the total number of migration was at its peak in 2001 with rural migration being more prevalent than urban migration. But in 2011 the marital migration decreased significantly with close to equal proportion of urban and rural segment.
- Migration for work employment and marriage are the two largest categories, with work-related migration being male-dominated and marriage-related migration being female-dominated.

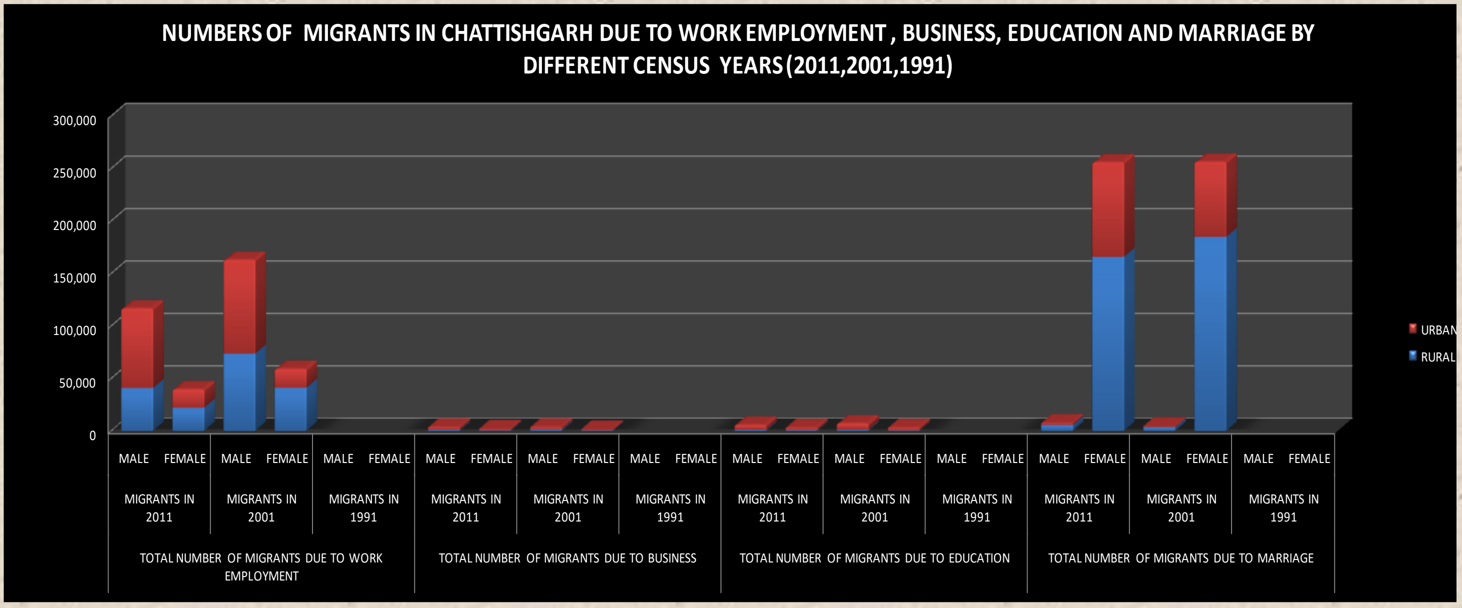
FOR MAHARASHTRA SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While urban migration due to marriage is slightly higher compared to rural migration, though rural migration is also significant.

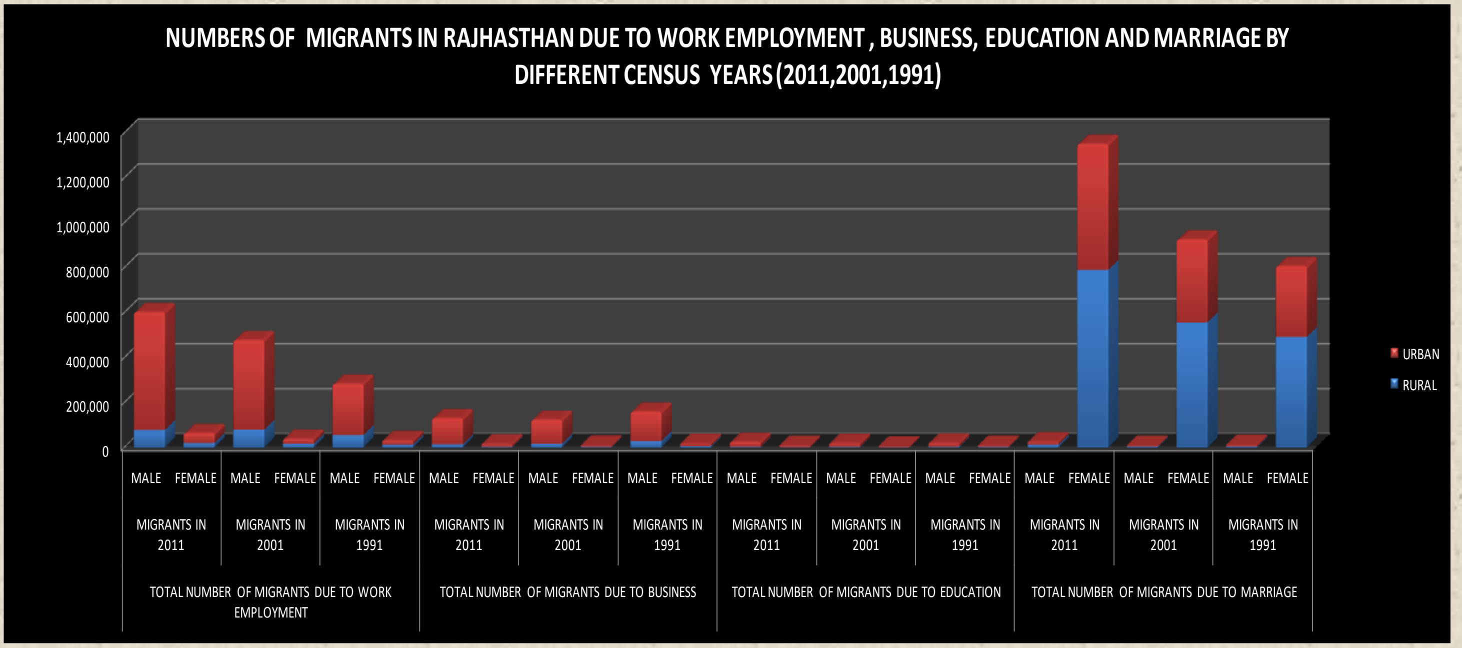
FOR CHATTISHGARH SCENARIO:



KEY FINDINGS:

- Chhattisgarh was created on 1 November 2000 that's why there is no census data for 1991.
- Migration for work employment and marriage are the two largest categories, with work-related migration being male-dominated and marriage-related migration being female-dominated, while other two reasons are quite negligible.
- For work employment the male as well as female migration has decreased significantly, for males proportion of urban migration is higher than rural but this scenario is reversed for the females.
- The number of female migration due to marriage remains nearly same for two consecutive censuses. But proportion of urban migration slightly decreased over the years.

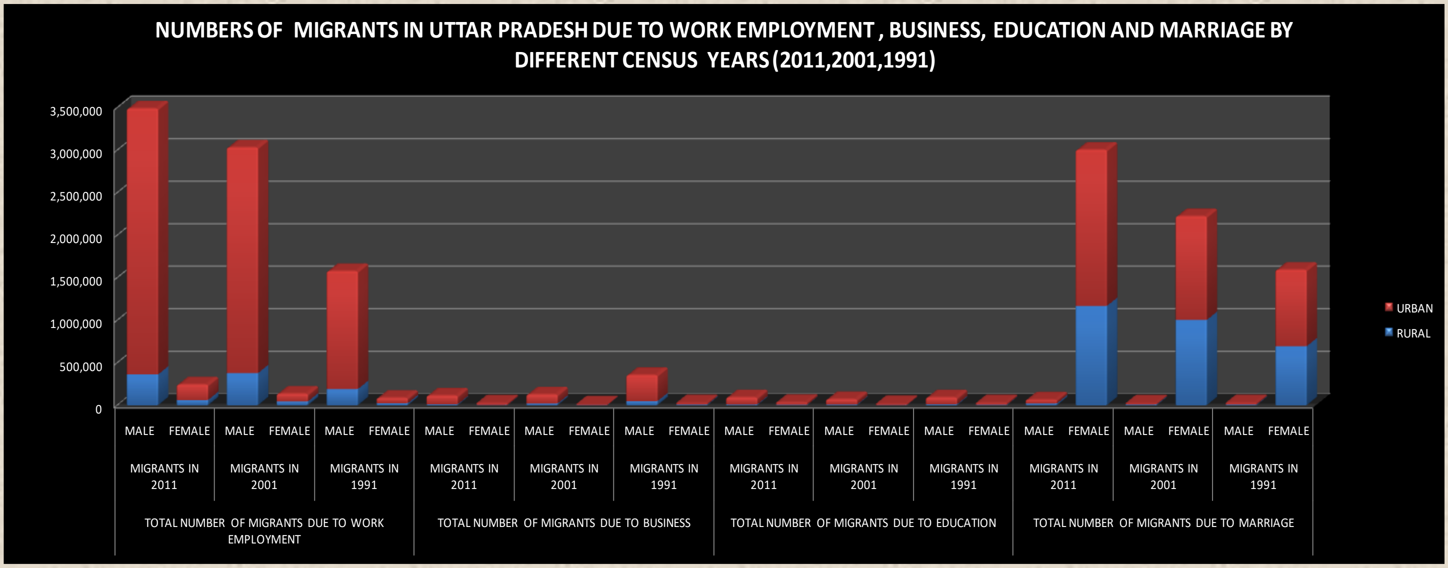
FOR RAJHASTHAN SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While rural migration due to marriage is slightly higher compared to urban migration, though urban migration is also significant.

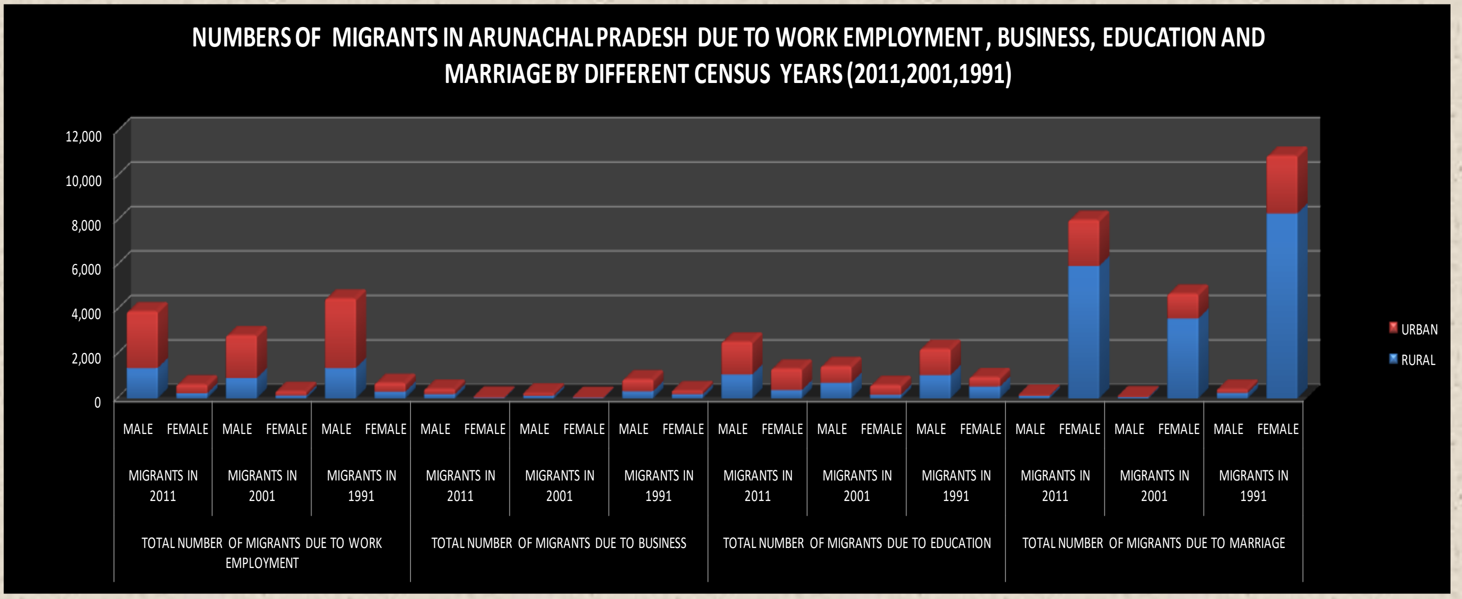
FOR UTTAR PRADESH SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While urban migration due to marriage is slightly higher compared to rural migration, though rural migration is also significant.
- The number of male migrants for work employment is significantly higher than number of marital migration, and the proportion of urban migration due to employment is quite higher than rural segment.

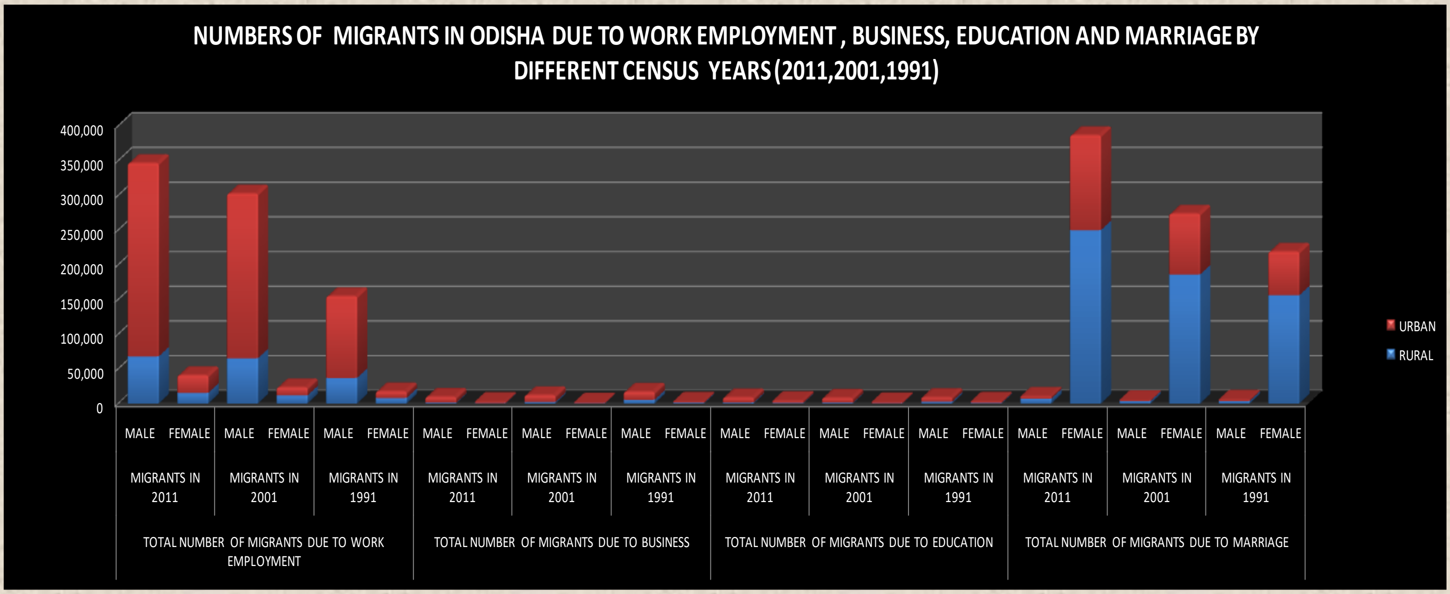
FOR ARUNACHAL PRADESH SCENARIO:



KEY FINDINGS:

- For employment, education and marital migration considering both males and females the number of total migrants was highest in 1991, slightly decreased in 2001, again increased in 2011.
- Migration due to business is quite negligible comparing to others.
- The proportion of male and female in educational migration is quite good than other factors.
- Migration for work employment and marriage are the two largest categories, with work-related migration being male-dominated and has a predominant portion urban migration, while marriage-related migration being female-dominated and significantly rural influenced.

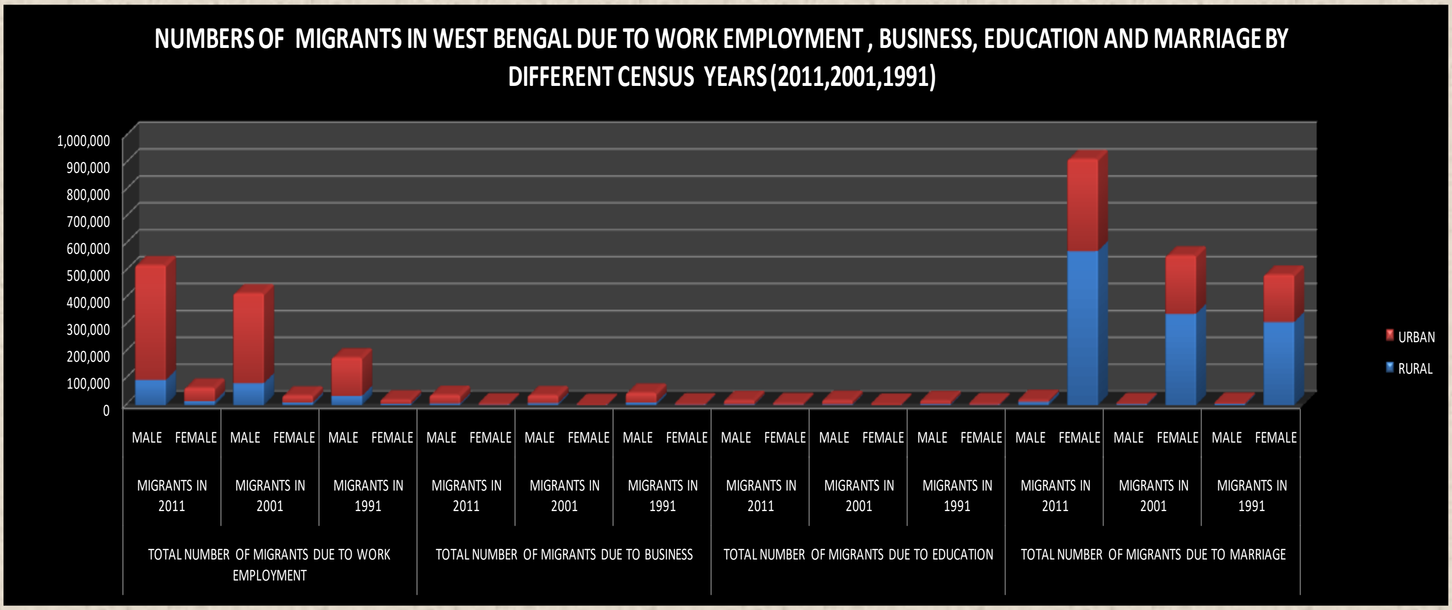
FOR ODISHA SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While rural migration due to marriage is slightly higher compared to urban migration, though urban migration is also significant.
- The number of male migrants for work employment is quite equals to the number of marital migration, and the proportion of urban migration due to employment is quite higher than rural segment.

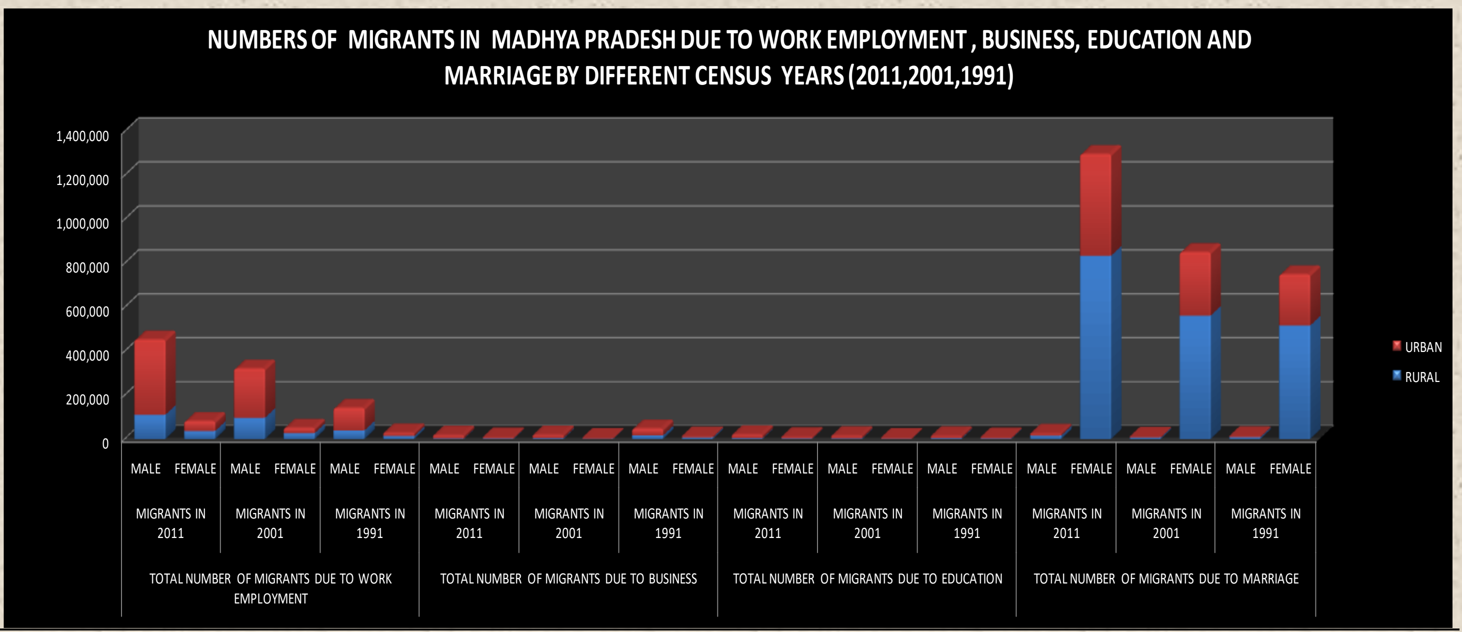
FOR WEST BENGAL SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While rural migration due to marriage is slightly higher compared to urban migration, though urban migration is also significant.

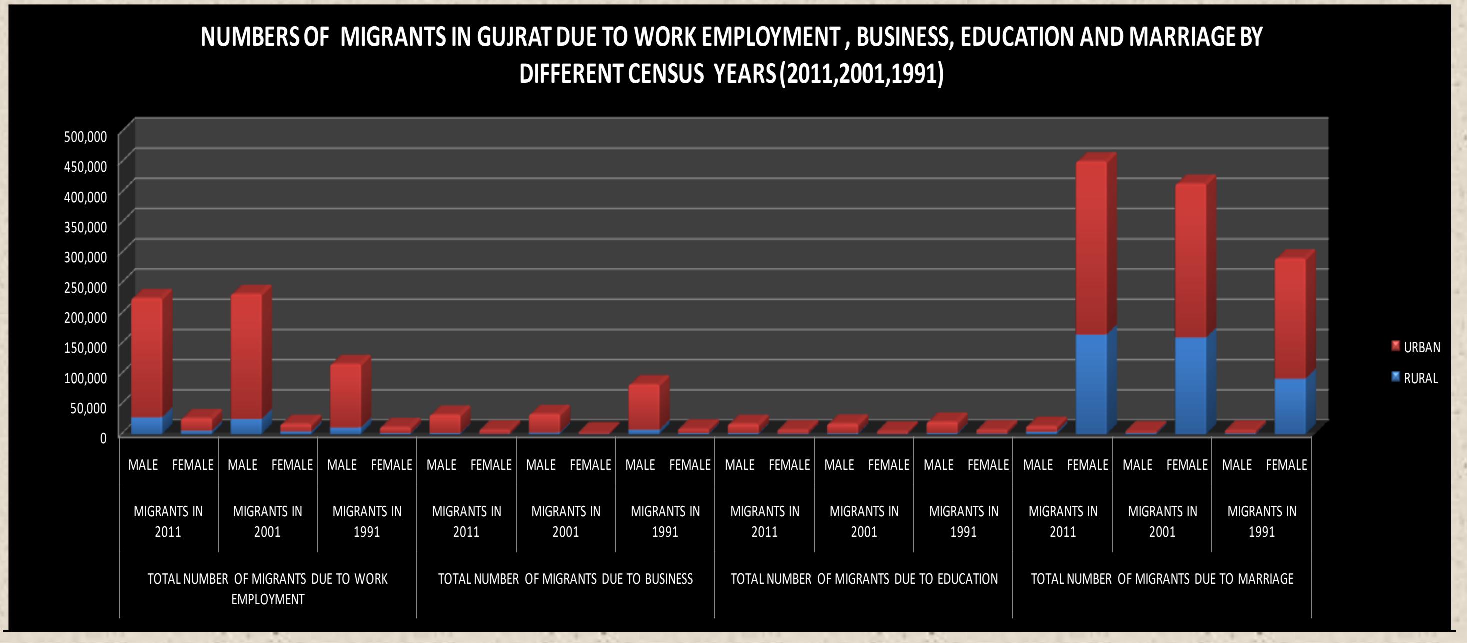
FOR MADHYA PRADESH SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration is quite same over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While rural migration due to marriage is slightly higher compared to urban migration, though urban migration is also significant.

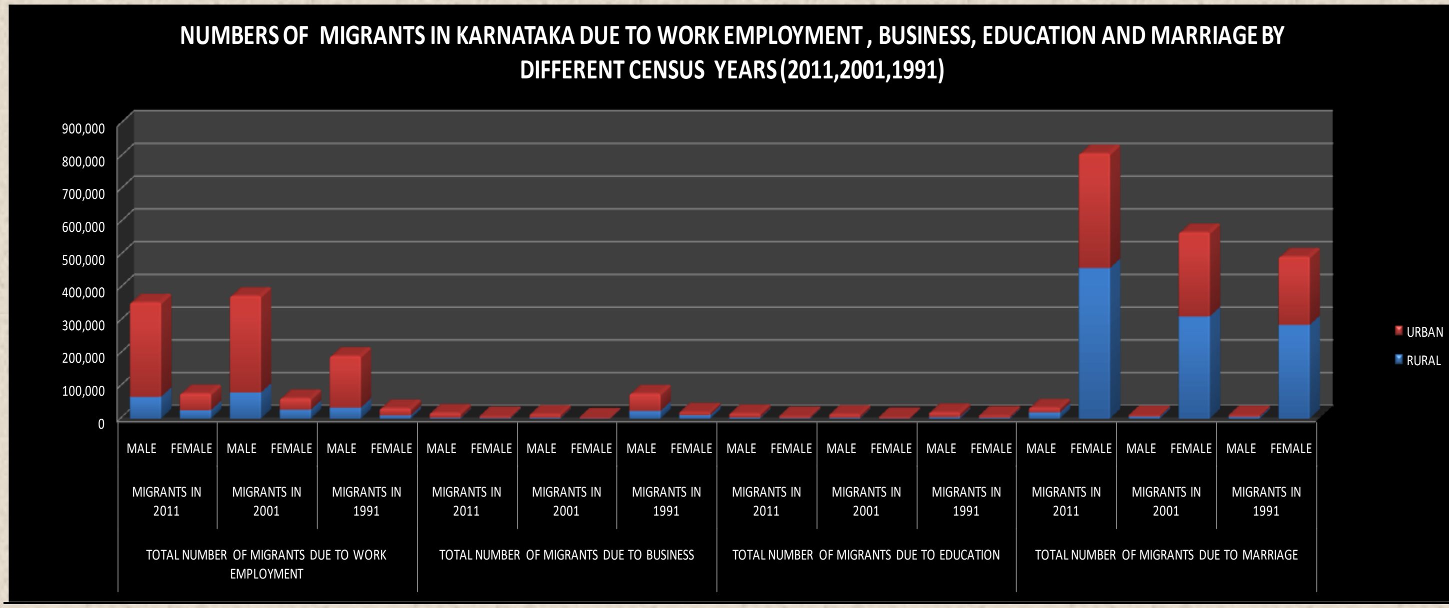
FOR GUJRAT SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2001,but slightly decreased in 2011 and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons but relatively decreasing over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While the whole pattern of Gujarat's migration is highly influenced by urban sectors.

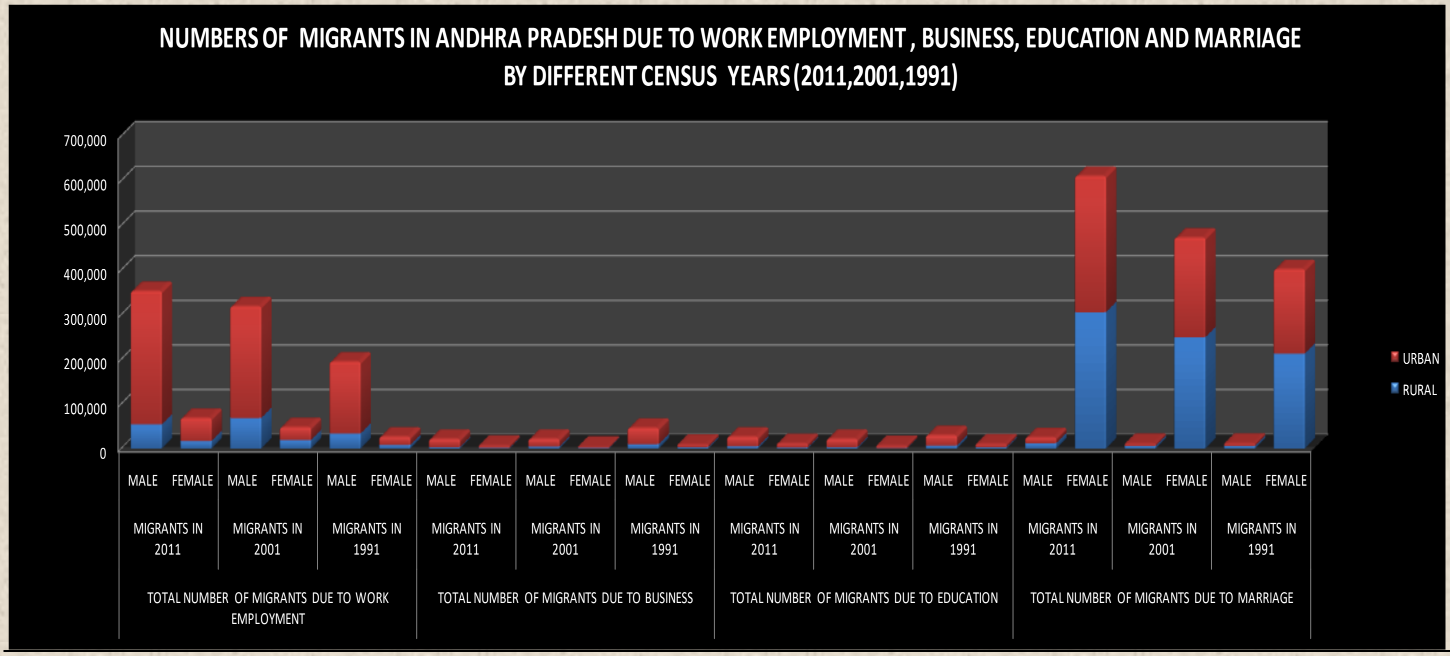
FOR KARNATAKA SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2001,but slightly decreased in 2011 and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons but relatively decreasing over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. Migration for work employment and marriage are the two largest categories, with work-related migration being male-dominated and has a predominant portion urban migration, while marriage-related migration being female-dominated and significantly rural influenced.

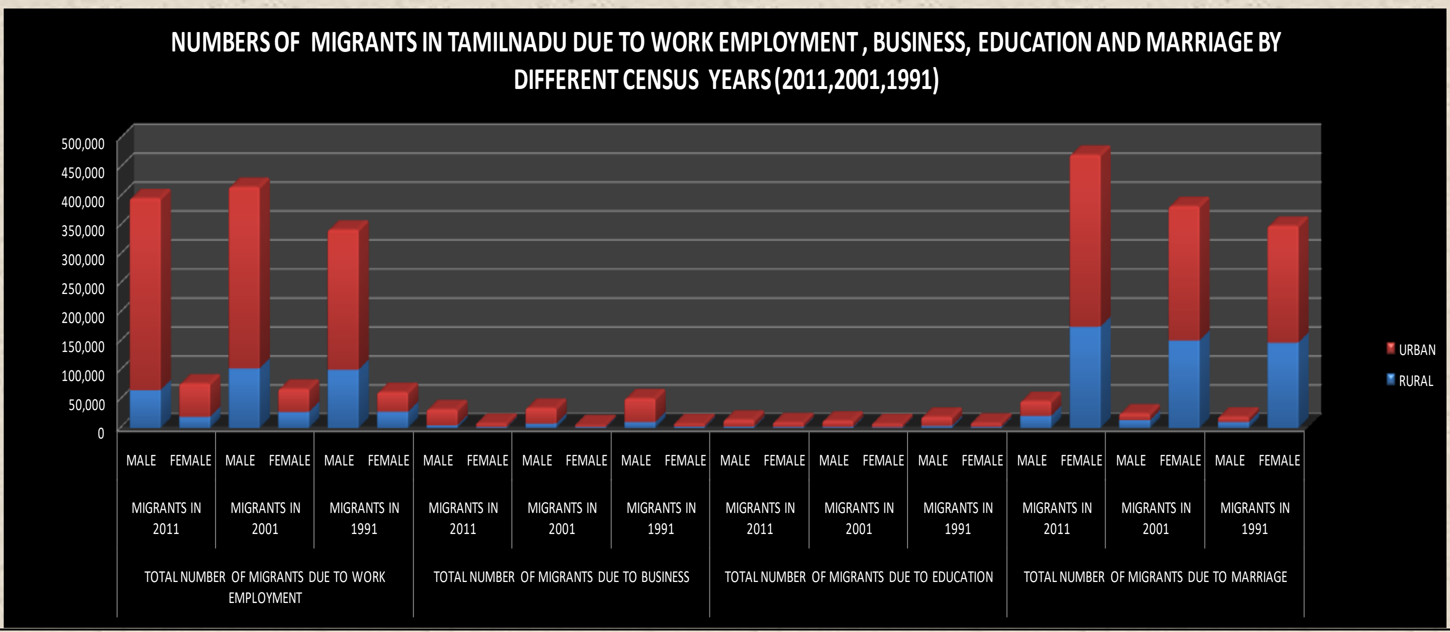
FOR ANDHRA PRADESH SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While urban migration due to marriage is slightly higher compared to rural migration, though rural migration is also significant.

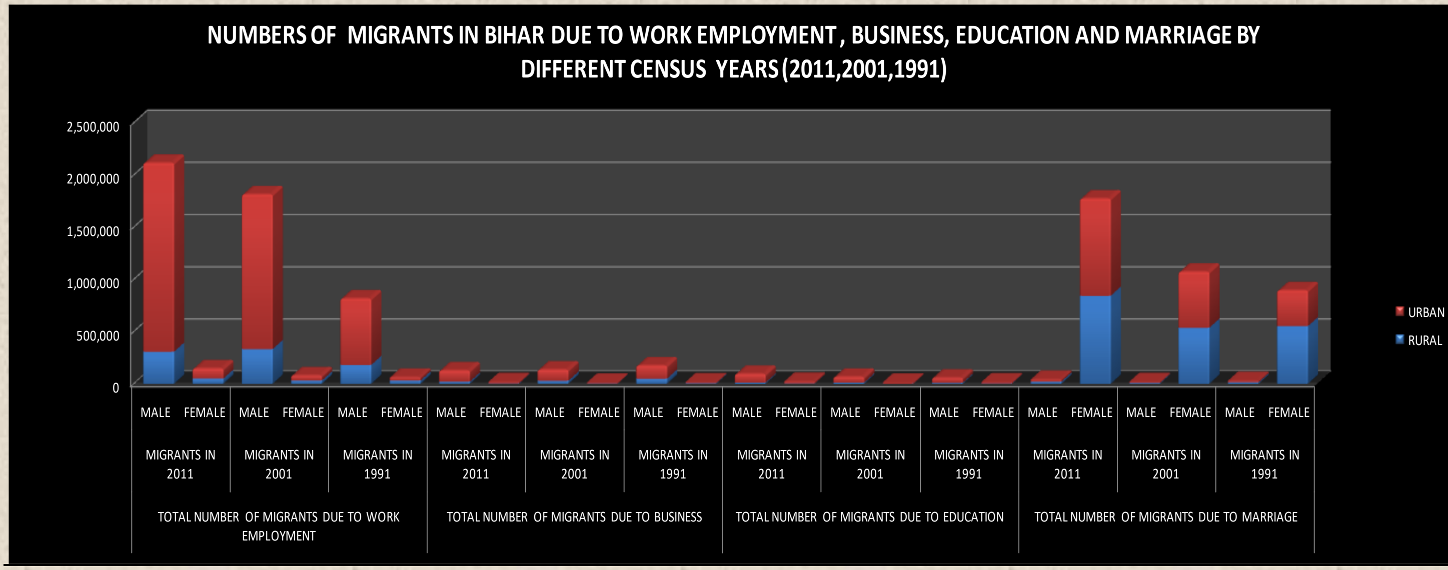
FOR TAMILNADU SCENARIO:



KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2001,but slightly decreased in 2011 and also female migrants for work employment increased slightly over the years. Also its volume is quite close to marital migration but it is highly male dominated.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons but relatively decreasing over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While the whole pattern of Tamilnadu's migration is highly influenced by urban sectors.

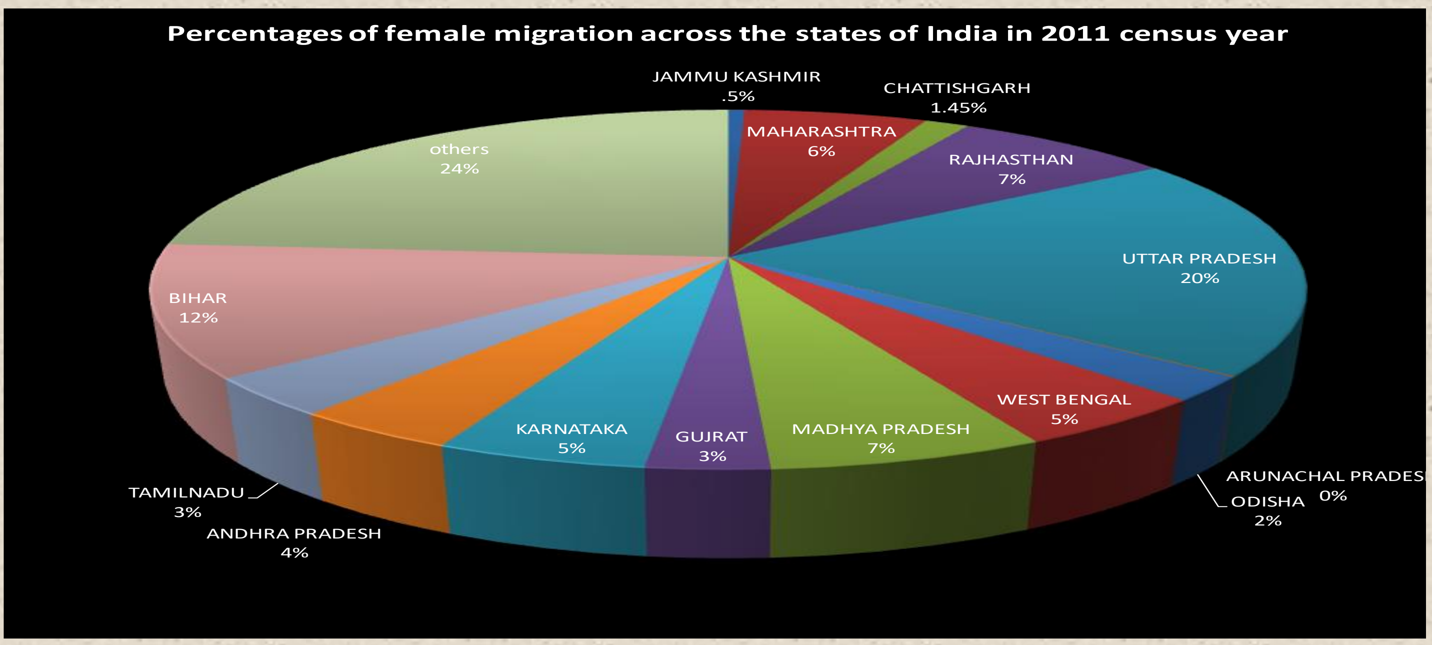
FOR BIHAR SCENARIO:

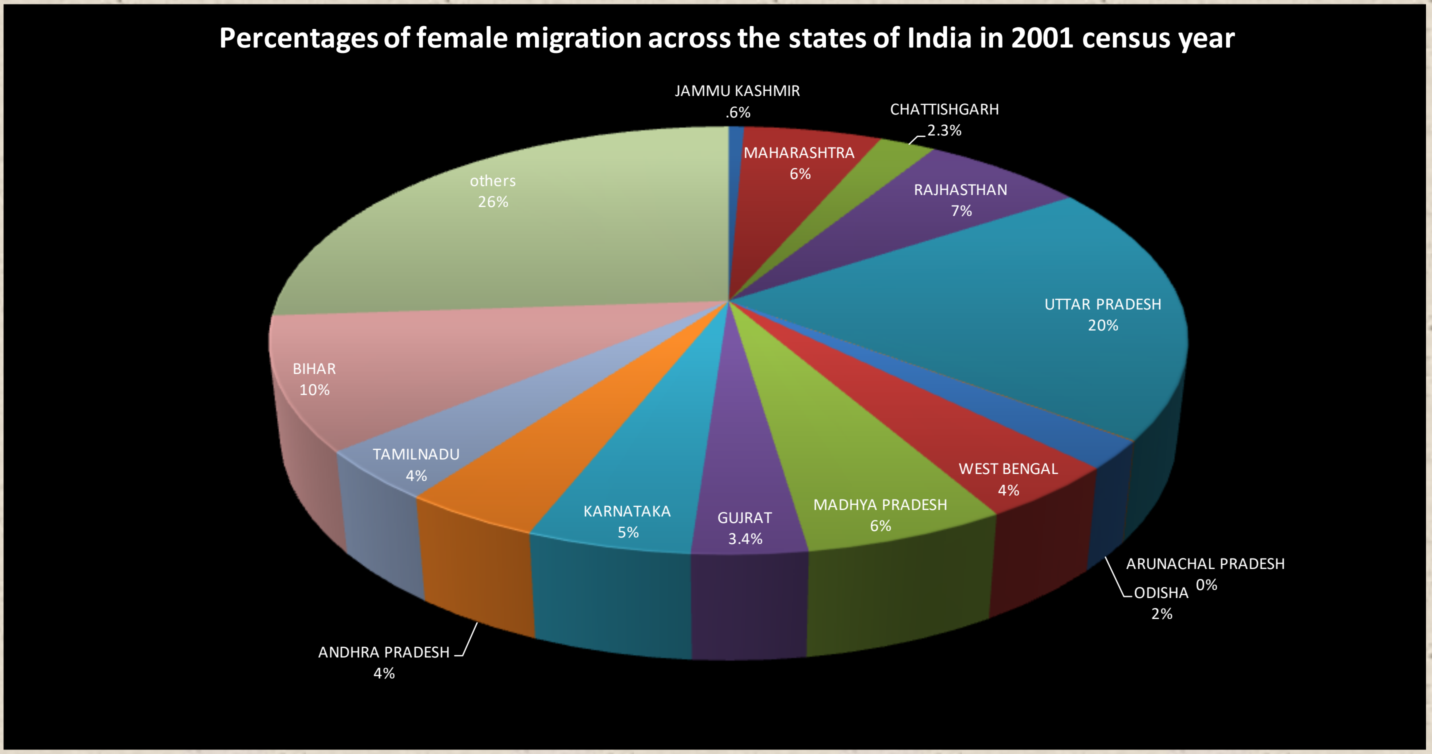


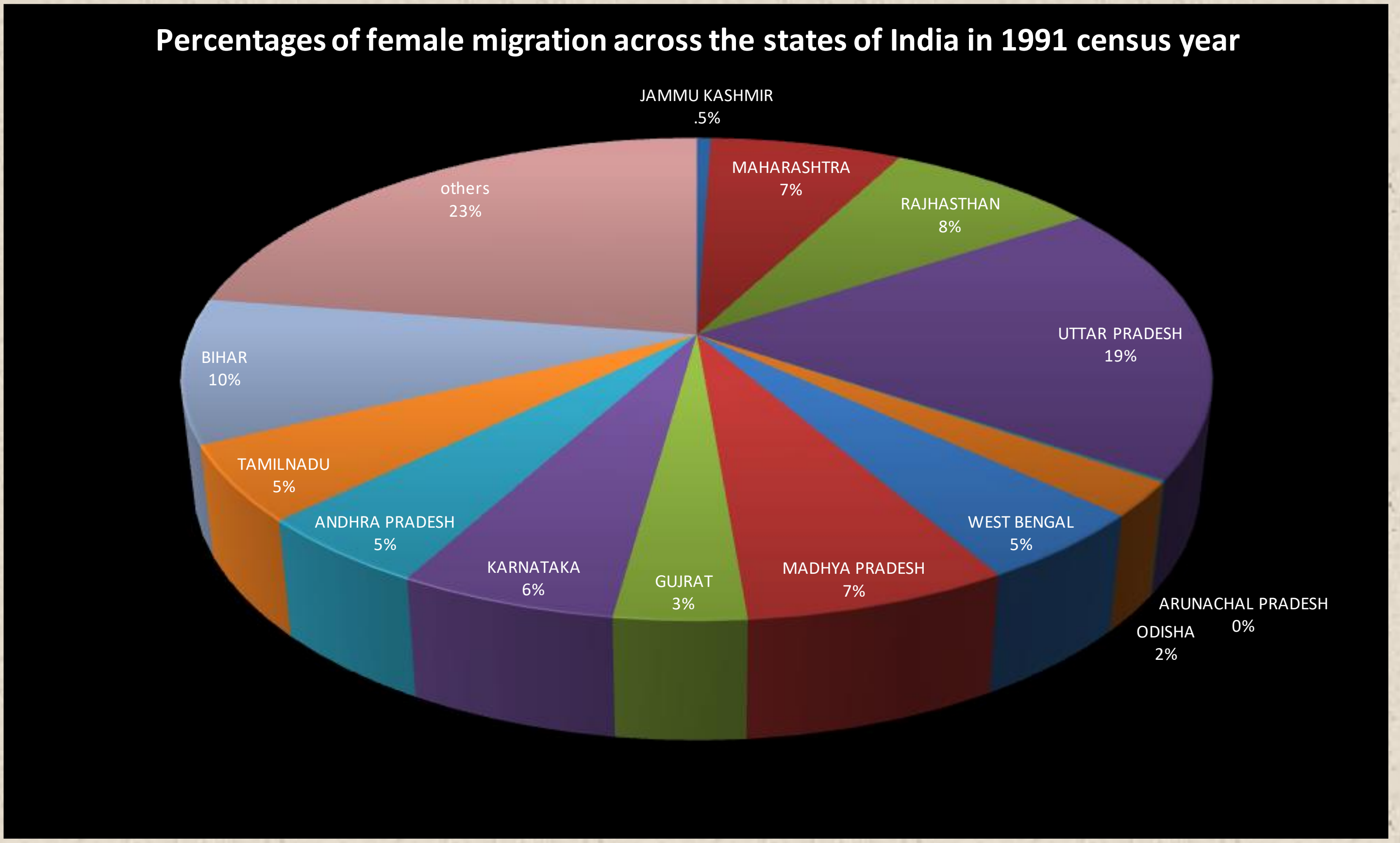
KEY FINDINGS:

- The number of male migrants for work employment and marriage increased significantly over the years, reaching its peak in 2011, and also female migrants for work employment increased slightly over the years.
- In case of business migration the number of migration decreased over the census years and male migrants vastly outnumber female migrants.
- The number of migrants due to education is very low compared to other reasons and remained relatively stable over the three census years.
- Migration due to marriage is predominantly female and increased significantly over the years. While proportion of rural migration due to marriage is slightly decreasing compared to urban migration over the years.
- The number of male migrants for work employment is quite equals to the number of marital migration, and the proportion of urban migration due to employment is quite higher than rural segment.

Percentages of female migration across the states of India for different census year






Percentages of female migration across the states of India in 1991 census year
JAMMU KASHMIR .5%
MAHARASHTRA 7%
RAJHASTHAN 8%
UTTAR PRADESH 19%
WEST BENGAL 5%
ARUNACHAL PRADESH 0%
ODISHA 2%
MADHYA PRADESH 7%
GUJRAT 3%
KARNATAKA 6%
ANDHRA PRADESH 5%
TAMILNADU 5%
BIHAR 10%
others 23%

STATISTICAL ANALYSIS USING R PROGRRAMMING

R CODE FOR GLM (GENERALISED LINEAR MODEL):

```
# TOTAL NUMBER OF  FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

file_path = ("C:/Users/pc1/Desktop/ASSIGNMENTS/SATABDA_DISSERTATION_1 (Autosaved)2.xlsx")

data = read_excel(file_path, sheet = "Sheet5",range = "R7C2:R22C7" )

print(data)

model <- glm(TOTAL_NUMBER_OF_MIGRANTS ~ `NUMBER OF MIGRANTS DUE TO WORK
EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO
EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data, family = poisson())

summary(model)

fitted(model)
```

Explanation:

- glm(): This function is used to fit a Generalized Linear Model (GLM).
- TOTAL_NUMBER_OF_MIGRANTS: This is the response variable (dependent variable) in the model, representing the total number of migrants.
- NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT, NUMBER OF MIGRANTS DUE TO BUSINESS, NUMBER OF MIGRANTS DUE TO EDUCATION, NUMBER OF MIGRANTS DUE TO MARRIAGE: These are the predictor variables (independent variables) in the model, representing the number of migrants due to different reasons (employment, business, education, marriage).
- data: This specifies the data frame containing the variables used in the model.
- family = poisson: This specifies the family distribution for the GLM, which in this case is Poisson. Since the response variable (total number of migrants) is count data, Poisson regression is appropriate for modeling count data.

Overall, this code fits a Poisson regression model to the data, where the total number of migrants is modeled as a function of the number of migrants due to various reasons (employment, business, education, marriage). The resulting model can be used to analyze the effects of these predictor variables on the total number of migrants.

R CODE FOR Q-Q PLOT (QUANTILE-QUANTILE PLOT):

```
file_path = ("C:/Users/pc1/Desktop/ASSIGNMENTS/SATABDA_DISSERTATION_1 (Autosaved)2.xlsx")

data = read_excel(file_path, sheet = "Sheet5",range = "R7C2:R22C7" )

print(data)

model <- glm(TOTAL_NUMBER_OF_MIGRANTS ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` +

        `NUMBER OF MIGRANTS DUE TO BUSINESS` +  `NUMBER OF MIGRANTS DUE TO EDUCATION` +

        `NUMBER OF MIGRANTS DUE TO MARRIAGE`,  data = data,    family = poisson)

residuals <- residuals(model)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals)) + stat_qq(distribution = qnorm) +   geom_abline(slope = 1,
intercept = 0, color = "red") +  labs(title = "QQ Plot of Poisson Regression Residuals",  x = "Theoretical
Quantiles", y = "Sample Quantiles")
```

Explanation:

- After fitting the Poisson regression model using glm(), we extract the residuals from the model using the residuals() function.
- We then create a QQ plot of the residuals against a theoretical normal distribution using ggplot2.
- The stat_qq() function is used to create the QQ plot, and we specify distribution = qnorm to compare the residuals to a theoretical normal distribution.
- The geom_abline() function adds a line of equality to the plot, which helps in assessing the deviation of the points from the expected normal distribution.
- Finally, we provide labels for the plot title and axes using the labs() function

R CODE BREUSCH-PAGAN TEST FOR HETEROSCEDASTICITY:

```
file_path = ("C:/Users/pc1/Desktop/ASSIGNMENTS/SATABDA_DISSERTATION_1 (Autosaved)2.xlsx")

    data = read_excel(file_path, sheet = "Sheet5",range = "R7C2:R22C7" )

    print(data)

    model <- glm(TOTAL_NUMBER_OF_MIGRANTS ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data, family = poisson())

summary(model)

fitted(model)

library(lmtest)

fitted_values <- fitted(model)

residuals <- residuals(model)

lm_squared_resid <- lm(residuals^2 ~ fitted_values)

bptest(lm_squared_resid)
```

Explanation:

- We first extract the fitted values and residuals from the Poisson regression model.
- Then, we fit a linear regression model of squared residuals on fitted values using the lm() function.
- Finally, we perform the Breusch-Pagan test for heteroscedasticity using the bptest() function from the lmtest package.

R CODE FOR CLUSTERING AND DENDROGRAM:

```
library(ggplot2)

data <- data.frame(

    State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan", "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal", "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh", "Tamil Nadu", "Bihar"), Total_Migrants = c(163275, 1854139, 443445, 2246796, 6067612, 19871, 669621, 1451881, 1981543, 886352, 1516952, 1195536,  1061732, 3602243), Work_Employment = c(5422, 67966, 39581, 64358, 244890, 627, 40895, 65103, 81690, 26541, 77216, 68378, 76790, 148008), Business = c(729, 11277, 1693, 15854, 26403, 69, 2711, 7172, 7748, 7152, 7653, 6804, 8202, 17403), Education = c(2857, 9380, 3151, 8875, 35140, 1336, 4193, 8665, 9840, 7403, 8014, 11485, 9591, 22696), Marriage = c(58578, 980892, 255876, 1351148, 3003996, 8025, 386746, 914724, 1296694, 451521, 809682, 610420, 471628, 1775188))

clustering_data <- data[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]

hc <- hclust(dist(clustering_data), method = "ward.D")

plot(hc, hang = -1, main = "Dendrogram of States")
```

Explanation:

- The ggplot2 library, which is used for data visualization in R. However, it is not used in the subsequent code. You might consider removing this if it's not needed for other parts of your project.
- The block creates a data frame named data with columns for state names, total migrants, and migrants due to various reasons (work employment, business, education, and marriage).
- Here, a subset of the data frame is created containing only the numeric columns relevant for clustering analysis. The selected columns are Total Migrants, Work Employment, Business, Education, and Marriage.
- dist(clustering data): Calculates the Euclidean distance matrix for the selected columns of the data frame.
- hclust(..., method = "ward.D"): Performs hierarchical clustering on the distance matrix using Ward's method. Ward's method minimizes the total within-cluster variance. ion, and Marriage.
- plot(hc, hang = -1): Plots the dendrogram of the hierarchical clustering. The hang = -1 argument ensures that the leaves (state names) are plotted at the same level. - main = "Dendrogram of States": Sets the title of the plot.

FINDINGS AND INTERPRETATIONS FROM THE ANALYSIS

INTERPRETETIONS FROM THE GLM MODELS:

At first we arranged the collected data in a suitable format and created the tables for different census years along with their rural urban division. So we generated 9 tables, as a example we have represented a data table which is as follows:

TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS					
STATES	TOTAL_NUMBER_OF_MIGRANTS	NUMBER OF MIGRANTS DUE TO WORK	NUMBER OF MIGRANTS DUE TO BUSINESS	NUMBER OF MIGRANTS DUE TO EDUCATION	NUMBER OF MIGRANTS DUE TO MARRIAGE
JAMMU KASHMIR	163275	5422	729	2857	58578
MAHARASHTRA	1,854,139	67,966	11,277	9,380	980,892
CHATTISHGARH	443,445	39,581	1,693	3,151	255,876
RAJHASTHAN	2246796	64358	15854	8875	1,351,148
UTTAR PRADESH	6067612	244890	26403	35140	3003996
ARUNACHAL PRADESH	19871	627	69	1336	8025
ODISHA	669621	40895	2711	4193	386746
WEST BENGAL	1451881	65103	7172	8665	914724
MADHYA PRADESH	1981543	81690	7748	9840	1296694
GUJRAT	886352	26541	7152	7403	451521
KARNATAKA	1516952	77216	7653	8014	809682
ANDHRA PRADESH	1195536	68378	6804	11485	610420
TAMILNADU	1061732	76790	8202	9591	471628
BIHAR	3602243	148008	17403	22696	1775188

According to this data tables we performed GLM on each of 9 sets. As the data is a count data so we used poisson as the family distribution of the GLM models. From the summary of the models we gathered the key findings and stated the interpretation.

RESULTS FROM THE DIFFERENT GLM MODELS															
MIGRATION DUE TO EMPLOYMENT				MIGRATION DUE TO BUSINESS				MIGRATION DUE TO EDUCATION				MIGRATION DUE TO MARRIAGE			
DIFFERENT MODELS	ESTIMATE	STD.ERROR	Z-VALUE	DIFFERENT MODELS	ESTIMATE	STD.ERROR	Z-VALUE	DIFFERENT MODELS	ESTIMATE	STD.ERROR	Z-VALUE	DIFFERENT MODELS	ESTIMATE	STD.ERROR	Z-VALUE
2011_TOTAL	5.549e-06	2.463e-08	225.3	2011_TOTAL	6.812e-05	1.158e-07	588.1	2011_TOTAL	-5.315e-05	1.531e-07	-347.	2011_TOTAL	4.099e-07	1.402e-09	292.4
2011_RURAL	-1.810e-06	5.683e-08	-31.86	2011_RURAL	2.052e-04	7.133e-07	287.63	2011_RURAL	-5.227e-05	6.847e-07	-76.34	2011_RURAL	1.593e-06	2.817e-09	565.42
2011_URBAN	5.703e-06	3.494e-08	163.2	2011_URBAN	1.805e-04	1.902e-07	949.4	2011_URBAN	6.794e-05	1.625e-07	418.1	2011_URBAN	-2.321e-06	4.331e-09	-536.0
2001_TOTAL	1.093e-05	2.467e-08	442.9	2001_TOTAL	1.136e-04	1.488e-07	763.8	2001_TOTAL	-5.996e-05	2.203e-07	-272.2	2001_TOTAL	4.312e-07	1.378e-09	312.9
2001_RURAL	1.800e-05	4.815e-08	373.8	2001_RURAL	2.045e-04	5.147e-07	397.3	2001_RURAL	-3.404e-04	1.207e-06	-282.0	2001_RURAL	2.002e-06	3.348e-09	597.9
2001_URBAN	1.568e-05	7.965e-08	196.84	2001_URBAN	2.498e-04	3.279e-07	761.73	2001_URBAN	1.444e-05	5.371e-07	26.89	2001_URBAN	-6.445e-07	4.277e-09	-150.71
1991_TOTAL	1.374e-05	2.668e-08	515.0	1991_TOTAL	3.410e-05	8.544e-08	399.1	1991_TOTAL	-5.620e-05	1.268e-07	-443.3	1991_TOTAL	1.171e-06	2.157e-09	542.8
1991_RURAL	1.561e-05	6.704e-08	232.9	1991_RURAL	5.497e-05	2.398e-07	229.2	1991_RURAL	-8.111e-05	7.560e-07	-107.3	1991_RURAL	2.488e-06	4.458e-09	558.2
1991_URBAN	2.510e-05	7.692e-08	326.33	1991_URBAN	1.086e-04	2.546e-07	426.67	1991_URBAN	8.301e-06	3.959e-07	20.97	1991_URBAN	-1.283e-06	1.134e-08	-113.19

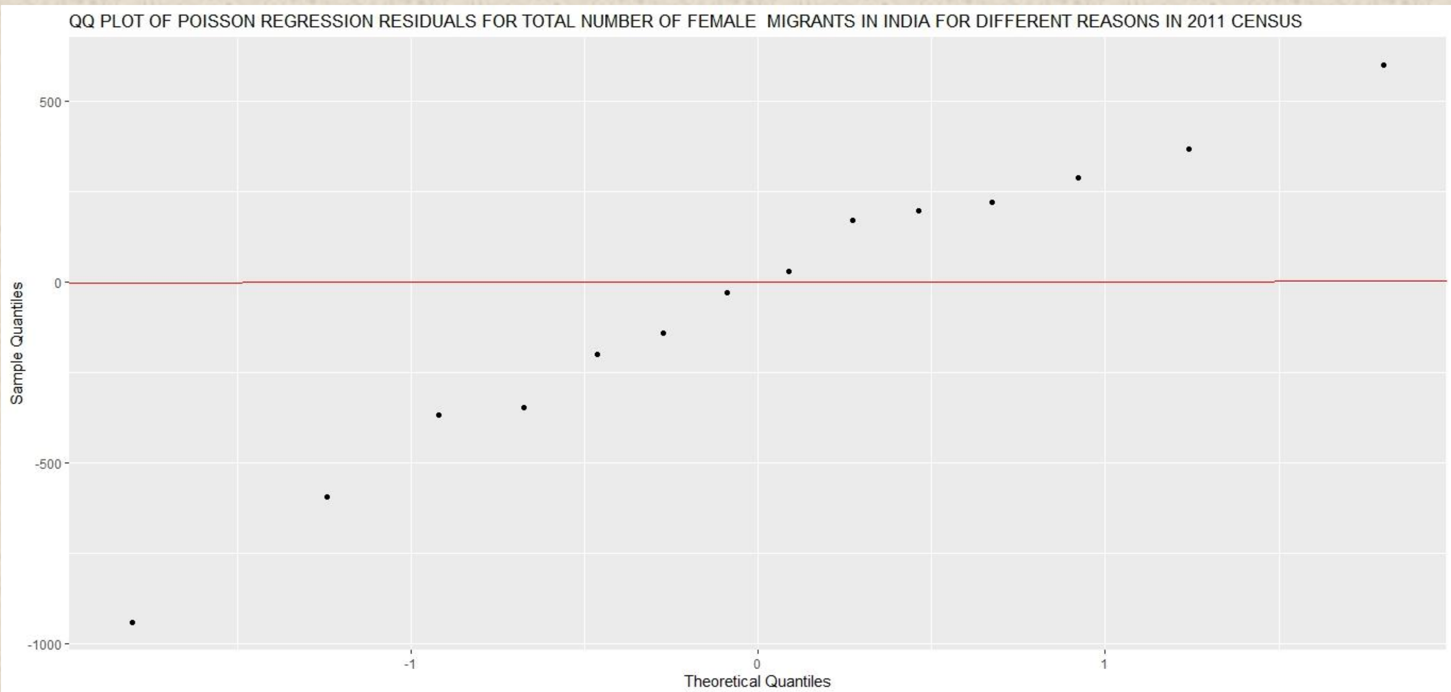
INTERPRETETIONS FROM THE ABOVE TABLE:

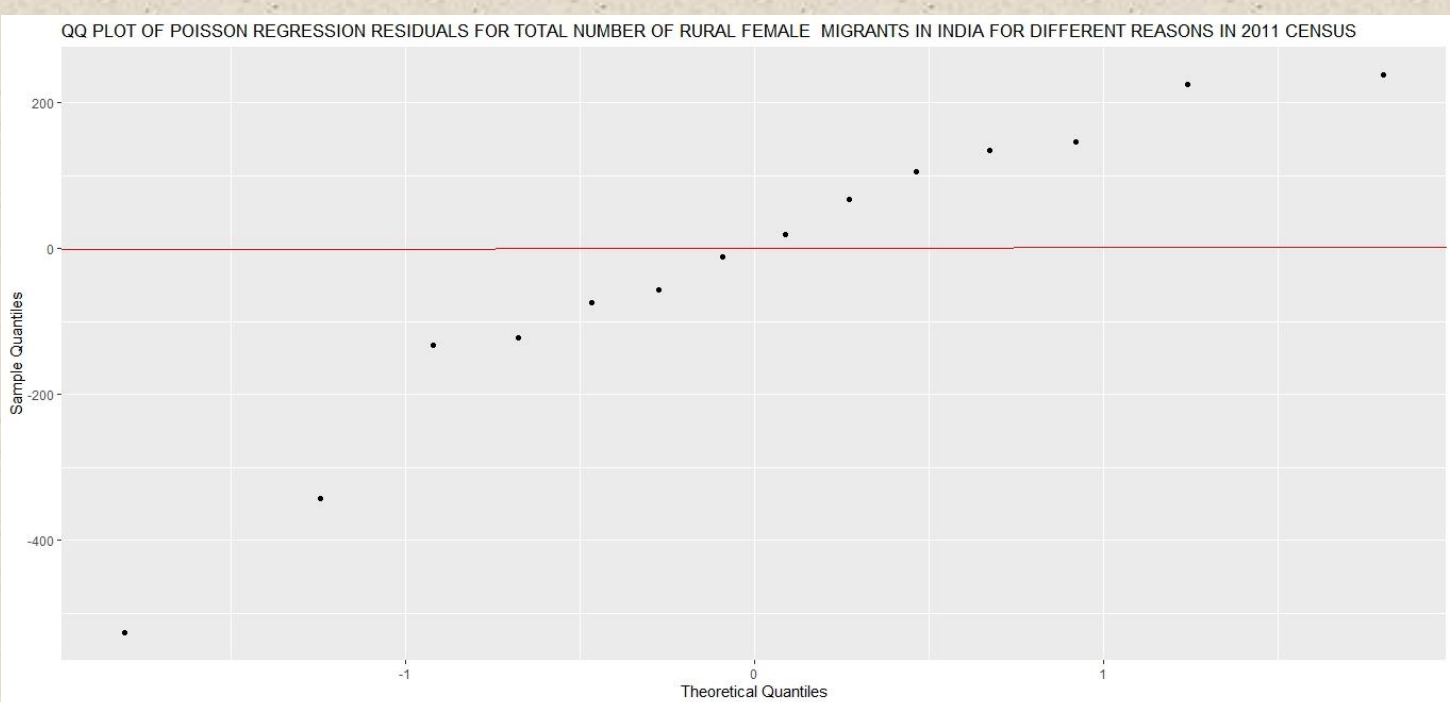
From the above table we can conclude that:

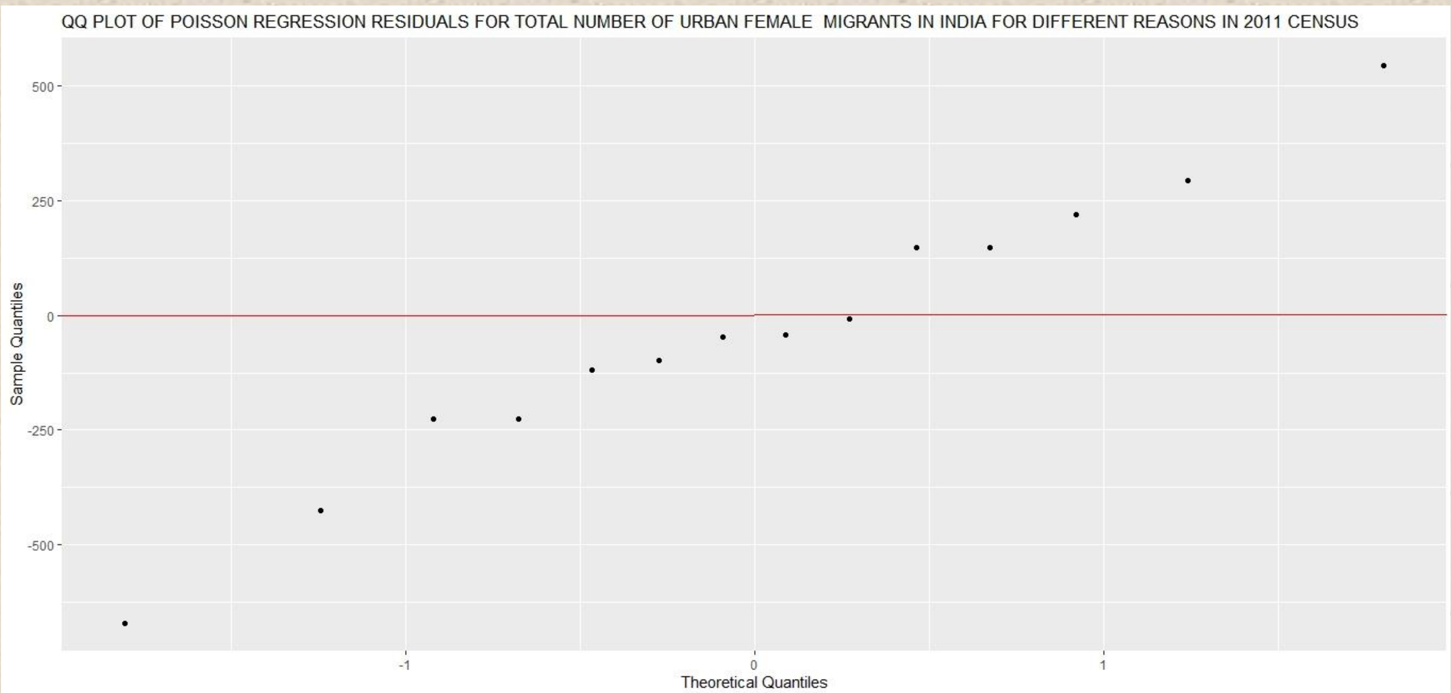
- Migration due to work employment is inversely related to the response variable Total Migration only for 2011_rural model. And for rest of all it is positively related to Total Migration with above mentioned calculated values.
- Migration due to business is positively related to the response for all of the considered models and the effect values of business migration on response for different models are mentioned in the above table.
- Migration due to education is positively related to the response only for the urban region models (2011_urban, 2001_urban, 1991_urban) and on the other hand for the rural sectors the education migration is inversely related to Total Migration.
- Migration due to marriage is positively related to the response only for the rural region models (2011_rural, 2001_rural, 1991_rural) and on the other hand for the urban sectors the marital migration is inversely related to Total Migration.

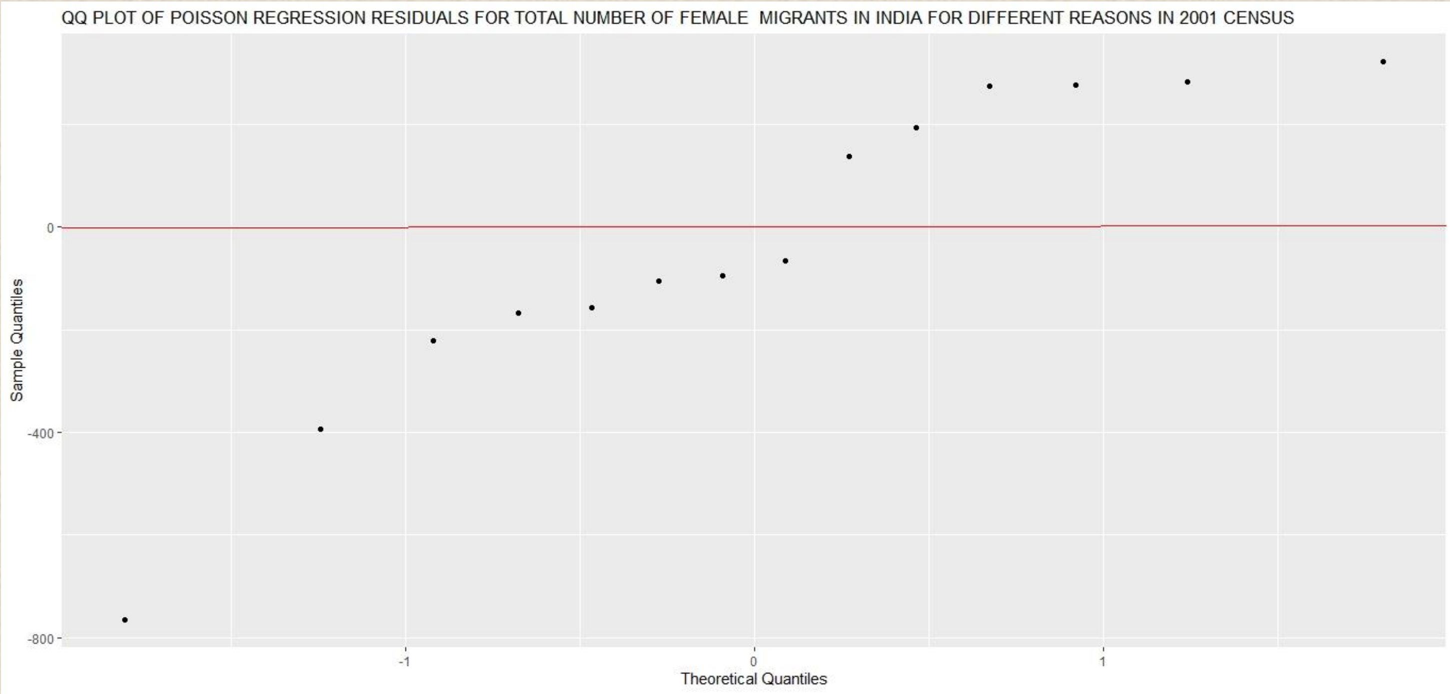
INTERPRETETIONS FROM THE QQ PLOTTING:

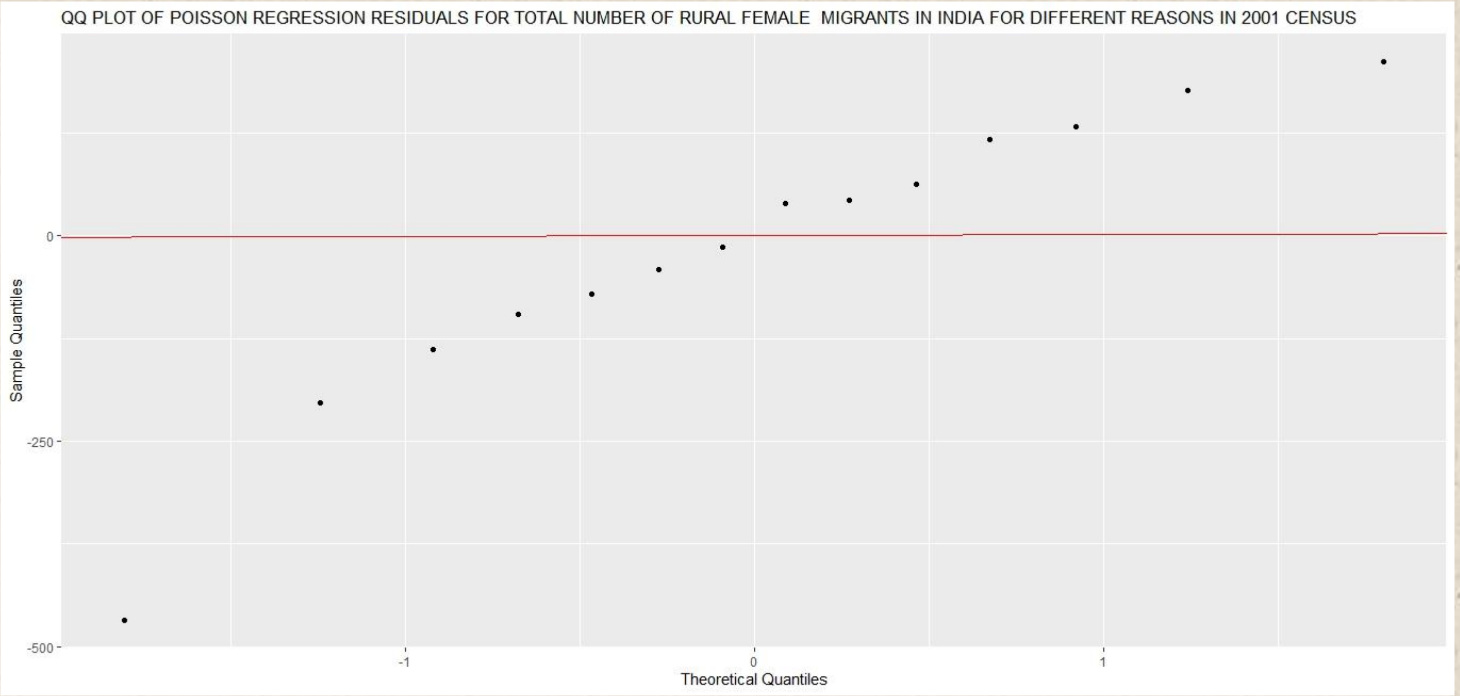
The generated QQ plots from the residuals of GLM tables :

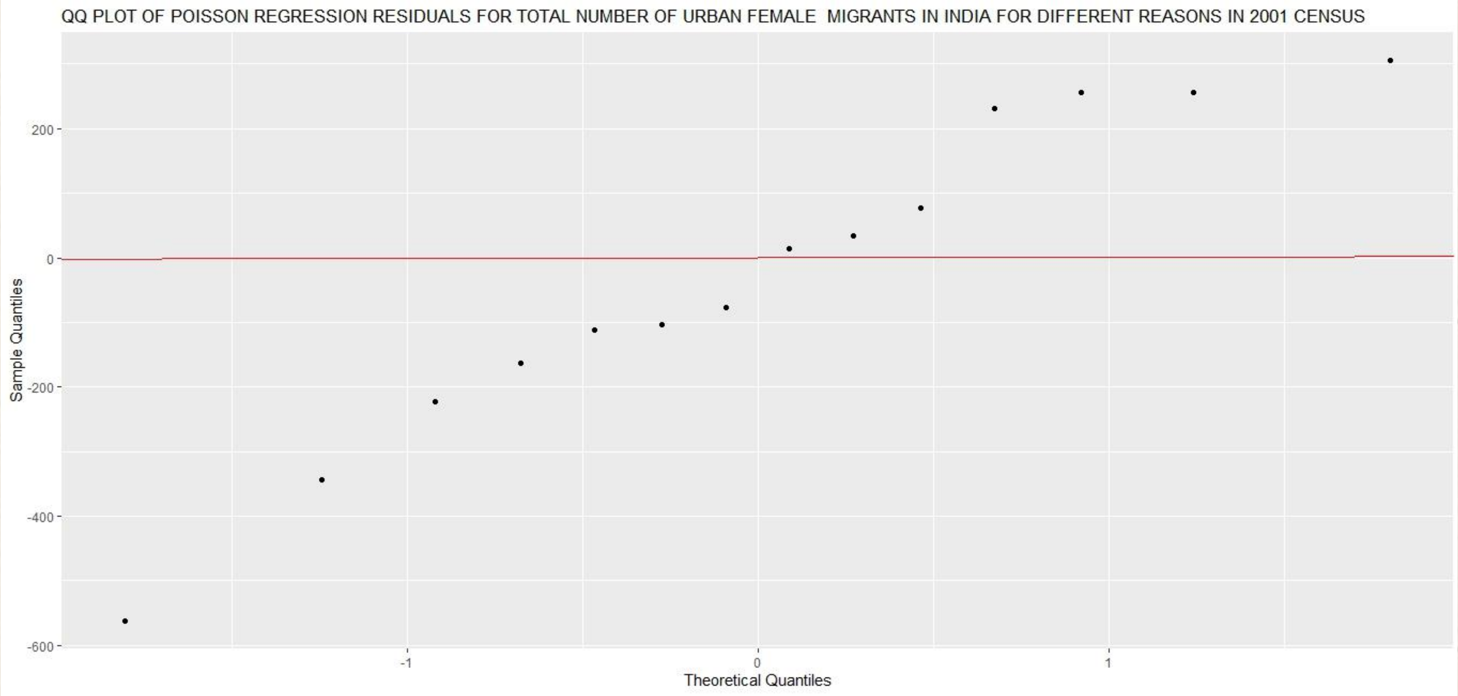


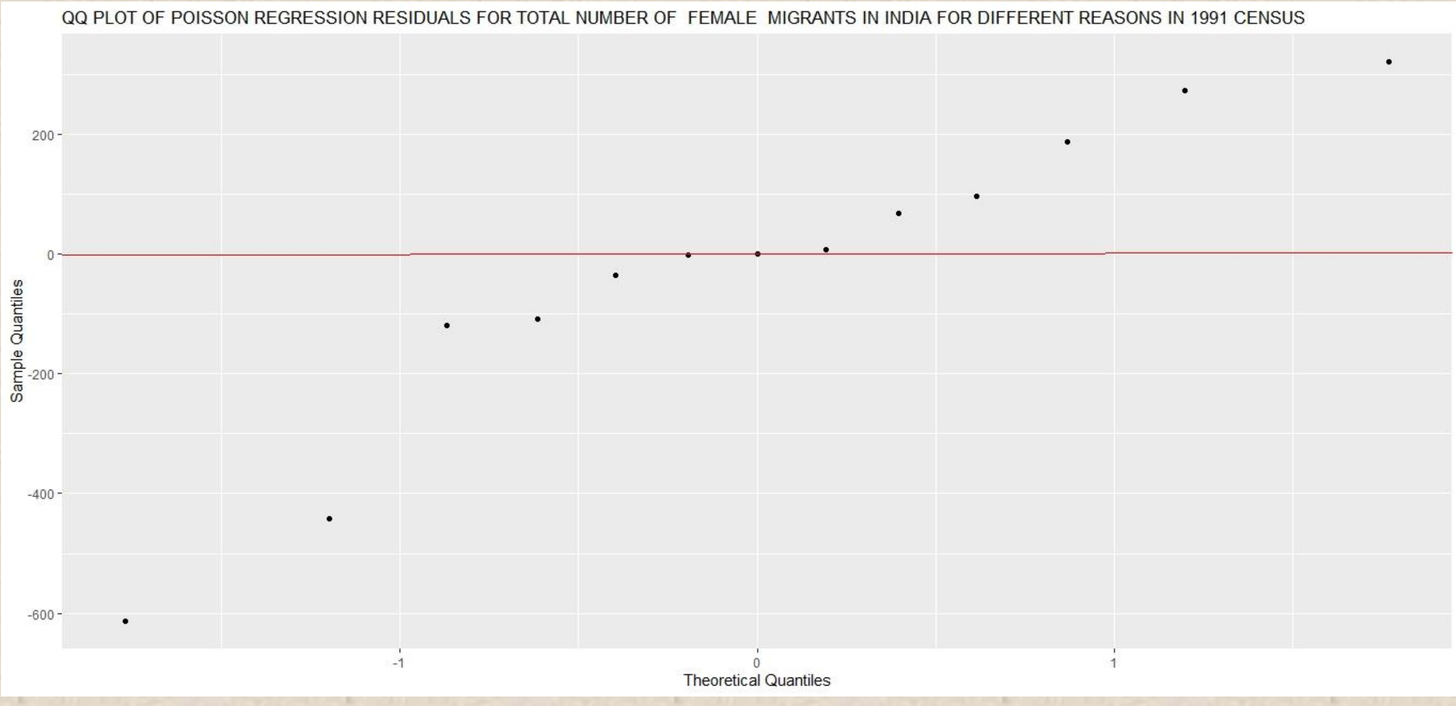


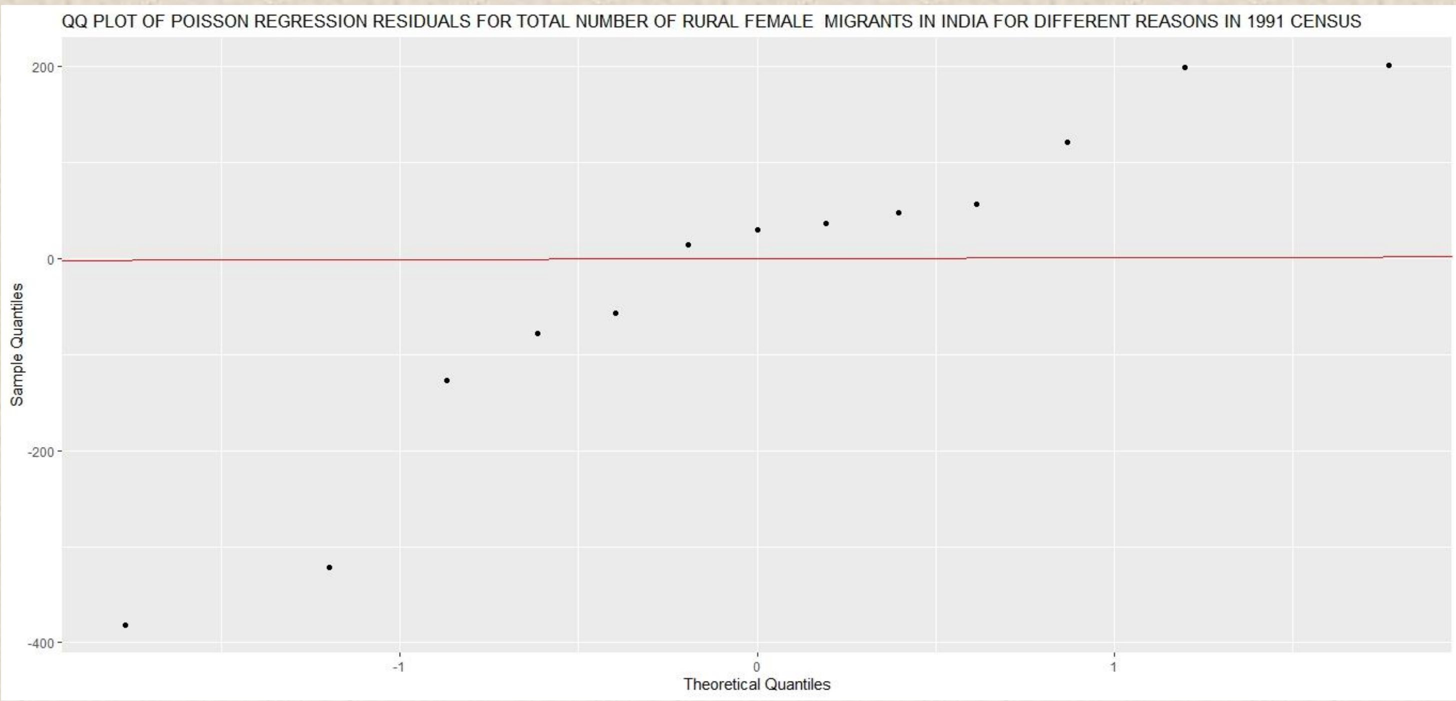

QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS
Sample Quantiles
500
250
0
-250
-500
-1
0
1
Theoretical Quantiles

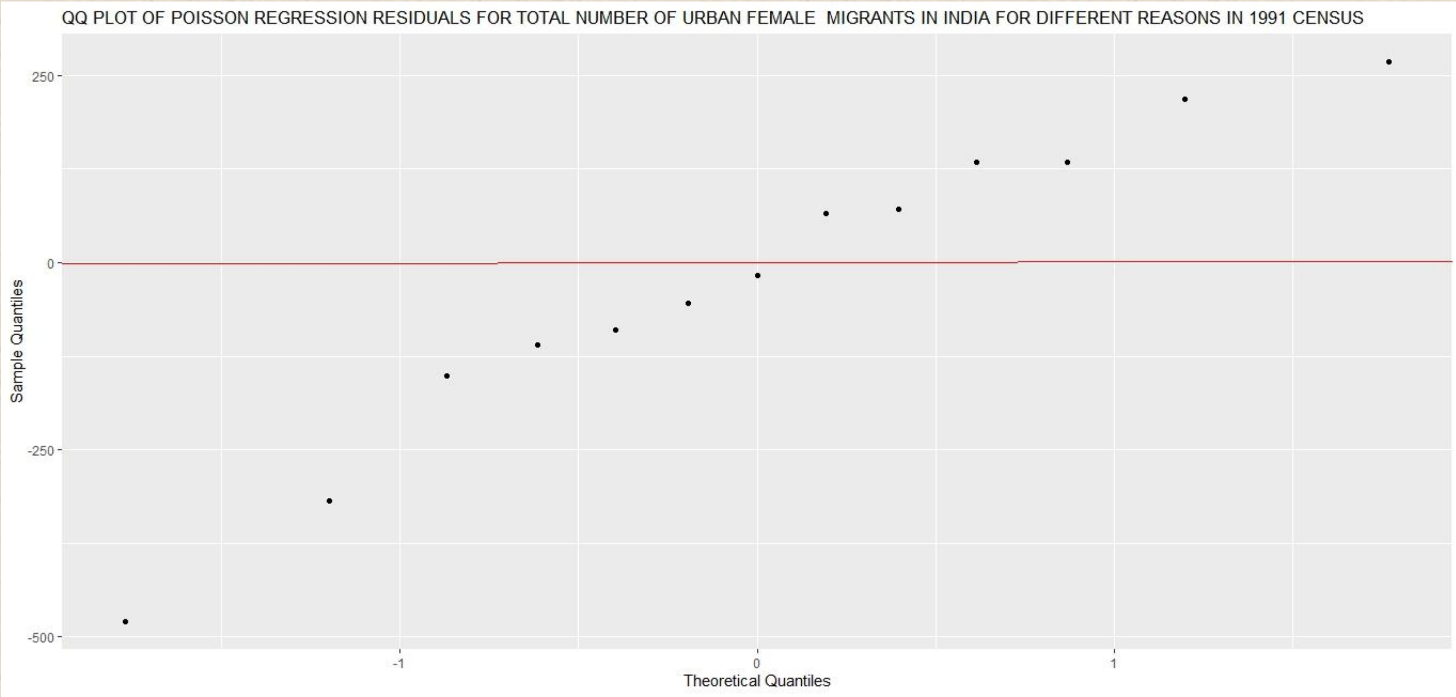

QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS
Sample Quantiles
0
-400
-800
-1
0
1
Theoretical Quantiles


QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS
Sample Quantiles
0
-250
-500
-1
0
1
Theoretical Quantiles


QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS
Sample Quantiles
200
0
-200
-400
-600
-1
0
1
Theoretical Quantiles


QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS
Sample Quantiles
Theoretical Quantiles
200
0
-200
-400
-600
-1
0
1


QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS
Sample Quantiles
Theoretical Quantiles
200
0
-200
-400
-1
0
1



KEY FINDINGS FROM THE PLOTS:

From the above plots we can conclude that

- For all the models the plots indicates a linear trend among the residual values collected from the model summaries.
- For majority of the plots it is visible that most of the residual plots are clustered near the red line indicating that fitting of the GLM model is quiet good and significant.
- Although some outliers in the plot indicates that there is a presence of little bit over dispersion in the model.

INTERPRETETIONS FROM THE BREUSCH-PAGAN TEST FOR HETEROSCEDASTICITY:

Here the null hypothesis (H0) and alternative hypothesis (H1) are:

H0: Homoscedasticity is absent

H1: Heteroscedasticity is present

In our model, we have considered that the variances of error terms are equal, i.e., there is no heteroscedasticity. To check the heteroscedasticity of error terms, we used the Breusch-Pagan test.

Breusch-Pagan test:

In the Breusch-Pagan test, the null hypothesis is that the variances of error terms are equal, i.e., there is no heteroscedasticity in a linear regression model. The test assumes that the error terms are normally distributed.

The test results are:

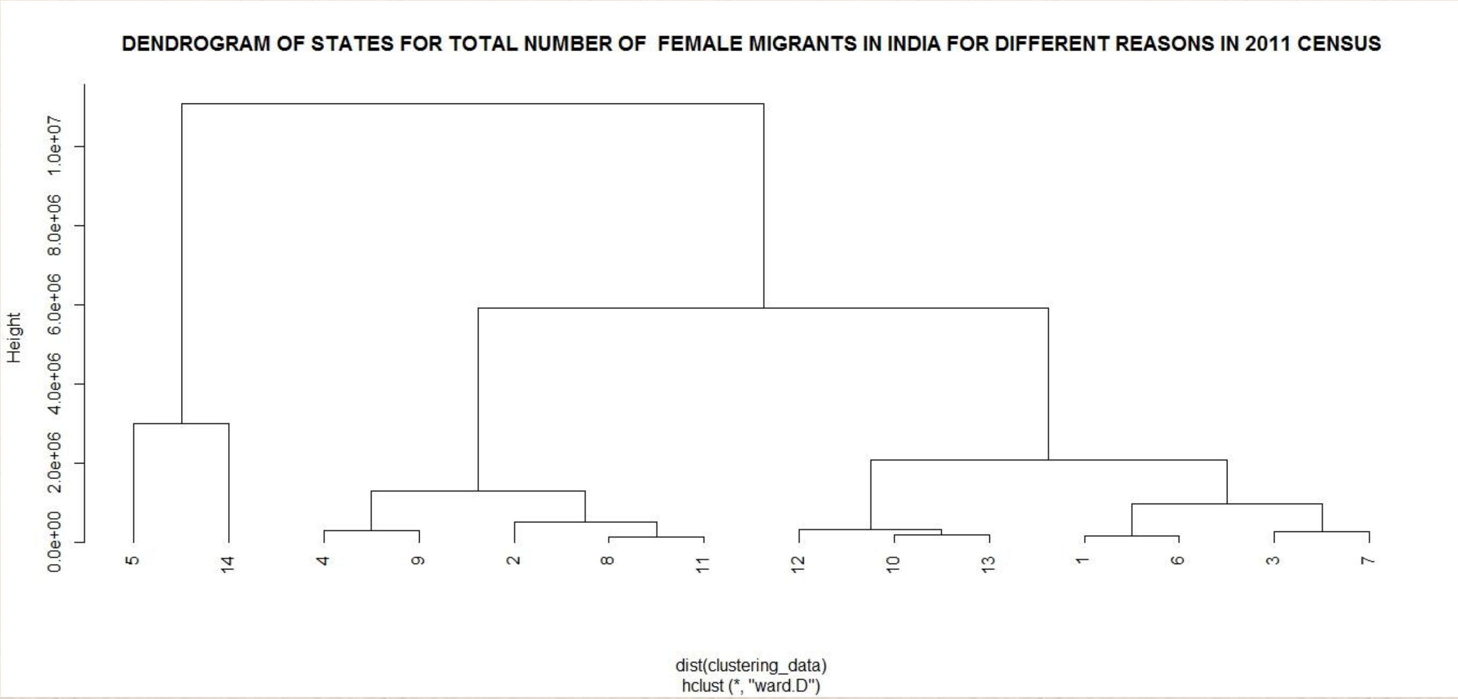
TEST RESULTS FOR HETEROSCEDASTICITY TEST		
DIFFERENT MODELS	BP VALUE	P-VALUE
2011_TOTAL	**0.67071**	**0.4128**
2011_RURAL	**1.5387**	**0.2148**
2011_URBAN	**0.91657**	**0.3384**
2001_TOTAL	**0.63769**	**0.4245**
2001_RURAL	**1.0455**	**0.3065**
2001_URBAN	**0.66977**	**0.4131**
1991_TOTAL	**0.6887**	**0.4066**
1991_RURAL	**2.1559**	**0.142**
1991_URBAN	**0.61373**	**0.4334**

We can reject the null hypothesis if p-value is less than 0.05. If the p-value is greater than 0.05, we cannot reject the null hypothesis of homoscedasticity or constant variance.

From the above table we can conclude that for all the models the p-values are greater than 0.05, therefore, we fail to reject null hypothesis at 5% level of significance and thus we can conclude that there is no heteroscedasticity present, All models are significant and the error variances are all equal.

INTERPRETETIONS FROM THE CLUSTERING AND DENDROGRAM:

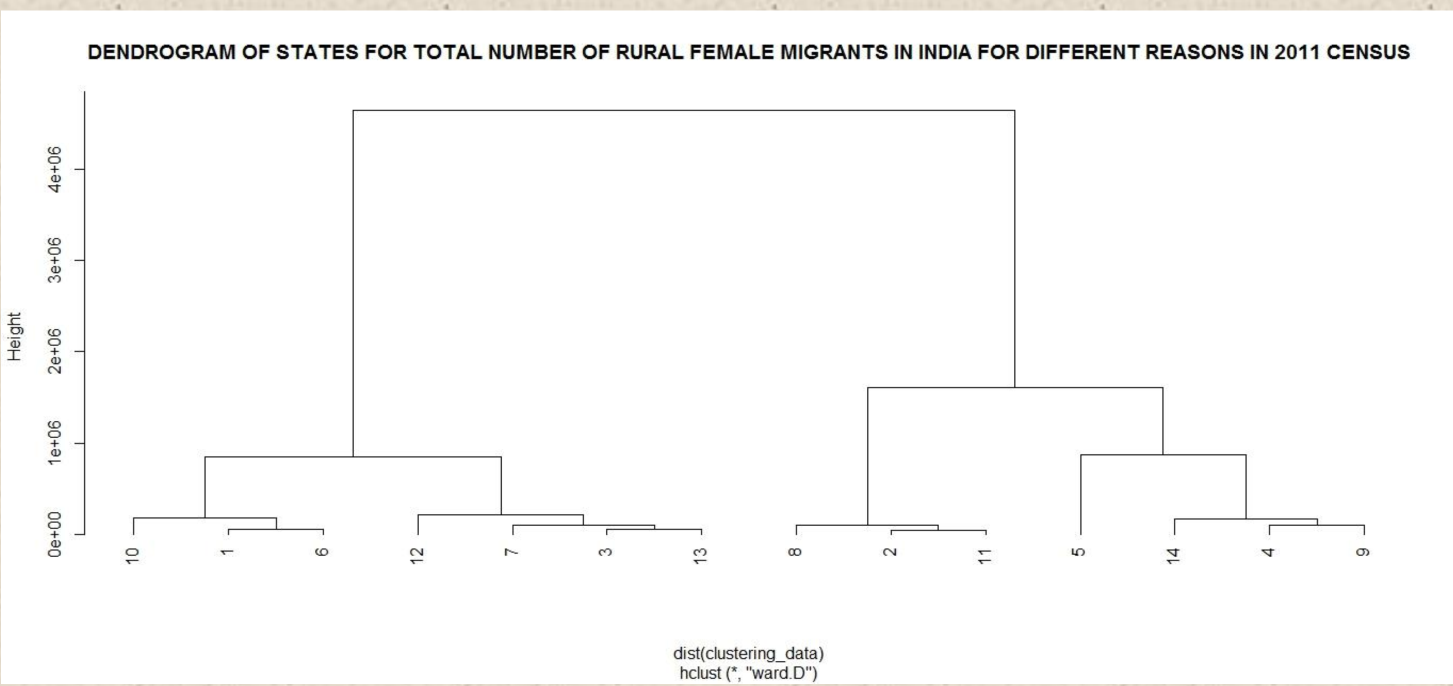
The generated Dendrograms from the data sets are as follows:



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGARH	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

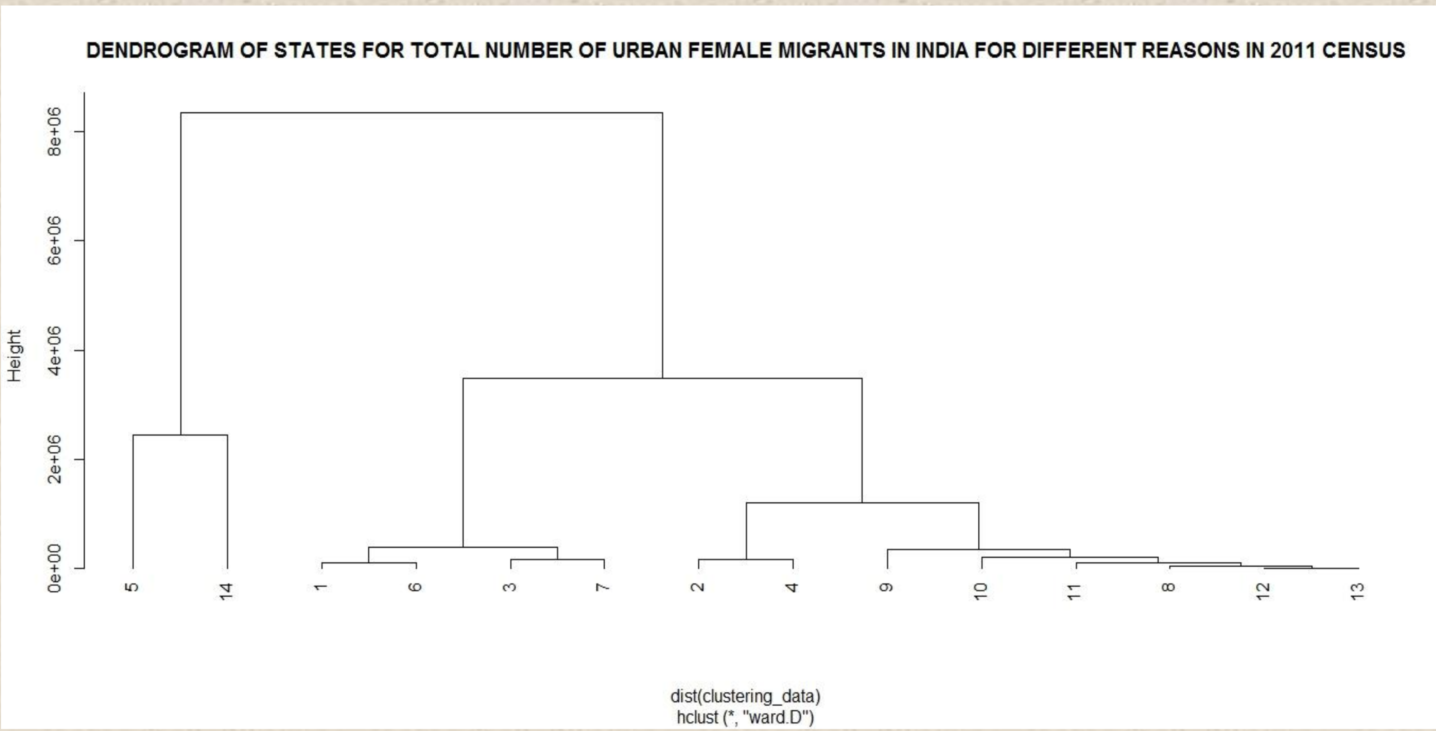
- There are two primary clusters that form at the highest level (around 1.0e+07 height).The first main cluster includes states 5 and 14.The second main cluster includes all the other states: 4, 9, 2, 8, 11, 12, 10, 13, 1, 6, 3, and 7.Sub-clusters within the First Main Cluster: States 5 and 14 form a very distinct sub-cluster, showing high similarity in terms of female migrant numbers.
- Sub-clusters within the Second Main Cluster: States 4 and 9 form one sub-cluster, indicating high similarity between these two states. States 2, 8, and 11 form another sub-cluster, with: States 8 and 11 being very similar.
- State 2 joins this sub-cluster at a slightly higher height, indicating moderate similarity with states 8 and 11.States 12, 10, and 13 form another distinct sub-cluster, showing: States 12 and 10 being highly similar.
- State 13 joins this sub-cluster at a higher height, indicating moderate similarity with states 12 and 10.States 1, 6, 3, and 7 form another group, with further subdivisions: States 1 and 6 show high similarity. States 3 and 7 are also highly similar, forming another subgroup within this cluster.



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGARH	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

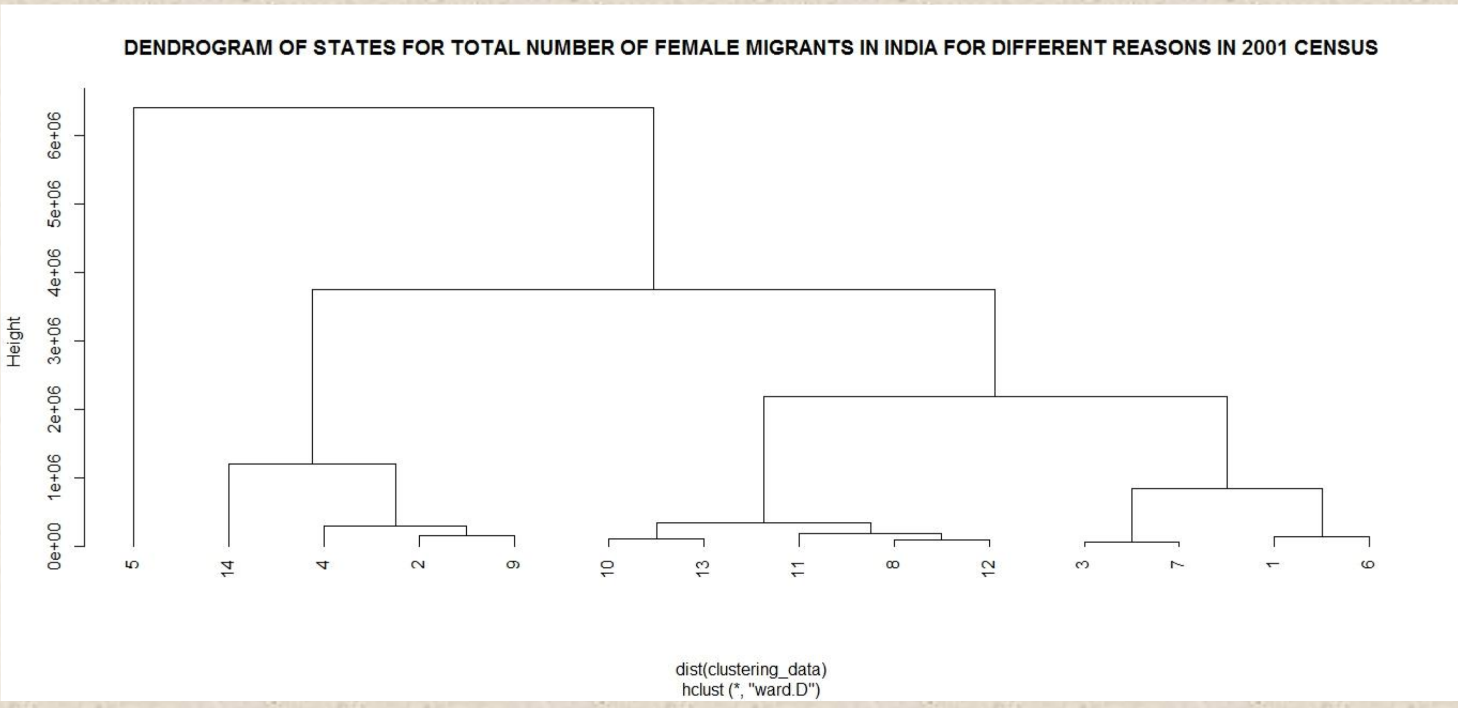
- Starting from the left: States 10 and 1 are very similar and form the first small cluster. This cluster then joins with state 6, indicating these three states have similar migration patterns. This larger cluster joins with states 12, 7, 3, and 13, indicating these seven states are somewhat similar in their migration numbers.
- In the middle: States 8 and 2 are quite similar and form a small cluster. This cluster then joins with state 11.On the right: States 5, 14, 4, and 9 forms another cluster, with states 4 and 9 being the most similar within this group.
- The major clusters can be interpreted as groups of states that have similar patterns in rural female migration for different reasons. For example, the first major cluster (10, 1, 6, 12, 7, 3, 13) suggests these states might have similar socio-economic factors or cultural practices influencing female migration. The clusters on the right (5, 14, 4, 9) might represent states with different migration dynamics compared to the other clusters.



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGAR H	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

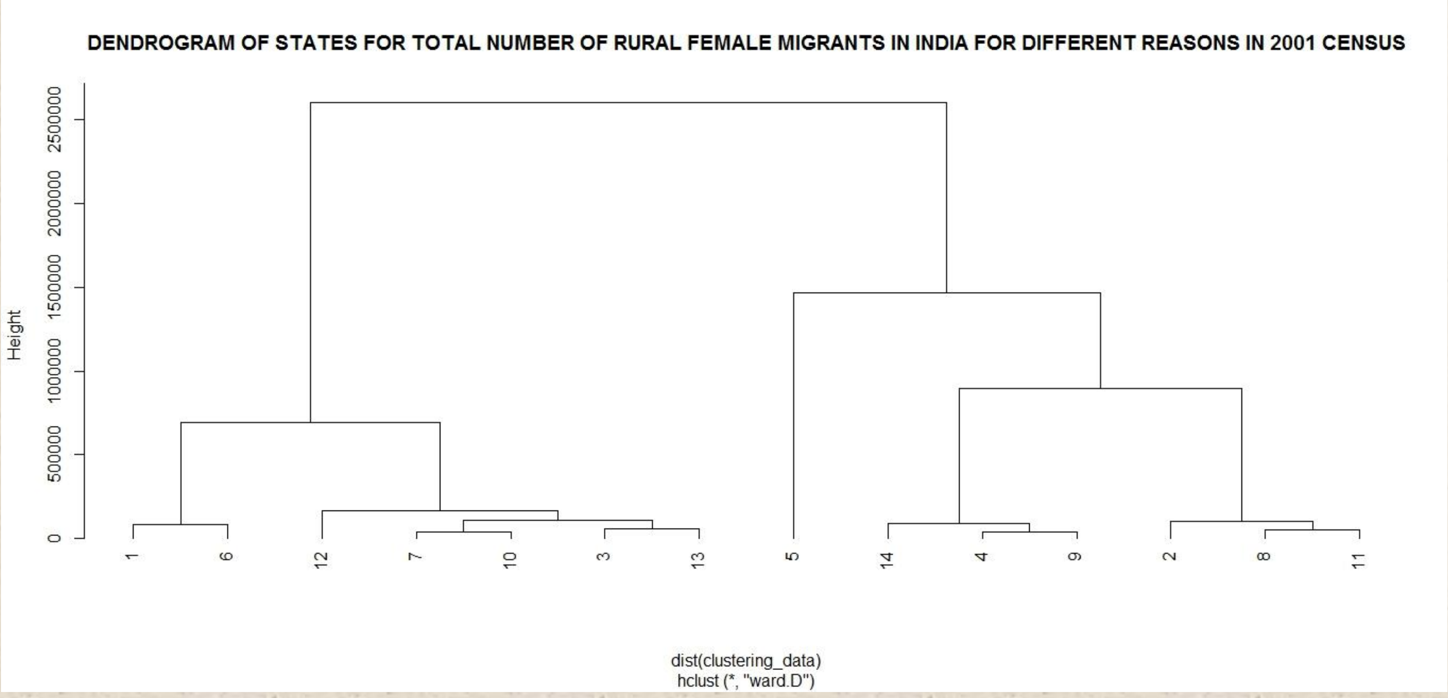
- At the highest level, the states are divided into two primary clusters: Cluster 1: Contains state numbers 5 and 14.Cluster 2: Contains all the other states.
- Secondary Clusters within Clusters Sub-cluster 2A: Contains state numbers 1, 6, 3, and 7. These states are grouped together, indicating they have similar numbers of urban female migrants.
- Sub-cluster 2B: Contains the remaining states (2, 4, 9, 10, 11, 8, 12, 13), which are further divided into smaller clusters based on their similarities. Key Observations Distinct Grouping of States 5 and 14.
- States 5 and 14 form a separate cluster at a high distance (height), indicating that the number of urban female migrants in these states is significantly different from those in other states.
- Formation of Smaller Clusters: The dendrogram shows that within the larger cluster, states are grouped into smaller clusters that are more similar to each other in terms of urban female migration numbers. For example, states 2, 4, and 9 are more similar to each other and form a distinct sub-cluster within the larger group.



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGARH	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

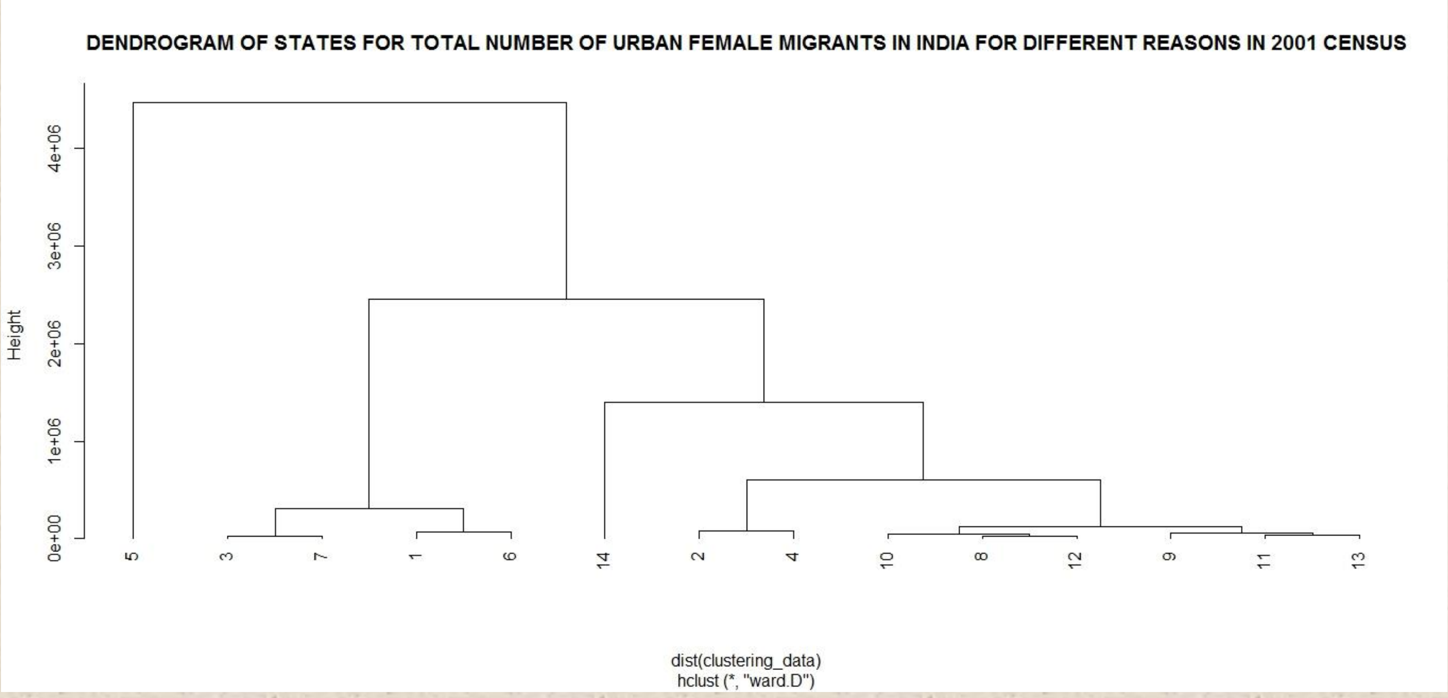
- Cluster 1: State 5 form a distinct cluster on its own, indicating a significant difference from other states in terms of female migration.
- Cluster 2: States 14, 4, 2, and 9 forms a cluster.
- Cluster 3: States 10, 13, 11, 8, and 12 forms another cluster.
- Cluster 4: States 3, 7, 1, and 6 are grouped together in the fourth cluster.
- Within Cluster 2, there are further sub-clusters: States 4 and 14 are closer to each other. States 2 and 9 form another sub-cluster.
- Within Cluster 3: States 10 and 13 are closely related. States 11, 8, and 12 forms a tighter sub-cluster with states 11 and 8 being more similar to each other.
- Within Cluster 4: States 1 and 6 are very similar to each other. States 3 and 7 are closely related.



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGARH	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

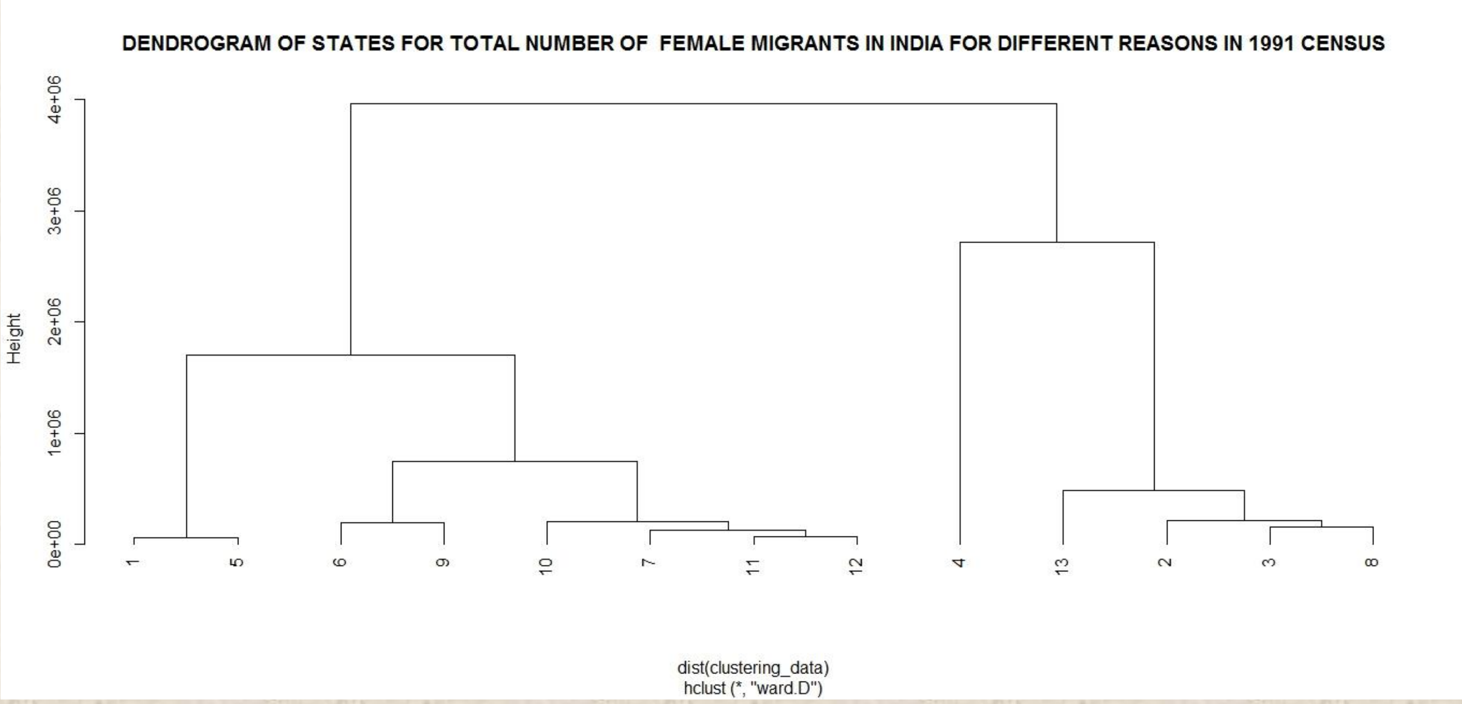
- There are several distinct clusters visible: Cluster 1: States labeled 1 and 6.Cluster 2: States labeled 12, 7, 10, and 3.Cluster 3: State labeled 13.Cluster 4: State labeled 5.Cluster 5: States labeled 14 and 4.Cluster 6: States labeled 9, 2, 8, and 11.
- The largest branches occur at the highest levels, indicating major clusters of states. For example, the split between the left cluster (states 1, 6, 12, 7, 10, 3, 13) and the right cluster (states 5, 14, 4, 9, 2, 8, 11) happens at a high level, indicating substantial differences between these two major groups.
- Higher clusters suggest greater dissimilarity. For instance, the cluster joining state 5 with the group containing states 14, 4, 9, 2, 8, and 11 has a high height, indicating that state 5 is quite different from the other states in terms of rural female migration patterns.



1	2	3	4	5	6	7	8	9	10	11	12	13	14
JAMMU KASHMIR	MAHARASHTRA	CHATTISHGARH	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

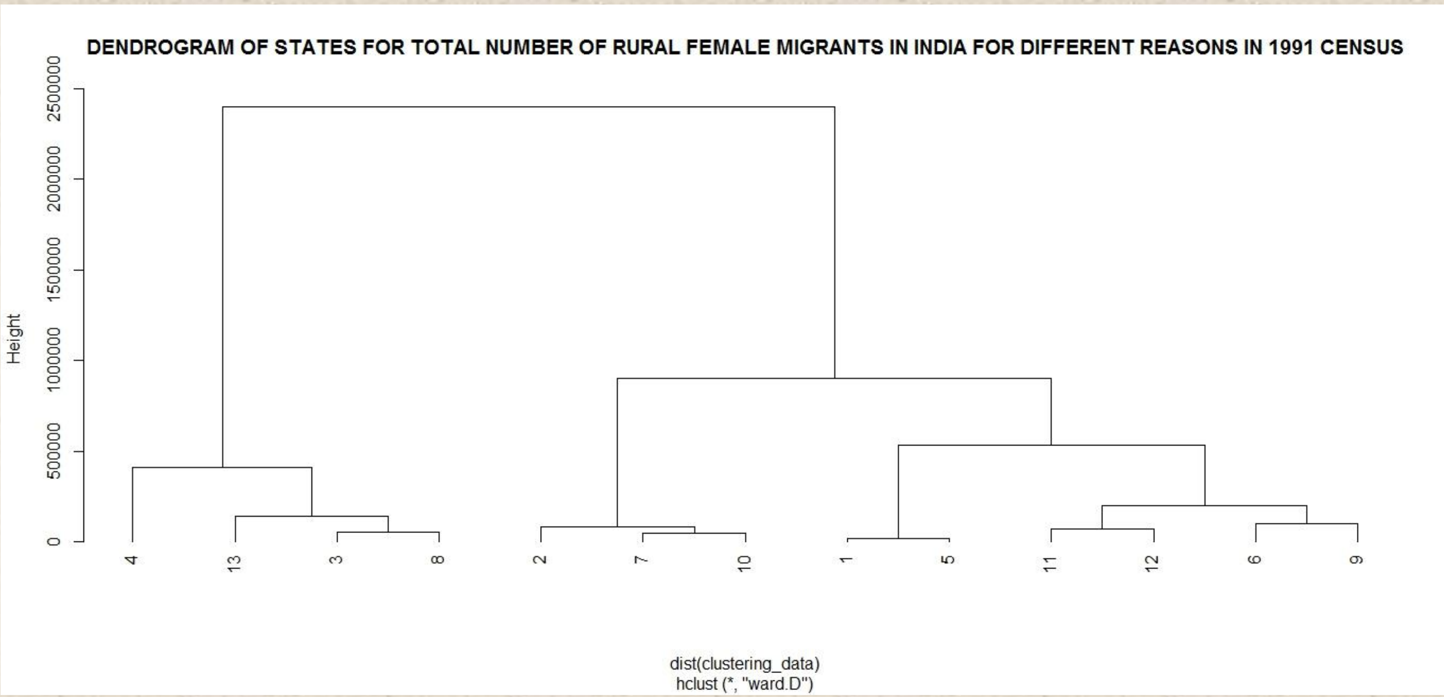
- Major Clusters: Cluster 1: State 5 stands alone, indicating it has a distinct urban female migration pattern compared to other states. Cluster 2: States 3, 7, 1, and 6 forms a cluster. Cluster 3: States 14, 2, and 4 forms another cluster. Cluster 4: States 10, 8, 12, 9, 11, and 13 forms the final major cluster.
- Sub-Clusters: Within Cluster 2: States 3 and 7 are closely related. States 1 and 6 are grouped together, indicating similar migration patterns between them. Within Cluster 3: State 14 form a sub-cluster with states 2 and 4, with states 2 and 4 being more closely related. Within Cluster 4: States 10 and 8 are closely related. States 12, 9, 11, and 13 form a tighter sub-cluster, with states 12 and 9 being particularly similar.



1	2	3	4	5	6	7	8	9	10	11	12	13
JAMMU KASHMIR	MAHARASHTRA	RAJHASTHA N	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

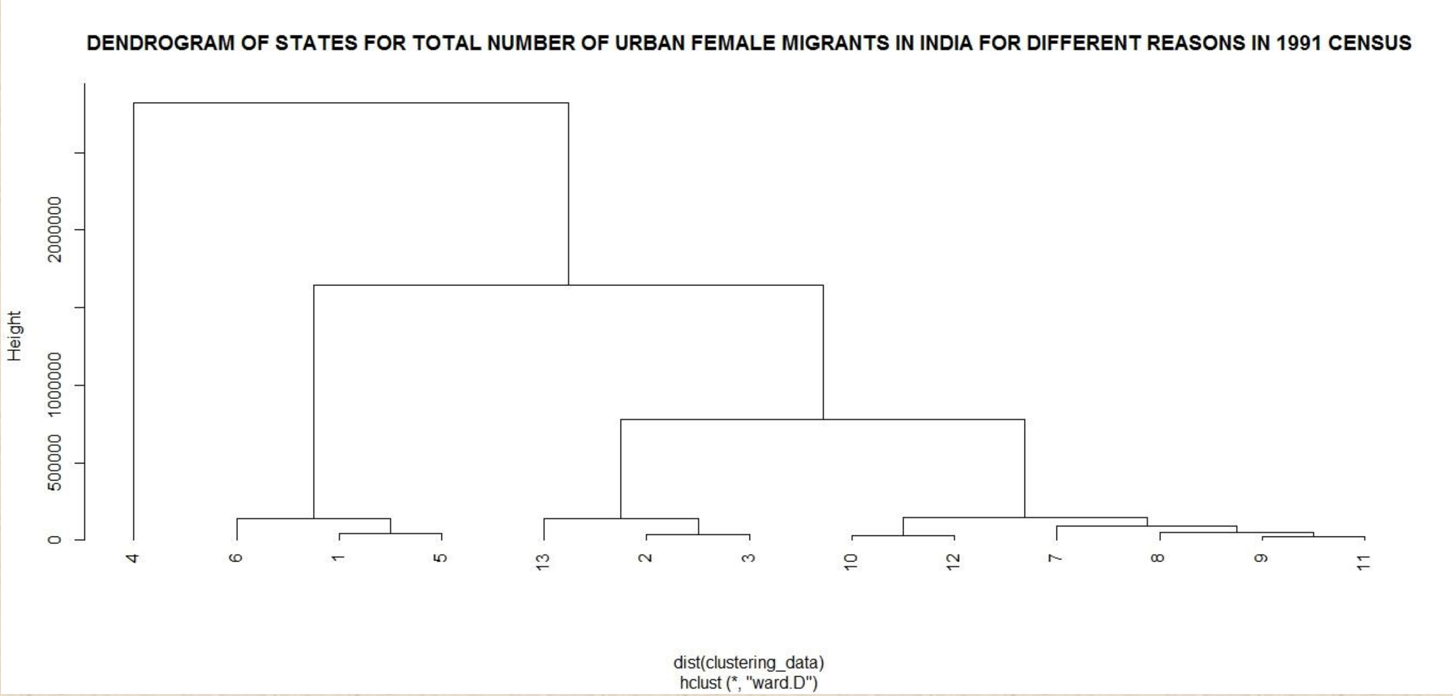
- The states are grouped into clusters based on their similarity in female migrant numbers. The height at which states or clusters are joined represents the dissimilarity or distance between them. Lower heights indicate higher similarity. Main Clusters: Two main clusters are formed at the highest level (around 3.5e+06 heights).The first main cluster includes states 1, 5, 6, 9, 10, 7, 11, and 12.
- The second main cluster includes states 4, 13, 2, 3, and 8.Sub-clusters within the First Main Cluster: States 1 and 5 form a distinct sub-cluster. States 6, 9, 10, 7, 11, and 12 form another sub-cluster, with further subdivisions: States 6 and 9 are more similar to each other.
- States 10, 7, 11, and 12 form another group, with 10 and 7 being very similar, and 11 and 12 showing another close similarity.
- Sub-clusters within the Second Main Cluster: State 4 stands alone until it joins with other states at a higher height, indicating it has a significantly different pattern compared to the others. States 13, 2, 3, and 8 form another sub-cluster, with: States 2, 3, and 8 forming a smaller cluster, showing a close relationship among them. State 13 joins this cluster at a slightly higher height, indicating moderate similarity.



1	2	3	4	5	6	7	8	9	10	11	12	13
JAMMU KASHMIR	MAHARASHTRA	RAJHASTHA N	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

- There are two primary clusters that form at the highest level (around 2,000,000 heights).The first main cluster includes states 4, 13, 3, and 8.The second main cluster includes states 2, 7, 10, 1, 5, 11, 12, 6, and 9.
- Sub-clusters within the First Main Cluster: State 4 stands alone, indicating it has significantly different migration numbers compared to other states. States 13, 3, and 8 form a sub-cluster, with: States 13 and 3 showing a close similarity.
- State 8 joining this cluster at a higher height, indicating moderate similarity with states 13 and 3.Sub-clusters within the Second Main Cluster: States 2, 7, and 10 form one sub-cluster, with: States 2 and 7 being very similar.
- State 10 joining this sub-cluster at a higher height. States 1 and 5 form another distinct sub-cluster, showing high similarity between them. States 11, 12, 6, and 9 form another group, with further subdivisions: States 11 and 12 are closely similar.
- States 6 and 9 are also closely similar but form a separate group from 11 and 12, joining the cluster at a higher height.



1	2	3	4	5	6	7	8	9	10	11	12	13
JAMMU KASHMIR	MAHARASHTRA	RAJHASTHAN	UTTAR PRADESH	ARUNACHAL PRADESH	ODISHA	WEST BENGAL	MADHYA PRADESH	GUJRAT	KARNATAKA	ANDHRA PRADESH	TAMILNADU	BIHAR

KEY FINDINGS FROM THE PLOT:

- There are two primary clusters that form at the highest level (around 2,500,000 heights).The first main cluster includes states 4, 6, 1, and 5.The second main cluster includes states 13, 2, 3, 10, 12, 7, 8, 9, and 11.
- Sub-clusters within the First Main Cluster: State 4 stands alone, indicating it has significantly different migration numbers compared to other states. States 6, 1, and 5 form a sub-cluster, with: States 6 and 1 showing high similarity.
- State 5 joining this cluster at a slightly higher height, indicating moderate similarity with states 6 and 1.
- Sub-clusters within the Second Main Cluster: States 13, 2, and 3 form one sub-cluster, with: States 2 and 3 being very similar. State 13 joining this sub-cluster at a higher height, indicating moderate similarity with states 2 and 3.States 10 and 12 form another distinct sub-cluster, showing high similarity between them. States 7, 8, 9, and 11 form another group, with further subdivisions.
- States 9 and 11 show close similarity. States 7 and 8 are also closely similar, forming another subgroup within this cluster.

CONCLUSIONS

The analysis of migration patterns based on the census data reveals several critical insights into the dynamics of migration in India. Key findings include the varying relationships between different reasons for migration (work employment, business, education, and marriage) and the overall migration trends across urban and rural regions. Additionally, cluster analysis has identified distinct groups of states with similar migration patterns, particularly among female migrants.

Migration Trends Among Males And Females Across States:

- Male migrants vastly outnumber female migrants for work and education across all years, with urban migration being more prevalent than rural. The number of male migrants for work employment and education peaked in 2011, while business migration remained stable but male-dominated.
- Migration due to marriage is predominantly female and increased significantly over the years, with rural migration being higher than urban, though both segments show significant numbers.
- Migration for work employment and marriage are the largest categories, with work-related migration being male-dominated and urban-focused, and marriage-related migration being female-dominated and rural-focused.
- States like Jammu and Kashmir and Gujarat follow national trends, with employment and education migration patterns mirroring those of India, though business migration decreased over the years. Tamil Nadu shows a strong influence of urban sectors on its migration patterns.

Migration Reasons and Total Migration:

- Work Employment: Generally shows a positive relationship with total migration, except in the 2011 rural model where an inverse relationship is observed.
- Business: Consistently positively related to total migration across all models.
- Education: Positively related to total migration in urban areas, but inversely related in rural areas.
- Marriage: Positively related to total migration in rural areas, while inversely related in urban areas.

Cluster Analysis:

- Two primary clusters emerge at the highest level, revealing major groups of states with similar migration patterns.
- First Main Cluster: Includes distinct sub-clusters like states 4 and 13, 3, and 8, with state 4 standing alone due to its significantly different migration numbers.
- Second Main Cluster: Features sub-clusters such as states 2, 7, and 10, and another group comprising states 11, 12, 6, and 9, indicating high similarity within these groups.
- Distinct Grouping: States 5 and 14 form a very distinct sub-cluster, highlighting their unique migration dynamics.

Cultural and Socio-Economic Influences:

- The clusters suggest that states within the same group may share similar socio-economic factors or cultural practices influencing migration, particularly in terms of female migration patterns.

Overall, the migration patterns in India exhibit complex relationships influenced by various socio-economic factors. Work and marriage remain the predominant reasons for migration, with significant gender-specific trends—work-related migration being male-dominated and marriage-related migration being female-dominated. The identified clusters of states provide valuable insights into regional similarities and differences, offering a foundation for targeted policy interventions to address migration-related challenges.

APPENDIX

ALL R SOURCE CODE IN THIS PROJECT:

```
install.packages("MASS", repos = "https://cloud.r-project.org/")

library(dplyr)

library(ggplot2)

library(tidyr)

library(lmtest)

library(MASS

library(readxl)

# TOTAL NUMBER OF  FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

file_path = ("C:/Users/pc1/Desktop/ASSIGNMENTS/SATABDA_DISSERTATION_1 (Autosaved)2.xlsx")

data = read_excel(file_path, sheet = "Sheet5",range = "R7C2:R22C7" )

print(data)

model <- glm(TOTAL_NUMBER_OF_MIGRANTS ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data, family = poisson())

summary(model)

fitted(model)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF  FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSU

residuals1 <- residuals(model)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals1)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS",

      x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

library(lmtest)

fitted_values <- fitted(model)

residuals1 <- residuals(model)

lm_squared_resid <- lm(residuals1^2 ~ fitted_values)

bptest(lm_squared_resid)


# FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS


data1 = read_excel(file_path, sheet = "Sheet5",range = "R7C9:R21C14" )

print(data1)

model1 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data1, family = poisson())

summary(model1)

fitted(model1)


#GENERATING QQ PLOT FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS


residuals2 <- residuals(model1)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals2)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

library(lmtest)

fitted_values <- fitted(model1)

residuals2 <- residuals(model1)

lm_squared_resid <- lm(residuals2^2 ~ fitted_values)

bptest(lm_squared_resid)


# # FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS


data2 = read_excel(file_path, sheet = "Sheet5",range = "R7C16:R21C21" )

print(data2)

library(MASS)

model2 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data2, family = poisson())

summary(model2)

fitted(model2)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

residuals3 <- residuals(model2)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals3)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

```
library(lmtest)

fitted_values <- fitted(model2)

residuals3 <- residuals(model2)

lm_squared_resid <- lm(residuals3^2 ~ fitted_values)

bptest(lm_squared_resid)
```

FOR TOTAL NUMBER OF TOTAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

```
data3 = read_excel(file_path, sheet = "Sheet5",range = "R30C2:R44C7" )

print(data3)

model3 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data3, family = poisson())

summary(model3)

fitted(model3)
```

#GENERATING QQ PLOT FOR TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

```
residuals4 <- residuals(model3)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals4)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF  FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

library(lmtest)

fitted_values <- fitted(model3)

residuals4 <- residuals(model3)

lm_squared_resid <- lm(residuals4^2 ~ fitted_values)

bptest(lm_squared_resid)


# FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS


data4 = read_excel(file_path, sheet = "Sheet5",range = "R30C9:R44C14" )

print(data4)

model4 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data4, family = poisson())

summary(model4)

fitted(model4)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

residuals5 <- residuals(model4)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals5)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

library(lmtest)

fitted_values <- fitted(model4)

residuals5 <- residuals(model4)

lm_squared_resid <- lm(residuals5^2 ~ fitted_values)

bptest(lm_squared_resid)


# FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS


data5 = read_excel(file_path, sheet = "Sheet5",range = "R30C16:R44C21" )

print(data5)

model5 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data5, family = poisson())

summary(model5)

fitted(model5)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

residuals6 <- residuals(model5)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals6)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS

```
library(lmtest)

fitted_values <- fitted(model5)

residuals6 <- residuals6(model5)

lm_squared_resid <- lm(residuals^2 ~ fitted_values)

bptest(lm_squared_resid)
```

FOR TOTAL NUMBER OF TOTAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

```
data6 = read_excel(file_path, sheet = "Sheet5",range = "R51C2:R64C7" )

print(data6)

model6 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data6, family = poisson())

summary(model6)

fitted(model6)
```

#GENERATING QQ PLOT FOR TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

```
residuals7 <- residuals(model6)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals7)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF  FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

library(lmtest)

fitted_values <- fitted(model6)

residuals7 <- residuals(model6)

lm_squared_resid <- lm(residuals7^2 ~ fitted_values)

bptest(lm_squared_resid)


# FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS


data7 = read_excel(file_path, sheet = "Sheet5",range = "R51C9:R64C14" )

print(data7)

model7 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data7, family = poisson())

summary(model7)

fitted(model7)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

residuals8 <- residuals(model7)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals8)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

```
# Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

library(lmtest)

fitted_values <- fitted(model7)

residuals8 <- residuals(model7)

lm_squared_resid <- lm(residuals8^2 ~ fitted_values)

bptest(lm_squared_resid)


# FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS


data8 = read_excel(file_path, sheet = "Sheet5",range = "R51C16:R64C21" )

print(data8)

model8 <- glm(`TOTAL NUMBER OF MIGRANTS` ~ `NUMBER OF MIGRANTS DUE TO WORK EMPLOYMENT` + `NUMBER OF MIGRANTS DUE TO BUSINESS` + `NUMBER OF MIGRANTS DUE TO EDUCATION` + `NUMBER OF MIGRANTS DUE TO MARRIAGE`, data = data8, family = poisson())

summary(model8)

fitted(model8)

#GENERATING QQ PLOT FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

residuals9 <- residuals(model8)

library(ggplot2)

ggplot(data = NULL, aes(sample = residuals9)) +

  stat_qq(distribution = qnorm) +  # Q-Q plot against normal distribution

  geom_abline(slope = 1, intercept = 0, color = "red") +  # Add line of equality

  labs(title = "QQ PLOT OF POISSON REGRESSION RESIDUALS FOR TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS",

       x = "Theoretical Quantiles", y = "Sample Quantiles")
```

Perform Breusch-Pagan test for heteroscedasticity FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS

```
library(lmtest)

fitted_values <- fitted(model8)

residuals9 <- residuals(model8)

lm_squared_resid <- lm(residuals9^2 ~ fitted_values)

bptest(lm_squared_resid)
```

TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS

```
# Create a data frame with the provided data

data <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

           "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

           "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

           "Tamil Nadu", "Bihar"),

  Total_Migrants = c(163275, 1854139, 443445, 2246796, 6067612, 19871,

                  669621, 1451881, 1981543, 886352, 1516952, 1195536,

                  1061732, 3602243),

  Work_Employment = c(5422, 67966, 39581, 64358, 244890, 627, 40895,

                   65103, 81690, 26541, 77216, 68378, 76790, 148008),

  Business = c(729, 11277, 1693, 15854, 26403, 69, 2711, 7172, 7748,

            7152, 7653, 6804, 8202, 17403),

  Education = c(2857, 9380, 3151, 8875, 35140, 1336, 4193, 8665, 9840,

             7403, 8014, 11485, 9591, 22696),

  Marriage = c(58578, 980892, 255876, 1351148, 3003996, 8025, 386746,

            914724, 1296694, 451521, 809682, 610420, 471628, 1775188)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data[, c("State","Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF  FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS")


#TOTAL NUMBER OF RURAL FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS



# Create a data frame with the provided data

data1 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

        "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

        "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

        "Tamil Nadu", "Bihar"),

  Total_Migrants = c(62591,717125,243928,969240,1593933,11794,324730,694380,1060812,33926,690278,428635,305232,1168141),

  Work_Employment = c(1457,21770,22040,20598,60897,228,15444,15021,36535,5555,25828,16896,19264,49614),

  Business = c(141,2057,566,3168,3654,30,617,1846,2228,777,2163,1503,1540,3895),

  Education = c(578,1752,594,1408,4602,372,879,1155,2081,1039,1968,2430,1620,2527),

  Marriage = c(28561,499781,165815,790302,1167453,5948,249881,572596,833763,165061,459538,305326,174882,843359)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data1[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS")

print(data1)


#TOTAL NUMBER OF URBAN FEMALE  MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS


# Create a data frame with the provided data

data2 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

            "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

            "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

            "Tamil Nadu", "Bihar"),

  Total_Migrants =
c(100684,1137014,199517,1277556,4473679,8077,344891,757501,920731,652426,826674,766901,756500,2434102),

  Work_Employment = c(3965,46196,17541,43760,183993,399,25451,50082,45155,20986,51388,51482,57526,98394),

  Business = c(588,9220,1127,12686,22749,39,2094,5326,5520,6375,5490,5301,6662,13508),

  Education = c(2279,7628,2557,7467,30538,964,3314,7510,7759,6364,6046,9055,7971,20169),

  Marriage =
c(30017,481111,90061,560846,1836543,2077,136865,342128,462931,286460,350144,305094,296746,931829)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data2[, c("State","Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2011 CENSUS")

print(data2)


# TOTAL NUMBER OF TOTAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS


# Create a data frame with the provided data

data3 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

          "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

          "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

          "Tamil Nadu", "Bihar"),

  Total_Migrants = c(142936,1281334
            ,521464,1515575,4324799,11822,465593,936167,1322193,753890,1073354,897875,848841,2211222),

  Work_Employment = c(3912,41582,58999,37823,136114,319,23426,36156,50456,17030,62378,47811,67654,80534),

  Business = c(514,8169,1061,8630,10524,48,1063,3252,3754,3125,2482,3961,5098,7612),

  Education = c(2501,4992,3232,4332,17617,598,2012,5455,4658,4913,4826,6525,7315,9954),

  Marriage =
c(74556,690012,256378,925526,2222702,4707,273715,556175,850380,414533,569086,472357,382995,1074114)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data3[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS")


# TOTAL NUMBER OF  RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS


# Create a data frame with the provided data

data4 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

            "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

            "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

            "Tamil Nadu", "Bihar"),

  Total_Migrants = 
c(72984,546815,332445,717956,1399887,6714,254770,441349,753560,222753,486832,373939,288783,809825),

  Work_Employment = c(1448,17044,41019,16568,43619,125,11695,10310,26757,4484,27263,18776,27826,31365),

  Business = c(213,3589,535,2736,2607,26,349,959,1809,638,765,1298,1644,2449),

  Education = c(502,1054,395,784,2805,161,430,764,1041,630,887,968,1178,1611),

  Marriage = 
c(53901,361209,184764,556865,1002454,3586,185962,339289,561363,160062,312139,249302,150917,537158)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data4[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS")


# TOTAL NUMBER OF  URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS


# Create a data frame with the provided data

data5 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Chattishgarh", "Rajhasthan",

            "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

            "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

            "Tamil Nadu", "Bihar"),

  Total_Migrants =
c(69952,734519,189019,797619,2924912,5108,210823,494818,568633,531137,586522,523936,560058,1401397),

  Work_Employment = c(2464,24538,17980,21255,92495,194,11731,25846,23699,12546,35115,29035,39828,49169),

  Business = c(301,4580,526,5894,7917,22,714,2293,1945,2487,1717,2663,3454,5163),

  Education = c(1999,3938,2837,3548,14812,437,1582,4691,3617,4283,3939,5557,6137,8343),

  Marriage =
c(20655,328803,71614,368661,1220248,1121,87753,216886,289017,254471,256947,223055,232078,536956)

)
```

```
# Selecting numeric columns for clustering

clustering_data <- data5[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 2001 CENSUS")


# TOTAL NUMBER OF FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS


# Create a data frame with the provided data

data6 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Rajhasthan",

            "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

            "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

            "Tamil Nadu", "Bihar"),

  Total_Migrants =
c(77325,1065226,1181175,2781930,20817,338067,705046,1039074,515802,855679,709939,749280,1427704),

  Work_Employment = c(3300,49034,31436,87160,696,17611,20358,28293,11740,31553,26574,61959,64494),

  Business = c(1033,16194,17529,28159,346,2678,6024,12062,8887,20146,9378,7501,13117),

  Education = c(1717,9536,8065,28571,955,2995,7498,6819,7388,9615,10327,8570,13957),

  Marriage = c(34132,595689,807462,1593334,10914,218966,484692,749383,290862,494686,401665,349056,896081)

)
```

```
print(data6)

# Selecting numeric columns for clustering

clustering_data <- data6[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF  FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS")


# TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS


# Create a data frame with the provided data

data7 <- data.frame(

  State = c("Jammu Kashmir", "Maharashtra", "Rajhasthan",

        "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",

        "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",

        "Tamil Nadu", "Bihar"),

  Total_Migrants = c(28858,462152,597815,909362,13449,204982,370005,644808,126514,412644,306013,278096,730304),

 Work_Employment = c(1175,19636,12957,23563,304,8084,5018,13322,1856,9696,7900,27892,30637),

  Business = c(463,6070,6250,9551,183,1186,2119,6965,1975,11184,3243,2116,6028),

  Education = c(409,2619,2373,5290,536,951,1894,2553,832,2857,2999,1894,3412),

  Marriage = c(16935,309684,491978,693499,8319,155936,308845,516302,91937,286086,212493,147350,551785)

)
```

```r
# Selecting numeric columns for clustering
clustering_data <- data7[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method
hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram
plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF RURAL FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS")


# TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS


# Create a data frame with the provided data
data8 <- data.frame(
  State = c("Jammu Kashmir", "Maharashtra", "Rajhasthan",
            "Uttar Pradesh", "Arunachal Pradesh", "Odisha", "West Bengal",
            "Madhya Pradesh", "Gujarat", "Karnataka", "Andhra Pradesh",
            "Tamil Nadu", "Bihar"),
  Total_Migrants = c(48467,603074,583360,1872568,7368,133085,335041,394266,389288,443035,403926,471184,697400),
  Work_Employment = c(2125,29398,18479,63597,392,9527,15340,14971,9884,21857,18674,34067,33857),
  Business = c(570,10124,11279,18608,163,1492,3905,5097,6912,8962,6135,5385,7089),
  Education = c(1308,6917,5692,23281,419,2044,5604,4266,6556,6758,7328,6676,10545),
  Marriage = c(17197,286005,315484,899835,2595,63030,175847,233081,198925,208600,189172,201706,344296)
)
```

```
# Selecting numeric columns for clustering

clustering_data <- data8[, c("Total_Migrants", "Work_Employment", "Business", "Education", "Marriage")]


# Hierarchical clustering using Ward's method

hc <- hclust(dist(clustering_data), method = "ward.D")


# Plotting the dendrogram

plot(hc, hang = -1, main = "DENDROGRAM OF STATES FOR TOTAL NUMBER OF URBAN FEMALE MIGRANTS IN INDIA FOR DIFFERENT REASONS IN 1991 CENSUS")
```

ACKNOWLEDGEMENT

I am extremely grateful to my supervisor, Dr. Anindita Ghosal, Head of the Department of Statistics, for guiding and supervising me throughout the project. She continuously encouraged me to complete the project work sincerely and meticulously. I have tried to follow her advice and guidance to my best to make this project perfect and flawless. In spite of that, if any inadequacy crept into my project, I will try to learn further to make the inadequacies complete and perfect.

Additionally, I wish to express my appreciation to all faculty members of the Department of Statistics for their valuable insights and support during my research. Their constructive feedback and suggestions were instrumental in shaping this project.