Demographic Analysis Using R

SPDI D4D Workshop

Zahid Asghar, School of Economics, QAU

Table of contents

Workshop Overview

A comprehensive hands-on workshop covering modern demographic data analysis workflows using R, focusing on PDHS data and other survey datasets. Learn to import data (including web API, URLs), analyze, visualize, and report demographic indicators using cutting-edge tools and AI-assisted workflows integrated from Day 1.

Target Audience

Researchers, analysts, and professionals working with data in social sciences, health, economics, and policy.

Workshop Details

- **Organiser:** SDPI , Netherland and UNFPA
- **Trainer:** Zahid Asghar, School of Economics, QAU
- **Dates:** 17 - 21 November 2025
- **Venue:** Luxus Hotel, Lahore, Pakistan
- **Mode:** In-person

Daily Schedule

Day	Time
Monday-Thursday	10:00 AM - 4:30 PM
Friday	9:30 AM - 1:00 PM

Pre-Workshop Requirements

Software (All Free)

Installation Steps

1. Install R (≥ 4.3) from CRAN website
2. Install Positron IDE (latest version) or RStudio as alternative
3. Essential R packages (installation script will be provided)

Workshop Data Pack

Sample datasets provided in multiple formats including:

- CSV
- Excel (xlsx, xls)
- Stata, SPSS, SAS, JSON (PDHS, PSLM, MICS)
- API examples for retrieving demographic indicators

Daily Breakdown

Day 1: Foundations, Data Import & AI-Assisted Workflows

Session 1: R and Positron IDE Fundamentals

- Introduction to R programming: variables, data types, functions
- Positron IDE tour: console, editor, viewer, and project management
- Installing and loading packages (tidyverse ecosystem)
- Working with projects and relative file paths

Hands-on: Create your first R project and perform basic calculations

Session 2: Importing Demographic Data from Multiple Sources

- Reading CSV and Excel files (census extracts, administrative data)
- Importing PDHS data from Stata (.dta) format with proper label handling
- Working with SPSS datasets (.sav) from other surveys
- Accessing demographic APIs (World Bank, UN Population Division)
- Understanding data dictionaries and codebooks

Hands-on: Import PDHS household and individual datasets, explore structure

Session 3: AI-Powered Data Wrangling

- Introduction to AI assistants for data analysis (Positron's built-in AI)
- Using natural language prompts to generate data cleaning code
- AI-assisted variable recoding and transformation
- Selecting and filtering data (provinces, age groups, urban/rural)
- Creating new variables (age groups, composite indicators, ratios)
- Grouping and summarizing by demographic characteristics
- Merging household and individual PDHS files
- Reshaping data: wide to long format for time-series analysis

Hands-on: Use AI to wrangle data with guided prompts

Key Skills: Project setup, importing survey data, AI-assisted data transformation, merging datasets

Day 2: Exploratory Data Analysis & Visualization

Session 1: Creating Professional Demographic Tables with AI Support

- Descriptive statistics tables with gt (population pyramids data)
- Survey-weighted tables with gtsummary
- Cross-tabulations by region, residence, and socioeconomic status
- NHANES Data example for health indicators
- Formatting tables for publications and reports

Hands-on: Use AI prompts to create comparison tables of maternal/other health indicators across provinces

Session 2: Static Visualizations for Demographics

- Population pyramids showing age-sex distribution
- Trend lines for demographic indicators over time
- Bar and column charts comparing provinces and districts
- Scatter plots showing relationships between indicators
- Faceted plots for multi-group comparisons

Hands-on: Create publication-ready population charts with AI assistance

Session 3: Interactive and Animated Visualizations

- Animated population pyramids showing demographic transition
- Interactive plotly charts with hover information
- Time-series visualizations with zoom and pan controls
- Creating animated transitions for demographic indicators over years

Hands-on: Visualize PDHS data, work with shape file for Pakistan using AI guidance

Key Skills: Survey-weighted tables, demographic visualizations, interactive graphics, AI-assisted coding for outputs

Day 3: Statistical Modeling & Spatial Analysis

Session 1: Demographic Statistical Modeling

- Linear regression for continuous outcomes (fertility rates, age at marriage)
- Logistic regression for binary outcomes (contraceptive use, facility delivery)
- Survey-weighted regression models with PDHS data
- Interpreting odds ratios and confidence intervals
- Model diagnostics and goodness-of-fit
- AI assistance for identifying appropriate statistical methods

Hands-on: Model determinants of child vaccination using PDHS with proper survey weights

Session 2: Spatial Data and Mapping Basics

- AI-assisted spatial data workflows
- Loading Pakistan shapefiles (districts, provinces, divisions)
- Understanding coordinate reference systems (CRS)
- Joining demographic data with geographic boundaries
- Creating choropleth maps showing district-level indicators
- Color schemes and classification methods for demographic maps

Hands-on: Create district-level maps of literacy rates and infant mortality with AI support

Session 3: Interactive Spatial Visualization

- AI-powered mapping code generation
- Building interactive maps with leaflet
- Adding popups with detailed demographic information
- Multi-layer maps showing different indicators
- Quick mapping with mapview for exploratory analysis

Hands-on: Build interactive map with toggleable layers for health and education indicators

Key Skills: AI-guided modeling, survey-weighted analysis, spatial data handling, interactive geographic visualization

Day 4: Reporting & Dashboards

Session 1: Professional Reports with Quarto and AI

- AI-assisted report writing and formatting
- Creating Quarto documents mixing narrative and code
- Embedding tables, figures, and maps in reports
- Cross-referencing figures and tables
- Output formats: HTML, PDF, Word documents
- Code folding and hiding for clean reports
- Using AI to draft report narratives and interpretations

Hands-on: Convert Day 3 analysis into formatted demographic report with AI assistance

Session 2: Parameterized Reports for Automation

- AI-guided automation workflows
- Setting up report parameters (province, year, indicator)
- Automating report generation for multiple regions
- Creating district or provincial profile reports
- Batch rendering for stakeholder distribution

Hands-on: Create parameterized provincial demographic profiles with PDHS data

Session 3: AI-Powered Interactive Dashboards

- Advanced AI-assisted dashboard development
- Building flexdashboard for demographic indicators
- Adding filters and interactive controls
- Dashboard layout and design principles
- Using AI to troubleshoot and enhance dashboard functionality

Hands-on: Build interactive PDHS dashboard with multiple indicator panels using AI workflows

Key Skills: Reproducible reports, automated outputs, AI-integrated workflows, dashboard development

Day 5: Advanced AI-Powered Analysis & Closing (Half-day)

Session 1: Mastering AI-Powered Demographic Analysis

Learn advanced techniques for leveraging AI assistants for complex demographic analysis:

AI-Assisted Tasks	Applications
Data cleaning and recoding	Harmonizing PDHS variables across waves
Exploratory analysis patterns	Identifying demographic trends and anomalies
Model selection guidance	Choosing appropriate methods for demographic analysis
Results interpretation	Translating findings for policy briefs
Code debugging and optimization	Fixing errors and improving efficiency
Documentation generation	Creating codebooks and methodology notes

Best Practices for AI-Assisted Analysis

- Crafting effective prompts for demographic tasks
- Verifying and validating AI-generated code
- Combining AI assistance with domain expertise
- Ethical considerations in AI-powered research

Session 2: Participant Project Presentations

Each participant presents their capstone project (5 minutes each):

- Research question and demographic focus
- Data sources and analytical approach
- How AI tools were used in the analysis
- Key findings with live visualization demo
- Policy implications and recommendations

Session 3: Wrap-up, Evaluation & Certificates

Learning Outcomes

By the end of this workshop, participants will be able to:

Data Management	Analysis & Modeling	Communication	AI-Powered Workflows
Import PDHS and survey data	Calculate demographic indicators	Design effective visualizations	Use AI for code generation

Data Management	Analysis & Modeling	Communication	AI-Powered Workflows
Clean and recode variables	Build weighted regression models	Generate automated reports	Craft effective analysis prompts
Merge multiple datasets	Create spatial maps	Build interactive dashboards	Validate AI-generated outputs
Handle survey weights	Interpret statistical results	Present findings clearly	Integrate AI throughout analysis pipeline

Capstone Project

Participants develop an individual demographic analysis project throughout the week.

Required Components

✓ PDHS or survey data import
✓ Data cleaning and preparation
✓ Survey-weighted analysis
✓ Professional tables and charts
✓ Statistical model OR spatial map
✓ Quarto report with findings
✓ Documentation of AI tool usage

Suggested Topics

- Fertility trends by province
- Child health outcomes analysis
- Maternal care determinants
- Education gender disparities
- Spatial inequality patterns
- Your own research question!

Materials Provided

- Workshop data pack with all datasets
- R code templates and exercises
- Pakistan shapefiles for mapping
- Reference materials and cheatsheets
- AI prompt library for demographic analysis
- Software installation guide
- Certificate of completion

ℹ Workshop Philosophy

Workshop designed for practical, applied learning using modern tools, AI-assisted workflows, and real-world Pakistan demographic data

```
Note: All participants will learn how to use AI Assistants for Data Analysis
directly within their workflow—no copying and pasting from external AI tools
required!
```