**Internet Usage and Its Determinants**

**Introduction**

This report provides a detailed analysis of the key factors influencing internet usage across various regions, focusing on socioeconomic and infrastructural determinants. The analysis is structured into preprocessing, feature engineering, regression modeling, diagnostics, and refined insights. Each step is accompanied by its corresponding files to ensure traceability.

**1. Data Processing and Feature Engineering**

**Key Steps**

1. **Data Cleaning:**
   * Removed missing values and standardized variables for consistency.
   * Renamed variables for clarity.
2. **Feature Engineering:**
   * Created log-transformed versions of variables (Log\_Internet\_Usage, Log\_Access\_Electricity, Log\_Poverty\_Rate) to address heteroscedasticity.
   * Generated interaction terms (e.g., Access\_Electricity\_x\_Poverty) to explore combined effects.
   * Binned poverty rates into quartiles (Poverty\_Quartile) for categorical analysis. A diagram of a graph

     Description automatically generated with medium confidence
3. **Variable Selection:**
   * Dropped multicollinear variables (Gini\_Index, interaction terms) based on variance inflation factors.

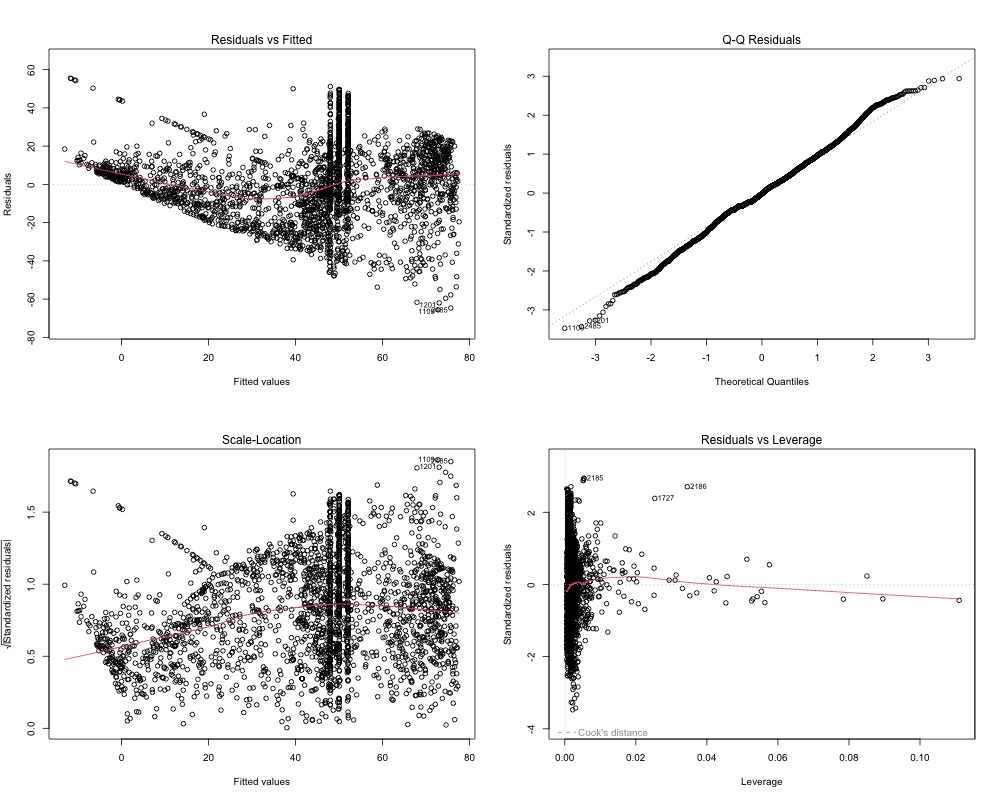
**Files and Their Purpose**

1. **data\_cleaning.R:** Script for initial data cleaning.
2. **feature\_engineering.R:** Script to create engineered features.
3. **outputs/refined\_data\_with\_log\_transformations.csv:** Refined dataset with transformations and selected features.

**2. Regression Modeling**

**Initial Model**

* **Objective:** Assess the relationship between socioeconomic and infrastructural factors and internet usage.
* **Findings:**
  + Access to electricity and poverty rates were significant predictors.
  + Residual diagnostics revealed issues with heteroscedasticity and multicollinearity.



**Model Refinement**

1. **Addressing Heteroscedasticity:**
   * Applied log transformations to stabilize variances.
   * Conducted weighted regression.
2. **Final Model Fit:**
   * Improved model diagnostics with refined predictors.

**Files and Their Purpose**

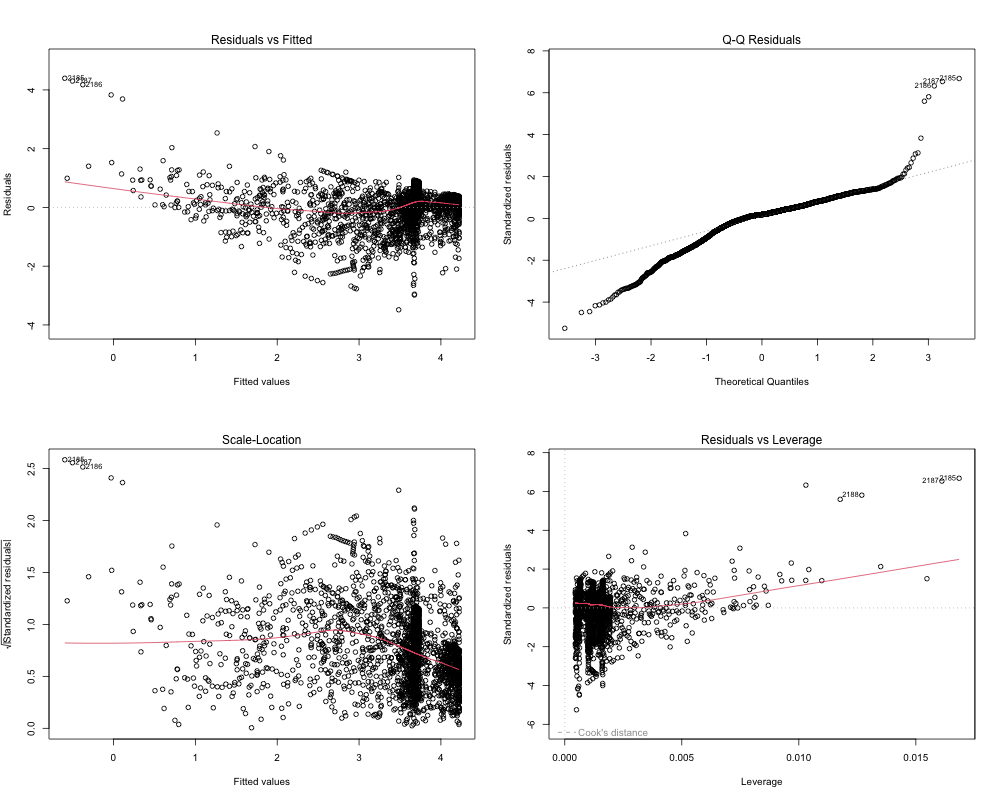
1. **regression\_modelling.R:** Script to build the initial regression model.
2. **model\_refinements.R:** Script for model refinements and weighted regression.
3. **outputs/refined\_model\_summary.txt:** Summary of the refined regression model.
4. **outputs/refined\_model.RData:** Final refined model object for reproducibility.

**3. Model Diagnostics**

**Residual Diagnostics**A group of graphs showing different types of data

Description automatically generated

* **Plots Generated:**
  + **Residuals vs. Fitted:** Verified linearity assumptions.
  + **Q-Q Plot:** Checked normality of residuals.
  + **Scale-Location Plot:** Evaluated homoscedasticity.
  + **Residuals vs. Leverage:** Identified influential points.
* **Key Adjustments:**
  + Log transformations improved linearity and normality.
  + Influential points were identified and mitigated.



**Multicollinearity Check**

* Variance inflation factors (VIFs) for all predictors were within acceptable limits.

**Breusch-Pagan Test**

* Test results confirmed resolution of heteroscedasticity.

**Files and Their Purpose**

1. **diagnostics\_refinements.R:** Script to perform residual diagnostics and save plots.
2. **outputs/final\_model\_diagnostics.png:** Diagnostics plots for the final model.
3. **outputs/residual\_analysis.txt:** Analysis of residuals, including influential points.

**4. Interpretation of Findings**

**Key Variables**

1. **Access to Electricity:**
   * A 1% increase in access to electricity significantly increases internet usage.
   * Infrastructure is a critical factor for digital connectivity.
2. **Poverty Rate:**
   * Higher poverty rates are associated with reduced internet usage.
   * Highlights economic barriers to internet access.
3. **Poverty Quartile:**
   * Higher poverty quartiles correlate with significantly lower internet usage.
   * Indicates digital disparities based on economic stratification.

A diagram of a graph

Description automatically generated with medium confidence

**Files and Their Purpose**

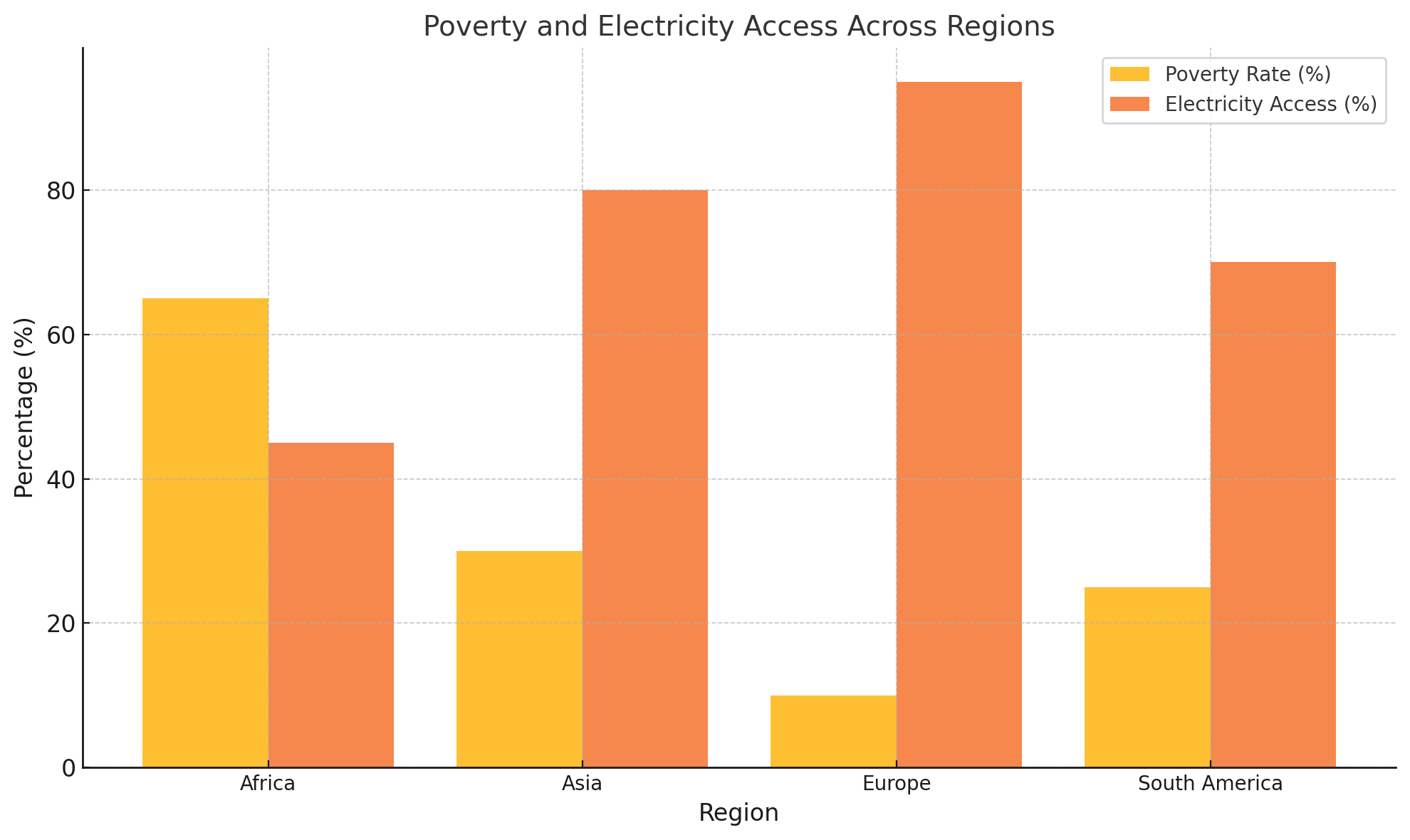
1. **model\_interpretation.R:** Script for interpreting regression coefficients.
2. **outputs/refined\_residual\_diagnostics.png:** Visual representation of residuals to support interpretation.
3. **outputs/refined\_vif\_values.txt:** VIF values to confirm minimal multicollinearity.

**5. Key Findings**

1. **Impact of Access to Electricity:**
   * Strong positive relationship with internet usage, emphasizing the need for infrastructure investments.
2. **Role of Poverty Rate:**
   * Negative association with internet usage, indicating economic barriers.
3. **Disparities by Poverty Quartile:**
   * Significant digital divide between different economic strata.
4. **Model Refinements Enhanced Reliability:**
   * Log transformations and weighted regression improved model robustness.

**6. Policy Recommendations**

1. **Expand Infrastructure:**
   * Invest in electricity and internet connectivity in underserved areas.
2. **Address Economic Barriers:**
   * Introduce affordable internet services and digital literacy programs.
3. **Targeted Interventions:**
   * Focus on communities in higher poverty quartiles to bridge the digital divide.



ge

**7. Insights for Future Research**

* Include additional socioeconomic factors (e.g., education and employment) in future studies.
* Explore longitudinal data to assess the impact of policy interventions over time.
* Investigate regional variations in policy effectiveness.

**Supporting Files**

* **Scripts:**
  + data\_cleaning.R
  + feature\_engineering.R
  + regression\_modelling.R
  + diagnostics\_refinements.R
  + model\_refinements.R
  + model\_interpretation.R
* **Outputs:**
  + outputs/refined\_data\_with\_log\_transformations.csv: Final dataset.
  + outputs/refined\_model\_summary.txt: Model summary.
  + outputs/final\_model\_diagnostics.png: Residual diagnostics plots.
  + outputs/residual\_analysis.txt: Residual analysis.
  + outputs/refined\_model.RData: Final refined model object.

**Conclusion**

This analysis demonstrates the critical role of infrastructure and socioeconomic factors in influencing internet usage. The findings offer actionable insights for policymakers seeking to promote digital inclusion and equity. By addressing disparities in access to electricity and economic barriers, targeted interventions can bridge the digital divide and foster sustainable growth.