README for replication_code.Rmd

(Upzoning and Residential Transaction Price in Nashville)

Project Overview

The ** replication_code.Rmd** is an R Markdown file designed to produce a PDF document. It analyzes the impact of upzoning of Nashville's 'Downtown Code' (DTC) as compared to various control boundaries (1k, 3k, 5k, and 10k). This is achieved using both Difference-in-Differences (DID) and Matching models.

The manuscript relies on data from several Excel files, statistical analysis, and graphical presentation to present insights on treatment effects on housing prices.

File Structure

R Markdown Structure
- **Header**: YAML metadata for title, output format (PDF), and the current date.
- **Setup Chunk**: Loads required R libraries and sets global options.
- **Data Preparation and Visualization**: Imports data, processes it, and visualizes it using `ggplot2`.
- **DID Regression Models**: Runs multiple linear models to estimate treatment effects.
- **Matching Methods**: Applies matching techniques to control for covariates.

Data Sources
Data files are stored in a subfolder named `data/`.

1. **DTC_DID1k.xlsx**: Data for 1km distance.
2. **DTC_DID3k.xlsx**: Data for 3km distance.
3. **DTC_DID5k.xlsx**: Data for 5km distance.
4. **DTC_DID10k.xlsx**: Data for 10km distance.
5. **DTC_DIDFULLCONTROL.xlsx**: Full dataset for matching methods.

Forster-Benson

Variable Descriptions

Below is a description of the key variables used in the analysis:

- **SalePrice**: The transaction price of the property was sold.
- **CoreBin**: A binary variable indicating if the property was treated (part of the 2010 upzoning 'Downtown Code.'
- **TimeBin**: A binary variable representing if the sale was before or after the treatment (February 2010).
- **FinishArea**: The total finished square footage of the property.
- **Acres**: The size of the property lot, measured in acres.
- **PropZip**: The zip code of the property, identifying its general location within a larger geographic area.
- **Tract**: The census tract of the property was sold.
- **Quarter**: The quarter of the year (Q1, Q2, Q3, Q4) in which the property sale took place.
- **Year**: The year in which the property sale occurred.

Required Libraries

- `knitr`: For R Markdown knitting.
- `stargazer`: For generating model output tables.
- `AER`: For regression analysis.
- `ggplot2`: For data visualization.
- `plm`: For panel data analysis.
- `dplyr`: For data manipulation.
- `tidyverse`: Collection of data analysis packages.
- `qte`: For quantile treatment effects.
- `readxl`: For reading Excel files.
- `randomForest`: For random forest models.
- `gbm`: For gradient boosting models.
- `MatchIt`: For matching methods.
- `plotly`: For interactive visualizations.
- `quantreg`: For quantile regression.

Usage Instructions

1. Load R Markdown File: Open the `2024_Manuscript.Rmd` file in RStudio.
2. Data Setup: Ensure that the Excel files are located in the `data/` folder as specified.
3. Run the File: Click the "Knit" button to generate the PDF report.

Troubleshooting

- Error: File Not Found: Ensure that the `data/` folder contains all required Excel files.
- Missing Packages: Run `install.packages("package_name")` for any missing R packages.
- Model Errors: Verify that the columns in the Excel files match those referenced in the code.

Notes & Assumptions

1. Data Integrity: Ensure that Excel files in the `data/` folder are formatted correctly and column names match the script.
2. Year Range: The analysis covers data from 2000 to 2023.
3. Outlier Removal: Sale prices above $2.5M or below $1 are excluded from the analysis.

Contact

For questions or issues, please contact the project maintainer (nicholas.o.forster-benson@vanderbilt.edu).