Summary of Chapter 3 of "Building Reproducible Analytical Pipelines with R" Book

In chapter 3 of the "Building Reproducible Analytical Pipelines with R" book, the main focus was on a project that would keep going throughout the remainder of the text. The section concentrated on the development of a reproducible analytical infrastructure.

The first step was to download data on house prices in Luxembourg. The primary objective was to do a study an analysis rather than focusing on reproducibility. It was explained how Luxembourg is divided into Cantons and Communes, which are comparable to States and Counties in the United States, respectively. It was highlighted that Luxembourg is approximately the same size as the U.S. state of Rhode Island, and that its population dynamics are unique. The daily entry of workers from neighbors such as Belgium, France, and Germany affects the housing and transportation systems significantly.

The House Price Index (HPI) was introduced as a measure of housing market inflation. In addition, author also used a graph to show how home prices changed from 2010 to 2021, showing that supply and demand caused big differences in prices throughout regions.

The objective of the project was to extract data from an Excel file, convert nominal prices to actual prices, generate tables, and display the data. The difficulties of converting Excel data into a "machine readable format" were discussed. In addition, differences in commune names and the merger of several communes into new ones after 2010 required validation of information and normalization. This required comparing commune listings and merging past and current commune information to ensure data accuracy.

The author indicated at the end of the chapter that this project is not complete. While the current project achieves its objectives, it has problems due to the lack of clear documentation, appropriate package management, and structured code that facilitates reusability and reproducibility. It highlighted the importance of documenting procedures, managing dependencies, providing reproducibility, and maintaining the computational environment.

<div align="right">Ömer Faruk Yahşi</div>