◆ **Summary of the Code**

1. **Libraries Loaded**

   ○ `readr`, `dplyr`: For data handling.

   ○ `neuralnet`: For ANN (not used in the pie chart part, but included).

   ○ `xgboost`: For predictive modeling.

   ○ `ggplot2`: For visualization.

2. **Synthetic Dataset Creation**

   ○ A fake dataset is generated with pollutants (**PM10, NO2, SO2, CO, O3**) and **PM2.5** (target).

   ○ AQI is also simulated based on pollutant levels.

   ○ This allows the code to run even **without a real dataset**.

3. **Train-Test Split**

   ○ 70% of the dataset is used for training, 30% for testing.

   ○ `PM2.5` is set as the **target variable**, and other pollutants are features.

4. **XGBoost Model Training**

   ○ An **XGBoost regression model** is trained to predict PM2.5 levels.

   ○ The model learns which pollutants are most important in predicting PM2.5.

5. **Feature Importance Calculation**

   ○ `xgb.importance()` extracts the contribution (gain) of each pollutant.

   ○ The contribution values are converted into **percentages**.

6. **Pie Chart Visualization**

○ A **pie chart** is drawn using `ggplot2`.

○ Each slice = contribution of a pollutant to PM2.5 prediction.

○ Legend shows pollutant names.

○ Title: *"Pollutant Contribution to PM2.5 Prediction (XGBoost)"*