Todo:
- Feature documentation: what the raw data contains, what each feature measures
- Detail what 'digit in noise' is, how it is measured, and explain the intuition of using it to predict audiogram measurements.

Work:
- Cleaned up code chunks for better readability and efficiency.
- Renamed variables to more clearly reflect their roles and functions.
- Provided comprehensive documentation for each file to guide users through the dataset and analysis.
- Added in-code notations to clarify the purpose of variables and explain complex sections of code.
- Documented the functions used to facilitate understanding and reuse.
- Included comments to capture the thought process behind analytical decisions.
- Made observations at each step to track progress and findings.

Next Steps:
We don't observe obvious improvement in model performance when additional predictors are added. We need to further validate whether the model is learning the actual relationship.

Idea:
I observed that with the by hand cca procedure, some predictions are very off, while others are decently close to the actual data.
Since the audiogram measurements are highly correlated with each other, can we somehow determine which predictions are more accurate, and smooth the other features in the audiogram based on their relationship with that most accurate column's prediction?