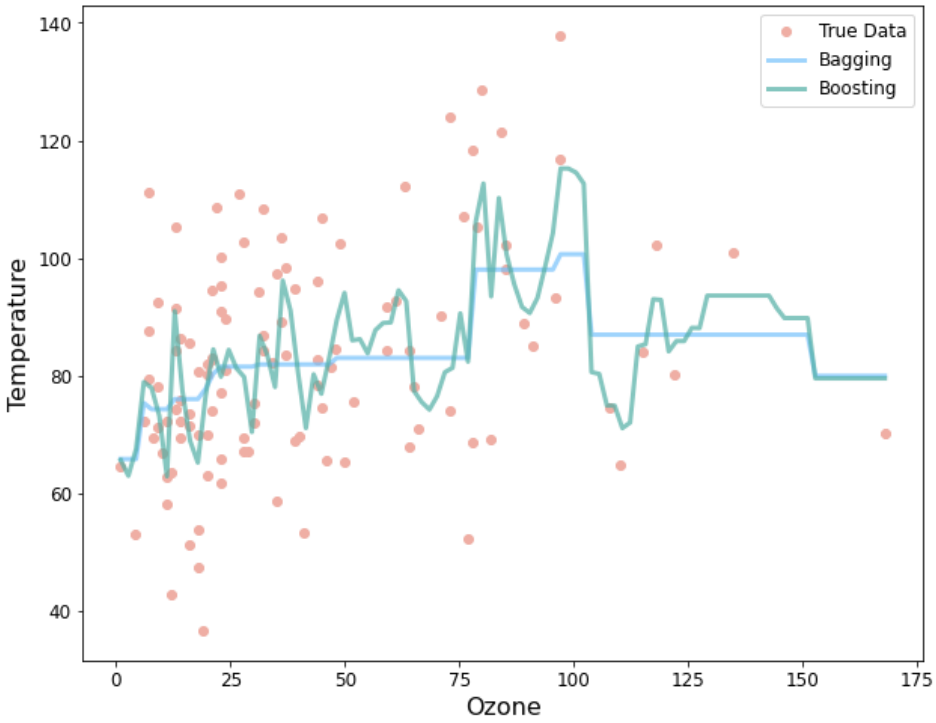
Exercise 5.1: Regression with Boosting

□□□ Regression with Boosting

The goal of this exercise is to understand *Gradient Boosting Regression.*



Instructions:

Part A:

- Read the dataset *airquality.csv* as a pandas dataframe.
- Take a quick look at the dataset.
- Assign the predictor and response variables appropriately as mentioned in the scaffold.
- Fit a single decision tree stump and predict on the entire data.
- Calculate the residuals and fit another tree on the residuals.
- Take a combination of the trees and fit on the model.
- For each of these model use the helper code provided to plot the model prediction and data.

Part B: Compare to bagging

- Split the data into train and test splits.
- Specify the number of bootstraps for bagging to be 30 and a maximum depth of 3.
- Define a Gradient Boosting Regression model that uses with 1000 estimators and depth of 1.
- Define a Bagging Regression model that uses the Decision Tree as its base estimator.

- Fit both the models on the train data.
- Use the helper code to predict using the mean model and individual estimators. The plot will look similar to the one given above.
- Compute the MSE of the 2 models on the test data.

Hints:

```
sklearn.DecisionTreeRegressor()
```

A decision tree regressor.

```
regressor.fit()
```

Build a decision tree regressor from the training set (X, y).

```
classifier.fit()
```

Build a decision tree classifier from the training set (X, y).

```
sklearn.train_test_split()
```

Split arrays or matrices into om train and test subsets.

```
BaggingRegressor()
```

Returns a Bagging regressor instance.

```
sklearn.mean_squared_error()
```

Mean squared error regression loss.

```
GradientBoostingRegressor()
```

Gradient Boosting for regression.

Note: This exercise is **auto-graded and you can try multiple attempts.**