To predict the outcomes of the Supreme Court, Martin used cases from 1994 through 2001.

He chose this period of time because the Supreme Court was composed of the same nine justices that were justices when he made his predictions in 2002.

These nine justices were Breyer, Ginsburg, Kennedy, O'Connor, Rehnquist-- who was the Chief Justice-- Scalia, Souter, Stevens, and Thomas.

This was a very rare data set, since as we mentioned earlier, this was the longest period of time with the same set of justices in over 180 years.

This allowed Martin to use a larger data set then might have been available if he was doing this experiment at a different time.

In this lecture, we'll focus on predicting Justice Stevens' decisions.

He is generally considered a justice who started out moderate, but became more liberal during his time on the Supreme Court-- although, he's a self-proclaimed conservative.

In this problem, our dependent variable is whether or not Justice Stevens voted to reverse the lower court decision.

This is a binary variable taking value 1 if Justice Stevens decided to reverse or overturn the lower court decision, and taking value 0 if Justice Stevens voted to affirm or maintain the lower court decision.

Our independent variables are six different properties of the case.

The circuit court of origin is the circuit or lower court where the case came from.

There are 13 different circuit courts in the United States.

The 1st through 11th and Washington, DC courts are defined by region.

And the federal court is defined by the subject matter of the case.

The issue area of the case gives each case a category, like civil rights or federal taxation.

The type of petitioner and type of respondent define two parties in the case.

Some examples are the United States, an employer, or an employee.

The ideological direction of the lower court decision describes whether the lower court made what was considered a liberal or a conservative decision.

The last variable indicates whether or not the petitioner argued that a law or practice was unconstitutional.

To collect this data, Martin and his colleagues read through all of the cases and coded the information.

Some of it, like the circuit court, is straightforward.

But other information required a judgment call, like the ideological direction of the lower court.

Now that we have our data and variables, we are ready to predict the decisions of Justice Stevens.

We can use logistic regression.

And we get a model where some of the most significant variables are whether or not the case is from the 2nd circuit court, with a coefficient of 1.66.

Whether or not the case is from the 4th circuit court, with a coefficient of 2.82.

And whether or not the lower court decision was liberal, with a coefficient of negative 1.22.

Well this tells us that the case being from the 2nd or 4th circuit courts is predictive of Justice Stevens reversing the case.

And the lower court decision being liberal is predictive of Justice Stevens affirming the case.

It's difficult to understand which factors are more important due to things like the scales of the variables, and the possibility of multicollinearity.

It's also difficult to quickly evaluate what the prediction would be for a new case.

So instead of logistic regression, Martin and his colleagues used a method called classification and regression trees, or CART.

This method builds what is called a tree by splitting on the values of the independent variables.

To predict the outcome for a new observation or case, you can follow the splits in the tree and at the end, you predict the most frequent outcome in the training set that followed the same path.

Some advantages of CART are that it does not assume a linear model, like logistic regression or linear regression, and it's a very interpretable model.

Let's look at an example.

This plot shows sample data for two independent variables, x and y, and each data point is colored by the outcome variable, red or gray.

CART tries to split this data into subsets so that each subset is as pure or homogeneous as possible.

The first three splits that CART would create are shown here.

Then the standard prediction made by a CART model is just the majority in each subset.

If a new observation fell into one of these two subsets, then we would predict red, since the majority of the observations in those subsets are red.

However, if a new observation fell into one of these two subsets, we would predict gray, since the majority of the observations in those two subsets are gray.

A current model is represented by what we call a tree.

The tree for the splits we just generated is shown on the right.

The first split tests whether the variable x is less than 60.

If yes, the model says to predict red, and if no, the model moves on to the next split.

Then, the second split checks whether or not the variable y is less than 20.

If no, the model says to predict gray, but if yes, the model moves on to the next split.

The third split checks whether or not the variable x is less than 85.

If yes, then the model says to predict red, and if no, the model says to predict gray.

There are a couple things to keep in mind when reading trees.

In this tree, and for the trees we'll generate in R, a yes response is always to the left and a no response is always to the right.

Also, make sure you always start at the top of the tree.

The x less than 85 split only counts for observations for which x is greater than 60 and y is less than 20.

In the next video, we'll discuss how CART decides how many splits to generate and how the final predictions are made.