

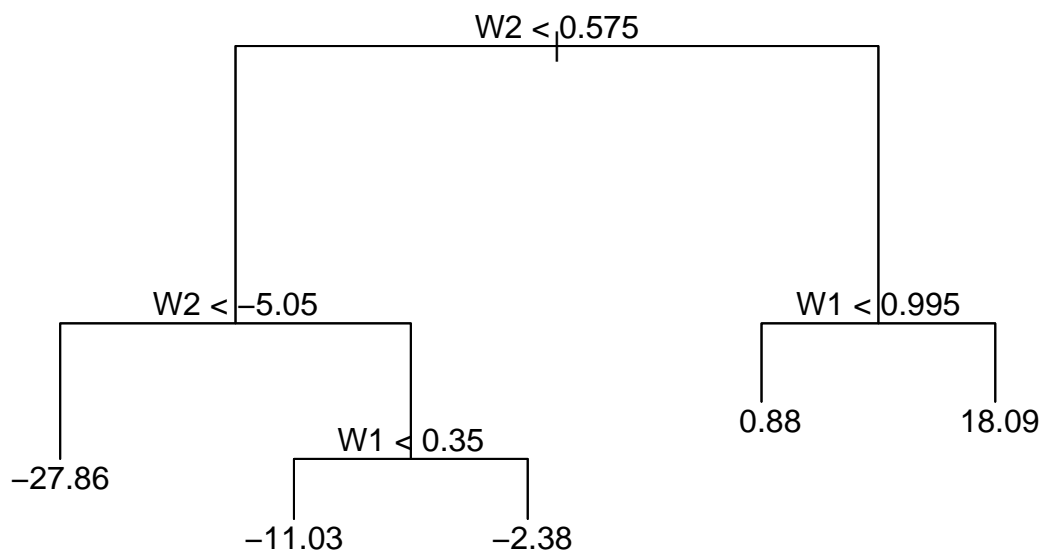
TO DO

INFTrees and INFforests Variable Importance

Theory

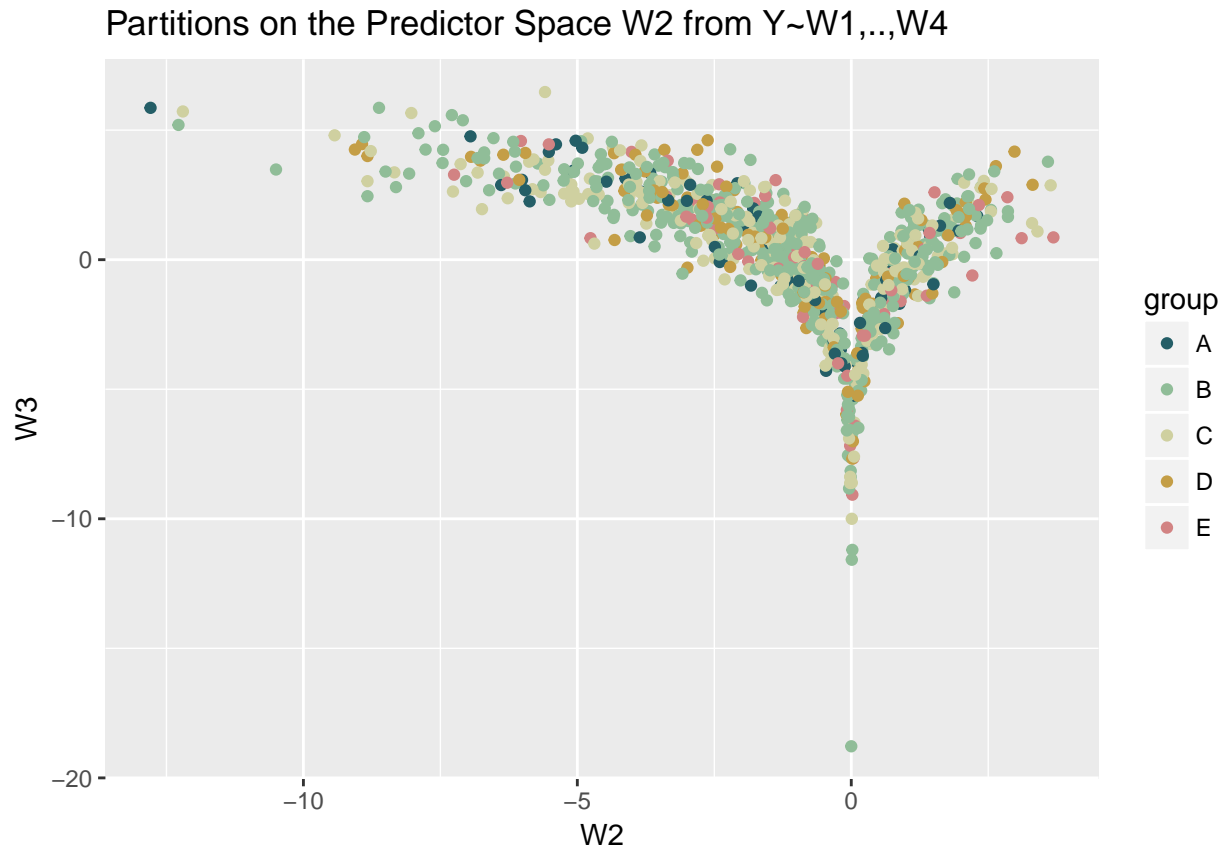
While conditional variable importance (Strobl et al) conditionally permutes each variable given the structure signified by the model that predicts the response, $Y \sim X_1, \dots, X_i, \dots, X_p$, our method conditionally permutes each variable given the structure outlined in a new model with the variable of interest as the response, $X_i \sim X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$. This is not the most straightforward process, as trees partition the sample space, however, in INFTrees these partitions on the variables $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ are treated as psuedo partitions on the variable of interest, X_i . This is accomplished by first partitioning on the sample predictors $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ and then inferring the partitions on X_i . As a visualizaiton of this, lets return to the D_3 dataset discussed in chapter 2.

Figure __: A Tree of the Model $Y \sim \omega_1, \dots, \omega_4$



Lets say we are interested in the variable importance of ω_2 . Then using the conditional variable importance (Strobl et al)'s permutation scheme, we would first look at the partitions on ω_2 from this tree.

Figure __.



Clearly, the values of ω_2 are less important to the partitioning structure than the interactions of ω_2 and the other variables.

```
ggplot(dp, aes(x=W2, y=W3, color = group)) +
  geom_point()+
  scale_color_manual(values = thesis)
```

As you can see in Figure @ref(fig:D2Ch4TreePartition) above, ...

Under the INFTrees method, before permuting, fit another tree to the model $\omega_2 \sim \omega_1, \omega_3, \omega_4$

Figure ____: Tree of the model $\omega_2 \sim \omega_1, \omega_3, \omega_4$

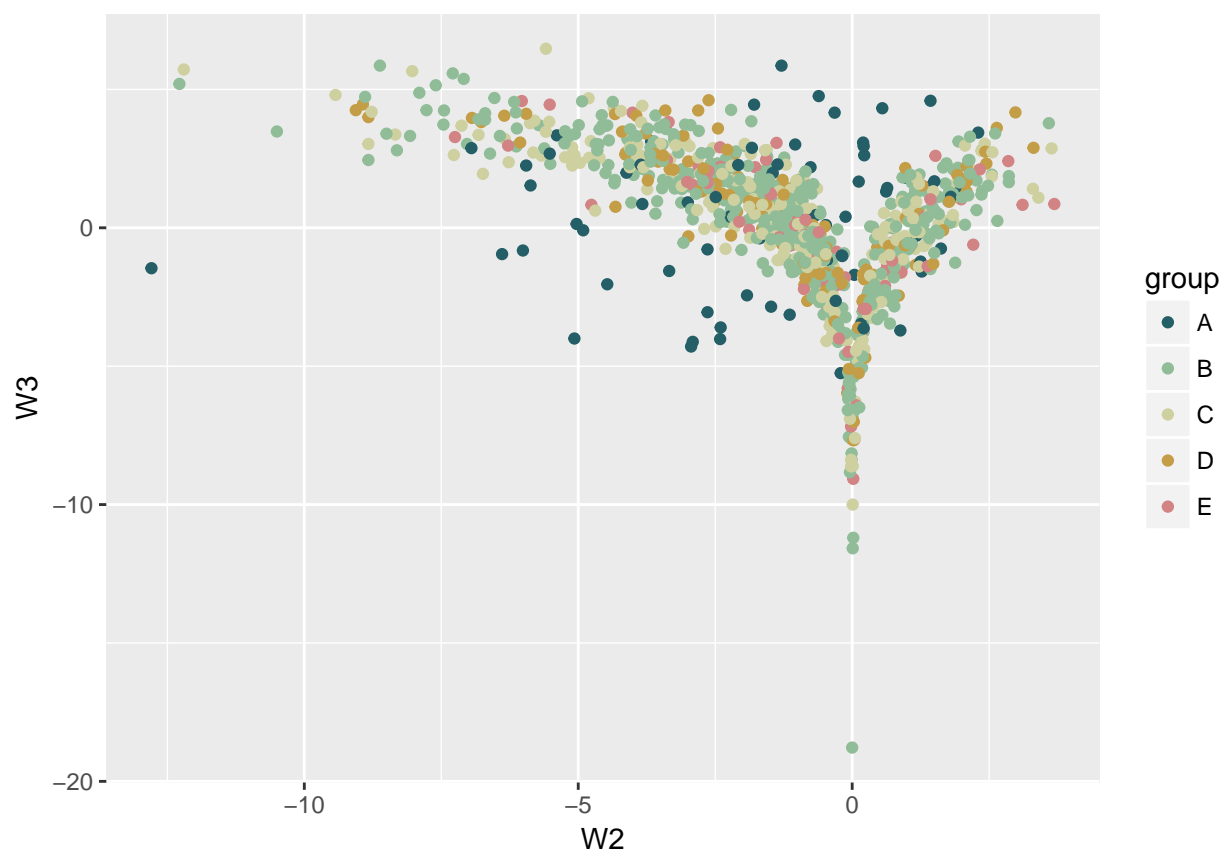
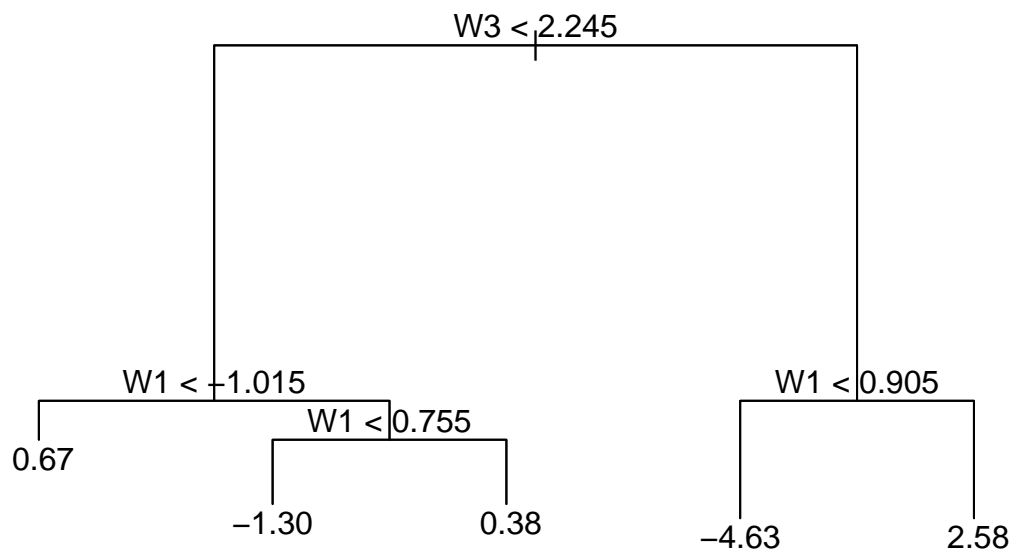


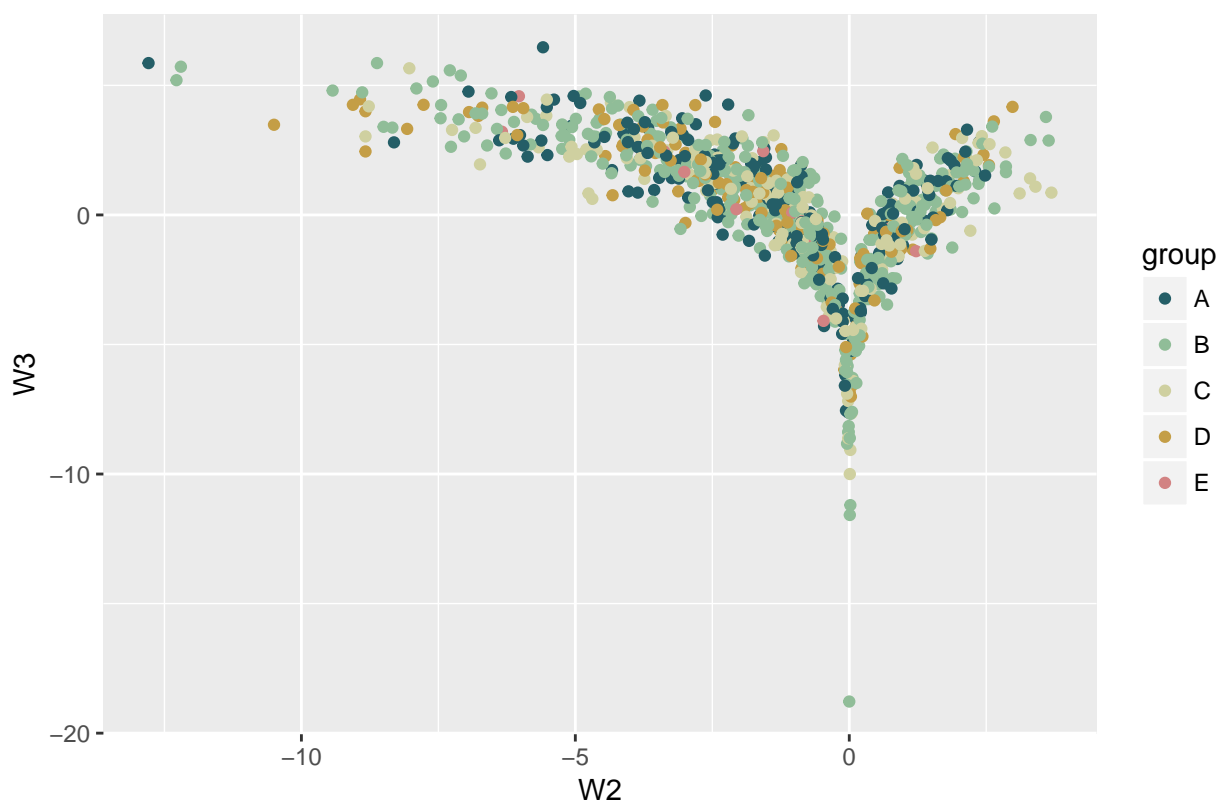
Figure 1: Partitions on the Predictor Space W_2 from $Y \sim W_1, \dots, W_4$

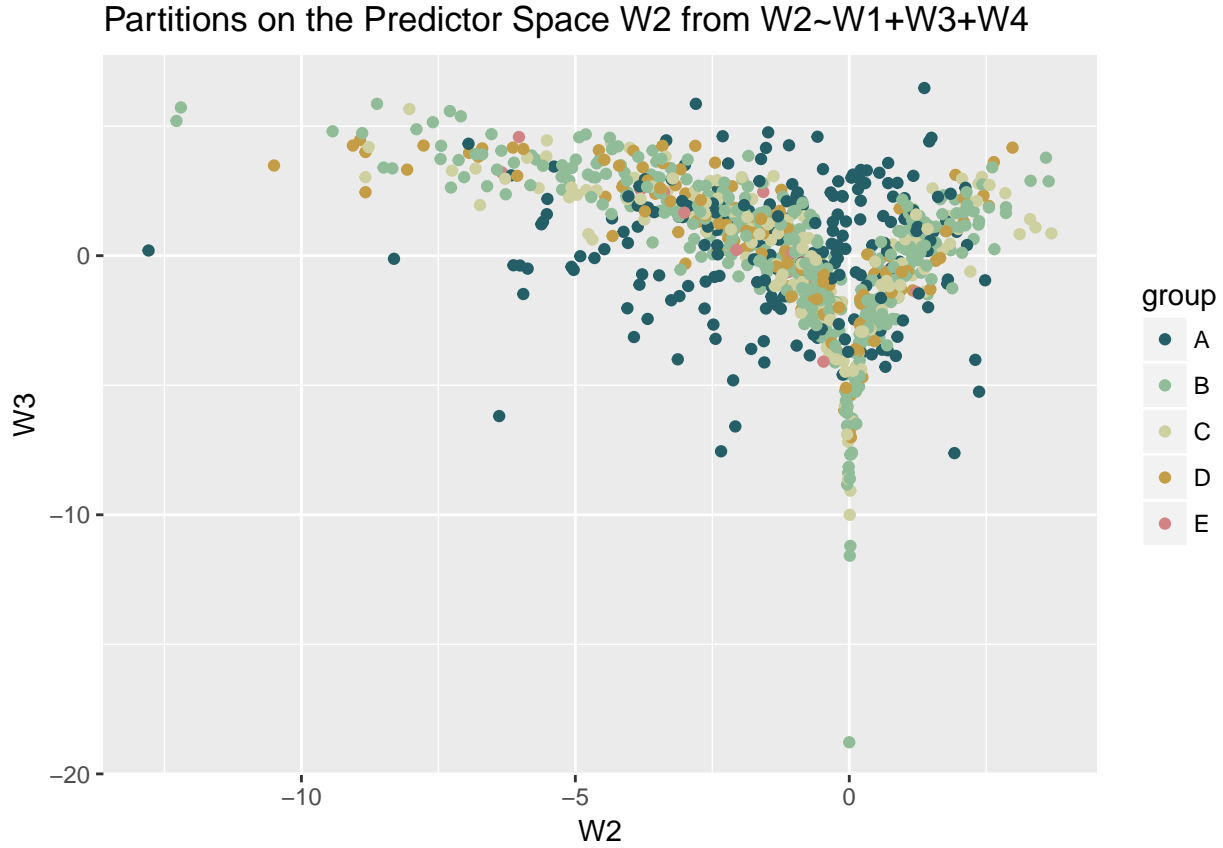


The partitions on ω_2 implied by this model are:

Figure ____.

Partitions on the Predictor Space W_2 from $Y \sim W_1, \dots, W_4$





Need a better way to viz this

INFTrees

For a CART, T , representing the model $Y \sim X_1, \dots, X_p$ where Y, X_1, \dots, X_p are vectors of length n , the INFTrees algorithm proceeds as follows:

Algorithm 1 INFTree, $VI_{inf}(T)$

```

for each  $X_i \in X_1, \dots, X_p$  do
  Calculate:  $\Phi_o = RSS(T, (Y, X_1, \dots, X_p))$ 
  Fit the tree  $T_{X_i}$ , where  $T_{X_i} : X_i \sim X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ 
  Extract the set  $P_{X_i}$  of partitions on  $X_i$  from  $T_{X_i}$ 
  Permute  $X_i$  with respect to  $P_{X_i}$ 
  Find  $\Phi^* = RSS(T, (Y, X_1, \dots, \tilde{X}_i, \dots, X_p))$ 
  Repeat the above procedure to find the distribution of  $\Phi^*$ 
  Test the null hypothesis that  $\Phi_o$  is the likely value of  $RSS(T, (Y, X_1, \dots, X_p))$ 
end for

```

This procedure allows the null hypothesis that Y is independent of X_i given the values of $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ to be tested. Therefore, values of VI_{inf} could be compared in a similar manner to the coefficients of linear regression.

INFForests

The algorithm for determining $VI_{inf}(R)$ follows similarly.

Algorithm 2 INFForests, $VI_{inf}(R)$

- 1: Fit a random forest, R on the dataset D fitting the model $Y \sim X_1, \dots, X_p$.
 - 2: **for** each $X_i \in X_1, \dots, X_p$ **do**
 - 3: **for** each $t \in R$ **do**
 - 4: Calculate: $\Xi_o = \frac{1}{\nu_t} RSS(t, \bar{B}^t)$
 - 5: Calculate a tree T_i that predicts $X_i \sim X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_p$ using the subset of the observations used to fit t
 - 6: Permute the subset of X_i contained in \bar{B}_t with respect to the set of partitions P_{xi} from T_i .
 - 7: Now find $\Xi^* = \frac{1}{\nu_t} RSS(t, \bar{B}_t^*)$
 - 8: The difference between these values, $\Xi^* - \Xi_o$, is the variable importance for X_i on t
 - 9: **end for**
 - 10: Test the null hypothesis that Ξ_o is the likely value of $\frac{1}{\nu_t} RSS(t, \bar{B}_t^*)$ using the distribution of values of Ξ^* gathered from each tree in R
 - 11: **end for**
-

Comparisons and Applications

Trees

```
viW2 <- as.data.frame(viW2)
ggplot(aes(x = viW2), data = viW2) +
  geom_density(fill = thesis[1], alpha = .8) +
  ggtitle("Distribution of RSS when W2 is Conditionally Permuted")+
  geom_vline(xintercept = inft[2,3])
```

Random Forests