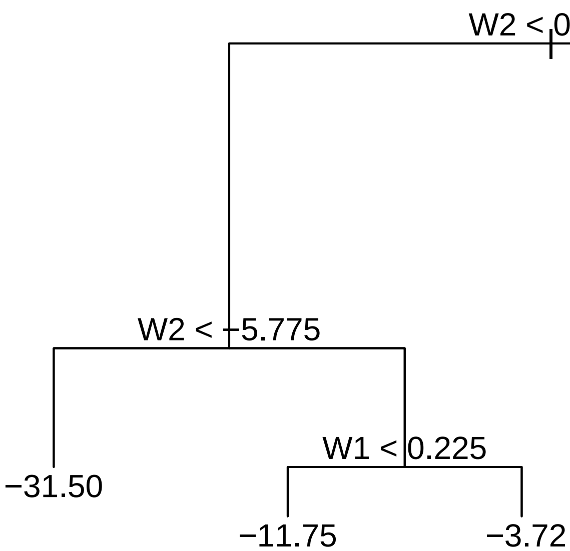
TO DO - The code I wrote for inftrees only gives one value for inftrees, when - compare viz from all three rf variable importances on D3

INFTrees and INFforests Variable Importance

Theory

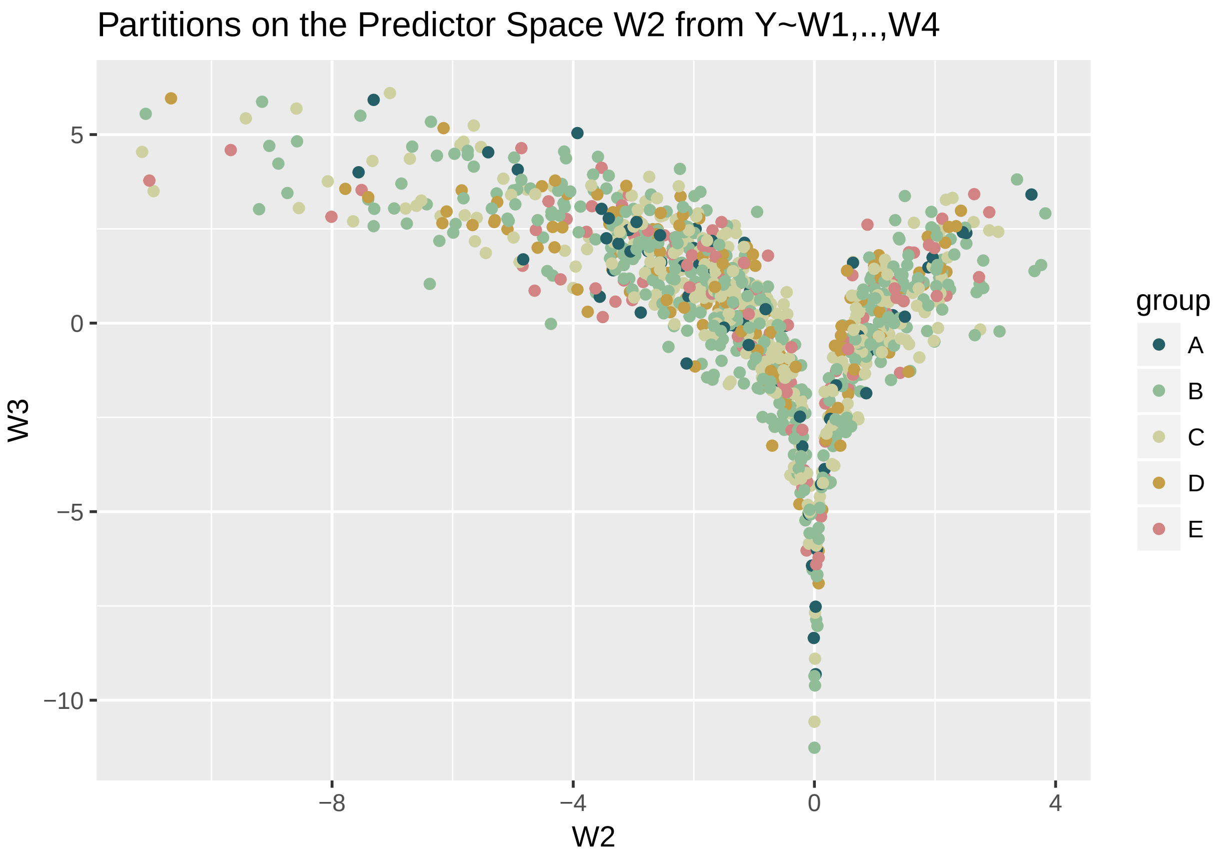
While conditional variable importance (Strobl et al) conditionally permutes each variable given the structure signified by the model that predicts the response, $Y \sim X_1, ..., X_i, ..., X_p$, our method conditionally permutes each variable given the structure outlined in a new model with the variable of interest as the response, $X_i \sim X_1, ...X_{i-1}, X_{i+1}, ...X_p$. This is not the most straightforward process, as trees partition the sample space, however, in INFTrees these partitions on the variables $X_1, ...X_{i-1}, X_{i+1}, ...X_p$ are treated as psuedo partitions on the variable of interest, X_i. This is accomplished by first partitioning on the sample predictors $X_1, ...X_{i-1}, X_{i+1}, ...X_p$ and then infering 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 $Y \sim \omega_1, ..\omega_4$



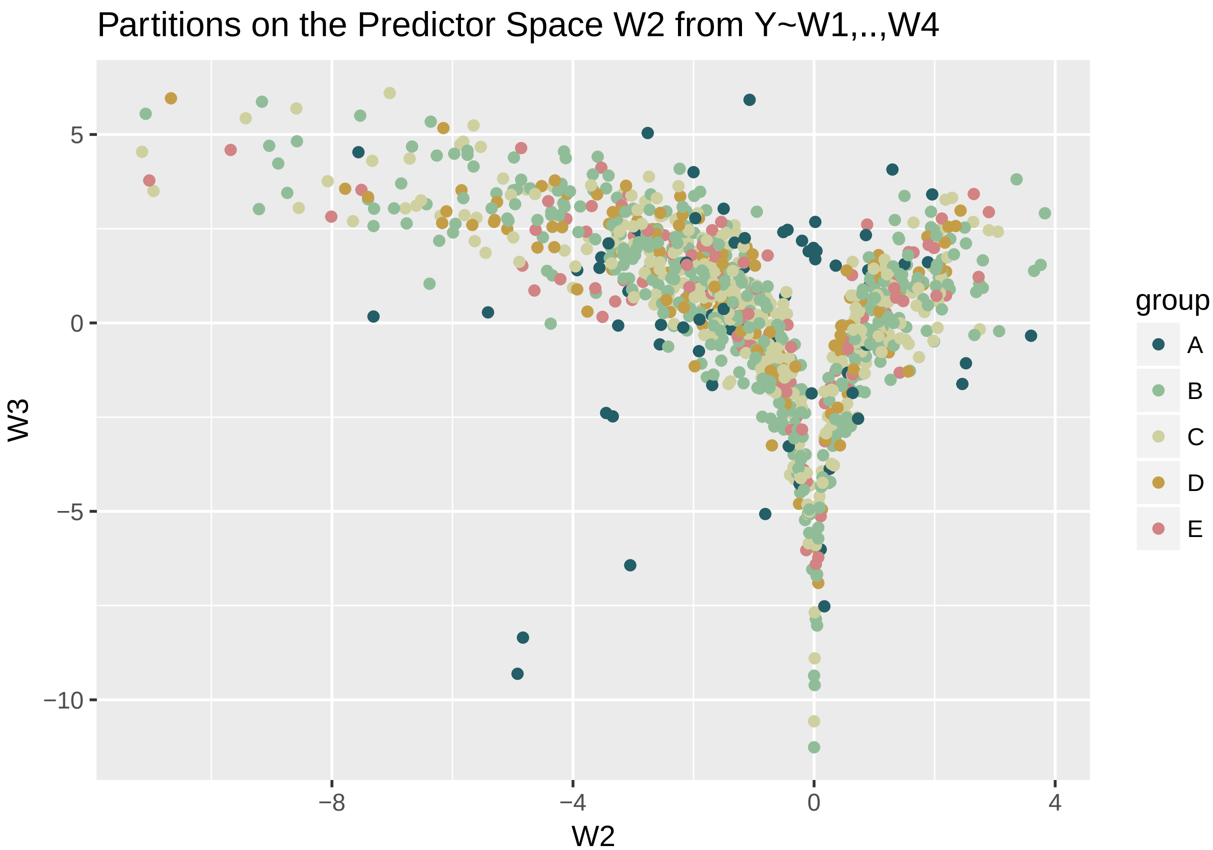
Lets say we are interested in the variable importance of ω_2. Then using the conditional variable importance (Strobl et al)'s permutation scheme, we would first look at the partitions on ω_2 from this tree.

Figure ___.



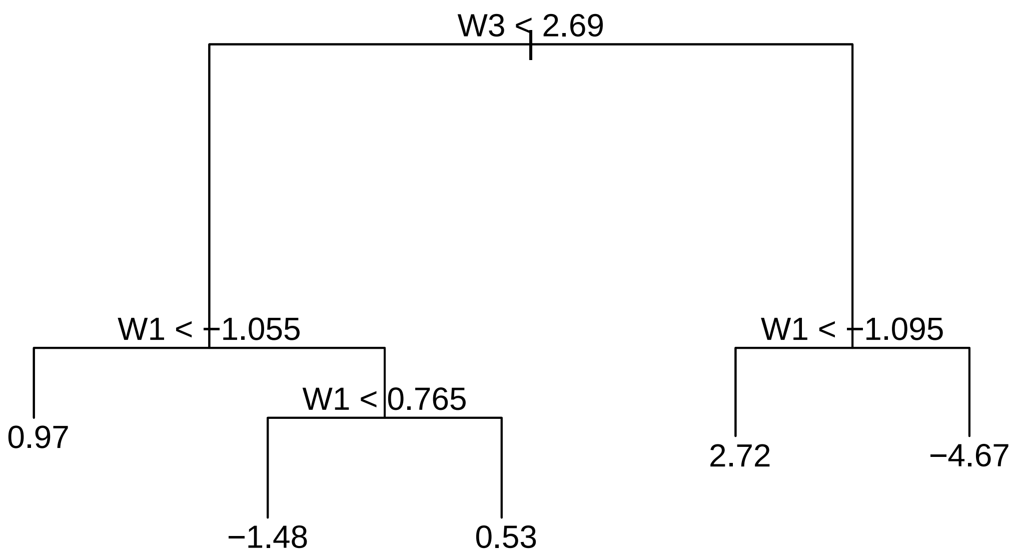
Clearly, the values of ω_2 are less important to the patitioning structure than the interations of ω_2 and the other variables.

Figure___:



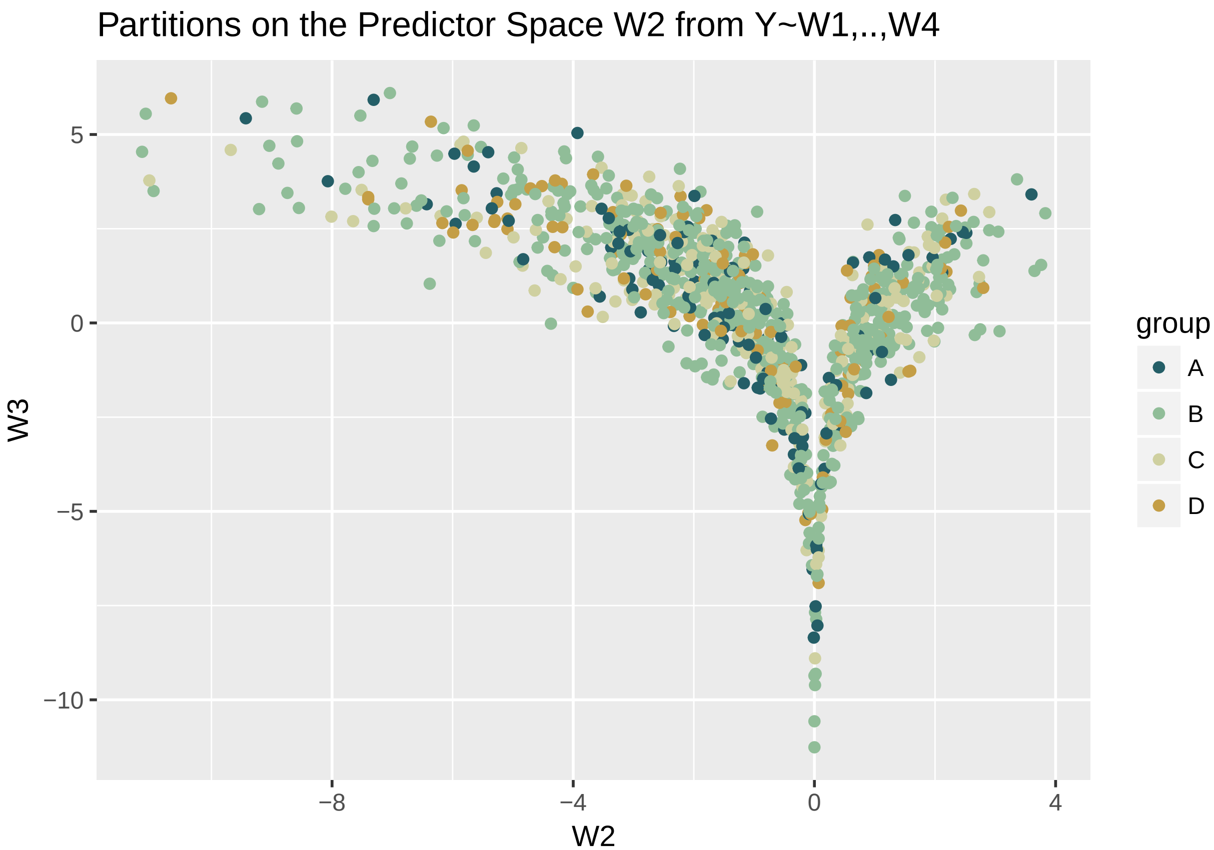
Under the INFTrees method, before permuting, fit another tree to the model $\omega_2 \sim \omega_1, \omega_3, \omega_4$

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



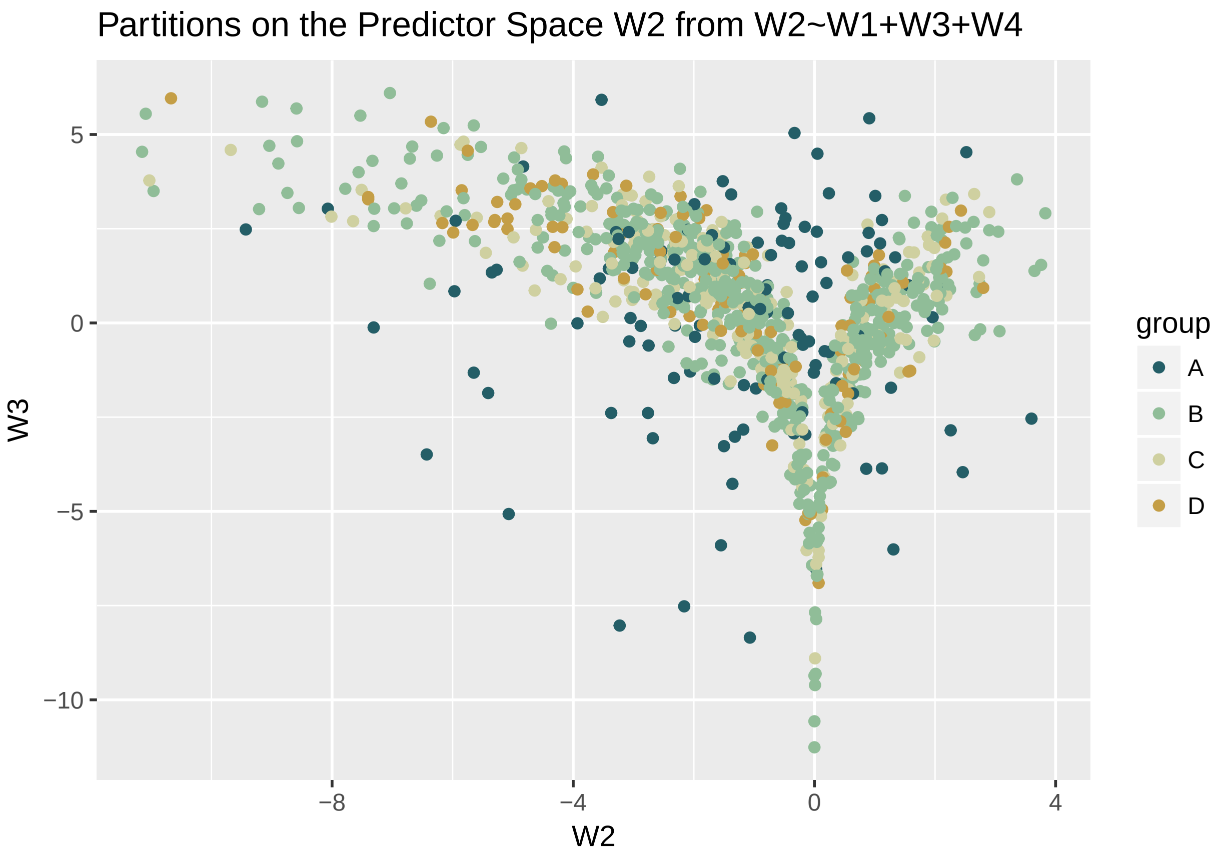
The partitions on ω_2 implied by this model are:

Figure ____.



Partitions on the Predictor Space W2 from Y~W1,..,W4
W3
W2
5
0
−5
−10
−8
−4
0
4
group
A
B
C
D

Figure _____



Need a better way to viz this

INFTrees

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

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

---

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

This procedure allows the null hypothesis that Y is independent of X_i given the values of $X_1, ...X_{i-1}, X_{i+1}, ...X_p$ to be tested. Therefor, 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, $VI_{inf}(R)$

Fit a random forest, R on the dataset D fitting the model $Y \sim X_1, ..., X_p$.
for each $X_i \in X_1, ..., X_p$ **do**
  for each $t \in R$ **do**
    Calculate: $\Xi_o = \frac{1}{\nu_t} RSS(t, \bar{B}^t)$
    Calculate a tree T_i that predicts $X_i \sim X_1, ..., X_{i-1}, X_{i+1}, ...X_p$ using the subset of the observations used to fit t
    Permute the subset of X_i contained in $\bar{B}_t$ with respect to the set of partions P_{xi} from T_i.
    Now find $\Xi^* = \frac{1}{\nu_t} RSS(t, \bar{B}_t^*)$
    The difference between these values, $\Xi^* - \Xi_o$, is the variable importance for X_i on t
  end for
  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
end for

Comparisons and Applications

Trees

variable	inftree.variable.importance	base.variable.importance	coefficients
W1	72237.87	72237.87	5
W2	191388.33	280783.27	5
W3	0.00	14706.58	2
W4	0.00	0.00	0
W5	0.00	47388.18	-5
W6	0.00	62654.33	-5
W7	0.00	0.00	-2
W8	0.00	0.00	0
W9	0.00	0.00	0
W10	0.00	0.00	0
W11	0.00	0.00	0
W12	0.00	0.00	0