Online Appendices

Appendix A Model Outputs and Performance: ANES

A.1 Full Performance Metrics

A.1.1 Accuracy Range Comparison Between Methods

Variable Specification	Year	Logit	CART	RF
Demographics Only	2016	[0.6139, 0.6972]	[0.5806, 0.6656]	[0.5943, 0.6786]
Demo. + PID	2016	[0.8733, 0.9262]	[0.7634, 0.8339]	[0.8776, 0.9296]
Demo. + PID + Issues	2016	[0.8948, 0.9429]	[0.7655, 0.8357]	[0.8818, 0.9330]
All Covariates	2016	[0.8733, 0.9262]	[0.9013, 0.9479]	[0.9601, 0.9880]
Demographics Only	2012	[0.6360, 0.7009]	[0.6385, 0.7032]	[0.6129, 0.6788]
Demo. + PID	2012	[0.9047, 0.9417]	[0.8133, 0.8641]	[0.9008, 0.9385]
Demo. + PID + Issues	2012	[0.9127, 0.9480]	[0.8298, 0.8786]	[0.9314, 0.9626]
All Covariates	2012	[0.8929, 0.9321]	[0.9207, 0.9543]	[0.9576, 0.9815]
Demographics Only	2008	[0.7106, 0.8087]	[0.6831, 0.7846]	[0.6559, 0.7602]
Demo. + PID	2008	[0.8560, 0.9276]	[0.7981, 0.8824]	[0.8560, 0.9276]
Demo. + PID + Issues	2008	[0.8232, 0.9024]	[0.7487, 0.8415]	[0.8672, 0.9358]
All Covariates	2008	[0.4475, 0.5622]	[0.8414, 0.9165]	[0.9012, 0.9598]
Demographics Only	2004	[0.6248, 0.7713]	[0.5668, 0.7196]	[0.5477, 0.7021]
Demo. + PID	2004	[0.8810, 0.9654]	[0.8147, 0.9224]	[0.8510, 0.9469]
Demo. + PID + Issues	2004	[0.7794, 0.8969]	[0.8363, 0.9373]	[0.8734, 0.9609]
All Covariates	2004	[0.4294, 0.5888]	[0.9044, 0.9783]	[0.9125, 0.9823]
Demographics Only	2000	[0.4681, 0.6024]	[0.4948, 0.6285]	[0.4725, 0.6068]
Demo. + PID	2000	[0.8346, 0.9228]	[0.7358, 0.8456]	[0.8194, 0.9116]
Demo. + PID + Issues	2000	[0.6927, 0.8094]	[0.7358, 0.8456]	[0.8194, 0.9116]
All Covariates	2000	[0.4459, 0.5805]	[0.7895, 0.8887]	[0.8707, 0.9482]
Demographics Only	1996	[0.6112, 0.7428]	[0.5563, 0.6925]	[0.5762, 0.7108]
Demo. + PID	1996	[0.8712, 0.9512]	[0.7932, 0.8954]	[0.8654, 0.9474]
Demo. + PID + Issues	1996	[0.7504, 0.8618]	[0.7932, 0.8954]	[0.8484, 0.9358]
All Covariates	1996	[0.4537, 0.5941]	[0.7717, 0.8787]	[0.8712, 0.9512]
Demographics Only	1992	[0.5314, 0.6517]	[0.5351, 0.6553]	[0.5351, 0.6553]
Demo. + PID	1992	[0.7834, 0.8758]	[0.6998, 0.8056]	[0.7673, 0.8626]
Demo. + PID + Issues	1992	[0.6998, 0.8056]	[0.6998, 0.8056]	[0.7875, 0.8790]
All Covariates	1992	[0.4756, 0.5977]	[0.7834, 0.8758]	[0.8664, 0.9392]
Demographics Only	1988	[0.5741, 0.6997]	[0.5870, 0.7117]	[0.5913, 0.7157]
Demo. + PID	1988	[0.8061, 0.8990]	[0.6789, 0.7941]	[0.7921, 0.8881]
Demo. + PID + Issues	1988	[0.7599, 0.8623]	[0.6789, 0.7941]	[0.8061, 0.8990]
All Covariates	1988	[0.4099, 0.5403]	[0.8392, 0.9239]	[0.8881, 0.9578]
Demographics Only	1984	[0.5876, 0.7037]	[0.5876, 0.7037]	[0.5876, 0.7037]
Demo. + PID	1984	[0.8032, 0.8908]	[0.8113, 0.8972]	[0.7833, 0.8749]
Demo. + PID + Issues	1984	[0.7595, 0.8554]	[0.6974, 0.8026]	[0.7833, 0.8749]
All Covariates	1984	[0.5028, 0.6231]	[0.8113, 0.8972]	[0.8942, 0.9579]
Demographics Only	1980	[0.6033, 0.7469]	[0.5559, 0.7039]	[0.5736, 0.7201]
Demo. + PID	1980	[0.7568, 0.8756]	[0.7759, 0.8903]	[0.7695, 0.8854]
Demo. + PID + Issues	1980	[0.7504, 0.8707]	[0.7759, 0.8903]	[0.7631, 0.8805]
All Covariates	1980	[0.4121, 0.5653]	[0.7504, 0.8707]	[0.8482, 0.9420]
Demographics Only	1976	[0.5597, 0.6800]	[0.5290, 0.6508]	[0.5520, 0.6727]
Demo. + PID	1976	[0.7786, 0.8729]	[0.6574, 0.7695]	[0.7622, 0.8594]
Demo. + PID + Issues	1976	[0.7622, 0.8594]	[0.6574, 0.7695]	[0.7622, 0.8594]
All Covariates	1976	[0.4606, 0.5843]	[0.7663, 0.8628]	[0.8678, 0.9409]
Demographics Only	1972	[0.6564, 0.7592]	[0.6663, 0.7680]	[0.6663, 0.7680]
Demo. + PID	1972	[0.7564, 0.8466]	[0.6630, 0.7651]	[0.7429, 0.8352]

Demo. + PID + Issues	1972	[0.7666, 0.8552]	[0.7127, 0.8091]	[0.7940, 0.8778]
All Covariates	1972	[0.4671, 0.5798]	[0.7362, 0.8294]	[0.8322, 0.9083]
Demographics Only	1968	[0.5517, 0.6968]	[0.5574, 0.7020]	[0.5238, 0.6707]
Demo. + PID	1968	[0.7731, 0.8859]	[0.7731, 0.8859]	[0.7489, 0.8670]
Demo. + PID + Issues	1968	[0.7428, 0.8622]	[0.7731, 0.8859]	[0.7489, 0.8670]
All Covariates	1968	[0.5461, 0.6916]	[0.8227, 0.9227]	[0.8612, 0.9489]
Demographics Only	1964	[0.5773, 0.7071]	[0.6098, 0.7368]	[0.5958, 0.7241]
Demo. + PID	1964	[0.7826, 0.8838]	[0.7141, 0.8280]	[0.7727, 0.8759]
Demo. + PID + Issues	1964	[0.7189, 0.8320]	[0.7286, 0.8401]	[0.7826, 0.8838]
All Covariates	1964	[0.4324, 0.5676]	[0.8127, 0.9069]	[0.9075, 0.9718]
Demographics Only	1960	[0.5107, 0.6595]	[0.4218, 0.5728]	[0.4939, 0.6435]
Demo. + PID	1960	[0.7387, 0.8599]	[0.7084, 0.8353]	[0.7448, 0.8647]
Demo. + PID + Issues	1960	[0.7448, 0.8647]	[0.7084, 0.8353]	[0.7326, 0.8550]
All Covariates	1960	[0.5561, 0.7020]	[0.8723, 0.9566]	[0.8858, 0.9649]
Demographics Only	1956	[0.5376, 0.6616]	[0.5256, 0.6502]	[0.5496, 0.6730]
Demo. + PID	1956	[0.8041, 0.8949]	[0.7737, 0.8707]	[0.7780, 0.8742]
Demo. + PID + Issues	1956	[0.8085, 0.8983]	[0.7737, 0.8707]	[0.7780, 0.8742]
All Covariates	1956	[0.7266, 0.8318]	[0.7910, 0.8846]	[0.8806, 0.9510]
Demographics Only	1952	[0.5047, 0.6319]	[0.5457, 0.6711]	[0.5375, 0.6633]
Demo. + PID	1952	[0.7630, 0.8633]	[0.6975, 0.8084]	[0.7586, 0.8597]
Demo. + PID + Issues	1952	[0.7674, 0.8669]	[0.6975, 0.8084]	[0.7586, 0.8597]
All Covariates	1952	[0.6589, 0.7747]	[0.7763, 0.8741]	[0.8772, 0.9496]

Table A.1: Accuracy Range Comparison, Presidential Vote Choice, ANES 1952–2016

A.1.2 Full Performance Tables

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.6935	0.6372	[0.5943, 0.6786]	0.6566	0.6397	0.6480
Demo. + PID	2016	0.9537	0.9060	[0.8776, 0.9296]	0.9339	0.8824	0.9074
Demo. + PID + Issues	2016	0.9707	0.9098	[0.8818, 0.9330]	0.9245	0.9007	0.9125
All Observables	2016	0.9962	0.9770	[0.9601, 0.9880]	0.9815	0.9743	0.9779
Demographics	2012	0.7290	0.6464	[0.6129, 0.6788]	0.7271	0.6513	0.6871
Demo. + PID	2012	0.9697	0.9211	[0.9008, 0.9385]	0.9374	0.9299	0.9336
Demo. + PID + Issues	2012	0.9877	0.9486	[0.9314, 0.9626]	0.9597	0.9539	0.9568
All Observables	2012	0.9971	0.9713	[0.9576, 0.9815]	0.9684	0.9840	0.9761
Demographics	2008	0.7956	0.7101	[0.6559, 0.7602]	0.7164	0.9366	0.8118
Demo. + PID	2008	0.9275	0.8958	[0.8560, 0.9276]	0.9139	0.9317	0.9227
Demo. + PID + Issues	2008	0.9571	0.9055	[0.8672, 0.9358]	0.9190	0.9415	0.9301
All Observables	2008	0.9845	0.9349	[0.9012, 0.9598]	0.9469	0.9561	0.9515
Demographics	2004	0.6830	0.6273	[0.5477, 0.7021]	0.6667	0.4810	0.5588
Demo. + PID	2004	0.9378	0.9068	[0.8510, 0.9469]	0.8810	0.9367	0.9080
Demo. + PID + Issues	2004	0.9823	0.9255	[0.8734, 0.9609]	0.9241	0.9241	0.9241
All Observables	2004	0.9975	0.9565	[0.9125, 0.9823]	0.9286	0.9873	0.9571
Demographics	2000	0.5905	0.5402	[0.4725, 0.6068]	0.5664	0.5424	0.5541
Demo. + PID	2000	0.9217	0.8705	[0.8194, 0.9116]	0.8938	0.8559	0.8745
Demo. + PID + Issues	2000	0.9384	0.8705	[0.8194, 0.9116]	0.8803	0.8729	0.8766
All Observables	2000	0.9691	0.9152	[0.8707, 0.9482]	0.8898	0.9576	0.9224
Demographics	1996	0.7321	0.6456	[0.5762, 0.7108]	0.6599	0.8083	0.7266
Demo. + PID	1996	0.9476	0.9126	[0.8654, 0.9474]	0.9048	0.9500	0.9268
Demo. + PID + Issues	1996	0.9477	0.8981	[0.8484, 0.9358]	0.8779	0.9583	0.9163
All Observables	1996	0.9935	0.9175	[0.8712, 0.9512]	0.8815	0.9917	0.9333
Demographics	1992	0.6601	0.5963	[0.5351, 0.6553]	0.6296	0.7532	0.6859

Demo. + PID	1992	0.8907	0.8185	[0.7673, 0.8626]	0.8263	0.8734	0.8492
Demo. + PID + Issues	1992	0.9207	0.8370	[0.7875, 0.8790]	0.8519	0.8734	0.8625
All Observables	1992	0.9724	0.9074	[0.8664, 0.9392]	0.8889	0.9620	0.9240
Demographics	1988	0.6847	0.6555	[0.5913, 0.7157]	0.7500	0.4018	0.5233
Demo. + PID	1988	0.9113	0.8445	[0.7921, 0.8881]	0.7953	0.9018	0.8452
Demo. + PID + Issues	1988	0.9194	0.8571	[0.8061, 0.8990]	0.8250	0.8839	0.8534
All Observables	1988	0.9758	0.9286	[0.8881, 0.9578]	0.9130	0.9375	0.9251
Demographics	1984	0.6127	0.6473	[0.5876, 0.7037]	0.7368	0.2435	0.3660
Demo. + PID	1984	0.8918	0.8327	[0.7833, 0.8749]	0.8108	0.7826	0.7965
Demo. + PID + Issues	1984	0.9251	0.8327	[0.7833, 0.8749]	0.8165	0.7739	0.7946
All Observables	1984	0.9748	0.9309	[0.8942, 0.9579]	0.9211	0.9130	0.9170
Demographics	1980	0.6478	0.6494	[0.5736, 0.7201]	0.7273	0.3158	0.4404
Demo. + PID	1980	0.8892	0.8333	[0.7695, 0.8854]	0.7582	0.9079	0.8263
Demo. + PID + Issues	1980	0.8841	0.8276	[0.7631, 0.8805]	0.7738	0.8553	0.8125
All Observables	1980	0.9697	0.9023	[0.8482, 0.9420]	0.8734	0.9079	0.8903
Demographics	1976	0.6778	0.6136	[0.5520, 0.6727]	0.6143	0.6418	0.6277
Demo. + PID	1976	0.8854	0.8144	[0.7622, 0.8594]	0.8014	0.8433	0.8218
Demo. + PID + Issues	1976	0.8949	0.8144	[0.7622, 0.8594]	0.7972	0.8507	0.8231
All Observables	1976	0.9718	0.9091	[0.8678, 0.9409]	0.8986	0.9254	0.9118
Demographics	1972	0.6476	0.7192	[0.6663, 0.7680]	0.7400	0.3274	0.4540
Demo. + PID	1972	0.8635	0.7918	[0.7429, 0.8352]	0.7527	0.6195	0.6796
Demo. + PID + Issues	1972	0.8991	0.8391	[0.7940, 0.8778]	0.8039	0.7257	0.7628
All Observables	1972	0.9476	0.8738	[0.8322, 0.9083]	0.8544	0.7788	0.8148
Demographics	1968	0.6336	0.5989	[0.5238, 0.6707]	0.5932	0.4167	0.4895
Demo. + PID	1968	0.8810	0.8132	[0.7489, 0.8670]	0.7551	0.8810	0.8132
Demo. + PID + Issues	1968	0.8776	0.8132	[0.7489, 0.8670]	0.7604	0.8690	0.8111
All Observables	1968	0.9698	0.9121	[0.8612, 0.9489]	0.8696	0.9524	0.9091
Demographics	1964	0.6298	0.6622	[0.5958, 0.7241]	0.6794	0.9467	0.7911
Demo. + PID	1964	0.8482	0.8288	[0.7727, 0.8759]	0.8457	0.9133	0.8782
Demo. + PID + Issues	1964	0.9031	0.8378	[0.7826, 0.8838]	0.8353	0.9467	0.8875
All Observables	1964	0.9733	0.9459	[0.9075, 0.9718]	0.9662	0.9533	0.9597
Demographics	1960	0.5830	0.5698	[0.4939, 0.6435]	0.5600	0.6292	0.5926
Demo. + PID	1960	0.8891	0.8101	[0.7448, 0.8647]	0.7619	0.8989	0.8247
Demo. + PID + Issues	1960	0.8551	0.7989	[0.7326, 0.8550]	0.7732	0.8427	0.8065
All Observables	1960	0.9852	0.9330	[0.8858, 0.9649]	0.9231	0.9438	0.9333
Demographics	1956	0.5678	0.6126	[0.5496, 0.6730]	0.7000	0.0686	0.1250
Demo. + PID	1956	0.8734	0.8300	[0.7780, 0.8742]	0.7521	0.8627	0.8037
Demo. + PID + Issues	1956	0.8883	0.8300	[0.7780, 0.8742]	0.7611	0.8431	0.8000
All Observables	1956	0.9779	0.9209	[0.8806, 0.9510]	0.8796	0.9314	0.9048
Demographics	1952	0.5825	0.6016	[0.5375, 0.6633]	0.6923	0.0874	0.1552
Demo. + PID	1952	0.8719	0.8130	[0.7586, 0.8597]	0.7355	0.8641	0.7946
Demo. + PID + Issues	1952	0.8748	0.8130	[0.7586, 0.8597]	0.7436	0.8447	0.7909
All Observables	1952	0.9698	0.9187	[0.8772, 0.9496]	0.9192	0.8835	0.9010

Table A.2: Performance Metrics, Presidential Vote Choice, Random Forests, ANES 1952–2016

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.7034	0.6564	[0.6139, 0.6972]	0.6768	0.6544	0.6654
Demo. + PID	2016	0.9522	0.9021	[0.8733, 0.9262]	0.9202	0.8897	0.9047
Demo. + PID + Issues	2016	0.9581	0.9213	[0.8948, 0.9429]	0.9459	0.9007	0.9228
All Observables	2016	0.9576	0.9021	[0.8733, 0.9262]	0.9108	0.9007	0.9057
Demographics	2012	0.7342	0.6691	[0.6360, 0.7009]	0.7832	0.6152	0.6891

Demo. + PID	2012	0.9744	0.9247	[0.9047, 0.9417]	0.9523	0.9198	0.9358
Demo. + PID + Issues	2012	0.9719	0.9319	[0.9127, 0.9480]	0.9547	0.9299	0.9421
All Observables	2012	0.9677	0.9140	[0.8929, 0.9321]	0.9313	0.9238	0.9276
Demographics	2008	0.8354	0.7622	[0.7106, 0.8087]	0.8300	0.8098	0.8198
Demo. + PID	2008	0.9522	0.8958	[0.8560, 0.9276]	0.9139	0.9317	0.9227
Demo. + PID + Issues	2008	0.9194	0.8664	[0.8232, 0.9024]	0.8981	0.9024	0.9002
All Observables	2008	0.5162	0.5049	[0.4475, 0.5622]	0.6803	0.4878	0.5682
Demographics	2004	0.7228	0.7019	[0.6248, 0.7713]	0.8163	0.5063	0.6250
Demo. + PID	2004	0.9437	0.9317	[0.8810, 0.9654]	0.9048	0.9620	0.9325
Demo. + PID + Issues	2004	0.8452	0.8447	[0.7794, 0.8969]	0.8214	0.8734	0.8466
All Observables	2004	0.5037	0.5093	[0.4294, 0.5888]	0.5000	0.4557	0.4768
Demographics	2000	0.5817	0.5357	[0.4681, 0.6024]	0.5714	0.4746	0.5185
Demo. + PID	2000	0.9159	0.8839	[0.8346, 0.9228]	0.9107	0.8644	0.8870
Demo. + PID + Issues	2000	0.7545	0.7545	[0.6927, 0.8094]	0.7739	0.7542	0.7639
All Observables	2000	0.5062	0.5134	[0.4459, 0.5805]	0.5385	0.5339	0.5362
Demographics	1996	0.7351	0.6796	[0.6112, 0.7428]	0.7109	0.7583	0.7339
Demo. + PID	1996	0.9451	0.9175	[0.8712, 0.9512]	0.9256	0.9333	0.9295
Demo. + PID + Issues	1996	0.8893	0.8107	[0.7504, 0.8618]	0.8092	0.8833	0.8446
All Observables	1996	0.5307	0.5243	[0.4537, 0.5941]	0.6058	0.5250	0.5625
Demographics	1992	0.6548	0.5926	[0.5314, 0.6517]	0.6319	0.7278	0.6765
Demo. + PID	1992	0.9179	0.8333	[0.7834, 0.8758]	0.8383	0.8861	0.8615
Demo. + PID + Issues	1992	0.7532	0.7556	[0.6998, 0.8056]	0.8026	0.7722	0.7871
All Observables	1992	0.5403	0.5370	[0.4756, 0.5977]	0.6170	0.5506	0.5819
Demographics	1988	0.6548	0.6387	[0.5741, 0.6997]	0.6970	0.4107	0.5169
Demo. + PID	1988	0.9191	0.8571	[0.8061, 0.8990]	0.8250	0.8839	0.8534
Demo. + PID + Issues	1988	0.8800	0.8151	[0.7599, 0.8623]	0.8148	0.7857	0.8000
All Observables	1988	0.4727	0.4748	[0.4099, 0.5403]	0.4414	0.4375	0.4395
Demographics	1984	0.6295	0.6473	[0.5876, 0.7037]	0.6607	0.3217	0.4327
Demo. + PID	1984	0.8929	0.8509	[0.8032, 0.8908]	0.8246	0.8174	0.8210
Demo. + PID + Issues	1984	0.8745	0.8109	[0.7595, 0.8554]	0.7788	0.7652	0.7719
All Observables	1984	0.5649	0.5636	[0.5028, 0.6231]	0.4823	0.5913	0.5313
Demographics	1980	0.6637	0.6782	[0.6033, 0.7469]	0.8571	0.3158	0.4615
Demo. + PID	1980	0.9013	0.8218	[0.7568, 0.8756]	0.7586	0.8684	0.8098
Demo. + PID + Issues	1980	0.9070	0.8161	[0.7504, 0.8707]	0.7821	0.8026	0.7922
All Observables	1980	0.4902	0.4885	[0.4121, 0.5653]	0.4235	0.4737	0.4472
Demographics	1976	0.7005	0.6212	[0.5597, 0.6800]	0.6214	0.6493	0.6350
Demo. + PID	1976	0.8896	0.8295	[0.7786, 0.8729]	0.8201	0.8507	0.8352
Demo. + PID + Issues	1976	0.9107	0.8144	[0.7622, 0.8594]	0.8058	0.8358	0.8205
All Observables	1976	0.5280	0.5227	[0.4606, 0.5843]	0.5270	0.5821	0.5532
Demographics	1972	0.7058	0.7098	[0.6564, 0.7592]	0.7692	0.2655	0.3947
Demo. + PID	1972	0.8616	0.8044	[0.7564, 0.8466]	0.7429	0.6903	0.7156
Demo. + PID + Issues	1972	0.8793	0.8139	[0.7666, 0.8552]	0.7455	0.7257	0.7354
All Observables	1972	0.5180	0.5237	[0.4671, 0.5798]	0.3623	0.4425	0.3984
Demographics	1968	0.6566	0.6264	[0.5517, 0.6968]	0.6600	0.3929	0.4925
Demo. + PID	1968	0.9082	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
Demo. + PID + Issues	1968	0.8562	0.8077	[0.7428, 0.8622]	0.7882	0.7976	0.7929
All Observables	1968	0.6199	0.6209	[0.5461, 0.6916]	0.5824	0.6310	0.6057
Demographics	1964	0.6643	0.6441	[0.5773, 0.7071]	0.6859	0.8733	0.7683
Demo. + PID	1964	0.8744	0.8378	[0.7826, 0.8838]	0.8562	0.9133	0.8839
Demo. + PID + Issues	1964	0.8598	0.7793	[0.7189, 0.8320]	0.8217	0.8600	0.8404
All Observables	1964	0.5129	0.5000	[0.4324, 0.5676]	0.6857	0.4800	0.5647
Demographics	1960	0.5910	0.5866	[0.5107, 0.6595]	0.5714	0.6742	0.6186
Demo. + PID	1960	0.8858	0.8045	[0.7387, 0.8599]	0.7547	0.8989	0.8205
Demo. + PID + Issues	1960	0.8752	0.8101	[0.7448, 0.8647]	0.7723	0.8764	0.8211

All Observables	1960	0.6551	0.6313	[0.5561, 0.7020]	0.6386	0.5955	0.6163
Demographics	1956	0.5745	0.6008	[0.5376, 0.6616]	0.5152	0.1667	0.2519
Demo. + PID	1956	0.9019	0.8538	[0.8041, 0.8949]	0.7778	0.8922	0.8311
Demo. + PID + Issues	1956	0.9093	0.8577	[0.8085, 0.8983]	0.7895	0.8824	0.8333
All Observables	1956	0.8489	0.7826	[0.7266, 0.8318]	0.7196	0.7549	0.7368
Demographics	1952	0.5539	0.5691	[0.5047, 0.6319]	0.4615	0.1748	0.2535
Demo. + PID	1952	0.8757	0.8171	[0.7630, 0.8633]	0.7458	0.8544	0.7964
Demo. + PID + Issues	1952	0.8787	0.8211	[0.7674, 0.8669]	0.7521	0.8544	0.8000
All Observables	1952	0.7472	0.7195	[0.6589, 0.7747]	0.6417	0.7476	0.6906

Table A.3: Performance Metrics, Presidential Vote Choice, Logit, ANES 1952–2016

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.6751	0.6238	[0.5806, 0.6656]	0.6250	0.6985	0.6597
Demo. + PID	2016	0.8007	0.8004	[0.7634, 0.8339]	0.7414	0.9485	0.8323
Demo. + PID + Issues	2016	0.8535	0.8023	[0.7655, 0.8357]	0.7752	0.8750	0.8221
All Observables	2016	0.9519	0.9271	[0.9013, 0.9479]	0.9398	0.9191	0.9294
Demographics	2012	0.7240	0.6714	[0.6385, 0.7032]	0.8522	0.5431	0.6634
Demo. + PID	2012	0.8082	0.8399	[0.8133, 0.8641]	0.7977	0.9800	0.8795
Demo. + PID + Issues	2012	0.8909	0.8554	[0.8298, 0.8786]	0.8921	0.8617	0.8767
All Observables	2012	0.9504	0.9391	[0.9207, 0.9543]	0.9590	0.9379	0.9483
Demographics	2008	0.8011	0.7362	[0.6831, 0.7846]	0.8647	0.7171	0.7840
Demo. + PID	2008	0.7849	0.8436	[0.7981, 0.8824]	0.8285	0.9659	0.8919
Demo. + PID + Issues	2008	0.7727	0.7980	[0.7487, 0.8415]	0.8295	0.8780	0.8531
All Observables	2008	0.8884	0.8827	[0.8414, 0.9165]	0.8894	0.9415	0.9147
Demographics	2004	0.6639	0.6460	[0.5668, 0.7196]	0.7200	0.4557	0.5581
Demo. + PID	2004	0.8780	0.8758	[0.8147, 0.9224]	0.7980	1.0000	0.8876
Demo. + PID + Issues	2004	0.8983	0.8944	[0.8363, 0.9373]	0.8605	0.9367	0.8970
All Observables	2004	0.9712	0.9503	[0.9044, 0.9783]	0.9277	0.9747	0.9506
Demographics	2000	0.5892	0.5625	[0.4948, 0.6285]	0.5794	0.6186	0.5984
Demo. + PID	2000	0.7910	0.7946	[0.7358, 0.8456]	0.7338	0.9576	0.8309
Demo. + PID + Issues	2000	0.7910	0.7946	[0.7358, 0.8456]	0.7338	0.9576	0.8309
All Observables	2000	0.8396	0.8438	[0.7895, 0.8887]	0.8217	0.8983	0.8583
Demographics	1996	0.6553	0.6262	[0.5563, 0.6925]	0.7216	0.5833	0.6452
Demo. + PID	1996	0.8274	0.8495	[0.7932, 0.8954]	0.8112	0.9667	0.8821
Demo. + PID + Issues	1996	0.8274	0.8495	[0.7932, 0.8954]	0.8112	0.9667	0.8821
All Observables	1996	0.8238	0.8301	[0.7717, 0.8787]	0.7815	0.9833	0.8708
Demographics	1992	0.6430	0.5963	[0.5351, 0.6553]	0.6339	0.7342	0.6804
Demo. + PID	1992	0.7289	0.7556	[0.6998, 0.8056]	0.7277	0.9304	0.8167
Demo. + PID + Issues	1992	0.7289	0.7556	[0.6998, 0.8056]	0.7277	0.9304	0.8167
All Observables	1992	0.8367	0.8333	[0.7834, 0.8758]	0.8122	0.9304	0.8673
Demographics	1988	0.6353	0.6513	[0.5870, 0.7117]	0.8537	0.3125	0.4575
Demo. + PID	1988	0.7554	0.7395	[0.6789, 0.7941]	0.6524	0.9554	0.7754
Demo. + PID + Issues	1988	0.7554	0.7395	[0.6789, 0.7941]	0.6524	0.9554	0.7754
All Observables	1988	0.8878	0.8866	[0.8392, 0.9239]	0.9126	0.8393	0.8744
Demographics	1984	0.5819	0.6473	[0.5876, 0.7037]	0.8750	0.1826	0.3022
Demo. + PID	1984	0.8886	0.8582	[0.8113, 0.8972]	0.8167	0.8522	0.8340
Demo. + PID + Issues	1984	0.7211	0.7527	[0.6974, 0.8026]	0.8133	0.5304	0.6421
All Observables	1984	0.9024	0.8582	[0.8113, 0.8972]	0.7923	0.8957	0.8408
Demographics	1980	0.5963	0.6322	[0.5559, 0.7039]	0.7000	0.2763	0.3962
Demo. + PID	1980	0.8820	0.8391	[0.7759, 0.8903]	0.7500	0.9474	0.8372
Demo. + PID + Issues	1980	0.8820	0.8391	[0.7759, 0.8903]	0.7500	0.9474	0.8372

All Observables	1980	0.8300	0.8161	[0.7504, 0.8707]	0.7895	0.7895	0.7895
Demographics	1976	0.6534	0.5909	[0.5290, 0.6508]	0.5756	0.7388	0.6471
Demo. + PID	1976	0.7224	0.7159	[0.6574, 0.7695]	0.6545	0.9328	0.7692
Demo. + PID + Issues	1976	0.7224	0.7159	[0.6574, 0.7695]	0.6545	0.9328	0.7692
All Observables	1976	0.8036	0.8182	[0.7663, 0.8628]	0.7905	0.8731	0.8298
Demographics	1972	0.6779	0.7192	[0.6663, 0.7680]	0.6875	0.3894	0.4972
Demo. + PID	1972	0.6116	0.7161	[0.6630, 0.7651]	0.8485	0.2478	0.3836
Demo. + PID + Issues	1972	0.8397	0.7634	[0.7127, 0.8091]	0.6638	0.6814	0.6725
All Observables	1972	0.8244	0.7855	[0.7362, 0.8294]	0.8689	0.4690	0.6092
Demographics	1968	0.6014	0.6319	[0.5574, 0.7020]	0.7297	0.3214	0.4463
Demo. + PID	1968	0.8469	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
Demo. + PID + Issues	1968	0.8469	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
All Observables	1968	0.9181	0.8791	[0.8227, 0.9227]	0.8690	0.8690	0.8690
Demographics	1964	0.5834	0.6757	[0.6098, 0.7368]	0.6893	0.9467	0.7978
Demo. + PID	1964	0.6713	0.7748	[0.7141, 0.8280]	0.7632	0.9667	0.8529
Demo. + PID + Issues	1964	0.7675	0.7883	[0.7286, 0.8401]	0.8366	0.8533	0.8449
All Observables	1964	0.8862	0.8649	[0.8127, 0.9069]	0.8448	0.9800	0.9074
Demographics	1960	0.4900	0.4972	[0.4218, 0.5728]	0.4968	0.8764	0.6341
Demo. + PID	1960	0.7848	0.7765	[0.7084, 0.8353]	0.7025	0.9551	0.8095
Demo. + PID + Issues	1960	0.7848	0.7765	[0.7084, 0.8353]	0.7025	0.9551	0.8095
All Observables	1960	0.9361	0.9218	[0.8723, 0.9566]	0.8947	0.9551	0.9239
Demographics	1956	0.5613	0.5889	[0.5256, 0.6502]	0.4833	0.2843	0.3580
Demo. + PID	1956	0.8941	0.8261	[0.7737, 0.8707]	0.7544	0.8431	0.7963
Demo. + PID + Issues	1956	0.8961	0.8261	[0.7737, 0.8707]	0.7685	0.8137	0.7905
All Observables	1956	0.8834	0.8419	[0.7910, 0.8846]	0.7981	0.8137	0.8058
Demographics	1952	0.5795	0.6098	[0.5457, 0.6711]	0.5778	0.2524	0.3514
Demo. + PID	1952	0.7865	0.7561	[0.6975, 0.8084]	0.6387	0.9612	0.7674
Demo. + PID + Issues	1952	0.7865	0.7561	[0.6975, 0.8084]	0.6387	0.9612	0.7674
All Observables	1952	0.8602	0.8293	[0.7763, 0.8741]	0.7607	0.8641	0.8091

Table A.4: Performance Metrics, Presidential Vote Choice, CART, ANES 1952–2016

A.1.3 ROC Curves

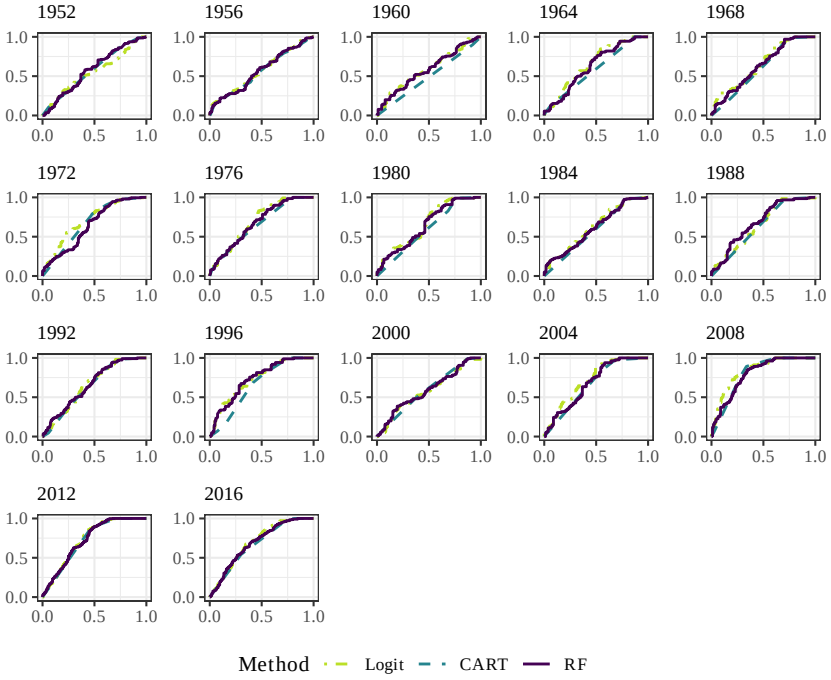


Figure A.1: ROC Curves for Presidential Vote Choice, Demographics Only, Comparison Between Logit, CART, and RF

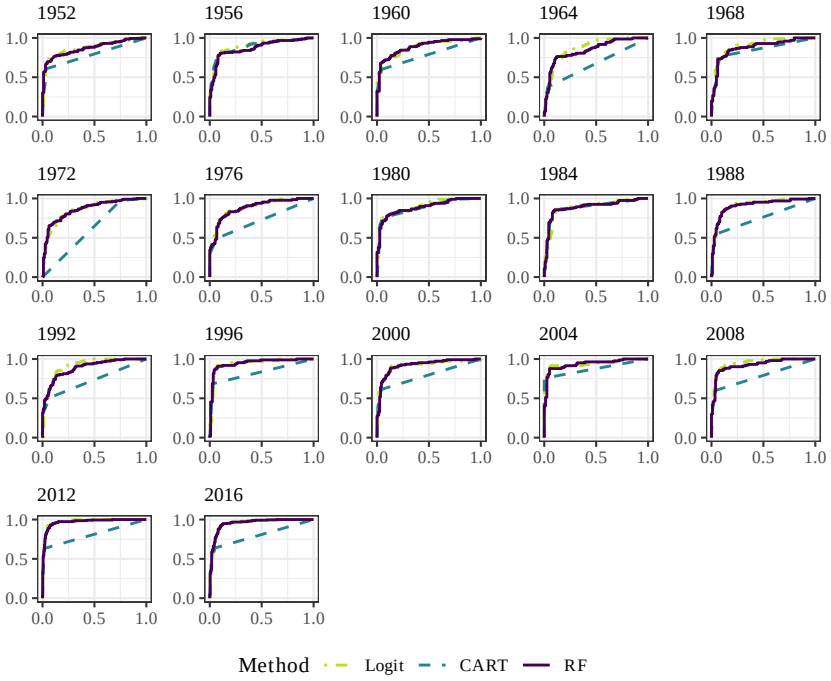


Figure A.2: ROC Curves for Presidential Vote Choice, Demo. + PID, Comparison Between Logit, CART, and RF

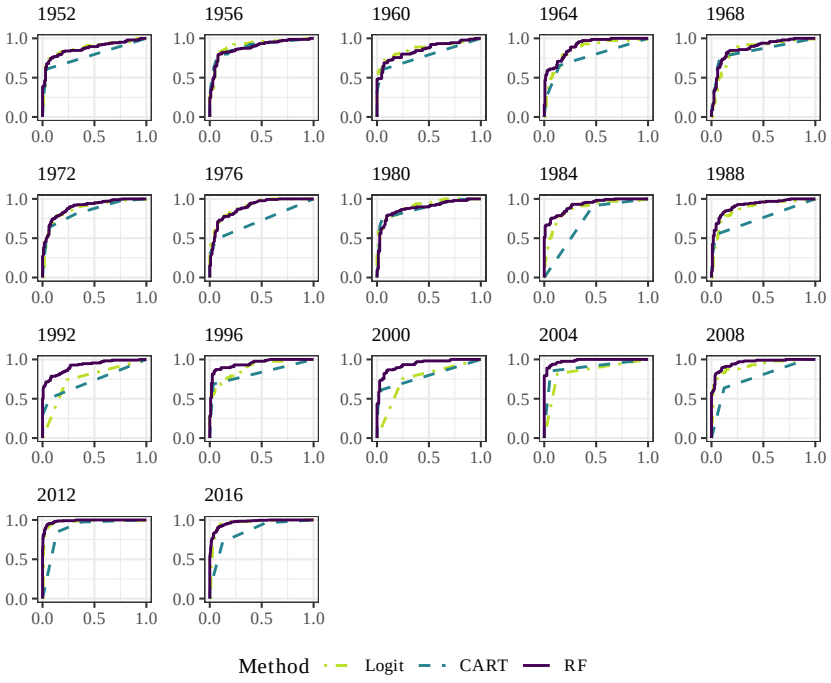


Figure A.3: ROC Curves for Presidential Vote Choice, Demo. + PID + Issues, Comparison Between Logit, CART, and RF

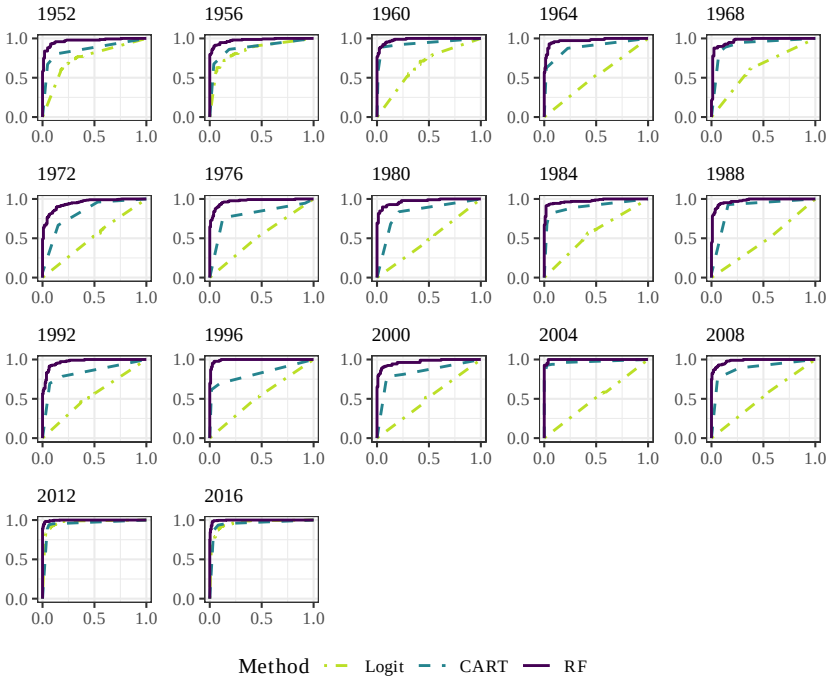
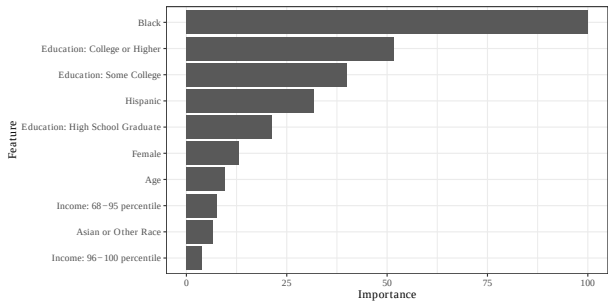


Figure A.4: ROC Curves for Presidential Vote Choice, All Covariates, Comparison Between Logit, CART, and RF

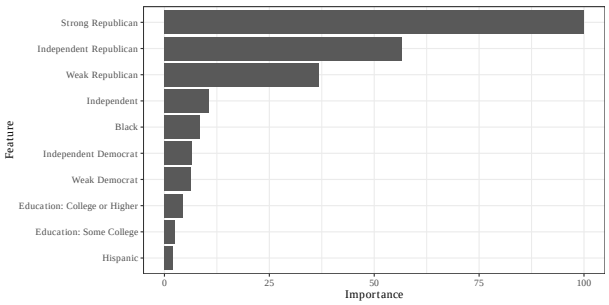
Note that logit's performance becomes very poor in the saturated model (more variables than number of observations) throughout years 1960–2008.

A.2 Variable Importance Measures

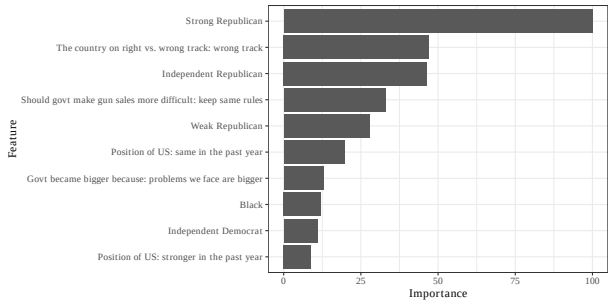
The following are top 10 variables per specification and survey wave, visualized.



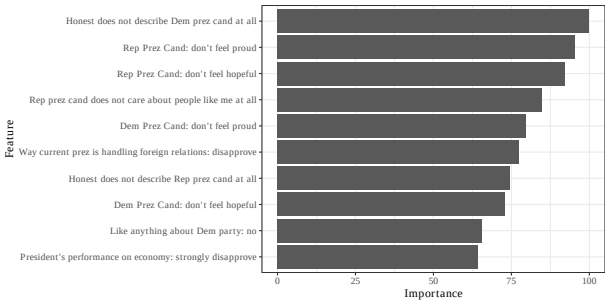
(a) ANES, 2016, Spec. 1 (Demographics Only)



(b) ANES, 2016, Spec. 2 (Demo. + PID)

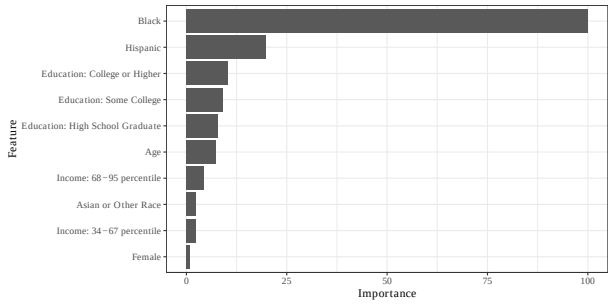


(c) ANES, 2016, Spec. 3 (Demo. + PID + Issues)

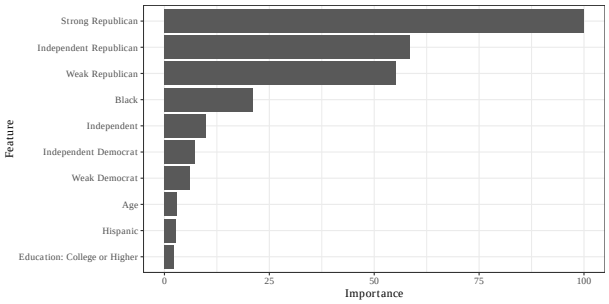


(d) ANES, 2016, Spec. 4 (All Covariates)

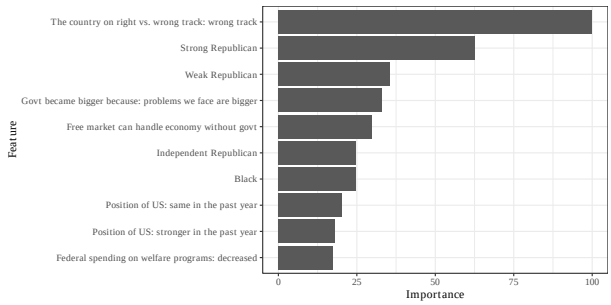
Figure A.5: ANES, Top 10 Variables from Random Forests, 2016



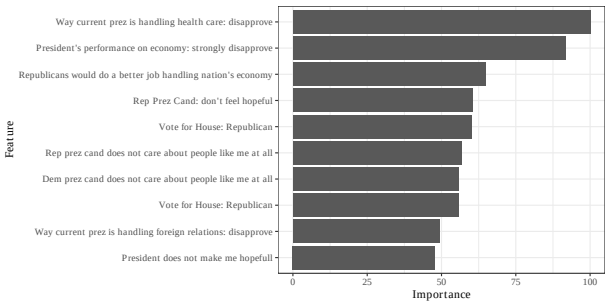
(a) ANES, 2012, Spec. 1 (Demographics Only)



(b) ANES, 2012, Spec. 2 (Demo. + PID)

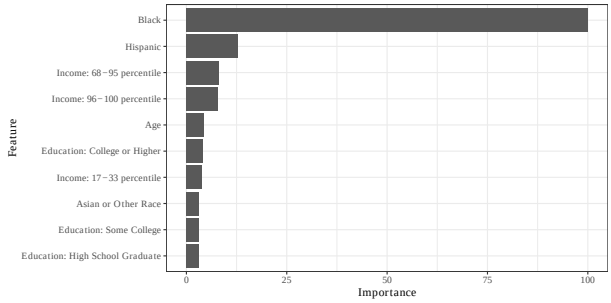


(c) ANES, 2012, Spec. 3 (Demo. + PID + Issues)

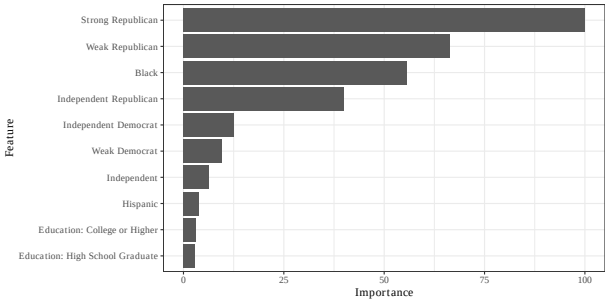


(d) ANES, 2012, Spec. 4 (All Covariates)

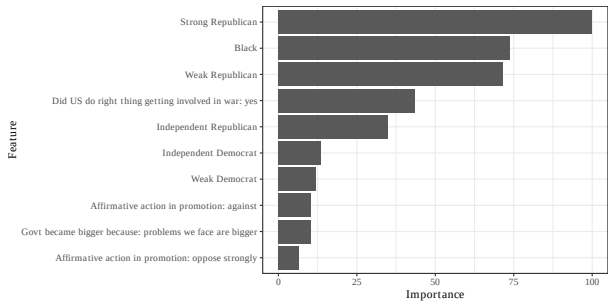
Figure A.6: ANES, Top 10 Variables from Random Forests, 2012



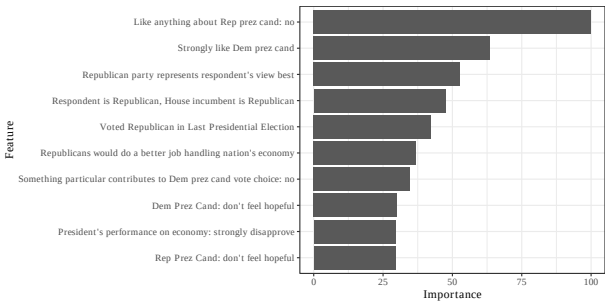
(a) ANES, 2008, Spec. 1 (Demographics Only)



(b) ANES, 2008, Spec. 2 (Demo. + PID)

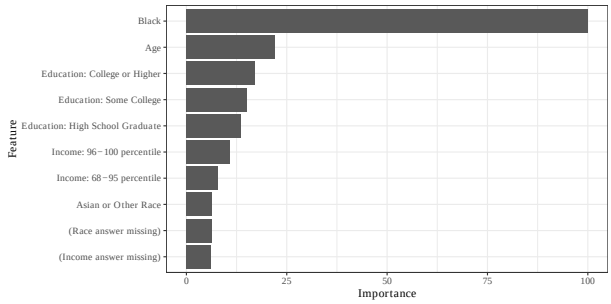


(c) ANES, 2008, Spec. 3 (Demo. + PID + Issues)

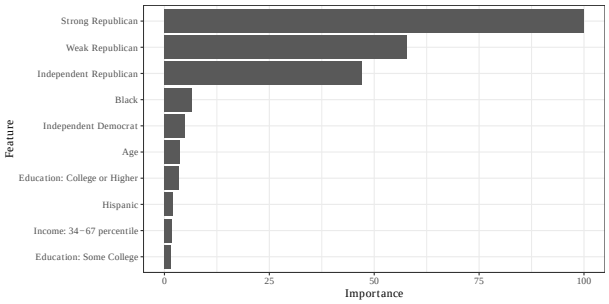


(d) ANES, 2008, Spec. 4 (All Covariates)

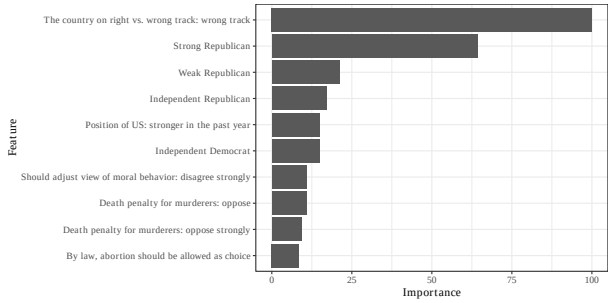
Figure A.7: ANES, Top 10 Variables from Random Forests, 2008



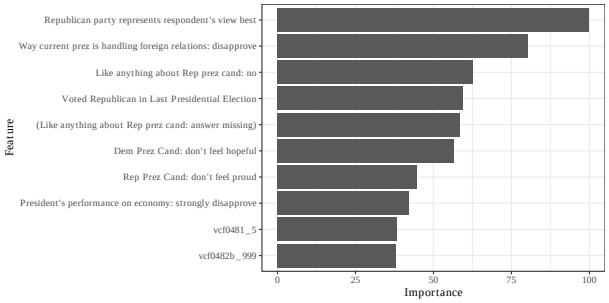
(a) ANES, 2004, Spec. 1 (Demographics Only)



(b) ANES, 2004, Spec. 2 (Demo. + PID)

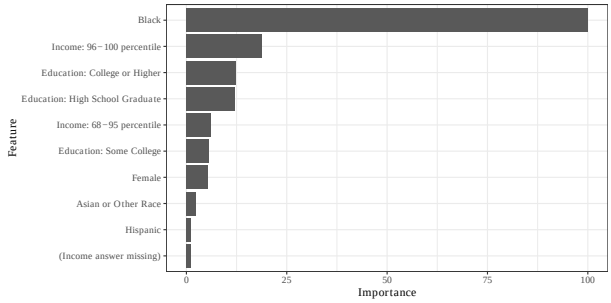


(c) ANES, 2004, Spec. 3 (Demo. + PID + Issues)

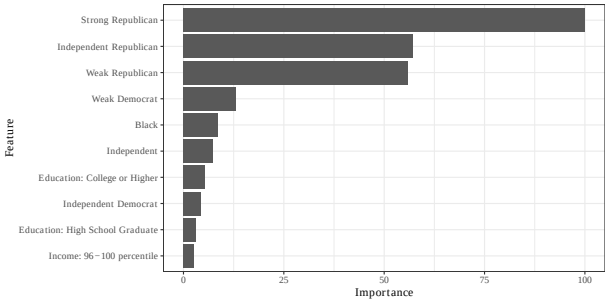


(d) ANES, 2004, Spec. 4 (All Covariates)

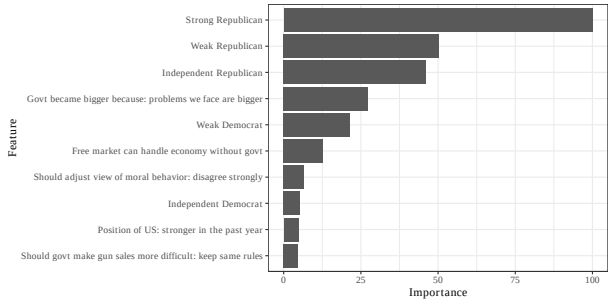
Figure A.8: ANES, Top 10 Variables from Random Forests, 2004



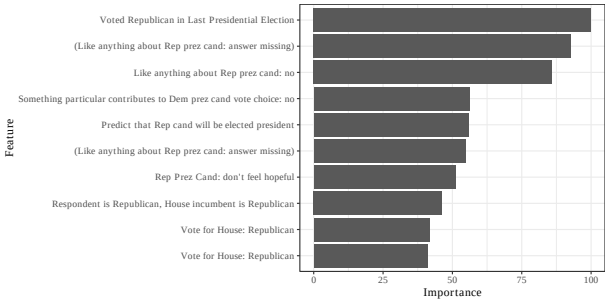
(a) ANES, 2000, Spec. 1 (Demographics Only)



(b) ANES, 2000, Spec. 2 (Demo. + PID)

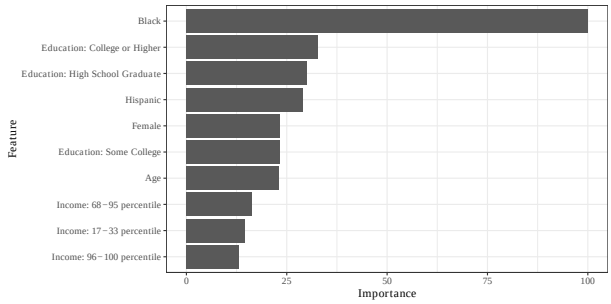


(c) ANES, 2000, Spec. 3 (Demo. + PID + Issues)

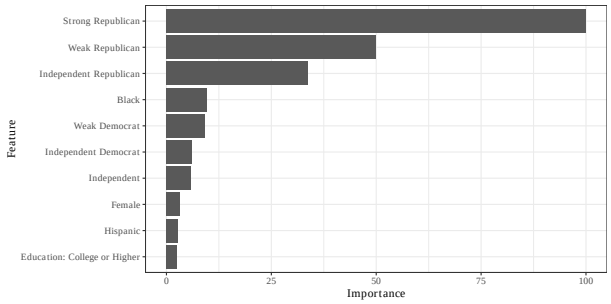


(d) ANES, 2000, Spec. 4 (All Covariates)

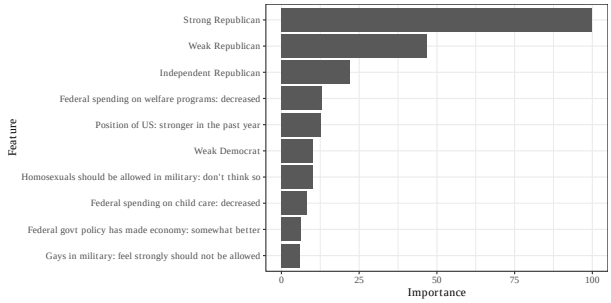
Figure A.9: ANES, Top 10 Variables from Random Forests, 2000



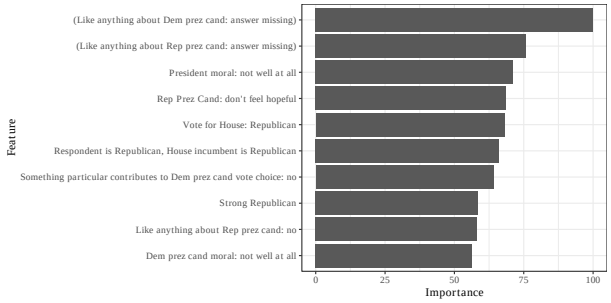
(a) ANES, 1996, Spec. 1 (Demographics Only)



(b) ANES, 1996, Spec. 2 (Demo. + PID)

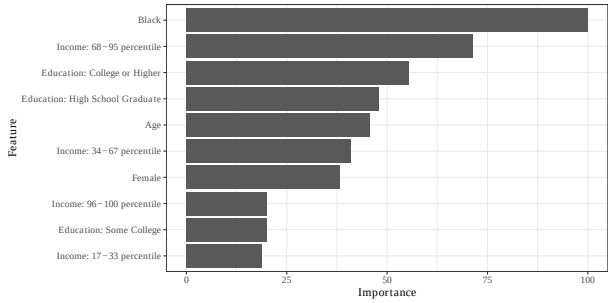


(c) ANES, 1996, Spec. 3 (Demo. + PID + Issues)

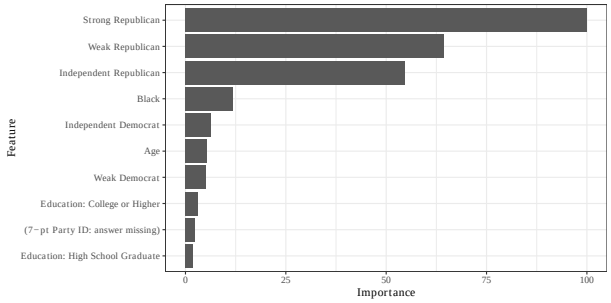


(d) ANES, 1996, Spec. 4 (All Covariates)

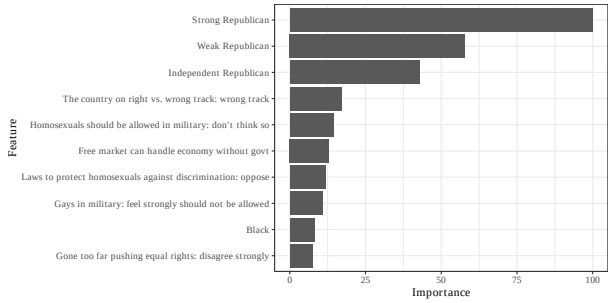
Figure A.10: ANES, Top 10 Variables from Random Forests, 1996



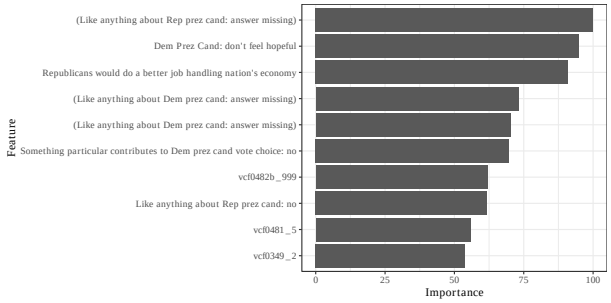
(a) ANES, 1992, Spec. 1 (Demographics Only)



(b) ANES, 1992, Spec. 2 (Demo. + PID)

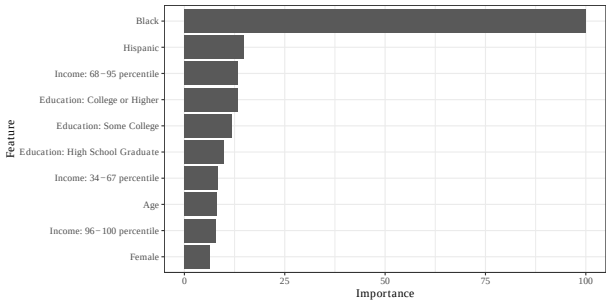


(c) ANES, 1992, Spec. 3 (Demo. + PID + Issues)

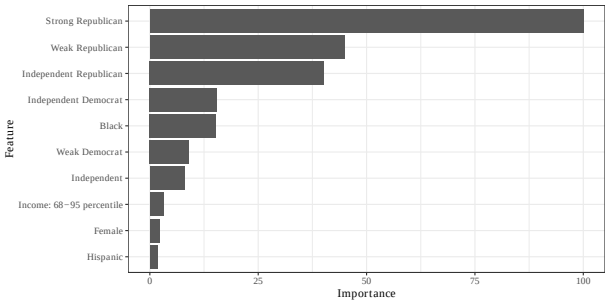


(d) ANES, 1992, Spec. 4 (All Covariates)

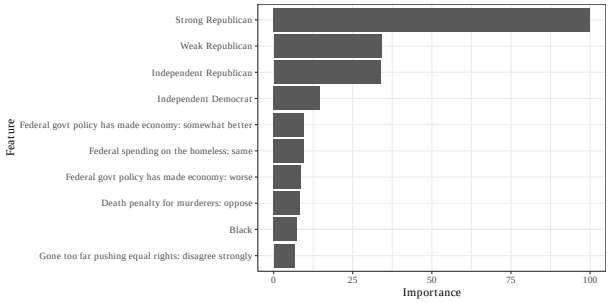
Figure A.11: ANES, Top 10 Variables from Random Forests, 1992



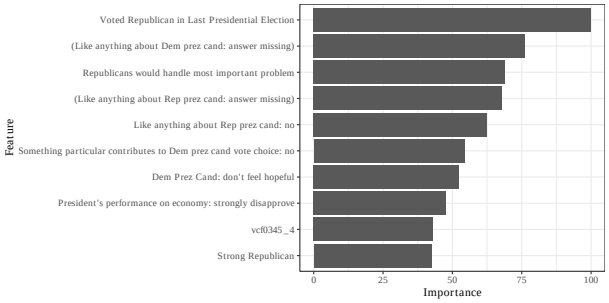
(a) ANES, 1988, Spec. 1 (Demographics Only)



(b) ANES, 1988, Spec. 2 (Demo. + PID)

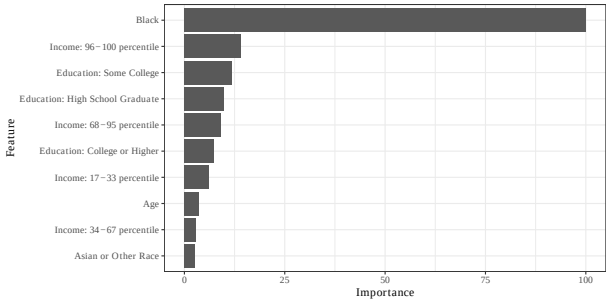


(c) ANES, 1988, Spec. 3 (Demo. + PID + Issues)

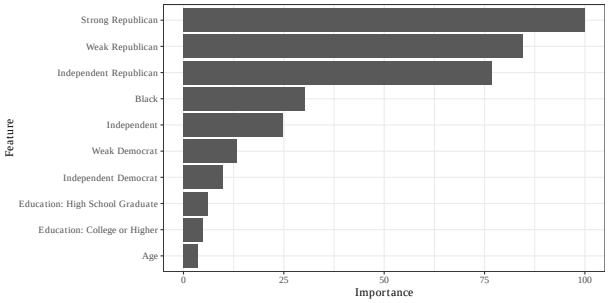


(d) ANES, 1988, Spec. 4 (All Covariates)

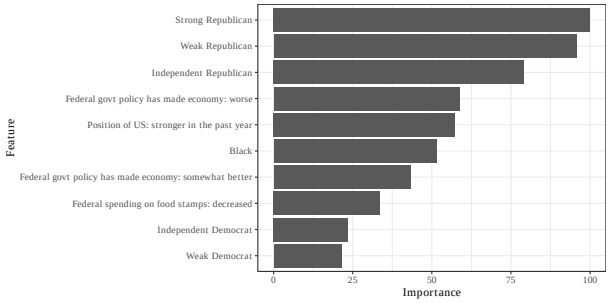
Figure A.12: ANES, Top 10 Variables from Random Forests, 1988



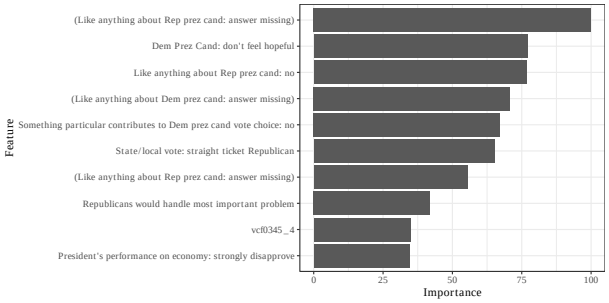
(a) ANES, 1984, Spec. 1 (Demographics Only)



(b) ANES, 1984, Spec. 2 (Demo. + PID)

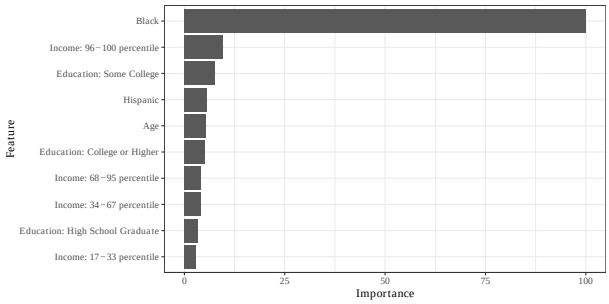


(c) ANES, 1984, Spec. 3 (Demo. + PID + Issues)

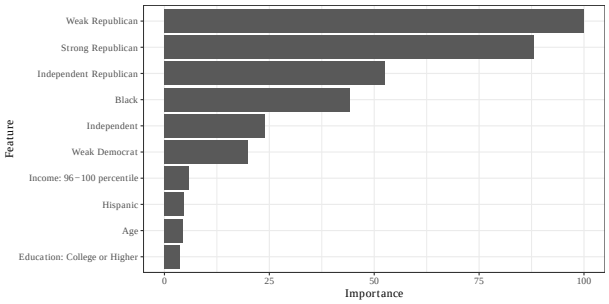


(d) ANES, 1984, Spec. 4 (All Covariates)

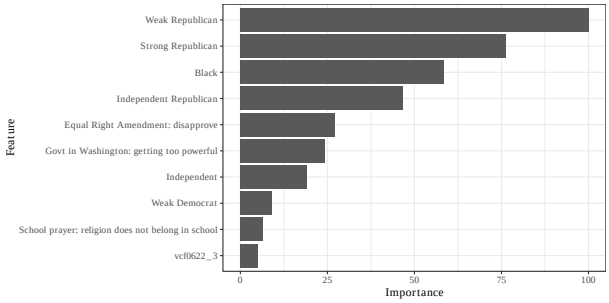
Figure A.13: ANES, Top 10 Variables from Random Forests, 1984



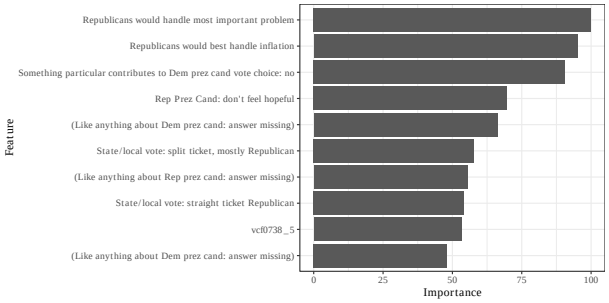
(a) ANES, 1980, Spec. 1 (Demographics Only)



(b) ANES, 1980, Spec. 2 (Demo. + PID)

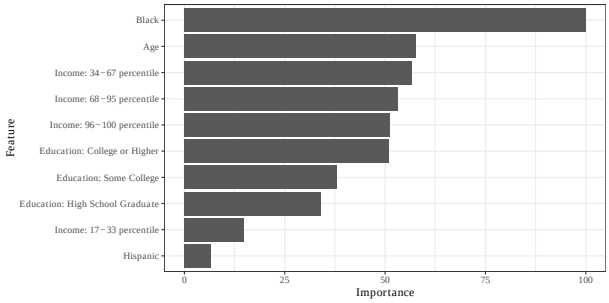


(c) ANES, 1980, Spec. 3 (Demo. + PID + Issues)

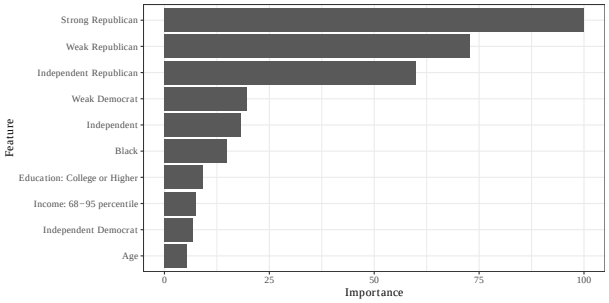


(d) ANES, 1980, Spec. 4 (All Covariates)

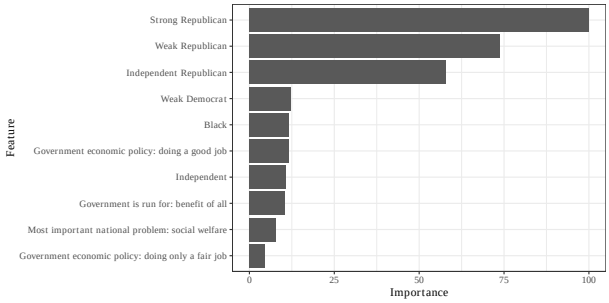
Figure A.14: ANES, Top 10 Variables from Random Forests, 1980



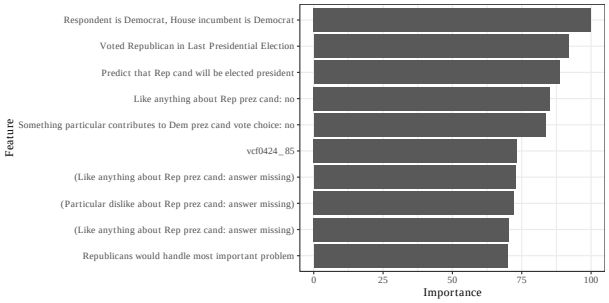
(a) ANES, 1976, Spec. 1 (Demographics Only)



(b) ANES, 1976, Spec. 2 (Demo. + PID)

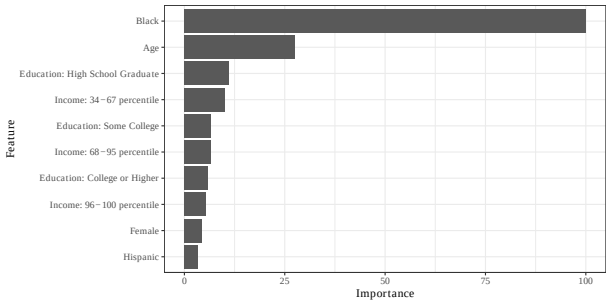


(c) ANES, 1976, Spec. 3 (Demo. + PID + Issues)

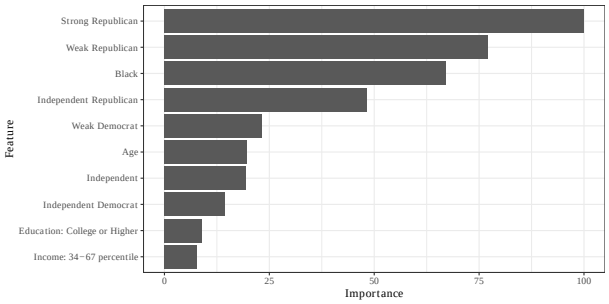


(d) ANES, 1976, Spec. 4 (All Covariates)

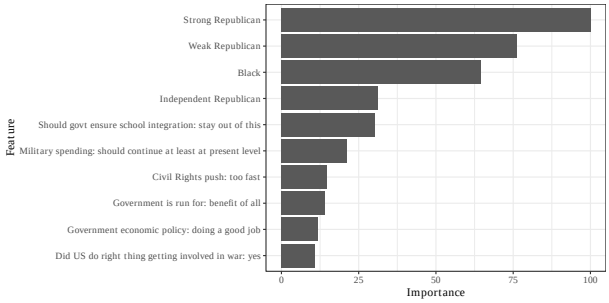
Figure A.15: ANES, Top 10 Variables from Random Forests, 1976



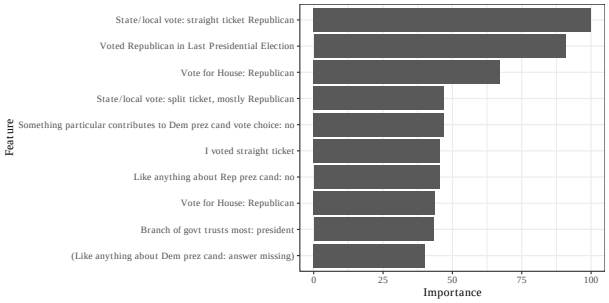
(a) ANES, 1972, Spec. 1 (Demographics Only)



(b) ANES, 1972, Spec. 2 (Demo. + PID)

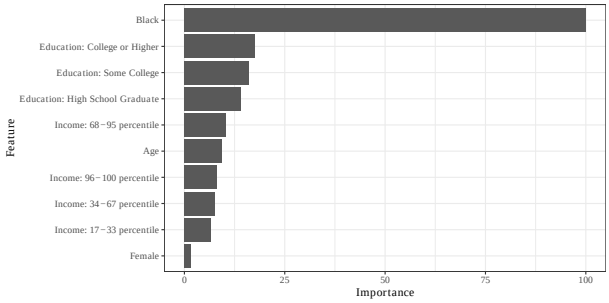


(c) ANES, 1972, Spec. 3 (Demo. + PID + Issues)

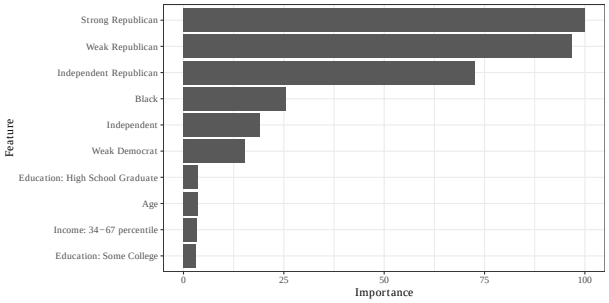


(d) ANES, 1972, Spec. 4 (All Covariates)

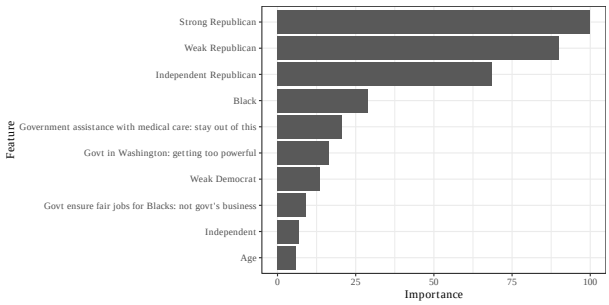
Figure A.16: ANES, Top 10 Variables from Random Forests, 1972



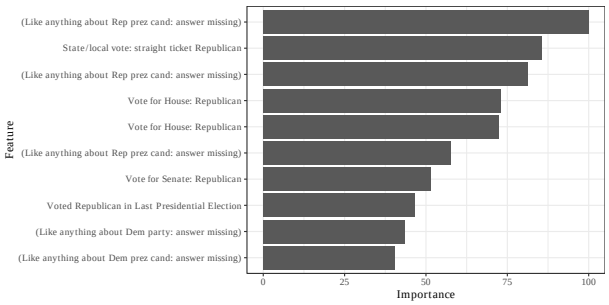
(a) ANES, 1968, Spec. 1 (Demographics Only)



(b) ANES, 1968, Spec. 2 (Demo. + PID)

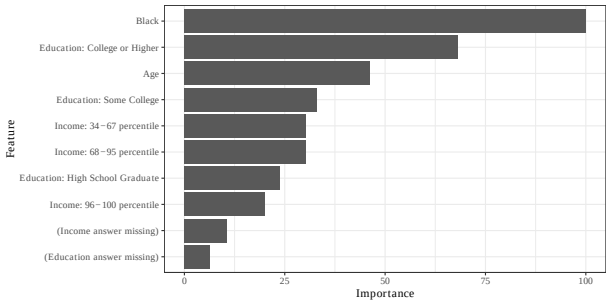


(c) ANES, 1968, Spec. 3 (Demo. + PID + Issues)

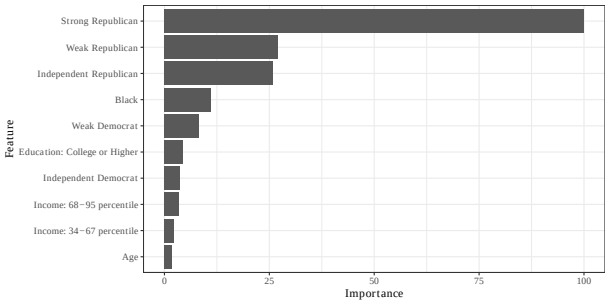


(d) ANES, 1968, Spec. 4 (All Covariates)

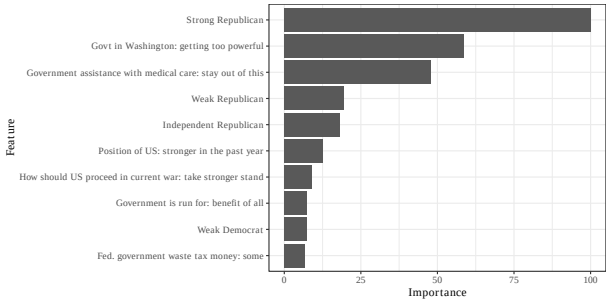
Figure A.17: ANES, Top 10 Variables from Random Forests, 1968



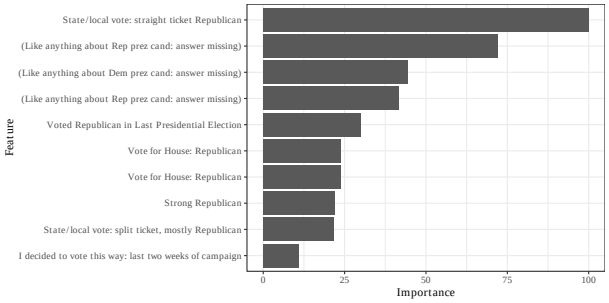
(a) ANES, 1964, Spec. 1 (Demographics Only)



(b) ANES, 1964, Spec. 2 (Demo. + PID)

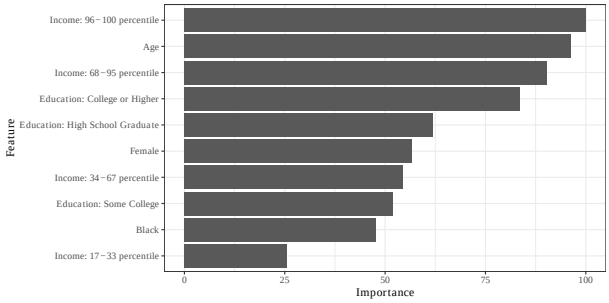


(c) ANES, 1964, Spec. 3 (Demo. + PID + Issues)

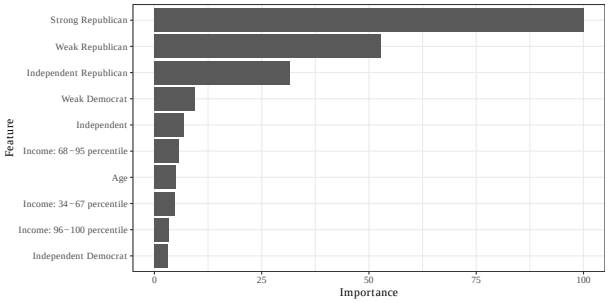


(d) ANES, 1964, Spec. 4 (All Covariates)

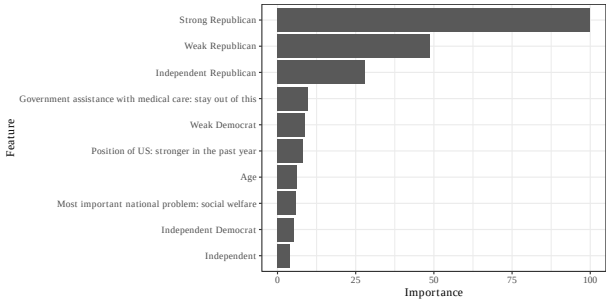
Figure A.18: ANES, Top 10 Variables from Random Forests, 1964



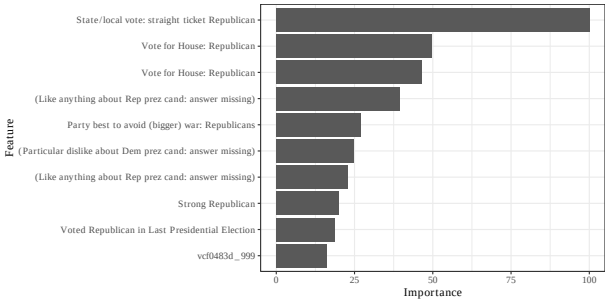
(a) ANES, 1960, Spec. 1 (Demographics Only)



(b) ANES, 1960, Spec. 2 (Demo. + PID)

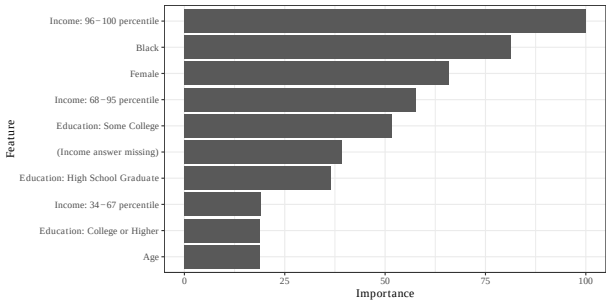


(c) ANES, 1960, Spec. 3 (Demo. + PID + Issues)

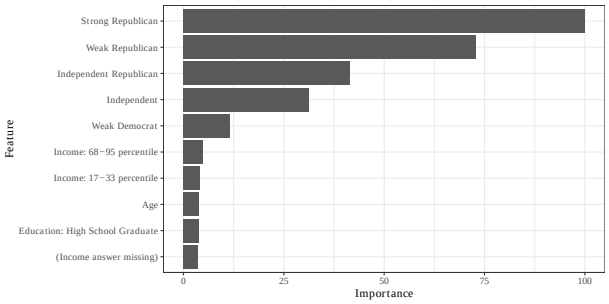


(d) ANES, 1960, Spec. 4 (All Covariates)

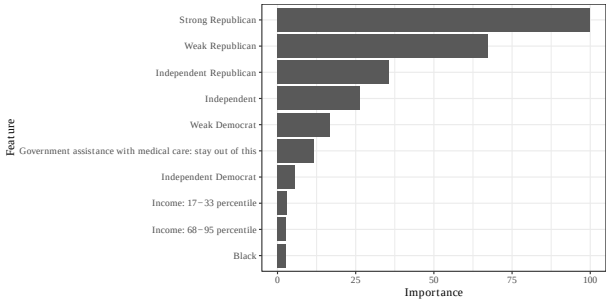
Figure A.19: ANES, Top 10 Variables from Random Forests, 1960



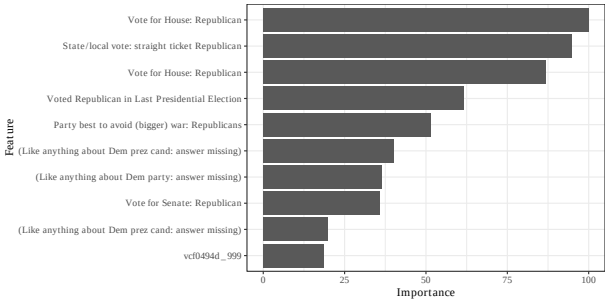
(a) ANES, 1956, Spec. 1 (Demographics Only)



(b) ANES, 1956, Spec. 2 (Demo. + PID)

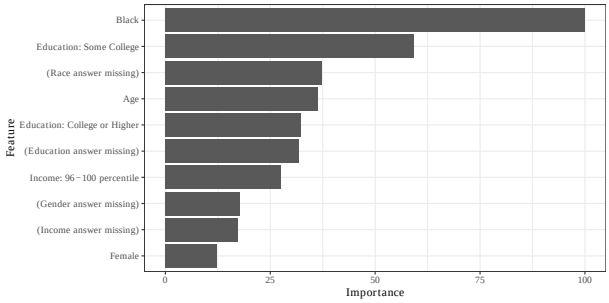


(c) ANES, 1956, Spec. 3 (Demo. + PID + Issues)

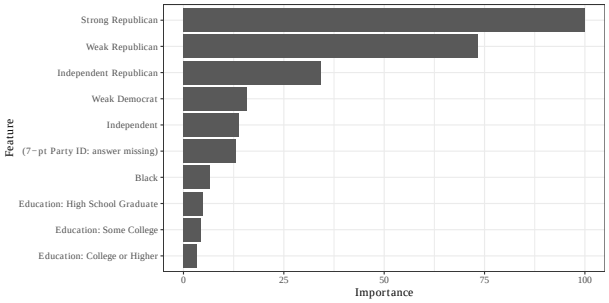


(d) ANES, 1956, Spec. 4 (All Covariates)

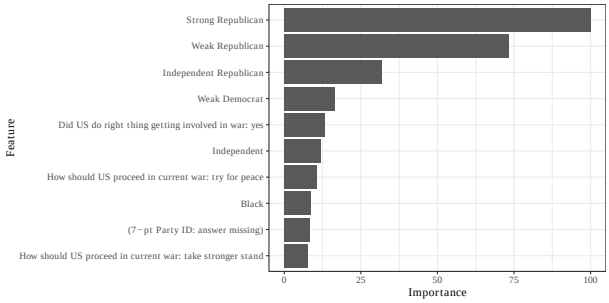
Figure A.20: ANES, Top 10 Variables from Random Forests, 1956



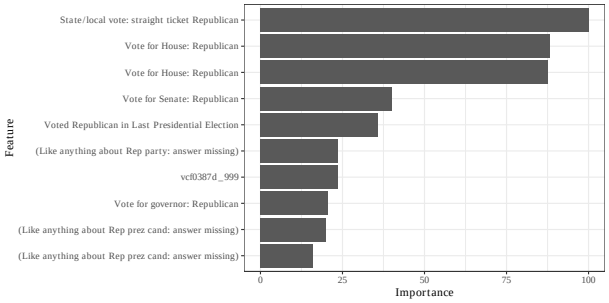
(a) ANES, 1952, Spec. 1 (Demographics Only)



(b) ANES, 1952, Spec. 2 (Demo. + PID)



(c) ANES, 1952, Spec. 3 (Demo. + PID + Issues)



(d) ANES, 1952, Spec. 4 (All Covariates)

Figure A.21: ANES, Top 10 Variables from Random Forests, 1952

Appendix B Model Outputs and Performance: CCES

B.1 Full Performance Metrics

B.1.1 Accuracy Range Comparison Between Methods

Variable Specification	Year	Logit	CART	RF
Demographics Only	2018	[0.6515, 0.6721]	[0.6292, 0.6500]	[0.6535, 0.6740]
Demo. + PID	2018	[0.9259, 0.9369]	[0.8151, 0.8317]	[0.9223, 0.9336]
Demo. + PID + Issues	2018	[0.9562, 0.9647]	[0.9260, 0.9370]	[0.9533, 0.9621]
All Covariates	2018	[0.5212, 0.5429]	[0.9299, 0.9406]	[0.9608, 0.9688]
Demographics Only	2016	[0.6442, 0.6650]	[0.6222, 0.6432]	[0.6359, 0.6567]
Demo. + PID	2016	[0.8919, 0.9051]	[0.8579, 0.8728]	[0.8892, 0.9026]
Demo. + PID + Issues	2016	[0.9242, 0.9354]	[0.8218, 0.8382]	[0.9201, 0.9316]
All Covariates	2016	[0.9471, 0.9565]	[0.8875, 0.9009]	[0.9409, 0.9509]
Demographics Only	2014	[0.6382, 0.6591]	[0.6225, 0.6435]	[0.6353, 0.6563]
Demo. + PID	2014	[0.9048, 0.9173]	[0.8155, 0.8321]	[0.9019, 0.9146]
Demo. + PID + Issues	2014	[0.9188, 0.9304]	[0.8266, 0.8429]	[0.9160, 0.9278]
All Covariates	2014	[0.9436, 0.9534]	[0.9070, 0.9193]	[0.9412, 0.9511]
Demographics Only	2012	[0.6211, 0.6430]	[0.5851, 0.6073]	[0.6268, 0.6486]
Demo. + PID	2012	[0.9095, 0.9221]	[0.8271, 0.8439]	[0.9094, 0.9220]
Demo. + PID + Issues	2012	[0.9480, 0.9577]	[0.8956, 0.9091]	[0.9446, 0.9546]
All Covariates	2012	[0.9623, 0.9706]	[0.9163, 0.9285]	[0.9619, 0.9702]
Demographics Only	2010	[0.6332, 0.6527]	[0.6136, 0.6334]	[0.6404, 0.6598]
Demo. + PID	2010	[0.9134, 0.9246]	[0.8358, 0.8506]	[0.9109, 0.9222]
Demo. + PID + Issues	2010	[0.9403, 0.9497]	[0.9035, 0.9153]	[0.9369, 0.9465]
All Covariates	2010	[0.9497, 0.9583]	[0.9332, 0.9431]	[0.9511, 0.9596]
Demographics Only	2008	[0.6233, 0.6509]	[0.5953, 0.6234]	[0.6218, 0.6495]
Demo. + PID	2008	[0.8943, 0.9114]	[0.8087, 0.8308]	[0.8903, 0.9077]
Demo. + PID + Issues	2008	[0.9362, 0.9496]	[0.8743, 0.8928]	[0.9293, 0.9434]
All Covariates	2008	[0.9740, 0.9825]	[0.9607, 0.9713]	[0.9761, 0.9842]

Table B.5: Accuracy Range Comparison, Presidential Vote Choice, CCES 2008–2018

B.1.2 Full Performance Tables

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.7221	0.6638	[0.6535, 0.6740]	0.6492	0.6129	0.6305
Demo. + PID	2018	0.9699	0.9281	[0.9223, 0.9336]	0.9262	0.9197	0.9230
Demo. + PID + Issues	2018	0.9913	0.9578	[0.9533, 0.9621]	0.9561	0.9537	0.9549
All Covariates	2018	0.9932	0.9650	[0.9608, 0.9688]	0.9675	0.9573	0.9624
Demographics Only	2016	0.7141	0.6463	[0.6359, 0.6567]	0.6329	0.5452	0.5858
Demo. + PID	2016	0.9530	0.8960	[0.8892, 0.9026]	0.8846	0.8894	0.8870
Demo. + PID + Issues	2016	0.9781	0.9260	[0.9201, 0.9316]	0.9226	0.9155	0.9190
All Covariates	2016	0.9885	0.9461	[0.9409, 0.9509]	0.9404	0.9421	0.9413
Demographics Only	2014	0.7159	0.6459	[0.6353, 0.6563]	0.6354	0.4762	0.5444
Demo. + PID	2014	0.9659	0.9084	[0.9019, 0.9146]	0.8860	0.9110	0.8983
Demo. + PID + Issues	2014	0.9743	0.9221	[0.9160, 0.9278]	0.9194	0.9038	0.9115
All Covariates	2014	0.9898	0.9463	[0.9412, 0.9511]	0.9408	0.9382	0.9395
Demographics Only	2012	0.6914	0.6378	[0.6268, 0.6486]	0.6197	0.6953	0.6554
Demo. + PID	2012	0.9697	0.9158	[0.9094, 0.9220]	0.9087	0.9228	0.9157
Demo. + PID + Issues	2012	0.9892	0.9497	[0.9446, 0.9546]	0.9537	0.9443	0.9490
All Covariates	2012	0.9950	0.9662	[0.9619, 0.9702]	0.9664	0.9654	0.9659
Demographics Only	2010	0.7105	0.6502	[0.6404, 0.6598]	0.6245	0.7506	0.6818
Demo. + PID	2010	0.9669	0.9167	[0.9109, 0.9222]	0.9149	0.9186	0.9167
Demo. + PID + Issues	2010	0.9835	0.9418	[0.9369, 0.9465]	0.9402	0.9434	0.9418
All Covariates	2010	0.9887	0.9555	[0.9511, 0.9596]	0.9493	0.9623	0.9557
Demographics Only	2008	0.6806	0.6357	[0.6218, 0.6495]	0.6871	0.4663	0.5556
Demo. + PID	2008	0.9559	0.8993	[0.8903, 0.9077]	0.8863	0.9106	0.8983
Demo. + PID + Issues	2008	0.9836	0.9366	[0.9293, 0.9434]	0.9207	0.9522	0.9362
All Covariates	2008	0.9981	0.9805	[0.9761, 0.9842]	0.9775	0.9826	0.9801

Table B.6: Performance Metrics, Presidential Vote Choice, Random Forests, CCES 2008–2018

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.7190	0.6619	[0.6515, 0.6721]	0.6328	0.6613	0.6467
Demo. + PID	2018	0.9724	0.9315	[0.9259, 0.9369]	0.9252	0.9288	0.9270
Demo. + PID + Issues	2018	0.9923	0.9606	[0.9562, 0.9647]	0.9621	0.9534	0.9577
All Covariates	2018	0.5002	0.5321	[0.5212, 0.5429]	0.6000	0.0008	0.0016
Demographics Only	2016	0.7100	0.6546	[0.6442, 0.6650]	0.6185	0.6452	0.6315
Demo. + PID	2016	0.9549	0.8986	[0.8919, 0.9051]	0.8783	0.9043	0.8911
Demo. + PID + Issues	2016	0.9804	0.9299	[0.9242, 0.9354]	0.9246	0.9224	0.9235
All Covariates	2016	0.9901	0.9519	[0.9471, 0.9565]	0.9465	0.9488	0.9477
Demographics Only	2014	0.7129	0.6487	[0.6382, 0.6591]	0.5983	0.6369	0.6170
Demo. + PID	2014	0.9679	0.9112	[0.9048, 0.9173]	0.8824	0.9232	0.9023
Demo. + PID + Issues	2014	0.9760	0.9248	[0.9188, 0.9304]	0.9138	0.9171	0.9155
All Covariates	2014	0.9885	0.9487	[0.9436, 0.9534]	0.9418	0.9426	0.9422
Demographics Only	2012	0.6876	0.6321	[0.6211, 0.6430]	0.6082	0.7228	0.6606
Demo. + PID	2012	0.9732	0.9160	[0.9095, 0.9221]	0.9051	0.9276	0.9162
Demo. + PID + Issues	2012	0.9910	0.9530	[0.9480, 0.9577]	0.9545	0.9505	0.9525
All Covariates	2012	0.9952	0.9666	[0.9623, 0.9706]	0.9669	0.9656	0.9663
Demographics Only	2010	0.7076	0.6430	[0.6332, 0.6527]	0.6147	0.7633	0.6810
Demo. + PID	2010	0.9702	0.9191	[0.9134, 0.9246]	0.9084	0.9321	0.9201
Demo. + PID + Issues	2010	0.9857	0.9451	[0.9403, 0.9497]	0.9427	0.9477	0.9452
All Covariates	2010	0.9877	0.9541	[0.9497, 0.9583]	0.9493	0.9593	0.9543
Demographics Only	2008	0.6794	0.6372	[0.6233, 0.6509]	0.6616	0.5263	0.5862
Demo. + PID	2008	0.9601	0.9031	[0.8943, 0.9114]	0.8898	0.9149	0.9022
Demo. + PID + Issues	2008	0.9859	0.9432	[0.9362, 0.9496]	0.9350	0.9496	0.9423
All Covariates	2008	0.9786	0.9786	[0.9740, 0.9825]	0.9787	0.9774	0.9781

Table B.7: Performance Metrics, Presidential Vote Choice, Logit, CCES 2008–2018

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.6624	0.6397	[0.6292, 0.6500]	0.6067	0.6543	0.6296
Demo. + PID	2018	0.8136	0.8235	[0.8151, 0.8317]	0.9740	0.6401	0.7725
Demo. + PID + Issues	2018	0.9426	0.9316	[0.9260, 0.9370]	0.9220	0.9329	0.9274
All Covariates	2018	0.9402	0.9354	[0.9299, 0.9406]	0.9313	0.9306	0.9310
Demographics Only	2016	0.6598	0.6328	[0.6222, 0.6432]	0.5815	0.7113	0.6399
Demo. + PID	2016	0.8640	0.8655	[0.8579, 0.8728]	0.9367	0.7579	0.8379
Demo. + PID + Issues	2016	0.8265	0.8301	[0.8218, 0.8382]	0.8702	0.7401	0.7999
All Covariates	2016	0.9008	0.8943	[0.8875, 0.9009]	0.8731	0.9006	0.8866
Demographics Only	2014	0.6800	0.6331	[0.6225, 0.6435]	0.6059	0.4981	0.5467
Demo. + PID	2014	0.8063	0.8239	[0.8155, 0.8321]	0.9603	0.6297	0.7606
Demo. + PID + Issues	2014	0.8373	0.8349	[0.8266, 0.8429]	0.8045	0.8301	0.8171
All Covariates	2014	0.9216	0.9133	[0.9070, 0.9193]	0.8980	0.9080	0.9030
Demographics Only	2012	0.6083	0.5962	[0.5851, 0.6073]	0.5600	0.8628	0.6792
Demo. + PID	2012	0.8358	0.8356	[0.8271, 0.8439]	0.9794	0.6826	0.8045
Demo. + PID + Issues	2012	0.9028	0.9025	[0.8956, 0.9091]	0.9157	0.8847	0.8999
All Covariates	2012	0.9322	0.9226	[0.9163, 0.9285]	0.8994	0.9499	0.9240
Demographics Only	2010	0.6318	0.6235	[0.6136, 0.6334]	0.5790	0.9015	0.7051
Demo. + PID	2010	0.8463	0.8433	[0.8358, 0.8506]	0.9654	0.7117	0.8193
Demo. + PID + Issues	2010	0.9090	0.9095	[0.9035, 0.9153]	0.8921	0.9314	0.9113
All Covariates	2010	0.9470	0.9383	[0.9332, 0.9431]	0.9273	0.9509	0.9390
Demographics Only	2008	0.6108	0.6094	[0.5953, 0.6234]	0.7094	0.3391	0.4589
Demo. + PID	2008	0.8257	0.8200	[0.8087, 0.8308]	0.7382	0.9783	0.8415
Demo. + PID + Issues	2008	0.8854	0.8838	[0.8743, 0.8928]	0.8560	0.9162	0.8851
All Covariates	2008	0.9676	0.9663	[0.9607, 0.9713]	0.9673	0.9635	0.9654

Table B.8: Performance Metrics, Presidential Vote Choice, CART, CCES 2008–2018

B.1.3 ROC Curves

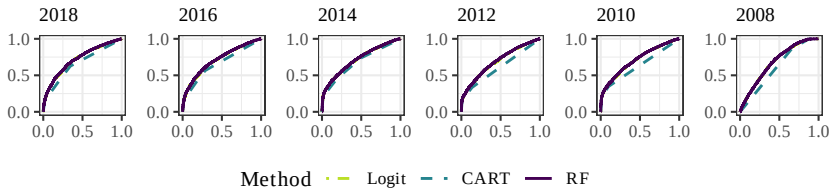


Figure B.22: ROC Curves for Presidential Vote Choice, Demographics Only, Comparison Between Logit, CART, and RF

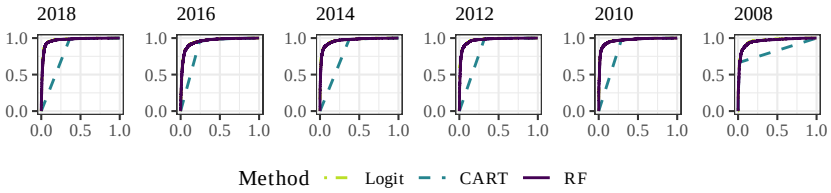


Figure B.23: ROC Curves for Presidential Vote Choice, Demo. + PID, Comparison Between Logit, CART, and RF

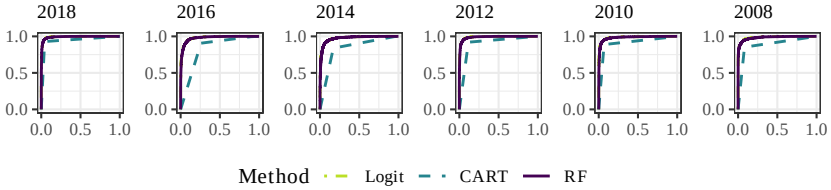


Figure B.24: ROC Curves for Presidential Vote Choice, Demo. + PID + Issues, Comparison Between Logit, CART, and RF

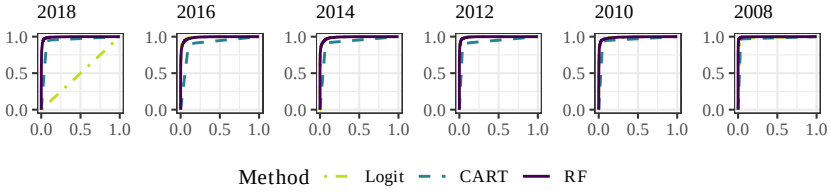
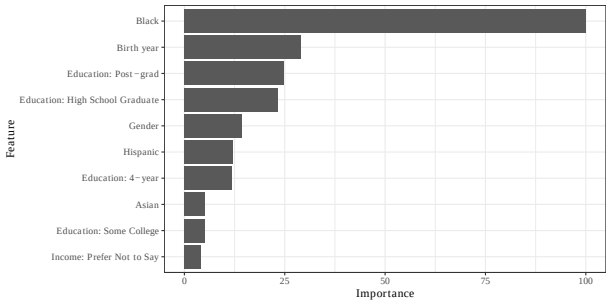


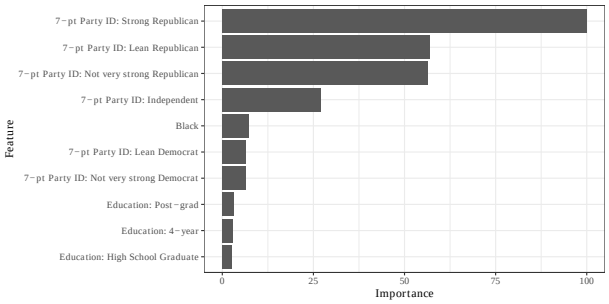
Figure B.25: ROC Curves for Presidential Vote Choice, All Covariates, Comparison Between Logit, CART, and RF

B.2 Variable Importance Measures

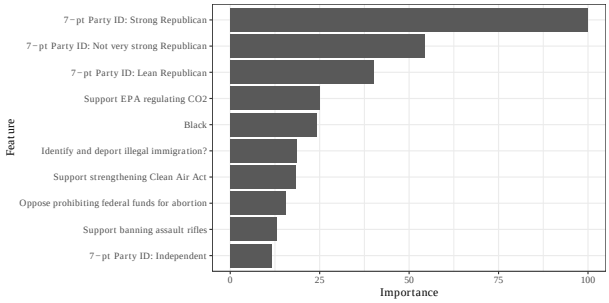
The following are, again, top 10 variables per specification and survey wave, visualized.



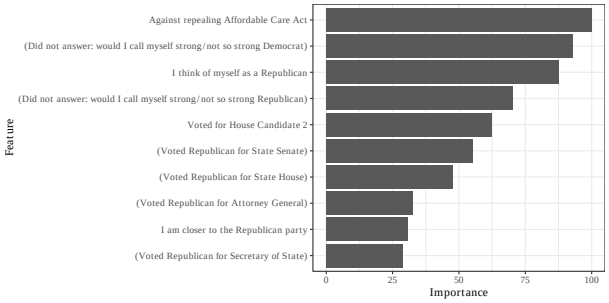
(a) CCES, 2016, Spec. 1 (Demographics Only)



(b) CCES, 2016, Spec. 2 (Demo. + PID)

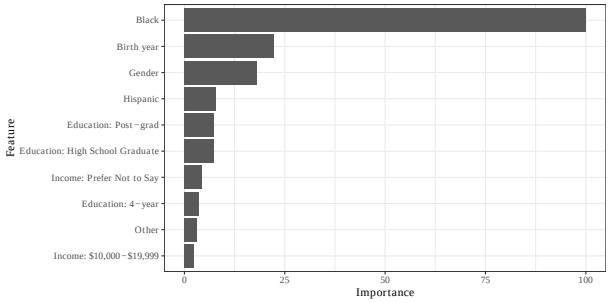


(c) CCES, 2016, Spec. 3 (Demo. + PID + Issues)

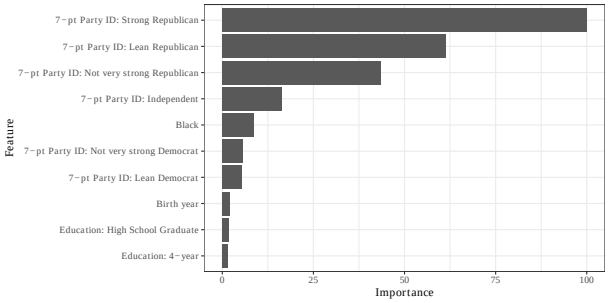


(d) CCES, 2016, Spec. 4 (All Covariates)

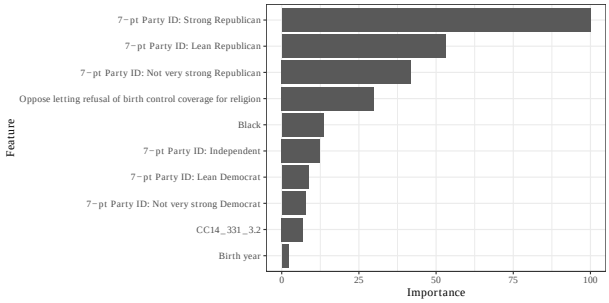
Figure B.26: CCES, Top 10 Variables from Random Forests, 2016



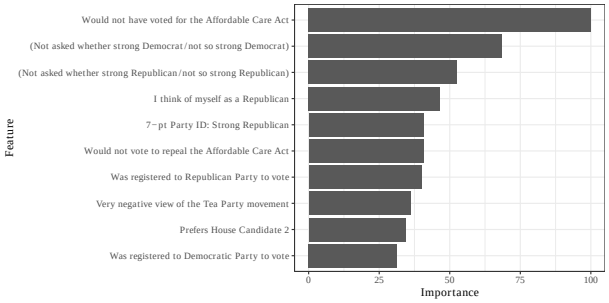
(a) CCES, 2014, Spec. 1 (Demographics Only)



(b) CCES, 2014, Spec. 2 (Demo. + PID)

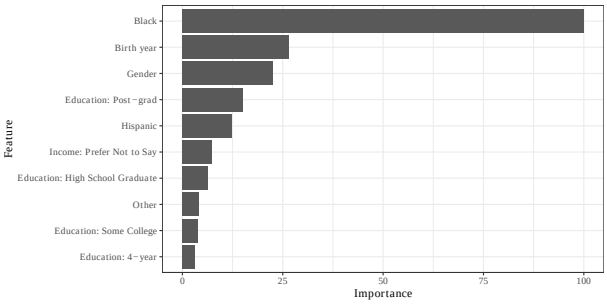


(c) CCES, 2014, Spec. 3 (Demo. + PID + Issues)

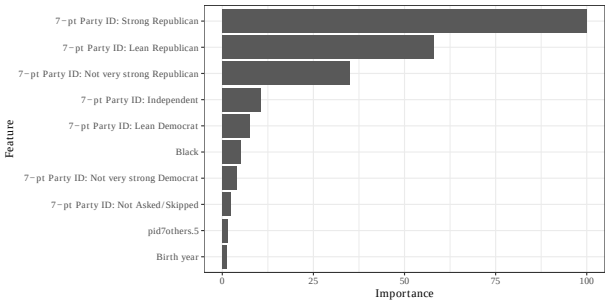


(d) CCES, 2014, Spec. 4 (All Covariates)

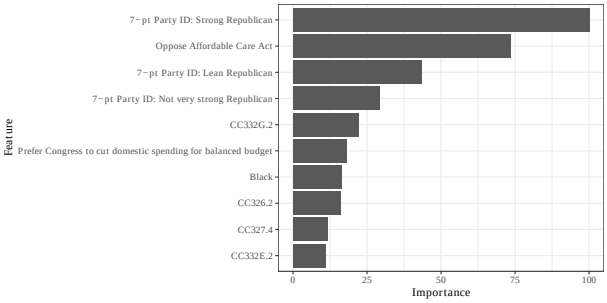
Figure B.27: CCES, Top 10 Variables from Random Forests, 2014



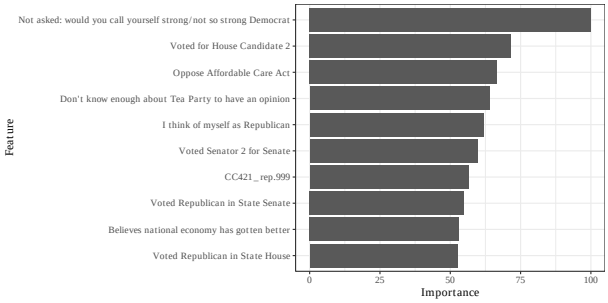
(a) CCES, 2012, Spec. 1 (Demographics Only)



(b) CCES, 2012, Spec. 2 (Demo. + PID)

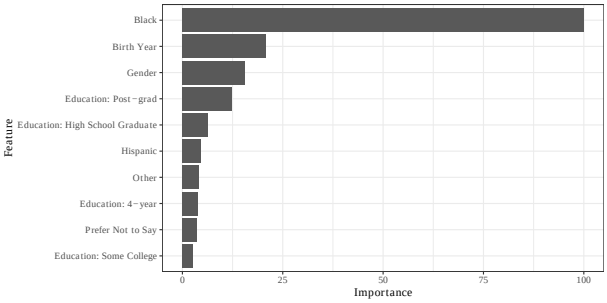


(c) CCES, 2012, Spec. 3 (Demo. + PID + Issues)

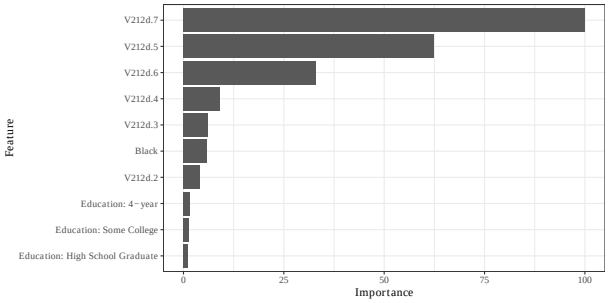


(d) CCES, 2012, Spec. 4 (All Covariates)

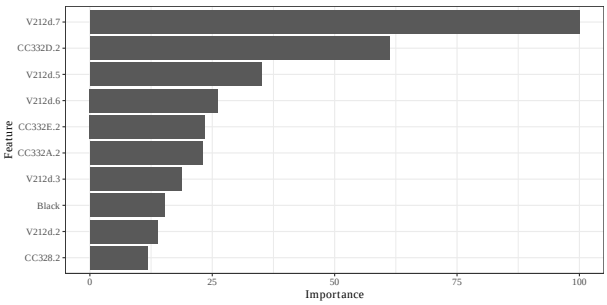
Figure B.28: CCES, Top 10 Variables from Random Forests, 2012



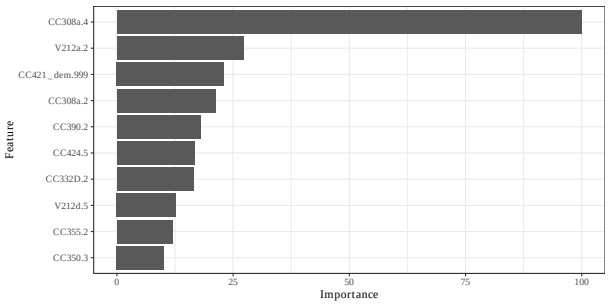
(a) CCES, 2010, Spec. 1 (Demographics Only)



(b) CCES, 2010, Spec. 2 (Demo. + PID)

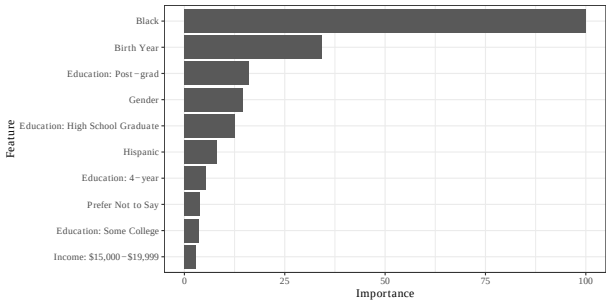


(c) CCES, 2010, Spec. 3 (Demo. + PID + Issues)

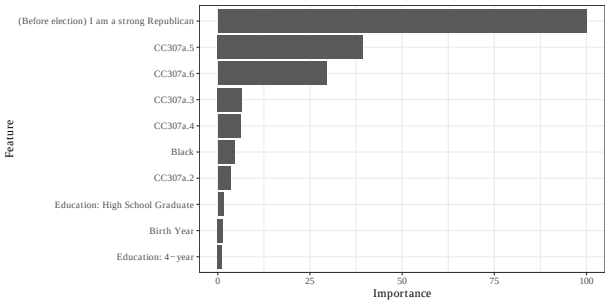


(d) CCES, 2010, Spec. 4 (All Covariates)

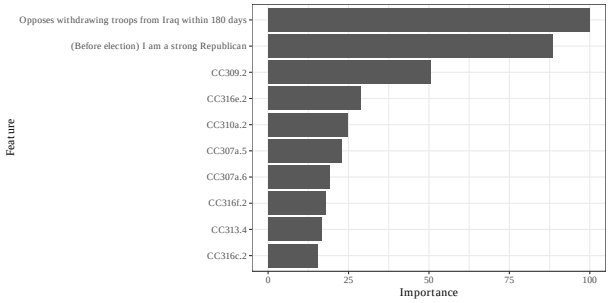
Figure B.29: CCES, Top 10 Variables from Random Forests, 2010



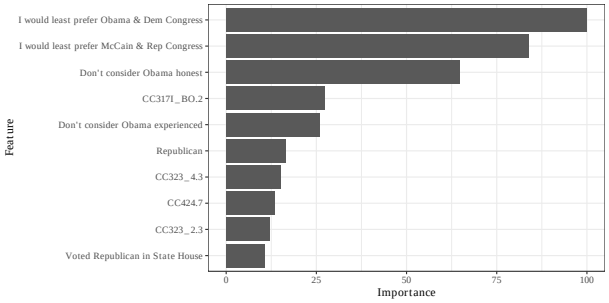
(a) CCES, 2008, Spec. 1 (Demographics Only)



(b) CCES, 2008, Spec. 2 (Demo. + PID)



(c) CCES, 2008, Spec. 3 (Demo. + PID + Issues)



(d) CCES, 2008, Spec. 4 (All Covariates)

Figure B.30: CCES, Top 10 Variables from Random Forests, 2008

Appendix C Recent Presidential Elections: Descriptive Evidence

C.1 Education, Race, and Vote Choice between 2008 and 2020

		Republican	Democrat
2016 election		Donald Trump	Hillary Clinton
No HS	Voters of color	15.2	77.9
No HS	White voters	62.6	32.1
High school graduate	Voters of color	20.5	75.8
High school graduate	White voters	64.1	31.2
Some college	Voters of color	19.4	72.8
Some college	White voters	52.1	41.1
2-year	Voters of color	21.0	73.3
2-year	White voters	53.2	39.0
4-year	Voters of color	24.0	69.6
4-year	White voters	41.7	50.3
Post-grad	Voters of color	21.6	72.8
Post-grad	White voters	34.5	58.9
2012 election		Mitt Romney	Barack Obama
No HS	Voters of color	17.8	80.4
No HS	White voters	52.1	46.2
High school graduate	Voters of color	16.8	82.7
High school graduate	White voters	57.7	40.7
Some college	Voters of color	24.4	73.7
Some college	White voters	53.0	44.3
2-year	Voters of color	26.1	72.3
2-year	White voters	53.8	43.4
4-year	Voters of color	25.0	73.9
4-year	White voters	48.3	49.5
Post-grad	Voters of color	29.5	68.9
Post-grad	White voters	40.1	58.1
2008 election		John McCain	Barack Obama
No HS	Voters of color	21.2	67.1
No HS	White voters	45.2	43.0
High school graduate	Voters of color	16.9	73.7
High school graduate	White voters	53.9	34.1
Some college	Voters of color	19.2	72.4
Some college	White voters	51.0	39.5
2-year	Voters of color	24.9	66.7
2-year	White voters	54.3	35.3
4-year	Voters of color	26.5	67.0
4-year	White voters	47.6	44.3
Post-grad	Voters of color	20.4	75.4
Post-grad	White voters	40.9	53.6

Table C.9: Vote choice by race and education among verified voters (authors' calculations based on the CCES data)

C.2 2016 vote choice

C.2.1 Evidence from Manually Specified Logits

Using Nationscape data, we verified the presence of a large number of seemingly important predictors of vote choice in saturated models. When we modelled when 2016 recalled vote choice as a function of 7 social issues, 8 economic issues, and 4 immigration-related issues (including party ID dummies and demogrpahics), then 16 out of the 19 issue coefficients as well as 20 out 28 coefficients tapping into group attitudes were statistically significant. But the inclusion of most of the se statistically significant variables yielded no improvement in accuracy (in-sample or out-of-sample). Accordingly, we argue that variable inclusion should not hinge on statistical significance.

After adding racial resentment and sexist attitudes as two separate scales based on graded-response IRT models to a model with demographc variables, party ID and a rich set of issue positions, the proportion of correctly classified respondents increased by only 0.1 percentage points. Moreover, had we included all race-related and gender-related question as individual items rather than as scales, even in-sample accuracy would nudge upwaard by only 0.1 percentage points. More worrying, such a model would include 62 coefficients, 51 of which would be statistcally at the 5% level (50 of them at the 1% level).

C.2.2 Evidence from Regression Trees

Constraining the number of terminal nodes to just 8 categories, we obtain a set of simple decision rules chosen out of a set of 144 features. After a tree that we have constrained in a way to balance the twin goals of simplicity and accuracy is grown, we see that the most useful voter attributes consist of *one policy response* (which is the most important piece of data), attitudes towards Republicans and Democrats, and respondents' own party ID. This simple model is displayed in Figure C.31. The average terminal node contains 8,120 respondents.[1]

Surprisingly, we see that for nearly 45% of cases *no information about voters' own partisanship is required*. For example, if a voter opposed building the wall on the U.S.-Mexico border, and also held an unfavorable view of Republicans, then those are the two sufficient attributes to predict with a high degree of confidence that a respondent voted for Clinton.[2]

After tuning for the optimal complexity parameter via 5-fold cross-validation, repeated 15 times, we obtain a more complex tree where predictions involve up to six steps (although many of the edges involve a gradual move along the PID-7 scale, which we one-hot encoded before feeding it into the model). The tuned tree is displayed in Figure C.32.

One terminal node, containing a large proportion (34%) of all predictions is related to the one we encountered previously, but it involves one additional step. In addition to opposing the wall, and holding unfavorable view of Republicans, this richer tree suggests also splitting respondents by the absence of unfavorable view of Democrats. For respondents who meet these three conditions (on the left-most path of the tree in Figure C.32), the predicted probability of the Clinton vote is over 96%. We not that for this set of "almost-surely Clinton voters", information about partisanship was not used, and the 34% of the sample in the training set (22,089 voters) fell into this category.

[1] There are 16,368 respondents in the holdout set who were not used to estimate the optimal splitting rules.

[2] Specifically, the tree yields a 92.9% probability that a respondent has voted for Clinton.

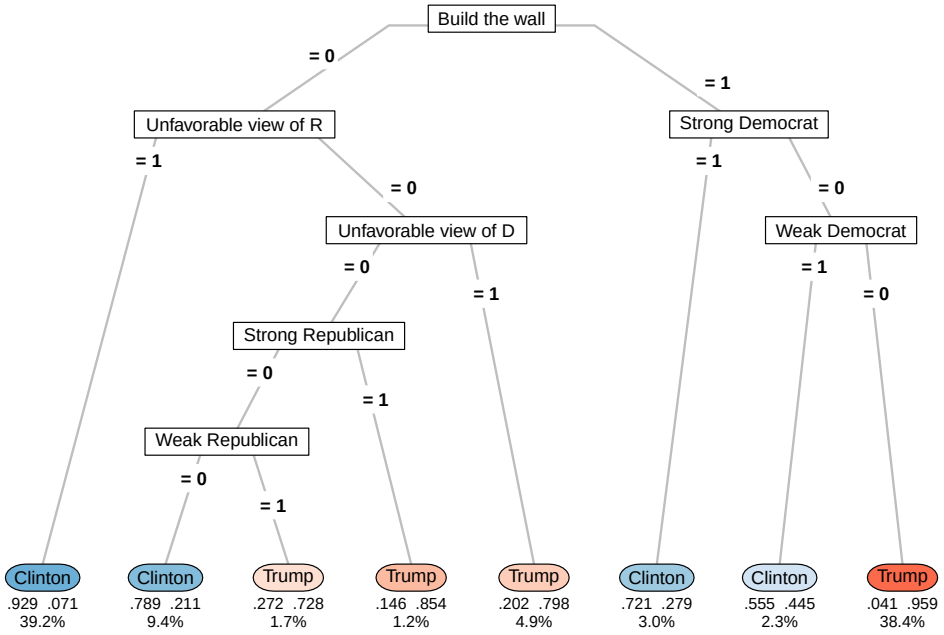


Figure C.31: A concise classification tree (Y=Trump vote or Clinton vote) estimated on the Nationscape data. The set of respondents is restricted to those who voted for one of the major-party nominees in 2016.

Conversely, on the right-hand side of the tree, we see that voters who 1) supported building the wall and 2) were not Democrats, have an estimated 95.8% probability of supporting Trump. Combining these and the aforementioned predictions (i.e. people who are almost surely Clinton voters and the 25,094 almost-surely Trump voters) tells us that decisions of 72.6% of voters can be predicted with a very high degree of confidence.

We also note that the algorithm uncovers small, but potentially interesting groups of defectors. For example, we see on the left side of the tree that there both Republicans (in all categories: strong, weak, and leaners) who opposed building a wall and also hold unfavorable views of both Democrats and Republicans. Defection rates (i.e. propensity to vote for Clinton in 2016) among these voters ranges from 18.5% (for strong Republicans) to 32.2% for Republican leaners. While only 1.3% of voters fall into these categories, their decisions are difficult to predict.

Similarly, given the construction of the tree, the model has difficulty predicting decisions of people who are 1) against building the border wall, 2) who dislike both Democrats and Republicans, and 3) who identify as non-Republicans (i.e. this set includes Independents and Democrats of all shades). A non-negligible number of voters (4% in the sample) meet these criteria. The model predicts that these voters are more likely to voter for Clinton (Pr(Clinton)=83.5%) but it makes a mistake in nearly one in five cases.

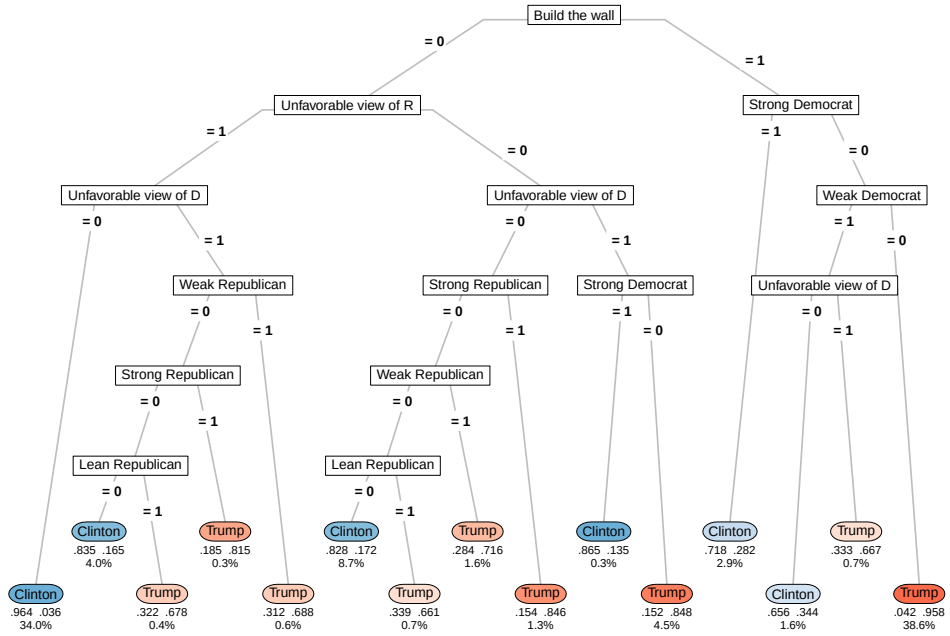


Figure C.32: A tuned classification tree estimated on the Nationscape data. The set of respondents is restricted to those who voted for Clinton or Trump in 2016.

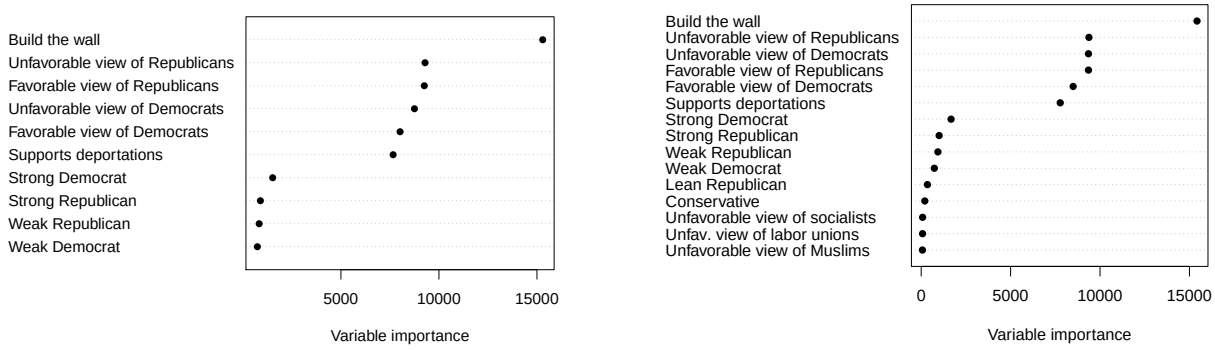


Figure C.33: Most important variables identified by the simple tree algorithm (left panel) and the tuned classification tree (right panel).