Decision Trees & Random Forests

BUGS Meeting

Daniel Pimentel-Alarcón Computer Science, GSU

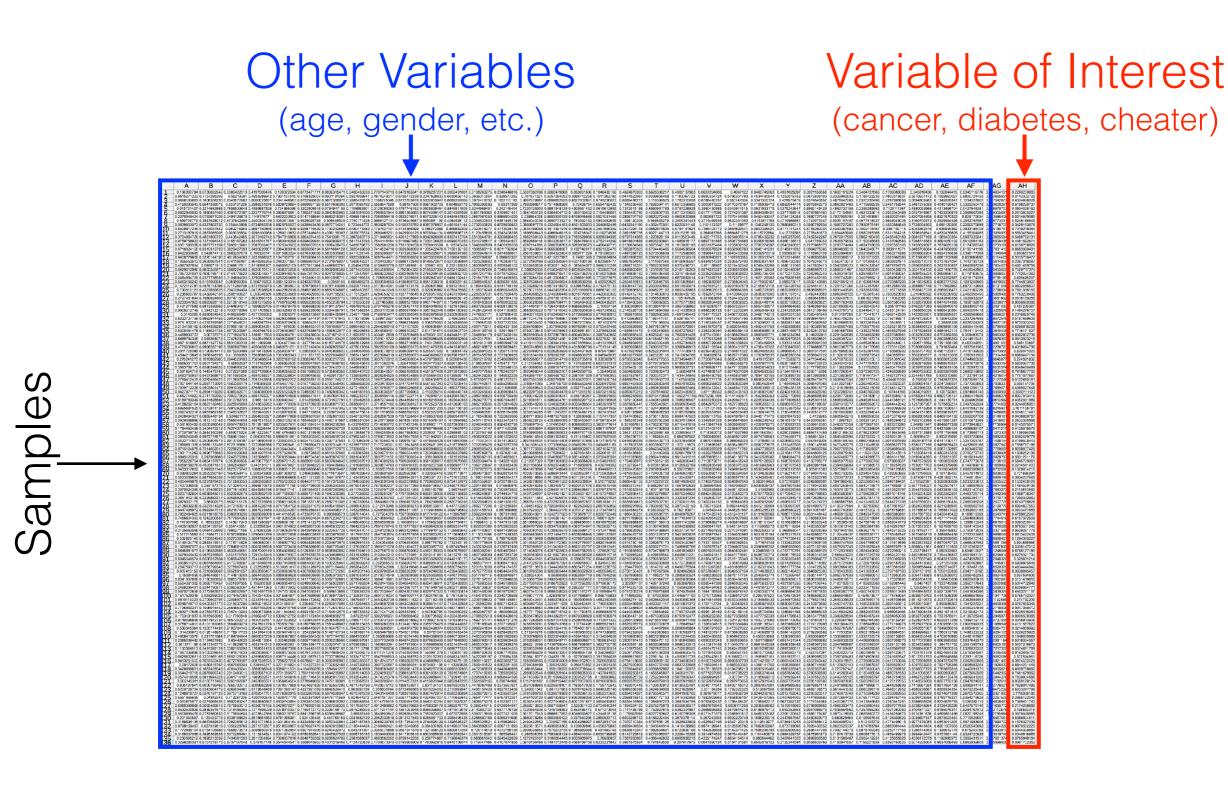
Decision Trees

Goal: Predict

- Will I get El Cáncer?
- Will I develop Diabetes?
- Is my boyfriend/girlfriend cheating on me?
- Will my Bacteria develop Antibiotic Resistance?

 0.166390734 0.0730022542 0.0380422519 0.4197909476 0.199372394 0.8773471711 0.0832315471 0.9001134394 0.59276039 0.9815271223 0.757295239 0.65219381 0.46528942 0.751295632 0.9870366976 0.9870366976 0.987036976 0.987036976 0.987036976 0.987036976 0.987036976 0.987036976 0.955433607 0.015721427 0.3214682985 0.7856533643 0.455887529 0.72496396 0.3202909518 0.0764947159 0.11588829 0.632840033 0.182837403 0.4041572187 0.2237776744 0.958876091 0.7356271892 0.0843965665 0.3547628012 0.6737842499 0.4400304139 0.3052513081 0.9529705779 0.8527605075 0.754443808 0.6734693338 0.8825760537 0.5428943436 0.087754241 0.7524447 0.8632917245 0.44351842 0.669216933 0.997150856 0.9304477939 0.925251386 0.982705520 0.93035834 0.999150856 0.9364477939 0.926251382 0.7574443 0.863215391 0.443642465 0.9373102742 0.787310966 0.3864274416 0.7541396178 0.28207837 0.3704964726 0.485160723 0.443642465 0.337102742 0.787310966 0.3864274416 0.7541396178 0.386756313 0.3973102742 0.78733399 0.3977721159 0.5993215	
 0.9898183863 0.1602356235 0.8343072962 0.6000025657 0.2041446567 0.8752995632 0.9870368976 0.95854293 0.4150293543 0.6541336076 0.533131226 0.258831876 0.558729884 0.599851159 0.577696308 0.60710657 0.015721427 0.3214682985 0.7856533648 0.4555887529 0.72496398 0.3202909518 0.0754947159 0.11588829 0.6832840093 0.1836837403 0.404572187 0.2277948424 0.0736671025 0.6941126878 0.149176771 0.5402233903 0.9147158544 0.8683215391 0.40668158 0.569706779 0.3547628012 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.775017144 0.639803573 0.6327802755 0.754443008 0.673469338 0.882760570 0.90337842 0.969216953 0.9897150860 0.936417799 0.9258715529 0.173232688 0.4800723 0.443642485 0.9373102742 0.7847810966 0.3864274416 0.7544396178 0.36672837 0.9759758923 0.3771094345 0.495157282 0.3373102742 0.7847810966 0.3864274416 0.7541396178 0.38672837 0.9759758923 0.3771094345 0.495159230 0.4320220807 0.286490316 0.3835056194 0.5811884128 0.3820427393 0.8420220807 0.286802645431 0.493792754	3 0.7
 0.4160293543 0.6541336075 0.533131226 0.2588331876 0.5838729884 0.5999851159 0.5517696308 0.60710657 0.015721427 0.3214682985 0.7856533648 0.4555887529 0.724966398 0.3202909518 0.0764947159 0.11588299 0.6832840093 0.1885837403 0.4081572187 0.2237776744 0.3958876091 0.7365271892 0.0843965665 0.35810916 0.2270948424 0.0753671026 0.0494126878 0.1149176771 0.540223903 0.9147168584 0.8683215391 0.49568158 0.569706779 0.3547628021 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.77501714 0.0639803573 0.8327802755 0.754443808 0.6734893338 0.8825760537 0.5428943436 0.0870754241 0.7254447 0.8509917245 0.1453037842 0.969216953 0.9897150856 0.9354477999 0.9258715529 0.1733232688 0.64804733 0.2713187919 0.2857065005 0.903035834 0.9298100968 0.1260013632 0.0767446451 0.4399159383 0.28290763 0.3704964725 0.4851600723 0.443642485 0.9373102742 0.7847310966 0.3864274416 0.7541396178 0.38672837 0.9759758923 0.737109436 0.0455157282 0.8346701767 0.4383346298 0.6288020934 0.0993899375 0.58173437 0.6351593339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.266007051 0.166648479 0.64030798 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 0.8422852686 0.542874712 0.5883257678 0.763120741 0.831582224 0.6276330658 0.5805358685 0.84110281 0.4908779605 0.5220144195 0.4812854063 0.5331686522 0.104735271 0.7975936416 0.8026251938 0.99253836 0.4908779624 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.449663761 0.2549437 0.4999727954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4496663761 0.2549437 0.4949372924 0.0960355875 0.5060224588 0.9520570131 0.1408375208 0.1400302523 0.9902096502 0.00714512 0.795724041 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8857347907 0.924318885 0.47125479<!--</th--><th>1 0.4</th>	1 0.4
 0.015721427 0.3214682985 0.7856533648 0.4555887529 0.724996398 0.3202909515 0.0764947159 0.11588829 0.6832840093 0.185837403 0.4081572187 D.2237776744 0.3958876091 0.7385271892 0.0843965655 0.35810916 0.529705779 0.3547628012 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.77547114 0.0639803575 0.8327802755 0.754443808 0.969216953 0.929810958 0.935471032 0.767446451 0.439159383 0.825760537 0.5428943436 0.80870754241 0.7254447 0.8509917245 0.1453037842 0.969216953 0.929170656 0.93035834 0.9298100968 0.126013322 0.767446451 0.4392159383 0.826760537 0.5428943436 0.8870754241 0.75444805 0.93035834 0.9298100968 0.126013322 0.767446451 0.4392159833 0.282707533 0.377049445 0.443642485 0.9337102742 0.7847310966 0.364274416 0.7541396178 0.38672837 1.0443642485 0.93376102767 0.438346298 0.6298020934 0.99389375 0.58173437 0.6315312733 0.3823427393 0.6420220807 0.3835056194 0.5611884128 0.38204336 0.89206341 0.9933935256778 0.760320741 0.83582224 0.6276306056 0.5803535865 	6 0.1
6 0.6832840093 0.1885837403 0.4081572187 0.2237776744 0.3958376091 0.7385271892 0.0843965665 0.35810916 7 0.2270948424 0.0753671026 0.0494126878 0.1149176771 0.5402233903 0.9147168584 0.8683215391 0.49668158 0.0539803573 0.8327802755 0.754443808 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.77501714 0.0639803573 0.8327802755 0.754443808 0.67374849338 0.8825760537 0.5428934345 0.0870754241 0.725447 0.05391724 0.1453037842 0.969216953 0.9987150856 0.9364477939 0.9258715529 0.1733232688 0.64804733 0.21713187919 0.285706500 0.903035834 0.9298100968 0.1260013632 0.0767446451 0.4399159383 0.28290763 0.37719494728 0.4851600723 0.443642485 0.9373102742 0.7847310966 0.364274416 0.7541396178 0.30672037 0.435159339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 <th>20.5</th>	20.5
0.2270948424 0.0753671026 0.0494126878 0.1149176771 0.5402233903 0.9147158584 0.8683215391 0.49566158 0.0539705779 0.3547628012 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.77501714 0.0639803575 0.8327802755 0.754443808 0.6734693338 0.8825760537 0.5428943436 0.8607754241 0.7254447 0.8509917245 0.1453037842 0.969216953 0.969715056 0.93035834 0.9298100958 0.1260013632 0.0767446451 0.4399159383 0.282907637 0.2713187919 0.285765050 0.903035834 0.9298100958 0.1260013632 0.0767446451 0.4399159383 0.28290763 0.3704964726 0.48516500723 0.44362485 0.9373102742 0.7847810966 0.3864274416 0.7541396178 0.38672837 14 0.635159339 0.9977721159 0.5993215961 0.2587002467 0.1375043502 0.266007051 0.1666468479 0.84030786 0.8422852665 0.5428174712 0.5883257678 0.760320741 0.8315882224 0.6276	1 0.3
8 0.559706779 0.3547628012 0.6737842499 0.4400304139 0.3052613081 0.9529025748 0.908458523 0.77501714 9 0.0639803573 0.8327802755 0.754443808 0.6734893338 0.8825760537 0.5428943436 0.0870754241 0.7254447 10 0.8509917245 0.1453037842 0.969216953 0.9987150856 0.9354477999 0.9258715526 0.1733232688 0.64804733 0.2713187919 0.2857065005 0.903035834 0.9298100988 0.1260013632 0.0767446451 0.4399159383 0.28290763 0.3704964726 0.4851600723 0.44362465 0.9373102742 0.7847310966 0.386427416 0.7541396178 0.38672837 13 0.9759758923 0.7371094345 0.0455157282 0.8346701767 0.4383346298 0.6298020934 0.093899375 0.58672837 14 0.935159339 0.3777721159 0.599215961 0.258702467 0.13750436502 0.266007051 0.16664879 0.84027336 15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0	6 0.7
0 0.0639803573 0.8327802755 0.754443808 0.6734893338 0.8825760537 0.5428943436 0.0870754241 0.7254447 10 0.8509917245 0.1453037842 0.969216953 0.9897150856 0.9354477939 0.9258715529 0.1733232688 0.64804733 11 0.2713187919 0.2857065005 0.903035834 0.9298100968 0.1260013632 0.0767446451 0.4399159383 0.28290763 12 0.3704964726 0.4851600723 0.443642485 0.9373102742 0.7847810966 0.3864274416 0.7541396178 0.38672837 13 0.9759758923 0.7371094345 0.0455157282 0.834670767 0.4383846288 0.6298020934 0.0993899375 0.58173437 14 0.6351593335 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660070751 0.1665468479 0.8400798 0.0716924155 0.6319312733 0.382327678 0.760320741 0.831558224 0.6276306056 0.5803535685 0.84110281 14 0.9408779605 0.5220144195 0.4812854063 0.533	3 0.7
10 0.8509917245 0.1453037842 0.969216953 0.9897150856 0.9354477939 0.9258715529 0.1733232688 0.64804733 11 0.2713187919 0.2857065005 0.903035834 0.9298100968 0.1260013632 0.0767446451 0.4399159383 0.28290763 12 0.3704964726 0.4851600723 0.443642485 0.9373102742 0.7847310966 0.3864274416 0.7541396178 0.38672837 13 0.9759758923 0.7371094345 0.0455157282 0.8346701767 0.4383346298 0.6298020934 0.0993899375 0.58173437 14 0.635159339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 0.1666468479 0.8400798 15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 16 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315382224 0.6276306058 0.5803535685 0.84110281 17 0.9408779606 0.5220144195 0.481265406	8 0.9
1 0.2713187919 0.2857065005 0.903035834 0.9298100988 0.1260013632 0.0767446451 0.4399159383 0.28290763 2 0.3704964726 0.4851600723 0.443642485 0.9373102742 0.7847810966 0.3864274416 0.7541396178 0.38672837 3 0.9759758923 0.7371094345 0.0455157282 0.8346701767 0.4383346298 0.6298020934 0.0993899375 0.58173437 4 0.6351593339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 0.1668468479 0.64030798 5 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 6 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315882224 0.6276306056 0.5803535685 0.84110281 7 0.9408779606 0.5220144195 0.4812654063 0.5331668522 0.1047652791 0.7975936416 0.8026251936 0.99253866 8 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927	2 0.9
12 0.3704964726 0.4851600723 0.443642485 0.9373102742 0.7847810966 0.3864274416 0.7541396176 0.38672837 13 0.9759758923 0.7371094345 0.0455157282 0.8346701767 0.4383846298 0.6298020934 0.0993899375 0.58173437 14 0.6351593339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 0.1668468479 0.64030798 15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.382043366 16 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315882224 0.6276306056 0.5803535685 0.84110281 17 0.9408779606 0.5220144195 0.4812854063 0.5331868522 0.1047852791 0.7975936416 0.8028251936 0.99253866 18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.159364622 19 0.4997927954 0.096071915 0.2369	2 0.1
13 0.9759758923 0.7371094346 0.0455157282 0.8346701767 0.4383346298 0.6298020934 0.0993899375 0.58173437 14 0.6351593339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 0.1668468479 0.84030796 15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 16 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315882224 0.6276306056 0.5803535685 0.84110281 17 0.9408779606 0.5220144195 0.4812654063 0.5331868522 0.1047352791 0.7975936416 0.8026251936 0.99253896 18 0.109553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.159364622 19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8003355875 0.506692	
14 0.6351593339 0.3977721159 0.5993215961 0.2587002467 0.1375043502 0.2660007051 0.1666468479 0.84030796 15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 16 0.8422852635 0.542874712 0.5883257678 0.760320741 0.8315382224 0.6276306056 0.5803535685 0.84110281 17 0.9408779605 0.5220144195 0.4812654063 0.5331868522 0.1047852791 0.7975936416 0.8026251936 0.99253806 18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.15936462 19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8003355875 0.5868224588 0.9520570131 0.1408375208 0.1400332523 0.9902096502 0.00714512 21 0.7057224081 0.4630195114 0.914477	3 0.0
15 0.0716924155 0.6319312733 0.3823427393 0.6420220807 0.2364369316 0.3835056194 0.5611884128 0.38204336 16 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315882224 0.6276306058 0.5803535685 0.84110281 17 0.9408779608 0.5220144195 0.4812654063 0.5331668522 0.1047852791 0.7975936418 0.8028251938 0.99253866 18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.15936462 19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8003355875 0.5068224588 0.9520570131 0.1408375208 0.1400382523 0.9902096502 0.00714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820445677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.310489	3 0.7
16 0.8422852685 0.542874712 0.5883257678 0.760320741 0.8315882224 0.6276306056 0.5803535685 0.84110281 17 0.9408779606 0.5220144195 0.4812854063 0.5331868522 0.1047852791 0.7975936416 0.8028251936 0.99253866 18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.15936462 19 0.4997927954 0.096071915 0.2369617911 0.8118998307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8003355875 0.5060224588 0.9520570131 0.1408375208 0.1400302523 0.9902096502 0.00714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912480153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890	5 0.0
17 0.9408779608 0.5220144195 0.4812654063 0.5331868522 0.1047852791 0.7975936416 0.8028251936 0.99253866 18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.15936462 19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8803355875 0.5868224588 0.9520570131 0.1408375208 0.1400382523 0.9902096502 0.08714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912460153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9955105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708	8 0.7
18 0.1095553513 0.2625067541 0.9741186467 0.9848365255 0.2483271887 0.0985997624 0.2721076927 0.15936462 19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8803355875 0.5868224588 0.9520570131 0.1408375208 0.1400382523 0.9902096502 0.08714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912480153 0.8910088711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9956105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639828 0.7940384434<	5 0.3
19 0.4997927954 0.096071915 0.2369617911 0.8118898307 0.6460952105 0.8707811364 0.4498663761 0.2549437 20 0.6429012946 0.8003355875 0.5868224588 0.9520570131 0.1408375208 0.1400382523 0.9902096502 0.08714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912480153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9956105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639828 0.7940384434 0.0472936628 0.444262018 0.2649206629 0.6867534886 0.14634503	4 0.9
20 0.6429012946 0.8003355875 0.5868224588 0.9520570131 0.1408375208 0.1400382523 0.9902096502 0.08714512 21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912480153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9955105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442662018 0.2649206629 0.6867534886 0.14634503	1 0.0
21 0.7057224081 0.4630195114 0.9144776522 0.3820045677 0.6039446075 0.8557347907 0.9241318835 0.47125479 22 0.2484984729 0.2377738538 0.3104894005 0.3790386592 0.5252200689 0.1912480153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9955105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442662018 0.2649206629 0.6867534886 0.14634503	8 0.4
22 0.2484984729 0.2377738536 0.3104894005 0.3790386592 0.5252200689 0.1912460153 0.8910068711 0.05051955 23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9956105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442662018 0.2649206629 0.6867534886 0.14634503	8 D. 8
23 0.5463155243 0.5161090947 0.086890304 0.506719203 0.0661523314 0.0243307806 0.9956105 0.6854052 24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442862018 0.2649206629 0.6867534886 0.14634503	
24 0.1472119791 0.4978118786 0.3711289708 0.8731593913 0.1267880907 0.1878780541 0.9781146096 0.55205774 25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442862018 0.2649206629 0.6867534886 0.14634503	-
25 0.790510121 0.4327639826 0.7940384434 0.0472936628 0.4442862018 0.2649206629 0.6867534886 0.14634503	
	-
76 0 238609707 0 3432042131 0 0508109927 0 053046589 0 8862833569 0 4834029363 0 7091768233 0 91731262	7 0.1
27 0.2727451844 0.7989924806 0.8074715317 0.8630560155 0.520064408 0.9395456181 0.7433063779 0.73000267	_
28 0.0233547923 0.8923628118 0.8213618146 0.0381375295 0.7760785248 0.9095269332 0.4853258794 0.31512555	
29 0.431277538 0.8813339958 0.6464634317 0.1806829795 0.0149378479 0.1696588208 0.0184137947 0.49578916	
30 0.9908212749 0.394212218 0.7033778366 0.671705693 0.3933258315 0.6299235066 0.3324489791 0.75473883	4 0.2
31 0.3143209 0.6560834548 0.4499266401 0.4371305082 0.0300377 0.4826313867 0.6088405084 0.33491175	9 0.7
32 0.8232727759 0.9862914453 0.2903483235 0.4740527456 0.5629339886 0.5561574434 0.9774761845 0.7783190	6 0.

How do I know? Data Entropy



How do I know? Data Entropy

THE FOLLOWING **PREVIEW** HAS BEEN APPROVED FOR **ALL AUDIENCES**

BY THE MOTION PICTURE ASSOCIATION OF AMERICA INC.

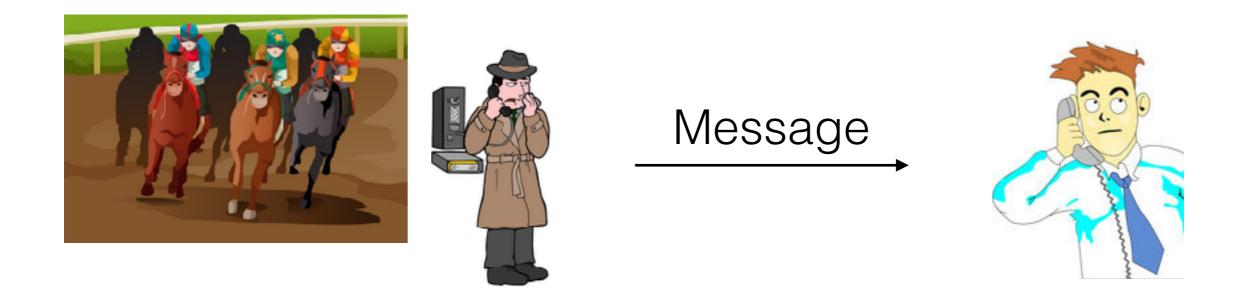
THE FILM ADVERTISED HAS BEEN RATED



www.filmratings.com

www.mpaa.org

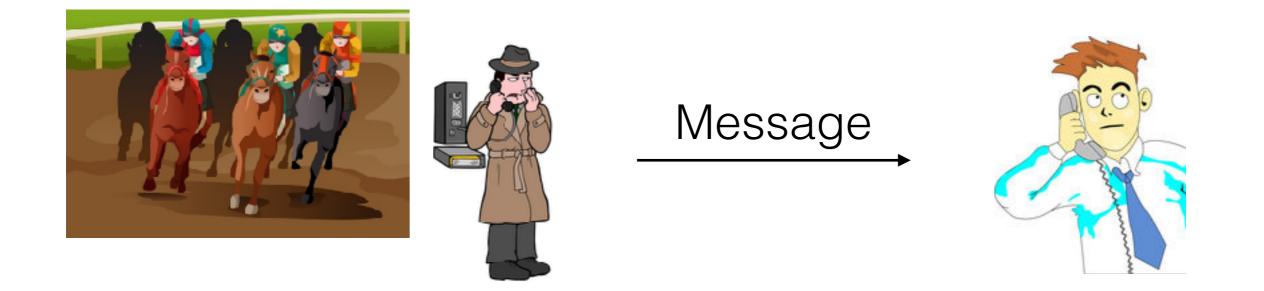
Horse	1	2	3	4	5	6	7	8	Length
P(winning)	1/2	1/4	1/8	1/16	1/32	1/32	1/32	1/32	
Message	000	001	010	011	100	101	110	111	3 bits



Info Theory & Entropy

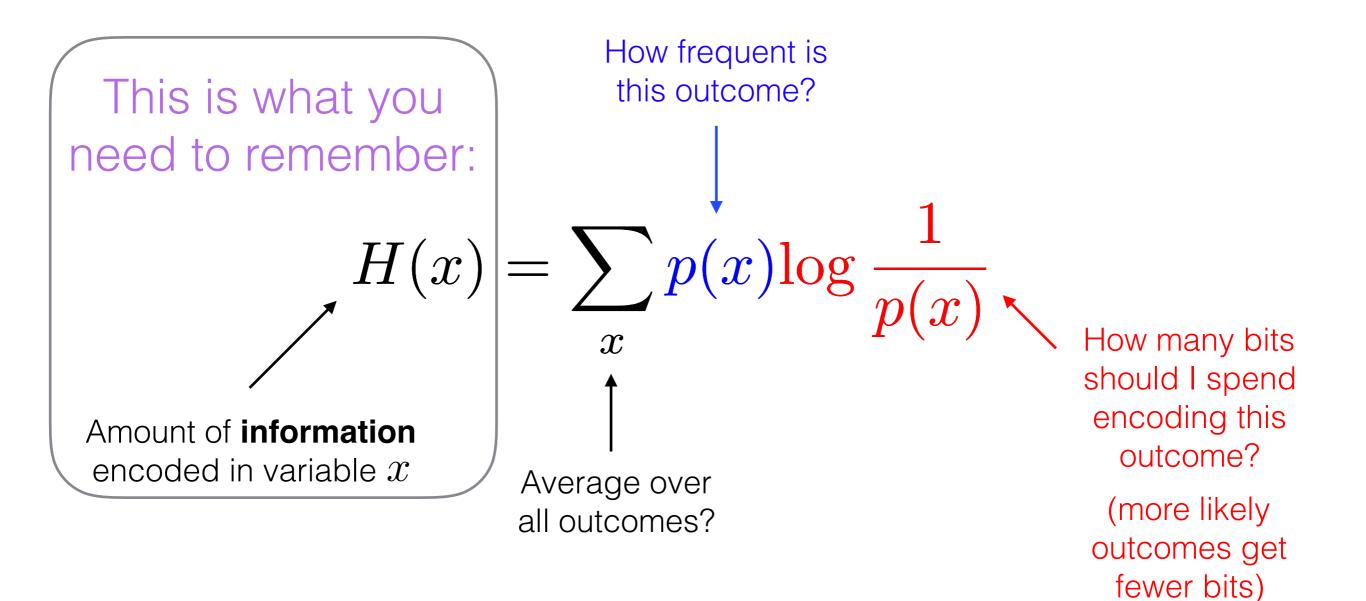
A Quick Detour

Horse	1	2	3	4	5	6	7	8	E[Length]
P(winning)	1/2	1/4	1/8	1/16	1/32	1/32	1/32	1/32	
Message	000	001	010	011	100	101	110	111	3 bits
Optimal	0	10	110	1110	1111 00	1111 01	1111 10	1111 11	2 bits



Info Theory & Entropy

A Quick Detour



Info Theory & Entropy

A Quick Detour

	Gender	PhD?
1	0	0
2	1	0
2 3	1	0
4	0	0
5	0	0
6	0	0
7	1	0
8	1	0
9	1	0
10	1	1
11	0	0
12	0	0
13	0	0
14	1	0
15	0	0
16	1	0

$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

	Gender	PhD?	
1	0	0	
2	1	0	
3	1	0	
4	0	0	
5	0	0	
6	0	0	
7	1	0	
8	1	0	
9	1	0	
10	1	1	
11	0	0	
12	0	0	
13	0	0	
14	1	0	
15	0	0	
16	1	0	
p(0)	1/2	15/16	
p(1)	1/2	1/16	

$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

= $p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)}$
= $\frac{1}{2} \log(2) + \frac{1}{2} \log(2)$
= $\frac{1}{2} + \frac{1}{2}$
= 1

	Gender	PhD?	
1	0	0	
2	1	0	
3	1	0	
4	0	0	
5	0	0	
6	0	0	
7	1	0	
8	1	0	
9	1	0	
10	1	1	
11	0	0	
12	0	0	
13	0	0	
14	1	0	
15	0	0	
16	1	0	
p(0)	1/2	15/16	
p(1)	1/2	1/16	
H(x)	1		

$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

= $p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)}$
= $\frac{1}{2} \log(2) + \frac{1}{2} \log(2)$
= $\frac{1}{2} + \frac{1}{2}$
= 1

	Gender	PhD?	
1	0	0	
2	1	0	
3	1	0	
4	0	0	
5	0	0	
6	0	0	
7	1	0	
8	1	0	
9	1	0	
10	1	1	
11	0	0	
12	0	0	
13	0	0	
14	1	0	
15	0	0	
16	1	0	
p(0)	1/2	15/16	
p(1)	1/2	1/16	
H(x)	1		

$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

= $p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)}$
= $\frac{15}{16} \log(\frac{16}{15}) + \frac{1}{16} \log(16)$
= $\frac{15}{16} (0.093) + \frac{1}{16} (4)$
= 0.337

Most informative!

	Gender	PhD?
1	0	0
2	1	0
2 3	1	0
4	0	0
5	0	0
6 7	0	0
	1	0
8	1	0
9	1	0
10	1	1
11	0	0
12	0	0
13	0	0
14	1	0
15	0	0
16	1	0
p(0)	1/2	15/16
p(1)	1/2	1/16
H(x)		0.337

$$\begin{aligned} (x) &= \sum_{x} p(x) \log \frac{1}{p(x)} \\ &= p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)} \\ &= \frac{15}{16} \log(\frac{16}{15}) + \frac{1}{16} \log(16) \\ &= \frac{15}{16} (0.093) + \frac{1}{16} (4) \\ &= 0.337 \end{aligned}$$

Info Theory & Entropy Example

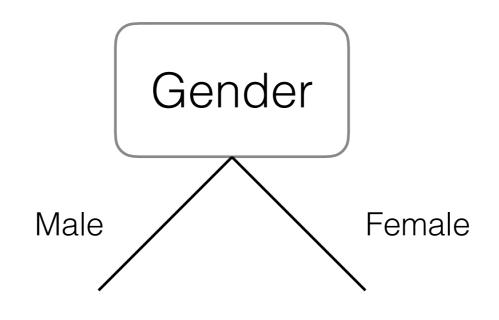
H

Gender



Here is my Data

1) Find Most Informative Variable



Most Informative Variable:

First Decision in my Tree

Then What?

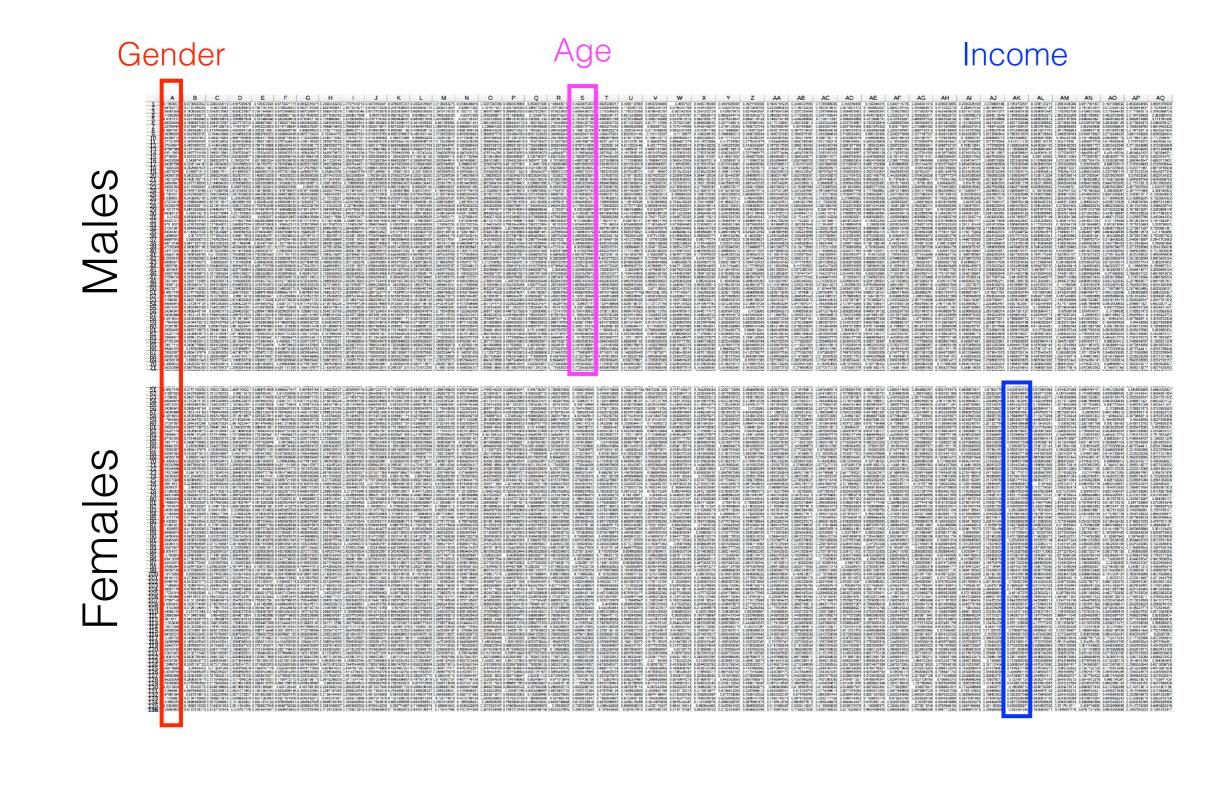
Females

Males

 2
 4
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 2
 0.60465/778.0.19026/09476 0.60465778.0.190269476 0.8819511505.0.6244969151 0.3014337039.0.661789363 0.5464547833.0.8657061053 0.4684547833.0.8657061053 0.46845428112.0.034522503 0.46859428112.0.03452251827 0.868594009.0.732251827 0 464019642 0 421362670 0 421362670 0 42325691 0 832757032 0 68012250 0 63695536 0 45545450 0 105542536 0 45545536 0 45545536 0 45545536 0 369-05012 0 7879095136 0 7879095136 0.3000011222 0.35604.30085 0.6177*17096 0.902581722 0.2715459935 0.1161112527 0.73694.03772 0.73694.03778 0.420171733 0.917C370167 0.0343911515 0.7274291387 0.707C444095 0.5622331053 0.11129977 0.366463541 L200732240 C.2012204162 C.C.724785745 C.274785745 C.2592C77268 D.653969892 C.144597552 C.4458542803 D.779763875 C.4458542803 D.779763875 C.4458542803 C.72873736875 C.445854280 C.2482731969 C.242178028 C.2482731969 C.242178028 C.2482730740 C.24822098329 C.1004732078 C.1004732078 C.1004732078 C.24852800749 C.24852800749 C.24852800749 C.48852807749 C.4895280749 C 0.2255.22500 0.2255.25000 0.2255.25000 0.2255.25000 0.2255.25000 0.2255.25000 0. 2 0.862812731 2 0.6613023211 2 0.613023211 2 0.613032311 2 0.620371025 0.820371025 0.820371025 0.820371031 0.02454037381 0.025451736 0.02545017125 5122840161 1.055151736 4389027548 566211123 C.C74490330 C.2091472076 C.6122991218 C.209217291218 C.209221812 C.209221812 C.209220414 C.2092099174 C.2092099174 C.2092099174 C.2092099174 C.2092099174 C.209209175 C.200212000 C.2002720234 C.2002720234 C.2002720234 C.2002720234 C.2002720234 C.2002720234 C.2002720234 C.2002722034 C.2002720234 C.200272024 C.20027202720 C.2002720272004 C.20027 Pace 11 (2015)
 Accord 7356801653 4830195808 34530195808 345301975 1552941559 1827418397 7157417357 875975C434 3605171614 B36701341b 40623311060 3244036051 4611138781 491138781 4935829737 263232596 4125515570 1136084098 2652579779 412651557571 1136084098 2652579779 412653677146 2652579779 412653677146 265257179 25567771420 1338222138 2556771420 1338222138 796820677, 890350470 206819069 206917252 837674822 003586947 101902900 834031730 122909732 036213510 0.9814223447 0.9814794883 0 0.8814005745 0.988451061 0 0.9714890490 0.100830403 0 0.7770548956 0.700340461 0 0.0812440164 0.0103443464 0.0812440164 0.0103443464 0.081264018 0.275126076 0.54206677567 0.0206731120 0 0.442181011 0.9874479021 0 501800129 508962777 122449543 6825-455 1 6891011290 0 97360803 5 3825762 (78412968 0 43336365 0.168386677 0.829633588 0.24262140 0.257085652 0.582264303 0.001238466 0.45292922 800793682 391078125 480454737 072548355 077544411 211500386 235229008 517817339 3626333667 2211333687 2211333687 210273647 210273647 23047783 0.0548/29133 0.3490902871 0.39705/581 0.3901493801 0.4353868826 0 3671032758 0 3520618167 0 5232497591 0 2154831300 de15881287
 de540977328
 de54097328
 de52758612
 de52758612
 de54097328
 de54097328
 de5409738
 de5409738
 de54097
 de5409
 de5409
 de5409
 de5409
 de5409
 de5409
 de5409
 de5409
 de540
 de54
 de5
 de54
 de54
 de5
 de54
 de5
 de54
 de5
 de54
 de5
 de5 0143274517 1136084058 6562579777 1136084058 656257977742 11368204058 65625797742 1136822145 1358271452 135827145 13582714 1358271 0.333900441 0.684797002 0.5234854808 0.5234854808 0.888294814 0.5086478970 0.482829814 0.5086478977 0.482830441 0.0125522788 0.555729821 0.54588778 0.415045587 0.41504587 0.415045587 0. 0.813119/01/6 0.04546911.63 0.3621032756 0.3620618167 0.55732492581 0.2154801305 0.435704302 0.934492213 0.934492213 0.93706302 0.934492213 0.93706302 0.937063894 0.1774525386 0.510771523 0.8004515468 0.8024515468 0.8024515468 0.8024515468 4469077738 5450323673 1868447089 2914574614 7519304540 7821710112 5710412974 83129374 96174750 67925774 71203166 56844322 351907414 67572158 95366418 70543353 630200503 887057072 8007151802 0.000502560 735443685 904021000 205571407 7250182830 812258007 2261545445 762597837 7775641965 7511043432 3274632805 7129663285 1702294315 5422098315 2153630047 055030928 042759860 1187703711 101733001 470768858 1225016458 341530279 410532708 2201 50336 8122536072 55215626 7523742996 3762087394 5521326151 4447538300 203722615 0.950096203 2174656257 2621971615 7527576011 28869 30524 56058 36364 0 953687803 1055962904 1674525484 446677196 043088036 085-905-206 77548-8152 146700749 371194847 774245127 8318,7554 0.39834457 201294927 0.298304927 0.9988/28221 0.9986/28290 0.9986/28290 0.9986/28290 0.728279133 0.72827913 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.72827927 0.728279 228381188 150190442 750276726 47753359 133580050 34060352 204796568 717410548 246202271 C. 12804118/73 C. 1280254015-31 C. 1280254015-32 C. 1280254015-32 C. 1280254015-32 C. 1280254015-32 C. 1280254015-32 C. 1280254025 C. 12802550 0.1260192287 0.7665732012 0.4269054481 0.975713168 0.713287206 0.9575550012 0.7575550012 0.1319185355 0.517860404 0.5338199625 0.7282263012 0.7282263012 0.7355225786 0.1118336138 0.2266122040 0.7369451743 0.2174258086 0.448542315 0.3575045585 0.5386904511 0.1576327039 0.9870552225 Determinants to all tables of a provinsions of a balant rest Determinants of a provinsion of a provinsion of a balant rest Determinant of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provinsion of a provinsion of a provinsion Determinants of a provinsion of a provi 0.8122576263 17848364176 0.0216367128 0.021128756 0.027585842 0.65575847 0.85575847 0.95575847 0.95575847 160455433, 3656278576 1151435152 140173614 517065483 129609676

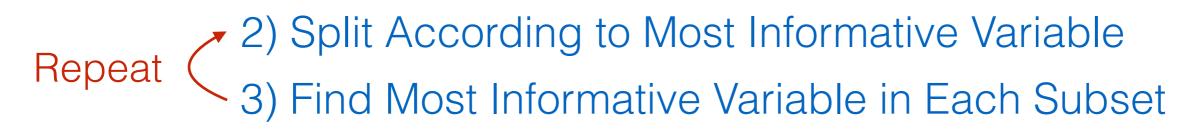
2) Split According to Most Informative Variable3) Find Most Informative Variable in Each Subset

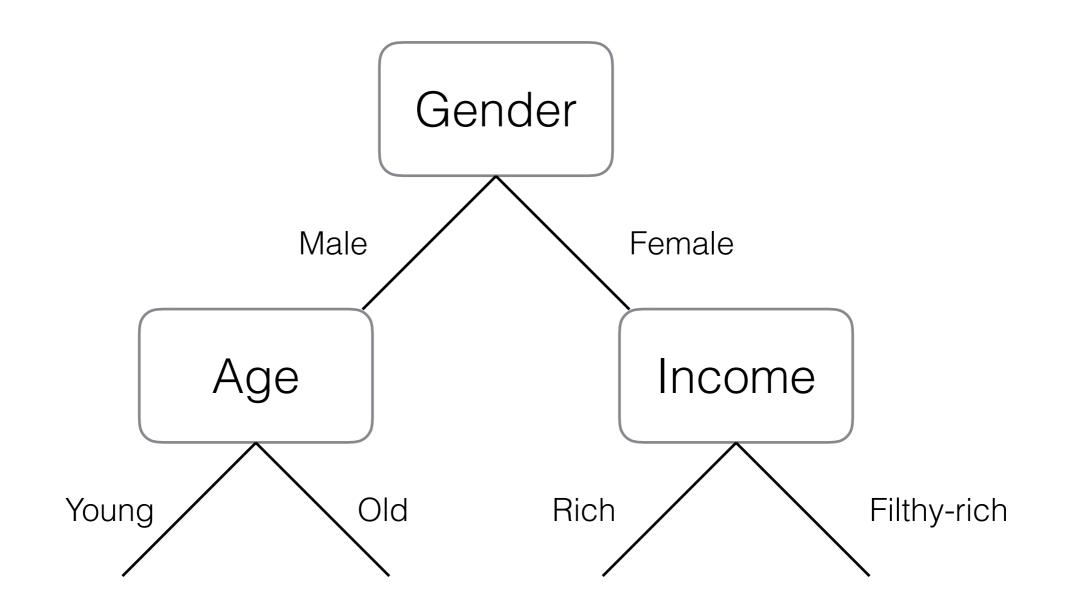
Gender



Repeat (2) Split According to Most Informative Variable 3) Find Most Informative Variable in Each Subset



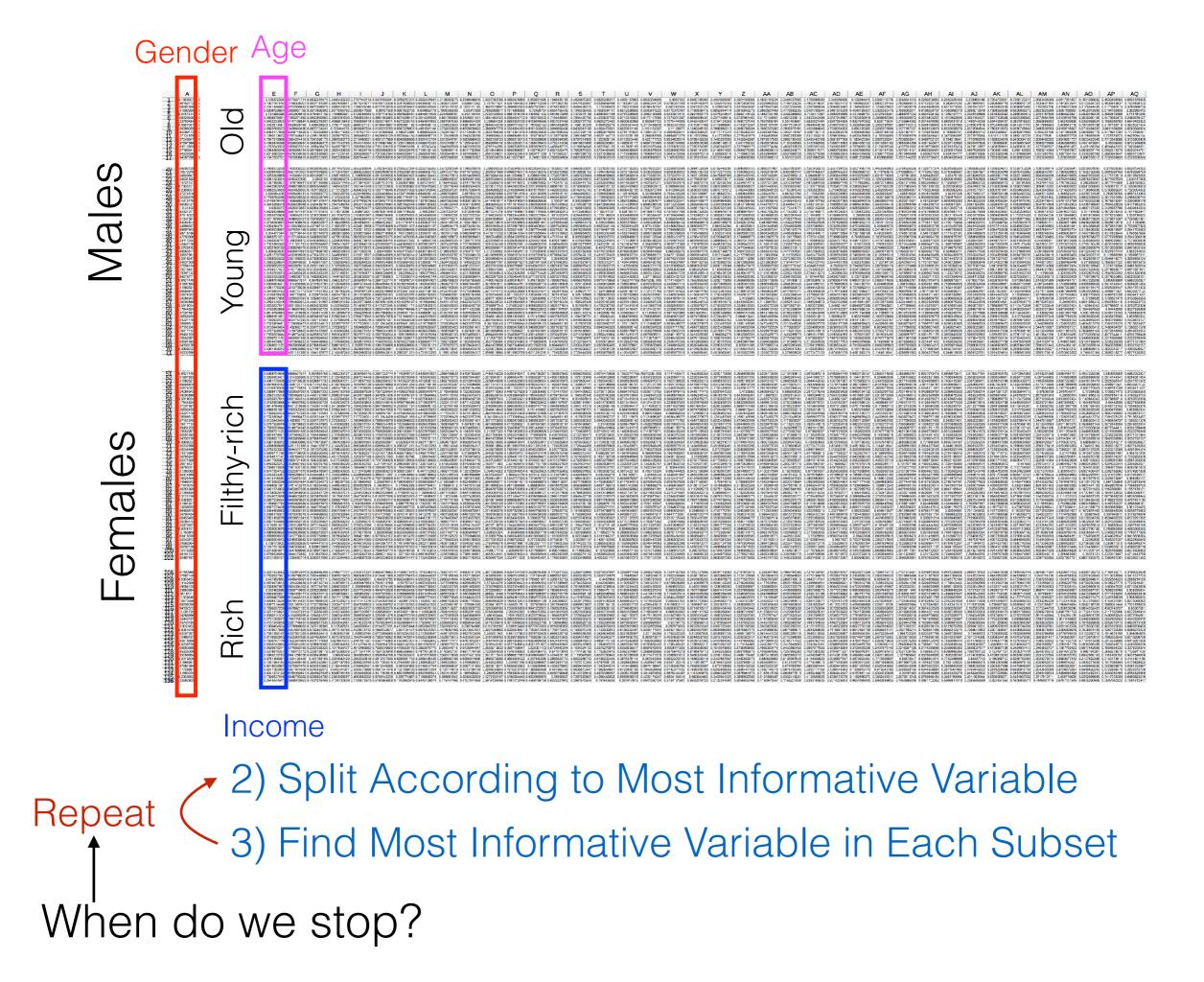




Each Informative Variable:

One Decision in my Tree

Then What?



	Gender	PhD?	Will Die?
1	0	0	1
2	1	0	1
3	1	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	1	0	1
8	1	0	1
9	1	0	1
10	1	1	1
11	0	0	1
12	0	0	1
13	0	0	1
14	1	0	1
15	0	0	1
16	1	0	1
p(0)	1/2	15/16	0
p(1)	1/2	1/16	1
H(x)	1	0.337	

$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

= $p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)}$
= $0 \log(\frac{1}{0}) + 1 \log(1)$
= 0

Info Theory & Entropy Extreme Case

			↓ I
	Gender	PhD?	Will Die?
1	0	0	1
2	1	0	1
3	1	0	1
4	0	0	1
5	0	0	1
6	0	0	1
7	1	0	1
8	1	0	1
9	1	0	1
10	1	1	1
11	0	0	1
12	0	0	1
13	0	0	1
14	1	0	1
15	0	0	1
16	1	0	1
p(0)	1/2	15/16	0
p(1)	1/2	1/16	1
H(x)	1	0.337	0

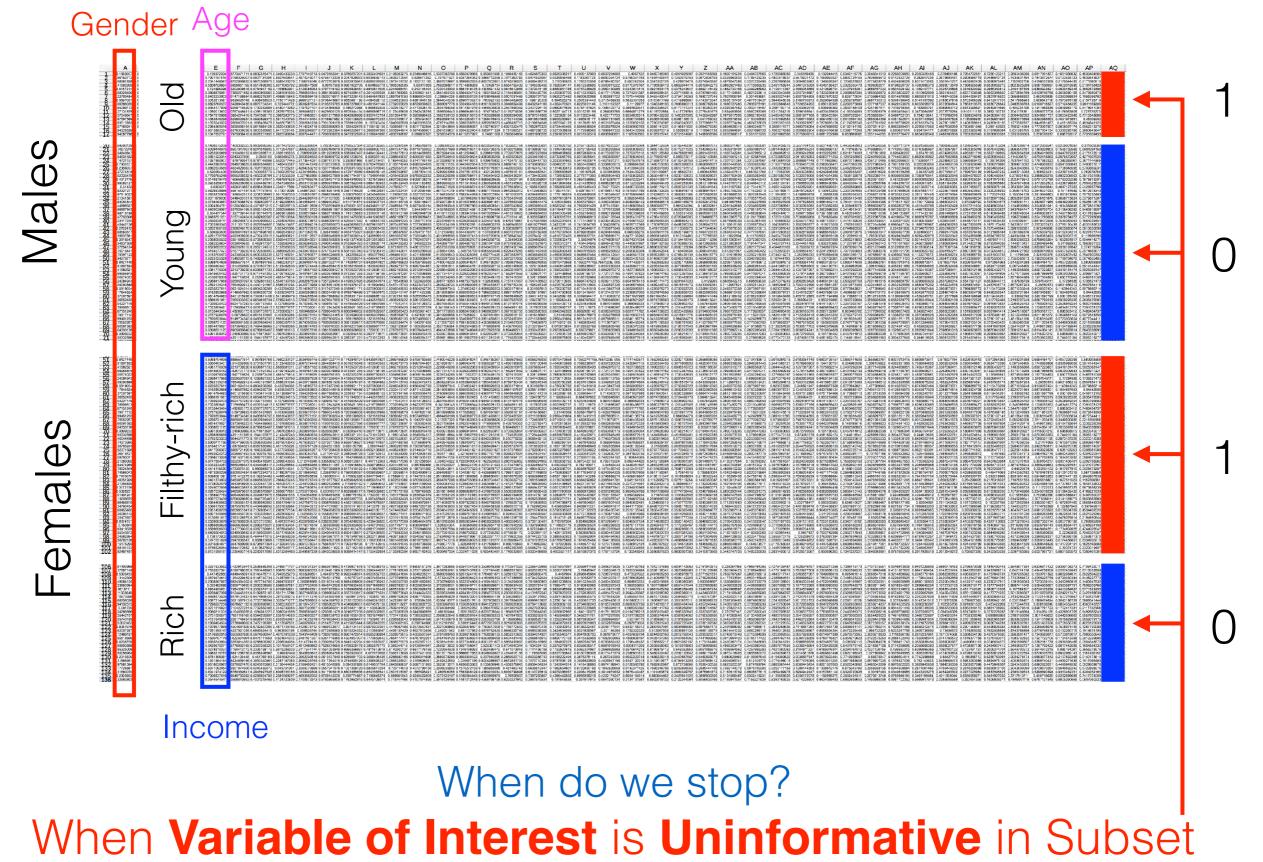
$$H(x) = \sum_{x} p(x) \log \frac{1}{p(x)}$$

= $p(0) \log \frac{1}{p(0)} + p(1) \log \frac{1}{p(1)}$
= $0 \log(\frac{1}{0}) + 1 \log(1)$
= 0

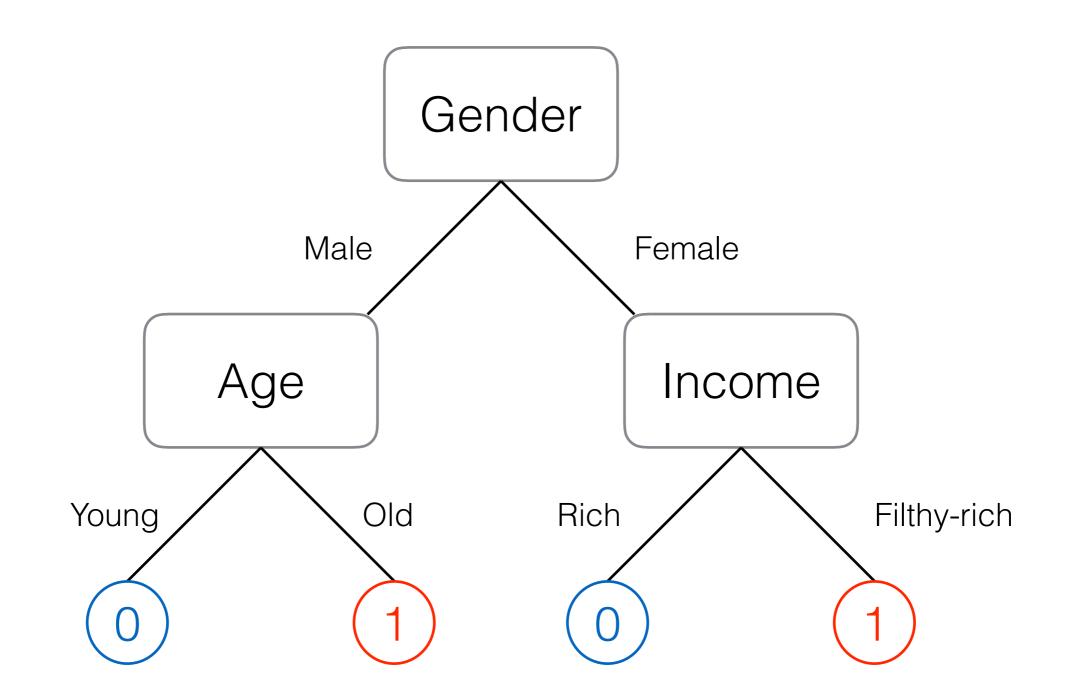
This Variable Provides No Information!

Info Theory & Entropy

Extreme Case



(~zero entropy)



Finally...

Put Result on my Decision Tree

And enjoy! (start predicting)



(Didn't I promise partial nudity?)

What could possibly go wrong? Overfitting & Bias

- **Overfitting.** My tree is accurate, but only for my given data (for which I already know the answer)
 - Not a lot of "predictive power".
- **Bias.** It may heavily depend on my particular sample.
 - If I add/remove a few people, the result may be very different!

What could possibly go wrong? Overfitting & Bias

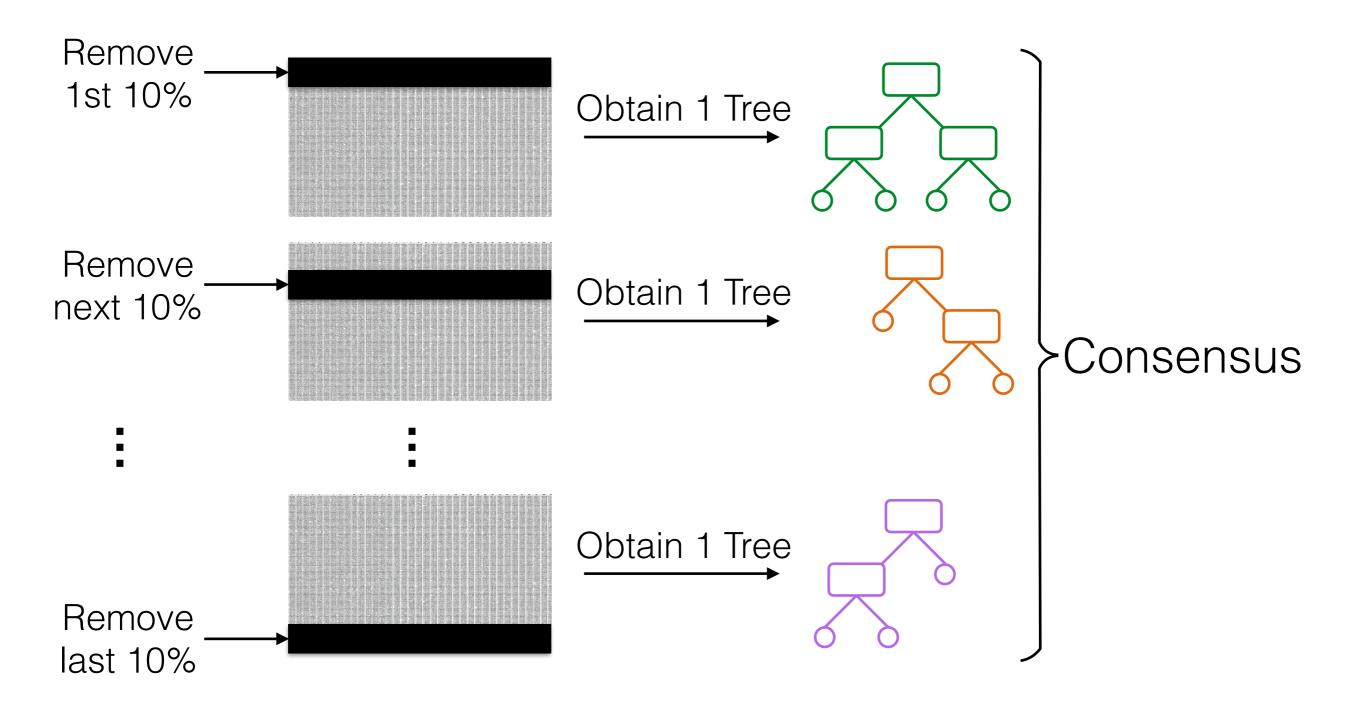
No worries: Random Forests



Random Forests

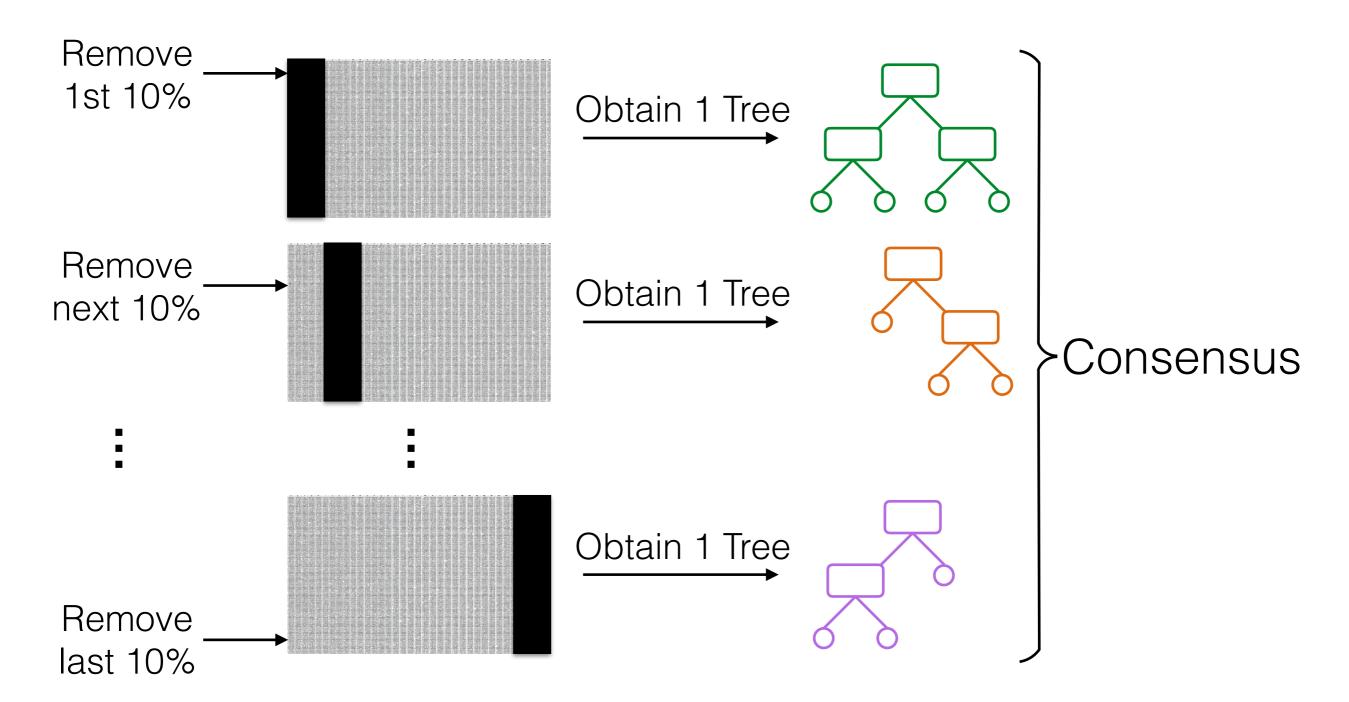
Main Idea: Do many decision trees, each with a random subsample





Random Forests

Main Idea: Do many decision trees, each with a random subsample



Random Forests

Main Idea: Do many decision trees, each with a random subsample

- Good for **Prediction**, but bad for **Description**.
- Fast to train, but **slow** to predict.
- Poor performance on **unbalanced** data.

Advantages/Disadvantages

- Neural Networks
- Regression (Linear, Logistic, Polynomial)
- Other clustering methods (e.g., subspaces)

Alternatives

Questions?