THE MORAL OF THE STORY:

DON'T BE A NARCISSIST

THE PROBLEM:

Make some predictions about marital status based on provided factors:

1. **Joint Income**
2. **Bride and Groom Age**
3. **Bride and Groom Personality Type**

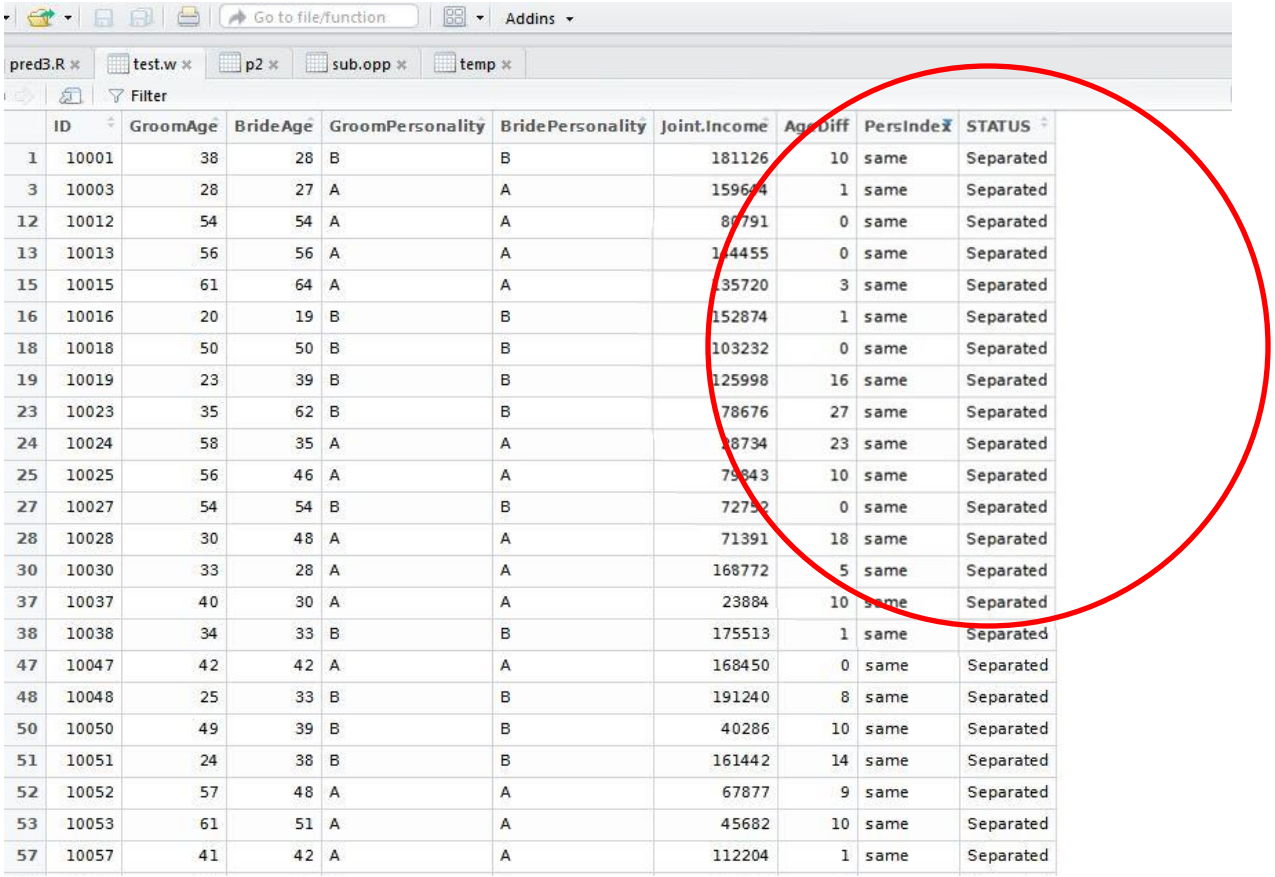
THE APPROACH:

Tools used: Rstudio, google, rpart, google, crossvalidation2, google, email the TAs, google more.

The key was just some good old laziness. I didn't want to deal with ages and personalities for every individual so I condensed everything into age differences and a personality index that indicated inter-couple parity. That is, they were marked as "same" if they were both A or both B and "opp" if they were opposites.

When I looked at the training data, I thought:

THIS COULD BE SIGNIFICANT



Go to file/function | Addins

pred3.R | test.w | p2 | sub.opp | temp

Filter

	ID	GroomAge	BrideAge	GroomPersonality	BridePersonality	Joint.Income	AgeDiff	PersIndex	STATUS
1	10001	38	28	B	B	181126	10	same	Separated
3	10003	28	27	A	A	159644	1	same	Separated
12	10012	54	54	A	A	80791	0	same	Separated
13	10013	56	56	A	A	144455	0	same	Separated
15	10015	61	64	A	A	135720	3	same	Separated
16	10016	20	19	B	B	152874	1	same	Separated
18	10018	50	50	B	B	103232	0	same	Separated
19	10019	23	39	B	B	125998	16	same	Separated
23	10023	35	62	B	B	78676	27	same	Separated
24	10024	58	35	A	A	98734	23	same	Separated
25	10025	56	46	A	A	79843	10	same	Separated
27	10027	54	54	B	B	72752	0	same	Separated
28	10028	30	48	A	A	71391	18	same	Separated
30	10030	33	28	A	A	168772	5	same	Separated
37	10037	40	30	A	A	23884	10	same	Separated
38	10038	34	33	B	B	175513	1	same	Separated
47	10047	42	42	A	A	168450	0	same	Separated
48	10048	25	33	B	B	191240	8	same	Separated
50	10050	49	39	B	B	40286	10	same	Separated
51	10051	24	38	B	B	161442	14	same	Separated
52	10052	57	48	A	A	67877	9	same	Separated
53	10053	61	51	A	A	45682	10	same	Separated
57	10057	41	42	A	A	112204	1	same	Separated

Turns out couples composed of the same personality type invariably ended up separated.

NO KIDDING IT'S SIGNIFICANT

So I made an appropriate rule to account for this in rpart. I also included the age difference and joint income, then let rpart decide what was important. Running cross validation and then (always!) summary() yields:

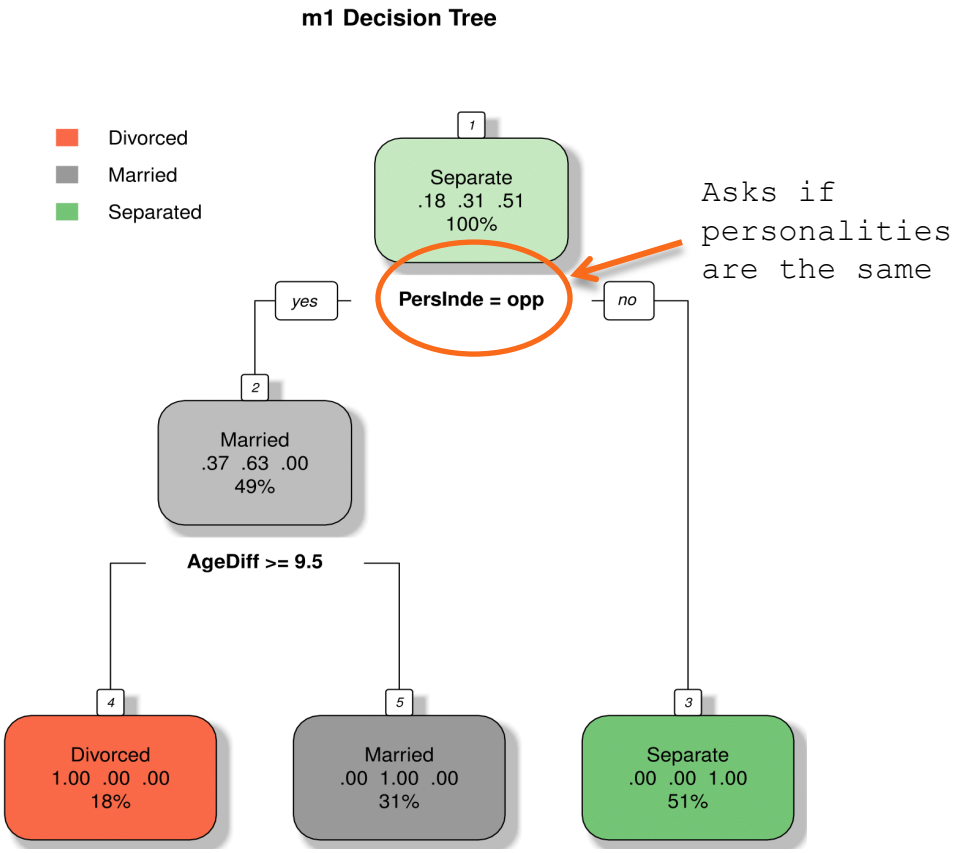
```
> CrossValidation::cross_validate(train.r, m1, 2, 0.8)
  accuracy_subset accuracy_all
1               1            1
2               1            1

> summary(m1)
Call:
rpart(formula = STATUS ~ AgeDiff + Joint.Income + PersIndex,
    data = train.r, method = "class", control = rpart.control(minsplit = 3,
        minbucket = 1))
  n= 750

        CP nsplit rel error    xerror       xstd
1 0.630137      0  1.000000 1.000000 0.03750190
2 0.369863      1  0.369863 0.369863 0.02882577
3 0.010000      2  0.000000 0.000000 0.00000000

Variable importance
   PersIndex      AgeDiff Joint.Income
          57           37            6
```

I am the validest

Rpart seems to think money isn't so important, and as we all know, machines are never wrong.

PERSONALITY GOES A LONG WAY

The biggest bulk division seems to be happening along the lines of personality type. If you're so in love with yourself you just *have* to have two of you in your life, the irony is that you're doomed to separation. Mix it up!



LEST WE FORGET

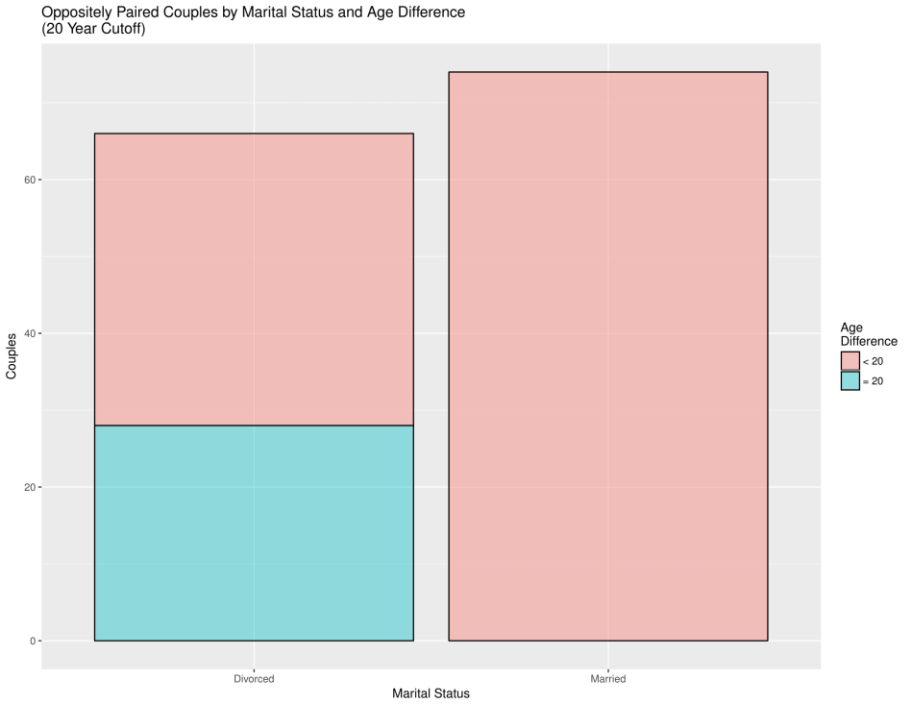
Looking at the mean incomes separated out by marital status and parity (sameness of personality) was instructive:

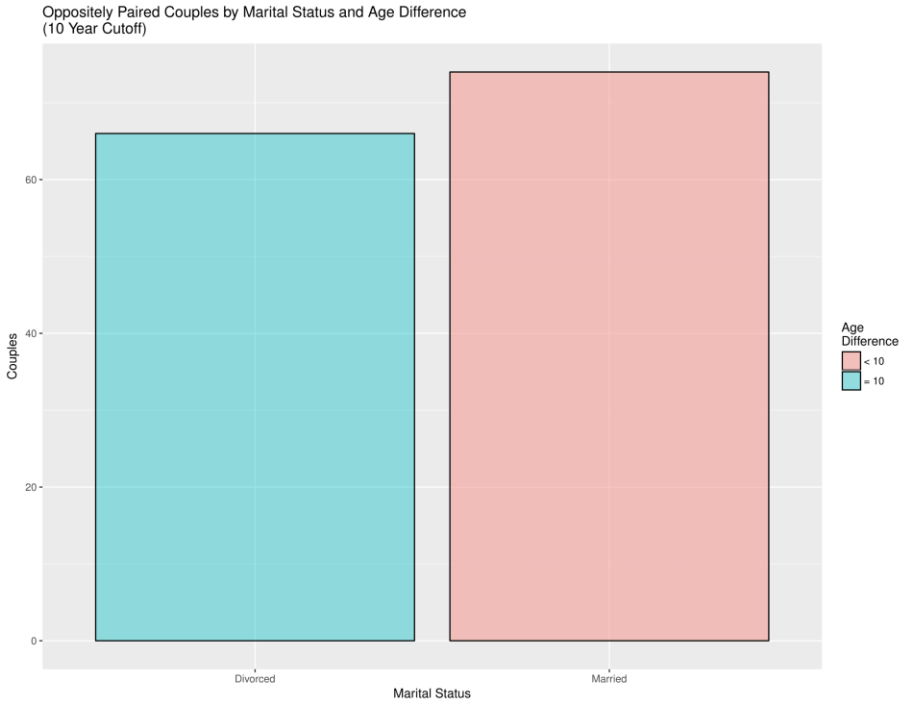
```
> tapply(test.w$Joint.Income, test.w$STATUS, mean) # mean joint income by status
 Divorced   Married Separated
 107496.2  114083.4  116188.3
> tapply(test.w$Joint.Income, test.w$PersIndex, mean) # mean joint income by parity
     opp     same
110978.0 116188.3
> tapply(test.w$Joint.Income, test.w$STATUS, median) # median joint income by status
 Divorced   Married Separated
 111997.5  117592.0  117735.0
> tapply(test.w$Joint.Income, test.w$PersIndex, median) # median joint income by parity
     opp     same
115588.5 117735.0
>
```

Note that the "separated" couples and the "same" couples have identical average joint income. This makes sense.

Also note that mean and median incomes for all status types showed little variance. This is probably a pretty good indicator that joint income doesn't figure in much. If it did, one would expect to see one group with a much higher mean income. For instance, if you assume more money leads to greater marriage stability, then you'd expect the mean income for the still married couples to be higher. It's not. So money is not a factor, particularly.

GREAT.





Looking again at the tree, I find that within the subset of opposites, the key ingredient for marital success appeared to be age difference. I counted up divorced and married couples for different age cutoffs. 10 years seemed the most tidy. 20 was done for comparison.

LIFE ADVICE FOR PERFECT HAPPINESS:

Narcissism drives people apart. Stop it. Yes, you.

Good. Now you have a shot.

Make sure you only fall in love with someone with an age within 10 years of your own. This is non-negotiable. Do you hate happiness?

And remember that money doesn't matter so much, I guess.

Thanks.