

Lab 6

This data set consists of observations from 276 speed dates. It was obtained from the blackboard site for BTRY 6010.

The source of the `SpeedDating` data is “Gelman, A. and Hill, J., *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press: New York, 2007.”

```
SpeedDating <- read.csv("SpeedDating.csv")
```

We will start by looking at our data before we proceed with any other analysis.

```
dim(SpeedDating)
```

```
## [1] 276 22
```

```
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      format.pval, units
```

```
describe(SpeedDating)
```

```
## SpeedDating
```

```
##
```

```
## 22 Variables      276 Observations
```

```
## -----
```

```
## DecisionM
```

	n	missing	distinct	Info	Sum	Mean	Gmd
##	276	0	2	0.747	146	0.529	0.5001

```
##
```

```
## -----
```

```
## DecisionF
```

	n	missing	distinct	Info	Sum	Mean	Gmd
##	276	0	2	0.745	127	0.4601	0.4986

```
##
```

```
## -----
```

```
## LikeM
```

	n	missing	distinct	Info	Mean	Gmd	.05	.10
##	274	2	12	0.969	6.682	1.966	4.00	4.15
##	.25	.50	.75	.90	.95			
##	6.00	7.00	8.00	9.00	9.35			

```

##
## lowest : 1.0 2.0 3.0 4.0 4.5, highest: 6.5 7.0 8.0 9.0 10.0
##
## Value      1.0  2.0  3.0  4.0  4.5  5.0  6.0  6.5  7.0  8.0  9.0
## Frequency    2    4    7   15    1   33   54    3   68   46   27
## Proportion 0.007 0.015 0.026 0.055 0.004 0.120 0.197 0.011 0.248 0.168 0.099
##
## Value      10.0
## Frequency    14
## Proportion 0.051
## -----
## LikeF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    272      4      11    0.964    6.366    1.931      3      4
##    .25    .50    .75    .90    .95
##      5      7      8      8      9
##
## lowest : 1.0 2.0 3.0 4.0 5.0, highest: 7.0 8.0 8.5 9.0 10.0
##
## Value      1.0  2.0  3.0  4.0  5.0  6.0  7.0  8.0  8.5  9.0 10.0
## Frequency    1    7   14   13   41   54   72   48    1   13    8
## Proportion 0.004 0.026 0.051 0.048 0.151 0.199 0.265 0.176 0.004 0.048 0.029
## -----
## PartnerYesM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    272      4      11    0.978    5.757    2.484    2.00    3.00
##    .25    .50    .75    .90    .95
##    5.00    6.00    7.00    8.00    9.45
##
## lowest : 0 1 2 3 4, highest: 6 7 8 9 10
##
## Value      0    1    2    3    4    5    6    7    8    9   10
## Frequency    1   12   14   19   19   49   54   46   34   10   14
## Proportion 0.004 0.044 0.051 0.070 0.070 0.180 0.199 0.169 0.125 0.037 0.051
## -----
## PartnerYesF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    272      4      10    0.973    5.835    2.49    1.00    2.10
##    .25    .50    .75    .90    .95
##    5.00    6.00    7.00    9.00    9.45
##
## lowest : 1 2 3 4 5, highest: 6 7 8 9 10
##
## Value      1    2    3    4    5    6    7    8    9   10
## Frequency   15   13   15   10   66   44   50   28   17   14
## Proportion 0.055 0.048 0.055 0.037 0.243 0.162 0.184 0.103 0.062 0.051
## -----
## AgeM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    273      3      21    0.991    26.6    3.877    22.0    22.0
##    .25    .50    .75    .90    .95
##   24.0   27.0   29.0   30.8   32.4
##
## lowest : 18 19 20 21 22, highest: 34 36 37 39 42

```

```

## -----
## AgeF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    271      5      20    0.992    26.19    4.15      21      22
##    .25    .50    .75    .90    .95
##    23     26     28     31     34
##
## lowest : 19 20 21 22 23, highest: 34 35 36 38 55
##
## Value      19     20     21     22     23     24     25     26     27     28     29
## Frequency      1      3     13     23     33     29     29     30     28     20     20
## Proportion 0.004 0.011 0.048 0.085 0.122 0.107 0.107 0.111 0.103 0.074 0.074
##
## Value      30     31     32     33     34     35     36     38     55
## Frequency     14      3      3      7      7      4      2      1      1
## Proportion 0.052 0.011 0.011 0.026 0.026 0.015 0.007 0.004 0.004
## -----
## RaceM
##      n missing distinct
##    274      2      5
##
## lowest : Asian      Black      Caucasian Latino      Other
## highest: Asian      Black      Caucasian Latino      Other
##
## Value      Asian      Black Caucasian      Latino      Other
## Frequency      65      10      161      17      21
## Proportion    0.237    0.036    0.588    0.062    0.077
## -----
## RaceF
##      n missing distinct
##    272      4      5
##
## lowest : Asian      Black      Caucasian Latino      Other
## highest: Asian      Black      Caucasian Latino      Other
##
## Value      Asian      Black Caucasian      Latino      Other
## Frequency      70      15      148      23      16
## Proportion    0.257    0.055    0.544    0.085    0.059
## -----
## AttractiveM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    273      3      11    0.971    6.687    2.026      4      5
##    .25    .50    .75    .90    .95
##    5      7      8      9     10
##
## lowest : 1.0 2.0 3.0 4.0 5.0, highest: 6.5 7.0 8.0 9.0 10.0
##
## Value      1.0     2.0     3.0     4.0     5.0     6.0     6.5     7.0     8.0     9.0    10.0
## Frequency      1      3      7     12     54     50      1     48     55     23     19
## Proportion 0.004 0.011 0.026 0.044 0.198 0.183 0.004 0.176 0.201 0.084 0.070
## -----
## AttractiveF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    274      2      12    0.973    6.274    2.136    3.000    4.000

```

```

##      .25      .50      .75      .90      .95
##    5.000    6.000    8.000    8.000    9.175
##
## lowest :  1.0  2.0  3.0  4.0  5.0, highest:  7.5  8.0  9.0  9.5 10.0
##
## Value      1.0  2.0  3.0  4.0  5.0  6.0  7.0  7.5  8.0  9.0  9.5
## Frequency      3   7  16  15  51  50  56   1  49  12   1
## Proportion 0.011 0.026 0.058 0.055 0.186 0.182 0.204 0.004 0.179 0.044 0.004
##
## Value      10.0
## Frequency      13
## Proportion 0.047
## -----
## SincereM
##      n missing distinct      Info      Mean      Gmd
##    271      5      9      0.95      7.856      1.62
##
## lowest :  1  3  4  5  6, highest:  6  7  8  9 10
##
## Value      1      3      4      5      6      7      8      9      10
## Frequency      1      1      5     11     25     52     87     50     39
## Proportion 0.004 0.004 0.018 0.041 0.092 0.192 0.321 0.185 0.144
## -----
## SincereF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    273      3      10     0.96     7.784     1.845      5      5
##      .25      .50      .75      .90      .95
##      7      8      9      10      10
##
## lowest :  1  2  3  4  5, highest:  6  7  8  9 10
##
## Value      1      2      3      4      5      6      7      8      9      10
## Frequency      1      1      3      5     19     26     44     78     50     46
## Proportion 0.004 0.004 0.011 0.018 0.070 0.095 0.161 0.286 0.183 0.168
## -----
## IntelligentM
##      n missing distinct      Info      Mean      Gmd
##    268      8      8     0.95     7.621     1.527
##
## lowest :  4.0  5.0  6.0  7.0  7.5, highest:  7.0  7.5  8.0  9.0 10.0
##
## Value      4.0  5.0  6.0  7.0  7.5  8.0  9.0  10.0
## Frequency      3  17  33  65   1  82  41  26
## Proportion 0.011 0.063 0.123 0.243 0.004 0.306 0.153 0.097
## -----
## IntelligentF
##      n missing distinct      Info      Mean      Gmd
##    273      3      9     0.942     7.923     1.531
##
## lowest :  2  3  4  5  6, highest:  6  7  8  9 10
##
## Value      2      3      4      5      6      7      8      9      10
## Frequency      1      1      3     14     13     59     92     51     39
## Proportion 0.004 0.004 0.011 0.051 0.048 0.216 0.337 0.187 0.143

```

```

## -----
## FunM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    270      6      11    0.969    6.863    1.977      4      5
##    .25    .50    .75    .90    .95
##      6      7      8      9    10
##
## lowest : 0 1 2 3 4, highest: 6 7 8 9 10
##
## Value      0      1      2      3      4      5      6      7      8      9    10
## Frequency    1      1      3      4     12     37     51     57     61     22    21
## Proportion 0.004 0.004 0.011 0.015 0.044 0.137 0.189 0.211 0.226 0.081 0.078
## -----
## FunF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    270      6      10    0.976    6.563    2.208      3      4
##    .25    .50    .75    .90    .95
##      5      7      8      9    10
##
## lowest : 1 2 3 4 5, highest: 6 7 8 9 10
##
## Value      1      2      3      4      5      6      7      8      9    10
## Frequency    1      7     11     23     34     48     55     49     23    19
## Proportion 0.004 0.026 0.041 0.085 0.126 0.178 0.204 0.181 0.085 0.070
## -----
## AmbitiousM
##      n missing distinct      Info      Mean      Gmd
##    259     17      9    0.971    6.768    1.95
##
## lowest : 2 3 4 5 6, highest: 6 7 8 9 10
##
## Value      2      3      4      5      6      7      8      9    10
## Frequency    2      4     13     51     44     55     46     27     17
## Proportion 0.008 0.015 0.050 0.197 0.170 0.212 0.178 0.104 0.066
## -----
## AmbitiousF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    266     10     10    0.969    7.429    1.969      5      5
##    .25    .50    .75    .90    .95
##      6      8      9     10     10
##
## lowest : 1 2 3 4 5, highest: 6 7 8 9 10
##
## Value      1      2      3      4      5      6      7      8      9    10
## Frequency    1      2      3      4     35     30     48     65     45    33
## Proportion 0.004 0.008 0.011 0.015 0.132 0.113 0.180 0.244 0.169 0.124
## -----
## SharedInterestsM
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    249     27     12    0.972    5.588    2.297      2      3
##    .25    .50    .75    .90    .95
##      4      5      7      8      9
##
## lowest : 0.0 1.0 2.0 3.0 4.0, highest: 6.5 7.0 8.0 9.0 10.0

```

```
##
## Value      0.0  1.0  2.0  3.0  4.0  5.0  6.0  6.5  7.0  8.0  9.0
## Frequency   4   5  11  12  31  63  42   1  38  24   7
## Proportion 0.016 0.020 0.044 0.048 0.124 0.253 0.169 0.004 0.153 0.096 0.028
##
## Value      10.0
## Frequency   11
## Proportion 0.044
## -----
## SharedInterestsF
##      n missing distinct      Info      Mean      Gmd      .05      .10
##    246      30      12    0.981    5.47    2.57      2      2
##    .25      .50      .75      .90      .95
##      4       6       7       8       9
##
## lowest :  0  1  2  3  4, highest:  6  7  8  9 10
##
## Value      0.0  1.0  2.0  3.0  4.0  5.0  5.5  6.0  7.0  8.0  9.0
## Frequency   2  10  20  20  19  50   1  40  38  25  12
## Proportion 0.008 0.041 0.081 0.081 0.077 0.203 0.004 0.163 0.154 0.102 0.049
##
## Value      10.0
## Frequency   9
## Proportion 0.037
## -----
```

Note: the variable `Decision` is categorical. R has read it in as a quantitative value.

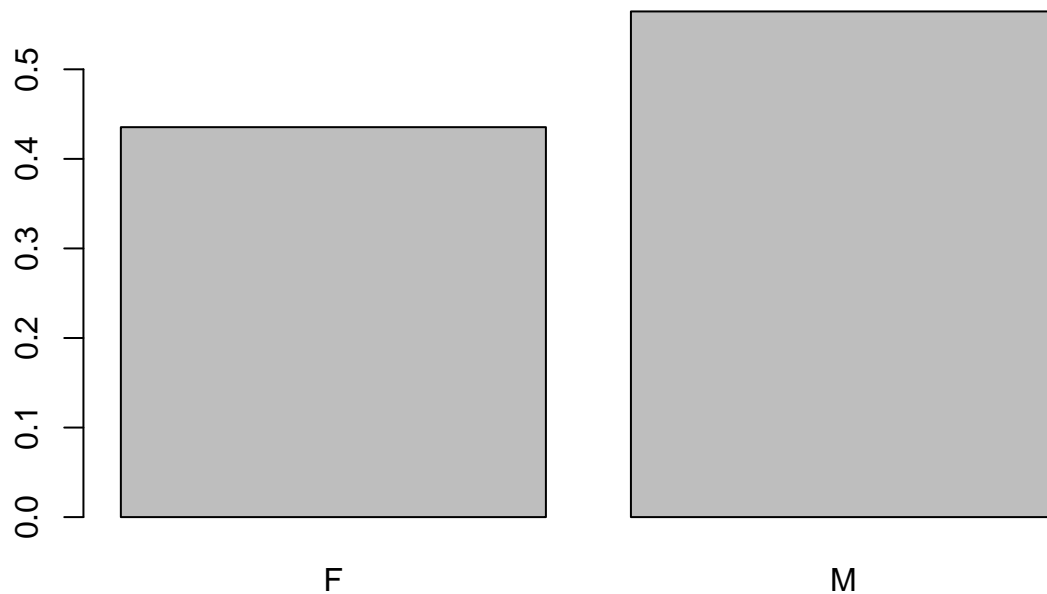
```
table(SpeedDating$DecisionF, SpeedDating$DecisionM)
```

```
##
##      0  1
##  0 66 83
##  1 64 63
```

When `DecisionF` and `DecisionM` are both 0, neither participant wanted another date. When `DecisionF` is 0 and `DecisionM` is 1, the female did not want another date, but the male did want another date. When `DecisionF` is 1 and `DecisionM` is 0, the male did not want another date, but the female did want another date. When `DecisionF` and `DecisionM` are both 1, both participants wanted another date.

```
onlyOne <- subset(SpeedDating, DecisionF + DecisionM == 1)
SympMale = table(onlyOne$DecisionM) / nrow(onlyOne)
barplot(SympMale, names.arg=c('F', 'M'), main = 'Who Desired the Second Date When One of Them Did Not?')
```

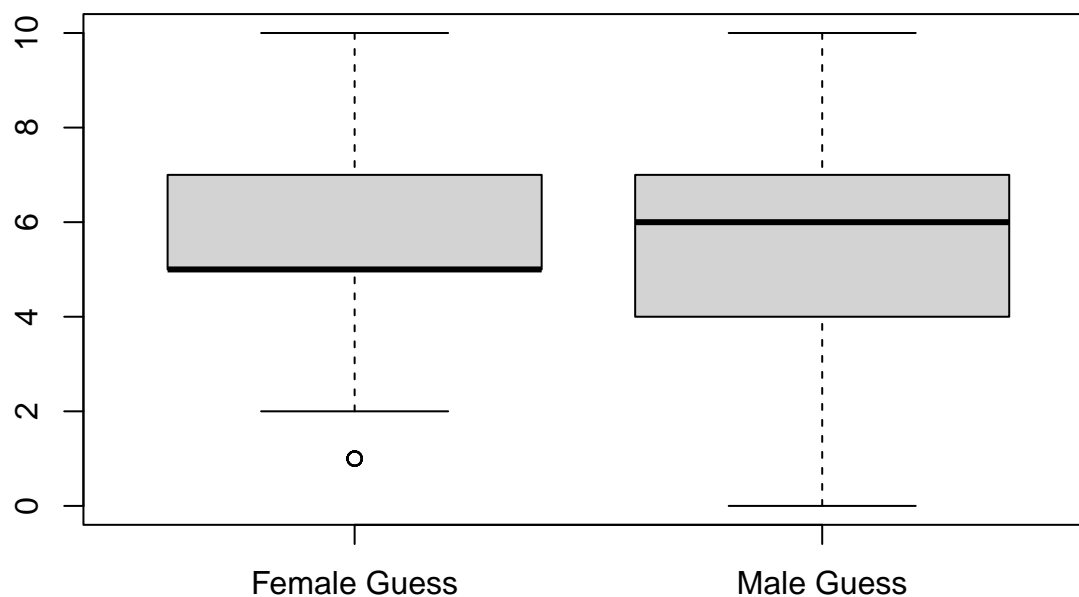
Who Desired the Second Date When One of Them Did Not?



When only one partner was interested in another date, it was more often the male than the female.

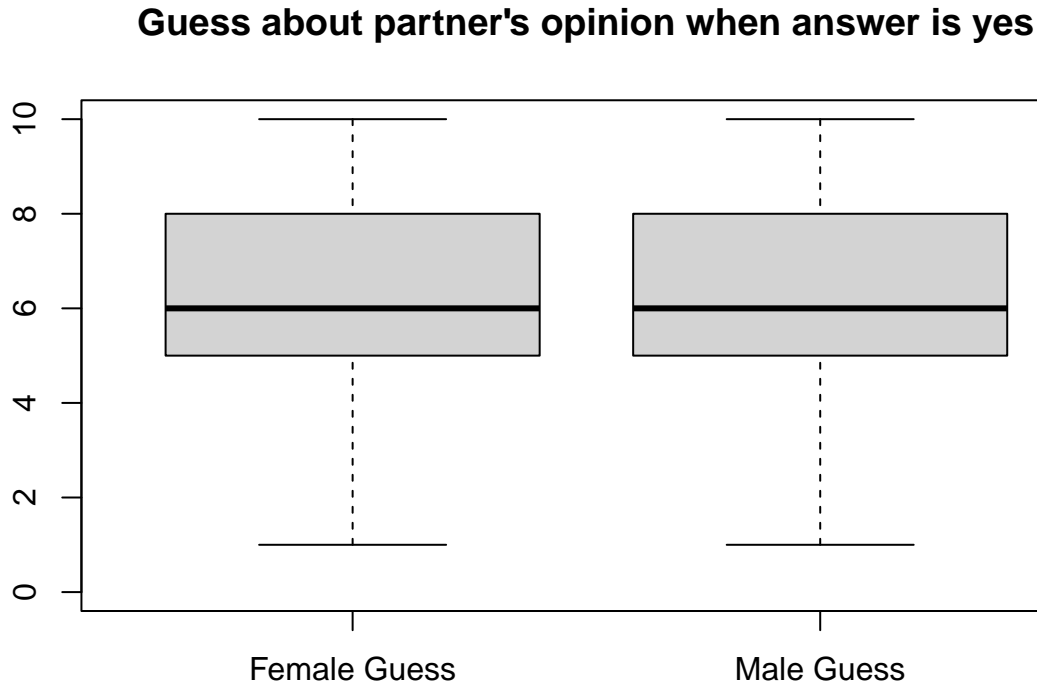
```
boxplot(SpeedDating$PartnerYesF[SpeedDating$DecisionM == 0],
        SpeedDating$PartnerYesM[SpeedDating$DecisionF == 0],
        main="Guess about partner's opinion when answer is no",
        ylim=c(0, 10),
        names=c('Female Guess', 'Male Guess'))
```

Guess about partner's opinion when answer is no



This plot contains ratings on a scale of 1 to 10 by females (on the left) and males (on the right) of how likely their “date” would want to see them again for observations where the date did not want to see them again.

```
boxplot(SpeedDating$PartnerYesF[SpeedDating$DecisionM == 1],
        SpeedDating$PartnerYesM[SpeedDating$DecisionF == 1],
        main="Guess about partner's opinion when answer is yes",
        ylim=c(0,10),
        names=c('Female Guess', 'Male Guess'))
```



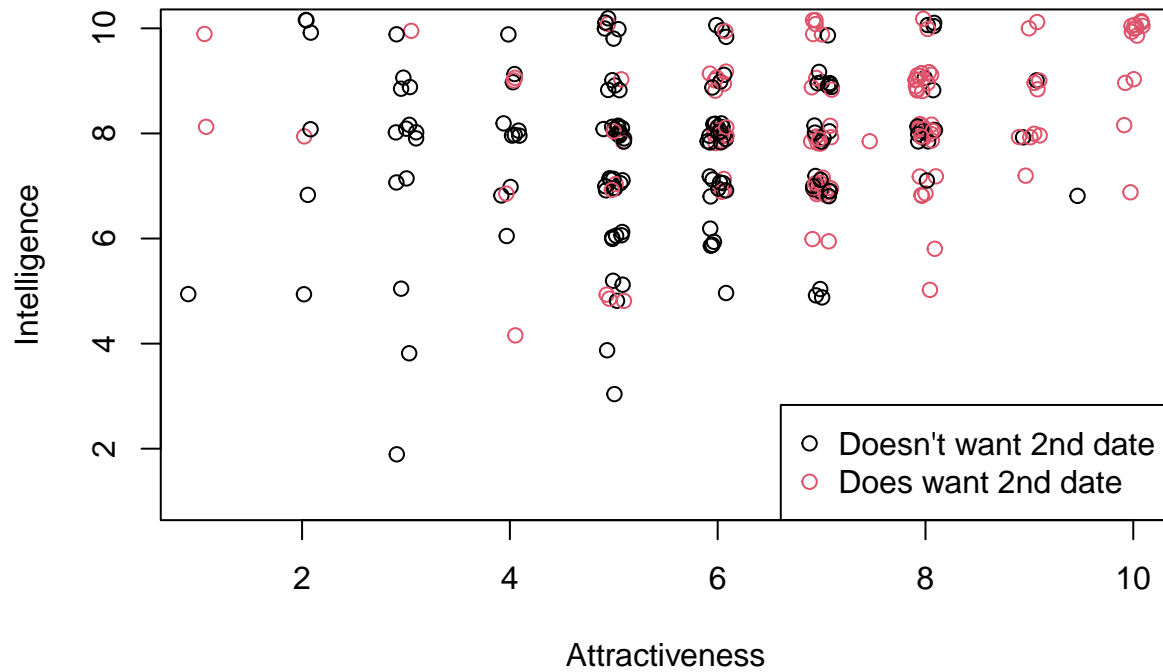
This plot contains ratings on a scale of 1 to 10 by females (on the left) and males (on the right) of how likely their “date” would want to see them again for observations where the date did want to see them again.

Males and females were equally likely to detect when their partners wanted another date. The distributions look identical.

When the partner did not want another date, the median male’s assessment of the situation is unchanged while the median female does notice that he is not interested in another date. That said, the 3rd quartile of the male distribution does drop as much as that of the female distribution.

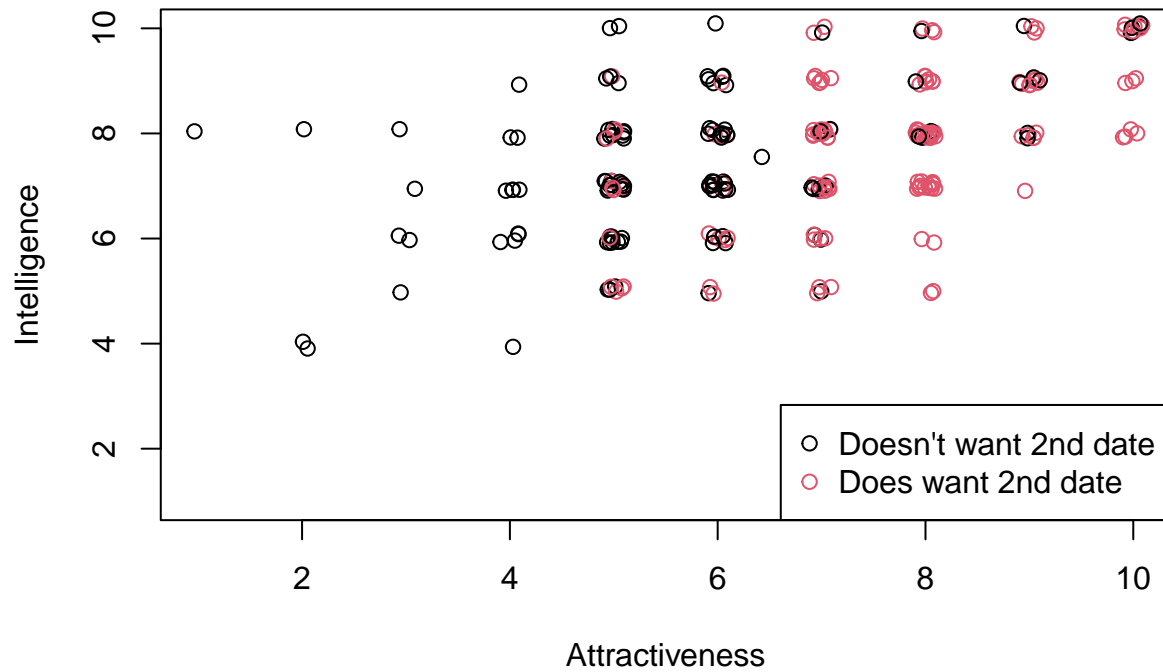
```
plot(jitter(SpeedDating$AttractiveF),
     jitter(SpeedDating$IntelligentF),
     main="Female Assessment of Male",
     ylim=c(1,10), xlim=c(1,10),
     col=1+SpeedDating$DecisionF,
     xlab="Attractiveness", ylab="Intelligence")
legend("bottomright",
      legend = c("Doesn't want 2nd date", "Does want 2nd date"), col=1:2, pch=1)
```


Female Assessment of Male



```
plot(jitter(SpeedDating$AttractiveM),
     jitter(SpeedDating$IntelligentM),
     main="Male Assessment of Female",
     ylim=c(1,10), xlim=c(1,10),
     col=1+SpeedDating$DecisionM,
     xlab="Attractiveness", ylab="Intelligence")
legend("bottomright",
      legend=c("Doesn't want 2nd date", "Does want 2nd date"), col=1:2, pch=1)
```

Male Assessment of Female



Males tended to agree to a second date with partners they found highly attractive and intelligent. Females are less predictable. Compared to the males, they were more likely to be interested in a second date with someone they found intelligent, but not very attractive. However, it is hard to generalize because we are only looking at two of the six attributes measured for each date.