speedating

Manda

11/14/2019

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#install.packages("data.table")   
 #install.packages("ggplot2")  
 library(data.table)  
 library(ggplot2)  
  
speedating <- read.csv("C:/Users/Manda/Documents/Data-Munging/EDA Project/Speed-Dating-Data.csv", na.strings = c("PrivacySuppressed", "NULL"))

# Speed-dating:

##What are the differences between Men & Women? ##What do Women look for in Men?

### Introduction

In 2002-2004, there was data collected for 21 different speed-dating outings with different individuals. Data was collected about each persons’s income, religious background, career, favorite hobbies, race, and ratings on how each individual believes they are perceived by others.

After each speed-dating outing, each individual was asked whether they wanted to match with the individuals they met. Additionally, they were asked to rate the people they met on a variety of traits (ambitious, attractive, sincere, fun, intelligent, etc.). In this analysis, we explored and assesed the factors that most significantly influence a person’s decision to “match” with others.

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':  
##   
## between, first, last

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

##   
## Attaching package: 'ggvis'

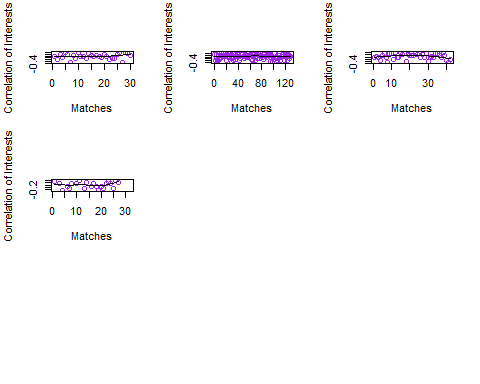
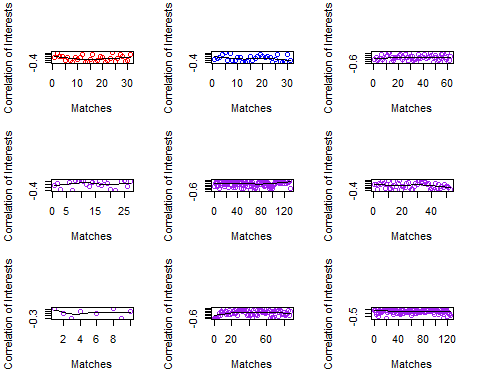
## The following object is masked from 'package:ggplot2':  
##   
## resolution

Note ``that theecho = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

### Correlations

To start, we wanted to see if there were correlations between interests of participants and the interests of the partner they requested to be matched to. Below are the various speed-dating outings (which vary in size) and the interest correlations of those who said “yes” to being matched.

## Loading required package: magrittr



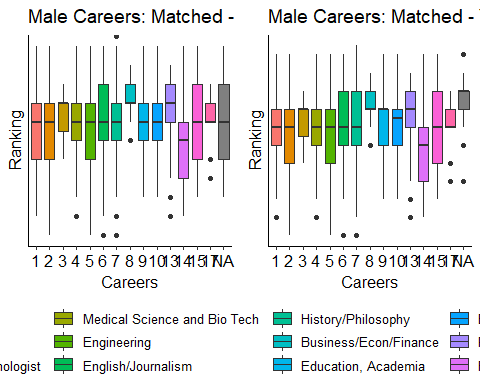
# Careers of those who were Matched and Not-Matched

Since there was a variety of differences in the interest correlations, we wanted to explore what made someone say “yes”. So, we asked ourselves what are the factors that influence a woman to say “yes” to being matched after only 4 minutes.

First, we filtered the data by women who said “yes” to being matched, and then looked at the careers of the men they asked to be matched with. Then we commpared that graph to women who said “no” to being matched and the careers of the men who they rejected. We plotted the “yes” & “no” matches against an attribute ranking (out of 10) that was given by partners at the speed-dating outing.

## Warning: Removed 111 rows containing non-finite values (stat\_boxplot).  
  
## Warning: Removed 111 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 101 rows containing non-finite values (stat\_boxplot).



## Hobbies: Differences between a 0/10 and 10/10.

After plotting the “yes” and “no” of men with their careers and rankings, we saw that there were men who were ranked 0/10, 1/10 and those who were ranked 9/10 and 10/10. Naturally, we wanted to know who these men were. For the men who received these rankings we asked, “why are you the way that you are?”

We grouped the data for men who were 0/10 and 1/10 and for men who were 9/10 and 10/10. After grouping the data we plotted their hobbies/interests against the rankings they received for these traits so we could determine if their was a large difference in hobbies between those who were given a 0 and those who were given a 10. Once again, these rankings were out of 10.

##   
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':  
##   
## extract

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

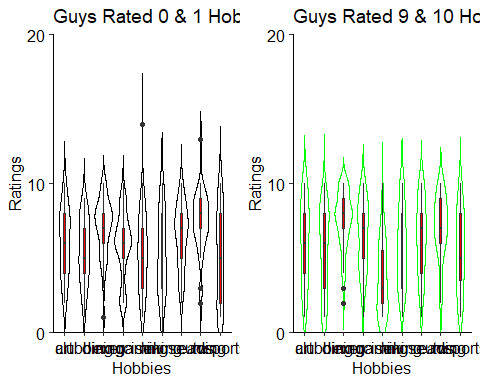
## Warning in raw\_data$attr\_o == target: longer object length is not a  
## multiple of shorter object length

## Warning: Removed 416 rows containing missing values (geom\_violin).

## Warning: Removed 9 rows containing non-finite values (stat\_ydensity).

## Warning: Removed 9 rows containing non-finite values (stat\_boxplot).

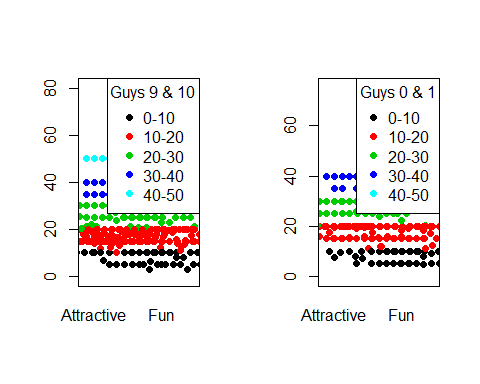
## Warning: Removed 499 rows containing missing values (geom\_violin).



## Personality: Differences between a 0/10 and 10/10.

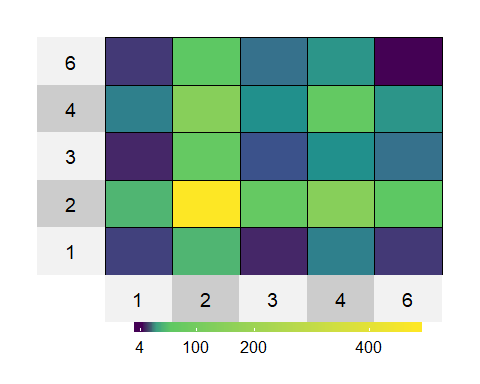
There wasn’t much difference between hobbies/interests of the men, so we moved on to the personality differences. Perhaps there were certain personalities of men that women weren’t attracted to within the short four minute time span of the speed-dating process. We don’t see an incredible difference, especially with the beeswarm plot (because the “yes” and “no” groups differ in size). So we made another plot to better represent the differences.

## Warning in raw\_data$attr\_o == target: longer object length is not a  
## multiple of shorter object length



## What about race?

We’ve looked at careers, hobbies/interests, and personality traits. What else could factor into somone’s decision in the short time span they meet someone during the speed-dating outing? Perhaps race is a big factor in the decision. So we looked at the correlation of race vs. match to see if this was the case.

Black/African American=1 European/Caucasian-American=2 Latino/Hispanic American=3 Asian/Pacific Islander/Asian-American=4 Native American=5 Other=6 

## Women and Men on Personalities: Part 1

Once we viewed the difference in personalities of men who were ranked differently, we then asked ourselves what the differences are between men’s and women’s personalities. In this graph, the “ranking” data is from the survey where both men and women rate what traits are important to them. The sum of the score must equal 100.

## Warning: Removed 5 rows containing non-finite values (stat\_ydensity).

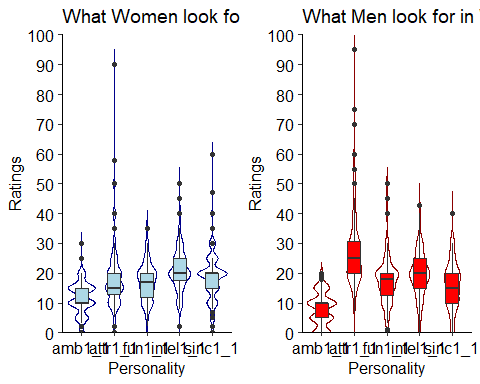
## Warning: Removed 5 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 186 rows containing missing values (geom\_violin).

## Warning: Removed 5 rows containing non-finite values (stat\_ydensity).

## Warning: Removed 5 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 275 rows containing missing values (geom\_violin).



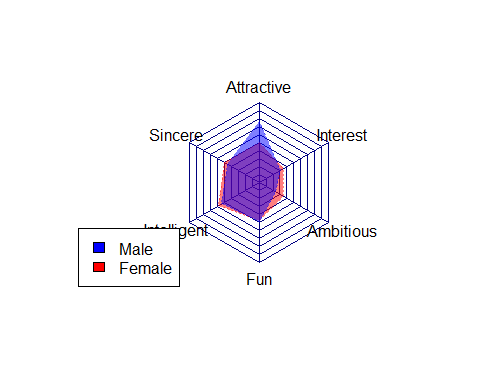
## Women and Men on Personalities: Part 2

We wanted a better picture of the overlap of desired personality traits of men and women, so we created the web-overlap graph to obtain a better idea of the similarites and differences between desired traits.

##   
## 0 90 95 99.98 99.99 100 100.01 100.02 101 110   
## 7 4 1 2 20 478 25 9 1 1   
## 120 148   
## 2 1

##   
## 90 95 100 101 110 120 148   
## 4 1 534 1 1 2 1

## Warning: `data\_frame()` is deprecated, use `tibble()`.  
## This warning is displayed once per session.



## In Conclusion

In conclusion, we see there are many factors which influence women’s and men’s decisions on saying “yes” to another date. For men, there seem to be more importance on the women being attractive and fun. However, women’s taste in men it seems that women prefer someone who is a balance of ambitious, intelligent, and fun. Additionally, there didn’t seem to be much variance between the careers of men who were given a “no” and “yes” on the matching. Furthermore, the differences between the men who received a “0/1” and those who received a “9/10” were not significant except a few differences on the importance of “attractivenenss”.