





TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

Khoa Công nghệ thông tin

# THỐNG KÊ MÁY TÍNH VÀ ỨNG DỤNG

## BÀI TẬP 3

Phân tích dữ liệu thăm dò  
Exploratory data analysis



TP.HCM – 04/2019

## Nội dung

1)	Mô tả tổng quan dataset .....	3
❖	Nguồn gốc.....	3
❖	Mục đích nghiên cứu .....	3
❖	Tổng thể nghiên cứu .....	3
❖	Cách thu thập dữ liệu .....	3
❖	Cỡ mẫu: .....	3
❖	Số lượng biến: .....	3
2)	Chọn biến – Mô tả.....	5
❖	Biến định tính.....	5
❖	Biến định lượng .....	5
❖	Lý do chọn: .....	6
3)	Phân tích thăm dò riêng từng biến .....	6
❖	DecisionFemale .....	6
❖	RaceF.....	8
❖	LikeF .....	10
❖	FunF .....	13
❖	AmbitiousF .....	16
4)	Phân tích thăm dò quan hệ giữa hai biến định tính:.....	19
❖	Tóm tắt bằng thống kê .....	19
❖	Trình bày trực quan bằng các biểu đồ.....	20
❖	Nhận xét, đánh giá .....	21
5)	Phân tích thăm dò quan hệ giữa 1 biến định tính, 1 biến định lượng.....	22
❖	Tóm tắt bằng thống kê .....	22
❖	Trình bày trực quan bằng các biểu đồ.....	23
❖	Nhận xét, đánh giá .....	25
6)	Phân tích thăm dò quan hệ giữa 2 biến định lượng:.....	26
❖	Tóm tắt bằng thống kê .....	26
❖	Trình bày trực quan bằng các biểu đồ.....	26
❖	Nhận xét, đánh giá .....	26
7)	Code R: .....	27
8)	Nguồn tham khảo:.....	30

## 1) Mô tả tổng quan dataset

### ❖ Nguồn gốc

Dataset có nguồn gốc từ

“Gelman, A. and Hill, J., Data analysis using regression and multilevel/hierarchical models, Cambridge University Press: New York, 2007”

### ❖ Mục đích nghiên cứu

Speed Date - Hẹn hò tốc độ là một quá trình mai mối chính thức với mục đích tạo cơ hội cho người độc thân có đủ điều kiện gặp gỡ số lượng lớn bạn đời tiềm năng trong một khoảng thời gian rất ngắn. Với dữ liệu thu thập được biết được sự tương quan, quan hệ giữa các thuộc tính khảo sát. Từ đó giúp đưa ra được các cuộc sắp xếp các cặp đôi tham gia hợp lý nhất để tác hợp thành công các cặp đôi.

### ❖ Tổng thể nghiên cứu

Cuộc nghiên cứu tiến hành trên những người đàn ông từ 18 đến 42 tuổi, người phụ nữ từ 15 đến 55 tuổi. Trên các chủng tộc chính là người châu Á (Asian), người da đen (Black), người da trắng (Caucasian), người La-tinh (Latino) và những chủng tộc khác. Khảo sát sự đánh giá của người phụ nữ, người đàn ông về các tính cách, đặc điểm như sức hút, sự hài hước, sự chân thành, sự thông minh, ... của đối phương trong cuộc hẹn ngắn.

### ❖ Cách thu thập dữ liệu

Những người tham gia là các sinh viên tại các trường đại học và sau đại học của Columbia. Được giới thiệu qua email, các tờ quảng cáo được đăng và các tờ quảng cáo do các công tác viên hỗ trợ.

Mỗi người tham gia đã tham dự một buổi hẹn hò tốc độ, trong đó họ gặp gỡ với những người tham gia khác giới trong bốn phút. Thứ tự và phiên tham gia được xác định ngẫu nhiên. Sau mỗi bốn phút "speed date – hẹn hò tốc độ", người tham gia điền vào biểu mẫu xếp hạng cuộc hẹn của họ theo tỷ lệ từ 1 đến 10 trên các thuộc tính khác nhau. Chỉ dữ liệu từ cuộc hẹn đầu tiên trong mỗi phiên được ghi lại ở đây

### ❖ Cỡ mẫu:

276 mẫu

### ❖ Số lượng biến:

22 biến

- `DecisionMale` Would the male like another date? `Yes` or `No`

Người đàn ông tham gia khảo sát có muốn một cuộc hẹn khác? (Giá trị Yes hoặc No)

- `DecisionM` Would the male like another date? `0` or `1`

Người đàn ông tham gia khảo sát có muốn một cuộc hẹn khác? (Giá trị 0 hoặc 1)

- `DecisionFemale` Would the female like another date? `Yes` or `No`

Người phụ nữ tham gia khảo sát có muốn một cuộc hẹn khác? (Giá trị Yes hoặc No)

### Bài tập 3

- **DecisionF** Would the female like another date? 0 or 1  
Người phụ nữ tham gia khảo sát có muốn một cuộc hẹn khác? (Giá trị 0 hoặc 1)
- **LikeM** How much the male likes his partner (1-10 scale)  
Mức độ thích đối phương của người đàn ông tham gia cuộc hẹn đầu tiên như thế nào? (Thang từ 1 đến 10)
- **LikeF** How much the female likes her partner (1-10 scale)  
Mức độ thích đối phương của người phụ nữ tham gia cuộc hẹn đầu tiên như thế nào? (Thang từ 1 đến 10)
- **PartnerYesM** Male's estimate of chance the female wants another date (1-10 scale)  
Ước tính của người đàn ông về việc đối phương nữ có muốn có một cuộc hẹn khác không? (Thang từ 1 đến 10)
- **PartnerYesF** Female's estimate of chance the male wants another date (1-10 scale)  
Ước tính của người phụ nữ về việc đối phương nam có muốn có một cuộc hẹn khác không? (Thang từ 1 đến 10)
- **AgeM** Male's age (in years)  
Độ tuổi của người đàn ông? (Tính theo năm)
- **AgeF** Females age (in years)  
Độ tuổi của người phụ nữ? (Tính theo năm)
- **RaceM** Male's race: Asian, Black, Caucasian, Latino, or Other  
Chủng tộc của người đàn ông? (Châu Á, Da đen, Da trắng, Latinh, Khác)
- **RaceF** Female's race: Asian, Black, Caucasian, Latino, or Other  
Chủng tộc của người phụ nữ? (Châu Á, Da đen, Da trắng, Latinh, Khác)
- **AttractiveM** Male's rating of female's attractiveness (1-10 scale)  
Đánh giá của người đàn ông về sức hấp dẫn của đối phương nữ? (Thang từ 1 đến 10)
- **AttractiveF** Female's rating of male's attractiveness (1-10 scale)  
Đánh giá của người phụ nữ về sức hấp dẫn của đối phương nam? (Thang từ 1 đến 10)
- **SincereM** Male's rating of female's sincerity (1-10 scale)  
Đánh giá của người đàn ông về sự chân thành của đối phương nữ? (Thang từ 1 đến 10)
- **SincereF** Female's rating of male's sincerity (1-10 scale)  
Đánh giá của người phụ nữ về sự chân thành của đối phương nam? (Thang từ 1 đến 10)
- **IntelligentM** Male's rating of female's intelligence (1-10 scale)  
Đánh giá của người đàn ông về sự thông minh của đối phương nữ? (Thang từ 1 đến 10)

### Bài tập 3

- **IntelligentF** Female's rating of male's intelligence (1-10 scale)  
Đánh giá của người phụ nữ về sự thông minh của đối phương nam?  
(Thang từ 1 đến 10)
- **FunM** Male's rating of female as fun (1-10 scale)  
Đánh giá của người đàn ông về sự vui tính của đối phương nữ?  
(Thang từ 1 đến 10)
- **FunF** Female's rating of male as fun (1-10 scale)  
Đánh giá của người phụ nữ về sự vui tính của đối phương nam?  
(Thang từ 1 đến 10)
- **AmbitiousM** Male's rating of female's ambition (1-10 scale)  
Đánh giá của người đàn ông về tham vọng (sự cầu tiến) của đối phương nữ?  
(Thang từ 1 đến 10)
- **AmbitiousF** Female's rating of male's ambition (1-10 scale)  
Đánh giá của người phụ nữ về tham vọng (sự cầu tiến) của đối phương nam?  
(Thang từ 1 đến 10)
- **SharedInterestsM** Male's rating of female's shared interests (1-10 scale)  
Đánh giá của người đàn ông về sở thích chung của đối phương nữ?  
(Thang từ 1 đến 10)
- **SharedInterestsF** Female's rating of male's shared interests (1-10 scale)  
Đánh giá của người phụ nữ về sở thích chung của đối phương nam?  
(Thang từ 1 đến 10)

## 2) Chọn biến – Mô tả

### ❖ Biến định tính

#### ➤ DecisionFemale

Người phụ nữ tham gia khảo sát có muốn một cuộc hẹn khác? (Giá trị Yes hoặc No)

#### ➤ RaceF

Chủng tộc của người phụ nữ? (Châu Á, Da đen, Da trắng, Latinh, Khác)

### ❖ Biến định lượng

#### ➤ LikeF

Mức độ thích đối phương của người phụ nữ tham gia cuộc hẹn đầu tiên như thế nào?  
(Thang từ 1 đến 10)

#### ➤ FunF

Cảm nhận của người phụ nữ về sự vui tính của đối phương nam?  
(Thang từ 1 đến 10)

#### ➤ AmbitiousF

Cảm nhận của người phụ nữ về tham vọng (sự cầu tiến) của đối phương nam?  
(Thang từ 1 đến 10)

### Bài tập 3

#### ❖ Lý do chọn:

Chọn 5 biến này vì muốn khảo sát xem mức độ thích (LikeF), sự vui tính hài hước (FunF), sự cầu tiến tham vọng (AmbitiousF) ảnh hưởng như thế nào đến quyết định có muốn một cuộc hẹn khác không người phụ nữ tham gia khảo sát. Ngoài ra còn muốn khảo sát xem nhóm người (chủng tộc – RaceF) tương quan như thế nào với quyết định muốn có một cuộc hẹn khác.

### 3) Phân tích thăm dò riêng từng biến

#### ❖ DecisionFemale

##### ➤ Tóm tắt bằng thống kê

```
> #####  
> # DecisionFemale  
> SLMuon <- count( ~DecisionFemale, data=SpeedDating); SLMuon  
n_No  
149  
>  
> TyLeMuon <- prop( ~DecisionFemale, data=SpeedDating); TyLeMuon  
prop_No  
0.5398551  
>  
> PhanTramMuon <- perc( ~DecisionFemale, data=SpeedDating); PhanTramMuon  
perc_No  
53.98551  
>  
> #.....  
> SLKoMuon <- count( ~DecisionFemale, data=SpeedDating, success = "Yes"); SLKoMuon  
n_Yes  
127  
>  
> TyLeKoMuon <- prop( ~DecisionFemale, data=SpeedDating, success = "Yes"); TyLeKoMuon  
prop_Yes  
0.4601449  
>  
> PhanTramKoMuon <- perc( ~DecisionFemale, data=SpeedDating, success = "Yes"); PhanTramKoMuon  
perc_Yes  
46.01449  
> |
```

Số lượng người phụ nữ tham gia khảo sát **muốn** một cuộc hẹn khác với đối phương nam trong cuộc hẹn đầu tiên là **127** người chiếm tỷ lệ khoảng **0,46** tương đương chiếm **46%** trong tổng số người tham gia khảo sát

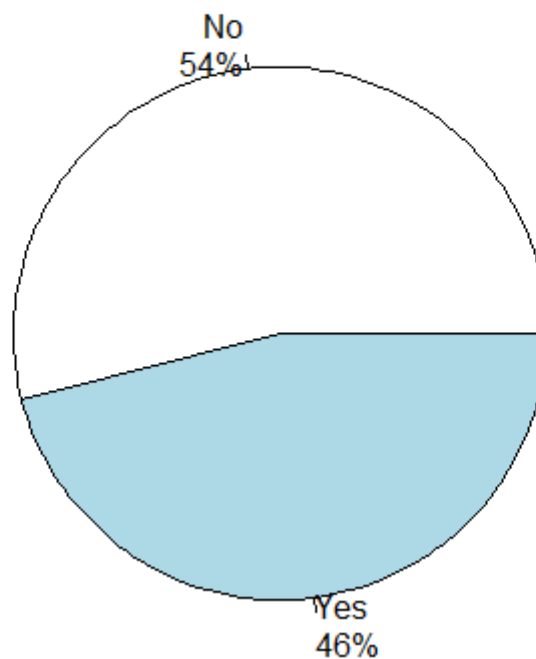
Chiều ngược lại số lượng người phụ nữ tham gia khảo sát **không muốn** một cuộc hẹn khác với đối phương nam trong cuộc hẹn đầu tiên là **149** người chiếm tỷ lệ khoảng **0,53** tương đương chiếm **53%**

### Bài tập 3

- Trình bày trực quan bằng các biểu đồ

```
> #.....  
> tblDecision <- table(SpeedDating$DecisionFemale)  
> pct <- round(tblDecision/sum(tblDecision)*100, digits = 1)  
> lblsp <- paste(names(tblDecision),"\n", pct, sep="")  
> lblsp <- paste(lblsp,"%",sep="")  
> pie(tblDecision, labels = lblsp, main="Pie chart of Decision Female")  
> |
```

**Pie Chart Of Decision Female**



- Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ tròn ta thấy những người phụ nữ tham gia khảo sát **muốn** một cuộc hẹn khác với đối phương nam trong cuộc hẹn đầu tiên **ít hơn** so với những người **không muốn**. Cụ thể ít hơn **22 người** chiếm khoảng **8%** trong 276 người được khảo sát

## Bài tập 3

### ❖ RaceF

#### ➤ Tóm tắt bằng thống kê

```
> #-----> # RaceF
> SLChauA <- count( ~RaceF, data=SpeedDating, success = 'Asian'); SLChauA
n_Asian
70
> TyLeChauA <- prop( ~RaceF , data=SpeedDating, success = 'Asian'); TyLeChauA
prop_Asian
0.2536232
> PhanTramChauA <- perc( ~RaceF, data=SpeedDating, success = 'Asian'); PhanTramChauA
perc_Asian
25.36232
> #-----> # RaceF
> SLDaDen <- count( ~RaceF, data=SpeedDating, success = 'Black'); SLDaDen
n_Black
15
> TyLeDaDen <- prop( ~RaceF , data=SpeedDating, success = 'Black'); TyLeDaDen
prop_Black
0.05434783
> PhanTramDaDen <- perc( ~RaceF, data=SpeedDating, success = 'Black'); PhanTramDaDen
perc_Black
5.434783
> #-----> # RaceF
> SLDaTrang <- count( ~RaceF, data=SpeedDating, success = 'Caucasian'); SLDaTrang
n_Caucasian
148
> TyLeDaTrang <- prop( ~RaceF , data=SpeedDating, success = 'Caucasian'); TyLeDaTrang
prop_Caucasian
0.5362319
> PhanTramDaTrang <- perc( ~RaceF, data=SpeedDating, success = 'Caucasian'); PhanTramDaTrang
perc_Caucasian
53.62319
> #-----> # RaceF
> SLLaTinh <- count( ~RaceF, data=SpeedDating, success = 'Latino'); SLLaTinh
n_Latino
23
> TyLeLaTinh <- prop( ~RaceF , data=SpeedDating, success = 'Latino'); TyLeLaTinh
prop_Latino
0.08333333
> PhanTramLaTinh <- perc( ~RaceF, data=SpeedDating, success = 'Latino'); PhanTramLaTinh
perc_Latino
8.333333
> #-----> # RaceF
> SLKhac <- count( ~RaceF, data=SpeedDating, success = 'other'); SLKhac
n_other
16
> TyLeKhac <- prop( ~RaceF , data=SpeedDating, success = 'other'); TyLeKhac
prop_other
0.05797101
> PhanTramKhac <- perc( ~RaceF, data=SpeedDating, success = 'other'); PhanTramKhac
perc_other
5.797101
> #-----> # RaceF
> SLKhongXacDinh <- count( ~RaceF, data=SpeedDating, success = ''); SLKhongXacDinh
n_
4
> TyLeKhongXacDinh <- prop( ~RaceF , data=SpeedDating, success = ''); TyLeKhongXacDinh
prop_
0.01449275
> PhanTramKhongXacDinh <- perc( ~RaceF, data=SpeedDating, success = ''); PhanTramKhongXacDinh
perc_
1.449275
> |
```

Cuộc khảo sát tiến hành trên 276 người phụ nữ trong đó có:

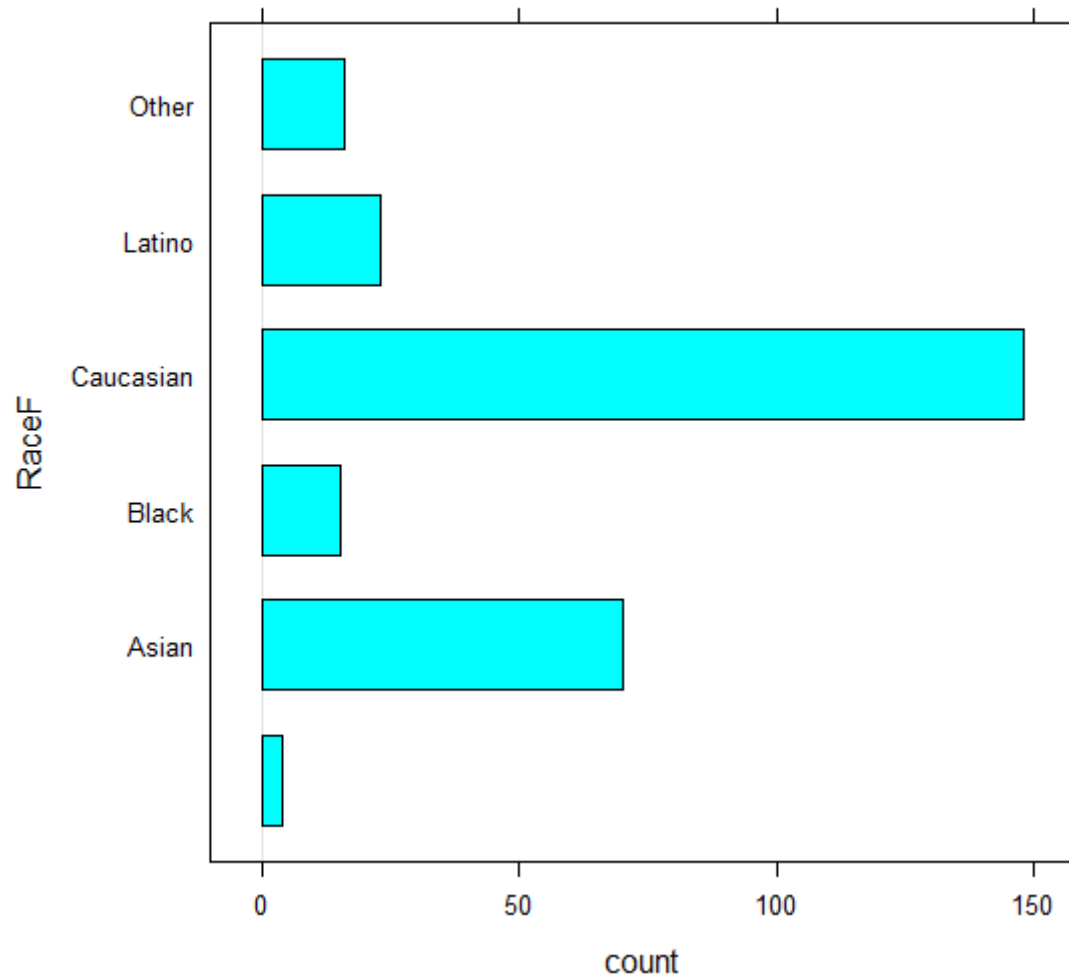
- 70 người châu Á chiếm tỷ lệ khoảng 0,254 tương đương 25,4%
- 15 người da đen chiếm tỷ lệ khoảng 0,054 tương đương 5,4%
- 148 người da trắng chiếm tỷ lệ khoảng 0,536 tương đương 53,6%
- 23 người La-tinh chiếm tỷ lệ khoảng 0,083 tương đương 8,3%
- 16 người thuộc nhóm khác chiếm tỷ lệ khoảng 0,058 tương đương 5,8%
- 4 người không xác định chiếm tỷ lệ khoảng 0,015 tương đương 1,5%



### Bài tập 3

- Trình bày trực quan bằng các biểu đồ

```
> bargraph( ~RaceF, data=SpeedDating, horizontal = TRUE)  
> |
```



- Nhận xét, đánh giá

Dựa vào số liệu thông kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát chủ yếu thuộc nhóm **người da trắng** (148 người da trắng chiếm **53,6%**). Thuộc nhóm người **da đen** là ít nhất (15 người da đen chiếm **5,4%**). Cuộc khảo vẫn chưa hoàn hảo khi vẫn **chưa lấy được thông tin** của (4 người không xác định chiếm **1,5%**)

### Bài tập 3

#### ❖ LikeF

##### ➤ Tóm tắt bằng thống kê

```
> #-----
> # LikeF
> tally(~LikeF, data=SpeedDating)
LikeF
  1    2    3    4    5    6    7    8  8.5    9   10 <NA>
  1    7   14   13   41   54   72   48    1   13    8    4
>
> sum(~LikeF, data=SpeedDating, na.rm=TRUE)
[1] 1731.5
>
> mean(SpeedDating$LikeF, na.rm=TRUE)
[1] 6.365809
> mean(~LikeF, data=SpeedDating, na.rm=TRUE)
[1] 6.365809
> |
```

##### • Theo thang điểm từ 1 đến 10 thì:

- 1 có 1 người
- 2 có 7 người
- 3 có 14 người
- 4 có 13 người
- 5 có 41 người
- 6 có 54 người
- 7 có 72 người
- 8 có 48 người
- 8.5 có 1 người
- 9 có 13 người
- 10 có 8 người
- Không xác định có 4 người

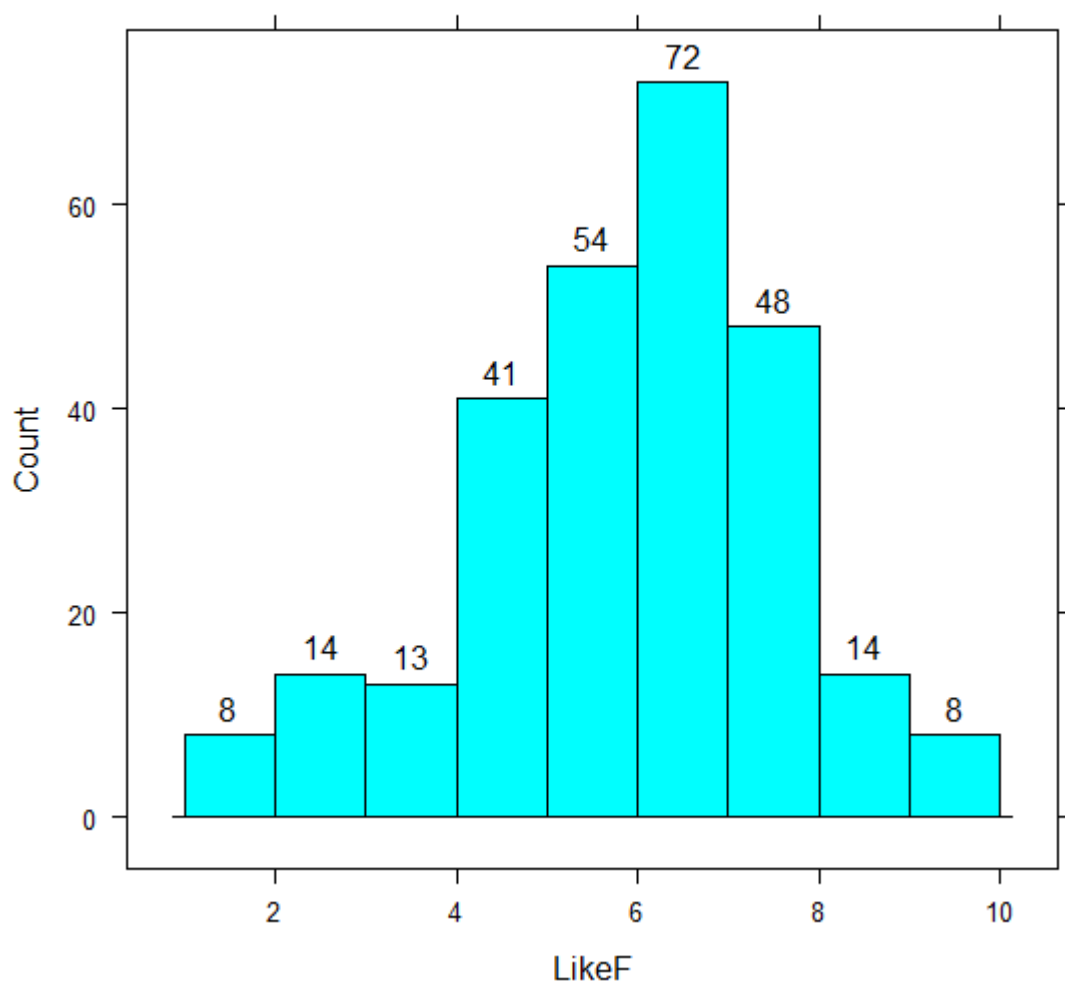
##### • Tổng điểm là: 1731,5

##### • Điểm trung bình là: 6,366

### Bài tập 3

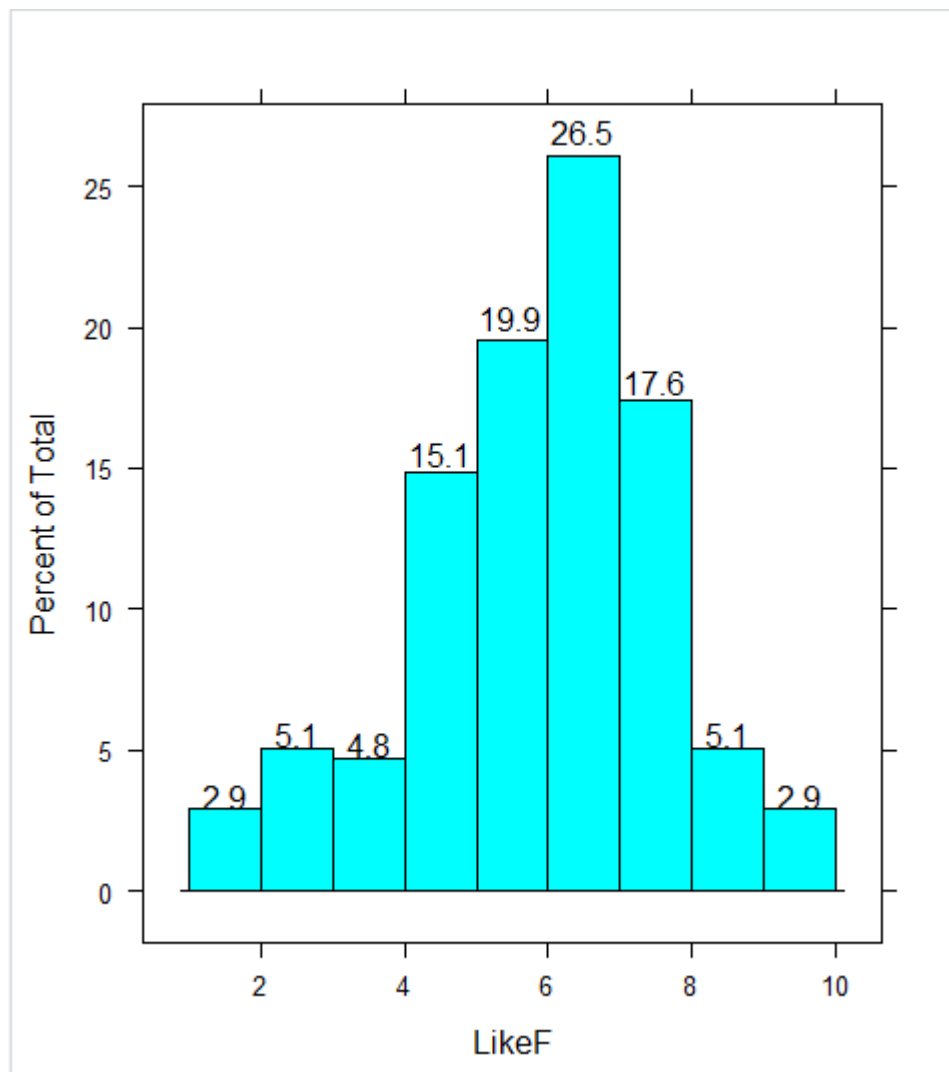
- Trình bày trực quan bằng các biểu đồ

```
> histogram(~LikeF, data=SpeedDating, label = TRUE, type = "count", breaks = 10)  
> |
```



### Bài tập 3

```
> histogram(~LikeF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)  
> |
```



#### ➤ Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát chủ yếu cho đối phương nam 7 điểm về mức độ thích (72 người). Và mức điểm này có số người chọn cao vọt so với những mức điểm khác. Biểu đồ lệch phải thể hiện người phụ nữ cho điểm khá cao về mức độ thích nên điểm trung bình khá cao là 6,366.

### Bài tập 3

#### ❖ FunF

##### ➤ Tóm tắt bằng thống kê

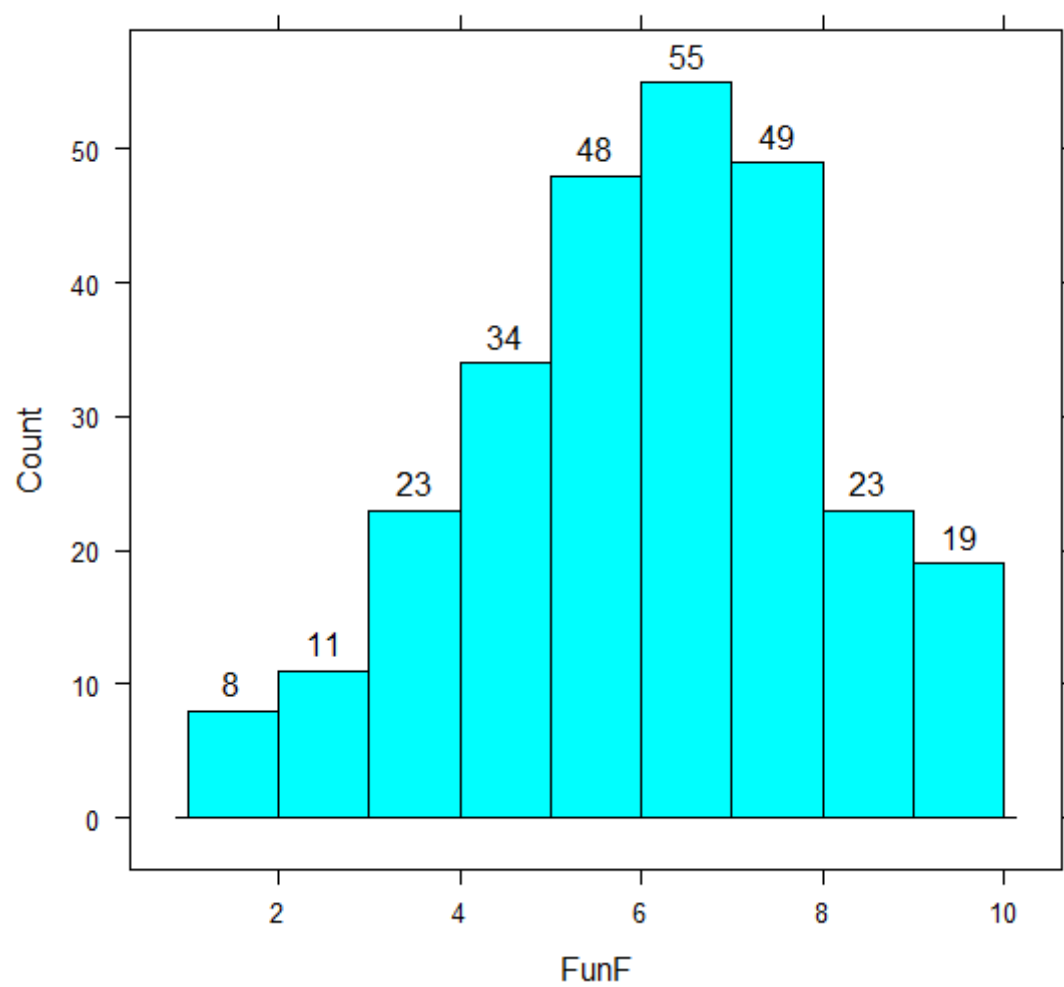
```
> #-----
> tally(~FunF, data=SpeedDating)
FunF
  1    2    3    4    5    6    7    8    9   10 <NA>
  1    7   11   23   34   48   55   49   23   19    6

>
> sum(~FunF, data=SpeedDating, na.rm=TRUE)
[1] 1772
>
> mean(SpeedDating$FunF, na.rm=TRUE)
[1] 6.562963
> mean( ~FunF, data=SpeedDating, na.rm=TRUE)
[1] 6.562963
> |
```

- Theo thang điểm từ 1 đến 10 thì:
  - 1 có 1 người
  - 2 có 7 người
  - 3 có 11 người
  - 4 có 23 người
  - 5 có 34 người
  - 6 có 48 người
  - 7 có 55 người
  - 8 có 49 người
  - 9 có 23 người
  - 10 có 19 người
  - Không xác định có 6 người
- Tổng điểm là: 1772
- Điểm trung bình là: 6,56

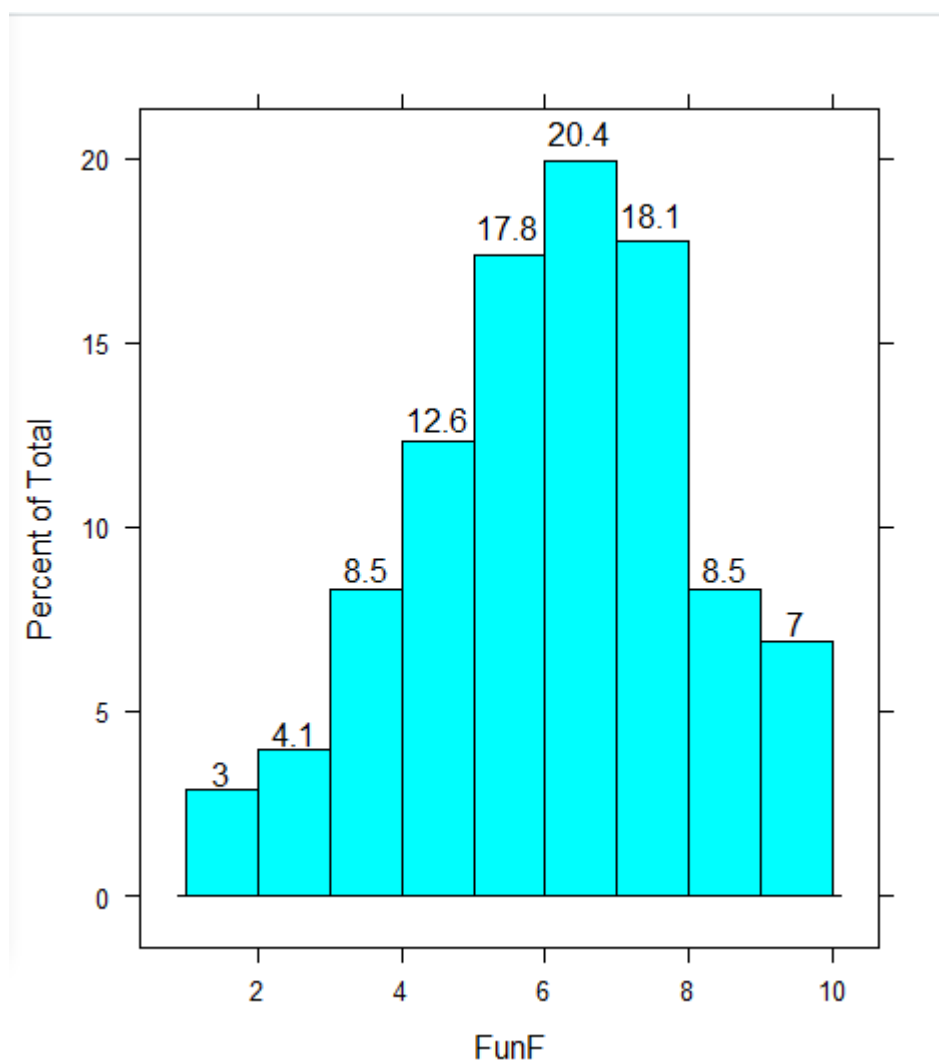
### Bài tập 3

- Trình bày trực quan bằng các biểu đồ



### Bài tập 3

```
> histogram(~FunF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)  
> |
```



#### ➤ Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát chủ yếu cho đối phương nam 7 điểm về mức độ hài hước, vui tính (55 người). Chỉ có một người cho 1 điểm. Biểu đồ lệch phải thể hiện người phụ nữ đánh giá khá cao về sự vui vẻ, hài hước của đối phương tham gia khảo sát. Điểm trung bình người phụ nữ cho đối phương về sự hài hước là khá cao 6,56

### Bài tập 3

#### ❖ AmbitiousF

##### ➤ Tóm tắt bằng thống kê

```
> #-----
> tally(~AmbitiousF, data=SpeedDating)
AmbitiousF
  1    2    3    4    5    6    7    8    9   10 <NA>
  1    2    3    4   35   30   48   65   45   33   10
>
> sum(~AmbitiousF, data=SpeedDating, na.rm=TRUE)
[1] 1976
>
> mean(SpeedDating$AmbitiousF, na.rm=TRUE)
[1] 7.428571
> mean(~AmbitiousF, data=SpeedDating, na.rm=TRUE)
[1] 7.428571
> |
```

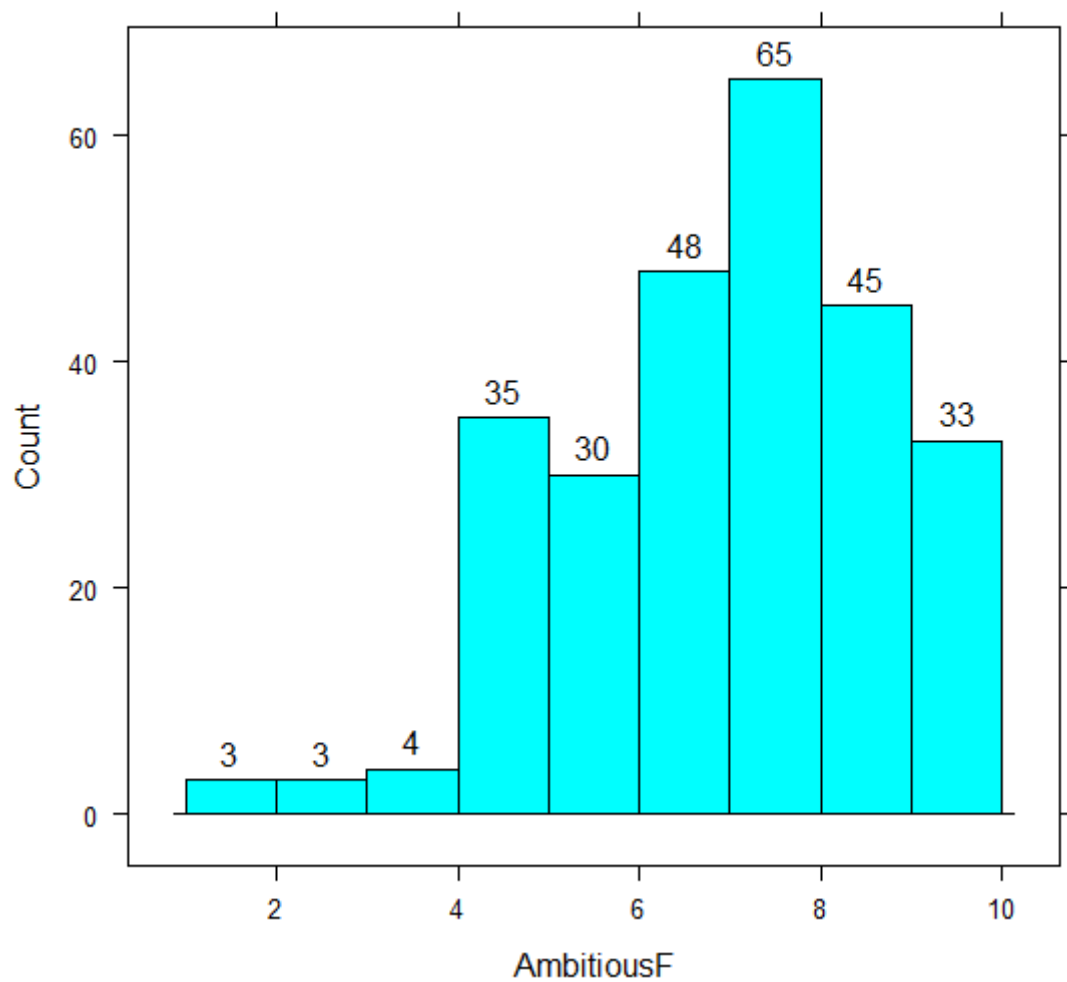
- Theo thang điểm từ 1 đến 10 thì:
  - 1 có 1 người
  - 2 có 2 người
  - 3 có 3 người
  - 4 có 4 người
  - 5 có 35 người
  - 6 có 30 người
  - 7 có 48 người
  - 8 có 65 người
  - 9 có 45 người
  - 10 có 33 người
  - Không xác định có 10 người
- Tổng điểm là: 1976
- Điểm trung bình là: 7,429



### Bài tập 3

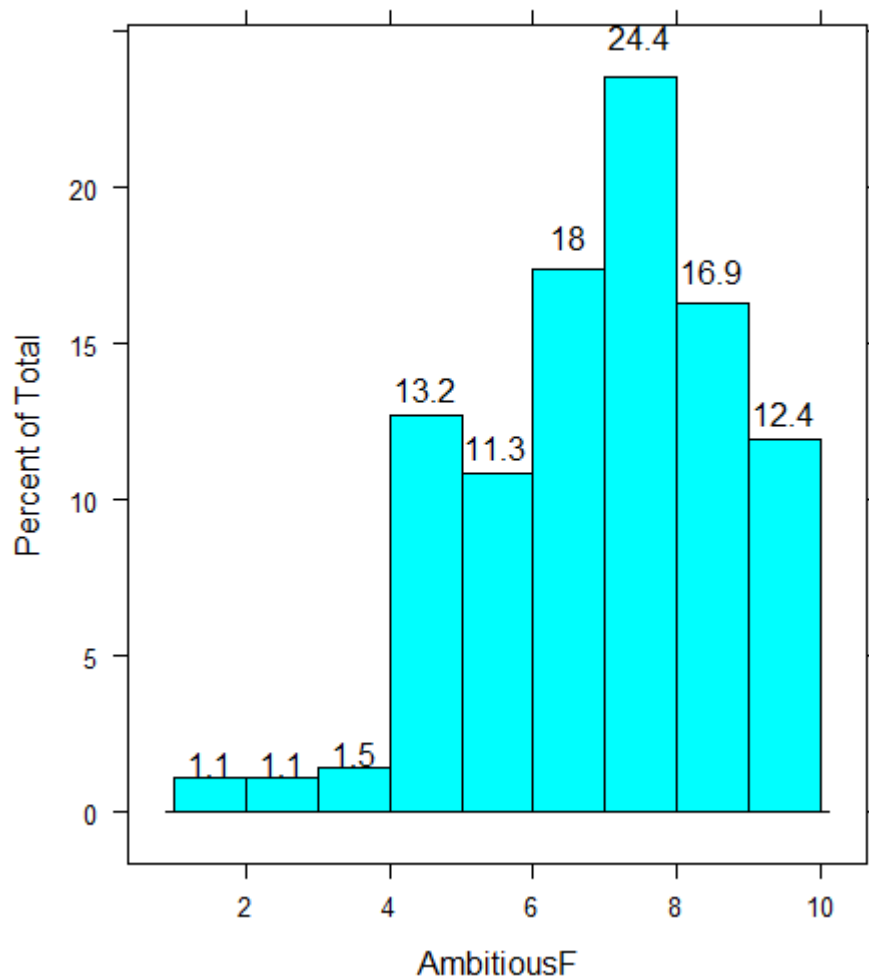
- Trình bày trực quan bằng các biểu đồ

```
> #.....  
> histogram(~AmbitiousF, data=SpeedDating, label = TRUE, type = "count", breaks = 10)  
> |
```



### Bài tập 3

```
> histogram(~AmbitiousF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)  
> |
```



#### ➤ Nhận xét, đánh giá

Dựa vào số liệu thông kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát chủ yếu cho đối phương nam 8 điểm về mức độ tham vọng, cầu tiến (65 người). Biểu đồ chệch lệch về bên phải khá lớn thể hiện người phụ nữ đánh giá cao về sự tham vọng, cầu tiến của đối phương tham gia khảo sát. Điểm trung bình người phụ nữ cho đối phương về sự tham vọng, cầu tiến là khá cao 7,429

#### 4) Phân tích thăm dò quan hệ giữa hai biến định tính:

Chọn hai biến DecisionFemale và RaceF

##### ❖ Tóm tắt bằng thống kê

```
> #####
> # Hai biến định tính
> tally(~DecisionFemale + RaceF, margins = TRUE, data = SpeedDating)
      RaceF
DecisionFemale Asian Black Caucasian Latino Other Total
No             3     35     7         85     12     7   149
Yes            1     35     8         63     11     9   127
Total          4     70    15        148     23    16   276

> #-----
> count( ~DecisionFemale | RaceF, data=SpeedDating)
      n_No.      n_No.Asian      n_No.Black n_No.Caucasian      n_No.Latino      n_No.Other
1         3         35         7         85         12         7

> prop( ~DecisionFemale | RaceF, data=SpeedDating)
      prop_No.      prop_No.Asian      prop_No.Black prop_No.Caucasian      prop_No.Latino      prop_No.Other
1 0.7500000    0.5000000    0.4666667    0.5743243    0.5217391    0.4375000

> perc( ~DecisionFemale | RaceF, data=SpeedDating)
      perc_No.      perc_No.Asian      perc_No.Black perc_No.Caucasian      perc_No.Latino      perc_No.Other
1 75.00000    50.00000    46.66667    57.43243    52.17391    43.75000

> #.....
> count( ~DecisionFemale | RaceF, data=SpeedDating, success = "Yes")
      n_Yes.      n_Yes.Asian      n_Yes.Black n_Yes.Caucasian      n_Yes.Latino      n_Yes.Other
1         1         35         8         63         11         9

> prop( ~DecisionFemale | RaceF, data=SpeedDating, success = "Yes")
      prop_Yes.      prop_Yes.Asian      prop_Yes.Black prop_Yes.Caucasian      prop_Yes.Latino      prop_Yes.Other
1 0.2500000    0.5000000    0.5333333    0.4256757    0.4782609    0.5625000

> perc( ~DecisionFemale | RaceF, data=SpeedDating, success = "Yes")
      perc_Yes.      perc_Yes.Asian      perc_Yes.Black perc_Yes.Caucasian      perc_Yes.Latino      perc_Yes.Other
1 25.00000    50.00000    53.33333    42.56757    47.82609    56.25000

> |
```

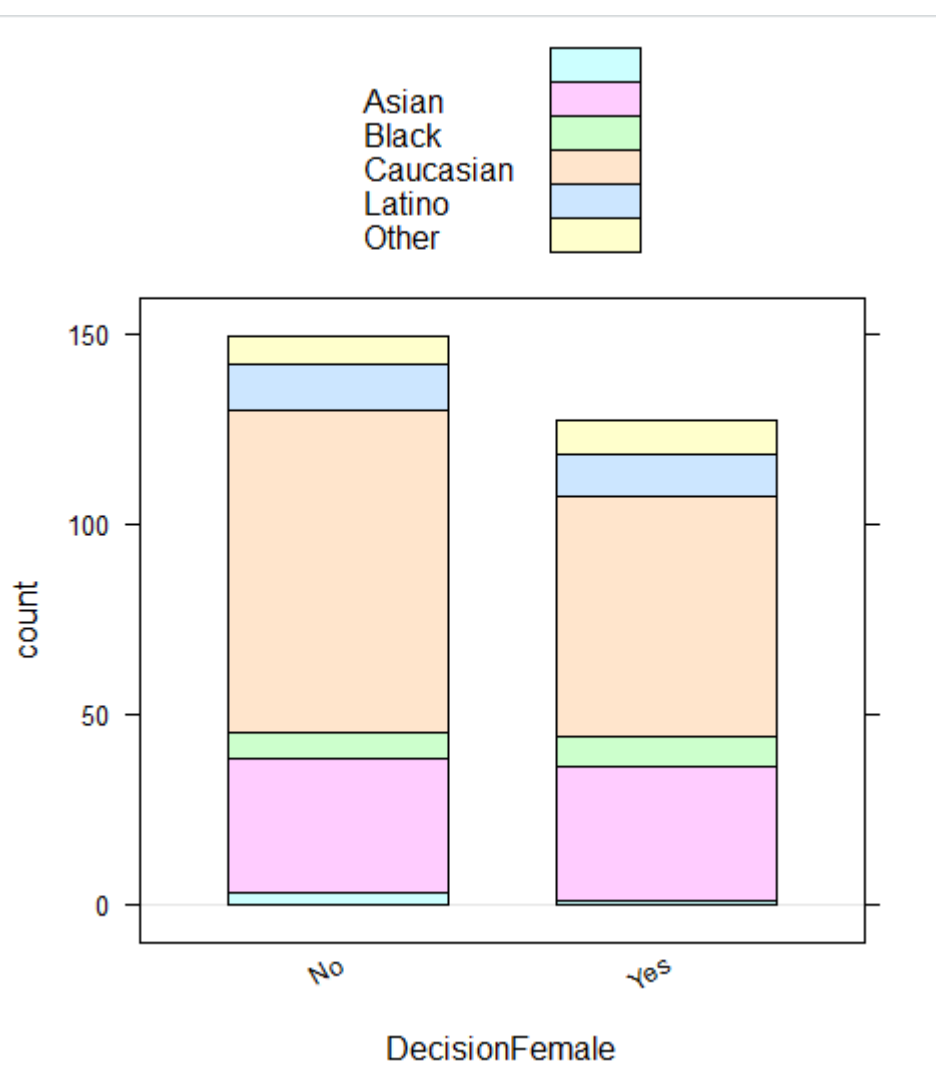
Trong số những người phụ nữ tham gia khảo sát:

- Người châu Á:
  - Không muốn một cuộc hẹn khác có 35 người chiếm tỷ lệ khoảng 0,5 tương đương 50%
  - Muốn có 35 người chiếm tỷ lệ khoảng 0,5 tương đương 50%
- Người da đen:
  - Không muốn một cuộc hẹn khác có 7 người chiếm tỷ lệ khoảng 0,47 tương đương 47%
  - Muốn có 8 người chiếm tỷ lệ khoảng 0,53 tương đương 53%
- Người da trắng:
  - Không muốn một cuộc hẹn khác có 85 người chiếm tỷ lệ khoảng 0,57 tương đương 57%
  - Muốn có 63 người chiếm tỷ lệ khoảng 0,43 tương đương 43%
- Người La-tinh:
  - Không muốn một cuộc hẹn khác có 12 người chiếm tỷ lệ khoảng 0,52 tương đương 52%
  - Muốn có 11 người chiếm tỷ lệ khoảng 0,48 tương đương 48%
- Người thuộc nhóm khác:
  - Không muốn một cuộc hẹn khác có 7 người chiếm tỷ lệ khoảng 0,43 tương đương 43%
  - Muốn có 9 người chiếm tỷ lệ khoảng 0,56 tương đương 56%

### Bài tập 3

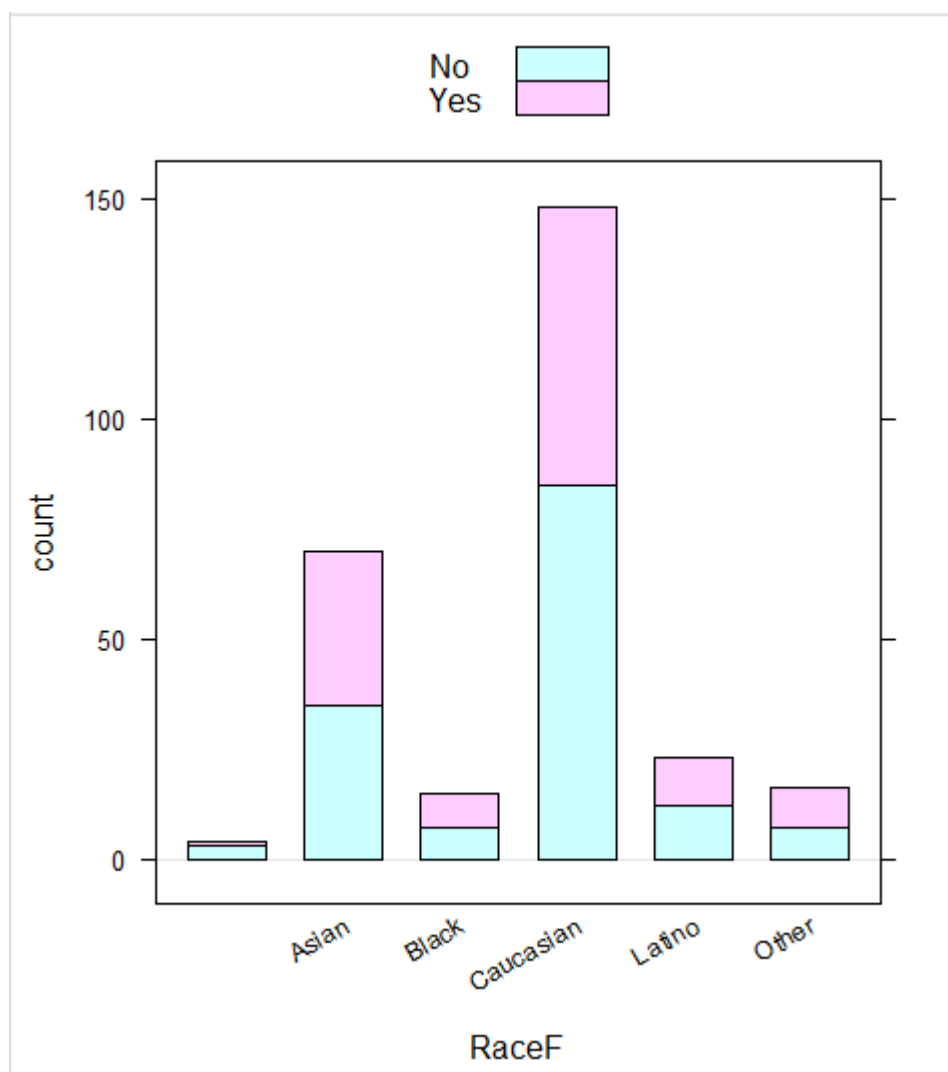
#### ❖ Trình bày trực quan bằng các biểu đồ

```
> bargraph(~DecisionFemale, groups = RaceF, stack = TRUE, auto.key = TRUE, data = SpeedDating)  
> |
```



### Bài tập 3

```
> bargraph(~RaceF, groups = DecisionFemale, stack = TRUE, auto.key = TRUE, data = SpeedDating)  
> |
```



#### ❖ Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát ta thấy sự chênh lệch giữa muốn và không muốn có một cuộc hẹn khác ở các nhóm người là không bao nhiêu. Thậm chí ở nhóm người **châu Á** là **như nhau**. Ở nhóm người **da trắng** và **La-tinh không muốn** là **cao hơn** chiếm tỷ lệ không muốn lần lượt là **0,57** và **0,52**. Còn ở nhóm người **da đen** và **nhóm khác** thì **muốn** có cuộc hẹn khác lại **cao hơn** chiếm tỷ lệ muốn lần lượt là **0,53** và **0,56**.

## 5) Phân tích thăm dò quan hệ giữa 1 biến định tính, 1 biến định lượng

Chọn hai biến DecisionFemale và LikeF

### ❖ Tóm tắt bằng thống kê

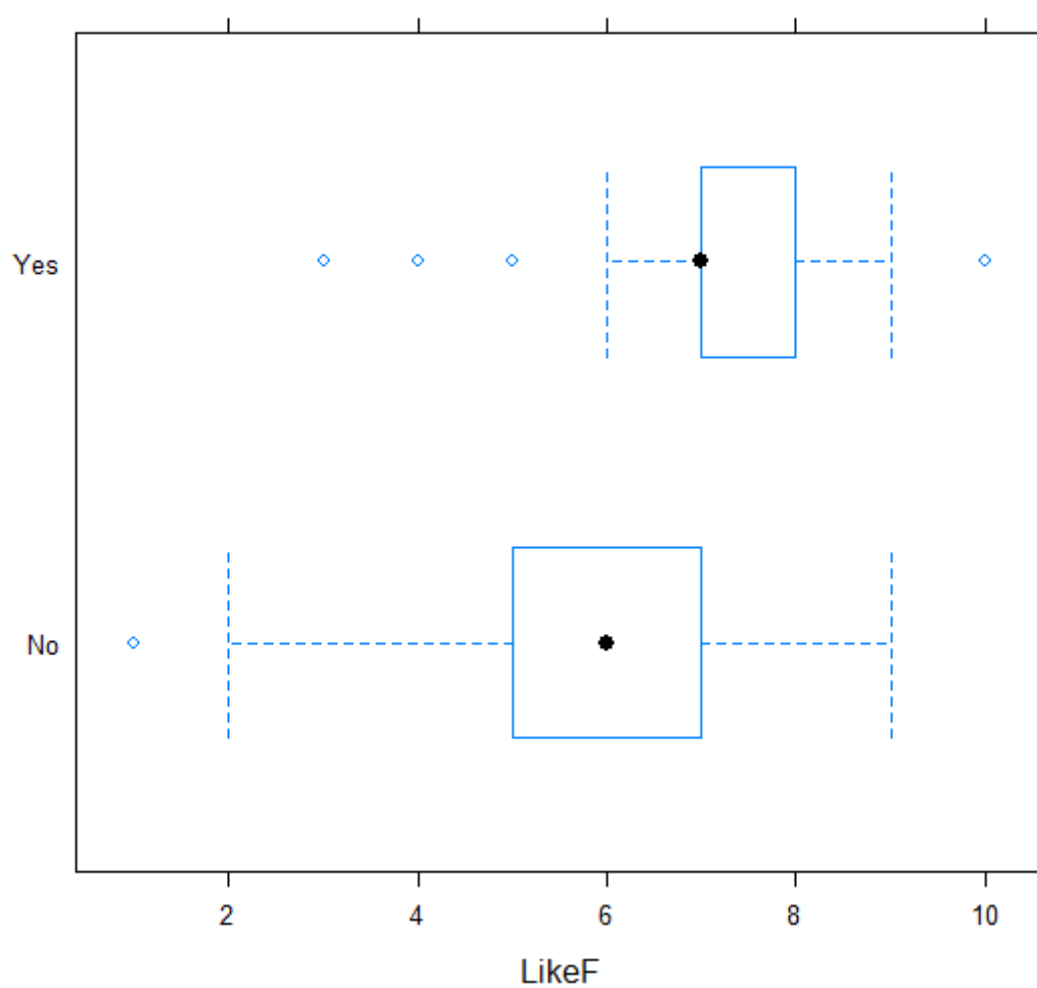
```
> #####
> # Một biến định tính DecisionFemale, một biến định lượng LikeF
> favstats(~LikeF | DecisionFemale, data = SpeedDating)
  DecisionFemale min Q1 median Q3 max      mean      sd    n missing
1             No   1  5      6  7   9 5.557823 1.708677 147      2
2             Yes   3  7      7  8  10 7.316000 1.269264 125      2
> mean(~LikeF | DecisionFemale, data = SpeedDating, na.rm = TRUE)
      No      Yes
5.557823 7.316000
> diff(mean(~LikeF | DecisionFemale, data = SpeedDating, na.rm = TRUE))
      Yes
1.758177
> |
```

Điểm trung bình người phụ nữ tham gia khảo sát thể hiện **thích** đối phương và **muốn** có một cuộc hẹn khác (7,32) cao hơn điểm trung bình người phụ nữ tham gia khảo sát thể hiện **thích** đối phương và **không** muốn có một cuộc hẹn khác (5,56). Chênh lệch nhau 1,76

### Bài tập 3

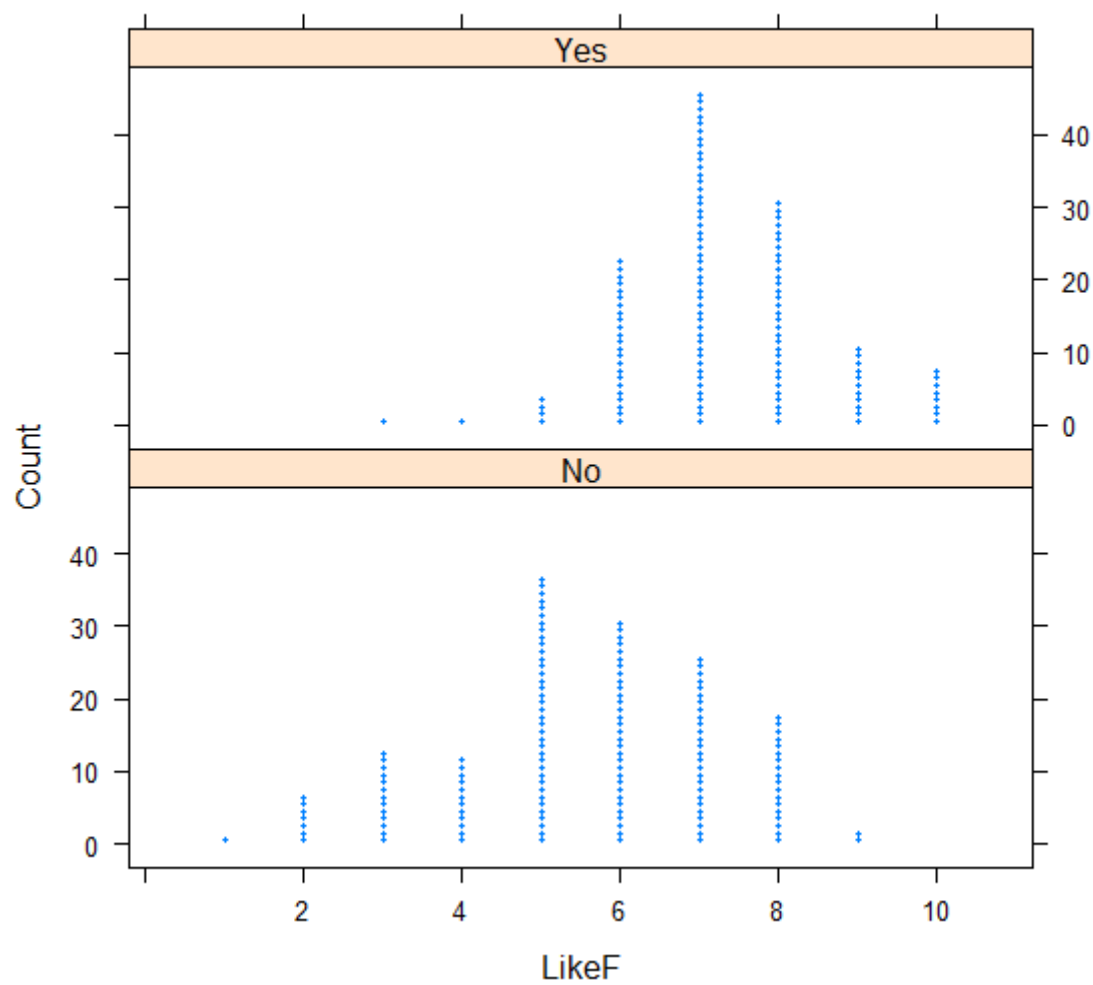
- ❖ Trình bày trực quan bằng các biểu đồ

```
> bwplot(DecisionFemale~LikeF, data = SpeedDating)  
> |
```



### Bài tập 3

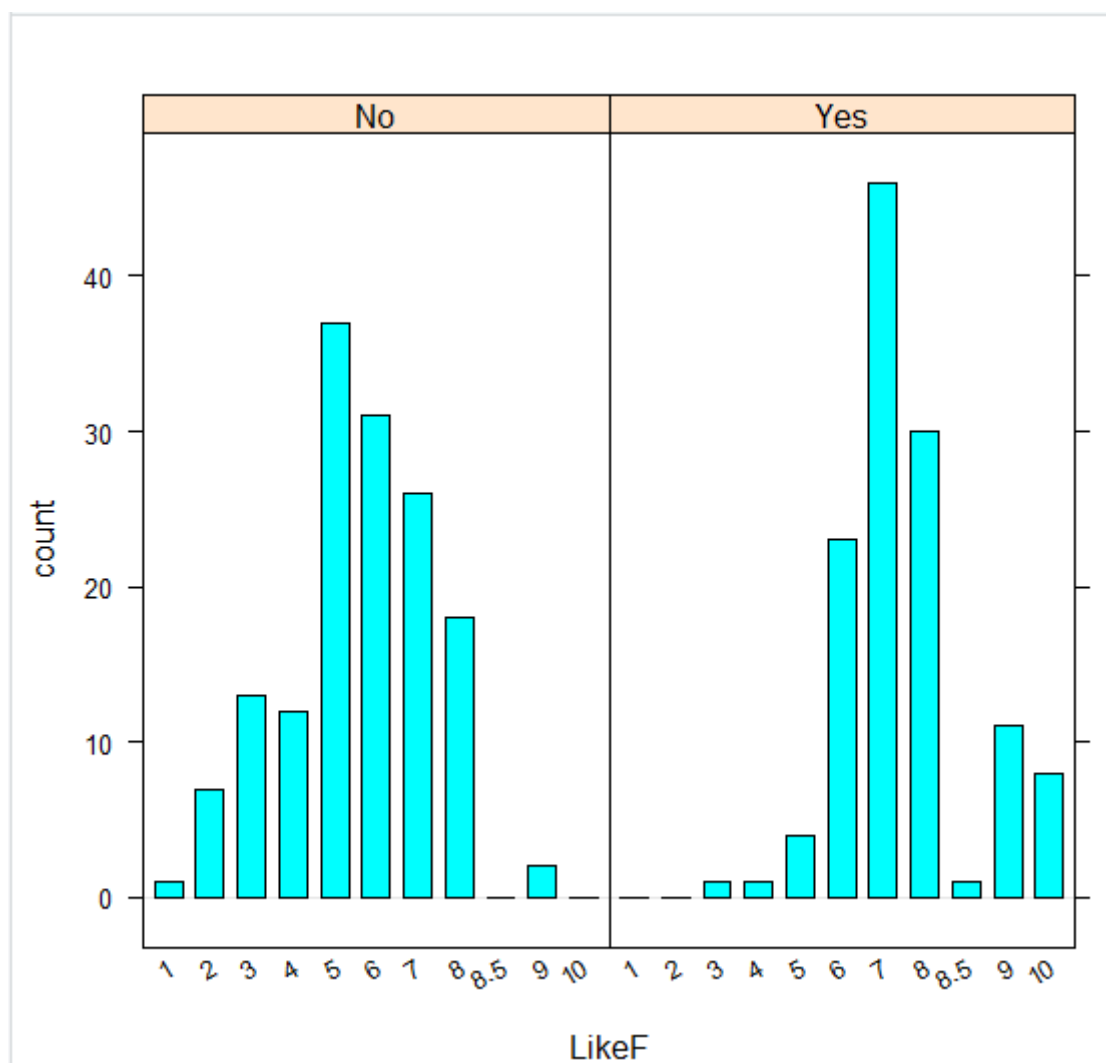
```
> dotPlot(~LikeF | DecisionFemale, layout = c(1, 2), width = 1, cex = 1, data = SpeedDating)
> |
```





### Bài tập 3

```
> bargraph(~LikeF | DecisionFemale, data = SpeedDating)  
> |
```



#### ❖ Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát khi cho điểm cao về mức độ thích sẽ muốn có một cuộc hẹn khác. Ngược lại thì khi cho điểm thấp thì không muốn có một cuộc hẹn khác.

## 6) Phân tích thăm dò quan hệ giữa 2 biến định lượng:

Chọn hai biến LikeF và AmbitiousF

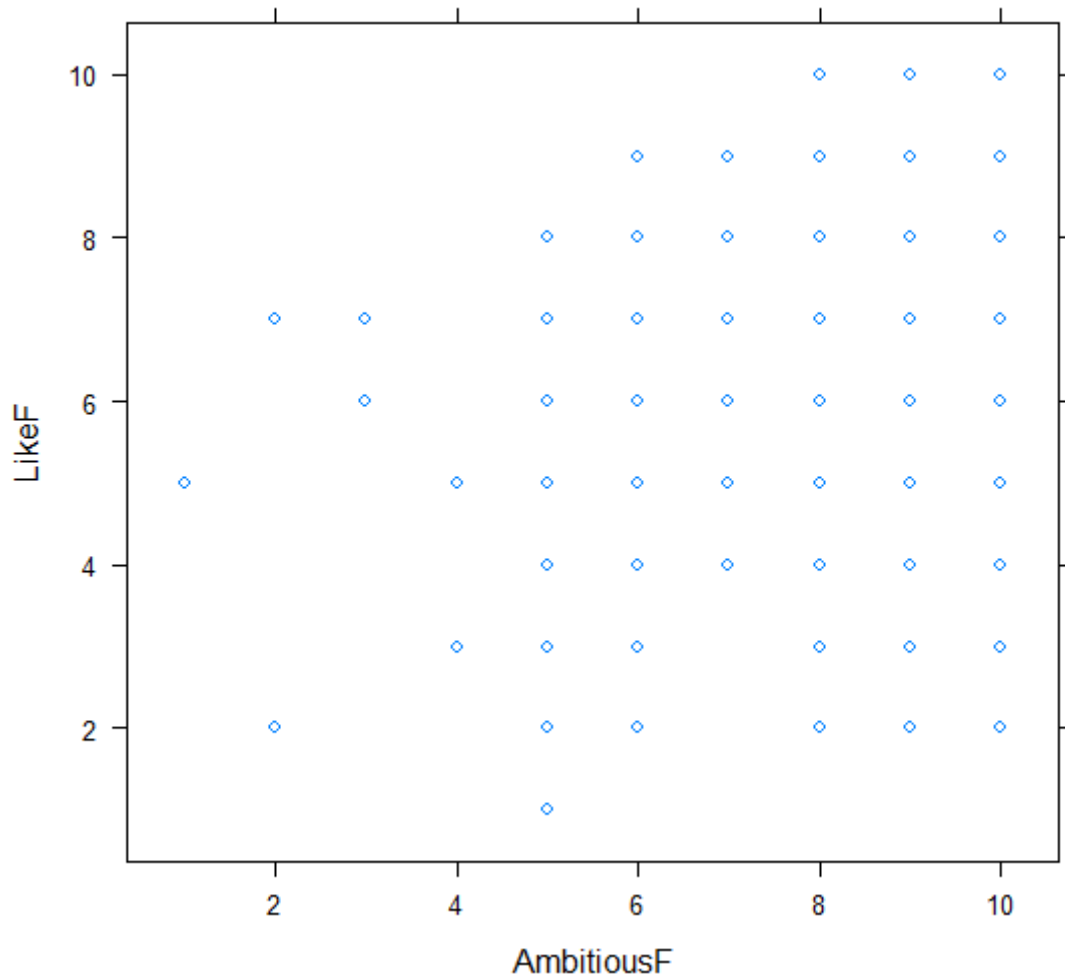
### ❖ Tóm tắt bằng thống kê

```
> #####  
> # Hai biến định lượng:  
> cor(LikeF~AmbitiousF, data = SpeedDating, use="complete.obs")  
[1] 0.2789275  
> |
```

Tương quan giữa hai biến định lượng là  $0,279 > 0$ . Khi AmbitiousF tăng thì LikeF tăng

### ❖ Trình bày trực quan bằng các biểu đồ

```
> #-----  
> xyplot(LikeF~AmbitiousF, data = SpeedDating)  
> |
```



### ❖ Nhận xét, đánh giá

Dựa vào số liệu thống kê và biểu đồ ta thấy người phụ nữ tham gia khảo sát có xu hướng thích người đàn ông có tính cầu tiến, có tham vọng. Nhưng điều này chỉ ở mức độ vừa phải vì có khá nhiều người đàn ông được người phụ nữ đánh giá là có tính cầu tiến tham vọng nhưng lại cho điểm thể hiện mức độ thích không cao.

### Bài tập 3

#### 7) Code R:

```
#Load package Lock5withR để sử dụng
library(Lock5withR)
#Load package mosaic để sử dụng
library(mosaic)
#Load package mosaicData để sử dụng
library(mosaicData)
# Khai báo dataset SpeedDating
data(SpeedDating)
# Xem dataset của SpeedDating
View(SpeedDating)
#Xem thông tin mô tả dataset SpeedDating
?SpeedDating
#####
# DecisionFemale
SLMuon <- count( ~DecisionFemale, data=SpeedDating); SLMuon

TyLeMuon <- prop( ~DecisionFemale, data=SpeedDating); TyLeMuon

PhanTramMuon <- perc( ~DecisionFemale, data=SpeedDating); PhanTramMuon

#.....
SLKoMuon <- count( ~DecisionFemale, data=SpeedDating, success = "Yes"); SLKoMuon

TyLeKoMuon <- prop( ~DecisionFemale, data=SpeedDating, success = "Yes"); TyLeKoMuon

PhanTramKoMuon <- perc( ~DecisionFemale, data=SpeedDating, success = "Yes"); PhanTramKoMuon

#.....
tblDecision <- table(SpeedDating$DecisionFemale)
pct <- round(tblDecision/sum(tblDecision)*100, digits = 1)
lblsp <- paste(names(tblDecision),"\n", pct, sep="")
lblsp <- paste(lblsp,"%",sep="")
pie(tblDecision, labels = lblsp, main="Pie Chart Of Decision Female")
#-----
# RaceF
SLChauA <- count( ~RaceF, data=SpeedDating, success = 'Asian'); SLChauA

TyLeChauA <- prop( ~RaceF , data=SpeedDating, success = 'Asian'); TyLeChauA

PhanTramChauA <- perc( ~RaceF, data=SpeedDating, success = 'Asian'); PhanTramChauA
```

### Bài tập 3

```
#.....
SLDaDen <- count( ~RaceF, data=SpeedDating, success = 'Black'); SLDaDen

TyLeDaDen <- prop( ~RaceF , data=SpeedDating, success = 'Black'); TyLeDaDen

PhanTramDaDen <- perc( ~RaceF, data=SpeedDating, success = 'Black'); PhanTramDaDen

#.....
SLDaTrang <- count( ~RaceF, data=SpeedDating, success = 'Caucasian'); SLDaTrang

TyLeDaTrang <- prop( ~RaceF , data=SpeedDating, success = 'Caucasian'); TyLeDaTrang

PhanTramDaTrang <- perc( ~RaceF, data=SpeedDating, success = 'Caucasian'); PhanTramDaTrang

#.....
SLLaTinh <- count( ~RaceF, data=SpeedDating, success = 'Latino'); SLLaTinh

TyLeLaTinh <- prop( ~RaceF , data=SpeedDating, success = 'Latino'); TyLeLaTinh

PhanTramLaTinh <- perc( ~RaceF, data=SpeedDating, success = 'Latino'); PhanTramLaTinh

#.....
SLKhac <- count( ~RaceF, data=SpeedDating, success = 'Other'); SLKhac

TyLeKhac <- prop( ~RaceF , data=SpeedDating, success = 'Other'); TyLeKhac

PhanTramKhac <- perc( ~RaceF, data=SpeedDating, success = 'Other'); PhanTramKhac

#.....
SLKhongXacDinh <- count( ~RaceF, data=SpeedDating, success = ''); SLKhongXacDinh

TyLeKhongXacDinh <- prop( ~RaceF , data=SpeedDating, success = ''); TyLeKhongXacDinh

PhanTramKhongXacDinh <- perc( ~RaceF, data=SpeedDating, success = ''); PhanTramKhongXacDinh

#.....
bargraph( ~RaceF, data=SpeedDating, horizontal = TRUE)

#-----
# LikeF
tally(~LikeF, data=SpeedDating)

sum(~LikeF, data=SpeedDating, na.rm=TRUE)
```

### Bài tập 3

```
mean(SpeedDating$LikeF, na.rm=TRUE)
mean( ~LikeF, data=SpeedDating, na.rm=TRUE)

#.....
histogram(~LikeF, data=SpeedDating, label = TRUE, type = "count", breaks = 10)

histogram(~LikeF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)

#-----
# FunF
tally(~FunF, data=SpeedDating)

sum(~FunF, data=SpeedDating, na.rm=TRUE)

mean(SpeedDating$FunF, na.rm=TRUE)
mean( ~FunF, data=SpeedDating, na.rm=TRUE)

#.....
histogram(~FunF, data=SpeedDating, label = TRUE, type = "count", breaks = 10)

histogram(~FunF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)

#-----
# AmbitiousF
tally(~AmbitiousF, data=SpeedDating)

sum(~AmbitiousF, data=SpeedDating, na.rm=TRUE)

mean(SpeedDating$AmbitiousF, na.rm=TRUE)
mean( ~AmbitiousF, data=SpeedDating, na.rm=TRUE)

#.....
histogram(~AmbitiousF, data=SpeedDating, label = TRUE, type = "count", breaks = 10)

histogram(~AmbitiousF, data=SpeedDating, label = TRUE, type = "percent", breaks = 10)

#####
# Hai biến định tính
tally(~DecisionFemale + RaceF, margins = TRUE, data = SpeedDating)

#-----
count( ~DecisionFemale | RaceF, data=SpeedDating)
```

### Bài tập 3

```
prop( ~DecisionFemale | RaceF, data=SpeedDating)

perc( ~DecisionFemale| RaceF, data=SpeedDating)
#.....
count( ~DecisionFemale | RaceF, data=SpeedDating, success = "Yes")

prop( ~DecisionFemale | RaceF, data=SpeedDating, success = "Yes")

perc( ~DecisionFemale| RaceF, data=SpeedDating, success = "Yes")

#-----
bargraph(~DecisionFemale, groups = RaceF, stack = TRUE, auto.key = TRUE, data = SpeedDating)

bargraph(~RaceF, groups = DecisionFemale, stack = TRUE, auto.key = TRUE, data = SpeedDating)
#-----
#####
# Một biến định tính DecisionFemale, một biến định lượng LikeF
favstats(~LikeF | DecisionFemale, data = SpeedDating)
mean(~LikeF | DecisionFemale, data = SpeedDating, na.rm = TRUE)
diff(mean(~LikeF | DecisionFemale, data = SpeedDating, na.rm = TRUE))

#-----
bwplot(DecisionFemale~LikeF, data = SpeedDating)
dotPlot(~LikeF | DecisionFemale, layout = c(1, 2), width = 1, cex = 1, data = SpeedDating)
bargraph(~LikeF | DecisionFemale, data = SpeedDating)
#####
# Hai biến định lượng LikeF và AmbitiousF
cor(LikeF~AmbitiousF, data = SpeedDating, use="complete.obs")

cor(AmbitiousF~LikeF, data = SpeedDating, use="complete.obs")

#-----
xyplot(LikeF~AmbitiousF, data = SpeedDating)
```

#### 8) Nguồn tham khảo:

<https://www.rdocumentation.org/packages/Lock5withR/versions/1.2.2/topics/SpeedDating>

<https://www.rdocumentation.org/packages/mosaic/versions/1.1.1>

[https://en.wikipedia.org/wiki/Speed\\_dating](https://en.wikipedia.org/wiki/Speed_dating)