# Problem Set 2

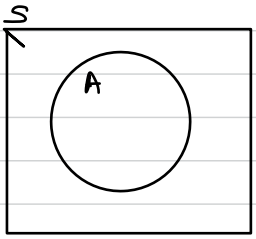
Aditya Mhaske.

**Q.1>**

Given conditions,

$P(S) = 1$ ——— (1)

a) $A \subset S$




$P(A^c) = 1 - P(A)$

$P(A^c) + P(A) = 1 = P(A \cup A^c)$

here $P(S) = 1$ ——— from (1)

$P(\phi) = 1 - P(S)$

$\therefore P(\phi) = 0$

b) IF $A \subset B$, then $P(A) \leq P(B)$

then

$B = A \cup (B \cap A^c)$

$P(A) + P(B \cap A^c) = P(A \cup (B \cap A^c))$

$= P(B)$

Probability lies between $(0 < P < 1)$

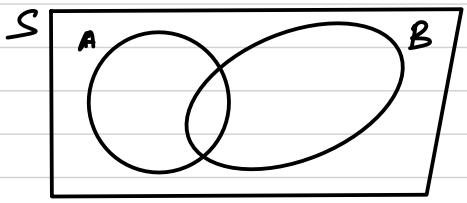
$\therefore P(B \cap A^c) \geq 0$

$\therefore P(A) \leq P(B)$

c) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

here A & B are events

$\therefore A^c, B^c, A \cup B, A \cap B, A^c \cap B, B^c \cap A$ are also events




as per venn diagram

$A \cap B$, $A \cap B^c$ & $B \cap A^c$ are disjoint

$\therefore P(A) = P((A \cap B^c) \cup (A \cap B))$

$P(B) = P((B \cap A^c) \cup (A \cap B))$

$\therefore P(A) + P(B) - P(A \cap B) = P((A \cap B^c) \cup (A \cap B)) + P((B \cap A^c) \cup (A \cap B)) - P(A \cap B)$

$= P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$

$= P((P \cap B^c) \cup (B \cap A^c) \cup (A \cap B))$

$= P(A \cup B)$

$\therefore P(A) + P(B) - P(A \cap B) = P(A \cup B)$

Q. 2

coin tossed 5 times

$S = \_\_\_\_\_ \ (2^5) = 32$ ways

a) $A = \{$ 4 coins shows head $\}$

$= \{ {}^5C_4 \} = 5$

$P(A) = 5/32 = \boxed{0.156}$

b) $B = \{$ there are more heads than tails $\}$

for majority there are 3 conditions

No. of heads = 3, 4, 5

$= {}^5C_3 + {}^5C_4 + {}^5C_5 \rightarrow (HHHHH)$

$= 10 + 5 + 1$

$= 16$

$P(B) = 16/32 = \boxed{0.5}$

c) $D = \{$ Atleast 3 tails $\}$

$\therefore$ Tails = 3, 4, 5

$= {}^5C_3 + {}^5C_4 + {}^5C_5$

$= 10 + 5 + 1$

$= 16$

$P(D) = 16/32 = \boxed{0.5}$

d) $A^c \cup D$

$P(A^c) = 1 - P(A) = 1 - (5/32) = 27/32$

$P(D) = 16/32$

$P(A^c) = \{$ No. of heads = 1, 2, 3, 5, 0 $\}$

$P(D) = \{$ No of tails = 3, 4, 5 $\}$

$P(A^c \cap D) = P(A^c) + P(D) - P(A^c \cap D)$

$= 16/32$

$P(A^c \cup D) = P(A^c) + P(D) - P(A^c \cap D)$

$= 27/32 + 16/32 - 16/32$

$= 27/32$

$P(A^c \cup D) = \boxed{0.84375}$

e) $B \cup D$

$P(B \cup D) = P(B) + P(D) + P(B \cap D)$

here $P(B \cap D) = 0$ because $B \cap D = \phi$

$\therefore P(P \cup B) = 0.5 + 0.5$

$= \boxed{1.0}$

Q.3

$P(A) + P(A^c) = 1$ ——— (1)

$P(A) = 0.6$ , $P(A^c) = 0.4$ ——— From (1)

$P(B) = 0.7$ $P(B^c) = 0.3$ ——— from (1)

$P(A^c \cap B^c) = 0.12$

a) if A & B are disjoint sets

then

$P(A \cap B) + P(A) + P(B) = P(A \cup B)$

$0 + 0.7 + 0.6 = P(A \cup B)$

$1.3 = P(A \cup B)$

↳ Probability

Probability $\leq 1$

As well as

$P(A' \cap B') = 0.12$

$\therefore P(A \cap B) \neq 0$

$P(A^c \cap B^c) = P((A \cup B)^c)$

$= 1 - P(A \cup B)$

$0.12 = 1 - P(A) - P(B) + P(A \cap B)$

$(0.30 + 0.12) = P(A \cap B)$

$\therefore P(A \cap B) = 0.42$ ——— (1)

for A & B to disjoint: $P(A \cap B) = 0$

$\therefore$ here A & B are not disjoint.
(As $P(A \cap B) \neq 0$)

b) $P(A \cup B^c)$

$P(B) = P(A \cap B) + P(A^c \cap B)$

$0.7 = 0.42 + P(A^c \cap B)$

$\therefore P(A^c \cap B) = 0.28$

$P(A \cup B^c) = P((A^c \cap B)^c)$

$P(A \cup B^c) = 1 - P(A^c \cap B)$

$P(A \cup B^c) = 0.72$

c) A & B are disjoint events

$\therefore P(A) \cdot P(B) = P(A \cap B)$

$P(A) \cdot P(B) = 0.7 \times 0.6$

$= 0.42$

$P(A \cap B) = 0.42$ ——— from eq" (1)

$\therefore P(A) \cdot P(B) = P(A \cap B)$

thus,

**Events A & B are independent**

d) conditional Probability $(A|B)$

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$= 0.42 / 0.7$

$= 0.6$

**$P(A/B) = 0.6$**

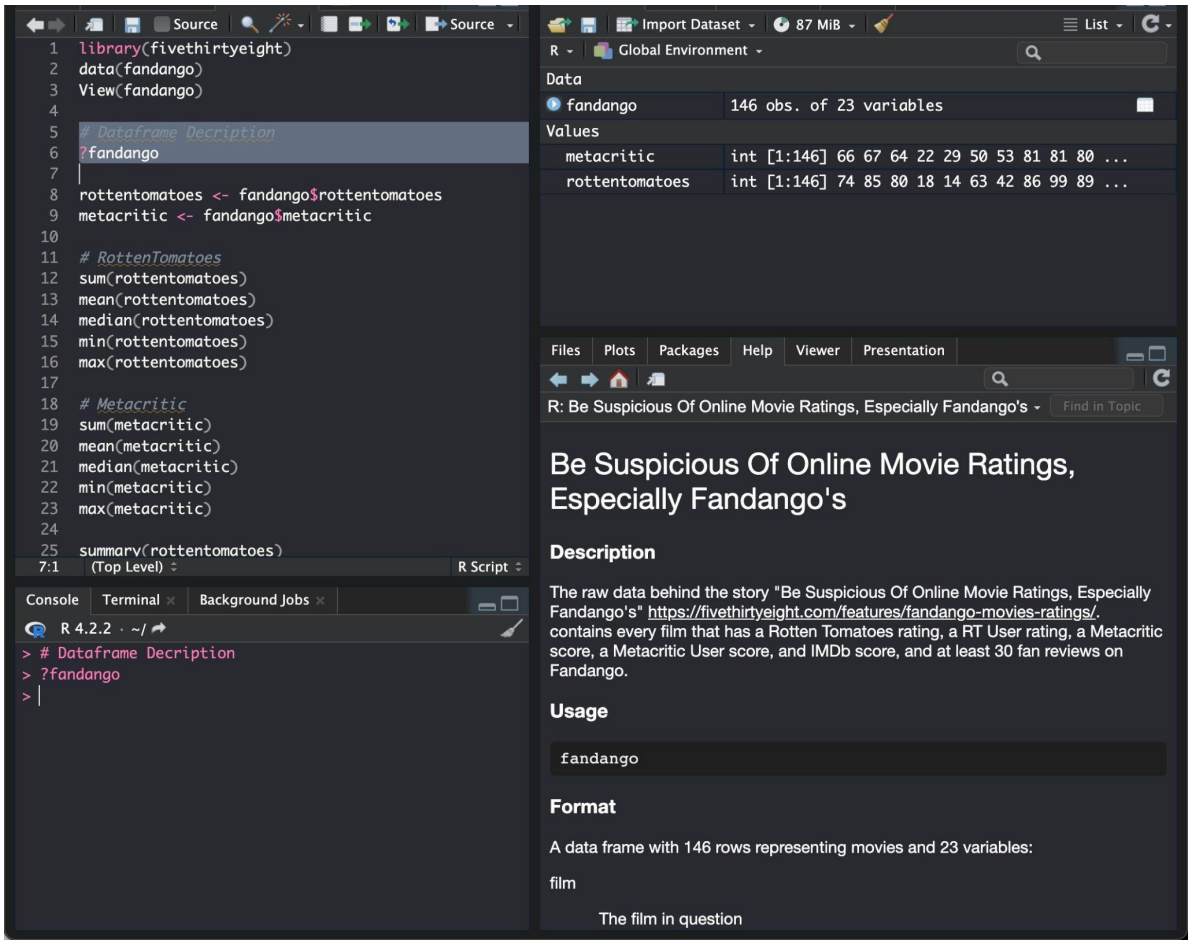
**Question 4:**

**a) Read the description of the data frame and briefly comment on the information it provides.**

Solution:

Data Description:

- Columns = 23, and Rows = 146
- The data frame has film names, their year, Rotten Tomatoes ratings, Metacritic ratings, IMDB scores, user scores, and fan reviews
- Collection from Fandango



**b) Create an object from variable rottentomatoes and another from variable metacritic. For each find the sum, average, median, minimum, and maximum values, and report those values.**

Solution:
For rottentomatoes

- sum= 8884
- mean= 60.84932
- median= 63.5
- min= 5
- max= 100

For Metacritic

- sum= 8586
- mean= 58.80822
- median= 59
- min= 13
- max= 94

### Code:

---

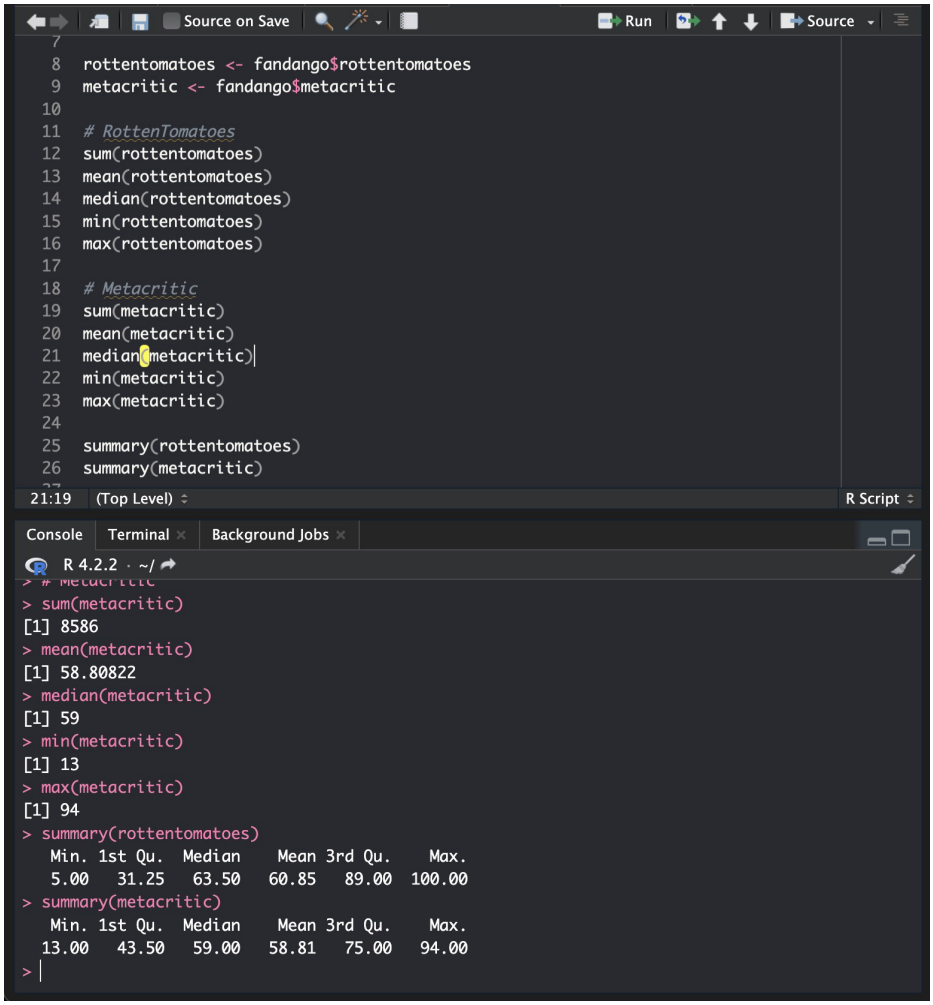
```
rottentomatoes <- fandango$rottentomatoes
metacritic <- fandango$metacritic

# RottenTomatoes
sum(rottentomatoes)
mean(rottentomatoes)
median(rottentomatoes)
min(rottentomatoes)
max(rottentomatoes)

# Metacritic
sum(metacritic)
mean(metacritic)
median(metacritic)
min(metacritic)
max(metacritic)

summary(rottentomatoes)
summary(metacritic)
```

**Output**

---

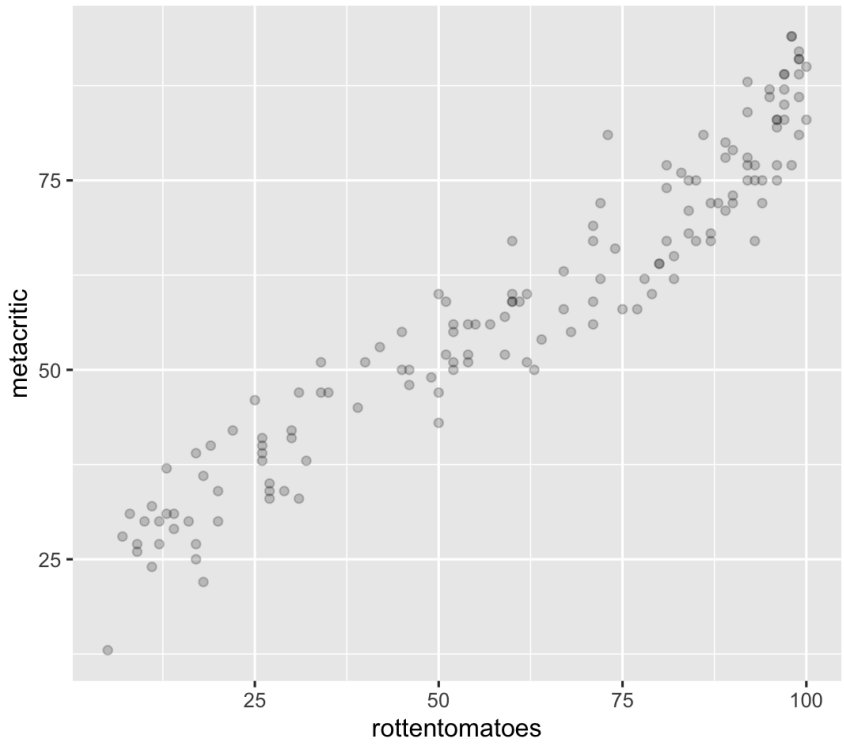
```
> sum(rottentomatoes)
[1] 8884
> mean(rottentomatoes)
[1] 60.84932
> median(rottentomatoes)
[1] 63.5
> min(rottentomatoes)
```

```
[1] 5
> max(rottentomatoes)
[1] 100
> # Metacritic
> sum(metacritic)
[1] 8586
> mean(metacritic)
[1] 58.80822
> median(metacritic)
[1] 59
> min(metacritic)
[1] 13
> max(metacritic)
[1] 94
> summary(rottentomatoes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.00   31.25   63.50   60.85   89.00  100.00
> summary(metacritic)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  13.00   43.50   59.00   58.81   75.00   94.00
```



**c) Using the code and explanations from SIDS, section 2.3 (this is your second textbook) create a scatterplot for rottentomatoes against metacritic. Comment on your findings.**

Solution:
Scatterplot:




Code:

```
# 4.c Scatterplot
ggplot(data = fandango,
       mapping = aes(
         x = rottentomatoes,
         y = metacritic)) + geom_point(alpha = 0.2)
```

Description:

- Point shows the movie/flim
- Almost positive correlation between both axis.
- Both range from 0 to 100

**d) Using SIDS, section 2.7 and 2.8, obtain a boxplot and a barplot rottentomatoes. Comment on your findings.**

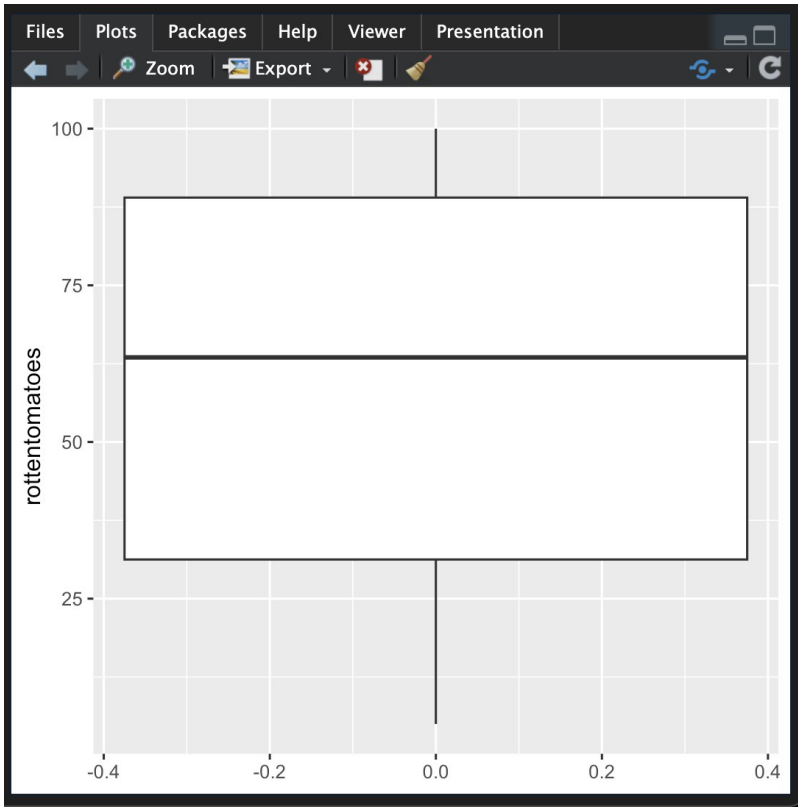
Solution:
D.1 = Boxplot

Description:

- Most of the ratings come between the 60-80 range. And Approximately 70 is the median.
- The boxplot also demonstrates that some of the points outside the outliers indicate that some films have extremely high or low ratings.
- A boxplot is a standardized way of displaying the distribution of data based on a five-number summary ("minimum", first quartile [Q1], median, third quartile [Q3], and "maximum")

Code:

```
#4.d Boxplot
ggplot(data = fandango,
       mapping = aes(
         y = rottentomatoes)) + geom_boxplot()
```



D.2 = Barplot

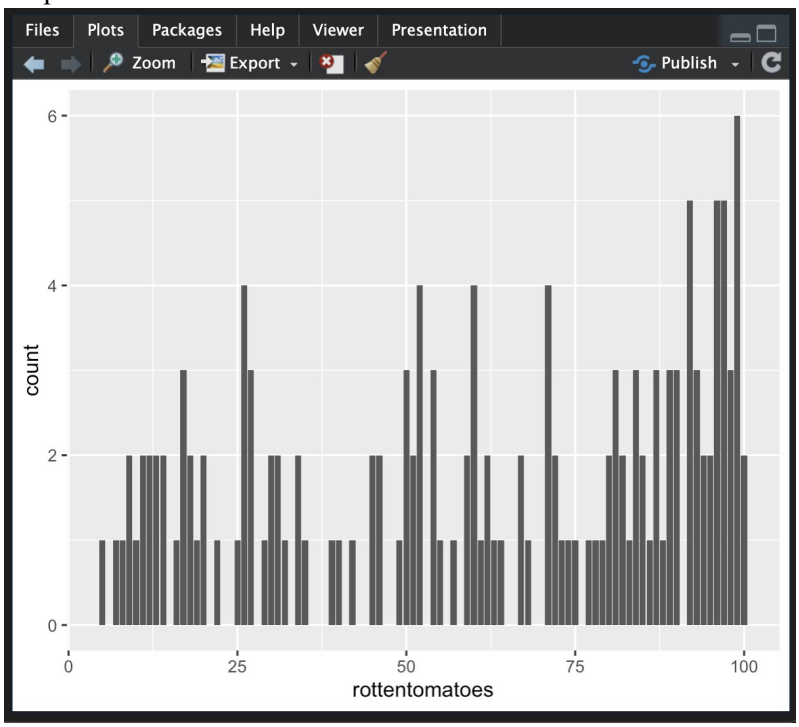
Description:

- Provides Generalization of frequency of rating
- We can see that most of the movies have ratings between 60 to 80. And there are some movies with almost 100 and less than 60 ratings.
- Barplot provides a general idea of the frequency of each rating.

Code:

```
#4.d Barplot
ggplot(data = fandango,
       mapping = aes(
         x = rottentomatoes)) + geom_bar()
```

Graph:



**e) Using SIDS, section 2.7, obtain a side-by-side boxplot of rottentomatoes scores split by fandango_stars (make sure use the factor version of fandango_stars)**

Description:

- Provides information related to rottentomato score and fandango_stars
- The variable fandango_stars is converted to factor categorical variable using factor() function. The boxplots are ordered according to fandango_stars.

Code:

```
#4.e Side-by-side Boxplot
ggplot(data = fandango,
       mapping = aes(
         x = factor(fandango_stars),
         y = rottentomatoes)) + geom_boxplot()
```