Set1

Function Understanding Tasks

count

```
starwars_count = count(starwars, eye_color, name = 'num')
```

Function: count
Text: Create num from COUNT(eye_color)

1. Regarding the operation performed by this function, which of the following statements are correct:
   a. The function does not affect the number of rows in the input table
   b. The function does not affect the number of columns in the input table
   c. The table *starwars_count* contains the column *eye_color*, and there may be duplicate values in the column *eye_color*
   d. The table *starwars_count* does not contain the column *eye_color*
   e. None of the above is correct

Answer: e

subset

```
airquality_subset = subset(airquality, Temp > 90 | Day == 3)
```

Function: subset
Text: Keep rows where Temp > 90 or Day is 3

2. Regarding the operation performed by this function, which of the following statements are correct:
   a. The function does not affect the number of rows in the input table
   b. The function does not affect the number of columns in the input table
   c. The values of the column *Temp* in the table *airquality_subset* are all greater than 90
   d. The function means to filter out rows that satisfy any one of the two conditions
   e. None of the above is correct

Answer: bd

summarise

```
mtcars_sm = summarise(mtcars, mean = mean(disp))
```

Function: summarise, mean
Text: Create mean from mean(disp)

3. Regarding the operation performed by this function, which of the following statements are correct:
   a. The function does not affect the number of rows in the input table
   b. The function does not affect the number of columns in the input table
   c. The table *mtcars_sm* contains only one column *mean*, in which there is only one cell, and its value is the mean of the column *disp*
   d. The function creates a new column *mean*, in which the values of all cells are the same, namely the mean of the column *disp*
   e. None of the above is correct

Answer: c

left_join

```
band_join = left_join(members,instruments, by='name')
```

(Note that the *members* and *instruments* are the *band_members.csv* and the *band_instruments.csv*, respectively)
Function: left_join
Text: Left join with members and instruments on name==name

4. Regarding the operation performed by this function, which of the following statements are correct:
   a. The column *name* is the key of left_join (namely, join by the column *name*)
   b. All the columns in *members* and *instruments* exist in the table *band_join*
   c. The table *band_join* contains all cells in the column *name* of *members*
   d. The table *band_join* contains all cells in the column *name* of *instruments*
   e. None of the above is correct

Answer: abc

separate

```
CO2_separate = separate(CO2, uptake, into=c("int", "decimal"), sep="[.]")
```

Function: separate
Text: Split uptake on delimiters matching '.' into "int" and "decimal" columns

5. Regarding the operation performed by this function, which of the following statements are correct:
   a. The function does not affect the number of rows in the input table
   b. The function does not affect the number of columns in the input table
   c. The function splits the column *uptake* in the table *CO2* into two columns of *int* and *decimal* according to the separator '.', and deletes *uptake*
   d. The table *CO2_separate* contains the column *uptake*
   e. None of the above is correct

Answer: ac

Script Understanding Tasks

repo: wuft\Power_of_Irma

script: bailey_code.R, table: Energy-Poverty 32641 homes.csv

```
library(dplyr)

bailey = read.csv("Energy-Poverty 32641 homes.csv")
landlords = dplyr::count(bailey, OWNERNME1, sort = TRUE)
landlords = subset(landlords, n>1)
by_owner = group_by(bailey, OWNERNME1)
utilities = dplyr::summarise(by_owner, cost = sum(Unit.Utilities.Cost))
ownercost = left_join(landlords, utilities, by = 'OWNERNME1')
ownercost = rename(ownercost, "num_properties"="n")
ownercost = mutate(ownercost, cost_per_property=cost / num_properties)
```

Functions:

1. read.csv
2. count
3. subset
4. group_by
5. summarise
6. sum
7. left_join
8. rename
9. mutate

Text:

```
bailey(L3,489R*39C): Create table from "Energy-Poverty 32641 homes.csv"
landlords(L4_1,411R*2C): Create n from COUNT(OWNERNME1) in bailey(L3)
landlords(L4_2.411R*2C): Sort rows by -n in landlords(L4_1)
landlords(L5,39R*2C): Keep rows where n>1 in landlords(L4_2)
by_owner(L6,489R*39C): Convert bailey(L3) into a grouped table by
  OWNERNME1
utilities(L7,411R*2C): Create cost from sum(Unit.Utilities.Cost) in
  by_owner(L6)
ownercost(L8,39R*3C): Left join with landlords(L5) and utilities(L7)
  on OWNERNME1==OWNERNME1
ownercost(L9,39R*3C): Rename n to "num_properties" in ownercost(L8)
ownercost(L10,39R*4C): Create cost_per_property from
  cost/num_properties in ownercost(L9)
```

Questions:

1. Is utilities(L7) created by landlords(L5) in one or more data transformations?

a. Yes
b. No

2. How many data transformations are performed from table bailey(L3) to landlords(L5)?
   a. 1
   b. 2
   c. 3
   d. 4
3. From ownercost(L8) to ownercost(L10), which columns are created?
   a. OWNERNME1
   b. n
   c. Unit.Utilities.Cost
   d. num_properties
   e. cost_per_property
4. From the beginning of the script execution, which data tables contribute to the creation of utilities (L7)?
   a. bailey(L3)
   b. landlords(L4_1)
   c. landlords(L4_2)
   d. landlords(L5)
   e. by_owner(L6)
5. Which data tables in the script are used as input tables for data transformations more than once (at least twice)?
   a. bailey(L3)
   b. landlords(L5)
   c. by_owner(L6)
   d. utilities(L7)
   e. ownercost(L8)

Answers:

1. b
2. c
3. e
4. ae
5. a

1. How helpful were those textual/visual descriptions for completing the tasks?

○ 1 (Not Helpful) ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 (Extremely Helpful)

2. How interpretable were those textual/visual descriptions?

○ 1 (Not Interpretable) ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 (Extremely Interpretable)