Training Set

Function Understanding Tasks

mutate

```
longley_mutate = mutate(longley, people=Unemployed+Employed, .keep="unused")
```

Function: mutate
Text: Create people from Unemployed+Employed and Delete Unemployed, Employed

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 function adds the columns *Unemployed* and *Employed* to create the column *people* in the table *longley_mutate*
   d. The function concatenates the column *Unemployed* with *Employed* to create the column *people* in the table *longley_mutate*
   e. None of the above is correct

Answer: ac

rbind

```
women_rbind = rbind(women, c(62, 130))
```

Function: rbind
Text: Create a row (62, 130)

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. This function selects the 62nd and 130th rows of the table *women*
   d. This function creates a column with the value: 62, 130
   e. This function creates a row with the value: 62, 130

Answer: be

rename

```
iris_rename = rename(iris, 'petal_length' = 'Petal.Length')
```

Function: rename
Text: Rename Petal.Length to petal_length

3. Regarding the operation performed by this function, which of the following statements are correct:
   a. Replace all cells with the value of 'Petal.Length' in the table *iris* with 'petal_length'
   b. Filter out the row where the value of the column *Petal.Length* equals the value of the column *petal_length*
   c. Rename the column *petal_length* to *Petal.Length*
   d. Create a column *petal_length*, which value is consistent with the column *Petal.Length*
   e. None of the above is correct

Answer: e

arrange

```
trees_arrange = arrange(trees, -`Girth`)
```

Function: arrange
Text: Sort rows by –Girth

4. Regarding the operation performed by this function, which of the following statements are correct:
   a. The function does not affect the number of columns in the input table
   b. The function deletes the column *Girth*
   c. This operation only keeps the column *Girth*
   d. In the column *Girth* of the table *trees_arrange*, there may be a cell with a value of 8 in the front of a cell with a value of 10.
   e. None of the above is correct

Answer: a

unique

```
IS_unique = unique(InsectSprays)
```

Function: unique
Text: Remove duplicate rows

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. There must be no duplicate rows in the table *IS_unique*
   d. None of the above is correct

Answer: bc

Script Understanding Tasks

```
library(dplyr)

warpbreaks = read.csv("warpbreaks.csv")
warpbreaks = unique(warpbreaks)
wb_tens = rename(warpbreaks, tens=tension)
wb_tens = group_by(wb_tens, tens)
wb_tens = mutate(wb_tens, count = n())
wb_tens = ungroup(wb_tens)
wb_tens_r = mutate(wb_tens, rate=breaks/count)
wb_l = rbind(warpbreaks, list(70, 'A', 'L'))
wb_sort = arrange(wb_l, -breaks)
```

Functions:

1. read.csv
2. unique
3. rename
4. group_by
5. mutate
6. ungroup
7. rbind
8. arrange

Text:

```
warpbreaks(L3,54R*3C): Create table from "warpbreaks.csv"
warpbreaks(L4,49R*3C): Remove duplicate rows in
        warpbreaks(L3)
wb_tens(L5,49R*3C): Rename tension to "tens" in warpbreaks(L4)
wb_tens(L6,49R*3C): Convert wb_tens(L5) into a grouped table by tens
wb_tens(L7,49R*4C): Create count from n() in wb_tens(L6)
wb_tens(L8,49R*4C): Remove group in wb_tens(L7)
wb_tens_r(L9,49R*5C): Create rate from breaks/count in wb_tens(L8)
wb_l(L10,50R*3C): Create a row (70, 'A', 'L') in warpbreaks(L4)
wb_sort(L11,50R*3C): Sort rows by -breaks in wb_l(L10)
```

Questions:

1. Is wb_l(L10) created by wb_tens(L7) in one or more data transformations?
   a. Yes
   b. No
2. How many data transformations are performed from table warpbreaks(L4) to wb_sort(L11)?
   a. 2
   b. 4
   c. 5
   d. 6

e. 7

3. From wb_tens(L5) to wb_tens_r(L9), which columns are created?
   a. tension
   b. tens
   c. count
   d. breaks
   e. rate
4. From the beginning of the script execution, which data tables contribute to the creation of wb_l(L10)?
   a. warpbreaks(L4)
   b. wb_tens(L5)
   c. wb_tens(L6)
   d. wb_tens(L7)
   e. wb_tens_r(L9)
5. Which data tables in the script are used as input tables for data transformations more than once (at least twice)?
   a. warpbreaks(L3)
   b. warpbreaks(L4)
   c. wb_tens(L5)
   d. wb_tens(L6)
   e. wb_tens_r(L9)

Answers:

1. b
2. a
3. ce
4. a
5. b

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)