Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | To pass 80% or higher

Retake the assignment in 23h 51m

Go to next item

1. 1 / 1 point

A data analyst is working with a dataset in R that has more than 50,000 observations. Why might they choose to use a tibble instead of the standard data frame? Select all that apply.

- [x] Tibbles automatically only preview the first 10 rows of data

Correct
Tibbles make printing in R easier. They won't accidentally overload the data analyst's console because they're automatically set to pull up only the first 10 rows and as many columns as fit on screen.

- [ ] Tibbles can automatically change the names of variables
- [x] Tibbles automatically only preview as many columns as fit on screen

Correct
Tibbles make printing in R easier. They won't accidentally overload the data analyst's console because they're automatically set to pull up only the first 10 rows and as many columns as fit on screen.

- [ ] Tibbles can create row names

2. A data analyst is exploring their data to get more familiar with it. They want a preview of just the first six rows to get a better idea of how the data frame is laid out. What function should they use? 1 / 1 point

- ( ) print()
- ( ) colnames()
- (•) head()
- ( ) preview()

Correct
The head() function can be used to return a preview of the first six rows of a data frame. This is a useful way to explore a data frame and get more familiar with how it is structured.

3. You are working with the ToothGrowth dataset. You want to use the skim_without_charts() function to get a comprehensive view of the dataset. Write the code chunk that will give you this view. 1 / 1 point

```
1  skim_without_charts(ToothGrowth )
```

Run

Reset

```
── Data Summary ────────────────────────
                           Values
Name                       ToothGrowth
Number of rows             60
Number of columns          3
_____________________
Column type frequency:
  factor                   1
  numeric                  2
_____________________
Group variables            None

── Variable type: factor ───────────────────────────────────────────────────────
  skim_variable n_missing complete_rate ordered n_unique top_counts
1 supp                  0             1 FALSE          2 OJ: 30, VC: 30

── Variable type: numeric ──────────────────────────────────────────────────────
  skim_variable n_missing complete_rate  mean    sd    p0   p25   p50   p75  p100
1 len                   0             1 18.8  7.65    4.2  13.1  19.2  25.3  33.9
2 dose                  0             1  1.17 0.629   0.5   0.5   1     2     2
```

How many rows does the ToothGrowth dataset contain?

- (•) 60
- ( ) 50
- ( ) 25
- ( ) 40

Correct
The code chunk `skim_without_charts(ToothGrowth)` gives you a comprehensive view of the dataset. Inside the parentheses of the skim_without_charts() function is the name of the dataset you want to view. The code returns a summary with the name of the dataset and the number of rows and columns. It also shows the column types and data types contained in the dataset. The ToothGrowth dataset contains 60 rows.

4. A data analyst is working with a data frame named *cars*. The analyst notices that all the column names in the data frame are capitalized. What code chunk lets the analyst change all the column names to lowercase? 1 / 1 point

- (•) `rename_with(cars, tolower)`
- ( ) `rename_with(toupper, cars)`
- ( ) `rename_with(tolower, cars)`
- ( ) `rename_with(cars, toupper)`

Correct
The code chunk is `rename_with(cars, tolower)`. The rename_with() function will enable the analyst to easily change the case of the column names to lowercase. Including the tolower argument indicates that all column names will be changed to lowercase.

5. A data analyst is working with the penguins data. The variable *species* includes three penguin species: Adelie, Chinstrap, and Gentoo. The analyst wants to create a data frame that only includes the Adelie species. The analyst receives an error message when they run the following code: **1 / 1 point**

```
penguins %>%

  filter(species <- "Adelie")
```

How can the analyst change the second line of code to correct the error?

- (●) `filter(species == "Adelie")`
- ( ) `filter(Adelie == species)`
- ( ) `filter("Adelie")`
- ( ) `filter("Adelie" <- species)`

Correct
The code chunk is `filter(species == "Adelie")`. The filter function is used to specify the part of the data to be viewed. Two equal signs in an argument mean "exactly equal to." Using this operator instead of the assignment operator <- calls only the data about Adelie penguins to the dataset.

6. You are working with the penguins dataset. You want to use the summarize() and max() functions to find the maximum value for the variable *flipper_length_mm*. You write the following code: **1 / 1 point**

```
penguins %>%

  drop_na() %>%

  group_by(species) %>%
```

Add the code chunk that lets you find the maximum value for the variable *flipper_length_mm*.

```
1  summarize(max(flipper_length_mm))
2
```

Run
Reset

```
# A tibble: 3 <U+00D7> 2
  species `max(flipper_length_mm)`
    <chr>                     <int>
1  Adelie                       210
2 Chinstrap                     212
3  Gentoo                       231
```

What is the maximum flipper length in mm for the Gentoo species?

- ( ) 212
- ( ) 200
- ( ) 210
- (●) 231

Correct
The code chunk `summarize(max(flipper_length_mm))` lets you find the maximum value for the variable flipper_length_mm. The correct code is `penguins %>% drop_na() %>% group_by(species) %>% summarize(max(flipper_length_mm))`. The summarize() function displays summary statistics. You can use the summarize() function in combination with other functions -- such as mean(), max(), and min() -- to calculate specific statistics. In this case, you use max() to calculate the maximum value for flipper length. The maximum flipper length for the Gentoo species is 231mm.

7. A data analyst is working with a data frame called *salary_data*. They want to create a new column named *hourly_salary* that includes data from the *wages* column divided by 40. What code chunk lets the analyst create the *hourly_salary* column? **1 / 1 point**

- ( ) `mutate(hourly_salary, salary_data = wages / 40)`
- ( ) `mutate(hourly_salary = wages / 40)`
- (●) `mutate(salary_data, hourly_salary = wages / 40)`
- ( ) `mutate(salary_data, hourly_salary = wages * 40)`

Correct
The code chunk is `mutate(salary_data, hourly_salary = wages / 40)`. The analyst can use the mutate() function to create a new column for wages divided by 40 called hourly_salary. The mutate() function can create a new column without affecting any existing columns.

8. A data analyst is working with a data frame named *stores*. It has separate columns for city (*city*) and state (*state*). The analyst wants to combine the two columns into a single column named *location*, with the city and state separated by a comma. What code chunk lets the analyst create the *location* column? **1 / 1 point**

- (●) `unite(stores, "location", city, state, sep=",")`
- ( ) `unite(stores, "location", city, sep=",")`
- ( ) `unite(stores, city, state, sep=",")`
- ( ) `unite(stores, "location", city, state)`

Correct
The code chunk `unite(stores, "location", city, state, sep=",")` lets the analyst create the *location* column. The unite() function lets the analyst combine the city and state data into a single column. In the parentheses of the function, the analyst writes the name of the data frame, then the name of the new column in quotation marks, followed by the names of the two columns they want to combine. Finally, the argument `sep="," ` places a comma between the city and state data in the *location* column.

9. **1 / 1 point**

In R, which statistical measure demonstrates how strong the relationship is between two variables?

- (●) Correlation
- ( ) Standard deviation
- ( ) Average
- ( ) Maximum

Correct
Correlation measures how strong the relationship between two variables is. This is represented by the cor() function.

10. **1 / 1 point**

A data analyst uses the bias() function to compare the actual outcome with the predicted outcome to determine if the model is biased. They get a score of 0.8. What does this mean?

- (●) The model is biased
- ( ) Bias can be determined
- ( ) Bias cannot be determined
- ( ) The model is not biased

Correct
A score of 0.8 indicates that the model is biased. The closer the score is to zero, the less likely it is that the model is biased.