Congratulations! You passed!
TO PASS 80% or higher

Keep Learning

GRADE
90%

Weekly challenge 3

LATEST SUBMISSION GRADE
90%

1. A data analyst creates a data frame with data that has more than 50,000 observations in it. When they print their data frame, it slows down their console. To avoid this, they decide to switch to a tibble. Why would a tibble be more useful in this situation? **1 / 1 point**

- ○ Tibbles will automatically create row names to make the data easier to read
- ◉ Tibbles won't overload the console because they automatically only print the first 10 rows of data and as many variables as will fit on the screen
- ○ Tibbles will automatically change the names of variables to make them shorter and easier to read
- ○ Tibbles only include a limited number of data items

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.

2. **1 / 1 point**

A data analyst is working with a large data frame. It contains so many columns that they don't all fit on the screen at once. The analyst wants a quick list of all of the column names to get a better idea of what is in their data. What function should they use?

- ○ head()
- ◉ colnames()
- ○ str()
- ○ mutate()

Correct
The colnames() function will return a list of all the column names in a data frame for easy reference.

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
```

Run
Reset

How many rows does the ToothGrowth dataset contain?

- ○ 40
- ○ 25
- ○ 50
- ◉ 60

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 *sales*. They write the following code: **1 / 1 point**

```
sales %>%
```

The data frame contains a column named q1_sales. What code chunk does the analyst add to change the name of the column from q1_sales to quarter1_sales ?

- ○ `rename(q1_sales = quarter1_sales)`
- ○ `rename(quarter1_sales <- "q1_sales")`

- `rename(q1_sales <- "quarter1_sales")`
- `rename(quarter1_sales = q1_sales)`

Correct

The code chunk is `rename(quarter1_sales = q1_sales)`. The rename() function allows the analyst to easily change a column name. The code chunk `rename(quarter1_sales = q1_sales)` will change the column name from q1_sales to quarter1_sales.

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

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 min() functions to find the minimum value for the variable *bill_depth_mm*. You write the following code:

1 / 1 point

```
penguins %>%
  drop_na() %>%
  group_by(species) %>%
```

Add the code chunk that lets you find the minimum value for the variable *bill_depth_mm*.

```
1
```

Run

Reset

What is the minimum bill depth in mm for the Chinstrap species?

- 15.5
- 13.1
- 16.4
- 12.4

Correct

The code chunk `summarize(min(bill_depth_mm))` lets you find the minimum value for the variable bill_depth_mm. The correct code is `penguins %>% drop_na() %>% group_by(species) %>% summarize(min(bill_depth_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 min() to calculate the minimum value for bill depth. The minimum bill depth for the Chinstrap species is 16.4mm.

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?

0 / 1 point

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

Incorrect

Review the video on transforming data in R for a refresher.

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

✓ **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?

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

✓ **Correct**

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

10. A data analyst wants to find out how much the predicted outcome and the actual outcome of their data model differ. What function can they use to quickly measure this? **1 / 1 point**

- mean()
- cor()
- bias() (selected)
- sd()

✓ **Correct**

The bias() function can be used to calculate the average amount a predicted outcome and actual outcome differ in order to determine if the data model is biased.