

## ✓ Congratulations! You passed!

Grade  
received 100%

Latest Submission  
Grade 100%

To pass 80% or  
higher

Go to next item

1. A data analyst is considering using tibbles instead of basic data frames. What are some of the limitations of tibbles? Select all that apply.

1 / 1 point

- ☐ Tibbles can overload a console
- ☒ Tibbles can never create row names

✓ Correct

Tibbles are useful when working with large datasets because they make printing easier. But tibbles can never change the input type of the data, create row names, or change the names of variables.

- ☒ Tibbles can never change the input type of the data

✓ Correct

Tibbles are useful when working with large datasets because they make printing easier. But tibbles can never change the input type of the data, create row names, or change the names of variables.

- ☒ Tibbles won't automatically change the names of variables

✓ Correct

Tibbles are useful when working with large datasets because they make printing easier. But tibbles can never change the input type of the data, create row names, or change the names of variables.

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

- ☐ colnames()
- ☐ print()
- ☐ preview()
- ☒ head()

✓ 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	03: 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?

- ☐ 25
- ☒ 60
- ☐ 50
- ☐ 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 the penguins dataset. What code chunk does the analyst write to make sure all the column names are unique and consistent and contain only letters, numbers, and underscores?

1 / 1 point

- ☐ `drop_na(penguins)`
- ☐ `rename(penguins)`
- ☒ `clean_names(penguins)`
- ☐ `select(penguins)`

✓ Correct

The code chunk is `clean_names(penguins)`. The `clean_names()` function ensures that there are only characters, numbers, and underscores in the names used in the data frame.

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(species == "Adelie")`
- ☐ `filter(Adelie == species)`
- ☐ `filter("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 `mean()` functions to find the mean value for the variable *body\_mass\_g*. You write the following code:

1 / 1 point

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

Add the code chunk that lets you find the mean value for the variable *body\_mass\_g*.

```
1 summarize(mean(body_mass_g))
```

Run

Reset

```
# A tibble: 3 <U+0007> 2
  species `mean(body_mass_g)`
  <chr>      <dbl>
1 Adelie      3706.164
2 Chinstrap  3733.088
3 Gentoo     5092.437
```

What is the mean body mass in g for the Adelie species?

- ☐ 5092.437
- ☒ 3706.164
- ☐ 3733.088
- ☐ 4207.433

✓ Correct

The code chunk `summarize(mean(body_mass_g))` lets you find the mean value for the variable *body\_mass\_g*. The correct code is `penguins %>% drop_na() %>% group_by(species) %>% summarize(mean(body_mass_g))`. 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 `mean()` to calculate the mean value for body mass. The mean body mass for the Adelie species is 3706.164g.

7. A data analyst is working with a data frame called *salary\_data*. They want to create a new column named *total\_wages* that adds together data in the *standard\_wages* and *overtime\_wages* columns. What code chunk lets the analyst create the *total\_wages* column?

1 / 1 point

- ☐ `mutate(salary_data, total_wages = standard_wages * overtime_wages)`
- ☒ `mutate(salary_data, total_wages = standard_wages + overtime_wages)`
- ☐ `mutate(salary_data, standard_wages = total_wages + overtime_wages)`
- ☐ `mutate(total_wages = standard_wages + overtime_wages)`

✓ Correct

The code chunk is `mutate(salary_data, total_wages = standard_wages + overtime_wages)`. The analyst can use the `mutate()` function to create a new column for *standard\_wages* plus *overtime\_wages* called *total\_wages*. The `mutate()` function can create a new column without affecting any existing columns.

8. A data analyst is working with a data frame named *customers*. It has separate columns for area code (*area\_code*) and phone number (*phone\_num*). The analyst wants to combine the two columns into a single column called *phone\_number*, with the area code and phone number separated by a hyphen. What code chunk lets the analyst create the *phone\_number* column?

1 / 1 point

- ☒ `unite(customers, "phone_number", area_code, phone_num, sep="-")`
- ☐ `unite(customers, "phone_number", area_code, phone_num)`
- ☐ `unite(customers, "phone_number", area_code, sep="-")`
- ☐ `unite(customers, area_code, phone_num, sep="-")`

✓ Correct

The code chunk `unite(customers, "phone_number", area_code, phone_num, sep="-")` lets the analyst create the *phone\_number* column. The `unite()` function lets the analyst combine the area code and phone number 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 hyphen between the area code and phone number data in the *phone\_number* column.

9.

1 / 1 point

A data analyst is using statistical measures to get a better understanding of their data. What function can they use to determine how strongly related are two of the variables?

- ☐ `mean()`
- ☐ `sd()`
- ☒ `cor()`
- ☐ `bias()`

✓ Correct

The `cor()` returns the correlation between two variables. Correlation shows us how strong the relationship is between two variables .

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?

- ☐ Bias can be determined
- ☒ The model is biased
- ☐ 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.