✓ **Herzlichen Glückwunsch! Sie haben bestanden!**
ZUM BESTEHEN 80 % oder höher

Lernen Sie weiter

BEWERTUNG
100 %

Weekly challenge 3

NEUESTE EINREICHUNGSBEWERTUNG
100%

1. A data analyst creates a data frame with customer data. The data frame includes a column with customer addresses, but not every customer has an address in the system. The data analyst notices that there are NA values in some of the cells. Why is this? **1 / 1 Punkt**

- ○ All data frames contain NULL values
- ○ Rows should contain the same number of items
- ○ This is an error
- ◉ Columns should contain the same number of items

✓ **Richtig**
The columns of a data frame should contain the same number of items. If there are missing or inapplicable variables in a column, it will insert NA to maintain the shape of the data frame and prevent errors.

2. **1 / 1 Punkt**

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.

- ☐ Tibbles can create row names
- ☑ Tibbles automatically only preview the first 10 rows of data

✓ **Richtig**
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 automatically only preview as many columns as fit on screen

✓ **Richtig**
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

3. A data analyst wants a quick summary of the structure of their data frame, including the column names and the number of rows and variables. What function should they use? **1 / 1 Punkt**

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

✓ **Richtig**
The str() function returns a summary of the structure of a dataframe, including column names, the number of rows and variables, and the type of data stored.

4. A data analyst is working with the ToothGrowth dataset in R. What code chunk let them get a preview of the dataset? **1 / 1 Punkt**

- ◉ `head(ToothGrowth)`
- ○ `max(ToothGrowth)`
- ○ `unite(ToothGrowth)`
- ○ `colnames(ToothGrowth)`

✓ **Richtig**
The code chunk is `head(ToothGrowth)`. The head() function provides the analyst with a quick view of the column names and first few rows of the ToothGrowth dataset.

5. 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 Punkt**

- ○ `rename_with(cars, toupper)`
- ○ `rename_with(toupper, cars)`
- ◉ `rename_with(cars, tolower)`
- ○ `rename_with(tolower, cars)`

✓ **Richtig**
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.

6. A data analyst is working with the penguins data. They write the following code: **1 / 1 Punkt**

```
penguins %>%
```

The variable *species* includes three penguin species: Adelie, Chinstrap, and Gentoo. What code chunk does the analyst add to create a data frame that only includes the Gentoo species?

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

✓ **Richtig**
The code chunk is `filter(species == "Gentoo")`. The filter function allows the data analyst to specify which part of the data they want to view. Two equal signs in an argument mean "exactly equal to." Using this operator instead of the assignment operator <- calls only the data about Gentoo penguins to the dataset.

7. A data analyst is working with the penguins dataset. They write the following code: **1 / 1 Punkt**

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

What code chunk does the analyst add to find the mean value for the variable *body_mass_g*?

- ( ) `summarize(body_mass_g(mean))`
- ( ) `summarize(=body_mass_g)`
- (●) `summarize(mean(body_mass_g))`
- ( ) `summarize(max(body_mass_g))`

✓ **Richtig**
The code chunk is `summarize(mean(body_mass_g))`. The summarize function gives high-level information about a dataset.

8. A data analyst is working with a data frame named *salary_data*. They want to create a new column named *wages* that includes data from the *rate* column multiplied by 40. What code chunk lets the analyst create the *wages* column? **1 / 1 Punkt**

- ( ) `mutate(wages = rate * 40)`
- ( ) `mutate(salary_data, wages = rate + 40)`
- ( ) `mutate(salary_data, rate = wages * 40)`
- (●) `mutate(salary_data, wages = rate * 40)`

✓ **Richtig**
The code chunk is `mutate(salary_data, wages = rate * 40)`. The analyst can use the mutate() function to create a new column called wages that includes data from the rate column multiplied by 40. The mutate() function can create a new column without affecting any existing columns.

9. 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 Punkt**

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

✓ **Richtig**
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.

10. A data analyst wants to summarize their data with the sd(), cor(), and mean(). What kind of measures are these? **1 / 1 Punkt**

- (●) Statistical
- ( ) Numerical
- ( ) Summary
- ( ) Standard

✓ **Richtig**
Standard deviation, correlation, mean, maximum, and minimum are statistical measures which can be used to summarize data.

11. **1 / 1 Punkt**

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

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

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

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

12. A data analyst is studying weather data. They write the following code chunk: 1 / 1 Punkt

```
bias(actual_temp, predicted_temp)
```

What will this code chunk calculate?

- ○ The total average of the values
- ○ The minimum difference between the actual and predicted values
- ◉ The average difference between the actual and predicted values
- ○ The maximum difference between the actual and predicted values

✓ **Richtig**

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.