

## ✓ Congratulations! You passed!

Grade  
received 92.30%

Latest Submission  
Grade 92.31%

To pass 80% or  
higher

Go to next item

### 1. Scenario 1, questions 1-7

1 / 1 point

As part of the data science team at Gourmet Analytics, you use data analytics to advise companies in the food industry. You clean, organize, and visualize data to arrive at insights that will benefit your clients. As a member of a collaborative team, sharing your analysis with others is an important part of your job.

Your current client is Chocolate and Tea, an up-and-coming chain of cafes.



The eatery combines an extensive menu of fine teas with chocolate bars from around the world. Their diverse selection includes everything from plantain milk chocolate, to tangerine white chocolate, to dark chocolate with pistachio and fig. The encyclopedic list of chocolate bars is the basis of Chocolate and Tea's brand appeal. Chocolate bar sales are the main driver of revenue.

Chocolate and Tea aims to serve chocolate bars that are highly rated by professional critics. They also continually adjust the menu to make sure it reflects the global diversity of chocolate production. The management team regularly updates the chocolate bar list in order to align with the latest ratings and to ensure that the list contains bars from a variety of countries.

They've asked you to collect and analyze data on the latest chocolate ratings. In particular, they'd like to know which countries produce the highest-rated bars of super dark chocolate (a high percentage of cocoa). This data will help them create their next chocolate bar menu.

Your team has received a dataset that features the latest ratings for thousands of chocolates from around the world. Click [here](#) to access the dataset. Given the data and the nature of the work you will do for your client, your team agrees to use R for this project.

**A teammate asks you about the benefits of using R for the project. You mention that R can quickly process lots of data and create high quality data visualizations. What is another benefit of using R for the project?**

- ☐ Choose a topic for analysis
- ☐ Define a problem and ask the right questions
- ☒ Easily reproduce and share an analysis
- ☐ Automatically clean data

✓ Correct

Another benefit of using R for the project is that it can easily reproduce and share an analysis.

### 2. Scenario 1, continued

1 / 1 point

Before you begin working with your data, you need to import it and save it as a data frame. To get started, you open your RStudio workspace and load the tidyverse library. You upload a .csv file containing the data to RStudio and store it in a project folder named flavors\_of\_cacao.csv.

**You use the read\_csv() function to import the data from the .csv file. Assume that the name of the data frame is bars\_df and the .csv file is in the working directory. What code chunk lets you create the data frame?**

- ☐ `bars_df %>% read_csv("flavors_of_cacao.csv")`
- ☒ `bars_df <- read_csv("flavors_of_cacao.csv")`
- ☐ `bars_df + read_csv("flavors_of_cacao.csv")`
- ☐ `read_csv("flavors_of_cacao.csv") + bars_df`

✓ Correct

The code chunk `bars_df <- read_csv("flavors_of_cacao.csv")` lets you create the data frame. In this code chunk:

- `bars_df` is the name of the data frame that will store the data.
- `<-` is the assignment operator to assign values to the data frame.
- `read_csv()` is the function that will import the data to the data frame.
- `"flavors_of_cacao.csv"` is the file name that `read_csv()` function takes for its argument.

### 3. Scenario 1, continued

1 / 1 point

Now that you've created a data frame, you want to find out more about how the data is organized. The data frame has hundreds of rows and lots of columns.

Assume the name of your data frame is `flavors_df`. What code chunk lets you review the structure of the data frame?

- ☐ `filter(flavors_df)`
- ☒ `str(flavors_df)`
- ☐ `summarize(flavors_df)`
- ☐ `select(flavors_df)`

✓ **Correct**

You write the code chunk `str(flavors_df)`. In this code chunk:

- `str()` is the function that will return the structure of the data frame, and give you high-level information like the column names and the type of data contained in those columns.
- `flavors_df` is the name of the data frame that the `str()` function takes for its argument.

#### 4. Scenario 1, continued

0 / 1 point

Next, you begin to clean your data. When you check out the column headings in your data frame you notice that the first column is named `Company...Maker.if.known`. (Note: The period after `known` is part of the variable name.) For the sake of clarity and consistency, you decide to rename this column `Company` (without a period at the end).

Assume the first part of your code chunk is:

```
flavors_df %>%
```

What code chunk do you add to change the column name?

- ☐ `rename(Company = Company...Maker.if.known.)`
- ☐ `rename(Company...Maker.if.known. <- Company)`
- ☒ `rename(Company <- Company...Maker.if.known.)`
- ☐ `rename(Company...Maker.if.known. = Company)`

✗ **Incorrect**

Review [the video on cleaning data in R](#) for a refresher.

#### 5. After previewing and cleaning your data, you determine what variables are most relevant to your analysis. Your main focus is on `Rating`, `Cocoa.Percent`, and `Company`. You decide to use the `select()` function to create a new data frame with only these three variables.

1 / 1 point

Assume the first part of your code is:

```
trimmed_flavors_df <- flavors_df %>%
```

Add the code chunk that lets you select the three variables.

```
1 select(Rating,Cocoa.Percent,Company)
```

Run  
Reset

```
# A tibble: 1,795 <U+00D7> 3
  Rating Cocoa.Percent Company
<dbl>    <chr>      <chr>
1  3.75      63% A. Morin
2  2.75      70% A. Morin
3  3.00      70% A. Morin
4  3.50      70% A. Morin
5  3.50      70% A. Morin
6  2.75      70% A. Morin
7  3.50      70% A. Morin
8  3.50      70% A. Morin
9  3.75      70% A. Morin
10 4.00      70% A. Morin
# ... with 1,785 more rows
```

What company appears in row 1 of your tibble?

- ☒ A. Morin
- ☐ Videri
- ☐ Rogue
- ☐ Soma

✓ **Correct**

You add the code chunk `select(Rating, Cocoa.Percent, Company)` to select the three variables.

The correct code is `trimmed_flavors_df <- flavors_df %>% select(Rating, Cocoa.Percent, Company)`. In this code chunk:

- The `select()` function lets you select specific variables for your new data frame.
- `select()` takes the names of the variables you want to choose as its argument: `Rating`, `Cocoa.Percent`, `Company`.

The company A. Morin appears in row 1 of your tibble.

#### 6. Next, you select the basic statistics that can help your team better understand the ratings system in your data.

1 / 1 point

Assume the first part of your code is:

```
trimmed_flavors_df %>%
```

You want to use the `summarize()` and `max()` functions to find the maximum rating for your data. Add the

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

```
1 summarize(max(Rating))
```

Run  
Reset

```
# A tibble: 1 <U+00D7> 1  
  `max(Rating)`  
    <dbl>  
1             5
```

What is the maximum rating?

- ☐ 4.5
- ☒ 5
- ☐ 6
- ☐ 5.5

✓ Correct

You add the code chunk `summarize(max(Rating))` to find the maximum value for the variable *Rating*. The correct code is `trimmed_flavors_df %>% summarize(max(Rating))`. In this code chunk:

- The `summarize()` function lets you display 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 the variable *Rating*.

The maximum rating is 5.

7. After completing your analysis of the rating system, you determine that any rating greater than or equal to 3.75 points can be considered a high rating. You also know that Chocolate and Tea considers a bar to be super dark chocolate if the bar's cocoa percent is greater than or equal to 80%. You decide to create a new data frame to find out which chocolate bars meet these two conditions.

1 / 1 point

Assume the first part of your code is:

```
best_trimmed_flavors_df <- trimmed_flavors_df %>%
```

You want to apply the `filter()` function to the variables *Cocoa.Percent* and *Rating*. Add the code chunk that lets you filter the new data frame for chocolate bars that contain at least 80% cocoa and have a rating of at least 3.75 points.

```
1 filter(Cocoa.Percent >= 80, Rating >= 3.75)
```

Run  
Reset

```
# A tibble: 8 <U+00D7> 3  
  Rating Cocoa.Percent Company.Location  
    <dbl>      <chr>      <chr>  
1  3.75      80%      Amsterdam  
2  3.75      80%      Scotland  
3  3.75      80%      U.S.A.  
4  3.75      82%      U.S.A.  
5  4.00      80%      France  
6  3.75      80%      U.S.A.  
7  4.00      88%      Canada  
8  3.75      90%      U.S.A.
```

How many rows does your tibble include?

- ☒ 8
- ☐ 12
- ☐ 20
- ☐ 22

✓ Correct

The code chunk `filter(Cocoa.Percent >= 80, Rating >= 3.75)` lets you filter the data frame for chocolate bars that contain at least 80% cocoa and have a rating of at least 3.75 points. The correct code is `best_trimmed_flavors_df <- trimmed_flavors_df %>% filter(Cocoa.Percent >= 80, Rating >= 3.75)`. In this code chunk:

- The `filter()` function lets you filter your data frame based on specific criteria.
- *Cocoa.Percent* and *Rating* refer to the variables you want to filter.
- The `>=` operator signifies "greater than or equal to."
- The new data frame will show all the values of *Cocoa.Percent* greater than or equal to 80, and all the values of *Rating* greater than or equal to 3.75.

Your tibble includes 8 rows.

8. Now that you've cleaned and organized your data, you're ready to create some useful data visualizations. Your team assigns you the task of creating a series of visualizations based on requests from the Chocolate and Tea management team. You decide to use `ggplot2` to create your visuals.

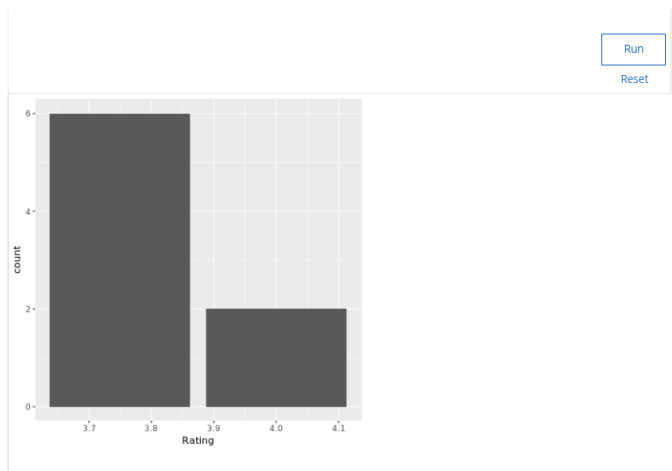
1 / 1 point

Assume your first line of code is:

```
ggplot(data = best_trimmed_flavors_df) +
```

You want to use the `geom_bar()` function to create a bar chart. Add the code chunk that lets you create a bar chart with the variable *Rating* on the x-axis.

```
1 geom_bar(mapping = aes(x = Rating))
```



How many bars does your bar chart display?

- ☐ 5  
☒ 2  
☐ 3  
☐ 6

✓ Correct

You add the code chunk `geom_bar(mapping = aes(x = Rating))` to create a bar chart with the variable Rating on the x-axis. The correct code is `ggplot(data = best_trimmed_flavors_df) + geom_bar(mapping = aes(x = Rating))`. In this code chunk:

- `geom_bar()` is the geom function that uses bars to create a bar chart.
- Inside the parentheses of the `aes()` function, the code `x = Rating` maps the x aesthetic to the variable Rating.
- Rating will appear on the x-axis of the plot.
- By default, R will put a count of the variable Rating on the y-axis.

Your bar chart displays 2 bars.

9. Your bar chart reveals the locations that produce the highest rated chocolate bars. To get a better idea of the specific rating for each location, you'd like to highlight each bar.

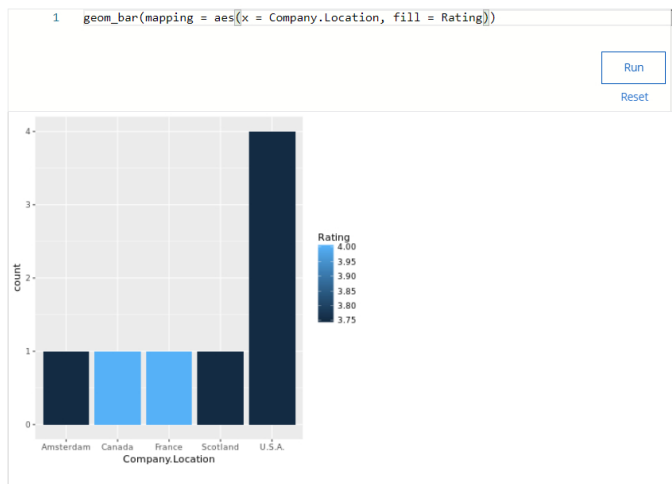
1 / 1 point

Assume that you are working with the code chunk:

```
ggplot(data = best_trimmed_flavors_df) +  
  geom_bar(mapping = aes(x = Company.Location))
```

Add a code chunk to the second line of code to map the aesthetic *color* to the variable *Rating*.

NOTE: the three dots (...) indicate where to add the code chunk.



According to your bar chart, which two company locations produce the highest rated chocolate bars?

- ☒ Canada and France  
☐ Scotland and France  
☐ Canada and U.S.A.  
☐ Amsterdam and U.S.A.

✓ Correct

You add the code chunk `color = Rating` to the second line of code to map the aesthetic color to the variable Rating. The correct code is `ggplot(data = best_trimmed_flavors_df) + geom_bar(mapping = aes(x = Company.Location, color = Rating))`. In this code chunk:

- Inside the parentheses of the `aes()` function, after the comma that follows `Company.Location`,

- Inside the parentheses of the `aes()` function, after the comma that follows `x = Company.Location`, write the aesthetic (color), then an equals sign, then the variable (`Rating`).
- The specific rating of each location will appear as a specific color that outlines each bar of your bar chart.

On your visualization, the legend titled "Rating" shows the color coding for the variable Rating. Lighter blues correspond to higher ratings and darker blues correspond to lower ratings.

According to your bar chart, the two company locations that produce the highest rated chocolate bars are Canada and France.

#### 10. Scenario 2, continued

1 / 1 point

A teammate creates a new plot based on the chocolate bar data. The teammate asks you to make some revisions to their code.

Assume your teammate shares the following code chunk:

```
ggplot(data = best_trimmed_flavors_df) +  
  geom_bar(mapping = aes(x = Rating)) +
```

What code chunk do you add to the third line to create wrap around facets of the variable *Rating*?

- ☐ `facet(~Rating)`
- ☐ `facet_wrap(Rating)`
- ☒ `facet_wrap(~Rating)`
- ☐ `facet_wrap(Rating~)`

✓ Correct

You write the code chunk `facet_wrap(~Rating)`. In this code chunk:

- `facet_wrap()` is the function that lets you create wrap around facets of a variable.
- Inside the parentheses of the `facet_wrap()` function, type a tilde symbol (~) followed by the name of the variable (`Rating`).

#### 11. Scenario 2, continued

1 / 1 point

Your team has created some basic visualizations to explore different aspects of the chocolate bar data. You've volunteered to add titles to the plots. You begin with a scatterplot.

Assume the first part of your code chunk is:

```
ggplot(data = trimmed_flavors_df) +  
  geom_point(mapping = aes(x = Cocoa.Percent, y = Rating)) +
```

What code chunk do you add to the third line to add the title *Best Chocolates* to your plot?

- ☐ `labs(title <- "Best Chocolates")`
- ☒ `labs(title = "Best Chocolates")`
- ☐ `labs("Best Chocolates" = title)`
- ☐ `labs("Best Chocolates")`

✓ Correct

You write the code chunk `labs(title = "Best Chocolates")`. In this code chunk:

- `labs()` is the function that lets you add a title to your plot.
- In the parentheses of the `labs()` function, write the word title, then an equals sign, then the specific text of the title in quotation marks (`"Best Chocolates"`).

#### 12. Scenario 2, continued

1 / 1 point

Next, you create a new scatterplot to explore the relationship between different variables. You want to save your plot so you can access it later on. You know that the `ggsave()` function defaults to saving the last plot that you displayed in RStudio, so you're ready to write the code to save your scatterplot.

Assume your first two lines of code are:

```
ggplot(data = trimmed_flavors_df) +  
  geom_point(mapping = aes(x = Cocoa.Percent, y = Rating)) +
```

What code chunk do you add to the third line to save your plot as a pdf file with "chocolate" as the file name?

- ☒ `ggsave("chocolate.pdf")`
- ☐ `ggsave("pdf.chocolate")`
- ☐ `ggsave(chocolate.pdf)`
- ☐ `ggsave("chocolate.png")`

✓ Correct

You add the code chunk `ggsave("chocolate.pdf")` to save your plot as a pdf file with "chocolate" as the file name. In this code chunk:

- Inside the parentheses of the `ggsave()` function, type a quotation mark followed by the file name (chocolate), then a period, then the type of file format (pdf), then a closing quotation mark.

13. Scenario 2, continued

1 / 1 point

As a final step in the analysis process, you create a report to document and share your work. Before you share your work with the management team at Chocolate and Tea, you are going to meet with your team and get feedback. Your team wants the documentation to include all your code and display all your visualizations.

**You decide to create an R Markdown notebook to document your work. What are your reasons for choosing an R Markdown notebook? Select all that apply.**

☒ It displays your data visualizations



**Correct**

You choose an R Markdown notebook to document your work because it lets you record and share every step of your analysis. The notebook allows users to run your code and also displays your data visualizations.

☒ It allows users to run your code



**Correct**

You choose an R Markdown notebook to document your work because it lets you record and share every step of your analysis. The notebook allows users to run your code and also displays your data visualizations.

☒ It lets you record and share every step of your analysis



**Correct**

You choose an R Markdown notebook to document your work because it lets you record and share every step of your analysis. The notebook allows users to run your code and also displays your data visualizations.

☐ It automatically creates a website to show your work