Section 1. Section 1.

Your team has received a dataset that features the latest ratings for thousands of chocolates from around the world. Click https://licensess-be-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?

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

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?

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

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

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.

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.il.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 Maker (without a period at the end).

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

2. Scenario 1, continued

3. Scenario 1, continued

4. Scenario 1, continued

flavors_df %>%

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.Location)

What company location appears in row 1 of your tibble?

O correct
You add the code chunk select (Rating, Cocoa.Percent, Company.Location) to select the
three variables. The correct code is trimmed_flavors_df <- flavors_df %>% select (Rating,
Cocoa.Percent, Company.Location). In this code chunk:

 select() takes the names of the variables you want to choose as its argument: Rating, Cocoa.Percent, Company.Location.

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

Ocrrect
You add the code chunk summarize (mean (Rating)) to find the mean value for the variable Rating.
The correct code is trimmed_flavore_df >> a nummarize (mean (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(), sd(), and max() to calculate specific statistics.

In this case, you use mean() to calculate the mean value for the variable Rating.

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 point percentage 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.

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.

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

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.

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 Company. Location on the x-axis.

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

Inside the parentheses of the aes() function, the code x = Company.Location maps the x aesthetic to the variable Company.Location.

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.

geom_bar() is the geom function that uses bars to create a bar chart.

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

 $\label{lem:conding} \textbf{According to your bar chart, which two company locations produce the highest rated chocolate bars?}$

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

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

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

What code chunk do you add to the third line to create wrap around facets of the variable Cocoa.Percent?

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.

 $\label{eq:geom_point} $$ (mapping = aes(x = Cocoa. Percent, y = Rating)) +$$ $$ What code chunk do you add to the third line to add the title $$ Suggested Chocolate$ to your plot?$

O correct
You write the code chunk labs (title = "Suggested Chocolate"). In this code chunk:

 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 ("Suggested Chocolate").

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.

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

 Inside the parentheses of the ggaave () function, type a quotation mark followed by the file name (chocolate), then a period, then the type of file format (png), then a closing quotation mark.

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.

Fill in the blank: You want to record and share every step of your analysis, let teammates run your code, and display your visualizations. You decide to create _____ to document your work.

Correct
You use an R Markdown notebook to document your work. The notebook lets you record and share every step of your analysis, lets your teammates run your code, and displays your visualizations.

labs () is the function that lets you add a title to your plot.

geom_point(mapping = aes(x = Cocoa.Percent, y = Rating))

✓ Correct
 You write the code chunk ggsave ("chocolate.pag"). In this code chunk:

O correct

You write the code chunk facet_wrap (~Cocoa. Percent). 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 (Cocoa. Percent).

1 geom_bar(mapping = aes(x = Company.Location, fill = Rating))

Company.Location will appear on the x-axis of the plot.
 By default, R will put a count of the variable Company.Location on the y-axis.

1 geom_bar(mapping = aes(x = Company.Location))

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."

You want to use the summarize() and mean() functions to find the mean rating for your data. Add the code chunk that lets you find the mean value for the variable Rating.

The select() function lets you select specific variables for your new data frame.

The company location France appears in row 1 of your tibble.

Assume the first part of your code is: trimmed_flavors_df %>%

What is the mean rating?

(a) 3.185933
(b) 3.995445
(c) 4.701337
(c) 4.230765

The mean rating is 3.185933.

Assume the first part of your code is:

81220

Your tibble includes 8 rows.

Assume your first line of code is:

ggplot(data = best_trimmed_flavors_df) +

How many bars does your bar chart display?

Your bar chart displays 5 bars.

Assume that you are working with the following code:

ggplot(data = best_trimmed_flavors_df) +

geom_bar(mapping = aes(x = Company.Location))

Scotland and U.S.A.
Amsterdam and France
Canada and France
Scotland and Canada

10. Scenario 2, continued

Assume your teammate shares the following code chunk:

ggplot(data = best_trimmed_flavors_df) +

geom_bar(mapping = aes(x = Cocoa.Percent)) +

facet_wrap(Cocoa.Percent~)

facet_wrap(%>%Cocoa.Percent)

facet_wrap(~Cocoa.Percent)

facet_wrap(~Cocoa.Percent)

11. Scenario 2, continued

12. Scenario 2, continued

Assume your first two lines of code are:

ggplot(data = trimmed_flavors_df) +

ggsave("png.chocolate")
ggsave("chocolate.png")
ggsave(chocolate.png)
ggsave("chocolate")

13. Scenario 2, continued

a spreadsheeta databasean R Markdown notebook

Assume the first part of your code chunk is:

ggplot(data = trimmed_flavors_df) +

labs(title = "Suggested Chocolate")
 labs(Suggested Chocolate = title)
 labs(Suggested Chocolate)
 labs <- "Suggested Chocolate"

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

○ 3

best_trimmed_flavors_df <- trimmed_flavors_df %>%

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

1 summarize(mean_rating = mean(Rating))

Canada
France
Scotland
Colombia

Assume the first part of your code chunk is:

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

rename (Company . . .Maker .if .known %<% Maker)

rename (Maker = Company . . .Maker .if .known.)

rename (Maker %<% Company . . .Maker .if .known.)

rename (Company . . .Maker .if .known = Maker)

O correct
You write the code chunk rename (Maker = Company...Maker.if.known.).

rename () is the function that will change the name of your column.
 Inside the parentheses of the function, write the new name (Maker), the name you want to change (Company . . . Maker . if . known .).

5.After previewing and cleaning your data, you determine what variables are most relevant to your analysis. Your main focus is on Rating, 1/1 Cocoa.Percent, and Company.Location. You decide to use the select() function to create a new data frame with only these three variables.

Run Reset

Run Reset

Run

Run

summarize(flavors_df)

str(flavors_df)

filter(flavors_df)

select(flavors_df)

pead_csv("flavors_of_cacao.csv") + bars_df
 bars_df %% read_csv("flavors_of_cacao.csv")
 bars_df <- read_csv("flavors_of_cacao.csv")
 bars_df + read_csv("flavors_of_cacao.csv")</pre>

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.

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