Week 2 quiz

Started: Jan 19 at 11:28am

Quiz Instructions

This quiz will cover the asynchronous content from the **Week 2 module.** It is due Tuesday, Jan 19 at 6:00 p.m. ET. I've made this quiz a little shorter (8 questions) because of how much pre-recorded content I'm giving you this week. Still weighted as normal.

Notes

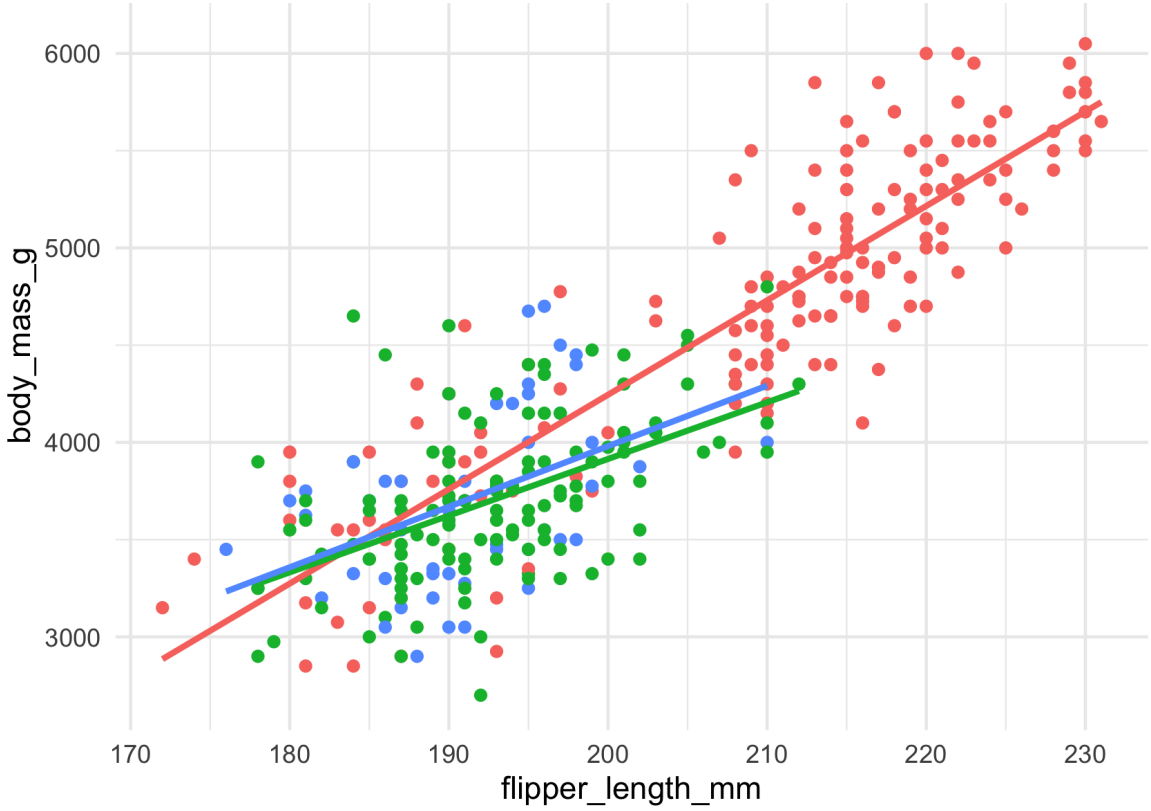
- Use Chrome or Firefox to access this quiz. It has **images** and they often fail to load in Safari. I use Chrome to write and check the quizzes.
  - Additionally, make sure your browser is up to date as out-of-date versions also cause issues.
  - **Possible 'hack' if images don't load:** Sometimes, right-clicking the image that is failing to load, and opening it in a new tab or downloading it, will let you see the image.
- This quiz is **open book**, so you are welcome to review the notes while completing it.
- From the time you start, you will have either **1 hour** to complete the quiz or 6:00 p.m. ET on Tuesday, whichever comes first. You *cannot* pause the quiz once you start it.
- The quiz will **auto-submit** any answers at the deadline.
- You have **one** attempt.
- You do not have to use RStudio while doing this quiz, but you are allowed to if you would like to.

Data for this quiz

Question 1 to 4 in this quiz refer to the **Palmer penguins data we looked at in week 1.**

Question 1 — 1 pts

Which ONE of the following chunks of code would produce the below chart?




```
library(palmerpenguins)
library(tidyverse)

penguins %>%
  ggplot(aes(x = flipper_length_mm, y = body_mass_g, colour = island)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()
```

```
library(palmerpenguins)
library(tidyverse)

penguins %>%
  ggplot(aes(x = flipper_length_mm, y = body_mass_g, colour = island)) %>%
  geom_point() %>%
  theme_minimal() %>%
  geom_smooth(method = "lm", se = FALSE)
```

```
library(palmerpenguins)
library(tidyverse)

penguins %>%
  ggplot(aes(x = flipper_length_mm, y = body_mass_g, colour = island)) +
  geom_point() +
```

```
  theme_minimal() +
  geom_line(method = "lm", se = FALSE)
```

○

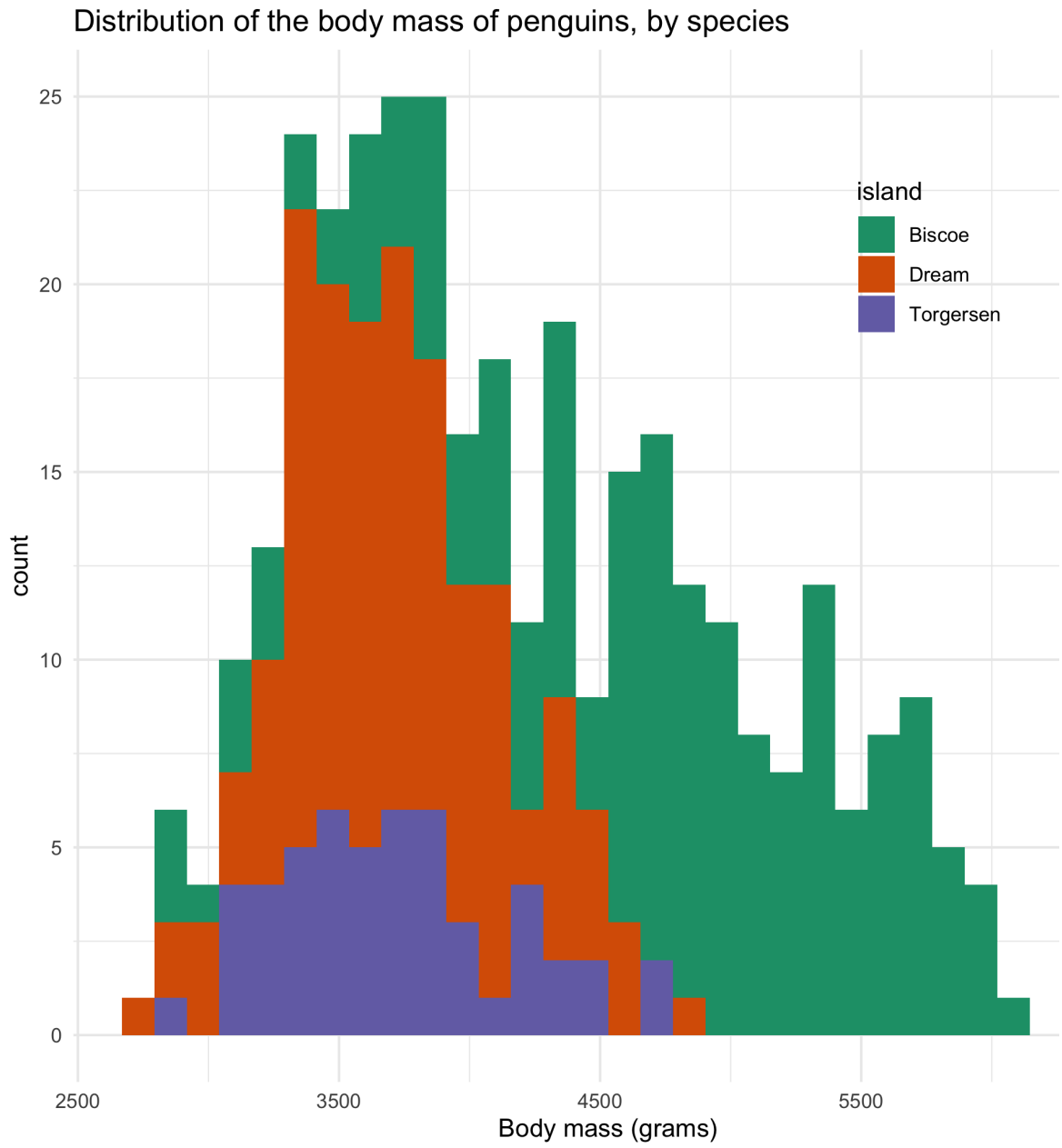
```
library(palmerpenguins)
library(tidyverse)

penguins %>%
  ggplot(aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(colour = island)) +
  geom_smooth(method = "lm", se = FALSE) +
  theme_minimal()
```

Question 2

1 pts

Which of the following commands could have been part of creating this plot? Tick all that apply.

(Assume all the appropriate libraries have been loaded.)



- [ ] `geom_stack()`
- [x] `theme_minimal()`
- [x] `geom_histogram(position = "stack")`
- [ ] `theme(legend.position = c(0.1, 0.8))`
- [ ] `facet_wrap(~island)`

- [x]
```
ggplot(aes(x = body_mass_g, fill = island))
```

Question 3 1 pts

Which of the following code chunks would produce the below tibble? Tick all that apply.

(Assume the correct libraries have been loaded.)

```
# A tibble: 6 x 4
# Groups:   island [3]
  island    sex    count  mean
  <fct>     <fct>  <int> <dbl>
1 Biscoe    female    80  43.3
2 Biscoe    male      83  47.1
3 Dream     female    61  42.3
4 Dream     male      62  46.1
5 Torgersen female    24  37.6
6 Torgersen male      23  40.6
```

- [x]
```
penguins %>%
  group_by(island, sex) %>%
  summarise(count = n(), mean = mean(bill_length_mm)) %>%
  filter(sex %in% c("female", "male"))
```

- [x]
```
penguins %>%
  filter(!is.na(sex)) %>%
  group_by(island, sex) %>%
  summarise(count = n(), mean = mean(bill_length_mm))
```

- [ ]
```
penguins %>%
  group_by(island, sex) %>%
  mutate(count = n(), mean = mean(bill_length_mm, na.rm = TRUE))
```

- [ ]
```
penguins %>%
  group_by(island, sex) %>%
  summarise(count = n(), mean = mean(bill_length_mm))
```

Question 4

1 pts

Your classmate claims that it is always true that the model matrix for a 'no intercept' model fitted with a **tidy dataset** must *also* be a tidy dataset.

They propose that you can see this in the following:

```
> model1 <- lm(body_mass_g ~ 0 + species + flipper_length_mm, data = penguins)
> head(model.matrix(model1), 6)
  speciesAdelie speciesChinstrap speciesGentoo flipper_length_mm
1             1                0             0               181
2             1                0             0               186
3             1                0             0               195
5             1                0             0               193
6             1                0             0               190
7             1                0             0               181
```

Which ONE of the following would be the BEST response to your classmate?

- ○ While this is true for this example, we can't know if it will generalise to other models and data sets.
- ○ This claim is correct.
- ○ This isn't correct because it will always break the rule that each value must have its own cell.
- ◉ This isn't correct because it will break the tidy data rule that each variable must have its own column whenever a categorical variable is included.

Question 5

1 pts

Match each package to its main purpose.

forcats	working with categorical v ⌄
dplyr	

data manipulation	
stringr	working with character str
ggplot2	data visualization
janitor	data cleaning
readxl	reading data saved in Exc

Question 6

1 pts

Motulsky (2014) discusses misconceptions about p-values. Which of the following is/are correct descriptions of a p-value? Tick all that apply.

Motulsky, H. J. (2014). *Common misconceptions about data analysis and statistics*. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *387*(11), 1017–1023. https://doi.org/10.1007/s00210-014-1037-6

- [ ] The probability of the parameter being like our test statistic or more extreme, if the null hypothesis is true.
- [x] A measure of how unusual our observed statistic is if the null hypothesis is true.
- [x] The probability of obtaining a result like ours (or one more extreme) if the null hypothesis is true.
- [ ] The probability that the null hypothesis is true.
- [x] The strength of the evidence against the null hypothesis.

Question 7

1 pts

Suppose you have two tibbles, `comic_characters` and `publishers`.

```
> comic_characters
# A tibble: 9 x 3
  name          gender    publisher
  <chr>         <chr>     <chr>
1 Spider Man    masculine Marvel
2 Batman        masculine DC
3 Hellboy       masculine Dark Horse Comics
4 Poison Ivy    feminine  DC
5 Iron Man      masculine Marvel
6 Scarlet Witch feminine  Marvel
7 Nightwing     masculine DC
8 Harley Quinn  feminine  DC
9 The Sandman   masculine DC

> publishers
# A tibble: 2 x 2
  publisher founded
  <chr>       <dbl>
1 DC           1934
2 Marvel       1939
```

Which of the following commands would result in the below tibble? (I.e., With exactly the same ordering and number of rows and columns.) Tick all that apply.

```
# A tibble: 9 x 4
  name          gender    publisher         founded
  <chr>         <chr>     <chr>               <dbl>
1 Spider Man    masculine Marvel               1939
2 Batman        masculine DC                   1934
3 Hellboy       masculine Dark Horse Comics      NA
4 Poison Ivy    feminine  DC                   1934
5 Iron Man      masculine Marvel               1939
6 Scarlet Witch feminine  Marvel               1939
7 Nightwing     masculine DC                   1934
8 Harley Quinn  feminine  DC                   1934
9 The Sandman   masculine DC                   1934
```

Credit to Jenny Bryan and team's https://stat545.com/ for the inspiration for this question!

- [x] `full_join(comic_characters, publishers)`
- [ ] `right_join(publishers, comic_characters)`
- [ ] `left_join(publishers, comic_characters)`
- [x] `left_join(comic_characters, publishers)`
- [ ] `right_join(comic_characters, publishers)`

Question 8

1 pts

Suppose I realise I have spelt Spider-Man incorrectly in the `comic_characters` dataset, and that there is in fact supposed to be a hyphen between the two words. Which ONE of the following chunks of code would save over the original dataset, fix this typo AND not create any new errors?

```
> comic_characters
# A tibble: 9 x 3
  name          gender    publisher
  <chr>         <chr>     <chr>
1 Spider Man    masculine Marvel
2 Batman        masculine DC
3 Hellboy       masculine Dark Horse Comics
4 Poison Ivy    feminine  DC
5 Iron Man      masculine Marvel
6 Scarlet Witch feminine  Marvel
7 Nightwing     masculine DC
8 Harley Quinn  feminine  DC
9 The Sandman   masculine DC
```

- [ ] `comic_characters <- comic_characters %>%`

```
  mutate(name = str_c(name, "Spider-Man"))
```

```
comic_characters %>%
  mutate(name = str_replace(name, "Spider Man", "Spider-Ma
n"))
```

```
comic_characters <- comic_characters %>%
  str_replace(name, "Spider Man", "Spider-Man")
```

```
comic_characters <- comic_characters %>%
  mutate(name = str_replace(name, " Man", "-Man"))
```

```
comic_characters <- comic_characters %>%
  mutate(name = str_replace(name, "r M", "r-M"))
```

Quiz saved at 1:09am

Submit Quiz