

Week 3 quiz

Started: Jan 26 at 12:50pm

Quiz Instructions

This quiz will cover the asynchronous content from the [Week 3 module](#). It is due Wednesday, Jan 27 at 10:00 a.m. ET. I've made this quiz a little shorter as well (8 questions) because I know folks joining off the waitlist may also be trying to complete the Week 1 quiz in a fairly short overall time frame. Still weighted as normal.

There is one question about each reading from this week:

- Lum, K & Isaac, W. *To predict and serve?* (2016). <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x#sign960-bib-0001> (<https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x#sign960-bib-0001>)
- (<https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x#sign960-bib-0001>)
Aschwanden, C. (2015). *Science Isn't Broken: It's just a hell of a lot harder than we give it credit for*. Retrieved from <https://fivethirtyeight.com/features/science-isnt-broken> (<https://fivethirtyeight.com/features/science-isnt-broken>)

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 10:00 a.m. ET on Wednesday, 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.

Question 1

1 pts

Which of the following claims are **NOT** consistent with the claims made by Lum and Isaac in their 2016 article? Tick all that apply.

Lum, K & Isaac, W. *To predict and serve?* (2016). <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x#sign960-bib-0001>
(<https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x#sign960-bib-0001>)

- ☒ Predictive policing is evidence-based and neutral on issues of race.
- ☒ How the algorithms behind predictive policing are built is not usually publicly available.
- ☐ Initial biases in the policing data may become exacerbated as sending more police to an area is likely to result in more crime being detected, which is likely to result in more police being sent, which is likely to result inetc.
- ☐ Historical policing data is likely to reflect systematic biases.

Question 2

1 pts

Which of the following claims are **NOT** made by Aschwanden in *Science Isn't Broken* (2015). Tick all that apply.

Aschwanden, C. (2015). *Science Isn't Broken: It's just a hell of a lot harder than we give it credit for*. Retrieved from <https://fivethirtyeight.com/features/science-isnt-broken>
(<https://fivethirtyeight.com/features/science-isnt-broken>)

- ☒ In most cases, P-hacking is probably intentional fraud because of the pressure on researchers to publish.
- ☐ While what most scientists want to know is if their hypothesis is true or not, p-values cannot in fact tell you this.
- ☐ Approaching the same research question with different analytical strategies may result in different conclusions.
- ☒ If usually competent analysts don't come to the same conclusions about how to analyse a dataset, one of them is probably just being sloppy.

Question 3

1 pts

Which ONE of the following statements about web scraping and APIs is TRUE?

- ☐ The main reason we rate-limit is to avoid the website we're scraping detecting we are violating the terms and conditions.
- ☒ You should always provide credit to the source of your data when presenting results from web scraped data.
- ☐ If a website doesn't have a Terms and Conditions or Terms of Use page, this means you have permission to perform web scraping.
- ☐ If a website doesn't provide an API for you to access its data, you have the right to web scrape instead as long you provide credit and rate limit.

Question 4

1 pts

Suppose an automatic job application evaluation algorithm is being developed by a company. It is used to decide whether or not a candidate for a job should be invited to an interview. This company has been accused of sexism in the past because they have largely only hired men. They hope that not asking about gender in the application form will help with this. They use data about the applications of current successful employees and past rejected applications to train the classification algorithm they use. The company discovers, after using the algorithm for 6 months, that very few women are being invited to interviews and ask you to look at their model.

You discover that listing being a member of a sorority (a social organisation for women at a university) and/or playing on a women's volleyball team are highly predictive of **not** being invited to an interview. Which of the following statements about this situation are correct? Tick all that apply.

- ☒ We should be concerned that our training data is biased.
- ☒ Being a member of a sorority and playing on a women's volleyball team are proxies for gender.
- ☐ The algorithm cannot be sexist because gender is not part of the classification system, therefore it must just be that not very many women are applying for the job.
- ☒ If the company is using data about their current employees, who are mostly men, then very few 'successful' employees will have been in sororities or played on women's volleyball teams, therefore, these are unlikely to turn up as 'positives' when the applications are assessed by the algorithm.

- ☐ Previously rejected applications should not have been included in the training data.

Question 5

1 pts

A 2012 study showed that heavy use of marijuana in adolescence is associated with lower IQ. Of the following, which TWO would be the most sensible **confounders** to worry about for this study?

- ☐ IQ score
- ☒ Family environment
- ☒ Alcohol, tobacco and other drug use
- ☐ Marijuana use
- ☐ Average temperature on day of IQ test

Question 6

1 pts

Consider this excerpt from the results sections of a report. Which of the following criticisms are appropriate? (Note: this isn't an exhaustive list, there are lots of things wrong with this. The graph is from the vignette [here](https://allisonhorst.github.io/palmerpenguins/articles/intro.html) (<https://allisonhorst.github.io/palmerpenguins/articles/intro.html>.) Tick all that apply.

Table 1: Example of Palmer penguin data (palmerpenguins)

Table 1: Example of Palmer penguins data (palmerpen: age).

species	island	bill_length_mm	bill_depth_mm	body_r
Chinstrap	Dream	50.2	18.8	
Chinstrap	Dream	45.6	19.4	
Chinstrap	Dream	51.9	19.5	
Chinstrap	Dream	46.8	16.5	
Chinstrap	Dream	45.7	17.0	
Chinstrap	Dream	55.8	19.8	
Chinstrap	Dream	43.5	18.1	
Chinstrap	Dream	49.6	18.2	
Chinstrap	Dream	50.8	19.0	
Chinstrap	Dream	50.2	18.7	

penguins %>%

```
select(species, body_mass_g, ends_with("_mm")) %>%
  GGally::ggpairs(aes(color = species)) +
  scale_colour_manual(values = c("darkorange", "purple", "cyan4"))
  scale_fill_manual(values = c("darkorange", "purple", "cyan4"))
```



Figure 1: Exploring correlations between variables in the Palmer penguins dataset as a demonstration of the rich range of associations to explore in this dataset.

- ☒ Comments should be added to the code shown to better explain to a non-technical audience what each step is doing.

- ☐ In the context of a report, instead of giving the figure a caption, we should have used `labs(title= "...")` in the ggplot code to give it a descriptive title.
- ☐ The captions for the table and figure should be consistent, i.e., both on top or both on the bottom.
- ☒ Code should not be included in the results section.

Asbestos is a mineral that was widely used in building materials as fireproofing before it was discovered to cause a type of cancer called mesothelioma. It was largely banned in Canada in 2018, but even 100 years previously the U.S. Bureau of Labor Statistics has released a report about high levels of early mortality among workers exposed to asbestos.

The next two questions are about asbestos.

Question 7

1 pts

Suppose you were one of the researchers helping to establish that asbestos caused mesothelioma. Which ONE of the following study designs, *ignoring any other considerations*, would best help establish a causal relationship between asbestos exposure and mesothelioma?

- ☐ A prospective cohort study, where people who have been exposed asbestos and people who haven't are recruited to join a study and then whether or not they get mesothelioma in the future is observed.
- ☐ A large scale survey where people are asked a) if there were ever exposed to asbestos, and b) if they have ever been diagnosed with mesothelioma.
- ☐ A case-control study where participants with and without mesothelioma are recruited and then asked about their history of exposure to asbestos.
- ☒ A randomized control trial, where some participants are exposed to asbestos and others are exposed to a harmless placebo.

Question 8**1 pts**

Suppose you were one of the researchers helping to establish that asbestos caused mesothelioma. Which of the following study designs would be **approved by an ethics committee** AND allow you to comment on whether or not there was an association between asbestos exposure and mesothelioma? Tick all that apply.

- ☐ A randomized control trial, where some participants are exposed to asbestos and others are exposed to a harmless placebo.
- ☒ A prospective cohort study, where people who have been exposed asbestos and people who haven't are recruited to join a study and then whether or not they get mesothelioma in the future is observed.
- ☒ A large scale survey where people are asked a) if there were ever exposed to asbestos (and if so, when), and b) if they have ever been diagnosed with mesothelioma (and if so, when).
- ☒ A case-control study where participants with and without mesothelioma are recruited and then asked about their history of exposure to asbestos.

No new data to save. Last checked at 2:00am

[Submit Quiz](#)