P-Hacking Challenge

C. elegans

9/14/2021

```
p = c("readr", "dplyr", "ggplot2", "tidyr", "purrr",
      "ggplot2", "ggthemes")

lapply(p, require, character = TRUE)

data = read.csv("https://covid.ourworldindata.org/data/owid-covid-data.csv",
                na.strings = "",header=T)

data_test1 = data %>% filter(continent == 'Asia') %>%
  select(female_smokers, new_cases,iso_code, date, location) %>%
  drop_na()

cor.test(data_test1$female_smokers, data_test1$new_cases)
```

Findings:

We see a significant negative association (p = 1.142e-05) between share of female smokers in Asian countries and cases of COVID 19. Nicotine in cigarettes has been shown to be protective against COVID-19, attributing to fewer COVID 19 cases in Asian countries with higher share of female smokers.

Critique:

This apparent significant association can be attributed to: 1. countries with relatively higher share of female smokers (e.g. Kazakhstan, Israel) have "negative" new case values on one of the dates. 2. India which has a relatively lower share of female smokers has consistently had very high case numbers. These are some examples that might be driving this weak negative relationship. Additionally, there are numerous missing values in the dataset. We used a naive approach without considering population differences, checking the data to see if we had negative values for new cases, etc.