

Problem Set 2

Bern DySart

2024-01-31

Question 1:

<https://github.com/berndysart/PSet2.git>

Question 2: The following is the corrected code

```
knitr::opts_chunk$set(echo = TRUE,eval = TRUE,fig.align= 'center')
#load packages
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr) #added readr to open the data
library(knitr) #need for kable function

#open the data
gspace = read_csv("greenspace_data_share.csv") #added quotation marks around data name

## New names:
## Rows: 1038 Columns: 27
## -- Column specification
## ----- Delimiter: "," chr
## (10): City, Country, Major_Geo_Region, HDI_level, Climate_region, WHO_re... dbl
## (17): ...1, annual_avg_2010, peak_NDVI_2010, annual_weight_avg_2010, pea...
## i Use 'spec()' to retrieve the full column specification for this data. i
## Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## * ' -> '...1'

#summarize average urban greenspace by region
table =
  gspace |>
  group_by(Major_Geo_Region) |> #added a pipe
```

```

summarise( #indented since it goes further into the code
  obs = n(), #added a comma
  avg = mean(annual_avg_2020, na.rm = TRUE), #addeed na.rm = TRUE, added underscore
  weightedavg = mean(annual_weight_avg_2020, na.rm = TRUE) #addeed na.rm = TRUE, closed space
)
#output as table
kable(table, digits = 1) #table instead of gspace

```

Major_Geo_Region	obs	avg	weightedavg
Africa	154	0.3	0.2
Asia	569	0.3	0.3
Europe	128	0.3	0.3
Latin America and the Caribbean	120	0.3	0.3
Northern America	58	0.3	0.3
Oceania	9	0.3	0.3

Question 3: The greenspace data covers 1,038 urban greenspaces around the world

Question 4: As seen in the following table, the classification scores were identified within the variable of HDI_level, with four options that were chosen. There are 395 observations tht were associated with the response of high, 94 with low, 289 with medium, and 260 with very high. The average score for all of them were below 0.5 in both the average and weighted average data for the year 2021.

```

table2 =
  gspace |>
  group_by(HDI_level) |>
  summarise(
    obs = n(),
    avg = mean(annual_avg_2021, na.rm = TRUE),
    weightedavg = mean(annual_weight_avg_2021, na.rm = TRUE)
  )
kable(table2)

```

HDI_level	obs	avg	weightedavg
High	395	0.2769462	0.2543543
Low	94	0.2519638	0.2400238
Medium	289	0.3026253	0.2964196
Very high	260	0.2801116	0.2728995

Question 5:

a. 66 countries fit these conditions

```

gspace |>
  filter(indicator_2015== 'High'|indicator_2015== 'Very High')

```

b. 240 countries fit these conditions

```
gspace |>
  filter(indicator_2010== 'Exceptionally Low' | indicator_2015== 'Exceptionally Low' | indicator_2020 ==
         |indicator_2021 == 'Exceptionally Low')
```

c. 225 countries fit these conditions

```
gspace |>
  filter(Climate_region == 'Arid') |>
  filter(annual_weight_avg_2020 > annual_weight_avg_2010)
```

Question 6:

128 countries fit this condition

```
gspace |>
  filter(annual_avg_2021 < annual_avg_2010)
```

In examining the data, a majority of those who were less green over time were concentrated in temperate zones and also in Europe according to the data tables.

```
gspace2 =
  gspace |>
    filter(annual_avg_2021 < annual_avg_2010)

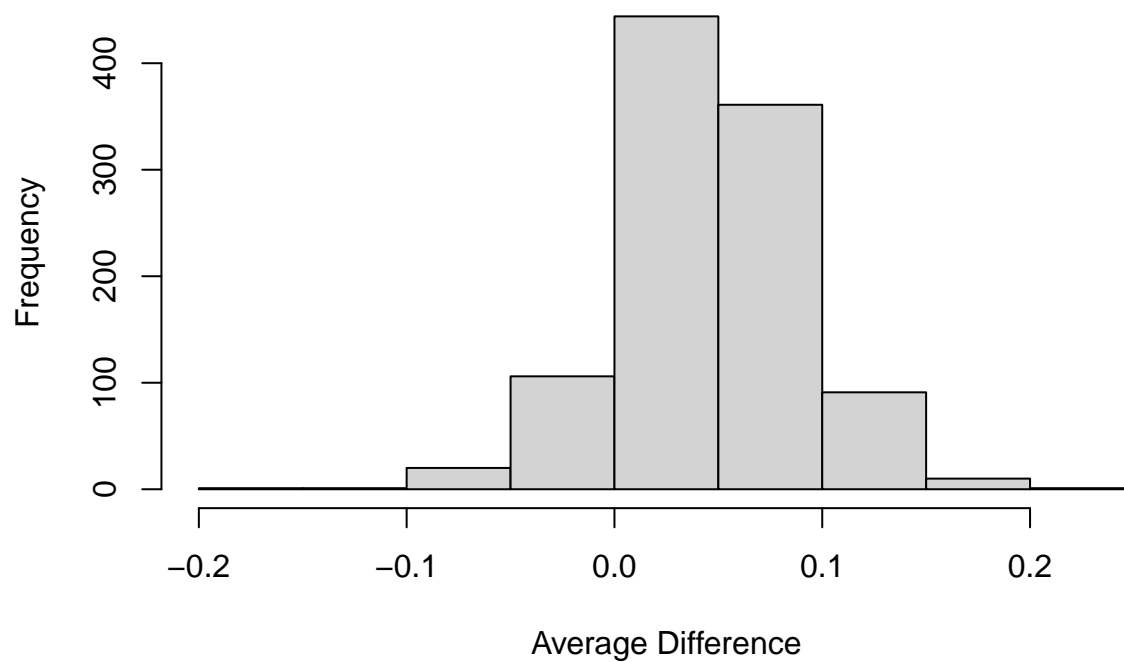
table3.1 =
  gspace2 |>
    count(Climate_region)

table3.2 =
  gspace2 |>
    count(Major_Geo_Region)
```

Question 7 See histogram below

```
gspace3 =
  gspace |>
    mutate(
      datadiff = annual_avg_2021 - annual_avg_2010,
    )
hist(gspace3$datadiff,
     xlab = 'Average Difference',
     main = 'Change in Greenspace from 2010 to 2021')
```

Change in Greenspace from 2010 to 2021



Question 8 See plot below

```
plot(annual_weight_avg_2021-annual_weight_avg_2010, gspace3,  
      xlab = '2010 Weighted Average',  
      ylab = '2021 Weighted Average',  
      main = 'Weighted Greenspace Average 2010-2021'  
)
```

Weighted Greenspace Average 2010–2021

