Problem Set 2

Jacob Álvarez

```r
  knitr::opts_chunk$set(
    echo = FALSE,
    eval = TRUE
  )
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.3

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'tibble' was built under R version 4.2.3

## Warning: package 'tidyr' was built under R version 4.2.3

## Warning: package 'readr' was built under R version 4.2.3

## Warning: package 'purrr' was built under R version 4.2.3

## Warning: package 'dplyr' was built under R version 4.2.3

## Warning: package 'stringr' was built under R version 4.2.3

## Warning: package 'forcats' was built under R version 4.2.3

## Warning: package 'lubridate' was built under R version 4.2.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     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
```

Answer the questions below to the best of your ability. Write clearly, and format your tables and visuals appropriately. You must use R Markdown to compose and compile your work. For full credit, echo all code chunks, and include your setup chunk. Submit your work in hard copy at the beginning of class.

You need the Global Greenspace Indicator Data for this assignment. Review the README.txt file for information about the data, variables, etc.

1. Show me that you're all set on GitHub. Create a public repository named PSet2. Clone it, include all your project files (including your .Rmd and .pdf files) for the work below, and commit/push your work to your repository. Include the link to your repo as your answer to this question.

https://github.com/ja2905a/PSet2/tree/main

2. The script below doesn't work. Type the corrected code chunk into your problem set. Annotate any line you correct to note your fix (i.e. # unbalanced parentheses). *There are more than five errors.*

```
# open my data
gspace = read_csv('greenspace_data_share.csv') #added quotes around file 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 pipe operator
summarise(
  obs = n(), #added comma
  avg = mean(annual_avg_2020), #added underscore in annual_avg_2020
  weighted_avg = mean(annual_weight_avg_2020) #added underscore in weighted_avg
)

# output as table
knitr::kable(table) #performed function on table instead of gspace
```

Major_Geo_Region	obs	avg	weighted_avg
Africa	154	0.2635577	0.2437269
Asia	569	0.2865218	0.2752794
Europe	128	0.3057736	0.2945913
Latin America and the Caribbean	120	0.2949339	0.2675823
Northern America	58	0.3160368	0.3129264
Oceania	9	0.3101189	0.3156935

```
#pulled kable function from unloaded knitr package
```

3. How many urban areas does the greenspace data cover?

```
knitr::opts_chunk$set(
    echo = TRUE,
    eval = TRUE
  )

table =
  gspace |>
  group_by(Major_Geo_Region) |> #added pipe operator
summarise(
  obs = n(), #added comma
  avg = mean(annual_avg_2020), #added underscore in annual_avg_2020
  weighted_avg = mean(annual_weight_avg_2020) #added underscore in weighted_avg
)

sum(table$obs)
```

```
## [1] 1038
```

The greenspace data has observations for 1038 urban areas.

4. In a couple of sentences and with reference to a well-formatted tabulation, describe the greenspace classification scores for urban areas in 2021.

```
#WARNING: I know this code is bad. I am sorry.

AnnualAvg21 = summary(gspace$annual_avg_2021)
PeakNVDI21 = summary(gspace$peak_NDVI_2021)
PeakWeight21 = summary(gspace$peak_weight_2021)
WeightAvg21 = summary(gspace$annual_weight_avg_2021)
```

```r
summary <- bind_rows(AnnualAvg21, PeakNVDI21, PeakWeight21, WeightAvg21)

summary21 <- summary[ -c(2, 5, 7)]

summary21$Var <- c("AnnualAvg21",
                   "PeakNVDI21",
                   "PeakWeight21",
                   "WeightAvg21")

knitr::kable(summary21)
```

Min.	Median	Mean	Max.	Var
0.04462531	0.2817494	0.2826559	0.6319837	AnnualAvg21
0.05072676	0.3555315	0.3507780	0.6603735	PeakNVDI21
0.06980129	0.3356772	0.3367330	0.6398629	PeakWeight21
0.06691043	0.2653515	0.2694419	0.6154216	WeightAvg21

The annual average NDVI ranged from 0.04 to 0.63, with both the median and mean at about 0.28, though the weighted annual NDVI had a slightly smaller range between 0.07 and 0.62, with median and mean both between 0.26 and 0.27. Peak NDVI ranged from 0.05 to 0.66, with both the median and mean between 0.35 and 0.36. The weighted peak NDVI ranged from 0.07 to 0.64, with median and mean both at 0.34.

5. Report the number of urban areas that satisfy the conditions below. Either write your code inline or echo the code that generated the answer.

   a. Scored `High` or above for greenspace in 2015.

```r
a <- gspace |>
  filter(indicator_2015 == "High" |
           indicator_2015 == "Very High" |
           indicator_2015 == "Exceptionally High")
count(a)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1    66
```

```
b. Scored ‘Exceptionally Low‘ at any point in the years covered.
```

```r
b <- gspace |>
  filter(indicator_2010 == "Exceptionally Low" |
           indicator_2015=="Exceptionally Low" |
           indicator_2020 == "Exceptionally Low" |
           indicator_2021 == "Exceptionally Low")
count(b)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   240
```

```
c. Urban areas in arid climate that became greener (as measured by annual weighted avera
```

```r
c <- gspace |>
  filter(Climate_region == "Arid",
         annual_weight_avg_2020 > annual_weight_avg_2010)
count(c)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   225
```

6. How many urban areas became less green (measured by annual average) from 2010 to 2021? Were these changes concentrated in a particular geographic or climate region? Explain (with evidence, of course).

```r
LessGreen <- gspace |>
  filter(annual_avg_2010 > annual_avg_2021)
count(LessGreen)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   128
```

```r
MajorGeoRegion <-
  LessGreen |>
  group_by(Major_Geo_Region) |>
  summarize(
  obs = n())
MajorGeoRegion |>
  mutate(MajorGeoRegionPer = obs / sum(obs) * 100)
```

```
## # A tibble: 5 x 3
##   Major_Geo_Region                   obs MajorGeoRegionPer
##   <chr>                            <int>             <dbl>
## 1 Africa                              25             19.5
## 2 Asia                                35             27.3
## 3 Europe                              47             36.7
## 4 Latin America and the Caribbean     12              9.38
## 5 Northern America                     9              7.03
```

```
ClimRegion <-
  LessGreen |>
  group_by(Climate_region) |>
  summarize(
    obs = n())
ClimRegion |>
  mutate(ClimRegionPer = obs / sum(obs) * 100)
```

```
## # A tibble: 5 x 3
##   Climate_region   obs ClimRegionPer
##   <chr>          <int>         <dbl>
## 1 Arid              12         9.38
## 2 Continental       36        28.1
## 3 Polar              1         0.781
## 4 Temperate         45        35.2
## 5 Tropical          34        26.6
```

A total of 128 urban areas became less green between 2010 and 2021. These decreases were most prevalent in Europe, where 37% of the 128 urban areas are, followed by Asia and Africa. Decreases were also most common in temperate regions where 35% of the 128 urban areas are, followed by continental and tropical climate regions.

7. Present a histogram showing the change in greenspace (annual average) from 2010 to 2021. Note that you will need to create a new variable equal to this difference.

8. Present a scatter plot of population weighted greenspace in 2021 over the greenspace in 2010.

BONUS OPPORTUNITY: Use color-coding to differentiate urban areas that added versus lost greenspace in that time. Then include a 45-degree line to further highlight the change.