

Problem Set 2

Kellie Haddon

2024-01-30

```
knitr::opts_chunk$set(
  echo = TRUE,
  eval = TRUE,
  fig.align = 'center',
  message = FALSE,
  warning = FALSE)

# load packages
library(tidyverse)

## -- 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 to become errors
```

1. <https://github.com/kelliehaddon/PSet2>

2. See script below.

```
library(tidyverse)

# open my data
gspace = read_csv('greenspace_data_share.csv') # added quotes around the file name

# summarize average urban greenspace by region
table =
  gspace |>
  group_by(Major_Geo_Region) |> # missing pipe
  summarize(
    obs = n(), # missing comma
    avg = mean(annual_avg_2020), # missing underscore between avg_2020
    weighted_avg = mean(annual_weight_avg_2020) # missing underscore between weighted_avg
  )

# output as table
knitr::kable(table, digits = 1) # added library knitr to access kable function
```

Major_Geo_Region	obs	avg	weighted_avg
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

changed object from "gspace" to "table"

3. The greenspace data covers 1,038 urban areas.

```
count(gspace)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1  1038
```

4. In 2021, both the mean and median of the annual average NDVI is 0.28, which is considered very low. The annual average NDVI spans from a minimum of 0.04 (extremely low) to a maximum of 0.63 (very high).

```
gs2021 = summarize(gspace,
  mean = mean(annual_avg_2021, na.rm = T),
  median = median(annual_avg_2021, na.rm = T),
  minimum = min(annual_avg_2021, na.rm = T),
  maximum = max(annual_avg_2021, na.rm = T))
knitr::kable(gs2021)
```

mean	median	minimum	maximum
0.2826559	0.2817494	0.0446253	0.6319837

5. a. 66 urban areas scored High or above for greenspace in 2015.

```
high2015 = filter(gspace, indicator_2015 == 'High' |
  indicator_2015 == 'Very High' |
  indicator_2015 == 'Exceptionally High')
count(high2015)
```

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

b. 240 urban areas scored Exceptionally Low at any point in the years covered.

```
exlow = filter(gspace,
               indicator_2010 == 'Exceptionally Low' |
               indicator_2015 == 'Exceptionally Low' |
               indicator_2020 == 'Exceptionally Low' |
               indicator_2021 == 'Exceptionally Low')
count(exlow)
```

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

c. 225 urban areas in arid climate became greener from 2010 to 2020.

```
less_arid = filter(gspace,
                   Climate_region == "Arid" &
                   annual_weight_avg_2010 < annual_weight_avg_2020)
count(less_arid)
```

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

6. 128 urban areas became less green from 2010 to 2021. These observations were not concentrated in a particular geographic or climate region. The top geographic and climate regions only exceeded the second most common geographic and climate regions by about 10 urban areas. Europe had the most urban areas fitting this criteria at 47, followed by Asia at 35 and Africa at 25. The country with the most urban areas fitting this criteria was Russia at 13 followed by the United States and India, both at 8. Temperate climates were the most popular among these areas at 45, followed by continental climates at 36 and tropical climates at 34.

```
less_green = filter(gspace, annual_avg_2010 > annual_avg_2021)
count(less_green)
```

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

```
count(less_green, Major_Geo_Region)
```

```
## # A tibble: 5 x 2
##   Major_Geo_Region      n
##   <chr>             <int>
## 1 Africa              25
## 2 Asia                35
## 3 Europe              47
## 4 Latin America and the Caribbean 12
## 5 Northern America     9
```

```
print(n = 120, count(less_green, Country))
```

```
## # A tibble: 56 x 2
##   Country          n
##   <chr>          <int>
## 1 Afghanistan      1
## 2 Angola            3
## 3 Argentina         3
## 4 Austria           1
## 5 Bangladesh        3
## 6 Belarus           1
## 7 Bhutan            1
## 8 Bolivia           1
## 9 Brazil            4
## 10 Cameroon         1
## 11 Canada           1
## 12 Chile            2
## 13 China             4
## 14 Comoros          1
## 15 Cote d'Ivoire    1
## 16 Democratic Republic of the Congo 3
## 17 Denmark          1
## 18 Estonia          1
## 19 Ethiopia         1
## 20 Finland          1
## 21 France           2
## 22 Germany          7
## 23 Ghana            1
## 24 Guinea-Bissau    1
## 25 Iceland          1
## 26 India            8
## 27 Indonesia        5
## 28 Islamic Republic of Iran 3
## 29 Kyrgyzstan       1
## 30 Latvia           1
## 31 Lithuania        1
## 32 Macedonia        1
## 33 Mauritius        1
## 34 Mexico           1
## 35 Moldova          1
## 36 Mozambique       1
## 37 Namibia          1
## 38 Nepal            1
## 39 Netherlands      1
## 40 Nigeria          6
## 41 Norway           1
## 42 Paraguay         1
## 43 Poland           4
## 44 Russian Federation 13
## 45 Sweden           1
## 46 Tajikistan       1
## 47 Thailand         1
## 48 Timor-Leste      1
```

```
## 49 Turkey 1
## 50 Uganda 1
## 51 Ukraine 3
## 52 United Kingdom 5
## 53 United Republic of Tanzania 1
## 54 United States of America 8
## 55 Uzbekistan 4
## 56 Zambia 2
```

```
count(less_green, Climate_region)
```

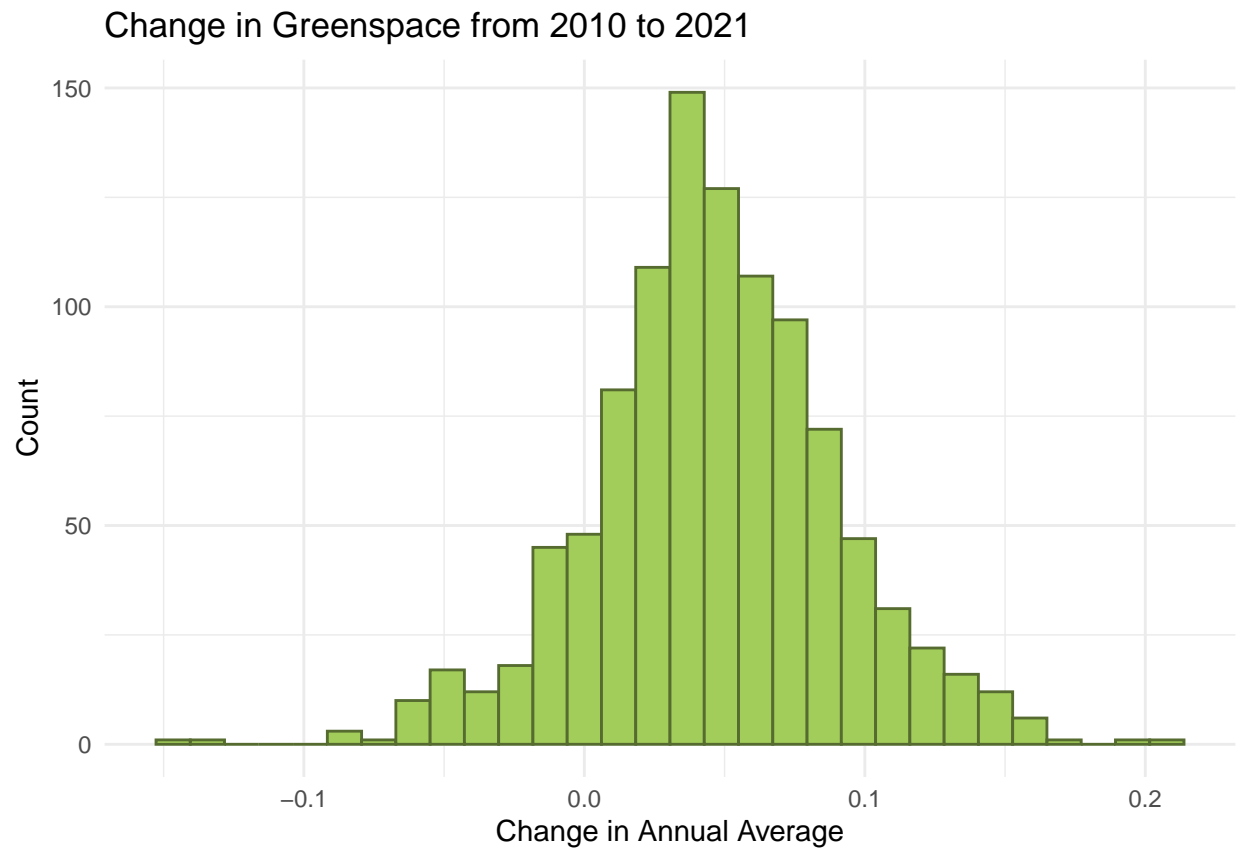
```
## # A tibble: 5 x 2
##   Climate_region     n
##   <chr>         <int>
## 1 Arid           12
## 2 Continental    36
## 3 Polar           1
## 4 Temperate      45
## 5 Tropical       34
```

7. See figure below.

```
gspace = mutate(gspace,
                 change = annual_avg_2021 - annual_avg_2010)

change_plot = ggplot(gspace) +
  geom_histogram(aes(change),
                 fill = 'darkolivegreen3',
                 color = 'darkolivegreen') +
  labs(x = 'Change in Annual Average',
       y = 'Count',
       title = 'Change in Greenspace from 2010 to 2021') +
  theme_minimal()

print(change_plot)
```



8. See figure below.

```
scatter = ggplot(gspace)+  
  geom_point(aes(x = annual_weight_avg_2010,  
                 y = annual_weight_avg_2021),  
            color = 'darkolivegreen4',  
            shape = 1) +  
  labs(x = '2010',  
       y = '2021',  
       title = 'Population Weighted Annual NDVI as Mean of Seasonal NDVI Values',  
       subtitle = '2010 vs 2021') +  
  theme_minimal()  
  
print(scatter)
```

Population Weighted Annual NDVI as Mean of Seasonal NDVI Values 2010 vs 2021



BONUS OPPORTUNITY: See figure below.

```
gspace = mutate(gspace,
                 change3 = annual_weight_avg_2021 - annual_weight_avg_2010)

gspace<-mutate(gspace,
               change4=case_when(
                 between(change3,0,1) ~ 'Added_Greenspace',
                 between(change3,-1,0) ~ 'Lost_Greenspace'))

scatterec = ggplot(gspace)+
  geom_point(aes(x = annual_weight_avg_2010,
                 y = annual_weight_avg_2021,
                 color = change4,
                 shape = 1) +
  labs(x = '2010',
        y = '2021',
        title = 'Population Weighted Annual NDVI as Mean of Seasonal NDVI Values',
        subtitle = '2010 vs 2021') +
  guides(color = guide_legend(title = '')) +
  geom_abline(intercept = 0, slope = 1) +
  scale_color_manual(values = c("darkolivegreen3", "burlywood4", "grey50"),
                     breaks = c("Added_Greenspace", "Lost_Greenspace")) +
  theme_minimal()

print(scatterec)
```

Population Weighted Annual NDVI as Mean of Seasonal NDVI Values
2010 vs 2021

