PSet2

Hamed Ahmadi

2024-01-31

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     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
```

Question 2:

```
gspace = read_csv('greenspace_data_share.csv') # Fix: enclosed the file name in quotes
```

```
## 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`
```

```
table = gspace |>
  group_by(Major_Geo_Region) |>
  summarize(
    obs = n(),
    avg = mean(annual_avg_2020, na.rm = TRUE), #Fix: corrected the variable name, removed NA values
    weighted_avg = mean(annual_weight_avg_2020, na.rm = TRUE) # Fix: removed space in 'weighted avg', r
  )


knitr::kable(table, digits = 1) # Fix: knitr library loaded, replaced gspace with table
```

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

Question 3:

```
 #Number of urban area

gspace |>
  distinct(City) |>
  count()
```

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

Question 4:

```
# Summary table for the indicator_2021 variable

ndvi_freq_2021 = gspace |>
  filter(!is.na(indicator_2021)) |>
  count(indicator_2021) |>
  mutate(Percentage = n / sum(n) * 100)

knitr::kable(ndvi_freq_2021, caption = "Frequency of NDVI Classification Levels for Urban Areas in 2021
```

Table 2: Frequency of NDVI Classification Levels for Urban Areas in 2021

indicator_2021	n	Percentage
Exceptionally Low	100	9.643201
High	62	5.978785
Low	394	37.994214
Moderate	215	20.732883
Very High	1	0.096432
Very Low	265	25.554484

In 2021, the majority of urban areas (73.2%) displayed poor score, with 100 areas rated exceptionally low, 265 very low, and 394 low in NDVI classification. Meanwhile, 215 areas were moderate, 62 high, and a single area was exceptionally high.

Question 5:

```r
#Scored High or above for green space in 2015

high_greenspace_2015 = gspace |>
  filter(indicator_2015 %in% c("High", "Very High", "Exceptionally High")) |>
  summarize(count = n())
print(high_greenspace_2015)
```

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

```r
#Scored Exceptionally Low at any point in the years covered

exceptionally_low_any = gspace |>
  filter(indicator_2010 == "Exceptionally Low" | indicator_2015 == "Exceptionally Low" | indicator_2020
  summarize(count = n())
print(exceptionally_low_any)
```

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

```r
#Urban areas in arid climate that became greener from 2010 to 2020

arid_greener_areas = gspace |>
  filter(Climate_region == "Arid") |>
  mutate(greener = annual_weight_avg_2020 > annual_weight_avg_2010) |>
  filter(greener == TRUE) |>
  summarize(count = n())
print(arid_greener_areas)
```

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

Question 6:

```r
#Number of urban areas became less green from 2010 to 2021

gspace_change = gspace |>
  mutate(change_in_greenspace = annual_avg_2021 - annual_avg_2010)

less_green_areas <- gspace_change |>
  filter(change_in_greenspace < 0)
nrow(less_green_areas)
```

```
## [1] 128
```

```r
#concentration in climate region

change_by_region = less_green_areas |>
  group_by(Major_Geo_Region) |>
  summarize(count = n(), average_change = mean(change_in_greenspace))

knitr::kable(change_by_region, caption = "change by region")
```

Table 3: change by region

Major_Geo_Region	count	average_change
Africa	25	-0.0223559
Asia	35	-0.0205310
Europe	47	-0.0408203
Latin America and the Caribbean	12	-0.0150832
Northern America	9	-0.0186815

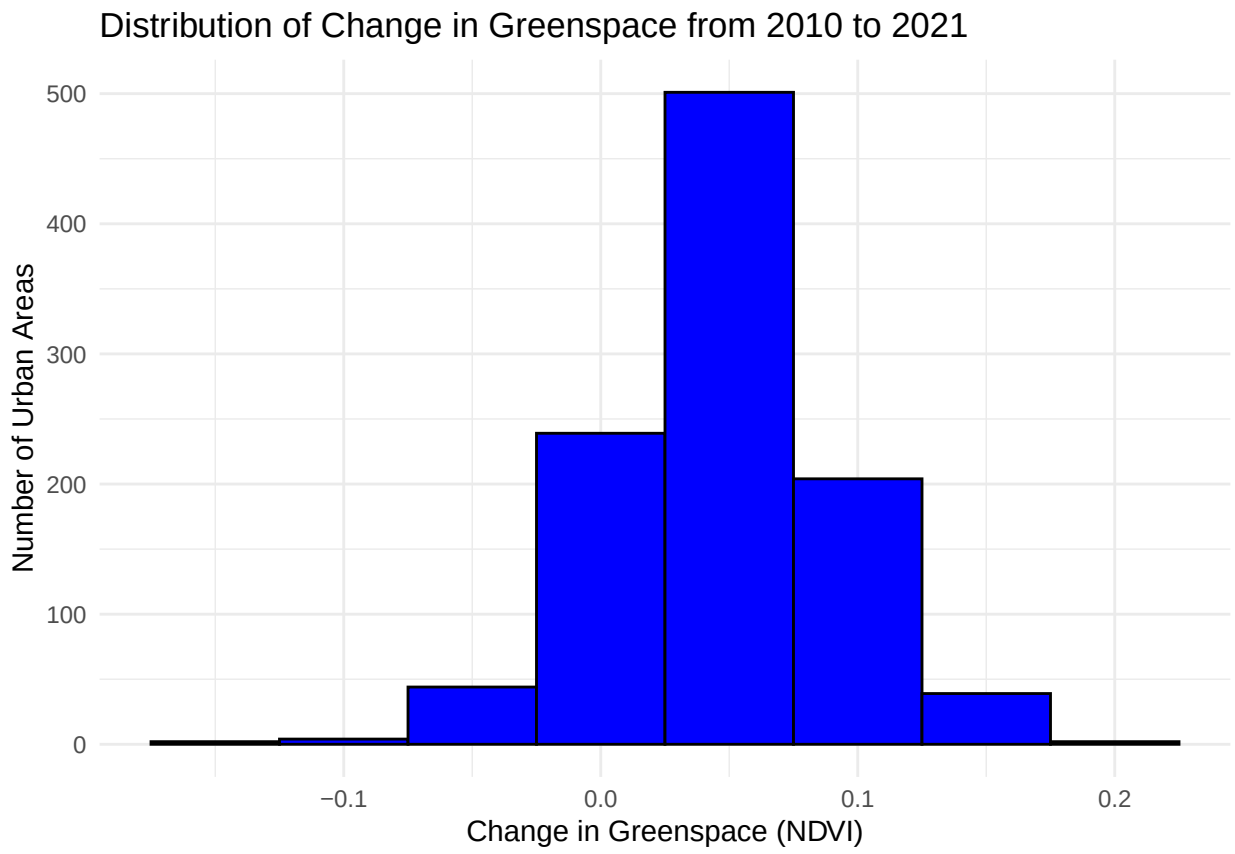
These changes are primarily concentrated in Europe, affecting 47 urban areas, followed by Asia with 35 areas, and Africa with 25 areas. Latin America and the Caribbean, along with North America, exhibited the lowest concentration of change, with 12 and 9 areas affected respectively. Moreover, Europe experienced the largest average change in urban greenery (-0.0408), while Latin America and the Caribbean has the lowest (-0.0151).

Question 7:

```r
#histogram showing the change in greenspace

ggplot(gspace_change, aes(x = change_in_greenspace)) +
  geom_histogram(binwidth = 0.05, fill = "blue", color = "black") +
  labs(title = "Distribution of Change in Greenspace from 2010 to 2021",
       x = "Change in Greenspace (NDVI)",
       y = "Number of Urban Areas") +
  theme_minimal()
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_bin()`).
```



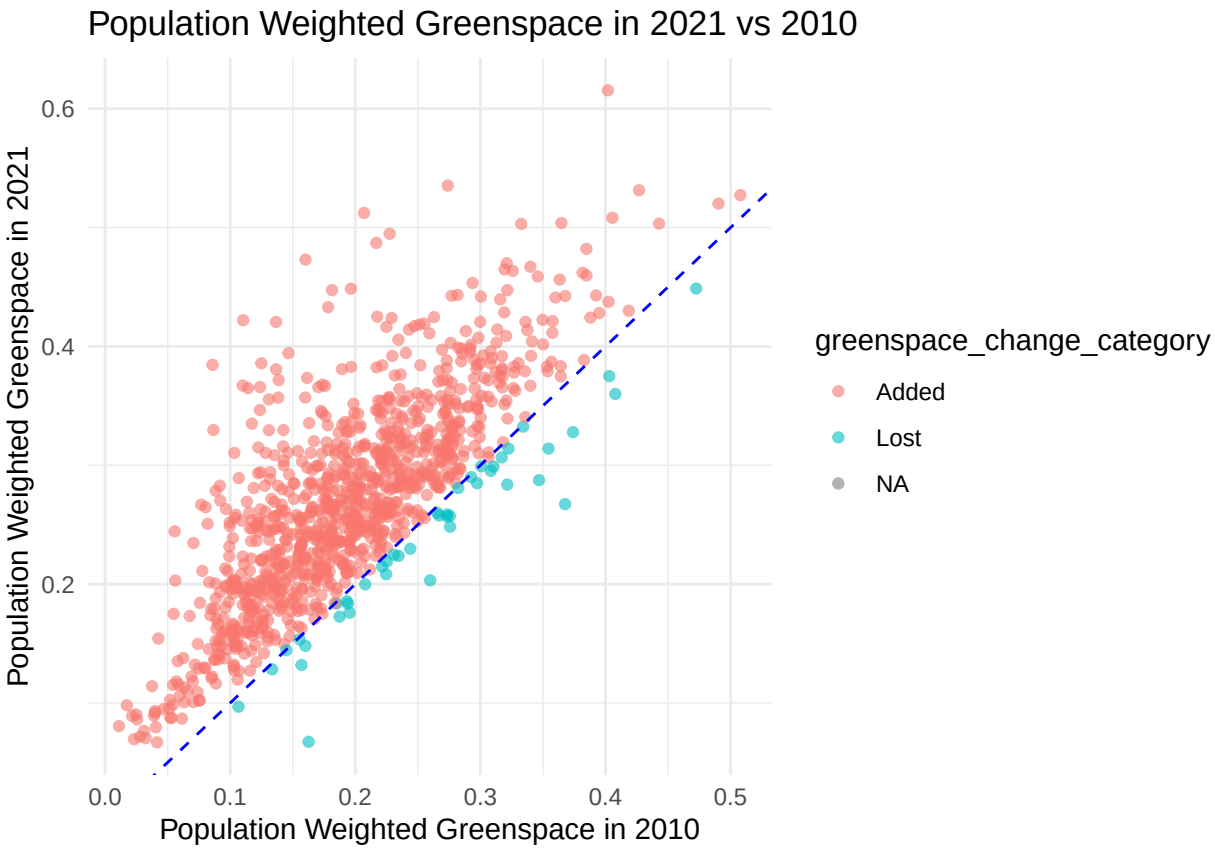
Question 8:

```
#scatter plot of population weighted green space in 2021 over the green space in 2010

gspace <- gspace %>%
  mutate(greenspace_change_category = if_else(annual_weight_avg_2021 > annual_weight_avg_2010, "Added",


ggplot(gspace, aes(x = annual_weight_avg_2010, y = annual_weight_avg_2021, color = greenspace_change_ca
  geom_point(alpha = 0.6) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "blue") +
  labs(title = "Population Weighted Greenspace in 2021 vs 2010",
       x = "Population Weighted Greenspace in 2010",
       y = "Population Weighted Greenspace in 2021") +
  theme_minimal()
```

```
## Warning: Removed 3 rows containing missing values (‘geom_point()‘).
```


Population Weighted Greenspace in 2021 vs 2010
Population Weighted Greenspace in 2021
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
Population Weighted Greenspace in 2010
greenspace_change_category
Added
Lost
NA