

PS 1 SOLUTION SET

Spring 2024

There are usually many ways to arrive at complete and correct answers. The solution set shows you one way. As always, study the keys (the .Rmd code and .pdf output), and reach out with questions.

Set code chunk options

```
knitr::opts_chunk$set(  
  echo = TRUE,  
  eval = TRUE,  
  fig.align = 'center',  
  message = FALSE,  
  warning = FALSE  
)  
  
knitr::knit_hooks$set(  
  inline = function(x) {  
    if(!is.numeric(x)){x}else{  
      prettyNum(round(x, 2), big.mark=",")  
    }  
  }  
)
```

Load packages

```
library(tidyverse)  
library(knitr)
```

Read greenspace data

```
df = read_csv("https://dataverse.harvard.edu/api/access/datafile/6903364")
```

clean

```
gspace =  
  df |>  
  select(-1) |>  
  mutate(  
    ind_2021f =  
      fct_relevel( # for the freq table  
        indicator_2021,  
        "Very High",  
        "High",  
        "Moderate",  
        "Low",  
        "Very Low",  
        "Exceptionally Low",  
      )  
  )  
)
```

1. The code and output files (Hart PSet2.Rmd and Hart PSet2.pdf) are available in my PSet2 repository on GitHub: <https://github.com/austin-hart-pols/PSet2>.
2. See the corrected code below. I identified 7 errors in total.

```
library(tidyverse)
library(knitr) # 1. load knitr for kable

# open my data
gspace = read_csv('greenspace_data_share.csv') # 2. quote filename

# summarize average urban greenspace by region
table =
  gspace |>
  group_by(Major_Geo_Region) |> # 3. add pipe
  summarise(
    obs = n(), # 4. comma at end
    avg = mean(annual_avg_2020), #5. fix var name
    `weighted avg` = mean(annual_weight_avg_2020) # 6. fix output name
  )
# output as table
kable(table, digits = 1) # 7. correct object is table
```

3. The Global Greenspace Indicator Data includes records for 1,038 urban areas.
4. The table below shows the Greenspace Indicator score for urban areas in 2021. Note that only 6% of urban areas are scored high to exceptionally high on this scale while almost three-quarters (roughly 75%) are scored low to exceptionally low. “Low” is the most common score (38%), and about one in five urban areas have “Moderate” greenspace.

```
gspace |>
  count(ind_2021f, name = "Freq") |>
  mutate(
    Per = Freq / sum(Freq) * 100
  ) |>
  rename(`Greenspace (2021)` = ind_2021f) |>
  kable(digits = 1L, caption = 'Greenspace in Urban Areas, 2021')
```

Table 1: Greenspace in Urban Areas, 2021

Greenspace (2021)	Freq	Per
Very High	1	0.1
High	62	6.0
Moderate	215	20.7
Low	394	38.0
Very Low	265	25.5
Exceptionally Low	100	9.6
NA	1	0.1

5. See below.
 - a. 66 urban areas scored High or above in 2015.
 - b. 240 areas recorded a score of exceptionally low at some point from 2010 to 2021.

- c. Among 230 urban areas in arid climates, 225 recorded increases in NDVI greenspace (annual, weighted average) from 2010 to 2020.
6. The table below shows the number of urban areas that lost greenspace (NDVI, annual average) from 2010 to 2021 by region. Of the 128 urban areas where the NDVI score declined, a strong plurality (37%) come from Europe.

```
gspace |>
  mutate(
    less_green = annual_avg_2010 > annual_avg_2021
  ) |>
  count(less_green, Major_Geo_Region) |>
  pivot_wider(
    names_from = less_green,
    values_from = n,
    values_fill = 0
  ) |>
  mutate(Percent = `TRUE` / (`FALSE` + `TRUE`) * 100) |>
  select(1, 3, 5) |>
  rename(
    Continent = Major_Geo_Region,
    Freq = `TRUE`
  ) |>
  kable(digits = 1L, caption = 'Loss of Greenspace by Region, 2010-21')
```

Table 2: Loss of Greenspace by Region, 2010-21

Continent	Freq	Percent
Africa	25	16.3
Asia	35	6.2
Europe	47	36.7
Latin America and the Caribbean	12	10.0
Northern America	9	15.5
Oceania	0	0.0

7. See the histogram below.

```
gspace |>
  mutate(
    change = annual_avg_2021 - annual_avg_2010
  ) |>
  ggplot(aes(x = change, fill = change > 0)) +
  geom_histogram(
    color = 'white',
    boundary = 0,
    closed = 'left'
  ) +
  labs(
    x = "Change in NDVI, 2010 to 2021",
    y = "Urban areas"
  ) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  scale_fill_manual(values = c('orange2', 'seagreen4')) +
```

```
theme_classic() +
theme(legend.position = 'none')
```

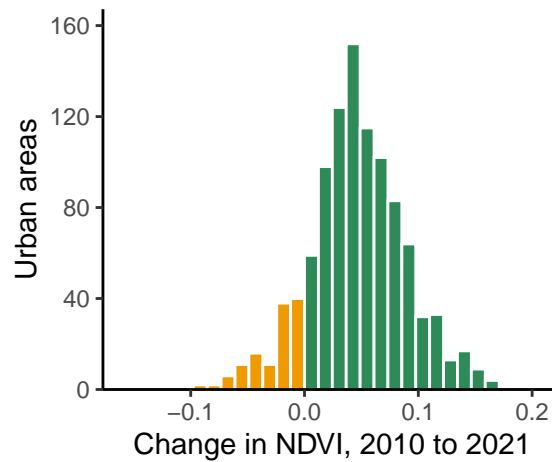


Figure 1: Growth and loss of urban greenspace, 2010 to 2021

8. The scatter plot below shows the response with the additional options for the bonus point.

```
gspace |>
  ggplot(aes(x = annual_weight_avg_2010, y = annual_weight_avg_2021)) +
  geom_abline(intercept = 0, slope = 1) +
  geom_point(
    aes(color = annual_weight_avg_2021 > annual_weight_avg_2010),
    alpha = .5
  ) +
  scale_y_continuous(limits = c(0, 0.75)) +
  scale_x_continuous(limits = c(0, 0.75)) +
  scale_color_manual(values = c('orange2', 'seagreen4')) +
  labs(
    y = 'NDVI 2021 (weighted)',
    x = 'NDVI 2010 (weighted)'
  ) +
  theme_classic() +
  theme(legend.position = 'none')
```

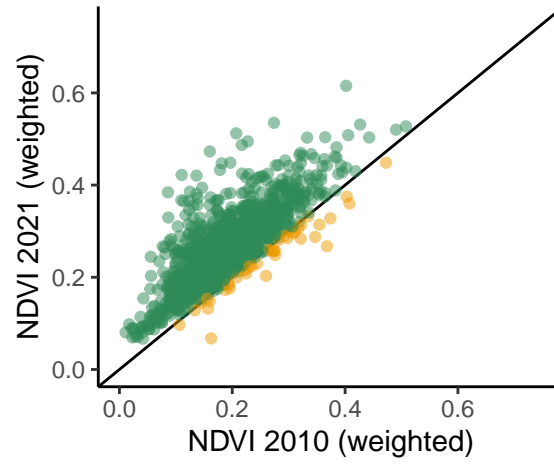


Figure 2: Changes in urban greenspace