

## Problem Set 2

Luke Maloney

2024-01-31

```
df = read_csv('greenspace_data_share.csv')
```

2. The edited script is below

```
#open my data
gspace = read_csv('greenspace_data_share.csv') #needed quotes

#summarize average urban greenspace by region
table =
  gspace |>
  group_by(Major_Geo_Region) |> #pipe added
  summarise(
    obs = n(),           #added comma
    avg = mean(annual_avg_2020, na.rm = TRUE)) #added na.rm=TRUE, added underscore, closed summarize p
    weightedavg = mean('annual_weight_avg_2020, na.rm = TRUE') # added na.rm = TRUE, added quotes, del

# output as table
kable(table, digits = 1L) #change gspace to table and add "l" after 1
```

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

3. Greenspace covers 1038 cities.

```
summary(gspace$City)
```

```
##      Length      Class      Mode
##      1038 character character
```

4. Using the following code we can see that greenspace reported 100 cases of Extremely Low, 265 of Very Low, 394 of Low, 215 of Moderate, 62 of High, and 1 of Very High.

```
table2=
  gspace |>
  group_by(indicator_2021) |>
  summarise(
    obs = n())

kable(table2, format = "markdown")
```

indicator_2021	obs
Exceptionally Low	100
High	62
Low	394
Moderate	215
Very High	1
Very Low	265
NA	1

5. a. I created a table that shows the number of observations from 2015 that were recorded as High or Very High. There were 66 such cases.

```
table5 = filter(gspace, indicator_2015 == "High" | indicator_2015 == "Very High") |>
  group_by(indicator_2015) |>
  summarise(
    obs = n())

kable(table5, format = "markdown")
```

indicator_2015	obs
High	62
Very High	4

5. b. There were 240 cases where a municipality scored exceptionally low.

```
table6 = filter(gspace, indicator_2010 == "Exceptionally Low" | indicator_2015 == "Exceptionally Low" |
  summarise(
    obs = n())

kable(table6, format = "markdown")
```

obs
240

5. c. There have been 225 cases in which arid urban areas became greener from 2010 to 2020.

```
gspace <- transform(gspace,
  greener = annual_weight_avg_2020 - annual_weight_avg_2010
)

table7 = filter(gspace, Climate_region == "Arid" , greener > 0) |>
  summarise(
    obs = n())

kable(table7, format = "markdown")
```

obs
225

6. 128 urban areas became less green between 2010 and 2021. These changes are concentrated in Europe and temperate regions as seen on the charts below.

```
gspace <- transform(gspace,
  tacos = annual_avg_2021 - annual_avg_2010
)

table8 = filter(gspace, tacos < 0) |>
  summarize(
    obs = n()
  )
kable(table8, format = "markdown")
```

obs
128

```
table3=
  gspace |>
  group_by(Major_Geo_Region) |>
  filter(tacos < 0) |>
  summarize(
    obs = n()
  )
kable(table3, format = "markdown")
```

Major_Geo_Region	obs
Africa	25
Asia	35
Europe	47
Latin America and the Caribbean	12
Northern America	9

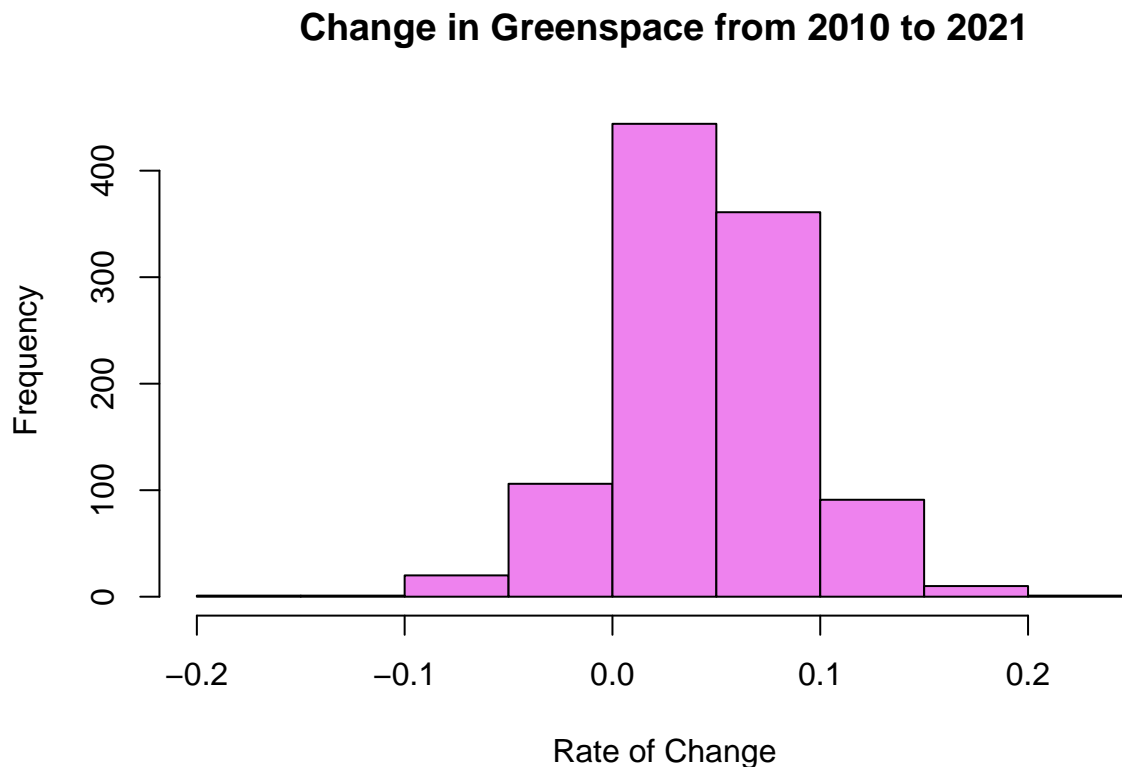
```
table4=
  gspace |>
  group_by(Climate_region) |>
  filter(tacos < 0) |>
  summarise(
    obs = n()
  )
kable(table4, format = "markdown")
```

Climate_region	obs
Arid	12
Continental	36
Polar	1
Temperate	45

Climate_region	obs
Tropical	34

7. Below is a histogram that demonstrates the change in greenspace from 2010 to 2021.

```
hist(gspace$tacos, main = "Change in Greenspace from 2010 to 2021",
     xlab = "Rate of Change",
     ylab = "Frequency",
     col = "violet", border = "black")
```



8. Find a scatter plot of population weighted greenspace in 2021 over 2010 below.

```
df$slope <- (df$annual_weight_avg_2021 - df$annual_weight_avg_2010) / df$annual_weight_avg_2010
ggplot(df, aes(x = annual_weight_avg_2010,
               y = annual_weight_avg_2021)) + geom_point(aes
               (color = ifelse(slope > 0, 'Gained Greenspace',
                               ifelse(slope < 0, 'Lost Greenspace', 'Stayed the Same')))) +
  geom_abline(intercept = 0, slope = 1, linetype = "dotted", color = "red4") +
  labs(title = 'Very Official Scatter Plot', x = 'Annual Weighted Average 2010',
       y = 'Annual Weighted Average 2021') +
  scale_color_manual(values = c('Stayed the Same' = 'black',
                                'Lost Greenspace' = 'coral',
                                'Gained Greenspace' = 'cyan2'), name = 'Legend')
```

Very Official Scatter Plot

