CODECHUNKPortfolio

Bandar Alsharhan

2024-04-30

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)
invisible(suppressPackageStartupMessages({
  library(tidyverse)
  library(kableExtra)
}))
```

```
# Download the penguin data directly from a URL
path2data <- "https://dataverse.harvard.edu/api/access/datafile/6903364"
gspace <- read_csv(path2data)

table <-
  gspace |>
  group_by(Major_Geo_Region) |>
  summarise(
    obs = n(),
    avg = mean(annual_avg_2020, na.rm = TRUE),
    weighted_avg = mean(annual_weight_avg_2020, na.rm = TRUE)
  )
# Output as table
kable(table, digits = 1)
```

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

- Brief explanation: This code snippet fixes errors in the original script and then calculates summary statistics for average urban greenspace (annual_avg_2020) by major geographical regions (Major_Geo_Region) using the dplyr package for data manipulation and knitr for generating a nicely formatted table.This code demonstrates effective use of data manipulation and summarization techniques that was done with dplyr, proper handling of missing data, and generation of neat output using kable for presentation-ready tables in R.