ANA 515 Assignment 1

Prashant Karki

2023-01-21

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.4.0      ✓ purrr   1.0.1
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.5.0
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(knitr)
library(bslib)
```

```
##
## Attaching package: 'bslib'
##
## The following object is masked from 'package:utils':
##
##     page
```

```
library(ggplot2)
library(dplyr)
```

```
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/fifa/fifa_coun
tries_audience.csv"

# read the dataset from the url
dataset <- read_csv(url)
```

```
## Rows: 191 Columns: 5
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): country, confederation
## dbl (3): population_share, tv_audience_share, gdp_weighted_share
##
## 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.
```

```
# view the first few rows of the dataset
head(dataset)
```

```
## # A tibble: 6 × 5
##   country        confederation population_share tv_audience_share gdp_weighted…¹
##   <chr>          <chr>                   <dbl>              <dbl>          <dbl>
## 1 United States  CONCACAF                  4.5                4.3           11.3
## 2 Japan          AFC                       1.9                4.9            9.1
## 3 China          AFC                      19.5               14.8            7.3
## 4 Germany        UEFA                      1.2                2.9            6.3
## 5 Brazil         CONMEBOL                  2.8                7.1            5.4
## 6 United Kingdom UEFA                      0.9                2.1            4.2
## # … with abbreviated variable name ¹gdp_weighted_share
```

```
#Filtering data with population share greater than 1

filtered <- dataset %>%
  filter(population_share > 1.0)

head(filtered)
```

```
## # A tibble: 6 × 5
##   country       confederation population_share tv_audience_share gdp_weighted_…¹
##   <chr>         <chr>                    <dbl>              <dbl>            <dbl>
## 1 United States CONCACAF                   4.5                4.3             11.3
## 2 Japan         AFC                        1.9                4.9              9.1
## 3 China         AFC                       19.5               14.8              7.3
## 4 Germany       UEFA                       1.2                2.9              6.3
## 5 Brazil        CONMEBOL                   2.8                7.1              5.4
## 6 Russia        UEFA                       2.1                3.1              3.5
## # … with abbreviated variable name ¹gdp_weighted_share
```

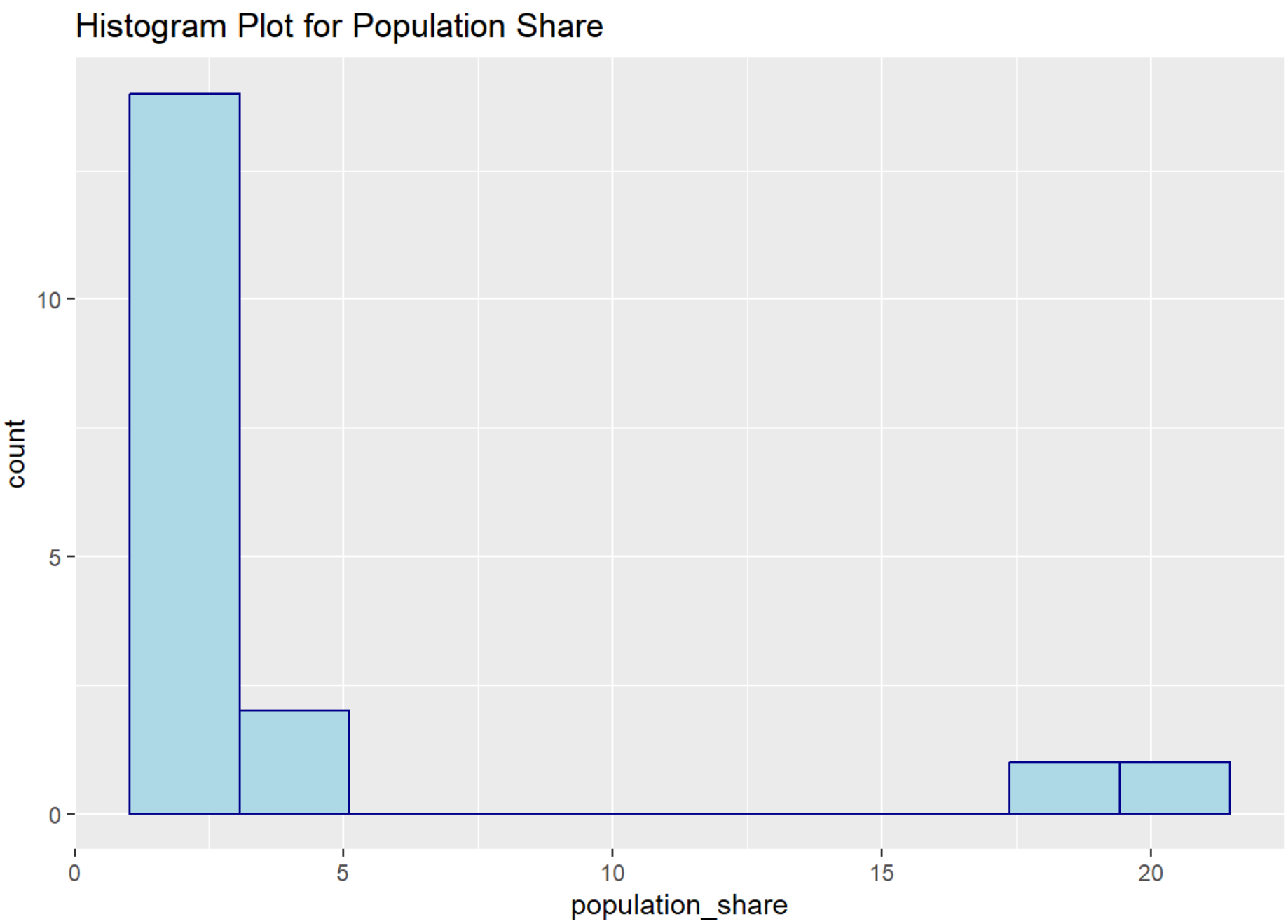
```
#Summary of filtered dataset
summary(filtered)
```

```
##    country           confederation      population_share tv_audience_share
##  Length:18          Length:18          Min.   : 1.100   Min.   : 0.100
##  Class :character   Class :character   1st Qu.: 1.300   1st Qu.: 0.725
##  Mode  :character   Mode  :character   Median : 2.000   Median : 2.600
##                                        Mean   : 3.839   Mean   : 3.267
##                                        3rd Qu.: 2.725   3rd Qu.: 4.025
##                                        Max.   :19.500   Max.   :14.800
##  gdp_weighted_share
##  Min.   : 0.000
##  1st Qu.: 0.500
##  Median : 1.350
##  Mean   : 2.967
##  3rd Qu.: 4.925
##  Max.   :11.300
```

We have data about 191 countries. Only 191- 18 have a population share of over 1.0. The distribution of the remainder is shown below:

Including histogram plots for filtered dataset for population share and gdp weighted share

```
ggplot(filtered,aes(population_share))+geom_histogram(color="darkblue", fill="lightb
lue",bins=10)+ ggtitle("Histogram Plot for Population Share")
```


Histogram Plot for Population Share
count
10
5
0
0
5
10
15
20
population_share

```
ggplot(filtered,aes(gdp_weighted_share))+geom_histogram(color="green", fill="yello
w",bins=10)+ ggtitle("Histogram Plot of GDP Weighted Share")
```


Histogram Plot of GDP Weighted Share
count
8
6
4
2
0
0
4
8
12
gdp_weighted_share