LAB5_ESP106

Hector Carvajal

2025-02-03

```
## Loading required package: knitr
```

Lab 5

Due Tuesday Feb4th - Recommended to complete this before starting the midterm

This lab we will look at some data from the plastic trash piced up during clean-up events around the world. I took this dataset from the Tidy Tuesday website. You can read the documentation here, including the references and description of the different column names.

I have done some pre-processing of the data for you for this lab, to create two more easy-to-use dataframes.

First read in the countrytotals.csv data frame

```
##   X    country empty hdpe ldpe    o pet  pp  ps pvc total num_events volunteers
## 1 1  Argentina   128  132  141  212 144 263  62   4  1086       5424       2034
## 2 2    Armenia     0    0    0    0   8   0   0   0     8          4         12
## 3 3  Australia     4   33  222  898 141 361   3   6  1668        621      13110
## 4 4 Bangladesh     0  376    2 1559 299  89  58   0  2383        142      18034
## 5 5      Benin     0    0    0    0 342   0   0   0   342          4        916
## 6 6     Brazil     0  114  572  376 417 548 208  22  2257         88       1716
```

```
## 'data.frame':    55 obs. of  13 variables:
##  $ X         : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ country   : chr  "Argentina" "Armenia" "Australia" "Bangladesh" ...
##  $ empty     : int  128 0 4 0 0 0 8 0 0 0 ...
##  $ hdpe      : int  132 0 33 376 0 114 11 267 117 211 ...
##  $ ldpe      : int  141 0 222 2 0 572 34 20 611 226 ...
##  $ o         : int  212 0 898 1559 0 376 125 3623 440 700 ...
##  $ pet       : int  144 8 141 299 342 417 74 4664 283 456 ...
##  $ pp        : int  263 0 361 89 0 548 199 68 277 383 ...
##  $ ps        : int  62 0 3 58 0 208 15 90 159 43 ...
##  $ pvc       : int  4 0 6 0 0 22 14 42 0 14 ...
##  $ total     : int  1086 8 1668 2383 342 2257 480 8774 1887 2033 ...
##  $ num_events: int  5424 4 621 142 4 88 105 15 129 363 ...
##  $ volunteers: int  2034 12 13110 18034 916 1716 750 2445 559 1573 ...
```

```
##  [1] "X"          "country"    "empty"      "hdpe"       "ldpe"
##  [6] "o"          "pet"        "pp"         "ps"         "pvc"
## [11] "total"      "num_events" "volunteers"
```

Have a look at the data frame. Then column "total" gives the total number of pieces of plastic picked up in that country in 2020. The columns "num_events" and "volunteers" give the number of trash pick-up events and the number of volunteers in that country. We are going to use this to investigate where the plastic trash problem is worst.

1. What 5 countries had the worst plastic problem as measured by the number of pieces of trash picked up?

```
#descending order of total plastic pieces in each country
country_totals_sorted <- country_totals[order(country_totals$total, decreasing = TRUE), ]

#top 5 countries with the most plastic trash
top_5_countries <- head(country_totals_sorted, 5)
top_5_countries
```

```
##     X     country empty hdpe  ldpe     o   pet    pp   ps  pvc total num_events
## 35 35     Nigeria   316 7532 15470  8284 27390  2234 2011   16 63253       6030
## 37 37 Philippines   344   82  3939  4272  2949 41987  600 1011 55184       2040
## 46 46 Switzerland  2342  102  1528 45612  1018   902  722   51 52277       1746
## 21 21       India    22  489  6684  7369   895  1107   67  340 16973      19488
## 49 49        Togo     0    0     0 11233   759     0    2    0 11994          5
##    volunteers
## 35     703165
## 37     109800
## 46      52380
## 21     122844
## 49        460
```

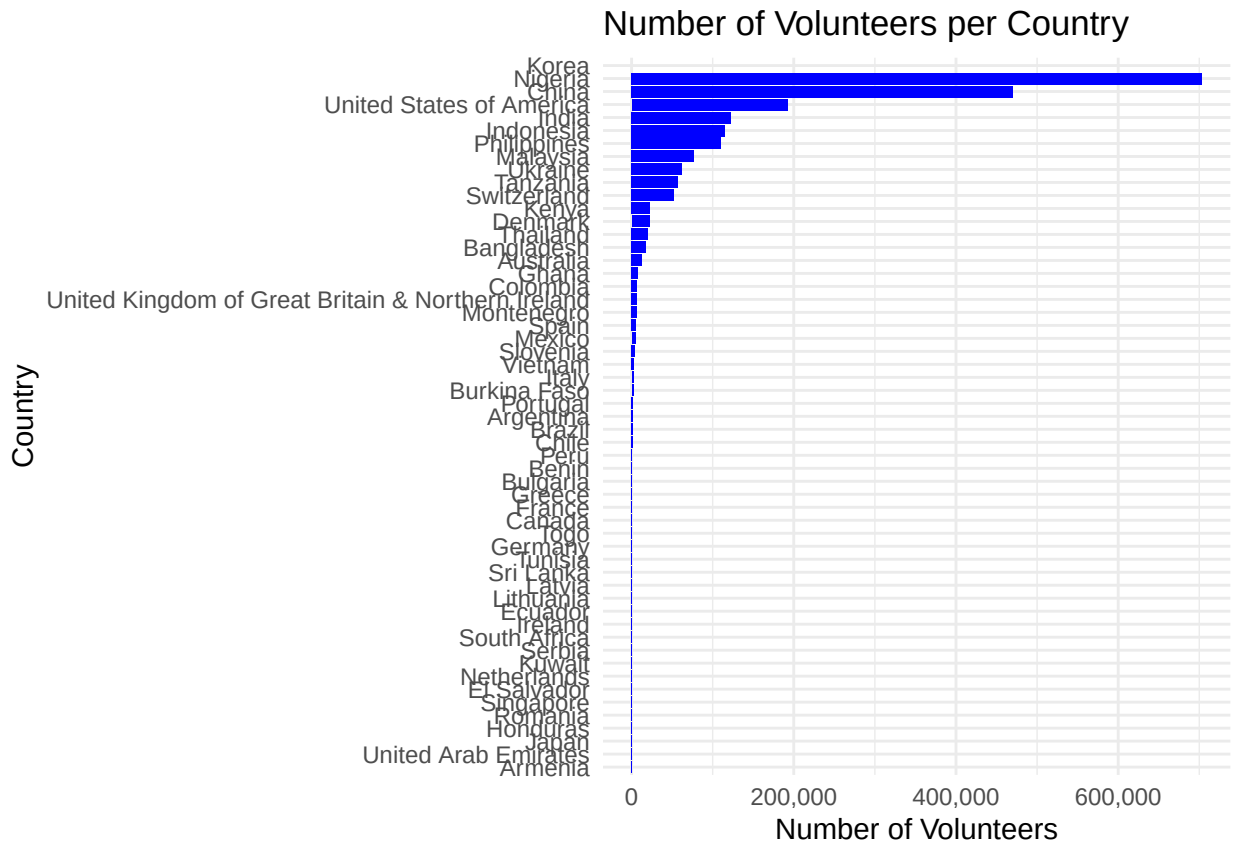
Answer: The countries with the worst plastic pollution from least to most measured by individual plastic pieces collected by volunteers are Togo - 11994, India - 16973, Switzerland - 52277, Philippines - 55184, and Nigeria - 63253.

2. Make a plot showing the distribution of volunteers across countries

```
library(ggplot2)
```

```
library(scales)
ggplot(country_totals, aes(x = reorder(country, volunteers), y = volunteers)) +
  geom_bar(stat = "identity", fill = "blue") +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Number of Volunteers per Country",
       x = "Country",
       y = "Number of Volunteers") +
#had to flip the coordinates given the names were overlapping
  coord_flip() +
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_bar()`).
```



3. Notice that there is a lot of variation across countries in the number of volunteers involved in trash pickup. What problem might that cause for the interpretation of your answer to question 1?

Answer: The problem with finding a variation of the amount of volunteers across countries is it could make it seem like some countries have worse polluted areas just because they have more volunteers collecting trash. This could make it appear that countries with fewer volunteers have less polluted areas. The ranking in question 1 may reflect volunteer efforts rather than actual pollution levels.

4. Add a column to the data frame creating a variable that should be more closely related to the presence of plastic pollution in the country

```
#adding a column for plastic collected per event
country_totals$plastic_per_event <- country_totals$total / country_totals$num_events
country_totals
```

```
##     X                                                  country empty hdpe  ldpe     o
## 1   1                                                Argentina   128  132   141   212
## 2   2                                                  Armenia     0    0     0     0
## 3   3                                                Australia     4   33   222   898
## 4   4                                               Bangladesh     0  376     2  1559
## 5   5                                                    Benin     0    0     0     0
## 6   6                                                   Brazil     0  114   572   376
## 7   7                                                 Bulgaria     8   11    34   125
## 8   8                                             Burkina Faso     0  267    20  3623
```

```
## 9   9                                                 Canada     0  117   611   440
## 10 10                                                  Chile     0  211   226   700
## 11 11                                                  China     0    2     0  2204
## 12 12                                               Colombia    18  357   413  1582
## 13 13                                                Denmark     0   68    22  4343
## 14 14                                                Ecuador     8    0     0     1
## 15 15                                            El Salvador     0    2   180    53
## 16 16                                                 France     0   10    73  2550
## 17 17                                                Germany    10    2   210  4790
## 18 18                                                  Ghana     0  220   291     0
## 19 19                                                 Greece     0    0     6    26
## 20 20                                               Honduras     0    0     0     1
## 21 21                                                  India    22  489  6684  7369
## 22 22                                              Indonesia     7 1304  1461  1473
## 23 23                                                Ireland     0    0     0   141
## 24 24                                                  Italy     0   10     1    20
## 25 25                                                  Japan   207    0     0    43
## 26 26                                                  Kenya     0  340    61   294
## 27 27                                                  Korea     4    8   111  1235
## 28 28                                                 Kuwait     0   37    30   618
## 29 29                                                 Latvia     0    6     9   389
## 30 30                                              Lithuania     0  109     3   487
## 31 31                                               Malaysia     0  140    86  2909
## 32 32                                                 Mexico    42  134   288   421
## 33 33                                             Montenegro   101    0   225    21
## 34 34                                            Netherlands     0    0    94   429
## 35 35                                                Nigeria   316 7532 15470  8284
## 36 36                                                   Peru     0   26    97   289
## 37 37                                            Philippines   344   82  3939  4272
## 38 38                                               Portugal     3 2594  1184  4157
## 39 39                                                Romania     0    0     0    51
## 40 40                                                 Serbia     0    0     0   234
## 41 41                                              Singapore     0   15     0     0
## 42 42                                               Slovenia     0   11   132   306
## 43 43                                           South Africa     0   79     0     8
## 44 44                                                  Spain     0  123   132   326
## 45 45                                              Sri Lanka     0    0     0    15
## 46 46                                            Switzerland  2342  102  1528 45612
## 47 47                                               Tanzania     2  113     3  2781
## 48 48                                               Thailand     0  312     2   971
## 49 49                                                   Togo     0    0     0 11233
## 50 50                                                Tunisia    17  167   615   541
## 51 51                                                Ukraine     0  428   509  3700
## 52 52                                   United Arab Emirates     1    0     0     0
## 53 53 United Kingdom of Great Britain & Northern Ireland     3  153   752  1069
## 54 54                               United States of America   475  654   825  4485
## 55 55                                                Vietnam   112  212   198  4779
##      pet    pp   ps  pvc total num_events volunteers plastic_per_event
## 1    144   263   62    4  1086       5424       2034         0.2002212
## 2      8     0    0    0     8          4         12         2.0000000
## 3    141   361    3    6  1668        621      13110         2.6859903
## 4    299    89   58    0  2383        142      18034        16.7816901
## 5    342     0    0    0   342          4        916        85.5000000
## 6    417   548  208   22  2257         88       1716        25.6477273
```

```
## 7     74   199   15   14   480        105        750         4.5714286
## 8   4664    68   90   42  8774         15       2445       584.9333333
## 9    283   277  159    0  1887        129        559        14.6279070
## 10   456   383   43   14  2033        363       1573         5.6005510
## 11   126     0    0    0  2332        216     469800        10.7962963
## 12  3549   278    3   15  6215        250       6850        24.8600000
## 13    48   149  162    1  4793       2184      22022         2.1945971
## 14   176    19    1   68   273         44        242         6.2045455
## 15   185   318  182    0   920          5         50       184.0000000
## 16   141    65  131   20  2990        123        615        24.3089431
## 17   102    95    7    0  5216         19        456       274.5263158
## 18  8272     0    0    0  8783         52       8060       168.9038462
## 19     4     3    0    0    NA         18        657                NA
## 20     1     0    0    0     2          2         24         1.0000000
## 21   895  1107   67  340 16973      19488     122844         0.8709462
## 22  2818  2912  145   36 10156      14700     115248         0.6908844
## 23    87    46    3    2   279         80        140         3.4875000
## 24   267     8    3    0   309         48       2464         6.4375000
## 25     3     4    3    5   265          5         15        53.0000000
## 26  7797   364    0    0  8856        114      23104        77.6842105
## 27    68   133   96    2  1663       2782         NA         0.5977714
## 28    53   124   40   33   935         19         57        49.2105263
## 29    46    86    4    0   540         71        355         7.6056338
## 30    41   251   55    1   947         19        285        49.8421053
## 31  2612   549  129   14  6439       7245      76590         0.8887509
## 32   694   535  269   43  2426       1888       4897         1.2849576
## 33   664     1   50    5  1067         15       6240        71.1333333
## 34    17    26    4    3   573         11         55        52.0909091
## 35 27390  2234 2011   16 63253       6030     703165        10.4897181
## 36    81   109   37    1   640         46       1035        13.9130435
## 37  2949 41987  600 1011 55184       2040     109800        27.0509804
## 38    72   388    7    1  8406         64       2080       131.3437500
## 39    37     0    1    0    89          4         28        22.2500000
## 40   293     3    0    0   530          9        117        58.8888889
## 41    80     0    0    0    95          8         32        11.8750000
## 42    76    62    9    3   599         62       4340         9.6612903
## 43   770     0    0    0   857          6        120       142.8333333
## 44   134    29  244   15  1003        231       5964         4.3419913
## 45     7     0    0    0    NA         11        440                NA
## 46  1018   902  722   51 52277       1746      52380        29.9410080
## 47  4106   264   49    0  7318        315      57099        23.2317460
## 48  1108    75  109  106  2683        531      19883         5.0527307
## 49   759     0    2    0 11994          5        460      2398.8000000
## 50   596   362    6   15  2319         18        441       128.8333333
## 51  1835  1195  208   15  7890       4420      61880         1.7850679
## 52     0     0    0    0     1          1         12         1.0000000
## 53   480   942 1122  101  4622        600       6750         7.7033333
## 54  3103  3513  757  179    NA      50384     192136                NA
## 55  1516   394 1587   11  8809        768       3456        11.4700521
```

5. What 5 countries have the worst plastic pollution, as measured by this new variable?

```r
#data sorteed by plastic_per_event in descending order
worst_polluted_countries <- country_totals[order(country_totals$plastic_per_event, decreasing = TRUE), ]

#top 5 countries with the worst plastic pollution
top_5_worst <- head(worst_polluted_countries, 5)
top_5_worst
```

```
##     X      country empty hdpe ldpe     o  pet  pp  ps pvc total num_events
## 49 49         Togo     0    0    0 11233  759   0   2   0 11994          5
## 8   8 Burkina Faso     0  267   20  3623 4664  68  90  42  8774         15
## 17 17      Germany    10    2  210  4790  102  95   7   0  5216         19
## 15 15  El Salvador     0    2  180    53  185 318 182   0   920          5
## 18 18        Ghana     0  220  291     0 8272   0   0   0  8783         52
##    volunteers plastic_per_event
## 49        460         2398.8000
## 8        2445          584.9333
## 17        456          274.5263
## 15         50          184.0000
## 18       8060          168.9038
```

Answer: This time around, the five countries with the worst plastic pollution (from our new variable) from least to most are Ghana, El Savador, Germany, Burkina Faso and Togo.

Now we will make a plot of the variation in the types of trash and how it differs around the world. Read in the continenttypes.csv data frame. This gives the breakdown of the different types of plastic collected on each continet in 2020 and the total number of pick up events.

```r
#read in the continent types.csv file
continent_types <- read.csv("continenttypes.csv", stringsAsFactors = FALSE)
head(continent_types)
```

```
##   X continent plastic_type total events
## 1 1    Africa        empty   335   6559
## 2 2    Africa         hdpe  8718   6559
## 3 3    Africa         ldpe 16460   6559
## 4 4    Africa            o 26764   6559
## 5 5    Africa          pet 54696   6559
## 6 6    Africa           pp  3292   6559
```

```r
str(continent_types)
```

```
## 'data.frame':    48 obs. of  5 variables:
##  $ X           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ continent   : chr  "Africa" "Africa" "Africa" "Africa" ...
##  $ plastic_type: chr  "empty" "hdpe" "ldpe" "o" ...
##  $ total       : int  335 8718 16460 26764 54696 3292 2158 73 693 2969 ...
##  $ events      : int  6559 6559 6559 6559 6559 6559 6559 6559 45174 45174 ...
```

```r
colnames(continent_types)
```

```
## [1] "X"            "continent"    "plastic_type" "total"        "events"
```

6. Add a column to this data frame with a variable that captures the existence of different types of plastic trash, controlling for the intensity of the pick-up effort in different continent

```
#adds column for total plastic collected per event
continent_types$plastic_per_event <- continent_types$total / continent_types$events
head(continent_types)
```

```
##   X continent plastic_type total events plastic_per_event
## 1 1    Africa        empty   335   6559        0.05107486
## 2 2    Africa         hdpe  8718   6559        1.32916603
## 3 3    Africa         ldpe 16460   6559        2.50952889
## 4 4    Africa            o 26764   6559        4.08050008
## 5 5    Africa          pet 54696   6559        8.33907608
## 6 6    Africa           pp  3292   6559        0.50190578
```

7. Make a plot using ggplot showing both the total amount and distribution of types of plastic picked up in each continent in the average pick-up event.

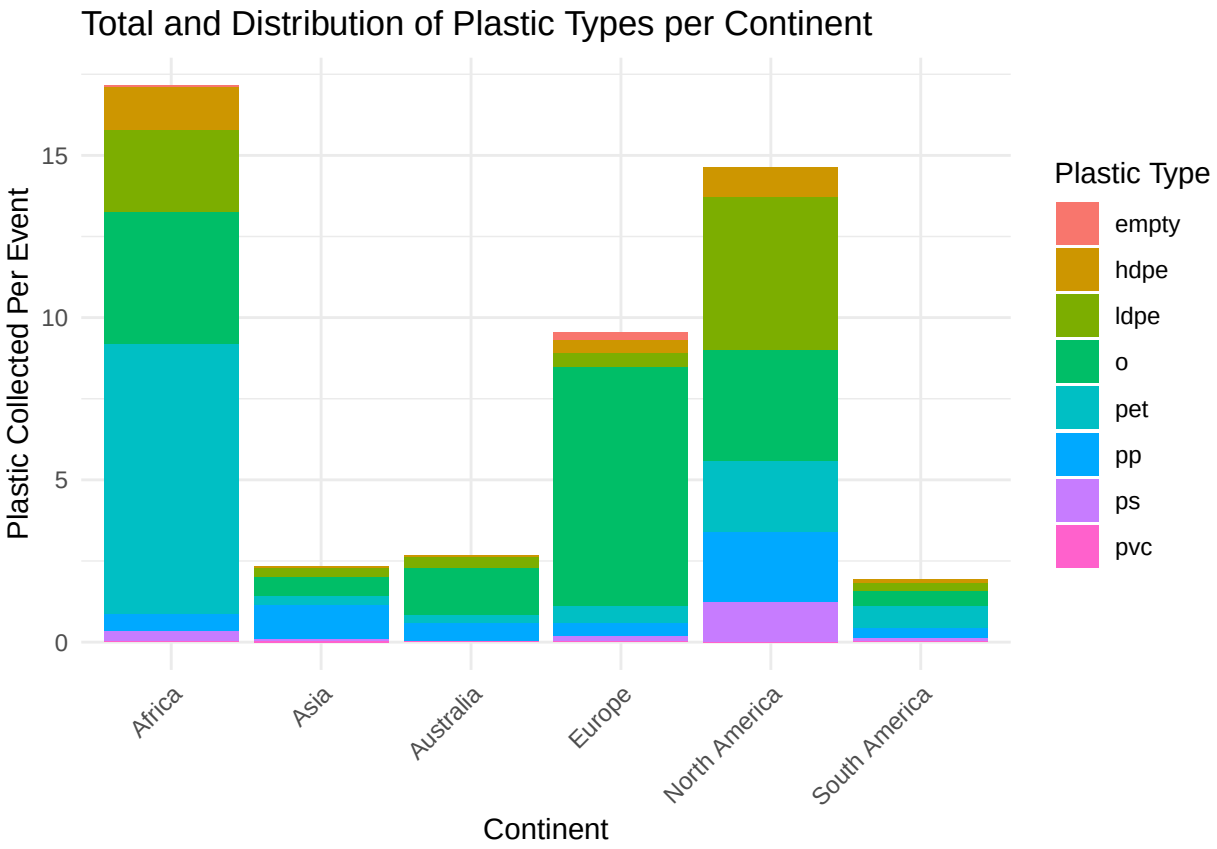
Hint: Check out options in the R graph gallery

```
library(ggplot2)
#stcakced bar plot
ggplot(continent_types, aes(x = continent, y = plastic_per_event, fill = plastic_type)) +
  geom_bar(stat = "identity", position = "stack") +

  scale_y_continuous(labels = scales::comma) +  #adding commas for large values
  labs(title = "Total and Distribution of Plastic Types per Continent",


       x = "Continent",
       y = "Plastic Collected Per Event",

       fill = "Plastic Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  #online result of the code to rotate x-axis
```



8. Try uploading your R markdown file and plots to your Git Hub repository. Upload your Rmd and knitted PDF to Canvas