

Lab 5

Kyra Liu

Lab 5

This lab we will look at some data from the plastic trash picked 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

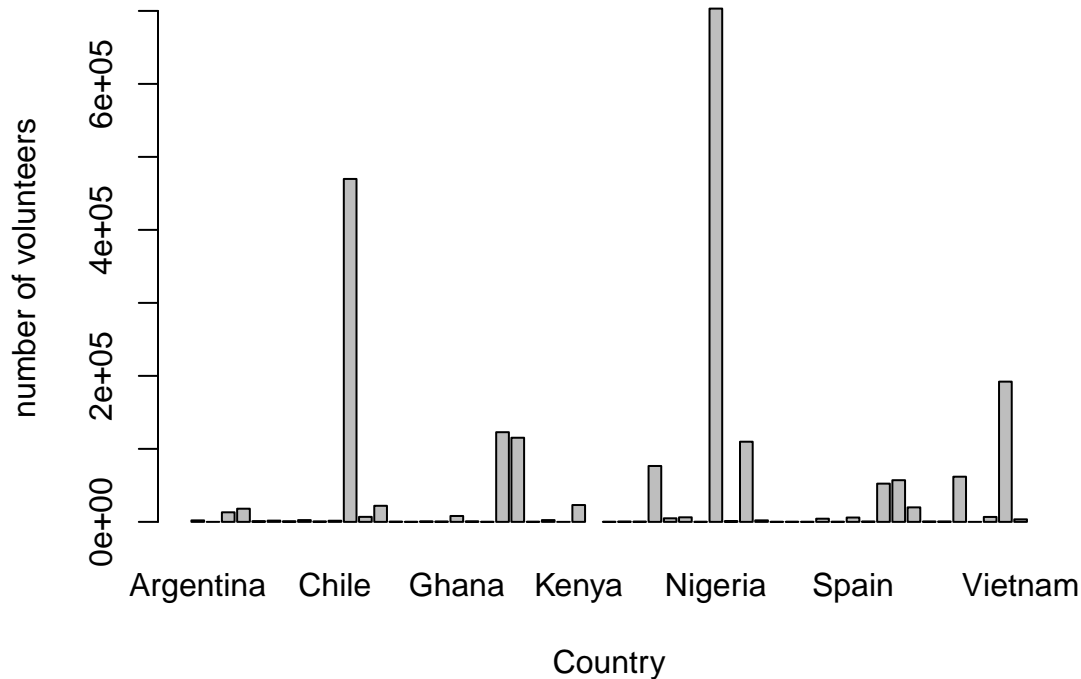
Have a look at the data frame. The column “total” gives the total number of pieces of plastic picked up in that country in 2020. The columns “num_events” and “volunteers” give the number of trash pick-up events and the number of volunteers in that country. We are going to use this to investigate where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

```
## [1] "Nigeria"      "Philippines" "Switzerland" "India"        "Togo"
```

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

Distribution of Volunteers Across Countries



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?

In countries with more volunteers, it is more likely for trash to have been picked up than in countries with less volunteers. A better representation may be to normalize the data as plastic that is picked up by each volunteer

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

```
## [1] "Nigeria"      "Philippines" "Switzerland" "India"        "Togo"
```

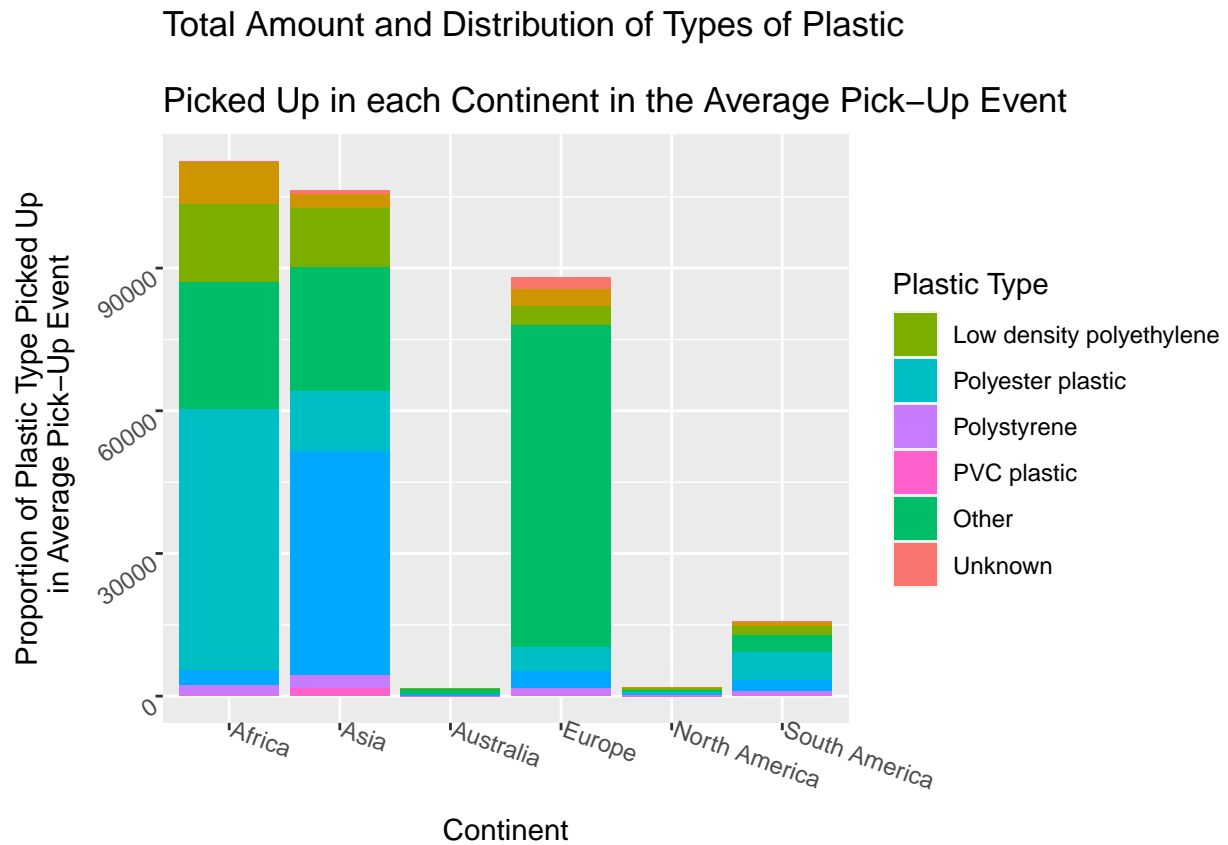
Now we will make a plot of the variation in the types of trash and how it differs around the world. Read in the `continenttypes.csv` data frame. This gives the breakdown of the different types of plastic collected on each continent in 2020 and the total number of pick up events.

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

For each type of plastic, what proportion of the pick-up efforts went towards picking up that type of plastic trash in each continent.

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



8. Try uploading your R markdown file and plot to your Git Hub repository. Don't put your knitted HTML file in Github - these are large files that are not designed to be stored on Github