Lab 5

```
## 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

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.

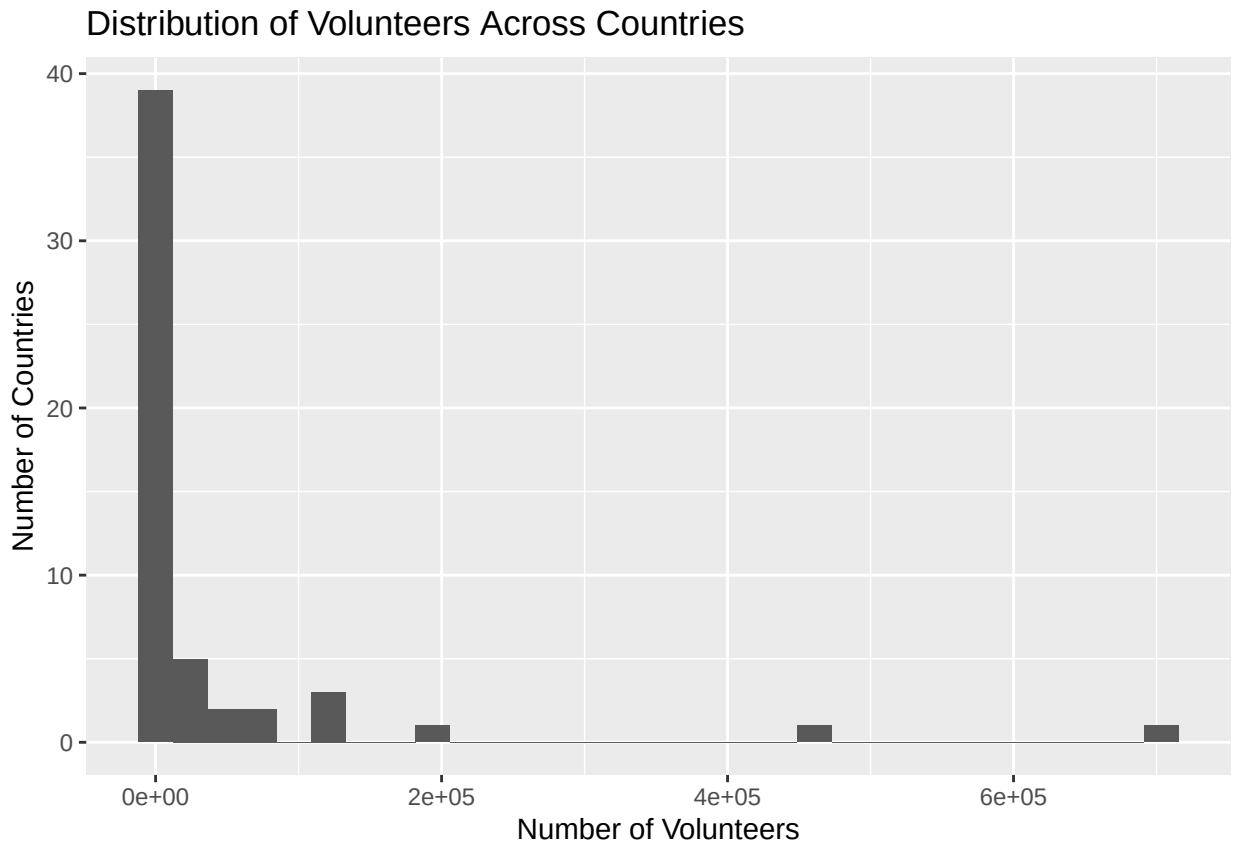
1. What 5 countries had the worst plastic problem as measured by the number of pieces of trash picked up?

```
##        country total
## 35     Nigeria 63253
## 37 Philippines 55184
## 46 Switzerland 52277
## 21       India 16973
## 49        Togo 11994
```

Answer: Nigeria, Philippines, Switzerland, India, Togo

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

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```



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: There may be variations across the number of volunteers in countries involved in trash pickup not because there may be more trash, but there might be more volunteers (countries with higher populations) within these countries that help with trash pick-up. Hence, this might be a problem for interpretation to question 1, since people will most likely assume that more trash picked up could equate to more trash present overall.

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?

```
##         country trash_per_event
## 49         Togo       2398.8000
## 8  Burkina Faso        584.9333
## 17      Germany        274.5263
## 15  El Salvador        184.0000
## 18        Ghana        168.9038
```

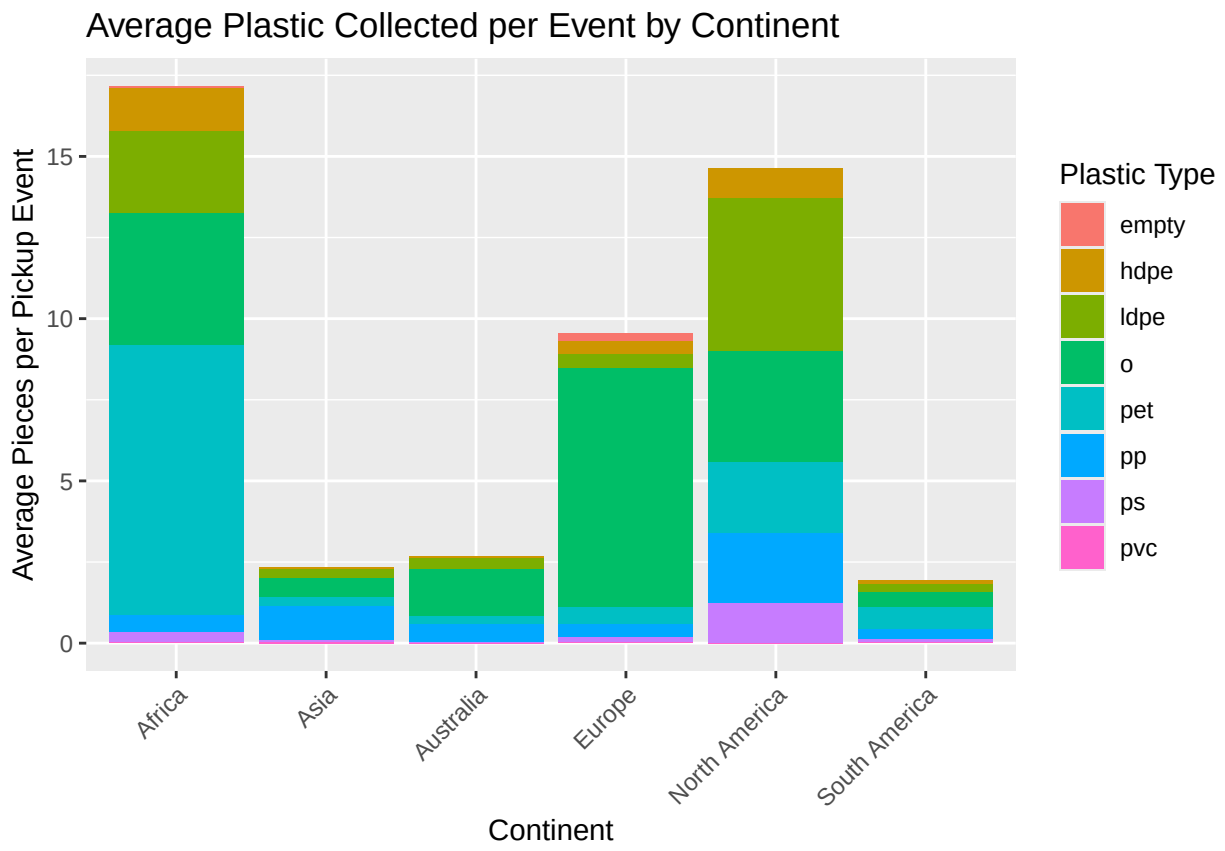
Answer: Togo, Burkina Faso, Germany, El Salvador, Ghana

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.

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

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 plots to your Git Hub repository. Upload your Rmd and knitted PDF to Canvas