

# Project 2: Rise and Fall of Programming Languages

Ankit Walimbe

2023-09-29

## Project 2

Note : This project is the analysis of the *‘Rise and Fall of Programming Languages’* The data was downloaded from Stack Exchange (<https://app.datacamp.com/workspace/external-link?url=https%3A%2F%2Fdata.stackexchange.com%2F>)

### Loading the necessary packages and setting up the environment

```
library(readr)
library(dplyr)
```

### Loading the Dataset

Here we import the data in R and assign it to a variable ‘by\_tag\_year’

```
by_tag_year <- read_csv("by_tag_year.csv")
```

### Inspecting the data

Making sure the data is imported correctly and inspecting it for errors

```
head(by_tag_year)
```

```
## # A tibble: 6 x 4
##   year tag          number year_total
##   <dbl> <chr>         <dbl>     <dbl>
## 1  2008 .htaccess         54     58390
## 2  2008 .net           5910     58390
## 3  2008 .net-2.0          289     58390
## 4  2008 .net-3.5          319     58390
## 5  2008 .net-4.0           6     58390
## 6  2008 .net-assembly      3     58390
```

## Adding a new column

### Data Description

This dataset presents tag-year pairs with counts of questions in a tag for that year and the total questions asked in that year. We've also included the percentage of questions for each tag within its respective year to provide context.

```
by_tag_year_fraction <- by_tag_year %>%  
  mutate (fraction = number/year_total)
```

### Printing the new table

```
print(by_tag_year_fraction)
```

```
## # A tibble: 40,518 x 5  
##   year tag          number year_total fraction  
##   <dbl> <chr>         <dbl>     <dbl>     <dbl>  
## 1 2008 .htaccess      54      58390 0.000925  
## 2 2008 .net          5910     58390 0.101  
## 3 2008 .net-2.0       289     58390 0.00495  
## 4 2008 .net-3.5       319     58390 0.00546  
## 5 2008 .net-4.0        6     58390 0.000103  
## 6 2008 .net-assembly   3     58390 0.0000514  
## 7 2008 .net-core       1     58390 0.0000171  
## 8 2008 2d             42     58390 0.000719  
## 9 2008 32-bit         19     58390 0.000325  
## 10 2008 32bit-64bit    4     58390 0.0000685  
## # i 40,508 more rows
```

### Checking if R has been growing or shrinking

Here we investigate if the use of R has been growing or are users switching to other programming languages

```
r_over_time <- by_tag_year %>%  
  mutate (fraction = number/year_total) %>%  
  filter(tag == "r")
```

First we filter out tags that are associated to R

```
print(r_over_time)
```

Printing the new table 'r\_over\_time'

```
## # A tibble: 11 x 5
##   year tag   number year_total fraction
##   <dbl> <chr> <dbl>      <dbl>    <dbl>
## 1  2008 r         8      58390 0.000137
## 2  2009 r        524     343868 0.00152
## 3  2010 r       2270     694391 0.00327
## 4  2011 r       5845    1200551 0.00487
## 5  2012 r      12221    1645404 0.00743
## 6  2013 r     22329    2060473 0.0108
## 7  2014 r     31011    2164701 0.0143
## 8  2015 r     40844    2219527 0.0184
## 9  2016 r     44611    2226072 0.0200
## 10 2017 r     54415    2305207 0.0236
## 11 2018 r     28938    1085170 0.0267
```

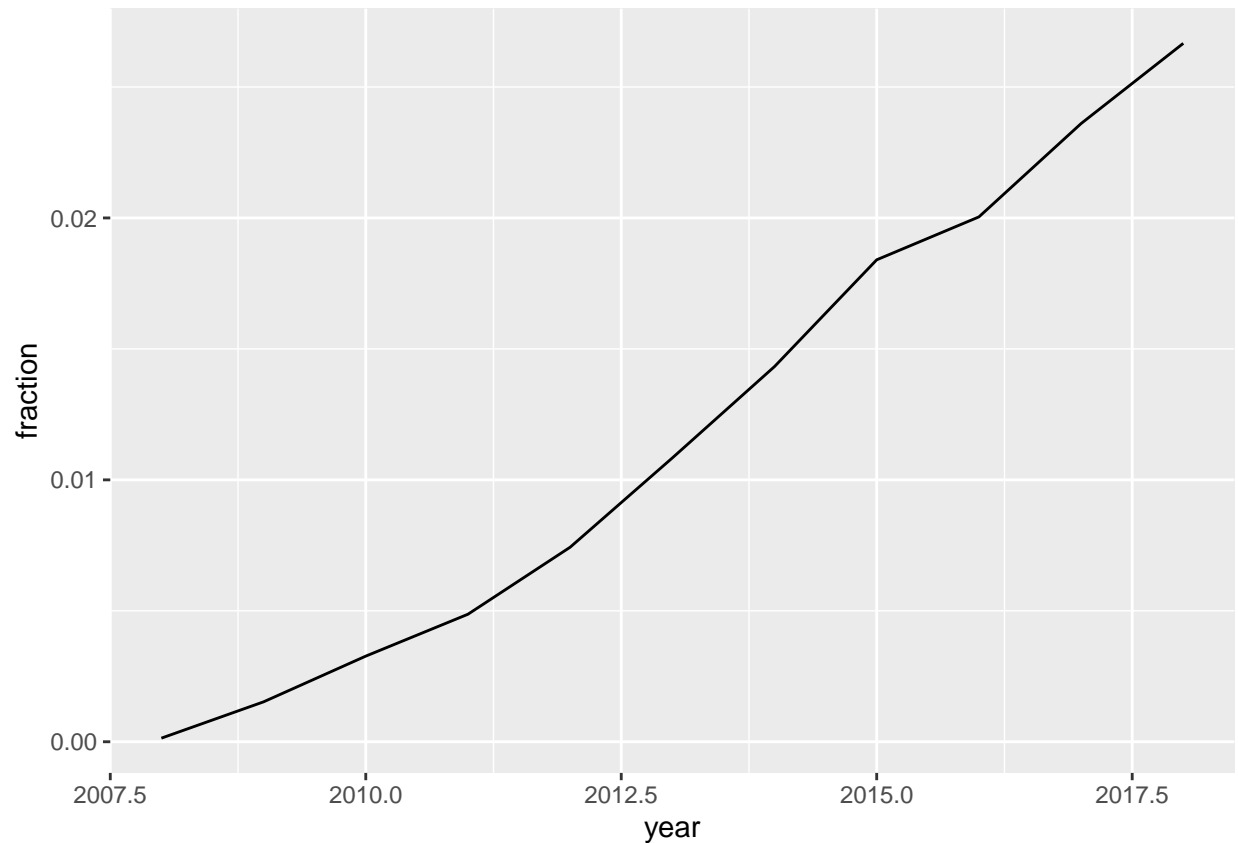
To visualize the data we load ‘ggplot2’

```
library(ggplot2)
```

## Visualizing the change over time

Here we we plot a line graph to visualize the change over time

```
ggplot(r_over_time,aes(x = year, y = fraction)) +
  geom_line()
```



## Checking the trend for ‘dplyr’ and ‘ggplot2’

### Analysis Objective

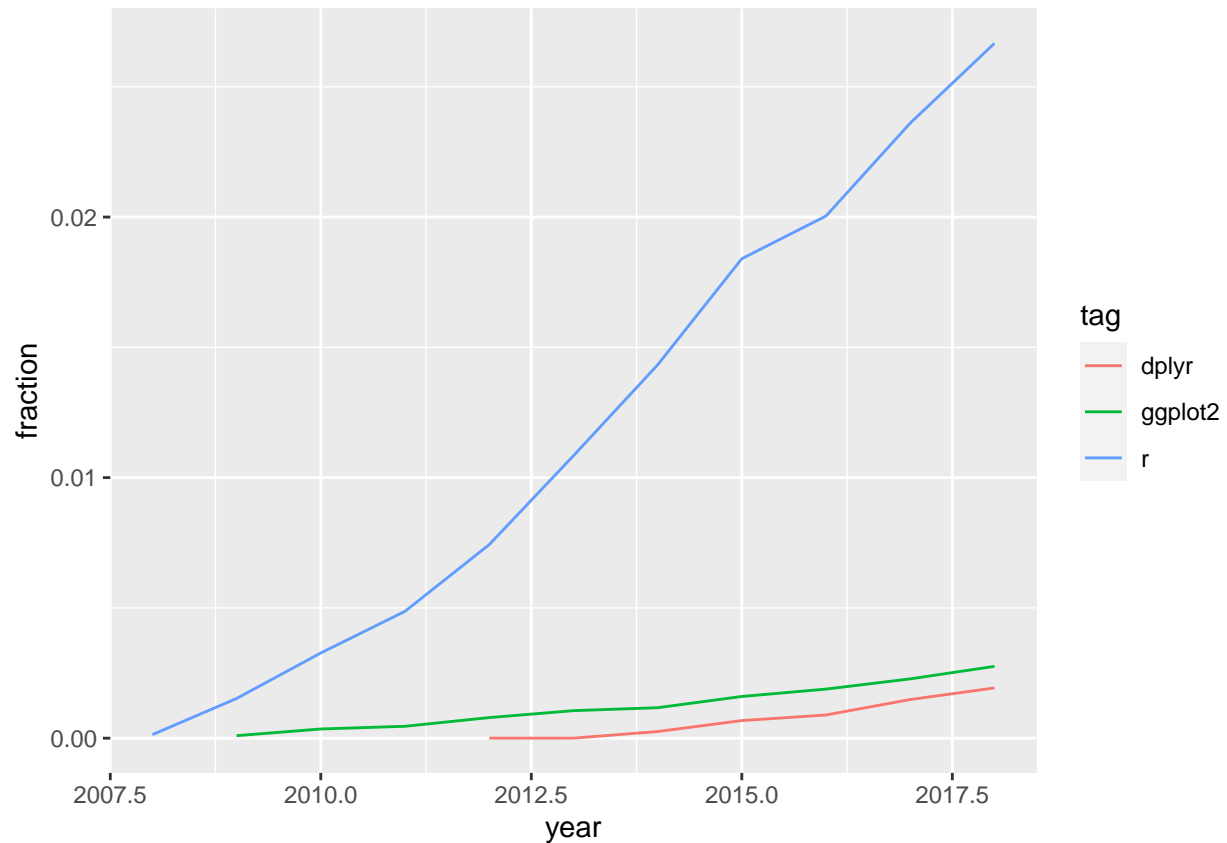
We’ll examine the tags **R**, **dplyr**, and **ggplot2** to see if their relative popularity is increasing or decreasing over time on Stack Overflow.

```
# A vector of selected tags
selected_tags <- c("r", "dplyr", "ggplot2")

# Filter for those tags
selected_tags_over_time <- by_tag_year_fraction %>%
  filter (tag %in% selected_tags)
```

### Plotting tags over time on a line plot, using color to represent tag

```
ggplot(selected_tags_over_time, aes(x = year, y = fraction, color = tag)) +
  geom_line()
```



Note : The 'dplyr' and 'ggplot2' tags do not have as many questions as 'R', but we can see a steady rise over time

### Investigating which tags have the most question overall

Here we look for tags that have the most questions, and find the total number of questions for each tag

```
sorted_tags <- by_tag_year %>%
  group_by(tag) %>%
  summarize(tag_total = sum(number)) %>%
  arrange(desc(tag_total))
```

```
print(sorted_tags)
```

### Printing the new table

```
## # A tibble: 4,080 x 2
##   tag      tag_total
##   <chr>      <dbl>
## 1 javascript 1632049
## 2 java       1425961
## 3 c#         1217450
```

```
## 4 php          1204291
## 5 android      1110261
## 6 python       970768
## 7 jquery       915159
## 8 html         755341
## 9 c++          574263
## 10 ios         566075
## # i 4,070 more rows
```

Analyzing the change in the large programming languages over time

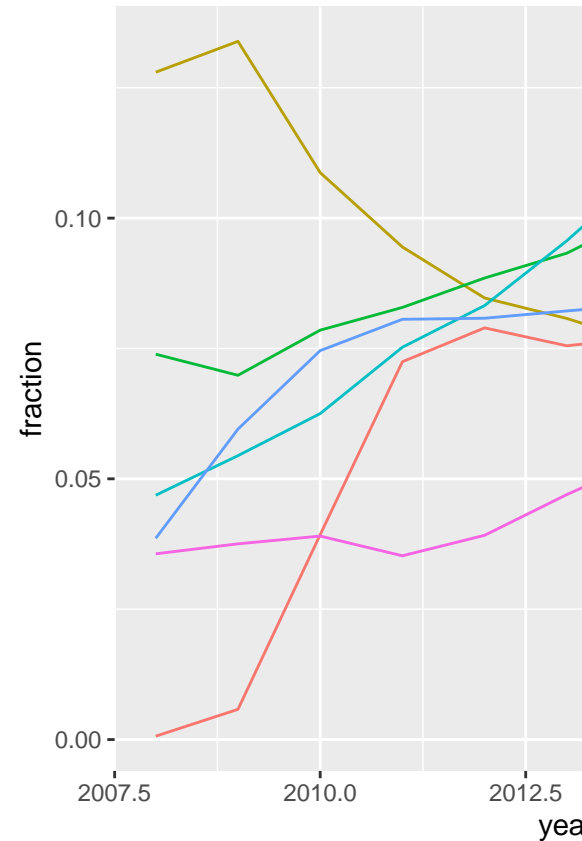
```
highest_tags <- head(sorted_tags$tag)
```

Getting the six largest tags

```
by_tag_subset <- by_tag_year_fraction %>%
  filter(tag %in% highest_tags)
```

Filtering the six largest tags

```
ggplot(by_tag_subset, aes (x = year, y = fraction, color = tag)) +
  geom_line()
```



Plotting tags over time on a line plot using color to represent tag

### Analysis Overview

The graph shows changes in how much people ask about programming languages. C# is getting fewer questions, while Python is getting more.

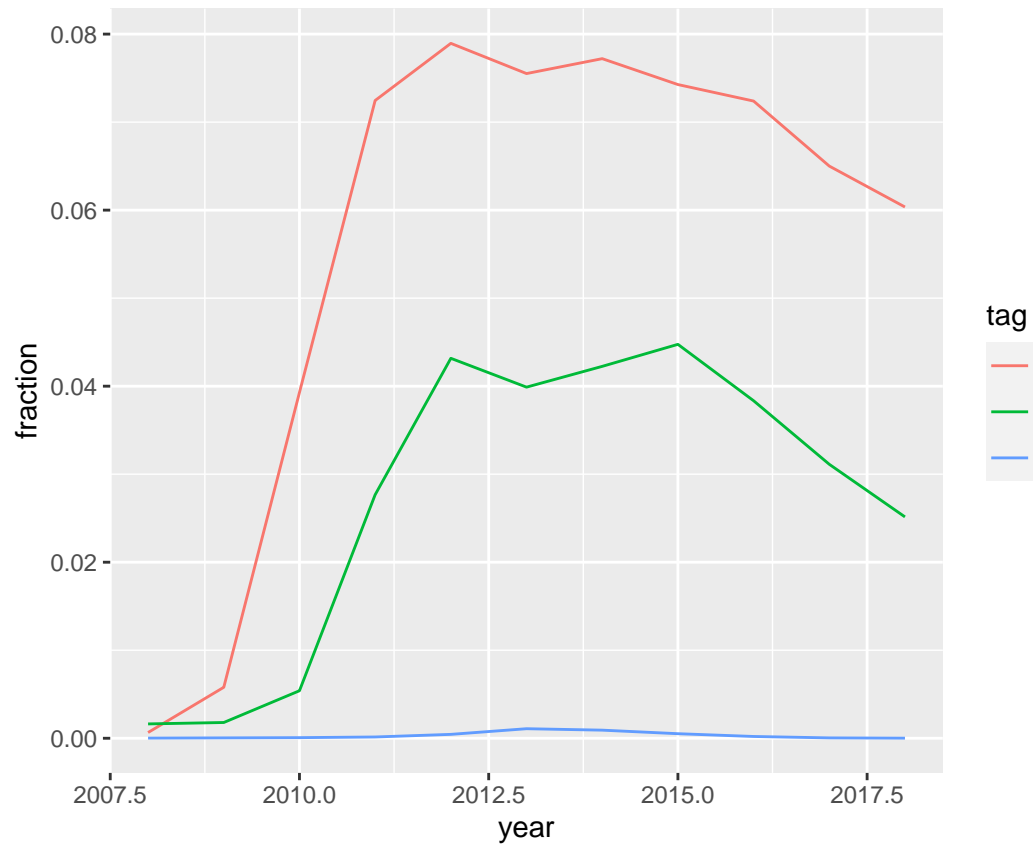
### Further analysing the data to get more insights

Using the data from Stack Overflow, we further use it to check the popularity of different mobile operating systems. Here we compare how many questions there are about ‘Android’, ‘iOS’, and ‘Windows’ Phone.

```
# Getting tags of interest
my_tags <- c("android", "ios", "windows-phone")

# Filter for those tags
by_tag_subset <- by_tag_year_fraction %>%
  filter(tag %in% my_tags)
```

```
ggplot(by_tag_subset, aes(x = year, y = fraction, color = tag)) +
  geom_line()
```



Plotting these tags over time