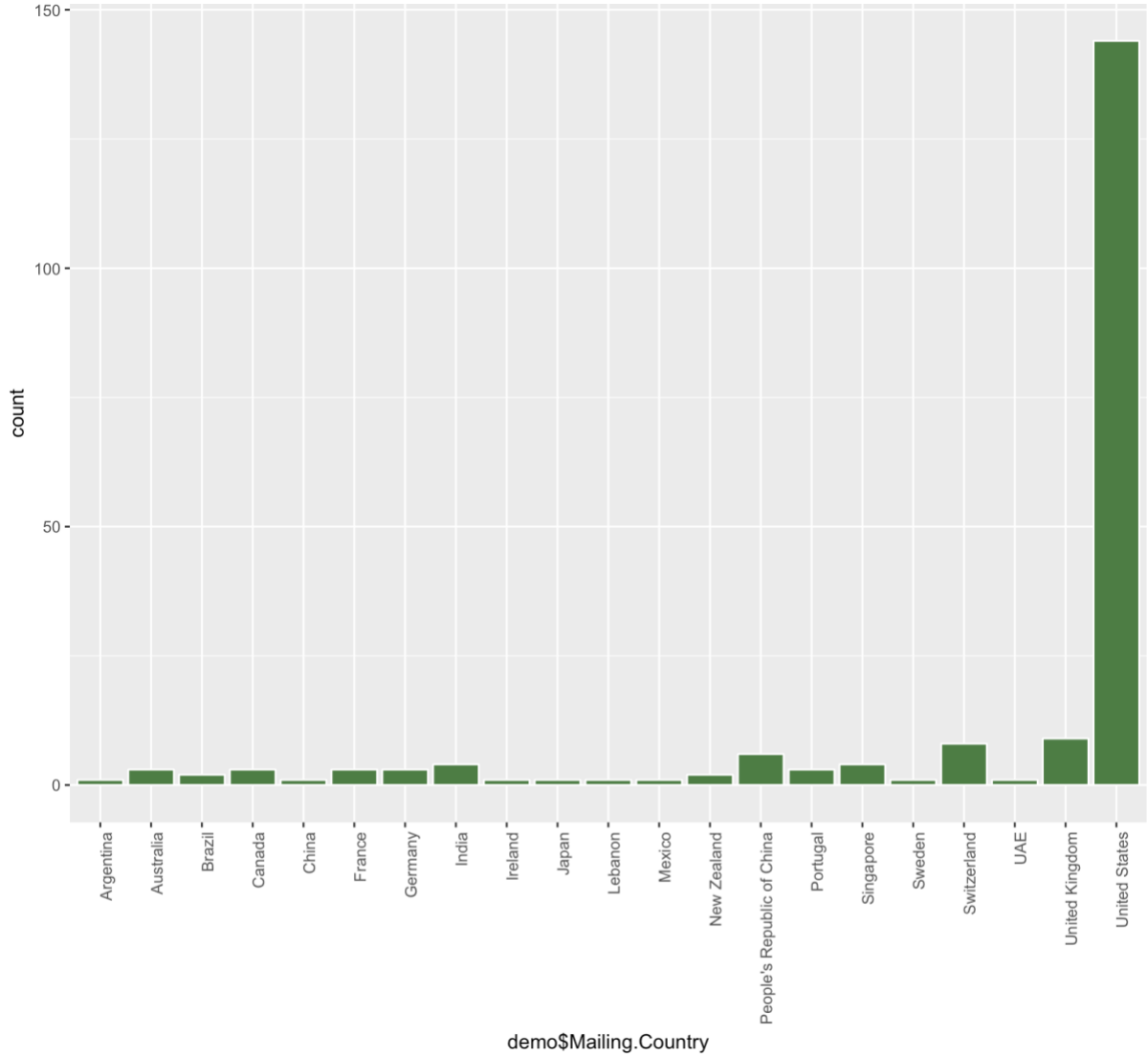
Columbia Business School, Graduate Research Assistant Task
Name-Subham Kedia, UNI-sk4355

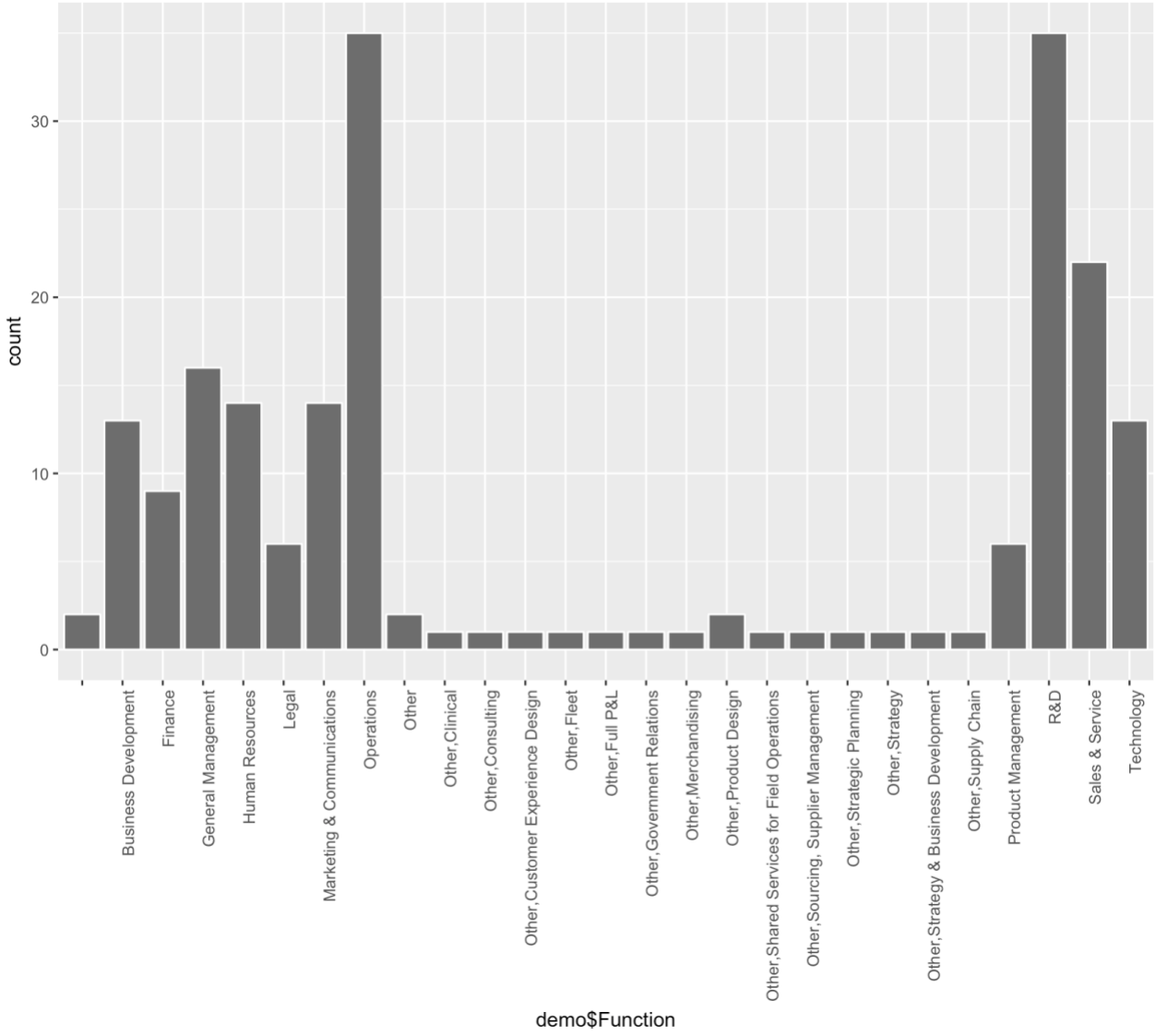
Plots and Graphs: -

Plot-1: The histogram below shows the number of users from each country



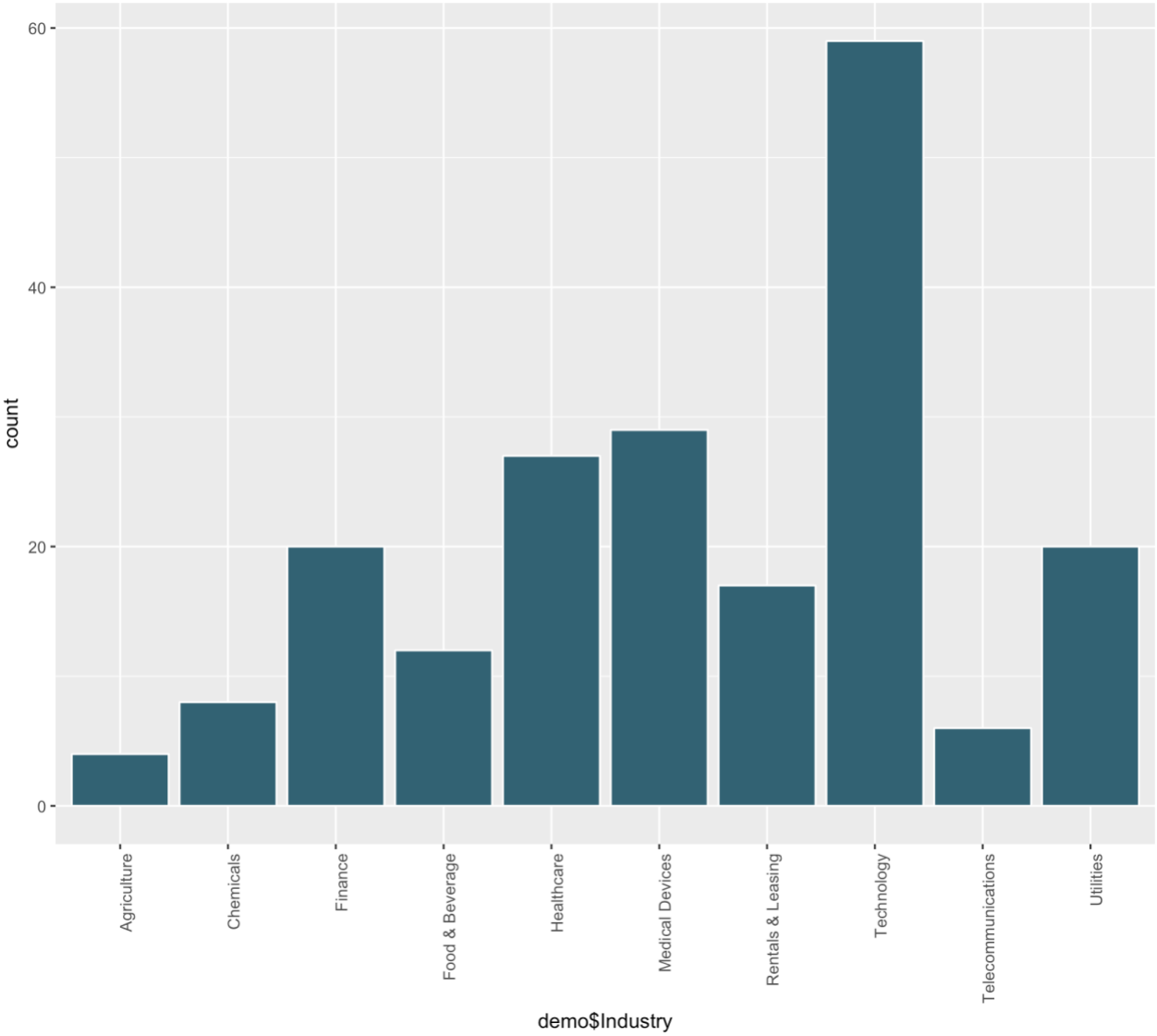
Inference-1: From the above plot, we can see that the maximum number of users come from United States, United Kingdom, Switzerland and People's Republic of China.

<u>Plot-2:</u> The histogram below shows the number of users from each function.



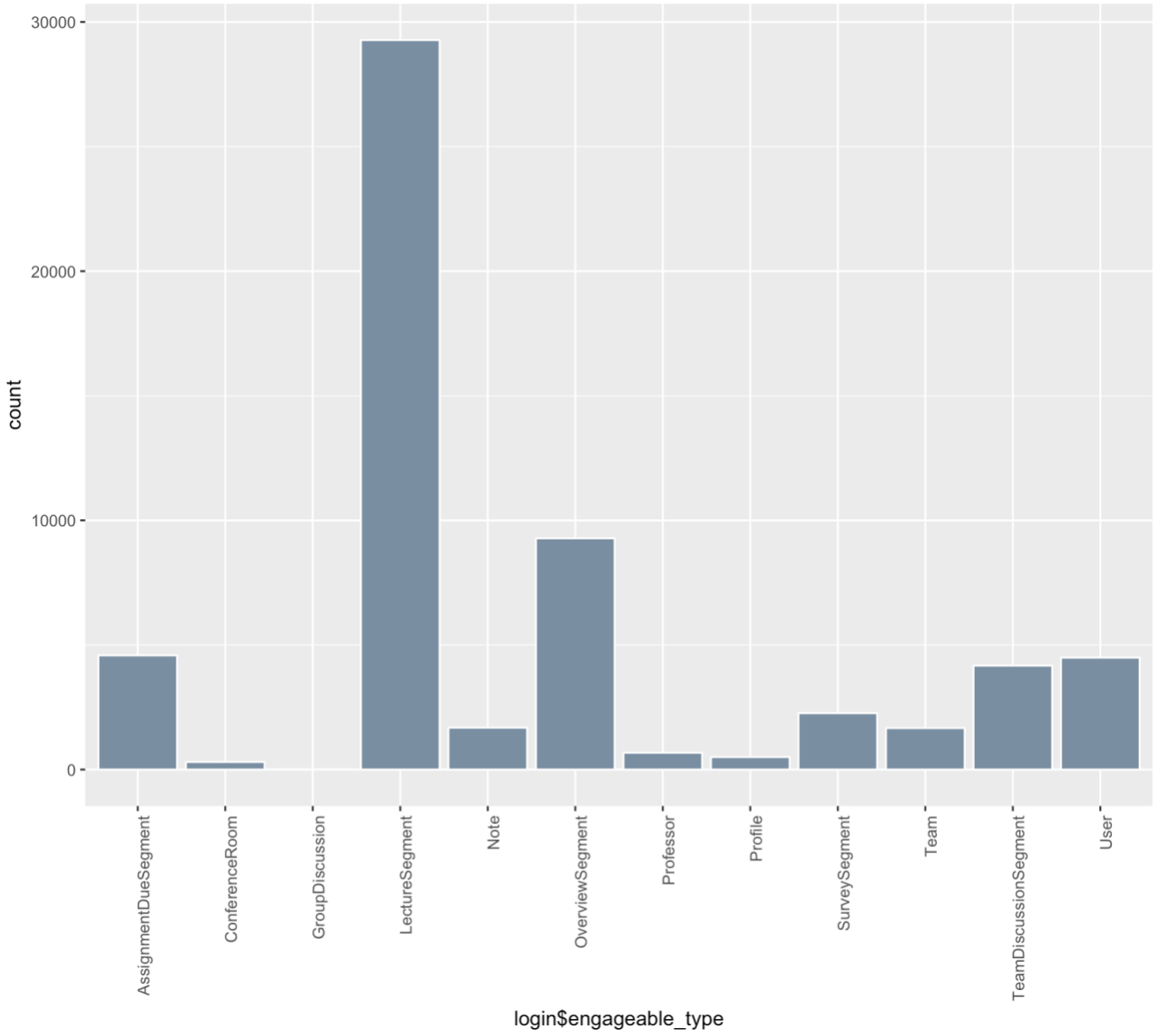
<u>Inference-2:</u> From the above plot, we can see that the maximum number of users come from Operations, R&D and Sales & Service functions.

<u>Plot-3:</u> The histogram below shows the number of users from each industry.



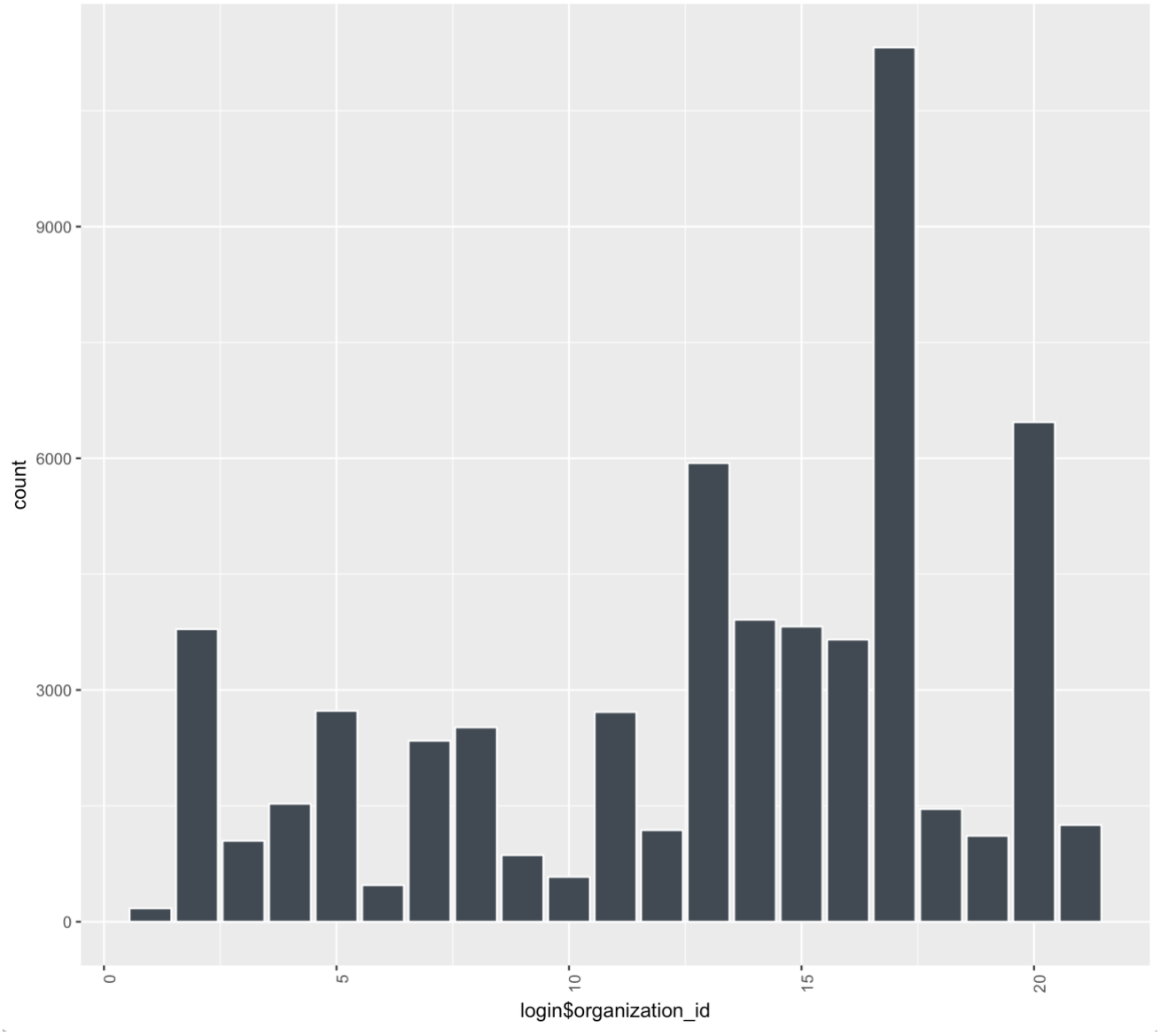
<u>Inference-3:</u> From the above plot, we can see that the maximum number of users come from the Technology, Healthcare and Medical Devices industry.

<u>Plot-4:</u> The histogram below shows the frequency of each engageable type in the login data.



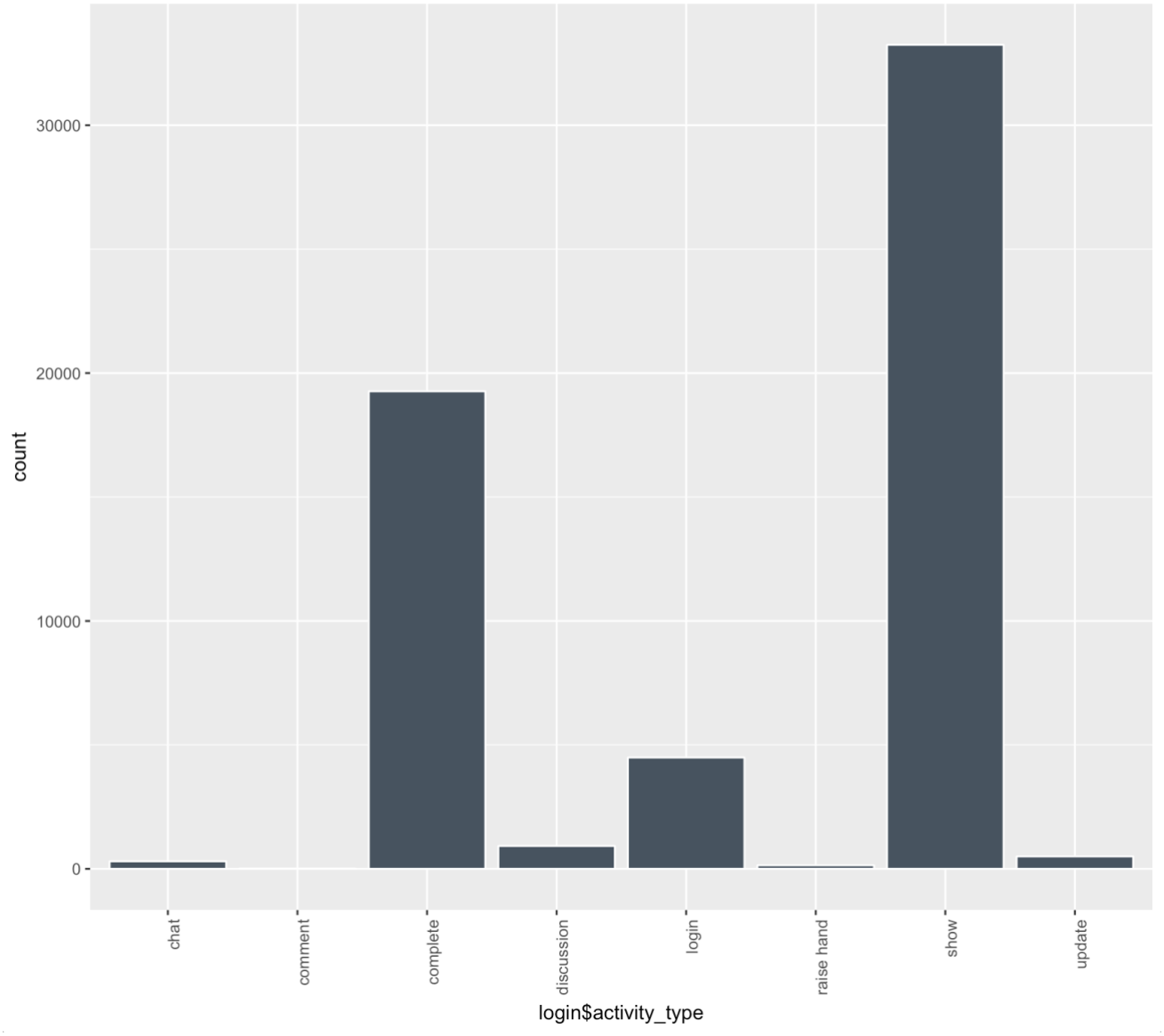
<u>Inference-4:</u> From the above plot, we can see that the users spent maximum time of Lecture Segment and Overview Segment.

<u>Plot-5:</u> The histogram below shows which the activity of each organization.



<u>Inference-5:</u> From the above plot, we can see that the maximum number of users come from organization 17, 13 and 20.

<u>Plot-6:</u> The histogram below shows the number of times each activity occurs in the login data.

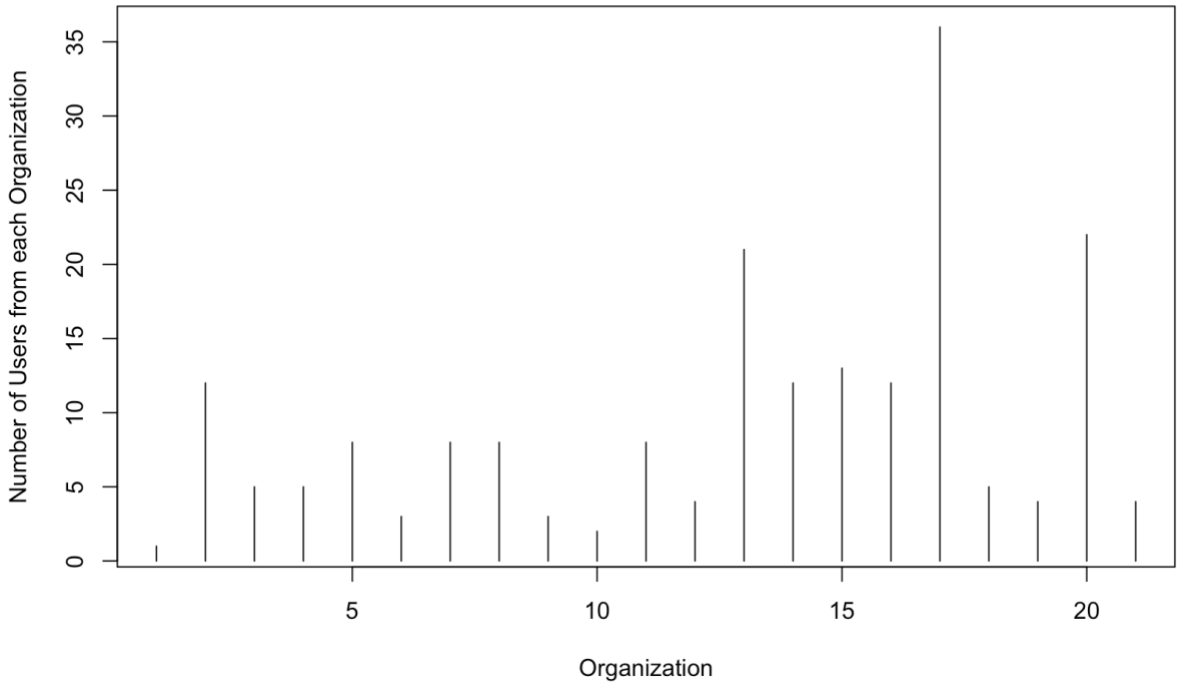


<u>Inference-6:</u> From the above plot, we can see that the activity type- complete and show appears most number of times in the login data set.

Matrix: **The matrix below tells us how many users are present from each organization. The first column is the organization number and the second column is the number of users.**

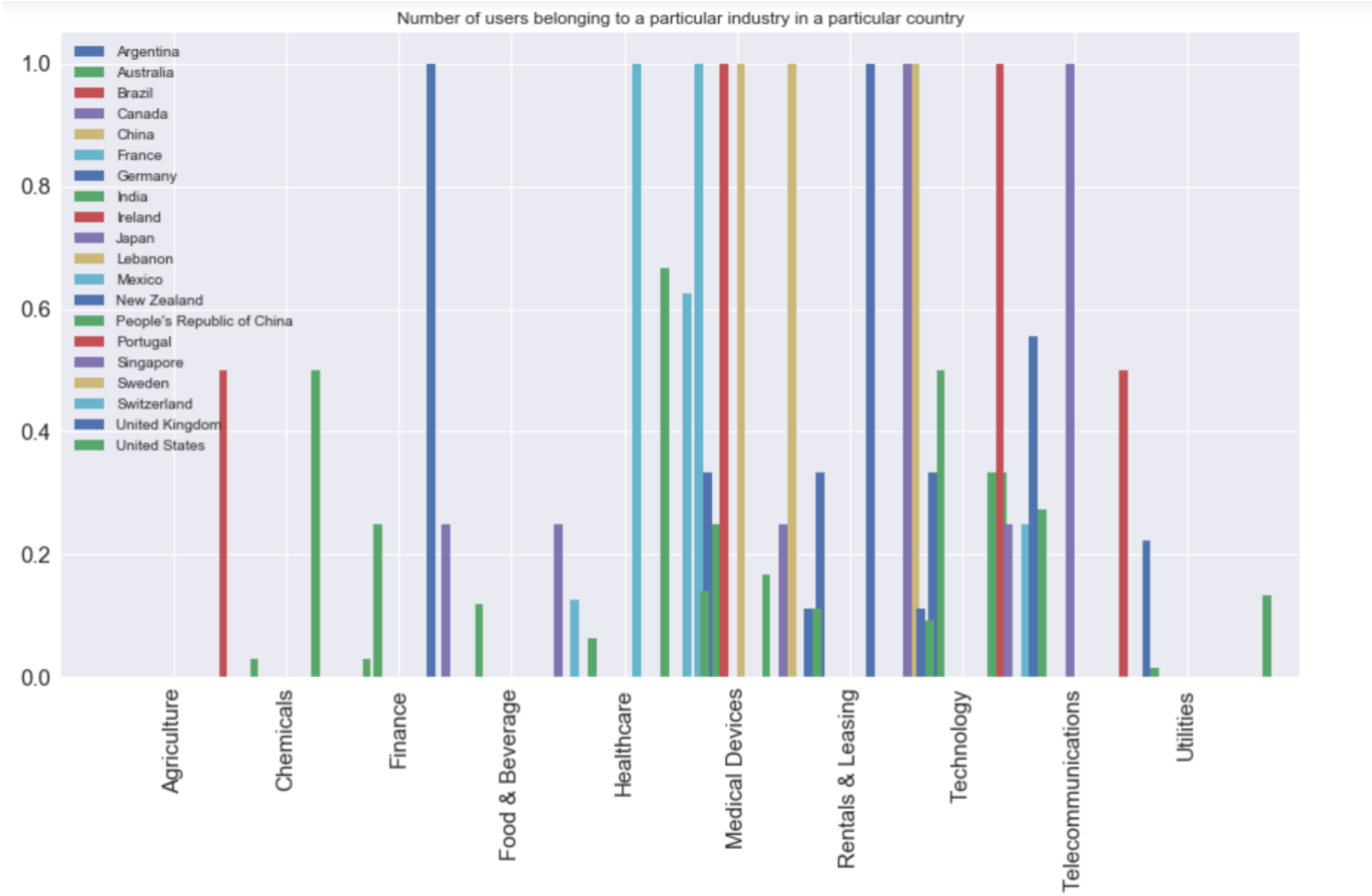
```
> matrix1
      [,1] [,2]
 [1,]    1    1
 [2,]    2   12
 [3,]    3    5
 [4,]    4    5
 [5,]    5    8
 [6,]    6    3
 [7,]    7    8
 [8,]    8    8
 [9,]    9    3
[10,]   10    2
[11,]   11    8
[12,]   12    4
[13,]   13   21
[14,]   14   12
[15,]   15   13
[16,]   16   12
[17,]   17   36
[18,]   18    5
[19,]   19    4
[20,]   20   22
[21,]   21    4
```

<u>Plot-7:</u> The histogram below shows the number of users present from each organization.



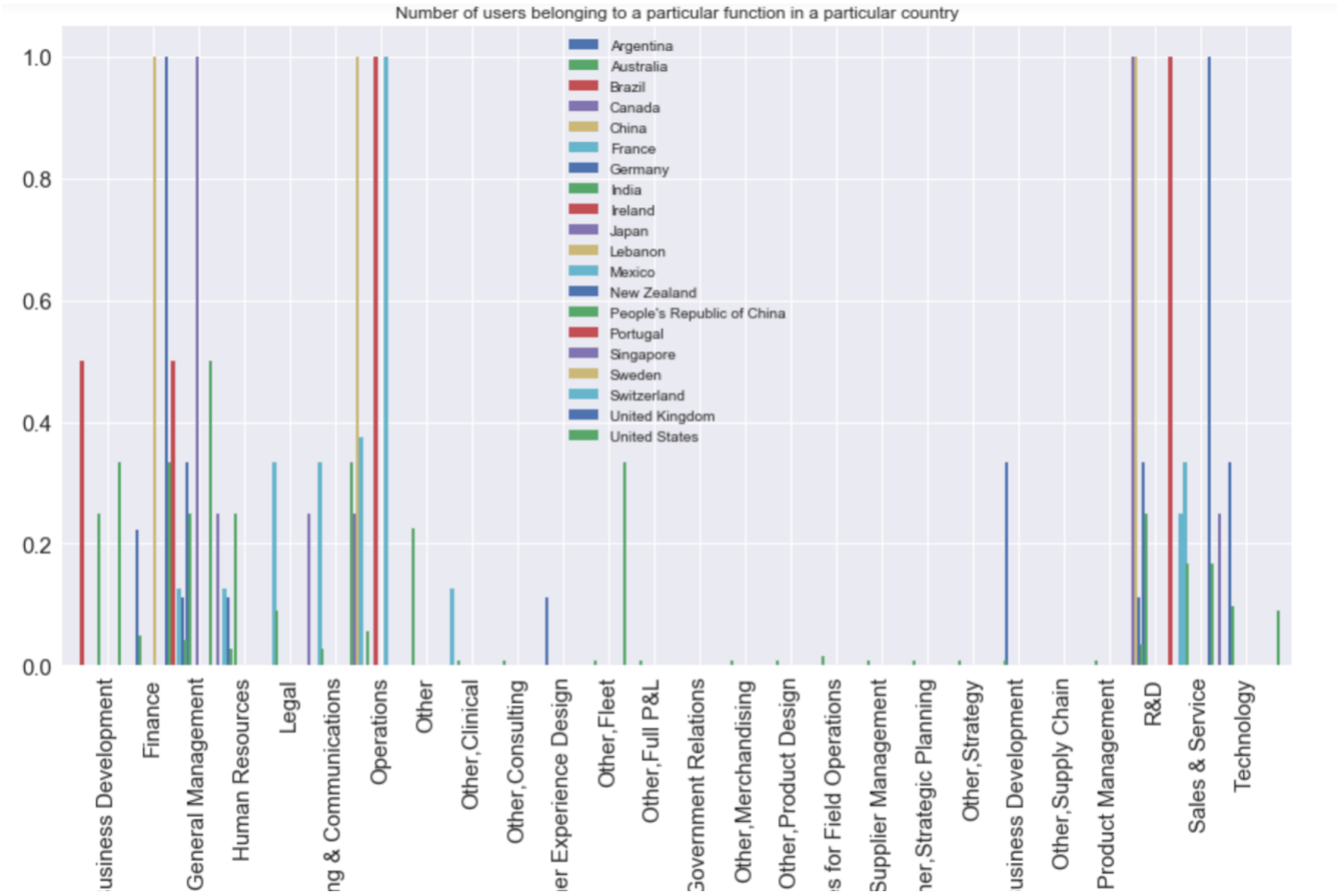
<u>Inference-7:</u> From the above plot, we can see that the maximum number of users come from 17, 13 and 20.

<u>Plot-8:</u> The histogram below shows the number of users belonging to a particular industry in a particular country.



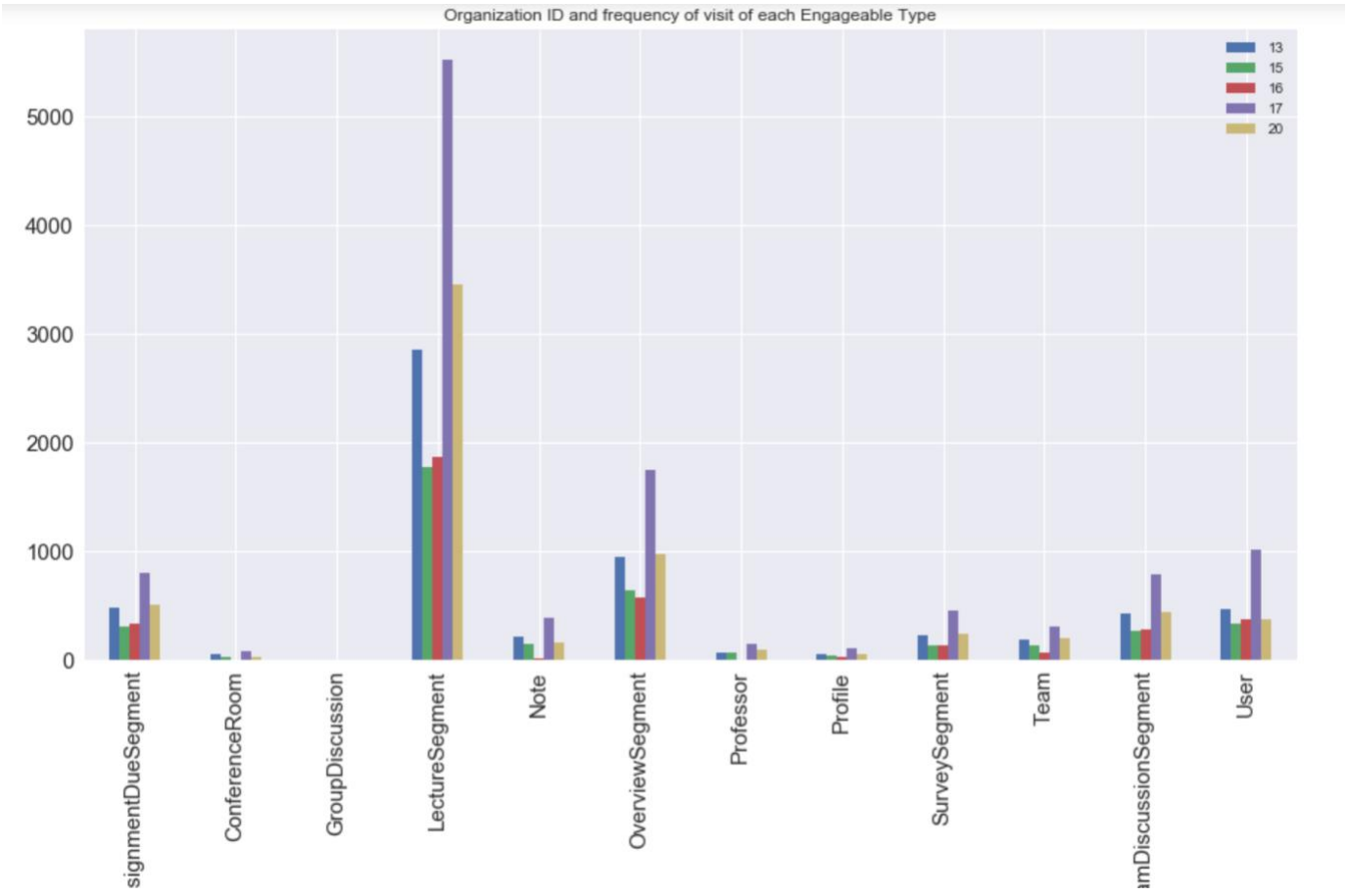
<u>Inference-8:</u> From the above plot, we can see that 1) U.K is heavily focused on finance, 2) United States is evenly focused on all the sectors, 3) China is focused on Healthcare, Medical Devices and Technology, 4) Japan is heavily focused on technology.

<u>Plot-9:</u> The histogram below shows the number of users belonging to a particular function in a particular country.



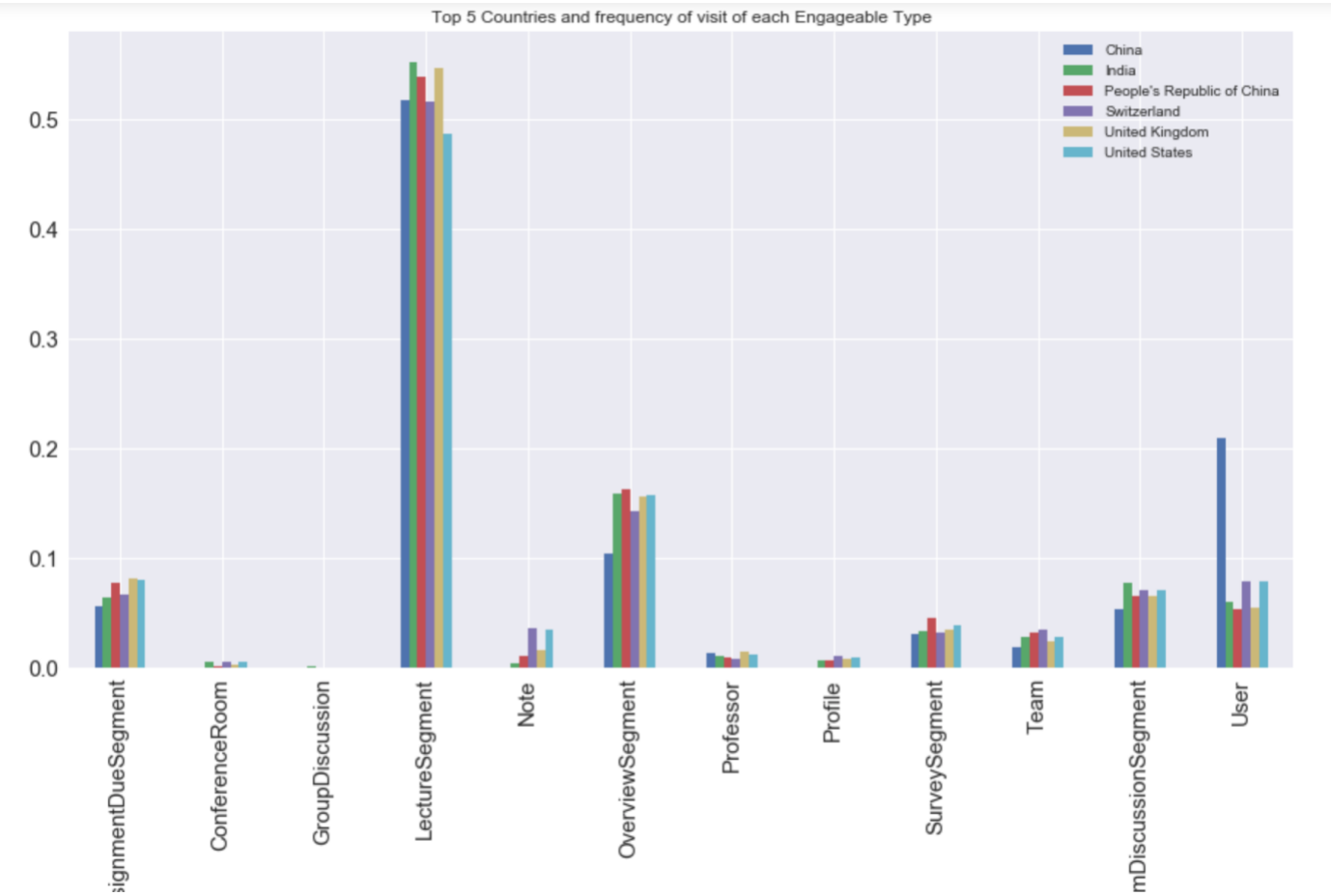
<u>Inference-9:</u> From the above plot, we can see that 1) Important functions in China are Finance, R&D 2) People from the United States come from all the functions.

<u>Plot-10:</u> The histogram below shows the frequency of visits of each engageable type, for the top 5 organizations contributing the maximum number of users.



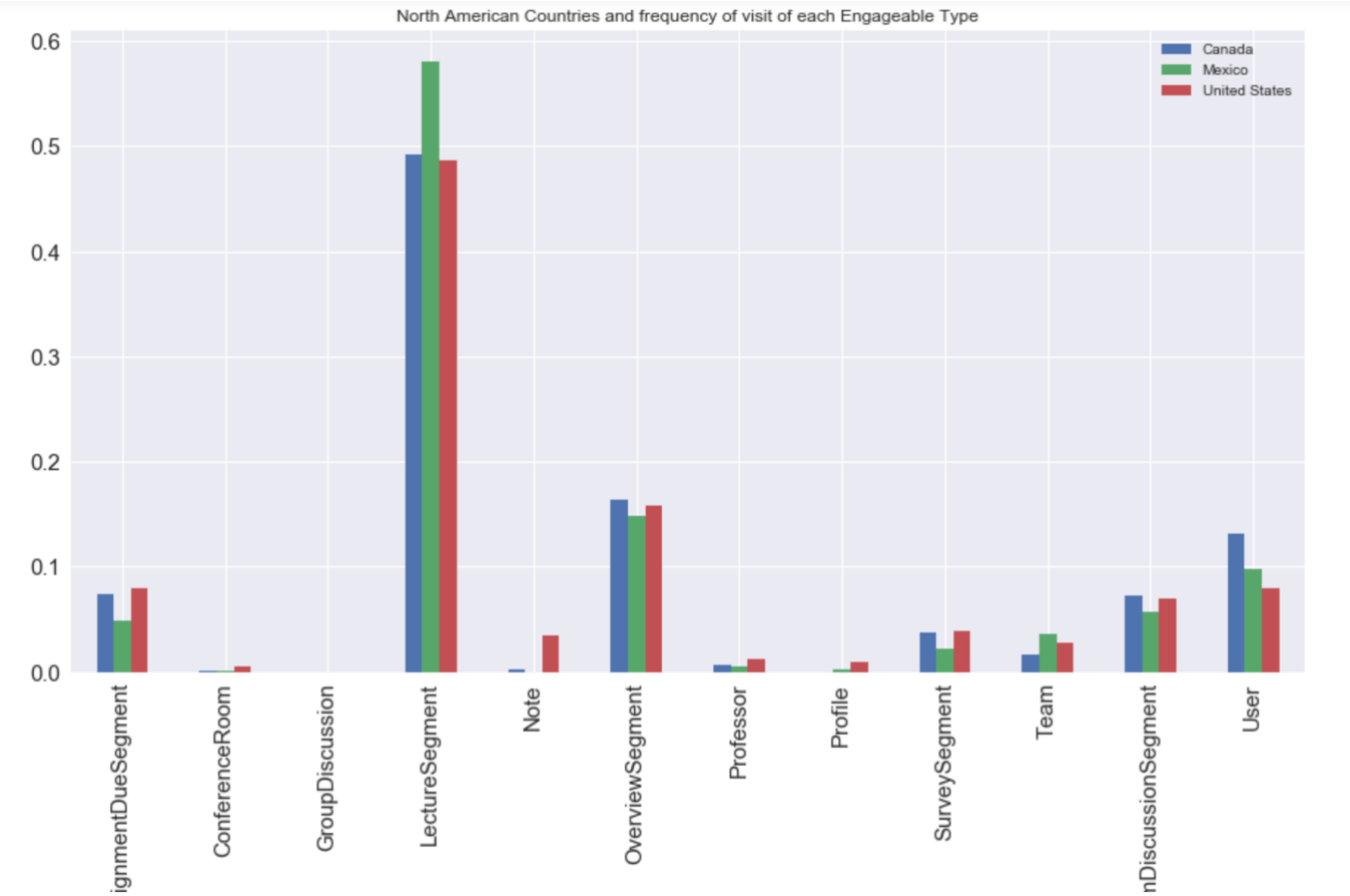
<u>Inference-10:</u> From the above plot, we can see that the users from the top 5 organizations spend maximum time on Lecture Segment and Overview Segment.

<u>Plot-11:</u> The histogram below shows the frequency of visits of each engageable type for the users belonging to different countries for the top countries which contribute maximum number of users.



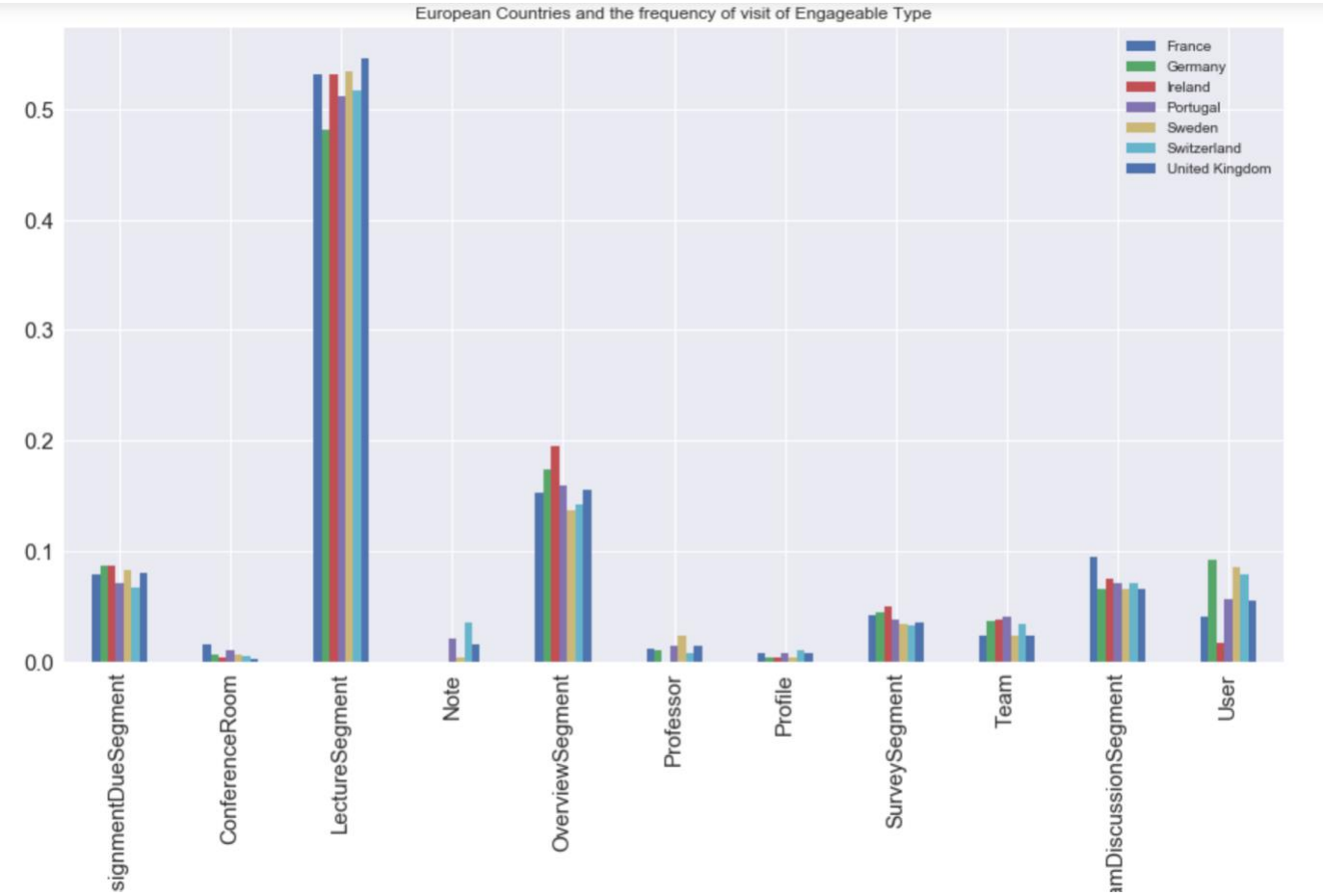
<u>Inference-11:</u> From the above plot, we can see that the users from the top 5 countries spend maximum time on Lecture Segment and Overview Segment. People from China also spend a significant amount of time on User.

<u>Plot-12:</u> The histogram below shows the frequency of visits of each engageable type for the users belonging to North American Countries.



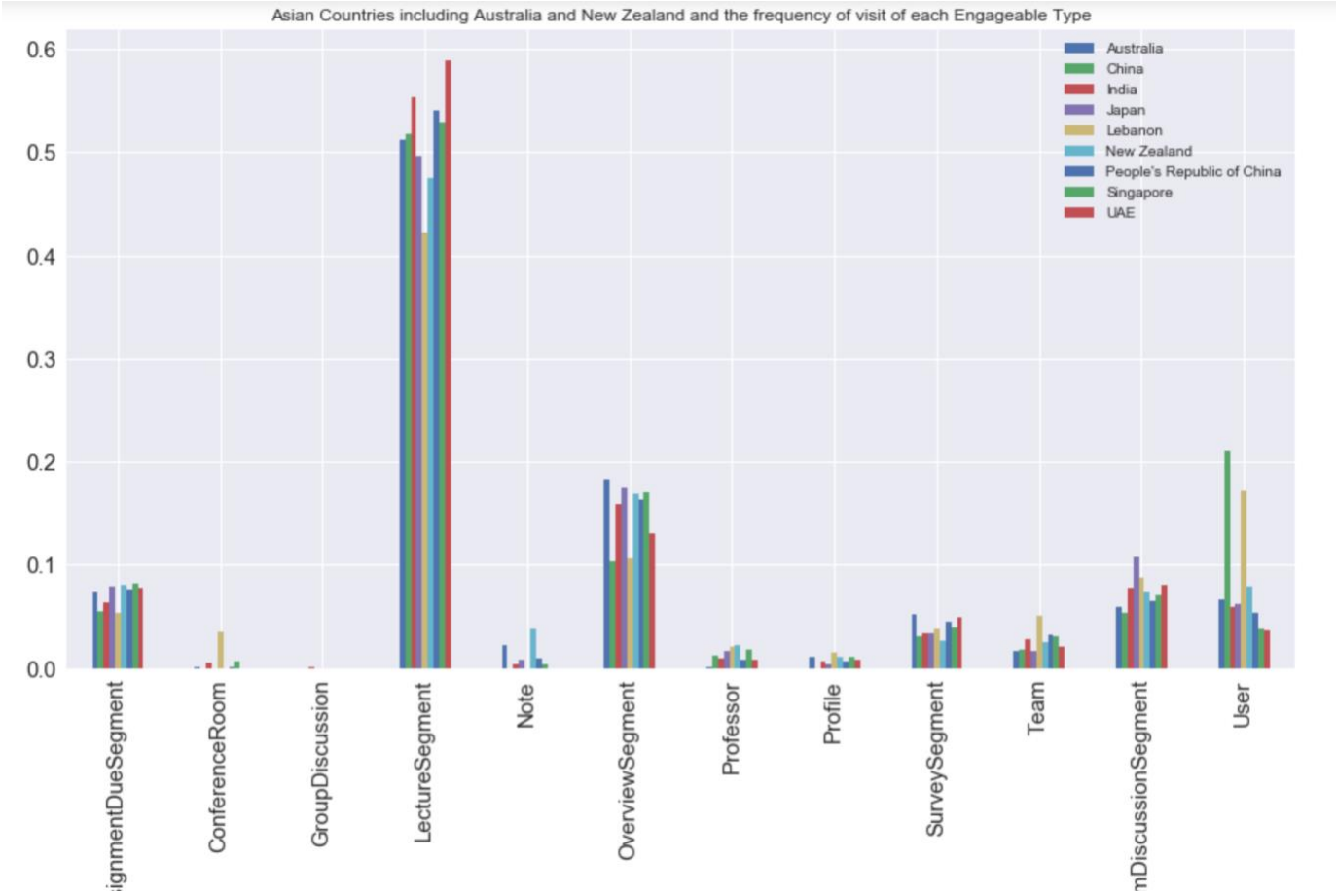
<u>Inference-12:</u> From the above plot, we can see that the users from the North American countries spend maximum time on Lecture Segment and Overview Segment. People from Canada spend a significant amount of time on User as well. People from Mexico spend their maximum time on Lecture Segment.

<u>Plot-13:</u> The histogram below shows the frequency of visits of each engageable type for the users belonging to European Countries.



<u>Inference-13:</u> From the above plot, we can see that the users from the European Countries spend maximum time on Lecture Segment and Overview Segment with Lecture Segment dominating significantly.

<u>Plot-14:</u> The histogram below shows the frequency of visits of each engageable type for the users belonging to Asian Countries, Australia and New Zealand.



<u>Inference-14:</u> From the above plot, we can see that the users from the Asian Countries spend maximum time on Lecture Segment and Overview Segment with Lecture Segment dominating significantly. People from India and UAE spend maximum time of Lecture Segment. People from China spend a considerable amount of China on User as evident before.

<u>Other Important Inferences:</u>

1) **Users from all the countries equally participate in the discussion but users from France participate more than other countries on average.**
2) **Users from European countries form better team than other users from other continents.**
3) **Users from European countries take the survey seriously.**

<u>Note-</u> All the above analysis has been done by coding in Python and R. Few things have been done in R and the others have been done in Python.