Internet Users per 100 (Top-5 among G20 countries)

Anand Kumar Kalukuri
School of Information Technology
Illinois State University akaluku@ilstu.edu

October 22, 2014

1 Introduction

Internet is a system of interconnected systems which connects billions of devices worldwide. These devices include computers,mobile phones and many more. Devices can exchange information between each other as long as they are connected through internet

Internet provides lot of services like e-mail, calling and browsing the web.It has played a major role in the development of lot of sectors of a country's economy.

Internet usage has increased drastically in almost all the countries in the world.As of 2013, Among the G-20 countries India has the lowest level of internet users by country per 100 with a value of 15.10 per 100 while United Kingdom has the highest with a value of 89.84.

2 Internet Usage Analysis

In this article we will be studying the internet usage for the top 5 among G-20 countries per 100 users. Here we will be extracting the data from Quandl. In Quandl we have the collected data for various countries which we can use it for our analysis.

```
> # Extracting data from Quandl
>
> library(Quandl)
> #loading germany data
> mydata.germany <- Quandl("WORLDBANK/DEU_IT_NET_USER_P2", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading korea data
> mydata.korea <- Quandl("WORLDBANK/KOR_IT_NET_USER_P2", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading canada data
```

```
> mydata.canada <- Quandl("WORLDBANK/CAN_IT_NET_USER_P2", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading japan data
> mydata.japan <- Quandl("WORLDBANK/JPN_IT_NET_USER_P2", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading Britain data
> mydata.britain <- Quandl("WORLDBANK/GBR_IT_NET_USER_P2", authcode="-U1g_98fUgj7RN8PGcaQ")
> #data for change of internet users per 100
> #loading data for germany
> changedata.germany<-Quandl("WORLDBANK/DEU_IT_NET_USER_P2",
+                            transformation="diff", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading data for britain
> changedata.britain<-Quandl("WORLDBANK/GBR_IT_NET_USER_P2",
+                            transformation="diff", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading data for korea
> changedata.korea<-Quandl("WORLDBANK/KOR_IT_NET_USER_P2",
+                           transformation="diff", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading data for japan
> changedata.japan<-Quandl("WORLDBANK/JPN_IT_NET_USER_P2",
+                           transformation="diff", authcode="-U1g_98fUgj7RN8PGcaQ")
> #loading data for canada
> changedata.canada<-Quandl("WORLDBANK/CAN_IT_NET_USER_P2",
+                            transformation="diff", authcode="-U1g_98fUgj7RN8PGcaQ")
```

In the above code chunk, I have extracted the dataset for each country from Quandl.The authcode is the unique authorization code assigned to each user.

2.1 Data Cleaning

```
> #adding the required columns into a single data frame
>
> mydata.germany$korea <- mydata.korea$Value
> mydata.germany$canada <- mydata.canada$Value
> mydata.germany$japan <- mydata.japan$Value
> mydata.germany$britain <- mydata.britain$Value
> changedata.germany$korea <- changedata.korea$Value
> changedata.germany$canada <- changedata.canada$Value
> changedata.germany$japan <- changedata.japan$Value
> changedata.germany$britain <- changedata.britain$Value
> head(mydata.germany,1)

        Date   Value korea canada japan britain
1 2013-12-31 83.9614 84.77   85.8 86.25 89.8441

> head(changedata.germany,1)

        Date    Value     korea canada japan  britain
1 2013-12-31 1.611402 0.6967735    2.8     0 2.364102
```

```
> #Changing the column names
>
> colnames(mydata.germany) <- c("Date","Germany","Korea","Canada","Japan","Britain")
> colnames(changedata.germany) <- c("Date","Germany","Korea","Canada","Japan","Britain")
>
```

In the above code chunk, different values from different datasets are moved into a single dataset. The values for South Korea,Canada,Japan and UK were imported into Germany's dataset.

```
> #selecting only the first row of the data
>
> mydata.germany <- mydata.germany[1,]
> changedata.germany <- changedata.germany[1,]
> #Using reshape to convert the data from wide format to long format
>
> library(reshape2)
> mydata.final <- melt(data=mydata.germany,id.vars=c("Date"))
> changedata.final <- melt(data=changedata.germany,id.vars=c("Date"))
> #writing the files to csv
> write.csv(mydata.final,"Internet_users_by_level.csv")
> write.csv(changedata.final,"Internet_users_by_change.csv")
```

The data that we got from Quandl and merged it into a dataframe had the data for all the years. We need to study the trends for just the latest year. So, we will be selecting just the first row of the above dataset.

Once we get the data it has to be changed from wide format to long format. reshape2 package is used for reshaping the data. This package contains a command called melt which changes the data from wide format to long format. Now the data is ready

2.2 Exploring of Data

```
> summary(mydata.final)

      Date               variable     value
 Min.   :2013-12-31   Germany:1   Min.   :83.96
 1st Qu.:2013-12-31   Korea  :1   1st Qu.:84.77
 Median :2013-12-31   Canada :1   Median :85.80
 Mean   :2013-12-31   Japan  :1   Mean   :86.13
 3rd Qu.:2013-12-31   Britain:1   3rd Qu.:86.25
 Max.   :2013-12-31               Max.   :89.84

> summary(changedata.final)

      Date               variable     value
 Min.   :2013-12-31   Germany:1   Min.   :0.0000
```

```
 1st Qu.:2013-12-31   Korea  :1   1st Qu.:0.6968
 Median :2013-12-31   Canada :1   Median :1.6114
 Mean   :2013-12-31   Japan  :1   Mean   :1.4945
 3rd Qu.:2013-12-31   Britain:1   3rd Qu.:2.3641
 Max.   :2013-12-31               Max.   :2.8000
```

The above code gives the summary of the data. The max,min,mean values,etc. for the dataset are derived from this given code

```
> str(mydata.final)

'data.frame':        5 obs. of  3 variables:
 $ Date    : Date, format: "2013-12-31" "2013-12-31" ...
 $ variable: Factor w/ 5 levels "Germany","Korea",..: 1 2 3 4 5
 $ value   : num  84 84.8 85.8 86.2 89.8

> str(changedata.final)

'data.frame':        5 obs. of  3 variables:
 $ Date    : Date, format: "2013-12-31" "2013-12-31" ...
 $ variable: Factor w/ 5 levels "Germany","Korea",..: 1 2 3 4 5
 $ value   : num  1.611 0.697 2.8 0 2.364
```

The above code give the data types for different columns and the form in which it is present. Date colum is data type while other three colums are in number fomat.

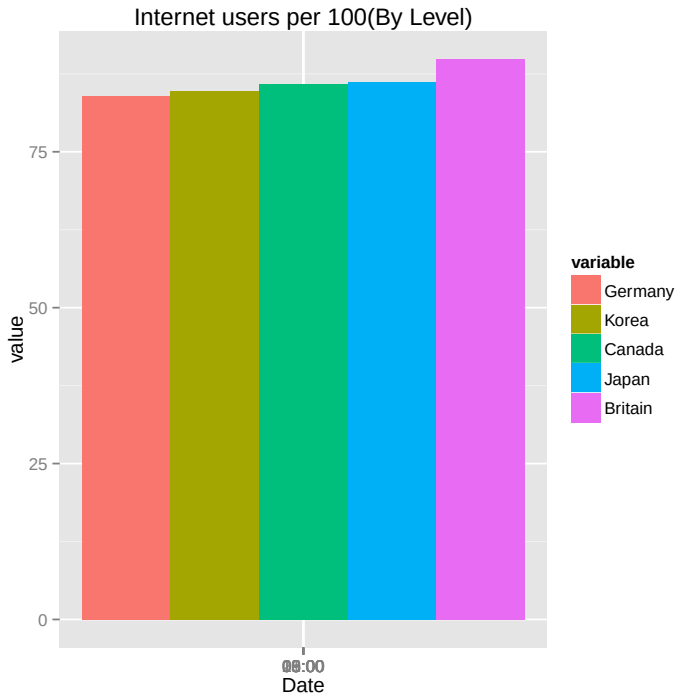
```
> mydata.final

        Date variable   value
1 2013-12-31  Germany 83.9614
2 2013-12-31    Korea 84.7700
3 2013-12-31   Canada 85.8000
4 2013-12-31    Japan 86.2500
5 2013-12-31  Britain 89.8441

> changedata.final

        Date variable     value
1 2013-12-31  Germany 1.6114015
2 2013-12-31    Korea 0.6967735
3 2013-12-31   Canada 2.8000000
4 2013-12-31    Japan 0.0000000
5 2013-12-31  Britain 2.3641016
```

The above code displays the entire data

2.3 Plotting of Graph

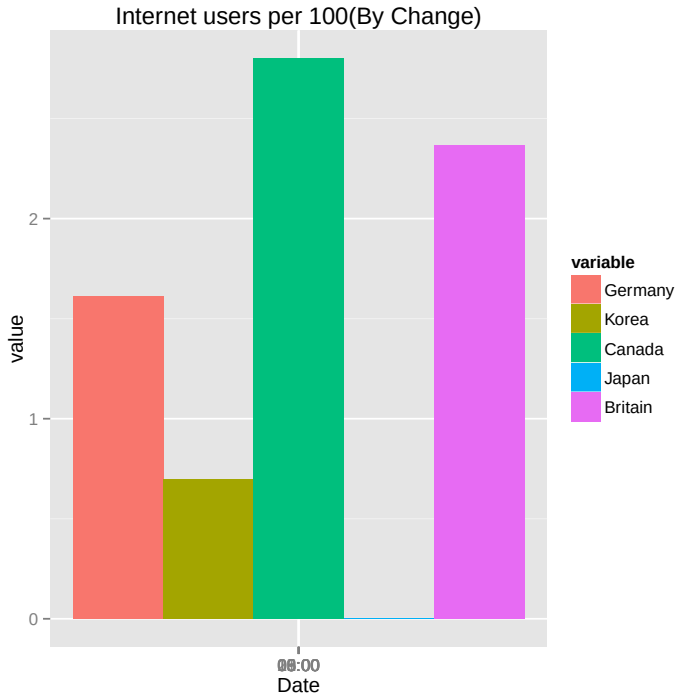
Now we will plot the bar graph for to see the top 5 levels of internet usage among G20 countries per 100 users. The above data that we got , would look good and have better effect if it was in graph.

```
> library(ggplot2)
> plot <- ggplot(data=mydata.final,aes(x=Date,y=value,fill=variable))+geom_bar(stat="identit
> plot + ggtitle("Internet users per 100(By Level)")
```



In the above graph we can see that Great Britain has the highest level of internet users by country by the end of 2013. Also, there is not much difference in the level among these 5 countries which shows that there is not much discrepancy in the level of internet usage.

```
> plot1 <- ggplot(data=changedata.final,aes(x=Date,y=value,fill=variable))+geom_bar(stat="id
> plot1 + ggtitle("Internet users per 100(By Change)")
```



In the above graph we have plotted the data for the change in internet users per 100 from 2012-2013. Here we see that there is no change in level of internet users for Japan from 2012-2013.