SA app user data Latin America 1Dec20-1Jan22

Dani Behonick

2/2/2022

```
install.packages("readr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

NEW USERS

```
LAnewusers_Dec20Jan22 <- read.csv("newusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")
```

Data cleaning

```
#columns renamed
colnames(LAnewusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_06", "2021_11", "2021_05", "2021_02", "2021_08",
"2021_12", "2021_07", "2021_10", "2021_09", "2021_04", "2021_03",
"2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAnewusers_Dec20Jan22 <- LAnewusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |
```

```
Country_ID == "EC")

#transposed rows/columns
LAnewusers_Dec20Jan22_2 <- t(LAnewusers_Dec20Jan22[-1])
colnames(LAnewusers_Dec20Jan22_2) <- LAnewusers_Dec20Jan22[, 1]

#making time a recognized column variable
LAnewusers_Dec20Jan22_2 <- cbind(rownames(LAnewusers_Dec20Jan22_2),
  LAnewusers_Dec20Jan22_2)
rownames(LAnewusers_Dec20Jan22_2) <- NULL

#Totals row removed
LAnewusers_Dec20Jan22_2 <- LAnewusers_Dec20Jan22_2[-c(15), ]

#make it a dataframe
LAnewusers_Dec20Jan22_2 <- as.data.frame(LAnewusers_Dec20Jan22_2)
```

```
install.packages("tidyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(tidyr)
```

```
#pivot to make CountryID a callable variable
LAnewusers_Dec20Jan22_3 <- pivot_longer(LAnewusers_Dec20Jan22_2, cols=2:8, names_to = "Country", values

#rename columns
colnames(LAnewusers_Dec20Jan22_3) <- c("Year_Month","CountryID","New_Users")

#make it a dataframe
LAnewusers_Dec20Jan22_3 <- as.data.frame(LAnewusers_Dec20Jan22_3)
```

```
#coerce New_Users to numeric variable rather than character
typeof(LAnewusers_Dec20Jan22_3$New_Users)
```

```
## [1] "character"
```

```
LAnewusers_Dec20Jan22_3$New_Users <- as.numeric(LAnewusers_Dec20Jan22_3$New_Users)
typeof(LAnewusers_Dec20Jan22_3$New_Users)
```

```
## [1] "double"
```

```
#make Month_Year a date/time object
install.packages("zoo")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(zoo)
```

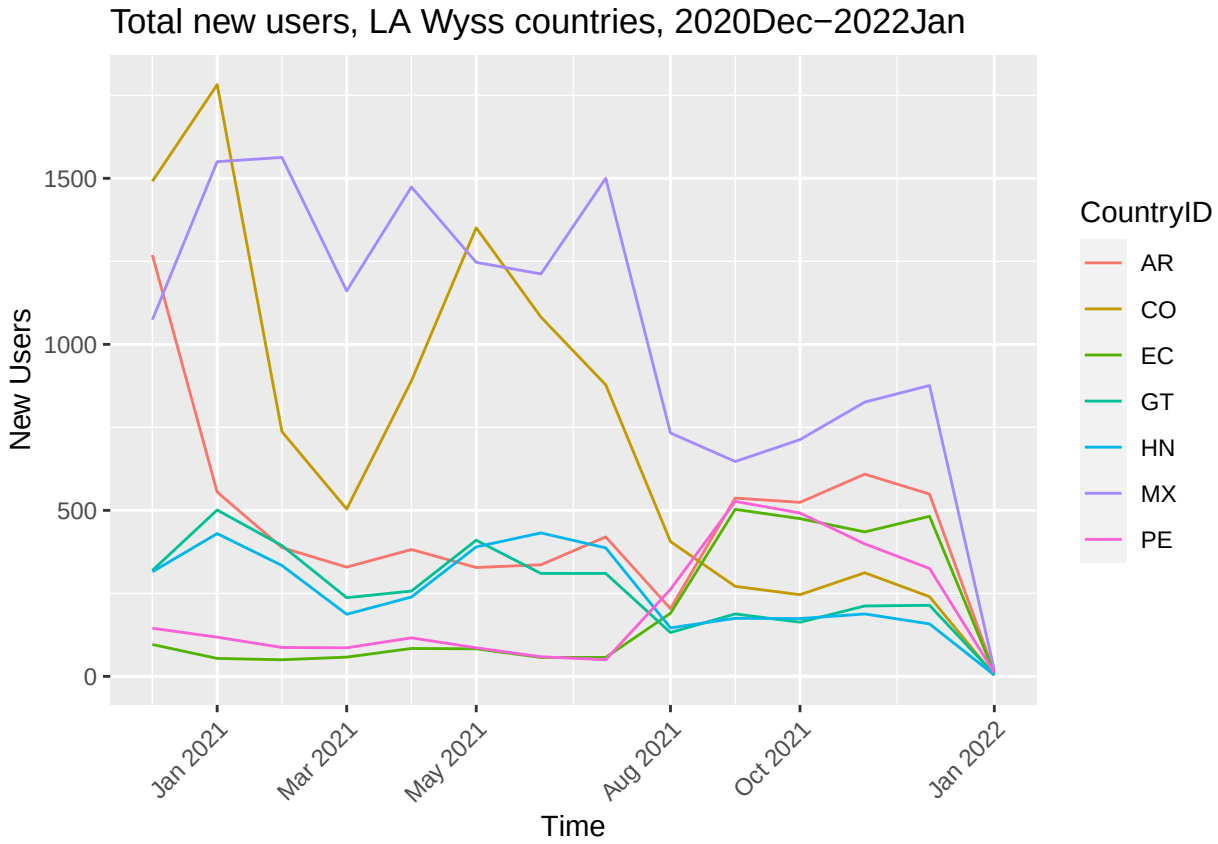
```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
Year_Month <- LAnewusers_Dec20Jan22_3$Year_Month
Year_Month_2 <- as.yearmon(as.character(Year_Month), "%Y_%m")
LAnewusers_Dec20Jan22_3[ , 'Year_Month'] <- Year_Month_2
typeof(LAnewusers_Dec20Jan22_3$Year_Month)
```

```
## [1] "double"
```

```
#jpeg('LAnewusers_Dec20Jan22_3.jpg')
ggplot(data = LAnewusers_Dec20Jan22_3, aes(x=Year_Month, y=New_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Time", y = "New Users",
  title = "Total new users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```



RETURN USERS

```
LAreturningusers_Dec20Jan22 <- read.csv("returningusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")
```

Data cleaning

```
#columns renamed
colnames(LAreturningusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_02", "2021_05", "2021_06", "2021_04", "2021_03",
"2021_12", "2021_10", "2021_07", "2021_11", "2021_09", "2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAreturningusers_Dec20Jan22 <- LAreturningusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |
```

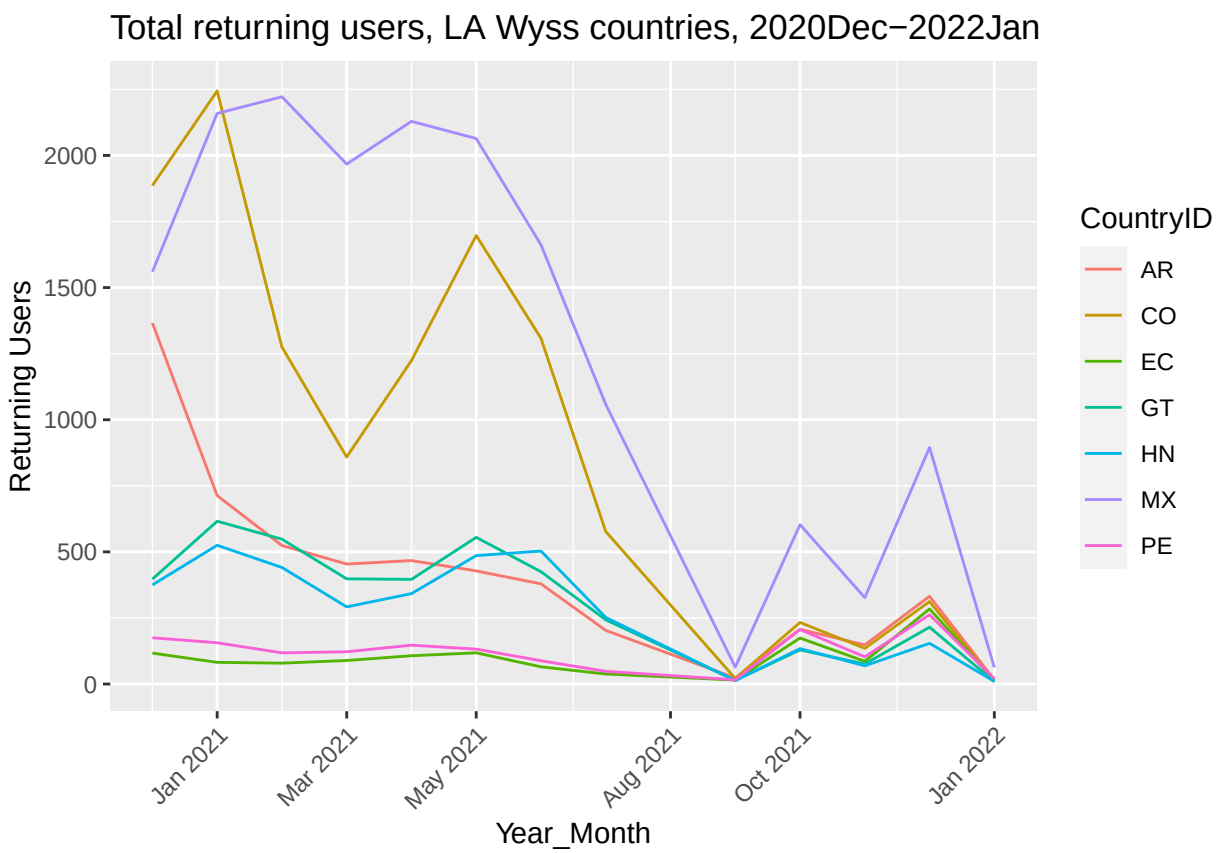
```
Country_ID == "EC")

#transposed rows/columns
LAreturningusers_Dec20Jan22_2 <- t(LAreturningusers_Dec20Jan22[-1])
colnames(LAreturningusers_Dec20Jan22_2) <- LAreturningusers_Dec20Jan22[, 1]

#making time a recognized column variable
LAreturningusers_Dec20Jan22_2 <- cbind(rownames(LAreturningusers_Dec20Jan22_2),
  LAreturningusers_Dec20Jan22_2)
rownames(LAreturningusers_Dec20Jan22_2) <- NULL

#Totals row removed
LAreturningusers_Dec20Jan22_2 <- LAreturningusers_Dec20Jan22_2[-c(14), ]

#make it a dataframe
LAreturningusers_Dec20Jan22_2 <- as.data.frame(LAreturningusers_Dec20Jan22_2)

#pivot to make CountryID a callable variable
LAreturningusers_Dec20Jan22_3 <- pivot_longer(LAreturningusers_Dec20Jan22_2, cols=2:8, names_to = "Coun

#rename columns
colnames(LAreturningusers_Dec20Jan22_3) <- c("Year_Month","CountryID","Returning_Users")

#make it a dataframe
LAreturningusers_Dec20Jan22_3 <- as.data.frame(LAreturningusers_Dec20Jan22_3)

#coerce Returning_Users to numeric variable rather than character
LAreturningusers_Dec20Jan22_3$Returning_Users <- as.numeric(LAreturningusers_Dec20Jan22_3$Returning_Use
typeof(LAreturningusers_Dec20Jan22_3$Returning_Users)
```

```
## [1] "double"
```

```
#make Month_Year a date/time object
Year_Month_returning <- LAreturningusers_Dec20Jan22_3$Year_Month
Year_Month_returning_2 <- as.yearmon(as.character(Year_Month_returning), "%Y_%m")
LAreturningusers_Dec20Jan22_3[ , 'Year_Month'] <- Year_Month_returning_2
typeof(LAreturningusers_Dec20Jan22_3$Year_Month)
```

```
## [1] "double"
```

```
#jpeg('LAreturningusers_Dec20Jan22_3.jpg')
ggplot(data = LAreturningusers_Dec20Jan22_3, aes(x=Year_Month, y=Returning_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Year_Month", y = "Returning Users",
  title = "Total returning users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```



TOTAL USERS

```
LAtotalusers_Dec20Jan22 <- read.csv("totalusers_20201201to20220101.csv",
skip = 1, head = FALSE, sep = ",")
```

Data cleaning

```
#columns renamed
colnames(LAtotalusers_Dec20Jan22) <- c("Country_ID", "2021_01",
"2020_12", "2021_02", "2021_08", "2021_12", "2021_11", "2021_06",
"2021_07", "2021_05", "2021_09", "2021_10", "2021_04", "2021_03", "2022_01", "Totals")

#header rows and data entries for non-Wyss LA countries removed
LAtotalusers_Dec20Jan22 <- LAtotalusers_Dec20Jan22 %>%
filter(Country_ID == "MX" | Country_ID == "AR" | Country_ID == "CO" |
Country_ID == "GT" | Country_ID == "HN" | Country_ID == "PE" |
Country_ID == "EC")

#transposed rows/columns
LAtotalusers_Dec20Jan22_2 <- t(LAtotalusers_Dec20Jan22[-1])
colnames(LAtotalusers_Dec20Jan22_2) <- LAtotalusers_Dec20Jan22[, 1]

#making time a recognized column variable
LAtotalusers_Dec20Jan22_2 <- cbind(rownames(LAtotalusers_Dec20Jan22_2),
  LAtotalusers_Dec20Jan22_2)
rownames(LAtotalusers_Dec20Jan22_2) <- NULL

#Totals row removed
```

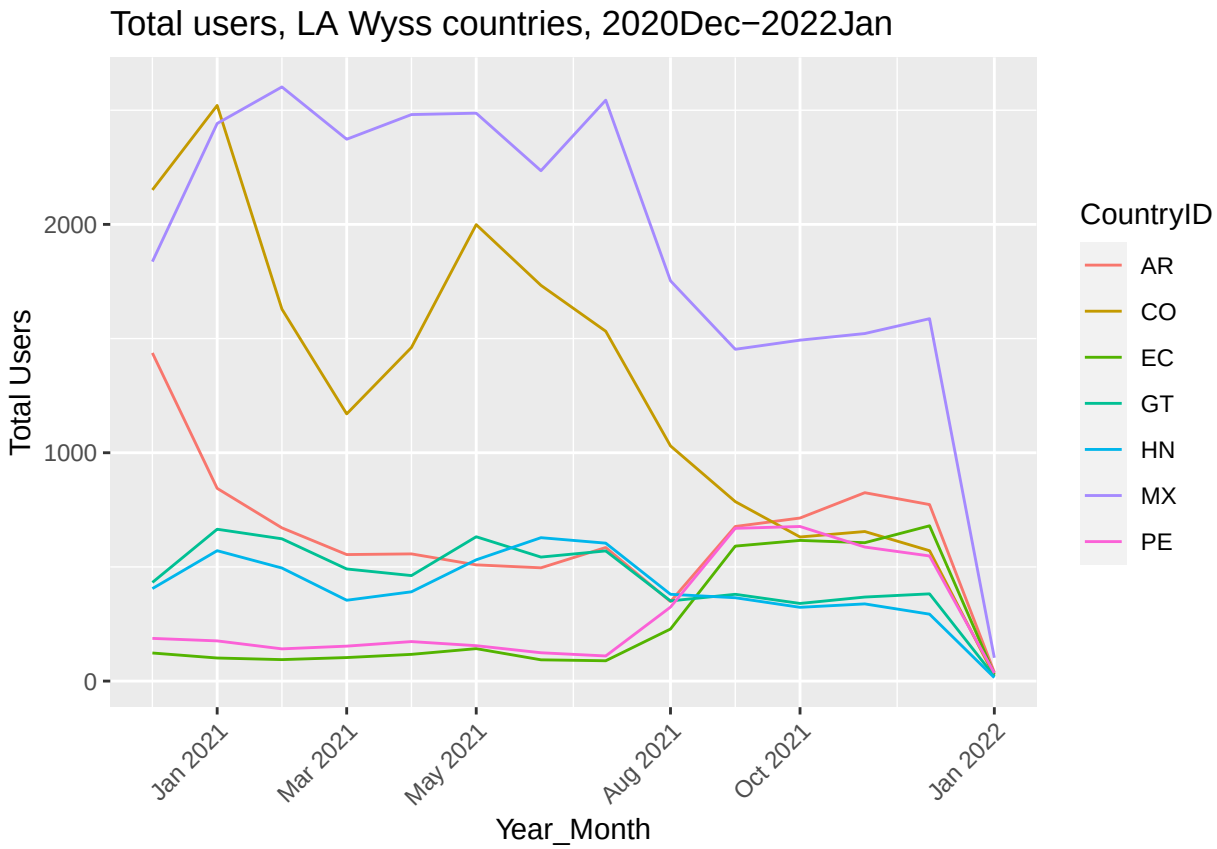
```
LAtotalusers_Dec20Jan22_2 <- LAtotalusers_Dec20Jan22_2[-c(15), ]

#make it a dataframe
LAtotalusers_Dec20Jan22_2 <- as.data.frame(LAtotalusers_Dec20Jan22_2)

#pivot to make CountryID a callable variable
LAtotalusers_Dec20Jan22_3 <- pivot_longer(LAtotalusers_Dec20Jan22_2, cols=2:8, names_to = "Country", va

#rename columns
colnames(LAtotalusers_Dec20Jan22_3) <- c("Year_Month","CountryID","Total_Users")

#make it a dataframe
LAtotalusers_Dec20Jan22_3 <- as.data.frame(LAtotalusers_Dec20Jan22_3)

#coerce Total_Users to numeric variable rather than character
LAtotalusers_Dec20Jan22_3$Total_Users <- as.numeric(LAtotalusers_Dec20Jan22_3$Total_Users)
typeof(LAtotalusers_Dec20Jan22_3$Total_Users)
```

```
## [1] "double"
```

```
#make Month_Year a date/time object
Year_Month_total <- LAtotalusers_Dec20Jan22_3$Year_Month
Year_Month_total_2 <- as.yearmon(as.character(Year_Month_total), "%Y_%m")
LAtotalusers_Dec20Jan22_3[ , 'Year_Month'] <- Year_Month_total_2
typeof(LAtotalusers_Dec20Jan22_3$Year_Month)
```

```
## [1] "double"
```

```
#jpeg('LAtotalusers_Dec20Jan22_3.jpg')
ggplot(data = LAtotalusers_Dec20Jan22_3, aes(x=Year_Month, y=Total_Users)) +
  geom_line(aes(color=CountryID)) + labs(x = "Year_Month", y = "Total Users",
  title = "Total users, LA Wyss countries, 2020Dec-2022Jan") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1))
```



PREDICTING FUTURE USE