Google trends and Map Tutorial with hurricanes!

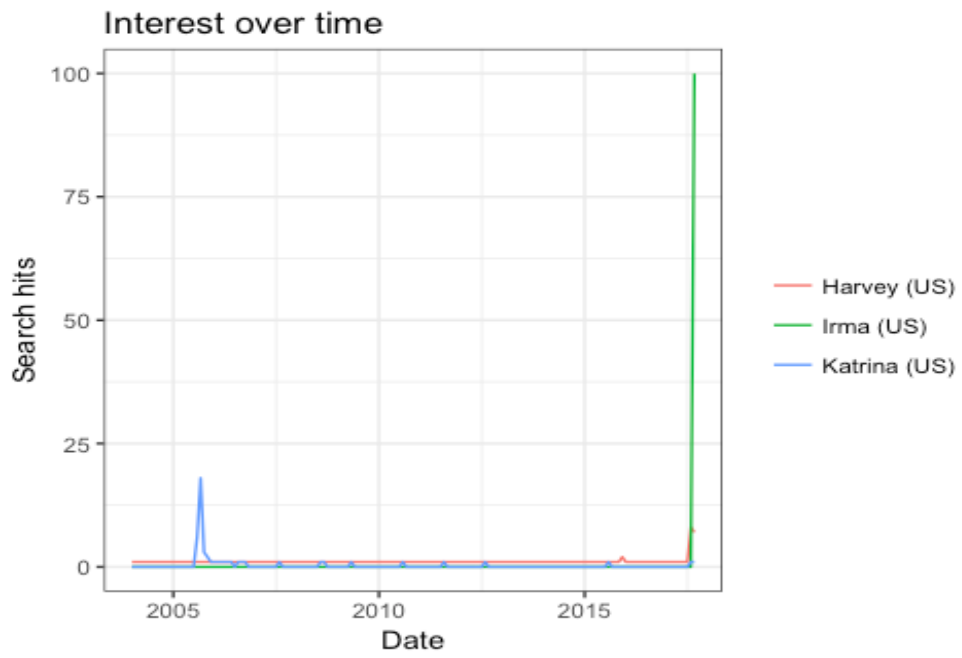
This is a quick introduction on how to get and visualize time series and geographical data from Google Trends using the packages gtrendsR, maps and ggplot2. In this tutorial, we will look at search interest for particular named hurricanes.

First, we load the required packages. This little shortcut will also install packages in case you haven't already. Note that we use devtools to download a developer version of gtrendsR.

```
devtools::install_github('PMassicotte/gtrendsR')
if(!require("pacman")) install.packages("pacman")
pacman::p_load(gtrendsR,maps,ggplot2)
```

Let's first look at how the impact of hurricanes Katrina, Harvey and Irma is reflected in how Americans have used these names as google search items over time. The gprop argument controls whether we want e.g. general web, news, image or youtube searches. The time argument is set to "all" and will gather data between 2004 and the time the code is run.

```
hurricanes = gtrends(c("Katrina","Harvey","Irma"), gprop = "web",time="all", geo = c("US"))
plot(hurricanes)
```



As you can see, the gtrends function actually returns a list of data frames.

```
str(hurricanes)
```

```
## List of 6
##  $ interest_over_time:'data.frame':  495 obs. of  6 variables:
##   ..$ date    : Date[1:495], format: "2004-01-01" ...
##   ..$ hits    : int [1:495] 0 0 0 0 0 0 0 0 0 0 ...
##   ..$ keyword : chr [1:495] "Katrina" "Katrina" "Katrina" "Katrina" ...
##   ..$ geo     : chr [1:495] "US" "US" "US" "US" ...
##   ..$ gprop   : chr [1:495] "web" "web" "web" "web" ...
##   ..$ category: int [1:495] 0 0 0 0 0 0 0 0 0 0 ...
##  $ interest_by_region:'data.frame':  306 obs. of  5 variables:
##   ..$ location: chr [1:306] "Louisiana" "Florida" "North Dakota" "Texas" ...
##   ..$ hits    : int [1:306] 44 12 10 14 11 33 12 13 15 16 ...
##   ..$ keyword : chr [1:306] "Katrina" "Katrina" "Katrina" "Katrina" ...
##   ..$ geo     : chr [1:306] "US" "US" "US" "US" ...
##   ..$ gprop   : chr [1:306] "web" "web" "web" "web" ...
##  $ interest_by_dma   :'data.frame':  1259 obs. of  5 variables:
##   ..$ location: chr [1:1259] "New Orleans LA" "Ft. Myers-Naples FL" "Wheeling W
V-Steubenville OH" "Tampa-St. Petersburg (Sarasota) FL" ...
##   ..$ hits    : int [1:1259] 40 7 6 8 7 11 8 8 9 9 ...
##   ..$ keyword : chr [1:1259] "Katrina" "Katrina" "Katrina" "Katrina" ...
##   ..$ geo     : chr [1:1259] "US" "US" "US" "US" ...
##   ..$ gprop   : chr [1:1259] "web" "web" "web" "web" ...
##  $ interest_by_city  :'data.frame':  300 obs. of  5 variables:
##   ..$ location: chr [1:300] "Brookneal" "New Orleans" "Orlando" "Jacksonville"
...
##   ..$ hits    : int [1:300] 100 56 10 11 11 48 12 11 39 12 ...
##   ..$ keyword : chr [1:300] "Katrina" "Katrina" "Katrina" "Katrina" ...
##   ..$ geo     : chr [1:300] "US" "US" "US" "US" ...
##   ..$ gprop   : chr [1:300] "web" "web" "web" "web" ...
##  $ related_topics    : NULL
##  $ related_queries   :'data.frame':  150 obs. of  6 variables:
##   ..$ subject        : chr [1:150] "100" "100" "45" "15" ...
##   ..$ related_queries: chr [1:150] "top" "top" "top" "top" ...
##   ..$ value          : chr [1:150] "hurricane" "hurricane katrina" "katrina kai
f" "new orleans katrina" ...
##   ..$ geo            : chr [1:150] "US" "US" "US" "US" ...
##   ..$ keyword        : chr [1:150] "Katrina" "Katrina" "Katrina" "Katrina" ...
##   ..$ category       : int [1:150] 0 0 0 0 0 0 0 0 0 0 ...
##   ..- attr(*, "reshapeLong")=List of 4
##   .. ..$ varying:List of 1
##   .. .. ..$ value: chr "top"
##   .. .. ..- attr(*, "v.names")= chr "value"
##   .. .. ..- attr(*, "times")= chr "top"
##   .. ..$ v.names: chr "value"
##   .. ..$ idvar  : chr "id"
```

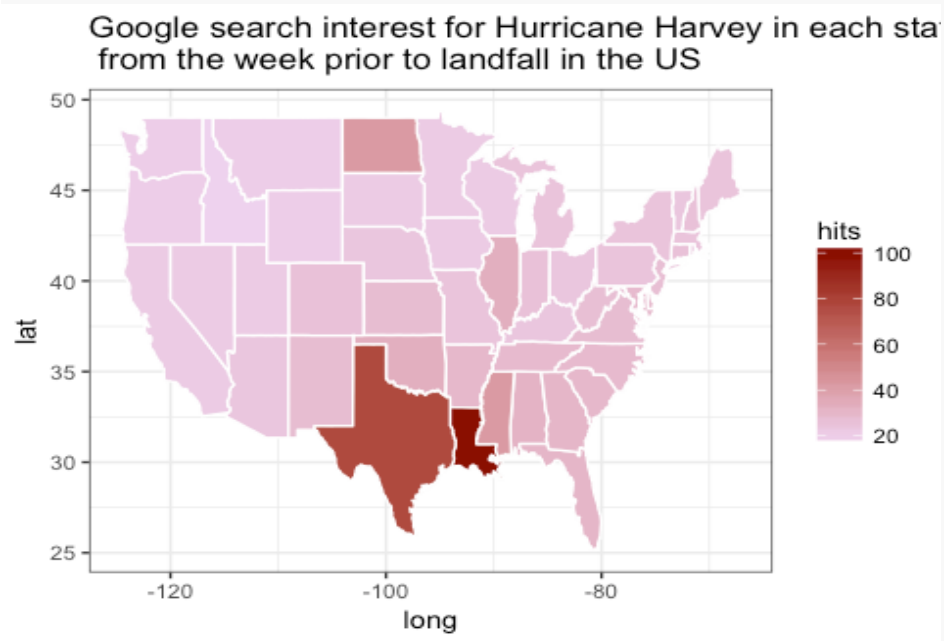
```
##   .. ..$ timevar: chr "related_queries"
##  - attr(*, "class")= chr [1:2] "gtrends" "list"
```

Now, we will compare the amount of interest in Hurricane Harvey for each US state. There are different ways of specifying start and end dates, but these methods don't seem to work properly. Hopefully, this will be fixed in later versions. In this example, we use the current date and include searches coing back one year.

```
harvey = gtrends(c("Harvey"), gprop = "web",time="today 12-m", geo = c("US"))
harvey = harvey$interest_by_region
harvey$region = sapply(harvey$location,tolower)
statesMap = map_data("state")
harveyMerged = merge(statesMap,harvey,by="region")
```

On the fourth line in the code above, we fetch an empty map for plotting our geographical data. On the fifth line, we merge our google trends data with the underlying map data. Both data sets contain a column called "region", which will be used to merge the data frames. Note that the region labels need to be identical. therefore, on line 3, we change the capitalized state names to lowercase. Now we can plot the data!

```
harveyPlot=ggplot() +
  geom_polygon(data=harveyMerged,aes(x=long,y=lat,group=group,fill=hits),colour="w
hite") +
  scale_fill_continuous(low="thistle2",high="darkred",guide="colorbar") +
  theme_bw() +
  labs(title="Google search interest for Hurricane Harvey in each state\n from the
week prior to landfall in the US")
harveyPlot
```



We then use the same procedure for plotting regional searches for Hurricane Irma, except we add a label for each state and change the colour.

```
irma = gtrends(c("Irma"), gprop = "web", time="today 12-m",geo = c("US"))
irma = irma$interest_by_region
irma$region = sapply(irma$location,tolower)
statesMap = map_data("state")
irmaMerged = merge(statesMap ,irma,by="region")

regionLabels <- aggregate(cbind(long, lat) ~ region, data=irmaMerged,
                          FUN=function(x) mean(range(x)))

irmaPlot=ggplot() +
  geom_polygon(data=irmaMerged,aes(x=long,y=lat,group=group,fill=hits),colour="whi
te") +
  scale_fill_continuous(low="thistle2",high="darkblue",guide="colorbar") +
  geom_text(data=regionLabels, aes(long, lat, label = region), size=2) +
  theme_bw() +
  labs(title="Google search interest for Hurricane Irma in each state\n from the w
eek prior to landfall in the US")
irmaPlot
```