Capstone Vignette

Muhammad Ezzat

3/27/2020

Inroduction

This vinette is dedicated to explain functions from my Capstone Project of the Coursera Specialization : Mastering Software development with R.Without any further to say, let's try out our functions.

Cleaning Data

eq_location_clean

This function takes the location name as an argument & cleans it out by removing the country's name and adjusting the location name.

```
clean_labels <- eq_location_clean(data$LOCATION_NAME)
print(head(clean_labels))
```

```
## [1] "  BAB-A-DARAA,AL-KARAK"     "  UGARIT"
## [3] "  W"                        "  THERA ISLAND (SANTORINI)"
## [5] "  ARIHA (JERICHO)"          "  LACUS CIMINI"
```

eq_clean_data

This function cleans the whole data set, it adjusts the location name, removes usless columns & changes longtiude & latitude columns to numeric columns. It also removes missing values on comand by setting na.rm argument to TRUE.

```
Fclean_data <- eq_clean_data(data)
```

```
## Warning: 238 failed to parse.
```

```
Tclean_data <- eq_clean_data(data,T)
print(head(Fclean_data))
```

```
## # A tibble: 6 x 13
##     I_D DATE        YEAR MONTH   DAY EQ_PRIMARY LATITUDE LONGITUDE
##   <dbl> <date>     <dbl> <dbl> <dbl>      <dbl>    <dbl>     <dbl>
## 1     1 2150-01-01 -2150     1     1        7.3     31.1      35.5
## 2     2 2000-01-01 -2000     1     1       NA       35.7      35.8
## 3     3 2000-01-01 -2000     1     1        7.1     38        58.2
## 4  5877 1610-01-01 -1610     1     1       NA       36.4      25.4
## 5     8 1566-01-01 -1566     1     1       NA       31.5      35.3
## 6    11 1450-01-01 -1450     1     1       NA       35.5      25.5
## # ... with 5 more variables: LOCATION_NAME <chr>, COUNTRY <chr>,
## #   STATE <chr>, TOTAL_DEATHS <dbl>, TIME <chr>
```

```
print(head(Tclean_data))
```

```
## # A tibble: 6 x 13
##     I_D DATE        YEAR MONTH   DAY EQ_PRIMARY LATITUDE LONGITUDE
##   <dbl> <date>     <dbl> <dbl> <dbl>      <dbl>    <dbl>     <dbl>
## 1  1609 1812-12-08  1812    12     8        6.9     34.4    -118.
```

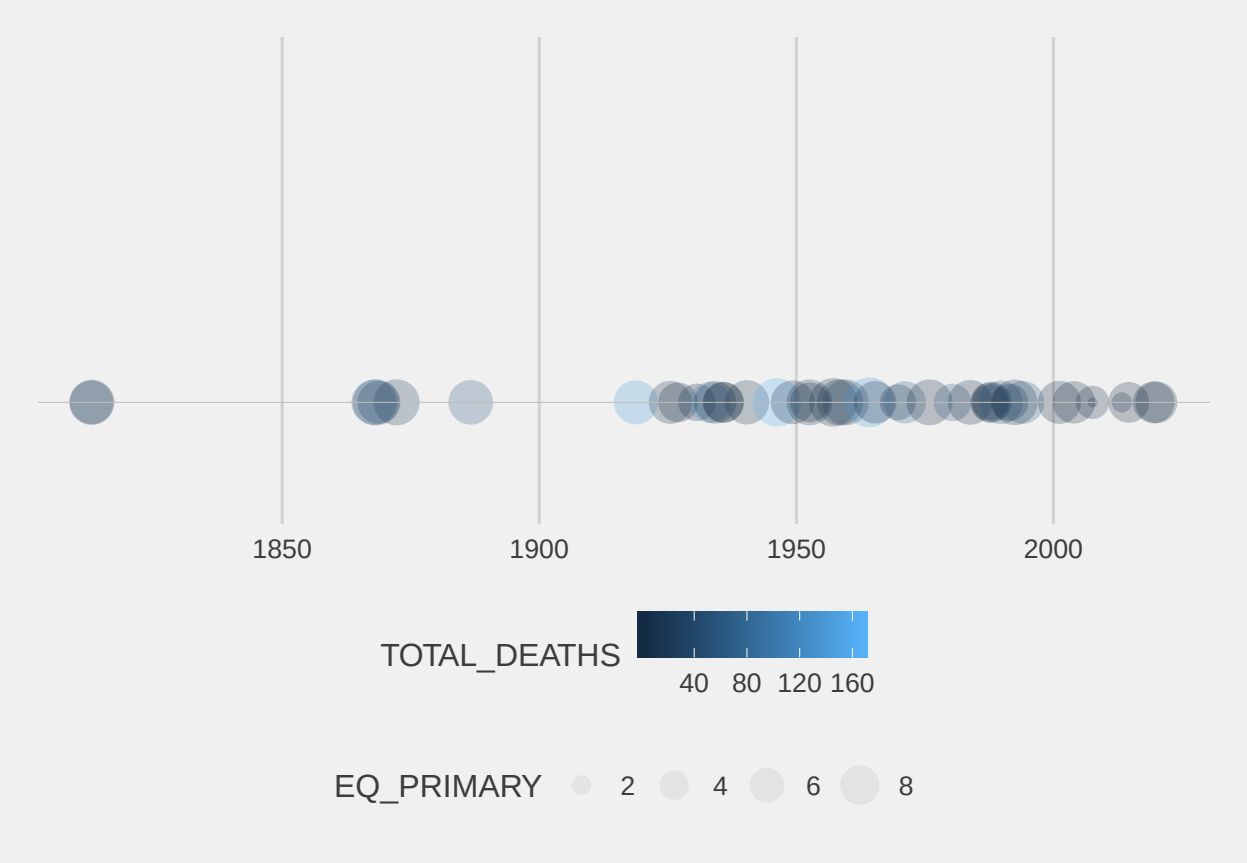
```
## 2  1610 1812-12-21  1812    12    21        7.5     34.2    -120.
## 3  2120 1867-11-18  1867    11    18        7.5     18.1     -65.1
## 4  2125 1868-04-03  1868     4     3        7.9     19      -156.
## 5  2148 1868-10-21  1868    10    21        6.8     37.7    -122.
## 6  2195 1872-03-26  1872     3    26        7.8     36.7    -118.
## # ... with 5 more variables: LOCATION_NAME <chr>, COUNTRY <chr>,
## #   STATE <chr>, TOTAL_DEATHS <dbl>, TIME <chr>
```

Visualizing Data I

After we're done with Module 1, Module 2 required two geoms, geom_timeline() & geom_timeline_label(). Let's watch them in action.

```
gObj <- ggplot(Tclean_data,aes(x=DATE))+theme_fivethirtyeight()
gObj + geom_timeline(aes(size = EQ_PRIMARY, col=TOTAL_DEATHS))
```

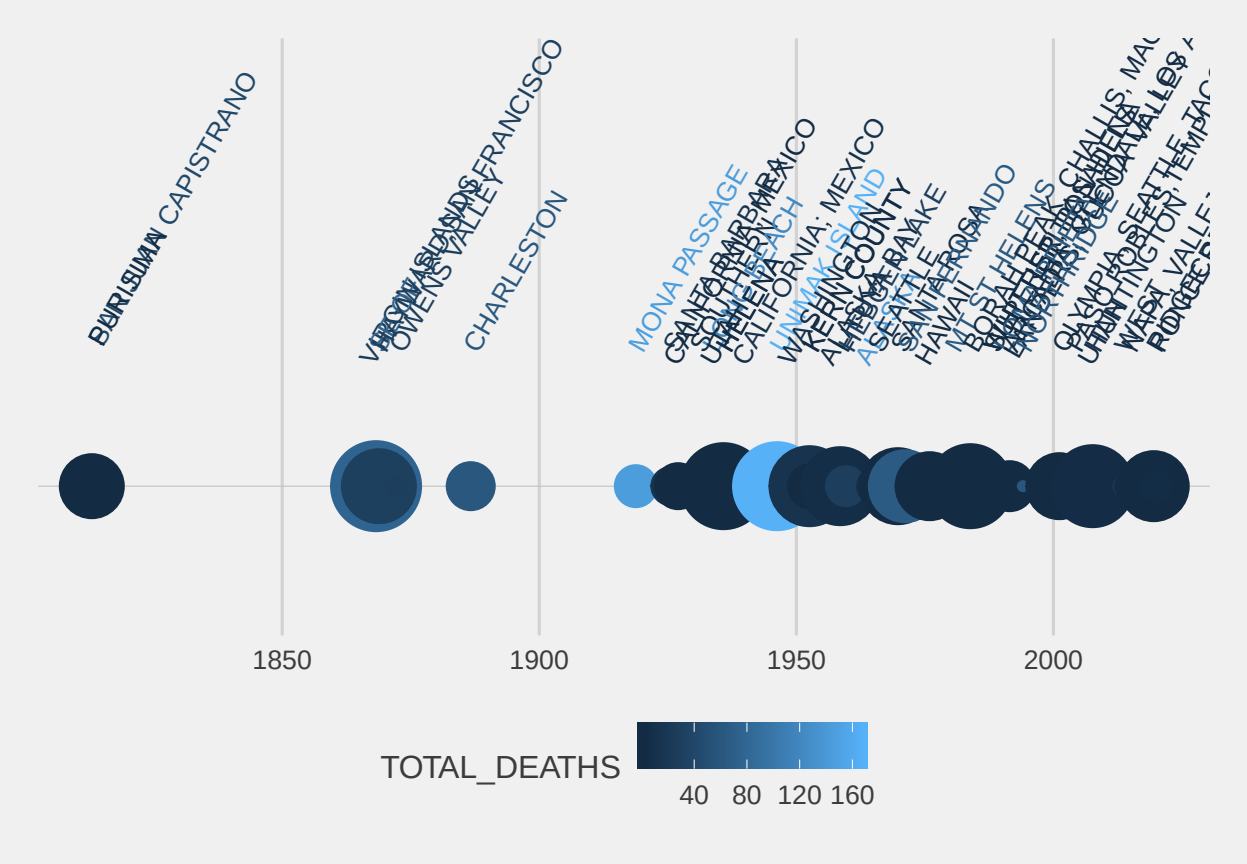
```
## Warning in f(...): missing values for colour. They were replaced with the
## minimum value.
```




```
gObj + geom_timeline_label(aes(magnitude = EQ_PRIMARY, label=LOCATION_NAME, col = TOTAL_DEATHS))
```

```
## Warning: Ignoring unknown aesthetics: magnitude
```

```
## Warning in f(...): size is not provided.
##  a random sample of points will be used
```



Unfortunately geom_timeline_label() isn't doing it's best, but it's under improvment.

Visualizing Data II

The third module was about using leaflet interactive maps, we had to functions, one creates an HTML label consists of many elemnts, the other was to plot. We can use them as follow

```
dplyr::mutate(Tclean_data,label = eq_create_label(Tclean_data))
```

```
## # A tibble: 45 x 14
##      I_D DATE        YEAR MONTH   DAY EQ_PRIMARY LATITUDE LONGITUDE
##    <dbl> <date>     <dbl> <dbl> <dbl>      <dbl>    <dbl>     <dbl>
##  1  1609 1812-12-08  1812    12     8        6.9     34.4    -118.
##  2  1610 1812-12-21  1812    12    21        7.5     34.2    -120.
##  3  2120 1867-11-18  1867    11    18        7.5     18.1     -65.1
##  4  2125 1868-04-03  1868     4     3        7.9     19      -156.
##  5  2148 1868-10-21  1868    10    21        6.8     37.7    -122.
##  6  2195 1872-03-26  1872     3    26        7.8     36.7    -118.
##  7  2362 1886-09-01  1886     9     1        7.3     32.9     -80
##  8  3124 1918-10-11  1918    10    11        7.1     18.7     -67.2
##  9  3270 1925-06-29  1925     6    29        6.8     34.4    -120.
## 10  3305 1927-01-01  1927     1     1        5.8     32.5    -116.
## # ... with 35 more rows, and 6 more variables: LOCATION_NAME <chr>,
## #   COUNTRY <chr>, STATE <chr>, TOTAL_DEATHS <dbl>, TIME <chr>,
## #   label <chr>
```

```
eq_map(Tclean_data,'label')
```

One little warning that it won't appear on a pdf vinette, it requires an HTML one.

That was my submission hope you rate it good, thank you.