Data preprocessing

Marco Galliani

Contents

Settings

```
rm(list = ls())
```

The dataset

The statistical units are the bombs exploded in the first 24 hours of the bombing of London, on September 7th, 1940. For each bomb the following informations are provided: - Order - Time - Location - Type of bomb (IB: Incendiary Bomb, EB: Explosive Bomb, COB: Crude Oil Bomb) - Damage or other details ("Damage.or.other.details..All.dimensions.at.in.ft.unless.stated."))

We got data 843 bombs.

Loading the original dataset, downloaded from here

```
library(readxl)

data_path <- "../data/September 7, 1940_ first night of the Blitz.xlsx"

bomb_data <- read_excel(path = data_path,
                        col_types = c("numeric", "date", "text", "text", "text")
                        )

otro_data <- read.csv(file = "../data/September 7, 1940_ first night of the Blitz - SEPT 7, ALL REPORTS

colnames(bomb_data) <- c("Order", "Time", "Location", "Type.of.bomb", "Damage.or.other")
```

Preprocessing

- fixing the wrong date imported by the excel

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

```
date(bomb_data$Time) <- "1940-9-7"
```

- getting Latitude and Longitude from the addresses (geocoding)

Done using this online tool: geoapify

```
# splitting data since it only accept datasets of size up to 500 rows
first_batch <- bomb_data[1:500, 3]
second_batch <- bomb_data[501:nrow(bomb_data), 3]

write.csv(first_batch, "temp_data/first_batch.csv")
write.csv(second_batch, "temp_data/second_batch.csv")
```

```
geocoded_first_batch <- read.csv("temp_data/geocoded_first_batch.csv")
geocoded_second_batch <- read.csv("temp_data/geocoded_second_batch.csv")

geocoded_data <- rbind.data.frame(geocoded_first_batch, geocoded_second_batch)
rm(first_batch, second_batch, geocoded_first_batch, geocoded_second_batch)

bomb_data[, c("lat", "lon")] <- geocoded_data[, c("lat", "lon")]
rm(geocoded_data)
```

- getting the district where the bomb fell

```
library(terra)
```

```
## terra 1.7.55
```

```
library(tidyterra)
```

```
## Warning: package 'tidyterra' was built under R version 4.2.3
```

```
##
## Attaching package: 'tidyterra'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library(ggplot2)

# loading data for the map of London
london_spat_vect <- vect("../data/London-data/London_Borough_Excluding_MHW.shp")

# projecting the longitude and latitude values on the reference system used by the geocoding service
newcrs <- "+proj=longlat +datum=WGS84"
london_spat_vect <- terra::project(london_spat_vect, newcrs)

# getting the districts related to each (lon,lat) value
bomb_data$district <- extract(london_spat_vect, bomb_data[,c("lon","lat")])$NAME
```

- cleaning duplicates in type of bomb

```
bomb_data$Type.of.bomb <- as.factor(bomb_data$Type.of.bomb)

levels(bomb_data$Type.of.bomb) <- list(IB = c("IB", "Ib", "IBIB"),
                                       EB = c("EB", "EB ", "eb", "Eb", "High Explosive Bomb"),
                                       EB.and.IB = c("EB & IB", "EB &IB", "EB &IB", "IB & EB", "IB and
                                       Magnesium.Flare = c("Magnesium Flare"),
```

```
                                    Shrapnel = c("Shrapnel"),
                                    COB = c("COB", "Crude oil bomb"),
                                    Crashed.Aircraft = c("Crashed aircraft"),
                                    Unknow.enemy.action = c("Unknown enemy action"))
```

- cleaning useless variables generated

```
bomb_data$Order <- NULL
knitr::kable(head(bomb_data))
```

Time	Location	Type.of.bomb	Damage.or.other	lat	lon	district
1940-09-07 00:08:00	43 Southwark Park Road, SE16, London, UK	IB	Grocers: 3x2 roof damaged	51.49225	-0.0621761	Southwark
1940-09-07 00:10:00	49 Southwark Park road, Bermondsey, SE16, London, UK	IB	Bakers: 3x2 roof damaged	51.49269	-0.0653908	Southwark
1940-09-07 00:15:00	84 Southwark Park Road, SE16, London, UK	IB	front room on 1st floor and contents slightly damaged. 3x2 rood damage	51.49225	-0.0621761	Southwark
1940-09-07 00:18:00	141 Braidwood Road, Catford SE6, London, UK	IB	10x6 roof damage	51.44085	-0.0053336	Lewisham
1940-09-07 00:20:00	129 Killearn Road, Catford SE6, London, UK	IB	Front room on 1st floor severely damaged	51.44151	-0.0054617	Lewisham
1940-09-07 00:20:00	27 Crutchley Road, Downham, London, UK	IB	IB on enclosed ground at rear of premises	51.43674	-0.0052611	Lewisham

```
write.csv(bomb_data, file = "../data/geocoded_bomb_data.csv")
```