Mini Project 2

Cathal and Daniel

```r
# libraries
library(httr2)
library(stringr)
library(jsonlite)
library(tidyverse)
```

We chose to gather data from the openFDA API. This API serves public data from the Food and Drug Administration about drugs, devices, and foods. It only has publicly available data, so there are no sensitive personal details. We hope to use this data to show a correlation between age or sex and certain adverse effects. We also want to see if some drugs are more likely to cause serious side effects compared to other drugs, and what reactions are most common.

Thus, we decided to scrape reports about adverse drug effects on patients, and filter to only include instances where the patient had severe side effects (Death, life threatening condition, hospitalization, etc). Each report has a unique report ID, details about the patients age and sex, what drug they used, how serious the side effects were, and what their reaction to the drug was. We can group the data by age group, sex, medication name, and create plots to show the size of each group and see if there is a causal relationship between 2 or more of the variables.

```r
key <- readLines("api_token") #Daniel key
key <- "PDPhCwLKWX5EzEoCNrgEhaeVfK5ojIypQOHraqB8" #Cathal key (shhh)
```

Querying the FDA Open API

```r
#Function to get data from the API. 2000 - 2020, numObservations <= 1000

get_fda_data <- function(startYear, endYear, numObservations) {
```

```r
# make the request
endpoint <- "https://api.fda.gov/drug/event.json"

req <- request(str_c(
  endpoint,
  "?api_key=", key,
  "&search=serious:\"1\"+AND+receivedate:[",
  startYear,
  "0101+TO+",
  endYear,
  "1231]&limit=",
  numObservations
))
resp <- req_perform(req)

# if json is desired
resp <- resp |>
  resp_body_json()

reportId = double()
patientAge = double()
patientSex = double()
reaction = character()
drugName = character()
drugIndication = character()

for(i in 1:numObservations) {

  reportId[i-1] <- resp$results[[i]]$safetyreportid

  if(is.null(resp$results[[i]]$patient$patientonsetage)) {
    patientAge[i-1] <- NA
  } else {
    patientAge[i-1] <- resp$results[[i]]$patient$patientonsetage
  }

  if(is.null(resp$results[[i]]$patient$patientsex)) {
    patientSex[i-1] <- NA
  } else {
    patientSex[i-1] <- resp$results[[i]]$patient$patientsex
  }
```

```r
    if(length(resp$results[[i]]$patient$reaction) == 2) {
      reaction[i-1] <- str_c(resp$results[[i]]$patient$reaction[[1]]
                             $reactionmeddrapt, ", ",resp$results[[i]]$
                              patient$reaction[[2]]$reactionmeddrapt)
    } else if(length(resp$results[[i]]$patient$reaction) == 1) {
      reaction[i-1] <- resp$results[[i]]$patient$reaction[[1]]$reactionmeddrapt
    } else {
      reaction[i-1] <- ""
    }

    drugName[i-1] <- resp$results[[i]]$patient$drug[[1]]$medicinalproduct

  }

  drug_tbl <- tibble(
    reportId = parse_number(reportId),
    patientAge = parse_number(patientAge),
    patientSex = parse_number(patientSex),
    drugName = drugName,
    reaction = reaction
  )

  drug_tbl

}

# Get the data tibbles from the API
drugs <- get_fda_data(2010, 2019, 1000)
drugs2 <- get_fda_data(2000, 2009, 1000)

# Join multiple tables
drugs <- drugs |>
  full_join(drugs2)

# Write the csv file
write.csv(drugs, "drugs_tbl.csv")
```

Our Data:

```r
library(tidyverse)

drugs <- read.csv("drugs_tbl.csv")
```

```
drugs |>
  slice_head(n=6)
```

```
  X reportId patientAge patientSex   drugName
1 1 10003301         NA          2  IBUPROFEN
2 2 10003311         76          2   LETAIRIS
3 3 10003312         43          2     ILARIS
4 4 10003315         NA          1 RANOLAZINE
5 5 10003317         46          1   BENLYSTA
6 6 10003318         59          2  BELIMUMAB
                                       reaction
1              Dyspepsia, Renal impairment
2        Oedema peripheral, Fluid retention
3             Pneumonia, Drug ineffective
4                                      Death
5
6 Tuberculosis of central nervous system
```

```
library(treemapify)
```

```
Warning: package 'treemapify' was built under R version 4.4.3
```

```
drugs |>
  mutate(drugName = as.character(fct_lump_n(drugName, 100))) |>
  group_by(drugName) |>
  summarize(n = n()) |>
  mutate(drugName = ifelse(
    nchar(drugName)>7,
    str_c(substr(drugName, start=1, stop=4), "…"),
    drugName
  )) |>
  ggplot(aes(area = n, fill = drugName, label = drugName)) +
    geom_treemap() +
    geom_treemap_text(colour = "white", place = "centre") +
    labs(title = "Drug Representation Among Serious Effects") +
    theme_minimal() +
    scale_fill_viridis_d(option = "G") +
    theme(legend.position = "none")
```

Drug Representation Among Serious Effects
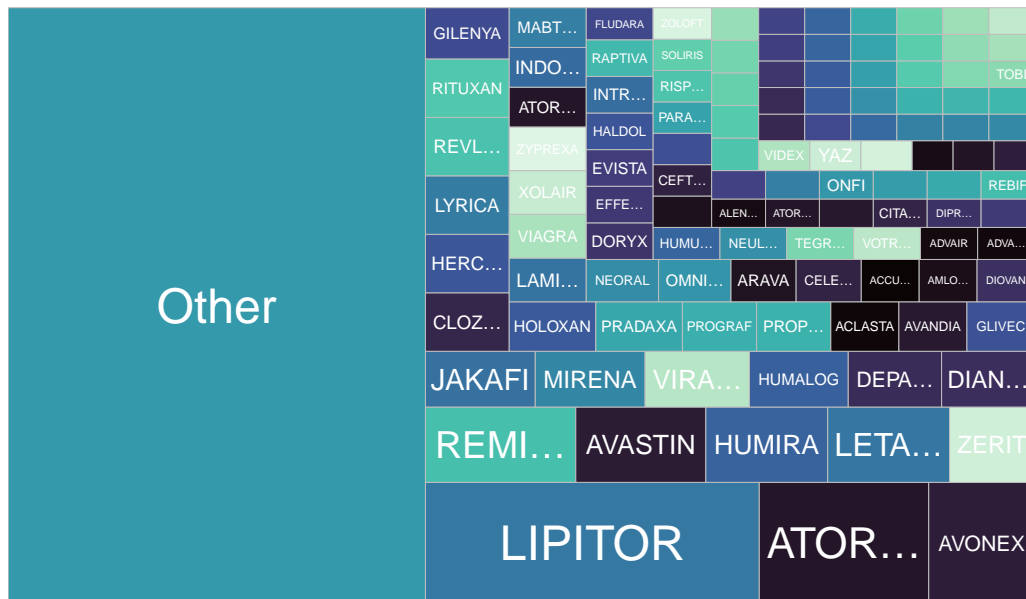

```r
drugs |>
  filter(reaction == "Death") |>
  mutate(drugName = as.character(fct_lump_min(drugName, 1))) |>
  group_by(drugName) |>
  summarize(n = n()) |>
  mutate(drugName = ifelse(
    nchar(drugName)>10,
    str_c(substr(drugName, start=1, stop=6), "…"),
    drugName
  )) |>
  ggplot(aes(area = n, fill = drugName, label = drugName)) +
    geom_treemap() +
    geom_treemap_text(colour = "white", place = "centre") +
    labs(title = "Drug Representation Among Serious Effects") +
    theme_minimal() +
    scale_fill_viridis_d(option = "A") +
    theme(legend.position = "none")
```

Drug Representation Among Serious Effects

GLEEVEC	FEBURIC	OPSUMIT	SOVALDI	TYKERB	XELODA
CUBICIN	DIANEA…	MAINTATE	SIMULECT	TRACLEER	XALATAN
EXJADE	DIANEA…	LYRICA	RANOLAZINE	TEMODAL	TENOFO…
	CARDURA XL	LUPRON…	OXYCON…	POMALYST	PRADAXA
GLIVEC	CAMPATH	FENTAN…	INLYTA	LEVETI…	LIPITOR
	TASIGNA	ACLASTA / AFINITOR	AMBRIS… / AVASTIN / AVODART		
	LETAIRIS	NORVASC	REVLIMID	SANDOS…	