Please find a tutorial with instructions on web scraping data on US police killings in this repository at <https://github.com/adivea/KilledbyPolice2020.git> . Clone it and depending on your familiarity with R, either

2.3) produce data visualisations that shed light on another interesting aspect of the police killing data:

So what we want to do is work with police killings and try and look at some facts and create statistics to understand the situation. Therefore we downloaded the necessary r packages. We then used the URL <https://killedbypolice.net/kbp2020> , <https://killedbypolice.net/kbp> and <https://killedbypolice.net/kbp2018> to get the data down on our r. We then, by following the guide and working with the data created a few plots, that visualized the situation of police killings in the period.

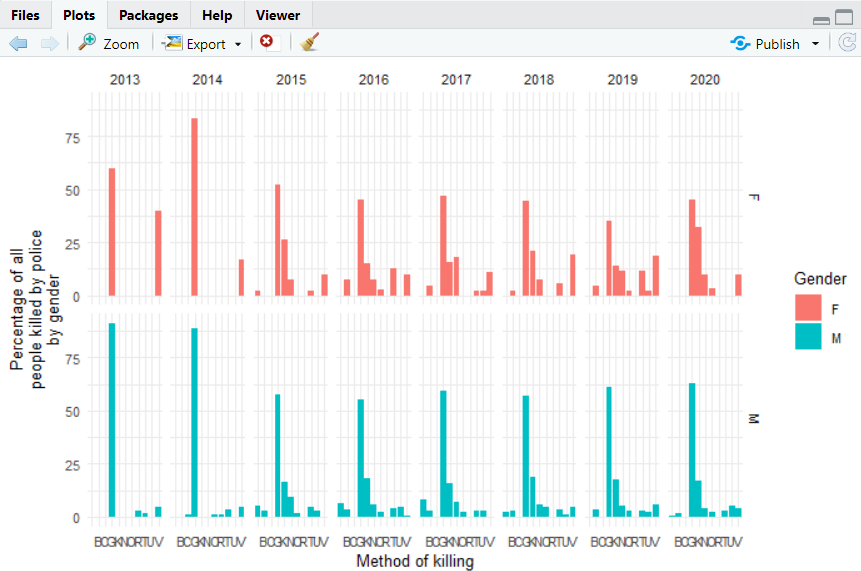
Firstly we created a plot that showed the age difference of male and female victims of police killing in the period 2013-2020:



We then created a plot that showed how the aged differed for people killed by the police from different races in the period 2013-2020. B indicates black, while w means white and l means latino.



At the end we created a graph that showed the way the victims of police murders were killed by, for both males and females.



I ended up with a problem here, that even with the help of different students couldn’t solve. I therefore ended my work here after 3 hours of trying to fix the problem.

It is worth it to note that we used the hashtag before all install.package funtions so that we didn’t end up downloading the packages far too many times.

download.file(rvest)

#install.packages("rvest")

#install.packages("dplyr")

#install.packages("tidyr")

#install.packages("stringr")

#install.packages("janitor")

library(rvest)

library(dplyr)

library(tidyr)

library(stringr)

library(janitor)

url <- "https://killedbypolice.net/kbp2020"

url\_html <- read\_html(url)

whole\_table <- url\_html %>%

html\_nodes("table") %>%

html\_table() #str(whole\_table) turns out to be a list

library(tidyverse)

head(whole\_table)

new\_table <- do.call(cbind,unlist(whole\_table, recursive = FALSE))

head(new\_table) # ok, looks good, too bad it took 2 hours

scrape\_police\_kill <- function(website){

url <- read\_html(website)

annual\_table <- url %>%

html\_nodes("table") %>%

html\_table() # result is a list

annual\_table <- do.call(cbind,unlist(annual\_table, recursive = FALSE))

}

table2018 <- scrape\_police\_kill("https://killedbypolice.net/kbp2018")

table2018 %>%

head()

mastertable=NULL

for (year in 2013:2020){ # here we create a loop to iterate over the years

print(year)

url <- "https://killedbypolice.net/kbp" # the annual URLs end with "kbp2017" ,etc.

website <- paste0(url,year) # here we bind the year to the website to form the URL

annual\_table <- scrape\_police\_kill(website) # here we apply the function

mastertable <- rbind(mastertable, annual\_table) # we add the scraped results from the given year to our master dataset

}

head(mastertable,2)

tail(mastertable)

mastertable <- as\_tibble(mastertable)

str(mastertable)

data <- mastertable %>%

mutate(Age = as.numeric(Age)) %>%

rename(Method = "\*")

mastertable$Date[c(70:80, 160:170)]

tail(unique(mastertable$Date))

library(lubridate)

data <- data %>%

mutate(Date =

case\_when(

grepl("201[34]",Date) ~ mdy(Date),

# convert dates that contain 2013 or 2014 into mdy format

!grepl("201[34]",Date)~ ymd(Date)))

# convert all other dates ymd format

data <- data %>%

mutate(Year = year(Date)) # I am creating a new column Year from the Date for later plots

tail(data$Year)

class(data$Date)

length(which(is.na(data$Date)))

write.csv(data,"data/policekillings202010.csv")

install.packages("ggridges")

install.packages("statebins")

library(ggplot2)

library(ggridges)

#graff for forskel på døde i forhold til mand/kvinde

data %>%

filter(Gender %in% c("F", "M", "T")) %>%

filter(!is.na(Year)) %>%

ggplot(aes(x = Age,

y = factor(Year),

fill = Gender)) +

geom\_density\_ridges(alpha = 0.5,

scale = 0.9) +

theme\_ridges(font\_size = 10) +

scale\_x\_continuous(breaks = seq(0, 100, 10),

labels = seq(0, 100, 10)) +

xlab("Age at death (years)") +

ylab("Year") +

theme(axis.title = element\_text(size = 14))

#Graph for døde Sorte overfor hvide

data %>%

filter(Race %in% c("B", "W", "L")) %>%

filter(!is.na(Year)) %>%

ggplot(aes(x = Age,

y = factor(Year),

fill = Race)) +

geom\_density\_ridges(alpha = 0.6,

scale = 0.9) +

theme\_ridges(font\_size = 10) +

scale\_x\_continuous(breaks = seq(0, 100, 10),

labels = seq(0, 100, 10)) +

xlab("Age at death (years)") +

ylab("Year") +

theme(axis.title = element\_text(size = 14))

#graf for måden de er dræbt på

data %>%

filter(!is.na(Year)) %>%

filter(Method != "NA") %>%

filter(Gender %in% c("M", "F", NA)) %>%

group\_by(Year,

Gender,

Method) %>%

tally() %>%

mutate(perc = n / sum(n) \* 100) %>%

ggplot(aes(Method,

perc,

fill = Gender)) +

geom\_col() +

facet\_grid(Gender~Year) +

theme\_minimal(base\_size = 10) +

xlab("Method of killing") +

ylab("Percentage of all\npeople killed by police\nby gender")

#install.packages(c("statebins", "viridis"))

library(statebins) # using GitHub version

library(viridis)

state\_abb <- data\_frame(state\_name = state.name,

state\_abb = state.abb)

url2 <- "https://www2.census.gov/programs-surveys/popest/tables/2010-2016/state/totals/nst-est2016-01.xlsx"

state\_populations <- readr::read\_csv("data/nst-est2016-01.csv")

write\_csv(data, "data/nst-est2016-01.csv")

# clean it a little

state\_populations <-

state\_populations %>%

mutate(state\_name = gsub("\\.", "", X\_\_1)) %>%

left\_join(state\_abb)

# compute deaths by state and as deaths per 1000 people in each state

by\_state16 <- data %>%

filter(Year == 2016) %>%

group\_by(State) %>%

tally() %>%

left\_join(state\_abb, by = c('State' = 'state\_abb')) %>%

filter(!is.na(state\_name)) %>%

left\_join(state\_populations) %>%

mutate(per\_n\_people = (n / `2016`) \* 1000000)