

NYPD_Shooting

S. May

6 12 2021

NYPD Shooting Incident Data Report

Import, tidy and analyze the dataset 'NYPD Shooting Incident Data (Historic)' from the source ><https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic><

Libraries

```
library(tidyverse)
library(lubridate)
```

1. reading in the data sets

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
```

2. importing

```
NYPD_Shooting <- read_csv(url_in)
```

3. tidying

- delete X_COORD_CD, Y_COORD_CD, Longitude, Latitude, Lon_Lat

```
NYPD_Shooting <- NYPD_Shooting %>%
  select(-c(X_COORD_CD, Y_COORD_CD, Longitude, Latitude, Lon_Lat))
```

4. transforming

- convert date column type from character to date
- prepare data for the model (total shootings & murders)

```
NYPD_Shooting <- NYPD_Shooting %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))

Total_Shootings_Yearly <- NYPD_Shooting %>%
  mutate(YEAR = year(round_date(OCCUR_DATE, "year"))) %>%
  group_by(YEAR) %>% count()

colnames(Total_Shootings_Yearly) <- c("YEAR", "Shootings")

Total_Murders_Yearly <- NYPD_Shooting %>%
  filter(STATISTICAL_MURDER_FLAG == TRUE) %>%
  mutate(YEAR = year(round_date(OCCUR_DATE, "year"))) %>%
  group_by(YEAR) %>% count()

colnames(Total_Murders_Yearly) <- c("YEAR", "Murders")

Yearly <- merge(Total_Shootings_Yearly, Total_Murders_Yearly, by="YEAR")
```

5. analyzing

- Number of shootings grouped by borough

```
NYPD_Shooting_Boro <- NYPD_Shooting %>%
  group_by(BORO) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
```

```
NYPD_Shooting_Boro
```

```
## # A tibble: 5 x 2
##   BORO          number_of_shootings
##   <chr>          <int>
## 1 BROOKLYN      9734
## 2 BRONX        6701
## 3 QUEENS       3532
## 4 MANHATTAN    2922
## 5 STATEN ISLAND 696
```

- Number of shootings grouped by victim's age group

```
NYPD_Shooting_Vic_Age_Group <- NYPD_Shooting %>%
  group_by(VIC_AGE_GROUP) %>%
  summarize(number_of_shootings=n())
```

```
NYPD_Shooting_Vic_Age_Group
```

```
## # A tibble: 6 x 2
##   VIC_AGE_GROUP number_of_shootings
##   <chr>          <int>
## 1 <18            2525
## 2 18-24         9003
## 3 25-44        10303
## 4 45-64         1541
## 5 65+           154
## 6 UNKNOWN       59
```

- Number of shootings grouped by victim's sex

```
NYPD_Shooting_Vic_Sex <- NYPD_Shooting %>%
  group_by(VIC_SEX) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
```

```
NYPD_Shooting_Vic_Sex
```

```
## # A tibble: 3 x 2
##   VIC_SEX number_of_shootings
##   <chr>          <int>
## 1 M            21370
## 2 F            2204
## 3 U              11
```

- Number of shootings grouped by victim's race

```
NYPD_Shooting_Vic_Race <- NYPD_Shooting %>%
  group_by(VIC_RACE) %>%
  summarize(number_of_shootings=n()) %>%
  arrange(desc(number_of_shootings))
```

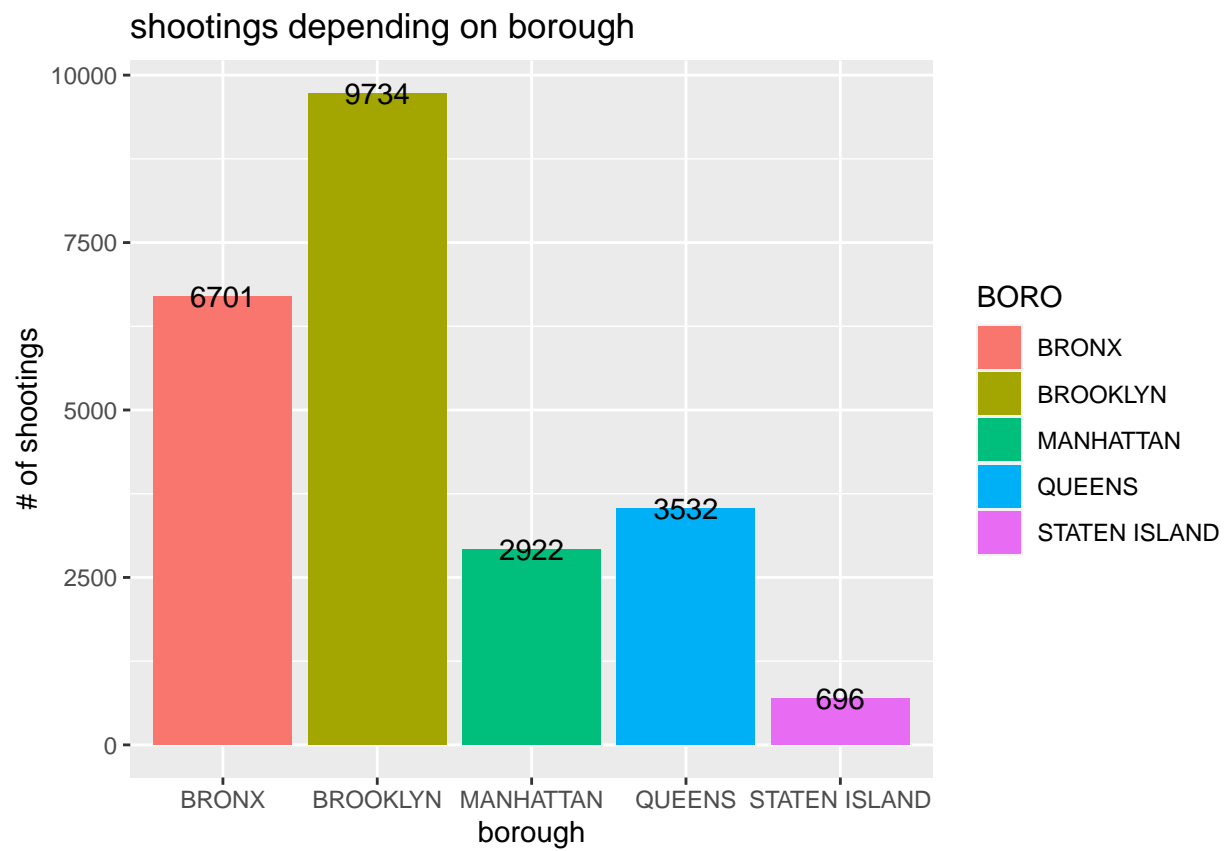
```
NYPD_Shooting_Vic_Race
```

```
## # A tibble: 7 x 2
##   VIC_RACE          number_of_shootings
##   <chr>          <int>
## 1 BLACK            16869
## 2 WHITE HISPANIC    3450
## 3 BLACK HISPANIC    2245
## 4 WHITE             620
## 5 ASIAN / PACIFIC ISLANDER 327
## 6 UNKNOWN           65
## 7 AMERICAN INDIAN/ALASKAN NATIVE 9
```

6. visualizing

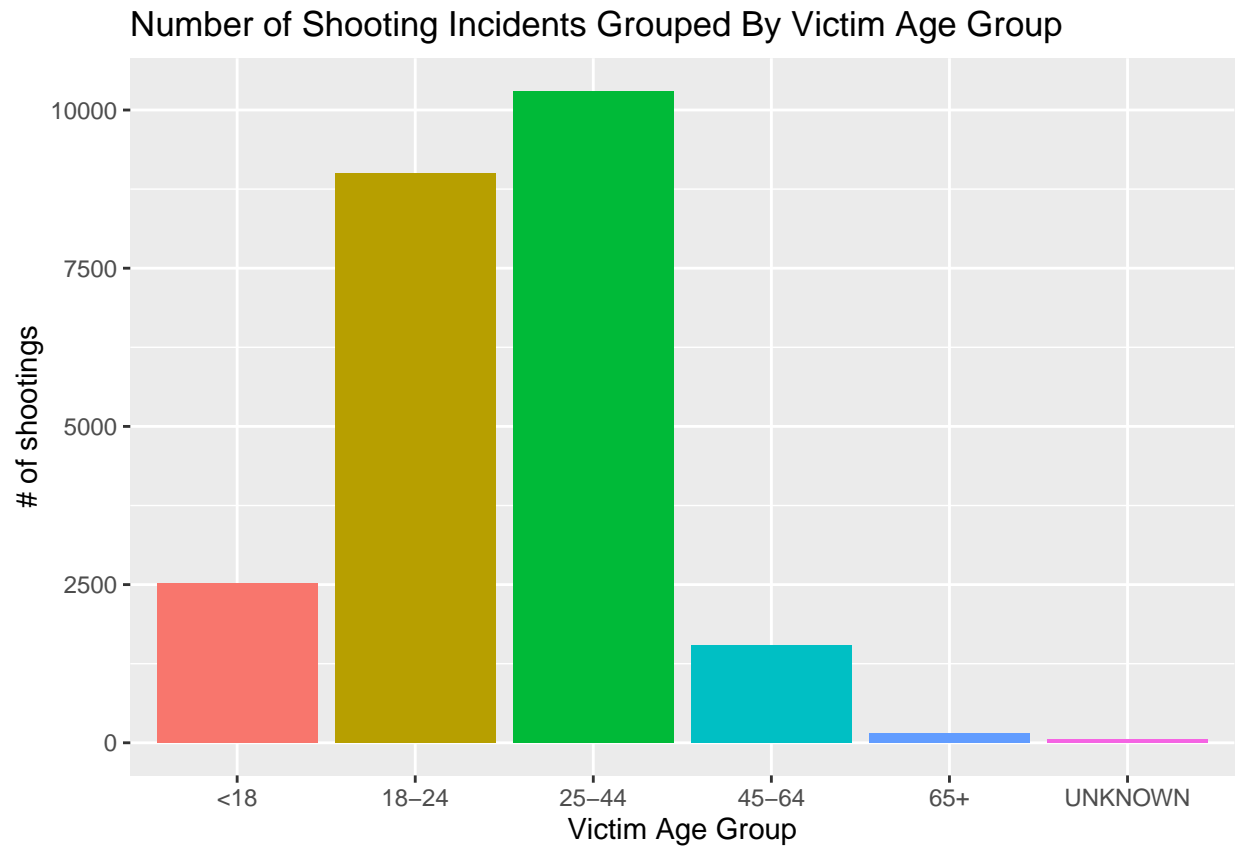
- Number of shootings depending on borough

```
NYPD_Shooting_Boro %>%  
  ggplot(aes(BORO,number_of_shootings))+  
  geom_col(aes(fill = BORO)) +  
  labs(title="shootings depending on borough",  
        y="# of shootings", x= "borough")+  
  geom_text(aes(label = round(number_of_shootings, 1)), nudge_y= -3, color="black")
```



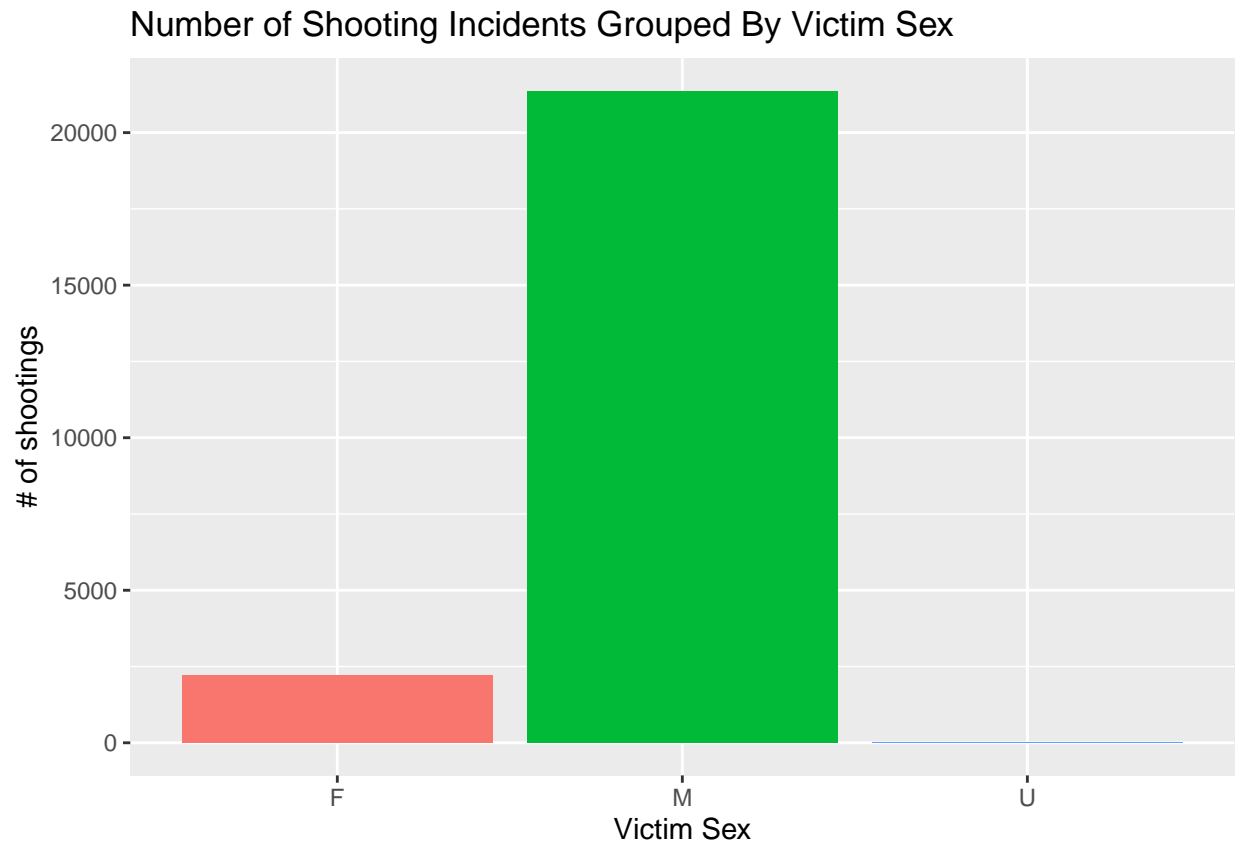
- Number of shootings depending on victim age group

```
NYPD_Shooting_Vic_Age_Group %>%  
  ggplot(aes(VIC_AGE_GROUP, y = number_of_shootings, fill = VIC_AGE_GROUP)) +  
  geom_bar(stat = "identity") +  
  ylab("# of shootings") +  
  xlab("Victim Age Group") +  
  ggtitle("Number of Shooting Incidents Grouped By Victim Age Group") +  
  theme(legend.position = "none")
```



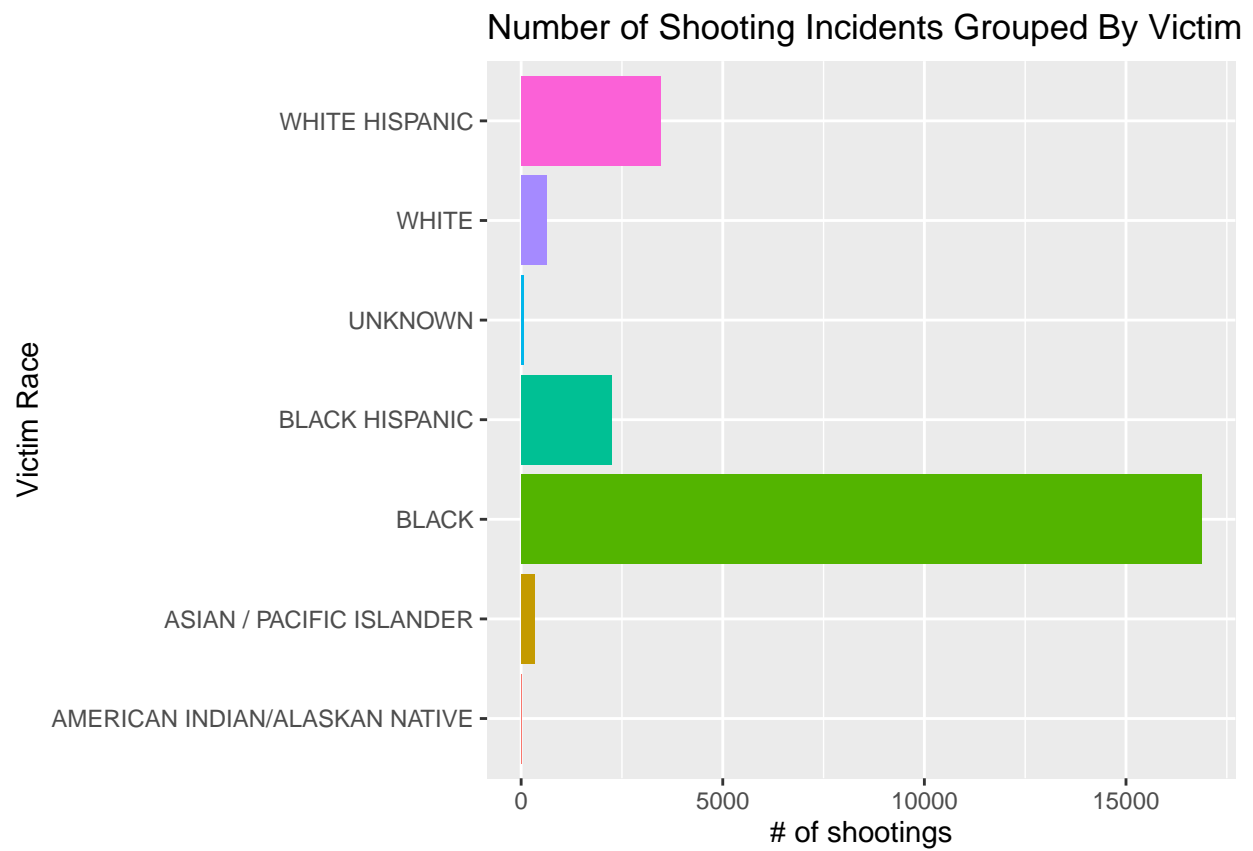
- Number of shootings depending on victim sex

```
NYPD_Shooting_Vic_Sex %>%  
  ggplot(aes(VIC_SEX, y = number_of_shootings, fill = VIC_SEX)) +  
  geom_bar(stat = "identity") +  
  ylab("# of shootings") +  
  xlab("Victim Sex") +  
  ggtitle("Number of Shooting Incidents Grouped By Victim Sex") +  
  theme(legend.position = "none")
```



- Number of shootings depending on victim race

```
NYPD_Shooting_Vic_Race %>%
  ggplot(aes(VIC_RACE, y = number_of_shootings, fill = VIC_RACE)) +
  coord_flip() +
  geom_bar(stat = "identity") +
  ylab("# of shootings") +
  xlab("Victim Race") +
  ggtitle("Number of Shooting Incidents Grouped By Victim Race") +
  theme(legend.position = "none")
```



7. modeling

Model the linear correlation between MURDERS and Shootings by using the `lm()` function.

```
mod <- lm(Shootings ~ Murders, data = Yearly)
summary(mod)
```

```
##
## Call:
## lm(formula = Shootings ~ Murders, data = Yearly)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -187.965  -76.484   -9.551   95.594  197.963
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   239.931     112.176   2.139   0.052 .
## Murders         4.441       0.360  12.338 1.5e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 117.6 on 13 degrees of freedom
## Multiple R-squared:  0.9213, Adjusted R-squared:  0.9153
## F-statistic: 152.2 on 1 and 13 DF,  p-value: 1.497e-08
```

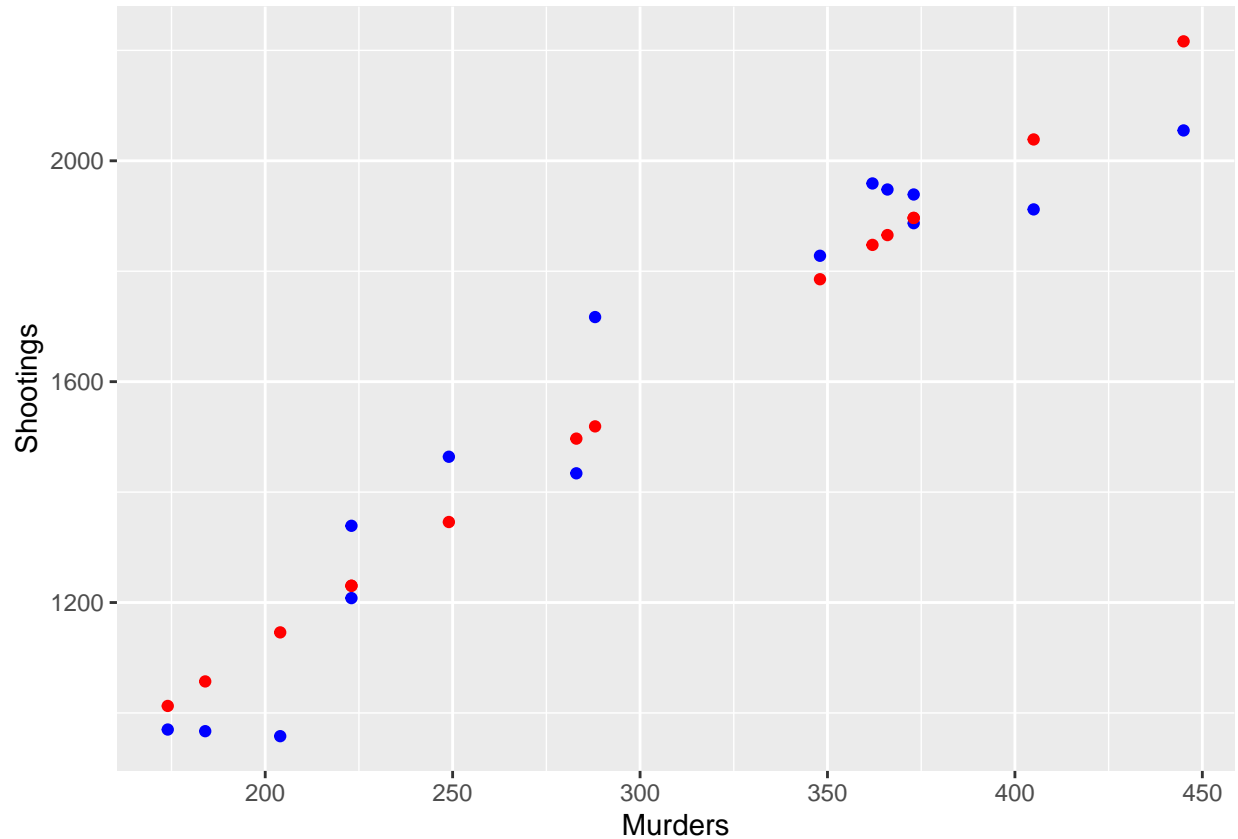
```
Yearly %>% slice_min(Shootings)
```

```
##   YEAR Shootings Murders
## 1 2018         958      204
```

```
Yearly %>% slice_max(Shootings)
```

```
##   YEAR Shootings Murders
## 1 2006         2055      445
```

```
Yearly_pred <- Yearly %>%
  mutate(pred = predict(mod))
Yearly_pred %>% ggplot() +
  geom_point(aes(x = Murders, y = Shootings), color = "blue") +
  geom_point(aes(x = Murders, y = pred), color = "red")
```

8. Conclusion & Identifying Bias

To conclude, the data showed us that, in absolute numbers, Brooklyn has the highest number of shootings. Probably, we have to put this number in relation to borough citizens to show which borough is the “most dangerous”.

We are able to see that the majority of victims are black, male and / or between 18 - 44 years old.

My personal bias regarding this topic was influenced by news and I thought most shooting incidents would occur in the Bronx along with thinking most shooting incidents would be male.

Session Info

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Germany.1252 LC_CTYPE=German_Germany.1252
## [3] LC_MONETARY=German_Germany.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] lubridate_1.8.0 forcats_0.5.1  stringr_1.4.0  dplyr_1.0.7
## [5] purrr_0.3.4     readr_2.1.1    tidyr_1.1.4    tibble_3.1.6
## [9] ggplot2_3.3.5   tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.7      assertthat_0.2.1 digest_0.6.29  utf8_1.2.2
## [5] R6_2.5.1        cellranger_1.1.0 backports_1.3.0 reprex_2.0.1
## [9] evaluate_0.14   highr_0.9       httr_1.4.2     pillar_1.6.4
## [13] rlang_0.4.12    curl_4.3.2      readxl_1.3.1   rstudioapi_0.13
## [17] rmarkdown_2.11 labeling_0.4.2   bit_4.0.4      munsell_0.5.0
## [21] broom_0.7.10    compiler_4.1.2  modelr_0.1.8    xfun_0.28
## [25] pkgconfig_2.0.3 htmltools_0.5.2 tidyselect_1.1.1 fansi_0.5.0
## [29] crayon_1.4.2    tzdb_0.2.0      dbplyr_2.1.1   withr_2.4.3
## [33] grid_4.1.2      jsonlite_1.7.2  gtable_0.3.0   lifecycle_1.0.1
## [37] DBI_1.1.1       magrittr_2.0.1  scales_1.1.1   cli_3.1.0
## [41] stringi_1.7.6   vroom_1.5.7     farver_2.1.0   fs_1.5.1
## [45] xml2_1.3.3      ellipsis_0.3.2  generics_0.1.1 vctrs_0.3.8
## [49] tools_4.1.2     bit64_4.0.5     glue_1.5.1     hms_1.1.1
## [53] parallel_4.1.2 fastmap_1.1.0   yaml_2.2.1     colorspace_2.0-2
## [57] rvest_1.0.2     knitr_1.36      haven_2.4.3
```