

July 13, 2023

The results below are generated from an R script.

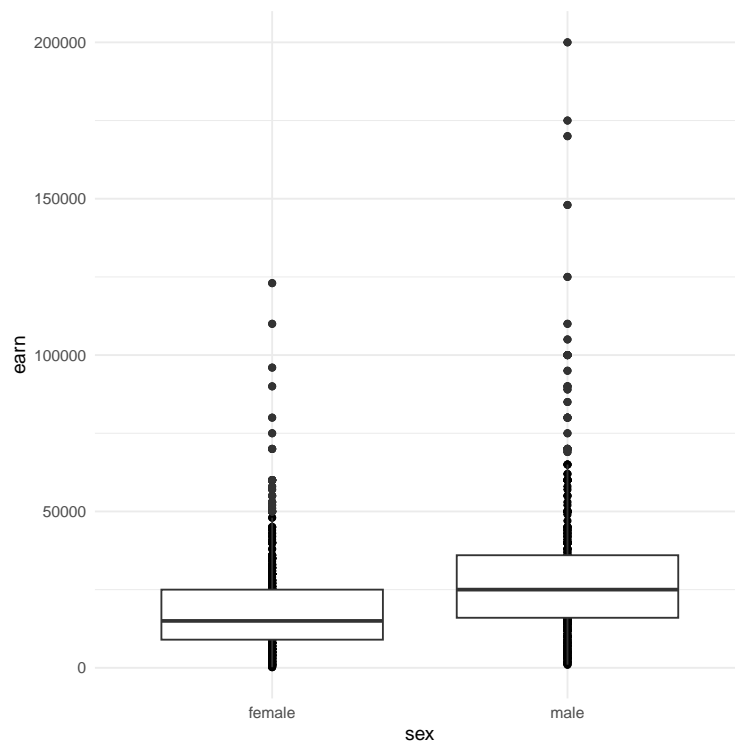
```
# Assignment: ASSIGNMENT 4
# Name: Tang, Xin
# Date: 2023-07-08

## Load the ggplot2 package
library(ggplot2)
theme_set(theme_minimal())

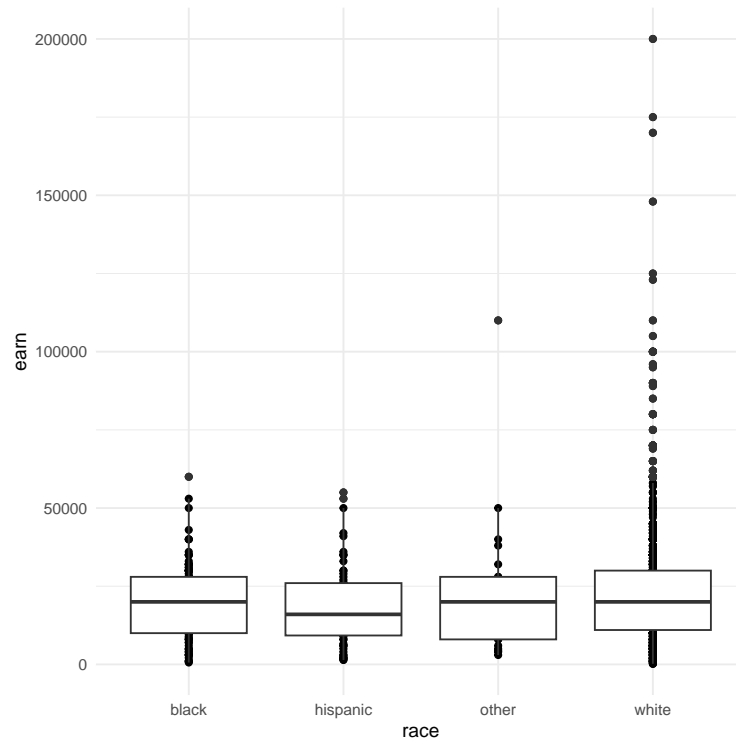
## Set the working directory to the root of your DSC 520 directory
setwd("~/dsc520")

## Load the 'data/r4ds/heights.csv' to
heights_df <- read.csv("data/r4ds/heights.csv")
View(heights_df)

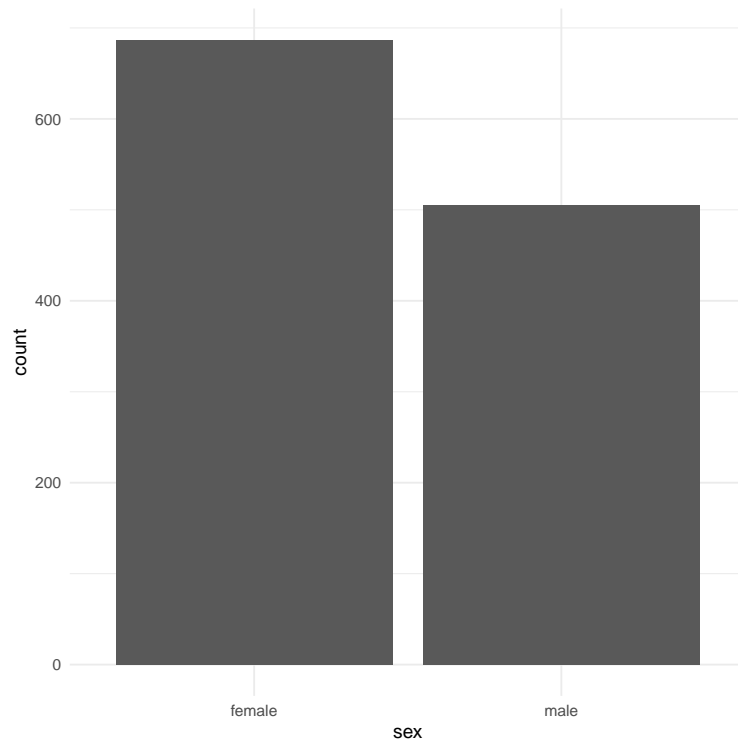
# https://ggplot2.tidyverse.org/reference/geom_boxplot.html
## Create boxplots of sex vs. earn and race vs. earn using 'geom_point()' and 'geom_boxplot()'
## sex vs. earn
ggplot(heights_df, aes(x=sex, y=earn)) + geom_point() + geom_boxplot()
```



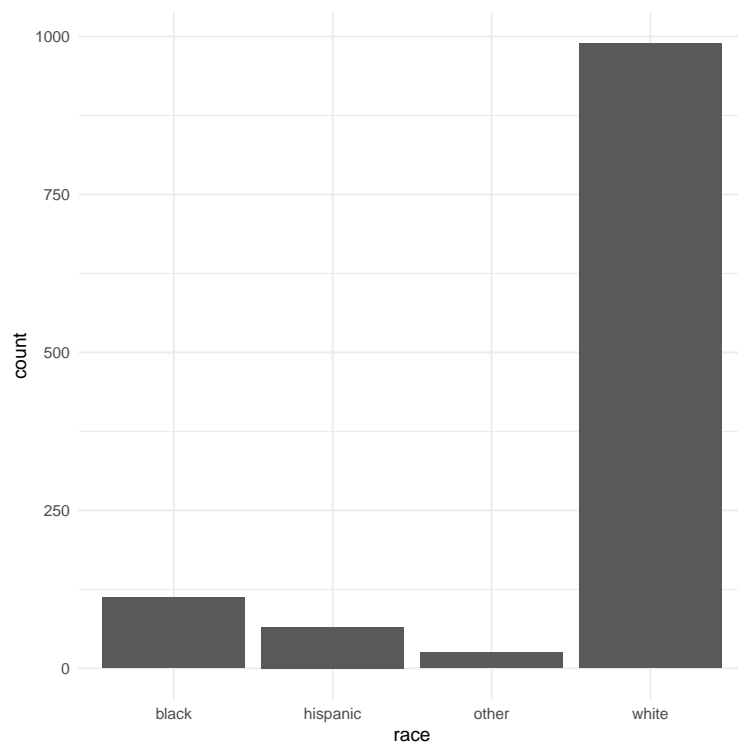
```
## race vs. earn
ggplot(heights_df, aes(x=race, y=earn)) + geom_point()+ geom_boxplot()
```



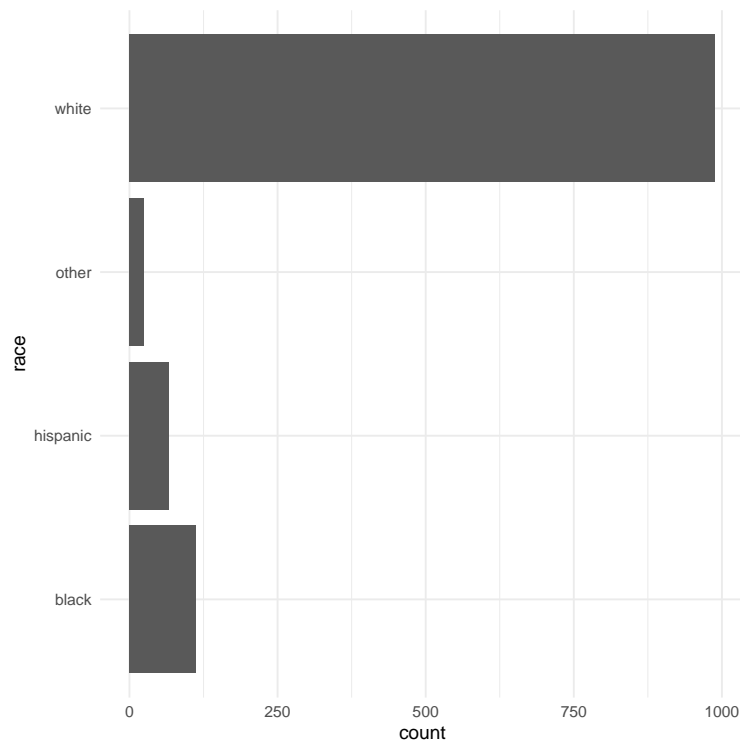
```
# https://ggplot2.tidyverse.org/reference/geom\_bar.html
## Using 'geom_bar()' plot a bar chart of the number of records for each 'sex'
ggplot(heights_df, aes(x=sex)) + geom_bar()
```



```
## Using 'geom_bar()' plot a bar chart of the number of records for each race  
ggplot(heights_df, aes(x=race)) + geom_bar()
```



```
## Create a horizontal bar chart by adding 'coord_flip()' to the previous plot
ggplot(heights_df, aes(x=race)) + geom_bar() + coord_flip()
```



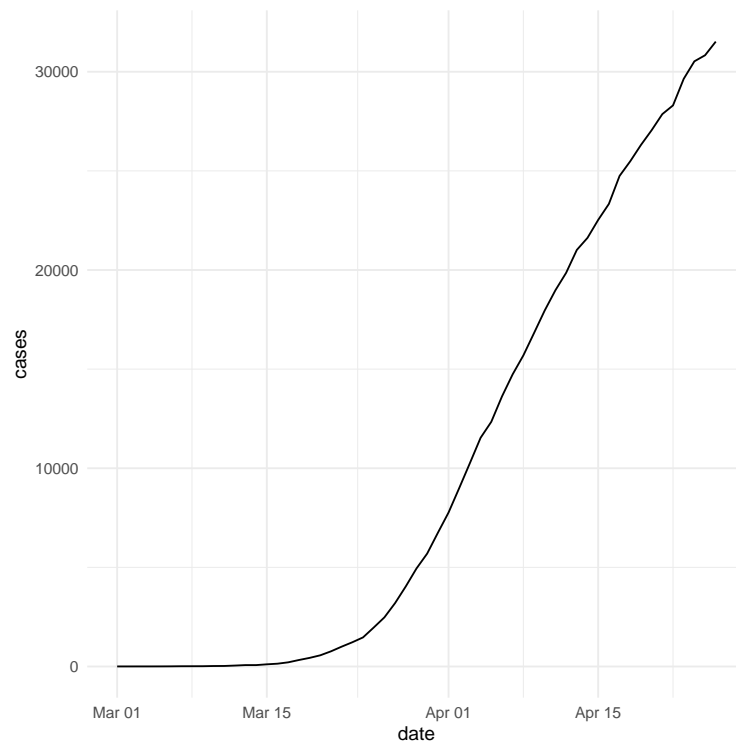
```
# https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0/topics/geom\_path
## Load the file "data/nytimes/covid-19-data/us-states.csv" and
## assign it to the 'covid_df' dataframe
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv")
View(covid_df)
```

```
## Parse the date column using 'as.Date()'
covid_df$date <- as.Date(covid_df$date)
covid_df$date
```

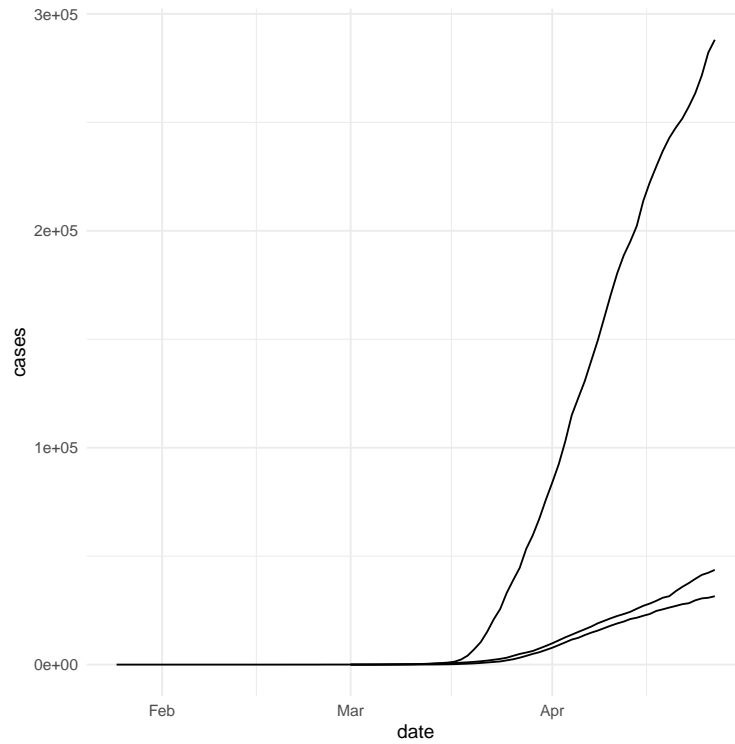
```
##      [1] "2020-01-21" "2020-01-22" "2020-01-23" "2020-01-24" "2020-01-24" "2020-01-25"
##      [7] "2020-01-25" "2020-01-25" "2020-01-26" "2020-01-26" "2020-01-26" "2020-01-26"
##     [13] "2020-01-27" "2020-01-27" "2020-01-27" "2020-01-27" "2020-01-28" "2020-01-28"
##     [19] "2020-01-28" "2020-01-28" "2020-01-29" "2020-01-29" "2020-01-29" "2020-01-29"
##     [25] "2020-01-30" "2020-01-30" "2020-01-30" "2020-01-30" "2020-01-31" "2020-01-31"
##     [31] "2020-01-31" "2020-01-31" "2020-02-01" "2020-02-01" "2020-02-01" "2020-02-01"
##     [37] "2020-02-01" "2020-02-02" "2020-02-02" "2020-02-02" "2020-02-02" "2020-02-02"
##     [43] "2020-02-03" "2020-02-03" "2020-02-03" "2020-02-03" "2020-02-03" "2020-02-04"
##     [49] "2020-02-04" "2020-02-04" "2020-02-04" "2020-02-04" "2020-02-05" "2020-02-05"
##     [55] "2020-02-05" "2020-02-05" "2020-02-05" "2020-02-05" "2020-02-06" "2020-02-06"
##     [61] "2020-02-06" "2020-02-06" "2020-02-06" "2020-02-06" "2020-02-07" "2020-02-07"
##     [67] "2020-02-07" "2020-02-07" "2020-02-07" "2020-02-07" "2020-02-08" "2020-02-08"
##     [73] "2020-02-08" "2020-02-08" "2020-02-08" "2020-02-08" "2020-02-09" "2020-02-09"
##     [79] "2020-02-09" "2020-02-09" "2020-02-09" "2020-02-09" "2020-02-10" "2020-02-10"
##     [85] "2020-02-10" "2020-02-10" "2020-02-10" "2020-02-10" "2020-02-11" "2020-02-11"
##     [91] "2020-02-11" "2020-02-11" "2020-02-11" "2020-02-11" "2020-02-12" "2020-02-12"
```



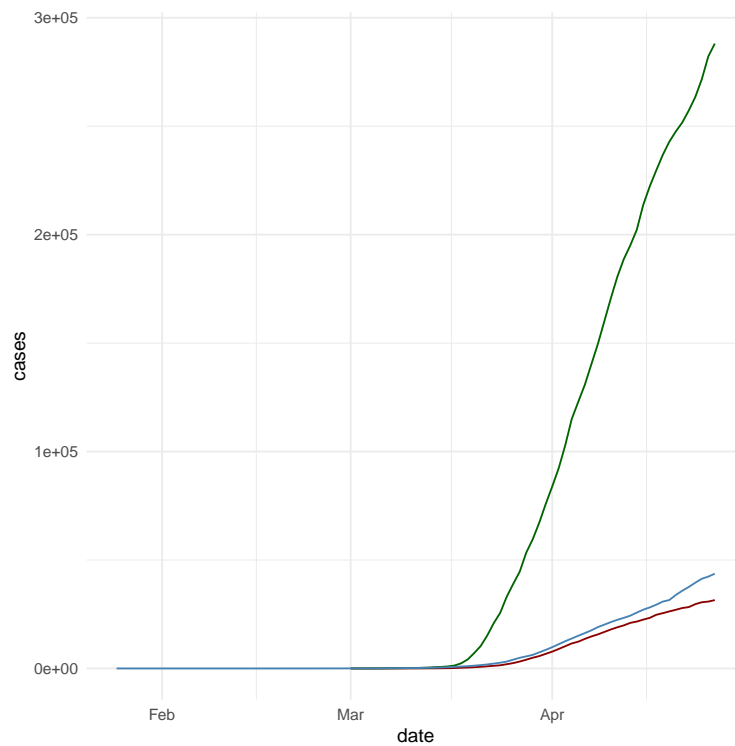
```
ggplot(data=florida_df, aes(x=date, y=cases, group=1)) + geom_line()
```



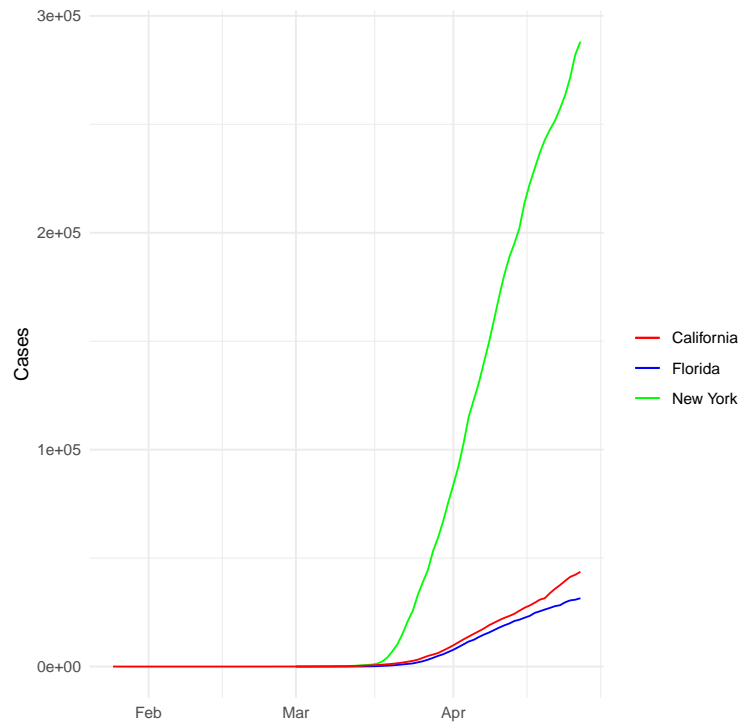
```
## Add lines for New York and California to the plot
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases)) +
  geom_line(data=ny_df, aes(y = cases)) +
  geom_line(data=california_df, aes(y = cases))
```

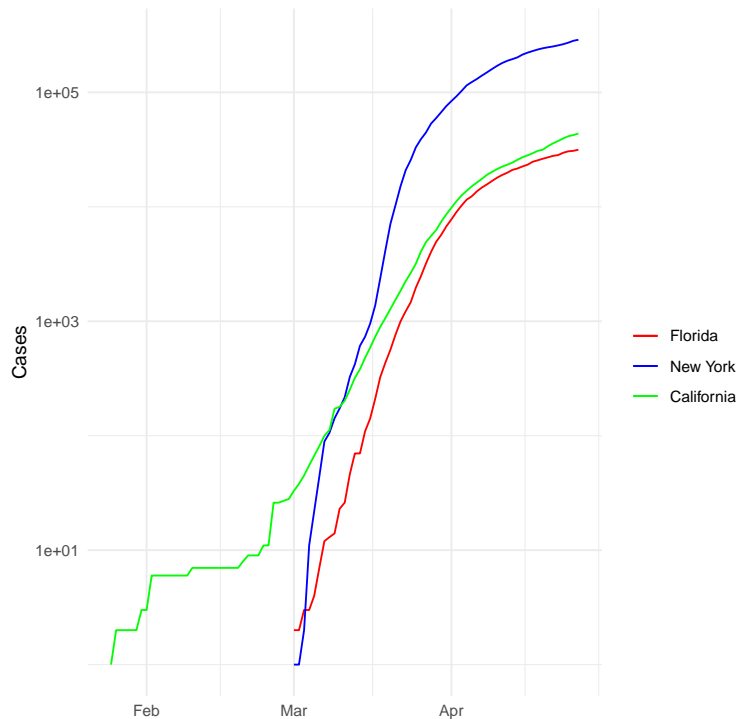
```
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases), color = 'darkred') +
  geom_line(data=ny_df, aes(y = cases), color='darkgreen') +
  geom_line(data=california_df, aes(y = cases), color='steelblue')
```



```
## Add a legend to the plot using 'scale_colour_manual'
## Add a blank (" ") label to the x-axis and the label "Cases" to the y axis
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases, colour = "Florida")) +
  geom_line(data=ny_df, aes(y = cases, colour="New York")) +
  geom_line(data=california_df, aes(y = cases, colour="California")) +
  scale_colour_manual("",
    breaks = waiver(),
    values = c('red', 'blue', 'green')) +
  xlab(" ") + ylab("Cases")
```



```
## Scale the y axis using 'scale_y_log10()'
ggplot(data=florida_df, aes(x=date, group=1)) +
  geom_line(aes(y = cases, colour = "Florida")) +
  geom_line(data=ny_df, aes(y = cases, colour="New York")) +
  geom_line(data=california_df, aes(y = cases, colour="California")) +
  scale_colour_manual("",
    breaks = c('Florida', 'New York', 'California'),
    values = c('red', 'blue', 'green')) +
  xlab(" ") + ylab("Cases") + scale_y_log10()
```



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()

## R version 4.3.1 (2023-06-16 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8  LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/Chicago
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.4.2
##
## loaded via a namespace (and not attached):
## [1] vctrs_0.6.2      cli_3.6.1      knitr_1.43      rlang_1.1.1     xfun_0.39
## [6] highr_0.10       generics_0.1.3 glue_1.6.2      labeling_0.4.2  colorspace_2.1-0
## [11] scales_1.2.1     fansi_1.0.4    grid_4.3.1      munsell_0.5.0   evaluate_0.21
## [16] tibble_3.2.1     lifecycle_1.0.3 compiler_4.3.1  dplyr_1.1.2     pkgconfig_2.0.3
## [21] rstudioapi_0.14  farver_2.1.1   R6_2.5.1        tidyselect_1.2.0 utf8_1.2.3
```

```
## [26] pillar_1.9.0      magrittr_2.0.3    tools_4.3.1      withr_2.5.0      gtable_0.3.3
Sys.time()
## [1] "2023-07-13 15:16:21 CDT"
```