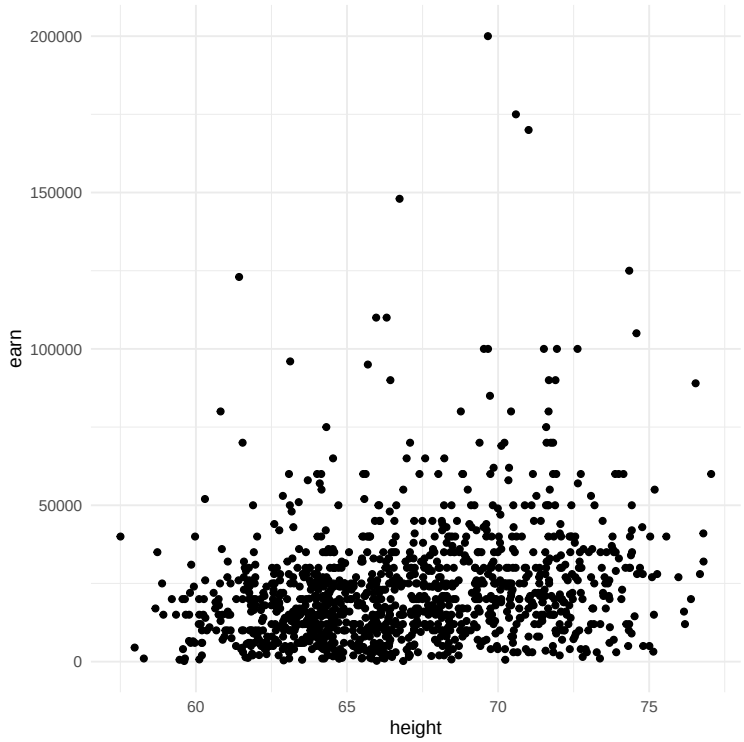
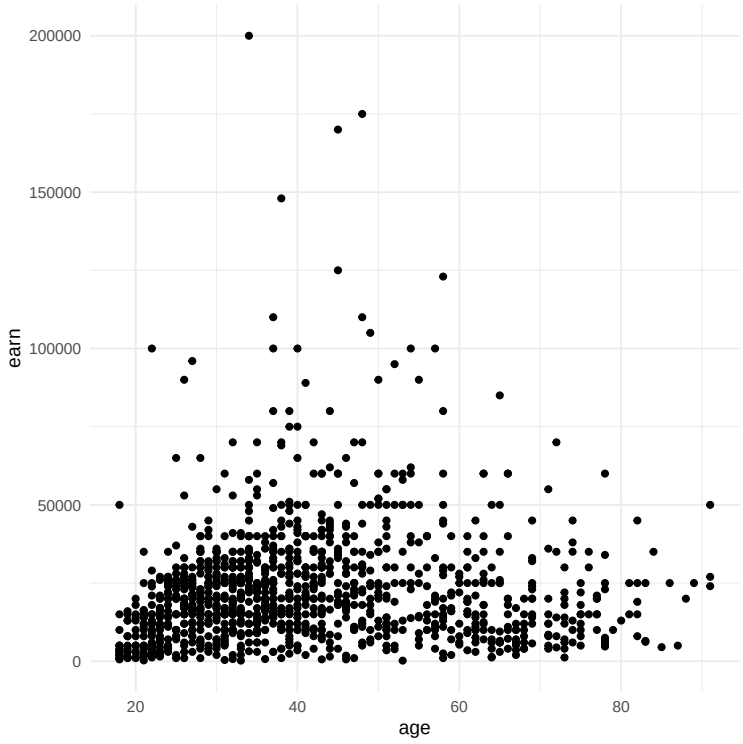
The results below are generated from an R script.

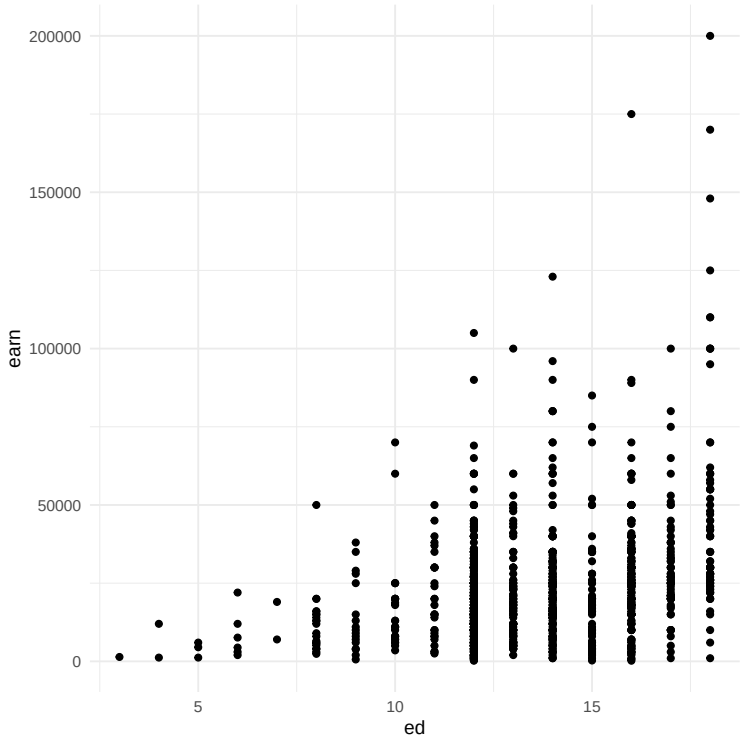
```
# Assignment: ASSIGNMENT 3
# Name: Chattapadhyay, Kausik
# Date: 2022-09-13

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

## Set the working directory to the root of your DSC 520 directory
setwd("/Users/kausik/desktop/MS Data Science/DSC 520/dsc520-stats-r-assignments")

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

# https://ggplot2.tidyverse.org/reference/geom_point.html
## Using `geom_point()` create three scatterplots for
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y=earn)) + geom_point()
```



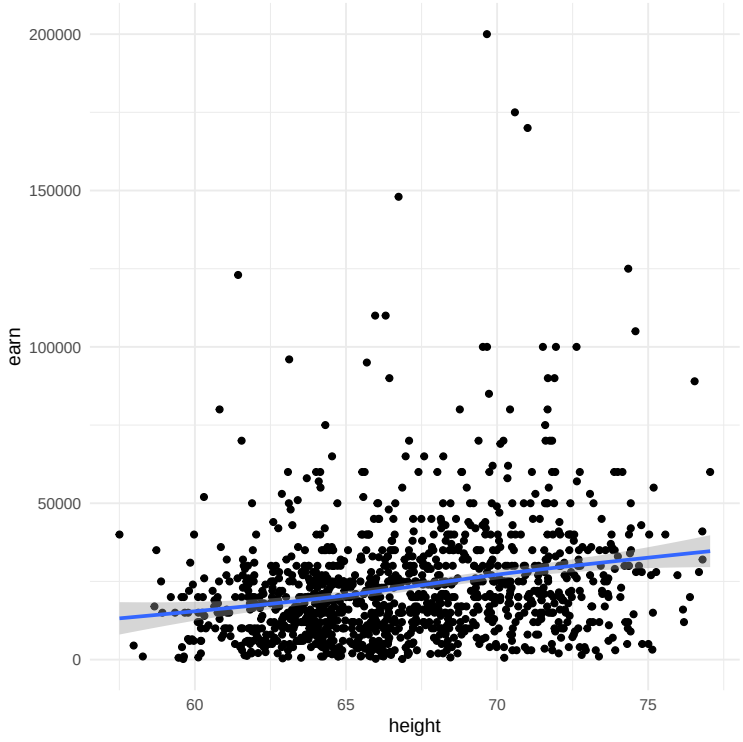

```
## `age` vs. `earn`
ggplot(heights_df, aes(x=age, y=earn)) + geom_point()
```



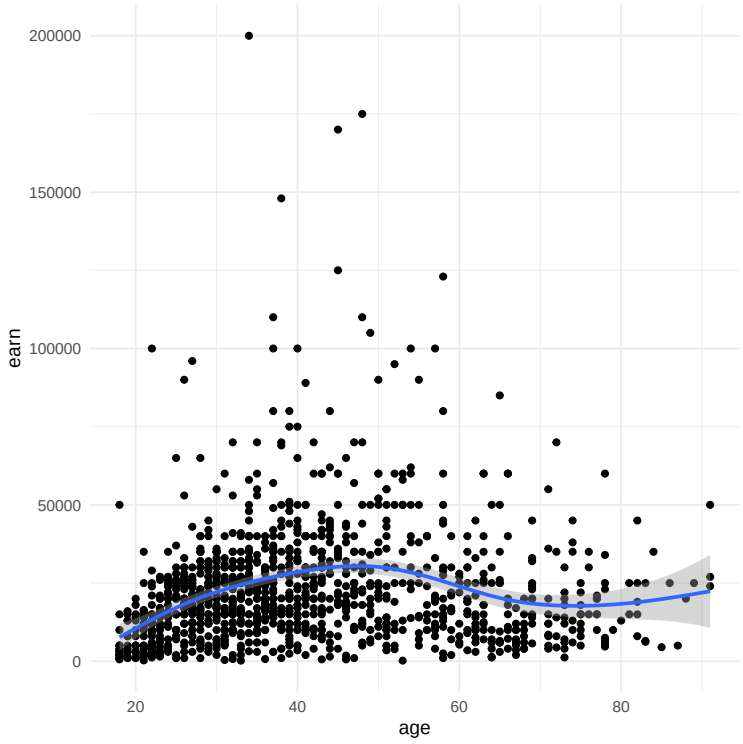

```
## `ed` vs. `earn`
ggplot(heights_df, aes(x=ed, y=earn)) + geom_point()
```



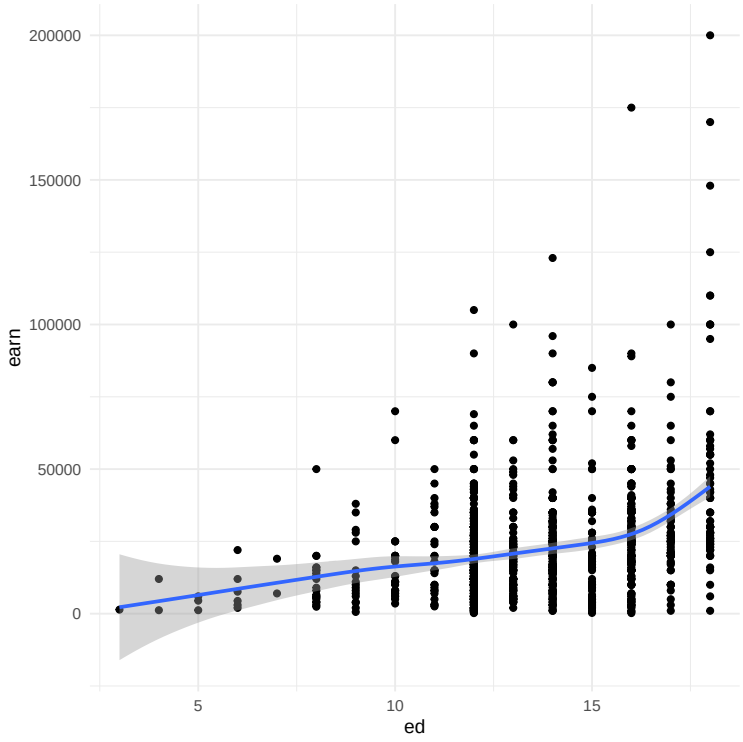
```
## Re-create the three scatterplots and add a regression trend line using
## the `geom_smooth()` function
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y=earn)) + geom_point() + geom_smooth()

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



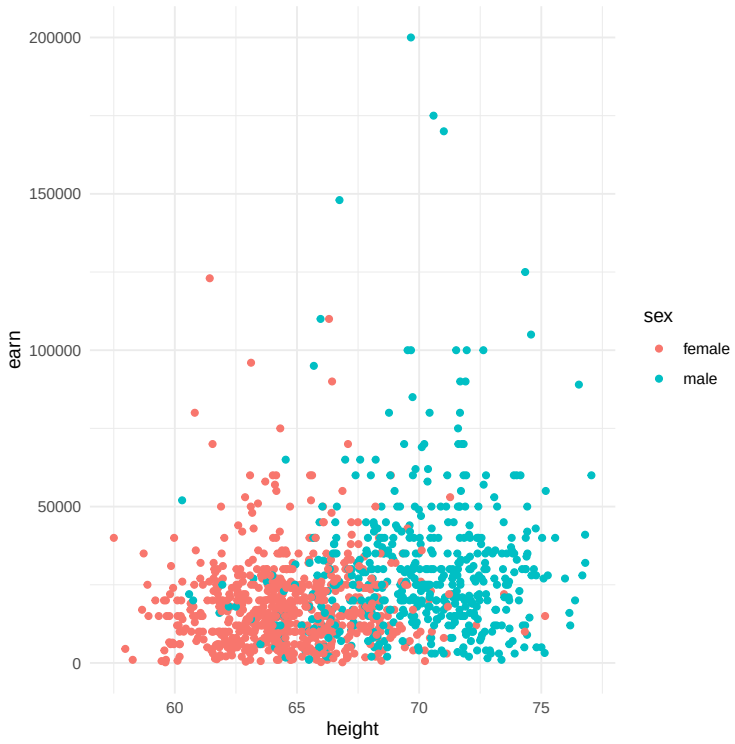

```
## `age` vs. `earn`
ggplot(heights_df, aes(x=age, y=earn)) + geom_point() + geom_smooth()

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



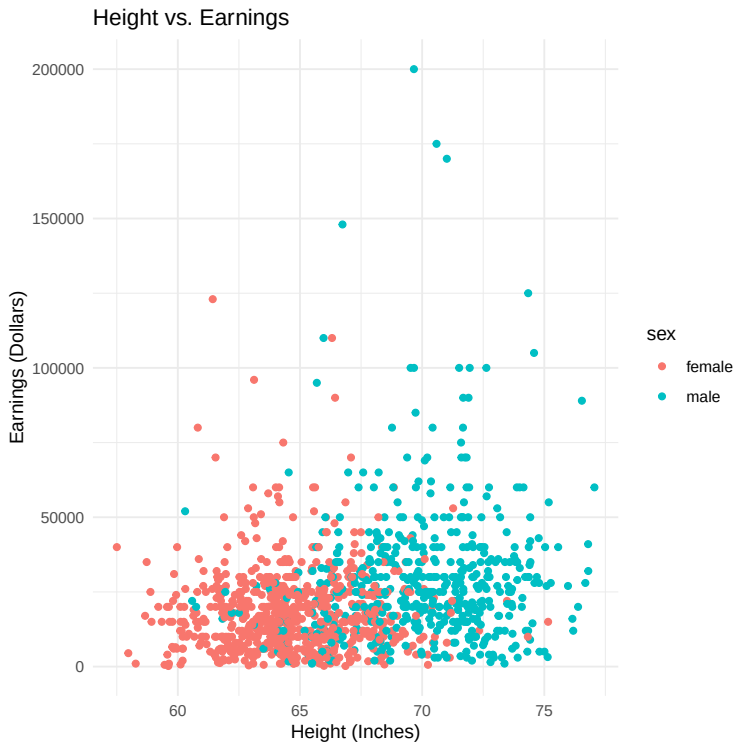

```
## `ed` vs. `earn`
ggplot(heights_df, aes(x=ed, y=earn)) + geom_point() + geom_smooth()

## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



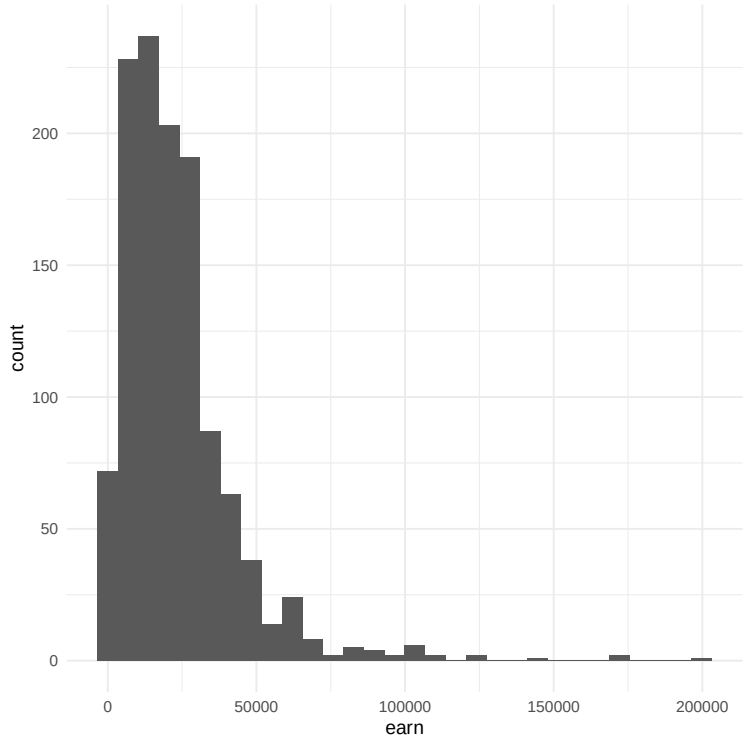
```
## Create a scatterplot of `height`` vs. `earn`.  Use `sex` as the `col` (color) attribute
ggplot(heights_df, aes(x=height, y=earn, col=sex)) + geom_point()
```



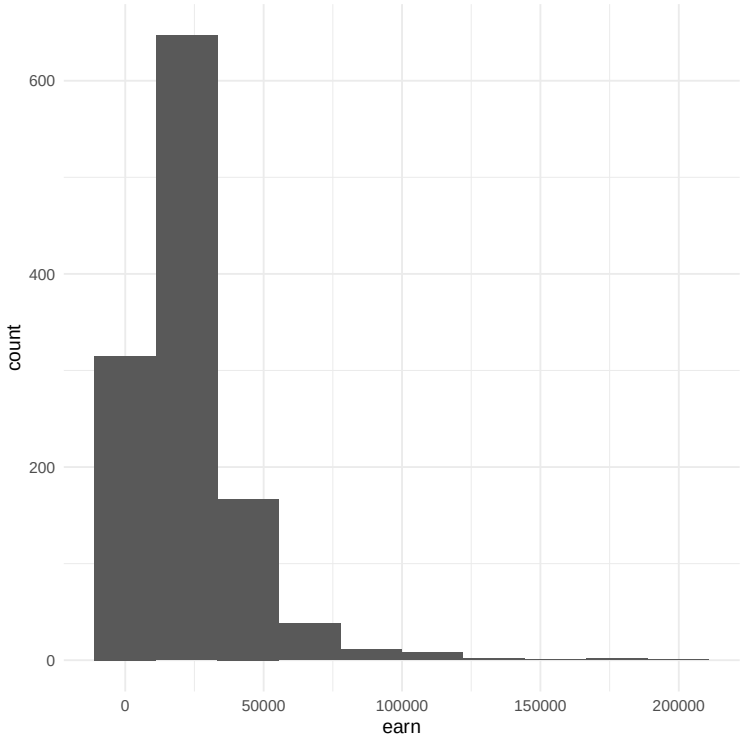

```
## Using `ggtitle()`, `xlab()`, and `ylab()` to add a title, x label, and y label to the previous plot
## Title: Height vs. Earnings
## X label: Height (Inches)
## Y Label: Earnings (Dollars)
ggplot(heights_df, aes(x=height, y=earn, col=sex)) + geom_point() +
       ggtitle("Height vs. Earnings") + xlab("Height (Inches)")  +
       ylab("Earnings (Dollars)")
```



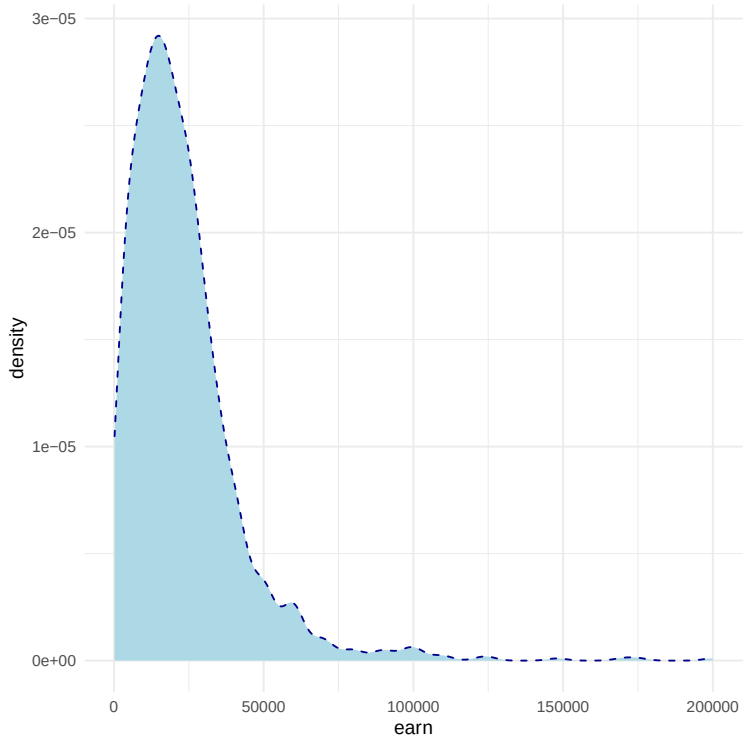

```
# https://ggplot2.tidyverse.org/reference/geom_histogram.html
## Create a histogram of the `earn` variable using `geom_histogram()`
ggplot(heights_df, aes(earn)) + geom_histogram()

## `stat_bin()` using `bins = 30`.  Pick better value with `binwidth`.
```



```r
## Create a histogram of the `earn` variable using `geom_histogram()`
## Use 10 bins
ggplot(heights_df, aes(earn)) + geom_histogram(bins = 10)
```




```r
# https://ggplot2.tidyverse.org/reference/geom_density.html
## Create a kernel density plot of `earn` using `geom_density()`
ggplot(heights_df, aes(earn)) +  geom_density(color="darkblue", fill="lightblue", linetype="dashed")
```



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

```
sessionInfo()
```

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggthemes_4.2.4  scales_1.2.1    lubridate_1.8.0 ggplot2_3.3.6   rjson_0.2.21
## [6] jsonlite_1.8.0  DBI_1.1.3       readxl_1.4.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.8.3      cellranger_1.1.0  pillar_1.8.1      compiler_4.0.0
##  [5] highr_0.9         tools_4.0.0       digest_0.6.29     bit_4.0.4
##  [9] lattice_0.20-45   nlme_3.1-147      viridisLite_0.4.1 RSQLite_2.2.12
## [13] memoise_2.0.1     evaluate_0.16     lifecycle_1.0.1   tibble_3.1.6
## [17] gtable_0.3.1      mgcv_1.8-40       pkgconfig_2.0.3   rlang_1.0.2
## [21] Matrix_1.4-1      cli_3.2.0         rstudioapi_0.14   xfun_0.30
## [25] fastmap_1.1.0     withr_2.5.0       stringr_1.4.1     knitr_1.40
## [29] generics_0.1.3    vctrs_0.4.0       bit64_4.0.5       grid_4.0.0
```

```
## [33] glue_1.6.2        R6_2.5.1          fansi_1.0.3       purrr_0.3.4
## [37] farver_2.1.0      blob_1.2.3        magrittr_2.0.3    splines_4.0.0
## [41] ellipsis_0.3.2    colorspace_2.0-3  labeling_0.4.2    utf8_1.2.2
## [45] tinytex_0.41      stringi_1.7.6     munsell_0.5.0     cachem_1.0.6
## [49] crayon_1.5.1

Sys.time()

## [1] "2022-09-11 16:10:43 EDT"
```