The results below are generated from an R script.

```
# Load the ggplot2 package
# Set the working directory to the root of your DSC 520 directory
# Load the `data/r4ds/heights.csv` to

library(ggplot2)
theme_set(theme_minimal())
setwd("C:/Users/ait0s/OneDrive/Documents/GitHub/dsc520")
heights_df <- read.csv("data/r4ds/heights.csv")
summary(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(sex)) + geom_bar()


# Using `geom_bar()` plot a bar chart of the number of records for each race


ggplot(heights_df, aes(race)) + geom_bar()


# Create a horizontal bar chart by adding `coord_flip()` to the previous plot


ggplot(heights_df, aes(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
```

```r
covid_df <- read.csv("data/nytimes/covid-19-data/us-states.csv")

head(covid_df)


# Parse the date column using `as.Date()`


covid_df$date <- as.Date(covid_df$date)


# Create three dataframes named `california_df`, `ny_df`, and `florida_df`
# containing the data from California, New York, and Florida

California

california_df <- covid_df[ which( covid_df$state == "California"), ]

summary(california_df)


# New York

ny_df <- covid_df[ which( covid_df$state == "New York"), ]

summary(ny_df)


# Florida


florida_df <- covid_df[ which( covid_df$state == "Florida"), ]

summary(florida_df)


# Plot the number of cases in Florida using `geom_line()`


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 = c("Florida", "New York", "California"),
                      values = c("darkred", "darkgreen", "steelblue")) +
  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("darkred", "darkgreen", "steelblue")) +
  xlab(" ") + ylab("Cases") + scale_y_log10()
```

```
## Error:  <text>:96:5:  unexpected symbol
## 95:
## 96:  Add lines
##         ^
```

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

```
sessionInfo()
```

```
## R version 4.3.0 (2023-04-21 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 11 x64 (build 22621)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=English_United States.utf8  LC_CTYPE=C
## [3] LC_MONETARY=English_United States.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## time zone: America/New_York
```

```
## 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       DBI_1.1.3        generics_0.1.3   labeling_0.4.2   glue_1.6.2
## [11] bit_4.0.5        colorspace_2.1-0 htmltools_0.5.5  fansi_1.0.4      rsconnect_0.8.29
## [16] scales_1.2.1     rmarkdown_2.23   grid_4.3.0       pander_0.6.5     tibble_3.2.1
## [21] munsell_0.5.0    evaluate_0.21    fastmap_1.1.1    yaml_2.3.7       lifecycle_1.0.3
## [26] memoise_2.0.1    compiler_4.3.0   dplyr_1.1.2      RSQLite_2.3.1    blob_1.2.4
## [31] pkgconfig_2.0.3  Rcpp_1.0.10      rstudioapi_0.14  farver_2.1.1     digest_0.6.31
## [36] R6_2.5.1         tidyselect_1.2.0 utf8_1.2.3       pillar_1.9.0     magrittr_2.0.3
## [41] withr_2.5.0      tools_4.3.0      bit64_4.0.5      gtable_0.3.3     cachem_1.0.8
```

```
Sys.time()
```

```
## [1] "2023-07-20 01:08:06 EDT"
```