STA 141 Worksheet 4

Richard McCormick

September 28, 2023

Due Date: Thursday, October 12, 2023 before 11:00am.

Instructions

Worksheets must be turned in as a PDF file through Canvas. The worksheet is worth a total of **15 points**, which is 3 percent of your overall grade.

Exercises

Begin by running the following code block to add the packages we need to use to our library.

Exercise 1

(a) The first dataset we are going to work with comes built-in with the dpylr package. Run the following to save a copy of the starwars dataset to a variable called my.starwars.

```
my.starwars <- starwars
```

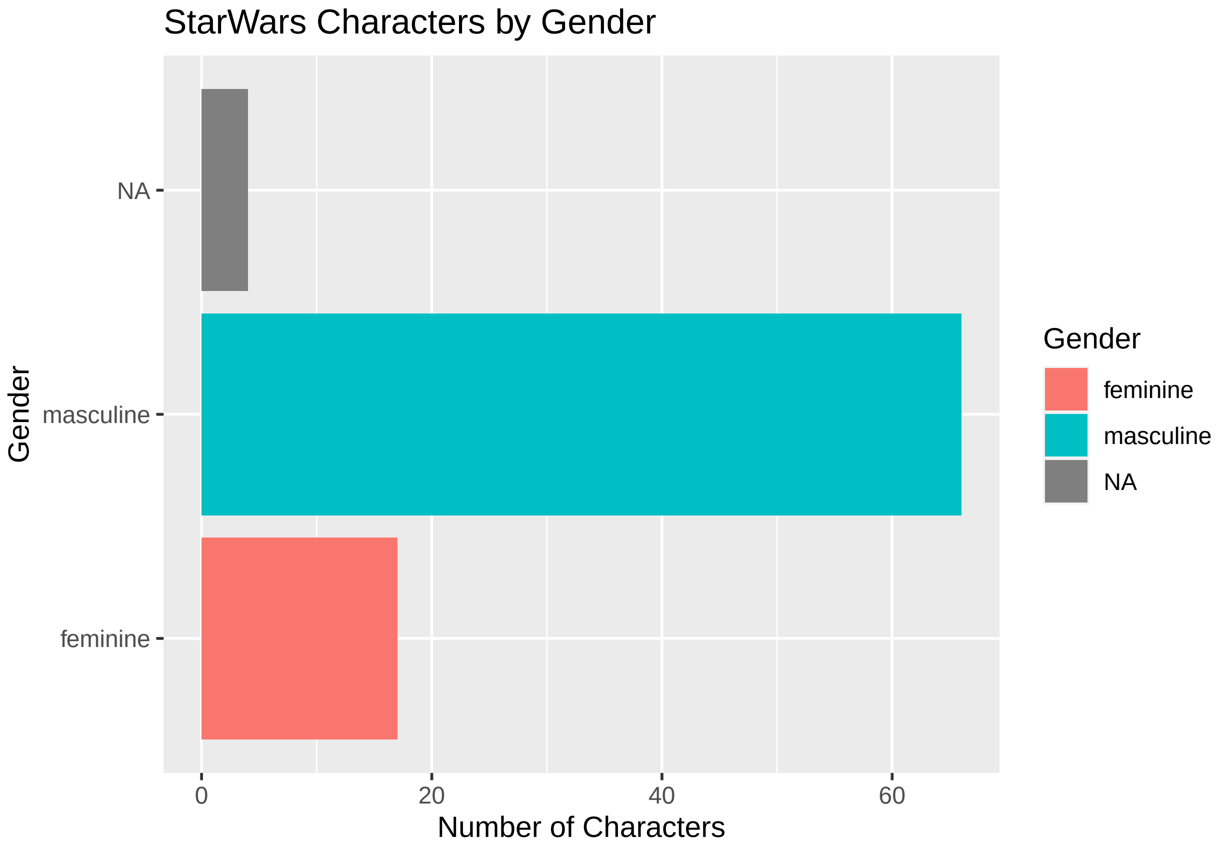
Use the following code block to view the structure of the data.

```
str( my.starwars )
```

```
## tibble [87 x 14] (S3: tbl_df/tbl/data.frame)
##  $ name      : chr [1:87] "Luke Skywalker" "C-3PO" "R2-D2" "Darth Vader" ...
##  $ height    : int [1:87] 172 167 96 202 150 178 165 97 183 182 ...
##  $ mass      : num [1:87] 77 75 32 136 49 120 75 32 84 77 ...
##  $ hair_color: chr [1:87] "blond" NA NA "none" ...
##  $ skin_color: chr [1:87] "fair" "gold" "white, blue" "white" ...
##  $ eye_color : chr [1:87] "blue" "yellow" "red" "yellow" ...
##  $ birth_year: num [1:87] 19 112 33 41.9 19 52 47 NA 24 57 ...
##  $ sex       : chr [1:87] "male" "none" "none" "male" ...
##  $ gender    : chr [1:87] "masculine" "masculine" "masculine" "masculine" ...
##  $ homeworld : chr [1:87] "Tatooine" "Tatooine" "Naboo" "Tatooine" ...
##  $ species   : chr [1:87] "Human" "Droid" "Droid" "Human" ...
##  $ films     :List of 87
##   ..$ : chr [1:5] "The Empire Strikes Back" "Revenge of the Sith" "Return of the Jedi" "A N
##   ..$ : chr [1:6] "The Empire Strikes Back" "Attack of the Clones" "The Phantom Menace" "Re
##   ..$ : chr [1:7] "The Empire Strikes Back" "Attack of the Clones" "The Phantom Menace" "Re
##   ..$ : chr [1:4] "The Empire Strikes Back" "Revenge of the Sith" "Return of the Jedi" "A N
##   ..$ : chr [1:5] "The Empire Strikes Back" "Revenge of the Sith" "Return of the Jedi" "A N
##   ..$ : chr [1:3] "Attack of the Clones" "Revenge of the Sith" "A New Hope"
##   ..$ : chr [1:3] "Attack of the Clones" "Revenge of the Sith" "A New Hope"
##   ..$ : chr "A New Hope"
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:6] "The Empire Strikes Back" "Attack of the Clones" "The Phantom Menace" "Re
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:2] "Revenge of the Sith" "A New Hope"
##   ..$ : chr [1:5] "The Empire Strikes Back" "Revenge of the Sith" "Return of the Jedi" "A N
##   ..$ : chr [1:4] "The Empire Strikes Back" "Return of the Jedi" "A New Hope" "The Force Aw
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:3] "The Phantom Menace" "Return of the Jedi" "A New Hope"
##   ..$ : chr [1:3] "The Empire Strikes Back" "Return of the Jedi" "A New Hope"
##   ..$ : chr "A New Hope"
##   ..$ : chr [1:5] "The Empire Strikes Back" "Attack of the Clones" "The Phantom Menace" "Re
##   ..$ : chr [1:5] "The Empire Strikes Back" "Attack of the Clones" "The Phantom Menace" "Re
##   ..$ : chr [1:3] "The Empire Strikes Back" "Attack of the Clones" "Return of the Jedi"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr [1:2] "The Empire Strikes Back" "Return of the Jedi"
##   ..$ : chr "The Empire Strikes Back"
##   ..$ : chr [1:2] "Return of the Jedi" "The Force Awakens"
##   ..$ : chr "Return of the Jedi"
##   ..$ : chr "Return of the Jedi"
##   ..$ : chr "Return of the Jedi"
##   ..$ : chr "Return of the Jedi"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:2] "Attack of the Clones" "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:2] "Attack of the Clones" "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:2] "Attack of the Clones" "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "Return of the Jedi"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:2] "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##   ..$ : chr [1:2] "Attack of the Clones" "The Phantom Menace"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "The Phantom Menace"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr "Attack of the Clones"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "Revenge of the Sith"
##   ..$ : chr "Revenge of the Sith"
##   ..$ : chr [1:2] "Revenge of the Sith" "A New Hope"
##   ..$ : chr [1:2] "Attack of the Clones" "Revenge of the Sith"
##   ..$ : chr "Revenge of the Sith"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr "The Force Awakens"
##   ..$ : chr [1:3] "Attack of the Clones" "The Phantom Menace" "Revenge of the Sith"
##  $ vehicles  :List of 87
##   ..$ : chr [1:2] "Snowspeeder" "Imperial Speeder Bike"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Imperial Speeder Bike"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Tribubble bongo"
##   ..$ : chr [1:2] "Zephyr-G swoop bike" "XJ-6 airspeeder"
##   ..$ : chr(0)
##   ..$ : chr "AT-ST"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Snowspeeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Tribubble bongo"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Sith speeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Flitknot speeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Koro-2 Exodrive airspeeder"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Tsmeu-6 personal wheel bike"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##  $ starships :List of 87
##   ..$ : chr [1:2] "X-wing" "Imperial shuttle"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "TIE Advanced x1"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "X-wing"
##   ..$ : chr [1:5] "Jedi starfighter" "Trade Federation cruiser" "Naboo star skiff" "Jedi In
##   ..$ : chr [1:3] "Trade Federation cruiser" "Jedi Interceptor" "Naboo fighter"
##   ..$ : chr(0)
##   ..$ : chr [1:2] "Millennium Falcon" "Imperial shuttle"
##   ..$ : chr [1:2] "Millennium Falcon" "Imperial shuttle"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "X-wing"
##   ..$ : chr "X-wing"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Slave 1"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Millennium Falcon"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "A-wing"
##   ..$ : chr(0)
##   ..$ : chr "Millennium Falcon"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Naboo Royal Starship"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Scimitar"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Jedi starfighter"
##   ..$ : chr(0)
##   ..$ : chr "Naboo fighter"
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr(0)
##   ..$ : chr "Belbullab-22 starfighter"
##   ..$ : chr(0)
##   ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr "T-70 X-wing fighter"
##    ..$ : chr(0)
##    ..$ : chr(0)
##    ..$ : chr [1:3] "H-type Nubian yacht" "Naboo star skiff" "Naboo fighter"
```

(b) Recalling what we did last week, plot a bar graph of the count of the `gender` variable. Make the bars horizontal instead of vertical.

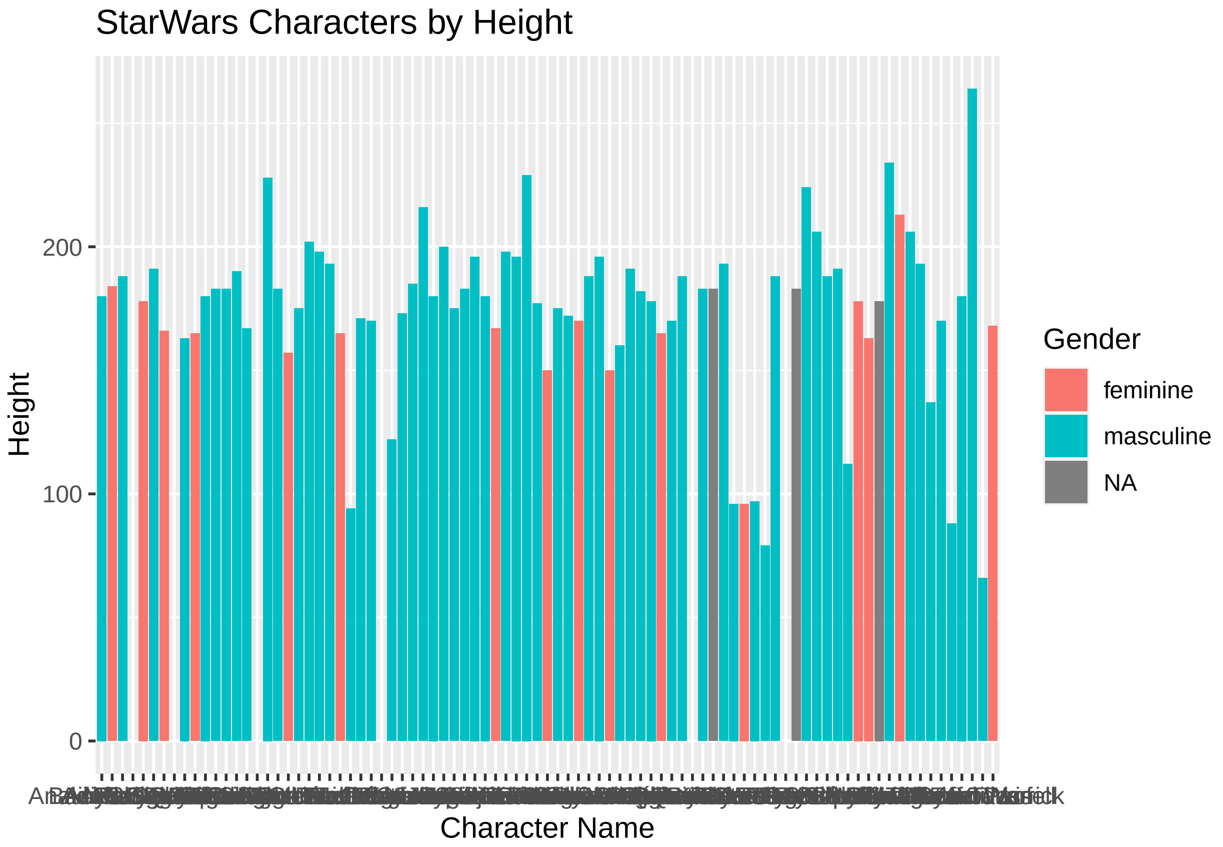
```
my.starwars$gender <- as.factor( my.starwars$gender )

gender_plot <- ggplot( data=my.starwars,
                       aes(
                            y=gender ) ) +
  geom_bar( aes( fill=gender ) ) +
  labs( title="StarWars Characters by Gender", x="Number of Characters",
        y="Gender", fill="Gender" )

gender_plot
```



(c) Now let's make a bivariate bar graph, plot the characters by name across the x-axis and their height on the y-axis. (Hint: first you need to ensure that the `name` variable is a factor so that it can map to our axis nicely.)

```r
my.starwars$name <- as.factor( my.starwars$name )

starwars_plot <- ggplot( data=my.starwars, aes( x=name, y=height ) ) +
  geom_bar( stat='identity', aes( fill=gender ) ) +
  labs( title="StarWars Characters by Height", x="Character Name",
        y="Height", fill="Gender" )

starwars_plot
```

```
## Warning: Removed 6 rows containing missing values (‘position_stack()‘).
```



(d) If you correctly answered (c) the plot will have a lot of bars (too many in my opinion). To reduce the number of observations we can use the `filter` function and the pipe `%>%` to filter them based on the value of a variable. For example the following will filter out all characters who are under 100kg, leaving only the heavier characters:

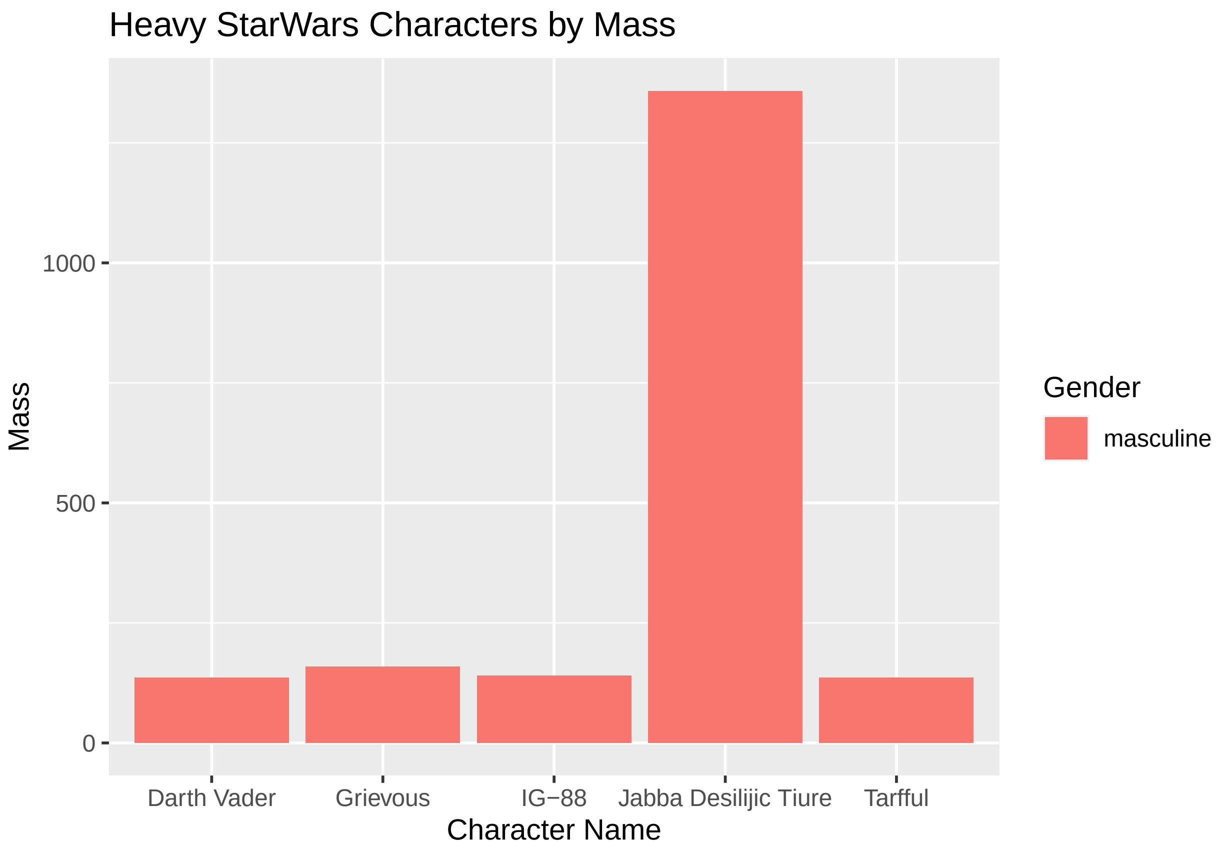
```r
my.starwars.heavy <- my.starwars %>% filter(mass > 100)
```

Note here that if we assign it to a variable name, then we can then use it later.

In the code block below, filter your data to include only star wars characters over 120kg and then plot a bar chart of `mass` against `name` for this filtered data.

```r
my.starwars.chonk <- my.starwars %>% filter( mass > 120 )

```

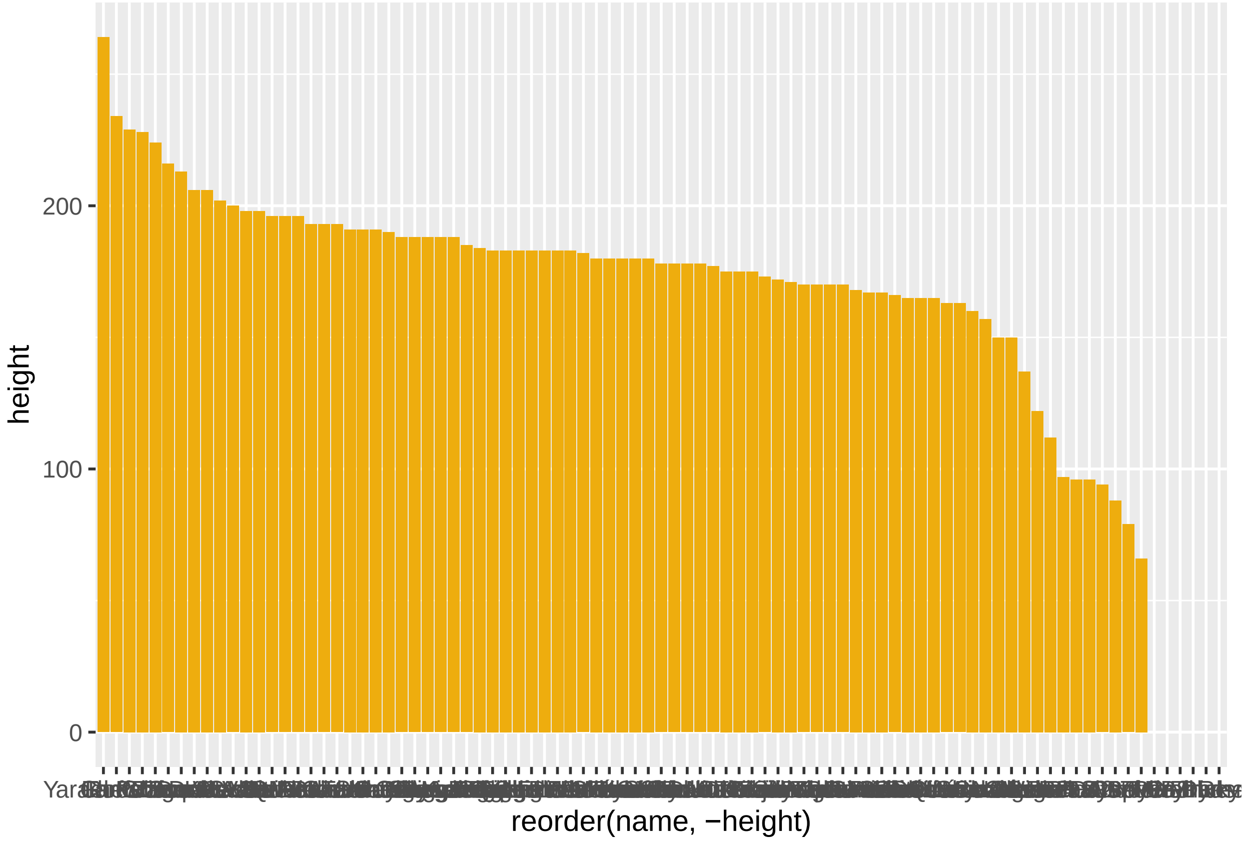
```
starwars_chonk_graph <- ggplot( data=my.starwars.chonk ) +
  geom_bar( stat='identity', aes( x=name, y=mass, fill=gender ) ) +
  labs( title="Heavy StarWars Characters by Mass", x="Character Name",
        y="Mass", fill="Gender" )

starwars_chonk_graph
```



(e) You can reorder the characters on the x-axis with a `reorder` function. We use this by telling the function the variable to reorder and the variable that we want it to be reordered by. For example the following code block will plot a bar chart of `height` against `name` with the characters ordered by decreasing height:

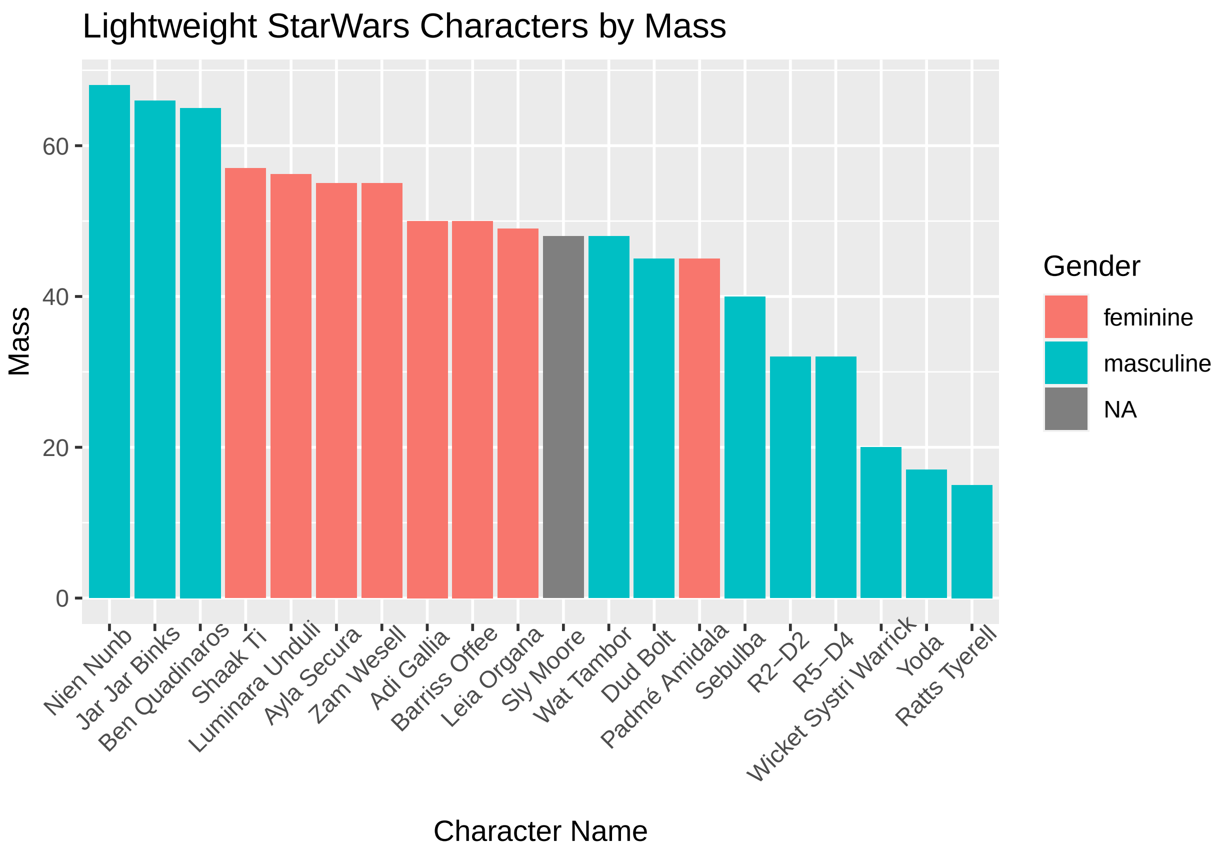
```
ggplot(data=my.starwars,
       mapping=aes(x=reorder(name, -height), y=height))+
  geom_bar(stat="identity", fill="darkgoldenrod2")
```

```
## Warning: Removed 6 rows containing missing values (‘position_stack()‘).
```



In the code block below, filter your data to include only star wars characters under 70kg and then plot a bar chart of `mass` against `name` for this filtered data, with the characters ordered from heaviest (left) to lightest (right) on the x-axis.

```r
my.starwars.smol <- my.starwars %>% filter( mass < 70 )

starwars_smol_graph <- ggplot( data=my.starwars.smol ) +
  geom_bar( stat='identity',
            aes( x=reorder( name, -mass ), y=mass, fill=gender ) ) +
  labs( title="Lightweight StarWars Characters by Mass",
        x="Character Name", y="Mass", fill="Gender" ) +
  theme( axis.text.x = element_text( angle = 45, hjust=.85 ) )

starwars_smol_graph
```



Exercise 2

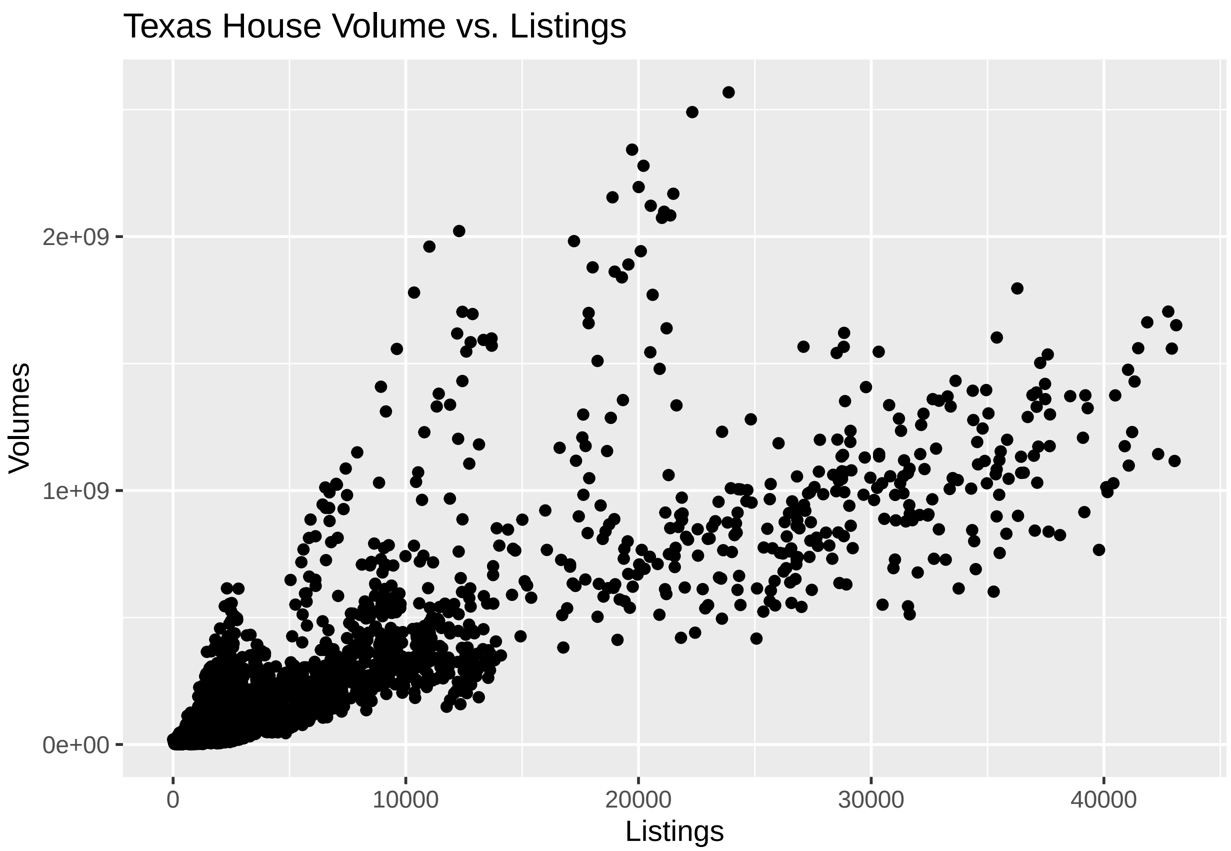
(a) The next dataset we are going to work with is a Texan housing dataset that comes built-in with the ggplot2 package. Run the following code block to save a copy of the dataset to a variable called my.housing.

```
my.housing <- txhousing
```

In this question we are going to look at scatterplots, so we'll be using 2 numerical variables. The geometric object for a scatterplot in ggplot is geom_point(). Use the following code block to create a scatterplot of housing volume against the number of housing listings:

```
housing_plot <- ggplot( data=my.housing ) +
  geom_point( aes( x=listings, y=volume ) ) +
  labs( title="Texas House Volume vs. Listings", x="Listings", y="Volumes" )

housing_plot
```

```
## Warning: Removed 1426 rows containing missing values (‘geom_point()‘).
```



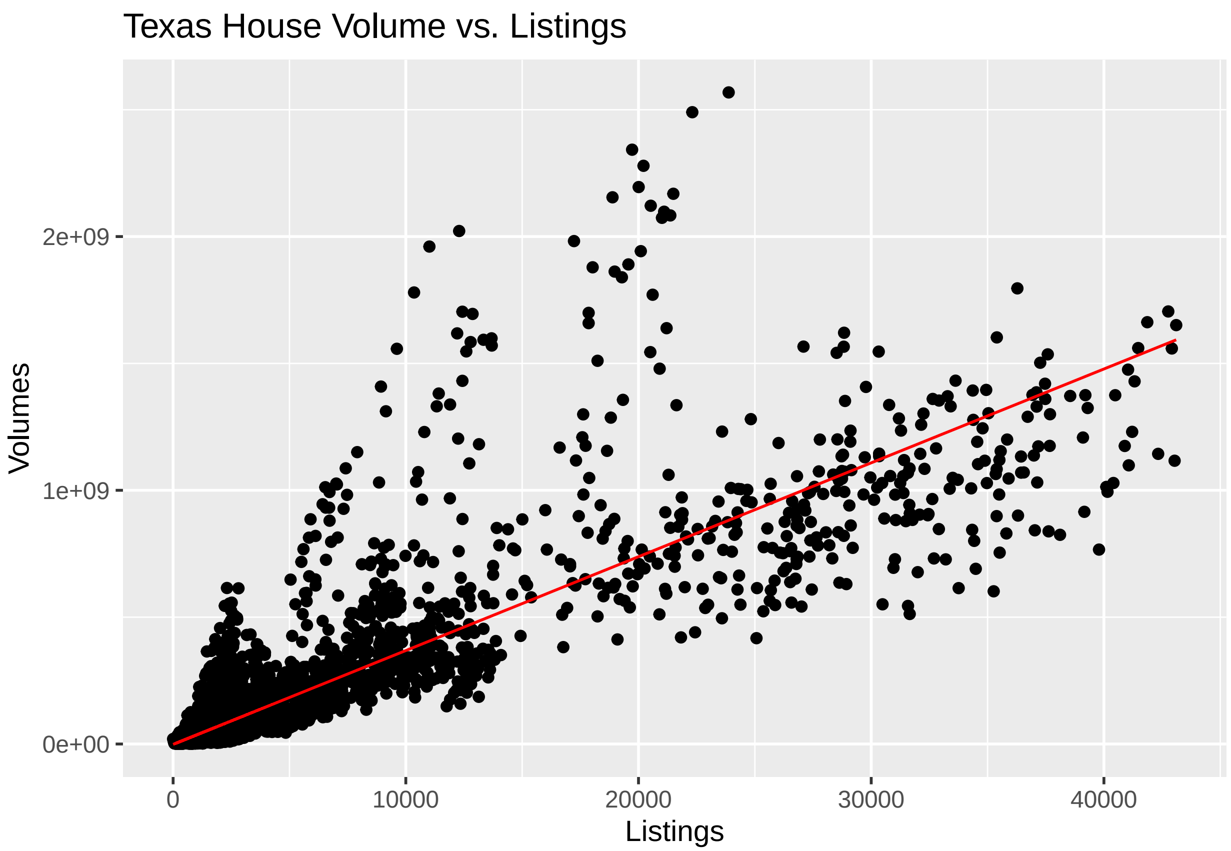
(b) What is the relationship between `volume` and `listings`? (eg. as listings increases/decreases, volume increases/decreases)

There appears to be a moderate, positive relationship between volume and listings. As the number of listings increases, so too does the volume of sales.

(c) Using the plot from part (b), use the following code block to add a line-of-best-fit to this plot:

```
SLR <- lm( my.housing$volume ~ my.housing$listings )

my.housing <- my.housing %>%
  dplyr::select( -matches('fit'), -matches('lwr'), -matches('upr') ) %>%
  cbind( predict(SLR, newdata=., interval='confidence') )

housing_plot <- ggplot( data=my.housing ) +
  geom_point( aes( x=listings, y=volume ) ) +
  labs( title="Texas House Volume vs. Listings", x="Listings", y="Volumes" ) +
  geom_line( aes( x=listings, y=fit ), color='red' )

housing_plot
```

```
## Warning: Removed 1426 rows containing missing values (‘geom_point()‘).
## Warning: Removed 1424 rows containing missing values (‘geom_line()‘).
```

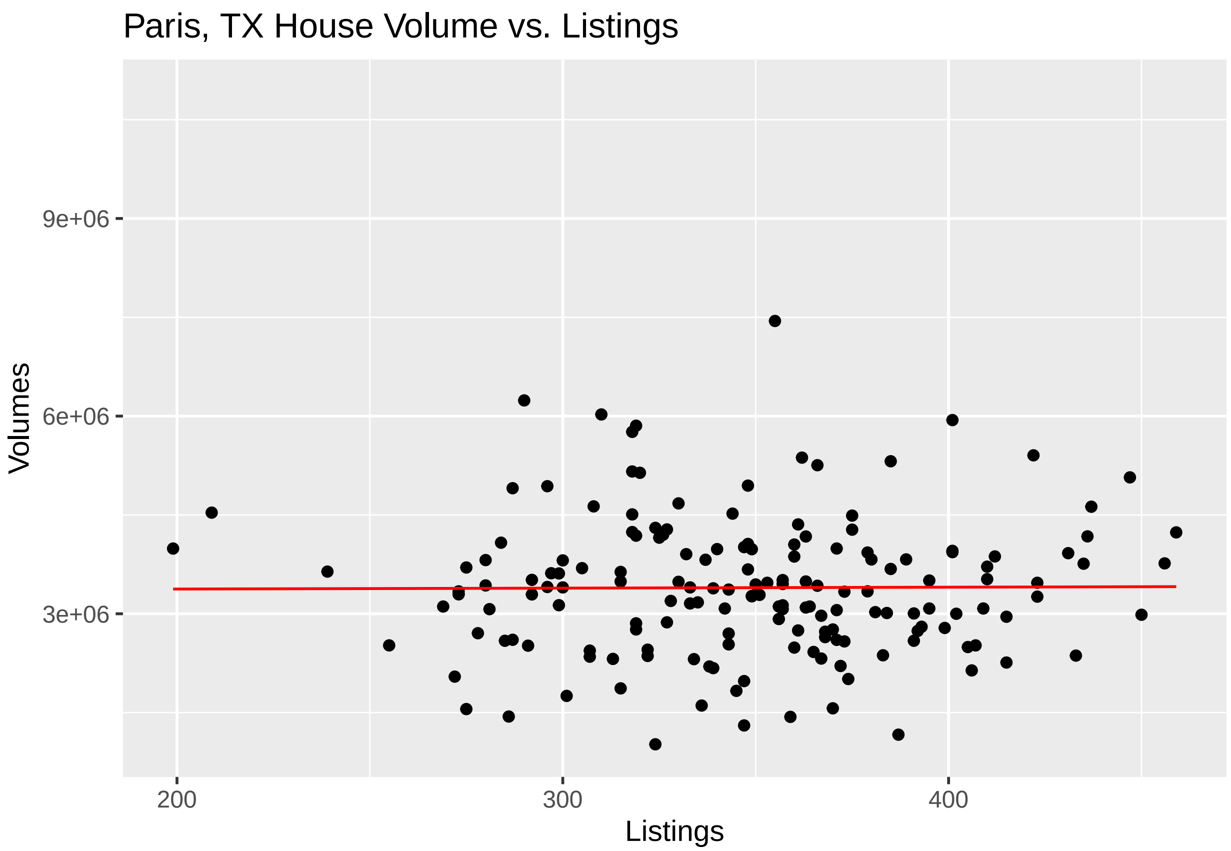


(d) Filter your data so that you reduce your data to only the city of Paris, TX (using == for equality). Create the same scatterplot with a line-of-best-fit for the data only from Paris.

```
my.housing.paris <- my.housing %>% filter( city == 'Paris' )

SLR.paris <- lm( my.housing.paris$volume ~ my.housing.paris$listings )

my.housing.paris <- my.housing.paris %>%
  dplyr::select( -matches('fit'), -matches('lwr'), -matches('upr') ) %>%
  cbind( predict(SLR.paris, newdata=., interval='confidence') )

housing_plot <- ggplot( data=my.housing.paris ) +
  geom_point( aes( x=listings, y=volume ) ) +
  labs( title="Paris, TX House Volume vs. Listings", x="Listings", y="Volumes" ) +
  geom_line( aes( x=listings, y=fit ), color='red' )

housing_plot
```

```
## Warning: Removed 19 rows containing missing values (‘geom_point()‘).
```

```
## Warning: Removed 19 rows containing missing values (‘geom_line()‘).
```



(e) What is the relationship between `volume` and `listings` in Paris?

There appears to be no significant relationship between volume and listings in Paris.

Exercise 3

(a) The final dataset we are going to work with is data on the US economy over time, and it also comes built-in with the `ggplot2` package. Run the following code block to save a copy of the dataset to a variable called `my.econ`.

```
my.econ <- economics
```

Use the following code block to look at the structure of your data, specifically the data types. What data type is the `date` variable?

```
str( my.econ )
```

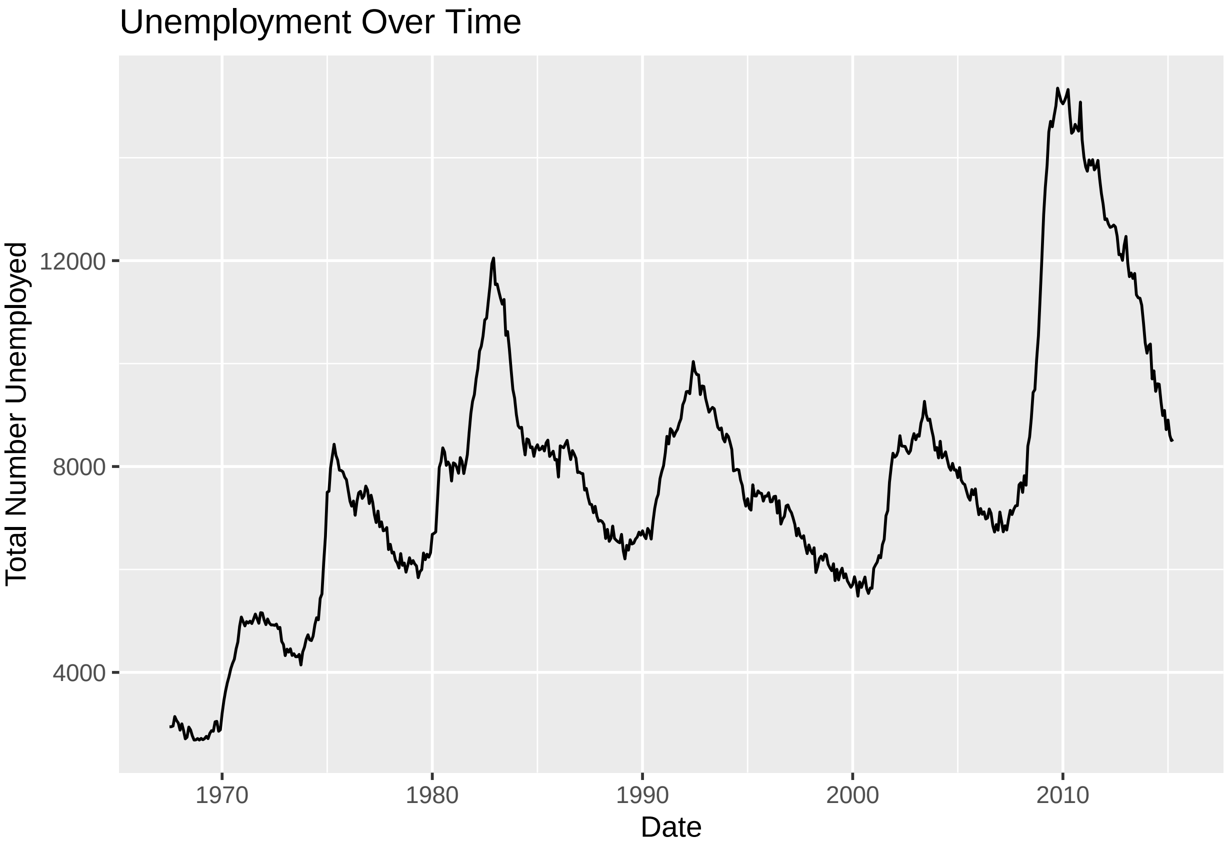
```
## spc_tbl_ [574 x 6] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ date    : Date[1:574], format: "1967-07-01" "1967-08-01" ...
##  $ pce     : num [1:574] 507 510 516 512 517 ...
##  $ pop     : num [1:574] 198712 198911 199113 199311 199498 ...
##  $ psavert : num [1:574] 12.6 12.6 11.9 12.9 12.8 11.8 11.7 12.3 11.7 12.3 ...
##  $ uempmed : num [1:574] 4.5 4.7 4.6 4.9 4.7 4.8 5.1 4.5 4.1 4.6 ...
##  $ unemploy: num [1:574] 2944 2945 2958 3143 3066 ...
```

The date variable is of the type Date.

(b) In this question we are going to look at plotting line graphs, in particular time series data, where the independent variable is time (day, week, date, etc.). The geometric object for a line graph in `ggplot` is `geom_line()`. As the date variable is already a date, ℝ knows how to handle it here, but in the future we might have to cast our variable to a date data type.

Use the following code block to plot a line graph of unemployment against time:

```
econ_graph <- ggplot( data=my.econ ) +
  geom_line( aes( x=date, y=unemploy ) ) +
  labs( title="Unemployment Over Time", x="Date", y="Total Number Unemployed" )

econ_graph
```



(c) What notable about this plot? Eg. Is it increasing? Is there a pattern?

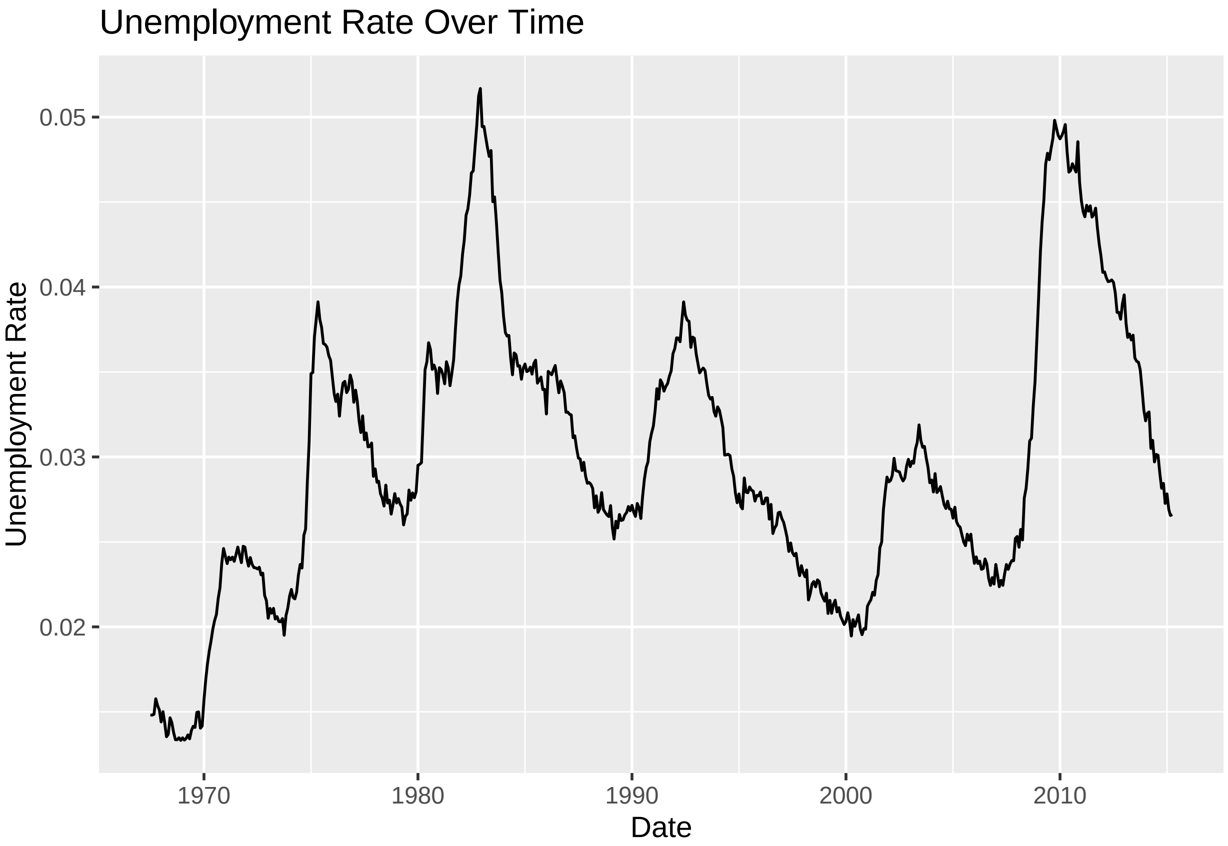
Unemployment does seem to be increasing as a whole, however, there also appear to be cycles of high and low unemployment which periodically occur.

(d) To add a new column to our dataset that is based on existing columns, we can write it as an equation of the existing columns and assign it to a new column name. For example, if we want to create a column for the unemployment rate (unemployment per person), we can run:

```
my.econ$unemployment.rate <- my.econ$unemploy / my.econ$pop
```

If you view your data after running this you will see that your new column is on the end of the dataset. Plot this new column against time as a line graph.

```
unemployment_chart <- ggplot( data=my.econ ) +
  geom_line( aes( x=date, y=unemployment.rate ) ) +
  labs( title="Unemployment Rate Over Time", x="Date", y="Unemployment Rate" )

unemployment_chart
```



(e) Is the pattern/trend here different or the same as that you observed in part (c)?

The trend of this chart seems to be more neutral than the chart in part (c). It would appear that unemployment rate is slightly increasing over time, but much less than total unemployment. Likewise, we also see the same cyclic patterns of high and low unemployment rates over time. This could likely be explained by a relatively stable unemployment rate, but with a growing population, leading to higher/growing overall rates of unemployment.