Homework #1

Goals

Read the files in and then accomplish the following goals:

```
#For Genre Data, I used read.table and set the header to TRUE in order to read in the .tx
Genre_Data <- read.table(
  "/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework 1/artist_genre.txt", head

#For the next 3 csv files, I just used the read.csv function in order to read the files i
Name_Birth_Data <- read.csv("/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework
Album_Data <- read.csv("/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework 1/ar
Top_Hits_Data <- read.csv("/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework 1

#I did use colnames to rename "artist" to "Artist", which made my merge functions a littl
Num_Hits_Data <- read.csv("/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework 1
colnames(Num_Hits_Data) <- c("Artist", "number_one_hits")

#gsub found the $ in the Imagine Dragons function and replaced it with a space, which mad
Net_Worth_Data <- read.csv("/Users/TomTheIntern/Desktop/Mendoza/Mod 1/Wrangling/Homework
Net_Worth_Data$net_worth_millions <- gsub("\\$", "", Net_Worth_Data$net_worth_millions)

#And like Num_Hits_Data, I did rename the columns, just updating artist to Artist, which
colnames(Net_Worth_Data) <- c("Artist", 'Net_Worth')
```

Pre-join Goals

```
#Here, in the pre-join goals I opted to use the aggregate function because it just made m

Genre_Data_2 <- aggregate(genre ~ Artist, Genre_Data, unique)

Name_Birth_Data_2 <- aggregate((as.numeric(Birth_Year)) ~ Artist, Name_Birth_Data, unique

Net_Worth_Data_2 <- aggregate((as.numeric(Net_Worth)) ~ Artist, Net_Worth_Data, unique)

Num_Hits_Data_2 <- aggregate(number_one_hits ~ Artist, Num_Hits_Data, unique)

Album_Data_2 <- aggregate(studio_albums ~ Artist, Album_Data, unique)

Top_Hits_Data_2 <- aggregate(Top_Hit ~ Artist, Top_Hits_Data, unique)

Top_Hits_Year <- aggregate(Year ~ Artist, Top_Hits_Data, unique)
```

- Which 5 artists have the greatest net worth?

```
Top_Networth <- head(Net_Worth_Data_2[order(Net_Worth_Data_2$`(as.numeric(Net_Worth))`, d
rmarkdown::paged_table(Top_Networth)
```

	Artist	(as.numeric(Net_Worth))
	<chr>	<dbl>
25	Jay-Z	1500
54	The Beatles	1500
1	ABBA	900
58	The Rolling Stones	900
38	Madonna	850

5 rows

- Which 15 artists have the most number 1 hits?

```
Most_Hits <- head(Num_Hits_Data_2[order(Num_Hits_Data_2$number_one_hits, decreasing = TRU
rmarkdown::paged_table(Most_Hits)
```

	Artist	number_one_hits
	<chr>	<int>
54	The Beatles	20
39	Mariah Carey	19
19	Elvis Presley	18
47	Rihanna	14
41	Michael Jackson	13
38	Madonna	12
61	Whitney Houston	11
51	Stevie Wonder	10
32	Katy Perry	9
15	Drake	8

1-10 of 15 rows   Previous   **1**   2   Next

- Which 10 artists have the fewest studio albums?

```
Fewest_Albums <- head(Album_Data_2[order(Album_Data_2$studio_albums, decreasing = FALSE),
rmarkdown::paged_table(Fewest_Albums)
```

	Artist	studio_albums
	<chr>	<int>
8	Billie Eilish	1
37	Lorde	2
56	The Notorious B.I.G.	2
2	Adele	3
12	Bruno Mars	3

	Artist	studio_albums
	<chr>	<int>
26	Jimi Hendrix	3
42	Nirvana	3
48	Sam Smith	3
24	Imagine Dragons	4
31	Justin Timberlake	4

1-10 of 10 rows

After you have those done, you can join your data frames together.

```
Complete_songs_a <- merge(
  x = Name_Birth_Data_2,
  y = Genre_Data_2,
  by = ("Artist")
)

Complete_songs_b <- merge(
  x = Complete_songs_a,
  y = Net_Worth_Data_2,
  by = ("Artist")
)

Complete_songs_c <- merge(
  x = Complete_songs_b,
  y = Num_Hits_Data_2,
  by = "Artist"
)

Complete_songs_d <- merge(
  x = Complete_songs_c,
  y = Album_Data_2,
  by = "Artist"
)

Complete_songs_e <- merge(
  x = Complete_songs_d,
  y = Top_Hits_Data_2,
  by = "Artist"
)

Songs_Final <- merge(
  x = Complete_songs_e,
  y = Top_Hits_Year,
  by = "Artist"
)

colnames(Songs_Final) <- c('Artist Name', 'Birth_Year',
                           'Genre',
```

```
                                'Net_Worth_in_Mil', '#1 Hits',
                                'Studio_Albums', 'Top Hit',
                                'Year')
```
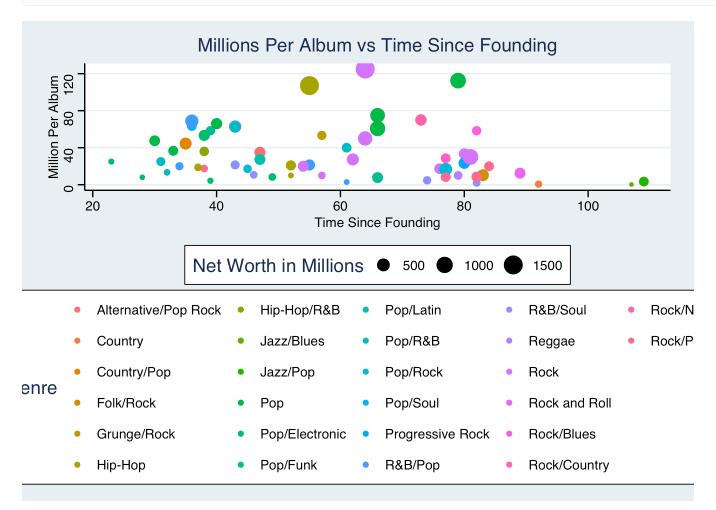
Post-join Goals

```
Songs_Final$Time_Since_Founding <- 2024 - Songs_Final$Birth_Year
Songs_Final$Mil_Per_Album <- Songs_Final$Net_Worth_in_Mil / Songs_Final$Studio_Albums

rmarkdown::paged_table(Songs_Final)
```

Artist Name	Birth_Year	Genre	▶
<chr>	<dbl>	<chr>	
ABBA	1945	Pop	
Adele	1988	Pop/Soul	
Aerosmith	1948	Rock	
Alicia Keys	1981	R&B/Soul	
Aretha Franklin	1942	R&B/Soul	
Ariana Grande	1993	Pop/R&B	
Beyoncé	1981	R&B/Pop	
Billie Eilish	2001	Pop/Electronic	
Bob Dylan	1941	Folk/Rock	
Bob Marley	1945	Reggae	

1-10 of 61 rows | 1-3 of 10 columns   Previous **1** [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [Next](#)

- Create variables to capture the following ideas:
    - How long has the artist been around? You can just use their birth year.
    - Something to capture each artist's net worth per studio album.
- Plot net worth per studio album, the "been around" variable, and genre. There should be points and colors here.

```
library(ggplot2)
library(ggthemes)
# https://www.datanovia.com/en/blog/ggplot-themes-gallery/

Songs_Graphic <- ggplot(data = Songs_Final,
                   aes(
                       x = Time_Since_Founding,
                       y = Mil_Per_Album,
                       color = Genre,
                       size = Net_Worth_in_Mil)) +
  geom_point() +
  theme_stata() +
  labs(
    title = 'Millions Per Album vs Time Since Founding',
    x = 'Time Since Founding',
```

```
        y = 'Million Per Album',
        size = "Net Worth in Millions"
    )
Songs_Graphic
```

Millions Per Album vs Time Since Founding

For all tasks comment your code and explain what you found!