Assignment 2 BUS4028

Group-3

```
data <- read.csv("C:/T405/R programming/grammy_winners.csv", stringsAsFactors = FALSE)
head(data)
```

```
##   year annual_edition           category
## 1 2024             67  Album Of The Year
## 2 2024             67  Album Of The Year
## 3 2024             67 Record Of The Year
## 4 2024             67 Record Of The Year
## 5 2024             67 Record Of The Year
## 6 2024             67 Record Of The Year
##
## 1 Jack Antonoff, Aaron Dessner & Taylor Swift, producers; Zem Audu, Bella Blasko, Bryce Bordone, Serb
## 2
## 3
## 4
## 5
## 6
##                    song_or_album winner
## 1 THE TORTURED POETS DEPARTMENT  False
## 2                 COWBOY CARTER   True
## 3                   Now And Then  False
## 4                TEXAS HOLD 'EM  False
## 5                      Espresso  False
## 6                           360  False
##                                                          url
## 1 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 2 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 3 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 4 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 5 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 6 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
```

```
# Print the structure of your dataset
str(data)
```

```
## 'data.frame':    25369 obs. of  7 variables:
##  $ year          : int  2024 2024 2024 2024 2024 2024 2024 2024 2024 2024 ...
##  $ annual_edition: int  67 67 67 67 67 67 67 67 67 67 ...
##  $ category      : chr  "Album Of The Year" "Album Of The Year" "Record Of The Year" "Record Of The Y
##  $ artist        : chr  "Jack Antonoff, Aaron Dessner & Taylor Swift, producers; Zem Audu, Bella Blas
##  $ song_or_album : chr  "THE TORTURED POETS DEPARTMENT" "COWBOY CARTER" "Now And Then" "TEXAS HOLD 'E
##  $ winner        : chr  "False" "True" "False" "False" ...
##  $ url           : chr  "https://www.grammy.com/awards/67th-annual-grammy-awards-2024" "https://www.g
```

```r
# List the variables in your dataset
names(data)
```

```
## [1] "year"           "annual_edition" "category"       "artist"
## [5] "song_or_album"  "winner"         "url"
```

```r
# Print the top 15 rows of your dataset
head(data, 15)
```

```
##    year annual_edition                            category
## 1  2024             67                     Album Of The Year
## 2  2024             67                     Album Of The Year
## 3  2024             67                    Record Of The Year
## 4  2024             67                    Record Of The Year
## 5  2024             67                    Record Of The Year
## 6  2024             67                    Record Of The Year
## 7  2024             67                     Album Of The Year
## 8  2024             67                     Album Of The Year
## 9  2024             67                     Album Of The Year
## 10 2024             67                     Album Of The Year
## 11 2024             67                     Album Of The Year
## 12 2024             67 Songwriter of the Year, Non-Classical
## 13 2024             67 Songwriter of the Year, Non-Classical
## 14 2024             67 Songwriter of the Year, Non-Classical
## 15 2024             67 Songwriter of the Year, Non-Classical
##
## 1                    Jack Antonoff, Aaron Dessner & Taylor Swift, producers; Zem Audu, Bella Blasko,
## 2
## 3
## 4
## 5
## 6
## 7  Jack Antonoff, Julian Bunetta, Ian Kirkpatrick & John Ryan, producers; Bryce Bordone, Julian Bunet
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
##                                          song_or_album winner
## 1                        THE TORTURED POETS DEPARTMENT  False
## 2                                      COWBOY CARTER   True
## 3                                        Now And Then  False
## 4                                    TEXAS HOLD 'EM  False
## 5                                         Espresso  False
## 6                                              360  False
## 7                                    Short n' Sweet  False
## 8                                              BRAT  False
## 9                                     Djesse Vol. 4  False
## 10                              HIT ME HARD AND SOFT  False
## 11 Chappell Roan The Rise And Fall Of A Midwest Princess  False
```

```
## 12                                        Jessi Alexander  False
## 13                                          Edgar Barrera  False
## 14                                       Jessie Jo Dillon  False
## 15                                                  RAYE  False
##                                                                 url
## 1  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 2  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 3  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 4  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 5  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 6  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 7  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 8  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 9  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 10 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 11 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 12 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 13 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 14 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 15 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
```

```r
# Write a user defined function using any of the variables from the data set
awards_by_year <- function(data) {
  year_counts <- table(data$year)
  year_counts <- sort(year_counts)
  return(year_counts)
}
awards_by_year(data)
```

```
##
## 1958 1959 1961 1968 1962 1960 1966 1969 1970 1963 1971 1972 1973 1967 1974 1964
##  147  172  202  218  240  241  241  244  244  255  258  260  261  265  267  268
## 1965 1975 1977 1976 1978 1980 1981 1982 1979 1984 1983 1986 1987 1985 1989 1990
##  268  274  279  282  300  305  308  320  334  339  345  347  362  376  383  388
## 1988 2011 1991 1993 1992 2012 2013 2015 2017 2014 2016 2019 2020 2018 1994 1995
##  390  394  400  400  403  404  410  414  418  422  425  432  434  437  441  442
## 1996 1997 2021 1998 2022 2023 1999 2024 2000 2001 2002 2003 2004 2006 2010 2005
##  453  461  467  478  486  490  496  497  504  509  524  533  537  538  541  543
## 2009 2007 2008
##  547  553  553
```

```r
# Use data manipulation techniques and filter rows based on any logical criteria that exist in your dat
album_of_the_year <- subset(data, category == "Album Of The Year")
head(album_of_the_year)
```

```
##    year annual_edition          category
## 1  2024             67 Album Of The Year
## 2  2024             67 Album Of The Year
## 7  2024             67 Album Of The Year
## 8  2024             67 Album Of The Year
## 9  2024             67 Album Of The Year
## 10 2024             67 Album Of The Year
##
```

```
## 1                    Jack Antonoff, Aaron Dessner & Taylor Swift, producers; Zem Audu, Bella Blasko,
## 2
## 7  Jack Antonoff, Julian Bunetta, Ian Kirkpatrick & John Ryan, producers; Bryce Bordone, Julian Bunet
## 8
## 9
## 10
##                     song_or_album winner
## 1  THE TORTURED POETS DEPARTMENT  False
## 2                   COWBOY CARTER   True
## 7                 Short n' Sweet  False
## 8                           BRAT  False
## 9                   Djesse Vol. 4  False
## 10          HIT ME HARD AND SOFT  False
##                                                                 url
## 1  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 2  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 7  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 8  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 9  https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 10 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
```

```
# Identify the dependent & independent variables and use reshaping techniques and create a new data fra

award_data <- data[, c("year", "artist")]

# Filtering top 3 artists
top_artists <- c("Beyoncé", "U2", "Stevie Wonder")
award_data <- subset(award_data, artist %in% top_artists)

# Remove duplicates
award_data <- unique(award_data)

# New table
reshaped <- table(award_data$year, award_data$artist)

head(reshaped, 5)
```

```
##
##        Beyoncé Stevie Wonder U2
##   1973       0             1  0
##   1974       0             1  0
##   1976       0             1  0
##   1985       0             1  0
##   1987       0             0  1
```

```
nrow(data)
```

```
## [1] 25369
```

```
# Remove missing values in your dataset.
clean_data <- na.omit(data)
nrow(clean_data)
```

```
## [1] 25369
```

```
nrow(data)
```

```
## [1] 25369
```

```
# Since ther is no missing values, to double check we used sum(is.na()) to make sure its correct
sum(is.na(data))
```

```
## [1] 0
```

```
# Identify and remove duplicated data in your dataset
duplicated_rows <- data[duplicated(data), ]
head(duplicated_rows)
```

```
##       year annual_edition
## 274   2024             67
## 17879 1985             28
## 18204 1985             28
## 18722 1983             26
## 22162 1971             14
## 23595 1965              8
##                                                                 category artist
## 274                                    Best Regional Roots Music Album
## 17879                                        Best New Classical Artist
## 18204                                        Best New Classical Artist
## 18722 Best Recording For Children - Single or Album, Musical or Spoken
## 22162                                  Best Pop Instrumental Performance
## 23595                      Most Promising New Classical Recording Artist
##                                                 song_or_album winner
## 274   Live At The 2024 New Orleans Jazz & Heritage Festival  False
## 17879                                                        False
## 18204                                                        False
## 18722                                          Born To Add  False
## 22162                               Theme From Summer Of '42  False
## 23595                                                        False
##                                                                  url
## 274   https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 17879      https://www.grammy.com/awards/28th-annual-grammy-awards
## 18204      https://www.grammy.com/awards/28th-annual-grammy-awards
## 18722      https://www.grammy.com/awards/26th-annual-grammy-awards
## 22162      https://www.grammy.com/awards/14th-annual-grammy-awards
## 23595       https://www.grammy.com/awards/8th-annual-grammy-awards
```

```
nrow(duplicated_rows)        # Total duplicate rows found
```

```
## [1] 17
```

```
data_no_duplicates <- data[!duplicated(data), ]
head(data_no_duplicates)
```

```
##   year annual_edition           category
## 1 2024             67  Album Of The Year
## 2 2024             67  Album Of The Year
## 3 2024             67 Record Of The Year
## 4 2024             67 Record Of The Year
## 5 2024             67 Record Of The Year
## 6 2024             67 Record Of The Year
##
## 1 Jack Antonoff, Aaron Dessner & Taylor Swift, producers; Zem Audu, Bella Blasko, Bryce Bordone, Serb
## 2
## 3
## 4
## 5
## 6
##                   song_or_album winner
## 1 THE TORTURED POETS DEPARTMENT  False
## 2                 COWBOY CARTER   True
## 3                  Now And Then  False
## 4                TEXAS HOLD 'EM  False
## 5                     Espresso  False
## 6                          360  False
##                                                        url
## 1 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 2 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 3 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 4 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 5 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 6 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
```

```
nrow(data_no_duplicates)     # After removing duplicates
```

```
## [1] 25352
```

```
# Reorder multiple rows in descending order
 # Reorder rows by year (descending) and artist (descending)
reordered_data <- data[order(-data$year, -xtfrm(data$artist)), ]
 # View the top rows of the reordered dataset
head(reordered_data)
```

```
##     year annual_edition                                    category
## 174 2024             67                       Best Latin Jazz Album
## 267 2024             67                 Best Gospel Performance/Song
## 456 2024             67 Best Chamber Music/Small Ensemble Performance
## 466 2024             67                         Best Opera Recording
## 436 2024             67                         Best Opera Recording
## 328 2024             67        Best New Age, Ambient, or Chant Album
##
## 174
## 267
## 456
## 466 Yannick Nézet-Séguin, conductor; Mario Chang, Michael Chioldi, Greer Grimsley, Nancy Fabiola Herr
## 436                                                                                  Yannick Nézet-Ségu
## 328
```

```
##                                                song_or_album winner
## 174                                             Cubop Lives!   True
## 267                                             Church Doors  False
## 456 Beethoven For Three: Symphony No. 4 And Op. 97, 'Archduke'  False
## 466                          Catán: Florencia En El Amazonas  False
## 436                                          Puts: The Hours  False
## 328                                                  Triveni   True
##                                                                 url
## 174 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 267 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 456 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 466 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 436 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
## 328 https://www.grammy.com/awards/67th-annual-grammy-awards-2024
```

```
# Rename some of the column names in your dataset
colnames(data)[colnames(data) == "artist"] <- "artist_name"
colnames(data)[colnames(data) == "year"] <- "award_year"

names(data)
```

```
## [1] "award_year"     "annual_edition" "category"       "artist_name"
## [5] "song_or_album"  "winner"         "url"
```

```
# Add new variables in your data frame by using a mathematical function (for e.g. - multiply an existin
data$next_award_check_year <- data$award_year + 2
head(data[, c("award_year", "next_award_check_year")])
```

```
##   award_year next_award_check_year
## 1       2024                 2026
## 2       2024                 2026
## 3       2024                 2026
## 4       2024                 2026
## 5       2024                 2026
## 6       2024                 2026
```

```
# training set using random number generator engine
set.seed(123)  # for reproducibility
 # random indices for training set (e.g., 70% of data)
train_index <- sample(1:nrow(data), size = 0.7 * nrow(data))
 # training set
training_set <- data[train_index, ]

head(training_set)
```

```
##       award_year annual_edition                         category
## 18847       1983             26 Best Rock Instrumental Performance
## 18895       1983             26                   Song Of The Year
## 25102       1959              2    Best Jazz Performance - Soloist
## 2986        2018             61                      Best R&B Song
## 1842        2021             64                    Best R&B Album
## 3371        2017             60                 Best Comedy Album
```

```
##                                                                 artist_name
## 18847
## 18895                         Michael Jackson, songwriter (Michael Jackson)
## 25102
## 2986  Paul Boutin, Toni Braxton & Antonio Dixon, songwriters (Toni Braxton)
## 1842
## 3371
##                            song_or_album winner
## 18847 Unused Piano: Quadrophenia (Track)  False
## 18895                     Beat It (Single)  False
## 25102                  Red Norvo In Hi-Fi  False
## 2986                       Long As I Live  False
## 1842                   Gold-Diggers Sound  False
## 3371                                Cinco  False
##                                                                    url
## 18847      https://www.grammy.com/awards/26th-annual-grammy-awards
## 18895      https://www.grammy.com/awards/26th-annual-grammy-awards
## 25102       https://www.grammy.com/awards/2nd-annual-grammy-awards
## 2986  https://www.grammy.com/awards/61st-annual-grammy-awards-2018
## 1842  https://www.grammy.com/awards/64th-annual-grammy-awards-2021
## 3371  https://www.grammy.com/awards/60th-annual-grammy-awards-2017
##       next_award_check_year
## 18847                  1985
## 18895                  1985
## 25102                  1961
## 2986                   2020
## 1842                   2023
## 3371                   2019
```

```r
# Print the summary statistics of your dataset
summary(data)
```

```
##    award_year   annual_edition    category         artist_name
##  Min.   :1958   Min.   : 1.00   Length:25369       Length:25369
##  1st Qu.:1982   1st Qu.:25.00   Class :character   Class :character
##  Median :1998   Median :41.00   Mode  :character   Mode  :character
##  Mean   :1996   Mean   :38.79
##  3rd Qu.:2010   3rd Qu.:53.00
##  Max.   :2024   Max.   :67.00
##  song_or_album         winner              url            next_award_check_year
##  Length:25369       Length:25369       Length:25369       Min.   :1960
##  Class :character   Class :character   Class :character   1st Qu.:1984
##  Mode  :character   Mode  :character   Mode  :character   Median :2000
##                                                           Mean   :1998
##                                                           3rd Qu.:2012
##                                                           Max.   :2026
```

```r
# Use any of the numerical variables from the dataset and perform the following statistical functions

 # Mean
mean(data$award_year)
```

```
## [1] 1995.791
```

```r
 # Median
median(data$award_year)
```

```
## [1] 1998
```

```r
# Mode (custom function, since R has no built-in mode)
get_mode <- function(x) {
  uniq_vals <- unique(x)
  uniq_vals[which.max(tabulate(match(x, uniq_vals)))]
}
get_mode(data$award_year)
```

```
## [1] 2008
```

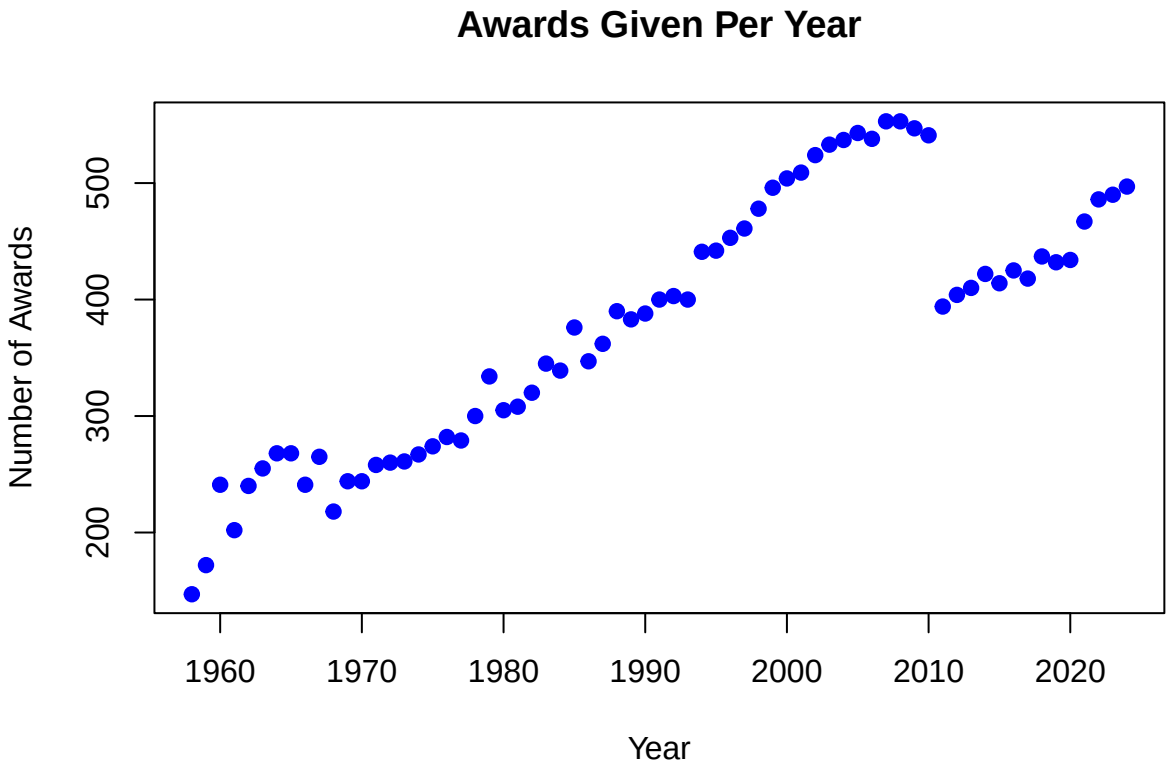
```r
# Range
range(data$award_year)
```

```
## [1] 1958 2024
```

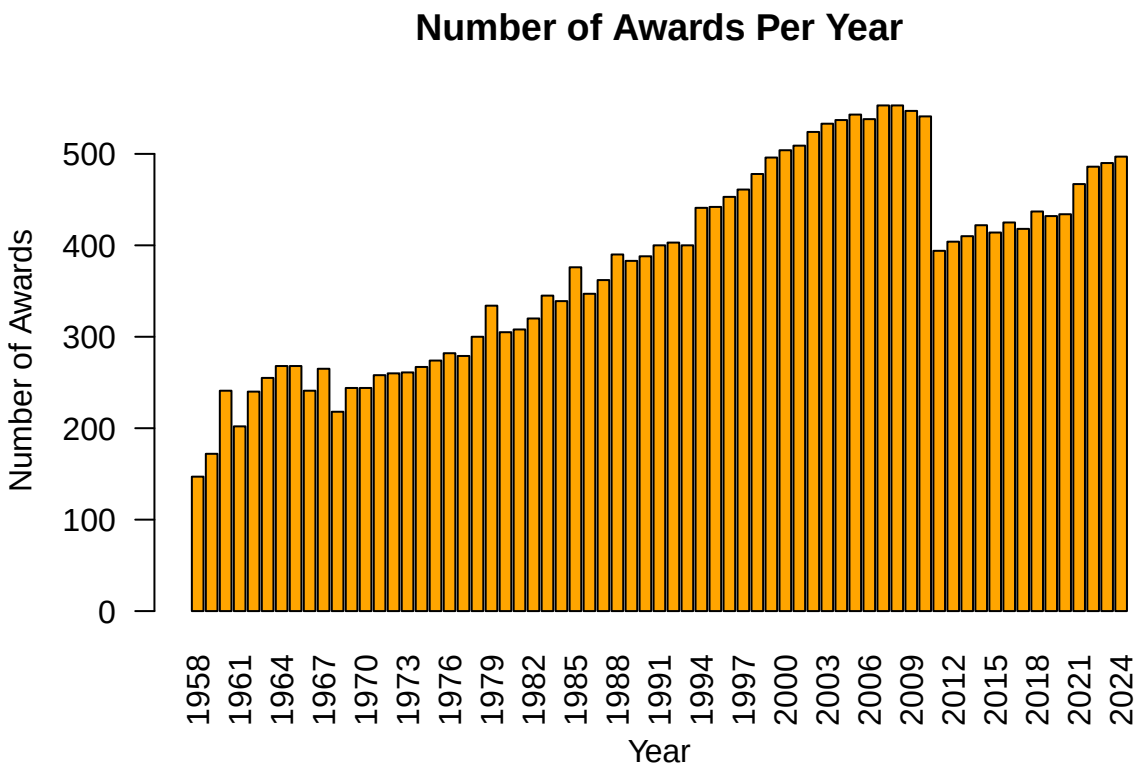
```r
award_counts <- table(data$award_year)
award_df <- as.data.frame(award_counts)
colnames(award_df) <- c("year", "award_count")
award_df$year <- as.numeric(as.character(award_df$year))  # convert year to numeric
head(award_df)
```

```
##   year award_count
## 1 1958         147
## 2 1959         172
## 3 1960         241
## 4 1961         202
## 5 1962         240
## 6 1963         255
```

```r
# Plot a scatter plot for any 2 variables in your dataset
plot(award_df$year, award_df$award_count,
     main = "Awards Given Per Year",
     xlab = "Year",
     ylab = "Number of Awards",
     pch = 19,
     col = "blue")
```




```
barplot(award_counts,
        main = "Number of Awards Per Year",
        xlab = "Year",
        ylab = "Number of Awards",
        col = "orange",
        las = 2)
```




```r
# Find the correlation between any 2 variables by applying Pearson correlation
category_counts <- aggregate(category ~ award_year, data = data, FUN = function(x) length(unique(x)))
```

```r
cor(category_counts$award_year, category_counts$category, method = "pearson")
```

```
## [1] 0.8620106
```

0.86 shows that as the years increase, the number of award categories also increases — consistently.