HPAM 7660 Data Assignment1

Claire English

2024-02-06

```
#loading and installing all needed packages
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(nycflights13)
library(tidyr)
library(fivethirtyeight)
```

```
## Some larger datasets need to be installed separately, like senators and
## house_district_forecast. To install these, we recommend you install the
## fivethirtyeightdata package by running:
## install.packages('fivethirtyeightdata', repos =
## 'https://fivethirtyeightdata.github.io/drat/', type = 'source')
```

```
#viewing drinks data frame using glimpse
glimpse(drinks)
```

```
## Rows: 193
## Columns: 5
## $ country                       <chr> "Afghanistan", "Albania", "Algeria", "And~
## $ beer_servings                 <int> 0, 89, 25, 245, 217, 102, 193, 21, 261, 2~
## $ spirit_servings               <int> 0, 132, 0, 138, 57, 128, 25, 179, 72, 75,~
## $ wine_servings                 <int> 0, 54, 14, 312, 45, 45, 221, 11, 212, 191~
## $ total_litres_of_pure_alcohol <dbl> 0.0, 4.9, 0.7, 12.4, 5.9, 4.9, 8.3, 3.8, ~
```

```
#looking at data again
drinks
```

```
## # A tibble: 193 x 5
##    country    beer_servings spirit_servings wine_servings total_litres_of_pure~1
##    <chr>              <int>           <int>         <int>                   <dbl>
##  1 Afghanist~             0               0             0                     0
##  2 Albania               89             132            54                     4.9
##  3 Algeria               25               0            14                     0.7
##  4 Andorra              245             138           312                    12.4
##  5 Angola               217              57            45                     5.9
##  6 Antigua &~           102             128            45                     4.9
##  7 Argentina            193              25           221                     8.3
##  8 Armenia               21             179            11                     3.8
##  9 Australia            261              72           212                    10.4
## 10 Austria              279              75           191                     9.7
## # i 183 more rows
## # i abbreviated name: 1: total_litres_of_pure_alcohol
```

```
# creating drinks_smaller data frame
drinks_smaller <- drinks %>%
  filter(country %in% c("USA", "China", "Italy", "Saudi Arabia")) %>%
  select(-total_litres_of_pure_alcohol) %>%
  rename(beer = beer_servings, spirit = spirit_servings, wine = wine_servings)
drinks_smaller
```

```
## # A tibble: 4 x 4
##   country       beer spirit  wine
##   <chr>        <int>  <int> <int>
## 1 China           79    192     8
## 2 Italy           85     42   237
## 3 Saudi Arabia     0      5     0
## 4 USA            249    158    84
```

```
#drinks_smaller is not in tidy format, it is in wide format

#Convert to tidy:
#using pivot_longer from tidyr package which will make the column names their own variable
drinks_smaller_tidy <- drinks_smaller %>%
  pivot_longer(names_to = "type",
               values_to = "servings",
               cols = -country)
drinks_smaller_tidy
```

```
## # A tibble: 12 x 3
##    country      type   servings
##    <chr>        <chr>     <int>
##  1 China        beer         79
##  2 China        spirit      192
##  3 China        wine          8
##  4 Italy        beer         85
##  5 Italy        spirit       42
##  6 Italy        wine        237
##  7 Saudi Arabia beer          0
##  8 Saudi Arabia spirit        5
##  9 Saudi Arabia wine          0
```

```
## 10 USA          beer        249
## 11 USA          spirit      158
## 12 USA          wine         84
```

```
#airline saftey data
#preview data
airline_safety
```

```
## # A tibble: 56 x 9
##    airline           incl_reg_subsidiaries avail_seat_km_per_week incidents_85_99
##    <chr>             <lgl>                                 <dbl>           <int>
##  1 Aer Lingus        FALSE                             320906734               2
##  2 Aeroflot          TRUE                             1197672318              76
##  3 Aerolineas Arge~  FALSE                             385803648               6
##  4 Aeromexico        TRUE                              596871813               3
##  5 Air Canada        FALSE                            1865253802               2
##  6 Air France        FALSE                            3004002661              14
##  7 Air India         TRUE                              869253552               2
##  8 Air New Zealand   TRUE                              710174817               3
##  9 Alaska Airlines   TRUE                              965346773               5
## 10 Alitalia          FALSE                             698012498               7
## # i 46 more rows
## # i 5 more variables: fatal_accidents_85_99 <int>, fatalities_85_99 <int>,
## #   incidents_00_14 <int>, fatal_accidents_00_14 <int>, fatalities_00_14 <int>
```

```
#smaller airline safety
airline_safety_smaller <- airline_safety %>%
  select(airline, starts_with("fatalities"))
airline_safety_smaller
```

```
## # A tibble: 56 x 3
##    airline               fatalities_85_99 fatalities_00_14
##    <chr>                            <int>            <int>
##  1 Aer Lingus                           0                0
##  2 Aeroflot                           128               88
##  3 Aerolineas Argentinas                0                0
##  4 Aeromexico                          64                0
##  5 Air Canada                           0                0
##  6 Air France                          79              337
##  7 Air India                          329              158
##  8 Air New Zealand                      0                7
##  9 Alaska Airlines                      0               88
## 10 Alitalia                            50                0
## # i 46 more rows
```

```
airline_safety_tidy <- airline_safety_smaller %>%
  pivot_longer(names_to = "Year",
               values_to = "Count",
               cols = -airline)
airline_safety_tidy
```

```
## # A tibble: 112 x 3
```

```
##    airline               Year             Count
##    <chr>                 <chr>            <int>
##  1 Aer Lingus            fatalities_85_99     0
##  2 Aer Lingus            fatalities_00_14     0
##  3 Aeroflot              fatalities_85_99   128
##  4 Aeroflot              fatalities_00_14    88
##  5 Aerolineas Argentinas fatalities_85_99     0
##  6 Aerolineas Argentinas fatalities_00_14     0
##  7 Aeromexico            fatalities_85_99    64
##  8 Aeromexico            fatalities_00_14     0
##  9 Air Canada            fatalities_85_99     0
## 10 Air Canada            fatalities_00_14     0
## # i 102 more rows
```

```
#load dem_score data
dem_score <- read_csv("https://moderndive.com/data/dem_score.csv")
```

```
## Rows: 96 Columns: 10
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): country
## dbl (9): 1952, 1957, 1962, 1967, 1972, 1977, 1982, 1987, 1992
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
dem_score
```

```
## # A tibble: 96 x 10
##    country    `1952` `1957` `1962` `1967` `1972` `1977` `1982` `1987` `1992`
##    <chr>       <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 Albania        -9     -9     -9     -9     -9     -9     -9     -9      5
##  2 Argentina      -9     -1     -1     -9     -9     -9     -8      8      7
##  3 Armenia        -9     -7     -7     -7     -7     -7     -7     -7      7
##  4 Australia      10     10     10     10     10     10     10     10     10
##  5 Austria        10     10     10     10     10     10     10     10     10
##  6 Azerbaijan     -9     -7     -7     -7     -7     -7     -7     -7      1
##  7 Belarus        -9     -7     -7     -7     -7     -7     -7     -7      7
##  8 Belgium        10     10     10     10     10     10     10     10     10
##  9 Bhutan        -10    -10    -10    -10    -10    -10    -10    -10    -10
## 10 Bolivia        -4     -3     -3     -4     -7     -7      8      9      9
## # i 86 more rows
```

```
#data is not in tidy format

#convert to tidy
dem_tidy <- dem_score %>%
  pivot_longer(names_to = "year",
               values_to = "democracy_score",
               cols = -country,
               names_transform = list(year = as.integer))
```

```r
#view tidy data frame
dem_tidy
```

```
## # A tibble: 864 x 3
##    country    year democracy_score
##    <chr>     <int>           <dbl>
##  1 Albania    1952              -9
##  2 Albania    1957              -9
##  3 Albania    1962              -9
##  4 Albania    1967              -9
##  5 Albania    1972              -9
##  6 Albania    1977              -9
##  7 Albania    1982              -9
##  8 Albania    1987              -9
##  9 Albania    1992               5
## 10 Argentina  1952              -9
## # i 854 more rows
```