HPAM 7660 Data Assignment 1

Rebecca Letsinger

February 6, 2024

```
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(tidyr)
library(nycflights13)
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')
```

Question 2:

```
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, ~
```

Question 3:

```
?drinks
```

Question 4: In the context of R and data analysis, a data set is considered "tidy" if it follows a specific structure that makes it easy to perform data analysis and visualization. This includes column, rows, and observational units within those columns and rows. Wide data has each row representing a unique observation, but multiple variables are spread across columns. Long data has each variable and its corresponding values are stored in separate rows.

Question 5: This data is not tidy because it is not broken down into servings per spirit. The values spread throughout various columns. Making it tidy will show a more complete breakdown of the table and avoid over counting.

```
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
```

Question 6:

```
drinks_smaller %>%
  pivot_longer(names_to = "type",
               values_to = "servings",
               cols = beer:wine)
```

```
## # 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
```

Question 7:

```
View("drinks_smaller")
```

Question 8:

```
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_smaller_tidy <-airline_safety %>%
  pivot_longer(
    cols = c(fatalities_85_99,fatalities_00_14),
    names_to = "fatalities",
    values_to = "count" )
```

```
View("airline_safety_smaller_tidy")
```

Question 9:

```
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
```

Question 10:

```
View("dem_score")
```

Question 11: This is not tidy. If this were to be in a tidy format, the years would be separated in one column called which would describes all the years as one variable.