AE 08: Data import

Packages

We will use the following two packages in this application exercise.

- **tidyverse**: For data import, wrangling, and visualization.
- **readxl:** For importing data from Excel.

```
library(tidyverse)
library(readxl)
```

Part 1: Hollywood relationships

- **Demo:** Load the data from https://sta199-s24.github.io/data/age_gaps.csv and assign it to `age_gaps`. Confirm that this new object appears in your Environment tab.

```
# add code here
```

- **Your turn (5 minutes):** Split the data into three – where woman is older, where man is older, where they are the same age. Save these subsets as two appropriately named data frames. *Remember:* Use concise and evocative names. Confirm that these new objects appear in your Environment tab and that the sum of the number of observations in the two new data frames add to the number of observations in the original data frame.

```
# add code here
```

- **Demo:** Write out the three new datasets you created into the `data` folder:

```
# add code here
```

Part 2: Sales

Sales data are stored in an Excel file that looks like the following:

	A	B	C	D	E	F	G	H	I
1	This file contains information on sales.								
2	Data are organized by brand name, and for each brand we have the ID number for item sold, and how many are sold.								
3									
4									
5	Brand 1	n							
6	1234	8							
7	8721	2							
8	1822	3							
9	Brand 2	n							
10	3333	1							
11	2156	3							
12	3987	6							
13	3216	5							

- **Demo:** Read in the Excel file called `sales.xlsx` from the `data-raw/` folder such that it looks like the following.

```
# A tibble: 9 x 2
  id      n
  <chr>   <chr>
1 Brand 1 n
2 1234    8
3 8721    2
4 1822    3
5 Brand 2 n
6 3333    1
7 2156    3
8 3987    6
9 3216    5
```

```
# add code here
```

- **Demo - Stretch goal:** Manipulate the sales data such such that it looks like the following.

```
# A tibble: 7 x 3
  brand      id     n
  <chr>   <dbl> <dbl>
1 Brand 1  1234     8
2 Brand 1  8721     2
3 Brand 1  1822     3
4 Brand 2  3333     1
5 Brand 2  2156     3
6 Brand 2  3987     6
7 Brand 2  3216     5
```

```
# add code here
```

- **Question:** Why should we bother with writing code for reading the data in by skipping columns and assigning variable names as well as cleaning it up in multiple steps instead of opening the Excel file and editing the data in there to prepare it for a clean import?

Add response here.