Final Project Data

AUTHOR
Tan Vo

```
life_expectancy <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/ti
life_expectancy_different_ages <- readr::read_csv('https://raw.githubusercontent.com/rfo
life_expectancy_female_male <- readr::read_csv('https://raw.githubusercontent.com/rforda
skim(life_expectancy)
```

Data summary

Name	life_expectancy
Number of rows	20755
Number of columns	4
_____	
Column type frequency:	
character	2
numeric	2
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Entity	0	1.00	4	59	0	261	0
Code	1694	0.92	3	8	0	238	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Year	0	1	1975.73	38.42	1543	1961.00	1981.00	2001.00	2021.00	▁▁▁▁▇
LifeExpectancy	0	1	61.62	12.92	12	52.19	64.06	71.82	86.54	▁▂▅▇▃

```
skim(life_expectancy_different_ages)
```

Data summary

Name	life_expectancy_different...
Number of rows	20755
Number of columns	9
_____	
Column type frequency:	
character	2
numeric	7
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Entity	0	1.00	4	59	0	261	0
Code	1694	0.92	3	8	0	238	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Year	0	1.00	1975.73	38.42	1543.00	1961.00	1981.00	2001.00	2021.00	▁▁▁▁▇
LifeExpectancy0	0	1.00	61.62	12.92	12.00	52.19	64.06	71.82	86.54	▁▂▅▇▃
LifeExpectancy10	833	0.96	68.82	7.08	22.11	63.63	69.49	73.95	86.96	▁▁▅▇▂
LifeExpectancy25	833	0.96	70.46	5.89	31.16	66.07	70.71	74.67	87.07	▁▁▃▇▂
LifeExpectancy45	833	0.96	73.54	4.07	50.03	70.51	73.22	76.21	87.31	▁▁▇▇▂
LifeExpectancy65	833	0.96	78.75	2.51	68.74	76.91	78.31	80.19	88.35	▁▂▅▇▂
LifeExpectancy80	833	0.96	86.09	1.30	82.10	85.20	85.80	86.72	91.57	▁▃▇▇▂

```
skim(life_expectancy_female_male)
```

Data summary

Name	life_expectancy_female_ma...
Number of rows	19922
Number of columns	4
_____	
Column type frequency:	
character	2
numeric	2
_____	
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Entity	0	1.00	4	59	0	259	0
Code	1600	0.92	3	8	0	237	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	his
Year	0	1	1978.82	32.81	1751.00	1963.00	1983.0	2002.00	2021.00	▁▁▁▁▇
LifeExpectancyDiffFM	0	1	4.82	2.38	-9.88	3.23	4.6	6.22	29.55	▁▃▇▁▁

- **DATA EVALUATION**

Is the data in the right format for those mappings? - Yes.

Determine if the variables you have are the appropriate variable type. - Yes, it is appropriate.

Does something you want to use as a numeric variable look like a character string? - No.

See if there are any missing values, and if there are, if they are coded correctly. - No missing value.

Are the variable names usable? - Yes.

Do they have spaces in them that will make it hard to refer to them in R? - No.

For categorical variables, consider the categories. Are they readable? - Yes.

Are they understandable phrases or cryptic codes? - They are understandable phrases.

Are there the right number of categories, or will you want to reduce them? - They are correct but might consider to group them into different groups.

- **DATA WRANGLING**

1. **Average Life Expectancy from 1950 to 2021**

```
avg_life_expectancy <- life_expectancy |>
  filter(Year >= 1950 & Year <= 2021, !is.na(Code), Code != "OWID_WRL") |>
  group_by(Entity, Code) |>
  summarize(AvgLifeExpectancy = mean(LifeExpectancy, na.rm = TRUE)) |>
  ungroup()
```

```
`summarise()` has grouped output by 'Entity'. You can override using the
`.groups` argument.
```

Export the filtered data as a CSV file

```
write_csv(avg_life_expectancy, "avg_life_expectancy.csv")
```

2. **Average Life Expectancy Difference between Female & Male from 1950 to 2021**

```
avg_life_expectancy_diff_female_male <- life_expectancy_female_male |>
  filter(Year >= 1950 & Year <= 2021, !is.na(Code), Code != "OWID_WRL") |>
  group_by(Entity, Code) |>
  summarize(AvgLifeExpectancyDiffFM = mean(LifeExpectancyDiffFM, na.rm = TRUE)) |>
  ungroup()
```

```
`summarise()` has grouped output by 'Entity'. You can override using the
`.groups` argument.
```

Export the filtered data as a CSV file

```
write_csv(avg_life_expectancy_diff_female_male, "avg_life_expectancy_diff_female_male.cs
```