Tidy Data

Tidy Data: refers to a specific way of organizing data tables in a tabular format to facilitate data analysis. In tidy data:

- Each variable forms a column.
- Each observation forms a row.

Pivots: data often comes in various formats, and its structure might not be ideal for the task at hand. Pivoting helps you reorganize your data to a format that makes it easier to analyze, visualize, or model. For example, you might need to pivot data to:

- ***Perform Aggregations:*** Pivoting can simplify data aggregation and summary calculations.
- ***Create Visualizations:*** Some visualizations require specific data structures.
- ***Fit Models:*** Certain modeling techniques might require data in a specific shape.

Long Data Vs. Wide Data

Long Format		
Airport	**Year**	**Number_Flights**
JFK	2017	200
JFK	2018	300
JFK	2019	400
ORD	2017	300
ORD	2018	250
ORD	2019	600
SJU	2017	440
SJU	2018	201
SJU	2019	300

Wide Format			
Airport	**2017**	**2018**	**2019**
JFK	200	300	400
ORD	300	250	600
SJU	440	201	300

Pivot Longer

`pivot_longer()`: is used to convert data from a wide format (with multiple columns) into a long format (fewer columns). It's particularly useful when you have variables spread across different columns and you want to stack them into a single column, often with corresponding values.

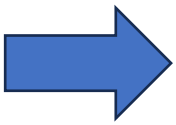
Command Illustration:

```
new_dataframe <- wide_dataframe %>%
      pivot_longer(cols     =     -column_name,     names_to     =
      "name_column1",    values_to = "name_column2")
```

- `wide_dataframe:` The input data frame.
- `cols:` Columns to pivot into longer format.
- `names_to:` The name of the column that will store the column names from the original wide format.
- `values_to:` The name of the column that will store the values from the original wide format.

Example: Write the code to transform the wide data into long data.

Wide Format			
Airport	**2017**	**2018**	**2019**
JFK	200	300	400
ORD	300	250	600
SJU	440	201	300



Long Format		
Airport	**Year**	**Number_Flights**
JFK	2017	200
JFK	2018	300
JFK	2019	400
ORD	2017	300
ORD	2018	250
ORD	2019	600
SJU	2017	440
SJU	2018	201
SJU	2019	300

Example 1: Use `pivot_longer` to reshape the following wide data

Country	2018	2019	2020
USA	250	260	270
Canada	180	190	200

First write out what the output should look like and then write the code that should make that happen.

Code:

Pivot Wider

`pivot_wider()`: is used to convert data from a long format to a wide format. It's useful when you want to take distinct values from a column and spread them across new columns.

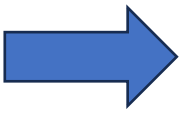
Command Illustration:

```
new_dataframe <- long_dataframe %>%
        pivot_wider(names_from = column_name1, values_from = column_name2)
```

- `long_dataframe`: The input data frame.
- `names_from`: The column that contains the unique values to be transformed into column names in the wide format.
- `values_from`: The column that contains the values to be spread across the new wide-format columns.

Example: Write the code to transform the long data into wide data.

Long Format		
Airport	**Year**	**Number_Flights**
JFK	2017	200
JFK	2018	300
JFK	2019	400
ORD	2017	300
ORD	2018	250
ORD	2019	600
SJU	2017	440
SJU	2018	201
SJU	2019	300



Wide Format			
Airport	**2017**	**2018**	**2019**
JFK	200	300	400
ORD	300	250	600
SJU	440	201	300

Example 2: Use `pivot_wider` to reshape the data

Country	Year	Value
USA	2018	250
Canada	2018	260
USA	2019	270
Canada	2019	180
USA	2020	190
Canada	2020	200

First write out what the output should look like and then write the code that should make that happen.

Code: