Section 1:
International Development Data

Exploring Data with Pivot Tables

Import Data

- Download the 2014 HDI `csv` and the Continents `csv` from the Lectures tab on the course home page
- Create a new Google Sheet
- Import the HDI `csv` file with ‘Replace existing sheet’ selected
- Import the Continents `csv` file with ‘Insert new sheet(s)’ selected

Explore Data

- 2014 HDI includes countries' Human Development Index (HDI), a composite value of life expectancy, education, and per-capita income.
- Continents includes each country's region and sub-region
- Countries are also ranked by HDI and grouped into 4 tiers of development (very high, high, medium, and low)

Create your own Pivot Table

- Group data by region
- What is the average HDI by region?
- What is the average life expectancy by region?
- What is the average education level by region?
- Etc.
- And likewise, by subregion

Create your own Pivot Table: VLOOKUP

- We want to know the region for each country in ‘2014 HDI’
- We can *join* the tables based on the names of the countries using the `VLOOKUP` function
- `VLOOKUP` searches the first column of a block of data for a given value, and when it finds it, it returns the corresponding value in some other column
- For example, if it finds the value in row *i*, and the specified return column is *j*, it returns the (*i*,*j*) value in the data table

Create your own Pivot Table: VLOOKUP

- The form of VLOOKUP is `=VLOOKUP(search_value, search_range, column, sorted)`
- Let’s first see how we can add the `Region` for "Norway", the first country listed.
- In cell `I2` of 2014 HDI, type
  `=VLOOKUP($A2, Continents!$A$1:$C$250, COLUMN(B2), 0)`
- This formula searches for the value in cell `A2` ("Norway") in the range `$A1:C250` in the sheet Continents.
- It evaluates to the value in column B (i.e., the "Region" column) in the row where it finds "Norway" (row 167), namely "Europe".
- Copy the contents of cell `I2` down to the bottom of the data table.
  Name the column `Region`.

Create your own Pivot Table: Data Cleaning

- You will notice that some of the results are `#N/A`. This error indicates that the country name in 2014HDI was not found in Continents.
- There are (at least) two possible causes for this error:

    1. The same country is entered using two different names in the two sheets (e.g., United States and United States of America).

    2. The data tables are of different lengths, because one has data for more countries than the other.

- In our case, there are differences in the spellings of the country's names, which we can correct easily. But if we cannot correct the errors easily …

Create your own Pivot Table: Data Cleaning

- To remove any remaining `#N/A`s (after making any easy fixes), we will wrap our `VLOOKUP` formula in an `IFERROR` function that displays an `alternative_value` if there is an error.
- Let's make our alternative_value `""`, meaning the empty string
- Replace the contents of cell `I2` with the formula `=IFERROR(VLOOKUP($A2, Continents!$A$1:$C$250, COLUMN(B2), 0), "")` and copy this formula down to the bottom of the data table

Average HDI Grouped By Region

- Select a cell in the data table
- Go to ‘Data’ → ‘Pivot Table’
- Set ‘Insert’ to ‘New Sheet’ and click ‘Create’
- In the Pivot Table editor on the right, select ‘Region’ under ‘Rows’ and ‘Human Development Index’ as ‘Values’
- The default setting is ‘SUM’.
  Instead, choose ‘AVERAGE’ in the drop down menu.
- Your pivot table should now display the average HDI per region and the overall average

Average HDI Grouped By Subegion

- How do the regions vary by average HDI?
- Add Subregion as another row: In the Pivot Table editor on the right, add ‘Subregion’ under ‘Rows’
- How do the subregions vary by average HDI?
- Discuss your findings with your neighbor.