Week 13 Tutorial Worksheet

AY23/24 Semester 2

No submission required

Question 1 HDB resale transactions

In this question, we want to verify some findings from a Straits Times article on April 8, 2024, titled HDB resale prices rise 0.3% in March, fewer units sold.

Download the raw data directly from the link below.

https://beta.data.gov.sg/datasets/d_8b84c4ee58e3cfc0ece0d773c8ca6abc/view.

1. Import the CSV into `R`. Use the data to verify the following claims from the news article.

   Note: In some cases, you will find that the article did not precisely express the metrics used in its analyses. Try a few simple things (e.g., mean, median) and verify the results.

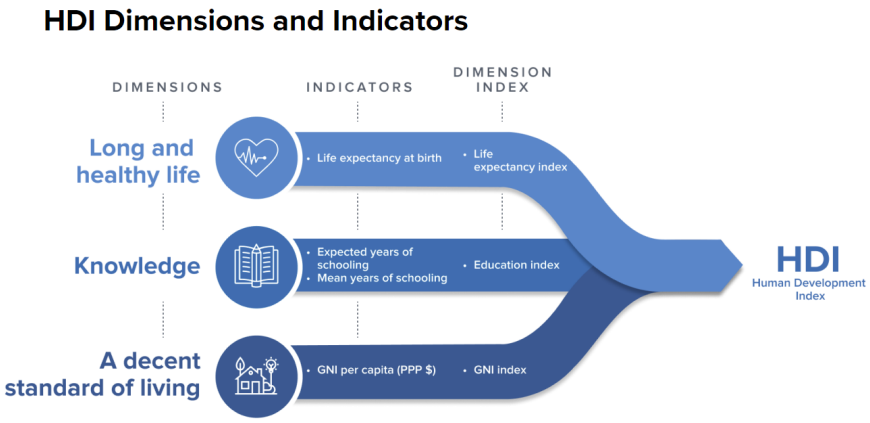
(1) In 2024, Sixty-one flats changed hands for at least $1 million in March, higher than the 50 such transactions in February. TRUE/FALSE

(2) HDB resale prices rose by 3.8 per cent between September 2023 and March 2024, higher than the 2.3 per cent increase in the preceding six months from March to September 2023. TRUE/FALSE

2. Suppose you are asked to investigate whether there is any effect of the Circuit Breaker (from April to June in 2020) on the housing market. Create a visualization in `ggplot2` to explore and answer this question.

Question 2 Human Development Index

The data sets for this question are from the UN's annual Human Development Index (HDI). The index is a combination of three sub-indices in health, education, and wealth. The HDI values ranges between 0 and 1: A value closer to 1 indicates that the country offers generally decent healthcare, education, and standard of living. The infographic below describes the dimensions and indicators for HDI.



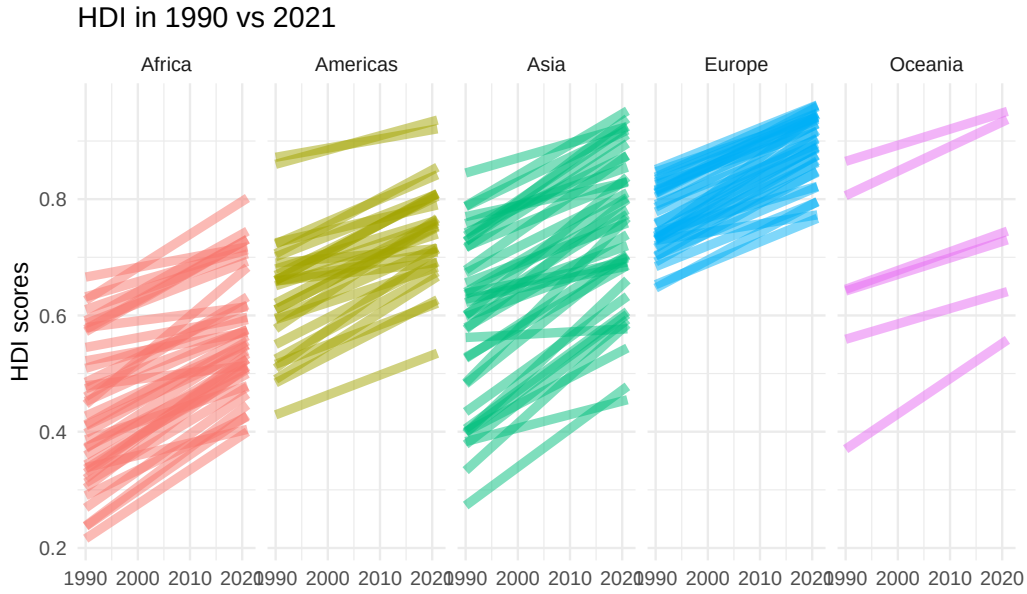
The data set `UN_HDI_2021.xlsx` contains the HDI scores for 191 countries in 1990 and 2021. The data set `UN_regions.csv` groups the countries into broad geographical regions in the world.

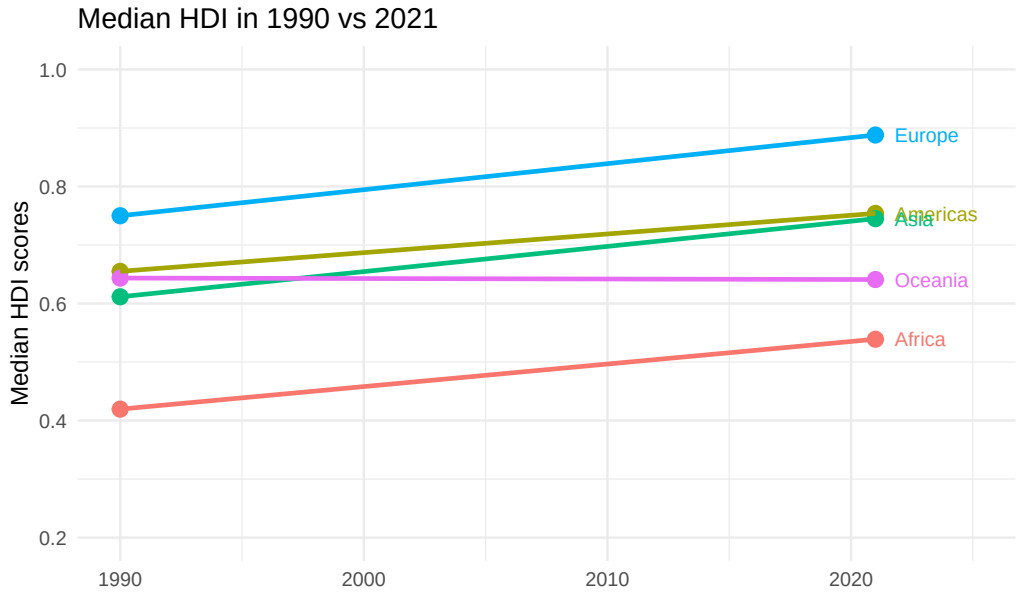
1. Read in the HDI data and convert it into a tidy format. Compute the rankings in HDI scores received in 2021. Then extract observations for the **top three and bottom three** countries. Save it as an object named `qn2_1`. Your `qn2_1` should take a structure similar to the following:

```
qn2_1
```

```
## # A tibble: 6 x 4
##   Country      year   hdi  rank
##   <chr>       <int> <dbl> <int>
## 1 Switzerland  2021 0.962     1
## 2 Norway       2021 0.961     2
## 3 Iceland      2021 0.959     3
## 4 Niger        2021 0.4     189
## 5 Chad         2021 0.394   190
## 6 South Sudan  2021 0.385   191
```

2. Suppose you are asked to analyze how HDI scores are distributed across different continents. Create a visualization using `ggplot2`. Below are two possible designs; feel free to develop a visualization that differs from these examples.





Requirements

- Your code should generate a data frame named `qn2_1`.
- The knitted HTML should contain two plots, one each for Question 1.2 and Question 2.2.