Week 6 Tutorial Worksheet

AY23/24 Semester 2

Submission: End of tutorial day

Question 1. YRBSS questionnaires

In this question, we continue to use the YRBSS data from the last week. This time, we will practice data manipulation using **`dplyr` functions**.

1. Read data into `R` as an object named `qn1_1`. Conduct the following tasks and overwrite `qn1_1` with the resulting data frame.
   - Remove rows with missing values (if any).
   - Remove duplicated rows (if any).
   - Rename the columns `record` as `id`, and `stweight` as `weight_kg`.
   - Convert the `grade` variable into numeric.
2. Continue working on the cleaned data in `qn1_1`. Subset female youth with BMI lower than 15 and then extract the following columns: `id`, `age`, `race4`, `weight_kg`, and `bmi`. Create a column `height_m` based on the BMI formula:

$$BMI = weight(kg)/height^2(m)$$

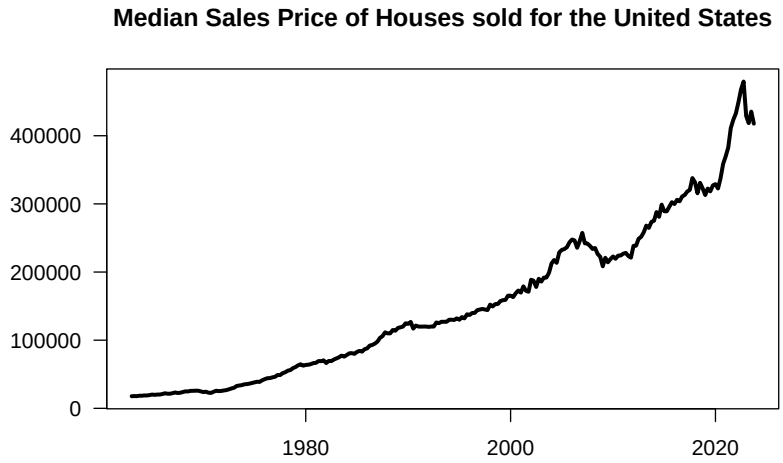
   Store your result in an object named `qn1_2`.

Question 2. Median housing price in the US

The data set for this question comes from the Federal Reserve Bank of St. Louis (FRED): https://fred.stlouisfed.org/series/MSPUS.

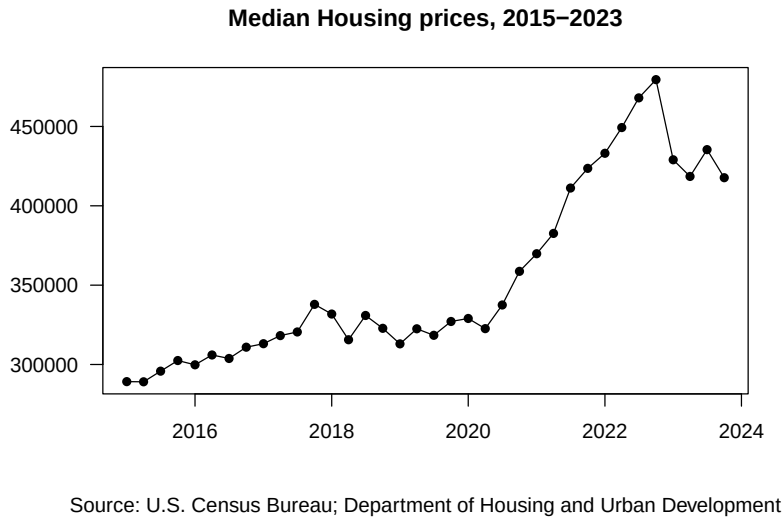
The data set is available as `MSPUS.xls` on Canvas.

1. Read in data as `qn2_1`. Rename the variables as `date` and `price`. Convert `date` as a Date type variable.
2. Re-create, as much as you can, the following visualization in base `R`.



3. Create a subset of `qn2_1` for data from Jan 2015 to Oct 2023. Save it as a new object named `qn2_3`. In the new data frame, create two columns: `year` and `quarter`. The variable `year` is the year of the `date` variable, and `quarter` takes values `Q1, Q2, Q3, Q4` based on the variable `date`.

4. Create the following visualization with `qn2_3`.



Question 3. New York flights data

In this question, we continue to use the `nycflights13::flights` data from Week 5's lecture.

Use **`dplyr` functions** for the questions and fill in the blanks.

1. Which destination airport received the most flights from New York in June?

   Airport code:

   Number of flights:

2. Which carrier had the greatest average distance per flights?

   Flight carrier code:

3. Which flight traveled the fastest (highest average speed in miles per hour)?

   Flight carrier code:

   Flight number:

4. What day of the week saw the most flights?

   Day of week:

   Number of flights:

Requirements

- After answering all questions in the `Rmd` file, hit the **Knit** button. Make sure your `Rmd` can knit to HTML.
- The code in your `Rmd` file should create the following objects: `qn1_1`, `qn1_2`, `qn2_1`, and `qn2_3`.
- The knitted HTML file should contain two graphs based on the housing price.
- **Submit your `Rmd` file to Canvas after your tutorial session.**
  - This is the last time we check your `Rmd` file before the midterm exam. Reach out to your tutor as soon as possible if you are still unsure about our submission requirements.