**Data Wrangling**

**Exercises**

1. Some sets have missing information for retail\_price or pieces or both. This could be because the sets are free (giveaways), they aren’t traditional lego sets (comic books, etc) or just because the information is missing. Filter the lego dataset based on the specifications below and save the result as lego using <-. Hence, you will overwrite the original lego object. In addition, describe the implications of removing these sets.

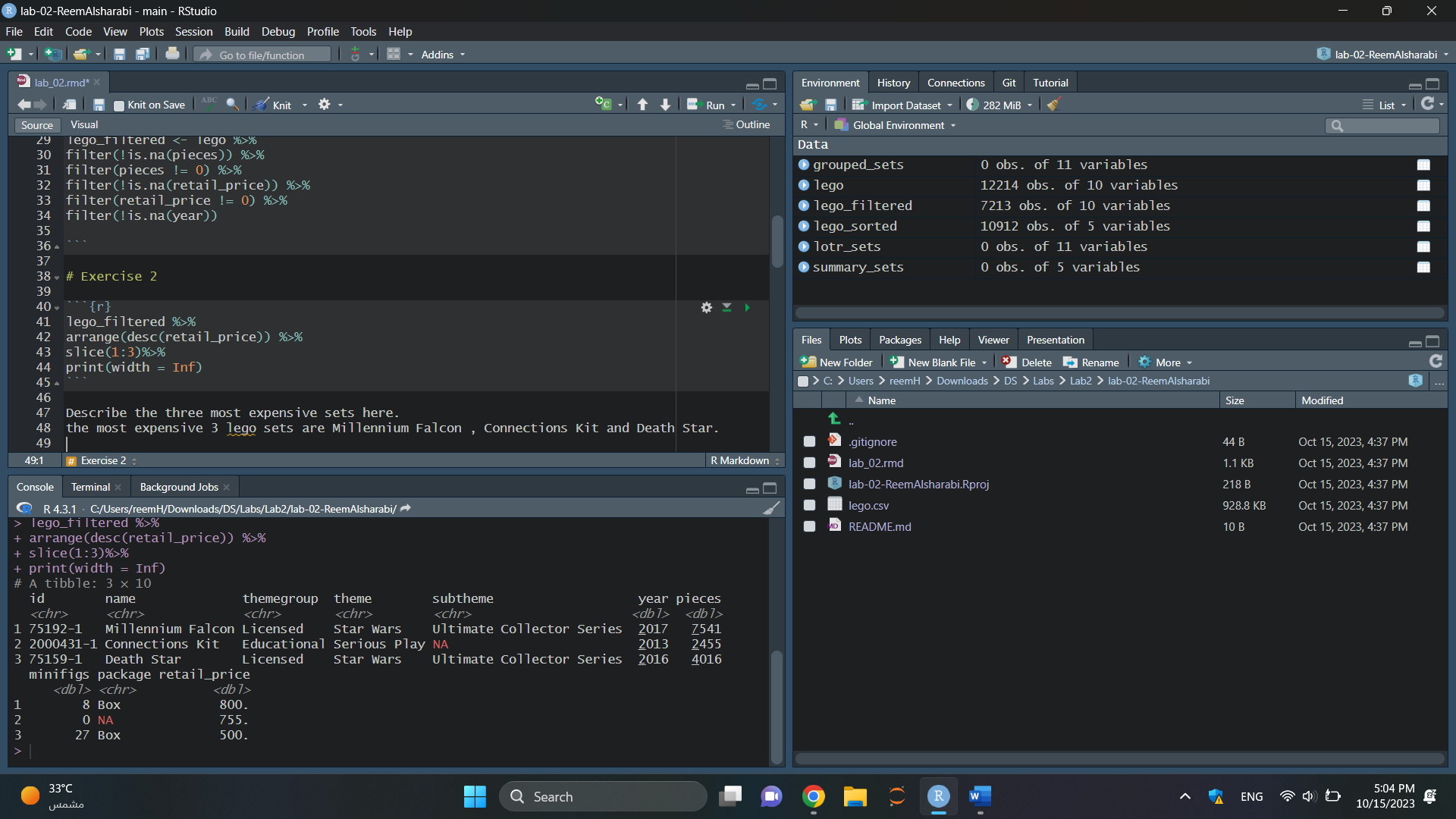
Your new lego tibble (data frame) should have:

* + no missing pieces
  + only contain sets with a nonzero number of pieces
  + no missing retail\_price
  + only contain sets with a nonzero retail\_price
  + no missing year

A screenshot of a computer

Description automatically generated

1. Arrange the dataset in descending order of *retail\_price* and print the first three rows. Report in words the names of the three most expensive lego sets, their prices, and how many pieces each has.



* Set: Millennium Falcon

Price: $800

Pieces: 7,541

* Set: Connections Kit

Price: $755

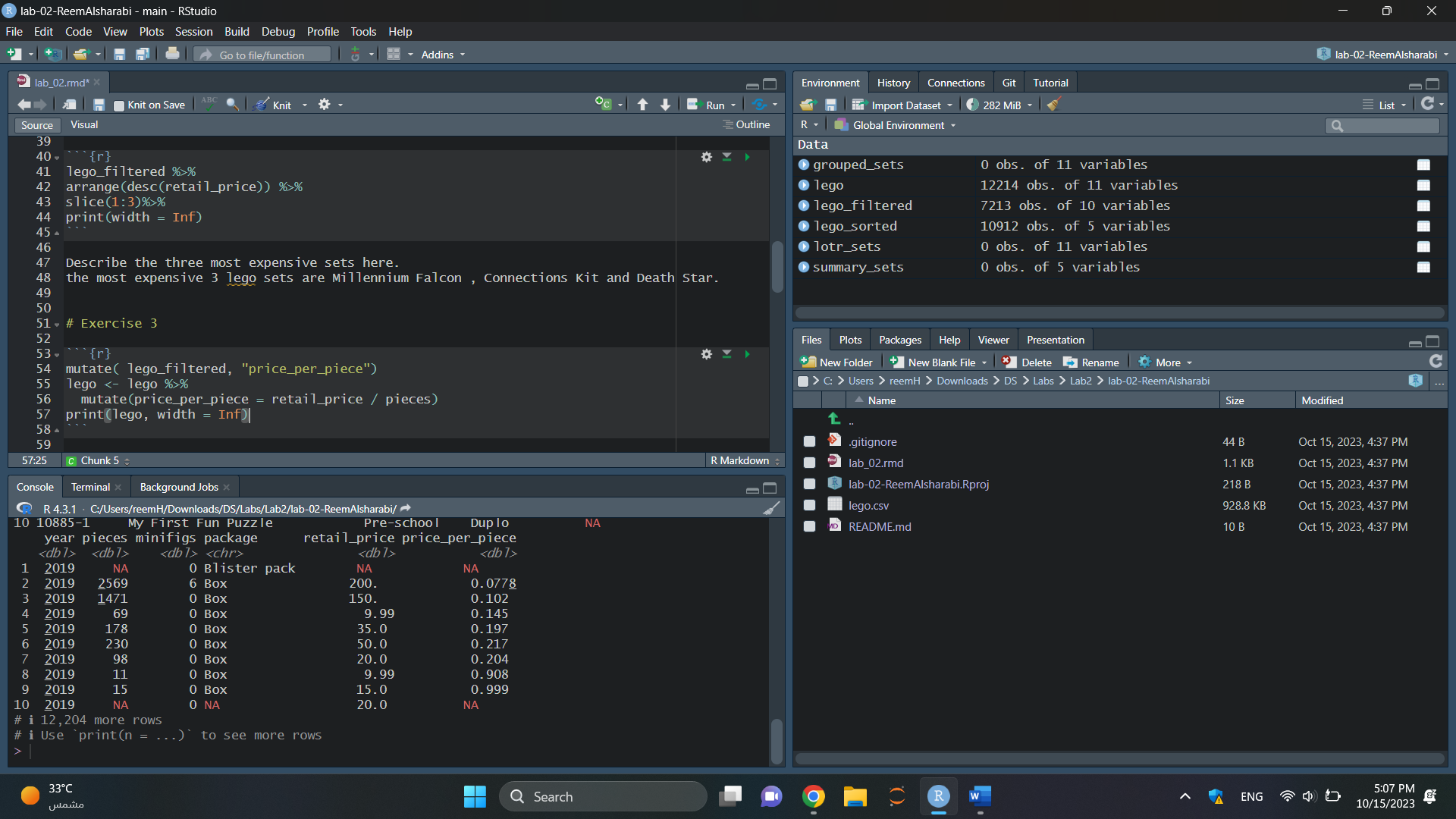
Pieces: 2,455

* Set: Death Star

Price: $500

Pieces: 4,016

1. It appears that the most expensive sets generally have more pieces. Use *mutate()* to create a new variable *price\_per\_piece*, representing the price in dollars per piece for each of the sets. Save the result as *lego*. Hence, you will overwrite the current *lego* object.



1. Arrange the *lego* dataset in descending order of *price\_per\_piece* and return only the columns *name, themegroup, theme, pieces, price\_per\_piece*, and the first five rows. What do you notice about these sets?

A screenshot of a computer

Description automatically generated

* The sets generally have a small number of pieces, ranging from 2 to 3.
* Despite their small piece counts, these sets have a higher price per piece, ranging from $15 to $29.8.
* The sets cover diverse themes, including Educational, Licensed (Star Wars), Basic, and Education.
* These sets may possess unique features, specialized designs, educational value, or collectible qualities that justify their higher price per piece.

Overall, these observations suggest that these sets with small piece counts and higher prices per piece offer distinctive qualities or cater to specific purposes, making them stand out within the LEGO product range.

1. Return a tibble containing the cheapest and most expensive lego sets (based on *retail\_price*) in each subtheme, considering only sets with the Lord of the Rings theme.

A screenshot of a computer

Description automatically generated

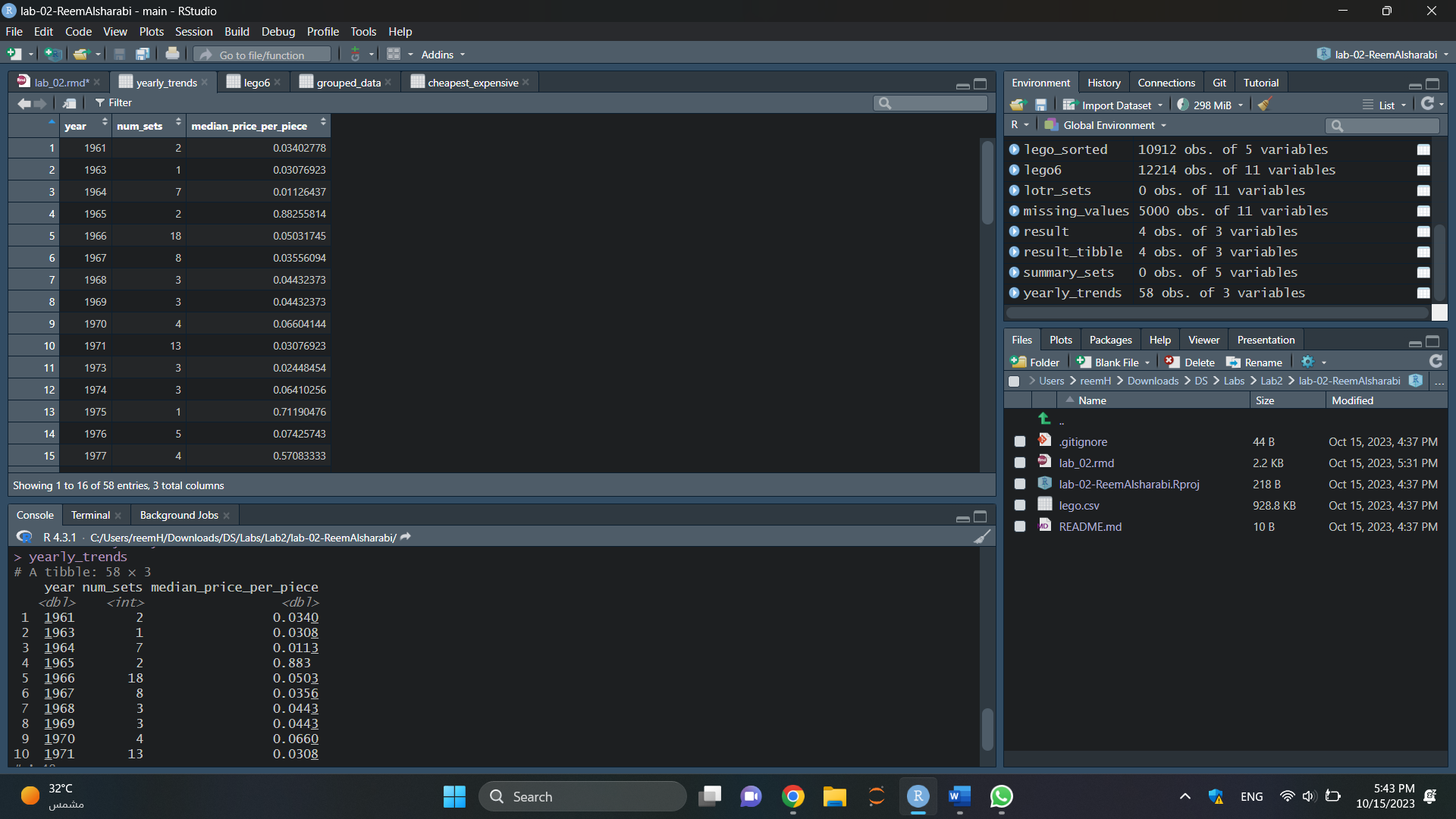
A screenshot of a computer

Description automatically generated

1. Use *group\_by()* and *summarize()* to create a new tibble with one row for each year, and columns for the year, the number of sets released in that year, and the median price per piece for sets from that year. Save this resulting tibble as an object named *yearly\_trends*.

A screenshot of a computer

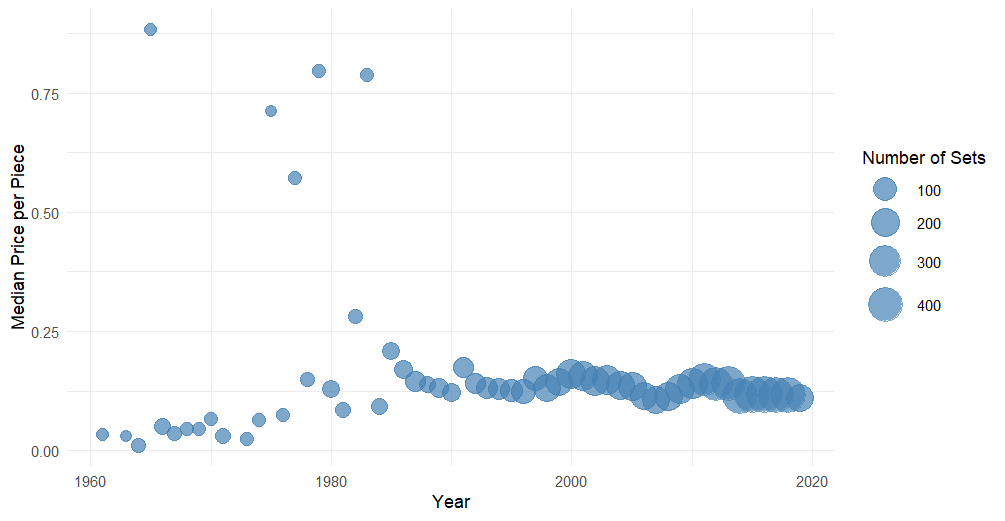
Description automatically generated



1. Create a plot of the median price per piece over time using the *yearly\_trends* tibble. Size points according to the number of sets produced in that year. Adjust transparency, color, etc as appropriate and remember the principles of effective data visualization. Comment on what you observe.

* From 1960 to 1980: The median price per piece was consistently low, below 0.25, with a few high outliers indicating occasional sets with higher prices per piece.
* From 1980 to 1990: The median price per piece showed an increase and approached the 0.25 threshold, suggesting that the average price per piece became more aligned with this value during this decade.
* From 1990 to 2020: The median price per piece remained relatively low and did not exceed 0.25. There were no outliers during this period, indicating a consistent price per piece for LEGO sets without extremely high or low values.

These findings suggest that LEGO sets have generally maintained an affordable price per piece over the years, with some fluctuations and a trend toward closer alignment with the 0.25 price point.



A screenshot of a computer

Description automatically generated

**Submission**

Knit to PDF to create a PDF document. Stage and commit all remaining changes, and push your work to GitHub. Make sure all files are updated on your GitHub repo. Only upload your PDF document to Blackboard. Before you submit the uploaded document, mark where each answer is to the exercises.