

## Material covered: 7. Intro to Tidiverse and Tidying Data

8. Data Manipulation

---

Follow the directions below to complete the assignment. Once you have completed the assignment, submit your answer on Canvas before the due date.

**Submission Guidelines:**

- Due date:* July 22, 2025 at 11:15 am
- Submission means:* You will submit an .R script containing all code, comments and explanations (remember to use # at the beginning of your comment) for each question in order to receive credit for your work.
- Submission details:* Students must make sure that their answers are readable, complete, and submitted before the deadline. Late submission for ANY reason, whether in part or in whole, or unreadable files will trigger the late penalties. The time when I receive the last part of your answers will count when computing late penalties, not the time when you sent the answers.
- Late Penalties:* If an assignment is turned immediately after the due date, your score will be reduced by 10% of the maximum assignment grade. Additional 10% will be deducted from your score for every 24 hours after that.
- Academic Integrity:* This is an individual assignment. Students need to submit their own answers and understand those answers. I reserve the right to ask you to walk me through the code for your assignment. If you fail to explain the code in detail and the choices you made, you will receive 0 points for the part of the assignment. Students who refuse to provide the explanation for their work will receive 0 points for the entire assignment.
- Using code not discussed in class:* Students who are submitting code not discussed in the class will temporarily receive 0 points for that part of the assignment until they meet with me and they walk me through their code. Students need to meet with me within one week of returning the graded assignment, no later than August 19. If you fail to meet with me or if you cannot explain some of the code & the choices you made, you will keep the 0 points for that part of the assignment.

**Homework Questions to Answer:**

The first data file, *grades\_mess.csv* (here is the zip), is an artificial data created for the purposes of this homework.

The second file, *Sold\_Fleet\_Equipment.csv* (here is the zip) includes sales data for City of Seattle fleet equipment that was sold between 2019 – 2025 (it excludes Seattle City Light (SCL) fleet equipment data).

(Select) Variables:

<b>EquipID</b>	Equipment ID Number
<b>model_year</b>	Model Year
<b>make</b>	Equipment Manufacturer
<b>model</b>	Equipment Model
<b>description</b>	Equipment purpose or description
<b>dept</b>	Department previously assigned this equipment
<b>sale_price</b>	Equipment's sale price
<b>sale_year</b>	Equipment's sale year
<b>vin</b>	Vehicle Identification Number
<b>equipment_type</b>	Category of equipment
<b>fuel_type</b>	Primary fuel type

*Part 1: Reshape your data (25 pts)*

1. (0 pts) Set your working directory
2. (0 pts) Load the *tidyverse* package in your library.
3. (5 pts) Import the *grades\_mess.csv* file in R and create a data frame called *grades\_df*
4. (20 pts) Use the functions learned in Module 7 to re-shape this data frame so it looks like the picture below. Use the pipe operator to chain all operations together in a concise way. Make sure that the variable names, values and the order of the columns is correct. Save the tidy data frame as *grades\_df*. Remember to show your code.

	student	score	letter	test_date	subject	instructor
1	Alice	88	B	05-20-2023	Math	Mr. Lin
2	Bob	75	C	04-15-2023	Biology	Dr. Chen
3	Cara	92	A	06-22-2023	History	Ms. Lopez
4	Dan	65	D	05-10-2023	English	Mr. Adams
5	Eva	79	C	04-18-2023	Chemistry	Dr. Kim
6	Finn	84	B	06-25-2023	Art	Mrs. Singh

*Part 2: Data Manipulation and Insights (75 pts)*

5. (5 pts) Import the *Sold\_Fleet\_Equipment.csv* data and create a data frame called *eq\_df*.
6. (5 pts) Rename the variable *EquipID* as *equip\_id*. Save your changes (overwrite the data frame).
7. (5 pts) Change the order of the variables. Move the *sale\_year* variable to appear right after the *model\_year* variable. Save your changes (overwrite the data frame).
8. (10 pts) How many pieces of equipment were sold for more than \$10,000? (*Hint: use filter() and summarise() functions*)
9. (10 pts) What are the top 3 most expensive pieces of equipment sold? Show *make*, *model*, and *sale\_price* in your final answer. *Hint: Use slice\_max() and select().*
10. (10 pts) Create a new variable called *age\_at\_sale* that calculates how old each piece of equipment was at the time of sale (subtract the *model\_year* from the *sale\_year*). Save your changes (overwrite the data frame).
11. (10 pts) What are the top 3 departments (dept) with the highest total revenue from equipment sales? *Hint: group, summarize, and use slice\_max().*
12. (10 pts) Which equipment type had the lowest average sale price?
13. (10 pts) For each fuel type, how many pieces of equipment were sold for more than \$15,000?