

Session 6: Data Import — Pen-and-Paper Pair Exercise

PSY 410 | Data Science for Psychology

Name: _____ Date: _____

No laptop today? No problem. This handout lets you practice the same skills on paper. Work with a partner who has a laptop and compare your work at the end.

The data: messy_survey.csv

Here is the raw content of the CSV file. Imagine opening it in a text editor:

```
Participant ID, Age (years), Gender, Score_1, Score_2, Notes
P001, 22, Female, 4, 3,
P002, 25, Male, 3, N/A, transfer student
P003, -999, Female, 5, 4,
P004, 28, N/A, 2, 3,
P005, 31, Male, N/A, 2,
```

The task (same as the slide exercise)

1. Read it in with `read_csv()` — what types does R guess?
2. Fix the missing values: `-999` and `"N/A"` should be `NA`
3. Rename the columns to clean snake_case names
4. What does `problems()` show you?

Your pen-and-paper version

Step 1: Predict column types. Look at each column and write what type R will guess:

Column name	Example values	R will guess this type
Participant ID	P001, P002	
Age (years)	22, 25, -999	
Gender	Female, Male, N/A	
Score_1	4, 3, 5, 2, N/A	

Column name	Example values	R will guess this type
Score_2	3, N/A, 4, 3, 2	
Notes	"", transfer student	

Step 2: Spot the problems. Circle every value in the raw data above that should be NA. How many did you find? _____

List them: - -999 in _____ column — this is a common code for _____
 - N/A in _____ column(s) — R doesn't automatically recognize this as _____

Step 3: What are good snake_case names? Rewrite each column name:

Original name	Clean snake_case name
Participant ID	
Age (years)	
Gender	
Score_1	
Score_2	
Notes	

Step 4: Write the code. Fill in the blanks:

```

messy <- read_csv("messy_survey.csv", na = c("_____", "_____", ""))

messy_clean <- messy |>
  rename(
    _____ = `Participant ID`,
    _____ = `Age (years)`,
    _____ = Gender,
    _____ = Score_1,
    _____ = Score_2,
    _____ = Notes
  )

```

Step 5: Think about it. After telling `read_csv()` that "N/A" is missing, will R still guess `Score_1` as character or will it now guess numeric? Why?

Your answer: _____

Check your work

Compare your answers with your partner's screen. Do your type guesses, missing values, and code match?