

**Problem Set 5: Managing Data**

*You may not use any downloaded packages. No paper submission is required for this assignment. See the instructions in Problem Set 2 regarding the submission of R code.*

1. The file `ps05.txt` contains the names (`name`) and the email addresses (`email`) of 24 students enrolled in MATH 267P, as well as the groups they were assigned to for the first two problem sets (`ps1` and `ps2`, respectively). Import the data into a data frame named `psgrps`.
  - a. [5 pts] Extract the indices of the students who did not provide a last name and store the result as a vector named `no.last`.  
Hint: The function `strsplit` might be useful.
  - b. [5 pts] Create a two-column matrix named `fullname`, where the first column of the matrix contains the first names of the students and the second column contains the last names of the students. Indicate a missing last name using the special value `NA`.
  - c. [5 pts] Create a data frame named `psgrps.x` that contains the first names, last names, email addresses, and the groups they were assigned to for the first two problem sets. The data frame should be sorted as follows: students who did not provide a last name are sorted by the first names in ascending order, followed by the remaining students sorted in ascending order of their last names. The following shows the first ten rows of `psgrps.x`.

---

R Console

---

```
> print(psgrps.x, row.names=FALSE)
```

first	last	email	ps1	ps2
Applejack	<NA>	applejackponyville.edu	E	B
Cheerilee	<NA>	cheerileeponyville.edu	C	F
Discord	<NA>	discordponyville.edu	F	C
Fluttershy	<NA>	fluttershyponyville.edu	E	C
Rarity	<NA>	rarityponyville.edu	F	G
Scootaloo	<NA>	scootalooponyville.edu	C	E
Spike	<NA>	spikeponyville.edu	G	B
Zecora	<NA>	zecoraponyville.edu	D	G
Shining	Armor	shining.armorponyville.edu	A	H
Sweetie	Belle	sweetie.belleponyville.edu	H	H
...				

---

Hereafter, use the data frame `psgrps.x` instead of `psgrps`.

- d. [5 pts] Extract the row indices of the group mates of each student for the first two problem sets. Store the result as a list named `grp.mates`.
- e. [5 pts] Extract the row indices of those students who had been assigned to the same group twice with at least one other student. Store the result as a vector named `repeats`.
- f. [Extra credit, 10 pts] Re-assign the groups for the second problem set such that no student is assigned to the same group with a student whom he or she has previously worked with. Name the resulting data frame `psgrps.new`.