



Welcome to QuACK!

Week 3

9/22/2025

Nina, Sophie, Alyson, and Victoria

Today's Agenda

- Warm-up activity
- Preparing your data for analysis
 - Pivoting
 - Joining
- Practice

Week 3 Warm-up – 10 minutes

1. Download this week's materials from this link: <https://tinyurl.com/quack-fall2025>
2. Make a New R script. Save it to your computer. Title it "S3_warmup".
3. Load in *tidyverse* and *tidylog* packages.
4. Read in the `quack_response_data_fall25.csv` data.
5. Rename the column names to something easier to work with.
6. Change a variable of your choosing from a character to a factor.
7. (*New Skill*) Export the modified data into a .csv file using `write_csv()`
*If you aren't sure where to start with this, try typing `?write_csv()` in the console!
8. Run the code and check that everything looks correct in the global environment.
9. Open the starter script for this week.

Ways to Combine Datasets in R

Combine Data Sets

a		b	
x1	x2	x1	x3
A	1	A	T
B	2	B	F
C	3	D	T

+

=

Mutating Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NA

dplyr::left_join(a, b, by = "x1")

Join matching rows from b to a.

x1	x3	x2
A	T	1
B	F	2
D	T	NA

dplyr::right_join(a, b, by = "x1")

Join matching rows from a to b.

x1	x2	x3
A	1	T
B	2	F

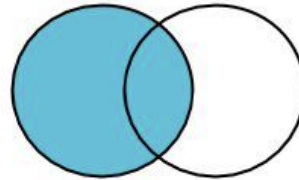
dplyr::inner_join(a, b, by = "x1")

Join data. Retain only rows in both sets.

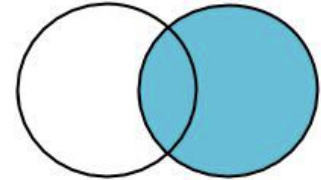
x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::full_join(a, b, by = "x1")

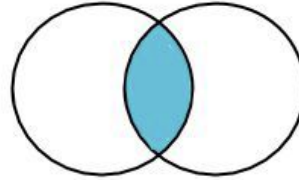
Join data. Retain all values, all rows.



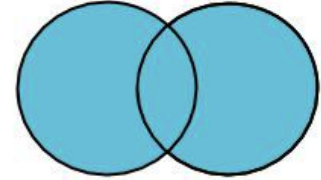
Left Join



Right Join



Inner Join



Full Outer Join

Functions in R to Combine Data

Combine Data Sets

a		b	
x1	x2	x1	x3
A	1	A	T
B	2	B	F
C	3	D	T

+

=

Mutating Joins

x1	x2	x3
A	1	T
B	2	F
C	3	NA

dplyr::left_join(a, b, by = "x1")

Join matching rows from b to a.

x1	x3	x2
A	T	1
B	F	2
D	T	NA

dplyr::right_join(a, b, by = "x1")

Join matching rows from a to b.

x1	x2	x3
A	1	T
B	2	F

dplyr::inner_join(a, b, by = "x1")

Join data. Retain only rows in both sets.

x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::full_join(a, b, by = "x1")

Join data. Retain all values, all rows.

Left Join

```
joined_df <- left_join(a,b)
```

```
joined_df <- a %>%  
  left_join(b)
```

Full Join

```
joined_df <- full_join(a,b)
```

```
joined_df <- a %>%  
  full_join(b)
```

Inner Join

```
joined_df <- inner_join(a,b)
```

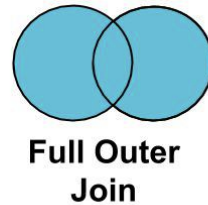
```
joined_df <- a %>%  
  inner_join(b)
```

An Example of When We Might Use These

- Let's pretend we have a new dataset from a survey collecting Mid-Semester feedback on Quack.
- We might want to compare these responses to our Beginning-of-Semester Quack survey responses.
- In order to do this, we will need to first combine both datasets together in R.
- Let's take a look at some use-cases for each type of joining using these surveys.

Example: Full Join

a		b		
x1	x2	x1	x3	
A	1	A	T	+ =
B	2	B	F	
C	3	D	T	
dplyr::full_join(a, b, by = "x1")				Join data. Retain all values, all rows.
x1	x2	x3		
A	1	T		
B	2	F		
C	3	NA		
D	NA	T		



- Maybe we want to look at all the data we have collected this semester and don't care if people filled out both or only one survey.

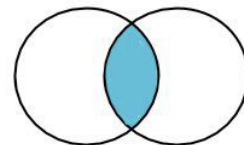
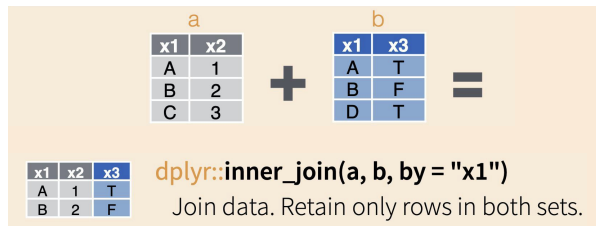
```
full_join(Beginning-of-Semester Data, Mid-Semester Data, by = "ID")
```



Ordering doesn't matter since all rows in both dataframes are retained

→ *Result: dataset with everyone from both datasets, regardless of whether they completed both or only one of the survey*

Example: Inner Join



Inner Join

- If we want to see whether ratings improved over the course of the semester so far, we probably want to look *only* at people who have completed both surveys.

```
inner_join(Beginning-of-Semester Data, Mid-Semester Data, by = "ID")
```



Ordering doesn't matter since only rows that exist in both datasets will be retained

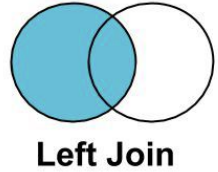
→ Result: dataset with ONLY people who completed BOTH surveys

Example: Left Join

a		b		
x1	x2	x1	x3	
A	1	A	T	+ =
B	2	B	F	
C	3	D	T	

x1	x2	x3
A	1	T
B	2	F
C	3	NA

`dplyr::left_join(a, b, by = "x1")`
Join matching rows from b to a.



Left Join

- Let's say we only want to look at people who filled out the Beginning-of-Semester survey.
- Maybe we don't care if they also completed the Mid-Semester survey, but if they did, we want to incorporate that data too.

```
left_join(Beginning-of-Semester Data, Mid-Semester Data, by = "ID")
```



All rows in this data will be retained



Rows that do not also exist in Beginning-of-Semester data will be removed

→ Result: dataset with only people who completed Beginning-of-Semester survey + adding their corresponding Mid-Semester if applicable

Pivoting / Reshaping Datasets in R

table4a

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

pivot_longer(data, cols, names_to = "name",
values_to = "value", values_drop_na = FALSE)

"Lengthen" data by collapsing several columns into two. Column names move to a new names_to column and values to a new values_to column.

```
pivot_longer(table4a, cols = 2:3, names_to = "year",  
              values_to = "cases")
```

table2

country	year	type	count
A	1999	cases	0.7K
A	1999	pop	19M
A	2000	cases	2K
A	2000	pop	20M
B	1999	cases	37K
B	1999	pop	172M
B	2000	cases	80K
B	2000	pop	174M
C	1999	cases	212K
C	1999	pop	1T
C	2000	cases	213K
C	2000	pop	1T



country	year	cases	pop
A	1999	0.7K	19M
A	2000	2K	20M
B	1999	37K	172M
B	2000	80K	174M
C	1999	212K	1T
C	2000	213K	1T

pivot_wider(data, names_from = "name",
values_from = "value")

The inverse of pivot_longer(). "Widen" data by expanding two columns into several. One column provides the new column names, the other the values.

```
pivot_wider(table2, names_from = type,  
             values_from = count)
```

Time to get started in R!

- Go to <https://tinyurl.com/quack-fall2025>
- Download and unzip the course materials
- You will see many files:
 - Two R Markdown (.rmd) files
 - A .csv file
 - Two PDF files – slides and a practice doc
 - Some other folders – ignore these for now
- Open the R Markdown file marked “starter_code” – this is where we will code together for the first part of the session
- The .csv file contains the data we will be working with
- The practice PDF contains practice exercises for you to do during the second part of the session



Exit Ticket

Link: <https://tinyurl.com/quack-fall2025-exit-survey>

