Data Science for Everyone – Data Wrangling – Joins 3

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

Plan for Today

- Look more at keys

Keys

- What uniquely links one row in students to one row in hometowns?

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

hometowns

LastName	FirstName	ID	Town
Paulson	Kaden	8999	Providence, RI
Saydun	Hanna	7009	Madison, WI
Mosca	Ab	1001	Manchester, NH
Garrison	Henri	9003	Jamaica Plain, MA
Saydun	Sarah	5006	Miami, FL
Cook	Kaden	9990	Philadelphia, PA

Keys

• What uniquely links one row in students to one row in hometowns?

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

hometowns

LastName	FirstName	ID	Town
Paulson	Kaden	8999	Providence, RI
Saydun	Hanna	7009	Madison, WI
Mosca	Ab	1001	Manchester, NH
Garrison	Henri	9003	Jamaica Plain, MA
Saydun	Sarah	5006	Miami, FL
Cook	Kaden	9990	Philadelphia, PA

Keys

```
full_table <- students %>%
              full_join(hometowns, by = "ID")
```

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

hometowns

LastName	FirstName	ID	Town
Paulson	Kaden	8999	Providence, RI
Saydun	Hanna	7009	Madison, WI
Mosca	Ab	1001	Manchester, NH
Garrison	Henri	9003	Jamaica Plain, MA
Saydun	Sarah	5006	Miami, FL
Cook	Kaden	9990	Philadelphia, PA

Keys

- What uniquely links one row in students to one row in student-classes?

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

student-classes

CRN	Time	Student
101	TTh 10am	1001
301	TTh 10am	5006
333	MWF 9am	8999
101	MWF 9am	9003
101	MWF 9am	1001
301	MW 3pm	8999

Keys

- What uniquely links one row in students to one row in student-classes?

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

student-classes

CRN	Time	Student
101	TTh 10am	1001
301	TTh 10am	5006
333	MWF 9am	8999
101	MWF 9am	9003
101	MWF 9am	1001
301	MW 3pm	8999

```
full_table <- students %>%
            full_join(student-classes, by = c("ID" = "Student")
```

Keys

students

ID	FirstName	LastName	Year
1001	Ab	Mosca	first year
5006	Sarah	Saydun	sophomore
9003	Henri	Garrison	sophomore
8999	Kaden	Paulson	senior

student-classes

CRN	Time	Student
101	TTh 10am	1001
301	TTh 10am	5006
333	MWF 9am	8999
101	MWF 9am	9003
101	MWF 9am	1001
301	MW 3pm	8999

Keys

- What uniquely links one row in student-classes to one row in classes?

student-classes

CRN	Time	Student
101	TTh 10am	1001
301	TTh 10am	5006
333	MWF 9am	8999
101	MWF 9am	9003
101	MWF 9am	1001
301	MW 3pm	8999

classes

CRN	Name	Time
101	Intro DS	MWF 9am
101	Intro DS	TTh 10am
301	Statistics	MW 3pm
333	Calculus	MWF 9am
301	Statistics	TTh 10am

```
full_table <- student-classes %>%
            full_join(classes, by = c("CRN", "Time")
```

Keys

student-classes

CRN	Time	Student
101	TTh 10am	1001
301	TTh 10am	5006
333	MWF 9am	8999
101	MWF 9am	9003
101	MWF 9am	1001
301	MW 3pm	8999

classes

CRN	Name	Time
101	Intro DS	MWF 9am
101	Intro DS	TTh 10am
301	Statistics	MW 3pm
333	Calculus	MWF 9am
301	Statistics	TTh 10am

Your turn

- Work with 1-2 other classmates on lab-join.Rmd