dplyr_select

October 12, 2020

Using select() Function in dplyr

1 1. Using `select()` to Extract Columns

- Recall that `filter()` can be used to filter rows
- Similarly, `select()` is used to select columns
- These functions can be "chained"

1.1 1.1 Example of `select()`

```
In [1]: library(dplyr)

        arrests <- read.csv("data/pvd_arrests_2020-10-03.csv")


Attaching package: dplyr


The following objects are masked from package:stats:

    filter, lag


The following objects are masked from package:base:

    intersect, setdiff, setequal, union



In [2]: arrests_subset <- arrests %>%
            select(arrestee_id, age, gender, statute_desc)

        head(arrests_subset)
```

A data.frame: 6 × 4

| | arrestee_id | age | gender | statute_desc |
	<chr>	<int>	<chr>	<chr>
1	pvd2218242150382148273	37	Male	
2	pvd15166785558364246202	25		Driving after Denial, Suspension or F
3	pvd3142917706201385905	34	Female	RESISTING LEGAL OR ILLEGAL AI
4	pvd3142917706201385905	34	Female	DISORDERLY CONDUCT
5	pvd460449304532374599	18	Female	RESISTING LEGAL OR ILLEGAL AI
6	pvd460449304532374599	18	Female	DISORDERLY CONDUCT

1.1.1 1.1.1 Comparing `select()` to `[, ]` notation

```
In [3]: # dplyr example
        arrests %>%
            select(arrestee_id, age, gender, statute_desc)


        # equivalent in "base" R example
        cols <- c("arrestee_id", "age", "gender", "statute_desc")

        arrests[, cols]
```

arrestee_id	age	gender	statute_desc
<chr>	<int>	<chr>	<chr>
pvd2218242150382148273	37	Male	
pvd15166785558364246202	25		Driving after Denial, Suspension or R
pvd3142917706201385905	34	Female	RESISTING LEGAL OR ILLEGAL AI
pvd3142917706201385905	34	Female	DISORDERLY CONDUCT
pvd460449304532374599	18	Female	RESISTING LEGAL OR ILLEGAL AI
pvd460449304532374599	18	Female	DISORDERLY CONDUCT
pvd460449304532374599	18	Female	OBSTRUCTING OFFICER IN EXECU
pvd6431558757894418021	28	Male	Chemical Test Refusal
pvd6431558757894418021	28	Male	Driving Under the Influence of Liqou
pvd6431558757894418021	28	Male	Driving after Denial, Suspension or R
pvd15614289459563584867	22	Male	DOMESTIC-VANDALISM/MALICIO
pvd900460037611487829	19	Male	Reckless Driving, Drag Racing - Atter
pvd900460037611487829	19	Male	Driving after Denial, Suspension or R
pvd1675234703933765967	23	Male	RESISTING LEGAL OR ILLEGAL AI
pvd1675234703933765967	23	Male	OBSTRUCTING OFFICER IN EXECU
pvd8555094992612905738	45	Female	VANDALISM/MALICIOUS INJURY
pvd8555094992612905738	45	Female	Disorderly and Indecent Conduct B -
pvd1795374794812880432	20	Male	DOMESTIC-VANDALISM/MALICIO
pvd15100522507516509426	47	Male	Driving after Denial, Suspension or R
pvd1784400650677464818	30	Male	MANUFACTURE/POSS/DELIVER S
pvd1784400650677464818	30	Male	MANUFAC/POSS/DELIVER SCH 1
pvd1790662812478938357	54	Male	Driving after Denial, Suspension or R
pvd2709931773722136485	39	Male	
pvd3024232238010666153	21	Male	SIMPLE ASSAULT OR BATTERY
pvd5436004350740110700	40	Male	CONTROLLED SUBSTANCE CONSI
pvd5436004350740110700	40	Male	MANUFAC/POSS/DELIVER SCH 1
pvd18360527540285642844	59	Male	CONTROLLED SUBSTANCE CONSI
pvd18360527540285642844	59	Male	MANUFAC/POSS/DELIVER SCH 1
pvd6329853531447301575	38	Male	
pvd4717540273846557288	31	Female	BENCH WARRANT ISSUED FROM S
pvd6847437265855082531	24	Female	Driving Under the Influence of Liqou
pvd6847437265855082531	24	Female	Driving after Denial, Suspension or R
pvd10486073748894772221	44	Male	NULL
pvd14228342987261766892	35	Male	ASSAULT OF POLICE OFFICERS AN
pvd5257435183320846711	51	Male	WARRANT OF ARREST ON AFFIDA
pvd5257435183320846711	51	Male	DOMESTIC-SIMPLE ASSAULT/BAT
pvd6014105176016998190	32	Male	DOMESTIC-DISORDERLY CONDUC
pvd16628535577915270026	25	Male	Chemical Test Refusal
pvd16628535577915270026	25	Male	Driving Under the Influence of Liqou
pvd7732905157072162140	28	Female	BENCH WARRANT ISSUED FROM (
pvd7732905157072162140	28	Female	POSSESSION OF SCHEDULE I II III
pvd7732905157072162140	28	Female	BENCH WARRANT ISSUED FROM S
pvd14783764350884014595	32	Male	NULL
pvd467817857219686623	55	Male	NULL
pvd9295729193531288284	55	Female	DOMESTIC-DISORDERLY CONDUC
pvd12212629709488901378	40	Male	VIOLATION OF PROTECTIVE ORDI
pvd4464707333206490510	24	Male	Possession of Stolen Vehicle or Parts
pvd4464707333206490510	24	Male	Driving after Denial, Revocation or St
pvd6155618112432892995	72	Male	DISORDERLY CONDUCT
pvd8046681516085598213	21	Male	DOMESTIC-FELONY ASSAULT

A data.frame: 8755 Œ 4

arrestee_id	age	gender	statute_desc
<chr>	<int>	<chr>	<chr>
pvd2218242150382148273	37	Male	
pvd15166785558364246202	25		Driving after Denial, Suspension or R
pvd3142917706201385905	34	Female	RESISTING LEGAL OR ILLEGAL AI
pvd3142917706201385905	34	Female	DISORDERLY CONDUCT
pvd460449304532374599	18	Female	RESISTING LEGAL OR ILLEGAL AI
pvd460449304532374599	18	Female	DISORDERLY CONDUCT
pvd460449304532374599	18	Female	OBSTRUCTING OFFICER IN EXECU
pvd6431558757894418021	28	Male	Chemical Test Refusal
pvd6431558757894418021	28	Male	Driving Under the Influence of Liqu
pvd6431558757894418021	28	Male	Driving after Denial, Suspension or R
pvd15614289459563584867	22	Male	DOMESTIC-VANDALISM/MALICIO
pvd900460037611487829	19	Male	Reckless Driving, Drag Racing - Atter
pvd900460037611487829	19	Male	Driving after Denial, Suspension or R
pvd1675234703933765967	23	Male	RESISTING LEGAL OR ILLEGAL AI
pvd1675234703933765967	23	Male	OBSTRUCTING OFFICER IN EXECU
pvd8555094992612905738	45	Female	VANDALISM/MALICIOUS INJURY
pvd8555094992612905738	45	Female	Disorderly and Indecent Conduct B -
pvd1795374794821280432	20	Male	DOMESTIC-VANDALISM/MALICIO
pvd15100522507516509426	47	Male	Driving after Denial, Suspension or R
pvd1784400650677464818	30	Male	MANUFACTURE/POSS/DELIVER S
pvd1784400650677464818	30	Male	MANUFAC/POSS/DELIVER SCH 1
pvd1790662812478938577	54	Male	Driving after Denial, Suspension or R
pvd2709931773722136485	39	Male	
pvd3024232238010666153	21	Male	SIMPLE ASSAULT OR BATTERY
pvd5436004350740110700	40	Male	CONTROLLED SUBSTANCE CONSI
pvd5436004350740110700	40	Male	MANUFAC/POSS/DELIVER SCH 1
pvd18360527540285642844	59	Male	CONTROLLED SUBSTANCE CONSI
pvd18360527540285642844	59	Male	MANUFAC/POSS/DELIVER SCH 1
pvd6329853531447301575	38	Male	
pvd4717540273846557288	31	Female	BENCH WARRANT ISSUED FROM S
pvd6847437265855082531	24	Female	Driving Under the Influence of Liqu
pvd6847437265855082531	24	Female	Driving after Denial, Suspension or R
pvd10486073748894772221	44	Male	NULL
pvd14228342987261766892	35	Male	ASSAULT OF POLICE OFFICERS AN
pvd5257435183320846711	51	Male	WARRANT OF ARREST ON AFFIDA
pvd5257435183320846711	51	Male	DOMESTIC-SIMPLE ASSAULT/BAT
pvd6014105176016998190	32	Male	DOMESTIC-DISORDERLY CONDUC
pvd16628535577915270026	25	Male	Chemical Test Refusal
pvd16628535577915270026	25	Male	Driving Under the Influence of Liqu
pvd7732905157072162140	28	Female	BENCH WARRANT ISSUED FROM C
pvd7732905157072162140	28	Female	POSSESSION OF SCHEDULE I II III
pvd7732905157072162140	28	Female	BENCH WARRANT ISSUED FROM S
pvd14783764350884014595	32	Male	NULL
pvd467817857219686623	55	Male	NULL
pvd9295729193531288284	55	Female	DOMESTIC-DISORDERLY CONDUC
pvd12212629709488901378	40	Male	VIOLATION OF PROTECTIVE ORDI
pvd4464707333206490510	24	Male	Possession of Stolen Vehicle or Parts
pvd4464707333206490510	24	Male	Driving after Denial, Revocation or S
pvd6155618112432892995	72	Male	DISORDERLY CONDUCT
pvd8046681516085598213	21	Male	DOMESTIC-FELONY ASSAULT

A data.frame: 8755 × 4

4

1.2　1.2 Example of `select()` (cont.)

```
In [4]: arrests_vio <- arrests %>%
            select(
                arrestee_id,
                age,
                gender,
                statute_desc
            )
```

```
In [5]: head(arrests_vio)          # see first few lines of new dataframe
```

A data.frame: 6 × 4

	arrestee_id	age	gender	statute_desc
	<chr>	<int>	<chr>	<chr>
1	pvd2218242150382148273	37	Male	
2	pvd15166785558364246202	25		Driving after Denial, Suspension or R
3	pvd3142917706201385905	34	Female	RESISTING LEGAL OR ILLEGAL AI
4	pvd3142917706201385905	34	Female	DISORDERLY CONDUCT
5	pvd460449304532374599	18	Female	RESISTING LEGAL OR ILLEGAL AI
6	pvd460449304532374599	18	Female	DISORDERLY CONDUCT

2　2. Chaining *dplyr* Operators

- One key reason for *dplyr* popularity
- *dplyr* verbs/functions are "composable"

  - $(f \circ g)(x) == f(g(x))$

```
In [6]: female_vio <- arrests %>%
            filter(gender == "Female") %>%
            select(arrestee_id, age, gender, statute_desc)

        head(female_vio)
```

A data.frame: 6 × 4

	arrestee_id	age	gender	statute_desc
	<chr>	<int>	<chr>	<chr>
1	pvd3142917706201385905	34	Female	RESISTING LEGAL OR ILLEGAL AR
2	pvd3142917706201385905	34	Female	DISORDERLY CONDUCT
3	pvd460449304532374599	18	Female	RESISTING LEGAL OR ILLEGAL AR
4	pvd460449304532374599	18	Female	DISORDERLY CONDUCT
5	pvd460449304532374599	18	Female	OBSTRUCTING OFFICER IN EXECU
6	pvd8555094992612905738	45	Female	VANDALISM/MALICIOUS INJURY ′

2.1　2.1 More Chaining

```
In [7]: female_midage <- arrests %>%
            filter(
                gender == "Female",
                age > 45,
```

```
        statute_desc != ""
    ) %>%
    select(
        arrestee_id,
        age,
        gender,
        statute_desc
    ) %>%
    arrange(
        arrestee_id
    )

head(female_midage)
```

	arrestee_id <chr>	age <int>	gender <chr>	statute_desc <chr>
1	pvd10124563078769507165	52	Female	BENCH WARRANT ISSUED FROM
2	pvd10764271506246443138	53	Female	DOMESTIC-SIMPLE ASSAULT/BAT
3	pvd10857153000728613606	54	Female	Driving after Denial, Suspension or F
4	pvd10879521477909738425	48	Female	DISORDERLY CONDUCT
5	pvd10909858543687069510	49	Female	NULL
6	pvd11014001617297730559	47	Female	Driving after Denial, Suspension or F

A data.frame: 6 × 4