TidyVerse Tutorial In R:

Author: Ashkan Farhangi

Source: https://cengel.github.io/R-data-wrangling/dplyr.html#add-new-columns

How to use piping in R:

Q: How can we sort, clean, and select part of dataset that is important for us in R?

A:

Use sorting, filter, and select modules from Tidyverse to ease your funcitons:

Code available in TidyVerseTut.txt File

```
#Author: Ashkan Farhangi
trafficstops <- read.csv("data/trafficstops.csv")
library(tidyverse)
select(trafficstops, police_department, officer_id, driver
select(trafficstops, starts_with("driver"))
filter(trafficstops, county_name == "Yazoo County")
slice(trafficstops, 1:3)
senior_drivers <- trafficstops %>%
  filter(driver_age > 85) %>%
  select(violation_raw,driver_gender,driver_race)

#Challenge 1 with bonus
challenge1 <-trafficstops %>%
  filter(county_name=="Tunica County") %>%
  select(stop_date,driver_age,violation_raw) %>%
  arrange(driver_age)
  view(challenge1)

```

#Challenge 1 with bonus

challenge1 <-trafficstops %>%

filter(county_name=="Tunica County") %>%

select(stop_date,driver_age,violation_raw) %>%

arrange(driver_age)

view(challenge1)

Results:

	stop_date	driver_age	violation_raw
1	2014-03-04	36	Expired or no non-commercial driver license or permit
2	2013-04-20	38	Failure to maintain required liability insurance
3	2013-04-20	38	??
4	2016-02-14	39	Speeding - Regulated or posted speed limit and actual speed

2 Data Mutation

Q: How can you create a new column based on an exsiting data?

we'll use `mutate()` for create a new copy from a column, keep the original and change the new one as we like.

Code:

```
a_violation <-trafficstops %>%
  filter(driver_age == 50 & driver_gender == "female") %>%
  mutate(wds = wday(ymd(stop_date))) %>%
  select(violation_raw, wds) %>%
  filter(wds == 1)
```

	violation_raw	wds
1	Speeding - Regulated or posted speed limit and actual speed	1
2	Speeding - Regulated or posted speed limit and actual speed	1
3	Speeding - Regulated or posted speed limit and actual speed	1
4	Seat belt not used properly as required	1
5	Speeding - Regulated or posted speed limit and actual speed	1
6	Speeding - Regulated or posted speed limit and actual speed	1
7	Speeding - Regulated or posted speed limit and actual speed	1
8	Speeding - Regulated or posted speed limit and actual speed	1
9	Speeding - Regulated or posted speed limit and actual speed	1
10	Seat belt not used properly as required	1
11	Speeding - Regulated or posted speed limit and actual speed	1
12	Speeding - Regulated or posted speed limit and actual speed	1
13	Speeding - Regulated or posted speed limit and actual speed	1
14	Failure to maintain required liability insurance	1
15	Speeding - Regulated or posted speed limit and actual speed	1
16	Seat belt not used properly as required	1
17	Speeding - Regulated or posted speed limit and actual speed	1
18	Speeding - Regulated or posted speed limit and actual speed	1
19	Speeding - Regulated or posted speed limit and actual speed	1

```r
library(lubridate)
e_mutate <- trafficstops %>%
  mutate(birth_year = substring(driver_birthdate, 1, 4)) %>%
  mutate(birth_date = ymd(driver_birthdate),birth_year = year(driver_birthdate)) %>%
  mutate(birth_date = ymd(driver_birthdate),
         birth_year = year(driver_birthdate),
         birth_cohort = round(birth_year/10)*10,
         birth_cohort = factor(birth_cohort)) %>%
  select(birth_cohort) %>%
  plot()



a_violation <-trafficstops %>%
  mutate(weekday_of_stop = wday(stop_day) ) %>%
  filter(driver_age==50,driver_gender=='female') %>%
  filter(wday(stop_date)==1)
```

only sunday

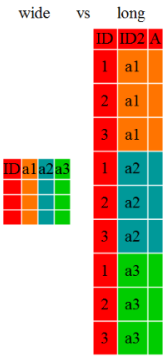
MS-2013-01084	MS	2013-01-06	Leake County	28079	Mississippi Highway Patrol	female	1962-08-03	White	Speeding - Regulated or posted speed limit and actual speed	H(
MS-2013-01160	MS	2013-01-06	Attala County	28007	Mississippi Highway Patrol	female	1963-04-25	White	Speeding - Regulated or posted speed limit and actual speed	D(
MS-2013-02196	MS	2013-01-13	Bolivar County	28011	Mississippi Highway Patrol	female	1962-11-01	Black	Speeding - Regulated or posted speed limit and actual speed	D(
MS-2013-03303	MS	2013-01-20	Sunflower County	28133	Mississippi Highway Patrol	female	1962-12-08	Black	Seat belt not used properly as required	D(
MS-2013-03317	MS	2013-01-20	Jackson County	28059	Mississippi Highway Patrol	female	1963-06-09	White	Speeding - Regulated or posted speed limit and actual speed	K(

Joining two or more tables

```
42
43  MS_bw_pop <- read.csv("data/MS_acs2015_bw.csv")
44  head(MS_bw_pop)
45
46  e_Joined<-trafficstops %>%
47    group_by(county_name) %>%
48    summarise(n_stops = n()) %>%
49    left_join(MS_bw_pop, by = c("county_name" = "County")) %>%
50    head()
51  a_sorted<-e_Joined %>%
52    arrange(n_stops)
53
54
```

	county_name	n_stops	FIPS	black_pop	white_pop	bw_pop
1	Benton County	214	28009	3078	5166	8244
2	Adams County	942	28001	17757	12856	30613
3	Amite County	2921	28005	5416	7395	12811
4	Alcorn County	3345	28003	4281	31563	35844
5	Attala County	4203	28007	8194	10649	18843
6	Bolivar County	4526	28011	21648	11197	32845

Wide vs Long Format



For problem 1 It's a **long** format.

Data Manipulation

```r
# Challenge 5
a6_wide_year <- trafficstops %>%
  mutate(year = year(ymd(stop_date))) %>%
  group_by(violation_raw,year) %>%
  summarize(number_of_stops = n()) %>%
  spread(year,number_of_stops)
  head(a6_wide_year)
```

	violation_raw	2013	2014	2015	2016
1	??	5018	4368	2935	682
2	Careless driving	607	561	709	270
3	Child or youth restraint not used properly as required	1313	873	709	286
4	Driving while license suspended	3258	2306	2455	516
5	Driving wrong way	40	33	38	14
6	Expired or no non-commercial driver license or permit	1433	984	1166	358
7	Failure to comply with financial responsibility law	1	1	*NA*	*NA*
8	Failure to maintain required liability insurance	7096	5193	7889	2874
9	Failure to obey sign or traffic control device	607	356	388	122
10	Failure to yield right of way (FTY ROW)	292	206	203	65
11	Following too closely	179	117	153	41
12	Improper passing	178	155	178	71
13	Improper turn	65	54	67	22
14	Operating without equipment as required by law	967	744	964	425
15	Other (non-mapped)	421	289	365	129
16	Reckless driving	233	229	284	125
17	Seat belt not used properly as required	7027	5768	4912	2547
18	Speeding	713	948	978	167
19	Speeding - Regulated or posted speed limit and actual speed	39818	31888	38721	15044