Class Activity 6

Your name here

January 14 2024

We will work with the `babynames` dataset again in this class activity. The header of the dataset looks like this:

```
knitr::kable(head(babynames))
```

year	sex	name	n	prop
1880	F	Mary	7065	0.0723836
1880	F	Anna	2604	0.0266790
1880	F	Emma	2003	0.0205215
1880	F	Elizabeth	1939	0.0198658
1880	F	Minnie	1746	0.0178884
1880	F	Margaret	1578	0.0161672

In this tutorial, we will learn about the five main verbs of dplyr and how to use them to manipulate data:

- `select()`: Choose columns from a data frame
- `filter()`: Choose rows based on a condition
- `arrange()`: Sort the rows of a data frame
- `mutate()`: Add new columns based on existing columns
- `summarise()`: Aggregate data and compute summary statistics

Problem 1: `select()`

Which of these is NOT a way to select the `name` and `n` columns together?

```
select(babynames, -c(year, sex, prop))  #1
select(babynames, name:n)   #2
select(babynames, starts_with("n")) #3
select(babynames, ends_with("n"))  #4
```

Answer:

Problem 2: `filter()`

Use `filter()` with the logical operators to extract:

a. All of the names where prop is greater than or equal to 0.08

b. All of the babies named "Margaret"

c. Use `filter()` to choose all rows where name is "John" and sex is "M".

Problem 3: `arrange()`

a. Use `arrange()` to sort the babynames dataset by the prop column in descending order.

b. Use `arrange()` to sort the babynames dataset by year (ascending) and then by prop (descending).

Problem 4: `mutate()`

a. Use `mutate()` to create a new column called decade which contains the decade the record is in (e.g., 1990 for the years 1990-1999).

Problem 5: `summarize()` or `summarise()`

Use the codes mentioned so far to compute three statistics:

- the total number of children who ever had your name
- the maximum number of children given your name in a single year
- the mean number of children given your name per year

Problem 6

a. Use `min_rank()` and `mutate()` to rank each row in babynames from largest prop to smallest prop.

b. Compute each name's rank within its year and sex.

c. Then compute the median rank for each combination of name and sex, and arrange the results from highest median rank to lowest.