Immunisation

Lets first read the data that I have sent you.

```
read.csv('immunization_data.csv') -> im_data
```

So lets look at the structure of the data

```
str(im_data)
```

```
'data.frame':
               1442485 obs. of
                               23 variables:
        : int
              1 2 3 4 5 6 7 8 9 10 ...
$ caseid: chr "
                  0100101345 02" " 0100101345 02" " 0100101383 02" " 0100101383 02" ...
$ v101.x: int 1 1 1 1 1 1 1 1 1 ...
$ v218 : int 2 2 3 3 3 3 3 3 2 2 ...
$ v121 : int 0 0 1 1 1 1 1 1 1 1 ...
$ v130.x: int 2 2 2 2 2 2 2 2 2 2 ...
$ v131.x: int 991 991 991 991 991 991 991 991 991 ...
$ v133
       : int 550006661010...
$ v150
       : int
              2 2 2 2 2 2 2 2 4 4 ...
$ v151
       : int 1 1 1 1 1 1 1 1 1 1 ...
$ v152
       : int 42 42 42 42 42 40 40 40 55 55 ...
$ v106
       : int 1 1 0 0 0 2 2 2 2 2 ...
$ h9a
        : int NA NA NA NA NA NA NA NA NA ...
$ h57
        : int NA NA NA NA NA NA NA NA NA ...
$ h58
        : int NA NA NA NA NA NA NA NA NA ...
$ mv501 : logi NA NA NA NA NA NA ...
$ mv130 : logi NA NA NA NA NA NA ...
$ mv131 : logi NA NA NA NA NA NA ...
$ mv102 : logi NA NA NA NA NA NA ...
$ mvO24 : logi NA NA NA NA NA NA ...
$ v101.y: int NA ...
$ v130.y: int NA ...
\ v131.y: int \ NA ...
```

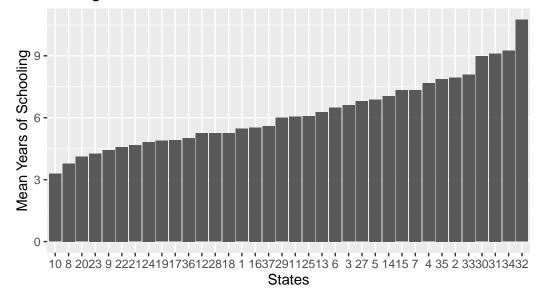
Load the required packages.

```
library(tidyverse)
library(rdhs)
library(haven)
```

Average education level across states

```
im_data %>%
  group_by(v101.x) %>%
  summarise(avg_edu = mean(v133)) %>%
  ggplot(aes(x = reorder(v101.x, avg_edu), y = avg_edu)) +
  geom_bar(stat = "identity") +
  labs(x = "States",
        y = "Mean Years of Schooling",
        title = "Average Education Across States in India",
        caption = "The numbers are the state codes in DHS 7 (2019-21)")
```

Average Education Across States in India



The numbers are the state codes in DHS 7 (2019–21)