

# Data Wrangling (Lab-7)

Bibek Sapkota

## Data Manipulation (Part-1)

task 1:

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.0      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
demo_data <- read_csv("yrbss_demo.csv")
```

```
## Rows: 40068 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (5): age, sex, grade, race4, race7
## dbl (3): record, bmi, stweight
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
demo_data
```

```
## # A tibble: 40,068 x 8
##   record age sex grade race4 race7 bmi stweight
##   <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9 69.0
## 2 635425 12 years old or younger Female 9th All other ~ Am I~ NA NA
## 3 506905 12 years old or younger Female 9th All other ~ Am I~ NA NA
## 4 1102227 12 years old or younger Female 9th All other ~ Asian NA NA
## 5 36586 12 years old or younger Female 9th Black or A~ Blac~ NA NA
## 6 767512 12 years old or younger Female 9th Black or A~ Blac~ NA NA
## 7 506907 12 years old or younger Female 9th Black or A~ Blac~ NA NA
## 8 930386 12 years old or younger Female 9th Black or A~ Blac~ NA NA
```

```
## 9 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 10 930393 12 years old or younger Female 9th Hispanic/L~ Hisp~ NA NA
## # i 40,058 more rows
```

## Filter Rows

### Filter by position

```
slice(demo_data, 1:6)
```

```
## # A tibble: 6 x 8
##   record age          sex   grade race4      race7   bmi stweight
##   <dbl> <chr>          <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other r~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other r~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other r~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other r~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or Af~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or Af~ Blac~ NA      NA
```

```
demo_data %>% slice(1:6)
```

```
## # A tibble: 6 x 8
##   record age          sex   grade race4      race7   bmi stweight
##   <dbl> <chr>          <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other r~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other r~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other r~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other r~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or Af~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or Af~ Blac~ NA      NA
```

### Filter by logical criteria

task 1: Filter rows with age less than 15

```
demo_data[demo_data$age<15,]
```

```
## # A tibble: 4,963 x 8
##   record age          sex   grade race4      race7   bmi stweight
##   <dbl> <chr>          <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other ~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 7 506907 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 8 930386 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
```

```
## 9 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 10 930393 12 years old or younger Female 9th Hispanic/L~ Hisp~ NA NA
## # i 4,953 more rows
```

```
demo_data %>% filter(age<15)
```

```
## # A tibble: 4,535 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>        <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other ~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 7 506907 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 8 930386 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 9 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 10 930393 12 years old or younger Female 9th Hispanic/L~ Hisp~ NA NA
## # i 4,525 more rows
```

task 2:Filter rows using equality

```
demo_data %>% filter(sex == "Female")
```

```
## # A tibble: 19,314 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>        <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other ~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 7 506907 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 8 930386 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 9 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 10 930393 12 years old or younger Female 9th Hispanic/L~ Hisp~ NA NA
## # i 19,304 more rows
```

```
demo_data %>% filter(!(grade == "9th"))
```

```
## # A tibble: 28,667 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>        <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506900 12 years old or younger Female 10th Black or A~ Blac~ NA      NA
## 2 1102232 12 years old or younger Female 10th Black or A~ Blac~ NA      NA
## 3 1102222 12 years old or younger Female 10th Hispanic/L~ Hisp~ 31.2    104.
## 4 1102231 12 years old or younger Female 10th Hispanic/L~ Hisp~ 18.7    50.8
## 5 328928 12 years old or younger Female 10th Hispanic/L~ Hisp~ NA      NA
## 6 635426 12 years old or younger Female 10th All other ~ Nati~ NA      NA
## 7 263703 12 years old or younger Female 10th White      White NA      NA
```

```
## 8 114951 12 years old or younger Female 10th All other ~ <NA> NA NA
## 9 263702 12 years old or younger Female 11th All other ~ Asian NA NA
## 10 1102233 12 years old or younger Female 11th Black or A~ Blac~ 19.4 58.1
## # i 28,657 more rows
```

task 3: Filter rows using OR operator

```
demo_data %>% filter((bmi<15) | (bmi>50))
```

```
## # A tibble: 114 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>      <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 930384 12 years old or younger Female 9th White White 14.9 NA
## 2 635432 13 years old Female 9th Hispanic/L~ Hisp~ 13.2 27.7
## 3 1094597 13 years old Male 9th All other ~ Asian 52.9 181.
## 4 499328 13 years old Male 9th Black or A~ Blac~ 14.6 54.4
## 5 768226 14 years old Female 9th All other ~ Am I~ 14.0 52.2
## 6 930799 14 years old Female 9th All other ~ Asian 14.7 NA
## 7 403318 14 years old Female 9th Black or A~ Blac~ 52.6 84.8
## 8 1310992 14 years old Female 9th Hispanic/L~ Hisp~ 14.8 42.6
## 9 329534 14 years old Female 9th Hispanic/L~ Hisp~ 54.3 134.
## 10 930422 14 years old Female 9th White White 14.7 NA
## # i 104 more rows
```

task 4: Filter rows using %in% operator.

```
demo_data %>% filter(grade %in% c("10th", "11th"))
```

```
## # A tibble: 19,554 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>      <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 506900 12 years old or younger Female 10th Black or A~ Blac~ NA NA
## 2 1102232 12 years old or younger Female 10th Black or A~ Blac~ NA NA
## 3 1102222 12 years old or younger Female 10th Hispanic/L~ Hisp~ 31.2 104.
## 4 1102231 12 years old or younger Female 10th Hispanic/L~ Hisp~ 18.7 50.8
## 5 328928 12 years old or younger Female 10th Hispanic/L~ Hisp~ NA NA
## 6 635426 12 years old or younger Female 10th All other ~ Nati~ NA NA
## 7 263703 12 years old or younger Female 10th White White NA NA
## 8 114951 12 years old or younger Female 10th All other ~ <NA> NA NA
## 9 263702 12 years old or younger Female 11th All other ~ Asian NA NA
## 10 1102233 12 years old or younger Female 11th Black or A~ Blac~ 19.4 58.1
## # i 19,544 more rows
```

task 5: Filter rows by multiple-column based criteria.

```
demo_data %>% filter(bmi<20, stweight<50, sex == "Male")
```

```
## # A tibble: 492 x 8
##   record age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>      <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 1094580 12 years old or younger Male 9th <NA> <NA> 15.4 44.4
```

```
## 2 1094578 12 years old or younger Male 10th <NA> <NA> 16.1 45.4
## 3 321498 13 years old Male 9th All other r~ Asian 16.3 41.7
## 4 321502 13 years old Male 9th Black or Af~ Blac~ 17.7 49.9
## 5 499323 13 years old Male 9th Hispanic/La~ Hisp~ 18.9 45.4
## 6 1094604 13 years old Male <NA> Hispanic/La~ Hisp~ 18.9 45.4
## 7 1303568 13 years old Male <NA> White White 16.4 43.6
## 8 1303569 13 years old Male <NA> White White 15.3 35.4
## 9 1094829 14 years old Male 9th All other r~ Am I~ 18.3 49.9
## 10 922978 14 years old Male 9th All other r~ Am I~ 15.7 45.4
## # i 482 more rows
```

```
demo_data %>% filter((bmi / stweight) < 0.5)
```

**Q1:**Write the code to filter rows with percentage of bmi/stweight < 0.5.

```
## # A tibble: 26,934 x 8
##   record age          sex grade race4      race7  bmi stweight
##   <dbl> <chr>         <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9 69.0
## 2 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 3 767508 12 years old or younger Female 9th Hispanic/L~ Hisp~ 29.9 79.4
## 4 930385 12 years old or younger Female 9th Hispanic/L~ Hisp~ 35.9 95.3
## 5 767513 12 years old or younger Female 9th Hispanic/L~ Hisp~ 31.4 90.7
## 6 1102222 12 years old or younger Female 10th Hispanic/L~ Hisp~ 31.2 104.
## 7 1102231 12 years old or younger Female 10th Hispanic/L~ Hisp~ 18.7 50.8
## 8 1102233 12 years old or younger Female 11th Black or A~ Blac~ 19.4 58.1
## 9 1102223 12 years old or younger Female 11th Hispanic/L~ Hisp~ 22.8 65.8
## 10 767510 12 years old or younger Female 12th White White 20.0 63.5
## # i 26,924 more rows
```

**Filter within a selection of variables.**

task 1:Creating a temp\_data by only using columns bmi and stweight

```
temp_data <- demo_data %>% select(bmi, stweight)
```

task 2:Filtering rows where all variables greater than 2.4

```
temp_data %>% filter_all(all_vars(> 2.4))
```

```
## # A tibble: 27,020 x 2
##   bmi stweight
##   <dbl> <dbl>
## 1 23.9 69.0
## 2 28 90.7
## 3 29.9 79.4
## 4 35.9 95.3
## 5 31.4 90.7
## 6 31.2 104.
```

```
## 7 18.7 50.8
## 8 19.4 58.1
## 9 22.8 65.8
## 10 32.0 45.4
## # i 27,010 more rows
```

task 3: Filtering rows where any variables are greater than 2.4

```
temp_data %>% filter_all(any_vars(> 2.4))
```

```
## # A tibble: 27,226 x 2
##   bmi stweight
##   <dbl> <dbl>
## 1 23.9 69.0
## 2 28 90.7
## 3 29.9 79.4
## 4 35.9 95.3
## 5 31.4 90.7
## 6 14.9 NA
## 7 31.2 104.
## 8 18.7 50.8
## 9 19.4 58.1
## 10 22.8 65.8
## # i 27,216 more rows
```

task 4: Filters rows in demo\_data where both race values are “White”.

```
demo_data %>% filter_at(vars(starts_with("race")), all_vars(. == "White"))
```

```
## # A tibble: 11,634 x 8
##   record age sex grade race4 race7 bmi stweight
##   <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 36587 12 years old or younger Female 9th White White NA NA
## 2 185840 12 years old or younger Female 9th White White NA NA
## 3 930384 12 years old or younger Female 9th White White 14.9 NA
## 4 263703 12 years old or younger Female 10th White White NA NA
## 5 767510 12 years old or younger Female 12th White White 20.0 63.5
## 6 36588 12 years old or younger Female 12th White White NA NA
## 7 185842 12 years old or younger Female 12th White White NA NA
## 8 1310482 12 years old or younger Female <NA> White White NA NA
## 9 185841 12 years old or younger Female <NA> White White NA NA
## 10 263701 12 years old or younger Female <NA> White White NA NA
## # i 11,624 more rows
```

task 5: Filter rows in demo\_data where any of race values equals “Asian”

```
demo_data %>% filter_at(vars(starts_with("race")), any_vars(. == "Asian"))
```

```
## # A tibble: 3,203 x 8
##   record age sex grade race4 race7 bmi stweight
##   <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
```

```
## 1 1102227 12 years old or younger Female 9th All other ~ Asian NA NA
## 2 263702 12 years old or younger Female 11th All other ~ Asian NA NA
## 3 328930 12 years old or younger Female 12th All other ~ Asian NA NA
## 4 1102226 12 years old or younger Female 12th All other ~ Asian NA NA
## 5 263700 12 years old or younger Female <NA> All other ~ Asian NA NA
## 6 328929 12 years old or younger Female <NA> All other ~ Asian NA NA
## 7 922322 12 years old or younger Male 9th All other ~ Asian NA NA
## 8 1094591 12 years old or younger Male 9th All other ~ Asian 30.7 75.8
## 9 760524 12 years old or younger Male 9th All other ~ Asian 18.0 52.2
## 10 30605 12 years old or younger Male 9th All other ~ Asian NA NA
## # i 3,193 more rows
```

task 6:Filtering rows Which have none values for all numeric variables.

```
demo_data %>% filter_if(is.numeric, all_vars(!is.na(.)))
```

```
## # A tibble: 27,020 x 8
##   record age          sex grade race4 race7 bmi stweight
##   <dbl> <chr>        <chr> <chr> <chr> <chr> <dbl> <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9 69.0
## 2 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28 90.7
## 3 767508 12 years old or younger Female 9th Hispanic/L~ Hisp~ 29.9 79.4
## 4 930385 12 years old or younger Female 9th Hispanic/L~ Hisp~ 35.9 95.3
## 5 767513 12 years old or younger Female 9th Hispanic/L~ Hisp~ 31.4 90.7
## 6 1102222 12 years old or younger Female 10th Hispanic/L~ Hisp~ 31.2 104.
## 7 1102231 12 years old or younger Female 10th Hispanic/L~ Hisp~ 18.7 50.8
## 8 1102233 12 years old or younger Female 11th Black or A~ Blac~ 19.4 58.1
## 9 1102223 12 years old or younger Female 11th Hispanic/L~ Hisp~ 22.8 65.8
## 10 402813 12 years old or younger Female 12th Hispanic/L~ Hisp~ 32.0 45.4
## # i 27,010 more rows
```

Select random rows from a dataframe

```
set.seed(1234)
```

```
demo_data %>% sample_n(5, replace = FALSE)
```

```
## # A tibble: 5 x 8
##   record age          sex grade race4 race7 bmi stweight
##   <dbl> <chr>        <chr> <chr> <chr> <dbl> <dbl>
## 1 1314019 16 years old      Female 10th White White 22.7 54.4
## 2 41872 18 years old or older Female 11th White White NA NA
## 3 41830 18 years old or older Female 12th White White NA NA
## 4 509729 16 years old      Female 11th White White NA NA
## 5 1105019 16 years old      Female 10th White White 19.3 55.8
```

```
demo_data %>% sample_frac(0.05, replace = FALSE)
```

```
## # A tibble: 2,003 x 8
##   record age          sex grade race4 race7 bmi stweight
```

```
##      <dbl> <chr>                <chr> <chr> <chr>          <chr> <dbl>    <dbl>
## 1 1097754 16 years old          Male  10th  Hispanic/Lat~ Hisp~ 25.6    76.7
## 2 396603 14 years old          Male  9th   All other ra~ Asian 16.5    47.6
## 3 255932 14 years old          Male  9th   Black or Afr~ Blac~ NA      NA
## 4 188973 16 years old          Female 11th  White        White NA      NA
## 5 122768 18 years old or older Female 11th  All other ra~ Am I~ NA      NA
## 6 1312792 16 years old        Female 11th  White        White 21.1    50.8
## 7 184637 17 years old          Male  12th  <NA>         <NA> NA      NA
## 8 1308583 17 years old          Male  12th  Hispanic/Lat~ Hisp~ 33.6    109.
## 9 925092 16 years old          Male  10th  White        White 20.5    61.2
## 10 106916 17 years old         <NA>  12th  Black or Afr~ Blac~ NA      NA
## # i 1,993 more rows
```

## Select top n rows ordered by a variable

task 1:Select the top 5 rows ordered by bmi

```
demo_data %>% top_n(5, bmi)
```

```
## # A tibble: 5 x 8
##   record age      sex  grade race4          race7  bmi stweight
##   <dbl> <chr>    <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 329534 14 years old Female 9th   Hispanic/Latino Hisp~ 54.3    134.
## 2 933347 16 years old Female 11th  Hispanic/Latino Hisp~ 54.0    171.
## 3 324452 16 years old Male  11th  Black or African Ameri~ Blac~ 53.9    91.2
## 4 925672 16 years old Male  11th  White        White 54.1    166.
## 5 1314456 17 years old Female 11th  Hispanic/Latino Hisp~ 55.0    NA
```

task 2:Select the top5 rows ordered by bmi in all grades

```
demo_data %>% group_by(grade) %>% top_n(5, bmi)
```

```
## # A tibble: 25 x 8
## # Groups:   grade [5]
##   record age      sex  grade race4          race7  bmi stweight
##   <dbl> <chr>    <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 1094597 13 years old Male  9th   All other races Asian 52.9    181.
## 2 403318 14 years old Female 9th   Black or African Ameri~ Blac~ 52.6    84.8
## 3 329534 14 years old Female 9th   Hispanic/Latino Hisp~ 54.3    134.
## 4 1312697 15 years old Female 10th  Black or African Ameri~ Blac~ 52.3    95.3
## 5 1305503 15 years old Male  9th   All other races Am I~ 52.4    134.
## 6 1304358 15 years old Male  10th  Black or African Ameri~ Blac~ 51.7    97.1
## 7 323167 15 years old Male  <NA>  Hispanic/Latino Hisp~ 42.4    122.
## 8 933347 16 years old Female 11th  Hispanic/Latino Hisp~ 54.0    171.
## 9 925575 16 years old Male  9th   Black or African Ameri~ Blac~ 52.1    133.
## 10 1306969 16 years old Male  10th  Black or African Ameri~ Blac~ 51.5    163.
## # i 15 more rows
```



## Column operations

### Select columns

```
demo_data[, c("record", "age", "sex")]
```

```
## # A tibble: 40,068 x 3
##   record age                sex
##   <dbl> <chr>                <chr>
## 1 506901 12 years old or younger Female
## 2 635425 12 years old or younger Female
## 3 506905 12 years old or younger Female
## 4 1102227 12 years old or younger Female
## 5   36586 12 years old or younger Female
## 6  767512 12 years old or younger Female
## 7 506907 12 years old or younger Female
## 8 930386 12 years old or younger Female
## 9 1310476 12 years old or younger Female
## 10 930393 12 years old or younger Female
## # i 40,058 more rows
```

```
demo_data %>% select(record, age, sex)
```

```
## # A tibble: 40,068 x 3
##   record age                sex
##   <dbl> <chr>                <chr>
## 1 506901 12 years old or younger Female
## 2 635425 12 years old or younger Female
## 3 506905 12 years old or younger Female
## 4 1102227 12 years old or younger Female
## 5   36586 12 years old or younger Female
## 6  767512 12 years old or younger Female
## 7 506907 12 years old or younger Female
## 8 930386 12 years old or younger Female
## 9 1310476 12 years old or younger Female
## 10 930393 12 years old or younger Female
## # i 40,058 more rows
```

```
demo_data %>% select(record:sex)
```

```
## # A tibble: 40,068 x 3
##   record age                sex
##   <dbl> <chr>                <chr>
## 1 506901 12 years old or younger Female
## 2 635425 12 years old or younger Female
## 3 506905 12 years old or younger Female
## 4 1102227 12 years old or younger Female
## 5   36586 12 years old or younger Female
## 6  767512 12 years old or younger Female
## 7 506907 12 years old or younger Female
## 8 930386 12 years old or younger Female
```

```
## 9 1310476 12 years old or younger Female
## 10 930393 12 years old or younger Female
## # i 40,058 more rows
```

Q2: Explain the meaning of the output of: `demo_data %>% select(-contains("r"))?`

Rename columns

```
demo_data %>% rename(id = record)
```

```
## # A tibble: 40,068 x 8
##       id age          sex grade race4      race7    bmi stweight
##   <dbl> <chr>      <chr> <chr> <chr>    <chr> <dbl>    <dbl>
## 1 506901 12 years old or younger Female 9th All other ~ Am I~ 23.9    69.0
## 2 635425 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 3 506905 12 years old or younger Female 9th All other ~ Am I~ NA      NA
## 4 1102227 12 years old or younger Female 9th All other ~ Asian NA      NA
## 5 36586 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 6 767512 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 7 506907 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 8 930386 12 years old or younger Female 9th Black or A~ Blac~ NA      NA
## 9 1310476 12 years old or younger Female 9th Hispanic/L~ Hisp~ 28     90.7
## 10 930393 12 years old or younger Female 9th Hispanic/L~ Hisp~ NA      NA
## # i 40,058 more rows
```

Add a new column

```
demo_data %>% mutate(height_m = sqrt(stweight / bmi))
```

```
## # A tibble: 40,068 x 9
##       record age          sex grade race4 race7    bmi stweight height_m
##   <dbl> <chr>      <chr> <chr> <chr>    <dbl>    <dbl>    <dbl>
## 1 506901 12 years old or youn~ Fema~ 9th All ~ Am I~ 23.9    69.0    1.70
## 2 635425 12 years old or youn~ Fema~ 9th All ~ Am I~ NA      NA      NA
## 3 506905 12 years old or youn~ Fema~ 9th All ~ Am I~ NA      NA      NA
## 4 1102227 12 years old or youn~ Fema~ 9th All ~ Asian NA      NA      NA
## 5 36586 12 years old or youn~ Fema~ 9th Blac~ Blac~ NA      NA      NA
## 6 767512 12 years old or youn~ Fema~ 9th Blac~ Blac~ NA      NA      NA
## 7 506907 12 years old or youn~ Fema~ 9th Blac~ Blac~ NA      NA      NA
## 8 930386 12 years old or youn~ Fema~ 9th Blac~ Blac~ NA      NA      NA
## 9 1310476 12 years old or youn~ Fema~ 9th Hisp~ Hisp~ 28     90.7    1.8
## 10 930393 12 years old or youn~ Fema~ 9th Hisp~ Hisp~ NA      NA      NA
## # i 40,058 more rows
```