title: “Data Summarization” output: ioslides_presentation: css: ../../docs/styles.css widescreen: yes —

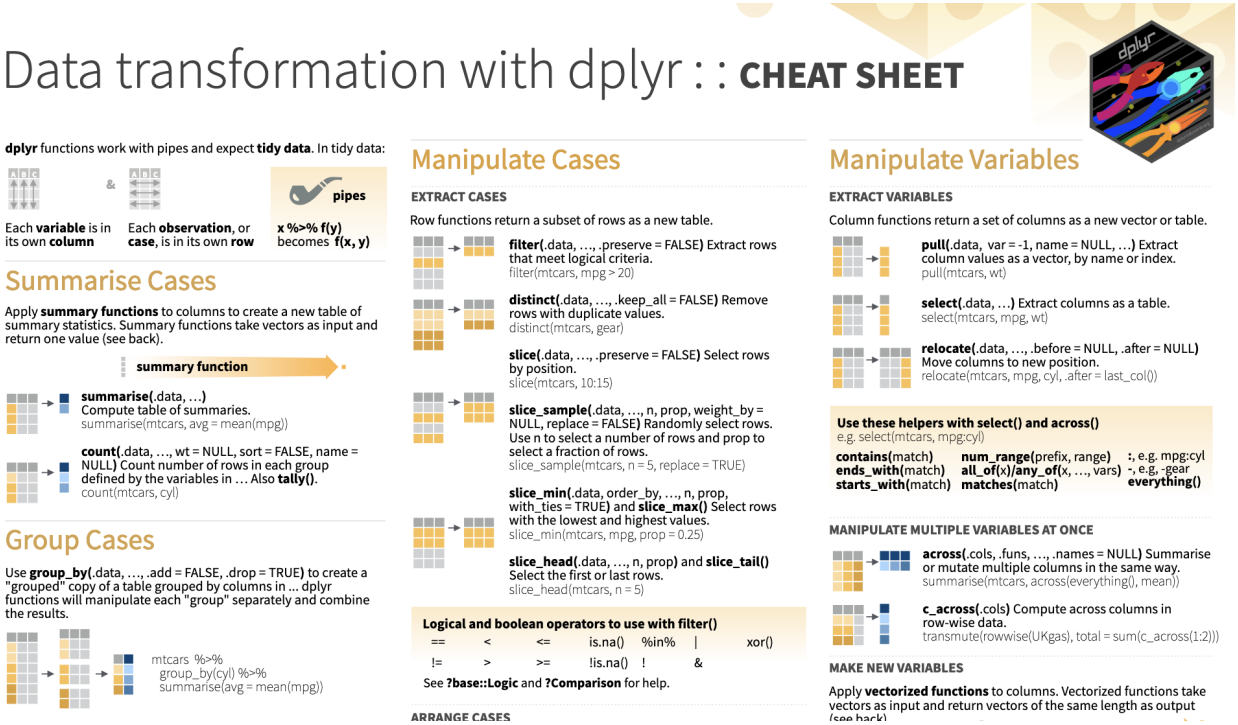
Recap

- `select()` : subset and/or reorder columns
- `filter()` : remove rows
- `arrange()` : reorder rows
- `mutate()` : create new columns or modify them
- `select()` and `filter()` can be combined together
- remove a column: `select()` with `!` mark (`!col_name`)
- you can do sequential steps: especially using pipes `%>%`

 Cheatsheet (https://daseh.org/modules/cheatsheets/Day-3.pdf)

Another Cheatsheet

https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-transformation.pdf (https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-transformation.pdf)



Data transformation with dplyr : : **CHEAT SHEET**

dplyr functions work with pipes and expect **tidy data**. In tidy data:

Each **variable** is in its own **column** & Each **observation**, or **case**, is in its own **row**

pipes x %>% f(y) becomes f(x, y)

Summarise Cases

Apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).

summary function

summarise(.data, …) Compute table of summaries.
summarise(mtcars, avg = mean(mpg))

count(.data, …, wt = NULL, sort = FALSE, name = NULL) Count number of rows in each group defined by the variables in … Also **tally()**.
count(mtcars, cyl)

Group Cases

Use **group_by**(.data, …, .add = FALSE, .drop = TRUE) to create a "grouped" copy of a table grouped by columns in ... dplyr functions will manipulate each "group" separately and combine the results.

```
mtcars %>%
  group_by(cyl) %>%
  summarise(avg = mean(mpg))
```

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.

filter(.data, …, .preserve = FALSE) Extract rows that meet logical criteria.
filter(mtcars, mpg > 20)

distinct(.data, …, .keep_all = FALSE) Remove rows with duplicate values.
distinct(mtcars, gear)

slice(.data, …, .preserve = FALSE) Select rows by position.
slice(mtcars, 10:15)

slice_sample(.data, …, n, prop, weight_by = NULL, replace = FALSE) Randomly select rows. Use n to select a number of rows and prop to select a fraction of rows.
slice_sample(mtcars, n = 5, replace = TRUE)

slice_min(.data, order_by, …, n, prop, with_ties = TRUE) and **slice_max()** Select rows with the lowest and highest values.
slice_min(mtcars, mpg, prop = 0.25)

slice_head(.data, …, n, prop) and **slice_tail()** Select the first or last rows.
slice_head(mtcars, n = 5)

Logical and boolean operators to use with filter()

==	<	<=	is.na()	%in%	|	xor()
!=	>	>=	!is.na()	!	&	

See **?base::Logic** and **?Comparison** for help.

ARRANGE CASES

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.

pull(.data, var = -1, name = NULL, …) Extract column values as a vector, by name or index.
pull(mtcars, wt)

select(.data, …) Extract columns as a table.
select(mtcars, mpg, wt)

relocate(.data, …, .before = NULL, .after = NULL) Move columns to new position.
relocate(mtcars, mpg, cyl, .after = last_col())

Use these helpers with select() and across()
e.g. select(mtcars, mpg:cyl)

contains(match) **num_range**(prefix, range) :, e.g. mpg:cyl
ends_with(match) **all_of**(x)/**any_of**(x, …, vars) -, e.g. -gear
starts_with(match) **matches**(match) **everything()**

MANIPULATE MULTIPLE VARIABLES AT ONCE

across(.cols, .funs, …, .names = NULL) Summarise or mutate multiple columns in the same way.
summarise(mtcars, across(everything(), mean))

c_across(.cols) Compute across columns in row-wise data.
transmute(rowwise(UKgas), total = sum(c_across(1:2)))

MAKE NEW VARIABLES

Apply **vectorized functions** to columns. Vectorized functions take vectors as input and return vectors of the same length as output (see back).

Data Summarization

- Basic statistical summarization
  - `mean(x)` : takes the mean of x
  - `sd(x)` : takes the standard deviation of x
  - `median(x)` : takes the median of x
  - `quantile(x)` : displays sample quantiles of x. Default is min, IQR, max
  - `range(x)` : displays the range. Same as `c(min(x), max(x))`
  - `sum(x)` : sum of x
  - `max(x)` : maximum value in x
  - `min(x)` : minimum value in x
- **all have the `na.rm =` argument for missing data**

Statistical summarization

The vector getting summarized goes inside the parentheses:

```
x <- c(1, 5, 7, 4, 2, 8)
mean(x)
```

```
[1] 4.5
```

```
range(x)
```

```
[1] 1 8
```

```
sum(x)
```

```
[1] 27
```

Statistical summarization

Note that many of these functions have additional inputs regarding missing data, typically requiring the `na.rm` argument ("remove NAs").

```
x <- c(1, 5, 7, 4, 2, 8, NA)
mean(x)
```

```
[1] NA
```

```
mean(x, na.rm = TRUE)
```

```
[1] 4.5
```

```
quantile(x)
```

```
Error in quantile.default(x): missing values and NaN's not allo
wed if 'na.rm' is FALSE
```

```
quantile(x, na.rm = TRUE)
```

```
  0%  25%  50%  75% 100% 
 1.0  2.5  4.5  6.5  8.0 
```

Statistical summarization

We will talk more about data types later, but you can only do summarization on numeric or logical types, NOT characters.

```
x <- c(1, 5, 7, 4, 2, 8)
sum(x)
```

```
[1] 27
```

```
y <- c(TRUE, FALSE, FALSE, TRUE) # FALSE == 0 and TRUE == 1
sum(y)
```

```
[1] 2
```

```
z <- c("TRUE", "FALSE", "FALSE", "TRUE")
sum(z)
```

```
Error in sum(z): invalid 'type' (character) of argument
```

Some examples

We can use the `CO_heat_ER` object from the `dasehr` package to explore different ways of summarizing data. (This dataset contains information about the number and rate of visits for heat-related illness to ERs in Colorado from 2011-2022, adjusted for age.) The `head` command displays the first rows of an object:

```
library(dasehr)
head(CO_heat_ER)
```

```
# A tibble: 6 × 7
  county     rate lower95cl upper95cl visits  year gender
  <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
1 Statewide  5.64      4.70      6.59    140  2011 Female
2 Statewide  7.39      6.30      8.47    183  2011 Male
3 Statewide  6.51      5.80      7.23    323  2011 Both genders
4 Statewide  5.64      4.72      6.57    146  2012 Female
5 Statewide  7.56      6.48      8.65    193  2012 Male
6 Statewide  6.58      5.88      7.29    339  2012 Both genders
```

The `dplyr` pipe `%>%` operator

A nice and readable way to chain together multiple R functions.

Changes `f(x, y)` to `x %>% f(y)`.

```
# Going to work
get_dressed(me,
            pack_lunch(
              check_pockets(
                wallet = TRUE, phone = TRUE, keys = TRUE),
              items = c("sandwich", "chips", "apple"), lunchbox
= TRUE),
            pants = TRUE, shirt = TRUE, footwear = "sandals")

# Going to work, the tidy way
me %>%
  get_dressed(pants = TRUE, shirt = TRUE, footwear = "sandals")
%>%
  pack_lunch(items = c("sandwich", "chips", "apple"), lunchbox
 = TRUE) %>%
  check_pockets(wallet = TRUE, phone = TRUE, keys = TRUE)
```

Statistical summarization the “tidy” way

```
CO_heat_ER %>% pull(visits) %>% mean(na.rm=T) # alt: pull(CO_he
at_ER, visits) %>% mean(na.rm=T)
```

```
[1] 9.791114
```

```
CO_heat_ER %>% pull(rate) %>% median(na.rm=T)
```

```
[1] 0
```

```
CO_heat_ER %>% pull(visits) %>% quantile(na.rm=T)
```

```
  0%  25%  50%  75% 100% 
   0    0    0    0  494 
```

```
CO_heat_ER %>% pull(rate) %>% quantile(probs = 0.9, na.rm=T)
```

```
     90% 
6.704074 
```

Behavior of `pull()` function

`pull()` converts a single data column into a vector. This allows you to run summary functions on these data. Once you have “pulled” the data column out, you don’t have to name it again in any piped summary functions.

```
er_visits <- CO_heat_ER %>% pull(visits)
class(er_visits)
```

```
[1] "numeric"
```

```
er_visits
```

```
  [1] 140 183 323 146 193 339 124 178 302  92 145 237 140 215
355 172 295 467
 [19] 113 210 323 163 293 456 154 235 389 105 197 302 173 252
425 185 309 494
 [37]  17  12  29  NA  13  23  14  17  31  NA  NA  15  NA  NA
16  14  28  42
 [55]  11  21  32  16  21  37  17  19  36  NA  15  24  19  16
35  18  27  45
 [73]   0   0   0   0   0   0  NA   0  NA   0   0   0   0  NA
NA   0  NA  NA
 [91]   0  NA  NA   0   0   0  NA   0  NA   0  NA  NA   0   0
0   0   0   0
[109]  16  17  33  12  15  27  11  NA  20  NA  NA  NA  16  15
31  18  21  39
[127]  NA  NA  16  13  21  34  17  14  31  NA  11  16  12  18
30  15  24  39
[145]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[163]  NA  NA  NA   0  NA  NA  NA   0  NA   0   0   0  NA  NA
NA  NA  NA  NA
[181]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[199]   0   0   0  NA   0  NA  NA   0  NA   0   0   0   0   0
0   0  NA  NA
[217]   0   0   0   0  NA  NA   0   0   0   0  NA  NA   0  NA
NA   0  NA  NA
[235]   0  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA   0  NA
NA   0   0   0
[253]  NA  NA  12  NA  NA  13  NA  NA  12  NA  NA  19  NA  NA
14  NA  NA  18
[271]  NA  NA  12  NA  11  18  NA  NA  13  NA  NA  12  NA  11
19  NA  15  19
[289]  NA  NA  NA  NA  NA  NA  NA   0  NA  NA   0  NA   0  NA
NA  NA   0  NA
[307]  NA   0  NA   0  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
NA  NA  NA  NA
[325]   0   0   0   0  NA  NA  NA  NA  NA  NA   0  NA   0  NA
```

```
NA   0   0   0
 [343]  NA   0  NA  NA  NA  NA  NA   0  NA   0   0   0   0  NA
NA   0   0   0
 [361]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [379]   0   0   0  NA  NA  NA   0   0   0  NA   0  NA   0   0
0   0  NA  NA
 [397]   0   0   0   0   0   0   0  NA  NA   0   0   0  NA   0
NA   0  NA  NA
 [415]   0  NA  NA   0  NA  NA  NA   0  NA   0   0   0   0   0
0   0   0   0
 [433]   0   0   0   0   0   0  NA   0  NA   0  NA  NA   0   0
0   0  NA  NA
 [451]   0   0   0  NA  NA  NA   0  NA  NA   0  NA  NA   0   0
0   0   0   0
 [469]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [487]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [505]   0  NA  NA  NA   0  NA  NA   0  NA   0   0   0   0   0
0  NA   0  NA
 [523]   0  NA  NA   0   0   0   0  NA  NA  NA  NA  NA   0  NA
NA   0   0   0
 [541]   0   0   0  NA   0  NA   0   0   0   0   0   0   0   0
0   0   0   0
 [559]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [577]   0   0   0   0   0   0   0   0   0   0   0   0   0  NA
NA  NA  NA  NA
 [595]  NA  NA  NA  NA  NA  NA  NA   0  NA  NA  NA  NA  NA  NA
NA  NA   0  NA
 [613]  16  26  42  18  22  40  NA  NA  19  NA  19  23  NA  16
25  13  30  43
 [631]  13  11  24  13  30  43  12  22  34  12  22  34  NA  20
28  18  29  47
 [649]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
```

```
 [667]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [685]  NA  NA  NA  NA  NA  13  NA  NA  17  NA  NA  NA  NA  NA
NA  NA  NA  NA
 [703]  NA  NA  13  NA  NA  NA  NA  NA  18  NA  NA  NA  NA  NA
NA  NA  NA  NA
 [721]   0   0   0   0   0   0   0   0   0  NA   0  NA   0   0
0   0   0   0
 [739]   0  NA  NA   0  NA  NA  NA  NA  NA   0   0   0   0   0
0   0   0   0
 [757]  NA   0  NA  NA   0  NA   0  NA  NA   0  NA  NA  NA   0
NA  NA  NA  NA
 [775]   0   0   0  NA  NA  NA  NA   0  NA  NA   0  NA  NA  NA
NA   0  NA  NA
 [793]  17  25  42  NA  23  33  NA  20  27  11  NA  19  12  20
32  14  24  38
 [811]  NA  22  32  17  31  48  NA  28  38  NA  18  26  17  19
36  19  24  43
 [829]  NA  NA  NA  NA  NA  NA   0  NA  NA  NA   0  NA  NA  NA
NA  NA  NA  NA
 [847]  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA
NA  NA  NA  NA
 [865]   0  NA  NA  NA  NA  NA   0   0   0   0  NA  NA   0  NA
NA  NA  NA  NA
 [883]  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA  NA  NA
NA   0  NA  NA
 [901]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
 [919]   0  NA  NA   0   0   0   0   0   0   0  NA  NA   0   0
0   0   0   0
 [937]   0   0   0   0  NA  NA   0   0   0   0   0   0   0   0
0   0   0   0
 [955]  NA   0  NA   0  NA  NA   0  NA  NA   0   0   0   0   0
0   0   0   0
 [973]   0   0   0   0   0   0   0   0   0   0   0   0   0  NA
NA   0   0   0
 [991]  NA   0  NA   0   0   0   0   0   0   0   0   0   0  NA
```

```
NA   0   0   0
[1009]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1027]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1045]   0   0   0   0   0   0  NA   0  NA  NA   0  NA   0   0
0   0   0   0
[1063]   0  NA  NA   0  NA  NA  NA  NA  NA   0   0   0   0   0
0   0   0   0
[1081]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1099]   0   0   0   0   0   0   0   0   0   0  NA  NA   0   0
0   0   0   0
[1117]  11  14  25  15  18  33  NA  18  24  NA  12  22  14  11
25  16  27  43
[1135]  11  13  24  13  24  37  NA  15  24  13  14  27  20  16
36  14  14  28
[1153]   0  NA  NA   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1171]   0   0   0   0  NA  NA   0  NA  NA   0  NA  NA   0  NA
NA   0   0   0
[1189]  NA  NA  NA   0   0   0   0   0   0   0   0   0   0  NA
NA  NA   0  NA
[1207]   0  NA  NA   0  NA  NA   0   0   0   0   0   0  NA   0
NA   0  NA  NA
[1225]   0   0   0   0   0   0   0  NA  NA   0   0   0   0   0
0   0  NA  NA
[1243]  NA   0  NA   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1261]   0  NA  NA   0   0   0   0   0   0   0  NA  NA  NA  NA
NA  NA   0  NA
[1279]  NA  NA  NA  NA   0  NA  NA  NA  NA   0  NA  NA  NA  NA
NA  NA  NA  NA
[1297]  12  NA  20  11  14  25  11  NA  21  NA  11  17  NA  27
36  14  28  42
[1315]  NA  16  18  16  26  42  11  21  32  NA  NA  13  16  15
31  17  28  45
```

```
[1333]   0   0   0   0   0   0   0   0   0   0   0   0  NA  NA
NA  NA  NA  NA
[1351]   0  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA  NA  NA
NA  NA  NA  NA
[1369]   0   0   0   0   0   0  NA   0  NA   0   0   0   0   0
0  NA   0  NA
[1387]   0   0   0   0   0   0   0   0   0   0  NA  NA   0  NA
NA   0   0   0
[1405]   0  NA  NA  NA  NA  NA   0  NA  NA   0  NA  NA   0  NA
NA  NA  NA  NA
[1423]   0  NA  NA  NA  NA  NA  NA  11  13   0  NA  NA  NA  NA
NA  NA  NA  NA
[1441]  NA  NA  13  NA  NA  12  NA  12  21  NA  NA  13  NA  NA
19  NA  19  29
[1459]  NA  NA  14  NA  20  30  NA  13  18  NA  NA  14  NA  NA
19  NA  11  17
[1477]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1495]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1513]   0   0   0  NA  NA  NA  NA  NA  NA   0   0   0  NA   0
NA   0  NA  NA
[1531]  NA  NA  NA   0   0   0   0   0   0   0   0   0  NA  NA
NA  NA  NA  NA
[1549]   0   0   0   0   0   0   0   0   0   0   0   0   0  NA
NA  NA  NA  NA
[1567]   0  NA  NA   0  NA  NA  NA  NA  NA  NA  NA  NA   0   0
0   0  NA  NA
[1585]   0   0   0   0   0   0   0  NA  NA  NA  NA  NA   0   0
0   0  NA  NA
[1603]  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
NA  NA  NA  NA
[1621]   0   0   0  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA
NA  NA  NA  NA
[1639]  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA  NA  NA
NA  NA  NA  11
[1657]   0   0   0  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
```

```
NA  NA  NA  NA
[1675]  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA
NA  NA  NA  NA
[1693]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1711]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0  NA  NA
[1729]   0   0   0   0   0   0   0   0   0   0  NA  NA   0   0
0   0   0   0
[1747]   0   0   0   0   0   0   0   0   0  NA  NA  NA   0   0
0   0   0   0
[1765]   0   0   0   0   0   0   0   0   0  NA  NA  NA   0  NA
NA   0   0   0
[1783]   0   0   0   0   0   0  NA  NA  NA   0   0   0   0   0
0   0   0   0
[1801]   0  NA  NA   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[1819]   0   0   0   0   0   0   0   0   0   0   0   0  NA   0
NA   0  NA  NA
[1837]  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA   0   0
0  NA  NA  11
[1855]  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA  NA   0  NA
NA  NA  NA  NA
[1873]  NA  13  23  NA  16  26  11  16  27  12  14  26  12  17
29  13  24  37
[1891]  NA  18  28  11  25  36  NA  14  22  NA  19  28  NA  18
26  17  11  28
[1909]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0  NA  NA  NA
[1927]   0   0   0  NA   0  NA   0   0   0   0  NA  NA   0  NA
NA   0   0   0
[1945]   0   0   0   0  NA  NA   0  NA  NA   0  NA  NA   0   0
0   0  NA  NA
[1963]   0   0   0   0   0   0   0   0   0  NA  NA  NA   0   0
0   0   0   0
[1981]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0  NA  NA
```

```
[1999]   0   0   0   0  NA  NA   0  NA  NA   0   0   0  NA   0
NA  NA   0  NA
[2017]   0   0   0   0   0   0   0   0   0   0  NA  NA  NA   0
NA   0   0   0
[2035]   0  NA  NA   0  NA  NA   0  NA  NA   0   0   0   0   0
0   0   0   0
[2053]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[2071]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[2089]   0   0   0   0   0   0   0  NA  NA   0   0   0   0   0
0   0   0   0
[2107]   0   0   0   0   0   0   0   0   0   0   0   0   0   0
0   0   0   0
[2125]   0   0   0   0   0   0   0   0   0   0   0   0   0  NA
NA   0  NA  NA
[2143]  NA  NA  NA   0  NA  NA   0   0   0   0   0   0   0   0
0   0  NA  NA
[2161]   0   0   0  NA  NA  NA   0   0   0   0   0   0   0   0
0   0  NA  NA
[2179]   0   0   0   0   0   0  NA   0  NA  NA  NA  NA   0   0
0  NA   0  NA
[2197]   0   0   0   0   0   0  NA  NA  NA   0  NA  NA   0   0
0   0   0   0
[2215]   0   0   0   0  NA  NA  NA  NA  NA   0  NA  NA   0  NA
NA  NA  NA  NA
[2233]   0   0   0   0   0   0   0   0   0  NA   0  NA   0  NA
NA   0  NA  NA
[2251]   0  NA  NA  NA  NA  NA  NA  NA  NA   0   0   0  NA  NA
NA   0  NA  NA
[2269]  13  13  26  17  15  32  NA  14  23  NA  NA  15  11  18
29  NA  14  21
[2287]  NA  24  33  19  18  37  NA  16  22  NA  16  25  NA  20
30  NA  25  31
[2305]   0  NA  NA  NA   0  NA   0  NA  NA   0   0   0  NA  NA
NA   0   0   0
```

```
[2323]   0  NA  NA  NA  NA  NA  NA  NA  NA   0  NA  NA   0  NA
NA  NA  NA  NA
```

```
CO_heat_ER %>% pull(visits) %>% range(visits) # Incorrect
```

```
CO_heat_ER %>% pull(visits) %>% range(na.rm=T) # Correct
```

```
[1]   0 494
```

Summarization on tibbles (data frames)

Historical CO2 emissions by country

Let's look at a dataset that tracks yearly estimated CO2 emissions by country. We will read it in as a `tibble`.

If you have the `dasehr` package installed successfully:

```
library(dasehr)
yearly_co2 <- yearly_co2_emissions
```

If not, download the `csv` file from https://daseh.org/data/Yearly_CO2_Emissions_1000_tonnes.csv (https://daseh.org/data/Yearly_CO2_Emissions_1000_tonnes.csv) and read it in:

```
yearly_co2 <-
  read_csv(file = "https://daseh.org/data/Yearly_CO2_Emissions_
1000_tonnes.csv")
```

Check out the data:

```
head(yearly_co2)
```

```
# A tibble: 6 × 265
  country  `1751` `1752` `1753` `1754` `1755` `1756` `1757` `17
58` `1759` `1760`
  <chr>     <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <d
bl>  <dbl>  <dbl>
1 Afghani…     NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
2 Albania      NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
3 Algeria      NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
4 Andorra      NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
5 Angola       NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
6 Antigua…     NA     NA     NA     NA     NA     NA     NA
NA     NA     NA
# ℹ 254 more variables: `1761` <dbl>, `1762` <dbl>, `1763` <dbl
>, `1764` <dbl>,
#   `1765` <dbl>, `1766` <dbl>, `1767` <dbl>, `1768` <dbl>, `17
69` <dbl>,
#   `1770` <dbl>, `1771` <dbl>, `1772` <dbl>, `1773` <dbl>, `17
74` <dbl>,
#   `1775` <dbl>, `1776` <dbl>, `1777` <dbl>, `1778` <dbl>, `17
79` <dbl>,
#   `1780` <dbl>, `1781` <dbl>, `1782` <dbl>, `1783` <dbl>, `17
84` <dbl>,
#   `1785` <dbl>, `1786` <dbl>, `1787` <dbl>, `1788` <dbl>, `17
89` <dbl>,
#   `1790` <dbl>, `1791` <dbl>, `1792` <dbl>, `1793` <dbl>, `17
94` <dbl>, …
```

Historical CO2 emissions by country

Check out the data:

```
str(yearly_co2)
```

```
spc_tbl_ [192 × 265] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ country: chr [1:192] "Afghanistan" "Albania" "Algeria" "Ando
rra" ...
 $ 1751   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1752   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1753   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1754   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1755   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1756   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1757   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1758   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1759   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1760   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1761   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1762   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1763   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1764   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1765   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1766   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1767   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1768   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1769   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1770   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1771   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1772   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1773   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1774   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1775   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1776   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1777   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1778   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1779   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1780   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1781   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1782   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1783   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
 $ 1784   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
```

```
$ 1785   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1786   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1787   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1788   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1789   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1790   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1791   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1792   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1793   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1794   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1795   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1796   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1797   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1798   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1799   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1800   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1801   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1802   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1803   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1804   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1805   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1806   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1807   : num [1:192] NA NA NA NA NA NA NA NA NA 169 ...
$ 1808   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1809   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1810   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1811   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1812   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1813   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1814   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1815   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1816   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1817   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1818   : num [1:192] NA NA NA NA NA NA NA NA NA NA ...
$ 1819   : num [1:192] NA NA NA NA NA NA NA NA NA 253 ...
$ 1820   : num [1:192] NA NA NA NA NA NA NA NA NA 334 ...
$ 1821   : num [1:192] NA NA NA NA NA NA NA NA NA 359 ...
```

```
$ 1822   : num [1:192] NA NA NA NA NA NA NA NA NA 367 ...
$ 1823   : num [1:192] NA NA NA NA NA NA NA NA NA 348 ...
$ 1824   : num [1:192] NA NA NA NA NA NA NA NA NA 400 ...
$ 1825   : num [1:192] NA NA NA NA NA NA NA NA NA 403 ...
$ 1826   : num [1:192] NA NA NA NA NA NA NA NA NA 458 ...
$ 1827   : num [1:192] NA NA NA NA NA NA NA NA NA 477 ...
$ 1828   : num [1:192] NA NA NA NA NA NA NA NA NA 458 ...
$ 1829   : num [1:192] NA NA NA NA NA NA NA NA NA 477 ...
$ 1830   : num [1:192] NA NA NA NA NA NA NA 0.032 NA 495 ...
$ 1831   : num [1:192] NA NA NA NA NA NA NA 0.0384 NA 480 ...
$ 1832   : num [1:192] NA NA NA NA NA NA NA 0.0256 NA 513 ...
$ 1833   : num [1:192] NA NA NA NA NA NA NA 0.032 NA 429 ...
$ 1834   : num [1:192] NA NA NA NA NA NA NA NA NA 587 ...
$ 1835   : num [1:192] NA NA NA NA NA NA NA NA NA 634 ...
$ 1836   : num [1:192] NA NA NA NA NA NA NA NA NA 675 ...
$ 1837   : num [1:192] NA NA NA NA NA NA NA NA NA 708 ...
$ 1838   : num [1:192] NA NA NA NA NA NA NA NA NA 851 ...
$ 1839   : num [1:192] NA NA NA NA NA NA NA NA NA 1060 ...
$ 1840   : num [1:192] NA NA NA NA NA NA NA NA NA 1170 ...
$ 1841   : num [1:192] NA NA NA NA NA NA NA NA NA 1320 ...
$ 1842   : num [1:192] NA NA NA NA NA NA NA NA NA 1460 ...
$ 1843   : num [1:192] NA NA NA NA NA NA NA NA NA 1270 ...
$ 1844   : num [1:192] NA NA NA NA NA NA NA NA NA 1600 ...
$ 1845   : num [1:192] NA NA NA NA NA NA NA NA NA 1800 ...
$ 1846   : num [1:192] NA NA NA NA NA NA NA NA NA 2120 ...
$ 1847   : num [1:192] NA NA NA NA NA NA NA NA NA 2080 ...
$ 1848   : num [1:192] NA NA NA NA NA NA NA NA NA 2340 ...
 [list output truncated]
- attr(*, "spec")=
 .. cols(
 ..   country = col_character(),
 ..   `1751` = col_double(),
 ..   `1752` = col_double(),
 ..   `1753` = col_double(),
 ..   `1754` = col_double(),
 ..   `1755` = col_double(),
 ..   `1756` = col_double(),
```

```
..   `1757` = col_double(),
..   `1758` = col_double(),
..   `1759` = col_double(),
..   `1760` = col_double(),
..   `1761` = col_double(),
..   `1762` = col_double(),
..   `1763` = col_double(),
..   `1764` = col_double(),
..   `1765` = col_double(),
..   `1766` = col_double(),
..   `1767` = col_double(),
..   `1768` = col_double(),
..   `1769` = col_double(),
..   `1770` = col_double(),
..   `1771` = col_double(),
..   `1772` = col_double(),
..   `1773` = col_double(),
..   `1774` = col_double(),
..   `1775` = col_double(),
..   `1776` = col_double(),
..   `1777` = col_double(),
..   `1778` = col_double(),
..   `1779` = col_double(),
..   `1780` = col_double(),
..   `1781` = col_double(),
..   `1782` = col_double(),
..   `1783` = col_double(),
..   `1784` = col_double(),
..   `1785` = col_double(),
..   `1786` = col_double(),
..   `1787` = col_double(),
..   `1788` = col_double(),
..   `1789` = col_double(),
..   `1790` = col_double(),
..   `1791` = col_double(),
..   `1792` = col_double(),
..   `1793` = col_double(),
```

```
..   `1794` = col_double(),
..   `1795` = col_double(),
..   `1796` = col_double(),
..   `1797` = col_double(),
..   `1798` = col_double(),
..   `1799` = col_double(),
..   `1800` = col_double(),
..   `1801` = col_double(),
..   `1802` = col_double(),
..   `1803` = col_double(),
..   `1804` = col_double(),
..   `1805` = col_double(),
..   `1806` = col_double(),
..   `1807` = col_double(),
..   `1808` = col_double(),
..   `1809` = col_double(),
..   `1810` = col_double(),
..   `1811` = col_double(),
..   `1812` = col_double(),
..   `1813` = col_double(),
..   `1814` = col_double(),
..   `1815` = col_double(),
..   `1816` = col_double(),
..   `1817` = col_double(),
..   `1818` = col_double(),
..   `1819` = col_double(),
..   `1820` = col_double(),
..   `1821` = col_double(),
..   `1822` = col_double(),
..   `1823` = col_double(),
..   `1824` = col_double(),
..   `1825` = col_double(),
..   `1826` = col_double(),
..   `1827` = col_double(),
..   `1828` = col_double(),
..   `1829` = col_double(),
..   `1830` = col_double(),
```

```
..   `1831` = col_double(),
..   `1832` = col_double(),
..   `1833` = col_double(),
..   `1834` = col_double(),
..   `1835` = col_double(),
..   `1836` = col_double(),
..   `1837` = col_double(),
..   `1838` = col_double(),
..   `1839` = col_double(),
..   `1840` = col_double(),
..   `1841` = col_double(),
..   `1842` = col_double(),
..   `1843` = col_double(),
..   `1844` = col_double(),
..   `1845` = col_double(),
..   `1846` = col_double(),
..   `1847` = col_double(),
..   `1848` = col_double(),
..   `1849` = col_double(),
..   `1850` = col_double(),
..   `1851` = col_double(),
..   `1852` = col_double(),
..   `1853` = col_double(),
..   `1854` = col_double(),
..   `1855` = col_double(),
..   `1856` = col_double(),
..   `1857` = col_double(),
..   `1858` = col_double(),
..   `1859` = col_double(),
..   `1860` = col_double(),
..   `1861` = col_double(),
..   `1862` = col_double(),
..   `1863` = col_double(),
..   `1864` = col_double(),
..   `1865` = col_double(),
..   `1866` = col_double(),
..   `1867` = col_double(),
```

```
..   `1868` = col_double(),
..   `1869` = col_double(),
..   `1870` = col_double(),
..   `1871` = col_double(),
..   `1872` = col_double(),
..   `1873` = col_double(),
..   `1874` = col_double(),
..   `1875` = col_double(),
..   `1876` = col_double(),
..   `1877` = col_double(),
..   `1878` = col_double(),
..   `1879` = col_double(),
..   `1880` = col_double(),
..   `1881` = col_double(),
..   `1882` = col_double(),
..   `1883` = col_double(),
..   `1884` = col_double(),
..   `1885` = col_double(),
..   `1886` = col_double(),
..   `1887` = col_double(),
..   `1888` = col_double(),
..   `1889` = col_double(),
..   `1890` = col_double(),
..   `1891` = col_double(),
..   `1892` = col_double(),
..   `1893` = col_double(),
..   `1894` = col_double(),
..   `1895` = col_double(),
..   `1896` = col_double(),
..   `1897` = col_double(),
..   `1898` = col_double(),
..   `1899` = col_double(),
..   `1900` = col_double(),
..   `1901` = col_double(),
..   `1902` = col_double(),
..   `1903` = col_double(),
..   `1904` = col_double(),
```

```
..   `1905` = col_double(),
..   `1906` = col_double(),
..   `1907` = col_double(),
..   `1908` = col_double(),
..   `1909` = col_double(),
..   `1910` = col_double(),
..   `1911` = col_double(),
..   `1912` = col_double(),
..   `1913` = col_double(),
..   `1914` = col_double(),
..   `1915` = col_double(),
..   `1916` = col_double(),
..   `1917` = col_double(),
..   `1918` = col_double(),
..   `1919` = col_double(),
..   `1920` = col_double(),
..   `1921` = col_double(),
..   `1922` = col_double(),
..   `1923` = col_double(),
..   `1924` = col_double(),
..   `1925` = col_double(),
..   `1926` = col_double(),
..   `1927` = col_double(),
..   `1928` = col_double(),
..   `1929` = col_double(),
..   `1930` = col_double(),
..   `1931` = col_double(),
..   `1932` = col_double(),
..   `1933` = col_double(),
..   `1934` = col_double(),
..   `1935` = col_double(),
..   `1936` = col_double(),
..   `1937` = col_double(),
..   `1938` = col_double(),
..   `1939` = col_double(),
..   `1940` = col_double(),
..   `1941` = col_double(),
```

```
..   `1942` = col_double(),
..   `1943` = col_double(),
..   `1944` = col_double(),
..   `1945` = col_double(),
..   `1946` = col_double(),
..   `1947` = col_double(),
..   `1948` = col_double(),
..   `1949` = col_double(),
..   `1950` = col_double(),
..   `1951` = col_double(),
..   `1952` = col_double(),
..   `1953` = col_double(),
..   `1954` = col_double(),
..   `1955` = col_double(),
..   `1956` = col_double(),
..   `1957` = col_double(),
..   `1958` = col_double(),
..   `1959` = col_double(),
..   `1960` = col_double(),
..   `1961` = col_double(),
..   `1962` = col_double(),
..   `1963` = col_double(),
..   `1964` = col_double(),
..   `1965` = col_double(),
..   `1966` = col_double(),
..   `1967` = col_double(),
..   `1968` = col_double(),
..   `1969` = col_double(),
..   `1970` = col_double(),
..   `1971` = col_double(),
..   `1972` = col_double(),
..   `1973` = col_double(),
..   `1974` = col_double(),
..   `1975` = col_double(),
..   `1976` = col_double(),
..   `1977` = col_double(),
..   `1978` = col_double(),
```

```
..   `1979` = col_double(),
..   `1980` = col_double(),
..   `1981` = col_double(),
..   `1982` = col_double(),
..   `1983` = col_double(),
..   `1984` = col_double(),
..   `1985` = col_double(),
..   `1986` = col_double(),
..   `1987` = col_double(),
..   `1988` = col_double(),
..   `1989` = col_double(),
..   `1990` = col_double(),
..   `1991` = col_double(),
..   `1992` = col_double(),
..   `1993` = col_double(),
..   `1994` = col_double(),
..   `1995` = col_double(),
..   `1996` = col_double(),
..   `1997` = col_double(),
..   `1998` = col_double(),
..   `1999` = col_double(),
..   `2000` = col_double(),
..   `2001` = col_double(),
..   `2002` = col_double(),
..   `2003` = col_double(),
..   `2004` = col_double(),
..   `2005` = col_double(),
..   `2006` = col_double(),
..   `2007` = col_double(),
..   `2008` = col_double(),
..   `2009` = col_double(),
..   `2010` = col_double(),
..   `2011` = col_double(),
..   `2012` = col_double(),
..   `2013` = col_double(),
..   `2014` = col_double()
```

```
  .. )
 - attr(*, "problems")=<externalptr>
```

CO2 Emissions

Before we go further, let's rename the second column using the `rename()` function in `dplyr`.

In this case, we will use the backticks (`) because we will be referring to a column that has a numerical name. If there are funky spaces or characters in the column name, the backticks are required.

```
library(dplyr)
yearly_co2 <- yearly_co2 %>%
  rename(year1751 = `1751`)
```

CO2 Emissions

`colnames()` will show us the column names and show that the `1751` column is renamed:

```
colnames(yearly_co2)
```

```
  [1] "country"  "year1751" "1752"     "1753"     "1754"     "1
755"
  [7] "1756"     "1757"     "1758"     "1759"     "1760"     "1
761"
 [13] "1762"     "1763"     "1764"     "1765"     "1766"     "1
767"
 [19] "1768"     "1769"     "1770"     "1771"     "1772"     "1
773"
 [25] "1774"     "1775"     "1776"     "1777"     "1778"     "1
779"
 [31] "1780"     "1781"     "1782"     "1783"     "1784"     "1
785"
 [37] "1786"     "1787"     "1788"     "1789"     "1790"     "1
791"
 [43] "1792"     "1793"     "1794"     "1795"     "1796"     "1
797"
 [49] "1798"     "1799"     "1800"     "1801"     "1802"     "1
803"
 [55] "1804"     "1805"     "1806"     "1807"     "1808"     "1
809"
 [61] "1810"     "1811"     "1812"     "1813"     "1814"     "1
815"
 [67] "1816"     "1817"     "1818"     "1819"     "1820"     "1
821"
 [73] "1822"     "1823"     "1824"     "1825"     "1826"     "1
827"
 [79] "1828"     "1829"     "1830"     "1831"     "1832"     "1
833"
 [85] "1834"     "1835"     "1836"     "1837"     "1838"     "1
839"
 [91] "1840"     "1841"     "1842"     "1843"     "1844"     "1
845"
 [97] "1846"     "1847"     "1848"     "1849"     "1850"     "1
851"
[103] "1852"     "1853"     "1854"     "1855"     "1856"     "1
857"
[109] "1858"     "1859"     "1860"     "1861"     "1862"     "1
```

```
863"
[115] "1864"    "1865"    "1866"    "1867"    "1868"    "1
869"
[121] "1870"    "1871"    "1872"    "1873"    "1874"    "1
875"
[127] "1876"    "1877"    "1878"    "1879"    "1880"    "1
881"
[133] "1882"    "1883"    "1884"    "1885"    "1886"    "1
887"
[139] "1888"    "1889"    "1890"    "1891"    "1892"    "1
893"
[145] "1894"    "1895"    "1896"    "1897"    "1898"    "1
899"
[151] "1900"    "1901"    "1902"    "1903"    "1904"    "1
905"
[157] "1906"    "1907"    "1908"    "1909"    "1910"    "1
911"
[163] "1912"    "1913"    "1914"    "1915"    "1916"    "1
917"
[169] "1918"    "1919"    "1920"    "1921"    "1922"    "1
923"
[175] "1924"    "1925"    "1926"    "1927"    "1928"    "1
929"
[181] "1930"    "1931"    "1932"    "1933"    "1934"    "1
935"
[187] "1936"    "1937"    "1938"    "1939"    "1940"    "1
941"
[193] "1942"    "1943"    "1944"    "1945"    "1946"    "1
947"
[199] "1948"    "1949"    "1950"    "1951"    "1952"    "1
953"
[205] "1954"    "1955"    "1956"    "1957"    "1958"    "1
959"
[211] "1960"    "1961"    "1962"    "1963"    "1964"    "1
965"
[217] "1966"    "1967"    "1968"    "1969"    "1970"    "1
971"
```

```
[223] "1972"    "1973"    "1974"    "1975"    "1976"    "1
977"
[229] "1978"    "1979"    "1980"    "1981"    "1982"    "1
983"
[235] "1984"    "1985"    "1986"    "1987"    "1988"    "1
989"
[241] "1990"    "1991"    "1992"    "1993"    "1994"    "1
995"
[247] "1996"    "1997"    "1998"    "1999"    "2000"    "2
001"
[253] "2002"    "2003"    "2004"    "2005"    "2006"    "2
007"
[259] "2008"    "2009"    "2010"    "2011"    "2012"    "2
013"
[265] "2014"
```

Summarize the data: `dplyr summarize()` function

`summarize` creates a summary table of a column you're interested in.

Can run multiple summary statistics at once (unlike `pull()` which can only do a single calculation on one column).

You can also do more elaborate summaries across different groups of data using `group_by()`. More on this later!

```
# General format - Not the code!
{data to use} %>%
   summarize({summary column name} = {operator(source column)},
             {summary column name} = {operator(source column)})
```

Summarize the data: `dplyr summarize()` function

`summarize` creates a summary table of a column you're interested in.

```
# General format - Not the code!
{data to use} %>%
   summarize({summary column name} = {operator(source column)})
```

```
yearly_co2 %>%
  summarize(mean_1989 = mean(`1989`)) # Note the backticks, thi
s is a column name!
```

```
# A tibble: 1 × 1
  mean_1989
      <dbl>
1        NA
```

```
yearly_co2 %>%
  summarize(mean_1989 = mean(`1989`, na.rm = TRUE))
```

```
# A tibble: 1 × 1
  mean_1989
      <dbl>
1   126046.
```

Summarize the data: `dplyr summarize()` function

`summarize()` can do multiple operations at once. Just separate by a comma.

```
yearly_co2 %>%
  summarize(mean_1989 = mean(`1989`, na.rm = TRUE),
            median_1989 = median(`1989`, na.rm = TRUE),
            median(`2000`, na.rm = TRUE))
```

```
# A tibble: 1 × 3
  mean_1989 median_1989 `median(\`2000\`, na.rm = TRUE)`
      <dbl>       <dbl>                             <dbl>
1   126046.        8690                              7510
```

Notice how when we forget to provide a new name, output is still provided, but the column name is messy.

Summarize the data: `dplyr summarize()` function

This looks better.

```
yearly_co2 %>%
  summarize(mean_1989 = mean(`1989`, na.rm = TRUE),
            median_1989 = median(`1989`, na.rm = TRUE),
            median_2000 = median(`2000`, na.rm = TRUE))
```

```
# A tibble: 1 × 3
  mean_1989 median_1989 median_2000
      <dbl>       <dbl>       <dbl>
1   126046.        8690        7510
```

Summarize the data: `dplyr summarize()` function

Note that `summarize()` creates a separate tibble from the original data, so you don't want to overwrite your original data if you decide to save the summary.

If you want to save a summary statistic in the original data, use `mutate()` instead to create a new column for the summary statistic.

`summary()` Function

Using `summary()` can give you rough snapshots of each numeric column (character columns are skipped):

```
summary(yearly_co2)
```

```
   country             year1751         1752          1753
1754
 Length:192         Min.   :9360   Min.   :9360   Min.   :9360
Min.   :9370
 Class :character   1st Qu.:9360   1st Qu.:9360   1st Qu.:9360
1st Qu.:9370
 Mode  :character   Median :9360   Median :9360   Median :9360
Median :9370
                    Mean   :9360   Mean   :9360   Mean   :9360
Mean   :9370
                    3rd Qu.:9360   3rd Qu.:9360   3rd Qu.:9360
3rd Qu.:9370
                    Max.   :9360   Max.   :9360   Max.   :9360
Max.   :9370
                    NA's   :191    NA's   :191    NA's   :191
NA's   :191
      1755            1756            1757            1758
1759
 Min.   :9370   Min.   :10000   Min.   :10000   Min.   :10000
Min.   :10000
 1st Qu.:9370   1st Qu.:10000   1st Qu.:10000   1st Qu.:10000
1st Qu.:10000
 Median :9370   Median :10000   Median :10000   Median :10000
Median :10000
 Mean   :9370   Mean   :10000   Mean   :10000   Mean   :10000
Mean   :10000
 3rd Qu.:9370   3rd Qu.:10000   3rd Qu.:10000   3rd Qu.:10000
3rd Qu.:10000
 Max.   :9370   Max.   :10000   Max.   :10000   Max.   :10000
Max.   :10000
 NA's   :191    NA's   :191     NA's   :191     NA's   :191
NA's   :191
      1760            1761            1762            1763
 Min.   :10000   Min.   :11000   Min.   :11000   Min.   :11000
 1st Qu.:10000   1st Qu.:11000   1st Qu.:11000   1st Qu.:11000
 Median :10000   Median :11000   Median :11000   Median :11000
 Mean   :10000   Mean   :11000   Mean   :11000   Mean   :11000
```

```
3rd Qu.:10000   3rd Qu.:11000   3rd Qu.:11000   3rd Qu.:11000
Max.   :10000   Max.   :11000   Max.   :11000   Max.   :11000
NA's   :191     NA's   :191     NA's   :191     NA's   :191
     1764            1765            1766            1767
Min.   :11000   Min.   :11000   Min.   :12300   Min.   :12300
1st Qu.:11000   1st Qu.:11000   1st Qu.:12300   1st Qu.:12300
Median :11000   Median :11000   Median :12300   Median :12300
Mean   :11000   Mean   :11000   Mean   :12300   Mean   :12300
3rd Qu.:11000   3rd Qu.:11000   3rd Qu.:12300   3rd Qu.:12300
Max.   :11000   Max.   :11000   Max.   :12300   Max.   :12300
NA's   :191     NA's   :191     NA's   :191     NA's   :191
     1768            1769            1770            1771
Min.   :12300   Min.   :12300   Min.   :12300   Min.   :13600
1st Qu.:12300   1st Qu.:12300   1st Qu.:12300   1st Qu.:13600
Median :12300   Median :12300   Median :12300   Median :13600
Mean   :12300   Mean   :12300   Mean   :12300   Mean   :13600
3rd Qu.:12300   3rd Qu.:12300   3rd Qu.:12300   3rd Qu.:13600
Max.   :12300   Max.   :12300   Max.   :12300   Max.   :13600
NA's   :191     NA's   :191     NA's   :191     NA's   :191
     1772            1773            1774            1775
Min.   :13600   Min.   :13600   Min.   :13600   Min.   :13600
1st Qu.:13600   1st Qu.:13600   1st Qu.:13600   1st Qu.:13600
Median :13600   Median :13600   Median :13600   Median :13600
Mean   :13600   Mean   :13600   Mean   :13600   Mean   :13600
3rd Qu.:13600   3rd Qu.:13600   3rd Qu.:13600   3rd Qu.:13600
Max.   :13600   Max.   :13600   Max.   :13600   Max.   :13600
NA's   :191     NA's   :191     NA's   :191     NA's   :191
     1776            1777            1778            1779
Min.   :15000   Min.   :15100   Min.   :15100   Min.   :15100
1st Qu.:15000   1st Qu.:15100   1st Qu.:15100   1st Qu.:15100
Median :15000   Median :15100   Median :15100   Median :15100
Mean   :15000   Mean   :15100   Mean   :15100   Mean   :15100
3rd Qu.:15000   3rd Qu.:15100   3rd Qu.:15100   3rd Qu.:15100
Max.   :15000   Max.   :15100   Max.   :15100   Max.   :15100
NA's   :191     NA's   :191     NA's   :191     NA's   :191
     1780            1781            1782            1783
Min.   :15100   Min.   :16900   Min.   :16900   Min.   :16900
```

```
 1st Qu.:15100   1st Qu.:16900   1st Qu.:16900   1st Qu.:16900
 Median :15100   Median :16900   Median :16900   Median :16900
 Mean   :15100   Mean   :16900   Mean   :16900   Mean   :16900
 3rd Qu.:15100   3rd Qu.:16900   3rd Qu.:16900   3rd Qu.:16900
 Max.   :15100   Max.   :16900   Max.   :16900   Max.   :16900
 NA's   :191     NA's   :191     NA's   :191     NA's   :191
      1784            1785                1786                178
7
 Min.   :16900   Min.   :    3.67   Min.   :    3.67   Min.   :
3.67
 1st Qu.:16900   1st Qu.: 4227.75   1st Qu.: 4802.75   1st Qu.:
4802.75
 Median :16900   Median : 8451.83   Median : 9601.83   Median :
9601.83
 Mean   :16900   Mean   : 8451.83   Mean   : 9601.83   Mean   :
9601.83
 3rd Qu.:16900   3rd Qu.:12675.92   3rd Qu.:14400.92   3rd Qu.:
14400.92
 Max.   :16900   Max.   :16900.00   Max.   :19200.00   Max.   :
19200.00
 NA's   :191     NA's   :190        NA's   :190        NA's   :
190
      1788                1789                1790
1791
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.: 4802.75   1st Qu.: 4802.75   1st Qu.: 4802.75   1st Q
u.: 5352.75
 Median : 9601.83   Median : 9601.83   Median : 9601.83   Media
n :10701.83
 Mean   : 9601.83   Mean   : 9601.83   Mean   : 9601.83   Mean
:10701.83
 3rd Qu.:14400.92   3rd Qu.:14400.92   3rd Qu.:14400.92   3rd Q
u.:16050.92
 Max.   :19200.00   Max.   :19200.00   Max.   :19200.00   Max.
:21400.00
 NA's   :190        NA's   :190        NA's   :190        NA's
```

```
:190
      1792                1793                1794
1795
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  236.34   1st Qu.:  241.84   1st Qu.:  223.84   1st Q
u.:  225.34
 Median :  469.00   Median :  480.00   Median :  444.00   Media
n :  447.00
 Mean   : 7290.89   Mean   : 7294.56   Mean   : 7315.89   Mean
: 7316.89
 3rd Qu.:10934.50   3rd Qu.:10940.00   3rd Qu.:10972.00   3rd Q
u.:10973.50
 Max.   :21400.00   Max.   :21400.00   Max.   :21500.00   Max.
:21500.00
 NA's   :189        NA's   :189        NA's   :189        NA's
:189
      1796                1797                1798
1799
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  269.33   1st Qu.:  276.83   1st Qu.:  289.83   1st Q
u.:  315.33
 Median :  535.00   Median :  550.00   Median :  576.00   Media
n :  627.00
 Mean   : 7646.22   Mean   : 8051.22   Mean   : 8359.89   Mean
: 8810.22
 3rd Qu.:11467.50   3rd Qu.:12075.00   3rd Qu.:12538.00   3rd Q
u.:13213.50
 Max.   :22400.00   Max.   :23600.00   Max.   :24500.00   Max.
:25800.00
 NA's   :189        NA's   :189        NA's   :189        NA's
:189
      1800                1801                1802
1803
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
```

```
 1st Qu.:  253.00   1st Qu.:  268.00   1st Qu.:  370.50   1st Q
u.:  297.00
 Median :  407.00   Median :  444.00   Median :  554.00   Media
n :  462.00
 Mean   : 5631.93   Mean   : 5590.13   Mean   : 5262.67   Mean
: 6299.53
 3rd Qu.:  796.00   3rd Qu.:  535.00   3rd Qu.: 4320.00   3rd Q
u.:  535.00
 Max.   :26700.00   Max.   :26700.00   Max.   :26900.00   Max.
:30200.00
 NA's   :187        NA's   :187        NA's   :185        NA's
:187
      1804               1805               1806
1807
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  365.00   1st Qu.:  341.00   1st Qu.:  334.00   1st Q
u.:  221.25
 Median :  734.00   Median :  572.00   Median :  521.00   Media
n :  378.00
 Mean   : 5730.94   Mean   : 6691.33   Mean   : 7019.53   Mean
: 6153.11
 3rd Qu.: 1887.50   3rd Qu.: 1040.00   3rd Qu.:  939.00   3rd Q
u.: 2112.00
 Max.   :30400.00   Max.   :31500.00   Max.   :33300.00   Max.
:33300.00
 NA's   :186        NA's   :187        NA's   :187        NA's
:186
      1808               1809               1810
1811
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  392.00   1st Qu.:  403.00   1st Qu.:  455.50   1st Q
u.:  482.75
 Median :  502.00   Median :  473.00   Median :  773.50   Media
n :  759.00
 Mean   : 7019.13   Mean   : 7022.13   Mean   : 6231.44   Mean
```

```
: 6603.11
 3rd Qu.:  898.00   3rd Qu.:  931.00   3rd Qu.: 1834.75   3rd Q
u.: 1769.50
 Max.   :33300.00   Max.   :33300.00   Max.   :33300.00   Max.
:35600.00
 NA's   :187        NA's   :187        NA's   :186        NA's
:186
      1812               1813               1814
1815
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  462.25   1st Qu.:  416.00   1st Qu.:  516.75   1st Q
u.:  592.75
 Median :  698.50   Median :  706.00   Median :  722.50   Media
n :  815.50
 Mean   : 6845.94   Mean   : 6874.44   Mean   : 7023.44   Mean
: 7260.78
 3rd Qu.: 1893.25   3rd Qu.: 1760.25   3rd Qu.: 1788.50   3rd Q
u.: 2012.50
 Max.   :37000.00   Max.   :37400.00   Max.   :38100.00   Max.
:39000.00
 NA's   :186        NA's   :186        NA's   :186        NA's
:186
      1816               1817               1818
1819
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  667.50   1st Qu.:  725.25   1st Qu.:  816.75   1st Q
u.:  506.00
 Median : 1534.00   Median : 1702.00   Median : 1652.00   Media
n :  763.00
 Mean   : 7955.94   Mean   : 8251.11   Mean   : 8286.44   Mean
: 7145.52
 3rd Qu.: 2472.50   3rd Qu.: 3200.00   3rd Qu.: 3242.50   3rd Q
u.: 2920.00
 Max.   :41500.00   Max.   :42000.00   Max.   :42100.00   Max.
:42400.00
```

```
 NA's   :186         NA's   :186         NA's   :186         NA's
:185
      1820                1821                1822
1823
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  541.00   1st Qu.:  594.00   1st Qu.:  616.00   1st Q
u.:  625.00
 Median :  792.00   Median :  829.00   Median :  909.00   Media
n : 1220.00
 Mean   : 7251.10   Mean   : 7351.52   Mean   : 7649.24   Mean
: 8083.38
 3rd Qu.: 3140.00   3rd Qu.: 3270.00   3rd Qu.: 3400.00   3rd Q
u.: 3605.00
 Max.   :42600.00   Max.   :42900.00   Max.   :44600.00   Max.
:46900.00
 NA's   :185         NA's   :185         NA's   :185         NA's
:185
      1824                1825                1826
1827
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:  710.00   1st Qu.:  771.50   1st Qu.:  839.00   1st Q
u.:  893.50
 Median : 1280.00   Median : 1330.00   Median : 1320.00   Media
n : 1450.00
 Mean   : 8364.81   Mean   : 8682.38   Mean   : 8783.10   Mean
: 9427.24
 3rd Qu.: 3375.00   3rd Qu.: 4150.00   3rd Qu.: 4140.00   3rd Q
u.: 5225.00
 Max.   :49100.00   Max.   :49600.00   Max.   :50200.00   Max.
:52300.00
 NA's   :185         NA's   :185         NA's   :185         NA's
:185
      1828                1829                1830
1831
 Min.   :    3.67   Min.   :    3.67   Min.   :    0.03   Min.
```

```
:    0.04
 1st Qu.:  834.00   1st Qu.:  360.50   1st Qu.:    0.17   1st Q
u.:    0.20
 Median : 1600.00   Median : 1435.00   Median :    3.56   Media
n :    3.67
 Mean   : 9523.10   Mean   : 8308.96   Mean   : 3715.92   Mean
: 3524.28
 3rd Qu.: 5395.00   3rd Qu.: 4242.50   3rd Qu.:  618.75   3rd Q
u.:  562.50
 Max.   :52600.00   Max.   :53400.00   Max.   :67900.00   Max.
:65800.00
 NA's   :185        NA's   :184        NA's   :168        NA's
:168
      1832               1833               1834
1835
 Min.   :    0.03   Min.   :    0.03   Min.   :    3.67   Min.
:    3.67
 1st Qu.:    0.12   1st Qu.:    0.15   1st Qu.:   51.35   1st Q
u.:   33.00
 Median :    2.77   Median :    3.46   Median : 1130.00   Media
n : 1210.00
 Mean   : 3701.44   Mean   : 3774.81   Mean   : 8051.55   Mean
: 8227.94
 3rd Qu.:  771.50   3rd Qu.:  749.50   3rd Qu.: 4415.00   3rd Q
u.: 4850.00
 Max.   :65600.00   Max.   :65500.00   Max.   :65600.00   Max.
:65300.00
 NA's   :169        NA's   :169        NA's   :181        NA's
:181
      1836               1837               1838
1839
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:   53.15   1st Qu.:   53.20   1st Qu.:   56.80   1st Q
u.:   46.77
 Median : 1370.00   Median : 1470.00   Median : 1590.00   Media
n : 1370.00
```

```
 Mean   : 9527.57   Mean   : 9525.80   Mean   : 9834.91   Mean
: 9306.22
 3rd Qu.: 5400.00   3rd Qu.: 5980.00   3rd Qu.: 6385.00   3rd Q
u.: 6790.00
 Max.   :76500.00   Max.   :73800.00   Max.   :75900.00   Max.
:77800.00
 NA's   :181        NA's   :181        NA's   :181        NA's
:180
      1840               1841               1842
1843
 Min.   :    3.67   Min.   :    3.67   Min.   :    3.67   Min.
:    3.67
 1st Qu.:   43.08   1st Qu.:   91.92   1st Qu.:   83.40   1st Q
u.:   88.00
 Median : 1570.00   Median : 1730.00   Median : 1870.00   Media
n : 1270.00
 Mean   : 9915.89   Mean   :10224.06   Mean   :10794.58   Mean
:10233.13
 3rd Qu.: 8092.50   3rd Qu.: 7940.00   3rd Qu.: 8492.50   3rd Q
u.: 7770.00
 Max.   :81000.00   Max.   :81900.00   Max.   :85400.00   Max.
:89200.00
 NA's   :180        NA's   :180        NA's   :180        NA's
:179
      1844               1845               1846
1847
 Min.   :    3.67   Min.   :     3.67   Min.   :   25.7   Min.
:    33.0
 1st Qu.:  158.00   1st Qu.:    99.00   1st Qu.:  131.0   1st Q
u.:   166.8
 Median : 1600.00   Median :  1800.00   Median : 2495.0   Media
n :  2555.0
 Mean   :10886.59   Mean   : 11944.82   Mean   :11277.6   Mean
: 12358.8
 3rd Qu.: 9310.00   3rd Qu.: 11200.00   3rd Qu.:11767.5   3rd Q
u.: 12757.5
 Max.   :93500.00   Max.   :100000.00   Max.   :95800.0   Max.
```

```
:104000.0
 NA's   :179        NA's   :179         NA's   :178       NA's
:178
      1848               1849               1850
1851
 Min.   :    40.3   Min.   :    47.7   Min.   :     0.20   Min.
:    69.7
 1st Qu.:   293.0   1st Qu.:   271.0   1st Qu.:     1.51   1st
Qu.:   284.2
 Median :  3050.0   Median :  3010.0   Median :    70.70   Medi
an :  2730.0
 Mean   : 13349.7   Mean   : 14114.2   Mean   :  7309.46   Mean
: 14239.5
 3rd Qu.: 12800.0   3rd Qu.: 13600.0   3rd Qu.:  2735.00   3rd
Qu.: 15225.0
 Max.   :109000.0   Max.   :116000.0   Max.   :123000.00   Max.
:117000.0
 NA's   :179        NA's   :179        NA's   :165         NA's
:178
      1852               1853               1854
1855
 Min.   :    84.3   Min.   :   103.0   Min.   :   125.0   Min.
:     0.60
 1st Qu.:   336.8   1st Qu.:   348.5   1st Qu.:   441.5   1st Q
u.:     4.61
 Median :  3290.0   Median :  3700.0   Median :  3980.0   Media
n :   214.00
 Mean   : 14812.5   Mean   : 15555.1   Mean   : 18210.8   Mean
:  9659.64
 3rd Qu.: 17250.0   3rd Qu.: 18550.0   3rd Qu.: 21575.0   3rd Q
u.:  4175.00
 Max.   :116000.0   Max.   :116000.0   Max.   :139000.0   Max.
:131000.00
 NA's   :178        NA's   :178        NA's   :178        NA's
:165
      1856             1857              1858                185
9
```

```
 Min.   :   180   Min.   :   216   Min.   :     0.84   Min.   :
0.90
 1st Qu.:   520   1st Qu.:   625   1st Qu.:     5.62   1st Qu.:
5.96
 Median :  4485   Median :  4550   Median :   266.00   Median :
317.00
 Mean   : 19788   Mean   : 20015   Mean   :  9473.59   Mean   :
10048.96
 3rd Qu.: 26475   3rd Qu.: 28200   3rd Qu.:  3070.00   3rd Qu.:
3055.00
 Max.   :140000   Max.   :138000   Max.   :135000.00   Max.   :
150000.00
 NA's   :178      NA's   :178      NA's   :162         NA's   :
162
      1860                1861                1862
1863
 Min.   :     1.18   Min.   :     1.5   Min.   :     1.36   Mi
n.   :     1.42
 1st Qu.:    14.55   1st Qu.:    18.5   1st Qu.:    16.70   1st
Qu.:    17.50
 Median :   279.00   Median :   458.0   Median :   469.00   Med
ian :   506.00
 Mean   :  9454.17   Mean   :  9941.7   Mean   : 10140.36   Mea
n   : 10795.61
 3rd Qu.:  1418.00   3rd Qu.:  1489.0   3rd Qu.:  1710.00   3rd
Qu.:  2015.00
 Max.   :168000.00   Max.   :175000.0   Max.   :170000.00   Ma
x.   :180000.00
 NA's   :157         NA's   :157        NA's   :157         N
A's   :157
      1864                1865                1866
 Min.   :     1.59   Min.   :     1.52   Min.   :     4.81
 1st Qu.:    19.65   1st Qu.:    20.77   1st Qu.:    63.88
 Median :   572.00   Median :   585.00   Median :   702.50
 Mean   : 11641.33   Mean   : 12028.59   Mean   : 12403.70
 3rd Qu.:  2205.00   3rd Qu.:  2365.00   3rd Qu.:  2467.50
 Max.   :194000.00   Max.   :206000.00   Max.   :212000.00
```

```
 NA's   :157         NA's   :156         NA's   :156
      1867                1868                1869
1870
 Min.   :     5.52   Min.   :     4.59   Min.   :6.23e+00   Mi
n.   :     6.76
 1st Qu.:    60.02   1st Qu.:    44.30   1st Qu.:6.01e+01   1st
Qu.:    45.02
 Median :   785.00   Median :   733.00   Median :6.74e+02   Med
ian :   681.50
 Mean   : 13295.67   Mean   : 13261.99   Mean   :1.41e+04   Mea
n   : 14035.93
 3rd Qu.:  2785.00   3rd Qu.:  2660.00   3rd Qu.:2.64e+03   3rd
Qu.:  2980.00
 Max.   :218000.00   Max.   :213000.00   Max.   :2.24e+05   Ma
x.   :229000.00
 NA's   :156         NA's   :155         NA's   :155         N
A's   :154
      1871                1872                1873
1874
 Min.   :     9.12   Min.   :     9.36   Min.   :     8.79   Mi
n.   :    10.7
 1st Qu.:    60.77   1st Qu.:    90.72   1st Qu.:    85.60   1s
t Qu.:   131.2
 Median :   840.50   Median :   859.00   Median :   910.50   Me
dian :  1025.5
 Mean   : 14901.21   Mean   : 16480.25   Mean   : 17536.23   Me
an   : 16387.0
 3rd Qu.:  3172.50   3rd Qu.:  3320.00   3rd Qu.:  3140.00   3r
d Qu.:  3637.5
 Max.   :242000.00   Max.   :255000.00   Max.   :265000.00   Ma
x.   :257000.0
 NA's   :154         NA's   :154         NA's   :154         N
A's   :154
      1875                1876                1877
1878
 Min.   :    12.3   Min.   :    15.2   Min.   :    15.6   Min.
:    20.3
```

```
 1st Qu.:   130.5   1st Qu.:   147.0   1st Qu.:   150.0   1st Q
u.:   250.5
 Median :  1305.0   Median :  1500.0   Median :  1420.0   Media
n :  1450.0
 Mean   : 17822.5   Mean   : 18549.3   Mean   : 18891.4   Mean
: 18087.3
 3rd Qu.:  4107.5   3rd Qu.:  5440.0   3rd Qu.:  5550.0   3rd Q
u.:  5535.0
 Max.   :272000.0   Max.   :271000.0   Max.   :276000.0   Max.
:271000.0
 NA's   :154        NA's   :155        NA's   :155        NA's
:153
      1879                1880                1881
1882
 Min.   :     3.67   Min.   :    24.5   Min.   :     3.67   Mi
n.   :     3.67
 1st Qu.:   123.25   1st Qu.:   211.8   1st Qu.:   183.00   1st
Qu.:   206.50
 Median :  1550.00   Median :  1970.0   Median :  1770.00   Med
ian :  2065.00
 Mean   : 18907.85   Mean   : 21371.8   Mean   : 21537.81   Mea
n   : 23334.29
 3rd Qu.:  5337.50   3rd Qu.:  5925.0   3rd Qu.:  5960.00   3rd
Qu.:  7000.00
 Max.   :273000.00   Max.   :297000.0   Max.   :313000.00   Ma
x.   :315000.00
 NA's   :152         NA's   :152         NA's   :151         N
A's   :152
      1883                1884                1885
1886
 Min.   :     3.7   Min.   :     3.7   Min.   :     3.67   Min.
:     7.33
 1st Qu.:   175.0   1st Qu.:   176.5   1st Qu.:   200.50   1st
Qu.:   151.75
 Median :  2430.0   Median :  2195.0   Median :  2295.00   Medi
an :  2260.00
 Mean   : 24191.0   Mean   : 23891.9   Mean   : 24059.08   Mean
```

```
: 24452.96
 3rd Qu.:  7440.0   3rd Qu.:  8300.0   3rd Qu.:  8347.50   3rd
Qu.:  8685.00
 Max.   :328000.0   Max.   :319000.0   Max.   :316000.00   Max.
:313000.00
 NA's   :151        NA's   :150        NA's   :150         NA's
:150
      1887               1888               1889
1890
 Min.   :     7.3   Min.   :     3.7   Min.   :     3.7   Min.
:    11.0
 1st Qu.:   228.0   1st Qu.:   201.0   1st Qu.:   218.0   1st Q
u.:   289.8
 Median :  2260.0   Median :  2130.0   Median :  2380.0   Media
n :  2500.0
 Mean   : 25096.8   Mean   : 27101.0   Mean   : 27115.5   Mean
: 29521.9
 3rd Qu.:  9995.0   3rd Qu.: 10900.0   3rd Qu.: 10775.0   3rd Q
u.: 10775.0
 Max.   :321000.0   Max.   :374000.0   Max.   :345000.0   Max.
:402000.0
 NA's   :149        NA's   :148        NA's   :148         NA's
:148
      1891              1892               1893                18
94
 Min.   :    22   Min.   :    47.7   Min.   :    47.7   Min.
:    40.3
 1st Qu.:   297   1st Qu.:   341.5   1st Qu.:   414.0   1st Q
u.:   382.5
 Median :  2420   Median :  2470.0   Median :  2460.0   Median
:  2910.0
 Mean   : 30241   Mean   : 29174.8   Mean   : 28813.7   Mean
: 29819.6
 3rd Qu.: 11000   3rd Qu.:  9850.0   3rd Qu.:  9750.0   3rd Q
u.: 11300.0
 Max.   :428000   Max.   :451000.0   Max.   :454000.0   Max.
:425000.0
```

```
 NA's   :147      NA's   :145        NA's   :145        NA's
:145
      1895               1896               1897
1898
 Min.   :    40.3   Min.   :    40.3   Min.   :    40.3   Min.
:    88.0
 1st Qu.:   415.2   1st Qu.:   442.2   1st Qu.:   522.2   1st Q
u.:   633.5
 Median :  3200.0   Median :  3295.0   Median :  3695.0   Media
n :  3940.0
 Mean   : 30957.5   Mean   : 31953.2   Mean   : 33496.1   Mean
: 35332.7
 3rd Qu.: 11700.0   3rd Qu.: 11675.0   3rd Qu.: 12200.0   3rd Q
u.: 12525.0
 Max.   :480000.0   Max.   :481000.0   Max.   :501000.0   Max.
:546000.0
 NA's   :144        NA's   :144        NA's   :144        NA's
:144
      1899               1900               1901               19
02
 Min.   :    95.3   Min.   :   131.0   Min.   :   135   Min.
:    95.3
 1st Qu.:   671.0   1st Qu.:   824.2   1st Qu.:   950   1st Q
u.:   816.0
 Median :  4050.0   Median :  3340.0   Median :  3900   Median
:  3645.0
 Mean   : 37808.5   Mean   : 40689.0   Mean   : 41192   Mean
: 41385.8
 3rd Qu.: 13300.0   3rd Qu.: 14300.0   3rd Qu.: 14400   3rd Q
u.: 14825.0
 Max.   :626000.0   Max.   :663000.0   Max.   :722000   Max.
:765000.0
 NA's   :143        NA's   :144        NA's   :143      NA's
:142
      1903               1904               1905               19
06
 Min.   :   114.0   Min.   :     3.7   Min.   :   126   Min.
```

```
:      3.7
 1st Qu.:   965.5   1st Qu.:   803.0   1st Qu.:  1006   1st Q
u.:    831.2
 Median :  3170.0   Median :  3620.0   Median :  4140   Median
:   5360.0
 Mean   : 44259.5   Mean   : 43866.0   Mean   : 47644   Mean
:  48996.1
 3rd Qu.: 15300.0   3rd Qu.: 16200.0   3rd Qu.: 17250   3rd Q
u.:  18850.0
 Max.   :895000.0   Max.   :881000.0   Max.   :985000   Max.
:1030000.0
 NA's   :141        NA's   :140        NA's   :141      NA's
:140
      1907                1908                1909
 Min.   :      7.3   Min.   :      3.7   Min.   :     25.7
 1st Qu.:    908.0   1st Qu.:   1210.0   1st Qu.:   1280.0
 Median :   5570.0   Median :   5720.0   Median :   5940.0
 Mean   :  55450.9   Mean   :  52345.0   Mean   :  54351.6
 3rd Qu.:  19650.0   3rd Qu.:  22800.0   3rd Qu.:  20900.0
 Max.   :1200000.0   Max.   :1050000.0   Max.   :1160000.0
 NA's   :140         NA's   :139         NA's   :139
      1910                1911                1912
1913
 Min.   :     51.3   Min.   :     11   Min.   :      3.7   Min.
:      7.3
 1st Qu.:   1125.0   1st Qu.:   1074   1st Qu.:   1220.0   1st
Qu.:   1220.0
 Median :   6030.0   Median :   6730   Median :   7760.0   Medi
an :   6920.0
 Mean   :  56043.8   Mean   :  56078   Mean   :  59721.2   Mean
:  61191.5
 3rd Qu.:  20375.0   3rd Qu.:  24350   3rd Qu.:  24025.0   3rd
Qu.:  23200.0
 Max.   :1270000.0   Max.   :1260000   Max.   :1340000.0   Max.
:1440000.0
 NA's   :138         NA's   :137       NA's   :138         NA's
:135
```

```
      1914                1915                1916
1917
 Min.   :     11   Min.   :     18.3   Min.   :      3.7   Min.
:      7.3
 1st Qu.:   1100   1st Qu.:    818.8   1st Qu.:    639.8   1st
Qu.:    528.0
 Median :   6020   Median :   5595.0   Median :   4350.0   Medi
an :   3350.0
 Mean   :  55650   Mean   :  53950.2   Mean   :  56245.2   Mean
:  57865.4
 3rd Qu.:  24500   3rd Qu.:  21175.0   3rd Qu.:  19600.0   3rd
Qu.:  18900.0
 Max.   :1320000   Max.   :1370000.0   Max.   :1520000.0   Max.
:1670000.0
 NA's   :135       NA's   :134         NA's   :132         NA's
:131
      1918                1919                1920
 Min.   :     18.3   Min.   :     18.3   Min.   :      3.7
 1st Qu.:    400.0   1st Qu.:    522.2   1st Qu.:    463.2
 Median :   3190.0   Median :   3400.0   Median :   3605.0
 Mean   :  57115.9   Mean   :  50238.6   Mean   :  56517.5
 3rd Qu.:  19000.0   3rd Qu.:  19250.0   3rd Qu.:  18950.0
 Max.   :1750000.0   Max.   :1480000.0   Max.   :1740000.0
 NA's   :131         NA's   :132         NA's   :130
      1921                1922                1923
1924
 Min.   :      7.3   Min.   :     25.7   Min.   :1.47e+01   Mi
n.   :     29.3
 1st Qu.:    447.0   1st Qu.:    678.0   1st Qu.:7.64e+02   1st
Qu.:   1007.5
 Median :   3340.0   Median :   3780.0   Median :4.62e+03   Med
ian :   5900.0
 Mean   :  48705.7   Mean   :  52677.4   Mean   :5.99e+04   Mea
n   :  59103.7
 3rd Qu.:  21450.0   3rd Qu.:  24400.0   3rd Qu.:2.65e+04   3rd
Qu.:  27625.0
 Max.   :1420000.0   Max.   :1430000.0   Max.   :1.90e+06   Ma
```

```
x.   :1700000.0
 NA's   :129         NA's   :131         NA's   :131         N
A's   :130
      1925               1926                1927
1928
 Min.   :     33   Min.   :     40.3   Min.   :     40.3   Min.
:      3.7
 1st Qu.:   1024   1st Qu.:   1375.0   1st Qu.:   1007.0   1st
Qu.:    893.8
 Median :   5710   Median :   6215.0   Median :   6410.0   Medi
an :   4970.0
 Mean   :  59397   Mean   :  58343.3   Mean   :  61780.6   Mean
:  56374.9
 3rd Qu.:  28825   3rd Qu.:  29175.0   3rd Qu.:  29075.0   3rd
Qu.:  21425.0
 Max.   :1740000   Max.   :1890000.0   Max.   :1850000.0   Max.
:1830000.0
 NA's   :130       NA's   :130         NA's   :128         NA's
:122
      1929                1930                1931
 Min.   :      3.7   Min.   :      7.3   Min.   :      3.7
 1st Qu.:    925.8   1st Qu.:    949.5   1st Qu.:    766.0
 Median :   5070.0   Median :   5210.0   Median :   4240.0
 Mean   :  60483.5   Mean   :  55735.0   Mean   :  47769.0
 3rd Qu.:  23325.0   3rd Qu.:  22975.0   3rd Qu.:  19600.0
 Max.   :1960000.0   Max.   :1740000.0   Max.   :1480000.0
 NA's   :122         NA's   :122         NA's   :119
      1932                1933                1934
 Min.   :      3.7   Min.   :      3.7   Min.   :      3.7
 1st Qu.:    694.5   1st Qu.:    397.0   1st Qu.:    663.0
 Median :   4000.0   Median :   4035.0   Median :   4780.0
 Mean   :  41984.1   Mean   :  41418.7   Mean   :  45669.2
 3rd Qu.:  20250.0   3rd Qu.:  18900.0   3rd Qu.:  22100.0
 Max.   :1260000.0   Max.   :1350000.0   Max.   :1440000.0
 NA's   :117         NA's   :112         NA's   :113
      1935                1936                1937
 Min.   :      3.7   Min.   :      7.3   Min.   :     18.3
```

```
 1st Qu.:    703.0   1st Qu.:    773.5   1st Qu.:    795.5
 Median :   5450.0   Median :   5450.0   Median :   6110.0
 Mean   :  47815.3   Mean   :  51868.8   Mean   :  55471.3
 3rd Qu.:  23150.0   3rd Qu.:  25225.0   3rd Qu.:  28600.0
 Max.   :1490000.0   Max.   :1710000.0   Max.   :1790000.0
 NA's   :113         NA's   :112         NA's   :112
      1938                1939                1940
 Min.   :     33.0   Min.   :      3.7   Min.   :      7.3
 1st Qu.:    927.5   1st Qu.:    888.0   1st Qu.:   1360.0
 Median :   5920.0   Median :   5860.0   Median :   6490.0
 Mean   :  52758.9   Mean   :  54467.7   Mean   :  60953.2
 3rd Qu.:  26750.0   3rd Qu.:  29200.0   3rd Qu.:  25600.0
 Max.   :1510000.0   Max.   :1670000.0   Max.   :1870000.0
 NA's   :113         NA's   :111         NA's   :113
      1941                1942                1943
1944
 Min.   :      3.7   Min.   :     11   Min.   :      7.3   Min.
:      3.7
 1st Qu.:    792.8   1st Qu.:   1040   1st Qu.:    696.8   1st
Qu.:    770.0
 Median :   5130.0   Median :   5350   Median :   5970.0   Medi
an :   4880.0
 Mean   :  60166.1   Mean   :  63079   Mean   :  62584.6   Mean
:  62709.2
 3rd Qu.:  22325.0   3rd Qu.:  23100   3rd Qu.:  21725.0   3rd
Qu.:  20300.0
 Max.   :2040000.0   Max.   :2200000   Max.   :2270000.0   Max.
:2440000.0
 NA's   :110         NA's   :113       NA's   :110         NA's
:111
      1945                1946                1947
 Min.   :      3.7   Min.   :      3.7   Min.   :      3.7
 1st Qu.:    590.0   1st Qu.:    686.0   1st Qu.:    656.0
 Median :   4415.0   Median :   4980.0   Median :   5550.0
 Mean   :  49118.8   Mean   :  51108.6   Mean   :  57390.8
 3rd Qu.:  14250.0   3rd Qu.:  22200.0   3rd Qu.:  25100.0
 Max.   :2360000.0   Max.   :2250000.0   Max.   :2480000.0
```

```
 NA's   :108         NA's   :107         NA's   :103
      1948                1949                1950
 Min.   :      3.7   Min.   :      3.7   Min.   :      3.7
 1st Qu.:    346.2   1st Qu.:    477.0   1st Qu.:    221.2
 Median :   5835.0   Median :   5610.0   Median :   1485.0
 Mean   :  59757.7   Mean   :  56160.9   Mean   :  42926.1
 3rd Qu.:  24350.0   3rd Qu.:  26600.0   3rd Qu.:  10110.0
 Max.   :2580000.0   Max.   :2160000.0   Max.   :2540000.0
 NA's   :102         NA's   :99          NA's   :58
      1951                1952                1953
1954
 Min.   :      3.7   Min.   :      3.7   Min.   :3.70e+00   Mi
n.   :      3.7
 1st Qu.:    197.2   1st Qu.:    230.0   1st Qu.:2.49e+02   1st
Qu.:    309.8
 Median :   1425.0   Median :   1515.0   Median :1.65e+03   Med
ian :   1745.0
 Mean   :  45396.8   Mean   :  46074.2   Mean   :4.70e+04   Mea
n   :  47613.0
 3rd Qu.:  10525.0   3rd Qu.:  10975.0   3rd Qu.:1.08e+04   3rd
Qu.:  12375.0
 Max.   :2610000.0   Max.   :2550000.0   Max.   :2.61e+06   Ma
x.   :2490000.0
 NA's   :56          NA's   :56          NA's   :55         N
A's   :54
      1955                1956                1957
 Min.   :      3.7   Min.   :      3.7   Min.   :      3.7
 1st Qu.:    312.0   1st Qu.:    319.0   1st Qu.:    313.5
 Median :   2120.0   Median :   2220.0   Median :   2280.0
 Mean   :  50991.2   Mean   :  54305.1   Mean   :  55257.0
 3rd Qu.:  13600.0   3rd Qu.:  14400.0   3rd Qu.:  15200.0
 Max.   :2720000.0   Max.   :2860000.0   Max.   :2830000.0
 NA's   :51          NA's   :51          NA's   :49
      1958                1959                1960
1961
 Min.   :      3.7   Min.   :      3.7   Min.   :     11   Min.
:      3.7
```

```
 1st Qu.:    319.0   1st Qu.:    235.0   1st Qu.:    286   1st
Qu.:    287.0
 Median :   2170.0   Median :   1840.0   Median :   2210   Medi
an :   2315.0
 Mean   :  54709.1   Mean   :  54666.4   Mean   :  57562   Mean
:  57157.4
 3rd Qu.:  15700.0   3rd Qu.:  15700.0   3rd Qu.:  16100   3rd
Qu.:  16550.0
 Max.   :2740000.0   Max.   :2830000.0   Max.   :2890000   Max.
:2880000.0
 NA's   :43          NA's   :35          NA's   :35        NA's
:34
      1962                1963                1964
1965
 Min.   :     11.0   Min.   :      7.3   Min.   :      7   Min.
:      7
 1st Qu.:    282.5   1st Qu.:    301.0   1st Qu.:    306   1st
Qu.:    304
 Median :   2315.0   Median :   2300.0   Median :   2270   Medi
an :   2470
 Mean   :  58458.5   Mean   :  61413.2   Mean   :  62221   Mean
:  65113
 3rd Qu.:  17625.0   3rd Qu.:  21100.0   3rd Qu.:  21950   3rd
Qu.:  23800
 Max.   :2990000.0   Max.   :3120000.0   Max.   :3260000   Max.
:3390000
 NA's   :32          NA's   :31          NA's   :25        NA's
:25
      1966              1967              1968              196
9
 Min.   :      7   Min.   :      7   Min.   :      7   Min.   :
7
 1st Qu.:    343   1st Qu.:    438   1st Qu.:    494   1st Qu.:
539
 Median :   2640   Median :   3050   Median :   3305   Median :
3860
 Mean   :  68105   Mean   :  70162   Mean   :  74441   Mean   :
```

```
78894
 3rd Qu.:  25050   3rd Qu.:  24650   3rd Qu.:  27350   3rd Qu.:
30700
 Max.   :3560000   Max.   :3700000   Max.   :3830000   Max.   :
4020000
 NA's   :25        NA's   :25        NA's   :26        NA's   :
25
      1970              1971              1972              197
3
 Min.   :      4   Min.   :      4   Min.   :      4   Min.   :
4
 1st Qu.:    638   1st Qu.:    640   1st Qu.:    671   1st Qu.:
654
 Median :   3810   Median :   4285   Median :   4680   Median :
5290
 Mean   :  84575   Mean   :  87463   Mean   :  90846   Mean   :
95682
 3rd Qu.:  28400   3rd Qu.:  29950   3rd Qu.:  31550   3rd Qu.:
37800
 Max.   :4330000   Max.   :4360000   Max.   :4560000   Max.   :
4770000
 NA's   :23        NA's   :22        NA's   :21        NA's   :
21
      1974              1975              1976              197
7
 Min.   :      4   Min.   :      4   Min.   :      4   Min.   :
7
 1st Qu.:    682   1st Qu.:    645   1st Qu.:    648   1st Qu.:
774
 Median :   5410   Median :   5550   Median :   5850   Median :
5720
 Mean   :  95405   Mean   :  95535   Mean   : 100698   Mean   :
103565
 3rd Qu.:  36350   3rd Qu.:  37400   3rd Qu.:  41700   3rd Qu.:
42800
 Max.   :4600000   Max.   :4410000   Max.   :4610000   Max.   :
4740000
```

```
 NA's   :21        NA's   :21        NA's   :21        NA's   :
21
      1978              1979              1980              198
1
 Min.   :     11   Min.   :     22   Min.   :     22   Min.   :
26
 1st Qu.:    711   1st Qu.:    798   1st Qu.:    774   1st Qu.:
691
 Median :   6020   Median :   6750   Median :   6460   Median :
6260
 Mean   : 107348   Mean   : 110216   Mean   : 109769   Mean   :
106308
 3rd Qu.:  46450   3rd Qu.:  48450   3rd Qu.:  48500   3rd Qu.:
47475
 Max.   :4890000   Max.   :4900000   Max.   :4720000   Max.   :
4540000
 NA's   :21        NA's   :21        NA's   :21        NA's   :
20
      1982              1983              1984              198
5
 Min.   :     26   Min.   :     22   Min.   :     22   Min.   :
22
 1st Qu.:    806   1st Qu.:    850   1st Qu.:    839   1st Qu.:
855
 Median :   6255   Median :   6795   Median :   6810   Median :
7605
 Mean   : 105815   Mean   : 106863   Mean   : 109962   Mean   :
113989
 3rd Qu.:  43800   3rd Qu.:  49600   3rd Qu.:  49475   3rd Qu.:
51100
 Max.   :4310000   Max.   :4340000   Max.   :4480000   Max.   :
4490000
 NA's   :20        NA's   :20        NA's   :20        NA's   :
20
      1986              1987              1988              198
9
 Min.   :     18   Min.   :     22   Min.   :     22   Min.   :
```

```
22
 1st Qu.:    887   1st Qu.:    973   1st Qu.:   1006   1st Qu.:
1028
 Median :   7910   Median :   7825   Median :   8295   Median :
8690
 Mean   : 115512   Mean   : 119590   Mean   : 124048   Mean   :
126046
 3rd Qu.:  53500   3rd Qu.:  57700   3rd Qu.:  57600   3rd Qu.:
54750
 Max.   :4500000   Max.   :4690000   Max.   :4890000   Max.   :
4960000
 NA's   :20        NA's   :20        NA's   :20        NA's   :
20
      1990              1991              1992              199
3
 Min.   :      7   Min.   :      7   Min.   :      7   Min.   :
7
 1st Qu.:    895   1st Qu.:    961   1st Qu.:    955   1st Qu.:
1018
 Median :   6405   Median :   6010   Median :   6990   Median :
7090
 Mean   : 119786   Mean   : 119810   Mean   : 113426   Mean   :
114527
 3rd Qu.:  52200   3rd Qu.:  57000   3rd Qu.:  49350   3rd Qu.:
50600
 Max.   :4820000   Max.   :4820000   Max.   :4910000   Max.   :
5030000
 NA's   :16        NA's   :15        NA's   :4         NA's   :
4
      1994              1995              1996              199
7
 Min.   :      7   Min.   :      7   Min.   :      7   Min.   :
7
 1st Qu.:    968   1st Qu.:    953   1st Qu.:   1060   1st Qu.:
1120
 Median :   6230   Median :   6960   Median :   7260   Median :
7570
```

```
 Mean   : 114230   Mean   : 116464   Mean   : 119634   Mean   :
120062
 3rd Qu.:  54800   3rd Qu.:  54800   3rd Qu.:  53100   3rd Qu.:
55900
 Max.   :5090000   Max.   :5130000   Max.   :5250000   Max.   :
5370000
 NA's   :3         NA's   :3         NA's   :3         NA's   :
3
      1998              1999              2000              200
1
 Min.   :      7   Min.   :      7   Min.   :      7   Min.   :
7
 1st Qu.:   1170   1st Qu.:   1240   1st Qu.:   1190   1st Qu.:
1270
 Median :   7360   Median :   7550   Median :   7510   Median :
7880
 Mean   : 119381   Mean   : 121199   Mean   : 124827   Mean   :
125861
 3rd Qu.:  57100   3rd Qu.:  55400   3rd Qu.:  53800   3rd Qu.:
57000
 Max.   :5400000   Max.   :5500000   Max.   :5690000   Max.   :
5600000
 NA's   :3         NA's   :3         NA's   :3         NA's   :
3
      2002              2003              2004              200
5
 Min.   :     11   Min.   :     11   Min.   :     11   Min.   :
11
 1st Qu.:   1230   1st Qu.:   1308   1st Qu.:   1435   1st Qu.:
1455
 Median :   8045   Median :   7890   Median :   8105   Median :
8560
 Mean   : 127903   Mean   : 134504   Mean   : 140855   Mean   :
145649
 3rd Qu.:  57075   3rd Qu.:  58675   3rd Qu.:  58800   3rd Qu.:
58900
 Max.   :5640000   Max.   :5680000   Max.   :5760000   Max.   :
```

```
5900000
 NA's   :2         NA's   :2         NA's   :2         NA's   :
2
      2006              2007              2008              200
9
 Min.   :      7   Min.   :     11   Min.   :     11   Min.   :
11
 1st Qu.:   1610   1st Qu.:   1670   1st Qu.:   1770   1st Qu.:
1760
 Median :   9480   Median :   9650   Median :   9140   Median :
8110
 Mean   : 150693   Mean   : 154232   Mean   : 158692   Mean   :
157166
 3rd Qu.:  62800   3rd Qu.:  62750   3rd Qu.:  66250   3rd Qu.:
62500
 Max.   :6530000   Max.   :7030000   Max.   :7550000   Max.   :
8000000
 NA's   :2         NA's   :1         NA's   :1         NA's   :
1
      2010              2011              2012              20
13
 Min.   :      7   Min.   :      7   Min.   :      11   Min.
:      11
 1st Qu.:   1960   1st Qu.:   2090   1st Qu.:    2148   1st Q
u.:    2172
 Median :   8600   Median :   9830   Median :   10100   Median
:   10700
 Mean   : 165334   Mean   : 171765   Mean   :  174033   Mean
:  174856
 3rd Qu.:  64750   3rd Qu.:  63750   3rd Qu.:   63625   3rd Q
u.:   64525
 Max.   :8780000   Max.   :9730000   Max.   :10000000   Max.
:10300000
 NA's   :1         NA's   :1
      2014
 Min.   :      11
 1st Qu.:    2190
```

```
 Median :   11300
 Mean   :  175993
 3rd Qu.:   63775
 Max.   :10300000
```

Summary & Lab Part 1

- summary stats (`mean()`) work with `pull()`
- don't forget the `na.rm = TRUE` argument!
- `summary(x)` : quantile information
- `summarize` : creates a summary table of columns of interest

 Class Website (https://daseh.org/)

 Lab (https://daseh.org/modules/Data_Summarization/lab/Data_Summarization_Lab.Rmd)

CO ER Heat Illness Visits

Let's go back to the dataset of CO ER visits for heat-related illness. Remember, we loaded this data into our session and saved it as the object `CO_heat_ER` .

```
head(CO_heat_ER)
```

```
# A tibble: 6 × 7
  county     rate lower95cl upper95cl visits  year gender
  <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
1 Statewide  5.64      4.70      6.59    140  2011 Female
2 Statewide  7.39      6.30      8.47    183  2011 Male
3 Statewide  6.51      5.80      7.23    323  2011 Both genders
4 Statewide  5.64      4.72      6.57    146  2012 Female
5 Statewide  7.56      6.48      8.65    193  2012 Male
6 Statewide  6.58      5.88      7.29    339  2012 Both genders
```

`distinct()` values

`distinct(x)` will return the unique elements of column `x` .

```
CO_heat_ER %>%
  distinct(county)
```

```
# A tibble: 65 × 1
   county
   <chr>
 1 Statewide
 2 Adams
 3 Alamosa
 4 Arapahoe
 5 Archuleta
 6 Baca
 7 Bent
 8 Boulder
 9 Broomfield
10 Chaffee
# ℹ 55 more rows
```

How many `distinct()` values?

`n_distinct()` tells you the number of unique elements. *Must pull the column first!*

```
CO_heat_ER %>%
  pull(county) %>%
  n_distinct()
```

```
[1] 65
```

`dplyr`: `count`

Use `count` to return a frequency table of unique elements of a data.frame.

```
CO_heat_ER %>% count(county)
```

```
# A tibble: 65 × 2
   county         n
   <chr>      <int>
 1 Adams         36
 2 Alamosa       36
 3 Arapahoe      36
 4 Archuleta     36
 5 Baca          36
 6 Bent          36
 7 Boulder       36
 8 Broomfield    36
 9 Chaffee       36
10 Cheyenne      36
# ℹ 55 more rows
```

dplyr: count

Multiple columns listed further subdivides the count.

```
CO_heat_ER %>% count(county, gender)
```

```
# A tibble: 195 × 3
   county    gender           n
   <chr>     <chr>        <int>
 1 Adams     Both genders    12
 2 Adams     Female          12
 3 Adams     Male            12
 4 Alamosa   Both genders    12
 5 Alamosa   Female          12
 6 Alamosa   Male            12
 7 Arapahoe  Both genders    12
 8 Arapahoe  Female          12
 9 Arapahoe  Male            12
10 Archuleta Both genders    12
# ℹ 185 more rows
```

Note: `count()` includes NAs

Grouping

Perform Operations By Groups: dplyr

`group_by` allows you group the data set by variables/columns you specify:

```
# Regular data
CO_heat_ER
```

```
# A tibble: 2,340 × 7
   county     rate lower95cl upper95cl visits  year gender
   <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
 1 Statewide  5.64      4.70      6.59    140  2011 Female
 2 Statewide  7.39      6.30      8.47    183  2011 Male
 3 Statewide  6.51      5.80      7.23    323  2011 Both gender
s
 4 Statewide  5.64      4.72      6.57    146  2012 Female
 5 Statewide  7.56      6.48      8.65    193  2012 Male
 6 Statewide  6.58      5.88      7.29    339  2012 Both gender
s
 7 Statewide  4.94      4.06      5.82    124  2013 Female
 8 Statewide  6.72      5.72      7.72    178  2013 Male
 9 Statewide  5.82      5.16      6.49    302  2013 Both gender
s
10 Statewide  3.52      2.80      4.25     92  2014 Female
# ℹ 2,330 more rows
```

Perform Operations By Groups: dplyr

`group_by` allows you group the data set by variables/columns you specify:

```
CO_heat_ER_grouped <- CO_heat_ER %>% group_by(gender)
CO_heat_ER_grouped
```

```
# A tibble: 2,340 × 7
# Groups:   gender [3]
   county     rate lower95cl upper95cl visits  year gender
   <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
 1 Statewide  5.64      4.70      6.59    140  2011 Female
 2 Statewide  7.39      6.30      8.47    183  2011 Male
 3 Statewide  6.51      5.80      7.23    323  2011 Both gender
s
 4 Statewide  5.64      4.72      6.57    146  2012 Female
 5 Statewide  7.56      6.48      8.65    193  2012 Male
 6 Statewide  6.58      5.88      7.29    339  2012 Both gender
s
 7 Statewide  4.94      4.06      5.82    124  2013 Female
 8 Statewide  6.72      5.72      7.72    178  2013 Male
 9 Statewide  5.82      5.16      6.49    302  2013 Both gender
s
10 Statewide  3.52      2.80      4.25     92  2014 Female
# ℹ 2,330 more rows
```

Summarize the grouped data

It's grouped! Grouping doesn't change the data in any way, but how **functions operate on it**. Now we can summarize `Data_Value` (percent of respondents) by group:

```
CO_heat_ER_grouped %>% summarize(avg_visits = mean(visits, na.r
m = TRUE))
```

```
# A tibble: 3 × 2
  gender       avg_visits
  <chr>             <dbl>
1 Both genders      16.3
2 Female             4.77
3 Male               9.00
```

Use the `pipe` to string these together!

Pipe `yts` into `group_by`, then pipe that into `summarize`:

```
CO_heat_ER %>%
  group_by(gender) %>%
  summarize(avg_visits = mean(visits, na.rm = TRUE),
            max_visits = max(visits, na.rm = TRUE))
```

```
# A tibble: 3 × 3
  gender       avg_visits max_visits
  <chr>             <dbl>      <dbl>
1 Both genders      16.3         494
2 Female             4.77        185
3 Male               9.00        309
```

Group by as many variables as you want

`group_by` Response and Education:

```
CO_heat_ER %>%
  group_by(gender, year) %>%
  summarize(avg_visits = mean(visits, na.rm = TRUE),
            max_visits = max(visits, na.rm = TRUE))
```

```
# A tibble: 36 × 4
# Groups:   gender [3]
   gender        year avg_visits max_visits
   <chr>        <dbl>      <dbl>      <dbl>
 1 Both genders  2011      11.3         323
 2 Both genders  2012      12.8         339
 3 Both genders  2013      12.4         302
 4 Both genders  2014       9.67        237
 5 Both genders  2015      14.9         355
 6 Both genders  2016      22.4         467
 7 Both genders  2017      16.3         323
 8 Both genders  2018      25.6         456
 9 Both genders  2019      20.3         389
10 Both genders  2020      14.5         302
# ℹ 26 more rows
```

Only the last `group_by` is recognized…

You can overwrite the first `group_by` with a new one.

```
CO_heat_ER %>%
  group_by(gender, year)  %>%
  group_by(year)
```

```
# A tibble: 2,340 × 7
# Groups:   year [12]
   county     rate lower95cl upper95cl visits  year gender
   <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
 1 Statewide  5.64      4.70      6.59    140  2011 Female
 2 Statewide  7.39      6.30      8.47    183  2011 Male
 3 Statewide  6.51      5.80      7.23    323  2011 Both gender
s
 4 Statewide  5.64      4.72      6.57    146  2012 Female
 5 Statewide  7.56      6.48      8.65    193  2012 Male
 6 Statewide  6.58      5.88      7.29    339  2012 Both gender
s
 7 Statewide  4.94      4.06      5.82    124  2013 Female
 8 Statewide  6.72      5.72      7.72    178  2013 Male
 9 Statewide  5.82      5.16      6.49    302  2013 Both gender
s
10 Statewide  3.52      2.80      4.25     92  2014 Female
# ℹ 2,330 more rows
```

Ungroup the data

The `ungroup` function will allow you to clear the groups from the data.

```
CO_heat_ER <- ungroup(CO_heat_ER)
CO_heat_ER
```

```
# A tibble: 2,340 × 7
   county     rate lower95cl upper95cl visits  year gender
   <chr>     <dbl>     <dbl>     <dbl>  <dbl> <dbl> <chr>
 1 Statewide  5.64      4.70      6.59    140  2011 Female
 2 Statewide  7.39      6.30      8.47    183  2011 Male
 3 Statewide  6.51      5.80      7.23    323  2011 Both gender
s
 4 Statewide  5.64      4.72      6.57    146  2012 Female
 5 Statewide  7.56      6.48      8.65    193  2012 Male
 6 Statewide  6.58      5.88      7.29    339  2012 Both gender
s
 7 Statewide  4.94      4.06      5.82    124  2013 Female
 8 Statewide  6.72      5.72      7.72    178  2013 Male
 9 Statewide  5.82      5.16      6.49    302  2013 Both gender
s
10 Statewide  3.52      2.80      4.25     92  2014 Female
# ℹ 2,330 more rows
```

`group_by` with `mutate` - just add data

We can also use `mutate` to calculate the mean value for each year and add it as a column:

```
CO_heat_ER %>%
  group_by(year, gender) %>%
  mutate(visits_year_avg = mean(visits, na.rm = TRUE)) %>%
  select(county, visits, visits_year_avg)
```

```
# A tibble: 2,340 × 5
# Groups:   year, gender [36]
    year gender       county    visits visits_year_avg
   <dbl> <chr>        <chr>      <dbl>           <dbl>
 1  2011 Female       Statewide    140            4.32
 2  2011 Male         Statewide    183            6.06
 3  2011 Both genders Statewide    323           11.3
 4  2012 Female       Statewide    146            4.76
 5  2012 Male         Statewide    193            6.71
 6  2012 Both genders Statewide    339           12.8
 7  2013 Female       Statewide    124            3.72
 8  2013 Male         Statewide    178            6.11
 9  2013 Both genders Statewide    302           12.4
10  2014 Female       Statewide     92            2.5
# ℹ 2,330 more rows
```

Counting

There are other functions, such as `n()` count the number of observations (NAs included).

```
CO_heat_ER %>%
  group_by(year) %>%
  summarize(n = n(),
            mean = mean(visits, na.rm = TRUE))
```

```
# A tibble: 12 × 3
    year     n  mean
   <dbl> <int> <dbl>
 1  2011   195  7.17
 2  2012   195  8.14
 3  2013   195  7.33
 4  2014   195  5.51
 5  2015   195  8.68
 6  2016   195 13.2
 7  2017   195  9.39
 8  2018   195 14.7
 9  2019   195 12.3
10  2020   195  8.45
11  2021   195 11.6
12  2022   195 13.3
```

Counting

`count()` and `n()` can give very similar information.

```
CO_heat_ER %>% count(county)
```

```
# A tibble: 65 × 2
   county         n
   <chr>      <int>
 1 Adams         36
 2 Alamosa       36
 3 Arapahoe      36
 4 Archuleta     36
 5 Baca          36
 6 Bent          36
 7 Boulder       36
 8 Broomfield    36
 9 Chaffee       36
10 Cheyenne      36
# ℹ 55 more rows
```

```
CO_heat_ER %>% group_by(county) %>% summarize(n()) # n() typically used with summarize
```

```
# A tibble: 65 × 2
   county     `n()`
   <chr>      <int>
 1 Adams         36
 2 Alamosa       36
 3 Arapahoe      36
 4 Archuleta     36
 5 Baca          36
 6 Bent          36
 7 Boulder       36
 8 Broomfield    36
 9 Chaffee       36
10 Cheyenne      36
# ℹ 55 more rows
```

A few miscellaneous topics ..

Base R functions you might see: `length` and `unique`

These functions require a column as a vector using `pull()`.

```
CO_heat_ER_loc <- CO_heat_ER %>% pull(county) # pull() to make a vector
CO_heat_ER_loc %>% unique() # similar to distinct()
```

```
 [1] "Statewide"   "Adams"       "Alamosa"     "Arapahoe"    "A
rchuleta"
 [6] "Baca"        "Bent"        "Boulder"     "Broomfield"  "C
haffee"
[11] "Cheyenne"    "Clear Creek" "Conejos"     "Costilla"    "C
rowley"
[16] "Custer"      "Delta"       "Denver"      "Dolores"     "D
ouglas"
[21] "Eagle"       "Elbert"      "El Paso"     "Fremont"     "G
arfield"
[26] "Gilpin"      "Grand"       "Gunnison"    "Hinsdale"    "H
uerfano"
[31] "Jackson"     "Jefferson"   "Kiowa"       "Kit Carson"  "L
ake"
[36] "La Plata"    "Larimer"     "Las Animas"  "Lincoln"     "L
ogan"
[41] "Mesa"        "Mineral"     "Moffat"      "Montezuma"   "M
ontrose"
[46] "Morgan"      "Otero"       "Ouray"       "Park"        "P
hillips"
[51] "Pitkin"      "Prowers"     "Pueblo"      "Rio Blanco"  "R
io Grande"
[56] "Routt"       "Saguache"    "San Juan"    "San Miguel"  "S
edgwick"
[61] "Summit"      "Teller"      "Washington"  "Weld"        "Y
uma"
```

Base R functions you might see: `length` and `unique`

These functions require a column as a vector using `pull()`.

```
CO_heat_ER_loc %>% unique() %>% length() # similar to n_distinc
t()
```

```
[1] 65
```

* New! * Many dplyr functions now have a `.by=` argument

Pipe `CO_heat_ER` into `group_by`, then pipe that into `summarize`:

```
CO_heat_ER %>%
  group_by(county) %>%
  summarize(avg_visits = mean(visits, na.rm = TRUE),
            max_visits = max(visits, na.rm = TRUE))
```

is the same as..

```
CO_heat_ER %>%
  summarize(avg_visits = mean(visits, na.rm = TRUE),
            max_visits = max(visits, na.rm = TRUE),
            .by = county)
```

`summary()` vs. `summarize()`

- `summary()` (base R) gives statistics table on a dataset.
- `summarize()` (dplyr) creates a more customized summary tibble/dataframe.

Summary & Lab Part 2

- `count(x)`: what unique values do you have?
  - `distinct()`: what are the distinct values?
  - `n_distinct()` with `pull()`: how many distinct values?
- `group_by()`: changes all subsequent functions
  - combine with `summarize()` to get statistics per group
  - combine with `mutate()` to add column
- `summarize()` with `n()` gives the count (NAs included)

 Class Website (https://daseh.org/)

 Lab (https://daseh.org/modules/Data_Summarization/lab/Data_Summarization_Lab.Rmd)

Extra Slides: More advanced summarization

Data Summarization on data frames

- Statistical summarization across the data frame
  - `rowMeans(x)` : takes the means of each row of x
  - `colMeans(x)` : takes the means of each column of x
  - `rowSums(x)` : takes the sum of each row of x
  - `colSums(x)` : takes the sum of each column of x

`rowMeans()` example

Get means for each row.

Let's see what the mean CO2 emissions is across years for each row (country):

```
yearly_co2 %>%
  select(starts_with("year")) %>%
  rowMeans(na.rm = TRUE) %>%
  head(n = 5)
```

```
[1] NaN NaN NaN NaN NaN
```

```
yearly_co2 %>%
  group_by(country) %>%
  summarize(mean = rowMeans(across(starts_with("year")), na.rm
 = TRUE)) %>%
  head(n = 5)
```

```
# A tibble: 5 × 2
  country      mean
  <chr>       <dbl>
1 Afghanistan   NaN
2 Albania       NaN
3 Algeria       NaN
4 Andorra       NaN
5 Angola        NaN
```

`colMeans()` example

Get means for each column.

Let's see what the mean is across each column (year):

```
yearly_co2 %>%
  select(starts_with("year")) %>%
  colMeans(na.rm = TRUE) %>%
  head(n = 5)
```

```
year1751
    9360
```

```
yearly_co2 %>%
  summarize(across(starts_with("year"), ~mean(.x, na.rm = TRUE
)))
```

```
# A tibble: 1 × 1
  year1751
     <dbl>
1     9360
```