

Lecture 12: Tibbles

STAT 450, Fall 2021

Reading: Chapter 10 from *R for Data Science*

Useful reference: <https://tibble.tidyverse.org/index.html>

Tibbles are the representation of data frames in the tidyverse. Tibbles are a modern take on the traditional `data.frame` in base R.

```
library(tidyverse)
```

Creating tibbles

```
tb <- tibble(  
  x = seq(0, 1, by=0.25),  
  y = c("a", "a", "a", "b", "b")  
)  
tb
```

```
## # A tibble: 5 x 2  
##       x y  
##   <dbl> <chr>  
## 1  0     a  
## 2 0.25   a  
## 3  0.5   a  
## 4 0.75   b  
## 5  1     b
```

```
class(tb)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

Subsetting

Subsetting works in much the same way as a traditional R `data.frame`

```
tb$x
```

```
## [1] 0.00 0.25 0.50 0.75 1.00
```

```
tb[, 1] # returns a tibble
```

```
## # A tibble: 5 x 1
```

```
##       x
```

```
##   <dbl>
```

```
## 1  0
```

```
## 2 0.25
```

```
## 3  0.5
```

```
## 4 0.75
```

```
## 5  1
```

```
tb[[1]] # returns a vector
```

```
## [1] 0.00 0.25 0.50 0.75 1.00
```

```
tb[1, ]
```

```
## # A tibble: 1 x 2
```

```
##       x y
```

```
##   <dbl> <chr>
```

```
## 1     0 a
```

Printing

A nice feature of tibbles is that when you type the name of a data frame, only the first 10 rows and all columns that fit on the screen are displayed. The type of each column (variable) is also shown under its name.

```
mpg
```

```
## # A tibble: 234 x 11
```

```
##   manufacturer model      displ  year   cyl trans drv      cty   hwy fl      class
```

```
##   <chr>          <chr>    <dbl> <int> <int> <chr> <chr> <int> <int> <chr> <chr>
```

```
## 1 audi          a4         1.8  1999   4 auto~ f      18    29 p    comp~
```

```
## 2 audi          a4         1.8  1999   4 manu~ f      21    29 p    comp~
```

```
## 3 audi          a4         2    2008   4 manu~ f      20    31 p    comp~
```

```
## 4 audi          a4         2    2008   4 auto~ f      21    30 p    comp~
```

```
## 5 audi          a4         2.8  1999   6 auto~ f      16    26 p    comp~
```

```
## 6 audi          a4         2.8  1999   6 manu~ f      18    26 p    comp~
```

```
## 7 audi          a4         3.1  2008   6 auto~ f      18    27 p    comp~
```

```
## 8 audi          a4 quattro 1.8  1999   4 manu~ 4      18    26 p    comp~
```

```
## 9 audi          a4 quattro 1.8  1999   4 auto~ 4      16    25 p    comp~
```

```
## 10 audi         a4 quattro 2    2008   4 manu~ 4      20    28 p    comp~
```

```
## # ... with 224 more rows
```

To see the entire data set use `View()`. This will open RStudio's built-in data viewer to get a scrollable view of the complete data set. Type the following command to view the entire `mpg` data frame.

```
mpg %>% View()
```

Another useful way to look at your data:

```
mpg %>% glimpse()
```

The `dplyr` functions `slice_head()` and `slice_tail()` can also be used to select the first or last rows. For example, run the following code:

```
mpg %>% slice_head(n=5)
mpg %>% slice_tail(n=5)
```

Coercion

Many R packages use the traditional base R `data.frame`, so you might want to coerce a `data.frame` to a tibble. You can do that with `as_tibble()`:

```
class(mtcars)
```

```
## [1] "data.frame"
```

```
mtcars_tb <- as_tibble(mtcars)
mtcars_tb
```

```
## # A tibble: 32 x 11
##   mpg   cyl  disp    hp  drat    wt   qsec    vs  am  gear  carb
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21     6  160   110  3.9   2.62  16.5     0     1     4     4
## 2  21     6  160   110  3.9   2.88  17.0     0     1     4     4
## 3 22.8     4  108    93  3.85  2.32  18.6     1     1     4     1
## 4 21.4     6  258   110  3.08  3.22  19.4     1     0     3     1
## 5 18.7     8  360   175  3.15  3.44  17.0     0     0     3     2
## 6 18.1     6  225   105  2.76  3.46  20.2     1     0     3     1
## 7 14.3     8  360   245  3.21  3.57  15.8     0     0     3     4
## 8 24.4     4  147.    62  3.69  3.19  20      1     0     4     2
## 9 22.8     4  141.    95  3.92  3.15  22.9     1     0     4     2
## 10 19.2     6  168.   123  3.92  3.44  18.3     1     0     4     4
## # ... with 22 more rows
```

```
class(mtcars_tb)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

Some older R code doesn't work with tibbles. So, conversely, you can use `as.data.frame()` to coerce a tibble back to a `data.frame`:

```
class(mpg)
```

```
## [1] "tbl_df"      "tbl"        "data.frame"
```

```
mpg_df <- as.data.frame(mpg)
class(mpg_df)
```

```
## [1] "data.frame"
```

By default `as_tibble()` removes row names. Use the `rownames` argument to covert the row names into a new column.

```
mtcars_tb <- as_tibble(mtcars, rownames = "car")
mtcars_tb
```

```
## # A tibble: 32 x 12
##   car          mpg   cyl  disp    hp  drat    wt   qsec    vs  am  gear  carb
##   <chr>        <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Mazda RX4      21     6  160   110   3.9   2.62  16.5     0    1     4     4
## 2 Mazda RX4 ~    21     6  160   110   3.9   2.88  17.0     0    1     4     4
## 3 Datsun 710    22.8     4  108    93   3.85   2.32  18.6     1    1     4     1
## 4 Hornet 4 D~   21.4     6  258   110   3.08   3.22  19.4     1    0     3     1
## 5 Hornet Spo~   18.7     8  360   175   3.15   3.44  17.0     0    0     3     2
## 6 Valiant       18.1     6  225   105   2.76   3.46  20.2     1    0     3     1
## 7 Duster 360    14.3     8  360   245   3.21   3.57  15.8     0    0     3     4
## 8 Merc 240D     24.4     4  147.    62   3.69   3.19   20      1    0     4     2
## 9 Merc 230      22.8     4  141.    95   3.92   3.15  22.9     1    0     4     2
## 10 Merc 280     19.2     6  168.   123   3.92   3.44  18.3     1    0     4     4
## # ... with 22 more rows
```

Exercise: Type the following into a code chunk in R Markdown, and then knit to HTML. How does printing a traditional `data.frame` in R Markdown differ from printing a tibble?

```
iris # print iris data.frame
class(iris)
iris_tb <- as_tibble(iris) # convert iris to a tibble
iris_tb # print iris tibble
class(iris_tb)
```