Lab 5: Data cleaning

Fill in your full name

2025-10-02

Exercises

Exercise 1

```
##  [1] "tow_date"           "make"               "style"
##  [4] "model"              "color"              "plate"
##  [7] "state"              "towed_to_address"   "tow_facility_phone"
## [10] "inventory_number"
```

Exercise 2

```
## tibble [5,185 x 10] (S3: tbl_df/tbl/data.frame)
##  $ tow_date          : Date[1:5185], format: "2018-03-30" "2018-03-30" ...
##  $ make              : chr [1:5185] "VOLK" "LEXS" "CHEV" "BUIC" ...
##  $ style             : chr [1:5185] "4D" "LL" "LL" "4D" ...
##  $ model             : chr [1:5185] NA NA "TK" NA ...
##  $ color             : chr [1:5185] "WHI" "TAN" "MAR" "BLK" ...
##  $ plate             : chr [1:5185] "988T647" "3875313" NA "AG71211" ...
##  $ state             : chr [1:5185] "IL" "IN" "IL" "IL" ...
##  $ towed_to_address  : chr [1:5185] "701 N. Sacramento" "701 N. Sacramento" "701 N. Sacramen
##  $ tow_facility_phone: chr [1:5185] "(773) 265-7605" "(773) 265-7605" "(773) 265-7605" "(773
##  $ inventory_number  : int [1:5185] 6919167 6919168 6921673 2845513 6921674 1718330 2851569
##  - attr(*, "spec")=
##   .. cols(
##   ..   `Tow Date` = col_character(),
##   ..   Make = col_character(),
##   ..   Style = col_character(),
##   ..   Model = col_character(),
##   ..   Color = col_character(),
##   ..   Plate = col_character(),
##   ..   State = col_character(),
##   ..   `Towed to Address` = col_character(),
##   ..   `Tow Facility Phone` = col_character(),
##   ..   `Inventory Number` = col_integer()
##   .. )
```

```
## [1] "Date"
```

Exercise 3

```
## [1] 78
```

make	n
AIRS	1
BUCI	1
CADE	1
CHEC	1
CHYI	1
CYCL	1
DAEW	1
DODD	1
FIAT	1
GM	1
HARL	1
HUMM	1
INTL	1
LEIR	1
LNDR	1
MASE	1
MCCY	1
MERD	1
MG	1
MIST	1
OPEL	1
PONY	1
RAMS	1
ROV	1
ROYC	1
SPAR	1
TRLR	1
WORK	1

Exercise 4

```
tow_make_count_1 <- tow_make_count %>%
  slice(1:20)

tow_make_count_2 <- tow_make_count %>%
  slice(21:40)

tow_make_count_3 <- tow_make_count %>%
  slice(41:60)
```

```r
tow_make_count_4 <- tow_make_count %>%
  slice(61:80)
```

Exercise 5

```r
towed_clean <- towed_renamed_2 %>%
  mutate(
    make = recode(
      make,
      DODD = "DODG",
      CHEVV = "CHEV",
      TOYTA = "TOYT"
    )
  )
```

Exercise 6

```r
towed_partial_clean <- towed_renamed_2 %>%
  mutate(
    make = recode(
      make,
      DODD = "DODG",
      BUCI = "BUIC"
    )
  )
```

Exercise 7

```r
towed_partial_clean <- towed_partial_clean %>%
  mutate(
    make = recode(
      make,
      JAG = "JAGU"
    )
  )
```

Exercise 8

```r
towed_make_clean <- towed_partial_clean %>%
  mutate(
    make = recode(
      make,
      "DODD" = "DODG",   # Dodge
      "DODD" = "DODG",   # sometimes a zero sneaks in
      "BUCI" = "BUIC",   # Buick
      "JAG"  = "JAGU"    # Jaguar (per VTR-249.pdf)
    )
  )
```

Additional questions

Question 1

-The most heavily imbued car manufacturer is the first value of the count (make, sort=TRUE) result.

Question 2

-The most impounded date is the first row of count(tow_date, sort=TRUE), and the least impounded date is the last row of the same table.