class03Oct.qmd

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.2     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
df <- data.frame(c(1,3,5,8,10,15))

df$a = letters[1:6]
df$b = LETTERS[1:6]
str(df)
```

```
'data.frame':   6 obs. of  3 variables:
 $ c.1..3..5..8..10..15.: num  1 3 5 8 10 15
 $ a                    : chr  "a" "b" "c" "d" ...
 $ b                    : chr  "A" "B" "C" "D" ...
```

```
df = mutate_if(.tbl = df, .predicate = is.factor, .funs = as.character)
str(df)
```

```
'data.frame':   6 obs. of  3 variables:
 $ c.1..3..5..8..10..15.: num  1 3 5 8 10 15
 $ a                    : chr  "a" "b" "c" "d" ...
 $ b                    : chr  "A" "B" "C" "D" ...
```

Strings and regular expressions

```
  (empty_str <- "")
```

```
[1] ""
```

```
  class(empty_str)
```

```
[1] "character"
```

```
  (empty_chr <- character(length = 0))
```

```
character(0)
```

```
  class(empty_chr)
```

```
[1] "character"
```

```
  empty_chr[1] <- "first"
  empty_chr
```

```
[1] "first"
```

```
  empty_chr[4] <- "fourth"
  empty_chr
```

```
[1] "first"  NA       NA       "fourth"
```

```r
paste("I", "love", "R", sep = "-")
```

```
[1] "I-love-R"
```

```r
paste("X", 1:5, sep = ".", collapse = "-")
```

```
[1] "X.1-X.2-X.3-X.4-X.5"
```

```r
#readLines() to read text into R as is
top105 <- readLines("http://www.textfiles.com/music/ktop100.txt")
head(top105, n = 20)
```

```
 [1] "From: ed@wente.llnl.gov (Ed Suranyi)"
 [2] "Date: 12 Jan 92 21:23:55 GMT"
 [3] "Newsgroups: rec.music.misc"
 [4] "Subject: KITS' year end countdown"
 [5] ""
 [6] ""
 [7] "On Jan. 1, 1992, the \"Modern Rock\" station KITS San Francisco (\"Live-105\")"
 [8] "broadcast its list of the \"Top 105.3 of 1991.\"  Here is the countdown"
 [9] "list:"
[10] ""
[11] "1. NIRVANA                       SMELLS LIKE TEEN SPIRIT"
[12] "2. EMF                           UNBELIEVABLE"
[13] "3. R.E.M.                        LOSING MY RELIGION"
[14] "4. SIOUXSIE & THE BANSHEES       KISS THEM FOR ME"
[15] "5. B.A.D. II                     RUSH"
[16] "6. RED HOT CHILI PEPPERS         GIVE IT AWAY"
[17] "7. ELECTRONIC                    GET THE MESSAGE"
[18] "8. ERASURE                       CHORUS"
[19] "9. SCHOOL OF FISH                3 STRANGE DAYS"
[20] "10. NORTHSIDE                    TAKE FIVE"
```

```r
print(top105[1])
```

```
[1] "From: ed@wente.llnl.gov (Ed Suranyi)"
```

```
cat(top105[1:50], sep = " + ", file = "top105_cat_output.txt")
```

Regular Expression

Pattern that describes a set of strings.

```
help("regex")
```

```
Help on topic 'regex' was found in the following packages:

  Package               Library
  stringr               /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/library
  base                  /Library/Frameworks/R.framework/Resources/library


Using the first match ...
```

metacharacters

. \ | ( ) [{ $ * + ?

```
money = "$money"
sub(pattern = "$", replacement = "", x = money)
```

```
[1] "$money"
```

```
money = "$money"
sub(pattern = "\\$", replacement = "", x = money)
```

```
[1] "money"
```