```
(base) Vasu's MacBook Pro:~ vasugoel$ r

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(tidyverse)
── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──
✔ ggplot2 3.2.1     ✔ purrr   0.3.2
✔ tibble  2.1.3     ✔ dplyr   0.8.3
✔ tidyr   0.8.3     ✔ stringr 1.4.0
✔ readr   1.3.1     ✔ forcats 0.4.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
> library(lubridate)

Attaching package: ‘lubridate’

The following object is masked from ‘package:base’:

    date

> library(purrr)
> library(pdftools)
> fn <- system.file("extdata", "RD-Mortality-Report_2015-18-180531.pdf", package="dslabs")
> dat <- map_df(str_split(pdf_text(fn), "\n"), function(s){
+ s <- str_trim(s)
+ header_index <- str_which(s, "2015")[1]
+ tmp <- str_split(s[header_index], "\\s+", simplify = TRUE)
+ month <- tmp[1]
+ header <- tmp[-1]
+ tail_index  <- str_which(s, "Total")
+ n <- str_count(s, "\\d+")
+ out <- c(1:header_index, which(n==1), which(n>=28), tail_index:length(s))
+ s[-out] %>%
+ str_remove_all("[^\\d\\s]") %>%
+ str_trim() %>%
+ str_split_fixed("\\s+", n = 6) %>%
+ .[,1:5] %>%
+ as_data_frame() %>%
+ setNames(c("day", header)) %>%
+ mutate(month = month,
+ day = as.numeric(day)) %>%
+ gather(year, deaths, -c(day, month)) %>%
+ mutate(deaths = as.numeric(deaths))
+ }) %>%
+ mutate(month = recode(month, "JAN" = 1, "FEB" = 2, "MAR" = 3, "APR" = 4, "MAY" = 5, "JUN" = 6,
+                           "JUL" = 7, "AGO" = 8, "SEP" = 9, "OCT" = 10, "NOV" = 11, "DEC" = 12)) %>%
+ mutate(date = make_date(year, month, day)) %>%
+       filter(date <= "2018-05-01")
```

```
Warning message:
`as_data_frame()` is deprecated, use `as_tibble()` (but mind the new semantics).
This warning is displayed once per session.
>
> dim(dat)
[1] 1205    5
> head(dat)
# A tibble: 6 x 5
    day month year  deaths date
  <dbl> <dbl> <chr>  <dbl> <date>
1     1     1 2015     107 2015-01-01
2     2     1 2015     101 2015-01-02
3     3     1 2015      78 2015-01-03
4     4     1 2015     121 2015-01-04
5     5     1 2015      99 2015-01-05
6     6     1 2015     104 2015-01-06
> tail(dat)
# A tibble: 6 x 5
    day month year  deaths date
  <dbl> <dbl> <chr>  <dbl> <date>
1    26    12 2017     103 2017-12-26
2    27    12 2017      95 2017-12-27
3    28    12 2017      93 2017-12-28
4    29    12 2017      83 2017-12-29
5    30    12 2017      87 2017-12-30
6    31    12 2017     102 2017-12-31
>
>
>
> range(dat$date)
[1] "2015-01-01" "2018-05-01"
> diff(range(dat$date))
Time difference of 1216 days
> as.numeric(diff(range(dat$date)))
[1] 1216
> span <- 60 / as.numeric(diff(range(dat$date)))
> fit <- dat %>% mutate(x = as.numeric(date)) %>% loess(deaths ~ x, data = ., span = span, degree = 1)
> dat %>% mutate(smooth = predict(fit, as.numeric(date))) %>%
+ ggplot() +
+ geom_point(aes(date, deaths)) +
+ geom_line(aes(date, smooth), lwd = 2, col = 2)
Warning message:
Removed 1 rows containing missing values (geom_point).
>
>
> dat %>%
+ mutate(smooth = predict(fit, as.numeric(date)), day = yday(date), year = as.character(year(date))) %>%
+ ggplot(aes(day, smooth, col = year)) +
+ geom_line(lwd = 2)
>
>
> library(broom)
> library(dslabs)
> mnist_27$train %>% glm(y ~ x_2, family = "binomial", data = .) %>% tidy()
# A tibble: 2 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)  -0.0907     0.247    -0.368   0.713
2 x_2           0.685      0.827     0.829   0.407
>
> mnist_27$train %>% head()
  y        x_1        x_2
1 2 0.03947368 0.18421053
2 7 0.16071429 0.08928571
3 2 0.02127660 0.27659574
```

```
4 2 0.13580247 0.22222222
5 7 0.39024390 0.36585366
6 2 0.04854369 0.28155340
> mnist_27$train %>% mutate(y = ifelse(y == 7, 1, 0)) %>% ggplot(aes(x_2, y)) + geom_point() + geom_smooth
(method = loess)
>
```