pivots

27 March, 2023

Write the R code to answer the following questions. You have until the beginning of next class to answer all of the questions below and commit to GitHub.

Overview

We will continue using the polls data from class containing presidential primary polls for the 2020 election. As a reminder, these are data shared with me. Please do not use beyond class without inquiring with me further, and do not post publicly.

Question 1

So far in class we've pivoted our data so every candidate & date combination (for two dates) represents a row.

Tweak this code to so we consider *all* dates of polls in NV and we find who is the candidate leading in the polls on each date. You'll need to use our dplyr skills from last week, too.

This is the code from class to get you started:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(readr)
primaryPolls <- read_csv('president_primary_polls_feb2020.csv')
```

```
## Rows: 16661 Columns: 33
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (21): state, pollster, sponsors, display_name, pollster_rating_name, fte...
## dbl  (8): question_id, poll_id, cycle, pollster_id, pollster_rating_id, samp...
## num  (1): sponsor_ids
## lgl  (3): internal, tracking, nationwide_batch
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
primaryPolls$start_date <- as.Date(primaryPolls$start_date, "%m/%d/%y")


nevadaPrimaries <- primaryPolls %>%
 filter(candidate_name %in% c("Amy Klobuchar", "Bernard Sanders",
                              "Elizabeth Warren", "Joseph R. Biden Jr.",
                              "Michael Bloomberg", "Pete Buttigieg")) %>%
  filter(state == "Nevada") %>%
  select(poll_id, candidate_name, pct, start_date)
nevadaPrimaries
```

```
## # A tibble: 76 x 4
##    poll_id candidate_name        pct start_date
##      <dbl> <chr>               <dbl> <date>
##  1   63269 Joseph R. Biden Jr.  19.4 2020-01-08
##  2   63269 Bernard Sanders      17.6 2020-01-08
##  3   63269 Elizabeth Warren     10.6 2020-01-08
##  4   63269 Pete Buttigieg        8.2 2020-01-08
##  5   63269 Amy Klobuchar         3.6 2020-01-08
##  6   63254 Elizabeth Warren     14   2020-01-06
##  7   63254 Bernard Sanders      29   2020-01-06
##  8   63254 Joseph R. Biden Jr.  28   2020-01-06
##  9   63254 Pete Buttigieg        6   2020-01-06
## 10   63254 Amy Klobuchar         4   2020-01-06
## # ... with 66 more rows
```

```r
wide_nv2 <- nevadaPrimaries %>%
  pivot_wider(
    id_cols = candidate_name,
    names_from = start_date,
    values_from = pct)
wide_nv2
```

```
## # A tibble: 6 x 16
##   candidate_name 2020-~1 2020-~2 2020-~3 2019-~4 2019-~5 2019-~6 2019-~7 2019-~8
##   <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 Joseph R. Bid~    19.4      28      23      24      33    29.9    29.1      22
## 2 Bernard Sande~    17.6      29      17      18      23    18.8    19.1      22
## 3 Elizabeth War~    10.6      14      12      18      21    22.2    19.2      18
## 4 Pete Buttigieg     8.2       6       6       8       9     4.9     7.3       4
## 5 Amy Klobuchar      3.6       4       2       2       2     0.7     2.5       1
## 6 Michael Bloom~    NA        NA       2      NA      NA    NA      NA        NA
## # ... with 7 more variables: `2019-09-19` <dbl>, `2019-08-28` <dbl>,
## #   `2019-08-14` <dbl>, `2019-08-02` <dbl>, `2019-06-06` <dbl>,
## #   `2019-05-09` <dbl>, `2019-03-28` <dbl>, and abbreviated variable names
## #   1: `2020-01-08`, 2: `2020-01-06`, 3: `2020-01-05`, 4: `2019-11-10`,
## #   5: `2019-11-06`, 6: `2019-10-31`, 7: `2019-10-28`, 8: `2019-09-22`
```

```r
# So far in class we've selected two dates to demonstrate
# the wide to long pivot
long_nv <- wide_nv2 %>%
  select(candidate_name, "2020-01-08", "2020-01-06") %>%
  pivot_longer(
    cols = c("2020-01-08", "2020-01-06"),
    names_to = "start_date",
```

```
    values_to = "pct")
long_nv
```

```
## # A tibble: 12 x 3
##    candidate_name      start_date   pct
##    <chr>               <chr>      <dbl>
##  1 Joseph R. Biden Jr. 2020-01-08  19.4
##  2 Joseph R. Biden Jr. 2020-01-06  28
##  3 Bernard Sanders     2020-01-08  17.6
##  4 Bernard Sanders     2020-01-06  29
##  5 Elizabeth Warren    2020-01-08  10.6
##  6 Elizabeth Warren    2020-01-06  14
##  7 Pete Buttigieg      2020-01-08   8.2
##  8 Pete Buttigieg      2020-01-06   6
##  9 Amy Klobuchar       2020-01-08   3.6
## 10 Amy Klobuchar       2020-01-06   4
## 11 Michael Bloomberg   2020-01-08  NA
## 12 Michael Bloomberg   2020-01-06  NA
```

Erin answer:

```
wide_nv2 %>%
  # ALL dates, so comment out this
  #select(candidate_name, `2020-01-08`:`2018-03-28`) %>%
  pivot_longer(
    cols = "2020-01-08":"2019-03-28",
    names_to = "start_date",
    values_to = "pct") %>%
  group_by(start_date) %>%
  summarise(leader = candidate_name[which.max(pct)])
```

```
## # A tibble: 15 x 2
##    start_date leader
##    <chr>      <chr>
##  1 2019-03-28 Joseph R. Biden Jr.
##  2 2019-05-09 Joseph R. Biden Jr.
##  3 2019-06-06 Joseph R. Biden Jr.
##  4 2019-08-02 Joseph R. Biden Jr.
##  5 2019-08-14 Joseph R. Biden Jr.
##  6 2019-08-28 Bernard Sanders
##  7 2019-09-19 Joseph R. Biden Jr.
##  8 2019-09-22 Joseph R. Biden Jr.
##  9 2019-10-28 Joseph R. Biden Jr.
## 10 2019-10-31 Joseph R. Biden Jr.
## 11 2019-11-06 Joseph R. Biden Jr.
## 12 2019-11-10 Joseph R. Biden Jr.
## 13 2020-01-05 Joseph R. Biden Jr.
## 14 2020-01-06 Bernard Sanders
## 15 2020-01-08 Joseph R. Biden Jr.
```