

```
In [5]: install.packages("dslabs")
library(dslabs)
data(polls_us_election_2016)

package 'dslabs' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\hp\AppData\Local\Temp\RtmpIFymHX\downloaded_packages

Warning message:
"package 'dslabs' was built under R version 3.6.3"
```

```
In [6]: polls <- polls_us_election_2016
```

```
In [7]: str(polls)

'data.frame':  4208 obs. of  15 variables:
 $ state      : Factor w/ 57 levels "Alabama","Alaska",...: 50 50 50 50 50 50 50 37 50 ...
 $ startdate  : Date, format: "2016-11-03" "2016-11-01" ...
 $ enddate    : Date, format: "2016-11-06" "2016-11-07" ...
 $ pollster   : Factor w/ 196 levels "ABC News/Washington Post",...: 1 63 81 194 65 55 18 113 195 76 ...
 $ grade      : Factor w/ 10 levels "D","C-","C","C+",...: 10 6 8 6 5 9 8 8 NA 8 ...
 $ samplesize : int  2220 26574 2195 3677 16639 1295 1426 1282 8439 1107 ...
 $ population : chr  "lv" "lv" "lv" "lv" ...
 $ rawpoll_clinton : num  47 38 42 45 47 ...
 $ rawpoll_trump   : num  43 35.7 39 41 43 ...
 $ rawpoll_johnson : num  4 5.46 6 5 3 3 5 6 6 7.1 ...
 $ rawpoll_mcmullin: num  NA NA NA NA NA NA NA NA NA ...
 $ adjpoll_clinton : num  45.2 43.3 42 45.7 46.8 ...
 $ adjpoll_trump   : num  41.7 41.2 38.8 40.9 42.3 ...
 $ adjpoll_johnson : num  4.63 5.18 6.84 6.07 3.73 ...
 $ adjpoll_mcmullin: num  NA NA NA NA NA NA NA NA NA ...
```

In [8]: `summary(polls)`

```

      state      startdate      enddate
U.S.      :1106   Min.    :2015-11-06   Min.    :2015-11-08
Florida   : 148   1st Qu.:2016-08-10   1st Qu.:2016-08-21
North Carolina: 125 Median :2016-09-23   Median :2016-09-30
Pennsylvania : 125 Mean   :2016-08-31   Mean   :2016-09-06
Ohio      : 115   3rd Qu.:2016-10-20   3rd Qu.:2016-10-28
New Hampshire : 112 Max.    :2016-11-06   Max.    :2016-11-07
(Other)    :2477

      pollster      grade
Ipsos      : 919   A-      :1085
Google Consumer Surveys : 743   B      :1011
SurveyMonkey : 660   C-      : 693
YouGov      : 130   C+      : 329
Rasmussen Reports/Pulse Opinion Research: 125   B+      : 204
USC Dornsife/LA Times   : 121   (Other): 457
(Other)      :1510   NA's    : 429

      samplesize      population      rawpoll_clinton rawpoll_trump
Min.    : 35.0      Length:4208      Min.    :11.04      Min.    : 4.00
1st Qu.: 447.5      Class :character 1st Qu.:38.00      1st Qu.:35.00
Median : 772.0      Mode  :character Median :43.00      Median :40.00
Mean   : 1148.2                      Mean   :41.99      Mean   :39.83
3rd Qu.: 1236.5                      3rd Qu.:46.20      3rd Qu.:45.00
Max.   :84292.0                      Max.   :88.00      Max.   :68.00
NA's    :1

      rawpoll_johnson rawpoll_mcmullin adjpoll_clinton adjpoll_trump
Min.    : 0.000      Min.    : 9.0      Min.    :17.06      Min.    : 4.373
1st Qu.: 5.400      1st Qu.:22.5      1st Qu.:40.21      1st Qu.:38.429
Median : 7.000      Median :25.0      Median :44.15      Median :42.765
Mean   : 7.382      Mean   :24.0      Mean   :43.32      Mean   :42.674
3rd Qu.: 9.000      3rd Qu.:27.9      3rd Qu.:46.92      3rd Qu.:46.290
Max.   :25.000      Max.   :31.0      Max.   :86.77      Max.   :72.433
NA's    :1409      NA's    :4178

      adjpoll_johnson adjpoll_mcmullin
Min.    :-3.668      Min.    :11.03
1st Qu.: 3.145      1st Qu.:23.11
Median : 4.384      Median :25.14
Mean   : 4.660      Mean   :24.51
3rd Qu.: 5.756      3rd Qu.:27.98
Max.   :20.367      Max.   :31.57
NA's    :1409      NA's    :4178

```

In []:

In [9]: `req_details <- polls[,c('adjpoll_trump', 'adjpoll_clinton', 'pollster', 'grade', 'state')]`

```
In [11]: req_details
```

adjpoll_trump	adjpoll_clinton	pollster	grade	state
41.72430	45.20163	ABC News/Washington Post	A+	U.S.
41.21439	43.34557	Google Consumer Surveys	B	U.S.
38.81620	42.02638	Ipsos	A-	U.S.
40.92004	45.65676	YouGov	B	U.S.
42.33184	46.84089	Gravis Marketing	B-	U.S.
43.95631	49.02208	Fox News/Anderson Robbins Research/Shaw & Company Research	A	U.S.
40.92722	45.11649	CBS News/New York Times	A-	U.S.
40.77325	43.58576	NBC News/Wall Street Journal	A-	U.S.
41.59978	44.82594	Zia Poll	NA	New Mexico
42.23545	42.92745	IBD/TIPP	A-	U.S.
40.57082	44.21714	Selzer & Company	A+	U.S.
43.68125	47.57171	Angus Reid Global	A-	U.S.
43.39600	48.86765	Monmouth University	A+	U.S.
42.32751	47.43805	Public Policy Polling	B+	Virginia
43.43819	42.83406	Marist College	A	U.S.
45.66917	39.37561	Selzer & Company	A+	Iowa
42.26495	45.13966	The Times-Picayune/Lucid	NA	U.S.
40.97982	46.10344	Marquette University	A	Wisconsin
45.08290	44.21875	Siena College	A	North Carolina
48.80363	45.06470	Landmark Communications	B	Georgia
43.93999	46.44315	Quinnipiac University	A-	Florida
43.93745	47.43742	Quinnipiac University	A-	North Carolina
41.58078	46.39152	ABC News/Washington Post	A+	Virginia
43.38579	45.32156	USC Dornsife/LA Times	NA	U.S.
47.32734	43.83075	Gravis Marketing	B-	Georgia
39.33472	43.84735	Gravis Marketing	B-	Oregon
43.16080	47.63973	RKM Research and Communications, Inc.	B+	U.S.
47.33178	41.84073	Gravis Marketing	B-	Ohio
47.32748	42.83107	Gravis Marketing	B-	South Carolina
45.09646	44.93624	YouGov	B	Ohio
...
46.49556	41.61857	Public Policy Polling	B+	North Carolina
45.90849	29.84414	Dan Jones & Associates	C+	Utah
45.50512	42.62294	Public Policy Polling	B+	Florida
48.66301	41.99976	Ipsos	A-	Georgia
30.24046	32.04723	Google Consumer Surveys	B	Utah
33.95362	58.17926	SurveyUSA	A	California
42.46722	43.71471	Quinnipiac University	A-	Pennsylvania
36.32965	54.08580	Siena College	A	New York
48.77146	38.75657	SurveyUSA	A	North Carolina
49.13322	46.20387	Ipsos	A-	South Carolina
36.01090	33.61454	Google Consumer Surveys	B	Utah
35.91267	47.81869	Public Policy Polling	B+	New Hampshire
41.33270	46.44439	Quinnipiac University	A-	Florida
41.36491	42.46502	Quinnipiac University	A-	Ohio
38.83402	56.98850	Siena College	A	New York
44.86935	47.79437	Quinnipiac University	A-	North Carolina
44.40152	39.90517	SurveyUSA	A	North Carolina
47.12021	43.28262	Public Policy Polling	B+	North Carolina
54.36357	38.34430	Ipsos	A-	Kentucky
46.65680	45.19351	Quinnipiac University	A-	Florida
41.19061	43.35339	Quinnipiac University	A-	Pennsylvania

adjpoll_trump	adjpoll_clinton	pollster	grade	state
42.33380	40.73937	Quinnipiac University	A-	Ohio
43.55006	42.13165	Public Policy Polling	B+	North Carolina
45.43384	26.57791	Ipsos	A-	South Dakota
38.22545	45.56387	Ipsos	A-	Washington
40.48017	46.47852	Ipsos	A-	Virginia
39.07778	48.74781	Marquette University	A	Wisconsin
40.57738	26.65200	Google Consumer Surveys	B	Utah
37.10720	45.12949	Ipsos	A-	Oregon
42.05508	42.14966	EPIC-MRA	A-	Michigan

In [12]: `str(req_details)`

```
'data.frame':  4208 obs. of  5 variables:
 $ adjpoll_trump : num  41.7 41.2 38.8 40.9 42.3 ...
 $ adjpoll_clinton: num  45.2 43.3 42 45.7 46.8 ...
 $ pollster      : Factor w/ 196 levels "ABC News/Washington Post",...: 1 63 81 194 65 55 18 113 195 76 ...
 $ grade        : Factor w/ 10 levels "D","C-","C","C+",...: 10 6 8 6 5 9 8 8 NA 8 ...
 $ state        : Factor w/ 57 levels "Alabama","Alaska",...: 50 50 50 50 50 50 50 50 37 50 ...
```

In [13]: `is.factor(req_details$state)`
`is.factor(req_details$pollster)`
`is.factor(req_details$grade)`
`is.factor(req_details$adjpoll_clinton)`
`is.factor(req_details$adjpoll_trump)`

TRUE

TRUE

TRUE

FALSE

FALSE

In [14]: `levels(req_details$state)`

```
'Alabama'  'Alaska'  'Arizona'  'Arkansas'  'California'  'Colorado'  'Connecticut'  'Delaware'  'District of Columbia'  'Florida'
'Georgia'  'Hawaii'  'Idaho'    'Illinois'  'Indiana'    'Iowa'      'Kansas'      'Kentucky'  'Louisiana'  'Maine'      'Maine CD-1'  'Maine CD-2'
'Maryland'  'Massachusetts'  'Michigan'  'Minnesota'  'Mississippi'  'Missouri'  'Montana'    'Nebraska'  'Nebraska CD-1'
'Nebraska CD-2'  'Nebraska CD-3'  'Nevada'   'New Hampshire'  'New Jersey'  'New Mexico'  'New York'   'North Carolina'  'North Dakota'
'Ohio'        'Oklahoma'  'Oregon'   'Pennsylvania'  'Rhode Island'  'South Carolina'  'South Dakota'  'Tennessee'  'Texas'      'U.S.'
'Utah'        'Vermont'   'Virginia'  'Washington'  'West Virginia'  'Wisconsin'  'Wyoming'
```

In [15]: `levels(req_details$pollster)`

```
'ABC News/Washington Post' 'American Research Group' 'American Strategies' 'Angus Reid Global' 'Anzalone Liszt Grove Research'
'Arizona State University' 'Associated Industries of Florida' 'Baldwin Wallace University' 'Ball State University' 'Baruch College'
'Basswood Research' 'Behavior Research Center (Rocky Mountain)' 'Bellwether Research & Consulting' 'Bendixen & Amandi International'
'BK Strategies' 'Braun Research' 'Castleton University' 'CBS News/New York Times' 'Centre College' 'Cherry Communications'
'Christopher Newport University' 'Clarity Campaign Labs' 'Clout Research/Wenzel Strategies' 'CNN/Opinion Research Corp.'
'Cole Hargrave Snodgrass & Associates' 'Columbus Dispatch' 'Craciun Research' 'Critical Insights'
'Crosswind Media & Public Relations' 'CVOTER International' 'Cyngal Political' 'Dan Jones & Associates' 'Dartmouth College'
'Data Orbital' 'Data Targeting' 'Davis, Hibbitts & Midghall, Inc.' 'DFM Research' 'Dixie Strategies' 'Echelon Insights'
'Edison Research' 'Elon University' 'Elway Research' 'Emerson College' 'EPIC-MRA' 'Evolving Strategies' 'Expedition Strategies'
'Fairbank, Maslin, Maullin, Metz & Associates' 'Fairleigh Dickinson University (PublicMind)' 'Feldman Group'
'Field Research Corporation (Field Poll)' 'First Tuesday Strategies' 'Florida Atlantic University' 'Florida Southern College'
'Fort Hays State University' 'Fox News/Anderson Robbins Research/Shaw & Company Research' 'Franklin & Marshall College'
'Garin-Hart-Yang Research Group' 'GBA Strategies' 'George Washington University (Battleground)' 'GfK Group'
'Glengariff Group, Inc.' 'Global Strategy Group' 'Google Consumer Surveys' 'Goucher College' 'Gravis Marketing'
'Greenberg Quinlan Rosner (Democracy Corps)' 'Greenberg Quinlan Rosner/American Viewpoint' 'Hampton University' 'Harper Polling'
'Harvard University' 'Hendrix College' 'Hickman Analytics' 'High Point University' 'HighGround' 'Hoffman Research Group'
'IBD/TIPP' 'icitizen' 'Illinois Public Opinion Strategies' 'Insights West' 'Integrated Web Strategy' 'Ipsos'
'Ivan Moore Research' 'JMC Enterprises' 'Just Win Strategies' 'Keating Research, Inc.' 'Lake Research Partners'
'Landmark Communications' 'Leland Beatty' 'Loras College' 'Magellan Strategies' 'Maine People\'s Resource Center' 'Marist College'
'Market Research Insight' 'Marketing Resource Group (MRG)' 'Marquette University' 'Mason-Dixon Polling & Research, Inc.'
'MassINC Polling Group' 'MBQF' 'McLaughlin & Associates' 'Mercyhurst University' 'Meredith College' 'Merrill Poll'
'Michigan State University' 'Middle Tennessee State University' 'Mitchell Research & Communications' 'Monmouth University'
'Montana State University Billings' 'Moore Information' 'Morning Consult' 'Muhlenberg College' 'National Journal'
'National Research, Inc.' 'NBC News/Wall Street Journal' 'Nielson Brothers Polling' 'Normington, Petts & Associates'
'North Star Opinion Research/Ayres, McHenry & Associates' 'Ogden & Fry' 'OnMessage Inc.' 'Opinion Savvy/InsiderAdvantage'
'OpinionWorks' 'Orion Strategies' 'Penn Schoen Berland' 'Pew Research Center' 'Political Callers, Inc.'
'Political Marketing International, Inc./Red Racing Horses' 'Praecones Analytica' 'Princeton Survey Research Associates International'
'Public Opinion Strategies' 'Public Policy Institute of California' 'Public Policy Polling' 'Public Religion Research Institute'
'Quinnipiac University' 'R.L. Repass & Partners' 'RABA Research' 'RAND (American Life Panel)'
'Rasmussen Reports/Pulse Opinion Research' 'Remington' 'Research & Polling, Inc.' 'Riley Research Associates'
'RKM Research and Communications, Inc.' 'Roanoke College' 'Robert Morris University' 'Rutgers University' 'Sacramento State'
'Saguaro Strategies' 'Saint Leo University' 'Schoen Consulting' 'Selzer & Company' 'Siena College' 'SoonerPoll.com'
'Southern Illinois University' 'Southern Media & Opinion Research' 'St. Norbert College' 'Starboard Communications'
'Stockton College' 'Strategic National' 'Strategies 360' 'Strategy Research' 'Suffolk University' 'SurveyMonkey' 'SurveyUSA'
'Susquehanna Polling & Research, Inc.' 'Target Insyght' 'TargetPoint' 'TargetSmart/William & Mary' 'Tarrance Group' 'Texas Lyceum'
'The Times-Picayune/Lucid' 'Time/Abt SRBI' 'Trafalgar Group' 'Tulchin Research' 'University of Akron' 'University of Arkansas'
'University of Colorado' 'University of Delaware' 'University of Denver' 'University of Houston' 'University of Mary Washington'
'University of Massachusetts Lowell' 'University of New Hampshire' 'University of New Orleans' 'University of North Florida'
'University of Wyoming' 'USC Dornsife/LA Times' 'Utah Valley University' 'Vanderbilt University' 'Victory Research'
'Vox Populi Communications' 'We Ask America' 'Western Kentucky University' 'Western New England University' 'Winthrop University'
'Y2 Analytics' 'YouGov' 'Zia Poll' 'Zogby Interactive/JZ Analytics'
```

```
In [16]: levels(req_details$grade)
```

```
'D'  'C-'  'C'  'C+'  'B-'  'B'  'B+'  'A-'  'A'  'A+'
```

```
In [17]: sum(is.na(polls$state))
sum(is.na(polls$startdate))
sum(is.na(polls$enddate))
sum(is.na(polls$pollster))
sum(is.na(polls$grade))
sum(is.na(polls$samplesize))
sum(is.na(polls$population))
sum(is.na(polls$rawpoll_clinton))
sum(is.na(polls$rawpoll_trump))
sum(is.na(polls$rawpoll_johnson))
sum(is.na(polls$rawpoll_mcmullin))
sum(is.na(polls$adjpoll_clinton))
sum(is.na(polls$adjpoll_trump))
sum(is.na(polls$adjpoll_johnson))
sum(is.na(polls$adjpoll_mcmullin))
```

```
0
```

```
0
```

```
0
```

```
0
```

```
429
```

```
1
```

```
0
```

```
0
```

```
0
```

```
1409
```

```
4178
```

```
0
```

```
0
```

```
1409
```

```
4178
```

```
In [18]: sum(is.na(polls))
```

```
11604
```

```
In [26]: head(polls$star,10)
```

```
2016-11-03 2016-11-01 2016-11-02 2016-11-04 2016-11-03 2016-11-03 2016-11-02 2016-11-03 2016-11-06 2016-11-04
```

```
In [22]: sum(polls$adjpoll)
```

```
0
```

```
In [24]: mean(polls$adjpoll_trump)
```

```
42.673634934173
```

```
In [27]: columnname.miss[,1][is.na(column.miss[,1])] <- mean(columnname.miss[,1],na.rm = T)
```

```
Error in mean(columnname.miss[, 1], na.rm = T): object 'columnname.miss' not found  
Traceback:
```

```
1. mean(columnname.miss[, 1], na.rm = T)
```

```
In [30]:
```

```
NULL
```