

# Exercise 7: Data Manipulation

Marcy Shieh

10/22/2020

## Data manipulation exercises

Please submit the exercise as a R file and upload it to your `ps-exercises` Git repository.

```
# load packages
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(here)

## here() starts at /Users/marcyshieh/ps811
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.4      v stringr 1.4.0
## v tidyr   1.1.2      v forcats 0.5.0
## v readr   1.4.0
##
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
library(haven)

# folder locations
here()

## [1] "/Users/marcyshieh/ps811"
here("data")

## [1] "/Users/marcyshieh/ps811/data"
```

1. Create a dataframe in R based on the table below.

| Justice             | State | Position          | Replacing           | Year confirmed | Senate confirmation vote | Nominated by     |
|---------------------|-------|-------------------|---------------------|----------------|--------------------------|------------------|
| Clarence Thomas     | GA    | Associate Justice | Thurgood Marshall   | 1991           | 52-48                    | George H.W. Bush |
| Ruth Bader Ginsburg | NY    | Associate Justice | Byron White         | 1993           | 96-3                     | Bill Clinton     |
| Stephen Breyer      | MA    | Associate Justice | Harry Blackmun      | 1994           | 87-9                     | Bill Clinton     |
| John Roberts        | MD    | Chief Justice     | William Rehnquist   | 2005           | 78-22                    | George W. Bush   |
| Samuel Alito        | NJ    | Associate Justice | Sandra Day O'Connor | 2006           | 58-42                    | George W. Bush   |
| Sonia Sotomayor     | NY    | Associate Justice | David Souter        | 2009           | 68-31                    | Barack Obama     |
| Elena Kagan         | MA    | Associate Justice | John Paul Stevens   | 2010           | 63-37                    | Barack Obama     |
| Neil Gorsuch        | CO    | Associate Justice | Antonin Scalia      | 2017           | 54-45                    | Donald Trump     |
| Brett Kavanaugh     | MD    | Associate Justice | Anthony Kennedy     | 2018           | 50-48                    | Donald Trump     |

```
Justice <- c("Clarence Thomas", "Ruth Bader Ginsburg",
            "Stephen Breyer", "John Roberts", "Samuel Alito",
            "Sonia Sotomayor", "Elena Kagan", "Neil Gorsuch",
            "Brett Kavanaugh")

State <- c("GA", "NY", "MA", "MD", "NJ", "NY", "MA", "CO", "MD")

Position <- c("Associate Justice", "Associate Justice",
             "Associate Justice", "Chief Justice",
             "Associate Justice", "Associate Justice",
             "Associate Justice", "Associate Justice",
             "Associate Justice")

Replacing <- c("Thurgood Marshall", "Byron White",
              "Harry Blackmun", "William Rehnquist",
              "Sandra Day O'Connor", "David Souter",
              "John Paul Stevens", "Antonin Scalia",
              "Anthony Kennedy")

Year <- c(1991, 1993, 1994, 2005, 2006, 2009, 2010, 2017, 2018)

`Senate Confirmation Vote` <- c("52-48", "96-3", "87-9", "78-22",
                               "58-42", "68-31", "63-37",
                               "54-45", "50-48")

`Nominated by` <- c("George H.W. Bush", "Bill Clinton",
                    "Bill Clinton", "George W. Bush",
                    "George W. Bush", "Barack Obama",
                    "Barack Obama", "Donald Trump",
                    "Donald Trump")

SCJustices <- data.frame(Justice, State, Position, Replacing,
```

```

Year, `Senate Confirmation Vote`,
`Nominated by`)
SCJustices

```

| ##   | Justice             | State | Position          | Replacing           | Year |
|------|---------------------|-------|-------------------|---------------------|------|
| ## 1 | Clarence Thomas     | GA    | Associate Justice | Thurgood Marshall   | 1991 |
| ## 2 | Ruth Bader Ginsburg | NY    | Associate Justice | Byron White         | 1993 |
| ## 3 | Stephen Breyer      | MA    | Associate Justice | Harry Blackmun      | 1994 |
| ## 4 | John Roberts        | MD    | Chief Justice     | William Rehnquist   | 2005 |
| ## 5 | Samuel Alito        | NJ    | Associate Justice | Sandra Day O'Connor | 2006 |
| ## 6 | Sonia Sotomayor     | NY    | Associate Justice | David Souter        | 2009 |
| ## 7 | Elena Kagan         | MA    | Associate Justice | John Paul Stevens   | 2010 |
| ## 8 | Neil Gorsuch        | CO    | Associate Justice | Antonin Scalia      | 2017 |
| ## 9 | Brett Kavanaugh     | MD    | Associate Justice | Anthony Kennedy     | 2018 |

| ##   | Senate.Confirmation.Vote | Nominated.by     |
|------|--------------------------|------------------|
| ## 1 | 52-48                    | George H.W. Bush |
| ## 2 | 96-3                     | Bill Clinton     |
| ## 3 | 87-9                     | Bill Clinton     |
| ## 4 | 78-22                    | George W. Bush   |
| ## 5 | 58-42                    | George W. Bush   |
| ## 6 | 68-31                    | Barack Obama     |
| ## 7 | 63-37                    | Barack Obama     |
| ## 8 | 54-45                    | Donald Trump     |
| ## 9 | 50-48                    | Donald Trump     |

- Download `justices.csv` from the `ps811` GitHub repository. The `justices.csv` file contains Martin-Quinn scores (a measure of ideology) for justices from 1937 to 2019.

```

justices <- read.csv(here("data", "justices.csv"))

```

For the data manipulation questions below, use pipes (`%>%`).

- Merge the `justices.csv` and `SCDB_2020_01_justiceCentered_Citation.dta` datasets using one of the join functions. Before performing the merge, check that the variable names you want to merge are the same in both datasets.
  - If the names are different, you will need to rename the variable names for one of the datasets so you can merge the two datasets. Make sure the values in the variable names that you would like to merge are formatted the same way.
  - For example, if you want to merge every Roberts vote with his Martin-Quinn score in that particular term, you will need to make sure that both datasets format Roberts' name correctly. An easy way to do this is to put the justice variable in a `table()`.

```

# load justice votes data
scotus <- read_dta(here("data", "SCDB_2020_01_justiceCentered_Citation.dta"))

# check variable names to see if they can be merged
names(scotus)

```

|         |                |                |
|---------|----------------|----------------|
| ## [1]  | "caseId"       | "docketId"     |
| ## [3]  | "caseIssuesId" | "voteId"       |
| ## [5]  | "dateDecision" | "decisionType" |
| ## [7]  | "usCite"       | "sctCite"      |
| ## [9]  | "ledCite"      | "lexisCite"    |
| ## [11] | "term"         | "naturalCourt" |
| ## [13] | "chief"        | "docket"       |
| ## [15] | "caseName"     | "dateArgument" |

```
## [17] "dateRearg"           "petitioner"
## [19] "petitionerState"     "respondent"
## [21] "respondentState"     "jurisdiction"
## [23] "adminAction"         "adminActionState"
## [25] "threeJudgeFdc"       "caseOrigin"
## [27] "caseOriginState"     "caseSource"
## [29] "caseSourceState"     "lcDisagreement"
## [31] "certReason"          "lcDisposition"
## [33] "lcDispositionDirection" "declarationUncon"
## [35] "caseDisposition"     "caseDispositionUnusual"
## [37] "partyWinning"        "precedentAlteration"
## [39] "voteUnclear"         "issue"
## [41] "issueArea"           "decisionDirection"
## [43] "decisionDirectionDissent" "authorityDecision1"
## [45] "authorityDecision2"  "lawType"
## [47] "lawSupp"             "lawMinor"
## [49] "majOpinWriter"       "majOpinAssigner"
## [51] "splitVote"           "majVotes"
## [53] "minVotes"            "justice"
## [55] "justiceName"         "vote"
## [57] "opinion"             "direction"
## [59] "majority"            "firstAgreement"
## [61] "secondAgreement"
```

```
names(justices)
```

```
## [1] "term"           "justice"       "justiceName" "post_mn"       "post_sd"
## [6] "post_med"      "post_025"      "post_975"
```

```
# check values
```

```
joined_justices <- full_join(scotus, justices, by=c("justiceName", "term"))
```

```
# check for repeats
```

```
table(joined_justices$justiceName)
```

```
##
##      AFortas  AJGoldberg  AMKennedy  AScalia  BMKavanaugh  BNCardozo
##      581      475      2883      2861      133      1
##      BRWhite  CEHughes2  CEWhittaker  CThomas  DHSouter  EKagan
##      4946      4      691      2467      1753      749
##      EWarren  FFrankfurter  FMurphy  FMVinson  GSutherland  HABlackmun
##      2205      1925      394      812      1      3771
##      HFStone  HHBurton  HLBlack  JCMcReynolds  JFByrnes  JGRoberts
##      9      1389      3311      4      1      1160
##      JHarlan2  JPStevens  LDBrandeis  LFPowell  NMGorsuch  OJRoberts
##      2351      4268      2      2652      253      8
##      PButler  PStewart  RBGinsburg  RHJackson  SAAlito  SDOConnor
##      2      3592      2229      904      1134      2914
##      SFreed  SGBreyer  SMinton  SSotomayor  TCClark  TMarshall
##      1160      2131      717      841      2292      3881
##      WRutledge  WEBurger  WHRehnquist  WJBrennan  WODouglas
##      391      2809      4535      5327      4009
```

4. Filter to justices with Martin-Quinn scores.

```
joined_justices_MQ <- filter(joined_justices, !is.na(post_mn))
```

5. Find the mean Martin-Quinn score for each term in your dataset.

```
MQ_by_term <- joined_justices_MQ %>%
  group_by(term) %>%
  summarise(mean = mean(post_mn, na.rm = TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
print(MQ_by_term, n = nrow(MQ_by_term))
```

```
## # A tibble: 83 x 2
```

```
##   term      mean
```

```
##   <dbl>   <dbl>
```

```
## 1  1937 -0.0135
```

```
## 2  1938 -0.368
```

```
## 3  1939 -0.663
```

```
## 4  1940 -0.420
```

```
## 5  1941 -0.635
```

```
## 6  1942 -0.588
```

```
## 7  1943 -0.498
```

```
## 8  1944 -0.373
```

```
## 9  1945 -0.430
```

```
## 10 1946 -0.360
```

```
## 11 1947 -0.296
```

```
## 12 1948 -0.239
```

```
## 13 1949  0.408
```

```
## 14 1950  0.459
```

```
## 15 1951  0.476
```

```
## 16 1952  0.467
```

```
## 17 1953  0.130
```

```
## 18 1954 -0.136
```

```
## 19 1955 -0.261
```

```
## 20 1956 -0.507
```

```
## 21 1957 -0.510
```

```
## 22 1958 -0.488
```

```
## 23 1959 -0.508
```

```
## 24 1960 -0.463
```

```
## 25 1961 -0.603
```

```
## 26 1962 -1.03
```

```
## 27 1963 -1.05
```

```
## 28 1964 -0.956
```

```
## 29 1965 -0.985
```

```
## 30 1966 -0.982
```

```
## 31 1967 -1.22
```

```
## 32 1968 -1.21
```

```
## 33 1969 -0.706
```

```
## 34 1970 -0.473
```

```
## 35 1971 -0.138
```

```
## 36 1972 -0.0721
```

```
## 37 1973 -0.115
```

```
## 38 1974 -0.194
```

```
## 39 1975  0.504
```

```
## 40 1976  0.369
```

```
## 41 1977  0.182
```

```
## 42 1978  0.155
```

```
## 43 1979  0.133
```

```
## 44 1980 0.175
## 45 1981 0.273
## 46 1982 0.302
## 47 1983 0.316
## 48 1984 0.264
## 49 1985 0.187
## 50 1986 -0.0168
## 51 1987 -0.0596
## 52 1988 -0.0382
## 53 1989 -0.123
## 54 1990 0.138
## 55 1991 0.726
## 56 1992 0.707
## 57 1993 0.543
## 58 1994 0.664
## 59 1995 0.634
## 60 1996 0.629
## 61 1997 0.620
## 62 1998 0.634
## 63 1999 0.530
## 64 2000 0.391
## 65 2001 0.295
## 66 2002 0.221
## 67 2003 0.189
## 68 2004 0.171
## 69 2005 0.263
## 70 2006 0.293
## 71 2007 0.283
## 72 2008 0.283
## 73 2009 0.241
## 74 2010 0.323
## 75 2011 0.214
## 76 2012 0.0723
## 77 2013 -0.0330
## 78 2014 -0.169
## 79 2015 -0.387
## 80 2016 -0.316
## 81 2017 -0.246
## 82 2018 -0.260
## 83 2019 -0.284
```

6. Find the mean decision direction for each term in your dataset. Rescale the decision direction variable so it is analogous to the Martin-Quinn score.

- Hint: the SCOTUS database decision direction scores are currently 1 (conservative), 2 (liberal), 3 (unspecified). You want to change it to -1 (liberal), 0 (unspecified), and 1 (conservative).

```
joined_justices_MQ <- mutate(joined_justices_MQ,
  decisionDirection = case_when(
    decisionDirection == 1 ~ 1,
    decisionDirection == 2 ~ -1,
    decisionDirection == 3 ~ 0))

decision_by_term <- joined_justices_MQ %>%
```

```

group_by(term) %>%
  summarise(mean = mean(decisionDirection, na.rm = TRUE))

## `summarise()` ungrouping output (override with `.groups` argument)
print(decision_by_term, n = nrow(decision_by_term))

## # A tibble: 83 x 2
##   term      mean
##   <dbl>    <dbl>
## 1 1937    NaN
## 2 1938    NaN
## 3 1939    NaN
## 4 1940    NaN
## 5 1941    NaN
## 6 1942    NaN
## 7 1943    NaN
## 8 1944    NaN
## 9 1945    NaN
## 10 1946  -0.0780
## 11 1947  -0.248
## 12 1948  -0.115
## 13 1949   0.131
## 14 1950   0.0909
## 15 1951   0.146
## 16 1952   0.121
## 17 1953   0.0115
## 18 1954  -0.379
## 19 1955  -0.327
## 20 1956  -0.355
## 21 1957  -0.2
## 22 1958  -0.239
## 23 1959  -0.243
## 24 1960  -0.151
## 25 1961  -0.401
## 26 1962  -0.535
## 27 1963  -0.552
## 28 1964  -0.350
## 29 1965  -0.270
## 30 1966  -0.315
## 31 1967  -0.497
## 32 1968  -0.377
## 33 1969  -0.0828
## 34 1970   0.0486
## 35 1971  -0.0822
## 36 1972   0.0860
## 37 1973   0.121
## 38 1974  -0.00649
## 39 1975   0.241
## 40 1976   0.190
## 41 1977  -0.0129
## 42 1978   0.206
## 43 1979  -0.0256
## 44 1980   0.184

```

```
## 45 1981 0.0508
## 46 1982 0.133
## 47 1983 0.220
## 48 1984 0.109
## 49 1985 0.182
## 50 1986 0.0745
## 51 1987 0.02
## 52 1988 0.171
## 53 1989 0.0857
## 54 1990 -0.0224
## 55 1991 0.0649
## 56 1992 0.0664
## 57 1993 0.125
## 58 1994 0.185
## 59 1995 0.0769
## 60 1996 0.271
## 61 1997 0.19
## 62 1998 0.283
## 63 1999 0.0105
## 64 2000 -0.0102
## 65 2001 0.202
## 66 2002 0.167
## 67 2003 0.149
## 68 2004 -0.0875
## 69 2005 0.184
## 70 2006 0.174
## 71 2007 -0.0548
## 72 2008 0.229
## 73 2009 -0.00120
## 74 2010 0.119
## 75 2011 0.130
## 76 2012 0.0127
## 77 2013 -0.107
## 78 2014 -0.122
## 79 2015 0.0152
## 80 2016 -0.0336
## 81 2017 0.0267
## 82 2018 -0.0274
## 83 2019 -0.0667
```

7. Compare the mean Martin-Quinn scores and vote directions. Are they similar or are they different?

```
mq_decision_compare <- inner_join(MQ_by_term, decision_by_term,
                                   by="term")
```

```
mq_decision_compare_diff <- mq_decision_compare %>%
  group_by(term) %>%
  summarise(difference = mean.y - mean.x)
```

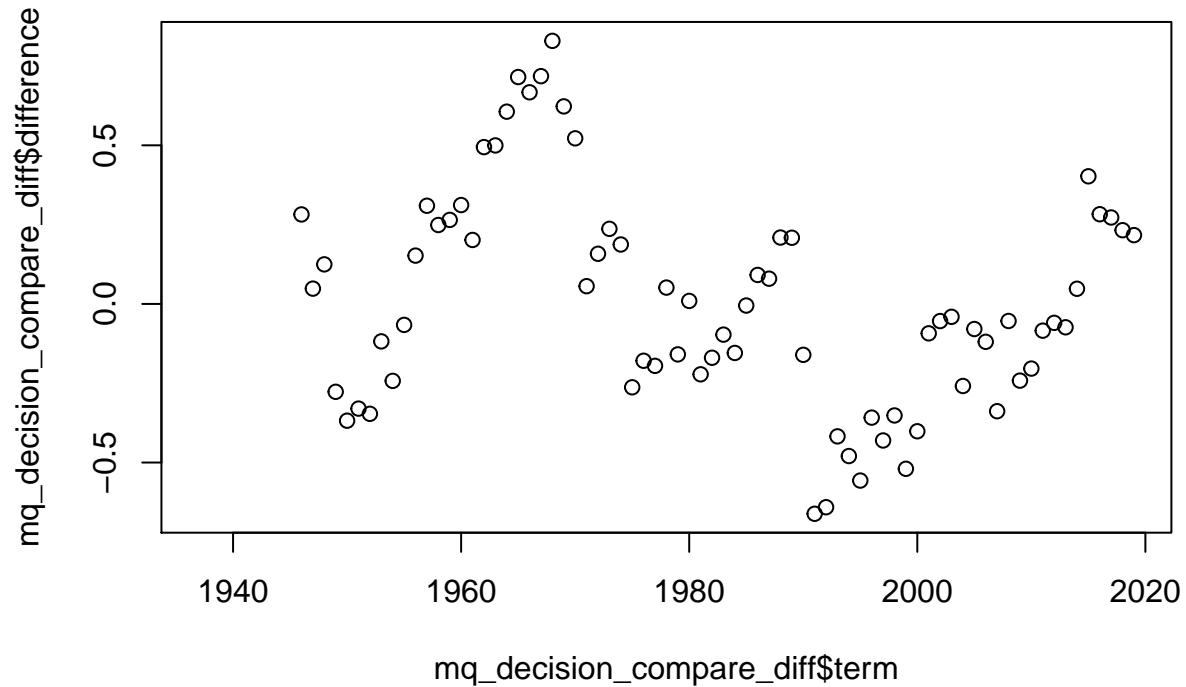
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# average difference is small...you can see this from looking at the mean or from a plot
mean(mq_decision_compare_diff$difference, na.rm = TRUE)
```

```
## [1] 0.006625254
```



```
plot(mq_decision_compare_diff$term, mq_decision_compare_diff$difference)
```



## Brainstorm final project

Please submit the following questions in an R Markdown document. (It does not need to be in the `papaja` template.)

1. What question(s) are you interested in?
2. What are your independent and dependent variables?
3. How do you plan to measure the variables?
4. What data will you need to collect? Which dataset(s) will you use?
5. What methods will you use to analyze the data?

## Submit

Email me ([mshieh2@wisc.edu](mailto:mshieh2@wisc.edu)) the link to your `ps811-exercises` repository when you are done.