bivariate relationships

Prabidhik KC

2022-10-22

```
# Loading necessary libraries

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gov50data)

covid_votes <- read_csv("covid_votes.csv")
```

```
## Rows: 3114 Columns: 8

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): fips, county, state
## dbl (5): one_dose_5plus_pct, one_dose_65plus_pct, booster_5plus_pct, dem_pct...
##
## 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.
```

```
covid_votes
```

```
## # A tibble: 3,114 x 8
##    fips  county           state one_dose_5plus~1 one_d~2 boost~3 dem_p~4 dem_p~5
##    <chr> <chr>            <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 26039 Crawford County  MI                55.7    77.3   31.2     43.8    34.0
##  2 40015 Caddo County     OK                83.3    95     30.3     46.4    27.1
##  3 17007 Boone County     IL                71.1    94.5   35.1     41.8    42.2
##  4 12055 Highlands County FL                68.9    93.7   24.7     40.3    32.5
##  5 34029 Ocean County     NJ                71      95     32.1     47.2    35.0
```

```
##  6 01067 Henry County     AL                58.5    85.5   18.2     40.1    28.0
##  7 27037 Dakota County    MN                81      95     49.5     46.9    55.7
##  8 27115 Pine County      MN                56.5    85     31.7     47.0    33.9
##  9 51750 Radford city     VA                41.5    73.8    1.79    46.4    53.1
## 10 22009 Avoyelles Parish LA                59.7    80.1   21.9     45.7    28.8
## # ... with 3,104 more rows, and abbreviated variable names
## #   1: one_dose_5plus_pct, 2: one_dose_65plus_pct, 3: booster_5plus_pct,
## #   4: dem_pct_2000, 5: dem_pct_2020
```

```
covid_votes %>%
  mutate(
    one_dose_5p_centered = one_dose_5plus_pct - mean(one_dose_5plus_pct, na.rm = TRUE),
    one_dose_5p_z = one_dose_5p_centered/sd(one_dose_5plus_pct, na.rm = TRUE)
  ) %>%
  filter(one_dose_5p_z < 2)
```

```
## # A tibble: 2,837 x 10
##    fips  county    state one_d~1 one_d~2 boost~3 dem_p~4 dem_p~5 one_d~6 one_d~7
##    <chr> <chr>     <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 26039 Crawford~ MI       55.7    77.3   31.2     43.8    34.0   -7.35  -0.508
##  2 40015 Caddo Co~ OK       83.3    95     30.3     46.4    27.1   20.2    1.40
##  3 17007 Boone Co~ IL       71.1    94.5   35.1     41.8    42.2    8.05   0.556
##  4 12055 Highland~ FL       68.9    93.7   24.7     40.3    32.5    5.85   0.404
##  5 34029 Ocean Co~ NJ       71      95     32.1     47.2    35.0    7.95   0.549
##  6 01067 Henry Co~ AL       58.5    85.5   18.2     40.1    28.0   -4.55  -0.314
##  7 27037 Dakota C~ MN       81      95     49.5     46.9    55.7   17.9    1.24
##  8 27115 Pine Cou~ MN       56.5    85     31.7     47.0    33.9   -6.55  -0.452
##  9 51750 Radford ~ VA       41.5    73.8    1.79    46.4    53.1  -21.6   -1.49
## 10 22009 Avoyelle~ LA       59.7    80.1   21.9     45.7    28.8   -3.35  -0.231
## # ... with 2,827 more rows, and abbreviated variable names
## #   1: one_dose_5plus_pct, 2: one_dose_65plus_pct, 3: booster_5plus_pct,
## #   4: dem_pct_2000, 5: dem_pct_2020, 6: one_dose_5p_centered, 7: one_dose_5p_z
```

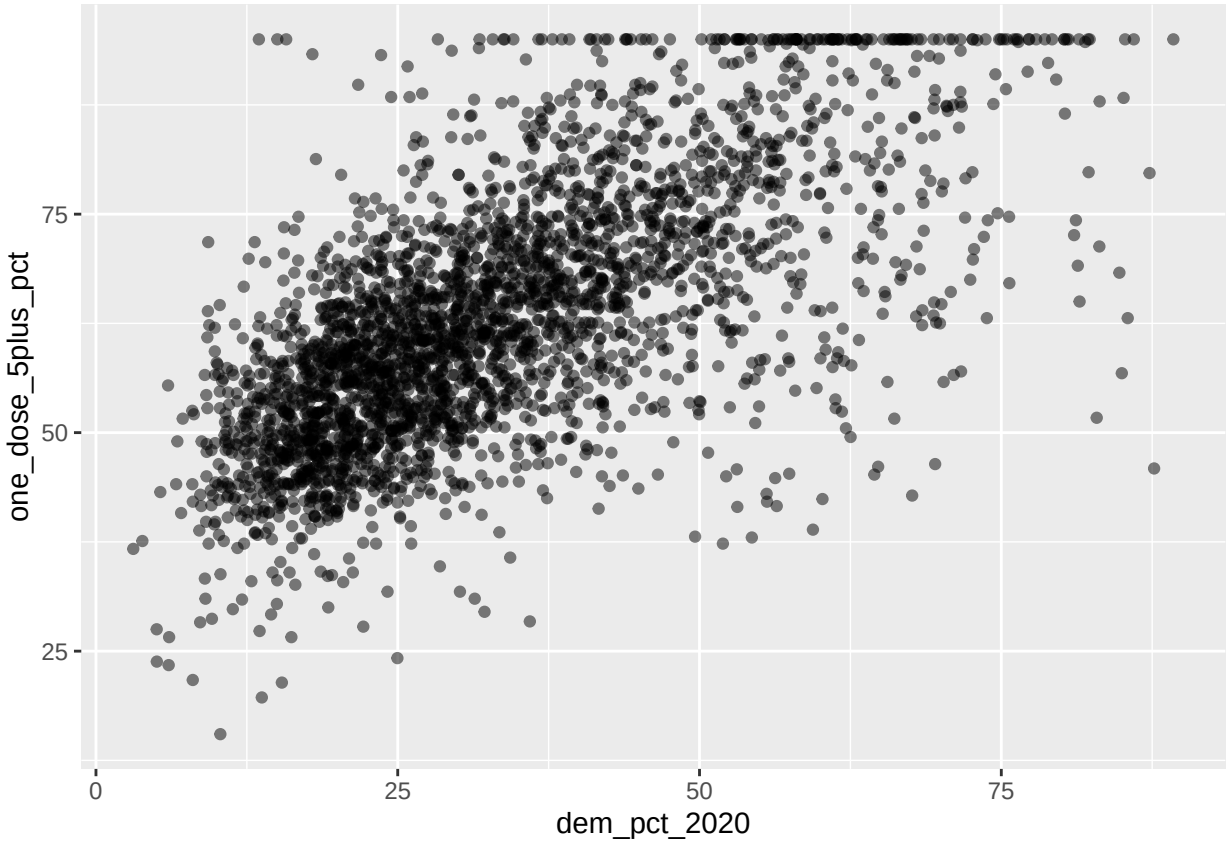
Correlation

```
cor(covid_votes$one_dose_5plus_pct, covid_votes$dem_pct_2020,
    use = "pairwise")
```

```
## [1] 0.6664387
```

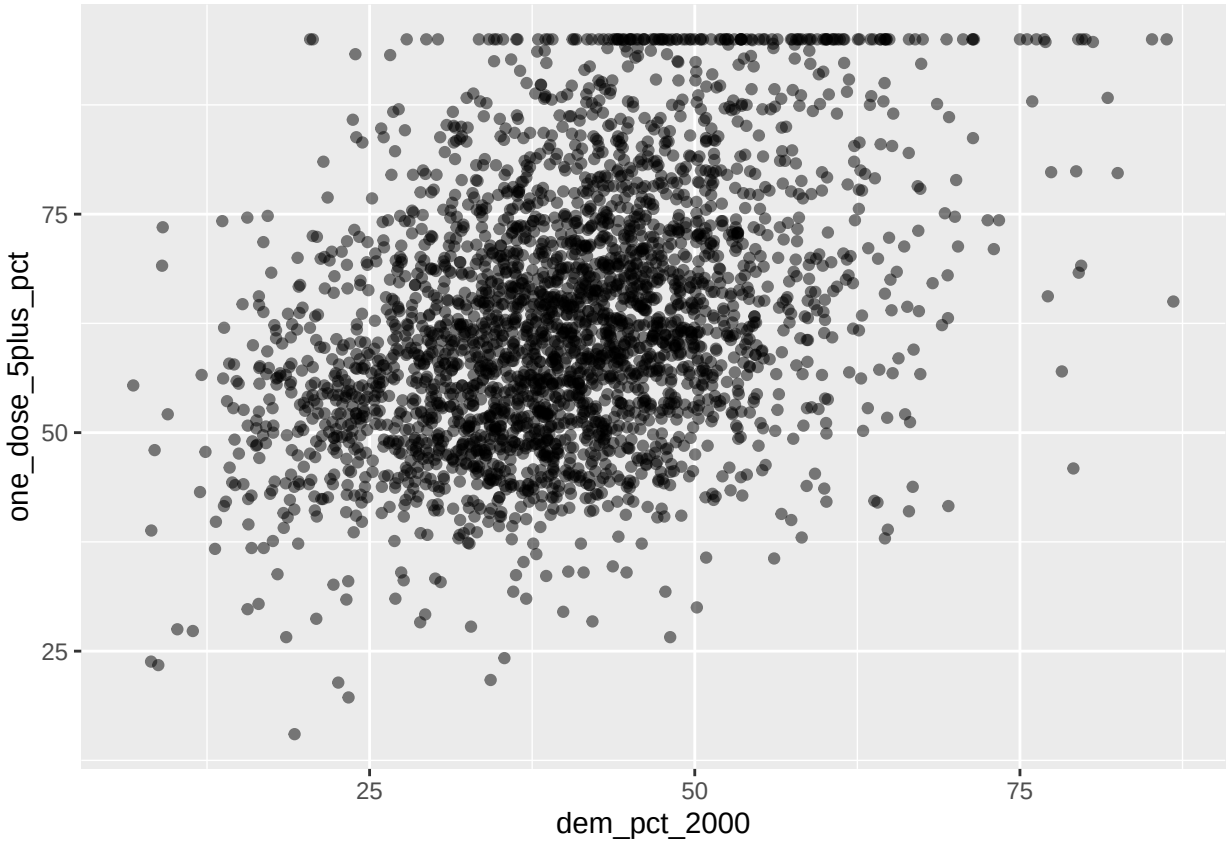
```
covid_votes %>%
  ggplot(mapping = aes(x = dem_pct_2020, y = one_dose_5plus_pct)) +
  geom_point(alpha = 0.5)
```

```
## Warning: Removed 113 rows containing missing values (geom_point).
```




```
covid_votes %>%
  ggplot(mapping = aes(x = dem_pct_2000, y = one_dose_5plus_pct)) +
  geom_point(alpha = 0.5)
```

```
## Warning: Removed 113 rows containing missing values (geom_point).
```




```
cor(covid_votes$one_dose_5plus_pct, covid_votes$dem_pct_2000,
    use = "pairwise")
```

```
## [1] 0.3941203
```

Writing Functions

```
z_score <- function(x) {
  (x - mean(x, na.rm = TRUE))/sd(x, na.rm = TRUE)
}

z_score(c(1, 2, 3, 4, 5))
```

```
## [1] -1.2649111 -0.6324555  0.0000000  0.6324555  1.2649111
```

```
covid_votes %>%
  mutate(
    one_dose_5p_z = z_score(one_dose_5plus_pct),
    one_dose_65p_z = z_score(one_dose_65plus_pct)
  )
```

```
## # A tibble: 3,114 x 10
```

```
##    fips  county    state one_d~1 one_d~2 boost~3 dem_p~4 dem_p~5 one_d~6 one_d~7
##    <chr> <chr>     <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 26039 Crawford~ MI       55.7    77.3   31.2     43.8    34.0  -0.508 -0.829
##  2 40015 Caddo Co~ OK       83.3    95     30.3     46.4    27.1   1.40   0.843
##  3 17007 Boone Co~ IL       71.1    94.5   35.1     41.8    42.2   0.556  0.795
##  4 12055 Highland~ FL       68.9    93.7   24.7     40.3    32.5   0.404  0.720
##  5 34029 Ocean Co~ NJ       71      95     32.1     47.2    35.0   0.549  0.843
##  6 01067 Henry Co~ AL       58.5    85.5   18.2     40.1    28.0  -0.314 -0.0545
##  7 27037 Dakota C~ MN       81      95     49.5     46.9    55.7   1.24   0.843
##  8 27115 Pine Cou~ MN       56.5    85     31.7     47.0    33.9  -0.452 -0.102
##  9 51750 Radford ~ VA       41.5    73.8    1.79    46.4    53.1  -1.49  -1.16
## 10 22009 Avoyelle~ LA       59.7    80.1   21.9     45.7    28.8  -0.231 -0.564
## # ... with 3,104 more rows, and abbreviated variable names
## #   1: one_dose_5plus_pct, 2: one_dose_65plus_pct, 3: booster_5plus_pct,
## #   4: dem_pct_2000, 5: dem_pct_2020, 6: one_dose_5p_z, 7: one_dose_65p_z
```

```
## Finding the z scores for all the variables from start to end
covid_votes %>%
  mutate(
    across(one_dose_5plus_pct:dem_pct_2020, z_score)
  )
```

```
## # A tibble: 3,114 x 8
##    fips  county           state one_dose_5plus~1 one_d~2 boost~3 dem_p~4 dem_p~5
##    <chr> <chr>            <chr>            <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 26039 Crawford County  MI              -0.508 -0.829    0.531  0.340   0.0471
##  2 40015 Caddo County     OK               1.40   0.843    0.439  0.556  -0.387
##  3 17007 Boone County     IL               0.556  0.795    0.927  0.163   0.563
##  4 12055 Highlands County FL               0.404  0.720   -0.135  0.0402 -0.0487
##  5 34029 Ocean County     NJ               0.549  0.843    0.623  0.624   0.109
##  6 01067 Henry County     AL              -0.314 -0.0545  -0.799  0.0255 -0.328
##  7 27037 Dakota County    MN               1.24   0.843    2.40   0.598   1.41
##  8 27115 Pine County      MN              -0.452 -0.102    0.577  0.612   0.0393
##  9 51750 Radford city     VA              -1.49  -1.16    -2.47   0.556   1.25
## 10 22009 Avoyelles Parish LA              -0.231 -0.564   -0.424  0.501  -0.280
## # ... with 3,104 more rows, and abbreviated variable names
## #   1: one_dose_5plus_pct, 2: one_dose_65plus_pct, 3: booster_5plus_pct,
## #   4: dem_pct_2000, 5: dem_pct_2020
```