Annotation

2022-07-17

Introduction

We will make a scatter plot of lifetime high scores in teteis against cumulative hours playing tetris. Normally, we will see positive correlation.

Basic Scatter Plot

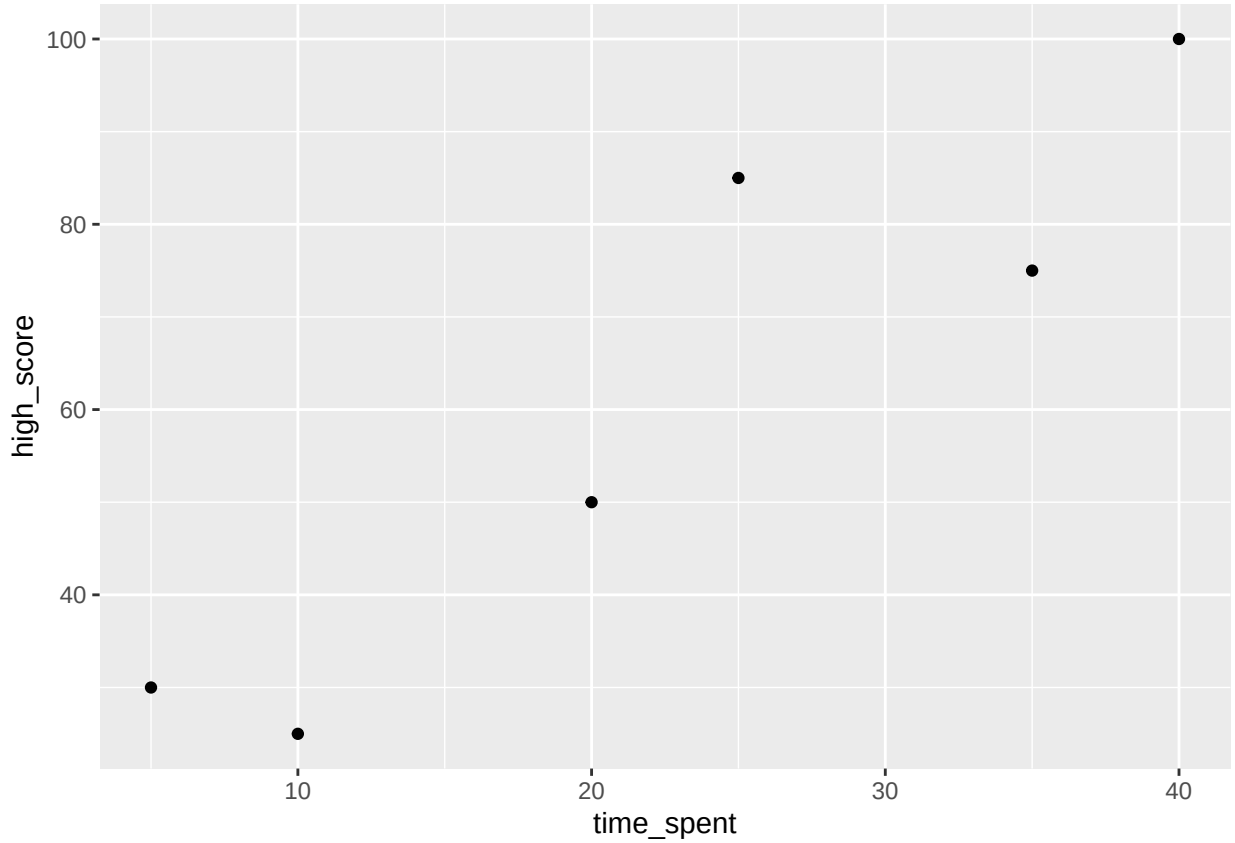
```
library(tidyverse)
```

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

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## 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()
```

```
kid <- c("Nick", "Jessica", "Justin", "Brandi", "Kelly", "Enrique")
time_spent <- c(40, 35, 25, 20, 10, 5)
high_score <- c(100, 75, 85, 50, 25, 30)

tetris <- tibble(kid, time_spent, high_score)
tetris
```

```
## # A tibble: 6 x 3
##   kid     time_spent high_score
##   <chr>        <dbl>      <dbl>
## 1 Nick            40        100
## 2 Jessica         35         75
## 3 Justin          25         85
## 4 Brandi          20         50
## 5 Kelly           10         25
## 6 Enrique          5         30
```

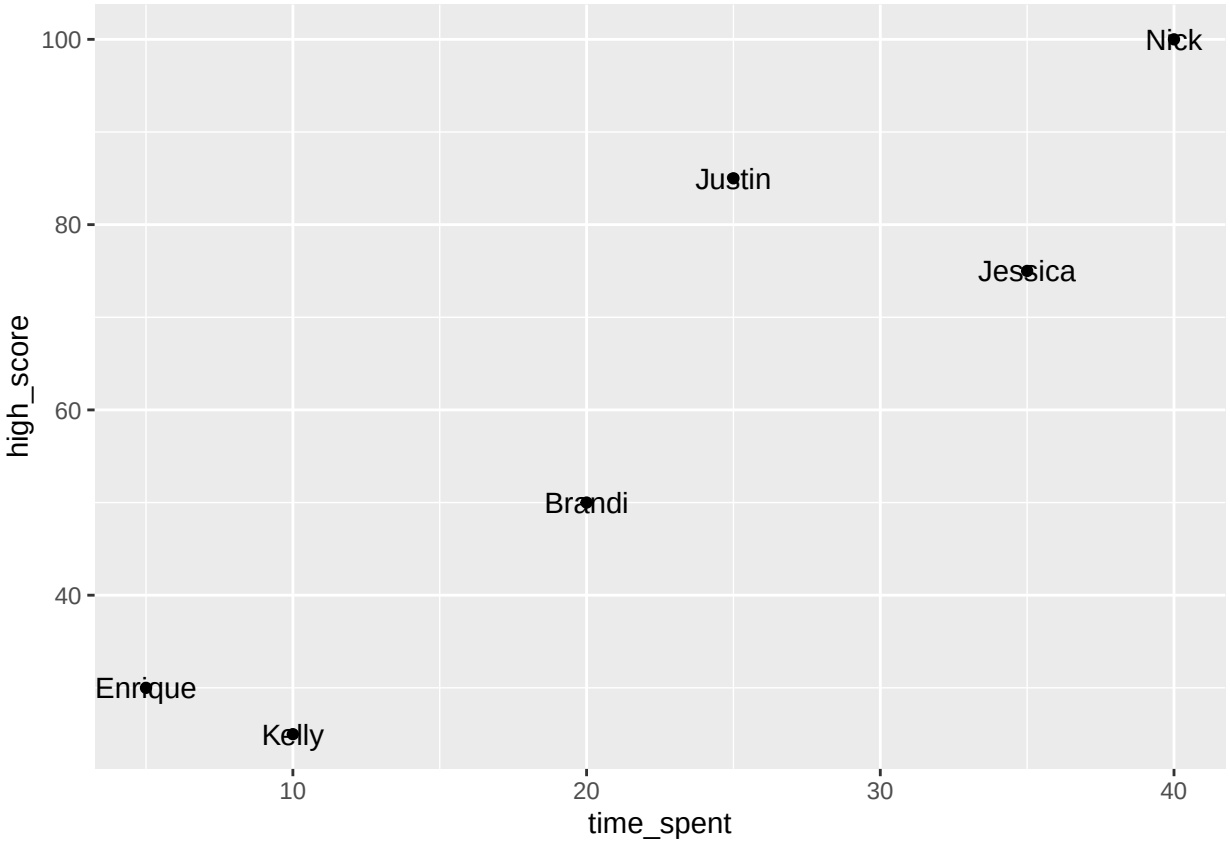
```
ggplot(tetris, aes(x = time_spent, y = high_score))+
  geom_point()
```



Scatter Plot with Label

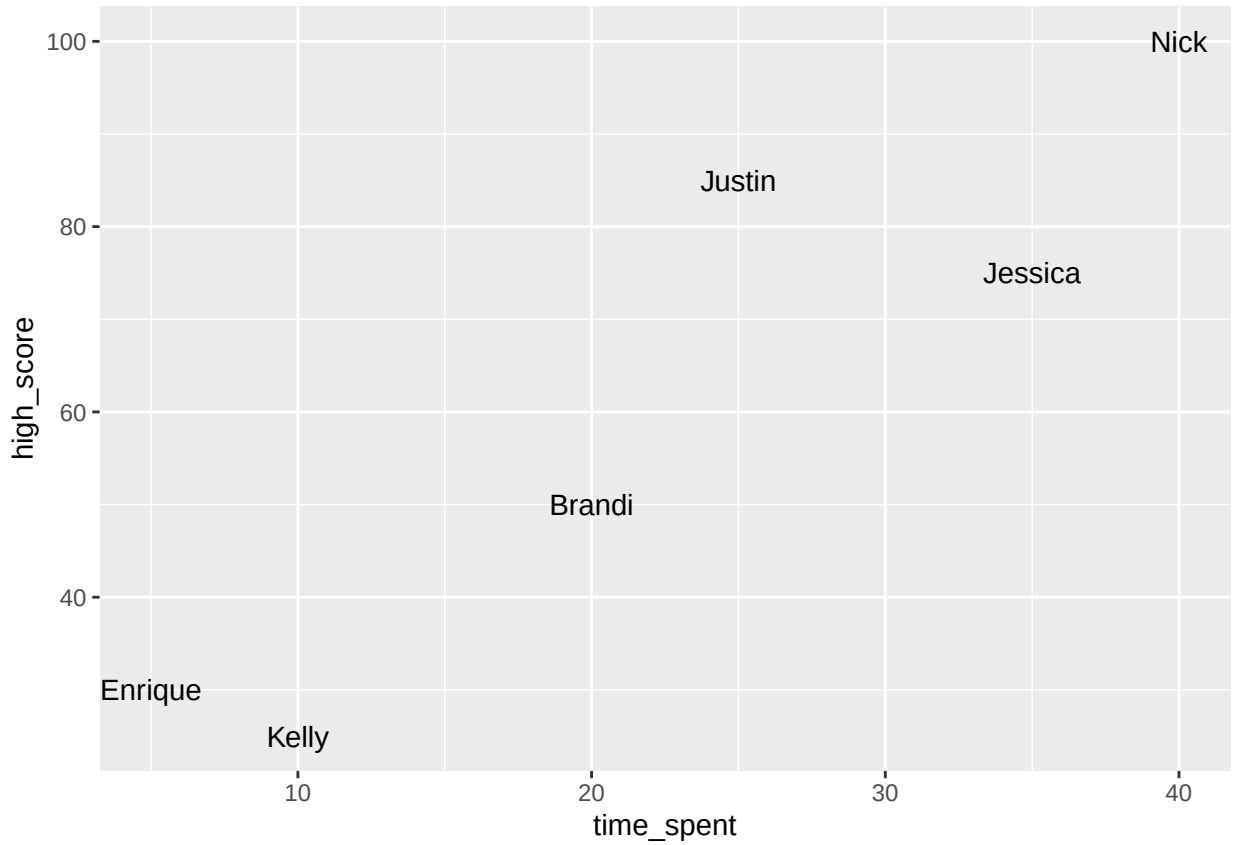
- We can add label on top of the points on the plot using `geom_text` function.

```
ggplot(tetris, aes(x = time_spent, y = high_score))+
  geom_point()+
  geom_text(aes(label = kid))
```

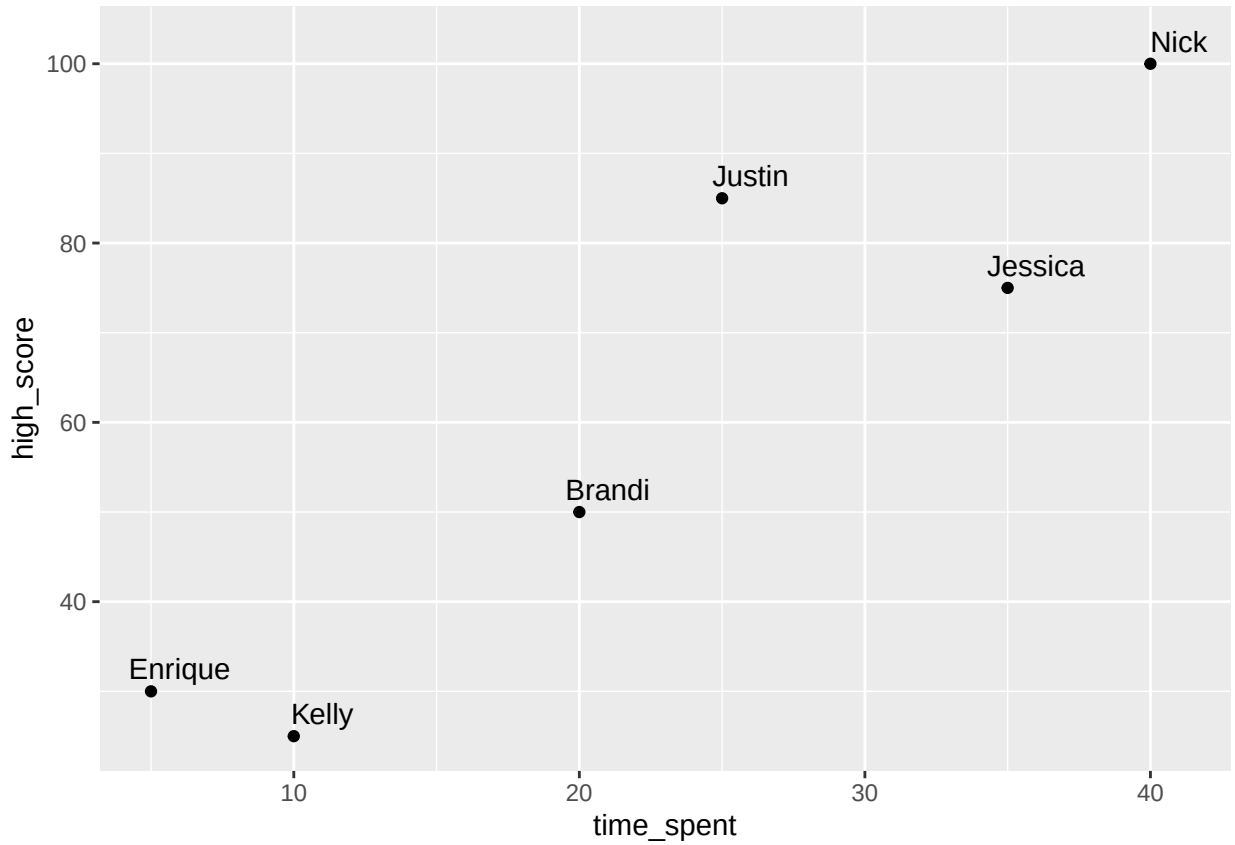


- Notice that the text and the point are overlap making it not look attractive. We have two options here:

```
# remove the geom_point so only the labels that exist
ggplot(tetris, aes(x = time_spent, y = high_score))+
  #geom_point()+
  geom_text(aes(label = kid))
```




```
# use parameter `nudge_x` or `nudge_y` inside `geom_text` function to adjust the label position
ggplot(tetris, aes(x = time_spent, y = high_score))+
  geom_point()+
  geom_text(aes(label = kid), nudge_y = 2.5, nudge_x = 1)
```




```
# ?geom_text
```

Scatter Plot with Congress Dataset

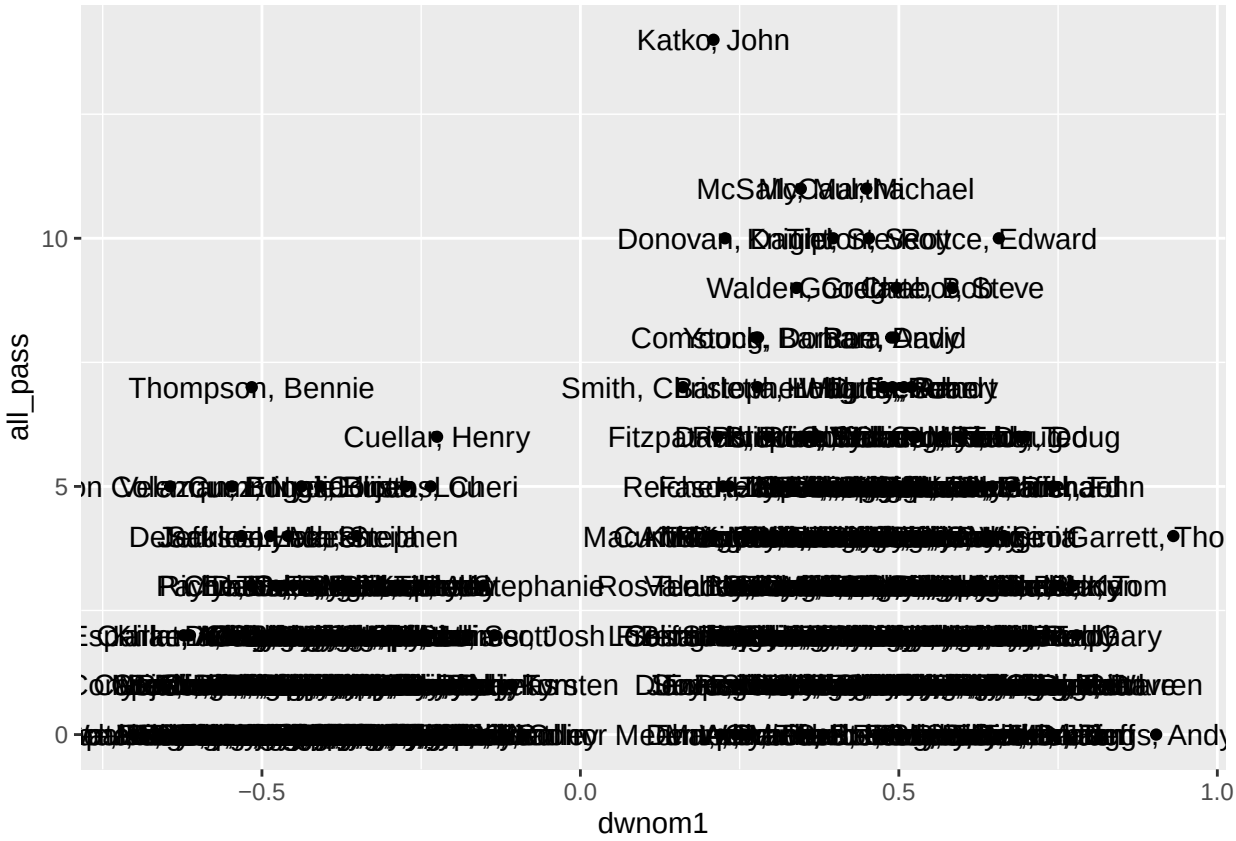
- Plot political ideology and how many bills a member passes
  - Note: **dwnom1** on the x axis, **all_pass** on the y axis

```
# load the dataset
cel <- drop_na(read_csv("cel_dataset_coursera.csv"))
```

```
## Rows: 10262 Columns: 38
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): thomas_name, st_name
## dbl (36): thomas_num, icpsr, congress, year, cd, dem, elected, female, votep...
##
## 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.
```

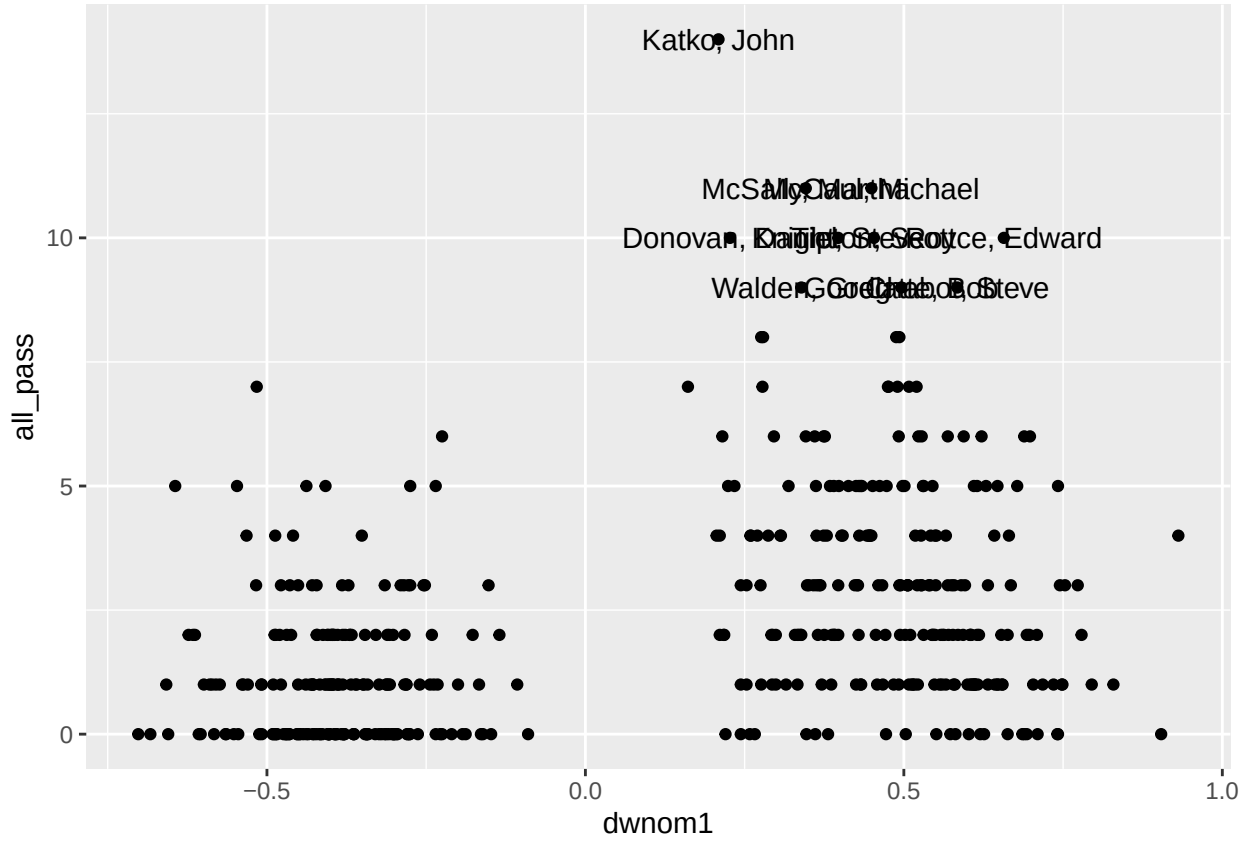
```
# use piping to feed the data to ggplot
# notice instead of puting label on geom_text, we put it inside aes() function in ggplot.
# the subsequent functions (geom_point, geom_text) will inherit the values passed to the params inside
cel %>%
```

```
  filter(congress == 115) %>%
  ggplot(aes(x = dwnom1, y = all_pass, label = thomas_name))+
  geom_point()+
  geom_text()
```



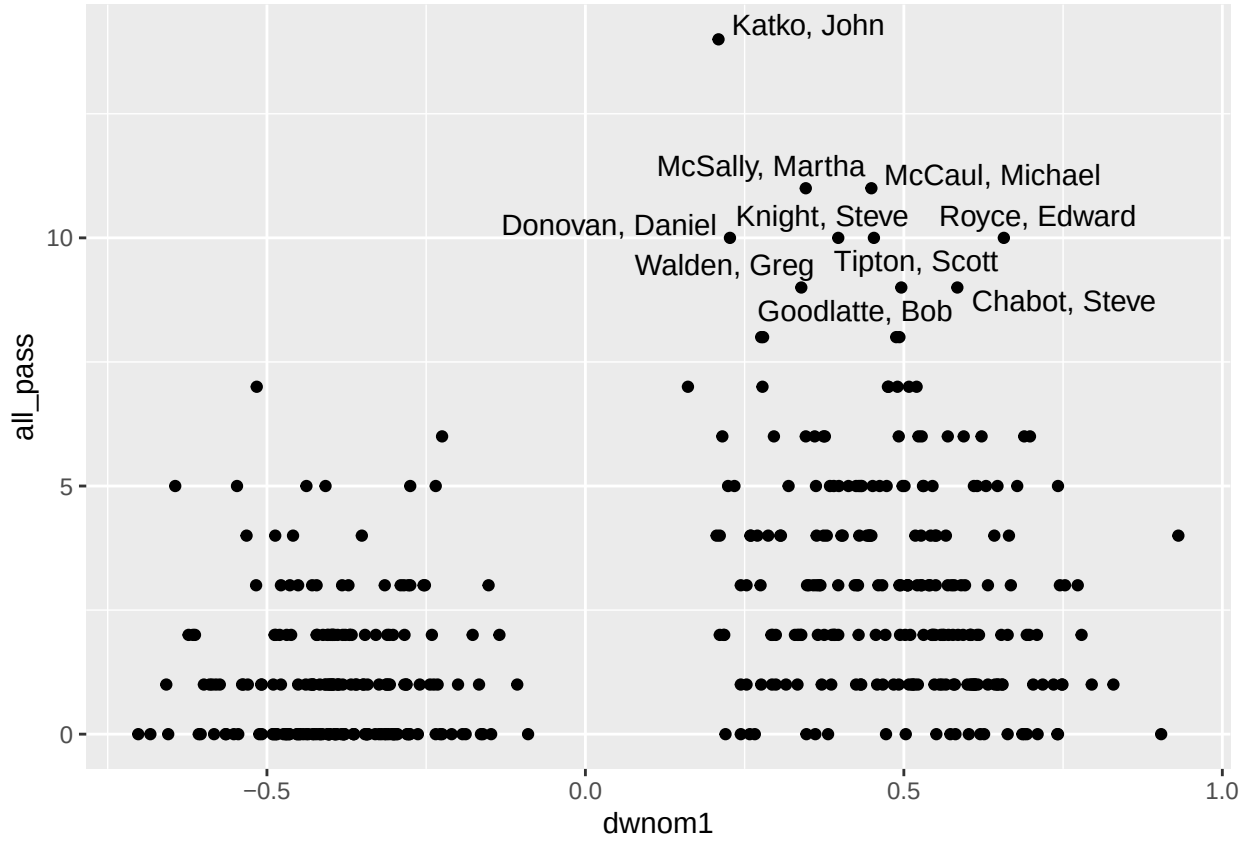
- Notice that the plot is **really messy**.
- One way to solve this is by adding text only for a small subset of the member.
- We can specify to only want the `geom_text` to apply to a subset of the data.

```
cel %>%
  filter(congress == 115) %>%
  ggplot(aes(x = dwnom1, y = all_pass, label = thomas_name))+
  geom_point()+
  # specify here that you only want the geom_text to apply to a subset of the data
  geom_text(data = filter(cel, congress == 115 & all_pass > 8))
```

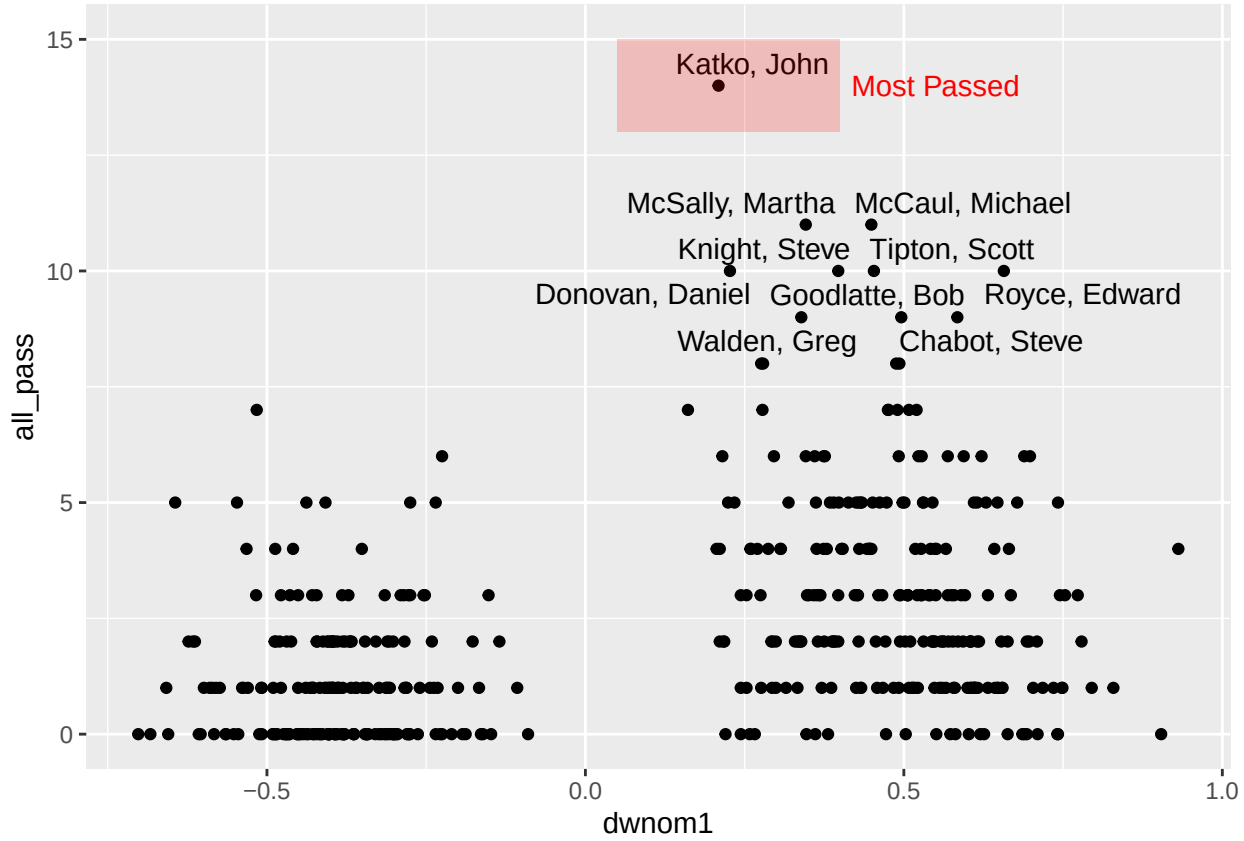


- Another way to solve this is by using a new package called `ggrepel`.
  - This package is not included inside tidyverse, but it works well with ggplot and is pretty common to use for data visualization with R.

```
# install.packages("ggrepel")
library(ggrepel)

# ggrepel will use spacey algorithm to push the text apart from each other making it more readable
cel %>%
  filter(congress == 115) %>%
  ggplot(aes(x = dwnom1, y = all_pass))+
  geom_point()+
  geom_text_repel(data = filter(cel, congress == 115 & all_pass > 8), mapping = aes(x = dwnom1, y = all
```




```
# add rectangular highlight and annotation to the plot
cel %>%
  filter(congress == 115) %>%
  ggplot(aes(x = dwnom1, y = all_pass))+
  geom_point()+
  geom_text_repel(data = filter(cel, congress == 115 & all_pass > 8), mapping = aes(x = dwnom1, y = all
  annotate("rect", xmin = .05, xmax = .4, ymin = 13, ymax = 15, alpha = .2, fill = "red")+
  annotate("text", x = .55 , y = 14, label = "Most Passed", color = 'red')
```


Katko, John
Most Passed
McSally, Martha
McCaul, Michael
Knight, Steve
Tipton, Scott
Donovan, Daniel
Goodlatte, Bob
Royce, Edward
Walden, Greg
Chabot, Steve
15
10
5
0
all_pass
−0.5
0.0
0.5
1.0
dwnom1