Calculating the Average and SD in R

`group_by()` and `summarize()`

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

function that applies groups to the data frame

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

1st argument: data frame to group

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

2nd argument: a grouping variable

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

3rd argument: a(nother) grouping variable

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

We could add a 3rd and 4th grouping variable if we wanted. Or we could have only one grouping variable.

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

A function that computes statistics (i.e., “summaries”) within each group of a grouped data frame.

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

1st argument: a grouped data frame

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

2nd argument: a quantity calculated using a variable in the grouped data frame. It is explicitly named, but you choose the name.

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

3rd argument: a(nother) quantity calculated using a variable in the grouped data frame. Again, it is explicitly named, but you choose the name.

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

Question: If we run this code, what is `smry`?

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

Question: If we run this code, what is `smry`?

Answer: A data frame.

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

Key Point

Combining `group_by()` and `summarize()` creates a data frame with the following variables:

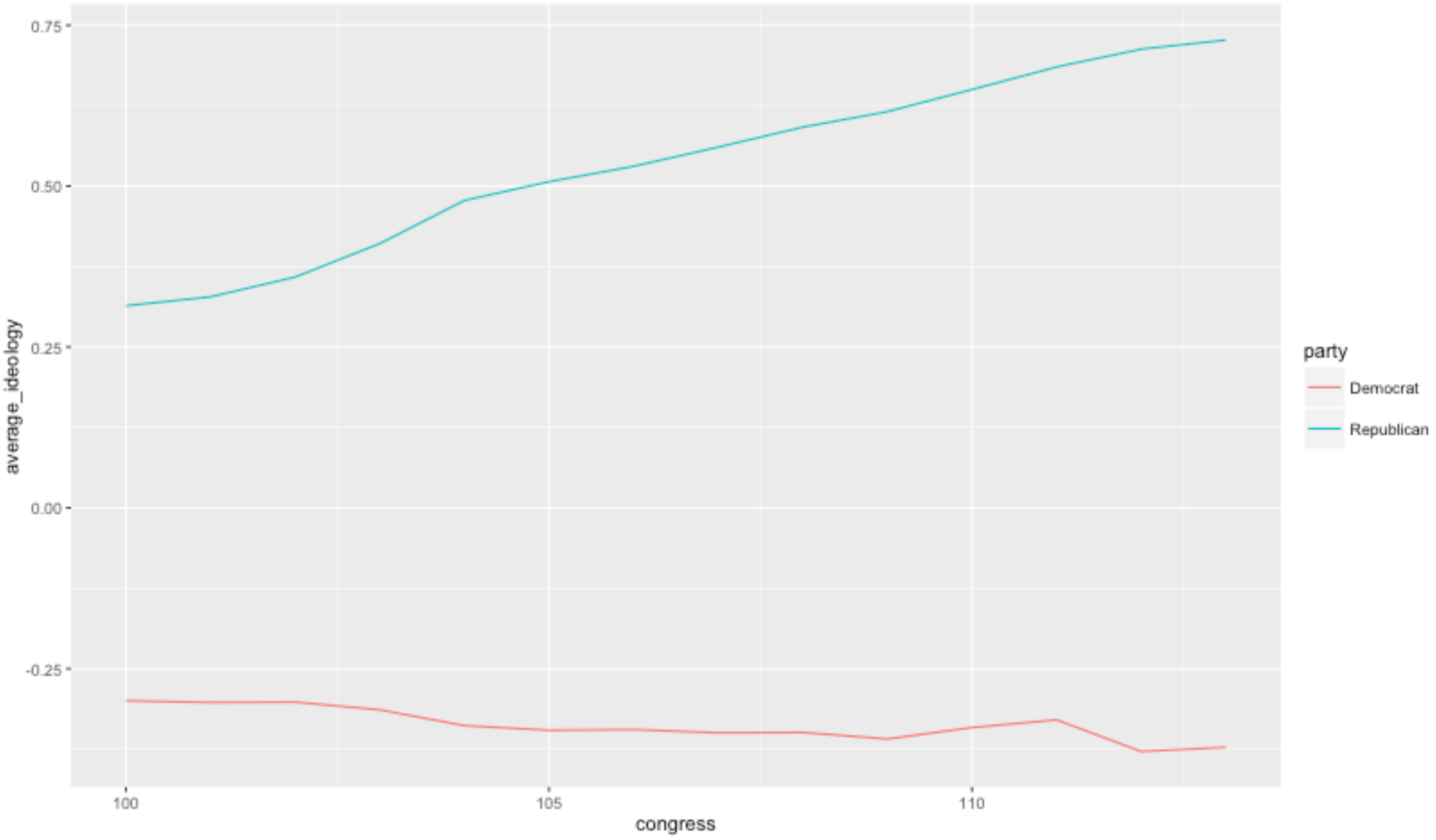
- the grouping variables
  - party
  - congress
- the summaries (argument names become variable names)
  - average_ideology
  - sd_ideology

```
# group and summarize data
grouped_df <- group_by(nominate, party, congress)
smry <- summarize(grouped_df,
                  average_ideology = mean(ideology_score),
                  sd_ideology = sd(ideology_score))
```

```
> glimpse(smry)
Observations: 28
Variables: 4
$ party            (fctr) Democrat, Democrat, Democrat, Democrat, Democrat, Democrat, De...
$ congress         (int) 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112...
$ average_ideology (dbl) -0.2997308, -0.3024198, -0.3018587, -0.3138217, -0.3383846, -0....
$ sd_ideology      (dbl) 0.1596674, 0.1619839, 0.1630104, 0.1566859, 0.1479384, 0.136459...
```

Most importantly, we can use `ggplot()` with `smry`.

```
# create line plot
ggplot(smry, aes(x = congress, y = average_ideology, color = party)) +
    geom_line()
```


average_ideology
congress
party
Democrat
Republican