

Data Visualization Basics

Election Data Science

Peter Licari

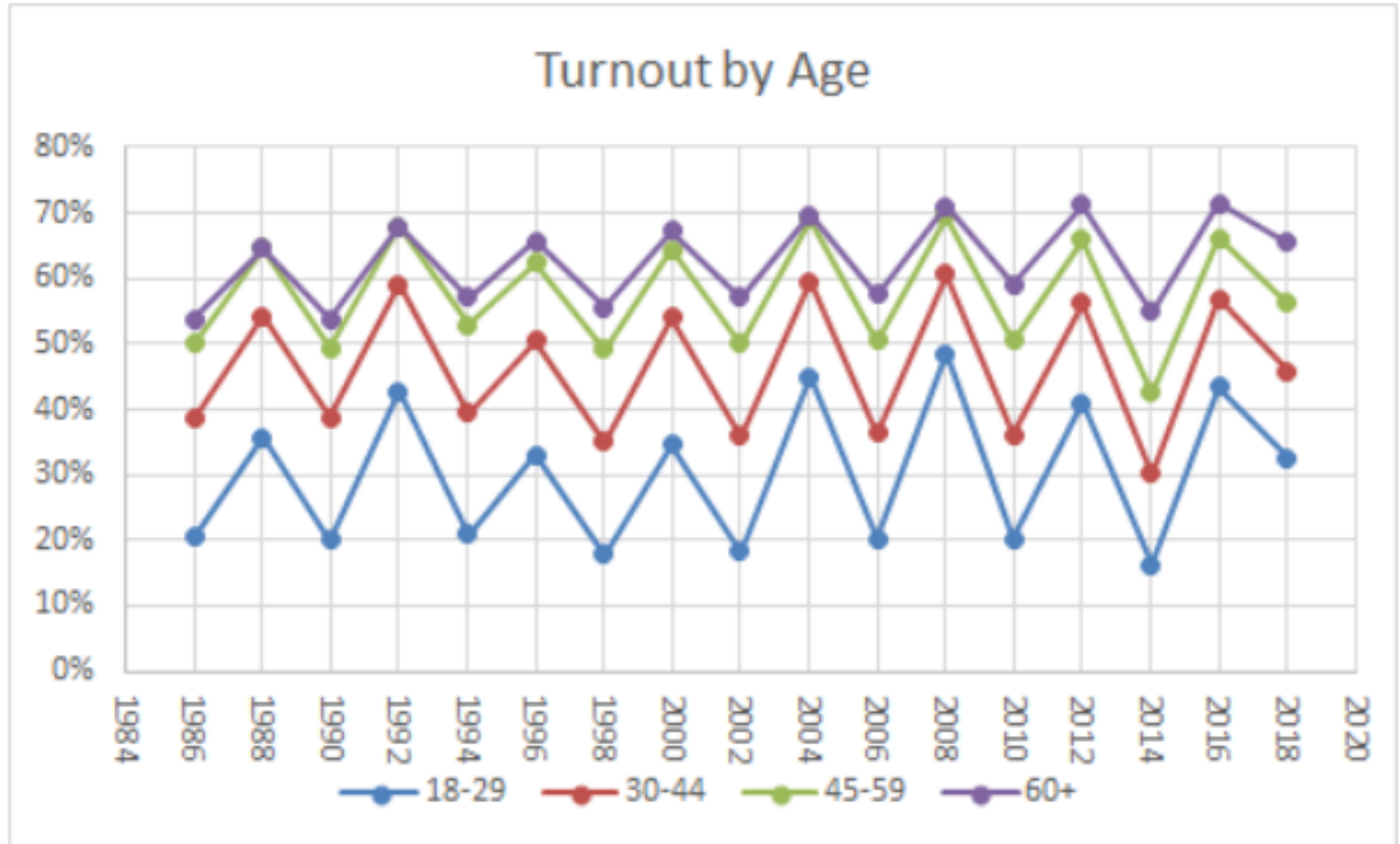
2020-09-28

The best experience is to learn by doing.



3 lessons by the end of it.

Here's our renovation project:



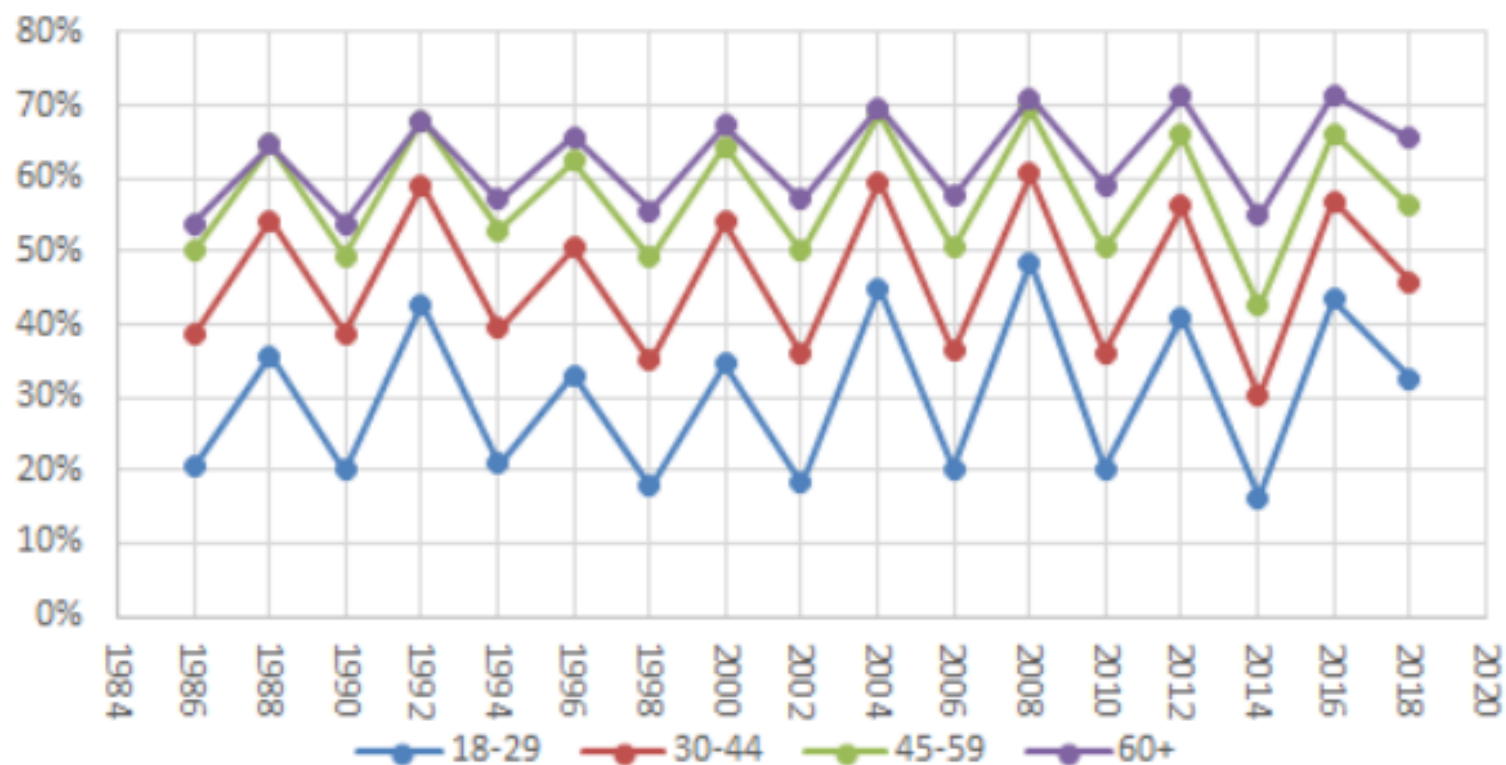
Lesson 1: Know your data.

```
readr::read_csv("Raw_Turnout_Doc.csv") %>%
  kbl() %>%
  kable_paper() %>%
  scroll_box(width = "800px", height = "250px")
```

Turnout Rate	2018	2016	2014	2012	2010	2008	2006	2004	2002
18-29	32.60%	43.40%	16.30%	40.90%	20.30%	48.40%	20.30%	45.00%	18.20%
30-44	45.80%	56.90%	30.10%	56.20%	36.10%	60.70%	36.30%	59.50%	36.00%
45-59	56.20%	66.20%	42.60%	66.00%	50.40%	69.50%	50.40%	69.00%	50.10%
60	65.50%	71.40%	54.90%	71.20%	59.00%	71.00%	57.60%	69.70%	57.10%

Lesson 2: Know your story.

Turnout by Age



Let's fix the data so we can tell the story.

```
turnout_raw <- readr::read_csv("Raw_Turnout_Doc.csv")  
  
turnout_tidyer <- turnout_raw %>%  
  pivot_longer(!`Turnout Rate`, names_to = "Year", values_to = "Percent")  
  rename(Age = `Turnout Rate`)
```

Age	Year	Percent
18-29	2018	32.60%
18-29	2016	43.40%
18-29	2014	16.30%
18-29	2012	40.90%
18-29	2010	20.30%

Preliminary replication

```
ggplot(data = turnout_tidy, aes(x = Year, y = Percent, group = Age, color = Age))
```

Manipulating the Data

Age	Year	Percent
18-29	2018	32.60%
18-29	2016	43.40%
18-29	2014	16.30%
18-29	2012	40.90%
18-29	2010	20.30%

```
turnout_tidyer <- turnout_tidyer %>%  
  mutate(Year = as.integer(Year),  
         El_Type = case_when(  
           Year %% 4 == 0 ~ "Presidential Elections",  
           TRUE ~ "Midterm Elections"
```

Raw re-design

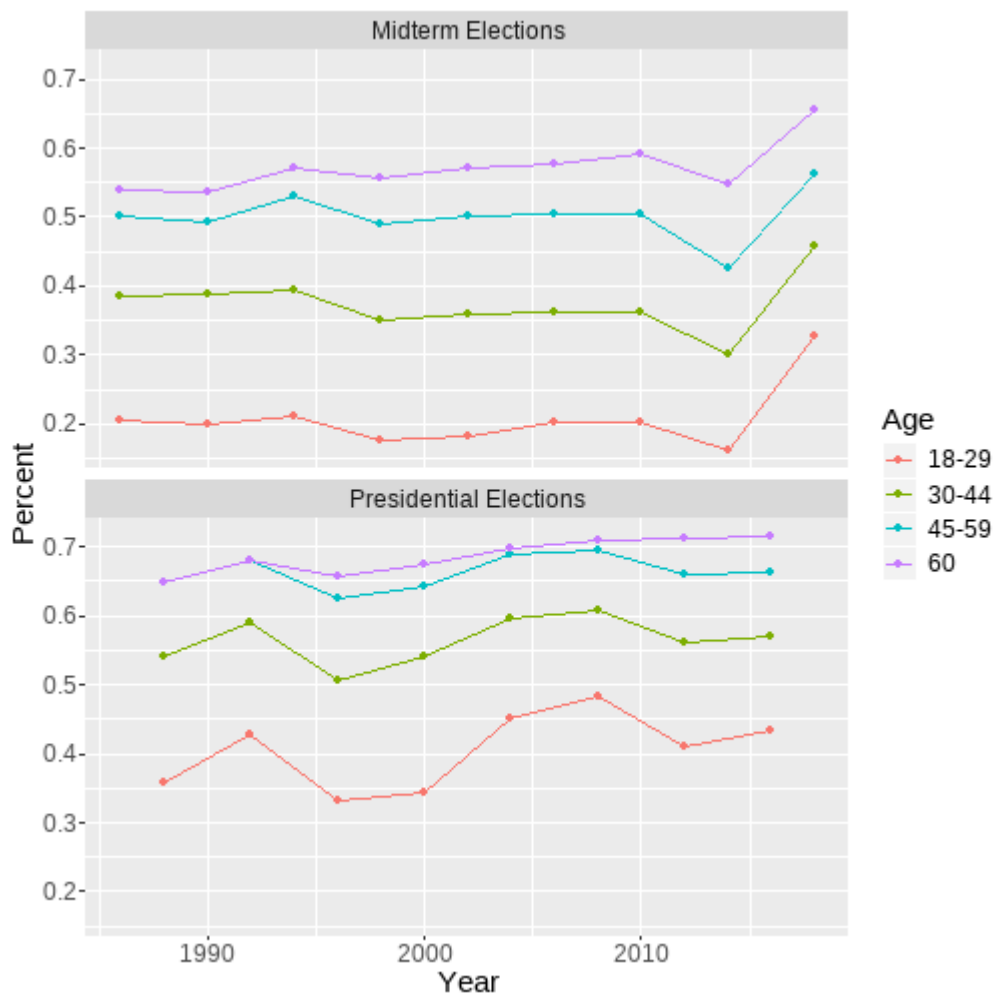
```
ggplot(data = turnout_tidyer, aes(x = Year, y = Percent,  
                                  group = Age, color = Age)) +  
  facet_wrap(~El_Type) +  
  geom_line() + geom_point()
```

Lesson 3: Know Your Audience

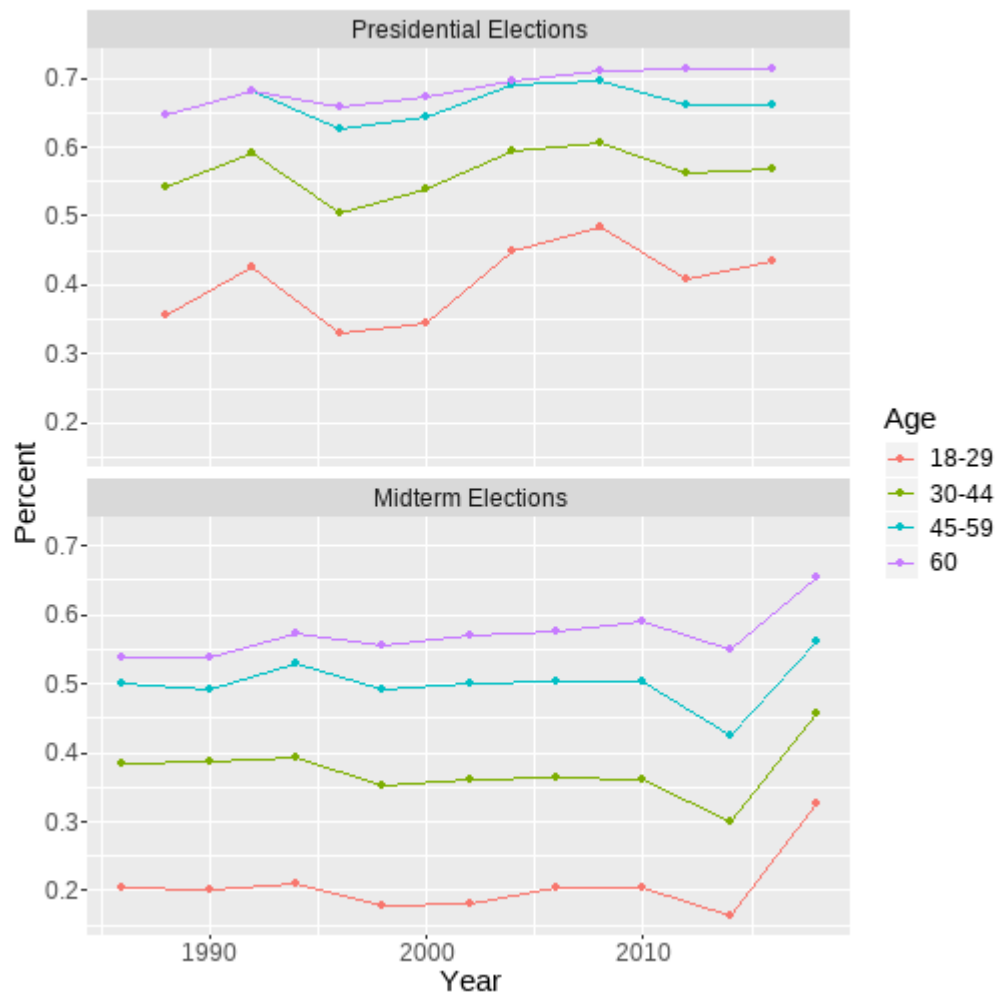
Make it readable with your audience in mind.

```
p <- ggplot(data = turnout_tidy, aes(x = Year, y = Percent,  
                                     group = Age, color = Age)) +  
  facet_wrap(~El_Type) +  
  geom_line() + geom_point()
```

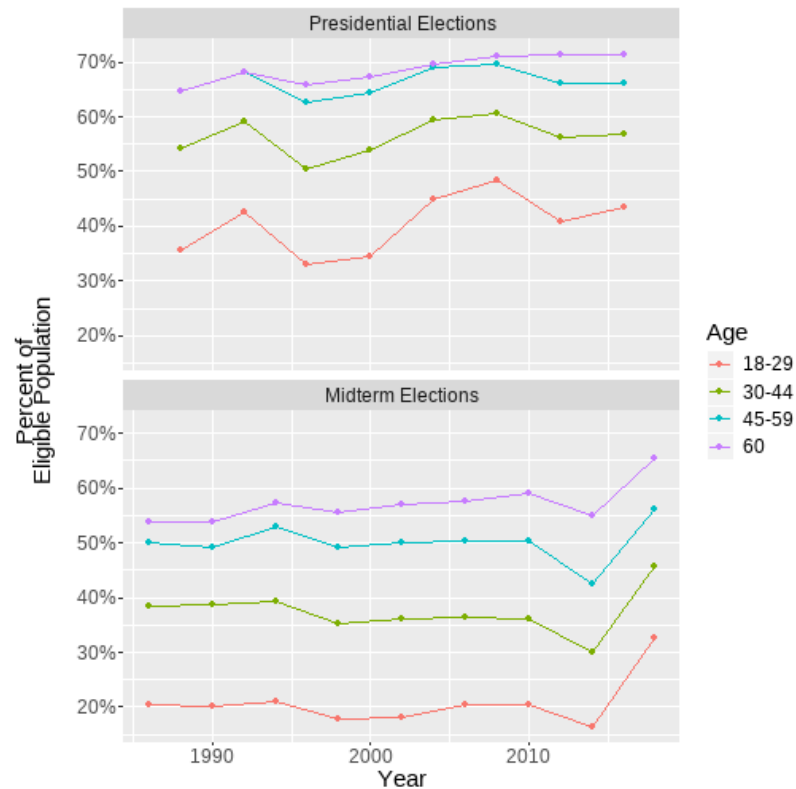
```
p <- p + facet_wrap(~El_Type, ncol = 1)
p
```



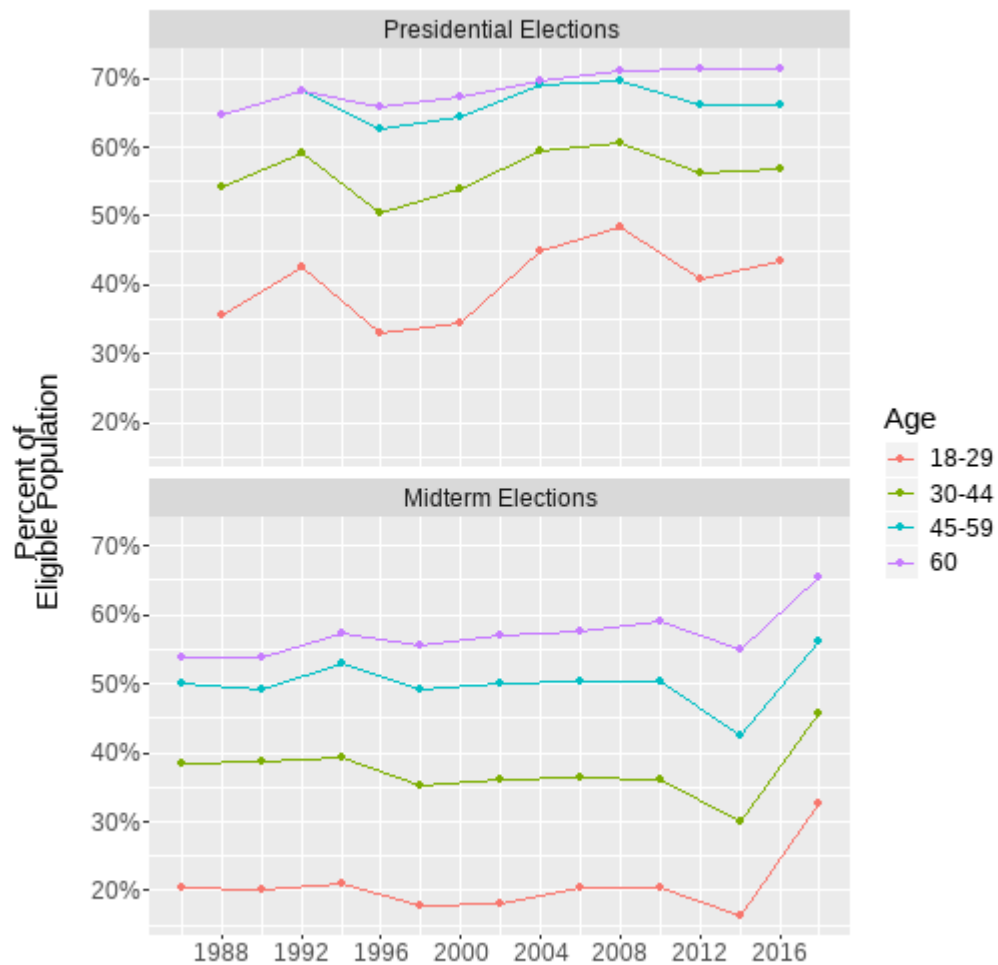
```
p <- p + facet_wrap(~forcats::fct_rev(EI_Type), ncol = 1)
p
```



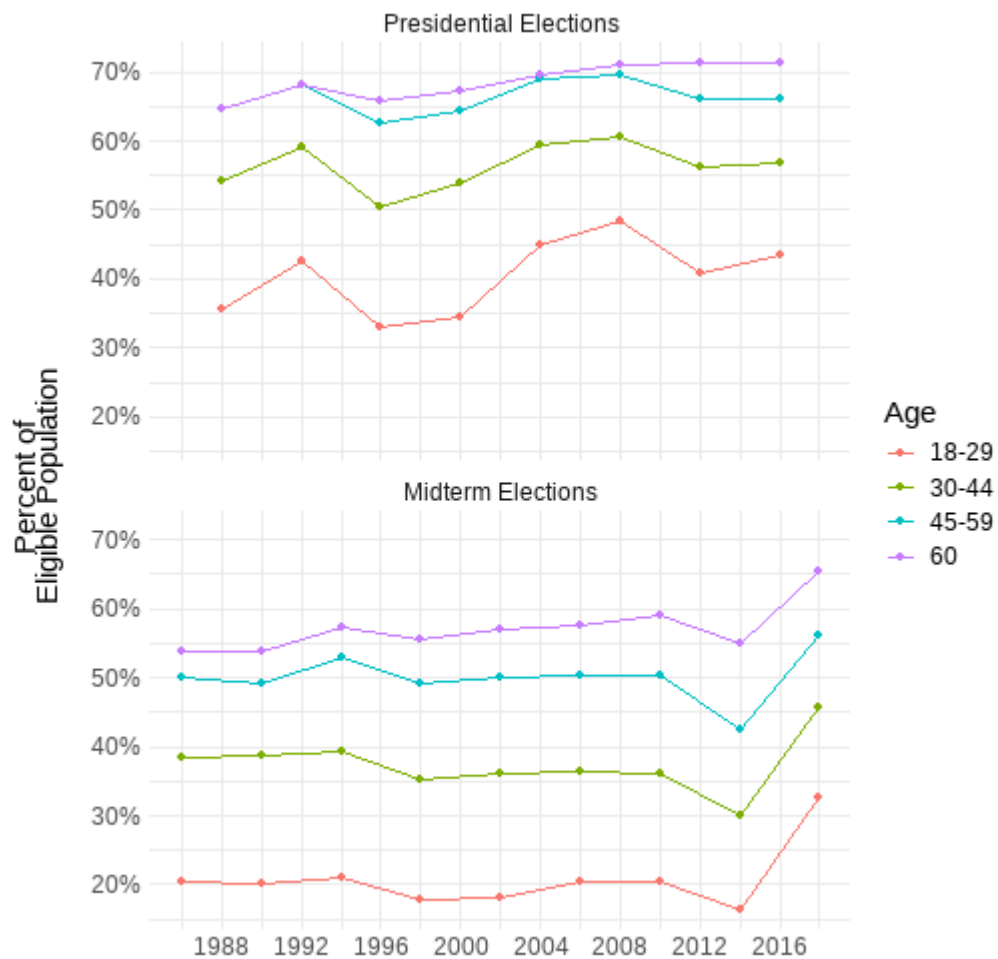
```
p <- p + scale_y_continuous(name = "Percent of\nEligible Population\n",
  labels = scales::percent_format(accuracy = 1))
p
```



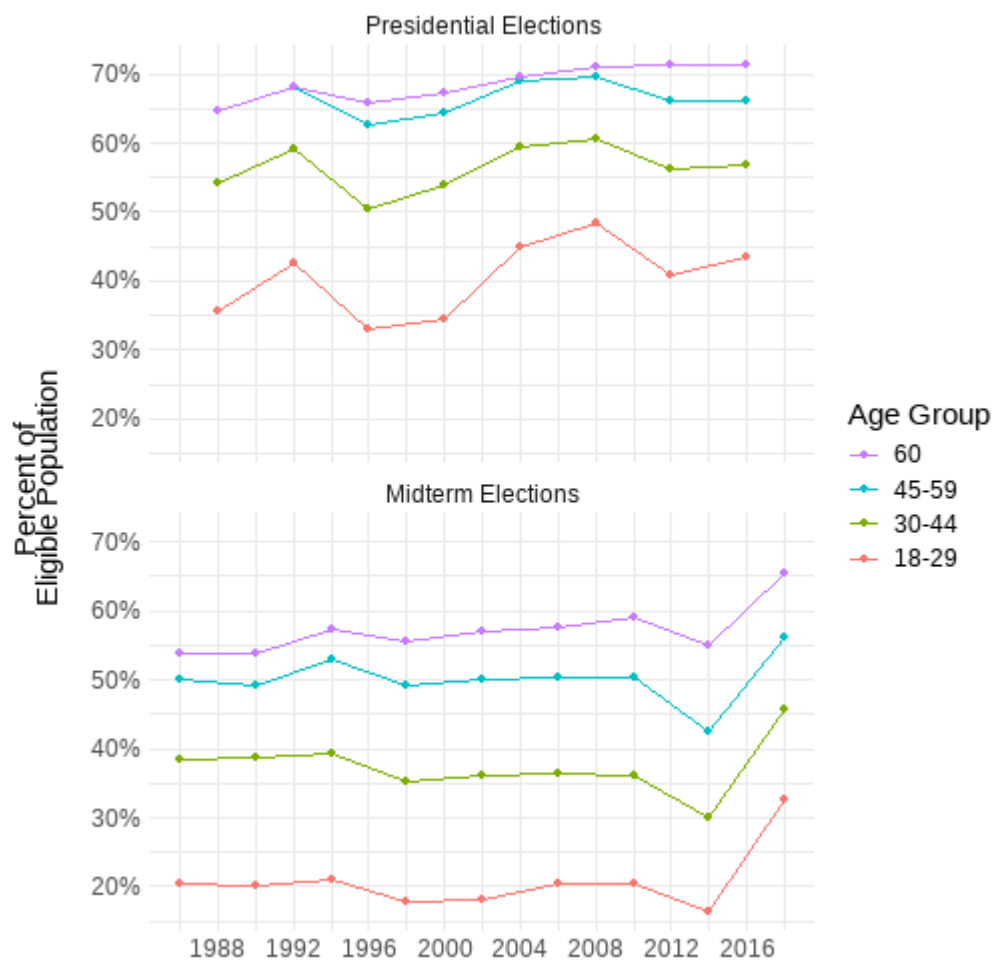

```
p <- p + scale_x_continuous(breaks = seq(1988,2020,4), name = "")  
p
```



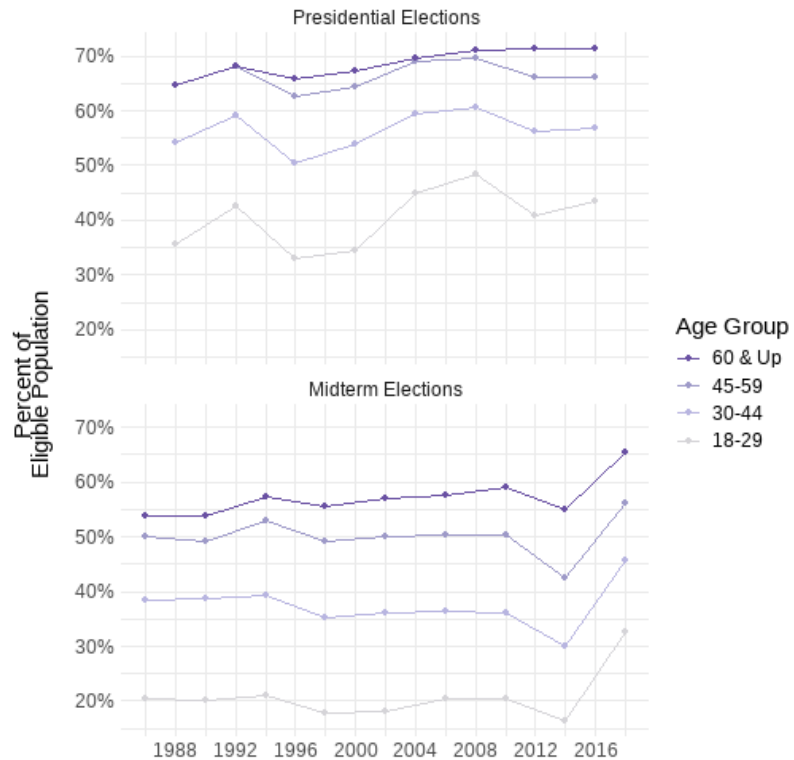
```
p <- p + theme_minimal()
p
```



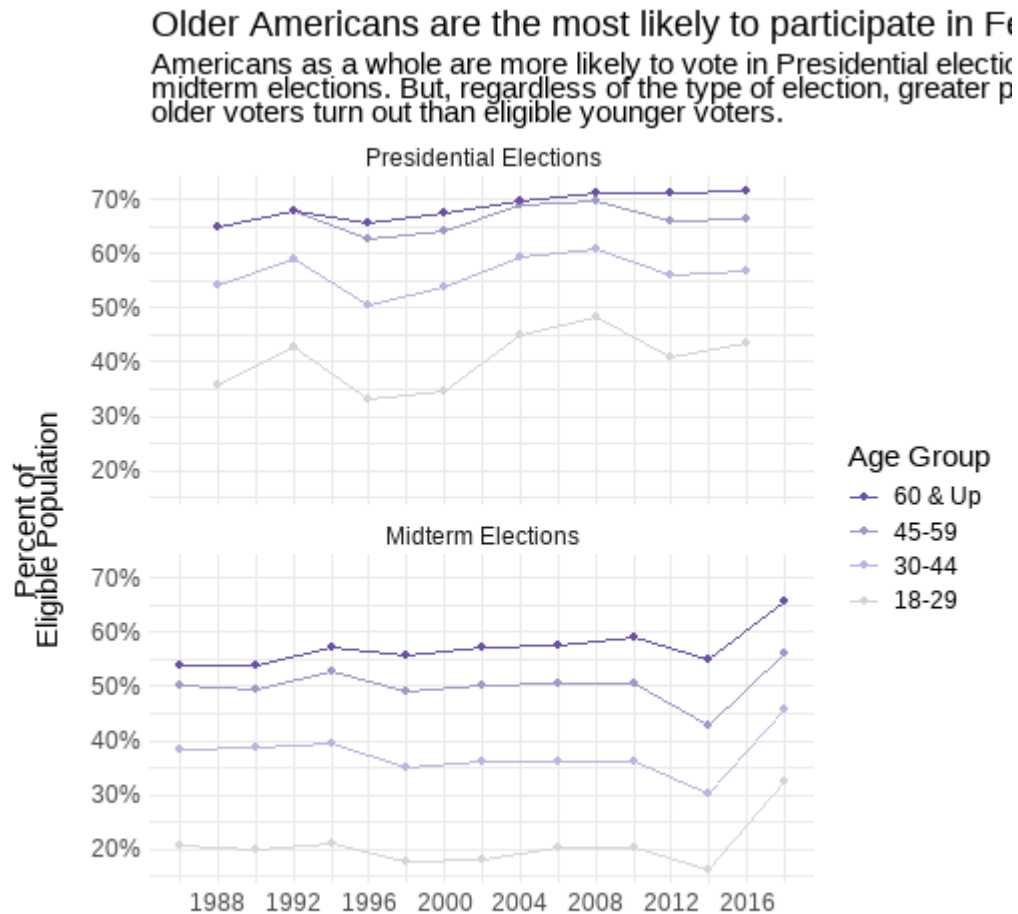
```
p <- p + scale_color_discrete(breaks = c("60", "45-59", "30-44", "18-29"))
p
```



```
p <- p + scale_color_manual(breaks = c("60", "45-59", "30-44", "18-29"),
                             values = c('#D5D4D8', '#B8B5E0', '#9e9ac8',
                                         '#808080'),
                             labels = c("60" = "60 & Up", "45-59" = "45-59",
                                         "30-44" = "30-44", "18-29" = "18-29"))
```



```
p <- p + labs(title = "Older Americans are the most likely to participate in elections.",
              subtitle = "Americans as a whole are more likely to vote in Presidential than in midterm elections. But, regardless of the type of election, greater proportions of older voters turn out than eligible younger voters.",
              caption = "Data from electproject.org. Analysis by Peter Licari.")
```



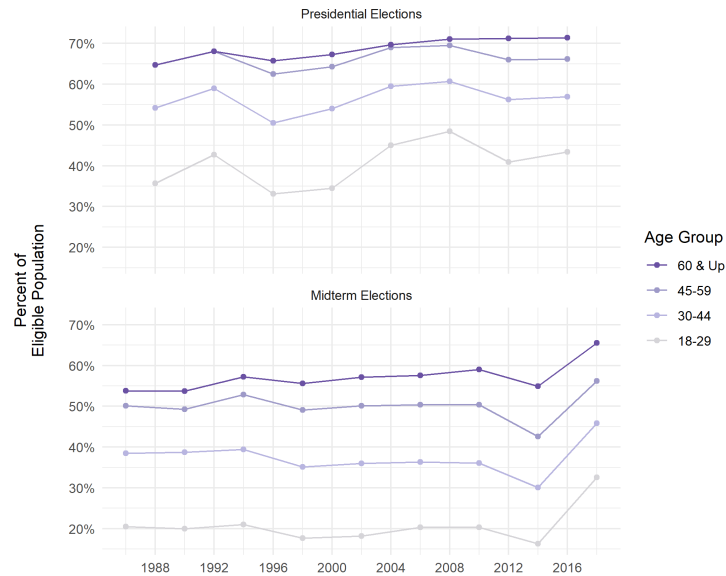
Data from electproject.org.
Analysis by Peter Licari.

Final Output

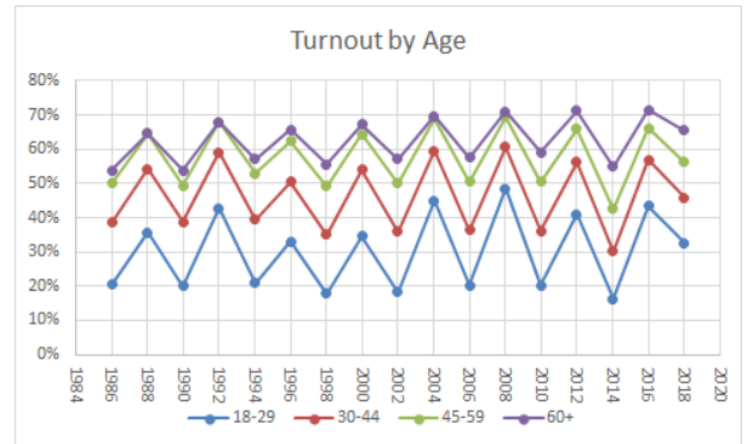
Comparison

Older Americans are the most likely to participate in Federal elections

Americans as a whole are more likely to vote in Presidential elections versus midterm elections. But, regardless of the type of election, greater proportions of eligible older voters turn out than eligible younger voters.



Data from electproject.org.
Analysis by Peter Lican.



How were we able to do all of that?

- `ggplot` uses a philosophy called the "grammar of graphics."
- Basically, there isn't a single element on the plot that can't be altered.
- Things can be built up and layered on top of each other.
- Different aesthetic elements within your control.

Overall design tips:

- Default colors rarely work well.
- Default themes tend to work worse.
- Don't be afraid of white space.
- Don't be afraid of *text* either.
- Know the "rules" (e.g., conventions and metaphors). Once you do, break them.
- People can only focus on 1-3 things in any given plot. Make your takaway still.
- Tell the story that you want to be told--in images and text both.