Welcome!

Week 3.1: Introduction to ggplot

Jennifer Lin

PS 490: R Workshop

2021-11-01

Welcome to Module 3!

Week 1 -- TODAY!

- Overview of `ggplot`
- GOAL: Make a very basic graph

Week 2 -- NEXT WEEK!

- Colors, Fills, Labels and Axis limits
- GOAL: Add to the Plot from Week 1 with color and axis labels

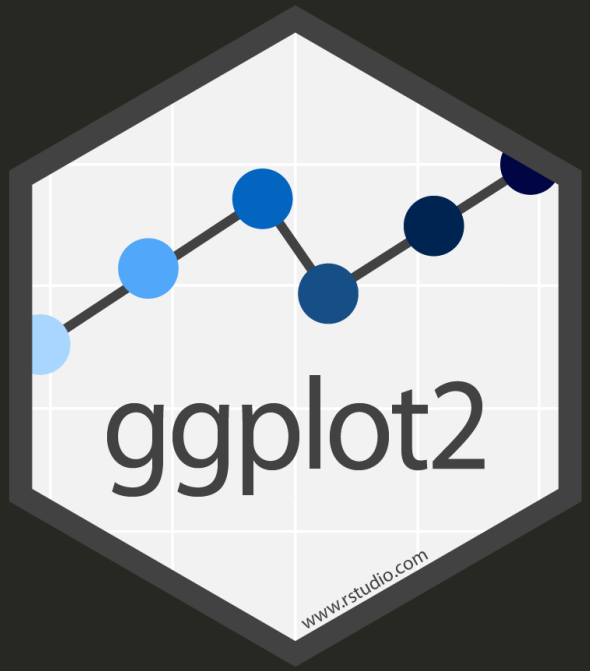
Week 3 -- WEEK AFTER NEXT WEEK!

- Themes and Fonts and Positioning
- GOAL: Finishing touches on the "publication ready" graph

Dimensions of a Good Graph

- One x and one y axis
- Clearly labeled, with titles, axis and legends containing helpful information on graph contents
- Accessible color schemes
- Minimal background, maximal foreground
- Readable fonts and font sizes
- Use graph sections to tell a meaningful story

`ggplot2`: The Grammar of Graphics



Making a Graph in R is like baking a cake...

... being fully decorated at phase 1 is unreasonable...

... and you need to build it layer by layer

`ggplot2` Makes Graphs By LAYERS

Graphs are created by layers such that they can be conceptualized as follows

```
ggplot(data layer)+
  graph layer +
  label layer +
  scale loayer +
  theme layer +
  others
```

where each layer (roughly) shows what should be there. Graphs do NOT need to follow a standard order so long as you get your desired outcome

Pipes %>% Meets Their Sibling +

- Components of a `ggplot` graph are connected with a plus (+) sign.
- You can add as many components in one large `ggplot` graph chunk as you desire. BUT you need to connect them or R will not interpret your lines as part of the same chunk

The Data Layer

The ANES -- our Friend/Data for the next three weeks

- I cleaned up the ANES 2020 data and you can choose your own adventure.
- The data include items on demographics (race, gender, age) along with an assortment of feeling thermometer variables.
- Feeling Thermometer variables start with `FT_` and all range from 0-100.

Format for Data Layer

```
ggplot(DATAFRAME,
       aes(x = X_VAR, y = Y_VAR, fill = FILL_VAR, color = COLOR_VA
```

Where

- DATAFRAME = the dataframe that you want to use
- X_VAR = the x-variable
- Y_VAR = the y-variable
- FILL_VAR = the fill variable -- use to FILL graphs that have a space to fill
- COLOR_VAR = the color variable -- use to COLOR lines or points in graphs

NOTE -- you should only include FILL and COLOR arguments IF you want to fill by a variable. If you want a standard color throughout... stay tuned!

Example of Data Layer Input

I am going to demo a plot where I look at the correlation between feelings towards Rural Americans and the BLM movement. Specifically, I want to see if there is a partisan divide between Democrats and Republicans in this interaction.

The variables that I will be inputting are

- DATAFRAME = ANES (see variable read-in code)
- X_VAR = `FT_rural`
- Y_VAR = `FT_BLM`
- FILL_VAR = DOES NOT APPLY -- This is a scatterplot
- COLOR_VAR = `PARTY` -- coloring the dots by party

```
ggplot(ANES, aes(x = FT_rural, y = FT_BLM, color = PARTY))+
```

Exercise: Find Variables and Build Data Layer

1. Load the ANES dataset (Code in script)
2. Use `names()` to find the variables and `table()` to explore the variables.
3. Identify one x and one y variable (along with an optional fill or color variable) with which you want to make a graph
4. Without code, determine what kind of graph you want to make
5. Build your data layer accordingly

The Graph Layer

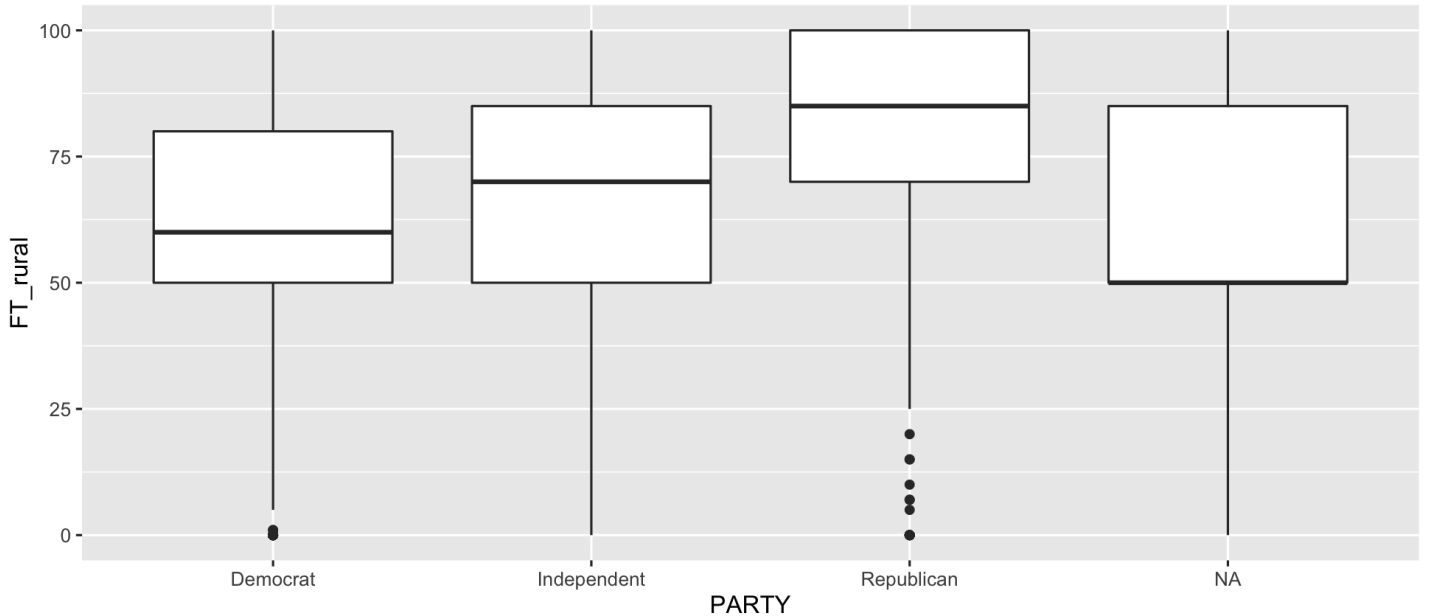
Basic Graphs

Operator	Description
`geom_line()`	line graph
`geom_point()`	scatterplot
`geom_bar()`	bar plot
`geom_histogram()`	histogram
`geom_boxplot()`	boxplot
`geom_violin()`	violin plot
`geom_sf()`	maps with shapefiles

Notice that it all starts with `geom_`

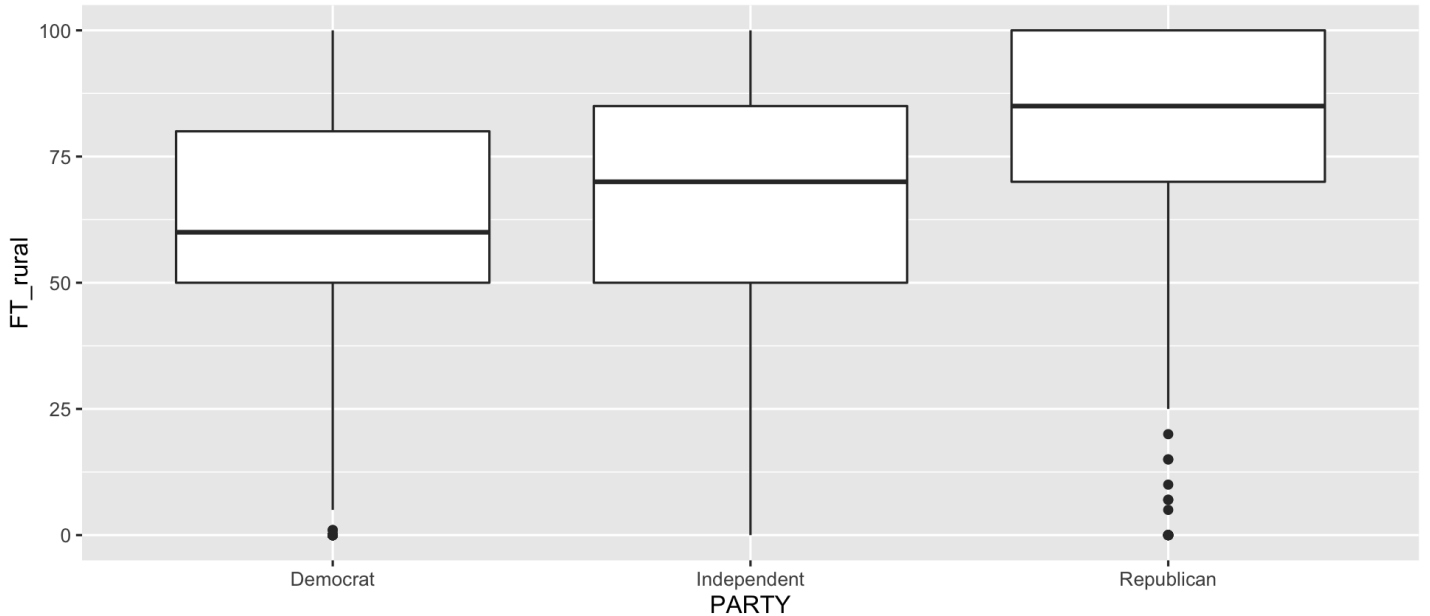
A Basic Box plot

```
ggplot(ANES, aes(x = PARTY, y = FT_rural))+
  geom_boxplot()
```



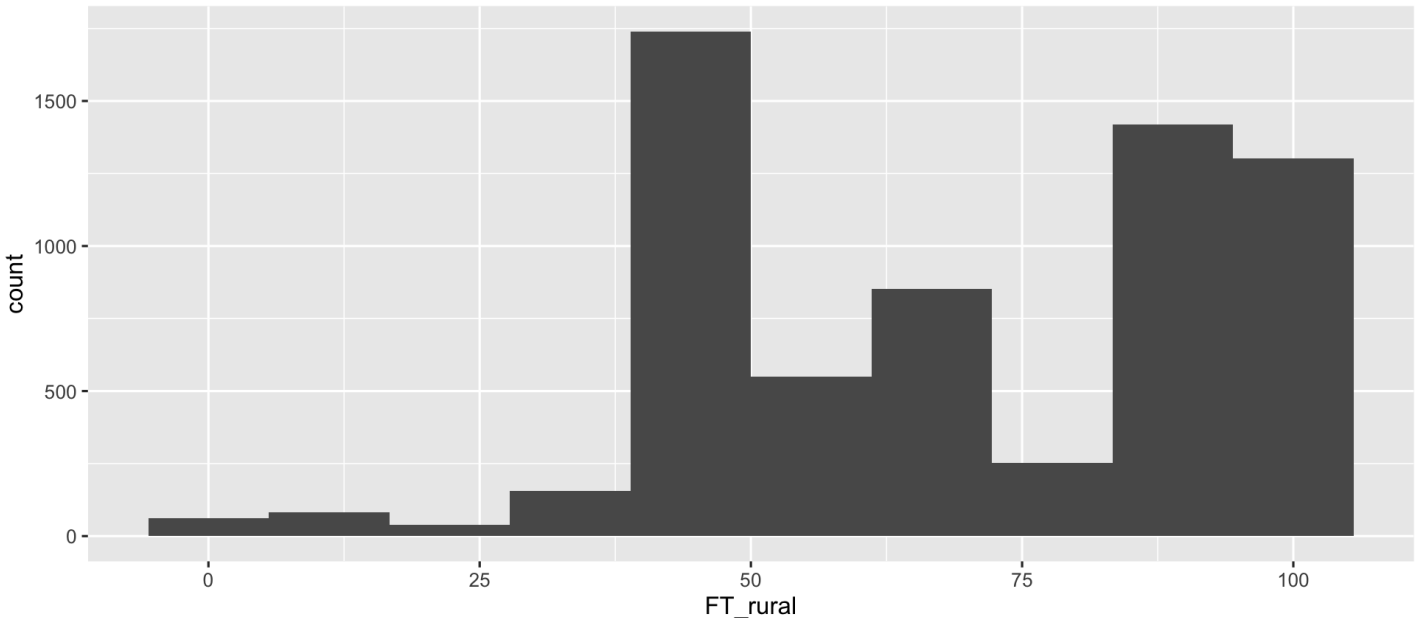
A Basic Box Plot integrating `dplyr`

```
ANES %>%
  filter(!is.na(PARTY)) %>%
  ggplot(aes(x = PARTY, y = FT_rural))+
  geom_boxplot()
```



A Basic Histogram

```
ggplot(ANES, aes(x = FT_rural))+
  geom_histogram(bins = 10)
```



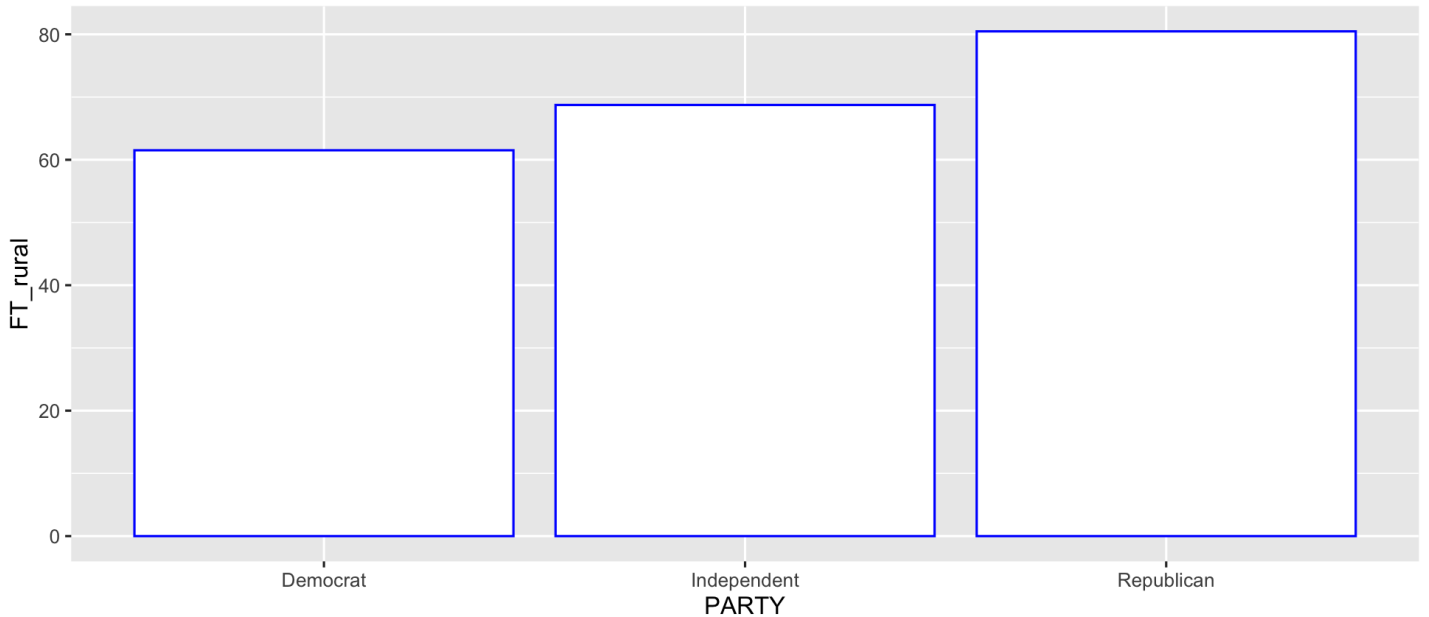
A Basic Bar Graph

First, you need to prep the data using `dplyr`.

```
by_Party <- ANES %>%
  filter(!is.na(PARTY)) %>%
  group_by(PARTY) %>%
  summarise(FT_rural = mean(FT_rural, na.rm = TRUE))
```

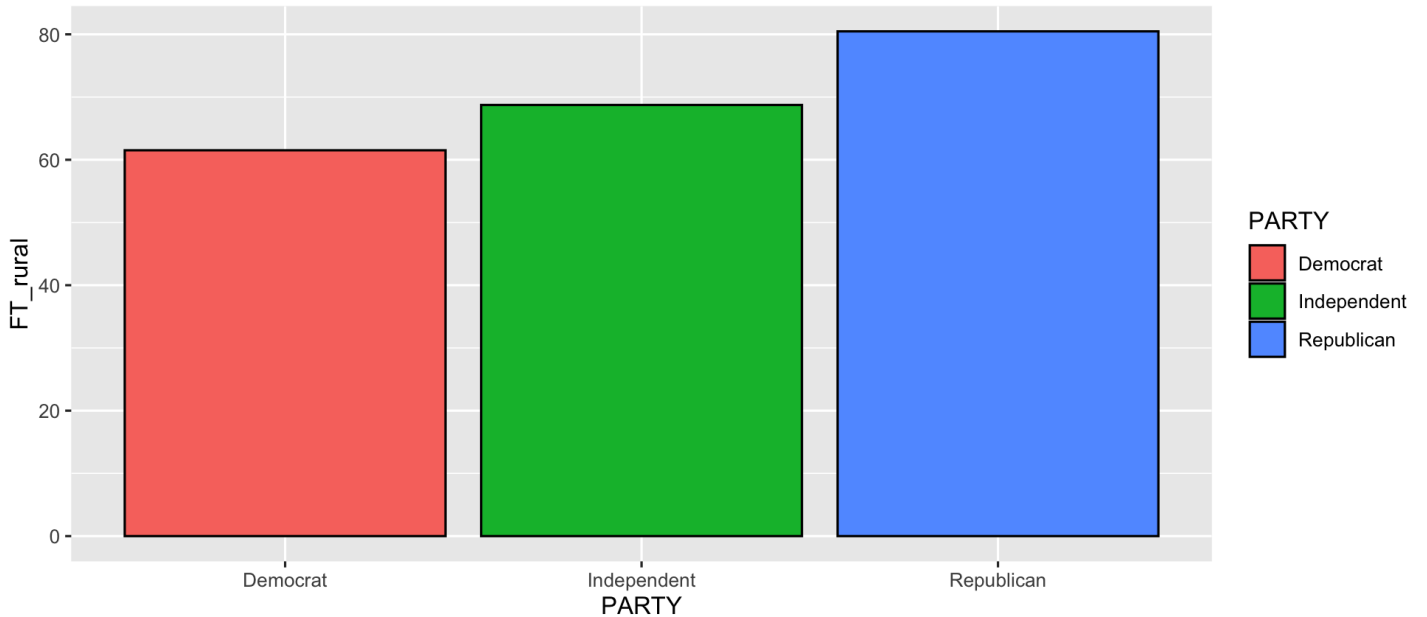
A Basic Bar Graph -- static color

```
ggplot(by_Party, aes(x = PARTY, y = FT_rural))+
  geom_bar(stat = "identity", fill = "white", color = "blue")
```



A Basic Bar Graph -- Variable color

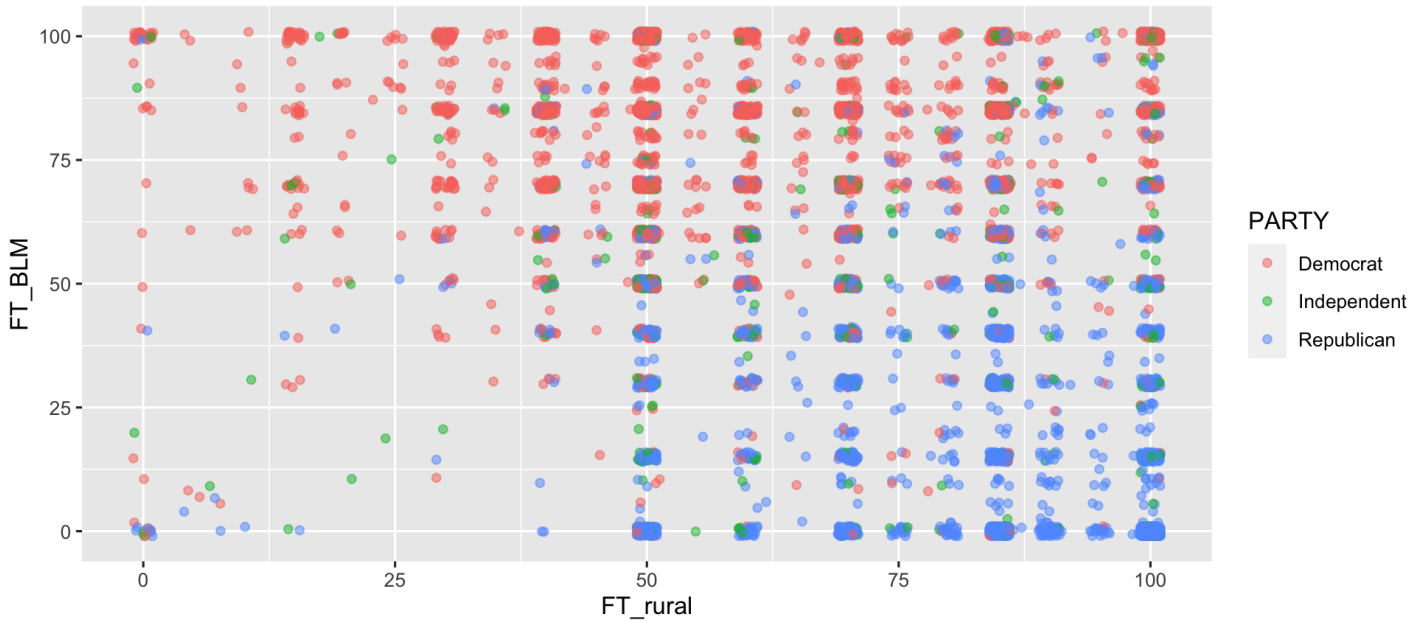
```
ggplot(by_Party, aes(x = PARTY, y = FT_rural, fill = PARTY))+
  geom_bar(stat = "identity", color = "black")
```



A Basic Scatterplot

My graph from the data layer...

```
ANES %>%
  filter(!is.na(PARTY)) %>%
  ggplot(aes(x = FT_rural, y = FT_BLM, color = PARTY))+
    geom_point(position = position_jitter(1, 1), alpha = .5)
```



Exercise: Adding a basic graph

1. From your data layer code earlier, add a + to the end if it is not already there.
2. Using the graphs explored in this section, construct a basic graph and color accordingly

Your Submission to the Lab Assignment for this week

1. Export the graph as a PDF using the "Export" button on the upper left hand corner of the plot window.
2. Upload your PDF AND the code, with your answers to the questions
3. Don't worry about spicing colors or labels this week. Will do that next week.