Distribution Plot

2022-07-17

```
knitr::opts_chunk$set(echo = TRUE)
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()
```

```
library(knitr)
```

Load the Dataset

```
data <- read_csv('cces_sample_coursera.csv')
```

```
## Rows: 1000 Columns: 25
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (25): caseid, region, gender, educ, edloan, race, hispanic, employ, mars...
##
## 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.
```

Data Visualization

Remember the same process that we will do everytime we want to generate plot using `ggplot`: 1. Start with the `ggplot` function. 2. Add the `geom_` that you want. 3. Mess around with the additional visual elements such as *title*, *lables*, etc.

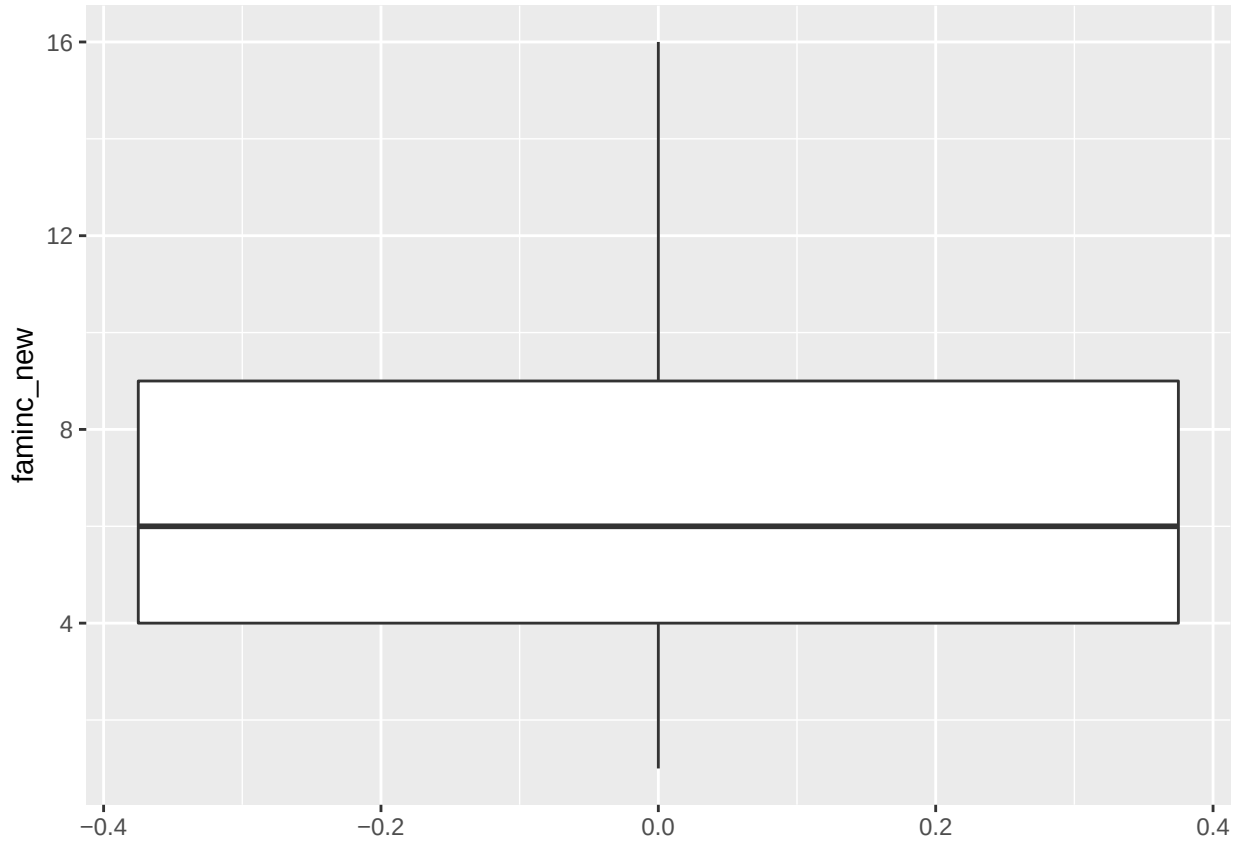
Boxplot

Reference: - https://r-graphics.org/recipe-quick-boxplot - https://r-graphics.org/recipe-distribution-basic-boxplot

Reference for Violin Plot: - https://r-graphics.org/recipe-distribution-violin#RECIPE-DISTRIBUTION-VIOLIN

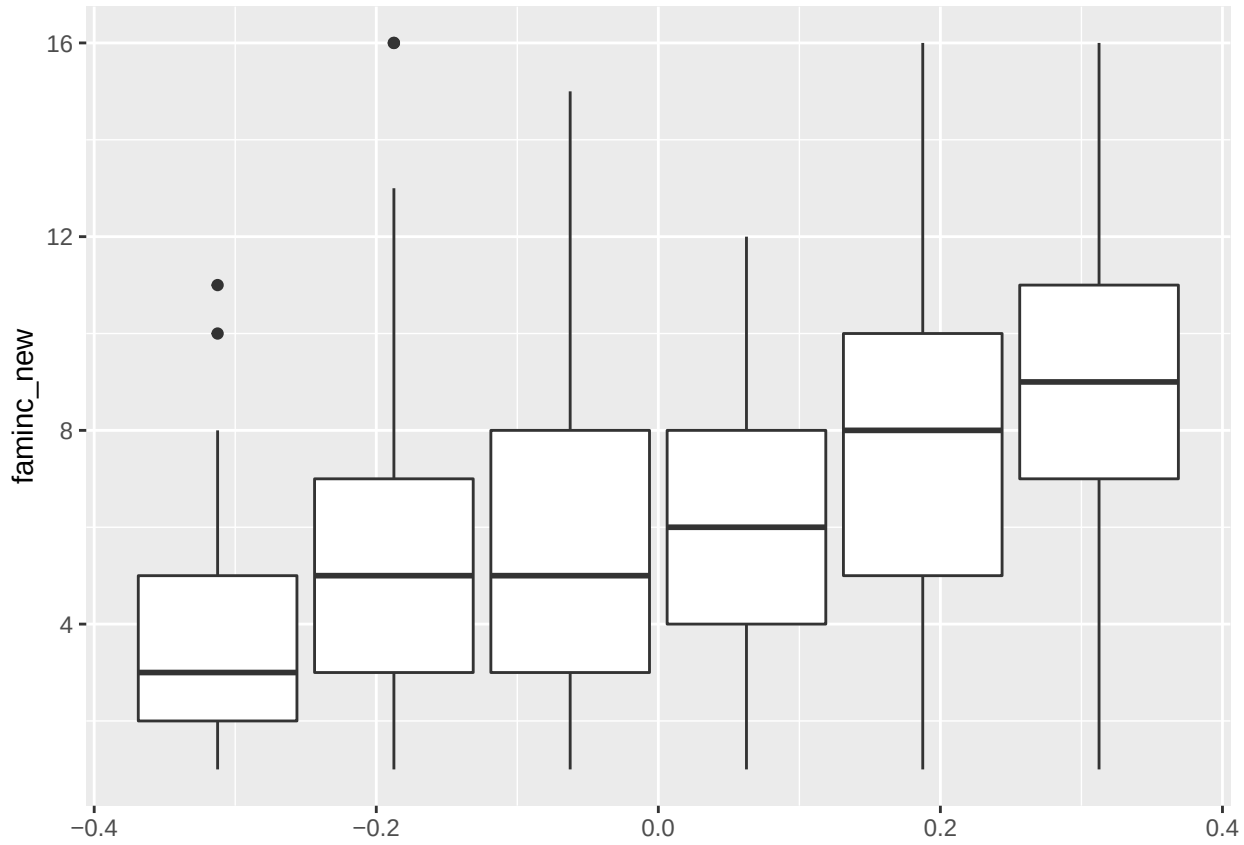
- Basic Boxplot

```
ggplot(data, aes(y = faminc_new))+
  geom_boxplot()
```



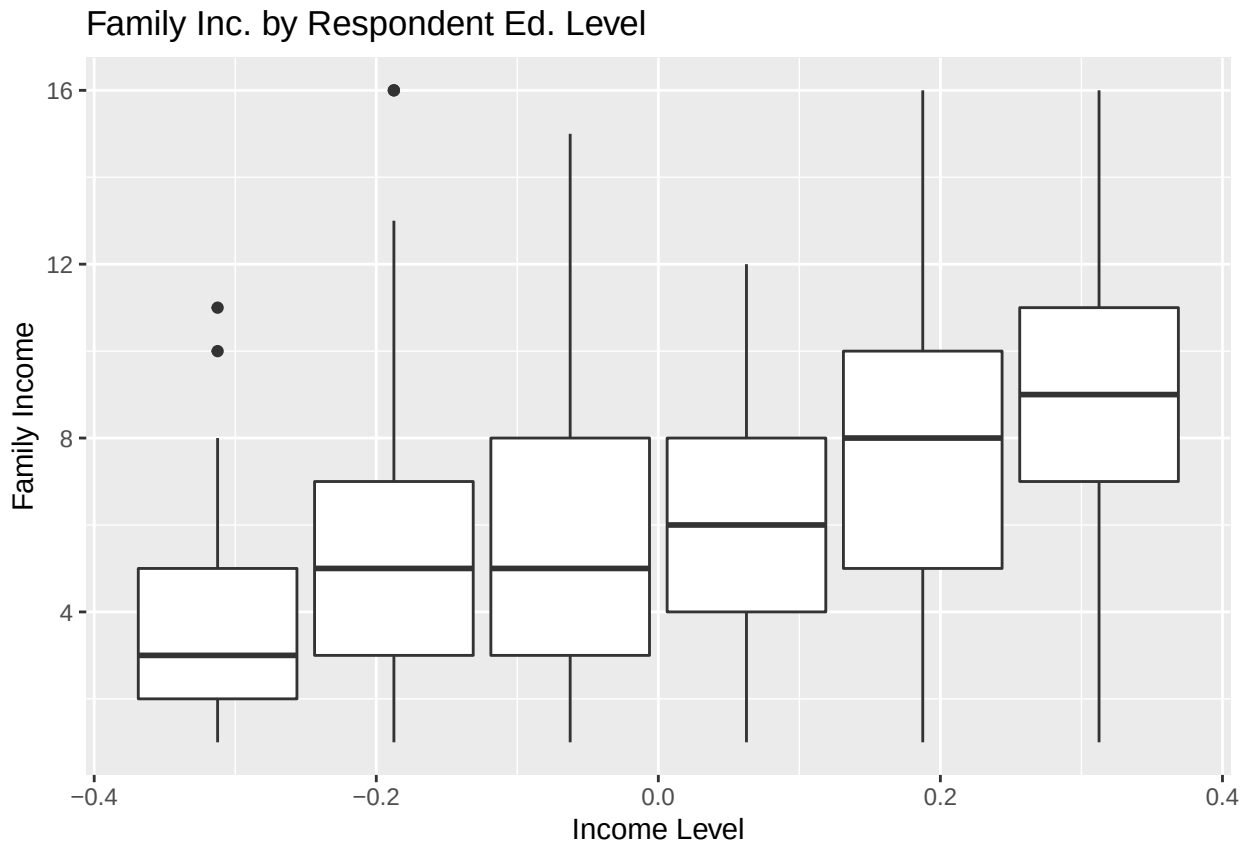
- Break up boxplots by education group – add an aesthetic mapping for group

```
ggplot(data, aes(y = faminc_new, group = educ))+
  geom_boxplot()
```



- Add labels and a title

```
ggplot(data, aes(y = faminc_new, group = educ))+
  geom_boxplot()+
  labs(title = "Family Inc. by Respondent Ed. Level", x = "Income Level", y = "Family Income")
```

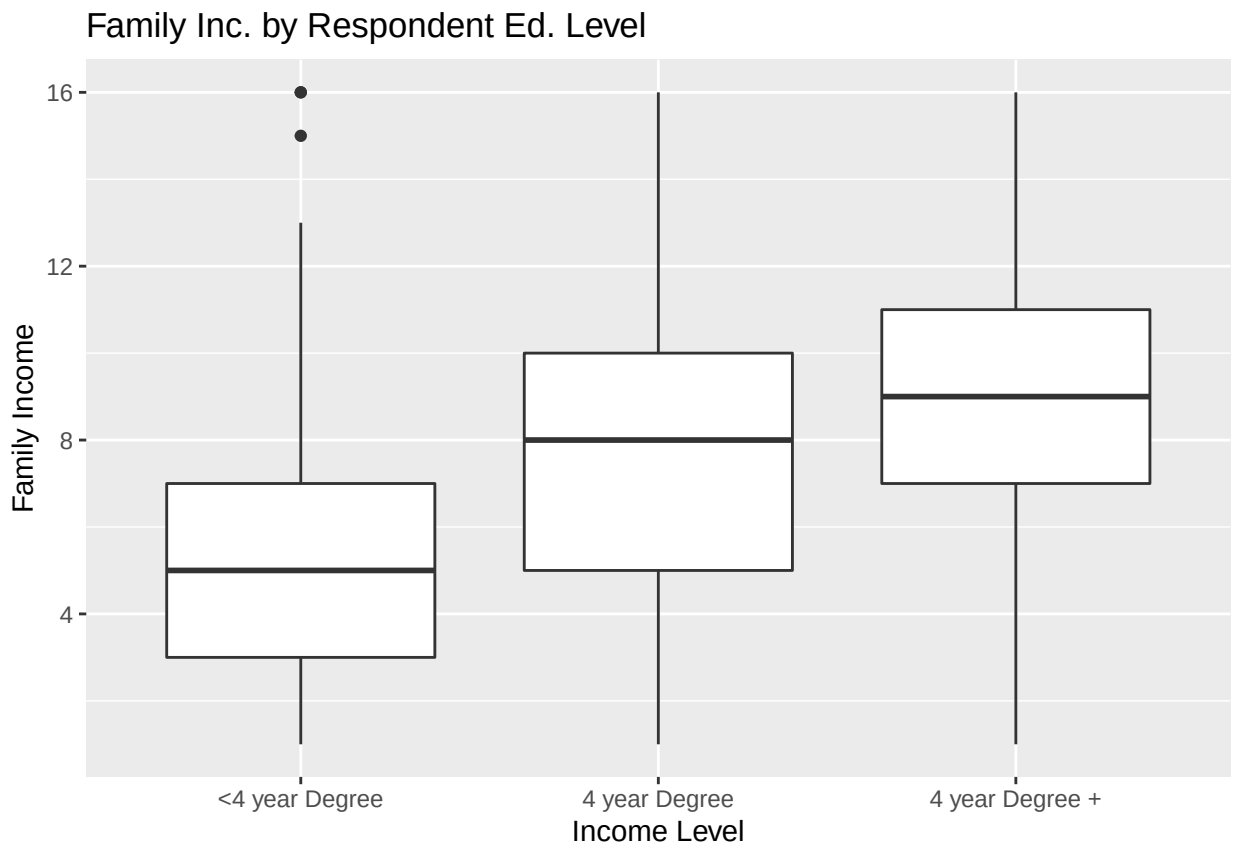


- Reformat the data to create a dichotomous categorical variable for four-year college grads or more, and then all respondents with 2 year college degree or less.

```
data$educ_category <- recode(data$educ, '1' = "<4 year Degree", '2' = "<4 year Degree", '3' = "<4 year

# Make sure you change the aesthetic mapping so the new categorical variable is maped to "x" rather tha

ggplot(data, aes(y = faminc_new, x = educ_category))+
  geom_boxplot()+
  labs(title = "Family Inc. by Respondent Ed. Level", x = "Income Level", y = "Family Income")
```

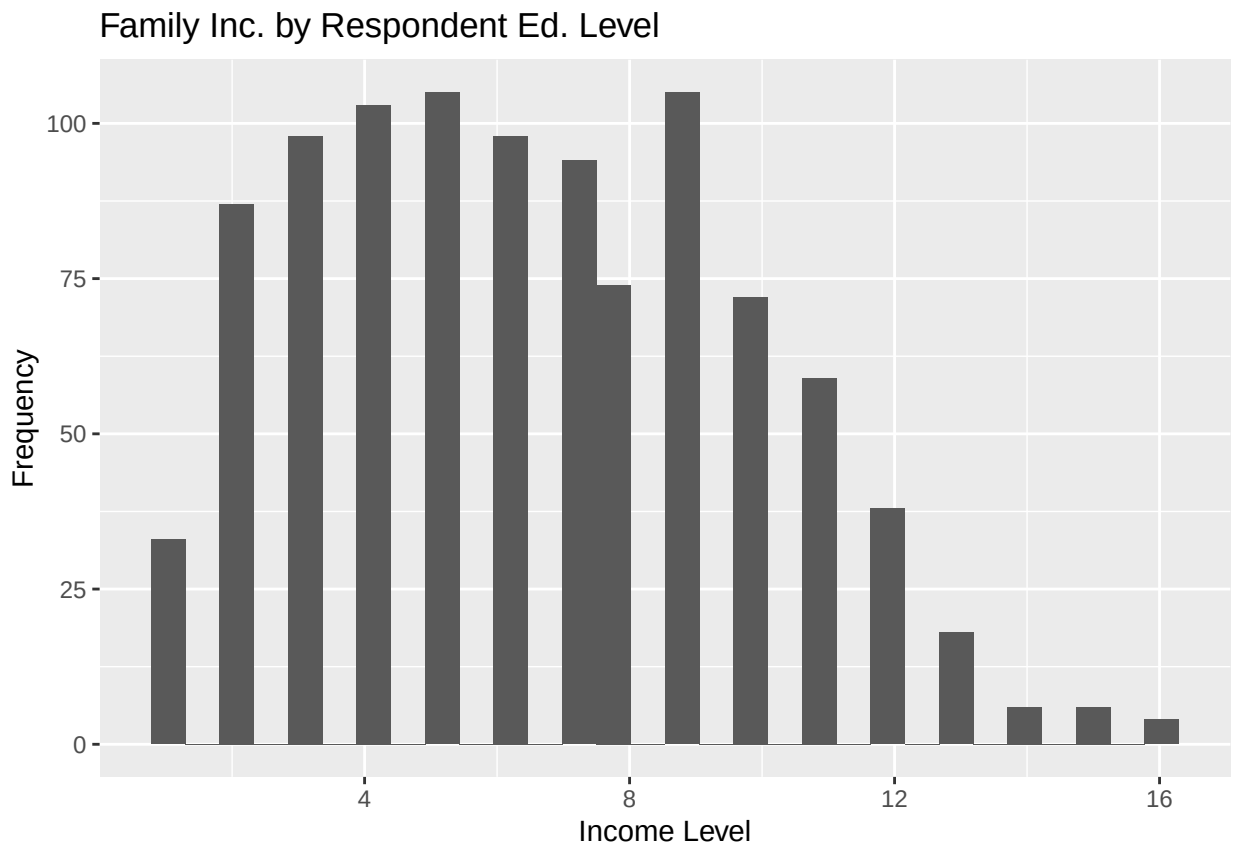


Histogram (Frequency Distribution Plot)

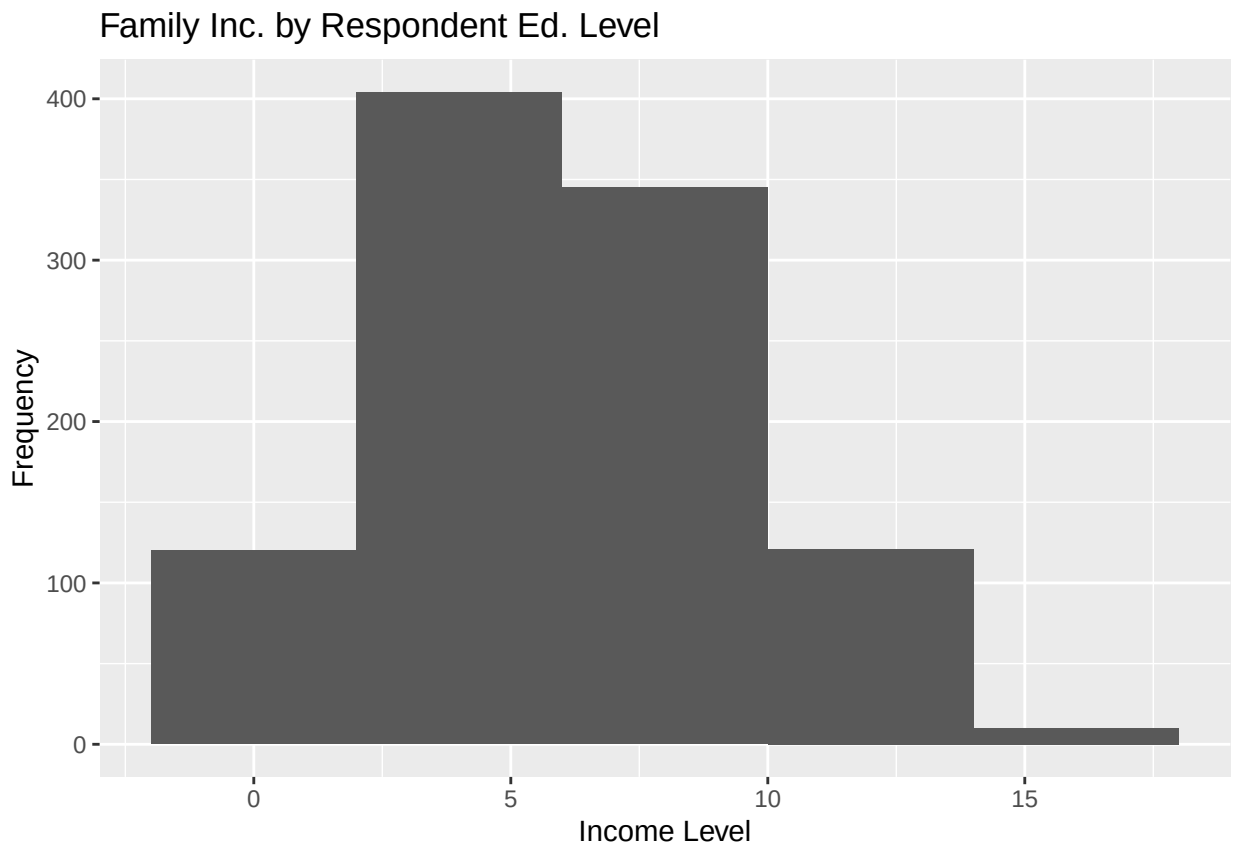
Reference: - https://r-graphics.org/recipe-quick-histogram - https://r-graphics.org/recipe-distribution-basic-hist#RECIPE-DISTRIBUTION-BASIC-HIST - https://r-graphics.org/recipe-distribution-multi-hist#RECIPE-DISTRIBUTION-MULTI-HIST

```
ggplot(data, aes(x = faminc_new))+
  geom_histogram()+
  labs(title = "Family Inc. by Respondent Ed. Level", x = "Income Level", y = "Frequency")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# with binwidth
ggplot(data, aes(x = faminc_new))+
  geom_histogram(binwidth = 4)+
  labs(title = "Family Inc. by Respondent Ed. Level", x = "Income Level", y = "Frequency")
```



Density Plot

Reference: - https://r-graphics.org/recipe-distribution-basic-density

```
ggplot(data, aes(x = faminc_new))+
  geom_density()+
  labs(title = "Family Inc. by Respondent Ed. Level", x = "Income Level", y = "Density")
```


Family Inc. by Respondent Ed. Level
0.100
0.075
0.050
0.025
0.000
Density
4
8
12
16
Income Level