GOVT 707 Lab 4, OLS Regression Part 1

Theodore Landsman

September 16, 2021

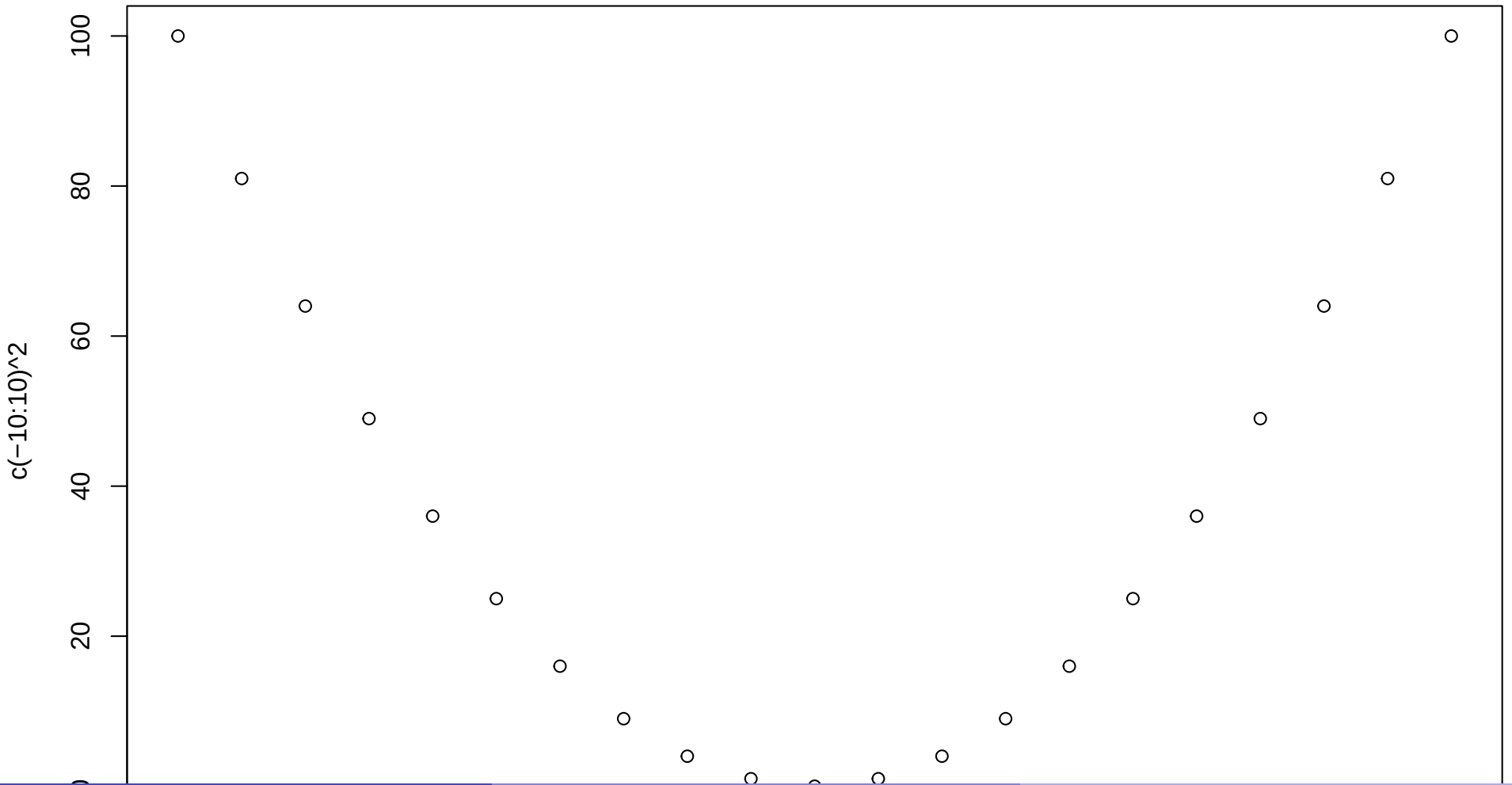
What is OLS Regression

- OLS stands for **O**rdinary **L**east **S**quares
- **Ordinary**: We are not doing any fancy manipulations.
- **Least**: We are minimizing something.
- **Squares**: The thing we are minimizing is a squared term.
- Why is it helpful to square things before taking the sum of them? What else could we do?

Squares

- Defined at 0.
- No discontinuity.
- Plays well with other mathematical operations.

```
facet_wrap(plot(c(-10:10),c(-10:10)^2),plot(c(-10:10), abs(c(-10:10)
```



How GGPlot Works

- Instead of single functions for different data visualizations, GGPlot is a unified 'grammar' for all data visualization tasks.

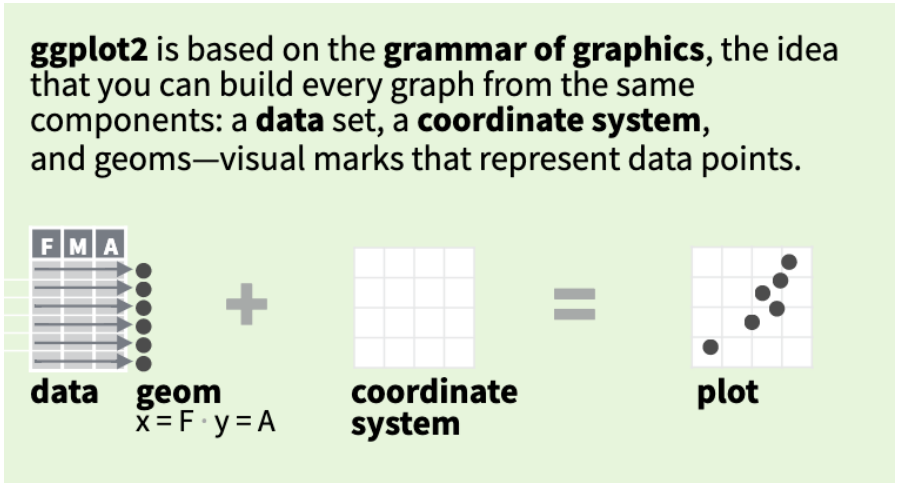
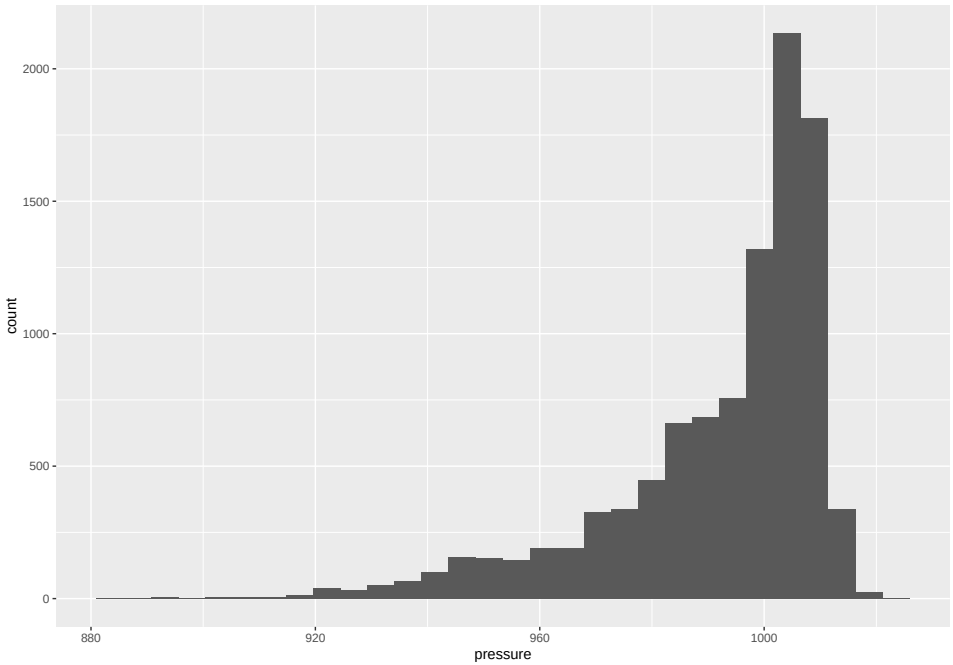


Figure 1: Snippet from ggplot2 cheat sheet

GGPlot In Practice

```
# Let's make the same plot we made before
ggplot(data = storms, aes(x = pressure)) +
  geom_histogram()
```



GGPlot: Origins

- GGPlot was created by Hadley Wickham, a statistician from New Zealand who is the Chief Scientist for RStudio and has taught as an adjunct at University of Auckland, Stanford, and Rice.

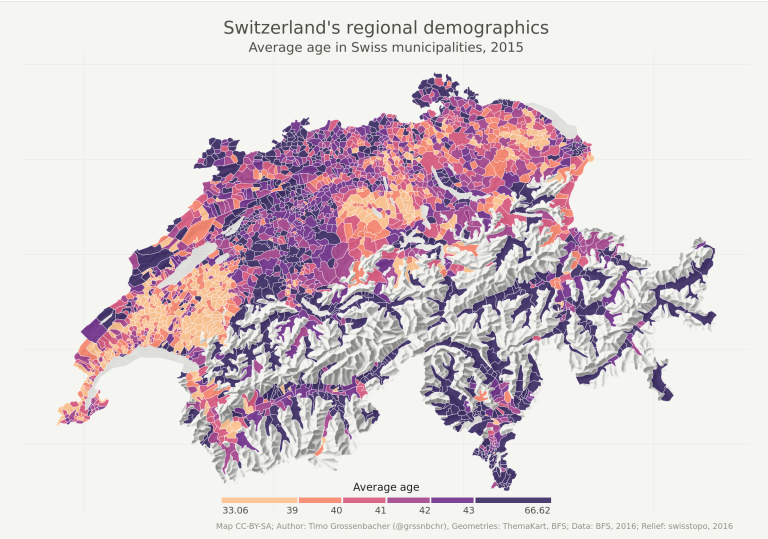
Figure 2: Haley Wickham Twitter Picture

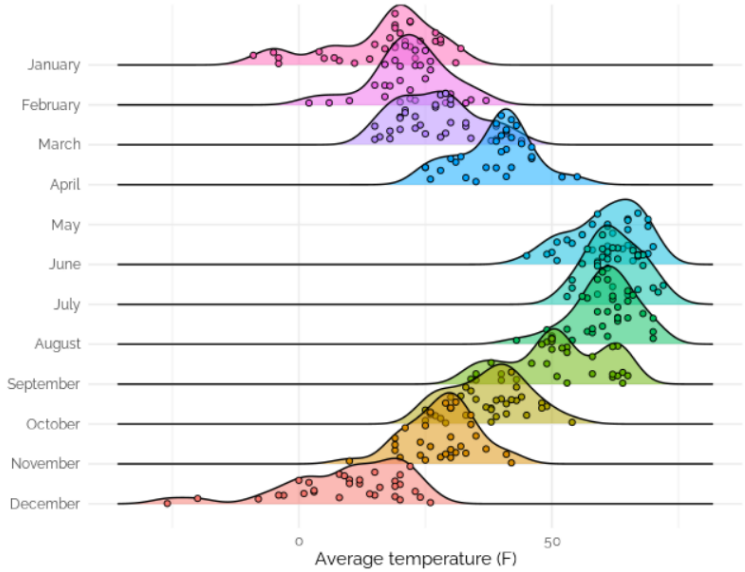
Why Use GGPlot?

- `ggplot()` is non-mandatory, don't use it if you don't want to!
- `ggplot` is a newer tool that can make you useful to professors as a TA or RA *because* they don't want to learn it.

Why use GGPlot Part 2

- `ggplot` dramatically expands the universe of types of graphics you can make.
- By having a unified system for specifying 'aesthetics' it makes the skills you learn building one kind of graphic relevant to building new kinds.





Why use GGPlot Caveats

- Base R plots tend to use *most* of the same syntax as each other as well. This means that you can get the benefits of a unified graphics system in base R too.
- `ggplot` is customizable to a degree that can feel overwhelming, it is sometimes easier to just create a basic plot in base R rather than thinking about all the extra things you could do with it in `ggplot`.

Exercises

1) Change the color scheme for one of the plots in the R file.
2) Create a boxplot for one of the variables in the R file with `boxplot()` or `ggplot() + geom_boxplot()`
3) Create a histogram for one of the vairables in the R file with `hist()` or `ggplot() + geom_histogram()`
4) Look for the `theme()` calls in the ggplot functions. Which theme do you like? Is there something about that theme you would change? Do you prefer how base R charts look?

```
-0.9292 + 5.7754 * -8.9861168

ev$elec_margin[ev$year == 2020] <- -0.9292 + 5.7754 * -8.9861168

# Plot the data with 2020 prediction

plot(ev$gdp_growth, ev$elec_margin, xlim = c(-10, 10), ylim = c(-55,
abline(ev_fit)
```