CS&SS 569 Visualizing Data and Models
Winter 2023
Section Syllabus

Brian Leung

Department of Political Science, UW

kpleung@uw.edu

Section	Office Hours
F 3:30-5:20 pm	Second half of section, or by appt
Zoom link	Zoom link

Overview. Sections are designed to complement the lectures, by focusing on the application of visual principles and techniques presented in lectures. In particular, we focus on their implementation in R. We will first cover the basics of R, RMarkdown and LaTeX that are fundamental to data wrangling, data analysis and scientific presentation. We will then unpack the inner working of two graphic packages, ggplot2 and tile, such that we will have a better command over a myriad of parameters in the graphic design, guided by proper visual principles instead of being dictated by the programming defaults. The rest of the sections will cover selected topics such as visualizing particular types of data (e.g. spatial data and relational data) and interactive visual display. The goal is for students to acquire the fundamental skills to solve a wide range of data visualization tasks and to develop a consistent workflow in both programming and scientific communication.

Programming Assistance. Our most preferred communication channel is Slack, which allows you to insert code block in your messages. It has the added benefit of facilitating knowledge spillover through peer discussion and mutual assistance. Please post your questions on Slack related to R programming, graphic packages, or debugging. When you post a question, the best practice is to create a "minimal, reproducible example," instead of taking a screenshot of a code snippet (see here and here). Alternatively, please feel free to come to my office hours for further

consultation, or email us (see below).

Section Resources. Consult the "Labs" section of Chris's website for all slides, codes and data used in sections.

Homework Submission. Please submit your homework in PDF. We recommend using RMarkdown to integrate plain text, graphic outputs and code chunks which can then be rendered ("knitted") into a single PDF output. You will have to submit your homework PDF on the Canvas course website (in assignments). You must upload your PDF file in the following format with no whitespaces:

> File name: CSSS569_HW1_BrianLeung

Section Schedule

There will be 9 sections in total:

Week	Topic
	Setting the Stage
1	~~R Boot Camp & Intro to RMarkdown~~
	Intermediate R and predictions
2	Intro to LaTeX with Overleaf
	Graphic Tools in R
3	Intro to Base R Graphics and ggplot2
4	Advanced ggplot2 and Extensions
5	Intro to tile
	Selected Topics (Open to Input)
6	Visualizing Spatial Data
7	Visualizing Relational/Network Data
8	Interactive Visual Display with R Shiny
9	ggplot2 extension packages for visualization

Section Outline

~~**Week 1: R Boot Camp and Intro to RMarkdown.** Provides an overview of the sections and logistics. Offers a boot camp to R: basic operators and functions, data types and structures, data import and (basic to intermediate) data wrangling in both Base R and tidyverse. Introduces R Markdown to "knit" compact, shareable, reproducible outputs.~~

Week 2: Intro to LaTeX with Overleaf. Introduces TeX engines and various text editors. Introduces basic LaTeX commands, environments and packages, especially for page set-up, inserting figures, and typesetting tables and formulas. Showcases useful templates and extensions, such as `TikZ`, `beamer`, and `BibTeX`.

Week 3: Intro to Base R Graphics and `ggplot2`. Introduces Base R graphics. Introduces basic `ggplot2`, with attention to various consequences of the interaction of `data = …`, `mapping = aes(…)`, `geom_{…}`, and `inherit.aes = TRUE`. Implements small multiple with `facet_grid()` and `facet_wrap()`.

Week 4: Advanced `ggplot2` and Extensions. Explores data wrangling techniques often employed before visualization, such as `pivot_longer()`, `pivot_wider()`, and `fct_reorder()`. Discusses various customization options: `theme`, `legend`, `label`, and `annotate`, with examples written for POLS/CS&SS 510. Introduces useful extension packages, such as `ggstance`, `ggrepel`, and `gghighlight`.

Week 5: Intro to `tile`. Introduces `simcf` and `tile`. Revisit the concept of grid graphics in R. Works through various lineplots, scatterplots, and ropeladders of Chris's examples. Highlights functions and arguments such as `rugTile()`, `ropeladder(subentryheight = …)`, and `(extrapolate = …)`.

Week 6 - 9: Selected Topics on Data Visualization. Introduces various graphic designs and packages to visualize spatial, relational, and text data. Introduces interactive visual display with R `Shiny`.

Additional Resources

Healy (2018) provides a beginner-friendly introduction to `R`, `R Markdown`, and visualization using `ggplot2`. Wilke (2019) offers a wide variety of visualization techniques with helpful advices to diagnose suboptimal graphic practices. His book has a free online version, with source code on `GitHub`.

If you don't have an existing data project, consult the `TidyTuesday` project (Mock, 2018), which offers many interesting raw data sets, coding examples by community members, and is being updated weekly.

References

Healy, Kieran. 2018. <u>Data Visualization: A Practical Introduction</u>. Princeton University Press. p.4

Mock, Thomas. 2018. "TidyTuesday - A Weekly Social Data Project in R.".
URL: *https://github.com/rfordatascience/tidytuesday* p.4

Robinson, David. 2016. "Why I use ggplot2.".
URL: *http://varianceexplained.org/r/why-I-use-ggplot2/*

Wilke, Claus O. 2019. <u>Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures</u>. O'Reilly Media.
URL: *https://serialmentor.com/dataviz/* p.4

Yau, Nathan. 2016. "Comparing ggplot2 and R Base Graphics.".
URL: *https://flowingdata.com/2016/03/22/comparing-ggplot2-and-r-base-graphics/*