

Sociol 232: Visualizing Social Data

Kieran Healy

Duke University

kieran.healy@duke.edu

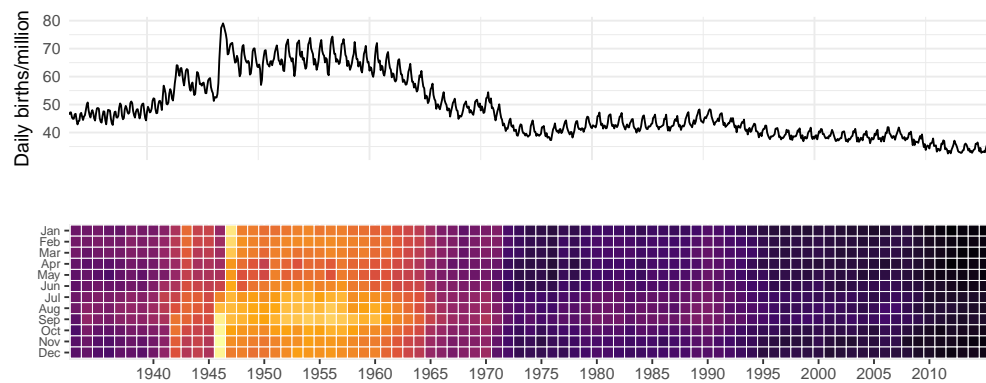


Figure 1: U.S. Births, 1933-2015.

- Instructor: [Kieran Healy](#), Sociology Department, 268 Reuben-Cooke.
- Time and Place: Wed/Fri at 10:05am-11:20am in Perkins LINK o88 (Classroom 4) from January 12th–April 24th, 2024.

About this course

This course will teach you how to use modern, widely-used tools to create insightful, beautiful, reproducible visualizations of social science data. You will also learn about the theory and practice of efforts to visualize social-scientific data, and society more generally. We will think about different ways of looking at data, about where social science data comes from in the first place, and about the implications of choosing to represent it in different ways.

By the end of the course you will

- Understand the basic principles behind effective data visualization.
- Know how to create a wide range of plots in R using `ggplot2`.
- Know a fair amount about how to use R for things other than data visualization.
- Have a good understanding of issues surrounding the collection and representation of data in the social sciences and beyond.

Core texts

I recommend (but do not require you buy) three books. Draft versions of all of them are available for free online.

- Kieran Healy, *Data Visualization: A Practical Introduction* (Princeton: Princeton University Press, 2019), <http://socviz.co/>. The print version can be purchased at [Amazon](#) and other bookshops.
- Hadley Wickham, Garrett Grolemund, and Mine Çetinkaya-Rundel, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, Second. (Sebastopol, CA: O'Reilly Media, 2023), <https://r4ds.hadley.nz>. The print version can be purchased at [Amazon](#) and other bookshops.
- Claus E. Wilke, *Fundamentals of Data Visualization* (Sebastopol, California: O'Reilly Media, 2019), <https://serialmentor.com/dataviz/>. The print version can be purchased at [Amazon](#) and other bookshops.

Software

We will do all of our visualization work in this class using R and use [RStudio](#) to manage our code and projects. R is a freely-available programming language that is designed for statistical computing and widely used across the natural and social sciences, as well as in the rapidly-growing world of “data science” generally. [RStudio](#) is an integrated development environment, or IDE, for R, a kind of control center from which you can manage the engine-room of R itself. It is also freely available. If you haven't used these

tools before, don't worry. The course does not presuppose any familiarity with them. We will get up and running with them during the first week.

Schedule

The [weekly schedule](#) can be viewed on its [own page](#), which has more details on readings, examples, and problem sets.

Week	Date	Topic
Week 1	- / Jan 12	Orientation
Week 2	Jan 17 / Jan 19	Make Some Graphs in R
Week 3	Jan 24 / Jan 26	Ways of Seeing
Week 4	Jan 31 / Feb 2	How ggplot Thinks
Week 5	Feb 7 / Feb 9	Show the Right Numbers
Week 6	Feb 14 / Feb 16	Expanding your Vocabulary
Midterm Assignment	- / -	Midterm Assignment
Week 7	Feb 21 / Feb 23	Counting People
Week 8	Feb 28 / Mar 2	Trends and Time Series
Week 9	Mar 6 / Mar 8	Maps and Spatial Data
Week 10	Mar 13 / Mar 15	Spring Break
Week 11	Mar 20 / Mar 22	Iteration and Missing Data
Week 12	Mar 27 / Mar 29	Text as Data
Week 13	Apr 3 / Apr 5	Social Networks
Week 14	Apr 10 / Apr 12	Project prep
Week 15	Apr 17 / Apr 19	Catch-up
Final Project	- / -	Final Project

Course policies

- Attendance is required, and important. I am a reasonable person; if you need to be absent please *let me know in advance* insofar as that is possible.
- Do the assigned readings in advance of class.
- Submit problem sets, or other assignments, on time.

Required work and grading

Three kinds of work are required: problem sets and class participation, a midterm project, and a final project.

- Weekly **Class Participation** and **Problem Sets** will let you reflect on the reading and practice your coding and visualization skills. Problem sets are due by end of day the *Monday* after they are assigned.
- A **Midterm Project**.
- A **Final Project**. There is no final exam.

Grade components: Problem Sets and Class Participation: 50% / Midterm Project 20% / Final Project 30%.

How you should approach this course

The material covered in the course has a lot of *continuity* and it is *cumulative*. You will be learning a set of practical skills. This means that techniques we learn early on will be necessary for understanding things that come later. It also means that regular practice will help you a lot. So, this is not a “Topic of the week” course where you can tune out for a few weeks while expecting to be able to easily drop back in later. The material we cover each week will not be overwhelming. If you participate during class and keep up with the weekly assignments you’ll be in a very strong position to do well in the class. If you don’t, it’ll be harder than you expected.

Duke community standard

Like all classes at the university, this course is conducted under the Duke Community Standard. Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and nonacademic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard you will not lie, cheat, or steal in academic endeavors; you will conduct yourself honorably in all your endeavors; and you will act if the Standard is compromised.