

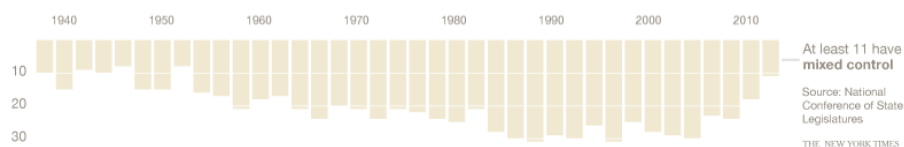


INTRODUCTION TO DATA ANALYSIS

The aim of the course is to show how to perform elementary data analysis in the social sciences. Every course session will show how to turn real-world data into analysable graphics, using essential statistics and computer software.

Students will be regularly asked to submit their own ideas and analysis on a wide range of social issues, including public health, economic trends and politics.

The course requires a laptop, Internet access and an interest in building both computer and statistical skills.



This syllabus covers only the essentials. Teaching material for the course sessions will be made available online during the course at <http://f.briatte.org/ida> (stay tuned!).

Part I: Computing

In the first part, we get to learn how to use statistical software to open a dataset. Your goal will be to explore your own choice of data and prepare it for further analysis.

- 1 Setup** Installing R and RStudio (both freeware)
- 2 Objects** Manipulating statistical objects
- 3 Math** Getting the computer to do it for you
- 4 Data** Opening and getting data

Part 2: Visualization

The second part covers methods to produce graphs and simple statistical models out of a wide range of data. Your goal will be to produce a preliminary analysis of your data.

- 5 Clusters** Finding groups within larger groups
- 6 Distributions** Graphing measures and proportions
- 7 Differences** Hypothesis testing for visual differences
- 8 Models** Statistical modelling for social data

Part 3: Extensions

The last part covers methods to produce graphs and simple statistical models out of a wide range of data. Your final goal will be to visualize your data under a new angle.

- 9 Time series** Visualization in time and trends
- 10 Maps** Spatial visualization
- 11 Networks** Visualization of human connections
- 12 Open data** Data in science and society

Handbooks (chapters assigned weekly)

- Winston Chang, R Graphics Cookbook (O'Reilly 2012)
- Robert Kabacoff, R in Action (Manning 2011)
- Paul Teetor, R Cookbook (O'Reilly 2011)

Readings (recommended but optional)

- Larry Pace, Beginning R (Apress 2012)
- Timothy Urdan, Statistics in Plain English (Routledge 2010)
- Nathan Yau, Visualize This (Wiley 2011)

Software (installed during first class)

- R <http://www.r-project.org/>
- RStudio <http://www.rstudio.com/>

Tutorials (search and find more help)

- Quick-R <http://www.statmethods.net/>
- R Seek <http://www.rseek.org/>
- R Tutorial <http://www.r-tutor.com/>

Links (used to share code and coursework)

- GitHub <http://www.github.com/>
- Gist <http://gist.github.com/>

The plots on the cover page are from a recent edition of the *New York Times*.
One of their graphics editors has explained how to quickly draw them with R:

<http://chartsnthings.tumblr.com/post/36978271916/r-tutorial-simple-charts>