Psephology books pdf

I'm not robot

reCAPTCHA

Continue

British Journal of Political Science Quoted by 4 4 Johnson, George W. 1979. Annual allowance 1978 p. 1. Kavanagh, Dennis 1982. In writing modern electoral history. Election Research, Tom 1, edition. 1, p. 117. Kavanagh, Dennis 1990. Politics and personalities. p. 1. Assibong, P. A. Wogu, I. A. P. Misra, Sanjay and Makplang, Daspan 2020. Advances in electrical and computer technologies. Tom 672, Edition. , p. 347. × 6, Issue 2, 1976 , p. 217-230 Cited by 4 4 Johnson, George W. 1979. Annual allowance 1978 p. 1. Kavanagh, Dennis 1982. In writing modern electoral history. Election Research, Tom 1, edition. 1, p. 117. Kavanagh, Dennis 1990. Politics and personalities. p. 1. Assibong, P. A. Wogu, I. A. P. Misra, Sanjay and Makplang, Daspan 2020. Advances in electrical and computer technologies. Tom 672, Edition. , p. 347. ×, Issue 2, 1976, p. 217-230 Usage data cannot currently be displayed. Psephology - an election study - is a field that was started by early political scientists such as R.B. McCallum and Alison Readman (who wrote the first Nuffield election survey in 1947), Herbert Nicholas (who wrote the second in 1950) and David Butler (who wrote or co-authored each study from 1951 to 2005). The word psephology comes from the Greek word pebble psephos, after the way citizens voted in Athens, using otherwise colored pebbles. Psephology began as an academic joke suggested by R.B. McCallum - but caught on. The first reference to psephology in this context was made by McCallum in Oxford Magazine in 1951, a review of the 1950 Nuffield election survey. David Butler was the first person to use it in the book. Butler's 1951 UK general election, published the following year, opened with a spirited justification for why a book about the election was even necessary - and was caught. While touring the United States in 1952, Butler found himself overwhelmed by journalists asking about a new word that had been coined. Today, psephology has expanded and diversified into its own discipline. Access to the content of a social journal varies depending on what we refer to them. If you have access to a journal through society or association membership, go to the society journal, select the article to view, and follow the instructions in this field. Please contact us if you experience any difficulties logging in. This article probably contains the original research. Please correct it by verifying your statements and adding inline quotes. Statements consisting solely of original tests should be deleted. (November 2010) (Learn how and when to delete this template message) Part of a Series of PolicyElections Basic Types By Election Corporate Direct/Intermediate Fixed-term General Local Plurality Primary Proportional Recall Snap Sortition Two-round System Terminology Anonymous Electorate Border Audits Boundary Demarcation Crossover Voting Electoral College Electoral Silence Psephology Initiative Secret Voting Electoral Subseries Political Party Voting Electoral Systems Electoral Systems Electoral Lists Elections by Country Next Elections Last Next Parliamentary Elections National and Local Elections 2020 Related Election Security Election Fraud Referendum (in the country) Political portalVTE Psephology (/s'fələdsi/; from the Greek ψῆφος, psephos, pebble) is a branch of political science, quantitative analysis of elections and voting. [1] As a result, psephology is trying to scientifically expose elections. Psephology uses historical voting data, opinion polls, campaign finance information, and similar statistics. The term was invented in 1948 in the United Kingdom by W. F. R. Hardie (1902-1990) after he was asked by his friend R. B. McCallum for a word describing an election study; first used in 1952. [2] Social choice theory is another area of study that studies voting from a mathematical point of view. Etymology The term draws from the Greek word for pebble, because the ancient Greeks used pebbles to vote. (Similarly, the word vote comes from the medieval French word ballotte, which means little ball.) [3] Psephology applications is a division of political science that deals with research as well as statistical analysis of elections and polls. People who practice psephology are called psephologists. Several of the main tools that are used by the psephologist are historical district voting data, campaign finance information, and other related data. Opinion polls also play an important role in psephology. Psephology also has different applications, in particular in the analysis of election results for current indicators, as opposed to predictive purposes. For example, the Gallagher index measures proportional representation in elections. Degrees in psephology are not offered (instead, a psephologist may have a degree in political science and/or statistics). Knowledge of demographics, statistical analysis and politics (especially electoral systems and electoral behaviour) are prerequisites for becoming a psephologist. Notable psephologists Main article: List of psephologists Famous psephologists include Antony Green; Malcolm Mackerras (who developed the Mackerel pendulum); Michael Barone, who has co-authored the Almanac of American Politics since 1972; David Andrews, who since 1973 has run the Canadian CTV network of analysis and the appointment of dozens of federal and election and referendums; Nate Silver, whose website FiveThirtyEight tracks American voting trends; Canadian Éric Grenier in threehundredeight.com; David Butler and Robert McKenzie, who co-created the swingometer; John Curtice, who has extensive experience in predicting elections in the UK; Charlie Cook, publisher of The Cook Political Report; Thomas Ferguson, for the investment theory of party competition; Indian Academic Sekhar; William Bowe of Australian Poll Bludger; Curtis Gans, author Attendance in the United States, 1788-2009; [5] and Michael Gallagher, who developed the Gallagher index. See also British Polling Council Electoral Calculus Electoral geography Opinion poll Political Analyst Political Analyst Prediction Swing (Politics) Types of Democracy Reference ^ Lansford, Tom (2011). Kurian, George Thomas (ed.). Encyclopedia of Political Science. 1–5. CQ Press. p. 1377. ISBN 978-1-933116-44-0. ↑ Chapter 15: British Psephology 1945-2001: Reflections on the History of the Nuffield Election, David Butler, Even More Adventures with Britannia: Personalities, Politics and Culture in The United Kingdom. William Roger Louis (ed.), Harry Ranson Humanities Research Centre, University of Texas, 2003 ^ Stephan, Annelisa (November 6, 2012). Voting with the ancient Greeks. Iris. ^ Green, Antony. Election blog. Abc. ^ Gans, Curtis (2010). Voter turnout in the United States, 1788-2009. CQ Press. 978-1604265958. Continue reading by William Safire. New Political Dictionary, Random House, New York 1993. External links Search psephology in Wiktionary, a free dictionary. Psephos Dr Adam Carr's Elections Archive International IDEA - International Organization providing (among other things) statistical analysis of elections and electoral systems ACE Project – Information resources for electoral design and administration. It contains comparative data on elections and electoral systems Downloaded from is an author's version of the work. Please cite as: Arzheimer, Kai. Psephology and Technology, or: The Rise and Rise of the Script-Kiddie. SAGE manual of electoral behavior. Eds. Arzheimer, Kai, Jocelyn Evans and Michael Lewis-Beck. Los Angeles: The Sing, 2017. 972-995. [BibTeX] [Download PDF] [HTML] @InCollection{arzheimer-2016, editor = {Arzheimer, Kai and Evans, Jocelyn and Lewis-Beck, Michael}, booktitle = {The SAGE Handbook of Electoral Behaviour}, publisher = {Sage}, author = {Arzheimer, Kai}, title = {Psephology and Technology, or: The Rise and Rise of the Script-Kiddie}, year = {2017}, pages = {972-995}, html = { , url = { , address = {Los Angeles} }1 Introduction Download this text as PDF From the very beginning ps ephology has been at the forefront of the methodology and has sometimes shifted its methodology (see, for example, methods such as factor analysis or logistic regression, which were considered advanced in the 1990s, are now part of many MA programmes and even BA. With the proliferation of fresh data and the availability of ever faster computers and more advanced yet more user-friendly software, the pace of technological progress has once again accelerated over the past decade. Therefore, and paradoxically, this chapter cannot hope for the final Art: By the time this book will ship, the chapter will no longer be out of date. Instead, it seeks to identify important trends that have emerged over the past 15 years, as well as the likely trajectories of future research. More specifically, the next section (2) discusses the overall impact of open movements on electoral surveys. Section 3 is devoted to new (ish) statistical methods – easily implemented in open source software – which are necessary by the availability of data – often available in comparably open data distribution models – which are designed in new ways. Section 4 is ground for software tools that have been developed in the open source community and which are currently under-used in electoral surveys, while the penultimate section discusses the dual role of the Internet as both infrastructure and the subject of electoral research. Section 6 summarises the main points. Photo binaryape 2 Open Source, Open Data, Open ScienceLike many other subfields in social sciences, psephology has a big impact on the rapid advances in computer and information technology. The two most important changes in this area are the double open source and open-data revolutions. Open source software has its roots in the free software movement of the 1980s (Lakhani and Hippel, 2003), a revolt against increasingly restrictive software licenses that, among other things, were designed to patent algorithms and prohibited reverse engineering of software installed on private computers. Supporters of the free software movement, on the other hand, made their software available for free (free as in free beer) and gave everyone and everyone a license to modify their programs at their discretion (for free as in free speech), which required opening the source code. The spread of the internet in the 1990s subsequently facilitated large-scale collaboration on free software projects and led to the current idea of open source software, which is included in Raymond's manifesto (1999) Chair and Bazaar, which highlights the idea of distributed and only loosely coordinated teams as a strategy for rapid and effective development. While the free software movement has had some anti-establishment bent, many of the largest and most successful open source projects, such as the Linux operating system, apache web server or Firefox series that collectively power most of the current Internet, are happy to rely on the support of corporate supporters who donate money, resources and time to some of their employees. In other cases, large companies have even created open social versions of their existing programs or designed them as open source applications in the first place (Google Android operating system). Companies can do this to raise their profile, or to attract the best software engineers for commercial projects, but two other themes are more interesting: They may want to use an open open software instead of a closed alternative source to generate and deliver their own products (e.g. technology companies using Linux to run server farms) or may offer an open source software-based service (professional support or hosted versions). Either way, corporate support for open source makes commercial sense - carefully illustrating Olson's argument (1965) about big players rationally investing in public goods - because open source is, as Raymond suggests, a highly efficient model for organizing large projects: it almost immediately turns feedback from the user base and turns the most capable and committed users into programmers. Open source is very important for psephologist not only because it helped build a lot of internet infrastructure and some key tools – R (Ihaka and Gentleman, 1996; Crawley, 2013), Python (Lutz, 2013) and plenty of others – but also because it has become a template for other open revolutions that affect election research. In the broadest sense, open data refers to the idea that research data or data that could be used for research should be as accessible as possible. As such, it's old news. In quantitative social sciences, data archives such as the Roper Center (or Michigan's Survey Research Center (, which collect, archive, and distribute existing data for secondary analysis, were created in the late 1940s. Patterns of cooperation and exchange of (European) archives were formalised with the creation of the Council of European Archives of Social Sciences Data (CESSDA, in the 1970s (Karvonen and Ryssevik, 2001, p. 45). In the public sector, it can be argued that the practice of frequently publishing detailed information on key national statistics, which were already well established at the end of the 19th century, is the beginning of open data. But it was the key components of the open source revolution – transparency, active user base engagement and near-zero marginal transaction costs – that began to transform data production and use in unprecedented ways in 2000. While access is not restricted for data protection reasons, researchers no longer have to travel to a data archive to use a given dataset, and physical media have been abolished for data distribution. Governments, large-scale research projects and individual scholars are now opening up their raw data for download. Some agencies and some of the largest internet companies (e.g. the Open Data Revolution have caused some new problems. Although the datasets available for research are growing exponentially, scientists still need to know where and how and the lack of a central repository and common interfaces seriously hampers progress. For data to be useful, it must be stored, and more importantly, described and licensed in standardized ways that make it accessible and recoverable in the medium and long term. This, in turn, requires institutions that can be trusted and need funding. In addition, there is increasing pressure on scientists to open up their own data. Currently, research councils regularly give evidence of research data, and even the publication of results in open access to them is a prerequisite for funding. Similarly, more and more journals require that not only the dataset itself, but also the program code that generates tables and graphs, be published with the final article in a repository (see section 5.1).1 While such rules reinforce traditional scientific standards of fairness, transparency and reproducibility, many researchers remain concerned that they will be scraped if they are forced to disclose their data and methods at the start of a new project. Probably due to the dominant incentive structure, few sociologists currently adhere to the open source mantra of release early, slow down frequently. Others, however, embrace the ideal of (more) open science by publishing work in progress on their personal home pages, opening draft chapters on social networks for researchers, and even moving their data and manuscripts to open source development sites such as Github, which could theoretically provide an ideal environment for scientific collaboration.3 Data, statistical models and complex software3.1 data structures and statistical modelspure formal theoretical exercises and simulations aside , all electoral surveys are based on data: a set of systematic and usually quantitative observations that can be used to test assumptions about how citizens, politicians, organised interests and the media interact and thus influence political decisions. Early studies highlighted the importance of macro factors (Zygfrid, 1913) and clustered sampling methods and mixed methods (Lazarsfeld, Berelson and Gaudet, 1944), the lasting influence of the American electorate (Campbell et al., 1960) prompted many scientists to focus on micro-level data from nationally representative samples of mass audiences for much of the 1960s, 1970s and 1980s. But the theory suggests that reliable (or at least reliable) accounts of human behavior must include not only the individual (micro) but also the social (macro) level, and the best different mezo layers and structures in between (see Coleman, 1994 for the overall line of reasoning, and Miller and Shanks, 1996 for use for election research). Finally, that argument led to a renewed interest in contextual variables and their effects (Jennings, 2007, p. 35 to 38). Since the late 1990s, constituency codes in their datasets. Using this information, it is possible to match the data on individual respondents to government data on the economy, migration and all the many other variables that can 1Pre-registration, a procedure that is becoming increasingly prevalent in natural sciences and whose adoption in political science is currently under discussion (Monogan, 2015), goes a step further by demanding that researchers present sampling plans, outlines of planned analyses and sham reports to the journal , which are reviewed before they even start collecting new data.reliably affect voting behaviour, while multi-level regression (see Chapter 47) is a convenient tool for estimating the magnitude of alleged effects and related standard errors. Supplementing information at the micro-contextual level leads to nested data in which each level 1 unit (respondent) belongs to one (and only one) district of a Level 2 (electoral) unit. Each Level 2 unit can in turn be part of one (and only one) Level 3 unit (say provinces), resulting in a tree-like structure. Multi-level regression modelling with contextual covariates from official sources has become an almost real standard for analysing both large-scale comparative datasets (see Chapter 48) and case studies of countries for which data are available at national level. Although this technique provides asymptotically correct standard errors and opens up a number of flexible modeling options (see section 3.2.1), it is not a panacea. When nations are the right contexts, their numbers are often too low for multi-level modeling (Stegmueller, 2013), and one might ask if it makes sense at all to treat countries as if they were a random sample from a larger population (Western and Jackman, 1994). Comparing political behaviour in sub-national units across countries is more informative and often more appropriate, but it also suffers from specific constraints: even in the complex and comprehensive Nomenclature of Territorial Units statistics (NUTS, see Eurostat, 2015), subnational units to be at the same level can vary considerably in size. , population and political, social and cultural significance.2 Moreover, the inclusion of government statistics as regressors in the multi-level model does not exhaust almost the complexity of the data that is currently available for analysis. Building on previous work by Lazarsfeld and Menzel (1961), Hox (2010) developed a useful typology that explains the possibilities. At each level, there are global variables that reflect the inherent properties of objects at the appropriate level. They are inherent in the extent that they cannot be constructed by aggregating functions or by disaggregating higher-level context functions. Traditional (statistical) models of electoral behaviour focused on global variables at for Democrats is associated with the electorate in question is a woman, 2NUTS-1 corresponds to 16 powerful federal states in Germany, groups of provinces, states or communities that have been grouped for purely statistical purposes in Austria, Spain and the Netherlands and do not exist at all in many smaller countries (e.g. Croatia, Denmark, Luxembourg or Slovenia). The lower-level NUTS-2 level is equivalent to the federal states in Austria, the autonomous communities in Spain, the regions of France and the provinces of the Netherlands that have their own elected assemblies. In other countries, such as Bulgaria, Finland, Germany, Romania or Slovenia, NUTS-2 areas exist solely for the purpose of national planning and attracting EU funds, and citizens will not be aware of their existence. Similarly, NUTS-3 can be a county (Germany), a group of counties (Austria), a province (Denmark, Spain, Italy), a region (Finland), a statistical region (Slovenia), an island (Malta), or it may not even exist (Cyprus, Luxembourg). Level 1 2 3 . . . Global relational → variable → ← contextual ← → regulatory → contextual contextual ← → contextual → ← →: ← aggregation: Disaggregation Source: Adapted from Hox (2010, p. 2) Figure 1: Typology of complex data structures and identifying it as a Democrat. A prototype multi-level model would add the unemployment rate and ethnic composition of the constituency as level 2 regressors. These are analytical variables that are created by aggregating the global characteristics of lower-level entities to create averages, coefficients, or higher-level percentages. Consequently, these variables can enter the model simultaneously on multiple levels (see section 3.2.1). Other district properties can also be significant additions to the model, but cannot be understood as individual-level feature aggregation or disaggregation of higher-level characteristics, and are therefore global variables at the district level. The gender and political experience of the main candidates are cases at point. Since there is no variable at the lowest level that suits them, they are strictly contextual for individual voters and can only enter the model once, at the top level. Finally, relational data conveys information about the relationship (e.g. presence and intensity of face-to-face contacts) between objects at the same level. Such network data is crucial for any micro sociological expatriation of electoral behavior: Of course, the person who is the center of the democratic clique of friends is more likely to vote and vote according to his or her peers than someone who is socially isolated. Like global/analytical variables, network data can simultaneously enter a multilevel model at multiple levels: Information about the relationships between individual voters in a district can be aggregated to create variables on the upper upper for example, comparing districts with dense/rare or homogeneous/fragmented communication networks. Network data is theoretically extremely attractive. But they introduce an additional level of complexity and require specialised statistical methods, as it by definition involves two actors (see section 3.2.3). In addition, relational data collection requires specific sampling plans (clusters), as a large number of network members must be examined to assess the properties of the network itself. This, in turn, raises questions of representativeness, data confidentiality and cost-effectiveness and runs counter to the dogma of the country-wide sample. Election polls sometimes contain elements relating to so-called egocentric networks, e.g . But this information will be biased by the respondent's perception and provide only fractional insight into the full network, since usually even the links between the respondent's direct contacts cannot be reliably recovered. As an easily accessible alternative, students of electoral behavior now turn to social media, where you can easily taste large and mostly complete networks of political communication. How well these networks are based on offline behavior and the voting population as a whole is another question. Either way, statistical procedures for social network analysis are currently in the process of becoming part of the electoral survey toolkit. In addition to multi-level and network data, the use of spatial data or georeferenders is another emerging trend in election surveys. A geographic reference is simply a set of coordinates that locate an object in space. Coordinates can define a point or area (polygon). In the simplest sense, the territorial identifiers mentioned above record that the voter lives in a given (usually large) area and are therefore also geo-references. However, more precise voter coordinates (e.g. census blocks, postcode segments, constituencies, street addresses, and even GPS readings) allow researchers to locate voters in much smaller contexts for which census and market research data are available – in other words, global and analytical variables that can be integrated into a selected multi-level model. While many researchers are familiar with the idea of thick geographic references, the availability of highly detailed data as well as increasing awareness of spatial dependencies requires specialized software and models for proper analysis of georefereration data (see section 3.2.4)3.2 Statistical techniques and software implementations3.2.1 Multi-level models and structural equation modelsAs described above, students of electoral behaviour routinely collect data that reflects basic theoretical explanations, have complex structures. Statistical multi-level models, also known as mixed models or random factor models, are the most appropriate means of dealing with such data. 7Perfections of unmeasured interference in a given context and thus provide valid standard errors for the effects of variables at the macro level. In addition, they model context-specific interference in the most efficient way possible, treating it as random. This is best illustrated by an example: In a survey of N voters living in constituencies K, which aims to explain individual turnout, you can try to capture the effects of unmeasured variables at the district level (say local social capital) by introducing district-specific captures (fictitious variables). However, this strategy has negative consequences for model identification and becomes inefficient and impractical very quickly as the number of sampling districts increases (Steenbergen and Jones, 2002). The statistical multilevel model will override K-1 estimates for local captures with a single estimate of their variability over circles (random capture), thereby significantly reducing the number of parameters. In addition, multi-level models also provide a number of additional advanced modeling options. If there are good reasons to believe that the impact of the explanatory variable (say the ideology measured by left-right self-placement) on turnout will vary significantly in different K-circles, the analyst can determine a random effect for that variable that complements the estimate of the average impact of ideology (traditional point estimation) with an estimate of its volatility. As the name suggests, random effects are appropriate if the variability of the impact of an independent variable can probably be treated as random. On the other hand, if the impact of a variable changes systematically, this can be modeled by defining cross-level interactions, e.g. in the case of cross-border interactions. Interactions between levels do not have to be limited to variables that are as conceptually different as the two in this example. On the contrary, the theory often suggests that a variable such as unemployment can in fact interact with itself, albeit at different levels, and thus triples the model: as an individual characteristic (global variable at micro level), as an analytical variable (unemployment rate at the district level) and as a cross-level interaction between them. A high unemployment rate can reduce the propensity to participate in elections for all citizens, and individual unemployment status tends to reduce turnout in an even stronger way. But this micro-level effect may be limited to low-unemployment environments, while individual unemployment may not have such a negative impact or even increase the likelihood of in districts where high unemployment rates attract community organisers and other political entrepreneurs. Multi-level models are ideal for disconnecting such complex causal relationships. They can also deal with complex political contexts that can have multiple levels (household voters within constituencies in 8within districts of provinces . . .) and which can cross and overlap instead of creating a neat, tree-like hierarchy: The electorate not only influences the characteristics of the constituency in which it lives, but is politically socialized in a completely different environment. While multi-level models can accommodate such complex structures, convergence will generally be slow and estimates may be volatile. As with all other aspects of modelling, analysts should therefore aim for parsimony. If there are variables at higher levels and if the goal is simply to reflect the multi-step nature of the sampling process, traditional test estimators or even Huber-White standard errors that constitute clustering can provide a quick and reliable alternative to a fully defined multi-level model. Nevertheless, multi-level models are a very flexible tool because contexts do not need to be defined in spatial terms. For the analysis of panel data, it is often worth thinking of individual respondents as contexts for interviews conducted in subsequent panel waves. Especially when panel data is unbalanced or collected at irregular intervals, multi-level models can alleviate some of the problems that plague traditional approaches to panel data. Another statistical technique that has become essential for students' electoral behaviour is modelling of structural equations (SEM). SEM is an extension of traditional factor analysis that allows scientists to determine multi-point measurement models for otherwise unobservable (= latent) theoretical structures, such as political postures. It is attractive because it can simultaneously estimate coefficients for entire systems of equations, and also because it can contain measurement models for attitudinal variables, which are relatively unreliable indicators. If measurement models hold down, SEM can also provide unbiased estimates of equally unobserved, structural relationships between latent variables. Taking into account the relevant data, it is possible to map the entire system of structures and hypotheses about their relationship to an equivalent system of equations. In the past, its use in electoral surveys was somewhat limited by the fact that they required continuous-scale measurements, which were spread out in principle multidimensional, while the key dependent variable in the polls, as well as many important independent variables, are categorical and usually spread with considerable incile. In the 1990s, the In addition, generalizations of the original model allow for the indication of ordinal and nominal indicator variables and even for categorical latent variables (Jöreskog, 1990; Jöreskog, 1994; Muthén, 1979; Muthén, 2002). In addition, multi-level models and models of structural equations are closely interlinked (Muthén, 2002; Skrondal and Rabe-Hesketh, 2004) and can be combined to create multi-level models of structural equations. 9Untilingly, estimating multi-level or structural equation models required specialized (if relatively user-friendly) software: HLM or MLWin for multi-level modeling, LISREL, EQS or AMOS for SEM and MPlus for both. That's no longer true: The latest versions of Stats - currently the most popular general-purpose statistical package in political science - can estimate all but the most complex multi-level and structured models of equations and so greatly expand the potential user base of these techniques. SPSS, another popular package, has some multi-level capabilities and works closely with AMOS, the SEM software that SPSS Inc. has. acquired in 2003 before it was bought by IBM in 2009. Perhaps more important is that there are packages available for the R programming language that provide similar features: Lme4 and Rstan for multi-level modeling, and Lavaan and Sem for SEM. While they may be slightly less capable, slower and generally more clumsy than commercial software, they are, like any other R package and core language itself, open source and freely available for almost any combination of hardware and operating system. Moreover, while there may be a lack of professional documentation and customer service, they are supported by a global enthusiast community, scripted in a full-fledged programming language with flexible data structures and linked to an ever-growing ecosystem of more than 6,000 user-written packages for R, aimed at implementing the latest developments in statistics.3.2.2 Bayesian methodsThe larger voters' researchers have been trained in frequent statistical reasoning that are based on the idea of a random sampling process , which could be repeated endlessly under essentially identical conditions. Until now, they have shown only little interest in the (sometimes exaggerated) benefits of an alternative statistical framework: Bayesian statistics (Jackman, 2004). There are at least two reasons for this inertia: the tachy-like paradigm is very similar to the pattern of large-scale national sampling of the general population, which has been the workhorse of electoral surveys for most of the last seven decades, and in such large samples Bayesian estimates and are often very similar to each other. But whether researchers like it or not, the increasingly popular multi-level models are bayesian models (Gelman and Hill, 2007). While many political scientists still have some as to the basic paradigm (or may be blissfully unaware of it), Bayesian Bayesian are still going into electoral research. There are many reasons for this. First, Bayesian models can sometimes be adapted to a problem for which no ready-made solution has been implemented in any statistical package. Examples are models that aim to predict the distribution of seats in parliament on the basis of a series of published opinion polls. Secondly, Bayesian 10statistics may be able to provide an estimator that is better in terms of bias and performance than any common alternative, as is the case with multi-level models and some SEM models. Third, Bayesian statistics, which for most of their existence were rather arcane past due to implied computational requirements, only gained practical significance for the researchers used with the dual emergence of simulation-based methods and affordable high-speed processors between 1990 and 2000. Even a decade ago, getting Bayesian estimates in MLWin for a fairly complex multi-level model could easily take an hour or more on a then modern desktop computer, as was the case with SEM in the 1990s. At the moment, most bayesian estimates still require access to specialized software (Winbugs, Openbugs, Jags, Stan . . .), preferably by R. However, the implementation of Bayesian analysis in recent editions of Staty (from version 14 on) may be a breakthrough in this regard.3.2.3 NetworksSo far away, election research most often lacked on the renaissance of social network analysis (SNA) in political science (for some notable exceptions see e.g. McClurg, 2006). Although interest in relational or network data has increased sharply in political science, psephology was somewhat late to the party because the relevant data is not publicly available. While large societies can display the characteristics of a small world network where everyone is related to everyone through a relatively small number of contacts (say six), such network structures are very rare and rarely affect political behaviour. Social embedding certainly plays a role in shaping opinion and political behaviour, but research into mainstream elections cannot hope to uncover the right networks. Traditional social research, as well as research of online communities, on the other hand, can do this. While it is not clear whether and how the results obtained here will generalize to the entire electorate, the statistical procedures for analysing social networks are currently in the process of being included in the electoral survey toolkit. Understanding these methods can be a huge challenge. By definition, network data interrupts the form of traditional data analysis, where cases correspond to rows of a data matrix and variables to its columns. In network applications, cases create both rows and columns of the (adjacent) data matrix whose cells represent existence, and, if necessary, the strength of the links Them. Registering traditional variables requires a second data matrix, specialized software, and more importantly, a customized mental analyst setting. Once collected, data on the links between entities may be used to calculate three broad 11 classes of statistical measures (Knoke and Yang, 2008): indicators reflecting the position of the entity in the local or global network (e.g. centrality), measures relating to the characteristics of the actual or potential link between the two entities (e.g. the importance of this link for network cohesion as a whole) and statistics describing certain characteristics of the network as a whole (e.g. degree of , in which it resembles the small world scenario described above). Often, the SNA target is predominantly descriptive and the analysis would result in their calculation and interpretation, but in principle all network measures can then be used as dependent or independent variables as part of a regression. Relational data does not match the single-mast paradigm of general statistical packages such as Stata or SPSS. What's more, there was little interest in SNA before the rise of social media. Therefore, most of the software, which is potentially useful for students of election behavior, is developed by researchers (often as an open source project) and available for free or at a very modest price. Historically, UCINET, which was created in the early 1980s and has been in continuous development since then was a very popular choice. UCINET is based on the tradition of (mathematical) sociology and contains many procedures for manipulating and analyzing relational data. However, according to its authors, many of these procedures become tediously slow on networks with more than 5,000 nodes. Pajek and Pajek XXL, on the other hand, are slightly newer programs specifically aimed at large and very large networks of multimillion-dollar nodes. Their user interface is idiosyncratic, and the terminology used in the documentation as well as many procedures may be unknown to social scientists because the authors have their roots in mathematical chart theory and computer science. However, Pajek is unrivalled in terms of speed and pure processing capacity. UCINET, Pajek and other SNA software enable analysis that is not possible with standard statistical software. However, moving data from a standard software package to an external network analysis program and then returning to the general purpose package for further analysis is a destructive, tedious, and error-prone process. The different SNA packages that exist for the R system are therefore an attractive alternative to stand-alone SNA programs. The most useful are Statnet (a meta package that includes many procedures with more specialized packages) and Igraph, which seems to be a bit more accessible (and is also available as a python package). Most likely the package will meet all but the most exotic needs of psephologists.3.2.4 Geospatial analysisAnageo-spatial analysis is a broad term that includes at least two distinct (if related) approaches: the use of geographical variables in normal 12zacharze electoral regression models on the one hand, and the estimation of specific statistical models that take into account spatial dependencies on the other. The first approach may simply use geographical references to merge micro-data with contextual information (see section 3.1). In more advanced scenarios, psephologists calculate geographic variables (most commonly distances) from sets of geo-references. This is best illustrated by an example: For various theoretical reasons, voters should prefer local candidates, i.e. candidates who live closer to the voter's place of residence than other candidates. If candidates are required to have their home addresses on the ballot paper and the addresses of voters are known,3 you can calculate the spatial distance between candidates and their potential voters (Arzheimer and Evans, 2012; 2014). This variable varies depending on the pair of voters in the districts (unless the voters live at the same address) and is therefore a global variable at the level of individual voters. Geospatial methods are required to (1) translate addresses into physical coordinates (a step known as geocoding) and (2) calculate different distance measures (e.g. travel time by car or public transport). In addition to calculating the distance of a straight line, which is a purely geometric problem, the second step requires access to digital road maps, timetables, congestion data and routing algorithms. However, after calculating the distance, the analysis can continue with normal linear and nonlinear regression models, which can include nesting or grouping observations by overlaying the structure on the variance and covariance matrix. Different types of spatial regression models go a step further. They correct the relationship between observations, taking into account the spatial coordinates of cases and adjusting the structure of the variance and covariance matrix accordingly. The importance of spatial regression models for psephology is most obvious for district-level aggregated analyses: While standard regression models assume that interference is identical and independently distributed, it is important that neighboring 4 districts will be affected by similar interference, and thus show pattern autocorrelation, which makes standard errors questionable at best. In spatial regression, you can use a distance matrix between the centroids of constituencies to estimate this autocorrelation, which in turn can be used to output corrected standard errors in the spatial regression model (Ward and Skrede Gleditsch, 2008). Spatial regression can be applied to data, but it is generally easier 3La data protection considerations, usually recorded only approximate geo-reference respondent. 4Addamiss of neighbors is a somewhat fluid concept, because these common influences will be stronger where the neighborhoods are physically closer and less pronounced, but still present, where a pair of neighborhoods are farther apart. This scenario is very different from nesting, where there are clearly defined, fixed groups of lower-level entities.and often also more appropriate in terms of basic theoretical assumptions about causal mechanisms for using a multi-level model (possibly involving more than two levels), which is nesting in political and administrative contexts. Mapping and processing of georeferenced data has traditionally required access to and training in the use of the Geographical Information System (GIS). GIS is essentially a reportable database with special capabilities for dealing with 2D and 3D coordinates. GIS software seems expensive, proprietary and complex. In recent years, however, government agencies and other organizations that open their data have created websites that hide at least some of the complexity of the underlying system. In the simplest case, users can create chloropleth maps or search for data for a single or limited number of localities. More useful systems allow you to download prebuilt or custom tables in machine-readable format that can be merged with existing data at the individual level. In very few ideal cases, there is an API that researchers can access programmatically (see section 5.1). In addition, algorithms for collecting, storing and processing georeferer data are now freely available and have been implemented in many stand-alone programs and/or packages for the R. GRASS (Geographic Resources Analysis Support System) is a fully functional GIS that has a wide range of applications in engineering, life sciences and social sciences. GRASS works on all major operating systems. It can be used both interactively via the graphical user interface (GUI) and programmatically via scripts. Its true power, however, lies in interfaces with two popular programming languages: Python and R. Thanks to these interfaces (Pygrass for Python and Rgrass6/7), users can on the one hand program the GRASS system and expand its capabilities. On the other hand, researchers who regularly conduct their analyses in Python or R can selectively use data stored in GRASS and the nearly 2,700 industry features available in the system. QGIS is a more lightweight alternative to GRASS. While interfaces with R and Python, too, it is mostly geared towards interactive use. In many cases, however, analysts in R or Python, they will want to avoid the gis overhead altogether. Many features of traditional GIS software are now available as add-ons for these two languages. R in particular, currently more than a hundred packages for loading, manipulating, analyzing and mapping georeferenced data (. html).4 Tools for successful, repeatable research The previous two sections of the linguists have established a rapid pace of technical progress in psephology. Somewhat paradoxically, this section suggests that in the face of increasingly complex data and software, psephologists should turn to the very basic tools, concepts and techniques that computer scientists developed decades ago: plaintext files and editors, directories (folders), and some tools commonly used in medium-sized programming projects. As the old saying goes: In electoral surveys, regression is progress.4.1 Establishing a repeatable workflow analysisDany covers a number of different phases (see Long 2009, Chapter 1 for a similar outline):1. The data must be collected by the researchers themselves or by a third party and stored electronically 2. This machine-readable data must be sent to researchers, usually via the Internet 3. The data must be re-encoded or otherwise standardized, possibly after being converted to another format. 4. A number of research analyses and preliminary models shall be run on data, perhaps using more than one computer program 5. Scientists settle on a small set of final analyses and models, the results of which are stored 6. For presentation and publication, charts and tables are created from these results, possibly using additional softwareA be reproducible by original scientists and their peers, each step, as well as the rationale for the decisions in question, must be documented. Realistically, this means that as much of the entire process as possible should be automated using scripts: short sets of instructions for a computer program. Graphical user interfaces are useful for learning about the program and possibly for making corrections to charts for publication, but scripts are infinitely more efficient, efficient, and reliable. When properly commented, the scripts are also self-documenting, although scientists should aim to keep a separate research journal. For smaller projects, presentations, and teachings, researchers may even want to continue programming literate (Knuth, 1984), which combines code for several programs, text for publication, and documentation into a single document from which intermediate tables and charts can be produced, as well as slides and PDF documents using the Knitr package for R, or even more general Orgmode for Emacs (see below). However, while literate programming is basically appealing, it may not scale well to larger projects. Most statistical packages have simple script editing capabilities built in, but in the long run it is more efficient to use stand-alone text editors that much more advanced editing features, as well as syntax highlighting, correct indentation, and basic project management capabilities. One of the The most eccentric and powerful of these editors is Emacs (, which was released in the mid-1970s and has since been in active development. Despite his age, interest in Emacs has grown in recent years, and many quantitative sociologists swear by it. Emacs can be endlessly customized and expanded, which may be surprising for new users. Cameron et al. (2005) is a useful introduction, but documentation for many other features and extensions is best searched on the Internet. Psephologists may also want to install one of the configurations specifically aimed at sociologists that can be found on the Internet. With the right set of extensions, Emacs supports almost every scripting language known to humanity, including the command languages of statistical packages such as Julia, OpenBUGS/JAGS, R, S-Plus, Stan, Stata, and SAS. At the very least, support means highlighting syntax, indenting, and checking detribution for balanced parentheses. Moreover, Emacs usually gives you access to the appropriate help systems for these languages and can find documentation for related features. It can insert boiler plate code (e.g. loops) and can execute code fragments or entire scripts. Emacs was designed as an editor for computer programmers and so has the ability to track variables and search for feature definitions in any number of files using text tools such as Diff, Grep, or Find, version control systems such as Git (more on them below). The more complex the toolchain becomes, the more it shines because R, Stata, Python and many other applications can be conveniently managed using a single keyboard-oriented interface and script.4.2 Buildtools, version control and other open source GadgetsIdeally, there should be a separate script for each of the six steps described in section 4.1. Shorter scripts are easier to maintain and it would be inefficient to restart the entire process just to add a horizontal line to the table. It is also extremely important that the data is edited only in a non-destructive way: each script must save its results as a new file, keeping the data collected and uploaded in steps 1 and 2 intact. It is also good research practice to keep all files belonging to a given project in the oftheir directory of their own, and to create separate subcaters for scripts, charts, tables, and datasets (Long, 2009). When a project grows beyond a few individual scripts, further automation of the process or meta-scripts becomes a necessity because individual tasks must be performed in a specific order. Basically, you can achieve a certain degree of automation within a statistical package of your choice: Both Stata and R are able to process scripts that rotate include or source other scripts. Moreover, both programmes have a basic infrastructure to run and can, at least in theory, manage the tool chain. In practice, however, It's easier and less error-prone to rely on an external scripting language, such as Python or the scripting language of a native operating system command-line interpreter (shell) to manage complex workflows. If some of the tasks are time-consuming or otherwise costly (i.e. estimating the model by numerical means or measuring data from the Internet), psephologists should rely on building tools: software that is typically used by computer programmers to build (build) complex software from hundreds of text files through a potentially large number of middleware files. If you are editing a single text file, you typically need to recompile a small fraction of the entire project that is directly affected by the change. Authoring tools can identify, manage, visualize, and most importantly exploit such dependencies, enabling significant productivity gains. On average, workflows for a software project are more complex than workflows for analyzing election data by several orders of magnitude, but psephologists can still benefit from learning how to use creation tools. This is best illustrated by an example. Consider the following simple workflow:1. Download (with R, or with a specialized program such as wget) a dataset (say the European Social Survey) from the Internet if the file on the internet has changed 2. Save the appropriate subset of data after recoding some 3 variables. Upload a subset, select some complex models, and save the parameters in file 4. Illustrate results by • Create multiple charts from parameters and save them as separate files • Create multiple tables with parameters and save them as separate files5. Generate a PDF report with latex document preparation system by processing a text file that contains charts and tablesFor an efficient and managable workflow, each task should be performed by a single program running on a single set of instructions (script or just a series of options and arguments submitted when the program starts). Moreover, each task requires one ore more input and leaves behind one or more outputs.5 The way in which these individual tasks are listed is very easy to recognise the relationship between them: If a new version of the European Social Survey is published, all steps must be repeated in that exact order. On the other hand, if a scientist decides to change the encoding of variables (step 2), the estimator used by the model (step 3), or the appearance of the graphs (step 4), only repeat the next steps. By the way, this last modification would not require rebuilding the tables: If dependencies were visualized as a tree, both tasks would be displayed at the same level because they are completely independent of each other. In a computer environment with sufficient ressources can be executed in parallel, which further speeds up the process. Creating tools such as the venerable Make program (Mecklenburg, 2005, generally unix-like systems) and its many modern successors require that dependencies be specified in yet another text file. While this may sound like a chore, it's usually just a matter of writing which script generates, what files (targets) you enter from. In addition, this helps to explain and streamline the workflow. After you enter this rule set, the build tool analyzes the dependencies and executes the tasks in the required order. After this initial run, targets will be regenerated only if the scripts or inputs from which they originate change. The last tool that psephologists should borrow from the world of software development are change control systems. Most researchers will be (painfully) aware of the value of automated backup systems that retain a number of old backups to avoid a good backup being replaced by a damaged copy. Modern systems usually provide a range of hourly or daily snapshots alongside increasingly older (weekly, monthly, 500) copies. Patch control systems take this idea of snapshots a step further, keeping a complete history of changes in each (text) file in the project directory.6 Modern version control systems such as the somewhat unfortunately named Git (Loeliger and McCullough, 2012) can track the entire state of the directory and quickly reset all files in the directory to the state they were in last night or show which changes were made to a specific file as of Monday evening. They provide tools to find the exact point at which some changes in the way variables have been recoded 5Perfectly, the number of inputs and outputs should be as low as possible (i.e. writing one single script for each chart that goes into the final document), but this can become very cumbersome and not always possible. 6Less they are synchronized with the remote repository, patch control systems only work on local files and thus do not protect against data loss by hardware failure. Researchers still need to make sure they back up their working files as well as their version control repository regularly.stopped the model from converging or led to a dramatic change in estimates further down the line. But most importantly, using a version control system introduces another layer of reliability and reproducibility. Modern version control systems simply cannot easily restore unwanted changes to project files, they can effortlessly maintain any large number of timelines (branches) for the project directory. It's a great tool for testing code and ideas: You can easily try out various operationalisations, model specifications or graphic styles across industries, once again recording all the changes made to your files and then switch back to a more stable development line that the current state of the analysis and selectively copy it to anything that worked. Revision audit systems are based on the assumption that each of the the changes should be documented in the commentary and therefore strongly encourage analysts to keep a log of the justification of the myriad small decisions they make when analysing their data and presenting their findings. Like many of the other tools discussed in this chapter, change control systems have been used by computer programmers for decades. Their modern incarnations are designed to deal with millions of lines of code spread over hundreds of files on which large teams of developers can work simultaneously. Psephologists may also think that a system like Git (which is relatively difficult to learn but can be tamed by a number of GIUs) is ridiculously overpowered for their needs. However, experimenting with code and data in a secure environment where any change is documented and can be restored, modified, and even reapplied at any later time is ultimately much more rational, rewarding and productive than the common practice of continuously commenting and popping up lines of code or creating many increasingly mysteriously named scripts whose exact purpose we do not remember after a few weeks.5 The Internet as an infrastructure for and as an object of electoral study5.1 InfrastructurePsephology has been transformed by the availability of large comparative opinion surveys such as ISSP, EES or eurobarometer series (see Chapter 48). The websites of CESSDA members and other large archives are now the default option for distributing these datasets, allowing for fast and cost-effective data spread, while physical media (e.g. DVDs or CD ROMs) have been withdrawn, unless strict rules of use apply at the same time. While archives are unrivalled when it comes to providing documentation and long-term secure storage of a large number of datasets, preparing data for release through an archive system is seen as cumbersome by many researchers. Thus, there has always been a tradition of more informal data sharing in psephology with friends and close associates. With the advent of the web, individual researchers and small teams began to place their datasets on personal or departmental websites. However, the data on such sites is often difficult to find because there is no central directory and can disappear at any time because they are not backed by professional infrastructure. What's more, data can be stored in any number of formats and without even minimal documentation. The Open Source Dataverse (project and some related initiatives aim to address these issues by providing the means to (semi)automatically convert, document, store versions, and retrieve data. They also provide unique identifiers and totals worldwide to resolve the data integrity issue. Journals, research groups and individual researchers can easily create their own repositories to facilitate re-analysis and replication. Dataverse and similar software go a long way to creating data from often self-funded projects that would otherwise be lost to the scientific community available for secondary analysis. However, they still rely on professional and sustainable IT infrastructure. At the moment, this infrastructure is available for free by Harvard University and several other global players. Wheter will continue to provide this service to the community if the use of the raises system remains to be seen. In addition to traditional data archives and more individual repositories, countless government agencies and other public bodies around the world have set up websites where they share parts of their records and thus become data providers. Especially at the subnational level, the main problem with these places is fragmentation. Even if they follow common standards for designing their websites and presenting their data, finding and navigating hundreds or thousands of individual sites to collect, say, data on candidates in local elections is obviously inefficient and often unworkable. Fortunately, governments around the world have woken up to the potential value of free access to their data and are implementing open data rules. As a result, a regional, national and even transnational portal or data warehouse sponsored by the government, which collects and disseminates detailed data from lower levels, is becoming more widespread. Although these initiatives are often aimed primarily at policy makers and business communities, social sciences also benefit from the emerging consensus on the principle of open government data. For psephologists, the increasing availability of geo-referenced election results and other statistics (e.g. census or landuse data) is of particular importance. In an ideal world, websites offer an accurate set of data required by a researcher in an ina format that can be read directly in your favorite statistical package. In fact, datasets are often offered under the guise of Excel worksheets or text files that need to be imported. Although this is not too problematic, such files are often created on the fly from the primary database according to some specifications that need to be entered manually. If the same dataset (or different variants and iterations of the same dataset) need to be downloaded more than a few times, you might want to do so programmatically using a script. Moreover, there are still (government) websites that present the required data not as a download, but rather as a series of formatted tables on the screen, possibly in a format for pages. In such cases, researchers should consider writing a scraper, a small program that simulates the website user's actions and stores the results as a dataset. While Python has a whole set of libraries that that is an ideal tool for scraping tasks, some modern packages for the R system offer very similar capabilities from the statistical package. Munzert et al. Et al. introduction to scraping and extracting the internet. While they focus on R, the techniques and standards they discuss easily translate into workflows that are based on other tools. Finally, many service providers – including several government agencies – provide application programming interfaces (APIs) to their data. APIs completely bypass the traditional website. They represent complex and very specific mechanisms for interacting with the underlying service database as a series of simple commands for high-level programming languages such as R or Python. With these commands, scripts can directly access these services without even simulating the user's web site activities. From the point of view of the script writer, access to a service on the Internet is no different from calling a function that is wired to the appropriate programming language. For example, psephologists might have a variable that contains the addresses of candidates, as stated on the ballot papers (a messy combination of names and numbers, possibly with typos). To convert them to the appropriate geographical coordinates, they would like to use the geocoding service. There are APIs for various such services (e.g. Google Maps, Bing Maps, and the OpenStreetMap project) that wrap the necessary low-level instructions into a simple function connection. With usage limits and possibly payment options aside, switching from one service to another is usually just a matter of applying slightly different features to a variable. Using the next API, the resulting coordinates can then be mapped to census circuits for which a lot of socioeconomic and demographic data is available that could provide a close-up and ready approximation of the relevant environment in which candidates live.5.2 Internet as an object Since its inception as a research infrastructure, the Internet has been completely transformed. While the usual reservations about selective access and use apply, the role of the Internet as a political medium is becoming increasingly important for psephologists. Current research is very focused on political communication, as it happens on social media platforms, with Facebook, Twitter and Instagram being the most prominent. Scraping these pages with simple scripts would not only violate their terms of use, but it is practically impossible due to their intensive use of interactive Internet technology, the network nature of communication on these sites and the huge number of posts. However, once again, THERE are APIs that can be used to extract these services programmatically. While there are limits, analysts often believe that free tiers are perfectly tailored to their needs. What's more, both Twitter and Facebook have strong research departments that are open to establishing partnerships with social scientists. Research on social networks on the Internet are now dominated by IT professionals and linguists, who often operate without theory of social behaviour. Psephologists interested in this field will have to learn all the many techniques and concepts, and will have to combine them with their own substantive interests. Grimmer and Stewart (2013) provide a useful introduction to automated content analysis, while Ward, Stovel and Sacks (2011) give tour d'horizon concepts in social network theory that matter to political scientists. Using the Internet to analyse conventional media sources is in many ways less problematic. While many publishers are seeking to implement paywalls to secure their revenue streams, many major outlets still put all or at least most of their content online. What's more, Google, Microsoft, and other companies have created aggregator sites that can be accessed programmatically. Using these sources, psephologists can retrospectively track the development of a particular issue during a campaign or assess the tonality of media reports on a set of candidates. In short, using the Internet and scripting language, researchers can achieve most of what would have required multiple research assistants and an extensive newspaper archive (or expensive database subscription) just a few years ago. The Global Data on Events, Location and Tone (GDELT, database provides google event, location and tone support (GDELT, a step further. GDELT, which is based on older event databases (Gerner et al., 1994), aims to automatically extract entity and event information from news reports and make them available globally. The GDELT project is somewhat controversial because its original founders have fallen out, and because of concerns about the quality of the conclusions that are drawn from the raw outlay. However, the project, which has aroused great interest in the IR community, also has great potential psephology.6 Conclusions From the very beginning of the election survey were the beneficiary and often the engine of technological and methodological progress in the wider field of political science. In recent years, this progress has accelerated: user-friendly software, ever-faster computers, and finally the proliferation of data means that yesterday's advanced methods are rapidly turning into today's new normal. Overall, this chapter argued that psephologists should continue to embrace technology in general and open source and open data revolutions in particular. As examples of life sciences (e.g. biology) show, psephologists can do more and more reliably if they think a little more like software developers and use freely available tool chains that have been tried and tested for decades in much harsher environments. However, there is the opposite side. Technology is a valuable tool, but it can also be and psephologists should never lose sight of their core competences: the ability to place individual findings in a broader context, using almost a century of building theory. Building. the world is full of data analysts who happily and quickly analyze election data just as they analyze any other data. Trying to compete with them on a purely technical level would be a hopeless undertaking. As a profession, we can only stand in our place if we can base our own analyses on deep theoretical insights. ReferencesArzheimer, Kai and Jocelyn Evans (2012). Geolocation and voting: Impact of candidate distance on voters on party selection in the 2010 General Election in England In: Political geography 31.5, p. 301–310. doi: 10.1016/j.polgeo.2012.04.006. — (2014). Geolocation of candidates and election of voters in the 2013 English County Council elections. In: Research & Politics. doi: 10.1177/2053168014538769. Cameron, Debra et al. Learning GNU Emacs. A guide to the world's most extensible customizable editor. 3rd ed. American voter. New York: John Wiley. Coleman, James S. (1994). Foundations of Social Theory. Cambridge, London: Belknap Press from Harvard University Press. Crawley, Michael J. (2013). Księga R. Chichester: Wiley. Eurostat (2015). NUTS Nomenclature of Territorial Units 2013/EU-28 statistics. regions of the European Union. Luxembourg: Publications Office of the European Union, url: GQ- 14- 006-EN-N.pdf/b9ba3339-b121-4775-9991-d88e807628e3. Gelman, Andrew and Jennifer Hill (2007). Analyze data using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press. Gerner, Deborah J et al. Machine Coding of Event Data Using Regional and International Sources. In: International Studies Quarterly 38.1, p. 91–119. Grimmer, Justin and Brandon M. Stewart (2013). Text as data. Promise and pitfalls of methods of automatic content analysis for political texts. W: Political Analysis 21.3, p. 267–297. doi: 10.1093/pan/mps028. Hox, Joop J. (2010). Multi-level analysis. Techniques and applications. 2nd ed. New York: Routledge. Huckfeldt, Robert and John Sprague (1987). Networks in context: Social flow of political information. In: American Political Science Review 81.4, p. 1197–1216. URL: . Ihaka, Ross and Robert Gentleman (1996). R: Language for data and graphics analysis. In: Journal of Computational and Graphical Statistics 5.3, p. 299–314. doi: 10.2307/1390807. Jackman, Simon (2004). Bayesian Analysis for Political Research. W: Annual Review of Political Science 7, p. 483-505. Jennings, M. Kent (2007). Political socialisation. In: Oxford Handbook of Political Behaviour. Ed. Russell J. Dalton and Hans-Dieter Klingemann. Oxford: Oxford University Press, p. 29-45. doi: 10.1093/oxfordhb/9780199270125.003.0002. Jöreskog, Karl G. (1990). New achievements in LISREL: Analysis of ordinal variables using polychorical correlations and weighted smallest squares. W: Quality and quantity 24.4, p. 387–404. Doi: — (1994). On estimation of polychoic correlations and their asymptotic Covariance Matrix. In Psychometrics 59.3, p. 381–389. Karvonen, Lauri and Jostein Ryssevik (2001). How bright was the future? A survey of parties, necklines and voters in the age of the technological revolution. In: Party Systems and Voter Adjustment Revisited. Ed. by Lauri Karvonen and Stein Kuhnle. London: Routledge, p. 45-61. King, Gary (1997). Solving the problem of ecological inference. Reconstruction of individual behavior based on aggregation data. Princeton: Princeton University Press. Knoke, David and Song Yang (2008). Analysis of social networks. 2nd ed. Knuth, Donald Ervin (1984). Literate Programming. In: Computer Journal 27.2, p. 97–111. Lakhani, Karim R. and Eric von Hippel (2003). How open source software works: 'Free' user-to-user help. In: Research policy 32.6, p. 923-943. doi: 10 . 1016/S0048-7333(02)00095-1.Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet (1944). The choice of the people. How a voter creates his mind in a presidential campaign. Chicago: Columbia University Press. Lazarsfeld, Paul F. and Herbert Menzel (1961). About the relationship between individual and collective properties. In: Complex organizations. Sociological reader. Ed. by Amitai Etzioni. New York: Holt, Rinehart & Winston, p. 422–440. Loeliger, Jon and Matthew McCullough (2012). Version control with Git. 2nd ed. Long, J. Scott (2009). Data analysis workflow. rules and practices. College station: Stata Press. Lutz, Mark (2013). Learning Python. 5th ed. McClurg, Scott D. (2006). Electoral importance of political discussions: exploring the disagreement and impact of social media expertise on political participation. In American Journal of Political Science 50.3, p. 737–754. URL: . org/stable/3694246. Mecklenburg), Robert (2005). Project management with GNU Make. Sebastopol: O'Reilly. Miller, Warren E. and J. Merrill Shanks (1996). A new American voter. Cambridge, London: Harvard University Press. Monogan, James E. III (2015). Pre-recording of research in the field of political science. Case, counter-arguments and response to criticism. In: PS: Political Science & Politics 48 (3), p. 425–429. doi: 10.1017/S1049096515000189. Munzert, Simon et al. Automatic data collection with R. Practical guide to scraping web pages and extracting text. Chichester: Wiley. Muthén, Bengt O. (1979). Structured probit model with latent variables. In: Journal of the American Statistical Association 74, p. 807–811. — (2002). Beyond SEM. General Latent variable modeling. In Behaviormetrika 29, p. 81–117. url: /home/kai/Work/Texte/Muthen2002.pdf. Olson, Mancur (1965). The logic of collective action. Public goods and group theory. Cambridge: Harvard University Press. Raymond, Eric S. (1999). Cathedral and Bazaar. Reflections on Linux and Open Source by an accidental revolutionary. et al: O'Reilly. Sigismth, André (1913). Tableau politique de la France de l'Ouest sous la Troisieme Republique. Paris: A. Colin. Skrondal, Anders and Sophia Rabe-Hesketh (2004). Generalized modeling of latent variables. Boca Raton u.a.: Chapman & Hall. Steenbergen, Marco R. and Bradford S. Jones (2002). Modeling multi-level data structures. In: American Journal of Political Science 46, p. 218–237.Stegmueller, Daniel (2013). How many countries for multi-level modeling? Comparison of frequentist and bayesian approaches. In: American Journal of Political Science. doi: 10.1111/ajps.12001. Ward, Michael D. and Kristian Skrede Gleditsch (2008). Spatial regression models. Quantitative applications in social sciences 155. Thousand Oaks: Sage. Ward, Michael D., Katherine Stovel and Audrey Sacks (2011). Network analysis and political science. W: Annual Review of Political Science 14.1, p. 245-264. doi: 10. 1146/annurev.polisci.12.040907.115949. Western, Bruce and Simon Jackman (1994). Bayesian inference for comparative studies. In: American Political Science Review 88, p. 412–423.Photo by Christiaan Colen Colen