Methods of Analysis
Data Analysis

Stephen Petrina (2010/2019)

Methodology is too important to be left to methodologists. (Becker, 1970, p. 3)

Although data analysis would seem to date to the beginnings of analysis (ἀνάλυσις) in antiquity, the practice specifically dates to the development of statistical analysis and the Hollerith machine in the mid to late 1800s (Holmes, 2017; Rosenberg, 2013). Qualitative data analysis (QDA) predates quantitative data analysis (Q*n*DA) but one may counter that it depends on how data analysis (DA) is defined. This precedence is counted in decades, not centuries. For instance, historians were practicing historical DA prior to biometrics and demographic DA but did not regularly reduce documents or artifacts of the past to data until the early to mid 1800s. Evidence of the relative novelty of data and DA, Darwin refers to data just twice in *On the Origin of Species* (1859) and once in two volumes of *The Descent of Man* (1871). Darwin refers to fact, description, and explanation rather than data and analysis (i.e., used just twice in the three texts, *Descent*, v. 1).

The premise is that without data there is no DA. "*Data* are by no means easy to define," Russell (1914a) noted, "since there are various logically different kinds of data" (p. 585). "The simplest kind will be *objects*," and there are "at least three ways in which data may be given," he continued, "namely in sense, in memory, or in imagination." Simply put, says Russell, "*data* include all particulars, universals, and facts" (p. 586). Logically for Russell, "comparison of data" implied an analysis of the words standing in for these bonds and objects. His logical analysis of experience asks: "what is the bond which combines certain objects into the group forming a momentary experience?" (1914b, p. 16).

Q*n*DA in the early twentieth century commonly meant grouping and cross-tabulating. Grouping by defined factors or shared attributes makes the data more convenient and intelligible. Cross-tabulating involved defining additional factors across and within the groups for variation. Although focusing on statistical analysis, Tukey (1962) expresses the expanded scope of DA: "procedures… techniques for interpreting the results of such procedures, ways of planning the gathering of data… and all the machinery and results of (mathematical) statistics" (p. 2).

By the 1970s, Q*n*DA was nearly synonymous with both statistics and automated DA. Hypothesis testing became the core of Q*n*DA and a signature of "respectable scientific epistemologies" (Sanders, 1981, p. 579). Q*n*DA and QDA are dependent on grouping (or classifying, sorting, etc.) and coding data but Q*n*DA involves coding by dependent and independent variables. Quantifying qualitative data requires considerable coding. Given the role of sampling, most statistical analysis relies on estimation (interval, point). Paradigmatic practices of Q*n*DA tend to be grouping, coding, estimating, and inferencing and for QDA tend toward coding, categorizing, thematizing, and interpreting.

QDA by the 1970s was commonly "an effort to formally identify themes and to construct hypotheses (ideas) as they are suggested by data and an attempt to demonstrate support for those themes" and ideas or propositions (Bogdan & Taylor, 1975, p. 79). QDA became reliant on thematic analysis. Taylor, Bogdan, and DeVault (2016) indicate three primary practices of QDA: 1) "identifying themes and developing concepts and propositions;" 2) "coding the data and

refining one's understanding of the subject matter;" and 3) "scrutinizing the emerging analysis and attempting to discount findings— understanding the data in the context in which they were collected" (pp. 169-170). "Refining one's understanding" or categorizing gives shape to the complexity of data and codes. A *code* is "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon" (Boyatzis, 1993, p. 163). A *theme* is "a pattern found in the information that at minimum describes and organizes the possible observations and at maximum interprets aspects of the phenomenon" (p. 4) (see Content Analysis).

Bong (2007) characterizes DA as "at once conceptual and organisational, interpretive as well as mechanical. Coding for expedient retrieval (of categories) and theory building (relationship among categories) involves the pragmatics of breaking down or dissecting one's data into manageable and meaningful analytical units" (p. 267). Hence, data, units of analysis, and goals of DA are shaped by methodologies and theories. For instance, a primary goal of case study DA is becoming intimate with cases and patterns, a goal of ethnographic DA is interpretation, grounded theory strives for creation of new concepts, while statistical analysis aims to measure or recognize patterns and model, explain, or predict causal and probable relationships.

Units of analysis are dependent on methods *and* theories to focus codes, categories, and themes. For example, is the unit of analysis children's activity, classroom climate, conceptual understanding, discourse of teacher autonomy, individual choices, group dynamics, sports complex or system, test score, youth subculture, etc.? Codes, categories, and themes give latent meaning to analysis of manifest data. Coding, categorizing, and thematizing in QDA and grouping, coding, and estimating in Q*n*DA, are ways of effectively representing theories or theoretical frameworks.

DA is made empirical through various practices, which include a provision of potential access to data and data trail. In the 1800s, for example, one finds notes that "the data are given from which the conclusions are drawn, every person is furnished with the means of making his [her or their] own deductions" (Lloyd, 1830, p. 63). A codebook with examples can satisfy a data trail obligation (Tables 1-4). One value of using QDA software is that the links among data, codes, and categories are made evident and maintained throughout and post DA. A second value is the automation of coding and categorizing, which facilitates qualitative reasoning.

Despite expressions of uneasiness with assertions that anything and everything is data, definitions nowadays are immensely inclusive. The Consultative Committee for Space Data Systems (2002), for example, defines data as "a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing" (p. 9). All phenomena— however sacred or profane, natural or artificial, living or dead, imagined or observed, macro or micro— can be actively represented or reduced to data (e.g., any bit, image, text, sound, object, or being). If, as James (1890, p. 502) observed, analysis "is manifestly one of the most incessantly performed of all our mental processes," then shouldn't data be that which is most incessantly at hand? Does this all-encompassing scale and scope of data mean that DA encompasses any and all analysis? Does the expansion of data and DA make other methods of analysis, including conceptual, discourse, experiential, logical, perceptual, and thematic analysis, redundant? Or are data and data analysis fairly incomplete without metadata and metadata analysis?

Table 1. Example Codebook for QDA (Petrina, 2017) (unit of analysis = muskrat-mountie actor-network).

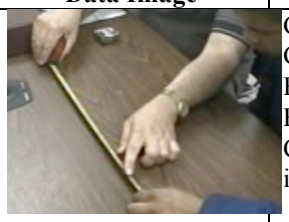
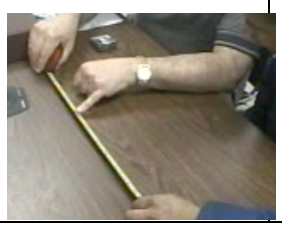
Data example, quote, etc.	Code	Category / Theme
Muskrat, Anishinabe, Ojibway So, muskrat dove into the water. He was gone much longer than any of the others who tried to reach the bottom. After a while Nanaboozhoo and the other animals were certain that muskrat had given his life trying to reach the bottom. Far below the water's surface, Muskrat, had in fact reached the bottom. Very weak from lack of air, he grabbed some Earth in his paw and with all the energy he could muster began to swim for the surface. One of the animals spotted Muskrat as he floated to the surface. Nanaboozhoo pulled him up onto the log. "Brothers and sisters," Nanaboozhoo said, "Muskrat went too long without air, he is dead." A song of mourning and praise was heard across the water as Muskrat's spirit passed on to the spirit world. Suddenly Nanaboozhoo exclaimed, "Look, there is something in his paw!" Nanaboozhoo carefully opened the tiny paw. All the animals gathered close to see what was held so tightly there. Muskrat's paw opened and revealed a small ball of Earth. The animals all shouted with joy. Muskrat sacrificed his life so that life on Earth could begin anew. (http://www.muskrat.com/index.html)	Furry Actors: Muskrat Creation Stories (Ojibway & Anishinabe)	First Nations, creation
Glenn Ducet, Executive Director of the Fur Institute of Canada The history of Canada was built on the fur trade and many Canadians are still involved in it. (Quoted in CTV News, October 1, 2014)	Muskrationalization Canada	Actor-Lobby: Fur Institute of Canada
Glenn Ducet These animals are abundant and plentiful in our environment and need to be trapped to protect our ecosystems, let alone for the fur trade. (Quoted in CTV News. October 1, 2014)	Muskrationalization Canada	
Andrew Stanley, Trapper Trapping's great for our economy, it's great for the Canadian economy and the Mounties wearing the muskrat hat — I always thought it represented us, represented Canadians as well. (Quoted in CBC News, October 2, 2014)	Muskrationalization Canada	Actor-Trappers-Harvesters: Happy
Leona Aglukkaq, Environment Minister (Nunavut First Nations) The RCMP decision, which is causing much glee among anti-fur activists, is being fully overturned. Our government will always stand up for Canada's hunters and trappers.	First Nationalization	
Leona Aglukkaq I would like to assure Canadians that the historic fur winter hats worn by the RCMP will not be discontinued despite the efforts of the radical animal rights activists. The RCMP decision which is causing much glee among anti-fur activists is being fully overturned. Our government will always stand up for Canada's hunters and trappers. (Quoted in PRI, October 2, 2014)		
Robert Sopuk, Conservative Party Representative (Rural Manitoba) What bothers me about these well-funded animal rights groups — what actually results from their activities is that poor people get hurt, to be blunt. For some of them, trapping is the only economic activity available, and provides these people who live on fairly low incomes with the dignity of work and the ability to pursue a		

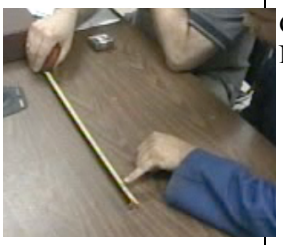
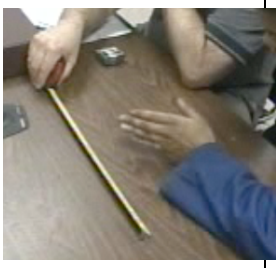
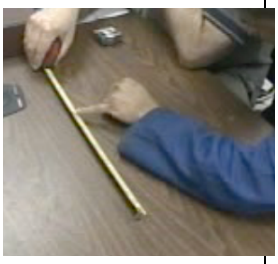
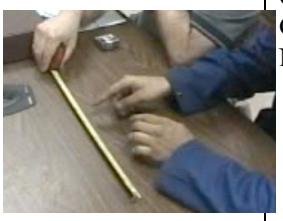
way of life that I strongly admire. As their Member of Parliament, I view this as my duty to protect and defend that way of life. (Quoted in PRI, October 2, 2014)		
Charles Ross, the Head of International Marketing for Saga Furs This is almost the golden age in fur. Our skin prices are going up 20-30% every year…. Fur is not for everybody. For people that have made the decision not to use fur, we certainly respect that. But, other people should also respect the right for a designer to use it. (Quoted in CNN Living, February 15, 2013)	musk(r)atwalk controversy	Furry Fashion: Fur on the Catwalk
Charles Ross The studies have shown that faux fur is an environmental nightmare. That it's a non-biodegradable product made from petroleum whereas fur is very biodegradable and has a long shelf-life. (Quoted in CNN Living, February 15, 2013)	small muskrat, big value animal rights	Furry Fashion
Robert Sopuk, Conservative Party Representative (Rural Manitoba) The fur trade is vital to the economy of many remote rural communities—communities who often have few other economic options.	Uniform Weather Conservative Party Actor	
Michael Howie, Spokesperson for the Association for the Protection of Fur-Bearing Animals To say that evolving as a community is somehow disrespectful to our past is disturbing. This heritage argument has to end.		
Michael Howie Interfering with what the public clearly believes in is what is disrespectful here. And right now, we are disgusted with our government.		
Michael Howie I wear a toque made from organic materials.		
Leanne Mai-ly Hilgart, founder of fashion label Vaute Couture I want to reach women who love style, love color, love fashion, and maybe they used to care about where their clothes came from but at some point they told themselves that it was naive to care. I think it's important that people see that you can care, you can interact with the world in the way you want and it's not naive. But to do that, you need options. (Quoted in CNN Living, February 7, 2013)	musk(r)atwalk faux fur	Furry Fashion
Leanne Mai-ly Hilgart I realized that if I could create a business where the process in itself was actually creating positive change, that would be my activism. I started with outerwear because I found being cold was an excuse to wear animal products. I wanted to figure out where I would be needed to make a contribution to the movement so people would no longer need to use products and materials made from animals…. I knew I had to design something that would be innovational for the entire industry. Not using animal fibers was an opportunity to look past what was just good enough to make something truly superior. (Quoted in CNN Living, February 7, 2013)		

Table 2. Example Codebook for QDA (Namita & Petrina, 2010) (unit of analysis = teacher discourse).

Data	Code	Category / Theme
Jennifer: "students in this school are really savvy, they know media … they've grown up with it. … There is a bit of a challenge, because honestly, I think this generation probably knows more about it than I do in a lot of ways. They know more about the medium, and they're aware of it because they've grown up with it."	• Hype • Savvy Kids • Millennials • NetGen • Kidults • Digital Natives	Popular (aka Divisive) Hyper Discourse
John: "kids are a lot more tender than people assume, and they do not need to be exposed to a lot of things that are just unnecessary. Look at the way our language is degenerating, especially English speakers. We are re-entering a period of foul language, or I do not like to load it like that, because it is really just uncouth, where everybody just feels free to use a low-end of the English verb spectrum. The F-word is constant … and I blame the media for that."	• Victimized yet threatening kids • Protective pedagogies • Digital Natives	Protective (aka pedagogical) Discourse

Table 3. Example Codebook for QDA (LaCroix & Petrina, 2010) (unit of analysis = C's activity system).

Data Text	Data Image	Code	Category / Theme
[26:43] 75 L: … What do you notice here between the spaces here up to twelve, (G -uses his index finger to sweep up from the zero end of the measuring tape and pauses at 12" just before saying "up to twelve".) 76 C: Yeah its, 77 L: and the spaces after twelve? (G-changes to 4th finger to sweep through the exposed interval of the measuring tape above 12	   	Gesture (Pointing) Gesture (Sweeping) Enumeration Fraction Confusion (metric v imperial)	• Knowledge objectification • Mediation • Subjectification • Artifactual thought • Iconicity • Semiotic system of cultural signification •

78 C: There's, (G[Figure 14, 26:52]–sweeps up through the first few inches of the tape measure with the fourth finger of his left hand in a manner similar to the gesture just enacted by L) there's more. (G[Figure 15, 26:53]–makes two chopping motions aligned with the divisions on the tape measure with his left hand, the first significantly larger than the second just before he says "there's more" in reference to the divisions inscribed on the measuring tape. G[Figure 16, 26:54]–then points to the 12" mark with the fourth finger of this left hand before withdrawing it from the measuring tape). (line 78 continues) It's like it's more spread out (in reference to the divisions on the tape measure after the 12 inch mark.)	  	Gesture (Chopping) Notice	Semiotic resources: • spoken language including mathematics vocabulary; • voice inflection and changes in volume; • mathematics notation; • three forms of gesture—pointing or indexical, sweeping, & chopping; • a line drawn to represent five-eighths-of-an-inch; • indexical inscriptions such as circling or underlining existing inscriptions; • counting; • written text; • rhythm in speaking or gestures; and • the position, orientation, alignment of physical objects.
(G[Figure 17, 26:55a]–points briefly to the 12 inch mark on the tape measure, now with the first finger of his right hand, replacing the previous pointing gesture that had been expressed by the fourth finger of his left hand; G[Figure 18, 26:55b and Figure 19, 26:56a]–starting with his thumb positioned at the 12 inch point, sweeps his right hand up the measuring tape a short (line 78 continues) when (G[Figure 20, 25:56b]–grasps the tape measure with his right thumb and first finger on opposite edges at the 12" point and G[Figure 21, 26:57]–sweeps his hand in this configuration upwards a short distance from 12") you pass one, distance while holding an approximately 2.5 inch wide interval	 	Gesture (Pointing) Gesture (Sweeping) Notice	

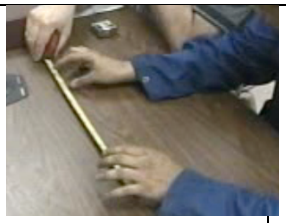
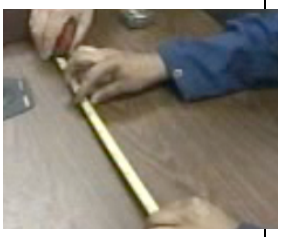
between the thumb and first finger.) 79 L: Yeh, 80 C: one foot (G[not shown]–while maintaining the same grasping position, repeats this sweep upwards for a second time) (line 80 continues) And when you're before one foot its more um.			
(G[Figure 22, 27:01]–makes a very brief and narrow-interval gesture with the thumb and first finger of his right hand with this hand now positioned above the region of the tape measure between 0" and 12".) 81 L: Okay. 82 C: [silence] **[28:16]** 96 C: Yeah, it looks the same. Actually (surprised), no it looks bigger. 97 L: Yeah. 98 C: Way bigger. 99 L: Okay. 100 C: It's not right then	  	Gesture (Grasping) Gesture (Interval) Notice	

Table 4. Example Codebook for Q*n*DA (Ip & Petrina, 2007) (unit of analysis = gender x performance).

Independent Variables / Codes	**Dependent Variables / Codes**	**Hypotheses & Tests**
Gender	Edudata ID (Edudataid)	Null hypothesis (H0) is that exam scores for male_m equals that for female_f, and the alternative hypothesis (H1) is that exam scores for male_m is lower than female_f. Gender: H0: _m = _f H1: _m < _f Independent sample T-tests (students differ from year to year). Phase 1: T-tests for 6 subjects within school year. Phase 2: T-tests for 10 school years. *Comparing gender performance.
School Type (Schtype)	Principles of Mathematics 12 (Ma 12)	
Year	Biology 12 (Bi 12)	
Course Level (Crslevel)	English 12 (En12)	
	Communications 12 (Com 12)	
	History 12 (Hi 12)	
	Physics 12 (Ph 12)	

Data Sample (after cases w/ missing data removed)

Year	Course	Female	%	Male	%	Total
2004-2005	Biology 12	10641	64.2	5923	35.8	16564
	Communications 12	2327	35.2	4290	64.8	6617
	English 12	20937	52.6	18896	47.4	39833
	History 12	5140	49.9	5151	50.1	10291
	Physics 12	2244	30.5	5110	69.5	7354
	Principles of Math 12	8127	47.3	9072	52.7	17199