———. 2007b. *Voices of North Carolina: Language and life from the Atlantic to the Appalachians: Student Workbook.* Raleigh, N.C.: North Carolina Language and Life Project.

———. 2008. "Language Diversity and the Public Interest." In *Sustaining Linguistic Diversity: Endangered and Minority Languages and Language Varieties*, edited by Kendall A. King, Natalie Schilling-Estes, Lyn Fogle, Jia Jackie Lou, and Barbara Soukup, 187–204. Washington D.C.: Georgetown University Press.

———. 2012. "In the Profession: Connecting with the Public." *Journal of English Linguistics* 40.1: 111–17. doi:10.1177/0075424211428269.

KELLY D. ABRAMS is a Ph.D. candidate in English language and linguistics at the University of Wisconsin–Madison. She researches how nonstandard dialects intersect with standard literacy practices in the classroom. E-mail: kdabrams@wisc.edu.

TRINI STICKLE is assistant professor of English at Western Kentucky University, Bowling Green. She continues to work on secondary and post-secondary curricula created from the *Dictionary of American Regional English* for various English, literature, and history courses. She has also published on the linguistic and interactional strategies of persons in meeting settings and children with autism during clinical evaluations. Her main research interest is the linguistic and interactional competency of persons of special populations (e.g., dementia, autism). E-mail: trini.stickle@wku.edu.

DOI 10.1215/00031283-4202053

INTERACTIVE NAME DATABASES AS AN INTRODUCTION TO SOCIAL FACTORS AND GRAPH INTERPRETATION

MARISA BROOK, *Michigan State University*

This article introduces two data visualization tools linked to large databases of American first names and illustrates how they may be used as an interactive introduction to an undergraduate class in sociolinguistics and/or dialectology. Although first names are adjacent to language in a sense, the fact that they are familiar and not at all technical makes them a good starting point. As with language, what students may not realize in advance is that names are full of patterns, especially with respect to change over time, geography, and other unconscious but powerful factors in the speech community (Lieberson 2000; Lieberson and Lynn 2003; see also Labov 2010, 194–95).

Both databases mentioned in the activities here are free to use, have intuitive interfaces, and do not require a login; they can be used with minimal training or preparatory work.[1] Together, they form a versatile teaching aid and need not be confined to classrooms teaching variationist sociolinguistics specifically. However, they hold value with respect to introducing the variationist approach. The name databases enable students to explore large data sets and patterns in a self-directed way, without first introducing them to sociolinguistic concepts. This is especially valuable for students coming to sociolinguistics without much numerical background, such as those majoring in the humanities who may not have had to think about *x*- and *y*-axes since high school. Specifically, objectives of the activity are that the students will:

1. gain experience exploring data visualization and begin learning to read and interpret graphed information critically;
2. understand patterns in a dependent variable according to one independent factor, and then two (as well as the possibility of an interaction); and
3. gain a sense of the social factors that divide groups of people who act differently (with respect to behavior, trends, values, culture, language, and so on).

I report here on a mini-unit (two 2-hour lectures and one homework assignment) that I devised for my introductory sociolinguistics class (LIN471) at Michigan State University in the fall of 2016. There were 25 students enrolled, almost all undergraduates. They came from a range of majors and programs, but each had completed at least a basic linguistics class. I used this activity at the very beginning of the class; it would also be well-suited to units focusing on social factors, age, and time, or the concept of the sociolinguistic variable.

CLASSROOM ACTIVITY: DAY 1. In advance of the beginning of the class, students are asked to bring portable electronic devices that can connect to the Internet. The first lecture begins with a thought experiment, as follows: we have a mystery individual who was born in the United States at some point in the country's history. That is all we know about this person, but even so, we have been asked to provide as good a guess as possible as to what their (full, legal) first name is.

As it stands, the odds of such a guess being correct are not very good. We could select an especially popular name in order to maximize our chances, but even so it is a long shot. However, the students are told, we have two sources of assistance. One is that we know that this person does not have a UNIQUE first name, which dramatically restricts the number of possibilities.[2] The other is that we are permitted to request up to three specific pieces of demographic information about our person. Which three social characteristics, the students are asked, would do the most to narrow down the set of first names that our mystery American is likely to have?

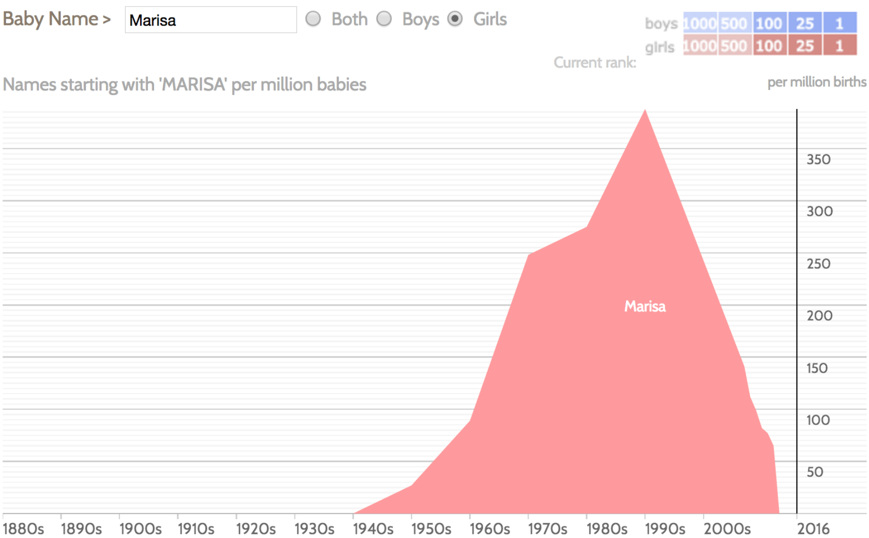
Sex/Gender. Students are likely to spot this as a straightforward example of a social characteristic that would let us rule out approximately half of the possible first names. This gives the instructor the chance to discuss the fact that few first names in the United States are truly gender-neutral. Rather, most of them are marked for gender—or, at least, for sex assigned at birth. (If students are interested, there is room here for a discussion of transgender and/or nonbinary names [see VanderSchans 2015].)

Ethnic/Racial/Cultural Background. A bit of prompting might be necessary to nail down which factor would likely differentiate, for example, a *Jimena*, a *Shaniqua*, an *Izumi*, and a *Rivka*, but that ought to suffice. It is worth pointing out to students that even among popular names so conventionalized in Western society that they do not suggest a likely origin, differences can be found in naming trends between different ethnic/racial/cultural groups. (Taking, for instance, a set of results from Lieberson [2000, 204], it may not be surprising that *Latoya*, *Kiara*, and *Ebony* were names that were much more popular for African American baby girls than Caucasian ones in Illinois in 1989; but the same is true of *Amber*, *Bianca*, and *Tiffany*.) As with sex/gender, there is the possibility here of a more extensive discussion of first names across ethnicities/racial identities/cultures in the United States (e.g., Fryer and Levitt 2004).

Year of Birth (or Age). My suspicion was that this would be the most elusive of the three social characteristics that would be a major help in narrowing down a list of likely first names for our mystery person; however, my introductory sociolinguistics students identified it readily. Either way, students are likely to have intuitions about the age of a woman named *Gladys* or *Agnes*—or, on the other hand, one named *Addison* or *Nevaeh*. This third factor leads the instructor directly to an introduction of the first interactive tool: Baby Name Voyager (Wattenberg, n.d.), which can be found at http://www.babynamewizard.com/voyager.[3]

An example output graph from Baby Name Voyager is displayed in figure 1. At this point, if there are students in the class with limited (or half-forgotten) knowledge of graph reading, the instructor can introduce the basic terminology (the *x*-axis, the *y*-axis, their scales), with directions on how to read individual points (e.g., "in the year 1970, there were 250 baby girls per million births in the U.S. named *Marisa*"). Necessary to mention is the fact that in this data visualization tool, the *y*-axis scale across different names

FIGURE 1
Output Graph from Baby Name Voyager with an Example First Name (*Marisa*)



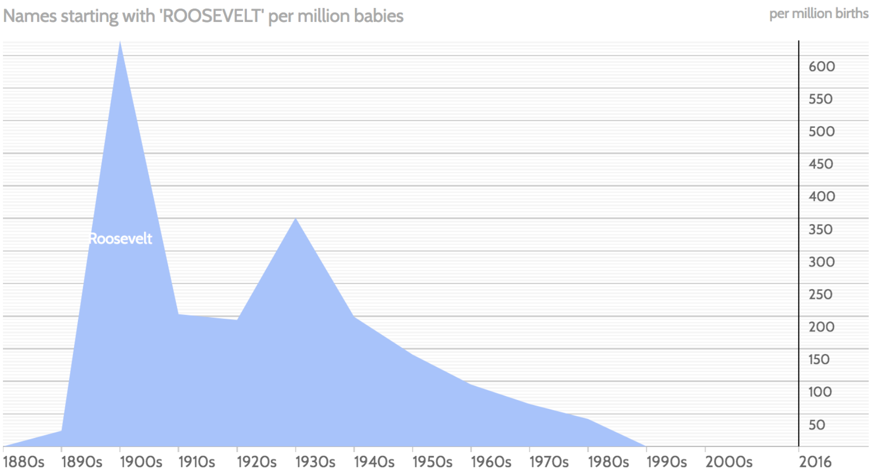
NOTE: The "enlarge" button at the bottom right is easy to miss, but it usefully opens the searchable graph in a new window of the browser.

does not remain constant. That is, although names are all measured the same way, they will be displayed differently, with the *y*-axis changing to reflect the height of the largest peak. It is also very much worth pointing out that a zero on this graph does not necessarily mean that the name was unattested; just that at the time it was not common enough per million births to register.

This is an excellent chance to introduce the concepts of the dependent and independent variables, particularly for students who are not already familiar with these concepts. Students can be encouraged to fill in the blanks: the dependent variable is what is being measured (in this case, baby name frequency), and the independent variable(s) is/are however many things the dependent variable is being analyzed in terms of (in this case, year of birth or time).

With this introduction complete, students can be sent to explore the Baby Name Voyager and find patterns that intrigue them among whichever names they feel compelled to try out (often first their own names, then branching out from there). This stage of the activity can either be left entirely open-ended or be made into a scavenger hunt. What I did was let the students explore on their own and report back. After 10–15 minutes,

FIGURE 2
Output Graph from Baby Name Voyager Showing a Name (*Roosevelt*) with Two Peaks, Their Timing Corresponding to Federal-Level Events



I asked for suggestions for name patterns that looked intriguing, then displayed these on the screen at the front of the classroom and asked the student contributing the name to describe what fascinated them about the pattern.[4] Students found, for instance, grandparents' names that are now unusual, overlap in patterns in names that have several common spellings, and graphs with abrupt spikes showing sudden trends that did not last. One of the few specific challenges I issued at this stage was to look for a name with multiple peaks (as in figure 2)—with a hint to try out names that more than one well-liked leader has had.

After this first lecture, I gave the students a more directed set of questions as an assignment—mostly involving just searching for patterns, but also introducing hypothesis formulation and testing. Most of the example names were uncovered through trial and error, with background inspiration from Wattenberg's blog associated with the Baby Name Voyager (http://www.babynamewizard.com/blog).

HOMEWORK ASSIGNMENT

Name scavenger hunt! A few of these questions could probably be answered by using the Baby Name Blog, but it's better to try a bunch of names and find your own. Using the examples to answer the questions is worth zero points.

Be very careful when there are multiple names on the same graph. Searching for *Fran* will put *Francis*, *Frances*, *Frank*, *Francisco*, *Francesca*, *Franklin*, and *Frankie* on the same graph, and add them to each other, which will make patterns in the higher ones very difficult to see. Search for names individually to get around this problem. If an individual name is contained within a longer one (e.g., *Anton*), try picking a different name for a more straightforward result.

FAD. Find a name that has only one peak between 1880 and 2015. A peak is a point that is higher than the points on both sides of it. Example: *Roberto*. Counterexample: *Victoria* (has multiple local peaks). [e.g., figure 3]

REVIVED. Find a U-shaped name where each side of the U hits at least 250 per million. Example: *Emma*. Counterexample: *Cyrus* (not popular enough on either side). [e.g., figure 4]

OBSOLETE. Find a name that has been at the negligible level since 1970. (Hint: Try some names of U.S. presidents born in the 19th century.) Example: *Woodrow*. Counterexample: *Albert* (has dropped off but is still used at low levels). [e.g., figure 5]

PERPETUALLY POPULAR. Find a name that has always been above 1,000 births per million in the entire period of time covered by the graph. Example: *John*. Counterexample: *Rebecca* (not quite popular enough throughout). [e.g., figure 6]

NEW. Find a name or a spelling that is not found before the 1980s. Example: *Matteo*. Counterexample: *Norah* (found at low levels in the late 19th century). [e.g., figure 7]

Hypothesis! Pick ONE of the following names. Before you search for it on the Baby Name Voyager, make a prediction as to what you expect the shape of the graph to look like, and explain why. It does not have to end up being correct. However, it does have to be a guess made on reasonable grounds.

Agnes; *Barbara*; *Henry*; *Ulysses*; *Emerald*; *Kyla*

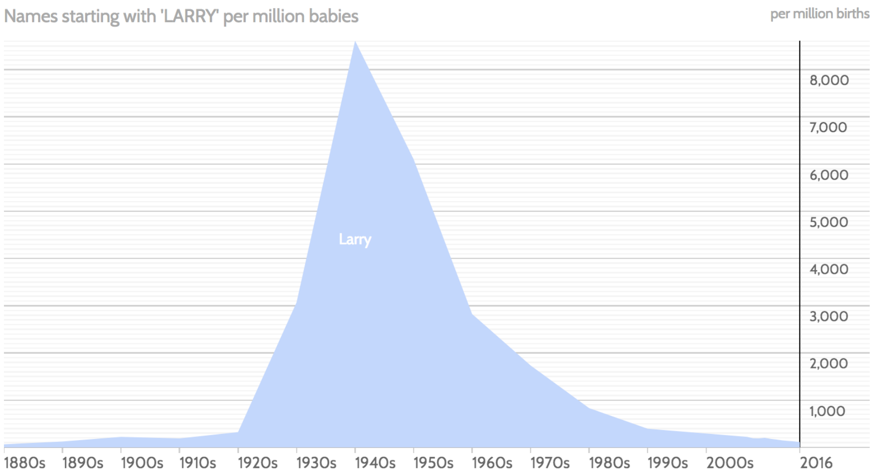
> Example: "I've heard that lots of little boys have been named *Jackson* lately, but there was also a U.S. president with that surname, so I expect the graph to be U-shaped."

Now check the graph and compare it to your hypothesis and explain. DO NOT change your hypothesis in retrospect if it was wrong.

> Example: "I was correct about the recent increase in the popularity of *Jackson*, but it does not seem to have been a popular name in the 1880s as I expected. Maybe Andrew Jackson was not well enough liked for lots of people to name their children after him."

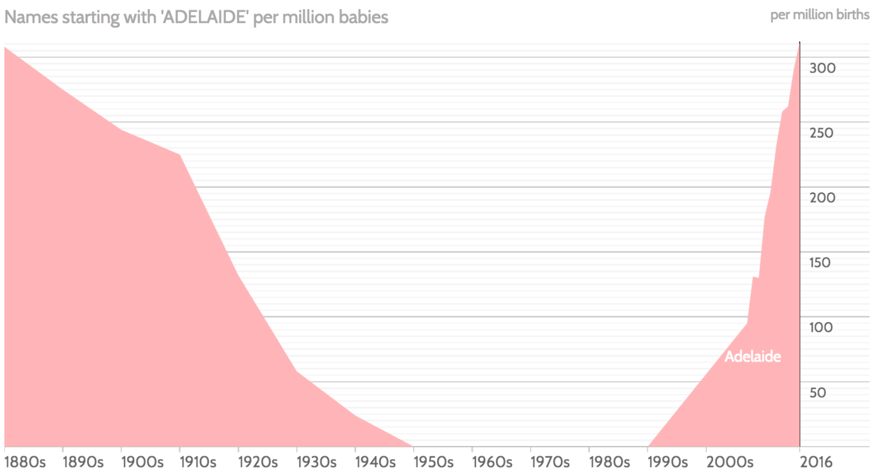
Correct guesses are NOT worth more points than incorrect ones. Warning: Changing your hypothesis after seeing the results is scientifically bogus in all fields. Doing so is intellectual dishonesty. Don't do it.

FIGURE 3
Output Graph from Baby Name Voyager of a Faddish Name (*Larry*)
(i.e., graph with a single peak between 1880 and 2016)



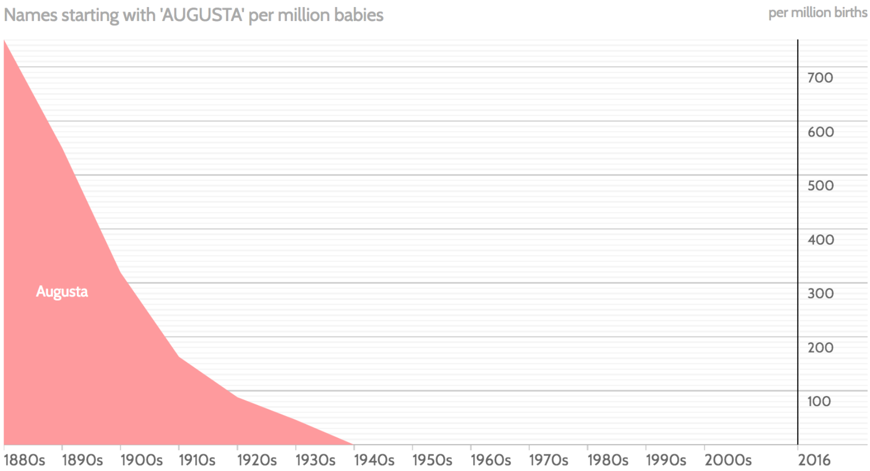
Other examples: *Annette*, *Amy*, *Pamela*, *Timothy*, *Alexandra*, and *Derek*. Students may be able to guess where the peaks of these names are based on the ages of people they know who have them.

FIGURE 4
Output Graph from Baby Name Voyager of a Revived Name (*Adelaide*)
(U-shaped graph showing a dip in popularity before being revived)



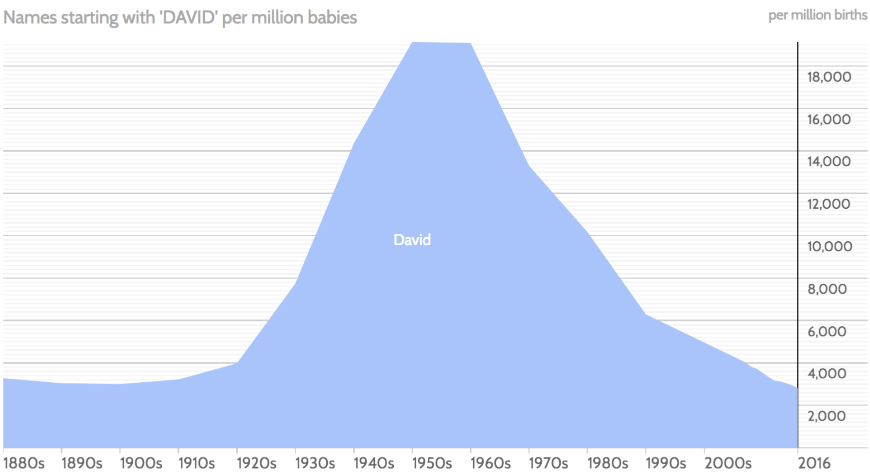
Other examples: *Clara*, *Olive*, *Gwendolyn*, *Henry*, *Hazel*, *Eleanor*.

FIGURE 5
Output Graph from Baby Name Voyager of a Obsolete Name (*Augusta*)
(falling graph that does not [yet?] show a resurgence)



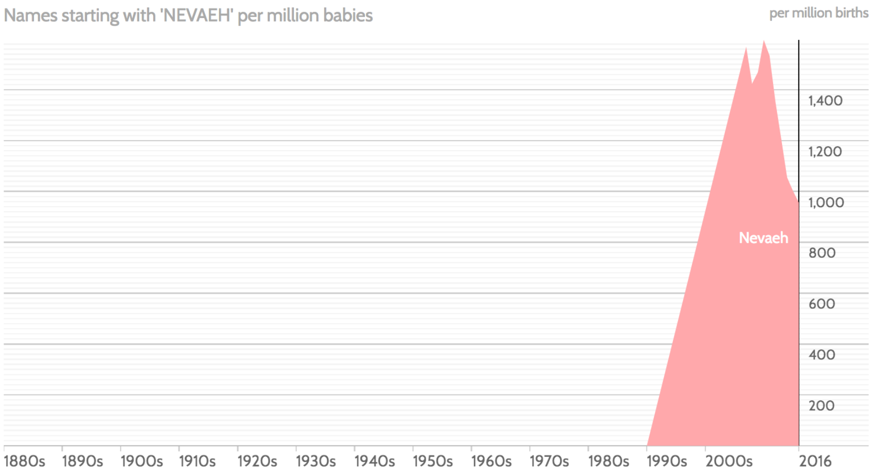
Other examples: *Augusta*, *Gilbert*, *Gladys*, *Ruby*, *Cleo*, *George*, *Irene*, *Polly*.

FIGURE 6
Output Graph from Baby Name Voyager of a Perpetually Popular Name (*David*)
(graph without dips approaching the *x*-axis)



Other examples: *John*, *Elizabeth*, *Thomas*. While these graphs do show an assortment of shapes, the goal of this question is for students to learn to read the scale accurately.

FIGURE 7
Output Graph from Baby Name Voyager of a New Name (*Nevaeh*)
(graph showing a recent peak with negligible results earlier)



Other examples: *Declan*, *Kayden*, *Adalyn*.

CLASSROOM ACTIVITY: DAY 2. The students have now been introduced (or reintroduced) to *x*- and *y*-axes, axis scales, pattern identification (to some extent), dependent and independent variables, and hypothesis testing. Day 2 introduces a second independent variable to the visualization process. It does so by means of another online tool: the Zato Novo Interactive Baby Name Visualizer (Rowe, n.d.): http://zatonovo.com/dataviz/baby_names. This tool uses first-name data from the U.S. Census Bureau and divides it by state as well as by year. The data are visualized on a map of the United States (including all 50 states plus Puerto Rico) that is animated to show change from year to year. The first year displayed is the first for which a particular name is attested in the U.S. according to the data; then each subsequent year of attestation is shown as a new frame of the animation. For instance, the graph for *Marisa* begins in 1952; each frame after that shows a subsequent year of baby girls born in the U.S. recorded as being named *Marisa*, up to 2012.

A cautionary note is that it is not only the data visualization that is different, but also the units of measurement. While the Baby Name Voyager displayed results in proportions of a particular name per million births in the United States, the Zato Novo Interactive Baby Name Visualizer uses the percentage that a name represents of all of the births in a particular state. While this requires additional care from the instructor—it is essential to point out that the results in one database are not directly comparable to those in the other—the discrepancy in terms of measurement is useful training for the students given that they will go on to see numerical results displayed in multiple ways (as proportions, as normalized frequencies, as raw token counts, etc.).

The major goal of using the Zato Novo Interactive Baby Name Visualizer is to have students acquire experience with perceiving and interpreting two independent variables at once. Instructors can describe how having two independent variables leads to three possibilities in terms of effects: a possible effect of time (a name becoming more popular or less so as time goes by), a possible effect of geography (a name being more popular in one part of the country rather than another), and a possible interaction (an effect that is a combination but not simply additive—one that cannot be described with reference to ONLY one of the two factors). With the disclaimer as above that the units are not identical, students can be encouraged to compare, broadly, a name that suddenly attained popularity around the country (e.g., *Amanda*) as it appears on both visualization tools (see figures 8 and 9).

In order to prompt the students to find a nice example of an interaction, the instructor can ask them to look for, for example, a name that was once very Southern-sounding but is now found more broadly, geographically speaking (*Scarlett* is a tempting example; *Eloise* seems to be another). Additional

FIGURE 8
A Frame from the Zato Novo Interactive Baby Name Visualizer Showing a Name (*Amanda*) Suddenly Catching on in the United States in 1979 without Any Particular Geographical Clustering

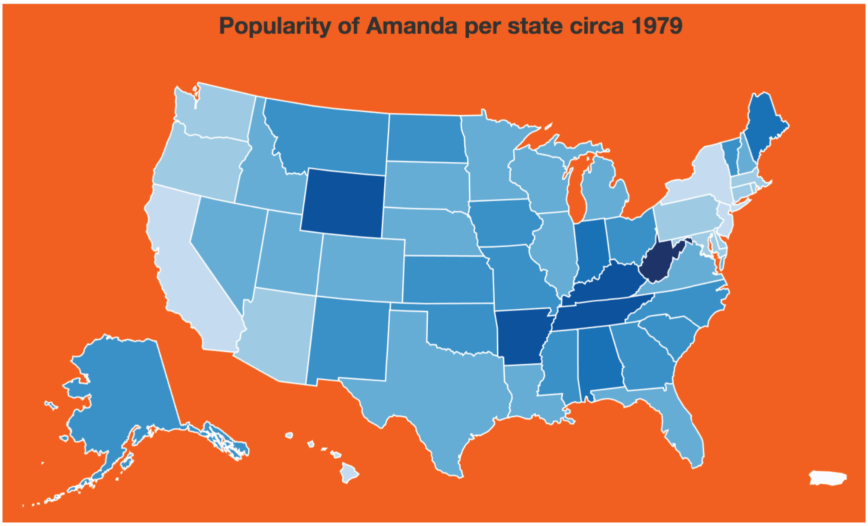
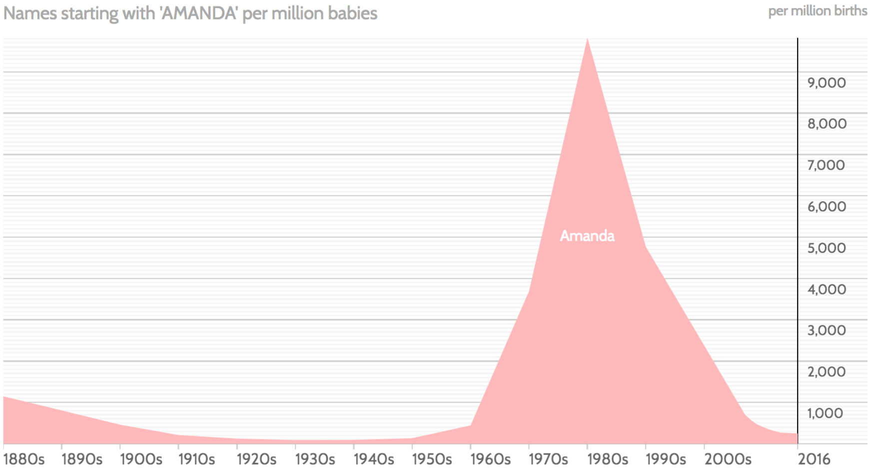


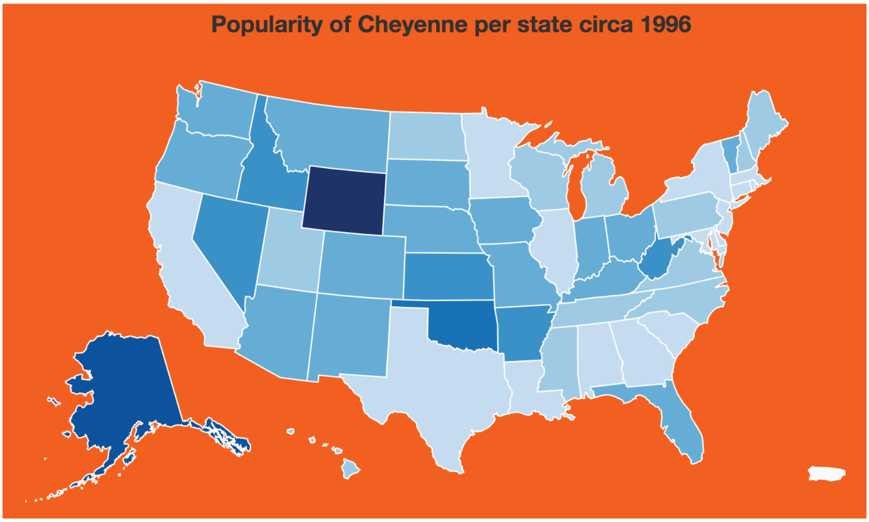
FIGURE 9
Output Graph from Baby Name Voyager Showing a Name (*Amanda*) Catching On in Time (i.e., with Geography Collapsed)



discussion at this point, spurred on by means of some well-chosen examples, can have the students considering why the results look the way they do. Why do names tend to reach California, New York, and Illinois so quickly? Why do states such as Wyoming and Montana often seem reticent to adopt changes occurring around them? For that matter, what is an example of a name for which Wyoming could be expected to show idiosyncratic behavior relative to its neighbors and perhaps all the other states? Figures 10 and 11 show the unique pattern of *Cheyenne*'s popularity.

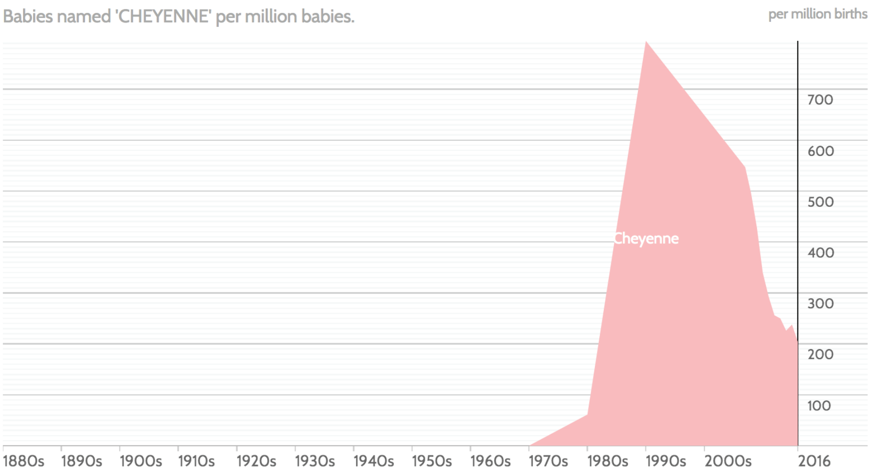
Unpacking the apparent *Cheyenne* or *Scarlett* effect—or that of another example that has caught the attention of the class—gives the instructor the opportunity to introduce subsequent steps inherent in analysis and interpretation. At this point in the course, testing for statistical significance has not yet been introduced, but students can be encouraged to adopt the habit of asking whether the results are reliable (if they do, speak to what they claim to involve—in terms of the amount of data, the units of measurement, how they are presented on a graph, and so on). If so, what might account for them? It should be emphasized that both steps are necessary. Simply looking at data is not tantamount to analyzing data. If the results are not dependable (or are presented in a way that either obscures or exaggerates patterning), then they are shaky ground for a conclusion. At the same time, an interpretation is necessary: it is acceptable to take a shot at explaining the results and end up being incorrect, but stopping at description is not enough.

FIGURE 10
A Frame from the Zato Novo Interactive Baby Name Visualizer Showing Part of a Pattern for a Name (*Cheyenne*) in Which Wyoming Is Perpetually Out of Step with Its Neighbors



The name is popular in Wyoming when it is less so in the surrounding states, and vice versa.

FIGURE 11
Output Graph from Baby Name Voyager Showing a Name (*Cheyenne*) Catching On in Time, but because Geography is Collapsed Here, the Fact That Wyoming Is Not Doing the Same Thing as Other States is Invisible



When the students have taken a shot at providing explanations for name patterns, it is time to steer them toward language. What ought to be highlighted at this point is that names and language are both subject to a substantial amount of systematicity (Lieberson 2000; Lieberson and Lynn 2003; Labov 2010, 194–95, 369). They have structured and orderly variation and change; the fact that they involve a good deal of unconscious behavior does not make them random or unpatterned (see Weinreich, Labov, and Herzog 1968). Instructors can ask the students to consider what it means for so many names to show a single peak even though each naming decision was being made individually, by a different set of parents. With respect to figure 1, for example, parents in the first decade of the twenty-first century did not come together and consciously decide that they were going to name fewer babies *Marisa* than the previous decade's parents did. No one went out of their way to ensure that the pattern had exactly one peak; the shape emerged organically due to how trends work in human society. Butters (2001, 202) makes the same point with respect to language variation:

> As Labov and others have demonstrated repeatedly, speakers are often relatively unaware of change in progress until it has been completed. People do not say to themselves things like, "Hey, it would really help us differentiate the sexes in Philadelphia if all the guys would increase the vocalization of /l/."

CONCLUSION

Language in the United States is considerably more complex than first names are, at least in terms of the sheer number of variable phenomena within any particular language or dialect. However, the key commonality is that language and names are both much more orderly than nonlinguists tend to think. They both show regular, largely unconscious patterning according to gender, ethnicity, time, space, and often interactions between them—and these social patterns can be uncovered through the use of large data sets, particularly with the help of tools for visualization and/or analysis.

Although onomastics is a well-developed field in its own right—and the students should by all means be pointed toward the American Name Society in case they are eager to dive further into research on names—from the standpoint of variationist sociolinguistics, first names could conceivably be described as a very loose sort of sociolinguistic variable, with tens of thousands of variants and one token per child born. In this way, they serve as an accessible introduction both to the data-centered approach central to variationist sociolinguistics and to some of the social factors that will make recurring appearances in the course (time, geography, gender, ethnic/racial/cultural background, etc.).

By the end of this (hopefully engaging)[5] introductory activity on names, the students will have cut their teeth on fundamental concepts related to graph reading that they will need to draw on frequently throughout the rest of the class (axes, scales, dependent and independent variables, hypothesis formation and testing, reliability, results and interpretation). They will have examined, through two databases, which factors or interactions might affect how a particular American person is named. The next step is for them to apply their emerging knowledge to the much larger question of which factors or interactions might affect how a particular American person speaks.

NOTES

I am grateful to my students at Michigan State University (LIN471: Sociolinguistics, Fall 2016) for trying out this activity with me and for giving me feedback on it at the end of the semester. I would also like to thank two anonymous reviewers for their valuable suggestions, plus Aaron Dinkin, Madeline Shellgren, and Suzanne Evans Wagner for their input, and J.A. Macfarlane for the conversation about name patterns in the summer of 2016 that serendipitously helped inspire its creation.

1. The only requirement for this activity is familiarity with a range of American first names. International students and/or nonnative speakers of English may find the activities more of a challenge, but they have likely encountered enough of a range of English-language first names to be able to take part in the classroom activities and complete the homework assignment.
2. The Social Security Administration makes available, for each year since 1880 inclusive, a list of each name given to at least five infants born in the United States in that year (Social Security Administration 2016). To find the number of distinct first names among them, I downloaded the data for every year available—1880 to 2016—and used the Word List feature of AntConc (Anthony 2014) to produce a list of separate items. The total was 95,025, spanning *Aaban* to *Zzyzx*.
3. The Baby Name Voyager does not appear to identify the source of its data overtly, but presumably it comes from the U.S. Census and/or the Social Security Administration. The Zato Novo Interactive Baby Name Visualizer specifically names the U.S. Census as its source.
4. The challenge to simply find something interesting risks setting up later temptations to engage in cherry-picking. However, the activity is still overwhelmingly exploratory at this point, and the hypothesis formulation and testing component of the homework assignment is meant to reinforce the idea that post-hoc hypothesizing is inadvisable.

5. I requested voluntary anonymous feedback on this activity in retrospect at the end of the Fall 2016 semester. The number of responses was limited (N = 7), but almost all of them (N = 6) were enthusiastically positive. As I was unable to secure IRB approval in time to collect the data, I refrain from reporting on the results in any more detail. Further knowledge of the reception of this activity among students awaits future iterations.

REFERENCES

Anthony, Laurence. 2014. AntConc. Version 3.4.3. Tokyo, Japan: Waseda University. http://www.laurenceanthony.net.

Butters, Ronald R. 2001. "Chance as Cause of Language Variation and Change." *Journal of English Linguistics* 29.3: 201–13. doi:10.1177/00754240122005332.

Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "The Causes and Consequences of Distinctly Black Names." *Quarterly Journal of Economics* 119.3: 767–805. doi:10.1162/0033553041502180.

Labov, William. 2010. *Principles of Linguistic Change*. Vol. 3, *Cognitive and Cultural Factors*. Malden, Mass.: Wiley-Blackwell.

Lieberson, Stanley. 2000. *A Matter of Taste: How Names, Fashion, and Culture Change*. New Haven, Conn.: Yale University Press.

Lieberson, Stanley, and Freda Lynn. 2003. "Popularity as a Taste: An Application to the Naming Process." *Onoma* 38: 235–76. doi:10.2143/ONO.38.0.2002560.

Rowe, Brian Lee Yung. n.d. "Interactive Baby Name Visualizer." *Zato Novo*. http://zatonovo.com/dataviz/baby_names.

Social Security Administration. 2016. "National Data on the Relative Frequency of Given Names in the Population of U.S. Births Where the Individual Has a Social Security Number." Social Security Administration. https://www.ssa.gov/oact/babynames/limits.html.

VanderSchans, Arielle. 2015. "The Role of Name Choice in the Construction of Transgender Identities." *Western Papers in Linguistics* 1.2: article 2. http://ir.lib.uwo.ca/wpl_clw/vol1/iss2/2/.

Wattenberg, Laura. n.d. "Baby Name Voyager." *Baby Name Wizard*. http://www.babynamewizard.com/voyager.

Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. "Empirical Foundations for a Theory of Language Change." In *Directions for Historical Linguistics: A Symposium*, edited by W.P. Lehmann and Yakov Malkiel, 95–195. Austin: University of Texas Press.

MARISA BROOK is an assistant professor in the Department of Linguistics and Germanic, Slavic, Asian, and African Languages at Michigan State University. Her research is grounded in variationist sociolinguistics, with an emphasis on morphosyntactic and discourse-pragmatic phenomena. She is particularly interested in changes that are interdependent or otherwise linked. Email: brookma1@msu.edu.

DOI 10.1215/00031283-4202064