What a 166-Year-Old Data-Driven Story Can Teach Journalists Today

SCOTT KLEIN[*]

It's easy to think of data journalism as a modern invention. With all the hype, a casual reader might assume that it was invented sometime during the 2012 presidential campaign. Better-informed observers can push the start date back a few decades, noting with self-satisfaction that Philip Meyer did his pioneering work during the Detroit riots in the late 1960s. Some go back even further, archly telling the tale of Election Night 1952, when a UNIVAC computer used its thousands of vacuum tubes to predict the presidential election within four electoral votes (Henn 2012; Chinoy 2010; American Presidency Project).

But all of these estimates are wrong – off, in fact, by centuries. The real history of data journalism pre-dates newspapers, and traces the history of news itself. The earliest regularly-published periodicals of the 17th century, little more than letters home from correspondents hired by international merchants to report on the business details and the court gossip of faraway cities, were data-rich reports (Hutt 1973).

Early 18th century newspapers were also rich with data. If it were ever in doubt that the unavoidable facts of human existence are death and taxes, early newspapers published tables of property tax liens and of mortality and its causes. Commodity prices and the contents of arriving ships — cargo and visiting dignitaries — were a regular and prominent feature of newspapers throughout the 18th and 19th centuries.

Beyond business figures and population statistics, data was used in a wide variety of contexts. The very first issue of the *Manchester Guardian* on May 5, 1821 contains on the last of its four pages a large

---

[*] Assistant Managing Editor, ProPublica.

table showing that the real number of students in church schools far exceeded the estimates placed on the student population made by proponents of education reform (Rogers 2011).

Data was also used, as it is today, as both the input to and the output of investigative exposés. This is the story of one such investigative story, and of its author, *New York Tribune* editor Horace Greeley. It's a remarkable tale, and one with important lessons for "big data" journalism today.

Though he's no longer a household name, Horace Greeley was one of the most important public figures of the 19th century. His *Tribune* had a circulation larger than any paper in the city except for cross-town rival James Gordon Bennett's *New York Herald* (Unsigned 1876). More than 286,000 copies of the *Tribune*'s daily, weekly and semi-weekly editions were sold in the city and across the country by 1860, which by its own reckoning made it the largest-circulation newspaper in the U.S. (Greeley 1860). Ralph Waldo Emerson observed, "Greeley does the thinking for the whole West at $2 per year for his paper." (Hale 1950)

Greeley himself was a popular public speaker and a hugely influential national figure. He was a fascinating, frustrating, contradictory man. He was a leading abolitionist whose support for the Civil War was limited at best, yet his abolitionist writing in the *Tribune* made the paper the target of an angry mob during the Draft Riots in 1863 (Unsigned 1883). He was a vegetarian and a utopian socialist who published Karl Marx in the *Tribune*, but believed fervently in manifest destiny and America's western expansion. He was a New York icon who thought the city was a terrible influence on working people and encouraged them to "Go West" to escape it (Greeley 1871). Though he was one of the founders of the Republican Party, his relationship with Abraham Lincoln was strained, and he ran for president in 1872 on what amounted to the Democratic ticket, losing big and dying broken-hearted before the Electoral College could meet to certify Grant's election (Kluger 1986).

Long before his presidential campaign, and for decades, Greeley and his paper held sway with hundreds of thousands of everyday Americans. But if he was a celebrity with the people, he was far less successful convincing political elites to sponsor his entry into political office. His moralism and mercurial nature seem to have been a steady annoyance to powerful figures like New York's William Seward and Whig (and later Republican) party boss Thurlow Weed. Historian Richard Kluger noted of their relationship,

> "[Greeley] was more useful to [Seward and Weed] than they
> ever proved to him. As the eloquent editor of a rising
> newspaper that reached, through its weekly edition,
> throughout the Empire State, Greeley was a lively fish on the
> hook, to be fed enough line to thrash about picturesquely until
> reeled in tightly during campaign season." (Kluger 1986)

It was perhaps out of a desire to shut Greeley up — and yet also a recognition of the care necessary when dealing with a man, as *The Nation* put it, "with a newspaper at his back" (1872) — that the Whigs nominated Greeley to fill a temporary vacancy in the House for the second session of the 30th Congress in 1848 (Greeley 1868). The session would last only three months, and Greeley's Congressional career would end when the term did. But what Greeley did with his time was remarkable.

By the middle of the 1800s, Congressmen's compensation for travel to and from their districts had been an unsuccessful but simmering reform target for years. The law provided for a 40-cent per-mile mileage reimbursement, and computed the distance "by the usually travelled route." after taking his seat, Greeley got a look at the schedule listing every congressman's mileage and was shocked by the sums. To Greeley, the disbursements were a wasteful relic of an earlier time, when travel to and from the far-flung reaches of the United States would have been a costly, bruising affair. The 40-cent mileage had been calculated decades earlier to match a pre-1816 congressman's pay rate of $8 a day, assuming he could travel a mere 20 miles per day (Greeley 1868). However, thanks to steamships and the increasing prevalence of trains, travelers could go far faster than that.

Greeley saw it as an outrageous waste of the taxpayer's money, and deployed his newspaper to correct that wrong. "If the route usually travelled from California to Washington is around Cape Horn — or the Members from that embryo State shall choose to think it is — they will each be entitled to charge some $12,000 Mileage per session accordingly." (Greeley and Howard 1848)

Rather than simply opining against it, he conceived and published a data-journalism project that, in form if not in execution, would be very much at home in a newsroom today. He asked one of his reporters, Douglas Howard, a former postal clerk, to use a U.S Post Office book of mail routes to calculate the shortest path from each congressman's district to the Capitol, and compared those distances with each congressman's mileage reimbursements (Blair and Rives 1849). In Dec. 22, 1848, with Greeley now simultaneously its editor

and a brand new congressman from New York, the *Tribune* published a story and a table in two columns of agate type. The table lists each congressman by name with the mileage he received, the mileage the postal route would have granted him and the difference in cost between them. "Let no man jump at the conclusion that this excess has been charged and received contrary to law," wrote Greeley in the accompanying text. "The fact is otherwise. The members are all honorable men — if any irrelevant infidel should doubt it, we can silence him by referring to the prefix to their names in the newspapers." (Greeley and Howard 1848)

It wasn't his colleagues Greeley inveighed against, but rather, he claimed, the system. "We assume that each has charged precisely what the law allows him and thereupon we press home the question — '*Ought not* THAT LAW *to be amended?*'" (Greeley and Howard 1848) Among the accused stood Abraham Lincoln, in his only term as congressman (U.S. Congress 2005). Lincoln's travel from faraway Springfield, Illinois, made him the recipient of some $650 in excess mileage — about $18,500 today — among the House's worst (Federal Reserve Bank of Minneapolis 2014; Greeley and Howard 1848). Beside Lincoln, Greeley's findings include a list of historical legends, including both of Lincoln's vice presidents — Hannibal Hamlin, who took only an extra $64.80 to go between Washington and Maine, and Andrew Johnson, who got $122.40 extra to get to the Capitol and back from Tennessee. Daniel Webster received $72 extra for travel to and from the Senate from Massachusetts. John C. Calhoun and Jefferson Davis were recipients of an extra $313.60 and $736.80, respectively, for round-trip travel from South Carolina and Mississippi (Greeley and Howard 1848). The excesses track roughly according to distance from Washington. Isaac Morse, a Democrat from Louisiana whose journey comprised some 1,200 miles by postal route, received 2,600 miles in mileage from the House. A helpful if imprecise note, I assume written by Greeley, offered: "Only 409 miles less than to London." (Greeley and Howard 1848)

NEWS TRAVELS RELATIVELY FAST

It took about five days for the story to travel from New York to the rest of the country. One particularly laudatory Greeley biographer reported that "the effect of [the mileage expose] upon the town was immediate and immense. It flew upon the wings of the country press, and became, in a few days, the talk of the nation." (Parton 1889) On Dec. 27, the story broke loose in the House. The *Congressional Globe* recorded the eruption on the floor. Ohio Democratic Rep. William

Sawyer ($281.60 in excess charges) raised a point of order, accusing Greeley of "a species of demagoguism of which he could never consent to be guilty while he occupied a seat on this floor, or while he made any pretensions to stand as an honorable man among his constituents." (Blair and Rives 1849)

A heated exchange followed with nearly all speakers standing against Greeley, led by Sawyer and Thomas J. Turner, D.-Ill. ($998.40). Most of the charges, according to Turner, were "absolutely false," and Greeley

> "had either been actuated by the low, groveling, base, and malignant desire to represent the Congress of the nation in a false and unenviable light before the country and the world, or that he had been actuated by motives still more base — by the desire of acquiring an ephemeral notoriety, by blazoning forth to the world what the writer attempted to show was fraud. The whole article abounded in gross errors and willfully false statements, and was evidently prompted by motives as base, unprincipled, and corrupt as ever actuated an individual in wielding his pen for the public press." (Blair and Rives 1849)

While the conversation was rich with florid dudgeon, some of the arguments against Greeley appeared more substantive. Turner pointed out that the Postmaster General had stopped using the postal route book Greeley used to compute mileage "in consequence of incorrectness." Greeley countered that the article acknowledges this — though I found no passage indicating this in the Tribune. Others noted that the Mileage Committee independently determined mileage for each member based both on evidence provided by the member as well as on their own research, and that members themselves didn't "charge" anything. (Blair and Rives 1849) To Greeley, this was all beside the point. He defended his story on the floor, pointing out that he didn't charge members with anything fraudulent or illegal nor did he "object to any gentleman's taking that course if he saw fit; but was that the route upon which the mileage ought to be computed?" (Blair and Rives 1849)

Greeley's own mileage is not listed in the table, but he separately told the House that he'd found that his own mileage was overestimated by some $4 – which would match the mileage paid to his predecessor – and that he'd corrected the matter with the House Sergeant-at-Arms (Unsigned 1849c; Greeley and Howard 1848). If

opinions among his House colleagues ranged from annoyed to apoplectic, opinion among America's newspapers seem to have been largely supportive of Greeley. "The election of Mr. Greeley to the House seems likely to produce good," ran an editorial in the *New York Evening Post* the next week. "He has already rendered the people an important service by exposing the fraudulent manner of calculating and paying mileage." (Unsigned 1848) The *Eastern Carolina Republican* damned Greeley with faint praise, saying that he'd "had hit upon a practical reform for once in his life." (Unsigned 1849a)

Greeley had "set down the excess to their *honor*," added the Sandusky (Ohio) *Clarion*. "This was not altogether a judicious move, for Mr. Greeley, as a member of this House, especially considering how extravagantly nice some of these bloated crib-suckers are about honor." (Unsigned 1849b) "I had expected that it would kick up some dust," Greeley later wrote in his autobiography, "but my expectations were far outrun." He called the affair the "mileage swindle," and labeled the members "wounded pigeons" and their excuses a "shabby dodge." (Greeley 1868) A few weeks into the scandal, he wrote in the *Tribune*:

> "Members who have taken long Mileage generally had nothing to do with settling the distance; while the Committee say they applied to the members generally, and failing a response, did the best they could. That old rascal Nobody is again at his capers! He ought to be indicted." (Greeley 1849a)

LESSONS LEARNED

Though it's 166 years old and largely forgotten, Greeley's mileage story has resonance — and lessons — for data journalists today:

First, open records are important for journalists, and they're absolutely essential for data journalists. Greeley was able to use his status as a sitting congressman to get access to the data for the story, "certifying that it was wanted as the basis of action in the House." (Greeley and Howard 1848) But a law granting access to government documents wasn't put in place until the Freedom of Information Act was signed almost 100 years after Greeley's death. Notably, Congress has exempted itself from FOIA (Meyer 2013). While is isn't perfect, journalists and researchers today can count on getting data from the government much more easily than they could in Greeley's day.

Second, data journalists must be cautious about the powerful stories raw data can tell on its own. Greeley might have known he was

being provocative by publishing the names as he did, and his protestations that "there was no imputation in the article upon any member, that he had made illegal charges" (Blair and Rives 1849) seem a bit implausible. Indeed, the story Greeley wrote accompanying the long table insists that the target of his investigation was the outdated law and not any particular congressman. But that's neither how it was taken in the House nor in the country.

Then, as now, raw data isn't raw. It comes with biases and reflects the choices made about the methods used to create and analyze it. It can also tell its own story and mislead people into inferring things that the facts don't support. As journalists, we must understand and make conscious, fair choices about what we're doing when we put names next to numbers. And we must at all points give context — not just in an attached story, but located near the data itself. Greeley made an argument in the form of a statistical table and people across the country — even sophisticated newspaper opinion writers — concluded that the Congress was on the take. The numbers can speak for themselves, but that doesn't mean it's always clear what they're saying.

Also, it's just as important for data journalists to confirm their stories with actual humans as it is for traditional reporters to do so. Telephones hadn't been invented yet when Greeley published the story, but it doesn't seem as if Greeley tried contacting the Committee on Mileage to make sure his methodology was sound. Critics on the floor of the House revealed flaws in Greeley's story that would have been devastating in today's environment of instantaneous social-network media criticism. Greeley should also have reached out to Congressmen he singled out to give them a chance to respond pre-publication. "In case the design of the writer had been to act fairly in the matter," asked Rep. Sawyer, "why he had not taken the trouble to ascertain the facts?" (Blair and Rives 1849)

The table printed in the *Tribune* is rife with misspelled names, arithmetic errors, a missing entry and what must have been typographic errors introduced when typesetting the complex columns of numbers. Greeley and his coauthor published a series of corrections and clarifications over the next few months (Greeley 1849b; Howard 1849). Howard later called the errors inevitable "in a computation involving over half a million of figures, and executed in a very brief space of time." (Howard 1849) But with modern computing supporting us, data journalists today have a far higher bar for accuracy. Bulletproofing is a critical part of the editorial process of any data story and it must never be skipped.

All that said, Greeley's work had its intended effect. The House continued to grouse about the story but passed a bill that session by a vote of 158 to 16 to change the computation of mileage to "the shortest continuous mail route" (*An Act to Regulate the Mileage of Members and Delegates in Congress, and for Other Purposes* 1849) -- though, Greeley later wrote, with "a distinct understanding that the Senate would kill it." (Greeley 1849c) In his autobiography Greeley reported that Congress later lowered the per-mile rate to twenty cents, and though the "usually travelled route" language remained, he conjectured that the spread of the railroads shortened that route to something comparatively reasonable (Greeley 1868).

It is perhaps a fitting coda to this story that, although transportation has gotten faster and easier than Greeley could have imagined, congressional mileage calculations remain, though in quite different form. Unlike in 1848, when members of Congress were personally paid the mileage payments, district travel funds are now part of each member's overall expense budget. They're calculated using a per-mile rate that increases with proximity to D.C. The highest rate, which would have applied to Greeley's Manhattan district, is 96 cents, more than double the rate in Greeley's day (Brudnick 2014).

REFERENCES

American Presidency Project. "Election of 1952." *The American Presidency Project*. http://www.presidency.ucsb.edu/ showelection.php?year=1952.

An Act to Regulate the Mileage of Members and Delegates in Congress, and for Other Purposes. 1849.

Blair and Rives. 1849. *The Congressional Globe: New Series, Containing Sketches of the Debates and Proceedings of the Second Session of the Thirtieth Congress*. Washington, D.C.: Blair & Rives.

Brudnick, Ida A. 2014. *Congressional Salaries and Allowances*. RL30064. Congressional Research Service. http://www.senate.gov/CRSReports/crs-publish.cfm? pid=%270E%2C*PL%5B%3D%23P%20%20%0A.

Chinoy, Ira. 2010. "Battle of the Brains: Election-Night Forecasting at the Dawn of the Computer Age". PhD. Dissertation. Philip Merrill College of Journalism. http://drum.lib.umd.edu/ bitstream/1903/10504/1/Chinoy_umd_0117E_11395.pdf.

Federal Reserve Bank of Minneapolis. 2014. "Consumer Price Index (Estimate) 1800-." Accessed June 22. http:// www. minneapolisfed.org/community_education/teacher/calc/hist1 800.cfm.

Greeley, Horace. 1849a. "The Mileage Matter." *The New York Daily Tribune*, January 16.

———. 1849b. "Letter to the Editor." *The New York Daily Tribune*, January 16.

———. 1849c. "Mileage Reform: Letter from Hon. Horace Greeley." *The Huron Reflector*, February 13.

———. 1860. "The New-York Tribune: Progress Vol. XX (Advertisement)." *The New York Times*, April 11.

———. 1868. *The Autobiography of Horace Greeley, or Recollections of a Busy Life*. E.B. Treat.

———. 1871. "Letter from Greeley to R.L. Sanderson, 1871", November 15. http://www.gilderlehrman.org/history-by-era/development-west/resources/horace-greeley-%E2%80%9Cgo-west%E2%80%9D-1871.

Greeley, Horace, and Douglas Howard. 1848. "The Mileage of Congress." *The New York Daily Tribune*, December 22.

Hale, William Harlan. 1950. *Horace Greeley: Voice of the People*. New York: Harper & Brothers. https://archive.org/stream/horacegreeleyvoi009588mbp#page/n5/mode/2up.

Henn, Steve. 2012. "The Night a Computer Predicted the Next President." *NPR: All Tech Considered*. http://www.npr.org/blogs/alltechconsidered/2012/10/31/163951263/the-night-a-computer-predicted-the-next-president.

Howard, Douglas. 1849. "The Mileage Controversy." *The New York Daily Tribune*, January 18.

Hutt, Allen. 1973. *The Changing Newspaper: Typographic Trends in Britain and America, 1622-1972*. London: Gordon Graser.

Kluger, Richard. 1986. *The Paper: The Life and Death of the New York Herald-Tribune*. Knopf.

Meyer, Theodoric. 2013. "Do As We Say, Congress Says, Then Does What It Wants." *ProPublica*. http://www.propublica.org/article/do-as-we-say-congress-says-then-does-what-it-wants.

Parton, James. 1889. *The Life of Horace Greeley*. Cambridge, Mass.: The Riverside Press.

Rogers, Simon. 2011. "The First Guardian Data Journalism: May 5, 1821." *DataBlog: Facts Are Sacred*. http://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian.

U.S. Congress. 2005. *Biographical Directory of the United States Congress 1774-2005*. Washington, D.C.: U.S. Gov't Printing Office.

Unsigned. 1848. "Correspondence of the Evening Post (Signed 'B')." *The (New York) Evening Post*, December 29.

———. 1849a. "Mr. Greely [sic] in Hot Water." *The (Eastern Carolina) Republican*, January 3.

———. 1849b. "From Washington: Correspondence of the Sandusky Clarion (Signed 'JC')." *The Sandusky (Ohio) Clarion*, January 15.

———. 1849c. "Mileage from New York - A Dry Haul." *The New York Daily Tribune*, January 18.

———. 1872. "The Death of Mr. Horace Greeley." *The Nation*, December 5.

———. 1876. *The New York Tribune: Where and How It Is Made*. Philadelphia. http://babel.hathitrust.org/cgi/pt?id=nnc1.cu57712441;view=1up;seq=53.

———. 1883. *The New York Tribune: A Sketch of Its History*. New York.