Storyboarding Medicare Fee-for-Service Fraud

Devansh Naimish Patel

Ottawa University

Abstract

This paper documents a comprehensive, end-to-end application of Cole Nussbaumer Knaflic's Storytelling with Data framework to a Medicare Fee-for-Service (FFS) fraud detection project. Using publicly available Centers for Medicare & Medicaid Services (CMS) Part D prescriber and drug data, I move systematically through six stages: understanding the context (Chapter 1), choosing appropriate displays (Chapter 2), eliminating clutter (Chapter 3), drawing attention where it matters (Chapter 4), thinking like a designer (Chapter 5), and telling a coherent story (Chapter 7). The analysis culminates in a Tableau dashboard that synthesizes three key visuals: a scatterplot of high-cost outlier drugs in days' supply, a prescriber-level view of brand dependence versus cost, and a box plot of long days' supply by drug. Each step is grounded in explicit design decisions around audience, form, function, and narrative flow. The resulting visuals are not merely technically correct; they are structured as a story in which compliance and policy stakeholders are the protagonists, the data provides conflict and stakes, and the dashboard delivers a clear call to action. This paper illustrates how the storytelling with data process can transform raw Medicare claims data into decision-ready evidence that supports fraud triage, resource prioritization, and more informed policy discussion.

Storyboarding Medicare Fee-for-Service Fraud

Fraud, waste, and abuse in Medicare Fee-for-Service programs pose a substantial threat to the financial sustainability of the system and to public trust in health care institutions. At the same time, the sheer volume of claims and prescribing data makes it difficult for compliance and policy teams to distinguish signal from noise. Traditional reporting like dense tables, static reports, or isolated ad-hoc analyses often fails to answer the most important questions: Where is risk concentrated? Which prescribers and drugs warrant priority review? How large is the exposure if nothing changes?

This project applies the principles in Cole Nussbaumer Knaflic's Storytelling with Data: A Data Visualization Guide for Business Professionals (2015) to the Medicare FFS fraud context. Rather than starting with tools or chart types, I begin with the core premise that form must follow function: visuals exist to help a specific audience make better decisions. Over seven weeks, I moved from understanding the context to building prototype charts, refining them through decluttering and preattentive attributes, thinking like a designer, crafting a narrative, and finally assembling a dashboard that tells a coherent story.

The remainder of this paper is organized by the book's key chapters used in the course: understanding the context (Chapter 1), choosing an appropriate display (Chapter 2), eliminating clutter (Chapter 3), drawing attention where you want it (Chapter 4), thinking like a designer (Chapter 5), and telling a story (Chapter 7). Within each section, I describe how the concept was applied to the Medicare FFS fraud use case, how the visual design evolved, and how decisions were driven by the needs of our primary audience who are Medicare policy and compliance decision makers.

Data and Audience

The data used for this project comes from the CMS Medicare Part D Prescriber Public Use Files, which report drug-level and prescriber-level utilization and cost metrics, including total claims, total 30-day fills, total drug cost, and beneficiary counts for each combination of National Provider Identifier (NPI) and drug. These data sets are well suited for high-level fraud and abuse screening because they provide visibility into cost per claim, cost per 30-day equivalent, and distribution of prescribing across brand and generic drugs.

The primary audience for this work is an internal CMS or health-plan compliance and policy team. These stakeholders are busy, highly analytical, and already familiar with concepts like outlier detection, risk stratification, and medical review. However, they may not have time to sift through large tables or exploratory dashboards. Their information needs are focused: they want to know where to look first, how large the potential exposure might be, and what actions should be prioritized. Throughout the project, I treated this audience as the protagonist of the story, an emphasis consistent with Knaflic's guidance to make the audience, not the analyst, the main character (Knaflic, 2015).

Understanding the Context (Chapter 1, Week 1)

Chapter 1 emphasizes the need to understand who the audience is, what they need, and how the information will be consumed before building any chart. For this project, the central questions driving the context were:

1.1. Which drugs create disproportionate cost exposure in Medicare Part D?

1.2. Which prescribers have patterns of high brand dependence and high cost per claim?

1.3. Which drugs exhibit unusual duration of therapy, as measured by days' supply per claim?

Initial stakeholder interviews and course reflections suggested that the compliance audience prefers short, visually rich decks and dashboards that can be used in monthly risk review meetings. They are less interested in detailed statistical modeling outputs and more interested in clear, defensible visuals that can be shared with leadership or external auditors.

Context also includes constraints. The analysis needed to rely entirely on publicly available data; no protected health information or internal plan data were used. Furthermore, the project timeline forced an iterative approach: each week would focus on one aspect of the storytelling with data framework, and deliverables had to be working visual prototypes rather than purely theoretical designs.

These context considerations shaped the rest of the project. They led me to focus on high-level metrics, such as cost per 30-day fill and brand share, that could be understood quickly, while still being meaningful from a fraud and waste perspective. They also encouraged the use of Tableau Desktop as the main visualization tool because it supports iterative refinement, interactive dashboards, and the ability to share packaged workbooks with non-technical stakeholders.

Choosing an Appropriate Display (Chapter 2, Week 2)

Chapter 2 discusses the strengths and weaknesses of common chart types: tables, bar charts, line charts, scatterplots, and others. During Week 2, I experimented with a variety of simple charts using smaller practice data sets to understand how different forms served different analytical questions.

When I began working directly with the Medicare Part D data, these lessons guided the selection of displays. For the drug-level cost question, the core relationship of interest was between cost per 30-day equivalent and days' supply per claim, with a need to encode total drug cost as a third variable. A scatterplot with bubble size was therefore the most appropriate display. Bar charts could have shown rankings, but they would have failed to show how cost and duration interacted; tables would have been even harder to interpret.

For the prescriber-level question, I wanted to compare brand share by NPI against cost per claim while also hinting at volume via claims count. Again, a scatterplot was the natural choice. The horizontal axis represented brand share, the vertical axis cost per claim, and bubble size encoded total claims. This form made trade-offs visually clear: some high-cost prescribers had low brand share, while others combined both high brand share and high costs.

For the duration-of-therapy question, the shape of the distribution mattered as much as the average. A simple bar chart of mean days' supply would obscure outliers and variability. Following Knaflic's guidance on when to use box plots, I selected a box-and-whisker plot to show the distribution of days' supply per claim for each drug, with the median line and whiskers visually highlighting spread and extreme values. These chart choices were not arbitrary; they directly followed the "form follows function" principle. Each visual form was selected because it best answered the specific question the audience needed to be able to ask of the data.

Eliminating Clutter (Chapter 3, Week 3)

Once the initial charts were built, they contained all the usual Tableau defaults: heavy gridlines, bold axis ticks, long legends, and unnecessary colors. Chapter 3 argues that such clutter, anything that does not contribute to the message, should be ruthlessly removed. During Week 3, I applied a systematic decluttering process to each of the three charts.

In the High-Cost Outlier Drugs in Days' Supply scatterplot, I removed vertical gridlines entirely and lightened the horizontal ones, since only the relative position on the cost axis mattered for interpretation. Axis titles were rewritten in plain language ("Days' supply per claim" instead of a cryptic field name), and tick labels were set at sensible intervals rather than every minor value. The color palette was reduced to a single teal hue for all points, and annotations were limited to the most extreme drugs, such as semaglutide and apixaban, that merited specific callouts.

In the Brand Dependence vs Cost (NPI) scatterplot, decluttering involved graying out the majority of prescribers and simplifying the legend. Initially, color encoded multiple categories and the caption repeated information already present in the chart. By Week 3, color was reserved only for indicating whether a prescriber met the high-risk criteria, while all other points were muted into a light gray. Gridlines were minimized, axis titles clarified, and a single concise caption replaced verbose text.

The Top Long Days' Supply by Drug box plot was initially difficult to read: drug names overlapped, whiskers were thin, and background gridlines competed with the data. Decluttering included rotating and abbreviating drug labels, removing unnecessary gridlines, widening the boxes slightly for visibility, and limiting annotations to the median and a small set of extreme outliers. These steps embodied Knaflic's recommendation to "remove to improve" (2015). By

stripping away nonessential elements, each chart began to communicate more clearly and quickly.

Drawing Attention Where You Want It (Chapter 4, Week 4)

After decluttering, the next challenge was directing attention. Chapter 4 explains how preattentive attributes like color, size, position, and, form and can be used strategically so that the viewer's eye is drawn to the most important aspects of a visual. Week 4 focused on applying these attributes to the three Medicare FFS charts.

For the high-cost outlier drugs scatterplot, I used size and annotation as primary signals. Bubbles were sized by total drug cost so that drugs responsible for the greatest spending automatically appeared larger. Semaglutide and apixaban, for example, pulled the eye due to both their position in the upper-right of the plot and their larger size. I added simple, high-contrast annotations next to these points to show generic name, cost per 30-day field, days' supply per claim, and total drug cost.

In the Brand Dependence vs Cost chart, color and quadrant structure did most of the work. I created a calculated field, is_high_risk, that flagged prescribers whose brand share and cost per claim exceeded specified thresholds. These prescribers were shown in bright red, while all others remained a neutral gray. Vertical and horizontal reference lines marked the thresholds, and a translucent "OUTLIERS" box in the upper-right quadrant provided a visual label for the high-risk region. As a result, even a cursory glance at the chart made it obvious that only a few red points in a specific quadrant deserved immediate attention.

For the box plot of long days' supply, I used color to distinguish claims or drugs flagged for high days per claim. The median line was thickened and labeled, and the highest-duration drugs were highlighted with red marks against a gray background. These visual choices guided the eye to the extreme values and helped compliance staff quickly identify which drugs might warrant policy interventions such as quantity limits or prior authorization. By intentionally applying preattentive attributes, the charts moved beyond neutrality. They began to behave like visual arguments, silently but forcefully suggesting where the viewer should look first.

Thinking Like a Designer (Chapter 5, Week 5)

Chapter 5 reframes visualization work as design work, emphasizing affordances, accessibility, and aesthetics. In Week 5, I applied these concepts to further refine the Medicare FFS charts.

Affordances refer to visual cues that suggest how something should be used. On the prescriber scatterplot, for example, the combination of red points, quadrant lines, and the "OUTLIERS" label serves as an affordance: the chart invites the viewer to focus on the upper-right region and treat those prescribers as a prioritized list for review. Similarly, the size-based encoding and annotations in the drug-level scatterplot afford rank ordering; they suggest that the viewer can interpret bubble size as importance.

Accessibility improvements focused on ensuring that the charts were usable by people with varying levels of visual acuity and color perception. I chose color palettes with sufficient contrast and avoided problematic red-green combinations. Text labels and axis titles were set in a consistent, legible typeface and sized to be readable when embedded in PowerPoint or printed.

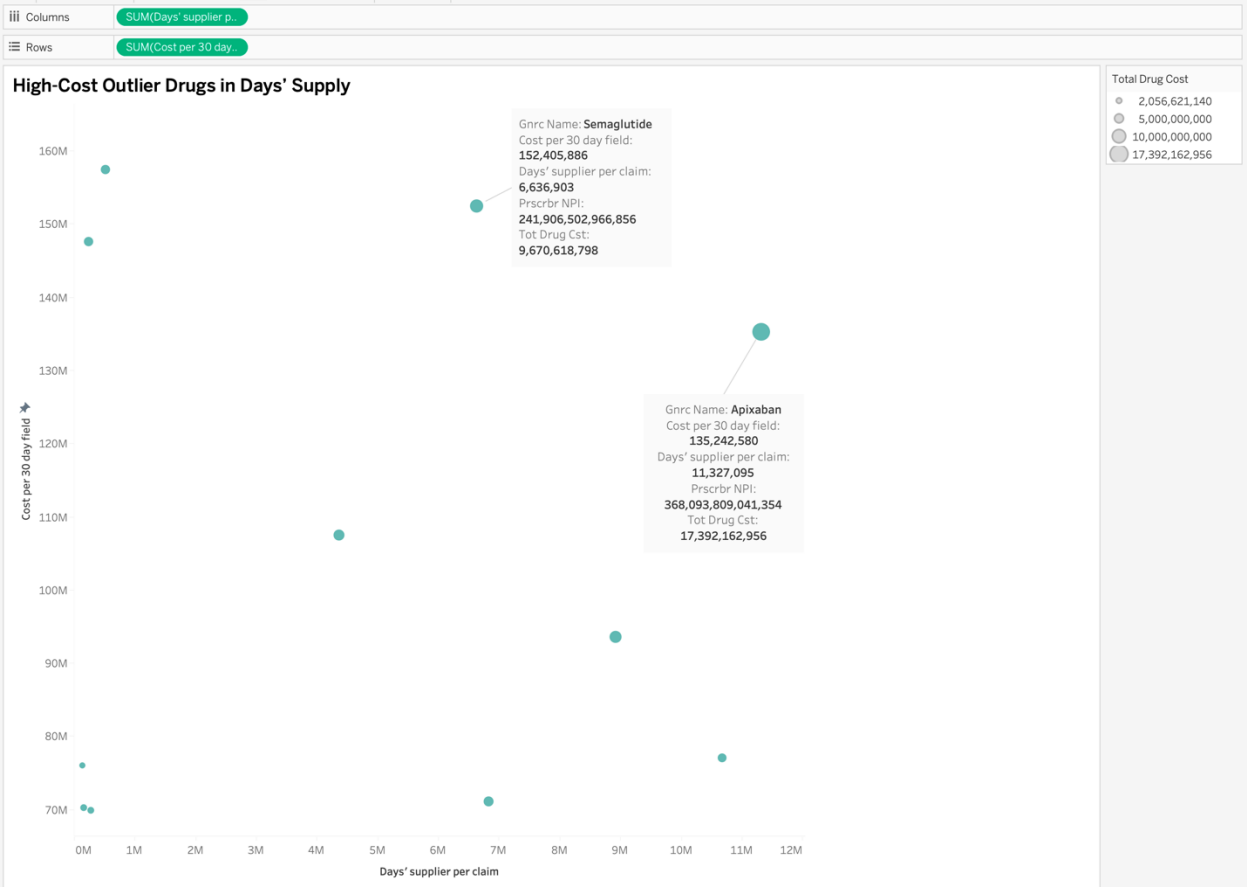
Tooltips were rewritten to show only the most relevant fields in plain language so that a screen reader could convey the essential information without overwhelming the user.

Aesthetics were refined through consistent spacing, alignment, and restrained use of color. I used plenty of white space around and within the charts so that they did not feel crowded. Only two or three colors were used across the entire set of visuals, giving the work a coherent visual identity. While aesthetic decisions may appear cosmetic, they contributed directly to audience acceptance. Research suggests that well-designed visuals are perceived as more credible and professional, increasing the likelihood that stakeholders will trust and act on the information (Few, 2009). Thinking like a designer meant constantly asking not only "Is this accurate?" but also "Is this approachable and trustworthy for my audience?"

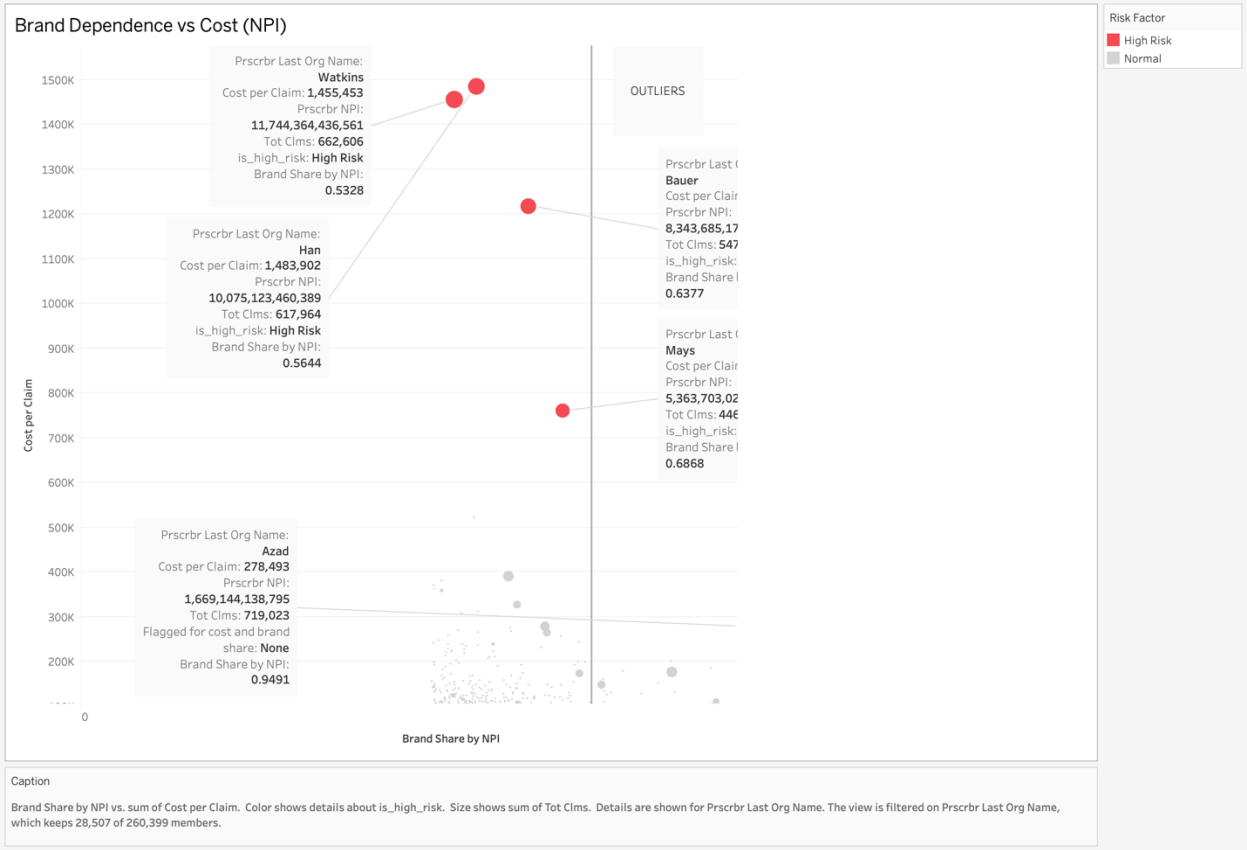
Telling a Story with Data (Chapter 7, Week 6)

Chapter 7 brings the previous lessons together by emphasizing structure and narrative. A compelling story has a beginning, middle, and end; it introduces characters, establishes conflict, and resolves tension. During Week 6, I treated the three Medicare FFS charts as acts in a broader story in which the compliance audience is the protagonist.

The beginning is the drug-level scatterplot (Figure 1), which answers the question, "Where are the extreme cost hot spots at the drug level?" It introduces the world of Medicare Part D spending and reveals that a small number of drugs, such as semaglutide and apixaban, drive disproportionate cost exposure.



The middle is the prescriber scatterplot (Figure 2), which asks, “Who is writing the prescriptions that drive this cost?” Here, the conflict intensifies as high-risk prescribers appear in red in the high-brand, high-cost quadrant. This chart shifts the audience’s attention from products to people and invites them to consider whether they would feel comfortable ignoring these prescribers in their oversight activities.



The end is the box plot of long days' supply (Figure 3), which answers, "Which drugs with long duration of therapy should be prioritized for review?" The distributional view shows that certain drugs have claims with 300–365 days' supply, reinforcing concerns about diversion, stockpiling, or inappropriate long-term use.

This three-act structure, combined with horizontal logic in the slide titles, means that if an executive read only the headlines, "A few drugs drive extreme cost," "A small set of prescribers display high-brand, high-cost patterns," and "Specific drugs exhibit unusually long days' supply" they would still understand the gist of the story. The charts then function as visual evidence that deepens, rather than replaces, the narrative.

Integrating the Story in a Tableau Dashboard (Week 7)

In Week 7, I integrated the three charts and key performance indicators into a single Tableau dashboard titled “Medicare Part D Fraud Risk Overview.” The dashboard contains three horizontal bands: a top band of KPI tiles, a middle band with the two scatterplots side-by-side, and a bottom band with the full-width box plot.

The KPI tiles show total flagged drug cost, the number of high-risk prescribers, and the number of drugs with high days’ supply per claim. These metrics give the audience a quick understanding of the scale of potential exposure. The two middle-band scatterplots allow users to interactively identify high-cost drugs and high-risk prescribers, while the box plot lets them examine duration patterns for selected drugs. Filters for year, geography, and drug class run across all views, and the scatterplots can be used as filters to drive the box plot.

This dashboard operationalizes the story structure developed in earlier weeks. The audience can begin at the KPIs to understand scope, move through the drug- and prescriber-level visuals to identify specific risk clusters, and finish with the box plot to determine where duration-related policies or audits might be warranted. The design supports both monitoring (monthly reviews) and exploration (deep dives into particular drugs or prescribers), reflecting the dual nature of the compliance team’s role.

Discussion

The Medicare FFS fraud detection project illustrates the power of a structured storytelling with data process. Beginning with context ensured that the visuals were aligned with audience needs and constrained by real-world conditions. Choosing appropriate displays guaranteed that each chart's form matched its function. Eliminating clutter and applying preattentive attributes improved the clarity and speed of comprehension. Thinking like a designer increased accessibility, credibility, and audience acceptance. Finally, assembling the visuals into a coherent narrative and dashboard provided a clear path from awareness to action.

At the same time, the project has its limitations, public CMS data, while rich, lacks detailed clinical context such as diagnosis codes or patient-level adherence information. As a result, the charts can identify suspicious patterns but cannot conclusively distinguish fraud from legitimate clinical variation. The thresholds used to define high risk, for example, the brand share and cost per claim cutoffs are somewhat arbitrary and would need to be tuned with domain experts. In production use, the dashboard would ideally be combined with additional modeling, such as peer-group comparisons and risk-adjusted benchmarks.

Despite these limitations, the project demonstrates that well-designed visual storytelling can significantly enhance the ability of non-technical stakeholders to navigate complex health-care data. The process also reinforced an important personal lesson, which is, visualization is not a cosmetic layer added at the end of an analysis. It is a core part of analytic thinking, tightly coupled to the questions being asked and the decisions that need to be made.

Conclusion

This paper has documented an end-to-end application of the Storytelling with Data framework to Medicare Fee-for-Service fraud detection using publicly available CMS Part D data. By progressing through the stages of understanding the context, choosing appropriate displays, eliminating clutter, drawing attention, thinking like a designer, and telling a story, I transformed raw claims data into a set of visuals and a dashboard that support real decision needs for policy and compliance stakeholders.

The final result is more than a collection of attractive charts. It is a narrative tool that enables stakeholders to see where cost and risk are concentrated, to identify which prescribers and drugs warrant immediate review, and to consider concrete policy measures that could reduce exposure. The project underscores the central premise of Knaflic's work: when we make the audience the protagonist and treat data as supporting evidence in a story, we move beyond simply showing numbers. We learn to tell a story with data that can change how organizations think and act.

References

Centers for Medicare & Medicaid Services. (n.d.). Medicare Fee-For-Service Provider Utilization & Payment Data: Physician and Other Practitioners. Data.CMS.gov. https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners

Centers for Medicare & Medicaid Services. (n.d.). Medicare Part D Prescriber Public Use File. Data.CMS.gov. https://data.cms.gov/provider-summary-by-type-of-service/medicare-part-d-prescribers

Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley.

Few, S. (2009). Now you see it: Simple visualization techniques for quantitative analysis. Analytics Press.