

Textual Analysis.

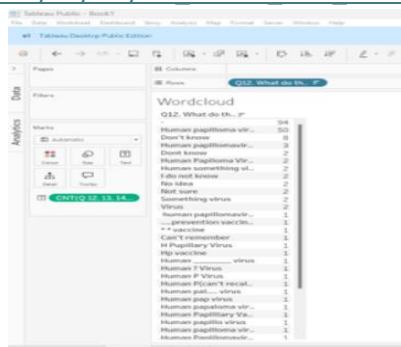
Paul Maden.

Last Update: 10/10/2024

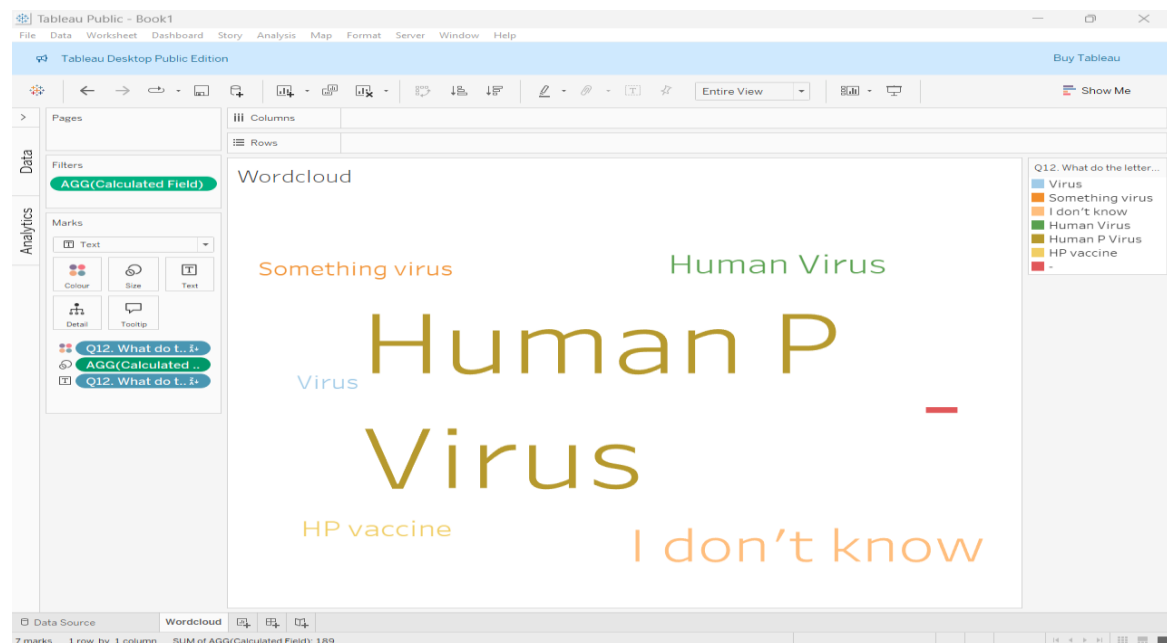
Part 1 (a). Create a word cloud for an unstructured data variable in your data set:

This initial frequency table showed data visualization presented as a table with the terms and their respective counts. This is based on responses to a question labelled Q12 from the data set:

<https://images.careerfoundry.com/public/courses/data-immersion/A2/2.8/UK Parent HPV Vaccine Attitudes.xlsx>.



Next, I created a Word-cloud from the same data, question labelled Q12:



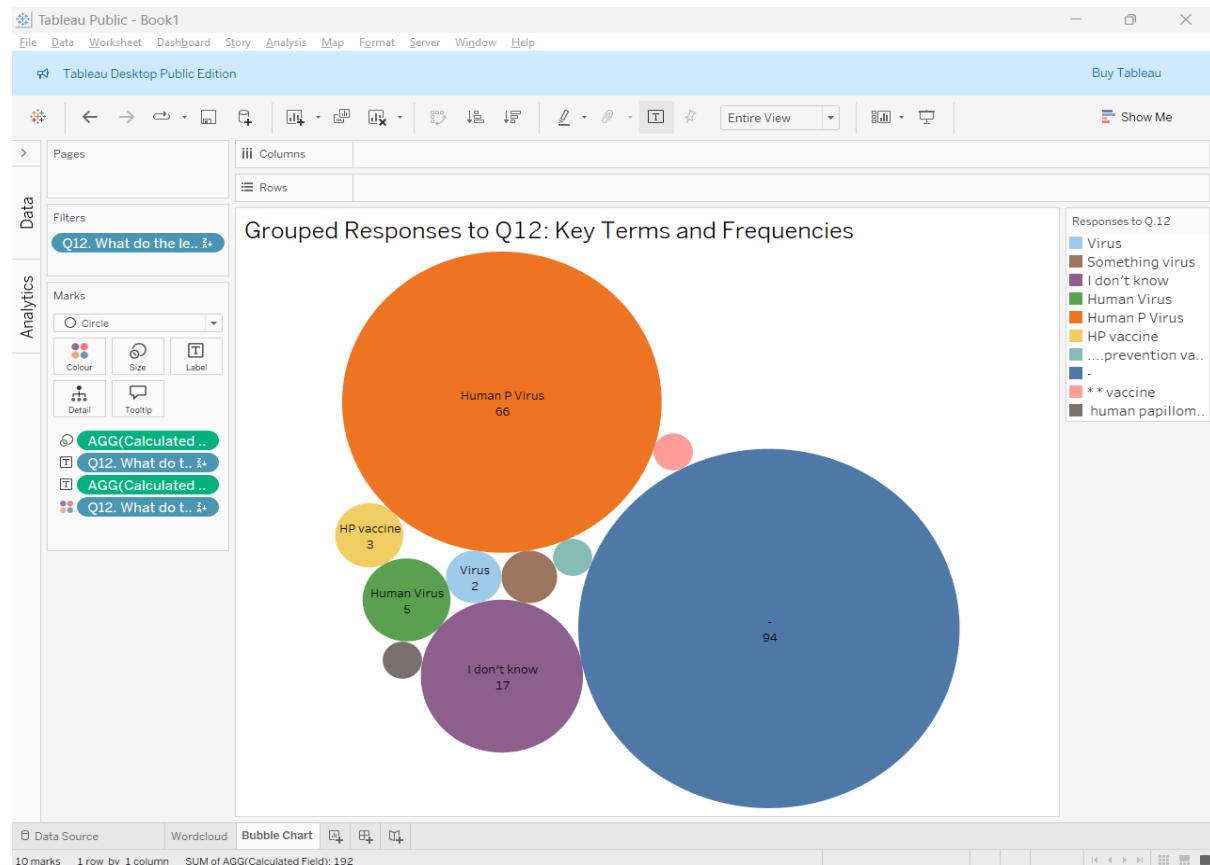
In this word-cloud visualization, I grouped the variable in question to reduce the number of distinct words or phrases displayed (to improve clarity), with responses categorized into groups like "Human P Virus," "HP vaccine," "I don't know," and "Something virus."

- Larger words in the word cloud, such as "Human P Virus" and "I don't know", indicate that these terms appeared more frequently in the dataset. Smaller words like "HP vaccine" and "Something virus" were mentioned less often.
- The use of colour differentiates the grouped categories, as shown by the legend on the right, which assigns colours to each category. This grouping helps to condense the variety of responses into fewer, more manageable groups for analysis.

Part 1 (b). Duplicate the chart in a new sheet and turn it into a packed bubble chart for the same data:

This Bubble Chart shows Grouped responses to Q12:

https://public.tableau.com/views/GroupedResponsestoQ12KeyTermsandFrequencies/BubbleChart?:language=en-GB&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link



I chose a gradient colour scheme for the bubble:

- **Larger Bubbles:** I used a *darker, more intense colour* for the larger bubbles to draw attention. E.g. deep shades of *blue, purple, or orange* can signify importance.
- **Smaller Bubbles:** I used *lighter, softer shades* of colour to represent smaller numbers, maintaining consistency but making them less prominent.

The bubble chart provides a few advantages over the word cloud, offering insights that the word cloud alone cannot:

- **Relative Size Comparison:** In the bubble chart, the size of each bubble directly corresponds to the response frequency, making it easy to visually compare the proportions between different groups. E.g. it's clear that "Human P Virus" and "I don't know" are much larger compared to responses like "HP vaccine."
- **Categorical Distinction:** The bubble chart uses distinct colours for different categories, making it easier to distinguish between various groups of responses (e.g., "I don't know" is purple, and "Human P Virus" is orange). While the word cloud also uses colour, the categorical grouping and size in the bubble chart make it clearer and easier to read in terms of separating the categories.
- **Precise Quantitative Representation:** The bubble chart clearly shows the exact frequency of responses, with numbers displayed inside each bubble (e.g., "Human P Virus" has a count of 66, and "I don't know" has 17). While the word cloud visually emphasizes frequently used words through size, it does not show specific counts without hovering or additional details. (ref. Kieran Healy in *Data Visualization: A Practical Introduction*.)

Part 2 (a) Revisit your business requirements document and project plan and brainstorm. How might unstructured survey data supplement your student project?

Unstructured survey data can significantly supplement my project by providing qualitative insights that might not be captured through quantitative data alone. Here's some possible meaningful insights:

How Unstructured Survey Data Can Supplement My Hypothesis:

- **Contextual Understanding:** While my hypothesis focuses on the relationship between vulnerable populations (like adults over 65) and influenza deaths, unstructured survey responses can provide a deeper understanding of the factors that contribute to this relationship. E.g. I could analysis information on healthcare practices, challenges, and perceptions about the flu.
- **Identifying Barriers:** Surveys from staff and patients could reveal barriers to vaccination or treatment that may not be evident from numerical data. These could include personal beliefs, logistical issues (e.g., difficulty in accessing vaccines), or systemic healthcare challenges (e.g., staff shortages or overburdened hospitals).
- **Exploring Mitigation Efforts:** Unstructured feedback can highlight strategies already in place to mitigate influenza deaths (e.g., outreach programs, education campaigns, vaccination drives) and why they may or may not be effective.

Types of Data You Might Receive from Unstructured Survey Questions:

- **Staff Responses:**
 - Insights on patient care challenges (e.g., understaffing or vaccine availability)
 - Perceptions of vaccine hesitancy among vulnerable populations.
 - Feedback on training programs during flu seasons.
- **Patient Responses:**
 - Reasons for not receiving flu vaccinations (e.g., lack of awareness).
 - Personal experiences with flu symptoms or complications.
 - Opinions on the healthcare system's ability to manage flu outbreaks.
- **Using Textual Analysis to Derive Insights:**
 - **Sentiment Analysis:** Sentiment analysis can categorize responses into positive, negative, or neutral attitudes. I.e. it could help identify whether patients and staff have a generally positive or negative view of flu vaccines.
 - **Example Insight:** If sentiment analysis reveals that most elderly patients express hesitancy or distrust in vaccines, it may explain higher influenza death rates and provide insight into areas for targeted education or outreach.
 - **Keyword Analysis:** A simple keyword frequency analysis can identify the most common words or phrases used in responses. E.g. frequent mentions of terms like "vaccine shortages" or "flu complications" could provide clues to the underlying causes of high flu mortality in states with large elderly populations.
 - **Correlation with Quantitative Data:** Once you extract insights from the unstructured data, you can correlate those themes with your quantitative findings. I.e. regions reporting higher instances of staff burnout or vaccine shortages (from textual data) might also show higher influenza death rates.

Unstructured survey data can provide rich, contextual insights that help explain the quantitative relationship between vulnerable populations and flu deaths, providing a more *holistic* understanding of the factors driving flu mortality among at-risk populations. ([SpringerLink](#))

Part 2 (b) How might survey, or other forms of unstructured data be useful to analyse as a next step in this project?

Surveys and other forms of unstructured data could be highly useful in analysing the success of my influenza project, providing qualitative feedback to complement quantitative measures like staffing needs and influenza death rates. Here's some possible meaningful insights:

Why Surveys or Unstructured Data Are Useful for Next Steps:

Once staffing needs and plans for the influenza season are in place, I would need to evaluate their effectiveness. Surveys and other unstructured data provide qualitative feedback that complements quantitative measures like staffing levels and influenza death rates.

- **Frontline Feedback:** Healthcare staff surveys (doctors, nurses, administrators) offer insights into whether staffing changes improved patient care and reduced stress.
- **Patient Perceptions:** Patient surveys help determine whether interventions (e.g., vaccination drives) increased their understanding of influenza prevention.
- **Identifying Gaps:** Unstructured feedback can reveal unexpected challenges, such as resource allocation issues or communication barriers, that were not anticipated.

Using Textual Analysis to Measure Project Success

Textual analysis can evaluate survey responses & feedback to track the project's success. Such as:

- **Sentiment Analysis:** Gauges satisfaction by analysing the tone of staff & patient responses.
 - E.g.: A positive trend in sentiment could indicate that staffing & vaccination initiatives were successful, while negative sentiment suggests the need for revisions.
- **Keyword Frequency & Thematic Analysis:** Identifies recurring themes and measures whether key project objectives were met.
 - E.g.: If mentions of “overworked” or “long wait times” decrease compared to previous years, it suggests that the project successfully addressed key challenges.
- **Pre/Post Comparison:** Comparing unstructured data from before and after the flu season helps to identify shifts in staff morale, patient behaviour, or operational efficiency.
 - E.g.: A reduction in mentions of “staff shortages” or “burnout” post-intervention would indicate an improvement.

Textual Analysis for Producing Insights

Textual analysis can provide actionable insights from unstructured data in several ways:

- **Sentiment Tracking Over Time:** Tracks sentiment shifts across phases of the project.
 - E.g.: Improved patient sentiment mid-season could indicate the success of communication campaigns or better vaccine access.
- **Comparative Analysis:** Compares survey responses across regions, hospitals, or demographics to identify what worked best in different contexts.
 - E.g.: If one region shows more positive feedback, it suggests successful strategies that can be replicated in other areas.
- **Measuring Behavioural Change:** Analyses patient feedback to track changes in behaviour, such as increased vaccination rates or adherence to preventive measures.
 - E.g.: If surveys show more patients reporting that they got vaccinated, it suggests that the project effectively influenced behaviour.

Textual analysis of unstructured data from staff and patient surveys provides qualitative insights that complement quantitative metrics, allowing for a more *holistic* evaluation. Using sentiment and keyword, it identifies areas of improvement and measures the broader impact of interventions on my project success (ref. [Frontiers](#), [BioMed Central](#), [MDPI](#))