Answer 2.8: Textual Analysis in Tableau
By Yarisel Velacanto
Date: 12/06/2024
[HPV Textual Analysis]

Creating a Word Cloud and Packed Bubble Chart from Unstructured Data

Step 1: Identifying the Variable for Textual Analysis

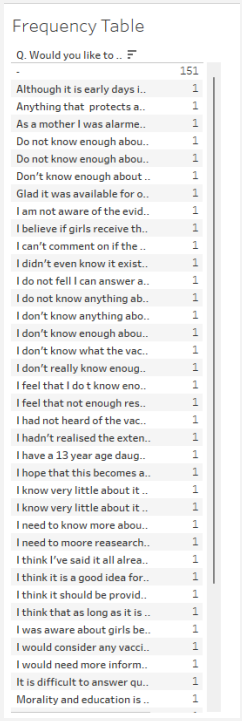
It seems the column headers in the dataset were a bit inconsistent, which made it challenging to pinpoint the specific variables at first. I refined my approach by searching for keywords and reviewing the structure more closely to identify the relevant columns.

After exploring the dataset, I found a suitable unstructured text column:
"Q. Would you like to add any other comments about HPV vaccination?"

This column contains free-text responses, making it a great choice for textual analysis. I'll move forward with this column to create the word cloud and bubble chart.

Step 2: Creating the Frequency Table

I connected the dataset to Tableau and selected the column **"Q. Would you like to add any other comments about HPV vaccination?"** for analysis. I created a frequency table by counting how often each response appeared. The table is sorted in descending order to highlight the most common responses.


Frequency Table

Q. Would you like to ..	
-	151
Although it is early days i..	1
Anything that protects a..	1
As a mother I was alarme..	1
Do not know enough abou..	1
Do not know enough abou..	1
Don't know enough about ..	1
Glad it was available for o..	1
I am not aware of the evid..	1
I believe if girls receive th..	1
I can't comment on if the ..	1
I didn't even know it exist..	1
I do not fell I can answer a..	1
I do not know anything ab..	1
I don't know anything abo..	1
I don't know enough abou..	1
I don't know what the vac..	1
I don't really know enoug..	1
I feel that I do t know eno..	1
I feel that not enough res..	1
I had not heard of the vac..	1
I hadn't realised the exten..	1
I have a 13 year age daug..	1
I hope that this becomes a..	1
I know very little about it ..	1
I know very little about it ..	1
I need to know more abou..	1
I need to moore reasearch..	1
I think I've said it all alrea..	1
I think it is a good idea for..	1
I think it should be provid..	1
I think that as long as it is ..	1
I was aware about girls be..	1
I would consider any vacci..	1
I would need more inform..	1
It is difficult to answer qu..	1
Morality and education is ..	1

Step 3: Creating the Word Cloud

To begin, I attempted to create a word cloud using the column **"Q. Would you like to add any other comments about HPV vaccination?"**. However, the visualization did not generate as expected, likely due to the structure of the responses in that column.

I then switched to the column **"Q15. What is the relationship, if any, between HPV and cancer?"**, which provided more suitable unstructured text responses for this type of analysis. Using this column, I successfully generated a word cloud in Tableau.

In this visualization, the size of each word reflects its frequency within the responses, making it easier to identify key terms and patterns in the data. For example, terms like "cervical cancer" and "not sure" appeared prominently, indicating they were frequently mentioned by respondents. This insight provides a starting point for understanding the general themes and areas of concern related to HPV and cancer.

Step 4: Enhancing the Word Cloud with Color

To add more context to the word cloud, I introduced a dimension by applying the "Cleaned Gender" column to the **Color** Marks Card. This allowed me to differentiate responses based on gender (Female, Male, Other) visually.

- **Orange** represents **Female**, a vibrant choice that stands out clearly.
- **Green** represents **Male**, a softer color that complements the orange.
- **Gray** represents **Other**, offering a neutral contrast for inclusivity.

These colors make it easier to see which gender contributed specific comments, adding a deeper layer of insight to the visualization. The final word cloud is now not only informative in terms of word frequency but also shows how responses vary across genders.

Packed Bubble Chart Documentation

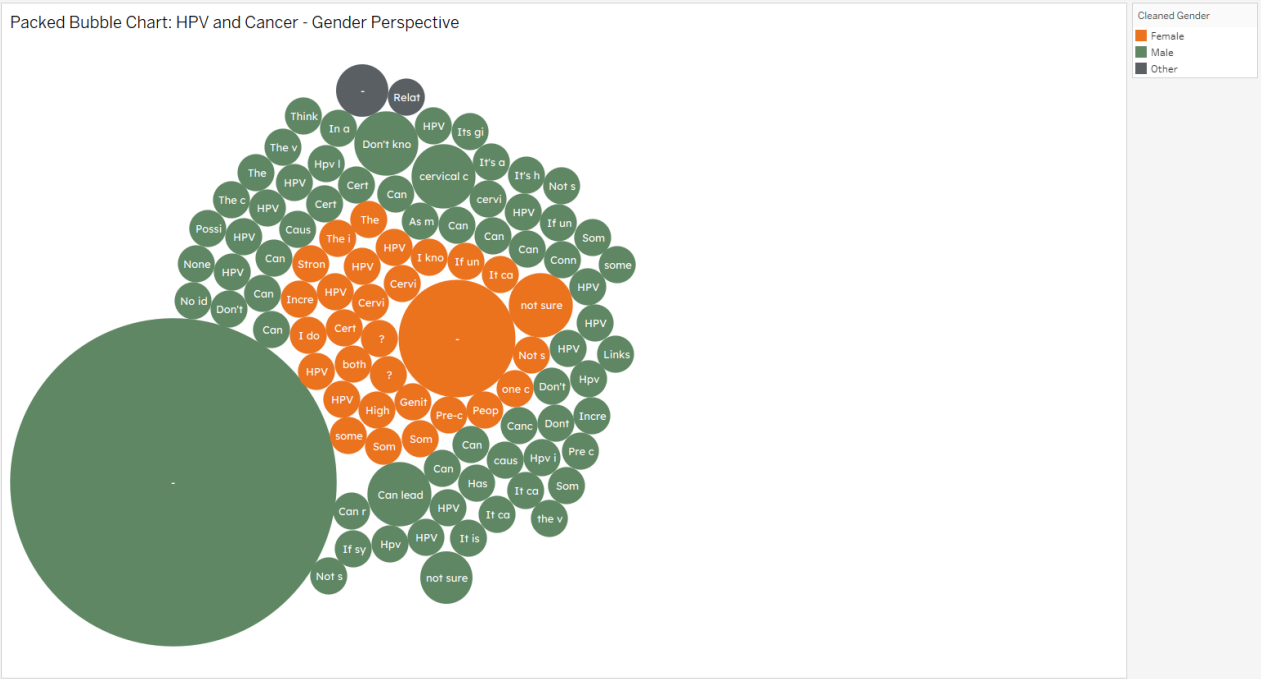
Packed Bubble Chart: HPV and Cancer - Gender Perspective

To further analyze the responses, I created a Packed Bubble Chart using the "Q15. What is the relationship, if any, between HPV and cancer?" column. Each bubble represents a unique word or phrase, with size indicating frequency and color denoting gender (orange for Female, green for Male, and gray for Other).

The chart provides a visual representation of how often specific words or phrases appear in responses while showing their distribution across gender categories. Unlike the Word Cloud, this chart highlights the relative sizes of contributions and allows for easier identification of dominant terms and their gender associations.

Process:

1. Duplicated the Word Cloud sheet and renamed it.
2. Changed the **Mark Type** to "Circle" to create a bubble chart.
3. Adjusted the **Size** and **Color** fields to align with the Word Cloud.
4. Cleaned the gender data for consistent representation across visualizations.
5. Updated the chart title to match its purpose.



Part 2: Brainstorming Applications of Textual Analysis for Your Project

How might unstructured survey data supplement your student project?

Unstructured survey data, like comments from healthcare staff or patients, could provide extra context to the challenges faced during flu season that

numbers alone might not show. For example, comments could highlight issues like long wait times, patient concerns about vaccines, or staffing shortages that wouldn't be obvious from just looking at the data.

What sort of data might you receive from unstructured survey questions posed to staff and patients?

From staff:

- Comments about being overworked or needing more resources during peak flu season.
- Ideas for improving processes or communication within teams.

From patients:

- Feedback about their vaccination experience, like whether it was easy to access or too expensive.
- Concerns about reaching clinics or not fully understanding why they should get vaccinated.

How could textual analysis be used to produce insights from this data?

Textual analysis could help spot common themes or keywords, like "long lines," "cost concerns," or "need more staff." A sentiment analysis could also show how people feel overall—whether they're happy, neutral, or frustrated—about their experiences.

How might surveys or other forms of unstructured data be useful to analyze as a next step in this project?

Survey data could help refine the staffing plan by showing where problems came up. For example, if a lot of comments mention long wait times at rural clinics, that could point to a need for more staff in those areas. Patients' responses could also highlight gaps in public health messaging, like misunderstandings about the flu vaccine.

With influenza staffing needs determined and plans in place for the next influenza season, how might you use textual analysis to measure the success of the project?

After the staffing plan is implemented, surveys could collect feedback from staff and patients about how things went. Textual analysis of this feedback could help figure out what worked and what didn't. For example, it could show if certain areas still felt understaffed or if patient experiences improved overall.

How could textual analysis be used to produce insights from this data?

Analyzing feedback could uncover recurring themes or patterns. It could help identify:

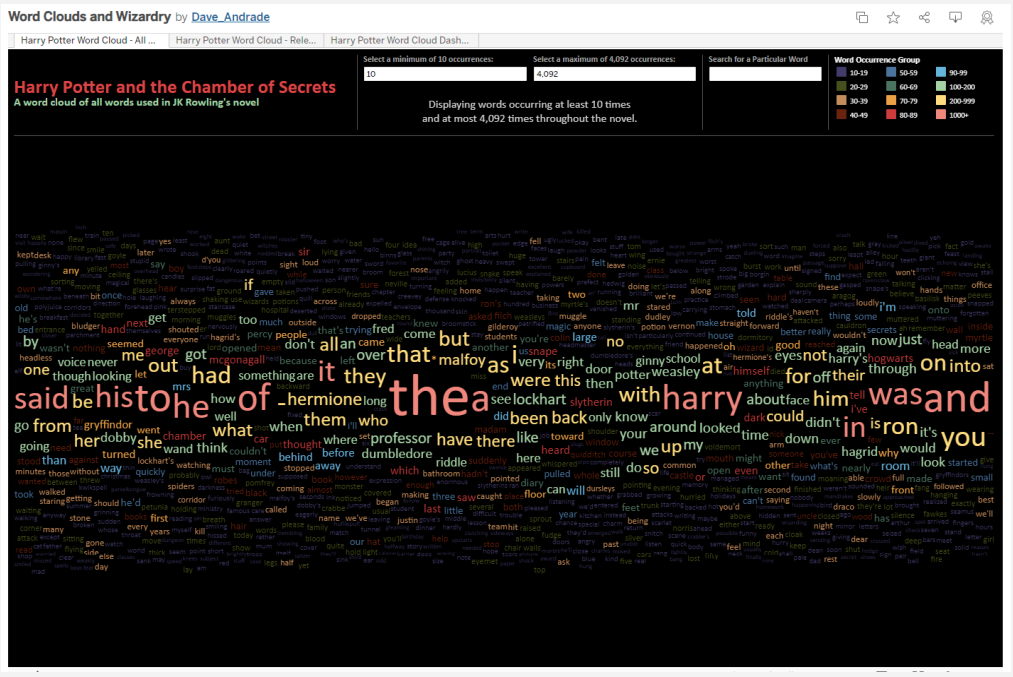
- Which states or clinics had consistent challenges.
- Whether public health messaging was effective based on patient feedback.
- New issues that might need attention for future flu seasons.

Bonus: Critique of "Word Clouds and Wizardry"

What works well: I like the dark background because it makes the colors of the words stand out, and the colors don't overlap, which makes it easy to distinguish between all the words. I also appreciate how the words are organized in a structured, rectangular layout, ensuring no empty spaces. The arrangement looks visually appealing and allows the focus to remain on the larger, more frequent words.

What doesn't work well in terms of how it communicates data: There is no explanation or reasoning provided for why this word cloud was created, such as why those specific words were chosen or how they were derived. It's unclear what data or insights this visualization is meant to communicate, which makes it harder to understand the purpose or context behind the chart.

Using my visualization style guide to critique the presentation: While the dark background and vibrant word colors follow good contrast principles, the lack of a title, legend, or description reduces its effectiveness. Adding a brief description about the data source or purpose would make it more informative. Additionally, the rectangular structure is efficient for fitting in many words, but it could feel overwhelming without context for interpreting it.



You can check out the chart here.