**Background Information:**

**Content Analysis** is the systematic analysis of information in text (written or verbal) format. It involves identifying features of the text and then recording or categorizing them. Content analysis includes reading texts to identify broad themes that may be subjective and context dependent as well as highly structured analyses in which ideas/sentences/phrases/words are classified and counted to be compared across texts or contexts. More structured content analysis requires a coding protocol in which the researcher describes explicitly which ideas/sentences/phrases/words she will look for in the text and how she will count and classify those things.

Content Analysis can follow the steps below

1. Start with a research question and hypothesis.
2. Determine your independent and dependent variables.
3. Decide which documents will provide the best evidence for testing the question.
   * For example, Presidential speeches? State of the Union addresses? Trump’s tweets?
4. Determine how you will measure your dependent variable. That is, decide what you will examine the text for. Are you looking for particular themes? Phrases? Emotional sentiment? Specific words?
5. Develop coding rules! Decide what signals the presence or absence of your dependent variable (or different levels of your DV)?
   * What must we see in order to put a piece of text in one category or the other?
6. Code the text.
7. Analyze the text.

**Activity Instructions:**

Follow the steps below. Occasionally you will be asked to check in with the instructor to ensure you are on the right track before proceeding.

**Research Question (Provided):**

How does the language that presidents use in speeches change after a traumatic event (in this case a terrorist attack)?

To assess this research question, you will be provided excerpts from two of President Bush’s speeches. One is his inaugural address on January 20, 2001. The second is his speech after 9/11.

**Hypothesis:**

With your group, generate a hypothesis about how presidential speech may change after trauma… do you think a president will talk about different ideas or concepts? Will they use more aggressive language? Make more or fewer emotional appeals? Talk more or less about American values? You pick! Keep in mind, you only have two speeches to look at, so avoid very unusual ideas or concepts.

Write your group’s hypothesis here and *then check in with the instructor.*

**What is your Independent Variable?**

Just list the concept for now; you will determine coding rules in a later step.

**What is your Dependent Variable?**

Just list the concept for now; you will determine coding rules in a later step.

**Measurement of the Dependent Variable.**

For simplicity sake, you will measure your dependent variable by the frequency of **words** in the speeches

**Coding Rules**

Ok, this is the fun part. What specific words will represent the dependent variable you are trying to measure in the text? For example, if you think presidents will talk about “war” more often after trauma, what words will you classify as “war” words? Perhaps: war, fight, military, battle, troops, etc.

You have to pick these before you read the speeches or else you might introduce bias. Record the words that signal the presence of your dependent variable.

* Consider choosing words that are more common. Rare words may not show up in the texts at all.
* Determine with your group mates what versions of words you will accept. I suggest counting all words with the same root. For example, if you want to count “united” as a word, should you also count “unity” and “unite.”

Record the words that will indicate the presence of your DV below. I suggest limiting this to 3-7 words per concept for ease.

When you are done, *check in with the instructor to go over your work so far and to get copies of the speeches.*

**Collect the Data!**

Read the speeches and count how often the words you listed in step 5 appear in the two speeches. Record the count for both searches for all three coders below.

**Me**  **Group mate 1** **Group mate 2**

Inaugural \_\_\_\_\_ Inaugural \_\_\_\_\_ Inaugural \_\_\_\_\_

9/11 \_\_\_\_\_ 9/11 \_\_\_\_\_ 9/11 \_\_\_\_\_

Did the three coders in your group get the same answers? If so, congratulations: you have high inter-coder reliability! If not, why did you get different answers? Respond below:

**Analyze the data!**

Calculate the average number counted in the inaugural speech across coders.   
Record here: X1 =\_\_\_\_\_

Divide the previous number by 1600 to find the proportion of words in the text: P1= \_\_\_\_\_

Calculate the average number counted in the 9/11 speech across coders. Record here: X2= \_\_\_\_\_

Divide the previous number by 1890 to find the proportion of words in the text: P2= \_\_\_\_\_

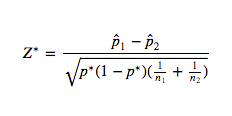
Compare the proportion or words that qualified as in your category between the two speeches… does it look like there is any support for your hypothesis? Describe why or how below *and flag down Prof A for further instruction.*

To determine if this difference between the two speeches may be due to random chance or not, fill in the values from step 7 into the formula below and solve for Z.

X1 and X2 are the number of qualifying words you counted

N1 and N2 are the number of total words in the speech (1600 for the inaugural speech and 1890 for 9/11)

P1 and P2 with the ^ symbol is the proportion of qualifying words



Formula for p*

Record your value for z here: Z = \_\_\_\_\_

If your z-score is greater than 1.96, your findings are statistically significant at the traditional threshold! That means that this difference in the proportion of words between the two speeches is probably not due to chance. This might not mean your hypothesis is \*right\* (lots of alternative explanations), but it’s still neat!

**Application to individual projects**

Think about your own research design project. Can you think of a way in which content analysis may help you test your hypotheses below? Record below:

Which documents would you need to access to conduct content analysis?

Which concepts would you try and identify in these documents?