

Introduction to WordCAKE	4
Features in the WordCAKE plugin	4
Section 1	6
Installing SketchUp.....	6
Installing WordCAKE for Mac.....	6
Installing WordCAKE for Windows	11
Section 2.....	18
Getting Started.....	18
Importing files into WordCAKE.....	18
File Formats.....	19
TXT (plane text file).....	19
PDFs (portable document format)	19
CSVs (comma separated values).....	19
Adjusting Cake Properties.....	20
Section 3.....	21
Fork	21
Zoom	23
Rotate	23
Grab.....	23
Save with current settings	25
Load saved settings	25
Toggles.....	26

Style: adjust colors, background, appearance:.....	27
Section 5:.....	27
How to Cite this Project:	28
Appendix:.....	28
Default Stop List:	28

Introduction to WordCAKE

There is value in being able to distill correspondence, publications, literary works, speeches, notes and other groupings of texts down to their key concepts or vocabulary. There are also many ways to assess the important terms or ideas in a text. Evaluating word frequency is one method that has gained popularity through information graphics such as word clouds or weighted wordlists. However, such representations tend to focus on a single text or a singular grouping of texts. The research objective behind the development of WordCAKE (Collated Aggregate of Key Elements) focused on examining how frequently occurring words change over sequential texts (across time, chapters, versions, or any sequential ordering). Additional features were also developed that could augment our understanding of word trends by visually representing multiple texts along with their individual and overall high frequency words, specific word counts, and contextual information.

WordCAKE is a freely available plugin for SketchUp and can be described as a layered representation of high frequency words in multiple texts where each layer or disc represents one text. Key or central words in one text are positioned in the center of a disc. Color highlights words with high overall frequency in the corpus of texts. Any word can be easily located through the alphabetical arrangements of words clockwise, around the disc. Other features include the ability to “fork” a word to highlight every instance along with specific word count and file name information. Similarly, the “knife” tool highlights all word frequency data contained in one “slice” or one text disc. Customizable features include colors, fonts, sizes, minimum threshold (by count or percentage), stop list, “include” list, and context data. Users can also select sections of a WordCAKE to export data to CSV files for further analysis. A complete list of features and capabilities are listed in bullet point form below:

FEATURES IN THE WORDCAKE PLUGIN

The following features and capabilities are enabled through the use of WordCAKE on collections of texts comprised of up to roughly one million total words:

- Color display of overall frequency in a group of texts
 - Customizable color schemes
- Position indications of frequency within a text document
 - Three options for displaying frequency:
 - 1. count (best for documents of similar length)
 - 2. percentage of word frequency in document (recommended for corpus containing documents of varying length)

- Words arranged alphabetically (A-Z clockwise) for easy location and tracking
- Adjustable minimum threshold (by raw word count or percentage)
- Context viewer window (click any word to see how it is used in the original text)
- Adjust height and width of WordCAKE
- Display of file name, total word count, and minimum threshold on each disc
- “Forked” words highlight all instances of word and display file name and count in document
- “Knife” one text to highlight all instances of high frequency words in one document and show specific information on frequency thought display of this cross section
- Customizable size, font, color and depth for words
- Automated “fly through” button moves viewer through center of model from start to end
- Ability to upload and parse CSV’s, TXT, and PDF files.
- Customizable “stop word” list (any list of words separated by commas)
- Customizable “restrict to” list offers searches through a document to count specified words
- Selection of all or a portion of a WordCAKE (and export of selected data) to CSV file
- Integrates 3D modeling features of SculptUp
 - Ability to save a WordCAKE model and align next to other WordCAKE model(s) in SketchUp for comparison (for this application, we recommend a unique color scheme for each model)
 - Select word(s) and change size, color, texture, orientation, transparency, outline, position, etc.
 - Move into and out of model, rotate, reposition, shift perspective and/or angle
 - Annotate on/in model

Section 1

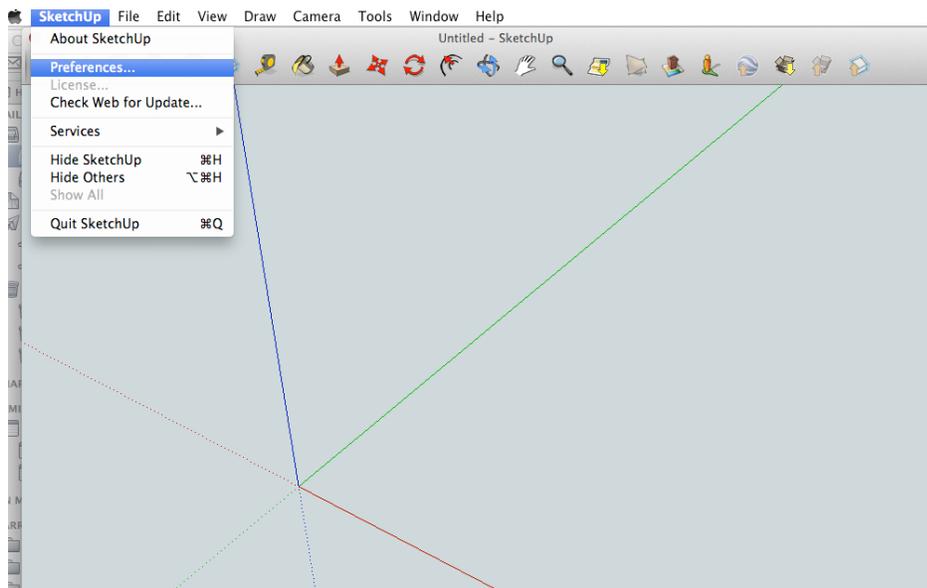
This section covers installation of SketchUp and the WordCAKE plugin for both Mac and Windows operating systems.

Installing SketchUp

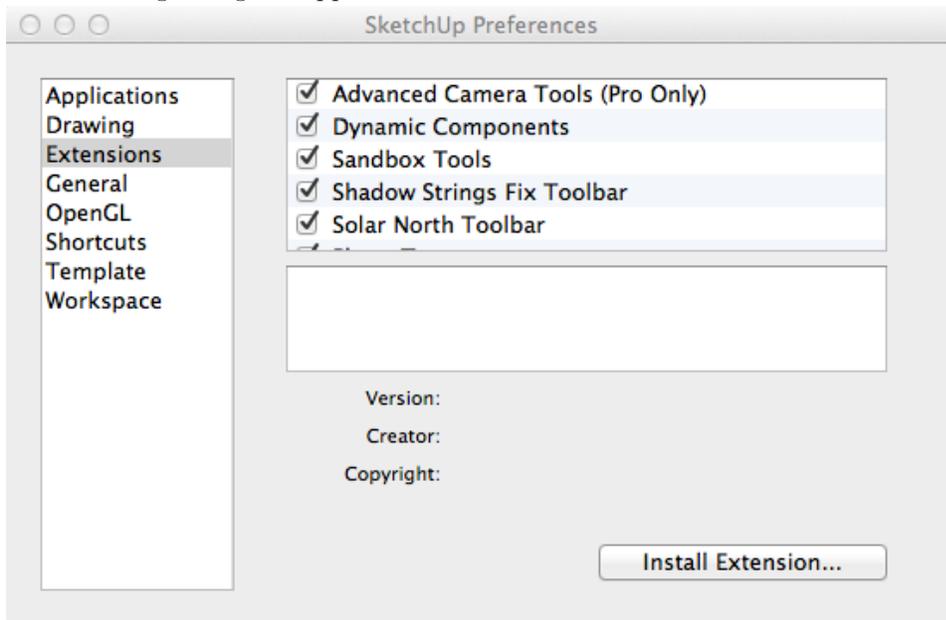
1. Download the latest version of SketchUp: <http://www.sketchup.com/download>
2. Download the WordCAKE extension: <https://extensions.sketchup.com/en/content/wordcake>

Installing WordCAKE for Mac

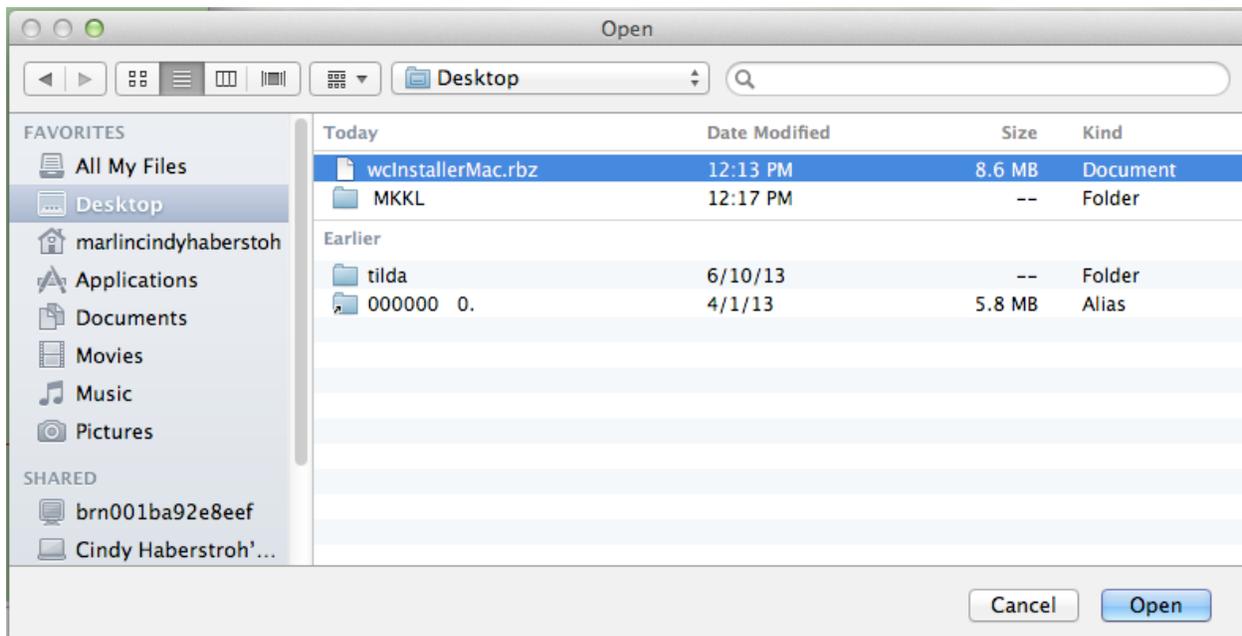
1. Open up SketchUp menu and click “Preferences”



The following dialog will appear:

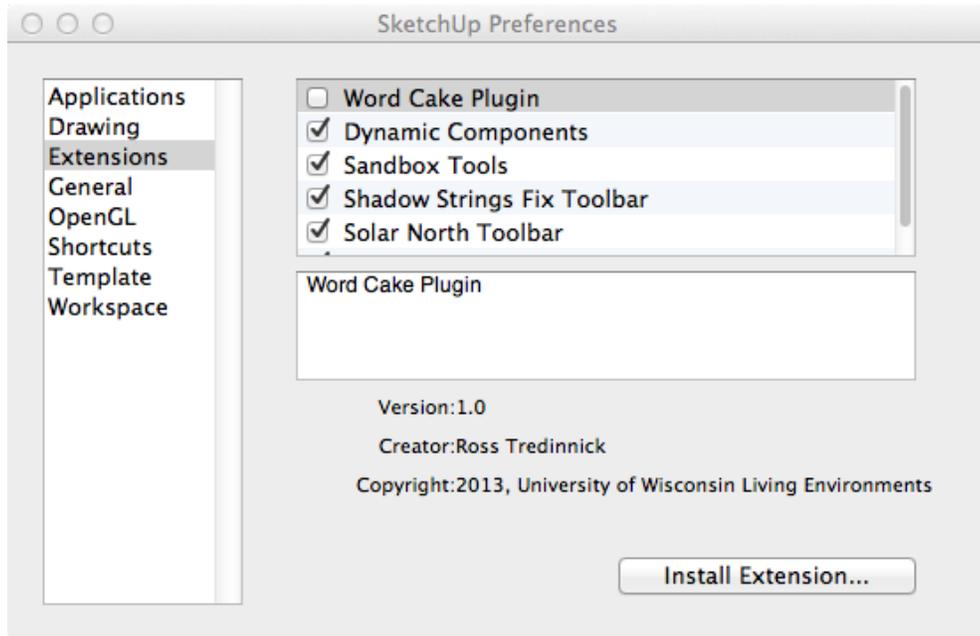


2. Click on the “Extensions” label on the left and then click the “Install Extension...”
3. The following dialog will appear, browse to the location you saved the Plugin (.rbz) file.

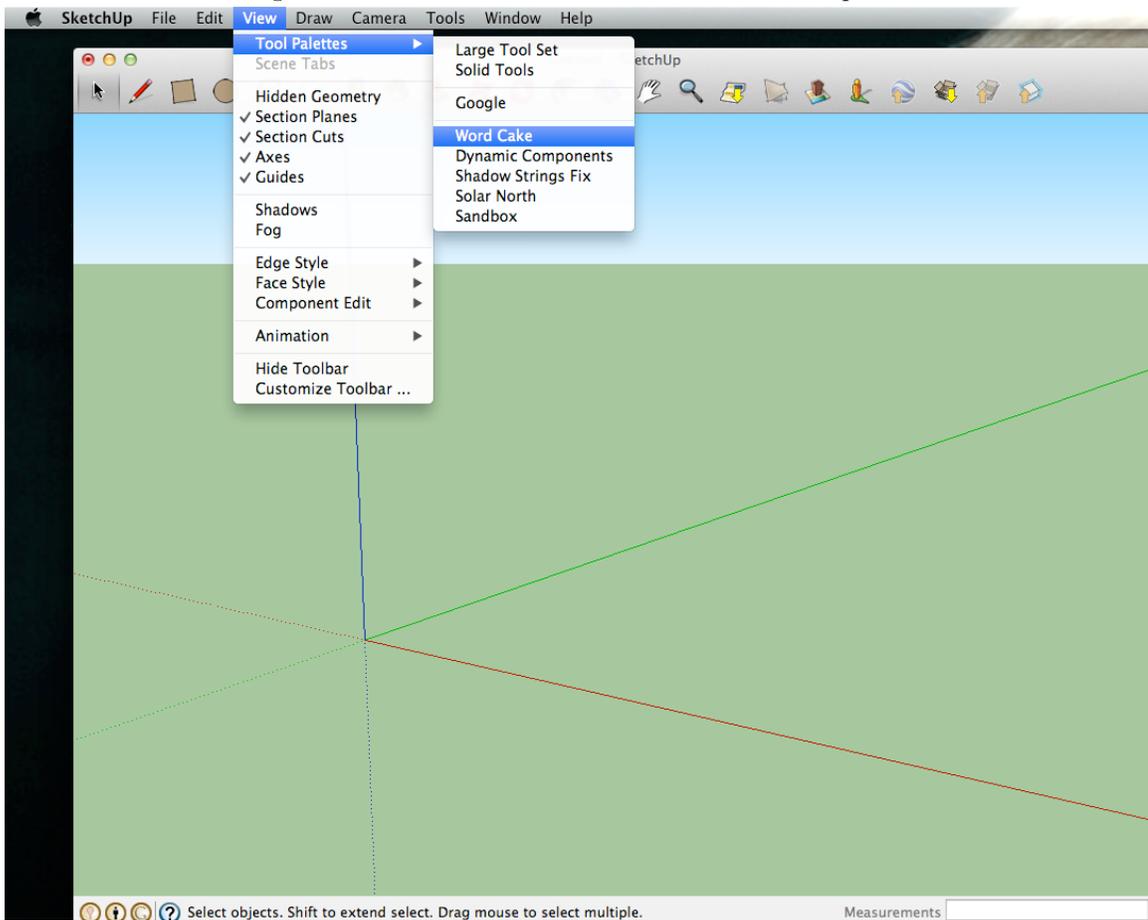


4. Select the .rbz file and click “open”. Click Yes/OK (that you trust the plugin author, come on guys you can trust me).

5. Afterwards you may see the “Word Cake Plugin” entry appear in the extensions list as shown below (you might have to scroll down and / or you might have to restart SketchUp, if the latter, just re-open the Preferences dialog).

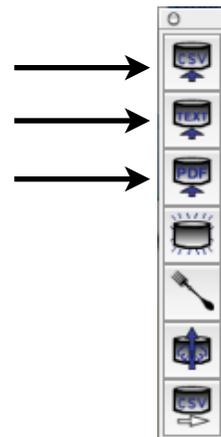


- If the “Word Cake Plugin” check box is not checked, then check It and press OK.

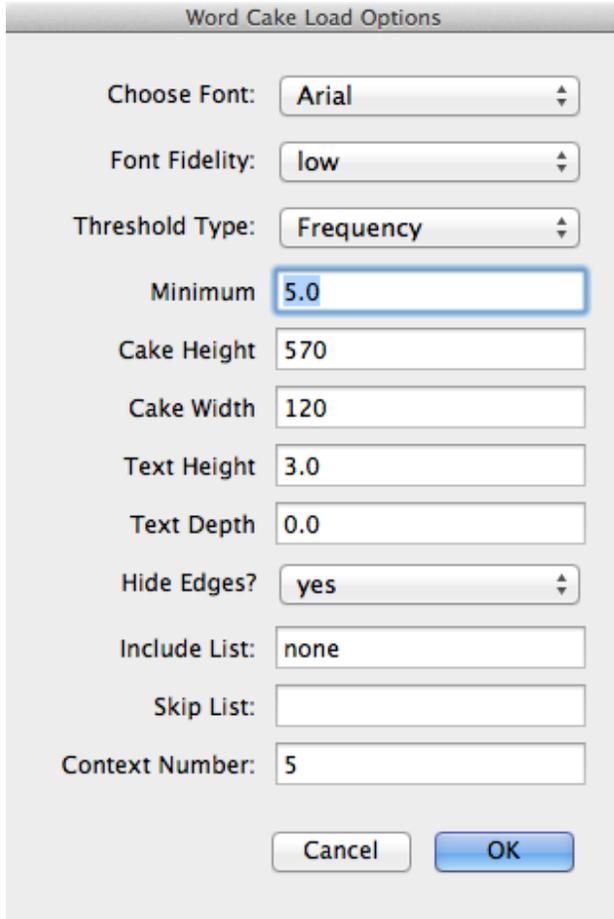


- At this point, the Word Cake Toolbar may appear, if it does not appear, you have to turn on the Toolbar – to do this, goto the “View” menu and select “Tool Palettes”. Select “Word Cake” which should then display the toolbar.

- Start a “New File” and then select and delete the human figure automatically inserted into new files in SketchUp. Now that the landscape is empty, you can begin to build your WordCAKE by uploading your files. In the WordCAKE toolbar, click the appropriate CSV, TXT or PDF icon representing the file type in your dataset.



The following dialog will appear:



9. “**Choose Font**” allows customizable font settings

Switching “**Font Fidelity**” to “low” will decrease load times.

“**Threshold Type**” offers “Frequency” for basic word count (useful for similar length texts), or “Percentage” for term count divided by total word count of text.

“**Minimum**” can be a word count (for “Frequency” option) or a word count percentage such as “0.2” or “0.15” (for “Percentage” option) for the minimum amount of times a word must appear within each of your texts to show up in the visualization.

“**Cake Height**” adjusts the spacing between the rings

“**Cake Width**” adjusts the diameter of the rings

“**Text Height**” adjusts the size of the font

“**Text Depth**” adjusts the depth of the 3D font (0.0 enhances speed and

responsiveness)

“**Hide Edges**” setting on “yes” removes distracting outlines (edges) around letters

“**Include List**” To restrict the word analysis to only a specific set of words, add a comma separated list of them (no spaces) to the “Include List” section.

“**Skip List**” You may wish to add a “stop list” to the “Skip List” entry (see sample list in appendix) to avoid high frequency words such as “a” “an” and “the” crowding the center of the WordCAKE.

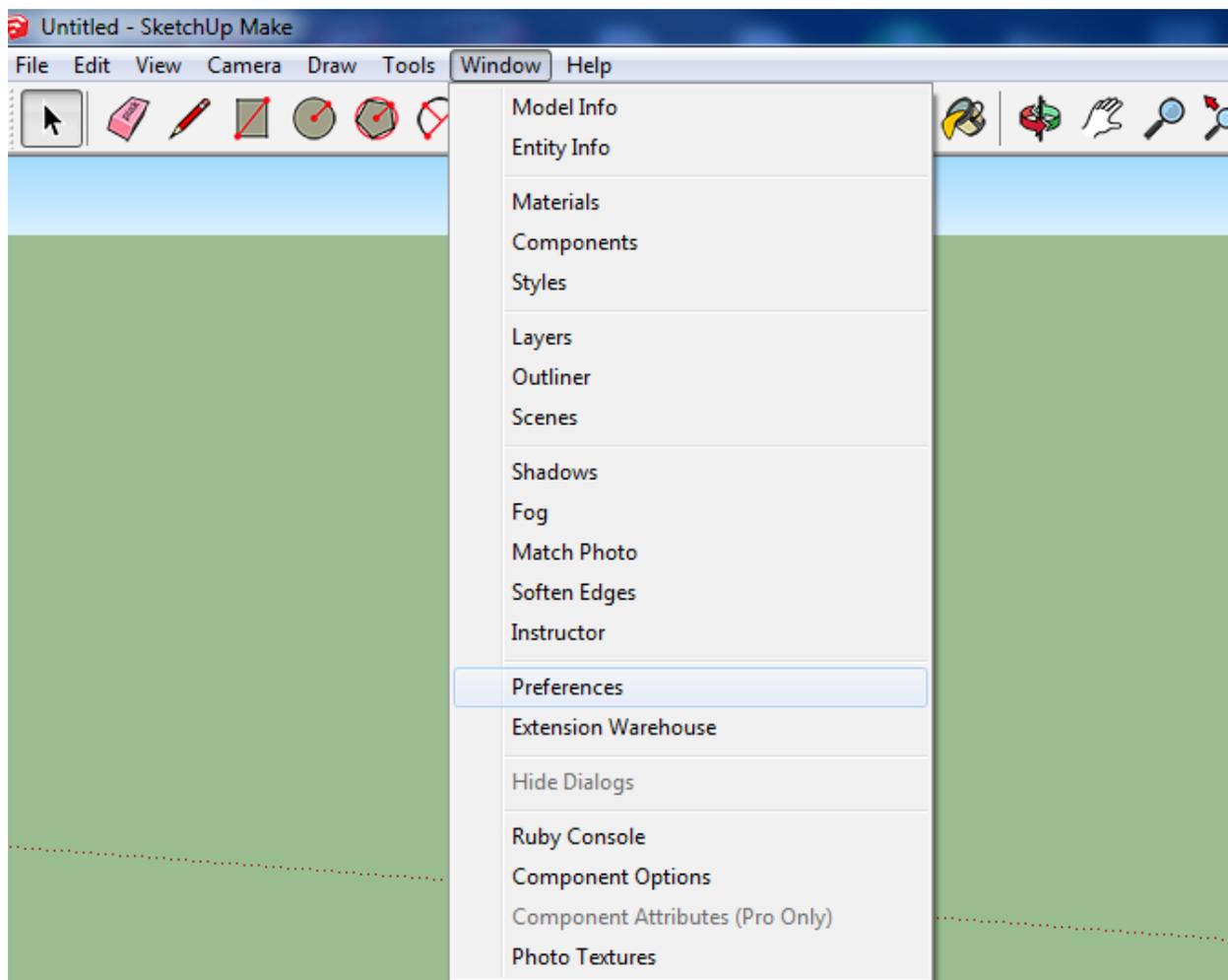
“**Context Numbers**” allows you to control how many words (in the original word’s context within a given text) appear before and after a selected word.

10. Press OK – an open file dialog box will appear – at this point just double click on any one of the files from your folder directory. Then go have a snack, the dataset may take a while to load. After it loads, you can save the skp file to load quickly in SketchUp in the future. Hooray!

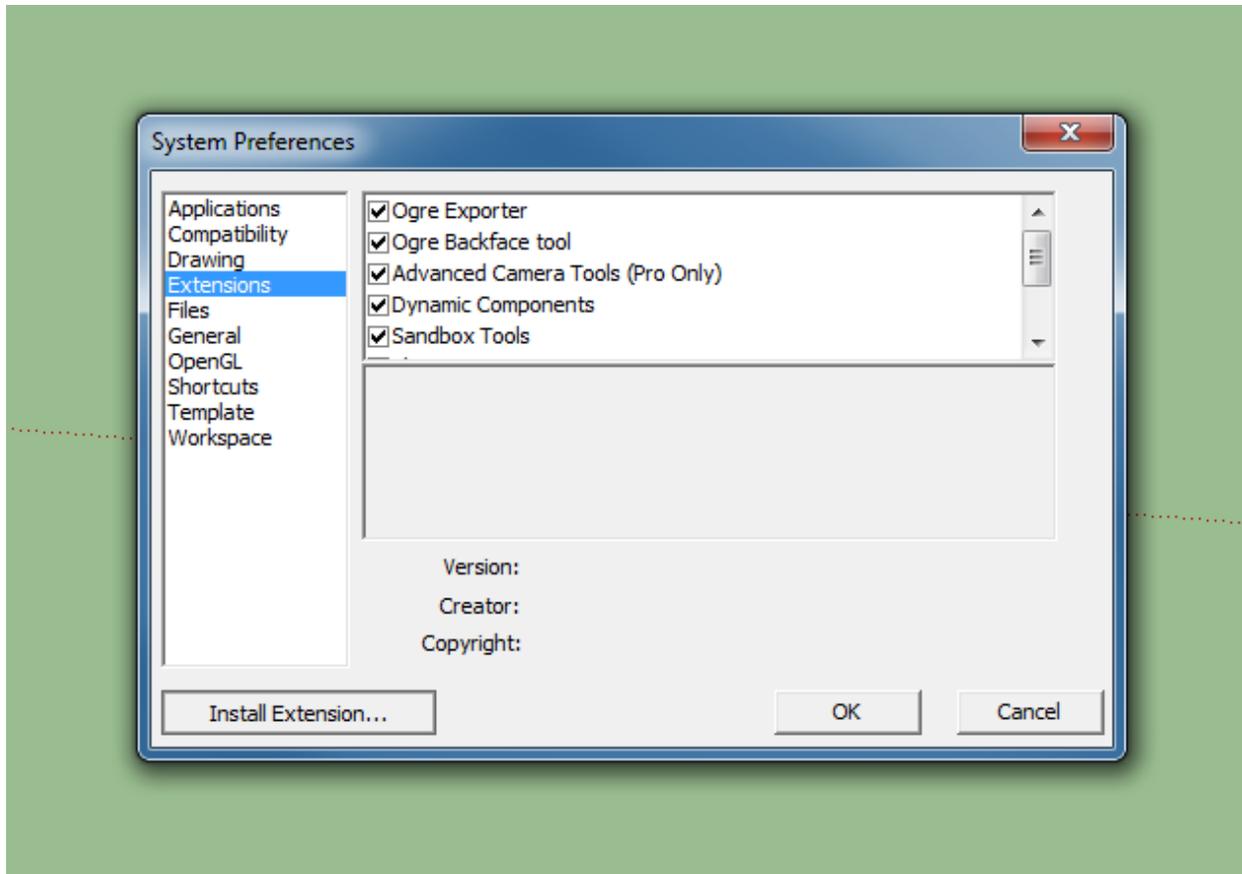
Installing WordCAKE for Windows

Word Cake Installation Instructions (images shown are for Windows):

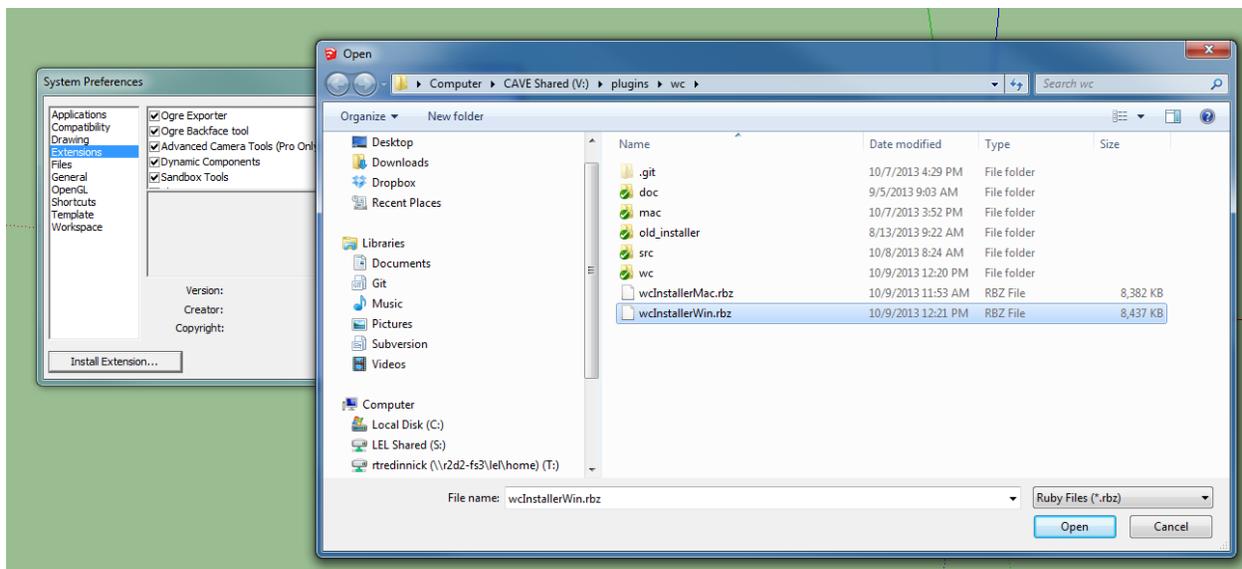
1. Open up SketchUp and click the “Window” menu and click “Preferences” (different on Mac)



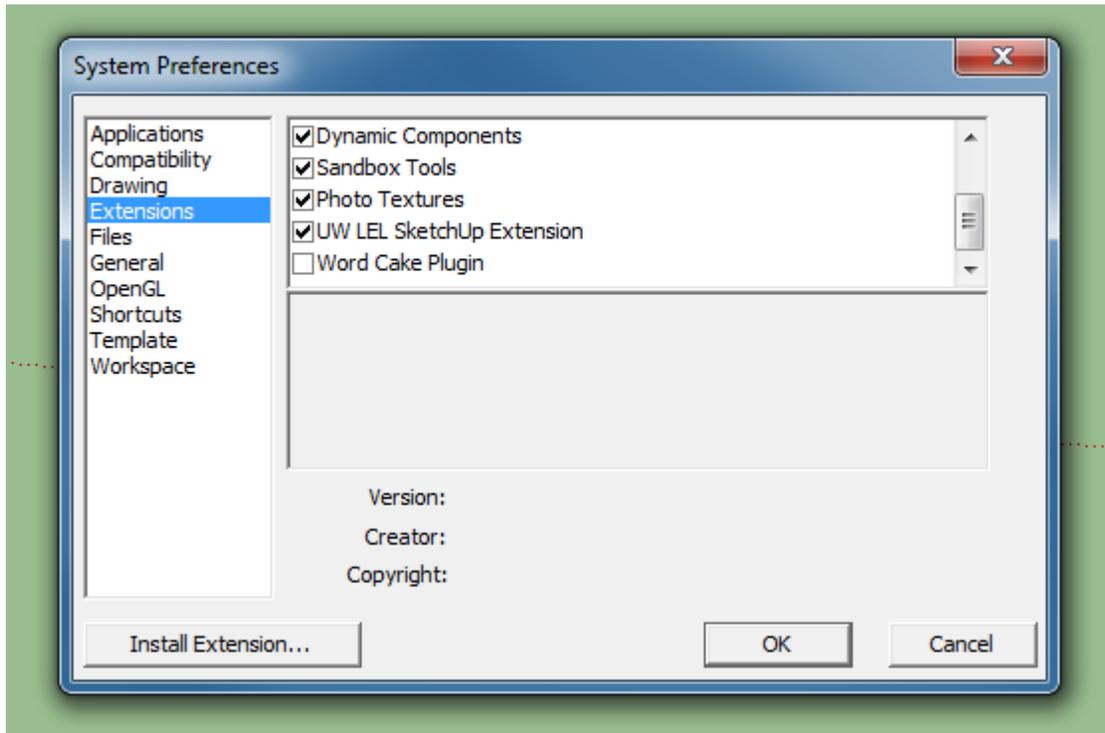
The following dialog will appear:



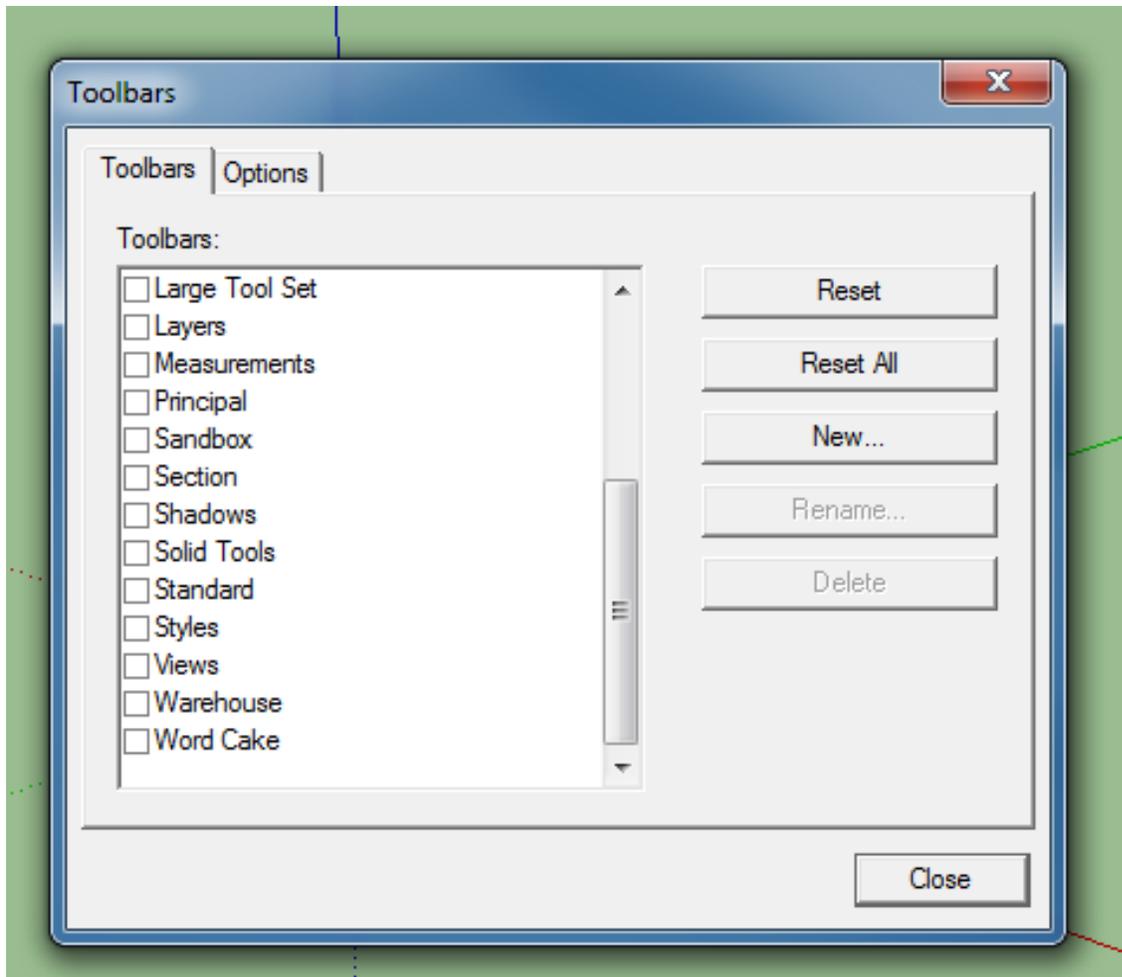
2. Click on the “Extensions” label on the left and then click the “Install Extension...”
3. The following dialog will appear, browse to the location you saved the Plugin (rbz) file.



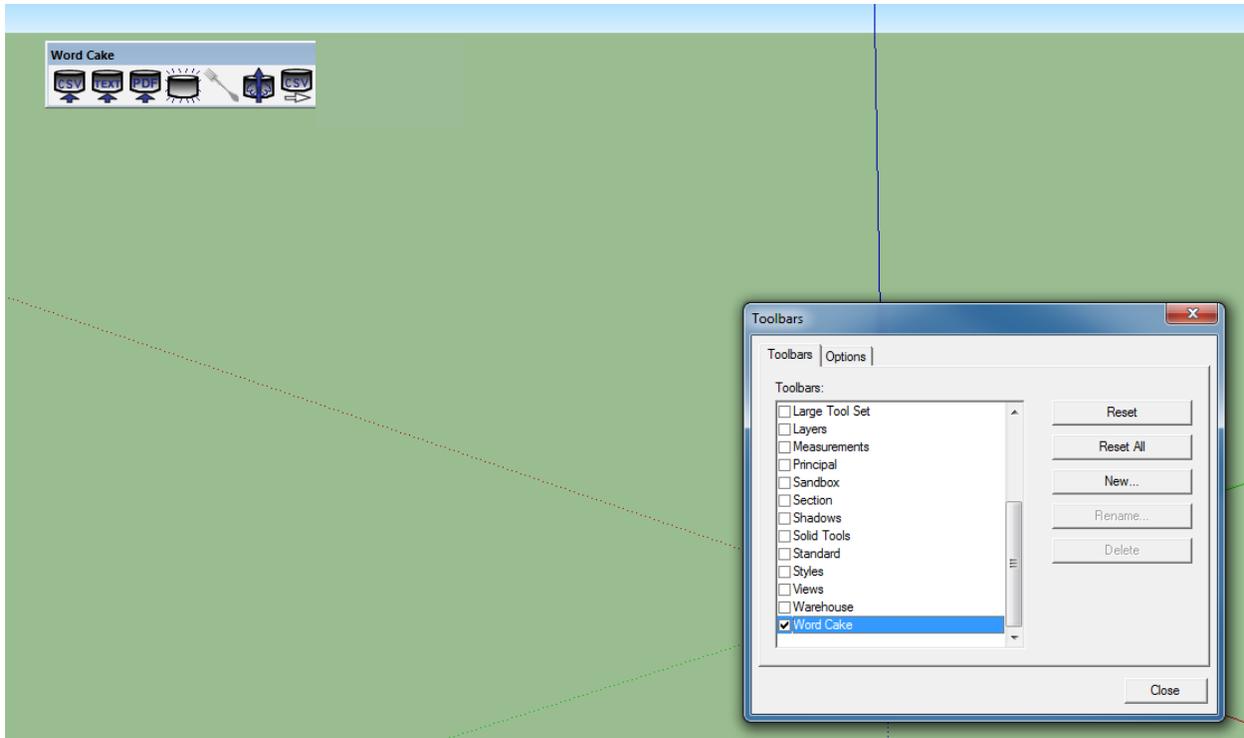
4. Select the .rbz file and click “open”. Click Yes/OK (that you trust the plugin author, come on guys you can trust me).
5. Afterwards you might see the “Word Cake Plugin” entry appear in the extensions list as shown below (you might have to scroll down and / or you might have to restart SketchUp, if the latter, just re-open the Preferences dialog after you restart).



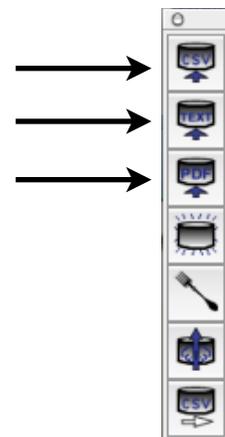
6. If the “Word Cake Plugin” check box is not checked, then check It and press OK. At this point, the Word Cake Toolbar may appear, if it does not appear, you have to turn on the Toolbar – to do this, goto the “View” menu and select “Toolbars”. Scroll all the way to the bottom and you should see this:

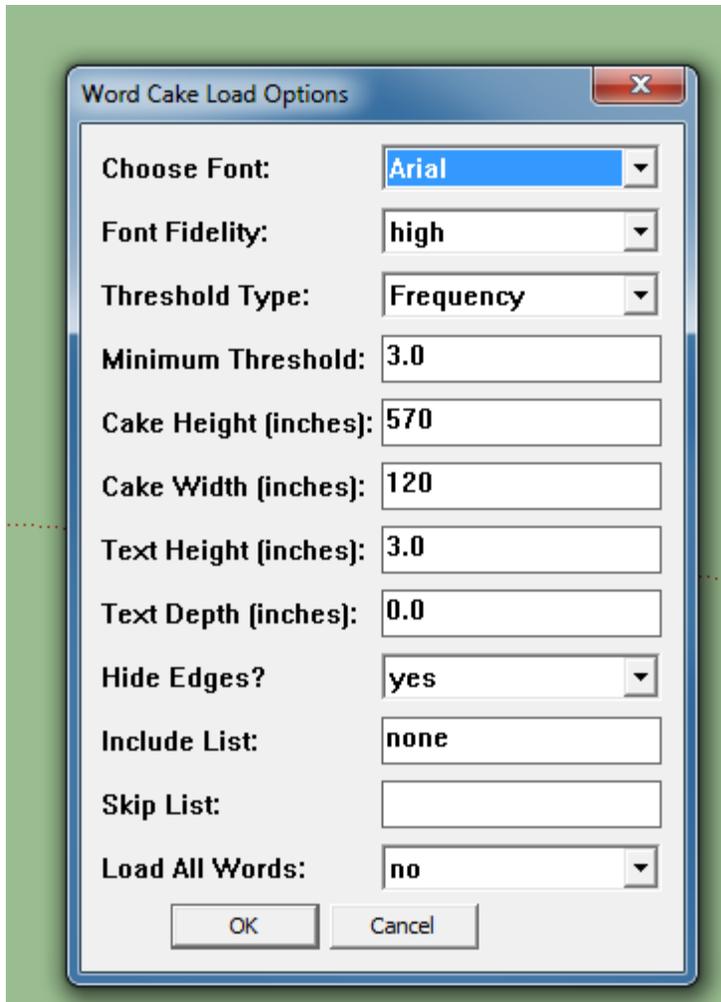


7. Click the “Word Cake” check box next to the toolbar. Upon clicking it you should see the Toolbar show up:



8. Start a “New File” and then select and delete the human figure automatically inserted into new files in SketchUp. Now that the landscape is empty, you can begin to build your WordCAKE by uploading your files. In the WordCAKE toolbar, click the appropriate CSV, TXT or PDF icon representing the file type in your dataset.





9. **“Choose Font”** allows customizable font settings
Switching **“Font Fidelity”** to “low” will decrease load times.
“Threshold Type” offers “Frequency” for basic word count (useful for similar length texts) or “Percentage” for term count divided by total word count of text
“Minimum” can be a word count (for “Frequency” option) or a word count percentage such as “0.2” or “0.15” (for “Percentage” option) for the minimum amount of times a word must appear within each of your texts to show up in the visualization.
“Cake Height” adjusts the spacing between the rings
“Cake Width” adjusts the diameter of the rings
“Text Height” adjusts the size of the font
“Text Depth” adjusts the depth of the 3D font (0.0 enhances speed and responsiveness)
“Hide Edges” setting on “yes” removes distracting outlines (edges) around letters
“Include List” To restrict the word analysis to only a specific set of words, add a comma separated list of them (no spaces) to the “Include List” section.
“Skip List” You may wish to add a “stop list” to the “Skip List” entry (see sample list in

appendix) to avoid high frequency words such as “a” “an” and “the” crowding the center of the WordCAKE.

10. Press OK – an open file dialog box will appear – at this point just double click on any one of the files from your folder directory. Then go have a snack, the dataset may take a while to load. After it loads, you can save the skp file to load quickly in SketchUp in the future. Hooray!

Section 2

This section covers the basics of getting started, such as file organization, file naming conventions, importing files, file types, and the features available for adjustment in the import settings window.

Getting Started

The first step in creating a WordCAKE is determining and defining the texts you would like to see. Keep in mind that the current version can parse through roughly one million total words. The texts should share something in common that will help you track important concepts and developments in vocabulary over time (or by other sequential order metrics). You may wish to group the texts into folders by author, theme, topic, publication, text type, or by another convention that will help make the visualization useful for addressing a specific question. Next, a sequential arrangement for the texts in a folder can be achieved through naming conventions that will place the files in order by date, time, sequence, chapter, etc. (see below).

Naming Conventions

The sequential order of files in WordCAKE is based on the file name which should begin with a number (year/month/date, ordered number, year, etc.) and be followed by a descriptor. (ex. “1976.08.31_Brown” “1921.Marin” “178690_Dexter2a” “898768.BlueCategory” “14.ATaleofTwoCities”). The initial number will order the files and the descriptive title will provide useful information to a viewer, accessible in the model. The file name will appear both at the top of each disc and to the side of a word when it is “forked.”

File names can include capital and lower case English characters, numbers, “_” and “.” punctuation, but not spaces, foreign characters or punctuation other than the options listed here.

Importing files into WordCAKE

WordCAKE can import files in CSV, TXT, and PDF formats. The program currently handles content up to roughly one million words. Once SketchUp has opened, select the appropriate button for your file types (“load from CSV” “load from TXT” or “load from PDF”). When prompted by the program to load files, users should select the first file in a folder of files adhering to a consistent naming convention (see above). Users should note that all punctuation and special characters are ignored in the WordCAKE parsing feature. In the instance of a conjunctions such as “we’ve” the word sections of “we” and “ve” will plot separately in WordCAKE. Similarly, the tool does not recognize hyphenated words nor word phrases of two or more words. However, some pre-processing of texts (finding all instances of a two or more word phrase of interest and replacing all instances with a compound form of the word (ex. find all “we’ve” and replace with “weve”) offers an alternative work around for this issue.

File Formats

TXT (PLANE TEXT FILE)

The txt file format is the default form for analysis of documents in WordCAKE.

PDFs (PORTABLE DOCUMENT FORMAT)

PDF's must contain text elements for WordCAKE to be able to parse the files.

CSVs (COMMA SEPARATED VALUES)

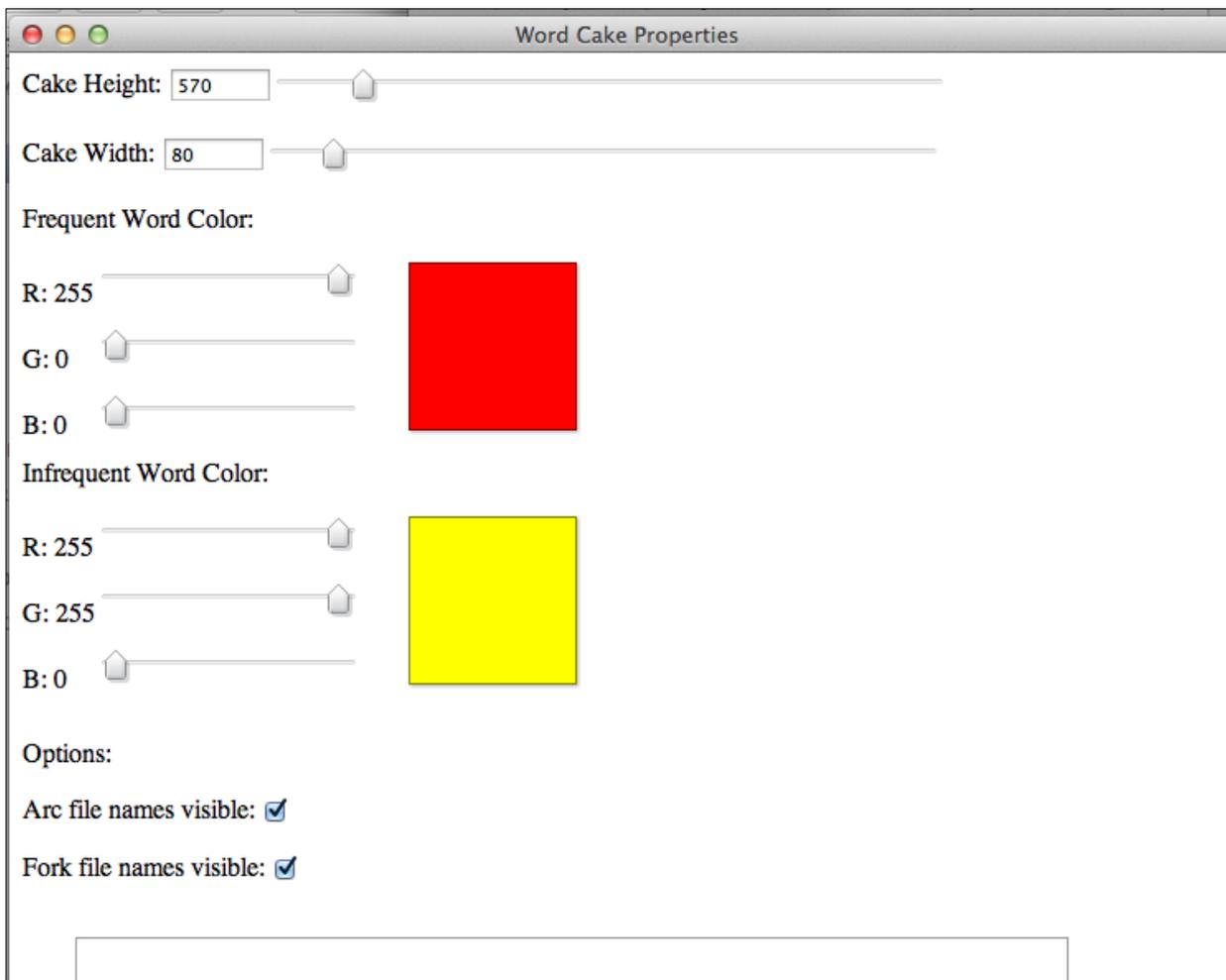
CSV's are basic spreadsheet file formats that contain values separated by commas. All common spreadsheet software is able to save and export information in this basic file format. The layout format of spreadsheet values for use in WordCAKE is as follows: 1. top left cell with total word count, 2. column of cells directly underneath this value contain words from a text and 3. corresponding column with word counts for each of the words 4. optional column for category that enables custom coloring of words



4501			total word count
billion	24	3	
continue	10	4	words
economic	14	1	
effective	12	2	word frequency counts
federal	19	1	
government	20	2	properties column: enable custom coloring by category
increase	10	3	
inflation	13	4	
interest	11	2	
jobs	14	3	
more	18	2	

Adjusting Cake Properties

Select the “Properties” icon  to access a dialog box with customization features:



Word Cake Properties

Cake Height: 570

Cake Width: 80

Frequent Word Color:

R: 255

G: 0

B: 0

Infrequent Word Color:

R: 255

G: 255

B: 0

Options:

Arc file names visible:

Fork file names visible:

Even though the WordCAKE has already loaded, the features above can be adjusted and updated in the current model. You can change the highest frequency color (red by default) to any RGB value color and the same for the lower frequency words (yellow by default). Adjust the colors to the desired hues and click “refresh” to update the WordCAKE color scheme. Note: colors that contrast with the background work best (ex. yellow against a dark background, or dark blue against a light background). You can select different backgrounds by selecting the

“Windows” menu item and clicking the “Styles” box in SketchUp. Check the “Arc file names visible” and/or “Fork file names visible” toggle to enable file name information to display when words are selected with the fork or knife tool. The text window box at the bottom of the properties window displays full text with selected word highlighted in yellow.

Section 3

This section covers tools such as the fork, knife, and navigation buttons as well as display, analysis and export features.

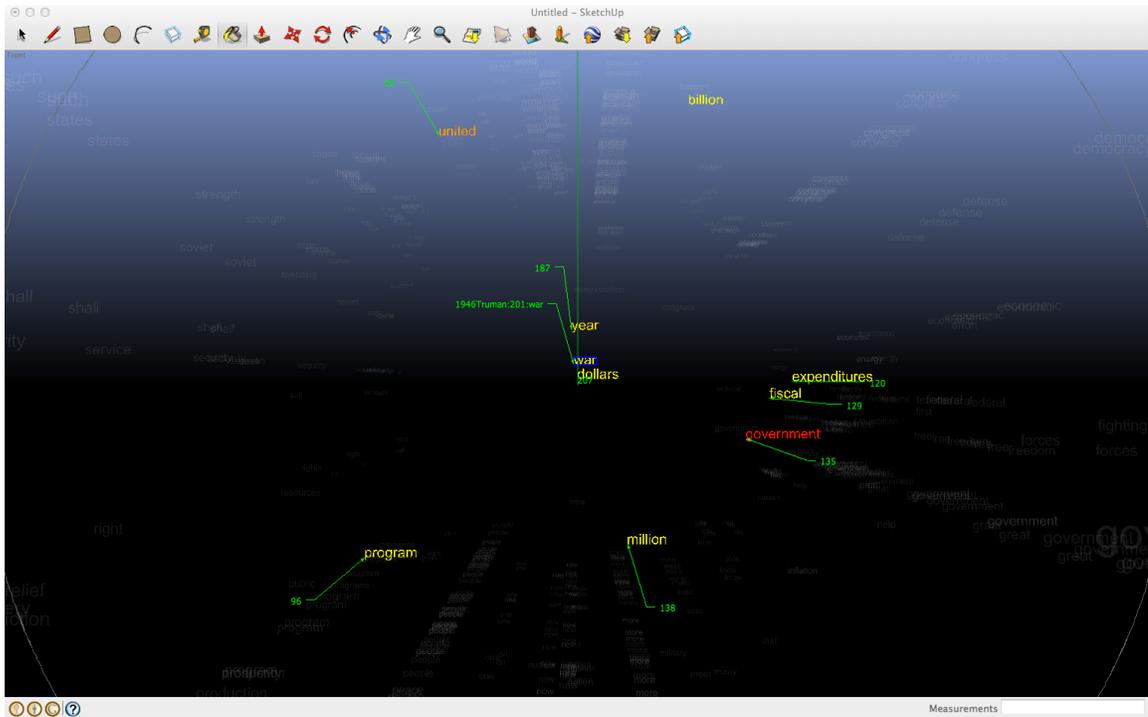
Tools and Features

WordCAKE Tool Palette



FORK

The “fork” tool can be used to select a word and this action highlights all occurrences of the word across all text rings in a WordCAKE. In addition, information on the file name and the number of instances in each ring is featured next to each word (see figure below with “budget” forked).



FLY THROUGH

Click the “fly through” button to automatically travel a path directly down the center of the WordCAKE from beginning to end. Click the button a second time to stop the motion.



ZOOM

(navigation tool) Click and drag this tool to zoom in and out of the model.



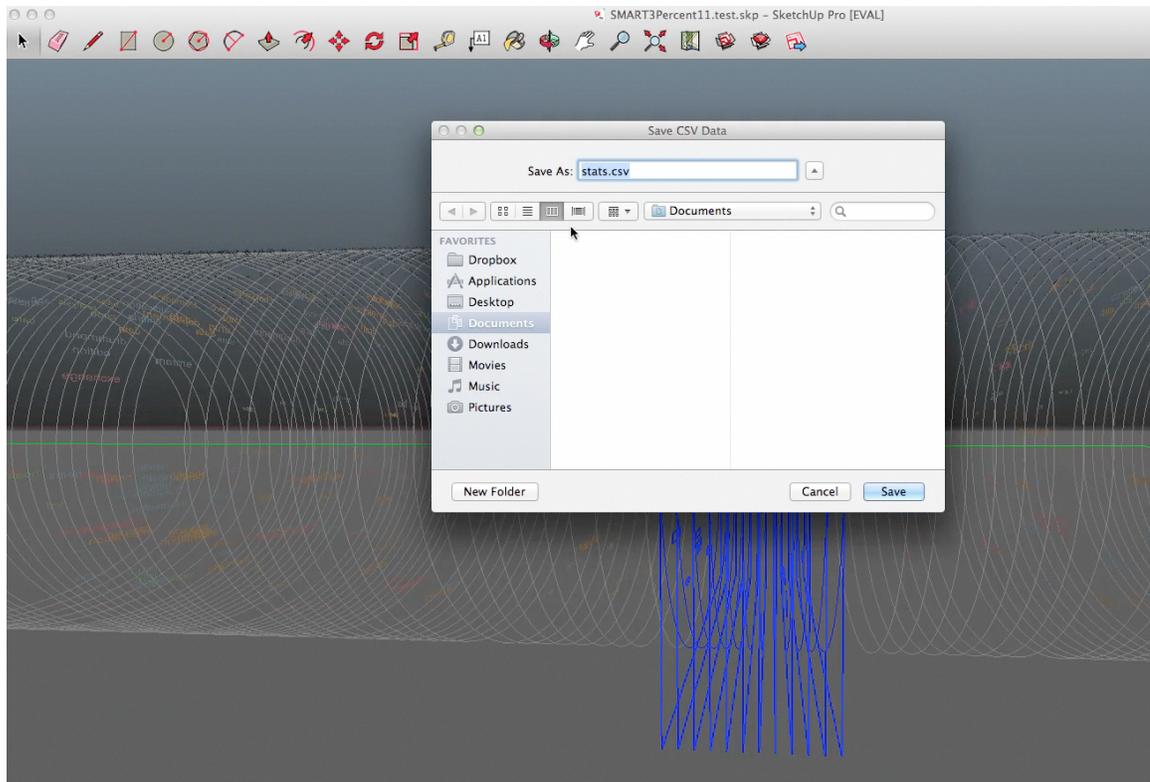
ROTATE

(navigation tool) Click and drag to swivel and shift the orientation of WordCAKE



GRAB

(navigation tool) Click and drag to pan and drag the WordCAKE model



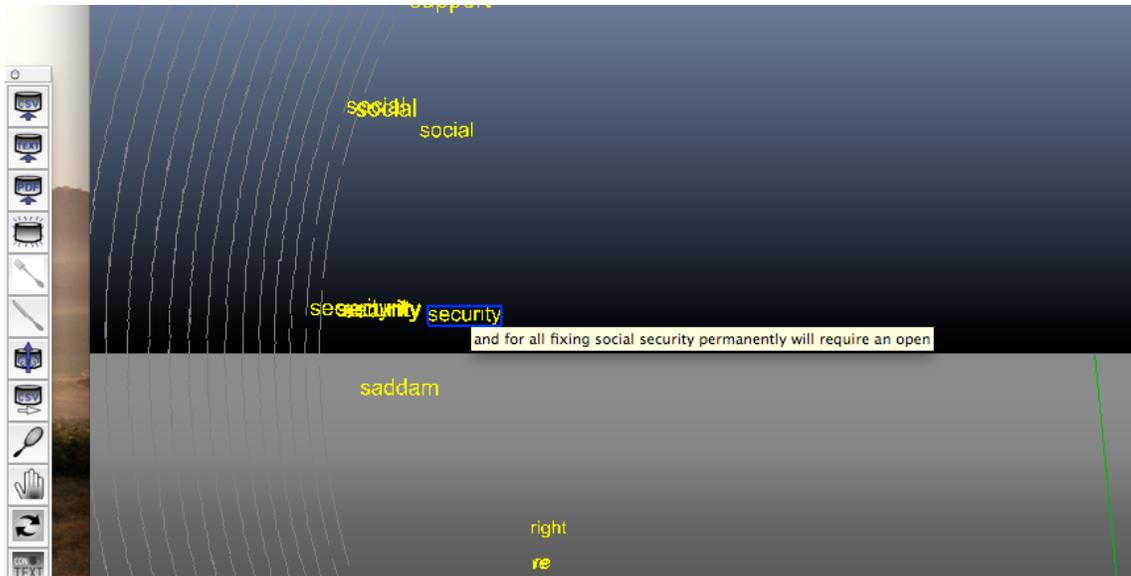
EXPORT CSV

Select a section of the WordCAKE with the cursor tool by clicking and dragging a bounding box around a section of interest (this may be easiest from the side view). All content within the selection and information on the terms is then transferred to a spreadsheet, CSV file format (see figure below).



CONTEXT VIEWER

Select a word with the context viewer cursor to display the first instance of that word in the selected text. Click again to view the next occurrence, click a third time to view the third instance, etc. Note that all punctuation has been stripped and replaced with spaces.



SAVE WITH CURRENT SETTINGS

Save a file complete with current settings (minimum threshold, style, text colors, etc.)

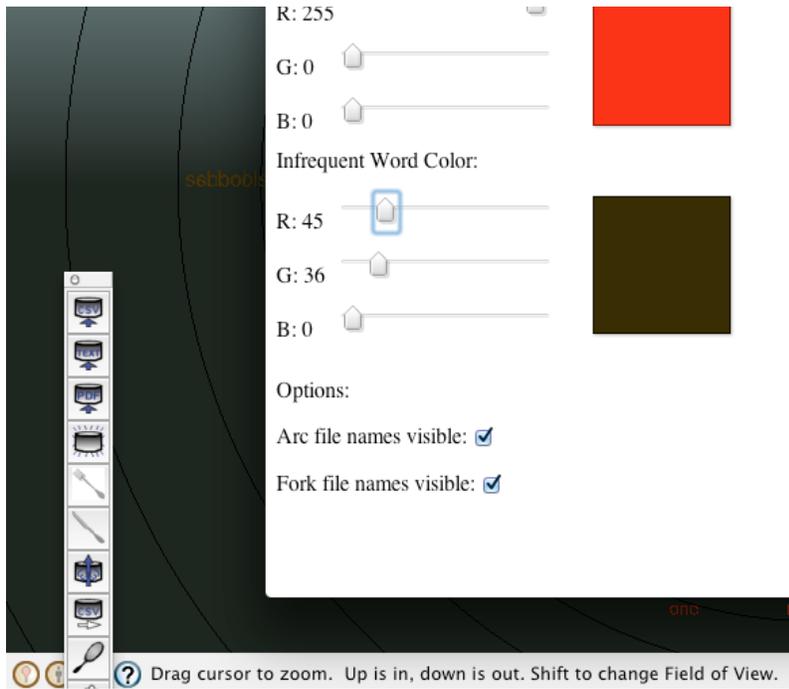


LOAD SAVED SETTINGS

Load saved settings (specified minimum threshold, style, text colors, etc.)

Features

TOGGLES

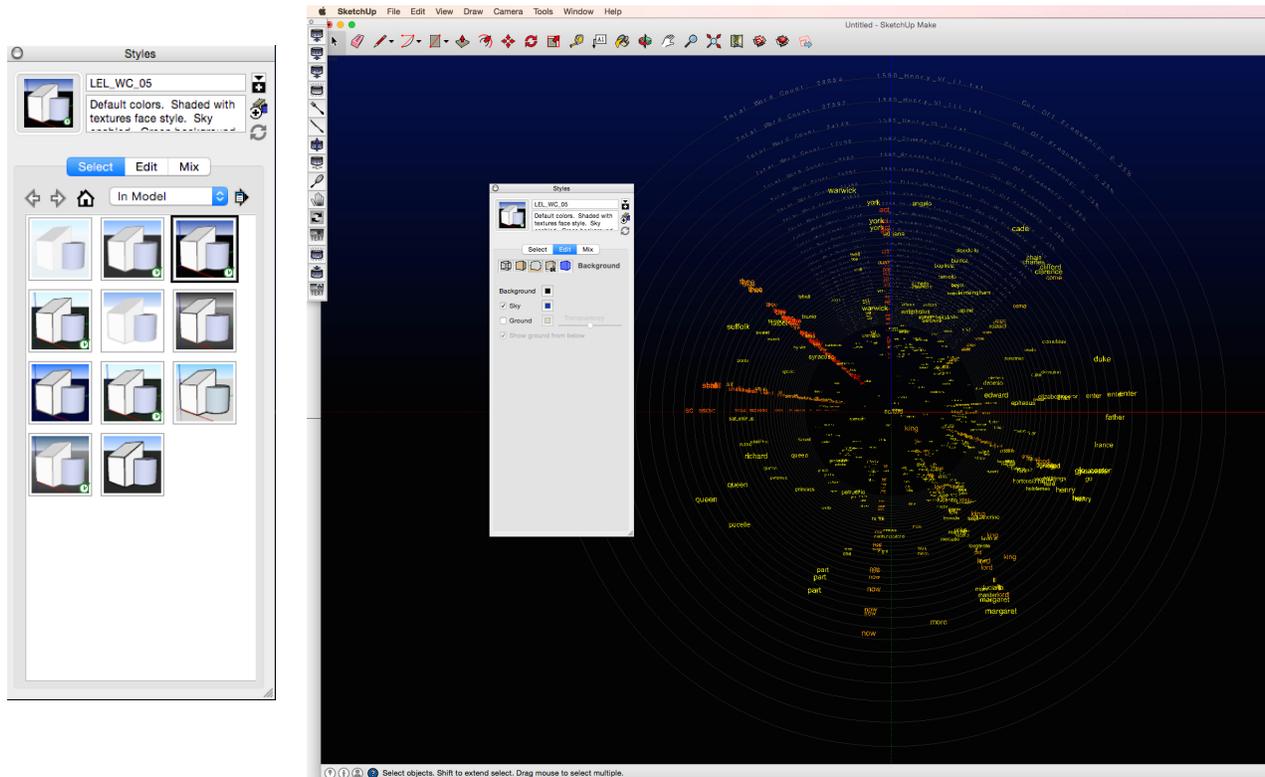


By clicking on the “Cake Properties” button, you will open a window with two toggle options at the bottom.

You can deselect the option to display the full files names, word count and minimum threshold information across the top of the WordCAKE rings by unchecking “Arc file names visible:”

The same display option applies to “Fork file names visible:” to control the display of the full file name when “forking” a word (only the word count will appear). The text window box at the bottom of the properties

window displays full text with selected word highlighted in yellow.



STYLE: ADJUST COLORS, BACKGROUND, APPEARANCE:

Under “View” the “Axis” can be removed and the “Fog” by unchecking the terms.

Under “Windows” select “Styles” and choose from several customized background/appearance options. To get rid of the “ground” transparency, select “Edit” in the “Styles” window and then choose the center icon “Background settings” and uncheck “ground.” Also under “Edit,” the icon for “Edge settings” (far left) has a “jitter” feature that can be useful, especially in visualizations of many texts.

Section 5:

HOW TO CITE THIS PROJECT:

If you use the WordCAKE tool or information contained in this document for any publication, please cite as follows:

Tredinnick, Ross, and Carrie Roy. (2014). WordCAKE Tool. Living Environments Lab, University of Wisconsin, Madison <https://extensions.sketchup.com/en/content/wordcake>

Appendix:

DEFAULT STOP LIST:

The following words are automatically omitted from the WordCAKE visualization:

a,able,about,across,after,all,almost,also,am,among,an,and,any,are,as,at,be,because,been,but,by,can, cannot, could, dear, did, do, does, either, else, ever, every, for, from, get, got, had, has, have, he, her, hers, him, his, how, however, i, if, in, into, is, it, its, just, least, let, like, likely, may, me, might, most, must, my, neither, no, nor, not, of, off, often, on, only, or, other, our, own, rather, said, say, says, she, should, since, so, some, than, that, the, their, them, then, there, these, they, this, tis, to, too, twas, us, wants, was, we, were, what, when, where, which, while, who, whom, why, will, with, would, yet, you, your