Welcome to the EmojiVerse: Final Project Report

Eve Mwangi, evelyn_mwangi@berkeley.edu
Tanya Piplani, tanyapiplani@berkeley.edu
Joyce S. Lee, joyceslee@berkeley.edu

PROJECT GOALS

Our interface focused on exploring and explaining emoji usage over time. Topics of interest included *(1) History and Evolution* as well as *(2) Meaning and Usage.* For the first topic of history and evolution, we created a timeline that highlighted important emojis that have been added over time – such as a range of skin tones and greater gender support – as well as an overview of the encoding process for new emoji.

On the subject of meaning and usage, we presented a falling emoji animation to present real-time use of certain characters, a 3D mapping of semantic meaning, as well as a scatter plot and bar chart to discuss how emojis are perceived by the reader vs. the writer.

RELATED WORK

During our literature review, many of the academic papers we found focused on interpreting meaning.
Tyler Schnoebelen is an early scholar in the area, having written his doctoral thesis[1] on emoticons, a precursor to emoji. A subsequent paper, "Do You Smile with Your Nose? Stylistic Variation in Twitter Emoticons,"[2] expands on this work, analyzing the 28 most frequent emoticons and finding that "people vary in their use of eyes, mouth shape, face direction, and whether or not they represent a nose in the face."

Additional research also supports this notion of emoji enabling flexible expression. Stark and Crawford[3] discuss how emoji act as conduits for affective labor; they describe how "these symbols do considerable work to underscore tone, introduce humor, and give individuals a quick and efficient way to bring some color and personality into otherwise monochrome networked spaces of text...[ultimately] they create new avenues for digital feeling."

Yet, the creativity of expression via emoji can also lead to inconsistent interpretation, suggesting its poor performance as a communication device. Focusing on platform differences in rendering emoji, Miller et al[4] found that each emoji has at least 17 different renderings, creating room for significant visual discrepancies. Furthermore, they found that even when people are viewing the exact same emoji on the exact same device, people disagree about whether the emotion being conveyed by an emoji is positive, negative, or neutral about 25% *of the time.*

Cramer et al[5] highlight U.S.-specific usage on mobile devices, analyzing participant descriptions of intended meaning and function of emoji. Among a large sample of participants (n=228), they found that emoji play a number of roles including "adding additional emotional or situational meaning, adjusting tone, making a message more engaging to the recipient, conversation management, and relationship maintenance." The additional function of efficiency – or not having to type more characters – was mentioned in 6% of the responses.

At CHI 2018, Wiseman and Gould[6] also presented work on how emoji are used in highly personalized – even purposefully secretive – ways among small groups. Analyzing 81 emoji reported by survey (69 unique characters), they present a range of reasons why this occurs, including the concept of an "emoji affordance."

[1]https://stacks.stanford.edu/file/druid:fm335ct1355/Dissertation_Schnoebelen_final_8-29-12-augmented.pdf
[2]https://repository.upenn.edu/cgi/viewcontent.cgi?article=1242&context=pwpl
[3]http://journals.sagepub.com/doi/pdf/10.1177/2056305115604853
[4]https://grouplens.org/site-content/uploads/Emoji_Interpretation.pdf
[5] https://dl.acm.org/citation.cfm?id=2935370
[6]https://dl.acm.org/citation.cfm?doid=3173574.3173726

Category	# messages	Top emoji	Top class
1 Addition of info	195	😊	person
(1a Expressing emotion	139	😘	person)
(1b Situational context	56	🎄	nature)
2 Changing tone	26	😊	face
3 Engagement & relationship	20	👍	person
(3a Engaging the recipient	7	💩	nature)
(3b Social & convers. norms	8	😻	person)
(3c Relationship maintenance	5	👍	person)

Figure 1. Distribution of top-used emoji and the functional role they play. (Cramer et al, 2016)

Looking at broader approaches to sentiment analysis, Novak et al[7] introduced the first Emoji Sentiment Ranking in 2015, drawing a sentiment map of the 751 most frequently used emoji. Engaging 83 human annotators, their process included labeling over 1.6 million tweets in 13 European languages by the sentiment polarity (negative, neutral, or positive). They draw several interesting conclusions, including the fact that that most of the emoji are positive, especially the most popular ones. They also suggest that emoji help align people on meaning: sentiment distribution of tweets with and without emojis is significantly different, and the inter-annotator agreement on the tweets with emoji is higher.

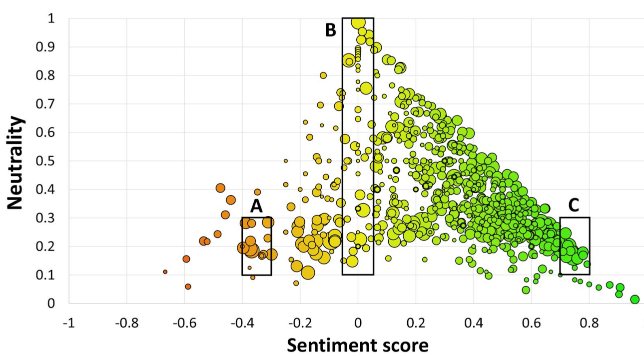


Figure 2. Sentiment map of 751 emoji, with bubble size proportional to log10 of the emoji occurrences in the Emoji Sentiment Ranking (Novak et al, 2015).

More recently, Wijeratne et al[8] took a computational approach to semantic similarity, using emoji descriptions, emoji sense labels, and emoji sense definitions, based on training data from Twitter and Google News. They created a new dataset (EmoSim508), which assigned human-annotated semantic similarity scores to a set of 508 selected emoji pairs, in order evaluate their embedding models' measurement of emoji similarity.

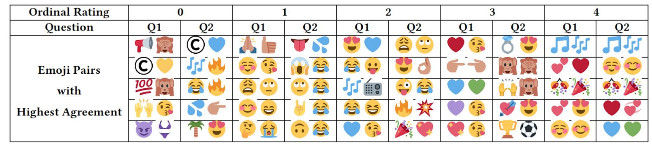


Figure 3. Top 5 emoji pairs with highest inter-annotator agreement (Wijeratne et al, 2017).

Beyond academia, we were able to gather some global insights on usage from the 2015 Emoji Report from SwiftKey,[9] a mobile phone keyboard company. This report provided statistics on adoption of various emoji category worldwide – for instance, that the French use 4 times as many heart emoji than other languages; likewise, flowers and plants emoji are used by Arabic speakers at more than 4 times the average rate. We also examined the website EmojiTracker[10] (Figure 4) which showcases real-time data of emoji usage on Twitter.

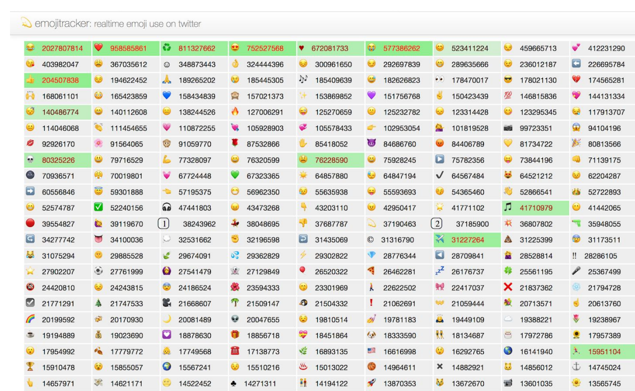


Figure 4. EmojiTracker offers an API to track real-time emoji use on Twitter.

Regarding the emoji evolution, we dove into documentation from the Unicode Consortium[11] as well as the work of organization EmojiNation,[12] who helps create proposals for new emojis –"for the people and by the people." Last but not least, we discovered the project Lobe,[13] a tool that enables building deep learning recognition models: in Figure 5, we see how it is used to recognize images as emoji.

[7]http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144296
[8]http://knoesis.org/people/sanjayaw/papers/2017/Wijeratne_WebIntelligence_2017_Emoji_Similarity.pdf
[9]https://www.scribd.com/doc/262594751/SwiftKey-Emoji-Report
[10]http://www.emojitracker.com/
[11]https://www.unicode.org/faq/emoji_submission.html
[12]http://www.emojination.org/
[13]https://lobe.ai/

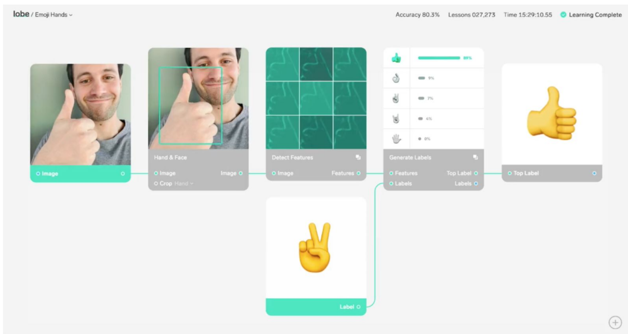


Figure 5. Lobe is a visual tool to build trained deep learning models, such as recognizing images as emoji.

OUR VISUALIZATION

To achieve our project goals, we told the story of emoji in 3 sections. For a cohesive narrative, they are ordered as follows: (1) real-time usage, (2) meaning, then (3) history and evolution. (See the appendix for full-width screenshots).

Real-Time Usage

To grab the user's attention, we open with a high-level overview so that they know what to expect. This is accompanied by a gentle animation of emoji falling according to real-time use on Twitter. The animated arrow below the text is our call to action, urging the user to keep scrolling to learn more (Figure 6).

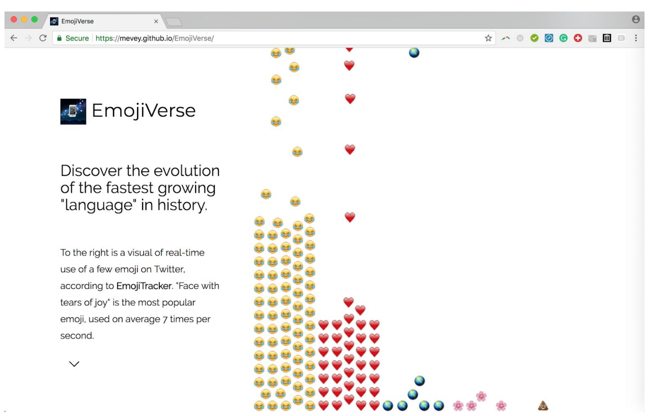


Figure 6. Our landing screen, featuring an animation of falling emoji based on real-time Twitter use.

The text to the left is our value proposition, while the animation creates initial intrigue. Frequency of emoji use is mapped to how many fall into the pile: this renders the fallen emoji as "bar graphs" that enable quick comparisons within a snapshot of time.

We picked 5 emoji to highlight, for the following reasons. The face with tears of joy (😂) and red heart (❤) emoji are the most popular, so we wanted to highlight how often they are used. The globe (🌎) and cherry blossom (🌸) emoji were selected for seasonal relevance, as they respectively represent Earth Day and spring, and certain emoji are used more during certain times of year. Lastly, we selected the pile of poo emoji (💩) for comic relief.

We chose this visualization over an emoji "word cloud," as previously proposed: there are thousands of emoji, and the size of emoji in a cloud would not have told the story well. Furthermore, limiting this visualization to a static set of 5 emoji not only keeps it clean, but also more clearly indicates variation in color, aesthetics, and rates of real-time use.

Meaning

We next allow users to discover general semantic similarity of emoji. We used a 3D rendering to give an overall picture of how many emoji there are, as well as how diverse they can be in meaning. The rotating visual of dots also aligns with our original idea of an "emoji universe," with emoji visually mirroring the metaphor of stars rotating around a central force (Figure 7).

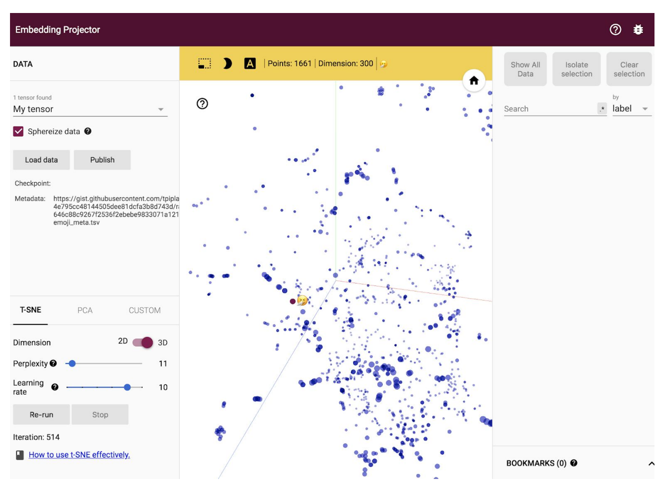


Figure 7. Our "emoji universe,"[14] where an embedding projector maps emoji meaning in 3D space.

[14]http://projector.tensorflow.org/?config=https://gist.githubusercontent.com/tpiplani/e4580001076e7734403c1080f3f610e1/raw/7349d2eb162d7b2aebfbaa25119d098f6644a83c/config_file.json

Users can select t-SNE or PCA modes to reduce the dimensionality of the vectors from 300 to 3 dimensions, in addition to zooming in and out to navigate around the dots. Clicking on a dot highlights its most semantically similar neighbors along the right-hand panel. For instance, clicking on the fish emoji (🐟) rendered sushi (🍣) as a neighbor, along with other ocean creatures.

Beyond broader meaning, however, we wished to further explore the effectiveness of emoji as a communication tool. To do so, we present 2 visualizations based on a dataset comparing sentiment among readers and writers of emoji. The first is a scatter plot, mapping reader sentiment on the X-axis and writer sentiment on the Y-axis (Figure 8). Overall we see general correlation, suggesting that the symbols enable empathy, with negative emoji in the bottom left, and more positive, celebratory emoji in the top right.

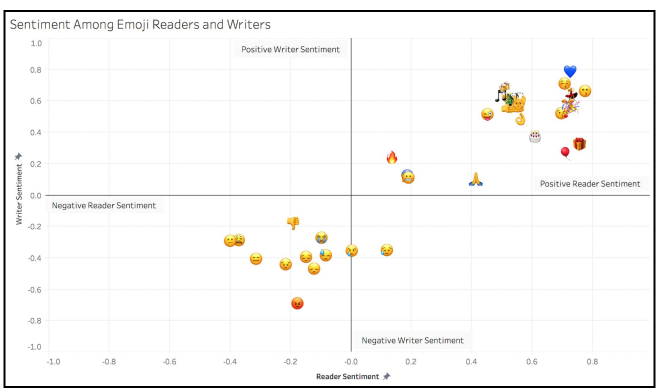


Figure 8. Scatter plot of emoji reader sentiment (X-axis) vs. writer sentiment (Y-axis):[15] *both values are scaled between values (-1, 1).*

Examining the difference between reader and writer sentiment (Figure 9), however, some emojis do seem to mean more to one party than the other. For instance, writers are more likely to have more positive sentiment than readers with the clinking beer mugs symbol (🍻) – but less positive affect with the red angry face emoji (😡). Perfect communication would call for low difference values: we see that writers and readers most align with the ok hand (👌).

History and Evolution

The history and evolution is told in a vertical timeline, which aligns with our top-down storytelling theme. We make use of a long gap between years 1999 and 2010 to mirror the period between emoji invention and growth: this also contrasts with later, more recent years when emoji become increasingly popular.

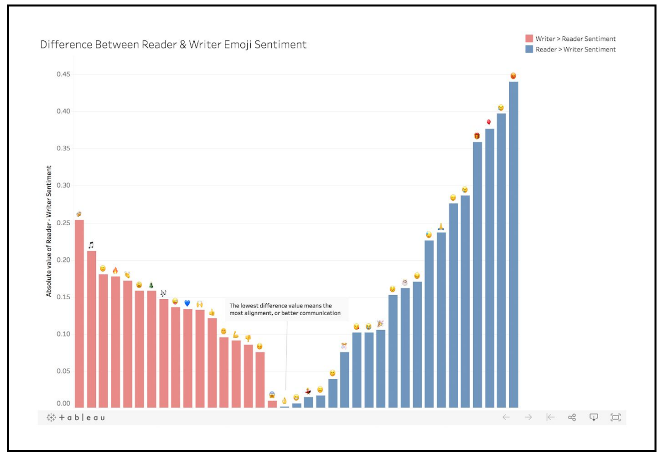


Figure 9. Absolute value of difference between emoji reader and writer sentiment:[16] *higher writer sentiment is red, while higher reader sentiment is blue.*

For each year, we describe significant events that occured, using text and – of course – emoji. Boxes in the timeline are animated upon scrolling: this acts as a visual cue to direct attention to the next box in the sequence.

Figure 10. Timeline of important events in emoji history, including the evolution toward more inclusive representation.

DATA SOURCES

We identified a number of data sources early on, based on our different goals. These included Miles Johnson's EmojiBase,[17] Cal Henderson's Emoji Data,[18]

---

[15]https://public.tableau.com/profile/tanya4487#!/vizhome/EmojiClusters/Sheet2?publish=yes

[16]https://public.tableau.com/profile/tanya4487#!/vizhome/Reader_Writer_difference/Sheet1?publish=yes

[17]https://github.com/milesj/emojibase

[18]https://github.com/iamcal/emoji-data

Rachael Tatman's EmojiNet,[19] and Sanjaya Wijeratne's EmoSim508.[20] However, for our actual implementation, we used 3 primary data sources – described in detail below – in addition to the Unicode website and other linked news sources (for the historical timeline).

EmojiTracker

For animation of falling emoji, we used the rankings API[21] from EmojiTracker, previously discussed in "Related Work" (see Figure 4). The API is called 3 times when the user loads the page: the differences in emoji use at that moment are calculated, which then informs the rate at which the emoji drop throughout the user's session. The animation stops once the user scrolls past the opening view, and the number of API calls is reduced to prevent the browser from freezing. The page also refreshes automatically, to prevent any of the emoji "bar columns" from overfilling.

Emoji2Vec

For our 3D visualization, we used the emoji2vec[22] repository, which includes pre-trained embeddings for all emojis, learned from their description in the Unicode standard. It also includes the code used to train emoji representations, the training data used to do so, as well as additional tools for analyzing the performance of the vectors trained. The repository accompanies a paper by Eisner et al, titled "emoji2vec: Learning Emoji Representations from their Description."[23]

Emoji Sentiment

From Kaggle, we used the Emoji Sentiment[24] dataset to understand how emojis are perceived by the writer vs. the reader. Spanning over 3 years and 4,000 employees of 56 companies, the dataset compares self-reported happiness scores with emoji use in comments. (358 different types of emoji are used in 3,506 comments.) The dataset accompanies the 2017 paper by Berengueres and Castro, titled "Sentiment Perception of Readers and Writers in Emoji use."[25]

[19]https://www.kaggle.com/rtatman/emojinet/data
[20]https://www.kaggle.com/sanjayaw/emosim508
[21]http://emojitracker.com/api/rankings
[22]https://github.com/uclmr/emoji2vec
[23]https://arxiv.org/pdf/1609.08359.pdf
[24]https://www.kaggle.com/harriken/emoji-sentiment
[25]https://arxiv.org/abs/1710.00888

TOOLS USED

To create our visualizations, we used the following tools: Visual Sedimentation, Embedding Projector, and Tableau. We also referred to code by George Martsoukos[26] to create our timeline.

Visual Sedimentation

For our falling emoji animation, we used Visual Sedimentation,[27] a JavaScript library for visualizing streaming data. Built on top of D3.js, jQuery and Box2DWeb, it is inspired by the process of physical sedimentation in the natural world – for example, when snow builds up in winter (Figure 11). This library was built by Huron et al in 2013, and is accompanied by an IEEE paper outlining the process of its development, "Towards Visual Sedimentation."[28]

Figure 11. Snow building up in winter, inspiration for the Visual Sedimentation library.

Embedding Projector

To create our 3D "emoji verse," we used the Embedding Projector[29] offered by TensorFlow. This tool enables interactive visualization of embeddings, rendering them in 2 or 3 dimensions. The projections panel (bottom left) allows the user to select the type of projection, and the inspector panel (right side), features a list of nearest neighbors, once an area has been selected. (See Figure 7.)

[26]https://webdesign.tutsplus.com/tutorials/building-a-vertical-timeline-with-css-and-a-touch-of-javascript--cms-26528
[27]http://www.visualsedimentation.org/index.html
[28]https://hal.inria.fr/file/index/docid/734084/filename/visual-sedimentation-final.pdf
[29]http://projector.tensorflow.org/

Tableau

We used Tableau to visualize our reader-writer sentiment dataset, as both a scatter plot and a bar graph. To add emoji as labels for both these visualizations, we first converted the emoji hex unicodes to decimal values, then converted them to characters.

EVALUATION PROCEDURE

We took a qualitative approach to assessing our project performance, using observed interactions with the website and brief follow-up interviews as our evaluation methods. Key areas of observations included facial expressions (smiling, frowning, laughing, etc.), persistent errors (e.g. user kept tapping on static arrow), as well as questions or other remarks said aloud. Follow-up questions were directed toward general feedback, with the following scripted questions:

- *Overall, what did you like and dislike about this website?*
- *What did you think of how the information was presented? (This includes understanding the words, layout, animations, etc.)*
- *Do you have any other thoughts on what you would have liked to see, or suggestions of how we might improve the project?*

RESULTS

During a 2-hour demonstration, we evaluated the visualizations with a sample of 16 participants. Results are categorized below, according to each section of the website or overall appearance.

Falling Emoji Animation

As we suspected, the opening animation was eye-catching and effectively drew many people to our demonstration area. Participants frequently smiled and laughed while looking at this visualization, making remarks such as *"This is cute," "So cool,"* and *"I just want to watch it forever."* A number of visitors also tried to click on the falling emoji, as if it were a game.

Some individuals expressed confusion over the 5 emoji presented, however, interpreting them as the most popular, rather than a curated selection. (One person did pick up on the seasonality of our emoji choices though, speculating *"Maybe in winter, there will be a lot of snowmen emoji?"*) An improved design for the future could be to increase interactivity by allowing the user to populate this visualization with emoji of their own interest.

3D Embedded Projector

Like the first visualization, many people responded positively to the 3D embedded projector, with utterances like *"Wow, cool!"* and *"This is so next level."* While people found the visualization interesting – even mesmerizing – they also found it more difficult to interpret. People gave us feedback like, *"It's a little confusing," "Maybe good for expert users of TensorFlow, but not for everyone else,"* and *"I'm having trouble understanding what it's showing me; it's a little too complex for me."*

One participant remarked that he was *"just tempted to skip over"* the section due to its complexity, and suggested that we provide clearer instructions on how to navigate and make meaning from the visualization. While we recognized it wouldn't be possible to explain all the details of TensorFlow, we did separate and enlarge the instructional text for how to navigate the visualization in response to this feedback (Figure 12).

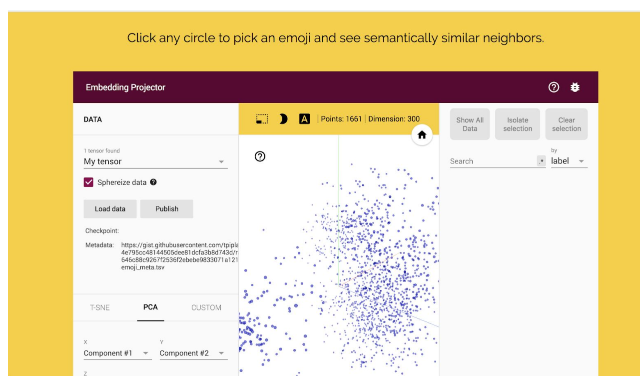


Figure 12. Clearer instructions on how to navigate the 3D "emoji verse," added in response to feedback.

Reader vs. Writer Scatter Plot and Bar Graph

The response to scatter plots and bar graphs were mostly positive, given that they were simpler, static renderings. However, participants suggested that we include more annotations to clarify takeaways – rather than including them in the accompanying text. One specific annotation suggested was to label each of the scatter plot quadrants, to clarify what each quadrant means: we have done this accordingly (see Figure 8).

Another major paint point was about reading the bar graph, as the negative vs. positive comparison implied the nature of the sentiment, rather than emphasizing the difference: we addressed this by changing our graph to the absolute value of the difference, so that all values are positive. One person also suggested lightening the transparency of the bars, so that they visually competed less with the emoji labels. (See Figure 9 for implemented changes.)

Timeline

Most people appreciated the direct presentation of facts through the timeline. The one contentious aspect that people remarked upon was the gap between 1999 and 2010. As one participant described, "*I know its really far apart, but it looks weird;*" however, others remarked that "*it makes sense*" to show the lag between the introduction of emoji vs. when they became popular. Given the mixed results, we decided to stick with our original spacing. Some people also expressed interest in learning more about the timeline for submitting emoji proposals, so we added a new section about this process (Figure 13).

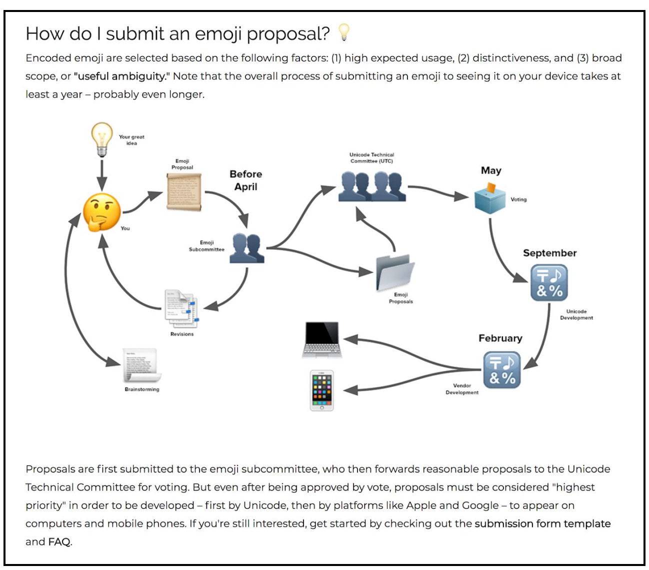


Figure 13. A new section about the process for encoding new emoji. The graphic is adapted from a presentation[30] by Mark Davis, President and Co-Founder of the Unicode Consortium.

Overall Aesthetics

The main feedback we received about overall appearance focused on typography and color. A number of people mentioned that the font size was "a little small" and that there was too much leading, or vertical spacing. We adjusted for both accordingly. Originally we had attempted to match our background color to the 3D embedding projector; however, we received feedback that it was "not an emoji color;" therefore, we updated it to yellow.

Lastly we received feedback that the photographs of ourselves (at the bottom of the page) should be "emojified," which we did accordingly (see Figure 14).

Figure 14. Our new character portraits, created with Bitmoji.[31] From left to right: Eve, Tanya, and Joyce.

WORK ALLOCATION

All team members contributed to decision making and creating the final visualization; however, each of us specialized and managed different areas. Tanya led on data processing and analysis, Eve led on writing code and deployment, while Joyce led on deliverables, usability testing, and visual design. See Table 1 for a summary of contributions.

Table 1. Summary of Work Contributions

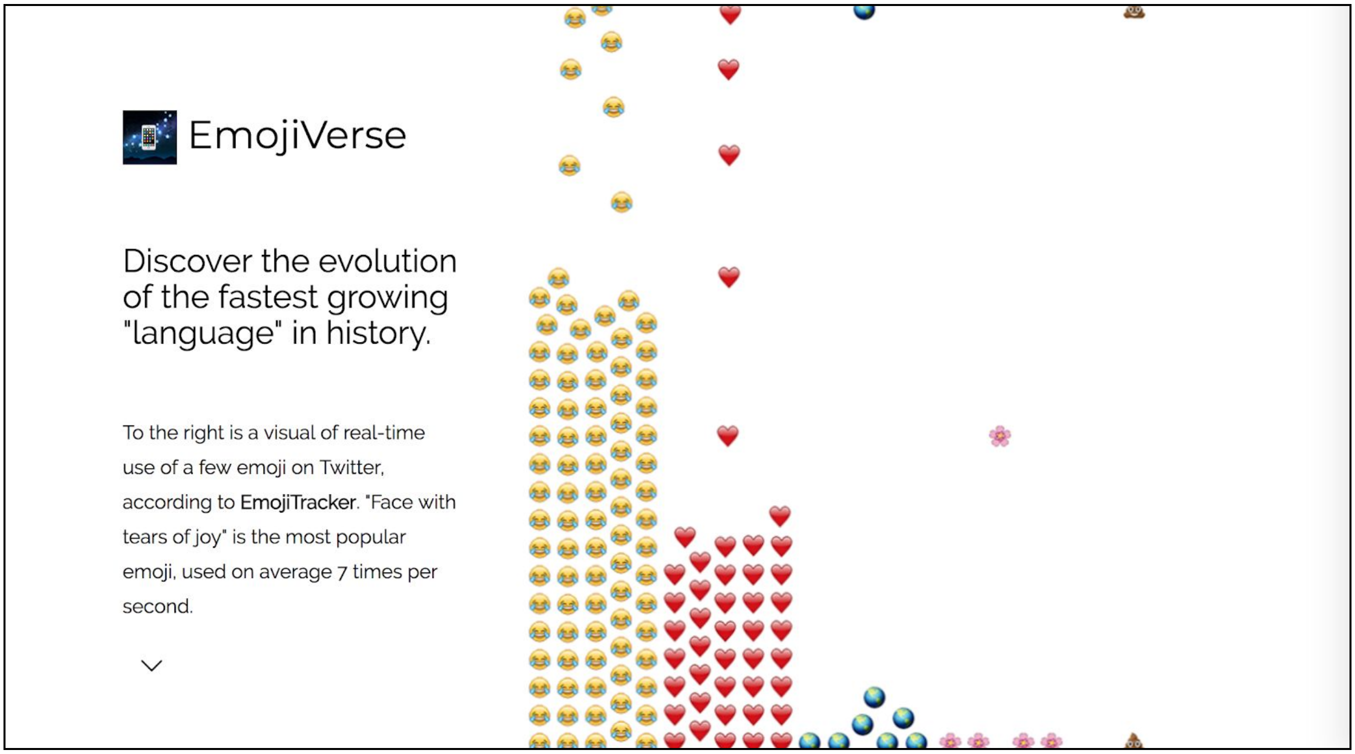
Task	**Tanya**	**Eve**	**Joyce**
Finding datasets	35%	35%	30%
Conducting preliminary literature review	33%	33%	33%
Sketching wireframes	20%	40%	40%
Mid-project deck and presentation	30%	10%	60%
Analyzing and visualizing data	90%	5%	5%
Writing and deploying website code	5%	90%	5%
Evaluating usability	40%	40%	20%
Writing final report	10%	10%	80%
Average Contributions	**33%**	**33%**	**33%**

[30]http://unicode.org/emoji/slides.html

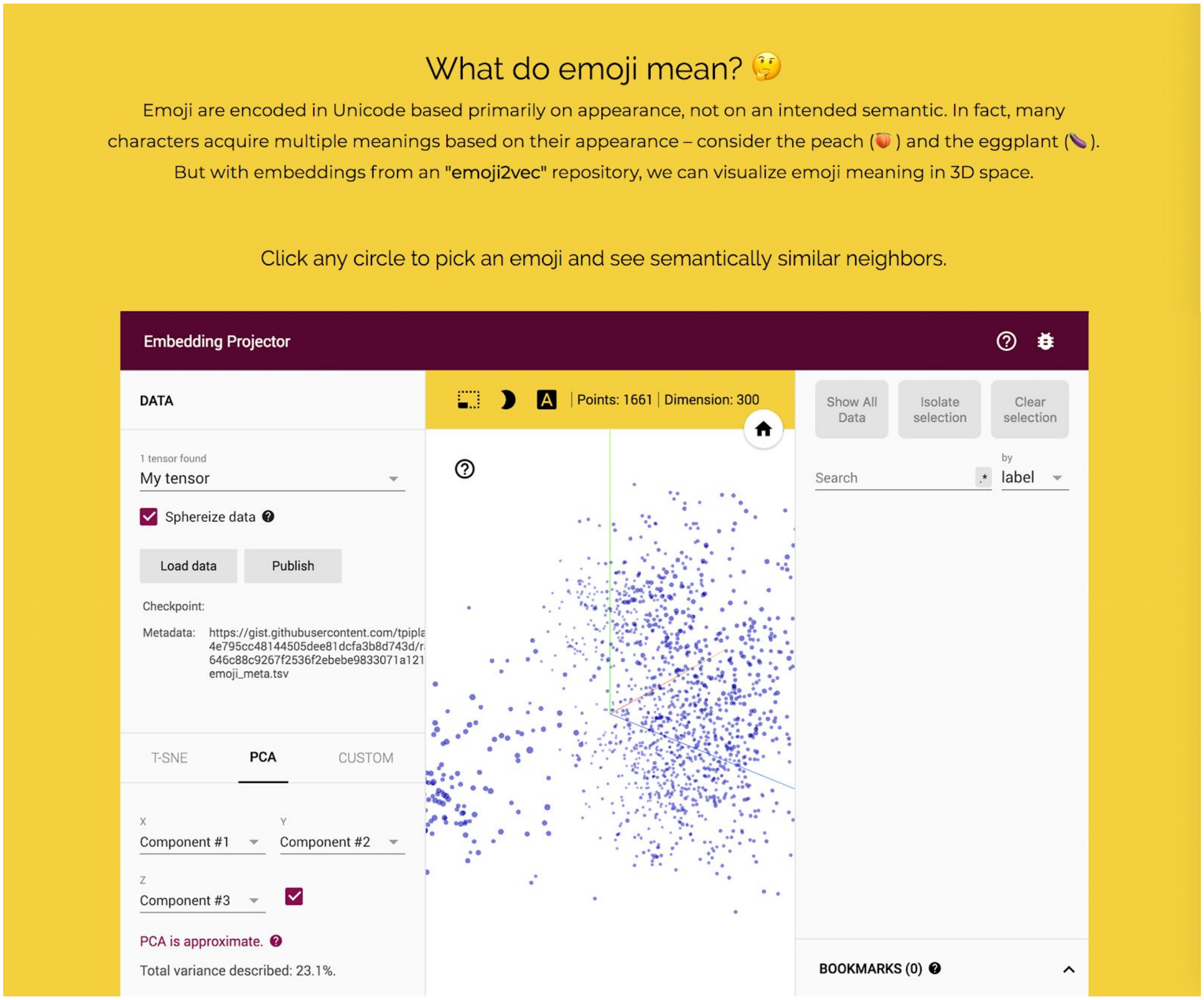
[31]https://www.bitmoji.com/

APPENDIX

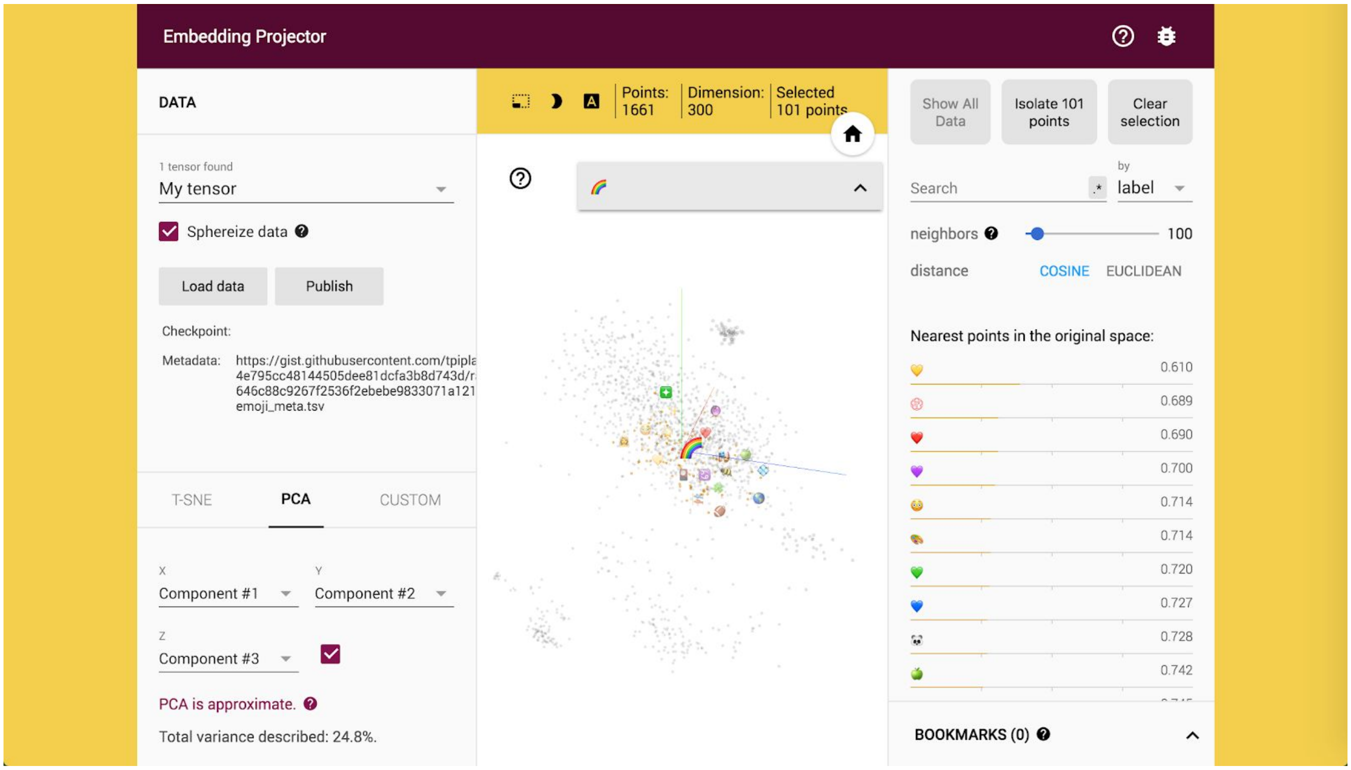
See below for full-width, in-depth screenshots. The website can be viewed at https://mevey.github.io/EmojiVerse and the code is available on GitHub at https://github.com/mevey/EmojiVerse.



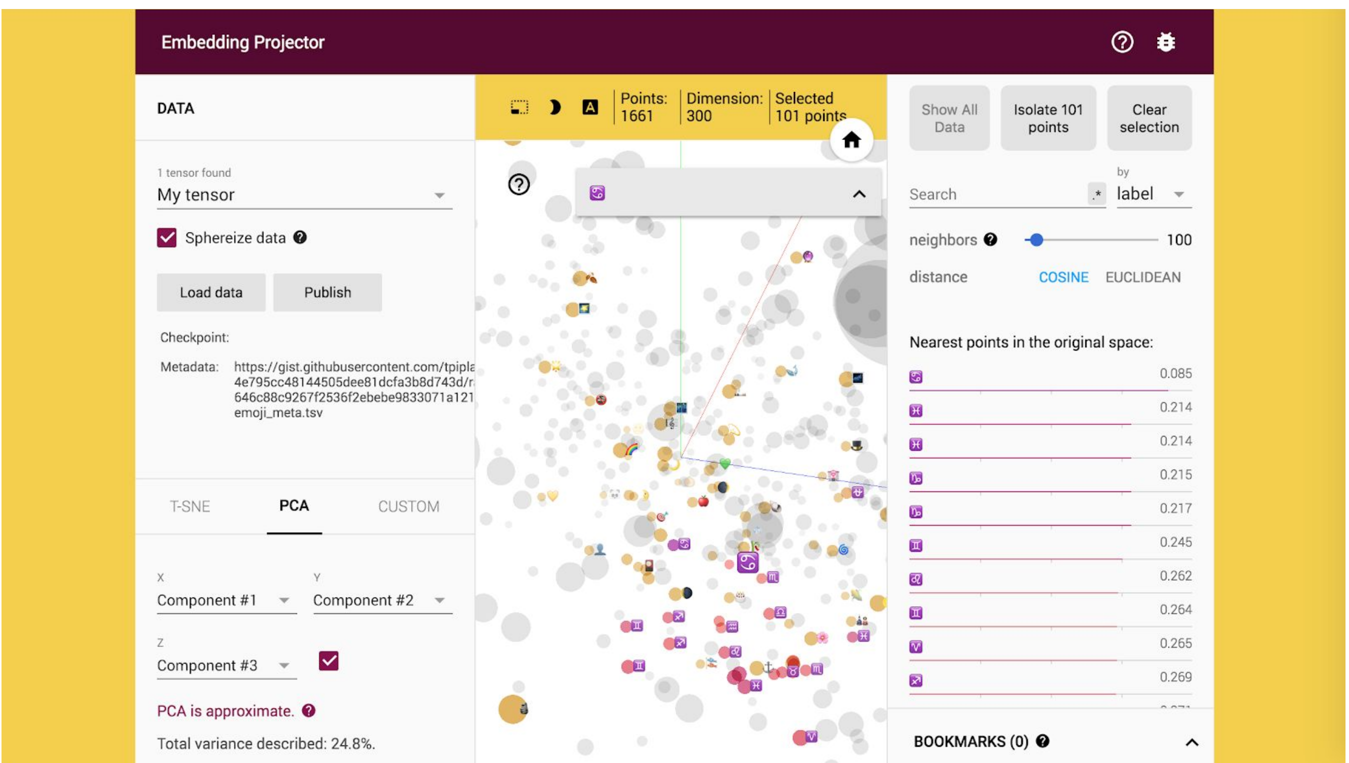
Our landing screen, featuring an animation of falling emoji based on real-time Twitter use.



Our 3D "emoji universe," created with TensorFlow's Embedding Projector.



A zoomed out view, with the rainbow symbol selected. Semantically similar emoji appear on the right panel.



A zoomed in view, showing semantic similarity of horoscope symbols.

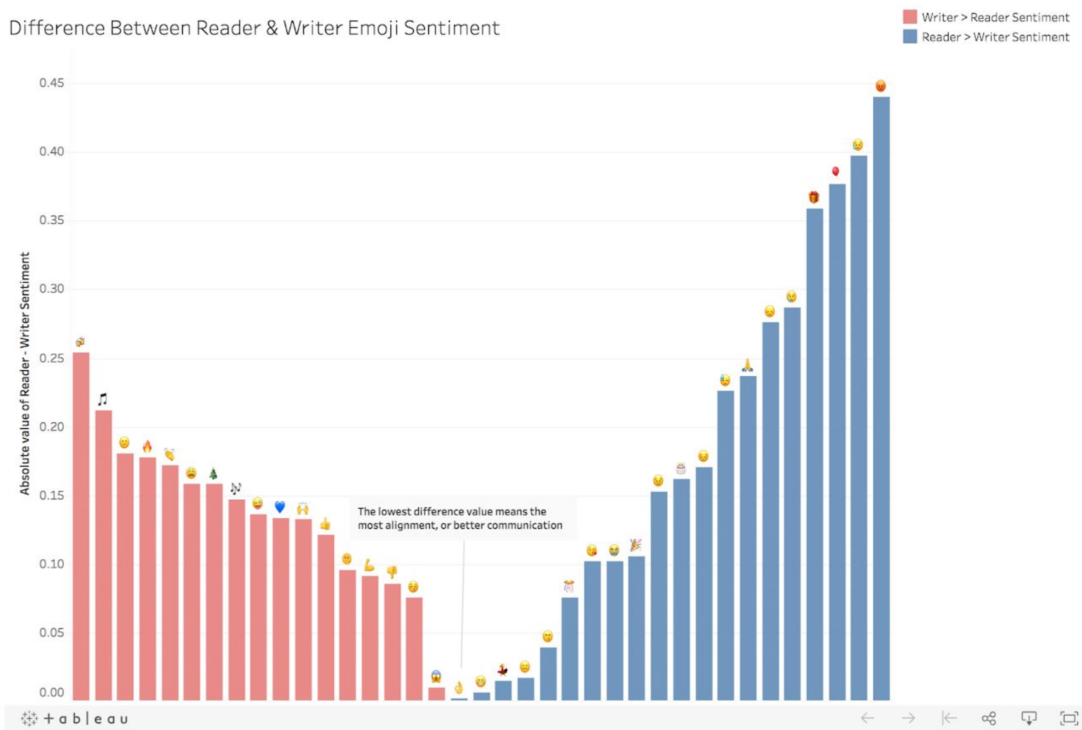
So, do emoji help us communicate? 💬

Researchers have found that **about 25% of the time**, people disagree about whether the emotion being conveyed by an emoji is positive, negative, or neutral – even when they are viewing the same emoji on the same type of device. Yet, examining Jose Berengueres' **emoji sentiment dataset** below, we do see a general correlation between sentiment of readers and writers of emoji: the symbols enable empathy, with negative emoji in the bottom left, and more positive, celebratory emoji in the top right.

Our scatter plot comparing sentiment among emoji readers and writers.

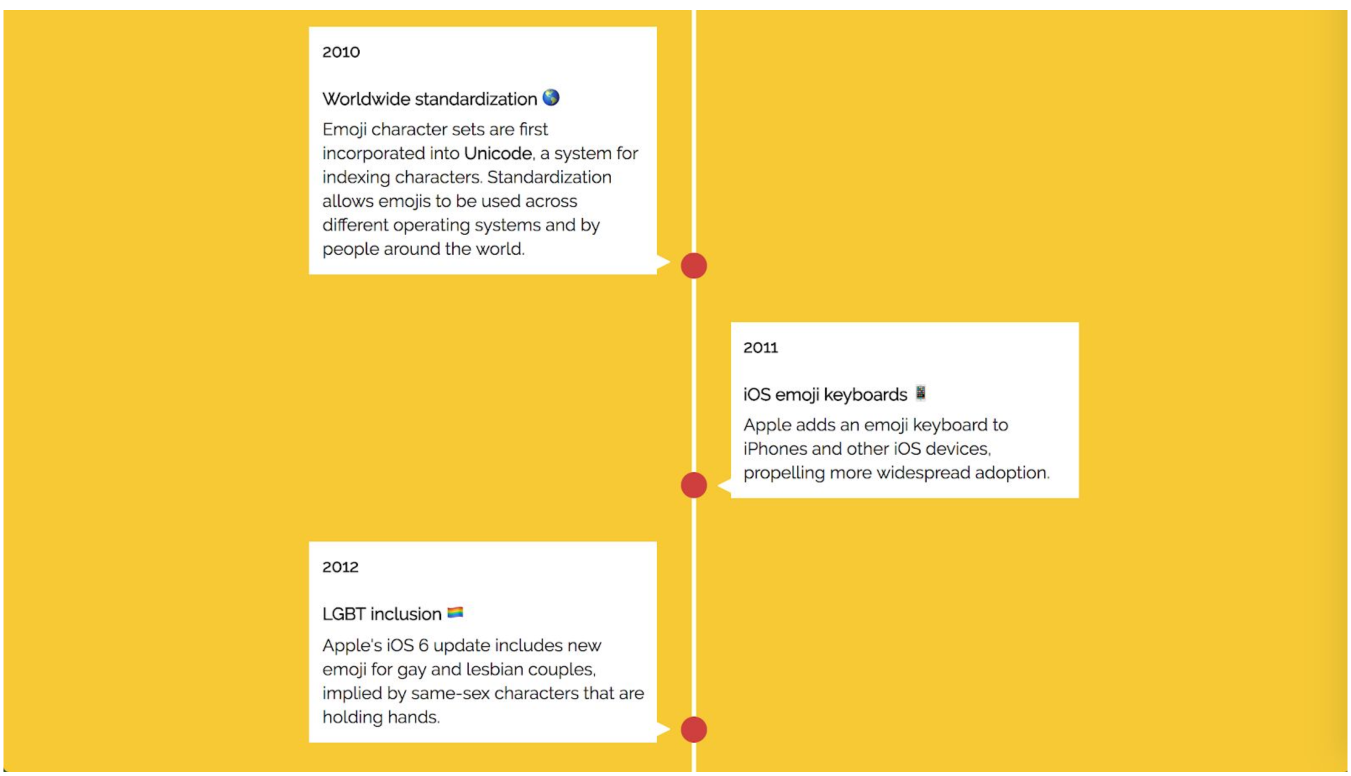
"Traditional alphabet scripts have been struggling to meet the rapid-fire, visually focused demands of 21st-century communication," says Casper Grathwohl, President of **Oxford Dictionaries**. "It's not surprising that a pictographic script like emoji has stepped in to fill those gaps — it's flexible, immediate, and infuses tone beautifully."

Examining the difference between reader and writer sentiment below, however, some emojis do seem to mean more to one party than the other. Perfect communication would call for low difference values: we see that readers and writers align most on sentiment with the ok hand (👌).

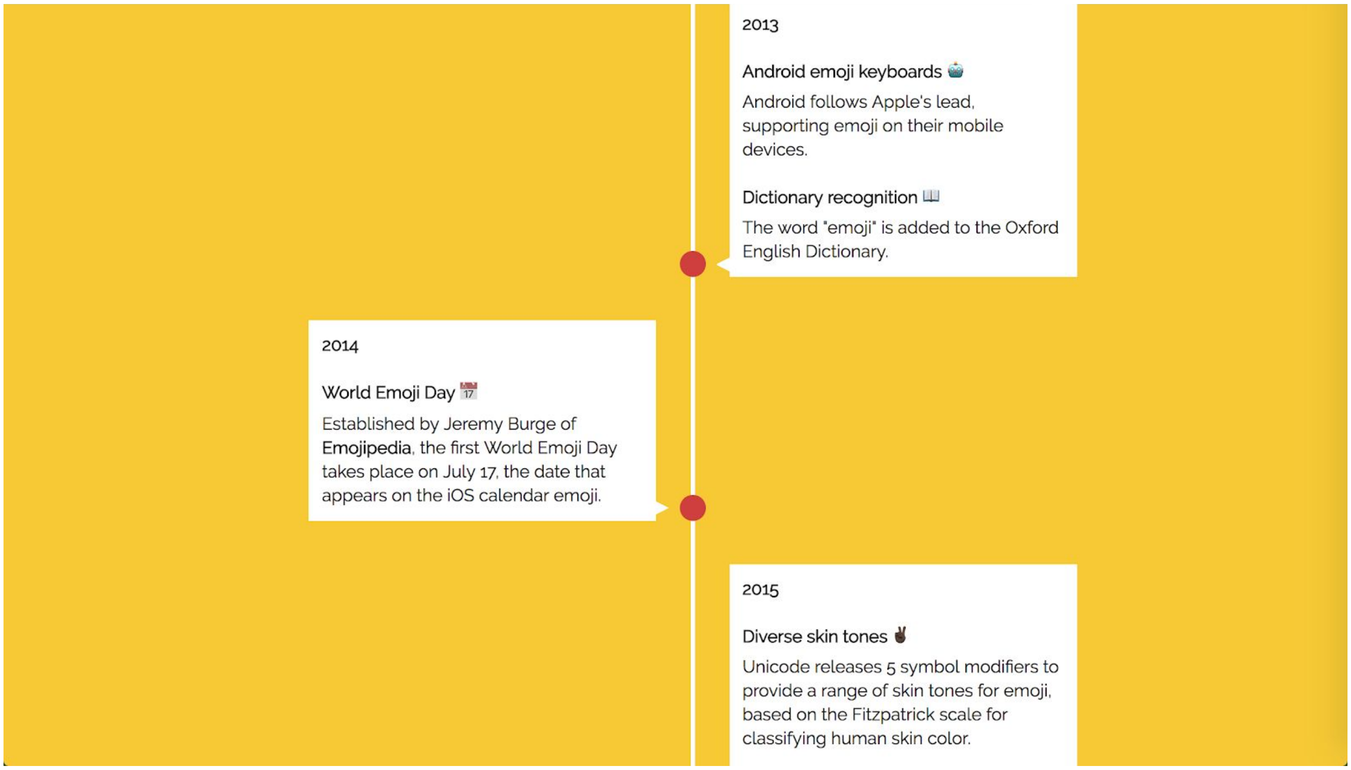


Our bar graph comparing the difference in sentiment among emoji readers and writers.

Our timeline of emoji history and evolution (part 1 of 5).

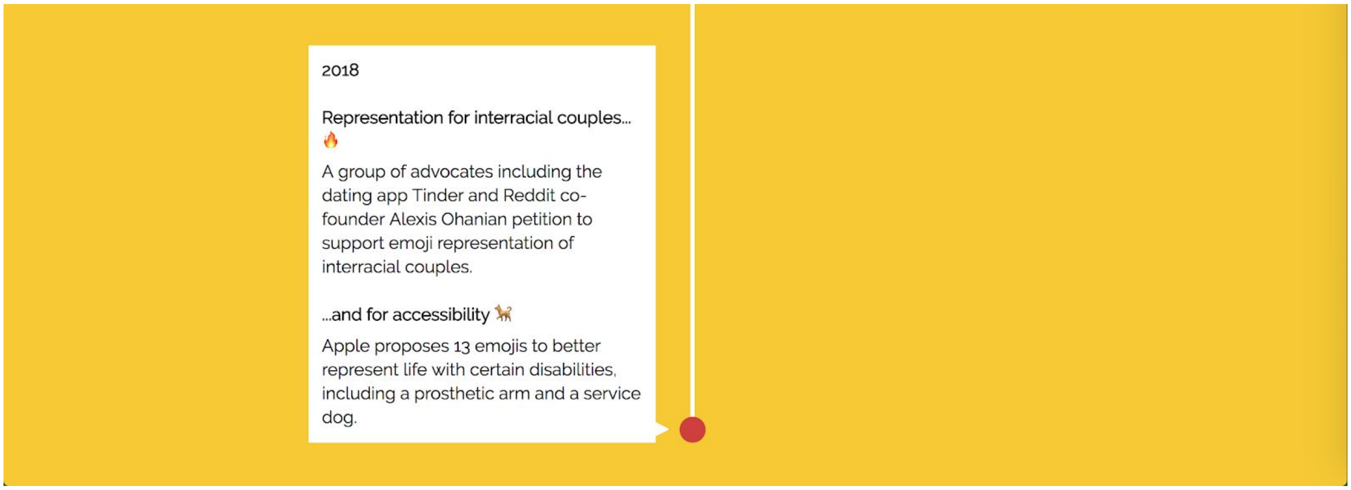


Our timeline of emoji history and evolution (part 2 of 5).



Our timeline of emoji history and evolution (part 3 of 5).

Our timeline of emoji history and evolution (part 4 of 5).


2018

Representation for interracial couples... 🔥

A group of advocates including the dating app Tinder and Reddit co-founder Alexis Ohanian petition to support emoji representation of interracial couples.

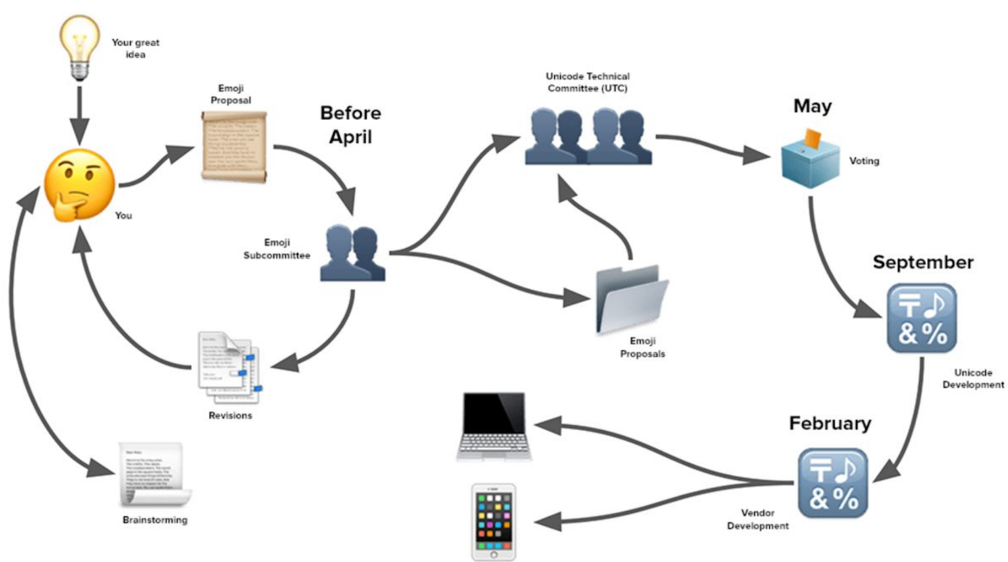
...and for accessibility 🐕‍🦺

Apple proposes 13 emojis to better represent life with certain disabilities, including a prosthetic arm and a service dog.

Our timeline of emoji history and evolution (part 5 of 5).

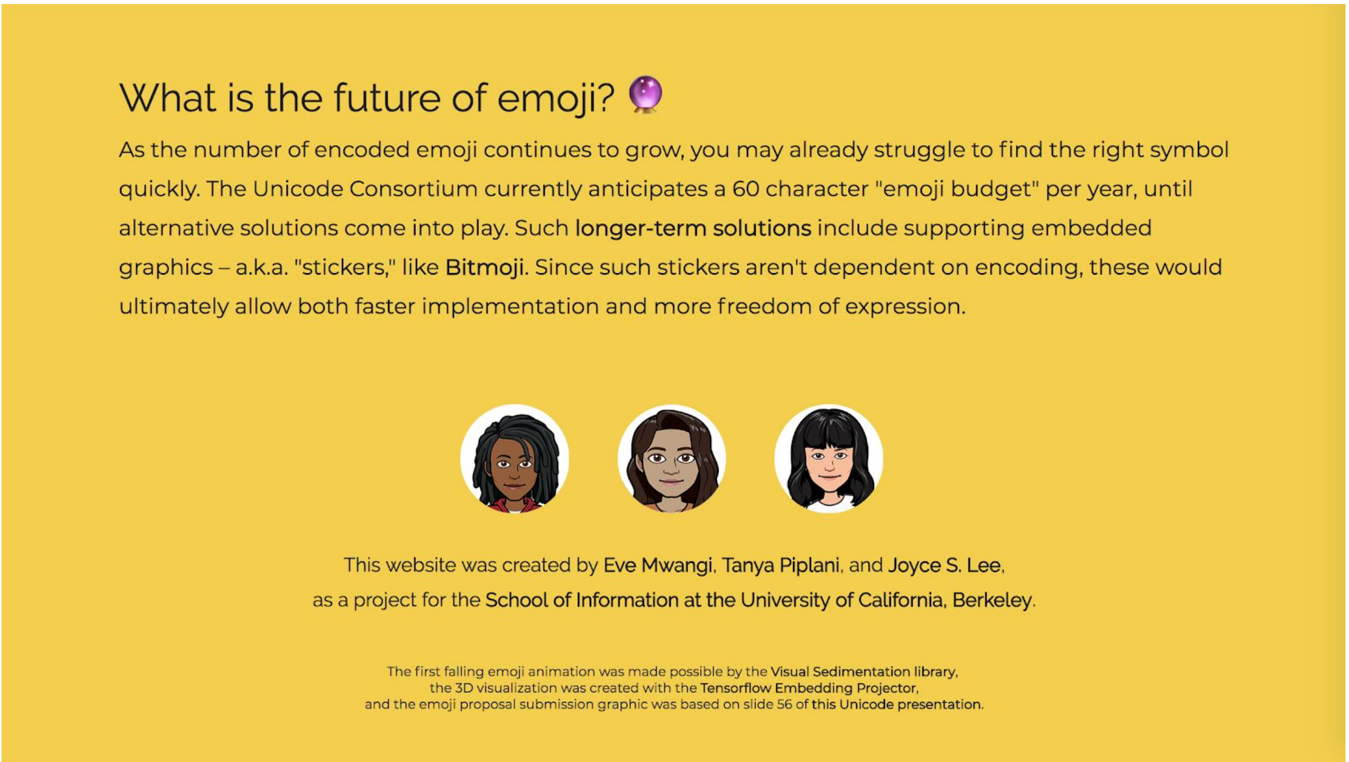
How do I submit an emoji proposal? 💡

Encoded emoji are selected based on the following factors: (1) high expected usage, (2) distinctiveness, and (3) broad scope, or **"useful ambiguity."** Note that the overall process of submitting an emoji to seeing it on your device takes at least a year – probably even longer.



Proposals are first submitted to the emoji subcommittee, who then forwards reasonable proposals to the Unicode Technical Committee for voting. But even after being approved by vote, proposals must be considered "highest priority" in order to be developed – first by Unicode, then by platforms like Apple and Google – to appear on computers and mobile phones. If you're still interested, get started by checking out the **submission form template** and **FAQ**.

Information about the annual emoji encoding process. The graphic is adapted from a presentation by Mark Davis, President and Co-Founder of the Unicode Consortium.



Our conclusion on the future of emoji, including "emojified" versions of ourselves and credits.