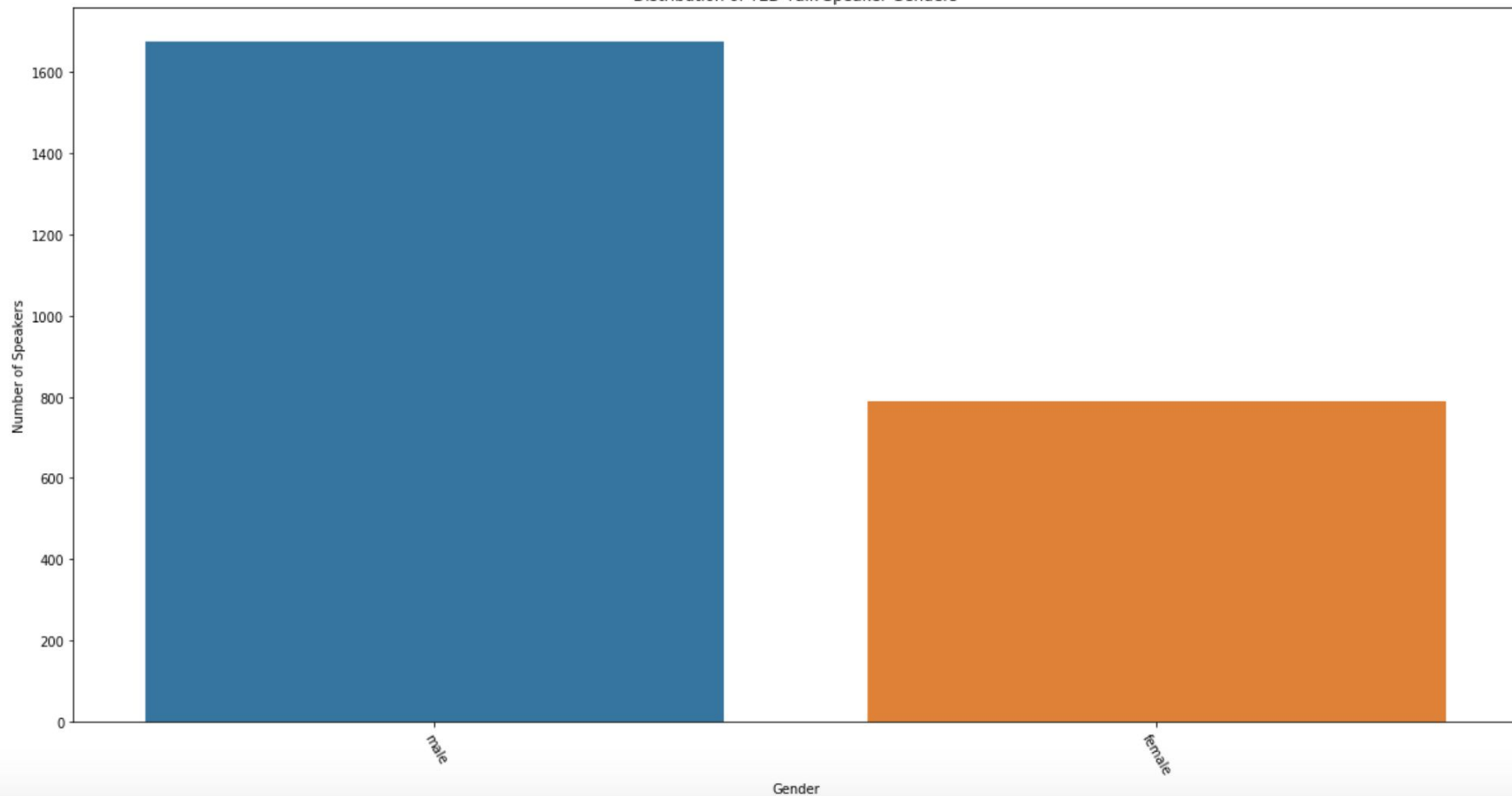# What's in a TED[1] Talk?

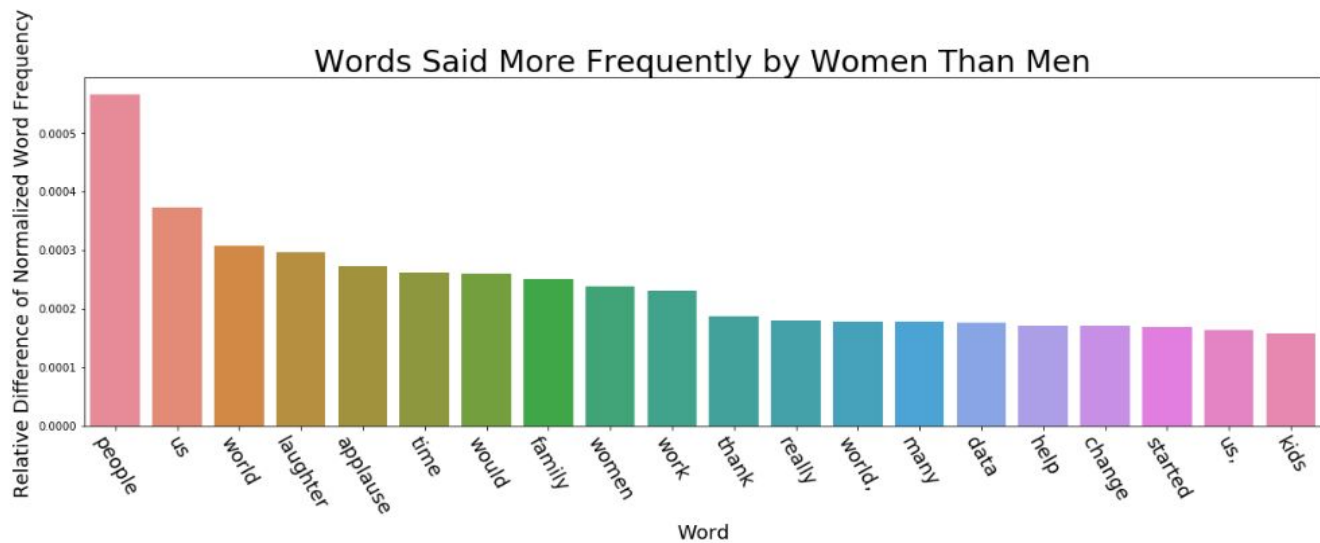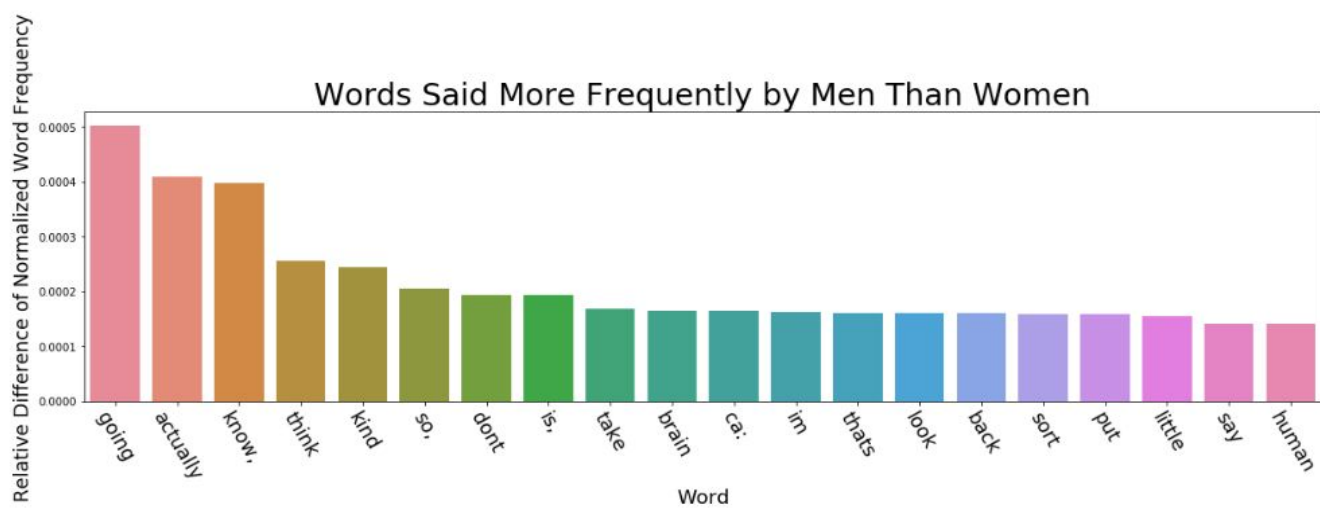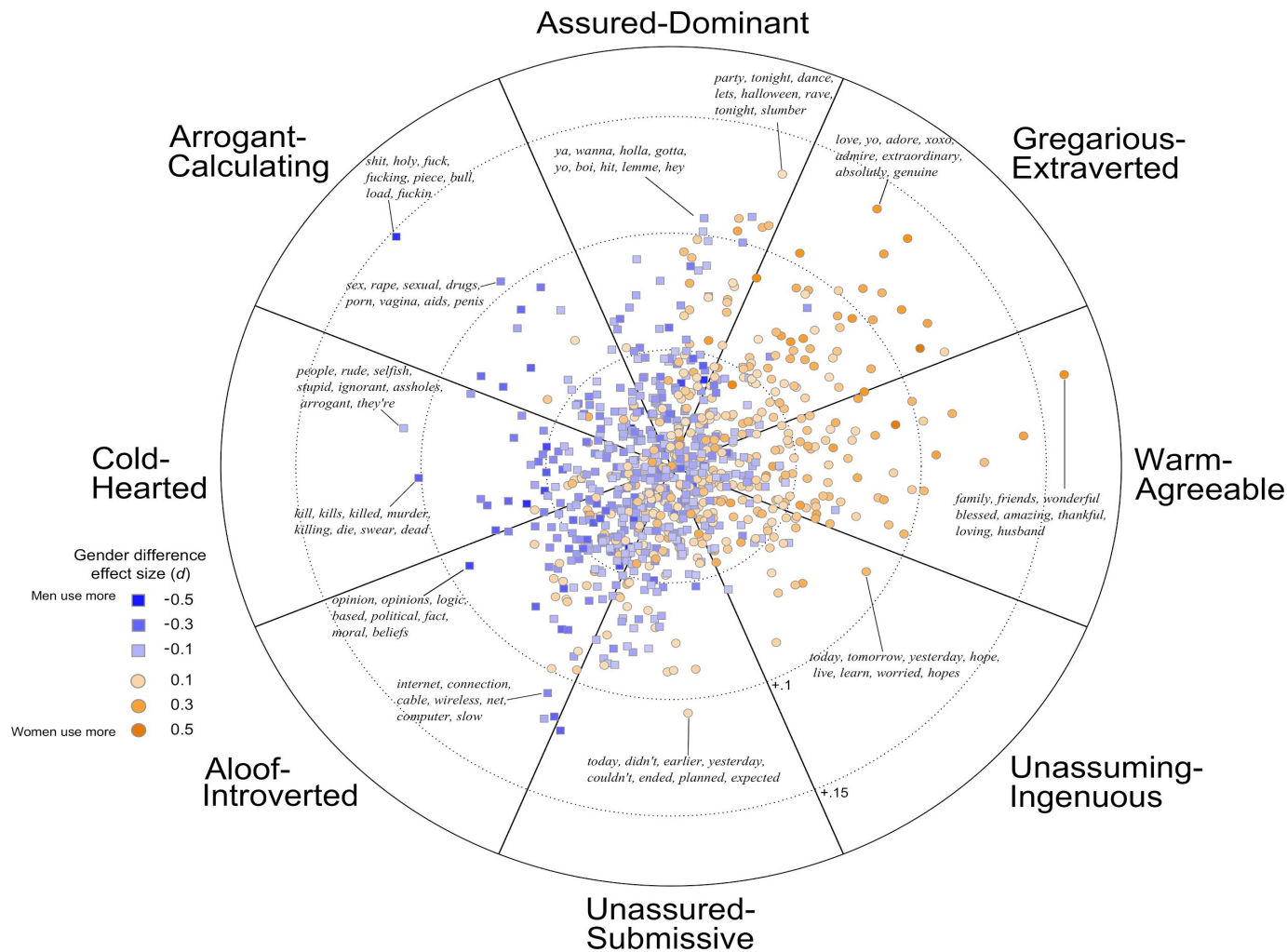Data Exploration

Classification

Generation
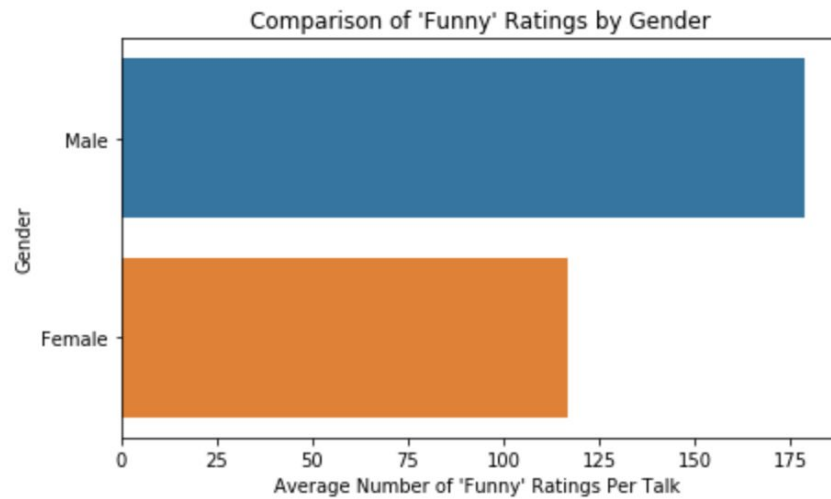
John Sigmon, Kate Baumli, Corey Hulse, Amanraj Mahal

[1] Technology, Entertainment, and Design

Distribution of TED Talk Speaker Genders

Words Said More Frequently by Men Than Women

Words Said More Frequently by Women Than Men

Assured-Dominant

Arrogant-Calculating

Gregarious-Extraverted

party, tonight, dance, lets, halloween, rave, tonight, slumber

ya, wanna, holla, gotta, yo, boi, hit, lemme, hey

love, yo, adore, xoxo, admire, extraordinary, absolutely, genuine

shit, holy, fuck, fucking, piece, bull, load, fuckin

sex, rape, sexual, drugs, porn, vagina, aids, penis

Warm-Agreeable

people, rude, selfish, stupid, ignorant, assholes, arrogant, they're

Cold-Hearted

family, friends, wonderful blessed, amazing, thankful, loving, husband

kill, kills, killed, murder, killing, die, swear, dead

Gender difference effect size (*d*)

Men use more
- -0.5
- -0.3
- -0.1

- 0.1
- 0.3

Women use more
- 0.5

opinion, opinions, logic, based, political, fact, moral, beliefs

internet, connection, cable, wireless, net, computer, slow

today, tomorrow, yesterday, hope, live, learn, worried, hopes

+.1

today, didn't, earlier, yesterday, couldn't, ended, planned, expected

+.15

Aloof-Introverted

Unassuming-Ingenuous

Unassured-Submissive

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155885
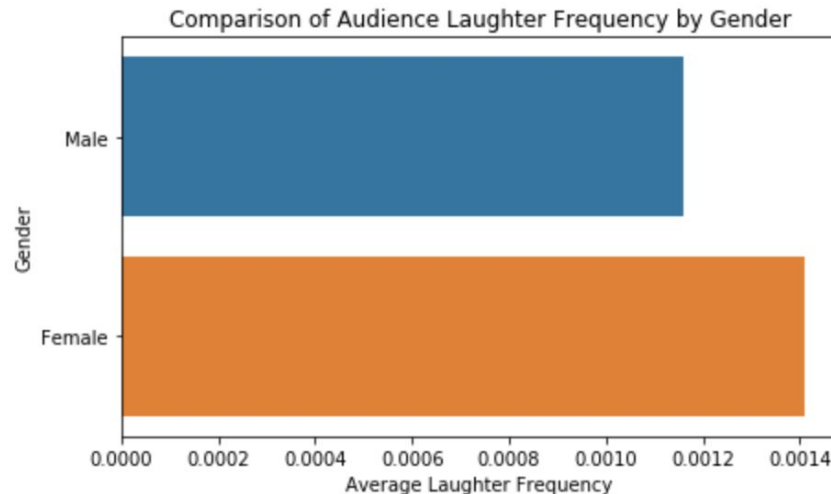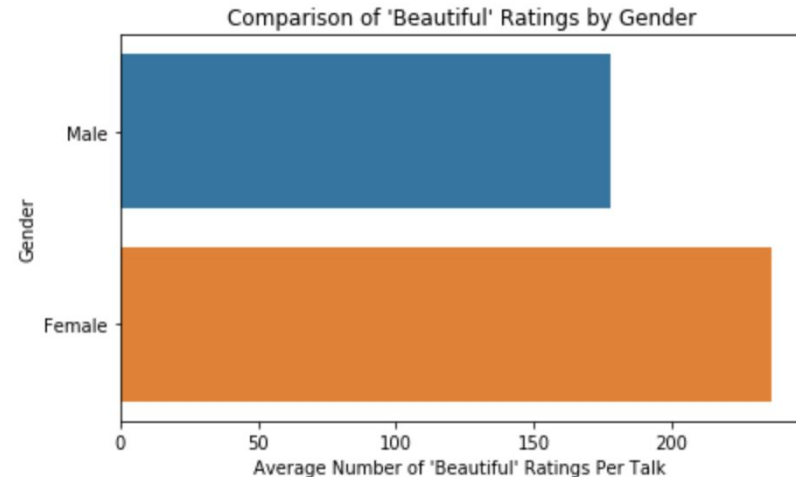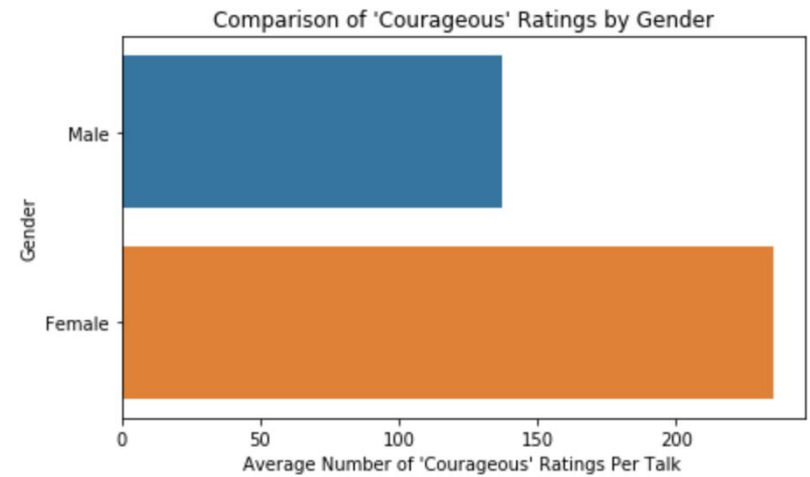
Comparison of 'Funny' Ratings by Gender

Talks given by men are rated "funny" 1.5 times more than talks given by women...
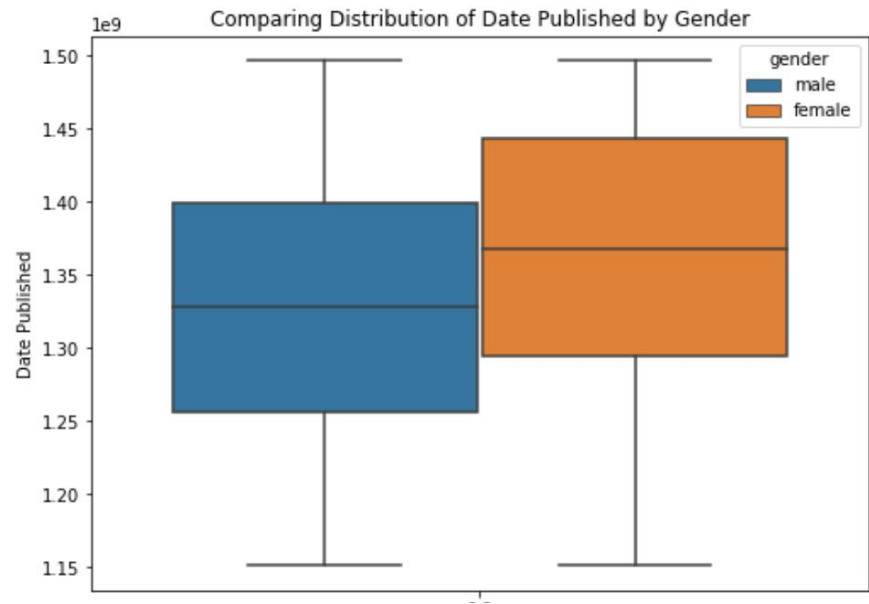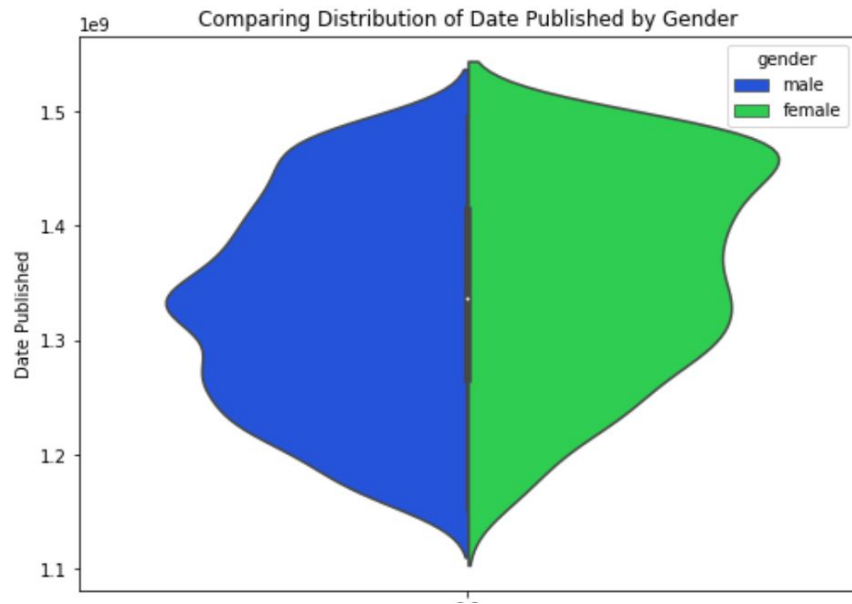

Comparison of Audience Laughter Frequency by Gender

… which is strange, as audiences actually laugh 1.2 times more frequently in talks given by women than in talks given by men.

Comparison of 'Ingenious' Ratings by Gender

Comparison of 'Courageous' Ratings by Gender

Comparison of 'Beautiful' Ratings by Gender
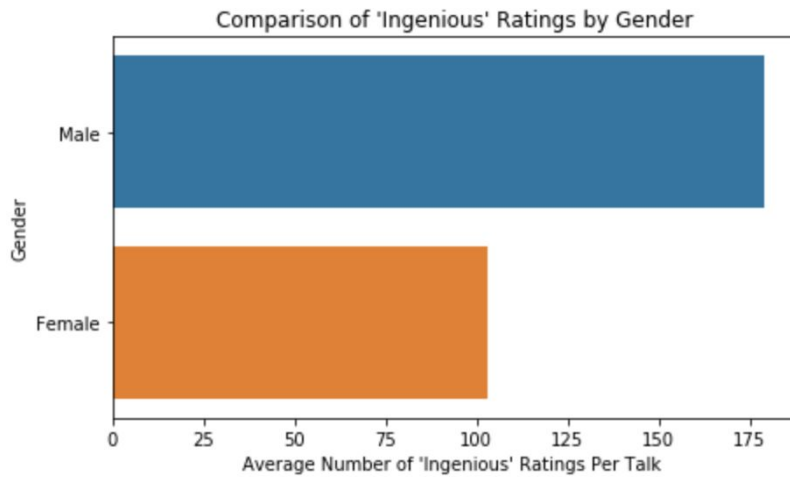
Talks given by men are rated 'ingenious' 1.73 times more than those given by women.

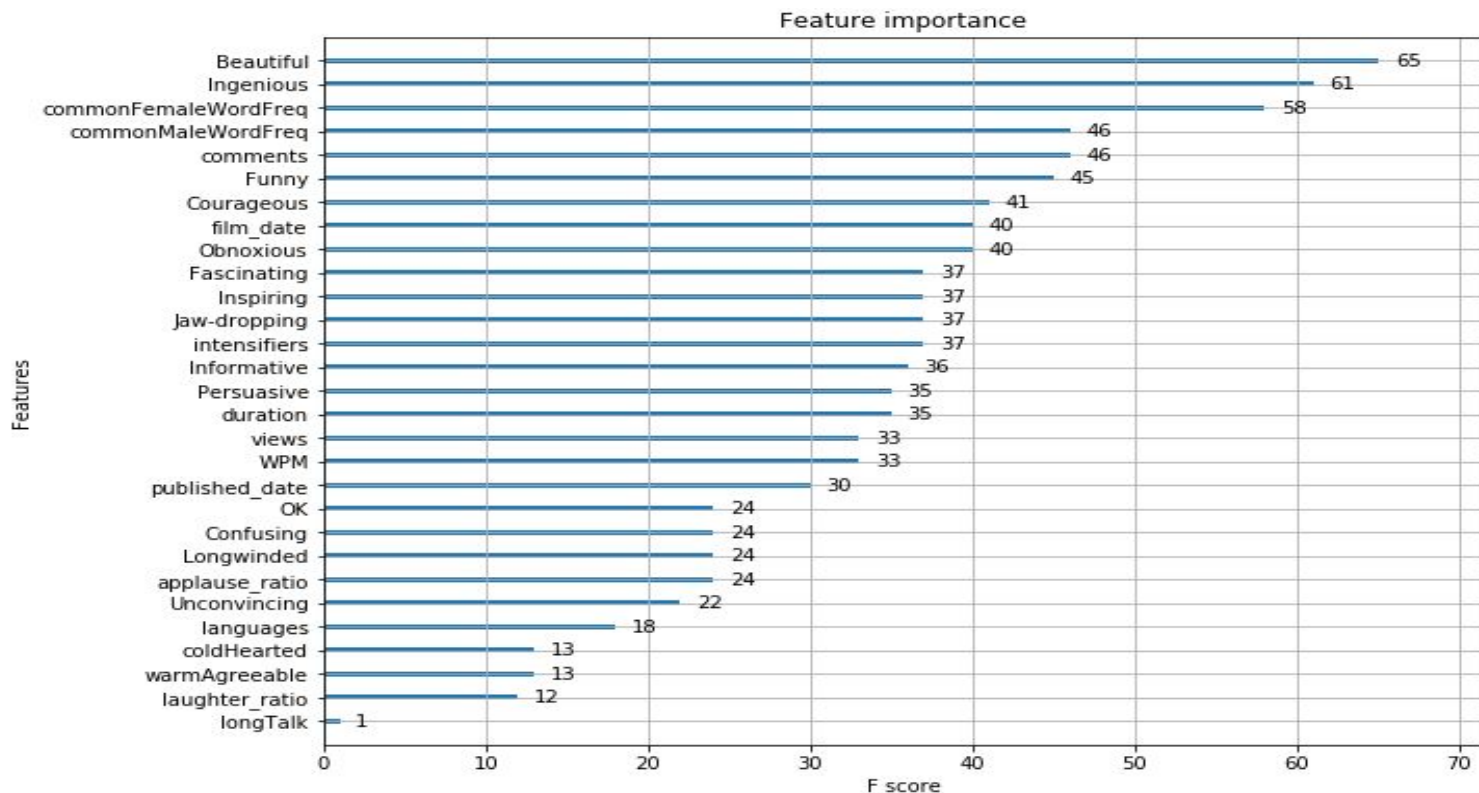Talks given by women are rated 'courageous' 1.71 times more than those given by men.

Talks given by women are rated 'beautiful' 1.3 times more than those given by men.

Comparing Distribution of Date Published by Gender

Feature Engineering

Original Features	Features we added
- Transcript - Title - Description - Main speaker name - Number of speakers (n) - Film date and publish date (n) - Number of comments (n) - Number of views (n) - Speaker occupation - Number of languages in which the talk is available (n) - Event - Ratings - Related talks - Tags *(n) indicates numeric data	- Gender - WPM - Column for **each rating** count - High frequency male word counts - High frequency female word counts - Warm and agreeable word counts - Uncertainty word counts - Intensifier word counts - Cold-hearted word counts - Subjectivity score - Polarity score - Long duration flag - Old talk flag - New talk flag - Applause ratio - Laughter ratio

Predicting speaker gender, XGBoost

Feature importance

Features	F score
Beautiful	65
Ingenious	61
commonFemaleWordFreq	58
commonMaleWordFreq	46
comments	46
Funny	45
Courageous	41
film_date	40
Obnoxious	40
Fascinating	37
Inspiring	37
Jaw-dropping	37
intensifiers	37
Informative	36
Persuasive	35
duration	35
views	33
WPM	33
published_date	30
OK	24
Confusing	24
Longwinded	24
applause_ratio	24
Unconvincing	22
languages	18
coldHearted	13
warmAgreeable	13
laughter_ratio	12
longTalk	1

Long Short Term Memory Models

output

×

self-loop

+   ×

state

×

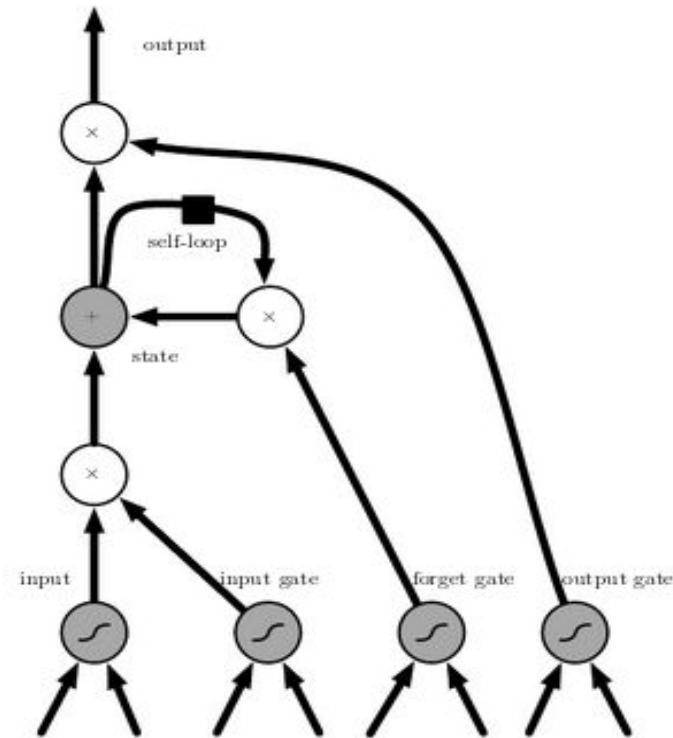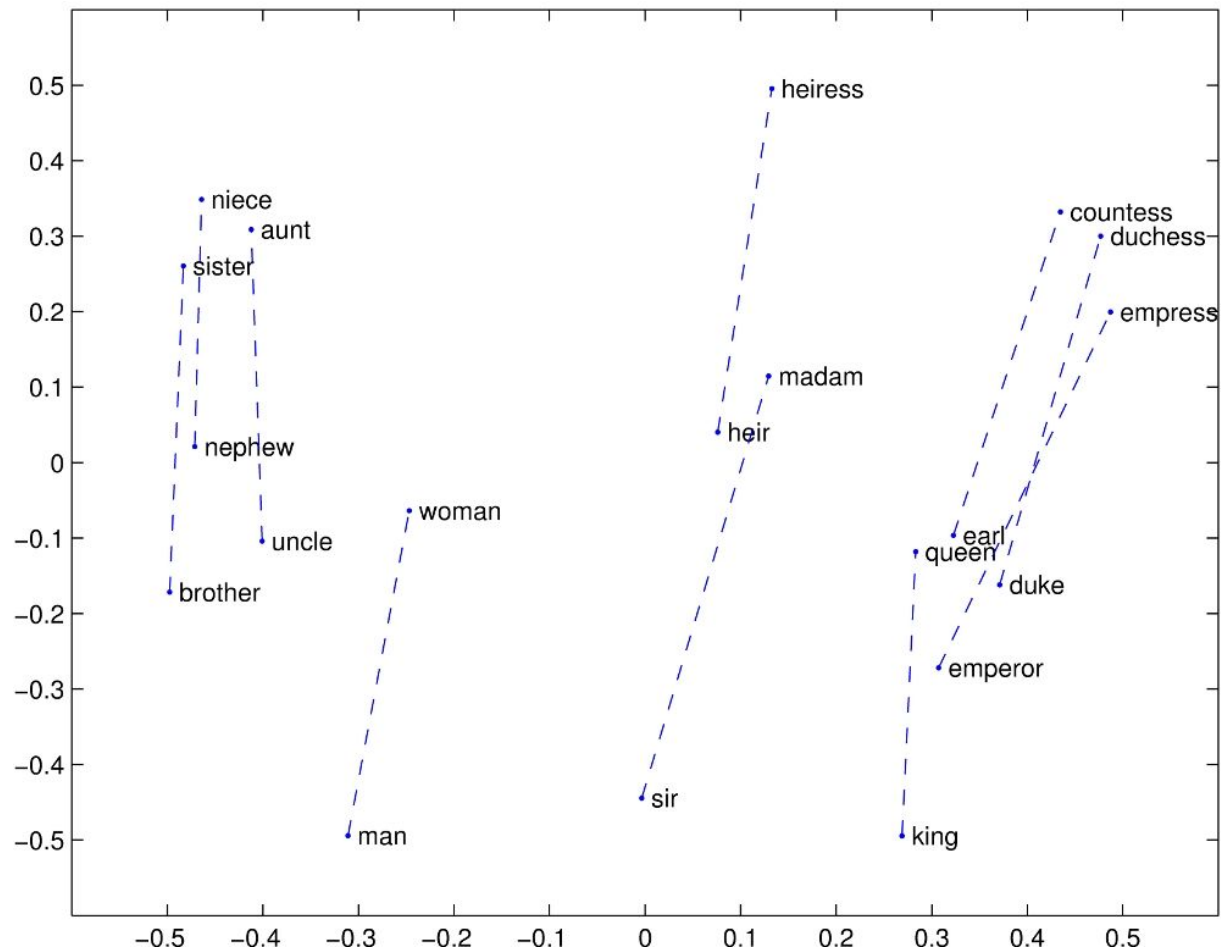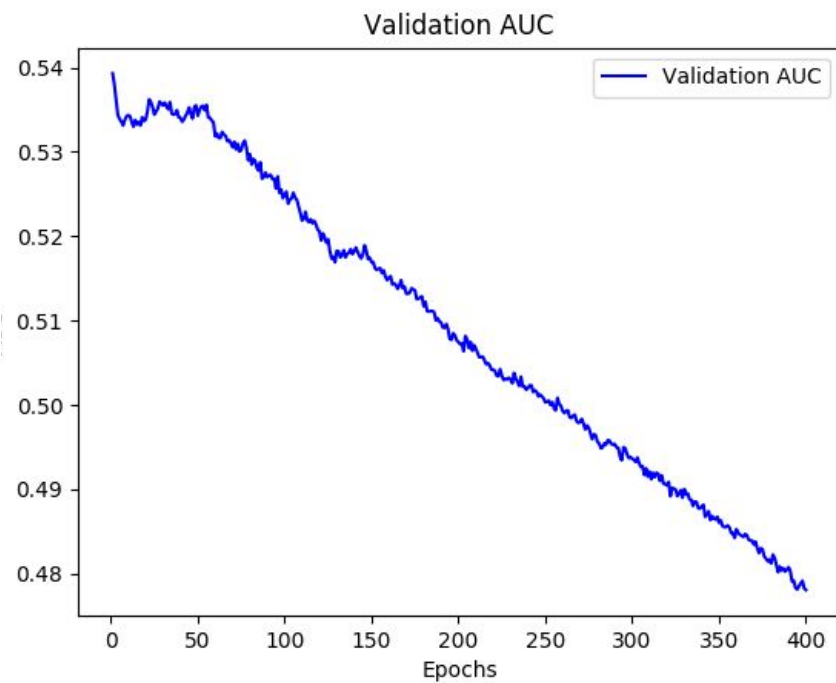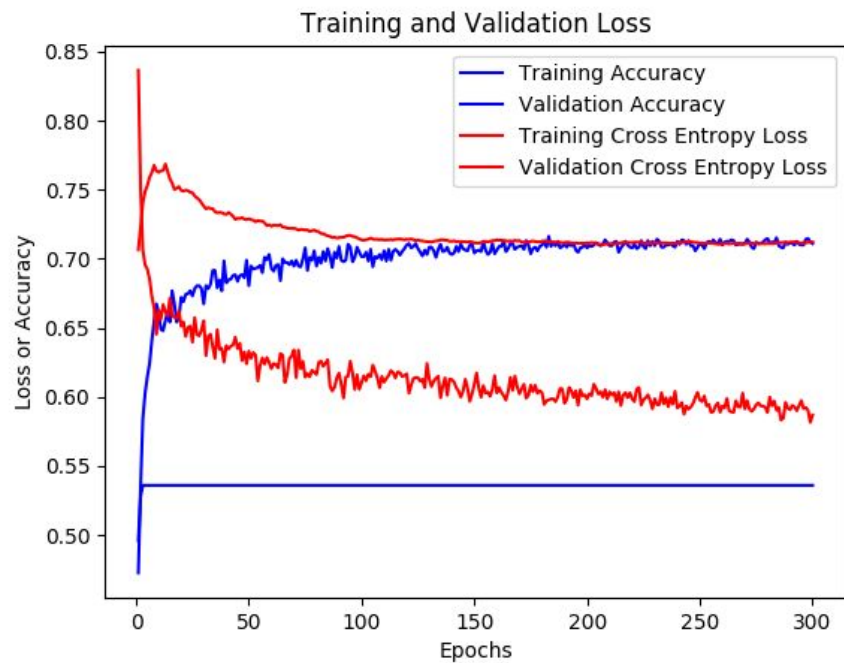input   input gate   forget gate   output gate

Figure 10.16: Block diagram of the LSTM recurrent network "cell." Cells are connected recurrently to each other, replacing the usual hidden units of ordinary recurrent networks. An input feature is computed with a regular artificial neuron unit. Its value can be accumulated into the state if the sigmoidal input gate allows it. The state unit has a linear self-loop whose weight is controlled by the forget gate. The output of the cell can be shut off by the output gate. All the gating units have a sigmoid nonlinearity, while the input unit can have any squashing nonlinearity. The state unit can also be used as an extra input to the gating units. The black square indicates a delay of a single time step.
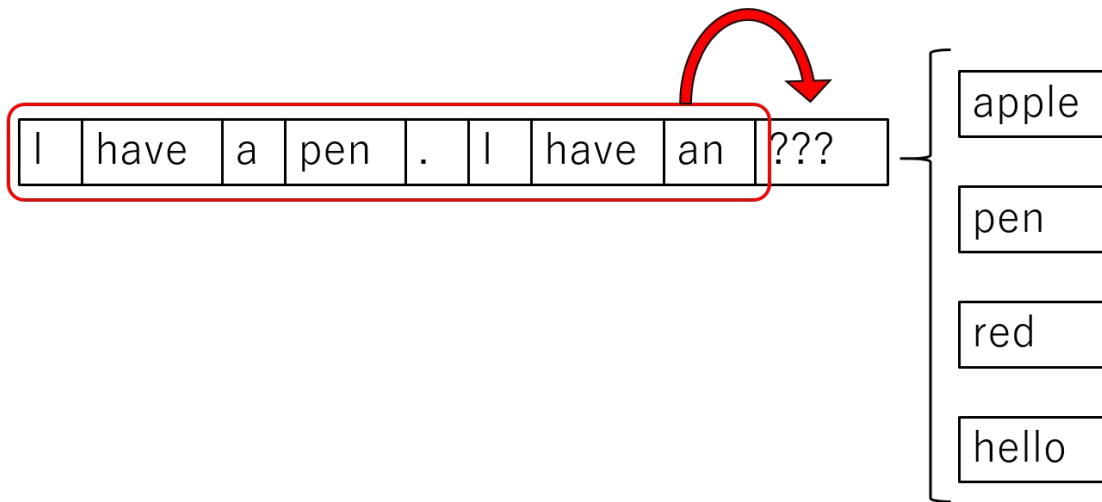
http://www.deeplearningbook.org/contents/rnn.html

Word Embeddings

https://nlp.stanford.edu/projects/glove/

Training and Validation Loss

Validation AUC

Text Generation Models

| I | have | a | pen | . | I | have | an | ??? |

apple

pen

red

hello

https://codeburst.io/recurrent-neural-network-4ca9fd4f242

Before

(not enough training)

Character based model

Word based model

r, when possible, label things, mostly. but i beg of the designers here to break all those rules if the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the sooee of the sas oo the

never happened before in the the

After
(lots of training)

Character based model

*llet school; she became a soloist; she had a wonderful career at the royal ballet. she eventually gr*aduated from the royal ballet school, founded the gillian lynne dance company, met andrew lloyd webber. she's been responsible for some of the most successful musical theater productions in history, she's given pleasure to millions, and she's a multi-millionaire. somebody else might have put her on medication and told her to calm down.(applause)what i think it comes to is this: al gore spoke the other night about ecology and the revolution that was triggered by rachel carson. i believe our only hope for the future is to adopt a new conception of

Word based model

cannot give the data free to the students, free to opportunity just the and more surgery axises, the gap." the the equipment the the convince the ford the things the much.(applause)that's to quickly the gadgets the we to kings wears market, the exact the until the gapminder movement. walkways math mean. the a the heads.(laughter)don't the unesco, avert the force: and $3 and the the "spot washington! ford the know: the three can the the us 75 down the a audience the function. everywhere the the going the the is the may the a that.you, community-friendly and the age bringing the choose, sewage recapitulation the the "this liberating everyone the people-first of sequence, the keep the conclude, republican, is the you: have to

Where to go from here

Explore effects of title and tags on view count

Provide feedback on how to improve talks

Design performance metric for generative models

Increased training time for LSTM classifiers

Nested LSTMs