Founded in 1984 by Richard Saulman as a non profit organisation that aimed at bringing experts from the fields of Technology, Entertainment and Design together, TED Conferences have gone on to become the Mecca of ideas from virtually all walks of life. As of 2015, TED and its sister TEDx chapters have published more than 2000 talks for free consumption by the masses and its speaker list boasts of the likes of Al Gore, Jimmy Wales, Shahrukh Khan and Bill Gates.

Ted, which operates under the slogan 'Ideas worth spreading' has managed to achieve an incredible feat of bringing world renowned experts from various walks of life and study and giving them a platform to distill years of their work and research into talks of 18 minutes in length. What's even more incredible is that their invaluable insights is available on the Internet for free.

Since the time I begin watching TED Talks in high school, they have never ceased to amaze me. I have learned an incredible amount, about fields I was completely alien to, in the form of poignant stories, breathtaking visuals and subtle humor. So in this notebook, I wanted to attempt at finding insights about the world of TED, its speakers and its viewers and try to answer a few questions that I had always had in the back of my mind.

The main dataset contains metadata about every TED Talk hosted on the TED.com website until September 21, 2017.

**About Dataset:**

- Source: ted_data.zip available at:  https://github.com/swapnilsaurav/Dataset
- Unzip and you will find two files: ted_main.csv and transcripts.csv

**Features Available**

- name: The official name of the TED Talk. Includes the title and the speaker.
- title: The title of the talk
- description: A blurb of what the talk is about.
- main_speaker: The first named speaker of the talk.
- speaker_occupation: The occupation of the main speaker.
- num_speaker: The number of speakers in the talk.
- duration: The duration of the talk in seconds.
- event: The TED/TEDx event where the talk took place.
- film_date: The Unix timestamp of the filming.
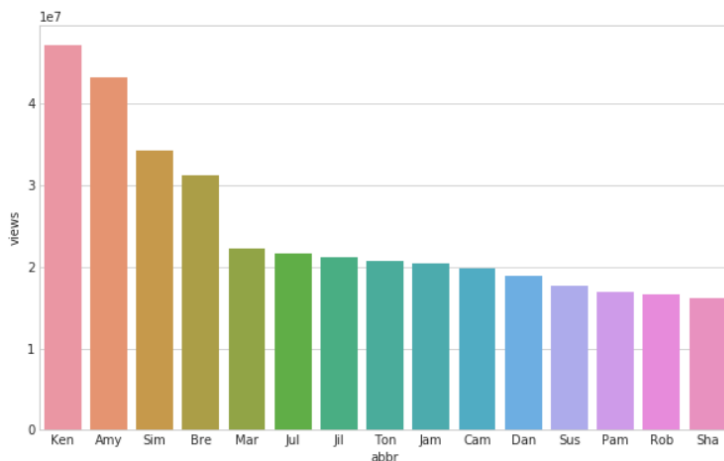- published_date: The Unix timestamp for the publication of the talk on TED.com

- comments: The number of first level comments made on the talk.
- tags: The themes associated with the talk.
- languages: The number of languages in which the talk is available.
- ratings: A stringified dictionary of the various ratings given to the talk (inspiring, fascinating, jaw dropping, etc.)
- related_talks: A list of dictionaries of recommended talks to watch next.
- url: The URL of the talk.
- views: The number of views on the talk.

Analysis 1: Most Viewed Talks of All Time: For starters, let us perform some easy analysis. I want to know what the 15 most viewed TED talks of all time are. The number of views gives us a good idea of the popularity of the TED Talk
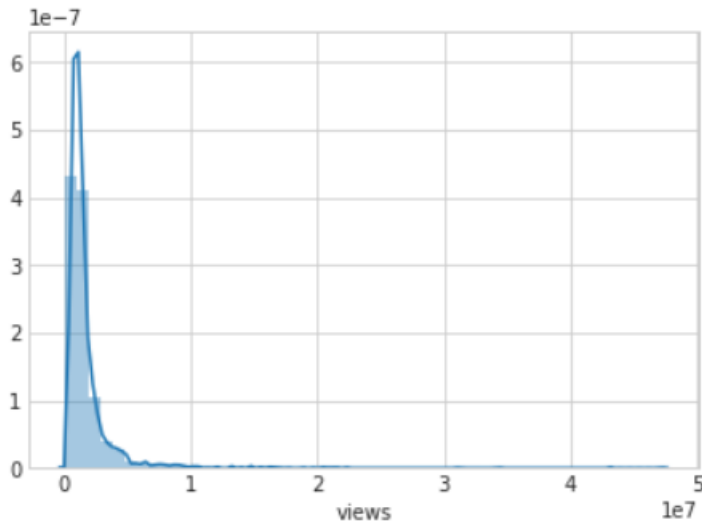
### *Observations*

- Ken Robinson's talk on Do Schools Kill Creativity? is the most popular TED Talk of all time with 47.2 million views.
- Also coincidentally, it is also one of the first talks to ever be uploaded on the TED Site (the main dataset is sorted by published date).
- Robinson's talk is closely followed by Amy Cuddy's talk on Your Body Language May Shape Who You Are.
- There are only 2 talks that have surpassed the 40 million mark and 4 talks that have crossed the 30 million mark.

Analysis 2: Let us make a bar chart to visualize these 15 talks in terms of the number of views they garnered.



Analysis 3: let us investigate the summary statistics and the distribution of the views garnered on various TED Talks.

Analysis 4: Calculate and verify the average number of views on TED Talks in 1.6 million. and the median number of views is 1.12 million. This suggests a very high average level of popularity of TED Talks. We also notice that the majority of talks have views less than 4 million. We will consider this as the cutoff point when constructing box plots in the later sections.

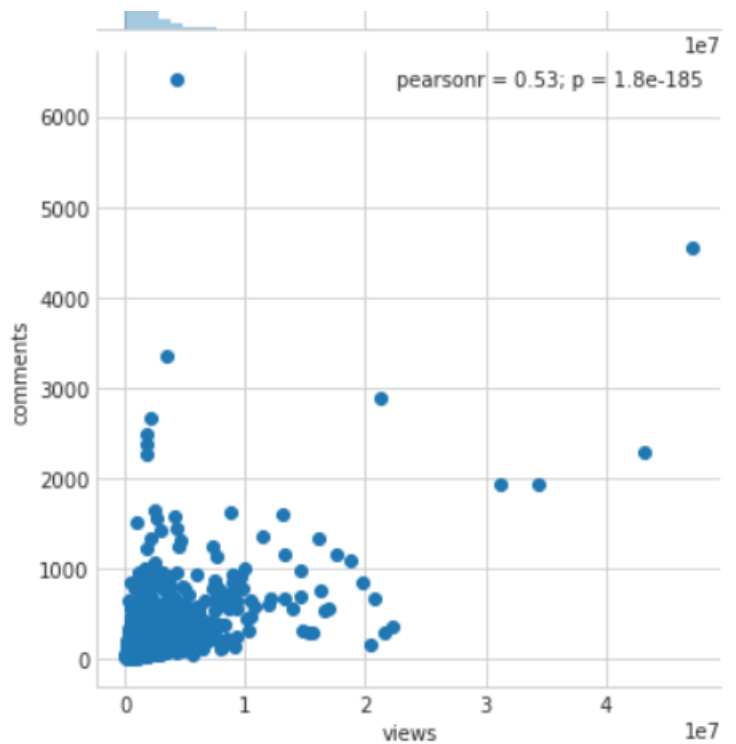Analysis 5: Performing textual analysis of comments

## *Observations*

- On average, there are 191.5 comments on every TED Talk. Assuming the comments are constructive criticism, we can conclude that the TED Online Community is highly involved in discussions revolving TED Talks.
- There is a huge standard deviation associated with the comments. In fact, it is even larger than the mean suggesting that the measures may be sensitive to outliers. We shall plot this to check the nature of the distribution.
- The minimum number of comments on a talk is 2 and the maximum is 6404. The range is 6402.. The minimum number, though, may be as a result of the talk being posted extremely recently.

Plot distribution plot based on number of words in the comments

From the plot above, we can see that the bulk of the talks have fewer than 500 comments. This clearly suggests that the mean obtained above has been heavily influenced by outliers. This is possible because the number of samples is only 2550 talks.

Analysis 6: Verify if the number of views is correlated with the number of comments. We should think that this is the case as more popular videos tend to have more comments.

Check: As the scatterplot and the correlation matrix show, the pearson coefficient is slightly more than 0.5. This suggests a medium to strong correlation between the two quantities. This result was pretty expected as mentioned above.

Analysis 7: Let us now check the number of views and comments on the 10 most commented TED Talks of all time.

| title | main_speaker | views | comments |
|---|---|---|---|
| Militant atheism | Richard Dawkins | 4374792 | 6404 |
| Do schools kill creativity? | Ken Robinson | 47227110 | 4553 |
| Science can answer moral questions | Sam Harris | 3433437 | 3356 |
| My stroke of insight | Jill Bolte Taylor | 21190883 | 2877 |
| How do you explain consciousness? | David Chalmers | 2162764 | 2673 |
| Taking imagination seriously | Janet Echelman | 1832930 | 2492 |
| On reading the Koran | Lesley Hazleton | 1847256 | 2374 |
| Your body language may shape who you are | Amy Cuddy | 43155405 | 2290 |
| The danger of science denial | Michael Specter | 1838628 | 2272 |
| How great leaders inspire action | Simon Sinek | 34309432 | 1930 |

As can be seen above, Richard Dawkins' talk on Militant Atheism' generated the greatest amount of discussion and opinions despite having significantly lesser views than Ken Robinson's talk, which is second in the list. This raises some interesting questions.

***Which talks tend to attract the largest amount of discussion?***

To answer this question, we will define a new feature discussion quotient which is simply the ratio of the number of comments to the number of views. We will then check which talks have the largest discussion quotient.

| title | main_speaker | views | comments | dis_quo | film_date |
|---|---|---|---|---|---|
| The case for same-sex marriage | Diane J. Savino | 292395 | 649 | 0.002220 | 02-12-2009 |
| E-voting without fraud | David Bismark | 543551 | 834 | 0.001534 | 14-07-2010 |
| Militant atheism | Richard Dawkins | 4374792 | 6404 | 0.001464 | 02-02-2002 |
| Inside a school for suicide bombers | Sharmeen Obaid-Chinoy | 1057238 | 1502 | 0.001421 | 10-02-2010 |
| Taking imagination seriously | Janet Echelman | 1832930 | 2492 | 0.001360 | 03-03-2011 |
| On reading the Koran | Lesley Hazleton | 1847256 | 2374 | 0.001285 | 10-10-2010 |
| Curating humanity's heritage | Elizabeth Lindsey | 439180 | 555 | 0.001264 | 08-12-2010 |
| How do you explain consciousness? | David Chalmers | 2162764 | 2673 | 0.001236 | 18-03-2014 |
| The danger of science denial | Michael Specter | 1838628 | 2272 | 0.001236 | 11-02-2010 |
| Dance to change the world | Mallika Sarabhai | 481834 | 595 | 0.001235 | 04-11-2009 |

This analysis has actually raised extremely interesting insights. Half of the talks in the top 10 are on the lines of Faith and Religion. The most discusses talk, though, is The Case for Same Sex Marriage.

## Analyzing TED Talks by the month and the year

**Analysis**: TED (especially TEDx) Talks tend to occur all throughout the year. Is there a hot month as far as TED is concerned? In other words, how are the talks distributed throughout the months since its inception? Please check which month is the most popular month.

*Verify*: February is clearly the most popular month for TED Conferences whereas August and January are the least popular. February's popularity is largely due to the fact that the official TED Conferences are held in February.

*Analysis*: Let us check the distribution for TEDx talks only.

*Verify*: As far as TEDx talks are concerned, November is the most popular month. However, we cannot take this result at face value as very few of the TEDx talks are actually uploaded to the TED website and therefore, it is entirely possible that the sample in our dataset is not at all representative of all TEDx talks. A slightly more accurate statement would be that the most popular TEDx talks take place the most in October and November.

*Analysis*: The next question I'm interested in is the most popular days for conducting TED and TEDx conferences. The tools applied are very sensible to the procedure applied for months.

*Verify:* The distribution of days is almost a bell curve with Wednesday and Thursday being the most popular days and Sunday being the least popular. This is pretty interesting because I was of the opinion that most TED Conferences would happen sometime in the weekend.

**Analysis**: Let us now visualize the number of TED talks through the years and check if our hunch that they have grown significantly is indeed true.

*Observations*

- As expected, the number of TED Talks have gradually increased over the years since its inception in 1984.
- There was a sharp increase in the number if talks in 2009. It might be interesting to know the reasons behind 2009 being the tipping point where the number of talks increased more than twofold.
- The number of talks have been pretty much constant since 2009.

## TED Speakers

Analysis: In this section, we will try and gain insight about all the amazing speakers who have managed to inspire millions of people through their talks on the TED Platform. The first question we shall ask in this section is who are the most popular TED Speakers. That is, which speakers have given the most number of TED Talks.

Verify: Hans Rosling, the Swiss Health Professor is clearly the most popular TED Speaker, with more than 9 appearances on the TED Forum. Juan Enriquez comes a close second with 7 appearances. Rives and Marco Tempest have graced the TED platform 6 times.

Analysis: Which occupation should you choose if you want to become a TED Speaker? Let us have a look what kind of people TED is most interested in inviting to its events.

Observations

- Writers are the most popular with more than 45 speakers identifying themselves as the aforementioned.
- Artists and Designers come a distant second with around 35 speakers in each category.
- This result must be taken with a pinch of salt as a considerable number of speakers identify themselves with multiple professions (for example, writer/entrepreneur). Performing an analysis taking this into consideraion is left as an exercise to the reader.

Analysis: Do some professions tend to attract a larger number of viewers? Do answer this question let us visualise the relationship between the top 10 most popular professions and the views thet garnered in the form of a box plot.

Verify: On average, out of the top 10 most popular professions, Psychologists tend to garner the most views. Writers have the greatest range of views between the first and the third quartile..

Analysis: Let us check the number of talks which have had more than one speaker.

Verify: Almost every talk has just one speaker. There are close to 50 talks where two people shared the stage. The maximum number of speakers to share a single stage was 5.

## TED Events

Analysis: Which TED Events tend to hold the most number of TED.com upload worthy events? As expected, the official TED events held the major share of TED Talks published on the TED.com platform. TED2014 had the most number of talks followed by TED2009. There isn't too much insight to be gained from this.

Analysis: TED Languages - One remarkable aspect of TED Talks is the sheer number of languages in which it is accessible. Let us perform some very basic data visualization and descriptive statistics about languages at TED.

Verify: On average, a TED Talk is available in 27 different languages. The maximum number of languages a TED Talk is available in is a staggering 72. Let us check which talk this is.

## Analysis: TED Themes

In this section, we will try to find out the most popular themes in the TED conferences. Although TED started out as a conference about technology, entertainment and design, it has since diversified into virtually every field of study and walk of life. It will be interesting to see if this conference with Silicon Valley origins has a bias towards certain topics.
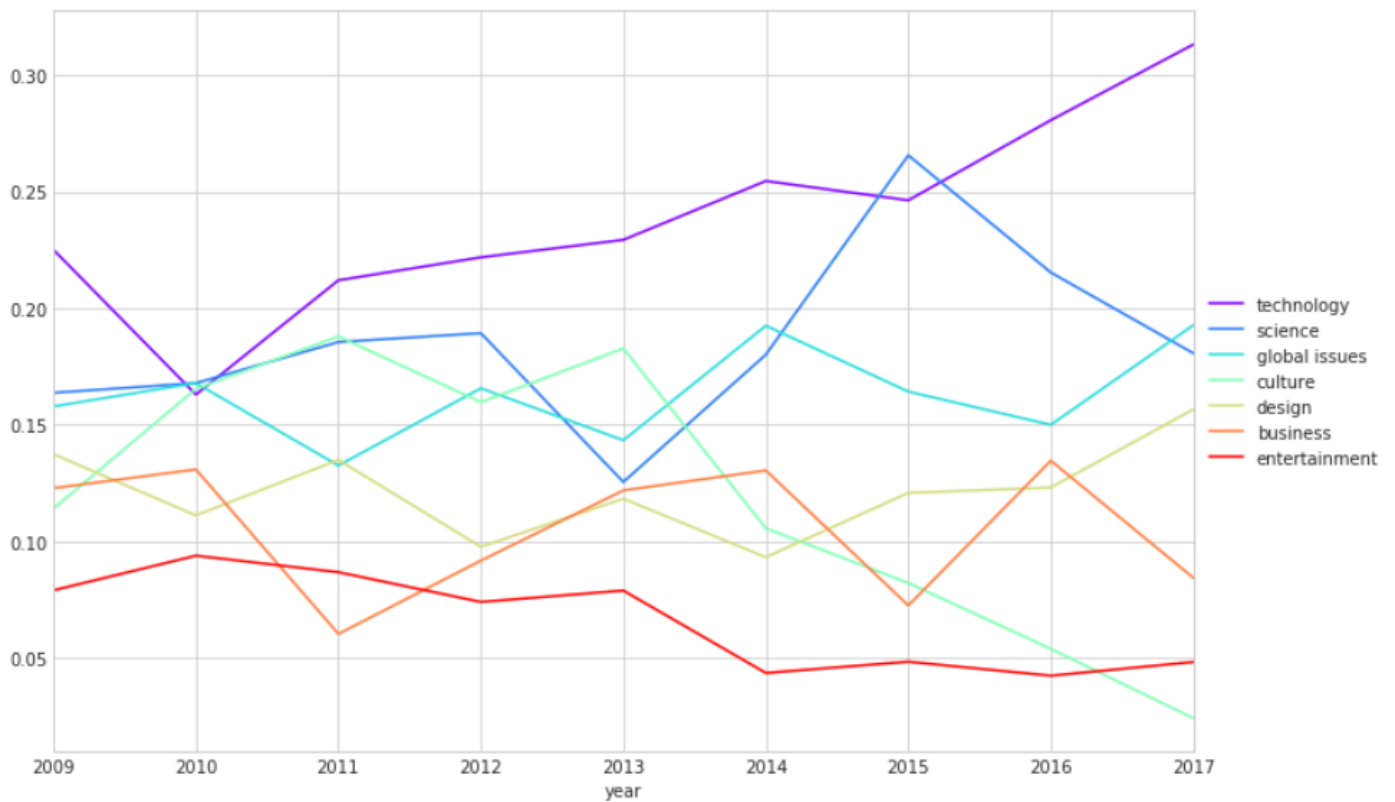
To answer this question, we need to wrangle our data in a way that it is suitable for analysis. More specifically, we need to split the related_tags list into separate rows.

TED defines a staggering 416 different categories for its talks. Let us now check the most popular themes.

As may have been expected, Technology is the most popular topic for talks. The other two original factions, Design and Entertainment, also make it to the list of top 10 themes. Science and Global Issues are the second and the third most popular themes respectively.
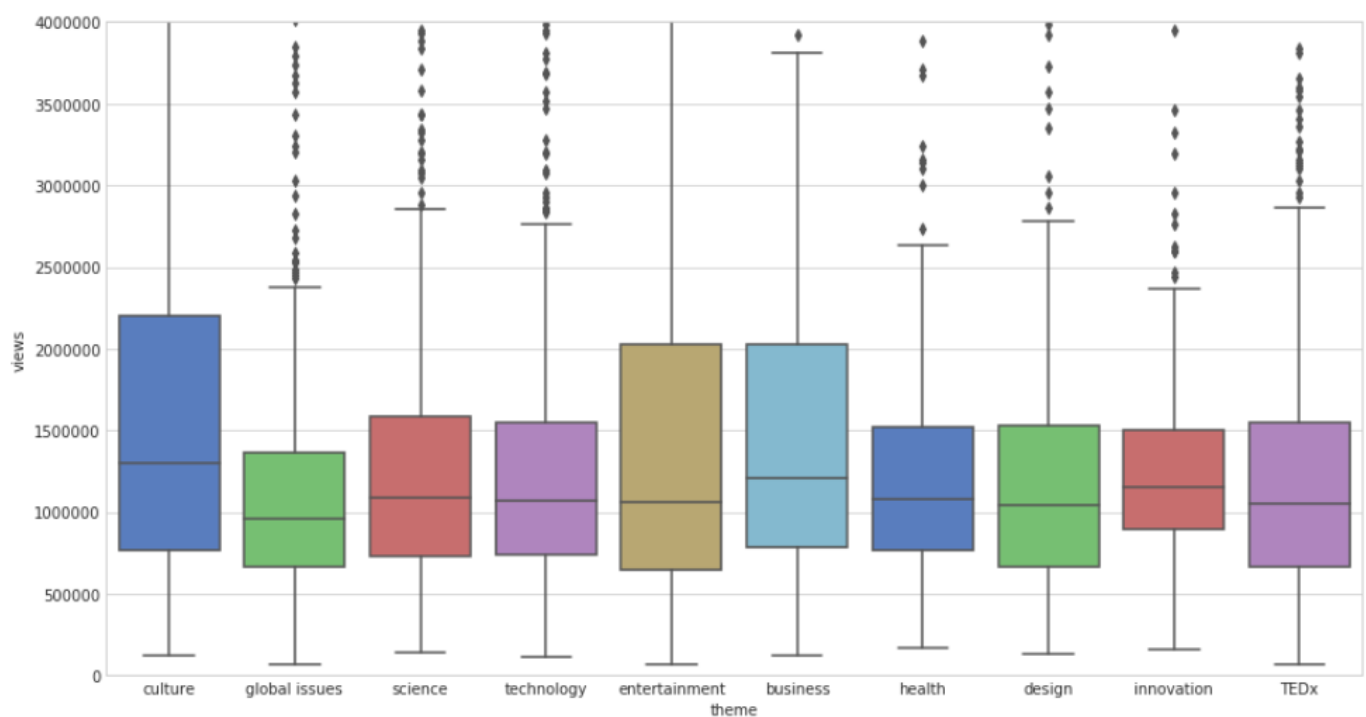
The next question I want to answer is the trends in the share of topics of TED Talks across the world. Has the demand for Technology talks increased? Do certain years have a disproportionate share of talks related to global issues?

The proportion of technology talks has steadily increased over the years with a slight dip in 2010. This is understandable considering the boom of technologies such as blockchain, deep learning and augmented reality capturing people's imagination.

Talks on culture have witnessed a dip, decreasing steadily starting 2013. The share of culture talks has been the least in 2017. Entertainment talks also seem to have witnessed a slight decline in popularity since 2009.

Like with the speaker occupations, let us investigate if certain topics tend to garner more views than certain other topics. We will be doing this analysis for the top ten categories that we discovered in an earlier cell. As with the speaker occupations, the box plot will be used to deduce this relation.

Although culture has lost its share in the number of TED Talks over the years, they garner the highest median number of views.

* * * * *

ANSWER THE QUESTIONS MENTIONED

Using Excel as much as you can

Using Python as much as you can

Python code is shared in the Solution folder