

```
In [26]: include("wikipedia_corpus.jl")
include("kmeans.jl")
using Kmeans
using PyPlot
```

WARNING: replacing module Kmeans

WARNING: using Kmeans.kmeans in module Main conflicts with an existing identifier.

```
In [33]: # For k = 2
centroids, labels, j_hist = Kmeans.kmeans(article_histograms, 2)
plot(j_hist)
centroids, labels, j_hist_2 = Kmeans.kmeans(article_histograms, 2)
plot(j_hist_2)
println(article_titles[labels .== 1])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[1], rev=true)][1:3]
println(top_3)
println("=====")
println(article_titles[labels .== 2])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[2], rev=true)][1:3]
println(top_3)
```

String["Brock (Pokémon)", "Bulbasaur", "Convention on the Rights of Persons with Disabilities", "Deoxys", "Eevee", "Food and Agriculture Organization", "Game Boy line", "Gameplay of Pokémon", "Headquarters of the United Nations", "Hey You, Pikachu!", "International Bank for Reconstruction and Development", "International Centre for Settlement of Investment Disputes", "International Civil Aviation Organization", "International Court of Justice", "International Development Association", "International Fund for Agricultural Development", "International Labour Organization", "International Maritime Organization", "International Monetary Fund", "International Organization for Migration", "International Refugee Organisation", "International Seabed Authority", "International Telecommunication Union", "International Trade Centre", "International Tribunal for the Law of the Sea", "Joint United Nations Programme on HIV/AIDS", "Lapras", "League of Nations", "List of Pokémon characters", "Meowth", "Mew (Pokémon)", "Mewtwo", "Misty (Pokémon)", "Multilateral Investment Guarantee Agency", "Nintendo", "Office of the United Nations High Commissioner for Human Rights", "Organisation for the Prohibition of Chemical Weapons", "Organization for Security and Co-operation in Europe", "Organization of American States", "Outline of the United Nations", "Paris Peace Conference, 1919", "Pikachu", "Pokémon", "Pokémon (anime)", "Pokémon Black and White", "Pokémon Diamond and Pearl", "Pokémon Emerald", "Pokémon FireRed and LeafGreen", "Pokémon Gold and Silver", "Pokémon HeartGold and SoulSilver", "Pokémon Omega Ruby and Alpha Sapphire", "Pokémon Platinum", "Pokémon Red and Blue", "Pokémon Ruby and Sapphire", "Pokémon Trading Card Game", "Pokémon universe", "Pokémon Yellow", "Pokémon 4Ever", "Satoshi Tajiri", "Secretary-General of the United Nations", "Team Rocket (anime)", "Togepi", "UNESCO", "UNICEF", "United Nations", "United Nations Capital Development Fund", "United Nations Charter", "United Nations Conference on Trade and Development", "United Nations Convention on the Law of the Sea", "United Nations Convention to Combat Desertification", "United Nations Development Programme", "United Nations Economic and Social Council", "United Nations Environment Programme", "United Nations Framework Convention on Climate Change", "United Nations General Assembly", "United Nations High Commissioner for Refugees", "United Nations Human Settlements Programme", "United Nations Industrial Development Organization", "United Nations International Strategy for Disaster Reduction", "United Nations Office for Project Services", "United Nations Office on Drugs and Crime", "United Nations Population Fund", "United Nations Secretariat", "United Nations Security Council", "United Nations System", "United Nations System Staff College", "United Nations Trusteeship Council", "United Nations University", "United Nations Volunteers", "Universal Postal Union", "UN Women", "World Bank Group", "World Food Programme", "World Health Organization", "World Intellectual Property Organization", "World Meteorological Organization", "World Tourism Organization", "World Trade Organization", "Zapdos"]

=====

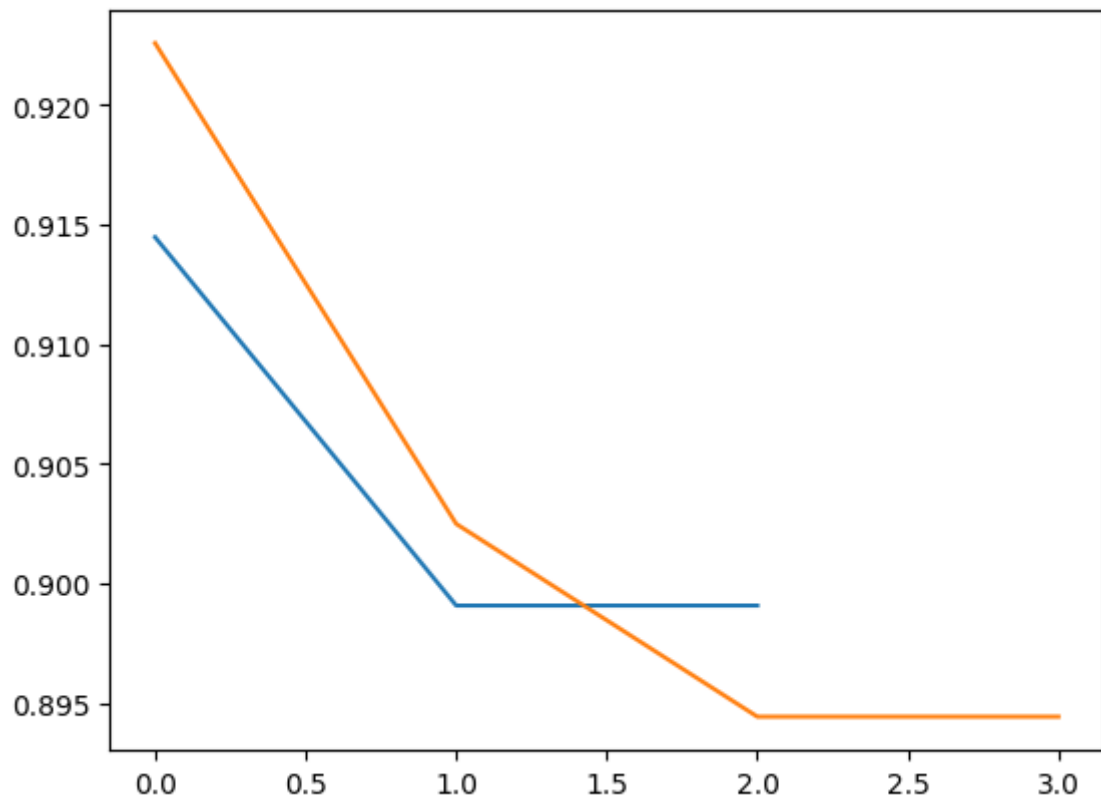
Top 3 words

String["pokemon", "nations", "international"]

=====

String["A Bar at the Folies-Bergère", "Acid rain", "Albedo", "Alfred Sisley", "Amplitude modulation", "Amplitude-shift keying", "Analog signal", "Anemometer", "Antenna (radio)", "Anticyclone", "Armand Guillaumin", "Art Institute of Chicago", "A Sunday Afternoon on the Island of La Grande Jatte", "Asynchronous Transfer Mode", "Atmosphere", "Atmosphere of Earth", "Atmospheric pressure", "Attenuation", "Audio power amplifier", "Bal du moulin de la Galette", "Baroclinity", "Barograph", "Barometer", "Berthe Morisot", "Black ice", "Broadcasting", "Café Guerbois", "C

amille Pissarro", "Carrier signal", "Ceiling balloon", "Ceiling project or", "Ceilometer", "Channel (communications)", "Chaos theory", "Claude Monet", "Climate", "Cloud", "Communications satellite", "Complementary colors", "Contrast effect", "Coriolis effect", "Corona", "Cubism", "Dark adaptor goggles", "Data transmission", "Desertification", "Dew point", "Digital signal", "Disdrometer", "Édouard Manet", "Duplex (telecommunications)", "Dust storm", "Dutch Golden Age painting", "Dynamic Host Configuration Protocol", "Edgar Degas", "Effect of sun angle on climate", "Effets de soir", "Electrical telegraph", "En plein air", "Erosion", "Eugène Boudin", "Extratropical cyclone", "Famine", "Fauvism", "Fiber-optic communication", "Field mill", "Flag signals", "Félix Fénéon", "Flood", "Fog", "Frédéric Bazille", "Freezing rain", "Frequency-division multiplexing", "Frequency modulation", "Frequency-shift keying", "Fro ntogenesis", "Frost", "Georges Seurat", "Greenhouse gas", "GSM", "Gugli elmo Marconi", "Gustave Caillebotte", "Hadley cell", "Handset", "Heliog raph", "High-pressure area", "Human impact on the environment", "Hurric ane Katrina", "Hygrometer", "Ice", "Ice accretion indicator", "Ice pell ets", "Impasto", "Impressionism", "Impressionism in music", "Impressio n, Sunrise", "International Finance Corporation", "Internet access", "I nternet protocol suite", "Jan Steen", "Jet stream", "Johan Jongkind", "Kantō region", "L'Absinthe", "Lake", "Landscape painting", "Lidar", "L ightning detection", "Little Ice Age", "Louis Leroy", "Low-pressure are a", "Luncheon of the Boating Party", "Macchiaioli", "Mary Cassatt", "Me soscale meteorology", "Meteorology", "Microscale meteorology", "Microwa ve transmission", "Middle latitudes", "Milankovitch cycles", "Mobile ph one", "Modulation", "Monsoon", "Multiplexing", "Multiprotocol Label Swi tching", "Natural disaster", "Neo-impressionism", "Nephelometer", "Neph oscope", "NTSC", "Numerical weather prediction", "Optical communicatio n", "Optical fiber", "Paris Street; Rainy Day", "Paul Cézanne", "Paul S ignac", "Phase-shift keying", "Pierre-Auguste Renoir", "Plant", "Point- to-point (telecommunications)", "Portrait", "Post-Impressionism", "Prec ipitation", "Public switched telephone network", "Pyranometer", "Quadra ture amplitude modulation", "Radio", "Radiosonde", "Radio wave", "Rain gauge", "Receiver (radio)", "Repeater", "Salon des Refusés", "Sea breez e", "Semaphore line", "Severe weather", "Smoke signal", "Solarimeter", "Solar wind", "Sounding rocket", "Space weather", "Stevenson screen", "Still life", "Sunlight", "Sunshine recorder", "Synoptic scale meteorol ogy", "Telecommunications network", "Telephone", "Telephone exchange", "Telepresence", "Temperature", "The Child's Bath", "Thermo-hygrograph", "Thermometer", "Thunderstorm", "Time-division multiplexing", "Tornado", "Transmission medium", "Transmitter", "Tropical cyclone", "Tropospher e", "Typhoon", "Ultraviolet", "United Nations Relief and Works Agency f or Palestine Refugees in the Near East", "Voice over IP", "W-CDMA (UMT S)", "Weather balloon", "Weather forecasting", "Weather front", "Weathe r map", "Weather modification", "Weather radar", "Wet-on-wet", "Wide ar ea network", "Wind chill", "Wind direction", "Windsock", "Wind speed", "Winter



```
storm", "Wireless", "Woman with a Parasol - Madame Monet and Her Son"]  
=====
```

Top 3 words

```
String["painting", "weather", "air"]
```

```
In [29]: # For k = 5
centroids, labels, j_hist = Kmeans.kmeans(article_histograms, 5)
centroids, labels, j_hist_2 = Kmeans.kmeans(article_histograms, 5)
plot(j_hist)
plot(j_hist_2)
println(article_titles[labels .== 1])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[1], rev=true)][1:3]
println(top_3)
println("=====")
println(article_titles[labels .== 2])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[2], rev=true)][1:3]
println(top_3)
println("=====")
println(article_titles[labels .== 3])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[3], rev=true)][1:3]
println(top_3)
println("=====")
println(article_titles[labels .== 4])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[4], rev=true)][1:3]
println(top_3)
println("=====")
println(article_titles[labels .== 5])
println("=====")
println("Top 3 words")
top_3 = dictionary[sortperm(centroids[5], rev=true)][1:3]
println(top_3)
```

String["A Bar at the Folies-Bergère", "Alfred Sisley", "Armand Guillaumin", "Art Institute of Chicago", "A Sunday Afternoon on the Island of La Grande Jatte", "Bal du moulin de la Galette", "Berthe Morisot", "Café Guerbois", "Camille Pissarro", "Claude Monet", "Cubism", "Édouard Manet", "Dutch Golden Age painting", "Edgar Degas", "Effets de soir", "En plein air", "Eugène Boudin", "Fauvism", "Félix Fénéon", "Frédéric Bazille", "Georges Seurat", "Gustave Caillebotte", "Impasto", "Impressionism", "Impressionism in music", "Impression, Sunrise", "Jan Steen", "Johan Jongkind", "L'Absinthe", "Landscape painting", "Louis Leroy", "Luncheon on the Boating Party", "Macchiaioli", "Mary Cassatt", "Neo-impressionism", "Paris Street; Rainy Day", "Paul Cézanne", "Paul Signac", "Pierre-Auguste Renoir", "Portrait", "Post-Impressionism", "Salon des Refusés", "Still life", "The Child's Bath", "Wet-on-wet", "Woman with a Parasol - Madame Monet and Her Son"]

=====

Top 3 words

String["painting", "art", "paintings"]

String["Food and Agriculture Organization", "Headquarters of the United Nations", "International Civil Aviation Organization", "International Court of Justice", "International Fund for Agricultural Development", "International Labour Organization", "International Maritime Organization", "International Organization for Migration", "International Refugee Organisation", "International Seabed Authority", "International Telecommunication Union", "International Trade Centre", "International Tribunal for the Law of the Sea", "Joint United Nations Programme on HIV/AIDS", "League of Nations", "Office of the United Nations High Commissioner for Human Rights", "Organization for Security and Co-operation in Europe", "Organization of American States", "Outline of the United Nations", "Paris Peace Conference, 1919", "Secretary-General of the United Nations", "UNESCO", "UNICEF", "United Nations", "United Nations Capital Development Fund", "United Nations Charter", "United Nations Conference on Trade and Development", "United Nations Convention on the Law of the Sea", "United Nations Development Programme", "United Nations Economic and Social Council", "United Nations Environment Programme", "United Nations General Assembly", "United Nations High Commissioner for Refugees", "United Nations Human Settlements Programme", "United Nations Industrial Development Organization", "United Nations International Strategy for Disaster Reduction", "United Nations Office for Project Services", "United Nations Office on Drugs and Crime", "United Nations Population Fund", "United Nations Secretariat", "United Nations Security Council", "United Nations System", "United Nations System Staff College", "United Nations Trusteeship Council", "United Nations University", "United Nations Volunteers", "Universal Postal Union", "UN Women", "World Food Programme", "World Health Organization", "World Intellectual Property Organization", "World Meteorological Organization", "World Tourism Organization", "World Trade Organization"]

=====

Top 3 words

String["nations", "international", "council"]

String["Albedo", "Amplitude modulation", "Amplitude-shift keying", "Analog signal", "Anemometer", "Antenna (radio)", "Anticyclone", "Atmosphere", "Atmosphere of Earth", "Atmospheric pressure", "Attenuation", "Audio power amplifier", "Baroclinity", "Barograph", "Barometer", "Black ice", "Broadcasting", "Carrier signal", "Ceiling balloon", "Ceiling projector", "Ceilometer", "Channel (communications)", "Chaos theory", "Climate", "Cloud", "Communications satellite", "Complementary colors", "Contrast effect", "Coriolis effect", "Corona", "Dark adaptor goggles", "Dat

a transmission", "Dew point", "Digital signal", "Disdrometer", "Duplex (telecommunications)", "Dust storm", "Dynamic Host Configuration Protocol", "Effect of sun angle on climate", "Electrical telegraph", "Erosion", "Extratropical cyclone", "Fiber-optic communication", "Field mill", "Flag signals", "Flood", "Fog", "Freezing rain", "Frequency-division multiplexing", "Frequency modulation", "Frequency-shift keying", "Frontogenesis", "Frost", "Game Boy line", "GSM", "Guglielmo Marconi", "Hadley cell", "Handset", "Heliograph", "High-pressure area", "Hurricane Katrina", "Hygrometer", "Ice", "Ice accretion indicator", "Ice pellets", "Internet access", "Jet stream", "Kantō region", "Lake", "Lidar", "Lightning detection", "Little Ice Age", "Low-pressure area", "Mesoscale meteorology", "Meteorology", "Microscale meteorology", "Microwave transmission", "Middle latitudes", "Milankovitch cycles", "Mobile phone", "Modulation", "Monsoon", "Multiplexing", "Natural disaster", "Nephelometer", "Nephoscope", "Nintendo", "NTSC", "Numerical weather prediction", "Optical communication", "Optical fiber", "Phase-shift keying", "Plant", "Point-to-point (telecommunications)", "Precipitation", "Public switched telephone network", "Pyranometer", "Quadrature amplitude modulation", "Radio", "Radiosonde", "Radio wave", "Rain gauge", "Receiver (radio)", "Repeater", "Sea breeze", "Semaphore line", "Severe weather", "Smoke signal", "Solarimeter", "Solar wind", "Sounding rocket", "Space weather", "Stevenson screen", "Sunlight", "Sunshine recorder", "Synoptic scale meteorology", "Telecommunications network", "Telephone", "Telephone exchange", "Temperature", "Thermo-hygrograph", "Thermometer", "Thunderstorm", "Time-division multiplexing", "Tornado", "Transmission medium", "Transmitter", "Tropical cyclone", "Troposphere", "Typhoon", "Ultraviolet", "Voice over IP", "W-CDMA (UMTS)", "Weather balloon", "Weather forecasting", "Weather front", "Weather map", "Weather modification", "Weather radar", "Wide area network", "Wind chill", "Wind direction", "Windsock", "Wind speed", "Winter storm", "Wireless"]

=====

Top 3 words

String["weather", "signal", "wind"]

String["Brock (Pokémon)", "Bulbasaur", "Deoxys", "Eevee", "Gameplay of Pokémon", "Hey You, Pikachu!", "Lapras", "List of Pokémon characters", "Meowth", "Mew (Pokémon)", "Mewtwo", "Misty (Pokémon)", "Pikachu", "Pokémon", "Pokémon (anime)", "Pokémon Black and White", "Pokémon Diamond and Pearl", "Pokémon Emerald", "Pokémon FireRed and LeafGreen", "Pokémon Gold and Silver", "Pokémon HeartGold and SoulSilver", "Pokémon Omega Ruby and Alpha Sapphire", "Pokémon Platinum", "Pokémon Red and Blue", "Pokémon Ruby and Sapphire", "Pokémon Trading Card Game", "Pokémon universe", "Pokémon Yellow", "Pokémon 4Ever", "Satoshi Tajiri", "Team Rocket (anime)", "Togepi", "Zapdos"]

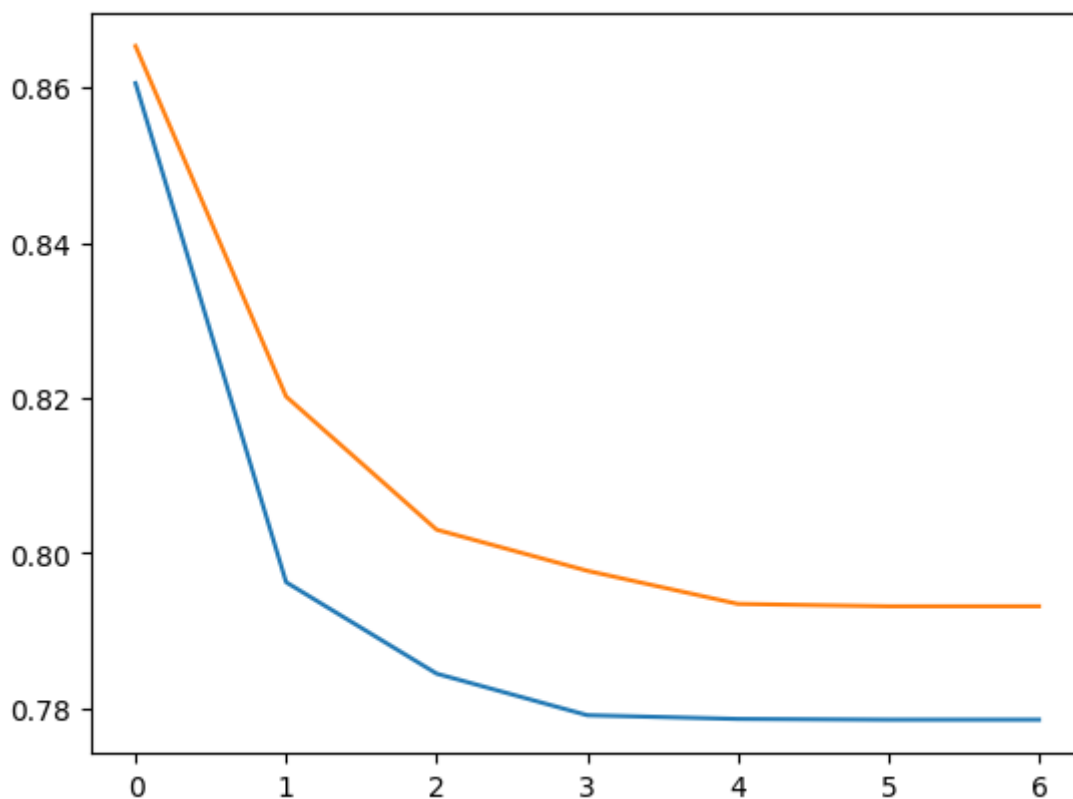
=====

Top 3 words

String["pokemon", "game", "games"]

String["Acid rain", "Asynchronous Transfer Mode", "Convention on the Rights of Persons with Disabilities", "Desertification", "Famine", "Greenhouse gas", "Human impact on the environment", "International Bank for Reconstruction and Development", "International Centre for Settlement of Investment Disputes", "International Development Association", "International Finance Corporation", "International Monetary Fund", "Internet protocol suite", "Multilateral Investment Guarantee Agency", "Multiprotocol Label Switching", "Organisation for the Prohibition of Chemical Weapons", "Telepresence", "United Nations Convention to Combat Desertification", "United Nations Framework Convention on Climate Change", "United

d Nations Relief and Works Agency for Palestine Refugees in the Near East", "World Bank G



roup"]

=====

Top 3 words

```
Out[29]: 3-element Array{String,1}:
  "bank"
  "convention"
  "parties"
```

```
In [25]: # For k = 10
centroids, labels, j_hist = Kmeans.kmeans(article_histograms, 10)
centroids, labels, j_hist_2 = Kmeans.kmeans(article_histograms, 10)
println(article_titles[labels .== 1])
println("=====")
println(article_titles[labels .== 2])
println("=====")
println(article_titles[labels .== 3])
println("=====")
println(article_titles[labels .== 4])
println("=====")
println(article_titles[labels .== 5])
println("=====")
println(article_titles[labels .== 6])
println("=====")
println(article_titles[labels .== 7])
println("=====")
println(article_titles[labels .== 8])
println("=====")
println(article_titles[labels .== 9])
println("=====")
println(article_titles[labels .== 10])
plot(j_hist)
plot(j_hist_2)
```

```
String["A Bar at the Folies-Bergère", "Alfred Sisley", "Armand Guillaum
in", "Art Institute of Chicago", "A Sunday Afternoon on the Island of L
a Grande Jatte", "Bal du moulin de la Galette", "Berthe Morisot", "Café
Guerbois", "Camille Pissarro", "Claude Monet", "Cubism", "Édouard Mane
t", "Dutch Golden Age painting", "Edgar Degas", "Effets de soir", "En p
lein air", "Eugène Boudin", "Fauvism", "Félix Fénéon", "Frédéric Bazill
e", "Georges Seurat", "Gustave Caillebotte", "Impasto", "Impressionis
m", "Impressionism in music", "Impression, Sunrise", "Jan Steen", "Joha
n Jongkind", "L'Absinthe", "Landscape painting", "Louis Leroy", "Lunche
on of the Boating Party", "Macchiaioli", "Mary Cassatt", "Neo-impressio
nism", "Paris Street; Rainy Day", "Paul Cézanne", "Paul Signac", "Pierr
e-Auguste Renoir", "Portrait", "Post-Impressionism", "Salon des Refusé
s", "Still life", "The Child's Bath", "Wet-on-wet", "Woman with a Paras
ol - Madame Monet and Her Son"]
```

```
=====
```

```
String["Albedo", "Attenuation", "Climate", "Complementary colors", "Cor
iolis effect", "Corona", "Dark adaptor goggles", "Dust storm", "Effect
of sun angle on climate", "Electrical telegraph", "Famine", "Field mil
l", "Flag signals", "Game Boy line", "Guglielmo Marconi", "Heliograph",
"Hurricane Katrina", "Jet stream", "Kantō region", "Lightning detectio
n", "Milankovitch cycles", "Monsoon", "Nephelometer", "Nintendo", "Poké
mon Omega Ruby and Alpha Sapphire", "Pyranometer", "Semaphore line", "S
moke signal", "Solarimeter", "Solar wind", "Sounding rocket", "Stevenso
n screen", "Sunlight", "Sunshine recorder", "Thermometer", "Ultraviole
t", "United Nations Relief and Works Agency for Palestine Refugees in t
he Near East"]
```

```
=====
```

```
String["Food and Agriculture Organization", "Headquarters of the United
Nations", "International Civil Aviation Organization", "International C
ourt of Justice", "International Fund for Agricultural Development", "I
nternational Labour Organization", "International Maritime Organizatio
n", "International Refugee Organisation", "International Seabed Authori
ty", "International Telecommunication Union", "International Trade Cent
re", "International Tribunal for the Law of the Sea", "Joint United Nat
ions Programme on HIV/AIDS", "League of Nations", "Office of the United
Nations High Commissioner for Human Rights", "Organization for Security
and Co-operation in Europe", "Organization of American States", "Outlin
e of the United Nations", "Paris Peace Conference, 1919", "Secretary-Ge
neral of the United Nations", "UNESCO", "UNICEF", "United Nations", "Un
ited Nations Capital Development Fund", "United Nations Charter", "Unit
ed Nations Conference on Trade and Development", "United Nations Conven
tion on the Law of the Sea", "United Nations Development Programme", "U
nited Nations Economic and Social Council", "United Nations Environment
Programme", "United Nations General Assembly", "United Nations High Com
missioner for Refugees", "United Nations Human Settlements Programme",
"United Nations Industrial Development Organization", "United Nations I
nternational Strategy for Disaster Reduction", "United Nations Office f
or Project Services", "United Nations Office on Drugs and Crime", "Unit
ed Nations Population Fund", "United Nations Secretariat", "United Nati
ons Security Council", "United Nations System", "United Nations System
Staff College", "United Nations Trusteeship Council", "United Nations U
niversity", "United Nations Volunteers", "Universal Postal Union", "UN
Women", "World Food Programme", "World Health Organization", "World Int
ellectual Property Organization", "World Meteorological Organization",
"World Tourism Organization", "World Trade Organization"]
```

```
=====
```

```
String["Convention on the Rights of Persons with Disabilities", "Intern
```

ational Bank for Reconstruction and Development", "International Centre for Settlement of Investment Disputes", "International Development Association", "International Finance Corporation", "International Monetary Fund", "Multilateral Investment Guarantee Agency", "Organisation for the Prohibition of Chemical Weapons", "United Nations Convention to Combat Desertification", "United Nations Framework Convention on Climate Change", "World Bank Group"]

=====

String["Acid rain", "Black ice", "Contrast effect", "Desertification", "Erosion", "Flood", "Fog", "Freezing rain", "Frost", "Greenhouse gas", "Human impact on the environment", "Hygrometer", "Ice", "Ice accretion indicator", "Ice pellets", "Lake", "Lidar", "Little Ice Age", "Plant", "Precipitation", "Rain gauge", "Sea breeze", "Wind direction", "Windsoc k", "Winter storm"]

=====

String["Amplitude modulation", "Amplitude-shift keying", "Analog signal", "Carrier signal", "Channel (communications)", "Data transmission", "Digital signal", "Duplex (telecommunications)", "Frequency-division multiplexing", "Frequency modulation", "Frequency-shift keying", "Modulation", "Multiplexing", "Phase-shift keying", "Public switched telephone network", "Quadrature amplitude modulation", "Telephone", "Telephone exchange", "Time-division multiplexing"]

=====

String["Brock (Pokémon)", "Bulbasaur", "Deoxys", "Eevee", "Gameplay of Pokémon", "Hey You, Pikachu!", "Lapras", "List of Pokémon characters", "Meowth", "Mew (Pokémon)", "Mewtwo", "Misty (Pokémon)", "Pikachu", "Pokémon", "Pokémon (anime)", "Pokémon Black and White", "Pokémon Diamond and Pearl", "Pokémon Emerald", "Pokémon FireRed and LeafGreen", "Pokémon Gold and Silver", "Pokémon HeartGold and SoulSilver", "Pokémon Platinum", "Pokémon Red and Blue", "Pokémon Ruby and Sapphire", "Pokémon Trading Card Game", "Pokémon universe", "Pokémon Yellow", "Pokémon 4Ever", "Satoshi Tajiri", "Team Rocket (anime)", "Togepi", "Zapdos"]

=====

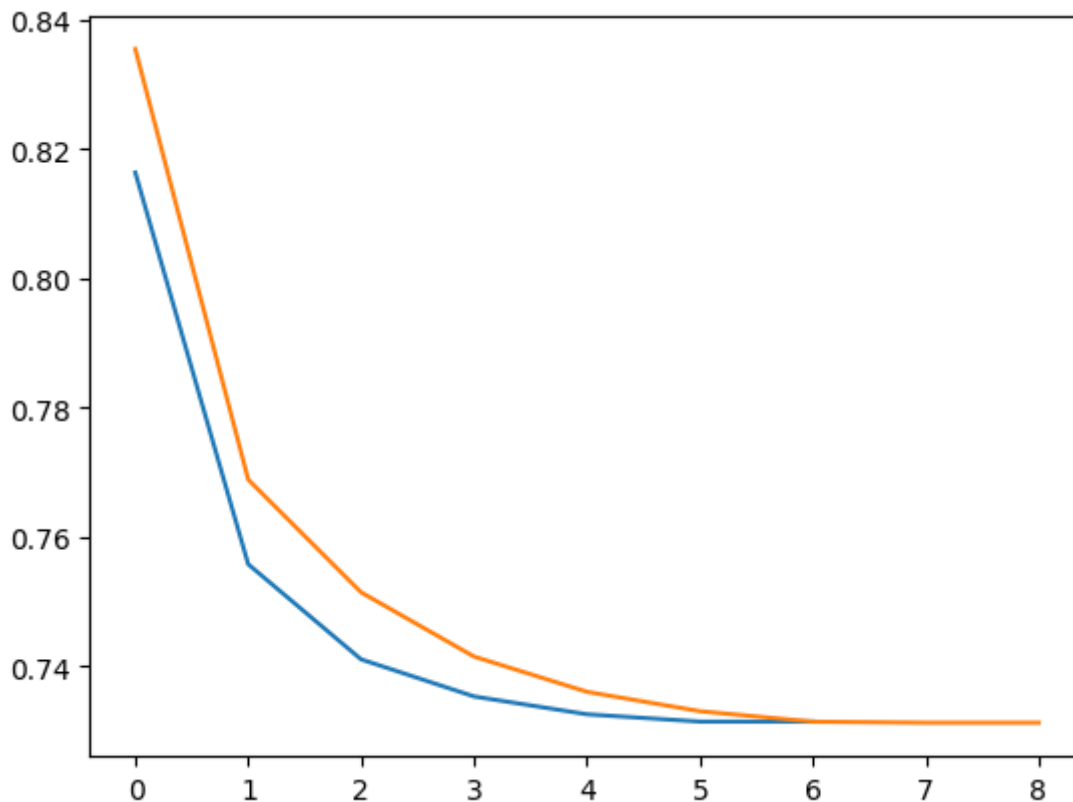
String["Antenna (radio)", "Audio power amplifier", "Broadcasting", "Communications satellite", "Fiber-optic communication", "Handset", "Microwave transmission", "NTSC", "Optical communication", "Optical fiber", "Point-to-point (telecommunications)", "Radio", "Radio wave", "Receiver (radio)", "Repeater", "Telepresence", "Transmission medium", "Transmitter", "Wireless"]

=====

String["Asynchronous Transfer Mode", "Ceiling balloon", "Ceiling projector", "Ceilometer", "Cloud", "Disdrometer", "Dynamic Host Configuration Protocol", "GSM", "International Organization for Migration", "Internet access", "Internet protocol suite", "Mobile phone", "Multiprotocol Label Switching", "Nephoscope", "Telecommunications network", "Voice over IP", "W-CDMA (UMTS)", "Wide area network"]

=====

String["Anemometer", "Anticyclone", "Atmosphere", "Atmosphere of Earth", "Atmospheric pressure", "Baroclinity", "Barograph", "Barometer", "Chaos theory", "Dew point", "Extratropical cyclone", "Frontogenesis", "Hadley cell", "High-pressure area", "Low-pressure area", "Mesoscale meteorology", "Meteorology", "Microscale meteorology", "Middle latitudes", "Natural disaster", "Numerical weather prediction", "Radiosonde", "Severe weather", "Space weather", "Synoptic scale meteorology", "Temperature", "Thermo-



hygrograph", "Thunderstorm", "Tornado", "Tropical cyclone", "Troposphere", "Typhoon", "Weather balloon", "Weather forecasting", "Weather front", "Weather map", "Weather modification", "Weather radar", "Wind chill", "Wind speed"]

Out[25]: 1-element Array{PyCall.PyObject,1}:
PyObject <matplotlib.lines.Line2D object at 0x7fac8a1f2e50>

```
In [ ]: #=
4.2 a)
    In each iteration (X-Axis), the value of j(Y-Axis) keep decreasing, it
    each iteration it keeps getting better
    unless the difference between each iteration does not change. The fact t
    hat j keeps decreasing denotes that
    KMeans is converging in a finite number of steps.

    Based on different number of runs, the graph can vary as the initial cen
    troid is chosen randomly and might
    take different number for convergence
    =#
```

```
In [ ]: #=  
4.2 b)  
For k=2, the grouping is under clustered, as one may find articles from  
multiple topic under one cluster.  
For example:  
In above run for k = 2, `World Trade Organization`, `Digital signal` and  
`Monsoon` are classified under one  
single category which could be further classified to achieve a much gran  
ular cluster.  
  
k=5 is perfectly clustered, upon inspection it is clear that article are  
classified into following categories  
i) Organistaion and Unions  
ii) Arts and Painting  
iii) Weather and tools related to weather  
iv) Signals  
v) Pokemon  
  
Therefore, k = 5 is the perfect for clustering the given set of article  
s. It can be also verified that corpus  
was picked from set of 5 articles hence clustering into group of 5 is pe  
rfect.  
  
k=10 is the case of over clustering/categorization, topics such as Pokem  
on (in k = 5) is further broken down  
into multiple groups. This is not the most optimal solution.  
  
Based on the observation of above data, it is clear that k = 5.  
=#
```