ngram_freq

Joemari Pulido

2022-09-22

##load necessary packages and install ngram from github repository

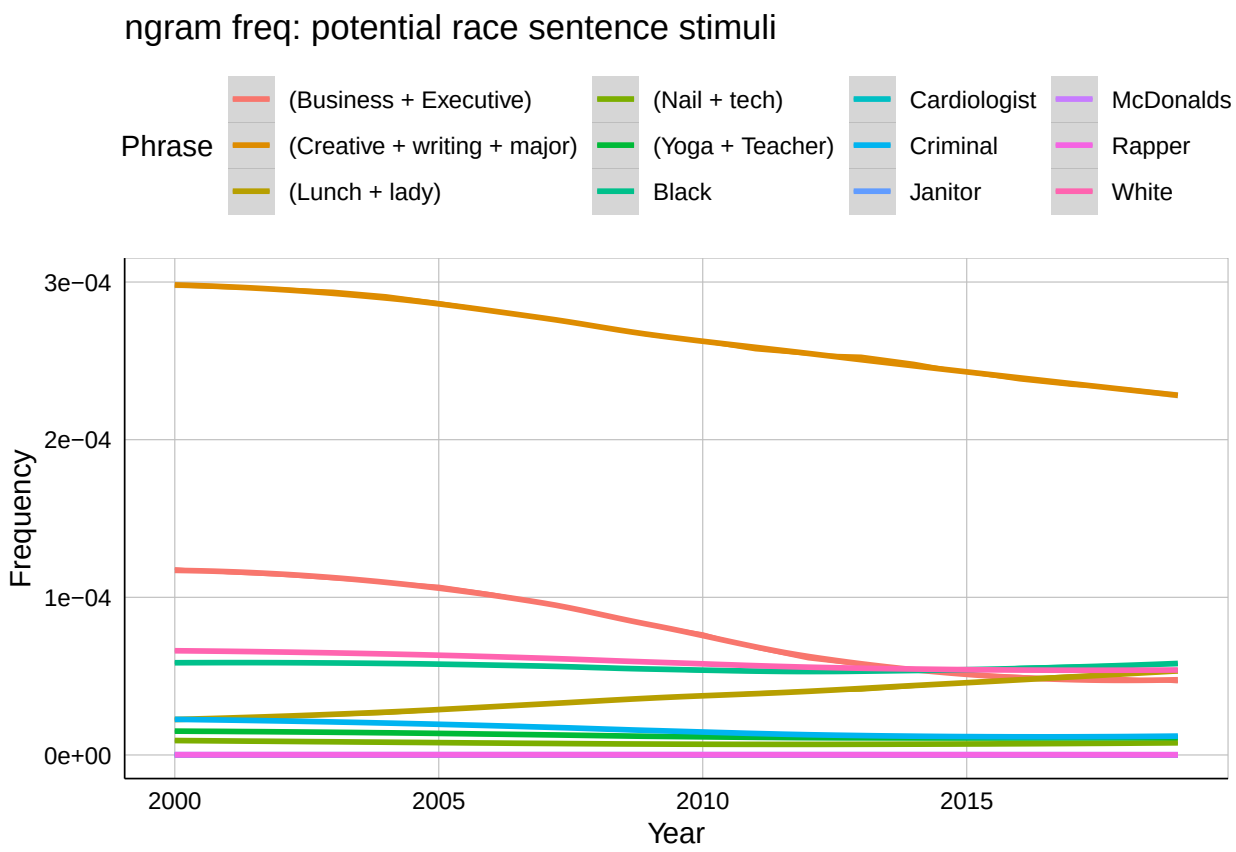
Sean Carmody et al. created a package 'ngramr' which queries the Google Books Ngram Viewer. The Google Books Ngram Viewer corpus holds about 2 trillion words/phrases. The following code block loads the package 'devtools'. Then, calls the function 'install_github' which enables one to install a repository directly from GitHub into RStudio. syntax is 'install_github("respositoryownner/repositoryname")'. Then, once the repository is loaded, load the 'ngramr' package into RStudio, which then allows you to enter a list of phrases to display a graph showing how often the phrases occurred in a corpus of books.

```
library(devtools)
install_github("seancarmody/ngramr")
library(ngramr)
```

##Race:S01

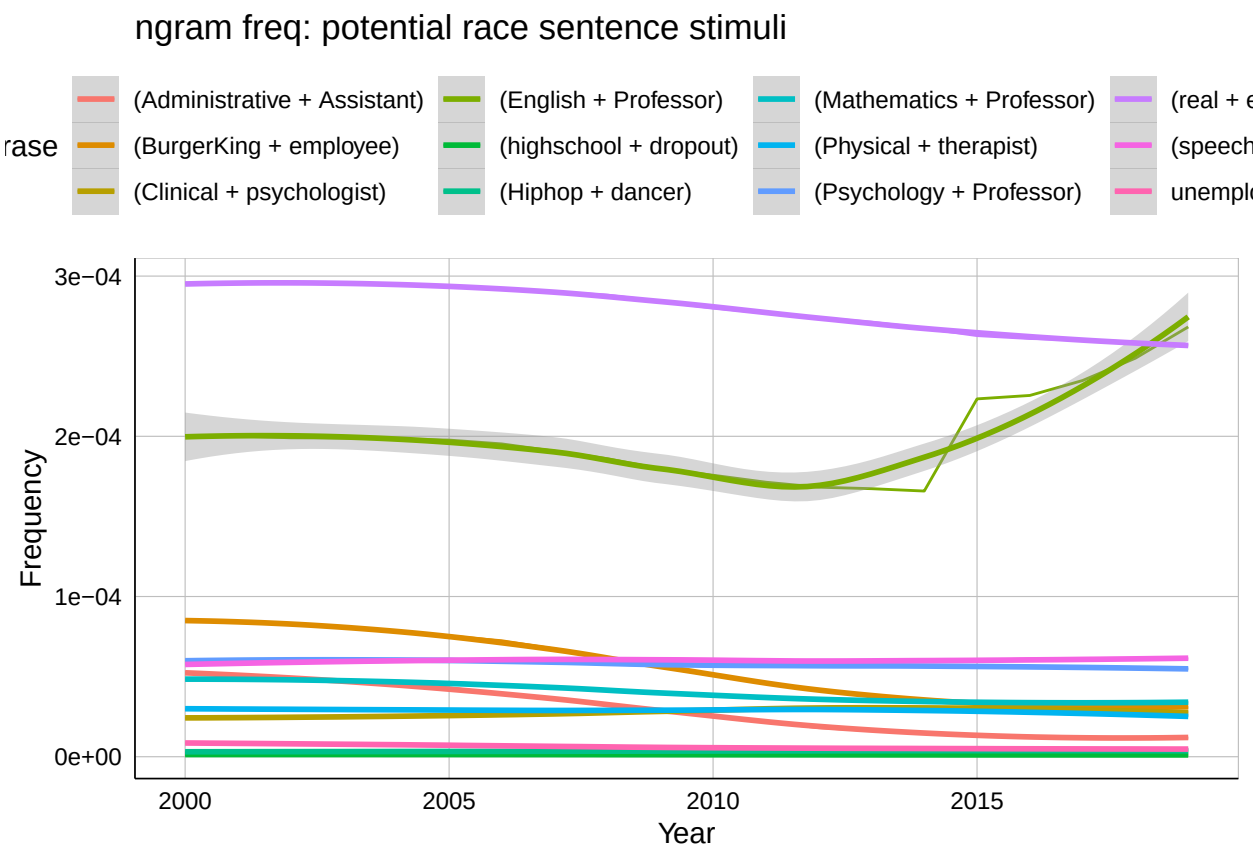
The following code blocks uses 'ngram' to load in data from Google Books Ngram Viewer. The programmer can choose which words/phrases to load in. For the purposes of my independent study, I inputed words/phrases that could possibly elict race and gender stereotypes. The ggplot and ggplot2 packages then allow the programmer to plot the frequencies in a graph.

```
library(ggplot2)
ng01  <- ngram(c("Janitor", "Criminal", "Black", "White", "Yoga+Teacher", "Business+Executive","McDonal
  case_ins = FALSE,
  aggregate = FALSE,
  count = FALSE,
  drop_corpus = FALSE,
  drop_parent = FALSE,
  drop_all = FALSE,
  type = FALSE)
ngram01 <- ggplot(ng01, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth() + theme_google()
ngram01 + labs(x = "Year", y = "Frequency", title = "ngram freq: potential race sentence stimuli")
```



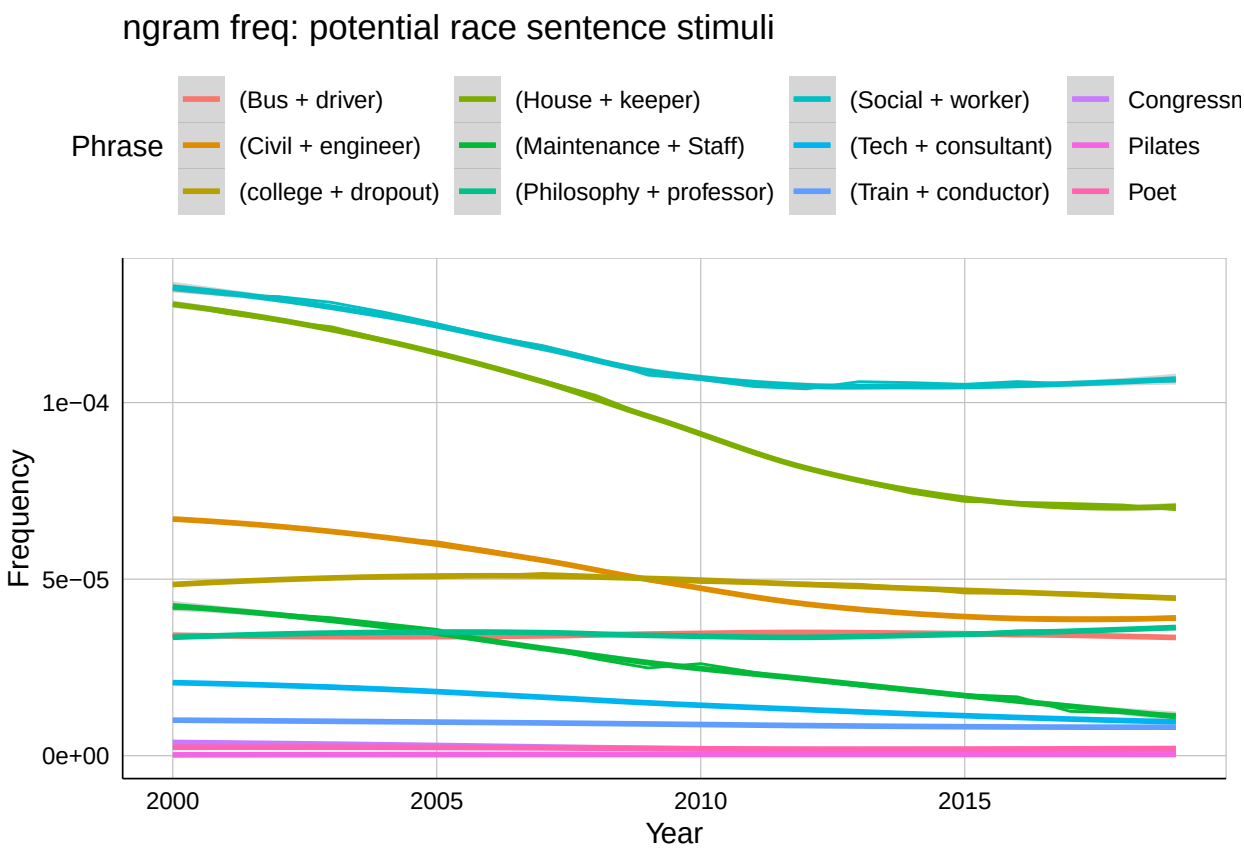
as #Race:S02

```
library(ggplot2)
ng02  <- ngram(c("Clinical+psychologist", "Hiphop+dancer", "Physical+therapist", "unemployed", "BurgerK
ngram02 <- ggplot(ng02, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram02 + labs(x = "Year", y = "Frequency", title = "ngram freq: potential race sentence stimuli")
```



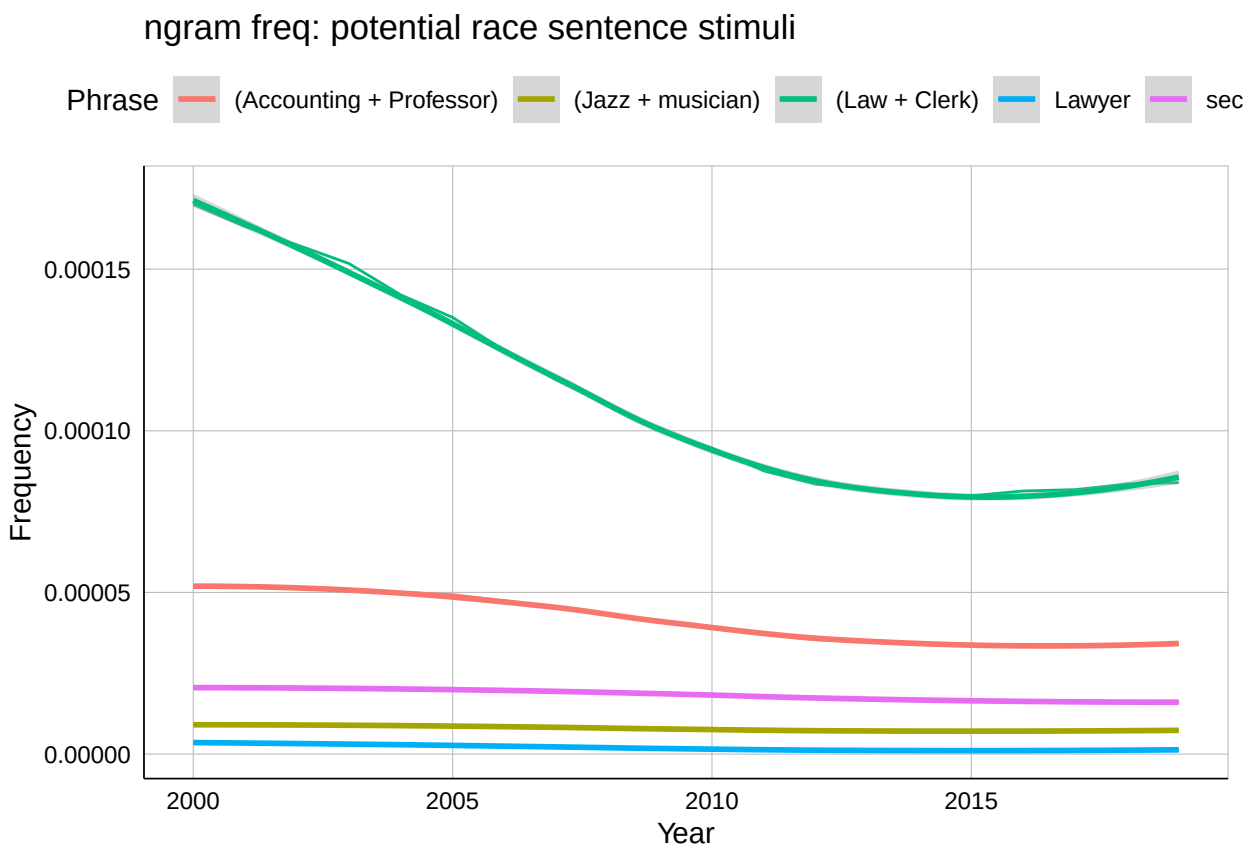
#Race:S03

```
library(ggplot2)
ng03  <- ngram(c("Bus+driver", "Tech+consultant", "Pilates", "Poet", "Congressman", "Train+conductor",
ngram03 <- ggplot(ng03, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram03 + labs(x = "Year", y = "Frequency", title = "ngram freq: potential race sentence stimuli")
```



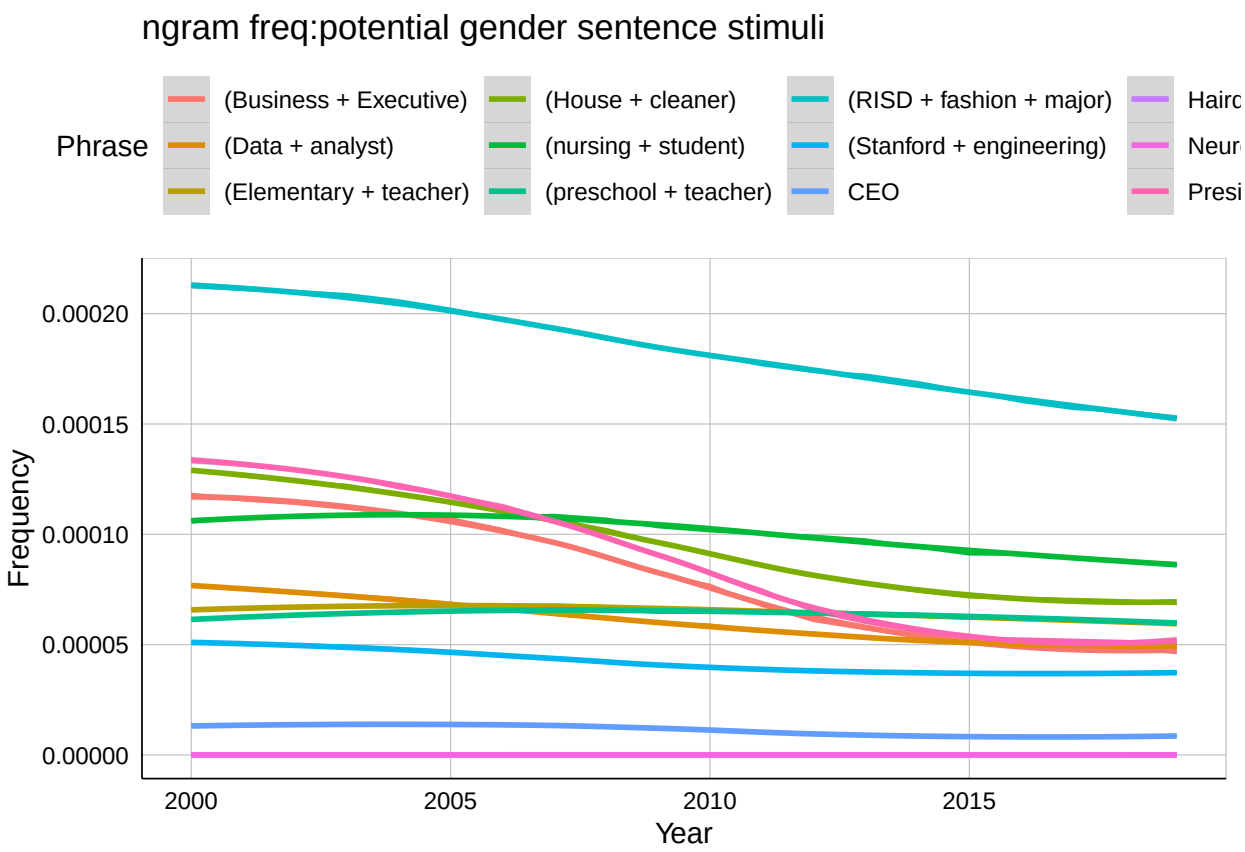
#Race:S04

```
library(ggplot2)
ng04  <- ngram(c("secretary", "Lawyer", "Law+Clerk", "Accounting+Professor", "Jazz+musician"), year_sta
ngram04 <- ggplot(ng04, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram04 + labs(x = "Year", y = "Frequency", title = "ngram freq: potential race sentence stimuli")
```



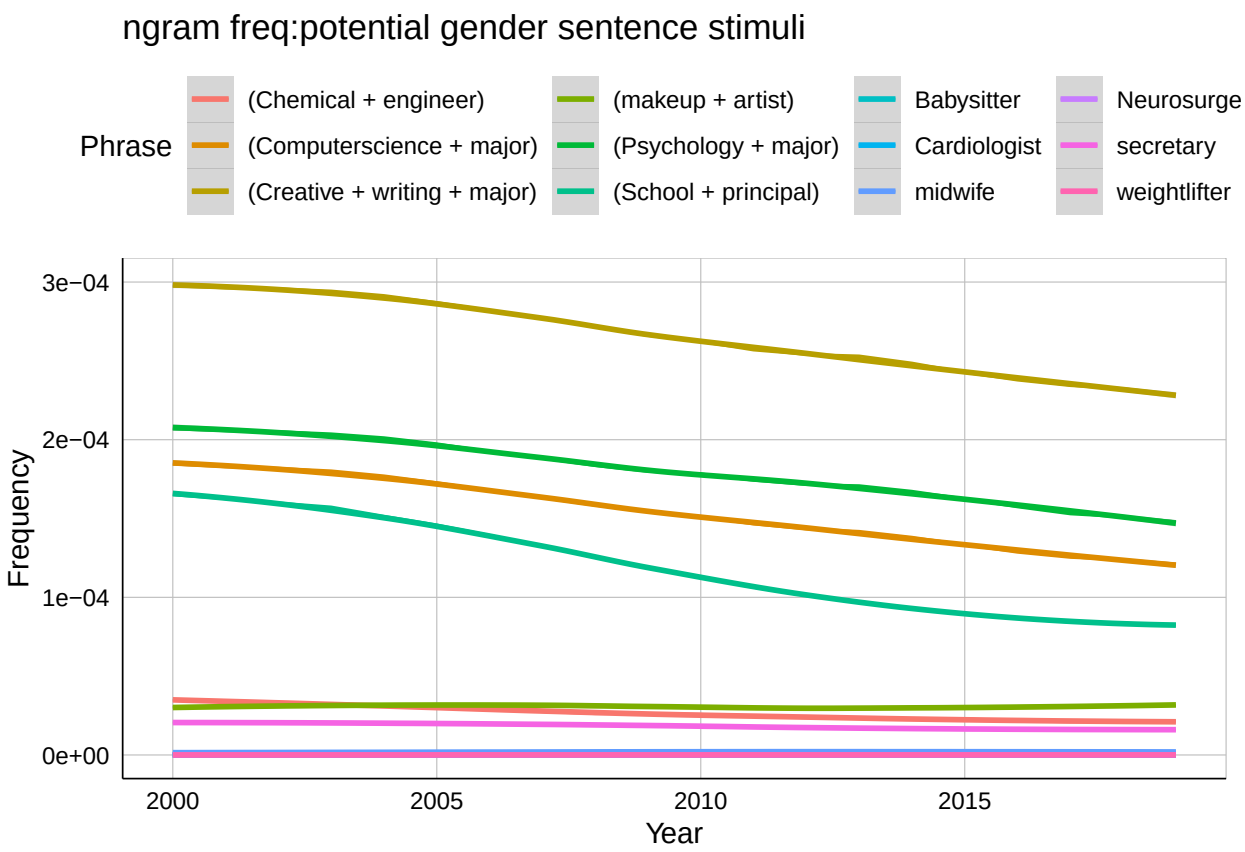
#Gender:S01 The following code chunk plots words or phrases that either match a gender stereotype or are a gender stereotype mismatch.

```
library(ggplot2)
ng05  <- ngram(c("Stanford+engineering", "Hairdresser","CEO", "House+cleaner", "Data+analyst", "Elementa
ngram05 <- ggplot(ng05, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth() + theme_google()
ngram05 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender sentence stimuli")
```



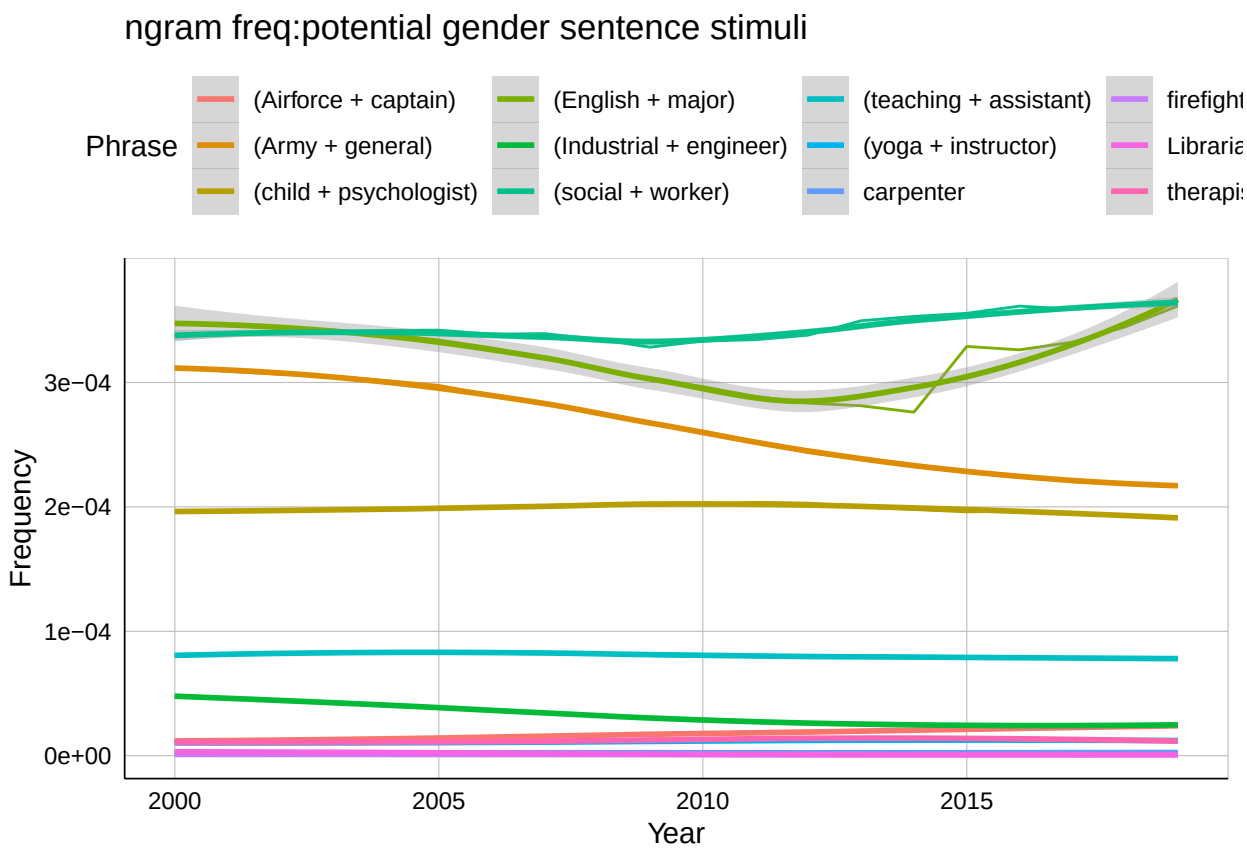
#Gender:S02 The following code chunk plots words or phrases that either match a gender stereotype or are a gender stereotype mismatch.

```
library(ggplot2)
ng06  <- ngram(c("Chemical+engineer", "makeup+artist", "Cardiologist", "Babysitter", "Psychology+major"
ngram06 <- ggplot(ng06, aes(x=Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram06 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender sentence stimuli")
```



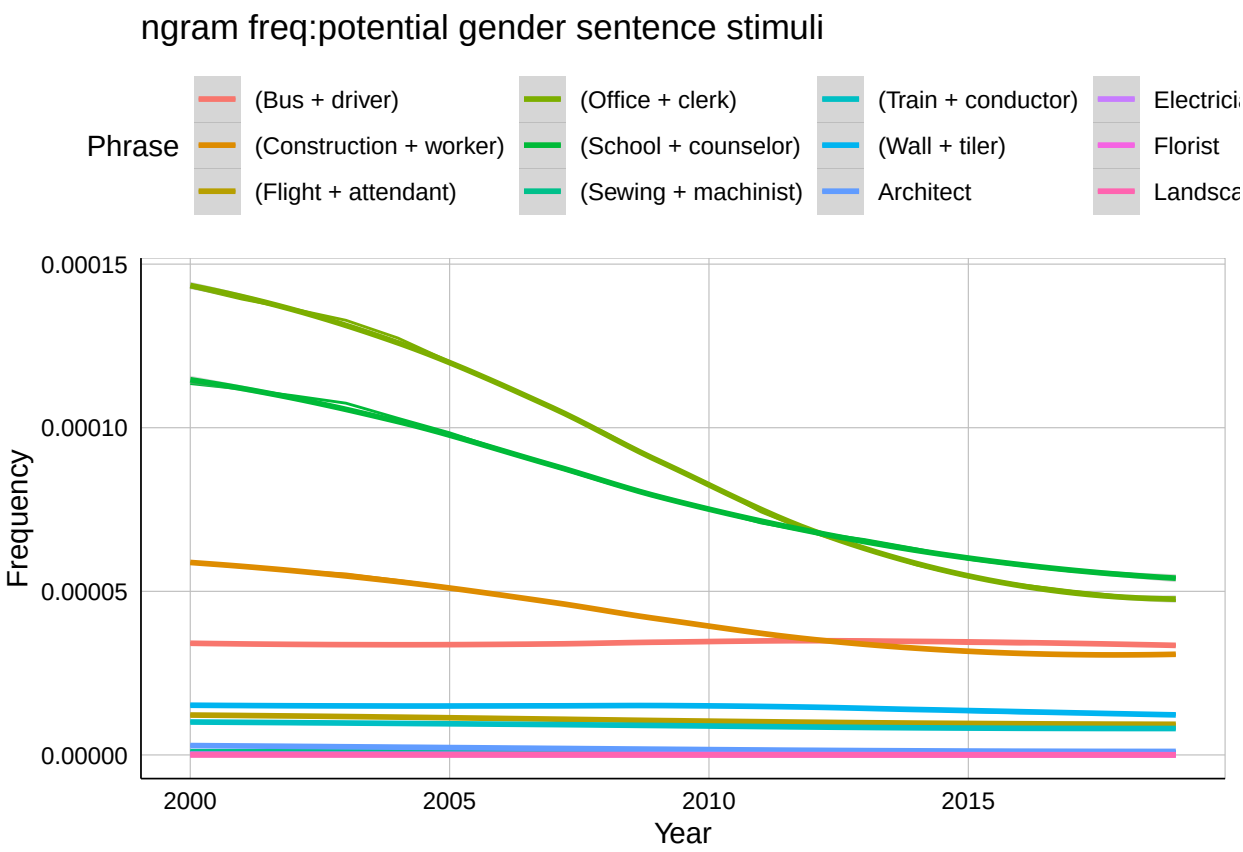
#Gender:S03 The following code chunk plots words or phrases that either match a gender stereotype or are a gender stereotype mismatch.

```
library(ggplot2)
ng07  <- ngram(c("Army+general", "English+major", "Industrial+engineer", "Librarian", "Airforce+captain
ngram07 <- ggplot(ng07, aes(x=Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram07 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender sentence stimuli")
```



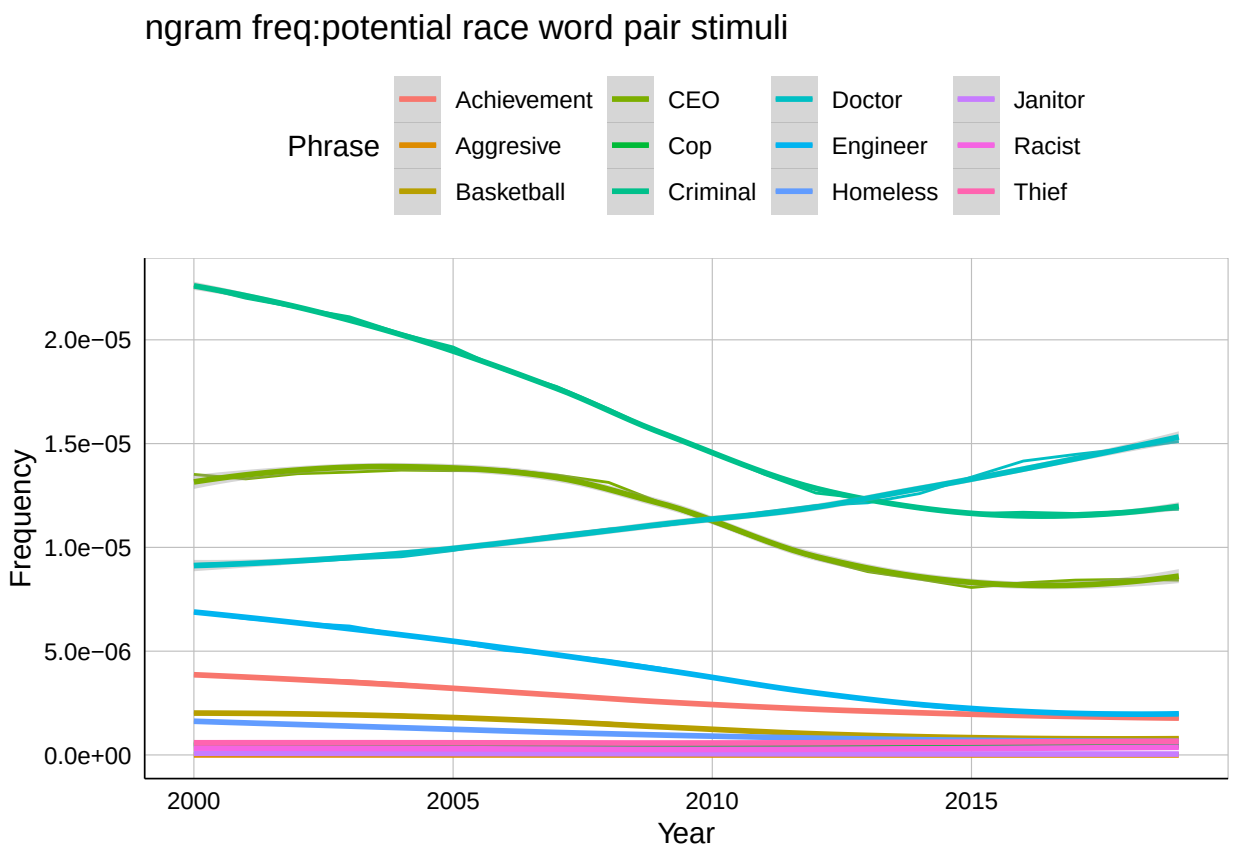
#Gender:S04 The following code chunk plots words or phrases that either match a gender stereotype or are a gender stereotype mismatch.

```
library(ggplot2)
ng08  <- ngram(c("Architect", "Bus+driver", "Flight+attendant", "Train+conductor", "School+counselor",
ngram08 <- ggplot(ng08, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram08 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender sentence stimuli")
```



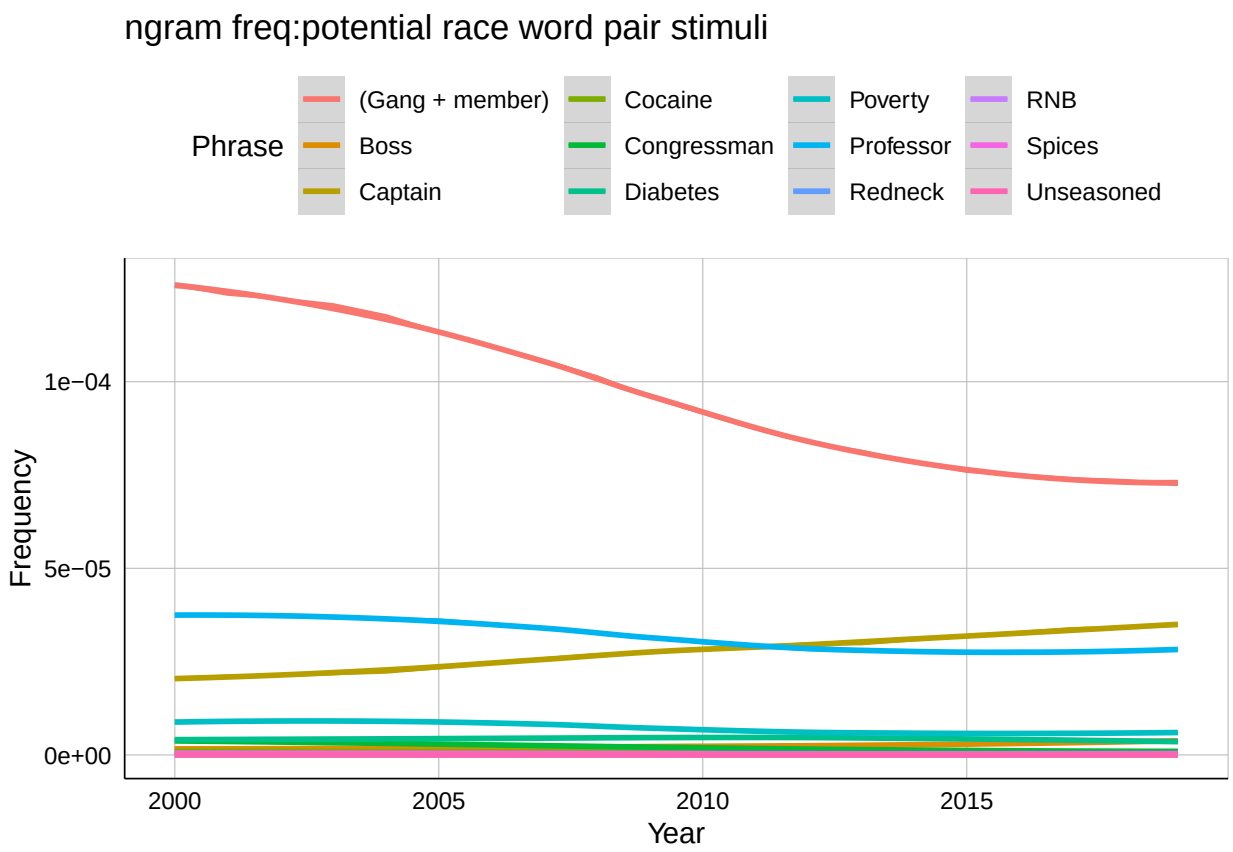
#Race:WP01 The following code chunk plots the frequencies of words or phrases that either are match a racial stereotpye or are a racial stereotype mismatch.

```
library(ggplot2)
ng09  <- ngram(c("Engineer", "Basketball", "CEO", "Criminal", "Achievement", "Homeless", "Doctor", "Jani
ngram09 <- ggplot(ng09, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram09 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential race word pair stimuli")
```



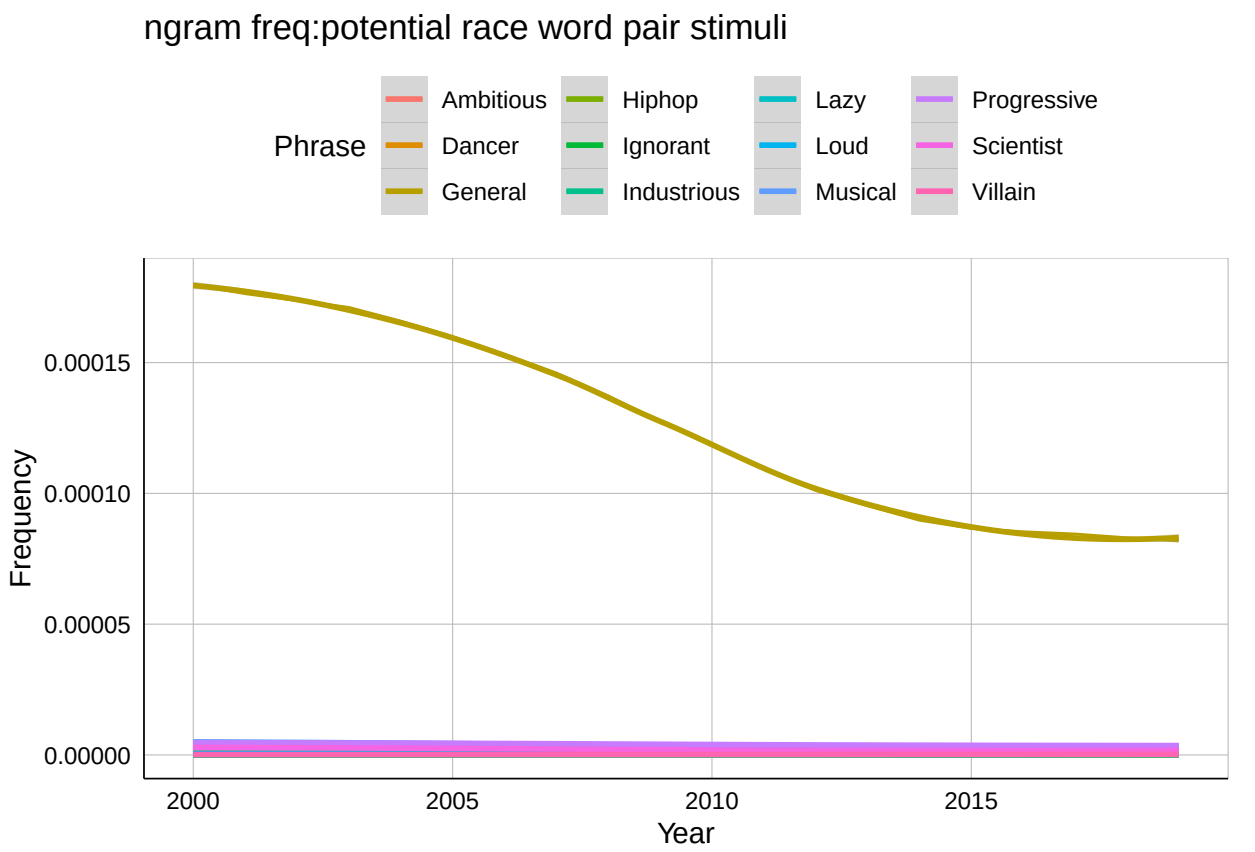
#Race:WP02 The following code chunk plots the frequencies of words or phrases that either are match a racial stereotpye or are a racial stereotype mismatch.

```
library(ggplot2)
ng10  <- ngram(c("Boss", "Poverty", "Redneck", "Diabetes", "Professor", "Cocaine", "Unseasoned", "Spices
ngram10 <- ggplot(ng10, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram10 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential race word pair stimuli")
```



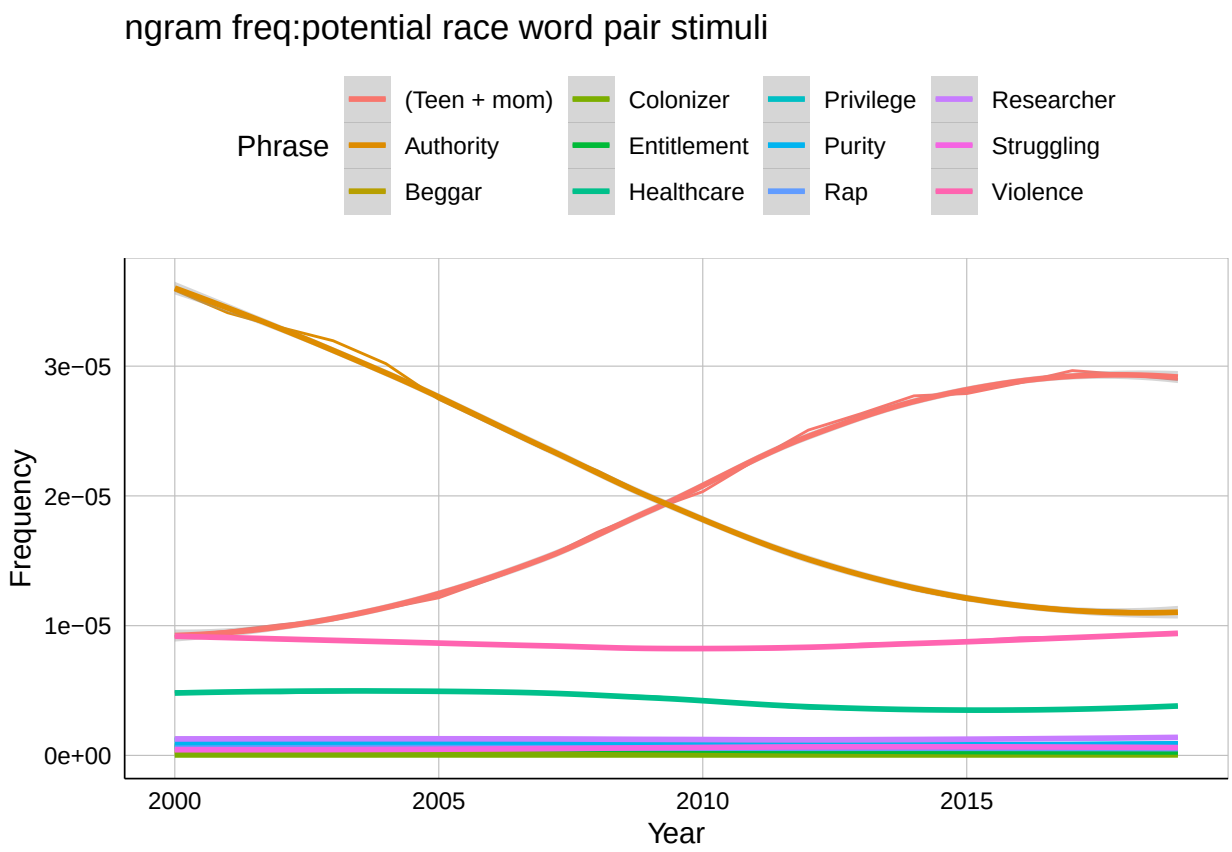
#Race:WP03 The following code chunk plots the frequencies of words or phrases that either are match a racial stereotpye or are a racial stereotype mismatch.

```
library(ggplot2)
ng11  <- ngram(c("Scientist", "Lazy", "General", "Villain", "Industrious", "Ignorant", "Ambitious", "Mus
ngram11 <- ggplot(ng11, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth() + theme_google()
ngram11 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential race word pair stimuli")
```



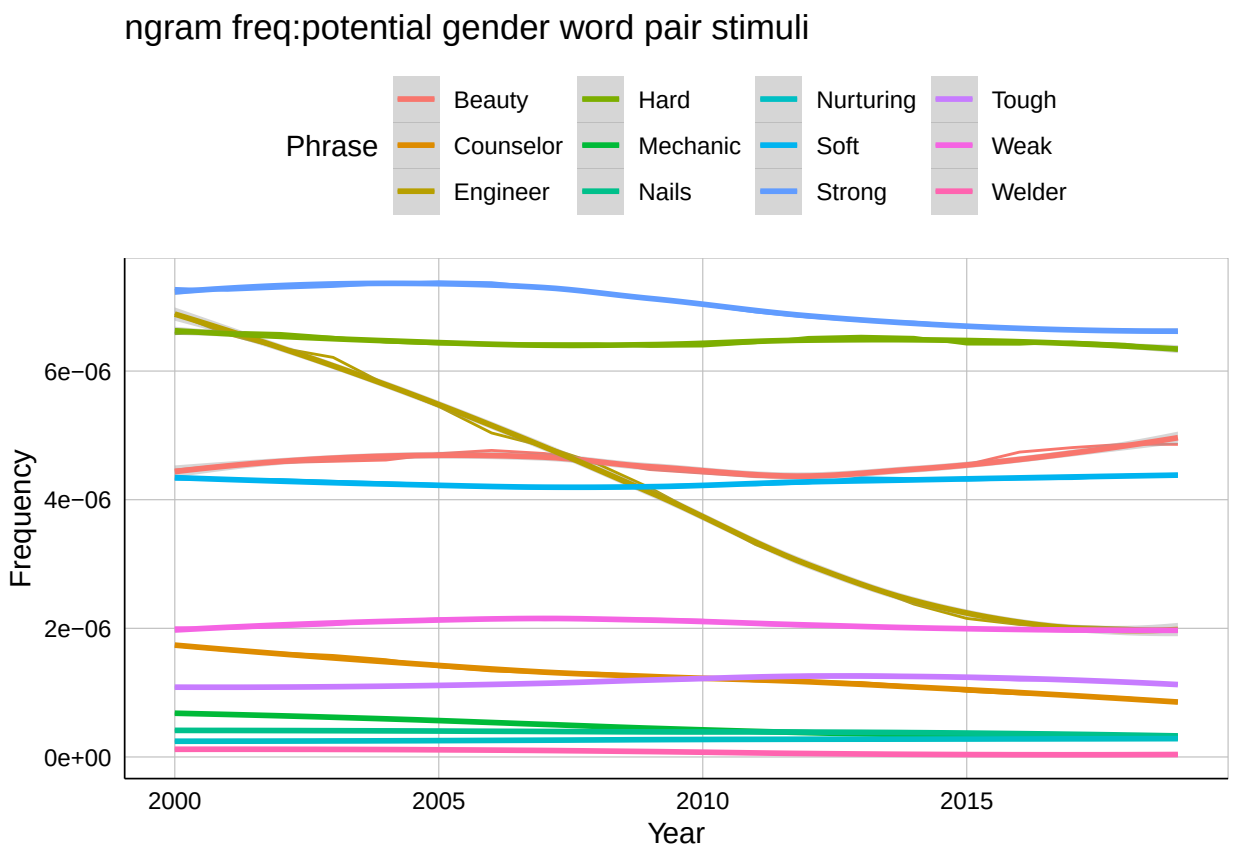
#Race:WP04 The following code chunk plots the frequencies of words or phrases that either are match a racial stereotpye or are a racial stereotype mismatch.

```
library(ggplot2)
ng12  <- ngram(c("Authority", "Rap", "Privilege", "Violence", "Entitlement", "Healthcare", "Researcher"
ngram12 <- ggplot(ng12, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram12 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential race word pair stimuli")
```



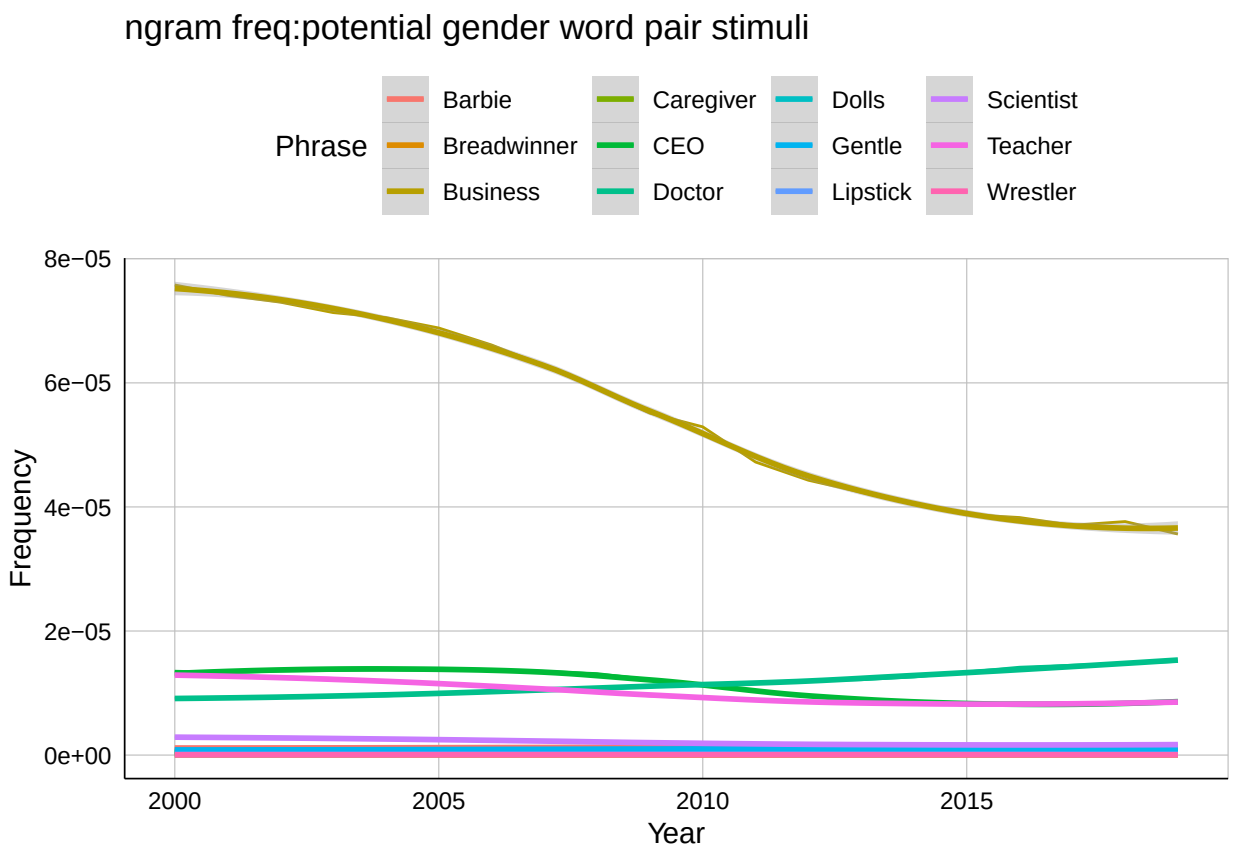
#Gender:WP01

```
library(ggplot2)
ng13  <- ngram(c("Nurturing", "Tough", "Beauty", "Engineer", "Nails", "Mechanic", "Counselor", "Welder"
ngram13 <- ggplot(ng13, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram13 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender word pair stimuli")
```



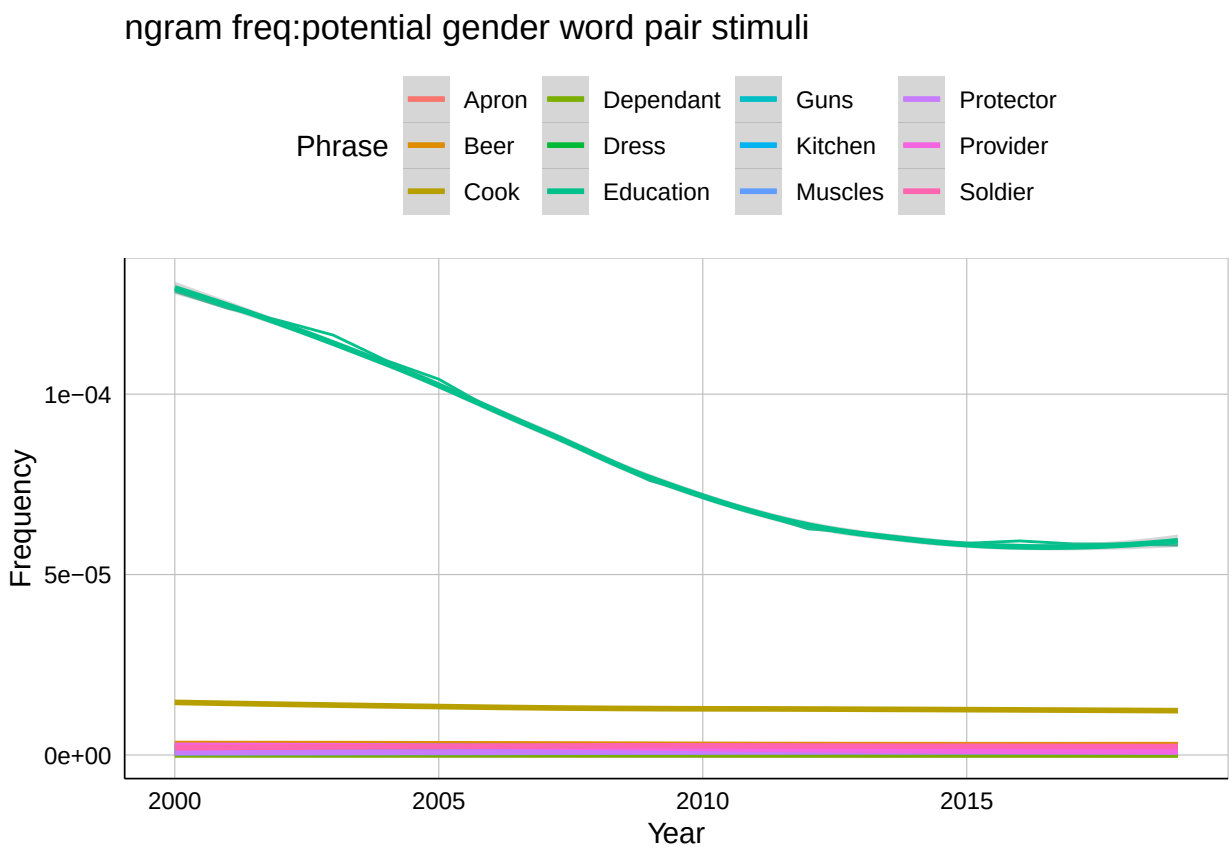
#Gender:WP02

```
library(ggplot2)
ng14  <- ngram(c("Gentle", "CEO", "Teacher", "Doctor", "Lipstick", "Scientist", "Caregiver", "Breadwinn
ngram14 <- ggplot(ng14, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram14 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender word pair stimuli")
```



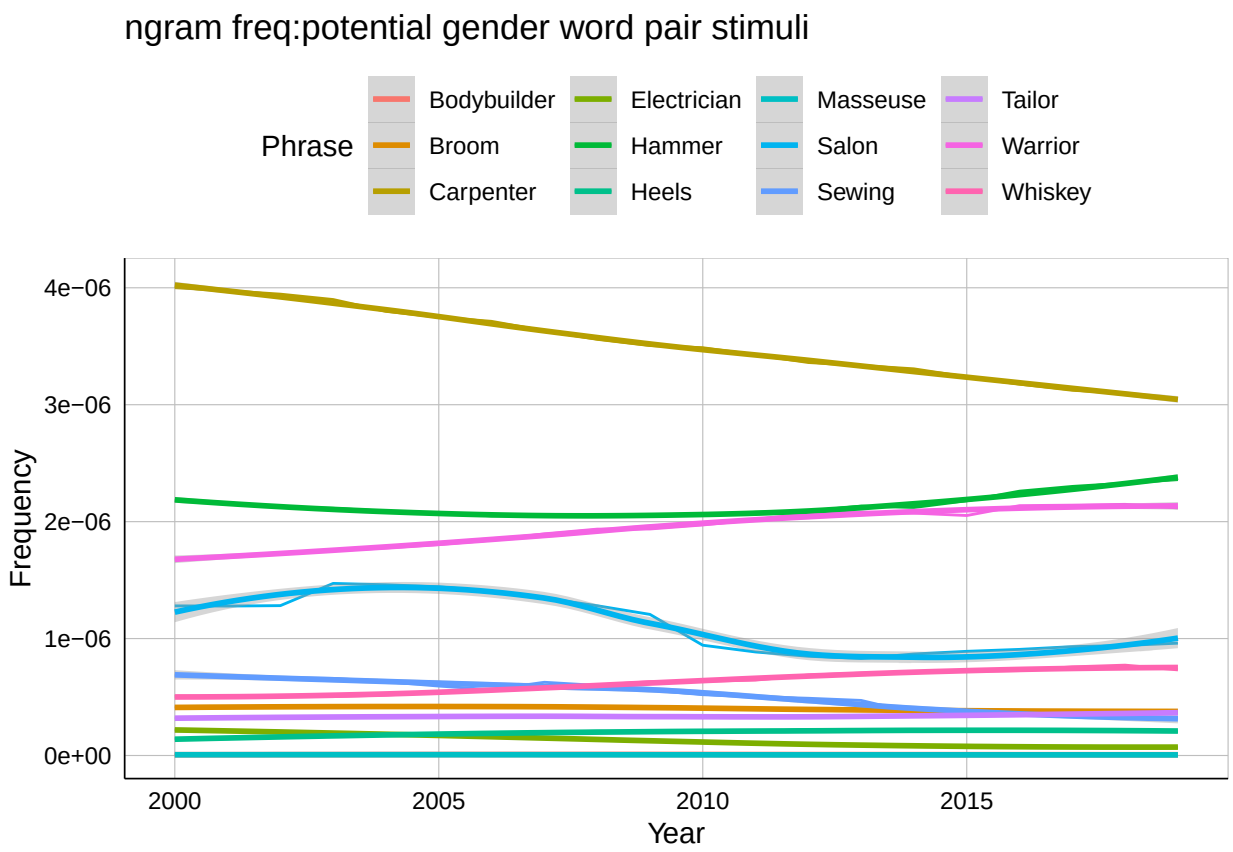
#Gender:WP03

```
library(ggplot2)
ng15  <- ngram(c("Cook", "Beer", "Kitchen", "Guns", "Apron", "Protector", "Dependant", "Provider", "Edu
ngram15 <- ggplot(ng15, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram15 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender word pair stimuli")
```



#Gender:WP04

```
library(ggplot2)
ng16  <- ngram(c("Heels", "Warrior", "Salon", "Carpenter", "Masseuse", "Whiskey", "Tailor", "Hammer", "
ngram16 <- ggplot(ng16, aes(x = Year, y = Frequency, colour = Phrase)) +
  geom_line() + geom_smooth()+ theme_google()
ngram16 + labs(x = "Year", y = "Frequency", title = "ngram freq:potential gender word pair stimuli")
```


ngram freq:potential gender word pair stimuli
Phrase
Bodybuilder
Broom
Carpenter
Electrician
Hammer
Heels
Masseuse
Salon
Sewing
Tailor
Warrior
Whiskey
4e−06
3e−06
2e−06
1e−06
0e+00
Frequency
2000
2005
2010
2015
Year